
Méthodologies pour la méta-modélisation

La conception d'un méta-modèle fiable nécessite traditionnellement plusieurs étapes clés. Elles sont résumées par le schéma représenté Figure 2.1. Certaines étapes sont complètement dictées par le cas d'application, notamment la définition des paramètres incertains, le choix de la réponse d'intérêt, la manière de l'évaluer ou encore l'utilisation qui est faite du méta-modèle. A l'inverse, la planification d'expériences, la construction du méta-modèle et sa validation sont des étapes génériques. Différentes méthodologies sont disponibles pour la réalisation de chacune d'entre elles. On se propose dans ce chapitre de décrire les méthodes utilisées pour ces étapes génériques dans le cadre de cette thèse, et en particulier les spécificités liées à la multi-fidélité.

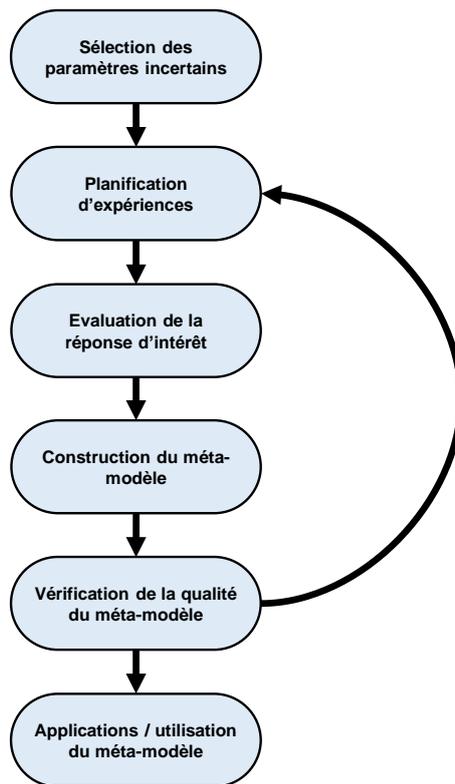


Figure 2.1 : Schéma traditionnel d'un workflow basé sur la méta-modélisation.

Tout d'abord, la planification d'expériences et ses enjeux sont brièvement introduits dans la partie 2.1. Dans un cadre classique de simple fidélité, les méthodes utilisées sont l'échantillonnage par hypercube latin pour la définition du plan d'expériences et le krigeage pour la construction du méta-modèle. Ces méthodes sont détaillées dans la partie 2.2. Celles utilisées dans un cadre multi-fidélité sont ensuite

décrites au sein de la partie 2.3. Il s'agit de l'échantillonnage par hypercubes latins imbriqués et du co-krigeage multi-fidélité. Finalement, l'estimation de la qualité des méta-modèles obtenus fait l'objet de la partie 2.4.

2.1 Introduction à la planification d'expériences

La planification d'expériences est un élément incontournable de la méta-modélisation. Elle consiste à choisir des points dans l'espace des paramètres où évaluer la réponse d'intérêt. L'ensemble des points d'expériences sélectionnés forme le plan d'expériences. La réponse d'intérêt est ensuite évaluée (*i.e.* simulée) pour chacun de ces points, formant l'ensemble des observations. Le plan d'expériences et ses observations sont les seules informations nécessaires à la construction d'un méta-modèle. La définition du plan d'expériences a donc un très fort impact sur la qualité du méta-modèle. De très nombreuses méthodes ont été proposées dans la littérature pour définir un plan d'expériences, notamment dans le cadre de la construction d'un méta-modèle. Ces méthodes sont largement détaillées dans les livres de [Koehler and Owen, 1996], [Fang et al., 2006] ou encore [Kleijnen, 2008].

Dans cette thèse, nous nous intéressons à la conception de méta-modèles prédictifs pour l'ensemble de l'espace sur lesquels ils sont définis. Pour cela, on considère des plans d'expériences exploratoires, pour lesquels les points d'expériences sont bien répartis dans l'espace. Certains plans d'expériences exploratoires, parmi les plus classiques, sont présentés ci-dessous. Pour la suite, on note d la dimension de l'espace des paramètres et n la taille du plan d'expériences, autrement dit le nombre d'expériences associées à ce plan. L'intervalle de variation des différents paramètres incertains est normalisé entre 0 et 1.

Plan d'expériences standard

Un premier type de plan d'expériences, dit standard ou tabulé, peut être considéré lorsque l'espace des paramètres est discrétisé de manière régulière. Chaque paramètre peut alors prendre k valeurs différentes, appelées facteurs. Dans ce cas, le plan d'expériences le plus intuitif est probablement le plan factoriel. Il consiste simplement à considérer un point d'expérience à chaque noeud du découpage ainsi obtenu. Un tel plan remplit relativement bien l'espace si un nombre de facteurs suffisant est considéré, mais peut également conduire à un nombre d'évaluations $n = k^d$ très important. Lorsque les simulations réalisées pour chacun des points du plan correspondent à des codes de calcul coûteux, le nombre n ne doit pas être trop important. Il faut donc que le nombre de facteurs reste faible ($k = 2$ ou 3).

De nombreux autres plans standards sont décrits dans la littérature, tels ceux introduits par [Box and Behnken, 1960], [Plackett and Burman, 1946] ou encore [Doehlert, 1970]. Une représentation en dimension 3 est donnée en exemple pour certains de ces plans avec la Figure 2.2. Les plans d'expériences standards sont souvent proposés dans les logiciels pour la construction de méta-modèles. En effet, ils ont l'avantage d'être simples à calculer et faciles à appréhender pour l'utilisateur. Toutefois, ils sont peu adaptés à la méta-modélisation car ils n'offrent pas une grande flexibilité dans le choix du nombre d'expériences n . De plus, comme pour les plans factoriels, n est généralement important lorsque la dimension d est élevée.

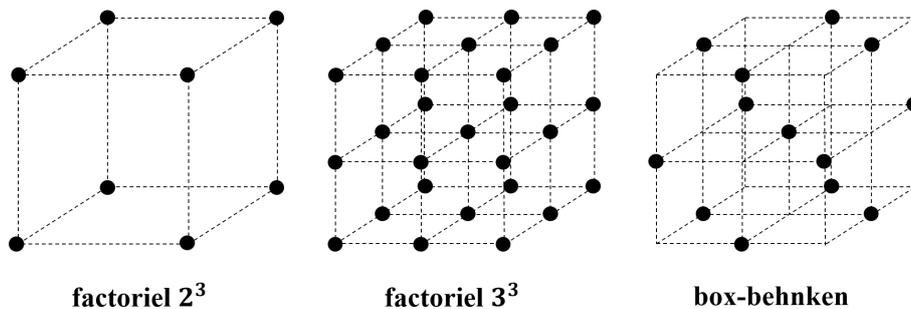


Figure 2.2 : Exemples de plan d'expériences standards représentés en dimension 3. A gauche et au centre, les plans factoriels considèrent un point d'expérience pour chacun des noeuds du découpage de l'espace. A droite, le plan de box-behnken définit des points d'expériences sur une partie des noeuds associés au découpage factoriel avec $k = 3$.

Plan *space-filling*

Considérons maintenant l'espace des paramètres comme continu. Les plans d'expériences *space-filling* désignent des plans d'expériences exploratoires particulièrement adaptés à la méta-modélisation [Pronzato and Müller, 2012]. Ce sont des plans pour lesquels le placement des points d'expériences est optimisé selon un critère géométrique ou statistique. L'optimisation est aussi dépendante du nombre d'expériences n et de la dimension d considérés.

Un critère géométrique couramment utilisé consiste à maximiser la distance minimale entre les différents points du plan d'expériences. On parle alors de critère « maximin » [Johnson et al., 1990]. Un autre critère cherche à minimiser la distance maximale entre les points du plan et les points de l'espace. En d'autres termes, quel que soit le point de l'espace considéré, il faut que la distance entre ce point et le point du plan le plus proche soit aussi petite que possible. Il s'agit du critère « minmax » [Johnson et al., 1990]. La Figure 2.3 donne des exemples de plans construits à partir de ces deux critères géométriques. Un inconvénient de ces plans est qu'ils peuvent être coûteux à construire, en particulier en grande dimension.

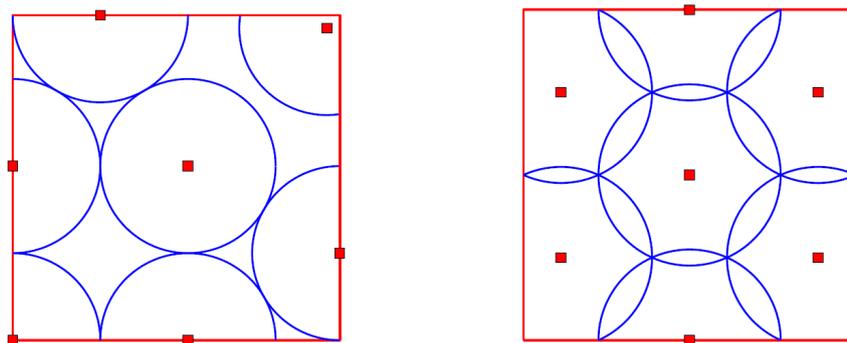


Figure 2.3 : Exemples de plan d'expériences de 7 points en dimension 2 construits selon les critères maximin (gauche) et minmax (droite). Source: [Pronzato and Müller, 2012] d'après <http://www.packomania.com/> et [Johnson et al., 1990].

Les critères statistiques employés sont généralement basés sur la discrédance [Fang, 2000]. Dans ce contexte, la discrédance désigne une mesure de l'écart entre la répartition des points du plan et une répartition uniforme. Elle peut se mesurer en comptant le nombre de points se trouvant à l'intérieur d'un volume défini par des intervalles de l'espace. Ainsi, plus la discrédance est faible, plus le nombre de points dans le volume tend à rester constant, quel que soit l'emplacement du volume dans l'espace.

Des plans d'expériences minimisant la discrédance peuvent être construits grâce à des suites à faible discrédance telles que les suites de [Sobol', 1967] ou [Halton, 1960]. On parle aussi de suites quasi aléatoires. Elles ont pour propriété de minimiser la discrédance lorsqu'elles tendent vers l'infini. Elles permettent ainsi de remplir l'espace de manière très régulière. Une comparaison entre un plan obtenu par tirage aléatoire (*i.e.* un plan non *space-filling*) et un plan obtenu par l'utilisation d'une suite de Sobol' est proposée sur la Figure 2.4. Les suites à faible discrédance sont fréquemment utilisées au sein de méthodes de Monte Carlo pour le calcul numérique d'intégrales.

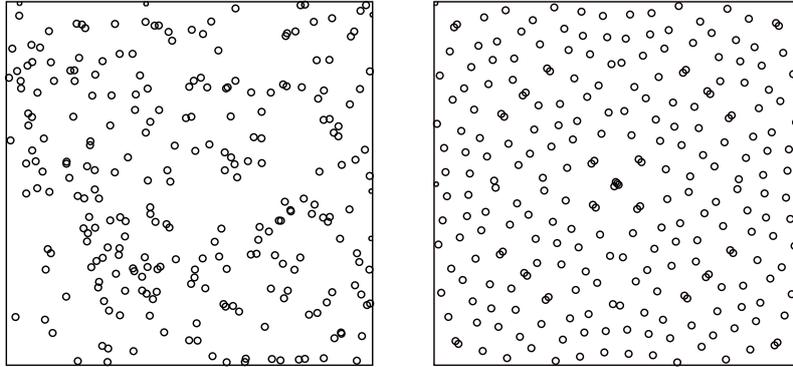


Figure 2.4 : A gauche, un plan d'expériences obtenu par tirage aléatoire. Le plan comporte de nombreux espaces vides et amas de points. A droite, un plan d'expériences issu d'une suite de Sobol'. La répartition des points est très régulière. Source : https://en.wikipedia.org/wiki/Sobol'_sequence, modifiée.

Une autre classe importante de plan d'expériences *space-filling* est obtenue par l'utilisation de l'échantillonnage par hypercube latin [Mckay et al., 1979]. C'est cette dernière méthode que nous utilisons dans cette thèse. Elle est décrite au sein de la section 2.2.1.

Pour plus de détails sur les différents plans d'expériences *space-filling* et leurs propriétés, le lecteur peut se référer à [Pronzato and Müller, 2012].

2.2 Méta-modélisation dans un cadre simple fidélité

Cette partie décrit les méthodes considérées dans cette thèse pour construire un méta-modèle dans un cadre simple fidélité. Ce cadre, le plus classique, signifie que la réponse d'intérêt n'est évaluée que sur un unique niveau de précision. Dans ce contexte, nous utilisons pour la planification d'expériences l'échantillonnage par hypercube latin décrit dans la section 2.2.1. Les méta-modèles sont ensuite construits selon la technique du krigeage dont la formulation est décrite dans la section 2.2.2.

2.2.1 Echantillonnage par hypercube latin

Les plans d'expériences considérés dans les différentes études réalisées ici sont issus d'un échantillonnage par hypercube latin, ou LHS pour *Latin Hypercube Sampling*. Cette technique, introduite en 1979 par [Mckay et al., 1979], est devenue populaire en ingénierie notamment après son utilisation par [Sacks et al., 1989] pour la planification et l'analyse d'expériences. Elle est l'une des approches les plus populaires pour la construction de plans d'expériences dans un cadre de méta-modélisation.

Principe

Une grille carrée est un carré latin si et seulement si elle possède un unique échantillon par colonne et par ligne. Un hypercube latin est la généralisation de ce concept à un nombre de dimensions quelconque. On considère que l'intervalle de variation des différents paramètres incertains est normalisé entre 0 et 1. Pour construire un LHS de n points en dimension d , chaque dimension est divisée de manière régulière en n intervalles : $[0, 1/n]$, $[1/n, 2/n]$, ..., $[(n-1)/n, 1]$. Ces intervalles définissent des strates. Les points du plan d'expériences sont ensuite tirés aléatoirement de manière à ce que chaque strate ne comporte qu'un seul point. Des exemples de LHS en dimension 2 sont donnés sur la Figure 2.5.

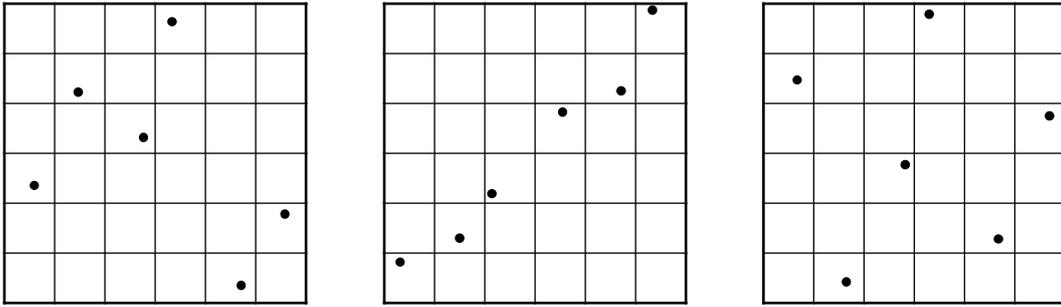


Figure 2.5 : Exemples de LHS à 6 points en dimension 2. Un point d'expérience au sein d'un LHS est tiré aléatoirement de manière à n'apparaître qu'une seule fois par strate (ou ligne/colonne en dimension 2). Un LHS ne remplit pas nécessairement l'espace de manière uniforme.

L'utilisation d'un LHS présente plusieurs avantages pour l'ingénieur. C'est un plan d'expériences pratique car il permet le choix du nombre désiré d'expériences, il est rapide à construire, et il se comporte bien en sous-projection. En particulier, lors d'une projection unidimensionnelle, les points sont bien répartis dans l'espace car par définition, il y a un point par strate. Par ailleurs, la construction du plan garantit aussi que deux points d'expériences ne peuvent pas se superposer lors d'une projection vers des sous-espaces de faibles dimensions. Ces dernières remarques sont importantes car elles signifient que si certains paramètres s'avèrent finalement non influents sur la réponse d'intérêt, alors la projection du LHS dans les dimensions restantes peut constituer un plan d'expériences toujours pertinent pour la construction d'un méta-modèle.

LHS optimisé

Un LHS ne définit pas en lui-même un plan d'expériences de type *space-filling*. En effet, il existe une infinité de LHS possibles pour n et d fixés. Tous ne sont pas de bons plans d'expériences, à l'image de l'exemple central de la Figure 2.5, qui laisse de très larges parties de l'espace inexploitées. Il est cependant possible de construire des LHS optimisés selon l'un des critères géométrique ou statistique présentés dans la partie 2.1. De tels LHS sont alors des plans d'expériences *space-filling* bien adaptés à la méta-modélisation.

Dans nos différentes études, nous utilisons exclusivement des LHS optimisés selon le critère maximin (maximisation de la distance minimale entre les points du plan d'expériences). Un exemple de tel LHS est donné par la Figure 2.6. Le LHS maximin a été introduit par [Morris and Mitchell, 1995] qui montrent la pertinence d'un tel plan d'expériences pour la construction de méta-modèles prédictifs par krigeage.

On remarque que la construction d'un LHS optimisé n'est pas nécessairement coûteuse en temps de calcul. Une méthode rapide pour construire un LHS optimisé consiste à générer un grand nombre de LHS, puis à sélectionner celui qui respecte le mieux le critère choisi. Par la suite, nous utiliserons

la fonction `maximinLHS` du package R `lhs` [Carnell, 2012], et le terme LHS désignera un LHS optimisé selon le critère maximin.

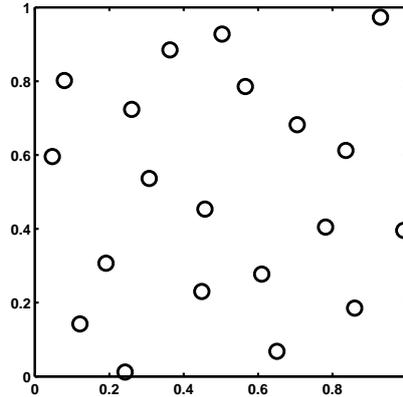


Figure 2.6 : Un LHS de 20 points en dimension 2 optimisé selon le critère maximin.

2.2.2 Méta-modèle de krigeage

Nous proposons dans cette section une description du krigeage inspirée notamment par les ouvrages de [Cressie, 1993], [Stein, 1999], [Santner et al., 2003] et [Rasmussen and Williams, 2006]. Le lecteur est invité à s’y référer pour plus de précisions.

Introduction

Le krigeage est une technique géostatistique qui permet d’interpoler la valeur d’une réponse à une position où elle n’est pas connue, à partir d’observations existantes de cette réponse. La théorie du krigeage a été formalisée par le mathématicien [Matheron, 1963]. Il s’est lui-même inspiré des travaux de [Krige, 1951], un ingénieur minier sud-africain qui a mis au point une méthode empirique d’estimation des teneurs en or.

Le krigeage est par la suite devenu une méthode classique pour la construction de méta-modèles, trouvant ainsi des applications dans de nombreux domaines, notamment les codes de calcul. Il est alors utilisé pour prédire la valeur d’une fonction, dont les évaluations sont coûteuses en temps de calcul, à partir des évaluations existantes de celle-ci. Cette méthode est également connue sous le nom de régression par processus Gaussien.

Soit une fonction $y(\mathbf{x})$ avec $\mathbf{x} \in \mathbb{R}^d$ et \mathbf{y} un vecteur de n valeurs observées de $y(\mathbf{x})$ sur $\mathbf{D} = \{\mathbf{x}_1; \dots; \mathbf{x}_n\}$, un plan d’expériences de dimension $n \times d$. On suppose alors que $y(\mathbf{x})$ est la réalisation d’un processus Gaussien $Y(\mathbf{x})$ de la forme suivante [Cressie, 1993] :

$$Y(\mathbf{x}) = h(\mathbf{x}) + Z(\mathbf{x}) + \epsilon(\mathbf{x}) \quad (2.1)$$

avec $h(\mathbf{x})$ une fonction qui approche la réponse moyenne de $Y(\mathbf{x})$, $Z(\mathbf{x})$ un terme représentant l’écart à la réponse moyenne et $\epsilon(\mathbf{x})$ un terme correspondant au bruit sur la réponse. $Z(\mathbf{x})$ est modélisé par un processus Gaussien de moyenne nulle et de fonction de covariance $K(\mathbf{x}, \mathbf{x}')$ telle que $K(\mathbf{x}, \mathbf{x}') = \text{Cov}(Z(\mathbf{x}), Z(\mathbf{x}')) = \sigma^2 R(\mathbf{x}, \mathbf{x}')$ avec σ^2 la variance du processus et $R(\mathbf{x}, \mathbf{x}')$ sa fonction de corrélation. $\epsilon(\mathbf{x})$ peut être modélisé par un processus Gaussien de moyenne nulle et indépendant de $Z(\mathbf{x})$.

La fonction de covariance

Afin d'obtenir le meilleur modèle possible, il est nécessaire que le comportement local de la fonction de covariance puisse refléter le comportement local de la réponse que l'on cherche à modéliser. Le choix de la fonction de covariance est ainsi un élément important du krigeage. De nombreuses fonctions de covariance sont proposées dans la littérature. Parmi elles, les plus communément utilisées sont les fonctions de covariance Matérn, Gaussienne, exponentielle ou encore sphérique. Une description de ces différentes fonctions ainsi que de leurs avantages et limites peut être trouvée dans [Chilès and Delfiner, 1999] ou [Rasmussen and Williams, 2006].

Les fonctions de covariance $K(\mathbf{x}, \mathbf{x}')$ peuvent souvent être exprimées comme un produit de fonctions de covariance monodimensionnelles k_j :

$$K(\mathbf{x}, \mathbf{x}') = \sigma^2 \prod_{j=1}^d k_j(x_j, x'_j) \quad (2.2)$$

Dans notre étude, nous avons choisi d'utiliser des fonctions de covariance Matérn [Matérn, 1986]. L'utilisation de ces dernières est notamment recommandée par [Stein, 1999] car elles permettent une grande flexibilité. Plus précisément, nous utilisons Matérn 5/2 qui s'écrit de la manière suivante :

$$k_j(x_j, x'_j) = \left(1 + \frac{\sqrt{5}|x_j - x'_j|}{\lambda_j} + \frac{5(x_j - x'_j)^2}{3\lambda_j^2} \right) \exp\left(-\frac{\sqrt{5}|x_j - x'_j|}{\lambda_j}\right) \quad (2.3)$$

Les paramètres de la fonction de covariance $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_d)$ et σ^2 sont également appelés hyper-paramètres afin de souligner qu'ils sont les paramètres d'un modèle qui n'est pas paramétrique. Ils doivent donc être estimés. σ^2 est un hyper-paramètre associé à la variance, il contrôle l'amplitude de variation du processus Gaussien. Similairement, $\boldsymbol{\lambda}$ est un vecteur d'hyper-paramètres qui peuvent être vus comme des longueurs de corrélation dans chacune des dimensions de l'espace, c'est-à-dire des distances limites en dessous desquelles les observations sont corrélées.

Formulation du krigeage

On se place dans les conditions du krigeage simple en supposant que la réponse moyenne est nulle (*i.e.* $h(\mathbf{x}) = 0$) et qu'elle n'est pas bruitée ($\epsilon(\mathbf{x}) = 0$).

Sachant la forme de la fonction de covariance K et les observations \mathbf{y} , le processus Gaussien $Y(\mathbf{x})$, conditionné aux données \mathbf{y} , a pour moyenne $y(\mathbf{x})$ et variance $s^2(\mathbf{x})$ définies par :

$$y(\mathbf{x}) = \mathbf{r}^\top(\mathbf{x})\mathbf{R}^{-1}\mathbf{y} \quad (2.4)$$

et

$$s^2(\mathbf{x}) = \sigma^2 \left(1 - \mathbf{r}^\top(\mathbf{x})\mathbf{R}^{-1}\mathbf{r}(\mathbf{x}) \right) \quad (2.5)$$

avec R la matrice de corrélation de \mathbf{D} et $\mathbf{r}(\mathbf{x}) = [R(\mathbf{x}, \mathbf{x}_1), \dots, R(\mathbf{x}, \mathbf{x}_n)]^\top$ le vecteur de corrélation entre \mathbf{x} et \mathbf{D} . On remarque que pour pouvoir calculer $y(\mathbf{x})$ et $s^2(\mathbf{x})$, il faut encore estimer les hyper-paramètres de la fonction de covariance K .

$y(\mathbf{x})$ et $s^2(\mathbf{x})$ sont communément appelés l'estimateur de krigeage et la variance de krigeage. La

variance de krigeage peut être vue comme l'erreur quadratique moyenne de l'estimateur de krigeage, on la note aussi MSE (*Mean Squared Error*). Cette estimation de l'erreur est valable seulement si la réponse approchée est bien la réalisation d'un processus Gaussien. En pratique, ce n'est jamais le cas pour nos applications. Ainsi, cette dénomination peut être trompeuse car elle ne représente pas l'erreur réelle du modèle. En revanche, elle peut être vue comme une mesure de distance entre le point \mathbf{x} et les points de \mathbf{D} . On note également que son expression ne dépend pas des observations \mathbf{y} .

On se place maintenant dans le cas où la réponse moyenne n'est pas nulle. On cherche à la modéliser par une fonction polynomiale de forme donnée :

$$h(\mathbf{x}) = \mathbf{h}^\top(\mathbf{x})\boldsymbol{\beta} = \sum_{j=1}^k h_j(\mathbf{x})\beta_j \quad (2.6)$$

avec $\mathbf{h}^\top = (h_1(\cdot), \dots, h_k(\cdot))$ une fonction de régression polynomiale qui doit être définie au préalable et $\boldsymbol{\beta} = (\beta_1, \dots, \beta_k)^\top$ les coefficients de régression associés qui doivent être estimés. Dans ce cas, on parle de krigeage universel.

On peut alors montrer que la moyenne de krigeage $y(\mathbf{x})$ et la variance de krigeage $s^2(\mathbf{x})$ s'expriment de la manière suivante :

$$y(\mathbf{x}) = \mathbf{h}^\top(\mathbf{x})\boldsymbol{\beta} + \mathbf{r}^\top(\mathbf{x})\mathbf{R}^{-1}(\mathbf{y} - \mathbf{H}\boldsymbol{\beta}) \quad (2.7)$$

et

$$s^2(\mathbf{x}) = \sigma^2 \left(1 - \mathbf{r}^\top(\mathbf{x})\mathbf{R}^{-1}\mathbf{r}(\mathbf{x}) + Q \right) \quad (2.8)$$

avec \mathbf{H} la matrice correspondant aux valeurs de $\mathbf{h}^\top(\mathbf{D})$ et Q un terme additionnel de variance lié à l'estimation des paramètres de la réponse moyenne. Son expression est obtenue par maximisation de la vraisemblance : $Q = \left(\mathbf{h}(\mathbf{x}) - \mathbf{H}^\top\mathbf{R}^{-1}\mathbf{r}(\mathbf{x}) \right)^\top \left(\mathbf{H}^\top\mathbf{R}^{-1}\mathbf{H} \right)^{-1} \left(\mathbf{h}(\mathbf{x}) - \mathbf{H}^\top\mathbf{R}^{-1}\mathbf{r}(\mathbf{x}) \right)$.

Estimation des hyper-paramètres

Les hyper-paramètres qui doivent être estimés pour caractériser le méta-modèle sont les longueurs de corrélation $\boldsymbol{\lambda}$, la variance σ^2 et les coefficients de régression $\boldsymbol{\beta}$. Leurs estimateurs, respectivement notés $\hat{\boldsymbol{\lambda}}$, $\hat{\sigma}^2$ et $\hat{\boldsymbol{\beta}}$, sont généralement calculés par maximisation de la fonction de vraisemblance ou MLE (*Maximum Likelihood Estimate*) [Mardia and Marshall, 1984]. En supposant que $Y(\mathbf{D})$ suit une loi normale multidimensionnelle, le logarithme de la fonction de vraisemblance L associée prend l'expression suivante en ignorant les termes constants [Santner et al., 2003] :

$$\log \left(L(\mathbf{y}|\boldsymbol{\beta}, \sigma^2, \boldsymbol{\lambda}) \right) = -\frac{1}{2} \left(n \log \sigma^2 + \log (\det(\mathbf{R})) + \frac{(\mathbf{y} - \mathbf{H}\boldsymbol{\beta})^\top \mathbf{R}^{-1} (\mathbf{y} - \mathbf{H}\boldsymbol{\beta})}{\sigma^2} \right) \quad (2.9)$$

Si on fixe $\boldsymbol{\lambda}$, les estimations de $\boldsymbol{\beta}$ et σ^2 par maximisation de la vraisemblance sont données par :

$$\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}(\boldsymbol{\lambda}) = \left(\mathbf{H}^\top\mathbf{R}^{-1}\mathbf{H} \right)^{-1} \mathbf{H}^\top\mathbf{R}^{-1}\mathbf{y} \quad (2.10)$$

et

$$\hat{\sigma}^2 = \hat{\sigma}^2(\boldsymbol{\lambda}) = \frac{1}{n} \left(\mathbf{y} - \mathbf{H}\hat{\boldsymbol{\beta}} \right)^\top \mathbf{R}^{-1} \left(\mathbf{y} - \mathbf{H}\hat{\boldsymbol{\beta}} \right) \quad (2.11)$$

En remplaçant $\boldsymbol{\beta}$ et σ^2 par leurs estimateurs dans l'équation 2.9, on trouve l'expression suivante

pour le logarithme de la fonction de vraisemblance :

$$\log(L(\boldsymbol{\lambda})) = -\frac{1}{2} \left(n \log \hat{\sigma}^2 + \log(\det(\mathbf{R})) + n \right) \quad (2.12)$$

Cette dernière expression dépend uniquement de $\boldsymbol{\lambda}$. L'estimation par maximum de vraisemblance de $\boldsymbol{\lambda}$ est alors obtenue en minimisant $n \log \hat{\sigma}^2 + \log(\det(\mathbf{R}))$.

Par la suite, nous construisons les méta-modèles de krigeage en utilisant le package R **DiceKriging** [Roustant et al., 2012].

2.3 Méta-modélisation dans un cadre multi-fidélité

Un cadre multi-fidélité signifie ici que la réponse d'intérêt est disponible à différents niveaux de précision. Ces différents niveaux de réponses peuvent alors être mis à profit lors de la méta-modélisation. Pour cela, le workflow résumant la conception d'un méta-modèle en simple fidélité (Figure 2.1) doit être adapté à la multi-fidélité. Dans un contexte d'une multi-fidélité à deux niveaux, le workflow peut ainsi être résumé par le schéma de la Figure 2.7. En particulier, les méthodes concernant la planification d'expériences ainsi que la construction du méta-modèle doivent être modifiées.

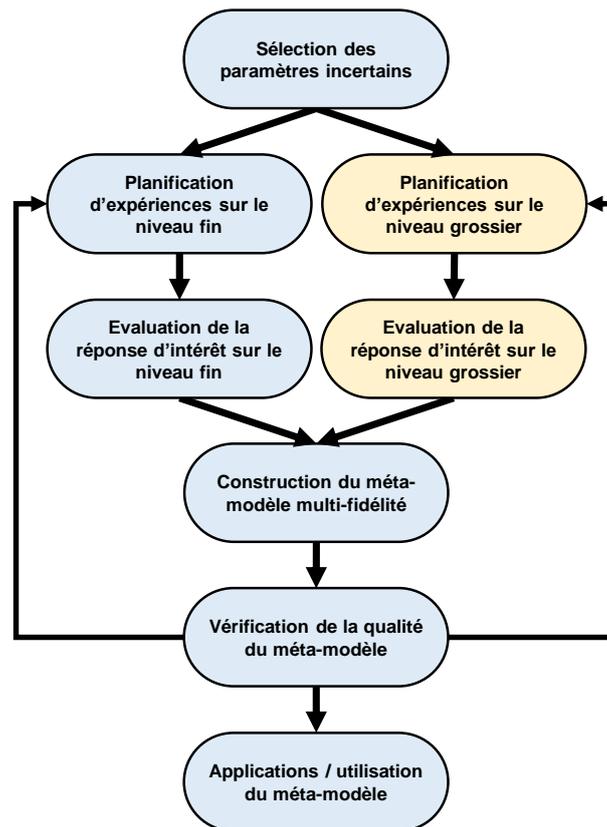


Figure 2.7 : Schéma d'un workflow basé sur la méta-modélisation dans un cadre multi-fidélité à deux niveaux.

La planification d'expériences doit notamment définir le niveau de fidélité sur lequel le point d'expérience doit être évalué. Nous utilisons pour cela l'échantillonnage par hypercubes latins imbriqués, tel que décrit dans la section 2.3.1. Le co-krigeage multi-fidélité est ensuite considéré pour la construc-

tion des méta-modèles (section 2.3.2). Il s'agit d'une extension du formalisme du krigeage permettant d'utiliser les observations obtenues à différents niveaux de résolution.

2.3.1 Echantillonnage par hypercubes latins imbriqués

Dans le cadre d'une approche multi-fidélité, la construction du plan d'expériences doit préciser pour chaque point le niveau de fidélité sur lequel la réponse y est évaluée. Il faut donc définir un plan d'expériences pour chaque niveau de fidélité considéré. On se place dans un cadre de multi-fidélité à deux niveaux, et on note \mathbf{D}_f le plan d'expériences pour le niveau fin et \mathbf{D}_g celui pour le niveau grossier. Il est possible d'utiliser des LHS de différentes tailles pour générer \mathbf{D}_f et \mathbf{D}_g et ainsi obtenir le plan d'expériences multi-fidélité de la taille souhaitée.

Pour la construction d'un méta-modèle par co-krigeage multi-fidélité, il est nécessaire d'avoir un plan d'expériences imbriqué, c'est-à-dire un plan d'expériences pour lequel les points sur les niveaux les plus fins sont aussi dans les plans d'expériences associés aux niveaux plus grossiers. Pour deux niveaux de fidélité, cela signifie que $\mathbf{D}_f \subset \mathbf{D}_g$.

On propose ici d'utiliser des LHS imbriqués selon le processus d'imbrication proposé dans [Le Gratiet and Garnier, 2014]. Ce dernier, illustré par la Figure 2.8, comprend les étapes suivantes :

1. Création d'un LHS pour le niveau fin (LHS fin).
2. Génération d'un ensemble de k LHS pour le niveau grossier (LHS grossiers).
3. Pour chacun de ces k LHS grossiers :
 - on cherche le point le plus proche de chaque point du LHS fin,
 - on calcule la distance séparant ces deux points et on somme les distances obtenues pour chaque couple de points afin de définir une mesure associée au LHS grossier.
4. On sélectionne le LHS grossier associé à la plus petite mesure.
5. Dans ce LHS, on remplace les points identifiés par les points du niveau fin correspondant. On obtient alors un LHS grossier modifié qui, avec le LHS fin calculé en 1, forme un couple de LHS imbriqués.

On note qu'une grande valeur de k permet de sélectionner un LHS grossier pour lequel de plus faibles modifications sont nécessaires. Ce LHS, une fois modifié, sera alors plus proche du LHS optimisé initial, et le LHS imbriqué correspondant sera donc de meilleure qualité. Par la suite, tous les LHS imbriqués utilisés sont construits avec $k = 1000$.

On remarque que cette procédure d'imbrication peut être appliquée de manière récursive afin de construire des LHS imbriqués pour un nombre quelconque de niveaux de fidélité. Pour cela, il suffit de l'appliquer en prenant le LHS grossier modifié obtenu lors de l'imbrication précédente comme nouveau LHS « fin ».

Comme les LHS, les LHS imbriqués sont très pratiques d'utilisation. Ils sont rapides à construire et permettent le choix au préalable du nombre de points d'expériences pour chacun des niveaux de fidélité considérés.

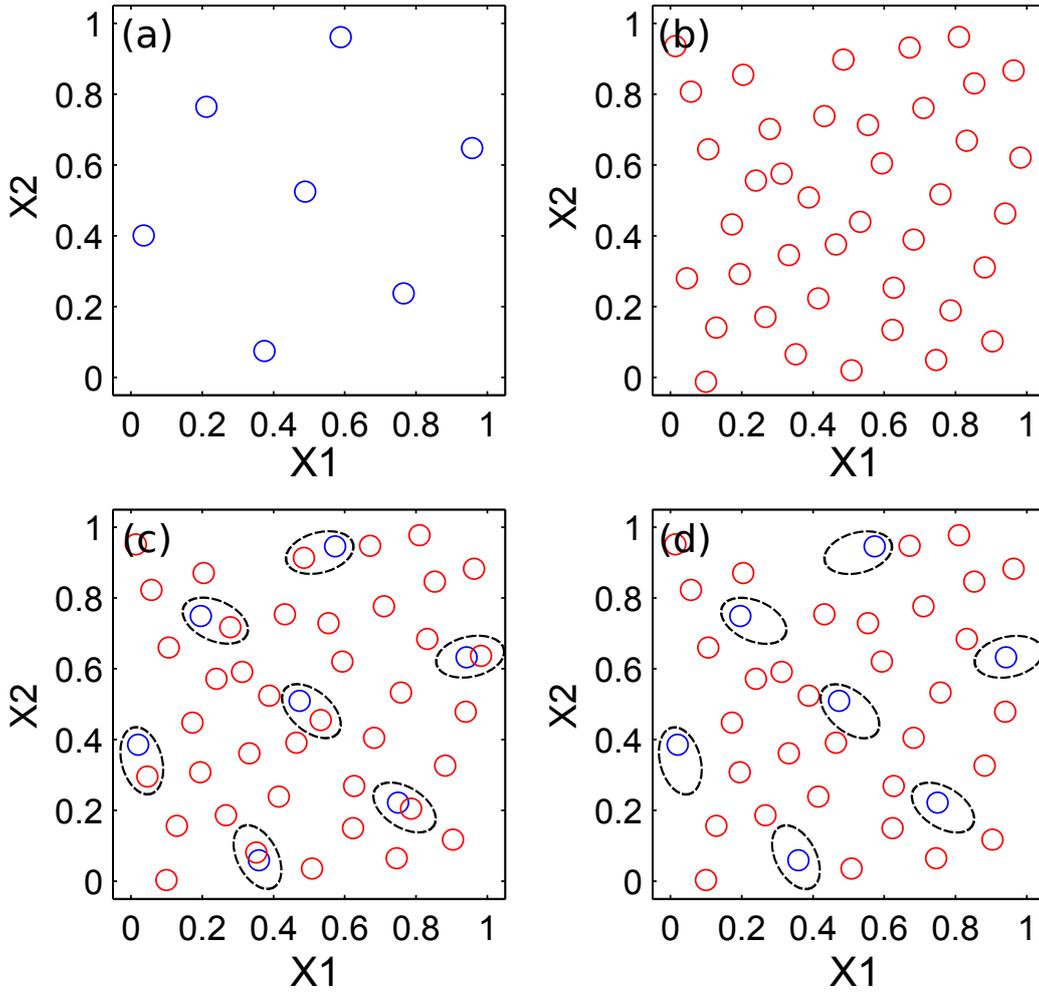


Figure 2.8 : Procédure d'imbrication de LHS pour deux niveaux de fidélité. (a) LHS fin, (b) LHS grossier, (c) on sélectionne les points grossiers les plus proches des points fins, (d) LHS grossier modifié résultant.

2.3.2 Méta-modèle de co-krigeage multi-fidélité

En pratique, un code de calcul peut souvent être lancé pour des niveaux de complexité différente. Différents niveaux de fidélité, correspondant à une modélisation plus ou moins fine de la réponse d'intérêt, peuvent ainsi être obtenus. L'idée est alors d'utiliser ces différents niveaux d'information au sein d'un méta-modèle pour approcher la réponse d'intérêt sur le niveau le plus fin, tout en réduisant les temps de calcul. L'approche multi-fidélité vise en effet à remplacer des évaluations sur le niveau de référence qui sont coûteuses en temps de calcul par des évaluations sur des niveaux plus rapides à évaluer. Un tel méta-modèle multi-fidélité peut être construit en adaptant le formalisme du krigeage à des systèmes multi-fidélité : on parle alors de co-krigeage multi-fidélité [Kennedy and O'Hagan, 2000a].

Le modèle auto-régressif proposé par [Kennedy and O'Hagan, 2000a] permet l'intégration d'un nombre quelconque s de niveaux de fidélité. Par souci de simplicité, nous considérons ici le cas où $s = 2$. Nous avons donc deux niveaux de fidélité que nous repérons par les indices f et g , pour le niveau fin et le niveau grossier, respectivement. Le niveau fin est le niveau de référence sur lequel on souhaite approcher la fonction d'intérêt : on a donc $y_f = y$. Sur le niveau grossier, y_g correspond à une réponse dégradée de la fonction d'intérêt.

On note $\mathbf{y}_f = y_f(\mathbf{D}_f)$ un vecteur de n_f valeurs observées de y sur le niveau fin aux points $\mathbf{D}_f =$

$\{\mathbf{x}_1; \dots; \mathbf{x}_{n_f}\}$, et $\mathbf{y}_g = y_g(\mathbf{D}_g)$ un vecteur de n_g valeurs observées sur le niveau grossier aux points $\mathbf{D}_g = \{\mathbf{x}_1; \dots; \mathbf{x}_{n_g}\}$. On choisit un plan d'expériences imbriqué, ainsi \mathbf{D}_f et \mathbf{D}_g respectent $\mathbf{D}_f \subset \mathbf{D}_g$ et $n_g > n_f$. Il est nécessaire que le plan d'expériences soit imbriqué pour pouvoir construire un méta-modèle de co-krigeage multi-fidélité. Toutefois, lorsque ce n'est pas le cas, il est toujours possible de se ramener à un plan imbriqué en estimant la réponse d'intérêt aux points d'imbrication par krigeage sur le niveau grossier.

Pour la formulation du co-krigeage multi-fidélité, on note \mathbf{D} l'ensemble des points d'observation de nos réponses pour les deux niveaux et \mathbf{y} le vecteur rassemblant ces observations :

$$\mathbf{D} = \begin{pmatrix} \mathbf{D}_g \\ \mathbf{D}_f \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_{n_g} \\ \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_{n_f} \end{pmatrix} \quad \text{et} \quad \mathbf{y} = \begin{pmatrix} y_g(\mathbf{D}_g) \\ y_f(\mathbf{D}_f) \end{pmatrix} = \begin{pmatrix} y_g(\mathbf{x}_1) \\ \vdots \\ y_g(\mathbf{x}_{n_g}) \\ y_f(\mathbf{x}_1) \\ \vdots \\ y_f(\mathbf{x}_{n_f}) \end{pmatrix} \quad (2.13)$$

Afin d'utiliser les données issues du niveau grossier, on formule un système qui va permettre d'intégrer la différence entre nos réponses aux niveaux fin et grossier. Plus précisément, on approche la réponse au niveau fin par un processus Gaussien $Y_f(\cdot)$ défini selon la combinaison linéaire suivante :

$$Y_f(\mathbf{x}) = \rho Y_g(\mathbf{x}) + Y_d(\mathbf{x}) \quad (2.14)$$

où ρ est un facteur d'échelle et $Y_d(\cdot)$ un processus Gaussien représentant la différence entre $\rho Y_g(\cdot)$ et $Y_f(\cdot)$.

Ce modèle auto-régressif est en réalité déduit de la propriété de Markov introduite par [Kennedy and O'Hagan, 2000a]:

$$\text{Cov}(Y_f(\mathbf{x}), Y_g(\mathbf{x}') | Y_g(\mathbf{x})) = 0 \quad \forall \mathbf{x} \neq \mathbf{x}' \quad (2.15)$$

Cette relation implique que si $Y_g(\mathbf{x})$ est connu, alors aucune nouvelle évaluation de $Y_g(\mathbf{x}')$ en $\mathbf{x}' \neq \mathbf{x}$ n'apporte d'information complémentaire sur $Y_f(\mathbf{x})$.

Formulation du co-krigeage multi-fidélité

Pour construire un modèle de co-krigeage multi-fidélité approchant y , la première étape consiste à construire un modèle de krigeage approchant le niveau grossier y_g . On note la matrice de covariance correspondante $K_g(\mathbf{x}, \mathbf{x}') = \text{Cov}(Y_g(\mathbf{x}), Y_g(\mathbf{x}')) = \sigma_g^2 R_g(\mathbf{x}, \mathbf{x}')$.

En reprenant les résultats pour le krigeage universel (Équation 2.7), on peut écrire que :

$$y_g(\mathbf{x}) = \mathbf{h}_g^\top(\mathbf{x})\boldsymbol{\beta}_g + \mathbf{r}_g^\top(\mathbf{x})\mathbf{R}_g^{-1} \left(\mathbf{y}_g - \mathbf{H}_g\boldsymbol{\beta}_g \right) \quad (2.16)$$

avec \mathbf{H}_g la matrice correspondant aux valeurs de $h_g^\top(\mathbf{D}_g)$ et \mathbf{R}_g la matrice de corrélation pour les points de \mathbf{D}_g selon la fonction K_g .

On construit ensuite le modèle de krigeage associé au processus Gaussien $Y_d(\cdot)$. La matrice de covariance correspondante est notée $K_d(\mathbf{x}, \mathbf{x}') = \text{Cov}(Y_d(\mathbf{x}), Y_d(\mathbf{x}')) = \sigma_d^2 R_d(\mathbf{x}, \mathbf{x}')$.

En posant $\mathbf{d} = \mathbf{y}_f - \rho y_g(\mathbf{D}_f)$, il vient que la moyenne de ce processus s'écrit :

$$y_d(\mathbf{x}) = \mathbf{h}_d^\top(\mathbf{x})\boldsymbol{\beta}_d + \mathbf{r}_d^\top(\mathbf{x})\mathbf{R}_d^{-1}(\mathbf{d} - \mathbf{H}_d\boldsymbol{\beta}_d) \quad (2.17)$$

avec \mathbf{H}_d la matrice correspondant aux valeurs de $h_g^\top(\mathbf{D}_f)$ et \mathbf{R}_d la matrice de corrélation entre les points de \mathbf{D}_f selon la fonction K_d .

Finalement, en utilisant les équations précédentes dans l'équation 2.14, il vient que le processus Gaussien $Y_f(\mathbf{x})$, conditionné aux données y , a pour moyenne

$$y_f(\mathbf{x}) = \mathbf{h}^\top(\mathbf{x})\boldsymbol{\beta} + \mathbf{k}^\top(\mathbf{x})\mathbf{K}^{-1}(\mathbf{y} - \mathbf{H}\boldsymbol{\beta}) \quad (2.18)$$

et pour variance

$$s_f^2(\mathbf{x}) = \rho^2\sigma_g^2 + \sigma_d^2 - \mathbf{k}^\top(\mathbf{x})\mathbf{K}^{-1}\mathbf{k}(\mathbf{x}) \quad (2.19)$$

avec $\mathbf{h}^\top(\mathbf{x}) = (\rho\mathbf{h}_g^\top(\mathbf{x}), \mathbf{h}_d^\top(\mathbf{x}))$, $\hat{\boldsymbol{\beta}} = \begin{pmatrix} \boldsymbol{\beta}_g \\ \boldsymbol{\beta}_d \end{pmatrix}$ et $\mathbf{G} = \begin{pmatrix} \mathbf{h}_g^\top(\mathbf{D}_g) & \mathbf{0} \\ \rho\mathbf{h}_g^\top(\mathbf{D}_f) & \mathbf{h}_d^\top(\mathbf{D}_f) \end{pmatrix}$. Le vecteur de covariance entre \mathbf{x} et \mathbf{D} s'écrit $\mathbf{k} = \begin{pmatrix} \rho\sigma_g^2 R_g(\mathbf{D}_g, \mathbf{x}) \\ \rho\sigma_g^2 R_g(\mathbf{D}_f, \mathbf{x}) + \sigma_d^2 R_d(\mathbf{D}_f, \mathbf{x}) \end{pmatrix}$ et K la matrice de covariance prend l'expression suivante :

$$\mathbf{K} = \begin{pmatrix} \sigma_g^2 \mathbf{R}_g & \rho\sigma_g^2 R_g(\mathbf{D}_g, \mathbf{D}_f) \\ \rho\sigma_g^2 R_g(\mathbf{D}_f, \mathbf{D}_g) & \rho^2\sigma_g^2 R_g(\mathbf{D}_f, \mathbf{D}_f) + \sigma_d^2 \mathbf{R}_d \end{pmatrix} \quad (2.20)$$

Estimation des hyper-paramètres

Pour compléter l'estimateur de co-krigeage multi-fidélité (Équation 2.18), un ensemble d'hyper-paramètres doit encore être estimé : $\boldsymbol{\beta}_g$, σ_g^2 , $\boldsymbol{\lambda}_g$, $\boldsymbol{\beta}_d$, σ_d^2 , $\boldsymbol{\lambda}_d$ et ρ . La condition d'indépendance entre $Y_g(\cdot)$ et $Y_d(\cdot)$ permet d'estimer séparément les paramètres relatifs au niveau grossier ($\boldsymbol{\beta}_g$, σ_g^2 , $\boldsymbol{\lambda}_g$) et ceux relatifs au niveau fin ($\rho, \boldsymbol{\beta}_d$, σ_d^2 , $\boldsymbol{\lambda}_d$). Comme suggéré par [Forrester et al., 2007], ces estimations peuvent se faire par maximisation de la fonction de vraisemblance, de façon similaire au krigeage.

En effet, pour $(\boldsymbol{\beta}_g, \sigma_g^2, \boldsymbol{\lambda}_g)$, en supposant que $Y_g(\mathbf{D}_g)$ suit une loi normale multidimensionnelle, le logarithme de la fonction de vraisemblance L_g associée s'écrit :

$$\log(L_g(\mathbf{y}_g | \boldsymbol{\beta}_g, \sigma_g^2, \boldsymbol{\lambda}_g)) = -\frac{1}{2}(n_g \log \sigma_g^2 + \log(\det(\mathbf{R}_g))) + \frac{(\mathbf{y}_g - \mathbf{H}_g\boldsymbol{\beta}_g)^\top \mathbf{R}_g^{-1}(\mathbf{y}_g - \mathbf{H}_g\boldsymbol{\beta}_g)}{\sigma_g^2} \quad (2.21)$$

En fixant $\boldsymbol{\lambda}_g$, les estimations de $\boldsymbol{\beta}_g$ et σ_g^2 par maximisation de la vraisemblance sont données par :

$$\hat{\boldsymbol{\beta}}_g = \hat{\boldsymbol{\beta}}_g(\boldsymbol{\lambda}_g) = (\mathbf{H}_g^\top \mathbf{R}_g^{-1} \mathbf{H}_g)^{-1} \mathbf{H}_g^\top \mathbf{R}_g^{-1} \mathbf{y}_g \quad (2.22)$$

et

$$\hat{\sigma}_g^2 = \hat{\sigma}_g^2(\boldsymbol{\lambda}_g) = \frac{1}{n_g} (\mathbf{y}_g - \mathbf{H}_g \hat{\boldsymbol{\beta}}_g)^\top \mathbf{R}_g^{-1} (\mathbf{y}_g - \mathbf{H}_g \hat{\boldsymbol{\beta}}_g) \quad (2.23)$$

En remplaçant $\boldsymbol{\beta}_g$ et σ_g^2 dans l'Equation 2.21 par leurs estimations respectives (Equations 2.22 et 2.23), on trouve l'expression suivante pour le logarithme de la fonction de vraisemblance :

$$\log \left(L_g(\mathbf{y}_g | \boldsymbol{\lambda}_g) \right) = -\frac{1}{2} \left(n_g \log \hat{\sigma}_g^2 + \log(\det(\mathbf{R}_g)) + n_g \right) \quad (2.24)$$

Cette expression dépend uniquement de $\boldsymbol{\lambda}_g$. L'estimation par maximum de vraisemblance de $\boldsymbol{\lambda}_g$ est finalement obtenue par la minimisation de $n_g \log \hat{\sigma}_g^2 + \log(\det(\mathbf{R}_g))$.

De la même manière, on peut estimer les quatre autres paramètres $(\rho, \boldsymbol{\beta}_d, \sigma_d^2, \boldsymbol{\lambda}_d)$ relatifs au modèle de krigeage appliqué à $\mathbf{d} = \mathbf{y}_f - \rho y_g(\mathbf{D}_f)$. Ainsi, en supposant que $Y_f(\mathbf{D}_f) - \rho Y_g(\mathbf{D}_f)$ suit une loi normale multidimensionnelle, le logarithme de la fonction de vraisemblance L_d associée s'écrit :

$$\log \left(L_d(\mathbf{d} | \rho, \boldsymbol{\beta}_d, \sigma_d^2, \boldsymbol{\lambda}_d) \right) = -\frac{1}{2} \left(n_f \log \sigma_d^2 + \log(\det(\mathbf{R}_d)) + \frac{(\mathbf{d} - \mathbf{H}_d \boldsymbol{\beta}_d)^\top \mathbf{R}_d^{-1} (\mathbf{d} - \mathbf{H}_d \boldsymbol{\beta}_d)}{\sigma_d^2} \right) \quad (2.25)$$

En fixant $\boldsymbol{\lambda}_d$, les estimations de $\boldsymbol{\beta}_d$ et σ_d^2 par maximisation de la vraisemblance sont données par :

$$\hat{\boldsymbol{\beta}}_d = \hat{\boldsymbol{\beta}}_d(\rho, \boldsymbol{\lambda}_d) = \left(\mathbf{H}_d^\top \mathbf{R}_d^{-1} \mathbf{H}_d \right)^{-1} \mathbf{H}_d^\top \mathbf{R}_d^{-1} \mathbf{d} \quad (2.26)$$

et

$$\hat{\sigma}_d^2 = \hat{\sigma}_d^2(\rho, \boldsymbol{\lambda}_d) = \frac{1}{n_f} \left(\mathbf{d} - \mathbf{H}_d \hat{\boldsymbol{\beta}}_d \right)^\top \mathbf{R}_d^{-1} \left(\mathbf{d} - \mathbf{H}_d \hat{\boldsymbol{\beta}}_d \right) \quad (2.27)$$

Puis, en remplaçant $\boldsymbol{\beta}_d$ et σ_d^2 dans l'Equation 2.25 par leurs estimations respectives (Equations 2.26 et 2.27), on trouve l'expression suivante pour le logarithme de la fonction de vraisemblance :

$$\log \left(L_d(\mathbf{d} | \rho, \boldsymbol{\lambda}_d) \right) = -\frac{1}{2} \left(n_f \log \hat{\sigma}_d^2 + \log(\det(\mathbf{R}_d)) + n_f \right) \quad (2.28)$$

Cette expression dépend uniquement de $\boldsymbol{\lambda}_d$ et ρ , qu'il est donc possible d'estimer en minimisant $n_f \log \hat{\sigma}_d^2 + \log(\det(\mathbf{R}_d))$. Pour davantage de détails sur l'estimation des hyper-paramètres d'un modèle de co-krigeage multi-fidélité par maximisation de la vraisemblance, le lecteur peut se référer à [Forrester et al., 2007].

Formulation récursive

Nous présentons ici la formulation récursive du modèle de co-krigeage multi-fidélité proposée par [Le Gratiet and Garnier, 2014]. Selon cette formulation, le modèle de co-krigeage multi-fidélité pour le niveau de fidélité s peut s'exprimer comme la somme d'un modèle de co-krigeage multi-fidélité au niveau $s-1$, et d'un modèle de krigeage de l'écart entre les deux niveaux. En outre, elle permet d'obtenir des expressions récursives pour la moyenne et la variance du co-krigeage multi-fidélité (Équations 2.29 et 2.30). Cette fois encore, nous nous concentrons sur le cas $s = 2$, où deux niveaux de fidélité sont considérés. On conserve également toutes les notations introduites précédemment.

Reprenant les équations 2.14 et 2.15, [Le Gratiet and Garnier, 2014] ont montré l'équivalence des expressions précédemment introduites (Équations 2.18 et 2.19) pour la moyenne $y_f(\mathbf{x})$ et la variance $s_f^2(\mathbf{x})$ du co-krigeage multi-fidélité avec les expressions suivantes :

$$y_f(\mathbf{x}) = \rho y_g(\mathbf{x}) + \mathbf{h}_d^\top(\mathbf{x}) \boldsymbol{\beta}_d + \mathbf{r}_d^\top(\mathbf{x}) \mathbf{R}_d^{-1} (\mathbf{d} - \mathbf{H}_d \boldsymbol{\beta}_d) \quad (2.29)$$

et

$$s_f^2(\mathbf{x}) = \rho^2 s_g^2(\mathbf{x}) + \sigma_d^2 \left(1 - \mathbf{r}_d^T(\mathbf{x}) \mathbf{R}_d^{-1} \mathbf{r}_d(\mathbf{x})\right) \quad (2.30)$$

Ce modèle récursif est toutefois plus général car il laisse la possibilité de renseigner un modèle de régression pour ρ , similairement à la modélisation proposée pour $h(\mathbf{x})$ (Equation 2.6). Le lecteur intéressé par cette approche pourra se référer aux travaux de [Le Gratiet and Garnier, 2014]. Une méthode Bayésienne d'estimation des hyper-paramètres associés à ce modèle est ensuite décrite dans [Le Gratiet, 2013a].

Un intérêt majeur de l'approche récursive est qu'elle réduit la complexité numérique du modèle de co-krigeage multi-fidélité au sens où elle réduit la taille des matrices de covariance à inverser. En effet, la récursivité implique que la construction d'un modèle de co-krigeage multi-fidélité pour s niveaux de fidélité est équivalente, en termes de complexité, à la construction de s modèles de krigeage successifs. Un autre avantage de la récursivité est que l'on dispose des estimateurs (moyenne et variance) du modèle pour les différents niveaux grossiers.

Par la suite, nous construisons les méta-modèles multi-fidélité en utilisant le modèle de co-krigeage multi-fidélité proposé par [Le Gratiet and Garnier, 2014]. Nous utilisons pour cela le package **R Mu-FiCokriging** [Le Gratiet, 2012] associé. Pour plus de détails sur le co-krigeage multi-fidélité, le lecteur peut se référer à la thèse de [Le Gratiet, 2013b].

2.4 Validation du méta-modèle

La vérification de la qualité du méta-modèle est également une étape importante dans une approche basée sur la méta-modélisation. En effet, un méta-modèle est pertinent seulement s'il est de bonne qualité, c'est-à-dire s'il prédit la réponse d'intérêt de manière fiable aux points de l'espace n'ayant pas été évalués. Il pourra alors être utilisé au sein de l'application envisagée. En revanche, si le méta-modèle n'est pas assez prédictif, il faut généralement reprendre son processus de construction en considérant un plan d'expériences de plus grande taille (*cf.* Figures 2.1 et 2.6). Ce dernier peut par exemple être augmenté en suivant une stratégie de planification séquentielle. Ce point fait l'objet du chapitre 5.

La qualité d'un méta-modèle peut être estimée en utilisant l'erreur de prédiction sur une base de test indépendante tel que décrit dans la section 2.4.1. Il est également possible de faire de la validation croisée comme expliqué dans la section 2.4.2.

2.4.1 Validation via une base de test

Pour estimer la qualité d'un méta-modèle, on peut évaluer la réponse d'intérêt en un ensemble de points indépendant du plan d'expériences utilisé pour la construction du méta-modèle - la base de test -, et calculer l'erreur de prédiction en ces points. Ces erreurs peuvent ensuite être regroupées dans un unique coefficient, par exemple le coefficient Q^2 défini par :

$$Q^2 = 1 - \frac{\sum_{i=1}^{n_{test}} (y(\mathbf{x}_i) - \hat{y}(\mathbf{x}_i))^2}{\sum_{i=1}^{n_{test}} (y(\mathbf{x}_i) - \bar{y})^2} \quad (2.31)$$

où on a noté $\mathbf{D}_{test} = \{\mathbf{x}_1; \dots; \mathbf{x}_{n_{test}}\}$ la base de test, $\hat{y}(\mathbf{x})$ l'estimateur de krigeage ou co-krigeage multi-fidélité de $y(\mathbf{x})$ (Equations 2.7 ou 2.29), et \bar{y} la valeur moyenne de $y(\mathbf{x})$ aux points de la base de test \mathbf{D}_{test} . Dans ce cas, ce coefficient peut être vu comme l'équivalent du coefficient de détermination classique R^2 appliqué aux résidus de prédiction sur la base de test.

On remarque que, plus un méta-modèle est prédictif, et plus le Q^2 est proche de 1. En pratique, on peut considérer qu'un méta-modèle est raisonnablement prédictif lorsque $Q^2 > 0,9$. Le Q^2 peut prendre des valeurs négatives, ce qui est caractéristique d'un mauvais méta-modèle puisque il est moins prédictif qu'une approximation par une constante égale à la réponse moyenne sur la base de test.

Le coefficient Q^2 est pratique car il est rapide à calculer et simple à interpréter. Toutefois, il faut garder en tête que sa valeur peut être fortement influencée par un résidu élevé pour un seul des points de la base de test. Un exemple de cas extrême serait d'obtenir un mauvais Q^2 alors qu'en réalité le méta-modèle est très prédictif pour l'ensemble des points de la base de test sauf un pour lequel il est très mauvais. Des vérifications complémentaires et régulières sont donc nécessaires pour détecter ce genre de cas.

2.4.2 Validation croisée

Les méthodes par validation croisée permettent d'estimer la fiabilité d'un méta-modèle en ne considérant que le plan d'expériences initial. On s'intéresse ici à la validation croisée lorsqu'une seule évaluation est mise à part, on parle aussi de LOO-CV (*Leave-One-Out Cross-Validation*). Prenons un plan d'expériences \mathbf{D} de n points et notons $y(\mathbf{x}_1), \dots, y(\mathbf{x}_n)$ les n évaluations correspondantes de la réponse d'intérêt $y(\mathbf{x})$. Le principe de la validation croisée consiste à supposer que l'évaluation i n'est pas connue et à construire un méta-modèle de y avec les $n - 1$ autres évaluations. On note $\hat{y}_{-i}(\mathbf{x})$ l'estimateur associé à ce méta-modèle.

On peut ensuite calculer l'erreur de validation croisée au point \mathbf{x}_i par $y(\mathbf{x}_i) - \hat{y}_{-i}(\mathbf{x}_i)$. En répétant cette opération pour les n points du plan d'expériences, on obtient une estimation de la fiabilité de notre méta-modèle. En particulier, ces erreurs peuvent être regroupées comme précédemment dans le calcul du coefficient Q^2 selon :

$$Q_{cv}^2 = 1 - \frac{\sum_{i=1}^n (y(\mathbf{x}_i) - \hat{y}_{-i}(\mathbf{x}_i))^2}{\sum_{i=1}^n (y(\mathbf{x}_i) - \bar{y})^2} \quad (2.32)$$

avec \bar{y} la valeur moyenne de $y(\mathbf{x})$ aux points du plan d'expériences \mathbf{D} .

Le coefficient Q_{cv}^2 s'interprète de la même manière que le Q^2 . Dans la pratique, nous avons observé que le Q_{cv}^2 prend des valeurs légèrement supérieures à celle du Q^2 pour un même plan d'expériences. On peut toutefois considérer le Q_{cv}^2 comme une bonne approximation du Q^2 . De plus, il présente l'avantage de pouvoir être calculé sans base de test supplémentaire, ce qui permet l'économie non négligeable de ces évaluations additionnelles.

La validation croisée nécessite néanmoins le calcul de n méta-modèles, ce qui peut avoir un coût important en temps de calcul pour un méta-modèle de krigeage. Cependant, différentes méthodes proposées par [Dubrule, 1983] [Rasmussen and Williams, 2006] [Zhang and Wang, 2010] proposent des raccourcis et des approximations afin de calculer rapidement les prédictions (et les variances associées) par validation croisée pour le krigeage. On peut par exemple éviter la ré-estimation de l'ensemble ou d'une partie des hyper-paramètres en supposant qu'une unique observation n'a que peu d'impact sur leur estimation.

La formulation récursive (Équation 2.29) permet d'étendre ces raccourcis au co-krigeage multi-fidélité. Ainsi, [Le Gratiet and Garnier, 2014] proposent une méthode de calcul rapide des prédictions et des variances par validation croisée pour le co-krigeage multi-fidélité. On note que, dans cette méthode, les hyper-paramètres λ ne sont pas ré-estimés. Par ailleurs, la validation croisée pour le co-krigeage multi-fidélité permet l'expression des résultats aux différents niveaux de fidélité. On peut donc

calculer la prédiction et la variance par validation au niveau grossier en considérant que l'observation à $\mathbf{x}_i \in \mathbf{D}_g$ est inconnue. Pour le calcul de ces résultats au niveau fin, on considère que l'observation à $\mathbf{x}_i \in \mathbf{D}_f$ est inconnue sur les deux niveaux de fidélité.

CHAPITRE 3

Méta-modélisation de sorties vectorielles et de la fonction objectif

Les propriétés d'intérêt considérées pour la méta-modélisation sont issues de la simulation d'écoulement. Ce sont donc des fonctions du temps, voire de l'espace, et leurs valeurs simulées peuvent être vues comme des vecteurs dont chaque composante correspond à un temps et/ou une maille donnée. La méta-modélisation de sorties vectorielles est ainsi un enjeu important en ingénierie de réservoir.

Ce chapitre est divisé en trois parties. Tout d'abord, nous résumons l'ensemble des propriétés et sorties d'intérêt de la simulation d'écoulement au sein de la partie 3.1. La méta-modélisation de ces sorties, généralement vectorielles, est un enjeu pour de nombreuses applications. Les méthodes proposées pour la méta-modélisation de sorties vectorielles dans des cadres simple ou multi-fidélité sont ensuite décrites en partie 3.2. Puis en partie 3.3, nous nous intéressons à la méta-modélisation de la fonction objectif quantifiant l'erreur entre données de production et réponses simulées. En particulier, nous proposons de l'approcher en utilisant les estimations des propriétés vectorielles à caler.

3.1 Propriétés d'intérêt de la simulation d'écoulement

En ingénierie de réservoir, on s'intéresse à l'évolution dynamique de la pression et des fluides dans le réservoir en réponse à un schéma de production. La grande majorité des sorties issues de la simulation d'écoulement sont donc des propriétés qui dépendent du temps. Ces propriétés sont calculées à chaque pas de temps considéré lors de la simulation d'écoulement. En pratique, on ne conserve leurs réponses qu'à certains temps, et notamment ceux pour lesquels on dispose de mesures de données dynamiques. Parmi les propriétés dynamiques d'intérêt, on peut distinguer les propriétés mesurées aux puits, les propriétés définies au niveau du réservoir, et les propriétés spatio-temporelles.

Les propriétés d'intérêt aux puits

Les propriétés d'intérêt aux puits regroupent notamment la pression en fond de puits (*bottomhole pressure*) et une mesure de la production des différents fluides, comme le débit d'huile. Pour l'eau et le gaz, on préfère souvent une expression de leur production sous la forme d'un rapport. Pour l'eau, on utilise la fraction d'eau (*watercut*). C'est une réponse prenant ses valeurs entre 0 et 1, qui désigne la part en eau produite par rapport à la part totale des fluides. Pour le gaz, on utilise fréquemment le ratio gaz/huile (*gaz/oil ratio*), qui désigne un rapport entre les volumes de gaz et d'huile produits en condition de pression et température de surface. Cette réponse permet de prendre en compte le gaz dissous au sein du pétrole brut produit. Selon le cas étudié et les options du simulateur d'écoulement,

de nombreuses autres sorties peuvent être simulées aux puits.

Les propriétés associées au réservoir

Ces propriétés représentent le comportement dynamique global au niveau du réservoir, et sont obtenues à partir des propriétés dynamiques aux différents puits. Il peut s'agir par exemple de débits ou de volumes cumulés pour un fluide donné, comme le volume cumulé de l'huile produite par le réservoir ou encore le débit total d'eau produite. Ces sorties fournissent des indicateurs globaux pouvant être utilisés dans des calculs économiques visant à estimer la valeur d'un schéma de production ou d'un emplacement pour un puits additionnel. Elles peuvent ainsi être utiles pour le choix d'une stratégie d'exploitation dans le cadre du développement du champ.

Les propriétés spatiales

Les propriétés spatiales sont définies dans chaque maille du réservoir. Il s'agit principalement de la distribution de la pression et des saturations en huile ou en eau dans le réservoir. La modélisation de ces propriétés est particulièrement intéressante lorsque l'on dispose de données sismiques répétées dans le temps, aussi appelées données de sismique 4D (variations d'impédance acoustique dans le temps par exemple). En effet, ces attributs sismiques peuvent être calculés via un modèle pétro-élastique à partir des cartes de pression et saturation fournies par la simulation d'écoulement. Les attributs sismiques obtenus peuvent ensuite être comparés aux données de sismique 4D au sein d'une fonction objectif adaptée et être pris en compte lors du processus de calage [Roggero et al., 2012].

La fonction objectif

La fonction objectif n'est pas une propriété de la simulation d'écoulement. C'est cependant une sortie d'intérêt majeure utilisée lors du calage d'historique afin de quantifier l'écart entre les données de production mesurées et celles simulées à l'aide du modèle réservoir. Nous en rappelons ci-dessous la définition introduite dans la section 1.2.2 :

$$FO(\mathbf{x}) = \frac{1}{2} \sum_{k=1}^K \omega_k \sum_{t=1}^{T(k)} \left(\frac{y_k(\mathbf{x}, t) - d_{kt}^{obs}}{\sigma_{kt}} \right)^2 \quad (3.1)$$

On rappelle que les données de production simulées sont notées $y(\mathbf{x}, t)$ et celles mesurées d^{obs} . Les indices $k \in [1, K]$ et $t \in [1, T(k)]$ font référence aux différentes sorties et aux différents temps pour lesquels ces sorties sont mesurées. Les coefficients ω et σ sont des poids définis par l'ingénieur. On note également que les données de production sont des réponses dynamiques. Leur dépendance au temps est indiquée par l'indice t . Les données de production sont généralement des propriétés aux puits, mais les propriétés associées au réservoir peuvent également être considérées dans le calcul de la fonction objectif.

Applications

La méta-modélisation de sorties vectorielles rend possible l'analyse de sensibilité simultanément sur les différents éléments du vecteur. Ainsi, dans [Douarce et al., 2014], une analyse de sensibilité réalisée au cours de la récupération de l'huile du réservoir permet de caractériser l'évolution de l'impact des différents surfactants et polymères testés dans le temps. De façon similaire, la méta-modélisation de