Planification et analyse de données spatio-temporelles

Papa Abdoulaye Faye

AUVERGNE
La Région juste et grande

Université Blaise Pascal

UNION EUROPÉENNE en AUVERGNE

***N°* d'Ordre : D.U. 2638**

UNIVERSITE BLAISE PASCAL

U.F.R. Sciences et Technologies

ECOLE DOCTORALE DES SCIENCES FONDAMENTALES

N° 844

THÈSE

présentée pour obtenir le grade

DOCTEUR D'UNIVERSITE

***Spécialité* : Mathématiques appliquées**

Par : **Papa Abdoulaye Faye**

TITRE DE LA THÈSE :

Planification et Analyse de données spatio-temporelles

Soutenue publiquement le 08 Décembre 2015, devant la commission d'examen.

Président : **Denys Pommeret**, I2M, Rapporteur

Examinateur :

M. Pierre Druilhet,	Professeur, UBP	Directeur de thèse
M. Nourddine Azzaoui,	Maître de Conférence, UBP	Co-directeur
Mme Anne-Françoise Yao,	Professeur, UBP	Co-directeur
M. Laurent Clavier,	Professeur, TELECOM Lille	Rapporteur
Mme Sophie Dabo-Niang,	Professeur, Université Lille 3	Rapporteur
Mme Céline Helbert,	Maître de Conférence, E.Central Lyon	Examinateur
M. Roland Gourvès,	Chef d'entreprise, Sol Solution	Membre invité

Remerciements

Comme ce travail n'est pas l'œuvre d'une seule personne, je tiens à remercier l'ensemble des acteurs, qui ont pu jouer un rôle dans son aboutissement.
En tout premier lieu, je voudrais citer mes directeurs de thèses et collègues les plus proches, Pierre Druilhet, Nourddine Azzaoui et Anne-Françoise Yao. Ils m'ont aidé à mener ce travail à son terme dans de très bonnes conditions grâce à leur disponibilité, leurs avis et leurs conseils et je les en remercie grandement.
Je voudrais également remercier Sophie Dabo-Niang, Denys Pommeret et Laurent Clavier, qui ont accepté d'être les rapporteurs de la thèse, ainsi que Céline Helbert et Roland Gourvès, qui ont accepté de participer au jury de soutenance.
Je remercie également l'ensemble du personnel du laboratoire de mathématiques et notamment François Bouchon pour son aide et pour sa disponibilité. Je ne peux me permettre d'oublier les entreprises du cluster e2ia (Sol Solution, Numtech, Athos Environnement, Biobasic Environnement) de m'avoir ouvert leurs portes pour s'imprégner de leurs besoins par rapport au projet. Merci à l'ensemble du personnel technique et administratif, qui a fait un très bon travail d'accompagnement et d'aide à des moments importants de la découverte du monde universitaire.
Merci également aux doctorants du laboratoire de mathématiques, Christèle, Jordane, Mahdi, Yacouba, Romuald, Victor et tous les autres, pour leur bonne humeur, leur implication dans la vie du laboratoire et les discussions diverses.
À l'extérieur du monde universitaire, d'autres personnes ont joué un rôle important de soutien scientifique, logistique ou moral. Il s'agit des amis, particulièrement ceux de Clermont-Ferrand, que je remercie infiniment. Je finis en remerciant sincèrement toutes les membres de ma famille d'avoir été à mes côtés pendant les moments les plus difficiles.

Je dédie ce travail à ma maman d'amour sans qui je ne serai pas là. Je le dédie aussi à la mémoire de mon père.

Table des matières

Chapitre 1

Introduction

Dans de nombreux domaines des sciences de l'environnement (géophysique, géologie, hydrologie, biologie,...) le phénomène étudié évolue suivant une dynamique spatio-temporelle. L'étude des mécanismes régissant ce processus requiert le développement de techniques statistiques propres. Cependant, du fait de la complexité du comportement de ces phénomènes, la construction de modèles spatio-temporels est souvent peu aisée. Plusieurs modèles existent dans la littérature. Citons par exemple les travaux précurseurs de Guting [51] et Tansel [124] ainsi que la revue bibliographique des modèles spatio-temporels de Pekelis [105] et plus récemment le livre de Cressie [25] sur la statistique pour les données spatio-temporelles. Cette branche de la statistique fournit des modèles et des méthodes permettant d'estimer les caractéristiques du processus sous-jacent à partir de données observées en des sites sélectionnés.

Comme en statistique classique, le coût de la collecte de données oblige souvent en pratique à restreindre le nombre d'observations. Une stratégie de sélection des sites d'observations à l'aide de plan d'expériences est donc indispensable pour la suite de l'étude. Par exemple, un choix optimal des emplacements des capteurs permettant l'observation de paramètres environnementaux conduira à une meilleure estimation du comportement du phénomène et donc à une meilleure prévision spatio-temporelle de valeurs non-observées.

Dans cette thèse, nous privilégions l'approche bayésienne car elle permet de combiner les différentes sources d'information : l'information spatiale apportée par les observations, l'information temporelle apportée par la boîte noire ainsi que l'information a priori connue du phénomène. Elle permet également de prendre en compte les incertitudes liées à la modélisation. Un autre intérêt de l'approche Bayésienne est de fournir de manière naturelle une cartographie de lois de probabilité du phénomène étudié au lieu d'une simple carte de prévision, ce qui permet de mieux

mesurer l'incertitude des prévisions.

Un des principaux objectifs de cette thèse est de proposer des alternatives aux méthodes déterministes numériques (approximation, maillage, équations aux dérivées partielles, boîte noire...) de prédiction/cartographie spatio-temporelle. Dans ce contexte, l'approche que nous proposons consiste à coupler des modèles numériques et statistiques pour obtenir une meilleure prédiction spatio-temporelle des phénomènes d'intérêt.

1.1 Objectifs de la thèse

Cette thèse a été financée par le conseil régional d'Auvergne et le FEDER (Fonds européen de développement économique régional). L'objectif était d'une part, de développer des techniques novatrices en terme de modélisation spatio-temporelle et d'autre part de fournir de nouveaux outils et algorithmes efficaces pouvant être utiles aux entreprises partenaires du Cluster E2IA Auvergne.

En effet, pour spatialiser leurs données, ces entreprises utilisent différents types de techniques (Interpolation linéaire, Krigeage, etc). Cependant, le périmètre d'application des techniques utilisées est souvent partiellement connu, et d'autres méthodes pourraient être mieux adaptées à leurs problématiques. De plus, bien que les données soient spatio-temporelles, les informations spatiales et temporelles sont traitées séparément. Ce projet de thèse a été mis en place pour fournir des solutions à ce type de problématique.

Plus précisément, les objectifs du projet étaient :

- Proposer des méthodes valides qui permettront la prédiction spatio-temporelle d'une variable d'intérêt en des sites ne disposant pas d'observations à partir des données disponibles tout en évaluant les incertitudes sur les valeurs prédites.
- Fournir un critère de construction de plans d'expériences optimaux qui tienne compte des connaissances *a priori* sur le phénomène étudié.
- Adapter les méthodes spatio-temporelles au contexte Bayésien pour tenir compte de l'information *a priori* qui existe sur le phénomène d'intérêt et sur le domaine d'étude.
- Proposer et implémenter des algorithmes séquentiels qui permettront de faire de la prédiction spatiale des valeurs non-observées du processus à différentes dates.
- Développer des méthodes génériques suffisamment flexibles pour s'adapter à la spécificité des applications rencontrées par chaque entreprise du cluster E2IA.

1.2 Structure de la thèse

Les différents points évoqués ci-dessus peuvent être regroupés en deux thèmes principaux :

- La modélisation bayésienne spatio-temporelle : elle consiste à donner des outils mathématiques *génériques* permettant de modéliser un phénomène évoluant dans l'espace et dans le temps. En effet, les outils développés dépendent beaucoup du format des données d'entrée, de leur origine, leur nombre, leur plan d'échantillonnage (les données équi-réparties en 3D, en 2D, image ...). La complexité et le nombre limité de données posent des problèmes de modélisation pour l'interpolation spatiale ; on est souvent amené à faire des hypothèses qui peuvent s'avérer réductrices d'information dans certains cas. Les interpolateurs utilisés vont des plus simples aux plus sophistiqués (krigeage ou autres méthodes stochastiques). Pour tirer profit de l'information existante sur le domaine spatial étudié et/ou le comportement du phénomène d'intérêt, les techniques bayésiennes ont été utilisées non seulement pour donner une estimation ponctuelle mais aussi des cartes de lois de probabilités, donnant ainsi la possibilité à l'utilisateur de calculer n'importe quel indicateur statistique concernant la variable d'intérêt.
- Le choix des sites d'observation ou plan d'expérience : Le déploiement spatial des sites d'observations utilise habituellement des modèles basés sur des champs homogènes. Cette hypothèse est loin d'être réaliste du faite de l'existence des régions vides ou ayant des densités variables selon les zones. La méthodologie des plans d'expériences consiste dans un premier temps à définir un critère de qualité globale d'un plan en fonction de l'objectif visé, puis, dans un deuxième temps, de chercher le plan maximisant ce critère à l'aide d'un algorithme d'optimisation.

Cette thèse est structurée en cinq chapitres :

Le chapitre 2 fait un rappel des principes généraux de la géostatistique, des propriétés de processus stochastiques spatiaux, en particulier gaussiens, et des méthodes de Krigeage.

Le krigeage étant la méthode d'interpolation spatiale la plus utilisée en géostatistique, nous détaillons ses différentes représentations paramétriques ainsi que son approche Bayésienne. Nous discutons également du choix des différentes lois a priori pour les paramètres du modèle et explicitons les lois a posteriori qui en découlent.

Le chapitre 3 est consacré dans un premier temps à la présentation de modèles d'évolution temporelle déterministe, tels que les équations différentielles ordinaires ou équation aux dérivées partielles. Nous détaillons trois exemples de système dynamique différentiel dont l'un sera utilisé

pour tester les méthodes proposées dans les chapitres suivants. Afin de garder une approche générale, nous qualifions de *boîte noire* tout modèle temporel déterministe, ce qui inclut les modèles dynamiques différentiels mais également toute modélisation implémentée sous forme de code numérique. Lorsque les paramètres de la boîte noire sont mal connus, on dira que boîte noire est *mal spécifiée*. Dans la deuxième partie de ce chapitre, nous faisons une revue de quelques modèles spatio-temporels comme le co-krigeage multifidélité et le filtre de Kalman.

Le chapitre 4 qui fait l'objet d'un article soumis est consacré à la construction d'un algorithme de krigeage spatio-temporel bayésien utilisant une boîte noire mal spécifiée. Dans cet article, la boîte noire est utilisée comme modèle temporel. Les informations a priori apportées par la boîte noire mal spécifiée sont utilisées pour proposer les lois a priori utilisées dans le krigeage bayésien. Nous montrons que cette approche améliore la qualité prédictive du krigeage Bayésien. L'algorithme permet ainsi d'obtenir de manière séquentielle une bonne prédiction de la variable d'intérêt dans le domaine d'étude.

Au chapitre 5, la boîte noire est utilisée comme modèle temporel, mais nous modifions la structure paramétrique du krigeage classique. La prédiction de la variable d'intérêt en des sites de non-observation se fait sans imposer une paramétrisation de la fonction covariance. En fait, au lieu de décrire cette dernière par trois paramètres, nous utilisons le complément de Schur pour l'estimer directement de manière empirique à partir de cartes simulées.

Le chapitre 6, est consacré aux plans d'expériences. Dans une première partie, nous faisons une revue bibliographique de ces techniques. Dans une deuxième partie, nous établissons un critère d'optimalité qui se base sur l'information que fournit la boîte noire. A l'aide d'un algorithme d'échange, nous cherchons un plan optimisant ce critère.

Chapitre 2

La Géostatistique

2.1 définition et notation

À la frontière entre les mathématiques et les domaines liés à l'environnement, la géostatistique est l'étude d'une variable régionalisée(VR). Cette dernière est toute fonction déterministe destinée à modéliser un phénomène qui se déploie dans l'espace et/ou le temps et y manifeste une certaine structure. Historiquement les premières utilisations du vocabulaire et du concept de variables régionalisées concernaient presque exclusivement la répartition des teneurs minéralisés dans un gisement minier [86, 88] ; cet outil a par la suite trouvé des applications dans des domaines aussi variés que la météorologie [95, 129], la sylviculture, la bathymétrie, la topographie (MNT), l'environnement [30, 35, 48, 58, 80, 110], l'agriculture de précision [118], l'halieutique [15], l'épidémiologie [93], le génie civil, l'hydrologie [76, 127], géologie [125], toute cartographie quantitative en général, etc.
En langage plus formel on dira que, pour un domaine d'étude $\mathcal{D} \subset \mathbb{R}^d$, $z(s)$, $s \in \mathcal{D}$ est une variable régionalisée si elle désigne la valeur au point s du phénomène étudié. Bien qu'une variable régionalisée dispose d'une haute fluctuation et d'un caractère imprévisible, elle doit cependant refléter à sa manière les caractéristiques structurelles du phénomène régionalisé étudié.

Il existe deux branches principales dans la géostatistique : la démarche transitive [88, 89] et celle intrinsèque [26, 27]. Les considérations et les hypothèses faites sur l'étude de la variable régionalisée indiquent quelle branche de la géostatistique on utilise. Pour la démarche transitive la variable régionalisée est considérée déterministe et son étude est faite sans hypothèse supplémentaire. Tandis-que pour la branche intrinsèque la variable régionalisée est vue comme une réalisation d'une fonction aléatoire $Z(s)$; appelée aussi processus stochastique. Toute valeur

régionalisée $z(s)$ est considérée comme une réalisation d'une variable aléatoire $Z(s)$, s étant un point du domaine $\mathcal{D}$. En plus en géostatistique intrinsèque on se donne généralement deux hypothèses clés lors de l'étude de la variable régionalisée que sont les hypothèses de stationnarité et d'ergodicité. La première permet de caractériser le phénomène à travers sa moyenne et sa covariance et la deuxième permet l'estimation empirique de ces deux la.

Malgré que la géostatistique transitive a l'avantage d'être simple, elle offre un spectre d'applications beaucoup plus restreint qu'en géostatistique intrinsèque. En outre elle se limite à certains types d'échantillonnage (e.g. régulier, aléatoire stratifié ou processus ponctuel) [15].
Dans la suite de cette thèse nous ne parlerons que de la géostatistique intrinsèque qui utilise un processus stochastique, en général gaussien, pour l'étude d'un phénomène régionalisé.

2.2 Processus aléatoire

Considérons un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$, un espace mesurable $(E, \mathcal{B}(E))$ et $\mathcal{D}$ un ensemble arbitraire. Un processus stochastique ou aléatoire $(Z(s), s \in \mathcal{D})$ est une collection de variables aléatoires définies sur $(\Omega, \mathcal{F}, \mathbb{P})$ et à valeur dans $(E, \mathcal{B}(E))$.

2.2.1 Processus stationnaire

Définition 2.2.1. *Un processus aléatoire $Z(.)$ est dit stationnaire de second ordre ou faiblement stationnaire ssi :*

a) L'espérance du processus est constante pour tout s appartenant à D

$$\mathbb{E}[Z(s)] = c, \quad \text{où } c \text{ une constante indépendante de } s. \tag{2.1}$$

b) La covariance de $Z(.)$ aux points s et $s+h$ existe et dépend uniquement de h ; le vecteur de translation entre ces points.

$$Cov(Z(s), Z(s+h)) = k(s, s+h) \triangleq C(h). \tag{2.2}$$

La fonction de covariance stationnaire $C(.)$ est nommée aussi ***covariogramme.***

L'une des propriétés importantes de la covariance est qu'elle est semi-définie positive, théorème de Bochner . Cette propriété est d'une importance capitale dans la théorie et la pratique. Nous la rappelons dans la définition suivante :

Définition 2.2.2. *Un noyau k est dit semi-définie positive si pour tout $(a_i)_{i=1,\dots,N} \in \mathbb{R}$, $N \in \mathbb{N}^*$ et tous $(s_i)_{i=1,\dots,N} \in D$, il satisfait la propriété suivante :*

$$\sum_{i,j=1}^{N} a_i a_j k(s_i, s_j) \geq 0$$

Remarques : Les fonctions de covariance stationnaire vérifient les propriétés suivantes [14, 120] :

1. Par la positivité nous avons $C(0) \geq 0$
2. Le covariogramme est paire $C(h) = C(-h)$
3. L'inégalité de Cauchy-Schwartz implique que $|C(h)| \leq C(0)$
4. Les covariances stationnaires vérifient une propriété de stabilité que nous résumons comme suit : si $C_1(.)$, $C_2(.)$ sont des covariances stationnaires, les fonctions suivantes le sont aussi :
   - $C(h) = a_1 C_1(h) + a_2 C_2(h)$, a_1 et $a_2 \geq 0$
   - $C(h) = C_1(ah)$, $a > 0$
   - $C(h) = C_1(h) C_2(h)$

La stationnarité du second ordre peut être limitée dans la mesure où, pour certains processus aléatoires, le covariogramme n'existe pas. Elle peut aussi être inadéquate, notamment sur les phénomènes régionalisés présentant une dispersion infinie (c'est-à-dire qui ont une variance infinie) ; c'est le cas par exemple des gisements d'or selon Krige [66]. Donc pour gagner en généralité et pour être en adéquation avec certains phénomènes à étudier, il peut préférable d'utiliser la stationnarité intrinsèque qui est définie comme suit :

Définition 2.2.3. *Un processus aléatoire $Z(.)$ est stationnaire intrinsèque ssi :*

a) L'espérance de tout accroissement $Z(s) - Z(s+h)$ est nulle

$$\mathbb{E}(Z(s) - Z(s+h)) = 0 \qquad \forall s \in D. \tag{2.3}$$

b) La variance de tout accroissement $Z(s) - Z(s+h)$ existe et dépend uniquement de h

$$Var(Z(s) - Z(s+h)) = 2\gamma(h) \qquad \forall s, s+h \in D. \tag{2.4}$$

La fonction $2\gamma(.)$ est nommée ***variogramme*** *et ne dépend que du vecteur de translation.*

Il y a un gain en généralité si on utilise la stationnarité intrinsèque par rapport à celle du second ordre parce que tout processus stationnaire du second ordre est intrinsèque. En effet dans le cas stationnaire de second ordre, l'équation

$$2\gamma(h) = 2(C(0) - C(h)), \tag{2.5}$$

relie le **variogramme** au **covariogramme**. L'implication inverse n'est pas vrai en général. Elle est vrai sauf si le variogramme est borné.

Exemple 2.2.1. *Le mouvement brownien* $(W(s), s \in \mathbb{R}^d)$ *vérifie cette propriété de stationnarité intrinsèque* $Var(W(s+h)-W(s)) = \|h\|$. *Par contre la* $Cov(W(s), W(s+h)) = (\|s\| + \|s+h\| - \|h\|)/2$ *ne dépend pas seulement de* h. *Donc le mouvement brownien est stationnaire intrinsèque mais pas du second ordre.*

Une propriété importante du variogramme est le fait qu'il soit une fonction conditionnellement semi-définie négative (de type négatif [13, chap. 3]). Une telle fonction est définie comme suit.

Définition 2.2.4. *Une fonction* $f : \mathbb{R} \to R^d$ *est dite conditionnellement semi-définie négative si* $\forall l$, $\forall s_1, \ldots, s_l \in \mathbb{R}^d$, $\forall \lambda_1, \ldots, \lambda_l \in \mathbb{R}$ *vérifiant* $\sum_{i=1}^{l} \lambda_i = 0$,

$$\sum_{i=1}^{l} \sum_{j=1}^{l} \lambda_i \lambda_j f(s_i - s_j) \leq 0$$

Le théorème suivant tiré de Cressie [27] caractérise bien cette propriété importante :

Théorème 2.2.1. *Toute fonction conditionnellement semi-définie négative est le variogramme d'un processus aléatoire intrinsèque.*

Le variogramme vérifie aussi les propriétés suivantes :

1. La parité : $\gamma(h) = \gamma(-h)$.
2. La positivité : $\gamma(h) \geq 0$.
3. Nullité en zéro $\gamma(0) = 0$.

Généralement le variogramme hérite de la plupart des propriétés de la covariance tout en gardant cette mesure de degré de liaison entre les variables régionalisées

2.2.2 Exemples de processus stationnaire

– **Bruit Blanc Fort** :

Il s'agit d'un processus aléatoire $(Z(s), s \in \mathcal{D})$ centré, dont les composantes sont indépendantes et identiquement distribuées (i.i.d). Ce processus est stationnaire intrinsèque sur $\mathcal{D}$

– **Bruit Blanc Faible** :

Les composantes de ce processus sont centrées, de variance finie et dé-corrélées : c'est-à-dire $\text{Var}(Z(s)) = \sigma^2 < \infty$ et $(\forall s \neq \tilde{s}, \text{Cov}(Z(s), Z(\tilde{s})) = 0)$. On montre facilement que ce processus est stationnaire intrinsèque

2.2.3 Processus Gaussiens

Définition 2.2.5. *Un processus stochastique est dit gaussien si toutes ses lois fini-dimensionnelles $\mathcal{L}(Z(s_1), \ldots, Z(s_n))$ sont gaussiennes. Autrement dit $(Z(s), s \in \mathcal{D})$ est gaussien ssi toute combinaison linéaire $a_1 Z(s_1) + \cdots + a_n Z(s_n)$ suit une loi normale, pour tous $n \in \mathbb{N}^*, s_1, \ldots, s_n \in \mathcal{D}$ et $a_1, \ldots, a_n \in \mathbb{R}$*

La loi d'un vecteur gaussien $(Z(s_1), \ldots, Z(s_n))$ est caractérisée par le vecteur des moyennes $(\mathbb{E}[Z(s_1)], \ldots, \mathbb{E}[Z(s_n)])$ et la matrice de variance-covariance $(\text{Cov}(Z(s_i), Z(s_j)))_{1 \leq i,j \leq n}$.
On comprend dés lors que toute la loi d'un processus gaussien est connue dés qu'on se donne la fonction moyenne

$$\mu(s) = \mathbb{E}[Z(s)] \qquad s \in \mathcal{D}. \tag{2.6}$$

et l'opérateur de covariance

$$k(s, \tilde{s}) = \text{Cov}(Z(s), Z(\tilde{s})) \qquad s, \tilde{s} \in \mathcal{D}. \tag{2.7}$$

En effet, la loi fini-dimensionnelle de $(Z(s_1), \ldots, Z(s_n))$ est alors la loi normale multi-dimensionnelle $\mathcal{N}(\mu_n, K_n)$ avec $\mu_n = (\mu(s_1), \ldots, \mu(s_n))$ et $K_n = (k(s_i, s_j))_{1 \leq i,j \leq n}$

La moyenne $\mu(s)$ d'un processus gaussien représente sa tendance. Par exemple dans un cadre de régression linéaire généralisée (GLM), la fonction moyenne a la forme $\mu(s) = m'(s)\beta$, avec $m'(s) = (m_1(s), \ldots, m_p(s))$ où $m'(s)$ est une famille de fonctions (base) dites aussi regrésseurs. L'importance ou le poids de chaque fonction est donnée à travers le vecteur β de taille $p \times 1$ appelé paramètre de régression.

La fonction de covariance $k(s, \tilde{s})$ est un **noyau défini positif**, c'est à dire pour tout $(a_i)_{i=1,\ldots,n} \in$

$\mathbb{R}$, $n \in \mathbb{N}^*$ et les différentes $(s_i)_{i=1,\dots,n} \in \mathcal{D}$, elle satisfait la propriété suivante :

$$\sum_{i,j=1}^{n} a_i a_j k(s_i, s_j) \geq 0,$$

et $\sum_{i,j=1}^{n} a_i a_j k(s_i, s_j) = 0$ si et seulement si $a_i = 0$ pour tout $i = 1, \dots, n$.

Le noyau de covariance décrit la structure de dépendance du processus gaussien $Z(s)$. Dans le cadre d'une régression, il est souvent considéré de la forme $k(s, \tilde{s}) = \sigma^2 r(s, \tilde{s}, \phi)$ où $r(s, \tilde{s}, \phi)$ est une fonction de corrélation, dépendante de deux paramètres ϕ et σ^2. Cette paramétrisation du noyau est inspirée par la structure de dépendance rencontrée dans la nature. Généralement cette dépendance décroît de manière plus ou moins rapide en fonction de la distance. Par exemple ϕ représente la plus grande distance à partir de laquelle on considère que les variables sont indépendantes et σ^2 est la variance fixe du processus.

Le noyau de covariance est certainement l'élément le plus important d'une régression de processus gaussien. En effet il contrôle le lissage du processus gaussien et donc la régularité de l'approximation de la fonction objective $z(s)$ [77].

2.3 Exemples de processus gaussiens

2.3.1 Mouvement Brownien

Soit $\mathcal{D} = \mathbb{R}_+$, le mouvement brownien standard $(W(s), s \geq 0)$ est un processus gaussien défini par sa moyenne $\mu(s) = \mathbb{E}[W(s)] = 0$ et sa covariance $\text{Cov}(W(s), W(\tilde{s})) = k(s, \tilde{s}) = \min(s, \tilde{s})$. On l'appelle aussi processus de Wiener.

2.3.2 Processus d'Ornstein-Uhlenbeck

Soit $\mathcal{D} = \mathbb{R}$, le processus d'Ornstein-Uhlenbeck est un processus gaussien centré défini par :

$$U(s) = \exp(-s/2) W(\exp(s))$$

où W est un mouvement Brownien. On montre facilement que $U(s) \sim \mathcal{N}(0; 1)$ car $Var(U(s)) = 1$. Sa fonction de covariance est donnée par :

$$k(s, \tilde{s}) = \exp(-|s - \tilde{s}|)/2$$

Elle ne dépend que de la différence $(s - \tilde{s})$, il s'agit bien d'un processus stationnaire de second d'ordre de fonction de covariance donnée par $C(h) = \exp(-|h|/2)$.

2.4 Interpolation spatiale

L'interpolation spatiale est une technique qui permet d'estimer la valeur d'un phénomène donné en des sites non échantillonnés d'un domaine $(\mathcal{D} \subset \mathbb{R}^d, d = (1, 2, 3))$, à partir des données recueillies aux sites d'observation, $(s_1, \ldots, s_n)$, avec $s_i \in \mathcal{D}$ et n étant le nombre de sites d'observations échantillonnés. Il existe deux manières de faire de l'interpolation spatiale :

- Méthode déterministe :

  Dans ce type d'interpolation la variable régionalisée n'est pas aléatoire. Une présentation détaillée sur ces techniques peut être dans [6]. Nous citons également, les méthodes barycentriques ([5], p.67), les méthodes d'interpolation par partitionnement de l'espace [115] et les splines ([27, p. 181], [130, p. 30]) ...
- Méthode stochastique :

  Dans cette technique le concept du hasard est incorporé dans l'étude du phénomène régionalisé. En effet contrairement à la vision déterministe, les méthodes stochastiques proposent toutes un modèle probabiliste. Pour formaliser le comportement du phénomène régionalisé, elles incluent un ou plusieurs termes d'erreurs aléatoires. Grâce à cette modélisation, des erreurs de prédictions peuvent être calculées. Parmi les méthodes stochastiques, nous citons entre autres les techniques de régression classique [73], de régression locale [40, 132] et de krigeage [4, 23, 43, 66, 74, 77, 86, 87, 91, 92, 94, 122, 126, 131].

Afin de tenir compte de la structure aléatoire et spatiale des données, nous allons nous intéressé, dans cette thèse, aux techniques du krigeage. Pour plus d'informations sur les autres techniques de régression classique et locale voir [6].

2.4.1 Le krigeage

Le krigeage doit son nom au chercheur français Matheron [86, 87] qui en plus d'avoir posé le formalisme de cette méthode d'interpolation spatiale, en a assuré son développement au centre de géostatistique de l'école des mines de Paris. Matheron s'est sans doute inspiré des travaux de l'ingénieur minier Krige [66] qui a été le précurseur de cette méthode. D'autres chercheurs, notamment le météorologue Gandin [43] se sont aussi inspirés des travaux de Krige pour établir

les fondements théoriques de cette méthode. Aujourd'hui c'est la terminologie proposée par Matheron qui est la plus adopté par la communauté de géostatistique. L'article de Cressie [23] qui parle de l'origine et l'historique de cette méthode peut en témoigner.

Le krigeage prend en compte la structure de dépendance spatiale des données, ce qui n'est pas le cas des deux autres méthodes d'interpolation spatiale stochastiques (régression classique ou locale) citées précédemment. L'un des avantages du krigeage est qu'en plus de permettre la prédiction de la variable aléatoire régionalisée en tout point, il fournit, en un coût de calcul faible, une estimation de l'erreur de prédiction de cette variable.

Le modèle de base du krigeage a le même formalisme que les modèles de régression de processus aléatoire gaussien défini précédemment. C'ést-à-dire à travers une moyenne et une fonction de covariance paramétriques. Il s'énonce comme suit :

$$Z(s) = \mu(s) + \delta(s) + \epsilon(s), \quad s \in \mathcal{D}, \tag{2.8}$$

où :

- La fonction $\mu(.) = (m(.))'\beta$, est la composante déterministe de $Z(.)$ ou sa tendance spatiale ; où $m(.)$ est une fonction connue à valeurs dans $\mathbb{R}^p$ et β est un vecteur à p coefficients inconnus à estimer. La structure de la fonction $m(.)$ est déterminée par notre connaissance *a priori* du phénomène à modéliser. C'est en général un vecteur de fonctions appartenant à une sous famille d'une base de l'espace fonctionnel (base de Fourier, Tchebychev, Lagrange). Pour des phénomènes ayant un structure homogène des auteurs comme [111] préconisent l'utilisation d'une moyenne constante $m(s) = 1$. Ils ont observé en pratique que cela donne un prédicteur satisfaisant et simplifie les calculs
- Le processus $(\delta(s), s \in \mathcal{D})$ est un processus gaussien faiblement stationnaire de moyenne nulle, de variance σ^2 et de fonction de corrélation $r(s, \tilde{s}, \phi)$ où ϕ est le paramètre de corrélation. Si le processus $\delta(s)$ est supposé être isotrope [1] alors la fonction de corrélation s'écrit sous la forme $r(\eta, \phi)$ où $\eta = \|s - \tilde{s}\|_2$ est la distance euclidienne entre s et $\tilde{s}$. Les paramètres σ^2, ϕ sont généralement inconnus et seront à estimer.
- $(\epsilon(s), s \in \mathcal{D})$ est un bruit blanc indépendant de $\delta(s)$. Il a pour variance $\text{Var}(\epsilon(s)) = \tau^2$ et représente les erreurs de mesure (tolérance des appareils...). Elle est généralement inconnue et devra être estimée.

Il est à préciser aussi qu'il existe différents types de krigeage, selon le choix et la spécification de

1. c'est-à-dire que sa structure de dépendance ne dépend pas de la direction (angle) mais seulement de la distance entre deux points

la tendance $\mu(.)$. Les plus classiques sont le krigeage simple, ordinaire ou universel. En effet on parle de

- **Krigeage simple** si on considère que $\mu(s) = \mu$ sur tout le domaine. Dans ce cas la constante μ est supposée connue. C'est typiquement le cas quand il s'agit de grandeurs physiques connus en moyenne et homogènes spatialement.
- **Krigeage ordinaire** si on prend $\mu(s) = \mu$ mais cette fois-ci la moyenne μ est inconnue et est constante localement. Ce modèle est largement utilisé en climatologie.
- **Krigeage universel** si on suppose que $\mu(s) = m(s)'\beta = \sum_{i=1}^{p} m_i(s)\beta_i$ est une combinaison linéaire de fonctions bases connues dépendant de la position $s \in \mathcal{D}$. Ce modèle plus général est adapté aux phénomènes présentant une certaine irrégularité ou de comportement extrême [27].

Il est à noter que la stationnarité intrinsèque n'impose pas l'existence de la variance du processus aléatoire $\delta(.)$, ni de supposer que l'espérance soit nulle mais seulement constante, contrairement à une stationnarité du second ordre. Dans le cas d'une stationnarité intrinsèque, le semi-variogramme $\gamma(.)$, qui est égale à la moitié du variogramme, sera utilisé dans la partie analyse variographique. Il s'agit d'une étape préalable du krigeage pour estimer la structure de dépendance spatiale du processus $\delta(.)$. En effet, même si le modèle (2.8) sur lequel se base le krigeage suppose sa connaissance, la structure de dépendance n'est pas généralement connue en pratique.Pour plus d'informations sur l'analyse variographique et le variogramme voir [6, 19, 27].

Dans la suite de ce rapport, nous ne considérons que des semi-variogrammes isotropes, c'est à dire qui ne dépendent que de la norme $\eta = \|h\|$ du vecteur de translation h entre deux points s, $s + h$. Pour le cas anisotrope c'est-à-dire le cas où la direction du vecteur de translation h est aussi prise en compte se référer par exemple à [49].

Pour la mise en œuvre du krigeage l'analyse variographique permet de deviner une paramétrisation de la structure de dépendance du processus $\delta(.)$. En effet elle permet de trouver d'une part la paramétrisation $k(s, \tilde{s}) = \sigma^2 r(\eta, \phi)$ et d'autre part la variance du bruit τ^2. Dans la terminologie du krigeage les paramètres τ^2 désigne l'effet de pépite, $\sigma^2 + \tau^2$ le palier et ϕ la portée et .

- **L'Effet de pépite** : Il capte les variation à très courte échelle, il est égale à $\lim_{h \to 0} \gamma(h)$. Dans le cadre du krigeage avec erreur de mesures l'effet de pépite sera associé au bruit blanc $\epsilon(.)$ donné dans (2.8).

– **Le Palier** : C'est la variance du processus aléatoire $(Z(s), s \in \mathcal{D})$ qui est fixe sous la conditions de stationnarité.

– **La Portée** : Il s'agit de la distance minimale où deux observations ne se ressemblent plus en moyenne et qui ne sont plus linéairement liées (covariance nulle). À cette distance, la valeur du variogramme correspond à la variance du processus $Z(.)$ (c'est-à-dire le palier).

Comme dit auparavant c'est dans l'analyse variographique que le semi-variogramme va être estimé à partir des données. Cressie [27, p. 69] utilise la méthode des moments pour estimer le semi-variogramme qui est dit expérimental ou empirique :

$$\widehat{\gamma}(\eta) = \frac{1}{2|N(\eta)|} \sum_{N(\eta)} (z(s_i) - z(s_j))^2, \tag{2.9}$$

où,

– les $z(s_i)$, $1 \leq i \leq n$ sont les données disponibles, n étant le nombre de données.

– L'ensemble $N(\eta) = \{(i,j), \text{tel que} \|s_i - s_j\| = \eta\}$

– La quantité $|N(\eta)|$ est le nombre de paires distincts de l'ensemble $N(\eta)$

Pour plus de détails sur le semi-variogramme empirique voir [5, 6, 27, 88].

D'après certains auteurs le semi-variogramme empirique présente des inconvénients, notamment Matheron, [88, p. 155] qui a fait remarqué que l'estimation du semi-variogramme empirique est biaisée. D'autres [6],[5, p. 126] disent qu'elle n'est pas fiable pour les grandes distances. Donc vu ces inconvénients évoqués ci-dessus pour le semi-variogramme empirique. Ce serait plus simple en pratique d'utiliser les modèles paramétriques du semi-variogramme. Les modèles les plus standards sont :

– Modéle Exponentiel d'effet de pépite c_0, de palier $c_0 + c$, de portée a et de portée pratique égale à $3a$:

$$\gamma(\eta) = c_0 + c(1 - \exp(-\frac{\eta}{a})) \qquad \forall \eta > 0$$

Ce modèle est illustré dans la figure suivante :

– Modéle Gaussien d'effet de pépite c_0, de palier $c_0 + c$, de portée a et de portée pratique

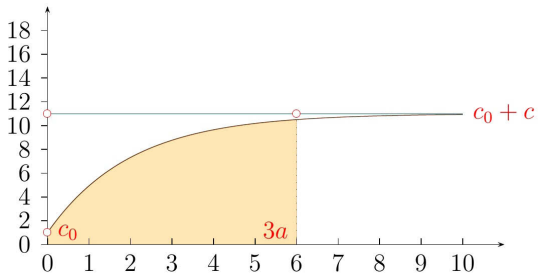

FIGURE 2.1 – variogramme exponentiel, $a = 2$, $c + c_0 = 11$, $c_0 = 1$.

égale à $3\sqrt{a}$:

$$\gamma(\eta) = c_0 + c(1 - \exp(-\frac{\eta^2}{a^2})) \qquad \forall \eta > 0$$

– Modéle Matérn [85]

$$\gamma(\eta) = c_0 + c(\left\{2^{\kappa-1}\Gamma(\kappa)\right\}^{-1} (\frac{\eta}{a})^{\kappa} K_{\kappa}(\frac{\eta}{a})) \qquad \forall \eta > 0$$

où $K_{\kappa}(.)$ désigne la fonction de Bessel modifiée d'ordre κ.

Pour plus de détails sur d'autres modèles du semi-variogramme voir [5, 6].

Une fois le modèle de semi-variogramme choisi, il faut ensuite utiliser ce dernier pour estimer le vecteur $\theta = (\beta, \sigma^2, \tau^2, \phi)$ des paramètres du krigeage. Avec τ^2 qui représente l'effet de pépite, $\sigma^2 + \tau^2$ le palier donc la variance de $Z(s)$, $\forall s \in \mathcal{D}$, et ϕ la portée. L'estimation du vecteur θ peut se faire par maximum de vraisemblance (voir par exemple [2, 98]). En effet, le vecteur des observations $Z_n = (Z(s_1), \ldots, Z(s_n))$ suit la loi gaussienne suivante :

$$[Z_n \mid \beta, \sigma^2, \phi, \tau^2] \sim \mathcal{N}\left(M\beta; \sum_{(\sigma^2,\phi,\tau^2)}\right), \tag{2.10}$$

où,

– $M = (m(s_1), \ldots, m(s_n))'$

– $\Sigma_{(\sigma^2,\phi,\tau^2)}$ est la matrice de variance covariance obtenue à partir du modèle du variogramme choisi.

Pour une écriture plus lisible, la matrice de variance-covariance peut être formulée comme $\Sigma_{(\sigma^2,\phi,\tau^2)} = \sigma^2 R(\phi) + \tau^2 I_n$, avec $R(\phi) = (r(s_i, s_j, \phi))_{1\leq i,j\leq n}$ et I_n est la matrice identité de taille n. Dans la pratique, il peut être préférable de travailler avec un effet de pépite relatif $\tau^2_{\mathcal{R}} = \dfrac{\tau^2}{\sigma^2}$, dans ce cas la matrice de variance covariance s'écrit $\Sigma_{(\sigma^2,\phi,\tau^2)} = \sigma^2(R(\phi) + \tau^2_{\mathcal{R}} I_n)$. L'effet de pépite relatif $\tau^2_{\mathcal{R}}$ est l'inverse de ce qu'on appelle rapport signal sur bruit (SNR)

Pour un processus gaussien, la **vraisemblance** correspondante aux observations Z_n s'écrit donc de la manière suivante :

$$f(Z_n = z_n \mid \theta) = \frac{1}{(2\pi)^{n/2}|\Sigma_{(\sigma\cdot\phi,\tau^2)}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(z_n - M\beta)'(\Sigma_{(\sigma\cdot\phi,\tau^2)})^{-1}(z_n - M\beta)\right\}, \qquad (2.11)$$

où $|.|$ représente le déterminant d'une matrice.

Pour simplifier les calculs il est préférable d'appliquer le logarithme qui est une fonction croissante sur la fonction de vraisemblance (2.11), ce qui donnera la fonction **log-vraisemblance** qui de manière équivalente permettra de calculer le maximum de vraisemblance $\widehat{\theta}$ correspondant aux observations Z_n.

$$\begin{aligned} l(\theta \mid Z_n = z_n) &= \log(f(z_n \mid \theta)) \\ &\propto -\frac{1}{2}\Big\{n\log(\sigma^2) + \log(|R(\phi) + \tau^2_{\mathcal{R}} I_n|) \\ &\quad + (z_n - M\beta)'(\sigma^2(R(\phi) + \tau^2_{\mathcal{R}}\mathcal{I}))^{-1}(z_n - M\beta)\Big\}. \end{aligned} \qquad (2.12)$$

Le maximum du log-vraisemblance (2.12) est atteint en un point $\theta = (\sigma^2, \phi, \beta, \tau^2)$ où les dérivées partielles sont nulles. Donc après avoir calculé les dérivées partielles pour β et σ^2, en cherchant leurs zéros, on obtient :

$$\widehat{\beta} = (M'(R(\phi) + \tau^2_{\mathcal{R}} I_n)^{-1} M)^{-1}(M'(R(\phi) + \tau^2_{\mathcal{R}} I_n)^{-1} Z_n), \qquad (2.13)$$

et

$$\widehat{\sigma^2} = \frac{1}{n}(Z_n - M\widehat{\beta})'(R(\phi) + \tau^2_{\mathcal{R}} I_n)^{-1}(Z_n - M\widehat{\beta}). \qquad (2.14)$$

les quantités $\widehat{\beta}$ et $\widehat{\sigma^2}$ ne pourront être estimés exactement qu'après avoir estimé $\widehat{\phi}$ et $\widehat{\tau^2_{\mathcal{R}}}$ car ils dépendent de ces deux paramètres.

Pour atteindre cet objectif et estimer $\widehat{\phi}$ et $\widehat{\tau^2_{\mathcal{R}}}$ nous remplaçons dans (2.12), β et σ^2 par les formules données respectivement dans (2.13) et (2.14). Ce qui donne la fonction de log-vraisemblance profilée suivante :

$$l(\widehat{\beta}, \widehat{\sigma^2}, \phi, \tau^2_{\mathcal{R}}) \propto -\frac{1}{2}\left\{n\log(\widehat{\sigma^2}) + \log(|R(\phi) + \tau^2_{\mathcal{R}} I_n|) + n\right\}, \tag{2.15}$$

qui ne dépendent que de ϕ et $\tau^2_{\mathcal{R}}$. L'estimation ϕ et $\tau^2_{\mathcal{R}}$ s'obtient de la manière suivante :

$$(\widehat{\phi}, \widehat{\tau^2_{\mathcal{R}}}) = \underset{\phi, \tau^2_{\mathcal{R}}}{\operatorname{argmax}} \left[-\frac{1}{2}\left\{n\log(\widehat{\sigma^2}) + \log(\det(R(\phi) + \tau^2_{\mathcal{R}} I_n))\right\}\right]. \tag{2.16}$$

Une fois obtenu $\widehat{\phi}$ et $\widehat{\tau^2_{\mathcal{R}}}$, ils permettent d'estimer les autres paramètres en remplaçant ϕ et $\tau^2_{\mathcal{R}}$ par leur estimation dans (2.13) et (2.14).

Une fois les paramètres du modèle estimés, il sera donc possible de prédire la variable régionalisée $Z(s)$ en tout point s du domaine $\mathcal{D}$ en utilisant le modèle 2.8.

Il est important de préciser que dans le cas où les paramètres du modèle sont connus ϕ, σ^2, β et τ^2, le krigeage est le meilleur prédicteur linéaire, non biaisé communément appelé en anglais **BLUP**(*Best linear Unbiased Predictor*). Le **BLUP** en un point non échantillonné $s_0 \in \mathcal{D}$ est :

$$\widehat{Z}(s_0) = m(s_0)'\beta + r_0(\phi)(R(\phi) + \tau^2_{\mathcal{R}} I_n)^{-1}(Z_n - M\beta), \tag{2.17}$$

avec $r_0(\phi) = (r(s_0 - s_i, \phi))_{1 \leq i \leq n}$.

Dans la pratique ces paramètres du modèle de krigeage sont inconnus et sont estimés comme précédemment. Dans ce cas le nom de **EBLUP**(Empirical Best linear Unbiased Predictor) est attribué au krigeage. Le **EBLUP** en un point non échantillonné $s_0 \in \mathcal{D}$ est :

$$\widehat{Z}(s_0) = m(s_0)'\widehat{\beta} + r_0(\widehat{\phi})(R(\widehat{\phi}) + \widehat{\tau^2}_{\mathcal{R}} I_n)^{-1}(Z_n - M\widehat{\beta}). \tag{2.18}$$

Il est aussi important de noter que si on tient compte des erreurs de mesure le krigeage n'est pas en soi un interpolateur du fait qu'il ne restitue pas les mêmes valeurs mesurées aux points d'observation s_i, $1 \leq i \leq n$. Par contre si on omet les erreurs de mesure, on restitue les valeurs observées aux sites d'observation ; dans ce cas le krigeage est considéré comme méthode d'interpolation.

Avec le krigeage on a l'estimation de la variable régionalisé $Z(s), \forall\, s \in \mathcal{D}$, de même que l'erreur commise correspondante. Par contre, l'un des inconvénients du krigeage est que l'incertitude dû à l'estimation des paramètres ϕ, τ^2, σ^2, β n'est pas prise en compte dans la prédiction de la variable régionalisée aux points non observés. D'où l'intérêt d'adopter une approche bayésienne pour remédier à cela.

2.4.2 Le krigeage Bayésien

L'inférence bayésienne permet, en quelque sorte, d'intégrer dans un modèle statistique, les connaissances qui existent sur le phénomène étudié ou bien sur le domaine spatiale où il doit être observé. Ces connaissances peuvent être les avis d'experts, des historiques existantes sur le comportement ou la distribution spatiale du phénomène ; par exemple la topographie du domaine... Évidemment ces connaissances ne décrivent pas parfaitement les fluctuations spatiales de la variable régionalisée. Elles sont même le plus souvent entachées de beaucoup d'incertitudes et d'erreurs. Pour réduire ces incertitudes, il est donc nécessaire de prélever la variable régionalisée en certains sites et de combiner ces observations aux connaissances *a priori* sur le phénomène et sur le domaine d'étude. Ceci permettra de mieux décrire le comportement spatial de la variable régionalisée. En résumé, l'approche bayésienne permet d'intégrer à l'interpolation spatiale de l'information *a priori* concernant la tendance de la variable régionalisée. Elle permet aussi la quantification de l'incertitude sur les paramètres du modèle, notamment à travers l'expertise (c'est-à-dire de leurs lois *a priori*) et des données recueillies. Au lieu d'avoir des estimations fixes, on aura donc des lois de distribution pour les paramètres du modèle et une distribution prédictive de la variable régionalisée. Pour plus de détails sur l'approche bayésienne se référer à [8, 45, 46, 108]. Ce qui a été dit auparavant pour l'approche bayésienne a été appliqué pour le krigeage et l'interpolation spatiale [34, 35, 53, 77, 100], d'où le nom du krigeage Bayésien. Avec cette approche les paramètres ϕ, β, σ^2 et τ^2 du modèle (2.8) sont eux aussi considérés comme des variables aléatoires. Donc la méthodologie bayesienne consite à leurs affecter des lois de probabilités ; elles sont communément appelées lois *a priori*. Le choix de ces lois n'est cependant pas évident et doit être en adéquation avec le phénomène à étudier.

2.4.2.1 Choix de l'information *a priori*

Dans cette partie, nous allons parler du choix de l'information *a priori* pour le krigeage Bayésien du fait de sa dépendance du phénomène à étudier et de ses conséquences sur le résultat attendu. En effet, dans l'approche bayesienne, l'information *a priori* est l'idée que l'on se fait

de la distribution des paramètres d'un modèle statistique avant toute observation de la variable régionalisée. Ce choix de l'information *a priori* n'est pas du tout évident et a un impact sur le résultat de l'approche bayésienne : c'est-à-dire sur la loi de prédiction de la variable régionalisée. Dans le cas où peu de données sont disponibles, les résultats peuvent être très dépendants de la nature des lois préalablement choisies. Donc pour l'approche bayésienne, la distribution *a priori* des paramètres d'un modèle a toujours été et continue d'être un vrai sujet de controverse ou de paradoxe. Dans la littérature, il existe différents types de lois *a priori* que l'on peut ranger dans deux catégories principales : les lois informatives et non informatives.

- Les lois *non informatives*, comme leur nom l'indique, sont privilégiées quand aucune connaissance n'est disponible sur le phénomène étudié. En effet, l'utilisation des *a priori* non informatifs permet de minimiser l'impact d'une éventuelle erreur du choix des paramètres sur l'inférence bayésienne. Dans ce cas, L'*a priori* de Jeffreys, exprimé en fonction de la matrice d'information de Fisher, est généralement le plus utilisé par la majorité des statisticiens [59, 60]. Dans le cadre de la régression linéaire multiple, la loi de Zellner, appelée aussi **"g-prior"**, est utilisé comme *a priori* sur les paramètre de la régression : il s'agit d'une loi gaussienne dont la matrice de covariance est égale à l'inverse de la matrice d'information de Fisher [136].
- Les lois *informatives* sont plus généralement utilisées dans le cas où il existe une connaissance même partielle du phénomène. Les lois conjuguées qui sont une classe générique de la famille exponentielle [9, 82, 82, 107] sont en général les plus utilisées. Ils sont caractérisés par deux quantités nommés hyperparamètres $\lambda \geq 0$ et η :

$$\pi(\theta \mid \lambda, \eta) \propto \exp(\theta.\eta - \lambda\psi(\theta)).$$

  Selon le choix des hyper-paramètres les *a priori* conjuguées peuvent avoir un fort impact sur l'inférence bayésienne. En effet, les profils et les valeurs de ces hyper-paramètres, peuvent contribuer à exclure ou à donner plus d'importance à des situations plutôt que d'autres. Il est à noter, également, qu'il est possible, à partir d'une classe de lois conjuguées, d'approcher l'*a priori* de Jeffreys (voir [16, 36]).

En ce qui concerne le modèle de krigeage (2.8), nous considérons en premier lieu le cas où il n'y a pas d'effet de pépite, c'est-à-dire le cas où $\tau^2 = 0$. En utilisant la factorisation de Bayes,

la densité de probabilité jointe des paramètres (ϕ, σ^2, β) peut être écrite comme suit :

$$\pi(\phi, \sigma^2, \beta) = \pi(\beta \mid \phi, \sigma^2)\pi(\sigma^2 \mid \phi)\pi(\phi). \tag{2.19}$$

Le choix des lois *a priori* non informatives les plus utilisées pour ces paramètres sont les suivant :

- Pour le paramètre de tendance β, un *a priori* plat **"flat prior"** est communément utilisé :

$$\pi(\beta) \propto 1. \tag{2.20}$$

  Une loi plate ne privilégie pas une plage de valeurs des paramètres mais leurs attribue une probabilité uniforme ou un poids égale.
- Pour $\sigma^2 = \mathrm{Var}(\delta(s))$ on prend une loi de la forme :

$$\pi(\sigma^2) \propto \frac{1}{\sigma^2},$$

  elle est communément appelé *a priori* réciproque **"reciprocal prior"**. Cette distribution donne plus d'importance aux petites variances et néglige les grandes variations.
- Pour le paramètre de corrélation ϕ on peut aussi choisir un *a priori* réciproque

$$\pi(\phi) \propto \frac{1}{\phi}.$$

Les lois citées ci-dessus sont souvent utilisées pour leur simplicité. Dans la litterature, d'autre lois *a priori* non informatives sont disponibles, pour plus de détails, voir par exemple [108].

En ce qui concerne les *a priori* informatifs, l'une des familles la plus utilisée pour le krigeage Bayesien est celle des lois conjuguées. Il s'agit de lois de probabilité stables par une certaine transformation linéaire, quadratique... par exemple la loi normale, la loi gamma, la loi inverse gamma... Parmi les *a priori* informatifs communément utilisée pour le krigeage nous énumérons :

$$\begin{aligned}
[\phi] &\sim \Gamma(a, b) \\
\left[\sigma^2 \mid \phi\right] &\sim \mathcal{X}^2_{ScI}(\nu, S^2) \\
\left[\beta \mid \sigma^2, \phi\right] &\sim \mathcal{N}(e, \sigma^2 V)
\end{aligned}$$

- Pour le paramètre ϕ on prend souvent la loi gamma $\Gamma(a, b)$ d'hyper-paramètres de forme a et d'échelle b.
- La loi conditionnelle de σ^2 sachant ϕ est la distribution inverse-$\mathcal{X}^2$ ayant ν degrés de liberté et de paramètre d'échelle S^2. Le symbol ScI provient du terme anglais **"Scaled inverse chi-squared"**.
- La loi conditionelle de β sachant ϕ et σ^2 suit une loi normale multivariée $\mathcal{N}(e, \sigma^2 V)$ avec e représente l'hyper-paramètre moyen et V représente l'hyper-paramètre matrice de covariance.

Il est à noter que pour ϕ fixé, les lois $\left[\sigma^2 \mid \phi\right] \sim \mathcal{X}^2_{ScI}(\nu, S^2)$ et $\left[\beta \mid \sigma^2, \phi\right] \sim \mathcal{N}(e, \sigma^2 V)$ sont considérés comme des *a priori* conjugués.

L'un des avantages des *a priori* conjuguées est qu'ils permettent d'avoir la même famille de lois pour la distribution *a posteriori* correspondante. On note que l'impact des *a priori* conjuguées sur l'inférence bayésienne peut être atténué en considérant leurs hyper-paramètres comme des variables aléatoires aussi, les lois sur les hyperparamétres sont communément appelés **hyperpriors**.

2.4.2.2 Distribution *a posteriori* des paramètres du modèle

La nomination "*a posteriori*" provient du fait que l'on calcule les lois des paramètres après avoir observé le phénomène étudié. En effet, comme les lois *a priori* sur $\theta = (\phi, \sigma^2, \beta)$ sont entachées d'incertitudes, l'idée est donc de diminuer au maximum ces incertitudes en y intégrant l'information apportée par le vecteur des observations $Z_n = (Z(s_1), \ldots, Z(s_n))$. Pour atteindre cet objectif, nous tirons profit du théorème de Bayes qui permet de calculer la distribution *a posteriori* en fonction des lois *a priori* et le modèle d'observation. Ici par exemple, pour θ, nous

obtenons la loi conditionnelle $[\theta \mid Z_n]$:

$$\pi(\theta \mid Z_n = z_n) = \frac{p(z_n \mid \theta)\pi(\theta)}{p(z_n)}, \tag{2.21}$$

cette densité ou distribution (2.21) est communément appelé densité de loi *a posteriori* de θ. La densité $p(z_n \mid \theta)$, dite aussi modèle d'observation, elle est donnée dans notre cas par le postulat du krigeage explicité dans l'équation (2.10). La quantité $\pi(\theta)$ est la densité de distribution *a priori* des paramètres du krigeage et $p(z_n)$ est une constante de normalisation. Dans la suite nous utiliserons la notation $\pi(\theta \mid z_n)$ au lieu de $\pi(\theta \mid Z_n = z_n)$.

Une fois les distributions *a priori* des paramètres du modèle de krigeage définies et les observations de la variable régionalisée aux points du plan d'expériences réalisées, il est possible de calculer les distributions *a posteriori* de ces paramètres. La forme de la distribution *a posteriori* d'un paramètre du modèle va dépendre du choix qu'on a fait sur les distributions *a priori* des paramètres. Dans le paragraphe suivant, nous donnons les distributions *a posteriori* des paramètres β, σ^2, ϕ correspondant aux distributions *a priori* choisies dans la section précédente pour ces paramètres (voir [34, 106] pour plus de détails).

La distribution *a posteriori* jointe des paramètres peut être factorisée de la manière suivante :

$$\pi(\beta, \sigma^2, \phi \mid z_n) = \pi(\beta \mid \sigma^2, \phi, z_n)\pi(\sigma^2 \mid \phi, z_n)\pi(\phi \mid z_n)$$

Nous rappelons ici que z_n est une réalisation du vecteur des observations Z_n.

Loi *a posteriori* pour β

1. **Cas d'*a priori* informatif**

   Si on choisit $[\beta \mid \sigma^2, \phi] \sim \mathcal{N}(e, \sigma^2 V)$ comme distribution *a priori* pour β, l'*a posteriori* correspondante est :

   $$\left[\beta \mid \sigma^2, \phi, z_n\right] \sim \mathcal{N}(\tilde{e}, \sigma^2 \tilde{V}),$$

   avec,

   $$\tilde{e} = \tilde{V}(V^{-1}e + M'R(\phi)^{-1}z_n) \tag{2.22}$$

   $$\tilde{V} = (V^{-1} + M'R(\phi)^{-1}M)^{-1}, \tag{2.23}$$

où,

- $M = (m(s_1), \ldots, m(s_n))'$
- $R(\phi) = r(s_i - s_j, \phi_t)_{1 \leq i,j \leq n}$, $r(.,.)$ étant la fonction de corrélation.

Dans ce qui suit pour simplifier les expressions on écrira tout simplement R en pensant à $R(\phi)$.

2. **Cas d'*a priori* non informatif**

   Si on choisit $\pi(\beta \mid \sigma^2, \phi) \propto 1$ comme *a priori*, la distribution *a posteriori* correspondante est :

$$\left[\beta \mid \sigma^2, \phi, z_n\right] \sim \mathcal{N}(\widehat{e}, \sigma^2 \widehat{V}),$$

   avec $\widehat{e}$ et $\widehat{V}$ qui peuvent être respectivement obtenu à partir de (2.22) et (2.23) en prenant le cas limite où $V^{-1} \equiv 0$. Donc on a :

$$\widehat{e} = (M' R^{-1} M)^{-1} M' R^{-1} z_n \tag{2.24}$$

$$\widehat{V} = \sigma^2 (M' R^{-1} M)^{-1}. \tag{2.25}$$

Loi *a posteriori* pour σ^2

1. **Cas d'a priori informatif**

   Si on choisit $[\sigma^2 \mid \phi] \sim \mathcal{X}^2_{ScI}(\nu, S^2)$ et $[\beta \mid \sigma^2, \phi] \sim \mathcal{N}(e, \sigma^2 V)$ comme *a priori*, la distribution *a posteriori* correspondante pour σ^2 est :

$$\left[\sigma^2 \mid \phi, \boldsymbol{z}\right] \sim \mathcal{X}^2_{ScI}(\tilde{\nu}, \tilde{S}^2)$$

   avec $\tilde{\nu} = \nu + n$ et

$$\tilde{S}^2 = \frac{\nu \tilde{V} + e' V^{-1} e + z_n' R^{-1} z_n - \tilde{e}' \tilde{V}^{-1} \tilde{e}}{\nu + n}. \tag{2.26}$$

2. **Cas d'*a priori* non informatif**

   Si on choisit $\pi(\sigma^2 \mid \phi) \propto \frac{1}{\sigma^2}$ et $\pi(\beta \mid \sigma^2, \phi) \propto 1$ comme *a priori*, la distribution *a posteriori* correspondante pour σ^2 est :

$$\left[\sigma^2 \mid \phi, z_n\right] \sim \mathcal{X}^2_{ScI}(n - p, \Lambda^2),$$

où p est la taille du vecteur β et

$$\Lambda^2 = \frac{1}{n-p}(z_n - M\widehat{e})' R^{-1}(z_n - M\widehat{e}). \tag{2.27}$$

Loi *a posteriori* pour ϕ

Pour le paramètre de corrélation ϕ, la densité de la distribution *a posteriori* se calcule comme suit :

$$\pi(\phi \mid z_n) \propto \frac{\pi(\beta, \sigma^2, \phi)\pi(z_n \mid \beta, \sigma^2, \phi)}{\pi(\beta \mid z_n, \sigma^2, \phi)\pi(\sigma^2 \mid z_n, \phi)}. \tag{2.28}$$

1. **Cas d'a priori informatif**

   Dans le cas où l'on a $\pi(\sigma^2) \sim \mathcal{X}^2_{ScI}(\nu, S^2)$ et $\pi(\beta) \sim \mathcal{N}(e, V)$, la distribution *a posteriori* correspondante pour ϕ est :

$$\pi(\phi \mid z_n) \propto \pi(\phi)|\tilde{V}|^{\frac{1}{2}}|R|^{-\frac{1}{2}}(\tilde{S}^2)^{-\frac{n+\nu}{2}}, \tag{2.29}$$

2. **Cas d'*a priori* non informatif**

   Dans le cas où l'on a $\pi(\beta \mid \sigma^2, \phi) \propto 1$ et $\pi(\sigma^2 \mid \phi) \propto \frac{1}{\sigma^2}$ comme *a priori*, la distribution *a posteriori* correspondante pour ϕ est :

$$\pi(\phi \mid z_n) \propto \pi(\phi)|\widehat{V}|^{\frac{1}{2}}|R|^{-\frac{1}{2}}(\Lambda^2)^{-\frac{n-p}{2}}. \tag{2.30}$$

   On note que dans les deux cas, on choisit en pratique pour $\pi(\phi)$ une distribution discrète dans un intervalle donné qui est pensé être une portée raisonnable.

   Dans le cas où l'on tient compte de l'effet de pépite, i.e. $\tau^2 > 0$, en général dans la pratique on utilise une distribution *a priori* jointe discrète pour ϕ et υ^2 qui est l'effet de pépite relatif. En effet dans la pratique au lieu d'utiliser l'effet de pépite absolu τ^2, il est beaucoup plus fréquent d'utiliser l'effet de pépite relatif $\upsilon^2 = \frac{\tau^2}{\sigma^2}$. L'effet de pépite est introduit en remplaçant tout simplement R par $(R + \upsilon^2 I_n)$, où I_n est la matrice identité.

2.4.2.3 Distribution prédictive de la variable régionalisée

Donc une fois la distribution *a posteriori* des paramètres du modèle de krigeage obtenue, on peut donc calculer la distribution prédictive de la variable régionalisé en tout point $s_0 \in \mathcal{D}$ non

échantillonné. Ceci en utilisant l'égalité suivante :

$$p(z(s_0) \mid z_n) = \int p(z(s_0) \mid z_n, \theta)\pi(\theta \mid z_n)d(\theta). \tag{2.31}$$

En pratique cette distribution est compliquée à calculer , donc l'idée est d'utiliser des techniques d'approximation tel que les *MCMC(Markov Chain Monte Carlo)* pour l' approcher (voir [84]). Donc avec les *MCMC* la distribution prédictive va être approchée en utilisant l'expression suivante [14] :

$$p(z(s_0) \mid z_n) \approx \frac{1}{k}\sum_{i=1}^{k} p(z(s_0) \mid z_n, \theta_i), \tag{2.32}$$

où θ_i, $i = 1 \dots k$ est un échantillon obtenu à partir des chaînes de Markov générée et suivent approximativement la loi $\pi(\theta \mid z_n)$.

l'équation (2.31) peut être réécrite en ces deux formes suivantes :

$$\begin{aligned} p(z(s_0)|z_n) &= \int\int\int p(z(s_0),\ \beta,\ \sigma^2|z_n,\ \phi)d\beta d\sigma^2\pi(\phi|z_n)d\phi \\ &= \int p(z(s_0)|z_n,\ \phi)\pi(\phi|z_n)d\phi, \end{aligned} \tag{2.33}$$

Avec $p(z(s_0)|z_n, \phi)$ qui est la densité de loi de Student multivariée suivante :

$$[Z(s_0)|z_n, \phi] \sim t_\nu(\mu(s_0),\ Q\Sigma], \tag{2.34}$$

donc on a :

$$\mu(s_0) = E[Z(s_0)|z_n, \phi] \quad \text{et} \quad \text{Var}[Z(s_0)|z_n, \phi] = \frac{\nu + n}{\nu + n - 2}Q\Sigma,$$

avec $\mu(s_0)$ et $Q\Sigma$ qui dépendent du choix de la distribution *a priori*.

Alors pour les *a priori* choisis ci-dessus on peut obtenir une expression analytique de l'équation (2.31)

1. **Cas d'*a priori* informatif**

   Dans le cas où on a des *a priori* conjugués pour les paramètres du krigeage, on a :

   Q qui est donné par l'équation (2.23),

$$\begin{aligned} \mu(s_0) &= (m(s_0) - r'R^{-1}M)(V^{-1} + M'R^{-1}M)^{-1}V^{-1}e \\ &\quad + [r'R^{-1} + (m(s_0) - r'R^{-1}M)(V^{-1} + M'R^{-1}M)^{-1}M'R^{-1}]\boldsymbol{z}, \end{aligned}$$

et $\Sigma = [R^0 - r'R^{-1}r + (m(s_0) - r'R^{-1}M)(V^{-1} + M'R^{-1}M)^{-1}(m(s_0) - r'R^{-1}M)']$,

où,

– $r = r(s_0 - s_i)_{1\leq i\leq n}$, n étant le nombre de points d'observation

– $R^0 = r(s_0 - s_0, \phi)$.

La densité $\pi(\phi \mid z_n)$ est donnée par l'équation (2.29)

2. **Cas d'*a priori* non informatif**

Dans le cas où on a des *a priori* impropres pour les paramètres du krigeage, on a :

$$Q \text{ qui est donné par l'équation } (2.27),$$

$$\mu(s_0) = (m_t(s_0) - r'R^{-1}M)\widehat{e} + r'R^{-1}z_n$$

et

$$\Sigma = R_0 - r'R^{-1}r + (m(s_0) - r'R^{-1}M)(M'R^{-1}M)^{-1}(m(s_0) - r'R^{-1}M)'.$$

La densité $\pi(\phi \mid z_n)$ est donnée par l'équation (2.30).

Chapitre 3

Modèle dynamique et spatio-temporelle

3.1 Modèle dynamique

Un modèle dynamique permet de décrire le comportement en termes mathématiques d'un système évoluant dans le temps. Ce dernier est une partie de la réalité concrète qu'on trouve pertinente dans un problème d'ingénierie et qu'on choisit d'isoler par la pensée. L'intérêt d'un modèle dynamique c'est de caractériser quantitativement l'évolution, au cours du temps, de l'état de ce système. Un modèle dynamique est en général formé d'équations aux dérivées partielles qui permettent au cours du temps de voir les variations spatiales d'un phénomène donné, et ceci de manière déterministe en général.

3.1.1 Système dynamique différentiel

L'idée de modéliser un système dynamique par des équations aux dérivées partielles vient du célèbre Isaac Newton [99] en 1687. En effet il utilisait les équations différentielles pour suivre l'évolution dans le temps d'un système physique. Depuis cette modélisation différentielle s'est étendue dans beaucoup domaine comme la mécanique des fluides [20, 116], la biologie, la météo etc...

Ci-dessous trois exemples de systèmes dynamiques modélisés par des équations différentielles. Le premier permet de modéliser la propagation de la chaleur dans un milieu donné. Le second permet de modéliser la mécanique des fluides comme par exemple les problèmes de pollution des eaux souterraines et le troisième permet de modéliser la dynamique des coccinelles et des

pucerons.

Exemple 3.1.1. *Équation de la chaleur [41]*

L'équation de la chaleur est une équation aux dérivées partielles parabolique, pour décrire le phénomène physique de conduction thermique. Un modèle simple d'équation de la chaleur, en l'absence d'un phénomène convection (transport de chaleur) s'écrit :

$$\begin{cases} \frac{\partial T}{\partial t}(x,t) - \ D\,\Delta T(x,t) & = 0, \quad \forall\, x \in \Omega \\ T(x,0) & = \ T_0(x), \quad \forall x \in \Omega \\ T(x,t) & = \ 0 \quad ,\forall\, x \in \partial\Omega \end{cases} \tag{3.1}$$

où Δ est l'opérateur Laplacien et D est le coefficient de diffusivité thermique, Ω un domaine de $\mathbb{R}^d$, $d = 1, 2, 3$, $\partial\Omega$ la frontière de Ω et $T(x,t)$ un champ de température sur ce domaine. La première équation de (3.1) représente le phénomène physique de conduction thermique. Dans le cas où il existe une source thermique dans le domaine (par exemple, une source radioactive qui serait placée à l'intérieur), on aurait :

$$\frac{\partial T}{\partial t}(x,t) - \ D\,\Delta T(x,t) \ = \mathcal{S}, \quad \forall\, x \in \Omega$$

où $\mathcal{S}$ est la source de chaleur.

La deuxième équation de (3.1) représente une condition initiale du système dynamique. La troisième équation représente une condition aux limites de Dirichlet sur le sur le bord du domaine. Cette troisième équation pourrait être remplacée par une condition de Neuman :

$$\forall\, x \in \partial\Omega\,, \quad \frac{\partial T(x,t)}{\partial n} \ = \ \vec{n}(x) \cdot \vec{\nabla} T(x,t) \ = \ 0,$$

où $\vec{n}(x)$ est le vecteur normal unitaire au point x.

Exemple 3.1.2. *Un phénomène de convection et diffusion [116]*

Soit Ω un domaine carré inclus dans R^2 et $(0,T)$ un intervalle de temps. Pour $\mathcal{S} = \mathcal{S}(x,y,t) \in L^2(\Omega \times (0,T))$, nous considérons l'équation de convection diffusion avec les deux conditions au bord de Dirichlet homogène.

$$\begin{cases} L(u) & := & \frac{\partial u}{\partial t} - \nu\Delta u + a_1\frac{\partial u}{\partial x} + a_2\frac{\partial u}{\partial y} = \mathcal{S}, \, in \; \Omega \times (0,T] \\ u(x,y,t) & = & 0 \; on \; \partial\Omega \times (0,T] \\ u(x,y,0) & = & u_0(x,y) \; in \; \Omega \end{cases} \tag{3.2}$$

a_1 et a_2 sont les coefficients de convection , ν est le coefficient de diffusion et $\mathcal{S}$ représente le terme source. a_1, a_2 sont donnés constants et $\nu > 0$. Ce type de modèle est utilisé dans beaucoup domaine d'application de la mécanique fluides par exemple dans les problémes de pollution des eaux souterraines ou aussi lors de la modélisation de l'écoulement du pétrole d'un réservoir donné. Avec ce modèle on peut déterminer l'état u du système dans le domaine Ω à chaque temps t donné.

Exemple 3.1.3. *Dynamique de coccinelles et des pucerons (voir [12])*
Les pucerons sont des insectes ravageurs permanents et redoutables pour les cultures de rosiers. Les coccinelles sont utilisées dans cette lutte biologique car elles se nourrissent de pucerons avec une grande voracité.

Soit le modèle d'état suivant qui exprime le bilan du nombre de pucerons et de coccinelles dans le temps :

$$\begin{cases} \dfrac{\partial x_1(t)(t)}{\partial t} = ax_1(t) - bx_1x_2(t) - cu(t)x_1(t) \\ \\ \dfrac{\partial x_2(t)}{\partial t} = dx_1(t)x_2(t) - ex_2(t) - fu(t)x_2(t). \end{cases} \tag{3.3}$$

avec $x_1(t)$ qui représente l'évolution du nombre pucerons et $x_2(t)$ est l'évolution du nombre de coccinelles et $u(t)$ est un taux épandage variable. Les termes a, b, c, d, e, f sont des constantes positives. le modéle d'état (3.3) a été construit sous les hypothèses suivantes :

- *en l'absence de coccinelles, la population de pucerons dispose d'assez de nourritures pour avoir une croissance exponentielle avec un taux spécifique de croissance constant ;*
- *les coccinelles dévorent d'autant plus de pucerons qu'ils sont nombreux ;*
- *la prédation par les coccinelles est la seule source de mortalité naturelle des pucerons ;*
- *les coccinelles ont un taux spécifique constant de mortalité naturelle ;*

Donc en résumé la modélisation dynamique différentielle permet de suivre l'évolution dans le temps d'un phénomène donné. Ceci de manière causale c'est à dire que son avenir ne dépend que de son passé ou du présent, et de manière déterministe c'est à dire qu'à une condition initiale donnée à l'instant présent va correspondre à chaque instant ultérieur un et un seul état futur

possible.

Les modèles dynamiques définis auparavant sont plus communément appelés des modèles de connaissance, car ils sont construit à partir d'une analyse physique, chimique, biologique, en appliquant les lois générales sur des principes (lois de la mécanique, de la physique quantique etc...) qui régissent les mécanismes intervenant dans les phénomènes étudiés. Cependant dans certains cas le phénomène à étudier est trop mal connu pour établir un modèle de connaissance suffisamment précis. Dans ces cas là, l'autre alternative c'est d'utiliser des modèles "Boîte noire".

3.1.2 Modèle "Boîte noire"

Un modèle "Boîte noire" permet d'étudier des phénomènes complexes sans connaissance ou hypothèse à propos de leurs compositions, structures ou parties internes. Il est construit essentiellement sur la base de mesures effectuées sur les entrées et les sorties du phénomène à modéliser. Le but d'un modèle "boîte noire" c'est soit de donner une description formelle des règles de transformation liant les entrées et les sorties ou bien exhiber un comportement qui approxime ce qui est observable à partir de la sortie de la boîte noire. Un modèle "boîte noire" peut être un processus chimique, un système mécanique, une formule mathématique, un algorithme etc. En résumé avec un modèle "boîte noire" on ne cherche pas à identifier ou à comprendre les mécanismes qui régissent le phénomène à étudier. On ajuste juste des fonctions de transfert entre variables d'entrée et variables de sortie. En général les modèles "boîte noire" sont très rapides donc important pour le temps réel. Par contre comme inconvénients c'est que leurs prédictions sont parfois imprécises. Aussi le fait que les paramètres de la fonction de transfert n'ont généralement pas de signification physique, donc si la physique change, il faut les redéfinir. Comme modèle "boîte noire" nous pouvons citer les travaux de David Sussillo [123] qui utilise une boîte noire pour modéliser les dynamiques basses-dimensionnelles dans les réseaux de neurones récurrents hauts-dimensionnels. Aussi les travaux de Piero Barone [11] qui utilise une boîte noire pour résoudre le problème d'approximation des exponentiels complexes.

3.2 Modélisation statistique spatio-temporelle

3.2.1 Définition

La modélisation statistique spatio-temporelle est une étape importante pour comprendre les mécanismes de certains phénomènes naturels, comme par exemple les phénomènes environnementaux, géophysiques, géologiques, hydrologiques, biologiques etc. En effet les concentrations

de polluants atmosphériques, les recueils météorologiques, les champs de précipitations, les paramètres hydrologiques etc... sont caractérisés par des variabilités spatiales et temporelles. Donc ces phénomènes sont le plus souvent considérés comme des processus aléatoires qu'on suppose généralement gaussiens. Ainsi les mesures de ces phénomènes en des sites d'observation sont vus comme des réalisations de fonctions aléatoires.
Donc l'idée c'est d'avoir des modèles valides qui permettront la prédiction spatio-temporelle de la variable d'intérêt en des sites non observés à partir des données disponibles et aussi d'évaluer les incertitudes sur les valeurs prédites.

La modélisation statistique spatio-temporelle est un domaine qui est en pleine expansion et qui touche aussi de nombreux secteurs d'activité. Nous allons donc faire une revue des différents travaux.

3.2.2 Revue des modèles statistiques spatio-temporels

L'émergence des travaux et de la recherche dans le domaine de la modélisation spatio-temporel a commencé environ il y a une vingtaine d'année de cela. En effet durant cette période il s'est avéré que cette modélisation étant une étape incontournable pour tout ce qui est la gestion et la manipulation des données qui varient aussi bien dans l'espace et le temps. Tout en sachant que ce n'est pas du tout évident de mettre en place des modèles spatio-temporels, ceci dû à la complexité des données impliquées. Comme l'a dit Pekelis [105] les travaux précurseurs dans ce domaine ont été l'œuvre de Guting [51] et de Tansel [124]. Depuis plusieurs approches et travaux ont été réalisés dans ce domaine. Nous pouvons citer les travaux de Gail Langran [75] qui a été le premier à voir l'influence du temps dans les Systèmes d'Information Géographique (**SIG**), de Yuan [135], de Roddick [109], de Pekelis [105] qui en 2004 a fait une revue de la littérature sur les modèles spatio-temporels, et plus récemment de Cressie [25] qui a écrit un livre sur les statistiques pour les données spatio-temporelles. Tous les travaux effectués dans ce domaine ont donné lieu à plusieurs modèles spatio-temporels dont deux ont particulièrement retenu notre attention que sont les modèles de Co-krigeage Multi-fidélité [77, 78] et de krigeage avec le filtre de kalman [24, 81].

3.2.2.1 Co-krigeage Multi-fidélité

On va d'abord faire dans une première partie une présentation du co-krigeage, puis en deuxième partie parler du multi-fidélité où on utilise le co-krigeage.

3.2.2.1.a Co-krigeage

Le Co-krigeage qui est une extension du krigeage est apparu en premier en 1999 dans les travaux de Chilès et Delfiner [18], puis en 2003 dans les travaux de Wackernagel [129]. Il permet de gérer le cas où la variable régionalisée est multivariée, c'est à dire le cas où $Z(s) \in \mathbb{R}^q$, $q \in \mathbb{N}^*$, $\forall s \in \mathcal{D}$. Donc dans le cas multivarié on a $Z(s) = (Z_1(s), \dots Z_q(s))$, de ce fait pour estimer une des composantes $Z_i(s)$, $1 \leq i \leq q$, il faudra tenir compte des autres composantes de la variable régionalisée multivariée. La composante que l'on veut estimer est appelée variable principale et les autres composantes de la variable régionalisée multivariée sont appelées des variables secondaires. En partant toujours du modèle de base du krigeage (2.8), on a $Z(s)$ qui suit une loi Gaussienne multivariée de moyenne :

$$\mu(.) = (\mu^1(s), ..., \mu^q(s)),$$

où $\mu^i(s) = (m_i(s))'\beta_i$ avec $m_i(.)$ est une fonction connue à valeurs dans $\mathbb{R}^{p_i}$ et β_i est un vecteur à p_i coefficients inconnus à estimer. La matrice de variance covariance croisée de $Z(s)$ est :

$$K(s, \tilde{s}; \theta) = \text{Cov}(Z(s), Z(\tilde{s}), \theta) = \begin{pmatrix} k_{11}(s, \tilde{s}; \theta_{11}) & \dots & k_{1q}(s, \tilde{s}; \theta_{1q}) \\ \vdots & \ddots & \vdots \\ k_{q1}(s, \tilde{s}; \theta_{q1}) & \dots & k_{qq}(s, \tilde{s}; \theta_{qq}) \end{pmatrix},$$

où

- $k_{ij}(s, \tilde{s}; \theta_{ij}) = \text{Cov}(Z_i(s), Z_j(\tilde{s}); \theta_{ij}),\ 1 \leq i, j \leq q$
- $\theta_{ij} = (\sigma_i^2, \sigma_j^2, \phi_i, \phi_j, \tau_i^2, \tau_j^2)$ représente le paramètre du noyau de covariance $k_{ij}(s, \tilde{s}; \theta_{ij})$
- $\theta = (\theta_{ij})_{1 \leq i,j \leq q}$

On note que $K(s, \tilde{s}, \theta)$ n'est pas forcément symétrique, c'est à dire qu'on peut avoir $k_{ij}(s, \tilde{s}; \theta_{ij}) \neq k_{ji}(s, \tilde{s}; \theta_{ji})$ même si $\theta_{ij} = \theta ji$. Il est important de préciser qu'il faut s'assurer que $K(s, \tilde{s}; \theta)$ est définie positive. Un moyen de s'en assurer c'est de vérifier que toute transformation linéaire du processus gaussien multivarié est un processus gaussien multivarié (voir [17, 50, 129]), ou bien effectué une analyse spectrale de la structure de covariance multivariée (voir [47]). On note aussi comme le suggère Wackernagel [128] qu'en co-krigeage, il est préférable d'utiliser le **co-variogramme** croisé parce que ça permet d'utiliser des variables mesurées à des localisations ne coïncidant pas pour chaque variable, et aussi de tenir compte de l'asymétrie de corrélations.

Alors qu'en krigeage comme ça été dit dans la section 2.4.1, il est recommandé d'utiliser le **variogramme**. Par contre l'inconvénient du **covariogramme** croisé comparer au **variogramme** croisé c'est qu'il requiert d'estimer les moyennes de chaque variable.

Dans ce qui suit, on va se mettre dans le cas où la variable régionalisée est bivariée, donc on a $q = 2$ et $Z(s) = (Z_1(s), Z_2(s))$ et on suppose que $K(s, \tilde{s}, \theta)$ est symétrique définie positive. Le prédicteur linéaire de $Z_2(s)$ s'écrit comme suit :

$$\widehat{Z}_2(s) = \sum_{i=1}^{n_2} \alpha_i Z_2(s_i^{(2)}) + \sum_{i=1}^{n_1} \lambda_i Z_1(s_i^{(1)}) = (\alpha)' Z_2^{n_2} + (\lambda)' Z_2^{n_1}, \tag{3.4}$$

n_2 et n_1 étant respectivement le nombre de points où les variables $Z_2(s)$ et $Z_1(s)$ ont été observées, $(s_i^{(2)}) \in \mathcal{D}$, $1 \leq i \leq n_2$, étant les points où $Z_2(s)$ ont été observés et $(s_i^{(1)}) \in \mathcal{D}$, $1 \leq i \leq n_1$, les points où $Z^1(s)$ ont été observés. Quand θ est connu, l'idée est d'avoir un estimateur sans biais, i.e $\mathrm{E}(Z_2(s)) = \mathrm{E}(\widehat{Z}_2(s))$, et aussi de trouver les coefficients $\alpha = (\alpha_1, \ldots, \alpha_{n_2})$ et $\lambda = (\lambda_1, \ldots, \lambda_{n_1})$ qui minimisent $\mathrm{Var}(Z_2(s) - \widehat{Z}_2(s))$. Donc après calcul on obtient l'estimateur de $Z_2(s)$ par cokrigeage suivante :

$$\widehat{Z}_2(s) = m_2'(s)\widehat{\beta}_2 + k_2(s) V_2^{-1} (Z^{(2)} - M^{(2)}\widehat{\beta}), \tag{3.5}$$

où

- $Z^{(2)} = (Z^{n_1}, Z^{n_2})$
- $\widehat{\beta} = \left((M^{(2)} V_2^{-1} M^{(2)}) \right)^{-1} \left(M^{(2)} \right)' V_2^{-1} Z^{(2)}$, et $\widehat{\beta}_2$ sont les p_2 premières composantes de $\widehat{\beta}$
- $M^{(2)} = \begin{pmatrix} M_2 & 0 \\ 0 & M_1 \end{pmatrix}$, avec $M_i = [m_i'(s_k^{(i)})]_{k=1,\ldots,n_i}$

En pratique θ est inconnu et il est possible de l'estimer par exemple par maximum de vraisemblance. Donc dans ce cas l'estimateur de $Z_2(s))$ sera donné comme suit :

$$\widehat{Z}_2(s) = m_2'(s)\widehat{\beta}_2 + \widehat{k}_2(s) \widehat{V}_2^{-1} (Z^{(2)} - M^{(2)}\widehat{\beta}), \tag{3.6}$$

avec

- $\widehat{k}_2'(s) = \left(\widehat{k}_{22}'(s), \widehat{k}_{21}'(s) \right)$, avec $\widehat{k}_{2j}(s) = [k_{2j}(s, s_k^{(j)}, \widehat{\theta}_{2j})]_{k=1,\ldots,n_j}$
- $\widehat{V}_2 = \begin{pmatrix} \widehat{K}_{22} & \widehat{K}_{21} \\ \widehat{K}_{12} & \widehat{K}_{11} \end{pmatrix}$, avec $\widehat{K}_{ij} = [k_{ij}(s_k^{(i)}, s_l^{(j)}, \widehat{\theta}_{ij})]_{k=1,\ldots,n_i \,;\, l=1,\ldots,n_j}$

On remarque comme pour le krigeage 2.4.1 qu'il est possible de faire du co-krigeage simple, ordinaire ou universel. Cela dépendra de la connaissance ou non de l'espérance de la variable principale et aussi l'hypothèse selon laquelle elle est constante ou une combinaison linéaire de la position s, $s \in \mathcal{D}$.

3.2.2.1.b Multi-fidélité

Dans cette partie, on va tenir compte de la variation temporelle du phénomène à étudier. On définit $Z_t(s)$, $s \in D$ la variable régionalisée d'un phénomène donné aux temps $t = 1, \ldots, T$, T étant le délai d'étude du phénomène.
A t fixé, $Z_t(.)$ est un processus gaussien définit par le modèle de base du krigeage (2.8). Le multi-fidélité permet d'évaluer le niveau de fidélité de $Z_t(s)$, $s \in D$, par rapport à la réalité, à différents temps t donnés. Il peut être obtenu en utilisant le modèle auto-régressif d'ordre 1 suggéré par Kennedy et O'Hagan [67], où la loi de $Z_t(s)$ connaissant $(Z_1(s),\ldots,Z_{t-1}(s))$ est la même que connaissant $Z_{t-1}(s)$. Le modèle auto-régressif d'ordre 1 est donné de la manière suivante :

$$Z_t(s) = \rho_{t-1}(s)Z_{t-1}(s) + \delta_t(s) + \epsilon_t(s) \quad t = 2, \ldots, T, \tag{3.7}$$

Il respecte les propriétés suivantes :

- La propriété de Markov, $\mathrm{Cov}(Z_t(s), Z_{t-1}(\tilde{s}) \mid Z_{t-}(s)) = 0 \quad \forall s \neq \tilde{s}$
- $\delta_t(.)$ est indépendant de $Z_{t-1}(.), \ldots, Z_1(.)$ et de $\epsilon_t(.), \ldots, \epsilon_1(.)$.
- $\rho_{t-1} = \dfrac{\mathrm{Cov}(Z_t(s), Z_{t-1}(s))}{\mathrm{Var}(Z_{t-1}(s))}$ représente le facteur d'échelle entre $Z_t(s)$ et $Z_{t-1}(s)$

Pour l'instant on suppose qu'il n'y a pas d'effet de pépite, donc τ_t^2 qui est la variance du bruit blanc $\epsilon_t(.)$ est égale à zéro. Le Gratiet [77, 78] suppose que $\rho_{t-1} = g_{t-1}\eta_{t-1}$, $t = 2, \ldots, T$, où $g_{t-1}(s) = (g^1_{\rho_{t-1}}, \ldots, g^{q_{t-1}}_{\rho_{t-1}})'$ est un vecteur de taille q_{t-1} de fonctions de régression et $\eta_{t-1} \in \mathbb{R}^{q_{t-1}}$. On a $k_t(s, \tilde{s}) = \mathrm{Cov}(\delta(s), \delta(\tilde{s})) = \sigma_t^2 r_t(s - \tilde{(s)}; \phi_t)$, où σ_t^2 est la variance du processus gaussien $\delta_t(.)$, et ϕ_t est le paramètre de corrélation de la fonction r_t. On a aussi $\mu_t(s) = m_t'(s)\beta_t$, ou $m_t(s)$ est un vecteur de fonction de régression de taille p_t et $\beta_t \in \mathbb{R}^p_t$. Donc pour $\sigma^2 = (\sigma_i^2)_{i=1,\ldots,t}$,

$\phi = (\phi_i)_{i=1,\dots,t}$, $\beta = (\beta_i)_{i=1,\dots,t}$, $\eta = (\eta_{i-1})_{i=2,\dots,t}$ connus on a :

$$\mathbb{E}[Z_t(s) \mid \sigma^2, \phi, \beta, \beta_\rho] = h_t'(s)\beta, \tag{3.8}$$

$$h_t'(s) = \left(\left(\prod_{i=1}^{t-1} \rho_i(s) \right) f_1'(s), \left(\prod_{i=2}^{t-1} \right) f_2'(s), \dots, \rho_{t-1}(s) f_{t-1}'(s), f_t'(s) \right), \tag{3.9}$$

et

$$\mathrm{Cov}(Z_t(s), Z_t(\tilde{s}) \mid \sigma^2, \phi, \beta, \beta_\rho) = \sum_{j=1}^{t} \sigma_j^2 \left(\prod_{i=j}^{t-1} \rho_i^2(s) \right) r_j(s - \tilde{s}, \phi_j). \tag{3.10}$$

Pour le cas où l'effet de pépite n'est pas supposé nul, voir [78, p. 207].
On note par P_t le plan d'expérience à l'instant t, Pour s'assurer d'avoir des expressions analytiques des paramètres à estimer, il est mieux de supposer $P_t \subseteq P_{t-1}$.

En supposant que $T = 2$ et $\rho_{t-1}(s) = \rho_{t-1}$, $\forall t \in \{1, \dots, T\}$ on a le modèle de co-krigeage multi-fidélité qui s'écrit comme suit :

$$Z_2(s) = \rho_1 Z_1(s) + \delta_2(s), \quad s \in \mathcal{D}. \tag{3.11}$$

Les paramètres du modèle (3.11) sont $(\beta_1, \beta_2, \rho_1)$, (ϕ_1, ϕ_2). Connaissant ces paramètres et les observations $\boldsymbol{Z} = (Z(P_1), Z(P_2))$, la distribution prédictive de $Z_2(s)$, $s \in \mathcal{D}$ suit une gaussienne dont la moyenne et la variance sont exprimés sous forme analytique. Mais en pratique les paramètres cités ci-dessus ne sont pas connus et il faut les estimer à partir des observations $\mathbf{Z}$. Une des façon d'estimer ces paramètres est d'utiliser des méthodes bayésiennes. Pour plus de détails (voir [78, p. 92])

3.2.2.2 Krigeage avec filtre de Kalman

On va d'abord faire une présentation de ce qui est le filtrage en statistique, ensuite faire une présentation du filtre de Kalman et de ses extensions, et pour finir présenter le cas où l'on associe avec le krigeage et le filtre de Kalman pour faire de la modélisation spatio-temporelle.

3.2.2.2.a Filtrage en statistique

Le filtrage consiste à estimer les variable d'états d'un système dynamique que l'on observe partiellement à travers une série de mesures. Les états $(\alpha_t)_{t\geq 0}$ du système sont soumis à une

évolution dans le temps décrite par une équation dynamique du type :

$$\alpha_t = f_t(\alpha_{t-1},\ \eta_t) \tag{3.12}$$

où le vecteur des états α_t prend ses valeurs dans $\mathbb{R}^p,\ p \geq 1$ étant un entier désignant le nombre d'états. La suite de variables aléatoires (v.a.) $(\eta_t)_{t\geq 0}$ à valeurs dans $\mathbb{R}^q,\ q \geq 1$, modélise les perturbations aléatoires de la dynamique et l'imperfection du modèle. La fonction f_t est à valeurs dans $\mathbb{R}^p$. Ce système est partiellement observé à travers une série de mesures z_t fournies par un instrument. Ce processus d'observation est modélisé par l'équation suivante :

$$z_t = h(\alpha_t) + \epsilon_t \tag{3.13}$$

h est une fonction d'observation connue à valeurs dans $\mathbb{R}^m,\ m \geq 1$ et $(\epsilon_t)_{t\geq 0}$ une suite de v.a. décrivant les erreurs de mesures dont l'intensité est fonction des imperfections des capteurs et du rapport signal sur bruit. Étant donné la série d'observations bruitées $\{z_0,\ z_1,\ z_2,\ \cdots z_t\}$, on souhaite estimer l'état du système α_t.
Les équations (3.12) et (3.13) combinées constituent un modèle espace-état. Parmi les modèles espace-état on peut citer les modèles ARIMA, les modèles à composantes inobservables, les modèles à tendances stochastiques et les modèles à coefficients aléatoires.

La problématique du filtrage en statistique c'est l'estimation de la probabilité conditionnelle de α_t sachant les mesures passées $\{z_0,\ z_1,\ z_2,\ \cdots z_t\}$, qu'on note $p_t = p(\alpha_t|z_{0:t})$. Un filtre optimal est une suite de la probabilité conditionnelle $(p_t)_{t\geq 0}$. Parmi les filtres optimaux on peut citer le filtre de Kalman qui est le filtre optimal pour les modèles linéaires gaussiens.

3.2.2.2.b Le filtre de Kalman et ses extensions

1. **le filtre de Kalman**

   Le filtre de Kalman dont le nom vient de R.E Kálmán, qui par son intérêt d'appliquer le concept de vecteurs d'état dans le problème de filtrage de Wiener, a mis en place ce filtre en 1960 [65]. Depuis le modèle et les techniques du filtre de Kalman ont beaucoup attiré et continue d'attirer l'attention des statisticiens du fait de leurs ressemblances avec les modèles de régression linéaire et l'analyse des séries temporelles. Le filtre de Kalman est d'une grande utilité dans beaucoup de domaine d'applications. Par exemple dans le

contrôle et l'estimation de trajectoire, on peut citer le livre de A.H. Jazwinski [57], et les travaux de Man Hong et Shao Cheng [55], qui grâce au filtre de Kalman ils suivent à la trace la température d'un réacteur biologique en état de marche entière durant un processus chimique. Il est utilisé dans la finance avec les travaux de Reinaldo Marques et al. intitulé « Restricted Kalman filter applied to dynamic style analysis of actuarial funds » [83]. Il a été appliqué aussi en mécanique des fluides avec les travaux de Shubham Srivastava et Tarek Echekk [117].

C'est le filtre optimal pour les modèles linéaires gaussiens de la forme suivante :

$$\begin{cases} \alpha_t &= F_t\alpha_{t-1} + B_t + \eta_t \\ z_t &= H_t\alpha_t + D_t + \epsilon_t \end{cases} \tag{3.14}$$

avec les hypothèses :

- $\alpha_t \in \mathbb{R}^p$ représente le vecteur d'état, et $z_t \in \mathbb{R}^n$ est le vecteur des mesures
- $F_t \in \mathcal{M}_{p,p}(\mathbb{R})$et $H_t \in \mathcal{M}_{n_)p}(\mathbb{R})$ sont des matrices déterministes et connues
- $B_t \in \mathbb{R}^p$ et $D_t \in \mathbb{R}^p$ sont des vecteurs connus
- les bruits η_t et ϵ_t à valeurs dans $\mathbb{R}^p$ et $\mathbb{R}^n$ sont des bruits blancs gaussiens de matrices de covariance respectives Q_t et R_t. η_t et ϵ_t sont mutuellement indépendants et ils sont aussi indépendants de la condition initiale α_0.
- La condition initiale α_0 suit une loi gaussienne de moyenne μ_0 et de matrice de covariance P_0.

Sous ces hypothèses, on montre que la loi jointe de $(\alpha_t,\ z_t)$ est gaussienne ce qui implique que la loi conditionnelle ou loi *a posteriori* p_t l'est également. Le filtre p_t est donc entièrement déterminé par la donnée de sa moyenne $\hat{\alpha}_t = \mathrm{E}(\alpha_t|z_{0:t})$ et de sa matrice de covariance $P_t = \mathrm{Cov}(\alpha_t|z_{0:t})$. Le filtre de Kalman est un algorithme qui permet de calculer analytiquement et récursivement ces 2 moments via les étapes de prédiction et de correction suivantes :

(a) *Prédiction*

$$\begin{cases} \hat{\alpha}_{t|t-1} &= F_t\hat{\alpha}_{t-1} + B_t \\ P_{t|t-1} &= F_tP_{t-1}F_t^T + Q_t \end{cases}$$

(b) *Correction*

$$\begin{cases} K_t &= P_{t|t-1}H_t^T[H_tP_{t|t-1}H_t^T + R_t]^{-1} \\ \hat{\alpha}_t &= \hat{\alpha}_{t|t-1} + K_t[z_t - (H_t\hat{\alpha}_{t|t-1} + D_t)] \\ P_t &= [I - K_tH_t]P_{t|t-1} \end{cases}$$

où, $\alpha_{t|t-1} = \mathbb{E}(\alpha_t|\alpha_{t-1}) + B_t = F_t\alpha_{t-1} + B_t$

Algorithme 1 : Filtre de Kalman

Le terme K_t désigne la matrice de gain de Kalman. La moyenne *a posteriori* fait intervenir la mesure z_t via le terme dit d'innovation $\psi_t = z_t - (H_t\hat{\alpha}_{t|t-1} + D_t)$. Notons que les matrices $(K_t)_{t\geq 0}$, $(P_t)_{t\geq 0}$ et $(P_{t|t-1})_{t\geq 0}$ ne dépendent pas des mesures $(z_t)_{t\geq 0}$ et peuvent être calculées hors ligne.

Lorsque le bruit de dynamique ou de mesure n'est plus gaussien, le filtre de Kalman est le filtre linéaire à variance minimale.

2. **le filtre de Kalman informationnel**

Le filtre de Kalman informationnel [3, p. 138], [137] est une implémentation qui propage l'inverse J_t de la matrice de covariance P_t, appelée matrice d'information.

Introduisons $J_{t|t-1} = P_{t|t-1}^{-1}$ et $J_t = P_t^{-1}$. Le calcul récursif de $J_{t|t-1}$ et J_t s'obtient en utilisant le lemme d'inversion matricielle suivant :

Lemme 3.2.1. *Soit A, B, C et D 4 matrices telles que les inverses $(A + BCD)^{-1}$, A^{-1} et C^{-1} existent.*

Alors,

$$(A + BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}$$

Donc en utilisant le lemme ci-dessus on a :

(a)

$$J_{t|t-1} = P_{t|t-1}^{-1} = [F_t P_{t-1} F_t^T + Q_t]^{-1} \tag{3.15}$$

$$J_{t|t-1} = [F_t J_{k-1}^{-1} F_t^T + Q_t]^{-1} \tag{3.16}$$

$$J_{t|t-1} = Q_t^{-1} - Q_t^{-1} F_t [F_t^T Q_t F_t + J_{t-1}]^{-1} F_t^T Q_t^{-1} \tag{3.17}$$

(b)

$$J_t = P_t^{-1} = [P_{t|t-1} - P_{t|t-\mathrm{i}} H_t^T (H_t P_{t|t-1} H_t^T + R_t)^{-1} H_t P_{t|t-1}]^{-1} \tag{3.18}$$

$$J_t = P_t^{-1} = (P_{t|t-1}^{-1} + H_t^T R_t^{-1} H_t) \tag{3.19}$$

$$J_t = P_t^{-1} = (J_{t|t-1} + H_t^T R_t^{-1} H_t) \tag{3.20}$$

La moyenne prédite se calcule de la même façon :

$$\hat{\alpha}_{t|t-1} = F_t \hat{\alpha}_{t-1} + B_t$$

La moyenne corrigée est obtenue en exprimant le gain de Kalman K_t en fonction de $J_{t|t-1}$:

$$K_t = J_{t|t-1} H_t^T [H_t J_{t|t-1} H_t^T + R_t]^{-1}$$

$$= [H_t R_t^{-1} H_t^T + J_{t|t-1}^{-1}]^{-1} H_t^T R_t^{-1}$$

d'après le lemme d'inversion matricielle suivant :

Lemme 3.2.2. *Soit Q et R deux matrices symétriques définies positives de dimensions respectives $p \times p$ et $n \times n$. Soit H une matrice $m \times d$. Alors,*

$$(H^T R^{-1} H + Q^{-1})^{-1} = Q - QH^T (HQH^T + R)^{-1} HQ,$$

et

$$(H^T R^{-1} H + Q^{-1})^{-1} H^T = QH^T (HQH^T + R)^{-1} R.$$

D'où

$$K_t = [H_t R_t^{-1} H_t^T + P_{t|t-1}^{-1}]^{-1} H_t^T R_t^{-1}$$

$$= J_t^{-1} H_t^T R_t^{-1}.$$

Finalement,

$$\hat{\alpha}_t = \hat{\alpha}_{t|t-1} + J_t^{-1} H_t^T R_t^{-1} [z_t - (H_t \hat{\alpha}_{t|t-1} + D_t)]$$

D'après (3.17) et (3.20), on observe que dans le filtre de Kalman informationnel, on inverse essentiellement des matrices $p \times p$ (exception faite de la matrice de covariance de mesure R_t) contrairement au filtre de Kalman standard qui nécessite l'inversion de matrices $n \times n$, où n est la dimension de la mesure et p celle de l'état. Dans le cas où $n >> p$, la version informationnelle peut se révéler avantageuse.

D'autre part, le filtre de Kalman informationnel permet de traiter le cas d'une incertitude initiale P_0 infinie, qui correspond à une information initiale nulle i.e. $J_0 = 0$, là où l'écriture $P_0 = \infty$ n'aurait pas de sens numériquement dans le filtre de Kalman standard.

Comme décrit ci-dessus le filtre de Kalman n'est applicable qu'à des systèmes linéaires. Pour remédier à cette restriction il est apparu dans la littérature des extensions du filtre Kalman qui traitent le cas des systèmes non linéaires. Ces extensions sont le filtre de Kalman étendu ou Extented Kalman Filter (**EKF**) en anglais qui linéarise les équations d'état et d'observation et Le filtre de Kalman d'ensemble, Ensemble Kalman Filter (**EnKF**) en anglais ou le filtre de Kalman sans parfum, Unscented Kalman Filter (**UKF**) qui sont des approximations du filtre de Kalman par la méthode monte-carlo.

3. le filtre de Kalman étendu(EKF)

L'EKF qui est une version non linéaire du filtre de Kalman est beaucoup appliqué en contrôle et estimation de trajectoire (voir [28], [56], [121], [133] [52],[72]). On considère le système dynamique dont les fonctions d'évolution et d'observation sont non-linéaires.

$$\begin{cases} \alpha_t &= f_t(\alpha_{t-1}) + \eta_t \\ z_t &= h_t(\alpha_t) + \epsilon_t \end{cases} \tag{3.21}$$

où, η_t et ϵ_t sont des bruits blancs gaussiens mutuellement indépendants, de variance respective Q_t et R_t, et indépendant de α_0.On suppose ne plus que la condition initiale α_0 est

gaussienne.

Dans le cas où les fonctions d'observation et d'état sont dérivables, on peut linéariser les équations de prédictions et de corrections respectivement autour de l'état corrigé $\hat{\alpha}_{t-1}$ et autour de l'état prédit $\hat{\alpha}_{t|t-1}$:

$$\alpha_t \approx f_t(\hat{\alpha}_{t-1}) + \nabla f_t(\hat{\alpha}_{t-1})(\alpha_{t-1} - \hat{\alpha}_{t-1}) + \eta_t$$

$$y_t \approx h(\hat{\alpha}_{t|t-1}) + \nabla h_t(\hat{\alpha}_{t|t-1})(\alpha_t - \hat{\alpha}_{t|t-1}) + \epsilon_t$$

où ∇f_t et ∇h_t représentent les matrices jacobiennes de f_t et h_t. Soit $F_{t-1} = \nabla f_t(\hat{\alpha}_{t-1})$ et $H_t = \nabla h_t(\hat{\alpha}_{t|t-1})$. Le modèle linéarisé est donc :

$$\begin{cases} \alpha_t &= F_t\alpha_{t-1} + f_t(\hat{\alpha}_{t-1}) - \nabla f_t(\hat{\alpha}_{t-1})\hat{\alpha}_{t-1} + w_t \\ z_t &= H_t\alpha_t + h_t(\hat{\alpha}_{t|t-1}) - \nabla h_t(\hat{\alpha}_{t|t-1})\hat{\alpha}_{t|t-1} + v_t \end{cases} \tag{3.22}$$

L'algorithme du filtre de Kalman étendu est alors le suivant :

(a) *Prédiction*

$$\begin{cases} F_{t-1} &= \nabla f_t(\hat{\alpha}_{t-1}) \\ \hat{\alpha}_{t|t-1} &= f_t(\hat{\alpha}_{t-1}) \\ P_{t|t-1} &= F_{t-1t-1}PF_t' + Q_t \end{cases}$$

(b) *Correction*

$$\begin{cases} H_t &= \nabla h_t(\hat{\alpha}_{t|t-1}) \\ K_t &= P_{t|t-1}H_t'(H_tP_{t|t-1}H_t' + R_t)^{-1} \\ \hat{\alpha}_t &= \hat{\alpha}_{t|t-1} + K_t[z_t - h\ (\hat{\alpha}_{t|t-1})] \\ P_t &= (I\ - K_tH_t)P_{t|t-1} \end{cases}$$

Algorithme 2 : Filtre de Kalman étendu

L'EKF est avantageux par la simplicité de sa mise en œuvre et sa faible complexité algorithmique. Ce pendant l'EKF présente deux principales difficultés : les modèles linéaires tangents peuvent être difficiles à dérivées et la première approximation des erreurs entre les états réels et estimés peut diriger à un filtre de faible performance et susceptible de

diverger [103]

4. **Le filtre de Kalman d'ensemble(EnKF) ou le filtre de Kalman sans parfum (UKF)**

l'EnKF qui est une approximation par la méthode de Monte Carlo du filtre de Kalman et l'UKF qui utilise une technique d'échantillonnage déterministe connu sous le nom de transformée sans parfum qui sera définit ci-dessous, sont utilisés dans beaucoup de domaines tel que dans la géologie, dans la géophysique, dans la climatologie, dans le contrôle de trajectoire etc... (voir [113] [117],[7],[33], [134]). l'UKF et l'EnK ont été introduits respectivement par Julier et Uhlmann [64] et [38]. Tous les deux permettent d'éviter l'étape de linéarisation de l'EKF qui peut poser des problémes numériques lorsque l'on souhaite calculer les jacobiennes des fonctions du modèle. Dans ce rapport on va seulement faire la présentation de l'UKF.

L'UKF est un filtre particulièrement utilisé pour les systèmes extrêmement non linéaires comme il a prouvé une précision du second d'ordre sur les erreurs de covariance. Il fonctionne en calculant successivement la moyenne et la covariance *a posteriori* de l'état à l'aide d'un nombre fini d'échantillons, appelés sigma-points, qu'on note χ_t, caractéristiques de la loi conditionnelle. Les sigma-points sont calculés à chaque itération pour éviter la diminution du rang dans l'erreur de covariance. L'algorithme de filtrage récursif de l'UKF a été obtenu en définissant l'état augmenté du système α_t^a, comme la concaténation de l'état α_t et des bruits η_t et ϵ_t à l'instant t. Soit $\alpha_t^a = (\alpha_t, \eta_t, \epsilon_t)'$, α_t^a est donc un vecteur de dimension $p_a = 2p + n$. On note :

$$P_t^a = \text{Cov } (\alpha_t^a | z_{0:t}) = \begin{pmatrix} P_t & 0 & 0 \\ 0 & Q_t & 0 \\ 0 & 0 & R_t \end{pmatrix}$$

La transformation sans parfum consiste à choisir les sigma points de tel sorte que leurs moyenne et covariance soit égales à la moyenne $\overline{\alpha}_{t-1}^a$ et la covariance P_t^a après l'analyse des itérations précédentes , voir([97],p.26 pour plus de détails).

L'UKF calcule donc à chaque instant la moyenne conditionnelle $\overline{\alpha}_t$ et la variance *a posteriori* P_t.

Initialisation ;

(a) $\overline{\alpha}_0 = \mathrm{E}(\alpha_0)$

(b) $P_0 = \mathrm{E}[(\alpha_0 - \hat{\alpha}_0)(\alpha_0 - \hat{\alpha}_0)']$

(c) $P_0^a = \begin{pmatrix} P_0 & 0 & 0 \\ 0 & Q_0 & 0 \\ 0 & 0 & R_0 \end{pmatrix}$

pour $t = 1 \to n$ **faire**

Calcul des sigma-points ; $\chi_{t-1}^{a,i} = \left[\overline{\alpha}_{t-1}^a \quad \overline{\alpha}_{t-1}^a \pm \sqrt{(p_a + \kappa)P_{t-1}^{a,i}}\right]$, $i = 1 \dots 2p_a$, et $\kappa = 3 - p$

où

$\sqrt{(p_a + \kappa)P_t^{a,i}}$ est la $i^{ème}$ colonne de la décomposition de Cholesky de $(p_a + \kappa)P_{t-1}^{a,i}$.

On a :

- $w_0 = \frac{\kappa}{d+\kappa}$
- $w_i = \frac{1}{2p+\kappa}$, $i = 2, \dots, 2p$

Prédiction ;

(a) $\chi_{t|t-1}^{\alpha,i} = f_t^i(\chi_{t-1}^{\alpha,i},\ \chi_{t-1}^{\eta,i})$

(b) $\overline{\alpha}_{t|t-1} = \sum_{i=0}^{2p_a} \omega_i \chi_{t|t-1}^{\alpha,i}$

(c) $P_{t|t-1} = \sum_{i=0}^{2p_a} \omega_i (\chi_{t|t-1}^{\alpha,i} - \overline{\alpha}_{t|t-1})(\chi_{t|t-1}^{\alpha,i} - \overline{\alpha}_{t|t-1})'$

(d) $z_{t|t-1}^i = h_t(\chi_{t-1}^{x,i},\ \chi_{t-1}^{\epsilon,i})$

(e) $\overline{z}_{t|t-1} = \sum_{i=0}^{2p_a} \omega_i z_{t|t-1}^i$

Correction ;

(a) $P_{t|t-1}^z = \sum_{i=0}^{2p_a} \omega_i (z_{t|t-1}^i - \overline{z}_{t|t-1})(z_{t|t-1}^i - \overline{z}_{t|t-1})'$

(b) $P_{t|t-1}^{\alpha z} = \sum_{i=0}^{2p_a} \omega_i (\chi_{t|t-1}^{\alpha,i} - \overline{\alpha}_{t|t-1})(z_{t|t-1}^i - \overline{z}_{t|t-1})'$

(c) $K_t = P_{t|t-1}^{\alpha z} P_{t|t-1}^{z,-1}$

(d) $\overline{\alpha}_t = \overline{\alpha}_{t|t-1} + K_t(z_t - \overline{z}_{t|t-1})$

(e) $P_t = P_{t|t-1} + K_t P_{t|t-1}^z K_t'$

fin

Algorithme 3 : Algorithme de l'UKF

Dans la plupart des applications, il a été observé que l'UKF permet d'estimer l'état du système avec plus de précision que l'EKF [64] avec un gain en robustesse appréciable. De plus, il ne nécessite pas de linéarisation du modèle, ce qui rend son implémentation plus simple que celle de l'EKF. Cependant son application se limite aux applications où la loi conditionnelle $p(\alpha_t|z_{0:t})$ est monomodale.

Pour plus de détails sur le filtre de Kalman (voir [97], [79], [81], [90]). Dans la partie qui suit on va parler de la modélisation spatio-temporelle qui associe le modèle de krigeage et le filtre de Kalman.

3.2.2.2.c Krigeage combiné au filtre de Kalman

Partant toujours du modèle de base du krigeage (2.8) où on intègre l'aspect temporel, on obtient l'équation suivante :

$$Z_t(s) = m_t'(s)\beta_t + \delta_t(s) + \epsilon_t(s), \quad s \in \mathcal{D}, \quad t = 0, 1, 2, \ldots, \tag{3.23}$$

Comme présenté dans la section 2.4.1, a t fixé les paramètres à estimer sont β_t, $\sigma_t^2 = \mathrm{Var}(\delta(s))$, ϕ_t qui est le paramètre de corrélation et $\tau_t^2 = \mathrm{Var}(\epsilon_t(s))$. On suppose connu σ_t^2, ϕ_t, τ_t^2 pour tout t donné, et on va donner l'estimation du paramètre $\beta_t \in \mathbb{R}^p$ par la méthode du filtre Kalman présentée dans la sous section précédente. le modèle espace-état est obtenu en associant l'équation de mesure (3.23) avec l'équation d'état donné comme suit :

$$\beta_t = F_t\beta_{t-1} + \eta_t, \tag{3.24}$$

où,

- $F_t \in \mathcal{M}_{(p,p)}$ est une matrice connue
- η_t est un bruit blanc à valeurs dans $\mathbb{R}^p$ de matrice de covariance respectives Q_t

Soit $(s_1, \ldots, s_n)$ les points où ont été observés le processus $Z_t(.)$ et $\boldsymbol{Z_t}$=$(Z_t(s_1), \ldots, Z_t(s_n))$ le vecteur des observations effectuées à ces points. On suppose aussi que $\beta_t \sim \mathcal{N}(\psi_t, P_{\beta_t})$, donc pour chaque t donné, on veut déterminer $\widehat{\beta}_{t|t} = \psi_t$ et $P_{t|t} = P_{\beta_t}$. En partant de l'état initiale $\widehat{\beta}_{0|-1} = \psi_0$ et $P_{0|-1} = P_{\beta_0}$, ψ_0 et P_{β_0} étant choisis de manière arbitraire, on peut calculer avec l'algorithme du filtre de Kalman décrit dans la sous section précédente $\widehat{\beta}_{t|t}$ et $P_{t|t}$ pour tout $t \geq 0$. Donc après calcul on a :

$$\widehat{\beta}_{t|t} = \widehat{\beta}_{t|t-1} + K_t(\boldsymbol{t} - M_t\widehat{\beta}_{t|t-1})$$

$$P_{t|t} = P_{t|t-1} - K_t M_t P_{t|t-1},$$

avec,

– $\boldsymbol{z_t}$ qui est une réalisation du vecteur aléatoire $\boldsymbol{Z_t}$

– $M_t = (m_t(s_1), \ldots, m_t(s_n))'$

– $K_t = P_{t|t-1}M_t' \left(M_t P_{t|t-1} M_t' + \Gamma_{(\sigma_t^2,\phi_t,\tau_t^2)}\right)^{-1}$

où $\Gamma_{(\sigma_t^2,\phi_t,\tau_t^2)} = \sigma_t^2 R(\phi_t) + \tau_t^2 I_n$, avec $R(\phi_t) = (r(s_i, s_j, \phi_t))_{1\leq i,j\leq n}$ la matrice de corrélation entre les points de mesure, et I_n est la matrice identité de taille n.

A la place du filtre de Kalman classique on peut utiliser comme la fait [21] le filtre informationnel décrit ci-dessous pour estimer le paramètre β_t. Pour cela on définit

$$\hat{a}_{t|w} = P_{t|w}^{-1}\hat{\beta}_{t|w},$$

et on écrit l'équation des équations du filtre informationnel dans les variables $(\hat{a}(t|t-1),\ P(t|t-1)^{-1})$ et $(\hat{a}(t|t),\ P(t|t)^{-1})$. Comme pour le filtre de Kalman classique, les équations du filtre informationnel ont une étape de prédiction puis une étape de correction.

– **Prédiction** : La prédiction au temps t avec les données obtenues au temps $t-1$ est

$$\hat{a}_{t|t-1} = P_{t|t-1}^{-1} F_t P_{t-1|t-1}\hat{a}_{t-1|t-1}\ ,$$

avec la matrice d'information

$$P_{t|t-1}^{-1} = (F_t P_{t-1|t-1} F_t' + R_t)^{-1}$$

– **Correction** : la prédiction optimale au temps t avec les données collectées au temps t peut être exprimé comme suit :

$$\hat{a}_{t|t} = \hat{a}_{t|t-1} + M_t(\Gamma_{(\sigma_t^2,\phi_t,\tau_t^2)})^{-1}\boldsymbol{z_t}.$$

Les équations du filtre informationnel fournit une manière itérative de calcul de $\hat{a}_{t|t}$ et $P_{t|t}^{-1}$. Ce qui est approprié pour une implémentation distribuée en réseau robotique.

A t fixé, une fois le paramètre β_t est estimé par le filtre de Kalman classique ou le filtre informationnel, il ne reste plus qu'à faire la prédiction de la variable régionalisé $Z_t(s)$, $s \in D$, en des points non échantillonnés du domaine D, en utilisant les techniques du krigeage. Pour plus de détails sur le krigeage combiné au filtre de Kalman voir [21], [81]

Chapitre 4

Krigeage Bayésien spatio-temporel avec boîte noire mal spécifiée

Nous proposons un nouvel algorithme pour la prédiction spatio-temporelle. A un temps donné t, nous utilisons un modèle de krigeage Bayésien (voir Section 2.4.2) pour la prédiction spatiale. L'évolution temporelle de t à $t+1$ est donnée par une boîte noire déterministe qui peut être un code numérique complexe ou une équation aux dérivées partielles. Comme souvent dans la pratique, la boîte noire est mal spécifiée, en ce sens que ses paramètres sont connues de manière imprécises, en ce sens qu'ils sont considérés comme des variables aléatoires. A l'instant t, nous utilisons la boîte noire afin d'obtenir une prédiction non précise du phénomène d'intérêt au temps t + 1. Cette prédiction obtenue avec la boîte noire sera utilisée comme information *a priori* pour estimer les hyperparamètres du krigeage Bayésien au temps t + 1 en utilisant des méthodes de Monte Carlo. Puis lorsque les nouvelles données au temps $t+1$ sont disponibles, on va les combiner avec la nouvelle distribution *a priori* des paramètres du modèle de krigeage pour obtenir une prédiction beaucoup plus précise du phénomène d'intérêt au temps $t+1$. Grâce à une application numérique, nous mettons en illustration notre algorithme spatio-temporel. A travers cet illustration nous montrons que notre méthode améliore les valeurs prédites par la boîte noire seulement.

Bayesian spatio-temporal kriging with misspecified black-box

Papa Abdoulaye FAYE, Pierre DRUILHET*, Nourddine AZZAOUI, Anne-Françoise YAO

Laboratoire de Mathématiques, Université Blaise Pascal, UMR CNRS 6620, Clermont-Ferrand

Abstract

We propose a new algorithm for spatio-temporal prediction. At a given time t, we use a Bayesian kriging model for spatial prediction. The temporal evolution from t to $t+1$ is given by a deterministic black-box which can be a complex numerical code or a partial differential equation. As often in practice, the black-box is misspecified, in the sense that its parameters are imprecisely known or may be varying randomly over time. At time t, we use the black-box to obtain a rough prediction at time $t+1$. When new data are available, the black-box is used to estimate the hyperparameters of the Bayesian kriging at time $t+1$ by using Monte Carlo methods. Through a numerical application, we show that our method improves the values predicted by the black-box only.

Keywords: `spatio-temporal prediction`, `Monte Carlo methods`, `Bayesian kriging`, `misspecified black-box`

1. Introduction

Spatio-temporal modeling is a fundamental step to understand the mechanisms that govern the evolution of a natural phenomenon. It arises when data are collected across time as well as space. The spatio-temporal models take into account temporal correlations as well as spatial correlations of data. It allows to reconstruct a phenomenon over a domain from a set of observed values. Such a reconstruction problem occurs in many areas such as climatology, meteorology, geology, atmospheric sciences, hydrology, environment, geography etc. For example, it occurs when one aim to predict an atmospheric pollutant from a monitoring network which provides data that are collected at regular intervals.

There exists many spatio-temporal models in spatial statistic literature, see Banerjee et al. (2004) or Cressie and Wikle (2011) for a review. For example Cressie and Wikle (2006) consider spatio-temporal kriging by using Kalman Filter methods. Irwin et al. (2002) use spatio-temporal nonlinear filtering based on hierarchical statistical models. Hengl et al. (2012) propose a procedure to interpolate daily mean temperature over a whole year period by using time series of auxiliary predictors. Stein (2005) considers a number of properties of space-time covariance functions and how these relate to the spatial-temporal interactions

*Corresponding author

Email address: `Pierre.DRUILHET@univ-bpclermont.fr` (Pierre DRUILHET)

of the process. Ip and Li (2015) construct valid parametric covariance models which are computationally estimable for univariate and multivariate spatio-temporal random fields. Le Gratiet (2013) proposes Bayesian hierarchical multi-fidelity methods with covariates.

The research on spatio-temporal modelling is ongoing to deal with problems that are more and more complex. The temporal models are often given by a deterministic black-box corresponding to the modelization of the studied phenomenon. However, in practice, the black-box is often misspecified and gives a rough prediction. On the other hand, the new data give accurate information but only around the sites of observations. In the paper, we propose to combine these spatial and temporal information.

At given time t, we use Bayesian kriging to obtain the spatial prediction of phenomenon of interest. The black-box is used to obtain a first prediction at time $t+1$ and also to derive new prior distributions of the Bayesian kriging at time $t+1$. The latter is based on the new data available at time $t+1$.

The article is structured as follows: Section 2 provides a description of Bayesian kriging. In Section 3, we present our spatio-temporal prediction procedure. In Section 4, we present a numerical application.

2. Spatial Bayesian kriging

In this section, we give a presentation of Bayesian kriging. Let $\left(Z(s) \in \mathbb{R}^p, s \in \mathcal{D} \subset \mathbb{R}^d\right)$ be a random Gaussian spatial process. Most often $d = 1$, 2 or 3. The random process is observed at a few number of sites and the aim is to predict the values of $Z(.)$ at some unobserved locations.

We recall that the kriging involves the construction of a linear predictor who takes into account of the structure covariance of $Z(.)$. Classically, the stochastic model associated with kriging is defined by:

$$Z(s) = \mu(s) + \delta(s) + \epsilon(s), \quad s \in \mathcal{D}, \tag{1}$$

where:

- $Z(s)$ is called the regionalized random variable.
- $\mu(s) := \beta' (m(s))$ (β' being the transpose of β), is the deterministic component of $Z(s)$; with $m(s) \in \mathbb{R}^p$ is a vector of a known basis functions and $\beta \in \mathbb{R}^p$ is an unknown coefficients vector to be estimated.
- $\delta(.)$ a stationary Gaussian random field with zero mean and correlation function defined by $g(s, \tilde{s}) = \mathrm{Cov}(\delta(s), \delta(\tilde{s}))/\sigma^2$, for s, $\tilde{s} \in \mathcal{D}$ and $\sigma^2 = \mathrm{Var}(\delta(s))$.
- $\epsilon(.) = (\epsilon(s), s \in \mathcal{D})$ is a spatial white noise independent of $\delta(.)$. For $s \in \mathcal{D}$ the measurement error $\mathrm{Var}(\epsilon(s)) = \tau^2$ is called nugget effect.

If $\delta(.)$ is isotropic, the correlation function is reduced to the correlogram $g(h) = g(s, \tilde{s})$, where $h = \|s - \tilde{s}\|_2$ is the euclidean distance between s and $\tilde{s}$. One of the most popular family of correlogram is the Matérn family defined by:

$$g(h, \alpha, \kappa) = \left\{2^{\kappa-1}\Gamma(\kappa)\right\}^{-1} \left(\frac{h}{\alpha}\right)^{\kappa} K_{\kappa}\left(\frac{h}{\alpha}\right), \tag{2}$$

with $K_\kappa(.)$ denotes the modified Bessel function of order κ (see Gneiting et al. (2010) for further details). The corresponding variogram is

$$\gamma(h) = \mathrm{Var}(Z(s+h) - Z(s)) = \tau^2 + \sigma^2 \left\{1 - g(h;\phi)\right\}.$$

We denote by ϕ the correlation parameter, here $\phi = (\kappa, \alpha)$ in (2). In that follows, we assume that $Z(.)$ is observed at the sites $s_1, \ldots s_n$ in $\mathcal{D}$. Let $Y = (Z(s_1), \ldots, Z(s_n))$ represents the corresponding observations.

In the Bayesian framework, we put prior distributions on the parameters $(\phi, \tau^2, \sigma^2, \beta)$. One interest is to take into account prior knowledge and uncertainty on the parameters estimation.

2.1. Prior distribution specification

We first consider the situation where $\tau^2 = 0$, i.e we assume that there is no nugget effect and that the prior distribution is separable:

$$\pi(\beta, \sigma^2, \phi) = \pi(\phi)\pi(\sigma^2)\pi(\beta). \tag{3}$$

For practical reasons, we choose the following priors

$$\begin{array}{rcl} [\phi] & \sim & \Gamma(a, b) \\ [\sigma^2] & \sim & \mathcal{X}^2_{ScI}(\nu, S^2) \\ [\beta] & \sim & \mathcal{N}(e, V) \end{array} \tag{4}$$

where

- $\Gamma(a, b)$ is a gamma distribution with shape parameter a and scale parameter b.
- $\mathcal{X}^2_{ScI}(\nu, S^2)$ is a scale inverse chi-squared distribution with ν represents the number of chi-squared degrees of freedom and S^2 the scaling parameter.
- $\mathcal{N}(e, V)$ is a multivariate Gaussian distribution where e and V respectively represent the mean and the standardized covariance hyper-parameters for the prior distribution for the mean (vector) parameter β.

At starting time, if there is no prior information about the behavior of the phenomenon, one can use a flat prior $\pi(\beta) \propto 1$ for the parameter β, which corresponds to the limit case $V^{-1} = 0$ and the Jeffrey's prior $\pi(\sigma^2) \propto \frac{1}{\sigma^2}$, for the parameter σ^2, which corresponds to $\nu = 0$ in (4), e.g. see Gelman et al. (2014); Diggle and Ribeiro (2002); Bernardo (1979); Bioche and Druilhet (2015).

2.2. Posterior distribution parameters and predictive distribution of regionalized variable

The posterior distributions for the prior (4) are given by, e.g. see Diggle and Ribeiro (2007):

$$\begin{array}{rcl} [\beta \mid \sigma^2, \phi, y] & \sim & \mathcal{N}(\tilde{e}, \sigma^2 \tilde{V}) \\ [\sigma^2 \mid \phi, y] & \sim & \mathcal{X}^2_{ScI}(\tilde{\nu}, \tilde{S}^2) \\ \pi(\phi \mid y) & \propto & \pi(\phi) |\widetilde{V}|^{\frac{1}{2}} |R(\phi)|^{-\frac{1}{2}} (S^2)^{-\frac{n+\nu}{2}}, \end{array} \tag{5}$$

with $\tilde{\nu} = \nu + n$, $\tilde{e} = \widetilde{V}(V^{-1}e + M'(R(\phi))^{-1}y)$, $\tilde{V} = (V^{-1} + M'(R(\phi))M)^{-1}$ and

$$\tilde{S}^2 = \frac{\nu\widetilde{V} + e'V^{-1}e + y'\left(R(\phi)\right)^{-1}y - \tilde{e}'\widetilde{V}^{-1}\tilde{e}}{\nu + n},$$

where $M = (m(s1), \ldots, m(s_n))$, $R(\phi)$ is the correlation matrix of $(s_1, \ldots, s_n)$, and y is a realization of Y.
To accommodate a positive nugget variance $\tau^2 > 0$, in practice we use a discrete joint prior for ϕ and υ^2, where $\upsilon^2 = \dfrac{\tau^2}{\sigma^2}$. A prior distribution for υ^2 can be:

$$[\upsilon^2] \sim \mathcal{IG}(c, d), \tag{6}$$

where $\mathcal{IG}(c, d)$ is an inverse-gamma distribution with shape parameter c and scale parameter d.
In this case, we replace $R(\phi)$ in the equations above by:

$$V(\phi, \upsilon^2) = R(\phi) + \upsilon^2\mathbb{I}_n,$$

where $\mathbb{I}_n$ is $n \times n$ identity matrix, n is the number of observations.

We used the posterior distribution of kriging parameters $\pi(\phi, \upsilon^2, \sigma^2, \beta \mid y)$ to obtain the predictive distribution at an unobserved site s:

$$p\left(z(s) \mid y\right) = \int p\left(z(s) \mid y, \phi, \upsilon^2, \sigma^2, \beta\right) \pi(\phi, \upsilon^2, \sigma^2, \beta \mid y)\, d(\phi, \upsilon^2, \sigma^2, \beta). \tag{7}$$

This Bayesian predictive distribution is an average of the predictive distributions for fixed value of $(\phi, \upsilon^2, \sigma^2, \beta)$, weighted with respect to the posterior distributions of $(\phi, \upsilon^2, \sigma^2, \beta)$. In practice, $[Z(s) \mid y]$ is not easy to obtain analytically, but it can be easily approximated by Monte Carlo methods. An estimate of $Z(s)$ is given by the mean, median or mode of $[Z(s) \mid y]$.

We now address the spatio-time prediction method based on the Bayesian kriging and the black-box. Actually, we deal with a process $(Z_t(s), s \in \mathcal{D})$ which is a stationary Gaussian process for each time $t \in \mathbb{R}^+$. As previously, we aim to predict the values of $Z_t(s)$ at unobserved sites s of $\mathcal{D}$.

3. Spatio-temporal modeling including misspecified black-box

A black-box is a tool which allows to follow the temporal evolution of complex dynamic systems. For example, in domains such as fluid mechanics, ecology or biology, the black-box may be an algorithm or a partial differential equation (PDE). For given inputs at time t, the black-box gives a prediction of the outputs at time $t + 1$. However, the black-box is often misspecified, i.e. its parameters are imprecisely known and are often considered as random variables.

We denote by $\mathcal{M}$ a mesh of $\mathcal{D}$, $\mathcal{M}_0 = (s_1, \ldots, s_n)$ a subset of $\mathcal{M}$, $Z_t^{\mathcal{M}_0} = (Z_t(s_1), \ldots, Z_t(s_n))$, the observations on $\mathcal{M}_0$ at time t and $\theta_t = (\beta_t, \sigma_t^2, \upsilon_t^2, \phi_t)$ the kriging parameter at time t. The idea is to predict the unobserved values of $Z_t(.)$ in the field $\mathcal{D}$ by combining Bayesian kriging with the information brought by the fuzzy temporal model. Now, we present our spatio-temporal procedure:

Step 1*: Bayesian kriging*

At time t, we obtain the predictive distribution of the regionalized variable $Z_t^{\mathcal{M}} = \{Z_t(s),\ s \in \mathcal{M}\}$ by Bayesian kriging where the prior distributions are obtained from time $t-1$:

$$p\left(z_t^{\mathcal{M}} \mid z_t^{\mathcal{M}_0}\right) = \int p\left(z_t^{\mathcal{M}} \mid z_t^{\mathcal{M}_0}, \theta_t\right) \pi(\theta_t \mid z_t^{\mathcal{M}_0})\ d(\theta_t), \tag{8}$$

Step 2*: temporal prediction maps*

From the black-box, denoted by f, and from distribution (8), the prediction distribution at time $t+1$ of $Z_{t+1}^{\mathcal{M}}$ is given by

$$\left[Z_{t+1}^{\mathcal{M}} \mid z_t^{\mathcal{M}_0}\right] = \left[f(Z_t^{\mathcal{M}}) \mid z_t^{\mathcal{M}_0}\right]. \tag{9}$$

In practice (9) is not accessible explicitly, but it can be easily approximated by using Monte Carlo methods as follows:

- First, do K simulations of (8) and get K spatial prediction maps $z_t^{\mathcal{M}[i]}$, $i = 1, \ldots, K$.
- Then, get K temporal prediction maps at time $t+1$ by $\widehat{z}_{t+1}^{\mathcal{M}[i]} = f\left(z_t^{\mathcal{M}[i]}\right)$, $i = 1, \ldots, K$.

Step 3*: prior distribution specification*

Using the K prevision maps, at time $t+1$, we can update the hyperparameters of the prior distribution for the Bayesian kriging at time $t+1$. For each map $\widehat{z}_{t+1}^{\mathcal{M}[i]}$, $i = 1, \ldots, K$, we compute the maximum likelihood estimate $\widehat{\theta}_{t+1}^{[i]} = \left(\widehat{\phi}_{t+1}^{[i]}, \widehat{\upsilon^2}_{t+1}^{[i]}, \widehat{\sigma^2}_{t+1}^{[i]}, \widehat{\beta}_{t+1}^{[i]}\right)$ of θ_{t+1} from the kriging model

$$\left[Z_{t+1}^{\mathcal{M}} \mid \theta_{t+1}\right] \sim \mathcal{N}\left(M_{t+1}\beta_{t+1}, \sigma_{t+1}^2\left(R(\phi_{t+1}) + \upsilon_{t+1}^2 \mathbb{I}_q\right)\right), \tag{10}$$

with $M_{t+1} = (m_{t+1}(s_1), \ldots, m_{t+1}(s_q))$, q is the number of points of $\mathcal{M}$, $R(\phi_{t+1})$ is the correlation matrix of $\mathcal{M}$ and $\mathbb{I}_q$ is $q \times q$ identity matrix.

From $\widehat{\theta}_{t+1}^{[i]}$, $i = 1, ..., K$ we can estimate the hyper-parameters a_{t+1}, b_{t+1}, c_{t+1}, d_{t+1}, e_{t+1}, V_{t+1}, ν_{t+1}, S_{t+1}^2 given in (4) and in (6) by moment methods.

- For the correlation parameter ϕ_{t+1} which follows a gamma prior distribution $\Gamma(a_{t+1}, b_{t+1})$:

$$\begin{cases} \widehat{a}_{t+1}\widehat{b}_{t+1} &= \dfrac{1}{K}\displaystyle\sum_{i=1}^{K} \widehat{\phi}_{t+1}^{[i]} \\ \widehat{a}_{t+1}\widehat{b}_{t+1}^2 &= \dfrac{1}{K}\displaystyle\sum_{i=1}^{K} \left(\widehat{\phi}_{t+1}^{[i]} - \sum_{j=1}^{K} \widehat{\phi}_{t+1}^{[j]}\right)^2. \end{cases}$$

- For the scale parameter σ^2_{t+1} which follows a inverse-chi-squared prior distribution $\mathcal{X}^2_{ScI}(\nu_{t+1}, S^2_{t+1})$:

$$\left\{\begin{array}{rcl} \dfrac{\widehat{\nu}_{t+1}\widehat{S^2}_{t+1}}{(\widehat{\nu}_{t+1}-2)} & = & \dfrac{1}{K}\displaystyle\sum_{i=1}^{K}\widehat{\sigma^2}^{[i]}_{t+1} \\ \dfrac{2\widehat{\nu}^2_{t+1}\widehat{S^2}^2_{t+1}}{(\widehat{\nu}_{t+1}-2)^2(\widehat{\nu}_{t+1}-4)} & = & \dfrac{1}{K}\displaystyle\sum_{i=1}^{K}\left(\widehat{\sigma^2}^{[i]}_{t+1}-\sum_{j=1}^{K}\widehat{\sigma^2}^{[j]}_{t+1}\right)^2. \end{array}\right.$$

- For the relative nugget parameter υ^2_{t+1} which follows a inverse-gamma prior distribution $\mathcal{IG}(c_{t+1}, d_{t+1})$:

$$\left\{\begin{array}{rcl} \dfrac{\widehat{d}_{t+1}}{(\widehat{c}_{t+1}-1)} & = & \dfrac{1}{K}\displaystyle\sum_{i=1}^{K}\widehat{\upsilon^2}^{[i]}_{t+1} \\ \dfrac{\widehat{d}^2_{t+1}}{(\widehat{c}_{t+1}-1)^2(\widehat{c}_{t+1}-2)} & = & \dfrac{1}{K}\displaystyle\sum_{i=1}^{K}\left(\widehat{\upsilon^2}^{[i]}_{t+1}-\sum_{j=1}^{K}\widehat{\upsilon^2}^{[j]}_{t+1}\right)^2. \end{array}\right.$$

- For the mean parameter β_{t+1} which follows a normal prior distribution $\mathcal{N}(e_{t+1}, V_{t+1})$:

$$\left\{\begin{array}{rcl} \widehat{e}_{t+1} & = & \dfrac{1}{K}\displaystyle\sum_{i=1}^{K}\widehat{\beta}^{[i]}_{t+1} \\ \widehat{V}_{t+1} & = & \dfrac{1}{K}\displaystyle\sum_{i=1}^{K}\left(\widehat{\beta}^{[i]}_{t+1}(\widehat{\beta}^{[i]}_{t+1})' - \widehat{e}_{t+1}(\widehat{e}_{t+1})'\right) \end{array}\right.$$

The procedure is summarized in Figure 1.

4. Numerical application

Here, we consider an application of our procedure in a convection-diffusion modeling problem which occurs in fluid mechanics such as groundwater pollution problem or flow of oil from a well's reservoir, see Dehghan (2005); Efendiev and Durlofsky (2003) or Song and Wu (2010).

Let $\mathcal{D}$ be a square domain $(a,b)^2$ in $\mathbb{R}^2$ and $(0,T)$ be a time interval. Given a function $\mathcal{S} = \mathcal{S}(x,y,t) \in L^2(\mathcal{D}\times(0,T))$, we consider the two-dimensional time-dependent convection-diffusion equation with homogeneous Dirichlet boundary condition.

$$\left\{\begin{array}{rcl} L(u) & := & \frac{\partial u}{\partial t} - \nu\Delta u + \varphi_1\frac{\partial u}{\partial x} + \varphi_2\frac{\partial u}{\partial y} = \mathcal{S}, \text{in } \mathcal{D}\times(0,T] \\ u(x,y,t) & = & 0 \text{ on } \partial\mathcal{D}\times(0,T] \\ u(x,y,0) & = & u_0(x,y) \text{ in } \mathcal{D} \end{array}\right. \tag{11}$$

where u is the state variable to be modelized, φ_1 and φ_2 are constants which represent the convection coefficients, and ν is the positive diffusion coefficient. As in Song and Wu (2010), we assume that the above problem admits a unique smooth solution.

The term $\mathcal{S}$, called the source term, represents a source which continually bring a quantity of pollution in the domain $(a,b)^2$ in the problem of interest of groundwater pollution. The

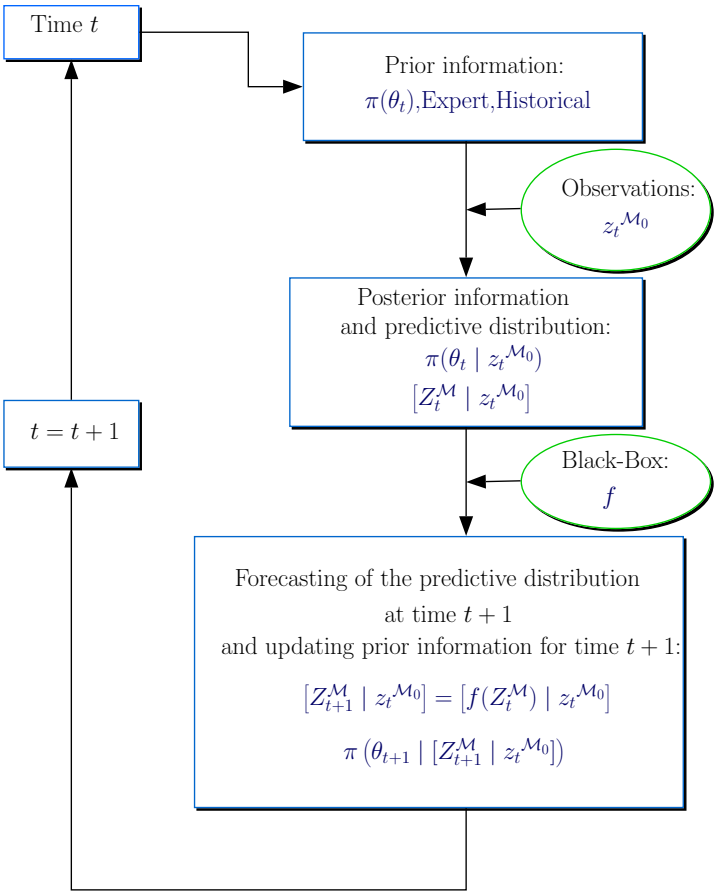

Figure 1: Spatio-temporal procedure

pollution's level due to $\mathcal{S}$ is not uniform in $(a,b)^2$ (higher in some parts of $(a,b)^2$ than in others). Once in the domain $(a,b)^2$, the quantity of pollution will change over time owing to the convection and diffusion phenomenon. Our aim is then to determine the state of the pollution, u, in $(a,b)^2$ at each time in view of Equation (11). The direction and velocity of pollution transport will depend on the values of φ_1 and φ_2, when the pollution's dispersion will depend on ν.

For our example, we assume that the pollution problem follows the PDE (11) with $(a,b)^2 = (0,1)^2$, $\varphi_1 = -0.09$, $\varphi_2 = -0.09$, $\nu = 0.001$, $\mathcal{S}(x,y,t) = (\frac{5}{4} - ((x-0.8)^2 + (y-\frac{1}{2})^2))$. We use a 24×24 mesh, say $\mathcal{M}$ to compute the values $u(x,y,t)$, $(x,y) \in \mathcal{M}$, and $t = 1, ..., 9$ which are displayed in Figure 3. We split the mesh points in two sets (see Figure 2): the set of observation sites, say $\mathcal{M}_0$, which correspond to the measurement points in real data applications and the set of sites of non-observed values.

Here, the black-box $f_{\varphi_1,\varphi_2,\nu}$ is defined by the same PDE (11) but the parameters $\varphi_1, \varphi_2, \nu$ are imprecisely known and considered as random with

$$\varphi_1, \varphi_2 \sim \mathcal{N}(-0.1, (0.01)^2) \text{ and } \nu \sim \mathcal{N}(0.006, (0.001)^2). \tag{12}$$

At the starting time, $t = 1$, we do a Bayesian kriging (Step 1) with the arbitrary values for the hyperparameters: $a_1 = 3$, $b_1 = 2$, $\nu_1 = 4$, $S_1^2 = 5$, $c_1 = 3$, $d_1 = 1$ and a flat prior for β_1 (see subsection 2.1). The Bayesian kriging was done using the package geoR of the R software, but an additional programming has been necessary to incorporate the priors for correlation and relative nugget parameters specified in (4) and in (6). At Step 2, we draw K maps $U_1^{[i]} = \{u^{[i]}(x, y, 1),\ (x, y) \in \mathcal{M}\}$ from the Bayesian kriging and K values for $\varphi_1^{[i]}, \varphi_2^{[i]}$ and $\nu^{[i]}$ according to (12), $i = 1, ..., K$, with $K = 10$. Then, we get 10 predictive maps $\widehat{U}_2^{[i]} = f_{\varphi_1^{[i]}, \varphi_2^{[i]}, \nu^{[i]}}(U_1^{[i]})$, $i = 1, ..., 10$. At Step 3, we use the maps $\widehat{U}_2^{[i]}$, $i = 1, .., 10$, to estimate the hyperparameters of the Bayesian kriging that will be done at time $t = 2$ when updated observations will be available.

Recursively, for $t = 2, \ldots, 9$, we follow steps 1 to 3, using the prior distribution obtained at the previous time.

At time t, after Step 2, we have a first prediction map of $u(x, y, t + 1)$ given by $\widehat{B}_{t+1} = \frac{1}{10} \sum_{i=1}^{10} \widehat{U}_{t+1}^{[i]}$ from the black-box. At time $t + 1$, after Step 1, we obtain a second prediction map $\widehat{u}_{t+1}$ obtained by the Bayesian kriging. The prediction errors of each method is evaluated resp. by $|\widehat{B}_{t+1}(x, y) - u(x, y, t + 1)|$ and $|\widehat{u}_{t+1}(x, y) - u(x, y, t + 1)|$. They are displayed in Figure 4 and 5, for $t = 2, \ldots, 9$.

These results show that our algorithm outperform the black-box prediction $\widehat{B}(x, y, t)$. Indeed, for any time $t = 2, \ldots, 9$, we see that the prediction errors of our algorithm are smaller or equal to those of the misspecified black-box in the domain $(0, 1)^2$. For example in the subdomain $(0, 0.6) \times (0, 0.2)$ the prediction errors of our algorithm are any time smaller than those of the misspecified black-box. The outperformance of our procedure means that for $t = 2, \ldots, 9$ the conditions (number of measurement sites, priors) required by Bayesian kriging are satisfactory to improve the prediction given by the misspecified black-box.

The code ran in approximately 72 minutes on a laptop computer (Processor: Intel(R) Core(TM) i5-4200 CPU @ 1.60GHz 2.30Ghz, RAM: 8Go). However, the duration can be largely reduced by parallelization of the code.

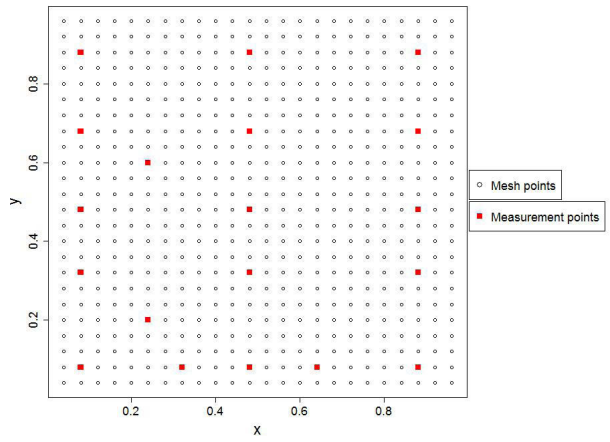

Figure 2: Mesh of domain $(0,1)^2$

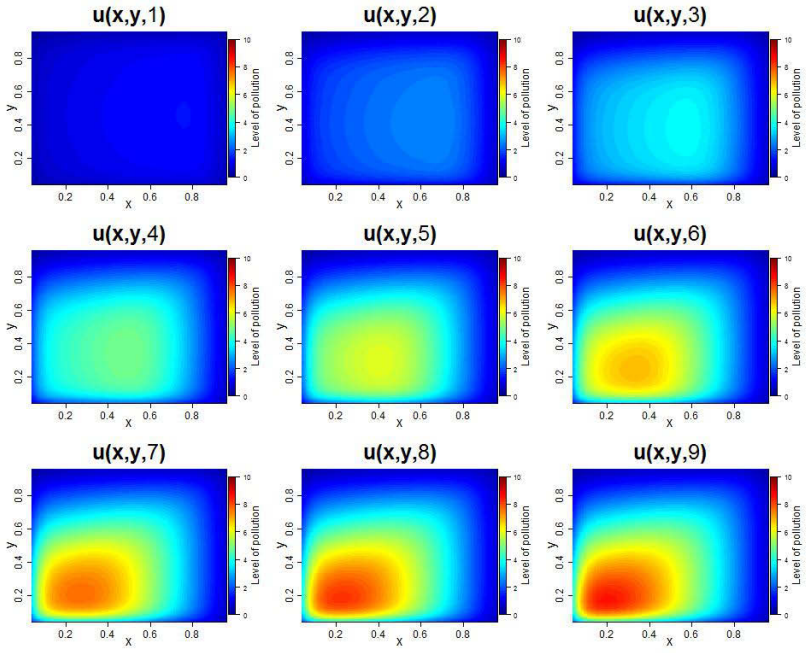

Figure 3: Evolution of real values of pollution to $t = 1$ at $t = 9$

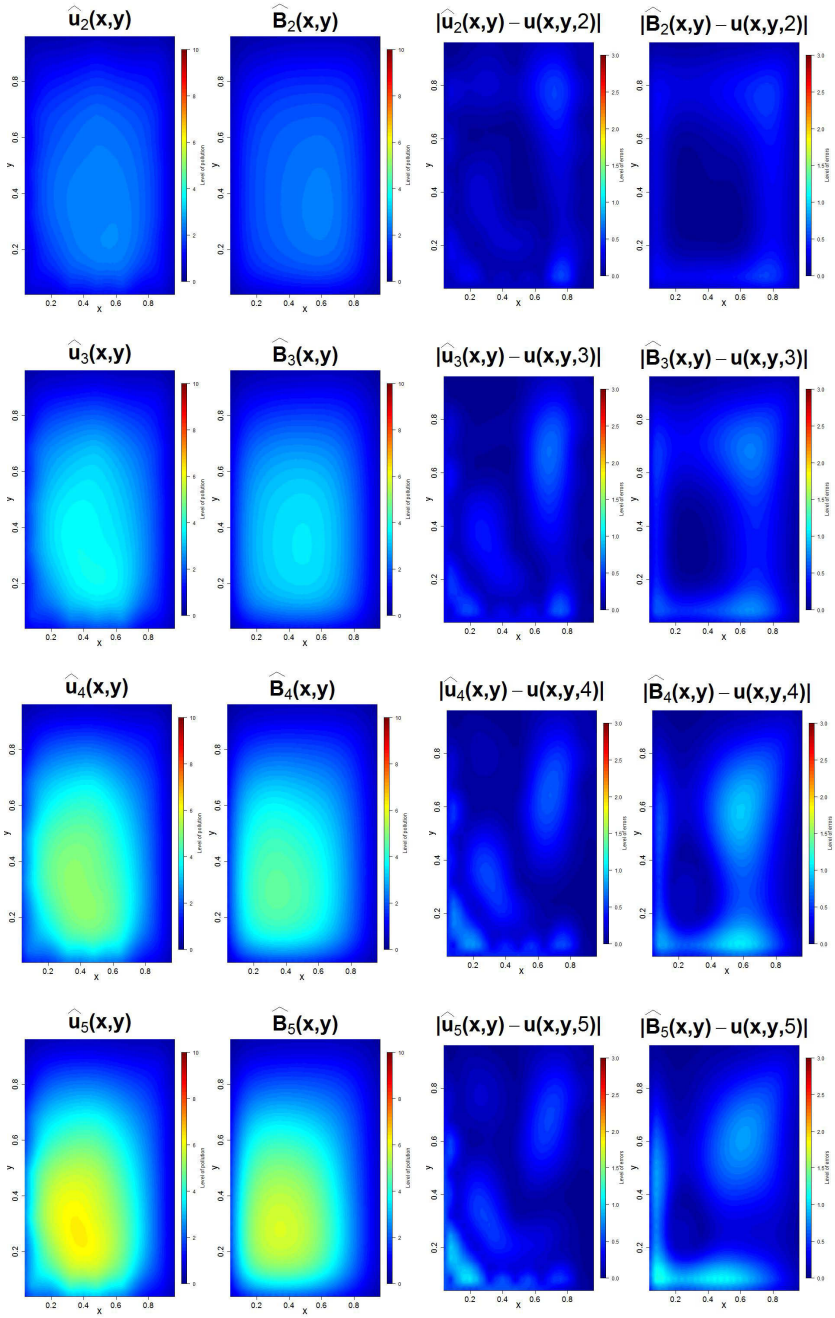

Figure 4: $\widehat{u}_t(x,y)$: Predicted values by our method, $\widehat{B}_t(x,y)$: Predicted values by misspecified black-box, $|\widehat{u}_t(x,y) - u(x,y,t)|$ predicted errors by our method, $|\widehat{B}_t(x,y) - u(x,y,t)|$ predicted errors by misspecified black-box, $t = 2, ..., 5$.

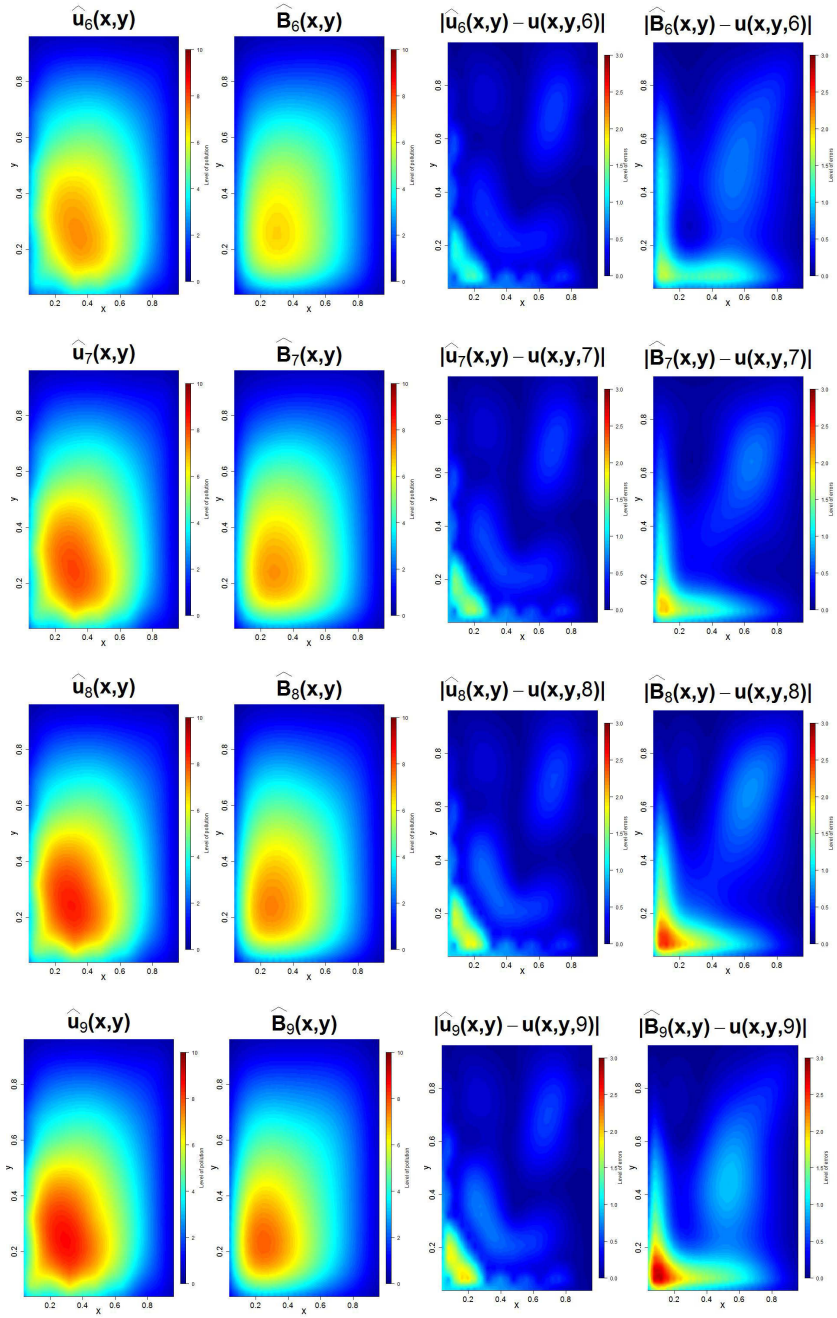

Figure 5: $\widehat{u}_t(x,y)$: Predicted values by our method, $\widehat{B}_t(x,y)$: Predicted values by misspecified black-box, $|\widehat{u}_t(x,y) - u(x,y,t)|$ predicted errors by our method, $|\widehat{B}_t(x,y) - u(x,y,t)|$ predicted errors by misspecified black-box, $t = 6, ..., 9$.

5. Conclusion

We have presented a spatio-temporal prediction algorithm which combines two kind of information: the information brought by the temporal model and the information brought

by the spatial kriging model. Both informations are fuzzy. The temporal model may be imperfect due to the imprecise knowledge of the inputs and/or because the black box represents a rough approximation of the underlying temporal model, inducing bias or extra-variability. On the other hand, the kriging model give accurate information locally around the observation sites but not on the whole region of interest. Our procedure use the temporal model to provide the prior information for the spatial model and therefore to combine two informations of different kind.

We think that this algorithm, or any variant, will find a wide scope of applicability in environmental science where time forward prediction and monitoring of spatio-temporal processes is a main activity, since it is easy to implement. The generality of the algorithm is due to the fact that the temporal evolution of the deterministic part of any physical phenomenon can be modelized by a black-box.

An application of our algorithm was done by using a convection-diffusion model which occurs in many domains of environmental sciences as problem of groundwater pollution. We used data simulated with a two-dimensional time-dependent convection-diffusion equation with homogeneous Dirichlet boundary condition. The black-box use the same equation but with misspecified coefficients. We have tested the performance of our method by comparing the prediction map obtained by the black-box at time t+1, before having the new data, with the one obtained by Bayesian kriging with updated prior once the data are observed.

Acknowledgments

This work has been funded in part by Auvergne Region and in by European Regional Development Fund.

References

Banerjee, S., Gelfand, A. E., Carlin, B. P., 2004. Hierarchical modeling and analysis for spatial data. Crc Press.

Bernardo, J.-M., 1979. Reference posterior distributions for Bayesian inference. J. Roy. Statist. Soc. Ser. B 41 (2), 113–147, with discussion.

Bioche, C., Druilhet, P., 2015. Approximation of improper priors. Bernoulli To appear.

Cressie, N., Wikle, C. K., 2006. Space-Time Kalman Filter. John Wiley & Sons, Ltd.

Cressie, N., Wikle, C. K., 2011. Statistics for spatio-temporal data. John Wiley & Sons.

Dehghan, M., 2005. On the numerical solution of the one-dimensional convection-diffusion equation. Mathematical Problems in Engineering 2005 (1), 61–74.

Diggle, P. J., Ribeiro, P. J., 2002. Bayesian inference in gaussian model-based geostatistics. Geographical and Environmental Modelling 6 (2), 129–146.

Diggle, P. J., Ribeiro, P. J., 2007. Model-based geostatistics. Springer Series in Statistics. Springer, New York.

Efendiev, Y., Durlofsky, L., 2003. A generalized convection-diffusion model for subgrid transport in porous media. Multiscale Modeling & Simulation 1 (3), 504–526.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., Rubin, D. B., 2014. Bayesian data analysis, 3rd Edition. Texts in Statistical Science Series. CRC Press, Boca Raton, FL.

Gneiting, T., Kleiber, W., Schlather, M., 2010. Matérn cross-covariance functions for multivariate random fields. Journal of the American Statistical Association 105 (491).

Hengl, T., Heuvelink, G. B., Tadić, M. P., Pebesma, E. J., 2012. Spatio-temporal prediction of daily temperatures using time-series of modis lst images. Theoretical and applied climatology 107 (1-2), 265–277.

Ip, R. H., Li, W., 2015. Time varying spatio-temporal covariance models. Spatial Statistics To appear.

Irwin, M. E., Cressie, N., Johannesson, G., 2002. Spatial-temporal nonlinear filtering based on hierarchical statistical models. Test 11 (2), 249–280.

Le Gratiet, L., 2013. Bayesian analysis of hierarchical multifidelity codes. SIAM/ASA J. Uncertain. Quantif. 1 (1), 244–269.

Song, L.-J., Wu, Y.-J., 2010. A modified Crank-Nicolson scheme with incremental unknowns for convection dominated diffusion equations. Appl. Math. Comput. 215 (9), 3293–3301.

Stein, M. L., 2005. Space-time covariance functions. J. Amer. Statist. Assoc. 100 (469), 310–321.

Chapitre 5

Estimation spatio-temporelle avec un modèle semi-empirique utilisant une boîte noire et la méthode du complément de Schur

5.1 Le complément de Schur

En 1917, dans l'article de I.Schur [112] apparaissait un lemme qui stipule que si

$$M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$$

est une matrice carrée et A est une sous-matrice non singulière de M alors

$$|M| = |A||D - CA^{-1}B|, \tag{5.1}$$

où $|.|$ représente le déterminant d'une matrice. Ce lemme a depuis pris le nom de « formule du déterminant de Schur », en anglais « Schur's determinal formula ». La matrice $D - CA^{-1}B$ a été nommé par E. V. Haynsworth [54] le complément de Schur de A dans M et est notée par (M/A). Cette notation a été inspirée par le fait qu'on a :

$$|M/A| = |M|/|A|.$$

Le complément de Schur apparaît dans plusieurs secteurs de la statistique, et d'ailleurs comme l'a remarqué W.Cottle [22, p. 192] la distribution normale multivariée fournit un exemple d'application naturelle du complément de Schur. Soit le vecteur aléatoire

$$X = \begin{pmatrix} X_1 \\ X_2, \end{pmatrix},$$

qui suit une loi normale multivariée, de moyenne

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix},$$

et de matrice de covariance

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$

où Σ_{22} est une matrice définie positive. Alors la distribution conditionnelle de X_1 sachant X_2 est une normale multivariée avec comme moyenne :

$$\nu_1 = \mu_1 + \Sigma_{12}(\Sigma_{22})^{-1}(X_2 - \mu_2),$$

et comme matrice de covariance le complément de Schur de Σ_{22} dans Σ

$$(\Sigma/\Sigma_{22}) = \Sigma_{11} - \Sigma_{12}(\Sigma_{22})^{-1}\Sigma_{21}.$$

Pour la preuve de ce résultat voir par exemple [102, p. 273].

5.2 Estimation spatio-temporelle d'un processus gaussien en utilisant une boîte noire et le complément de Schur

Soit $Z_t(.)$ un processus gaussien stationnaire pour tout $t \geq 0$, localisé dans un domaine $\mathcal{D} \subset \mathbb{R}^d$, $d = 1, 2, 3$. Soit $\mathcal{M}$ un maillage de taille $m \in \mathbb{N}^*$ du domaine $\mathcal{D}$, $\mathcal{M}_0 = (s_1, \ldots, s_n) \subset \mathcal{M}$, le plan d'expériences, $Z_t^{\mathcal{M}_0} = (Z_t(s_1), \ldots, Z_t(s_n))$ les observations correspondant à l'instant t. Nous notons que nous supposons qu'il n'y pas d'erreurs de mesure, c'est-à-dire la variance du bruit blanc gaussien est nulle.

On a :

$$Z_t^{\mathcal{M}} \sim \mathcal{N}(\mu_t^{\mathcal{M}} \in \mathbb{R}^m, \Sigma_t),$$

avec $Z_t^{\mathcal{M}} = \{Z_t(s), s \in \mathcal{M}\}$.

L'idée est d'obtenir à partir des observations de $Z_t(.)$ aux points $\mathcal{M}_0$, la prédiction spatiale de $Z_t(.)$ aux points de $\mathcal{M}$ qui n'ont pas été observés. On utilise ensuite la boîte noire pour avoir une prévision temporelle, i.e. au temps $t+1$ de cette prédiction de $Z_t(.)$. A noter que contrairement au chapitre 4 la boîte noire ici n'est pas supposé mal spécifiée, i.e. que ces paramètres sont bien connus. Les résultats obtenus grâce à la boîte noire sont utilisés comme information *a priori* pour obtenir une prédiction meilleure de $Z_{t+1}(.)$ une fois que les observations de $Z_{t+1}(.)$ au points du plan d'expériences $\mathcal{M}_0$ sont disponibles. C'est le complément de Schur qui va nous permettre de combiner l'information apporté par la boîte noire et les nouvelles données au temps $t+1$.

La procédure de notre méthode est la suivante :

Étape 1 : Krigeage Bayésien

Au tout début, $t = 1$, on dispose que de $Z_1^{\mathcal{M}_0}$ les observations de $Z_1(.)$ au points du plan d'expériences $\mathcal{M}_0$. Grâce au krigeage Bayésien présenté à la sous section 2.4.2 et en utilisant des lois *a priori* impropres pour les paramètres du krigeage on obtient une prédiction de $Z_1(.)$ en tout point de M :

$$[Z_1^{\mathcal{M}} | z_1^{\mathcal{M}_0}] \tag{5.2}$$

Étape 2 : Prédiction temporelle des cartes avec la boîte noire

Comme au chapitre 3, la boîte noire notée par f va être utilisée pour le passage du temps t au temps $t+1$. Elle va permettre d'obtenir une estimation de la prédiction spatiale du phénomène au temps $t+1$ avec la formule suivante :

$$[Z_{t+1}^{\mathcal{M}} | z_t^{\mathcal{M}_0}] = [f(Z_t^{\mathcal{M}}) | z_t^{\mathcal{M}_0}]. \tag{5.3}$$

La distribution (5.3) n'étant pas accessible explicitement, elle va être approchée par des méthodes de Monte Carlo. La procédure est la suivante :

1. On fait K simulations de la prédiction spatiale obtenu à l'instant t, ce qui va donner K cartes de prédiction $z_t^{\mathcal{M}[i]}$, $i = 1, \ldots, K$

2. Chaque carte parmi les K cartes de prédiction simulées est considérée comme variable d'entrée de la boîte noire f, qui va fournir comme sortie la carte correspondante au temps $t+1$

$$\widehat{z}_{t+1}^{\mathcal{M}[i]} = f(z_t^{\mathcal{M}[i]}).$$

On suppose que :

$$[Z_{t+1}^{\mathcal{M}} | z_t^{\mathcal{M}_0}] \sim \mathcal{N}(\overline{\mu}_{t+1}, \overline{\Sigma}_{t+1}), \tag{5.4}$$

où $\overline{\mu}_{t+1}$ est la moyenne empirique des K cartes :

$$\overline{\mu}_{t+1} = \frac{1}{K} \Sigma_{i=1}^{K} \widehat{z}_{t+1}^{\mathcal{M}[i]}.$$

De même $\overline{\Sigma}_{t+1}$ peut être calculé de manière empirique en utilisant la formule de la variance empirique. Par contre le problème qui se pose en utilisant la variance empirique pour calculer la matrice de variance-covariance des cartes est qu'on a une forte probabilité d'obtenir une matrice mal conditionnée, à moins d'avoir un nombre de simulation K qui est largement supérieur à la taille m du maillage $\mathcal{M}$. Pour avoir une matrice de variance-covariance mieux conditionnée tel que soit la taille du maillage et le nombre de simulations K, on peut utiliser la méthode proposée par Karim M. Abadir et al [1]. Dans leur article **Design-free estimation of variance matrices**, ils font une décomposition orthogonale de la matrice de variance-covariance obtenue avec la variance empirique pour ensuite en déduire une meilleur estimation de cette dernière. En effet ils ont voulu exploiter le fait qu'une matrice orthogonale n'est jamais mal conditionnée et donc se focaliser sur l'amélioration de l'estimation des valeurs propres qui est la source du mauvais conditionnement. Ils estiment les vecteurs propres à partir juste d'une partie des données, puis les utilisent pour transformer approximativement les données en séries orthogonales qui délivreront un estimateur bien conditionné de la matrice de variance-covariance.

Nous faisons ci-dessous une description détaillée de leur méthode :

Soit Z un vecteur aléatoire de taille m et Σ la matrice de variance covariance de z qu'on ne connaît pas. Donc l'idée c'est d'avoir une bonne estimation de Σ. Pour ce faire on tire K échantillons de Z, $X' = (z^1, \ldots, z^K)$ une matrice de taille $K \times m$. Nous décomposons la matrice X' en deux composantes $X' = (X'_1, X'_2)$, où X'_1 et X'_2 sont respectivement des matrices de taille

$H \times m$ et $(K-H) \times m$, avec $K > H$.

Nous notons par $\widetilde{\Sigma}$ une estimation de Σ en utilisant la variance empirique :

$$\widetilde{\Sigma} := \frac{1}{K} X' M_K X,$$

où $M_K := \mathbb{I}_K - \frac{1}{K}\mathbf{1}_K\mathbf{1}'_K$, avec $\mathbb{I}_K$ la matrice identité de taille K et $\mathbf{1}_K$ un vecteur de 1 de taille K. La décomposition orthogonale de la matrice symétrique $\widetilde{\Sigma}$ est :

$$\widetilde{\Sigma} = \widetilde{P}\widetilde{\Lambda}\widetilde{P}', \tag{5.5}$$

où $\widetilde{P}$ est la matrice orthogonale des vecteurs propres et $\widetilde{\Lambda}$ une matrice diagonale des valeurs propres de $\widetilde{\Sigma}$. Sachant que $\widetilde{\Lambda}$ est la source du mauvais conditionnement, l'idée est donc de trouver une meilleur estimation de Λ. L'équation (5.5) permet de réécrire $\widetilde{\Lambda}$ comme suit :

$$\widetilde{\Lambda} = \widetilde{P}'\widetilde{\Sigma}\widetilde{P} = \mathrm{diag}(\mathrm{var}(\widetilde{p}'_1 Z), \ldots, \mathrm{var}(\widetilde{p}'_K Z)) \tag{5.6}$$

où les $(\widetilde{p}_i)_{i=1,\ldots,K}$ sont les vecteurs propres de $\widetilde{\Sigma}$.

Pour avoir une bonne estimation de Σ, on va d'abord utiliser les H simulations de Z choisies parmi le nombre total de simulations K. Nous allons calculer la variance empirique de ces H simulations avec la formule suivante :

$$\widetilde{\Sigma}_1 = \frac{1}{H} Z'_1 M_H Z_1 = \widetilde{P}_1 \widetilde{\lambda}_1 \widetilde{P}'_1. \tag{5.7}$$

Nous allons ensuite utiliser $\widetilde{P}_1$ et les $K-H$ simulations qui restent pour avoir une estimation bien conditionnée de Σ. Pour ce faire, nous allons d'abord estimer Λ avec l'équation suivante :

$$\overline{\Lambda} = \frac{1}{K-H} \mathrm{diag}(\widetilde{P}'_1 X'_2 M_{K-H} X_2 \widetilde{P}_1) \tag{5.8}$$

$$\overline{\Lambda} = \mathrm{diag}(\widetilde{P}'_1 \widetilde{\Sigma}_2 \widetilde{P}_1) \tag{5.9}$$

Alors avec $\overline{\Lambda}$ qui est une meilleure estimation de Λ, on peut avoir grâce à l'équation (5.5) une meilleur estimation (bien conditionnée) de la matrice de variance-covariance Σ :

$$\overline{\Sigma} = \widetilde{P}\overline{\Lambda}\widetilde{P}' \tag{5.10}$$

Avec la méthode d'estimation de la variance ci-dessus nous sommes assurés d'avoir une estimation non singulière de la matrice de variance-covariance. A noter que la précision de cette estimation n'est pas indépendant du choix des H simulations parmi le nombre totale de simulations K. Pour plus de détails voir [1].

Étape 3 : Correction avec les observations $Z_{t+1}^{\mathcal{M}_0}$

Après avoir simulé une échantillon de K cartes de prédiction au temps $t+1$ et calculer la moyenne $\overline{\mu}_{t+1}$ et la matrice de variance-covariance $\overline{\Sigma}_{t+1}$ de ces K cartes. Nous allons maintenant les utilisés comme information *a priori* et les combiner avec les observations de $Z_{t+1}(.)$ aux points de $\mathcal{M}_0$ pour obtenir une prédiction plus précise de $Z_{t+1}(.)$ dans le domaine $\mathcal{D}$. Ceci grâce à la méthode du complément de Schur décrit à la sous section 5.1. Le procédé est le suivant :

Nous notons par $\mathcal{M}_1 = \mathcal{M} - \mathcal{M}_0$ les points du maillage qui n'ont pas été observés. On décompose $\overline{\mu}_{t+1}$ et $\overline{\Sigma}_{t+1}$ de la manière suivante :

$$\overline{\mu}_{t+1} = \begin{pmatrix} \overline{\mu}_{t+1}^{\mathcal{M}_0} \\ \\ \overline{\mu}_{t+1}^{\mathcal{M}_1} \end{pmatrix},$$

et

$$\overline{\Sigma}_{t+1} = \begin{pmatrix} \overline{\Sigma}_{t+1}^{00} & \overline{\Sigma}_{t+1}^{01} \\ \\ \overline{\Sigma}_{t+1}^{10} & \overline{\Sigma}_{t+1}^{11}, \end{pmatrix}$$

Où $\overline{\Sigma}_{t+1}^{00}$ est la matrice de variance-covariance pour les points de $\mathcal{M}_0$, $\overline{\Sigma}_{t+1}^{01}$ est la matrice de covariance entre les points de $\mathcal{M}_0$ et de $\mathcal{M}_1$, $\overline{\Sigma}_{t+1}^{10}$ est la transposé de $\overline{\Sigma}_{t+1}^{01}$, et $\overline{\Sigma}_{t+1}^{11}$ est la matrice de variance-covariance pour les points de $\mathcal{M}_1$.

En utilisant la méthode du complément de Schur nous estimons le vecteur d'espérance et la matrice de variance-covariance correspondante des points non observés.

$$\widehat{\mu}_{t+1}^{\mathcal{M}_1} = \mathbb{E}(Z_{t+1}^{\mathcal{M}_1} | z_{t+1}^{\mathcal{M}_0}) = \overline{\mu}_{t+1}^{\mathcal{M}_1} + (\overline{\Sigma}_{t+1}^{10})(\overline{\Sigma}_{t+1}^{00})^{-1}(z_{t+1}^{\mathcal{M}_0} - \overline{\mu}_{t+1}^{\mathcal{M}_0})$$

$$\begin{aligned}\widehat{\Sigma}_{t+1}^{\mathcal{M}_1} &= \mathrm{Var}(Z_{t+1}^{\mathcal{M}_1}|z_{t+1}^{\mathcal{M}_0}) \\ &= (\overline{\Sigma}_{t+1}/\overline{\Sigma}_{t+1}^{00}) \\ &= \overline{\Sigma}_{t+1}^{11} - \overline{\Sigma}_{t+1}^{10}(\overline{\Sigma}_{t+1}^{00})^{-1}\overline{\Sigma}_{t+1}^{01}.\end{aligned}$$

Nous avons :

$$[Z_{t+1}^{\mathcal{M}_1}|z_{t+1}^{\mathcal{M}_0}] \sim \mathcal{N}(\widehat{\mu}_{t+1}^{\mathcal{M}_1}, \widehat{\Sigma}_{t+1}^{\mathcal{M}_1}). \tag{5.11}$$

Ainsi de suite pour tout $t \geq 2$, on revient à l'étape 2 en faisant $t = t + 1$ et en faisant K simulations de l'équation (5.11) . Puis nous utilisons les nouvelles données pour obtenir la prédiction du processus à différents temps donnés.

5.3 Application numérique et résultats

Pour illustrer la méthode, nous reprenons le modèle de convection-diffusion du chapitre 3 [31, 37, 116].

Soit $\mathcal{D}$ un domaine carré $(a, b)^2$ dans $\mathbb{R}^2$ et $(0, T)$ est un intervalle de temps. Pour le terme source donné $\mathcal{S} = \mathcal{S}(x, y, t) \in L^2(\Omega \times (0, T))$, on considère l'équation suivante en deux dimensions $(2D)$ de convection diffusion qui dépend du temps avec des conditions au bord de Dirichlet homogène.

$$\begin{cases} L(u) &:= \frac{\partial u}{\partial t} - \nu\Delta u + a_1\frac{\partial u}{\partial x} + a_2\frac{\partial u}{\partial y} = \mathcal{S}, \text{in } \Omega \times (0, T] \\ u(x, y, t) &= 0 \text{ on } \partial\Omega \times (0, T] \\ u(x, y, 0) &= u_0(x, y) \text{ in } \Omega \end{cases} \tag{5.12}$$

où u est l'état du système dynamique, a_1 et a_2 sont des constantes qui représentent les coefficients de convection, et ν est une constante positive appelée le coefficient de diffusion. Comme dans [116], on suppose que le problème ci-dessous admet une solution lisse et unique.
Cette équation de convection et diffusion s'applique dans beaucoup de domaine de la mécanique des fluides notamment dans les problèmes de pollution des eaux souterraines, elle est utilisée aussi pour modéliser l'écoulement du pétrole ou d'autres fluides à partir du réservoir au puits.

A partir de l'équation aux dérivées partielles (EDP) (5.12), il est possible de déterminer l'état u du système dans le domaine $(a, b)^2$ pour chaque temps t donné.

Par exemple Nous considérons que nous avons un problème de pollution d'eaux souterraines

et que $\mathcal{S}$ est la source qui continuellement apporte une quantité de pollution dans le domaine $(a,b)^2$. Nous choisissons que le terme source $\mathcal{S}$ s'écrit sous cette forme :

$$\mathcal{S} = \mathcal{S}(x,y,t) = (\frac{5}{4} - ((x-0.8)^2 + (y-\frac{1}{2})^2)), \tag{5.13}$$

Avec l'équation (5.13) nous considérons que la quantité de pollution apportée par la source $\mathcal{S}$ ne varie pas en fonction du temps et est plus grande sur certaines parties de $(a,b)^2$ que dans d'autres. L pollution qu'apporte continuellement la source $\mathcal{S}$ évoluera dans le domaine $(a,b)^2$ en fonction du temps, ceci dû au phénomène de convection et diffusion. La direction et la vitesse de déplacement de la pollution dépendra des valeurs de a_1 et a_2, quand à la diffusion elle dépendra du coefficient ν.

Notre but est de montrer l'efficacité de notre méthode en obtenant une estimation précise de l'état u de la pollution dans $(a,b)^2$ à chaque temps t donné.

Pour ce faire nous allons utilisé une grille 24×24, nommé $\mathcal{M}$, du domaine $(a,b)^2$ pour calculer les valeurs $u(x,y,t)$, $(x,y) \in \mathcal{M}$. Pour les points du maillage $\mathcal{M}$ illustrés dans la Figure 5.1, on a résolu l'EDP (5.12) avec : $(a,b)^2 = (0,1)^2$, $a_1 = -0.09$, $a_2 = -0.09$, $\nu = 0.001$ et avec la source $\mathcal{S}$ donnée ci-dessus (5.13) pour obtenir les valeurs simulées de u pour $t = 1, \ldots, 9$ (voir Figure 5.2). Les valeurs de u obtenues avec cette simulation sont considérées comme les vraies valeurs pour les différents temps $t = 1, \ldots, 9$.

Donc pour tester l'efficacité de la méthode qui a été mise en place, nous avons séparé les points du maillage $\mathcal{M}$ en deux parties (voir Figure 5.1) :

- Les sites d'observations correspondant à $\mathcal{M}_0$
- Les sites où u est inconnu pour $t = 1, \ldots, 9$.

La boîte noire notée par f est définie par le même EDP (5.12) avec aussi $a_1 = -0.09$, $a_2 = -0.09$, $\nu = 0.001$.

Le procédé pour prédire les valeurs de u aux sites non observés est donné comme suit :

- Étape 1 : Au tout début à $t = 1$, nous récupérons les valeurs de u aux sites d'observations pour faire du krigeage Bayésien en prenant des *a priori* impropres pour les paramètres de krigeage afin obtenir la distribution spatiale de $u(x,y,1)$, $(x,y) \in \mathcal{M}$.

- Étape 2 : Nous tire une échantillon de K cartes $U_1^{[i]} = \{u^{[i]}(x,y,1),\ (x,y) \in \mathcal{M}\}$ de la distribution de $u(x,y,1)$ obtenue avec le krigeage Bayésien. Chacune de ces K cartes de

prédiction est utilisé comme entrée pour la boîte noire, notéé par f, pour obtenir la carte correspondante au temps $t = 2$

$$\widehat{u}^{[i]}(x, y, 2) = f(u^{[i]}(x, y, 1)). \tag{5.14}$$

Puis nous calculons la moyenne et la matrice de variance-covariance de ces cartes de prédiction obtenues grâce à la boîte noire à l'instant $t = 2$. La variance des cartes simulées que nous nommons par $\overline{\Sigma}_2$ sera calculée à partir de la méthode d'estimation de la matrice de variance-covariance de [1] et qui a été expliquée à la sous section ci-dessus. Le vecteur de moyenne des cartes $\overline{u}_2$ est lui calculé en utilisant la moyenne empirique :

$$\overline{u}_2 = \sum_{i}^{K} \frac{1}{K} \widehat{u}^{[i]}(x, y, 2).$$

– Étape 3 : La moyenne $\overline{u}_2$ et la matrice de variance-covariance $\overline{\Sigma}_2$ des cartes obtenues avec la boîte noire seront utilisées comme information *a priori* afin d'avoir une prédiction plus précise de $u(x, y, 2)$. Pour ce faire nous allons utiliser la méthode du complément de Schur présentée ci-dessus qui va nous permettre de combiner l'information *a priori* que nous a fournit la boîte noire avec les nouvelles observations $u(x, y, 2)$ obtenues aux points du plan d'expériences $\mathcal{M}_0$.

Pour appliquer la méthode du complément de Schur, on réécrit $\overline{u}_2$ et $\overline{\Sigma}_2$ sous ces formes :

$$\overline{\mu}_2 = \begin{pmatrix} \overline{\mu}_2^{\mathcal{M}_0} \\ \\ \overline{\mu}_2^{\mathcal{M}_1} \end{pmatrix},$$

et

$$\overline{\Sigma}_2 = \begin{pmatrix} \overline{\Sigma}_2^{00} & \overline{\Sigma}_2^{01} \\ \\ \overline{\Sigma}_2^{10} & \overline{\Sigma}_2^{11} \end{pmatrix}.$$

En appliquant le complément de Schur on a

$$\widehat{u}_2^{\mathcal{M}_1} = \mathbb{E}(u_2^{\mathcal{M}_1} | u_2^{\mathcal{M}_0}) = \overline{\mu}_2^{\mathcal{M}_1} + (\overline{\Sigma}_2^{10})(\overline{\Sigma}_2^{00})^{-1}(u_2^{\mathcal{M}_0} - \overline{\mu}_2^{\mathcal{M}_0})$$

$$\widehat{\Sigma}_2^{\mathcal{M}_1} = \mathrm{Var}(u_2^{\mathcal{M}_1}|u_2^{\mathcal{M}_0}) = (\overline{\Sigma}_2/\overline{\Sigma}_2^{00})$$
$$= \overline{\Sigma}_2^{11} - \overline{\Sigma}_2^{10}(\overline{\Sigma}_2^{00})^{-1}\overline{\Sigma}_2^{01},$$

où $u_2^{\mathcal{M}_0}$ sont les observations de $u(x, y, 2)$ aux points de $\mathcal{M}_0$.

Nous avons :

$$[u_2^{\mathcal{M}_1}|u_2^{\mathcal{M}_0}] \sim \mathcal{N}(\widehat{u}_2^{\mathcal{M}_1}, \widehat{\Sigma}_2^{\mathcal{M}_1}). \tag{5.15}$$

La carte d'estimation par notre méthode de la pollution au temps $t = 2$ est donc $\widehat{u}_2(x, y) = \begin{pmatrix} u_2^{\mathcal{M}_0} \\ \\ \widehat{u}_2^{\mathcal{M}_1} \end{pmatrix}$, $(x, y) \in \mathcal{M}$.

Ainsi de suite, on répète l'étape 2 et 3 pour obtenir les cartes d'estimation de la pollution au temps $t = 3, \ldots, 9$. À l'étape 2, on tire K cartes de la distribution de u_t obtenu à l'instant t et à l'étape 3 on utilise les nouvelles données disponibles à l'instant $t+1$ pour obtenir la distribution de u_{t+1} et une estimation de la moyenne $\widehat{u}_{t+1}(x, y), (x, y) \in \mathcal{M}$.

Sachant qu'en général une EDP est utilisée pour prédire l'état d'un phénomène dans un domaine $\mathcal{D}$ au fur du temps, nous allons comparer la prédiction faite par notre méthode et celle faite par la boîte noire qui est ici représentée par l'EDP (5.12). L'estimation faite avec la boîte noire est $\widehat{B}_t(x, y) = \overline{u}_t$. Les erreurs de prédiction de la boîte noire et de notre méthode sont respectivement données par $|\widehat{B}_{t+1}(x, y) - u(x, y, t+1)|$ et $|\widehat{u}_{t+1}(x, y) - u(x, y, t+1)|$. Les résultats obtenus (voir Figure 5.3) montrent que la prédiction faite par notre méthode améliore celle faite par la boîte noire au même temps t donné.

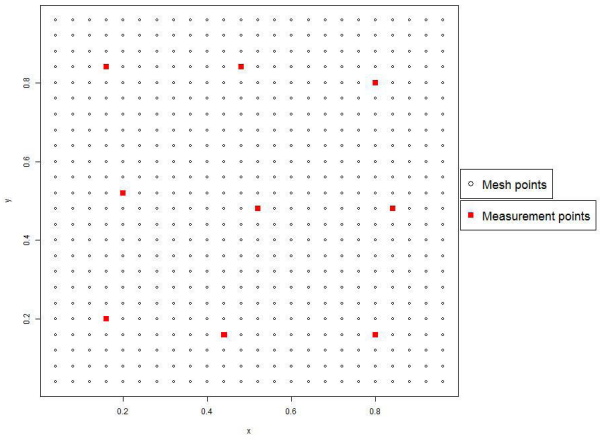

FIGURE 5.1 – Plans minimax sur une grille de 576 points

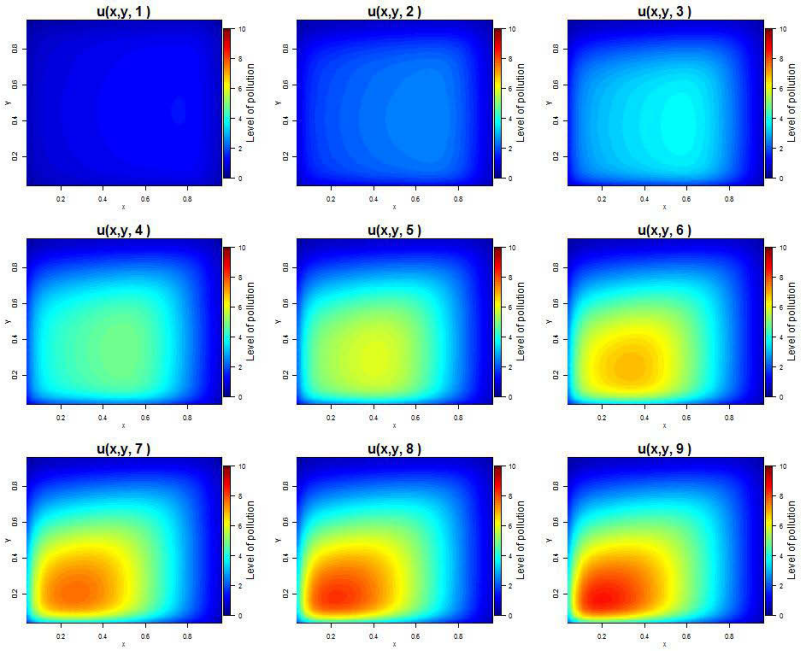

FIGURE 5.2 – Evolution of real values of pollution to $t = 1$ at $t = 9$

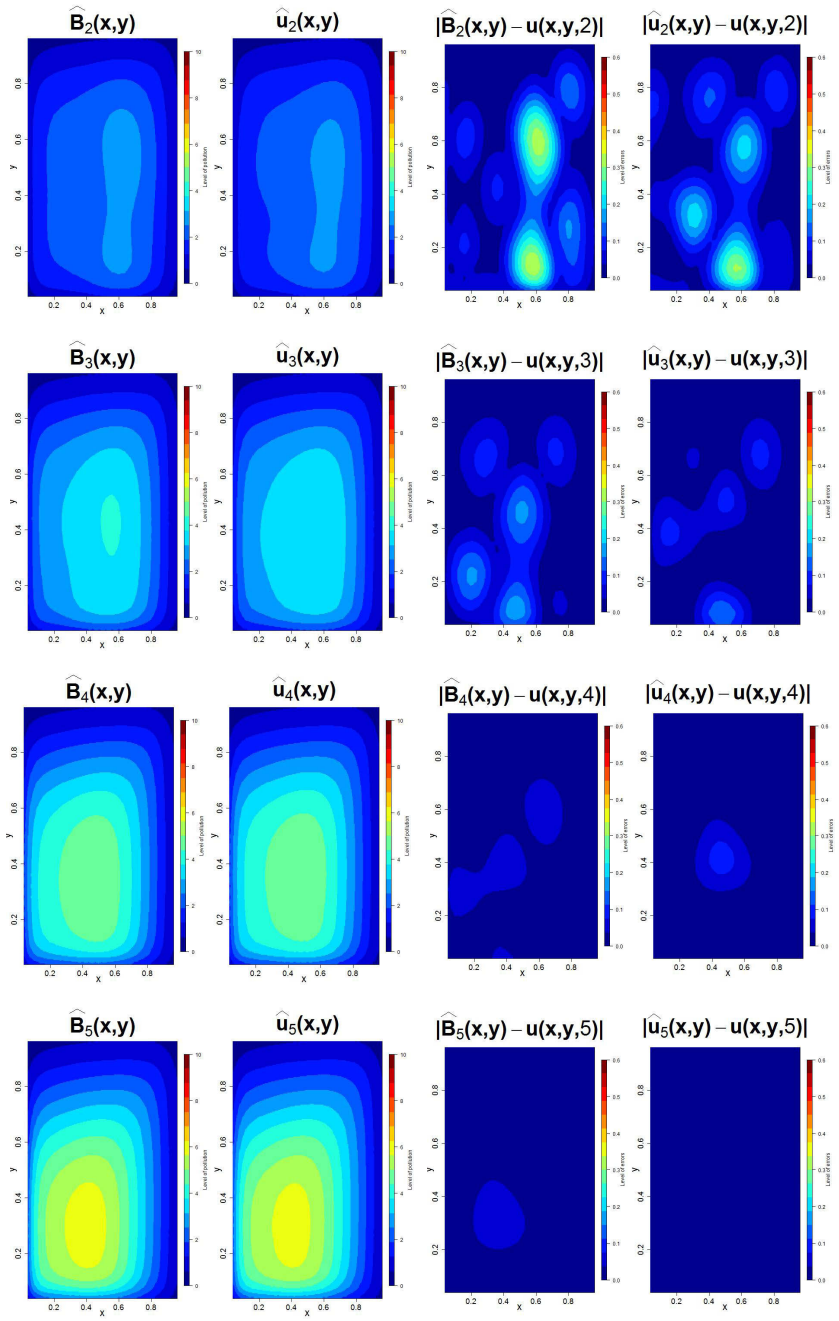

FIGURE 5.3 – $\widehat{u}_t(x,y)$: Valeurs prédites par notre méthode, $\widehat{B}_t(x,y)$: Valeurs prédites par la boîte noire, $|\widehat{u}_t(x,y) - u(x,y,t)|$ erreurs de prédictions de notre méthode, $|\widehat{B}_t(x,y) - u(x,y,t)|$ erreurs de prédiction de la boîte noire, $t = 2, ..., 5$.

Chapitre 6

Plans d'expériences

Dans les parties précédentes, le plan d'expériences où était effectué les observations du processus gaussien $Z_t(.)$ localisé dans le domaine d'étude $\mathcal{D} \subset \mathbb{R}^d, d = \{1, 2, 3\}$, était supposé connu. A présent, la question du choix de ce plan se pose. L'idée est de choisir un plan d'expériences optimal qui permettra d'obtenir le maximum d'information sur le comportement de $Z_t(.)$ dans le domaine $\mathcal{D}$.
Nous allons d'abord dans une première partie faire une revue bibliographique des plans d'expériences, ensuite dans une deuxième partie nous allons présenter la technique que nous avons mise en place pour obtenir les plans d'expériences qui sont les mieux adaptés à notre algorithme statistique.

6.1 Revue bibliographique des plans d'expériences

Historiquement la théorie de l'optimalité en plans d'expériences a été développée dans le cadre du modèle linéaire : régression linéaire et analyse de la variance. Les travaux précurseurs pour l'optimisation de plans d'expériences sont l'œuvre de Kiefer [68, 69, 70], Kiefer et Wolfowitz [71] et de Federov [39]. Ces auteurs à travers leurs travaux ont mise en place des critères d'optimalité qui permettent de construire des plans d'expériences pour modèles linéaires. En général l'optimisation de l'expérimentation physique d'un phénomène nécessite plusieurs réplications des observations aux sites du plan d'expériences pour tenir compte de la variabilité du phénomène. Parmi les critères d'optimalité pour plans d'expériences pour modèle linéaire nous pouvons citer les critères de D-optimalité, G-optimalité, A-optimalité, E-optimalité et J-optimalité. Pour une revue récente et détaillé des méthodes d'optimisation des plans d'expériences pour modèle linéaire se référer à l'article de Jean-Pierre Gauchi [44].

Cependant pour des raisons de coût de nombreux phénomènes scientifiques sont étudiés, non plus via l'expérimentation physique, mais à l'aide de modèles numériques. Malgré les progrès des outils informatiques, l'étude numérique de certains phénomènes physiques complexes reste très coûteuse en temps de calcul. Il est donc nécessaire de faire appel à des méthodes de planification d'expériences, d'où la construction des plans d'expériences numériques.

Plans d'expériences numériques

D'après Jourdan [62], l'expérimentation numérique tel que les surfaces de réponse (voir [32]) consiste à fixer un vecteur de valeur pour les variables d'entrée du simulateur puis à récolter la ou les réponses de celui-ci. Il est ensuite étudié le comportement de cette réponse en fonction des variations des variables d'entrée. L'objectif de l'étude n'est pas la réponse du simulateur mais le phénomène simulé. En effet en utilisant une méthode statistique représentant au mieux cette réponse il est possible d'estimer le comportement du phénomène simulé. Ce pendant pour des raisons de coût financier et technique, le nombre d'expériences (sites d'observation du phénomène) nécessaires pour estimer au mieux le comportement d'un phénomène semble difficile à réaliser. Donc les deux questions qui se posent sont :

- Comment choisir au mieux les valeurs des variables d'entrée de façon à récolter un maximum d'informations sur le comportement du phénomène d'intérêt en un minimum de simulations ?
- Quel modèle statistique est approprié aux réponses du simulateur.

Certaines particularités des expériences numériques par rapport aux expérimentations physiques doivent être prises en compte pour le traitement de l'information :

- Les expériences sont deterministes, c'est à dire que deux simulations avec deux jeux de variables d'entrée identiques donnent la même réponse
- Les variables d'entrée sont très nombreuses. En effet, aux variables liées au phénomènes physique, viennent s'ajouter celles dues au modèle numérique (par exemple une taille de maillage).

Le modèle statistique choisit est alors utilisé pour prédire la réponse du simulateur en des sites non testés par des simulations. Sachant que nous avons fait le choix d'utiliser le krigeage présenté

à la sous section 2.4.1 comme modèle statistique spatiale, nous allons faire une présentation des plans d'expériences pour le krigeage. Initialement les plans développés pour le krigeage étaient des plans optimaux pour certains critères statistiques, on peut citer :

- L'erreur quadratique moyenne intégrée (IMSE) [111] qui va sélectionner le plan qui minimise l'écart quadratique moyen intégré entre la réponse du simulateur et sa prédiction.
- le critère d'entropie qui permet de mesurer la quantité d'information fournie par la simulation [114] et [29].

Par la suite sont apparus les plans à marges uniformes tels que les hypercubes latins ou les plans orthogonaux [119] pour lesquels les points doivent être proche de la distribution uniforme en projection sur un axe ou une face du domaine. Les suites de faible discrépance [101] ont aussi fait leur apparition et ils stipules que la distribution empirique des points doit être proche de la distribution uniforme.

Ces plans sont construits indépendamment du modèle notamment de la structure de covariance choisie et leurs points représentent au mieux le domaine expérimental , d'où leur nom de «space filling designs». Par contre ils ne répondent pas nécessairement aux critères statistiques cités ci-dessus.

Les plans qui sont donc actuellement utilisés en krigeage sont ceux qui, à l'intérieur d'une classe de « space filling designs», optimisent un critère statistique. Cette technique permet ainsi :

- de s'assurer d'une bonne répartition spatiale des points du plan,
- d'avoir un plan optimal pour le modèle.

En pratique parmi les plans qui appartiennent à la classe de « space filling designs» les plus utilisés sont les hypercubes latins. Avec les plans en hypercube latin chaque arrête du domaine expérimental est divisé en n segments de même longueur de façon à obtenir un maillage du domaine. les plans hypercube latins sélectionnent alors n points parmi les n^d points de la grille, où d est la dimension du domaine, de façon à ce que les n niveaux des variables d'entrée soient testés une fois par les simulations. Les hypercubes latins présentent beaucoup d'avantages.

- Ils sont simple à construire. En effet, chaque colonne d'une hypercube est une permutation de $1, \ldots, n$.
- Les points sont uniformément distribués sur chaque axe du domaine.

La distribution uniforme sur chaque axe n'assure pas l'uniformité sur le domaine expérimental. Cependant, pour n fixé, ils existent $(n!)^d$ hypercubes latins possibles. Il est donc possible de sélectionner le plan optimisant un critère d'uniformité, par exemple de discrépance, ou bien un critère statistique, par exemple l'IMSE, l'entropie etc. Park [104] propose un algorithme d'échange pour

déterminer un hypercube optimal pour un critère donné. Le principe de l'algorithme d'échange est simple : pour un point donné du plan, on remplace ce point avec un de l'ensemble candidat en le choisissant soit de manière aléatoire ou en faisant une recherche exhaustive, puis on examine si cet échange permet d'améliorer le critère. Si l'amélioration est avérée, le nouveau point est adopté dans le plan et l'ancien point est ajouté à l'ensemble des candidats. Ce processus continue jusqu'à ce qu'il n'y ait plus d'échange possible ou que le nombre d'itérations préalablement fixé soit atteint. Ainsi la convergence de l'algorithme est assurée, mais aucune garantie n'est donnée d'avoir généré le meilleur plan possible. L'algorithme peut considérer uniquement les échanges entre les points candidats et les plus proches voisins, ce qui permet de diminuer le temps de calcul tout particulièrement quand l'ensemble candidat est grand. Il paraît clair qu'en grande dimension ces plans ne pourront être de bonne qualité vis-à-vis du remplissage de l'espace. En effet, le procédé calculatoire qui utilise des points aléatoires à la place d'une grille régulière ne permet pas d'écarter au mieux les points. Une extension de cet algorithme consiste à fixer des points du plan, c'est-à-dire des points qui ne peuvent être échangés. Ceci est nécessaire dans une approche séquentielle i.e. si l'on veut accumuler une suite de plans en augmentant la taille et en conservant les points déjà générés (voir [42, 62] plus de détails sur l'algorithme d'échange). D'autres algorithmes tel que celui de recuit simulé [10, 96] permet aussi de construire des plans d'expériences optimaux. Le principe de base de ces algorithmes consiste à déterminer un sous-ensemble de points, à partir d'un grand ensemble de candidats, qui améliore un critère géométrique de remplissage de l'espace. Pour le modèle de krigeage Collombier et Jourdan [63] ont montré qu'un hypercube optimal est robuste aux variations du paramètre de corrélation.

Parmi les plans expériences qui font partie de la classe des hypercubes latins nous pouvons citer les plans minimax ou maximin qui sont fondés sur des critères de distance établis par Jonhson et al. [61]. Un plan minimax est défini comme suit :

Définition 6.1.1. *Un plan d'expérience $P = (s_1, \ldots, s_n) \subset \mathcal{D}^n$ est dit minimax s'il minimise*

$$h_P = \sup_{s \in D} \min_{1 \leq i \leq n} \|s - s_i\|_2.$$

Un plan minimax assure qu'en tout point du domaine $\mathcal{D}$, on sera jamais trop loin d'un point du plan d'expérience P. Le principe du critère minimax est de mesurer la distance maximale entre un point du domaine expérimental n'appartenant pas au plan, qui varie, et les points du plan d'expériences. Un plan maximin est défini comme suit :

Définition 6.1.2. *Un plan d'expérience $P = (s_1, \ldots, s_n) \subset \mathcal{D}^n$ est dit maximin s'il maximise*

$$\delta_P = \min_{1 \leq i,j \leq n} \|s_i - s_j\|_2.$$

Un plan maximin espace de manière optimale ses points afin d'éviter les réplications.

Parmi plusieurs plans maximin, sont privilégiés ceux de plus petit indice c'est à dire ceux qui ont un nombre minimal de paires réalisant la distance δ_P. D'aprés [10], puisque P est un espace compact et les fonctions $P \to h_P$ et $P \to \delta_P$ sont continues, l'existence des plans minimax et maximim est garantie. Il n'y a cependant pas toujours unicité de la solution.
Afin de calculer les critères, on prend généralement pour les points de $\mathcal{D}$ une grille régulière. Aussi, en grande dimension il n'est plus envisageable de considérer ce critère sans prendre pour l'ensemble $\mathcal{D}$ un plan aléatoire.
La figure 6.1 donne l'exemple d'un plan minimax construit sur une grille régulière de 576 points, à l'aide de la fonction **cover.design** de *R*.

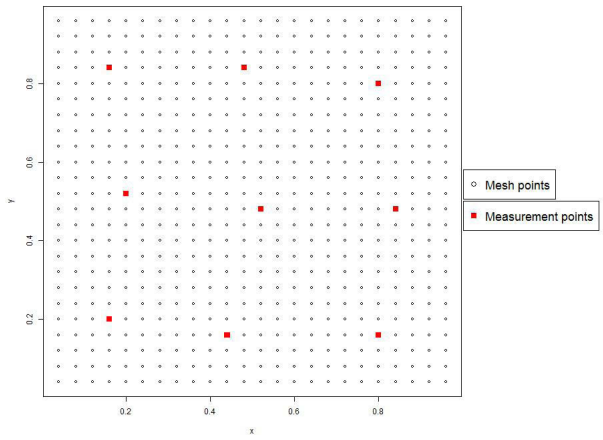

FIGURE 6.1 – Plans minimax sur une grille de 576 points

En ce qui nous concerne, nous avons construit un critère d'optimalité qui permet de construire des plans d'expériences numériques adaptés à nos algorithmes de prédiction spatio-temporelle. Ce critère conçu permet de planifier les simulations de sorte que la quantification des incertitudes sur la réponse soit la meilleure possible.

6.2 Plans d'expériences numériques adaptés à nos algorithmes statistiques

Dans cette section nous allons dans un premier temps décrire en détail le critère d'optimalité que nous avons mise en place pour construire des plans d'expériences numériques adaptés à nos deux méthodes statistiques de prédiction spatio-temporelle (krigeage Bayésien avec boîte noire mal spécifiée et modèle semi-empirique avec boîte noire). Puis nous allons donner la méthode de construction de ces plans d'expériences numériques en utilisant ce critère d'optimalité.

Nous rappelons tous d'abord que ce soit pour la méthode du krigeage Bayésien avec boîte noire mal spécifiée ou pour le modèle semi-empirique avec boîte noire, nous considérons que nous avons un maillage $\mathcal{M}$ du domaine $\mathcal{D}$. L'objectif étant d'estimer la valeur de la variable d'intérêt $Z_t(s)$ en tout point s appartenant $\mathcal{M}$, pour chaque temps t donné, à partir de l'observation de cette variable en quelques points de $\mathcal{M}$. Une fois la prédiction spatiale du processus gaussien $Z_t(.)$ obtenue, nous utilisons la boîte noire pour obtenir une prédiction temporelle de ce processus au temps $t+1$.

Nous rappelons ci-dessous la procédure de prédiction temporelle du processus par la boîte noire :

Soit $\mathcal{M}_0^t = \{(s_1, \ldots, s_n), (s_i)_{1<i<n} \in \mathcal{M}\}$ le plan d'expérience à l'instant t, $Z_t^{\mathcal{M}_0^t} = (Z_t(s_1), \ldots, Z_t(s_n))$ les observations de $Z_t(.)$ correspondantes et

$$[Z_t^{\mathcal{M}} | z_t^{\mathcal{M}_0^t}] \sim \mathcal{N}(\mu_t^{\mathcal{M}}, \Sigma_t), \tag{6.1}$$

la distribution prédictive de $Z_t(s)$, $\forall s \in \mathcal{M}$. Une prévision temporelle de cette distribution prédictive, au temps $t+1$, s'obtient en utilisant la boîte noire notée par f :

$$[Z_{t+1}^{\mathcal{M}} \mid z_t^{\mathcal{M}_0^t}] = [f(Z_t^{\mathcal{M}}) \mid z_t^{\mathcal{M}_0^t}]. \tag{6.2}$$

Étant donné que f n'est pas connue explicitement en général, il n'est pas possible d'obtenir une expression analytique de la loi de $[Z_{t+1}^{\mathcal{M}} \mid z_t^{\mathcal{M}_0}]$. Nous allons donc approcher la loi de 6.2 par les méthodes de **Monte Carlo** en procédant comme suit :

1. On fait K simulations de la prédiction spatiale 6.1 obtenue à l'instant t, ce qui va donner K cartes de prédiction $z_t^{\mathcal{M}[i]}$, $i = 1, \ldots, K$

2. Chacune des K cartes de prédiction simulées est considérée comme variable d'entrée de la boîte noire f, qui va fournir comme sortie la carte correspondante au temps $t+1$

$$\widehat{z}_{t+1}^{\mathcal{M}[i]} = f(z_t^{\mathcal{M}[i]}). \tag{6.3}$$

Puis nous calculons la moyenne et la matrice de variance-covariance des K cartes obtenues à l'instant $t+1$. La moyenne $\overline{\mu}_{t+1}$ des K cartes est calculée de manière empirique :

$$\overline{\mu}_{t+1} = \frac{1}{K}\Sigma_{i=1}^{K}\widehat{z}_{t+1}^{\mathcal{M}[i]}.$$

Quand à la matrice de variance-covariance $\overline{\Sigma}_{t+1}$, elle est calculée par exemple en utilisant la méthode de Karim M. Abadir et al [1] expliquée à la section 5.1 du chapitre 5.

La loi approchée de $[Z_{t+1}^{\mathcal{M}} \mid z_t^{\mathcal{M}_0^t}]$ est donc :

$$[Z_{t+1}^{\mathcal{M}} \mid z_t^{\mathcal{M}_0^t}] \sim \mathcal{N}(\overline{\mu}_{t+1}, \overline{\Sigma}_{t+1})$$

Nous allons maintenant détailler dans la partie suivante le procédé d'établissement du critère d'optimalité pour la construction du plan d'expérience $\mathcal{M}_0^{t+1}$ optimal, en utilisant la moyenne $\overline{\mu}_{t+1}$ et la matrice de variance-covariance $\overline{\Sigma}_{t+1}$ obtenue grâce à la boîte noire.

6.2.1 Critère d'optimalité

Soit $\mathcal{M}_0^{t+1} = (s_1, \ldots, s_n)$ un plan d'expérience donné à l'instant $t+1$, nous définissons le critère théorique par la formule suivante :

$$\max_{s\in\mathcal{M}}\{\sqrt{\mathrm{Var}(Z_{t+1}(s)|z_t^{\mathcal{M}_0^t}, z_{t+1}^{\mathcal{M}_0^{t+1}})} + \rho\cdot\mathbb{E}(Z_{t+1}(s)|z_t^{\mathcal{M}_0^t}, z_{t+1}^{\mathcal{M}_0^{t+1}})\}, \tag{6.4}$$

où ρ est une constante qui appartient à $\mathbb{R}$.

L'idée du critère du maximum est de repérer la zone où l'incertitude sur l'estimation du phénomène est la plus grande, tout en tenant compte de la valeur espérée du phénomène à travers le paramètre ρ. Par exemple dans les problèmes de pollution en prenant $\rho = 1$, le critère du maximum permettrait de repérer la zone, où la pollution ajouté à l'incertitude sur son estimation est la plus grande. La valeur fournit par le critère permet d'évaluer de la qualité du plan $\mathcal{M}_0^{t+1} = (s_1, \ldots, s_n)$. Cependant, les valeurs de $z_{t+1}^{\mathcal{M}_0^{t+1}}$ apparaissant dans le critère (6.4) ne sont pas encore observées.

Le terme $\text{Var}(Z_{t+1}(s)|z_t^{\mathcal{M}_0^t}, z_{t+1}^{\mathcal{M}_0^{t+1}})$ qui apparaît dans l'équation (6.4) ne dépend que de $z_t^{\mathcal{M}_0^t}$ et du plan choisit $\mathcal{M}_0^{t+1}$, mais pas de $z_{t+1}^{\mathcal{M}_0^{t+1}}$. Donc, nous pouvons écrire :

$$\text{Var}(Z_{t+1}(s)|z_t^{\mathcal{M}_0^t}, z_{t+1}^{\mathcal{M}_0^{t+1}}) = \text{Var}(Z_{t+1}(s)|z_t^{\mathcal{M}_0^t}, \mathcal{M}_0^{t+1}).$$

Sachant qu'aux points observés $\mathcal{M}_0^{t+1}$ à $t+1$, la variance conditionnelle est nulle, nous allons seulement donner l'estimation de la variance aux points de $\mathcal{M}_1^{t+1} = \mathcal{M} \setminus \mathcal{M}_0^{t+1}$. Pour ce faire nous allons utiliser la matrice de variance-covariance $\overline{\Sigma}_{t+1}$ obtenue grâce à la boîte noire. Nous réécrivons $\overline{\Sigma}_{t+1}$ sous la forme de matrice en bloc :

$$\overline{\Sigma}_{t+1} = \begin{pmatrix} \overline{\Sigma}_{t+1}^{00} & \overline{\Sigma}_{t+1}^{01} \\ \\ \overline{\Sigma}_{t+1}^{10} & \overline{\Sigma}_{t+1}^{11}, \end{pmatrix}$$

où,

- $\overline{\Sigma}_{t+1}^{00}$ est la matrice de variance-covariance pour les points de $\mathcal{M}_0^{t+1}$
- $\overline{\Sigma}_{t+1}^{01}$ et $\overline{\Sigma}_{t+1}^{01}$ sont les matrices de covariances entre les points de $\mathcal{M}_1^{t+1}$ et de $\mathcal{M}_0^{t+1}$. La matrice $\overline{\Sigma}_{t+1}^{01}$ est la transposée $(\overline{\Sigma}_{t+1}^{01})$
- $\overline{\Sigma}_{t+1}^{11}$ est la matrice de variance-covariance des points de $\mathcal{M}_{t+1}^1$.

En utilisant le complément de Schur (voir Section 5.1), nous avons :

$$\begin{aligned} \Sigma_{t+1}^{\mathcal{M}_1^{t+1}} &= (\overline{\Sigma}_{t+1}/\overline{\Sigma}_{t+1}^{00}) \\ &= \overline{\Sigma}_{t+1}^{11} - \overline{\Sigma}_{t+1}^{10}(\overline{\Sigma}_{t+1}^{00})^{-1}\overline{\Sigma}_{t+1}^{01}. \end{aligned} \tag{6.5}$$

Les valeurs $\text{Var}(Z_{t+1}(s)|z_t^{\mathcal{M}_0^t}, \mathcal{M}_0^{t+1})$, $s \in \mathcal{M}_1^{t+1}$ correspondent donc aux éléments diagonaux de $\Sigma_{t+1}^{\mathcal{M}_1^{t+1}}$.

Le terme d'espérance $\mathbb{E}(Z_{t+1}(s)|z_t^{\mathcal{M}_0^t}, z_{t+1}^{\mathcal{M}_0^{t+1}})$ quand à lui sera remplacée par son approximation

$$\overline{\mu}_{t+1}(s) = \frac{1}{K}\Sigma_{i=1}^{K}\widehat{z}_{t+1}^{[i]}(s)$$

où $\widehat{z}_{t+1}^{[i]}(s)$ est obtenu à partir de (6.3).

On peut donc remplacer le critèe théorique (6.4) par le critère C_ρ défini par :

$$C_\rho(\mathcal{M}_0^{t+1}) = \max_{s\in\mathcal{M}}\{\sqrt{\Sigma_{t+1}(s,s)} + \rho\cdot\overline{\mu}_{t+1}(s). \tag{6.6}$$

Comme dit auparavant la valeur du critère C_ρ permet d'évaluer la qualité du plan d'expérience $\mathcal{M}_0^{t+1}$. Le plan qui minimise le plus la valeur du critère C_ρ est considéré comme plan optimal et est appelé plan minimax.

Remarque : La valeur $\overline{\mu}_{t+1}(s)$ correspond également à

$$\overline{\mu}_{t+1}(s) = \mathbb{E}(Z_{t+1}(s)|z_t^{\mathcal{M}_0^t}, \overline{\mu}_{t+1}^{\mathcal{M}_0^{t+1}}).$$

En effet en utilisant la formule du complément de Schur nous avons :

$$\mathbb{E}(Z_{t+1}^{\mathcal{M}_1^{t+1}}|z_t^{\mathcal{M}_0^t}, \overline{\mu}_{t+1}^{\mathcal{M}_0^{t+1}}) = \overline{\mu}_{t+1}^{\mathcal{M}_1^{t+1}} + (\overline{\Sigma}_{t+1}^{10})(\overline{\Sigma}_{t+1}^{00})^{-1}(\overline{\mu}_{t+1}^{\mathcal{M}_0^{t+1}} - \overline{\mu}_{t+1}^{\mathcal{M}_0^{t+1}}) = \overline{\mu}_{t+1}^{\mathcal{M}_1^{t+1}}.$$

6.2.2 Construction des plans d'expériences minimax

A l'instant $t+1$, nous cherchons un plan minimax $\mathcal{M}_0^{t+1}$ basé sur $\overline{\mu}_{t+1}$ et $\overline{\Sigma}_{t+1}$, l'information que nous apporte la boîte noire. Selon le phénomène à étudier, nous pouvons être confrontés à deux cas de figure pour la construction du plan d'expérience minimax $\mathcal{M}_0^{t+1}$.

1. La première est d'ajouter un point d'observation de plus au plan minimax $\mathcal{M}_0^t$ utilisé à t. Prenons par exemple le cas où, à l'instant t, nous avions des capteurs de pollution fixes placés dans un domaine donné et qu'a l'instant $t+1$ nous désirons placé un capteur supplémentaire dans un site de la ville, où le degré de pollution et l'incertitude sur l'estimation de cette dernière est la plus grande.
2. La deuxième est de construire complètement un nouveau plan minimax.
   Prenons par exemple le cas des sondages dans les nappes phréatiques pour estimer le niveau de pollution de la nappe, ou bien la densité d'une espèce toxique. Les résultats obtenus avec les n sondages réalisés à l'instant t vont être utilisés pour réaliser n nouveaux sondages à l'instant $t+1$.

6.2.2.1 Rajout d'un point d'observation au plan minimax $\mathcal{M}_0^t$

Ce point d'observation supplémentaire sera pris parmi les points du maillage $\mathcal{M}_1^t = \mathcal{M} \setminus \mathcal{M}_0^t$. Le procédé suivant qui est une extension de l'algorithme d'échange permet de construire le plan minimax $\mathcal{M}_0^{t+1}$ en rajoutant un point au plan minimax $\mathcal{M}_0^t$. Nous notons par $Maxiter$ le nombre maximale d'itération.

Initialisation ;

$i = 0$;

On choisit le point $s^{(0)}$ de $\mathcal{M}_1^t$ tel que :

$$s^{(0)} = \tilde{s} = \underset{s \in \mathcal{M}_1^t}{\operatorname{argmax}} \{\sqrt{\overline{\Sigma}_{t+1}(s,s)} + \rho \times \overline{\mu}_{t+1}(s)\}.$$

On définit par $P^{(0)} = \mathcal{M}_0^t \cup \{\tilde{s}\}$ le plan obtenu en ajoutant le point $\tilde{s}$ au plan minimax $\mathcal{M}_0^t$;

tant que $i \leq Maxtier$ **faire**

On tire au hasard un point s_i de $\tilde{\mathcal{M}}_1^{t+1} = \mathcal{M} \setminus (\mathcal{M}_0^t \cup \{s_0, \ldots, s_{i-1}\})$;

On ajoute ce point s_i au plan existant $\mathcal{M}_0^t$ pour obtenir le plan $\mathcal{M}_*^{t+1} = \mathcal{M}_0^t \cup \{s_i\}$;

si $C_\rho(\mathcal{M}_*^{t+1}) < C_\rho(P^{(i)})$ **alors**

$s^{(i+1)} = s_i$

sinon

$s^{(i+1)} = \tilde{s}$

fin

$P^{(i+1)} = P^{(i)} \cup \{s^{(i+1)}\}$;

$i = i + 1$;

fin

A la sortie de la boucle tant que on obtient $\mathcal{M}_0^{t+1} = P^{(Maxiter)}$.

Algorithme 4 : Algorithme d'ajout d'un point

Si le temps de calcul n'est pas trop long, on peut faire une recherche exhaustive, i.e au lieu de tirer au hasard un point de $\tilde{\mathcal{M}}_1^{t+1} = \mathcal{M} \setminus P^{(i)}$, on va chercher une recherche sur tous les points de $\tilde{\mathcal{M}}_1^{t+1} = \mathcal{M} \setminus P^{(i)}$, auquel cas on obtient le plan minimax.

6.2.2.2 Construction d'un nouveau plan

La construction complète du plan minimax $\mathcal{M}_0^{t+1}$ va se faire en utilisant l'algorithme d'échange. Ces deux algorithmes permettent sous certaines conditions de converger vers un optimum global, i.e vers le plan minimax cherché.

Initialisation ;

$i = 0$, $P^{(i)}$;

tant que $i \leq Maxtier$ **faire**

On a $\widehat{P}^{(i)} = \mathcal{M} \setminus P^{(i)}$;

On tire au hasard un point $\tilde{s}$ de $P^{(i)}$, qu'on enlève pour obtenir $\tilde{P}^{(i)} = \widehat{P}^{(i)} \setminus \{\tilde{s}\}$

On tire au hasard un point s^* de $\widehat{P}^{(i)}$;

On ajoute $s*$ à $\tilde{P}^{(i)}$ pour obtenir ${P_*}^{(i)} = \tilde{P}^{(i)} \cup \{s^*\}$;

si $C_\rho^{{P_*}^{(i)}} < C_\rho^{P^{(i)}}$ **alors**

$P^{(i+1)} = {P_*}^{(i)}$

sinon

$P^{(i+1)} = P^{(i)}$

fin

i= i +1 ;

fin

A la sortie de la boucle tant que on obtient le plan d'expérience minimax $\mathcal{M}_0^{t+1} = P^{(Maxiter)}$.

Algorithme 5 : Algorithme d'échange

Le plan de départ $P^{(0)}$ peut être définit comme le plan minimax $\mathcal{M}_0^t$ obtenu à l'instant t, ou être construit de la manière suivante :

$$
\begin{aligned}
s_1 &= \underset{s\in\mathcal{M}}{\operatorname{argmax}}\{\sigma(s|z_t^{\mathcal{M}_0^t}) + \rho\cdot\overline{\mu}_{t+1}(s)\}\\
s_2 &= \underset{s\in\mathcal{M}\setminus\{s_1\}}{\operatorname{argmax}}\{\sigma(s|z_t^{\mathcal{M}_0^t}, z_{t+1}(s_1)) + \rho\cdot\overline{\mu}_{t+1}(s)\}\\
&\vdots\\
s_i &= \underset{s\in\mathcal{M}\setminus\{s_1,s_2,\ldots,s_{i-1}\}}{\operatorname{argmax}}\{\sigma(s|z_t^{\mathcal{M}_0^t}, z_{t+1}(s_1),\ldots,z_{t+1}(s_{i-1})) + \rho\cdot\overline{\mu}_{t+1}(s)\}\\
&\vdots\\
s_n &= \underset{s\in\mathcal{M}\setminus\{s_1,s_2,\ldots,s_{n-1}\}}{\operatorname{argmax}}\{\sigma(s|z_t^{\mathcal{M}_0^t}, z_{t+1}(s_1),\ldots,z_{t+1}(s_{n-1})) + \rho\cdot\overline{\mu}_{t+1}(s)\},
\end{aligned}
$$

où $\sigma(s|A) = \sqrt{\mathrm{Var}(Z_{t+1}(s)|A)}$.

Nous précisons que le plan de départ peut avoir une influence sur la convergence de l'algorithme d'échange vers un optimum global.

6.3 Application numérique

Nous allons faire des illustrations de notre méthode de construction de plans d'expérience pour ces deux cas de figures :

- **le cas où on désire ajouter un site d'observation de plus au plan existant :** Ce cas correspond à des situations où l'on sait que le phénomène ne varie pas beaucoup au cours du temps, mais par contre il y'a besoin d'ajouter un capteur de plus dans le domaine d'étude pour avoir une prédiction spatiale plus précise du phénomène à un temps t donné.
- **le cas où on reconstruit un nouveau plan pour chaque temps t donné :**
  Ce cas correspond à des situations où l'on sait que le phénomène varie plutôt rapidement au cours du temps. Donc pour chaque temps t donné on a besoin de reconstruire complètement un nouveau plan d'expériences où observer le phénomène d'intérêt.

Dans les deux cas notre but est de placer le ou les sites d'observation pour chaque temps t donné, dans la zone où l'incertitude sur l'estimation du phénomène est la plus grande, tout en tenant compte de la valeur espérée du phénomène à travers le paramètre ρ. Nous considérons toujours le problème de pollution décrit par l'équation de convection-diffusion qui dépend du

temps, avec une condition au bord de Dirichlet homogène :

$$\begin{cases} L(u) & := & \frac{\partial u}{\partial t} - \nu\Delta u + a_1\frac{\partial u}{\partial x} + a_2\frac{\partial u}{\partial y} = \mathcal{S}, \text{in } \Omega \times (0,T] \\ u(x,y,t) & = & 0 \text{ on } \partial\Omega \times (0,T] \\ u(x,y,0) & = & u_0(x,y) \text{ in } \Omega \end{cases} \tag{6.7}$$

Nous rappelons que $\mathcal{D}$ est un domaine carré $(a,b)^2 \subset \mathbb{R}^2$, $(0,T]$ est un intervalle de temps, $\mathcal{S} = \mathcal{S}(x,y,t) \in L^2(\Omega \times (0,T))$ est le terme source, u est l'état du système dynamique , a_1 et a_2 sont des constantes qui représentent les coefficients de convection et ν est une constante positive appelée le coefficient de diffusion. Pour les points du maillage $\mathcal{M}$, nous avons simulé les valeurs de u pour $t = 1, \ldots, 9$, en résolvant l'équation (6.7) avec $(a,b)^2 = (0,1)^2$, $a_1 = -0.09$, $a_2 = -0.09$, $\nu = 0.001$ et avec la source $\mathcal{S}$ donnée par (5.13) (voir Figure 5.2). Les valeurs de u obtenues aux différents temps t donnés sont considérées comme les vraies valeurs.

Pour tester l'efficacité de la méthode de construction de plans d'expériences que nous avons mise en place, nous allons à chaque temps t donné choisir quelques points de $\mathcal{M}$ que nous considérons comme des capteurs ou sondages et où nous allons observer les vraies valeurs de la pollution u. Les points de $\mathcal{M}$ choisis vont constituer le plan d'expériences noté par $\mathcal{M}_0^t$, à un temps t donné. Nous allons donc utiliser le modèle statistique qui à été présenté au chapitre 5, et notre méthode de construction de plans d'expériences pour estimer de manière précise les vraies valeurs de u aux sites non observés, à chaque temps t donné. Nous précisons que la boîte noire est définie par la même EDP (5.12) avec aussi $a_1 = -0.09$, $a_2 = -0.09$, $\nu = 0.001$. Nous n'allons pas revenir sur la méthode d'estimation de u aux sites non observés dans cette partie. Ceci a été déjà fait Section 5.3 du chapitre 5. Nous mettons l'accent sur la méthode de placement des capteurs ou des sondages (sites d'observation) pour chaque temps t donné, et l'impact de leurs choix de placement sur l'estimation de u à chaque temps t donné.

Nous faisons en premier l'illustration du cas où on prend $\rho = 0.1$, ce cas permet de tenir en compte de la valeur espérée du phénomène en plus de l'incertitude sur l'estimation de ce dernier, pour le choix de placement du ou des sites observations. Puis nous allons illustrer le cas où $\rho = 0$, ce cas signifie que seule l'incertitude sur l'estimation du phénomène est prise en compte pour le choix du ou des capteurs ou sondages.

6.3.1 Rajout d'un point d'observation

- $|\widehat{B}_t(x,y) - u(x,y,t)|$: Erreur de prédiction de la boîte noire au point (x,y)
- $|\widehat{u}_t(x,y) - u(x,y,t)|$: Erreur de prédictions de notre méthode au point (x,y)
- $\widehat{B}_t(x,y)$: Valeur prédite par la boîte noire au point (x,y)
- $\widehat{\Sigma}_t(x,y)$: Variance obtenue avec la boîte noire au point (x,y)

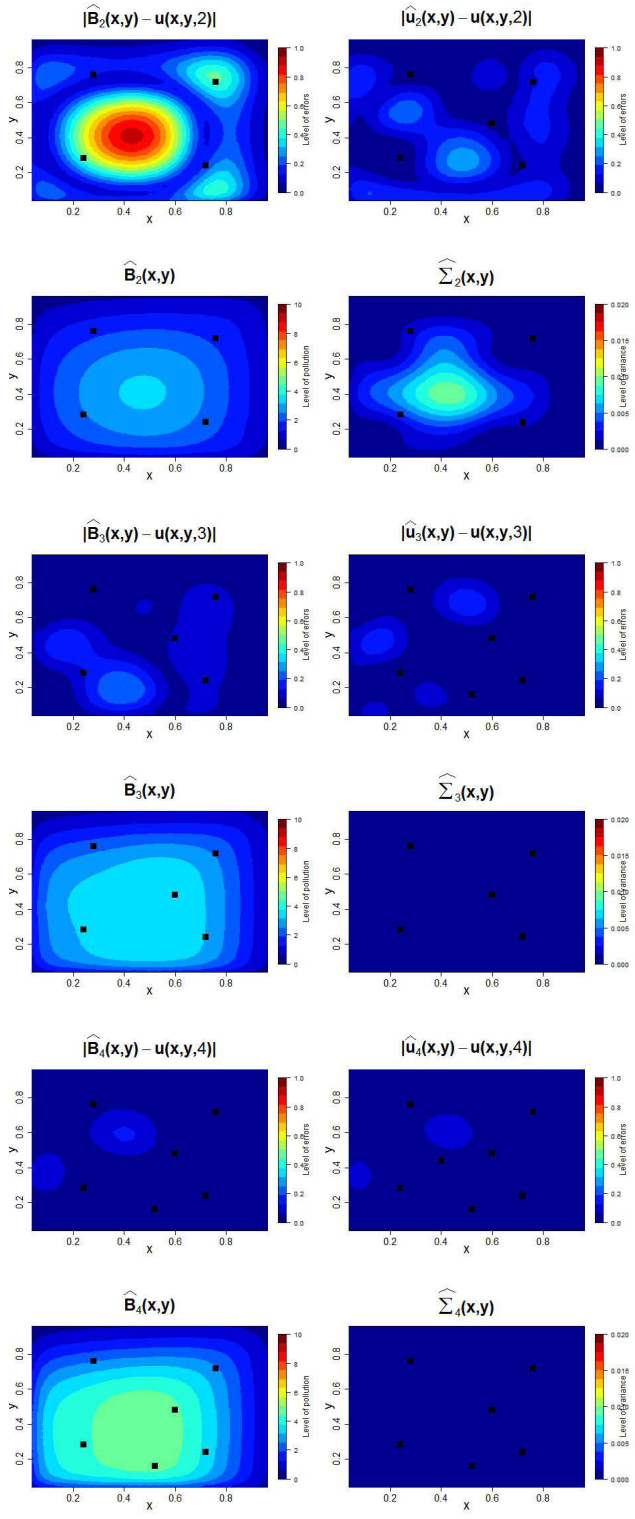

FIGURE 6.2 – $\rho = 0.1$, $\mathcal{M}_0^{t-1}$ est représenté sur $\widehat{B}_t(x,y)$, $\widehat{\Sigma}_t(x,y)$ et $|\widehat{B}_t(x,y) - u(x,y,t)|$, $\mathcal{M}_0^t$ est représenté sur $|\widehat{u}_t(x,y) - u(x,y,t)|$, $t = 2, ..., 4$.

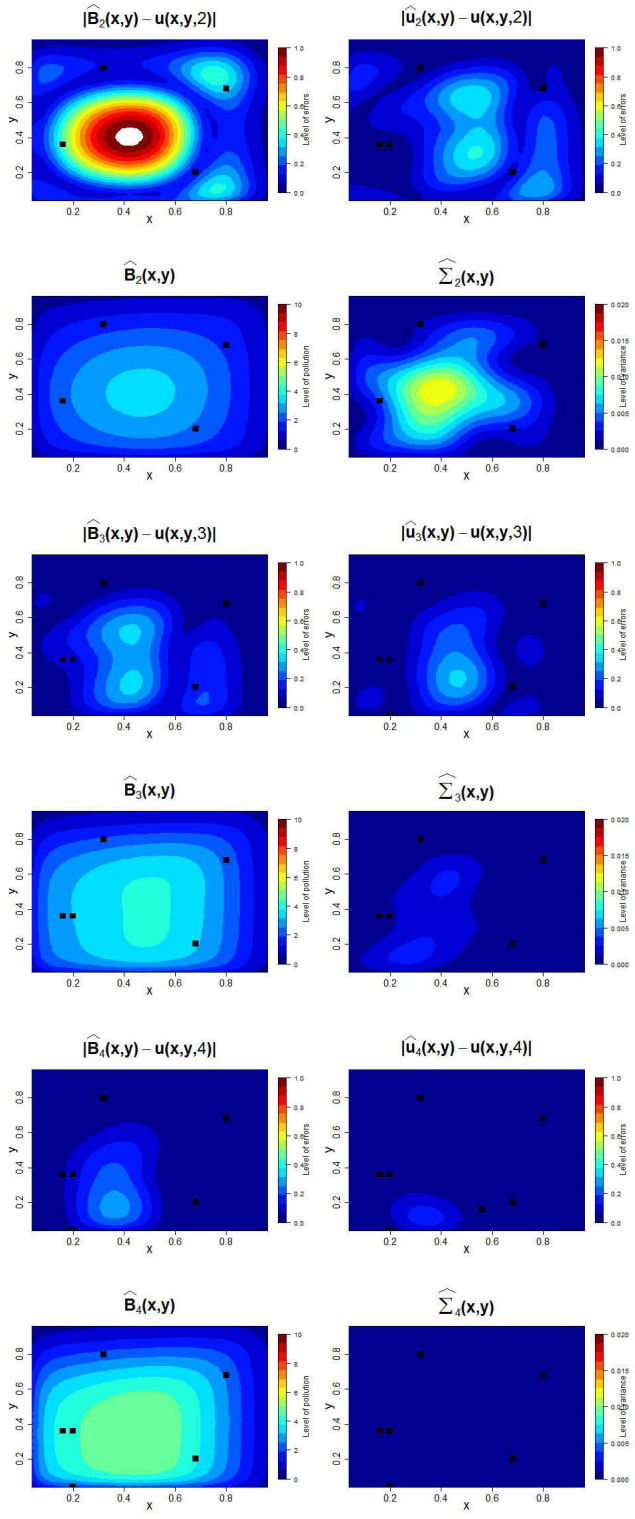

FIGURE 6.3 – $\rho = 0$, $\mathcal{M}_0^{t-1}$ est représenté sur $\widehat{B}_t(x,y)$, $\widehat{\Sigma}_t(x,y)$ et $|\widehat{B}_t(x,y) - u(x,y,t)|$, $\mathcal{M}_0^t$ est représenté sur $|\widehat{u}_t(x,y) - u(x,y,t)|$, $t = 2, ..., 4$.

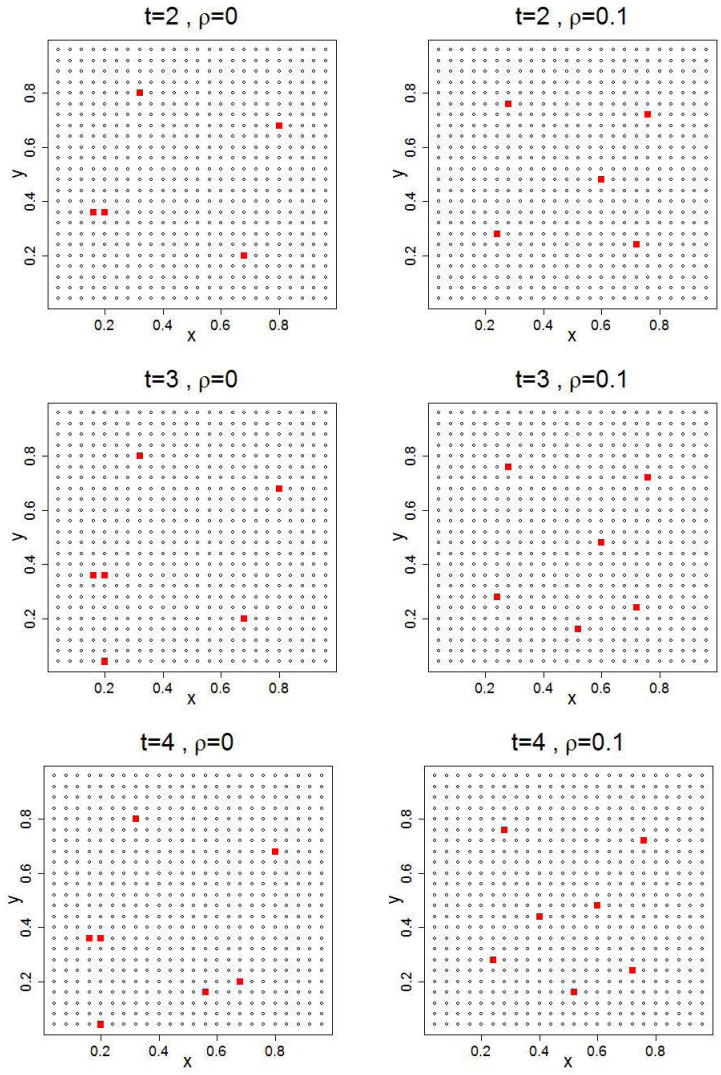

FIGURE 6.4 – Dynamique des points (rajout d'un point à chaque temps t), à gauche $\rho = 0$, à droite $\rho = 0.1$.

6.3.2 Construction d'un nouveau plan

- $|\widehat{B}_t(x,y) - u(x,y,t)|$: Erreur de prédiction de la boîte noire au point (x,y)
- $|\widehat{u}_t(x,y) - u(x,y,t)|$: Erreur de prédictions de notre méthode au point (x,y)
- $\widehat{B}_t(x,y)$: Valeur prédite par la boîte noire au point (x,y)
- $\widehat{\Sigma}_t(x,y)$: Variance obtenue avec la boîte noire au point (x,y)

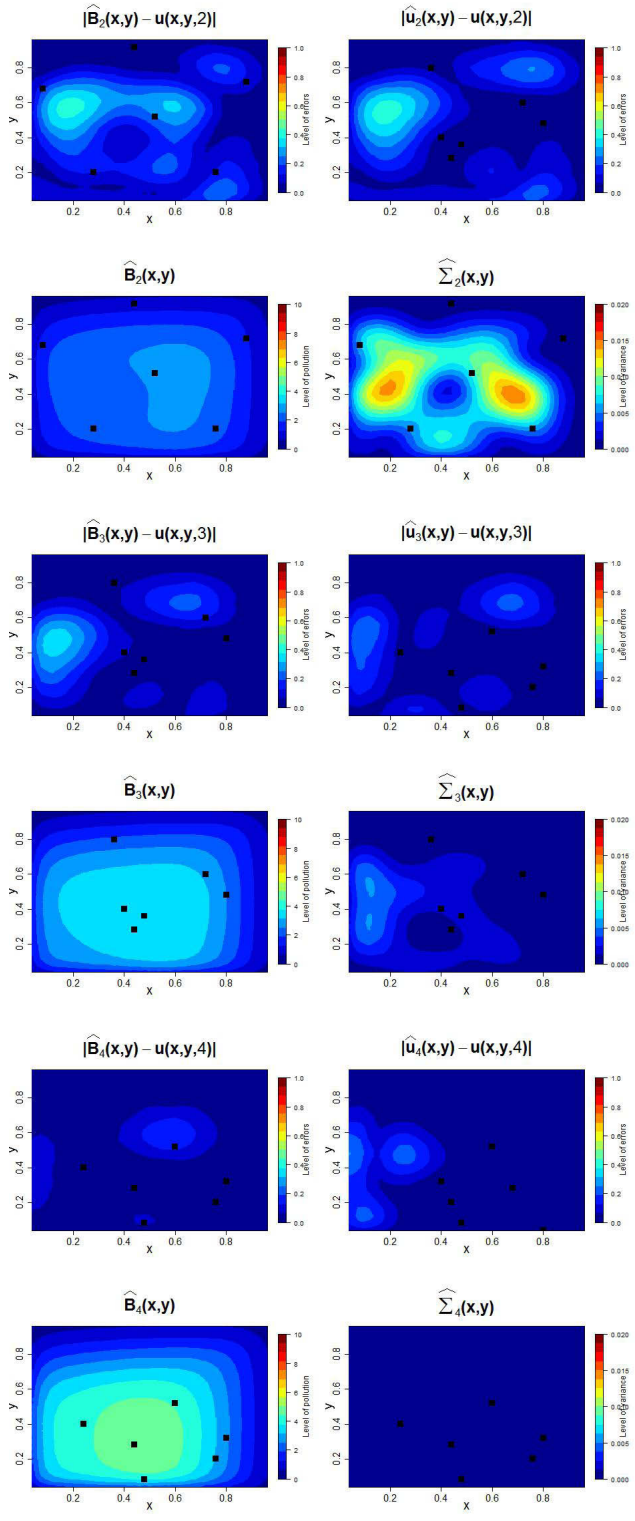

FIGURE 6.5 – $\rho = 0.1$, $\mathcal{M}_0^{t-1}$ est représenté sur $\widehat{B}_t(x,y)$, $\widehat{\Sigma}_t(x,y)$ et $|\widehat{B}_t(x,y) - u(x,y,t)|$, $\mathcal{M}_0^t$ est représenté sur $|\widehat{u}_t(x,y) - u(x,y,t)|$, $t = 2, ..., 4$.

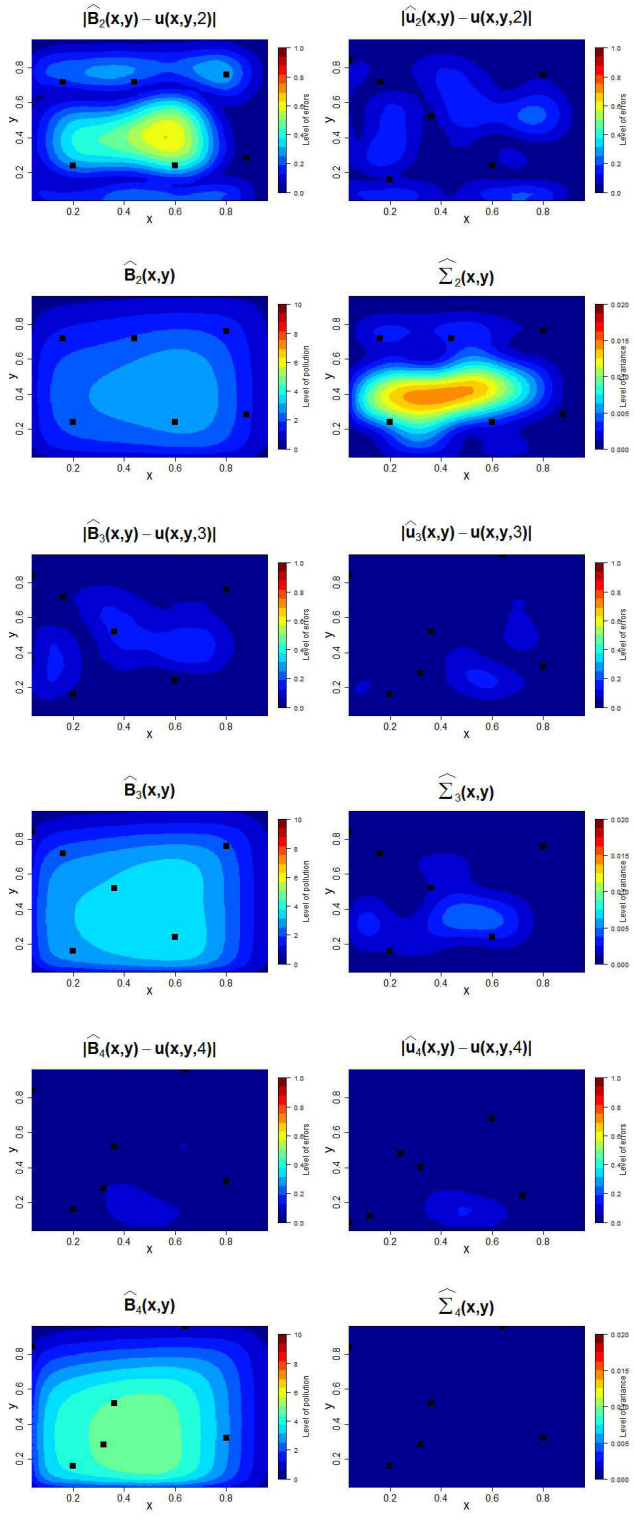

FIGURE 6.6 – $\rho = 0$, $\mathcal{M}_0^{t-1}$ est représenté sur $\widehat{B}_t(x,y)$, $\widehat{\Sigma}_t(x,y)$ et $|\widehat{B}_t(x,y) - u(x,y,t)|$, $\mathcal{M}_0^t$ est représenté sur $|\widehat{u}_t(x,y) - u(x,y,t)|$, $t = 2, ..., 4$.

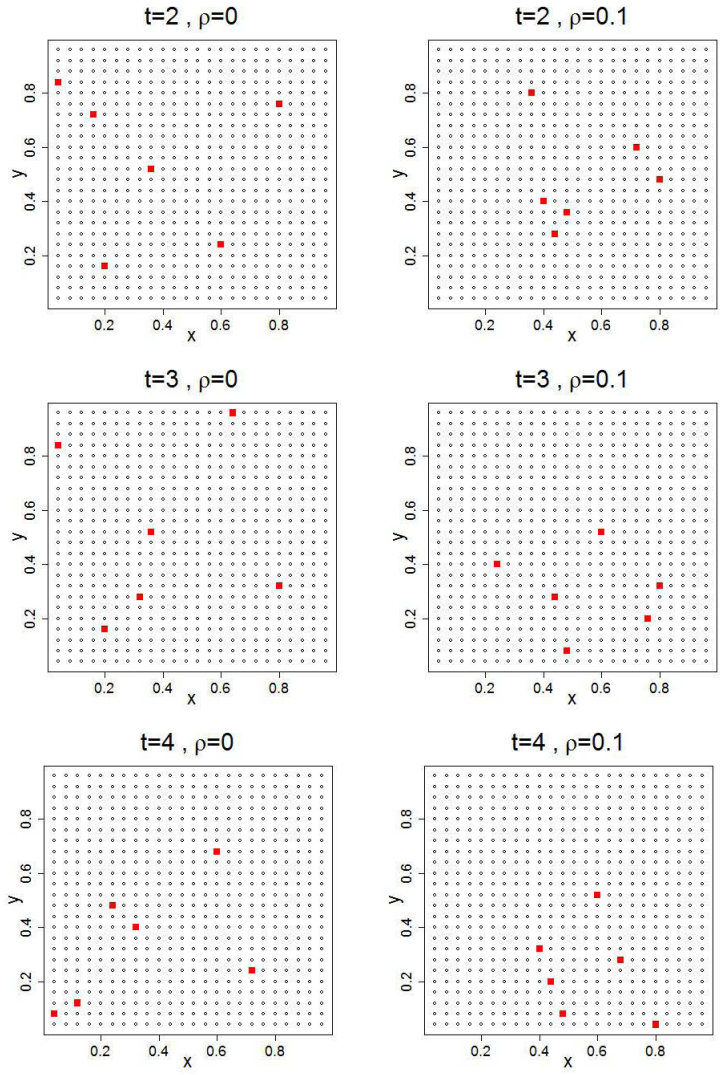

FIGURE 6.7 – Dynamique des points (construction d'un nouveau plan à chaque temps t), à gauche $\rho = 0$, à droite $\rho = 0.1$.

Comme dit auparavant ces simulations ont été faites en supposant qu'il n'y a pas d'erreurs de mesure, c'est-à-dire que les appareils (capteurs, sondes) sont de bonnes qualités et mesurent parfaitement la quantité de pollution exacte à chaque temps t donné. Dans le cas où l'on suppose qu'il y a des erreurs de mesure, la méthode ne marche pas parfaitement, nous avons constaté que pour certains temps t la prédiction faite par la boîte noire est meilleur, ceci dû au fait que la méthode semi-empirique n'assure pas un lissage suffisant lors de la prédiction.

Chapitre 7

Conclusion

Dans cette thèse nous avons proposé de nouvelles approches du traitement statistique de données spatio-temporelles. Afin de rester dans un cadre général et répondre ainsi aux problématiques diverses des industriels du secteur de l'environnement partenaire du projet, nous avons proposé un cadre général où la modélisation temporelle était représentée par une "boîte noire". Celle-ci pouvant être une représentation exacte du phénomène sous-jacent ou une approximation plus ou moins grossière.

Les méthodes Bayésiennes permettent de prendre en compte des sources de variabilité ou d'imprécision différentes : dans notre cas, il s'agissait de prendre en compte la connaissance imparfaite des données spatiales, représentée par un modèle de Krigeage, et l'incertitude sur la modélisation temporelle.

Un premier algorithme a été développé sous R et présenté au chapitre 4. Il a montré son efficacité sur des données simulées à l'aide d'une boîte noire mal spécifiée. Dans cette algorithme, nous nous somme essentiellement concentré sur la modélisation de la variabilité spatiale en spécifiant les loi a priori par des méthodes de Monte-Carlo. Une amélioration possible serait de mieux prendre en compte la modélisation de la moyenne spatiale surtout lorsque la boîte noire représente une assez bonne approximation de la réalité.

Nous avons ensuite, au chapitre 5, proposé une méthode basée sur une estimation semi-empirique de la moyenne et variance du champs spatial à l'aide de la boîte noire. Même si l'étude à partir de données simulées que nous avons traité donne des résultats satisfaisant, la méthode s'avère peu robuste à une mauvaise spécification de la boîte noire. Une amélioration de la méthode pourrait se faire en obtenant une estimation plus robuste de la matrice de variance-covariance du champs spatial.

Enfin, nous avons proposé un algorithme de construction de plan d'expériences pour données spatio-temporelles. Contrairement aux algorithmes de type "space filling" tendant à avoir des points les plus équi-répartis possibles, les plans que nous proposons tiennent compte des connaissances a priori du phénomène spatial. Une autre innovation importante a été de pendre en compte dans le critère d'optimalité la valeur prédite du phénomène étudié en plus de son incertitude. En effet, si l'objectif est par exemple de suivre la diffusion d'une source de pollution, il n'est pas toujours nécessaire d'avoir une connaissance fine du phénomène spatial aux endroits où l'on sait que la pollution sera faible et on préfèrera se concentrer sur les endroits où le niveau de pollution sera modéré ou fort. Même si nous n'avons appliqué notre approche qu'aux prévisions basée sur les méthodes semi-empiriques développées au chapitre 5, la méthode de construction proposée s'applique à tout type de situation, du moment que l'on a une connaissance a priori du champs spatial à étudier.

Résumé

Planification et Analyse de données spatio-temporelles

La Modélisation spatio-temporelle permet la prédiction d'une variable régionalisée à des sites non observés du domaine d'étude, basée sur l'observation de cette variable en quelques sites du domaine à différents temps t donnés. Dans cette thèse, l'approche que nous avons proposé consiste à coupler des modèles numériques et statistiques. En effet en privilégiant l'approche bayésienne nous avons combiné les différentes sources d'information : l'information spatiale apportée par les observations, l'information temporelle apportée par la boîte noire ainsi que l'information a priori connue du phénomène. Ce qui permet une meilleure prédiction et une bonne quantification de l'incertitude sur la prédiction. Nous avons aussi proposé un nouveau critère d'optimalité de plans d'expérience incorporant d'une part le contrôle de l'incertitude en chaque point du domaine et d'autre part la valeur espérée du phénomène.

Mots Clés : *prédiction spatio-temporelle*, *Bayésien*, *information spatiale*, *information temporelle*, *boîte noire*, *a priori*, *plans d'expérience*, *critère d'optimalité*.

Abstract

Design and analysis of spatio-temporal data

Spatio-temporal modeling allows to make the prediction of a regionalized variable at unobserved points of a given field, based on the observations of this variable at some points of field at different times. In this thesis, we proposed a approach which combine numerical and statistical models. Indeed by using the Bayesian methods we combined the different sources of information : spatial information provided by the observations, temporal information provided by the black-box and the prior information on the phenomenon of interest. This approach allowed us to have a good prediction of the variable of interest and a good quantification of incertitude on this prediction. We also proposed a new method to construct experimental design by establishing a optimality criterion based on the uncertainty and the expected value of the phenomenon.

Keywords : *spatio-temporal prediction*, *Bayesian*, *spatial information*, *temporal information*, *black-box*, *prior information*, *experimental design*, *optimality criterion*.

Bibliographie

[1] Karim M Abadir, Walter Distaso, and Filip Žikeš. Design-free estimation of variance matrices. *Journal of Econometrics*, 181(2) :165–180, 2014. 66, 68, 71, 81

[2] John Aldrich et al. Ra fisher and the making of maximum likelihood 1912-1922. *Statistical Science*, 12(3) :162–176, 1997. 17

[3] Brian DO Anderson and John B Moore. *Optimal filtering*. Courier Corporation, 2012. 40

[4] Bruce Ankenman, Barry L. Nelson, and Jeremy Staum. Stochastic kriging for simulation metamodeling. *Oper. Res.*, 58(2) :371–382, 2010. 13

[5] Michel Arnaud and Xavier Emery. *Estimation et interpolation spatiale : méthodes déterministes et méthodes géostatistiques*. Hermes science publications, 2000. 13, 16, 17

[6] Sophie Baillargeon. *Le krigeage : revue de la théorie et application à l'interpolation spatiale de données de précipitations*. PhD thesis, Université Laval, 2005. 13, 15, 16, 17

[7] Mikio Bando, Yukihiro Kawamata, and Toshiyuki Aoki. Dynamic sensor bias correction for attitude estimation using unscented Kalman filter in autonomous vehicle. In *Proceedings of the 42nd ISCIE International Symposium on Stochastic Systems Theory and its Applications*, pages 33–39. Inst. Syst. Control Inform. Engrs. (ISCIE), Kyoto, 2011. 44

[8] Sudipto Banerjee, Alan E Gelfand, and Bradley P Carlin. *Hierarchical modeling and analysis for spatial data*. Crc Press, 2004. 20

[9] Jarrett J Barber. Bayesian core : A practical approach to computational bayesian statistics. *Journal of the American Statistical Association*, 103(481) :432–433, 2008. 21

[10] Pierre Barbillon. *Méthodes d'interpolation à noyaux pour l'approximation de fonctions type boîte noire coûteuses*. PhD thesis, Université Paris Sud-Paris XI, 2010. 78, 79

[11] Piero Barone. A black box method for solving the complex exponentials approximation problem. *Digit. Signal Process.*, 23(1) :49–64, 2013. 32

[12] G Bastin and V Wertz. Modélisation et analyse des systemes dynamiques. *Lectures notes, Louvain School of Engineering*, 2013. 31

[13] Christian Berg. Positive definite and related functions on semigroups. *The Analytical and Topological Theory of Semigroups : Trends and Developments*, 1 :253, 1990. 10

[14] Régis Bettinger. *Inversion d'un système par krigeage : application à la synthèse des catalyseurs à haut débit*. PhD thesis, Université de Nice Sophia Antipolis, 2009. 9, 27

[15] Nicolas Bez. Transitive geostatistics and statistics per individual : a relevant framework for assessing resources with diffuse limits. *Journal de la Société française de statistique*, 148(1) :53–75, 2007. 7, 8

[16] Christele Bioche and Pierre Druilhet. Approximation of improper prior by vague priors. 30 pages., November 2013. 21

[17] Phillip Boyle and Marcus Frean. Dependent gaussian processes. *Advances in neural information processing systems*, 17 :217–224, 2005. 34

[18] Jean-Paul Chilès and Pierre Delfiner. *Geostatistics*. Wiley Series in Probability and Statistics : Applied Probability and Statistics. John Wiley & Sons, Inc., New York, 1999. Modeling spatial uncertainty, A Wiley-Interscience Publication. 34

[19] George Christakos. On the problem of permissible covariance and variogram models. *Water Resources Research*, 20(2) :251–265, 1984. 15

[20] Faruk Civan. Waterflooding of naturally fractured reservoirs : an efficient simulation approach. In *Production Operations Symposium*, pages 395–407, 1993. 29

[21] Jorge Cortés. Distributed Kriged Kalman filter for spatial estimation. *IEEE Trans. Automat. Control*, 54(12) :2816–2827, 2009. 47, 48

[22] Richard W. Cottle. Manifestations of the Schur complement. *Linear Algebra and Appl.*, 8 :189–211, 1974. 64

[23] Noel Cressie. The origins of kriging. *Mathematical Geology*, 22(3) :239–252, 1990. 13, 14

[24] Noel Cressie and Christopher K Wikle. Space-time kalman filter. *Encyclopedia of environmetrics*, 2002. 33

[25] Noel Cressie and Christopher K Wikle. *Statistics for spatio-temporal data*. John Wiley & Sons, 2011. 3, 33

[26] Noel A. C. Cressie. *Statistics for spatial data*. Wiley Series in Probability and Mathematical Statistics : Applied Probability and Statistics. John Wiley & Sons, Inc., New York, 1991. A Wiley-Interscience Publication. 7

[27] Noel AC Cressie. Statistics for spatial data, revised edition, 1993. 7, 10, 13, 15, 16

[28] Javier Cuadrado, Daniel Dopico, Jose A. Perez, and Roland Pastorino. Automotive observers based on multibody models and the extended Kalman filter. *Multibody Syst. Dyn.*, 27(1) :3–19, 2012. 42

[29] Carla Currin, Toby Mitchell, Max Morris, and Don Ylvisaker. Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments. *Journal of the American Statistical Association*, 86(416) :953–963, 1991. 77

[30] A. C. Davison, R. Huser, and E. Thibaud. Geostatistics of dependent and asymptotically independent extremes. *Math. Geosci.*, 45(5) :511–529, 2013. 7

[31] Mehdi Dehghan. On the numerical solution of the one-dimensional convection-diffusion equation. *Mathematical Problems in Engineering*, 2005(1) :61–74, 2005. 69

[32] JP Dejean, G Blanc, et al. Managing uncertainties on production predictions using integrated statistical methods. In *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers, 1999. 76

[33] Xin Ping Deng, Jian Hua Zheng, and Dong Gao. Application research of modified unscented Kalman filter in X-ray pulsar navigation. *Trans. Beijing Inst. Tech.*, 32(1) :51–55, 2012. 44

[34] Peter J. Diggle and Paulo J. Ribeiro, Jr. *Model-based geostatistics*. Springer Series in Statistics. Springer, New York, 2007. 20, 24

[35] Peter J Diggle and Paulo J Ribeiro Jr. Bayesian inference in gaussian model-based geostatistics. *Geographical and Environmental Modelling*, 6(2) :129–146, 2002. 7, 20

[36] Pierre Druilhet, Denys Pommeret, et al. Invariant conjugate analysis for exponential families. *Bayesian Analysis*, 7(4) :903–916, 2012. 21

[37] Y Efendiev and LJ Durlofsky. A generalized convection-diffusion model for subgrid transport in porous media. *Multiscale Modeling & Simulation*, 1(3) :504–526, 2003. 69

[38] Geir Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using monte carlo methods to forecast error statistics. *Journal of Geophysical Research : Oceans (1978–2012)*, 99(C5) :10143–10162, 1994. 44

[39] V. V. Fedorov. *Theory of optimal experiments*. Academic Press, New York-London, 1972. Translated from the Russian and edited by W. J. Studden and E. M. Klimko, Probability and Mathematical Statistics, No. 12. 75

[40] A Stewart Fotheringham, Chris Brunsdon, and Martin Charlton. *Geographically weighted regression : the analysis of spatially varying relationships*. John Wiley & Sons, 2003. 13

[41] Jean Baptiste Joseph Fourier. *Théorie analytique de la chaleur*. Cambridge Library Collection. Cambridge University Press, Cambridge, 2009. Reprint of the 1822 original, Previously published by Éditions Jacques Gabay, Paris, 1988 [MR1414430]. 30

[42] Jessica Franco. *Planification d'expériences numériques en phase exploratoire pour la simulation des phénomènes complexes*. PhD thesis, Ecole Nationale Supérieure des Mines de Saint-Etienne, 2008. 78

[43] Lev Semenovich Gandin and R Hardin. *Objective analysis of meteorological fields*, volume 242. Israel program for scientific translations Jerusalem, 1965. 13

[44] Jean-Pierre Gauchi. Plans d'expériences optimaux : un exposé didactique. *revue Modulad*, 33 :139–162, 2005. 75

[45] A Gelman, J Carlin, and H Stern. D. rubin, bayesian data analysis, 1995. 20

[46] Walter R Gilks, Sylvia Richardson, and David J Spiegelhalter. *Markov chain Monte Carlo in practice*, volume 2. CRC press, 1996. 20

[47] Tilmann Gneiting, William Kleiber, and Martin Schlather. Matérn cross-covariance functions for multivariate random fields. *Journal of the American Statistical Association*, 105(491), 2010. 34

[48] J. Jaime Gómez-Hernández, Ana Horta, and Nicolas Jeanée. Geostatistics for environmental applications. *Spat. Stat.*, 5 :1–2, 2013. 7

[49] Pierre Goovaerts. *Geostatistics for natural resources evaluation*. Oxford university press, 1997. 15

[50] Michel Goulard and Marc Voltz. Linear coregionalization model : tools for estimation and choice of cross-variogram matrix. *Mathematical Geology*, 24(3) :269–286, 1992. 34

[51] Ralf Hartmut Güting. An introduction to spatial database systems. *The VLDB Journal The International Journal on Very Large Data Bases*, 3(4) :357–399, 1994. 3, 33

[52] Chingiz Hajiyev and Halil Ersin Soken. Robust adaptive unscented Kalman filter for attitude estimation of pico satellites. *Internat. J. Adapt. Control Signal Process.*, 28(2) :107–120, 2014. 42

[53] Mark S Handcock and Michael L Stein. A bayesian analysis of kriging. *Technometrics*, 35(4) :403–410, 1993. 20

[54] Emilie V. Haynsworth. Determination of the inertia of a partitioned Hermitian matrix. *Linear Algebra and Appl.*, 1(1) :73–81, 1968. 63

[55] Man Hong and Shao Cheng. Model predictive control based on Kalman filter for constrained Hammerstein-Wiener systems. *Math. Probl. Eng.*, pages Art. ID 104702, 6, 2013. 39

[56] Rui Huang, Sachin C. Patwardhan, and Lorenz T. Biegler. Robust stability of nonlinear model predictive control with extended Kalman filter and target setting. *Internat. J. Robust Nonlinear Control*, 23(11) :1240–1264, 2013. 42

[57] Andrew H Jazwinski. *Stochastic processes and filtering theory*. Courier Corporation, 2007. 39

[58] Stephen J Jeffrey, John O Carter, Keith B Moodie, and Alan R Beswick. Using spatial interpolation to construct a comprehensive archive of australian climate data. *Environmental Modelling & Software*, 16(4) :309–330, 2001. 7

[59] Harold Jeffreys. *Theory of Probability*. Oxford University Press, Oxford, 1939. 21

[60] Harold Jeffreys. *Theory of probability*. Oxford Classic Texts in the Physical Sciences. The Clarendon Press, Oxford University Press, New York, 1998. Reprint of the 1983 edition. 21

[61] M. E. Johnson, L. M. Moore, and D. Ylvisaker. Minimax and maximin distance designs. *J. Statist. Plann. Inference*, 26(2) :131–148, 1990. 78

[62] Astrid Jourdan. Planification d'experiences numeriques. *Revue Modulad*, 63(33), 2005. 76, 78

[63] Astrid Jourdan and Dominique Collombier. Régression trigonométrique et plans d'échantillonnage pour expériences simulées. *Revue de statistique appliquée*, 49(2) :5–25, 2001. 78

[64] S Julier and J Uhlmann. A new extension of the kalman filter to nonlinear systems. int. symp. *Aerospace/Defense Sensing*, 1997. 44, 46

[65] Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. *Journal of Fluids Engineering*, 82(1) :35–45, 1960. 38

[66] DG Kbiob. A statistical approach to some basic mine valuation problems on the witwatersrand. *Journal of Chemical, Metallurgical, and Mining Society of South Africa*, 1951. 9, 13

[67] Marc C Kennedy and Anthony O'Hagan. Predicting the output from a complex computer code when fast approximations are available. *Biometrika*, 87(1) :1–13, 2000. 36

[68] J. Kiefer. Optimum experimental designs. *J. Roy. Statist. Soc. Ser. B*, 21 :272–319, 1959. 75

[69] J. Kiefer. Optimum experimental designs. In *Actes du Congrès International des Mathématiciens (Nice, 1970), Tome 3*, pages 249–254. Gauthier-Villars, Paris, 1971. 75

[70] J. Kiefer. General equivalence theory for optimum designs (approximate theory). *Ann. Statist.*, 2 :849–879, 1974. 75

[71] J. Kiefer and J. Wolfowitz. Optimum designs in regression problems. *Ann. Math. Statist.*, 30 :271–294, 1959. 75

[72] Gregor Klančar, Luka Teslić, and Igor Škrjanc. Mobile-robot pose estimation and environment mapping using an extended Kalman filter. *Internat. J. Systems Sci.*, 45(12) :2603–2618, 2014. 42

[73] B Klinkenberg and N Waters. Spatial interpolation i. *Goodchild, M. et Kemp, K., éditeurs : NCGIA Core Curriculum in GIS. National Center for Geographic Information and Analysis, University of California, Santa Barbara CA.[En ligne], http ://www. geog. ubc. ca/courses/klink/gis. notes/ncgia/u40. html (Page consultée le 28 avril 2005)*, 1990. 13

[74] Darongsae Kwon. *Kriging prediction with estimated covariances*. ProQuest LLC, Ann Arbor, MI, 2012. Thesis (Ph.D.)–The University of Chicago. 13

[75] Gail Langran. *Time in geographic information systems*. CRC Press, 1992. 33

[76] Théophile Lasm. *HYDROGÉOLOGIE DES RÉSERVOIRS FRACTURÉS DE SOCLE : ANALYSES STATISTIQUE ET GÉOSTATISTIQUE DE LA FRACTURATION ET DES PROPRIÉTÉS HYDRAULIQUES. APPLICATION À LA RÉGION DES MONTAGNES DE CÔTE D'IVOIRE(DOMAINE ARCHÉEN)*. PhD thesis, 2000. 7

[77] Loic Le Gratiet. Bayesian analysis of hierarchical multifidelity codes. 2013. 12, 13, 20, 33, 36

[78] Loic Le Gratiet. *Multi-fidelity Gaussian process regression for computer experiments*. PhD thesis, Université Paris-Diderot-Paris VII, 2013. 33, 36, 37

[79] Matthieu Lemoine and Florian Pelgrin. Introduction aux modèles espace-état et au filtre de kalman. *Revue de l'OFCE*, (3) :203–229, 2003. 46

[80] Sabrina Maggio, Claudia Cappello, and Daniela Pellegrino. GIS and geostatistics for supporting environmental analyses in space-time. In *Statistical methods for spatial planning and monitoring*, Contrib. Statist., pages 77–92. Physica-Verlag/Springer, Milan, 2013. 7

[81] Kanti V. Mardia, Colin Goodall, Edwin J. Redfern, and Francisco J. Alonso. The Kriged Kalman filter. *Test*, 7(2) :217–285, 1998. With comments and a rejoinder by the authors. 33, 46, 48

[82] Jean-Michel Marin and Christian Robert. *Bayesian Core : A Practical Approach to Computational Bayesian Statistics : A Practical Approach to Computational Bayesian Statistics*. Springer Science & Business Media, 2007. 21

[83] Reinaldo Marques, Adrian Pizzinga, and Luciano Vereda. Restricted Kalman filter applied to dynamic style analysis of actuarial funds. *Appl. Stoch. Models Bus. Ind.*, 28(6) :558–570, 2012. 39

[84] Jay D Martin and Timothy W Simpson. A monte carlo simulation of the kriging model. In *10th AIAA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, pages 2004–4483, 2004. 27

[85] Bertil Matérn. *Spatial variation*, volume 36 of *Lecture Notes in Statistics*. Springer-Verlag, Berlin, second edition, 1986. With a Swedish summary. 17

[86] Georges Matheron. *Traité de géostatistique appliquée. 1 (1962)*, volume 1. Editions Technip, 1962. 7, 13

[87] Georges Matheron. Principles of geostatistics. *Economic geology*, 58(8) :1246–1266, 1963. 13

[88] Georges Matheron. *La théorie des variables régionalisées et ses applications*. Ecole Nationale Supérieure des Mines de Paris, 1970. 7, 16

[89] Georges Matheron. *The theory of regionalized variables and its applications*, volume 5. École national supérieure des mines, 1971. 7

[90] Richard J Meinhold and Nozer D Singpurwalla. Understanding the kalman filter. *The American Statistician*, 37(2) :123–127, 1983. 46

[91] Alessandra Menafoglio and Piercesare Secchi. A universal kriging predictor for spatially dependent functional data of a Hilbert space. *Electron. J. Stat.*, 7 :2209–2240, 2013. 13

[92] Ghanim Mhmood Dhaher and Muhammad Hisyam Lee. A comparison between the performance of kriging and cokriging in spatial estimation with application. *Matematika (Johor Bahru)*, 29(1B) :33–41, 2013. 13

[93] Sueli Aparecida Mingoti, Arlene Guimarães Leite, and Gilmar Rosa. Describing the total number of diagnosed cases of AIDS by means of geostatistics. *Rev. Mat. Estatíst.*, 24(1) :61–76, 2006. 7

[94] José Mira and María Jesús Sánchez. Prediction of deterministic functions : an application of a Gaussian kriging model to a time series outlier problem. *Comput. Statist. Data Anal.*, 44(3) :477–491, 2004. 13

[95] Pascal Monestiez, Dominique Courault, Denis Allard, and Françoise Ruget. Spatial interpolation of air temperature using environmental context : application to a crop model. *Environ. Ecol. Stat.*, 8(4) :297–309, 2001. 7

[96] Max D Morris and Toby J Mitchell. Exploratory designs for computational experiments. *Journal of statistical planning and inference*, 43(3) :381–402, 1995. 78

[97] A Murangira. *Nouvelles approches en filtrage particulaire. Application au recalage de la navigation inertielle*. PhD thesis, Université de Technologie de Troyes-UTT, 2014. 44, 46

[98] In Jae Myung. Tutorial on maximum likelihood estimation. *Journal of mathematical Psychology*, 47(1) :90–100, 2003. 17

[99] Isaac Newton. Philosophiæ naturalis principia mathematica (mathematical principles of natural philosophy). *London (1687)*, 1965. 29

[100] Szu Hui Ng and Jun Yin. Bayesian kriging analysis and design for stochastic simulations. *ACM Trans. Model. Comput. Simul.*, 22(3) :Art. 17, 26, 2012. 20

[101] Harald Niederreiter. *Random number generation and quasi-Monte Carlo methods*, volume 63 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992. 77

[102] Diane Valérie Ouellette. Schur complements and statistics. *Linear Algebra Appl.*, 36 :187–295, 1981. 64

[103] Lauren E Padilla and Clarence W Rowley. An adaptive-covariance-rank algorithm for the unscented kalman filter. In *Decision and Control (CDC), 2010 49th IEEE Conference on*, pages 1324–1329. IEEE, 2010. 44

[104] Jeong-Soo Park. Optimal Latin-hypercube designs for computer experiments. *J. Statist. Plann. Inference*, 39(1) :95–111, 1994. 77

[105] Nikos Pelekis, Babis Theodoulidis, Ioannis Kopanakis, and Yannis Theodoridis. Literature review of spatio-temporal database models. *The Knowledge Engineering Review*, 19(03) :235–274, 2004. 3, 33

[106] Paulo J Ribeiro Jr and Peter J Diggle. Technical report st-99-09. 2000. 24

[107] Christian Robert. *L'analyse statistique bayésienne*. Economica, 1992. 21

[108] Christian P. Robert. *The Bayesian choice*. Springer Texts in Statistics. Springer, New York, second edition, 2007. From decision-theoretic foundations to computational implementation. 20, 22

[109] John F Roddick and Myra Spiliopoulou. A bibliography of temporal, spatial and spatio-temporal data mining research. *ACM SIGKDD Explorations Newsletter*, 1(1) :34–38, 1999. 33

[110] Paritosh K. Roy and Syed S. Hossain. Predicting arsenic concentration in groundwater of Bangladesh using Bayesian geostatistical model. *Environ. Ecol. Stat.*, 21(3) :583–597, 2014. 7

[111] Jerome Sacks, William J. Welch, Toby J. Mitchell, and Henry P. Wynn. Design and analysis of computer experiments. *Statist. Sci.*, 4(4) :409–435, 1989. With comments and a rejoinder by the authors. 14, 77

[112] J Schur. Über potenzreihen, die im innern des einheitskreises beschränkt sind. *Journal für die reine und angewandte Mathematik*, 147 :205–232, 1917. 63

[113] Bogdan Sebacher, Remus Hanea, and Arnold Heemink. A probabilistic parametrization for geological uncertainty estimation using the ensemble Kalman filter (EnKF). *Comput. Geosci.*, 17(5) :813–832, 2013. 44

[114] Michael C Shewry and Henry P Wynn. Maximum entropy sampling. *Journal of applied statistics*, 14(2) :165–170, 1987. 77

[115] Robin Sibson. A brief description of natural neighbour interpolation. *Interpreting multivariate data*, 21, 1981. 13

[116] Lun-Ji Song and Yu-Jiang Wu. A modified Crank-Nicolson scheme with incremental unknowns for convection dominated diffusion equations. *Appl. Math. Comput.*, 215(9) :3293–3301, 2010. 29, 30, 69

[117] Shubham Srivastava and Tarek Echekki. A novel Kalman filter based approach for multiscale reacting flow simulations. *Comput. & Fluids*, 81 :1–9, 2013. 39, 44

[118] A. Stein. Analysis of space-time variability in agriculture and the environment with geostatistics. *Statist. Neerlandica*, 52(1) :18–41, 1998. 7

[119] Michael Stein. Large sample properties of simulations using Latin hypercube sampling. *Technometrics*, 29(2) :143–151, 1987. 77

[120] Michael L. Stein. *Interpolation of spatial data*. Springer Series in Statistics. Springer-Verlag, New York, 1999. Some theory for Kriging. 9

[121] Subchan and Tahiyatul Asfihani. The missile guidance estimation using extended Kalman filter-unknown input-without direct feedthrough (EKF-UI-WDF) method. *J. Indones. Math. Soc.*, 19(1) :1–14, 2013. 42

[122] A. Subramanyam and H. S. Pandalai. On the equivalence of the cokriging and kriging systems. *Math. Geol.*, 36(4) :507–523, 2004. 13

[123] David Sussillo and Omri Barak. Opening the black box : low-dimensional dynamics in high-dimensional recurrent neural networks. *Neural Comput.*, 25(3) :626–649, 2013. 32

[124] Abdullah Uz Tansel, James Clifford, Shashi Gadia, Sushil Jajodia, Arie Segev, and Richard Snodgrass. *Temporal databases : theory, design, and implementation.* Benjamin-Cummings Publishing Co., Inc., 1993. 3, 33

[125] J. A. Vargas-Guzmán. Transitive geostatistics for stepwise modeling across boundaries between rock regions. *Math. Geosci.*, 40(8) :861–873, 2008. 7

[126] E. Vazquez, E. Walter, and G. Fleury. Intrinsic Kriging and prior information. *Appl. Stoch. Models Bus. Ind.*, 21(2) :215–226, 2005. 13

[127] Rossmary Villegas, Oliver Dorn, Miguel Moscoso, and Manuel Kindelan. Characterization of reservoirs by evolving level set functions obtained from geostatistics. In *Progress in industrial mathematics at ECMI 2006*, volume 12 of *Math. Ind.*, pages 597–602. Springer, Berlin, 2008. 7

[128] Hans Wackernagel. Cokriging. In *Multivariate Geostatistics*, pages 144–151. Springer, 1995. 34

[129] Hans Wackernagel. *Multivariate geostatistics.* Springer, 2003. 7, 34

[130] G Wahba. Spline models for observational data, cbms-nsf regional conference series in applied mathematics, vol. 59. *SIAM, Philadelphia, PA*, 1990. 13

[131] Dennis J. J. Walvoort and Jaap J. de Gruijter. Compositional kriging : a spatial interpolation method for compositional data. *Math. Geol.*, 33(8) :951–966, 2001. 13

[132] Matt P Wand and M Chris Jones. Kernel smoothing, vol. 60 of. *Monographs on statistics and applied probability*, 1995. 13

[133] Boyu Yi, Longyun Kang, Kai Jiang, and Yujian Lin. A two-stage Kalman filter for sensorless direct torque controlled PM synchronous motor drive. *Math. Probl. Eng.*, pages Art. ID 768736, 12, 2013. 42

[134] Boyu Yi, Longyun Kang, Sinian Tao, Xianxian Zhao, and Zhaoxia Jing. Adaptive two-stage extended Kalman filter theory in application of sensorless control for permanent magnet synchronous motor. *Math. Probl. Eng.*, pages Art. ID 974974, 13, 2013. 44

[135] May Yuan. Temporal gis and spatio-temporal modeling. In *Proceedings of Third International Conference Workshop on Integrating GIS and Environment Modeling, Santa Fe, NM*, 1996. 33

[136] Arnold Zellner. On assessing prior distributions and bayesian regression analysis with g-prior distributions. *Bayesian inference and decision techniques : Essays in Honor of Bruno De Finetti*, 6 :233–243, 1986. 21

[137] Ying Zhang, Yeng Chai Soh, and Weihai Chen. Robust information filter for decentralized estimation. *Automatica J. IFAC*, 41(12) :2141–2146, 2005. 40