



PROJET DE RECHERCHE : KRIGEAGE

Compréhension et mise en place d'une méthode
d'interpolation appelée "krigeage"

Septembre 2023 - Décembre 2023

Samuel GAUDIN, Mathias GRAU, Alexandre VER HULST
Tuteur : Thierry KLEIN



REMERCIEMENTS

Nous tenons à exprimer notre plus sincère gratitude envers le **Professeur Thierry Klein** pour son encadrement exceptionnel et son dévouement tout au long de notre projet d'EA d'initiation à la recherche (dans le cadre de la thématique MAP511) sur les *Processus Aléatoires*.

Ses conseils avisés et son expertise ont été d'une aide précieuse et ont joué un rôle déterminant dans l'orientation et la réussite de notre travail. Sa disponibilité et sa réactivité face à nos interrogations et besoins ont grandement contribué à l'avancement de notre projet. Nous sommes profondément reconnaissants pour son soutien inestimable et pour l'impact positif qu'il a eu sur notre parcours académique et notre compréhension du sujet.

RÉSUMÉ

Les processus gaussiens constituent un outil fondamental dans le domaine de l'apprentissage statistique et dispose de nombreuses applications notamment en géostatistique.

On peut définir un processus gaussien comme un processus stochastique (une collection de variables aléatoires avec un index temporel ou spatial) de telle sorte que chaque collection finie de ces variables aléatoires suit une loi normale multidimensionnelle. Cette définition souligne la flexibilité et la puissance des processus gaussiens en tant qu'outil de modélisation.

Le krigeage, du nom du géostatisticien D.G. Krige, est une méthode de prédiction spatiale qui utilise les processus gaussiens pour interpoler une fonction inconnue. Il se divise principalement en deux types : le krigeage simple, qui suppose une moyenne constante et inconnue, et le krigeage universel, qui modélise la moyenne comme une combinaison linéaire de fonctions connues. Ces méthodes permettent de réaliser des prédictions et des simulations conditionnelles, en fournissant des estimations optimales et non biaisées de la variable d'intérêt à partir des données observées.

La simulation conditionnelle de processus gaussiens est une extension naturelle du krigeage. Elle permet de générer des réalisations possibles du processus qui sont cohérentes avec les observations existantes. Cette technique est essentielle pour évaluer l'incertitude spatiale et pour la reconstruction de fonctions initiales à partir de données partielles ou bruitées.

Dans cette perspective, la relation entre la fonction de covariance d'un processus gaussien et ses trajectoires est cruciale, car la fonction de covariance détermine les propriétés de régularité et de continuité des trajectoires du processus. En effet, la différentiabilité du noyau est intimement liée aux propriétés de régularité des trajectoires du processus. Par ailleurs, l'estimation des paramètres du noyau est cruciale pour optimiser la performance du krigeage. Des méthodes telles que le Maximum de Vraisemblance, ou la Validation Croisée sont couramment utilisées pour estimer ces paramètres, influençant directement la qualité des prédictions et simulations effectuées par le krigeage. Ces sujets en particulier feront l'objet de nos recherches dans ce papier.

ABSTRACT

Gaussian processes are a fundamental tool in the field of statistical learning and have many applications, particularly in geostatistics.

A Gaussian process can be defined as a stochastic process (a collection of random variables with a temporal or spatial index) such that each finite collection of these random variables follows a multidimensional normal distribution. This definition highlights the flexibility and power of Gaussian processes as a modelling tool.

Kriging, named after the geostatistician D.G. Krige, is a spatial prediction method that uses Gaussian processes to interpolate an unknown function. There are two main types: simple kriging, which assumes a constant, unknown mean, and universal kriging, which models the

mean as a linear combination of known functions. These methods enable predictions and conditional simulations to be carried out, providing optimal, unbiased estimates of the variable of interest from the observed data.

Conditional simulation of Gaussian processes is a natural extension of kriging. It is used to generate possible realisations of the process that are consistent with existing observations. This technique is essential for assessing spatial uncertainty and for reconstructing initial functions from partial or noisy data.

From this perspective, the relationship between the covariance function of a Gaussian process and its trajectories is crucial, as the covariance function determines the regularity and continuity properties of the process trajectories. Indeed, the differentiability of the kernel is closely linked to the regularity properties of the process trajectories. Furthermore, estimating the kernel parameters is crucial for optimising kriging performance. Methods such as Maximum Likelihood or Cross Validation are commonly used to estimate these parameters, directly influencing the quality of the predictions and simulations carried out by kriging. These topics in particular will be the focus of our research in this paper.

Keywords Kriging, Gaussian Processes, Kernel Regularity, Maximum Likelihood Estimation, Geostatistical Interpolation, Gaussian Processes.

CONTENTS

1	Processus Gaussiens et Krigeage	7
1.1	Processus Gaussiens, définition et premières propriétés	7
1.1.1	Processus aléatoire	7
1.1.2	Processus Gaussien	8
1.1.3	Noyau et fonction de covariance	8
1.2	Krigeage	9
1.2.1	Introduction et objectif	9
1.2.2	Propriétés et liens avec les processus gaussiens	10
1.2.3	Application au cas multidimensionnel	10
2	Régularité du kernel, régularité du processus gaussien	12
2.1	Continuité et différentiabilité au sens des moindres carrés	12
2.2	Liens entre le kernel et la régularité du processus gaussiens	13
2.3	Exemple important	14
3	Estimation paramétrique de la fonction de covariance pour des modèles de krigeage	15
3.1	Introduction	15
3.2	Notions générales	16
3.3	Estimation des paramètres d'une fonction de covariance	16
3.4	Estimateur du maximum de vraisemblance pour les paramètres	17
3.5	Application au Kernel Matérn	20

INTRODUCTION

MISE EN CONTEXTE

Nous allons dans ce paragraphe introduire un contexte d'application du krigeage, pour voir dans quel contexte celui-ci s'utilise.

Contexte : Selon l'organisation mondiale de la santé, près de 6,7 millions de décès prématurés par an sont causés par la pollution de l'air. Il est donc primordial de mener des études concrètes à l'intérieur des villes afin d'avoir une idée des taux de pollution à chaque endroit de la ville. Seulement, dans des villes de grande superficie, il est impossible de mesurer à chaque endroit la pollution.

Objectif : Connaître la pollution de l'air à chaque endroit de la ville.

Application du krigeage : Plusieurs étapes sont nécessaires. D'abord, il s'agit de recueillir un ensemble de données suffisantes auprès des stations de surveillance dans la ville. Ensuite, il s'agit d'analyser comment la pollution évolue avec la distance pour choisir la bonne corrélation distance-pollution. Enfin, nous interpolons à l'aide du krigeage puis intégrons les incertitudes de manière à connaître l'erreur type en chacun des points de la ville.

De nombreuses autres applications sont possibles, puisque tout problème d'interpolation peut se ramener à un problème de krigeage. Ainsi, voici quelques exemples succincts :

- *Dans les mines :* afin d'estimer la concentration de minéraux dans un gisement
- *Météo :* afin de faire des mesures à partir de stations météo dispersées au sein de la France
- *Agriculture :* pour mesurer plus précisément la fertilité des sols dans une étendue de prairies.

POURQUOI LE KRIGEAGE ?

La krigeage possède plusieurs qualités que l'on ne retrouve pas dans d'autres méthodes d'interpolation comme une interpolation de Lagrange par exemple :

- Prise en compte de la corrélation spatiale : à travers un noyau bien choisi, nous pouvons choisir comment les distances affectent le lien entre chaque point.
- De plus, un avantage indéniable du krigeage se situe dans le fait qu'il fournit des estimations non biaisées avec une variance d'erreur minimale. Les interpolations faites par krigeage sont ainsi les plus probables statistiquement au vu de l'ensemble de données.
- Le krigeage jouit également de la capacité à approximer des ensembles de données en plusieurs dimensions, comme l'exemple de la mesure de pollution dans les villes le montre.

1

PROCESSUS GAUSSIENS ET KRIGEAGE

1.1 PROCESSUS GAUSSIENS, DÉFINITION ET PREMIÈRES PROPRIÉTÉS

1.1.1 • PROCESSUS ALÉATOIRE

Nous commencerons ici par introduire les notions élémentaires concernant les processus aléatoires ainsi que les processus gaussiens. Nous nous placerons ici dans l'espace $\mathcal{D} \subset \mathbb{R}^d$.

Definition 1 Un processus à valeurs réelles sur \mathcal{D} est une application Y définie sur \mathcal{D} et telle que $Y(x)$ est une variable aléatoire sur l'espace de probabilités usuel $(\Omega, \mathcal{F}, \mathbb{P})$.

Pour chaque $\omega \in \Omega$, $Y(\omega, x)$ est une "réalisation" (ou "trajectoire") du processus Y .

Exemple 1 : On considère ici l'exemple suivant : U suit une loi uniforme sur $[-\pi, \pi]$ et on prend $Y(\omega, x) = \cos(U(\omega) + x)$. On remarque que à ω fixé, la période est fixée. ω détermine en revanche l'ordonnée à l'origine. On s'attend donc à obtenir des courbes similaires, mais translatées, comme la figure 1 le montre.

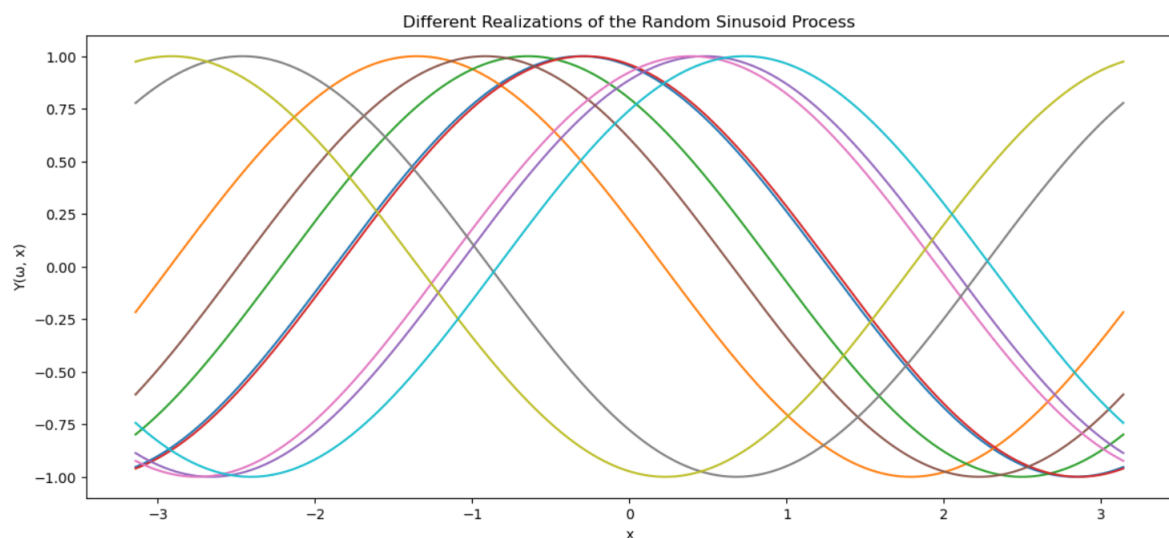


Figure 1: 10 réalisations différentes du processus précédent

Definition 2 Pour tous points $x^1, \dots, x^n \in \mathcal{D}$, la probabilité multidimensionnelle du vecteur $Y(x^1), \dots, Y(x^n)$ est la distribution en dimension finie du processus aléatoire Y

1.1.2 • PROCESSUS GAUSSIEN

Definition 3 On considère $n > 1$ ainsi qu'un vecteur aléatoire $\mathbf{y} = (y_1, \dots, y_n)$. On dit que \mathbf{y} est un vecteur aléatoire gaussien si l'une des deux conditions suivantes (équivalentes) est vérifiée :

1. Pour tous $a_1, \dots, a_n \in \mathbb{R}$, la variable aléatoire $\sum_{i=1}^n a_i y_i$ suit une distribution gaussienne
2. Il existe un vecteur \mathbf{m} de taille n et une matrice positive K de taille $n \times n$ tels que le vecteur \mathbf{y} possède une fonction caractéristique qui s'écrit $\mathbf{u} \rightarrow \exp(i\mathbf{u}^T \mathbf{m} - \frac{1}{2} \mathbf{u}^T K \mathbf{u})$

On note alors $\mathbf{y} \sim \mathcal{N}(\mathbf{m}, K)$

Definition 4 La densité de probabilité d'un vecteur gaussien \mathbf{X} de dimension n avec une moyenne vectorielle $\boldsymbol{\mu}$ et une matrice de covariance Σ est donnée par la formule suivante :

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})\right) \quad (1)$$

où $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ est un vecteur de l'espace euclidien \mathbb{R}^n .

Cette densité est caractéristique de la distribution gaussienne multidimensionnelle, définissant complètement les propriétés de corrélation et de variation des composants de \mathbf{X} .

1.1.3 • NOYAU ET FONCTION DE COVARIANCE

Dans le cadre des processus gaussiens, le choix du noyau (ou *kernel*) est crucial car il détermine la fonction de covariance du processus, influençant ainsi la nature de la corrélation entre les points de données. Différents noyaux peuvent être utilisés, chacun ayant des propriétés et des applications spécifiques :

1. **Noyau Linéaire** : Le noyau linéaire, défini par $K(x, x') = x^T x'$, est simple mais puissant, modélisant des relations linéaires entre les variables. Il est souvent utilisé dans les cas où une relation linéaire est supposée entre les points de données.
2. **Noyau Polynomial** : Défini par $K(x, x') = (x^T x' + c)^d$, où c et d sont des constantes, ce noyau introduit une non-linéarité et permet de modéliser des relations plus complexes. Il est particulièrement utile dans les situations où les relations entre les données sont de nature polynomiale.
3. **Noyau Gaussien (RBF - Radial Basis Function)** : Le noyau gaussien, $K(x, x') = \exp\left(-\frac{\|x - x'\|^2}{2\sigma^2}\right)$, où σ est un paramètre d'échelle, est l'un des plus populaires. Il modélise des relations lisses et est largement utilisé en raison de sa flexibilité et de sa capacité à gérer des relations non linéaires entre les points de données.
4. **Noyau Exponentiel** : Proche du noyau gaussien, le noyau exponentiel est défini par $K(x, x') = \exp\left(-\frac{\|x - x'\|}{\sigma}\right)$, offrant une alternative moins lisse que le noyau gaussien, et est souvent utilisé pour des processus avec des variations plus rapides.

5. **Noyau Matérn** : Le noyau Matérn, donné par $K(x, x') = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}\|x-x'\|}{\rho} \right)^\nu K_\nu \left(\frac{\sqrt{2\nu}\|x-x'\|}{\rho} \right)$, où ν et ρ sont des paramètres, permet de modéliser une gamme de régularités. K_ν est une fonction de Bessel modifiée de seconde espèce. Ce noyau est particulièrement utile pour capturer des corrélations avec différents degrés de régularité.

Definition 5 La matrice de covariance dans le cadre d'un processus gaussien est construite à partir du noyau choisi, qui détermine la relation de dépendance entre les différentes observations. Considérons un ensemble de points $\mathbf{X} = \{x_1, x_2, \dots, x_n\}$, la matrice de covariance \mathbf{K} est formée en évaluant le noyau pour chaque paire de points dans \mathbf{X} . Chaque élément K_{ij} de la matrice \mathbf{K} est calculé en appliquant le noyau sur la paire de points (x_i, x_j) , c'est-à-dire $K_{ij} = k(x_i, x_j)$. Ainsi, la matrice de covariance \mathbf{K} est définie comme suit :

$$\mathbf{K} = \begin{bmatrix} k(x_1, x_1) & k(x_1, x_2) & \cdots & k(x_1, x_n) \\ k(x_2, x_1) & k(x_2, x_2) & \cdots & k(x_2, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ k(x_n, x_1) & k(x_n, x_2) & \cdots & k(x_n, x_n) \end{bmatrix}$$

1.2 KRIGEAGE

1.2.1 • INTRODUCTION ET OBJECTIF

Le krigeage est une méthode d'interpolation qui prédit les valeurs d'une variable spatiale à des emplacements non échantillonnés en se basant sur une combinaison linéaire pondérée des valeurs observées, tout en tenant compte de la distance spatiale et des corrélations entre les points de données.

L'essence du krigeage repose sur l'hypothèse que la distance ou la similitude spatiale entre les points de données influence leurs valeurs. On utilise pour cela un kernel qui va nous permettre de déduire fonction de covariance, dérivée des processus gaussiens, pour quantifier cette interdépendance spatiale.

Le krigeage fournit une estimation optimale non biaisée, et permet de quantifier l'incertitude de cette estimation. Ce dernier aspect est crucial dans les applications pratiques, car il permet d'évaluer la fiabilité des prédictions.

Il existe plusieurs variantes de krigeage, chacune adaptée à différents types de données et structures de covariance. Les plus courantes et celle que nous étudieront par la suite sont :

- **Krigeage Simple** : Utilisé lorsque la moyenne de la fonction est considérée comme constante et inconnue dans la zone d'intérêt.
- **Krigeage Ordinaire** : Similaire au krigeage simple, mais il suppose une moyenne inconnue mais constante localement.
- **Krigeage Universel** : Applicable lorsque la moyenne de la fonction n'est ni constante ni connue, nécessitant l'utilisation de tendances déterministes dans le modèle.

1.2.2 • PROPRIÉTÉS ET LIENS AVEC LES PROCESSUS GAUSSIENS

Proposition 1 Nous considérons ici un vecteur de taille $n = n_1 + n_2$ de la forme

$$\begin{bmatrix} y^{(1)} \\ y^{(2)} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m^{(1)} \\ m^{(2)} \end{bmatrix}, \begin{bmatrix} K_{1,1} & K_{1,2} \\ K_{2,1} & K_{2,2} \end{bmatrix}\right)$$

. Alors, si $K_{1,1}$ est inversible, nous avons, conditionnellement à $y^{(1)} = v^1$, y_2 suit une loi de la forme :

$$\mathcal{N}(m^{(2)} K_{2,1} K_{1,1}^{-1} (v^{(1)} - m^{(1)}), K_{2,2} - K_{2,1} K_{1,1}^{-1} K_{1,2})$$

Remarque Notons que dans le cadre du krigeage, cette propriété s'avérera être très importante. En effet, nous remarquerons que $y^{(1)}$ représentera l'ensemble des données connues, tandis que $y^{(2)}$ représentera l'interpolation que l'on décide de réaliser. La matrice K est celle qu'on appellera 'kernel', et qui nous permettra de modéliser l'interaction entre les différents points.

1.2.3 • APPLICATION AU CAS MULTIDIMENSIONNEL

Il nous suffit donc de définir un kernel (Exponentiel, Gaussien, Matern, ...), de choisir une fonction moyenne (nulle dans le cas du krigeage simple), de calculer la matrice de covariance, obtenue pour les différents points (échantillons $x^{(1)}, \dots, x^{(n)}$ avec leur valeur $\mathbf{y} = (y_1, \dots, y_n)$ et les nouveaux points x_{new} pour lesquels nous souhaitons déterminer la valeur de avec un intervalle de confiance de 95%) et enfin de simuler un nouveau processus gaussien avec ces paramètres. On déduit de cela l'intervalle de confiance à 95% dans lequel la courbe initiale est censée se trouver. Cette méthode fonctionne pour tout type de fonctions notamment les fonctions de $\mathbb{R}^2 \rightarrow \mathbb{R}$

Nous prenons ici l'exemple d'une fonction :

$$f(x, y) = \begin{cases} \mathbb{R}^2 \rightarrow \mathbb{R} \\ x \mapsto \sin(0.4x) \cdot \sin(0.4y) \end{cases}$$

3D Plot with Kriging Predictions and 95% Confidence Interval

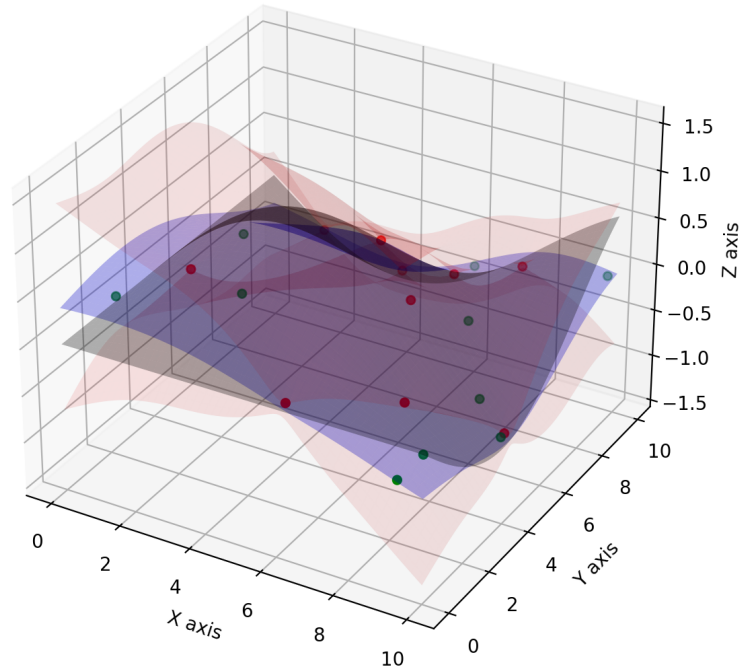


Figure 2: $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, $f(x, y) = \sin(0.4x) \cdot \sin(0.4y)$ avec 10 points (rouges) connus, en noir la courbe de f , en bleu l'espérance du processus gaussien, en rouge l'intervalle de confiance à 95%, en vert les points de test

Premièrement, il est notable que les points de test (indiqués en rouge) se situent bien dans l'intervalle de confiance prédit par le modèle de krigeage. Cela suggère une bonne adéquation du modèle de krigeage avec les données observées, témoignant de sa capacité à capturer la structure sous-jacente de la fonction f .

De plus, la fonction réelle f , dont on effectue le krigeage, est également bien modélisée, comme en témoigne la proximité entre les prédictions du krigeage et les valeurs réelles de la fonction. Cela démontre la précision et la fiabilité du krigeage en tant qu'outil d'interpolation.

Enfin, un point crucial à souligner est que, aux emplacements des 10 points d'échantillonnage, l'intervalle de confiance est pratiquement nul. Ceci confirme que le krigeage est bien une méthode d'interpolation : le modèle passe exactement par les points d'échantillonnage, garantissant une prédiction parfaite en ces points. Cette caractéristique est fondamentale pour comprendre pourquoi le krigeage est si largement utilisé pour l'interpolation spatiale, en particulier dans des contextes où la précision aux points échantillonnés est primordiale.

2

RÉGULARITÉ DU KERNEL, RÉGULARITÉ DU PROCESSUS GAUSSIEN

2.1 CONTINUITÉ ET DIFFÉRENTIABILITÉ AU SENS DES MOINDRES CARRÉS

Nous allons à présent aborder quelques définitions sur les continuités des processus, afin d'aboutir sur un théorème très puissant liant différentiabilité d'un processus et différentiabilité de sa fonction covariance.

Definition 6 Un processus à valeurs réelles sur \mathcal{D} est continue au sens des moindres carrés si pour tout $x^{(0)}$ dans \mathcal{D} , $Y(x)$ tend au sens des moindres carrés vers $Y(x^{(0)})$ quand x tend vers $x^{(0)}$, ie $\mathbb{E}((Y(x) - Y(x^{(0)}))^2) \xrightarrow{x \rightarrow x^{(0)}} 0$

Definition 7 Un processus gaussien Y est différentiable au sens des moindres carrés sur D si et seulement s'il existe d processus gaussiens $\frac{\partial}{\partial x_1} Y, \dots, \frac{\partial}{\partial x_d} Y$ tels que, pour tout $k \in \{1, \dots, d\}$, $x^{(0)} \in D$, avec $e^{(k)}$ le k -ième vecteur de base, la quantité

$$\frac{Y(x^{(0)} + he^{(k)}) - Y(x^{(0)})}{h}$$

converge vers $\frac{\partial}{\partial x_k} Y(x^{(0)})$ au sens des moindres carrés, lorsque le scalaire h tend vers zéro.

Definition 8 Un processus gaussien Y est différentiable k fois au sens des moindres carrés sur D si et seulement s'il est différentiable $k - 1$ fois au sens des moindres carrés, et pour tous $i_1, \dots, i_{k-1} \in \{1, \dots, d\}$, le processus gaussien

$$\frac{\partial}{\partial x_{i_{k-1}}} \dots \frac{\partial}{\partial x_{i_1}} Y$$

est différentiable au sens des moindres carrés, avec les processus dérivés au sens des moindres carrés

$$\frac{\partial}{\partial x_1} \frac{\partial}{\partial x_{i_{k-1}}} \dots \frac{\partial}{\partial x_{i_1}} Y, \dots, \frac{\partial}{\partial x_d} \frac{\partial}{\partial x_{i_{k-1}}} \dots \frac{\partial}{\partial x_{i_1}} Y.$$

2.2 LIENS ENTRE LE KERNEL ET LA RÉGULARITÉ DU PROCES- SUS GAUSSIENS

Proposition 2 *Un processus gaussien centré Y est continu en moyenne quadratique si et seulement si sa fonction de covariance est continue en chaque paire (x, x) , $x \in D$. De plus, si une fonction de covariance est continue en chaque paire (x, x) , $x \in D$, alors elle est continue sur $D \times D$. La preuve est admise ici*

Proposition 3 *Soit $D \subset \mathbb{R}^d$. Pour un processus gaussien centré Y , pour tout $k \in \mathbb{N}$, $i_1, \dots, i_k \in \{1, \dots, d\}$, si la fonction dérivée*

$$\frac{\partial^2}{\partial x_{i_1} \partial y_{i_1}} \cdots \frac{\partial^2}{\partial x_{i_k} \partial y_{i_k}} K$$

existe et est finie, alors le processus dérivé

$$\frac{\partial}{\partial x_{i_1}} \cdots \frac{\partial}{\partial x_{i_k}} Y$$

existe au sens des moindres carrés et est un processus gaussien

Preuve. La preuve se fait par récurrence, nous allons au sein de cette preuve montrer l'hérédité (le cas $k = 1$ étant très technique).

$$\text{Cov} \left(\frac{\partial Y}{\partial x_i}(x^{(1)}), \frac{\partial Y}{\partial x_i}(x^{(2)}) \right) = \frac{\partial^2 K(x^{(1)}, x^{(2)})}{\partial x_i \partial y_i}. \quad (2.2)$$

La relation 2.2 est montrée en écrivant

$$\begin{aligned} \text{Cov} \left(\frac{Y(x^{(1)} + he^{(i)}) - Y(x^{(1)})}{h}, \frac{Y(x^{(2)} + he^{(i)}) - Y(x^{(2)})}{h} \right) &= \frac{1}{h^2} K(x^{(1)} + he^{(i)}, x^{(2)} + he^{(i)}) \\ &\quad + \frac{1}{h^2} K(x^{(1)}, x^{(2)}) \\ &\quad - \frac{1}{h^2} K(x^{(1)} + he^{(i)}, x^{(2)}) \\ &\quad - \frac{1}{h^2} K(x^{(1)}, x^{(2)} + he^{(i)}). \end{aligned}$$

Maintenant, si deux variables aléatoires X_1 et X_2 convergent au sens des moindres carrés, leur covariance converge vers la covariance de leurs limites. D'où

$$\text{Cov} \left(\frac{\partial}{\partial x_i} Y(x^{(1)}), \frac{\partial}{\partial x_i} Y(x^{(2)}) \right) = \frac{\partial}{\partial x_i} \frac{\partial}{\partial y_i} K(x^{(1)}, x^{(2)}). \quad (2)$$

On utilise ensuite la propriété précédente pour conclure.

Remarque : Si on cherche à interpoler une fonction dont on connaît la régularité, il est donc cohérent de choisir un kernel deux fois plus régulier.

2.3 EXEMPLE IMPORTANT

Nous allons illustrer le théorème précédent avec l'aide d'un kernel (= matrice de covariance K) Matérn. Cet exemple est essentiel puisque le kernel Matérn va constituer une de nos bases de travail pour la suite de nos recherches. Le kernel Matérn est défini comme suivant :

$$K_{\text{Matérn}}(d) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu} d}{l} \right)^\nu K_\nu \left(\frac{\sqrt{2\nu} d}{l} \right)$$

. Dans cette fonction, nous avons :

- d : distance entre deux points considérés. Une fonction kernel doit être décroissante en d . En effet, plus deux points sont distants l'un de l'autre, moins l'image d'un point doit influencer l'image de l'autre
- σ^2 : Comme dans la plupart des modèles, le paramètre σ^2 est la variance du kernel. Au plus σ^2 est grand, au plus la largeur des trajectoires envisagée est grande.
- l : l est un paramètre d'échelle de longueur. Au plus l est grand, au plus Y corrèle deux points x_1 et x_2 . Ainsi, un l plus grand revient à avoir une fonction plus lisse
- ν : Paramètre de régularité. Y est k -fois différentiable au sens des moindres carrés et presque sûrement différentiable lorsque $\nu > k$.

La figure 3 permet de percevoir de manière pratique l'influence des différents paramètres sur les réalisations du processus. Ainsi, lorsque l'on va mettre en place un krigeage, il va s'agir d'adapter ces paramètres en fonction de la solution recherchée, des incertitudes que nous possédons sur nos mesures, de la forme attendue de solution, ...

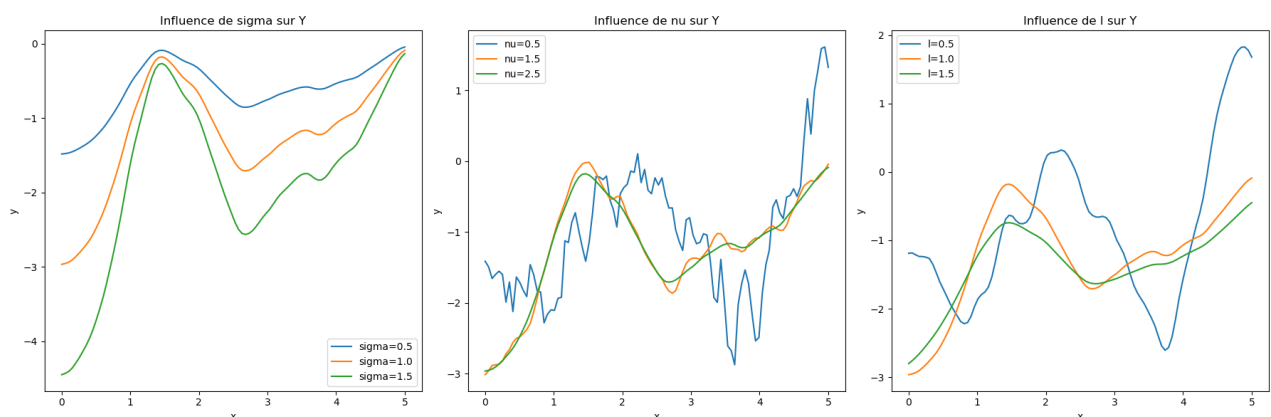


Figure 3: Analyse de l'Influence des Paramètres Sigma, Nu et Longueur d'Échelle sur les Réalisations d'un Processus Gaussien avec un Kernel Matern

Après cette brève introduction du kernel Matérn, revenons à l'application de la **Proposition 3**. Nous prendrons ici $\nu = 1.5$ de manière à ce que nos réalisations du processus Y soient de classe \mathcal{C}^1 .

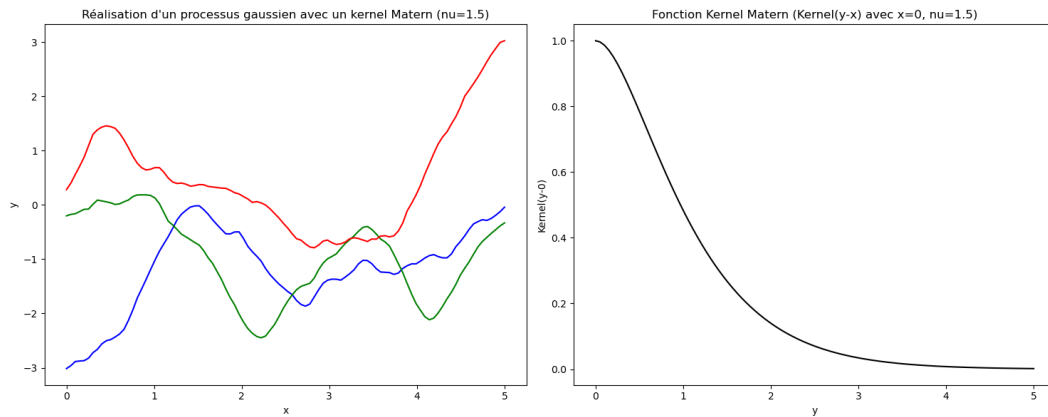


Figure 4: Réalisations de processus avec un kernel Matérn de classe \mathcal{C}^2 et graphe du kernel

Sur ce graphique, nous pouvons deviner que les réalisations de processus sont de classe \mathcal{C}^1 même s'il est évidemment impossible de le vérifier seulement à partir de ces graphes en pratiques.

3

ESTIMATION PARAMÉTRIQUE DE LA FONCTION DE COVARIANCE POUR DES MODÈLES DE KRIGEAGE

3.1 INTRODUCTION

Dans le cadre de notre étude sur les processus gaussiens et le krigeage, une question cruciale a rapidement émergé : quelle est l'influence du choix de la fonction de covariance et de ses paramètres sur les résultats obtenus ? À travers nos premières simulations et expériences de krigeage, nous avons observé que la nature de la fonction de covariance jouait un rôle déterminant dans les caractéristiques des fonctions estimées relativement aux fonctions initiales que l'on souhaitait retrouver, notamment en termes de régularité et de variations. Ce constat nous a conduit à nous intéresser de manière approfondie à l'estimation paramétrique de la fonction de covariance. Cette étape est fondamentale pour la modélisation précise et la prédiction efficace dans le cadre des processus gaussiens. Elle implique non seulement le choix d'une forme fonctionnelle adaptée mais aussi une estimation judicieuse des paramètres qui la définissent.

Cette partie de notre projet vise donc à explorer, d'un point de vue pratique, l'impact du choix des paramètres de la fonction de covariance, en soulignant leur influence significative sur la qualité et la pertinence des prédictions réalisées par le krigeage.

Nous considérons donc toujours un vecteur y de n observations scalaires aléatoires. Dans les deux sous-sections suivantes, nous donnons d'abord les définitions de base et les propriétés pour l'estimation du paramètre caractérisant la distribution inconnue de y . De plus, on se donne Y un processus gaussien, observé en $x^{(1)}, \dots, x^{(n)}$, avec le vecteur d'observation y . On a alors K la matrice de covariance de y et $k(x)$ le vecteur de covariance de Y entre $x^{(1)}, \dots, x^{(n)}$ et x . Alors, selon y , on a que le processus aléatoire Y est gaussien, avec une fonction moyenne $x \rightarrow m(x|y)$ et une fonction de covariance $(x_1, x_2) \rightarrow K(x_1, x_2|y)$.

3.2 NOTIONS GÉNÉRALES

Definition 9 Une famille paramétrique de distributions pour y est une famille paramétrique de distributions sur \mathbb{R}^n , définie par

$$\mathcal{P} = \{P_\psi, \psi \in \Psi\}$$

où P_ψ est une distribution sur \mathbb{R}^n et Ψ est un sous-ensemble de \mathbb{R}^p . À moins qu'il ne soit explicitement indiqué autrement, il existe $\psi^{(0)} \in \Psi$ tel que la distribution de y soit $P_{\psi^{(0)}}$. Nous soulignerons parfois cela en disant que le modèle \mathcal{P} est bien spécifié.

Definition 10 Un estimateur $\hat{\psi}$ est une fonction déterministe de \mathbb{R}^n vers Ψ . $\hat{\psi}(y)$ est l'estimation de ψ , selon le vecteur d'observations y .

Definition 11 Supposons que toutes les distributions P_ψ aient une fonction de densité de probabilité l_ψ sur \mathbb{R}^n . Alors, l'estimateur du Maximum de Vraisemblance (ML) $\hat{\psi}_{ML}$ de ψ est défini par

$$\hat{\psi}_{ML} \in \arg \max_{\psi \in \Psi} l_\psi(y)$$

L'estimateur ML est à la fois un estimateur M et un estimateur Z, sous des conditions de régularité sur la famille des fonctions de densité de probabilité $\{l_\psi, \psi \in \Psi\}$. L'estimateur ML est peut-être l'estimateur le plus étudié théoriquement, et le plus utilisé en pratique.

3.3 ESTIMATION DES PARAMÈTRES D'UNE FONCTION DE CO-VARIANCE

Il est déraisonnable de considérer toutes les fonctions de covariance possibles comme candidates pour le processus gaussien en question. Par conséquent, il est classique de supposer une famille paramétrique pour la fonction de covariance d'un processus gaussien Y . De plus, nous étudions spécialement le cas classique d'une famille de fonctions de covariance stationnaires. Ces deux remarques motivent la définition suivante d'une famille paramétrique de fonctions de covariance stationnaires.

Definition 12 Une famille paramétrique de fonctions de covariance stationnaires est de la forme

$$\{K_\psi, \psi \in \Psi\}$$

où K_ψ est une fonction de covariance stationnaire, et Ψ est un sous-ensemble de \mathbb{R}^p .

Puisque la variance $K_\psi(0)$ du processus gaussien stationnaire est constante, il est généralement judicieux de la considérer comme un paramètre explicite. Par conséquent, nous allons considérer la paramétrisation alternative de K_ψ dans la définition précédente,

$$\{\sigma^2 R_\theta, \sigma^2 > 0, \theta \in \Theta\}$$

où R_θ est une fonction de corrélation et Θ est un sous-ensemble de \mathbb{R}^{p-1} . La séparation explicite du super-paramètre de variance σ^2 et du super-paramètre de corrélation θ s'avèrera utile lorsque nous aborderons leur estimation.

3.4 ESTIMATEUR DU MAXIMUM DE VRAISEMBLANCE POUR LES PARAMÈTRES

Pour l'estimation, dans toutes les sous-sections, y est le vecteur des observations du processus gaussien Y en $x^{(1)}, \dots, x^{(n)}$. K_ψ est la matrice de covariance de y sous la fonction de covariance K_ψ et R_θ est la matrice de corrélation de y sous la fonction de corrélation R_θ . K_ψ et R_θ sont définies par $(K_\psi)_{i,j} = K_\psi(x^{(i)} - x^{(j)})$ et $(R_\theta)_{i,j} = R_\theta(x^{(i)} - x^{(j)})$. Nous supposons que, lorsque les points $x^{(1)}, \dots, x^{(n)}$ sont distincts, les matrices K_ψ et R_θ sont inversibles.

Considérons un processus gaussien observé en n points, notés $x^{(1)}, \dots, x^{(n)}$. Le vecteur des observations est représenté par $\mathbf{y} = [y(\mathbf{x}_1), y(\mathbf{x}_2), \dots, y(\mathbf{x}_n)]^\top$. L'objectif est d'estimer les paramètres de la fonction de covariance du processus gaussien.

Le principe du maximum de vraisemblance consiste à maximiser la probabilité des données observées \mathbf{y} sous le modèle gaussien. La fonction de vraisemblance pour un processus gaussien est donnée par :

$$\mathbf{L}(\psi) = \frac{1}{(2\pi)^{n/2} |K_\psi|^{1/2}} \exp \left(-\frac{1}{2} \mathbf{y}^\top K_\psi^{-1} \mathbf{y} \right) \quad (3)$$

où K_ψ est la matrice de covariance construite à partir des paramètres ψ et des données. Pour trouver les estimations des paramètres, nous maximisons le logarithme de la vraisemblance :

$$\ln \mathbf{L}(\psi) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \ln |K_\psi| - \frac{1}{2} \mathbf{y}^\top K_\psi^{-1} \mathbf{y} \quad (4)$$

L'optimisation de cette fonction par rapport à ψ nous fournit les estimations des paramètres de la fonction de covariance. Cette approche est cruciale car elle permet de calibrer le modèle de manière à capturer au mieux la structure de dépendance spatiale des données.

Krigeage Simple

Dans le cas du Krigeage simple, le critère de maximum de vraisemblance du vecteur d'observations \mathbf{y} dépend uniquement de ψ (après légère modification de la formule du maximum de vraisemblance) et consiste à minimiser,

$$L(\psi) := \frac{1}{n} [\ln |K_\psi| + \mathbf{y}^t K_\psi^{-1} \mathbf{y}]$$

Dans le cas du Krigeage simple, l'estimateur du maximum de vraisemblance du super-paramètre de covariance ψ dans la est

$$\hat{\psi}_{ML} \in \arg \min_{\psi \in \Psi} L(\psi)$$

Maintenant, dans le cas de la séparation des hyper-paramètres de variance et de corrélation dans la définition 12, le critère de vraisemblance devient

$$L(\sigma^2, \theta) := \frac{1}{n} \ln |\sigma^2 R_\theta| + \frac{1}{n} \frac{1}{\sigma^2} \mathbf{y}^t R_\theta^{-1} \mathbf{y} \quad (5)$$

Ainsi, l'optimisation par rapport à σ^2 , pour θ fixé, peut être effectuée explicitement. Cela élimine une dimension dans le problème d'optimisation numérique. Ceci est résumé dans la proposition suivante.

Proposition 4 *L'estimateur du maximum de vraisemblance de (σ^2, θ) est $\hat{\sigma}_{ML}^2, \hat{\theta}_{ML}$, avec :*

$$\hat{\theta}_{ML} \in \arg \min_{\theta \in \Theta} L(\theta) \quad (6)$$

avec

$$L(\theta) = \ln(\hat{\sigma}_{ML}^2(\theta)) + \frac{1}{n} \ln |R_\theta|$$

$$\sigma_{ML}^2(\theta) = \frac{1}{n} \mathbf{y}^t R_\theta^{-1} \mathbf{y}$$

et

$$\hat{\sigma}_{ML}^2 = \hat{\sigma}_{ML}^2(\hat{\theta}_{ML})$$

Krigeage Universel

Considérons maintenant le cas du Krigeage ordinaire ou universel, où la fonction moyenne n'est plus considérée nulle : nous notons par H la matrice $n \times m$ telle que

$$H_{i,j} = g_j(x^{(i)})$$

où la fonction moyenne est supposée être de la forme $\sum_{i=1}^m \beta_i g_i$, avec des fonctions connues g_i et des coefficients inconnus β_i . Nous notons également par h le vecteur $m \times 1$ tel que $h_j = g_j(x_{\text{new}})$. Cependant, nous utilisons toujours explicitement la décomposition σ^2, θ . Par conséquent, nous présenterons les équations du maximum de vraisemblance uniquement dans ce cas.

Le critère de vraisemblance dépend maintenant de β, σ^2 et θ et est donné par :

$$L(\beta, \sigma^2, \theta) := \frac{1}{n} \ln |\sigma^2 R_\theta| + \frac{1}{n\sigma^2} (\mathbf{y} - H\beta)^t R_\theta^{-1} (\mathbf{y} - H\beta) \quad (7)$$

De manière similaire au cas du Krigeage simple, le critère de vraisemblance de 7 peut être minimisé explicitement par rapport à β et σ^2 , en éliminant $m+1 = \dim(\beta)+1$ dimensions dans le problème d'optimisation numérique. Ceci est résumé dans la proposition suivante.

Proposition 5 *L'estimateur du maximum de vraisemblance de $(\hat{\beta}, \hat{\sigma}^2, \theta)$ est $(\hat{\beta}_{ML}, \hat{\sigma}_{ML}^2, \hat{\theta}_{ML})$, avec :*

$$\hat{\theta}_{ML} \in \arg \min_{\theta \in \Theta} L(\theta), \quad (8)$$

avec

$$L(\theta) = \ln(\hat{\sigma}_{ML}^2(\theta)) + \frac{1}{n} \ln |R_\theta|,$$

$$\hat{\sigma}_{ML}^2(\theta) = \frac{1}{n} (\mathbf{y} - H\hat{\beta}_{ML}(\theta))^t R_\theta^{-1} (\mathbf{y} - H\hat{\beta}_{ML}(\theta))^t$$

$$\hat{\beta}_{ML}(\theta) = (H^t R_\theta^{-1} H)^{-1} H^t R_\theta^{-1} \mathbf{y}$$

et

$$\hat{\beta}_{ML} = \hat{\beta}_{ML}(\hat{\theta}_{ML}), \hat{\sigma}_{ML}^2 = \hat{\sigma}_{ML}^2(\hat{\theta}_{ML}).$$

De plus, on peut aussi écrire:

$$\hat{\sigma}_{ML}^2(\theta) = \frac{1}{n} \mathbf{y}^t \Pi_\theta \mathbf{y},$$

avec :

$$\Pi_\theta = R_\theta^{-1} - R_\theta^{-1} H (H^t R_\theta^{-1} H)^{-1} H^t R_\theta^{-1}$$

3.5 APPLICATION AU KERNEL MATÉRN

Modèle de Matérn sur \mathbb{R}

La partie sur la différentiabilité du noyau et ses liens avec la différentiabilité en moyenne quadratique du processus gaussien donne la motivation pour un modèle de covariance où la régularité à zéro est ajustable, ou de manière équivalente, où le taux de disparition à $+\infty$ de la transformée de Fourier de la fonction de covariance est ajustable. Le modèle de Matérn satisfait à cela, et son utilisation systématique pour modéliser les processus gaussiens stationnaires est donc recommandée. On rappelle que le modèle de Matérn est paramétré par les hyper-paramètres $\sigma^2 > 0$, $l > 0$ et $\nu > 0$ et est défini par

$$\hat{K}(\omega) = \sigma^2 \frac{\Gamma(\nu + \frac{1}{2})(2\sqrt{\nu})^{2\nu}}{l^{2\nu} \sqrt{\pi} \Gamma(\nu)} \frac{1}{(\frac{4\nu}{\lambda^2} + \omega^2)^{\frac{1}{2} + \nu}}$$

On peut déduire que $\omega^2 k \hat{K}(\omega)$ est sommable dès que $\nu > k$. Par conséquent, ν est appelé l'hyper-paramètre de lissage et Y est k fois moyennement carré différentiable et k fois presque sûrement différentiable dès que $\nu > k$. Les deux autres hyper-paramètres σ^2 et l ont, nous trouvons, une interprétation plus claire après avoir donné l'équivalent temporel de (2.4),

$$K_{\text{Matérn}}(d) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu} d}{l} \right)^\nu K_\nu \left(\frac{\sqrt{2\nu} d}{l} \right)$$

Initialisation avec un fonction déterministe On se donne une fonction déterministe $f : x \mapsto \sin(x)$ et un ensemble de points d'observation en x_1, \dots, x_{10} , avec le vecteur d'observation y .

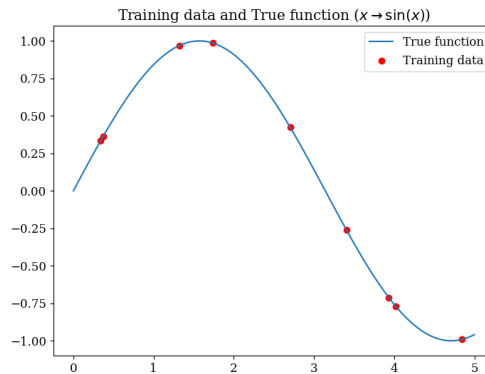


Figure 5: Fonction $f : x \mapsto \sin(x)$ et 10 points d'observation

Nous implémentons alors une méthode de minimisation pour obtenir les paramètres du kernel Matérn par maximum de vraisemblance et nous obtenons les résultats suivants.

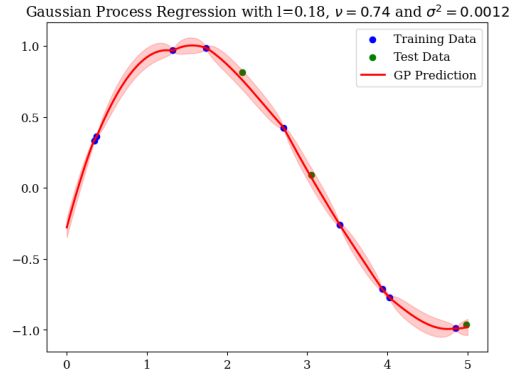


Figure 6: Krigeage de la fonction $f : x \mapsto \sin(x)$ avec les paramètres obtenus par maximum de vraisemblance

Extrapolation à l'interpolation de la réalisation d'un processus gaussien On définit un kernel Matérn avec des paramètres préféfinis. Nous pouvons alors logiquement espérer retrouver ces paramètres par une méthode de maximum de vraisemblance. Pour cela nous implémentons une réalisation d'un processus gaussien avec les paramètres $l = 1$, $\nu = \frac{5}{2}$ et $\sigma = 0.1$.

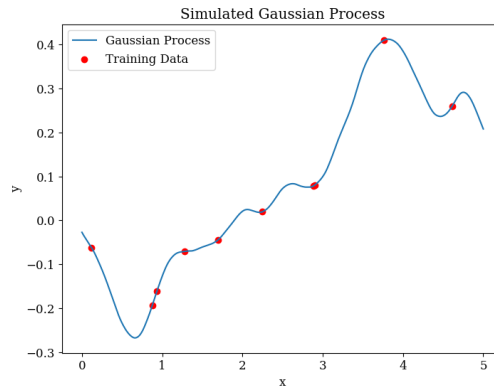


Figure 7: Réalisation d'un processus gaussien avec le kernel Matérn ($l = 1$, $\nu = \frac{5}{2}$ et $\sigma = 0.1$)

Nous appliquons ensuite la méthode de calcul par maximum de vraisemblance et nous obtenons le krigeage suivant :

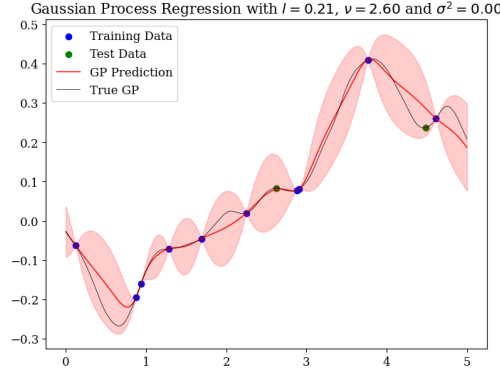


Figure 8: Krigeage d'un processus gaussien avec le kernel Matérn ($l = 1$, $\nu = \frac{5}{2}$ et $\sigma = 0.1$) par maximum de vraisemblance

Afin de comparer avec les paramètres initiaux, nous avons tenté d'effectuer le krigeage avec un kernel dont les paramètres sont exactement ceux utilisés pour réaliser la première fonction. Nous obtenons les résultats suivants.

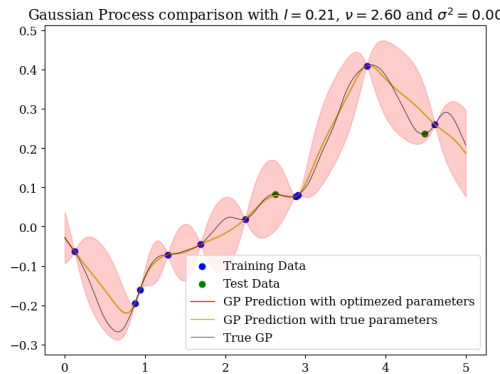


Figure 9: Comparaison du krigeage avec le vrai kernel Matérn ($l = 1$, $\nu = \frac{5}{2}$ et $\sigma = 0.1$) et celui obtenu par maximum de vraisemblance

On constate alors une faible différence pour les trajectoires de l'espérance des 2 kernels proposées mais afin de quantifier la différence par rapport à la vraie fonction qu'il fallait interpoler à l'origine nous pouvons regarder l'erreur quadratique des 2 propositions. Dans le cas du maximum de vraisemblance, l'erreur est $MSE = 0.001229$ tandis que pour le krigeage avec les vrais paramètres, l'erreur est de $MSE = 0.001235$. Nous avons ici gagné 0.5% de précision. Cependant pour d'autres exemples de réalisation, nous avons parfois eu jusqu'à 20% de précision supplémentaire par maximum de vraisemblance, par rapport au krigeage avec les "vrais" paramètres. Finalement, on peut essayer d'interpréter cela en sachant que le processus gaussien obtenu en entrée n'est qu'une réalisation d'un événement ω qui a abouti au processus gaussien dont on a la représentation. Or il se peut totalement que ce processus gaussien soit mieux approximé par un kernel matern aux paramètres différents de ceux introduits lors de la création du premier processus. Cela est confirmé par le calcul de la moyenne des erreurs

quadratiques entre le processus de base et un krigeage avec les paramètres optimisés d'une part et les vrais paramètres d'autres part.

CONCLUSION

La particularité du krigeage comme méthode d'interpolation est qu'elle fait intervenir le choix d'un kernel. Lors de notre étude, nous avons donc voulu comprendre comment choisir le kernel et quels sont les conséquences d'un tel choix sur la fonction reconstituée. Quant aux conséquences du choix de kernel sur la fonction reconstituée, la proposition 3 de la section 2.2 nous assure que la régularité du kernel implique une certaine régularité sur la fonction reconstituée. Pour choisir le kernel qui permet la meilleure interpolation possible, nous nous sommes intéressés notamment à la famille des kernels Matérn qui présentent deux avantages : c'est une famille large et diverse de kernels, et elle peut être paramétrisée par 3 paramètres. Nous avons vu que le choix du kernel revenait à un problème d'optimisation sur ces paramètres. Nous avons pu implémenter numériquement ce choix de kernel optimal et vérifier sa pertinence. Les résultats numériques sont convaincants : le choix du kernel comme solution du problème d'optimisation permet en effet une meilleure interpolation.

REFERENCES

- [1] M. Ebden *Gaussian Processes for Regression: A Quick Introduction*. August 2008.
- [2] F. Bachoc *Asymptotic analysis of maximum likelihood estimation of covariance parameters for Gaussian processes: an introduction with proofs*. Université de Toulouse, September 2020.
- [3] J. Wang *An Intuitive Tutorial to Gaussian Processes Regression*. Queen's University, April 2022.
- [4] J. Melo *Gaussian Processes for regression: a tutorial*. University of Porto.
- [5] F. Bachoc, E. Contal, H. Maatouk, D. Rulli re *Gaussian processes for computer experiments*. Institut de Math matiques de Toulouse, ENS Cachan, INRIA Rennes, ISFA Universit  Lyon 1, 2017.
- [6] J-C. Breton *Processus Gaussiens, Master IMA 2 me ann e*. Universit  de La Rochelle, September-December 2006.
- [7] F. Bachoc *Lecture notes Gaussian processes and sensitivity analysis for computer experiments*. University Paul Sabatier, January 2020.
- [8] F. Bachoc *Estimation param trique de la fonction de covariance dans le mod le de Krigeage par processus Gaussiens. Application   la quantification des incertitudes en simulation num rique*. Universit  Paris Diderot, Octobre 2013.