

ÉCOLE NATIONALE DES SCIENCES GEOGRAPHIQUES

COURS AU MASTÈRE SPÉCIALISÉ DÉSIGÉO

Introduction à la Géostatistique

Variographie, krigage, interpolation et simulation

Yann Méneroux

Année scolaire 2018-2019

Contact : [yann.meneroux\(at\)ign.fr](mailto:yann.meneroux(at)ign.fr)

Laboratoire en Sciences et Technologies de l'Information Géographique (LaSTIG)
Institut National de l'Information Géographique et Forestière (IGN)

Abstract

Ce document est un support de cours destiné aux étudiants du Mastère spécialisé Décision et Système d'Information Géolocalisée (Désigéo) de l'École Nationale des Sciences Géographiques. Il permet une introduction autodidacte à la Géostatistique linéaire, appliquée plus particulièrement au domaine de la Géomatique. Il comprend de nombreux exemples, des travaux dirigés à réaliser sur papier ainsi que des travaux pratiques sur machine (dans la langage de programmation R).

Les ressources nécessaires (jeux de données, code informatique...) peuvent être téléchargées à l'adresse suivante, dans la section *Mathématiques → Géostatistique* :

<http://cours-fad-public.ensg.eu/>

Notations

En règle générale, on note en majuscule les variables aléatoires et en minuscule les réalisations correspondantes. Ainsi, par exemple si X est une variable aléatoire distribuée suivant une loi normale, on note x_1, x_2, \dots, x_n des réalisations de X et :

$$m = \frac{1}{n} \sum_{i=1}^n x_i$$

est la moyenne empirique des n réalisations. Lorsqu'on souhaite étudier les propriétés statistiques de cette moyenne, on remplace les minuscules par des majuscules :

$$M = \frac{1}{n} \sum_{i=1}^n X_i$$

et M devient une variable aléatoire dont la loi dépend de celles des X_i .

Dans le problème modèle, on considère une fonction $z : \mathcal{D} \subset \mathbb{R}^2 \rightarrow \mathbb{R}$, représentant le relief du terrain : pour un site x du domaine d'étude \mathcal{D} , la variable réelle $z(x)$ désigne l'altitude du terrain en x . La variable aléatoire associée est $Z(x)$. La fonction de covariance du processus Z est notée C , et le variogramme est noté γ . Tous deux dépendent de h qui désigne suivant les cas, ou bien un vecteur de \mathcal{D} séparant deux sites x_i et x_j , ou bien simplement la norme de ce même vecteur lorsque le phénomène considéré est isotrope. De la même manière, la notation $x_i - x_j$ peut désigner à la fois la distance ou bien le vecteur séparant x_i et x_j .

Lorsqu'on considère des signaux aléatoires uni-dimensionnels (en général dans un but pédagogique), on note X le processus, et $X(t)$ sa valeur en un point t du domaine (par analogie avec les signaux temporels). Une réalisation x de X est donc une fonction classique de \mathbb{R} dans \mathbb{R} . Covariance et variogramme sont alors fonctions de l'écart τ entre les points considérés : $\tau = t_2 - t_1$.

Dans ce cours, nous utiliserons fréquemment 4 loi de probabilités : $\mathcal{N}(m, \sigma^2)$, la loi normale de moyenne m et d'écart-type σ ; $\mathcal{U}([a, b])$, la loi uniforme sur le segment $[a, b] \in \mathbb{R}$; $\mathcal{B}(n, p)$, la loi binomiale décrivant le nombre de succès de n épreuves de Bernoulli de probabilité p et $\mathcal{E}(\lambda)$, la loi exponentielle d'intensité λ (*i.e.* de moyenne $1/\lambda$). Ainsi, la notation $X \sim \mathcal{N}(0, 1)$ désigne une variable aléatoire distribuée suivant la loi normale standard, et avec un léger abus de notation, on écrira $x \sim \mathcal{N}(0, 1)$ pour désigner une variable réelle (fixée) ayant été tirée suivant la loi $\mathcal{N}(0, 1)$.

Pour un critère f donné, lorsqu'une variable θ peut prendre un ensemble de valeurs dans un ensemble défini Θ , on note θ^* une valeur optimale. Par exemple dans le cas où on cherche à minimiser f :

$$\theta^* \in \operatorname{argmin}_{\theta \in \Theta} f(\theta) = \{\theta \in \Theta \mid \forall t \in \Theta : f(\theta) \leq f(t)\}$$

Remarquons que l'ensemble argmin est non-vide si, et seulement si, f est bornée en valeurs inférieures et atteint sa borne. En général, f est continue, et Θ est un pavé (donc compact) de \mathbb{R}^p : l'ensemble argmin est donc non-vide et θ^* existe (mais n'est pas nécessairement unique).

Lorsqu'une fonction f définie sur \mathbb{R}^+ admet une limite (finie) l en l'infini, on note $f(\infty) = l$.

Pour une variable inconnue z donnée, on note \hat{z} son estimateur. Lorsqu'on souhaite estimer les propriétés statistiques de cet estimateur, on le considère comme une variable aléatoire et on le note \hat{Z} . L'erreur d'estimation, $\hat{Z} - Z$ est elle-même une variable aléatoire.

En général, on note en caractères gras les quantités matricielles et vectorielles. Ainsi, \mathbf{Z} est le vecteur contenant les valeurs prises par Z en n sites : x_1, x_2, \dots, x_n . Dans ce cas, nous n'opérons plus de distinction entre la variable aléatoire et ses réalisations.

Pour un évènement quelconque A , on note $\mathbb{1}_A$ la fonction indicatrice de A , c'est-à-dire telle que $\mathbb{1}_A(x)$ vaut 1 si x vérifie l'élément A et 0 sinon. Par exemple, pour un MNT z et un seuil $s \in \mathbb{R}$, la fonction indicatrice $\mathbb{1}_{z(x) \geq s}$ vaut 1 en tout point x d'altitude supérieure à s et 0 sinon.

Pour un ensemble A quelconque, on note $|A|$ le nombre d'éléments contenus dans A (appelé cardinal de A) et $\mathcal{P}(A)$ l'ensemble des parties de A , *i.e.* l'ensembles des ensembles $B \subseteq A$. Par exemple, si $A = \{1, 2, 3\}$, alors $\mathcal{P}(A) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, A\}$. On montre facilement que $|\mathcal{P}(A)| = 2^{|A|}$, d'où la notation parfois rencontrée : $\mathcal{P}(A) = 2^A$.

Lorsqu'elle existe, \mathbf{A}^{-1} est la matrice inverse de $\mathbf{A} \in \mathbb{R}^{n \times n}$, c'est-à-dire l'unique matrice de $\mathbb{R}^{n \times n}$ telle que $\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{I}_n$ où \mathbf{I}_n désigne la matrice identité de $\mathbb{R}^{n \times n}$. Par ailleurs, \mathbf{A}^T désigne la matrice transposée de \mathbf{A} : $(\mathbf{A}^T)_{ij} = (\mathbf{A})_{ji}$.

Si f est une fonction réelle de p variables : x_1, x_2, \dots, x_p , alors $\partial f / \partial x_i$ est la dérivée partielle de f par rapport à la i -eme variable. Le vecteur ∇f de terme général $(\nabla f)_i = \partial f / \partial x_i$ est le vecteur gradient de f . Si en plus f est à valeurs dans \mathbb{R}^m , on note f_1, f_2, \dots, f_m ses m composantes scalaires et alors la matrice $\mathbf{J} \in \mathbb{R}^{m \times p}$ de terme général $(\mathbf{J})_{ij} = \partial f_i / \partial x_j$ est la *matrice jacobienne* de f , c'est-à-dire la matrice composée des m vecteurs gradients des m composantes de f .

Si X et Y sont deux variables aléatoires, alors $\mathbb{P}(X = x | Y = y)$ désigne la probabilité conditionnelle de X sachant que Y prend la valeur y . Elle vaut par définition :

$$\mathbb{P}(X = x | Y = y) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)}$$

De la même manière on peut définir l'espérance conditionnelle $\mathbb{E}[X | Y = y]$ comme l'espérance de la variable X soumise à la loi conditionnelle $\mathbb{P}(X = x | Y = y)$.

Contents

1 Rappels statistiques	10
1.1 Espace probabilisé	10
1.2 Variable aléatoire réelle	11
1.3 Fonction de répartition	12
1.4 Densité de probabilité	14
1.5 Espérance	15
1.6 Variance	17
1.7 Covariance	19
1.8 Moments statistiques	23
1.9 Exemple de synthèse	25
2 Analyse variographique	29
2.1 Processus stochastique	29
2.2 Une première approche informelle	34
2.3 Les hypothèses fondamentales	35
2.3.1 La Stationnarité	35
2.3.2 L'ergodicité	38
2.3.3 Les hypothèses en pratique	41
2.4 Le variogramme	47
2.4.1 Variogramme expérimental	50
2.4.2 Interpréter le variogramme	59
2.4.3 Les modèles de variogramme	62
2.4.4 L'estimation du variogramme	70
2.4.5 Pour aller plus loin	77
2.5 Bilan	82
3 Interpolation par Krigeage	83
3.1 Introduction	83
3.2 Les contraintes du krigeage	90
3.2.1 Contrainte de linéarité	90
3.2.2 Contrainte d'autorisation	91
3.2.3 Contrainte d'universalité	92
3.2.4 Contrainte d'optimalité	93
3.2.5 Formulation du problème	93
3.3 Le krigeage	94
3.3.1 Le krigeage ordinaire	95
3.3.2 Le krigeage intrinsèque	97
3.4 Mise en oeuvre du krigeage	99
3.5 Propriété du krigeage	108
3.5.1 Exactitude	109
3.5.2 Multiplication du variogramme	109
3.5.3 Linéarité	110
3.5.4 Lissage	113

3.6	Formulation duale	114
3.7	Les variantes du krigage	116
3.7.1	Krigage simple	116
3.7.2	Krigage de la moyenne	117
3.7.3	Krigage par blocs	119
3.7.4	Krigage universel	123
3.7.5	Krigage avec dérive externe	126
3.7.6	Autres variantes	127
3.8	Bilan	130
4	Compléments	131
4.1	La validation croisée	131
4.2	Application à l'optimisation bayésienne	140
4.2.1	Introduction	140
4.2.2	Principe	141
4.2.3	Pour aller plus loin...	144
4.3	Simulations	146
4.3.1	Simulation non-conditionnelle	147
4.3.2	Simulation conditionnelle	157
4.4	Bilan	175
A	Régularité d'un processus stochastique	176
A.1	Définitions	176
A.2	Liens avec le variogramme	177
A.3	Régularisée d'un variogramme	178
B	Contre-exemples pratiques	180
B.1	Deux variables aléatoires décorrélées mais pas indépendantes	180
B.2	Une variable aléatoire qui ne possède pas d'espérance	181
B.3	Une variable aléatoire qui ne possède pas de variance	181
B.4	Un processus stationnaire mais non-ergodique	182
B.5	Un processus ergodique à l'ordre 1 mais pas à l'ordre 2	182
B.6	Un processus strictement intrinsèque	183
B.7	Un processus continu en m.q. dont toutes les réalisations sont discontinues	184

Introduction

Une manière simple et convenie de définir la Géostatistique pourrait être la suivante :

Ensemble de techniques et méthodes permettant d'étudier les phénomènes qui s'étendent dans l'espace et y présentent une organisation ou une structure.

Autrement dit, la Géostatistique permet de répondre à des questions sur les phénomènes possédant une certaine forme de continuité dans l'espace (nous parlerons par la suite d'*autocorrélation spatiale*), mais dont la complexité interdit toutefois toute approche purement analytique du problème. Ainsi, parmi de nombreux exemples de domaines d'application, nous pourrions citer l'interpolation continue des conditions météorologiques sur un territoire donné, la prospection minière (pétrole, or...), l'analyse de champs de déformation crustale ou encore l'inventaire forestier. En particulier, dans le domaine de la géomatique, les exemples peuvent se trouver à foison :

- Production de modèle numérique de terrain haute définition (par interpolation d'un semis d'observations locales type GPS ou LIDAR ([Yang et al., 2004](#))).
- Calcul d'isochrones pour un véhicule terrestre : une *isochrone* $\Gamma(t)$ est définie par le lieu des points accessibles en un temps t depuis un point de référence donné. L'ensemble des courbes Γ calculées pour une série de valeurs t_1, t_2, \dots, t_n est appelé une *carte isochrone*.
- Estimation (à l'échelle mondiale) des retards induits par la ionosphère sur les signaux GPS.
- Analyse des erreurs cartographiques de sources anciennes de données ([Dumenieu, 2015](#)).
- Calcul d'un modèle de geoïde¹ à partir d'observations combinées de GPS et de nivellation.
- Calcul d'un champ de déformation (modèle de grille élastique) pour mettre en correspondance plusieurs sources de données vectorielles (*e.g.* bati, parcelles cadastrales et réseau routier).
- Estimation d'un modèle numérique de bâtiments 3D ([Lafarge, 2007](#)).

Cette définition est pourtant quelque peu restrictive. Dans un cadre plus général, on pourrait définir la Géostatistique comme une méthode d'analyse et de manipulations de grandeurs continues et corrélées. En statistique classique, l'hypothèse d'indépendance des variables aléatoires permet une simplification des calculs. Par exemple, si X et Y sont deux variables aléatoires indépendantes, alors les variances sont additives :

$$\text{Var}[X + Y] = \text{Var}[X - Y] = \text{Var}[X] + \text{Var}[Y] \quad (1)$$

Traduite en termes d'écart-types, cette propriété donne une relation analogue au théorème de Pythagore en géométrie euclidienne : le carré de l'écart-type d'une somme d'erreurs indépendantes vaut la

¹Le geoïde est la surface de référence des altitudes, *i.e.* le niveau moyen des mers prolongé sous les continents.

somme des carrés des écarts-types des erreurs prises individuellement. Au sens mathématique strict, l'hypothèse d'indépendance est suffisante mais non nécessaire pour obtenir cette égalité. Il est suffisant que X et Y soient non-correlées (*i.e.* de covariance nulle) pour être valide.

Dans un cadre plus général, si les variables sont corrélées, on doit ajouter un terme correctif :

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] + 2\text{Cov}(X, Y) \quad (2)$$

On peut prendre un exemple assez édifiant en considérant un modèle numérique de terrain (MNT) défini sur une grille de résolution 5 m, sur laquelle on prend deux noeuds voisins d'altitudes X et Y . Si l'erreur typique d'altitude est évaluée à ± 1 m ($\sigma_X = \sigma_Y = 1$ et donc $\text{Var}[X] = \text{Var}[Y] = 1$) par les spécifications, alors l'erreur sur le dénivelé $X - Y$ entre 2 cellules voisines vaut $\sqrt{2} \approx 1.41$ m, soit une erreur typique de près de 15° sur la pente du terrain, ce qui rendrait le MNT inutilisable pour la plupart des applications. En pratique, l'erreur entâchant les mesures d'altitude possède une autocorrélation spatiale assez marquée², impliquant ainsi que deux cellules voisines sont erronées d'une quantité statistiquement semblable, réduisant ainsi d'autant l'erreur commise lors du calcul de la pente par différentiation.

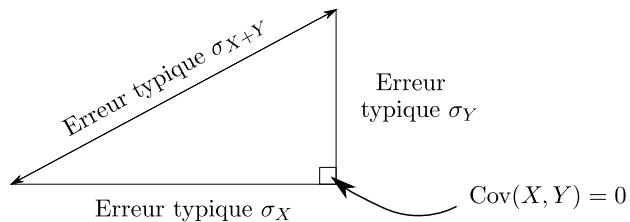


Fig. 1. Si X et Y sont des variables aléatoires décorrélées, l'erreur typique sur leur somme $S = X + Y$ est égale à la somme quadratique des erreurs individuelles : $\sigma_S^2 = \sigma_X^2 + \sigma_Y^2$

Le modèle des variables indépendantes est donc certes fort commode en première approximation pour simplifier les calculs, mais rapidement inopérant dès lors que les corrélations entre variables ne peuvent plus être négligées. La Géostatistique vient combler ce manque, et si son domaine d'application est bien souvent lié à la géographie en vertu de la première loi de Tobler³, on pourrait donc la définir, au risque de grossir un peu le trait, comme une méthode de traitement des corrélations.

Cette observation offre donc un cadre d'application beaucoup plus vaste à la Géostatistique. En particulier, les séries temporelles (*e.g.* les indices boursiers, la température en un lieu donné au fil des jours...), sont suffisamment autocorrélées pour qu'on ne puisse se permettre de ne pas en tenir compte. De manière similaire, l'espace de travail peut parfaitement être abstrait, par exemple, si on souhaite étudier les performances (*e.g.* le temps de calcul) d'un algorithme en fonction de ses inputs afin d'en déterminer un paramétrage optimal.

Notons que la Géostatistique est une discipline relativement ancienne, mais qu'elle a trouvé un essor particulier dans les années 50 sous l'impulsion de la prospective minière en Afrique du Sud, notamment avec les travaux précurseurs de Danie Gerhardus Krige : *A statistical approach to some basic mine valuation problems on the Witwatersrand* (Krige, 1951). Ces recherches ont par la suite été reprises

²Cette autocorrélation, souvent assez difficile à quantifier, se retrouve sous diverses formes dans tous les produits cartographiques, et résulte des autocorrélations individuelles des étapes de la chaîne de production : distorsion résiduelle de la caméra, erreur d'orthorectification des clichés, dérive locale d'un chantier topographique, bruit de généralisation...

³Tout interagit avec tout mais deux choses voisines ont plus de chances d'entrer en interaction que deux choses lointaines (Waldo Tobler, 1970).

et formalisées dans un cadre théorique à l’École des Mines par [Matheron \(1962\)](#) dans son *Traité de géostatistique appliquée*. Par nature, elles possèdent des interconnexions profondes avec le domaine du traitement des signaux aléatoires notamment via les filtres de Kolmogorov-Zurbenko (1941) et de Wiener (1949).

Ce cours est subdivisé en quatre parties. Après un bref rappel des prérequis statistiques dans un premier chapitre, nous verrons dans un deuxième temps comment définir et caractériser une variable régionalisée d’un point de vue stochastique⁴. Pour ce faire, nous utiliserons le *variogramme*, un outil simple et puissant, qui nous permettra de quantifier (jusqu’à un certain degré de précision) la ressemblance de la valeur prise par le phénomène en deux points donnés quelconques du domaine d’étude. Dans un troisième temps, nous étudierons le *krigeage* (et ses dérivées), une technique d’interpolation avancée, reposant sur l’analyse variographique effectuée en première partie. Sur ce point, il pourrait être instructif au préalable, de jeter un oeil aux méthodes d’interpolation déterministes dont on pourra trouver une présentation complète et pédagogique dans les ouvrages de [Mitas et Mitasova \(1999\)](#), [Arnaud et Emery \(2000\)](#) ou encore [Bosser \(2011\)](#). Enfin, dans un dernier chapitre, nous verrons quelques compléments aux notions vues dans les deux chapitres précédents : une méthode de validation numérique de la procédure d’interpolation, l’optimisation bayésienne, ainsi que la simulation qui joue un rôle important pour contourner certaines limitations du krigeage.

Dans tout ce document, nous nous référerons régulièrement au problème modèle suivant : on suppose que n points x_1, x_2, \dots, x_n d’un domaine \mathcal{D} quelconque ont été observés par GPS afin de déterminer leurs altitudes respectives z_1, z_2, \dots, z_n . Le problème consiste à produire une carte du relief du domaine, *i.e.* formellement de déterminer une fonction $z : \mathcal{D} \rightarrow \mathbb{R}$, qui à tout point $x \in \mathcal{D}$ associe l’altitude correspondante $z(x)$, avec $z(x_i) = z_i$ pour tout $i \in \{0, 1, \dots, n\}$.

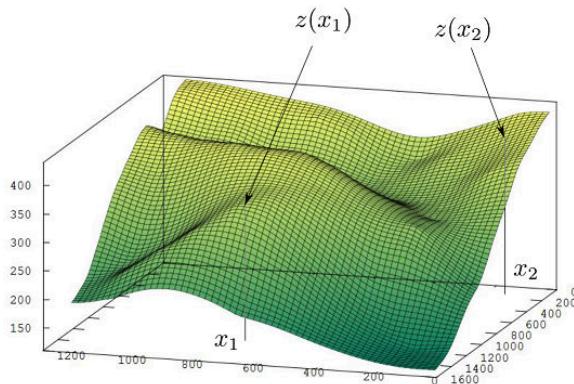


Fig. 2. Objet central de la géostatistique, la *variable régionalisée* est une fonction possédant une certaine organisation spatiale, mais dont la complexité interdit toute description analytique simple.

Trois remarques importantes avant de débuter l’étude :

- *Que faire quand on a plusieurs variables de sortie ?* En effet, ce cas de figure peut survenir, par exemple si on souhaite interpoler le champ de vecteur vitesse du vent dans un volume donné. La fonction f à déterminer est alors définie de \mathbb{R}^3 dans \mathbb{R}^3 (3 coordonnées d’espace en entrée, 3 coordonnées des vecteurs vitesse en sortie). En première approximation, on peut

⁴C'est-à-dire d'un point de vue aléatoire, par opposition aux méthodes déterministes qui ne requièrent aucun fondement probabiliste (*e.g* plus proche voisin, interpolation linéaire, splines de lissage...) .

s'en sortir en étudiant séparément chaque variable. On définit ainsi trois fonctions réelles scalaires f_x , f_y et f_z chacune prenant en entrée un point de \mathbb{R}^3 et retournant la composante du vecteur vitesse dans l'une des trois dimension d'espace. Formellement, nous avons ainsi décomposé un problème d'inférence d'une fonction à valeurs **vectorielles** $f : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ à trois problèmes indépendants d'estimation de fonctions **scalaires** $f_x, f_y, f_z : \mathbb{R}^3 \rightarrow \mathbb{R}$. Dans certains cas cependant, l'interdépendance des composantes individuelles ne rend pas souhaitable cette démarche, et on devra avoir recours à des méthodes plus spécifiques, telles que le co-krigeage, dont nous donnerons une présentation succincte dans le chapitre 3.

- Dans le cadre de ce document, toutes les variables entrant en jeu (sorties et entrées de f) doivent être **continues**, ou au moins **ordinales** (c'est-à-dire qu'elles doivent pouvoir être triées dans un ordre naturel). Exemples : température, vitesse du vent, altitude du relief, pourcentage de votants pour tel ou tel parti politique aux législatives... sont des valeurs continues. L'âge (en années) des arbres d'une forêt est une valeur discrète mais ordinale donc on peut utiliser les méthodes présentées dans ce document. En revanche, la couleur politique des maires de communes, la langue parlée par les habitants d'un pays, la marque de télévision des foyers... sont des variables catégorielles non-ordinales, et donc en l'état peu adaptées aux outils proposés par la Géostatistique.
- Dans ce document, nous nous limiterons à la **Géostatistique linéaire**, *i.e.* un pan du domaine qui se restreint à estimer les variables inconnues uniquement à l'aide de combinaison linéaires des observations effectuées. Cela peut paraître réducteur au premier regard, mais couvre déjà en réalité la grande majorité des besoins pratiques. Nous essayerons toutefois d'exhiber les limites de la Géostatistique linéaire, auxquelles nous proposerons des palliatifs dans le chapitre 4.



Fig. 3. Dannie G. Krige (à gauche), ingénieur minier sud-africain et Georges Matheron (à droite) mathématicien et géologue français : les pères fondateurs de la discipline.

Comment utiliser ce fascicule

Il existe deux options principales pour suivre ce cours. La première consiste à prendre les sections dans l'ordre du document (à l'exception du chapitre 1 qui pourra être passé par le lecteur disposant déjà d'une solide base en statistiques et probabilités), puis à considérer les activités comme des travaux pratiques, à réaliser si possible dans un premier temps sans la correction. La seconde manière de procéder, plus empirique, consiste à réaliser l'ensemble des activités (dans l'ordre indiqué par la numérotation

en chiffres romains), puis compléter les points bloquants en ciblant (à l'aide de l'index grammatical et de la table des matières) les sections de cours à parcourir. Une lecture complète du document peut alors être envisagée dans un deuxième temps lorsque les notions des bases vues dans les activités sont maîtrisées. Les sections les plus techniques du cours, pouvant allègrement être passées en première lecture, sont intitulées "Pour aller plus loin...".

Dans chaque activité nécessitant des manipulations informatiques, la partie centrale du code est regroupée dans un cadre gris, dont l'entête comprend le nom (et le numéro) de l'activité ainsi que le nom du fichier de script R correspondant dans le répertoire des corrections.

Attention : le copier-coller du code depuis le pdf ne fonctionne pas toujours correctement (ajout inopportun d'espaces, en particulier sur les mots-clés R). Il y a trois solutions principales pour résoudre ce problème :

- Faire du copier-coller et rectifier le code, éventuellement avec les messages d'erreur de R (nécessite une maîtrise minimale du langage).
- Recopier le code au clavier.
- Aller chercher à chaque étape la portion de code correspondante dans les corrigés.

Chapter 1

Rappels statistiques

Afin de pouvoir manipuler des quantités aléatoires, nous devons au préalable nous placer dans un cadre de travail adéquat. La géostatistique linéaire se circonscrit à l'utilisation des deux premiers moments statistiques. L'objectif de cette section est donc de rappeler les éléments théoriques nécessaires à la compréhension des notions d'*espérance* et de *covariance*. Le lecteur intéressé pourra trouver une présentation plus complète dans l'excellent ouvrage de [Lecoutre \(2002\)](#).

1.1 Espace probabilisé

Soit Ω un ensemble quelconque (éventuellement continu), représentant l'ensemble des **éventualités élémentaires** d'une expérience aléatoire. À partir de Ω , on construit l'ensemble¹ \mathcal{A} contenant tous les **événements** pouvant être produits par l'expérience aléatoire. Enfin, on doit définir une **mesure de probabilité** \mathbb{P} sur \mathcal{A} , afin d'attribuer une probabilité d'apparition à tout évènement $\alpha \in \mathcal{A}$:

$$\begin{aligned} \mathbb{P} &: \mathcal{A} \rightarrow \mathbb{R} \\ \alpha &\mapsto \mathbb{P}(\alpha) \end{aligned} \tag{1.1}$$

où \mathbb{P} respecte les 3 axiomes fondamentaux de la théorie des probabilités (axiomes de Kolmogorov) :

- Pour tout évènement $A \in \mathcal{A}$ $0 \leq \mathbb{P}(A) \leq 1$.
- $\mathbb{P}(\Omega) = 1$, autrement dit, la probabilité de l'évènement certain vaut 1.
- Soient A_1, A_2, \dots, A_n , une suite de n évènements incompatibles de \mathcal{A} , i.e. d'intersections mutuelles vides ($i \neq j \Rightarrow A_i \cap A_j = \emptyset$) :

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \mathbb{P}(A_i) \tag{1.2}$$

Ce troisième axiome stipule que la probabilité de l'union d'un ensemble d'évènements disjoints est toujours égale à la somme des probabilités d'apparition des évènements pris individuellement. Lorsque les évènements ne sont plus disjoints, la relation 1.2 se transforme en une inégalité, comme le montre l'exemple ci-après.

Exemple 1.1. L'Éducation Nationale indique qu'en 2010, 10.8 % des collégiens ont suivi l'option d'enseignement du Latin, et 7.1 % celle du Grec ancien. En notant A_L et A_G les évènements correspondants, on a donc $\mathbb{P}(A_L) = 0.108$ et $\mathbb{P}(A_G) = 0.071$. Pour autant, selon la même source, seuls 11.9

¹Formellement, \mathcal{A} est une tribu sur Ω , c'est-à-dire un sous-ensemble de l'ensemble $\mathcal{P}(\Omega)$ des parties de Ω , assurant ainsi certaines propriétés de stabilité nécessaires pour établir une théorie probabiliste. En pratique, lorsque Ω est discret, on peu prendre sans vergogne $\mathcal{A} = \mathcal{P}(\Omega)$.

% ont suivi au moins une des deux options : $\mathbb{P}(A_L \cup A_G) = 0.119 \leq \mathbb{P}(A_L) + \mathbb{P}(A_G) = 0.179$, puisqu'on doit retrancher à la somme la probabilité qu'un collégien ait choisi les deux options.

Dans le cas non-disjoint et avec deux évènements A et B , on obtient la formule bien connue :

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \quad (1.3)$$

Démonstration. On peut facilement décomposer $A \cup B$ en 3 sous-ensembles : $A \setminus B$, $B \setminus A$ et $A \cap B$, dont on montre facilement qu'ils sont deux-à-deux disjoints. Puis, on peut décomposer A en 2 sous-ensembles disjoints : $A \setminus B$ et $A \cap B$. On fait de même pour B . En appliquant le troisième axiome de Kolmogorov, il vient :

$$\begin{aligned} \mathbb{P}(A \cup B) &= \mathbb{P}((A \setminus B) \cup (B \setminus A) \cup (A \cap B)) \\ &= \mathbb{P}(A \setminus B) + \mathbb{P}(B \setminus A) + \mathbb{P}(A \cap B) \\ &= (\mathbb{P}(A) - \mathbb{P}(A \cap B)) + (\mathbb{P}(B) - \mathbb{P}(A \cap B)) + \mathbb{P}(A \cap B) \\ &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \end{aligned} \quad (1.4)$$

En combinant l'ensemble des éventualités élémentaires Ω , son ensemble d'évènements \mathcal{A} et une mesure de probabilité \mathbb{P} on obtient ainsi une structure mathématique appelée *espace probabilisé* $(\Omega, \mathcal{A}, \mathbb{P})$.

Exemple 1.2. On prend place à bord d'un véhicule et on s'intéresse à l'état du prochain feu tricolore rencontré. L'ensemble Ω est donc composé de trois éléments correspondants aux états possibles : vert (V), orange (O) et rouge (R), d'où $\Omega = \{V, O, R\}$. En particulier, le passage est autorisé $A_1 = \{V, O\}$ ou interdit $A_2 = \{R\}$. L'ensemble des évènements possibles est alors $\mathcal{A} = \{\emptyset, A_1, A_2, \Omega\}$. Si on suppose que chaque feu est programmé pour être au rouge pendant 60 % du temps, au vert pendant 35 % du temps et enfin à l'orange le temps restant, on en déduit la mesure de probabilité associée : $\mathbb{P}(\emptyset) = 0$, $\mathbb{P}(A_1) = 0.35 + 0.05 = 0.4$, $\mathbb{P}(A_2) = 0.6$ et $\mathbb{P}(\Omega) = 1$. On vérifiera alors à titre d'exercice que les 3 axiomes de Kolmogorov sont vérifiés.

Remarquons que le choix de l'ensemble des évènements \mathcal{A} à partir de Ω n'est pas unique. En fonction des besoins, on aurait pu proposer la modélisation suivante : le passage est sécurisé $B_1 = \{V\}$ ou alors, le passage est déconseillé, voire interdit $B_2 = \{O, R\}$, ce qui donne : $\mathcal{A} = \{\emptyset, B_1, B_2, \Omega\}$. Pour un niveau de granularité maximale, on peut combiner tous les évènements, en ajoutant donc $C_1 = \{R, V\}$ (les consignes données par le feu sont claires) et $C_2 = \{O\}$ (le passage est laissé à l'appréciation du conducteur). On obtient alors : $\mathcal{A} = \mathcal{P}(\Omega) = \{\emptyset, A_1, A_2, B_1, B_2, C_1, C_2, \Omega\}$. Nous laissons le soin au lecteur de calculer, à titre d'exercice, les probabilités des différents évènements.

En général, l'ensemble \mathcal{A} n'est pas défini explicitement. Dans ce cours, nous utiliserons le symbole $\omega \in \Omega$ pour désigner une réalisation stochastique quelconque, *i.e.* l'issue d'une expérience aléatoire.

1.2 Variable aléatoire réelle

Afin de quantifier les phénomènes étudiés, on cherche souvent à associer une valeur numérique à chaque issue possible de l'expérience. C'est ici qu'intervient la notion de **variable aléatoire**.

Formellement, on peut définir une *variable aléatoire réelle* X comme une fonction qui, à chaque éventualité élémentaire ω associe une valeur réelle $x = X(\omega)$.

$$\begin{aligned} X & : \Omega \rightarrow \mathbb{R} \\ \omega & \mapsto X(\omega) \end{aligned} \tag{1.5}$$

Traditionnellement, on note en majuscule une variable aléatoire, et en minuscule une réalisation donnée (quelconque) de cette variable.

Exemple 1.3. Dans une version simplifiée du Loto, un joueur choisit 5 numéros dans une grille de nombres allant de 1 à 49. On admet que les gains varient de la manière suivante en fonction du nombre de numéros corrects : 4 € (2 num), 21 € (3 num), 490 € (4 num) et 95200 € (5 num).

L'ensemble fondamental correspondant est donc $\Omega = \{1, 2, \dots, 49\}^5$ et il contient près de 300 millions d'éléments. Parmi tous les événements possibles, on note :

A_i : le tirage du jour contient i numéros sélectionnés par le joueur

et on s'intéresse plus particulièrement aux événements² A_2, A_3, A_4 et A_5 . On définit alors la variable aléatoire réelle X dénotant le gain du joueur :

$$X(\omega) = \begin{cases} 4 & \text{si } \omega \in A_2 \\ 21 & \text{si } \omega \in A_3 \\ 490 & \text{si } \omega \in A_4 \\ 95200 & \text{si } \omega \in A_5 \\ 0 & \text{sinon.} \end{cases}$$

Supposons que le choix du joueur est le suivant (37, 5, 48, 15, 29). Une réalisation aléatoire particulière, par exemple $\omega = (5, 42, 2, 29, 37)$ produira donc la variable $x = X(\omega) = 21$.

1.3 Fonction de répartition

Si la fonction X possède les propriétés théoriques adéquates (en particulier celle d'être mesurable), on peut transférer la loi de probabilité \mathbb{P} sur la variable aléatoire. Pour tout intervalle $A \subseteq \mathbb{R}$ (en toute généralité, A peut être un élément de la tribu borélienne $\mathcal{B}(\mathbb{R})$), ce qui inclut entre autres les intervalles), on peut définir la loi P_X :

$$P_X(A) = \mathbb{P}(X \in A) = \mathbb{P}(X^{-1}(A)) \tag{1.6}$$

où $X^{-1}(A) = \{\omega \in \Omega \mid X(\omega) \in A\}$, autrement dit, l'ensemble des événements élémentaires ω qui, lorsqu'ils sont réalisés, produisent une variable x inclue dans l'intervalle réel A .

La fonction P_X n'étant pas commode à manipuler (en particulier, il est très difficile de la représenter graphiquement puisqu'elle prend en entrée un intervalle de nombres réels), on la réduit à une seule variable scalaire en ne considérant que les intervalles réels du type $] -\infty; x]$. On obtient alors une fonction réelle classique : la *fonction de répartition*.

²Notons que tous ces événements ne comportent pas le même nombre d'éventualités ω . Par exemple, il n'y a qu'un seul tirage qui satisfasse A_5 , alors que A_4 par exemple comporte $5 \times 44 = 220$ éléments.

Définition 1 (Fonction de répartition)

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, et $X : \Omega \rightarrow \mathbb{R}$ une variable aléatoire réelle. On appelle *fonction de répartition* de X , la fonction F_X définie par :

$$F_X(x) = \mathbb{P}(X \in]-\infty; x]) = \mathbb{P}(X \leq x)$$

Lorsqu'il n'y a pas d'ambiguïté sur la variable concernée, nous noterons F cette fonction. Par ailleurs, à l'aide des axiomes de Kolmogorov, on démontre facilement les propriétés suivantes.

Propriété 2 (Fonction de répartition)

Soit X une variable aléatoire réelle et F_X sa fonction de répartition.

- (i) F_X est croissante et à valeurs dans $[0, 1]$.
- (ii) $\forall a, b \in \mathbb{R}$ avec $a \leq b$: $\mathbb{P}(X \in [a, b]) = F_X(b) - F_X(a)$
- (iii) $\lim_{x \rightarrow -\infty} F_X(x) = 0$ et $\lim_{x \rightarrow +\infty} F_X(x) = 1$

Le passage de la loi de probabilité P_X à la fonction de répartition F_X induit bien évidemment une perte d'information, mais la propriété (ii) en particulier, montre que la fonction de répartition permet de calculer la probabilité que la variable X appartienne à tout intervalle $I \subseteq \mathbb{R}$, ce qui est largement suffisant en pratique.

Exemple 1.4. On donne ci-dessous en figure 1.1 la fonction de répartition de l'erreur (en m) commise par un récepteur GPS de randonnée (*i.e.* la distance entre la position vraie et la position estimée). Cette fonction permet la détermination immédiate de plusieurs indicateurs, comme la médiane $m = F_X^{-1}(0.5) = 1.91$ m. On y observe également que 95% des erreurs sont inférieures à 6.63 m.

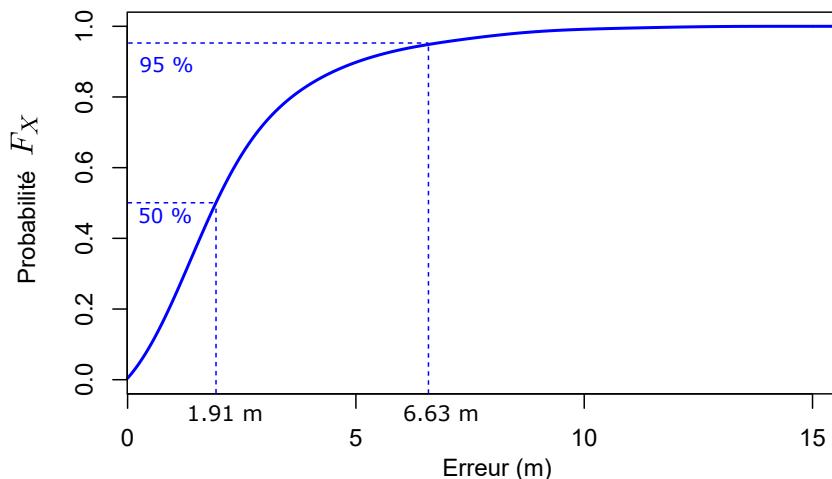


Fig. 1.1. Fonction de répartition F_X de la variable aléatoire réelle X d'erreur GPS (en m).

1.4 Densité de probabilité

La fonction de répartition est un outil pratique, mais ne donne cependant que des informations globales sur la loi \mathbb{P} (c'est-à-dire des valeurs de probabilités intégrées sur des intervalles plus ou moins larges). Si on reprend l'exemple de l'erreur commise par un GPS, on observe qu'il existe une infinité (qui plus est non dénombrable) de valeurs possibles entre 0 et 15 m. Chacune d'elles a donc logiquement une probabilité nulle d'apparaître. Qu'en est-il si on souhaite tout de même quantifier la contribution d'une valeur particulière x dans la loi totale ? On peut s'en sortir en se plaçant sur un voisinage très local de x , par exemple sur l'intervalle $I = [x; x + dx]$, puis en calculant la probabilité qu'une réalisation de X appartienne à I , rapportée à la largeur de I . Le point (ii) de la propriété 2 indique que cette probabilité s'exprime par $\mathbb{P}(X \in I) = F_X(x + dx) - F_X(x)$. Après division par la largeur dx de l'intervalle et passage à la limite $dx \rightarrow 0$, on obtient précisément la dérivée de F_X .

Définition 3 (Densité de probabilité)

Soit X une variable aléatoire réelle et F_X sa fonction de répartition. Si F_X est dérivable sur \mathbb{R} , alors on peut définir la densité de probabilité π par :

$$\pi(x) = \frac{dF_X}{dx}(x)$$

Remarque : inversement, étant donnée la densité f d'une variable aléatoire réelle, on peut retrouver sa fonction de répartition :

$$F(x) = \int_{-\infty}^x f(t)dt \quad (1.7)$$

Plus généralement on a les propriétés suivantes :

Propriété 4 (Densité de probabilité)

Soit X une variable aléatoire réelle de densité π , alors :

$$(i) \forall x \in \mathbb{R} \quad \pi(x) \geq 0.$$

$$(ii) \forall a, b \in \mathbb{R} \quad \int_a^b \pi(x)dx = \mathbb{P}(X \in [a, b])$$

$$(iii) \int_{\mathbb{R}} \pi(x)dx = 1$$

Preuves : (i) découle immédiatement du fait que F_X est croissante, d'après le point (i) de la propriété 2. Par ailleurs, d'après le théorème fondamental de l'analyse :

$$\int_a^b \pi(x)dx = \int_a^b \frac{dF_X}{dx}(x)dx = F(b) - F(a)$$

ce qui montre le point (ii). Enfin, en faisant tendre a et b vers l'infini, et à l'aide du point (iii) de la propriété 2, on obtient que l'intégrale de π sur \mathbb{R} vaut 1 (iii).

Attention : lorsqu'elle existe, la densité π est positive et son aire sous la courbe vaut 1. Mais il faut être vigilant au fait que π n'est pas un probabilité ! Elle peut prendre des valeurs supérieures à 1. Pour obtenir une probabilité, on doit intégrer π , comme stipulé par le point (ii) de la propriété 4.

Exemple 1.5. On reprend l'exemple 1.4, et on en déduit en figure 1.2 la densité de probabilité des erreurs commises par le GPS.

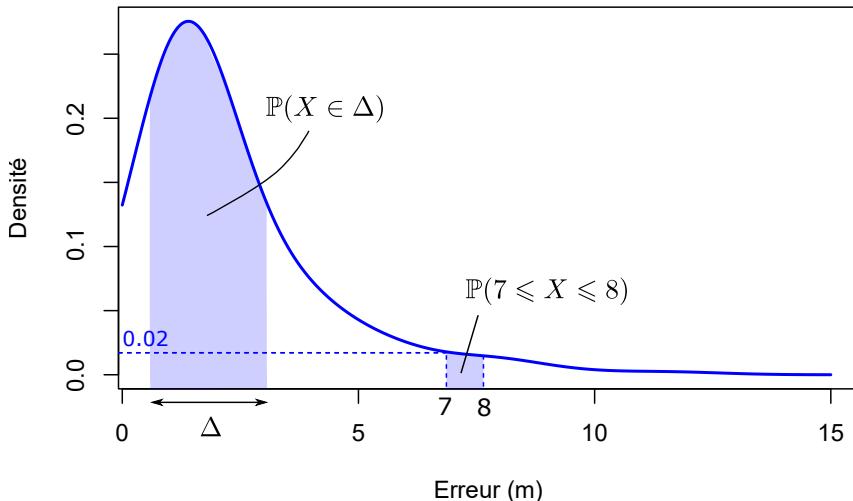


Fig. 1.2. Densité de probabilité π de la variable aléatoire réelle X d'erreur GPS (en m).

Lorsque l'intervalle est petit devant les variations de π , on peut approcher le calcul de la probabilité $\mathbb{P}(x \in \Delta x)$ par $\pi(x)\Delta x$. Par exemple, la probabilité que l'erreur commise soit comprise entre 7 et 8 m est : $\mathbb{P}(7 \leq X \leq 8) \approx \pi(7.5) \times (8 - 7) = 2\%$.

1.5 Espérance

Partant à nouveau de l'exemple 1.4, on peut chercher à déterminer la valeur moyenne des erreurs commises par le GPS. On appelle *espérance mathématique* d'une variable aléatoire X la moyenne des valeurs prises par X sur un très grand nombre de réalisations³.

Définition 5 (Espérance)

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et X une variable aléatoire (discrète ou continue). On définit l'*espérance* de X comme le centre de masse (lorsqu'il existe) du support de X suivant la mesure \mathbb{P} :

$$\mathbb{E}[X] = \int_{\Omega} X(\omega) d\mathbb{P}(\omega)$$

Suivant que la loi de X est discrète ou continue, la définition 5 admet une forme plus explicite :

³Cette caractérisation intuitive de l'espérance n'a en réalité rien de trivial, et constitue l'objet d'un théorème fondamental de la statistique : la *loi forte des grands nombres* (Lecoutre, 2002).

- Dans le cas **discret**, $\mathbb{E}[X]$ est la moyenne des valeurs x_i potentiellement prises par X pondérées par les probabilités d'apparition associées p_i :

$$\mathbb{E}[X] = \sum_{i=0}^{+\infty} x_i p_i \quad (1.8)$$

sous réserve que la suite de terme général $x_i p_i$ soit sommable.

- Dans le cas **continu**, si X admet π pour densité et si la fonction $x\pi(x)$ est intégrable :

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} x\pi(x)dx \quad (1.9)$$

Exemple 1.6. En reprenant l'exemple 1.3, on calcule la probabilité p_i d'obtenir l'événement A_i : *le tirage du jour contient i numéros sélectionnés par le joueur*. Le nombre total de tirages possibles est égal au nombre des façons de tirer 5 éléments parmi 49 : c'est donc le coefficient $\binom{49}{5}$. Par ailleurs, le nombre de tirages aléatoires ω vérifiant l'événement A_i est égal au nombre de manières de tirer i éléments parmi les 5 éléments choisis par le joueur, multiplié par le nombre de manières de choisir les $5 - i$ autres éléments parmi les 44 éléments restants :

$$p_i = \mathbb{P}(A_i) = \frac{\binom{5}{i} \binom{44}{5-i}}{\binom{49}{5}}$$

L'application numérique donne :

$$\mathbb{E}[X] = 4 \times \frac{132440}{1906884} + 21 \times \frac{9460}{1906884} + 490 \times \frac{440}{1906884} + 95200 \times \frac{10}{1906884}$$

D'où une espérance de gain⁴ de 0.99 €... à laquelle on doit soustraire le prix du ticket (2.20 €).

Exemple 1.7. Lorsque la variable aléatoire X peut prendre un nombre infini de modalités, évaluer l'espérance nécessite l'emploi de techniques de séries numériques ou de calcul intégral. Par exemple, considérons la loi binomiale $\mathcal{B}(n, p)$ qui donne la probabilité d'obtenir k succès à l'issue de la répétition de n expériences aléatoires identiques de probabilité de succès individuelle p :

$$P_n(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad (1.10)$$

On calcule alors l'espérance de la loi discrète 1.10 par :

$$\mathbb{E}[X] = \sum_{k=0}^{+\infty} k \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=1}^{+\infty} k \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

⁴En comptabilisant le *numéro chance*, l'espérance augmente mais reste toujours négative.

$$\begin{aligned}
&= np \sum_{k=1}^{+\infty} \frac{(n-1)!}{(k-1)!(n-1-(k-1))!} p^{k-1} (1-p)^{n-k} = np \sum_{k=0}^{+\infty} \frac{(n-1)!}{k!(n-1-k)!} p^k (1-p)^{n-1-k} \\
&= np \sum_{k=0}^{+\infty} P_{n-1}(X=k) = np
\end{aligned}$$

On procède de même avec une loi continue, par exemple la loi normale standard $\mathcal{N}(0, 1)$ (moyenne nulle et variance unitaire), définie par sa densité : $p(x) = \exp(-x^2/2)/\sqrt{2\pi}$

$$\mathbb{E}[X] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x \exp\left(-\frac{x^2}{2}\right) dx = \frac{1}{\sqrt{2\pi}} \left[-\exp\left(-\frac{x^2}{2}\right) \right]_{-\infty}^{+\infty} = 0$$

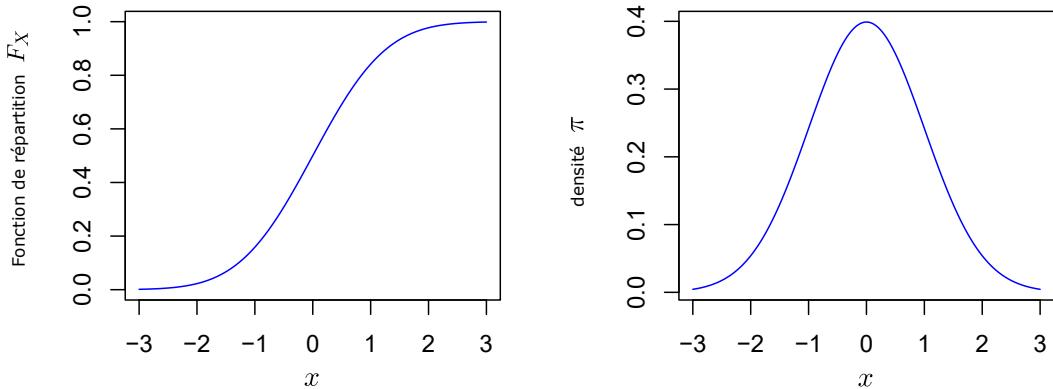


Fig. 1.3. Fonction de répartition et densité de probabilité de la loi normale standard.

Propriété 6 (Espérance)

Soient X et Y deux variables aléatoires réelles quelconques et $a \in \mathbb{R}$:

- (i) Espérance d'une constante : $\mathbb{E}[a] = a$
- (ii) Linéarité de l'espérance : $\mathbb{E}[aX + Y] = a\mathbb{E}[X] + \mathbb{E}[Y]$

Remarque : l'espérance étant un opérateur linéaire, en général, pour une fonction f non linéaire, on a $\mathbb{E}[f(X)] \neq f(\mathbb{E}[X])$ (la moyenne des carrés des notes d'étudiants par exemple, n'est en général pas égale au carré de la moyenne de la classe). De même, en général $\mathbb{E}[XY] \neq \mathbb{E}[X] \times \mathbb{E}[Y]$ (l'égalité est cependant vérifiée quand X et Y sont décorélées, cf section 1.7).

1.6 Variance

L'espérance d'une variable aléatoire permet de localiser le centre d'une distribution. Pour caractériser la dispersion de la distribution autour de cette valeur centrale, on a besoin d'un second indicateur. C'est précisément le rôle de la variance.

Définition 7 (Variance)

Soit X une variable aléatoire (discrète ou continue). On définit la *variance* de X comme la valeur moyenne (lorsqu'elle existe) du carré des écarts de X à son espérance :

$$\text{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

À nouveau, on a deux expressions explicites différentes suivant que la variable est discrète (à gauche) ou continue et à densité (à droite) :

$$\text{Var}[X] = \sum_{i=0}^{+\infty} p_i(x_i - \mathbb{E}[X])^2 \quad \text{Var}[X] = \int_{\mathbb{R}} (x - \mathbb{E}[X])^2 \pi(x) dx \quad (1.11)$$

Pour pouvoir donner une interprétation physique à $\text{Var}[X]$ on en extrait souvent la racine carrée :

Définition 8 (Ecart-type)

On appelle *écart-type* d'une variable aléatoire X la quantité :

$$\sigma_X = \sqrt{\text{Var}[X]}$$

Exemple 1.8. Un thermomètre (dont la précision de mesure sera supposée absolue, mais qui n'affiche que des valeurs entières) indique 17°C . On peut donc supposer que la température T réelle est distribuée suivant une loi uniforme entre 16.5° et 17.5° (on note $\mathcal{U}([16.5; 17.5])$ cette loi) :

$$\pi(t) = \begin{cases} 1 & \text{si } 16.5 \leq t < 17.5 \\ 0 & \text{sinon.} \end{cases}$$

On peut alors estimer l'incertitude sur la valeur affichée en calculant la variance de T :

$$\text{Var}[T] = \int_{\mathbb{R}} (t - \mathbb{E}[T])^2 \pi(t) dt = \int_{16.5}^{17.5} (t - 17)^2 dt$$

Par changement de variable $t \mapsto t + 17$, on réduit le problème à :

$$\text{Var}[T] = \int_{-1/2}^{+1/2} t^2 dt = \frac{1}{12}$$

L'écart-type de la mesure est donc égal à $\frac{1}{\sqrt{12}} \approx 0.29^\circ\text{C}$. Ce bruit sur la lecture d'un capteur est appelé *bruit de numérisation* ou encore *bruit d'arrondi*.

Propriété 9 (Variance)

Soient X et Y deux variables aléatoires réelles **indépendantes** et $a \in \mathbb{R}$:

(i) Variance d'une constante : $\text{Var}[a] = 0$

(ii) Combinaison linéaire indépendante : $\text{Var}[aX + Y] = a^2\text{Var}[X] + \text{Var}[Y]$

La propriété (ii) confirme la validité du changement de variable de l'exemple du thermomètre : $\text{Var}[T - 17] = \text{Var}[T] + \text{Var}[17] = \text{Var}[T]$. La variance est insensible à toute translation.

1.7 Covariance

Pour traiter le cas de la somme de variables dépendantes, on doit introduire la notion de **covariance**.

Définition 10 (Covariance)

Soit (X, Y) un couple de variables aléatoires réelles. On définit la *covariance* par la quantité :

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

Remarque 1. L'opérateur de covariance peut être considérée comme un produit scalaire de l'espace préhilbertien des variables aléatoires de carré intégrable. Il hérite donc de toutes les propriétés algébriques des produits scalaires. En particulier il est **symétrique** : $\text{Cov}(X, Y) = \text{Cov}(Y, X)$ et **bilinéaire** : $\text{Cov}(aX + Y, Z) = a\text{Cov}(X, Z) + \text{Cov}(Y, Z)$.

Remarque 2. Il est important de le mentionner, la covariance ne mesure que la dépendance **linéaire**. On peut exhiber deux variables dépendantes de covariance nulle, *e.g.* X et X^2 avec $X \sim \mathcal{U}([-1; 1])$. On pourra trouver un contre-exemple similaire dans [Hauchecorne \(2007\)](#), page 360.

Afin d'obtenir une mesure facilement interprétable du degré de dépendance, on peut normaliser le résultat par le produit des écarts-types de chacune des deux variables.

Définition 11 (Corrélation)

Soit (X, Y) un couple de variables aléatoires réelles. On définit la *corrélation* par la quantité :

$$\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

L'inégalité de Cauchy-Schwarz nous montre alors que ρ_{XY} est compris entre -1 et 1 .

Remarque sémantique : on parle d'*autocorrélation* (spatiale, temporelle ou autre) lorsque les deux variables ont été générées par le *même phénomène*. Par exemple si $Z(t)$ désigne l'altitude du Mont-Blanc

à l'année t , alors on pourra parler de l'autocorrélation entre $Z(t_1)$ et $Z(t_2)$. Lorsque le phénomène Z étudié est implicite, on pourra noter par abus de langage $\rho(t_1, t_2)$ et $\text{Cov}(t_1, t_2)$. Un couple de variables (X, Y) est dit décorrélé lorsque $\rho_{XY} = \text{Cov}(X, Y) = 0$. Attention au piège en pratique : $\text{Cov}(X, Y)$ proche de 0 n'implique pas ρ_{XY} proche de 0 et réciproquement ! Enfin, un phénomène est dit décorrélé lorsque les valeurs qu'il prend sont deux-à-deux décorélées.

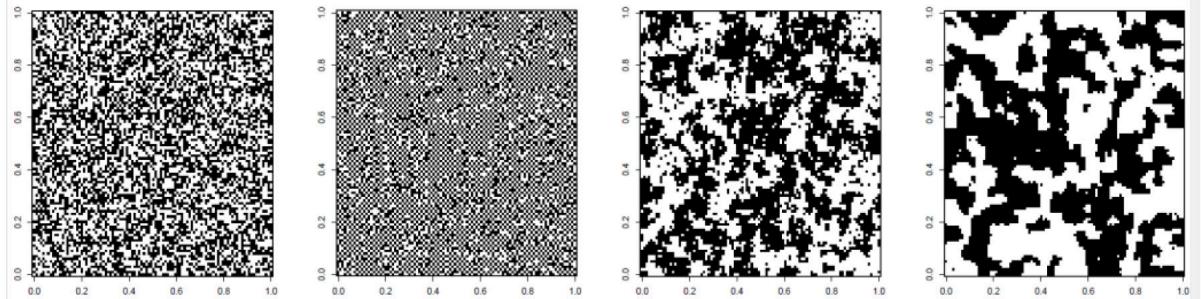


Fig. 1.4. De gauche à droite : champ aléatoire décorrélé ($\rho = 0$), champ anti-corrélé ($\rho < 0$), champ faiblement corrélé ($\rho > 0$) et champ fortement corrélé ($\rho \gg 0$).

Remarque : $\text{Var}[X] = \text{Cov}(X, X)$. Ainsi, pour un ensemble de n variables aléatoires X_1, X_2, \dots, X_n , on peut formuler toutes les informations de variances et de covariances dans une unique matrice carrée de dimension n et de terme général $(\Sigma)_{ij} = \text{Cov}(X_i, X_j)$. On l'appelle la **matrice de covariance**. Si les variables sont décorélées (et *a fortiori* si elles sont indépendantes), Σ est une matrice diagonale.

$$\Sigma_X = \begin{pmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_2) & \dots & \text{Cov}(X_1, X_n) \\ \text{Cov}(X_2, X_1) & \text{Var}(X_2) & \dots & \text{Cov}(X_2, X_n) \\ \vdots & \dots & \ddots & \vdots \\ \text{Cov}(X_n, X_1) & \dots & \dots & \text{Var}(X_n) \end{pmatrix} \quad (1.12)$$

La matrice de covariance est un outil fondamental dans le domaine de la propagation des incertitudes. Soit $\mathbf{X} = (X_1, X_2, \dots, X_n)^T$ un vecteur aléatoire (*i.e.* un vecteur contenant un ensemble de variables aléatoires). Si on se donne une matrice $\mathbf{A} \in \mathbb{R}^{m \times n}$, l'application $\mathbf{X} \mapsto \mathbf{AX}$ est une application linéaire transformant un groupe de n variables aléatoires en un second groupe de m variables aléatoires. Le théorème suivant permet de propager les incertitudes des variables d'entrée aux variables de sortie.

Théorème 12 (Propagation des variances)

Soit \mathbf{X} un vecteur aléatoire de matrice de covariance Σ_X et $\mathbf{A} \in \mathbb{R}^{m \times n}$. Alors, la matrice de covariance Σ_Y sur $\mathbf{Y} = \mathbf{AX}$ est :

$$\Sigma_Y = \mathbf{A}\Sigma_X\mathbf{A}^T$$

En particulier, en posant $A = [1, 1]$, on retrouve directement l'expression 2 de la variance d'une somme de 2 variables aléatoires corrélées : $\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] + 2\text{Cov}(X, Y)$

Que faire lorsque l'application n'est pas linéaire ? Soit f une fonction scalaire non-linéaire, et X une variable aléatoire de variance $\text{Var}(X)$. Pour évaluer la variance de $Y = f(X)$ une solution prête à l'emploi consiste à linéariser f autour d'une valeur de référence a :

$$f(X) = f(a) + f'(a)(X - a) + \frac{f''(a)}{2}(X - a)^2 + \dots + \frac{f^{(n)}(a)}{n!}(X - a)^n + o((X - a)^n)$$

Alors, par les propriétés de la variance, et en se limitant à un développement limité à l'ordre 2, on a :

$$\text{Var}[f(X)] \approx \text{Var}[f(a) + f'(a)(X - a)] = \text{Var}[f(a)] + f'(a)^2 \text{Var}[X - a] = f'(a)^2 \text{Var}[X]$$

Prenons le cas de $f : x \mapsto x^2$, et supposons que l'on travaille au voisinage de $a = 1$. On a alors :

$$\text{Var}[X^2] = (2a)^2 \text{Var}[X] = 4\text{Var}[X]$$

Bilan : si $x = 1 \pm 0.02$ (où 0.02 représente l'écart-type de x), alors $f(x) = x^2 = 1 \pm 0.04$.

On peut généraliser cette méthode pour une fonction $f : \mathbb{R}^n \mapsto \mathbb{R}^m$, dont on peut calculer la matrice jacobienne \mathbf{J} de terme général $(\mathbf{J})_{ij} = \frac{\partial f_i}{\partial x_j}$, où f_1, f_2, \dots, f_m sont les composantes de f .

Théorème 13 (Propagation des variances : cas non-linéaire)

Soit $f : \mathbb{R}^n \mapsto \mathbb{R}^m$, une fonction régulière de matrice jacobienne \mathbf{J} et \mathbf{X} un vecteur aléatoire de \mathbb{R}^n de matrice de covariance Σ_X :

$$\Sigma_{f(\mathbf{X})} = \mathbf{J}\Sigma_X\mathbf{J}^T$$

Exemple 1.9. On reprend l'exemple du thermomètre. Nous avons vu que le bruit de numérisation induit une erreur de lecture d'écart-type environ égal à trois dixièmes de degré. On fait un second relevé, et on lit à présent la valeur 20° , soit une augmentation de température ΔT de 3° . Quelle est l'incertitude sur la valeur de ΔT ?

On note $\sigma = 0.3$ l'incertitude de lecture sur une mesure, et X_1 et X_2 les deux mesures effectuées. Elles sont décorrélées, donc de matrice de covariance diagonale : $\Sigma = \sigma^2 I_2$ où I_2 est l'identité de \mathbb{R}^2 . En exprimant ΔT sous la forme matricielle $\Delta T = [-1 \ 1]\mathbf{X}$, le théorème de propagation des variances nous donne immédiatement la matrice de covariance (ne contenant qu'un élément) de ΔT :

$$\Sigma_{\Delta T} = [-1 \ 1] \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} = 2\sigma^2$$

On en déduit l'écart-type sur la différence de température : $\sigma_{\Delta T} = \sigma\sqrt{2}$. La différence de température mesurée est donc : $\Delta T = 3 (\pm 0.41) ^\circ C$.

Exemple 1.10. A l'issue d'un examen, un étudiant reçoit 5 notes (sur 20 points) : mathématiques (coeff 12), physique-chimie (coeff 7), informatique (coeff 7), français (coeff 4) et anglais (coeff 2). On suppose que l'écart-type de l'erreur d'évaluation de chaque copie est de 3 points. Par ailleurs un seul et même évaluateur corrige les copies de mathématiques et d'informatique, ce que nous modéliserons par une corrélation ($\rho = 50\%$) des erreurs d'évaluation sur ces deux épreuves. Calculer l'incertitude sur la moyenne de l'étudiant. Qu'en est-t-il si on suppose à présent les corrections de mathématique

et d'informatique comme étant parfaitement décorrélées ?

On exprime sous forme matricielle la moyenne de l'étudiant : $\mu = \frac{1}{32} \mathbf{A} \mathbf{X}$, avec \mathbf{A} le vecteur des coefficients et \mathbf{X} le vecteur aléatoire des notes obtenues. Par ailleurs, on a $\text{Var}[X] = \sigma^2 = 9$ et $\text{Cov}(X_i, X_j) = 0$ sauf pour le couple maths-info où on a : $\text{Cov}(X_i, X_j) = \rho_{ij}\sigma_{X_i}\sigma_{X_j} = \frac{\sigma^2}{2} = 4.5$

La matrice de covariance de X traduisant l'incertitude sur l'évaluation s'exprime par :

$$\Sigma_X = \begin{bmatrix} 9 & 0 & 4.5 & 0 & 0 \\ 0 & 9 & 0 & 0 & 0 \\ 4.5 & 0 & 9 & 0 & 0 \\ 0 & 0 & 0 & 9 & 0 \\ 0 & 0 & 0 & 0 & 9 \end{bmatrix}$$

Une application numérique nous donne immédiatement : $\text{Var}[\mu] = \frac{1}{32^2} \mathbf{A} \Sigma_X \mathbf{A}^T = 3.03$

La note moyenne de l'étudiant au module est donc entâchée d'une erreur d'évaluation d'écart-type $\sqrt{3.03} = 1.74$ points. En utilisant le graphique 1.3 de la fonction de répartition de la loi normale, on peut (en première approximation) annoncer qu'avec une probabilité de 95% l'impact de l'erreur d'évaluation sur la moyenne obtenue par l'étudiant à l'examen, sera inférieure à $2 \times 1.74 = 3.48$ points.

En supposant à présent que les copies de mathématiques et d'informatique sont corrigées par deux évaluateurs différents, la matrice Σ_X devient diagonale et on a :

$$\text{Var}[\mu] = \frac{1}{32^2} \mathbf{A} (9I_5) \mathbf{A}^T = \frac{9}{32^2} \mathbf{A} \mathbf{A}^T = 2.28$$

On remarque donc que l'erreur sur la moyenne d'examen est amplifiée par le fait qu'un même évaluateur corrige deux épreuves. Il s'agit d'une règle générale : les corrélations positives démultiplient les erreurs de calculs algébriques de type intégral (somme, cumul, moyenne...) et réduisent les erreurs de calculs différentiels (dérivées, pentes, vitesse...). L'inverse est aussi vérifié pour les corrélations négatives.

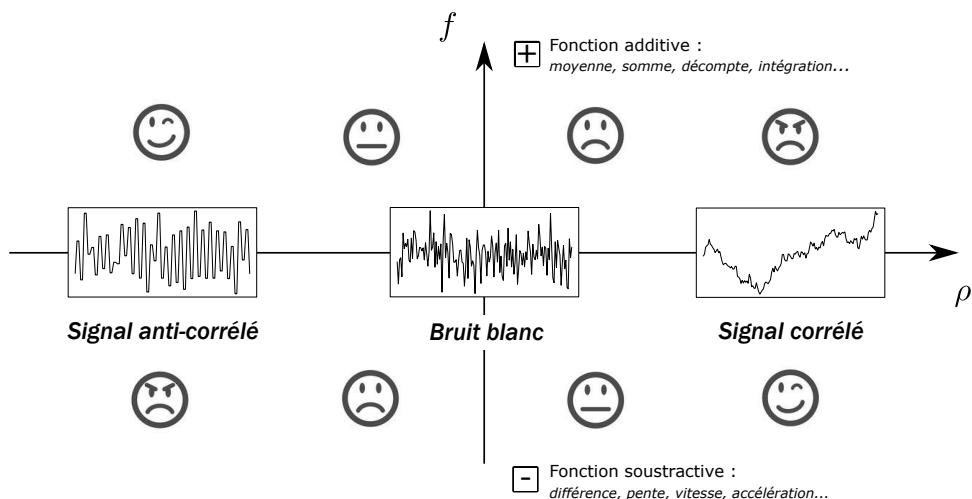


Fig. 1.5. Impact d'un bruit à corrélation négative (à gauche), décorrélé (au centre) et à corrélation positive (à droite) sur des applications de type additif (en haut) et soustractif (en bas).

1.8 Moments statistiques

Dans cette dernière section, qui pourra être passée en première lecture, nous généralisons les notions d'espérance et de variance, et nous tentons par la même occasion de montrer qu'elles ne constituent que les premiers coefficients d'une caractérisation des lois de probabilités par un développement en série de fonctions à base de quantités appelées *moments*. Cette discussion devrait nous permettre de comprendre le concept (et les limites) de la géostatistique dite linéaire.

Dans le but de simplifier la présentation, nous nous placerons à présent exclusivement dans le cas d'une variable aléatoire continue, la généralisation au cas discret étant immédiate.

Définition 14 (Moment statistique)

Soit X une variable aléatoire réelle. Alors, pour $k \in \mathbb{N}$, on définit le moment d'ordre k par :

$$m_k = \mathbb{E}[X^k]$$

Remarque : l'espérance correspond au moment m_1 . De même, le développement suivant permet d'exprimer la variance uniquement en fonction des 2 premiers moments :

$$\mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2 - 2X\mathbb{E}[X] + \mathbb{E}[X]^2] = \mathbb{E}[X^2] - 2\mathbb{E}[X]^2 + \mathbb{E}[X]^2 = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

$$\text{Var}[X] = m_2 - m_1^2 \quad (1.13)$$

Cette relation, plus fondamentale qu'il y paraît au premier abord, appelle deux observations :

- D'une part, d'un point de vue informatique, il est possible de calculer la variance (et donc l'écart-type) en une seule passe sur les données (là où l'utilisation de la formule classique nécessite 2 passes : une pour calculer la moyenne m_1 , et une pour calculer les écarts à m_1). On prendra garde cependant au risque accru d'erreurs d'arrondi numérique avec cette formulation.
- D'autre part, sur un plan plus théorique, si X représente l'amplitude d'une erreur commise, alors le terme d'erreur quadratique $\mathbb{E}[X^2]$ (aussi appelé MSE pour *Mean Square Error* dans la littérature anglo-saxonne), peut s'exprimer en fonction de deux composantes distinctes :

$$\mathbb{E}[X^2] = \mathbb{E}[X]^2 + \text{Var}[X]$$

En remarquant que $\mathbb{E}[X]$ est la moyenne des erreurs, c'est-à-dire le *biais*, on peut aussi écrire :

$$\mathbb{E}[X^2] = \text{Biais}^2 + \text{Variance} \quad (1.14)$$

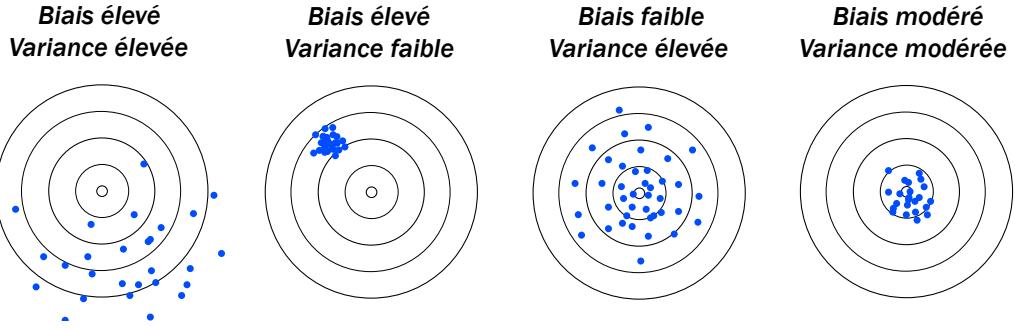


Fig. 1.6. Illustration des deux composantes de l'erreur de pointé dans un jeu de féchettes : le biais, *i.e.* l'erreur systématique, et la variance, *i.e.* l'incapacité à produire des jets identiques.

Définition 15 (Fonction génératrice des moments)

Soit X une variable aléatoire réelle. On appelle *fonction génératrice des moments*, la fonction $M_X : \mathbb{R} \rightarrow \mathbb{R}$ définie (lorsqu'elle existe) par :

$$M_X(t) = \mathbb{E}[e^{tX}]$$

On justifie *a posteriori* le nom de cette fonction par le fait que ses dérivées successives d'ordre k prises en l'origine coïncident exactement avec la suite $(m_k)_{k \in \mathbb{N}}$ des moment de X .

Propriété 16 (Fonction génératrice des moments)

Soit X une variable aléatoire réelle de fonction génératrice des moments M_X et $k \in \mathbb{N}$:

$$\left. \frac{d^k M_X(t)}{dt^k} \right|_{t=0} = \mathbb{E}[X^k]$$

Preuve : par linéarité des opérateurs $\frac{d}{dt}$ et \mathbb{E} :

$$\frac{d^k M_X(t)}{dt^k} = \frac{d^k \mathbb{E}[e^{tX}]}{dt^k} = \mathbb{E}\left[\frac{d^k e^{tX}}{dt^k}\right] = \mathbb{E}[X^k e^{tX}] \text{ et donc } \frac{d^k M_X}{dt^k}(0) = \mathbb{E}[X^k] = m_k$$

Théorème 17 (Relation univoque FGM \leftrightarrow DDP)

Soit $f : \mathbb{R} \rightarrow \mathbb{R}$ une fonction. Il existe au plus une unique densité de probabilité π admettant f pour fonction génératrice des moments, *i.e.* solution de l'équation :

$$f(t) = \int_{-\infty}^{+\infty} e^{tx} \pi(x) dx$$

Preuve : immédiate en posant $t = i\tau$ (avec $i^2 = -1$) et à l'aide de la transformation de Fourier.

Le théorème 17 revêt une importance capitale pour la compréhension du rôle des moments dans la caractérisation statistique d'une variable. Pour mieux s'en convaincre, effectuons une développement en série entière de la fonction exponentielle dans la définition 15 :

$$M_X(t) = \mathbb{E}[e^{tX}] = \mathbb{E}\left[1 + tX + \frac{t^2}{2}X^2 + \frac{t^3}{3!}X^3 + \dots\right]$$

À nouveau, par linéarité de l'espérance :

$$M_X(t) = 1 + t\mathbb{E}[X] + \frac{t^2}{2}\mathbb{E}[X^2] + \frac{t^3}{3!}\mathbb{E}[X^3] + \dots = 1 + tm + \frac{t^2m}{2} + \frac{t^3m}{3!} + \dots \quad (1.15)$$

Si on met bout-à-bout toutes les informations dont nous disposons jusqu'à présent :

- La fonction génératrice des moments se calcule par une somme ne nécessitant que la connaissance de la suite (infinie) des moments de X (équation 1.15).
- A partir de cette même fonction, le théorème 17 stipule qu'il est possible de reconstruire sans ambiguïté la densité de probabilité de X .
- L'intégrale 1.7 nous donne alors la fonction de répartition de X sur l'ensemble des réels.

Autrement dit, la suite des moments est une caractérisation complète⁵ d'une variable aléatoire réelle.

La géostatistique linéaire, qui se retrace à l'estimation de quantités inconnues par des combinaisons linéaires d'observations, ne nécessite en retour que l'emploi des 2 premiers moments. Tout se passe donc virtuellement comme si on opérait une troncature au second ordre d'un genre de développement analytique de la loi de la variable aléatoire.

1.9 Exemple de synthèse

Un usager attend le bus à une station desservie par 2 lignes différentes. Les bus y passent à un intervalle régulier de 10 minutes. Les 2 lignes sont supposées indépendantes. L'usager prendra le premier des deux bus qui arrive. Combien de temps doit-il attendre en moyenne ? On peut modéliser l'expérience par un ensemble fondamental $\Omega = [0, 10]^2$. La mesure de probabilité \mathbb{P} est celle d'une loi uniforme sur Ω . Soit X la variable aléatoire dénotant le temps d'attente de l'usager :

$$\begin{aligned} X &: [0, 10]^2 &\rightarrow \mathbb{R} \\ (\omega_1, \omega_2) &\mapsto \min(\omega_1, \omega_2) \end{aligned} \quad (1.16)$$

Soit $x \in [0, 10]$ une réalisation de la variable de temps d'attente. Une analyse graphique rapide (*cf* fig. 1.7) permet de déterminer la valeur de $F_X(x)$:

⁵Sous réserve d'existence de la fonction génératrice M_X , sinon l'unicité n'est pas garantie sur \mathbb{R} . On pourra trouver un contre-exemple avec la loi log-normale. C'est l'objet du problème de Hamburger (1920).

$$F_X(x) = P_X([0, x]) = \frac{x^2 + x(10 - x) + (10 - x)x}{100} = \frac{x(20 - x)}{100}$$

En particulier, on s'assure bien que F_X vérifie les propriétés des fonctions de répartitions : F_X est croissante, $F_X(-\infty) = F_X(0) = 0$ et $F_X(\infty) = F_X(10) = 1$.

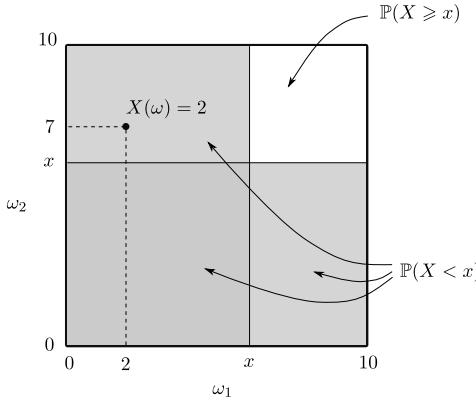


Fig. 1.7. Illustration de l'espace fondamental Ω . Chaque point $\omega = (\omega_1, \omega_2)$ représente une réalisation aléatoire, avec ω_i le temps d'attente du bus i . $X(\omega) = \min(\omega_1, \omega_2)$ est le temps d'attente de l'usager.

Plus généralement, on peut transposer la loi \mathbb{P} sur l'ensemble fondamental à une loi P_X sur la variable aléatoire X , définie pour tout intervalle $[a, b] \in [0, 10]$ par :

$$P_X([a, b]) = \mathbb{P}(\{(\omega_1, \omega_2) \in \Omega ; a \leq \min(\omega_1, \omega_2) \leq b\}) = \mathbb{P}([a, b]^2) = \frac{a^2 - b^2 + 20(b - a)}{100}$$

S'ensuit alors la densité de probabilité : $p_X(x) = \frac{dF_X}{dx}(x) = \frac{10-x}{50}$

Puis l'espérance du temps d'attente :

$$\mathbb{E}(X) = \int_0^{10} x p_X(x) dx = \frac{1}{50} \int_0^{10} x(10 - x) dx = \frac{10}{3}$$

L'usager attendra donc en moyenne 3 min 20. On peut évaluer de la même manière l'incertitude autour de cette valeur moyenne à l'aide de la notion de variance. On utilise la formule 1.13 : $\text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$ (pour profiter du fait que nous avons déjà calculé l'espérance).

$$\mathbb{E}[X^2] = \int_0^{10} x^2 p_X(x) dx = \frac{1}{50} \left[\frac{x^3}{3} - \frac{x^3}{4} \right]_0^{10} = \frac{50}{3}$$

D'où : $\text{Var}[X] = \frac{50}{3} - \frac{100}{9} = \frac{50}{9}$. On en déduit alors l'incertitude $\sigma_X = \sqrt{\text{Var}[X]} = \frac{5\sqrt{2}}{3} \approx 2.35$, soit un écart-type sur le temps d'attente de 2 min 21.

A titre d'exercice, on généralisera cette étude au cas d'un arrêt desservi par n lignes de bus. Par ailleurs, on peut pousser le scénario un peu plus loin, en supposant que le premier bus arrivant est systématiquement complet. L'usager doit alors attendre le second bus.

On pose $Y : (\omega_1, \omega_2) \mapsto \max(\omega_1, \omega_2)$ la variable aléatoire qui à une réalisation donnée associe le temps d'attente de cet usager. Par symétrie du problème, on trouve aisément la densité de Y :

$$p_Y(y) = 10 - p_X(y) = \frac{y}{50}$$

puis son espérance $\mathbb{E}[Y] = 10 - \mathbb{E}[X] = \frac{20}{3}$ et sa variance $\text{Var}[Y] = \text{Var}[X] = \frac{50}{9}$.

On peut également exprimer la covariance des temps d'attente entre les deux cas de figures. À nouveau, de manière similaire à l'équation 1.13, on cherche une expression alternative pour $\text{Cov}(X, Y)$, afin de réutiliser les résultats trouvés pour $\mathbb{E}[X]$ et $\mathbb{E}[Y]$:

$$\begin{aligned} \text{Cov}(X, Y) &= \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] \\ &= \mathbb{E}[XY - Y\mathbb{E}[X] - X\mathbb{E}[Y] + \mathbb{E}[X]\mathbb{E}[Y]] \\ &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \end{aligned} \tag{1.17}$$

Il ne nous reste alors plus qu'à évaluer l'espérance de la variable $Z = XY$. On doit exprimer la densité de probabilité $p(x, y)$. Malheureusement, X et Y n'étant pas indépendantes comme nous allons le voir, $p(x, y) \neq p(x)p(y)$. En revanche on peut écrire : $p(x, y) = p(x|y)p(y)$ où $p(x|y)$ est la **probabilité conditionnelle** de X sachant Y . Par exemple, si on sait que le second bus (Y) est arrivé après 8 minutes d'attente, alors on en déduit que le temps d'attente du premier bus suit nécessairement une loi uniforme entre 0 et 8 : $Y \sim \mathcal{U}([0, 8])$. D'où :

$$p(x|y) \sim \mathcal{U}([0, y]) \quad \text{et donc : } p(z) = p(x, y) = \left(\frac{1}{y}\right) \times \left(\frac{y}{50}\right) = \frac{1}{50}$$

On en déduit alors l'espérance de la variable $Z = XY$:

$$\mathbb{E}[Z] = \int_{\mathbb{R}} z p(z) dz = \int_{y=0}^{10} \int_{x=0}^y xy p(x, y) dx dy = \frac{1}{50} \int_{y=0}^{10} \int_{x=0}^y xy dx dy = 25$$

Par ailleurs, on évalue $\mathbb{E}[X]\mathbb{E}[Y] = \left(\frac{10}{3}\right)\left(\frac{20}{3}\right) = \frac{200}{9}$, ce qui nous permet de terminer le calcul :

$$\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 25 - \frac{200}{9} = \frac{25}{9}$$

Au passage, on peut déterminer la corrélation entre ces deux variables :

$$\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} = 50\%$$

Enfin, supposons que deux personnes attendent à la station, et que le bus ne contienne plus qu'une place libre. Calculons le retard R pris par l'un des usagers par rapport à l'autre.

Avec les notations utilisées jusqu'ici, on a $R = Y - X$. On cherche à calculer la moyenne et la variance de R . Nous disposons à ce stade de toutes les informations nécessaires. D'un part, l'espérance étant un opérateur linéaire, on peut écrire (quand bien même X et Y sont corrélées) :

$$\mathbb{E}[R] = \mathbb{E}[Y - X] = \mathbb{E}[Y] - \mathbb{E}[X] = \frac{20}{3} - \frac{10}{3} = 3.33 \text{ min}$$

Pour la variance, on utilise la formule de propagation donnée par le théorème 12 :

$$\text{Var}[Y - X] = \text{Var}[Y] + \text{Var}[X] - 2\text{Cov}(X, Y) = \frac{50}{9} + \frac{50}{9} - 2 \times \frac{25}{9} = \frac{50}{9}$$

On a donc un écart-type de 2 min 21.

On remarque que la moyenne et l'écart-type de l'avance d'un usager par rapport à l'autre, sont strictement égaux à ceux du temps d'attente de l'usager qui prend le premier bus, ce qui paraît intuitif. Tout se passe comme si R correspondait au temps d'attente d'un usager arrivant à la station au moment du départ de X .

Chapter 2

Analyse variographique

En tant qu'outil central de la Géostatistique, le *variogramme* constitue une signature probabiliste (partielle mais suffisante) de l'autocorrélation des phénomènes. Dans ce chapitre, nous définissons en premier lieu plus rigoureusement la notion de phénomène à l'aide des processus stochastiques. Nous établirons alors les conditions sous lesquelles il est possible d'en extraire un variogramme, dont nous étudierons les propriétés théoriques, avant de présenter la méthode standard d'estimation du variogramme (inconnu par nature) à partir d'un semis d'observations ponctuelles.

2.1 Processus stochastique

Reprendons notre problème modèle : nous disposons d'un semis d'observations ponctuelles d'altitudes z_1, z_2, \dots, z_n mesurées par GPS sur un domaine \mathcal{D} (par exemple un massif montagneux délimité). L'objectif du problème de production d'un Modèle Numérique de Terrain (MNT) consiste à estimer une fonction z retournant l'altitude en chaque point de \mathcal{D} et telle qu'en chaque point x_i sur lequel une observation z_i a été effectuée, la contrainte $z_i = z(x_i)$ soit respectée. Il s'agit d'un problème classique d'**interpolation spatiale**.

Nous l'avons dit en introduction, il existe un certain nombre de méthodes déterministes permettant de traiter ce problème : la méthode du plus proche voisin, l'interpolation linéaire, bilinéaire, pondérée par l'inverse des distances aux sites d'observations, les splines de lissages et de régression... ([Mitas et Mitasova, 1999](#); [Arnaud et Emery, 2000](#)). Toutes ces méthodes présentent l'avantage de ne nécessiter aucune hypothèse forte sur le phénomène à estimer (ici le relief du terrain). Nommément, la seule supposition effectuée est que l'altitude sur le domaine \mathcal{D} puisse être décrite par une fonction $z : \mathcal{D} \rightarrow \mathbb{R}$. En un couple de coordonnées cartographiques x donné, il ne peut donc y avoir qu'une seule altitude correspondante $z(x)$, ce qui interdit donc la modélisation des dévers, grottes, sous-terrains, tunnels... Cette fonction z est appelée une **variable régionalisée**.

En Géostatistique, nous devons ajouter une seconde couche d'abstraction à ce modèle primaire, afin de prendre en compte la nature stochastique du phénomène étudié. Nous supposons à présent que la variable régionalisée z est une réalisation $Z(\omega)$ d'une variable aléatoire fonctionnelle Z . Nous allons voir par la suite que ce niveau supplémentaire d'abstraction nécessite de poser des hypothèses un peu plus lourdes sur le phénomène.

Définition 18 (Processus stochastique)

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, et \mathcal{D} un espace quelconque. On appelle *processus stochastique* toute fonction Z :

$$Z : \mathcal{D} \times \Omega \rightarrow \mathbb{R}$$

Remarque : dans le problème modèle, \mathcal{D} est un sous-ensemble de \mathbb{R}^2 .

Cette variable aléatoire Z est appelée *fonction aléatoire* (FA), *processus stochastique* ou encore *signal aléatoire*. Quand le domaine \mathcal{D} contient plusieurs dimensions, on parle aussi de *champ aléatoire*.

La fonction Z est une fonction de 2 variables : x définie sur un domaine analytique quelconque (*i.e.* sur un espace physique perceptible, par exemple l'espace géographique) et ω définie sur une espace probabilisé. L'espace produit cartésien $\mathcal{D} \times \Omega$ est donc un espace dit *topo-probabiliste*. Il peut être instructif de regarder ce qu'il advient quand on fixe l'une des 2 variables :

- Pour une éventualité $\omega \in \Omega$ donnée, la réalisation $z(\cdot) = Z(\cdot, \omega)$ est une **trajectoire** du processus. C'est une fonction classique de $\mathcal{D} \rightarrow \mathbb{R}$: la **variable régionalisée**.
- En un lieu $x \in \mathcal{D}$ donné, la variable $Z(x, \cdot)$ est une variable aléatoire réelle classique, dont la valeur ne dépend plus que de la réalisation ω considérée.

Remarquons également qu'un processus stochastique n'est que la généralisation continue de la notion de vecteur aléatoire. Nous avons vu dans le chapitre 1 qu'il peut être avantageux de regrouper un ensemble de variables aléatoires Z_1, Z_2, \dots, Z_n au sein d'un unique vecteur $\mathbf{Z} = [Z_1, Z_2, \dots, Z_n]^T$. Dans le cadre d'un processus stochastique, les variables individuelles sont indexées par les positions (continues) du domaine : $Z(x_1), Z(x_2), \dots$. Informellement, Z peut donc être considéré comme un vecteur aléatoire contenant une infinité d'éléments.

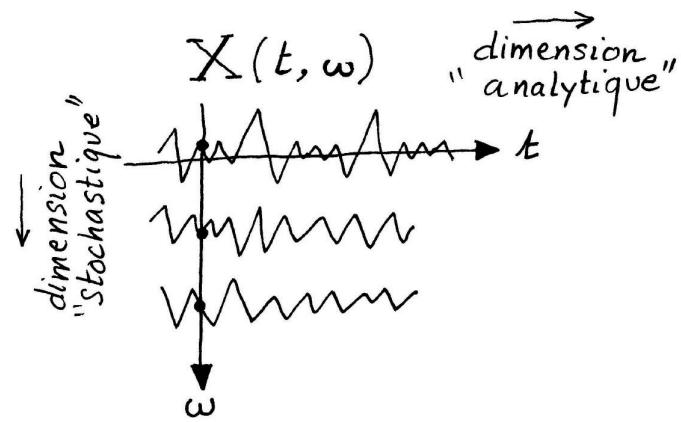


Fig. 2.1. 3 exemples de réalisations d'un processus stochastique $X : \mathbb{R} \times \Omega \rightarrow \mathbb{R}$, où \mathbb{R} représente la dimension *analytique* et Ω (en pratique non-spécifié) est la dimension *stochastique* que l'on peut considérer informellement comme un ensemble d'univers parallèles. Source : [Antoni \(2013\)](#).

Pour réaliser un processus stochastique, un protocole simple et ne nécessitant que peu de matériel (et que l'on pourra même faire en pensée si on ne souhaite pas heurter sa sensibilité écologique), consiste

à prendre une feuille de papier, à la froisser, puis à essayer de la remettre à peu près à plat sur la table. Le motif obtenu est alors parfaitement aléatoire, et en réitérant l'expérience, on obtiendra en toute probabilité une géométrie de pliage différente. Malgré tout, la forme et la longueur typique des lignes de brisure (*i.e.* certaines des caractéristiques d'autocorrélation spatiale du processus de froissage) vont rester fortement similaires d'une réalisation à l'autre. La même observation pourrait être effectuée sur le processus de formation du relief montagneux (si toutefois nous pouvions rejouer le temps en quelques sortes). Dans les deux cas d'exemple, les forces globales externes (pression des mains sur la feuille, convergence des plaques tectoniques) sont sensiblement les mêmes. C'est la présence de petites variations locales qui vont générer l'aléatoire des réalisations. Tout au long de ce cours, il pourra être intéressant de revenir à cette expérience de la feuille de papier, qui contrairement au processus orogénique, présente l'avantage de pouvoir être conceptuellement rejouée).

La figure 2.2 donne un exemple de 8 réalisations d'un processus stochastique de MNT. En réalité, une et une seule de ces réalisations correspond au relief d'une zone existant réellement. Les 7 autres sont des simulations informatiques programmées pour reproduire les caractéristiques statistiques du champ réel. La difficulté certaine à identifier le champ modèle parmi ces 8 réalisations, valide l'efficacité de la simulation. La réponse au problème sera donnée dans le chapitre 4.

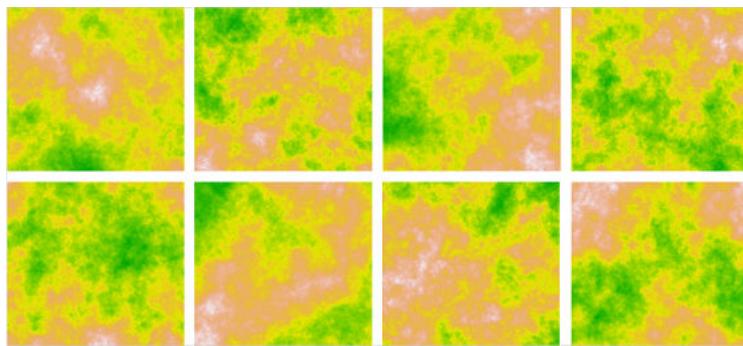


Fig. 2.2. 8 réalisations d'un processus stochastique $Z : \mathbb{R}^2 \times \Omega \rightarrow \mathbb{R}$ pour 8 tirages $\omega_1, \omega_2, \dots, \omega_8$. Le champs $z(\cdot) = Z(\cdot, \omega_i)$ est la carte de relief générée par la i -eme réalisation. La quantité $Z(x, \cdot)$ est une variable aléatoire traduisant l'incertitude sur l'altitude au lieu $x \in \mathbb{R}^2$ de la carte.

Activité I. Simulation informatique d'un processus stochastique

Bien évidemment, la nature finie de l'espace mémoire de la machine ne nous permet que de créer des processus discrets. Cette activité devrait cependant permettre de rendre plus concret la notion de processus stochastique.

Q1. On appelle *marche aléatoire* sur \mathbb{R} à temps discret¹ la réalisation d'une suite de variables aléatoires X_0, X_1, X_2, \dots définie à chaque pas de temps $t \in \mathbb{N}$ par :

$$X_t = \begin{cases} 0 & \text{si } t = 0 \\ X_{t-1} + \varepsilon_t & \text{sinon.} \end{cases} \quad (2.1)$$

où $\varepsilon_t \sim \mathcal{N}(\mu, \sigma^2)$ est une suite de variables aléatoires décorrélées, distribuées suivant une loi normale de moyenne μ et d'écart-type σ .

De manière informelle, X représente la position d'une personne se déplaçant sur un axe uni-dimensionnel en partant d'un position arbitrairement fixée à 0. À chaque pas de temps t , elle peut se déplacer d'une

¹On parle aussi de manière imagée de *marche de l'ivrogne*.

quantité ε_t (en arrière si $\varepsilon_t < 0$ et en avant sinon). La quantité ε_t est échantillonnée aléatoirement (et à chaque pas de temps t) suivant la loi normale (*cf* fig. 1.3). On pourra trouver une analyse théorique de ce processus dans Barret (2009).

Remarque : observons que ε_t est lui même un processus stochastique. Il est entièrement décorrélé et se rencontre sous le nom de *bruit blanc* dans littérature (Picinbono, 1998)).

On écrira une fonction `random_walk(N,m,s)` prenant en entrée le nombre N de pas de temps à générer ainsi que les paramètres (moyenne m et écart-type s) du processus de déplacement ε_t .

Script 1. Génération d'une marche aléatoire [code1.r]

```
random_walk = function(N,m,s){
  X = rep(0, N)                                # Initialisation

  for (i in 2:N){
    X[i] = X[i-1] + rnorm(1,m,s)      # Simulation
  }

  return(X)
}
```

Le code ci-dessus est une transcription directe de la définition 2.1. La fonction R `rnorm(n,m,s)` génère un vecteur de n réalisations d'une loi normale de moyenne m et d'écart-type s .

Lorsqu'un signal aléatoire est défini de manière récursive, à l'instar de 2.1, on parle de *représentation markovienne* du processus. Un processus admettant une représentation markovienne est qualifiée de *markovien*. Remarquons que 2.1 peut être programmé de manière plus concise à l'aide de la fonction `cumsum(Y)` qui effectue une somme cumulée des termes de Y : $X = \text{cumsum}(rnorm(N,m,s))$.

Q2. Après avoir spécifié les valeurs des paramètres (par exemple `N=500`, `m=0` et `s=1` pour commencer) représenter une réalisation du processus à l'aide de la fonction `plot` :

```
plot(random_walk(N,m,s), type='l')
```

Faire varier les paramètres et observer les réalisations produites. En particulier, on pourra modifier la tendance globale avec le paramètre m .

Q3. Pour un triplet de paramètres (N,m,s) fixé, représenter (sur le même graphe) 10 réalisations du processus stochastique X .

Q4. On se place au niveau d'un époque donnée `tps` (par exemple 150). Compiler dans un unique vecteur les valeurs prises par le processus X à l'abscisse `tps` sur 10 réalisations différentes.

Q5. Représenter la distribution des valeurs prises par le processus X à l'abscisse `tps`. Si L est le vecteur contenant les réalisations de X_{tps} , on pourra utiliser la commande : `plot(density(L))`.

Script 2. Génération d'une marche aléatoire [code1.r]

```

N = 500; m = 0; s = 1 # Paramètres

par(mfrow=c(1,2)) # Pour tracer 2 graphes

tps = 150; L=c()

plot(0,xlim=c(0,500),ylim=c(-100,100)) # On initialise le graphique

for (realisation in 1:10){ # On génère 10 réalisations
  X = random_walk(N,m,s)
  L = c(L, X[tps])
  lines(X) # Représentation graphique
}

abline(v=tps, lty=2)
plot(density(L)) # Distribution de X[tps]

```

Pour aller plus loin, on pourra étendre le code pour représenter simultanément les distributions au niveau de deux époques différentes, par exemple 100 et 350. La figure 2.3 illustre le résultat obtenu.

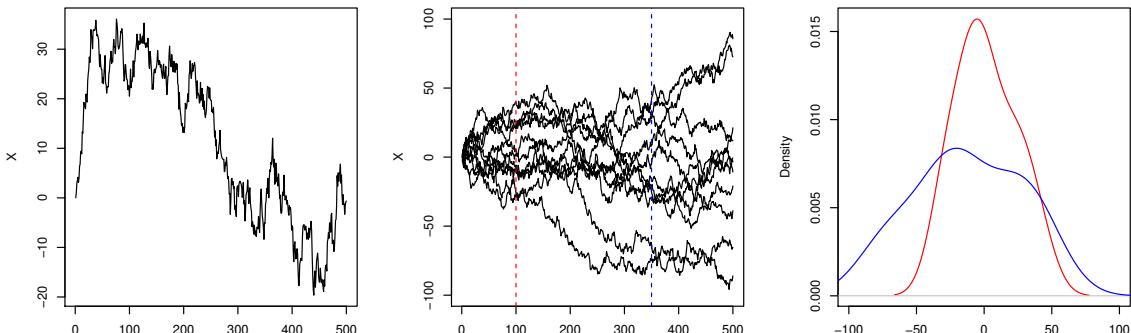


Fig. 2.3. Résultat de code1.r : génération d'une réalisation du processus aléatoire (à gauche), génération de 10 réalisations (au centre) avec les coupes à $t = 100$ (en rouge) et $t = 350$ (en bleu), et distributions de probabilités des variables X_t correspondantes (à droite).

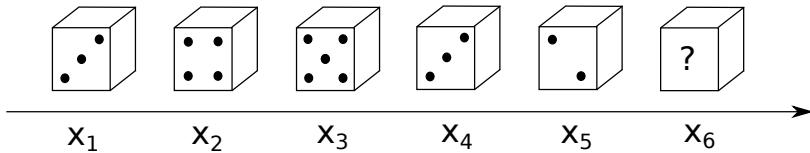
La figure 2.3 illustre bien le concept d'espace topo-probabiliste. À gauche, la réalisation $\omega \in \Omega$ est fixée, et on observe l'évolution de la réalisation sur son domaine analytique. À droite le lieu du domaine est fixé : $t = 100$ (en rouge) ou $t = 350$ (en bleu), et on observe la variabilité stochastique de X_t sur Ω , *i.e.* la densité de probabilité $\pi(X_t)$ à t fixé. Au centre, on essaye de combiner la représentation des deux espaces (\mathbb{R} en abscisse et Ω en superposant les courbes).

Par ailleurs, on observe que la variance $\text{Var}[X_t]$ augmente à mesure que t augmente (ce qui est assez intuitif, la position du mobile est d'autant plus incertaine que le temps passe). En revanche l'espérance $\mathbb{E}[X_t]$ est constante et égale à 0. Nous dirons par la suite que le processus est *stationnaire* à l'ordre 1 (*i.e.* pour son moment d'ordre 1, l'espérance) mais *non-stationnaire* à l'ordre 2.

Nous programmerons d'autres processus stochastiques plus loin.

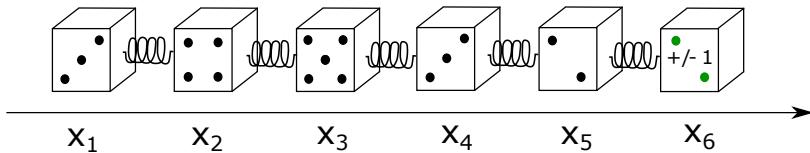
2.2 Une première approche informelle

Dans cette section, nous allons essayer d'acquérir une compréhension empirique de l'approche employée par la Géostatistique, en particulier pour résoudre le problème d'interpolation. Pour ce faire, nous utilisons une comparaison inspirée de [Arnaud et Emery \(2000\)](#) : on considère 6 expériences aléatoires (que nous représenterons de manière schématique par une collection de dés à jouer), dont la réalisation est observé pour 5 d'entre elles (x_1 à x_5) et inconnue pour la dernière (x_6). On nous demande de pronostiquer le résultat de cette dernière expérience.



En l'état des connaissances, on serait bien en peine de répondre. Probablement, le mieux que l'on puisse faire pour minimiser les risques d'un écart trop important à la vérité, ce serait de parier sur l'espérance a priori de chaque expérience aléatoire individuelle (*i.e.* 3.5 pour le cas du dé à 6 faces), ce qui est avouons-le, un prognostic relativement pauvre. L'écart-type associé, on pourra le vérifier à titre d'entraînement, vaut 1.71.

En revanche, qu'en est-il si on sait à présent que les résultats de l'expérience sont spatialement corrélés, ce que nous modéliserons graphiquement par des ressorts entre les différentes réalisations ?



On va alors intuitivement chercher à regarder ce qui se passe sur la portion observée pour quantifier cette corrélation. On y observe que les valeurs prises par deux réalisations voisines ne diffèrent (en règle générale) pas plus d'une unité. Une estimation sur x_6 serait alors vraisemblablement 2 ± 1 .

Généralisons cette analogie pour le cas de l'interpolation des valeurs d'un MNT (fig. 2.4). Le semis de relevés d'altitudes nous permet d'étudier les corrélations statistiques du phénomène.

Dans une seconde phase, le modèle de corrélation ainsi inféré va être utilisé pour prédire plus finement les valeurs inconnues (fig. 2.5). Nous étudierons cette seconde étape dans le chapitre 3.

Cette manière de procéder suppose toutefois deux hypothèses importantes :

- Collecter des informations aux points observés pour les transférer aux sites inconnus nécessite une uniformité des propriétés statistiques du phénomène sur tout le domaine d'étude. C'est la **stationnarité**.
- Nous disposons initialement de n mesures d'altitude, mais il ne faut pas oublier que ces observations appartiennent toutes à une unique réalisation $Z(\omega)$ du processus stochastique. Pouvoir inférer des informations de nature statistique à partir d'une unique réalisation n'est absolument

pas trivial, et requiert une seconde hypothèse un peu plus technique : l'**ergodicité**.

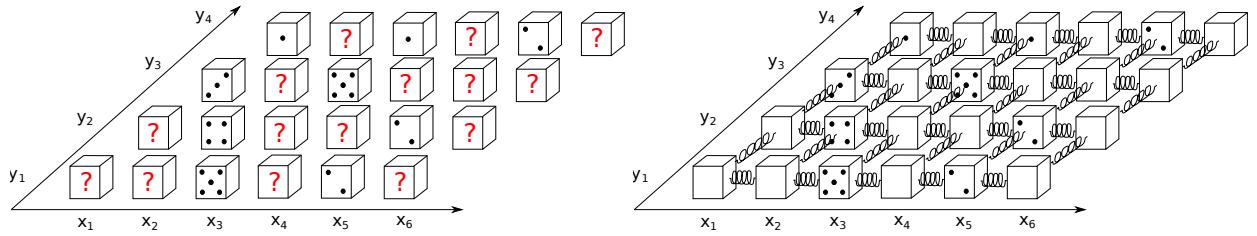


Fig. 2.4. À gauche : problème d'interpolation consistant à trouver les valeurs des réalisations inconnues (en rouge) à partir des sites observés (en noir). À droite : étude des corrélations entre les réalisations (uniquement à l'aide des observations).

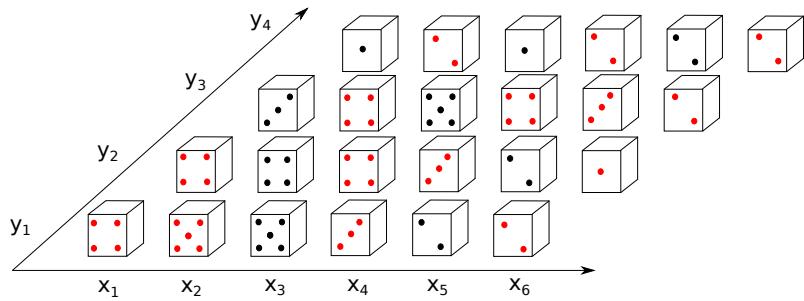


Fig. 2.5. Résultat de la procédure d'interpolation

2.3 Les hypothèses fondamentales

Un processus Z est entièrement déterminé par ses lois jointes *fini-dimensionnelles* :

$$F_{\mathbf{X}}(z_1, \dots, z_k) = \mathbb{P}[Z(x_1) \leq z_1, \dots, Z(x_k) \leq z_k] \quad (2.2)$$

où $\mathbf{X} = [x_1, x_2, \dots, x_n]$ est un vecteur (de dimension finie) répertoriant les n sites concernés par la loi fini-dimensionnelle. Remarquons que pour un domaine \mathcal{D} de taille finie p , il y a autant de lois $F_{\mathbf{X}}$ que de vecteurs \mathbf{X} possibles, à savoir 2^p , le nombre de partitions de \mathcal{D} . Pour un domaine continu (e.g. \mathbb{R}), la connaissance de la famille (infinie) de ces lois est impossible en pratique, et on choisit de caractériser Z uniquement à l'aide de ses 2 premiers moments : moyenne et covariance.

2.3.1 La Stationnarité

Informellement, la propriété de stationnarité d'un processus stochastique décrit l'uniformité de ses propriétés statistiques sur son domaine de définition analytique. Cela ne signifie en aucun cas qu'il est constant, mais seulement que ses réalisations ont les *mêmes chances de produire les mêmes motifs* en tout lieu x du domaine.

On distingue plusieurs types de stationnarité en fonction des propriétés ciblées.

Définition 19 (Stationnarité au sens fort (SSS))

Un processus stochastique est dit *stationnaire au sens fort*, ou au *sens strict* si, et seulement si, toutes ses lois jointes fini-dimensionnelles sont invariantes par translation $h \in \mathcal{D}$:

$$F_{\mathbf{X}}(z_1, z_2, \dots, z_n) = F_{\mathbf{X}+h}(z_1, z_2, \dots, z_n)$$

Remarque : h est un vecteur de même dimension que les points du domaine \mathcal{D} . Donc, dans le cas du problème modèle du MNT, h est un vecteur 2D, avec deux composante Est-Ouest et Sud-Nord.

L'hypothèse de stationnarité au sens fort est beaucoup trop lourde en pratique, puisque impossible à vérifier, même de manière approchée.

Définition 20 (Stationnarité à l'ordre 2 (WSS))

Un processus stochastique est dit *stationnaire au sens faible*, ou au *sens large*, si et seulement si, ses deux premiers moments existent et sont invariants par translation $h \in \mathcal{D}$:

- (i) $\mathbb{E}[Z(x)] = \mathbb{E}[Z(x + h)]$
- (ii) $\text{Cov}[Z(x_1), Z(x_2)] = \text{Cov}[Z(x_1 + h), Z(x_2 + h)]$

Le point (i) peut être reformulé comme suit : il existe une valeur m telle que $\mathbb{E}[Z(x)] = m$ pour tout point $x \in \mathcal{D}$. Par ailleurs, un processus vérifiant cette condition (i), sans nécessairement vérifier (ii), est qualifié de **stationnaire à l'ordre 1**.

Propriété 21 (Stationnarité à l'ordre 2)

Soit Z un processus stationnaire au sens faible, alors sa covariance ne dépend que du vecteur séparant les sites :

$$\text{Cov}[Z(x_1), Z(x_2)] = f(x_1 - x_2)$$

Preuve : Z étant stationnaire à l'ordre 2, sa covariance est invariante par translation $h \in \mathcal{D}$: $\text{Cov}[Z(x_1), Z(x_2)] = \text{Cov}[Z(x_1 + h), Z(x_2 + h)]$. En particulier, on peut choisir $h = -x_2$ et on obtient alors : $\text{Cov}[Z(x_1), Z(x_2)] = \text{Cov}[Z(x_1 - x_2), Z(x_2 - x_2)] = \text{Cov}[Z(x_1 - x_2), Z(0)]$. Cette fonction ne dépend que de $x_1 - x_2$, le vecteur séparant les deux sites, d'où le résultat.

Cette propriété (et sa démonstration) mettent clairement en évidence que la covariance des valeurs prises pas le phénomène entre deux sites séparés d'un vecteur \vec{v} , est égale à la covariance entre la valeur prise en l'origine et celle prise en un point \vec{v} . Autrement dit, la *ressemblance* statistique entre les valeurs prises en deux sites, ne dépend que de leur position relative, et non de l'origine du repère.

À titre d'exemple, le processus de marche aléatoire étudié dans l'activité I (fig. 2.3) est stationnaire à l'ordre 1 (la valeur moyenne des signaux est uniforme) mais non-stationnaire à l'ordre 2 (la variance,

traduite ici par la dispersion des réalisations, augmente le long de l'axe des abscisses). Il n'est donc pas stationnaire au sens faible, et *a fortiori*, il n'est pas non plus stationnaire au sens strict. Nous démontrerons rigoureusement cette affirmation plus loin.

Par commodité, nous ajoutons une hypothèse subsidiaire, qui n'est pas indispensable (et pourra facilement être relaxée), mais qui simplifiera la présentation : l'**isotropie**.

Définition 22 (Isotropie)

Soit Z un processus stationnaire au sens faible. On dit que Z est *isotrope* si, et seulement si, sa covariance ne dépend que de la distance séparant les sites :

$$\text{Cov}[Z(x_1), Z(x_2)] = C(\|x_1 - x_2\|)$$

On appelle **covariogramme** cette fonction C . C'est une fonction classique de \mathbb{R}^+ dans \mathbb{R}^+ , prenant en entrée la distance h_{ij} séparant deux sites x_i et x_j et retournant la covariance des valeurs prises par le phénomène en ces deux sites : $\text{Cov}[Z_i, Z_j]$. On donne un exemple de covariogramme en figure 2.6.

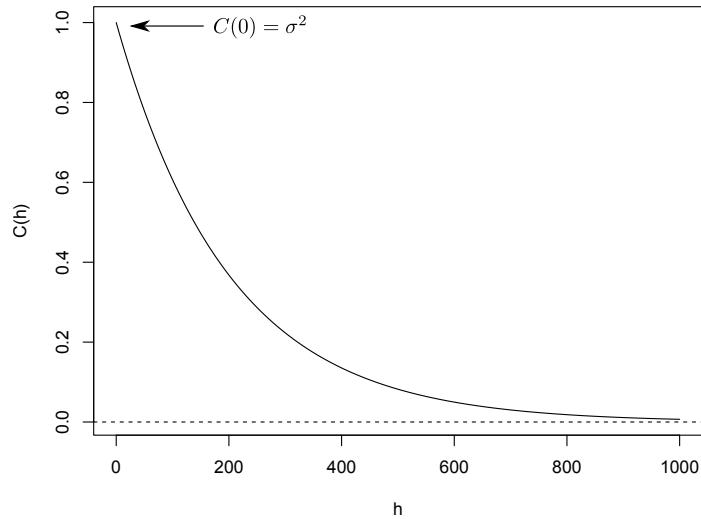


Fig. 2.6. Pour une distance h donnée séparant 2 lieux, le covariogramme $C(h)$ désigne la covariance (ou si on préfère la ressemblance) entre les valeurs prises par le phénomène en ces 2 lieux.

On a systématiquement $C(0) = \sigma_Z^2$, l'écart-type du processus. En effet, $C(0) = \text{Cov}[Z(x), Z(x)] = \text{Var}[Z(x)] = \text{Var}[Z]$ (par stationnarité du processus). En général², la fonction C tend vers 0 en l'infini (les variables tendent à se décorrélérer à mesure qu'elles sont éloignées géographiquement).

Plus le covariogramme décroît rapidement, plus la structure du phénomène sous-jacent est locale. Par exemple, on peut parfaitement imaginer que la figure 2.6 représente le covariogramme du relief sur une zone montagneuse. Pour $h > 600$ m, on observe que la corrélation entre sites est quasi-nulle. Cette valeur critique (que nous appellerons *portée* par la suite) correspond typiquement à la taille

²Il existe des contre-exemples artificiels, mais cette propriété est vraie en général pour des phénomènes *naturels*.

caractéristique d'une structure : vallée, colline...

2.3.2 L'ergodicité

Pour estimer le covariogramme en pratique, nous devons partir des observations z_1, z_2, \dots, z_n , qui ne sont que des mesures ponctuelles prélevées sur une **unique** réalisation. Il est important ici de bien saisir l'importance de cette limitation. Dans le cadre de la Géostatistique, les données collectées sur la variable régionalisée (altitudes, températures...) sont le résultat d'un double échantillonnage, d'abord dans l'espace probabiliste Ω , puisque seule une réalisation est disponible, puis dans l'espace analytique \mathcal{D} (*i.e.* l'espace physique réel) puisqu'on ne peut observer qu'un nombre fini de sites.

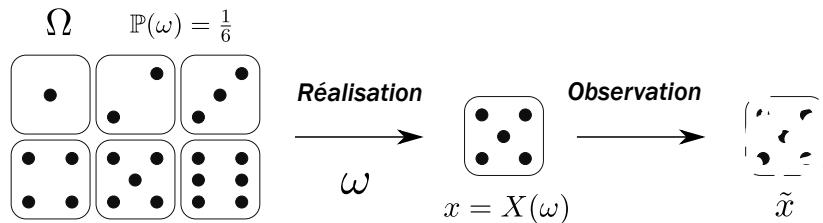


Fig. 2.7. Tout l'enjeu du problème consiste à inférer des informations statistiques sur un phénomène (modélisé par le processus stochastique X) à partir des données disponibles \tilde{x} , qui ne sont que des observations partielles d'une unique réalisation de ce phénomène.

Dès lors, se pose la question de savoir comment déduire des informations de nature statistique à partir d'une unique répétition d'une expérience aléatoire. L'idée centrale va consister (lorsque c'est possible) à compenser le manque de réalisations par l'étendue spatiale des informations collectées.

Imaginons que nous puissions observer une réalisation d'un processus (supposons par exemple un MNT), sur une domaine très grand (*e.g.* sur un carré de plusieurs dizaines de km de côté). Si le domaine d'étude \mathcal{D} est suffisamment grand, on peut alors envisager de le segmenter en un certain nombre de zones, nous permettant ainsi d'obtenir virtuellement plusieurs réalisations (quasi-indépendantes) du processus.

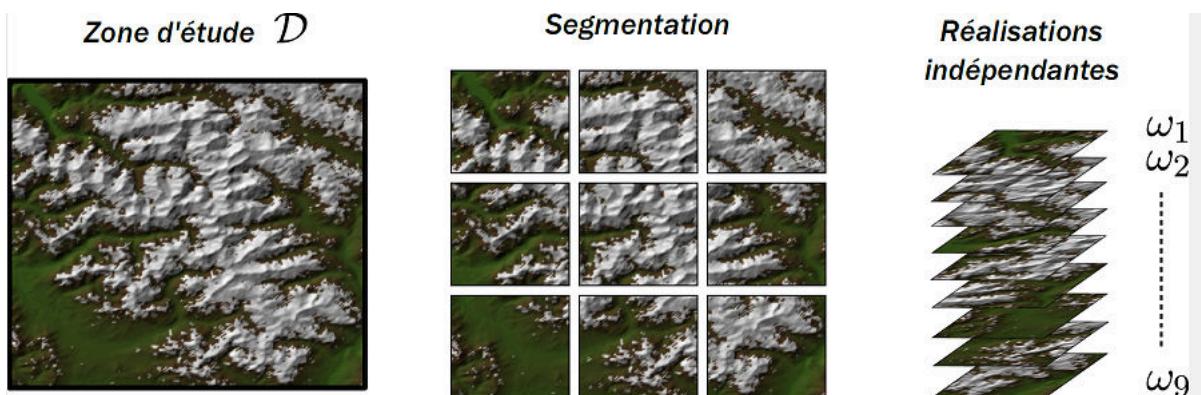


Fig. 2.8. Crédit artificielle de plusieurs réalisations $\omega_1, \omega_2, \dots$ du phénomène par segmentation.

Remarquons que le domaine initial doit être suffisamment grand pour que cette approximation ait une chance d'être valide. En effet, si par exemple on segmente à nouveau la réalisation ω_9 (portion du coin sud-est) du domaine de la figure 2.8, on sent bien intuitivement que les sous-zones qui en résulteront

ne posséderont alors plus les mêmes caractéristiques statistiques que ω_9 .

En quelques sortes, on échange de l'information sur le domaine géographique \mathcal{D} , contre de l'information sur l'espace probabiliste Ω . Pour que cette substitution soit rigoureusement correcte, le processus étudié doit posséder une propriété statistique supplémentaire : l'**ergodicité**.

L'hypothèse ergodique³ consiste à admettre que l'étude d'un processus sur son domaine analytique (e.g. au cours du temps pour un signal, ou sur l'espace géographique pour un MNT), apporte rigoureusement la même information qu'une suite de réalisations aléatoires.

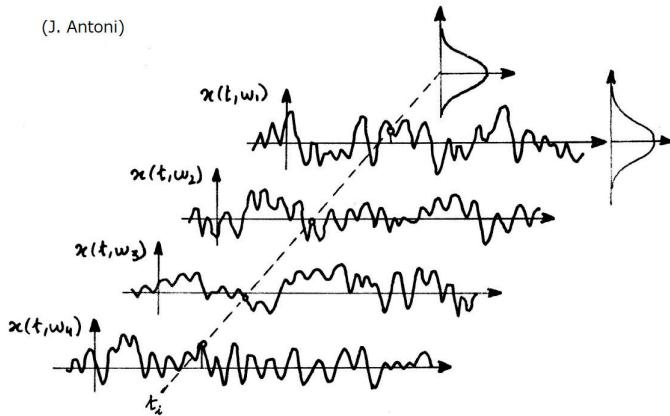


Fig. 2.9. Illustration du concept d'ergodicité : la distribution statistique des valeurs prises au cours du temps par une réalisation quelconque du signal, tend à devenir égale à la loi de probabilité d'un échantillon $X(t_i)$ pris à un temps t_i quelconque. Source : [Antoni \(2013\)](#)

Cette définition de l'ergodicité, dite *ergodicité au sens strict* est beaucoup trop forte en pratique. Nous nous restreindrons aux notions d'ergodicité au premier et second ordre.

Définition 23 (Ergodicité à l'ordre 1)

Soit Z un processus stochastique. On dit que Z est *ergodique à l'ordre 1*, ou encore *ergodique pour la moyenne*, si et seulement si, pour toute réalisation z du processus :

$$\frac{1}{|\mathcal{D}|} \int_{\mathcal{D}} z(x) dx \xrightarrow{|\mathcal{D}| \rightarrow \infty} \int_{\Omega} Z(x, \omega) d\mathbb{P}(\omega) = \mathbb{E}[Z]$$

avec $|\mathcal{D}|$ qui désigne la taille de \mathcal{D} (longueur, surface, volume...).

Autrement dit, la moyenne spatiale des échantillons d'une réalisation doit être égale à la moyenne statistique de la variable aléatoire désignant la valeur prise par le processus en un lieu donné. Dit de manière imagée, et en reprenant l'exemple de la feuille de papier froissée, on dira que ce processus est ergodique si la hauteur moyenne de la feuille (calculée en moyennant la hauteur de tous les points

³Ethymologiquement, le terme ergodique a été introduit dans le domaine de la physique statistique, pour désigner la propriété d'un ensemble de particules, dont l'étude statistique de la trajectoire de l'une d'entre elles (arbitrairement choisie) est représentative des caractéristiques (vitesse, accélérations, etc.) de l'ensemble des particules

d'une feuille après la réalisation d'une expérience aléatoire), est égale à la hauteur d'un point donné arbitraire (par exemple le centre) de la feuille, moyennée sur un grand nombre de réalisations.

Remarque : pour qu'une moyenne calculée sur tout le domaine analytique \mathcal{D} converge vers une unique valeur d'espérance qui ne dépende pas du lieu donné, nécessairement, pour être ergodique à l'ordre 1, un processus doit *a minima* être stationnaire à l'ordre 1.

Exemple 2.1. On peut donner un exemple simple de processus **non-ergodique** : le signal aléatoire constant sur chacune de ses réalisations $Z(x, \omega) = A(\omega)$. Concrètement : on souhaite générer aléatoirement le relief du terrain sur une zone donnée. Pour ce faire, on propose de tirer aléatoirement un nombre à l'aide d'un dé à 6 faces, puis on affecte la valeur obtenue à toutes les cellules du terrain. Le processus est bien stationnaire à l'ordre 1, toutes les cellules ont la même espérance d'altitude : $\mathbb{E}[Z] = 3.5$. Prenons une réalisation quelconque $\omega \in \Omega = \{1, 2, 3, 4, 5, 6\}$, par exemple $\omega = 2$, et calculons une moyenne spatiale du MNT générée. Toutes les cellules ayant pris la valeur 2, cette moyenne vaut également $2 \neq \mathbb{E}[Z]$, quelque soit la taille du terrain. Ce processus n'est donc pas ergodique.

De manière complètement analogue, on définit l'**ergodicité à l'ordre 2** comme étant la propriété d'un signal aléatoire dont chaque réalisation comporte individuellement toute l'information nécessaire au calcul de la covariance :

Définition 24 (Ergodicité à l'ordre 2)

Soit Z un processus stochastique ergodique à l'ordre 1. On dit que Z est *ergodique à l'ordre 2*, ou encore *ergodique pour la covariance*, si et seulement si, pour toute réalisation z du processus (de moyenne spatiale m_z sur \mathcal{D}), et pour tout vecteur de translation h :

$$\frac{1}{|\mathcal{D}|} \int_{\mathcal{D}} (z(x+h) - m_z)(z(x) - m_z) dx \xrightarrow[|\mathcal{D}| \rightarrow \infty]{} \text{Cov}[Z(x+h), Z(x)]$$

avec $|\mathcal{D}|$ qui désigne la taille de \mathcal{D} (longueur, surface, volume...).

Remarque : étant donné que Z est supposé être ergodique à l'ordre 1, l'ergodicité à l'ordre 2 peut être caractérisée par une relation asymptotique ne requérant pas la connaissance de m_z :

$$\frac{1}{|\mathcal{D}|} \int_{\mathcal{D}} z(x+h)z(x) dx \xrightarrow[|\mathcal{D}| \rightarrow \infty]{} \mathbb{E}[Z(x+h)Z(x)]$$

En effet, en partant du membre de gauche de la définition 24, on a :

$$\begin{aligned} \frac{1}{|\mathcal{D}|} \int_{\mathcal{D}} (z(x+h) - m_z)(z(x) - m_z) dx &= \frac{1}{|\mathcal{D}|} \int_{\mathcal{D}} z(x+h)z(x) dx - \frac{2m_z}{|\mathcal{D}|} \int_{\mathcal{D}} z(x) dx + \frac{1}{|\mathcal{D}|} \int_{\mathcal{D}} m_z^2 dx \\ &= \frac{1}{|\mathcal{D}|} \int_{\mathcal{D}} z(x+h)z(x) dx - 2m_z^2 + m_z^2 = \frac{1}{|\mathcal{D}|} \int_{\mathcal{D}} z(x+h)z(x) dx - m_z^2 \xrightarrow[|\mathcal{D}| \rightarrow \infty]{} \mathbb{E}[Z(x+h)Z(x)] - m_z^2 \end{aligned}$$

Et à droite : $\text{Cov}[Z(x+h), Z(x)] = \mathbb{E}[Z(x+h)Z(x)] - \mathbb{E}[Z(x+h)]\mathbb{E}[Z(x)] = \mathbb{E}[Z(x+h)Z(x)] - m_z^2$

2.3.3 Les hypothèses en pratique

Pour pouvoir appliquer les techniques de la Géostatistique linéaire, on doit être en mesure de calculer le covariogramme du phénomène étudié, qui donc en échange doit pouvoir être modélisé de manière satisfaisante par une processus stochastique stationnaire⁴ et ergodique au second ordre. L'isotropie est une propriété souhaitable pour alléger les calculs, mais non-indispensable. Si ces hypothèses sont parfaitement vérifiables (ou réfutables) pour des processus stochastiques formels (*i.e.* des signaux définis de manière théorique par des équations) comme nous allons le voir dans l'activité II, la situation est en revanche beaucoup plus compliquée dans le cas de phénomènes réels :

- La **stationnarité à l'ordre 1** est vérifiée lorsque le signal observé ne présente pas de tendance de fond. Cependant, on peut facilement rendre le processus stationnaire, en estimant la tendance, en l'éliminant, puis en appliquant l'analyse variographique sur le signal résiduel qui en principe doit être stationnaire pour la moyenne.
- La **stationnarité à l'ordre 2** stipule que la variabilité du signal doit être à peu près la même sur tout le domaine. Si ce n'est pas le cas (et si la taille du domaine est suffisamment importante pour le permettre), il faudra nécessairement segmenter le signal de sorte à le rendre approximativement stationnaire pour la covariance sur chaque zone.
- L'**ergodicité** (au premier et second ordre) est impossible à vérifier en pratique.
- L'**isotropie** stipule que le phénomène est insensible aux directions. En général, hormis en présence de pathologies géomorphologiques marquées (failles, rift...) un MNT peut être considéré comme isotrope.
- Dans tous les cas, on peut se référer à la littérature pour s'avoir si la classe de phénomènes étudiée (gisement d'or, altitude du terrain, pression atmosphérique, etc.) est reputée posséder les propriétés souhaitées.

Activité II. Stationnarité

Dans la section 2.1, nous avons défini le concept de processus stochastique, que nous avons manipulé dans l'activité I. La section 2.3 a énoncé les 2 hypothèses fondamentales pour pouvoir calculer le covariogramme d'un processus à partir d'un semis de points observés : la **stationnarité** et l'**ergodicité** au second ordre. Dans cette activité, nous proposons 6 exemples concrets de processus stochastiques uni-dimensionnels, pour lesquels nous allons chercher à valider ou réfuter l'hypothèse de stationnarité :

- $X_t = w_t$ avec $w_t \sim \mathcal{N}(\mu, \sigma)$
- $X_{t+1} = X_t + kw_t + b$ avec $(k, b) \in \mathbb{R}^2$ et $w_t \sim \mathcal{N}(0, 1)$
- $X_{t+1} = \rho X_t + c + w_t$ avec $(\rho, c) \in \mathbb{R}^2$ $w_t \sim \mathcal{N}(0, 1)$
- $X_t = A \cos(\omega t)$, $Y_t = B \sin(\omega t)$ et $Z_t = X_t + Y_t$, avec $\omega \in \mathbb{R}$ et A, B indépendantes $\sim \mathcal{N}(0, 1)$
- $X_t = r \cos(\omega t + \varphi)$, avec $(\omega, \varphi) \in \mathbb{R}^2$ et $r \sim \mathcal{N}(0, 1)$
- $X_t = r \cos(\omega t + \psi)$, avec $(\omega, r) \in \mathbb{R}^2$ et $\psi \sim \mathcal{U}(-\pi; \pi]$

⁴Par la suite, nous verrons que l'hypothèse moins exigeante de stationnarité intrinsèque est suffisante en pratique.

On rappelle que la notation $x \sim \mathcal{N}(\mu, \sigma)$ signifie que la variable x est distribuée aléatoirement suivant une loi normale de moyenne μ et d'écart-type σ . De même, $x \sim \mathcal{U}([a; b])$ signifie que x est distribuée aléatoirement et uniformément dans l'intervalle $[a, b]$.

L'objectif de ces travaux pratiques est de valider ou invalider les hypothèses suivantes :

Stationnarité	<input type="checkbox"/> Non	<input type="checkbox"/> À l'ordre 1
----------------------	------------------------------	--------------------------------------

On pourra utiliser au choix l'une des deux stratégies suivantes pour résoudre le problème : (1) exprimer analytiquement les espérances et les variances des processus ou bien (2) procéder de manière similaire à celle employée dans le script 1 (page 32), *i.e.* simuler informatiquement un certain nombre de réalisations du processus et les représenter graphiquement.

Correction :

On commence par donner la résolution analytique du problème pour chaque processus, puis on écrit le code nécessaire à la simulation informatique.

a) Il s'agit d'un bruit blanc classique. On a $\mathbb{E}[X_t] = \mathbb{E}[w_t] = \mu$, indépendant de t donc X est stationnaire à l'ordre 1. Par ailleurs, la covariance de deux échantillons séparés d'un intervalle τ est :

$$\text{Cov}(X_t, X_{t+\tau}) = \begin{cases} \sigma^2 & \text{si } \tau = 0 \\ 0 & \text{sinon.} \end{cases} \quad (2.3)$$

Cette covariance ne dépend que de τ donc le processus X est également stationnaire à l'ordre 2.

Stationnarité	<input type="checkbox"/> Non	<input checked="" type="checkbox"/> À l'ordre 1
----------------------	------------------------------	-------------------------------------------------

b) Le processus X est une marche aléatoire réelle à temps discret, similaire à celle définie par l'équation 2.1. On a : $\mathbb{E}[X_{t+1}] = \mathbb{E}[X_t + kw_t + b] = \mathbb{E}[X_t] + k\mathbb{E}[w_t] + b = \mathbb{E}[X_t] + b$. L'espérance du signal est donc constante au cours du temps si, et seulement si, $b = 0$. Dans ce cas, X est stationnaire à l'ordre 1. Sous l'hypothèse où $b = 0$, étudions à présent la variance du processus⁵ : $\text{Var}[X_{t+1}] = \text{Var}[X_t + kw_t] = \text{Var}[X_t] + k^2\text{Var}[w_t]$ (puisque les 2 termes sont indépendants) $= \text{Var}[X_t] + k^2$. Ici aussi, le processus ne peut être stationnaire à l'ordre 1 que si $k = 0$. Dans ce cas, l'équation du processus est réduite à : $X_{t+1} = X_t$ autrement dit, X est une suite constante, complètement déterminée par son premier terme. Elle est donc stationnaire au sens strict. Dans le cas général cependant, $k \neq 0$, $b \neq 0$ et le résultat est :

Stationnarité	<input checked="" type="checkbox"/> Non	<input type="checkbox"/> À l'ordre 1
----------------------	-----------------------------------------	--------------------------------------

Etudions quand même la covariance du processus pour le cas $k \neq 0$: $X_{t+1} = X_t + kw_t$. On a alors, pour un décalage $\tau \geq 0$: $X_{t+\tau} = X_t + kw_t + kw_{t+1} + \dots + kw_{t+\tau-1}$. D'où :

⁵Pour déterminer si le processus est stationnaire à l'ordre 2, c'est bien la covariance qu'il faut étudier, mais si on trouve que la variance est déjà non-stationnaire, on aura montré à moindre frais que X n'est pas stationnaire à l'ordre 2.

$$\begin{aligned}\text{Cov}(X_t, X_{t+\tau}) &= \text{Cov}\left(X_t, X_t + k \sum_{n=0}^{\tau-1} w_{t+n}\right) = \text{Cov}(X_t, X_t) + k \sum_{n=0}^{\tau-1} \text{Cov}(X_t, w_{t+n}) \\ &= \text{Var}[X_t] = \text{Var}[X_0] + tk^2\end{aligned}$$

Donc, en supposant déterministe le premier échantillon X_0 (comme c'est le cas dans 2.1) :

$$\text{Cov}(X_{t_1}, X_{t_2}) = k^2 \min(t_1, t_2) \quad (2.4)$$

La ressemblance entre deux échantillons X_{t_1} et X_{t_2} ne dépend donc pas de l'écart de temps $\tau = |t_1 - t_2|$ les séparant, mais uniquement de la date du premier des deux échantillons.

c) X est un processus autorégressif d'ordre 1 : AR(1). Il généralise les deux processus précédents : avec $\rho = 1$ on obtient une marche aléatoire, et avec $\rho = 0$ on a un bruit blanc gaussien.

$$\mathbb{E}[X_{t+1}] = \mathbb{E}[\rho X_t + c + w_t] = \rho \mathbb{E}[X_t] + c \quad (2.5)$$

$$\text{Var}[X_{t+1}] = \text{Var}[\rho X_t + c + w_t] = \rho^2 \text{Var}[X_t] + 1 \quad (2.6)$$

Le processus ne peut être stationnaire à l'ordre 2 que si $\mathbb{E}[X_{t+1}] = \mathbb{E}[X_t]$ et $\text{Var}[X_{t+1}] = \text{Var}[X_t]$. Les équations de récurrence 2.5 et 2.6 imposent donc les 2 conditions suivantes sur les paramètres (en supposant $\rho \neq 1$ dont le cas a déjà été étudié précédemment) :

$$\mathbb{E}[X_t] = \frac{c}{1 - \rho} \quad \text{Var}[X_t] = \frac{1}{1 - \rho^2} \quad (2.7)$$

La positivité de la quantité $\text{Var}[X_t]$ impose donc nécessairement la condition $\rho < 1$. Calculons à présent la covariance entre deux échantillons successifs X_t et X_{t+1} :

$$\begin{aligned}\text{Cov}(X_{t+1}, X_t) &= \text{Cov}(\rho X_t + c + w_t, X_t) = \rho \text{Cov}(X_t, X_t) = \rho \text{Var}[X_t] = \frac{\rho}{1 - \rho^2} \\ \text{Cov}(X_{t+2}, X_t) &= \text{Cov}(\rho X_{t+1} + c + w_{t+1}, X_t) = \rho \text{Cov}(X_{t+1}, X_t) = \frac{\rho^2}{1 - \rho^2} \\ \dots &= \dots \\ \text{Cov}(X_{t+\tau}, X_t) &= \text{Cov}(\rho X_{t+\tau-1} + c + w_{t+\tau-1}, X_t) = \rho \text{Cov}(X_{t+\tau}, X_t) = \frac{\rho^\tau}{1 - \rho^2}\end{aligned}$$

La covariance ne dépend que de l'écart τ entre les échantillons, et non de la date absolue t . Donc, le processus X est stationnaire au second ordre, sous les conditions :

$$\rho < 1 \text{ et } X_0 \text{ distribué suivant une loi de moyenne } \mu = \frac{c}{1 - \rho} \text{ et d'écart-type } \sigma = \sqrt{\frac{1}{1 - \rho^2}}$$

Stationnarité

Non

À l'ordre 1

À l'ordre 2

d) X et Y sont deux signaux sinusoïdaux d'amplitudes aléatoires, à phases fixées et mutuellement en opposition de phase. On a : $\mathbb{E}[X_t] = \mathbb{E}[A \cos(\omega t)] = \mathbb{E}[A] \cos(\omega t) = 0$, puisque A est d'espérance nulle. Le processus est donc stationnaire pour la moyenne. En revanche $\text{Var}[X_t] = \text{Var}[A] \cos^2(\omega t) = \cos^2(\omega t)$, dépendant de t et donc X n'est pas stationnaire à l'ordre 2. Par symétrie, on obtient la même conclusion pour le processus Y . Intéressons-nous à présent à la somme des deux processus :

$$\begin{aligned}\mathbb{E}[Z_t] &= \mathbb{E}[X_t + Y_t] = \mathbb{E}[A] \cos(\omega t) + \mathbb{E}[B] \sin(\omega t) = 0 \\ \text{Var}[Z_t] &= \text{Var}[X_t + Y_t] = \text{Var}[A] \cos^2(\omega t) + \text{Var}[B] \sin^2(\omega t) = \cos^2(\omega t) + \sin^2(\omega t) = 1\end{aligned}$$

Plus généralement, pour la covariance, en considérant s et t deux instants quelconques :

$$\begin{aligned}\text{Cov}(Z_s, Z_t) &= \text{Cov}(X_s + Y_s, X_t + Y_t) = \text{Cov}(X_s, X_t) + \text{Cov}(X_s, Y_t) + \text{Cov}(Y_s, X_t) + \text{Cov}(Y_s, Y_t) \\ &= \mathbb{E}[A^2 \cos(\omega s) \cos(\omega t)] + \mathbb{E}[A]\mathbb{E}[B](\cos(\omega s) \sin(\omega t) + \sin(\omega s) \cos(\omega t)) + \mathbb{E}[B^2 \sin(\omega s) \sin(\omega t)] \\ &= \mathbb{E}[A^2] \cos(\omega s) \cos(\omega t) + \mathbb{E}[B^2] \sin(\omega s) \sin(\omega t) \\ &= \cos(\omega s) \cos(\omega t) + \sin(\omega s) \sin(\omega t) \\ &= \cos(\omega(s - t)) = \cos(\omega\tau)\end{aligned}$$

La covariance entre deux échantillons Z_t et Z_s ne dépend donc que de l'écart $\tau = |s - t|$. Le processus Z est donc stationnaire à l'ordre 2 (malgré la non-stationnarité de ses 2 composantes additives).

Stationnarité de Z Non À l'ordre 1 À l'ordre 2

e) Il s'agit d'un signal sinusoïdal à phase et fréquence fixées et d'amplitude aléatoire.

$$\begin{aligned}\mathbb{E}[X_t] &= \mathbb{E}[r \cos(\omega t + \varphi)] = \mathbb{E}[r] \cos(\omega t + \varphi) = 0 \\ \text{Var}[X_t] &= \text{Var}[r] \cos^2(\omega t + \varphi) = \cos^2(\omega t + \varphi)\end{aligned}$$

La variance est fonction du temps t donc le processus X n'est pas stationnaire à l'ordre 2.

Stationnarité Non À l'ordre 1 À l'ordre 2

f) Il s'agit d'un signal sinusoïdal à fréquence et amplitude fixées et de déphasage aléatoire.

$$\begin{aligned}\mathbb{E}[X_t] &= r\mathbb{E}[\cos(\omega t + \psi)] = \frac{r}{2\pi} \int_{-\pi}^{\pi} \cos(\omega t + \psi) d\psi = \frac{r}{2\pi} [\sin(\omega t + \psi)]_{-\pi}^{\pi} = 0 \\ \text{Cov}(X_s, X_t) &= r^2 \mathbb{E}[\cos(\omega s + \psi) \cos(\omega t + \psi)] = \frac{r^2}{2\pi} \int_{-\pi}^{\pi} \cos(\omega s + \psi) \cos(\omega t + \psi) d\psi\end{aligned}$$

En utilisant l'identité trigonométrique $\cos p \cos q = \frac{1}{2}(\cos(p+q) + \cos(p-q))$, on obtient :

$$\text{Cov}(X_s, X_t) = \frac{r^2}{2\pi} \int_{-\pi}^{\pi} \cos(\omega(s+t) + 2\psi) d\psi + \frac{r^2}{2\pi} \int_{-\pi}^{\pi} \cos(\omega(s-t)) d\psi = r^2 \cos(\omega(s-t)) = r^2 \cos(\omega\tau)$$

Le processus X est donc stationnaire à l'ordre 2 :

Stationnarité

Non

À l'ordre 1

À l'ordre 2

On donne ci-dessous le code R pour programmer ces 6 processus et contrôler de manière expérimentale l'hypothèse de stationnarité. On utilise les paramètres par défaut suivants :

- a) $\mu = 0$ et $\sigma = 1$
- b) $k = 1$ et $b = 0$
- c) $\rho = 0.99$ (non-stationnarité pour $\rho \geq 1$) et $c = 1$ (C dans le code)
- d) $\omega = 1$ (noté w dans le code)
- e) $\omega = 1$ (noté w dans le code) et $\varphi = 0$
- f) $\omega = 1$ (noté w dans le code) et $r = 1$

On pourra générer n variables aléatoires gaussiennes μ, σ ou uniformes $\mathcal{U}([‐; +])$, respectivement avec les commandes R : `rnorm(n, mu, sigma)` et `runif(n, a, b)`.

Le script 3 donne la fonction `generate_process` permettant de simuler aléatoirement l'un des 6 processus dans un vecteur de longueur N. On peut alors représenter 500 échantillons, par exemple du processus (c) par :

```
plot(generate_process(3, 500), type="l")
```

Pour représenter R réalisations de chacun des 6 processus (simultanément dans 6 fenêtres graphiques distinctes), on pourra utiliser le script 4. Si besoin, pour faciliter la visualisation, on peut représenter chaque réalisation d'une couleur différente en ajoutant l'argument suivant à la fonction `lines`.

```
col=rgb(runif(1,0,1),runif(1,0,1),runif(1,0,1),1)
```

On peut alors tester de faire varier les paramètres des processus pour trouver les conditions sous lesquelles les hypothèses de stationnarité (au premier et second ordre) sont vérifiées.

Script 3. Stationnarité [code2.r]

```
# Parametres
C = 1; k = 1; b = 0; w = 1; r = 1; rho = 0.99; phi = 0; mu = 0; sigma = 1

par(mfrow=c(2,3))      # Pour tracer les 6 graphes dans une matrice [2x3]

# -----
# Fonction de generation des processus
# Entrees : indice i (1 a 6) du processus, longueur N
# Sortie : vecteur X de N echantillons du processus i
# -----
generate_process = function(id, N){

  T = 10*(0:(N-1))/(N-1)    # Initialisation du vecteur de pas de temps

  # Simulation
  if (id == 1){  # Bruit blanc gaussien
    X = rnorm(N,mu,sigma)
  }

  if (id == 2){  # Marche aleatoire discrete
    X = cumsum(k*rnorm(N,0,1) + rep(b,N))
  }

  if (id == 3){  # Processus auto-regressif AR(1)
    X = rep(0, N)
    X[1] = rnorm(1, C/(1-rho), 1/sqrt(1-rho*rho))
    for (i in 2:N){
      X[i] = rho*X[i-1] + C + rnorm(1,0,1)
    }
  }

  if (id == 4){  # Somme de signaux sinusoidaux d'amplitudes aleatoires
    S1 = rnorm(1,0,1)*cos(w*T)
    S2 = rnorm(1,0,1)*sin(w*T)
    X = S1+S2
  }

  if (id == 5){  # Signal sinusoidal d'amplitude aleatoire
    X = rnorm(1,0,1)*cos(w*T + phi)
  }

  if (id == 6){  # Signal sinusoidal a dephasage aleatoire
    X = r*cos(w*T + runif(1,-pi, pi))
  }

  return(X)
}
```

Script 4. Représentation graphique de la stationnarité [code2.r]

```

N = 500      # Longueur du processus
R = 10       # Nombre de réalisations

for (id in 1:6){
    X = generate_process(id, N)
    marge = c(-1,1)*(max(X)-min(X))
    plot(X, type="l", xlab="t", ylab="X", ylim=c(min(X), max(X))+marge)

    for (réalisation in 2:R){
        lines(generate_process(id,N))
    }
}

```

2.4 Le variogramme

Nous avons introduit la notion de **covariogramme** à la fin de la section 2.3.1. Pour un processus **stationnaire au second ordre** et **isotrope**, le covariogramme est une fonction C qui, à tout réel positif h associe la covariance $C(h)$ entre deux observations lorsque la seule information dont on dispose a priori est la distance h séparant les deux sites sur lesquelles ont été effectuées ces observations.

Il existe cependant des cas pratiques où la variance d'une loi n'existe pas. Il suffit par exemple de considérer, dans le cas uni-dimensionnel, la loi de densité π définie sur $[1; +\infty[$ par $\pi(x) = 2/x^3$, dont la moyenne vaut 1, mais dont la variance est infinie, comme illustré sur la figure 2.10. On pourra trouver deux autres exemples de lois de variance infinie dans [Hauchecorne \(2007\)](#).

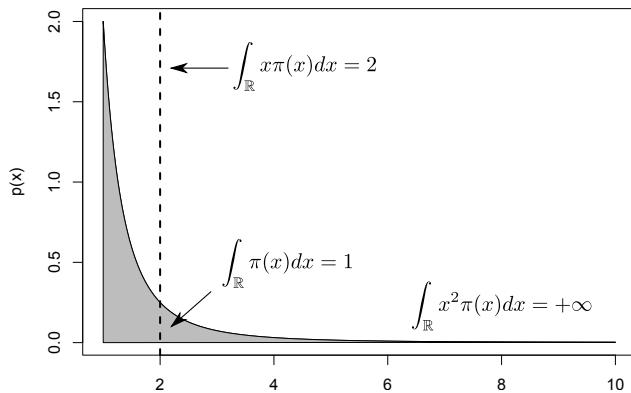


Fig. 2.10. Illustration d'une loi de probabilité de variance infinie : $\pi : x \mapsto 2/x^3$. Cette pathologie est caractéristique des lois dont la densité ne décroît pas suffisamment rapidement vers 0 à mesure que x augmente. Ces lois sont dites *à longue traîne*.

Notons que cette limitation, qui pourrait sembler théorique au premier abord, n'est absolument pas anecdote. [Arnaud et Emery \(2000\)](#) par exemple, rapportent le cas de densité de minerais dont la variance croît indéfiniment à mesure que l'extension spatiale du gisement considéré augmente.

On préfère donc utiliser un outil légèrement plus général : le **variogramme**, qui opère directement sur les accroissements : c'est l'hypothèse **intrinsèque**.

Définition 25 (Hypothèse intrinsèque et Variogramme)

Un processus stochastique Z est dit *intrinsèquement stationnaire* (ou *intrinsèque*) s'il est stationnaire pour la moyenne et si la variance de la différence entre deux sites x et $x + h$ du domaine \mathcal{D} distants de h , existe et ne dépend que de h :

$$\gamma(h) = \frac{1}{2}\text{Var}[Z(x+h) - Z(x)]$$

La fonction $\gamma : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ est appelée *variogramme* (ou parfois *semi-variogramme*) de Z .

Remarque : si Z n'est pas isotrope, alors h est un vecteur de \mathcal{D} , et le variogramme γ devient une fonction de \mathcal{D} dans \mathbb{R}^+ . Par la suite, sauf mention contraire, l'hypothèse d'isotropie sera implicitement supposée. Par ailleurs, nous motiverons plus loin l'introduction du facteur $1/2$.

À l'aide de l'identité 1.13, on obtient une expression plus interprétable de γ :

$$\gamma(h) = \frac{1}{2}\text{Var}[Z(x+h) - Z(x)] = \frac{1}{2}\left(\mathbb{E}[(Z(x+h) - Z(x))^2] - \mathbb{E}[Z(x+h) - Z(x)]^2\right)$$

Or, Z est stationnaire pour la moyenne donc $\mathbb{E}[Z(x+h) - Z(x)] = \mathbb{E}[Z(x+h)] - \mathbb{E}[Z(x)] = 0$ d'où :

$$\gamma(h) = \frac{1}{2}\mathbb{E}[(Z(x+h) - Z(x))^2] \quad (2.8)$$

Le variogramme indique donc, pour toute distance h , la moitié de l'espérance des écarts entre les valeurs prises par le processus en deux sites séparés de h .

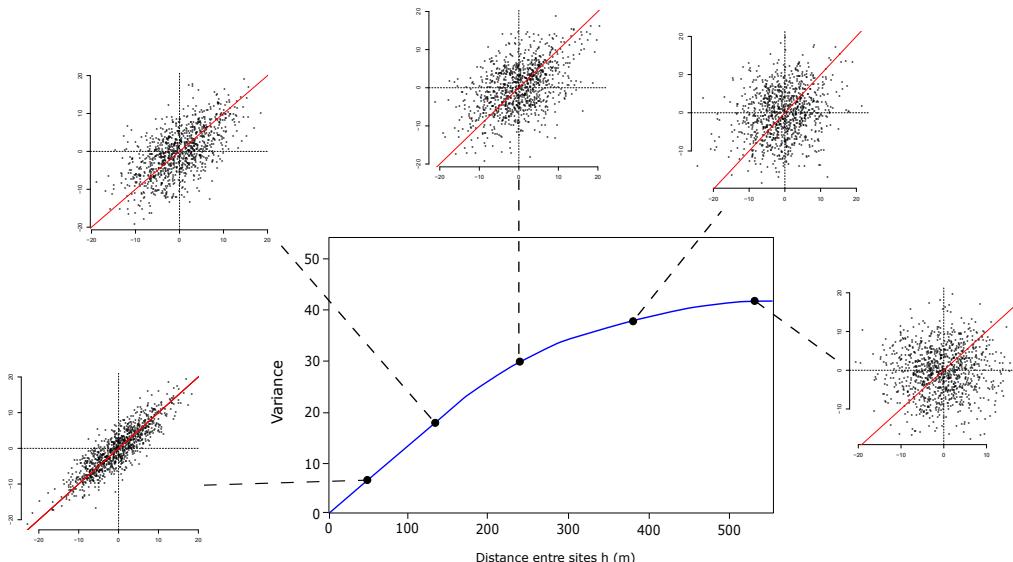


Fig. 2.11. Chaque valeur $\gamma(h)$ du variogramme traduit la dissemblance entre deux sites séparés d'une distance h . Plus $\gamma(h)$ est faible plus les valeurs prises sur les deux sites sont statistiquement *similaires*.

Là où le **covariogramme** $C(h)$ mesure la **ressemblance** entre sites, le **variogramme** $\gamma(h)$ est quant à lui un indicateur de la **dissemblance**. On s'attend donc à avoir une relation antagoniste entre les fonctions C et γ . C'est la relation fondamentale de la Géostatistique :

Théorème 26 (Relation fondamentale variogramme/covariogramme)

Pour un champ stationnaire au sens large, pour toute distance h , variogramme et covariogramme sont liés par la relation :

$$\gamma(h) = C(0) - C(h)$$

Preuve : en considérant x et $x + h$ deux sites quelconques séparés d'une distance h , on a : $2\gamma(h) = \text{Var}[Z(x) - Z(x + h)] = \text{Var}[Z(x)] + \text{Var}[Z(x + h)] - 2\text{Cov}(Z(x), Z(x + h)) = 2\text{Var}[Z] - 2C(h)$. En faisant remarquer que $\text{Var}[Z] = \text{Cov}(Z(x), Z(x)) = C(0)$ et en divisant les deux membres de l'équation par 2, on obtient la relation fondamentale de la Géostatistique.

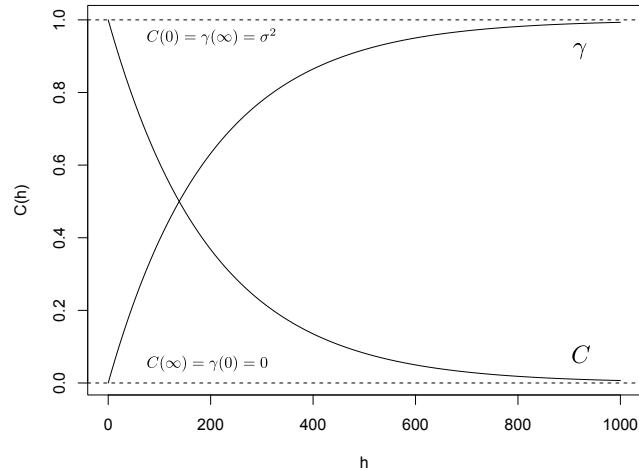


Fig. 2.12. Lorsque le covariogramme C existe, il est lié au variogramme γ par la relation $\gamma = C(0) - C(h)$. À l'infini, le variogramme atteint donc le pallier $C(0) = \sigma^2$.

Une manière simple de comprendre le variogramme est de considérer $\gamma(h)$ comme l'incertitude entourant la valeur prise par le phénomène en un lieu donné, étant donnée l'observation d'une valeur prise en un site distant de h . Cette incertitude est donc intuitivement maximale quand les deux sites sont séparés d'une distance qui tend à devenir infinie⁶. À l'inverse, si les deux sites sont voisins (typiquement quelques mètres de distance dans le cas du MNT), alors l'observation de $Z(x_1)$ détermine pratiquement la valeur prise par Z en x_2 .

Lorsqu'ils existent tous deux, variogramme et covariogramme contiennent la même information. S'ils sont donnés par un oracle (*i.e.* s'ils sont parfaitement connus et qu'il n'est pas nécessaire de les estimer à partir d'un semis d'observation), alors on obtient théoriquement les mêmes résultats indépendamment du choix effectué entre γ et C . En revanche, lorsqu'on doit les estimer, on peut démontrer que

⁶Dans ce cas, remarquer que $\gamma(h)$ vaut la variance a priori revient à considérer que les deux sites x_1 et x_2 sont séparés d'une distance h si importante, que les valeurs qui y sont prises par le phénomène ne sont plus corrélées. L'observation de $Z(x_1)$ n'apporte alors aucune information sur les valeurs que l'on peut espérer observer en x_2 et vis-versa.

l'estimateur de C est plus robuste mais biaisé. On préfère donc en pratique utiliser le variogramme γ , d'autant qu'il est plus générique (γ peut exister sans C alors que l'inverse n'est pas vrai).

Remarque : l'hypothèse de stationnarité **intrinsèque**, qui suppose uniquement la stationnarité au 1^{er} ordre ainsi que l'existence de la variance des accroissements (def. 25) est donc une version affaiblie de la stationnarité au 2nd ordre, qui est malgré tout suffisante pour travailler en Géostatistique.

Exemple 2.2. On propose ici un exemple de cas où seul le variogramme existe. On pose $Z_1 = A + C$ et $Z_2 = B + C$, avec $A, B \sim \mathcal{N}(0, 1)$ deux variables aléatoires gaussiennes indépendantes et C distribuée suivant une loi π n'admettant pas de variance (par exemple π est la loi à densité de la fig. 2.10). On a alors, par bilinéarité de l'opérateur de covariance : $\text{Cov}(Z_1, Z_2) = \text{Cov}(A, B) + \text{Cov}(A, C) + \text{Cov}(C, B) + \text{Cov}(C, C) = \text{Cov}(C, C)$ (puisque A , B et C sont indépendantes) = $\text{Var}[C]$ qui n'existe pas ! En revanche, la variance de l'accroissement est bien définie : $\text{Var}[Z_1 - Z_2] = \text{Var}[A - B] = \text{Var}[A] + \text{Var}[B] = 2$. Nous laissons le soin au lecteur de vérifier que si π est à variance finie, alors $\text{Cov}(Z_1, Z_2)$ existe et la relation énoncée par le théorème 26 est vérifiée. On pourra trouver un exemple moins artificiel de processus strictement intrinsèque dans l'annexe B.6.

2.4.1 Variogramme expérimental

Sous l'hypothèse d'ergodicité au second ordre, on peut estimer γ à partir d'un ensemble d'observations ponctuelles, prélevées sur une unique réalisation du processus. Son estimateur (que l'on note $\hat{\gamma}$, prononcé *gamma chapeau*) est appelé **variogramme expérimental**.

Pour construire le variogramme expérimental, on doit donc estimer les valeurs de $\gamma(h)$ pour toutes les distances h possibles, ce qui est bien évidemment impossible en pratique. On discrétise alors le problème en une séquence de nombres réels h_1, h_2, \dots régulièrement échantillonnés (*i.e.* $\forall k \in \mathbb{N}^*$ $h_k = h_0 + k\Delta h$, avec Δh et h_0 des paramètres à calibrer en fonction du nombre de données disponibles).

Par définition $\hat{\gamma}(200)$ par exemple, représente la dissemblance entre deux sites du terrain séparés de 200 m. Disposant d'un nombre limité d'observations, il est en réalité peu probable que nous soyons capable de trouver ne serait-ce qu'un couple de sites distants d'exactement 200 m. On doit donc s'autoriser une tolérance, par exemple on recherchera tous les couples de sites (x_i, x_j) séparés d'une distance comprise entre 150 et 250 m. Les couples sélectionnés sont alors utilisés pour estimer la quantité 2.8, ce qui nous donne un estimateur $\hat{\gamma}(h)$ pour $h = 200$ (en toute rigueur pour $h = 200 \pm 50$).

Formellement : disposant d'un semis de points d'observations $z(x_1), z(x_2), \dots, z(x_n)$ d'un phénomène Z , on peut estimer le variogramme de Z par :

$$\hat{\gamma}(h) = \frac{1}{2|N(h)|} \sum_{(i,j) \in N(h)} (z(x_i) - z(x_j))^2 \quad (2.9)$$

où $N(h)$ désigne l'ensemble des couples d'indices (i, j) tels que les sites x_i et x_j sont séparés d'une distance $h \pm \frac{dh}{2}$: $N(h) = \{(i, j), i < j \text{ t.q. } h - \frac{dh}{2} \leq \|x_i - x_j\| \leq h + \frac{dh}{2}\}$, et où $|N(h)|$ désigne le nombre de couples contenus dans l'ensemble $N(h)$.

Informellement : de manière imagée dans le cas du problème modèle du MNT, pour une distance horizontale h , le variogramme expérimental $\hat{\gamma}(h)$ vaut la moitié de la moyenne des carrés des dénivélés entre les points séparés (approximativement) d'une distance h .

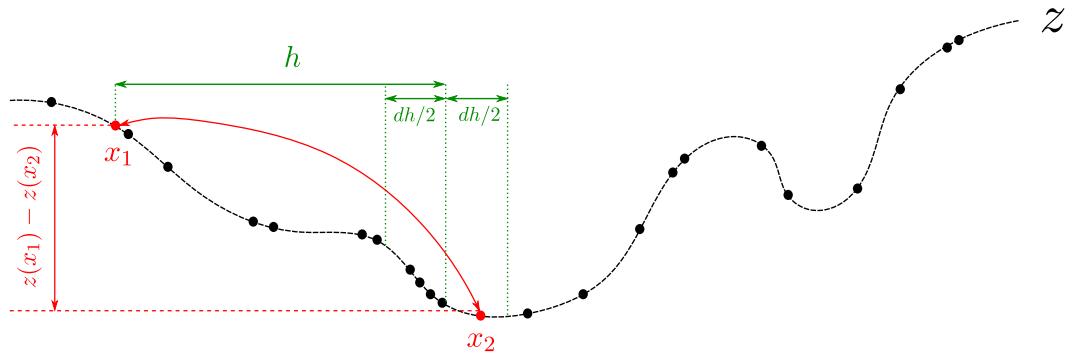
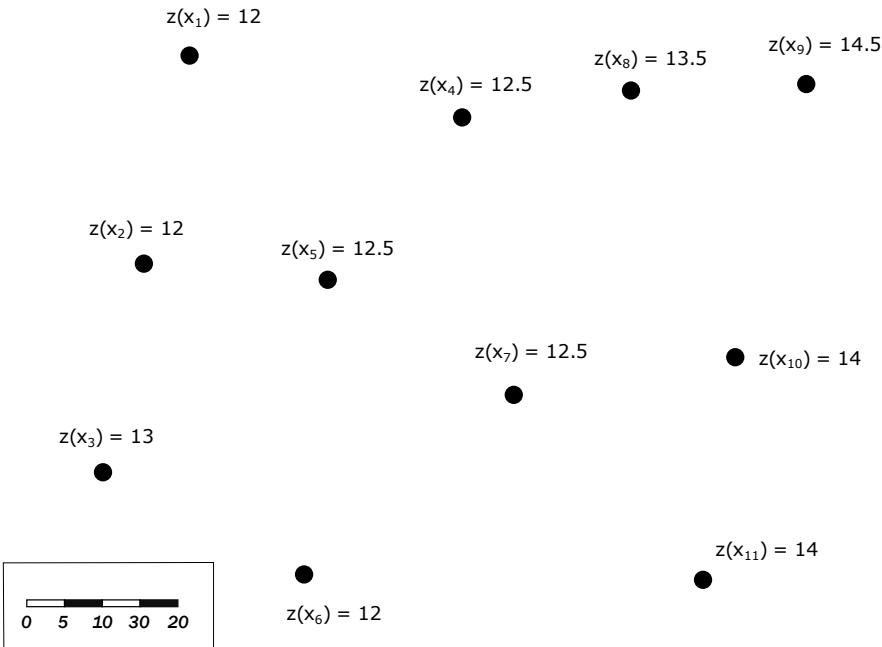


Fig. 2.13. Calcul du variogramme expérimental pour un ensemble d'observations ponctuelles collectés sur une réalisation z du processus. Le couple de sites (x_1, x_2) appartient à l'ensemble $N(h)$ des couples de sites approximativement séparés d'une distance h .

Activité III. Variogramme expérimental

Cette activité ne nécessite pas obligatoirement de ressources informatiques, mais seulement un décimètre, une calculatrice... et de la patience. On considère en semis d'observations (d'un phénomène Z quelconque), disposées comme représenté sur la figure ci-dessous, dont on donne l'échelle graphique (dans un système d'unités terrain arbitraire) en bas à gauche.



Sur une version papier de ce cours, l'échelle graphique doit être de 1 unité / mm, si bien que la distance entre x_1 et x_2 par exemple, doit valoir 28 cm = 28 unités (sous réserve que l'impression ait été réalisée en format A4). Sur une version numérique, on pourra ajuster l'échelle du lecteur PDF de sorte à obtenir une conversion aisée entre les tailles mesurées en mm sur l'écran, et le système d'unités terrain. Dans une optique plus professionnelle, on pourra sauvegarder l'image ci-dessous (capture d'écran ou autres) dans une taille arbitraire, puis mesurer les distances à l'aide d'un logiciel graphique (Paint, Inkscape...) et enfin normaliser les mesures obtenues à l'aide d'un produit en croix pour obtenir des distances dans le système d'unités terrain.

Calculer le variogramme expérimental du semis de points. On utilisera un pas de discréétisation $\Delta h = 10$ unités, en démarrant de $h_0 = 5$: $h_k = 5 + k\Delta h$.

Correction. On commence par mesurer les distances (exprimées dans le système d'unités terrain) pour tous les couples de points, et on les regroupe dans une matrice de taille $n \times n$. Cette matrice étant symétrique, on ne présente que la partie triangulaire supérieure.

-	1	2	3	4	5	6	7	8	9	10	11
1	0	28	56	37	35	71	62	59	82	83	97
2	-	0	28	47	24	46	53	69	91	79	86
3	-	-	0	67	39	30	56	86	99	86	81
4	-	-	-	0	28	64	38	22	46	48	69
5	-	-	-	-	0	39	28	48	68	55	64
6	-	-	-	-	-	0	37	78	94	64	53
7	-	-	-	-	-	-	0	43	57	30	35
8	-	-	-	-	-	-	-	0	23	39	66
9	-	-	-	-	-	-	-	-	0	38	68
10	-	-	-	-	-	-	-	-	-	0	30
11	-	-	-	-	-	-	-	-	-	-	0

Il ne reste alors plus qu'à appliquer la formule 2.9 pour tous les pas h : 5, 15, 25, ... 95. Détailons un exemple avec le cas $h = 65$. Dans un premier temps, on récupère l'ensemble $N(65)$ de tous les couples (i, j) avec $i < j$, tels que la distance entre les deux sites x_i et x_j est comprise entre 60 et 70 mètres :

$$\begin{aligned} N(65) &= \{(i, j) \mid i < j \mid 60 \leq \|x_i - x_j\| \leq 70\} \\ &= \{(1, 7)(1, 8)(3, 4)(4, 6)(4, 11)(5, 9)(5, 11)(6, 10)(8, 11)(9, 11)\} \end{aligned}$$

On dénombre $|N(65)| = 10$ couples. Le calcul de $\hat{\gamma}(65)$ par 2.9 est alors immédiat :

$$\begin{aligned} \hat{\gamma}(65) &= \frac{1}{2 \times 10} \left[(z_1 - z_7)^2 + (z_1 - z_8)^2 + (z_3 - z_4)^2 + (z_4 - z_6)^2 + (z_4 - z_{11})^2 + (z_5 - z_9)^2 + (z_5 - z_{11})^2 + (z_6 - z_{10})^2 \right. \\ &\quad \left. + (z_8 - z_{11})^2 + (z_9 - z_{11})^2 \right] = 0.8 \end{aligned}$$

On procède de manière similaire pour les autres valeurs de h , ce qui donne :

h	5	15	25	35	45	55	65	75	85	95
$N(h)$	11	0	10	12	6	7	10	3	6	4
$\hat{\gamma}(h)$	0.0	-	0.33	0.30	0.71	1.0	0.8	1.04	1.38	2.34

Table 2.1: Nombre de couples $N(h)$ et variogramme expérimental $\hat{\gamma}(h)$ pour chaque valeur de distance h . Remarquons que $\gamma(15)$ n'a pas pu être estimé faute de données.

La figure 2.14 donne une représentation graphique du variogramme expérimental. La courbe en pointillés présente une régression polynomiale $f(h) = ah^2$ (avec $a = 2.5 \cdot 10^{-4}$) des points calculés. De cette manière, il sera possible d'estimer le variogramme pour toute distance h sur la plage $[0; 100]$. Nous reparlerons plus loin des fonctions de régression.

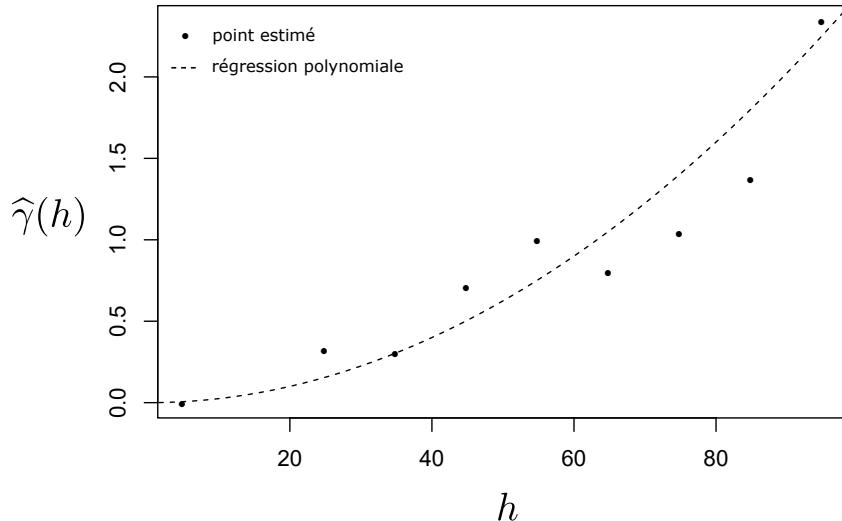


Fig. 2.14. Suite des points estimés à partir du semis d’observation et régression polynomiale.

À travers l'exemple de l'activité III, on mesure l'importance de calibrer soigneusement le pas de discréétisation Δh . Comme illustré sur la figure 2.15, lorsque le pas est trop petit, le variogramme est bien échantillonné, mais chaque point est calculé avec un faible nombre d'observations (à l'extrême, certains points peuvent même être impossibles à estimer, à l'instar du point $h = 15$ dans l'activité précédente). À l'inverse, si Δh est trop grand, chaque point est statistiquement très précis, mais on dispose de trop peu de points pour décrire fidèlement le comportement du variogramme théorique.

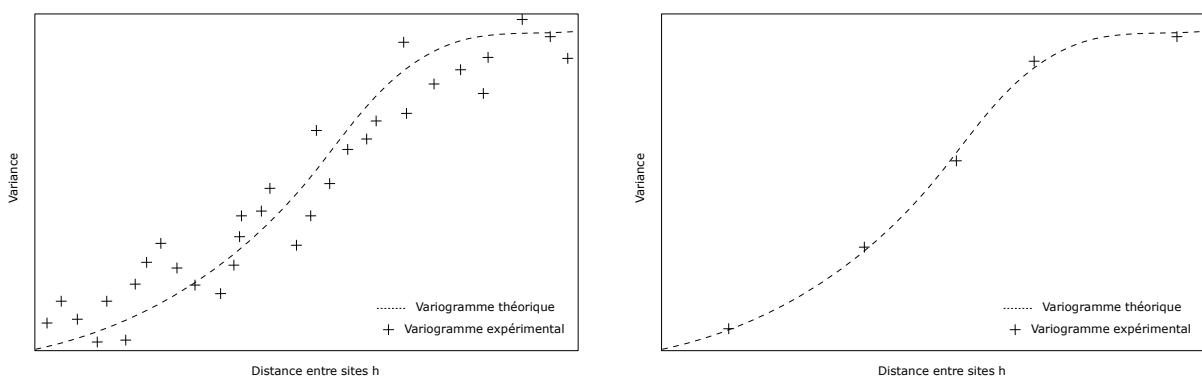


Fig. 2.15. Variogramme théorique inconnu (courbes pointillées) et estimations ponctuelles du variogramme empirique. À gauche : pas de discréétisation Δh trop fin, le variogramme estimé est résolu mais peu précis. À droite : Δh trop grand, le variogramme estimé est précis, mais peu résolu.

Le choix du paramètre Δh doit donc obéir à un compromis entre **robustesse** et **résolution**. Plus le nombre n d'observations est grand, plus on a de couples de distances, et plus on pourra se permettre

de réduire la taille de Δh . Une règle simple consiste à choisir Δh le plus fin possible tout en assurant la présence d'au moins 30 couples pour chaque point estimé sur le variogramme expérimental. En général, on procède par essai-erreur, en partant d'une résolution assez grossière, et en l'affinant jusqu'à ce que le variogramme obtenu commence à présenter des signes d'instabilité. En parallèle, on contrôlera les effectifs de couples dans chaque classe de distance.

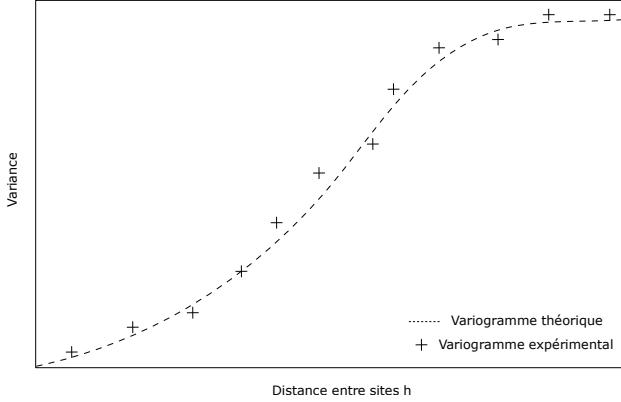


Fig. 2.16. Compromis idéal sur le pas de discréttisation Δh pour le calcul du variogramme expérimental. L'estimateur $\hat{\gamma}$ ainsi obtenu est relativement précis, et suffisamment résolu pour espérer pouvoir reconstruire le variogramme théorique en pointillé.

Disposant d'un estimateur $\hat{\gamma}$ du variogramme théorique inconnu γ , on doit pouvoir s'assurer de la qualité de celui-ci. Nous l'avons vu dans l'équation 1.14, la justesse d'un estimateur se mesure sur deux composantes : son **biais** (*i.e.* son erreur systématique quelque soit le jeu de données considéré) et sa **variance** (*i.e.* sa sensibilité aux petites variations aléatoires dans les données).

Théorème 27 (Biais du variogramme expérimental)

Le variogramme expérimental est un estimateur **non-biaisé** du variogramme théorique :

$$\forall h \in \mathbb{R}^+ \quad \mathbb{E}[\hat{\gamma}(h)] = \gamma(h)$$

Autrement dit, quelque soit la distance h considérée, la valeur prise par le variogramme expérimental $\hat{\gamma}$ en h n'a pas de raison d'être systématiquement en dessous ou au dessus de la valeur réelle $\gamma(h)$, ce qui est une propriété souhaitable pour tout estimateur digne de ce nom.

Preuve : nous donnons ici une ébauche de démonstration, faisant abstraction du fait que le calcul de $\hat{\gamma}(h)$ fait appel à des couples de sites séparés d'une distance qui ne vaut pas exactement h . En partant de la définition 2.9 du variogramme expérimental, et en remarquant que $N(h)$ est déterministe :

$$\begin{aligned} \mathbb{E}[\hat{\gamma}(h)] &= \frac{1}{2|N(h)|} \sum_{(i,j) \in N(h)} \mathbb{E}[(Z_i - Z_j)^2] = \frac{1}{2|N(h)|} \sum_{(i,j) \in N(h)} (\mathbb{E}[Z_i - Z_j]^2 + \text{Var}[Z_i - Z_j]) \\ &= \frac{1}{2|N(h)|} \sum_{(i,j) \in N(h)} \text{Var}[Z_i - Z_j] = \frac{1}{2|N(h)|} \times |N(h)|\gamma(h) = \gamma(h) \end{aligned}$$

où le passage de la première à la deuxième ligne résulte du fait que le processus est stationnaire, et donc que l'espérance de ses accroissement est nulle : $\mathbb{E}[Z_i - Z_j] = \mathbb{E}[Z_i] - \mathbb{E}[Z_j] = \mathbb{E}[Z] - \mathbb{E}[Z] = 0$.

Théorème 28 (Variance du variogramme expérimental)

La dispersion du variogramme expérimental autour du variogramme théorique, est inversement proportionnelle au nombre $N(h)$ de couples de sites utilisés dans l'estimation :

$$\text{Var}[\hat{\gamma}(h)] \propto \frac{\gamma(h)^2}{N(h)}$$

Nous ne donnerons pas de démonstration du théorème 28, mais on peut remarquer qu'il s'agit d'une application plus ou moins directe du théorème 12, dit de *propagation des variances*.

Le variogramme expérimental est donc d'autant moins précis que h augmente. En effet, on peut montrer que, dans le cas où les sites sont répartis uniformément, le nombre de couples de sites séparés d'une distance (plus ou moins la marge de tolérance) décroît à mesure que la distance h s'approche du diamètre de la zone considérée (*i.e.* de la plus grande distance séparant deux sites sur la zone). En conséquence $N(h)$ diminue, et l'incertitude sur l'estimation $\hat{\gamma}(h)$ augmente naturellement, comme illustré sur la figure 2.17.

En pratique, on considérera que le variogramme expérimental n'est plus un estimateur robuste pour des distances de séparation h supérieures à la moitié du diamètre du domaine \mathcal{D} .

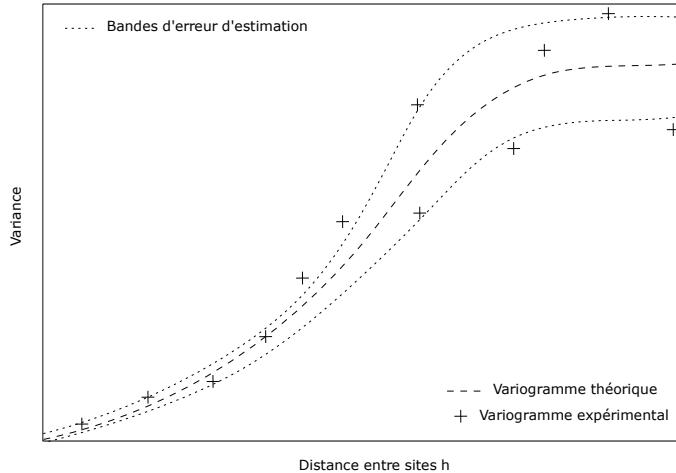


Fig. 2.17. Le variogramme expérimental est d'autant plus incertain que h augmente.

Exemple 2.3. Pour 150 sites observés sur une zone montagneuse de $1 \text{ km} \times 1 \text{ km}$, le diamètre du domaine vaut $\sqrt{2} \approx 1.4 \text{ km}$. On ne calculera le variogramme expérimental que pour $h \in [0; 0.7] \text{ km}$.

Activité IV. Variogramme expérimental (bis repetita)

Dans cette activité, nous allons écrire un code R pour calculer un variogramme expérimental, à partir de données réelles d'altitude. Pour collecter les données, nous allons utiliser l'application *Google Maps find altitude*, disponible gratuitement en ligne à l'adresse suivante :

<https://www.daftlogic.com/sandbox-google-maps-find-altitude.htm>

En cas de difficulté bloquante dans la phase de collecte des données, on pourra récupérer directement l'un des deux fichiers `sample_data1.txt` ou `sample_data2.txt` disponibles dans le répertoire de matériel pédagogique. Dans ce cas, on passera directement à la question Q4.

Q1. Accéder à l'application *Google Maps find altitude* et positionner la carte sur une zone rectangulaire (de préférence montagneuse ou au moins vallonnée), de taille au plus de l'ordre de quelques km de côté.

Q2. Effectuer un clic bref sur la carte pour saisir un point d'observation. Un marqueur bleu doit alors apparaître. Repéter l'opération jusqu'à obtenir un nombre n de points. Plus la zone sélectionnée est grande, plus il faudra un nombre important d'observations. Pour cette activité, un semis de l'ordre de 50 observations devrait être suffisant, cependant les activités du chapitre 3 nécessiteront un nombre plus conséquent de données pour construire des MNT relativement fidèles à la réalité. Dans cette optique, on pourra dès maintenant saisir 200 à 300 points d'observation. On essaiera de produire une couverture homogène de la zone, mais sans nécessairement former une grille régulière.

Q3. Les observations collectées sont listées dans une fenêtre de texte, un peu plus bas dans la page web. Copier le contenu de cette fenêtre dans un fichier `sample_data.txt`.

Q4. Charger le contenu du fichier dans l'environnement R, avec la commande :

```
obs = read.csv("sample_data.txt")
```

Visualiser les valeurs des différents champs. On remarque que les positions de points sont exprimées en coordonnées géographiques (degrés décimaux). Cela va poser problème par la suite pour calculer des distances entre points. On va donc convertir ces données dans un système de coordonnées planes :

```
R = 6378137.0
obs$longitude = obs$longitude*pi/180.0
obs$latitude = obs$latitude*pi/180.0
X = (obs$longitude - min(obs$longitude))*cos(mean(obs$latitude))*R
Y = (obs$latitude - min(obs$latitude))*R
Z = obs$meters
```

où R désigne le rayon moyen de la Terre (en m). Les deux lignes suivantes assurent la conversion des angles des degrés décimaux vers les radians. Le facteur $\cos(\text{mean}(\text{obs$latitude})) * R$ correspond à la longueur métrique d'un arc de parallèle de 1 radian, à la latitude moyenne $\text{mean}(\text{obs$latitude})$ du chantier. Par commodité, les coordonnées métriques des points sont stockées dans 3 vecteurs numériques de même longueur : X, Y et Z.

Q5. Récupérer les coordonnées extrémales et en déduire le diamètre L de la zone d'étude.

Q6. Calculer et représenter le variogramme en 30 points h_1, h_2, \dots, h_{30} régulièrement espacés dans l'intervalle $[0, \frac{L}{2}]$.

Correction :

Script 5. Variogramme expérimental (bis repetita) [code3.r]

```

n = length(X) # Nombre d'observations

# Calcul du diametre de la zone
dX = max(X)-min(X);
dY = max(Y)-min(Y);
diam = sqrt(dX*dX + dY*dY)

# Vecteur de distances a evaluer
dmax = diam/2
dh = dmax/30
H = (dh/2) + (0:30)*dh

# -----
# Calcul du variogramme
# -----
GAMMA = rep(0, length(H))
for (k in 1:length(H)){
  ck = 0
  for (i in 1:n){
    for (j in 1:n){
      dX = X[i]-X[j];
      dY = Y[i]-Y[j];
      d = sqrt(dX*dX+dY*dY)
      if ((d >= H[k]-dh/2) && (d <= H[k]+dh/2)){
        dZ = Z[i]-Z[j]
        GAMMA[k] = GAMMA[k] + dZ*dZ
        ck = ck + 1
      }
    }
  }
  cat(paste("Step", k, H[k], ck, GAMMA[k], "\r\n", sep=" "))
  GAMMA[k] = GAMMA[k]/(2*ck)
}

plot(H, GAMMA, pch=3)

```

On remarque que dans le code proposé ci-dessus, on parcourt chaque couple (i, j) de sites, deux fois, ainsi que tous les couples (i, i) , qui ne contribuent pourtant pas au calcul du variogramme. On peut accélérer le code (d'un facteur 2 environ), en ajustant adéquatement les paramètres de boucle :

```

for (i in 1:(n-1)){
  for (j in (i+1):n){
    # On traite ici le couple (i,j)...
  }
}

```

Pour accélérer davantage le calcul du variogramme (qui peut être une étape très chronophage lorsque

plusieurs milliers de points d'observation sont disponibles), on peut adopter la stratégie suivante : plutôt que de parcourir tous les couples de sites pour chaque valeur h en laquelle on souhaite évaluer le variogramme, une solution alternative consiste à ne parcourir qu'une seule fois l'ensemble des couples d'observations, puis pour chaque couple, évaluer sa distance et l'affecter à la classe correspondante. En retour, notons que cela nécessite de garder en mémoire le décompte des nombres de couples utilisés dans l'estimation pour chaque valeur de h .

Script 6. Version optimisée [code3.r]

```
GAMMA = rep(0, length(H))
COUNT = rep(0, length(H))

for (i in 1:(n-1)){
  for (j in (i+1):n){
    dX = X[i]-X[j]; dY = Y[i]-Y[j]; dZ = Z[i]-Z[j];
    d = sqrt(dX*dX+dY*dY)
    if (d > max(H)+dh/2){
      next
    }
    id = round((d+dh/2)/dh)
    GAMMA[id] = GAMMA[id] + dZ*dZ
    COUNT[id] = COUNT[id] + 1
  }
}
GAMMA = GAMMA / (2*COUNT)
```

Avec cette nouvelle version, sur le jeu de données `sample_data1.txt` (393 observations), on divise par 10 environ le temps de calcul. On pourra alors inspecter visuellement le variogramme. En particulier, on observe que les 3 derniers points semblent dévier de la tendance globale. On entre dans la zone non-robuste dont nous parlions précédemment.

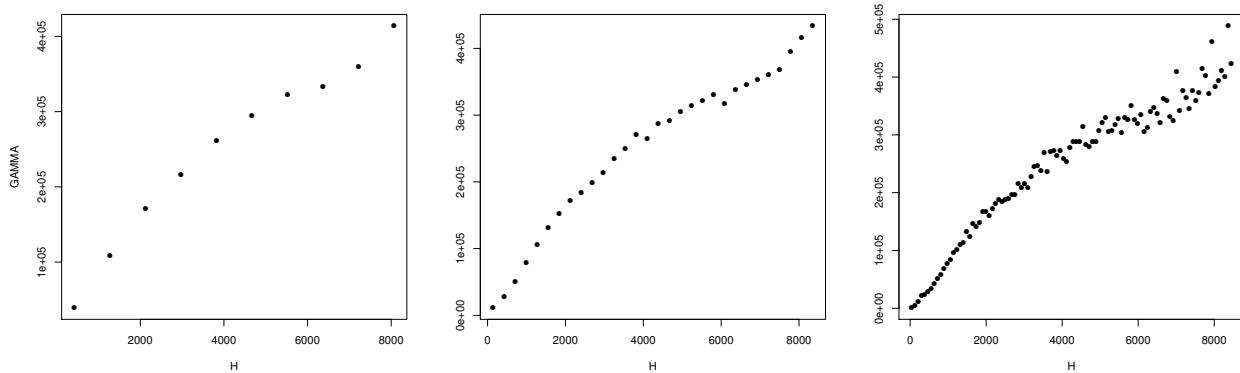


Fig. 2.18. Calcul du variogramme pour 3 pas de discréttisation différents. À gauche : $\Delta h = 850$ m (10 échantillons) - le variogramme est peu résolu, mais les points calculés sont précis (excepté probablement le dernier point). À droite : $\Delta h = 85$ m (100 échantillons) - le variogramme est très résolu, mais complètement instable. Au centre : $\Delta h = 280$ m (30 échantillons) - compromis idéal.

2.4.2 Interpréter le variogramme

Etre capable de calculer programmer informatiquement le calcul d'un variogramme est certes une compétence intéressante, mais pas indispensable, dans la mesure il existe déjà de nombreuses bibliothèques⁷ qui le font très bien dans les principaux langages de programmation. En revanche, il est capital de savoir interpréter graphiquement un variogramme.

On donne ci-dessous en figure 2.19 la représentation schématique d'un variogramme borné.

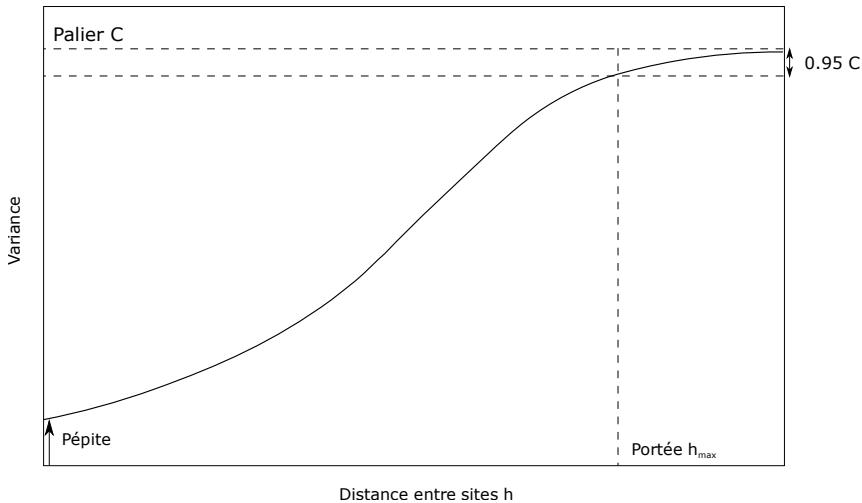


Fig. 2.19. Les 3 éléments caractéristiques d'un variogramme : palier C , portée h_{max} et pépite η .

Le palier

Nous avons expliqué précédemment qu'un variogramme pouvait tendre vers l'infini à mesure que h augmente. Il s'agit d'un cas particulier où la covariance n'existe pas. Dans la grande majorité des cas rencontrés en pratique, le variogramme converge vers une valeur C , le **palier**, correspondant à la dissemblance maximale entre sites lointains. Informellement, on peut dire que lorsque deux sites x_i et x_j sont séparés d'une grande distance, les valeurs Z_i et Z_j prises par le phénomène respectivement en chacun des sites, deviennent complètement décorrélées. La covariance $C(h)$ tend vers 0, et le théorème 26 nous montre alors que le variogramme $\gamma(h)$ tend à devenir égal à $C(0) = \text{Var}[Z]$. Autrement dit, la connaissance de Z_i n'apporte aucune information (statistiquement, on dit que cette connaissance ne réduit pas la variance) sur Z_j , et réciproquement.

Dans le cas pratique de la figure 2.18, on peut récupérer une approximation du palier en moyennant par exemple les 5 dernières valeurs de γ (pour h compris entre 7.2 et 8.4 km) :

```
C = mean(GAMMA[26:30])      On trouve 394 678 m2
sqrtC = sqrt(C)              On trouve 628.234 m
```

On compare alors cette valeur à l'écart-type *a priori* du champ, que l'on approxime par l'écart-type des observations d'altitudes :

```
sigma = sd(Z)    On trouve 556.495 m
```

⁷Par exemple `gstat` pour le langage R, `variogram` (Matlab), `VARFIT` (FORTRAN) ou encore `scikit-gstat` (Python).

Définition 29 (Palier)

Le palier C d'un variogramme γ (lorsqu'il existe) est la limite de γ en l'infini :

$$C = \lim_{h \rightarrow +\infty} \gamma(h)$$

À nouveau, ici aussi, la locution "lorsqu'il existe" précise que C peut prendre une valeur infinie, auquel cas on ne peut pas lui donner de sens physique.

La portée

La valeur à partir de laquelle le variogramme atteint à peu près son palier est appelée la *portée*. Elle représente la distance à partir de laquelle le processus n'est plus autocorrélé. On retient traditionnellement la première valeur qui dépasse 95% de la valeur du palier (notons que si cette règle simpliste fonctionne dans la grande majorité des cas, elle peut poser problème avec les variogrammes dits *à effet de trou*, comme nous le verrons plus loin) :

```
hmax = H[min(which(GAMMA > 0.95*C))]
```

On obtient une portée de 7.8 km environ (7786 m). Dans le cas d'un MNT, on s'attend donc à trouver des montagnes et vallées de taille caractéristique (on pourrait parler de périodicité spatiale) 8 km.

Définition 30 (Portée)

Pour un seuil de tolérance $\varepsilon > 0$ fixé, la portée d'un variogramme γ de palier C est la plus petite valeur h_{max} telle que :

$$\forall h \geq h_{max} : 1 - \varepsilon \leq \frac{\gamma(h)}{C} \leq 1 + \varepsilon$$

La pépite

Tout variogramme est théoriquement nul en 0. En effet, d'après la définition 25 :

$$\gamma(0) = \frac{1}{2} \text{Var}[Z(x) - Z(x)] = 0$$

En revanche, rien n'interdit d'avoir une discontinuité en 0, *i.e.* un saut faisant passer le variogramme de la valeur 0 à une valeur η , pour $h > 0$ aussi petit soit-il.

Cette valeur η , appelée *pépite*⁸ (ou *nugget* en anglais) dénote la présence d'un bruit de fond dans le processus. D'un point de vue variographique, cela signifie que l'accroissement entre deux sites infinitésimamente voisins (mais distincts) x_1 et x_2 aura toujours une variabilité résiduelle, ce qui, au niveau des réalisations, se traduit par des discontinuités locales dans la variable régionalisée.

⁸En référence à la variographie des gisments d'or

Définition 31 (Pépite)

Soit γ un variogramme quelconque (non-nécessairement borné). On appelle *pépite* du variogramme, la quantité (éventuellement nulle) :

$$\eta = \lim_{h \rightarrow 0^+} \gamma(h)$$

Remarque : lorsque la pépite est non nulle, on dit que le variogramme, et par extension le processus stochastique ainsi que ses réalisations, sont *pépitiques*.

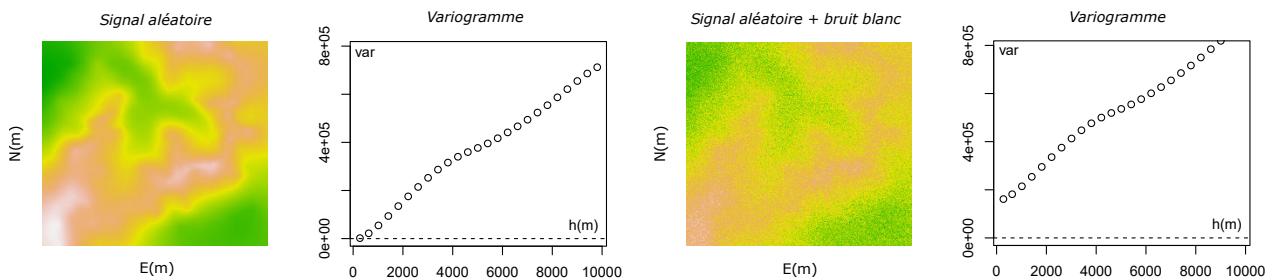


Fig. 2.20. Illustration de l'effet de pépite. À gauche : réalisation d'un processus stochastique sans effet de pépite et variogramme associé. À droite : idem avec effet de pépite. On mesure $\eta \approx 2.10^5$, donc l'écart type du bruit additif vaut $\sigma = \sqrt{\eta} = \sqrt{2.10^5} \approx 440$ m.

Notons que l'effet de pépite peut être utilisé pour modéliser le comportement d'un phénomène, mais aussi (et surtout) pour modéliser un bruit dans le processus d'observation. Par exemple, si on souhaite construire un MNT à partir de mesures d'altitudes effectuées avec un GPS de randonnée (précision ± 5 m), on pourra alors utiliser un variogramme pépitique avec $\eta = 5^2 = 25$ m².

Pour aller plus loin...

Le comportement du variogramme au voisinage de l'origine (et plus spécifiquement son rythme de croissance) fournit en réalité beaucoup d'information sur la régularité des réalisations. L'effet de pépite n'en est qu'un cas particulier, comme le montre la propriété suivante :

Propriété 32 (Comportement du variogramme à l'origine)

La pente en 0 du variogramme reflète le degré de régularité de la variable régionalisée :

- **Parabolique** : ($\gamma'(0) = 0$), la variable régionalisée est **dérivable**.
- **Linéaire** : ($\gamma'(0) = a$), la variable régionalisée est **continue** mais **non-dérivable**.
- **Discontinu** : ($\gamma'(0) = +\infty$) la variable régionalisée est **discontinue**.

Remarque : chaque réalisation du processus étant par nature différente, la fonction dont on cherche à quantifier la régularité est une fonction aléatoire. Les notions de continuité et de dérivabilité telles qu'on les connaît ne sont alors plus opérantes, et on devrait en toute rigueur introduire les notions probabilistes de continuité et dérivabilité dites *en moyenne quadratique*. Ici, par abus de langage on dit qu'une variable régionalisée est continue (resp. dérivable) lorsque les réalisations du processus associés à cette variable sont statistiquement localement continues (resp. dérivables). L'annexe A donne quelques clés supplémentaires pour saisir cette notion. On pourra trouver une démonstration de cette propriété dans [Picinbono \(1998\)](#) ou encore dans [Allard \(2012\)](#).

Grossièrement, le comportement à l'origine du variogramme est un indicateur du rythme avec lequel les valeurs prises par un phénomène en deux sites x_1 et x_2 infinitésimamente proches se décorrèlent à mesure qu'on les éloigne mutuellement. Un comportement **parabolique** indique une pente de γ nulle en 0, et donc que γ croît lentement à l'origine, à l'instar de la fonction parabolique $x \mapsto x^2$. Les réalisations du processus sont alors suffisamment lisses pour être dérivables. Dans le cas d'un comportement **linéaire**, $\gamma'(0) = a$ la décorrélation se produit à un rythme plus rapide, les réalisations restent continues mais ne sont plus dérivables. Enfin, remarquons qu'un comportement **discontinu** est synonyme d'effet de pépite. Il s'agit du cas limite du comportement linéaire pour $a = +\infty$.

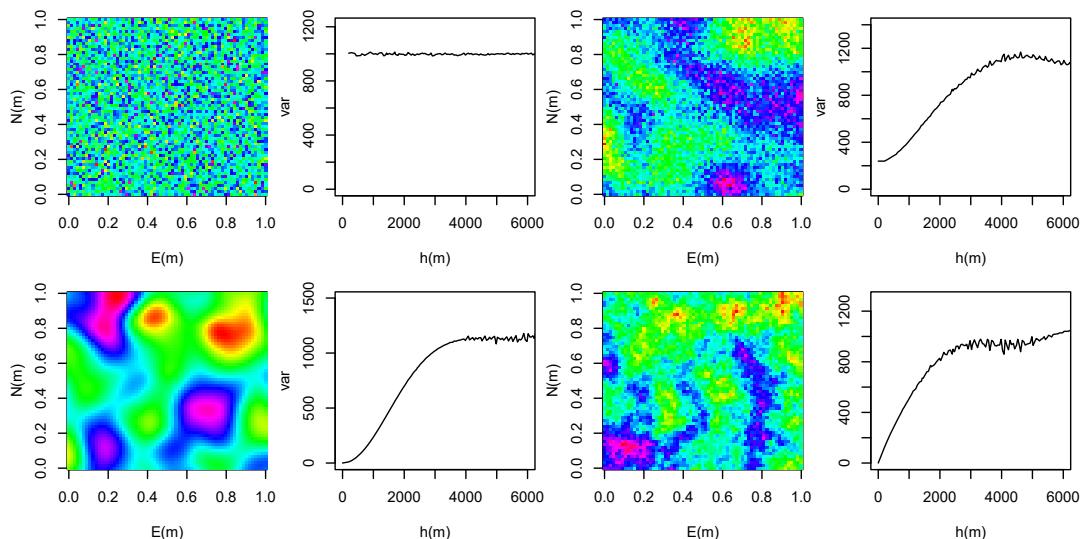


Fig. 2.21. Illustration du lien entre la régularité de la variable régionalisée (à gauche en couleur) et le comportement à l'origine de son variogramme (à droite en noir et blanc). En haut à gauche : pépite pur. En haut à droite : comportement parabolique + pépite. En bas à gauche : comportement parabolique. En bas à droite : comportement linéaire.

La figure 2.21 illustre l'importance du comportement à l'origine du variogramme sur l'aspect local (lisse ou erratique) de ses réalisations. En particulier, on remarque que le champ à comportement linéaire (en bas à droite) ressemble dans une certaine mesure au comportement pépitique et parabolique (en haut à droite). Parfois, seule l'analyse variographique permet de différencier deux processus dont les réalisations semblent similaires au premier abord.

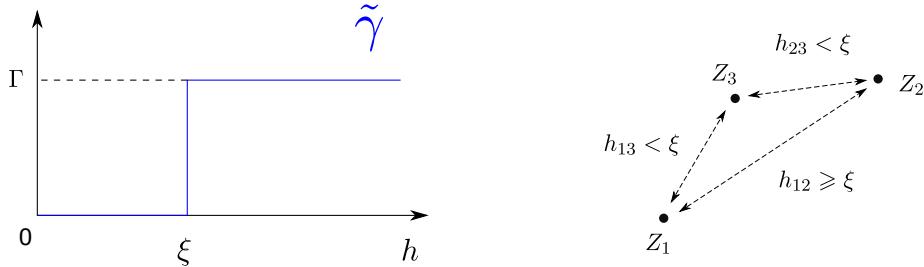
2.4.3 Les modèles de variogramme

Considérons une fonction quelconque $f : \mathbb{R}^+ \rightarrow \mathbb{R}^+$. Nous l'avons vu précédemment, la contrainte $f(0) = 0$ est une condition nécessaire pour que f puisse être un variogramme. Est-ce une condition suffisante ? La réponse est non : la nature statistique du variogramme impose que f possède une propriété supplémentaire. Voyons cela à travers un exemple simple.

Exemple 2.4. On considère l'exemple d'une fonction $\tilde{\gamma}$ définie par :

$$\tilde{\gamma}(h) = \begin{cases} 0 & \text{si } h < \xi \\ \Gamma & \text{sinon.} \end{cases} \quad (2.10)$$

avec Γ et ξ deux valeurs réelles positives quelconques. Prenons 3 sites, positionnés comme ci-dessous (à droite), pour lesquels on note Z_1 , Z_2 et Z_3 les valeurs prises par le phénomène de variogramme $\tilde{\gamma}$.



On relève alors une contradiction : les variables (Z_1, Z_3) d'une part et (Z_2, Z_3) d'autre part, étant séparées d'une distance inférieure au seuil ξ , elles sont complètement corrélées et elles sont donc liées par une relation affine. On montre alors aisément que Z_1 et Z_2 sont nécessairement aussi liées par une relation affine, et donc également parfaitement corrélées. Or, Z_1 et Z_2 sont séparées d'une distance supérieure à ξ , donc $\text{Cov}(Z_1, Z_2) = \tilde{\gamma}(\infty) - \tilde{\gamma}(h_{12}) = 0$, ce qui est une contradiction⁹. La fonction $\tilde{\gamma}$ ne peut donc pas être un modèle de variogramme.

D'après l'exemple 2.4, il apparaît clairement qu'on ne peut pas choisir n'importe quelle fonction pour modéliser le variogramme. Plus formellement, γ doit vérifier une propriété de *défini-négativité conditionnelle*, qui assure que les quantités algébriques calculées à l'aide du variogramme et auxquelles on peut donner le sens d'une variance, sont toujours positives Allard (2012). Nous reviendrons sur ce point quelque peu technique dans la section 2.4.5.

En conséquence, le variogramme doit appartenir à une classe particulière de modèles, dont nous donnons les principaux représentants dans cette section.

Le modèle linéaire

$$\gamma(h) = kh \quad (2.11)$$

où k est une constante réelle positive.

Le variogramme linéaire permet de modéliser une décorrélation progressive des valeurs prises par le phénomène, à mesure que les sites considérés sont éloignés spatialement. Il ne converge pas vers une valeur palier, et n'a donc pas non plus de valeur de portée définie. En conséquence, il n'existe pas de covariogramme associé¹⁰. Son comportement à l'origine est bien évidemment linéaire.

Le modèle linéaire présente l'avantage d'être extrêmement simple à estimer. Par ailleurs, dans certaines applications comme l'interpolation par krigeage (*cf* chapitre 3), le résultat obtenu est invariant par une multiplication du variogramme par un scalaire. Dans cette optique, le coefficient k n'a pas

⁹Sauf si ξ est nul, auquel cas on est réduit au variogramme purement pépitiques, qui est parfaitement valide

¹⁰Ce modèle de variogramme convient particulièrement pour les processus stochastiques ne respectant que l'hypothèse de stationnarité intrinsèque, et non celle de stationnarité au sens large, comme expliqué dans l'exemple 2.2.

besoin d'être estimé, et on pourra utiliser arbitrairement le modèle de variogramme prêt-à-l'emploi $\gamma : h \mapsto h$. Le modèle linéaire devra être utilisé pour les processus stochastiques dont la dispersion explose à mesure que grandit la taille de la zone d'étude (Arnaud et Emery, 2000). Il convient en particulier pour la modélisation des gisements de cuivre.

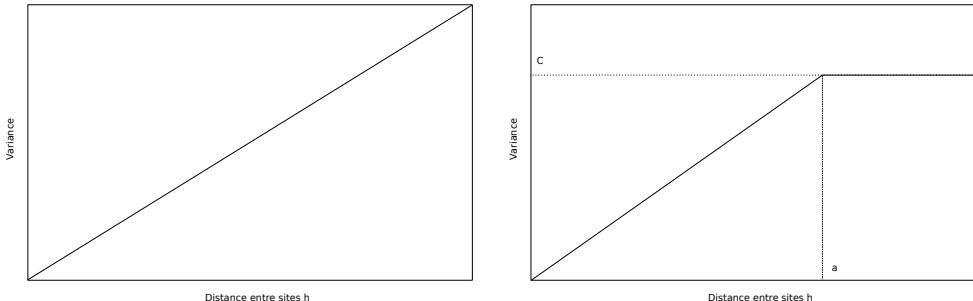


Fig. 2.22. Variogramme linéaire (à gauche) et linéaire avec palier (à droite).

Si $k = 0$, le variogramme dégénère en $\gamma : h \mapsto 0$, traduisant un champ aléatoire constant.

Le modèle linéaire avec palier

$$\gamma(h) = \begin{cases} \frac{C}{a}h & \text{si } h \leq a \\ C & \text{sinon.} \end{cases} \quad (2.12)$$

où C et a sont deux constantes réelles positives.

Dans une version simplifiée, comme nous le verrons dans l'activité V, les paramètres du modèle sont ajustés en déterminant empiriquement la portée a , puis en effectuant une régression linéaire sur la portion $[0, a]$. Ce modèle présente l'avantage de fournir une approximation relativement fiable de modèles analytiquement plus complexes.

Ce variogramme dégénère en un modèle pépitique pur quand a tend à s'annuler.

Le modèle puissance

$$\gamma(h) = kh^\theta \quad \text{avec } 0 < \theta \leq 2 \quad (2.13)$$

Quand $\theta = 1$, on retrouve le modèle linéaire. Pour $\theta > 1$, on a $\gamma'(h) = \theta kh^{\theta-1}$ et donc $\gamma'(0) = 0$, le variogramme a un comportement parabolique à l'origine, traduisant ainsi des réalisations plus lisses que le modèle linéaire. Pour $\theta < 1$, $\gamma'(0)$ est infini, le comportement à l'origine est quasi-pépitique. Tout comme pour le modèle linéaire, la covariance n'existe pas.

L'estimation des paramètres est immédiate par régression affine dans l'espace log-log. Un intérêt principal de ce modèle est de pouvoir modéliser des comportements à l'origine parabolique, linéaire et quasi-pépitique avec un unique jeu de paramètres, ce qui offre une grande variabilité de cas d'application. Le variogramme puissance devrait donc être utilisé lorsque l'on n'a *a priori* aucune idée du comportement variographique du phénomène étudié. Remarquons que le cas spécial $\theta = 2$ bien que licite

(comme nous le verrons dans la section 2.4.5), résulte en un processus non-ergodique, et donc inutilisable en Géostatistique.

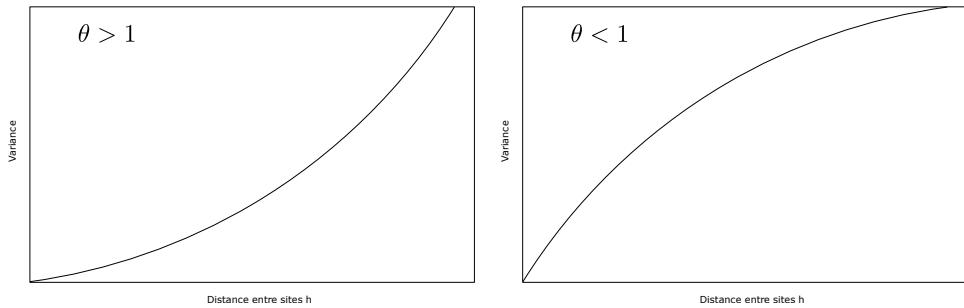


Fig. 2.23. Variogramme puissance au comportement supra-linéaire (à gauche) et sub-linéaire (à droite). Le comportement à l'origine parabolique du variogramme de gauche produira statistiquement des réalisations plus lisses que celles du variogramme de droite.

On pourra trouver un exemple pédagogique d'utilisation de ce variogramme, avec $\theta \approx 1.31$ pour la modélisation de l'indice de porosité du sol en fonction de la profondeur (Gringarten et Deutsch, 2001).

Le modèle exponentiel

$$\gamma(h) = C(1 - e^{-\frac{h}{a}}) \quad (2.14)$$

où C et a sont deux constantes réelles positives.

Le palier n'est jamais rigoureusement atteint. La portée est définie comme la distance à laquelle on atteint 95% de la limite asymptotique C du variogramme. Elle vaut donc $h_{max} = 3a$.

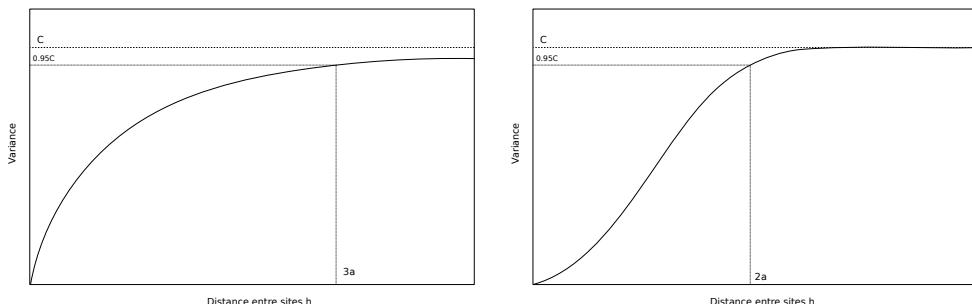


Fig. 2.24. Variogramme exponentiel (à gauche) et gaussien (à droite).

Par ailleurs $\gamma'(h) = \frac{C}{a}e^{-\frac{h}{a}}$ et $\gamma'(0) = \frac{C}{a} > 0$, le comportement à l'origine est linéaire.

Le modèle exponentiel convient en particulier pour modéliser les erreurs GPS, et pourra donc dans ce cadre être utilisé par exemple pour améliorer les systèmes de guidage (Grejner-Brzezinska et al., 2005) et de navigation (Meneroux, 2019).

Il caractérise notamment un bruit rouge, c'est-à-dire dont la puissance décroît comme le carré de la fréquence. En 1827, le botaniste écossais Robert Brown, découvre que les particules du pollen sont

soumise à un tel type de bruit. En particulier, il est solution de l'équation de Langevin, caractérisant le mouvement d'une particule ponctuelle soumise à un frottement proportionnel à la vitesse ainsi qu'à une force de collision aléatoire. Ce bruit, depuis rebaptisé *bruit brownien*, sera plus tard à la base des hypothèses nécessaires à l'établissement d'une loi célèbre reliant température, pression et volume d'un gaz, dite loi des gaz parfaits. En acoustique, un signal brownien évoque le bruit des vagues et aurait des vertus thérapeutiques contre l'hyperacousie.

Le modèle gaussien

C'est avec le modèle exponentiel, l'un des deux modèles les plus couramment utilisés.

$$\gamma(h) = C(1 - e^{-\frac{h^2}{a^2}}) \quad (2.15)$$

où C et a sont deux constantes réelles positives.

La portée est définie comme la distance à laquelle on atteint 95% de la limite asymptotique C du variogramme. Elle vaut $h_{max} = \sqrt{\ln(20)}a \approx 1.73a$.

Par ailleurs, $\gamma'(h) = \frac{2Ch}{a^2} \exp(-h^2/a^2)$ et $\gamma'(0) = 0$, le comportement à l'origine est parabolique. Plus généralement, on montre facilement que la dérivée n -ème $\gamma^{(n)}$ du modèle, est nulle en l'origine, pour tout entier $n \geq 0$, d'où l'aspect très lisse des réalisations d'un processus de variogramme γ , ce qui convient en particulier à la modélisation topographique.

Par ailleurs, la fonction gaussienne possède de nombreuses propriétés élégantes, voire fondamentales¹¹ (qu'il serait malheureusement trop long d'exposer ici), lui conférant ainsi un rôle central à la fois en théorie des probabilités et en traitement du signal. C'est donc tout naturellement qu'on la retrouve à une place de choix dans la liste des modèles de variogramme.

Le modèle sphérique

$$\gamma(h) = \begin{cases} C\left(\frac{3}{2}\frac{h}{a} - \frac{1}{2}\frac{h^3}{a^3}\right) & \text{si } h \leq a \\ C & \text{sinon.} \end{cases} \quad (2.16)$$

où C et a sont deux constantes réelles positives.

Ici aussi, on retrouve le modèle pépitique pur quand a tend à s'annuler. Par ailleurs, on a : $\gamma'(h) = C\left(\frac{3}{2a} - 3\frac{h^2}{a^3}\right)$ et donc $\gamma'(0) = \frac{3}{2a} > 0$, le comportement est linéaire à l'origine.

Le variogramme sphérique a une interprétation physique simple : il modélise les corrélations entre point d'un espace 3D, en supposant que l'aire d'influence de chaque point est une sphère de diamètre a . La corrélation entre deux points est définie comme étant égale au volume de l'intersection des

¹¹En particulier, le théorème central limite, stipule que toute somme de variables aléatoires identiquement distribuées converge vers la loi gaussienne, comme l'illustre parfaitement l'expérience de la planche de Galton. La théorie de l'entropie démontre que c'est la loi la moins arbitraire que l'on puisse choisir lorsqu'on ne connaît que l'espérance et la variance d'une variable aléatoire. On peut également montrer que l'observation d'un signal sur une fenêtre gaussienne réalise le meilleur compromis pour le déterminer à la fois d'un point de vue temporel et fréquentiel...

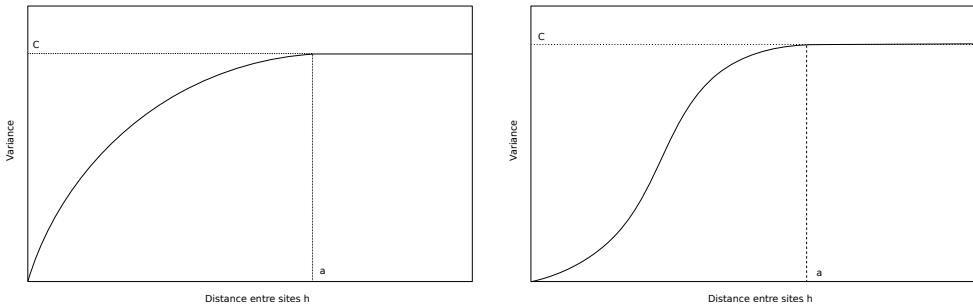
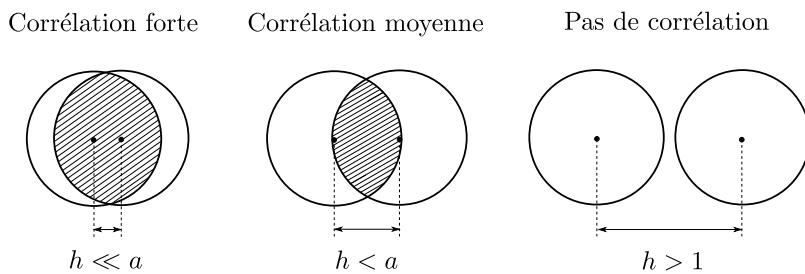


Fig. 2.25. Variogramme sphérique (à gauche) et cubique (à droite).

intérieurs de deux sphères d'influence. Lorsque les points sont séparés d'une distance supérieure à a , on retrouve naturellement que la corrélation est nulle, et le variogramme atteint son palier. Le variogramme sphérique est donc un équivalent tri-dimensionnel du variogramme linéaire avec palier, qui lui mesure la corrélation entre deux points par la longueur de l'intersection des intérieurs de deux segments de longueur a (Allard, 2012).



Le modèle sphérique peut être employé comme une version analytiquement simplifiée du variogramme exponentiel. Il se généralise pour toutes les dimensions d'espace.

Le modèle sphérique

$$\gamma(h) = \begin{cases} C \left(7 \frac{h^2}{a^2} - \frac{35}{4} \frac{h^3}{a^3} + \frac{7}{2} \frac{h^5}{a^5} - \frac{3}{4} \frac{h^7}{a^7} \right) & \text{si } h \leq a \\ C & \text{sinon.} \end{cases} \quad (2.17)$$

où C et a sont deux constantes réelles positives.

On a $\gamma'(0) = 0$ donc le comportement à l'origine est parabolique. Le modèle cubique, qui partage avec le modèle sphérique la propriété d'atteindre son palier, mais qui diffère par son comportement à l'origine, peut être considéré comme une version simplifiée du variogramme gaussien.

Le modèle à effet de trou

$$\gamma(h) = C \left(1 - \frac{\sin(h/a)}{h/a} \right) \quad (2.18)$$

où C et a sont deux constantes réelles positives.

Ici aussi, le palier n'est jamais rigoureusement atteint. La portée peut être définie au niveau de l'une des premières intersections de la courbe avec la limite asymptotique C du variogramme. Différentes mesures du phénomène peuvent être effectuée sur le variogramme (abscisse du premier pic, écart inter-pics, distance du premier creux, etc.) et traduites en termes de caractéristiques de la variable régionalisée (largeur typique des ondulations, etc.). Pour plus d'informations nous renvoyons le lecteur au travail complet de [Pyrcz et Deutsch \(2003\)](#).

Le comportement à l'origine est parabolique : $\gamma'(0) = 0$.

Physiquement, nous savons depuis le début du XIX^e siècle et les études de Joseph Fourier, que tout signal peut être considéré comme une somme (potentiellement infinie) de sinusoïdes : la nature est intrinsèquement oscillatoire. Un phénomène dont la périodicité nous est apparente est un signal dont n'ont été conservé que les basses fréquences. Comme nous le verrons dans le paragraphe [2.4.5](#), le variogramme à effet de trou ne décrit pas autre chose : c'est la caractérisation d'un modèle pépitique pur passé en entrée d'un filtre passe-bas idéal.

Le modèle à effet de trou est adapté pour modéliser les phénomènes (en particulier spatio-temporels) cycliques : la pluviométrie, le niveau de la mer (et d'une manière générale tous les phénomènes liés d'une manière ou d'une autre à la rotation ou à l'orbite de la Terre), l'imagerie par résonnance magnétique, ainsi que certains phénomènes topographiques présentant une forme de périodicité spatiale (plissements, dunes...).

D'une manière plus générale, la littérature de référence désigne par l'appellation variogramme à effet de trou, tout modèle non-monotone ([Journel et Huijbregts, 1978](#)).

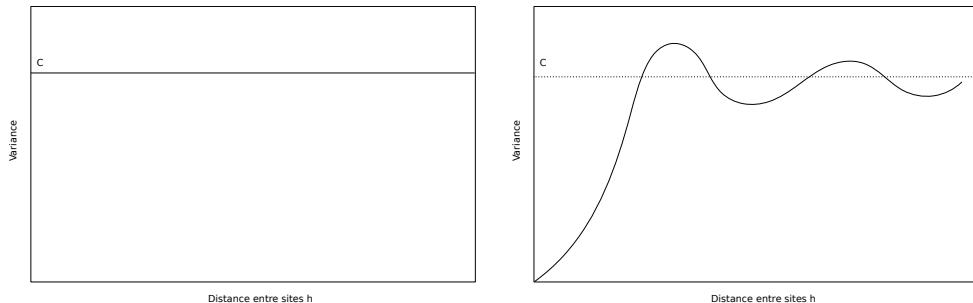


Fig. 2.26. Variogramme pépitique pur (à gauche) et variogramme à effet de trou (à droite).

Le modèle pépitique pur

C'est le cas d'un processus complètement décorrélé. On parle de bruit blanc au sens faible ([Picinbono, 1998](#)). Utilisé seul, il ne présente aucun intérêt.

$$\gamma(h) = \begin{cases} 0 & \text{si } h = 0 \\ C & \text{sinon.} \end{cases} \quad (2.19)$$

où C une constante réelle positive.

γ a un comportement discontinu à l'origine.

Le modèle à effet de pépite

Notons qu'il est possible de combiner l'effet de pépite à tous les modèles de variogrammes précédents. Par exemple, on peut fabriquer un variogramme cubique à effet de pépite.

$$\gamma(h) = \begin{cases} C(7\frac{h^2}{a^2} - \frac{35}{4}\frac{h^3}{a^3} + \frac{7}{2}\frac{h^5}{a^5} - \frac{3}{4}\frac{h^7}{a^7}) & \text{si } h \leq a \\ C & \text{sinon.} \end{cases} \quad (2.20)$$

où C , C_0 et a sont trois constantes réelles positives.

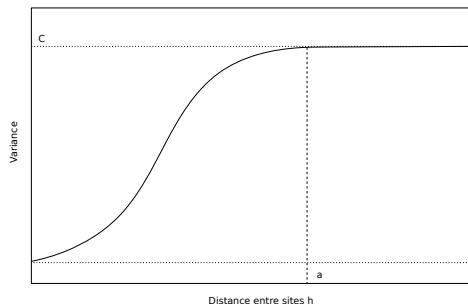


Fig. 2.27. Variogramme cubique avec effet de pépite.

Les gisments d'or par exemple, peuvent se modéliser par un variogramme au comportement erratique (fort effet de pépite), une faible portée et une croissance linéaire à l'origine, donc a priori avec un variogramme exponentiel ou sphérique avec pépite. Le modèle cubique ou gaussien à effet de pépite peut permettre de modéliser les erreurs de mesures (GPS par exemple) sur un MNT.

Les modèles gigognes

Les modèles de variogrammes sont additifs. Autrement dit, si γ_1 et γ_2 sont des variogrammes valides, alors $\gamma_1 + \gamma_2$ en est un aussi¹². Avec ce procédé, on peut facilement construire des variogrammes étagés, dits *variogrammes en gigogne*.

$$\gamma(h) = \gamma_1(h) + \gamma_2(h) + \dots + \gamma_n(h) \quad (2.21)$$

où $\gamma_1, \gamma_2, \dots, \gamma_n$ sont des modèles de variogramme licites.

Il est possible de regrouper ces modèles en trois classes (Baillargeon, 2005) : les modèles **sans palier** (linéaire, puissance) avec **palier exact** (linéaire avec palier, pépitique pur, sphérique et cubique) et enfin avec **palier asymptotique** (gaussien, exponentiel et à effet de trou). Il existe par ailleurs de nombreux autres modèles moins classiques : Matérn, Legendre, Bessel... qui généralisent et complètent les variogrammes ci-dessus.

¹²Notons que ceci n'est pas vrai pour la soustraction.

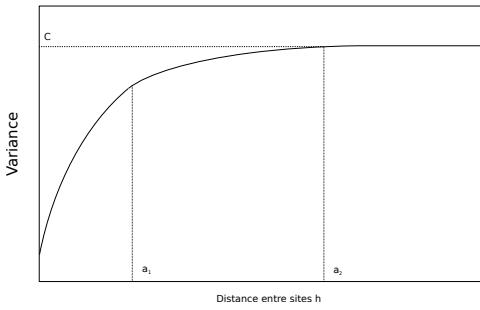


Fig. 2.28. Variogramme gigogne.

Nous présentons en figure 2.29 les 6 modèles de variogramme les plus fréquemment utilisés, avec à chaque fois un exemple de réalisation associée.

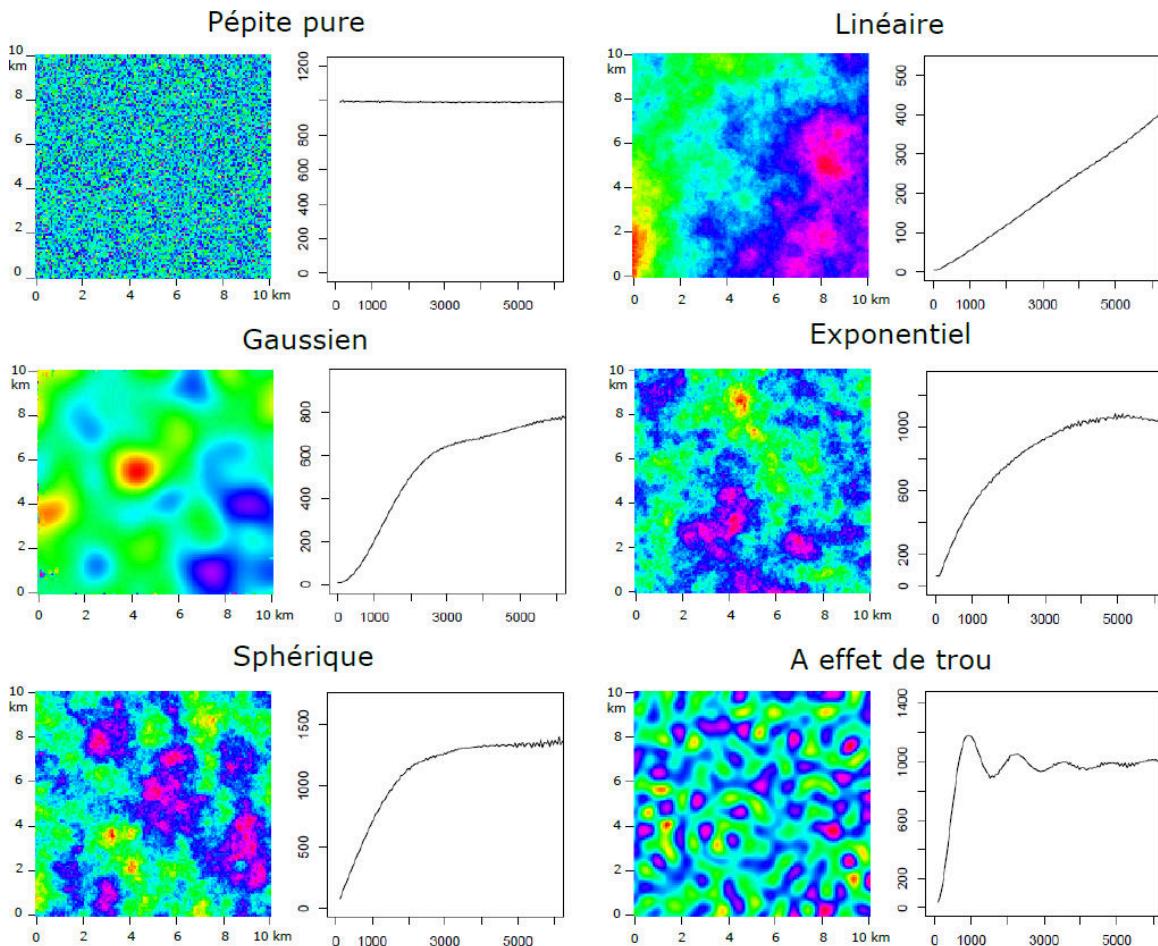


Fig. 2.29. Présentations de 6 modèles de variogramme avec un exemple de réalisation associée.

2.4.4 L'estimation du variogramme

Au bilan, nous avons manipulé jusqu'ici trois concepts différents du variogramme. Il peut être intéressant de les récapituler à ce stade :

- **Le variogramme théorique :** introduit au début de la section 2.4, variogramme réel du phénomène que l'on cherche à modéliser. De par le nombre fini de points d'observation, il s'agit

d'un objet conceptuel, inconnu en pratique, et que l'on cherchera à estimer.

- **Le variogramme expérimental :** ou variogramme empirique (2.4.1), est une estimation du variogramme théorique, calculée à l'aide des observations disponibles. C'est un estimateur bruité (variance statistique d'estimation) et fragmentaire (connu en quelques points seulement).
- **Les modèles de variogrammes :** bibliothèque de fonctions mathématiques satisfaisants les propriétés du variogramme théorique (défini-négativité conditionnelle), et dont les paramètres sont estimés à partir des observations (2.4.3). Les modèles de variogrammes permettent d'établir un pont entre le variogramme théorique insaisissable et son estimation bruitée, le variogramme expérimental.

Après avoir choisi un modèle de variogramme, on effectue l'estimation de ses paramètres à l'aide de méthodes statistiques classiques : moindres carrés linéaires, non-linéaires, pondérés... dont l'étude exhaustive sort malheureusement du cadre de ce cours, et au sujet desquelles nous ne saurions faire mieux que de renvoyer le lecteur intéressé à l'excellent ouvrage de [Sillard \(2001\)](#). Soulignons toutefois que cette phase d'estimation (que l'on résume sur la figure 2.30) est généralement entièrement prise en charge en boîte noire par les solutions logicielles.

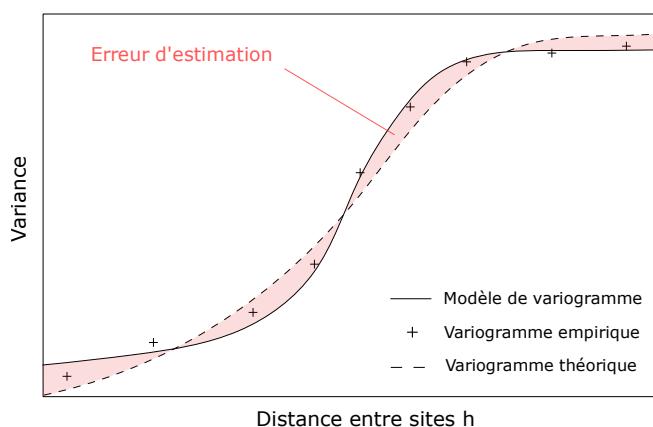


Fig. 2.30. Variogramme théorique inconnu (en ligne pointillée), variogramme expérimental (croix rouges), modèle de variogramme inféré par estimation paramétrique (trait plein) et erreur d'estimation.

Notons qu'il peut parfois se révéler difficile de choisir le bon modèle de variogramme. Certains travaux ont été menés en ce sens pour permettre une sélection automatique ([Barry et al., 1996](#)). En général, les approches employées consistent à tester l'ajustement sur une collection de modèles candidats. Le modèle retenu est alors celui qui offre le meilleur compromis entre parcimonie et qualité d'ajustement¹³.

Activité V. Sélection et estimation d'un modèle de variogramme

Dans cette activité nous allons apprendre concrètement comment choisir un bon modèle de variogramme et estimer ses paramètres (dans un premier temps *à la main*, puis informatiquement) en fonction des caractéristiques supposées de la variable régionalisée.

¹³Étant entendu qu'un modèle avec plus de degrés de liberté permet en général un meilleur ajustement, une approche classique pour éviter le sur-apprentissage consiste à pénaliser le nombre de paramètres du modèle.

On donne en figure 2.31 cinq exemples de réalisations d'un processus stochastiques (A à E), avec leurs cinq variogrammes empiriques associés (dans le désordre : I à V).

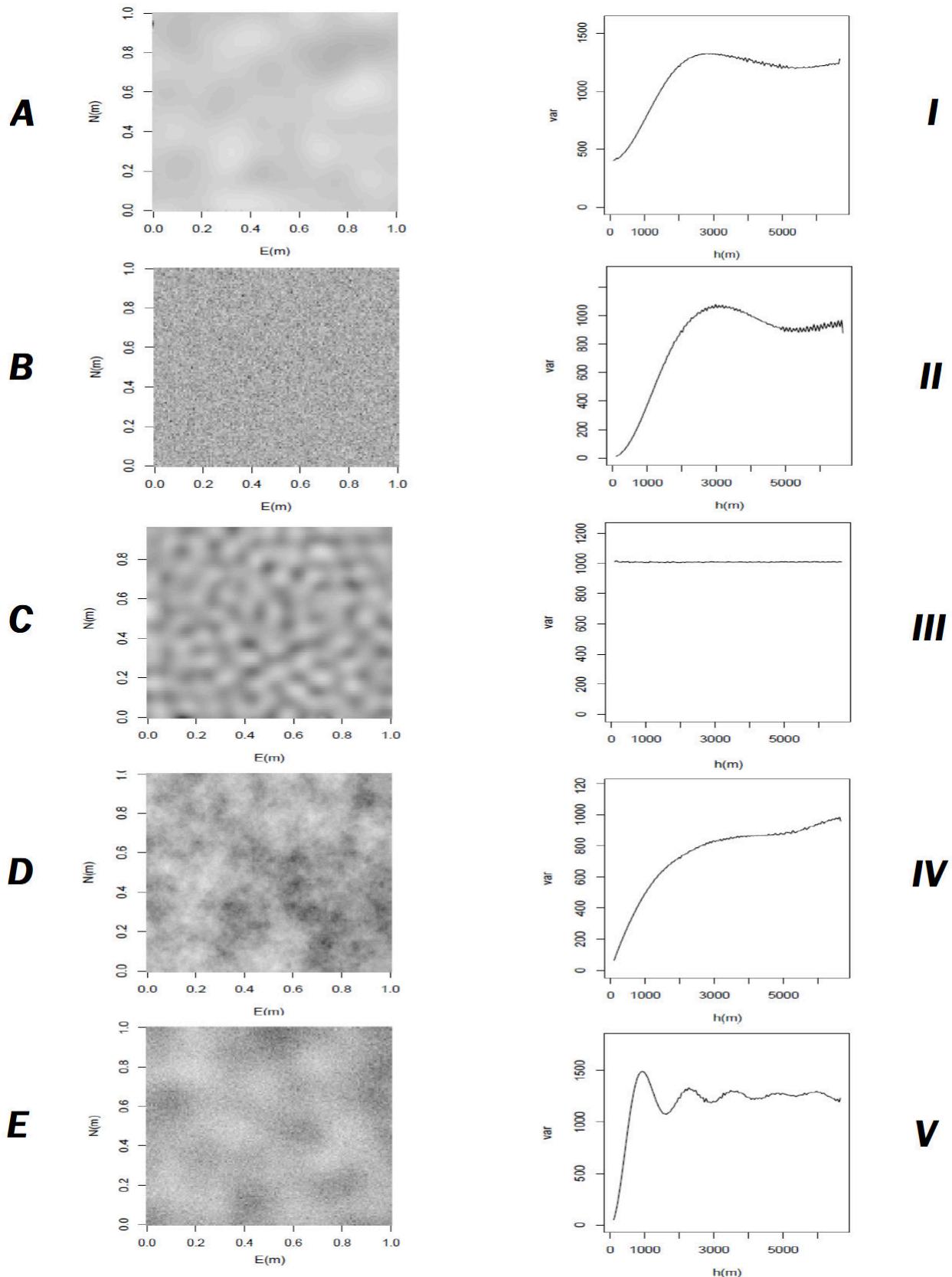


Fig. 2.31. 5 exemples de réalisations d'un processus stochastiques (colonne de gauche) et les 5 variogrammes empiriques associés (colonne de droite dans le désordre).

Q1. Associer chaque réalisation avec son variogramme correspondant.

Q2. On considère le variogramme expérimental de la fig. 2.32, construit à partir d'un semis de relevés d'altitude. Choisir (en justifiant) un modèle de variogramme, et le représenter sur le graphique.

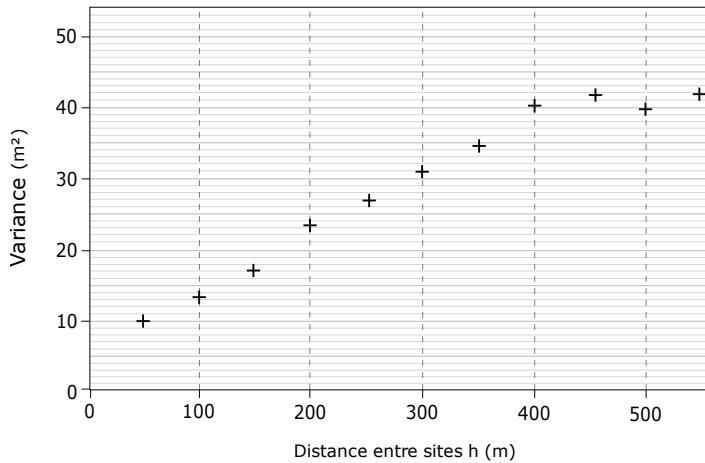


Fig. 2.32. Le variogramme expérimental d'un ensemble de relevés altimétriques.

Q3. Reprendre la question Q2 et calculer informatiquement les paramètres du variogramme.

Correction :

Q1. La réalisation B est un bruit blanc (signal sans aucune structure de corrélation spatiale). Son variogramme est donc à pépite pure, donc constant et égal à $\Gamma > 0$ sur $]0; +\infty[$ et nul en 0 : c'est le variogramme III. La réalisation C possède un motif d'ondulations périodiques, ce qui se modélise à l'aide d'un variogramme non-monotone dit à *effet de trou* : le modèle V. Parmi les 3 réalisations restantes A, D et E, seule la variable régionalisée E semble bruitée. On recherche donc un effet de pépite dans le variogramme, que l'on trouve dans le modèle I. Au passage, la représentation variographique nous permet de calculer le ration signal sur bruit (ou SNR pour *Signal to Noise Ratio* en anglais) du processus : c'est le rapport de la variance de l'information utile sur la variance du bruit¹⁴. En termes variographiques, c'est le rapport de la différence palier - pépite sur la pépite :

$$\text{SNR} = \frac{C - \eta}{\eta} = \frac{C}{\eta} - 1$$

Pour une pépite pure ($\eta = C$) et on retrouve bien un SNR nul (aucun signal dans le bruit). Ici, dans le cas du variogramme I, on a $C \approx 1200$ et $\eta \approx 450$, d'où un SNR de l'ordre de 1.7 dans la réalisation E (1.7× plus de signal que de bruit). Les SNR de tous les autres modèles (hormis I et III) sont infinis.

Enfin, entre les deux réalisations restantes (A et D), A paraît plus régulière, on lui associe donc II, pour son comportement parabolique à l'origine (notons que le processus ayant généré A semble être une version débruitée de celui ayant généré E). La réalisation D, plus irrégulière, correspond donc au

¹⁴Physiquement, cette quantité correspond également au rapport de la puissance du signal utile sur le bruit de fond. En télécommunication, plus le SNR est réduit, plus la réception du signal est délicate.

variogramme IV, au comportement clairement linéaire à l'origine.

Réalisation	Variogramme	Modèle	Portée (m)
A	II	Gaussien/Cubique	3000
B	III	Pépite pure	0
C	V	Effet de trou	5000
D	IV	Exp/Sphérique/Puissance	3000
E	I	[Gaussien/Cubique] + pépite	3000

Q2. Par simplicité, on propose d'utiliser une modèle de variogramme linéaire avec palier.

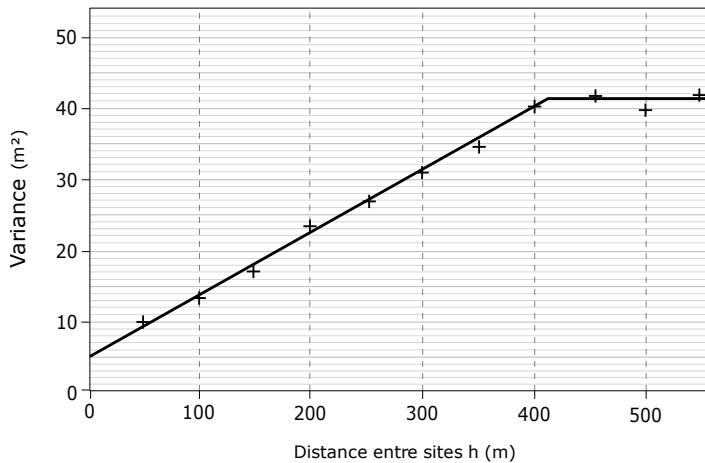


Fig. 2.33. Variogramme expérimental d'un ensemble de relevés altimétriques et ajustement empirique d'un modèle de variogramme linéaire avec palier.

On estime alors graphiquement : portée $a = 410$ m, pépite $\eta = 5 \text{ m}^2$ et palier $C = 41 \text{ m}^2$, soit une pente $(C - a)/\eta = 0.088 \text{ m}^{-1}$ et un ratio signal sur bruit de 7.2.

Remarquons que ce choix simplificateur a un prix : la topographie est en général assez lisse, alors que le modèle utilisé ici a un comportement linéaire à l'origine. En fonction de la géologie du milieu considéré, peut-être qu'un modèle gaussien ou cubique avec pépite serait plus adapté.

Q3. Informatiquement, on peut estimer les paramètres plus rigoureusement. On commence par relever les valeurs du graphe pour les consigner informatiquement :

```
G = c(10, 13.5, 17, 23.5, 27, 31, 34.5, 40.5, 42, 40, 42)
n = length(G); H = (1:n)*50
```

On peut alors estimer les paramètres optimaux du variogramme. On rappelle que la droite de régression $y(x) = \beta_1 x + \beta_2$ entre deux séries de nombres X et Y a pour paramètres optimaux :

$$\hat{\beta}_1 = \frac{\text{Cov}(X, Y)}{\text{Var}[X]} \quad \hat{\beta}_2 = \mathbb{E}[Y] - \hat{\beta}_1 \mathbb{E}[X] \quad (2.22)$$

On donne alors le code suivant, permettant, pour un entier a donné, d'effectuer la régression linéaire 2.22 sur les a premiers points du variogramme expérimental, puis de calculer la droite de régression horizontale (qui correspond simplement à l'ordonnée moyenne) des $(n - a)$ points restants. Le code

retourne (en plus des paramètres ajustés), un indice de la qualité de l'ajustement, le RMSE, calculé par la racine carrée de la moyenne des carrés des écarts entre le modèle et les points de données.

Script 7. Calcul des paramètres d'un variogramme [code4.r]

```
a = 7 # Test d'une coupure possible

# -----
# Estimation
# -----
slope = cov(H[1:a],G[1:a])/var(H[1:a])
intersept = mean(G[1:a])-slope*mean(H[1:a])

model1 = intersept + slope*H

palier = mean(G[(a+1):n])
model2 = rep(palier, n)

erreur = c(model1[1:a]-G[1:a], model2[(a+1):n]-G[(a+1):n])
rmse = sqrt(mean(erreur*erreur))

# Interpolation linéaire de la portée
id = min(which(model1 > palier))
weight = (palier-model1[id-1])/(model1[id]-model1[id-1])
scope = H[id-1]*(1-weight) + H[id]*weight

slope = (palier-intersept)/scope

# -----
# Sorties graphique
# -----
plot(H, G, pch=16)
lines(H, model1, lty=2)
lines(H, model2, lty=2)
abline(v=scope, lty=2)

# -----
# Sorties consoles
# -----
cat("-----\r\n")
cat(paste("Portee:", round(scope,2), "\r\n"))
cat(paste("Pente:", round(slope,3), "\r\n"))
cat(paste("Pepite:", round(intersept,2), "\r\n"))
cat(paste("Palier:", round(palier,2), "\r\n"))
cat("-----\r\n")
cat(paste("RMSE:", round(rmse,2), "\r\n"))
cat("-----\r\n")
```

On peut alors lancer le code ci-dessus pour toutes les valeurs de coupure possibles (nommément pour a allant de 2 à 10) et ne conserver que celle qui donne le meilleur ajustement (*i.e.* le RMSE le plus faible). La meilleure estimation est obtenue pour la coupure $a = 7$ (soit entre le 7eme et le 8eme

point), avec une erreur RMSE de 0.77 m^2 . Les paramètres estimés sont :

$$\text{Portee} = 421.73 \text{ m} \quad \text{Pente} = 0.085 / \text{m} \quad \text{Pepite} = 5.43 \text{ m}^2 \quad \text{Palier} = 41.12 \text{ m}^2$$

ce qui correspond à peu de chose près aux paramètres estimés à la main. La valeur du palier (41.12 m^2) nous donne l'amplitude verticale caractéristique de variation du champ : $\sigma_Z = \sqrt{41.12}$, soit de l'ordre de 6 à 7 m. Le bruit blanc additif a une amplitude caractéristique $\sigma = \sqrt{5.43}$, soit environ 2 m. La taille horizontale typique des éléments structurants du relief (colline, vallées...) serait d'environ 400 m.

Nous concluons cette activité avec une représentation pédagogique du variogramme, inspirée de l'excellent ouvrage de [Gringarten et Deutsch \(2001\)](#).

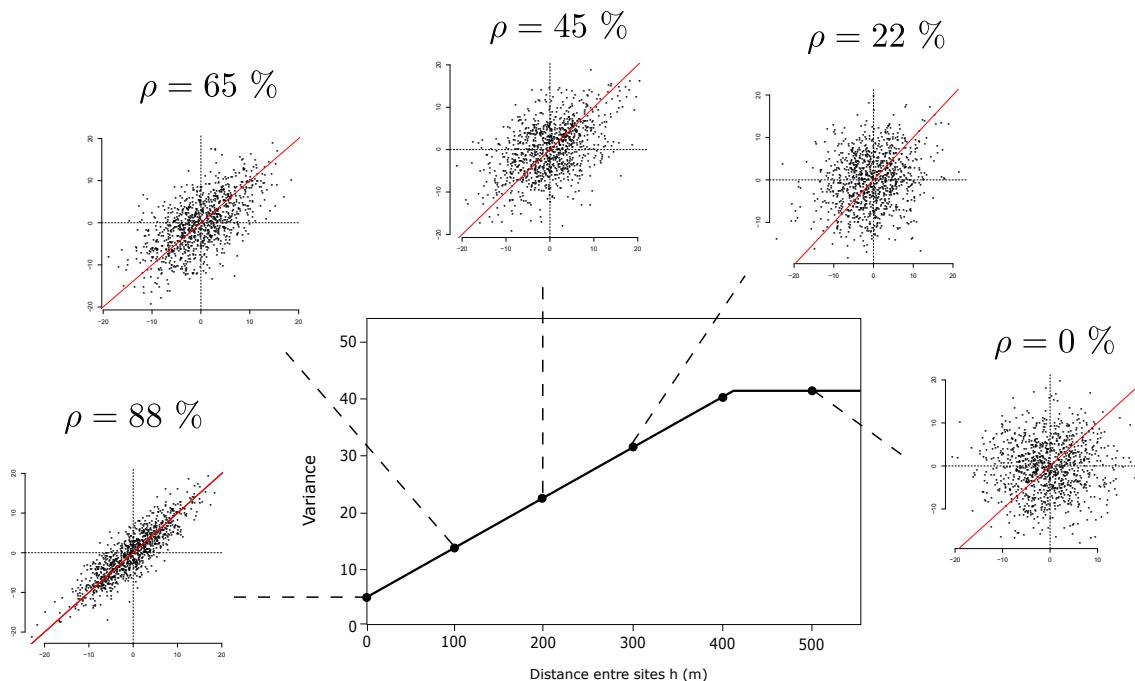


Fig. 2.34. Variogramme du processus, et corrélation linéaire ρ entre deux sites distants de $h = 0, 100, 200, 300$ et 500 m. Remarquons qu'à cause de l'effet de pépite, la corrélation n'est jamais parfaite, y compris pour deux sites infinitésimalement voisins ($h \approx 0$).

Le lecteur pourra vérifier les valeurs de corrélation, en utilisant la définition 11 et le théorème 26 :

$$\rho(h) = \frac{\text{Cov}(Z(x), Z(x + h))}{\sigma(Z(x))\sigma(Z(x + h))} = \frac{C(h)}{C(0)} = \frac{\gamma(\infty) - \gamma(h)}{\gamma(\infty)}$$

Par exemple, pour $h = 200$ m, on calcule à l'aide des paramètres estimés : $\gamma(200) = 0.085 \times 200 + 5.43 = 22.43$. Par ailleurs, l'asymptote du variogramme est $\gamma(\infty) = 41.12$, d'où la corrélation liant des observations effectuées en des sites séparés de 200 m :

$$\rho(200) = \frac{41.12 - 22.43}{41.12} \approx 45 \%$$

2.4.5 Pour aller plus loin

Anisotropie

Si le processus stochastique est anisotrope, son variogramme ne peut plus être décrit par une fonction d'une seule variable, puisqu'il dépend à présent à la fois de la distance h et de l'angle θ séparant les sites. L'estimation du variogramme expérimental se fait d'une manière analogue à 2.9 mais en ajoutant en plus une discrétisation sur l'angle θ . Formellement : disposant d'un semis de points d'observations $z(x_1), z(x_2), \dots, z(x_n)$ d'un phénomène **anisotrope** Z , on peut estimer le variogramme de Z par :

$$\hat{\gamma}(h, \theta) = \frac{1}{2|N(h, \theta)|} \sum_{(i,j) \in N(h, \theta)} (z(x_i) - z(x_j))^2 \quad (2.23)$$

où $N(h, \theta)$ désigne l'ensemble des couples d'indices (i, j) tels que les sites x_i et x_j sont séparés d'une distance $h \pm \frac{dh}{2}$ suivant une direction $\theta \pm \frac{d\theta}{2}$ et où $|N(h, \theta)|$ désigne le nombre de couples contenus dans $N(h, \theta)$. En général 2 directions (0° et 90°) ou 4 directions ($0^\circ, 45^\circ, 90^\circ$ et 135°) sont suffisantes¹⁵.

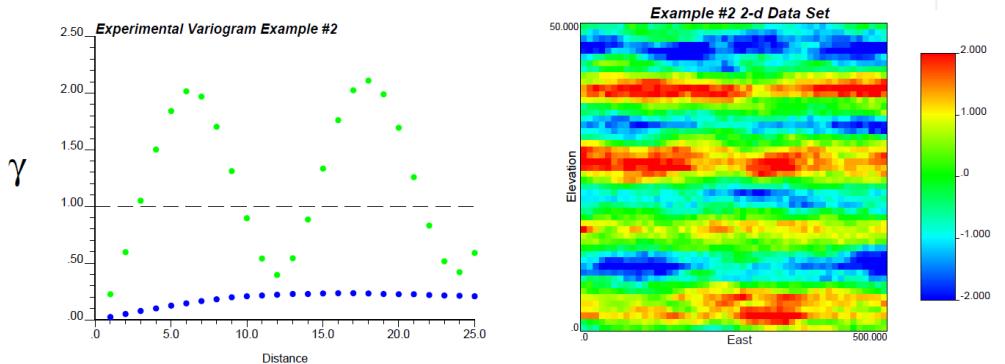


Fig. 2.35. Anisotropie d'un processus stochastique : le variogramme expérimental est calculé dans deux directions orthogonales : verticale (variogramme à effet de trou en vert) et horizontale (variogramme gaussien ou cubique en bleu). Source : [Pyrcz et Deutsch \(2003\)](#).

Les variogrammes anisotropiques permettent de modéliser des processus géologiques complexes (strates, failles, couches sédimentaires, etc.) mais peuvent aussi être utilisés pour gérer des processus spatio-temporels (la coordonnée *temps* étant en général sujette à un variogramme différent des coordonnées *espace*) ou encore les espaces non-euclidiens (par exemple localement en coordonnées géographiques pour prendre en compte la différence de métrique en latitude et en longitude). Parmi les différents types d'anisotropie, on distingue deux mode caractéristiques :

- **L'anisotropie géométrique**, dont les variogrammes sont de palier identiques mais de portées différentes suivant les directions, et qui peut être facilement corrigée par un changement de repère. Dans ce mode d'anisotropie, la variabilité préférentielle du processus dans certaines directions est souvent due à un mauvais choix de système de coordonnées.
- **L'anisotropie zonale**, aux variogrammes de portées identiques mais de palier différents, qui caractérise une anisotropie plus intrinsèquement liée au processus, et donc plus difficile à modéliser.

Le lecteur pourra trouver plus d'informations sur ce sujet dans ([Allard, 2012](#)).

¹⁵Les angles sont donnés à une rotation globale près, dépendant de l'orientation du repère.

De la validité d'un modèle variographique

Comment savoir si une fonction γ (nulle à l'origine) est un modèle de variogramme licite ?

En théorie, il suffit de s'assurer que γ est **conditionnellement définie-négative**, i.e. pour tout n -uplet de points du domaine : $(x_1, x_2, \dots, x_n) \in \mathcal{D}$ et pour tout n -uplet de poids $(\lambda_1, \lambda_2, \dots, \lambda_n) \in \mathbb{R}^n$:

$$\sum_{i=1}^n \lambda_i = 0 \Rightarrow \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \gamma(x_i - x_j) \leq 0 \quad (2.24)$$

Exemple 2.5. Reprenons l'exemple 2.4 de la fonction de Heaviside (fonction échelon unité), dont nous avons montré empiriquement qu'elle ne pouvait pas être un variogramme. On montre facilement que $\tilde{\gamma}$ ne vérifie pas la contrainte de défini-négativité conditionnelle. En effet, en reprenant les trois mêmes points x_1, x_2 et x_3 , avec $\|x_1 - x_3\| = \|x_2 - x_3\| < \xi$ et $\|x_1 - x_2\| \geq \xi$, on peut considérer le système de poids : $\lambda_1 = \lambda_2 = -1$ et $\lambda_3 = 2$. On a bien $\lambda_1 + \lambda_2 + \lambda_3 = 0$, et pour autant :

$$\sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \tilde{\gamma}(x_i - x_j) = \lambda_1 \lambda_2 \tilde{\gamma}(x_1 - x_2) = \Gamma > 0$$

donc la fonction $\tilde{\gamma}$ n'est pas CDN, et ne peut donc pas être un modèle de variogramme.

Si la contrainte 2.24 est parfois facile à réfuter, elle est en général très difficile à vérifier, et on pourra alors trouver un salut dans les deux théorèmes suivants. Tout d'abord, dans le cas où γ est bornée : si γ est un modèle de variogramme licite, alors on sait que la covariance C doit exister et la relation fondamentale 26 nous permet d'obtenir son expression. Il suffit alors de vérifier que C est bien une fonction de covariance licite, et pour cela, on peut utiliser le puissant théorème de Bochner :

Théorème 33 (Théorème de Bochner)

Une fonction $C : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ est la covariance d'un processus stationnaire au sens large si, et seulement si, elle est la transformée de Fourier d'une mesure positive finie sur \mathbb{R} :

$$C(h) = \int_{\mathbb{R}} e^{i\omega h} F(d\omega)$$

Autrement dit, dans la plupart des cas, pour savoir si une fonction γ quelconque bornée est un modèle de variogramme licite, il suffit :

- d'évaluer la fonction covariance candidate associée : $C(h) = \lim_{t \rightarrow \infty} \gamma(t) - \gamma(h)$
- de calculer la transformée de Fourier μ de C (Cottet, 2015) : $\mu = \mathcal{F}[C]$
- et enfin, de vérifier que μ est positive sur \mathbb{R} .

Remarque : pour calculer la transformation de Fourier sur l'intervalle réel, on symétrise la fonction de covariance candidate : $C(-h) = C(h)$.

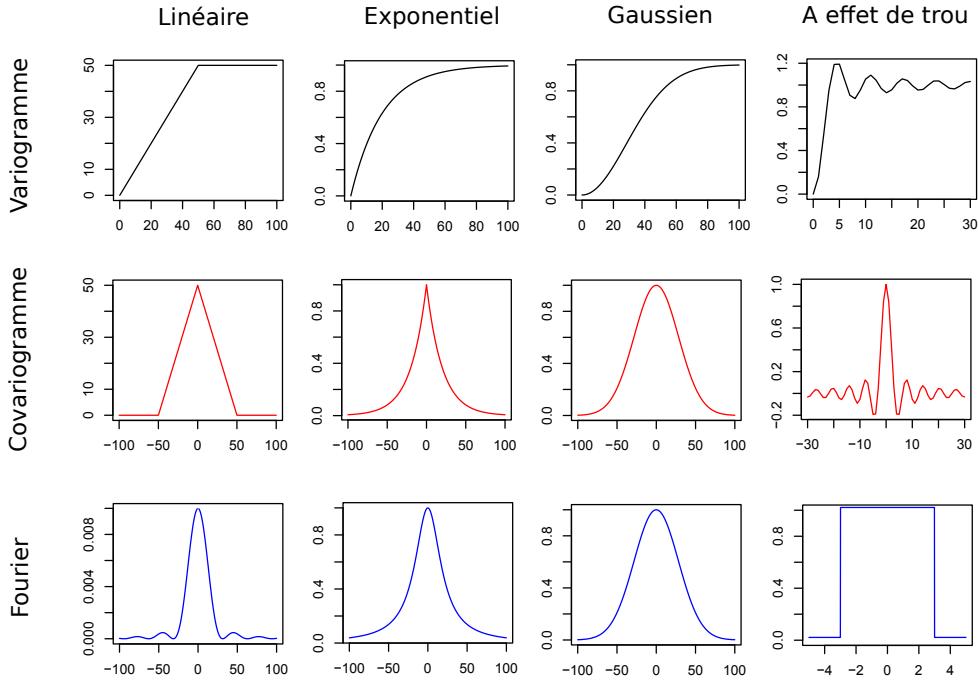


Fig. 2.36. Pour chaque fonction γ (ligne du haut) dont on souhaite vérifier la validité en tant que modèle de variogramme, on calcule le covariogramme candidat associé (ligne du milieu), et on vérifie que sa transformée de Fourier (ligne du bas) est une fonction positive.

On donne en figure 2.36 quatre exemples de vérifications.

Remarque : si C est bien une covariance, F est sa **densité spectrale de puissance**, c'est-à-dire la répartition probabiliste des puissances des signaux générés par le processus en fonction de la fréquence. Elle fournit donc de précieuses informations sur le processus.

Exemple 2.6. À nouveau, en reprenant l'exemple 2.4, si $\tilde{\gamma}$ était un variogramme licite, alors sa covariance associée serait (à un facteur d'échelle Γ près) la fonction porte Π_ξ , définie par :

$$\Pi_\xi(h) = \begin{cases} 1 & \text{si } |h| \leq \xi \\ 0 & \text{sinon.} \end{cases} \quad (2.25)$$

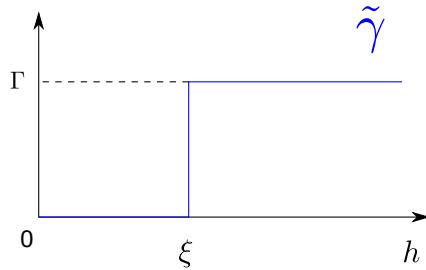


Fig. 2.37. La fonction échelon unité $\tilde{\gamma}$ n'est pas un modèle licite de variogramme.

Remarquons qu'il s'agit de la fonction en bas à droite de la figure 2.36. Sa transformation de Fourier est un sinus cardinal (par symétrie de la transformée de Fourier, on retrouve naturellement la fonction

au milieu à droite sur cette même figure), qui n'est pas une mesure positive. Donc la fonction porte n'est pas un modèle licite de fonction de covariance, et nous avons à présent montré rigoureusement que $\tilde{\gamma}$ ne peut pas être un variogramme.

Et pour les fonctions non bornées ?

Nous l'avons dit précédemment, un variogramme non-borné, est la caractérisation d'un processus stochastique stationnaire intrinsèque (et non stationnaire au sens large). Sa covariance n'existe pas et on ne peut donc pas utiliser la méthode ci-dessus. Nous pouvons alors avoir recours à un second théorème, qui généralise en quelques sortes le théorème de Bochner ([Allard, 2012](#)).

Théorème 34 (Théorème : validité d'un modèle de variogramme)

Une fonction $\gamma : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ (éventuellement non-bornée), telle que $\gamma(0) = 0$, est le variogramme d'un processus stochastique stationnaire intrinsèque si, et seulement si la fonction $\psi(h) = e^{-t\gamma(h)}$ est une fonction de covariance quelque soit $t > 0$.

Exemple 2.7. Montrons que la fonction puissance [2.13](#) : $\gamma : h \mapsto h^\theta$ est bien le variogramme d'un processus stochastique intrinsèque. On a bien $\gamma(0) = 0$. Par ailleurs, formons la fonction :

$$\psi(h) = \exp(-th^\theta)$$

Traitons alors différents cas en fonction de la valeur de θ :

- $\theta = 1$: $\psi(h)$ est une fonction exponentielle décroissante (c'est la fonction de la deuxième colonne de la figure [2.36](#)). Sa transformée de Fourier est une fonction lorentzienne, donc positive sur \mathbb{R} . Le variogramme linéaire est bien un modèle licite de variogramme.
- $\theta = 2$: $\psi(h)$ est une fonction gaussienne (c'est la fonction de la troisième colonne de la figure [2.36](#)). Sa transformée de Fourier est également une gaussienne, donc positive sur \mathbb{R} . Le variogramme quadratique est bien un modèle licite de variogramme.

Plus généralement, on montre que le modèle de variogramme en puissance est licite pour tout $\theta \in]0; 2]$. [Allard \(2012\)](#) propose une méthode ne nécessitant pas le calcul de la transformation de Fourier, pour montrer que la puissance θ ne peut pas être supérieure à 2. L'idée centrale du développement est de procéder de manière similaire à l'exemple [2.4](#) de la fonction échelon unité, en établissant une inégalité triangulaire sur le variogramme. Formellement, en considérant h et h' deux vecteurs du domaine :

$$\begin{aligned}\gamma(h + h') &= \frac{1}{2}\mathbb{E}\left[\left(Z_x - Z_{x+h+h'}\right)^2\right] = \frac{1}{2}\mathbb{E}\left[\left(Z_x - Z_{x+h} + Z_{x+h} - Z_{x+h+h'}\right)^2\right] \\ &= \frac{1}{2}\mathbb{E}\left[\left(Z_x - Z_{x+h}\right)^2 + \left(Z_{x+h} - Z_{x+h+h'}\right)^2 + 2\left(Z_x - Z_{x+h}\right)\left(Z_{x+h} - Z_{x+h+h'}\right)\right] \\ &= \gamma(h) + \gamma(h') + \mathbb{E}\left[\left(Z_x - Z_{x+h}\right)\left(Z_{x+h} - Z_{x+h+h'}\right)\right]\end{aligned}$$

Or, d'après l'inégalité de Cauchy-Schwarz on a :

$$\begin{aligned}\mathbb{E}[(Z_x - Z_{x+h})(Z_{x+h} - Z_{x+h+h'})] &= \text{Cov}(Z_x - Z_{x+h}, Z_{x+h} - Z_{x+h+h'}) \\ &\leq (\text{Var}[Z_x - Z_{x+h}] \times \text{Var}[Z_{x+h} - Z_{x+h+h'}])^{\frac{1}{2}}\end{aligned}\tag{2.26}$$

D'où la contrainte d'**inégalité triangulaire** sur le variogramme :

$$\gamma(h + h') \leq \gamma(h) + \gamma(h') + 2(\gamma(h)\gamma(h'))^{\frac{1}{2}}\tag{2.27}$$

En particulier, la fonction échelon unité (2.4) ne respecte pas cette contrainte puisque :

$$\Gamma = \tilde{\gamma}(\xi) = \tilde{\gamma}\left(\frac{\xi}{2} + \frac{\xi}{2}\right) > \tilde{\gamma}\left(\frac{\xi}{2}\right) + \tilde{\gamma}\left(\frac{\xi}{2}\right) + 2\sqrt{\tilde{\gamma}\left(\frac{\xi}{2}\right)\tilde{\gamma}\left(\frac{\xi}{2}\right)} = 0$$

De manière plus générale, l'inégalité 2.27 montre que le variogramme ne peut pas croître plus rapidement qu'une parabole : $\gamma(h) = \mathcal{O}(h^2)$. En effet, en considérant $h = h'$ on obtient : $\gamma(2h) \leq 4\gamma(h)$, et donc, si $\gamma(h) = h^\theta$, on a (pour $h > 0$) : $2^\theta h^\theta \leq 4h^\theta \Leftrightarrow \theta \leq 2$.

Il est intéressant de constater que dans le cas limite $\theta = 2$, l'inéquation 2.26 nous montre que les accroissements du processus sont complètement corrélés, et on aboutit à un cas dégénéré où les réalisations sont rigoureusement des plans inclinés, de pentes aléatoires. Par exemple, on peut considérer le processus uni-dimensionnel $X(t) = At$, avec $A \sim \mathcal{N}(0, 1)$. On vérifie alors facilement que X est bien stationnaire intrinsèque : $\mathbb{E}[X(t+h) - X(t)] = \mathbb{E}[Ah] = 0$ et $\text{Var}[X(t+h) - X(t)] = \text{Var}[Ah] = h^2$, qui est donc bien un variogramme quadratique. Le cas $\theta = 2$ est un modèle licite de variogramme, mais d'aucune utilité en pratique. Lors de utilisation d'un modèle en puissance, on veillera à bien contraindre l'ajustement paramétrique de θ dans l'intervalle ouvert $]0; 2[$.

Exemple 2.8. La linéarité de la transformation de Fourier nous dit que si C_1 et C_2 sont deux fonctions de covariance, alors leur somme $C_1 + C_2$ en est une aussi. L'implication est moins triviale pour les variogrammes, mais les théorème 33 et 34 nous permettent d'arriver à nos fins.

Soient γ_1 et γ_2 deux variogrammes licites. Considérons la fonction $\gamma : h \mapsto \gamma_1(h) + \gamma_2(h)$. La condition $\gamma(0) = 0$ est trivialement vérifiée puisque $\gamma_1(0) = \gamma_2(0) = 0$. Par ailleurs, formons la fonction : $\psi(h) = \exp(-t\gamma(h))$. On a alors : $\psi(h) = \exp(-t(\gamma_1(h) + \gamma_2(h))) = \exp(-t\gamma_1(h))\exp(-t\gamma_2(h)) = \psi_1(h)\psi_2(h)$, où $\psi_i(h)$ est, d'après le théorème 34, une fonction de covariance. Le théorème 33 nous dit alors qu'il existe deux mesures positives μ_1 et μ_2 dont ψ_1 et ψ_2 sont respectivement les transformées de Fourier. Par analogie de la multiplication et du produit de convolution (dit théorème de Plancherel), la fonction ψ en tant que produit de transformée de Fourier de fonctions, est aussi la transformée de Fourier de leur produit de convolution) :

$$\psi = \psi_1 \times \psi_2 = \mathcal{F}[\mu_1] \times \mathcal{F}[\mu_2] = \mathcal{F}[\mu_1 * \mu_2]$$

Par ailleurs, μ_1 et μ_2 étant des mesures positives, leur produit de convolution $\mu = \mu_1 * \mu_2$ l'est aussi et ψ , en tant que transformée de Fourier d'une mesure positive est une fonction de covariance. À nouveau, en utilisant le théorème 34, la fonction γ est donc un variogramme licite. On pourra généraliser ce raisonnement, par récurrence, pour une somme de n variogrammes. Nous avons donc montré que le variogramme en gigogne 2.28 est valide dès lors que ses n composantes individuelles le sont également.

2.5 Bilan

Le variogramme, qui indique pour toute distance de séparation h , la variance de la différence entre deux sites distants de h , caractérise (à l'ordre 2) le phénomène à étudier. En particulier, le comportement à l'origine du variogramme traduit la régularité du phénomène (continuité et dérivabilité) tandis que la valeur numérique de sa portée est un indicateur de la taille caractéristique des éléments (*e.g.* vallées, collines) structurant ses réalisations.

On utilise le semis d'observations ponctuelles pour estimer un variogramme expérimental, moyennant quelques hypothèses théoriques sur le processus étudié, à savoir la stationnarité et l'ergodicité, ce à quoi on ajoute souvent par commodité l'isotropie. La quantité d'observations étant toujours insuffisante, on régularise l'estimation avec une bibliothèque de fonctions paramétriques permettant d'assurer que la fonction estimée possède bien le sens physique d'une variance statistique. Le modèle de variogramme est choisi par le praticien en fonction de la structure du variogramme expérimental, ainsi que des connaissances disponibles *a priori* sur le phénomène étudié, tandis que les paramètres sont ajustés automatiquement de sorte à minimiser l'erreur d'estimation.

Tous les calculs subséquents (en particulier ceux liés au krigeage, comme nous allons le voir dans le chapitre 3) sont menés à partir du modèle de variogramme inféré (cohérence avec la théorie, stabilité numérique des matrices, positivité des variances d'estimation...).

Chapter 3

Interpolation par Krigeage

Dans ce chapitre, nous traitons le problème modèle énoncé dans l'introduction : disposant d'un semis d'observations d'altitude z_1, z_2, \dots, z_n en n sites x_1, x_2, \dots, x_n , déterminer une fonction interpolante $z : \mathcal{D} \rightarrow \mathbb{R}$ qui en chaque lieu x du domaine retourne une estimation $z(x)$ de l'altitude, avec la contrainte que la surface z coïncide¹ aux observations au niveau des sites : $\forall i \in \llbracket 1..n \rrbracket, z_i = (x_i)$.

Nous l'avons vu précédemment, le krigeage, comme toutes les méthodes de Géostatistique, aborde le problème d'un point de vue probabiliste, en posant l'hypothèse que la **variable régionalisée** z est **une** réalisation d'un **processus stochastique** Z , dont on suppose qu'il possède les propriétés minimales nécessaires au calcul de son **variogramme** à partir d'un semis d'observations ponctuelles.

Nous supposons disposer d'un modèle de variogramme γ dont les paramètres ont été estimés avec les techniques présentées dans le chapitre 2. Le krigeage utilise alors les informations fournies par γ pour construire un estimateur de variance minimale de z .

Nous montrons en premier lieu et de manière empirique, comment la théorie des processus stochastiques peut être employée pour traiter le problème d'interpolation. Nous présentons ensuite les équations de krigeage et illustrons la méthode à l'aide de plusieurs cas concrets. Nous concluons enfin le chapitre en présentant quelques variantes et dérivées de la technique.

3.1 Introduction

Pour montrer empiriquement comment les processus stochastiques (et la connaissance fine des variogrammes associés) peut permettre de résoudre un problème d'interpolation, nous proposons ci-dessous un activité sur machine. Notons qu'elle requiert un peu plus d'aisance dans le langage R que les activités précédentes, mais apporte une perspective intéressante sur les fondements de ce chapitre.

Activité VI. Une première approche de l'interpolation

Pour en faciliter la prise en main de cette activité, nous progressons pas à pas avec une correction après chaque question. Par ailleurs, pour rendre le code reproductible (et ainsi permettre au lecteur d'obtenir exactement les mêmes résultats que ceux présentés en guise de correction) nous suggérons d'utiliser une graine informatique : `set.seed(123)`. Cette commande, à insérer tout en haut du code, permet de fixer l'initialisation du générateur de nombres aléatoires. En conséquence, pour une valeur de *graine* donnée (ici 123 par exemple), les résultats obtenus seront identiques sur plusieurs exécutions du code (et ne devrait en principe pas dépendre de la machine sur laquelle le code est exécuté). A la fin de cette activité, il sera possible de changer la valeur de la graine, ou même de supprimer

¹Nous relaxerons quelque peu cette contrainte par la suite pour modéliser le cas où les observations sont imprécises. La contrainte stipulera alors que la variable $z_i - z(x_i)$ possède un écart-type égal à celui de l'erreur entâchant les observations.

complètement cette ligne, pour pouvoir observer différentes réalisations aléatoires.

Dans un premier temps, nous allons programmer un processus stochastique légèrement plus sophistiqué que la marche aléatoire 2.1 de l'activité I. Pour ce faire, nous allons avoir besoin d'une fonction (qui malheureusement n'existe pas dans la bibliothèque standard de R) permettant de décaler les éléments d'un vecteur d'un nombre donné de rangs vers la gauche ou vers la droite :

```
shift = function(x, lag) {
  n = length(x)
  xnew = rep(0, n)
  if (lag < 0) {
    xnew[1:(n-abs(lag))] = x[(abs(lag)+1):n]
  } else if (lag > 0) {
    xnew[(lag+1):n] = x[1:(n-lag)]
  } else {
    xnew = x
  }
  return(xnew)
}
```

On pourra trouver plus de détails sur le fonctionnement interne du module `shift` sur le site internet de [Richards \(2016\)](#). Pour l'heure, il est suffisant de savoir que, pour un vecteur $X = \text{c}(1, 2, 3, 4, 5, 6)$ par exemple, l'instruction `shift(X, 2)` va retourner le vecteur X décalé de 2 éléments vers la droite : $X = [0, 0, 1, 2, 3, 4]$. De même lorsque le paramètre est négatif, `shift(X, -3)` par exemple, va retourner le vecteur X décalé de 3 éléments vers la gauche $X = [4, 5, 6, 0, 0, 0]$.

Cette fonction va nous permettre de construire des versions translatées d'un même signal. Par exemple, plaçons-nous sur l'intervalle $[0; 1000]$ et considérons une fonction gaussienne (ici notée h) centrée au milieu de l'intervalle, et d'écart-type (l'échelle) 100 :

```
x = 0:999
h = exp(-((500-x)/100)**2)
```

Avec la fonction `shift`, on peut construire des réplications translatées de h :

```
plot(x,h,type="l") # On trace la gaussienne de base
lines(x,shift(h,150),col="blue",lty=2) # Translation vers la droite (en bleu)
lines(x,shift(h,220),col="blue",lty=2) # Translation vers la droite (en bleu)
lines(x,shift(h,-210),col="blue",lty=2) # Translation vers la gauche (en bleu)
```

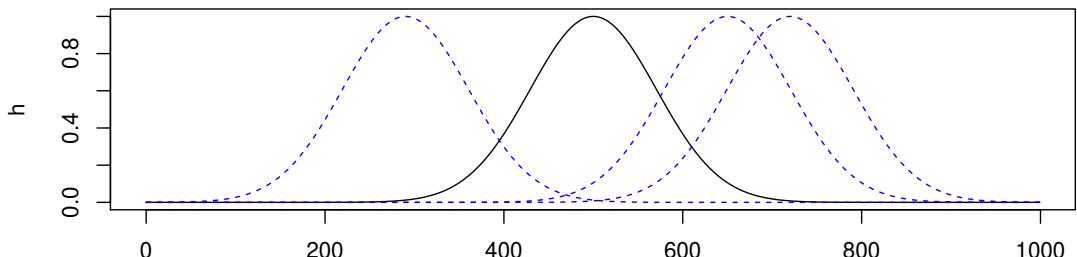


Fig. 3.1. Tracé de la gaussienne de base (en noir) et de 3 réplications translatées (en pointillés).

En général, on appelle **noyau** une fonction symétrique destinée à être répliquée, translatée et changée d'échelle. Ici, h est donc un noyau gaussien. On peut alors construire un processus stochastique en sommant un certain nombre N de réplications de h translatées et amplifiées aléatoirement :

$$Z(x) = \sum_{i=1}^N x_i h(x - \tau_i) \quad (3.1)$$

où x_i est distribuée suivant la loi normale standard, et τ_i est distribuée suivant la loi uniforme sur la plage des translations possibles [-500; 500].

Q1. Programmer le processus Z sous forme d'une fonction `generate` prenant en entrée un noyau h de taille quelconque (dans notre cadre nous travaillerons toujours sur des signaux de taille 1000), et générant une réalisation aléatoire.

```
generate = function(h){
  X = rep(0,length(h))
  for (i in 1:100){
    d = sample((-length(h):length(h))/2, 1)
    X = X + rnorm(1,0,1)*shift(h,d)
  }
  return(X)
}
```

La fonction `generate` somme 100 réplications de la gaussienne, translatées (fonction `sample`) et amplifiées (fonction `rnorm`) aléatoirement.

Q2. Sur un même graphique, tracer 5 réalisations du processus Z . On pourra réitérer l'expérience en changeant l'échelle du noyau h , par exemple à 50 ou 200.

```
par(mfrow=c(1,3)) # Pour tracer trois graphiques

h = exp(-((500-x)/200)**2) # Definition du noyau gaussien 1
plot(c(), xlim=c(0,1000), ylim=c(-10,10)) # Initialisation
for (i in 1:5){lines(generate(h), type="l")}

# Representation des realisations

h = exp(-((500-x)/100)**2) # Definition du noyau gaussien 2
plot(c(), xlim=c(0,1000), ylim=c(-10,10)) # Initialisation
for (i in 1:5){lines(generate(h), type="l")}

# Representation des realisations

h = exp(-((500-x)/50)**2) # Definition du noyau gaussien 3
plot(c(), xlim=c(0,1000), ylim=c(-10,10)) # Initialisation
for (i in 1:5){lines(generate(h), type="l")}

# Representation des realisations
```

Par souci de clarté, on ne représente ci-dessous en figure 3.2 qu'une seule réalisation par graphique. Remarquons que la portée d'autocorrélation des processus est, comme on pouvait s'y attendre, d'autant plus grande que l'échelle du noyau l'est aussi.

Q3. Pour la suite, et jusqu'à la question Q5 nous n'utiliserons que la version d'échelle 100 du noyau.

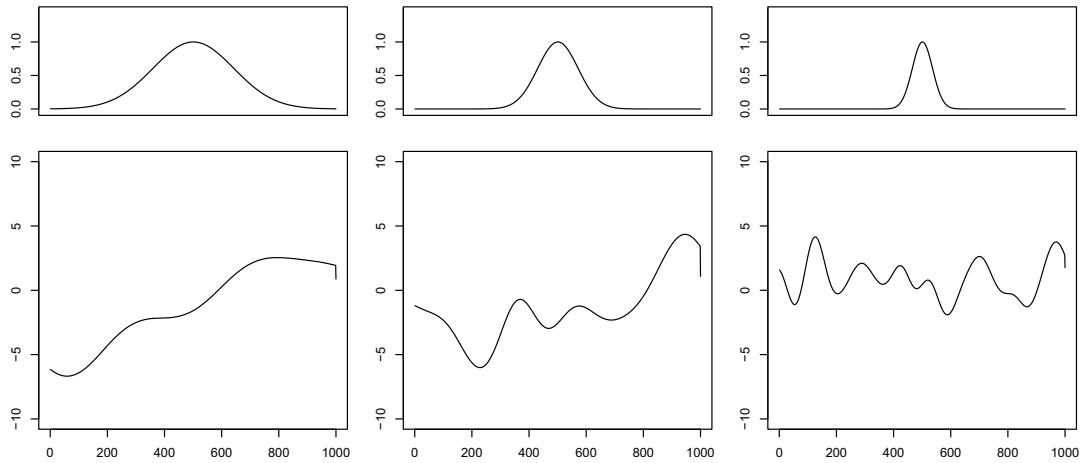


Fig. 3.2. Trois réalisations d'un signal aléatoire (en bas) et les noyaux de base associés (en haut). Échelles des noyaux : 200 (à gauche), 100 (au centre) et 50 (à droite).

Nous cherchons à présent à estimer une fonction inconnue z sur l'intervalle $[0, 1000]$. Pour ce faire, nous disposons de trois observations : $z(300) = 1$, $z(500) = 5$ et $z(800) = -1$, et le problème est donc réduit à un problème d'interpolation. Nous allons montrer ici qu'il est possible de résoudre ce problème avec des simulations aléatoires. Commençons par tracer ces données dans un nouveau graphique :

```
SITES = c(300,500,800)
OBS = c(1,5,-3)
plot(SITES, OBS, pch=16, ylim=c(-10,10))
```

En utilisant directement la fonction `generate`, créer 5000 réalisations du processus aléatoire, et ne représenter dans le graphique que celles qui satisfont les contraintes d'interpolation à ε près, *i.e.* uniquement les réalisations z telles que les valeurs prises en chacun des 3 points observés (vecteur `SITES`) soient à une distance verticale inférieure à ε des valeurs observées (vecteur `OBS`) : $|z(x_i) - z_i| \leq \varepsilon$. On utilisera une valeur seuil $\varepsilon = 1$.

```
plot(SITES, OBS, pch=16, ylim=c(-10,10))

epsilon = 1
for (i in 1:5000){
  if (i %% 1000== 0){print(i)}
  X = generate(h)
  ok = TRUE
  for (j in 1:length(OBS)){
    ok = ok && (abs(X[SITES[j]]-OBS[j])<epsilon)
  }
  if (ok){
    lines(X, type="l")
  }
}
```

Quelques explications sur le code ci-dessus : pour chacune des 5000 répétitions, on génère une réalisations X . On déclare alors la variable `ok` permettant de s'assurer que la réalisation courante X vérifie les contraintes d'interpolation. Pour chaque point de donnée j , on teste la contrainte. Si cette dernière n'est pas vérifiée sur l'un (ou plus) des points du vecteur `SITES`, alors on ne représente pas la réalisation. Notons que l'instruction `if (i %% 1000== 0){print(i)}` permet de suivre la trace du calcul en cours, qui peut prendre un certain temps (en fonction des ressources de la machine utilisée).

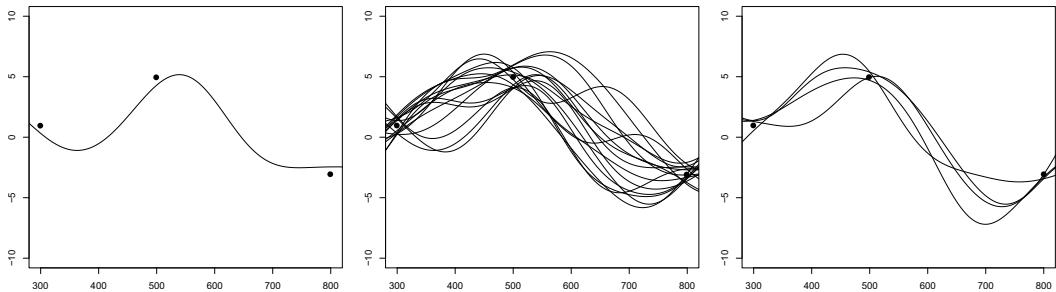


Fig. 3.3. À gauche : une réalisation respectant la contrainte d’interpolation à ε près. Au centre : 16 réalisations (sur 5000 tirages) respectant la contrainte d’interpolation. À droite : 4 réalisations (sur 5000 tirages) respectant la contrainte d’interpolation pour une tolérance plus stricte ($\varepsilon = 0.5$).

Si on le souhaite, on pourra être plus exigeant sur la tolérance, en choisissant par exemple $\varepsilon = 0.5$, mais attention, le nombre de solutions chute très rapidement à mesure que la tolérance est plus stricte, comme illustré sur la figure 3.3.

Q4. Toujours pour un nombre total de 5000 réalisations, représenter la courbe moyenne de l’ensemble des réalisations respectant la contrainte d’interpolation. Formellement, si z_i désigne la i -eme réalisation valide, on calculera la courbe moyenne μ par la moyenne point-à-point des z_i :

$$\mu(x) = \frac{1}{n} \sum_{i=1}^n z_i(x) \quad (3.2)$$

où n désigne le nombre de réalisations respectant la contrainte d’interpolation. Dans la littérature de référence, la courbe μ est appellée *moyenne d’ensemble*. On peut également calculer les courbes d’incertitude :

$$L(x) = \mu(x) - \sigma(x) \quad (3.3)$$

$$U(x) = \mu(x) + \sigma(x) \quad (3.4)$$

où la fonction σ désigne l’écart-type point-à-point des z_i :

$$\sigma^2(x) = \frac{1}{N} \sum_{i=1}^n (z_i(x) - \mu(x))^2 \quad (3.5)$$

On complète le code précédent en incrémentant à chaque itération de boucle trois variables fonctionnelles : la moyenne des courbes (INTERPOLATION), la moyenne des carrés des courbes INTERPOLATION2 et un décompte des réalisations valides (COUNT).

```
plot(SITES, OBS, pch=16, ylim=c(-10,10))

INTERPOLATION = rep(0,1000)
INTERPOLATION2= rep(0,1000)
COUNT = 0

epsilon = 1
for (i in 1:5000){
  if (i %% 1000== 0){print(i)}
  X = generate(h)
```

```

ok = TRUE
for (j in 1:length(OBS)){
  ok = ok && (abs(X[SITES[j]]-OBS[j])<epsilon)
}
if (ok){
  INTERPOLATION = INTERPOLATION + X
  INTERPOLATION2= INTERPOLATION2+ X**2
  COUNT = COUNT + 1
  lines(X, type="l")
}
}

INTERPOLATION = INTERPOLATION/COUNT
INTERPOLATION2= INTERPOLATION2/COUNT
SIGMA = sqrt(INTERPOLATION2- INTERPOLATION**2)

lines(INTERPOLATION, col="red", lwd=2)
lines(INTERPOLATION+SIGMA, col="blue", lwd=2)
lines(INTERPOLATION-SIGMA, col="blue", lwd=2)

```

Remarquons que le calcul de la courbe d'écart-type a été effectué à partir de la formulation 1.13 de la variance, ne nécessitant qu'une seule itération de boucle sur les données : $\sigma^2 = \mathbb{E}[z^2] - \mathbb{E}[z]^2$. On obtient alors le résultat de la figure suivante :

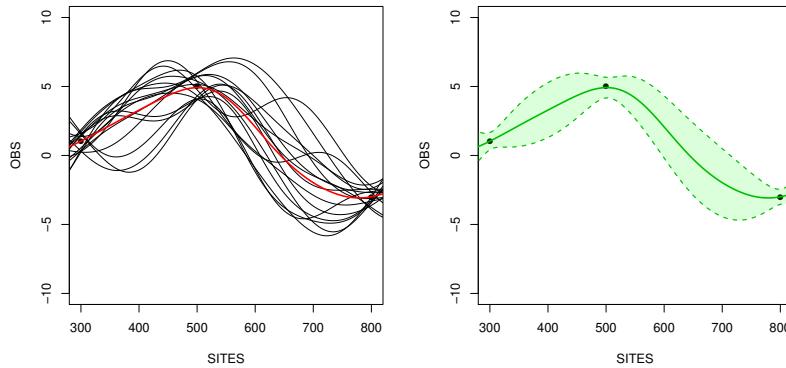


Fig. 3.4. À gauche : les 16 réalisations respectant la contrainte d'interpolation et la moyenne d'ensemble de ces réalisations (en rouge). À droite : le résultat de la procédure d'interpolation : la courbe interpolée (en trait plein vert) et les bandes de confiance (en pointillés). La vraie courbe (inconnue) se situe probablement dans la zone verte.

Q5. Lorsque le code la question Q4 est opérationnel, réitérer la procédure d'interpolation pour le jeu de données suivant (contenant un point supplémentaire) :

```

SITES = c(300,450,500,800)
OBS = c(1,1,5,-3)

```

et en générant un total de 100 000 réalisations (l'exécution du code peut alors prendre quelques minutes). Relancer la procédure pour plusieurs échelles de noyaux, par exemple 100, 50 et 25.

Pour ce faire, il suffit de changer la définition du noyau, de manière similaire au code ayant permis l'obtention de la figure 3.2 :

```

h = exp(-((500-x)/100)**2)      # Definition du noyau gaussien 1
h = exp(-((500-x)/50)**2)        # Definition du noyau gaussien 2
h = exp(-((500-x)/25)**2)        # Definition du noyau gaussien 3

```

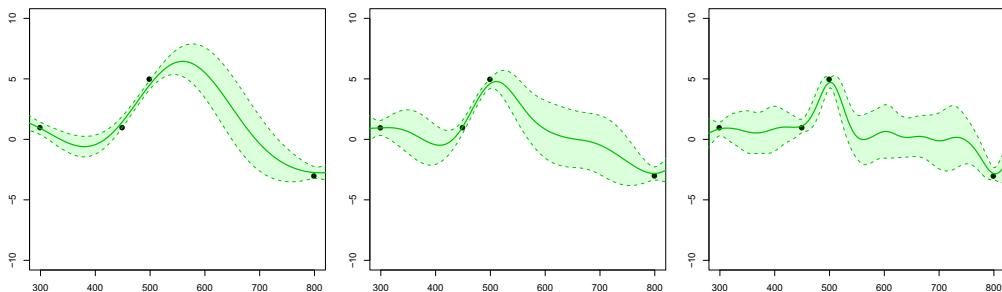


Fig. 3.5. Résultat de la procédure d’interpolation pour 3 noyaux d’échelles différentes : 100 (à gauche), 50 (au centre) et 25 (à droite).

Remarque. Pour un noyau d’échelle 200, le jeu de points d’interpolation est trop contraignant pour pouvoir être facilement interpolé par un processus si régulier. Le programme ne trouve donc pas de candidat en un temps raisonnable, et on ne peut pas calculer de courbe moyenne. Nous atteignons malheureusement là les limites de la méthode par simulation.

Bilan :

Dans cette activité, nous avons montré comment la simulation d’un processus pouvait être utilisée pour interpoler un semis de points d’observations. Quelques remarques s’imposent toutefois :

- La solution du problème d’interpolation est une courbe moyenne des réalisations. D’après les propriétés statistiques de la moyenne, nous pouvons dire qu’il s’agit de la courbe qui minimise l’écart quadratique moyen de l’estimation par rapport à la vérité. Nous verrons par la suite que le krigage fonctionne de manière similaire : il propose la solution **minimisant la variance de l’erreur d’estimation**. Le calcul sera cependant effectué algébriquement et ne nécessitera donc pas de simulations.
- Le recours à la simulation permettra toutefois (et moyennant un coût de calcul important) de contourner quelques limitations du krigage. Nous verrons cela plus en détail dans la partie consacrée aux **simulations conditionnelles** dans le chapitre 4.
- La figure 3.5 nous montre l’importance de bien modéliser le **variogramme** du processus stochastique. Ici, pour trois variogrammes de portées différentes (*i.e.* pour trois processus générés par des noyaux d’échelles différentes), nous obtenons trois interpolations sensiblement différentes. Concrètement, employée avec un variogramme de mauvaise qualité (*i.e.* peu fidèle à la réalité), les techniques de Géostatistique de ce chapitre 3 ne donneront pas de résultats sensiblement meilleurs que les techniques d’interpolation non-probabilistes (splines, etc.).
- Le graphique de droite de la fig. 3.5 révèle une limite naturelle de la méthode. Si le semis d’observations est trop lâche par rapport à la **portée du variogramme**, l’interpolation donne des résultats très médiocres entre les sites d’observation. C’est naturel, par exemple en $x = 650$,

les deux sites d'observations les plus proches (500 et 800) sont situés bien au-delà de la portée du variogramme (qui est de l'ordre de 3 fois le facteur d'échelle du noyau gaussien, soit environ 75). L'estimation n'a donc pas d'autre choix que de retourner l'espérance *a priori* du processus, à savoir 0. Nous verrons un phénomène similaire lorsque nous parlerons du krigage simple (3.7.1).

- Le graphique de gauche de la fig. 3.5 nous le montre clairement : l'interpolation **n'est pas convexe**, autrement dit, les valeurs estimées peuvent être au dessus ou en deçà des valeurs extrêmes des observations, ce qui est une propriété souhaitable que ne partagent pas, par exemple, les méthodes plus basiques telles que l'interpolation linéaire ou en inverse des distances.

3.2 Les contraintes du krigage

Le krigage est une technique géostatistique d'estimation locale, initialement mise au point par Danie G. Krige en 1951 et théorisé plus par Georges Matheron (1962). Il permet d'estimer la valeur prise par la variable régionalisée en un site non-échantillonné, en tenant en compte des corrélations avec les valeurs prises en d'autres sites. C'est donc une méthode probabiliste d'interpolation de données, qui contrairement à ses homologues déterministes, présente l'avantage de quantifier également l'incertitude formelle de l'estimation en chaque point.

L'interpolation par krigage est intrinsèquement liée à l'analyse variographique du phénomène à étudier mais n'est pas sa seule application pratique, nous en verrons d'autres dans le chapitre 4.

Nous avons vu dans l'activité VI comment la définition d'un processus stochastique permet, par *moyennage* d'un ensemble de simulations, d'interpoler un semis de points d'observation. La solution obtenue peut alors être considérée comme la plus représentative (ou la plus centrale plutôt devrait-on dire) d'une collection de variables régionalisées passant toutes par les points à interpoler et possédant les propriétés statistiques (moyenne et covariance) du processus modélisant le phénomène. C'est à peu de chose près ce que propose de faire le krigage avec une différence capitale² toutefois : le krigage pose algébriquement les contraintes souhaitées sur la solution, et la résolution du problème se fait donc à l'aide des techniques matricielles classiques.

Dans la présentation qui suit, nous montrons comment interpoler la fonction en un point x précis du domaine d'étude. La résolution complète du problème d'interpolation nécessitera alors d'itérer la méthode sur une grille de valeurs, comme nous le ferons dans l'activité VIII.

3.2.1 Contrainte de linéarité

Moyenne et covariance ne sont que les 2 premières composantes d'une représentation de la loi de probabilité à base de moments. Lorsqu'on décide de ne travailler qu'avec ces deux quantités, on opère en quelques sortes une *linéarisation* de la loi complète du processus. Avec cette information réduite, il devient alors impossible de manipuler des quantités qui ne soient pas des combinaisons linéaires de variables aléatoires. Par conséquent, le krigage ne peut être qu'un estimateur linéaire :

²Nous verrons d'autres différences plus précisément dans le chapitre 4

Contrainte 1 (Linéarité)

La valeur $\hat{z}(x)$ interpolée au site x est une combinaison linéaire des observations :

$$\hat{z}(x) = \sum_{i=1}^n w_i z(x_i) = \mathbf{W}^T \mathbf{Z}$$

où $w_i \in \mathbb{R}$ correspond au poids affecté à l'observation $z(x_i)$.

Pour une résolution matricielle du problème, nous noterons : $\mathbf{W} = [w_1, w_2, \dots, w_n]^T$, le vecteur de \mathbb{R}^n contenant les poids de la combinaison linéaire, et $\mathbf{Z} = [z(x_1), z(x_2), \dots, z(x_n)]^T$ le vecteur de \mathbb{R}^n contenant (dans le même ordre) les observations. Intuitivement, plus une observation sera proche du site à interpoler, plus elle aura des chances de recevoir une pondération importante. Cette formulation paraît assez peu rigoureuse : en pratique, et c'est là un atout du krigage, la valeur du poids affecté à une observation va aussi dépendre de la configuration géométrique de toutes les autres observations. En particulier, une observation proche mais redondante peu recevoir un poids plus faible qu'une observation plus lointaine mais isolée.

Dès lors, l'objectif du problème est clair : estimer $\hat{z}(x)$ revient à estimer le vecteur \mathbf{W} . Attention : ce vecteur \mathbf{W} va dépendre du point x sur lequel on réalise l'interpolation ! En toute rigueur, on devrait écrire $\mathbf{W}(x)$, mais nous ferons abstraction de x pour ne pas alourdir la présentation.

Remarquons qu'aucune contrainte n'impose que les poids w_i soient positifs. Imaginons un instant le cas d'un étudiant qui calcule la moyenne de ses deux notes 10 et 15, avec les coefficients respectifs -1 et 2, comme si la première note était en quelques sortes *récursive*. On dit que le krigage est un estimateur **non-convexe**³ ce qui est une propriété très intéressante pour une interpolation, comme nous le verrons plus loin.

La solution $\hat{z}(x)$ obtenue est appelée *valeur krigée* en x .

3.2.2 Contrainte d'autorisation

C'est certainement la moins intuitive des 4 contraintes, certainement parce qu'elle ne semble pas présenter d'intérêt au premier abord. La contrainte précédente implique que, une fois les poids w_i déterminés, l'erreur d'estimation en x s'exprime par :

$$\hat{z}(x) - z(x) = \sum_{i=1}^n w_i z_i - z_0 \tag{3.6}$$

Pour pouvoir contrôler statistiquement la qualité du krigage, on doit être capable de calculer la moyenne et la variance de cette erreur. En effet, si la variance ne peut s'exprimer, comment peut-on avoir des garanties sur l'exactitude du résultat ? Calculer une solution munie d'une variance infinie n'a aucun sens en pratique, et ne présente guère plus d'intérêt que d'essayer de deviner à l'aveugle la valeur à estimer. Il est donc important de s'assurer que la combinaison linéaire 3.6 possède une espérance et une variance finies. Une telle combinaison est dite *autorisée*.

³En géométrie, un ensemble est dit convexe lorsque 2 de ses points quelconques peuvent toujours être reliés *en ligne droite* sans sortir de l'ensemble. Un disque par exemple est un ensemble convexe, contrairement à un croissant de lune. Ici, dire que l'estimateur est convexe signifierait que les valeurs qu'il prend sont comprises entre la plus basse et la plus haute des valeurs observées. Clairement, l'interpolateur du graphe de gauche de la fig. 3.2 est non-convexe.

Contrainte 2 (Autorisation)

L'erreur d'estimation $\widehat{Z}(x) - Z(x)$ doit être d'espérance et de variance finies.

Lorsque le variogramme est borné, nous l'avons dit précédemment, le processus est stationnaire au sens large, sa covariance existe et on peut la relier au variogramme par la relation fondamentale 26. Le théorème 12 de propagation des incertitudes, nous montre alors comment calculer la variance d'une combinaison linéaire à l'aide de la matrice de covariance. La contrainte d'autorisation ne pose donc aucun problème. En revanche, nous verrons plus loin dans la section 3.3.2 que lorsque le variogramme diverge vers l'infini (*i.e.* quand le processus vérifie seulement l'hypothèse plus faible de stationnarité intrinsèque), seules certaines combinaisons de poids sont autorisées. Plus précisément, nous verrons que les combinaisons linéaires autorisées forment un sous-espace vectoriel de \mathbb{R}^n .

Remarque : la contrainte d'autorisation n'a donc d'implication pratique que pour les variogrammes non-bornés. Elle ne doit pas être négligée pour autant, car c'est elle qui garantit la validité statistique des calculs.

3.2.3 Contrainte d'universalité

On souhaite que l'estimateur ne produise pas d'erreur systématique. Statistiquement, on cherche donc que l'espérance de l'erreur (signée) soit nulle.

Contrainte 3 (Universalité)

L'erreur d'estimation doit être nulle en moyenne : $\mathbb{E}[\widehat{Z}(x) - Z(x)] = 0$.

Autrement dit, sur un nombre infini de répétitions de l'expérience, la moyenne des valeurs estimées par le krigeage est exactement égale à la valeur réelle.

Algébriquement, cette contrainte se traduit par :

$$\begin{aligned}\mathbb{E}[\widehat{Z}(x) - Z(x)] &= \mathbb{E}\left[\sum_{i=1}^n w_i Z(x_i)\right] - \mathbb{E}[Z(x)] = \sum_{i=1}^n w_i \mathbb{E}[Z(x_i)] - \mathbb{E}[Z(x)] \\ &= \sum_{i=1}^n w_i \mathbb{E}[Z] - \mathbb{E}[Z] = \mathbb{E}[Z] \left(\sum_{i=1}^n w_i - 1\right) = 0 \Leftrightarrow \sum_{i=1}^n w_i = 1\end{aligned}$$

La contrainte d'universalité impose donc que la somme des poids soit égale à 1. Soyons clair quant à l'implication pratique : si le variogramme est borné, le processus est stationnaire au sens large, la contrainte d'autorisation n'a aucun effet, mais la contrainte d'universalité impose de travailler avec des combinaisons de poids total unitaire. Si le variogramme est non-borné, le processus est seulement intrinsèquement stationnaire, et dans ce cas, les contraintes d'autorisation et d'universalité ont toutes deux pour effet d'imposer que la somme des poids soit unitaire.

Ayant fixé cette contrainte de non-biais, seule la composante de variance va déterminer la qualité de l'estimateur, comme le montre la formulation 1.13 de l'erreur quadratique moyenne. C'est précisément le rôle de la quatrième et dernière contrainte de minimiser cette quantité.

3.2.4 Contrainte d'optimalité

Parmi tous les estimateurs qui sont des combinaisons linéaires autorisées et sans biais on retient celle qui minimise la variance de l'erreur.

Contrainte 4 (Optimalité)

La variance de l'estimateur $\text{Var}[\widehat{Z}(x) - Z(x)]$ doit être minimale.

Empiriquement, on peut dire qu'on cherche l'estimateur le plus précis possible.

3.2.5 Formulation du problème

En combinant les quatre contraintes précédentes, on recherche un estimateur qui soit une combinaison linéaire autorisée, non-baisée et dont l'erreur est de variance minimale. Dans la littérature anglo-saxonne, on rencontre fréquemment l'acronyme d'*estimateur BLUE*, pour *Best Linear Unbiased Estimator*, qui fournit un bonne astuce mnémotechnique, mais qui possède l'inconvénient de reléguer la contrainte d'autorisation au second plan.

Remarquons que ces quatre contraintes sont emboîtées : le contrainte d'autorisation ne permet de travailler qu'avec des combinaisons linéaires, ce qui suppose donc implicitement que la contrainte de linéarité est vérifiée. Les contraintes d'universalité et d'optimalité requièrent quant à elles de pouvoir calculer les deux premiers moments de l'erreur d'estimation, ce qui suppose la contrainte d'autorisation. Enfin, la variance de l'erreur d'estimation, que minimise la contrainte d'optimalité, n'est vraiment un indicateur de la précision que si l'erreur est de moyenne nulle, ce qui suppose donc que la contrainte d'universalité est elle aussi vérifiée.

D'un point de vue mathématique, la recherche de la combinaison optimale \mathbf{W}^* s'écrit sous forme d'un problème de minimisation sous contrainte :

$$\mathbf{W}^* \in \underset{\mathbf{W} \in \mathbb{R}^n}{\operatorname{argmin}} \text{Var} \left[\sum_{i=1}^n w_i Z(x_i) - z(x) \right] \quad \text{sous la contrainte : } \sum_{i=1}^n w_i Z(x_i) = 1 \quad (3.7)$$

où $\operatorname{argmin} f(x)$ désigne l'ensemble des valeurs x telles que la fonction $f(x)$ est minimale :

$$\operatorname{argmin}_{x \in A} f(x) = \{x \in A \mid \forall y \in A, f(x) \leq f(y)\}$$

Notons que cet ensemble n'est pas nécessairement réduit à un élément. Si c'est le cas, les contraintes de krigeage suscitées ne préconisent que le choix arbitraire d'une solution optimale (parmi potentiellement de nombreuses autres). En pratique, excepté dans quelques cas pathologiques (par exemple lorsque deux sites sont confondus), la solution du problème est unique.

3.3 Le krigeage

Le problème 3.7 se résout à l'aide des multiplicateurs de Lagrange, dont nous effectuons un bref rappel ici. Pour plus d'information, on pourra consulter le cours d'optimisation de Dumont et al. (2019).

Considérons deux fonctions réelles à p variables $\mathbf{x} = (x_1, x_2, \dots, x_p) \in \mathbb{R}^p$: une fonction *objectif* (aussi appelée fonction de *coût*) : $f : \mathbb{R}^p \rightarrow \mathbb{R}$, ainsi qu'une fonction *contrainte* $g : \mathbb{R}^p \rightarrow \mathbb{R}$. Le problème suivant, consiste à minimiser f sur \mathbb{R}^p , sous la contrainte g :

$$\mathbf{x}^* \in \operatorname{argmin}_{\mathbf{x} \in \mathbb{R}^p} f(\mathbf{x}) \quad \text{sous la contrainte : } g(\mathbf{x}) = 0 \quad (3.8)$$

On suppose que f et g possèdent des dérivées partielles, et on note $\mathcal{L} : \mathbb{R}^{p+1} \rightarrow \mathbb{R}$ la fonction *lagrangienne* :

$$\mathcal{L}(\mathbf{x}, \lambda) = f(\mathbf{x}) - \lambda g(\mathbf{x}) \quad (3.9)$$

où λ est un réel appelé *multiplicateur de Lagrange*. On utilise alors le théorème suivant :

Théorème 35 (Minimisation par multiplicateurs de Lagrange)

Si \mathbf{x}_0 est un point d'extremum local de f , aussi appelé *point stationnaire*, alors :

$$\nabla \mathcal{L}(\mathbf{x}_0) = 0$$

Remarque : ∇ est l'opérateur gradient, qui à une fonction dérivable f prenant p variable en entrée, associe le vecteur de \mathbb{R}^p de ses p dérivées partielles :

$$\nabla : f \mapsto \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_p} \end{pmatrix}$$

Considérant ∇f comme un vecteur de l'espace à p dimensions, le théorème 35 nous dit que, qu'en un point \mathbf{x}_0 solution du problème, il existe un réel λ tel que $\nabla f(\mathbf{x}_0) = \lambda \nabla g(\mathbf{x}_0)$, autrement dit, le gradient de la surface à minimiser est perpendiculaire à la ligne de contrainte, comme l'illustre la figure 3.6. C'est assez intuitif, si ce n'était pas le cas, il serait possible de réduire davantage la fonction objectif en se déplaçant sur la ligne de contrainte, ce qui est contradictoire puisque \mathbf{x}_0 est un minimum local.

Remarque : le théorème 35 ne donne que les conditions nécessaires. Une fois les points stationnaires trouvés, on doit en théorie vérifier que l'un d'eux est bien un extrema.

Exemple 3.1. En guise de mise en pratique, considérons une nappe de terrain (infinie) de la forme d'un paraboloïde de révolution, et définie par l'équation : $z = f(x, y) = x^2 + y^2$. Nous savons qu'une ligne à haute tension parcours le terrain en ligne droite suivant l'équation $y = 2x + 75$. Déterminer un point minimal de cette ligne.

On pose g , la fonction de contrainte définie par : $g(x, y) = y - 2x - 75$. On forme alors le lagrangien, à 3 paramètres x, y et le multiplicateur de Lagrange λ :

$$\mathcal{L}(x, y, \lambda) = f(x, y) - \lambda g(x, y) = x^2 + y^2 - \lambda(y - 2x - 75)$$

On calcule alors le gradient de \mathcal{L} : $\nabla \mathcal{L}(x, y, \lambda) = \begin{pmatrix} \frac{\partial \mathcal{L}}{\partial x} \\ \frac{\partial \mathcal{L}}{\partial y} \\ \frac{\partial \mathcal{L}}{\partial \lambda} \end{pmatrix} = \begin{pmatrix} 2x + 2\lambda \\ 2y - \lambda \\ y - 2x - 75 \end{pmatrix}$

On résout enfin l'équation vectorielle $\nabla \mathcal{L}(x, y, \lambda) = 0$ pour obtenir une solution : $\mathbf{x}^* = (-30, 15)$.

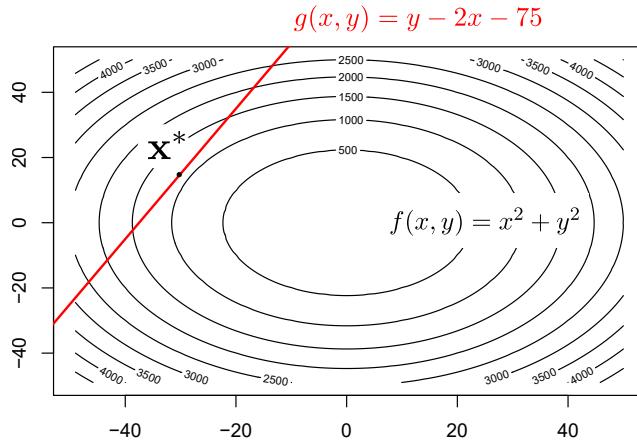


Fig. 3.6. Le terrain f (isolignes noires), la contrainte g (en rouge) et le point optimal $\mathbf{x}^* = (-30, 15)$. On voit clairement, que le gradient du terrain est orthogonal à la ligne de contrainte en \mathbf{x}^* .

Dans le cadre du problème initial 3.7, la fonction f correspond à la variance de l'erreur d'estimation, et la fonction g traduit la contrainte imposant que la somme des poids soit nulle. En notant $z_i = z(x_i)$ les relevés effectués au niveau des sites, on forme alors le lagrangien $\mathcal{L} : \mathbb{R}^{n+1} \rightarrow \mathbb{R}$:

$$\mathcal{L}(w_1, w_2, \dots, w_n, \lambda) = \text{Var}\left[\sum_{i=1}^n w_i Z_i - z_0\right] - 2\lambda\left(\sum_{i=1}^n w_i - 1\right) \quad (3.10)$$

où $\lambda \in \mathbb{R}$ est un multiplicateur de Lagrange, variable latente inconnue du problème. Le facteur 2 devant le λ est purement arbitraire, et permettra par la suite de simplifier les calculs.

3.3.1 Le krigeage ordinaire

Supposons dans un premier temps que Z est un processus stationnaire au sens large. On peut alors utiliser la covariance et en vertu du théorème 26 on : $C(h) = C(0) - \gamma(h)$, où $C(0) = \text{Var}[Z]$ est le palier du variogramme. D'où :

$$\begin{aligned} \text{Var}\left[\sum_{i=1}^n w_i Z_i - z_0\right] &= \text{Var}\left[\sum_{i=1}^n w_i Z_i\right] + \text{Var}[z_0] - 2\text{Cov}\left(\sum_{i=1}^n w_i Z_i, z_0\right) = \sum_{i=1}^n \sum_{j=1}^n w_i w_j \text{Cov}(Z_i, Z_j) \\ &+ \text{Var}[z_0] - 2 \sum_{i=1}^n w_i \text{Cov}(Z_i, z_0) = \sum_{i=1}^n \sum_{j=1}^n w_i w_j (\text{Var}[Z] - \gamma(h_{ij})) + \text{Var}[z_0] - 2 \sum_{i=1}^n w_i (\text{Var}[Z] - \gamma(h_{i0})) \end{aligned}$$

où h_{ij} désigne la distance entre les sites d'observation x_i et x_j . De même h_{i0} désigne la distance entre le site d'observation x_i et le site x sur lequel on souhaite calculer l'estimateur. En remarquant que $\sum_i \sum_j w_i w_j \text{Var}[Z] = \text{Var}[Z] \sum_i \sum_j w_i w_j = \text{Var}[Z] \sum_i w_i (\sum_j w_j) = \text{Var}[Z] \sum_i w_i = \text{Var}[Z]$, puisque les sommes de poids valent 1, et en abrégant $\gamma_{ij} = \gamma(h_{ij})$, on obtient :

$$\begin{aligned} \text{Var}\left[\sum_{i=1}^n w_i Z_i - Z_0\right] &= \sum_{i=1}^n \sum_{j=1}^n w_i w_j \text{Var}[Z] - \sum_{i=1}^n \sum_{j=1}^n w_i w_j \gamma_{ij} + \text{Var}[Z] - 2 \sum_{i=1}^n w_i \text{Var}[Z] + 2 \sum_{i=1}^n w_i \gamma_{i0} \\ &= \text{Var}[Z] - \sum_{i=1}^n \sum_{j=1}^n w_i w_j \gamma_{ij} + \text{Var}[Z] - 2 \text{Var}[Z] + 2 \sum_{i=1}^n w_i \gamma_{i0} = 2 \sum_{i=1}^n w_i \gamma_{i0} - \sum_{i=1}^n \sum_{j=1}^n w_i w_j \gamma_{ij} \end{aligned}$$

En ajoutant le terme de contrainte, nous obtenons une expression plus explicite du lagrangien :

$$\mathcal{L}(\mathbf{W}, \lambda) = 2 \sum_{i=1}^n w_i \gamma_{i0} - \sum_{i=1}^n \sum_{j=1}^n w_i w_j \gamma_{ij} - 2\lambda \left(\sum_{i=1}^n w_i - 1 \right)$$

Il ne reste alors plus qu'à calculer les dérivées partielles (en w_i et λ , soit $n+1$ dérivées au total). La dérivée partielle par rapport à un poids w_i arbitraire s'écrit :

$$\frac{\partial \mathcal{L}}{\partial w_i} = 2\gamma_{i0} - 2 \sum_{j=1}^n w_j \gamma_{ij} - 2\lambda$$

En divisant cette équation par 2, puis ajoutant l'équation sur la dérivée partielle en λ on obtient le système d'équations de krigeage de $n+1$ équations à $n+1$ inconnues :

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial w_i} = \gamma_{i0} - \sum_{j=1}^n w_j \gamma_{ij} - \lambda = 0 \\ \frac{\partial \mathcal{L}}{\partial \lambda} = \sum_{i=1}^n w_i - 1 = 0 \end{cases} \quad (3.11)$$

Remarquons que la dernière équation n'est rien d'autre que la contrainte de somme des poids unitaire (contrainte d'autorisation et/ou d'universalité). La résolution de 3.11 s'effectue finalement à l'aide de l'algèbre linéaire : on note $\boldsymbol{\Gamma} \in \mathbb{R}^{n \times n}$ la matrice des variances entre sites observés : $\Gamma_{ij} = \gamma_{ij} = \gamma(h_{ij}) = \gamma(\|x_i - x_j\|) \quad \forall i, j \in [1..n]$

$$\boldsymbol{\Gamma} = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \dots & \gamma_{1n} \\ \gamma_{21} & \gamma_{22} & \dots & \gamma_{2n} \\ \dots & \dots & \dots & \dots \\ \gamma_{n1} & \gamma_{n2} & \dots & \gamma_{nn} \end{bmatrix}$$

Grâce à la modélisation variographique, cette matrice $\boldsymbol{\Gamma}$ possède nécessairement les propriétés algébriques adéquates, et nous n'aurons pas de mauvaise surprise par la suite.

De même on note $\boldsymbol{\Gamma}_0 \in \mathbb{R}^n$ le vecteur des variances entre les sites observés et le site à estimer x_0 :

$$\boldsymbol{\Gamma}_0 = [\gamma_{10} \quad \gamma_{20} \quad \dots \quad \gamma_{n0}]^T$$

On vérifie alors facilement que 3.11 s'écrit matriciellement sous la forme :

$$\begin{bmatrix} \boldsymbol{\Gamma} & \mathbf{1} \\ \mathbf{1}^T & 0 \end{bmatrix} \begin{bmatrix} \mathbf{W} \\ \lambda \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Gamma}_0 \\ 1 \end{bmatrix} \quad \Rightarrow \quad \begin{bmatrix} \mathbf{W} \\ \lambda \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Gamma} & \mathbf{1} \\ \mathbf{1}^T & 0 \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{\Gamma}_0 \\ 1 \end{bmatrix}$$

où $\mathbf{1} = [1, 1, \dots, 1]^T$ est un vecteur colonne contenant n fois la valeur 1.

Remarque : la matrice à inverser ne dépend que des configurations des sites d'observation. Elle peut donc être inversée *une fois pour toutes* en amont de la procédure d'interpolation.

On récupère alors les n premières composantes du vecteur solution : w_1, w_2, \dots, w_n et on peut calculer l'estimateur du krigeage en x par :

$$\hat{z}(x) = \sum_{i=1}^n w_i z(x_i) \tag{3.12}$$

La variance de l'erreur associée, appelée **variance de krigeage**, s'exprime par :

$$\text{Var}[\hat{Z}(x) - Z(x)] = 2 \sum_{i=1}^n w_i \gamma_{i0} - \sum_{i=1}^n \sum_{j=1}^n w_i w_j \gamma_{ij} = \sum_{i=1}^n w_i \gamma_{i0} - \lambda \tag{3.13}$$

où la dernière égalité résulte d'une transformation opérée grâce à la première équation du système 3.11.

Elle traduit la précision de l'estimation, au sens de l'écart quadratique moyen entre la variable estimée et la valeur vraie au site considéré. Elle dépend donc naturellement du site sur lequel l'interpolation est effectuée. Intuitivement, plus le site est éloigné des observations, plus sa variance va être élevée.

Remarquons que la variance d'estimation en un site ne dépend que de sa configuration géométrique relativement aux autres sites, et non des valeurs observées. La variance peut donc être calculée *a priori*, avant d'effectuer les observations sur le terrain, ce qui constitue un atout indéniable pour valider en amont les plans d'expérimentation.

3.3.2 Le krigeage intrinsèque

Qu'en est-il dans le cas strictement intrinsèque ? Le passage du variogramme à la covariance grâce au théorème 26 n'est plus possible, et on doit de surcroit s'assurer que les combinaisons linéaires manipulées sont autorisées, comme le stipule la contrainte 2. On considère l'erreur de l'estimation 3.6. Remarquons qu'il s'agit bien d'une combinaison linéaire des observations $(Z_1, Z_2, \dots, Z_n, Z_0)$ par les poids $(w_1, w_2, \dots, w_n, -1)$.

Pour pouvoir calculer l'espérance et la variance de cette quantité, on doit être capable d'exprimer variables aléatoires sous forme d'accroissement, *i.e.* sous la forme $Z_i - Z_j$ avec i et j deux indices quelconques, puisque sous l'hypothèse intrinsèque, l'existence de l'espérance et de la variance ne sont garanties que pour les accroissements. On vérifie facilement que ceci n'est possible que si la somme des poids de la combinaison est nulle. En effet, dans ce cas, la somme des w_i est nécessairement égale à 1 et on peut écrire :

$$\varepsilon = \sum_{i=1}^n w_i Z_i - Z_0 = \sum_{i=1}^n w_i Z_i - \sum_{i=1}^n w_i Z_0 = \sum_{i=1}^n w_i (Z_i - Z_0)$$

Sous cette formulation, on dérive facilement l'expression de l'espérance de l'erreur d'estimation :

$$\mathbb{E}[\varepsilon] = \mathbb{E}\left[\sum_{i=1}^n w_i (Z_i - Z_0)\right] = \sum_{i=1}^n w_i \mathbb{E}[Z_i - Z_0] = 0$$

où la seconde égalité résulte de la linéarité de l'espérance et de l'existence des espérances des accroissements $Z_i - Z_0$, tandis que la dernière égalité résulte de la stationnarité de ces accroissements. On peut faire de même pour la variance :

$$\begin{aligned} \text{Var}[\varepsilon] &= \text{Var}\left[\sum_{i=1}^n w_i (Z_i - Z_0)\right] = \mathbb{E}\left[\left(\sum_{i=1}^n w_i (Z_i - Z_0)\right)^2\right] \\ &= \mathbb{E}\left[\sum_{i=1}^n \sum_{j=1}^n w_i w_j (Z_i - Z_0)(Z_j - Z_0)\right] = \sum_{i=1}^n \sum_{j=1}^n w_i w_j \mathbb{E}[(Z_i - Z_0)(Z_j - Z_0)] \end{aligned}$$

Par ailleurs, on peut écrire : $2\gamma_{ij} = \text{Var}[Z_i - Z_j] = \text{Var}[Z_i - Z_0 + Z_0 - Z_j] = \mathbb{E}[(Z_i - Z_0 + Z_0 - Z_j)^2] = \mathbb{E}[(Z_i - Z_0)^2] + \mathbb{E}[(Z_j - Z_0)^2] - 2\mathbb{E}[(Z_i - Z_0)(Z_j - Z_0)]$ et donc :

$$\mathbb{E}[(Z_i - Z_0)(Z_j - Z_0)] = -\gamma_{ij} + \gamma_{i0} + \gamma_{j0}$$

En substituant dans l'expression de $\text{Var}[\varepsilon]$, on obtient une formulation ne dépendant que de γ :

$$\text{Var}[\varepsilon] = \sum_{i=1}^n \sum_{j=1}^n w_i w_j \gamma_{i0} + \sum_{i=1}^n \sum_{j=1}^n w_i w_j \gamma_{j0} - \sum_{i=1}^n \sum_{j=1}^n w_i w_j \gamma_{ij} = \sum_{i=1}^n w_i \gamma_{i0} + \sum_{j=1}^n w_j \gamma_{j0} - \sum_{i=1}^n \sum_{j=1}^n w_i w_j \gamma_{ij}$$

Et donc, en faisant le changement de la variable muette $j \mapsto i$ dans la seconde sommation, on obtient :

$$\text{Var}[\varepsilon] = 2 \sum_{i=1}^n w_i \gamma_{i0} - \sum_{i=1}^n \sum_{j=1}^n w_i w_j \gamma_{ij}$$

On retrouve exactement l'expression 3.13 de la variance d'erreur d'estimation du krigeage ordinaire. L'équation lagrangienne à minimiser sera donc exactement la même et on peut conclure :

Le krigeage ordinaire (i.e. sous l'hypothèse de stationnarité au sens large du processus) et le krigeage intrinsèque (sous l'hypothèse plus faible de stationnarité des incrément du processus) ont exactement la même solution lorsqu'elle est exprimée en termes de variogramme.

On comprend ici que le variogramme est un outil plus générique que la covariance. Remarquons toutefois une différence dans le sens donné aux contraintes : dans le cas **stationnaire**, la contrainte d'autorisation n'a aucun impact, puisque toutes les combinaisons linéaires sont autorisées. En revanche, la contrainte d'universalité stipule que la somme des poids doit valoir 1 pour que l'estimateur soit non-biaisé. Dans le cas **intrinsèque** la contrainte d'autorisation stipule que les combinaisons linéaires dont on souhaite exprimer l'espérance ou la variance doivent être de poids total nul (afin de pouvoir les exprimer comme des incrément). Pour pouvoir calculer la variance de l'erreur, la somme des poids total doit donc valoir 1 et la contrainte d'autorisation est redondante avec la contrainte d'universalité.

Au passage, on peut mener les mêmes développements pour une combinaison linéaire quelconque des observations, et on trouve :

$$\text{Var} \left[\sum_{i=1}^n w_i Z_i \right] = - \sum_{i=1}^n \sum_{j=1}^n w_i w_j \gamma_{ij} \quad (3.14)$$

La variance étant une quantité nécessairement positive, on retrouve la contrainte 2.24 de définitivité conditionnelle du variogramme :

$$\sum_{i=1}^n w_i = 0 \Rightarrow \sum_{i=1}^n \sum_{j=1}^n w_i w_j \gamma_{ij} \leq 0$$

3.4 Mise en oeuvre du krigeage

Dans cette section, nous voyons des exemples pratiques d'interpolation par krigeage à partir d'un semis de points d'observation, dans un premier temps en mode local (une seul estimation) et *à la main* (activité VII). Nous verrons un exemple plus opérationnel par la suite, dans lequel nous résolverons le problème modèle et construirons notre premier MNT par krigeage (activité VIII).

La méthodologie est la suivante (en supposant que γ est un modèle de variogramme licite, dont les paramètres ont été estimés éventuellement à partir d'un variogramme expérimental) :

- Calcul de la matrice \mathbf{H} contenant toutes les distances $h_{ij} = \|x_i - x_j\|$ entre les sites observés (en général, cette matrice a déjà été évaluée précédemment lors de l'estimation du variogramme expérimental), et du vecteur \mathbf{H}_0 contenant les distances $h_{i0} = \|x_i - x_0\|$ entre les sites observés et le site à interpoler.
- Transformation de \mathbf{H} et \mathbf{H}_0 par la fonction γ : $\mathbf{\Gamma} = \gamma(\mathbf{H})$ et $\mathbf{\Gamma}_0 = \gamma(\mathbf{H}_0)$.

- Formation de la matrice \mathbf{A} en complétant \mathbf{H} par une colonne et une ligne de 1 et en ajoutant un 0 dans le coin inférieur droit.
- Formation de la matrice \mathbf{B} en complétant le vecteur \mathbf{H}_0 par un 1.
- Calcul de la solution \mathbf{X} du système : $\mathbf{AX} = \mathbf{B}$ et récupération de ses n premiers éléments dans un vecteur colonne de poids \mathbf{W} .
- Calcul de l'estimation par le produit scalaire : $\hat{z}(x_0) = \mathbf{W}^T \mathbf{Z}$
- Calcul de la variance d'estimation associée : $\sigma^2 = \mathbf{W}^T \mathbf{\Gamma}_0 - \lambda$, où λ est le dernier terme de \mathbf{X} .

Activité VII. Un krigage en 7 lignes de code

On considère le jeu de données, composé de 5 sites x_1, x_2, \dots, x_5 disposés comme représenté sur la figure 3.7 ci-dessous. Une observation z_i (d'un phénomène quelconque Z) a été relevée sur chaque site x_i .

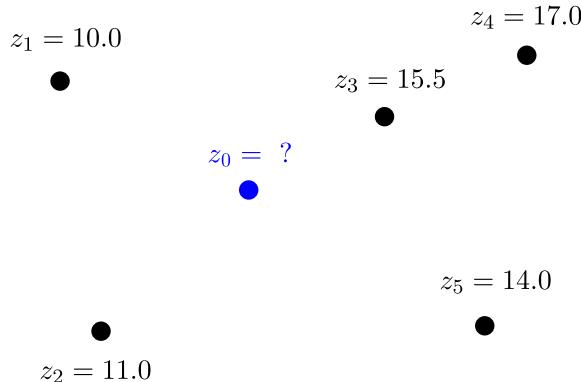


Fig. 3.7. On donne 5 points d'observation. L'objectif du problème consiste à estimer z_0 .

Par ailleurs, on donne les distances h_{ij} entre les 5 sites dans une matrice \mathbf{H} :

$$\mathbf{H} = \begin{bmatrix} 0.0 & 24.2 & 30.5 & 44.0 & 46.3 \\ 24.2 & 0.0 & 33.0 & 47.4 & 36.2 \\ 30.5 & 33.0 & 0.0 & 14.8 & 22.1 \\ 44.0 & 47.4 & 14.8 & 0.0 & 25.7 \\ 46.3 & 36.2 & 22.1 & 25.7 & 0.0 \end{bmatrix}$$

Par exemple, $H_{23} = 33$ signifie que les sites x_i et x_j sont distants de 33 m. On donne également les distances h_{i0} entre ces 5 sites et le site à interpoler : $\mathbf{H}_0 = [20.6 \ 19.0 \ 14.4 \ 29.1 \ 25.8]^T$. Nous utiliserons un variogramme linéaire avec palier :

$$\gamma(h) = \begin{cases} h & \text{si } h \leq 40 \\ 40 & \text{sinon.} \end{cases} \quad (3.15)$$

Toutes les opérations effectuées ci-dessous peuvent être reproduites par le lecteur, à l'aide d'un langage de programmation disposant des fonctionnalités classiques d'algèbre linéaire (somme, produit et

inversion de matrice notamment). Le code R correspondant est donné à la fin de ce paragraphe.

On calcule tout d'abord la matrice $\boldsymbol{\Gamma} = \gamma(\mathbf{H})$, ce qui consiste simplement à seuiller les éléments supérieurs à 40. On forme ensuite la matrice \mathbf{A} en complétant $\boldsymbol{\Gamma}$ avec une colonne de 1, une ligne de 1 et un 0 dans le coin inférieur droit.

$$\boldsymbol{\Gamma} = \begin{bmatrix} 0.0 & 24.2 & 30.5 & \mathbf{40.0} & \mathbf{40.0} \\ 24.2 & 0.0 & 33.0 & \mathbf{40.0} & 36.2 \\ 30.5 & 33.0 & 0.0 & 14.8 & 22.1 \\ \mathbf{40.0} & \mathbf{40.0} & 14.8 & 0.0 & 25.7 \\ \mathbf{40.0} & 36.2 & 22.1 & 25.7 & 0.0 \end{bmatrix} \quad \mathbf{A} = \begin{bmatrix} 0.0 & 24.2 & 30.5 & 40.0 & 40.0 & \mathbf{1} \\ 24.2 & 0.0 & 33.0 & 40.0 & 36.2 & \mathbf{1} \\ 30.5 & 33.0 & 0.0 & 14.8 & 22.1 & \mathbf{1} \\ 40.0 & 40.0 & 14.8 & 0.0 & 25.7 & \mathbf{1} \\ 40.0 & 36.2 & 22.1 & 25.7 & 0.0 & \mathbf{1} \\ \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{0} \end{bmatrix}$$

Par souci de clarté, on figure en gras les éléments ajoutés ou modifiés à chaque étape du calcul.

On effectue alors la même opération pour le vecteur \mathbf{H}_0 :

$$\boldsymbol{\Gamma}_0 = \gamma(\mathbf{H}_0) = [20.6 \ 19.0 \ 14.4 \ 29.1 \ 25.8]^T$$

$$\mathbf{B} = [20.6 \ 19.0 \ 14.4 \ 29.1 \ 25.8 \ \mathbf{1}]^T$$

On résout informatiquement le système $\mathbf{AX} = \mathbf{B}$, en rappelant que $\mathbf{X} = [\mathbf{W}^T, \lambda]^T$ est un vecteur composé des $n = 5$ poids et du multiplicateur de Lagrange :

$$\mathbf{X} = \mathbf{A}^{-1}\mathbf{B} = [0.195 \ 0.304 \ 0.553 \ -0.147 \ 0.093 \ -1.501]^T$$

$$\mathbf{W} = [0.195 \ 0.304 \ 0.553 \ -0.147 \ 0.093]^T$$

L'estimation finale s'obtient alors immédiatement par combinaison linéaire :

$$\hat{z}(x_0) = \mathbf{W}^T \mathbf{Z} = \sum_{i=1}^5 w_i z_i \approx 12.7$$

ainsi que l'écart-type associé : $\sigma^2 = \mathbf{W}^T \mathbf{T}_0 - \lambda = 17.38 = 4.16$

Le résultat final de l'estimation est donc $\hat{z}(x_0) = 12.7 \pm 4.16$. Notons que l'écart-type associé est relativement élevé, ce à quoi on pouvait s'attendre étant donné le faible nombre d'observations disponibles.

Comment interpréter la variance de l'erreur d'estimation en pratique ? C'est une question qui n'admet pas de réponse pleinement satisfaisante. Le phénomène étudié et le jeu de données à disposition sont par nature uniques, et la notion de variance d'une estimation autour de sa valeur vraie ne fait pas nécessairement sens. En statistique classique, on préfère souvent exprimer l'incertitude sous la forme d'un intervalle de confiance : par exemple, on dira qu'avec une risque de 5%, l'altitude en un point x_0 du terrain vaut \hat{z} plus ou moins une marge d'erreur Δz . Cependant, pour définir l'intervalle de

confiance, on doit connaître la loi de probabilité de l'estimateur, ce qui en Géostatistique est impossible du fait de la complexité de cette loi. On peut alors poser l'hypothèse (plus ou moins valide en fonction des situations) que l'erreur d'estimation est distribuée suivant une loi normale. La largeur de l'intervalle de confiance est alors complètement déterminée par l'écart-type. En particulier, pour un risque de 5%, on $\Delta z = 1.96\sigma$. Dans cet exemple pédagogique on a $\sigma = 4.16$, et donc on pourra dire qu'avec un taux de confiance de 95%, la valeur vraie est comprise entre 4.38 et 21.02.

On donne ci-dessous le code R permettant de reproduire les calculs. Notons que la multiplication matricielle s'écrit avec l'opérateur `%*%`.

Script 8. Un krigeage en 7 lignes de code [code6.r]

```
# -----
# Donnees
# -----
H = matrix(
  c(0.00, 24.2, 30.5, 44.0, 46.3,
    24.2, 0.00, 33.0, 47.4, 36.2,
    30.5, 33.0, 0.00, 14.8, 22.1,
    44.0, 47.4, 14.8, 0.00, 25.7,
    46.3, 36.2, 22.1, 25.7, 0.00),
  nrow=5, ncol=5)

h = matrix(c(20.6,19.0,14.4,29.1,25.8), nrow=5, ncol=1)

Z = matrix(c(10,11,15.5,17,14), nrow=5, ncol=1)
# -----


# Variogramme
H[which(H > 40)] = 40
h[which(h > 40)] = 40

# Completion de 1 et 0
A = rbind(H,rep(1,5))
A = cbind(A,c(rep(1,5),0))
B = rbind(h,1)

# Resolution
W = solve(A,B)[1:5]
z0 = (W%*%Z)[1]
```

Remarquons que les instructions `solve(A,B)` et `solve(A)%*%B` sont équivalentes (aux erreurs d'arrondi près), mais si la seconde est plus proche de la notation formelle $\mathbf{A}^{-1}\mathbf{B}$, la première reste la plus efficace d'un point de vue informatique. L'indexation `[1:5]` qui s'y ajoute permet de ne récupérer que les 5 premiers éléments du vecteur résultat, à savoir les 5 poids solution. L'instruction `which(H > 40)` permet de lister tous les indices de la matrice **H** contenant un élément supérieur à 40. La commande `H[which(H > 40)] = 40` permet donc de *plafonner* les éléments de la matrice **H** à la valeur 40.

Quelques remarques sur le résultat de cette activité :

- La somme des poids affectés aux mesures est bien égale à 1, mais comme nous l'avions fait observer précédemment, l'estimation n'est en général pas convexe et certains poids peuvent être

négatifs comme c'est le cas ici pour $w_4 = -0.147$. La figure 3.7 nous révèle que le site x_4 est en quelques sortes *caché* derrière le site x_3 . On parle d'**effet d'écran**. C'est donc x_3 qui va principalement déterminer la valeur interpolée ($w_3 = 0.553$, qui est le plus élevé des poids). Le site x_3 , aligné avec x_0 et x_5 , agit comme un pivot, et on comprend que la rigidité de la surface à interpoler conduit z_5 à avoir une influence répulsive : pour une valeur de pivot z_3 fixée, plus z_5 est élevée, plus z_0 est faible et inversement. Cet exemple simple nous permet donc de comprendre l'intérêt capital d'autoriser des combinaisons linéaires non-convexes des observations.

- On remarque que la matrice \mathbf{A} ne dépend pas des observations z_i effectivement mesurées mais uniquement des positions x_i des sites d'observation. Dans un contexte où on souhaiterait estimer l'intégralité du domaine d'étude (comme nous allons le faire dans l'activité suivante), il peut être avantageux de calculer l'inverse de \mathbf{A} une seule fois en amont. Chaque estimation ne nécessite alors plus que de calculer \mathbf{H}_0 , \mathbf{B} , puis d'effectuer la multiplication matricielle $\mathbf{A}^{-1}\mathbf{B}$.
- Lorsqu'on dispose d'un très grand nombre d'observations (typiquement plusieurs dizaine de milliers), le calcul de l'inversion peut être relativement coûteux, d'autant que les observations lointaines se verront très vraisemblablement affecter des poids négligeables. Une solution classique consiste à retirer, en amont de la procédure d'interpolation de x_0 , tous les sites situés à une distance supérieure à un certain seuil (par exemple 3 fois la portée effective du variogramme). On parle de *voisinage glissant*. En contre-partie, cela nécessite de recalculer la matrice \mathbf{A} (et donc aussi son inverse) pour chaque nouvelle estimation.

Activité VIII. Calcul d'un Modèle Numérique de Terrain

On considère un semis de points sur lesquels ont été effectués un relevé d'altitude. Chaque ligne du jeu de données contient : les coordonnées géographiques (longitude et latitude) d'un point, et l'altitude (en mètres) mesurée sur ce point. L'objectif du problème consiste à construire un modèle numérique de terrain sur la zone d'étude (on définit ici la zone d'étude comme le rectangle englobant dont les coins sont définis par les coordonnées extrémales du semis de points d'observation).

Ici aussi, tout comme pour l'activité IV on pourra réutiliser le jeu de données saisi sur l'application Google Maps, ou alors récupérer l'un des deux fichiers de données disponibles dans le répertoire de matériel pédagogique. Pour les exemples ci-dessous, nous travaillerons avec le fichier `sample_data2.txt`.

Dans cette activité, on utilisera un modèle de variogramme linéaire (sans palier) : $\gamma : h \mapsto h$.

Correction :

Comme pour l'activité IV, on commence par charger les données et convertir les coordonnées dans un système métrique de coordonnées plane (cette étape est importante pour que les deux dimensions Ouest-Est et Nord-Sud aient la même importance dans le calcul des distances entre sites) :

```
# Import des fichiers
obs = read.csv("sample_data1.txt", sep=",", header=1)

# Conversion en coordonnées planes
R = 6378137.0
obs$longitude = obs$longitude*pi/180.0
obs$latitude = obs$latitude*pi/180.0
obs$longitude = (obs$longitude - min(obs$longitude))*cos(mean(obs$latitude))*R
obs$latitude = (obs$latitude - min(obs$latitude))*R
```

Pour simplifier les manipulations, nous séparons les coordonnées des sites et les relevés effectués :

```
# Table des sites
sites = obs[c("longitude", "latitude")]

# Table des observations
Z = obs["meters"]
Z = rbind(Z, 0)

# Nombre de données
N = nrow(sites)
```

Remarquons qu'un 0 est ajouté à la fin du vecteur d'observations, de sorte que le produit de Z par le vecteur solution du système de krigeage élimine automatiquement le multiplicateur de Lagrange dans le calcul de l'estimation.

On définit ensuite l'emprise géométrique de la zone d'étude, ainsi que la résolution de la grille d'interpolation. Cette résolution est exprimée dans l'unité des points d'observation, donc en mètres ici dans notre cas puisque nous avons projeté les coordonnées des sites.

```
# Resolution
resolution = 25

# Calcul de l'emprise
xmin = min(sites$longitude)
xmax = max(sites$longitude)
ymin = min(sites$latitude)
ymax = max(sites$latitude)
```

On définit alors l'objet informatique `zhat` qui va contenir le MNT estimé. On définit également une seconde grille `vhat` pour stocker l'incertitude de la valeur d'altitude estimée en chaque cellule.

```
# Parcours des sites
DX = (xmin/resolution):(xmax/resolution)*resolution
DY = (ymin/resolution):(ymax/resolution)*resolution

# Output
zhat = matrix(ncol=length(DX), nrow=length(DY))
vhat = matrix(ncol=length(DX), nrow=length(DY))
```

À ce stade, il peut être judicieux de tester le code ci-dessus, pour vérifier que les matrices `zhat` et `vhat` ont des tailles raisonnables, à l'aide des fonctions `nrow` et `ncol` (typiquement ces deux valeurs ne devraient pas dépasser 100 éléments pour commencer). Si besoin, ne pas hésiter à augmenter la valeur de la variable `resolution`.

Remarquons que pour cet exemple, nous considérons dans un premier temps un variogramme identité : $\gamma : h \mapsto h$, et donc l'application du variogramme aux distances h calculées est une étape transparente. Malgré tout, dans le but de pouvoir changer cette fonction par la suite, nous construisons la fonction γ :

```
vario = function(h){
  return(h)
}
```

On effectue les calculs de krigeage qui sont indépendants du site à interpoler, à savoir : le calcul de la matrice `D` des distances entre sites observés, le passage par le variogramme et l'ajout des colonnes/lignes de 1 pour compléter le système. On pré-calcule également l'inverse de la matrice obtenue.

```
# Calcul de la matrice de distance
D = data.matrix(dist(sites, diag=TRUE, upper=TRUE))
D = vario(D)

# Completion avec des 1
D = rbind(D, rep(1,N))
D = cbind(D, c(rep(1,N),0))

# Pre-calcul de l'inversion
I = solve(D)
```

L'instruction `dist(A, diag=TRUE, upper=TRUE)`, où `A` est un tableau de données contenant n points dans un espace de p dimensions, retourne un tableau de taille $n \times n$ contenant les distances euclidiennes entre tous les couples de points. Combinée avec l'instruction `data.matrix`, la commande globale permet de retourner ces distances sous forme d'une matrice.

Il ne reste alors plus qu'à parcourir la grille, ce qui se fait à l'aide d'une double boucle sur les éléments de `DX` et `DY`. Chaque itération de la boucle interne correspond à l'estimation d'un noeud de la grille.

Script 9. Calcul d'un Modèle Numérique de Terrain [code7.r]

```
for (ix in 1:length(DX)){
  for (iy in 1:length(DY)){

    # Position
    x = DX[ix]
    y = DY[iy]

    # Distance entre les sites et (x,y)
    D0 = sqrt((sites$longitude - x)**2+(sites$latitude - y)**2)
    D0 = vario(D0)

    # Completion avec un 1
    D0 = c(D0, 1)

    # Resolution
    W = (I %*% D0)

    # Estimation
    zhat[iy,ix] = sum(W*Z)
    vhat[iy,ix] = sum(W*D0) - W[N+1]

  }
}
```

Le calcul pouvant prendre un certain temps, on pourra introduire les deux lignes suivantes entre le

boucle externe et le boucle interne, pour visualiser à l'écran l'état du traitement :

```
# Barre de progression
pourcent = floor(100*ix/length(DX)); cat("Computing interpolation:",min(pourcent,100), "%\r")
A l'issue du calcul, on pourra visualiser le MNT (et son incertitude associée) :
```

```
par(mfrow=c(1,2))
image(DX, DY, t(zhat), col=terrain.colors(255), xlab="Estimation")
contour(DX, DY, t(zhat), add=TRUE, nlevels=20)
image(DX, DY, t(vhat), col=terrain.colors(8), xlab="Variance")
```

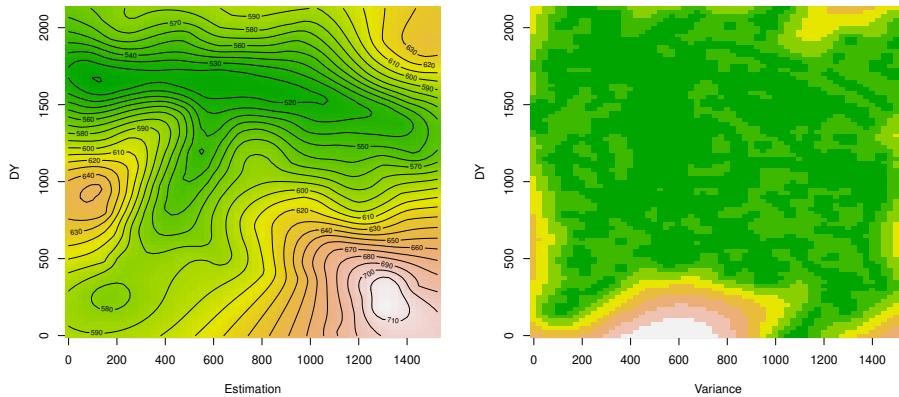


Fig. 3.8. Estimation par krigage d'un MNT (à gauche) et de son incertitude associée (à droite).

On remarque que l'incertitude de la valeur interpolée augmente à mesure qu'on s'éloigne des sites observés, ce qui paraît intuitif. Relancer le calcul avec différents modèles de variogrammes. En particulier, on testera plusieurs variogrammes linéaires avec des coefficients variés, *e.g.* $\gamma(h) = 2h$, $\gamma(h) = 3h$, etc. Observe-t-on des différences dans les résultats produits ? Peut-on l'expliquer ?

Pour gérer plusieurs variogrammes on doit compléter la fonction `vario`. Pour sélectionner un modèle de variogramme, il suffit de décommenter la ligne correspondante. Lorsqu'aucune ligne n'est décommentée (exceptée l'instruction `return` bien entendu), le variogramme retenu est le modèle de base $\gamma : h \mapsto h$.

```
vario = function(h){

  #h = 2*h                      # Lineaire coeff 2
  #h = 100*h                     # Lineaire coeff 100

  #h[which(h>8000)] = 8000      # Lineaire avec palier et portee 8000m
  #h[which(h>4000)] = 4000      # Lineaire avec palier et portee 4000m
  #h[which(h>2000)] = 2000      # Lineaire avec palier et portee 2000m

  #h = h**0.1                    # Puissance 0.1
  #h = h**1.9                    # Puissance 1.9
  #h = 1-exp(-h/1000)           # Exponentiel avec palier et portee 1000m

  return(h)
}
```

On pourra également tester le modèle gaussien si on le souhaite, mais en faisant attention au fait que celui-ci a souvent tendance à produire des erreurs d'arrondi numérique, menant à l'impossibilité d'inverser le système de krigage. On pourra contourner ce problème en introduisant une pépite de petite amplitude (*e.g.* ici 10^{-3}) dans le modèle.

```
pepite = 10**(-3)
y = pepite+1-exp(-h**2/578**2)
y[which(h == 0)] = 0; h = y
```

La portée $a = 578$ est calculée en divisant la portée pratique souhaitée (ici 1 km) par 1.73, comme expliqué dans la section 2.15. Une autre solution consiste à utiliser un modèle cubique.

On pourra alors effectuer des comparaisons entre les résultats. En particulier, si on dispose de deux estimations `zhat1` et `zhat2` (par exemple calculées avec 2 modèles variographiques différents), on pourra calculer le champ différentiel `diff=abs(zhat1-zhat2)` et :

- le représenter à l'aide de la fonction `image`
- calculer l'erreur quadratique moyenne `sqrt(mean(diff**2))` ou l'erreur maximale `max(diff)` entre les deux estimations

En effectuant ce test, on montre expérimentalement que le coefficient de la pente du modèle linéaire n'a absolument aucun impact sur les valeurs estimées, mais seulement sur les incertitudes associées. Nous tenterons d'expliquer cette observation dans la section suivante.

On peut ensuite tester d'autres modèles de variogramme, en particulier le modèle linéaire avec palier, dont nous avions montré dans la section 2.4.2 qu'il était un bon candidat pour le jeu de données `sample_data2.txt`. Nous avions même estimé la valeur du palier à 7.7 km. Ici nous considérerons donc un variogramme linéaire avec palier et portée de 8 km. Pour tester l'influence du paramètre de portée sur les résultats, nous testons également avec des variogrammes de différentes portées, ainsi que pour d'autres modèles.

La figure 3.9 montre les MNT obtenus pour 6 modèles de variogramme différents : sur la première ligne : linéaire avec palier et portées 8 km, 2 km et 500 m, et sur la seconde ligne : modèle puissance avec $\theta = 0.1$ et $\theta = 1.9$ (on rappelle que ce modèle n'est défini que pour θ strictement compris entre 0 et 2), et modèle exponentiel avec portée pratique de 8 km.

On observe les faits suivants :

- La valeur de la portée du variogramme linéaire n'est pas critique jusqu'à 3 ou 4 km de portée. Pour des valeurs inférieures, la qualité du modèle obtenu se dégrade rapidement.
- Le résultats obtenus pour les deux paramètres $\theta = 0.1$ et $\theta = 1.9$ de variogramme en puissance sont assez similaires. On remarque cependant que le modèle $\theta = 0.1$ semble moins lisse, ce qui s'explique par son comportement linéaire à l'origine, contrairement au modèle $\theta = 1.9$ qui est parabolique.
- Le variogramme exponentiel avec portée de 8 km semble donner des résultats très proches du modèle linéaire avec palier et portée de 8 km. Cela s'explique par les formes semblables de ces

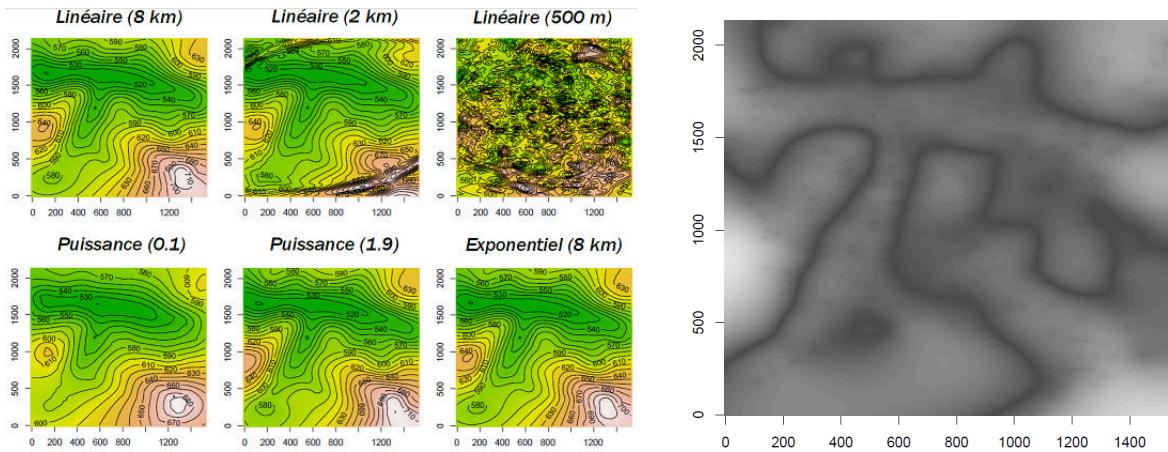


Fig. 3.9. À gauche : Estimation par krigage d'un MNT avec différents variogramme. À droite : différence des estimations entre les deux modèles puissance $\theta = 0.1$ et $\theta = 1.9$. Plus la couleur est claire, plus la différence absolue entre les deux estimations est grande.

deux variogrammes. Le modèle linéaire avec palier constitue donc en général une bonne approximation du modèle exponentiel (et dont les paramètres sont plus faciles à inférer comme nous l'avons vu dans l'activité IV).

On peut faire des comparaisons numériques entre les champs estimés :

Modèle 1	Modèle 2	Erreurs RMSE (m)	Erreurs max (m)
Linéaire (8 km)	Exponentiel (8 km)	0.33	4.41
Puissance (0.1)	Puissance (1.9)	13.94	62
Linéaire (8 km)	Linéaire (500 m)	44.89	188

Il peut être assez instructif également d'effectuer une comparaison visuelle du champ calculé par la différence des deux modèles en puissance (fig. 3.9 à droite). On observe que les différences sont maximales au niveau des zones où les courbures sont les plus prononcées (sommet et fond de vallée), ce qui peut s'expliquer par la différence de régularité des deux modèles de variogramme (linéaire à l'origine pour $\theta \leq 1$ et parabolique à l'origine pour $\theta > 1$). À l'inverse, les différences sont quasi-nulles au niveau des points observés mais aussi au niveau des lignes d'inflexion.

Ces quelques expérimentations montrent que le choix du variogramme n'est pas critique, tant que le modèle choisi reste fidèle à la structure sous-jacente des données, estimée par le variogramme expérimental. C'est ce que montre en particulier la comparaison entre le modèle linéaire avec palier et le modèle exponentiel, dont les différences obtenues sur l'estimation sont en général plutôt faibles. En revanche, lorsqu'un mauvais modèle est sélectionné, ou alors que les paramètres du modèle sont trop grossièrement estimés, la qualité de l'estimation se dégrade très rapidement.

3.5 Propriété du krigage

Nous l'avons vu dans la section 3.3.2, les solutions du krigage ordinaire et du krigage intrinsèque possèdent exactement la même formulation algébrique. Dans cette section nous parlerons donc de krigage sans plus de précision.

3.5.1 Exactitude

Un interpolateur est dit exact lorsqu'il passe exactement par toutes les données qui ont été utilisées pour le construire. Si on reprend l'exemple pédagogique de l'activité VI, nous avions clairement un interpolateur non-exact, sauf si on fixe la tolérance ε à 0. Formellement :

Propriété 36 (Exactitude du krigeage)

Le krigeage est un interpolateur exact, autrement dit, pour tout jeu de données $\{x_i, z_i\}_{i=1..n}$, la valeur krigée en tout point x_i est égale à la valeur observée : $\hat{z}(x_i) = z_i$.

Preuve. En un point x_i donné, la solution $\hat{z}(x_i) = z_i$ vérifie les 4 contraintes de krigeage. En effet, z_i est bien une combinaison linéaire des observations de poids total unitaire donc les contraintes de linéarité, d'universalité et d'autorisation sont vérifiées. Par ailleurs, l'erreur d'estimation étant nulle lorsqu'on choisit z_i , la contrainte d'optimalité est également vérifiée. En toute rigueur, il resterait à vérifier que la solution $\hat{z}(x_i) = z_i$ est l'unique solution possible au système de krigeage. Lorsque tous les sites sont distincts, on peut montrer que la solution du système est unique, z_i est donc la seule solution admissible et l'interpolateur est nécessairement exact.

Cette propriété peut être désavantageuse lorsque les observations (ou que les positions des sites sur lesquelles elles ont été effectuées) sont bruitées. Une méthode pour contourner ce problème, consiste à utiliser un variogramme à pépite. Dans ce cas, comme nous l'avons vu dans la section 2.20 (ou plus généralement dans la propriété 32), les réalisations du processus sont discontinues, et il en va de même pour la variable régionalisée estimée par krigeage. La valeur de la pépite est calculée à partir de la variance du bruit de l'appareil ayant été utilisé pour relever les observations. Par exemple, pour un GPS grand public d'erreur typique sur l'altitude $\sigma = 5$ m, on choisira une pépite $\eta = \sigma^2 = 25$ m².

3.5.2 Multiplication du variogramme

Propriété 37 (Multiplication du variogramme par un scalaire)

Soient $\alpha \in \mathbb{R}^{+*}$. Considérons γ et $\tilde{\gamma}$ deux variogrammes tels que $\tilde{\gamma}(h) = \alpha\gamma(h)$. Alors, en tout point x du champ \mathcal{D} :

- les estimations de krigeage menées avec ces deux variogrammes coïncident
- les variance d'estimations sont égales à un facteur α près

Preuve. Notons $\hat{z}(x; \gamma) = \sum_{i=1}^n w_i z_i$ l'estimation calculée par krigeage en x , où w_i désigne les poids solution du système de krigeage 3.11, formé avec le variogramme γ . En multipliant les n premières équations de ce système par $\alpha \neq 0$, on ne change pas la solution et on obtient :

$$\alpha\gamma_{i0} - \alpha \sum_{j=1}^n w_j \gamma_{ij} - \alpha\lambda = 0 \Leftrightarrow \tilde{\gamma}_{i0} - \sum_{j=1}^n w_j \tilde{\gamma}_{ij} - \lambda_2 = 0$$

avec la notation : $\lambda_2 = \alpha\lambda$. On retrouve l'équation du système de krigeage pour le variogramme $\tilde{\gamma}$. La solution est donc invariante par multiplication par un scalaire (strictement positif).

Autrement dit, multiplier le variogramme par une constante $\alpha > 0$ ne modifie pas les résultats d'estimation. C'est assez intuitif puisque toutes les relations de dépendance statistique aux autres points sont multipliées par un même facteur, il n'y a donc pas de raison que certaines observations prennent plus de poids que d'autres, et l'estimation finale en est donc invariante. En revanche, l'incertitude autour de cette estimation est multipliée par α :

$$\text{Var}[\hat{z}(x, \tilde{\gamma}) - z(x)] = \sum_{i=1}^n w_i \tilde{\gamma}_{i0} - \lambda_2 = \sum_{i=1}^n w_i \alpha \gamma_{i0} - \alpha \lambda = \alpha \text{Var}[\hat{z}(x, \gamma) - z(x)]$$

On pourra se servir de cette propriété, par exemple si on ne souhaite calculer que l'estimateur (et non son incertitude) et qu'on sait que le variogramme est linéaire, on pourra alors se dispenser de chercher à estimer la pente de ce variogramme. Pour tous les variogrammes à palier (linéaire, exponentiel, gaussien, cubique, sphérique...), la valeur numérique du palier n'a pas d'importance dans ce cas, et il est suffisant de connaître la valeur portée. De même, pour un processus entaché d'un bruit décorrélaté, la valeur numérique de la pépite n'a pas d'importance, seul compte le ratio signal sur bruit, qui parfois est plus simple à estimer.

3.5.3 Linéarité

Soit \mathcal{AL} une application linéaire quelconque, prenant en entrée tout ou partie d'une variable régionalisée et retournant un résultat (sous forme d'un nombre, d'un vecteur ou même d'une variable régionalisée). Alors, le krigeage du résultat de \mathcal{AL} ou l'application de \mathcal{AL} à l'estimation par krigeage produisent exactement les mêmes résultats :

Propriété 38 (Linéarité du krigeage)

Pour toute application linéaire \mathcal{AL} :

$$\mathcal{AL}[\hat{z}] = \widehat{\mathcal{AL}[z]}$$

où \hat{q} dénote la valeur krigée d'une quantité inconnue q .

Cette propriété possède de nombreuses applications pratiques :

- **Sommes** : si X et Y représentent deux phénomènes, alors le krigeage d'observations ponctuelles de la variable $X + Y$ donne le même résultat que la somme des krigeages d'observations de X et Y prises séparément. Il en va de même pour la différence : par exemple, si Z_1 traduit le relief d'un terrain donné au temps t_1 , et Z_2 celui au temps t_2 , alors le krigeage des différences d'altitudes $Z_2 - Z_1$ mesurées en plusieurs points, est égal à la différence des estimations par krigeage de Z_1 et Z_2 .
- **Moyennes** : le krigeage de la moyenne de n phénomènes est égal à la moyenne des krigeages des n phénomènes pris séparément.
- **Dérivation** : la dérivation est une opération linéaire. Le krigeage des pentes du terrain est donc égal à la pente d'un terrain estimé par krigeage. Par ailleurs, on pourra énoncer ici une relation (démontrée dans l'annexe A) liant le variogramme d'un processus avec celui de sa dérivée :

Dérivation d'un processus. Soit X un processus stochastique uni-dimensionnel stationnaire au sens large, dont la covariance C_X est dérivable deux fois en l'origine. Alors, la covariance $C_{X'}$ de la dérivée X' de X vaut :

$$C_{X'}(\tau) = -\frac{d^2 C_X(\tau)}{d\tau^2}$$

On pourra généraliser cette propriété pour des processus définis sur un domaine à plusieurs dimensions, en considérant des dérivées uni-directionnelles.

Au niveau du variogramme on a :

$$\gamma_{X'}(\tau) = C_{X'}(0) - C_{X'}(\tau) = K + \frac{d^2 C_X(\tau)}{d\tau^2} = K - \frac{d^2 \gamma_X(\tau)}{d\tau^2}$$

où K est une constante que l'on peut facilement déterminer avec la contrainte : $\gamma(0) = 0$.

En guise d'exemple d'application, supposons qu'on souhaite estimer la pente du terrain sur une zone donnée dont on ne connaît que des observations de pente⁴. Si par ailleurs on sait grâce à des études antérieures que le MNT sur ce même type de terrain peut-être modélisé par un variogramme gaussien, alors pour modéliser le processus de pente, on pourra s'orienter vers le choix du modèle de variogramme suivant :

$$\gamma(h) = K - \frac{d^2}{dh^2} \left[1 - e^{-\frac{h^2}{a^2}} \right] = 1 - \left(1 - \frac{h^2}{a^2} \right) e^{-\frac{h^2}{a^2}}$$

C'est un variogramme non-monotone, donc à *effet de trou*. Cela traduit l'anti-corrélation des pentes à une distance caractéristique a (fig. 3.10). Par ailleurs, il est intéressant de remarquer que le variogramme dérivé γ a aussi un comportement parabolique à l'origine, ce qui signifie que la pente est elle-même dérivable. Plus généralement, un processus stochastique est dérivable à l'ordre n si et seulement si son variogramme est $2n$ fois dérivable (Allard, 2012). Un processus de variogramme gaussien est donc infiniment dérivable, ce qui explique son aspect très lisse.

- **Intégration** : en reprenant l'exemple des MNT Z_1 et Z_2 à 2 dates successives, on définit l'application \mathcal{AL} comme étant l'intégrale de la différence : $\int Z_2 - Z_1$, où l'intégrale est prise sur le domaine d'étude. Alors le résultat de \mathcal{AL} n'est autre que le volume du terrain perdu (ou gagné) entre les deux dates. L'application \mathcal{AL} étant linéaire, on sait alors que l'estimation de ce volume par intégration du champ krigé des différences mesurées, est correcte.
- **Convolution** : en pratique, les observations effectuées ne sont pas strictement ponctuelles. Par exemple, pour mesurer la teneur d'un élément minéralogique dans le sol en un lieu donné, on doit procéder à l'analyse chimique d'un bloc de terre d'extension spatiale non-nulle. En fonction des techniques employées (et de l'élément recherché), il peut arriver que la taille de bloc

⁴Par exemple pour la surveillance volcanologique, où les mesures sont fournies par un réseau d'inclinomètres.

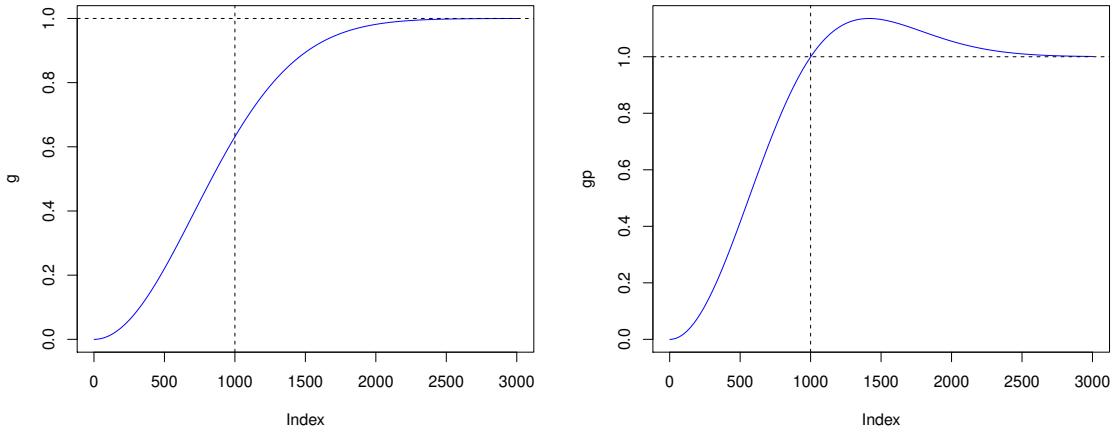


Fig. 3.10. À gauche : variogramme gaussien du terrain (portée $a = 1000$ m et palier arbitrairement fixé à 1.0). À droite : variogramme dérivée seconde de la pente du terrain.

ne puisse plus être négligée devant l'écart typique entre les observations. Dans ce cas, on peut considérer qu'une mesure en un site x est en réalité une moyenne spatiale sur une zone centrée en x , ce qui mathématiquement se traduit par une convolution, et donc une opération à nouveau linéaire. Si les observations sont des convolutions de la variable régionalisée, alors le krigage de ces observations sera égal à la convolution du krigage des mesures ponctuelles. Ici aussi, on peut citer un théorème bien pratique liant les covariances d'un champ avant et après convolution :

Covariance de convolutions. Soit X un processus stochastique stationnaire au sens large, de covariance C_X . On note Y la convolution de X sur le domaine d'étude \mathcal{D} par un noyau h :

$$Y(t) = (X * h)(t) = \int_{\mathcal{D}} X(\tau)h(t - \tau)d\tau$$

Alors, la covariance du champ Y est liée à celle de X par :

$$C_Y(\tau) = C_{X*h}(\tau) = (C_X * h * h)(\tau)$$

Pour un processus 1D et dans le cas particulier où les observations sont des moyennes, h est une fonction porte et l'autoconvolution $h * h = \Pi * \Pi$ est une fonction triangle de support égal au double de la zone sur laquelle la moyenne est calculée. La covariance du champ lissé Y est donc égale au produit de convolution de C_Y par une fonction triangle. Plus les blocs de mesures sont grands, plus la fonction triangle a un support large et plus la covariance de Y est déformée par rapport à celle de X : le champ est lissé. À l'inverse, si la mesure est parfaitement ponctuelle, h est le dirac δ , et la covariance de Y s'exprime par : $C_Y = C_X * \delta * \delta = C_X$ (le dirac est l'élément neutre de la convolution) et on retrouve bien que les deux champs sont de covariances égales.

Ces considérations appartiennent à une théorie plus générale de la Géostatistique, dite **théorie de la régularisation**, dont l'objectif consiste à transférer les propriétés d'un champ aléatoire au processus d'observation de ce champ sur des *blocs* d'extension spatiale non-nulle. Nous en reparlerons dans la section 3.7.3 lorsque nous présenterons le principe de l'estimation globale.

3.5.4 Lissage

Le champ interpolé par krigage est en général plus lisse que la variable régionalisée vraie (et inconnue). De manière formelle :

Propriété 39 (Effet de lissage)

La variance d'une variable krigée \widehat{Z} en tout point donné x_0 du champ est plus faible que celle de la variable aléatoire prise par le processus en ce même point :

$$\text{Var}[\widehat{Z}(x_0)] \leq \text{Var}[Z(x_0)]$$

La démonstration formelle de cette propriété dans le cas du krigage ordinaire n'est pas triviale, mais elle peut être mise en évidence à l'aide d'expérimentations, par exemple ci-dessous pour un processus 1D (fig. 3.11). Sur cette figure, on observe également deux phénomènes : (1) l'interpolation tend à devenir rapidement mauvaise sur les bords du domaine⁵ et (2) l'interpolée est non-convexe (on le voit en particulier autour de l'abscisse 675, où la courbe rouge dépasse le maxima des observations).

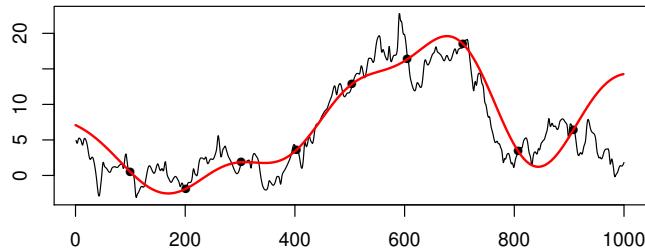


Fig. 3.11. Interpolation par krigage (en rouge) d'un signal aléatoire 1D (trait noir) à partir d'un échantillon de 1% de ses points (points noirs).

Remarquons que cette déformation ne résulte pas d'un manque de connaissance du variogramme, mais d'une propriété intrinsèque au krigage. On représente par exemple ci-dessous en figure 3.12 une expérimentation, où nous avons utilisé un champ (connu de manière exhaustif) dont nous pouvons calculer une estimation relativement fidèle du variogramme γ . Cette estimation γ est alors utilisée comme modèle de base dans une estimation par krigage de l'ensemble du champ à partir d'un sous-échantillon (environ une centaine de points). On observe alors que l'interpolée est plus lisse que le champ réel, et en particulier, on voit que son variogramme possède un comportement à l'origine parabolique, à comparer avec le comportement linéaire à l'origine du variogramme de base.

L'estimateur par krigage est victime d'un phénomène analogue à la non-représentativité de la moyenne. Il est construit de sorte à limiter l'erreur de prédiction. En ce sens, la variable régionalisée estimée est la plus centrale possible, et n'est donc pas une réalisation quelconque parmi tant d'autres. Paradoxalement, cette contrainte d'optimalité empêche l'estimateur d'être pleinement représentatif du phénomène. On peut faire une comparaison assez simple : si une entreprise emploie 9 ouvriers au salaire mensuel de 2000 €, à la tête desquels on place un PDG à 22 000 €, le revenu mensuel

⁵D'une manière générale, hors de l'enveloppe convexe du semis d'observations. On parle alors d'extrapolation.

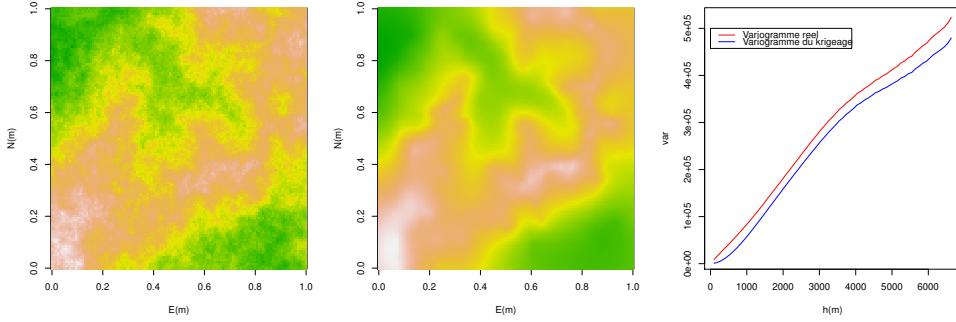


Fig. 3.12. Le champ vrai (à gauche) et son estimation par krigage à partir d'un semis constitué d'une centaine de points d'observation (au centre). À droite : le variogramme du champ vrai (en rouge) et celui du champ interpolé (en bleu).

moyen des salariés s'élève à 4000 €, ce qui n'est absolument pas représentatif, ni des ouvriers, ni du directeur. Pourtant, s'il fallait parier sur le revenu d'un salarié pris au hasard parmi les 10 membres de l'entreprise, c'est bien cette valeur moyenne de 4000 € qui minimiserait la variance de l'erreur.

Autrement dit, dans le cas du krigage, s'il fallait parier sur une valeur en x_0 , de manière à limiter le risque d'erreur quadratique, une décision rationnelle serait de prendre $\hat{z}(x_0)$. Pourtant, l'ensemble de l'interpolée \hat{z} ne constitue en général pas un bon représentant du phénomène. Cette observation justifiera par la suite le recours aux méthodes de simulations.

3.6 Formulation duale

Comme nous l'avons vu dans la section 3.3 ainsi que dans l'activité VIII, le calcul de l'estimation \hat{z} en un site x_0 donné, nécessite de multiplier une matrice \mathbf{A}^{-1} (l'opération d'inversion ayant été effectuée une seule fois en amont du parcours des sites à interpoler) par un vecteur \mathbf{B} (dépendant du modèle de variogramme et de x_0) puis à effectuer le produit scalaire du résultat obtenu par le vecteur $\mathbf{Z} = [z_{1,2}, \dots, z_n, 0]^T$. Dans le formalisme matriciel, l'estimée s'écrit :

$$\hat{z} = \langle \mathbf{A}^{-1}\mathbf{B}, \mathbf{Z} \rangle$$

où $\langle u, v \rangle$ désigne le produit scalaire de deux vecteurs u et v de \mathbb{R}^n . On peut alors écrire :

$$\hat{z} = \langle \mathbf{Z}, \mathbf{A}^{-1}\mathbf{B} \rangle = \mathbf{Z}^T \mathbf{A}^{-1}\mathbf{B} = \mathbf{Z}^T (\mathbf{A}^{-1})^T \mathbf{B} = (\mathbf{A}^{-1}\mathbf{Z})^T \mathbf{B} = \langle \mathbf{A}^{-1}\mathbf{Z}, \mathbf{B} \rangle$$

où la troisième égalité résulte du fait que \mathbf{A} , et donc \mathbf{A}^{-1} également, sont symétriques par construction. Avec cette nouvelle formulation, on peut traiter en amont tout ce qui ne dépend pas du site à interpoler (à savoir l'inverse de \mathbf{A} et le produit par \mathbf{Z}), puis à chaque itération de boucle, il ne reste plus qu'à effectuer le produit scalaire avec le vecteur \mathbf{B} .

On suppose que le domaine à interpoler est assez vaste pour que le temps de calcul en dehors de la boucle de parcours des sites soit négligeable (c'est en pratique le cas sur tous les problèmes concrets d'interpolation). Dans la version de base du krigage, chaque interpolation nécessite une multiplication d'une matrice carrée (de taille n) par un vecteur, soit n^2 multiplications, ce à quoi on ajoute un produit scalaire de n multiplications, soit un algorithme d'interpolation en $\mathcal{O}(n^2)$. Avec la version duale, chaque interpolation ne nécessite plus qu'un produit scalaire, soit $\mathcal{O}(n)$ multiplications.

En contre-partie, le vecteur $\mathbf{X} = [\mathbf{W}, \lambda]$, contenant les poids de la combinaison linéaire ainsi que le multiplicateur de Lagrange, n'étant plus disponible, on ne peut plus calculer la variance de l'estimation avec l'expression 3.13, ce qui limite l'intérêt pratique du krigeage dual.

En partant de l'activité VIII, le code ne nécessite que trois modifications :

- Au niveau du précalcul, on ajoute la multiplication par \mathbf{Z} :

```
# Pre-calcul de l'inversion
I = solve(D) %*% data.matrix(Z)
```

- Au niveau du calcul de l'estimation, suppression de $\mathbf{W} = (\mathbf{I} \%*% \mathbf{D}\theta)$ et modification :

```
# Estimation
zhat[iy, ix] = sum(t(I)*Dtheta)
```

- Et suppression du calcul de la variance de l'estimation :

```
vhat[iy, ix] = sum(W*Dtheta) - W[N+1]
```

À titre d'exemple, le calcul d'un MNT au pas de 25 m à partir des données du fichier `sample_data1.txt` (393 points d'observation) nécessite environ 3 minutes avec le krigeage de base, contre 13 secondes avec le krigeage dual. Pour une zone de taille fixée, le gain de calcul est proportionnel au nombre n de sites.

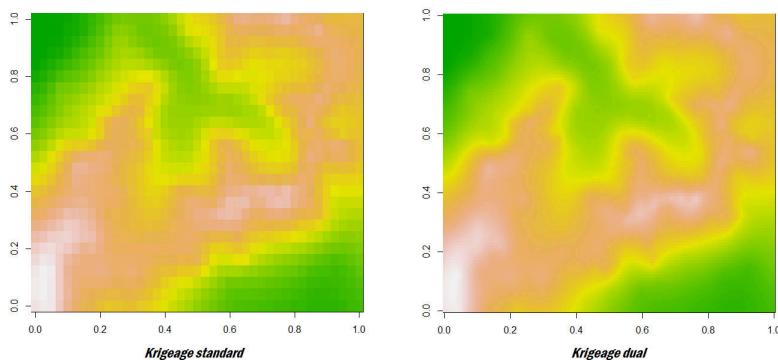


Fig. 3.13. Interpolation d'un MNT pour deux niveaux de résolution différents, à gauche par krigeage standard et à droite par krigeage dual, pour un temps de calcul identique de l'ordre de 1 seconde.

L'intérêt du krigeage dual est toutefois à relativiser au regard des recommandations de la littérature, qui préconise d'utiliser un nombre limité de sites pour chaque interpolation. Par exemple, la méthode des octants, utilisée pour chaque point à interpoler une zone de recherche ellipsoïdale (préférentiellement orientée suivant les anisotropies du champ) divisée en 8 secteurs de 45°. Les 3 points les plus proches sont alors sélectionnés dans chaque secteur, de sorte à limiter la taille du système de krigeage à une matrice de taille $8 \times 3 + 1 = 25$. En contre-partie, l'inversion doit être effectuée à chaque étape.

On pourra trouver plus d'informations sur la recherche de données par octants dans Zoraster (1996). Il s'agit d'un cas particulier de krigeage dit à **voisinage glissant**. Dans le cas contraire, lorsque le krigeage est effectué avec tous les points d'observations, on parle de krigeage à **voisinage unique**.

3.7 Les variantes du krigeage

Nous présentons ici plusieurs résolutions possibles du problème d'interpolation à l'aide de la technique présentée ci-dessus et pour diverses variations dans les hypothèses sous-jacentes. Nous verrons ainsi que le krigeage est un outil générique facilement personnalisable pour des besoins propres.

Précisons un peu la terminologie : on désigne en général par krigeage ordinaire à la fois le krigeage ordinaire et le kriegage intrinsèque puisque les deux méthodes sont identiques. On utilise le terme krigeage intrinsèque lorsqu'on souhaite insister sur le fait que le phénomène considéré est strictement intrinsèque. Par opposition, toutes les méthodes présentées ci-dessous sont des alternatives au krigeage ordinaire.

3.7.1 Krigeage simple

On se place dans les hypothèses du krigeage ordinaire, *i.e.* avec un processus stochastique Z stationnaire au sens large, observé en n sites ponctuels. Supposons que la moyenne $m = \mathbb{E}[Z]$ du processus soit connue, et supposons sans perte de généralité⁶ que m est nulle. La contrainte d'autorisation est nécessairement respectée puisque Z est stationnaire au sens large et donc toute combinaison linéaire est autorisée. Par ailleurs, la contrainte d'universalité impose :

$$\mathbb{E}[\widehat{Z}(x) - Z(x)] = \sum_{i=1}^n w_i \mathbb{E}[Z(x_i)] - \mathbb{E}[Z(x)] = m \sum_{i=1}^n w_i - m = 0$$

La contrainte d'universalité est donc automatiquement respectée (puisque $m = 0$) et le problème du krigeage est alors réduit à un problème d'optimisation sans contrainte, qu'on exprime plus facilement à l'aide de la fonction de covariance $C_{ij} = \gamma(\infty) - \gamma_{ij}$:

$$\begin{bmatrix} C_{11} & C_{12} & \dots & C_{1n} \\ C_{21} & C_{22} & \dots & C_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ C_{n1} & C_{n2} & \dots & C_{nn} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} = \begin{bmatrix} C_{01} \\ C_{02} \\ \vdots \\ C_{0n} \end{bmatrix} \quad (3.16)$$

La matrice de terme général C_{ij} est exactement la matrice de covariances Σ du vecteur aléatoire d'observations \mathbf{Z} . La solution s'exprime alors : $\mathbf{W} = \Sigma^{-1}\mathbf{C}_0$ et on retrouve un équivalent multidimensionnel du filtre de Wiener ([Barret, 2009](#)).

La variance de l'erreur d'estimation est la même que celle du krigeage ordinaire à laquelle on aurait retiré le multiplicateur de Lagrange λ . On peut montrer que $\lambda < 0$, et donc, pour un même jeu d'observations et en un point x_0 donné, la variance d'estimation du krigeage simple (KS) est inférieure à celle du krigeage ordinaire (KO). C'est intuitivement cohérent : la connaissance *a priori* de la moyenne guide le calcul et contribue à réduire l'incertitude finale sur l'estimation.

Pour bien visualiser la différence entre KS et KO, une méthode consiste à représenter le résultat d'une interpolation 1D avec un variogramme de portée inférieure à l'écart typique séparant deux observations (fig. [3.14](#)). En un site x_0 éloigné des observations (typiquement au point médian entre deux échantillons), les données ne sont plus significativement corrélées à $Z(x_0)$ et l'estimateur tend à retourner la valeur moyenne du champ. En revanche, dans les zones où la densité de points d'observation

⁶Si $m \neq 0$, on peut facilement se ramener au cas $m = 0$ en étudiant la variable auxiliaire $Y(x) = Z(x) - m$.

est grande au regard de la portée du variogramme, KS et KO donnent sensiblement les mêmes résultats.

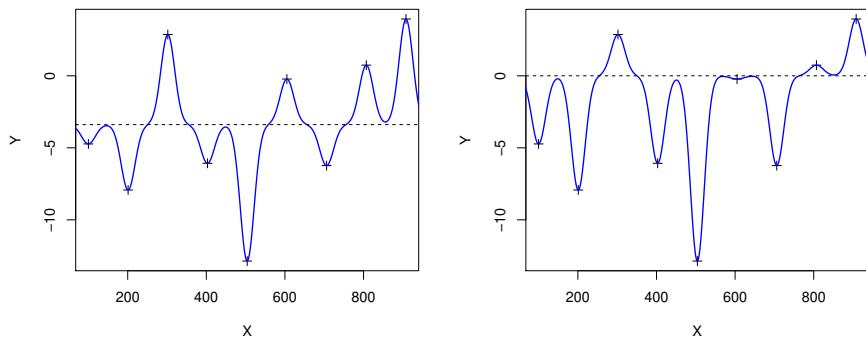


Fig. 3.14. À gauche : krigeage ordinaire (à moyenne inconnue). À droite : krigeage simple à moyenne $m = 0$. Dans les deux cas, le variogramme est un modèle gaussien de portée 45.

En pratique, il est très rare de connaître la moyenne d'un processus et l'emploi du krigeage simple reste relativement peu fréquent.

3.7.2 Krigeage de la moyenne

Si le KO permet d'estimer une variable régionalisée (et au passage sa moyenne) tandis que le KS nécessite en prérequis la connaissance de la moyenne pour estimer la variable, il doit exister une méthode complémentaire qui n'estime que la moyenne de la variable.

Comme pour les versions précédentes, on part des contraintes de krigeage. On cherche donc :

- Un estimateur de la moyenne m (inconnue) du processus sous la forme d'une combinaison linéaire des observations : $\hat{m} = \sum w_i Z_i$.
- On se place dans le cas stationnaire donc toutes les combinaisons linéaires sont autorisées.
- La contrainte d'**universalité** impose :

$$\mathbb{E}\left[\sum_{i=1}^n w_i Z_i - m\right] = \sum_{i=1}^n w_i \mathbb{E}[Z_i] - m = m\left(\sum_{i=1}^n w_i - 1\right) = 0$$

et donc, tout comme pour le krigeage ordinaire, on souhaite que la somme des poids soit unitaire.

- Enfin, l'**optimalité** consiste à minimiser la quantité :

$$\text{Var}[\hat{m} - m] = \text{Var}\left[\sum_{i=1}^n w_i Z_i - m\right]$$

Remarquons que m n'est pas une variable aléatoire mais bien un paramètre du modèle. Elle n'entre donc pas en jeu dans la variance, ce qui donne la simplification :

$$\text{Var}[\hat{m} - m] = \sum_{i=1}^n \sum_{j=1}^n w_i w_j C_{ij}$$

On forme alors le lagrangien du problème d'optimisation :

$$\mathcal{L}(\mathbf{W}, \lambda) = \sum_{i=1}^n \sum_{j=1}^n w_i w_j C_{ij} + 2\lambda \left(\sum_{i=1}^n w_i - 1 \right)$$

La résolution du problème avec les techniques présentées dans la section 3.3 donne le système :

$$\begin{bmatrix} C_{11} & C_{12} & \dots & C_{1n} & 1 \\ C_{21} & C_{22} & \dots & C_{2n} & 1 \\ \vdots & \vdots & \vdots & \vdots & 1 \\ C_{n1} & C_{n2} & \dots & C_{nn} & 1 \\ 1 & 1 & \dots & 1 & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \\ \lambda \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \quad (3.17)$$

La variance associée vaut :

$$\text{Var}[\hat{m} - m] = \text{Var}[\hat{m}] = \sum_{i=1}^n \sum_{j=1}^n w_i w_j C_{ij} = \sum_{i=1}^n w_i \left(\sum_{j=1}^n w_j C_{ij} \right)$$

Or pour tout indice i la somme $\sum w_j C_{ij}$ vaut $-\lambda$ d'après le système 3.17. Donc :

$$\text{Var}[\hat{m} - m] = -\lambda \sum_{i=1}^n w_i = -\lambda$$

Nécessairement $\lambda \leq 0$ (sinon, la variance perd son sens physique).

En notant KM le krigeage de la moyenne, on peut dire informellement que :

$$KO = KM + KS \quad (3.18)$$

Autrement dit, tant du point de vue de l'estimateur que de sa variance d'erreur, appliquer un krigeage ordinaire revient à : (1) estimer la moyenne \hat{m} du processus par le krigeage de la moyenne puis (2) à fixer $m = \hat{m}$ dans un krigeage simple⁷. La différence de variance entre KS et KO vaut le multiplicateur de Lagrange, qui correspond donc à la part de variance expliquée par la connaissance *a priori* de la moyenne du processus.

⁷En fait il s'agit d'un cas particulier d'application de la propriété 38 de linéarité du krigeage.

3.7.3 Krigeage par blocs

Nous avons vu comment estimer la moyenne m d'un processus stochastique stationnaire au sens large dans la section précédente. Rappelons que m est un paramètre du modèle, sa valeur vraie inconnue (si tant est qu'on veuille bien lui donner un sens) ne doit en théorie pas dépendre de la réalisation à étudier. Pour rappatrier ce concept dans le monde physique, on pourrait dire que m serait la valeur moyennée du phénomène si le domaine d'étude était de taille infinie. Or, ce qui nous intéresse en général, c'est la valeur moyenne prise par le phénomène sur la réalisation courante, éventuellement restreinte à un sous-domaine V .

On peut donner l'exemple d'application suivant : un prospecteur minier a fait réaliser des prélèvements⁸ de la teneur en cuivre du sol. Disposant d'un temps limité, il ne peut exploiter qu'une unique carrière d'emprise spatiale V délimitée a priori par le matériel et la main d'oeuvre à disposition. Dès lors, la question se pose du choix du site d'implantation. En général, le nombre de candidats est limité par des facteurs externes (fleuves, pentes, commodité d'accès...) et le prospecteur doit choisir le bloc qui lui permettra statistiquement d'extraire le plus de cuivre.

Une solution intuitive pourrait consister à subdiviser chaque bloc à l'aide d'une grille régulière, à appliquer le krigeage ordinaire en chaque noeud de cette grille, puis à sommer les estimations obtenues pour comparer les résultats. La propriété 38 de linéarité du krigeage nous assure la validité théorique de cette solution. Cependant, en plus de nécessiter potentiellement un temps de calcul considérable, cette démarche ne permet pas d'évaluer la variance d'estimation sur chaque bloc. Nous allons voir ici une manière de calculer la valeur moyenne (et sa variance) d'un bloc en une seule estimation.

Pour ce faire, nous devons d'abord introduire la notion de **covariance régularisée**. Soit $V \subseteq \mathcal{D}$ un sous-domaine du champ d'étude de forme quelconque (V peut même allègrement être non-connexe). La moyenne de la variable régionalisée z sur V s'exprime par :

$$z(V) = \frac{1}{|V|} \int_V z(x) dx \quad (3.19)$$

où $|V|$ désigne l'extension spatiale de V (longueur, aire, volume, etc. suivant la dimension de l'espace de travail) et dx est une parcelle infinitésimale du domaine V .

En remplaçant les minuscules par des majuscules, la quantité 3.19 devient une variable aléatoire et on peut calculer sa variance. Le théorème 12 de propagation des incertitudes nous indique comment calculer la variance d'une somme de variables corrélées. On peut proposer une expression analogue pour la variance d'une intégrale :

$$\text{Var}[Z(V)] = \frac{1}{|V|^2} \iint_{x,y \in V} C(x-y) dx dy \quad (3.20)$$

La démonstration s'établit facilement en utilisant la bilinéarité de l'opérateur de covariance, puis en transformant l'intégrale en une somme de Riemann et enfin par passage à la limite. Concrètement, la variance d'une moyenne sur une domaine est la moyenne des covariances entre tous les couples de points possibles du domaine.

⁸On suppose que ces prélèvements sont ponctuels, ou au moins d'extension spatiale négligeable devant les autres grandeurs entrant en jeu dans le problème : taille du domaine, taille du bloc, portée du variogramme...

Remarquons par ailleurs que pour un processus strictement intrinsèque, on peut réexprimer 3.20 uniquement à l'aide du variogramme, et on retrouve une expression analogue à 3.14.

Plus généralement, on peut définir l'opérateur de covariance entre deux sous-domaines :

Définition 40 (Covariance régularisée)

Soient V_1 et V_2 deux sous-domaines de \mathcal{D} potentiellement d'intersection non-vide. On définit la covariance régularisée \bar{C} de Z par la covariance des moyennes des valeurs prises par Z en V_1 et V_2 :

$$\bar{C}(V_1, V_2) = \text{Cov}(Z(V_1), Z(V_2)) = \frac{1}{|V_1||V_2|} \int_{V_1} \int_{V_2} C(x - y) dx dy$$

On retrouve bien entendu $\bar{C}(V, V) = \text{Var}[Z(V)]$.

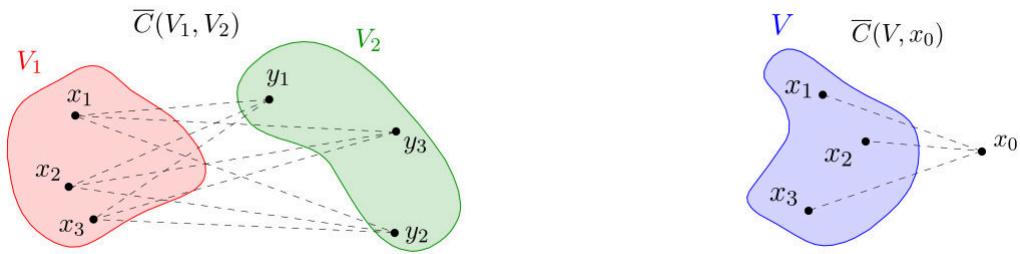


Fig. 3.15. La covariance régularisée \bar{C} est la moyenne des covariances entre tous les couples possibles de valeurs (x, y) avec $x \in V_1$ et $y \in V_2$. Le concept reste valide pour des ensembles discrets, par exemple à droite, dans le cas de la covariance régularisée entre un ensemble V et le singleton $\{x_0\}$.

Remarque 1. Si V_1 est un ensemble fini, on obtient une expression mêlant somme et intégrale :

$$\bar{C}(V_1, V_2) = \text{Cov}\left(\frac{1}{|V_1|} \sum_{i=1}^{|V_1|} Z(x_i), \frac{1}{|V_2|} \int_{V_2} Z(x) dx\right) = \frac{1}{|V_1||V_2|} \sum_{i=1}^{|V_1|} \int_{V_2} C(x - x_i) dx$$

où $|V_1|$ est le cardinal de l'ensemble fini de points $x_i \in V_1$. Lorsque V_1 et V_2 sont tous les deux discrets, on retrouve les expressions classiques de la covariance entre sommes de variables aléatoires.

Remarque 2. La notion de covariance régularisée se généralise au variogramme afin de traiter le cas des processus strictement intrinsèques : $\bar{\gamma}(V_1, V_2) = C(0) - \bar{C}(V_1, V_2)$.

Lorsque les deux blocs ont la même forme, par analogie avec le covariogramme, on peut définir le covariogramme régularisé \bar{C}_h qui mesure la ressemblance des valeurs moyennes prises par deux blocs identiques V séparés d'un vecteur de translation h .

Définition 41 (Covariogramme régularisée)

Soient V un sous-domaine de \mathcal{D} . On définit le *covariogramme régularisé* \bar{C}_V de Z par la covariance des valeurs moyennes prise par Z en V et V_h :

$$\bar{C}_V(h) = \bar{C}(V, V_h) \quad \text{avec : } V_h = \{x + h \mid x \in V\}$$

À nouveau, ici aussi, on peut définir le variogramme régularisé $\bar{\gamma}_V(h) = \bar{C}(V, V) - \bar{C}_V(h)$. De plus si Z est isotrope, \bar{C}_V et $\bar{\gamma}_V$ ne dépendent que de la norme du vecteur h .

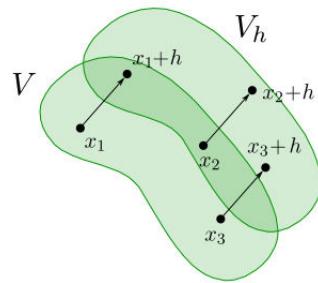


Fig. 3.16. Covariogramme régularisé \bar{C}_h d'un domaine V .

D'un point de vue graphique, la régularisation mène à aplatiser le variogramme au voisinage de l'origine. Pour des distances de séparation grande devant la taille caractéristique du bloc V , on retrouve le comportement du variogramme de Z avec un décalage qui peut s'interpréter de deux manières : (1) horizontalement, comme un allongement de l'ordre de $|V|$ de la portée du variogramme régularisé ou (2) verticalement comme une perte de variance dû au fait de moyenner sur des blocs.

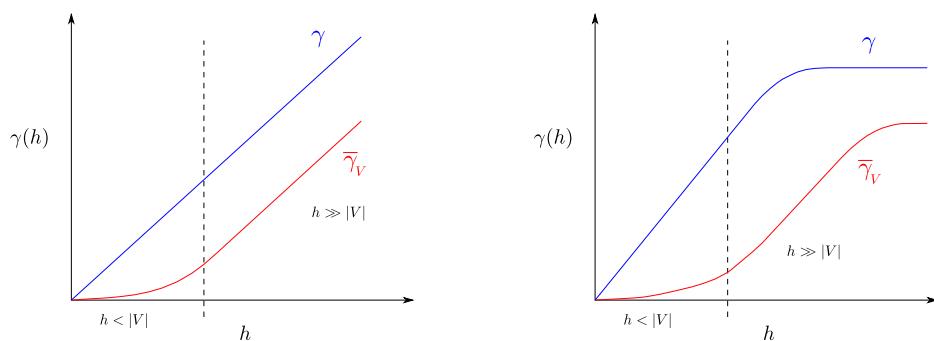


Fig. 3.17. Le variogramme du processus (en bleu) et sa régularisée (en rouge) dans le cas strictement intrinsèque (à gauche) et stationnaire au second ordre (à droite). La ligne verticale en pointillés représente la dimension $|V|$ du bloc.

Remarque 3. Le calcul formel des régularisées \bar{C}_V et $\bar{\gamma}_V$ est en général assez compliqué. Un exemple sur un cas simple est donné dans l'annexe A.3. On pourra avoir recours si besoin à un langage de calcul scientifique (R, Matlab...).

On peut alors construire un estimateur de la moyenne $Z(V)$ en utilisant le variogramme régularisé.

Plaçons-nous en premier lieu dans le cadre stationnaire, et examinons les contraintes de krigeage :

- **Linéarité** : on estime $z(V)$ par combinaison linéaire des sites observés : $\hat{z}(V) = \sum_{i=1}^n w_i z(x_i)$
- **Autorisation** : toutes les combinaisons linéaires sont autorisées
- **Universalité** :

$$\begin{aligned}\mathbb{E}[\hat{Z}(V) - Z(V)] &= \mathbb{E}\left[\sum_{i=1}^n w_i Z(x_i)\right] - \mathbb{E}\left[\frac{1}{|V|} \int_V Z(x) dx\right] = \sum_{i=1}^n w_i \mathbb{E}[Z] - \frac{1}{|V|} \int_V \mathbb{E}[Z(x)] dx \\ &= m\left(\sum_{i=1}^n w_i - \frac{1}{|V|} \int_V dx\right) = m\left(\sum_{i=1}^n w_i - 1\right)\end{aligned}$$

Donc seules les combinaisons de poids unitaires respectent la contrainte d'universalité.

- **Optimalité** : calculons la variance de l'erreur avec la notion de covariance régularisée :

$$\begin{aligned}\text{Var}[\hat{Z}(V) - Z(V)] &= \text{Var}[\hat{Z}(V)] + \text{Var}[Z(V)] - 2\text{Cov}(\hat{Z}(V), Z(V)) \\ &= \sum_{i=1}^n \sum_{j=1}^n w_i w_j C(x_i - x_j) + \bar{C}(V, V) - 2 \sum_{i=1}^n w_i \bar{C}(x_i, V)\end{aligned}$$

En formant le Lagrangien et en annulant toutes ses dérivées partielles (par rapport à w_i et λ), on trouve facilement le système de krigeage :

$$\begin{bmatrix} C(0) & C(x_1 - x_2) & \dots & C(x_1 - x_n) & 1 \\ C(x_2 - x_1) & C(0) & \dots & C(x_2 - x_n) & 1 \\ \vdots & \vdots & \vdots & \vdots & 1 \\ C(x_n - x_1) & C(x_n - x_2) & \dots & C(x_n - x_n) & 1 \\ 1 & 1 & \dots & 1 & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \\ \lambda \end{bmatrix} = \begin{bmatrix} \bar{C}(x_1, V) \\ \bar{C}(x_2, V) \\ \vdots \\ \bar{C}(x_n, V) \\ 1 \end{bmatrix} \quad (3.21)$$

On montre qu'un système identique peut-être formé avec γ et $\bar{\gamma}$ dans le cadre intrinsèque :

$$\begin{bmatrix} \gamma(0) & \gamma(x_1 - x_2) & \dots & \gamma(x_1 - x_n) & 1 \\ \gamma(x_2 - x_1) & \gamma(0) & \dots & \gamma(x_2 - x_n) & 1 \\ \vdots & \vdots & \vdots & \vdots & 1 \\ \gamma(x_n - x_1) & \gamma(x_n - x_2) & \dots & \gamma(x_n - x_n) & 1 \\ 1 & 1 & \dots & 1 & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \\ \lambda \end{bmatrix} = \begin{bmatrix} \bar{\gamma}(x_1, V) \\ \bar{\gamma}(x_2, V) \\ \vdots \\ \bar{\gamma}(x_n, V) \\ 1 \end{bmatrix} \quad (3.22)$$

La variance d'estimation vaut alors : $\text{Var}[\hat{Z}(V) - Z(V)] = \sum_{i=1}^n \bar{\gamma}(x_i, V) - \bar{\gamma}(V, V) - \lambda$

Si V est ponctuel, on retrouve les équations 3.11 et 3.13 du krigage ordinaire. Par ailleurs, on remarque que l'estimateur et sa variance ne font appel qu'au variogramme, ainsi qu'à sa régularisée entre un point x_i et un bloc. Le variogramme régularisé *bloc-à-bloc* $\bar{\gamma}(V_1, V_2)$ avec $V_1 \neq V_2$ (en général plus difficile à calculer) n'est pas nécessaire.

Cette extension laisse entrevoir les inombrables possibilités de généralisation du krigage. Ici, nous avons collecté des valeurs ponctuelles pour estimer une grandeur moyennée sur un bloc. Certains problèmes peuvent nécessiter à l'inverse d'estimer une grandeur ponctuelle à partir d'informations zonales. À titre d'exercice, on pourra s'interroger sur les deux applications suivantes. Pour chaque cas, on détaillera la système de krigage correspondant en suivant les 4 contraintes usuelles. On commencera par se placer dans le cas stationnaire au sens large, avant de généraliser au cas intrinsèque.

Exercice 3.1. Dans une région réputée riche en minerais, un terrain V est à vendre. Aucun prélèvement n'a été effectué par le passé sur V , si bien qu'un acheteur potentiel n'a aucune idée a priori de la rentabilité de son investissement. En revanche, on sait que n terrains V_1, V_2, \dots, V_n , de formes et de tailles diverses ont été exploitées par des concurrents, et on connaît pour chacun d'eux, la quantité totale de minerais qui en a été extraite. Par ailleurs, on suppose connaître le variogramme de la teneur minérale sur la région (par exemple après analyse d'une région voisine). Estimer la quantité de minerais contenue sur V , ainsi que la variance de l'erreur d'estimation.

Exercice 3.2. On suppose que la quantité estimée sur V a conduit le prospecteur à acheter le terrain. Pour rentabiliser rapidement son investissement, il souhaite commencer par exploiter la zone la plus prometteuse, mais sans perdre de temps à effectuer des sondages au préalable. Comment peut-il utiliser les valeurs de teneurs moyennes des terrains voisins pour estimer le point de V possédant la plus forte densité de minerais ? Nous reparlerons d'un problème similaire dans la section 4.2 traitant de l'optimisation bayésienne.

3.7.4 Krigage universel

Dans certains cas, le phénomène à étudier présente une tendance significative qui ne peut plus être considérée comme une manifestation aléatoire. Un exemple typique peut être trouvé en météorologie, si on cherche à interpoler des valeurs de températures (dans un volume d'atmosphère). On sait alors qu'à une fluctuation spatiale aléatoire des températures (dépendant essentiellement des conditions météorologiques du moment) s'ajoute une dérive déterministe de -6° par tranche de 1000 m d'altitude au dessus du sol.

L'hypothèse stationnaire n'est alors plus satisfaisante et on décompose le processus Z sous la forme d'une somme de deux fonctions :

$$Z(x) = Y(x) + m(x) \quad (3.23)$$

où Z est un processus stationnaire tel qu'étudié jusqu'à présent, et m est une fonction déterministe traduisant l'espérance du processus en un site x . Cependant, la covariance n'étant pas sensible à la moyenne, on vérifie aisément que la fonction de covariance de Z ne dépend que de la distance entre les sites. On modélise alors la tendance déterministe sous forme d'une somme de fonctions de base :

$$m(x) = \sum_{p=1}^l a_p f^p(x) \quad (3.24)$$

où les a_p sont des coefficients réels et où chaque fonction de base f_p ne dépend que d'une seule coordonnée. En règle générale, on prend des fonctions classiques comme des exponentielles ou des monômes x, x^2, x^3, \dots et on fixe par convention la première fonction de base $f_1(x) = 1$.

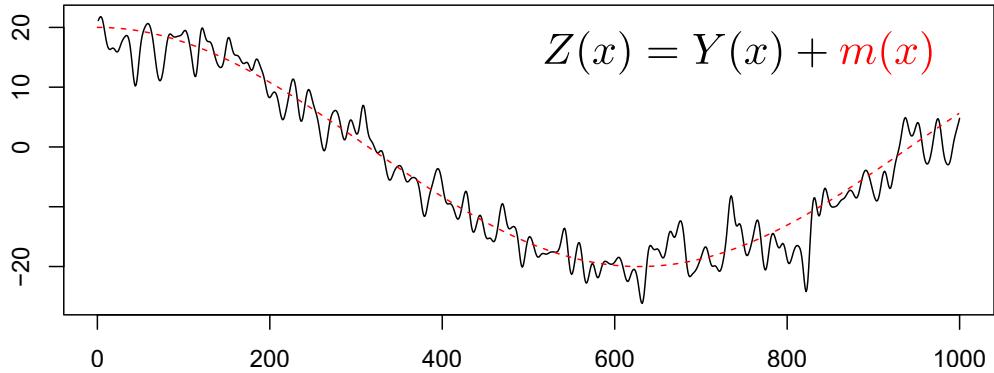


Fig. 3.18. Un processus stochastique non-stationnaire Z (en noir), modélisé comme somme d'un processus stationnaire Y et d'une tendance déterministe m (en rouge).

Le krigeage universel consiste alors à estimer simultanément la tendance m et la fluctuation aléatoire Y en x_0 . La résolution du problème s'effectue de manière similaire aux sections précédentes (en pratique, seule la contrainte d'universalité est fondamentalement différente, d'où le nom de cette variante du krigeage) et on obtient un système de $n + l$ inconnues :

$$\begin{bmatrix} C(0) & C(x_1 - x_2) & \dots & C(x_1 - x_n) & f^1(x_1) & \dots & f^l(x_1) \\ C(x_2 - x_1) & C(0) & \dots & C(x_2 - x_n) & f^1(x_2) & \dots & f^l(x_2) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ C(x_n - x_1) & C(x_n - x_2) & \dots & C(x_n - x_n) & f^1(x_n) & \dots & f^l(x_n) \\ f^1(x_1) & f^1(x_2) & \dots & f^1(x_n) & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ f^l(x_1) & f^l(x_2) & \dots & f^l(x_n) & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \\ \lambda_1 \\ \vdots \\ \lambda_l \end{bmatrix} = \begin{bmatrix} \bar{\gamma}(x_1 - x_0) \\ \bar{\gamma}(x_2 - x_0) \\ \vdots \\ \bar{\gamma}(x_n - x_0) \\ f^1(x_0) \\ \vdots \\ f^l(x_0) \end{bmatrix} \quad (3.25)$$

qui se généralise sans problème au cas intrinsèque en remplaçant comme à l'accoutumée C par $-\gamma$.

La variance de krigeage s'écrit :

$$\text{Var}[\hat{Z} - Z] = C(0) - \sum_{i=1}^n w_i C(x_i - x_0) - \sum_{p=1}^l \lambda_p f^p(x_0) \quad (3.26)$$

On peut donc considérer le krigeage universel comme la combinaison d'une régression linéaire et d'un krigeage ordinaire, permettant ainsi une estimation conjointe (et donc de qualité supérieure) des poids de krigeage et des paramètres du modèle de tendance (même si les coefficients a_p ne sont pas explicitement retournés par la méthode). En particulier, en l'absence de tendance, seule la fonction de base f^1 subsiste dans le modèle et les équations 3.25 et 3.26 se réduisent au krigeage ordinaire.

Prenons un exemple dans un espace à deux dimensions, avec $\mathbf{x}_0 = (x_0, y_0) \in \mathbb{R}^2$ un point à interpoler à partir d'un semis d'observations $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$. On suppose que la tendance est linéaire, *i.e.* la moyenne

du processus sur la zone est un plan incliné. On a alors 3 fonctions de base $f^1(x, y) = 1$, $f^2(x, y) = x$ et $f^3(x, y) = y$. En notant $h_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|_2$ où $\|\cdot\|_2$ est la distance euclidienne classique de \mathbb{R}^2 , le système de krigeage s'écrit :

$$\begin{bmatrix} C(0) & C(h) & \dots & C(h_{1n}) & 1 & x_1 & y_1 \\ C(h_{21}) & C(0) & \dots & C(h_{2n}) & 1 & x_2 & y_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \dots & \vdots \\ C(h_{n1}) & C(h_{n2}) & \dots & C(h_{nn}) & 1 & x_n & y_n \\ 1 & 1 & \dots & 1 & 0 & \dots & 0 \\ x_1 & x_2 & \vdots & x_n & 0 & \dots & \vdots \\ y_1 & y_2 & \vdots & y_n & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \\ \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix} = \begin{bmatrix} \bar{\gamma}(h_{10}) \\ \bar{\gamma}(h_{20}) \\ \vdots \\ \bar{\gamma}(h_{30}) \\ 1 \\ x_0 \\ y_0 \end{bmatrix} \quad (3.27)$$

En pratique, se pose toutefois la question de l'inférence du variogramme, puisque la tendance structurelle vient s'ajouter à l'estimateur. En effet :

$$\begin{aligned} \hat{\gamma}(h) &= \frac{1}{2} \text{Var}[Z_{x+h} - Z_x] = \frac{1}{2} \mathbb{E}[(Z_{x+h} - Z_x)^2] = \frac{1}{2} \mathbb{E}[(Y_{x+h} + m_{x+h} - Y_x - m_x)^2] \\ &= \frac{1}{2} \mathbb{E}[(Y_{x+h} - Y_x)^2] + \frac{1}{2} \mathbb{E}[(m_{x+h} - m_x)^2] + \mathbb{E}[Y_{x+h}m_{x+h} - Y_{x+h}m_x - Y_xm_{x+h} + Y_xm_x] \\ &= \gamma(h) + \frac{1}{2} (m(x+h) - m(x))^2 \end{aligned}$$

Le variogramme est donc biaisé d'une quantité fonction de la différence entre la valeur moyenne prise par le processus en x et $x + h$. Lorsque la tendance varie lentement à l'échelle de la portée du variogramme, on a $m(x + h) \approx m(x)$ et l'estimateur $\hat{\gamma}$ peut être satisfaisant. Sinon, il faudra recourir à d'autres stratégies. Nous en exposons quatre brièvement.

- Il y a des cas de figures où on sait a priori que la tendance est invariante par translation suivant une des coordonnées du repère. Notamment, [Emery \(2001\)](#) mentionne le cas de la bathymétrie sur le littoral. On sait que la tendance induit une augmentation de la profondeur moyenne à mesure que le sondeur s'éloigne de la côte, mais on peut supposer que les coupes de terrain parallèles à la ligne de côte ne sont globalement pas soumise à cette tendance. On peut donc estimer le variogramme uniquement suivant cette direction (qui peut être considérée simon stationnaire, au moins intrinsèque), puis supposer le phénomène isotrope.
- On peut rechercher une zone sur laquelle on sait que la tendance est de valeur relativement constante pour y estimer le variogramme à l'aide des méthodes classiques. Remarquons que si la zone est une coupe à coordonnée constante dans le repère, on se ramène à la première option.
- Une méthode itérative peut consister à calculer une première estimation \hat{m} de la tendance à l'aide de méthodes déterministes classiques, souvent à base de moindres carrés (régression polynomiales, splines de lissage...), puis à calculer le variogramme de la variable résidu $Z - \hat{m}$. On opère alors un krigeage universel sur la variable Z ([Bourgine, 1995](#)).
- Enfin, une dernière option, qui est sans doute la plus satisfaisante d'un point de vue théorique, consiste à se placer dans le cadre de l'hypothèse intrinsèque d'ordre k , et de ne considérer que des

accroissements insensibles à la tendance (quelque soit les valeurs prises par les paramètres a_p). Remarquons qu'il s'agit bien d'une généralisation de l'hypothèse intrinsèque qui elle ne travaille que sur des combinaisons linéaires de poids nul, et est donc insensible à la valeur moyenne. En retour, cet élargissement du champ théorique offre un choix plus vaste de fonctions de covariance (appelées covariances généralisées) dont les covariances de l'hypothèse stationnaire et les variogrammes de l'hypothèse intrinsèque ne sont que des cas particuliers. On pourra trouver plus de détails sur ce vaste sujet dans [Droesbeke et al. \(2006\)](#).

3.7.5 Krigeage avec dérive externe

Parfois, la tendance de la variable régionalisée est connue *a priori*, ou bien de manière analytique (résultat de la théorie générale ou d'études spécifiques antérieures), ou alors numériquement, par exemple si la variable d'intérêt fluctue autour d'une valeur dépendant linéairement d'un autre variable φ (la dérive externe). Pour que le krigeage à dérive externe (KDE) soit opérant, la fonction φ doit être connue sur une grille de résolution beaucoup plus fine que celle du semis d'observations, de sorte qu'il soit possible de connaître avec un bon niveau de précision la valeur prise par φ en tout point du domaine (et en particulier au niveau des sites d'observations et des sites à interpoler). Remarquons que contrairement au cas du krigeage universel, ici la tendance est parfaitement connue et calculable numériquement. On modélise alors le processus sous la forme :

$$Z(x) = Y(x) + a\varphi(x) + b$$

où Y est un processus stationnaire de moyenne nulle, $a, b \in \mathbb{R}$ sont des coefficients **inconnus** et φ est une fonction déterministe **connue en tout point du domaine**.

La connaissance de φ résulte en général de l'observation d'une variable annexe. [Bourgine \(1995\)](#) donne en particulier un exemple de cas d'application : en géothermie, la température $T(x)$ s'exprime par $Y(x) + aP(x) + b$ où P est la profondeur d'un site x . D'une manière générale, le KDE peut être utilisé pour calculer les fluctuations d'un variable autour d'un niveau de référence déterministe.

Nous laissons au lecteur le soin de vérifier que l'écriture des contraintes de krigeage produit un système d'équations identique à celui du krigeage ordinaire, avec un ligne supplémentaire pour assurer l'absence de biais, soit, en notant $\varphi_i = \varphi(x_i)$:

$$\begin{bmatrix} C_{11} & C_{12} & \dots & C_{1n} & 1 & \varphi_1 \\ C_{21} & C_{22} & \dots & C_{2n} & 1 & \varphi_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ C_{n1} & C_{n2} & \dots & C_{nn} & 1 & \varphi_n \\ 1 & 1 & \dots & 1 & 0 & 0 \\ \varphi_1 & \varphi_2 & \dots & \varphi_n & 0 & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \\ \lambda_1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} C_{01} \\ C_{02} \\ \vdots \\ C_{0n} \\ 1 \\ \varphi_0 \end{bmatrix} \quad (3.28)$$

Dans le cas où le modèle sous-jacent Y est intrinsèque C est remplacé par $-\gamma$.

Le KDE peut être particulièrement intéressant pour modéliser les discontinuités (à condition qu'elles soient connues *a priori*), comme illustré sur la figure [3.19](#).

Nous présentons ci-dessous un cas d'application concret issu des travaux de [Peraudin \(2003\)](#), dont l'objet était d'étudier la concentration spatiale du benzène (C_6H_6) sur la ville de Rouen, à partir de

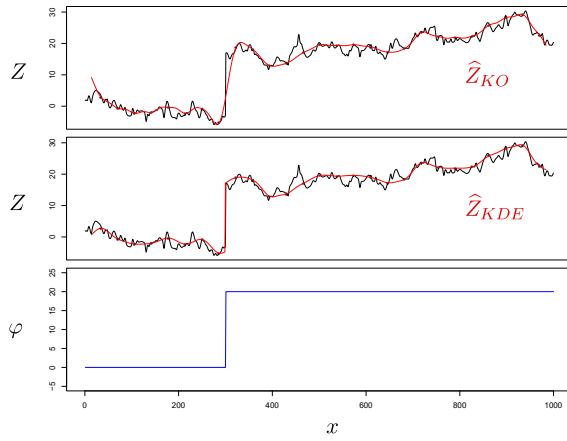


Fig. 3.19. Comparaison du résultat du krigeage ordinaire \hat{Z}_{KO} et du krigeage avec dérive externe \hat{Z}_{KDE} pour un processus 1D stationnaire avec dérive $m(x) = 20 \times \mathbb{1}_{x \geq 300}$. L'échantillonnage du semis d'observations a été opéré à une résolution de 1 point toutes les 35 unités.

l'observation de 73 sites. L'estimation a été effectuée par krigeage avec dérive externe, en considérant une variable annexe, décrivant la densité de population et l'altitude du terrain, deux facteurs qui peuvent avoir une influence indirecte sur la concentration de benzène. Le modèle est alors évalué objectivement par validation croisée (*cf* chapitre 4) et comparé aux estimations obtenus par un krigeage ordinaire. Le résultat de l'étude met en évidence une meilleure précision d'estimation avec le KDE.

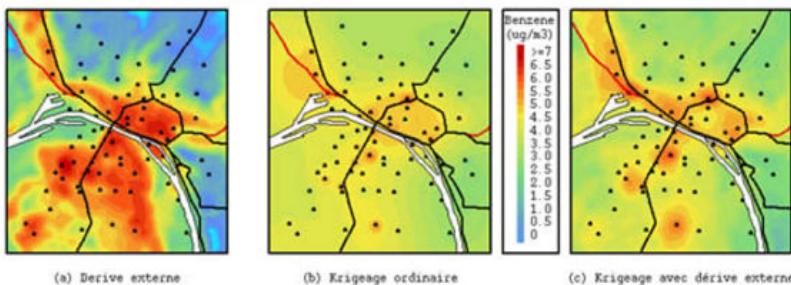


Fig. 3.20. Estimation de la concentration de C_6H_6 par KO (au centre) et par KDE (à droite) tenant compte de la densité de population et de la topographie (à gauche). Source : [Peraudin \(2003\)](#).

3.7.6 Autres variantes

Il existe de nombreuses autres variantes et extensions du krigeage, que nous ne pouvons malheureusement pas détailler ici. Nous citons cependant brièvement les plus utilisées en pratique :

- **Krigeage aléatoire** : lorsque les positions des sites sont elles-mêmes significativement bruitées⁹, on peut avoir recours au krigeage aléatoire qui suppose que les sites x_i sont aléatoirement et indépendamment distribués dans un voisinage local V_i . La résolution, qui s'opère classiquement à partir des 4 contraintes usuelles, permet d'aboutir à un système d'équations faisant intervenir le variogramme régularisé $\bar{\gamma}$ entre les voisinages (*cf* section 3.7.3 sur le krigeage par blocs). Notons que ce bruit de positionnement impacte également l'estimation du variogramme, et fait en général apparaître une pépite artificielle ([Emery, 2001](#)).

⁹Ce qui arrive souvent en bathymétrie par exemple comme le mentionne [Emery \(2001\)](#).

- **Co-krigeage :** nous avons vu comment utiliser le krigeage à dérive externe lorsque la tendance globale de la variable est perturbée par une variable annexe. Toutefois l'utilisation du KDE en nécessite la connaissance exhaustive (ou moins sur une grille de résolution suffisamment fine). Lorsque cette variable annexe n'est connue qu'au niveau de quelques sites d'observations, la géostatistique multivariée ([Rivoirard, 2003](#)) permet de traiter le problème en estimant les quantités inconnues à l'aide de combinaisons linéaires mixtes de la variable d'intérêt Z et de la variable annexe Y :

$$\hat{z}(x_0) = \sum_{i=1}^m v_i y_i + \sum_{i=1}^n w_i z_i \quad (3.29)$$

où m et n désignent respectivement les effectifs des observations des champs Y et Z . Moyennant la connaissance des variogrammes γ_Z et γ_Y ainsi que d'un variogramme croisé γ_{YZ} traduisant les influences mutuelles des deux variables, on peut déterminer la solution $\hat{z}(x_0)$ non-biaisée et de variance minimale. Dans le cas du co-krigeage simple et stationnaire, le vecteur des poids inconnus : $X = [\mathbf{v}, \mathbf{w}]^T = [v_1, v_2, \dots, v_n, w_1, w_2, \dots, w_n]^T$ est solution de l'équation matricielle :

$$\begin{bmatrix} C_Y & C_{YZ} \\ C_{YZ} & C_Z \end{bmatrix} \begin{bmatrix} \mathbf{v} \\ \mathbf{w} \end{bmatrix} = \begin{bmatrix} C_Y^0 \\ C_X^0 \end{bmatrix} \quad (3.30)$$

où $C_{AB}(h) = \gamma_{AB}(\infty) - \gamma_{AB}(h)$ dénote la covariance entre les deux variables A et B en deux sites x_i et x_j séparés d'une distance h , $C_A = C_{AA}$ et C_A^0 est la covariance de la variable A entre un site x_i et le site à interpoler x_0 . Dans un cas défavorable, γ_{YZ} a un comportement pépitique pur et la connaissance de la variable auxiliaire n'apporte rien à la résolution du problème. L'équation 3.30 se généralise pour le cas du cokrigeage ordinaire (avec deux multiplicateurs de Lagrange pour assurer la contrainte d'universalité) ainsi que dans le cas strictement intrinsèque (avec l'emploi exclusif du variogramme).

Le co-krigeage se généralise facilement au cas de p variables (une variable d'intérêt Z_1 et $p - 1$ variables annexes Z_2, Z_3, \dots, Z_p). L'estimateur 3.29 devient alors :

$$\hat{z}(x_0) = \sum_{i=1}^n \sum_{j=1}^p w_{ij} z_{ij} \quad (3.31)$$

où w_{ij} est le poids affecté à la i -eme observation de la variable Z_j , en général d'autant plus fort et que les sites x_0 et x_i sont proches, et que les variables Z_1 et Z_j sont corrélées. Remarquons que les sites ne coïncident pas nécessairement entre les différentes variables. Le cokrigeage constitue un outil de choix pour l'interpolation des champs vectoriels (avec des résultats en général plus précis qu'un découpage en n champs scalaires). On le rencontre fréquemment dans les études liées à la pollution atmosphérique, qui font souvent intervenir un grand nombre de variables corrélées telles que les concentrations d'une large gamme de particules et des paramètres météorologiques ([Deraisme et Bobbia, 2003](#)). Notons toutefois que le nombre de variogrammes à modéliser augmente quadratiquement avec p , ce qui limite en général le co-krigeage à un maximum pratique d'une dizaine de variables annexes).

Pour illustrer cette méthode, nous nous appuyons à nouveau sur les travaux de Peraudin (2003), qui a étudié la concentration de dioxyde d'azote (NO_2) sur la région Ile-de-France à partir des relevés fournis par 20 capteurs. Pour affiner les résultats, l'auteur propose d'ajouter en variable auxiliaire le logarithme de la concentration en NO_x .

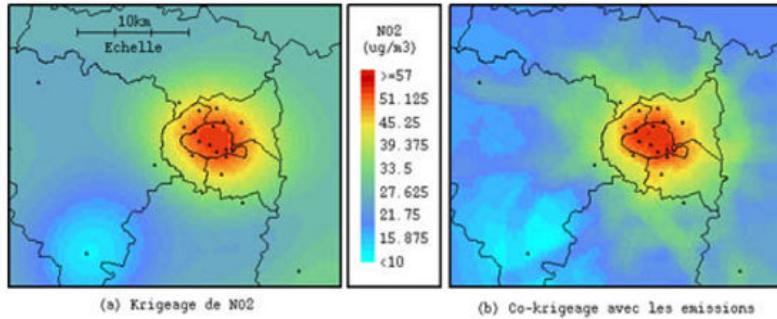


Fig. 3.21. Éstimation de la concentration de NO_2 par krigeage ordinaire (à gauche) et par co-krigeage croisé avec le logarithme des émissions de NO_x (à droite). Source : Peraudin (2003).

- **Krigeage d'indicatrice :** on peut utiliser toutes les techniques de krigeage vues précédemment pour interpoler une variable binaire, par exemple la présence d'un élément chimique dans le sol. La variable \hat{z} estimée peut être considérée comme étant l'espérance (conditionnée aux observations) d'une fonction indicatrice valant 1 si l'élément est présent et 0 sinon :

$$\hat{z}(x) = \mathbb{E}[Z(x)] = 0 \times \mathbb{P}[Z(x) = 0] + 1 \times \mathbb{P}[Z(x) = 1] = \mathbb{P}[Z(x) = 1]$$

Autrement, dit la valeur krigée en un site x peut être considérée comme la probabilité de présence de l'élément chimique en x sachant les observations. Remarquons toutefois que le krigeage peut produire des valeurs en dehors de l'intervalle $[0, 1]$, ce qui nécessite donc quelques opérations de régularisation pour que la valeur krigée puisse légitimement être considérée comme une probabilité en tout point du domaine.

De manière similaire, si Z est une variable continue, comme par exemple la teneur de l'élément, on peut la transformer en une variable binaire avec la fonction indicatrice $Y_c(x) = \mathbb{1}_{Z(x) \leq c}$, où $c \in \mathbb{R}$ est un seuil arbitrairement fixé. La fonction Y_c prend la valeur 0 en tous les sites où Z dépasse c et 1 sinon. Si on calcule un krigeage de la variable Y_c on obtient :

$$\hat{y}_c(x) = \mathbb{P}[Y(x) = 1] = \mathbb{P}[Z(x) \leq c] = F_{Y(x)}(c)$$

L'estimation en x correspond donc à la valeur prise en c par la fonction de répartition de la variable aléatoire $Y(x)$. On peut alors réitérer l'opération pour une série de valeurs régulières de seuils : c_1, c_2, \dots, c_n , permettant ainsi de reconstruire la fonction de répartition de $Z(x)$ en chaque site x , et ainsi d'en dériver des intervalles de confiance plus robustes que ceux qu'on peut obtenir sous l'hypothèse de normalité avec la variance de l'erreur de krigeage.

Remarquons toutefois que rien ne garantit la monotonie de la fonction de répartition ainsi estimée. En pratique, ici aussi, un post-traitement est nécessaire pour régulariser la solution.

Mentionnons également le fait que l'obtention de F a un coût de calcul important puisqu'un variogramme doit être calculé pour chaque valeur c du seuil. Par ailleurs, la fonction indicatrice Y_c étant par nature discontinue, le variogramme ne pourra pas être parabolique à l'origine, ce qui interdit d'emblée le choix des modèles gaussien et cubique.

- **Soft-kriging :** ou *krigeage mou*, il permet de combiner des observations formelles, avec des informations plus vagues. Par exemple, en bathymétrie, on peut combiner des informations de profondeur exactes (obtenues par sonar) avec des indications du type : en ce point, la profondeur est supérieure à x m, dans le cas de figure ou la portée du sondeur n'a pas permis d'y déterminer la profondeur. Le soft-kriging utilise en particulier des techniques de krigeage sous contraintes d'inégalités. Précisons que lorsque le nombre de contraintes floues reste modéré, les techniques de simulations présentées dans le chapitre 4 pourront avantageusement être employées.

3.8 Bilan

Dans ce chapitre nous avons vu comment utiliser la notion de processus stochastique pour estimer une variable régionalisée observée en seulement quelques points. En particulier, nous avons appris à utiliser les contraintes de krigeage pour dériver le système d'équations menant à un vecteur de poids solution. Rappelons que ces contraintes, qui sont la *marque de fabrique* du krigeage, sont toujours les mêmes : linéarité, autorisation, universalité et optimalité. Seules les hypothèses sur le modèle du processus vont mener ces contraintes à produire des équations différentes. En particulier, nous avons vu que le krigeage peut-être facilement personnalisé pour répondre à des besoins spécifiques.

Toutes les variantes du krigeage qui ont été étudiées dans ce chapitre (à l'exception du krigeage de la moyenne et du krigeage simple qui nécessitent impérativement une stationnarité du processus pour que la notion de valeur moyenne du champ ait un sens), peuvent s'exprimer aussi bien sous l'hypothèse stationnaire avec la covariance que sous l'hypothèse intrinsèque avec le variogramme. De même, si les observations sont prélevées sur des zones d'emprise spatiale non-négligeable, on peut avoir recours au krigeage par blocs avec le variogramme régularisé. Toutes ces caractéristiques sont combinables à volonté. Ainsi, on peut parler de co-krigeage intrinsèque par blocs d'indicatrices, ce qui illustre ici aussi la grande polyvalence de la méthode.

Par ailleurs, le krigeage permet de dériver une incertitude formelle sur la prédiction : la variance de l'erreur d'estimation, ce qui est une propriété fondamentale des méthodes probabilistes. De plus nous avons vu que cette variance ne dépend que de l'implantation des sites d'observation, et pas des mesures effectivement collectées, ce qui permet en amont de contrôler la qualité d'un plan d'expériences.

Enfin, si nous devions retenir une faiblesse capitale du krigeage, ce serait sans aucun doute son inaptitude à estimer des quantités non-linéaires, ce qui combiné à sa propriété de lissage, peut conduire à des estimations assez éloignées de la vérité. Cette limitation va motiver le besoin de simulations numériques, dont nous parlerons dans le chapitre suivant.

Chapter 4

Compléments

Dans ce dernier chapitre nous étudions quelques compléments utiles aux notions vues précédemment. En particulier, nous verrons une méthode objective de validation, qui peut s'apparenter à la méthode dite *Leave One Out Cross Validation* en apprentissage automatique. Nous verrons également comment utiliser le krigage pour trouver la valeur optimale d'une fonction numérique en minimisant le nombre d'observations nécessaires. Nous étudierons par ailleurs des techniques de simulations, conditionnelles et non-conditionnelles, offrant une vaste gamme de possibilités, allant de l'interpolation sous contrainte à l'étude de propagation d'erreurs, en passant par l'estimation de quantités non-linéaires, propriété capitale pour contourner les faiblesses du krigage. Nous conclurons ce cette section par une activité de synthèse, permettant de mettre en application (et de comparer) le krigage et les techniques de simulation, sur un problème concret.

4.1 La validation croisée

Comme nous l'avons vu dans le chapitre précédent, quelques soient les configurations d'hypothèses retenues (processus stationnaire ou intrinsèque, à moyenne connue ou inconnue, observations et variables à inférer ponctuelles ou par blocs, ...) il est possible de dériver une incertitude formelle sur la prédiction : la **variance de l'erreur d'estimation**. Le krigage d'indicatrice nous permet même d'aller plus loin et de connaître la fonction de répartition (et donc un intervalle de confiance sur la prédiction) de la variable estimée en chaque point.

Cependant, la propriété 37 nous montre que les estimations du krigage sont invariantes par multiplication scalaire du variogramme tandis que la variance de l'erreur d'estimation se retrouve affectée d'un facteur équivalent. Cela nous montre qu'il est toujours possible de réduire artificiellement l'incertitude formelle de la prédiction en divisant le variogramme par une constante. Une erreur dans la modélisation variographique (ou pire, en amont dans les hypothèses du modèle) peut donc rapidement conduire à la double sanction d'une variable mal estimée et d'une variance d'erreur optimistiquement faible. L'incertitude formelle doit donc être considérée comme un indicateur *a priori* de précision intéressant pour valider en amont un plan d'expérimentation. Néanmoins le modèle et les estimations obtenues doivent toujours être passées à l'épreuve d'une validation objective. C'est ce que nous étudions dans cette section.

On note $x \in \mathcal{D}$ un site quelconque du domaine, et $\mathbf{X} = [x_1, x_2, \dots x_n] \in \mathcal{D}^n$ et $\mathbf{Z} = [z_1, z_2, \dots z_n] \in \mathbb{R}^n$ les vecteurs contenant respectivement les sites et les observations qui y ont été relevées. Ces deux vecteurs ayant la même taille, on peut les regrouper dans une matrice $\mathbf{Y} = [\mathbf{X}, \mathbf{Z}]$ à n lignes, et telle que le vecteur ligne $\mathbf{y}_i = (x_i, z_i)$ désigne les données relatives au site i . Pour une configuration Θ de paramètres (modèle, palier, portée et pépite du variogramme, type de krigage, nombre d'observations utilisées pour chaque estimation...) on note $f(x, \mathbf{Y}, \Theta)$ l'estimateur par krigage de la variable $Z(x)$, à partir du semis d'observations \mathbf{Y} et avec le paramétrage Θ . Remarquons que le vecteur Θ peut alternativement être considéré comme un ensemble d'hypothèses. Par exemple, on

peut avoir : Θ_1 = [variogramme linéaire avec palier, pépite nulle, portée 200 m, moyenne connue...] et Θ_2 = [variogramme cubique, pépite nulle, portée 250 m, moyenne inconnue...]. L'objectif du problème consiste à trouver une mesure de qualité de prédiction permettant d'évaluer et de comparer les performances respectives des sets¹ d'hypothèses Θ_1 et Θ_2 afin de ne retenir que le meilleur.

Dans le chapitre précédent, nous avons défini l'erreur d'estimation en x par la quantité (aléatoire) :

$$E(x) = \hat{Z}(x) - Z(x) = f(x, \mathbf{Y}, \Theta) - Z(x) \quad (4.1)$$

Notons que l'erreur E est positive lorsque le krigeage sur-estime la valeur $Z(x)$ et négative sinon. Par ailleurs, il est intéressant de noter que E est lui-même un processus stochastique, clairement non-stationnaire (l'erreur étant statistiquement plus faible au voisinage des points du semis d'observations).

Pour estimer l'erreur E en x , on doit bien évidemment connaître la valeur vraie $Z(x)$. Sur des problèmes de benchmarks expérimentaux, la vérité terrain est connue de manière quasi-exhaustive, et les erreurs de toutes les grandeurs physiquement mesurables peuvent être évaluées. En revanche, dans un contexte opérationnel, la valeur vraie est justement inconnue puisqu'on cherche à l'estimer. La **validation croisée** apporte une solution à ce problème.

On note \mathbf{Y}_{-i} le jeu de données \mathbf{Y} dans lequel on a supprimé la ligne i . Le principe consiste alors à réestimer la valeur Z_i prise en x_i à partir des valeurs Z_{-i} prises en tous les autres sites. L'erreur d'estimation 4.1 se calcule alors par :

$$E_{LOOCV}(x_i) = \hat{Z}(x_i) - Z(x_i) = f(x_i, \mathbf{Y}_{-i}, \Theta) - Z(x_i) \quad (4.2)$$

La variance de l'erreur commise est alors calculée en moyennant 4.2 sur tous les sites observés :

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n \left[f(x_i, \mathbf{Y}_{-i}, \Theta) - Z(x_i) \right]^2 \quad (4.3)$$

et on peut calculer un écart-type d'estimation en extrayant la racine carrée de 4.3.

Il s'agit d'un cas particulier de validation croisée (CV). Dans une validation croisée à k *folds* on décompose l'ensemble des observations disponibles en k groupes de tailles homogènes. On utilise alors $k - 1$ groupes pour estimer les valeurs prises par le groupe restant, et on calcule l'erreur d'estimation en comparant les valeurs prédites aux valeurs observées. Le processus est répété (k fois au total) jusqu'à ce que toutes les observations aient été réestimées une fois. La validation 4.3 est donc un cas particulier de validation croisée avec : $k = n$. Comme chaque groupe ne comporte qu'une seule observation, on parle de *leave-one-out* cross validation (LOOCV).

Il existe une vaste littérature sur le sujet de la validation croisée (Bengio et Grandvalet, 2004), en particulier sur le contrôle statistique de son biais et de sa variance. En général, il est admis que plus les groupes sont d'effectifs réduits moins l'erreur d'estimation est biaisée, ce qui peut se comprendre

¹En théorie, l'ensemble des valeurs possibles de Θ est infini, et on ne peut comparer qu'un nombre fini de sets $\Theta_1, \Theta_2, \dots, \Theta_k$, qui seront choisis de manière à balayer l'ensemble des hypothèses crédibles pour le phénomène à modéliser.

intuitivement par le fait que $\mathbf{Y}_{-i} \approx \mathbf{Y}$ et donc :

$$E_{LOOCV}(x_i) = f(x_i, \mathbf{Y}_{-i}, \Theta) - Z(x_i) \approx f(x_i, \mathbf{Y}, \Theta) - Z(x_i) = E(x_i) \quad (4.4)$$

En revanche, dans cette configuration, les estimateurs sont fortement corrélés, *i.e.* même pour $i \neq j$: $f(x, \mathbf{Y}_{-i}, \Theta) \approx f(x, \mathbf{Y}_{-j}, \Theta)$, et donc la variance de E_{LOOCV} tend à augmenter (Efron, 1983).

Comme bien souvent, le compromis biais-variance 1.14 suggère que la configuration optimale correspond à des groupes d'effectif intermédiaire. En pratique, la conclusion semble moins directe, et est principalement fonction des dépendances entre les échantillons observés. Plusieurs travaux (Burman, 1989; Tian et al., 2007; Zhang et Yang, 2015) ont relevé des cas où biais et variances diminuent de manière monotone avec le nombre de groupes utilisés, indiquant ainsi que le LOOCV peut être optimal.

Remarquons également que la plupart de ces considérations concernent la communauté de l'apprentissage machine, dans laquelle les échantillons utilisés peuvent bien souvent être considérés comme indépendants, ce qui n'est pas le cas en Géostatistique.

En pratique, comme nous allons le voir dans l'activité suivante, avant de former la différence 4.2, on stocke les résultats dans deux vecteurs : $\hat{\mathbf{Z}}$, contenant les estimations par krigage et \mathbf{Z} contenant les valeurs réelles observées. On peut alors extraire de nombreux indicateurs de performance :

- Le **coefficient de corrélation** ρ entre les vecteurs $\hat{\mathbf{Z}}$ et \mathbf{Z} . Lorsque l'estimation est parfaite, on a $\hat{z}_i = z_i$ sur tous les sites, et donc $\rho = 1$. À l'inverse, plus ρ est proche de zéro (il peut même être négatif dans des cas particulièrement défavorables) moins l'estimation est satisfaisante. Le calcul de ρ s'accompagne souvent d'une représentation graphique du nuage de points (z_i, \hat{z}_i) .

En R, on peut le calculer avec la commande `cor(x,y)`. Lorsque le nombre de données est faible et qu'on souhaite caractériser la corrélation à l'aide d'indicateurs statistiques plus poussés, tels que la p-valeur qui permet de s'assurer que la corrélation obtenue n'est pas due au hasard, on pourra utiliser l'instruction `cor.test(x,y)`.

- La **distribution des erreurs** peut s'obtenir avec un histogramme `hist(E)` (en supposant que E est le vecteur contenant les écarts $\hat{z}_i - z_i$ ou avec une estimation lissée par noyaux (à manipuler avec précaution) : `density(E)`). On peut calculer de nombreux indicateurs à partir de E , notamment le biais `mean(E)`, l'erreur L_1 `mean(abs(E))`, le RMSE `sqrt(mean(E**2))` (aussi appelée erreur L_2), les valeurs extrêmes `min(E)` et `max(E)`, les quantiles `quantile(E)` et les indices des sites contenant des erreurs anormalement élevées `which(E > 1.96*sd(E))` (où ici, le facteur 1.96 est choisi en supposant la distribution des erreurs comme étant gaussienne). On pourra également représenter ces erreurs sur la carte pour inspecter leur distribution spatiale. Si les erreurs anormales sont groupées sur la carte, cela peut être le signe d'une valeur abérante dans les observations, ou bien encore de l'invalidité de l'hypothèse de stationnarité du processus.

- La part de **variance expliquée** :

$$1 - \frac{\sum_{i=1}^n (\hat{z}_i - z_i)^2}{\sum_{i=1}^n (\hat{z}_i - \bar{z})^2}$$

où \bar{z} désigne la moyenne des valeurs de z_i . Cet indicateur permet de mesurer le rapport de l'erreur commise avant et après l'estimation par krigage (étant entendu qu'avant l'estimation,

la meilleure prédiction qui puisse être faite est bien la moyenne \bar{z}). Ainsi, le dénominateur de la fraction indique l'erreur commise si on affectait en chaque site la valeur moyenne de tous les sites (c'est la variabilité intrinsèque de Z). Le numérateur quantifie lui l'erreur résiduelle.

- Enfin, l'**erreur normalisée** :

$$\kappa_i = \frac{|\hat{z}_i - z_i|}{\hat{\sigma}_i}$$

où $\hat{\sigma}_i$ désigne la racine carrée de la variance de l'erreur d'estimation pour le site i . Le coefficient κ_i permet donc de mesurer pour chaque site l'adéquation de l'estimation effectuée avec son incertitude formelle (évaluée elle aussi par le krigage). Si l'erreur $\hat{z}_i - z_i$ est trop grande relativement à l'écart-type $\hat{\sigma}_i$, cela signifie que la variance d'estimation est trop optimiste, ce qui est le signe d'une mauvaise analyse variographique. En général, on se contente d'examiner la proportion p des κ_i supérieurs à 1.96. À nouveau, ici aussi, sous l'hypothèse de normalité des erreurs, cette proportion ne devrait pas dépasser 5%. Lorsque le nombre d'échantillons est important (typiquement plusieurs milliers, on peut aussi évaluer la proportion des κ_i supérieurs à 3, et on s'assurera qu'elle est inférieure à 0.3 %). Par ailleurs, la représentation spatiale et fréquentielle des κ_i peut permettre d'identifier des données aberrantes.

Il revient alors à l'utilisateur de définir le ou les indicateurs les plus pertinents pour son application, afin de définir une métrique d'évaluation lui permettant de retenir le modèle le plus performant. Par exemple, si on retient la métrique L_1 , le choix de l'hypothèses optimale Θ^* se fera par :

$$\Theta^* = \underset{\Theta \in \mathcal{H}}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^n |f(x_i, \mathbf{Y}_{-i}, \Theta) - z(x_i)| \quad (4.5)$$

où $\mathcal{H} = \{\Theta_1, \Theta_2, \dots, \Theta_k\}$ est un ensemble réduit d'hypothèses sur lesquels on opère la comparaison.

Remarquons pour finir que la validation croisée LOOCV possède un coût de calcul important (notamment dû au fait que la matrice principale doit être inversée pour chaque estimation). Dans certains cas, si le nombre de données disponibles est considérable, on pourra opérer la procédure de LOOCV sur un sous-échantillon des sites.

Activité IX. La validation croisée en pratique

Dans cette activité, nous reprenons l'exemple d'interpolation d'un MNT étudié dans l'activité VIII et nous illustrons de manière concrète toutes les métriques brièvement présentées ci-dessus.

Nous commençons par reconditionner le code pour l'adapter à cette activité. En premier lieu, nous récupérons la portion des instructions permettant de lire le fichier de points d'observations et d'en transformer les coordonnées dans un repère métrique local. En particulier, pour simplifier le code, nous renommons les colonnes de la strcuture `sites` en `x`, `y`.

```
# Import des fichiers
obs = read.csv("sample_data2.txt", sep=",", header=1)
```

```

# Conversion en coordonnees planes
R = 6378137.0
obs$longitude = obs$longitude*pi/180.0
obs$latitude = obs$latitude*pi/180.0
obs$longitude = (obs$longitude - min(obs$longitude))*cos(mean(obs$latitude))*R
obs$latitude = (obs$latitude - min(obs$latitude))*R

# Table des sites
sites = obs[c("longitude", "latitude")]; names(sites) = c("x", "y")

# Table des observations
Z = obs["meters"]; Z = rbind(Z, 0)

```

Nous utilisons à présent un variogramme linéaire de portée $h_{max} = 7.7$ km avec palier $C = 395$ m², comme nous l'avions déterminé sur la figure 2.18 et dans la section 2.4.2.

```

vario = function(h){
  h[which(h > 7700)] = 7700; h=0.051*h
  return(h)
}

```

La pente de γ est calculée par : $C/h_{max} = 0.051$ m.

Pour simplifier la suite des manipulations, nous allons créer la fonction $f(x, \mathbf{Y}, \Theta)$ sous la forme `estimate(x,y,sites,Z)`, avec x un site du domaine (de coordonnées x, y dans le code), \mathbf{Y} un ensemble de semis d'observations (`sites` pour les coordonnées des sites et `Z` pour les observations qui y ont été effectuées) et Θ les paramètres du variogramme (qui ici n'interviennent qu'à travers la fonction globale `vario`). Remarquons que, tout comme dans l'activité VIII, le vecteur `Z` est complété par un 0 en prévision de la produit scalaire final avec le multiplicateur de Lagrange.

```

estimate = function(x,y,sites,Z){

  N = nrow(sites)

  # Calcul de la matrice de distance
  D = data.matrix(dist(sites, diag=TRUE, upper=TRUE))
  D = vario(D); D = rbind(D, rep(1,N)); D = cbind(D, c(rep(1,N),0))

  # Pre-calcul de l'inversion
  I = solve(D)

  # Distance entre les sites et (x,y)
  D0 = sqrt((sites$x - x)**2+(sites$y - y)**2)
  D0 = vario(D0); D0= c(D0, 1)

  # Resolution
  W = (I %*% D0)
  return(c(sum(W*Z), sum(W*D0) - W[N+1]))
}


```

La fonction `estimate` retourne la valeur prédite par krigeage $\hat{z}(x)$ ainsi que sa variance associée.

Il ne reste alors plus qu'à faire tourner une boucle sur les points du semis d'observations. Pour chaque point (de coordonnées (`sites$x`,`sites$y`) et en lequel a été observé la valeur `Z[i]`), on réestime la valeur prise par Z à partir des données amputées du site i : `sites[-i]` et `Z[-i]`.

Script 10. La validation croisée en pratique [code8.r]

```
N = nrow(sites)
zhat = rep(0,N)
variance = rep(0,N)
for (i in 1:N){
  pourcent = floor(100*i/N)
  cat("Computing interpolation : ",min(pourcent,100), "%\r")
  prediction = estimate(sites$x[i], sites$y[i], sites[-i,], Z[-i,])
  zhat[i] = prediction[1]
  variance[i] = prediction[2]
}

zvrai = Z[1:N,]

e = zvrai-zhat

cat("\r\n")
cat("Cross validation done\r\n")
```

À l'issue de l'exécution de ce morceau de code, les vecteurs `zvrai` et `zhat` contiennent respectivement les observations (réelles) des N sites et les valeurs réestimées par krigage. Idéalement, on souhaite que ces deux vecteurs soient les plus similaires possibles. On forme alors leur différence $e = zvrai - zhat$, le vecteur d'erreur. Remarquons que le calcul peut prendre un certain temps. Il est conseillé en premier lieu d'utiliser un jeu de données ne contenant pas plus de 500 observations. Nous avons alors tous les outils nécessaires pour étudier la précision de la procédure d'interpolation de l'activité VIII.

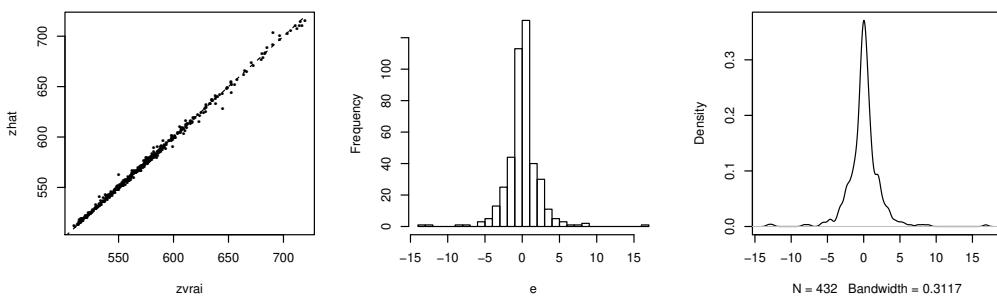


Fig. 4.1. À gauche : droite de corrélation entre les valeurs réelles (en abscisse) et prédites par krigage (en ordonnée). Au centre : histogramme des erreurs d'estimation pour $k = 30$ classes. À droite : densité de probabilité des erreurs d'estimation.

En premier lieu, on commence par représenter le nuage de points (z_i, \hat{z}_i) , à l'aide de l'instruction `plot(zvrai, zhat, pch=16, cex=.5)`. Pour aider la visualisation, on peut également représenter la première bissectrice du repère : `lines(1:max(zvrai), 1:max(zvrai), lty=2)`.

Alternativement, on peut représenter l'histogramme² des erreurs `hist(e,k)`, où k dénote le nombre de classes souhaitées, ou la densité des erreurs `plot(density(e))`.

Remarquons que la densité de probabilité des erreurs est calculée par la méthode des noyaux (Rosenblatt, 1956; Parzen, 1962), qui s'apparente un peu à ce qui a été fait dans l'activité VI : les données sont lissées à l'aide d'une fonction gaussienne. La largeur du noyau est un paramètre clé de la procédure, et est déterminée à partir des données (Nadaraya, 1965). Dans le cas des données du fichier `sample_data2.txt`, le noyau de lissage de la figure 4.1 a une largeur de 0.31 m environ. Plus le nombre de données augmente, plus le noyau a tendance à être compact et moins les données sont lissées par l'estimateur de la densité.

On peut alors calculer 4 indicateurs intéressants : le biais (3 cm), la moyenne des erreurs non-signées (1.17 m), la racine carrée de l'erreur quadratique moyenne (2.25 m) et le coefficient de corrélation (99.87 %). L'écart important résidant entre l'erreur L1 et l'erreur L2 indique la présence d'outliers dans l'estimation (*i.e.* d'erreurs anormalement grandes), qu'il faudra investiguer par la suite. La faible valeur de biais confirme que la contrainte d'universalité du krigeage a bien été respectée.

```
bias = mean(e)                      # Biais
emoy = sqrt(mean(abs(e)))           # Erreur L1
rmse = sqrt(mean(e**2))             # Erreur L2
rho = cor(zvrai, zhat)              # Biais
```

On peut également calculer la part de variance expliquée (99.7 %) :

```
var_explained = 1-mean(e**2)/var(zvrai)
```

On recherche dans un second temps les erreurs anormalement élevées, c'est-à-dire les indices des points en lesquels l'erreur d'estimation est supérieur à 1.96 fois l'écart-type des erreurs. On représente alors la distribution spatiale de ces sites (par rapport à tous les autres sites) pour rechercher un motif singulier.

```
id = which(e > 1.96*rmse)
plot(sites, pch=3); points(sites[id,], pch=16)
```

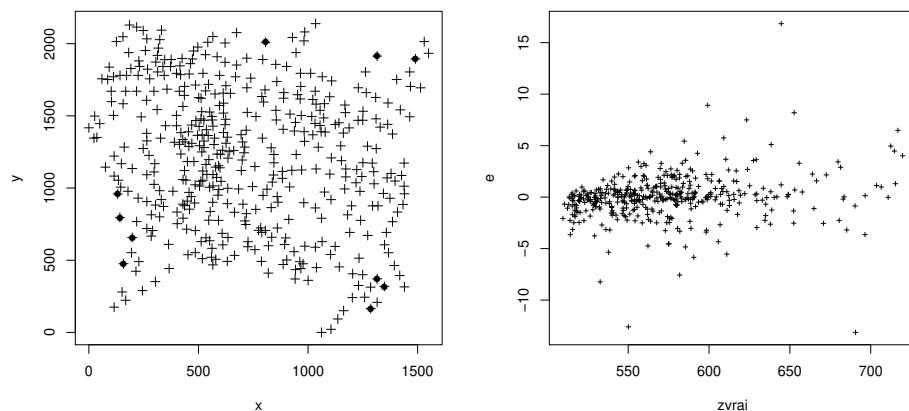


Fig. 4.2. À gauche : répartition des sites où l'erreur d'estimation est anormalement élevée (en losange plein). À droite : évolution de l'erreur en fonction de la valeur vraie.

²La fonction `hist` permet à la fois le calcul et le tracé de l'histogramme. La fonction `density` en revanche, nécessite d'être complétée par un appel à `plot`.

On trouve 10 sites en lesquels l'erreur est supérieure à $1.96 \times \text{rmse} = 4.42$ m. La figure 4.2 (à gauche) en particulier, montre que les points concernés sont situés en bord de domaine, ce qui explique les erreurs observées. On note cependant deux *clusters* de points, au milieu à gauche, et en bas à droite, ce qui peut inciter à aller vérifier si un des points voisins n'a pas été mal saisi.

Il peut être instructif également de représenter les erreurs commises en fonction de la valeur vraie (fig. 4.2 à droite) avec la commande `plot(zvrai, e)`. Sur ce graphique, on observe que l'erreur est en général d'autant plus élevée que la valeur à estimer est elle-même élevée. On parle d'*effet de proportionnalité* (Wackernagel, 2013b).

On peut généraliser le principe du graphique de gauche de la figure 4.2. Puisqu'on connaît l'erreur commise en chaque site x_i du domaine, on peut calculer une estimation par krigage de l'erreur sur l'ensemble du domaine. Pour connaître l'erreur en un site (x, y) quelconque, on cherche à interpoler les observations d'erreurs : $e(x_i) = \hat{z}(x_i) - z(x_i)$. Pour ce faire, il suffit d'appeler la fonction `estimate(x, y, sites, E)`, où E est un vecteur analogue à Z , dans lequel les valeurs d'altitudes sont remplacées par les erreurs d'estimations calculées par la validation croisée. On donne ci-dessous (fig. 4.3) les résultats obtenus.

On vérifie que l'erreur est à peu près homogène à l'intérieur du domaine. En particulier, les zones où l'erreur est la plus faible (en jaune orangé) correspondent aux zones de plus forte densité de points.

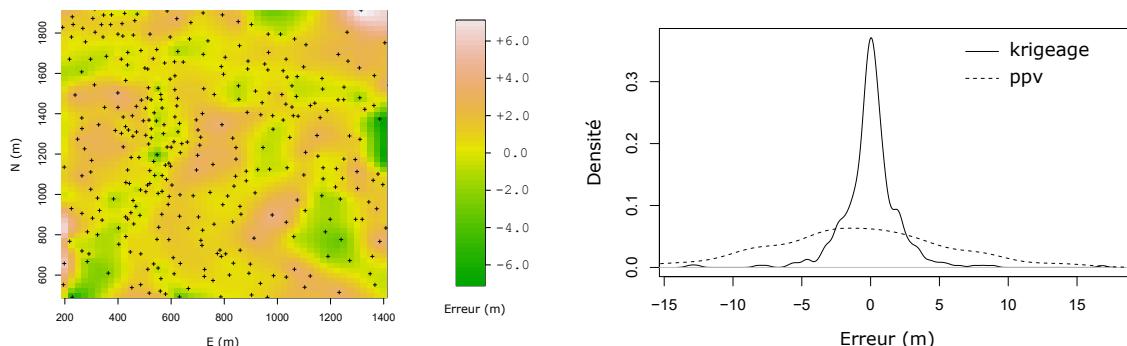


Fig. 4.3. À gauche : interpolation par krigage ordinaire des erreurs calculées par validation croisée. Le semis de points d'observations est représenté par des croix. À droite : densités des erreurs commises par le krigage (en trait plein) et par la méthode des plus proches voisins (en pointillé)

On peut également comparer les résultats obtenus avec une méthode d'interpolation déterministe, par exemple celle du plus proche voisin (ppv), qui consiste à affecter en un site x , la valeur observée $Z(x_i)$ du site i le plus proche de x . Ici, on ne cherche pas à effectuer l'interpolation ppv sur tout le champ mais seulement au niveau des sites. Pour ce faire, on commence par écrire une fonction `argmin` prenant en entrée un vecteur de nombres et renvoyant l'indice du minimum :

```
argmin = function(x){return(which.min(as.vector(x)))}
```

Par exemple, si `x = c(5, 8, 7, 6, 4, 1, 2, 4, 8, 6)`, alors `argmin(x)` va retourner la valeur 6, correspondant à l'indice de `x[6] = 1` qui est bien le minimum de la série. Notons que si `x` contient en plusieurs exemplaires le nombre qui réalise le minimum, alors la fonction retourne arbitrairement le plus petit indice de ces nombres.

On calcule par la suite la matrice des distances entre tous les sites : `M = data.matrix(dist(sites))`. Ainsi, le vecteur ligne `M[i,]` contient les distances séparant le site x_i à tous les autres sites. En passant ce vecteur dans la fonction `argmin` on doit pouvoir récupérer le site le plus proche. Mais il faut être vigilant : les entrées diagonales de la matrice `M` valent 0, et donc l'application directe de `argmin` à `M[i,]`

va nécessairement renvoyer l'entier i (en supposant qu'il n'y a pas deux sites confondus). On va donc artificiellement augmenter la diagonale de M , par exemple en posant :

```
M[which(M == 0)] = max(M)+1
```

qui va avoir pour effet de : (1) récupérer la distance maximale entre deux sites, (2) l'incrémenter d'une unité, et enfin (3) d'affecter le résultat obtenu aux éléments diagonaux de M . De cette manière, l'instruction `argmin(M[i,])` renverra bien le site x_j (avec $j \neq i$) le plus proche de x_i . La procédure de validation croisée de la méthode ppv est alors immédiate, en bouclant sur les sites :

```
eppv = rep(0,N)
```

```
for (i in 1:N){
  eppv[i] = Z[argmin(M[i,]),]-Z[i,]
}
```

En effet, si `argmin(M[i,])` est l'indice du ppv de x_i , alors, $Z[argmin(M[i,]),]$ est l'estimation $f_{\text{ppv}}(x_i, \mathbf{Y}_{-i})$ de $z(x_i)$ à partir des données \mathbf{Y} dont on a supprimé la i -eme ligne. Il ne reste alors plus qu'à représenter (sur la même) les histogrammes lissés des vecteurs d'erreur e et ppv :

```
plot(density(e))
lines(density(ppv), lty=2)
```

Le résultat obtenu est donné en figure 4.3 (à droite). En particulier, on peut comparer le RMSE de la procédure de krigeage (2.25 m pour mémoire) avec celui du ppv : 7.36 m, ce qui montre la nette supériorité du krigeage, y compris avec un choix arbitraire de variogramme. De même, on remarque que contrairement au krigeage, la méthode ppv est significativement biaisée (-1.13 m), ce qui peut probablement s'expliquer par la géométrie particulière des sites sur la surface du terrain.

Inspectons à présent la variance formelle de l'erreur d'estimation. Pour ce faire, on commence par former le vecteur κ des résidus normalisés : `kap = e/sqrt(variance)`. On cherche alors les outliers, avec la commande `which(kap > 1.96)`, qui retourne un vecteur de 14 éléments, ce qui signifie que 96.7% des erreurs sont comprises dans l'intervalle de confiance théorique à 95%, suggérant ainsi que les incertitudes formelles évaluées par le krigeage sont relativement fiables. À nouveau, une représentation graphique similaire à celle de la figure 4.2 permet de mettre en évidence le fait que les sites concernés sont situés en bord de domaine. Ici aussi, on identifie la présence de deux clusters de valeurs erronées.

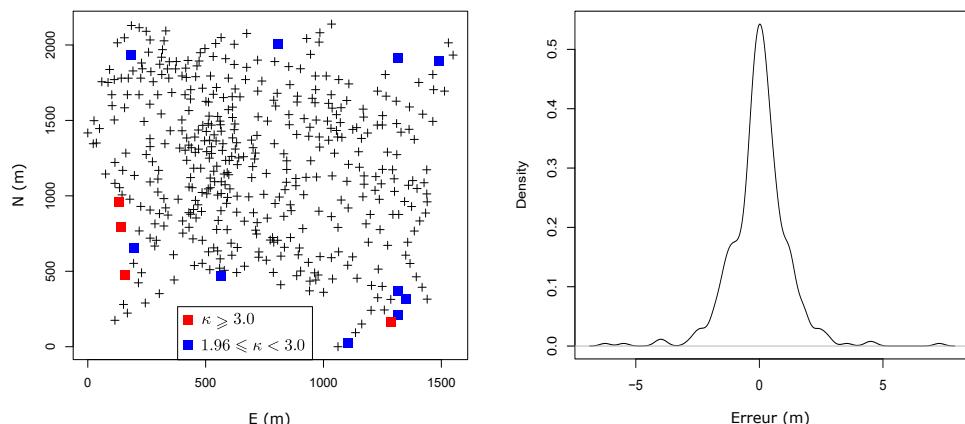


Fig. 4.4. À gauche : répartition des sites où l'erreur d'estimation est anormalement élevée ($\kappa \in [1.96; 3]$ en bleu et $\kappa \geq 3$ en rouge). À droite : distribution des erreurs normalisées κ .

4.2 Application à l'optimisation bayésienne

Dans cette section, nous allons présenter une utilisation concrète du krigeage pour trouver le maximum d'une fonction en minimisant le nombre d'observations effectuées.

4.2.1 Introduction

Formellement, on se donne une fonction inconnue $f : \mathcal{X} \rightarrow \mathbb{R}$, où \mathcal{X} représente l'espace des paramètres de la fonction. En général $\mathcal{X} \subseteq \mathbb{R}^p$. On note f^* la valeur optimale de f , et x^* un paramétrage de \mathcal{X} qui permet d'atteindre f^* :

$$x^* \in \operatorname{argmax}_{x \in \mathcal{X}} f(x) \quad f^* = f(x^*)$$

Dans la suite nous supposerons que \mathcal{X} est un domaine borné de \mathbb{R}^p , et que la fonction f est continue, si bien que x^* existe et f^* a une valeur finie. L'objectif du problème consiste à trouver le paramétrage optimal x^* en minimisant le nombre n d'observations .

La démarche employée ici consistera, pour chaque itération, à analyser les n observations effectuées jusque lors, à affiner les informations disponibles sur f , puis à choisir un nouveau point x_{n+1} à observer. Nous verrons par la suite que le choix de x_{n+1} peut répondre à plusieurs critères différents, mais en règle générale, il vise à répondre à 2 exigences complémentaires :

- **Exploitation** : on cherche à positionner x_{n+1} dans une zone à fort potentiel, *i.e.* dans une zone où ont déjà été observées des valeurs élevées de f .
- **Exploration** : on veut avoir une connaissance grossière du paysage $f(\mathcal{X})$, ce qui nécessite de positionner des sites d'observations dans les zones peu connues.

De manière plus concise, on cherche à placer l'observation x_{n+1} dans des zones qui sont à la fois **prometteuses et peu observées**. Nous verrons que ces deux objectifs en apparence contradictoires peuvent être combinés de manière élégante à l'aide des fonctions dites de *bande de confiance*.

Le problème du *bandit manchot* constitue certainement le problème modèle le plus simple : un joueur se trouve face à une rangée de machines à sous, qu'il peut utiliser N fois, en répartissant ses essais de la manière qu'il souhaite entre les différentes machines. Chaque machine possède une espérance de gain fixée et inconnue du joueur. Quelle stratégie le joueur doit-il adopter pour maximiser ses gains ?

Il est intéressant de remarquer que cette stratégie de compromis exploitation/exploration (parfois aussi appelée intensification/diversification) est présente dans de nombreux processus naturels. Par exemple, en génétique, la théorie de Darwin nous enseigne que seules les gènes les plus adaptés à l'environnement survivent aux générations suivantes, ce qui constitue une forme d'exploitation, complétée par l'exploration des mutations aléatoires. Les stratégies de chasse et de recherche de nourriture de nombreuses espèces animales sont également fondées sur ce principe ([Mehlhorn et al., 2015](#)).

Puisque le choix du plan d'expérience est fait simultanément à la procédure d'interpolation, on parle d'*optimisation séquentielle*. Par ailleurs, le processus de recherche consiste à considérer f comme une fonction aléatoire, puis à inférer ses paramètres de manière itérative avec les données disponibles au fil de l'eau, d'où l'appellation d'*optimisation bayésienne*.

Remarquons que si on cherche à minimiser la fonction f , toutes les considérations qui suivent peuvent être transposées sans problème en considérant la fonction $g = -f$.

Donnons en premier lieu quelques exemples concrets motivant le recours à l'optimisation séquentielle.

- Dans le domaine des essais cliniques, on peut souhaiter déterminer un protocole de soin optimal pour traiter une maladie. Chaque protocole est défini par un certains nombre de paramètres (composition et grammage du médicament, posologie, soins annexes...) et la réponse du patient est mesurée à l'issue du traitement. De par la nature statistique des évaluations, chaque test nécessite plusieurs dizaines de patients, et un temps considérable pour suivre le protocole et analyser les résultats. Dans ces conditions, il n'est pas envisageable de tester tous les protocoles possibles, et l'optimisation bayésienne nous permettra de définir séquentiellement un plan d'expérimentation $\mathbf{X} = \{x_1, x_2, \dots, x_n\}$, de sorte à trouver un protocole quasi-optimal $x^* \in \mathbf{X}$.
- En ingénierie, et en particulier dans l'étude de la résistance des matériaux, chaque expérimentation peut nécessiter la destruction d'un ouvrage, occasionnant ainsi des frais colossaux pour l'entreprise. Par exemple, trouver la structure x d'un véhicule permettant de limiter les dégâts causés par un choc frontal, nécessite de sacrifier n véhicules avec différentes structures x_1, x_2, \dots, x_n , et évaluer dans chaque cas les dégâts causés $f(x_1), f(x_2), \dots, f(x_n)$. On comprend alors aisément l'importance de minimiser le nombre n d'expérimentations.
- En ingénierie logicielle ou dans la recherche scientifique, on peut chercher à optimiser les performances d'un code f travaillant à partir d'un ensemble x d'arguments. Si le programme est relativement long à exécuter, on ne peut raisonnablement pas tester f pour tous les arguments x possibles. Ici aussi, l'optimisation séquentielle permettra de trouver un paramétrage satisfaisant du code tout en minimisant le temps de conception. On parle d'*hyper-paramétrage*.

D'une manière générale, l'optimisation bayésienne est un outil de choix dans tous les problèmes de recherche dans un espace dans lequel la collecte d'informations est longue, coûteuse ou risquée.

4.2.2 Principe

Pour illustrer la méthode, nous prendrons l'exemple d'une fonction f à une dimension, dont nous supposerons disposer de 3 observations z_1, z_2 et z_3 en trois points donnés x_1, x_2 et x_3 .

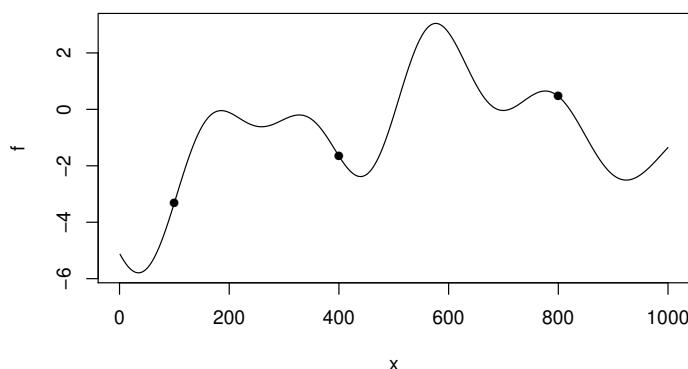


Fig. 4.5. Une fonction uni-dimensionnelle f et trois échantillons aléatoires $x_1, x_2, x_3 \in \mathcal{X}$.

Dans un premier temps nous considérerons que ces observations ne sont pas entâchées d'erreurs, autrement dit on a les égalités exactes $z_i = f(x_i)$ pour $i = 1, 2, 3$. Par ailleurs, on suppose avoir un modèle de variogramme γ caractérisant le processus stochastique dont f est une réalisation.

Connaissant γ ainsi qu'un semis de points d'observations, on peut caluler une estimation de f par krigeage en tout point du domaine \mathcal{X} . On note \hat{f} cette estimation et σ l'écart-type de l'erreur d'estimation. Insistons bien sur le fait que \hat{f} et σ sont des fonctions définie sur le même support que f . On les représente sur la figure 4.6

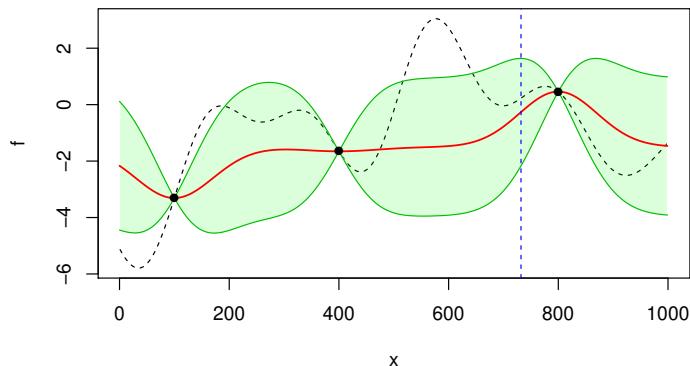


Fig. 4.6. La fonction f inconnue (en pointillé noir), son estimation \hat{f} par krigeage à l'aide des 3 premiers points de données (en rouge) et la bande de confiance associée (en vert).

On forme alors deux fonctions, traditionnellement notées U (pour UCB ou *Upper Confidence Band*) et L (pour LCB ou *Lower Confidence Band*) :

$$\begin{aligned} U(x) &= \hat{f}(x) + q\sigma(x) \\ L(x) &= \hat{f}(x) - q\sigma(x) \end{aligned} \tag{4.6}$$

où $q > 0$ est un coefficient de sureté. Plus q est grand, plus la marge d'erreur tolérée est grande. La zone comprise entre les courbes L et U est appelée la bande de confiance de \hat{f} : pour un point x donné quelconque, la valeur vraie de $f(x)$ appartient à l'intervalle $[L(x); U(x)]$ avec une probabilité dépendant uniquement du coefficient de sureté. En particulier, sous l'hypothèse normale et pour $q = 1.96$, cette probabilité vaut 95%.

On choisit alors un point x_4 qui maximise la bande de confiance supérieure, autrement dit, on cherche le site qui a le plus haut potentiel : $x_4 = 732$ (ligne verticale pointillée sur la fig. 4.6). On mesure alors la valeur prise par la fonction en x_4 et on ajoute cette nouvelle donnée aux observations, ce qui permet d'effectuer une nouvelle estimation par krigeage de f , puis d'en déduire un nouveau point x_5 et ainsi de suite.

Formellement, l'algorithme effectue à chaque itération les opérations suivantes :

- Calcul de \hat{f}_k , l'estimation par krigeage effectuée à l'aide des k premiers points observés, et de σ_k , l'écart-type correspondant.

- Détermination du point à observer à l'étape suivante :

$$x_{k+1} \in \operatorname{argmax}_{x \in \mathcal{X}} \hat{f}_k(x) + q\sigma_k(x) \quad (4.7)$$

- Calcul de $f(x_{k+1})$ et ajout dans les données.

L'algorithme est initialisé en tirant 2 ou 3 sites aléatoirement et se termine lorsque l'une des conditions suivantes est remplie :

- L'algorithme a convergé, *i.e.* la distance (dans \mathcal{X}) séparant deux points consécutifs est inférieure à un critère prédéterminé : $\|x_{k+1} - x_k\| \leq \varepsilon$.
- Le nombre maximal d'itération a été atteint.
- Une valeur $f(x_k)$ satisfaisante a été obtenue.

La figure 4.7 illustre le déroulement de 6 itérations de l'algorithme à partir de l'évaluation des 3 premiers points. Pour un critère d'arrêt fixé à $\varepsilon = 1$, la convergence est atteinte en $x_{10} = 579$, soit après 10 évaluations de la fonction objectif f . La position du maximum vrai de la fonction est $x^* = 578$, et l'erreur commise par l'algorithme est de l'ordre de la tolérance ε . L'erreur sur f^* , rapportée à l'amplitude de f , est de l'ordre de 0.02%. Pour obtenir un résultat d'une qualité équivalente avec une recherche exhaustive, il aurait fallut de l'ordre d'un millier d'évaluations de la fonction objectif.

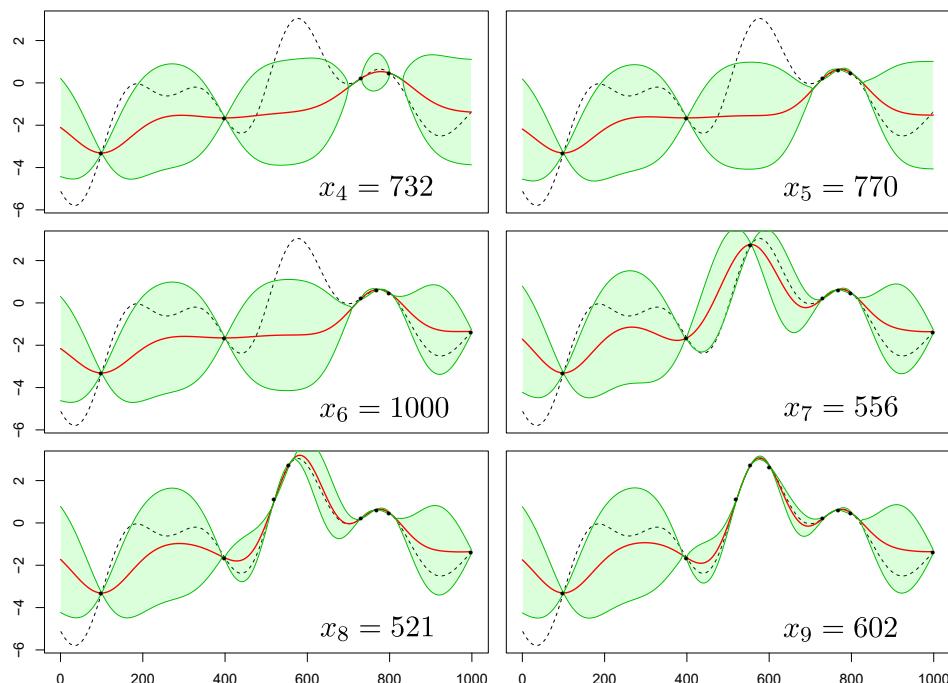


Fig. 4.7. 6 itérations de l'algorithme (après évaluations des 3 premiers points) pour un variogramme gaussien de portée 100 et un coefficient de sureté $q = 3$.

On remarquera que le coefficient de sureté q est un pondérateur entre les stratégies d'exploitation et d'exploration. Lorsque q est faible, la bande de confiance est relativement compacte autour de l'estimation, et les zones à fort potentiel ont tendance à être des zones où l'estimation est elle-même élevée. À l'extrême, si $q = 0$, et l'équation 4.7 dégénère en une recherche du point maximisant

l'estimateur \hat{f} sans tenir compte de l'incertitude σ . L'algorithme a ainsi tendance à échantillonner les zones de valeurs élevées, et donc à favoriser l'exploitation. À l'inverse, quand $q \rightarrow \infty$, l'incertitude σ est prépondérante dans le calcul de la bande U , l'algorithme ne tient plus compte de l'estimateur \hat{f} et adopte une stratégie purement exploratoire.

À ce titre il est instructif de rejouer le scénario d'optimisation de la figure 4.7 pour un coefficient de sureté $q = 2$. Le résultat est représenté sur la figure 4.8 et montre que l'algorithme converge vers la même valeur optimale, mais suivant un schéma différent. En particulier, il commence par exploiter plus intensivement la première zone prometteuse (aux alentours de $x = 800$) avant d'explorer la zone contenant le maximum final. La convergence est dans ce cas légèrement plus longue et est atteinte après 12 évaluations de f .

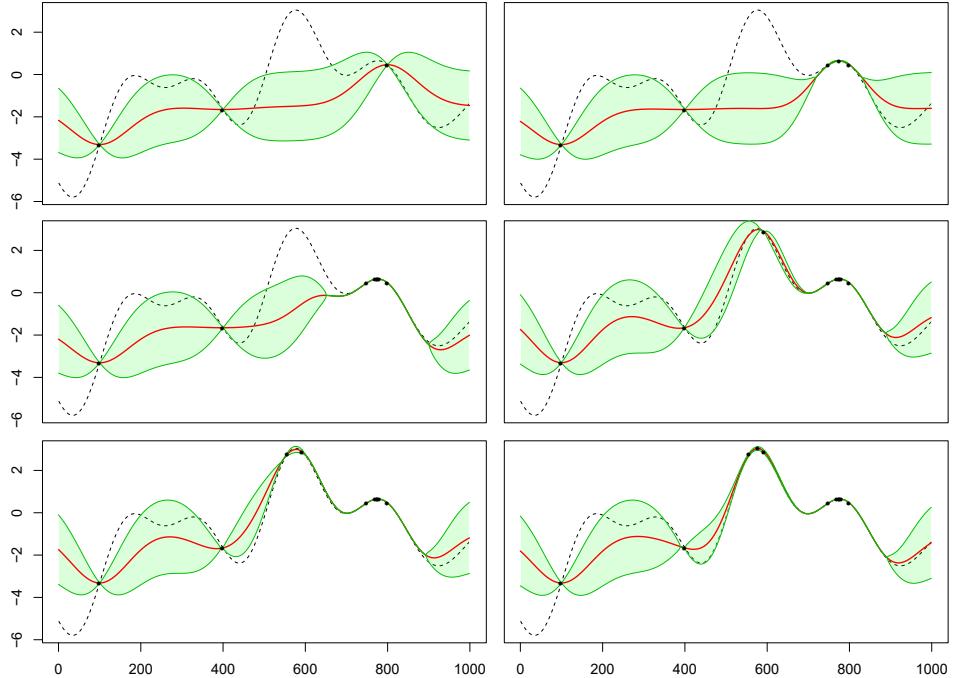


Fig. 4.8. 6 itérations de l'algorithme (après évaluations des 3 premiers points) pour un variogramme gaussien de portée 100 et un coefficient de sureté $q = 2$.

4.2.3 Pour aller plus loin...

Plusieurs améliorations peuvent être apportées à l'algorithme.

Critère de choix de x_{k+1}

La recherche du nouveau site à exploiter x_{k+1} se fait en maximisant une fonction analytique g , estimée à partir des k premières données. Ci-dessus, nous avons utilisée la fonction $g(x|x_1, x_2, \dots, x_k) = U(x)$, où $U(x)$ est définie de manière implicite par l'équation :

$$\mathbb{P}[f(x) \leq U(x)|x_1, x_2, \dots, x_k] = \alpha$$

avec α une marge de confiance, dépendant de q .

On peut utiliser d'autres critères, en notant f_k^* le maximum courant de f trouvé à l'itération k et en définissant la fonction g par

- **Probability of improvement :**

$$g(x|x_1, x_2, \dots, x_k) = \mathbb{P}[f(x) \geq f_k^* | x_1, x_2, \dots, x_k]$$

Maximiser g revient donc à trouver le point x_{k+1} du domaine tel que la probabilité que la valeur prise par f en x_{k+1} soit supérieure au maximum courant est la plus grande possible.

- **Expectancy of improvement :**

$$g(x|x_1, x_2, \dots, x_k) = \mathbb{E}[f(x) - f_k^* | x_1, x_2, \dots, x_k]$$

Maximiser g revient à trouver le point x_{k+1} du domaine tel que le gain moyen sur f par rapport au maximum courant soit le plus grand possible.

Remarquons que, contrairement à la définition $g(x) = U(x)$, dans les deux alternatives ci-dessus, la fonction g dépend directement du maximum courant. On pourra trouver plus de détails sur ces différents critères dans [Snoek et al. \(2012\)](#).

Bruit d'observation

Lorsque les mesures effectuées sont bruitées, c'est-à-dire que les données disponibles sont $z_i = f(x_i) \pm \sigma_\eta$ où σ_η est un bruit inhérent au processus d'observation, on peut utiliser un variogramme à pépite, comme illustré sur la figure 4.9. La variance d'estimation de f n'est donc jamais nulle, y compris au niveau des points d'observations où elle est justement égale à la valeur de la pépite.

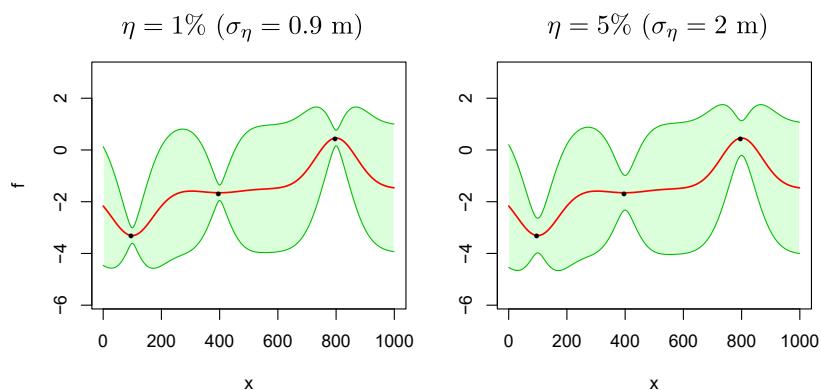


Fig. 4.9. Procédure d'optimisation bayésienne à partir de 3 échantillons observés et avec un variogramme gaussien de pépite $\eta = 1\%$ (à gauche) et $\eta = 5\%$ (à droite).

Activité X. Optimisation bayésienne par krigage

Dans cette activité nous utilisons une fonction f , échantillonnée en 1000 points et générée aléatoirement à l'aide de la fonction `generate`, définie dans l'activité VI et d'un noyau gaussien h d'écart-type 100. Écrire le code R permettant de trouver le maximum de f à l'aide de l'optimisation bayésienne. On utilisera pour ce faire un modèle de variogramme gaussien de portée 100, un nombre maximal de 10 itérations et un critère de convergence fixé à $\varepsilon = 1$.

Le code solution ne sera pas présenté ici, puisqu'il s'agit d'un assemblage de modules déjà expliqués dans les activités précédentes. On pourra trouver une correction dans le fichier `code9.r`.

4.3 Simulations

Nous avons vu précédemment dans la section 3.5.4 que l'interpolation par krigage d'un semis de points d'observations produit systématiquement des fonctions plus lisses que les réalisations typiques du processus ayant généré le semis. Cet effet était particulièrement visible sur les figure 3.11 et 3.12.

Or dans certaines problématiques (dont nous donnerons des exemples concrets par la suite), il est important de pouvoir disposer d'une reproduction fidèle des réalisations. C'est justement ce que proposent de faire les simulations. Notons qu'il existe deux types de simulations :

- La **simulation non-conditionnelle** permet, moyennant la connaissance d'un modèle de variogramme, de générer un nombre arbitraire de réalisations du processus sur un domaine \mathcal{D} .
- La **simulation conditionnelle** permet, moyennant la connaissance d'un modèle de variogramme et d'un semis de points d'observations, de générer sur un domaine \mathcal{D} un nombre arbitraire de réalisations du processus coïncidant toutes au niveau des points du semis d'observations.

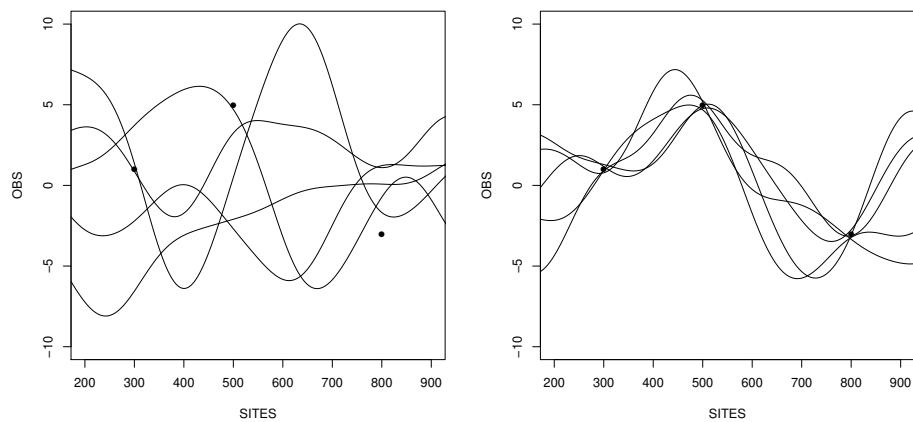


Fig. 4.10. À gauche : 5 simulations non-conditionnelles. À droite : 5 simulations conditionnées aux 3 points de données : $y(300) = 1$, $y(500) = 5$ et $y(800) = -3$. Dans les deux cas, un variogramme gaussien de portée 100 et de pépite d'écart-type 0.5 a été utilisé.

Notons que la distinction entre ces deux types de simulation est arbitraire et à vocation uniquement pédagogique : une simulation non-conditionnelle n'est autre qu'une simulation conditionnée à un ensemble vide d'observations. Malgré tout, les contextes nécessitant le recours à l'un ou l'autre de ces

types étant en général assez différents, nous maintiendrons cette distinction dans la suite du cours.

En théorie le domaine \mathcal{D} peut être discret (*e.g.* un semis de points, une grille régulière...) ou continu (*e.g.* un sous-domaine de $\mathbb{R}, \mathbb{R}^2\dots$). Néanmoins, d'un point de vue informatique, toutes les quantités adressables en machine étant nécessairement finies, toute simulation sur un support continu devra au préalable passer par une discrétisation de résolution aussi fine que les capacités de calculs le permettent. Ainsi, par exemple, la simulation d'un MNT se fait en générant une collection de variables aléatoires (corrélées via le variogramme γ) sur un semis de points d'observations en forme de grille régulière.

Remarquons également que nous avons déjà effectué des simulations au début du chapitre 3, plus spécifiquement dans l'activité VI. Notamment, la figure 3.3 présentent des exemples de simulations conditionnées³ à trois points observés.

4.3.1 Simulation non-conditionnelle

On considère le problème suivant : disposant du modèle de variogramme γ d'un processus Z défini sur un domaine \mathcal{D} , comment simuler aléatoirement des réalisations z ?

Comme précisé en introduction, on commence nécessairement par discréteriser le domaine \mathcal{D} , et nous noterons $x_1, x_2, \dots, x_n \in \mathcal{D}$ les sites ne lesquels on souhaite simuler Z .

Une solution intuitive pourrait consister à procéder itérativement et de manière similaire à ce que nous avons vu dans la section traitant de l'optimisation bayésienne. Supposons que nous soyons capable de simuler Z sur les k premiers sites. En utilisant ces k valeurs générées comme un semis de points d'observation, on peut calculer une estimation z_{k+1} par krigage de la valeur prise par le processus sur le $(k+1)$ -eme site. Grâce à la nature probabiliste du krigage, l'estimateur \hat{z}_{k+1} est accompagné d'une variance de l'erreur d'estimation σ_{k+1}^2 . Informellement, cela signifie que pour être cohérente avec les k premiers tirages, la valeur en x_{k+1} doit être distribuée suivant une loi normale de moyenne \hat{z}_{k+1} et de variance σ_{k+1}^2 . Il suffit donc de générer : $z_{k+1} \sim \mathcal{N}(\hat{z}_{k+1}, \sigma_{k+1}^2)$. On ajoute ce nouveau point aux observations et on continue ainsi de suite jusqu'à avoir simulé Z sur tous les sites. L'algorithme est initialisé en tirant arbitrairement la valeur du premier site : $z_1 \sim \mathcal{N}(0, \gamma(\infty))$.

Cette méthode possède l'avantage d'être conceptuellement simple et facilement interprétable, mais requiert un coût de calcul particulièrement élevé : chaque nouvelle simulation ponctuelle z_k nécessite de calculer une estimation par krigage à partir des $k - 1$ points précédents. Le coût d'une estimation ponctuelle par krigage étant un $\Theta(k^3)$, on en déduit la complexité de la procédure :

$$C(n) = \sum_{k=1}^n \Theta(k^3) = \Theta(n^4)$$

Le coût d'une simulation évolue donc comme la puissance quatrième du nombre de sites à simuler, ce qui est considérable. Il est possible de le réduire à l'aide des techniques de voisinage glissants. C'est en particulier l'approche employée par la bibliothèque `gstat` que nous allons utiliser à la fin de ce chapitre.

Dans un but pédagogique, nous présentons ici une méthode purement algébrique permettant de simuler les valeurs prises par tout les sites simultanément, moyennant l'hypothèse que le champ soit stationnaire, et qu'il soit donc possible de manipuler sa fonction de covariance. Pour ce faire, nous allons

³Pour abréger le temps de calcul, nous avions autorisé une tolérance ε , impliquant ainsi que les simulations ne passent pas exactement par les points conditionnans. Formellement, on peut considérer qu'il s'agit d'une simulation conditionnelle d'un processus à variogramme pépitique.

avoir recours au théorème 12 de propagation des variances : si \mathbf{A} est une matrice de $\mathbb{R}^{m \times n}$ et $\mathbf{X} \in \mathbb{R}^n$ un vecteur aléatoire de matrice de covariance Σ , alors $\mathbf{Y} = \mathbf{AX}$ a pour covariance $\mathbf{A}\Sigma\mathbf{A}^T$.

Soit Σ la matrice de terme général $(\Sigma)_{ij} = \gamma(0) - \gamma(h_{ij})$ avec $h_{ij} = \|x_i - x_j\|$. Comme γ est parfaitement connu, le calcul de la matrice Σ ne pose aucun problème.

On pose \mathbf{X} un vecteur de n nombres aléatoires indépendants et identiquement distribués suivant une loi normale standard. La matrice de covariance de \mathbf{X} est $\Sigma_{\mathbf{X}} = \mathbf{I}_n$, la matrice identité d'ordre n .

Enfin, considérons $\mathbf{A} \in \mathbb{R}^{m \times n}$ une matrice quelconque de réels. Le produit $\mathbf{Y} = \mathbf{AX}$ est une variable aléatoire et sa matrice de covariances vaut :

$$\Sigma_{\mathbf{Y}} = \mathbf{A}\Sigma_{\mathbf{X}}\mathbf{A}^T = \mathbf{AA}^T$$

Donc, pour que la variable \mathbf{Y} ait pour matrice de covariances Σ , il faut et il suffit que $\Sigma = \mathbf{AA}^T$, autrement dit, on souhaite que \mathbf{A} soit une décomposition de Cholesky de Σ .

Théorème 42 (Factorisation de Cholesky)

Soit \mathbf{A} une matrice symétrique définie-positive. Alors elle admet une *factorisation de Cholesky*, i.e. il existe une matrice \mathbf{L} triangulaire inférieure telle que :

$$\mathbf{A} = \mathbf{LL}^T$$

La matrice \mathbf{L} est en quelques sortes une *racine carrée* de \mathbf{A} .

La matrice Σ étant une matrice de covariance, elle est symétrique définie-positive⁴, et on peut donc en extraire une décomposition de Cholesky. Nous avons donc une méthode clés en main pour simuler le processus en n sites. On en donne les étapes ci-dessous :

- Former la matrice de covariance Σ entre les n sites (à l'aide de la fonction de covariance ou du variogramme γ). C'est une matrice carrée de taille n .
- Calculer sa décomposition de Cholesky, i.e. trouver une matrice \mathbf{L} telle que : $\Sigma = \mathbf{LL}^T$. Cette étape s'effectue à l'aide des bibliothèques de calcul matriciel, disponibles dans tous les langages de programmation. En particulier, en R on utilise la fonction `chol`.
- Calculer le vecteur $\mathbf{Z} = \mathbf{LX}$

Et c'est tout ! On sait d'après ce qui précède que \mathbf{Z} est un vecteur dont la covariance est celle prise par le processus au niveau des sites à simuler. Voyons cela en pratique dans l'activité qui suit.

⁴D'où l'importance de modéliser le variogramme avec une fonction licite, i.e. une fonction conditionnellement définie-négative, comme nous l'avons vu dans la section 2.4.5.

Activité XI. Une première simulation

On se place en 1D, avec un variogramme linéaire avec palier (portée 100) :

```
vario = function(h){  
  h[which(h > 100)] = 100  
  return(h)  
}
```

On crée alors une fonction **C** qui transforme le variogramme en covariance :

```
C = function(h){  
  return(100-vario(h))  
}
```

On crée ensuite 500 sites régulièrement espacés, puis on calcule les distances (le long de la droite réelle) entre ces sites et on forme la matrice de covariances **D** :

```
x = 0:500; n = length(x)  
D = C(data.matrix(dist(x, diag=TRUE, upper=TRUE)))
```

Le cœur de la simulation réside alors en trois lignes : la génération du vecteur normal **X**, la décomposition de Cholesky **L** et le produit matriciel **LX**.

Script 11. Une première simulation [code10.r]

```
X = rnorm(n,0,1)  
L = chol(D)  
Z = t(L) %*% X
```

Il ne reste plus qu'à tracer le processus simulé avec l'instruction `plot(x,Z,type="l")`. La figure 4.11 illustre une des simulations obtenues.

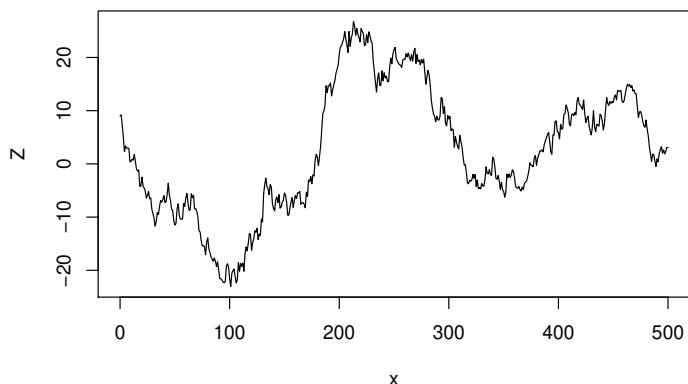


Fig. 4.11. Simulation non-conditionnelle d'un processus stochastique de variogramme linéaire avec palier et portée $h_{max} = 100$. Calcul réalisé par factorisation de Cholesky.

Remarquons que cette simulation n'a été effectuée qu'avec les fonctions de la librairie standard de R. Aucun appel à une bibliothèque externe n'a été requis. Néanmoins, cette apparente simplicité cache un

écueil majeur. Il faut être vigilant aux erreurs d'arrondi numériques. Si la décomposition de Cholesky se passe en général bien avec des variogrammes au comportement linéaire à l'origine, il n'en va pas de même avec des processus à variogramme gaussien et cubique, qui posent souvent des problèmes de stabilité numérique. Parfois, le problème peut être résolu en ajoutant une pépite infinitésimale dans le modèle, mais sans garantie.

Pour faire de la simulation sur des applications concrètes (avec différents variogrammes et avec potentiellement des milliers, voire des millions, de valeurs à générer simultanément, il est fortement recommandé de passer par une bibliothèque, telle que `gstat` ([Pebesma, 2020](#)) par exemple que nous utiliserons dans le cadre de la simulation non-conditionnelle.

En particulier, la simulation non-conditionnelle a permis de générer l'image 2.2, le Modèle Numérique de Terrain réel étant le deuxième à partir de la gauche sur la seconde ligne, relevé sur la Mer de Glace dans le massif du Mont-Blanc. On reconnaît au demeurant ce même terrain sur la figure 3.12 (à gauche). Les 7 autres réalisations de la figure 2.2 sont donc des simulations imitant la structure d'autocorrélation du terrain réel.

Voyons à présent quelques exemples d'applications de la simulation non-conditionnelle.

- **Analyse de sensibilité** : il s'agit d'un domaine d'application fondamental. Prenons un cas formel : on dispose d'une fonction f qui transforme un ensemble \mathbf{x} de données en une sortie $\mathbf{y} = f(\mathbf{x})$. Par exemple, \mathbf{x} peut être une collection de modèles numériques de bâtiments (sous forme de polygones 2D) et \mathbf{y} peut représenter l'aire visible depuis un point d'observation donné, comme illustré sur la figure 4.12 ci dessous.

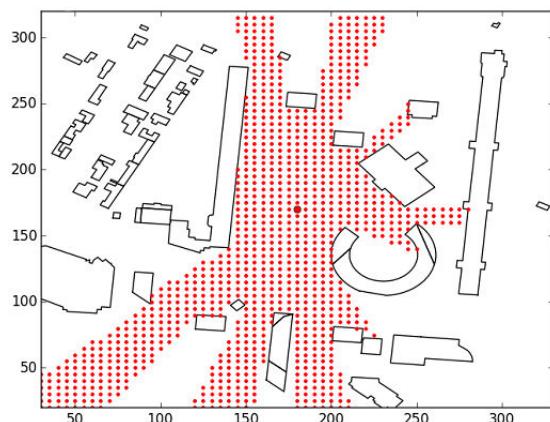


Fig. 4.12. Zone visible depuis un point d'observation. Données IGN BD TOPO[©] sur la commune d'Aubervilliers (93). La surface visible est évaluée à 22 275 m².

Cependant, en pratique, les données \mathbf{x} utilisées sont entachées d'erreurs, et il est important de contrôler l'impact de cette imprécision sur le résultat \mathbf{y} : on parle d'*analyse de sensibilité* de l'application f ([Saltelli et al., 2000](#)). Connaître l'influence du bruit de \mathbf{x} sur une application f est primordial pour (1) quantifier l'incertitude sur le résultat obtenu à l'issue de l'application, (2) déterminer les paramètres et données critiques sur lesquels on doit investir le plus d'efforts afin de les déterminer précisément et (3) déterminer la plage de fonctionnement de f , *i.e.* spécifier la qualité minimale des données utilisées en entrée garantissant le fonctionnement nominal de f .

Il existe dans la littérature trois grandes classes de méthodes pour propager l'erreur de \mathbf{x} à $f(\mathbf{x})$. La première consiste à déterminer les amplitudes de variation $\Delta\mathbf{x}$ sur les entrées, puis à les

transférer sur Δf à l'aide d'une différentielle totale :

$$\Delta f = \left| \frac{\partial f}{\partial x_1} \Delta x_1 \right| + \left| \frac{\partial f}{\partial x_2} \Delta x_2 \right| + \dots + \left| \frac{\partial f}{\partial x_n} \Delta x_n \right| \quad (4.8)$$

où Δx_i désigne l'intervalle de variation de la i -eme composante des données \mathbf{x} .

Exemple 4.1. Entre 11h29 et 11h40, un randonneur parcourt une distance qu'il évalue (à l'aide d'une carte papier et d'un décimètre) à $L = 1 \text{ km} \pm 50 \text{ m}$. On cherche à calculer l'incertitude sur sa vitesse moyenne. On évalue d'abord le temps de parcours : $t = t_2 - t_1 = 11 \text{ minutes}$, avec une incertitude ΔL de 2 minutes (1 min sur chacun de deux relevés chronométriques). Sa vitesse moyenne est $v = f(L, t) = L/t = 5.45 \text{ km/h}$. L'incertitude sur v s'obtient par 4.8 :

$$\Delta v = \left| \frac{\partial f}{\partial L} \Delta L \right| + \left| \frac{\partial f}{\partial t} \Delta t \right| = \left| \frac{\Delta L}{t} \right| + \left| \frac{L \Delta t}{t^2} \right| = 0.27 + 0.50 = 0.77 \text{ km.h}^{-1}$$

Sa vitesse v est donc estimée à $5.45 \pm 0.77 \text{ km/h}$. En particulier, on peut quantifier l'importance relative les deux sources d'incertitude : 35% de l'imprécision est due à l'erreur sur la mesure de distance et 65% à celle sur le relevé du temps.

Cette méthode, simple à mettre en oeuvre, nécessite toutefois que f soit suffisamment régulière pour être approchée par une linéarisation au voisinage du point de fonctionnement. Par ailleurs, une de ses limites majeures et de ne fournir qu'un intervalle absolu Δf , *i.e.* une plage de l'ensemble des valeurs potentiellement prises par f sans tenir compte de leurs probabilités d'apparition respectives, et va donc fournir en général des intervalles très larges et pessimistes.

Une autre méthode consiste à utiliser le théorème 12 de propagation des variances (ou le théorème 13 lorsque f est non linéaire). Ainsi, si f peut être fidèlement représentée par sa jacobienne \mathbf{J} et si la matrice de covariance des entrées est connue, alors on peut évaluer la variance de la sortie de l'application. Cette méthode présente l'avantage de fournir un intervalle probabiliste, donc toujours plus compact (mais plus représentatif) que l'intervalle absolu de la méthode de différentielle totale 4.8. Par ailleurs, elle permet de prendre en compte, via la matrice de covariance, les intercorrélations entre les différentes variables, ce qui est une propriété indispensable lorsque les données en entrées sont de nature géographique. On pourra trouver deux cas d'applications dans les exemples 1.9 et 1.10.

Cependant, si f est fortement non-linéaire (notamment avec des effets de seuil) alors la linéarisation de f n'est pas satisfaisante et la variance de sortie n'est plus réaliste. C'est le cas dans l'exemple donné en figure 4.12 : l'intervisibilité est un phénomène binaire (une zone est visible ou ne l'est pas mais il n'y a pas de situation intermédiaire) donc typiquement non-linéaire. Par ailleurs, lorsque f n'a pas d'expression analytique (c'est le cas ici aussi, et en général dans toutes les applications où f est calculée avec un algorithme complexe), ces deux premières méthodes d'analyse de sensibilité sont inopérantes.

Il reste alors une troisième approche, fondée sur des simulations. On introduit un bruit dans l'entrée \mathbf{x} , et on observe l'impact sur $f(\mathbf{x})$ pour un grand nombre de réalisations. On peut alors évaluer tous les indicateurs statistiques souhaités sur la population de sortie (biais, écart-type,

erreur RMSE, percentiles, valeurs extrêmes...). On peut réitérer l'expérience pour différents niveaux de bruits en entrée. Par exemple, sur la figure 4.13, on ajoute un bruit sur les coordonnées des bâtiments (d'écart-type $\sigma = 5$ m sur la rangée d'images supérieure, et $\sigma = 10$ m sur la rangée inférieure) et on étudie l'impact sur la surface visible calculée.

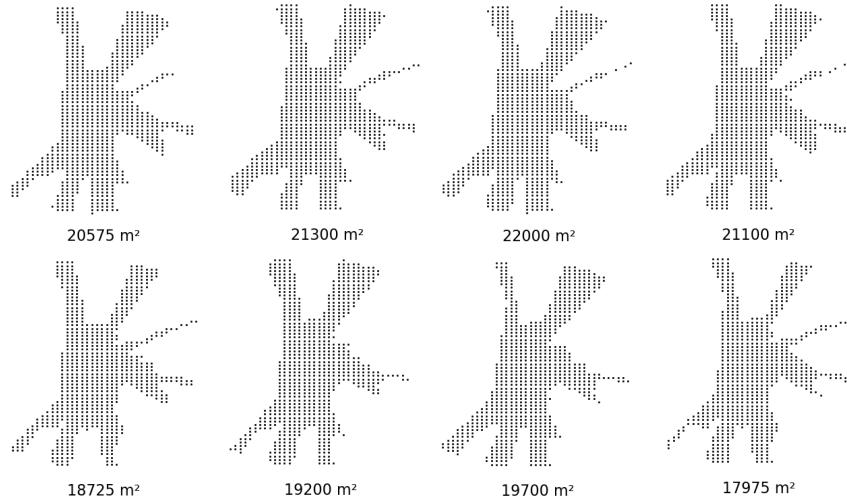


Fig. 4.13. Analyse de sensibilité de la surface visible pour deux niveaux de précision des bâtiments ($\sigma = 5$ m en haut en $\sigma = 10$ m en bas).

Les résultats montrent un RMSE de 1304 m^2 (soit environ 6% de l'aire totale) sur la surface visible pour une précision de 5 m en entrée. Cette valeur passe à 2484 m^2 (11% de l'aire totale) pour $\sigma = 10 \text{ m}$. Dans les deux cas, le biais d'estimation est négatif et représente environ les deux tiers de l'erreur total, ce qui laisse penser que le bruit sur les données induit plutôt une erreur systématique tendant à sous-estimer l'aire visible.

On pourra trouver dans la littérature des études similaires qui cherchent à quantifier l'impact de l'erreur géométrique de modèles de bâtiments sur l'évaluation du potentiel solaire (Biljecki et al., 2015) ou encore sur la prédiction des multi-trajets GPS (Zhang et al., 2018).

Malgré l'apparente simplicité de cette méthode, il faut veiller à ce que le bruit introduit soit représentatif de l'erreur typique entâchant les données d'entrée. Comme expliqué empiriquement sur la figure 1.5, l'utilisation d'un bruit blanc (*i.e.* non-corrélé) pour perturber l'entrée \mathbf{x} , peut mener à une sous-évaluation, ou au contraire à une sur-évaluation de l'impact sur $f(\mathbf{x})$. En particulier, quand f est plutôt de type additive (sommes, cumuls, moyennes, intégrations...) le bruit blanc a une tendance systématique à sous-estimer l'impact. À l'inverse, une fonction de type soustractive (difféentielle, pente, vitesse, accélération, écart...) aura tendance à être plus sensible au bruit blanc qu'à l'erreur typique entâchant les données en réalité. Dans l'exemple 4.13, le bruitage indépendant des coordonnées des sommets n'est pas une solution satisfaisante, et pour deux raisons principales : en premier lieu, le bruit blanc sur les coordonnées a tendance à facilement *bloquer* les lignes de vues. Dans la pratique, les erreurs entâchant les bâtiments sont spatialement corrélées, et les lignes de vues ont globalement tendance à être conservées (au moins jusqu'à la portée de corrélation des erreurs). Deuxièmement d'un point de vue plus pragmatique, l'introduction d'un bruit indépendant sur des données vectorielles (polylinéaires, polygones...), en plus de produire des formes irréalistes, peut résulter en des erreurs topologiques (auto-intersections, chevauchements...), risquant ainsi de planter la fonction f .

Si on dispose d'un modèle de variogramme γ caractérisant le type d'erreurs susceptibles d'entâcher les données, la simulation non-conditionnelle nous permet de générer un nombre arbitrairement

grand de versions bruitées de \mathbf{x} . On donne ci-dessous en figure 4.14 un exemple concret : on suppose que la donnée \mathbf{x} est un ensemble de profils de vitesse de véhicules. Chaque profil de vitesse $v_i : [0, 100] \rightarrow \mathbb{R}^+$ est une fonction qui à un site s d'un axe de route de longueur 100 m, associe la vitesse instantanée $v_i(s)$ du véhicule i à l'instant où il passe la position s . On dispose d'un profil pour chaque véhicule ayant emprunté l'axe sur un intervalle de temps donné. L'application f est un algorithme d'apprentissage statistique ayant été entraîné pour détecter des éléments de signalisation routière (feux tricolores, stops, passages piétons...) à partir de la collection de profils observés (Meneroux, 2019). L'objectif de l'analyse de sensibilité est de déterminer l'impact de la précision des profils sur le taux de détection.

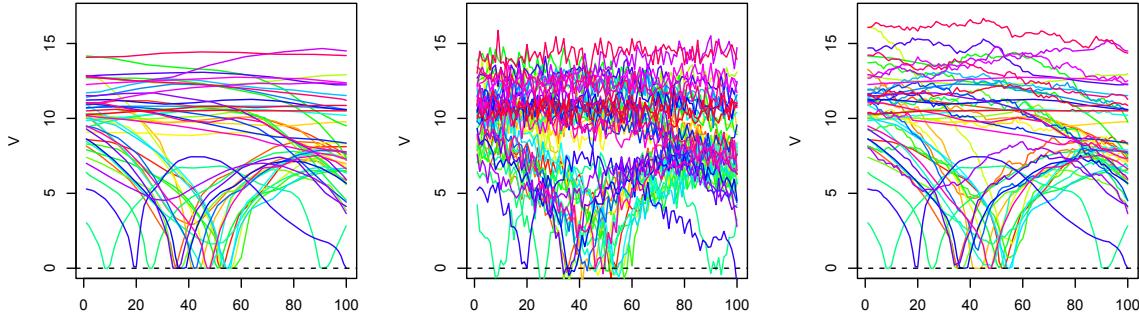


Fig. 4.14. À gauche : profils de vitesses spatiaux observés sur un axe de longueur 100 m. Au centre : perturbation $v(s) + \varepsilon(s)$ où ε est un bruit blanc. À droite : perturbation $v(s) + \varepsilon(s)$ où ε est un processus de bruit de variogramme exponentiel et de portée 600 m.

La figure 4.14 à gauche illustre les profils réels observés. La figure centrale représente les mêmes profils perturbés par un bruit blanc, *i.e.* qu'à chaque position $v(s)$ on ajoute une quantité aléatoire ε_s , indépendante des quantités ε_{s-1} et ε_{s+1} ajoutées aux points voisins de s . On ne peut espérer représenter fidèlement le bruit de mesure avec ce genre de modèle simpliste. Enfin, à droite, on bruite les profils avec bruit corrélé, de variogramme exponentiel d'une portée de l'ordre de 600 m. Les analyses effectuées ont permis de montrer que le taux de détection reste relativement stable jusqu'à un certain niveau de bruit σ_c puis chute ensuite rapidement. La valeur critique σ_c à laquelle s'opère le changement de régime définit la plage de fonctionnement nominal de f .

On pourra trouver de nombreux exemples dans la littérature, par exemple pour le contrôle de qualité des données géographiques linéaires (Vauglin, 1997), pour l'évaluation de la stabilité du calcul d'isochrones (Bonin, 1998), pour le recalage de points GPS sur le réseau routier (Méneroux et al., 2017), ou encore pour la mesure des surfaces sur un MNT (Li et al., 2018).

- **Réalité virtuelle** : dans un domaine plus ludique, on peut citer la génération aléatoire de terrain comme application potentielle de la simulation.
- **Test de corrélation** : pour introduire cette section, considérons un exemple tiré du site humoristique *spurious correlation*⁵. On donne ci-après deux séries temporelles \mathbf{x} et \mathbf{y} relevées sur 11 années consécutives (de 1999 à 2009), où \mathbf{x} désigne le nombre (annuel) d'accidents mortels impliquant un hélicoptère, et \mathbf{y} désigne la différence de points séparant les deux équipes finalistes du championnat de football américain (*Super Bowl*).

⁵<https://tylervigen.com/>

	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
x	59	64	56	48	79	75	42	49	47	69	43
y	15	7	27	3	27	3	3	11	12	3	4

On crée les données correspondantes dans l'environnement R avec les 2 instructions suivantes :

```
x = c(59,64,56,48,79,75,42,49,47,69,43)
y = c(15,7,27,3,27,3,3,11,12,3,4)
```

L'étude de la corrélation linéaire entre ces deux séries (à l'aide de la commande `cor(x,y)`) nous donne un coefficient $\rho = 0.30$, ce qui montre un certains degré de dépendance entre ces deux variables n'ayant pourtant en apparence aucun lien de causalité, ni aucun facteur sous-jacent commun. Peut-on alors imputer cette valeur de corrélation au hasard seul ?

Pour déterminer si un résultat est significatif, la méthode du bootstrap statistique ([Efron et Tibshirani, 1994](#)) possède une redoutable efficacité. Appliquée dans le contexte particulier de notre exemple, le principe en est le suivant : si la corrélation entre les séries **x** et **y** est imputable au hasard seul, alors on peut obtenir facilement (*i.e.* avec une probabilité non-négligeable, typiquement plus de 5%) une corrélation similaire, voire encore plus marquée, avec des données regénérées aléatoirement.

Dans un cadre plus général, pour démontrer la significativité statistique d'un résultat, on comptabilise le nombre de générations aléatoires qui produisent un résultat plus *singulier* ou plus *extrême* que le résultat effectivement observé. Pour régénérer aléatoirement les données, la méthode du bootstrap statistique préconise d'utiliser un tirage avec remise sur les données disponibles.

Concrètement, si **x** est la série de données observées, alors on peut considérer que la série rééchantillonnée avec remise `sample(x, length(x), replace=TRUE)` est une *réalisation alternative*, *i.e.* une série de valeurs, en général distincte de **x** mais possédant les mêmes propriétés statistiques que **x**. Il suffit donc de générer un grand nombre de réalisations de **x**, puis pour chacune d'elles, calculer sa corrélation empirique avec la série **y**, et enfin comptabiliser le nombre de cas pour lesquels la corrélation obtenue est supérieure à la corrélation réelle $\rho = 0.30$.

Nous choisissons ici de générer 1000 échantillons.

```
B = rep(0,1000)

for (i in 1:length(B)){
  B[i] = cor(sample(x, length(x), replace=TRUE), y)
}

nb = length(which(B > 0.3))
```

Remarquons que le choix de perturber le vecteur **x** est arbitraire. Nous aurions également pu perturber le vecteur **y**, ou même les deux vecteurs **x** et **y** simultanément.

A l'issue de l'expérimentation, nous obtenons `nb = 175` réalisations produisant une corrélation plus forte que 0.3, soit 17.5% des cas. Remarquons que si la corrélation est considérée en valeur absolue, alors 36.4% des cas produisent une corrélation plus éloignée de la valeur nulle que 0.3.

On peut donc raisonnablement conclure que la corrélation entre les accidents d'hélicoptères et les différences de scores du *Super Bowl* sur ces 11 années n'est pas significative. Autrement dit, soit la corrélation est fortuite, soit si elle existe vraiment, nous n'avons pas assez de données à disposition pour la mettre en évidence.

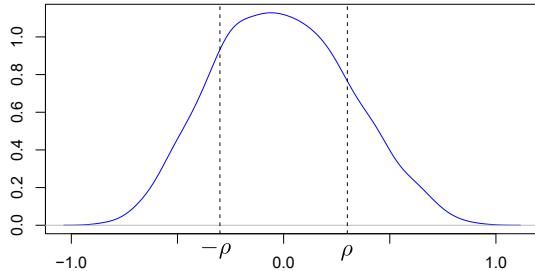


Fig. 4.15. Corrélations relevées sur 1000 simulations aléatoires de bootstrap et comparaison avec la corrélation réellement observée $\rho = 0.30$.

En effet, nous savons intuitivement que plus le nombre de données disponibles est grand, plus les corrélations fortuites sont improbables. À l'extrême, avec un échantillon de taille infini, la théorie nous montre que la corrélation entre deux variables indépendantes est rigoureusement égale à 0.

Cependant, à tailles d'échantillon similaires, la corrélation entre deux variables indépendantes paraît statistiquement plus forte (en valeur absolue) lorsque les phénomènes étudiés sont spatialement autocorrélés. Et plus la portée de cette autocorrélation est longue (relativement à l'emprise de la zone étudiée) plus l'effet est accentué. Dit de manière imagée, deux champs lisses vont apparaître plus significativement corrélés (ou anti-corrélés) qu'ils ne le sont en réalité. C'est ce qu'illustre la figure 4.16 : pour 4 portées de corrélation différentes (de 750 à 6000 m), 2 MNT indépendants ont été simulés et dans chaque cas on indique la corrélation ρ obtenue avec une probabilité de 5% (autrement dit, par exemple, si on simule 100 fois un couple de deux MNT indépendants et autocorrélés sur une portée de 6000 m, alors dans 5 cas la corrélation mesurée sur le couple sera en valeur absolue supérieure à 0.44). À titre de comparaison, la même corrélation calculée sur un couple de réalisations d'un bruit blanc n'excède que rarement 0.02.

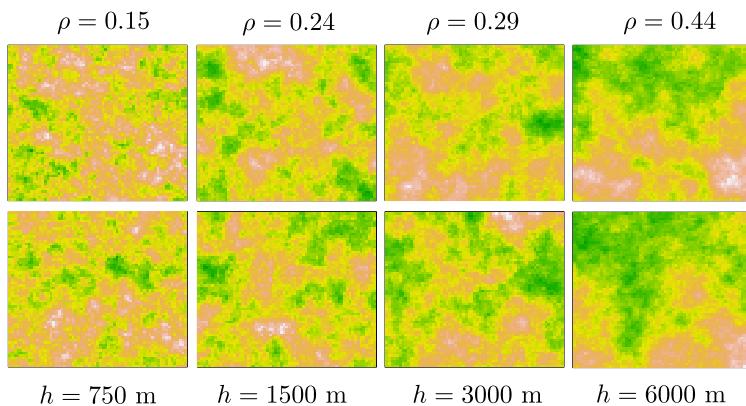


Fig. 4.16. Deux simulations z_1 et z_2 de variables régionalisées pour 4 portées de variogrammes différentes (de 750 à 6000 m), et valeur de corrélation $\rho(z_1, z_2)$ obtenue sur 5% des réalisations. La zone d'étude est un carré de 10×10 km.

Remarquons que l'autocorrélation ne posait pas de problème sur l'exemple considéré ci-dessus. En effet, on peut raisonnablement supposer que le nombre d'accidents mortels d'hélicoptères est

suffisamment rare pour être temporellement décorrélé. Pour le cas du Super Bowl, c'est plus compliqué, les mêmes équipes pouvant statistiquement se rencontrer plusieurs fois successivement en finale, rendant ainsi plus probable la reproduction de scores semblables. En pratique, la représentation graphique de la série y montre l'absence d'autocorrélation significative.

Comment procéder alors lorsque les variables en jeu sont significativement autocorrélées ? C'est précisément dans ce contexte que les techniques de simulation non-conditionnelle vues précédemment peuvent nous venir en aide.

Nous illustrerons la méthode avec un exemple tiré des travaux de stage de [Conte \(2018\)](#) réalisés à l'institut national de l'information géographique et forestière (IGN), et dont l'objectif était de quantifier le changement climatique en comparant la répartition des toponymes du houx à la répartition de l'espèce aujourd'hui. En effet, si l'étymologie d'un toponyme renvoie à une espèce végétale cela signifie probablement que cette espèce se trouvait dans les environs du toponyme à une certaine époque. Un examen toponymique poussé peut donc constituer une source de données ancienne sur la présence du houx dans le passé. Cependant, on se heurte bien vite à une problématique spécifique : la toponymie dépend de la géographie linguistique. Pour obtenir un résultat exhaustif et exploitable sur l'ensemble du territoire métropolitain, il est important au préalable de tester l'indépendance de la répartition des toponymes relatifs au houx (issus de toutes les langues) et la géographie de ces langues.

En pratique, ce questionnement est apparu après comparaison (sur la fig. 4.17) de la prévalence des toponymes liés au houx (à gauche) avec la carte des régions linguistiques (à droite). Notamment, on relève certains similarités de motifs, en particulier au niveau des régions catalane et basque. Sur le reste du territoire, il ne semble pas y avoir de dépendance significative, pour autant, le coefficient de corrélation ρ entre les deux champs vaut 0.71, ce qui semble traduire une dépendance plutôt marquée⁶.

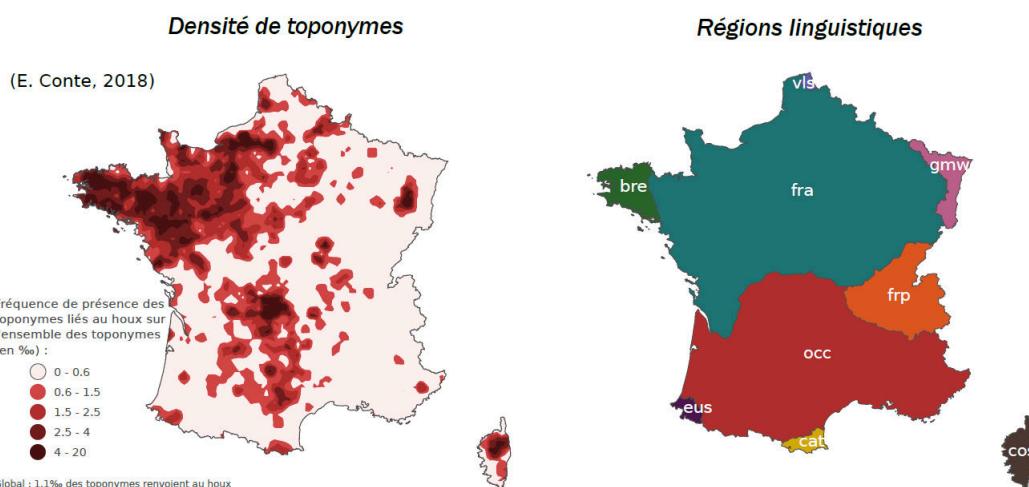


Fig. 4.17. À gauche : fréquence des toponymes relatifs au houx. À droite : régions linguistiques en France métropolitaine. D'après [Conte \(2018\)](#).

Pour tester la significativité de ce résultat, nous avons commencé par calculer le variogramme empirique du champ de prévalence des toponymes, auquel nous avons justé un modèle exponentiel (sans pépite et d'une portée de l'ordre de 1000 km, traduisant ainsi une autocorrélation du

⁶Remarquons que la variable des régions linguistiques étant catégorielle (et non ordinaire), nous avons dû utiliser une variante d'indice de corrélation, mais qui s'interprète de la même manière : de 0 à 1 à mesure que les deux variables étudiées sont mutuellement corrélées.

champ sur l'ensemble du territoire métropolitain). Il est alors possible de simuler des réalisations du processus ; nous en présentons trois (arbitrairement sélectionnées) ci-dessous.

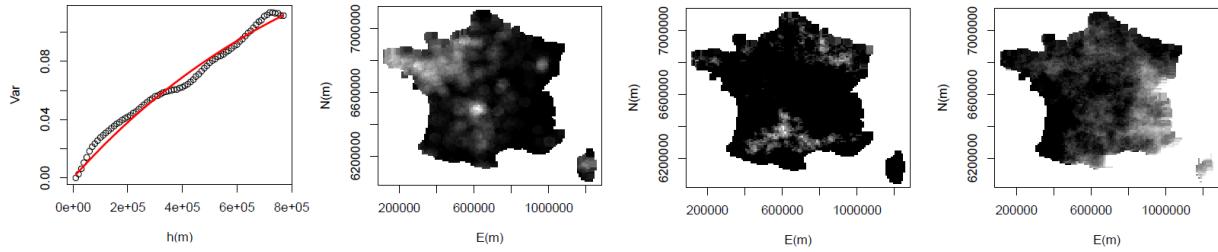


Fig. 4.18. Variogramme empirique et ajustement d'un modèle exponentiel (en haut à gauche) et 3 simulations non-conditionnelles du processus.

Un total de 1000 simulations ont été générées. Pour chaque réalisation, on calcule son indice de corrélation avec la carte des régions linguistiques, et on représente la distribution des valeurs prises par cet indice en figure 4.19. En particulier, on comptabilise 174 cas où la simulation a produit un indice de corrélation supérieur à 0.71, soit une probabilité de 17.4 %. On peut donc en conclure que les champs de régions linguistiques et de fréquence des toponymes du houx ne sont pas significativement corrélés.

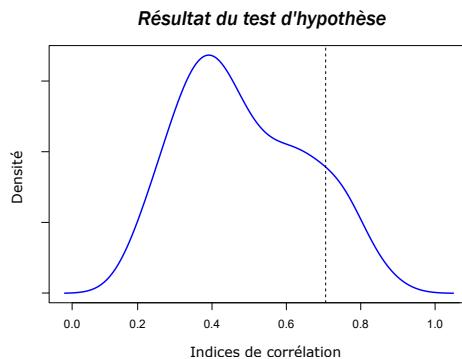


Fig. 4.19. Densité des indices de corrélation sur 1000 simulations aléatoires et comparaison avec l'indice observé (ligne pointillée verticale).

Pour conclure cette section, nous préciserons qu'il existe d'autres méthodes de simulations non-conditionnelles, notamment celle dite des *bandes tournantes*, à base de décomposition LU, ou encore à l'aide des méthodes de *Méthode de Monte-Carlo par chaînes de Markov* (Marcotte, 2003).

4.3.2 Simulation conditionnelle

On parle de simulation conditionnelle lorsque le champ généré aléatoirement est contraint à passer par un semis de points d'observations.

Formellement, on suppose avoir à disposition un semis d'observations $\mathbf{Z} \in \mathbb{R}^n$ en n sites \mathbf{x} , et on cherche à produire des réalisations aléatoires d'un champ Z stationnaire, de fonction de covariance $C(h) = \gamma(\infty) - \gamma(h)$ (la fonction γ étant donnée) entre m sites non-observés \mathbf{x}' et telles que pour tout $i \in \llbracket 1..n \rrbracket$, Z prenne la valeur \mathbf{Z}_i en \mathbf{x}_i .

Pour simuler le champ en m sites, comme pour la méthode non-conditionnelle, on commence par calculer la décomposition de Cholesky de la matrice de covariance Σ (de dimension $n + m$) prise au niveau de **tous les sites** (sites de conditionnement \mathbf{x} et sites à simuler \mathbf{x}') :

$$\Sigma = \begin{bmatrix} C(\mathbf{x}_i - \mathbf{x}_j) & C(\mathbf{x}_i - \mathbf{x}'_j) \\ C(\mathbf{x}'_i - \mathbf{x}_j) & C(\mathbf{x}'_i - \mathbf{x}'_j) \end{bmatrix} = \begin{bmatrix} C & C' \\ C' & C'' \end{bmatrix}$$

en notant C la matrice de covariance entre les n sites conditionnans \mathbf{x} , C'' la matrice de covariance entre les m sites à simuler \mathbf{x}' et C' la matrice de covariance croisée entre ces deux catégories de sites.

On calcule la factorisation de Cholesky de Σ , *i.e.* on trouve la matrice \mathbf{L} telle que :

$$\Sigma = \mathbf{L}\mathbf{L}^T$$

On échantillonne ensuite un vecteur aléatoire \mathbf{A} suivant une loi normale standard multivariée de dimension m et de covariance identité, et on forme le vecteur aléatoire \mathbf{B} , dont on précise la matrice de covariance $\Sigma_{\mathbf{B}}$ (en remarquant bien que la quantité $\mathbf{L}^{-1}\mathbf{Z}$ est déterministe) :

$$\mathbf{B} = \begin{bmatrix} \mathbf{L}^{-1}\mathbf{Z} \\ \mathbf{A} \end{bmatrix} \quad \Sigma_{\mathbf{B}} = \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{I}_n \end{bmatrix}$$

On vérifie alors facilement que le vecteur aléatoire $\mathbf{Y} = \mathbf{LB}$ est une simulation conditionnelle du processus. En effet :

- La matrice de covariance de \mathbf{Y} est (en vertu du théorème 12 de propagation des variances) :

$$\Sigma_{\mathbf{Y}} = \mathbf{L} \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{I}_n \end{bmatrix} \mathbf{L}^T = \begin{bmatrix} 0 & 0 \\ 0 & C'' \end{bmatrix}$$

ce qui correspond bien à la matrice de covariance entre les $n + m$ sites.

- Les n premières valeurs de \mathbf{Y} sont : $\mathbf{Y}_{1:n} = (\mathbf{LB})_{1:n} = \mathbf{L}\mathbf{L}^{-1}\mathbf{Z} = \mathbf{Z}$.

Il est intéressant de constater que la moyenne de k simulations conditionnées à \mathbf{Z} tend vers l'estimateur du krigeage simple 3.7.1 à mesure que k augmente . En effet on a :

$$\mathbb{E}[\mathbf{Y}] = \mathbf{L} \begin{bmatrix} \mathbb{E}[\mathbf{L}^{-1}\mathbf{Z}] \\ \mathbb{E}[\mathbf{A}] \end{bmatrix} = \mathbf{L} \begin{bmatrix} \mathbf{L}^{-1}\mathbf{Z} \\ 0 \end{bmatrix} = \begin{bmatrix} \mathbf{Z} \\ 0 \end{bmatrix}$$

Donc \mathbf{Y} est d'espérance nulle en dehors des sites observés. Or, on sait que la moyenne arithmétique minimise l'écart quadratique moyen aux données, donc la moyenne d'une infinité de simulations conditionnelles est un estimateur interpolant les données, de moyenne nulle en dehors des sites et de variance minimale : c'est donc bien un krigage simple des données. On pourra trouver une illustration graphique de cette propriété sur la figure 3.4.

Par ailleurs, remarquons que lorsque $n = 0$, le vecteur \mathbf{Z} est vide et on retrouve la méthode de simulation non-conditionnelle présentée dans la section précédente.

La figure 4.20 présente le résultat de la procédure de simulation conditionnelle sur un semis de 393 observations collectées sur le MNT de la vallée de Chamonix (jeu de données `sample_data1.txt`).

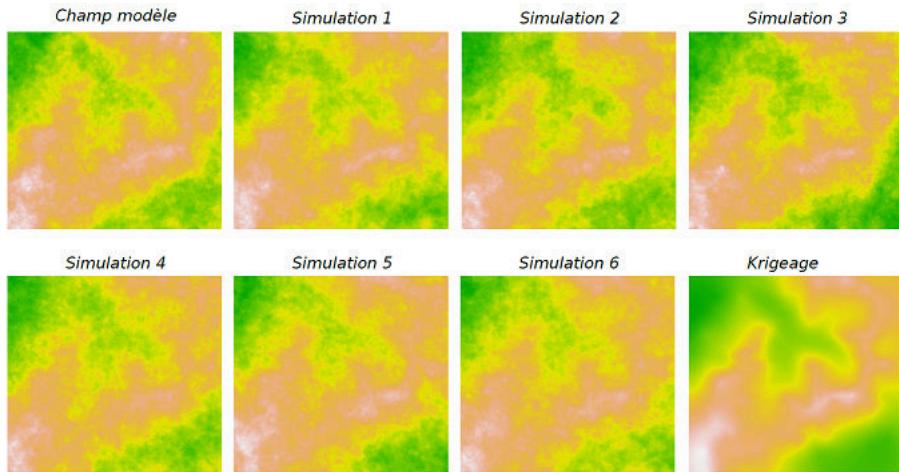


Fig. 4.20. 6 exemples de simulations conditionnelles à partir d'un semis d'observations collectées sur le champ modèle (en haut à gauche). En bas à droite : comparaison avec le résultat obtenu par krigage ordinaire du semis d'observations.

Le résultat obtenu est donc un exemple (parmi une infinité d'autres) de champ respectant à la fois la structure du phénomène et les observations. Contrairement au krigage, la solution obtenue ne possède pas de propriété de minimalité de la variance d'estimation. Une simulation donnée est donc a priori *moins précise* que la solution du krigage. En revanche, la simulation conditionnelle possède l'avantage de fournir une solution différente à chaque tirage. Lorsque les quantités à estimer sont fortement non-linéaires (*e.g.* effets de seuillage) la moyenne des solutions ainsi obtenues est en pratique plus robuste que la solution du krigage. En effet, nous l'avons vu précédemment, si \mathcal{F} est linéaire, alors l'application de \mathcal{F} à l'estimation \hat{x} est égale à l'estimation de $\mathcal{F}(x)$: $\mathcal{F}(\hat{x}) = \widehat{\mathcal{F}(x)}$

Cette propriété est fondamentale puisqu'elle indique que $\mathcal{F}(\hat{x})$ est un estimateur optimale de $\mathcal{F}(x)$, mais elle n'est garantie que si \mathcal{F} est linéaire. Or, en pratique, il existe de nombreux cas où la quantité d'intérêt est une fonction non-linéaire de la variable régionalisée. Un exemple classique est le **seuillage** : lorsqu'on souhaite estimer la surface totale dépassant une certaine valeur (*e.g.* altitude, risque acceptable, teneur en minéraux assurant la rentabilité économique...), alors, l'application \mathcal{F} s'écrit :

$$\mathcal{F}[z] = \int_{\mathcal{D}} \mathbb{1}_{z(x) \geq s} dx$$

où \mathcal{D} est le domaine d'étude et $s \in \mathbb{R}$ est le niveau du seuillage opéré.

Clairement, du fait de la fonction indicatrice $\mathbb{1}_{z(x) \geq s}$, l'application \mathcal{F} est non linéaire, et $\mathcal{F}[\widehat{z}]$ n'est pas un estimateur optimal de $\mathcal{F}[z]$. Autrement dit, le seuillage d'une variable krigée n'est en général pas égal au krigeage de la variable seuillée.

Par ailleurs, du fait de la propriété de lissage du krigeage, le seuillage d'une valeur krigée produira des valeurs systématiquement biaisées. Par exemple, dans le cadre d'une analyse de risque d'inondation en zone côtière, la surface inondable sera systématiquement sous-estimée. Nous verrons un exemple concret par la suite.

La capacité de la simulation conditionnelle à produire des estimations interpolantes moins lissées va ici être mise à contribution pour contourner les faiblesses de la Géostatistique linéaire. Nous allons voir trois exemples typiques (dont l'un sera étudié sous forme d'activité) : l'estimation de la surface d'une île par relevés bathymétriques ([Wackernagel, 2013a](#)), le calcul de la longueur d'un câble sous-marin ([Marcotte, 2003](#)) et l'évaluation d'une surface de danger en zone côtière.

Surface insulaire

Cet exemple est tiré des travaux de [Wackernagel \(2013a\)](#). Au total, 23 mesures bathymétriques ont été réalisées au large de l'Ile-d'Yeu en Vendée, suivant 8 profils rectilignes, comme illustré sur la figure 4.21 (à gauche). On dispose de mesures de profondeurs \mathbf{z} et des coordonnées géographiques \mathbf{x} associées et on suppose que les incertitudes sur ces observations sont suffisamment réduites pour être négligées.

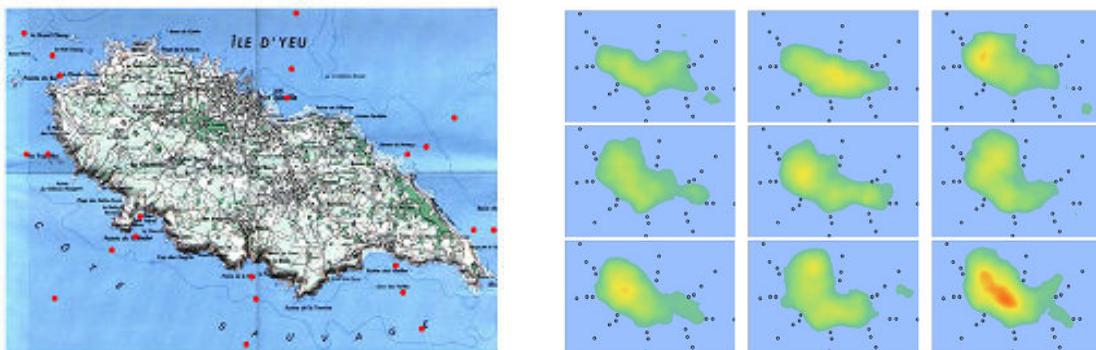


Fig. 4.21. À gauche : 23 mesures bathymétriques effectuées autour de l'Ile-d'Yeu (points rouges). À droite : 9 exemples (arbitrairement choisis) de simulations conditionnelles et seuillage des altitudes négatives (en bleu) pour extraire uniquement l'île. Source : [Wackernagel \(2013a\)](#)

L'objectif du problème consiste à estimer la surface de l'île.

Une première approche pourrait consister à interpoler les mesures bathymétriques (\mathbf{x}, \mathbf{z}) pour reconstruire le plancher marin z . La surface insulaire est alors intuitivement définie comme la surface (plane) totale d'altitude z supérieure à 0. Informatiquement, l'opération consiste donc à calculer un estimateur par krigeage \widehat{z} de l'altitude du terrain sur une grille de résolution r , puis à multiplier le nombre de cellules d'altitude positive par r^2 . Malheureusement, comme nous allons le voir, cette solution simpliste n'est pas satisfaisante : la quantité estimée est non-linéaire et on ne possède donc pas de garantie sur la robustesse de l'estimation.

Une méthode alternative consiste à calculer des simulations conditionnelles de Z , et à évaluer pour chacune d'elles une estimation de la surface insulaire, comme représenté sur la figure 4.21 à droite. Avec cette méthode, la surface totale estimée varie entre 15.4 et 31.9 km² (cette grande variabilité étant imputable au faible effectif du jeu de données). Cependant, en moyennant les estimations calculées sur les simulations, l'auteur de l'étude indique obtenir une valeur de surface de 23.2 km², à mettre en

perspective aux 23.3 km^2 de la vérité terrain. En comparaison, l'estimation directe à partir du krigage résulte en une valeur de 22.9 km^2 , soit une erreur 4 fois plus importante. Cette sous-estimation du krigage s'explique simplement par sa propriété de lissage : l'estimateur \hat{z} est plus lisse que la réalité, ce qui réduit donc la probabilité d'obtenir des valeurs extrêmes. Par ailleurs, la majorité de la zone d'étude étant maritime, la surface insulaire est justement définie par les valeurs extrêmes d'altitude. La proportion de cellules d'altitude positive est donc plus forte en réalité que dans l'estimation krigée. La moyenne de simulations ne présente en revanche pas ce défaut de biais systématique.

Grâce à la simulation conditionnelle, on peut obtenir une estimation fiable de la surface d'une île exclusivement à partir de mesures bathymétriques. À titre d'exercice, on pourra s'interroger sur la marche à suivre pour estimer, à l'inverse, la surface (ou encore le volume ou la profondeur maximale) d'un lac à partir de relevés d'altitudes effectués exclusivement sur terre.

Longueur de câble marin

On reste dans le domaine maritime, avec un exemple tiré du cours de ([Marcotte, 2003](#)) : une compagnie de télécommunication souhaite installer un câble sous-marin sur le plancher océanique, dont le profil est représenté sur la figure [4.22](#). On sait que la distance horizontale entre les deux points à connecter est de $l = 100 \text{ km}$. On souhaite évaluer la longueur de câble nécessaire. Pour ce faire, on effectue des relevés bathymétriques sur 20 sites échantillonnés régulièrement le long du trajet du câble.

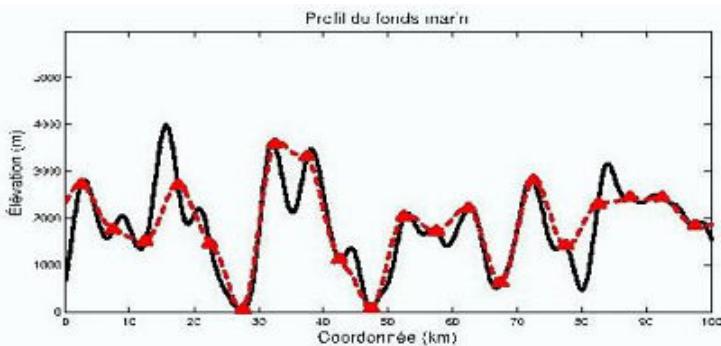


Fig. 4.22. Profil d'élévation du plancher océanique (en noir) et relevés bathymétriques (en rouge) échantillonnés tous les 5 km. Source : [Marcotte \(2003\)](#).

Connaissant le profil $z : [0, l] \rightarrow \mathbb{R}$ du fond marin qui à un lieu x associe la profondeur $z(x)$, la longueur L du câble est très clairement une opération non-linéaire :

$$L = \mathcal{F}[z] = \int_0^l \sqrt{dx^2 + dz^2} = \int_0^l \sqrt{1 + \left(\frac{dz}{dx}\right)^2} dx = \int_0^l \sqrt{1 + z'(x)^2} dx$$

On sait donc que le calcul de la longueur de la variable krigée $\mathcal{F}[\hat{z}]$ est un mauvais estimateur de L . En effet, avec cette méthode, on obtient 104.2 km , contre une longueur réelle de 110 km . La longueur nécessaire a donc été sous-estimée de près de 6%, ce qui peut entraîner des conséquences économiques relativement lourdes pour l'entreprise.

À nouveau, on peut contourner le problème en générant des simulations conditionnelles et en moyennant les longueurs de câble estimées sur toutes les réalisations, ce qui donne une longueur estimée $\hat{L} = 110 \text{ km}$, avec une intervalle de confiance à 95% associé : $[108.8; 113.5]$. L'estimation par moyenne des simulations conditionnelles est donc bien plus satisfaisante.

Activité XII. Évaluation du risque de crue par simulations

Dans cette activité, qui constitue en quelques sortes un travail de synthèse des notions vues dans ce cours, nous allons évaluer la surface inondable d'un terrain donné, dont nous disposons d'un semis de relevés d'altitude `heights.txt` (disponible dans le répertoire associé à ce cours, *cf* en deuxième de couverture). Pour contrôler la qualité de notre estimation, nous utiliserons également le MNT (de résolution $r = 250$ m) de la zone : `mnt.asc`. Ces deux fichiers peuvent être inspectés à l'aide d'un éditeur de text type *bloc-notes*.

Jusqu'à présent, dans un but purement pédagogique, nous n'avons utilisé que les fonctions de base du langage R. En pratique, le traitement numérique des problèmes concrets de Géostatistique présentent une multitude d'ecueils (erreurs d'arrondi, problème de conditionnement des matrices, non-convergence des régressions paramétriques...) rendant ainsi quasi-indispensable le recours à des librairies dédiées qui prennent en charge ces problèmes de manière transparente pour l'utilisateur.

Nous allons utiliser la librairie R `gstat` ([Pebesma, 2020](#)), qui pourra être installée à l'aide de l'instruction `install.packages("gstat")`, puis en sélectionnant un serveur dans la liste proposée. Pour pouvoir facilement manipuler des données géographiques, nous utiliserons également le package `sp` ([Pebesma et Bivand, 2005](#)) : `install.packages("sp")`. Pour activer ces librairies, on placera les deux lignes d'instructions suivantes en entête du code :

```
library("gstat")      # Librairie de Geostatistique  
library("sp")          # Librairie de gestion de donnees spatiales
```

Malgré le soin porté à la mise-à-jour de ce document, il est n'est pas impossible que certaines fonctionnalités de `gstat` présentées ci-dessous ne soient plus disponibles exactement sous le même formalisme syntaxique. On pourra en général facilement résoudre le problème en se référant au manuel d'utilisation ([Pebesma, 2001](#)) le plus récent.

Enfin, précisons que cette activité n'est en principe pas trop sensible au caractère aléatoire des réalisations. Malgré tout, dans un souci de reproductibilité et de débogage éventuel des problèmes, nous mentionnons que les résultats ci-dessous ont été obtenus avec la graine : `set.seed(1)`.

On considère une région côtière, d'altitude comprise entre 0 et 154 m, et soumise à un risque de montée des eaux, dont une analyse préalable a permis de montrer que tous les terrains situés à une altitude inférieure à 5 m pouvaient potentiellement être impactés. L'objectif de l'étude est de déterminer la surface totale (en km²) à évacuer.

Q1. Dans un premier temps, nous allons évaluer cette surface dans l'hypothèse où le MNT de la zone est connu. La valeur trouvée à l'issue de cette étape préliminaire constituera une vérité terrain pour pouvoir comparer par la suite les résultats obtenus respectivement par krigage et par simulations.

On commence par fixer deux paramètres : la résolution du MNT à disposition (`r`) et le seuil d'altitude en dessous duquel un terrain est situé en zone inondable (`threshold`). On définit également une fonction `estimate` permettant de calculer la surface inondable totale d'un MNT.

```
r = 250  
threshold = 5  
  
estimate = function(Z){  
  return(length(which(Z < threshold))*r**2/10**6)  
}
```

Notons que `threshold` est une variable globale. Sa modification entraîne automatiquement celle de la fonction `estimate`. Par ailleurs, soulignons la multiplication par r^2 (qui permet d'exprimer un nombre de cellules en une surface) puis la division par 10^6 (qui assure la conversion en km^2).

Charger le MNT du fichier `mnt.asc` et calculer la surface inondable sur la zone d'étude.

```
S = as.matrix(read.table("mnt.asc"))
svrai = estimate(S)
gx = (1:nrow(S))*r-r/2
gy = (1:ncol(S))*r-r/2
image(gx, gy, S, col=terrain.colors(255))
```

Les vecteurs `gx` et `gy` contiennent les coordonnées (suivant chacune des deux directions) des noeuds de la grille de MNT. On obtient une estimation `svrai` de 30.81 km^2 , soit environ 6% de la zone totale. Par ailleurs, on peut représenter graphiquement la zone inondable en réduisant la valeur d'altitude de tous les sites concernés de manière à marquer plus nettement la distinction dans l'échelle des couleurs, puis en choisissant un code couleur composite :

```
S[which(S<threshold)] = 0
map = c(topo.colors(255)[50:60], terrain.colors(255))
image(gx, gy, S, col=map)
```

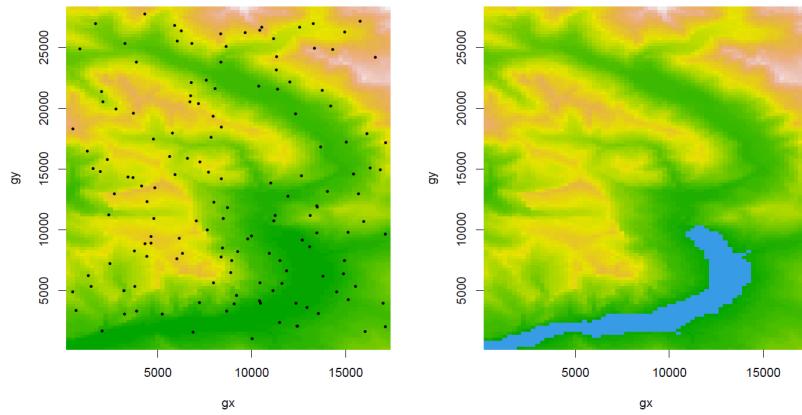


Fig. 4.23. À gauche : le MNT de la zone d'étude avec le semis de points observés. À droite : la zone inondable calculée à partir du MNT, représentant une surface totale de 30.81 km^2 .

Q2. On suppose à présent ne pas avoir de MNT à disposition. L'ingénieur chargé de l'étude décide alors de relever à l'aide d'un GPS professionnel de précision centimétrique un total de 150 points d'altitude répartis aléatoirement et uniformément sur la zone. Le résultat de la campagne est consigné dans le fichier `heights.txt`, dans lequel chaque ligne représente les coordonnées géographiques (X, Y) d'un site, exprimées dans une projection plane quelconque, et l'altitude Z qui y a été mesurée. L'objectif consiste à essayer d'estimer avec la meilleure précision possible, la surface de la zone inondable à partir de ces données réduites.

Charger le semis de points dans l'environnement R, et le représenter graphiquement pour contrôler l'homogénéité spatiale des sites. Calculer (en choisissant judicieusement les paramètres) le variogramme empirique du semis d'observations.

Le chargement du jeu de données ne pose aucun problème avec les fonctions de base R :

```
PTS = read.csv("heights.txt")
```

Pour pouvoir profiter pleinement des fonctionnalités de `gstat`, on doit spécifier que les colonnes `X` et `Y` du tableau `PTS` sont des coordonnées. Au passage, on en profite également pour grouper les observations d'altitudes dans un vecteur `Z` :

```
coordinates(PTS)=~X+Y  
Z = PTS$Z
```

Attention, ici en particulier, le caractère `~` est parfois récalcitrant au copier-coller.

Le choix des paramètres du variogramme se fait à l'aide des notions vues dans la section 2.4.1 : on commence par définir la plage de valeurs sur laquelle on va calculer le variogramme expérimental en prenant la moitié de l'emprise totale du semis de points :

```
D = sqrt((max(PTS$X)-min(PTS$X))**2+(max(PTS$Y)-min(PTS$Y))**2)/2
```

On trouve une valeur `D` de 15.7 km environ. Au delà de cette valeur, le nombre de couples d'observations disponibles est a priori trop faible pour garantir la robustesse statistique de l'estimateur du variogramme. On définit alors un pas de calcul $h = 1$ km, ce qui permet d'avoir un variogramme défini en une quinzaine de points, et permettra ainsi un ajustement de qualité. Par ailleurs, on vérifiera a posteriori que ce choix conduit à disposer d'un nombre suffisant de couples de points dans chaque intervalle. Le calcul du variogramme expérimental est alors immédiat avec `gstat` :

```
vario = variogram(Z~1, data=PTS, cutoff=D, width=1000)
```

Les différents arguments à spécifier sont :

- Une **tendance globale**, permettant de gérer le cas du krigeage à dérive externe, *i.e.* lorsque l'altitude est soumise à une tendance de fond que l'on souhaite prendre en compte dans l'analyse. Ici, le processus est supposé stationnaire, la moyenne de `Z` est donc une constante sur l'ensemble de la zone, ce qu'on exprime avec la formule : `z~1` (qui signifie que la moyenne de `Z` est proportionnelle à la constante 1, indépendante des coordonnées du site).
- Le **jeu de données** utilisé `PTS`, qui doit donc contenir une colonne `Z`, et un nombre arbitraire de colonnes ayant un format de coordonnées.
- La **plage de valeurs** sur laquelle est évaluée le variogramme : `cutoff=D`, (dans l'unité des coordonnées de `PTS`).
- La **Résolution** de calcul (dans l'unité des coordonnées de `PTS`) : ici 1000 m.

On peut alors représenter le variogramme expérimental :

```
plot(vario$dist, vario$gamma, pch=3)
```

Afin de s'assurer que le choix de la résolution $r = 1000$ m est raisonnable, on peut calculer et représenter le variogramme expérimental pour différentes valeurs de pas de discréétisation (fig. 4.24). On voit à nouveau apparaître le compromis biais-variance : lorsque le pas est trop fin, le variogramme est très

résolu mais insuffisamment robuste. À l'inverse, pour un pas trop lâche, le variogramme est précis mais peu résolu.

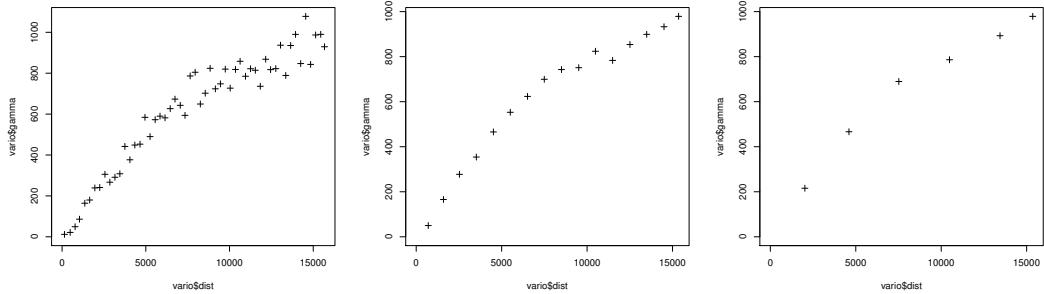


Fig. 4.24. Variogramme expérimental du semis de points d'altitude pour différents pas de discréttisation : 300 m (à gauche), 1 km (au centre) et 3 km (à droite).

Par ailleurs, il peut être instructif de représenter les effectifs de couples d'observations utilisés pour chaque point du variogramme expérimental. Ces nombres sont consignés dans le champ `np` du variogramme estimé, si bien que l'instruction suivante permet la représentation graphique des effectifs :

```
plot(vario$dist, vario$np, type="l")
```

Le résultat obtenu est représenté en figure 4.25 et met en évidence une chute brutale du nombre de couples disponibles à mesure que l'on s'approche de la moitié du diamètre du champ ($\simeq 15$ km). Plus spécifiquement, on pourra considérer que le variogramme expérimental n'est plus suffisamment robuste pour $h > 13$ km.

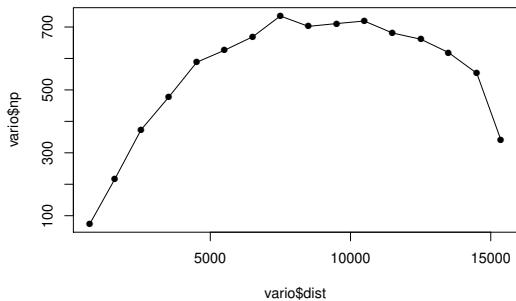


Fig. 4.25. Nombre de couples d'observations pour chaque classe de distance.

Q3. Proposer un modèle de variogramme et estimer ses paramètres.

Après inspection graphique du résultat, il paraît raisonnable de choisir un modèle de variogramme exponentiel (notons qu'un variogramme sphérique, ou même linéaire avec pallier semblent être également des solutions convenables). Dans `gstat`, chaque modèle est défini par un code de 3 lettres. On pourra consulter la liste des modèles disponibles avec l'instruction `vgm()`. La régression paramétrique se fait avec la commande `fit.variogram`, ce qui dans notre cas donne :

```
vmod = fit.variogram(vario, vgm("Exp"))
```

On pourra inspecter le résultatat vmod :

model	psill	range
Nug	0.000	0.00
Exp	1308.601	10797.69

Le résultatat standard de gstat est un variogramme en gigogne, *i.e.* une somme de plusieurs modèles de variogrammes différents (ici une pépite Nug et un modèle exponentiel Exp), avec pour chacun d'eux un paramètre de pallier (psill) et une portée (range). Remarquons que par définition, la portée de la pépite est nécessairement nulle. Par ailleurs, on remarque que le palier de la pépite est également nul, ce qui signifie que le modèle inféré ici est purement exponentiel. On peut représenter le résultatat de la régression (en superposition du variogramme expérimental) en réécrivant l'expression analytique 2.24 du modèle exponentiel avec $C = 1308.601 \text{ m}^2$ et $a = 10797.69 \text{ m}$:

```

h = 0:D
g = vmod$psill[1] + vmod$psill[2]*(1-exp(-h/vmod$range[2]))
plot(vario$dist, vario$gamma, pch=3)
lines(h, g, col="blue")

```

Remarquons que la partie pépitique vmod\$psill[1] n'a aucune influence ici. Le résultatat obtenu est représenté en figure 4.26 (à gauche).

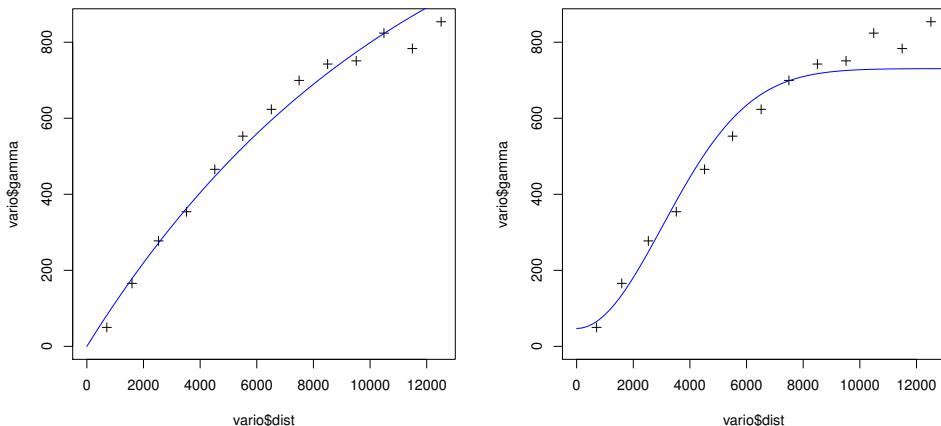


Fig. 4.26. Variogramme expérimental (pour h compris entre 0 et 13 km et un pas de 1 km) et régression paramétrique d'un modèle exponentiel (à gauche) ou gaussien (à droite).

Remarque. Nous avons vu précédemment que la topographie est en général suffisamment lisse pour être caractérisée par un variogramme au comportement parabolique à l'origine. Dès lors se pose la question du choix d'un modèle exponentiel, qui possède la caractéristique notable d'être linéaire à l'origine. En réalité, le comportement à l'origine (d'apparence linéaire) du variogramme expérimental s'explique par la densité trop faible de points d'observations, n'autorisant pas une représentation fine de l'allure du variogramme au voisinage de 0. Si on le souhaite, on peut malgré tout utiliser un modèle gaussien avec l'instruction :

```
vmod = fit.variogram(vario, vgm("Gau"))
```

et la représentation graphique se fait à l'aide de la suite de commandes :

```

h = 0:D
g = vmod$psill[1] + vmod$psill[2]*(1-exp(-(h/vmod$range[2])**2))
plot(vario$dist, vario$gamma, pch=3)
lines(h, g, col="blue")

```

Q4. Calculer par krigeage le relief du terrain à partir du semis de points observés et en déduire une estimation de la surface inondable.

Pour calculer le champ par krigeage, tout comme dans l'activité VIII (`code7.r`), on doit commencer par définir la grille d'interpolation. Afin de travailler sur une zone identique à la vérité terrain, on lui donne les mêmes paramètres (résolution et taille) que le MNT, et donc, nous allons réutiliser les vecteurs `gx` et `gy` définis précédemment dans la question Q1 :

```

GRID = expand.grid(x=gx, y=gy)
gridded(GRID)=~x+y

```

On peut alors effectuer l'estimation par krigeage à l'aide de la fonction `krige` :

```
krigeage = krige(Z~1, PTS, GRID, model=vmod, nmax=50, debug.level=-1)
```

À nouveau, ici aussi, la formule `Z~1` spécifie qu'on se place dans le cadre du krigeage ordinaire (sans dérive externe) et donc que la moyenne du processus Z est constante sur la zone d'étude. Les deux arguments suivants de la fonction `krige` sont : les données observées `PTS`, et l'ensemble des sites sur lesquels on souhaite réaliser l'estimation `GRID`. On spécifie ensuite le modèle théorique de variogramme utilisé `vmod`, puis le nombre maximal d'observations utilisées pour chaque estimation (par défaut, la fonction travaille en voisinage global et effectue chaque estimation avec toutes les observations, ce qui peut s'avérer relativement pénalisant en termes de temps de calcul). Ici, on travaille sur un voisinage glissant de 50 observations. Remarquons qu'il existe d'autres manières de paramétriser ce voisinage, notamment à l'aide d'un critère de distance maximale. Enfin, le paramètre `debug.level=-1` permet d'afficher l'état du calcul dans la console).

Pour représenter graphiquement le champ interpolé, on doit convertir la sortie de la fonction `krige` en une matrice `R` :

```
K = matrix(krigeage$var1.pred, length(gx), length(gy))
```

La représentation graphique se fait alors comme précédemment dans la question Q1, en remplaçant le MNT vérité terrain `S` par l'estimateur de krigeage `K` :

```
image(gx, gy, K, col=terrain.colors(255))
```

De même, le calcul de la surface inondable est immédiat à l'aide de la fonction `estimate` :

```
skrig = estimate(K)
```

On obtient une surface inondable estimée à 16.88 km^2 , soit une sous-estimation d'un facteur 2 par rapport à la valeur vraie `svrai`. On donne en figure 4.27 une représentation du krigeage (à gauche) et de la surface inondable estimée (à droite).

En particulier, on remarque bien que l'estimateur par krigeage est trop lisse par rapport à la réalité, impliquant ainsi une sous-estimation systématique des queues de distributions.

Il peut être intéressant de comparer l'estimateur par krigeage de la surface inondable, avec un simple produit en croix sur les observations, *i.e.* on calcule le ratio du nombre de sites en dessous de l'altitude

seuil, et on le multiplie par la surface totale de la zone d'étude :

```
length(which(PTS$Z < threshold))/length(Z)*max(gx)*max(gy)/10**6
```

On obtient un estimateur de 38.02 km^2 , soit une erreur de l'ordre de 23%, ce qui reste tout de même meilleur⁷ que l'estimateur du krigage.

En présence de fortes non-linéarités (ici c'est le cas puisqu'on calcule la quantité résultant d'un seuillage, par nature non-linéaire), nous avons vu que l'estimateur par krigage n'est plus optimal. On doit donc recourir aux simulations conditionnelles.

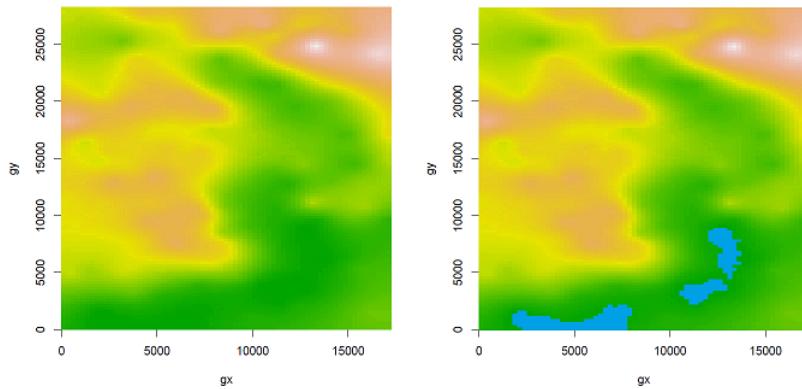


Fig. 4.27. À gauche : l'estimateur par krigage du MNT sur la zone d'étude. À droite : la zone inondable calculée à partir du MNT krigé, représentant une surface totale de 16.88 km^2 .

Q5. Calculer 50 simulations conditionnées au semis de points observés. Pour chaque simulation générée, on évaluera la surface de la zone inondable et on stockera la valeur obtenue. Calculer la moyenne des estimations obtenues (on pourra également dériver un écart-type et des bandes de confiance). Cette nouvelle valeur est-elle plus précise que l'estimateur du krigage ?

Pour effectuer des simulations conditionnelles avec la bibliothèque `gstat`, on utilise exactement la même fonction que pour le krigage à laquelle on ajoute une entrée `nsim` permettant de spécifier le nombre de simulations à calculer :

```
simulation = krige(Z~1, PTS, GRID, model=vmod, nmax=50, nsim=50, debug.level=-1)
```

En fonction des ressources de la machine utilisée, le calcul peut prendre jusqu'à quelques minutes. L'objet R `simulation` contient alors un jeu de 50 simulations réalisées sur `GRID`, de variogramme `vmod` et conditionnées aux observations `PTS`.

On représente en figure 4.28 les 8 premières simulations obtenues. Il faut bien garder à l'esprit que chaque réalisation d'une simulation conditionnelle est en général assez loin de la réalité. La force de la méthode réside dans le nombre, *i.e.* dans la combinaison de l'ensemble des réalisations.

Chaque simulation `i` peut-être récupérée à l'aide de l'instruction `simulation[[names(simulation)[i]]]`. On peut alors évaluer la surface inondable sur chacun des 50 simulations à l'aide d'une boucle :

⁷En réalité, si l'estimateur brut à partir du semis donne de meilleurs résultats que le krigage, cela s'explique par le fait que le tirage des points observés a été effectué suivant une loi rigoureusement uniforme. Si l'échantillonnage avait été plus inhomogène, comme c'est bien souvent le cas puisqu'il dépend de considérations pratiques telles que la facilité d'accès à un site, l'estimateur brut par produit en croix sur les observations aurait été bien plus mauvais.

```

SURFACES = rep(0, 50)

for (repetition in 1:length(SURFACES)){
  SURFACES[repetition] = estimate(simulation[[names(simulation)[repetition]]])
}

```

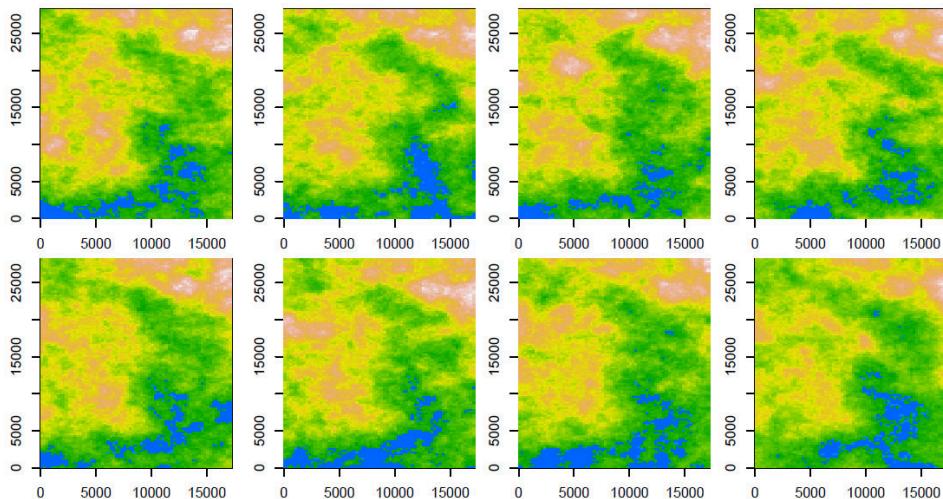


Fig. 4.28. 8 simulations conditionnelles et dans chaque cas, l'évaluations de la surface inondable.

À l'issue de l'exécution de cette boucle, le vecteur SURFACES contient les 50 estimations de surfaces inondables associées aux 50 simulations. On peut alors calculer plusieurs statistiques intéressantes à partir de cet échantillon :

- En premier lieu évidemment, la **moyenne** m du vecteur nous donne un estimateur robuste de la quantité non-linéaire recherchée (à condition que le nombre d'échantillons soit suffisant, ce dont on s'assurera en augmentant le nombre de simulations générées jusqu'à convergence de m) :

```
m = mean(SURFACES)
```

Nous trouvons ici une valeur de 29.29 km^2 , soit une erreur relative de l'ordre de 5%, donc bien plus faible que celle obtenue par le krigeage.

- Par ailleurs, un point positif important de la simulation est de disposer d'une population d'échantillons, ce qui permet d'évaluer finement l'incertitude entâchant la quantité estimée. Il existe deux manières de procéder : la première consiste à supposer que la quantité à estimer est distribuée suivant une loi normale puis à évaluer une bande de confiance en multipliant l'écart-type de la population par un coefficient dépendant du risque d'erreur souhaité (par exemple 1.96 si on souhaite calculer un intervalle de confiance à 95%) :

```
b_inf = m - 1.96*sd(SURFACES)
b_sup = m + 1.96*sd(SURFACES)
```

Cependant l'hypothèse de normalité n'est pas toujours valide, et dans certains cas, elle peut même conduire à des intervalles complètement irréalistes, voire physiquement incohérent (par exemple, on pourrait trouver une borne inférieure négative sur la surface inondable, ou encore

une borne supérieure qui dépasse la surface totale de la zone d'étude).

Dans ce cas, pourquoi ne pas profiter du fait de disposer d'un échantillon de réalisations pour estimer plus finement encore les bandes de confiance à l'aide de quantiles ?

```
b_inf = as.vector(quantile(SURFACES, 0.05))
b_sup = as.vector(quantile(SURFACES, 0.95))
```

Avec cette seconde méthode, on peut dire qu'avec une marge d'erreur de 5%, la valeur vraie de la surface inondable est comprise entre 21.3 et 36.2 km². En particulier, on remarque que l'estimation faite par le krigage est même en dehors de cet interval ! De même soulignons le fait que, à cause de la non-linéarité de la quantité à estimer, le krigage ne permettait pas de donner une incertitude formelle sur la zone inondable.

Un avantage indéniable de cette seconde option est de produire en général des intervalles disymétriques (environ 7 km² d'incertitude vers le haut, et 8 km² vers le bas), ce qui est une propriété intéressante, en particulier dans les cas où la quantité à estimer est proche⁸ de ses bornes physiques).

- Enfin pour synthétiser graphiquement le résultat de toutes les simulations, on peut calculer un histogramme ou une densité des valeurs estimées :

```
d = density(SURFACES)
plot(d$x, d$y, type="l", xlab="surface (km2)", ylab="freq")
```

Le résultat graphique obtenu est représenté en figure 4.29, sur lequel on voit clairement la prédominance de la méthode par simulations par rapport au krigage.

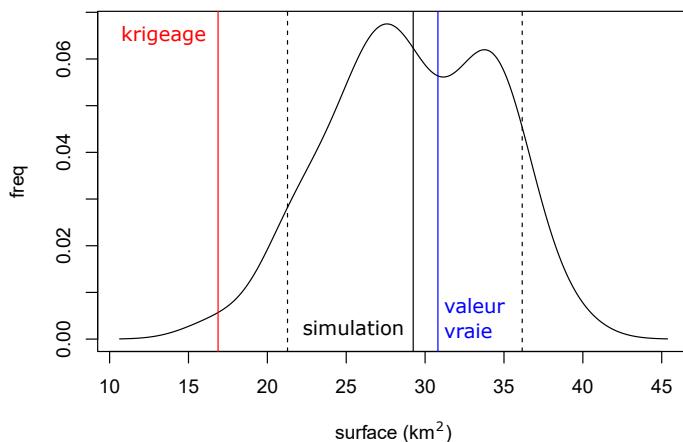


Fig. 4.29. Distribution des surfaces de zone inondable estimées sur les 50 simulations conditionnelles et moyenne des valeurs estimées (trait vertical noir). En rouge : valeur estimée par le krigage. En bleu : valeur vraie. En pointillés : bandes de confiance associées à l'estimation par simulation.

Du point de vue du décideur public, il peut être intéressant également de disposer d'une carte de probabilité d'inondation, *i.e.* un raster indiquant la probabilité que chaque cellule soit située à une

⁸Par exemple, quand la grandeur q à estimer est un pourcentage, si \hat{q} est proche de 0 ou de 100, on s'attend à ce que l'intervalle soit plus lâche du côté opposé à la borne.

altitude inférieure au niveau de crue prédict. Pour ce faire, on commence par générer un nombre plus important de simulations, typiquement 500 :

```
simulation = krige(Z~1, PTS, GRID, model=vmod, nmax=50, nsim=500, debug.level=-1)
```

Pour chaque simulation, on calcule alors une carte d'indicatrice $\mathbb{1}_{z < t}$ valant 1 si la cellule de MNT est située à une altitude z inférieure au niveau t de la crue et 0 sinon. Comme nous l'avons fait remarquer précédemment dans la section 3.7.6, la moyenne d'indicatrices $\mathbb{1}_{z < t}$ sur un grand nombre de réalisations est égale à la probabilité $\mathbb{P}[Z < t]$. On appelle RISQUE la matrice contenant cette valeur de probabilité.

```
RISQUE = matrix(simulation$sim1*0, length(gx), length(gy))
```

```
for (k in 1:500){
  VAL = matrix(simulation[[names(simulation)[k]]], length(gx), length(gy))
  Ik = 0 + (VAL < threshold)
  RISQUE = RISQUE + Ik
}
```

```
RISQUE = RISQUE/500
```

```
image(gx, gy, sqrt(RISQUE-0.001), col=rainbow(255))
```

Dans la représentation graphique, nous opérons la transformation $p \mapsto \sqrt{p - 0.001}$ pour deux raisons : (1) pour diluer le bleu sur l'ensemble des zones de forte probabilité d'inondation (typiquement quand p est supérieure à 50% de risque) et (2) pour représenter en blanc les zones pratiquement dépourvues de risques ($p < 0.1\%$). On donne le résultat graphique obtenu en figure 4.30.

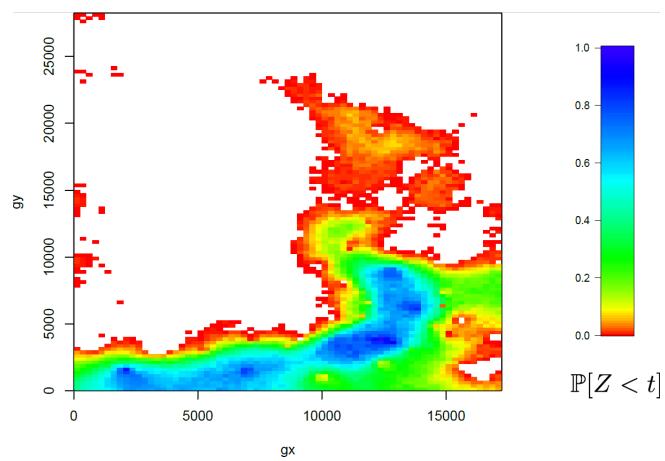


Fig. 4.30. Risque d'inondation pour une crue prédict à 5 m. La probabilité d'inondation évolue de 0 (cellules rouges) à 1 (cellules bleues). Les cellules blanches ont un risque nul d'être impactées.

Par exemple, supposons qu'un site stratégique (refuge, cellule de crise, site d'approvisionnement, centrale nucléaire, etc.) soit situé au point de coordonnées ($X = 11000$ m, $Y = 10000$ m). Une division par la résolution $r = 250$ m nous donne immédiatement les indice (44,40) du point dans la grille, et la valeur $\text{RISQUE}[44, 40] = 0.138$ nous indique alors que le site en question a une probabilité d'être en zone inondable de l'ordre de 14%.

Il est important de remarquer que, jusqu'à présent l'incertitude sur le fait d'être en zone inondable résulte uniquement d'un manque de connaissance du MNT. Dans un cadre plus réaliste, on doit ajouter

à cette incertitude, celle liée à la difficulté de prévision. En effet, le niveau de la crue n'est pas connu exactement en amont de l'évènement, en général on ne dispose que d'une estimation avec un degré d'incertitude associé. Dans certains cas, les études hydrographiques réalisées en amont sont suffisamment poussées pour permettre d'associer une loi de probabilité à cette variable inconnue, *e.g.* : le niveau atteint par la crue sera distribué uniformément entre 0 et 30 m... On est à présent dans un cadre bayésien, autrement dit, la variable de niveau de crue qui jusqu'à présent était considérée comme un paramètre (fixé à 5 m) est à présent une variable aléatoire. Cela ne pose aucun problème en simulation.

Une première méthode peut consister à générer des cartes de risque R_t pour un grand nombre de niveaux de crue $t = t_1, t_2, \dots$ différents, puis à calculer la moyenne de ces cartes pondérées par la probabilité $p(t)$ d'avoir une crue de niveau t :

$$R = \mathbb{P}[Z \leq t] = \int_{t=-\infty}^{+\infty} \mathbb{P}[Z \leq t] p(t) dt \approx \frac{\sum_k p(t_k) R_{t_k}}{\sum_k p(t_k)} \quad (4.9)$$

La carte R obtenue indique alors, en chaque lieu de la zone, la probabilité totale d'inondation (tenant en compte tous les niveaux de crues possibles).

Une seconde méthode, plus directe en termes de développement, exploite encore un peu plus la stratégie de type Monte-Carlo, en tirant le niveau de crue aléatoirement et pour chaque simulation suivant sa distribution p . Par exemple, si le niveau de crue est distribué uniformément entre 0 et 30 m, le code devient :

```
RISQUE = matrix(simulation$sim1*0, length(gx), length(gy))
T = runif(500,0,30)

for (k in 1:500){
  VAL = matrix(simulation[[names(simulation)[k]]], nx, ny)
  Ik = 0 + (VAL < threshold)
  RISQUE = RISQUE + Ik
}
RISQUE = RISQUE/500

image(gx, gy, sqrt(RISQUE-0.001), col=rainbow(255))
```

On donne en figure 4.31 le résultat de cette analyse bayésienne pour 4 scénarii différents sur la prédiction a priori du niveau de la crue. Dans le premier cas (a) on considère que la niveau de crue est parfaitement connu et égal à 15 m : formellement, il est donc distribué suivant une loi de dirac δ_{15} . Dans les trois autres cas (b à d), il est aléatoire et distribué respectivement suivant une loi uniforme entre -10 et 40 m, suivant une loi normale de moyenne et d'écart-type 15 m et suivant une loi exponentielle de moyenne 15 m.

Remarquons que dans ces trois derniers cas, la variable aléatoire de niveau de crue a approximativement la même moyenne et le même écart-type, autrement dit les trois scénarii pronostiquent le même niveau de crue en moyenne, et sont entâchés de la même quantité globale d'incertitude. Pour autant, les cartes de risques générées sont significativement différentes : en particulier, la loi exponentielle (d) étant une distribution *à queue lourde* (et donc favorisant l'apparition d'évènements rares), elle rend possible l'impact de la crue en des zones systématiquement épargnées par les autres distributions. À l'inverse, le risque de crue à basse altitude (typiquement dans la vallée) y est proportionnellement moins important.

De même, si on reprend le site stratégique (situé aux coordonnées $X = 11000$ m, $Y = 10000$ m), on obtient des probabilités d'inondation relativement différentes en fonction du modèle de prédition de crue utilisé : 43% avec les modèles déterministe et uniforme (a et b), 47% avec le modèle normal et 36% avec le modèle exponentiel.

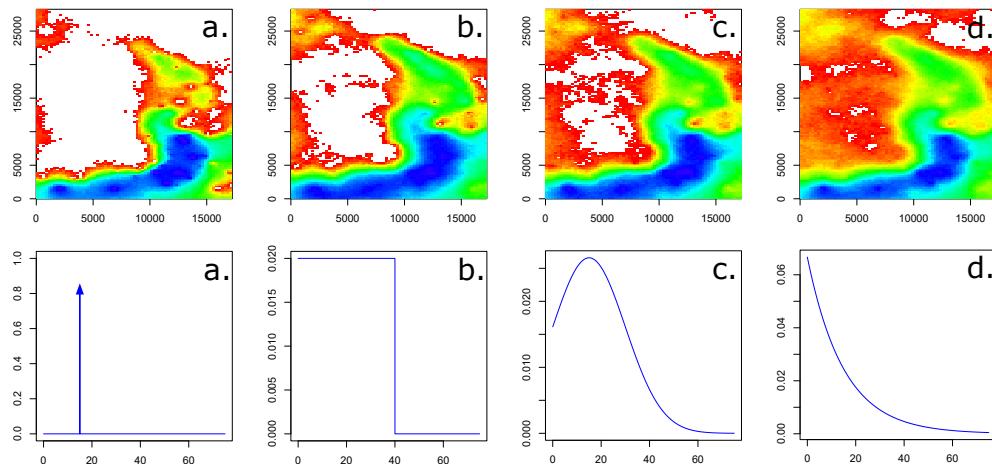


Fig. 4.31. Cartes de risque d'inondation (en haut) pour 4 prédictions hydrographiques différentes (en bas), suivant que le niveau de la crue est : (a) déterministe δ_{15} , (b) distribué suivant la loi uniforme $\mathcal{U}([-10; 40])$, (c) distribué suivant la loi normale $\mathcal{N}(15, 15^2)$ ou (d) distribué suivant une loi exponentielle $\mathcal{E}(1/15)$.

Un atout important de la simulation est qu'une même population de réalisations peut permettre de répondre à une large gamme de questions. En particulier, on peut évaluer des quantités qui ne possèdent même pas d'expression analytique simple, comme par exemple : déterminer la probabilité d'apparition d'enclaves après la crue, calculer la loi de probabilité de la plus grande surface épargnée, ou encore évaluer l'allongement moyen des temps de trajets en prenant en compte la coupure de ponts, etc. On illustre ce propos avec un exemple simple, en considérant une ligne électrique dont on donne ci-dessous les coordonnées planes du tracé (grossier) :

```
X = c(0, 5, 12, 15, 15.5, 17.5)*10**3
Y = c(24, 24.75, 20.3, 15, 14.5, 12)*10**3
```

Dans un premier temps, on affine la précision de la ligne en sur-échantillonnant ses coordonnées :

```
LINE = approx(X, Y, xout=seq(0, 17000, 100))
```

où la valeur 17000 représente la coordonnée X maximale de la zone et la valeur 100 représente le nombre de points souhaités en sortie. À l'issue de la procédure d'interpolation, l'objet `LINE` contient les coordonnées sur-échantillonnées `LINE$x` et `LINE$y` de la ligne. On peut alors convertir ces coordonnées en indice de cellules du MNT, en les divisant par la résolution r :

```
LX = floor(LINE$x/r)+1
LY = floor(LINE$y/r)+1
```

On suppose dans un premier temps que la ligne est positionnée à une hauteur de 3 m par rapport au sol :

```
hgt_line = 3
```

On peut alors écrire une fonction `coupure` permettant de tester si la ligne est impactée par la crue. Pour ce faire, on parcourt toutes les coordonnées de la ligne (d'où l'intérêt de l'avoir échantillonnée

suffisemment finement), et pour chaque point, on récupère son altitude dans le MNT et on y ajoute la hauteur de ligne `hgt_line`. Si pour au moins un point cette valeur est inférieure au niveau de la crue, alors la ligne est coupée.

```
coupure = function(Z){
  for (i in 1:length(LX)){
    if (Z[LX[i], LY[i]] + hgt_line < threshold){
      return(TRUE)
    }
  }
  return(FALSE)
}
```

Dans cette dernière section, on suppose à nouveau que le niveau de la crue `threshold` est déterministe et fixé à 5 m. On applique alors la fonction `coupure` sur chaque simulation générée et on calcule la proportion de cas où la ligne est coupée.

```
COUPURE = rep(FALSE, 500)

for (k in 1:500){
  Z = matrix(simulation[[names(simulation)[k]]], length(gx), length(gy))
  COUPURE[k] = coupure(Z)
}

proba_coupure = sum(COUPURE)/length(COUPURE)
```

Avec 500 simulations, on trouve une probabilité de coupure de 17.6 %. Pour complexifier l'exercice, on peut chercher à déterminer la hauteur d'installation `hgt_line` de la ligne (par rapport au sol) permettant de garantir que le risque de coupure en cas de crue est inférieur à 5%. Sur le même jeu de simulations, on trouve que la ligne doit être installée à une hauteur minimale de 9.10 m.

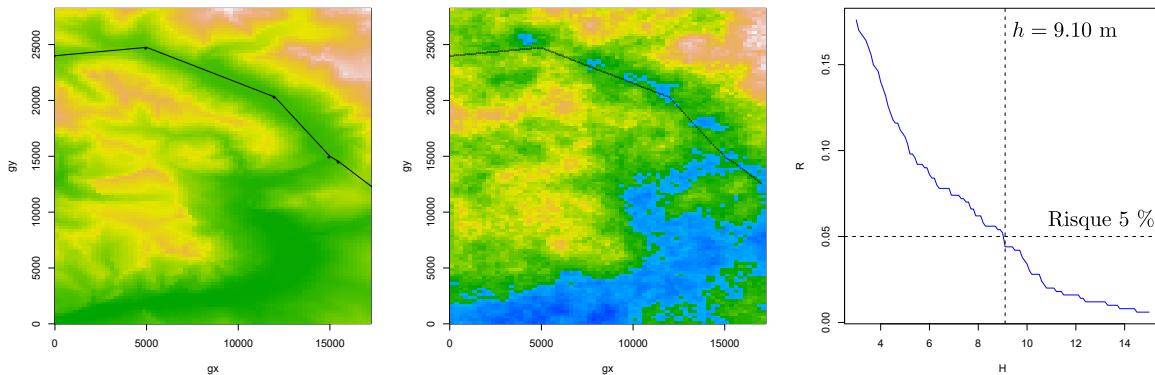


Fig. 4.32. À gauche : MNT réel et tracé de la ligne électrique. Au centre : un exemple de simulation avec la crue associée et discrétisation de la ligne électrique. À droite : probabilité de coupure de la ligne en fonction de sa hauteur par rapport au sol. L'installation est considérée comme sûre (risque inférieur à 5%) pour une hauteur de 9.10 m.

4.4 Bilan

En comparaison des méthodes déterministes, le krigeage possède l'avantage de permettre le calcul d'une variance de l'erreur d'estimation, qui peut être interprétée comme le degré d'incertitude entourant la prédiction effectuée. Cependant, si le modèle de variogramme utilisé n'est pas satisfaisant (variogramme expérimental mal estimé, choix non pertinent du modèle théorique de variogramme ou non-convergence de la régression paramétrique) on obtient en général des variances d'estimations complètement faussées et qui peuvent indûment faire présumer que l'estimation est d'une qualité meilleure qu'elle ne l'est réellement. Il est donc nécessaire de disposer de méthodes objectives de validation des prédictions effectuées par krigeage. Dans le cadre particulier de l'interpolation, la **validation croisée** consiste à réestimer les valeurs prises par le phénomène au niveau de chacun des points du semis d'observations, et à l'aide de toutes les autres données. La différence entre la valeur prédite et la valeur effectivement observée, moyennée sur l'ensemble de la zone, produit un bon estimateur de la qualité globale de la procédure.

Dans un seconde partie, nous avons étudié une application alternative du krigeage : l'**optimisation bayésienne** (ou séquentielle), qui consiste à alterner des étapes d'interpolation et d'observation pour déterminer l'optimum d'une fonction numérique f en minimisant le nombre de relevés nécessaires. À chaque étape, la méthode calcule une estimation \hat{f}_n de la fonction inconnue à l'aide des n points de données observés dans le passé. La position du $(n + 1)$ -eme point à observer est déterminée à l'aide de l'estimateur \hat{f}_n et de ses bandes de confiance, et on itère le processus jusqu'à un nombre maximal préfixé d'observations, ou alors jusqu'à atteindre un optimum satisfaisant. L'optimisation bayésienne est un outil de choix lorsque les observations sont coûteuses, chronophages ou risquées.

Enfin, dans une troisième section, nous avons introduit le concept important de simulations. Une **simulation conditionnelle** est une réalisation aléatoire d'un processus stochastique (de variogramme connu) passant par tous les points du jeu de données. Il s'agit donc en quelques sortes d'une réalité alternative compatibles avec les observations. Lorsque le semis de points est vide, la simulation n'est pas contrainte, et on parle de **simulation non-conditionnelle**. Nous avons vu comment effectuer ces deux types de simulations, à l'aide de la factorisation de Cholesky ou encore par krigeage séquentiel. Elles possèdent deux applications fondamentales : l'analyse de propagation d'erreurs et l'estimation de **quantités non-linéaires**. Dans ce second cardre en particulier, nous avons vu à travers de nombreux exemples, que le krigeage est systématiquement plus lisse que la réalité, induisant ainsi un biais dans l'estimation des dépassements de seuil. Nous avons vu dans le chapitre précédent que de nombreux estimateurs s'expriment de manière linéaire (somme, différence, moyenne, cumul, pente, convolution...), ce qui est remarquable eut égard au fait que ces fonctions linéaires ne représente qu'une infime partie de l'ensemble des fonctions calculables sur une variable régionalisée. En pratique, la majorité des grandeurs qui vont nous intéresser *in fine*, sont non-linéaires : surface dépassant un seuil, pente maximale, probabilité de coupure d'un réseau électrique par une inondation, étendue visible depuis un point d'observation... L'évaluation direct de ces quantités sur la variable krigée ne sera en général pas un bon estimateur. La technique de simulation conditionnelle, qui préconise de moyenner la quantité d'intérêt obtenue sur un grand nombre de réalisations, offre une solution simple à ce problème, permettant de s'épargner le recours à la Géostatistique non-linéaire.

Appendix A

Régularité d'un processus stochastique

Étudier la régularité d'un processus stochastique Z (en particulier sa continuité et sa dérivabilité) permet de se faire une idée de la régularité de ses réalisations z . Il faut avoir les idées claires : il s'agit bien d'une notion probabiliste. Dire qu'un processus stochastique est continu ne signifie pas que toutes ses réalisations sont systématiquement continues en tout point, comme le montre le contre-exemple B.7. L'implication est plus modeste et de manière informelle on pourrait l'énoncer comme suit : en un point donné du domaine, une réalisation va probablement être continue. Dans cette annexe, pour la clarté de l'exposé, nous nous limitons au cas d'un processus uni-dimensionnel X . La généralisation aux dimensions supérieures ne pose pas de problème¹.

A.1 Définitions

Une fonction réelle f est continue si la valeur qu'elle prend en deux points infinitésimamente proches tendent à devenir égales, *i.e.* si $f(t + \Delta t)$ tend vers $f(t)$ à mesure que Δt tend à s'annuler.

Pour une réalisation particulière x d'un processus X , cette définition a un sens, mais on souhaiterait être plus général, et pouvoir caractériser la capacité du processus X à produire des réalisations continues. On doit donc vérifier que la variable aléatoire $X(t + \Delta t)$ converge vers $X(t)$ en tout point t du domaine. Or, pour une variable aléatoire, il existe de nombreux modes de convergences différents (Lecoutre, 2002) : convergence en loi, en probabilité, presque sûre, en moyenne d'ordre r ... Pour caractériser la régularité des signaux aléatoires, on utilise en général la convergence en moyenne d'ordre 2, dite *en moyenne quadratique*, consistant à vérifier que l'espérance du carré de l'écart entre la variable et sa limite tend vers 0.

Continuité en moyenne quadratique. *Un processus stochastique $X : \mathbb{R} \times \Omega \rightarrow \mathbb{R}$ est dit continue en moyenne quadratique (m.q.) si, et seulement si, pour tout point $t \in \mathbb{R}$:*

$$\lim_{\Delta t \rightarrow 0} \mathbb{E}[(X(t + \Delta t) - X(t))^2] = 0 \quad (\text{A.1})$$

De la même manière, on peut définir la notion de dérivabilité en moyenne quadratique :

Dérivabilité en moyenne quadratique. *Un processus stochastique $X : \mathbb{R} \times \Omega \rightarrow \mathbb{R}$ est dit dérivable en moyenne quadratique si, et seulement si, pour tout point $t \in \mathbb{R}$:*

$$\lim_{\Delta t \rightarrow 0} \mathbb{E}\left[\left(\frac{X(t + \Delta t) - X(t)}{\Delta t}\right)^2\right] \quad (\text{A.2})$$

¹Notamment en se ramenant à des processus 1D avec des dérivées directionnelles.

existe et est une quantité finie $Z'(t)$: la dérivée du processus Z en t .

Remarquons que si Z est dérivable en moyenne quadratique, Z' est aussi un processus stochastique. On montre facilement que si Z est stationnaire, alors sa dérivée Z' l'est aussi. La réciproque n'est pas nécessairement vraie : il suffit de considérer l'exemple d'une marche aléatoire Z (non-stationnaire) dont les incrémentations Z' sont stationnaires.

A.2 Liens avec le variogramme

Les notions de continuité et de dérivabilité étant intrinsèquement locales, il n'est pas surprenant de le voir étroitement liées avec le comportement du variogramme (et de la fonction de covariance) au voisinage de l'origine (*i.e.* pour des distances infinitésimales entre sites).

Dans cette section, nous considérerons la cas particulier d'un processus stationnaire à l'ordre 2 et nous nous autoriserons donc à utiliser la fonction de covariance. Notons que les mêmes résultats peuvent être obtenus directement à partir du variogramme, c'est-à-dire pour un processus dont la stationnarité est strictement intrinsèque.

Théorème : *un processus stochastique X est continue en moyenne quadratique si, et seulement si, son variogramme γ est continu en 0.*

La démonstration est immédiate à partir de la définition A.1 et de l'expression 2.8 du variogramme :

$$\mathbb{E}\left[\left(X(t + \Delta t) - X(t)\right)^2\right] = 2\gamma(\Delta t)$$

et donc, la quantité A.1 converge quand $\Delta t \rightarrow 0$ si, et seulement si, $\gamma(t)$ tend vers $\gamma(0) = 0$, *i.e.* si le variogramme γ est continu en l'origine.

Par application de ce théorème, tout variogramme muni d'une composante pépitique (non-nulle) est irrévocablement discontinu en moyenne quadratique.

Il est intéressant de noter que la règle est exactement la même pour les processus stationnaires à l'ordre 2 : d'après le théorème 26, $C(h) = C(0) - \gamma(h)$, et donc $\gamma(h)$ tend vers 0 si et seulement si $C(h)$ tend vers $C(0)$, et donc la fonction de covariance doit être continue en l'origine.

Remarque. On peut montrer que le variogramme est continu en 0 si, et seulement si, il est continu sur \mathbb{R} . En effet, considérons la différence des valeurs prises par le variogramme pour deux écarts τ et τ' :

$$\begin{aligned} |\gamma(\tau) - \gamma(\tau')| &= |C(0) - C(\tau) - C(0) + C(\tau')| = |C(\tau) - C(\tau')| = \text{Cov}[X(\tau), X(0)] - \text{Cov}[X(\tau'), X(0)] \\ &= \text{Cov}[X(\tau) - X(\tau'), X(0)] \leqslant \sqrt{\text{Var}[X(\tau) - X(\tau')] \text{Var}[X(0)]} = \sigma_X \sqrt{2\gamma(\tau - \tau')} \end{aligned}$$

et donc, si $\gamma(\tau - \tau')$ tend vers 0 quand la quantité $\tau - \tau'$ s'annule, nécessairement $\gamma(\tau)$ tend vers $\gamma(\tau')$, ce qui montre la continuité de γ sur \mathbb{R} .

Considérons à présent un processus X continue en m.q. et examinons les conditions sous lesquelles la dérivabilité de X est garantie.

Théorème : le processus stochastique X est dérivable en m.q. si, et seulement si, son variogramme γ est deux fois dérivable en 0.

Pour la preuve, on considère le processus X' , différence finie de X :

$$X'(t) = \frac{X(t + \Delta t) - X(t)}{\Delta t}$$

Δt est une quantité finie, donc on peut utiliser X' même sans avoir au préalable démontré la dérivabilité de X . Par ailleurs X' est un accroissement, i.e une combinaison linéaire de variables aléatoires de poids total nul, on peut donc utiliser la covariance, même dans le cas intrinsèque :

$$\begin{aligned} C_{X'}(\tau) &= \text{Cov}[X'(t + \tau), X'(t)] = \frac{1}{\Delta t^2} \text{Cov}[X(t + \Delta t + \tau) - X(t + \tau) - X(t + \Delta t) + X(t)] \\ &= \frac{1}{\Delta t^2} \text{Cov}[X(\Delta t + \tau) - X(\tau) - X(\Delta t) + X(0)] = \frac{C_X(\Delta t + \tau) - 2C(\tau) + C_X(\tau - \Delta t)}{\Delta t^2} \end{aligned}$$

La fonction de covariance $C_{X'}(\tau)$ est donc égale à la différence finie d'ordre 2 de la covariance de X . Pour que la dérivée de X existe, il faut pouvoir faire tendre Δt vers 0. La différence finie d'ordre 2 tend donc vers une dérivée seconde, et le processus X est dérivable en m.q. si et seulement si sa fonction de covariance est deux fois dérivable. Le variogramme doit donc vérifier cette même propriété.

Remarque : tout comme pour la continuité, on peut montrer que cette condition est équivalente à la dérivabilité d'ordre 2 du variogramme en 0.

A.3 Régularisée d'un variogramme

Dans cette section, nous établissons un lien entre la notion de variogramme régularisé et la régularité du processus sous-jacent.

On se place en 1D sur la droite réelle et on considère un processus Z de variogramme linéaire $\gamma : h \mapsto h$. Soit $V = [0, 1]$ un sous-domaine de \mathbb{R} . Calculons le variogramme régularisé sur V (on ne considérera que le cas $h > 0$ puisque le variogramme est symétrique) : $\bar{\gamma}_V(h) = \bar{\gamma}(V, V_h) - \bar{\gamma}(V, V)$ avec :

$$\bar{\gamma}(V, V_h) = \frac{1}{|V||V_h|} \int_V \int_{V_h} \gamma(x - y) dx dy = \int_0^1 \int_0^1 |y + h - x| dx dy$$

Par ailleurs, en supposant dans un premier temps $h < 1$ (i.e. que les blocs V et V_h se recouvrent) on peut décomposer le domaine d'intégration en 3 parties en fonction du signe de $y + h - x$:

$$\bar{\gamma}_h(V) = \int_0^h \int_0^1 (y + h - x) dy dx + \int_h^1 \int_{x-h}^1 (y + h - x) dy dx + \int_h^1 \int_0^{x-h} (x - y - h) dy dx$$

On peut résoudre chaque intégrale individuellement et sommer les résultats obtenus. Alternativement, on peut considérer le problème de manière purement géométrique. L'objectif consiste à calculer le volume contenu entre le plan $z = 0$ et la surface d'équation $z(x, y) = |y + h - x|$. Remarquons que sans la valeur absolue, l'intégrale recherchée vaut h . Pour tenir compte de la valeur absolue, on doit ajouter à cette valeur moyenne deux fois le volume de la pyramide qui a été comptabilisée en dessous du plan $z = 0$. La base de cette pyramide est un triangle rectangle isocèle de côté et de hauteur $1 - h$.

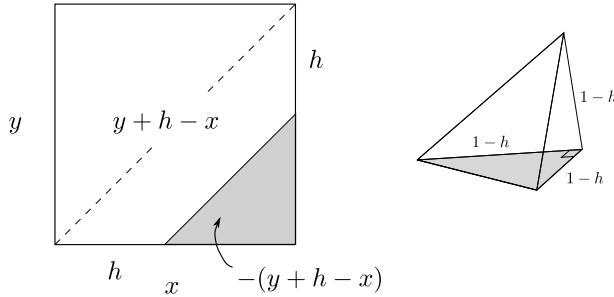


Fig. A.1. L'intégrale de $|y + h - x|$ vaut h plus deux fois le volume de la pyramide à droite.

On obtient alors : $\bar{\gamma}(V, V_h) = h + \frac{(1-h)^3}{3}$. Au passage, on en déduit que $\bar{\gamma}(V, V) = \frac{1}{3}$ et donc :

$$\bar{\gamma}_V(h) = h + \frac{(1-h)^3}{3} - \frac{1}{3} = h^2 - \frac{h^3}{3} \quad \text{si } h \leq 1$$

On trouve bien un comportement parabolique à l'origine. La régularisation du variogramme correspond à un lissage des réalisations, ce qui semble intuitif puisque le fait de moyenner le champ sur des blocs réduit d'autant la variabilité du processus.

Lorsque $h > 1$, la quantité $y+h-x$ est toujours positive et le calcul de l'intégrale donne immédiatement $\bar{\gamma}_V(h) = h - \frac{1}{3}$. On retrouve un comportement linéaire lorsque les moyennes sont calculé sur des blocs disjoints. En particulier, on vérifie que la fonction $\bar{\gamma}_V$ est bien continue en $h = 1$.

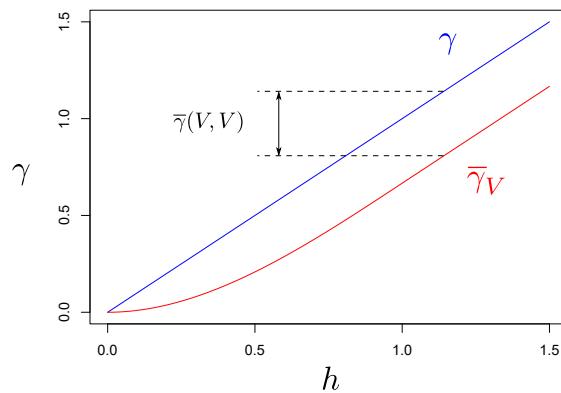


Fig. A.2. Variogramme d'un processus 1D (en bleu) et variogramme régularisé sur le bloc $V = [0, 1]$ (en rouge). Pour $h \geq 1$ le variogramme régularisé reprend son aspect linéaire.

Appendix B

Contre-exemples pratiques

B.1 Deux variables aléatoires décorrélées mais pas indépendantes

Prenons $X \sim \mathcal{U}([-1, 1])$, une variable aléatoire distribuée suivant une loi uniforme sur $[-1, 1]$, et posons $Y = X^2$. Le couple (X, Y) est un couple de variables aléatoires réelles. Étant donné que Y est défini en fonction de X , les deux variables ne sont pas indépendantes. Pour s'en convaincre, on peut regarder les probabilités conditionnelles :

$$\mathbb{P}[X = 1|Y = 1] = \mathbb{P}[X = 1|X^2 = 1] = \mathbb{P}[X = 1|X \in \{-1, 1\}] = \frac{1}{2}$$

Or, la loi de X étant continue, on a $\mathbb{P}[X = 1] = 0$, donc :

$$\mathbb{P}[X = 1|Y = 1] \neq \mathbb{P}[X = 1]$$

Les variables X et Y sont donc dépendantes. Calculons leur covariance :

$$\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \mathbb{E}[XY] = \mathbb{E}[X^3] = \frac{1}{2} \int_{-1}^{+1} x^3 dx = 0$$

où la seconde égalité résulte du fait que $\mathbb{E}[X] = 0$. On a donc :

$$\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}[X]\text{Var}[Y]}} = 0$$

Le couple de v.a. (X, Y) est donc dépendant et décorrélé à la fois. Cet exemple nous montre que l'indépendance est une notion plus forte que la décorrération. L'indépendance implique que $\rho_{XY} = 0$, mais la réciproque n'est pas vrai en général.

Le coefficient de corrélation et la covariance ne mesurent que la dépendance linéaire de deux variables. On peut rencontrer des schémas de dépendance plus complexes (comme ici en X^2), impliquant une décorrération des variables.

B.2 Une variable aléatoire qui ne possède pas d'espérance

On considère la loi de Cauchy centrée en 0 et de paramètre 1, définie par la densité :

$$p(x) = \frac{1}{\pi(x^2 + 1)}$$

On vérifie que π est bien une densité :

$$\int_{-\infty}^{+\infty} \pi(x) dx = \int_{-\infty}^{+\infty} \frac{dx}{\pi(x^2 + 1)} = \frac{1}{\pi} [\arctan(x)]_{-\infty}^{+\infty} = 1$$

Calculons l'espérance de la loi :

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} x\pi(x) dx = \int_{-\infty}^{+\infty} \frac{x}{\pi(x^2 + 1)} dx$$

Or $\frac{x}{\pi(x^2 + 1)} \sim \frac{1}{x}$ en l'infini, donc n'est pas intégrable. La queue de la distribution est trop lourde et donc l'espérance n'existe pas.

Cela peut sembler d'autant plus étonnant que la loi est symétrique et que sa médiane vaut donc 0. Pourtant, lorsqu'on tente d'appliquer la loi des grands nombres, on observe empiriquement que la moyenne d'une série de valeurs distribuées suivant cette loi ne converge pas vers 0. Informellement, on peut dire que la probabilité d'apparition de valeurs extrêmes n'est pas suffisamment faible pour que la moyenne des tirages se stabilise.

On peut facilement simuler une loi de Cauchy en R, par exemple en prenant le quotient de deux variables gaussiennes : `rnorm(1, 0, 1)/rnorm(1, 0, 1)`. On peut alors réaliser l'expérience ci-dessus avec l'instruction :

```
plot(cumsum(rnorm(N, 0, 1)/rnorm(N, 0, 1))/(1:N), type="l")
```

où N est un entier qu'on pourra faire croître progressivement, en commençant par une centaine. Ce résultat n'est pas contradictoire, puisque la loi des grands nombres ne s'applique que pour une suite de variables aléatoires dont l'espérance existe.

B.3 Une variable aléatoire qui ne possède pas de variance

De la même manière que pour le contre-exemple B.2, il faut trouver une loi dont le produit de la densité par x^2 n'est pas intégrable. On peut prendre la loi définie sur $[1; +\infty[$ par $p : x \mapsto 2/x^3$. On vérifie que p est bien une densité de probabilité :

$$\int_1^{+\infty} p(x) dx = \int_1^{+\infty} 2x^{-3} dx = \left[-x^{-2} \right]_1^{+\infty} = 1$$

De même, on vérifie également que l'espérance existe bien :

$$\mathbb{E}[X] = \int_1^{+\infty} xp(x)dx = \int_1^{+\infty} \frac{2}{x^2} dx = \left[-\frac{2}{x} \right]_1^{+\infty} = 2$$

On montre alors que la variance est infinie :

$$\text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = 2 \int_1^{+\infty} \frac{dx}{x} - 4 = 2 \lim_{x \rightarrow +\infty} \ln x - 4$$

La fonction $2/x$ n'est pas intégrable sur $[1; +\infty[$ donc la variance est infinie.

B.4 Un processus stationnaire mais non-ergodique

Il suffit de considérer le signal aléatoire constant sur chaque réalisation : $Z(x, \omega) = A(\omega)$, où A est une variable aléatoire réelle d'une loi quelconque, prenant au moins deux valeurs possibles différentes. La définition de Z ne dépend pas de x donc ce processus est clairement stationnaire (au sens strict). Il y a donc au moins une réalisation $\omega_0 \in \Omega$ telle que $A(\omega_0) \neq \mathbb{E}[A]$, et donc : la moyenne sur le domaine analytique de $Z(., \omega_0)$ vaut $A(\omega_0) \neq \mathbb{E}[A] = \mathbb{E}[Z(x)]$. Le processus Z n'est donc pas ergodique à l'ordre 1.

B.5 Un processus ergodique à l'ordre 1 mais pas à l'ordre 2

On peut considérer le processus $Z(x, \omega) = A(\omega)Y(x, \omega(x))$ où A est une variable aléatoire réelle d'une loi quelconque, prenant au moins deux valeurs possibles différentes, et $Y \sim \mathcal{N}(0, 1)$ est un bruit blanc. On a alors $\mathbb{E}[Z(x)] = \mathbb{E}[A]\mathbb{E}[Y(x)] = 0$ (puisque A et Y sont indépendantes). Par ailleurs, la moyenne analytique d'une réalisation quelconque $\omega_0 \in \Omega$ est :

$$\int_{\mathbb{R}} Z(x, \omega_0) dx = \int_{\mathbb{R}} A(\omega_0)Y(x, \omega(x)) dx = A(\omega_0) \int_{\mathbb{R}} Y(x, \omega(x)) dx = 0$$

Le processus est donc ergodique à l'ordre 1. En revanche, on a $\mathbb{E}[Z^2] = \text{Var}[Z(x, \omega)] = \text{Var}[A]\text{Var}[Y(x, \omega(x))] = \text{Var}[A]$. Or, le moment d'ordre 2 calculé sur une réalisation ω_0 est :

$$\int_{\mathbb{R}} Z(x, \omega_0)^2 dx = A(\omega_0)^2 \int_{\mathbb{R}} Y(x, \omega(x))^2 dx = A(\omega_0)^2$$

et en général, $A(\omega_0)^2 \neq \text{Var}[A]$. Le processus n'est donc pas ergodique à l'ordre 2.

B.6 Un processus strictement intrinsèque

Un processus stochastique X est dit *strictement intrinsèque* lorsqu'il vérifie l'hypothèse de stationnarité intrinsèque (def. 26) sans être stationnaire au second ordre (def. 20).

On peut reprendre l'exemple de la marche aléatoire de l'activité II, dont nous avions montré que la fonction de covariance valait :

$$\text{Cov}(X(t), X(t + \tau)) \propto \min(t, t + \tau) = t$$

en supposant l'intervalle $\tau \geq 0$, et où le symbole \propto signifie que la covariance est exprimée à un facteur multiplicatif près dépendant de la variance des incrément individuels. La covariance dépend du temps absolu t et pas uniquement de l'intervalle de temps τ séparant les deux échantillons. Le processus n'est donc pas stationnaire au second ordre¹.

Etudions alors le processus formé par les accroissements de X :

$$Y(t) = X(t) - X(t - 1)$$

On a : $\mathbb{E}[Y(t)] = \mathbb{E}[X(t)] - \mathbb{E}[X(t - 1)] = 0$. Par ailleurs :

$$\begin{aligned} \text{Cov}(Y(t), Y(t + \tau)) &= \text{Cov}(X(t) - X(t - 1), X(t + \tau) - X(t + \tau - 1)) = \text{Cov}(X(t), X(t + \tau)) - \\ &\text{Cov}(X(t), X(t + \tau - 1)) - \text{Cov}(X(t - 1), X(t + \tau)) + \text{Cov}(X(t - 1), X(t + \tau - 1)) \propto t - t - (t - 1) + (t - 1) = 0 \end{aligned}$$

Les incrément sont donc décorrélés et le processus Y est stationnaire au second ordre. X est donc un processus strictement intrinsèque. L'implication pratique en géostatistique est qu'on ne peut pas calculer le covariogramme de X puisqu'il ne dépend pas que de l'écart τ entre les sites. En revanche, le variogramme est bien défini et on a :

$$\gamma_X(\tau) = \frac{1}{2} \text{Var}[X(t + \tau) - X(t)] = \frac{1}{2} (\text{Var}[X(t + \tau)] + \text{Var}[X(t)] - 2\text{Cov}(X(t + \tau), X(t))) = \frac{1}{2} (t + \tau + t - 2t)$$

Et donc, le variogramme ne dépend bien que de l'écart τ :

$$\gamma_X(\tau) = \frac{\tau}{2}$$

C'est un variogramme linéaire sans palier (hypothèse intrinsèque donc variogramme non-borné). Son comportement à l'origine est linéaire, donc d'après l'annexe A, le processus X est continu mais non-dérivable en moyenne quadratique, ce qui explique l'aspect erratique d'une marche aléatoire.

¹Nous avions vu cependant dans l'activité II qu'il était tout de même stationnaire au premier ordre.

B.7 Un processus continu en m.q. dont toutes les réalisations sont discontinues

Le contre-exemple suivant, tiré de [Picinbono \(1998\)](#), est intéressant pour bien comprendre que la continuité en m.q. n'est qu'une notion probabiliste, elle ne s'applique donc pas systématiquement en tout point du domaine et pour toutes les réalisations.

On considère une variable aléatoire réelle $S \sim \mathcal{U}([0, 1])$, distribué suivant une loi uniforme sur l'intervalle $[0, 1]$, ainsi qu'un processus $X : [0, 1] \times \Omega \rightarrow \mathbb{R}$ défini par :

$$X_t = \begin{cases} 0 & \text{si } t \leq S \\ 1 & \text{sinon.} \end{cases} \quad (\text{B.1})$$

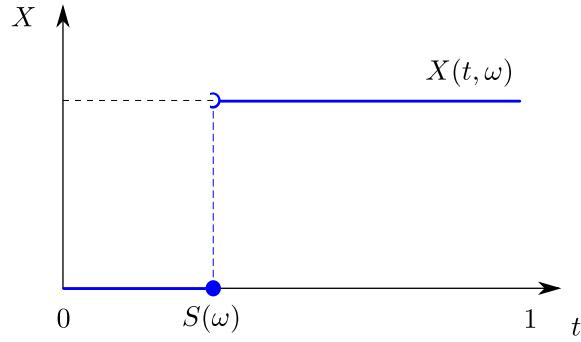


Fig. B.1. Processus stochastique X échelon unité au temps S aléatoire.

Très clairement, toute réalisation x de ce processus est discontinue (la position S du lieu où survient cette discontinuité étant toutefois aléatoire). Pourtant on peut montrer que le processus X est continu en m.q. : considérons la grandeur aléatoire $X(t) - X(t + \Delta t)$. Elle ne peut prendre que deux valeurs possibles, à savoir 1 si $s \in [t; t + \Delta t[$ et 0 sinon. Cette quantité est donc distribuée suivant une loi de Bernoulli dont la probabilité de succès est égale à la largeur de l'intervalle considéré : $\mathbb{P}[X(t) - X(t + \Delta t) = 1] = \Delta t$, d'où l'espérance quadratique :

$$\mathbb{E}[(X(t) - X(t + \Delta t))^2] = \mathbb{P}[X(t) - X(t + \Delta t) = 1] = \Delta t \xrightarrow{\Delta t \rightarrow 0} 0$$

Le processus est donc continu en m.q., malgré la discontinuité systématique de ses réalisations.

Index

- Accroissements, 48, 98
Analyse de sensibilité, 150
Anisotropie, 77
autocorrélation, 19
Autorisation (contrainte), 92
- Biais, 23, 54, 132, 137
Bilinéarité, 19
Bochner (théorème), 78
Bruit blanc, 68, 73
Bruit brownien, 66
Bruit de numérisation, 18, 21
- Cauchy-Schwarz (inégalité), 19
Champ aléatoire, 30
Champ scalaire, 8
Champ vectoriel, 7
Cholesky (factorisation), 148
Co-krigeage, 128
Coefficient de sureté, 142
Compromis biais-variance, 133
Compromis biais-variances, 23
Continuité en moyenne quadratique, 62
Convolution, 81, 111
Corrélation, 19
Covariance, 19
Covariance régularisée, 120
Covariogramme, 37
- Défini-négativité conditionnelle, 78
Dérivabilité en moyenne quadratique, 62, 111
Danie Gerhardus Krige, 6
Densité de probabilité, 14
Densité spectrale de puissance, 79
- Ecart-type, 18
Echelon unité, 78
Effet de lissage, 113
Ensemble fondamental, 12
Ergodicité à l'ordre 1, 39
Ergodicité à l'ordre 2, 40
Erreur d'estimation, 91
Erreur normalisée, 134
Erreur systématique, 24
Espérance, 15
Espace probabilisé, 10
Espace topo-probabilitiste, 30
- Eventualités élémentaires, 10
Exactitude, 109
Expérience aléatoire, 10
- Fonction aléatoire, 30
Fonction de répartition, 13
fonction génératrice des moments, 24
Fonction objectif, 94
Fourier, Joseph, 68
- Géostatistique multivariée, 128
Georges Matheron, 7
GPS, 7, 13, 15, 29, 61, 65, 69
Gradient, 94
- Heaviside (fonction, 78
- Interpolateur exact, 109
Interpolation, 29, 83
Intervalle de confiance, 101, 142
Intrinsèque (hypothèse), 48
Isotropie, 37
- Kolmogorov (axiomes), 10
Krigeage, 83
Krigeage aléatoire, 127
Krigeage avec dérive externe, 126
Krigeage d'indicatrice, 129
Krigeage de la moyenne, 117
Krigeage dual, 114
Krigeage intrinsèque, 97
Krigeage mou, 130
Krigeage ordinaire, 95
Krigeage par blocs, 119
Krigeage séquentiel, 147
Krigeage simple, 116
Krigeage universel, 123
- Lagragien, 94
Linéarité (contrainte), 91
Loi binomiale, 16
Loi jointe fini-dimensionnelle, 35
Loi log-normale, 25
Loi normale, 17, 31, 66
LOOCV, 132
- Méthodes déterministes, 29, 138
Marche aléatoire, 31, 36

Matrice de covariances, 20
MNT, 6, 7, 29, 56, 60, 69, 103
Modèle à effet de pépite, 69
Modèle à effet de trou, 67
Modèle cubique, 67
Modèle de variogramme, 62
Modèle exponentiel, 65
Modèle gaussien, 66
Modèle gigogne, 69
Modèle linéaire, 63
Modèle linéaire avec palier, 64
Modèle pépitique pur, 68
Modèle puissance, 64
Modèle sphérique, 66
Moments statistiques, 23
Multiplicateur de Lagrange, 94

Noyau, 85
noyau, 137

Optimalité (contrainte), 93
Optimisation bayésienne, 140

Pépite, 60, 145
Pépite pure, 73
Palier, 59
Parabolique, 61
Plus proche voisin, 138
Portée, 60
Probabilité, 10
Problème modèle, 7, 29
Processus markovien, 32
Processus stochastique, 30
Propagation d'incertitudes, 20

Régression linéaire, 74
Régularisation, 121
Ratio signal sur bruit, 73
RMSE, 75, 137

Semi-variogramme, 48
Signal aléatoire, 30
Simulation conditionnelle, 157
Simulation non-conditionnelle, 147
Simulations, 146
Splines, 29
Stationnarité, 35
Stationnarité à l'ordre 1, 36
Stationnarité à l'ordre 2, 36
Stationnarité au sens faible, 36
Stationnarité au sens fort, 36
Stationnarité au sens large, 36
Stationnarité au sens strict, 36

Transformation de Fourier, 78

Universalité (contrainte), 92
Validation croisée, 131
Variable aléatoire, 12
Variable auxiliaire, 128
Variable régionalisée, 7, 29, 30
Variables décorrélées, 20
Variables indépendantes, 20
Variance, 18
Variance expliquée, 133
Variogramme, 48
Variogramme expérimental, 50
Variogramme pépitique, 61
Variogramme régularisé, 120
Vecteur aléatoire, 30

Bibliography

- Allard, D. (2012). Statistiques spatiales: introduction à la géostatistique. *Lecture notes, University of Montpellier*.
- Antoni, J. (2013). *Cours de Signaux Aléatoires : Master MGEA, INSA-Lyon*.
- Arnaud, M. et Emery, X. (2000). *Estimation et interpolation spatiale: méthodes déterministes et méthodes géostatistiques*. Hermès.
- Baillargeon, S. (2005). Le krigage: revue de la théorie et application à l'interpolation spatiale de données de précipitations.
- Barret, M. (2009). *Traitemet statistique du signal: Estimation, filtrage de Wiener, méthodes récursives, détection*.
- Barry, R. P., Jay, M., et Hoef, V. (1996). Blackbox kriging: spatial prediction without specifying variogram models. *Journal of Agricultural, Biological, and Environmental Statistics*, pages 297–322.
- Bengio, Y. et Grandvalet, Y. (2004). No unbiased estimator of the variance of k-fold cross-validation. *Journal of machine learning research*, 5(Sep):1089–1105.
- Biljecki, F., Heuvelink, G. B., Ledoux, H., et Stoter, J. (2015). Propagation of positional error in 3d gis: estimation of the solar irradiation of building roofs. *International Journal of Geographical Information Science*, 29(12):2269–2294.
- Bonin, O. (1998). Attribute uncertainty propagation in vector geographic information systems: sensitivity analysis. In *Proceedings. Tenth International Conference on Scientific and Statistical Database Management (Cat. No. 98TB100243)*, pages 254–259. IEEE.
- Bosser, P. (2011). Interpolation spatiale.
- Bourgine, B. (1995). Logiciels géostatistiques de gdm, manuel de l'utilisateur. *août*, 37842.
- Burman, P. (1989). A comparative study of ordinary cross-validation, v-fold cross-validation and the repeated learning-testing methods. *Biometrika*, 76(3):503–514.
- Conte, E. (2018). Changement climatique et toponymie : écologie historique du houx et du buis à travers leurs traces toponymiques. *Mémoire de fin d'étude Master 2 Carthagéo*.
- Cottet, F. (2015). *Traitemet des signaux et acquisition de données-4e éd*. Dunod.
- Deraisme, J. et Bobbia, M. (2003). Geostatistics in the study of air pollution-related risks. *Environnement, Risques & Santé*, 2(3):168–175.
- Droesbeke, J.-J., Lejeune, M., et Saporta, G. (2006). *Analyse statistique des données spatiales*. Editions TECHNIP.
- Dumenieu, B. (2015). *Un système d'information géographique pour le suivi d'objets historiques urbains à travers l'espace et le temps*. PhD thesis, Paris, EHESS.

- Dumont, T., Léonard, C., Mary, X., et Mohamed, H. (2019). Cours d'optimisation. *Notes de cours de L1 - économie-Gestion - analyse 2*.
- Efron, B. (1983). Estimating the error rate of a prediction rule: improvement on cross-validation. *Journal of the American statistical association*, 78(382):316–331.
- Efron, B. et Tibshirani, R. J. (1994). *An introduction to the bootstrap*. CRC press.
- Emery, X. (2001). *Cours de Géostatistique linéaire*. Document de cours à l'Ecole des Mines.
- Grajner-Brzezinska, D., Toth, C., et Yi, Y. (2005). On improving navigation accuracy of gps/ins systems. *Photogrammetric engineering & remote sensing*, 71(4):377–389.
- Gringarten, E. et Deutsch, C. V. (2001). Teacher's aide variogram interpretation and modeling. *Mathematical Geology*, 33(4):507–534.
- Hauchecorne, B. (2007). *Les contre-exemples en mathématiques*. Ellipses Paris.
- Journel, A. G. et Huijbregts, C. J. (1978). *Mining geostatistics*, volume 600. Academic press London.
- Krige, D. G. (1951). A statistical approach to some basic mine valuation problems on the witwatersrand. *Journal of the Southern African Institute of Mining and Metallurgy*, 52(6):119–139.
- Lafarge, F. (2007). *Modèles stochastiques pour la reconstruction tridimensionnelle d'environnements urbains*. PhD thesis.
- Lecoutre, J.-P. (2002). *Statistique et probabilités*. Dunod.
- Li, C., Zhao, S., Wang, Q., et Shi, W. (2018). Uncertainty modeling and analysis of surface area calculation based on a regular grid digital elevation model (dem). *International Journal of Geographical Information Science*, 32(9):1837–1859.
- Marcotte, D. (2003). Cours de géostatistiques, École polytechnique - glq340.
- Matheron, G. (1962). *Traité de géostatistique appliquée. 1* (1962), volume 1. Editions Technip.
- Mehlhorn, K., Newell, B. R., Todd, P. M., Lee, M. D., Morgan, K., Braithwaite, V. A., Hausmann, D., Fiedler, K., et Gonzalez, C. (2015). Unpacking the exploration-exploitation tradeoff: A synthesis of human and animal literatures. *Decision*, 2(3):191.
- Meneroux, Y. (2019). *Méthodes d'apprentissage statistique pour la détection de la signalisation routière à partir de véhicules traceurs*. PhD thesis, Paris Est.
- Méneroux, Y., Le Guilcher, A., Orfila, O., Lusetti, B., Saint Pierre, G., et Mustière, S. (2017). Using surrogate road network for map-matching a sensitivity analysis of positional accuracy. *Geocomputation 2017*.
- Mitas, L. et Mitasova, H. (1999). Spatial interpolation. *Geographical information systems: principles, techniques, management and applications*, 1(2).
- Nadaraya, E. (1965). On non-parametric estimates of density functions and regression curves. *Theory of Probability & Its Applications*, 10(1):186–190.
- Parzen, E. (1962). On estimation of a probability density function and mode. *The annals of mathematical statistics*, 33(3):1065–1076.
- Pebesma, E. (2020). The meuse data set: a brief tutorial for the gstat r package.
- Pebesma, E. J. (2001). Gstat user's manual. *Dept. of Physical Geography, Utrecht University, Utrecht, The Netherlands*.
- Pebesma, E. J. et Bivand, R. S. (2005). Classes and methods for spatial data in R. *R News*, 5(2):9–13.

- Peraudin, J.-J. (2003). Surveillance de la qualité de l'air par cartographie : l'apport de la géostatistique.
- Picinbono, B. (1998). *Signaux aléatoires: avec problèmes résolus. Fonctions aléatoires et modèles.* Dunod.
- Pyrcz, M. et Deutsch, C. (2003). The whole story on the hole effect. *Geostatistical Association of Australasia, Newsletter*, 18:3–5.
- Richards, C. (2016). An r function to shift vectors by a specified lag, <https://clarkrichards.org/>.
- Rivoirard, J. (2003). *Cours de Géostatistique multivariable.* Centre de Géostatistique de Fontainebleau.
- Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, pages 832–837.
- Saltelli, A., Chan, K., et Scott, E. M., editors (2000). *Sensitivity analysis.* Wiley series in probability and statistics. J. Wiley & sons, New York, Chichester, Weinheim.
- Sillard, P. (2001). *Estimation par moindres carrés.* Hermès Science Publications.
- Snoek, J., Larochelle, H., et Adams, R. P. (2012). Practical bayesian optimization of machine learning algorithms. In *Advances in neural information processing systems*, pages 2951–2959.
- Tian, L., Cai, T., Goetghebeur, E., et Wei, L. (2007). Model evaluation based on the sampling distribution of estimated absolute prediction error. *Biometrika*, 94(2):297–311.
- Vauglin, F. (1997). *Modèles statistiques des imprécisions géométriques des objets géographiques linéaires.* PhD thesis, Université de Marne-la-Vallée.
- Wackernagel, H. (2013a). Basics in geostatistics 3: Geostatistical monte-carlo methods - conditional simulationconditional simulation.
- Wackernagel, H. (2013b). *Multivariate geostatistics: an introduction with applications.* Springer Science & Business Media.
- Yang, C.-S., Kao, S.-P., Lee, F.-B., et Hung, P.-S. (2004). Twelve different interpolation methods: A case study of surfer 8.0. In *Proceedings of the XXth ISPRS Congress*, volume 35, pages 778–785.
- Zhang, S., Lo, S., Chen, Y.-H., Walter, T., et Enge, P. (2018). Gnss multipath detection in urban environment using 3d building model. In *2018 IEEE/ION Position, Location and Navigation Symposium (PLANS)*, pages 1053–1058. IEEE.
- Zhang, Y. et Yang, Y. (2015). Cross-validation for selecting a model selection procedure. *Journal of Econometrics*, 187(1):95–112.
- Zoraster, S. (1996). Imposing geologic interpretations on computer-generated contours using distance transformations. *Mathematical geology*, 28(8):969–985.