
Ingénierie de réservoir

Dans ce chapitre, nous proposons une introduction générale à l'ingénierie de réservoir afin d'exposer le contexte dans lequel s'inscrit cette thèse. Nous résumons les étapes clés dans la construction de modèles numériques représentatifs d'un réservoir, en nous inspirant de l'ouvrage [Le Ravalec et al., 2014]. L'objectif de la première étape est la conception de modèles à partir desquels il est possible de simuler l'exploitation d'un réservoir. La partie 1.1 en rappelle les points essentiels. Puis vient en partie 1.2 une description du calage d'historique, le processus visant à calibrer les modèles de réservoir pour qu'ils respectent les données de production. La calibration des modèles de réservoir, puis leur utilisation pour l'optimisation de la production, sont des opérations nécessitant de nombreuses simulations. Les simulations d'écoulement étant généralement très coûteuses en temps de calcul (jusqu'à plusieurs heures par simulation), il est nécessaire de réduire les temps de calcul afin de pouvoir optimiser la production tout en respectant les contraintes de temps des études opérationnelles. Nous abordons la réduction des temps de calcul en partie 1.3. Finalement, nous délimitons la problématique de cette thèse et détaillons ses objectifs dans la partie 1.4.

1.1 Modélisation du réservoir

La construction d'un modèle numérique représentatif d'un réservoir passe par l'intégration dans ce modèle de toutes les données disponibles sur le réservoir. Ce processus se fait traditionnellement en deux temps : d'abord l'intégration des données constantes dans le temps, les données statiques, puis celle des données variant dans le temps, les données dynamiques. Nous détaillons tout d'abord l'intégration des données statiques. Pour cela, la section 1.1.1 décrit très brièvement les bases de la caractérisation et de la modélisation statique du réservoir. Il s'agit ici de construire un modèle de réservoir à partir des données statiques. La simulation des écoulements au sein du modèle construit fait ensuite l'objet de la section 1.1.2.

1.1.1 Caractérisation et modélisation statique du réservoir

La première étape consiste donc à utiliser les données statiques disponibles pour construire un premier modèle numérique représentatif de notre réservoir. Ces données statiques comprennent la sismique 3D, des diagraphies et des mesures sur carottes qui ont pu être extraites lors de forages. La Figure 1.1 propose des exemples illustrés des différentes données statiques permettant la caractérisation du réservoir. L'interprétation de ces données combinée aux connaissances géologiques du terrain permet la création d'un modèle dit géologique. On parle également de géo-modèle ou modèle statique.

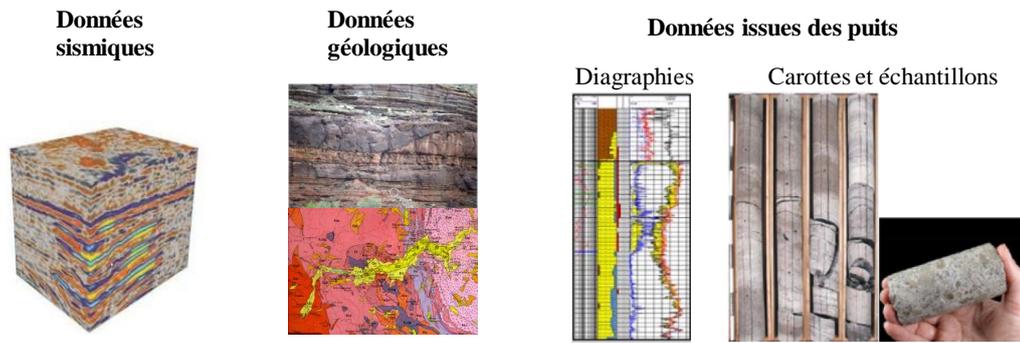


Figure 1.1 : Types de données statiques collectées pour caractériser les propriétés du réservoir. Source: [Le Ravalec et al., 2014], modifiée.

Données statiques

Les données statiques peuvent être mesurées directement ou indirectement. Elles renseignent des propriétés à des échelles très différentes allant du micromètre à l’hectomètre.

Un premier type de données, dites géologiques, fait référence aux connaissances géologiques préexistantes à l’échelle régionale ou du bassin. Souvent disponibles, ces données contribuent à l’identification des formations rocheuses comprises dans le réservoir étudié. Elles peuvent éventuellement apporter des indications sur la valeur de certaines propriétés pétrophysiques, ou sur la présence d’hétérogénéités ou structures particulières.

Les données sismiques sont des mesures indirectes de la structure du sous-sol. Elles sont obtenues par de la prospection sismique qui consiste à propager des ondes sismiques dans le sous-sol, puis à en enregistrer les échos provenant de la réflexion sur les structures géologiques en profondeur. Le traitement et l’analyse de ces enregistrements permettent la reconstitution d’une image du sous-sol. En particulier, ces données permettent de localiser les différents horizons et failles du réservoir. On peut ainsi en déduire la structure du réservoir.

Les données issues des puits sont les seules mesures directes disponibles pour caractériser le réservoir. Elles fournissent les valeurs de diverses propriétés pétrophysiques comme la porosité, la perméabilité, les saturations, la densité de la roche, la résistivité, *etc.* Les données de puits sont disponibles uniquement à l’emplacement des puits. A l’inverse des données sismiques, elles sont donc peu nombreuses et très localisées. Toutefois, elles ont une très bonne précision/résolution. Elles permettent de caractériser le réservoir pour de toutes petites échelles (de l’ordre du centimètre). Les données statiques aux puits peuvent être obtenues à partir de mesures sur des échantillons de roches extraits lors du forage, à savoir sur les carottes ou les boues de forage. Elles peuvent également provenir de diagraphies, c’est-à-dire obtenues par des outils de mesures parcourant les puits. Elles fournissent les variations de différentes propriétés pétrophysiques le long des puits, et permettent de distinguer les différents faciès, les différentes séquences stratigraphiques, *etc.*

Construction du modèle géologique

Il est possible de concevoir un modèle géologique représentatif du réservoir et respectant l’ensemble de ces données statiques. La structure du modèle est donnée par l’interprétation des données sismiques calées sur les mesures aux puits. Cette interprétation doit également tenir compte de l’information géologique régionale. La structure du réservoir (différentes couches, failles) est ensuite discrétisée selon un maillage tridimensionnel complexe composé de blocs élémentaires : on parle de grille réservoir. Afin

d'être la plus représentative possible des hétérogénéités observées sur le réservoir réel, cette grille est en général finement maillée. Elle peut ainsi être constituée de millions de blocs.

Pour compléter le modèle géologique, il faut ensuite spécifier les valeurs des propriétés pétrophysiques pour chacun des blocs de la grille. L'analyse des diagraphies et carottes permet de les renseigner aux puits. On peut ensuite se servir de techniques géostatistiques pour attribuer des valeurs de propriétés aux mailles non échantillonnées en s'appuyant sur l'analyse des données. Les méthodes géostatistiques permettent de générer des réalisations de propriétés pétrophysiques qui respectent les données existantes, et qui sont cohérentes entre elles (*e.g.* avec des corrélations entre la porosité et la perméabilité). On obtient alors un modèle géologique possible puisqu'il respecte l'ensemble des données statiques.

De nombreuses incertitudes

Grâce aux données sismiques, la structure d'un réservoir est relativement bien connue. En revanche, il est en pratique impossible de connaître les propriétés pétrophysiques en chaque point du réservoir ou pour chacune des mailles du modèle. En effet, le nombre de puits est généralement très réduit en raison de leur coût, alors que les réservoirs sont souvent très étendus (de l'ordre du kilomètre). Le nombre de données disponibles est donc très faible en comparaison du grand nombre de paramètres qu'il faut renseigner pour définir le modèle.

De plus, parmi les données disponibles, certaines sont mal réparties dans l'espace. Par exemple, les données de puits sont uniquement disponibles aux emplacements des puits. En conséquence, les réalisations géostatistiques basées sur ces données proposent des représentations qui, tout en restant plausibles, sont entachées d'incertitudes lorsque l'on considère des mailles se situant loin des puits.

Une autre source d'incertitude est liée à l'échelle des différentes mesures. En effet, les données de puits sont généralement mesurées à des échelles bien plus petites que l'échelle des blocs de la grille du modèle. Des mises à l'échelle sont alors indispensables pour calculer des propriétés équivalentes. De telles mises à l'échelle sont cependant difficiles à réaliser car elles dépendent du réservoir et de ses hétérogénéités. De plus, il existe peu de manières de vérifier leur pertinence, surtout lorsque l'on s'éloigne des puits.

Finalement, il existe de nombreuses autres sources d'incertitudes liées aux erreurs et approximations sur la mesure (*e.g.* incertitudes liées aux outils de mesure, à leur calibrage), le traitement des données (*e.g.* ré-échantillonnage des données sismiques), leur interprétation (*e.g.* conversion temps/profondeur) ou tout simplement la discrétisation du maillage [Vincent et al., 1999] [Corre et al., 2000] [Schulze-Riegert and Ghedan, 2007].

En résumé, les données disponibles sont très insuffisantes pour définir entièrement un modèle géologique. En conséquence, il existe une infinité de modèles géologiques pouvant respecter les données statiques. C'est pour cela que de nombreux paramètres du modèle peuvent être considérés comme incertains.

1.1.2 Simulation des écoulements

La construction de modèles géologiques permet d'estimer les volumes des ressources en place et donc de donner une première idée de la valeur économique d'un champ. Toutefois, apprécier les volumes en place ne suffit pas, il est fondamental d'en estimer la part techniquement récupérable. Le modèle géologique est donc fourni en entrée à un simulateur d'écoulement pour modéliser les déplacements de fluide au cours du temps. On simule alors la production suivant un schéma de production défini par

l'ingénieur réservoir.

Modèle géologique et modèle réservoir

Le temps de simulation d'un modèle dépend directement du nombre de mailles actives, c'est-à-dire du nombre de mailles dans lesquelles on simule les écoulements. Les mailles inactives correspondent à des mailles peu ou pas poreuses, elles servent généralement à délimiter les contours du réservoir. La grande finesse de la grille du modèle géologique devient alors un inconvénient majeur pour la simulation d'écoulement. En effet, les simulations pour un modèle comportant des millions de mailles vont impliquer un temps de calcul très important. Typiquement, la durée d'une seule simulation peut s'étendre de quelques heures jusqu'à plusieurs jours. Or, de nombreuses simulations sont généralement nécessaires pour proposer un plan de développement optimisant l'exploitation du champ. Dans de telles conditions, le coût total en temps de calcul devient rapidement prohibitif. C'est pourquoi le modèle géologique n'est généralement pas utilisé directement pour la simulation d'écoulement.

Une pratique courante consiste à construire un modèle plus grossier du réservoir, comportant un nombre réduit de mailles et sur lequel les simulations d'écoulement seront plus rapides. On parle alors de modèle réservoir, modèle dynamique ou encore modèle de simulation. Ce modèle est peuplé en propriétés pétrophysiques à partir du modèle géologique via une étape de mise à l'échelle. Un schéma représentant la mise à l'échelle d'une propriété pétrophysique de la grille d'un modèle fin vers la grille d'un modèle grossier est donné par la Figure 1.2. Le but est d'obtenir pour le modèle réservoir un comportement dynamique aussi proche que possible de celui du modèle géologique.

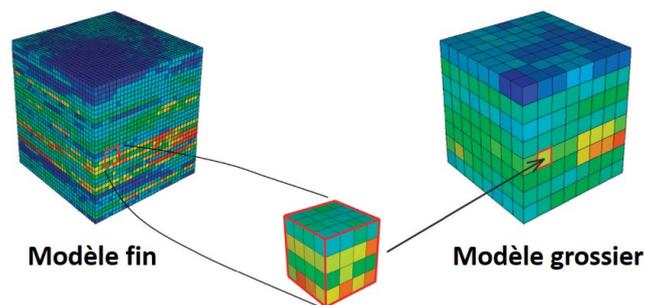


Figure 1.2 : Exemple de mise à l'échelle d'une propriété pétrophysique. A partir des valeurs de la propriété sur le modèle fin, on calcule la valeur équivalente de cette propriété pour le modèle grossier. Source: <http://petrowiki.org/>, modifiée.

Bien que couramment utilisée, la distinction entre modèle géologique et modèle réservoir n'est pas systématiquement nécessaire. Par exemple, si le modèle géologique est de taille raisonnable ou si l'on dispose d'une grande puissance de calcul, il est possible de considérer une unique taille de modèle. On utilise alors le modèle géologique et sa grille fine pour réaliser les simulations d'écoulement.

Simulation d'écoulement dans le modèle réservoir

Le modèle réservoir est ensuite fourni en entrée d'un simulateur d'écoulement afin de simuler l'évolution de la pression et des saturations dans le réservoir sous l'effet de la production. En plus du modèle réservoir, le simulateur a besoin qu'on lui spécifie les propriétés des fluides ainsi que le schéma de production. Puis, en résolvant la loi de conservation des masses, la loi de Darcy et les équations d'état à chaque pas de temps et pour chaque maille active, le simulateur d'écoulement calcule l'évolution au cours du temps des différentes propriétés des fluides dans tout le réservoir.

La simulation d'écoulement permet notamment de simuler la production aux puits en réponse au schéma de production, c'est-à-dire l'évolution des pressions et débits des différents fluides. La caractérisation et modélisation statique du réservoir, puis la simulation d'écoulement correspondent à la résolution du problème de modélisation direct résumé par la figure 1.3. Ce workflow permet déjà à l'ingénieur réservoir d'évaluer l'impact du schéma de production, de la position des puits, et ainsi de proposer un plan de développement pour les premières phases de l'exploitation du champ.

Cependant, les incertitudes existantes sur le modèle géologique se propagent sur le modèle de réservoir et finalement sur la précision des prédictions. Afin de les réduire, il est nécessaire d'incorporer les nouvelles données qui sont acquises au fur et à mesure de l'exploitation. Ces dernières peuvent être issues du forage de puits additionnels, d'acquisitions sismiques répétées au cours du temps, et surtout de mesures faites aux puits au cours de la production, les données de production. Les données de production sont des fonctions du temps. Il s'agit par exemple des débits de fluide comme le débit d'huile, ou bien de la pression au fond du puits. En pratique, on peut les voir comme des vecteurs dont chaque élément correspond à la valeur de la propriété à un temps donné. On parle ainsi de sorties vectorielles, ou propriétés d'intérêt, que l'on simule avec le simulateur d'écoulement. Une bonne concordance entre les données mesurées et les réponses simulées correspondantes conforte la pertinence des modèles géologique et réservoir. A l'inverse, de fortes différences indiquent une modélisation de mauvaise qualité, il faut alors revoir le modèle.

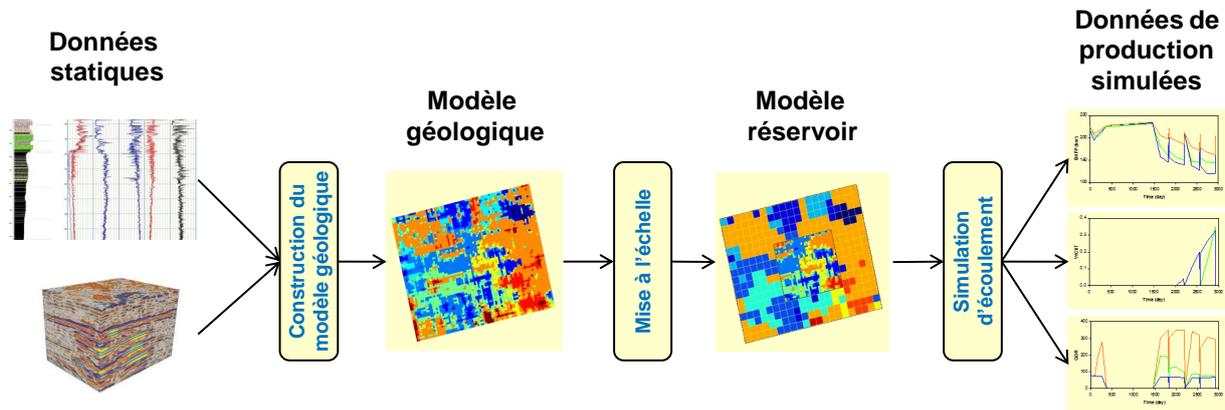


Figure 1.3 : Schéma de la modélisation d'un réservoir. A partir des données statiques, on construit des modèles qui permettent la simulation des données de production lors de l'exploitation du champ.

1.2 Calage d'historique

La modélisation directe comprend la phase de construction du modèle géologique, de mise à l'échelle si nécessaire, et de simulation des écoulements. Les données statiques sont intégrées au cours de la construction du modèle géologique. Toutefois, la modélisation directe ne garantit pas le respect des données de production. L'intégration des données dynamiques dans la modélisation correspond à la résolution d'un problème inverse : le calage d'historique. Le processus du calage d'historique est décrit dans la section 1.2.1. La définition de la fonction objectif et le choix de la paramétrisation sont deux étapes clefs de ce processus. Elles sont détaillées dans les sections 1.2.2 et 1.2.3, respectivement.

1.2.1 Principe

Les données dynamiques correspondent aux réponses dynamiques du réservoir lorsque celui-ci est en exploitation : les données de production ou historique de production. Contrairement à l'intégration des données statiques qui se fait directement lors de la construction des modèles géologiques, l'intégration des données dynamiques ne peut se faire qu'après la mise en place du workflow de simulation direct. Ce processus, appelé calage d'historique, consiste à ajuster différentes caractéristiques des modèles géologique et/ou réservoir jusqu'à ce que ces derniers reproduisent les données de production. Le calage d'historique est donc la résolution d'un problème inverse dans lequel on cherche à déduire une partie des caractéristiques du réservoir à partir des données dynamiques disponibles.

Traditionnellement, le calage d'historique est résolu à travers une approche variationnelle [Jacquard et al., 1965] résumée par le schéma de la Figure 1.4. Les principales étapes sont les suivantes :

1. On identifie des paramètres incertains intervenant dans la définition des modèles réservoir ou géologique. C'est la paramétrisation.
2. Si nécessaire, on fait la mise à l'échelle du modèle géologique, puis on effectue la simulation d'écoulement sur le modèle réservoir.
3. On quantifie l'écart entre les réponses dynamiques simulées et les données mesurées en calculant une fonction objectif.
4. On utilise un algorithme de minimisation pour ajuster les paramètres incertains afin de réduire la fonction objectif.
5. On répète les étapes 2 à 4 jusqu'à ce que le calage obtenu soit satisfaisant.

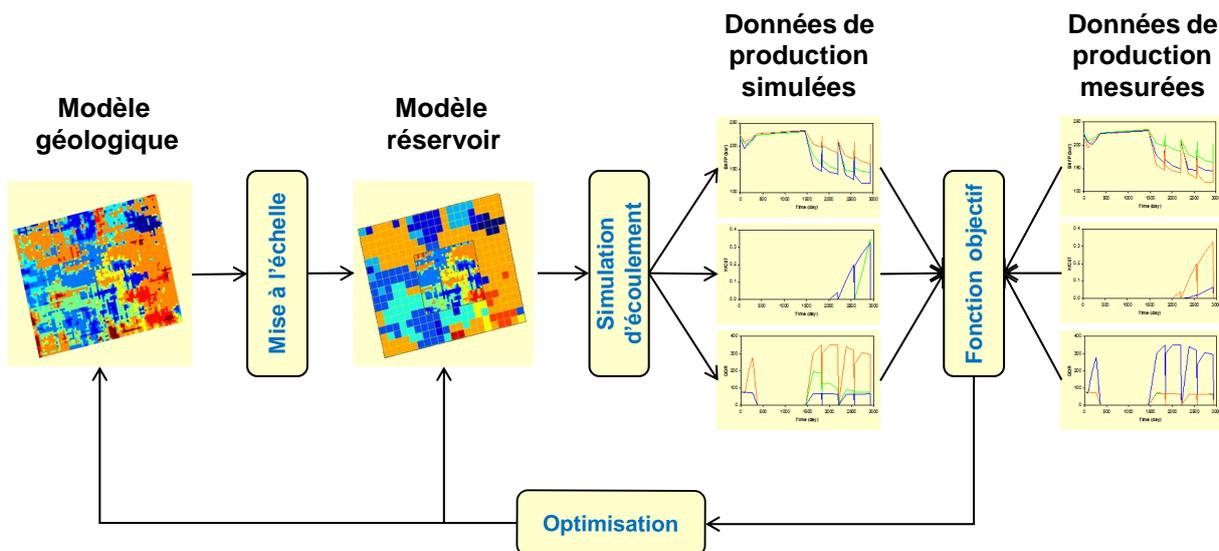


Figure 1.4 : Schéma d'un calage d'historique. Les modèles géologique et/ou réservoir sont modifiés jusqu'à ce que les données de production simulées reproduisent le mieux possible les données de production mesurées.

Un autre type d'approches basées sur des méthodes d'assimilation séquentielle avec notamment le filtre de Kalman d'ensemble (EnKF) [Lorentzen et al., 2003] [Evensen, 2009] a suscité un fort intérêt ces dernières années pour la résolution du calage d'historique. Ces méthodes considèrent un ensemble de réalisations pour le modèle réservoir, ce qui donne une distribution a priori. On résout alors le problème direct pour chaque réalisation jusqu'au temps où on a la première mesure. Tous les paramètres sont alors mis à jour à partir d'un krigeage. Le processus est ensuite répété jusqu'à ce que toutes les données soient assimilées. Cette approche a l'avantage de pouvoir incorporer de nouvelles données en continu et de fournir un ensemble de modèles calés.

De manière générale, le problème du calage d'historique a fait l'objet de nombreuses recherches ces dernières années, notamment pour l'automatiser, traiter un nombre de paramètres incertains toujours plus important et améliorer ses performances. [Oliver and Chen, 2011] passent en revue les dernières méthodes expérimentées pour le calage d'historique au niveau de la paramétrisation, du calcul de la fonction objectif, des algorithmes de calage, et de la quantification des incertitudes. Dans la suite de cette thèse, on se place dans le cadre d'un calage d'historique par méthodes variationnelles suivant le schéma de la Figure 1.4.

1.2.2 Fonction objectif

Nous notons \mathbf{x} un ensemble de paramètres incertains d'un modèle réservoir. On note aussi $y(\mathbf{x})$ les données de production correspondant à la simulation du réservoir défini par \mathbf{x} . Les données de production simulées peuvent alors être comparées aux données de production que l'on mesure lors de l'exploitation du réservoir réel, notées \mathbf{d}^{obs} . Cette comparaison se fait par le calcul d'une fonctionnelle appelée fonction objectif qui quantifie l'écart entre les données de production simulées et mesurées. La fonction objectif, ou FO , est généralement une différence par moindres carrés pondérés :

$$FO(\mathbf{x}) = \frac{1}{2}(\mathbf{y}(\mathbf{x}) - \mathbf{d}^{obs})^\top \mathbf{C}_D^{-1}(\mathbf{y}(\mathbf{x}) - \mathbf{d}^{obs}) \quad (1.1)$$

avec \mathbf{C}_D la matrice de covariance qui quantifie les différentes incertitudes sur les données de production.

En pratique, une matrice diagonale est souvent choisie pour la matrice de covariance \mathbf{C}_D , on suppose alors que les incertitudes sur les données de production sont indépendantes pour chaque donnée. Dans ce cas, une formulation équivalente souvent considérée pour le calcul de la fonction objectif s'exprime de la manière suivante:

$$FO(\mathbf{x}) = \frac{1}{2} \sum_{k=1}^K \omega_k \sum_{t=1}^T \left(\frac{y_k(\mathbf{x}, t) - d_{kt}^{obs}}{\sigma_{kt}} \right)^2 \quad (1.2)$$

avec k et t des indices faisant référence aux différentes sorties (*e.g.* un débit d'huile à un puits donné) et aux différents temps pour lesquels ces sorties sont mesurées. On considère ici K sorties différentes, toutes mesurées à T différents temps. Les coefficients ω et σ sont des poids. Les premiers sont assignés par sortie afin de rendre la somme de leurs erreurs comparable. En effet, la fonction objectif regroupe souvent des données de production avec des ordres de magnitude très différents, comme des mesures de pression, pouvant s'exprimer en centaines de bar, et des fractions d'eau (*i.e.* le ratio entre le volume d'eau sur le volume total produit) strictement comprises entre 0 et 1. Si aucune pondération n'est mise en place, la fonction objectif ne sera représentative que de la sortie présentant la plus forte magnitude. Finalement, σ désigne des poids pouvant être vus comme l'écart-type des erreurs de mesure sur les données de production.

La fonction objectif est nulle lorsque les données de production simulées sont identiques à celles observées. Le calage d'historique est donc équivalent à un problème de minimisation de la fonction objectif. En se donnant un jeu de valeurs initiales pour les paramètres incertains \mathbf{x} et en simulant les réponses dynamiques correspondantes, on peut calculer une valeur initiale de fonction objectif. Un algorithme d'optimisation, typiquement un algorithme de descente basé sur le calcul de gradients, peut alors être utilisé pour déterminer les valeurs de paramètres incertains qui permettent de minimiser la fonction objectif. A la fin de ce processus d'optimisation, on trouve généralement un minimum local. On peut alors réitérer ce processus jusqu'à trouver le minimum global. Toutefois ce dernier

est difficile à atteindre car les variations de la fonction objectif sont complexes. En pratique, lorsque la valeur de la fonction objectif est inférieure à un seuil prédéfini, on considère que le modèle est calé et qu'il respecte suffisamment bien les différentes données dynamiques. De nombreux algorithmes d'optimisation peuvent être envisagés pour la minimisation de la fonction objectif. L'algorithme de descente Quasi-Newton est souvent plébiscité pour son efficacité, mais de nombreux autres algorithmes basés sur des méthodes de descente [Makhlouf et al., 1993] [Li et al., 2001], de région de confiance [Subbey et al., 2004] [Christie et al., 2006], ou d'optimisation sans calcul de dérivées [Romero and Carter, 2001] [Schulze-Riegert et al., 2003], ont été appliqués avec succès au calage d'historique. Un point commun à toutes ces méthodes est qu'elles nécessitent un nombre conséquent de simulations d'écoulement.

Il est important de se rappeler que le processus de calage d'historique via la minimisation d'une fonction objectif correspond à la résolution d'un problème mal posé, où le nombre de paramètres du modèle est toujours bien plus important que le nombre de données indépendantes dont on dispose pour les contraindre [Tarantola, 2005]. Cela implique notamment qu'il peut ne pas y avoir de solution exacte ou bien qu'il existe potentiellement une infinité de modèles respectant les données. L'obtention d'un unique modèle réservoir calé n'est donc pas toujours satisfaisante. Il est alors possible d'adopter une approche probabiliste pour résoudre le calage d'historique. Le but d'une telle approche est de déterminer le modèle calé le plus probable et la distribution associée [Oliver et al., 2008] [Feraille and Marrel, 2012].

1.2.3 Paramétrisation

Comme évoqué dans la partie 1.1, de nombreux paramètres définissent les modèles du réservoir, mais très peu d'entre eux sont directement contraints par les données statiques. Avant de passer à la phase de calage, l'ingénieur doit donc spécifier quelles sont les propriétés du réservoir qu'il souhaite ajuster, et qui constituent les paramètres incertains du problème. L'idée est de choisir des paramètres influençant le comportement dynamique et dont la valeur est incertaine. Le choix des paramètres incertains, leur nombre et la définition de leur intervalle de variation va dépendre du cas d'étude, mais aussi des méthodes de calage d'historique utilisées.

En pratique, un nombre assez restreint de paramètres est souvent considéré afin de limiter la complexité des études et les temps de calculs associés. L'idée est alors de choisir ou définir des paramètres globaux (*e.g.* le volume poreux, la force de l'aquifère) ou des paramètres représentatifs d'un ensemble de mailles (des *patches* ou des couches sur lesquels on applique des coefficients multiplicateurs de porosité, perméabilité, *etc*) afin d'assurer un impact significatif sur le comportement dynamique du réservoir. Des méthodes d'analyse de sensibilité peuvent être envisagées à cette étape afin de sélectionner les paramètres les plus influents (*i.e.* les paramètres dont la perturbation impacte le plus la réponse d'intérêt) tout en écartant les paramètres ayant une influence négligeable.

Lorsque les paramètres incertains sont définis au niveau du modèle réservoir, on parle de calage d'historique traditionnel. Des choix courants de paramètres portent sur les propriétés d'écoulement identiques pour l'ensemble du réservoir (la densité des fluides, les saturations critiques, *etc*) ou sur des ensembles de mailles (multiplicateur de perméabilité/porosité par couche ou autour des puits, *etc*). Cette paramétrisation, dite de zonation, est très répandue car très simple à mettre en oeuvre.

Une paramétrisation au niveau du modèle géologique peut également être envisagée. On parle alors de calage d'historique avancé. Les paramètres choisis peuvent être ceux utilisés par les méthodes géostatistiques générant les réalisations de propriétés pétrophysiques (porosité, perméabilité, *etc*). Des paramètres tels que la moyenne, la portée, la variance d'une propriété pétrophysique ou les nombres aléatoires utilisés pour initier le processus de génération peuvent aussi être considérés. L'intérêt d'une paramétrisation au niveau du modèle géologique est qu'elle rend possible une modification des modèles

tout en continuant à honorer les données statiques, alors que ce n'est généralement pas le cas lorsque la paramétrisation est définie au niveau du modèle réservoir. Cependant, une paramétrisation du modèle géologique implique aussi une mise à l'échelle systématique afin de calculer le modèle réservoir correspondant au modèle géologique modifié, ce qui ajoute un coût supplémentaire en temps de calcul.

1.3 Réduction des temps de calcul

La conception de modèles représentatifs du réservoir passe par l'intégration de toutes les données disponibles. Cela nécessite entre autres le calcul de nombreuses simulations d'écoulement lors du calage d'historique. Les modèles calés servent ensuite de point de départ pour la recherche de plans de développement optimisés, ce qui est également coûteux en temps de simulation. Les temps de calculs associés aux simulations d'écoulement apparaissent donc comme une limitation majeure en ingénierie de réservoir. Une solution consiste à remplacer le simulateur d'écoulement par un méta-modèle (section 1.3.1). Un méta-modèle est une fonction rapide à évaluer qui permet d'approcher une sortie d'intérêt, par exemple une sortie du simulateur d'écoulement. Lorsque la sortie d'intérêt peut être simulée pour différents niveaux de fidélité, il est souvent possible d'accélérer la construction des méta-modèles en considérant la méta-modélisation multi-fidélité comme expliqué dans la section 1.3.2.

1.3.1 La méta-modélisation

Dans de nombreux domaines, on observe que les codes de calcul pour la modélisation et la simulation intègrent une physique de plus en plus détaillée. En contrepartie, ces codes impliquent de nombreux paramètres à calibrer et sont de plus en plus gourmands en temps de calcul, ce qui rend leur analyse très complexe. Ce type de problématique peut être traité de manière générale comme un problème de type « boîte noire ». L'idée est que le système n'est vu qu'à travers ses entrées et ses sorties (Figure 1.5). Le passage de l'un à l'autre (via le code de calcul) est alors considéré comme un processus inconnu, la boîte noire.



Figure 1.5 : Schéma d'un système « boîte noire ».

Ce type de représentation est souvent adopté pour les codes de calculs complexes, dont les évaluations sont coûteuses en temps de calcul. Appliqué à l'ingénierie de réservoir, l'entrée correspond aux paramètres incertains, la boîte noire fait référence à la simulation d'écoulement et la sortie désigne les réponses en production simulées. De ce point de vue, les problèmes rencontrés en ingénierie de réservoir (analyse de sensibilité, prédiction des données de production, calage d'historique/calibrage du code de calcul, *etc*) sont similaires aux problèmes généralement rencontrés lors de l'analyse de codes de calculs longs à évaluer. Une solution spécifiquement adaptée pour traiter ce type de problématique consiste à construire un méta-modèle.

Principe

Le principe d'un méta-modèle est de remplacer la boîte noire par une approximation mathématique afin d'approcher une sortie donnée, la réponse d'intérêt [Sacks et al., 1989], [Osio and Amon, 1996],

[Kleijnen and Sargent, 2000], [Fang et al., 2006]. On parle de méta-modèle car il est construit en se basant uniquement sur la relation entre entrées et sorties, et ne cherche pas à modéliser les processus intervenant au sein de la boîte noire. Selon les domaines dans lesquels ils sont utilisés, les méta-modèles sont aussi connus sous le nom de modèles *proxy*, surfaces de réponse, modèles de substitution ou encore émulateurs.

Un méta-modèle est construit à partir d'un nombre limité d'évaluations de la boîte noire. L'ensemble des points considérés en entrée pour la construction du méta-modèle est appelé plan d'expériences. Les valeurs en sortie associées à ce plan d'expériences sont appelées observations. L'intérêt du méta-modèle est de pouvoir prédire à moindre coût la valeur de la sortie d'intérêt pour des valeurs en entrée autres que celles du plan d'expériences. Si ces prédictions sont de bonne qualité, le méta-modèle peut ensuite être utilisé pour remplacer les appels à la boîte noire, et ainsi réaliser à coût réduit différentes études.

Il existe différentes méthodes pour construire un méta-modèle. Parmi les plus courantes, on trouve le krigeage, les polynômes, les splines et les réseaux de neurones. La méta-modélisation est un outil couramment utilisé pour l'analyse de problèmes de type boîte noire. Leur utilisation est très répandue dans de nombreux domaines dès lors que des codes de calcul lourds sont impliqués : météorologie, aéronautique, design de surfaces, ingénierie de réservoir...

Application à l'ingénierie de réservoir

L'ingénierie de réservoir est un domaine d'application possible de la méta-modélisation. Des méta-modèles peuvent être construits pour se substituer aux simulations d'écoulement coûteuses en temps de calcul. Ils peuvent ensuite être utilisés à différentes étapes du processus de calage, notamment lors de :

- l'analyse de sensibilité portant sur les paramètres incertains du modèle,
- la minimisation de la fonction objectif,
- la mise en place d'approches probabilistes.

Leur utilisation peut également intervenir dans un second temps lors de la planification du développement des champs : placement de nouveaux puits, optimisation du planning de production, *etc.*

La méta-modélisation dans le cadre de telles applications a été largement étudiée dans la littérature. En particulier, différentes méthodes de construction des méta-modèles ont été testées. Des polynômes sont considérés dans [Pan and Horne, 1998], [Peng and Gupta, 2004], des splines dans [Li and Friedmann, 2005], des réseaux de neurones dans [Queipo et al., 2000], [Guyaguler et al., 2000], [Cullick et al., 2006] et du krigeage dans [Pan and Horne, 1998], [Guyaguler et al., 2000], [Badru and Kabir, 2003], [Peng and Gupta, 2004], [Scheidt et al., 2007], [Slotte and Smorgrav, 2008], [Osterloh, 2008], [Feraille and Marrel, 2012], [Becerra et al., 2012], [Feraille, 2013], [Douarche et al., 2014]. Une comparaison de ces différentes méthodes appliquées à l'ingénierie de réservoir est également proposée dans [Yeten et al., 2005] et [Zubarev, 2009].

Parmi les différentes approches existantes, le krigeage présente plusieurs avantages pour l'ingénierie de réservoir. Tout d'abord, c'est un interpolateur. Le méta-modèle est donc exact aux points où la fonction est connue, ce qui est une propriété intéressante car les simulations d'écoulement sont déterministes. De plus, cette technique fournit une expression analytique de l'estimateur, ainsi qu'une expression de l'erreur sur l'estimateur. Ces expressions peuvent s'avérer utiles pour certaines applications importantes en ingénierie de réservoir telles que la propagation des incertitudes [Scheidt, 2006] ou l'analyse de sensibilité [Marrel, 2008]. Un défaut souvent relevé pour le krigeage par rapport aux autres méthodes est le temps nécessaire à la construction du modèle. L'amélioration des méthodes

d'estimation des hyper-paramètres a cependant permis de rendre le krigeage plus compétitif sur ce point [Kennedy and O'Hagan, 2000b] [Santner et al., 2003] [Rasmussen and Williams, 2006].

1.3.2 La méta-modélisation multi-fidélité

Dans de nombreux cas, il est facile d'obtenir, en un temps de calcul réduit, des évaluations dégradées d'une sortie d'intérêt. En effet, un code de calcul peut souvent être lancé pour différents niveaux de complexité, produisant ainsi plusieurs niveaux de réponses dont la précision varie. En ingénierie de réservoir, différents niveaux de fidélité peuvent facilement être obtenus en faisant varier la résolution des modèles réservoir, en considérant des simulations d'écoulement à physique dégradée, en dégradant les critères de convergence, *etc.*

Différentes techniques ont été proposées pour la création de méta-modèles à partir d'informations de fidélité variable. Elles sont décrites par [Huang et al., 2006] et [Forrester and Keane, 2009]. Parmi elles, les méta-modèles multi-fidélité basés sur le co-krigeage et proposés par [Kennedy and O'Hagan, 2000a] ont été largement étudiés, notamment dans l'aérospatiale [Forrester et al., 2007], [Kuya et al., 2011], [Han et al., 2012].

Le principe d'un méta-modèle multi-fidélité est le suivant. On suppose d'abord que l'on dispose de plusieurs niveaux de résolution sur lesquels on peut évaluer la réponse d'intérêt. Selon le niveau considéré, ces évaluations sont plus ou moins précises et plus ou moins coûteuses en temps de calcul. A partir d'un ensemble de réponses sur les différents niveaux de fidélité, on construit un méta-modèle multi-fidélité approchant le niveau de fidélité le plus précis. En favorisant les évaluations sur les niveaux grossiers, on peut espérer construire un méta-modèle prédictif plus rapidement que via une méta-modélisation basée uniquement sur le niveau le plus fin. Toutefois, cela suppose que les niveaux grossiers apportent suffisamment d'information sur le comportement dynamique au niveau fin.

Le co-krigeage multi-fidélité est une extension du krigeage. De ce point de vue, il présente les mêmes caractéristiques : interpolateur, expression analytique de l'estimateur, estimation de l'erreur sur l'estimateur. En outre, de nombreux résultats sur le krigeage peuvent être appliqués au co-krigeage multi-fidélité. C'est pour cela que celui-ci est très attractif pour des applications en ingénierie de réservoir. Les récents travaux de thèse de [Le Gratiet, 2013b] ont permis d'améliorer le modèle initialement suggéré par [Kennedy and O'Hagan, 2000a]. Ces améliorations permettent notamment une construction rapide et facile d'un modèle de co-krigeage multi-fidélité, en particulier grâce à l'introduction d'une formulation récursive du modèle et d'une méthode rapide d'estimation des hyper-paramètres.

1.4 Objectifs et organisation de la thèse

La méta-modélisation, et notamment le krigeage, apparaît comme un outil d'importance grandissante en ingénierie de réservoir. L'efficacité de la méta-modélisation dépend toutefois de la prédictivité obtenue avec le méta-modèle. En pratique, le nombre d'observations et donc de simulations nécessaires pour assurer une bonne qualité du méta-modèle est encore important. L'utilisation de méta-modèles multi-fidélité peut permettre de réduire ce temps, en réduisant le nombre de simulations sur le niveau le plus fin. Une telle solution semble d'autant plus judicieuse qu'il existe diverses manières d'obtenir des niveaux de fidélité dégradés en ingénierie de réservoir via la mise à l'échelle, la simplification de la modélisation, les approximations dans la simulation, *etc.* L'utilisation de méta-modèles multi-fidélité peut ainsi tirer profit des niveaux de fidélité dégradés disponibles afin d'obtenir plus rapidement un méta-modèle prédictif. Le co-krigeage multi-fidélité a déjà été appliqué avec succès dans [Le Ravalec-dupin, 2012] à la construction de cartes de qualité pour le placement de puits. Les différents niveaux de fidélité y sont obtenus par l'utilisation de grilles de résolutions différentes. Les résultats de ce travail

appelaient à être poursuivis afin de mieux appréhender et pouvoir pleinement exploiter le potentiel de la méta-modélisation multi-fidélité en ingénierie de réservoir.

1.4.1 Objectifs de la thèse

Dans cette thèse, nous proposons de tester de manière approfondie l'utilisation du co-krigeage multi-fidélité pour la construction de méta-modèles dédiés à des applications en ingénierie de réservoir, approchant les sorties d'intérêt de la simulation d'écoulement et en particulier la fonction objectif.

Les travaux entrepris au cours de cette thèse s'organisent autour de deux objectifs principaux :

1. Comparer co-krigeage multi-fidélité et krigeage pour la méta-modélisation à partir de plans d'expériences de différentes tailles.
2. Améliorer les performances de la méta-modélisation et exploiter le potentiel d'une approche multi-fidélité grâce à la planification séquentielle d'expériences.

Dans un premier temps, l'objectif est donc de déterminer quel est l'apport du co-krigeage multi-fidélité en termes de gain en temps de calcul par rapport au krigeage. Pour cela, une comparaison entre les deux approches a été réalisée sur le cas d'étude PUNQ. Cette comparaison a permis d'identifier dans quelles conditions l'approche multi-fidélité pouvait s'avérer plus pertinente et quels gains en temps pouvaient être espérés. Pour améliorer les performances de la méta-modélisation de sorties vectorielles, une méthode basée sur une réduction de la dimension via l'analyse en composantes principales a été testée en simple fidélité. Puis, une extension à un cadre multi-fidélité a été proposée. Les résultats obtenus ont aussi conduit à la mise en place d'une méthode adaptée à la méta-modélisation de la fonction objectif. Cette méthode vise à tirer profit de la multi-fidélité, mais permet également d'obtenir de meilleurs résultats en simple fidélité.

Dans un second temps, des méthodes de planification séquentielle d'expériences ont été proposées afin d'accélérer la construction de méta-modèles prédictifs dans des cadres simple et multi-fidélité. Ces méthodes de planification séquentielle d'expériences ont été appliquées à l'approximation de la fonction objectif, principalement sur le cas d'étude BRUGGE. Elles sont également adaptées à la méta-modélisation particulière suggérée pour la fonction objectif.

1.4.2 Organisation du manuscrit

Chapitre 2 : Méthodologies pour la méta-modélisation

Dans ce chapitre, nous présentons les étapes clés de la méta-modélisation : la planification d'expériences, la construction du méta-modèle et l'estimation de sa qualité en termes de prédictivité. Nous détaillons en particulier les approches de méta-modélisation par krigeage et co-krigeage multi-fidélité, ainsi que les méthodes utilisées pour la définition des plans d'expériences fixes dans les deux cas.

Chapitre 3 : Modélisation de sorties vectorielles et de la fonction objectif

Dans ce chapitre, nous proposons de construire des méta-modèles de sorties vectorielles. Pour cela, on considère une méthode permettant de réduire leur dimension : l'analyse en composantes principales. L'utilisation combinée de cette dernière avec le krigeage ou le co-krigeage multi-fidélité permet la construction rapide de méta-modèles pour des sorties vectorielles dans un cadre simple fidélité ou multi-fidélité, respectivement. Ces méthodes peuvent notamment être utilisées pour approcher les

différentes sorties dynamiques incluses dans le calcul de la fonction objectif. De cette manière, on obtient une approche vectorielle pour la méta-modélisation de la fonction objectif dont nous présentons les caractéristiques.

Chapitre 4 : Cas d'étude PUNQ

Dans ce chapitre, nous présentons le cas d'application PUNQ, inspiré du cas de référence PUNQ-S3. Le cas PUNQ multi-fidélité est le support d'une étude approfondie comparant krigeage et co-krigeage multi-fidélité pour la méta-modélisation de différentes sorties : scalaires, temporelles, spatiales, et fonction objectif. Les résultats de ces comparaisons mettent en évidence l'intérêt de la méthode proposée pour la méta-modélisation de la fonction objectif, en particulier dans un cadre multi-fidélité. Un exemple d'application est également proposé par la réalisation d'un calage d'historique sur ce cas.

Chapitre 5 : Planification séquentielle d'expériences

Dans ce chapitre, nous proposons des algorithmes de planification séquentielle d'expériences afin de construire rapidement des méta-modèles prédictifs sur l'ensemble de l'espace des paramètres. Des versions adaptées à un cadre simple et multi-fidélité sont détaillées pas à pas. Les algorithmes proposés sont notamment applicables aux méthodes précédemment introduites pour la méta-modélisation de sorties vectorielles et surtout de la fonction objectif. Des versions adaptées au cas où les simulations peuvent être calculées en parallèle sont également proposées.

Chapitre 6 : Cas d'étude BRUGGE

Dans ce chapitre, nous présentons le cas d'étude BRUGGE, inspiré du cas de référence Brugge, ainsi que la version multi-fidélité mise en place. Ce cas est utilisé ici pour tester les différents algorithmes séquentiels proposés pour la planification d'expériences dans des cadres simple fidélité et multi-fidélité. Un exemple d'application est également proposé par la réalisation d'une analyse de sensibilité sur ce cas. Les résultats obtenus soulignent la pertinence des algorithmes proposés ainsi que l'intérêt de la méta-modélisation multi-fidélité en ingénierie de réservoir.

Les chapitres 3 et 4 ont fait l'objet d'une publication : Thenon, A., Gervais, V., Le Ravalec, M., 2016, Multi-fidelity meta-modeling for reservoir engineering - application to history matching, Computational Geosciences, 20(6), 1231-1250, doi:10.1007/s10596-016-9587-y

Un article basé sur les chapitres 5 et 6 est en cours d'écriture:
Thenon, A., Gervais, V., Le Ravalec, M., Sequential design strategies for multi-fidelity meta-models in reservoir engineering

CHAPITRE 2

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.

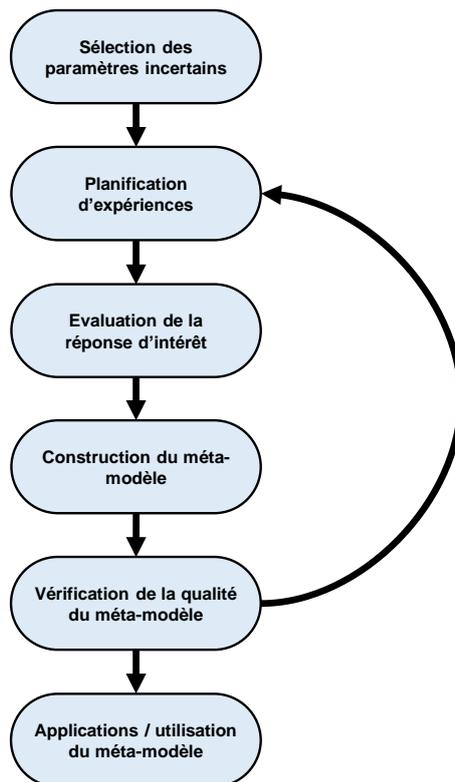


Figure 2.1 : Schéma traditionnel d'un workflow basé sur la méta-modélisation.

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.

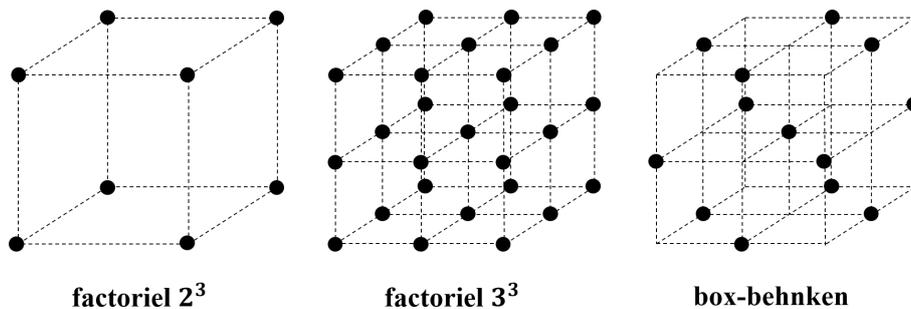


Figure 2.2 : Exemples de plan d'expériences standards représentés en dimension 3. A gauche et au centre, les plans factoriels considèrent un point d'expérience pour chacun des noeuds du découpage de l'espace. A droite, le plan de box-behnken définit des points d'expériences sur une partie des noeuds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

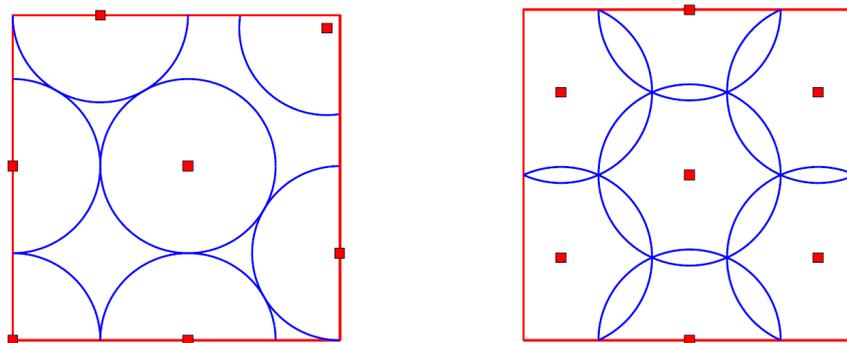


Figure 2.3 : Exemples de plan d'expériences de 7 points en dimension 2 construits selon les critères maximin (gauche) et minmax (droite). Source: [Pronzato and Müller, 2012] d'après <http://www.packomania.com/> et [Johnson et al., 1990].

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.

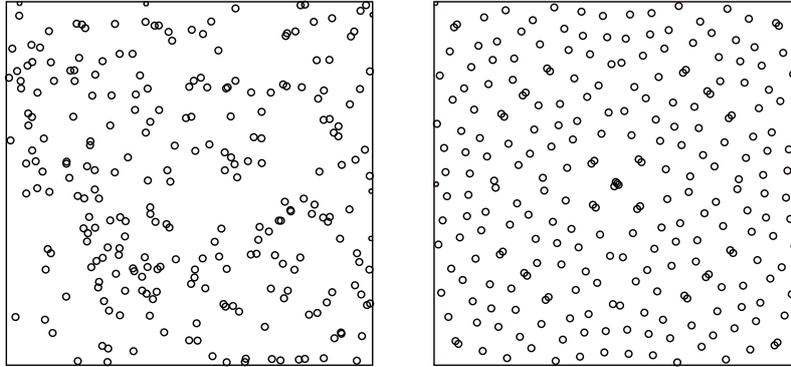


Figure 2.4 : A gauche, un plan d'expériences obtenu par tirage aléatoire. Le plan comporte de nombreux espaces vides et amas de points. A droite, un plan d'expériences issu d'une suite de Sobol'. La répartition des points est très régulière. Source : https://en.wikipedia.org/wiki/Sobol'_sequence, modifiée.

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.

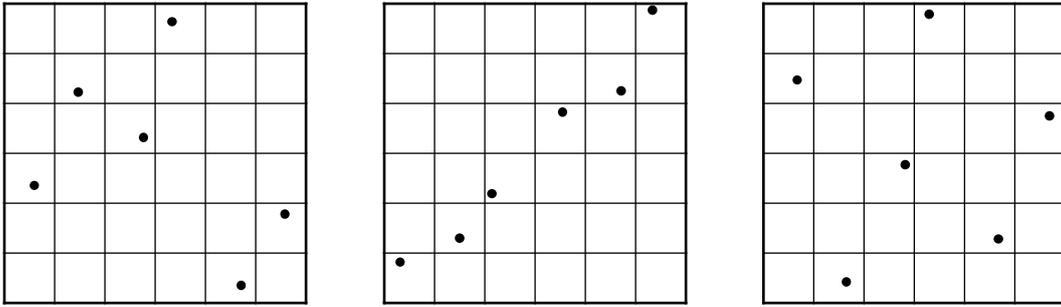


Figure 2.5 : Exemples de LHS à 6 points en dimension 2. Un point d'expérience au sein d'un LHS est tiré aléatoirement de manière à n'apparaître qu'une seule fois par strate (ou ligne/colonne en dimension 2). Un LHS ne remplit pas nécessairement l'espace de manière uniforme.

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