Méthodes numériques et modèle réduit de chimie tabulée pour la propagation d'incertitudes de cinétique chimique

Nicolas Dumont

université PARIS-SACLAY

CentraleSupélec

NNT : 2019SACLC037

Thèse de doctorat

Méthodes numériques et modèle réduit de chimie tabulée pour la propagation d'incertitudes de cinétique chimique

Thèse de doctorat de l'Université Paris-Saclay
préparée à CentraleSupélec

École doctorale n°579 Sciences mécaniques et énergétiques, matériau et géosciences (SMEMAG)
Spécialité de doctorat : Combustion

Thèse présentée et soutenue à Gif-syr-Yvette, le 08/07/2019, par

NICOLAS DUMONT

Composition du Jury :

Vincent Giovangigli Directeur de Recherche CNRS, Ecole Polytechnique (CMAP - UMR 7641)	Président
Pascale Domingo Directeur de Recherche CNRS (CORIA - UMR 6614)	Rapporteur
Alessandro Parente Professeur, Université Libre de Bruxelles	Rapporteur
Olivier Le Maitre Directeur de Recherche CNRS (LIMSI - UPR 3251)	Examinateur
Mélanie Rochoux Senior Research Scientist, CERFACS	Examinateur
Guillaume Vanhove Maître de Conférences, Université de Lille 1 (PC2A - UMR 8522)	Examinateur
Olivier Gicquel Professeur, CentraleSupélec (EM2C - UPR 288)	Directeur de thèse
Ronan Vicquelin Maître de Conférences, CentraleSupélec (EM2C - UPR 288)	Co-encadrant

Remerciements

Je souhaite tout d'abord remercier l'ensemble des membres de mon jury pour leur présence à ma soutenance, leur retour sur mon travail et les riches échanges durant la soutenance. Merci en particulier à Pascale Domingo et à Alessandro Parente pour avoir accepter la lourde tâche d'être les rapporteurs de mon manuscrit. Merci également à Vincent Giovangigli qui a accepté d'être le président de ce jury, ainsi qu'à Olivier Le Maitre, Mélanie Rochoux et Guillaume Vanhove pour leurs rôles d'examinateurs.

Je tiens également à remercier chaleureusement mes encadrants Olivier Gicquel et Ronan Vicquelin pour leur apport à ce travail et leur soutien indéfectible tout au long de la thèse. Merci à tous les deux pour m'avoir donné l'opportunité de réaliser à bien ce travail de thèse en m'aiguillant et en étant toujours disponibles. J'ai beaucoup appris grâce à vous et je vous en remercie.

Je voulais remercier l'ensemble du personnel de l'école doctorale pour m'avoir aidé et soutenu notamment durant la fin de ma thèse, en particulier Benoît Goyeau pour le temps qu'il a su m'accorder.

Je tiens également à remercier l'ensemble du personnel administratif et technique du laboratoire pour leur aide qu'ils offrent tous les jours à l'ensemble du personnel. Merci à tous ceux qui m'ont fait l'honneur de partager leur bureau, ainsi qu'à l'ensemble des doctorants, stagiaires, membres du laboratoire et membres des autres laboratoires avec qui j'ai pu partager durant ces années de thèse, que ce soit autour d'un café, d'une pause improvisée ou d'un match de basket ou de foot.

Enfin, merci à l'ensemble de mes amis et à ma famille pour m'avoir soutenu et supporté durant ces années de thèse comme ils l'ont toujours fait !

Abstract

Numerical simulation plays a key role in the field of combustion today, either in the research area by permitting a better understanding of phenomenons taking place inside reactive flows or in the development of industrial application by reducing designing cost of systems. Large Eddy Simulation is at the time the most suited tool for the simulation of reactive flows. Large Eddy Simulation of reactive flows is in practice only possible thanks to a modeling of different phenomenons :

— turbulence is modeled for small structures allowing to resolve only big structures which results in lower computational cost
— chemistry is modeled using reduction methods which allows to drastically reduce computational cost

The maturity of Large Eddy Simulation of reactive flows makes it today a reliable, predictive and promising tool. It now makes sense to focus on the impact of the parameters involved in the different models on the simulation results. This study of the impact of the modeling parameters can be seen from the perspective of uncertainties propagation, and can give interesting informations both from a practical side for the robust design of systems but also on the theoretical side in order to improve the models used and guide the experimental measurements to be made for the reliability improvement of these models.

The context of this thesis is the development of efficient methods allowing the propagation of uncertainties present in the chemical kinetic parameters of the reaction mechanisms within Large Eddy Simulation, these methods having to be non-intrusive in order to take advantage of the existence of the different computation codes which are tools requiring heavy means for their development. Such a propagation of uncertainties using a brute-force method suffers from the "curse of dimensionality" because of the large number of chemical kinetic parameters, implying a practical impossibility with the current means of computation which justifies the development of efficient methods.

The objective of the thesis is the development of a reduced model that can be used for uncertainties propagation in Large Eddy simulations. The handling and implementation of various tools resulting from the uncertainties propagation framework has been an essential preliminary work in this thesis in order to bring this knowledge and skills into the EM2C laboratory.

The method developed in this thesis for the propagation of chemical kinetic

parameters uncertainties is limited to chemistry models in which the advancement of the combustion process is summarized by the evolution of a progress variable given by a transport equation, the access to other informations being made through the use of a table. Through the study of the evolution of a constant pressure adiabatic reactor containing a homogeneous mixture of air and dihydrogen, it is shown that a large part of the uncertainties of such a system can be explained by the uncertainties of the progress variable. This makes it possible to define a chemical table that can be used to propagate uncertainties of chemical kinetic parameters in Large Eddy Simulations. The introduction of the uncertainties is then done only by the modeling of the source term present in the transport equation of the progress variable, which can be parameterized with the help of few uncertain parameters thus avoiding the "curse of dimensionality".

Résumé

La simulation numérique joue aujourd'hui un rôle majeur dans le domaine de la combustion, que ce soit au niveau de la recherche en offrant la possibilité de mieux comprendre les phénomènes ayant lieu au sein des écoulements réactifs ou au niveau du développement de nouveaux systèmes industriels par une diminution des coûts liés à la conception de ces systèmes. A l'heure actuelle, la simulation aux grandes échelles est l'outil le mieux adapté à la simulation numérique d'écoulements réactifs turbulents. Cette simulation aux grandes échelles d'écoulements réactifs n'est en pratique possible que grâce à une modélisation des différents phénomènes :

— la turbulence est modélisée pour les plus petites structures permettant de n'avoir à résoudre que les plus grandes structures de l'écoulement et ainsi réduire le coût de calcul
— la chimie des différentes espèces réactives est modélisée à l'aide de méthodes de réduction permettant de considérablement réduire le coût de calcul

La maturité de la simulation aux grandes échelles d'écoulements réactifs en fait aujourd'hui un outil fiable, prédictif et prometteur. Il fait désormais sens de s'intéresser à l'impact des paramètres impliqués dans les différents modèles sur le résultat de la simulation. Cette étude de l'impact des paramètres de modélisation peut être vue sous l'angle de la propagation d'incertitudes, et peut donner des informations intéressantes à la fois d'un côté pratique pour la conception robuste de systèmes mais également d'un côté théorique afin d'améliorer les modèles utilisées et d'orienter les mesures expérimentales à réaliser afin d'améliorer la fiabilité de ces modèles.

Le contexte de cette thèse est le développement de méthodes efficaces permettant la propagation d'incertitudes présentes dans les paramètres de cinétique chimique des mécanismes réactionnels au sein de simulation aux grandes échelles, ces méthodes devant être non intrusive afin de profiter de l'existence des différents codes de calcul qui sont des outils nécessitant de lourds moyens pour leur développement. Une telle propagation d'incertitude à l'aide d'une méthode de force brute souffre du "fléau de la dimension" du fait du grand nombre de paramètres de cinétique chimique, impliquant une impossibilité pratique avec les moyens de calculs actuels et justifiant le développement de méthodes efficaces.

L'objectif de la thèse est donc le développement d'un modèle réduit utilisable pour la propagation d'incertitudes dans la simulation aux grandes échelles. La prise en main et l'implémentation de différents outils issus de la propagation d'incertitudes a été un travail préliminaire indispensable dans cette thèse afin d'amener ces connaissances et compétences au sein du laboratoire EM2C.

La méthode développée dans cette thèse pour la propagation d'incertitudes des paramètres de cinétique chimique se restreint aux cas d'une modélisation de la chimie dans laquelle l'avancement du processus de combustion est résumé par l'évolution d'une variable d'avancement donnée par une équation de transport, l'accès aux autres informations se faisant grâce à l'utilisation d'une table. Au travers de l'étude de l'évolution d'un réacteur adiabatique à pression constante contenant un mélange homogène d'air et de dihydrogène, il est montré qu'une grande partie des incertitudes d'un tel système peuvent être expliquées grâce aux incertitudes de la variable d'avancement. Cela permet de définir une table chimique utilisable pour la propagation d'incertitudes des paramètres de cinétique chimique dans les simulations aux grandes échelles. L'introduction des incertitudes se fait alors uniquement par la modélisation du terme source présent dans l'équation de transport de la variable d'avancement, lequel peut être paramétré à l'aide de quelques paramètres incertains évitant ainsi le "fléau de la dimension".

Table des matières

Liste des tableaux

Table des figures

Chapitre 1

Introduction

1.1 Contexte

1.1.1 Enjeux du siècle

La révolution industrielle a indéniablement été pour les sociétés humaines modernes un tournant majeur qui a modifié en profondeur l'ensemble de l'organisation économique et sociale mondiale. Cette transformation radicale de nos sociétés s'est accompagnée de l'exploitation des ressources d'énergies fossiles et encore aujourd'hui, la stabilité et la prospérité de nos sociétés reposent sur la consommation massive de ces énergies fossiles. Deux problèmes majeurs se présentent cependant aujourd'hui du fait de cette consommation massive d'énergies fossiles. D'une part, la définition même des énergies fossiles implique que leur exploitation ne pourra se faire indéfiniment, et le pic de production de celle-ci est actuellement suspecté d'être en cours en ce qui concerne le pétrole comme le montre la figure 1.1.

Il est à noter que le pic pétrolier mondial pour le pétrole conventionnel est d'ores et déjà passé, et que l'augmentation de la production de pétrole ces dernières années est dûe à l'exploitation des ressources non conventionnelles, qui correspondent aux ressources en eau profonde (plus de 500 m de profondeur), aux huiles lourdes, aux sables bitumineux, mais également à l'huile de schiste obtenue par fracturation de la roche. Concernant le charbon et le gaz naturel, qui sont les autres principales ressources fossiles, le pic de production ne semble pas encore atteint, particulièrement pour le charbon dont les ressources mondiales prouvées correspondent à plus de 150 ans des ressources actuelles. Cependant, le pétrole représente actuellement plus du tiers de la consommation mondiale d'énergie primaire, en partie du fait de sa facilité d'utilisation pour l'ensemble des usages des transports, propre à son caractère liquide, plaçant le pic de production de pétrole comme un problème majeur. En plus de ce phénomène de raréfaction des ressources fossiles qui représentent encore aujourd'hui plus de 75% de la consommation d'énergie primaire humaine, s'ajoute l'impact climatique de la combustion massive de ressources fossiles lié à "l'effet de serre"

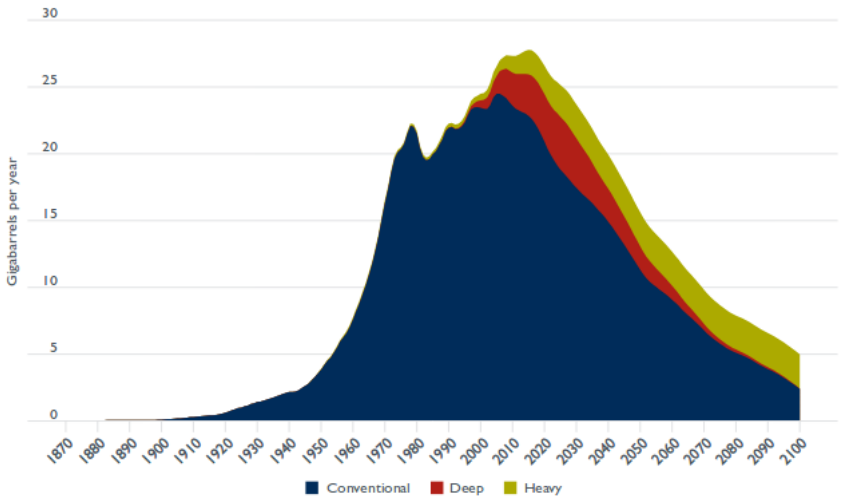

Figure 1.1 – *Simulation de la production de pétrole mondiale en milliard de barils par an pour la période* $1870-2100$*, répartie suivant le pétrole conventionnel, le pétrole profond et le pétrole lourd [1].*

induit par l'importante augmentation de la concentration de dioxyde de carbone dans l'atmosphère. Afin de n'avoir une hausse de la température moyenne globale de seulement 2^oC d'ici à 2100 par rapport à 1850, il ne faut pas rejeter dans l'atmosphère plus que la quantité de CO_2 déjà rejeté dans l'atmosphère depuis le début de la révolution industrielle, ce qui implique une réduction de nos émissions de gaz à effet de serre dès maintenant. A ce problème climatique s'ajoute également les pollutions environnementales et les problèmes sanitaires résultant des produits de la combustion de ces énergies fossiles, bien que la combustion de combustible non fossile tel que la biomasse peut également être source de pollutions atmosphériques et de problèmes sanitaires. L'ensemble de ces problèmes imposent la transition énergétique de nos économies afin de se passer à terme des énergies fossiles. Bien des fronts sont possibles afin de relever l'important défi que représente la transition énergétique, mais il est certain que les ressources fossiles joueront encore un rôle majeur dans le siècle à venir.

Compte tenu de la diminution à venir de la ressource pétrolière alors même que la demande ne cesse de croître, une approche consiste à considérer de nouveaux carburants liquides. Ces autres carburants liquides proviennent pour partie de l'exploitation des ressources fossiles gazières, mais également de la production de biocarburants issus de la biomasse. Sur la figure 1.2 est présenté une simulation de la production de l'ensemble des liquides sur une période allant de 1870 à 2100.

L'intégration d'autres liquides ne permet pas d'enrayer la diminution de la production de liquides, mais permet de repousser à plus tard la pénurie annoncée, au prix de la diversification des combustibles utilisées. L'amélioration de l'efficacité énergétique des systèmes de conversions énergétiques est éga-

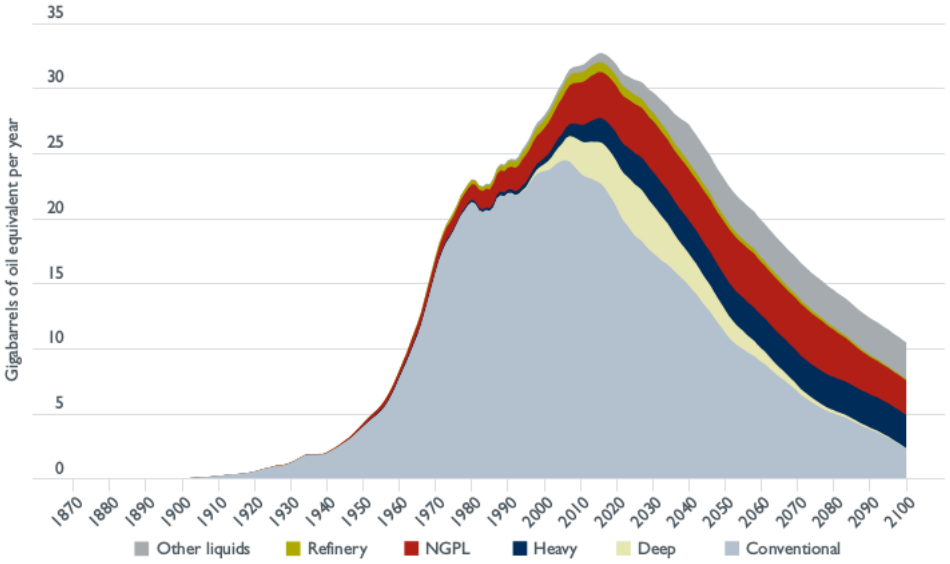

Figure 1.2 – *Simulation de la production de liquides mondiale en milliard de barils par an pour la période* $1870-2100$*, répartie suivant le pétrole conventionnel, le pétrole profond, le pétrole lourd, les liquides issues de l'exploitation gazière, les excédents issues du raffinage de pétroles lourds ainsi que les autres liquides constitués essentiellement de biocarburants [1].*

lement une solution permettant de retarder la pénurie annoncée de pétrole, permettant de faire plus avec moins. En parallèle, l'augmentation de l'efficacité énergétique des systèmes de conversion énergétique permet également une réduction de l'émission de CO_2 qui est globalement directement proportionnelle à la quantité consommée de combustible. En plus d'améliorer l'efficacité énergétique des systèmes de conversion énergétique, il est également nécessaire de réduire leurs pollutions afin de réduire l'impact environnemental et sanitaire de ceux-ci. L'amélioration de ces systèmes ne peut se faire qu'avec une connaissance fine des processus physiques et chimiques prenant place en leur sein, qui passe à la fois par le développement d'outils de diagnostiques mais également par la simulation numérique qui est devenue à ce jour un outil incontournable, à la fois dans le monde de la recherche pour l'amélioration de la compréhension des phénomènes, mais également dans le domaine industriel pour le prototypage des systèmes de conversions énergétiques.

1.1.2 Rôle de la simulation numérique

La collecte directe d'informations sur les phénomènes de combustion ayant lieu au sein des systèmes de conversion peut se faire de deux façons :

— avec des instruments de mesures placés au sein des systèmes, perturbant plus ou moins l'écoulement

— avec des diagnostics optiques nécessitant des accès optiques et donc

une modification du système

Les informations recueillies sont toujours parcellaires, la composition du mélange en tout point de l'écoulement n'étant par exemple pas connue. La simulation numérique est un moyen pour accéder à une plus grande quantité d'informations sur un système à un coût plus faible. La fiabilité des informations obtenues dépend cependant des modèles utilisés pour réaliser la simulation numérique, et de sa capacité à reproduire correctement les phénomènes de combustion. La maturité des outils de simulation numérique pour les phénomènes de combustion amène aujourd'hui à se pencher sur l'étude de la fiabilité de ces modèles. En particulier, la modélisation de la chimie des combustibles présente encore aujourd'hui des incertitudes significatives sur certains paramètres des modèles, même pour un carburant simple comme le dihydrogène. La complexité et la diversité des carburants risquant d'être utilisés à l'avenir amène donc à se questionner dès maintenant sur l'impact des incertitudes résidant dans la modélisation de la chimie, et en particulier de leur impact sur les résultats des simulations numériques des systèmes de conversion énergétiques.

Cette thèse s'inscrit dans le cadre du développement de méthodologie permettant l'étude de l'impact de la méconnaissance des processus chimiques sur le comportement des systèmes de conversion énergétique, avec pour but ultime de pouvoir aider à la prise de décision lors de la phase de conception de ces systèmes.

1.2 Prise en compte d'incertitudes

Cette section a pour objectif de cadrer l'objet d'étude, en particulier de quelles incertitudes cette thèse propose d'étudier l'impact. Ce cadrage est nécessaire puisque la nature des incertitudes détermine l'approche à envisager pour l'étude de leur impact.

1.2.1 Incertitudes épistémiques ou aléatoires

Il est courant de distinguer deux types d'incertitudes, notamment dans les domaines d'évaluation des risques et/ou de fiabilité [114]. Ces deux types d'incertitudes sont respectivement les incertitudes aléatoires et les incertitudes épistémiques.

Les incertitudes aléatoires correspondent aux incertitudes intrinsèques que peut posséder un système. Ces incertitudes intrinsèques sont par exemple présentes dans les systèmes régis par la mécanique quantique qui postule l'existence d'un hasard irréductible et qui à ce jour n'a été mis en défaut par aucune expérience. Il apparaît que par nature, certaines mesures sont nécessairement entachées d'incertitudes, et cela quel que soit l'appareil de mesure utilisé en vertu du principe d'incertitude de Heisenberg. En fait, même pour des systèmes non quantiques, il existe des incertitudes de mesures qui seront toujours présentes du fait qu'une précision infinie ne sera jamais accessible. L'erreur de mesure

commise peut entraîner une incapacité à prédire correctement l'état futur du système si celui-ci est chaotique par exemple. L'imprécision de la mesure en dehors du cas quantique peut encore une fois être considérée comme une incertitude aléatoire, même si ces incertitudes peuvent également être classées dans la seconde catégorie d'incertitudes que sont les incertitudes épistémiques.

Les incertitudes épistémiques correspondent aux incertitudes liées à un manque d'information sur le système physique étudié, mais qui peuvent être réduites en améliorant la connaissance du système. Les erreurs de mesures peuvent donc être considérées comme des incertitudes épistémiques, qui ne pourront cependant être réduites complètement. Une autre forme d'incertitudes épistémiques correspond à la modélisation mathématique utilisée, qui peut ne pas correspondre à la réalité physique en raison d'un manque de connaissance, mais aussi du fait de la volonté de simplifier un problème trop complexe pour qu'il puisse être résolu.

La différentiation entre ces deux types d'incertitudes est utile afin de savoir quels outils utiliser pour l'étude de l'impact de ces dernières. Différents formalismes mathématiques ont été développés et sont en cours d'étude constituant un domaine actif de recherche[114]. Les incertitudes épistémiques s'avèrent plus complexes à prendre en compte, notamment celles concernant les incertitudes sur la modélisation utilisée, et ne seront pas traitées dans le cadre de cette thèse. Seules les incertitudes aléatoires seront ici prises en compte, et plus particulièrement les incertitudes sur les paramètres de cinétique chimique des modélisations des processus de combustion. Plus particulièrement, l'accent sera mis sur l'impact de ces incertitudes, qui seront une information à priori donnée par le travail d'autres équipes de recherches, sur les résultats de simulations. Deux aspects sont donc à distinguer, le premier correspondant à la définition des incertitudes des paramètres de cinétique chimique et le second correspondant à l'étude de l'impact de ces incertitudes, ces deux aspects étant respectivement nommés "quantification d'incertitude" et "propagation d'incertitude" dans le cadre de cette thèse.

1.2.2 Quantification et propagation d'incertitudes

La quantification d'incertitudes est de manière générale l'activité consistant à chercher à caractériser les incertitudes présentes au sein de modélisations de la réalité. Dans le cas où un modèle a été choisi afin de représenter la réalité, il est possible de séparer les différentes grandeurs du modèle en deux catégories : les entrées et les sorties.

Les entrées correspondent aux paramètres du modèle ainsi qu'aux conditions initiales et limites du système nécessaires à l'obtention d'une solution. Ces entrées peuvent être entachées d'incertitudes, et la caractérisation de ces incertitudes correspond à ce que l'on appellera dans cette thèse la quantification d'incertitudes.

Les sorties correspondent aux grandeurs de la solution obtenue qui dé-

pendent des entrées. Les incertitudes de ces sorties dépendront donc des incertitudes des entrées, et la caractérisation des incertitudes des sorties, étant données les incertitudes des entrées, correspond à ce qu'on appellera dans cette thèse la propagation d'incertitudes.

L'objectif de ce travail de thèse est de développer des méthodes de propagation d'incertitudes efficaces dans le domaine de la simulation numérique aux grandes échelles de la combustion, reposant sur un travail de quantification d'incertitudes préalable qui n'est pas l'objet d'étude de cette thèse.

1.2.3 Rappel de théorie des probabilités

L'ensemble des considérations sur la quantification et la propagation d'incertitudes dans cette thèse reposent sur la théorie des probabilités. Une introduction à cette théorie et à ses concepts peut être trouvé dans de nombreux ouvrages [121].

L'objectif de la propagation d'incertitude est en pratique l'obtention d'informations stochastiques sur des quantités d'intérêts, qui peuvent être l'espérance aussi appelée moyenne de ces quantités d'intérêts, leur variance ou encore la probabilité que ces quantités d'intérêts ont d'être dans une plage de valeurs donnée. L'ensemble de ces informations peuvent se ramener à un calcul d'espérance. Ainsi, si l'on considère une grandeur représentée par une variable aléatoire G, la moyenne m_G de cette grandeur et sa variance σ_G^2 seront données par les expressions suivantes :

$$\begin{aligned} m_G &= E\left[G\right] \\ \sigma_G^2 &= E\left[(G - m_G)^2\right] \end{aligned} \tag{1.1}$$

La probabilité $P_A(G)$ que la quantité d'intérêt G prenne ses valeurs dans l'ensemble A est quant à elle donnée par l'expression suivante :

$$P_A(G) = E\left[\mathbb{1}_A(G)\right] \tag{1.2}$$

L'expression précédente fait intervenir la fonction indicatrice $\mathbb{1}_A$ de l'ensemble A, prenant la valeur 1 sur les éléments de A et 0 ailleurs. L'ensemble de ces espérances peuvent être exprimées sous forme d'intégrales, ainsi la moyenne m_G de G peut être exprimée à l'aide d'une intégrale de la façon suivante, dans laquelle π_G est la densité de probabilité de la variable aléatoire G :

$$m_G = \int_\Omega g\pi_G(g)dg \tag{1.3}$$

Les autres informations que sont σ_G^2 et $P_A(G)$ peuvent également être

obtenues à l'aide d'un calcul d'intégrale, faisant du calcul d'intégrale un outil essentiel à la propagation d'incertitudes. En pratique, le calcul exacte des intégrales rencontrées est souvent irréalisable, et l'estimation numérique de celle-ci devient alors obligatoire. Cette estimation numérique des intégrales rencontrées nécessite l'utilisation de méthodes numériques efficaces qui sont présentées dans la présente thèse.

1.2.4 Méthodes intrusives et non-intrusives

Cette section vise à introduire les différentes alternatives envisageables pour la propagation d'incertitudes. On considère un système physique, pour lequel on possède un modèle mathématiques $\mathcal{M}$, dépendant d'un vecteur de paramètres $\mathbf{d}$, et tel que le comportement du système physique est décrit par la solution $\mathbf{u}$, qui dépend du vecteur de paramètres $\mathbf{d}$, au travers de l'équation suivante :

$$\mathcal{M}(\mathbf{u}(\mathbf{d}), \mathbf{d}) = 0 \tag{1.4}$$

Éventuellement, la résolution de l'équation précédente n'est pas possible directement, et des méthodes numériques sont nécessaires à la construction d'une solution numérique approchée, solution de l'équation discrétisée suivante :

$$\mathcal{M}_{disc}(\mathbf{u_{disc}}(\mathbf{d}), \mathbf{d}) = 0 \tag{1.5}$$

Dans le cadre de cette thèse, l'objectif est la propagation d'incertitudes présentes dans une partie des paramètres du modèle, ce qui amène à considérer désormais un vecteur aléatoire $\mathbf{D}$, dont on suppose la loi de probabilité P_D connue, pour représenter les paramètres incertains. Deux approches peuvent alors être considérées.

La première consiste à changer le modèle mathématique afin de prendre en compte le caractère désormais aléatoire du problème directement dans la nouvelle formulation du problème, ce qui correspond à l'approche intrusive de la propagation d'incertitudes, et peut formellement s'écrire sous la forme :

$$\mathcal{M}_{disc,alea}(U_{disc}(\mathbf{D}), \mathbf{D}) = 0 \tag{1.6}$$

La solution de ce nouveau problème est désormais U_{disc}, qui est généralement un champ stochastique. La difficulté de telles méthodes est qu'elle nécessite l'introduction de la formulation $\mathcal{M}_{disc,alea}$ du problème. Cette nouvelle formulation peut introduire des difficultés notamment en terme de coût de calcul et de ressource mémoire qui peut s'avérer être limitant, et nécessite généralement le développement d'algorithme de résolution spécifiques [65]. Pour

cette raison, la méthode suivante lui est préférée dans cette thèse.

Cette autre méthode consiste à utiliser directement le modèle $\mathcal{M}_{disc}$, ce qui permet notamment de ne pas développer de nouveaux outils numériques lorsque ces outils sont déjà disponibles pour résoudre le problème déterministe. Différentes approches peuvent être considérées pour l'utilisation des méthodes non intrusives, mais toutes reposent sur de multiples résolutions de problèmes pouvant s'écrire :

$$\mathcal{M}_{disc}(\mathbf{u}_{\mathbf{disc}}^{(\mathbf{i})}(\mathbf{d}^{(\mathbf{i})}), \mathbf{d}^{(\mathbf{i})}) = 0 \tag{1.7}$$

Dans cette dernière expression, les $\mathbf{d_i}$ sont des valeurs pouvant être prises par le vecteur aléatoire $\mathbf{D}$. Par exemple, dans le cas de l'utilisation d'une méthode de Monte Carlo, ces valeurs correspondent aux valeurs $\mathbf{D}^{(\mathbf{i})}(\omega)$ où les vecteurs aléatoires $\mathbf{D}^{(\mathbf{i})}$ sont des vecteurs aléatoires indépendants entre eux et identiquement distribués suivant la loi du vecteur aléatoire $\mathbf{D}$. Une fois l'ensemble des solutions $\mathbf{u}_{\mathbf{disc}}^{(\mathbf{i})}$ obtenues, celles-ci sont post-traitées afin d'obtenir des informations statistiques sur la solution $\mathbf{U_{disc}}$ recherchées. Le coût d'une telle méthode est globalement proportionnel au nombre N de problèmes résolus, le nombre N de problèmes à considérer dépendant de la convergence de la méthode non-intrusive employée ainsi que de la précision désirée.

1.3 Enjeux et défis de la propagation d'incertitudes de cinétique chimique en LES

Les considérations générales précédentes s'appliquent à la quantification et la propagation d'incertitudes dans n'importe quel domaine. Des considérations spécifiques à chaque domaine peuvent cependant s'ajouter à ces considérations générales. Ces considérations spécifiques à la propagation d'incertitudes de cinétique chimique en simulation aux grandes échelles d'écoulements réactifs sont détaillées dans cette section.

1.3.1 La modélisation de la chimie en combustion

1.3.1.1 Schéma réactionnel et cinétique chimique

La combustion est un processus chimique au cours duquel des espèces réagissant entre elles sont consommées pour produire de nouvelles espèces. Ce processus de consommation et de production d'espèces chimiques peut être modélisé à l'aide d'un mécanisme réactionnel, contenant une liste de réactions chimiques élémentaires décrivant les différentes réactions possibles entre les

différentes espèces, qui s'écrit sous la forme :

$$\forall r \in [\![1, N_r]\!], \sum_k \nu'_{kr} \mathcal{M}_k \rightleftharpoons \sum_k \nu''_{kr} \mathcal{M}_k \tag{1.8}$$

Dans l'expression (1.8), $\mathcal{M}_k$ est l'espèce chimique k et ν'_{kr} et ν''_{kr} sont ses coefficients stoechiométriques pour la réaction r. L'expression (1.8) peut encore s'écrire comme dans (1.9).

$$\begin{cases} \forall r \in [\![1, N_r]\!], \sum_k \nu_{kr} \mathcal{M}_k = 0 \\ \nu_{kr} = \nu'_{kr} - \nu''_{kr} \end{cases} \tag{1.9}$$

La vitesse spécifique volumique de réaction $\mathcal{Q}_r$ de la réaction élémentaire r, caractérisant la vitesse de transformation des espèces chimiques réactives en leur produit, est donnée par l'expression suivante :

$$\mathcal{Q}_r = k_{fr} \prod_k [X_k]^{\nu'_{kr}} - k_{br} \prod_k [X_k]^{\nu''_{kr}} \tag{1.10}$$

Les facteurs k_{fr} et k_{br} sont respectivement la constante de vitesse directe et indirecte de la réaction. L'expression de ces constantes est généralement faite à l'aide de la loi d'Arrhenius, reposant à la fois sur des résultats empiriques ainsi que des considérations venant de la théorie cinétique, et qui s'exprime sous la forme :

$$k_{fr} = A_r T^{\beta_r} \exp\left(-\frac{E_r}{RT}\right) \tag{1.11}$$

Dans cette expression, T correspond à la température et R est la constante des gaz parfaits, les trois autres paramètres étant :

- le facteur pré-exponentiel A_r, modélisant les effets de géométrie et d'orientation des molécules durant les collisions.
- l'exposant β_r, caractérisant l'impact de l'excitation thermique des molécules.
- l'énergie d'activation E_r, caractérisant l'énergie minimum nécessaire pour que la réaction ait lieu.

D'autres lois plus complexes que la loi d'Arrhenius existent permettant de prendre d'autres aspects en considération, notamment les dépendances en pression [76, 44]. La constante de vitesse indirecte k_{br} est quant à elle obtenue comme le rapport entre la constante de vitesse directe k_{fr} et la constante

d'équilibre thermodynamique K_r^{eq} de la réaction :

$$k_{br} = \frac{k_{fr}}{K_r^{eq}} \tag{1.12}$$

Une fois les constantes de vitesse calculées, il est possible d'accéder au taux de réaction chimique $\dot{\omega}_k$ de l'espèce k, qui se déduit des vitesses spécifiques volumiques de réaction $\mathcal{Q}_r$ et utilise la masse volumique ρ du mélange et les masses molaires W_k des différentes espèces chimiques impliquées :

$$\rho\dot{\omega}_k = \sum_{r=1}^{N_r} W_k \nu_{kr} \mathcal{Q}_r \tag{1.13}$$

Ce taux de réaction chimique $\dot{\omega}_k$ est un ingrédient essentiel à l'étude des écoulements réactifs en tant que terme source des équations de transports associées aux fractions massiques Y_k des différentes espèces, qui en utilisant la convention de sommation d'Einstein sont de la forme :

$$\frac{\partial \rho Y_k}{\partial t} + \frac{\partial \rho u_j Y_k}{\partial x_j} + \frac{\partial \mathcal{J}_j^k}{\partial x_j} = \rho\dot{\omega}_k \tag{1.14}$$

Ces équations de transports font intervenir la masse volumique ρ du mélange, la vitesse u du fluide ainsi que le tenseur de diffusion $\mathcal{J}^k$ de l'espèce k dans le mélange. La place du terme source $\dot{\omega}_k$ dans cette dernière équation traduit l'action de la cinétique chimique sur l'ensemble de l'écoulement, et une modification de la cinétique chimique entraîne donc mécaniquement une modification de la solution des équations de l'écoulement.

L'étude de l'impact d'incertitudes dans la cinétique chimique sur l'écoulement nécessite en premier lieu une caractérisation de ces incertitudes, correspondant à la phase de quantification d'incertitudes. Ce travail de quantification d'incertitudes est réalisée par l'ensemble des chimistes de la communauté et est accessible pour les équipes du laboratoire via la littérature.

1.3.1.2 Modélisation des incertitudes dans la cinétique chimique

Les grandes étapes de la construction d'un schéma cinétique sont décrites dans [35], et peuvent être résumées de la sorte :

- Définir un mécanisme réactionnel, c'est à dire l'ensemble des réactions élémentaires pouvant avoir lieu
- Assigner des valeurs aux paramètres permettant le calcul des constantes de vitesses, par la littérature ou par de judicieuses estimations.
- Réaliser ou trouver des expériences dans la littérature qui dépendent d'une partie ou de toutes les constantes de vitesses.

— Calculer numériquement, en utilisant le schéma cinétique maintenant défini, les équations de transports associées aux expériences cibles choisies, donnant ainsi accès numériquement aux grandeurs observables mesurées expérimentalement. Déterminer également la sensibilité de ces grandeurs observables aux paramètres des constantes de vitesses.
— Choisir plusieurs cibles expérimentales sensibles à un sous ensembles des paramètres, et réaliser une optimisation sur ces paramètres.

Une fois le mécanisme réactionnel ainsi que la paramétrisation des constantes de vitesses associées à chacune des réactions élémentaires définies, la détermination des paramètres repose donc sur leur optimisation étant donnés des résultats cibles à atteindre définis par les expériences réalisées. Un tel problème d'optimisation est difficile du fait du nombre important de paramètres, et rien ne garantit l'unicité de la solution. Cette difficulté de détermination des paramètres de cinétiques chimiques, ainsi que les incertitudes sur les mesures expérimentales, engendre des incertitudes sur les paramètres des constantes de vitesses.

Dans la littérature, l'incertitude sur les constantes de vitesses k des schémas cinétiques est souvent résumée à l'aide d'un facteur d'incertitude f, qui est à priori dépendant de la température mais est en pratique souvent considéré constant sur l'ensemble du domaine de température [89]. Différentes définitions existent pour ce facteur d'incertitude, dont $f = k^0/k^{min} = k^{max}/k^0$ [142] et $f = \log_{10}(k^0/k^{min}) = \log_{10}(k^{max}/k^0)$ [132], expressions dans lesquelles k^0 correspond à la valeur nominale du paramètre k, alors que k^{min} et k^{max} correspondent aux valeurs minimales et maximales possibles. Suivant que l'on considère la première ou la seconde expression, la valeur du facteur d'incertitude f appartient à $[1, +\infty[$ ou à $[0, +\infty[$ respectivement. En fait, l'ensemble de ces définitions se retrouve dans le fait que la valeur de k se trouve avec forte probabilité dans l'intervalle $[k^0/f, k^0 f]$. Il est courant de donner du sens à ce qui est entendu par forte probabilité au sein de cet intervalle, et cela peut se traduire par une loi uniforme [142] sur celui-ci, mais il est plus d'usage de considérer une loi log-normale pour ce paramètre [132].

Le paramètre k suit généralement une loi d'Arrhenius, ou une version plus complexe de celle-ci, et l'incertitude sur ce paramètre peut être dû à une incertitude présente sur les paramètres de cette loi d'Arrhenius. En fait, il a été montré que la considération d'un facteur d'incertitude f indépendant de la température était équivalent à ce que seul le facteur pré-exponentiel A de la loi d'Arrhenius était incertain [89]. Cependant, l'hypothèse d'un facteur d'incertitude f ne dépendant pas de la température est peu physique, et est utilisée par manque d'informations sur les incertitudes [89]. Des méthodes ont ainsi été proposées afin d'obtenir une dépendance en température du facteur d'incertitude f, et considéraient l'ensemble ou une partie des paramètres de la loi d'Arrhenius comme incertains. En plus d'être incertains, des corrélations existaient entre les paramètres de la loi d'Arrhenius pour une même réaction, et l'objectif de ces méthodes étaient la détermination de la loi du vecteur aléatoire

de paramètres. Une première méthode [89] utilise une méthode d'optimisation afin de déterminer la matrice de covariance du vecteur aléatoire $(\ln(A), \beta, E/R)$ des paramètres à optimiser, et a été utilisée notamment sur un mécanisme du H_2/CO en optimisant 57 paramètres appartenant à 17 réactions différentes [93]. Pour cette méthode, seule la matrice de covariance est déterminée, ce qui n'est pas suffisant pour déterminer la loi de probabilité jointe des paramètres. Les auteurs proposent l'utilisation d'une loi gaussienne pour ce vecteur de paramètres. Une seconde méthode [90] utilise l'inférence bayésienne afin de déterminer la loi jointe des paramètres $(\ln(A), \ln(E))$ de la loi d'Arrhenius. Ces deux dernières méthodes permettent l'obtention de loi jointe des paramètres de la loi d'Arrhenius d'une réaction, mais ne considèrent pas de dépendances entre les paramètres de réactions différentes. Cette hypothèse d'indépendance des incertitudes entre les paramètres de réactions différentes est actuellement systématiquement faite dans l'ensemble des études de propagation d'incertitudes de cinétique chimique.

Tout comme un schéma cinétique est la donnée d'un mécanisme réactionnel et d'une loi pour la constante de vitesse de chaque réaction élémentaire ainsi qu'une valeur pour les paramètres de cette loi, on peut définir un schéma cinétique incertain comme la donnée d'un schéma cinétique et d'une loi de probabilité jointe pour l'ensemble des paramètres définissant les constantes de vitesses des réactions. Un tel schéma cinétique incertain est l'ingrédient de base pour la propagation d'incertitudes de cinétique chimique. A l'heure actuelle, la détermination des facteurs d'incertitudes n'est pas systématiquement réalisée pour l'ensemble des réactions des mécanismes réactionnels, et un travail important reste à faire au sein de la communauté afin d'obtenir ces informations. En l'absence d'informations ou en présence d'informations partielles, il est nécessaire de faire des hypothèses qui ne seront bien évidemment pas sans impact.

1.3.2 Prise en compte de la combustion en LES

Une fois choisi le mécanisme réactionnel, il est possible d'envisager une simulation aux grandes échelles d'un écoulement réactif. L'utilisation directe du mécanisme réactionnel implique de considérer une équation de transport par espèce chimique, ce qui s'avère être très coûteux en temps de calcul, rendant cela impossible à l'heure actuelle pour de nombreuses applications pratiques impliquant des carburants lourds. Afin de contourner ce coût de calcul prohibitif, une modélisation de la chimie peut être faite rendant possible les simulations aux grandes échelles d'écoulements réactifs. Actuellement, il est possible de distinguer deux grandes familles de modèles :

— la chimie réduite
— la chimie tabulée

1.3.2.1 Chimie réduite

1.3.2.1.1 Mécanismes squelettiques Un mécanisme squelettique est obtenu à partir d'un mécanisme détaillé en éliminant des espèces et des réactions qui ont un effet négligeable sur le phénomène d'intérêt. De nombreuses méthodes existent afin d'éliminer des réactions [60, 83, 143] et également pour l'élimination d'espèces [78, 133]. Avec un tel mécanisme squelettique, les informations concernant les espèces éliminées sont définitivement perdues, mais les taux de réactions des espèces principales ne sont pas significativement impactés de sorte que les caractéristiques macroscopique du phénomène de combustion (délai d'auto-allumage, réponse à l'étirement, vitesse de flamme laminaire, etc...) sont correctement décrites. Cependant, les mécanismes squelettiques sont souvent encore trop important pour être utilisés directement en LES, mais peuvent néanmoins être utilisés pour la construction de mécanismes plus réduits, ou même pour la construction de table chimique.

1.3.2.1.2 Mécanismes chimiques réduits L'objectif est ici d'obtenir des mécanismes chimiques réduits, qui sont des versions très simplifiées de la réelle chimie, capable de reproduire les caractéristiques d'intérêt du phénomène de combustion, tout en assurant de ne garder qu'un nombre restreint d'espèces (de l'ordre de la dizaine maximum) et de réactions afin d'avoir un coût de calcul raisonnable. Deux approches différentes existent pour la construction de mécanismes chimiques réduits.

1.3.2.1.2.1 Mécanismes analytiques La construction d'un mécanisme analytique [79, 99] consiste en la simplification d'un mécanisme squelettique à l'aide de l'approximation de l'état quasi-stationnaire pour certaines espèces et à l'hypothèse d'équilibre partiel pour certaines réactions.

L'approximation de l'état quasi-stationnaire consiste à considérer que si une espèce k est telle que son taux de création est faible devant son taux de consommation, alors on a $\dot{\omega}_k \approx 0$. Cela permet d'écrire la relation suivante :

$$\dot{\omega}_k = \sum_{r=1}^{N_r} \nu_{kr} \left[k_{fr} \prod_{i=1}^{N_s} [X_i]^{\nu'_{ir}} - k_{br} \prod_{i=1}^{N_s} [X_i]^{\nu''_{ir}} \right] = 0 \tag{1.15}$$

Cette dernière relation permet de calculer la concentration de l'espèce k à partir des concentrations des autres espèces, ce qui permet de ne plus considérer la résolution de l'équation de transport de l'espèce k.

L'hypothèse d'équilibre partiel permet de simplifier encore plus le mécanisme squelettique. Cette hypothèse peut être faite pour la réaction r dès lors que les temps caractéristiques associés au sens direct et inverse de cette réaction sont négligeables par rapport aux temps caractéristiques associés aux autres réactions. La réaction r est alors considérée en équilibre partiel, se traduisant par

la relation suivante.

$$\mathcal{Q}_r = k_{fr} \prod_{i=1}^{N_s} [X_i]^{\nu'_{ir}} - k_{br} \prod_{i=1}^{N_s} [X_i]^{\nu''_{ir}} = 0 \tag{1.16}$$

Les mécanismes analytiques permettent de conserver un nombre d'espèces plus important que le nombre d'équations de transports présentes. La réduction qui est effectuée dépend du mécanisme squelettique initiale et le nombre d'équations de transports à considérer restant peut être jugé trop important pour avoir une simulation à un coût raisonnable. Les mécanismes ajustés sont alors une alternative intéressante.

1.3.2.1.2.2 Mécanismes ajustés Contrairement aux mécanismes analytiques dont la réduction n'est pas prévisible à l'avance, les mécanismes ajustés [34, 57] sont conçus pour répondre aux besoins et aux contraintes de la simulation envisagée. De tels mécanismes consistent généralement en la considération des espèces majoritaires et éventuellement d'espèces d'intérêt uniquement, ainsi qu'un nombre de réactions très limité composé d'équations bilans. Les paramètres de réactions pour de tels schémas sont obtenus par optimisation afin de reproduire au mieux certaines caractéristiques du processus de combustion nécessaire à la LES, et cela pour un domaine thermodynamique qui se trouvera présent dans la LES à réaliser. L'avantage de tels mécanismes est le coût de calcul lié à leur utilisation comparés aux mécanismes analytiques. L'inconvénient est qu'ils ont été ajustés pour reproduire certaines caractéristiques dans des conditions précises, rendant leur utilisation possible uniquement dans des cas spécifiques. Il est de plus impossible d'avoir accès à des espèces minoritaires absentes d'un tel mécanisme par exemple, et une LES utilisant un tel mécanisme offre donc une moindre richesse d'informations que si un mécanisme analytique avait été utilisé.

La chimie tabulée est une alternative permettant de réduire le nombre d'équations de transport pour les espèces à une seule tout en permettant d'assurer une information riche en conservant par exemple l'ensemble des espèces chimiques.

1.3.2.2 Chimie tabulée

L'état d'un écoulement réactif gazeux peut être caractérisé par un champ de vitesse, un champ de température, un champ de pression ainsi qu'un champ de fraction massique des espèces chimiques présentes [102]. L'ensemble des états thermodynamiques rencontrés dans un tel écoulement réactif au cours du temps peuvent donc être caractérisés par l'ensemble des points suivant qui appar-

tiennent à un espace à $N_s + 2$ dimensions.

$$\psi(\mathbf{x}, t) = (Y_1(\mathbf{x}, t), \ldots, Y_{N_s}(\mathbf{x}, t), T(\mathbf{x}, t), P(\mathbf{x}, t)) \tag{1.17}$$

La chimie tabulée repose sur l'utilisation d'une variété de dimension $P < N_s + 2$ sur laquelle se situent approximativement les états $\psi(\mathbf{x}, t)$. Une telle variété correspond intuitivement à une variété d'équilibre pour les phénomènes chimiques suffisamment lents, une perturbation d'un point de cette variété en dehors de celle-ci relaxant très rapidement vers celle-ci.

Le principe de la tabulation de la cinétique chimique consiste à considérer une paramétrisation pour cette variété de dimension $P < N_s + 2$ qui est faite à l'aide de P paramètres de contrôles $(\phi_k)_{1 \leq k \leq P}$. L'état thermodynamique ψ dans l'espace des phases se situant sur la variété peut alors s'exprimer comme une fonction $\mathcal{G}$ de ces paramètres de contrôle :

$$\psi = \mathcal{G}(\phi_1, \ldots, \phi_P) \tag{1.18}$$

N'importe quelle grandeur $\mathcal{H}$ dépendant de l'état thermodynamique ψ peut également s'exprimer comme une fonction $\hat{\mathcal{H}}$ des paramètres de contrôles par :

$$\mathcal{H}(\psi) = \mathcal{H}(\mathcal{G}(\phi_1, \ldots, \phi_P)) = \hat{\mathcal{H}}(\phi_1, \ldots, \phi_P) \tag{1.19}$$

En pratique, les fonctions $\hat{\mathcal{H}}$ ne possèdent pas d'expression analytique simple, et une table est construite afin de contenir les valeurs prises par ces fonctions aux nœuds d'un maillage de l'espace des variables de contrôle, la reconstruction en tout autre point étant réalisée à l'aide d'une méthode d'interpolation.

Une des difficultés de la chimie tabulée est la détermination des paramètres de contrôles et de la fonction $\mathcal{G}$. Plusieurs méthodes ont été développées afin de déterminer les paramètres de contrôles ainsi que la fonction $\mathcal{G}$ caractérisant la variété en se basant sur des arguments mathématiques (tel que ILDM [80]) ou physiques [33]. Ces dernières méthodes reposent sur des configurations physiques et supposent que l'écoulement n'influence pas significativement le processus chimique de combustion, de sorte que la structure de la flamme au sein du système peut être décrite à l'aide d'une famille de prototypes de flammes simples. Suivant le mode de combustion rencontré, différents prototypes de flammes et différents paramètres de contrôles sont considérés [140].

La chimie tabulée est un outil puissant mais qui possède cependant quelques limitations. La détermination du mode de combustion n'est pas toujours simple et un mauvais choix du prototype de flammes peut engendrer des résultats incorrects. Un autre problème provient du nombre de paramètres de contrôles à

utiliser, qui peut augmenter rapidement avec la complexité du problème, causant des problèmes pratiques de mémoire pour le stockage de la table.

1.3.3 Propagation d'incertitudes en LES d'écoulements réactifs

Une revue critique des propagation d'incertitudes dans les écoulements réactifs est faite dans cette section, en mettant l'accent sur les méthodes jusqu'à présent investiguées pour propager les incertitudes tout en insistant sur les éventuelles hypothèses prises, ainsi que les justifications ou les conséquences de ces hypothèses. Seules les études pour lesquelles les incertitudes influencent l'écoulement seront répertoriées dans cette section, une étude ayant été menée avec pour objectif une estimation d'erreur dans la production de suies via un formalisme de propagation d'incertitude [88], mais dans laquelle les incertitudes introduites n'influençaient pas l'écoulement.

1.3.3.1 Incertitudes non liées à la cinétique chimique

Une étude [61] a été menée pour propager les incertitudes sur trois paramètres, qui sont le coefficient de Smagorinsky, ainsi que les nombres de Prandtl et de Schmidt turbulents, le tout sur la simulation aux grandes échelles d'une flamme stabilisée par un corps non profilé [82]. L'objectif de l'étude est la construction de surfaces de réponse de grandeurs d'intérêts en fonction de ces trois paramètres, permettant d'obtenir des informations, notamment sur le choix des paramètres des modèles de sous maille utilisés et l'impact de ces choix, à un moindre coût.

Les trois paramètres variables dans cette étude sont considérés comme des variables aléatoires indépendantes entre elles, et suivant chacune des lois uniformes sur un intervalle défini grâce aux valeurs extrêmes de ces paramètres utilisées dans la littérature. La dimension stochastique de l'étude est donc de 3, et le choix a été fait d'utiliser une méthode non intrusive pour la propagation d'incertitude, permettant l'utilisation de code déjà existant. La propagation d'incertitude consiste en le calcul de surfaces de réponses pour représenter les grandeurs d'intérêts à l'aide d'expansions en polynômes du chaos généralisés, les coefficients de ces expansions étant obtenus grâce à une méthode de cubature de Smolyak utilisant une quadrature imbriquée de Clenshaw-Curtis (qui seront présentées au chapitre 2). La tensorisation creuse de Smolyak a permis aux auteurs de ne considérer que 25 calculs LES pour le calcul des coefficients malgré la dimension stochastique de 3, chacun des calculs se voyant attribué une valeur pour chacun des trois paramètres donnée par le point de cubature. La qualité des surfaces de réponses est évaluées dans l'étude à l'aide de 18 calculs LES supplémentaires pour des valeurs de paramètres choisies aléatoirement, et les auteurs mettent l'accent sur l'utilisation simple et peu coûteuse des surfaces de réponses.

L'intérêt majeur de cette étude est de montrer la possibilité de la propa-

gation d'incertitudes dans la simulation aux grandes échelles d'un écoulement réactif en utilisant des méthodes non-intrusives efficaces pour un nombre de paramètres incertains raisonnable, et également de montrer les possibilités de construction de surfaces de réponse permettant d'obtenir des informations supplémentaires pour un coût de calcul négligeable. L'évaluation de l'erreur commise permet de donner un ordre d'idée de l'efficacité des méthodes employées, et montre la difficulté d'évaluation des résultats obtenus par de telles méthodes, celle-ci ayant nécessitée 18 calculs LES supplémentaires. Ce nombre est jugé faible par les auteurs, puisqu'il ne permet pas selon eux de trancher sur les causes de l'augmentation de l'erreur commise par la surface de réponse lorsque le degré maximal utilisé passe de 1 à 2. L'introduction d'incertitudes dans les paramètres choisis ne nécessitait pas de modélisation supplémentaire et les calculs pour la propagation d'incertitude pouvait donc être menés directement, ce qui n'est pas le cas lorsque des incertitudes dans les coefficients paramétrant la cinétique chimique sont considérées.

1.3.3.2 Incertitudes liées à la cinétique chimique

Une seule propagation d'incertitudes de cinétique chimique dans une simulation aux grandes échelles a été réalisée à ce jour [87]. L'étude a été faite sur la flamme de Sandia D [11], qui est une flamme jet turbulente pilotée partiellement pré-mélangée de méthane, pour laquelle de nombreuses mesures expérimentales sont disponibles. L'objectif de cette étude est d'obtenir l'impact des incertitudes de la chimie sur l'écoulement. Les incertitudes obtenues se trouvent être de l'ordre de grandeur des incertitudes de mesures, conduisant les auteurs à conclure que les incertitudes présentes au niveau de la cinétique chimique sont à l'heure actuelle trop importantes pour pouvoir valider correctement les modèles de combustion turbulente sur les résultats expérimentaux.

Encore une fois, l'étude réalisée propage les incertitudes grâce à une méthode non intrusive, permettant de réutiliser les méthodes et codes de calculs déjà disponibles pour les simulations déterministes habituelles. Chacune des simulations aux grandes échelles à effectuer utilise le modèle de flammelette quasi-stationnaire [100], pour lequel une flammelette dépend de la fraction de mélange Z et d'un taux de dissipation scalaire de référence χ_{ref}. Ainsi, chaque grandeur X (température, densité, fractions massiques ...) obtenue par la résolution de l'équation de la flammelette est une fonction de Z et de χ_{ref}. La modélisation de sous-mailles est effectuée en supposant que la fraction de mélange suit une loi bêta paramétrée par la fraction de mélange filtrée $\widetilde{Z}$ et la variance de la fraction de mélange de sous maille $\widetilde{Z''^2}$. Les grandeurs tabulées sont les grandeurs filtrées $\widetilde{X}$ données par l'expression (1.20), et dépendant de la fraction de mélange filtrée $\widetilde{Z}$, de la variance de la fraction de mélange de

sous-maille $\widetilde{Z''^2}$ et du taux de dissipation scalaire χ_{ref}.

$$\widetilde{X}(\widetilde{Z}, \widetilde{Z''^2}, \chi_{ref}) = \int_0^1 X(Z, \chi_{ref}) \beta(Z; \widetilde{Z}, \widetilde{Z''^2}) dZ \tag{1.20}$$

La fraction de mélange filtrée $\widetilde{Z}$ est obtenue à l'aide d'une équation de transport, alors que la variance de la fraction de mélange de sous maille $\widetilde{Z''^2}$ et le taux de dissipation scalaire χ_{ref} sont obtenues à l'aide de modèles algébriques.

Les incertitudes sont introduites via les coefficients $\mathbf{k_f} = (k_{fr})_{1 \leq r \leq N_r}$ pour les réactions directes qui sont considérés comme des variables aléatoires indépendantes entre elles, et suivant toutes une loi log-normale. Afin d'obtenir les grandeurs présentes dans la table données par l'expression (1.20), il est nécessaire d'obtenir les grandeurs non filtrées $X(Z, \chi_{ref}, \mathbf{k_f})$ dépendant désormais également du vecteur aléatoire $\mathbf{k_f}$. Les auteurs ont obtenu des statistiques des grandeurs filtrées de l'expression (1.20) à l'aide d'un échantillonnage par hypercube latin du vecteur aléatoire $\mathbf{k_f}$. Ce faisant, ils ont noté que la plupart des variables aléatoires $X(Z, \chi_{ref}, \mathbf{k_f})$ suivaient une loi normale. Cela les a motivé à considérer une loi normale pour la loi jointe de plusieurs de ces variables aléatoires. Il est à partir de là possible de considérer la version stochastique des grandeurs filtrées $\widetilde{X}$, qui sont désormais donnée par l'expression (1.21).

$$\widetilde{X}(\widetilde{Z}, \widetilde{Z''^2}, \chi_{ref}, \mathbf{k_f}) = \int_0^1 X(Z, \chi_{ref}, \mathbf{k_f}) \beta(Z; \widetilde{Z}, \widetilde{Z''^2}) dZ \tag{1.21}$$

Le calcul numérique de l'expression (1.23) est réalisé par les auteurs à l'aide d'une méthode de quadrature, donnant l'expression suivante :

$$\widetilde{X}(\widetilde{Z}, \widetilde{Z''^2}, \chi_{ref}, \mathbf{k_f}) \approx \sum_{i=1}^{N} p_i X(Z_i, \chi_{ref}, \mathbf{k_f}) \tag{1.22}$$

Les poids $(p_i)_{1 \leq i \leq N}$ ainsi que les nœuds de la quadrature $(Z_i)_{1 \leq i \leq N}$ ont à priori une dépendance en $\widetilde{Z}$ et $\widetilde{Z''^2}$ qui est omise dans l'expression précédente, bien que pour des raisons pratiques il est fort possible que les nœuds aient été fixés par les auteurs afin de limiter le nombre de calcul, ne laissant une dépendance qu'aux poids de la quadrature. L'explication de ce qui remplit la table utilisée dans le calcul LES est ensuite mentionné. Tout d'abord, l'espérance des variables aléatoires $\widetilde{X}(\widetilde{Z}, \widetilde{Z''^2}, \chi_{ref}, .)$ est placée dans la table, qui est elle même

calculée en utilisant la quadrature suivante :

$$E\left[\widetilde{X}(\widetilde{Z},\widetilde{Z''^2},\chi_{ref})\right] \approx \sum_{i=1}^{N} p_i(\widetilde{Z},\widetilde{Z''^2}) E\left[X(Z_i,\chi_{ref})\right] \tag{1.23}$$

Ainsi, à la place des variables $\widetilde{X}(\widetilde{Z},\widetilde{Z''^2},\chi_{ref})$ qui sont placés dans la table dans un cas déterministe, les espérances des variables aléatoires $\widetilde{X}(\widetilde{Z},\widetilde{Z''^2},\chi_{ref},.)$ correspondantes sont placées. En plus de cela, les matrices de covariances des différentes quantités sont également placées dans la table. Les auteurs précisent que celles-ci sont paramétrées par $\widetilde{Z}$, $\widetilde{Z''^2}$ et χ_{ref} laissant entendre que seule la covariance de différentes grandeurs en un même point de la table sont stockées, et non la covariance de deux grandeurs prises en des points différents de la table. Ainsi, en chaque point de la table paramétré par $\widetilde{Z}$, $\widetilde{Z''^2}$ et χ_{ref} sont stockées les matrices de covariances de termes générales $\left(Cov\left[\widetilde{X}_i,\widetilde{X}_j\right]\right)_{1\leq i,j\leq N_g}$, où N_g est le nombre de grandeurs stockées. La covariance entre deux grandeurs $\widetilde{X}$ et $\widetilde{Y}$ est également obtenue grâce à la quadrature précédemment introduite, de sorte qu'elle est approchée par l'expression suivante :

$$\begin{aligned} &Cov\left[\widetilde{X}(\widetilde{Z},\widetilde{Z''^2},\chi_{ref}),\widetilde{Y}(\widetilde{Z},\widetilde{Z''^2},\chi_{ref})\right] \\ &= \sum_{i=1}^{N} p_i(\widetilde{Z},\widetilde{Z''^2}) p_j(\widetilde{Z},\widetilde{Z''^2}) Cov\left[X(Z_i,\chi_{ref}),Y(Z_j,\chi_{ref})\right] \end{aligned} \tag{1.24}$$

Les auteurs font alors remarquer que dans l'expression (1.22), la nouvelle variable aléatoire définie est une somme de variables aléatoires normales, sans rien mentionner de plus. En fait, l'hypothèse précédemment faite de considérer une loi jointe normale pour une collection de variables aléatoires de quantité d'intérêts X obtenues à l'aide du calcul de la flammelette appliquées permet de considérer le vecteur aléatoire $(X(Z_i,\chi_{ref},.))_{1\leq i\leq N}$ comme gaussien, et de fait l'approximation de la variable aléatoire $\widetilde{X}(\widetilde{Z},\widetilde{Z''^2},\chi_{ref},.)$ donnée par l'expression (1.22) est elle même gaussienne comme combinaison linéaire de composantes d'un vecteur gaussien. Cette dernière remarque peut expliquer l'intérêt sous cette hypothèse de considérer uniquement dans la table les moments d'ordre 1 et 2 des variables aléatoires tabulées, ceux-ci les caractérisant dans l'hypothèse où ils sont gaussiens.

La construction de la table a permis de propager l'incertitude de la cinétique chimique aux quantités tabulées, à savoir les variables $\widetilde{X}$, que sont la température, la densité, les fractions massiques ... A ce stade, les auteurs affirment que la dimension stochastique est encore trop importante, mais qu'avec le modèle de flammelette utilisé, seules la masse volumique, la diffusivité molé-

culaire et la viscosité influencent l'écoulement. Ainsi, seules ces trois grandeurs sont à considérer comme incertaines pour la propagation d'incertitude par méthode non intrusive à travers le calcul LES. Afin d'être en mesure de réaliser l'étude à un coût relativement faible, les auteurs simplifient celle-ci en ne considérant que la masse volumique comme étant incertaine. En considérant que la température est une variable aléatoire normale, justifié par un histogramme de T qui suit effectivement une loi proche d'une loi normale, les auteurs choisissent de considérer le volume massique v, qui est proportionnel à la température en vertu de la loi des gaz parfait, plutôt que la masse volumique, et lui donne la paramétrisation (1.25), dans laquelle ξ est une variable aléatoire normale centrée réduite.

$$\widetilde{v}(\widetilde{Z}, \widetilde{Z''^2}, \chi_{ref}) = E\left[\widetilde{v}(\widetilde{Z}, \widetilde{Z''^2}, \chi_{ref})\right] + \sqrt{Var\left[\widetilde{v}(\widetilde{Z}, \widetilde{Z''^2}, \chi_{ref})\right]}\xi \quad (1.25)$$

La justification des auteurs pour cette paramétrisation est ici insuffisante. En effet, les auteurs partent du constat que les variables aléatoires $T(Z, \chi_{ref})$, obtenues pour différentes valeurs de Z et de χ_{ref} par les calculs de flammelettes, suivent chacune une loi normale avec une bonne approximation. Cela entraîne effectivement, sous l'hypothèse que la masse molaire du mélange est constante, que les variables aléatoires $v(Z, \chi_{ref})$ correspondant au volume massique suivent également une loi normale du fait de la proportionnalité entre la température et le volume massique. Il a été explicité précédemment qu'il était possible sous certaines hypothèses faites par les auteurs que les variables aléatoires $\widetilde{v}(\widetilde{Z}, \widetilde{Z''^2}, \chi_{ref})$ suivent des lois normales comme combinaison linéaire de composantes d'un vecteur gaussien. La paramétrisation des auteurs permet bien de garantir le caractère normal de chacune des variables aléatoires $\widetilde{v}(\widetilde{Z}, \widetilde{Z''^2}, \chi_{ref})$, mais elle fait l'hypothèse très forte que l'ensemble de ces variables aléatoires sont parfaitement corrélées deux à deux, alors même qu'aucune information n'est donnée à ce sujet par les auteurs. La bonne caractérisation des incertitudes dans cette table, notamment celle concernant le volume massique, doit se faire en considérant le processus stochastique $\widetilde{v}$, indexé par $\widetilde{Z}$, $\widetilde{Z''^2}$ et χ_{ref}, dans son ensemble plutôt que sur la loi des variables aléatoires $\widetilde{v}(\widetilde{Z}, \widetilde{Z''^2}, \chi_{ref})$ prises séparément. Leur forte hypothèse permet de se ramener à une dimension stochastique de 1, alors qu'un nombre plus important de variables aléatoires est sans doute nécessaire afin de caractériser correctement le processus stochastique $\widetilde{v}$.

La suite de leur article explique la méthode non intrusive utilisée, et notamment le calcul des statistiques des grandeurs d'intérêt. Pour cela, sept simulations aux grandes échelles ont été réalisées pour sept différentes valeurs de ξ, correspondant aux points de quadrature de Gauss-Hermite (introduit au chapitre 2), la variable ξ étant normale centrée réduite. Pour chacun de ces calculs LES, la table utilisée est la table précédemment décrite dans laquelle la

masse volumique a été remplacée par l'inverse du volume massique donné par l'expression (1.25), la variable ξ étant fixée par le point de quadrature correspondant. Cela permet de calculer l'espérance de n'importe quelle grandeur X considérée comme incertaine à cause des incertitudes dans la masse volumique par l'expression (1.26), les w_i et les ξ_i correspondant respectivement aux poids et aux nœuds de la quadrature de Gauss-Hermite.

$$E[X] = \sum_{i=1}^{N} w_i X(\xi_i) \tag{1.26}$$

Les auteurs font remarquer que deux sources d'incertitudes sont en fait à prendre en compte pour les grandeurs autres que celles influençant l'écoulement, c'est à dire autres que la masse volumique dans la présente étude. La première provient de l'incertitude dûe à l'effet sur l'écoulement de la masse volumique qui est incertaine, dénommée indirecte par les auteurs, alors que la seconde provient des incertitudes inhérentes à la grandeur étudiée, qui sont présentes dans la table, et qui est dénommée directe par les auteurs. Ces considérations ont motivé les auteurs à décomposer n'importe quelle grandeur X comme la somme d'une contribution X_D, correspondant aux effets des incertitudes indirectes, et une contribution X_S, correspondant aux effets des incertitudes directes. La contribution X_D correspond simplement au résultat donné par la simulation aux grandes échelles déterministes, alors que la contribution X_S est une variable aléatoire dont la loi est la loi jointe de la grandeur X conditionnée par la valeur du volume massique. La détermination de cette loi conditionnelle est facile en suivant les hypothèses de loi normale des auteurs, l'espérance et la matrice de covariance étant connues.

Cette étude permet d'introduire la propagation d'incertitude en simulations aux grandes échelles d'écoulement réactifs par l'intermédiaire de méthodes non intrusives, en s'appuyant notamment sur une table chimique permettant d'obtenir les incertitudes sur l'ensemble des grandeurs de la simulation, en permettant de s'intéresser à la fois à la contribution directe et à la contribution indirecte. Les justifications de cette étude sont cependant lacunaires, notamment celle concernant la paramétrisation utilisant une unique variable aléatoire pour le champ de volume massique incertain de la table, qui aurait mérité l'étude du processus stochastique concerné. La justification par les auteurs de l'utilisation de lois normales dans l'ensemble de l'étude est également très succincte, et il n'est pas montré qu'un tel choix n'influence effectivement pas significativement les résultats.

1.4 Objectifs de la thèse et méthodes numériques

1.4.1 Objectifs

Cette thèse a pour objectif principal le développement de méthodes permettant de propager les incertitudes présentes dans la cinétique chimique au sein de simulations aux grandes échelles d'écoulements réactifs, et cela via l'utilisation de méthodes non-intrusives afin de pouvoir utiliser les méthodes numériques et les codes déjà existants.

Une première difficulté pour la réalisation de cet objectif provient du nombre important de paramètres de cinétique chimique dans les mécanismes réactionnels, entraînant de fait un grand nombre de paramètres incertains lorsque l'ensemble de ces paramètres sont supposés incertains. Cela conduit donc à une grande dimension stochastique, pour lesquelles les méthodes de Monte Carlo, qui sont des méthodes non-intrusives, se révèlent les principales candidates pour la propagation d'incertitude mais à un coût en temps de calcul trop important les rendant inutilisables en pratique. Cette dernière remarque conduit à la nécessité d'une réduction de la dimension stochastique, afin de n'avoir une paramétrisation des incertitudes n'impliquant qu'un petit nombre de paramètres incertains permettant d'employer des méthodes non intrusives plus efficaces que les méthodes de Monte Carlo.

En fait, les simulations aux grandes échelles d'écoulement réactifs sont généralement trop coûteuses pour être réalisées dans le cas déterministe en utilisant le mécanisme réactionnel détaillé, et un modèle réduit, dénommé modèle de flamelettes, est souvent nécessaire et utilisé pour la modélisation de la chimie. L'ensemble des informations concernant la chimie se retrouve alors contenu dans ce modèle de flamelettes, ce qui implique que les incertitudes des paramètres de cinétique chimique doivent être introduites dans ce modèle réduit, étant des informations relatives à la chimie. Ce sont donc les incertitudes présentes dans ce modèle de flamelettes qu'il convient de caractériser et de paramétrer à l'aide d'un petit nombre de paramètres incertains afin de pouvoir employer des méthodes de propagation d'incertitudes non intrusives efficaces.

La caractérisation des incertitudes du modèle de flamelettes est réalisée à l'aide d'une propagation d'incertitudes des paramètres de cinétique chimique incertains du mécanisme réactionnel détaillé dans ce modèle de flamelettes. Le grand nombre de paramètres incertains n'est ici pas prohibitif puisque les calculs nécessaires pour la construction d'un modèle de flamelettes sont tels qu'ils autorisent l'utilisation de méthodes de type Monte Carlo. Les résultats de cette méthode de Monte Carlo sont alors à analyser afin de construire une caractérisation des incertitudes du modèle de flamelettes adéquate et impliquant un petit nombre de paramètres incertains afin de rendre possible l'utilisation de méthodes de propagation d'incertitudes non intrusives efficaces pour les simulations aux grandes échelles.

Afin de pouvoir appliquer cette méthodologie, différents prérequis sont

nécessaires, qu'il a fallu identifier afin d'être en mesure de s'approprier les méthodes et outils nécessaires, puis les implémenter pour répondre au besoin. Les différents sous-objectifs identifiés sont listés ci-après, ainsi qu'un bref descriptif des outils utilisés afin de les remplir.

1.4.2 Calculer des intégrales multiples

Dans de nombreux cas, le formalisme apporté par la théorie des probabilité permet de ramener l'obtention d'informations sur les incertitudes à un calcul d'espérance mathématique, qui correspond à un calcul d'intégrale. C'est notamment le cas pour le calcul de la moyenne ou de la variance d'une grandeur incertaine, qui sont en pratique les plus utilisées. Il est également possible, par exemple, d'exprimer la probabilité qu'une grandeur incertaine G, de densité de probabilité π_G, soit supérieure à une valeur critique g_c comme une intégrale :

$$P(G \geq g_c) = E\left[\mathbb{1}_{[g_c,+\infty[}(G)\right] = \int_{\mathbf{R}} \mathbb{1}_{[g_c,+\infty[}(g)\pi_G(g)dg \tag{1.27}$$

Les intégrales à calculer ne possèdent généralement pas d'expression analytique, et il est nécessaire de recourir à des méthodes numériques permettant l'estimation efficace de la valeur de ces intégrales. Suivant la dimension de l'espace sur lequel l'intégrale est à calculer, différentes approches doivent être envisagées afin d'avoir un coût de calcul le plus faible possible. En grande dimension (chimie détaillée), des méthodes probabilistes sont les seules possibles pour des raisons de coût de calcul alors qu'en dimension plus faible (typiquement inférieure à 5), des méthodes dites déterministes par opposition à probabiliste s'avèrent bien souvent plus efficaces. Ces deux familles de méthodes sont étudiées dans les chapitres 2 et 3.

1.4.3 Paramétrer les incertitudes du système chimique canonique à l'aide d'un jeu restreint de paramètres incertains

La caractérisation des incertitudes du modèle de flamelettes nécessaire à la propagation efficace des incertitudes de cinétique chimique au sein d'une simulation aux grandes échelles d'un écoulement réactif doit être faite à l'aide d'un nombre restreint de paramètres incertains. Cet aspect revient à réduire la dimension stochastique du système, qui est généralement un processus s'accompagnant d'une perte d'information concernant les incertitudes du système. Ce processus de réduction de la dimension stochastique est en fait un compromis qu'il faut trouver entre le coût de calcul, directement lié à la dimension stochastique retenue, et la quantité d'information retenue. Ces méthodes de réduction de la dimension stochastique sont présentées dans le chapitre 4.

1.4.4 Modéliser la loi de probabilité du jeu restreint de paramètres incertains

Certaines méthodes pour la réduction du nombre de paramètres incertains permettent la génération de nouveaux paramètres incertains, pour lesquelles des dépendances entre eux peuvent exister. La prise en compte de ces dépendances est parfois indispensable à la bonne reproduction des incertitudes du système considéré. Il est donc nécessaire de modéliser adéquatement la loi de probabilité du vecteur de paramètres incertains, afin de prendre en compte les dépendances entre les nouveaux paramètres incertains. Cet aspect est étudié dans le chapitre 5.

1.5 Organisation du manuscrit

Le manuscrit est organisé en deux parties. La première partie est focalisée sur la présentation des outils numériques et leur implémentation ainsi que des applications de certaines de ces méthodes dans d'autres domaines que la propagation d'incertitude.

La seconde partie se concentre sur l'utilisation de ces méthodes à un système chimique canonique "simple" afin de construire un modèle de chimie tabulée incertaine encapsulant les incertitudes de la cinétique chimique.

Partie I :

Le chapitre 2 est consacré aux méthodes de cubatures, permettant le calcul d'intégrales multiples en dimension faible (typiquement inférieure à 5). Des méthodes de quadratures classiques sont présentées et comparées, permettant des calculs d'intégrales efficaces pour des fonctions d'une seule variable. L'obtention de méthodes en dimension supérieure à 2 passe par la tensorisation de ces méthodes de quadratures. La comparaison de tensorisation pleine et creuse de Smolyak est faite, en terme d'erreur et de coût de calcul, ainsi qu'avec une implémentation d'un algorithme adaptatif utilisant une tensorisation creuse.

Le chapitre 3 est consacré aux méthodes de Monte Carlo, de Quasi-Monte Carlo et de Quasi-Monte Carlo randomisée. Ces méthodes permettent toutes le calcul d'intégrales multiples, préférablement en grande dimension (typiquement supérieure à 5). Une implémentation parallèle de ces méthodes a été introduite au cours de cette thèse, permettant entre autre la propagation d'incertitude au sein de systèmes chimiques "simples" tels les modèles de flamelettes, par l'obtention de statistiques permettant la caractérisation des incertitudes de ces systèmes en bénéficiant de la puissance de calcul de supercalculateur. Les taux de convergence théoriques sont présentés, et les différentes méthodes sont comparées sur des fonctions tests. Une utilisation de la méthode de Quasi-Monte Carlo randomisée pour la résolution de l'équation de transfert radiatif dans des

cas complexes est également a également fait l'objet d'une communication au congrès ASME et est présentée en annexe B.

Le chapitre 4 est consacré à des méthodes spectrales pour la représentation de variables aléatoires et de processus stochastiques. Une première partie présente ce qui est appelé dans la littérature les expansions en polynômes du chaos (PCE), pouvant être utilisées à la fois pour la représentation de variables aléatoires et de processus stochastiques. L'obtention de telles expansions est présentée, ainsi que leur usage pour l'analyse de sensibilité globale pouvant permettre la réduction du nombre de paramètres incertains. Une seconde partie est consacrée à l'expansion de Karhunen-Loève d'un processus stochastique ainsi qu'à son obtention par la méthode numérique de Nyström. Cette expansion permet d'obtenir une représentation spectrale d'un processus stochastique possédant la caractéristique de concentrer la plus grande partie de la variance de celui-ci dans les premiers termes de l'expansion, permettant une réduction de la dimension stochastique en tronquant cette expansion. L'application de PCE pour le calcul de surface de réponse de la vitesse axiale en fonction du diamètre de goutte au sein d'une simulation monodisperse d'un écoulement diphasique a fait l'objet d'une communication au congrès ICMF qui est présenté en annexe C.

Le chapitre 5 est consacré à l'estimation des densités de probabilités de vecteurs aléatoires de taille raisonnable (typiquement moins de 5 composantes) à l'aide de méthodes à noyaux. Ces méthodes sont dites non paramétriques, permettent l'estimation de la densité de probabilité d'un vecteur aléatoire à l'aide d'échantillons de celui-ci, sans présupposer une loi de probabilité pour ce vecteur aléatoire comme dans le cas de méthodes paramétriques.

Partie II :

Le chapitre 6 est consacré à la caractérisation de l'incertitude d'un système chimique canonique "simple", en l'occurrence un réacteur homogène adiabatique à pression constante d'un mélange d'air et d'hydrogène. Un tel système chimique canonique peut être utilisé pour la simulation de flammes pour lesquelles le phénomène d'auto-allumage est critique pour décrire leur comportement. Il est montré dans ce chapitre qu'il est possible d'introduire les incertitudes dans le système chimique canonique étudié quasiment uniquement à travers la variable d'avancement de la réaction classiquement utilisée en chimie tabulée.

Le chapitre 7 est consacré à la modélisation du terme source incertain de la variable d'avancement de la réaction, par lequel l'ensemble des incertitudes seront introduites, toujours dans le cas d'un réacteur homogène adiabatique à pression constante d'un mélange d'air et d'hydrogène. La modélisation du terme source incertain est faite dans ce chapitre à l'aide d'une expansion en

polynômes du chaos faisant intervenir un nombre restreint des paramètres incertains initiaux, la sélection des paramètres incertains gardés ayant été réalisée grâce à une analyse de sensibilité globale. La validation de l'utilisation de ce terme source modélisé est faite en s'assurant que les incertitudes sur la variable d'avancement de la réaction sont correctement reproduites.

Le chapitre 8 est consacré à la modélisation du terme source incertain de la variable d'avancement de la réaction, par lequel l'ensemble des incertitudes seront introduites, toujours dans le cas d'un réacteur homogène adiabatique à pression constante d'un mélange d'air et d'hydrogène, à l'aide de nouvelles variables aléatoires. Ces nouvelles variables aléatoires sont issues de l'expansion de Karhunen-Loève du terme source directement mais également de la variable d'avancement. Le terme source incertain est modélisé à l'aide de son expansion de Karhunen-Loève dans le cas où les nouvelles variables introduites sont issues de cette expansion, et à l'aide d'une expansion en polynôme du chaos lorsque les nouvelles variables aléatoires sont issues de l'expansion de Karhunen-Loève de la variable d'avancement. Dans l'ensemble des cas, la validation de l'utilisation du terme source incertain modélisé est faite en s'assurant que les incertitudes sur la variable d'avancement de la réaction sont correctement reproduites.

Première partie

Outils numériques pour la propagation d'incertitudes

Chapitre 2

Méthodes déterministes pour l'intégration numérique

Notions clés et apports du chapitre :

- Méthodes de quadrature et leurs convergences
- Comparaison des méthodes de quadrature
- Fléau de la dimension pour le calcul d'intégrales multiples
- Tensorisation pleine de méthodes de quadrature pour les intégrales en faible dimension ($d \lesssim 5$)
- Tensorisation creuse de Smolyak et algorithme adaptatif
- Comparaison des différentes tensorisations

Il a été mis en avant lors du chapitre introductif, que les informations statistiques sont obtenues par le calcul d'intégrale, la théorie des probabilités étant basées sur la théorie de la mesure. Le calcul d'intégrale ne peut que rarement se faire de manière analytique, et des méthodes de calcul numérique d'intégrale sont nécessaires. Suivant le nombre de variables dont dépend la fonction à intégrer, qui correspond au nombre de paramètres incertains du système, des approches différentes doivent être envisagées pour le calcul numérique de l'intégrale, dû à ce qui est communément appelé le fléau de la dimension. Dans ce chapitre vont être présentées des méthodes d'intégration numérique déterministes, par opposition à celles probabilistes, pour des fonctions dépendant d'un nombre de variables typiquement inférieur à cinq, pour lesquels les méthodes présentées sont, sous certaines conditions de régularités de la fonction à intégrer, plus efficaces en pratique que les méthodes qui seront présentées dans le chapitre suivant, consacré aux méthodes d'intégration numérique probabilistes plus efficaces pour des fonctions dépendant d'un grand nombre de variables, typiquement supérieur à cinq.

2.1 Intégration numérique de fonctions d'une variable

L'objectif de cette section est de détailler l'évaluation numérique de l'espérance ou moyenne $E_w[g]$, que l'on suppose exister, d'une fonction g d'une seule variable sur un intervalle I de $\mathbb{R}$ pondéré par une fonction positive et d'intégrale sur I finie w, qui est typiquement une densité de probabilité. Cette valeur moyenne de la fonction g pondérée par w peut s'exprimer comme une intégrale comme dans l'expression (1.3), de sorte que l'on a :

$$E_w[g] = \int_I g(t)w(t)dt \tag{2.1}$$

Il est possible de modifier l'expression précédente pour se ramener à un calcul d'intégrale sur le segment $[0,1]$, en effectuant un changement de variable :

$$\int_I g(t)w(t)dt = \int_{[0,W_\infty]} g(W^{-1}(v))dv = W_\infty \int_{[0,1]} g(W^{-1}(W_\infty u))du \tag{2.2}$$

Dans l'expression précédente, W_∞ est la limite en la borne supérieure de I (pouvant être $= \infty$) de la fonction W, qui est donnée par l'expression suivante :

$$\forall t \in I, W(t) = \int_{I\cap]-\infty,t]} w(t')dt' \tag{2.3}$$

Dans le cas où w est une densité de probabilité, W correspond à la fonction de répartition d'une variable aléatoire de densité de probabilité w. L'expression

(2.2) permet de déduire qu'il est possible de passer de l'intégrale sur un intervalle I pondéré par une fonction w_I à une intégrale sur un intervalle J par une fonction w_J, à l'aide d'un changement de variable.

L'évaluation numérique de l'intégrale d'une fonction d'une variable pondérée par une fonction de pondération est réalisé en utilisant l'évaluation de la fonction en certains points, et peut également impliquer l'évaluation de dérivées de la fonction en certains points, qui ne sont pas nécessairement les mêmes que ceux où la fonction a été évaluée. Dans les travaux présentés dans cette thèse, les dérivés des fonctions dont on souhaite évaluer l'intégrale ne sont pas accessibles. Il est possible d'approximer numériquement les dérivés à l'aide de différences finies, mais cela revient à évaluer la fonction en de nouveaux points. Pour cette raison, les seules méthodes d'évaluation numériques d'intégrales d'une fonction d'une variable pondérées par une fonction de pondération présentées dans cette section sont celles se servant de la valeur de la fonction en certains points, et qui approxime l'intégrale à l'aide d'une somme pondérée de ces valeurs (2.4).

$$\int_I g(t)dt \approx \sum_{i=1}^{N} w_i g(x_i) \tag{2.4}$$

Une formule de la forme de celle présentée en (2.4) s'appelle une formule de quadrature. Chaque formule de quadrature implique un ensemble de N poids $\{w_i : 1 \leq i \leq N(\}$ et N points $\{x_i : 1 \leq i \leq N\}$ de I. Ces poids et ces points d'évaluations dépendent de la fonction de pondération et sont indépendants de la fonction g à intégrer. Dans les cas d'applications envisagés dans la présente thèse, seul le coût d'évaluation de la fonction est important, et il est donc primordial de maintenir le nombre de points N présents dans la formule de quadrature le plus petit possible en s'assurant que l'évaluation de l'intégrale est suffisamment précise pour le cas d'application étudié. Il est pour cela primordial de choisir des formules de quadratures qui sont les meilleures possibles, c'est à dire qui sont les plus précises possibles avec un nombre N d'évaluations le plus petit possible.

2.1.1 Méthodes de quadratures simples

Les méthodes présentées dans cette section correspondent au cas d'un intervalle $I = (a, b)$ (les parenthèses signifiant que les bords de l'intervalle peuvent être indifféremment fermés ou ouverts) pour lequel la fonction de pondération est constante et égale à 1 sur l'ensemble de l'intervalle. Cela revient à considérer l'intégrale d'une fonction g entre a et b en dimension 1, qui peut être interprétée géométriquement comme l'aire comprise entre la courbe et l'axe des abscisses, comptée positivement si la fonction est positive, et négativement si celle-ci est négative, et entre les droites d'équations $x = a$ et $x = b$, comme visible sur la figure 2.1.

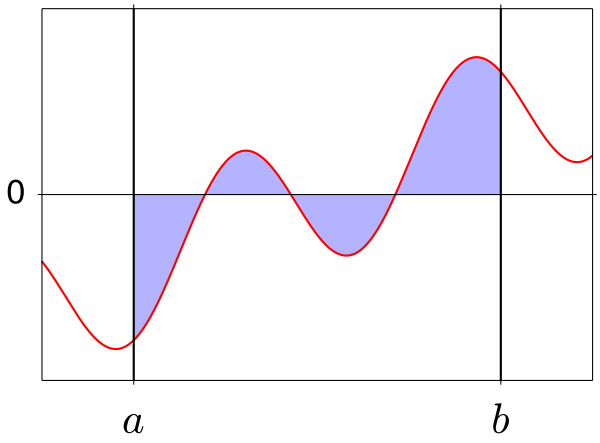

FIGURE 2.1 – *Illustration géométrique de l'intégrale d'une fonction en dimension* 1

2.1.1.1 Méthode du point médian

Une première façon d'approximer l'intégrale d'une fonction g entre a et b, est de considérer un rectangle de largeur $(b-a)$, et de hauteur la valeur de la fonction au point $\frac{a+b}{2}$, ce qui donne l'approximation (2.5) pour l'intégrale de g.

$$\int_a^b g(t)dt \approx (b-a)\, g\left(\frac{a+b}{2}\right) \tag{2.5}$$

Une illustration géométrique de cette première approximation est visible sur la figure 2.2.

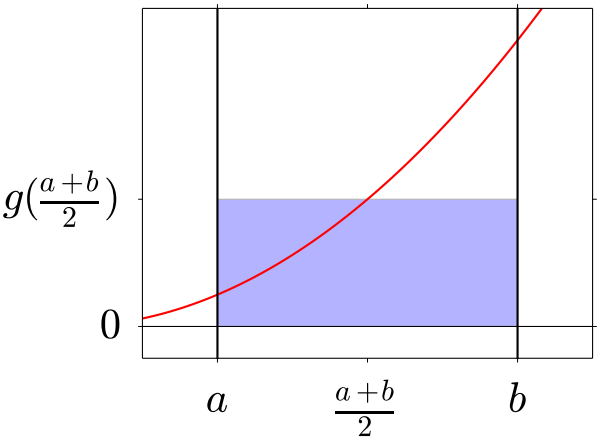

FIGURE 2.2 – *Approximation de l'intégrale d'une fonction g entre a et b par l'aire d'un rectangle*

Pour une fonction suffisamment régulière, au moins C^1, cette approximation sera d'autant plus exacte que a et b seront proches. Une idée est donc de découper l'intervalle $[a, b]$ en N sous intervalles de taille $\Delta = \frac{b-a}{N}$, en définissant les N points suivant, correspondant aux milieux de chacun des sous intervalles :

$$x_i = a + \frac{(2i+1)\Delta}{2}, i \in [\![1, N]\!] \tag{2.6}$$

La relation de Chasles pour l'intégrale permet alors d'écrire :

$$\int_a^b g(t)dt = \sum_{i=1}^{N} \int_{x_i - \frac{\Delta}{2}}^{x_i + \frac{\Delta}{2}} g(t)dt \tag{2.7}$$

En appliquant l'approximation définie dans (2.5) sur chacun des segments $\left[x_i - \frac{\Delta}{2}, x_i + \frac{\Delta}{2}\right]$, on obtient finalement comme approximation de l'intégrale de g entre a et b :

$$\int_a^b g(t)dt \approx \sum_{i=1}^{N} \Delta g\left(x_i\right) \tag{2.8}$$

Géométriquement, cela revient à approximer l'intégrale par l'aire de N rectangles de largeur Δ et de hauteur $g(x_i)$, comme présenté sur la figure 2.3.

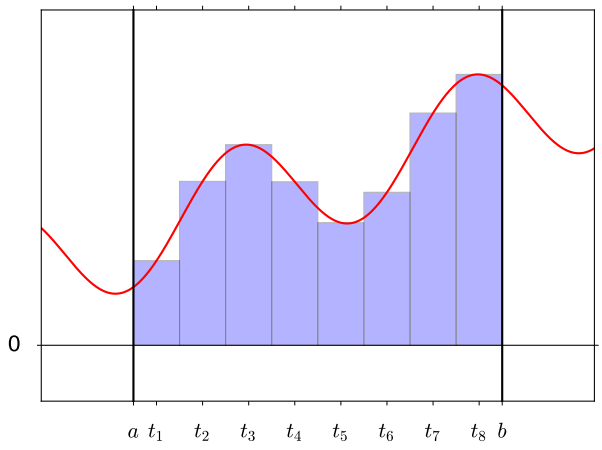

FIGURE 2.3 – *Approximation de l'intégrale d'une fonction par les aires de rectangles dont la hauteur est définie comme la valeur de la fonction au milieu de leur base.*

2.1.1.2 Méthodes des trapèzes

Une autre façon d'approximer l'intégrale de g entre a et b est d'approximer l'intégrale par l'aire du trapèze rectangle de hauteur $(b-a)$ et de bases de longueurs $g(a)$ et $g(b)$, comme présenté sur la figure 2.4. Cette méthode approxime l'intégrale par (2.9).

$$\int_a^b g(t)dt \approx (b-a)\frac{g(a)+g(b)}{2} \tag{2.9}$$

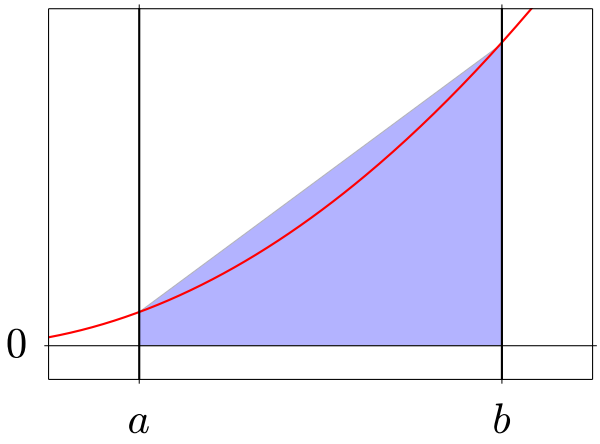

FIGURE 2.4 – *Approximation de l'intégrale d'une fonction g entre a et b par l'aire d'un trapèze*

Encore une fois, intuitivement l'approximation sera d'autant meilleure que les points a et b seront proches, et découper l'intervalle $[a,b]$ en N sous intervalles de taille $\Delta = \frac{b-a}{N}$ permet d'avoir une nouvelle approximation de l'intégrale de g entre a et b. Pour la méthode des trapèzes, les extrémités de chaque sous segment $\{x_i : 0 \leq i \leq N\}$ sont à considérer, avec $x_i = a + i\Delta$, et en appliquant l'approximation donnée par (2.9) sur chaque sous segment, on obtient alors la nouvelle approximation (2.10).

$$\int_a^b g(t)dt \approx \sum_{i=1}^{N}\int_{x_{i-1}}^{x_i} g(t)dt \approx \frac{\Delta}{2}g(x_0) + \sum_{i=1}^{N-1}\Delta g(x_i) + \frac{\Delta}{2}g(x_N) \tag{2.10}$$

Géométriquement, cela revient à approximer l'intégrale par l'aire de N trapèzes rectangles de hauteur Δ et de bases $g(x_i)$ et $g(x_{i+1})$, comme présenté sur la figure 2.5.

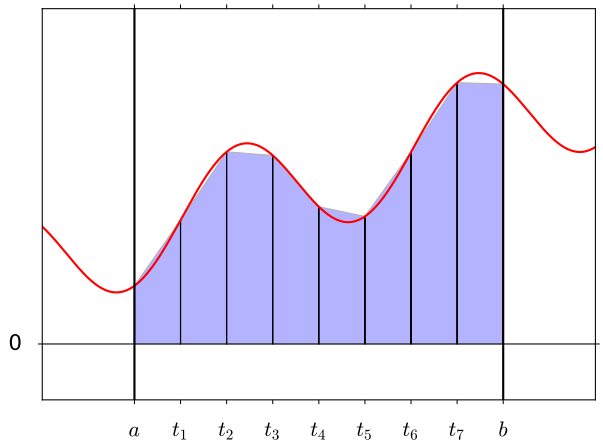

FIGURE 2.5 – *Approximation de l'intégrale d'une fonction par les aires de trapèzes rectangles.*

2.1.1.3 Formules de Newton-Cotes

Les méthodes du point médian et des trapèzes consistent à approximer la fonction sur chaque sous intervalles par des polynômes interpolants en des points remarquables, c'est à dire des polynômes qui ont la même valeur que la fonction en ces points remarquables, et à remplacer la valeur de l'intégrale de la fonction par la valeur de l'intégrale de ces polynômes. Dans le cadre de la méthode du point médian, les polynômes utilisés sont constants, donc de degré 0, et interpolant au milieu du segment. Pour la méthode des trapèzes, les polynômes utilisés sont de degré 1, et sont donc des fonctions affines qui interpolent la fonction au niveau des deux extrémités du sous intervalle.

Il est possible de continuer sur cette idée, en considérant cette fois des polynômes interpolants avec un degré p plus important. En considérant l'intégrale sur le segment $[a, b]$ de la fonction g, l'idée est de découper l'intervalle $[a, b]$ en p sous segments, et à interpoler la fonction g en chacune des extrémités des sous-segments à l'aide du polynôme interpolateur en ces points. Les polynômes interpolants considérés peuvent approcher correctement la fonction à intégrer sur l'intervalle $[a, b]$, de sorte qu'avec l'augmentation du degré p, les graphes de la fonction g et du polynôme interpolant se rapprochent, comme présenté sur la figure 2.6.

Comme les p sous segments définissent $p+1$ extrémités, le polynôme interpolateur sera bien de degré p. La construction d'un tel polynôme interpolateur peut se faire à l'aide des polynômes interpolateurs de Lagrange. Les polynômes interpolateurs de Lagrange basés sur l'ensemble de points $\{x_i : 0 \leq i \leq p\}$ sont

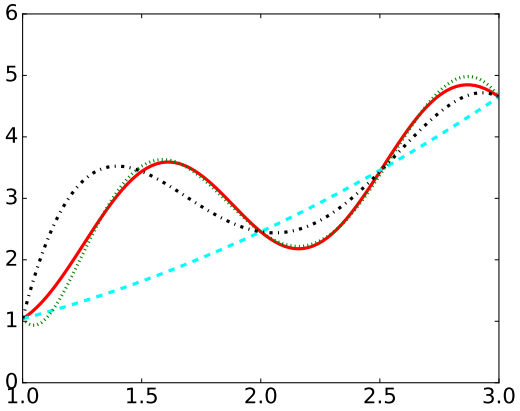

FIGURE 2.6 – *Fonction $g(x) = 1 + x + sin(5x)$ (ligne pleine) sur le segment $[1, 3]$ et polynômes interpolants utilisés pour les formules de Newton-Cotes, de degré 2 (tirets), degré 4 (tirets-points) et degré 6 (pointillés).*

les polynômes :

$$l_i(x) = \prod_{j=0, j\neq i}^{d} \frac{x - x_j}{x_i - x_j}, i \in [\![0, p]\!] \tag{2.11}$$

Ces polynômes possèdent la propriété que le $i - me$ polynômes est nul sur l'ensemble des points x_j avec j différent de i, et vaut 1 en x_i. Il est alors aisé de définir le polynôme interpolateur $L_g^{(p)}$ d'une fonctions g aux points $\{x_i : 0 \leq i \leq p\}$ par :

$$L_g(x) = \sum_{i=0}^{d} g(x_i) l_i(x) \tag{2.12}$$

Comme précédemment pour la méthode du point médian ou celle des trapèzes, l'intégrale de la fonction g est approximée par l'intégrale de l'interpolation polynomiale $L_g^{(p)}$, de sorte que :

$$\int_a^b g(t)dt \approx \int_a^b L_g^{(p)}(t)dt = \sum_{i=0}^{p} g(x_i) \int_a^b l_i(t)dt \tag{2.13}$$

Cette dernière expression peut également s'écrire :

$$\int_a^b g(t)dt \approx \sum_{i=0}^{p} g(x_i)w_i \tag{2.14}$$

Avec :

$$w_i = \int_a^b l_i(t)dt, i \in [\![0, p]\!] \tag{2.15}$$

Les poids w_i sont donc indépendants de la fonction g à intégrer, et ne dépendent à priori que du degré p choisi, et des bornes a et b. Il est en fait facile de montrer à l'aide d'un changement de variable affine en se reportant au segment $[0, 1]$ que les poids w_i ne dépendent que de la taille de l'intervalle $[a, b]$. Cela mène à l'expression (2.16) pour les poids w_i, impliquant les polynômes interpolateurs de Lagrange construits en découpant le segment $[0, 1]$ en p parties égales.

$$w_i = \int_a^b l_i(t)dt = (b-a)\int_0^1 l_i^{[0,1]}(t)dt = (b-a)w_i^{[0,1]} \tag{2.16}$$

Pour les méthodes du point médian ou des trapèzes, l'intervalle initial était subdivisé en N sous intervalles sur chacun desquels la méthode en question était appliquée. Il est clair que cela permet d'approcher mieux la fonction sur l'ensemble de l'intervalle d'intégration, les polynômes utilisés dans ces méthodes n'étant que des fonctions constantes ou affines. En augmentant le degré p des polynômes utilisés, on pourrait espérer approcher suffisamment bien la fonction à l'aide du polynôme interpolateur à mesure que l'on augmente le degré p de celui-ci, pour ne pas avoir à découper l'intervalle en N sous intervalles sur lesquels on appliquerait une formule de Newton-Cotes impliquant un polynôme de degré p. Il est en fait dangereux d'augmenter le degré p, car le polynôme interpolateur peut alors se mettre à osciller fortement et à s'éloigner de la fonction à mesure que l'on en augmente le degré p. Ce phénomène connu est appelé phénomène de Runge [30], et apparaît pour certaines fonctions, comme notamment la fonction de Runge présentée sur la figure 2.7 avec des polynômes interpolateurs de différents degrés. La norme infini de la différence de la fonction de Runge et du polynôme interpolateur utilisé pour les formules de Newton-Cotes sur $[-1, 1]$ augmente avec le degré p, et tend vers l'infini quand p tend vers l'infini.

Un moyen d'éviter ce phénomène, est de ne pas considérer des degrés p trop importants, en découpant l'intervalle d'intégration en N sous intervalles de tailles égales, et d'appliquer sur chacun de ces sous intervalles une méthode de

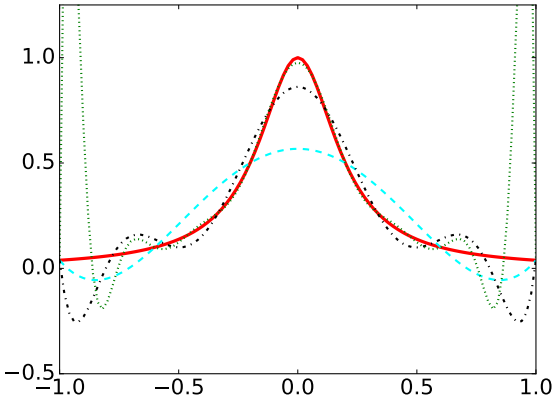

FIGURE 2.7 – *Fonction de Runge* $g(x) = 1/(1+25x^2)$ *(ligne pleine) sur le segment* $[-1,1]$ *et polynômes interpolants utilisés pour les formules de Newton-Cotes, de degré 5 (tirets), degré 9 (tirets-points) et de degré 15 (pointillés).*

Newton-Cotes impliquant un polynôme interpolateur de degré p fixé. On parle alors de méthode composite de degré p. Pour faire ainsi, le segment d'intégration $[a,b]$ est découpé en N sous segments de tailles égales de la formes $[x_{i-1}, x_i]$ pour $i \in [\![1, N]\!]$, et à l'intérieur de chacun de ces sous segments $[x_{i-1}, x_i]$ on considère $p+1$ points $x_{i,j}$ avec $j \in [\![0, p]\!]$ répartis uniformément dans $[x_{i-1}, x_i]$. Il est en fait possible de considérer ou non les extrémités des sous segments $[x_{i-1}, x_i]$ dans ces $p+1$ points, conduisant à deux types de méthodes de Newton-Cotes :

- les méthodes de Newton-Cotes fermées, dans lesquelles les points $x_{i,0}$ et $x_{i,p}$ sont respectivement égaux à x_{i-1} et x_i, les autres points étant répartis dans le segment $[x_{i-1}, x_i]$, le découpant en p segments de mêmes longueurs.
- les méthodes de Newton-Cotes ouvertes, dans lesquelles l'ensemble des points $x_{i,j}$ sont situés à l'intérieur du segment $[x_{i-1}, x_i]$. Les point $x_{i,0}$ et $x_{i,p}$ sont alors situés à une distance $\delta/2$ de x_{i-1} et x_i respectivement, et l'écart entre deux points $x_{i,j}$ consécutifs est de δ avec $\delta = (x_i - x_{i-1})/(p+1)$.

Il est en fait également possible de créer des méthodes ouvertes d'un côté et fermées de l'autre, mais celles-ci ne seront pas présentées dans cette thèse. En appliquant la formule de Newton-Cotes sur chaque segment $[x_{i-1}, x_i]$ de taille Δ et en utilisant la relation de Chasles pour l'intégrale, on obtient alors l'approximation (2.17) pour l'intégrale de g sur $[a,b]$.

$$\int_a^b g(t)dt \approx \sum_{i=1}^{n} \Delta \sum_{j=0}^{p} g(x_{i,j}) w_j^{[0,1]} \tag{2.17}$$

Un exemple de méthode composite de Newton-Cotes fermée bien connue est la méthode de Simpson, où un degré $p = 2$ est utilisé, et dont l'expression est la suivante :

$$\begin{aligned}\int_a^b g(t)dt \approx & \frac{b-a}{n}\sum_{i=1}^{n} g\left(a+(i-1)\frac{b-a}{n}\right) \\ & + 4\frac{b-a}{n}\sum_{i=1}^{n} g\left(a+(2i-1)\frac{b-a}{2n}\right) \\ & + \frac{b-a}{n}\sum_{i=1}^{n} g\left(a+i\frac{b-a}{n}\right)\end{aligned} \tag{2.18}$$

L'expression (2.17) montre qu'il est possible de jouer à la fois sur le nombre de découpage N du segment d'intégration, et sur le degré p des polynômes interpolateurs pour approcher l'intégrale de g. Un compromis doit être trouvé afin d'assurer une bonne convergence de l'approximation de l'intégrale, en utilisant un minimum de points. Pour pouvoir décider d'un tel compromis, des considérations théoriques concernant la convergence des formules de quadrature sont très utiles, et sont l'objet de la prochaine partie.

2.1.2 Convergence des formules de Newton Cotes

Une question importante concernant une formule de quadrature, est de savoir si celle-ci est convergente, c'est à dire si l'approximation de la valeur de l'intégrale devient meilleure à mesure que l'on augmente le nombre de points d'évaluations utilisés. Pour la méthode du point médian par exemple, la quadrature obtenue est une somme de Riemann qui convergera bien vers la valeur de l'intégrale de la fonction, à condition que celle-ci soit Riemann intégrable, ce qui est le cas pour les fonctions continues par morceaux par exemple. Une information plus intéressante est la vitesse de convergence de la méthode, c'est à dire le nombre de points d'évaluations N nécessaire pour obtenir une précision ϵ donnée. Pour cela, on s'intéresse à l'erreur $E(g)$ définie comme la différence de la valeur de l'intégrale et la valeur de l'approximation (2.19), écrite ici pour une quadrature quelconque utilisant N points d'évaluations.

$$E(g) = \int_a^b g(t)dt - \sum_{i=1}^{N} w_i g(x_i) \tag{2.19}$$

Concernant la méthode du point médian sur un sous segment $[x_{i-1}, x_i]$ de longueur Δ, en supposant que la fonction g est deux fois continûment dérivable, on peut appliquer la formule de Taylor-Lagrange en $m_i = \frac{x_{i-1}+x_i}{2}$ de g, qui

donne :

$$g(m_i + h) = g(m_i) + g'(m_i)h + g''(\xi)\frac{h^2}{2} \tag{2.20}$$

Dans la précédente équation, le point ξ appartient au segment $[m_i, m_i+h]$, et g'' étant supposée continue sur $[x_{i-1}, x_i]$, elle est majorée par sa norme infinie $||g''||_\infty$, ce qui permet d'écrire :

$$\left|g(m_i + h) - g(m_i) - g'(m_i)h\right| \leq ||g''||_\infty \frac{h^2}{2} \tag{2.21}$$

De cette dernière équation, on peut en déduire par intégration sur $[x_{i-1}, x_i]$ que :

$$\int_{x_{i-1}}^{x_i} \left[g(t) - g(m_i) - g'(m_i)(t - m_i)\right] dt = \int_{x_{i-1}}^{x_i} g(t)dt - g(m_i) \tag{2.22}$$

On peut également obtenir l'inégalité suivante à partir de (2.21) :

$$\begin{aligned}\left|\int_{x_{i-1}}^{x_i} [g(t) - g(m_i) - g'(m_i)(t - m_i)]dt\right| &\leq \int_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} \left|g(m_i + h) - g(m_i) - g'(m_i)h\right| dh \\ &\leq \frac{||g''||_\infty}{2} \int_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} h^2 dh \\ &= \frac{\Delta^3}{24}||g''||_\infty \end{aligned} \tag{2.23}$$

Une majoration de l'erreur d'intégration de la méthode du point médian peut être obtenue à l'aide de (2.22) et (2.23) sur le segment $[a, b]$ coupé en N sous segments de taille $\Delta = \frac{(b-a)}{N}$ chacun :

$$
\begin{aligned}
|E(g)| &= \left| \int_a^b g(t)dt - \sum_{i=1}^{N} g(m_i) \right| \\
&= \left| \sum_{i=1}^{N} \left(\int_{x_{i-1}}^{x_i} g(t)dt - g(m_i) \right) \right| \\
&\leq \sum_{i=1}^{N} \left| \int_{x_{i-1}}^{x_i} g(t)dt - g(m_i) \right| \\
&\leq \sum_{i=1}^{N} \frac{\Delta^3}{24} ||g''||_\infty = \frac{(b-a)^3}{24N^2} ||g''||_\infty \\
&= O\left(\frac{1}{N^2}\right) = O\left(\Delta^2\right)
\end{aligned} \tag{2.24}
$$

De cette dernière expression, on trouve que l'erreur de la méthode du point médian commise est inversement proportionnelle au carré du nombre N de points d'évaluation utilisés dans la méthode du point médian, ou de la même manière qu'elle est de l'ordre du carré du pas de discrétisation Δ utilisé. De plus, le terme de majoration fait intervenir la norme infinie de la dérivé seconde de la fonction à intégrer.

Des majorations similaires peuvent être obtenues d'une façon similaire pour les autres formules composites de Newton-Cotes de degré p et impliquant N points d'évaluation [105]. La majoration dépend de la parité du degré p utilisé. En supposant que la fonction g est C^{p+2} sur le segment d'intégration $[a, b]$, alors on a si p est pair :

$$
|E(g)| \leq M_p \frac{(b-a)^{p+3}}{N^{p+2}} ||g^{(p+2)}||_\infty \tag{2.25}
$$

Et si p est impair, la majoration devient :

$$
|E(g)| \leq M_p \frac{(b-a)^{p+2}}{N^{p+1}} ||g^{(p+1)}||_\infty \tag{2.26}
$$

Dans le cas de la méthode pour laquelle $p = 1$, l'erreur est majorée par l'expression suivante :

$$
|E(g)| \leq \frac{1}{12} \frac{(b-a)^3}{N^2} ||g^{(2)}||_\infty \tag{2.27}
$$

Dans le cas de la méthode de Simpson composite pour laquelle $p = 2$,

l'erreur est majorée par l'expression suivante :

$$|E(g)| \leq \frac{1}{180}\frac{(b-a)^5}{N^4}||g^{(4)}||_\infty \tag{2.28}$$

Pour les deux majorations dans le cas général, M_p ne dépend que du degré p utilisé dans la formule composite. La norme infini de la dérivé d'ordre $p+1$ ou $p+2$ apparaît également dans cette majoration. Des fonctions présentant des dérivées d'ordre élevé qui sont de plus en plus grande peuvent alors présenter le phénomène de Runge, les majorations de l'erreur précédentes ne tendant pas vers 0 avec le degré p qui augmente.

Augmenter le degré p des formules de Newton-Cotes peut donc permettre de réduire l'erreur d'intégration pour des fonctions se comportant suffisamment bien, mais sans plus d'informations sur la fonction à intégrer, il est impossible de prévoir si le phénomène de Runge apparaîtra et quel sera le degré p à utiliser dans une méthode composite qui convergera le plus rapidement possible. L'utilisation des méthodes de Newton-Cotes composite se fait donc rarement avec des grandes valeurs de p, typiquement plus faibles que 5, pour lesquelles les poids w_i sont positifs. Il existe cependant des alternatives pour obtenir des méthodes présentant une convergence plus rapide, et ces méthodes de quadrature peuvent être obtenues à partir de méthodes de Newton-Cotes composites de degré p faible. La méthode de Romberg est une de ces méthodes et se base sur la méthode des trapèzes.

2.1.3 Accélération de la convergence

La méthode de Romberg [111] est basée sur la formule d'intégration d'Euler-Maclaurin [110], qui permet d'obtenir un développement limité de l'erreur commise en fonction de la longueur Δ des sous segments utilisés dans la méthode des trapèzes, et sur l'application du procédé d'extrapolation de Richardson [108]. La formule d'intégration d'Euler-Maclaurin pour une fonction g qui est $p+2$ fois continûment dérivable et pour laquelle on souhaite approximer l'intégrale sur $[a,b]$ à l'aide d'une méthode des trapèzes à $N+1$ points stipule l'existence d'un point ξ de $[a,b]$ tel que :

$$\begin{aligned}\int_a^b g(t)dt &= \Delta\frac{g(a)+g(b)}{2} + \Delta\sum_{k=1}^{N-1} g(a+k\Delta) \\ &- \sum_{k=1}^{p} \Delta^{2k}\frac{b_{2k}}{(2k)!}\left(g^{(2k-1)}(b) - g^{(2k-1)}(a)\right) \\ &- (b-a)\Delta^{2p+2}\frac{b_{2p+2}}{(2p+2)!}g^{(2p+2)}(\xi)\end{aligned} \tag{2.29}$$

Les nombres b_k dans cette dernière formule sont les nombres de Bernouilli. En notant $T^{(N)}(g)$ la valeur obtenue à partir de la méthode des trapèzes à $N+1$

points, on trouve à partir de 2.29 que :

$$T^{(N)}(g) = \int_a^b g(t)dt + \sum_{k=1}^{p} c_k \Delta^{2k} + O(\Delta^{2p+2}) \tag{2.30}$$

Dans la précédente expression, les coefficients c_k ne dépendent pas du nombre N d'évaluations utilisées dans la méthode des trapèzes, et sont donnés par l'expression suivante :

$$c_k = \frac{b_{2k}}{(2k)!}\left(g^{(2k-1)}(a) - g^{(2k-1)}(b)\right) \tag{2.31}$$

Si le coefficient c_1 est non nul, on retrouve bien le résultat donné dans la section précédente sur la convergence de la méthode des trapèzes, qui donne une erreur en $O(\Delta^2) = O(N^{-2})$, et une erreur plus faible dans le cas où le coefficient c_1 est nul. La rapidité de la convergence est en fait donnée par le plus petit des c_k qui est non nul dans le développement précédent. L'idée de la méthode de Romberg est d'annuler les c_k afin d'obtenir une méthode avec une convergence plus rapide. En considérant une méthode des trapèzes avec une longueur de sous segment moitié plus petite, on trouve en appliquant la formule d'Euler Maclaurin :

$$\begin{aligned} T^{(2N)}(g) &= \int_a^b g(t)dt + \sum_{k=1}^{p} c_k \left(\frac{\Delta}{2}\right)^{2k} + O\left(\left(\frac{\Delta}{2}\right)^{2p+2}\right) \\ &= \int_a^b g(t)dt + \sum_{k=1}^{p} \frac{c_k}{4^k}\Delta^{2k} + O(\Delta^{2p+2}) \end{aligned} \tag{2.32}$$

En combinant (2.30) et (2.32) correctement, il est possible d'éliminer le terme en Δ^2. Cela donne une nouvelle méthode de quadrature R en $O(\Delta^4)$, dont le développement est le suivant :

$$\begin{aligned} R(g) &= \frac{4T^{(2N)}(g) - T^{(N)}(g)}{3} \\ &= \int_a^b g(t)dt + \sum_{k=2}^{p} \frac{c_k}{3}\left(\frac{1}{4^{k-1}} - 1\right)\Delta^{2k} + O(\Delta^{2p+2}) \end{aligned} \tag{2.33}$$

En suivant ce procédé, il est possible de construire récursivement une formule de quadrature avec une erreur qui varie comme $O\left(\left(2^k+1\right)^{-2k}\right)$, dès que la fonction est au moins $2k+2$ fois continûment dérivable.

On définit pour tout $N \in \mathbb{N}$ la formule de quadrature $R(N, 0)$, qui corres-

pond à la méthode des trapèzes à $2^N + 1$ points :

$$R(N, 0) = T^{(2^N)} \tag{2.34}$$

Puis, pour $0 < M \leq N$, on définit la formule de quadrature $R(N, M)$ récursivement en utilisant la formule de récurrence suivante :

$$R(N, M) = \frac{4^M R(N, M-1) - R(N-1, M-1)}{4^M - 1} \tag{2.35}$$

En procédant ainsi, la formule de quadrature $R(N, M)$ utilise $2^N + 1$ points, et son erreur est en $O\left((2^N + 1)^{-2(M+1)}\right)$. Au maximum, la méthode de Romberg permet donc d'obtenir une erreur en $O\left((2^N + 1)^{-2(N+1)}\right)$ pour un nombre de $2^N + 1$ points.

D'autres méthodes, qui ne seront pas détaillées et n'ont pas été utilisées, telles la règle TANH [116] ou IMT [53] utilisent le fait que dans la formule d'Euler Maclaurin, les coefficients c_k sont proportionnels à la différence des dérivés de la fonction aux extrémités du segment. L'idée de ces méthodes est d'annuler ces coefficients c_k en effectuant un changement de variable qui permet d'annuler l'ensemble des dérivées aux bords de l'intervalle, permettant d'atteindre une convergence très rapide pour la simple méthode des trapèzes.

L'ensemble des méthodes de quadrature présentées jusqu'à présent utilisent des points d'évaluations équidistants sur le segment d'intégration. Il est possible de construire d'autres méthodes de quadratures en enlevant cette contrainte d'équidistance des points d'évaluations, qui sont plus puissantes que l'ensemble des méthodes présentés jusqu'à présent.

2.1.4 Méthodes de quadratures avancées

Les méthodes de Newton-Cotes sont basées sur l'idée de substituer à la fonction dont on cherche à calculer l'intégrale des fonctions polynomiales par morceaux. Les poids des méthodes de quadrature sont tels que ces fonctions polynomiales sont alors intégrées parfaitement numériquement. De ce fait, il s'avère qu'une méthode de Newton-Cotes composite de degré p sera capable d'intégrer parfaitement les polynômes de degré inférieur ou égal à p. En fait, les méthodes composite de Newton-Cotes de degré p impair sont capables d'intégrer parfaitement des polynômes de degré inférieur ou égal à $p+1$, alors que pour un degré p pair, les polynômes intégrés parfaitement sont ceux de degré inférieur ou égal à p. Une méthode de quadrature capable d'intégrer parfaitement l'ensemble des polynômes de degré inférieur ou égal à p, mais pour laquelle il existe un polynôme de degré $p + 1$ qui n'est pas intégré exactement par la méthode de quadrature sera dite d'ordre p.

2.1.4.1 Méthode de Gauss

L'idée des méthodes de quadrature de Gauss [38] est de construire des méthodes de quadrature avec un ordre le plus élevé possible étant donné le nombre N de points d'évaluation. Dans un cadre général, ces méthodes sont destinées à approximer l'intégrale I de fonctions pondérées par une fonction de pondération w sur un intervalle J, de la forme de (2.36).

$$I = \int_J f(t)w(t)dt \tag{2.36}$$

La fonction de pondération w est une fonction que l'on suppose ici à valeurs strictement positive sur l'intervalle J, et telle que l'ensemble de ses moments existent, c'est à dire :

$$\forall n \in \mathbb{N}, \int_J |x^n|\, w(x)dx < +\infty \tag{2.37}$$

La densité de probabilité d'une variable aléatoire ayant tous ses moments qui existent est typiquement une fonction de pondération. Pour une fonction de pondération w donnée, il est possible de définir un produit scalaire sur l'ensemble des fonctions dont le carré pondéré par w est intégrable sur J, ce produit scalaire étant défini par (2.38).

$$(f, g) = \int_J f(t)g(t)w(t)dt \tag{2.38}$$

Une méthode de Gauss associée à la fonction de pondération w est alors basée sur une base de polynômes orthogonaux $(P_k)_{k\in\mathbb{N}}$ où chaque P_k est de degré k. On peut montrer que les éléments d'une telle base sont uniques à un facteur multiplicatif scalaire près, et que les points d'évaluation de la quadrature de Gauss impliquant N points sont les racines du polynôme P_N, ces racines étant toutes distinctes et à l'intérieur de l'intervalle J [45]. Les polynômes orthogonaux P_k sont données par une relation de récurrence impliquant trois polynômes successifs à partir de laquelle il est possible de construire les points d'évaluation et les poids de la quadrature de Gauss [45]. Il existe d'autres moyens de construire les points d'évaluation et les poids, notamment quand l'ensemble des moments de la fonction de pondération w sont connus. En effet, la connaissance des moments implique que l'on est capable de déterminer l'intégrale de n'importe quel polynôme comme combinaison linéaire de ces moments. Le produit scalaire (2.38) entre deux polynômes quelconques est donc connu, et il est donc possible de construire une base de polynômes orthogonaux $(P_k)_{k\in\mathbb{N}}$ à partir d'un procédé d'orthogonalisation de Gramm-Schmidt appliqué à la base canonique de l'espace des polynômes. Une fois les polynômes P_k construits, il est

possible de déterminer les racines du polynôme P_N, à l'aide d'une méthode de Newton ou d'une méthode par dichotomie par exemple, pour obtenir les points d'évaluation de la quadrature de Gauss à N points. Les poids w_i de la quadrature peuvent être ensuite simplement calculés en remarquant que l'intégrale de chacun des N polynômes interpolateurs de Lagrange $(L_i)_{1\leq i\leq N}$ basés sur les N points d'évaluation $(x_i)_{1\leq i\leq N}$ est calculée exactement par la quadrature, et vérifie :

$$\int_J L_i(t)w(t)dt = \sum_{j=1}^{N} w_j L_i(x_j) = \sum_{j=1}^{N} w_j \delta_{ij} = w_i \tag{2.39}$$

Une méthode de Gauss à N points d'évaluation appliquée à une fonction g qui est $2N$ fois continûment dérivable donne une erreur d'intégration $E(g)$ qui vérifie la relation (2.40) [45] :

$$\exists M \in \mathbb{R}_+^*, \exists \alpha \in J, |E(g)| \leq M \frac{g^{(2N)}(\alpha)}{(2N)!} \tag{2.40}$$

L'erreur commise par les méthodes de Gauss en $O(1/(2N)!)$ est donc sans commune mesure avec celle des méthodes de Newton-Cotes présentées précédemment, qui sont en $O(N^{-(p+1)}$ ou $O(N^{-(p+2)})$ où p est le degré des polynômes choisis pour la méthode.

Il est possible de construire des méthodes de Gauss pour n'importe quelle fonction de pondération w, certaines ayant un nom spécifique pour des fonctions de pondérations usuelles comme la méthode de Gauss-Legendre, la méthode de Gauss-Hermite ou encore la méthode de Gauss-Laguerre. Seule l'utilisation de la méthode de Gauss-Legendre sera illustrée ici, qui correspond à l'intervalle $J = [-1, 1]$, avec une fonction de pondération uniforme et prise égale à $\frac{1}{2}$ dans cette thèse, définissant donc une densité de probabilité. La construction effective des points et des poids de quadrature est réalisée en pratique grâce à l'algorithme de Golub et Welsch [45]. Il est important de noter qu'il est toujours possible de se ramener à une intégrale sur $[-1, 1]$ avec une fonction de pondération uniforme à l'aide d'un changement de variable.

Les méthodes de Gauss sont très puissantes, mais souffrent en pratique du problème qu'il est impossible de réutiliser les évaluations de la fonction utilisées pour une méthode impliquant N points d'évaluation dans une méthode impliquant M points avec $M > N$, ce qui a poussé au développement des méthodes de Gauss-Kronrod [64] pour pallier à ce problème, modifiant les méthodes de Gauss originales en ajoutant des points pouvant être réutilisées. Néanmoins, d'autres méthodes puissantes existent également basées sur d'autres considérations pour leur construction que les méthodes de Gauss ou de Gauss-Kronrod, qui sont présentées dans le paragraphe suivant.

2.1.4.2 Méthodes de Clenshaw-Curtis et de Féjer

Comme dit précédemment, un des problèmes des quadratures de Gauss est la non possibilité de réutiliser des points d'évaluations lorsqu'on change le nombre de points d'évaluations utilisés pour approximer l'intégrale. Les méthodes de Clenshaw-Curtis [22] et Fejér [32] possèdent la propriété intéressante de pouvoir réutiliser les points d'évaluations alors même qu'on augmente le nombre de points d'évaluation utilisés, tout en ayant une vitesse de convergence similaire au méthode de Gauss [134].

Ces méthodes sont basées sur le changement de variable (2.41) qui permet de rendre paire la fonction à intégrer.

$$\int_{-1}^{1} f(t)dt = \int_{0}^{\pi} f(\cos(\theta))\sin(\theta)d\theta \tag{2.41}$$

Il s'agit donc d'intégrer la fonction paire $f(\cos(\theta))$. L'idée pour calculer cette intégrale est de considérer la série de Fourier de $f(\cos(\theta))$ (2.42).

$$f(\cos(\theta)) = \frac{a_0}{2} + \sum_{k=1}^{+\infty} a_k \cos(k\theta) \tag{2.42}$$

En substituant à $f(\cos(\theta))$ sa série de Fourier, l'intégrale de droite dans (2.41) devient alors :

$$\int_{0}^{\pi} f(\cos(\theta))\sin(\theta)d\theta = a_0 + \sum_{k=1}^{+\infty} \frac{2a_{2k}}{1-(2k)^2} \tag{2.43}$$

De l'expression (2.43), il vient qu'il faut évaluer les coefficients a_k qui sont donnés par le calcul d'une intégrale (2.44) encore une fois.

$$a_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f\left(\cos(\theta)\right)\cos(k\theta)d\theta = \frac{2}{\pi}\int_{0}^{\pi} f\left(\cos(\theta)\right)\cos(k\theta)d\theta \tag{2.44}$$

Il s'avère que les intégrales (2.44) sont des intégrales de fonctions 2π-périodique sur une demi période, qui lorsqu'elles sont approximées avec une méthode des trapèzes donnent une erreur qui est bien plus faible, pour une fonction f suffisamment régulière [55], que l'erreur présentée précédemment pour la méthode des trapèzes. Les modes pouvant être correctement reproduit en vertu du théorème d'échantillonnage de Shannon-Nyquist [118] sont ceux pour lesquels la fréquence est inférieure à la moitié de la fréquence d'échantillonnage du signal. En se basant sur une méthode des trapèzes impliquant $N+1$ points, on découpe l'intervalle d'une demi période $[0, \pi]$ en N sous in-

tervalles, ce qui correspond à une fréquence d'échantillonnage $f_e = \pi/N$. Les modes que l'on peut conserver sont donc ceux pour lesquels la fréquence est inférieure $\pi/2N$, soit des valeurs de k inférieures à $N/2$.

Plusieurs méthodes de quadratures ont été développées, différant par la méthode de quadrature utilisées pour le calcul des coefficients a_k de la décomposition de Fourier de $f(\cos(\theta))$. En utilisant la méthode des trapèzes à $N+1$ points, on obtient la méthode de Clenshaw-Curtis, dont les poids et les abscisses des points d'évaluations [141] sont données par (2.45) et (2.46) respectivement.

$$x_i = cos\left(\frac{i\pi}{N}\right), i \in [\![0, N]\!] \tag{2.45}$$

$$\begin{cases} w_i = \frac{c_i}{N}\left(1 - \sum_{j=1}^{E[N/2]} \frac{b_j}{4j^2-1}\cos\left(\frac{2ji\pi}{N}\right)\right), i \in [\![0, N]\!] \\ b_j = \begin{cases} 1, j = N/2 \\ 2, j < N/2 \end{cases}, c_i = \begin{cases} 1, i = 0\ ou\ N \\ 2, sinon \end{cases} \end{cases} \tag{2.46}$$

La seconde méthode de Fejér est basée sur le même jeu de points (2.45) que la méthode de Clenshaw-Curtis, sans les points extrémales, c'est à dire -1 et 1. Les poids pour cette méthode utilisées avec $N-1$ points [141] sont donnés par (2.47).

$$w_i = \frac{4}{N}\sin\left(\frac{i\pi}{N}\right)\sum_{j=1}^{E[N/2]} \frac{\sin\left(\frac{(2j-1)i\pi}{N}\right)}{2j-1}, i \in [\![1, N-1]\!] \tag{2.47}$$

La première méthode de Fejér est pour sa part basée sur des points d'évaluation différents [141], donnés par (2.48) pour N points d'évaluations utilisés. Les poids [141] sont obtenus en considérant une méthode du point médian pour évaluer l'intégrale (2.44), et sont donnés par (2.49).

$$x_i = cos\left(\frac{(2i+1)\pi}{2N}\right), i \in [\![0, N-1]\!] \tag{2.48}$$

$$w_i = \frac{2}{N}\left(1 - \sum_{j=1}^{E[N/2]} \frac{1}{4j^2-1}\cos\left(\frac{2j(2i+1)\pi}{N}\right)\right), i \in [\![0, N-1]\!] \tag{2.49}$$

Ces trois méthodes de quadrature intègrent parfaitement l'ensemble des polynômes de degré $N-1$ et sont donc d'ordre $N-1$, où N est le nombre de

nœuds utilisés. Elles sont basées sur des nœuds de Tchebychev qui ne sont pas équidistants les uns des autres comme pour les formules de Newton-Cotes. En fait, ces méthodes peuvent aussi être obtenues en considérant les polynômes interpolateurs en les nœuds de Tchebychev utilisés pour chaque méthode [141]. Ces nœuds sont les racines des polynômes de Tchebychev, qui présentent la propriété intéressante d'être des points qui atténuent le plus fortement le phénomène de Runge [109], ce qui est un autre avantage de ces méthodes par rapport aux méthodes de Newton-Cotes. Un dernier point important à retenir est la possibilité de réutiliser les points d'évaluation entre la méthodes de Clenshaw-Curtis à $N+1$ points et celle à $2N+1$ points, et la même chose est possible avec la seconde méthode Fejér entre la méthode à $N-1$ points et celle à $2N-1$ points. Il n'est pas possible de faire de même avec la première méthode de Fejér.

2.1.5 Fonctions présentant des singularités

Hormis pour une partie des méthodes de Gauss, l'ensemble des méthode de quadratures présentées jusqu'à maintenant sont définies pour calculer l'intégrale sur un segment avec une fonction de pondération uniforme sur celui-ci. Comme précisé auparavant, il est possible de se ramener au calcul d'une intégrale de cette forme sous les bonnes conditions, qui seront toujours vérifiées dans cette thèse. Cependant, un tel changement de variable peut impliquer de transformer des bornes infinies en bornes finies par exemple, ce qui peut amener la fonction à intégrer à posséder des singularités aux bords du nouvel intervalle. Un exemple est présenté sur la figure 2.8 où un changement de variable ϕ est utilisé pour passer d'un cas où la fonction de pondération est $w(x) = e^{-x}$ sur l'intervalle $[0, +\infty[$ à un cas où la fonction de pondération est $w(u) = 1$ sur l'intervalle $[0, 1]$. La nouvelle fonction à intégrer présente une asymptote verticale en $u = 1$, les valeurs à l'infini de la fonction initiale étant ramenées en $u = 1$.

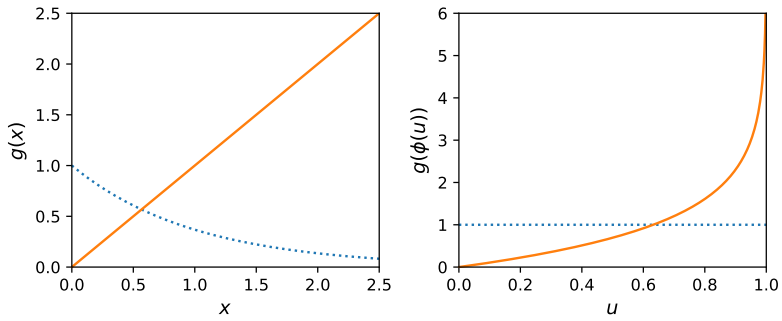

FIGURE 2.8 – *Gauche : fonction de pondération* $w(x) = e^{-x}$ *(pointillés) et fonction* $g(x) = x$ *(ligne pleine) sur* $[0, +\infty[$. *Droite : fonction de pondération* $w(u) = 1$ *(pointillés) et fonction* $g(\phi(u)) = \ln(\frac{1}{1-u})$ *(ligne pleine) définies sur* $[0, 1]$.

Dans les pires cas, la nouvelle fonction à intégrer n'est même pas défi-

nie sur les bords, comme dans l'exemple précédent, et cela implique qu'il sera impossible d'utiliser les méthodes de quadrature utilisant une évaluation de la fonction aux bords de l'intervalle, comme c'est le cas pour l'ensemble des méthodes de Newton-Cotes présentées ici, qui sont dites fermées. Il est possible de construire des méthodes de Newton-Cotes ouvertes n'utilisant pas les bords de l'intervalle d'intégration, mais cela ne sera pas utilisé dans la suite du fait du peu d'intérêt que présentent ces méthodes. Les méthodes de Gauss pour leur part sont toutes ouvertes [45], et sont donc applicables dans l'ensemble des cas. Finalement, concernant les méthodes de Clenshaw-Curtis et Fejér, il s'avère que les deux méthodes de Fejér sont ouvertes alors que la méthode de Clenshaw-Curtis ne l'est pas.

2.1.6 Comparaision des différentes quadratures

Les différentes méthodes de quadratures sont testées sur différentes fonctions, ne présentant pas toutes les mêmes régularités. Les fonctions utilisées pour comparer les méthodes de quadrature sont issues de [40], et sont toutes définies sur $[0, 1]$. Toutes les méthodes de quadrature introduites dans cette section étant définies sur $[-1, 1]$, une transformation affine a permis de se ramener à $[0, 1]$. Ces fonctions test sont présentées dans l'annexe A, les tests ayant été réalisés avec une valeur de u choisie aléatoirement entre 0 et 1 strictement (la même est cependant utilisé pour les différentes méthodes afin que la comparaison ait du sens), alors que pour a la valeur a été prise égale à 1. La valeur exacte de l'intégrale est pour sa part calculer analytiquement comme expliqué dans l'annexe A.

Les valeurs exactes des intégrales sur $[0, 1]$ pour l'ensemble de ces fonctions ont été calculées afin de pouvoir estimer l'erreur d'intégration relative E_{rel} commise par les différentes méthodes de quadrature. Sur la figure 2.9 sont représentées les valeurs absolues des erreurs relatives E_{rel} d'estimations de l'intégrale commises en fonction du nombre N de points utilisées dans la méthode de quadrature, pour les différentes fonctions test, les méthodes de quadrature utilisées étant des méthodes composites de Newton-Cotes fermées et ouvertes avec un degré p pour le polynôme d'interpolation égal à 1, 2 ou 4.

Pour les fonctions infiniment dérivables, c'est à dire ayant la dénomination "Oscillatory", "Product Peak", "Corner Peak" et "Gaussian", la valeur absolue de l'erreur relative $|E_{rel}|$ en fonction du nombre de points N pour chacune des méthodes de quadratures est une droite dans le diagramme log-log. Cela traduit le fait qu'il existe une convergence de la valeur absolue de l'erreur proportionnelle à N^{α}, où α est la pente de la droite. Les pentes des méthodes ouvertes et fermées sont les mêmes pour un degré p donné. De plus, les pentes pour chaque méthode sont indépendantes de la fonction à intégrer, ce qui traduit que la convergence est indépendante de la fonction à intégrer tant que celle-ci est suffisamment régulière pour la méthode de quadrature utilisée. Enfin, on peut voir que l'on retrouve les valeurs théoriques pour la pente des droites,

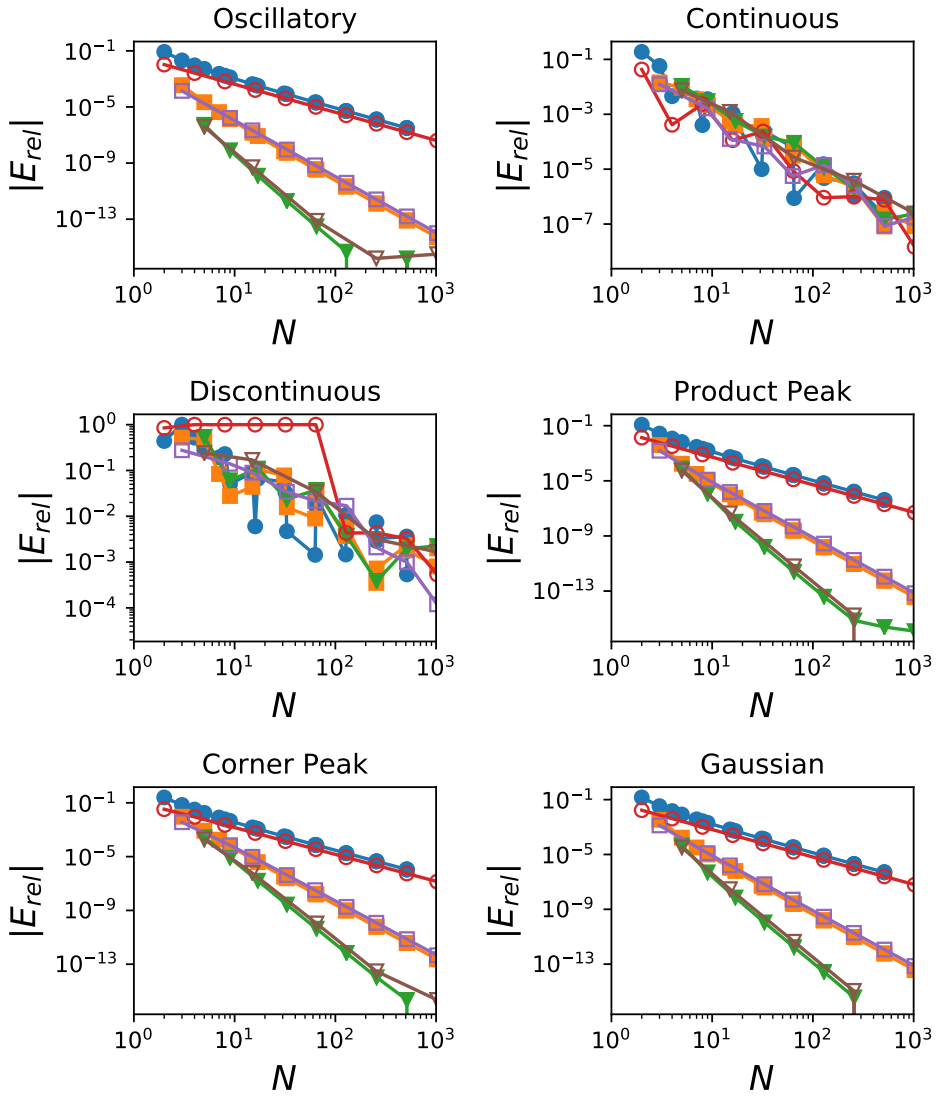

FIGURE 2.9 – *Valeur absolue de l'erreur relative E_{rel} d'intégration des différentes fonctions test en fonction du nombre de points N utilisés, en diagramme log-log. Les méthodes de quadratures utilisées sont des méthodes composites de Newton-Cotes fermées (symboles remplis en haut) et ouvertes (symboles remplis en bas), avec une valeur du degré p égale à 1 (cercle), égale à 2 (carré) et égale à 4 (triangle).*

c'est à dire $\alpha = -2$ pour $p = 1$, $\alpha = -4$ pour $p = 2$ et $\alpha = -6$ pour $p = 4$. Pour les fonctions présentant une régularité insuffisante du fait de la présence d'une singularité, non dérivable en un point pour la fonction "Continuous", et non continue en un point pour la fonction "Discontinuous", la convergence des différentes méthodes est sensiblement la même quelle que soit la méthode de

quadrature utilisée.

Sur la figure 2.10 sont représentées les valeurs absolues des erreurs relatives E_{rel} d'estimations de l'intégrale commises en fonction du nombre N de points utilisées dans la méthode de quadrature, les méthodes de quadrature utilisées étant celles de Romberg, de Fejér, de Clenshaw-Curtis, et de Gauss Legendre.

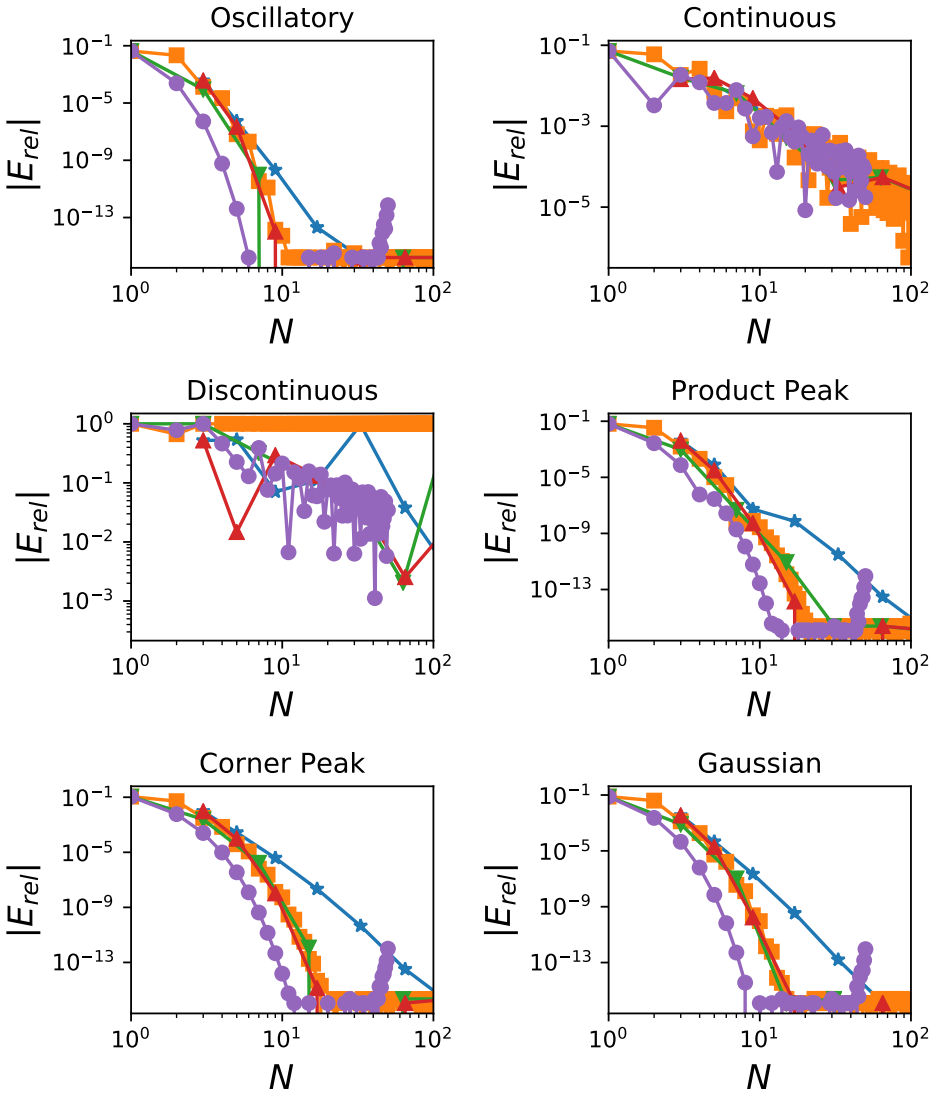

FIGURE 2.10 – *Valeur absolue de l'erreur relative E_{rel} d'intégration des différentes fonctions test en fonction du nombre de points N utilisés, en diagramme log-log. Les méthodes de quadratures utilisées sont la méthode de Romberg (étoile), la première méthode de Fejér (carré), la seconde méthode de Fejér (triangle bas), la méthode de Clenshaw-Curtis (triangle haut) et la méthode de Gauss Legendre (cercle).*

Encore une fois, le comportement est différent suivant que la fonction à intégrer est régulière ou non. Pour les quatre fonctions infiniment dérivables, la méthode de Gauss Legendre est la plus performante, capable d'intégrer ces fonctions en utilisant seulement une dizaine de points N au bruit numérique dû à la précision machine près qui est de l'ordre de 10^{-15} dans le cas de l'utilisation de flottants en double précision. Les méthodes de Fejér et de Clenshaw-Curtis sont toutes équivalentes, et légèrement moins performantes que la méthode de Gauss Legendre, mais avec une différence de seulement quelques points d'évaluations supplémentaires pour approximer l'intégrale avec une précision de l'ordre du bruit numérique. La méthode de Romberg quant à elle est moins performante que les méthodes précédentes, mais est capable d'intégrer les fonctions avec moins de 100 points d'évaluations jusqu'à une erreur de l'ordre du bruit numérique, ce qui est plus performant que les méthodes composites de Newton-Cotes précédentes. L'ensemble des courbes sont concaves tant qu'elles n'ont pas atteint le niveau du bruit numérique, ce qui traduit le fait qu'il arrive un moment où elles deviennent toutes plus performantes que n'importe quelle méthode qui aurait une erreur d'estimation de l'ordre de N^α avec α un nombre négatif fixé, comme c'est le cas pour les méthodes de Newton Cotes. Pour les fonctions présentant des irrégularités, les méthodes sont toutes équivalentes, et ont des performances qui sont sensiblement les mêmes que les méthodes de Newton Cotes.

Comme prévu par la théorie, les méthodes de Newton-Cotes sont bien moins puissantes que les méthodes de Gauss, de Clenshaw-Curtis et Fejér concernant les fonctions suffisamment régulières. Les méthodes de Gauss-Legendre, Clenshaw-Curtis et Fejér présentent des vitesses de convergences proches, bien que la méthode de Gauss-Legendre restent plus puissantes que ces dernières. Néanmoins, le fait de pouvoir réutiliser les nœuds pour la méthode de Clenshaw-Curtis et la seconde méthode de Fejér fait qu'il est en pratique souvent plus intéressant d'utiliser ces dernières, en particulier lorsque l'on souhaite utiliser ces méthodes pour intégrer des fonctions de plusieurs variables où des tensorisations creuses peuvent être utilisées. Dans le cas de fonctions qui ne sont pas suffisamment régulières, toutes les méthodes sont sensiblement équivalentes, ce qui ne donne aucune méthode comme étant préférentielle. Enfin, dans le cas où les points extrêmes de l'intervalle présentent une singularité faisant que la fonction ne puissent pas y être évaluée, des méthodes ouvertes sont à utiliser, excluant la méthode de Clenshaw-Curtis, de Romberg et les méthodes de Newton-Cotes fermées.

2.2 Intégration numérique de fonctions de plusieurs variables

L'objectif de cette partie est de présenter des méthodes efficaces permettant d'approximer numériquement l'intégrale de fonctions de d variables, avec

un nombre d restant faible, typiquement inférieur à 5. Dans le cadre d'un calcul d'espérance du type de l'expression (1.3), il sera montré dans le chapitre 3 qu'il est toujours possible de se ramener à des intégrales sur l'hypercube $[0,1]^d$ muni de la mesure de probabilité uniforme, et il est immédiat à partir de celui-ci de passer à l'hypercube $[-1,1]^d$ muni encore une fois de la mesure de probabilité uniforme. Dans la suite, on ne s'intéressera donc qu'à des fonctions définies sur $[-1,1]^d$ dont on cherche à approximer l'intégrale.

2.2.1 Exemple en dimension 2

On s'intéresse à une fonction réelle de deux variables, définies sur $[-1,1]^2$. Une telle fonction peut être représentée par une surface, comme présenté sur la figure 2.11.

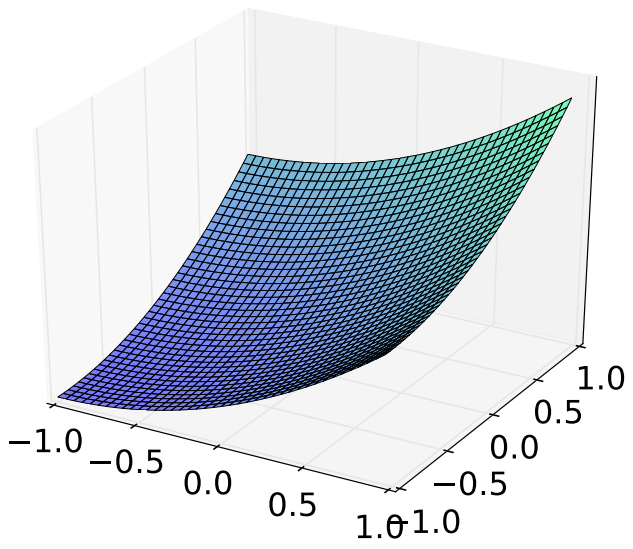

FIGURE 2.11 – *Graphe d'une fonction réelle à deux variables définies sur* $[-1,1]^2$

De manière analogue au cas de la dimension 1, l'intégrale de cette fonction peut être vue comme le volume compris entre la surface et le plan d'équation $z = 0$, compté positivement lorsque ce volume est au-dessus du plan $z = 0$, et compté négativement sinon.

Pour évaluer ce volume, il est possible de s'inspirer de ce qui a été fait pour les fonctions réelles d'une variable avec la méthode du point médian par exemple, où la fonction à intégrer était remplacée par une fonction en "escalier". Pour construire cette fonction en "escalier", l'intervalle de définition avait été coupé en segments de même taille, sur lesquels était assignée une va-

leur constante pour la fonction en "escalier", qui était la valeur de la fonction au milieu du segment dans le cas de la méthode du point médian. L'intégrale de cette fonction en escalier était alors la somme d'une aire de rectangles, qui est facilement calculable.

Encore une fois, l'idée est de découper l'ensemble de définition, qui est ici le carré $[-1,1]^2$. La manière la plus simple de découper ce carré est de le découper en rectangle de taille identique, à l'aide de coupes parallèles aux côtés du carré initial. Ensuite, on peut remplacer le calcul de l'intégrale de la fonction par le calcul de l'intégrale d'une fonction qui est constante sur chacun des sous rectangles, la valeur de cette fonction sur le rectangle étant la valeur de la fonction initiale au centre du rectangle. L'intégrale d'une telle fonction étant une somme de volume de parallélépipède rectangle, est facilement calculable.

Différentes illustrations d'une fonction initiale et de la fonction constante par morceaux sont présentées, avec différents découpages du carré initiale sur la figure 2.12. En haut à gauche, le carré initial n'a pas été découpé, en haut à droite et en bas à gauche, le carré initial a été découpé en deux suivant la première et la seconde variable, et en bas à droite le domaine de définition est découpé en quatre carrés identiques.

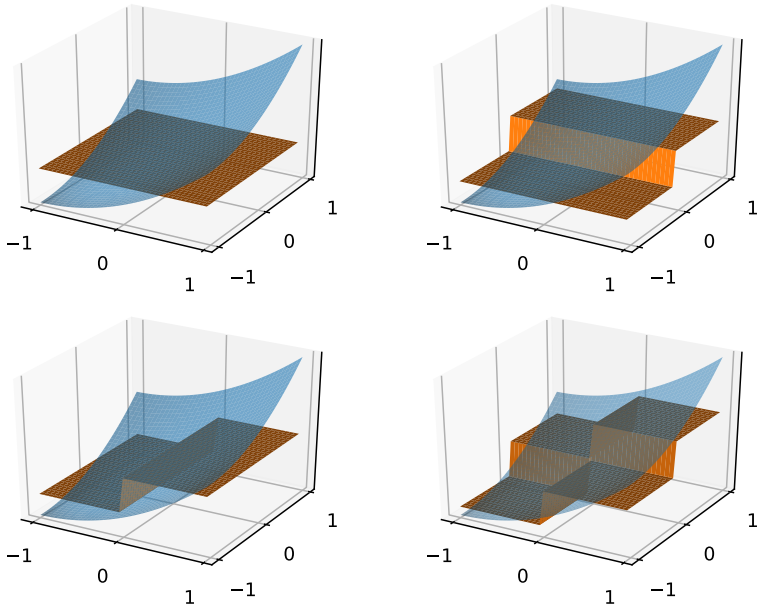

FIGURE 2.12 – *Fonction exemple et différentes fonctions par morceaux utilisés par la tensorisation de la méthode du point médian pour le calcul numérique de l'intégrale.*

On suppose maintenant que l'on coupe le carré initial en $N_x N_y$ rectangle, en découpant l'ensemble de définition de la première variable en N_x segments de taille égale, et celui de la seconde variable en N_y segments de taille égale également. Étant sur $[-1,1]^2$, chaque rectangle a donc une aire de $\frac{4}{N_x N_y}$. En

notant (x_i, y_j) le centre du rectangle placé sur la i-ème ligne et j-ème colonne, on a donc pour l'approximation de l'intégrale d'une fonction g par cette méthode :

$$\int_{-1}^{1}\int_{-1}^{1} g(x,y)dxdy \approx \sum_{i=1}^{N_x}\sum_{j=1}^{N_y} \frac{4}{N_x N_y} g(x_i, y_j) \tag{2.50}$$

Dans cette expression, on peut voir apparaître les abscisses d'évaluation de la méthode du point médian à N_x points pour la première variable, et celle à N_y points pour la seconde variable. De plus, cette expression peut se réécrire de façon à faire apparaître les poids de la méthode du point médian utilisant N_x éléments et celle utilisant N_y éléments, qui sont respectivement $\frac{2}{N_x}$ et $\frac{2}{N_y}$:

$$\int_{-1}^{1}\int_{-1}^{1} g(x,y)dxdy \approx \sum_{i=1}^{N_x}\sum_{j=1}^{N_y} \frac{2}{N_x}\frac{2}{N_y} g(x_i, y_j) \tag{2.51}$$

Cet exemple simple a permis de construire une méthode d'intégration numérique de fonctions réelles de deux variables à l'aide de la méthode du point médian définie en dimension 1. Il est possible de généraliser cette idée à plus de variables, et avec d'autres méthodes de quadrature. Un tel procédé est en fait une tensorisation de méthodes de quadrature à une variable.

2.2.2 Tensorisation pleine de méthodes de quadrature

Dans un soucis de simplicité et de clarté, seules des méthodes de quadratures sur l'hypercube $[-1,1]^p$ seront considérées à présent, ce qui sera présenté par la suite pouvant être généralisé.

On considère d méthodes de quadrature $Q^{(i)}$ sur $[-1,1]$, chacune caractérisée par une famille de N_i points d'évaluations $(x_j^{(i)})_{1\leq j\leq N_i}$ et une famille de N_i poids $(w_j^{(i)})_{1\leq j\leq N_i}$. A partir de ces méthodes de quadratures, on définit la méthode de cubature C permettant d'approximer l'intégrale d'une fonction g de d variables et définie sur $[-1,1]^d$ comme étant la tensorisation des quadratures $Q^{(i)}$ par :

$$\begin{aligned} C(g) &= \left(Q^{(1)} \otimes \cdots \otimes Q^{(d)}\right)(g) \\ &= \sum_{i_1=1}^{N_1} \cdots \sum_{i_d=1}^{N_d} w_{i_1}^{(1)} \ldots w_{i_d}^{(d)} g\left(x_{i_1}^{(1)}, \ldots, x_{i_d}^{(d)}\right) \end{aligned} \tag{2.52}$$

La cubature C requiert donc l'évaluation de g en $N = \prod_{i=1}^{d} N_i$ points, qui correspondent aux noeuds d'une grille définies par les quadratures $Q^{(i)}$. Des exemples de telles grilles sont présentés sur la gauche de la figure 2.17.

Lorsqu'une méthode de quadrature impliquant M points est utilisée selon toutes les dimensions, la taille de la cubature augmente exponentiellement avec la dimension d du problème, le nombre de points d'évaluation total dans la tensorisation pleine de celles-ci étant $N = M^d$. De plus, comme précisé dans [91], la précision de la méthode de cubature dans un tel cas n'est pas nécessairement meilleure que la précision de la méthode de quadrature utilisée sur chaque dimension, comme il est possible de s'en convaincre en considérant une fonction qui ne dépend que d'une seule des variables, pour laquelle la méthode de cubature aura donc la précision de la quadrature dans la direction de cette variable. Par exemple, en considérant une méthode des trapèzes, et une fonction deux fois continûment dérivable, l'erreur commise sera de l'ordre de M^{-2}, et en terme du nombre total $N = M^d$ de points d'évaluation de la cubature, elle pourra s'exprimer comme $N^{-2/d}$. Sur la figure 2.13 est représenté ce phénomène pour la tensorisation de la méthode des trapèzes, la dimension d variant de 1 à 5. On peut observer qu'à mesure que la dimension d augmente, la valeur absolue de la pente diminue et est bien égale à $2/d$, comme prévu théoriquement, la convergence se dégradant donc avec l'augmentation de la dimension d.

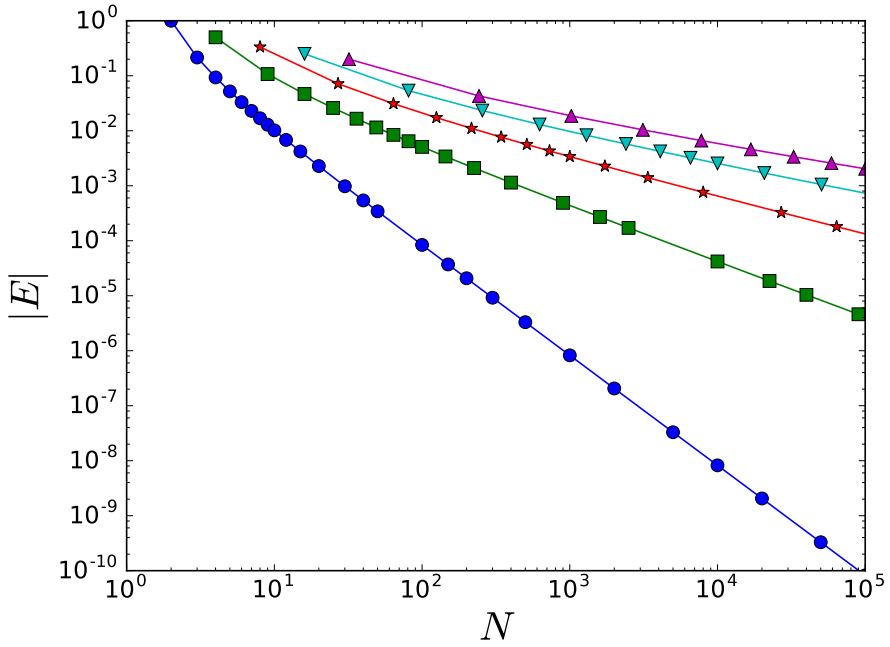

FIGURE 2.13 – *Valeur absolue de l'erreur relative E commise en utilisant une tensorisation uniforme de la méthode des trapèzes en fonction du nombre de points d'évaluation N, pour la fonction $g(\mathbf{x}) = \frac{1}{d}\sum_{i=1}^{d} \sin(\pi x_i)$ définie sur $[0,1]^d$. La dimension d prend les valeurs 1 (cercles), 2 (carrés), 3 (étoiles), 4 (triangles bas) et 5 (triangles hauts).*

Ce phénomène, qui voit l'erreur d'approximation se détériorer avec l'augmentation de la dimension d du problème est appelé le "fléau de la dimension" ("curse of dimensionality" en anglais), et est la raison principale de l'impossibilité d'appliquer des méthodes de cubatures en dimension trop importante, typiquement supérieur à 5. Cette impossibilité vient également du fait que le nombre de points d'évaluations augmente exponentiellement avec la dimension d. Néanmoins, il est possible de modérer les effets de l'augmentation de la dimension, en faisant appel à des tensorisations creuses à la place des tensorisations pleines, qui permettent d'avoir une augmentation du nombre de points avec la dimension plus faible, sans modifier significativement l'erreur d'approximation [58].

2.2.3 Tensorisation creuse

La tensorisation pleine de méthodes de quadrature produit des méthodes de cubature pour lesquelles les points d'évaluations occupent les nœuds d'un maillage structuré, dont le nombre augmente exponentiellement avec la dimension d. La tensorisation creuse de méthodes de quadrature permet de créer des méthodes de cubature avec un nombre de points d'évaluation moins important, en ne considérant qu'une partie des nœuds d'un ou plusieurs maillages structurés, permettant d'offrir un moindre coût de calcul que la tensorisation pleine. L'utilisation de quadratures dites imbriquées qui sont présentés dans le paragraphe suivant permet d'assurer que la tensorisation creuse implique des nœuds provenant d'un unique maillage structuré, permettant de réduire au minimum le nombre de points d'évaluation nécessaires.

2.2.3.1 Méthodes de quadrature imbriquées

Une famille de méthodes de quadrature $(Q_l)_{l\in\mathbb{N}}$ est dite imbriquée si les points d'évaluation de la quadrature Q_l, qu'on dira de niveau l, sont parmi les points d'évaluation de la quadrature Q_{l+1}. L'utilisation d'une famille de méthodes de quadrature imbriquées permet donc de raffiner la valeur de l'approximation de l'intégrale à calculer en limitant le nombre d'évaluations de la fonction à intégrer lors d'approximations successives.

L'ensemble des méthodes de quadratures présentées précédemment permettent d'obtenir des familles de méthodes de quadratures imbriquées, hormis les méthodes de Gauss ainsi que la première méthode de Fejér.

Les méthodes de Newton-Cotes, qu'elles soient composite ou non, peuvent permettre de construire une famille de quadratures imbriquées. Dans le cas où la méthode de Newton-Cotes considérée est fermée, il suffit de doubler le nombre points entre deux niveaux de quadratures, comme illustré sur la figure 2.14 pour la la méthode des trapèzes et la méthode de Newton-Cotes fermée composite impliquant un degré $p = 3$ pour les polynômes.

Dans le cas où la méthode de Newton-Cotes considérée est ouverte, il suffit

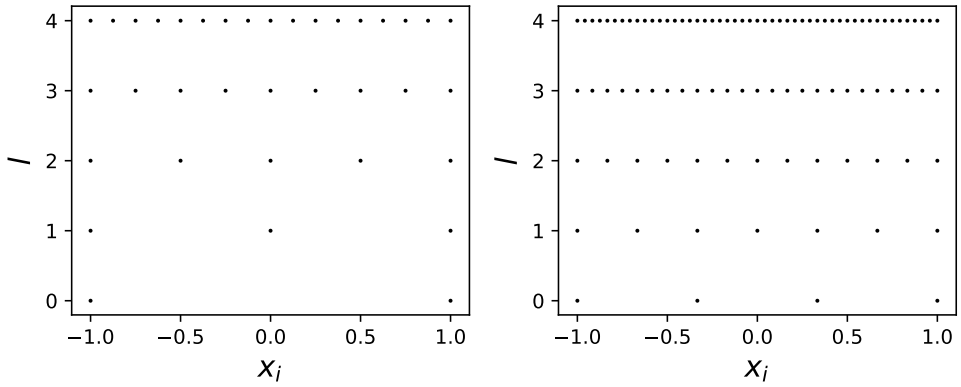

FIGURE 2.14 – *Abscisses des points d'évaluations x_i de quadratures imbriqués de Newton-Cotes fermées en fonction du niveau l. Gauche : méthode des trapèzes ($p = 1$). Droite : méthode de Newton-Cotes fermée avec $p = 3$.*

de tripler le nombre de points entre deux niveaux de quadratures, comme illustré sur la figure 2.15 pour la la méthode du point médian et la méthode de Newton-Cotes ouverte composites impliquant un degré $p = 2$ pour les polynômes.

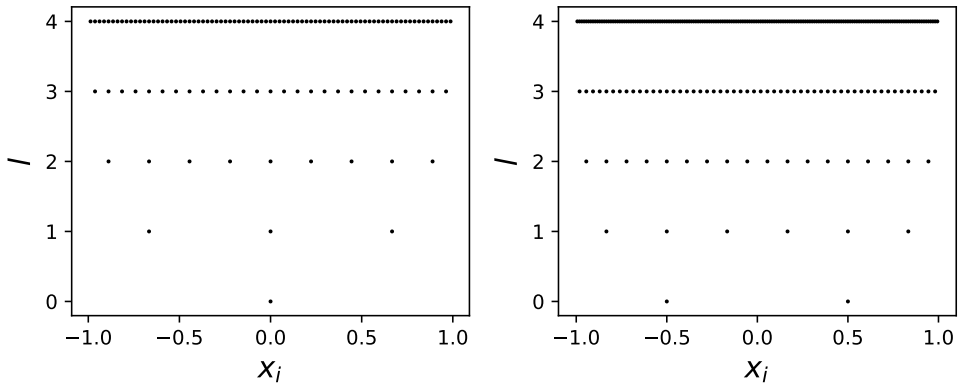

FIGURE 2.15 – *Abscisses des points d'évaluations x_i de quadratures imbriqués de Newton-Cotes ouvertes en fonction du niveau l. Gauche : méthode du point médian ($p = 0$). Droite : méthode de Newton-Cotes ouverte avec $p = 2$.*

Comme dit précédemment, les méthodes de Clenshaw-Curtis et la seconde méthode de Fejér peuvent également s'arranger afin d'obtenir une famille de méthodes de quadrature imbriquées. Pour cela, il suffit de considérer pour le niveau $l \geq 1$ la méthode de Clenshaw-Curtis avec un nombre de points d'évaluation égal à $N_l = 2^l + 1$, et pour la seconde méthode de Fejér celles avec un nombre de points d'évaluation égal à $N_l = 2^l - 1$. Les points d'évaluation en fonction du niveau pour Clenshaw-Curtis sont visibles sur la figure 2.16, les

points de l'extrémité devant être retiré pour obtenir ceux de la seconde méthode de Fejér.

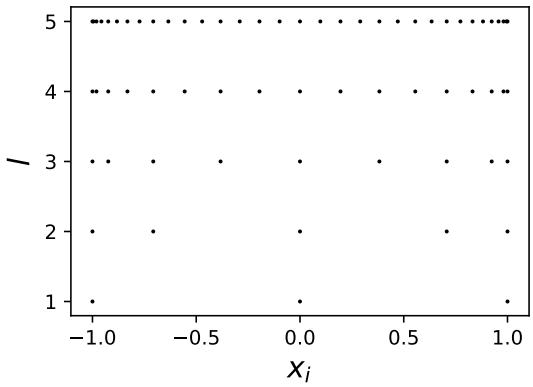

Figure 2.16 – *Abscisses des points d'évaluations x_i de quadratures imbriqués de Clenshaw-Curtis en fonction du niveau l.*

2.2.3.2 Méthode de Smolyak

Smolyak fût le premier à proposer une méthode de construction de grilles creuses [123]. La construction de Smolyak présentée ici utilise une famille de méthodes de quadrature imbriquées $(Q_l)_{l\in\mathbb{N}}$, chacune possédant son propre niveau l, à partir de laquelle on définit une famille d'opérateurs $(\Delta_l)_{l\in\mathbb{N}}$ comme :

$$\begin{cases} \Delta_0 = Q_0 \\ \forall l \in \mathbb{N}^*, \Delta_l = Q_l - Q_{l-1} \end{cases} \tag{2.53}$$

En notant P_l l'ensemble des points d'évaluations de la méthode de quadrature Q_l, et $P_l \setminus P_{l-1}$ l'ensemble des points d'évaluations de la méthode de quadrature Q_l n'appartenant pas à la méthode de quadrature Q_{l-1}, on trouve que l'opérateur Δ_l avec $l \geq 1$ appliqué à une fonction g donne :

$$\begin{aligned} \Delta_l(g) &= \sum_{x_j \in P_l} w_j^{(l)} g(x_j) - \sum_{x_j \in P_{l-1}} w_j^{(l-1)} g(x_j) \\ &= \sum_{x_j \in P_l \setminus P_{l-1}} w_j^{(l)} g(x_j) + \sum_{x_j \in P_{l-1}} (w_j^{(l)} - w_j^{(l-1)}) g(x_j) \end{aligned} \tag{2.54}$$

Avec cette construction de l'opérateur Δ_l, on remarque qu'il est possible

de reconstruire les méthodes de quadratures Q_l par télescopage :

$$Q_l = \sum_{j=0}^{l} \Delta_j \tag{2.55}$$

En fait, l'opérateur Δ_l peut être vu comme une correction à apporter à la méthode de quadrature Q_{l-1}, afin d'obtenir une meilleure approximation, qui correspond à celle donnée par la méthode de quadrature Q_l.

L'intérêt de la famille d'opérateur $(\Delta_l)_{l\in\mathbb{N}}$ se trouve dans la construction de méthode de cubature basée sur des grilles dites creuses. Pour cela, on considère ici pour chaque dimension i une méthode de quadrature $(Q_l^{(i)})_{l\in\mathbb{N}}$, à partir desquelles on peut construire des familles d'opérateurs $(\Delta_l^{(i)})_{l\in\mathbb{N}}$. Maintenant, étant donné un multi-indice $\mathbf{l} = (l_1, \ldots, l_d) \in \mathbb{N}^p$, on peut construire un opérateur $\Delta_{\mathbf{l}}$ multi-dimensionnel comme la tensorisation pleine des opérateurs $(\Delta_{l_k})_{1\leq k\leq d}$:

$$\Delta_{\mathbf{l}} = \left(\Delta_1^{(l_1)} \otimes \cdots \otimes \Delta_d^{(l_d)}\right) \tag{2.56}$$

Étant donné un entier k, la méthode de Smolyak consiste à construire la méthode de cubature suivante :

$$C_k = \sum_{\mathbf{l}\in\mathbb{N}^p, |\mathbf{l}|_1\leq k} \Delta_{\mathbf{l}} \tag{2.57}$$

Dans l'expression précédente, $|\mathbf{l}|_1$ correspond à la norme 1 de $\mathbf{l}$, définie par :

$$|\mathbf{l}|_1 = \sum_{i=1}^{d} l_i \tag{2.58}$$

L'entier k est donc ici le seul degré de liberté, qui permet de contrôler le nombre de points d'évaluation qui seront présents dans la grille. Ce degré de liberté k peut également servir à la définition d'une tensorisation pleine, en utilisant encore une fois le télescopage entre les opérateurs Δ, ce qui donne l'expression suivante, où la norme ∞ est utilisé :

$$Q_k^{(full)} = \sum_{\mathbf{l}\in\mathbb{N}^p, |\mathbf{l}|_\infty\leq k} \Delta_{\mathbf{l}} \tag{2.59}$$

Cette dernière expression permet d'exprimer la tensorisation pleine $Q_k^{(full)}$ à l'aide de la tensorisation creuse C_k associée à la méthode de Smolyak :

$$Q_k^{(full)} = C_k + \sum_{\substack{\mathbf{l}\in\mathbb{N}^p, |\mathbf{l}|_1 > k \\ \mathbf{l}\in\mathbb{N}^p, |\mathbf{l}|_\infty \leq k}} \Delta_{\mathbf{l}} \tag{2.60}$$

Les points d'évaluation présents uniquement dans les opérateurs $\Delta_{\mathbf{l}}$ de la somme de droite dans l'expression précédente correspondent aux points d'évaluations supplémentaires utilisés par la tensorisation pleine et non la tensorisation creuse. A droite de la figure 2.17 sont représentées des exemples de grilles creuses correspondant à l'ensemble des points d'évaluation de la méthode de cubature construite par la méthode de Smolyak avec une méthode de quadrature de Clenshaw-Curtis ainsi que la seconde méthode de Fejér. Ces grilles creuses sont à comparer aux grilles pleines associées. Le ratio du nombre de points d'évaluations entre la grille creuse et la grille pleine est bien plus faible pour la seconde méthode de Fejér que pour la méthode de Clenshaw-Curtis, du fait que les points extrêmes -1 et 1 ne sont pas présents. Ce fait est encore plus vrai en dimension plus grande, où l'ensemble des points sur les bords sont encore plus nombreux, leur nombre augmentant exponentiellement avec la dimension d.

Comme visible sur la figure 2.17, la méthode de Smolyak permet la construction de grille creuse qui sont isotrope lorsque les méthodes de quadrature selon chaque dimension sont les mêmes, ce qui est généralement le cas. Il est possible de construire des grilles non isotropes, qui peuvent avoir un intérêt pour des fonctions où une variable joue un rôle plus important que les autres. Ces grilles doivent cependant posséder certaines propriétés qui sont décrites dans la section suivante.

2.2.3.3 Algorithme adaptatif utilisant des grilles creuses

Comme expliqué dans [41], une méthode de cubature peut s'écrire sous la forme suivante :

$$C_{\mathcal{I}} = \sum_{\mathbf{l}\in\mathcal{I}} \Delta_{\mathbf{l}} \tag{2.61}$$

Dans cette expression, l'ensemble de multi-indices $\mathcal{I}$ doit vérifier certaines propriétés pour que la cubature $C_{\mathcal{I}}$ converge bien vers l'intégrale de la fonction à intégrer. En notant $e_1, \ldots, e_d$ les vecteurs de la base canonique de $\mathbf{R}^d$, on a les propriétés suivantes pour l'ensemble $\mathcal{I}$:

$$\begin{cases} (0,\ldots,0) \in \mathcal{I} \\ \forall \mathbf{l} = (l_1,\ldots,l_d) \in \mathcal{I}, \forall 1 \leq i \leq d, ((l_i > 0) \Rightarrow (\mathbf{l} - e_i \in \mathcal{I})) \end{cases} \tag{2.62}$$

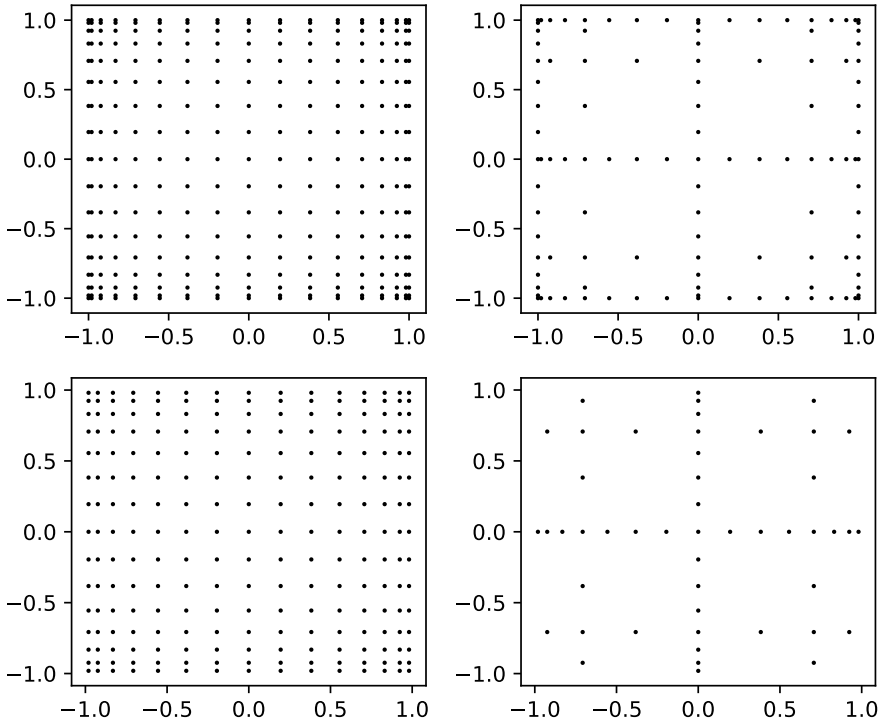

FIGURE 2.17 – *Exemples de tensorisations pleines (à gauche avec $|l|_\infty < 4$) et creuses construites à l'aide de la méthode de Smolyak (à droite avec $|l|_1 < 4$). En haut se trouvent celles construites avec la méthode de Clenshaw-Curtis et en bas se trouvent celles construites avec la seconde méthode de Fejér.*

Ces propriétés assurent que l'ensemble $\mathcal{I}$ ne possède pas de "trous", ce qui assure un bon télescopage des opérateurs Δ afin que la cubature concernée converge bien vers la valeur de l'intégrale recherchée.

Plusieurs choix sont possibles pour l'ensemble $\mathcal{I}$, comme par exemple les deux exemples suivants :

— $\mathcal{I} = \left\{(l_1, \ldots, l_d) \in \mathbb{N}^d : \forall 1 \leq i \leq d, l_i \leq k\right\}$
— $\mathcal{I} = \left\{(l_1, \ldots, l_d) \in \mathbb{N}^d : \sum_{i=1}^{d} l_i \leq k\right\}$

Ces deux ensembles vérifient bien les propriétés énoncées précédemment. Le premier de ces deux exemples correspond en fait à une tensorisation pleine. Le second exemple construit une grille creuse isotrope, correspondant à la méthode de Smolyak explicité dans la section précédente.

Le choix de l'allure de la grille de points, qui se fait à travers le choix de la forme de l'ensemble $\mathcal{I}$ a un impact sur le coût de calcul, à travers le nombre de points présents dans la grille, et également sur la précision du résultat. Des algorithmes adaptatif de construction pas à pas de l'ensemble $\mathcal{I}$ existent

cependant, ce qui évite d'avoir à faire un choix prédéfini sur la forme de cet ensemble.

Une construction de méthode de cubature à grille creuse adaptative a été proposée dans [41]. L'algorithme proposé construit pas à pas l'ensemble $\mathcal{I}$ en s'assurant qu'il respecte toujours les contraintes qui lui sont imposées.

Pour cela, on considère l'ensemble $\mathcal{O}$ des multi-indices admissibles, c'est à dire tels que pour $\mathbf{l} \in \mathcal{O}$, $\{\mathbf{l}\} \cup \mathcal{I}$ vérifie encore les propriétés (2.62). On considère également pour chaque multi-indice $\mathbf{k}$ un indicateur positif $g_{\mathbf{l}}$ prenant en compte à la fois la correction apportée par le multi-indice $\mathbf{l}$ à la valeur de l'intégrale recherchée, c'est à dire $\Delta_{\mathbf{l}}(f)$, et le coût de calcul associé au calcul de $\Delta_{\mathbf{l}}(f)$, c'est à dire le nombre de nouvelles évaluations de la fonction à intégrer nécessaires à ce calcul. Enfin, connaissant l'ensemble des indicateurs $g_{\mathbf{l}}$ pour $\mathbf{l}$ dans $\mathcal{O}$, on peut calculer un indicateur de convergence global η dépendant de ces $g_{\mathbf{l}}$, qui servira de critère d'arrêt. L'algorithme adaptatif est alors le suivant :

- Initialiser $\mathcal{I}$ avec l'élément $(0, \ldots, 0)$ et assigner $I = \Delta_{(0,\ldots,0)}(f)$.
- Construire l'ensemble des multi-indices admissibles $\mathcal{O}$.
- Calculer l'ensemble des $g_{\mathbf{l}}$ pour $\mathbf{l} \in \mathcal{O}$.
- Calculer l'indicateur de convergence global $\eta(\{g_{\mathbf{l}} : \mathbf{l} \in \mathcal{O}\})$.
- Tant que η est plus grand que le critère d'arrêt ϵ :
  1. Choisir le multi-indice admissible $\mathbf{l_{max}}$ dans $\mathcal{O}$ ayant le plus grand indicateur $g_{\mathbf{l_{max}}}$
  2. Ajouter $\mathbf{l_{max}}$ à $\mathcal{I}$ et mettre à jour $I : I \leftarrow I + \Delta_{\mathbf{l_{max}}}(f)$
  3. Ajouter les nouveaux multi-indices admissibles à $\mathcal{O}$, et calculer les indicateurs g associés.
  4. Calculer le nouvel indicateur de convergence global $\eta(\{g_{\mathbf{l}} : \mathbf{l} \in \mathcal{O}\})$.
- Renvoyer I

Il est important de noter que le critère de convergence utilisé ici ne permet pas d'estimer l'erreur commise entre l'évaluation numérique de l'intégrale et la vraie valeur de cette intégrale. Il estime simplement le progrès qui pourrait être fait à la prochaine itération, et juge qu'il y a convergence lorsque celui-ci est suffisamment petit, c'est à dire que l'évaluation numérique de l'intégrale ne change pas significativement d'une itération à l'autre.

2.3 Comparaison des méthodes de cubatures

Comme dans le cas des fonctions d'une variable, les fonctions utilisées pour comparer les méthodes de cubature sont également issues de [40], et sont toutes définies sur $[0, 1]^d$, où d est le nombre de variables considéré. Toutes les méthodes de cubature introduites dans cette section étant définies sur $[-1, 1]^d$,

une transformation affine a permis de se ramener à $[0,1]^d$. Ces fonctions test sont présentées dans l'annexe A.

Les nombres réels u_i ont été choisis aléatoirement entre 0 et 1, et de même pour les nombres a_i représentant la difficulté d'intégration, à la différence que leur somme est ensuite ramenée à 1. Ce choix aléatoire permet d'assurer que les fonctions tests ne présenteront pas des comportements identiques dans deux directions différentes, et impose donc une certaine anisotropie à celles-ci. Bien entendu, les mêmes valeurs de u_i et de a_i sont considérées pour comparer les différentes méthodes entre elles. Pour l'ensemble de ces fonctions, la valeur exacte de l'intégrale est accessible analytiquement, ce qui permet de calculer l'erreur commise par l'approximation numérique. Sur la figure 2.18 sont comparées la tensorisation creuse de Smolyak et la tensorisation pleine pour la méthode des trapèzes ainsi que la seconde méthode de Fejér pour une dimension $d = 3$ grâce à l'erreur relative commise en fonction du nombre d'évaluations de la fonction nécessaires N. Pour la méthode des trapèzes, la tensorisation creuse de Smolyak offre une convergence plus rapide comparée à la tensorisation pleine pour l'ensemble des fonctions testées. Concernant la seconde méthode de Fejér, celle-ci offre également une convergence plus rapide pour l'ensemble des fonctions testées, exceptées la fonction "Corner Peak", ce qui peut s'expliquer par le fait que la tensorisation creuse de Smolyak a tendance à concentrer les points au centre de l'hypercube $[-1,1]^d$, en plaçant moins dans les coins que la tensorisation pleine alors que l'information de cette fonction est située dans un coin. Néanmoins, l'avantage de la tensorisation creuse de Smolyak sur la tensorisation pleine est visible, et est d'autant plus important en pratique que la dimension d est grande et que l'évaluation de la fonction a un coût important, puisqu'il est alors beaucoup moins coûteux de passer d'un niveau k au suivant.

Sur la figure 2.19 sont tracées les valeurs absolues des erreurs relatives E_{rel} commises par chaque méthode et pour chacune des fonctions présentées dans l'annexe A en fonction du nombre d'évaluations de la fonction nécessaire N pour une tensorisation creuse de Smolyak et pour une valeur de la dimension d égale à 3.

Pour les fonctions régulières (Oscillatory, Product Peak, Corner Peak et Gaussian), la méthode de Clenshaw-Curtis et la seconde méthode de Fejér présente des convergences similaires, suivies de près par la méthode de Romberg puis la méthode des trapèzes qui offre une convergence plus lente comme attendu. Un avantage de la méthode de Fejér est que remplir les premiers niveaux demande un nombre moins important de points que pour les autres méthodes, permettant d'avoir des résultats même avec un faible nombre de points. Pour les fonctions présentant des irrégularités, les méthodes présentent des vitesses de convergences similaires, qui sont d'autant plus faibles que l'irrégularité est importante. La seconde méthode de Fejér semble cependant légèrement meilleure que les autres pour le type de fonction "Continuous", alors que pour le type de fonction "Discontinuous", aucune méthode ne semble meilleure qu'une autre.

Sur la figure 2.20 sont tracées les valeurs absolues des erreurs relatives E_{rel}

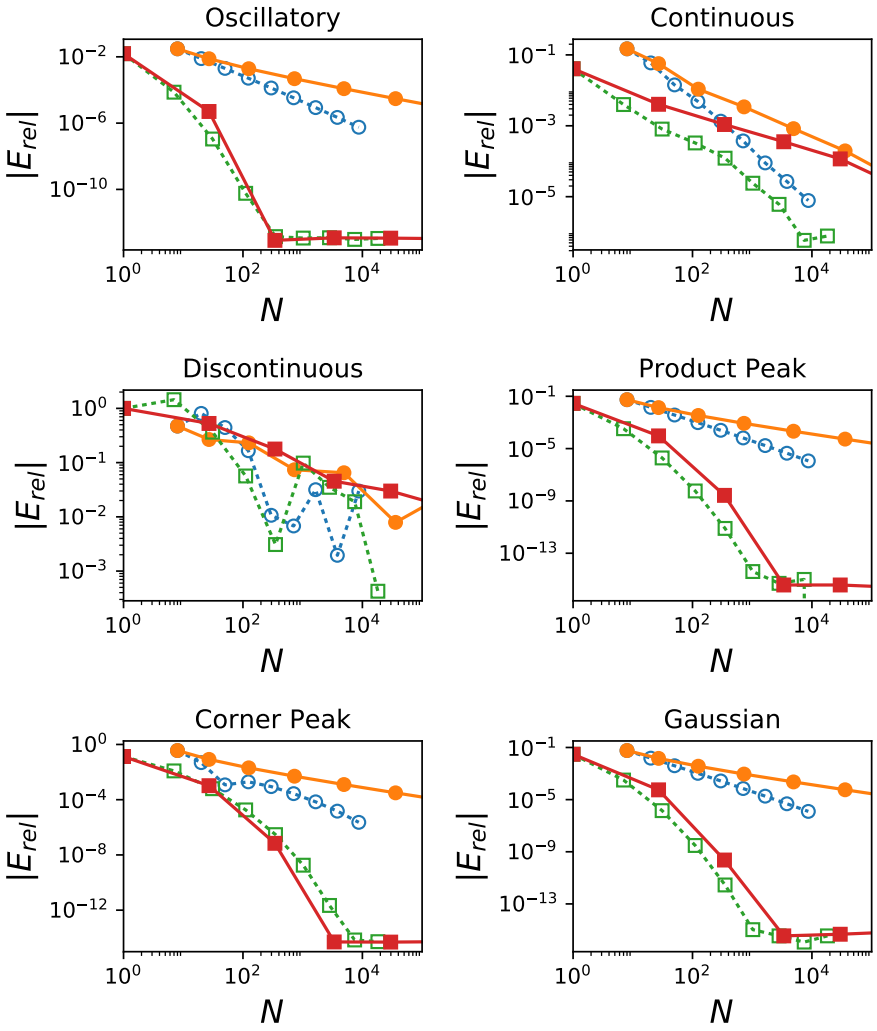

FIGURE 2.18 – *Valeur absolue de l'erreur relative E_{rel} obtenue par la tensorisation creuse de Smolyak (ligne pointillés et symboles creux) et la tensorisation pleine (ligne pleine et symboles pleins) pour les fonctions de test présentées dans l'annexe A pour $d = 3$, en fonction du nombre d'évaluations nécessaires N. Les quadratures imbriquées utilisées sont la méthode des trapèzes (cercles) ainsi que la seconde méthode de Fejér (carrés).*

commises pour les dimensions d de 2 à 5 pour chacune des fonctions présentées dans l'annexe A en fonction du nombre d'évaluations de la fonction nécessaire N pour une tensorisation creuse de Smolyak utilisée avec la seconde méthode de Fejér.

L'augmentation de la dimension d impacte la convergence comme attendu, à savoir que le nombre d'évaluations N augmente avec celle-ci. Cet effet de la

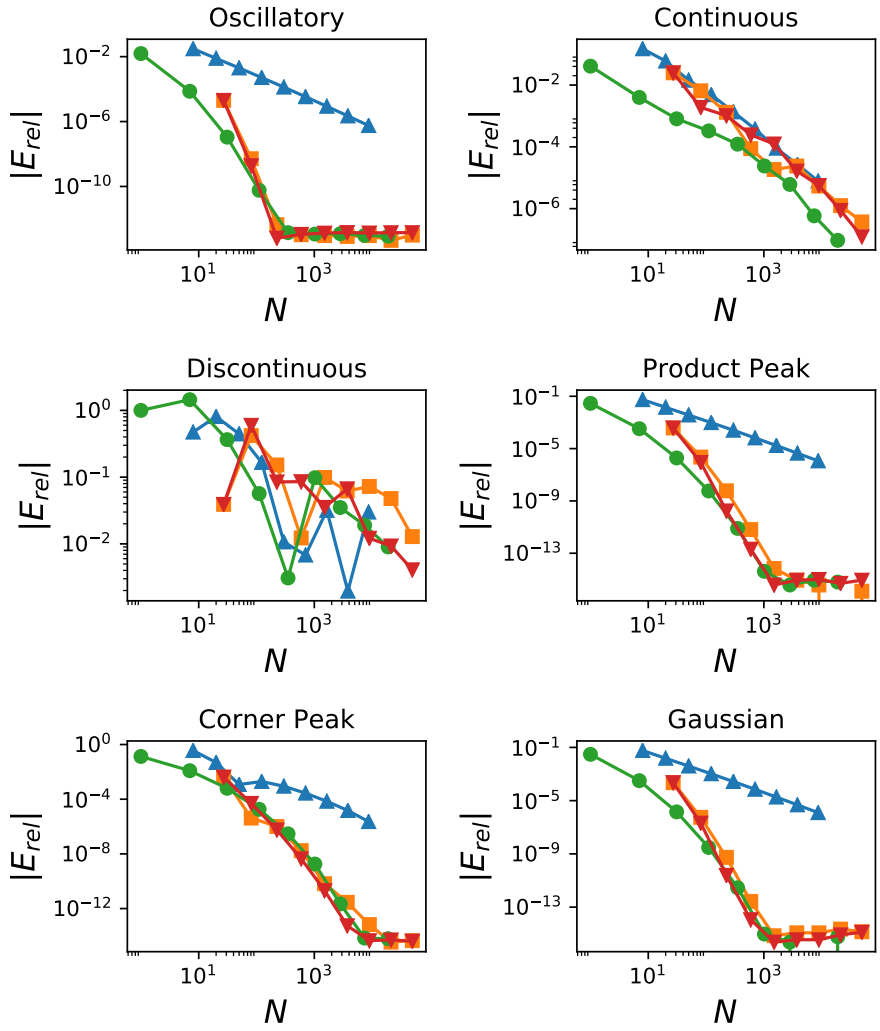

FIGURE 2.19 – *Valeur absolue de l'erreur relative E_{rel} obtenue par la tensorisation creuse de Smolyak pour les fonctions de test présentées dans l'annexe A pour $d = 3$, en fonction du nombre d'évaluations nécessaires N. Les quadratures imbriquées utilisées sont la méthode des trapèzes (triangles hauts), la méthode de Romberg (carrés), la méthode de Clenshaw-Curtis (triangles bas) et la seconde méthode de Fejér (ronds).*

dimension est présent que la tensorisation considérée soit creuse ou pleine, la convergence étant globalement la même pour ces deux modes de tensorisation comme le suggère la figure 2.18. Cependant, les résultats présentées sur la figure 2.21, présentant le nombre de points d'évaluations nécessaires N en fonction du niveau k, montrent que la tensorisation creuse de Smolyak offre l'avantage de pouvoir passer d'un niveau k au suivant pour un coût raisonnable comparé à la tensorisation pleine, et ce d'autant plus que la dimension d est

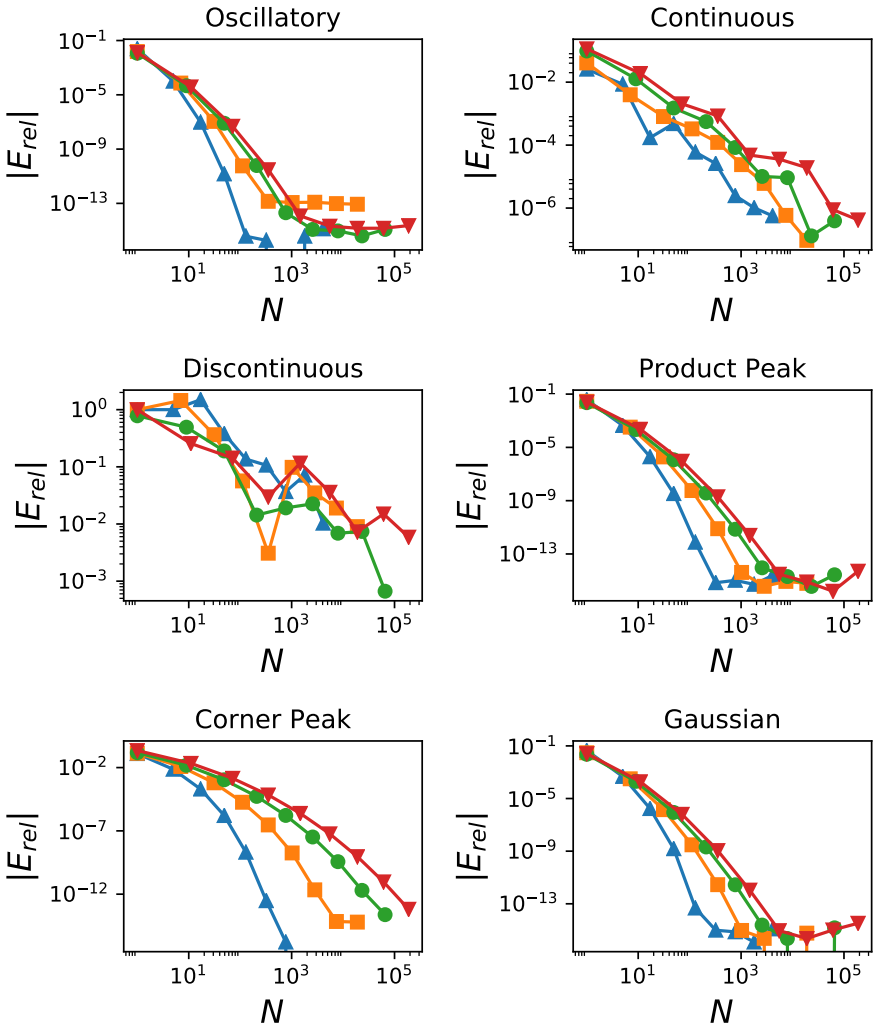

FIGURE 2.20 – *Valeur absolue de l'erreur relative E_{rel} obtenue par la tensorisation creuse de Smolyak pour les fonctions de test présentées dans l'annexe A pour la seconde méthode de Fejér pour différentes dimensions d, en fonction du nombre d'évaluations nécessaires N. Les dimensions d considérées sont $d = 2$ (triangles hauts), $d = 3$ (carrés), $d = 4$ (triangles bas) et $d = 5$ (ronds).*

importante. L'utilisation de la seconde méthode de Fejér avec une tensorisation creuse de Smolyak présente l'avantage sur les autres méthodes d'avoir un nombre de points d'évaluations N plus raisonnable pour les premiers niveaux, et cela d'autant plus que la dimension d est importante.

Sur la figure 2.22 sont comparées les convergences de la valeur absolue de l'erreur relative en fonction du nombre de points d'évaluations N, pour l'ensemble des fonctions test, pour la tensorisation creuse de Smolyak et l'al-

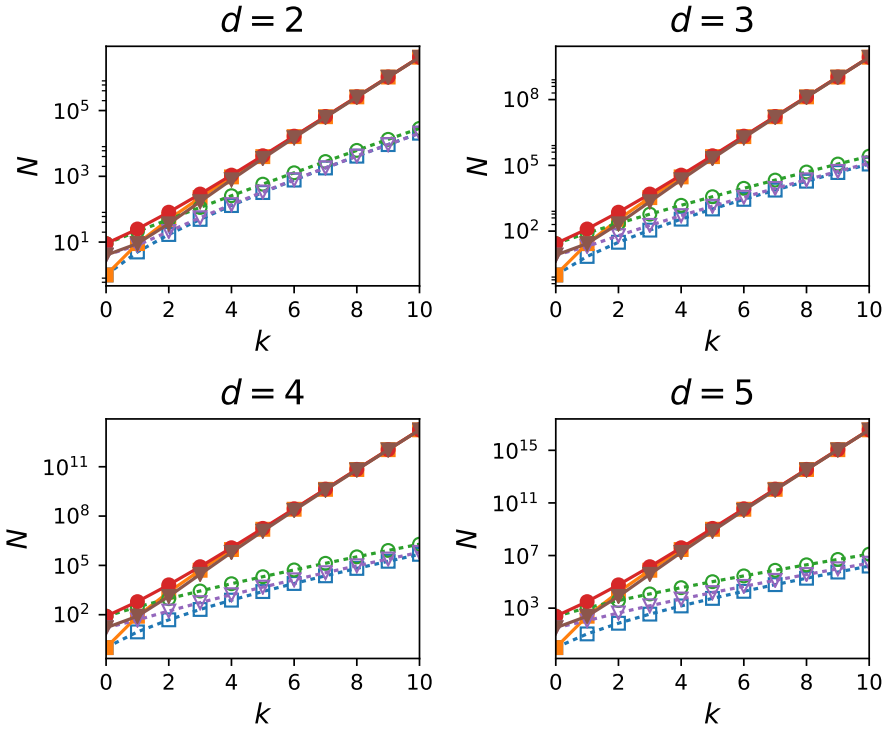

FIGURE 2.21 – *Nombres de points utilisées par la tensorisation creuse de Smolyak (ligne pointillé et symbole vide) et la tensorisation pleine (ligne pleine et symbole plein) utilisées avec différentes méthodes de quadratures : seconde méthode de Fejér (carrés), méthode de Clenshaw-Curtis (ronds) et méthode de Romberg (triangles), en fonction du degré de liberté k, pour une dimension d allant de 2 à 5.*

gorithme adaptatif, la seconde méthode de Fejér étant utilisée dans les deux cas et la dimension d étant de 2. Le choix du multi-indice à ajouter à $\mathcal{I}$ ici est le multi-indice qui offre la plus grande variation de l'approximation I parmi les multi-indices admissibles. L'algorithme adaptatif offre pour l'ensemble des fonctions régulières une convergence plus rapide que la tensorisation creuse de Smolyak, mais n'offre pas d'avantage (voire une convergence plus lente pour la fonction "continuous") pour les fonctions test présentant des singularités. Un avantage toujours présent avec l'algorithme adaptatif est qu'il ne nécessite pas le remplissage total des niveaux k comme la tensorisation creuse de Smolyak, ce qui permet de l'arrêter "plus souvent". Cet avantage est d'autant plus intéressant que la dimension d est importante, les niveaux k devenant de plus en plus coûteux à remplir.

Sur la figure 2.23 est présenté l'ensemble $\mathcal{I}$ construit après convergence de l'algorithme adaptatif pour la seconde méthode de Fejér sur la fonction Continuous en dimension 2. Le motif obtenu n'est pas un triangle comme ce serait le cas avec une tensorisation creuse de Smolyak, mais présente une forme

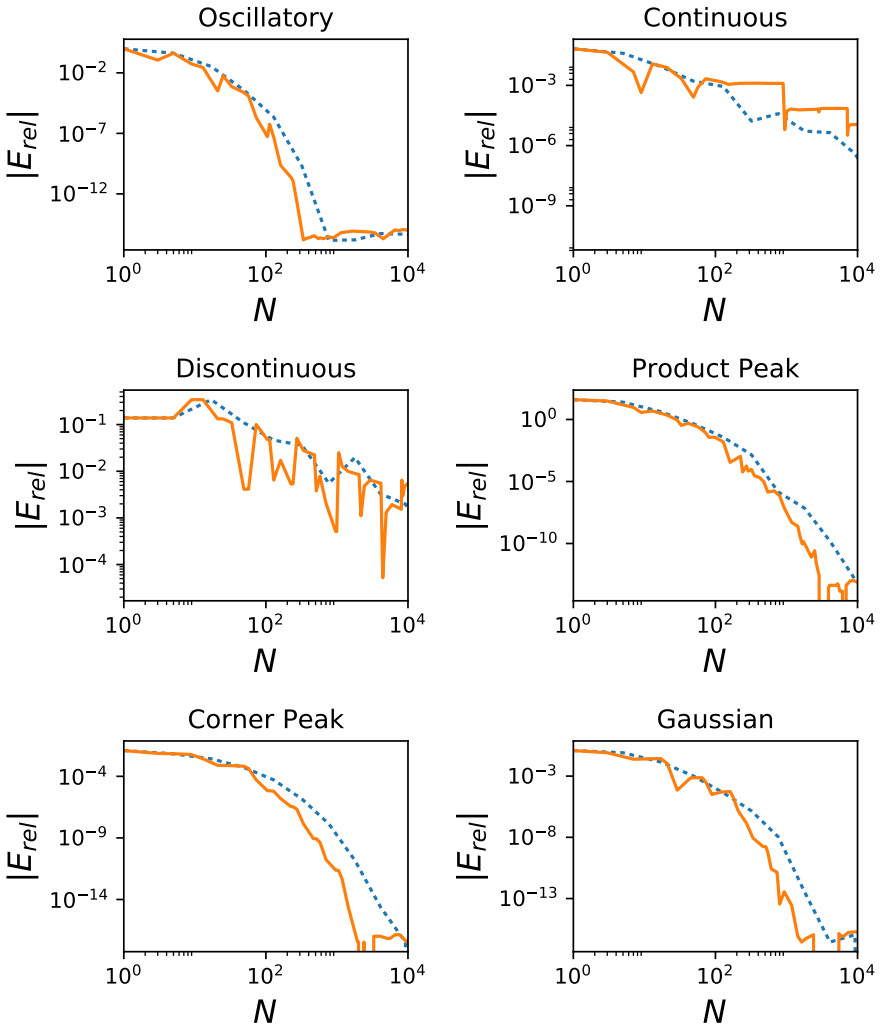

FIGURE 2.22 – *Valeur absolue de l'erreur relative E_{rel} obtenue par une tensorisation creuse de Smolyak (ligne pointillé) et obtenue par l'algorithme adaptatif (ligne pleine), les deux utilisant la seconde quadrature de Fejér, pour les fonctions de test présentées dans l'annexe A pour $d = 2$, en fonction du nombre d'évaluations nécessaires N.*

propre qui provient de la fonction à intégrer qui n'est pas isotrope.

Un algorithme adaptatif offre donc l'avantage de ne prendre en compte que les multi-indices qui vont significativement contribuer à l'approximation numérique de l'intégrale, et d'omettre ceux qui n'auraient eu que très peu d'impact, diminuant ainsi le coût de calcul. Il y a de plus une estimation de l'erreur d'intégration commise à chaque itération de l'algorithme, qui n'est cependant pas toujours représentatif de la véritable erreur commise.

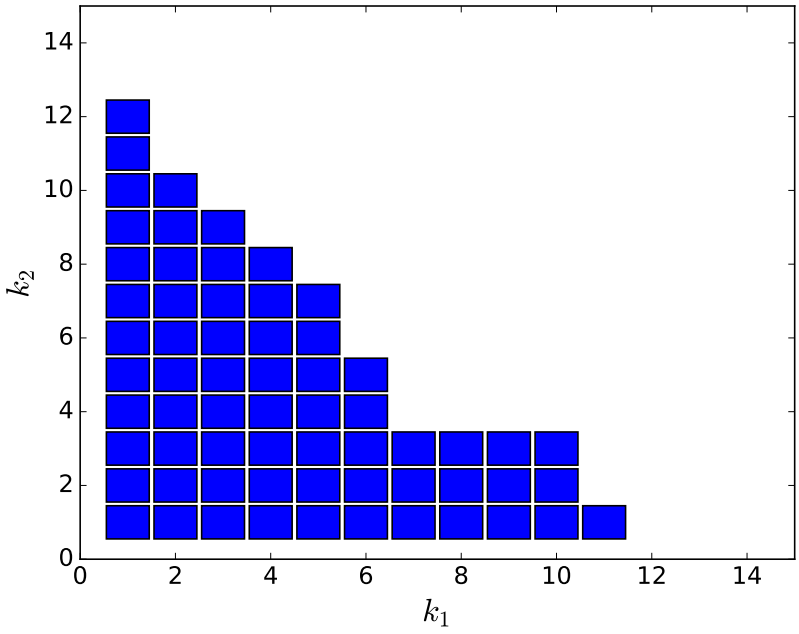

FIGURE 2.23 – *Multi-indices* $\mathbf{k} = (k_1, k_2)$ *présents pour le calcul de la fonction nommée Continuous dans l'annexe A en dimension* 2 *avec la seconde méthode de Fejér, avec une précision* ϵ *demandée de* 10^{-5}.

2.4 Conclusion

Dans ce chapitre ont été présentées des méthodes de quadrature permettant d'approximer numériquement des intégrales de fonctions d'une variable réelle. Différentes méthodes permettent d'approximer numériquement une intégrale, qui reposent toutes sur un nombre plus ou moins important d'évaluations de la fonction à intégrer. Pour des fonctions coûteuses à évaluer, tout l'enjeu est donc d'avoir la meilleure approximation possible avec un nombre minimum d'évaluations. Une approche simple utilisée par les méthodes de Newton-Cotes consiste à approcher la fonction par un polynôme interpolateur par morceaux et à approcher l'intégrale par la valeur de l'intégrale du polynôme par morceaux que l'on sait calculer exactement. La rapidité de la convergence de ces méthodes augmente à priori avec le degré utilisé pour le polynôme interpolateur, mais le phénomène de Runge ne permet en fait pas d'utiliser des degrés trop importants, qui en pratique se traduit par l'utilisation d'un degré des polynômes interpolateurs souvent inférieur à 5. Des méthodes plus efficaces existent cependant telles les méthodes de Gauss, Clenshaw-Curtis ou Fejér. Toutes ces méthodes offrent des convergences similaires, bien que les méthodes de Gauss restent les plus efficaces. Néanmoins, ces méthodes requièrent une régularité suffisante de la fonction à intégrer, et en cas de fonctions présentant une simple

singularité en un point, aucune méthode ne semble meilleure que les autres.

Il est possible d'utiliser les méthodes de quadrature de fonctions d'une variable réelle pour construire des méthodes numérique d'intégration pour des fonctions de plus d'une variable, par tensorisation des méthodes en 1 dimension. Le "fléau de la dimension" limite toutefois l'utilisation de la tensorisation à des fonctions ayant un nombre de variables raisonnable, typiquement de l'ordre de 5 au plus. Il est cependant possible de nuancer le "fléau de la dimension" par l'utilisation de tensorisation creuse, introduite par Smolyak pour la première fois, qui permettent d'avoir une croissance du nombre de points d'évaluation moins importante que dans le cas d'une simple tensorisation. Cette tensorisation creuse est idéalement à utiliser avec des méthodes de quadrature dites imbriquées, qui permettent de réutiliser des évaluations de la fonction plusieurs fois, plutôt que d'avoir des nouvelles évaluations à calculer. Parmi les méthodes de quadratures imbriquées, la méthode de Clenshaw-Curtis et la seconde méthode de Fejér en particulier offrent d'excellentes performances. Un algorithme adaptatif avec une estimation de l'erreur commise est également possible, qui permet encore un peu plus de limiter le nombre d'évaluations de la fonction et donc le coût de calcul associé à l'estimation numérique de l'intégrale.

Chapitre 3

Méthodes probabilistes et quasi-probabilistes pour le calcul numérique d'intégrale en grande dimension

Notions clés et apports du chapitre :

- Génération de variables et de vecteurs aléatoires de lois quelconques à partir de variables aléatoires uniformes et indépendantes sur $[0, 1]$
- Calcul d'intégrales en dimension quelconque avec estimation de l'erreur grâce à la méthode de Monte Carlo
- Suites à discrépance faible et méthode de Quasi-Monte Carlo pour le calcul d'intégrale
- Randomisation des méthodes de Quasi-Monte Carlo pour l'estimation de l'erreur
- Comparaison des méthodes de randomisation de la méthode de Quasi-Monte Carlo
- Comparaison des méthodes de Monte Carlo et de Quasi-Monte Carlo randomisé
- Méthodes de ré-échantillonnage (Jackknife et Bootstrap) pour l'estimation d'erreur sur des grandeurs statistiques "complexes", notamment dans le cas de l'utilisation de la méthode de Quasi-Monte Carlo randomisé

Dans le chapitre précédent, des méthodes ont été présentées pour l'approximation numérique d'intégrale de fonctions de plusieurs variables, avec cependant un nombre de variables réduit, typiquement plus petit que 5. Ces méthodes nécessitaient l'évaluation de la fonction à intégrer en un nombre plus ou moins important de points, ce nombre de points augmentant exponentiellement avec le nombre de variables de la fonction à intégrer. Ces méthodes devenaient ainsi vite trop coûteuses pour un grand nombre de variables. Les méthodes présentées dans ce chapitre pour approximer l'intégrale d'une fonction nécessitent également l'évaluation de la fonction à intégrer en certains points, mais diffèrent des méthodes présentées dans le chapitre précédent par le choix des points d'évaluations de la fonction. Ce choix différent permet d'avoir une convergence de l'approximation de l'intégrale qui, bien que moins rapide, est indépendant ou peu dépendant selon les méthodes, du nombre de variables dont dépend la fonction à intégrer.

3.1 Introduction

Les méthodes d'intégration numérique présentées dans ce chapitre visent à estimer une intégrale de la forme de celle de l'expression (1.3) à l'aide d'une expression de la forme :

$$\int_{\mathbb{R}^d} g(\mathbf{x})\pi_{\mathbf{X}}(\mathbf{x})d\mathbf{x} \approx \frac{1}{n}\sum_{i=1}^{n} g(\mathbf{x}_i) \tag{3.1}$$

Tout comme dans le cas des méthodes déterministes présentées au chapitre précédent, les méthodes consistent en l'évaluation de la fonction d'intérêt en un grand nombre de points pour ensuite approximer l'intégrale comme une somme pondérée de ces valeurs.

Le choix des points d'évaluations $(\mathbf{x}_i)_{1\leq i\leq n}$ dépend de la méthode utilisée. Dans le cas de la méthode de Monte Carlo, ces points correspondent à la réalisation d'un échantillon de variables aléatoires indépendantes entre elles alors que dans le cas de méthodes de Quasi-Monte Carlo, ces points possèdent une propriété de remplissage de l'espace permettant d'avoir une erreur d'évaluation de l'intégrale faible. Dans les deux cas, les points $(\mathbf{x}_i)_{1\leq i\leq n}$ sont construits à partir d'une séquence $(u_i)_{1\leq i\leq m}$ de points de $[0,1]$, qui ne possèdent pas les mêmes propriétés selon que la méthode utilisée est une méthode de Monte Carlo ou de Quasi-Monte Carlo.

Ce chapitre comprend quatre sections. La première présente les méthodes permettant de passer de séquences de points $(u_i)_{1\leq i\leq m}$ de $[0,1]$ aux points d'évaluations $(\mathbf{x}_i)_{1\leq i\leq n}$ ainsi qu'une description de l'implémentation parallèle mise en œuvre pour l'évaluation de l'intégrale d'une fonction g devant être évaluer un grand nombre de fois et pouvant être coûteuse à évaluer. La seconde section présente les résultats théoriques sur lesquels la méthode de Monte Carlo

repose ainsi que différentes méthodes permettant d'accélérer celle-ci. La troisième section présente les méthodes de Quasi-Monte Carlo, et plus particulièrement les méthodes de Quasi-Monte Carlo randomisée, ainsi que des méthodes par ré-échantillonnage permettant d'évaluer les erreurs commises. Enfin, la quatrième section présente une comparaison des différentes méthodes en utilisant les fonctions test de Genz présentées dans annexe A.

3.2 Génération de variables et de vecteurs aléatoires

Comme il sera expliqué dans les prochaines sections, les méthode de Monte Carlo et de Quasi-Monte Carlo randomisée repose pour le calcul d'une intégrale de la forme de l'expression (1.3) sur l'utilisation d'un échantillon $(\mathbf{X}_1, \dots, \mathbf{X}_n)$ de vecteurs aléatoires. Dans le cas de la méthode de Monte Carlo, ces vecteurs aléatoires sont indépendants entre eux, alors que pour les méthodes de Quasi-Monte Carlo randomisée, des dépendances existent entre eux. Dans les deux cas, cet échantillon est construit à partir d'un échantillon $(U_1, \dots, U_m)$ de variables aléatoires de $[0, 1]$, présentant ou non des dépendances selon que l'on s'intéresse à une méthode de Monte Carlo ou de Quasi-Monte Carlo randomisée.

Dans l'ensemble des cas, il est nécessaire d'avoir une source d'aléatoire (ou de pseudo-aléatoire) qui se fait à travers l'utilisation d'un générateur de nombres pseudo-aléatoires (PRNG pour Pseudo Random Number Generator), qui permet de générer une séquence de nombres mimant une réalisation d'un échantillon de variables aléatoire indépendantes et uniformes sur $[0, 1]$. Formellement, cela revient à considérer que le générateur de nombre pseudo-aléatoire renvoie une réalisation $(U_1(\omega), \dots, U_n(\omega))$ avec $\omega \in \Omega$, où $(U_1, \dots, U_n)$ est un échantillon de variables aléatoires réelles définies sur l'espace probabilisé $(\Omega, \mathcal{A}, P)$ qui sont indépendantes entre elles et suivent toutes une loi uniforme sur $[0, 1]$. Une fois cette séquence de nombres construite, celle-ci est utilisée dans le cas de la méthode de Monte Carlo pour construire une réalisation d'une famille de vecteurs aléatoires indépendants suivant la loi désirée, permettant d'accéder à des réalisations indépendantes de la quantité d'intérêt utilisées dans la méthode de Monte Carlo. Dans la méthode de Quasi-Monte Carlo randomisée, cette séquence de nombre permet de générer une réalisation $(\mathbf{U}_1(\omega), \dots, \mathbf{U}_n(\omega))$ avec $\omega \in \Omega$ d'un échantillon $(\mathbf{U}_1, \dots, \mathbf{U}_n)$ de vecteurs aléatoires possédant des dépendances entre eux, permettant d'accéder à des réalisations dépendantes de la quantité d'intérêt utilisées dans la méthode de Quasi-Monte Carlo randomisé.

3.2.1 Génération de nombres pseudo-aléatoires

3.2.1.1 Généralités sur les générateurs de nombres pseudo-aléatoires

Les générateurs de nombres aléatoires sont des algorithmes permettant de construire des séquences de nombres entre 0 et 1 avec pour propriété qu'une séquence de N nombres construite par un tel algorithme peut être vue comme

la réalisation d'un échantillon $(U_i)_{1\leq i\leq n}$ de variables aléatoires indépendantes et de loi uniforme sur $[0,1]$. Un des avantages de tels algorithmes est qu'ils permettent la répétabilité d'une expérience numérique, puisqu'une même séquence pourra être à nouveau générée, ce qui est également en pratique indispensable pour le débogage d'un code. Comme expliqué dans [66], un générateur de nombre aléatoire peut être formellement décrit comme une structure (S, T, τ, ξ, x_0) avec :

$$\begin{aligned} S &= \text{Espace d'états} \\ T &= \text{Espace de sortie} \\ \tau : S \to S &= \text{Fonction de transition} \\ \xi : S \to T &= \text{Fonction de sortie} \\ x_0 \in S &= \text{Graine} \end{aligned}$$

Une telle structure vise à être implémentée, et l'espace d'état S est donc un ensemble fini. La finitude de S implique que l'application successive de la fonction de transition τ à partir de la graine x_0 finit par renvoyer x_0. Étant donné τ et x_0, il est donc possible de définir une période ρ, qui correspondra à la période de notre générateur de nombres aléatoires. Pour des raisons d'implémentations, l'espace de sortie T est un sous ensemble de $[0,1]$, et contient des nombres flottants qui doivent être représentables en machine, et est donc également fini. La fonction ξ est simplement une fonction permettant de passer d'un état s de S à un nombre u entre 0 et 1 de T. Cette fonction peut ne pas être injective, l'espace S étant généralement plus grand que l'espace T.

La construction de telles séquences de nombres par ces algorithmes n'a rien d'aléatoire, et est purement déterministe puisque les fonctions τ et ξ sont déterministes. L'important est que ces séquences se comportent statistiquement de manière similaires à une vraie réalisation d'une famille de variables aléatoires *i.i.d.*, et cela est assuré par la réalisation de test statistiques sur ces séquences, testant l'uniformité de la distribution ou encore l'indépendance par exemple.

Un mauvais générateur aléatoire ou un générateur aléatoire utilisé dans de mauvaises conditions pourra produire des résultats erronés car la séquence utilisée ne mimera plus un comportement de variables aléatoires indépendantes et uniformes sur $[0,1]$. Par exemple, la présence d'une période ρ implique qu'une séquence de taille supérieure à ρ ne sera pas aléatoire, car présentant un cycle. En fait, certaines études [68, 70] ont montré que le rapport entre la longueur n de la séquence utilisée et la racine carré de la période ρ du générateur doit rester petit. Un bon générateur pour une application donnée doit donc avoir une période ρ supérieure au carré du nombre n de nombres nécessaires pour l'expérience numérique envisagée, ce qui exclut bien souvent certains générateurs aléatoires dont la période n'est que de $2^{32}-1$ par exemple. Un autre problème lors de l'exécution parallèle en mémoire distribuée d'un code de Monte Carlo, est le choix de la graine à utiliser pour chacun des processus. En effet, choisir la même graine pour chacun des processus entraînerait que les mêmes séquences

seraient générées pour chacun des processus, et l'indépendance serait donc perdue. Un bon choix de la graine doit donc être effectué pour chacun des processus afin qu'aucun recouvrement des séquences des différents processus entre elles ne soit présent. Pour s'assurer de cela, il faut une bonne connaissance de la fonction de transition τ, ainsi qu'une idée de la taille des séquences pour chacun des processus. En fait, une propriété intéressante de la fonction de transition τ permet de ne pas avoir à faire un tel choix. S'il est possible de faire un saut "rapide" dans l'espace d'état S, alors il est possible de s'arranger pour que l'ensemble des processus génère des séquences de nombres aléatoires qui mises bout à bout reconstituent une seule séquence. Ce procédé est décrit ultérieurement, et son implémentation a été réalisé en MPI. Il est important de noter qu'en parallélisant ainsi la génération de séquences de nombres, cela revient à générer globalement une seule séquence contiguë, plutôt qu'une séquence différente pour chacun des processus. Les tests des générateurs de nombres aléatoires sont généralement effectués sur une séquence, et non sur deux séquences choisies au "hasard" et mises bout à bout, et la parallélisation présentée permet de s'assurer que l'utilisation du générateur de nombres aléatoires se fait dans les mêmes conditions que les tests qui ont permis de valider statistiquement ce générateur. Un inconvénient reste cependant que pour une telle implémentation, la taille des sous-séquences générées par les processus doit être fixée, et les sauts seront nécessairement des multiples de cette taille.

Plusieurs générateurs de nombres aléatoires existent et sont présentés dans la littérature. Un générateur simple à implémenter, et possédant les propriétés précédentes et pour lequel de nombreux tests statistiques ont été passés avec succès [67] est le générateur MRG32k3a de L'écuyer. L'ensemble des simulations impliquant la méthode de Monte Carlo et de Quasi-Monte Carlo randomisé dans cette thèse ont été réalisées à l'aide de ce générateur de nombres aléatoires.

3.2.1.2 Génération de nombres aléatoires en parallèle

3.2.1.2.1 Implémentation séquentielle/parallèle de Monte-Carlo Les méthodes de Monte Carlo et de Quasi-Monte Carlo randomisé sont basées sur l'évaluation de la fonction d'intérêt en un grand nombre de points choisis au sein d'une séquence, aléatoire ou non. Lorsque la fonction d'intérêt est coûteuse à évaluer, l'évaluation de celle-ci représente la plus grande partie du temps de calcul, et l'occupation du processus au cours du temps pour une implémentation séquentielle a l'allure présentée figure 3.1.

Il est intéressant de paralléliser le travail d'évaluation de la fonction d'intérêt sur des machines massivement parallèles. Il est aisé de réaliser une implémentation parallèle, impliquant un processus maître et des processus esclaves. Les processus esclaves réalisent les évaluations de la fonction d'intérêt, et envoient ensuite leurs résultats au processus maître, qui s'occupe de calculer la valeur des estimateurs et des intervalles de confiance, afin de pouvoir décider quand le calcul est fini. Dans une telle implémentation, les processus esclaves

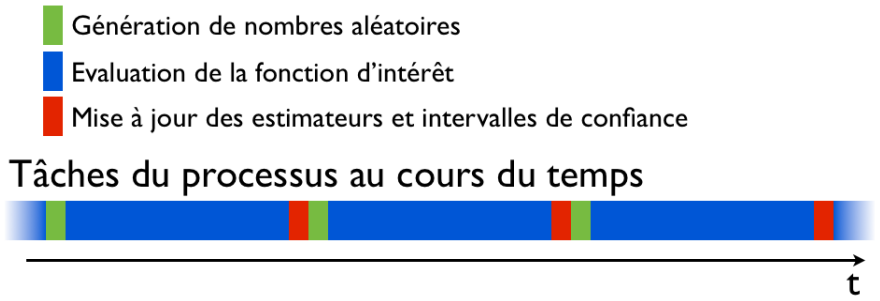

FIGURE 3.1 – *Occupation du temps par le processus pour une méthode de Monte-Carlo séquentielle*

ne communiquent pas entre eux, et communiquent seulement avec le processus maître, qui lui, communique avec tous les processus esclaves. L'occupation des différents processus au cours du temps a l'allure présentée figure 3.2 si la génération des nombres de la séquence se fait du côté des processus esclaves, ou l'allure présentée figure 3.3 si la génération des nombres de la séquence se fait du côté du processus maître.

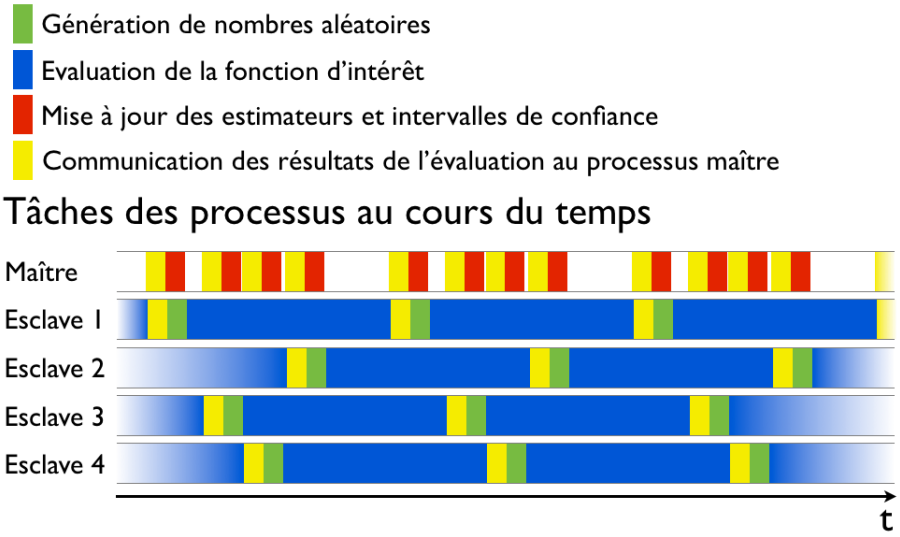

FIGURE 3.2 – *Occupation du temps par les différents processus pour une méthode de Monte-Carlo parallèle, avec la génération des nombres de la séquence côté processus esclaves*

L'illustration des figures 3.2 et 3.3 montre qu'il est préférable que ce soit le processus maître qui soit en attente des processus esclaves plutôt que le contraire, puisqu'il est alors le seul en attente contre plusieurs processus en attente, et cela afin que l'occupation générale des processus soit la plus importante possible. Il est donc important que le processus maître soit le moins

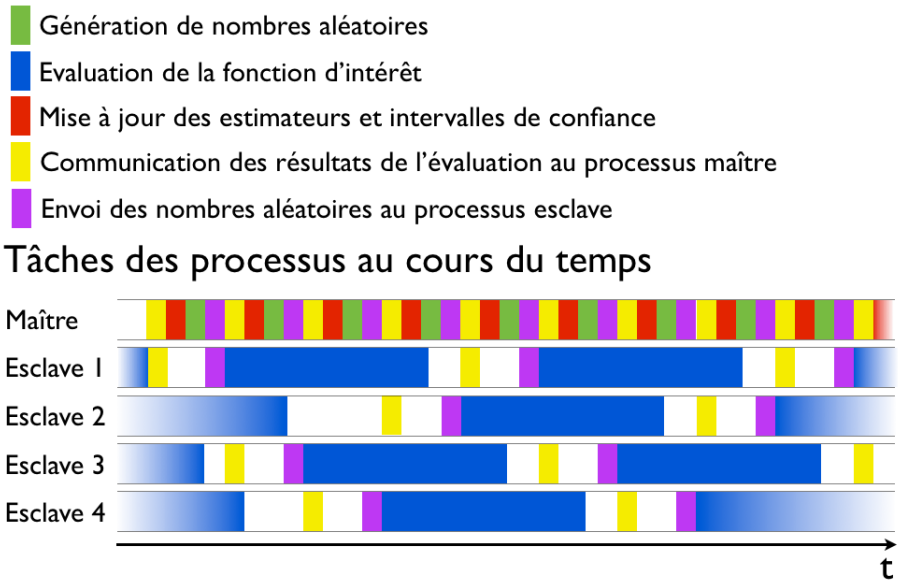

FIGURE 3.3 – *Occupation du temps par les différents processus pour une méthode de Monte-Carlo parallèle, avec la génération des nombres de la séquence côté processus maître*

chargé possible, et que les communications avec celui-ci soient réduites au minimum, et pour cela, les nombres de la séquence doivent être générés du côté des processus esclaves.

Il est donc nécessaire de paralléliser la génération des nombres de la séquence, afin d'avoir l'implémentation parallèle la plus efficace.

3.2.1.2.2 Générateur de nombres pseudo-aléatoires

3.2.1.2.2.1 Générateur MRG32k3a Le générateur MRG32k3a [67] est un générateur de nombres pseudo-aléatoires de type CMRG (Combined Multile Recursive Generator). Il a passé de nombreux tests théoriques et statistiques, et il est considéré comme un bon générateur pour réaliser des simulations au vu des connaissances et des moyens de calcul actuels, reproduisant une réalisation de variables aléatoires i.i.d. de loi uniforme sur $[0, 1]$. Il est caractérisé par une période longue au vu des moyens de calcul actuelle, qui est de l'ordre de 2^{191}. Ce générateur est défini par le jeu d'équations suivant :

$$\begin{cases} y_{1,n} = (a_{12}y_{1,n-2} + a_{13}y_{1,n-3}) \pmod{m_1}, \\ y_{1,n} = (a_{21}y_{2,n-1} + a_{23}y_{2,n-3}) \pmod{m_2}, \\ z_n = (y_{1,n} - y_{2,n}) \pmod{m_1}, \\ u_n = \frac{z_n}{m_1} \end{cases} \tag{3.2}$$

Les paramètres utilisés, qui assurent de bonnes propriétés au générateurs,

sont les suivants :

$$\begin{cases} a_{12} = 1403580, & a_{13} = -810728, & m_1 = 2^{32} - 209 \\ a_{12} = 527612, & a_{13} = -1370589, & m_2 = 2^{32} - 22853 \end{cases}$$

Les deux premières équations du système (3.2) peuvent également se mettre sous une forme matricielle, comme suit :

$$\begin{cases} Y_{1,n+1} = A_1 Y_{1,n} \pmod{m_1} \\ Y_{2,n+1} = A_2 Y_{2,n} \pmod{m_2} \end{cases} \tag{3.3}$$

Avec :

$$\begin{cases} A_1 = \begin{pmatrix} 0 & a_{12} & a_{13} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, & Y_{1,n} = \begin{pmatrix} y_{1,n} \\ y_{1,n-1} \\ y_{1,n-2} \end{pmatrix} \\ A_2 = \begin{pmatrix} a_{21} & 0 & a_{23} \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}, & Y_{2,n} = \begin{pmatrix} y_{2,n} \\ y_{2,n-1} \\ y_{2,n-2} \end{pmatrix} \end{cases}$$

Utiliser ce générateur revient à choisir une graine composée de deux vecteurs $Y_{1,-1}$ et $Y_{2,-1}$, et à ensuite appliquer le système d'équation (3.2) autant de fois que de nombres dans la séquence souhaités. La séquence de nombres renvoyée par le générateur ne dépendra alors que de la graine avec lequel il a été initialisé.

Les tests théoriques et statistiques réalisés lors de la validation de ce générateur assurent alors que quelque soit la graine choisie, la suite de nombres générées aura de bonnes propriétés, tant que sa taille reste raisonnable relativement à la période de ce générateur, ce qui sera le cas dans les simulations réalisées ici.

3.2.1.2.2.2 Difficulté de la parallélisation du générateur MRG32k3a

L'utilisation du générateur est par nature séquentielle, et elle est donc adaptée à une implémentation séquentielle sur un seul processus, à une implémentation parallèle du type de celle présentée sur la figure 3.3, et non à une implémentation parallèle sur plusieurs processus où chaque processus génère lui même ses nombres pseudo-aléatoires comme dans la figure 3.2. Or, l'implémentation parallèle où les processus esclaves génèrent leurs propres nombres aléatoires est potentiellement plus efficaces, car elle permet de décharger le processus maître de la génération de ces nombres pseudo-aléatoires ou quasi-aléatoires, et surtout à limiter les communications entre processus maître et esclaves, les nombres pseudo-aléatoires ou quasi-aléatoires générés n'ayant pas besoin d'être transmis aux processus esclaves.

Une première approche possible est d'initialiser les générateurs de chaque processus esclaves avec des graines différentes, afin qu'ils ne sortent pas les

mêmes séquences de nombres. Ce choix de graines doit être fait de telle manière que les séquences de nombres générées ne se recouvrent pas partiellement les unes des autres, étant toutes des séquences extraites de la séquence totale du générateur de longueur la période de celui-ci. En plus de ce problème, les tests statistiques réalisés pour valider le générateur ont été fait pour des séquences contiguës, et non pour des concaténations de séquences contiguës, comme c'est le cas lorsque les processus esclaves utilisent chacun leur séquence contiguë. On ne peut donc pas assurer que la concaténation des séquences de nombres générées par les esclaves aura des bonnes propriétés statistiques.

3.2.1.2.2.3 Bonne parallélisation du générateur MRG32k3a Pour éviter ces problèmes, il est possible de réaliser une implémentation parallèle similaire à une implémentation séquentielle sur un seul processus, c'est à dire que la concaténation de l'ensemble des séquences générées par les processus esclaves sera une séquence contiguë du générateur, dépendant uniquement de la graine ayant servie à l'initialisation. Afin de réaliser une telle implémentation, il faut simplement que le processus maître soit en mesure de savoir combien de nombres pseudo-aléatoires ont déjà été générées par l'ensemble des processus esclaves, ce qui est possible si par exemple les processus esclaves génèrent le même nombre de nombres pseudo-aléatoires entre deux communications avec le processus maître. Si cette condition est respectée, il suffit alors au processus maître de communiquer au processus esclave libre combien de nombres pseudo-aléatoires ont déjà été générés, afin que le processus esclave sache à partir de quel nombre de la séquence celui-ci doit générer les nombres aléatoires nécessaires pour l'évaluation de la fonction d'intérêt, tous les processus ayant utilisés la même graine pour l'initialisation de leur générateur.

Ainsi, il est nécessaire de savoir faire des sauts au sein de la séquence de nombres générée, et cela de manière efficace. Une manière naïve de faire un saut de taille n dans la séquence serait de générer les n nombres suivant, et de prendre le dernier, en jetant les $n-1$ autres. Cette méthode nécessite un nombre d'opérations arithmétique élémentaire proportionnel à n. Une méthode [15] basée sur l'exponentiation rapide permet de réduire ce coût, réalisant un saut de n nombres à l'aide d'un nombre d'opérations arithmétique élémentaire en $O(log(n))$.

Supposons que l'on soit dans l'état $Y_{i,n}$ pour $i \in [\![1,2]\!]$ ayant permis de générer le $n-ime$ nombre pseudo-aléatoire, et que l'on souhaite passer dans l'état $Y_{i,n+p}$. Une récurrence immédiate donne que :

$$\begin{cases} Y_{1,n+p} = A_1^p Y_{1,n} \pmod{m_1} \\ Y_{2,n+p} = A_2^p Y_{2,n} \pmod{m_2} \end{cases} \tag{3.4}$$

Il suffit donc de savoir calculer efficacement les puissances des matrices A_i. Pour cela, une technique d'exponentiation rapide peut être utilisée, qui aura un coût proportionnel à $log\,(p)$. Il est donc possible d'effectuer des sauts au sein

de la séquence à un coût raisonnable. L'allure de l'occupation du temps par les différents processus aura donc l'allure présentée sur la figure 3.4.

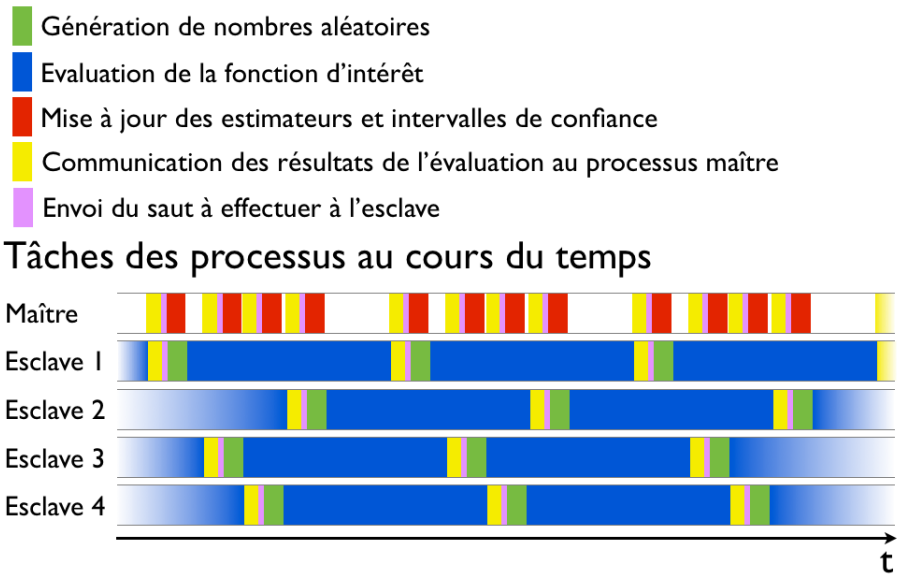

FIGURE 3.4 – *Occupation du temps par les différents processus pour une méthode de Monte-Carlo parallélisée de manière efficace et reposant sur le générateur MRG32k3a, pouvant s'étendre à une méthode de Quasi-Monte Carlo randomisée parallèle où les nombres quasi-aléatoires sont générés par les esclaves.*

Le schéma d'occupation du temps présenté par la figure 3.4 peut également se généraliser à une méthode de Quasi-Monte Carlo randomisée dans laquelle les points sont générés par les processus esclaves, qui est également l'implémentation qui a été utilisée dans cette thèse.

3.2.2 Génération de variables et vecteurs aléatoires

La partie précédente a permis de décrire comment générer des réalisations de suites de variables aléatoires uniformes sur $[0,1]$. Cependant, les méthodes de Monte Carlo et de Quasi-Monte Carlo randomisé nécessitent des suites de variables ou vecteurs aléatoires, suivant des lois différentes que la loi uniforme sur $[0,1]$ ou $[0,1]^d$. En effet, l'objectif des méthodes de Monte Carlo et de Quasi-Monte Carlo randomisé dans cette thèse est l'obtention d'une information statistique d'une quantité d'intérêt Q dépendant d'un ensemble de paramètres incertains modélisés par une variable ou un vecteur aléatoire $\mathbf{X}$, de densité de probabilité π, et pouvant être obtenue comme l'espérance d'une fonction de cette quantité d'intérêt. Par exemple, la moyenne de cette quantité d'intérêt

est approchée de la manière suivante :

$$E\left[Q(\mathbf{X})\right] \approx \frac{1}{n}\sum_{k=1}^{n} Q(\mathbf{X}_k) \tag{3.5}$$

Les variables ou vecteurs aléatoires $\mathbf{X}_k$ suivent une loi différente de la loi uniforme sur $[0,1]^d$, et ce sont eux qu'il est nécessaire d'obtenir en pratique. Fort heureusement, il existe des méthodes permettant de générer une variable ou un vecteur aléatoire suivant une loi donnée à partir d'une séquence, plus ou moins longues, de variables aléatoires indépendantes et suivant une loi uniforme sur $[0,1]$ [24]. La dépendance ou l'indépendance entre les variables ou vecteurs aléatoires $\mathbf{X}_k$ suivant que l'on utilise une méthode de Monte Carlo ou de Quasi-Monte Carlo randomisé résultera de la dépendance ou de l'indépendance des variables aléatoires sur $[0,1]$ utilisées.

3.2.2.1 Génération de variables aléatoires réelles

De nombreuses méthodes existent pour la génération de variables aléatoires, parmi lesquelles certaines sont spécifiques à une loi de probabilité, qui peuvent être trouvées dans l'ouvrage [24]. Les deux méthodes présentées ici sont des méthodes générales, pouvant s'appliquer à n'importe quelle loi de probabilité, sous réserve que l'on ait accès à la densité de probabilité de celle-ci pour la première, et que l'on ait accès à la fonction de répartition pour la seconde.

3.2.2.1.1 Méthode de rejet La méthode de rejet permet de générer une variable aléatoire de densité de probabilité π définie sur $\mathbb{R}$. Pour cela, il est nécessaire de posséder une majoration de cette densité de probabilité π par une autre densité de probabilité f de la forme suivante :

$$\forall x \in \mathbb{R}, \pi(x) \leq cf(x) \tag{3.6}$$

La densité de probabilité f doit être telle qu'il est simple de générer une variable aléatoire suivant celle-ci. Une fois cette majoration connue, l'algorithme de la méthode de rejet est le suivant :

- Générer un couple de variables aléatoires indépendantes (X,U), avec X une variable aléatoire de densité de probabilité f et U de loi uniforme sur $[0,1]$.
- Si $Ucf(X) > \pi(X)$, reprendre à l'étape précédente, sinon conserver X

La variable aléatoire X générer par cette méthode suit une loi de probabilité caractérisée par la densité de probabilité π. L'algorithme boucle tant que la condition $Ucf(X) \leq \pi(X)$ n'est pas remplie, impliquant qu'il est préférable d'avoir une valeur de c la plus faible possible. En fait, le nombre moyen de bouclage est égale à c, justifiant la nécessité d'avoir d'avoir une valeur de c la

plus petite possible. Une illustration de la méthode est présentée sur la figure 3.5 avec une loi bêta de paramètres $(5, 2)$ à générer en utilisant une densité de probabilité uniforme pour f, et avec deux valeurs de c. Le nombre de valeurs conservées pour une valeur de c plus petite est plus importante, et engendre donc une plus grande efficacité de la méthode. La zone rouge sur la figure représente la zone de rejet pour laquelle le point de coordonnée $(X, Ucf(X))$ est rejeté, alors que la zone bleue correspond à la zone d'acceptation. Dans le cas de la figure, la loi de $(X, Ucf(X)) = (X, cU)$ est une loi uniforme sur le rectangle $[0, 1] \times [0, c]$, entraînant que la probabilité d'acceptation est de $1/c$, qui correspond au rapport de l'aire de la zone d'acceptation sur l'aire du rectangle $[0, 1] \times [0, c]$. Cela entraîne que le nombre moyen d'essai avant d'accepter un point est de c, confirmant qu'il est préférable d'avoir une valeur de c la plus petite possible. Cette propriété peut se généraliser, entraînant que le nombre d'essai moyen nécessaire est de c lors de l'utilisation de la méthode de rejet.

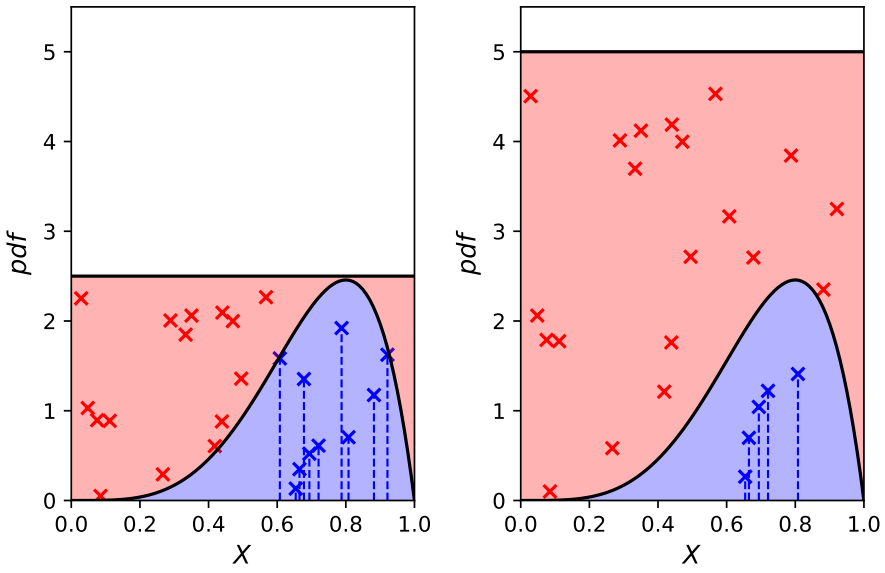

FIGURE 3.5 – *Illustration de la méthode de rejet pour la génération d'une variable aléatoire suivant une loi bêta de paramètres* $(5, 2)$*, en utilisant la densité uniforme pour la majoration avec deux constantes différentes c, valant* $2, 5$ *pour la figure de gauche et* 5 *pour la figure de droite. Un échantillon de* 25 *couples* (X, U) *ont été tirées qui sont les mêmes pour les deux figures, le point étant rejeté lorsque le point de coordonnées* (X, cU) *tombe dans la zone rouge, et gardé dans le cas contraire.*

Une des difficultés de la méthode de rejet est de trouver une bonne majoration de la densité d'intérêt, ce qui n'est pas toujours évident quand la densité d'intérêt est définie sur $\mathbb{R}$ tout entier par exemple où il faut s'assurer que la majoration soit encore valable au niveau des queues de la distribution de probabilité. Cette majoration doit de plus ne pas impliquer une constante c trop

importante afin de ne pas pénaliser la méthode. Une autre difficulté provient du fait qu'une partie des essais n'est pas conservé. Dans le cadre de la parallélisation proposée, le fait de rejeter une partie des essais entraîne une charge de travail pour les processus esclaves variables, nécessitant de modifier l'implémentation présentée précédemment.

Cette méthode est initialement prévue pour une utilisation dans le cadre de la méthode de Monte Carlo, mais une utilisation en méthode de Quasi-Monte Carlo peut également être effectuée, avec quelques modifications permettant de ne pas introduire de discontinuité pouvant nuire à la convergence de la méthode en effectuant une opération de lissage [86].

L'utilisation de cette méthode n'est cependant pas possible directement avec les méthodes de cubature présentées au chapitre précédent. Pour cette raison, la méthode d'inversion présentée dans le paragraphe suivant est préférable.

3.2.2.1.2 Méthode d'inversion Contrairement à la méthode de rejet qui nécessite la connaissance de la densité de probabilité, la méthode d'inversion requiert la connaissance de la fonction de répartition de la variable aléatoire que l'on souhaite générer. Bien entendu, la fonction de répartition peut s'obtenir en intégrant la densité de probabilité lorsque celle-ci existe, mais cela requiert une opération d'intégration parfois coûteuse, et on supposera donc dans la suite que la fonction de répartition est connue, impliquant que son coût d'évaluation est considéré comme raisonnable. On considère une variable aléatoire X de fonction de répartition F_X continue, et dont on définit l'inverse F_X^{-1} par la relation (3.7).

$$\forall u \in [0,1], F_X^{-1}(u) = \inf \{x \in \mathbb{R} | F_X(x) \geq u\} \tag{3.7}$$

La méthode d'inversion repose sur la propriété que pour une telle variable aléatoire X, la variable aléatoire $F_X^{-1}(U)$ où U suit une loi uniforme sur $[0,1]$ suit la loi de probabilité de X puisque :

$$\forall x \in \mathbb{R}, P(F_X^{-1}(U) \leq x) = P(U \leq F_X(x)) = F_X(x) \tag{3.8}$$

La relation (3.8) montre que la fonction de répartition de la variable aléatoire $F_X^{-1}(U)$ n'est autre que F_X, ce qui implique bien que $F_X^{-1}(U)$ suit la même loi de probabilité que X. La méthode d'inversion est illustrée sur la figure 3.6 avec à nouveau une variable aléatoire X de loi bêta de paramètres $(5,2)$. La figure montre comment on passe de points uniformément répartis sur l'axe des ordonnées correspondant à la variable aléatoire uniforme U à un ensemble de points qui ne sont plus uniformément répartis sur l'axe des abscisses correspondant à la variable aléatoire X, en inversant la fonction de répartition F_X.

En pratique, la difficulté est l'inversion de la fonction de répartition F_X qui peut être coûteuse. A partir du moment où la fonction F_X est connue, il

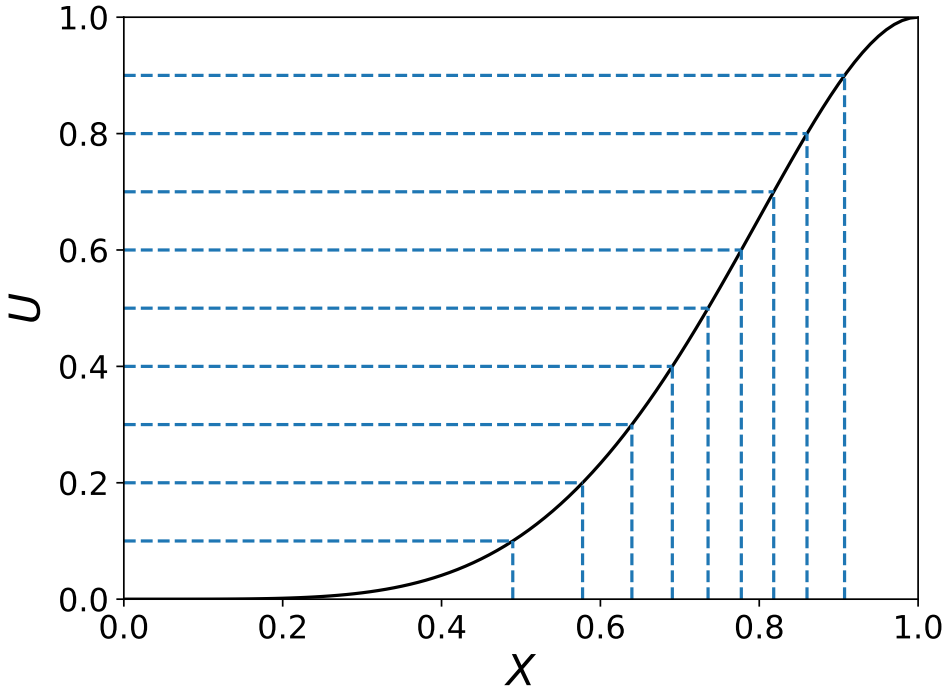

FIGURE 3.6 – *Illustration de la méthode d'inversion pour la génération d'une variable aléatoire X suivant une loi bêta de paramètres* $(5, 2)$. *La courbe correspond à la fonction de répartition* F_X.

est toujours possible de l'inverser à l'aide d'un algorithme de recherche de zéro d'une fonction, telle qu'une méthode par dichotomie ou une méthode de Newton par exemple [104]. De telles méthodes impliquent de nombreuses évaluations de F_X, ce qui constitue un potentiel aspect coûteux de la méthode d'inversion, d'autant plus que F_X est coûteuse à évaluer.

Un avantage de la méthode d'inversion est qu'elle est directement applicable à une méthode de Quasi-Monte Carlo randomisée ou même à une méthode de cubature. Elle peut en fait s'interpréter comme un changement de variable, permettant de passer d'une intégrale sur $\mathbb{R}$ muni de la fonction de pondération π à une intégrale sur $[0, 1]$ muni de la distribution uniforme.

3.2.2.2 Génération de vecteurs aléatoires

On considère un vecteur aléatoire $\mathbf{X}$ de $\mathbb{R}^d$. Dans le cas où les d composantes de ce vecteur aléatoire sont indépendantes entre elles, la génération de celui-ci revient à la génération de chacune des composantes séparément, et le problème revient donc à la génération de d variables aléatoires réelles, qui peut être réalisée en utilisant les méthodes de la section précédente.

L'objet est donc ici la génération de vecteurs aléatoires dont les compo-

santes présentent des dépendances entre elles. La méthode de rejet précédemment présentée peut se généraliser dans un tel cas dès qu'on a accès à la densité de probabilité π du vecteur aléatoire considéré et à une majoration de celle-ci de la forme de (3.6). Les propriétés de la méthode de rejet sont alors les mêmes que dans le cas d'une variable aléatoire. D'autres méthodes existent pour la génération de vecteurs aléatoires, parmi lesquelles celles basées sur l'utilisation de chaînes de Markov présentant souvent elles aussi une étape d'acceptation-rejet.

3.2.2.2.1 Utilisation de chaînes de Markov Le principe de base des méthodes utilisant les chaînes de Markov est de construire une chaîne de Markov $(\mathbf{X}_0, \mathbf{X}_1, \dots)$ prenant ses valeurs dans $\mathbb{R}^d$, dont la loi stationnaire est caractérisée par la densité de probabilité π, pour laquelle on cherche à générer des vecteurs aléatoires. Pour réaliser cette construction, l'idée est de choisir un point $\mathbf{X}_0$ selon une loi quelconque mais dont le support est inclus dans celui de π, et de construire par récurrence les éléments suivants de la chaîne à l'aide d'une relation de récurrence de la forme suivante :

$$\forall k \geq 0, \mathbf{X}_{k+1} = \psi(\mathbf{X}_k, \mathbf{U}_k) \tag{3.9}$$

Dans l'expression précedente, $\mathbf{U}_k$ correspond à un élément de $[0,1]^{m_k}$ avec m_k n'étant pas forcément constant ni égal à d, et ψ est une fonction qui assure que l'unique loi stationnaire de la chaîne de Markov est bien celle caractérisée par π. Ainsi, la construction de l'expression (3.9) permet de passer de points de $[0,1]$ à des points suivant la loi de probabilité caractérisée par π. Différents algorithmes existent permettant de construire une bonne fonction ψ, tels l'algorithme de Metropolis-Hasting [51] ou encore l'échantillonnage de Gibbs [39].

En tant que chaîne de Markov, l'état courant dépend uniquement de l'état précédent, et de ce fait une dépendance existe entre les états ce qui implique que les méthodes de Monte Carlo impliquant des chaînes de Markov (Markov Chain Monte Carlo) ne reposent pas sur les mêmes résultats théoriques que ceux qui seront présentés dans la section suivante, bien que des résultats similaires de convergence existent [56].

Les méthodes de Monte Carlo impliquant des chaînes de Markov peuvent être sous certaines conditions généralisées aux méthodes de Quasi-Monte Carlo et de Quasi-Monte Carlo randomisé [69, 21]. Néanmoins, ces méthodes ne peuvent être utilisées dans le cadre des méthodes de quadrature et de cubature, ce qui motive la méthode présentée dans le paragraphe suivant.

3.2.2.2.2 Transformation de Rosenblatt La généralisation de la méthode d'inversion au cas multidimensionnel correspond à la méthode d'inversion généralisée, aussi appelée la méthode des distributions conditionnelles [23]. Cette méthode consiste à décomposer la distribution de probabilité π sous la

forme du produit suivant :

$$\pi(x_1, \ldots, x_d) = \pi_1(x_1)\pi_{2|1}(x_2|x_1) \ldots \pi_{d|1,\ldots,d-1}(x_d|x_1, \ldots, x_{d-1}) \tag{3.10}$$

Étant donné un vecteur aléatoire $(U_1, \ldots, U_d)$ suivant une loi uniforme sur $[0,1]^d$, le vecteur aléatoire $(X_1, \ldots, X_d)$ définie par (3.11) a alors pour densité de probabilité π, F_π correspondant à la fonction de répartition associée à la densité de probabilité π.

$$\begin{cases} X_1 = F_{\pi_1}^{-1}(U_1) \\ \ldots \\ X_d = F_{\pi_{d|1,\ldots,d-1}}^{-1}(U_d|X_1, \ldots, X_{d-1}) \end{cases} \tag{3.11}$$

Cette méthode permet donc de passer d'un vecteur aléatoire $(U_1, \ldots, U_d)$ suivant une loi uniforme sur $[0,1]^d$ à un vecteur aléatoire $(X_1, \ldots, X_d)$ suivant la loi de probabilité définie par π. Les vecteurs aléatoires $(U_1, \ldots, U_d)$ et $(X_1, \ldots, X_d)$ sont alors relié par la relation (3.12), où T est la transformation permettant de passer de $(U_1, \ldots, U_d)$ à $(X_1, \ldots, X_d)$.

$$(X_1, \ldots, X_d) = T(U_1, \ldots, U_d) \tag{3.12}$$

Il devient donc possible d'écrire l'égalité d'espérances suivante pour une fonction g mesurable de d variables réelles :

$$E\left[g(X_1, \ldots, X_d)\right] = E\left[g(T(U_1, \ldots, U_d))\right] \tag{3.13}$$

Les deux membres de l'égalité (3.13) peuvent être exprimés sous forme d'intégrales, ce qui donne l'égalité suivante :

$$\int_{\mathbb{R}^d} g(x_1, \ldots, x_d)\pi(x_1, \ldots, x_d)dx_1 \ldots dx_d = \int_{[0,1]^d} g(T(u_1, \ldots, u_d))du_1 \ldots du_d \tag{3.14}$$

Cette dernière égalité traduit le fait que la transformation T, appelées transformation de Rosenblatt, revient à un changement de variable. Ainsi, il est toujours possible de se ramener à une intégrale sur l'hypercube $[0,1]^d$, ou de manière similaire à $[-1,1]^d$ comme dans le chapitre précédent pour l'utilisation de cubature. Les méthodes de Quasi-Monte Carlo et de Quasi-Monte Carlo randomisée peuvent naturellement être utilisée avec cette transformation, puisqu'elles sont initialement développée pour le calcul d'intégrale de la forme

du membre de droite de l'expression (3.14).

Un autre résultat lié à cette méthode et qui sera utilisé dans la suite, est qu'il est également possible à partir d'un vecteur aléatoire $(X_1, \ldots, X_d)$ de densité de probabilité π de construire un vecteur aléatoire $(U_1, \ldots, U_d)$ suivant une loi uniforme sur $[0,1]^d$ en utilisant la méthode dans l'autre sens [112].

La difficulté principale liée à la transformation de Rosenblatt est qu'il est nécessaire de connaître les fonctions de répartition conditionnelles de la loi de probabilité considérée. Le chapitre 5 sera consacré à des méthodes permettant d'estimer la densité de probabilité π de vecteurs aléatoires dont on possède simplement des réalisations, ces méthodes permettant également d'obtenir les fonctions de répartition conditionnelles sous certaines conditions en pratique.

3.3 Méthode de Monte Carlo

De manière générale, le terme de méthode de Monte Carlo désigne les manières d'obtenir des informations statistiques sur un système présentant un caractère aléatoire. L'objectif de la propagation d'incertitudes est d'obtenir des informations statistiques sur la sortie d'un système, étant donné un jeu de paramètres d'entrée incertains, ce qui s'avère être le cadre idéal pour les méthodes de Monte Carlo. Comme précisé dans le chapitre 1, de nombreuses informations statistiques d'intérêt peuvent être vu comme l'espérance mathématique d'une variable aléatoire correctement choisie, cette espérance pouvant elle même être vue comme le calcul d'une intégrale. La méthode de Monte-Carlo présentée dans ce chapitre est donc une méthode permettant d'estimer numériquement la valeur d'une espérance mathématique, ou d'une intégrale, et elles reposent sur des résultats théoriques de la théorie des probabilités.

3.3.1 Bases théoriques pour la méthode de Monte Carlo

3.3.1.1 Intuition

Le nom de méthode de Monte Carlo vient de la ville de Monte Carlo, célèbre pour les jeux de hasards qui y sont pratiqués. Un jeu de hasard simple consiste à miser sur le résultat d'un lancer de pièce de monnaie, qui donne un résultat pouvant prendre la valeur 'pile' ou la valeur 'face'. En considérant que la pièce est équilibrée, et que le lancer est aléatoire, l'intuition nous dit que la chance d'obtenir une des deux valeurs est la même. Dit autrement, le résultat d'une telle expérience aléatoire est 'pile' avec une probabilité de 1/2 et 'face' avec la même probabilité. Intuitivement, après un grand nombre de lancer, approximativement la moitié des résultats auront la valeur 'pile', et l'autre moitié la valeur 'face'. Autrement dit, la fréquence d'apparition d'un événement, définie comme le rapport entre le nombre de fois que cet événement a eu lieu et le nombre total de réalisations de l'expérience tend vers la probabilité que cet événement ait lieu.

La méthode de Monte Carlo reprend l'idée intuitive décrite précédemment avec l'expérience simple du lancer d'une pièce, et consiste à répéter plusieurs fois une expérience comprenant une partie de hasard. De l'ensemble de ces réalisations, diverses informations statistiques peuvent être obtenues, comme par exemple la probabilité qu'un événement se produise, qui sera obtenue à partir de la fréquence d'apparition de cet événement dans l'ensemble des réalisations. Des résultats théoriques permettent de justifier, sous certaines conditions, le fait que la fréquence d'apparition d'un événement tende vers la probabilité de cet événement, et sont l'objet des sections suivantes.

3.3.1.2 Estimateurs et méthodes de Monte Carlo

On se place dans le cadre de la théorie des probabilité brièvement discutée dans le chapitre 1, et on considère donc un espace probabilisé $(\Omega, \mathcal{A}, P)$. L'objectif est d'obtenir des informations sur une variable aléatoire X à valeur dans un espace vectoriel probabilisable $(E, \mathcal{E})$, qui peut par exemple être son espérance μ. Comme expliqué précédemment, la méthode de Monte Carlo repose sur la répétition d'une expérience aléatoire, ce qui correspond formellement pour l'expérience représentée par la variable aléatoire X à considérer un échantillon $(X_1, \ldots, X_n)$ de n variables aléatoires, chacune étant définie sur $(\Omega, \mathcal{A}, P)$, prenant ses valeurs dans $(E, \mathcal{E})$ et suivant la même loi que X. A partir de cet échantillon, il est possible de construire des estimateurs, qui sont simplement des fonctions de cet échantillon, et qui sont donc de la forme $f(X_1, \ldots, X_n)$. En tant que fonction d'un ensemble de variables aléatoires définies sur $(\Omega, \mathcal{A}, P)$, un estimateur est également une variable aléatoire définie sur $(\Omega, \mathcal{A}, P)$. En toute rigueur, il faut plutôt considérer une famille (f_n) de fonctions f dépendant chacune d'un échantillon de taille n, mais cette dépendance sera généralement omise dans la suite et on préférera l'utilisation de f à (f_n) pour simplifier l'écriture. Un exemple classique d'estimateur est l'estimateur de l'espérance S_n qui est la moyenne arithmétique de l'échantillon, définie par l'expression suivante :

$$S_n = \frac{1}{n} \sum_{i=1}^{n} X_i \tag{3.15}$$

Sous certaines conditions qui seront explicitées dans la section suivante, cet estimateur, qui est une variable aléatoire, converge vers l'espérance μ de X.

Les méthodes de Monte Carlo consiste donc à construire de tels estimateurs afin d'estimer les valeurs de grandeurs d'intérêt. Bien entendu, un estimateur sera d'autant plus intéressant qu'il convergera rapidement vers la vraie valeur à estimer. Un indicateur de l'erreur commise par un estimateur fréquemment utilisée est l'erreur quadratique moyenne (MSE pour Mean Squarred Error en anglais). Étant donné un estimateur réel $\hat{\theta}$ de la valeur réelle θ, l'erreur quadratique moyenne de $\hat{\theta}$ est définie comme l'espérance du carré de l'écart

entre la valeur estimée et la vraie valeur :

$$MSE(\hat{\theta}) = E\left[\left(\hat{\theta} - \theta\right)^2\right] \tag{3.16}$$

Un des intérêts de l'erreur quadratique moyenne est qu'elle se décompose sous la forme suivante, faisant intervenir le biais $E\left[\hat{\theta}\right] - \theta$ de l'estimateur ainsi que sa variance :

$$MSE(\hat{\theta}) = Bias\left[\hat{\theta}\right]^2 + Var\left[\hat{\theta}\right] \tag{3.17}$$

Cette dernière expression traduit le compromis biais-variance qui correspond à la nécessité pour réduire l'erreur quadratique moyenne de réduire simultanément le biais ainsi que la variance de l'estimateur. Dans le cas d'estimateurs sans biais, l'erreur quadratique moyenne de l'estimateur se ramène à la variance de l'estimateur. En pratique, le calcul de l'erreur quadratique moyenne, du biais ou de la variance peut s'avérer très coûteux puisqu'il repose sur un calcul d'espérance d'une variable aléatoire dépendant elle même de n variables aléatoires, et des considérations théoriques permettent généralement de construire des estimateurs dont on connaîtra par avance le comportement.

L'utilisation pratique d'un estimateur revient à obtenir une réalisation de celui-ci. Formellement, cela revient à choisir un élément arbitraire ω de Ω et d'évaluer $f(X_1(\omega), \ldots, X_n(\omega))$. $(X_1(\omega), \ldots, X_n(\omega))$ est alors appelé une réalisation de l'échantillon $(X_1, \ldots, X_n)$. Souvent, seule la réalisation d'un tel échantillon de variables aléatoires est accessible.

Dans les paragraphes suivant, la méthode de Monte Carlo présentées repose sur la considération d'échantillons $(X_1, \ldots, X_n)$ présentant une indépendance entre les variables aléatoires constituant l'échantillon, permettant d'obtenir de puissants résultats de convergence théorique.

3.3.1.3 Loi des grands nombres

Le premier résultat théorique important est la loi des grands nombres [62], qui donne des conditions suffisantes pour que la fréquence d'apparition d'un événement dans un ensemble d'expériences aléatoires tende vers la probabilité que cet événement se produise. L'énoncé de cette loi est le suivant :

Théorème 1 *Soient $(X_i)_{i\in\mathbb{N}^*}$ des variables aléatoires indépendantes et identiquement distribuées (i.i.d) d'espérance finie μ. Alors presque sûrement :*

$$\lim_{n\to+\infty} \frac{1}{n}\sum_{i=1}^{n} X_i = \mu \tag{3.18}$$

La probabilité d'un événement A sous la loi d'une variable aléatoire X, c'est à dire $P_X(A) = P(X \in A)$, peut également s'écrire comme l'espérance de la variable aléatoire $\mathbb{1}_A(X)$ où $\mathbb{1}_A$ est la fonction indicatrice de l'ensemble A, qui prend la valeur 1 sur A et 0 ailleurs. La loi des grands nombres donne donc un moyen de calculer cette probabilité en considérant une suite de variables aléatoires indépendantes entre elles $(X_i)_{i\in\mathbb{N}^*}$ et de même loi que X :

$$P_X(A) = E\left[\mathbb{1}_A(X)\right] = \lim_{n\to+\infty} \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_A(X_i) \tag{3.19}$$

Pour pouvoir appliquer la loi des grands nombres, plusieurs conditions doivent être réunies. Une première condition est que les variables aléatoires utilisées doivent toutes suivre la même loi. En pratique, pour vérifier cette condition lorsque l'on fait de la propagation d'incertitude à l'aide de méthodes non intrusives, il suffit juste de s'assurer que les paramètres incertains fournis en entrée suivent bien tous la même loi, ce qui assurera que les grandeurs en sortie suivent également toutes la même loi. En effet, si $\mathbf{X}$ et $\mathbf{Y}$ sont deux vecteurs aléatoires réels de même lois, et si f est une fonction mesurable, alors il est immédiat que $f(\mathbf{X})$ et $f(\mathbf{Y})$ ont la même loi. Ce cas de figure est typiquement rencontré lors de l'utilisation de méthode de propagation d'incertitude non intrusive, dont la méthode de Monte Carlo présentée ici fait partie.

Une autre condition à vérifier est l'indépendance des variables aléatoires X_i. L'importance de cette condition est illustrée sur la figure 3.7, où sont tracées les moyennes arithmétiques de deux suites de variables aléatoires $(U_i)_{i\in\mathbb{N}^*}$, dont les éléments sont indépendants entre eux et suivent tous une loi uniforme sur $[0,1]$, et $(\hat{U}_i)_{i\in\mathbb{N}*}$ qui est définie à partir de $(U_i)_{i\in\mathbb{N}^*}$ comme suit :

$$\begin{cases} \hat{U}_1 = U_1 \\ \forall i \geq 1, \hat{U}_{i+1} = \begin{cases} U_{i+1} \text{ si } \hat{U}_i < \frac{1}{2} \text{ et } U_{i+1} < \frac{1}{2} \\ 1 - U_{i+1} \text{ si } \hat{U}_i < \frac{1}{2} \text{ et } U_{i+1} \geq \frac{1}{2} \\ 1 - U_{i+1} \text{ si } \hat{U}_i \geq \frac{1}{2} \text{ et } U_{i+1} < \frac{1}{2} \\ U_{i+1} \text{ si } \hat{U}_i \geq \frac{1}{2} \text{ et } U_{i+1} \geq \frac{1}{2} \end{cases} \end{cases} \tag{3.20}$$

On peut montrer que chaque variable aléatoire $\hat{U}_i$ ainsi définie suit également une loi uniforme sur $[0,1]$, mais elles ne sont à l'évidence plus indépendantes entre elles. Le résultat de cette dépendance est que la moyenne arithmétique S_n ne converge pas vers la moyenne $\mu = 1/2$ de variables aléatoires uniformes, mais vers une autre valeur, comme visible sur la figure 3.7.

Considérer des dépendances entre les variables aléatoires de la suite peut donc modifier la valeur vers laquelle la moyenne arithmétique converge. L'introduction de dépendance n'empêche pas nécessairement la convergence vers la

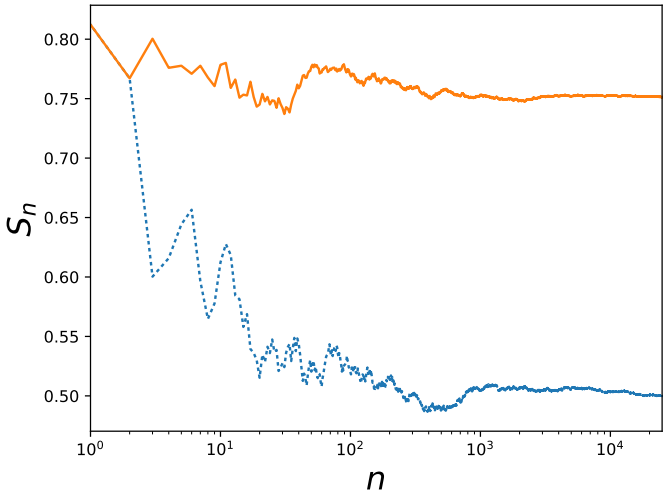

FIGURE 3.7 – *Moyennes arithmétiques S_n des suites de variables aléatoires $(U_i)_{i\in\mathbb{N}^*}$ (ligne pointillée) et $(\hat{U}_i)_{i\in\mathbb{N}^*}$ (ligne pleine).*

moyenne, comme dans le cas de processus stochastique discret ergodique [97]. Cependant, la convergence de la moyenne arithmétique dans un tel cas n'a pas lieu au sens presque sûr, mais dans un sens plus faible.

Enfin, une autre condition importante présente dans l'énoncé de la loi des grands nombres est l'existence d'un moment d'ordre 1 pour la loi que suivent l'ensemble des variables aléatoires de la suite. En l'absence d'un moment d'ordre 1, la moyenne arithmétique diverge. Sur la figure 3.8 sont représentées des réalisations de moyennes arithmétiques de suites de variables aléatoires i.i.d. suivant une loi de densité π_α sur $[1, +\infty[$ avec $\alpha < -1$, π_α étant définie par (3.21).

$$\pi_\alpha(x) = \frac{x^\alpha}{1+\alpha} \tag{3.21}$$

Une variable suivant une telle loi aura des moments d'ordre k pour k vérifiant $k < |\alpha| - 1$, les fonctions puissances étant intégrables sur $[1, +\infty[$ pour des valeurs de la puissance strictement inférieure à -1. La moyenne arithmétique des réalisations de S_n diverge bien quand n tend vers l'infini pour $\alpha = -1.5$, puisque la loi de densité $\pi_{-1.5}$ ne possède pas de moment d'ordre 1. En revanche, pour une valeur de α inférieure à -2, les réalisations de S_n convergent bien vers la moyenne μ_α, qui est donnée par (3.22).

$$\mu_\alpha = \frac{\alpha+1}{\alpha+2} \tag{3.22}$$

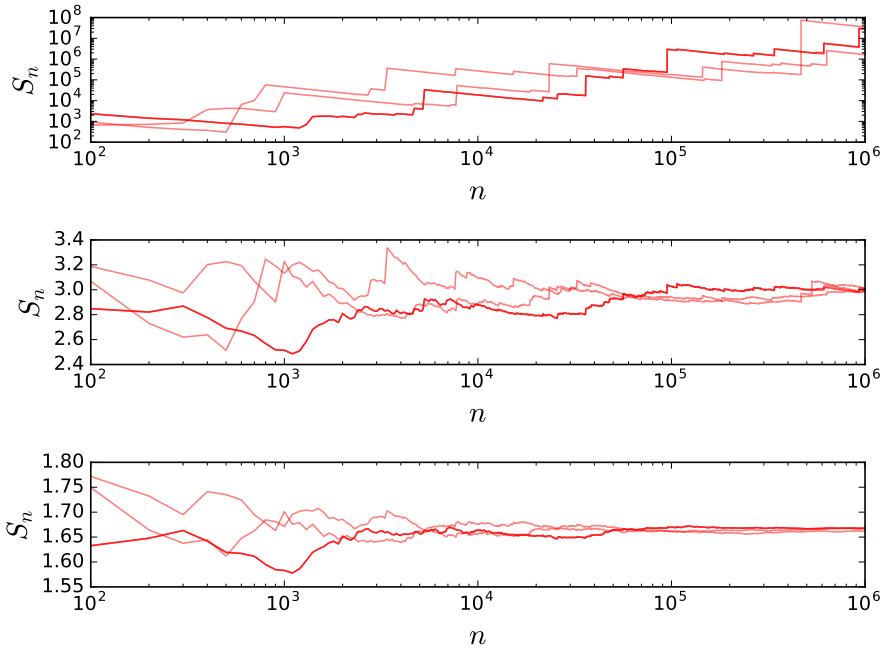

FIGURE 3.8 – *Réalisations de S_n (une courbe par réalisation) en fonction de n, S_n étant construit comme la moyenne arithmétique de variables aléatoires indépendantes et de lois définies par (3.21). Le graphe du haut présente des réalisations pour une valeur $\alpha = -1.5$, celui du milieu pour une valeur $\alpha = -2.5$ et celui du bas pour une valeur $\alpha = -3.5$*

Avant d'appliquer la méthode de Monte Carlo, il est donc important de s'assurer que la variable aléatoire que l'on considère possède bien une espérance finie. Un exemple simple d'une variable aléatoire ne possédant pas de moments d'ordre 1 est de considérer une variable aléatoire avec la densité (3.21) pour $\alpha < -1$. En dehors de cette condition de posséder un moment d'ordre 1, qui est une condition d'intégrabilité lorsqu'on transpose l'espérance en terme d'intégrale d'une fonction, il est important de remarquer qu'aucune condition sur la régularité de la fonction n'est requise pour avoir la convergence, comme c'était le cas pour les méthodes de quadrature présentées dans le chapitre 2.

La loi des grands nombres présentées dans cette section donne des conditions suffisantes pour réaliser en pratique une méthode de Monte Carlo, et obtenir certaines statistiques sur le système étudiée. Néanmoins, cette loi ne donne aucune indication sur la performance de la méthode de Monte Carlo, c'est à dire sur la convergence de S_n vers μ. Un autre résultat théorique important permet d'obtenir une telle information, et est l'objet de la prochaine section.

3.3.1.4 Théorème Centrale Limite

Le théorème central limite (TCL), dû à Polya [103], permet de déterminer la loi limite de la moyenne arithmétique S_n de variables aléatoires *i.i.d*, lorsque ce nombre de variables tend vers l'infini. Son énoncé est le suivant :

Théorème 2 *Soient $(X_i)_{i\in\mathbb{N}^*}$ une suite de variables aléatoires réelles i.i.d. d'espérance μ et de variance $0 < \sigma^2 < +\infty$. Soit la suite de variables aléatoires $(Z_n)_{n\in\mathbb{N}^*}$ définie par :*

$$Z_n = \frac{S_n - \mu}{\sigma\sqrt{n}} \tag{3.23}$$

Alors la suite de variables aléatoires $(Z_n)_{n\in\mathbf{N}^*}$ où Z_n converge en loi vers la loi normale centrée réduite $\mathcal{N}(0,1)$ lorsque n tend vers l'infini.

Comparées à la loi des grands nombres présentées précédemment, les conditions d'applications sont légèrement plus fortes. En effet, la loi des grands nombres exige seulement que la loi suivie par les variables possèdent un moment d'ordre 1, alors que pour le TCL, un moment d'ordre 2 est requis pour la loi suivie par les variables aléatoires de la suite.

Pour n suffisamment grand, la moyenne arithmétique S_n suit approximativement sous les conditions du TCL une loi normale de moyenne μ, d'écart-type $\frac{\sigma}{\sqrt{n}}$ et donc de variance $\frac{\sigma}{n}$. Le fait que l'écart-type évolue comme $n^{-1/2}$ conduit souvent à dire que la méthode de Monte Carlo offre une convergence inversement proportionnelle à $\sqrt{n}$. Les fractiles de la loi limite varie également comme $n^{-1/2}$. En effet, le f-fractile $x_f^{(n)}$ avec $f \in [0,1]$ pour la loi normale de moyenne μ et d'écart-type $\frac{\sigma}{\sqrt{n}}$ est défini par :

$$x_f^{(n)} = \mu + \frac{\sigma\sqrt{2}}{\sqrt{n}}\,\mathrm{erf}^{-1}(2f-1) \tag{3.24}$$

Il est donc possible de définir un intervalle $I_p^{(n)}$ centré sur μ, tel que la probabilité pour S_n de se trouver dans cet intervalle soit asymptotiquement de p, avec $p \in [0,1]$. La définition d'un tel intervalle est donné par (3.25).

$$I_p^{(n)} = \left[x_{\frac{1-p}{2}}^{(n)}, x_{\frac{1+p}{2}}^{(n)}\right] \tag{3.25}$$

La longueur de l'intervalle $I_p^{(n)}$ est alors également proportionnelle à $n^{-\frac{1}{2}}$. L'ensemble des points $(\ln(n), \ln(I_p^{(n)}))$ se placent donc sur une droite de pente $-1/2$, comme visible par la ligne pointillé sur la figure 3.9, où S_n est définie comme la moyenne arithmétique de variables indépendantes et uniforme sur $[0,1]$, pour lesquelles $\mu = 1/2$ et $\sigma^2 = 1/12$, et où la valeur de p est de 0.99.

Avec cette définition de S_n, la fonction de répartition de S_n est accessible analytiquement [48], permettant d'obtenir les fractiles de S_n et donc la longueur d'un intervalle centré (S_n a une loi symétrique par construction) $I_p(S_n)$ pour lequel la probabilité d'y trouver S_n est exactement p. La longueur de cet intervalle $I_p(S_n)$ est représentée par les triangles sur la figure 3.9, avec $p = 0.99$. On peut observer qu'avec l'augmentation de la valeur de n, la longueur de l'intervalle $I_p(S_n)$ représentée par les triangles tend à être égale à celle de l'intervalle $I_p^{(n)}$ correspondant à une loi normale avec pour variance $\frac{1}{12n}$ comme définie par le TCL appliquée à une suite de variables aléatoires indépendantes de lois uniformes sur $[0, 1]$. Malgré son caractère asymptotique, le TCL donne dans ce cas une bonne approximation très rapidement, puisque pour $n = 10$ les points sont d'ores et déjà visuellement sur la droite en pointillés.

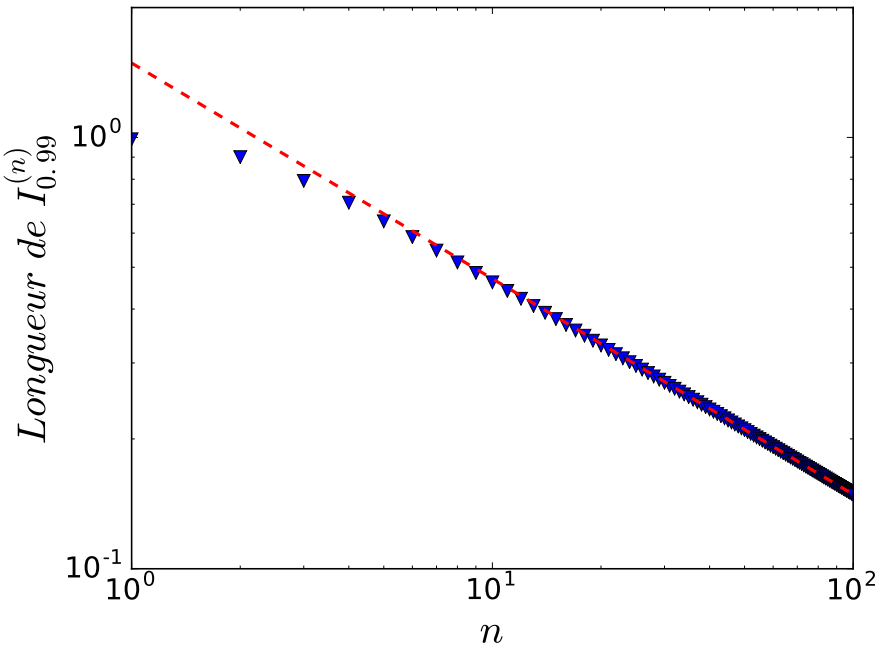

FIGURE 3.9 – *Longueur de l'intervalle centré $I_p(S_n)$ en fonction de n pour la moyenne arithmétique S_n de variables indépendantes et uniforme sur $[0, 1]$ (triangle), et longueur de l'intervalle centré de probabilité $p = 0.99$ pour la loi normale limite donnée par le TCL pour S_n (pointillés), en fonction du nombre n de termes présents dans la moyenne arithmétique.*

Un intervalle tel que $I_p(S_n)$ peut également être estimé numériquement en estimant ses bornes qui sont des fractiles empiriques, et cela même lorsque la variable aléatoire considérée ne possède pas de moment d'ordre 2. Pour cela, il suffit de lancer N simulations indépendantes permettant d'obtenir N réalisations $(s_n^{(i)})_{1\leq i\leq N}$ de S_n. Il faut ensuite ordonner l'ensemble de ces réalisations pour obtenir une liste ordonnée $(s_{n,ord}^{(i)})_{1\leq i\leq N}$. Le f-fractile empirique $x_{f,emp}^{(n)}$

de S_n peut alors être défini par la relation (3.26), comme barycentre des deux points de la liste ordonnées $(s^{(i)}_{n,ord})_{1\leq i\leq N}$ encadrant ce fractile.

$$x^{(n)}_{f,emp} = (fN - \lfloor fN \rfloor)s^{(\lfloor fN \rfloor)}_{n,ord} + (\lfloor fN \rfloor + 1 - fN)s^{(\lfloor fN \rfloor+1)}_{n,ord} \tag{3.26}$$

Dans l'expression (3.26), $\lfloor x \rfloor$ correspond à la partie entière inférieur de x. On est alors en mesure de construire l'ensemble des points $(\ln(n), \ln(I_{p,emp}(S_n)))$ pour différentes valeurs de n, qui se placent asymptotiquement sur une droite en vertu du TCL comme présenté sur la figure 3.9. Une régression linéaire sur l'ensemble de ces points permet de donner une estimation de la pente de cette droite.

La pente de la droite de régression des points $(\ln(n), \ln(I_{0.96,emp}(S_n)))$ pour n variant de 10 à 10^5 par pas de 10 a été calculée pour des variables aléatoires ayant pour densité π_α (3.21) pour différentes valeurs de α, en utilisant $N = 24,000$ échantillons. Le coefficient de corrélation pour l'ensemble étudié était toujours en valeur absolue supérieur à 0.98 ce qui indique que les points sont bien situés sur une droite. Sur la figure 3.10 est représentée la pente de la droite de régression en fonction de α. Les lois de probabilité considérées possèdent un moment d'ordre 2 pour α inférieur à -3 uniquement, impliquant que le TCL ne peut s'appliquer que dans ces cas. En pratique on observe que pour les valeurs de α comprise entre -3 et -2, la pente n'est pas égale à $-1/2$, et est plus faible. Pour ces valeurs de α, la variable aléatoire possède simplement un moment d'ordre 1. Ainsi, seul la loi des grands nombres s'applique, qui implique qu'il y a une convergence presque sûre vers la moyenne, la taille de l'intervalle doit donc tendre vers 0 malgré tout, ce qui se traduit par la pente négative. Pour des valeurs de α inférieure à -3.5, la valeur observée de la pente vaut bien $-1/2$ comme prévu par le TCL. Pour les valeurs comprises entre -3.5 et -3, la valeur de la pente n'est pas tout à fait égale à $-1/2$, le caractère asymptotique semblant plus long à atteindre à mesure que la valeur de α se rapproche de la valeur limite $\alpha = -3$.

Le TCL donne la rapidité de convergence de la méthode de Monte Carlo qui est inversement proportionnelle à la racine carré de la taille de l'échantillon utilisé. Dans la construction de la figure 3.10, un nombre important de réalisations de S_n ont été utilisées pour retrouver numériquement le résultat du TCL. En pratique, une seule réalisation de S_n est calculée car déjà suffisamment coûteuse, et des informations sur la convergence doivent être extraites de cette seule réalisation. Cette information sur la convergence de la méthode de Monte Carlo est donnée par le calcul d'intervalles de confiance, qui sont basées sur le TCL et d'autres résultats théoriques.

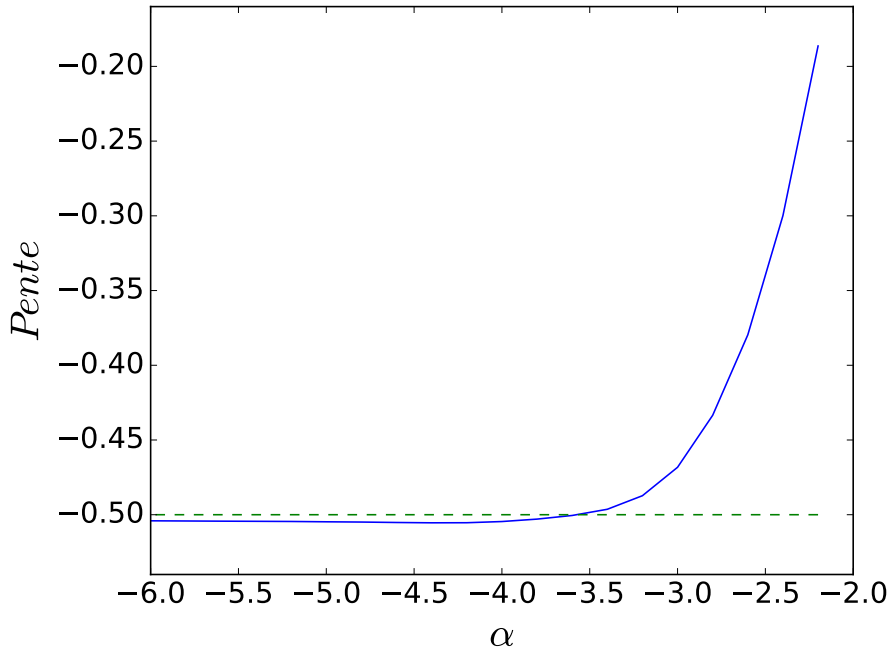

FIGURE 3.10 – *Pente de la droite de régression linéaire pour les points* $(\ln(n), \ln(I_{0.96,emp}(S_n)))$ *dans le cas d'une moyenne arithmétique de variables aléatoires indépendantes ayant une densité* π_α (3.21)*, pour différentes valeurs de* α*.*

3.3.1.5 Construction d'intervalles de confiance

En reprenant les notations de l'énoncé du TCL, on a que la suite de variables aléatoires $(Z_n)_{n\in\mathbb{N}^*}$ suit asymptotiquement une loi normale centrée réduite. La continuité de la fonction de répartition de la loi normale centrée réduite sur $\mathbb{R}$ suffit alors pour affirmer que pour tout intervalle I de $\mathbb{R}$, on a, Z suivant une loi normale centrée réduite :

$$\lim_{n\to+\infty} P(Z_n \in I) = P(Z \in I) \tag{3.27}$$

Si $I = [a, b]$ est un intervalle tel que $P(Z \in I) \geq 1 - \alpha$ avec $\alpha \in [0, 1]$, alors on peut en déduire l'inégalité suivante :

$$\lim_{n\to+\infty} P(Z_n \in I) \geq 1 - \alpha \tag{3.28}$$

La relation (3.28) peut se réécrire, en utilisant la définition de Z_n (3.23) à

α	Longueur de $I_{1-\alpha}$
0.5	1.349
0.25	2.301
0.1	3.290
0.05	3.920
0.01	5.152
0.001	6.581
0.00001	8.834

TABLE 3.1 – *Longueur des intervalles $I_{1-\alpha}$ définis par l'expression (3.30) pour différentes valeurs de α.*

partir de la moyenne arithmétique S_n, de μ et de σ :

$$\lim_{n\to+\infty} P\left(\mu \in \left[S_n - a\frac{\sigma}{\sqrt{n}}, S_n + b\frac{\sigma}{\sqrt{n}}\right]\right) \geq 1-\alpha \tag{3.29}$$

Dans la relation (3.29), il est important de noter que les bornes de l'intervalle $\left[S_n - a\frac{\sigma}{\sqrt{n}}, S_n + b\frac{\sigma}{\sqrt{n}}\right]$ dépendent de S_n, qui est une variable aléatoire, et cet intervalle est donc un intervalle aléatoire. La relation (3.29) traduit le fait que la probabilité pour μ de se trouver dans cet intervalle aléatoire est asymptotiquement supérieure à $1-\alpha$. Un tel intervalle aléatoire est appelé intervalle de confiance asymptotique pour μ de niveau $1-\alpha$, et est différent de l'intervalle donné par l'expression (3.25), qui n'était pas aléatoire et pour lequel on s'intéressait à la présence de S_n et non de μ à l'intérieur. En pratique, l'intervalle I considéré dans (3.27) est souvent centré et ne dépend que de α, le TCL impliquant la loi normale centrée réduite qui est symétrique. Ce dernier peut être défini à l'aide de la fonction de répartition de la loi normale centrée réduite Φ, et son expression est donnée en 3.30.

$$I_{1-\alpha} = \left[-\Phi^{-1}\left(1-\frac{\alpha}{2}\right), \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\right] \tag{3.30}$$

La fonction Φ ayant pour expression :

$$\Phi(x) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{x}{\sigma}\right)\right] \tag{3.31}$$

Dans le tableau 3.1 sont recensées plusieurs valeurs de α ainsi que les longueurs des intervalles $I_{1-\alpha}$ définis par l'expression (3.30).

En combinant l'expression (3.23), l'expression (3.27) et (3.30) il vient l'expression suivante concernant l'intervalle de confiance asymptotique de niveau

$1-\alpha$ pour S_n construit à l'aide du TCL :

$$\lim_{n\to+\infty} P\left(\mu \in \left[S_n - \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\frac{\sigma}{\sqrt{n}}, S_n + \Phi^{-1}\left(1-\frac{\alpha}{2}\right)\frac{\sigma}{\sqrt{n}}\right]\right) \geq 1-\alpha \tag{3.32}$$

Il est immédiat que la longueur de l'intervalle de confiance asymptotique définit par l'expression (3.32) est proportionnelle à celle de $I_{1-\alpha}$ défini par l'expression (3.30), et compte tenu des valeurs de longueur présente dans le tableau 3.1, on peut voir que pour passer d'une probabilité de ne pas trouver μ de 10% à 1%, il suffit de multiplier la taille de l'intervalle de confiance asymptotique par moins de 2, et pour passer de 1% à 0.00001%, il suffit également de multiplier la taille de l'intervalle de confiance asymptotique par moins de 2. Le choix de α n'est donc pas primordiale car il n'influencera pas significativement l'erreur d'estimation donnée par la longueur de l'intervalle, et on considérera généralement une valeur de α de 0.05 ou 0.01.

Il est donc possible de construire un intervalle de confiance asymptotique à partir du TCL en utilisant l'expression (3.32). Sur la figure 3.11 est présentée une réalisation de la moyenne arithmétique de variables aléatoires indépendantes et uniformes sur $[0,1]$, ainsi que des intervalles de confiance centrés obtenus grâce au TCL et à l'écart-type de la loi uniforme sur $[0,1]$ qui est connu. Comme prévu par la loi des grands nombres, la moyenne arithmétique tend bien vers la valeur moyenne de la loi uniforme, qui est 0.5. L'intervalle de confiance se resserre également avec l'augmentation de n, ce qui est attendu de par la construction de celui-ci. Pour une valeur de n fixé, on peut observer que la valeur moyenne n'est pas toujours contenue dans l'intervalle de confiance. En augmentant la valeur de $1-\alpha$, on augmente la chance de trouver la valeur moyenne dans l'intervalle de confiance, ce qui est bien la définition d'un intervalle de confiance. Enfin, la probabilité de ne pas trouver la valeur moyenne dans l'intervalle de confiance est divisée par 10 entre deux intervalles successifs, et pour autant, la largeur de l'intervalle de confiance n'augmente pas significativement, comme expliqué précédemment à l'aide du tableau 3.1.

Un ingrédient nécessaire à la construction d'un intervalle de confiance grâce au TCL est la donnée de l'écart-type, ou ce qui revient au même, la variance, de la variable aléatoire étudiée. En général, l'écart-type, tout comme l'espérance de la variable aléatoire, sont des quantités inconnues que l'on cherche à estimer à l'aide de la méthode de Monte Carlo. Tout comme l'espérance peut être estimée par la moyenne arithmétique S_n d'une suite de variables aléatoires i.i.d. $(X_i)_{i\in\mathbb{N}}$ d'après la loi des grands nombres, on peut montrer d'après cette même loi des grands nombres que la variance peut être estimée à l'aide de

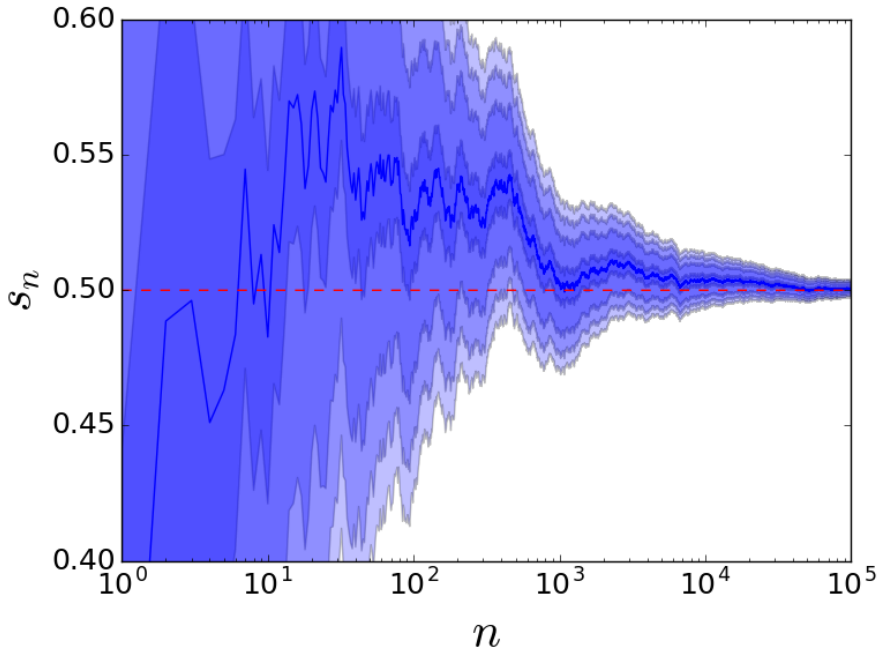

FIGURE 3.11 – *Réalisation d'une moyenne arithmétique de variables aléatoires indépendantes et uniforme sur* $[0,1]$*, et intervalles de confiance centrés à* 50%*,* 95%*,* 99.5% *et* 99.95% *construits à partir de l'écart-type théorique de la loi uniforme sur* $[0,1]$ *et de l'expression* (3.32)*.*

l'estimateur V_n qui a pour expression :.

$$V_n = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - S_n)^2 \tag{3.33}$$

Cet estimateur est sans biais, ce qui signifie que son espérance est exactement la valeur σ^2 qui est la variance de X. La convergence découlant de la loi des grands nombres, celle-ci a lieu de manière presque sûre vers la valeur de la variance des X_i. Le fait que la convergence ait lieu de manière presque sûre, permet de pouvoir utiliser le théorème de Slutsky [122] dont l'énoncé est le suivant :

Théorème 3 *Soient deux suites de variables aléatoires réelles* $(Y_n)_{n\in\mathbb{N}}$ *et* $(Z_n)_{n\in\mathbb{N}}$*,* Y *une variable aléatoire réelle et* $c \in \mathbb{R}$*, tels que :*

$$Y_n \xrightarrow{\mathcal{L}} Y \tag{3.34}$$

$$Z_n \xrightarrow{\mathcal{P}} c \tag{3.35}$$

Alors :

$$Y_n + Z_n \xrightarrow{\mathcal{L}} Y + c \tag{3.36}$$

$$Y_n Z_n \xrightarrow{\mathcal{L}} cY \tag{3.37}$$

En notant $V > 0$ la variance des variables aléatoires $(X_i)_{i\in\mathbb{N}}$, on trouve que $\frac{\sqrt{V}}{\sqrt{V_n}}$ converge presque sûrement, et donc en probabilité vers 1. De ce fait, le produit $\frac{\sqrt{V}}{\sqrt{V_n}} Z_n$, où Z_n est définie par (3.23), converge en loi vers une loi normale centrée réduite d'après la seconde partie de la conclusion du théorème de Slutsky. On peut donc écrire la relation (3.38), similaire à la relation (3.28).

$$\lim_{n\to+\infty} P(\frac{\sqrt{V}}{\sqrt{V_n}} Z_n \in I) \geq 1 - \alpha \tag{3.38}$$

Cette dernière relation permet d'obtenir un intervalle de confiance présenté dans (3.39), similaire à celui présent dans (3.29), mais impliquant cette fois une estimation de l'écart-type plutôt que la vraie valeur qui est généralement inconnue.

$$\lim_{n\to+\infty} P\left(\mu \in \left[S_n - a\frac{\sqrt{V_n}}{\sqrt{n}}, S_n + b\frac{\sqrt{V_n}}{\sqrt{n}}\right]\right) \geq 1 - \alpha \tag{3.39}$$

Sur la figure 3.12 est présentée la même réalisation de la moyenne arithmétique de variables aléatoires indépendantes et uniformes sur $[0, 1]$ que sur la figure 3.11, ainsi que des intervalles de confiance à 95% centrés obtenus grâce au TCL et à l'écart-type de la loi uniforme sur $[0, 1]$ obtenu grâce à (3.33), et en utilisant la vraie valeur pour l'écart-type. Bien que différents au début, on peut observer que les intervalles de confiance tendent à être identiques asymptotiquement, l'estimation de l'écart-type tendant vers la vraie valeur de celui-ci.

En pratique, l'estimation de l'erreur de la méthode de Monte Carlo pour le calcul de la moyenne est réalisée à l'aide de l'intervalle de confiance de l'expression (3.39), obtenu grâce au TCL et au théorème de Slutsky. Cette intervalle de confiance est asymptotique, de par le fait que le TCL est un résultat asymptotique et également par le fait que l'écart-type considéré a été remplacé par un estimateur de celui-ci. Sans plus d'informations, on estimera que pour des grands nombres de réalisations n, typiquement supérieurs à 10^4, ces résultats

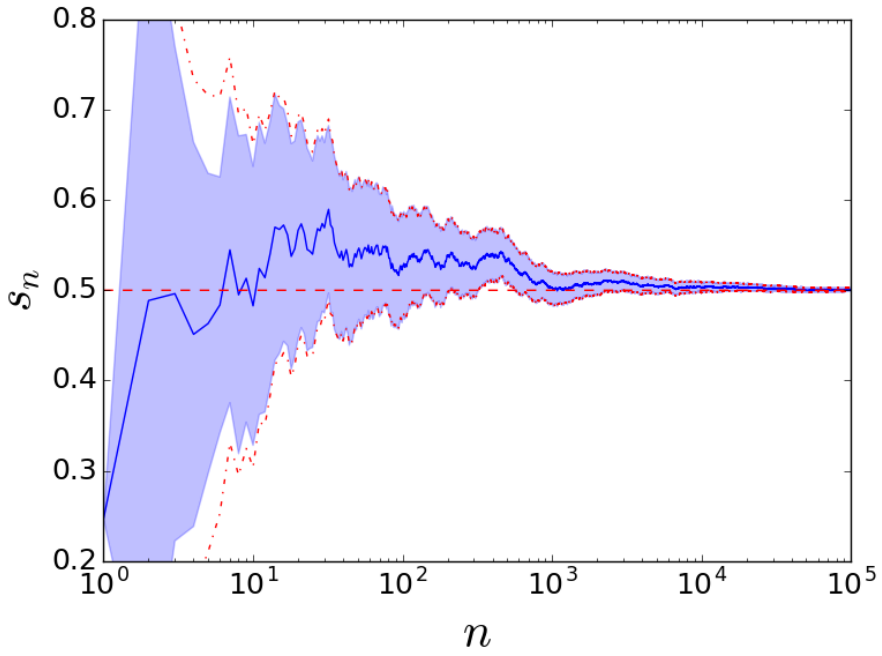

FIGURE 3.12 – *Réalisation d'une moyenne arithmétique de variables aléatoires indépendantes et uniforme sur* $[0,1]$*, et intervalle de confiance centré à* 95% *donné par* (3.39)*, et intervalle de confiance centré à* 95% *construit à partir de la vraie valeur de l'écart-type (pointillés) donné par* (3.32)*.*

asymptotiques sont valables. Dans le cadre des lois de densité π_α, le comportement asymptotique du TCL est plus long à atteindre pour les valeurs de α proche de -3, qui possèdent un moment d'ordre 2 de "justesse", et pas de moments d'ordre supérieur. La présence de moments d'ordre supérieur permet de définir des intervalles de confiances également pour ceux-ci, en appliquant la même méthode que pour la moyenne, qui est le moment d'ordre 1.

La construction d'intervalles de confiance permet de définir un critère d'arrêt pour la méthode de Monte Carlo, en imposant que des intervalles de confiance de quantités d'intérêt aient leurs longueurs qui passent sous un certain seuil. Ces quantités d'intérêts sont généralement fonction d'un vecteur de paramètres incertains, modélisé à l'aide d'un vecteur aléatoire dont la taille est la dimension stochastique du problème. Contrairement aux méthodes présentées dans le chapitre précédent, la vitesse de convergence de la méthode de Monte-Carlo donnée par le TCL ne dépend que de la variance de la quantité d'intérêt, et pas de la dimension stochastique du problème, ce qui rend possible son utilisation quelque soit la dimension stochastique.

3.3.2 Amélioration de la méthode de Monte Carlo

La méthode de Monte Carlo basique présentée précédemment repose sur les résultats théoriques donnés par le TCL, stipulant que la convergence de la méthode est proportionnelle à l'écart-type σ de la variable aléatoire dont on cherche l'espérance. Une perspective d'amélioration de la méthode de Monte Carlo est la réduction de l'écart-type σ par diverses considérations, permettant de fait d'accélérer la convergence de la méthode. Ces méthodes sont dénommées méthodes de réduction de variance, et sont brièvement discutées dans le prochain paragraphe.

3.3.2.1 Méthodes de réduction de variance

Une revue des principales méthodes de réduction de variance peut être trouvée dans [72]. Un point commun de l'ensemble des méthodes présentées est qu'elles utilisent toutes les résultats du Théorème Central Limite, et sont donc soumises aux hypothèses de celui-ci, en particulier l'hypothèse d'indépendance. De fait, le taux de convergence reste inchangé, la longueur d'un intervalle de confiance étant toujours proportionnelle à $1/\sqrt{n}$ où n est la taille de l'échantillon utilisé. L'obtention d'un résultat obtenu à l'aide d'une de ces méthodes deux fois plus précis nécessite donc, comme pour la méthode de Monte Carlo basique, une taille d'échantillon quatre fois plus importante. Néanmoins, les gains par rapport à la méthode de Monte Carlo basique peuvent ne pas être négligeable, rendant l'utilisation de telles méthodes intéressantes dans certaines applications.

La méthode de Monte Carlo vise à calculer l'espérance d'une quantité d'intérêt Q fonction de paramètres incertains que l'on modélise à l'aide d'un vecteur aléatoire $\mathbf{X}$. Dans la section précédente, une transformation T a été introduite, permettant de transformer une vecteur aléatoire $\mathbf{U}$ de loi uniforme sur $[0,1]^d$ en un vecteur aléatoire $\mathbf{X}$ de loi désirée, de sorte que l'espérance de Q peut s'exprimer sous les formes suivantes :

$$E[Q] = E[f(\mathbf{X})] = E\left[\tilde{f}(\mathbf{U})\right] \tag{3.40}$$

Dans l'expression précédente, f et $\tilde{f}$ sont liées par la relation $\tilde{f} = f \circ T$, $\circ$ correspondant à la composition des fonctions. Les méthodes de réduction de variance consiste à exprimer l'espérance de Q différemment des deux expressions présentes dans l'expression (3.40), tout en conservant son expression sous la forme d'un calcul d'une ou plusieurs espérances.

3.3.2.1.1 Variable de contrôle Une variable de contrôle dans le contexte de la méthode de Monte Carlo désigne une variable aléatoire C dont on connaît l'espérance μ_C, et qui est de préférence peu coûteuse à évaluer. La méthode

de réduction de la variance par variable de contrôle consiste à utiliser l'identité suivante, dans laquelle β est un nombre réel qui sera défini ultérieurement :

$$E[Q] = E[Q + \beta(\mu_C - C)] \tag{3.41}$$

Pour avoir accès à l'espérance de la variable aléatoire Q, il est préférable d'estimer le membre de droite à l'aide d'une méthode de Monte Carlo lorsque la variance de la variable aléatoire $Q + \beta(\mu_c - C)$ est inférieure à celle de la variable aléatoire Q. Or la variance de cette quantité est donnée par l'expression suivante :

$$Var(Q + \beta(\mu_c - C)) = Var(Q) + \beta^2 Var(C) - 2\beta Cov(Q, C) \tag{3.42}$$

Cette variance, qui dépend de β, peut être minimisée en choisissant la valeur suivante pour β :

$$\beta_{opt} = \frac{Cov(Q, C)}{Var(C)} \tag{3.43}$$

La variance vaut alors $(1 - Corr(Q, C)^2)Var(Q)$, et est donc d'autant plus petite que la variable de contrôle est corrélée (ou anti-corrélée) à la variable aléatoire Q. Plusieurs difficultés sont présentes pour la mise en place pratique de cette méthode. Tout d'abord, cette méthode nécessite un choix "intelligent" de la variable de contrôle. Une fois cette variable de contrôle choisie, il convient de déterminer la valeur de β_{opt}, ce qui peut se faire à l'aide de l'échantillon utilisé pour la méthode de Monte Carlo. Cependant, l'utilisation directe de cette échantillon pour la détermination de β_{opt} biaise l'estimation de l'espérance de Q, ainsi que de la variance de l'estimateur, et des méthodes plus complexes doivent alors être invoquées afin de retirer ces biais.

3.3.2.1.2 Échantillonnage préférentiel L'échantillonnage préférentiel repose sur la réécriture du calcul de l'espérance de Q sous la forme suivante :

$$\begin{aligned} E[Q] = E[f(\mathbf{X})(\,right] &= \int_{\mathbb{R}^d} f(\mathbf{x})\pi_X(\mathbf{x})d\mathbf{x} \\ &= \int_{\mathbb{R}^d} f(\mathbf{y})\frac{\pi_X(\mathbf{y})}{\pi_Y(\mathbf{y})}\pi_Y(\mathbf{y})d\mathbf{y} = E\left[f(\mathbf{Y})\frac{\pi_X(\mathbf{Y})}{\pi_Y(\mathbf{Y})}\right] \end{aligned} \tag{3.44}$$

L'espérance de Q revient alors au calcul de l'espérance de la variable aléatoire $f(\mathbf{Y})\pi_X(\mathbf{Y})/\pi_Y(\mathbf{Y})$, le vecteur aléatoire $\mathbf{Y}$ ayant une densité de probabilité π_Y, plutôt qu'au calcul de l'espérance de la variable aléatoire $f(\mathbf{X})$ où le vecteur aléatoire $\mathbf{X}$ a la densité de probabilité π_X. La réduction de la variance

passe par le choix d'une bonne densité de probabilité π_Y, permettant de minimiser la variance de la variable aléatoire $f(\mathbf{Y})\pi_X(\mathbf{Y})/\pi_Y(\mathbf{Y})$. Le choix optimal pour la distribution π_Y, permettant la plus grande réduction de variance, est le suivant :

$$\pi_Y^{(opt)}(\mathbf{y}) = \frac{|f(\mathbf{y})|\pi_X(\mathbf{y})}{\int_{\mathbb{R}^d} |f(\mathbf{y})|\pi_X(\mathbf{y})d\mathbf{y}} \tag{3.45}$$

Cependant, cette densité de probabilité est en général inconnue, et ne peut donc pas être utilisée en pratique. Il convient alors de faire un choix "intelligent" en fonction du problème à traiter, un mauvais choix pouvant entraîner une augmentation de la variance comparée au Monte Carlo basique.

3.3.2.1.3 Conditionnement La méthode de réduction de variance par conditionnement consiste en l'utilisation de la propriété suivante de l'espérance conditionnelle de Q par une variable aléatoire C, que l'on peut considérer dépendre elle aussi du vecteur aléatoire $\mathbf{X}$:

$$E[Q] = E[E[Q|C]] = E[E[f(\mathbf{X})|g(\mathbf{X})]] \tag{3.46}$$

Plutôt que d'estimer l'espérance de la variable aléatoire Q, l'espérance de la variable aléatoire $E[Q|C]$ est estimée par une méthode de Monte Carlo. La variance de cette nouvelle variable aléatoire est nécessairement plus petite que la variance de Q. En effet, la formule de décomposition de la variance énonce que :

$$Var(Q) = E[Var(Q|C)] + Var(E[Q|C]) \tag{3.47}$$

La variable aléatoire $Var(Q|C)$ étant tout le temps positive par définition, il s'en suit que l'on a toujours :

$$Var(E[Q|C]) \leq Var(Q) \tag{3.48}$$

La variable aléatoire $E[Q|C]$ peut s'écrire comme une fonction de la variable aléatoire C, de sorte qu'elle est en fait une fonction du vecteur aléatoire $\mathbf{X}$:

$$E[Q|C] = \psi(C) = \psi(g(\mathbf{X})) \tag{3.49}$$

En pratique, le conditionnement requiert de connaître une variable aléatoire C pour laquelle une expression de $E[Q|C]$ est connue, c'est à dire connaître

la fonction ψ.

3.3.2.1.4 Stratification La stratification consiste à partitionner l'espace en M événements disjoints $(\Omega_i)_{1\leq i\leq M}$, chacun de probabilité non nulle p_i, que l'on appelle strates. Cela permet d'exprimer l'espérance de la quantité Q sous la forme suivante :

$$E\left[Q\right] = E\left[\sum_{i=1}^{M} Q\mathbb{1}_{\Omega_i}\right] = \sum_{i=1}^{M} p_i E\left[Q|\Omega_i\right] \tag{3.50}$$

Ainsi, le calcul de l'espérance de Q revient à un calcul de M espérances, que sont les calculs des espérances des variables aléatoires $Q^{\Omega_i} = (Q|\Omega_i)$. Ces espérances peuvent chacune être estimées à l'aide d'une méthode de Monte Carlo impliquant n_i échantillons :

$$E\left[Q|\Omega_i\right] \approx \frac{1}{n_i}\sum_{k=1}^{n_i} Q_k^{\Omega_i} \tag{3.51}$$

L'estimation de l'espérance de la variable aléatoire Q est alors donnée par l'expression suivante :

$$E\left[Q\right] \approx \sum_{i=1}^{M} \frac{p_i}{n_i}\sum_{k=1}^{n_i} Q_k^{\Omega_i} \tag{3.52}$$

On peut montrer que la variance de l'expression précédente a pour expression :

$$Var\left(\sum_{i=1}^{M} \frac{p_i}{n_i}\sum_{k=1}^{n_i} Q_k^{\Omega_i}\right) = \sum_{i=1}^{M} \frac{p_i^2}{n_i} Var\left(X^{\Omega_i}\right) \tag{3.53}$$

La réduction de variance consiste alors à choisir correctement les différents nombres d'échantillons n_i à considérer par strates, afin que la variance précédente soit inférieure à la variance de l'estimateur basique impliquant un nombre d'échantillons $n = \sum_{i=1}^{M} n_i$. Un tel choix est toujours possible en pratique, permettant donc une réduction de variance et une accélération de la méthode.

3.3.2.1.5 Variables antithétiques La méthode des variables antithétiques consiste, dans une de ces versions, à calculer l'espérance de la variable aléatoire

Q de la manière suivante :

$$\begin{aligned} E[Q] = E\left[\tilde{f}(\mathbf{U})\right] &= \int_{[0,1]^d} \tilde{f}(\mathbf{u})d\mathbf{u} \\ &= \frac{1}{2}\left(\int_{[0,1]^d} \tilde{f}(\mathbf{u})d\mathbf{u} + \int_{[0,1]^d} \tilde{f}(\mathbf{1}-\mathbf{u})d\mathbf{u}\right) \\ &= E\left[\frac{\tilde{f}(\mathbf{U}) + \tilde{f}(\mathbf{1}-\mathbf{U})}{2}\right] \end{aligned} \tag{3.54}$$

Plutôt que de calculer l'espérance de $Q = \tilde{f}(\mathbf{U})$, la méthode des variables antithétiques suggère de calculer l'espérance de la variable aléatoire $\hat{Q} = \frac{\tilde{f}(\mathbf{U})+\tilde{f}(\mathbf{1}-\mathbf{U})}{2}$. Afin d'avoir une amélioration de la méthode, il est nécessaire que la variance de $\hat{Q}$ soit inférieure à celle de Q. Or, la variance de $\hat{Q}$ est donnée par l'expression suivante :

$$\begin{aligned} Var\left(\hat{Q}\right) &= \frac{1}{2}\left[Var\left(\tilde{f}(\mathbf{U})\right) + Var\left(\tilde{f}(\mathbf{1}-\mathbf{U})\right) + 2Cov\left(\tilde{f}(\mathbf{U}), \tilde{f}(\mathbf{U})\right)\right] \\ &= Var(Q) + Cov\left(\tilde{f}(\mathbf{U}), \tilde{f}(\mathbf{U})\right) \end{aligned} \tag{3.55}$$

La réduction de la variance avec cette méthode a donc lieu à l'unique condition que la covariance des variables aléatoires $\tilde{f}(\mathbf{U})$ et $\tilde{f}(\mathbf{1}-\mathbf{U})$ est négative. Une condition suffisante pour que cette covariance soit négative est que la fonction $\tilde{f}$ soit monotone en chacune de ses variables.

3.3.2.1.6 Synthèse des méthodes de réduction de variance Les trois premières méthodes de réduction de variance présentées nécessitent une étude préalable du problème à étudier, afin d'être en mesure d'appliquer la méthode ou de faire un choix "intelligent" permettant une réduction effective et significative de la variance. A cela s'ajoute un travail d'implémentation supplémentaire et une complexité de calcul pouvant être plus importante. La méthode de stratification offre la possibilité d'avoir simplement une réduction de variance systématique, mais une étude préalable reste cependant nécessaire afin de déterminer une partition "intelligente" de l'espace des événements permettant une réduction significative de la variance. La méthode des variables antithétiques offre l'avantage d'être utilisable directement et simplement. Cependant, elle n'offre pas systématiquement une réduction de la variance, et il peut être difficile de savoir à priori si une réduction de variance aura lieu pour un problème donné rendant son utilisation moins intéressante.

Les méthodes de réduction de variance peuvent offrir une réduction du

coup de la méthode de Monte Carlo, mais nécessitent une connaissance préalable du problème étudié et n'ont, pour cette raison, pas été envisagées dans cette thèse. En fait, l'amélioration de la méthode de Monte Carlo peut être effectuée en s'intéressant à l'échantillonnage utilisé, qui peut lui aussi permettre une réduction de la variance comme cela était déjà visible avec la méthode des variables antithétiques, qui peut être vue comme un échantillonnage utilisant des paires anti-corrélées, et donc ne respectant plus la propriété d'indépendance. Un échantillonnage non indépendant classiquement utilisé est l'échantillonnage par hypercube latin.

3.3.2.2 Échantillonnage par hypercube latin

L'échantillonnage par hypercube latin [85] est une méthode permettant d'échantillonner l'hypercube $[0,1]^d$ de manière plus "homogène" qu'avec un échantillonnage aléatoire classique. Cette méthode permet d'obtenir une réalisation d'une famille de vecteurs aléatoires $(\mathbf{U_i})_{1\leq i\leq N}$ identiquement distribués suivant une loi uniforme sur l'hypercube $[0,1]^d$, mais n'étant pas indépendants entre eux. Afin d'obtenir une telle construction, chaque côté de l'hypercube $[0,1]^d$ est découpé en N intervalles de tailles égales $\left(\left[\frac{i}{N},\frac{i+1}{N}\right]\right)_{i\in[\![0,N-1]\!]}$. On partitionne ainsi l'hypercube $[0,1]^d$ dans chaque direction j en N parties $S_i^{(j)}$ de la forme :

$$S_{N,i}^{(j)} = [0,1]^{j-1} \times \left[\frac{i}{N},\frac{i+1}{N}\right] \times [0,1]^{d-j-1} \tag{3.56}$$

Toutes ces parties $S_i^{(j)}$ ont le même poids de $1/n$ par rapport à la mesure uniforme sur l'hypercube $[0,1]^d$. Une famille de n vecteurs aléatoires $(\mathbf{U_i})_{i\in[\![1,n]\!]}$ suivant un échantillonnage d'hypercube latin sera telle que :

$$\forall i \in [\![1,n]\!], \forall j \in [\![1,d]\!], \exists! k \in [\![1,n]\!], U_k \in S_i^{(j)} \tag{3.57}$$

Cela se traduit par le fait que la probabilité pour que U_k et U_l se retrouvent ensemble dans une partie $S_i^{(j)}$ est nulle. Cela assure donc une certaine homogénéité dans l'échantillonnage, puisqu'on assure avec N points que les projections de ces points sur les côtés de l'hypercube occuperont l'ensemble des intervalles de la forme $\left[\frac{i}{n},\frac{i+1}{n}\right]$. Bien entendu, il n'existe plus avec cette méthode l'indépendance entre les vecteurs aléatoires $(\mathbf{U_i})_{i\in[\![1,n]\!]}$. Une réalisation d'un échantillonnage par hypercube latin en dimension 2 est présenté sur la figure 3.13.

La construction d'un échantillonnage par hypercube latin nécessite de connaître au préalable le nombre n de points que l'on veut dans cet échantillon, afin de pouvoir découper chaque direction en n sous-intervalles.

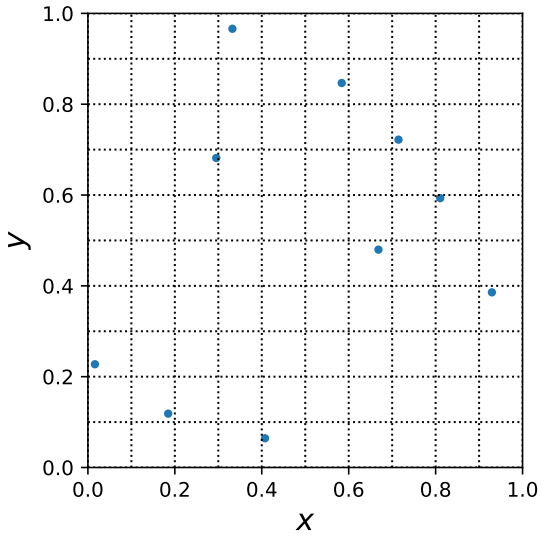

FIGURE 3.13 – *Exemple d'échantillonnage par hypercube latin contenant* 10 *points en dimension* 2.

Comme montré dans [85], si $(\mathbf{U_i})_{1\leq i\leq n}$ est une famille de variables aléatoires, chacune uniforme sur $[0,1]$, et vérifiant la propriété 3.57, on a presque sûrement la relation 3.58, où U est une variable aléatoire de loi uniforme sur $[0,1]$.

$$\lim_{n\to+\infty} \frac{1}{n}\sum_{i=1}^{n} f(U_i) = S_n^{LHS} = E[f(U)] \tag{3.58}$$

Il est également montré dans [85] que sous certaines conditions, la variance de la moyenne arithmétique S_n^{LHS} est plus faible que celle de la moyenne arithmétique S_n définie dans le cas de variables aléatoires i.i.d. suivant une loi uniforme sur $[0,1]$. Cela assure que dans un tel cas, l'utilisation d'un échantillonnage par hypercube latin assurera une convergence plus rapide que dans le cas de nombres aléatoires indépendants entre eux. Cependant, un des problème de cette méthode est que l'accès à la variance de l'estimateur S_n^{LHS}, qui est directement liée à la convergence de la méthode, ne peut pas se faire avec le seul échantillon de n points. On ne peut donc pas avoir d'informations concernant la convergence du calcul, tel qu'un intervalle de confiance par exemple, à moins de lancer plusieurs jeux d'échantillons par hypercube latin, ce qui fait perdre de l'intérêt à la méthode.

Concernant la construction d'un échantillon $(\mathbf{U_i})_{i\in[\![1,n]\!]}$ par hypercube latin, il est important que celui-ci doit assurer deux propriétés essentielles :

— les vecteurs $\mathbf{U_i}$ suivent bien une loi uniforme sur $[0,1]$

— la contrainte d'un seul point par strate $S_i^{(j)}$ est respectée

Afin d'assurer ces propriétés, la construction d'un échantillonnage par hypercube latin de n points en dimension d utilisent d permutations $(P_i)_{i\in[\![1,d]\!]}$ de l'ensemble $[\![1,n]\!]$. Chacune de ces permutations est choisie aléatoirement et de manière uniforme parmi les $n!$ permutations possibles. Ces permutations sont de plus choisis de manière indépendante. On considère également un échantillon $(\mathbf{Y_i})_{i\in[\![1,n]\!]}$ de vecteurs aléatoires i.i.d. de loi uniforme sur l'hypercube $[0,1]^d$.

Les vecteurs aléatoires $(\mathbf{U_i})_{i\in[\![1,n]\!]}$ définis par la relation suivante seront alors un échantillonnage par hypercube latin sur $[0,1]^d$:

$$\mathbf{U_i}_j = \frac{P_j(i)}{n} + \frac{\mathbf{Y_i}_j}{n} \tag{3.59}$$

Une des motivations au développement de l'échantillonnage par hypercube latin est qu'il permet de forcer une exploration "homogène" de chaque dimension de l'hypercube $[0,1]^d$, en partant du principe qu'apporter de l'information sur l'ensemble des d dimensions permettra d'avoir une meilleure approximation de l'intégrale multiple. Le bon fonctionnement de telle méthode d'intégration peut être reliée à la notion de dimension effective [72], qui sera mentionnée dans le chapitre 4. Cette notion d'homogénéité dans l'exploration de l'hypercube $[0,1]^d$ est ce qui a motivé en partie les méthodes de Quasi-Monte Carlo, qui permettent d'assurer une exploration de l'hypercube bien plus homogène que dans le cas d'un tirage aléatoire.

3.4 Méthode de Quasi-Monte Carlo

Les méthodes de Quasi-Monte Carlo permettent le calcul d'intégrales de grande dimension, qui sont typiquement définies sur l'hypercube $[0,1]^d$. Tout comme pour la méthode de Monte Carlo, l'intégrale d'une fonction f définie sur $[0,1]^d$ sera estimée par l'expression (3.60).

$$\int_{[0,1]^d} f(\mathbf{x})d\mathbf{x} \approx \frac{1}{n}\sum_{i=1}^{n} f(\mathbf{x}_i) \tag{3.60}$$

Dans le cas des méthodes de Monte Carlo, les points x_i de $[0,1]^d$ sont une réalisation d'un échantillon de variables aléatoires de distribution uniforme sur $[0,1]^d$, indépendantes ou non suivant qu'on considère un Monte Carlo standard ou un hypercube latin par exemple. Dans les méthodes de Quasi-Monte Carlo, les points x_i sont définis de manière déterministe, et il n'existe donc plus de caractère aléatoire. L'idée de ces méthodes est de remplir le plus "homogènement" possible l'hypercube $[0,1]^d$ à l'aide des points x_i.

3.4.1 Considérations théoriques

L'utilisation des méthodes de Quasi-Monte Carlo est motivée par l'inégalité de Koksma-Hlawka [52], qui stipule que pour une fonction f définie sur $[0,1]^d$ et possédant une variation de Hardy-Krause $V(f)$ finie et un ensemble de points $\{\mathbf{x_i} : 1 \leq i \leq n\}$, on a la relation (3.61).

$$\left| \int_{[0,1]^d} f(\mathbf{x}) d\mathbf{x} - \frac{1}{n} \sum_{i=1}^{n} f(\mathbf{x_i}) \right| \leq V(f) D^*(\mathbf{x_1}, \ldots, \mathbf{x_n}) \tag{3.61}$$

Dans l'expression (3.61), $D^*(\mathbf{x_1}, \ldots, \mathbf{x_n})$ est appelé la discrépance à l'origine de l'ensemble de points $\{\mathbf{x_i} : 1 \leq i \leq n\}$. Cette égalité étant valable pour n'importe quelle fonction f à variation de Hardy-Krause bornée, pour minimiser l'erreur commise, il convient de trouver une famille de points dont la discrépance à l'origine est la plus faible possible. La discrépance de l'ensemble de points $\{\mathbf{x_i} : 1 \leq i \leq n\}$ est définie par (3.62), dans laquelle # est la fonction cardinal d'un ensemble.

$$D^*(\mathbf{x_1}, \ldots, \mathbf{x_n}) = \sup_{\mathbf{u} \subset [0,1]^d} \left| \frac{\#\{i : \forall j, x_{i,j} \leq u_j\}}{n} - \prod_{j=1}^{d} u_j \right| \tag{3.62}$$

Pour calculer la discrépance à l'origine d'un ensemble de points, on s'intéresse donc à l'écart en valeur entre le volume d'un parallélépipède dont l'origine est un coin et dont les faces sont parallèles aux axes et la proportion des points dans ce parallélépipède. Cet écart est calculé pour l'ensemble de tous les parallélépipèdes possibles, et l'écart maximum est la discrépance à l'origine de cet ensemble de points. Ainsi, en minimisant la discrépance à l'origine de l'ensemble $\{\mathbf{x_i} : 1 \leq i \leq n\}$, on cherche donc à ce que le nombre de points présents dans n'importe quel pavé, dont les faces sont parallèles aux plans de coordonnées, délimité par l'origine et un point $\mathbf{u}$ soit "proportionnel" au volume de ce pavé, ce qui est bien une façon de traduire le fait que l'ensemble $\{\mathbf{x_i} : 1 \leq i \leq n\}$ remplit l'hypercube $[0,1]^d$ de manière homogène.

Une conjecture non encore démontrée pour $d > 2$ est qu'il existe une borne inférieure pour la discrépance à l'origine d'un ensemble $\{\mathbf{x_i} : 1 \leq i \leq n\}$ [72], donnée par la relation (3.63).

$$D^*(\mathbf{x_1}, \ldots, \mathbf{x_n}) \geq B_d \frac{(\ln n)^{d-1}}{n} \tag{3.63}$$

Cette borne inférieure supposée a motivé la dénomination d'ensembles de points $\{\mathbf{x_i} : 1 \leq i \leq n\}$ à discrépance faible, lorsque ceux-ci suivent une

majoration de la forme de celle de (3.64).

$$D^*(\mathbf{x_1}, \ldots, \mathbf{x_n}) \leq A\frac{(\ln n)^d}{n} \tag{3.64}$$

A titre de comparaison, la discrépance d'une séquence de n points se répartissant sur un maillage uniforme est de l'ordre de $n^{-\frac{1}{d}}$ [72], alors que la discrépance d'une séquence de n points tirés aléatoirement de manière uniforme et indépendante dans l'hypercube unité $[0,1]^d$ est elle de l'ordre de $\sqrt{\ln(\ln(n))}/\sqrt{n}$ [72]. L'échantillonnage aléatoire devient donc plus intéressant asymptotiquement que l'échantillonnage sur un maillage uniforme pour $d > 2$, et sa discrépance est indépendante de la dimension d, ce qui fait écho au fait que les résultats théoriques de la loi des grands nombres ou du TCL sont indépendants de la dimension stochastique d. Les bornes conjecturées pour la discrépance montrent bien qu'il y a la possibilité d'avoir des séquences de points qui présentent une discrépance asymptotique plus faible que dans le cas purement aléatoire, et cela quelque soit la valeur de la dimension.

Il existe différentes manières de construire des ensembles possédant une discrépance faible. Une manière consiste à utiliser des treillis [72], mais le nombre de points n est alors fixé, sans possibilité de le changer sans changer de treillis. D'autres manières consistent à construire des séquences de points, dont il a été montré qu'elles possèdent une discrépance faible. Le nombre de points de ces séquences n'est pas déterminé à l'avance et il est donc possible d'augmenter le nombre de point si nécessaire pour obtenir une meilleure convergence de l'intégrale à calculer.

3.4.2 Quelques suites à discrépance faible

Les constructions des séquences à discrépance faible qui vont être présentées sont telles que la construction du i-ème point de la séquence repose sur la décomposition de l'entier $i-1$ dans une base b. Ces construction reposent sur la fonction g_b définie, pour un entier i, par l'expression suivante :

$$g_b(i) = \sum_{l=0}^{+\infty} a_l(i) b^{-l-1}, \text{ avec } i = \sum_{l=0}^{+\infty} a_l(i) b^l \tag{3.65}$$

Il est immédiat que $g_b(i)$ est toujours compris entre 0 et 1, et les 16 premiers points de g_2, g_3, g_5 et g_7 sont présentés sur la figure (3.14).

Les points présentés sur la figure (3.14) correspondent aux suites de van der Corput [72] pour les 4 premiers entiers premiers, qui sont des suites à discrépance faible en dimension 1. La séquence de Halton généralise cette construction à une dimension plus grande.

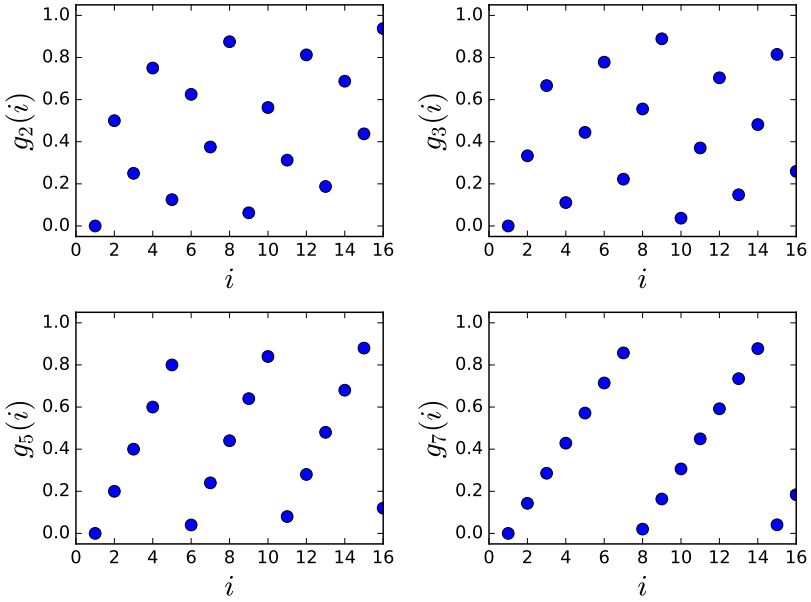

FIGURE 3.14 – *Valeurs des* 16 *premiers points pour, de droite à gauche et de haut en bas,* g_2, g_3, g_5 *et* g_7.

3.4.2.1 Séquence de Halton

La séquence de Halton [49] est une construction simple en dimension d, où une base b différente est utilisée pour chacune des d coordonnées. En effet, prendre une base identique b pour deux coordonnées différentes k et l impliquerait que les k-èmes et l-ièmes coordonnées de tous les points de la séquence seraient toutes identiques. La projection de la séquence sur le plan de coordonnées définies par les k-ième et l-ième dimensions se situera donc sur la diagonale du carré unité dans ce plan de coordonnée, ce qui empêchera donc un bon remplissage de l'hypercube unité $[0,1]^d$. Pour assurer de bonnes propriétés à la séquence, les nombres utilisés comme bases pour les différentes coordonnées doivent être premier entre eux. Pour cela, la séquence de Halton en dimension d est généralement basée sur les d premiers nombres premiers.

La suite de Halton possède une discrépance en $O\left(\frac{(\ln n)^d}{n}\right)$ [49]. Cependant, en grande dimension notamment, le remplissage de l'hypercube unité peut se révéler insatisfaisant en pratique, la taille de la séquence étant trop faible pour que le résultat asymptotique précédent se fasse sentir. En effet, comme montré sur la figure 3.15, les projections des points de la séquence sur certains plans de coordonnées ont tendance à se placer sur la diagonale du carré unité.

Le phénomène d'accumulation des projections des premiers points proches des diagonales est dû à la valeur très proche des bases b utilisées entre différentes dimensions. Pour empêcher ce phénomène, d'autres méthodes pour la

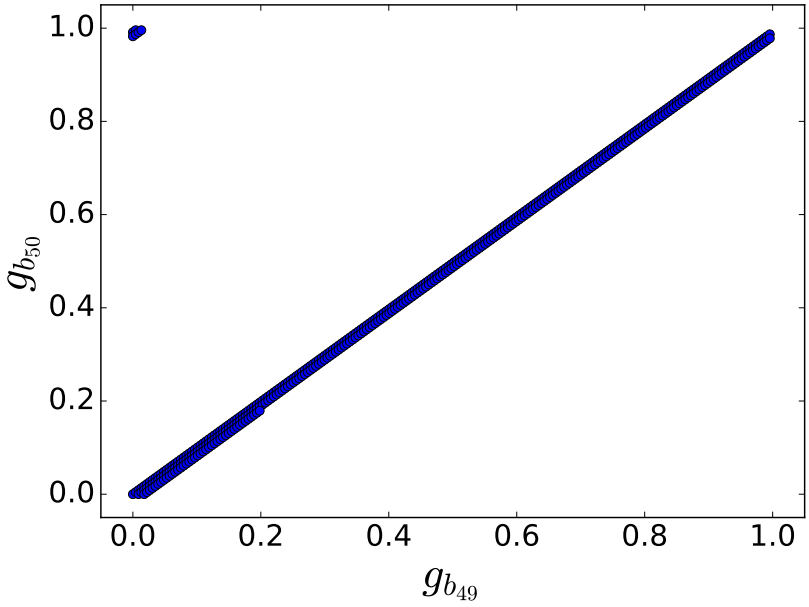

FIGURE 3.15 – 500 *premiers points de la suite de Halton pour les dimensions* 49 *et* 50 *possédant les bases* $b_{49} = 227$ *et* $b_{50} = 229$. *Du fait de la valeur très proches des deux bases, les premiers points de la séquence se concentrent sur la diagonale.*

définition des points doivent être employées, pouvant autoriser la même base b pour chaque coordonnée. De telles séquences, utilisant la même base b pour l'ensemble des coordonnées, sont appelées "digital sequences" en anglais, la séquence de Sobol étant l'une d'elles.

3.4.2.2 Séquence de Sobol

La séquence de Sobol est telle que le i-ème point implique l'utilisation de la décomposition en base $b = 2$ de l'entier $i - 1$ pour chaque coordonnée. Comme dit précédemment, en utilisant la même base b pour chaque coordonnée, il y a le risque d'avoir l'ensemble des points sur la diagonale de l'hypercube, ou d'avoir certaines projections de ces points sur les diagonales de sous hypercubes. Pour éviter cela, et afin d'avoir le meilleur remplissage possible de l'hypercube, il est important de transformer de manière différente et, le plus intelligemment possible, chacune des coordonnées. La transformation opérée sur chacune des coordonnées est décrite par (3.66), où $u_{i,j}$ est la j-ème coordonnée du i-ème point $\mathbf{u_i}$ de la séquence de Sobol.

$$u_{i,j} = \sum_{l=0}^{+\infty} \tilde{a}_{j,l}(i-1)b^{-l-1} \tag{3.66}$$

Dans l'expression (3.66), les chiffres $\tilde{a_{j,l}}(i-1)$ sont obtenues à l'aide des chiffres $a_j(i-1)$ présents dans la décomposition de $i-1$ en base 2 présentée dans l'expresson (3.65), comme une transformation linéaire de ceux-ci donnée par l'expression (3.67), les opérations prenant place dans le corps $\mathbb{Z}/2\mathbb{Z}$.

$$\tilde{a}_{j,l}(i-1) = \sum_{k=0}^{+\infty} c_{l,k} a_k(i-1) \tag{3.67}$$

La relation (3.67) peut être réécrite sous forme matricielle, impliquant pour chaque coordonnée j une matrice C_j de dimension ∞ dont les coefficients sont dans $\mathbb{Z}/2\mathbb{Z}$, de sorte que l'on a, en notant $A(i-1)$ le vecteur colonne contenant les $a_l(i-1)$, et $\tilde{A}_j(i-1)$ le vecteur colonne contenant les $\tilde{a}_{j,l}(i-1)$, la relation matricielle (3.68).

$$\tilde{A}_j(i-1) = C_j A(i-1) \tag{3.68}$$

En pratique, la dimension des matrices C_j n'est bien évidemment pas infinie, et il suffit de prendre comme dimension pour ces matrices l'entier k où $n = 2^k$ correspond au nombre maximum de points que l'on considère dans les séquences. En effet, les vecteurs $A(i-1)$ pour $i \leq 2^k$ contiennent nécessairement des 0 à partir de la ligne $k+1$, impliquant qu'il est inutile de considérer les coefficients des C_j avec des indices de lignes ou de colonnes supérieurs à k. Les matrices C_j sont appelées les matrices génératrices de la séquence de Sobol et définissent totalement les coordonnées des points de la séquence de Sobol. La construction de ces matrices repose pour chacune des coordonnées j, sur la donnée d'un polynôme primitif, aussi appelé polynôme minimal, P_j dans $\mathbb{Z}/2\mathbb{Z}$, de degré d_j, et de d_j nombres directionnels $(v_{r,j})_{1 \leq r \leq d_j}$ vérifiant la relation 3.69.

$$v_{r,j} = \frac{m_{r,j}}{2^r}, 0 \leq m_{r,j} \leq 2^r - 1, m_{r,j} \equiv 1[2] \tag{3.69}$$

A partir de ces nombres peuvent être construites les matrices génératrices C_j, comme explicité dans [16]. Le choix des polynômes primitifs ainsi que des nombres directionnels est essentiel pour avoir une séquence de Sobol de "bonne qualité". Un exemple de séquence de Sobol en dimension 2 est donné par la figure 3.16. Plusieurs critères peuvent être considérés pour choisir ceux-ci. Une propriété historique introduite par Sobol [127] sous la dénomination de 'Propriété A', impose une certaine homogénéité des points de la séquence. Une séquence vérifie la propriété A si, lorsque la séquence est découpée en blocs consécutifs de 2^d points, les 2^d points de ces blocs se répartissent uniformément dans les 2^d sous hypercubes identiques de l'hypercube de $[0,1]^d$. En pratique, cette propriété n'a que peu d'intérêt en grande dimension, puisque placer un

point dans chacun des 2^d sous hypercubes de $[0,1]^d$ est irréalisable en pratique en grande dimension.

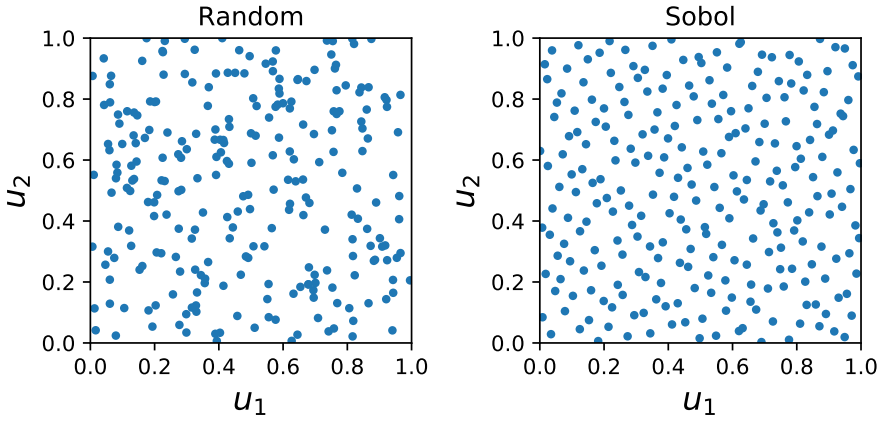

FIGURE 3.16 – *Gauche :* 256 *points tirés uniformément et indépendamment sur* $[0,1]^2$. *Droite :* 256 *premiers points d'une séquence de Sobol dans* $[0,1]^2$.

Dans [54], le choix est fait d'établir une relation d'ordre sur l'ensemble des polynômes primitifs de $\mathbb{Z}/2\mathbb{Z}$, ce qui ne laisse plus que la détermination des nombres directionnels à opérer. La détermination de ces nombres directionnels est effectuée à l'aide de l'optimisation d'un critère, qui vise à avoir l'ensemble des projections $2D$ des points de la séquence qui soit la plus uniforme possible, tout en ayant la propriété A, pour un nombre de points allant jusqu'à 2^{32}, et cela jusqu'à une dimension $d = 21101$, ce qui est plus que suffisant pour les applications considérées dans cette thèse. La bonne uniformité des projections $1D$ est assurée par la construction même des séquences de Sobol, alors que celle des projections $2D$ dépend du choix des polynômes primitifs et des nombres directionnels, qui peuvent mener à de mauvaises projections $2D$ comme montré dans [54]. Le choix de n'optimiser que les nombres directionnels pour les projections $2D$ ne garantit pas d'avoir des projections d'ordre supérieur de bonne qualité en terme d'uniformité. En pratique, les problèmes rencontrés ont souvent une dimension effective faible [72]. La faible dimension effective implique que la fonction peut être approximée par une somme de fonctions ne dépendant que de combinaison d'un sous-ensemble des variables initiales de petites tailles, et l'idée est que ces fonctions seront particulièrement bien intégrés si les projections des points de la séquence de Sobol sur l'hyperplan de coordonnées correspondant à ces sous-ensemble de variables est de bonne qualité. Un autre aspect de la construction de Joe et Kuo [54] est qu'il donne une légère, mais tout de même plus grande, importance aux premières dimensions lors de la détermination des nombres directionnels. En fait, les premières directions ont d'ores et déjà une plus grande importance de par le choix d'ordonnancement des polynômes primitifs, puisqu'elles correspondent aux polynômes primitifs de

plus petit degré, et que la qualité des projections de la séquence de points est liés à ces degrés [126]. Dans des cas où le nombre de variables est très important, il peut donc être souhaitable de placer les variables de plus grande importance en première, en opérant une permutation sur les variables d'entrée.

Les séquences de Sobol, tout comme celle de Halton ou toute méthode impliquant un ensemble de point à discrépance faible, permettent dans de nombreux cas d'obtenir une approximation d'une intégrale meilleure qu'avec la simple méthode de Monte Carlo basée sur des tirages de points aléatoires. En revanche, la méthode de Monte Carlo offre l'avantage d'obtenir une estimation de l'erreur commise ce qui n'est pas le cas avec les méthodes de Quasi-Monte Carlo. Il est cependant possible de randomiser les méthodes de Quasi-Monte Carlo afin d'obtenir une estimation de l'erreur commise.

3.4.3 Randomisation des suites à discrépance faible

L'objectif de la randomisation des suites à discrépance faible est d'obtenir une estimation de l'erreur commise à l'aide des considérations présentées dans la section 3.3. Pour faire cela, l'idée est de construire m suites à discrépances faibles $(x_{i,j})_{1\leq i\leq n, 1\leq j\leq m}$ de n points chacune, chacune étant une version randomisée indépendantes des autres, de la suite à discrépance faible originale considérée, de Halton ou de Sobol par exemple. Pour chacune de ces m suites, on peut évaluer l'intégrale d'une fonction f par l'expression 3.60, et obtenir ainsi m estimations $(I_{n,j})_{1\leq j\leq m}$ de cette intégrale. L'ensemble de ces estimations étant obtenues de manière aléatoire similairement et indépendamment les unes des autres, on peut considérer que les $I_{n,j}$ sont des réalisations de variables aléatoires indépendantes et identiquement distribuées, dont on peut estimer la variance σ_n^2 par l'expression (3.70).

$$\sigma_n^2 \approx \frac{1}{m-1}\sum_{j=1}^{m}(I_{n,j} - I_{mn})^2 \tag{3.70}$$

Dans l'expression (3.70), I_{mn} correspond à la moyenne arithmétique des $I_{n,j}$, qui est l'estimateur de l'intégrale de f obtenue à l'aide de la randomisation de la méthode de Quasi-Monte Carlo. I_{mn} étant une moyenne arithmétique de variables aléatoires indépendantes et identiquement distribuées, il est possible d'obtenir des intervalles de confiance sur la quantité à estimer en lui appliquant les résultats du TCL dont un ingrédient est la variance σ_{mn} de I_{mn}, qui vérifie la relation (3.71).

$$\sigma_{mn}^2 = \frac{1}{m}\sigma_n^2 \tag{3.71}$$

Cette randomisation des suites à discrépances faibles permet donc d'obte-

nir une estimation de l'erreur commise par les méthode de Quasi-Monte Carlo à travers l'écart-type grâce à la construction d'intervalles de confiance pour la quantité à estimer.

Pour pouvoir construire un estimateur tel que I_{mn}, il est essentiel de savoir randomiser correctement les suites à discrépance faible afin que les variables aléatoires $I_{n,j}$ possèdent les caractéristiques adéquates. Tout comme il existe plusieurs constructions possibles de suites à discrépance faible, il existe plusieurs techniques de randomisation de suites à discrépance faible. Une technique de randomisation n'est pas nécessairement applicable à l'ensemble des suites à discrépance faible, et le choix d'une technique de randomisation est en fait intimement liée à la suite à discrépance faible utilisée. Deux conditions sont importantes pour les techniques de randomisation de suites à discrépance faible. La première, essentielle, est que chaque point $\mathbf{U_i}$ de la séquence randomisée suive une loi de probabilité uniforme sur l'hypercube unité $[0,1]^d$. Cela assure que l'estimateur I_n ne soit pas biaisé, comme montré par la relation (3.72).

$$E[I_n] = \frac{1}{n}\sum_{i=1}^{n} E[f(\mathbf{U_i})] = \frac{1}{n}\sum_{i=1}^{n}\int_{[0,1]^d} f(\mathbf{u})d\mathbf{u} = \int_{[0,1]^d} f(\mathbf{u})d\mathbf{u} \qquad (3.72)$$

La seconde propriété importante est que la technique de randomisation conserve la discrépance faible de la séquence de points, c'est à dire que les propriétés concernant la discrépance faible d'une réalisation de la suite à discrépance faible randomisée doivent être les mêmes que celles de la suite à discrépance faible déterministe associée. Cette condition est essentielle pour ne pas affecter la convergence de la suite à discrépance faible lors de la randomisation, et ne pas se retrouver dans un cas similaire au Monte Carlo classique.

Différentes méthodes de randomisation sont présentées dans les paragraphes suivants, ainsi que certains résultats de convergence des méthodes de Quasi-Monte Carlo randomisé lors de l'utilisation de ces méthodes de randomisation.

3.4.3.1 "Shift" et "Digital Shift"

Comme dit précédemment, différentes techniques de randomisation existent. La plus simple est de considérer une translation, modulo 1 sur chaque composante afin de rester dans l'hypercube unité $[0,1]^d$, de l'ensemble des points $\mathbf{u_i}$ de la séquence, à l'aide d'un vecteur aléatoire $\mathbf{V}$ suivant une loi uniforme sur $[0,1]^d$. Cette technique est très simple à mettre en place, mais ne conserve pas nécessairement certaines propriétés, comme la propriété A de Sobol pour la séquence de Sobol par exemple [72]. Une méthode un peu plus avancée est celle nommée "digital shift" [72] en anglais, qui consiste encore une fois à se servir d'un vecteur aléatoire $\mathbf{V}$ suivant une loi uniforme sur $[0,1]^d$, mais plutôt que de translater les points $\mathbf{u_i}$ de la séquence à l'aide de ce vecteur comme dans la méthode précédente, on se sert des décompositions en base b_j de chacune des

composantes $u_{i,j}$ et V_j de $\mathbf{u_i}$ et $\mathbf{V}$, pour effectuer l'opération (3.73) qui permet d'obtenir la j-ième composante $U_{i,j}$ du point de la séquence randomisée, l'addition étant réalisée dans $\mathbb{Z}/b_j\mathbb{Z}$.

$$U_{i,j} = \sum_{l=1}^{+\infty} (u_{i,j,l} + V_{j,l}) b^{-l} \tag{3.73}$$

Cette transformation conserve notamment la propriété A de Sobol pour les séquences de Sobol. En fait, dans le cas particulier de la séquence de Sobol, où les b_j sont tous égaux à 2, cela revient en fait à appliquer une permutation de $\mathbb{Z}/2\mathbb{Z}$ à chaque chiffre de la décomposition en base 2, comme il a été fait pour améliorer la séquence de Halton. En effet, les permutations de $\mathbb{Z}/2\mathbb{Z}$ sont au nombre de 2, et l'addition de 0 dans $\mathbb{Z}/2\mathbb{Z}$ correspond à la permutation identité, alors que l'addition de 1 dans $\mathbb{Z}/2\mathbb{Z}$ correspond à la seconde permutation. Le vecteur aléatoire $\mathbf{V}$ suivant une loi uniforme sur $[0,1]^d$, cela revient en fait à appliquer une permutation choisie aléatoirement et de manière uniforme à chaque chiffre de la décomposition, ces permutations étant de plus indépendantes les unes des autres. Cette méthode consistant à appliquer à chacun des chiffres de la décomposition en base b_j de chaque composante des points de la séquence un ensemble de permutations, ayant été choisies aléatoirement, uniformément et indépendamment les unes des autres dans l'ensemble des permutations de $\mathbb{Z}/b_j\mathbb{Z}$ est appelée "random digit scrambling" [84].

Suivant l'idée d'utiliser des permutations sur les chiffres des décompositions en base b_j des composantes des points de la séquence, différentes méthodes ont été développées, qui sont associées à de nouveaux résultats de convergence pour l'estimation de l'intégrale à l'aide d'une méthode de Quasi-Monte Carlo randomisée.

3.4.3.2 "Full scrambling"

L'utilisation de permutations sur les chiffres de la décomposition en base b a permis l'émergence d'un ensemble de méthodes de randomisation, dénommée "scrambling" ou "nested scrambling" en anglais. Celles-ci consistent également à appliquer à chaque chiffre $u_{i,j,l}$ de la décomposition en base b_j de $u_{i,j}$ une permutation dans $\mathbb{Z}/b_j\mathbb{Z}$. Cependant, contrairement au "random digit scrambling", les permutations appliquées à chaque chiffre $u_{i,j,l}$ de la décomposition de $u_{i,j}$ dépendent des chiffres $u_{i,j,l'}$ présents dans la décomposition avant $u_{i,j,l}$, c'est à dire vérifiant $l' < l$. Formellement, la version randomisée $U_{i,j}$ de $u_{i,j}$ aura dans sa décomposition en base b_j les chiffres $U_{i,j,l}$, données par (3.74).

$$U_{i,j,l} = \pi^{i,j}_{u_{i,j,1},\ldots,u_{i,j,l-1}}(u_{i,j,l}) \tag{3.74}$$

Dans l'expression (3.74), la permutation $\pi^{i,j}_{u_{i,j,1},\ldots,u_{i,j,l-1}}$ devant être appliquée au l-ième chiffre de la décomposition de $u_{i,j}$ dans la base b_j dépend des chiffres précédents $u_{i,j,1},\ldots,u_{i,j,l-1}$ présents dans cette décomposition, comme spécifié par l'indiçage de celle-ci. Différentes méthodes de "scrambling" existent [95, 84, 138, 131], différant par leur complexité.

La plus complexe et originalement proposée par Owen [95], dénommée "full scrambling" en anglais, consiste à ne faire aucune restriction en choisissant aléatoirement, uniformément et indépendamment, l'ensemble des permutations $\pi^{i,j}_{u_{i,j,1},\ldots,u_{i,j,l}}$. Cette randomisation plus complexe permet, sous réserve d'une certaine condition de régularité de la fonction à intégrer, d'avoir la variance de l'estimateur I_n de l'ordre de $n^{-3}\ln(n)^{d-1}$, pour un nombre de points dans la séquence $n = b^k$ avec k un entier [96].

Une telle méthode de randomisation est coûteuse, en mémoire et/ou en calcul suivant que l'on décide de stocker ces permutations ou de les construire à la volée. Pour illustrer ce fait, considérons les 2^k premiers points d'une séquence de Sobol que l'on souhaite randomiser. Une propriété de la séquence de Sobol provenant du fait que toutes les matrices génératrices C_j sont triangulaires supérieures non singulières [72], est que la décomposition en base 2 de chaque composante j aura tous ses chiffres à partir du $k+1$-ième compris nuls, et les autres $((u_{i,j,1},\ldots,u_{i,j,k})_{i\in[\![0,2^k-1]\!]})$ parcourront l'ensemble des 2^k possibilités de $\{0,1\}^k$. Ainsi, il y aura la nécessité d'avoir un nombre de permutation par composante de 2^k-1, donné par (3.75).

$$\sum_{l=1}^{k-1} 2^l = 2^k - 1 \tag{3.75}$$

En plus de cela, une permutation doit également être appliquée à chaque chiffre $u_{i,j,l}$ avec $l > k$. Du fait que l'ensemble des chiffres $u_{i,j,l}$ sont nuls pour $l > k$, le l-ième chiffre $U_{i,j,l}$ de la décomposition de $U_{i,j}$ sera donné par (3.76).

$$U_{i,j,l} = \pi^{i,j}_{\underbrace{u_{i,j,1},\ldots,u_{i,j,k},0,\ldots,0}_{l-1 \text{ chiffres}}}(u_{i,j,l}) = \pi^{i,j}_{\underbrace{u_{i,j,1},\ldots,u_{i,j,k},0,\ldots,0}_{l-1 \text{ chiffres}}}(0) \tag{3.76}$$

L'expression (3.76) est vraie pour tous les chiffres de la décomposition de $U_{i,j}$ strictement après le k-ième, et l'ensemble de ces chiffres ne dépend que des valeurs prises par les k chiffres $(u_{i,j,1},\ldots,u_{i,j,k})$, de sorte que le reste $R_{i,j,k}$ défini par (3.77), ne dépend lui aussi que de la valeur de ces k chiffres.

$$R_{i,j,k} = \sum_{l=k+1}^{+\infty} U_{i,j,l} 2^{-l} \tag{3.77}$$

De plus, l'indépendance et l'uniformité du tirage de l'ensemble des permutations $u_{i,j,1}, \ldots, u_{i,j,k}, 0, \ldots, 0$ font que les $R_{i,j,k}$ sont des nombres aléatoires indépendants entre eux et suivant une loi uniforme sur $[0, 2^{-k}]$. Il faut donc, pour une composante j, en plus des $2^k - 1$ permutations, générer ou stocker 2^k nombres flottants tirés aléatoirement et indépendamment entre 0 et 2^{-k} pour avoir une réalisation d'une randomisation par la méthode de "full scrambling" d'une séquence de Sobol de 2^k points.

Comme dit précédemment, un choix possible est de stocker toutes ces informations préalablement. Dans le cas où l'on veut appliquer une méthode de Quasi-Monte Carlo en dimension d, impliquant m réalisations de séquence de Sobol randomisées par une méthode de "full-scrambling", et où un maximum de 2^k points par séquence sont générés, on a donc un nombre total $N_{d,m,k}^{perm}$ de permutations de $\mathbb{Z}/2\mathbb{Z}$ à utiliser donné par (3.78).

$$N_{d,m,k}^{perm} = md(2^k - 1) \tag{3.78}$$

En plus de cela, le nombre $N_{d,m,k}^{reste}$ de nombres aléatoires correspondant aux $R_{i,j,k}$ à utiliser est lui donné par (3.79).

$$N_{d,m,k}^{reste} = md2^k \tag{3.79}$$

Ces deux nombres sont donc sensiblement égaux, et croissent proportionnellement avec la dimension d et le nombre de points $n = 2^k$ utilisés dans les séquences. Pour une application où l'on prendrait une valeur, raisonnable mais minimale, de $m = 10$, en dimension $d = 100$, et où $k = 15$ de sorte que le nombre maximum de points dans la séquence serait $n = 32768$, le nombre de permutations et de nombres flottants à stocker serait de 32768000. Dans le cas de la séquence de Sobol, la base étant 2, les permutations peuvent être stockées sur un bits, de sorte que la quantité d'information à stocker pour les permutations serait de $4Mo$ dans cet exemple, alors que pour les flottants, en considérant que ceux-ci soit en double précision, et prennent chacun 8 octets, cela donnerait une quantité d'information de $250Mo$. On a donc un total de $254Mo$ d'espace nécessaire pour cet exemple. Si l'on venait à simplement doubler la valeur de d, ainsi que celle du nombre de points, la quantité nécessaire serait supérieur à $1Go$, et commence donc à atteindre les limites de la mémoire disponible par cœur sur certains calculateurs, d'autant plus qu'il peut être nécessaire d'avoir également de la mémoire dédiée à l'évaluation de la fonction par exemple. Dans le cas où la base b utilisé n'est plus 2, les nombres évoqués explose complètement, puisqu'on aurait alors besoin de mdb^k permutations et nombres flottants, les permutations nécessitant elles $\log_2(b!)$ bits pour être représentées.

Pour ne pas souffrir d'une grande consommation de mémoire, différentes

approches peuvent être envisagées. Dans une implémentation parallèle suivant le paradigme maître-esclave, où les processus esclaves réalisent l'évaluation de la fonction à intégrer et le processus maître construit les estimateurs, une solution est de faire générer les points de la séquence par le processus maître, qui serait seul consommateur de la mémoire de stockage de ces permutations et de ces nombres flottants. Néanmoins, cela alourdirait la charge de travail ainsi que les communications réalisées par le processus maître, ce qui pourrait être préjudiciable à la scalabilité du programme. Une autre méthode consiste à générer les permutations ainsi que les nombres flottants à la volée. Cette méthode n'utilise presque pas de mémoire, mais est en revanche plus coûteuse en temps de calcul. Néanmoins, le fait qu'il soit possible de faire des "sauts" rapide au sein du générateur de nombres aléatoires permet néanmoins de rendre cette méthode plus rapide, puisqu'il n'est pas nécessaire de parcourir complètement la séquence de plus de $md2^k$ nombres aléatoires nécessaires pour générer l'ensemble des permutations et nombres flottants nécessaires à cette méthode de randomisation. La stratégie utilisée dans l'implémentation du code est celle gourmande en mémoire, consistant à calculer pour chacun des processus l'ensemble des informations en début de calcul et à les stocker.

Le "full scrambling" permet d'obtenir des résultats de convergence plus intéressant que pour un simple "digital shift" pour des fonctions suffisamment régulière, l'estimateur I_n possédant alors une variance de l'ordre de $n^{-3}\ln(n)^{d-1}$ lorsque le point n est une puissance de 2. Cependant, il demande une implémentation complexe ainsi qu'un coût plus important en mémoire et en coût de calcul. Le résultat concernant la plus rapide convergence n'est en fait pas propre à la méthode de "full scrambling", et peut donc être obtenu par des méthodes plus simples.

3.4.3.3 "Linear scrambling"

Le coût important de la méthode de "full scrambling" a motivé le développement d'autres méthodes de "scrambling", modifiant suffisamment les points de la séquence pour permettre de conserver le résultat de la variance de l'estimateur I_n de l'ordre de $n^{-3}\ln(n)^{d-1}$ pour n une puissance de 2. En fait, il suffit que certaines propriétés soient présentes sur l'ensemble de permutations utilisées [84], qui sont vérifiées pour les méthodes dénommées "random linear scrambling". Celles-ci consistent à considérer des permutations de $\mathbb{Z}/b\mathbb{Z}$ construites comme des combinaisons linéaires dans $\mathbb{Z}/b\mathbb{Z}$ d'éléments de $\mathbb{Z}/b\mathbb{Z}$. Dans le cas de la séquence de Sobol, ces transformations consistent à considérer d matrices triangulaires inférieures inversibles R_j, et avec lesquelles vont être multipliées matriciellement à gauche les matrices génératrices C_j de la suite de Sobol. La relation (3.68) devient donc :

$$\tilde{A}_j(i-1) = R_j C_j A(i-1) \tag{3.80}$$

Le caractère triangulaire inférieur de chacune des matrices R_j est là pour traduire le caractère imbriqué de la méthode de scrambling, où la valeur du l-ième chiffre présent dans la décomposition en base b est permutée en fonction des valeurs des chiffres précédents. En terme de mémoire et de temps CPU, l'avantage d'une méthode de "linear scrambling" sur la méthode de "full scrambling", est immédiate, puisqu'il suffira de remplacer les matrices génératrices C_j originales par les matrices R_jC_j, calculées une fois au début du calcul, et le coût de génération des points de la séquence est ensuite sensiblement le même que pour la séquence déterministe originale. Concernant la construction des matrices R_j, différentes méthodes peuvent être utilisées et sont présentes dans la littérature [131, 94]. La méthode "I-binomial scrambling" [131] a été implémentée, pour laquelle les matrices R_j possèdent une unique valeur pour chacune des diagonales inférieures, ainsi que la méthode de "random linear scrambling", où l'ensemble des coefficients sous la diagonales sont choisis aléatoirement dans $\mathbb{Z}/b\mathbb{Z}$. Pour assurer que les points des séquences randomisées construites suivent bien une loi uniforme sur $[0,1]^d$, il est de plus nécessaire d'effectuer une opération de "digital shift" sur la séquence ainsi construite [84].

En pratique, si le nombre de points dans la séquence envisagée ne dépasse pas b^k, il suffit de construire le bloc supérieur gauche de dimension k des matrices C_j. En effet, seules les k premières composantes du vecteur colonne $A(i-1)$ seront possiblement non nulles, et donc seules les k première colonnes des C_j sont utiles, et celles-ci étant triangulaires supérieures, seules les k premières composantes des produit $C_jA(i-1)$ seront possiblement non nulles, impliquant que seules les k premières lignes des C_j sont utiles. En revanche, il est important de garder un bloc plus grand des matrices R_j. En ce qui concerne les colonnes, seules les k premières comptent, le produit $C_jA(i-1)$ ayant au plus des valeurs non nulles jusqu'à la k-ième composante. En revanche, le nombre de lignes à garder dans les R_j est plus important. Le caractère inversible des C_j et des R_j assurent que, à part pour $A(0)$ qui possèdent l'ensemble de ces composantes nulles et donc pour lequel les produits $R_jC_jA(0)$ renverront le vecteur nulle, au moins l'une des k première composantes de $A(i-1)$, et donc de $\tilde{A}_j(i-1) = R_jC_jA(i-1)$ sera non nulle, pour $1 < i \leq b^k$. De ce fait, on aura nécessairement que le nombre dont la décomposition en base b est constitué du vecteur $\tilde{A}_j(i-1)$ est supérieur à b^{-k}. En considérant une représentation flottante, et donc une arithmétique sur des nombres flottants, il y a une valeur ϵ telle que si x est un flottant quelconque plus petit que ϵ, alors $b^{-k} + x = b^{-k}$. Ainsi, il existe un entier l tel que $b^{-l+1} \geq \epsilon \geq b^{-l}$. Cet entier l est le nombre optimal de lignes à choisir pour les matrices R_j, assurant de bonnes propriétés de randomisation à celle-ci et une consommation mémoire et CPU optimale. Dans le cas de la séquence de Sobol, la base b utilisée est 2, et la valeur de l est donnée par $l = k + p$, où p est le nombre de bits définissant la précision de la représentation flottante. Par exemple, pour des flottants double précision suivant la norme IEEE 754, la valeur de p est 52.

3.4.3.4 Comparaison des méthodes de randomisations

Des comparaisons de ces différentes méthodes de randomisation ont été réalisées appliquées à la séquence de Sobol, pour une dimension $d = 100$ et pour l'ensemble des fonctions proposées par Genz [40], qui sont présentées dans l'annexe A. Parmi ces fonctions, seules les fonctions de type "Continuous" ainsi que les fonctions de type "Discontinuous" présentent des irrégularités, la première présentant une discontinuité de sa dérivé alors que la seconde présente une discontinuité. La convergence est obtenue en considérant des méthodes de Quasi-Monte Carlo randomisé impliquant $m = 10$ séquences de Sobol randomisées.

Sur la figure 3.17 sont représentées les évolutions de l'écart-type σ_n de l'estimateur I_n estimé grâce à la relation (3.70) pour les différentes méthodes de randomisations présentées précédemment, seules des valeurs de n correspondant à des puissances de 2 ayant été prises en compte. L'ensemble des courbes sont globalement des droites, ce qui a motivé le calcul des coefficients directeurs des régressions linéaires opérées sur les points dans le diagramme $\log - \log$ de ces figures. Ces coefficients directeurs sont répertoriés dans le tableau (3.2), une valeur α signifiant que la valeur de σ_n évolue proportionnellement à n^{α}.

Pour l'ensemble des fonctions, les méthodes de randomisation peuvent être classées en deux groupes : le premier concernant la méthode de "shift" et de "digital shift", et le second comportant les méthodes de "nested scrambling", à savoir la méthode de "I-binomial scrambling", la méthode de "random linear scrambling" et la méthode de "full scrambling". Pour les fonctions régulières, les méthodes de "shift" et de "digital shift" ont une convergence de l'ordre de n^{-1}, qui se trouve dégradée pour la fonction de type "discontinuous" présentant une discontinuité. Les méthodes impliquant un "nested scrambling" présentent de meilleures résultats pour les fonctions régulières, avec une convergence de l'ordre de $n^{-3/2}$ pour la fonction de type "gaussian". Pour les fonctions de type "oscillatory" ou encore "corner peak", la convergence est légèrement plus faible. Cela peut s'expliquer par la difficulté d'intégration de ces fonctions, qui implique que le régime asymptotique est plus long à atteindre. La présence d'irrégularités au niveau des fonctions à intégrer conduit à une dégradation de la vitesse de convergence, comme c'est le cas pour la fonction de type "discontinuous", la théorie ne garantissant pas une convergence aussi rapide pour de telles fonctions. Cette dégradation est d'autant plus marquée que l'irrégularité est importante, et une discontinuité dans la fonction entraîne que l'ensemble des méthodes possèdent les mêmes performances. Pour des grandes valeurs de n, les méthodes de "I-binomial scrambling" et de "random linear scrambling" semblent surpasser la méthode de "full scrambling" pour les fonctions de types "product peak" et "gaussian". Il semble que cet effet soit en fait dû à une mauvaise convergence de l'estimation de σ_n dû à un nombre m de séquences de Sobol randomisé trop faible, comme le suggèrera la figure (3.18).

En effet, un autre paramètre important lors de l'utilisation de la méthode

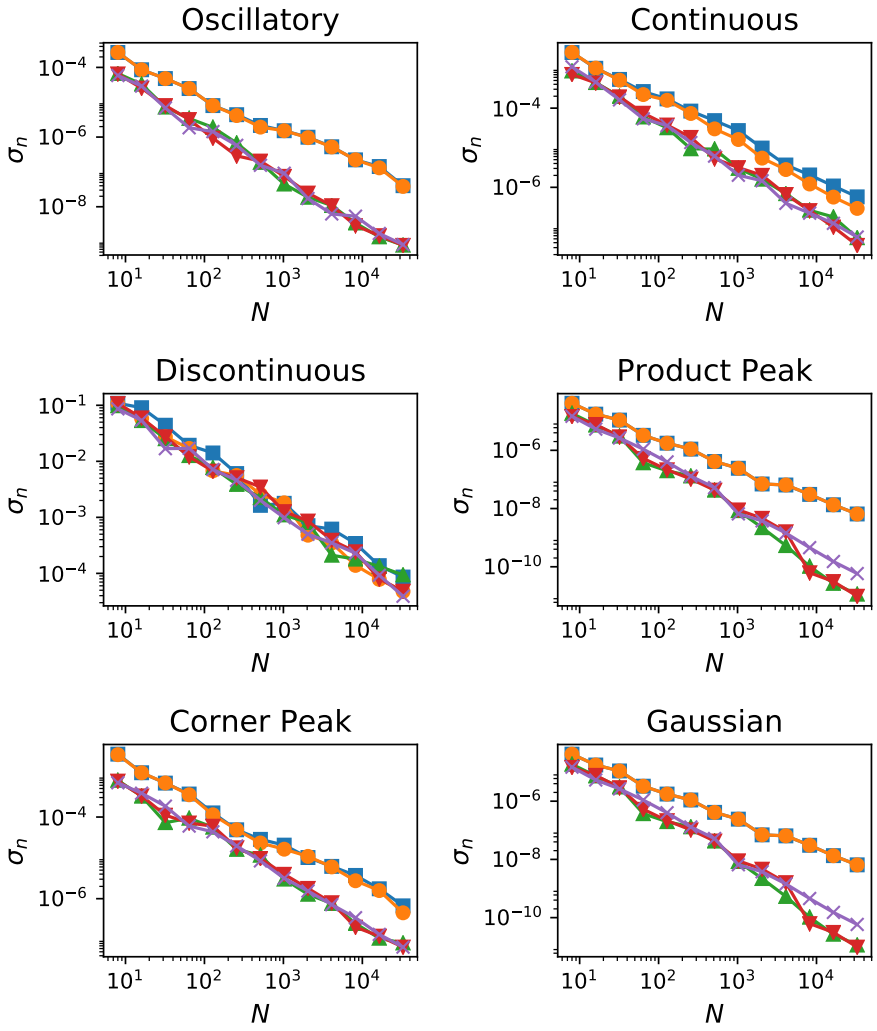

FIGURE 3.17 – *Estimations de σ_n en dimension 100 en fonction du nombre de points n utilisé dans les séquences de Sobol randomisées pour l'ensemble des fonctions test de Genz, les nombres n considérés étant tous de la forme $n = 2^k$ avec k entier. Les différentes méthodes de randomisations utilisées sont : "shift" (carrés), "digital shift" (ronds), "I-binomial scrambling" (triangle bas), "random linear scrambling" (pentagone) et "full scrambling" (croix).*

de Quasi-Monte Carlo randomisée est le nombre de réalisations m considérées pour estimer l'écart-type σ_n. Sur la figure 3.18 sont représentées des réalisations de l'estimateur de σ_n obtenues pour différentes valeurs de m, pour une valeur de n fixé à 4096. On observe pour l'ensemble des courbes qu'avec l'augmentation du nombre m de séquences de Sobol randomisées utilisées, les courbes se stabilisent et convergent vers une valeur, ce qui est le fruit de la consistance de l'estima-

	Shift	Dig. Shift	I Bin. Scr.	Rand. Lin. Scr.	Full Scr.
Oscillatory	-1.008	-1.018	-1.276	-1.286	-1.284
Continuous	-1.051	-1.079	-1.224	-1.220	-1.212
Discontinuous	-0.788	-0.831	-0.828	-0.820	-0.846
Product Peak	-1.025	-1.025	-1.507	-1.498	-1.494
Corner Peak	-0.995	-1.018	-1.193	-1.211	-1.204
Gaussian	-1.025	-1.025	-1.507	-1.498	-1.494

TABLE 3.2 – *Coefficient linéaire de la régression linéaire opérée sur l'ensemble de points* $(\log(n), \sigma_n)$ *avec* $n = 2^k$ *pour* k *allant de* 1 *à* 12 *en dimension* 100*, pour les différentes méthodes de randomisation de la séquence de Sobol et les différentes fonctions de Genz.*

teur de la variance. Le classement des méthodes en deux groupes se retrouvent encore pour les fonctions régulières, alors que l'ensemble des méthodes ont des performances similaires pour la fonction de type "discontinuous". Ces courbes permettent d'avoir une idée d'un bon choix de m, permettant d'avoir une estimation de σ_n suffisamment convergée. La stabilisation des courbes en fonction de m dépend à la fois du type de fonction à intégrer, ainsi que de la méthode de randomisation utilisée. Les courbes suggèrent que parmi les trois méthodes de "nested scrambling", la méthode de "full scrambling" permet d'avoir une estimation de σ_n plus vite convergée, notamment pour les fonctions de type "product peak" et "gaussian", les variations en fonction de m pour les deux autres méthodes impliquant plus qu'un rapport 10 entre la valeur minimale et maximale. Dans la littérature, une valeur minimale de m couramment admise est la valeur de 10 [72]. Augmenter la valeur de m permet d'avoir une meilleure estimation de σ_n, et permet donc d'avoir une meilleure estimation de la convergence, l'erreur dans certains cas pouvant être d'un ordre de grandeur. Cependant, pour un nombre total $N = nm$ donné, la convergence de la méthode sera dans le meilleur des cas pour les méthodes de "nested scrambling" de l'ordre de $n^{-3/2} \ln(n)^{d-1} m^{-1/2}$ d'après l'expression (3.71), et il est donc plus intéressant de ce point de vu de favoriser peu de longues séquences de Sobol randomisées que beaucoup de courtes séquences de Sobol randomisées. Afin de ne pas pénaliser la grande convergence de la méthode de Quasi-Monte Carlo randomisé, la valeur de $m = 10$ sera prise dans la suite, et la méthode de "full scrambling" sera utilisée le plus possible, puisque celle-ci semble globalement permettre une estimation plus fiable de σ_n pour des faibles valeurs de m. Le surcoût de calcul et de mémoire lié à l'utilisation de la méthode de "full scrambling" reste acceptable tant que la longueur maximale n des séquences de Sobol randomisée est raisonnable, et que la dimension d n'est pas trop importante.

Il ressort des comparaisons menées que les méthodes de randomisation préservant les propriétés d'homogénéité des séquences, celle de Sobol dans le cas présent, offrent des convergences plus rapides. En particulier, dans le cas de fonctions régulières, voire même légèrement irrégulières comme les fonctions

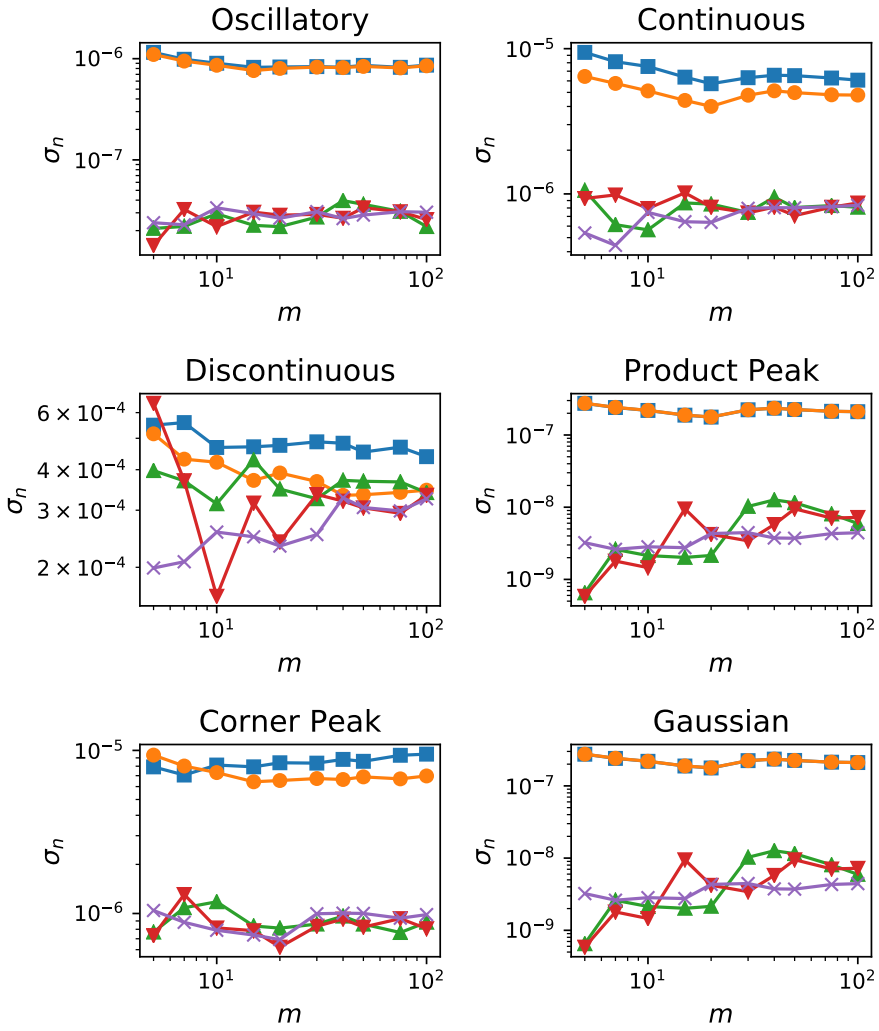

FIGURE 3.18 – *Valeurs des estimations de σ_n pour $n = 4096$ en dimension $d = 50$, en fonction du nombre m de réalisations de séquence de Sobol randomisées utilisées pour l'estimer, suivant les différentes méthodes présentées : "shift" (carrés), "digital shift" (ronds), "I-binomial scrambling" (triangle haut), "random linear scrambling" (triangle bas) et "full scrambling" (croix).*

de type "Continuous", les méthodes de "nested scrambling" que sont la méthode de "I-binomial scrambling", la méthode de "random linear scrambling" et la méthode de "full scrambling" surpassent l'ensemble des autres méthodes de randomisations, offrant un écart-type σ_n inférieur, et de manière non négligeable dans les exemples investigués, aux autres méthodes, pour des valeurs de n égales à des puissances de 2. Ces méthodes entraînent un comportement de σ_n comme $n^{-3/2}$ pour des fonctions suffisamment régulières, alors que les

autres méthodes entraînent un comportement de σ_n en n^{-1}. Le comportement de σ_n en $n^{-3/2}$ fait écho au résultat théorique précisant qu'il doit se comporter comme $n^{-3/2}\ln(n)^{d-1}$, la partie dépendant du logarithme semblant invisible en pratique dans les exemples investigués. Ce résultat théorique est un résultat asymptotique, et il semblerait que seul la partie en $n^{-3/2}$ puisse être conservée dans certains cas en pratique. Les méthodes de "I-binomial scrambling" ou de "random linear scrambling" sont les choix les plus judicieux, entraînant une convergence rapide à un faible coût de calcul et de mémoire, comparé à la méthode de "full scrambling" qui offre la même convergence à un coût beaucoup plus élevé. Cependant, la méthode de "full scrambling" semble permettre une meilleure estimation de σ_n pour un nombre m de séquences de Sobol randomisées faible. Concernant ce nombre m de réalisations de séquences randomisées à prendre en compte, celui-ci ne doit pas être pris inférieur à 10, mais ne doit pas non plus être trop important, car pour un nombre total N d'évaluations de la fonctions à intégrer, augmenter la valeur de m dégrade l'écart-type de l'estimateur de Quasi-Monte Carlo, et cela d'autant plus que la méthode de randomisation utilisées a un taux de convergence important.

Dans la présentation de la méthode de Quasi-Monte Carlo randomisée faite, les seules statistiques présentées ont été l'estimateur I_n ainsi que l'estimation de la variance σ_n, toutes deux réalisées obtenues à l'aide d'un échantillon restreint de m variables aléatoires. La considération d'autres statistiques pose la question de l'obtention de résultats de convergence sur ces statistiques avec cette taille d'échantillon m restreint. Les méthodes d'estimation par ré-échantillonnage sont un outil permettant de tels estimations, qui sont particulièrement utiles en pratique dans le cas d'une taille d'échantillon réduite, ce qui est le cas en Quasi-Monte Carlo randomisé, ce qui motive d'avoir placé la section suivante au sein des méthodes de Quasi-Monte Carlo malgré que celle-ci puisse également être utilisées avec une simple méthode de Monte Carlo.

3.4.4 Estimation d'erreur par ré-échantillonnage

Le calcul de l'espérance d'une variable aléatoire $Q = \tilde{f}(\mathbf{U})$, $\mathbf{U}$ suivant une loi uniforme sur $[0,1]^d$, avec une méthode de Quasi-Monte Carlo randomisé peut être vu comme le calcul de l'espérance d'une variable aléatoire $Q_{QMCR,n}$ définie par la relation (3.81) à l'aide d'une méthode de Monte Carlo.

$$Q_{QMCR,n} = \tilde{f}_{QMCR,n}(\mathbf{U}_1, \ldots, \mathbf{U}_n) = \frac{1}{n}\sum_{i=1}^{n} \tilde{f}(\mathbf{U}_i) \tag{3.81}$$

Dans l'expression (3.81), $\mathbf{U}_{QMCR,n} = (\mathbf{U}_i)_{1\leq i\leq n}$ est une suite à discrépance faible randomisée, et peut être vu comme le vecteur aléatoire dont est fonction la variable aléatoire $Q_{QMCR,n}$. L'estimateur de Quasi-Monte Carlo randomisé $\mu_Q^{(QMCR)}$ pour l'espérance de Q est finalement donné par l'expres-

sion (3.82), les variables aléatoires $Q^{(j)}_{QMCR,n}$ étant indépendantes et suivant la même loi que $Q_{QMCR,n}$.

$$\mu_Q^{QMCR} = \frac{1}{m}\sum_{j=1}^{m} Q^{(j)}_{QMCR,n} \tag{3.82}$$

En combinant les relations (3.81) et (3.82), il est clair que l'estimateur μ_Q^{QMCR} peut s'écrire comme une fonction des m suites à discrépance faible randomisées $\mathbf{U}^{(j)}_{QMCR,n} = (\mathbf{U}_i^{(j)})_{1\leq i\leq n}$:

$$\mu_Q^{(QMCR)} = T(\mathbf{U}^{(1)}_{QMCR,n}, \ldots, \mathbf{U}^{(m)}_{QMCR,n}) \tag{3.83}$$

Voir l'estimateur μ_Q^{QMCR} comme un estimateur de Monte Carlo permet d'obtenir des informations sur celui-ci par le biais du TCL. Ces informations permettent notamment d'estimer l'erreur commise.

Il est parfois nécessaire d'estimer des paramètres plus complexes que la simple espérance d'une variable aléatoire Q. On se place donc dans le contexte où l'on cherche à estimer une grandeur θ à l'aide d'une méthode de Quasi-Monte Carlo randomisé. Des exemples de tels paramètres seront donnés dans le chapitre suivant, avec notamment les indices de Sobol ou encore les valeurs propres de la décomposition de Karhunen-Loève d'un processus stochastique. On considère pour cela un estimateur $\hat{\theta}$ qui peut être écrit sous la forme suivante :

$$\hat{\theta} = T_m(\mathbf{U}^{(1)}_{QMCR,n}, \ldots, \mathbf{U}^{(m)}_{QMCR,n}) \tag{3.84}$$

L'expression (3.84) renvoie en fait à l'estimation d'un paramètre à l'aide d'une statistique dépendant de vecteurs aléatoires indépendants et suivant tous la même loi. Elle peut plus généralement s'écrire sous la forme de l'expression (3.85), où les $\mathbf{X}^{(j)}$ sont simplement des vecteurs aléatoires indépendants et de même loi, et peut également s'appliquer dans un autre contexte que la méthode de Quasi-Monte Carlo randomisé.

$$\hat{\theta}_m = T_m(\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(m)}) \tag{3.85}$$

Comme toute estimation d'un paramètre, il est intéressant d'avoir des méthodes permettant d'évaluer l'erreur commise par cette estimation, qui sont ici basées sur le principe de ré-échantillonnage.

3.4.4.1 Jackknife

La méthode de Jackknife a été introduite pour la première fois dans le but de réduire le biais [107, 106] d'une statistique telle celle présente dans l'expression (3.85), et a été amélioré par la suite [136] notamment en s'intéressant en plus à l'estimation de la variance dans le cadre de faibles échantillons, ce qui est précisément le cas dans la méthode de Quasi-Monte Carlo randomisé où le nombre m de séquences est considéré raisonnablement faible afin de ne pas pénaliser la convergence de la méthode.

L'intérêt premier de la méthode est la réduction de biais d'un estimateur. Pour montrer cela, on suppose que $T_m(\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(m)})$ possède un biais de la forme suivante :

$$E\left[T_m(\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(m)})\right] - \theta = \frac{a}{m} + \frac{b}{m^2} + O\left(\frac{1}{m^3}\right) \tag{3.86}$$

Les estimateurs $T_{m-1,i} = T_{m-1}(\mathbf{X}^{(1)}, \ldots, \mathbf{X}(i-1), \mathbf{X}^{(i+1)}, \ldots, \mathbf{X}^{(m)})$ sont construits à partir de l'échantillon initiale où a été retiré le i-ème échantillon, et l'estimateur $\bar{T}_m$ correspondant à leur moyenne arithmétique est également construit :

$$\bar{T}_m = \frac{1}{m}\sum_{i=1}^{m} T_{m-1,i} \tag{3.87}$$

Dans le cas où T_m dépend linéairement des $\mathbf{X}^{(j)}$, comme c'est le cas lorsque l'on cherche à estimer l'espérance de $\mathbf{X}$, $\bar{T}_m$ et T_m sont confondus, mais ce n'est pas vrai dans le cas général. L'estimateur de Jackknife est alors donné par l'expression suivante :

$$T_{jack} = mT_m - (m-1)\bar{T}_m \tag{3.88}$$

En utilisant la relation (3.86) pour calculer le biais de ce nouvel estimateur

de θ, il vient :

$$\begin{aligned} E\left[T_{jack}\right] - \theta &= m\left(E\left[T_m\right] - \theta\right) - (m-1)\left(E\left[\bar{T}_m\right] - \theta\right) \\ &= m\left(\frac{a}{m} + \frac{b}{m^2} + O\left(\frac{1}{m^3}\right)\right) \\ &- (m-1)\left(\frac{a}{m-1} + \frac{b}{(m-1)^2} + O\left(\frac{1}{(m-1)^3}\right)\right) \\ &= b\left(\frac{1}{m} - \frac{1}{m-1}\right) + O\left(\frac{1}{m^2}\right) \\ &= -\frac{b}{m(m-1)} + O\left(\frac{1}{m^2}\right) \end{aligned} \tag{3.89}$$

La considération de l'estimateur T_{jack} permet donc de réduire le biais par rapport à l'estimateur T_m initial, celui-ci étant d'ordre m^{-2} plutôt que m^{-1}. Pour s'intéresser à l'estimateur de la variance de T_m, il est nécessaire d'introduire les pseudo-valeurs $T^*_{m-1,i}$ [136] :

$$T^*_{m-1,i} = mT_m - (m-1)T_{m-1,i} \tag{3.90}$$

Il est alors immédiat que l'estimateur T_{jack} est la moyenne arithmétique des pseudo-valeurs $T^*_{m-1,i}$. La construction de l'estimateur jackknife de la variance pour T_m repose alors sur deux hypothèses, qui ne sont pas vraie en générale :

— Les pseudo-valeurs $T^*_{m-1,i}$ peuvent être traitées comme si elles étaient indépendantes et identiquement distribuées.

— La variance de T_{jack} est égale à la variance de $\sqrt{m}T_m$.

Avec ces deux hypothèses, la variance de T_m peut alors être approchée de la manière suivante :

$$Var\left(T_m\right) = \frac{1}{m}Var\left(T_{jack}\right) \approx \frac{1}{m(m-1)}\sum_{i=1}^{m}\left(T^*_{m-1,i} - T^2_{jack}\right) \tag{3.91}$$

L'expression précédente est celle de l'estimateur jackknife de la variance de T_m. La méthode présentée présente l'intérêt d'obtenir une estimation de la variance d'une statistique de la forme de celle de T_m, à un coût relativement raisonnable en terme de nombre d'évaluations de la statistiques T. Cependant, la méthode de jackknife fonctionne correctement à condition que la fonction T soit suffisamment régulière [119], et son utilisation est risquée dans le cas contraire, l'estimateur de la variance pouvant par exemple ne pas converger vers la vrai valeur de la variance. Des méthodes de jackknife plus avancées peuvent être utilisée [119], mais il est également possible de se tourner dans une telle

situation vers le bootstrap, qui est peut se voir comme une généralisation de la méthode de jackknife [28].

3.4.4.2 Bootstrap

L'idée de la méthode de bootstrap est postérieure à l'idée du jackknife [27], et son essor est notamment dû à l'augmentation de la puissance de calcul des machines rendant son utilisation pratique accessible. Là où le jackknife utilise la construction de m pseudo-valeurs construites à partir de m sous-échantillons d'une forme donnée des m échantillons disponibles, le bootstrap s'autorise à utiliser l'ensemble des sous-échantillons disponibles.

L'idée du bootstrap va être introduite ici à l'aide de la variance de la statistique $T_m\left(\mathbf{X}^{(1)},\ldots,\mathbf{X}^{(m)}\right)$ estimée précédemment à l'aide de la méthode de jackknife. Les vecteurs aléatoires $\mathbf{X}^{(i)}$ sont toujours supposés indépendants et identiquement distribués, selon une loi donnée par la mesure de probabilité P. L'expression de la variance de la statistique a l'expression suivante :

$$Var\left(T_m\right) = E_{P^{\otimes m}}\left[\left(T_m\left(\mathbf{X}^{(1)},\ldots,\mathbf{X}^{(m)}\right) - E_{P^{\otimes m}}\left[T_m\left(\mathbf{X}^{(1)},\ldots,\mathbf{X}^{(m)}\right)\right]\right)^2\right] \tag{3.92}$$

L'espérance est prise selon la loi produit $P^{\otimes m}$ puisque la statistique T_m dépend des m vecteurs aléatoires $\mathbf{X}^{(i)}$ tous de même loi et indépendants entre eux. La mesure de probabilité P est bien entendu généralement inconnue. Tout comme pour la méthode de jackknife, l'objectif est l'estimation de cette quantité en utilisant seulement une réalisation d'un échantillon $\left(\mathbf{X}^{(1)},\ldots,\mathbf{X}^{(m)}\right)$, l'obtention d'une réalisation étant supposée coûteuse, ce qui interdit l'utilisation d'une méthode de Monte Carlo pour évaluer l'expression (3.92). L'idée est de substituer à la mesure de probabilité P une mesure de probabilité $\hat{P}$ à partir de laquelle il est peu coûteux d'échantillonner, de sorte que l'approximation bootstrap de la variance de T_m est donnée par :

$$Var_{boot}\left(T_m\right) = E_{\hat{P}^{\otimes m}}\left[\left(T_m\left(\mathbf{X}^{(1)},\ldots,\mathbf{X}^{(m)}\right) - E_{\hat{P}^{\otimes m}}\left[T_m\left(\mathbf{X}^{(1)},\ldots,\mathbf{X}^{(m)}\right)\right]\right)^2\right] \tag{3.93}$$

Généralement, la mesure de probabilité $\hat{P}$ choisie est la mesure empirique $P_{emp,m}$ correspondant à l'échantillon $\left(\mathbf{X}^{(1)},\ldots,\mathbf{X}^{(m)}\right)$, et définie par :

$$P_{emp,m}(A) = \frac{1}{m}\sum_{i=1}^{m} \mathbb{1}_A\left(\mathbf{X}^{(i)}\right) \tag{3.94}$$

Échantillonner suivant la loi $P_{emp,m}$ revient à choisir aléatoirement et de manière uniforme un des $\mathbf{X}^{(i)}$, alors qu'échantillonner suivant la loi produit $P_{emp,m}^{\otimes m}$ revient à construire un échantillon $(\mathbf{X}_1^*, \ldots, \mathbf{X}_m^*)$ où chacun des $\mathbf{X}_i^*$ a été tiré indépendamment des autres et de manière uniforme parmi les $\left(\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(m)}\right)$, ce que l'on appelle couramment un tirage avec remise. Se faisant, il est possible d'estimer la variance $Var_{boot}(T_m)$ à l'aide d'une méthode de Monte Carlo, à l'aide de l'expression suivante où B tirages ont été effectués :

$$Var_{boot}(T_m) \approx \frac{1}{B}\sum_{b=1}^{B}\left(T_m(\mathbf{X}_{1,b}^*, \ldots, \mathbf{X}_{m,b}^*) - \frac{1}{B}\sum_{l=1}^{B} T_m(\mathbf{X}_{1,l}^*, \ldots, \mathbf{X}_{m,l}^*)\right)^2 \tag{3.95}$$

Le membre de droite de l'expression précédente tend bien vers le membre de gauche avec le nombre de tirages B qui tend vers l'infini.

Tout comme il a été possible d'obtenir une estimation de la variance de T_m, il est possible d'obtenir des estimations d'autres informations sur T_m en appliquant le même procédé au besoin. En particulier, le bootstrap permet d'accéder à la distribution bootstrap du paramètre T_m, qui s'obtient au travers de l'ensemble des évaluations $T_m(\mathbf{X}_{1,b}^*, \ldots, \mathbf{X}_{m,b}^*)$, et à partir desquelles il est possible de construire un histogramme ou des intervalles de confiance pour T_m.

Les résultats obtenus dépendent bien entendu du choix de la mesure de probabilité de substitution $\hat{P}$ qui sera toujours dans cette thèse celui de la mesure empirique $P_{emp,m}$. D'autres choix sont cependant possibles, notamment avec les méthodes présentées dans le chapitre 5.

Le bootstrap possède un avantage sur la méthode de jackknife de par le fait qu'il n'échoue pas pour déterminer la variance de T_m en cas de manque de régularité de la statistique T_m [119]. Il permet de plus d'accéder à la distribution de la statistique T_m, ce qui est moins évident avec la méthode de jackknife. Cependant, le bootstrap est plus coûteux que la méthode de jackknife et le jackknife peut s'avérer suffisant dans de nombreux cas.

Les méthodes de jackknife et de bootstrap sont des méthodes permettant d'obtenir des informations riches à partir d'un nombre restreint d'échantillons indépendants et identiquement distribués, ce qui est le cas dans l'usage de la méthode de Quasi-Monte Carlo randomisé où le nombre m de séquences randomisées considérées est intentionnellement faible. Bien que ces méthodes puissent également s'utiliser avec les échantillons générées par une méthode de Monte Carlo, il semble plus intéressant de les présenter comme un outil complémentaire à la méthode de Quasi-Monte Carlo randomisé, qui offre des performances supérieures à la méthode de Monte Carlo comme le montrent les résultats de la partie suivante.

3.5 Comparaison des différentes méthodes

Sur les figures 3.19 et 3.20 sont comparées les convergences des différentes méthodes d'intégrations présentées dans ce chapitre, sur les fonctions proposées par Genz [40], pour une valeur de la dimension d de 5 et 100 respectivement.

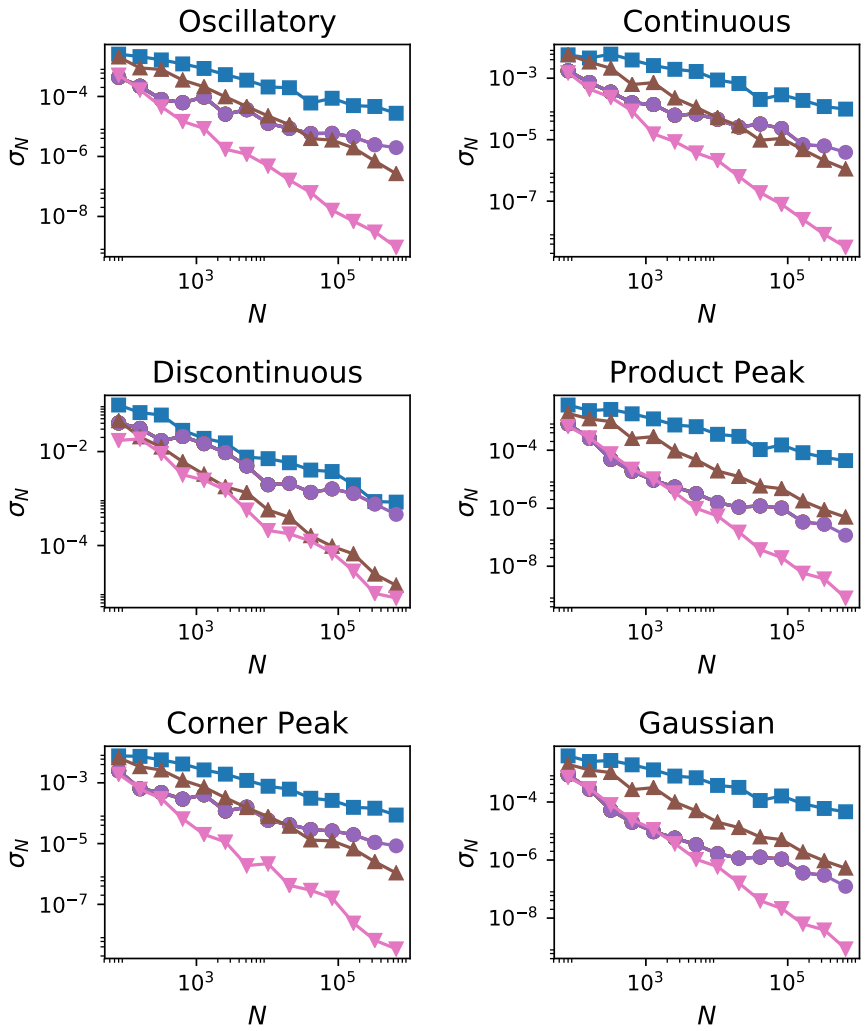

FIGURE 3.19 – *Comparaison de l'évolution de σ_N en fonction du nombre d'évaluations de la fonction $N = mn$, pour une valeur de m égale à 10, sur les six types de fonctions proposées par Genz dépendant de 5 variables. Les méthodes comparées sont : Monte-Carlo (carrés), Latin Hypercube Sampling (ronds), Séquence de Halton améliorée randomisée par un "shift" (triangle haut) et Séquence de Sobol randomisée par un "full scrambling" (triangle bas). Les valeurs de n pour les points présents sur le graphe sont des puissances de 2.*

Les méthodes comparées sont la méthode de Monte-Carlo brute, l'échan-

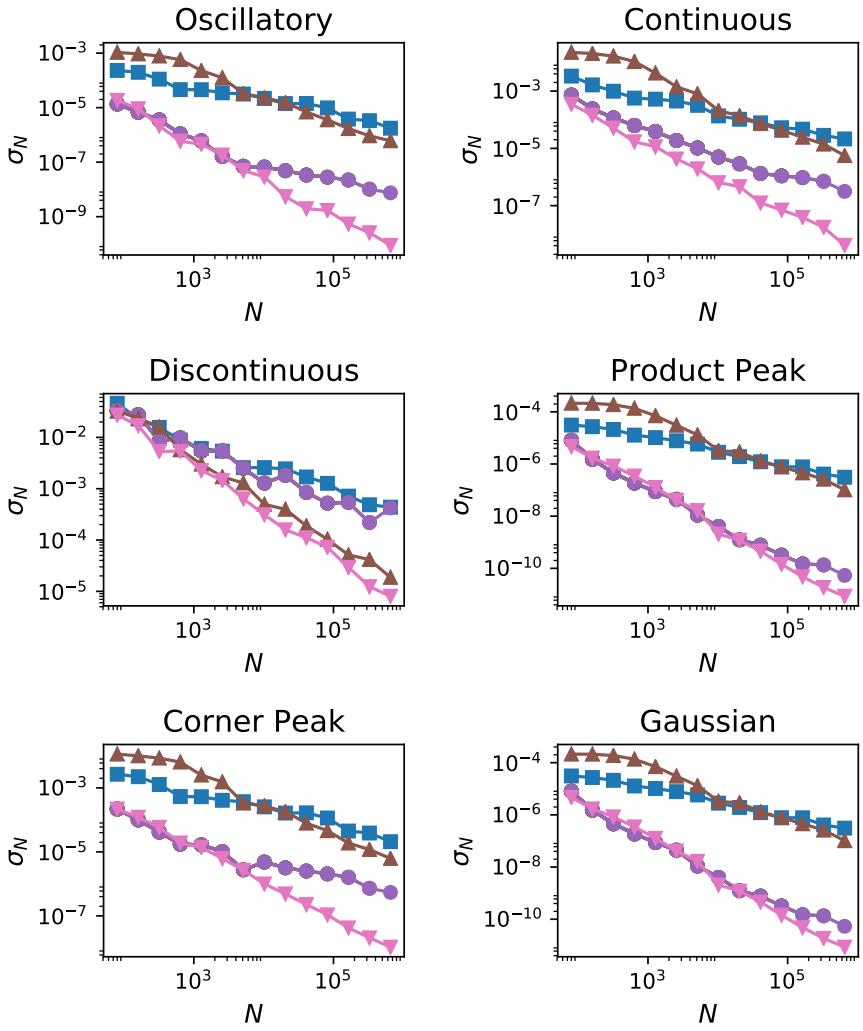

FIGURE 3.20 – *Comparaison de l'évolution de σ_N en fonction du nombre d'évaluations de la fonction $N = mn$, pour une valeur de m égale à 10, sur les six types de fonctions proposées par Genz dépendant de 100 variables. Les méthodes comparées sont : Monte-Carlo (carrés), Latin Hypercube Sampling (ronds), Séquence de Halton améliorée randomisée par un "shift" (triangle haut) et Séquence de Sobol randomisée par un "full scrambling" (triangle bas). Les valeurs de n pour les points présents sur le graphe sont des puissances de 2.*

tillonnage par hypercube latin, la méthode de Quasi-Monte Carlo avec une séquence de Halton améliorée à l'aide d'un "digital shift" et randomisée à l'aide d'un simple "shift", et la méthode de Quasi-Monte Carlo randomisée avec une séquence de Sobol randomisée à l'aide d'une méthode de "full scrambling". L'estimation de l'écart-type σ_N, avec $N = mn$, a été obtenue à l'aide de $m = 10$

réalisations pour les méthodes de Quasi-Monte Carlo randomisé, et la même méthode d'estimation de cet écart-type a été utilisée pour la méthode de LHS, m échantillonnages par hypercube latin indépendants chacun contenant n points ayant été utilisés pour estimer cet écart-type. Les nombres n de points d'évaluation dans chacune des m réalisations utilisés pour construire les graphes sont des puissances de 2, et les courbes correspondent donc au meilleur cas pour la séquence de Sobol. Dans l'ensemble des cas, la méthode de Quasi-Monte Carlo randomisé utilisant les séquences de Sobol surpasse toutes les autres alors que la méthode de Monte-Carlo brute est celle présentant les moins bonnes performances. Comme prévu par la théorie, la méthode de Monte-Carlo présente une convergence de σ_N inversement proportionnelle à la racine carré de N pour l'ensemble des cas étudiés. La méthode de Quasi-Monte Carlo randomisé utilisant des séquences de Halton est bien plus efficace en dimension 5, visible sur la figure 3.19, qu'en dimension 100 visible sur la figure 3.20. Il semble que le régime asymptotique pour cette méthode est d'autant plus long à se mettre en place que la dimension est importante. L'échantillonnage par hypercube latin présente dans l'ensemble des cas une convergence asymptotique similaire à la méthode de Monte-Carlo, mais présente une forte convergence pour les faibles valeurs de N, le faisant initialement rivaliser avec la méthode de Quasi-Monte Carlo utilisant des séquences de Sobol dans certains cas. Cependant, l'échantillonnage par hypercube latin dans sa version classique n'offre pas la possibilité d'ajouter des points afin d'améliorer la convergence, contrairement à la méthode de Quasi-Monte Carlo randomisé utilisant des séquences à discrépance faible. Les deux méthodes de Quasi-Monte Carlo offrent bien des convergences asymptotiques meilleures que les deux autres méthodes dans l'ensemble des cas. Dans le cas de la fonction de type "Discontinuous", les performances des deux méthodes de Quasi-Monte Carlo sont proches, bien que la séquence de Sobol soit légèrement meilleure.

3.6 Conclusion

Lorsque l'espace d'intégration est de dimension trop importante, typiquement plus grande que 5, les méthodes d'intégrations déterministes basées sur la tensorisation de quadrature sont inutilisables en pratique car trop coûteuses. La méthode de Monte Carlo, qui repose sur des considérations théoriques de la théorie des probabilités, permet le calcul d'intégrales de fonctions même très irrégulières, et possède un taux de convergence inversement proportionnel à la racine carré du nombre n d'évaluations de la fonction, indépendant de la dimension de l'espace sur lequel la fonction doit être intégrée, et qui est proportionnel à l'écart-type de cette fonction. Des méthodes, dites de réduction de variance, permettent d'accélérer la convergence de la méthode, en réduisant la variance et donc l'écart-type par une transformation de l'intégrale à calculer. Cependant, la dépendance du taux de convergence avec le nombre d'évaluations

reste inchangée, et est inversement proportionnelle à la racine carré du nombre n d'évaluations de la fonction à calculer. En particulier, cela implique que pour doubler la précision d'un résultat, il faut quadrupler le nombre d'évaluations de la fonction, et donc le coût du calcul.

La méthode de Monte Carlo présente donc l'avantage de pouvoir être utilisée en toute circonstances avec le même taux de convergence, mais ce taux de convergence reste cependant faible. Les méthodes de Quasi-Monte Carlo, utilisant un tirage de points à discrépance faible plutôt qu'un tirage de points aléatoire, ont été développées afin d'améliorer ce faible taux de convergence, en offrant la possibilité d'avoir cette fois une convergence asymptotique proportionnelle à n^{-1}, voire $n^{-3/2}$ où n est le nombre d'évaluations. La séquence de Sobol a été choisie pour ces travaux de thèse. La simple utilisation d'un tirage de points à discrépance faible ne permet cependant pas d'obtenir une estimation de l'erreur commise, mais une randomisation de ces points à discrépance faible permet d'obtenir une estimation de l'erreur. Plusieurs méthodes de randomisations existent, plus ou moins complexes, la complexité impliquant un coût de calcul et/ou de mémoire plus important. Les méthodes de randomisation impliquant un "nested scrambling" offrent une convergence meilleure que les autres méthodes, et dans le cas de fonctions suffisamment régulières, permettent d'obtenir un taux de convergence proportionnel à l'inverse du nombre d'évaluations n à la puissance $3/2$, quand ce nombre de points d'évaluations est une puissance de 2. Certaines de ces méthodes sont de plus très peu coûteuses en termes de calcul, tel le "I Binomial Scrambling" ou encore le "Random Linear Scrambling". La méthode de "Full scrambling" bien plus coûteuse que les deux précédentes méthodes semblent cependant permettre une estimation moins bruitée de l'écart-type σ_n. Dans le cas de fonctions présentant une régularité moindre, les méthodes de randomisation impliquant un "nested scrambling" perdent leur supériorité par rapport aux autres méthodes, mais restent tout de même plus efficace qu'une simple méthode de Monte Carlo.

Les méthodes de Quasi-Monte Carlo randomisées seront privilégiées dans la suite de la thèse. La supériorité des méthodes de Quasi-Monte Carlo randomisées implique de faire un compromis entre le nombre m de séquences utilisées, et le nombre n de points dans chacune de ces séquences : à coût de calcul égal, un faible nombre de séquences implique une convergence plus forte au prix d'une moins bonne estimation de l'erreur caractérisée par l'écart-type σ_n, qui peut devenir préjudiciable dans le cas d'un nombre m de séquences trop faible (typiquement inférieur à 10). L'estimation de certaines statistiques peut alors être réalisée avec une faible taille d'échantillon aléatoire, et des méthodes de ré-échantillonnage telles la méthode de Jackknife ou le bootstrap peuvent alors être avantageusement utilisées afin d'obtenir des informations de convergence de ces statistiques.

Les résultats de ce chapitre ont permis de réaliser une communication au sein de l'ASME présentée en annexe B. Le travail présenté consistait en l'utilisation des méthodes de Quasi-Monte Carlo randomisées introduites dans

ce chapitre pour la résolution de l'équation de transfert radiatif. L'utilisation de ces méthodes permet de réaliser la résolution de l'équation de transfert radiatif pour un coût de calcul 2 à 3 fois moins important que dans le cas de l'utilisation de la méthode de Monte Carlo. Cela représente un gain de temps significatif et une avancée prometteuse pour la réduction du temps de calcul des simulations multi-physiques couplées. Un papier sur cette même utilisation est également à paraître dans le *Journal of Quantitative Spectroscopy & Radiative Transfer (JQSRT).*

Chapitre 4

Méthodes spectrales pour la représentation de variables aléatoires et de processus stochastiques

Prérequis :

- *Calcul d'intégrales multiples (Cubature, Monte Carlo et Quasi-Monte Carlo randomisé)*
- *Estimation d'erreur par des méthodes de ré-échantillonnage*

Notions clés et apports du chapitre :

- *Expansion en Polynôme du Chaos (PCE) d'une variable aléatoire ou d'un processus stochastique*
- *Calcul de l'expansion en polynôme du chaos par des méthodes projectives basée sur le calcul d'intégrale*
- *Analyse de sensibilité globale d'une fonction d'un vecteur aléatoire (HDMR et indices de Sobol)*
- *Expansion de Karhunen-Loève (KLE) d'un processus stochastique*
- *Calcul de l'expansion de Karhunen-Loève d'un processus stochastique par la méthode de Nyström*
- *Étude des différentes sources d'erreur lors du calcul de l'expansion de Karhunen-Loève par la méthode de Nyström*

Dans le contexte de la quantification ou de la propagation d'incertitudes, l'objectif est de caractériser l'incertitude d'une ou plusieurs quantités d'intérêt du système étudié. Il a été vu dans le chapitre 1 que ces quantités d'intérêt pouvaient alors être considérées comme des variables aléatoires, si la quantité d'intérêt est la valeur d'une grandeur en un point et à un instant donné, ou des processus stochastiques si la quantité d'intérêt est un champ spatial, une évolution temporelle d'une grandeur, ou l'évolution temporelle d'un champ. Pour des raisons pratiques, la caractérisation de l'incertitude de la quantité d'intérêt se fait au travers d'une modélisation de celle-ci. Cette modélisation consiste dans cette thèse en l'obtention d'une représentation simple et peu coûteuse à évaluer de la quantité d'intérêt, comme une combinaison linéaire d'un nombre restreints de variables aléatoires dont on sait facilement et rapidement générer des échantillons. Cette étape de modélisation peut être vue comme une compression de l'information avec perte, la perte d'information étant telle qu'elle n'affecte pas significativement les résultats obtenus à l'aide de la représentation simplifiée de la quantité d'intérêt. La construction d'une telle représentation dépend du type de problème à traiter, que l'on peut distinguer en deux catégories.

Dans la première catégorie, seulement quelques informations sont connues sur la quantité d'intérêt à représenter, typiquement des moments statistiques ou un ensemble de réalisations de cette quantité d'intérêt obtenues grâce aux données d'un résultat d'expérience par exemple, expérience pouvant être numérique. Dans un tel cas de figure où seules des informations concernant la quantité d'intérêt sont disponibles, il n'est pas possible de relier la quantité d'intérêt aux variables aléatoires étant à la source de l'incertitude du système, et de nouvelles variables aléatoires doivent être construites pour modéliser l'incertitude de la quantité d'intérêt, ce qui correspond à de la quantification d'incertitudes. L'obtention de ces nouvelles variables aléatoires peut alors être obtenues à partir de différentes techniques, parmi lesquelles l'analyse en composantes principales (PCA pour Principal Components Analysis en anglais) ou encore l'analyse en composante indépendantes (ICA pour Independant Components analysis en anglais).

La seconde catégorie correspond aux cas où il est possible de calculer la quantité d'intérêt à partir des variables aléatoires d'entrée, étant à la source de l'incertitude du système, pour lesquelles l'incertitude a été caractérisée, correspondant donc à un cas de propagation d'incertitudes. Il est possible dans un tel cas d'envisager de construire une surface de réponse pour la quantité d'intérêt dépendant uniquement des variables aléatoires d'entrée. Un tel exemple de surface de réponse est donné par la figure 4.1 où une vitesse axiale, qui est la quantité d'intérêt, en un point d'un écoulement diphasique monodisperse au sein d'une géométrie cylindrique est obtenue en fonction de deux variables d'entrée que sont la distance à l'axe de symétrie et la taille des gouttes de liquides présentes dans l'écoulement.

Deux types de méthodes sont à envisager pour construire une surface de réponse dépendant des paramètres d'entrées.

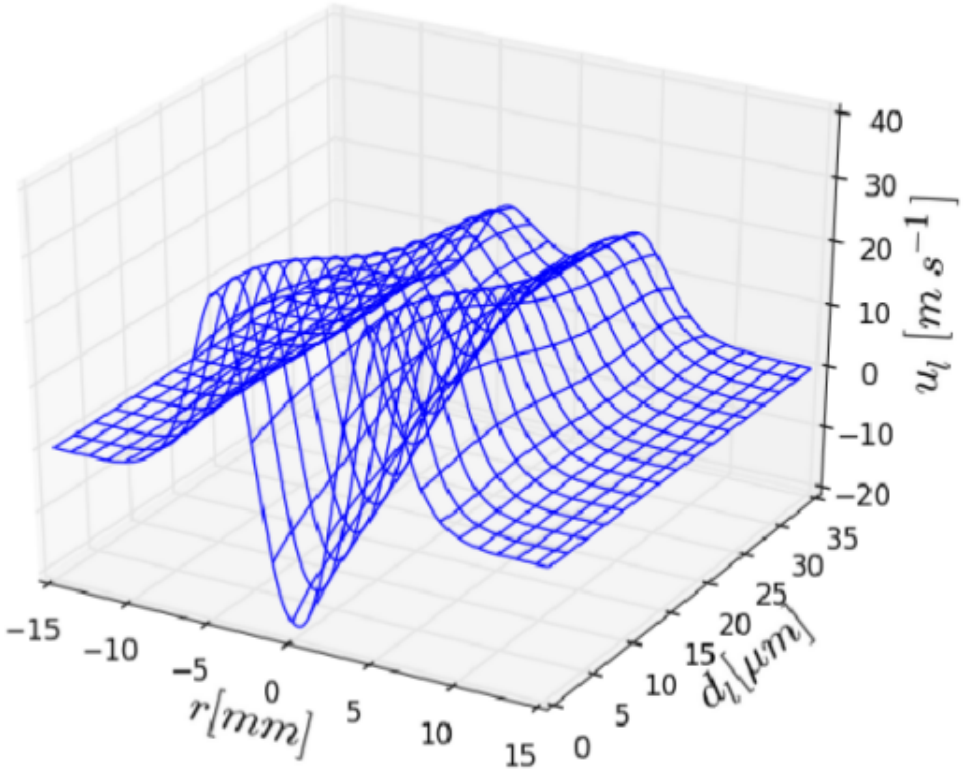

FIGURE 4.1 – *Exemple de surface de réponse obtenue par une expansion en polynôme du chaos, tiré de l'article présent à la fin de ce chapitre.*

- *La première, plus simple d'utilisation, correspond aux méthodes dites non-intrusives, où seuls sont utilisés des couples de valeurs* $(\mathbf{V_e}^{(i)}, QoI^{(i)})$, $\mathbf{V_e}$ *correspondant aux variables d'entrée et* QoI *à la quantité d'intérêt, pour la construction de la surface de réponse. Le passage des valeurs des variables d'entrée* $\mathbf{V_e}^{(i)}$*, qui sont des valeurs choisies à l'aide d'un échantillonnage présentées dans les chapitres précédents, aux valeurs de la quantité d'intérêt* $QoI^{(i)}$ *correspondantes est généralement réalisé à l'aide de multiples simulations pour lesquelles l'outil de simulation utilisé est généralement le même que celui utilisé pour des simulations déterministes classiques, ne nécessitant pas le développement de nouvelles méthodes ou de nouveaux outils numériques.*
- *La seconde correspond aux méthodes dites intrusives, pour lesquelles l'obtention de la surface de réponse va nécessiter de reformuler la résolution du système avec un formalisme stochastique, ne permettant pas l'utilisation des méthodes et outils utilisés dans le simple cas déterministe. Cette approche est de fait bien plus complexe puisqu'elle nécessite un nouveau formalisme, ainsi que de nouveaux outils numériques pour l'obtention de la réponse stochastique du système.*

Le choix a été fait de ne considérer que des méthodes non intrusives, car celles-ci sont plus simples et permettent l'utilisation d'outils numériques déjà largement utilisés pour la version déterministe des problèmes traités dans cette thèse.

4.1 Propagation d'incertitude via une expansion en polynômes du chaos

Les polynômes du chaos sont un outil permettant de construire une surface de réponse polynomiale en des variables aléatoires, pour approximer une variable aléatoire, réelle ou vectorielle, ou un processus stochastique d'intérêt. Avant de présenter les outils numériques pour la détermination de l'expansion en polynômes du chaos et l'utilisation pratique des polynômes du chaos dans cette thèse, des éléments théoriques vont être apportés dans cette partie, afin de présenter la provenance de cet outil mathématique que sont les polynômes du chaos, ainsi que leurs propriétés et leurs limitations.

4.1.1 Éléments théoriques concernant les polynômes du chaos

Comme précisé en introduction, la théorie des probabilités est un formalisme adapté à la prise en compte des incertitudes, et nous considérons dans la suite un espace probabilisé (Θ, Σ, P) sur lequel seront définie l'ensemble des variables aléatoires ou processus stochastiques considérés. On fait également l'hypothèse raisonnable pour des applications physiques que les variables aléatoires auxquelles nous nous intéressons sont de carré intégrable, c'est à dire qu'elles possèdent une variance. L'espace des variables aléatoires de carré intégrale $L^2(\Theta, P)$ offre un cadre intéressant, puisqu'il est naturellement muni d'un produit scalaire. Si U et V sont deux variables aléatoires réelles, le produit scalaire de U et V est défini comme l'espérance du produit UV.

4.1.1.1 Polynômes du chaos

Cameron et Martin ont montré [19] que, en considérant une famille dénombrable $\xi = \{\xi_i\}_{i\in\mathbb{N}^*}$ de variables gaussiennes centrées réduites et mutuellement orthogonales définies sur l'espace probabilisé (Θ, Σ, P), n'importe quelle variable aléatoire U de $L^2(\Theta, P)$ peut s'exprimer sous la forme :

$$U = u_0\Gamma_0 + \sum_{i_1=1}^{+\infty} u_{i_1}\Gamma_{i_1}(\xi_{i_1}) + \sum_{i_1=1}^{+\infty}\sum_{i_2=1}^{i_1} u_{i_1 i_2}\Gamma_{i_1 i_2}(\xi_{i_1}, \xi_{i_2}) + \sum_{i_1=1}^{+\infty}\sum_{i_2=1}^{i_1}\sum_{i_3=1}^{i_2} u_{i_1 i_2 i_3}\Gamma_{i_1 i_2 i_3}(\xi_{i_1}, \xi_{i_2}, \xi_{i_3}) + \ldots \tag{4.1}$$

Dans l'expression (4.1), les fonctions Γ_I, où $I = (i_1, \ldots, i_n)$ est une combinaison avec répétition (un ensemble d'éléments où la présence multiple d'un

élément est autorisée et où l'ordre n'a pas d'importance) d'éléments de $\mathbb{N}^*$*, sont des polynômes de degré totale* n *et qui dépendent des variables* $\xi_{i_1}, \ldots, \xi_{i_n}$*. De plus, l'ensemble des variables aléatoires* $\Gamma_I(\xi_{\mathbf{I}})$*, que l'on dénommera dans la suite abusivement polynômes, forment une famille orthogonale pour le produit scalaire naturel sur* $L^2(\Theta, P)$*. La représentation définie dans (4.1) a un sens pour la norme euclidienne associée au produit scalaire naturel de* $L^2(\Theta, P)$*, ce qui signifie que l'espace* $L^2(\Theta, P)$ *est un espace de Hilbert, dont une base hilbertienne est l'ensemble des polynômes* $\Gamma_I(\xi_{\mathbf{I}})$*.*

Les coefficients u_I *présents dans l'expansion (4.1) correspondent aux coefficients des projections orthogonales de la variable aléatoire* U *sur les variables aléatoires* $\Gamma_I(\xi_{\mathbf{I}})$*, et sont donnés par l'expression (4.2) :*

$$u_I = \frac{E\left[U\Gamma_I(\xi_{\mathbf{I}})\right]}{E\left[\Gamma_I(\xi_{\mathbf{I}})^2\right]} \tag{4.2}$$

L'expression (4.2) correspond statistiquement au coefficient de corrélation entre U *et* $\Gamma_I(\xi_{\mathbf{I}})$*. Dans le cas où la base de polynômes du chaos est orthonormale, cela revient juste à calculer l'espérance du produit de la variable aléatoire* U *avec* $\Gamma_I(\xi_{\mathbf{I}})$*, le dénominateur valant* 1*.*

En fait, une expansion tronquée de l'expansion (4.1) en ne considérant que des éléments indexés par des listes d'éléments $I \in \mathcal{I}$ *correspond à la projection orthogonale pour le produit scalaire de* $L^2(\Theta, P)$ *sur l'adhérence du sous espace vectoriel engendré par les variables aléatoires* $\Gamma_I(\xi_{\mathbf{I}})$ *avec* $I \in \mathcal{I}$*.*

Les polynômes $\Gamma_I(\xi_{\mathbf{I}})$ *sont appelés polynômes du chaos dans la littérature, en référence aux travaux précurseurs de Wiener [145]. Dans le cas où les variables aléatoires* $\{\xi_i\}_{i\in\mathbb{N}^*}$ *sont mutuellement indépendantes (la mutuelle indépendance est équivalente à la mutuelle orthogonalité dans le cas où les* ξ_i *sont gaussiennes, mais la construction qui suit est généralisable au cas non gaussien qui sera abordé plus tard), la construction des* $\Gamma_I(\xi_{\mathbf{I}})$ *se réalisent en deux temps. Tout d'abord, les variables* ξ_i *étant toutes indépendantes et identiquement distribuées, il est intéressant de construire une famille de polynômes orthogonaux d'une seule variable aléatoire suivant la loi des* ξ_i*. En prenant pour ces polynômes un degré croissant, cela conduit aux polynômes de Hermite [3]* $(H_k)_{k\in\mathbb{N}}$ *dans le cas où les* ξ_i *sont des gaussiennes centrées réduites. Ensuite, il est possible de construire les polynômes* $\Gamma_I(\xi_I)$ *à partir des polynômes* H_k*. En notant* n_I *le nombre d'éléments distincts dans* I*,* α_k *les éléments présents dans* I *et* β_k *leurs nombres d'apparition dans* I*, on définit* $\Gamma_I(\xi_I)$ *comme :*

$$\Gamma_I(\xi_I) = \prod_{k=1}^{n_I} H_{\alpha_k}(\xi_{\beta_k}) \tag{4.3}$$

Il est immédiat qu'ainsi construit, les polynômes $\Gamma_I(\xi_I)$ *sont mutuellement orthogonaux pour le produit scalaire naturel sur* $L^2(\Theta, P)$*.*

Distribution	Support	Polynômes
Gaussienne	$]-\infty,+\infty[$	Hermite
Uniforme	$[a,b]$	Legendre
Beta	$[a,b]$	Jacobi
Gamma	$[0,+\infty[$	Laguerre

TABLE 4.1 – *Lois de probabilités avec leurs supports et familles de polynômes orthogonaux correspondant.*

Dans le cas où l'indépendance mutuelle des variables aléatoires $(\xi_i)_{i\in\mathbb{N}}$ n'est pas vérifiée, il est encore possible de construire une base de polynômes [147], mais le caractère tensorisé des polynômes de plusieurs variables à partir de ceux monodimensionnels est perdu. Une alternative permet de conserver ce caractère tensorisé [128], mais la base ainsi construite, bien qu'orthogonale, n'est alors plus généralement polynomiale. Le fait de ne plus avoir de polynômes n'est pas nécessairement problématique, et on pourra dénommer les éléments de cette base des fonctions du chaos tout comme pour les polynômes de Hermite précédemment définis.

En fait, le résultat de Cameron et Martin ne considère que des polynômes en des variables aléatoires normales centrées et réduites, et la vitesse de convergence d'une expansion comme celle de l'expression (4.1) dépend de la variable aléatoire que l'on souhaite reconstruire [101]. Il peut donc être intéressant de s'intéresser à la reconstruction d'une variable aléatoire de $L^2(\Theta,P)$ comme une expansion du type de (4.1) impliquant d'autres variables que les $(\xi_i)_{i\in\mathbb{N}^*}$.

4.1.1.2 Polynômes du chaos généralisés

Xiu et al. [146] ont été les premiers à considérer l'ensemble des polynômes dérivant du schéma de Askey [9] pour établir une expansion similaire à (4.1) en utilisant des variables aléatoires $(\eta_i)_{i\in\mathbb{N}^*}$ non normales. On souhaite désormais approcher une variable aléatoire U de $L^2(\Theta,P)$ à l'aide d'une expansion de la forme :

$$U = u_0\Gamma_0 + \sum_{i_1=1}^{+\infty} u_{i_1}\Gamma_{i_1}(\eta_{i_1}) + \sum_{i_1=1}^{+\infty}\sum_{i_2=1}^{i_1} u_{i_1i_2}\Gamma_{i_1i_2}(\eta_{i_1},\eta_{i_2}) + \sum_{i_1=1}^{+\infty}\sum_{i_2=1}^{i_1}\sum_{i_3=1}^{i_2} u_{i_1i_2i_3}\Gamma_{i_1i_2i_3}(\eta_{i_1},\eta_{i_2},\eta_{i_3}) + \dots \tag{4.4}$$

Les polynômes Γ_I sont tels que les variables aléatoires $\Gamma_I(\eta_I)$ sont mutuellement orthogonales, et ils sont donc différents des polynômes du chaos rencontrés précédemment, comme montré dans le tableau 4.1 qui présente différentes variables aléatoires possibles ainsi que les polynômes orthogonaux associés.

Le choix des variables aléatoires à considérer peut être motivé par différentes raisons. Il peut être choisi en fonction de la loi des paramètres incertains

du problème étudié, ou encore en fonction de la forme attendu de la loi de la variable que l'on cherche à approcher. En effet, la convergence de l'expansion en pratique semble d'autant mieux fonctionner que les polynômes du chaos généralisés utilisés sont construits sur des variables aléatoires ayant une loi proche de la loi de la variable que l'on souhaite approximer, ce qui a d'ailleurs motivé l'utilisation de polynômes du chaos généralisés construit à partir de la loi de la variable à approcher [101].

Théoriquement, l'utilisation de polynômes du chaos généralisés, ne dérivant pas forcément du schéma de Askey, a été justifiée en terme de convergence sous certaines conditions sur les lois des $(\eta_i)_{i\in\mathbb{N}^*}$, qui peuvent ne pas tous suivre la même loi, mais qui doivent être telles que leur fonction de répartition est continue, et que tous les moments soient finis. Ernst et al. [31] donne des conditions suffisantes sur les variables aléatoires $(\eta_i)_{i\in\mathbb{N}^*}$ pour que l'expansion (4.4) converge pour la norme 2 sur $L^2(\Theta, P)$. En particulier, dans le cas où les variables $(\eta_i)_{i\in\mathbb{N}^*}$ sont mutuellement indépendantes, une condition nécessaire et suffisante pour avoir la convergence de l'expansion en polynômes du chaos généralisés (4.4) est qu'il y ait une unique solution au problème des moments [50] pour chacun des η_i.

La condition précédente n'est par exemple pas remplie pour des lois log-normales, pour lesquelles le problème des moments ne possède pas une unique solution. Un exemple pour lequel une expansion de la forme (4.4) en des variables aléatoires η_i indépendantes et suivant toutes une loi log-normale est explicité dans [31]. L'expansion bien que convergente, ne converge pas vers la variable aléatoire qu'elle doit approximer.

Le dernier exemple montre que le choix des variables aléatoires à utiliser pour le paramétrage de la surface de réponse, et donc les polynômes du chaos qui seront utilisés a une importance théorique pour être capable d'approximer correctement la variable aléatoire d'intérêt. Dans les cas rencontrés dans cette thèse, l'incertitude du système étudié provient de paramètres d'entrée, et il peut alors être intéressant de construire une surface de réponse dépendant de ces paramètres d'entrée.

4.1.1.3 Polynômes du chaos en les variables aléatoires d'entrée du problème

Les précédentes sections s'intéressaient à l'approximation d'une variable aléatoire de $L^2(\Theta, P)$ à l'aide de polynômes en des variables aléatoires faisant partie d'une famille dénombrable de variables aléatoires $(\eta_i)_{i\in\mathbb{N}^*}$. Sous de bonnes conditions, pour une famille de variables aléatoires $(\eta_i)_{i\in\mathbb{N}^*}$ donnée, il est possible d'approcher n'importe quelle variable aléatoire de $L^2(\Theta, P)$ par une expansion en polynômes du chaos en ces variables aléatoires.

Dans les problèmes de propagation d'incertitudes rencontrés dans cette thèse, la donnée de départ est un ensemble de n paramètres qui sont des va-

riables aléatoires réelles $(K_i)_{i\in[\![1,n]\!]}$.

$$K_i : (\Theta, \Sigma) \to (\mathbb{R}, \mathcal{B}(\mathbb{R})) \tag{4.5}$$

La réponse S *du système étudié dépend directement de ces paramètres d'entrée, et peut être vue comme une fonction* f *mesurables des* K_i *:*

$$S = f(K_1, \ldots, K_n) \tag{4.6}$$

S *est donc également une variable aléatoire définie sur l'espace probabilisé* (Θ, Σ, P)*, et il est donc possible de s'intéresser à des probabilités d'événements concernant la réponse du système* S*. On supposera de plus que* S *est un élément de* $L^2(\Theta, P)$.

Plutôt que de chercher à exprimer S *comme une expansion en polynômes du chaos d'une famille de variables aléatoires* $(\eta_i)_{i\in\mathbb{N}^*}$ *que l'on a préalablement choisie, il peut être intéressant d'exprimer* S *comme une expansion en polynôme du chaos des variables aléatoires dont il dépend, à savoir les variables aléatoires* $(K_i)_{i\in[\![1,n]\!]}$*. Dans [31] sont données des conditions suffisantes sur la loi des paramètres aléatoires* $(K_i)_{i\in[\![1,n]\!]}$ *pour que* S *possède une expansion en polynômes du chaos dépendants des* $(K_i)_{i\in[\![1,n]\!]}$*. Dans le cas où les paramètres aléatoires ne sont pas mutuellement indépendants, il est possible de construire une base de fonctions non polynomiales ne dépendant que des* $(K_i)_{i\in[\![1,n]\!]}$ *à partir des polynômes du chaos correspondant aux distributions marginales [128]. Dans le cas où la réponse du système est une variable aléatoire, elle peut donc en théorie être approchée aussi près que l'on veut, pour la norme hermitienne, par une expansion en polynômes du chaos construits à l'aide de la loi des paramètres incertains du système, sous certaines conditions sur cette loi. En fait, comme il a été vu précédemment dans le chapitre 3, il est possible par un changement de variables de considérer d'autres variables aléatoires d'entrée pour paramétrer l'incertitude du système, notamment des variables aléatoires suivant une loi uniforme sur* $[0, 1]$*. Différents choix de paramétrisation des incertitudes d'entrées mènent à différents polynômes du chaos, et donc à une convergence plus ou moins rapide, ainsi qu'une convergence possible ou peut être impossible, de l'expansion en polynômes du chaos.*

Dans le cas où la réponse du système étudié est une variable aléatoire, il est donc en général possible de l'exprimer comme une expansion en polynôme du chaos des paramètres d'entrée. Cependant, la réponse du système peut prendre des formes plus complexes qu'une simple variable aléatoire, comme un vecteur aléatoire ou un processus stochastique.

4.1.1.4 Polynôme du chaos pour des processus stochastiques

Un vecteur aléatoire pouvant être vu comme un processus stochastique indexé par un ensemble fini, le discours de cette section ne portera que sur les processus stochastique. On considère donc un processus stochastique U :

$$U : \Omega \times \Theta \to \mathbb{R} \tag{4.7}$$

Ω correspond à l'ensemble d'indexation du processus stochastique, que l'on nommera également espace déterministe dans la suite, qui peut par exemple être le temps ou l'espace. On suppose que pour chaque $x \in \Omega$, la variable aléatoire $U(x, .)$ possède une variance, c'est à dire qu'elle est un élément de $L^2(\Theta, P)$.

Ainsi, si l'on considère des polynômes du chaos $\Gamma_I(\mathbf{K_I})$ en l'ensemble des paramètres incertains $\mathbf{K}$, on peut construire une expansion pour $U(x, .)$ (en tant que variable aléatoire de carré intégrable) de la forme :

$$U(x, \mathbf{K}) = \sum_{I \in \mathcal{I}} u_I(x) \Gamma_I(\mathbf{K_I}) \tag{4.8}$$

On peut ainsi construire des fonctions $u_I : \Omega \to \mathbb{R}$ pour chacun des polynômes du chaos $\Gamma_I(\mathbf{K_I})$. Ces fonctions sont généralement appelées des modes du processus stochastiques, et elles traduisent à quel point le processus stochastique est corrélé en un point du domaine déterministe avec une des variables aléatoires $\Gamma_I(\mathbf{K_I})$.

4.1.2 Méthodes non intrusive pour la construction de PCE

On s'intéresse dans cette section à la construction numérique d'expansion en polynômes du Chaos en les variables aléatoires d'entrée d'une quantité d'intérêt du système étudié, qui correspond en pratique au calcul des coefficients de l'expansion. Plus particulièrement, seules les méthodes non intrusives seront présentées ici, laissant de côté les méthodes intrusives [65]. Les méthodes non intrusives sont plus simples de mise en œuvre que les méthodes intrusives, nécessitant simplement des évaluations de la quantité d'intérêt, là où les méthodes intrusives s'appuient sur une équation d'évolution de celle-ci.

Pour la suite, on suppose que le système est paramétré à l'aide d'un vecteur aléatoire ξ à valeurs dans $\mathbb{R}^d$ et de densité de probabilité π. On suppose également qu'on possède une famille libre de M polynômes orthonormaux $(P_k)_{1 \leq k \leq M}$ pour la mesure de probabilité définie par la densité de probabilité π.

4.1.2.1 Méthode projective par minimisation de distance

L'idée des méthodes projectives par minimisation d'une distance $||.||$ consiste à choisir au sein de l'espace vectoriel engendré par la famille libre de polynômes

orthonormaux l'élément le plus proche de la quantité d'intérêt Q que l'on souhaite représenter. Un élément de cet espace vectoriel est caractérisé par les coefficients $(\alpha_k)_{1\leq k\leq M}$ de sa combinaison linéaire des éléments de la base, de sorte que la solution au problème est formellement définie par :

$$(\alpha_1,\ldots,\alpha_M) = \underset{(\alpha_1,\ldots,\alpha_M)\in\mathbb{R}^d}{\operatorname{argmin}} \left|\left| Q - \sum_{k=1}^{M} \alpha_k P_k \right|\right| \tag{4.9}$$

En cas d'utilisation de la norme $||.||_2$ de l'espace des variables aléatoires de carrés intégrables, la minimisation précédente revient à la détermination de la projection orthogonale de la quantité Q sur le sous espace engendré par la base de polynômes orthonormaux $(P_k)_{1\leq k\leq M}$. Comme il est équivalent de minimiser la norme ou le carré de la norme, l'expression (4.9) dans le cadre de la norme $||.||_2$ peut se réécrire de la façon suivante :

$$\begin{aligned}(\alpha_1,\ldots,\alpha_M) &= \underset{(\alpha_1,\ldots,\alpha_M)\in\mathbb{R}^d}{\operatorname{argmin}} E\left[\left(Q - \sum_{k=1}^{M}\alpha_k P_k\right)^2\right] \\ &= \underset{(\alpha_1,\ldots,\alpha_M)\in\mathbb{R}^d}{\operatorname{argmin}} \int_{\mathbb{R}^d}\left(Q(\xi) - \sum_{k=1}^{M}\alpha_k P_k(\xi)\right)^2 \pi(\xi)d\xi\end{aligned} \tag{4.10}$$

L'intégrale dans l'expression (4.10) peut de manière générale s'évaluer à l'aide d'une formule de cubature impliquant des couples de poids et de points $(w^{(i)},\xi^{(i)})_{1\leq i\leq N}$, de sorte que le problème à résoudre devient :

$$(\alpha_1,\ldots,\alpha_M) \approx \underset{(\alpha_1,\ldots,\alpha_M)\in\mathbb{R}^d}{\operatorname{argmin}} \left(\sum_{i=1}^{N} w^{(i)}\left[Q(\xi^{(i)}) - \sum_{k=1}^{M}\alpha_k P_k(\xi^{(i)})\right]^2\right) \tag{4.11}$$

La minimisation à effectuer dans l'expression (4.11) est un problème au moindres carrés linéaire, qui peut se mettre sous la forme matricielle (4.12), où la matrice W est la matrice avec l'ensemble des poids $w^{(i)}$ sur sa diagonale, P est la matrice de taille $M\times N$ dont les coefficients sont donnés par $P_{ij} = P_j(\xi^{(i)})$, $\hat{\mathbf{s}}$ est le vecteur colonne de coefficients $\hat{s}_i = \sum_{j=1}^{N}\alpha_j P_j(\xi^{(i)})$, et $\mathbf{s}$ est le vecteur colonne de coefficients $s_i = Q(\xi^{(i)})$ [65].

$$\left(P^T W P\right)\hat{\mathbf{s}} = P^T W \mathbf{s} \tag{4.12}$$

Certaines conditions sont à remplir pour que le système matriciel précédent

soit bien posé, dépendant notamment du choix des points de cubature [65]. Dans le cas où la cubature est issue d'une méthode de Monte Carlo, la méthode se retrouve, dans la limite d'un grand nombre d'échantillons, être identique à la méthode de projection présentée dans la section suivante.

4.1.2.2 Méthode projective orthogonale

La méthode précédente revient à déterminer la projection orthogonale d'une quantité d'intérêt sur un sous-espace vectoriel engendré par une base de vecteur orthonormaux, grâce à la résolution d'un problème d'optimisation. Le calcul directe de cette projection orthogonale peut être effectué directement, en calculant simplement les coefficients $(\alpha_k)_{1\leq k\leq M}$ *de cette projection orthogonale à l'aide de leur définition :*

$$\alpha_k = E\left[P_k Q\right] = \int_{\mathbf{R}^d} P_k(\xi) Q(\xi) \pi(\xi) d\xi \tag{4.13}$$

Numériquement, le calcul de l'intégrale se fait à l'aide d'une méthode de cubature, de sorte que les coefficients α_k *sont estimés par :*

$$\alpha_k \approx \sum_{i=1}^{N} w^{(i)} P_k(\xi^{(i)}) Q(\xi^{(i)}) \tag{4.14}$$

La convergence de l'estimation dépend bien entendu de la méthode de cubature utilisée, et sera donc liée aux propriétés de celle-ci et potentiellement à la régularité de la quantité d'intérêt Q*, les polynômes* P_k *de la base orthonormale étant pour leur part infiniment dérivables. Dans le cas où une méthode de cubature est utilisée permettant d'intégrer parfaitement les polynômes considérés dans la base, cette méthode est équivalent à la méthode de projection présentée dans la section suivante où la même cubature aurait été utilisée [65].*

La méthode de cubature utilisée dans l'expression (4.14) peut bien entendu être une méthode de Monte Carlo ou de Quasi-Monte Carlo randomisée. Dans de tels cas, les poids $w^{(i)}$ *sont tous égaux à* $1/N$*, et la vitesse convergence des coefficients* α_k *est la convergence habituelle de ces méthodes. Une estimation de cette convergence peut être effectuée en pratique, en s'intéressant à la variance du produit* $P_k Q$ *qui peut être estimé dans le cas de la méthode de Monte Carlo ou de la méthode de Quasi-Monte Carlo randomisé.*

Cette méthode consistant à calculer directement les coefficients de projection α_k *à l'aide d'un calcul d'intégrale a été choisie dans la suite de cette thèse, pour sa simplicité de mise en œuvre du fait du développement des méthodes présentées dans les deux précédents chapitres, et également parce que les résultats de convergence de ces précédents chapitres s'appliquent directement à l'estimation des coefficients. La méthode de minimisation de distance précédemment*

présentée, légèrement plus complexe à mettre en place (nécessité de résolution d'un système matricielle), possède de plus la particularité de converger vers les mêmes résultats dans le cas de l'utilisation des méthodes de Monte-Carlo [65], et est similaire à cette même méthode dans le cadre de méthodes de cubatures intégrant parfaitement la base de polynômes orthonormaux, rendant les deux méthodes très proches.

4.2 Application des méthodes projectives à l'étude de sensibilité globale

L'analyse de sensibilité d'une grandeur d'intérêt Q dépendant d'un vecteur de paramètres $\mathbf{A}$ consiste à s'intéresser à l'impact des différentes composantes du vecteur de paramètres $\mathbf{A}$ sur cette quantité d'intérêt Q. Un des intérêt de l'analyse de sensibilité particulièrement intéressant dans cette thèse est qu'elle permet de réduire la taille du vecteur de paramètres $\mathbf{A}$ en ne gardant que les composantes ayant un impact significatif sur Q.

Deux types d'analyses de sensibilité peuvent être distinguées [129] :

- *l'analyse de sensibilité local, s'intéressant à l'impact sur la quantité d'intérêt Q de petites perturbations des composantes du vecteur de paramètres $\mathbf{A}$ autour d'une valeur nominale $\mathbf{A}^0$ de celui-ci.*
- *l'analyse de sensibilité globale, s'intéressant à l'impact sur la quantité d'intérêt Q des différentes composantes du vecteur de paramètres $\mathbf{A}$ lorsque celles-ci prennent l'ensemble des valeurs possibles.*

Dans la présente thèse, l'analyse de sensibilité globale est privilégiée. Celle-ci se résume à la mesure de l'impact sur la quantité d'intérêt de chacune des composantes du vecteur de paramètres $\mathbf{A}$ par le rapport de deux variances : la variance de la quantité d'intérêt restreinte à ne dépendre que de certaines composantes du vecteur $\mathbf{A}$ et la variance globale de cette quantité d'intérêt.

4.2.1 Décomposition de la variance

La décomposition de la variance se base sur un des résultats de Sobol [124], permettant d'exprimer une fonction f définie sur l'hypercube unité $[0,1]^d$ de la manière suivante :

$$f(\mathbf{x}) = \sum_{I \subseteq [\![1,d]\!]} f_I(\mathbf{x}_I) \tag{4.15}$$

Dans l'expression (4.15), $\mathbf{x}_I$ est un point de $[0,1]^{|I|}$, où $|I|$ correspond au cardinal de l'ensemble I, construit à partir du vecteur $\mathbf{x}$ de $[0,1]^d$ en gardant la composante i de celui-ci si et seulement si $i \in I$. Les termes f_I sont obtenues

de proche en proche :

$$\begin{cases} f_\emptyset = \int_{[0,1]^d} f(\mathbf{x}) d\mathbf{x} \\ f_I(\mathbf{x}_I) = \int_{[0,1]^{d-|I|}} f(\mathbf{x}) d\mathbf{x}_{I^c} - \sum_{J \subset I} f_J(\mathbf{x}_J) \end{cases} \tag{4.16}$$

Dans les expressions précédentes, I^c désigne le complémentaire de I dans $[\![1, d]\!]$. Les expressions précédentes peuvent être interprété en terme de théorie des probabilités. En effet, les intégrales dans (4.16) peuvent s'interpréter en terme de calcul d'espérance et d'espérance conditionnelle présentée dans (4.17), le vecteur aléatoire $\mathbf{X}$ ayant une loi uniforme sur $[0,1]^d$, donc de densité de probabilité $\pi_{\mathbf{X}}$ constante égale à 1 sur l'hypercube $[0,1]^d$.

$$\begin{cases} f_\emptyset = E\left[f(\mathbf{X}\right] \\ f_I(\mathbf{X}_I) = E\left[f(\mathbf{X})|\mathbf{X}_I\right] - \sum_{J \subset I} f_J(\mathbf{x}_J) \end{cases} \tag{4.17}$$

On peut montrer qu'ainsi construits, les termes f_I forment une famille orthogonale pour le produit scalaire définie par la relation (4.18).

$$(f, g) = \int_{[0,1]^d} f(\mathbf{x}) g(\mathbf{x}) d\mathbf{x} = E\left[f(\mathbf{X}) g(\mathbf{X})\right] \tag{4.18}$$

Un exemple d'une telle décomposition est donnée pour la fonction de deux variables $f(x_1, x_2) = (x_1 + x_2)^2$ ci-dessous :

$$\begin{cases} f(x_1, x_2) = x_1^2 + x_2^2 + 2x_1x_2 = \frac{7}{6} + (x_1^2 - \frac{1}{3}) + (x_2^2 - \frac{1}{3}) + (2x_1x_2 - \frac{1}{2}) \\ f_\emptyset = \frac{7}{6} \\ f_1(x_1) = x_1^2 - \frac{1}{3} \\ f_2(x_2) = x_2^2 - \frac{1}{2} \\ f_{12}(x_1, x_2) = 2x_1x_2 - \frac{1}{2} \end{cases} \tag{4.19}$$

Il est possible, à l'aide d'un simple changement de variable, de transposer les résultats précédents au cas d'une quantité d'intérêt Q dépendant d'un vecteur aléatoire $\mathbf{A}$ à composantes indépendantes, de densité de probabilité π, qui pourrait se décomposer selon la première ligne de (4.20), les deux autres lignes

correspondant à la définition de l'ensemble des termes.

$$\begin{cases} Q(\mathbf{A}) = \sum\limits_{I \subseteq [\![1,d]\!]} Q_I(\mathbf{A}_I) \\ Q_\emptyset = \int_{\mathbb{R}^d} Q(\mathbf{A})\pi(\mathbf{A})d\mathbf{A} = E\left[Q(\mathbf{A})\right] \\ Q_I(\mathbf{A}_I) = \displaystyle\int_{\mathbb{R}^{d-|I|}} Q(\mathbf{A})\pi_{I^c}(\mathbf{A}_{I^c})d\mathbf{A}_{I^c} - \sum_{J \subset I} Q_J(\mathbf{A}_J) \\ \qquad\qquad = E\left[Q(\mathbf{A})|\mathbf{A}_I\right] - \displaystyle\sum_{J \subset I} Q_J(\mathbf{A}_J) \end{cases} \tag{4.20}$$

Il est immédiat par construction que cette expansion dépend de la loi du vecteur aléatoire $\mathbf{A}$*, et que changer cette loi revient également à changer les fonctions* Q_I*, qui sont donc intimement liée à la loi de* $\mathbf{A}$*. Du fait de l'orthogonalité de l'ensemble des termes* $Q_I(\mathbf{A}_I)$ *deux à deux, la variance de* $Q(\mathbf{A})$ *s'exprime de la façon suivante :*

$$Var\left[Q(\mathbf{A})\right] = \sum_{\substack{I \subseteq [0,1]^d \\ I \neq \emptyset}} Var\left[Q_I(\mathbf{A}_I)\right] \tag{4.21}$$

Cette décomposition de la variance de $Q(\mathbf{A})$ *permet de définir les indices de sensibilités globaux, ou indices de Sobol* S_I *[125] :*

$$S_I = \frac{Var\left[Q_I(\mathbf{A}_I)\right]}{Var\left[Q(\mathbf{A})\right]} \tag{4.22}$$

Par construction, la somme de l'ensemble des indices de Sobol est égale à 1*. L'indice* S_I *correspond à la proportion de la variance de la variable aléatoire* $Q(\mathbf{A})$ *expliquée par l'interaction des composantes d'indice* $i \in I$*. Le nombre total d'indices de Sobol est de* $2^d - 1$*, ce qui implique très vite un très grand nombre d'indices à étudier pour des systèmes paramétrés par un nombre* d *conséquent de paramètres incertains. Pour cette raison, les indices* $S_{\{i\}}$*, appelés indices de Sobol du premier ordre, correspondant à la variance dû uniquement à la composante* i *du vecteur* $\mathbf{A}$*, sont en pratique ceux qui seront le plus étudiés, avec potentiellement ceux impliquant l'interaction de deux composantes du vecteur aléatoire* $\mathbf{A}$*.*

L'estimation des indices de Sobol passe par l'évaluation de la variance des variables aléatoires $Q_I(\mathbf{A}_I)$*, consistant en des calculs d'intégrales souvent en grande dimension. Une méthode d'estimation des indices de Sobol de premier ordre due à Sobol [125] implique un total de* $N(d+1)$ *évaluations de la quantité d'intérêt* Q*,* N *évaluations correspondant à l'évaluation de la moyenne, et* N *autres évaluations étant nécessaires pour le calcul de chacun des indices. La*

convergence de ces estimations est liée au nombre N d'évaluations, de la même façon que dans une méthode de Monte-Carlo. La méthode $FAST$ [115] permet le calcul des mêmes indices de Sobol du premier ordre, mais contrairement à la méthode précédente, les N évaluations de la quantité d'intérêt Q peuvent être utilisées pour le calcul de tous les indices.

Ces deux dernières méthodes permettent une évaluation directe des seules indices de Sobol de premier ordre, le coût de calcul minimum étant celui de la méthode $FAST$. Il est cependant possible d'estimer ces indices de Sobol du premier ordre, ainsi que les indices d'ordre supérieur, et cela en considérant N évaluations qui pourront, tout comme pour la méthode $FAST$, être utilisée pour le calcul de chacun des indices.

4.2.2 HDMR et polynômes du chaos

Dans la littérature, la décomposition d'une fonction sous une forme similaire à celle de l'expression (4.15), sans imposer les contraintes (4.16), est appelée HDMR (High Dimensional Model Representation) [74]. Il existe plusieurs formes d'HDMR [75], et dans le cas où les termes de cette représentation vérifient en plus la condition d'orthogonalité des termes définie par la relation (4.16), la décomposition est dénommée RS-HDMR (Random Sampling-High Dimensional Model Representation) [74]. Une autre forme d'HDMR est la Cut-HDMR [75], dont les termes sont définis à l'aide des relations suivantes et dans lesquelles $\mathbf{x}_0$ est un point de $[0,1]^d$:

$$\begin{cases} f_\emptyset = f(\mathbf{x}_0) \\ f_I(\mathbf{x}_I) = f(\mathbf{x}_I^{(0)}) - \sum\limits_{J \subset I} f_J(\mathbf{x}_J) \end{cases} \tag{4.23}$$

Dans l'expression précédente, $\mathbf{x}_I^{(0)}$ est le vecteur dont les composantes sont égales aux composantes de $\mathbf{x}$ pour $i \in I$ et à celles de $\mathbf{x}_0$ sinon. La Cut-HDMR en $\mathbf{x}_0 = (0,0)$ de la fonction $f(x_1, x_2) = (x_1 + x_2)^2$ dont la RS-HDMR est donné par (4.19) est la suivante :

$$\begin{cases} f(x_1, x_2) = x_1^2 + x_2^2 + 2x_1x_2 \\ f_\emptyset = 0 \\ f_1(x_1) = x_1^2 \\ f_2(x_2) = x_2^2 \\ f_{12}(x_1, x_2) = 2x_1x_2 \end{cases} \tag{4.24}$$

L'exemple choisi permet d'illustrer que les deux HDMR peuvent être différentes. Dans cette thèse, seule la RS-HDMR est considérée car c'est celle qui est nécessaire à l'obtention des indices de Sobol précédemment introduits. En effet,

la connaissance des termes de cette représentation permet la détermination des indices de Sobol associés. La dénomination de RS-HDMR est parfois réservé aux fonctions dépendant d'un vecteur aléatoire de loi uniforme sur l'hypercube unité $[0,1]^d$*. Dans la suite, le terme sera employé indifféremment pour toute fonction* Q *du vecteur de paramètres incertains* $\mathbf{A}$*, tant que les composantes de ce dernier vecteur sont indépendantes, étant donné qu'un simple changement de variable permet alors de se ramener au cas d'un vecteur aléatoire de loi uniforme sur l'hypercube unité* $[0,1]^d$*.*

Du fait de la définition des termes de la RS-HDMR d'une quantité Q*, une première idée pour approximer ceux-ci est l'utilisation d'une méthode de Monte-Carlo pour le calcul des intégrales. Cette approche est cependant très onéreuse en terme de calcul [74], ce qui motive l'idée d'approximer ces termes à l'aide d'une combinaison linéaire de fonctions d'une base. Les polynômes du Chaos présentés précédemment sont d'excellents candidats pour remplir cette tâche [74]. On considère donc une base orthonormale de polynôme du chaos* $(P_n(\mathbf{A}))_{n\in\mathbb{N}}$ *que l'on sépare en sous-ensemble* $(P_{I,k}(\mathbf{A}))_{1\leq k\leq+\infty}$ *pour chaque sous ensemble* I *de* $[\![1,d]\!]$*, les polynômes du chaos* $P_{I,k}(\mathbf{A})$ *dépendant exactement des composantes* i *du vecteur aléatoire* $\mathbf{A}$ *avec* $i \in I$*, et de fait vérifient la relation abusive suivante :*

$$P_{I,k}(\mathbf{A}) = P_{I,k}(\mathbf{A}_I) \tag{4.25}$$

La variable aléatoire $Q(\mathbf{A})$ *peut alors se décomposer selon cette base de polynômes du chaos sous la forme suivante :*

$$Q(\mathbf{A}) = \sum_{n=0}^{+\infty} \alpha_n P_n(\mathbf{A}) = \alpha_\emptyset + \sum_{I\in[\![1,d]\!],I\neq\emptyset} \left(\sum_{k=1}^{+\infty} \alpha_{I,k} P_{I,k}(\mathbf{A}_I) \right) \tag{4.26}$$

En injectant l'expression de $Q(\mathbf{A})$ *donnée par (4.26) dans le système (4.20), il est possible de montrer par récurrence les relations suivantes :*

$$\begin{cases} Q_\emptyset = \alpha_\emptyset \\ \forall I \subset [\![1,d]\!], I \neq \emptyset, Q_I(\mathbf{A}_I) = \sum\limits_{k=1}^{+\infty} \alpha_{I,k} P_{I,k}(\mathbf{A}_I) \end{cases} \tag{4.27}$$

En pratique, les expressions précédentes sont tronquées, seul un nombre M_I *de polynômes du chaos étant utilisés pour chacun des sous ensembles* I *de* $[\![1,d]\!]$*. Ce dernier système et la remarque précédente montre également qu'il est possible de construire de manière approchée la RS-HDMR à l'aide d'une expansion en polynômes du chaos. Dans l'expression (4.27), chaque coefficient* $\alpha_{I,k}$ *correspond au produit scalaire de* $Q(\mathbf{A})$ *avec le polynôme* $P_{I,k}(\mathbf{A}_I)$*, et est*

donc donné par l'expression suivante :

$$\alpha_{I,k} = E\left[P_{I,k}(\mathbf{A}_I)Q(\mathbf{A})\right] \tag{4.28}$$

L'enjeu est maintenant d'exprimer l'indice de Sobol S_I en fonction des coefficients $\alpha_{I,k}$ donnés par la relation précédente.

4.2.3 Estimation des indices de Sobol

En prenant la variance de chacun des membres de l'expression (4.27), et en se servant du caractère orthonormal de la famille des polynômes du chaos $P_{I,k}(\mathbf{A})$, il vient :

$$Var\left[Q_I(\mathbf{A}_I)\right] \approx \sum_{k=1}^{M_I} \alpha_{I,k}^2 \tag{4.29}$$

L'évaluation de cette variance repose donc sur l'estimation des termes $\alpha_{I,k}^2$. Dans le cas de l'utilisation d'une méthode de cubature, le terme $\alpha_{I,k}^2$ sera simplement pris comme le carré du résultat du calcul numérique de l'intégrale (4.28). Dans le cas de l'utilisation d'une méthode de Monte Carlo ou de Quasi-Monte Carlo randomisé, des estimateurs sans biais de ces termes sont utilisés dans cette thèse qui sont ceux donnés dans [73]. Dans le cas où une méthode de Monte Carlo est utilisée, pour laquelle on suppose l'utilisation d'un échantillon de taille N, l'estimation sans biais $\alpha_{I,k,MC}^2$ de $\alpha_{I,k}^2$ utilisée dans cette thèse est donnée par l'expression suivante :

$$\alpha_{I,k,MC}^2 = \frac{N}{N-1}\left[\left(\frac{1}{N}\sum_{i=1}^{N} P_{I,k}(\mathbf{A}_I^{(i)})Q(\mathbf{A}^{(i)})\right)^2 - \frac{1}{N^2}\sum_{i=1}^{N} P_{I,k}(\mathbf{A}_I^{(i)})^2 Q(\mathbf{A}^{(i)})^2\right] \tag{4.30}$$

Étant donné une estimation $\sigma_{Q,MC}^2$ de la variance de $Q(\mathbf{A})$ obtenue à l'aide des N échantillons de Monte Carlo, l'indice de Sobol $S_{I,MC}$ est alors estimé par :

$$S_{I,MC} = \frac{1}{\sigma_{Q,MC}^2}\sum_{k=1}^{M_I} \alpha_{I,k,MC}^2 \tag{4.31}$$

Dans le cas d'une méthode de Quasi-Monte Carlo randomisée, pour laquelle on suppose l'utilisation de Q séquences de N points chacune, l'estimation

sans biais $\alpha^2_{I,k,QMC}$ *est quant à elle donnée par l'expression suivante :*

$$\begin{aligned}\alpha^2_{I,k,QMC} = \frac{Q}{Q-1}\left[\frac{1}{NQ}\sum_{q=1}^{Q}\sum_{i=1}^{N} P_{I,k}(\mathbf{A}_I^{(i,q)})Q(\mathbf{A}^{(i,q)})\right]^2 \\ - \frac{1}{Q(Q-1)}\sum_{q=1}^{Q}\left(\sum_{i=1}^{N} P_{I,k}(\mathbf{A}_I^{(i,q)})Q(\mathbf{A}^{(i,q)})\right)^2\end{aligned} \tag{4.32}$$

Étant donné une estimation $\sigma^2_{Q,QMC}$ *de la variance de* $Q(\mathbf{A})$ *obtenue à l'aide des* Q *séquences de* N *échantillons de Quasi-Monte Carlo randomisé, l'estimation de l'indice de Sobol* $S_{I,QMC}$ *est alors donnée par :*

$$S_{I,QMC} = \frac{1}{\sigma^2_{Q,QMC}}\sum_{k=1}^{M_I}\alpha^2_{I,k,QMC} \tag{4.33}$$

L'utilisation d'une méthode de Quasi-Monte Carlo randomisée offre en général de meilleurs résultats que la méthode de Monte Carlo, et sera utilisée dans la suite de cette thèse pour l'analyse de sensibilité globale. Elle offre de plus, comme vu dans le chapitre précédent, de pouvoir obtenir rapidement une estimation de l'erreur commises sur les statistiques, telles que les indices de Sobol calculés ici, à l'aide de méthodes de Jack-knife ou de bootstrap.

Les outils présentés depuis le début de ce chapitre permettent, pour une quantité d'intérêt dépendant d'un ensemble de paramètres incertains, de déterminer les paramètres impactant le plus cette quantité d'intérêt afin de n'en garder qu'un nombre réduit. A partir de là, il est possible d'approximer la quantité d'intérêt comme une somme de polynômes en ces paramètres que l'on a conservé. Cette démarche contraint à se servir des paramètres initiaux, ce qui n'est pas toujours souhaitable, notamment lorsque ceux-ci ont tous un impact du même ordre ne permettant pas d'en éliminer un plutôt qu'un autre. La suite de ce chapitre s'intéresse à une méthode permettant de faire émerger de nouveaux paramètres incertains différents des paramètres initiaux, rangés par ordre d'importance décroissant de sorte que seuls les premiers peuvent être gardés.

4.3 Expansion de Karhunen-Loève

L'expansion de Karhunen-Loève [59], également connue sous le nom de POD (Proper Orthogonal Decomposition) ou PCA (Principal Component Analysis) en dimension finie, permet de représenter un processus stochastique de manière spectrale, optimale dans un certain sens, en utilisant les informations issues de sa fonction d'auto-corrélation.

4.3.1 Cadre du problème

On considère un processus stochastique réel U défini sur un espace $D \times \Omega$, où D est l'ensemble des paramètres déterministes, que nous supposerons compact, et $(\Omega, \mathcal{F}, P)$ est un espace probabilisé. On fait l'hypothèse que pour tout élément $\omega \in \Omega$, $U(.,\omega) \in L^2(D)$ qui est l'espace des fonctions de carré intégrable sur D. L'espace $L^2(D)$ est naturellement muni d'un produit scalaire $(.,.)$ définie par l'expression (4.34), auquel est associé la norme $||.||_D$, et qui fait de $L^2(D)$ un espace de Hilbert.

$$\forall u, v \in L^2(D), (u, v) = \int_D u(x)v(x)dx \tag{4.34}$$

Quitte à remplacer U par $U - E[U]$, on ne considère dans la suite que des processus stochastiques centrés, c'est à dire de moyenne nulle :

$$\forall x \in D, E[U(x,.)] = \int_\Omega U(x,\omega)P(d\omega) = 0 \tag{4.35}$$

On fait également l'hypothèse que U possède un moment d'ordre 2, c'est à dire que :

$$\forall x \in D, E\left[U(x,.)^2\right] < +\infty \tag{4.36}$$

Enfin, on fait l'hypothèse que U est continu au sens de (4.37).

$$\forall x, x' \in D, \lim_{x' \to x} E\left[U(x,.) - U(x',.)\right] = 0 \tag{4.37}$$

U possédant un moment d'ordre 2, il est possible de définir sa fonction d'auto-covariance C_{UU}, dont l'expression est donnée par (4.38) pour U centré.

$$\forall x, x' \in D, C_{UU}(x, x') = E\left[U(x,.)U(x',.)\right] \tag{4.38}$$

Sous les hypothèses faites précédemment, on peut montrer que la fonction d'auto-covariance C_{UU} est continue sur D^2 et qu'elle vérifie en plus la relation (4.39).

$$\int_D \int_D C_{UU}(x, x')dxdx' < +\infty \tag{4.39}$$

On peut alors définir l'opérateur linéaire K sur $L^2(D)$, basé sur le noyau

d'auto-covariance C_{UU}, et défini grâce à la relation (4.40).

$$\forall u \in L^2(D), \forall x \in D, (Ku)(x) = \int_D C_{UU}(x, x')u(x')dx' \tag{4.40}$$

Comme toute fonction d'auto-covariance, C_{UU} est symétrique et positive [139]. Le théorème de Mercer [6] s'applique donc à l'opérateur linéaire K, qui énonce que :

- *K possède une infinité dénombrable de valeurs propres positives $(\lambda_i)_{i\in\mathbb{N}^*}$, que l'on peut ranger en ordre décroissant : $\lambda_1 \geq \lambda_2 \geq \ldots$.De plus, la trace de K est finie (cf. (4.39)), et donc : $\sum_{i=1}^{+\infty} \lambda_i = \int_D \int_D C_{UU}(x, x')dxdx' < +\infty$*
- *Il existe une base hilbertienne (u_i) de $L^2(D)$ de fonctions propres pour K, qui sont continues lorsqu'elles sont associées à une valeur propre non nulle.*
- *C_{UU} admet la représentation $C_{UU}(x, x') = \sum_{i=1}^{+\infty} \lambda_i u_i(x)u_i(x')$, où la convergence a lieu de manière absolue et uniforme, et également pour la norme $||.||_D$.*

Les valeurs propres λ et les fonctions propres u de K, vérifiant $Ku = \lambda u$, sont solutions de l'équation intégrale de Fredholm du second-type suivante :

$$\forall x \in D, \int_D C_{UU}(x, x')u(x')dx' = \lambda u(x) \tag{4.41}$$

Possédant une base de Hilbert de $L^2(D)$, et $U(., \omega)$ appartenant à $L^2(D)$ pour tout $\omega \in \Omega$, il est possible d'exprimer U sous la forme d'une série présentée en (4.42), où la convergence de la série a lieue au sens de la norme $||.||_D$.

$$\forall \omega \in \Omega, \lim_{N\to+\infty} \left\| U(., \omega) - \sum_{i=1}^{N} \xi_i(\omega)u_i \right\|_D = 0 \tag{4.42}$$

Les coefficients $\xi(\omega)$ sont des variables aléatoires, qui correspondent aux coefficients de la projection orthogonale de $U(., \omega)$ sur la base des u_i, c'est à dire que :

$$\forall \omega \in \Omega, \xi_i(\omega) = (U(., \omega), u_i) \tag{4.43}$$

On peut montrer [6] que ces variables aléatoires sont centrées, qu'elles sont

deux à deux décorrélées et que leur variance est égale à la valeur propre associée :

$$\begin{cases} E\left[\xi_i\right] = 0 \\ E\left[\xi_i \xi_j\right] = \lambda_i \delta_{ij} \end{cases} \tag{4.44}$$

On peut également montrer [6] que l'erreur $\epsilon_n(x)$ défini dans (4.45) converge uniformément sur D, et donc également simplement pour tout point x de D.

$$\forall x \in D, \epsilon_n(x) = E\left[U(x) - \sum_{i=1}^{n} \xi_i u_i(x)\right] \tag{4.45}$$

Cette dernière convergence assure que pour chaque variable aléatoire $U(x,.)$, la série converge en norme 2 sur $L^2(\Omega)$, et donc également en probabilité et en loi. Si la valeur propre λ_i est strictement positive, il est possible de normer la variable aléatoire ξ_i pour obtenir une nouvelle variable aléatoire η_i centrée et de variance égale à 1. Cette nouvelle variable aléatoire est donc définie comme :

$$\forall \omega \in \Omega, \eta_i(\omega) = \frac{1}{\sqrt{\lambda_i}}(U(.,\omega), u_i) \tag{4.46}$$

Ces nouvelles variables aléatoires η_i sont centrées et réduites, et orthogonales deux à deux, c'est à dire décorrélées. On peut alors définir l'expansion de Karhunen-Loève du processus stochastique U, dont la convergence a lieue dans $L^2(\Omega)$, sous la forme généralement rencontrée, comme :

$$\forall x \in D, \forall \omega \in \Omega, U(x,\omega) = \sum_{i=1}^{+\infty} \sqrt{\lambda_i} u_i(x) \eta_i(\omega) \tag{4.47}$$

Si les λ_i sont nulles à partir d'un certain rang, les η_i ne peuvent plus être définis à partir de ce rang mais cela n'a que peu d'importance car alors l'expansion de Karhunen-Loève n'a qu'un nombre fini de termes et les termes correspondants à ces λ_i nulles n'apparaissent pas, ne nécessitant pas la définition des η_i.

4.3.2 Propriété de l'expansion de Karhunen-Loève

Une propriété intéressante de l'expansion de Karhunen-Loève d'un processus stochastique U est qu'elle est l'expansion sous forme de série similaire optimale dans un sens que nous allons préciser. Soit le processus stochastique U_n défini comme la troncation de l'expansion de Karhunen-Loève de U à l'ordre

n :

$$\forall x \in D, \forall \omega \in \Omega, U_n(x,\omega) = \sum_{i=1}^{n} \sqrt{\lambda_i} u_i(x) \eta_i(\omega) \tag{4.48}$$

Alors, U_n est l'expansion de n termes optimales [65] au sens où elle minimise l'erreur quadratique moyenne $r_n = E\left[\|U - U_n\|_D\right]$, erreur dont la valeur peut être obtenue grâce aux propriétés d'orthogonalité et qui est donnée par :

$$\begin{aligned} r_n &= E\left[\|U - U_n\|_D\right] = E\left[(U - U_n, U - U_n)\right] \\ &= E\left[\left(\sum_{i=n+1}^{+\infty} \sqrt{\lambda_i} u_i \eta_i, \sum_{i=n+1}^{+\infty} \sqrt{\lambda_i} u_i \eta_i\right)\right] \\ &= \sum_{i=n+1}^{+\infty} \sum_{j=n+1}^{+\infty} \sqrt{\lambda_i \lambda_j} (u_i, u_j) E\left[\eta_i \eta_j\right] \\ &= \sum_{i=n+1}^{+\infty} \lambda_i \end{aligned} \tag{4.49}$$

L'erreur quadratique moyenne entre le processus stochastique U et son expansion tronquée U_n à l'ordre n est donc directement reliée aux valeurs propres λ_i. Plus la décroissance de celle-ci est rapide, plus petite sera cette erreur et meilleure sera l'approximation. En fait, la somme des valeurs propres λ_i, qui est la trace de l'opérateur linéaire K, peut être reliée à la variance de U comme le montre le calcul suivant :

$$\begin{aligned} \int_D Var\left[U(x,.)\right] dx &= \int_D E\left[\left(\sum_{i=1}^{+\infty} \sqrt{\lambda_i} u_i(x) \eta_i\right)^2\right] dx \\ &= \int_D \sum_{i=1}^{+\infty} \sum_{j=1}^{+\infty} \sqrt{\lambda_i \lambda_j} u_i(x) u_j(x) E\left[\eta_i \eta_j\right] dx \\ &= \sum_{i=1}^{+\infty} \lambda_i \int_D u_i(x) dx = \sum_{i=1}^{+\infty} \lambda_i \end{aligned} \tag{4.50}$$

On peut calculer de même l'intégrale de la variance de U_n, et l'on trouve cette fois-ci :

$$\int_D Var\left[U(x,.)\right] dx = \sum_{i=1}^{n} \lambda_i \tag{4.51}$$

U_n reproduit donc une fraction α_n de l'intégrale sur D de la variance de U, fraction donnée par la relation (4.52).

$$\alpha_n = \frac{\sum_{i=1}^{n} \lambda_i}{\sum_{i=1}^{+\infty} \lambda_i} \tag{4.52}$$

De plus, l'optimalité de l'expansion de Karhunen-Loève implique que la somme des n premières valeurs propres est maximale, de sorte que la fraction α_n est maximale pour une expansion à n termes. Cette valeur de α_n peut servir d'indicateur pour savoir combien de termes doivent être gardés dans la somme. Cependant, d'autres critères peuvent rentrer en ligne de compte. En effet, cette reproduction de la variance de U sur D n'a aucune raison d'être uniforme, et son caractère maximal implique que l'ajout d'un terme à l'expansion de Karhunen-Loève aura tendance à vouloir reproduire le plus possible de l'intégrale de cette variance, ce qui pousse à reproduire en premier les zones de D où la variance de U est la plus importante. Si le phénomène étudié est sensible aux valeurs prises par le processus stochastique U en des points x de D où la variance de celui-ci est faible, un grand nombre de termes peuvent être nécessaire dans l'expansion de Karhunen-Loève pour être capable de reproduire ce phénomène. Il peut alors être intéressant de considérer un processus stochastique $\hat{U}$ correspondant à une modification de U. Ce phénomène sera exploité dans le chapitre 8.

Néanmoins, une fois tronquée à l'ordre n, l'expansion de Karhunen-Loève permet de reproduire un processus stochastique en ne nécessitant que n variables aléatoires décorrélées. Celles-ci, bien que décorrélées, ne sont en général pas indépendantes.

4.3.3 Méthode de résolution numérique de l'équation de Fredholm

La détermination de l'expansion de Karhunen-Loève repose sur la résolution de l'équation intégrale de Fredholm du second type (4.41). Cette même équation repose sur la connaissance de la fonction d'auto-covariance du processus stochastique que l'on veut reproduire, qui peut ou bien être donnée pour des raisons théoriques ou de modélisations, ou bien ne pas l'être. Si celle-ci n'est pas donnée, elle peut néanmoins être estimée à l'aide de réalisations du processus stochastique lui même, réalisations pouvant provenir de données expérimentales ou encore de données de simulations numériques. Une fois connue la fonction d'auto-covariance, il ne reste plus qu'à résoudre l'équation (4.41) pour obtenir les valeurs propres λ_i et les fonctions propres u_i présent dans l'expansion de Karhunen-Loève. Ce travail de résolution ne peut que très rarement s'effectuer de manière analytique, et il existe donc plusieurs méthodes de résolutions nu-

mériques ayant été développées qui sont présentées en détail dans [10], et qui vont être présentées dans cette section.

4.3.3.1 Méthodes à noyaux dégénérés

La méthode à noyaux dégénérés consiste à approximer le noyau C_{UU} dans l'équation intégral de Fredholm (4.41) par une suite de noyaux $C_{UU}^{(n)}$, dits dégénérés, convergeant vers le noyau C_{UU} et s'écrivant sous la forme :

$$\forall x, x' \in D, C_{UU}^{(n)}(x, x') = \sum_{k=1}^{n} \alpha_k^{(n)}(x)\beta_k^{(n)}(x') \tag{4.53}$$

La méthode repose sur le fait qu'une solution $(\lambda_i^{(n)}, u_i^{(n)})$ à l'équation intégrale de Fredholm du second type (4.54) impliquant le noyau $C_{UU}^{(n)}$ va converger vers la solution (λ_i, u_i) de l'équation de Fredholm (4.41), et cela d'autant plus vite que les noyaux convergent vite vers C_{UU}.

$$\forall x \in D, \int_D C_{UU}^{(n)}(x, x')u^{(n)}(x')dx' = \lambda^{(n)}u^{(n)}(x) \tag{4.54}$$

Des expressions (4.53) et (4.54), on peut exprimer $u^{(n)}$ comme :

$$\begin{aligned} \forall x \in D, u^{(n)}(x) &= \frac{1}{\lambda^{(n)}} \sum_{j=1}^{n} \alpha_j^{(n)}(x) \int_D \beta_j^{(n)}(x')u^{(n)}(x')dx' \\ &= \frac{1}{\lambda^{(n)}} \sum_{j=1}^{n} c_j^{(n)}\alpha_j^{(n)}(x) \end{aligned} \tag{4.55}$$

Dans l'expression (4.55) ont été introduit les nombres $c_j^{(n)}$ qui sont donnés par la relation (4.56), qui fait encore intervenir l'inconnue $u^{(n)}$.

$$\forall 1 \leq j \leq n, c_j^{(n)} = \int_D \beta_k^{(n)}(x')u^{(n)}(x')dx' \tag{4.56}$$

En multipliant les deux membres de (4.54) par $\beta_i^{(n)}(x)$, et en intégrant sur

D, on obtient pour le membre de gauche :

$$\begin{aligned}\int_D \beta_i^{(n)}(x) \int_D C_{UU}^{(n)}(x,x')u^{(n)}(x')dx'dx &= \int_D \beta_i^{(n)}(x) \int_D \sum_{j=1}^{n} \alpha_j^{(n)}(x)\beta_i^{(n)}(x')u^{(n)}(x')dx'dx \\ &= \sum_{j=1}^{n} \int_D \alpha_j^{(n)}(x)\beta_i^{(n)}(x)dx \int_D \beta_i^{(n)}(x')u^{(n)}(x')dx' \\ &= \sum_{j=1}^{n} (\alpha_j^{(n)}, \beta_i^{(n)})c_j^{(n)}\end{aligned} \tag{4.57}$$

En multipliant le membre de droite par $\beta_i^{(n)}(x)$, et en intégrant sur D, on obtient, en substituant $u^{(n)}$ par son expression dans (4.55) :

$$\lambda^{(n)} \int_D \beta_i^{(n)}(x)u^{(n)}(x)dx = \lambda^{(n)}c_i^{(n)} \tag{4.58}$$

En faisant pour toutes les valeurs de i entre 1 et n, on trouve finalement que les $c_j^{(n)}$ sont solutions du système linéaire suivant :

$$\forall 1 \leq i \leq n, \sum_{j=1}^{n} (\alpha_j^{(n)}, \beta_i^{(n)})c_j^{(n)} = \lambda^{(n)}c_i^{(n)} \tag{4.59}$$

En notant $A^{(n)}$ la matrice dont les coefficients sont $(\alpha_j^{(n)}, \beta_i^{(n)})$, et $C^{(n)}$ la matrice colonne contenant les coefficients $c_j^{(n)}$, ce système linéaire peut se réécrire sous une forme matricielle :

$$A^{(n)}C^{(n)} = \lambda^{(n)}C^{(n)} \tag{4.60}$$

Une fois choisies les fonctions $\alpha_j^{(n)}$ et $\beta_j^{(n)}$ qui composent les noyaux dégénérées, la méthode à noyaux dégénérés consiste donc en la résolution d'un problème linéaire aux valeurs propres de taille n. La matrice $A^{(n)}$ n'a ici aucune propriété particulière, notamment de symétrie, si l'on ne fait pas d'hypothèses sur les $\alpha_j^{(n)}$ et $\beta_j^{(n)}$.

4.3.3.2 Méthodes projectives

Les méthodes projectives peuvent être classées en deux grandes familles, que l'on va expliciter ici. Les méthodes de collocation et les méthodes de Galerkin.

L'idée dans ces deux familles de méthodes est de choisir une famille libre de n fonctions $(\phi_1, \ldots, \phi_n)$ de $L^2(D)$, et de rechercher une solution approchée $\hat{u}$ à l'équation (4.41) dans le sous-espace $\Phi_n = Vect(\phi_1, \ldots, \phi_n)$ engendré par ces fonctions, et qui a donc la forme :

$$\hat{u}^{(n)} = \sum_{i=1}^{n} d_i \phi_i \tag{4.61}$$

En substituant u à $\hat{u}$ dans (4.41), $\hat{u}$ n'étant pas à priori une solution exacte de l'équation, un résidu $r^{(n)}(x)$ subsistera, de sorte que l'on peut écrire :

$$\forall x \in D, \int_D C_{UU}(x, x')\hat{u}^{(n)}(x')dx' = \hat{\lambda}^{(n)}\hat{u}^{(n)}(x) + r^{(n)}(x) \tag{4.62}$$

La proximité de la solution approchée $(\hat{\lambda}^{(n)}, \hat{u}^{(n)})$ avec la vrai solution (λ, u) sera d'autant meilleure que le résidu $r^{(n)}$ sera proche de 0, et les méthodes qui suivent visent donc à rapprocher de 0 la valeur de celui-ci.

4.3.3.2.1 Méthodes de collocation *Les méthodes de collocation consistent à choisir un ensemble de points de collocation $(x_1, \ldots, x_m)$ de D, et à imposer $r^{(n)}(x_i) = 0$ pour l'ensemble des points. Cela revient donc à trouver les valeurs d_i telles que :*

$$\forall 1 \leq i \leq m, \sum_{j=1}^{n} d_j \left(\hat{\lambda}\phi_j(x_i) - \int_D C_{UU}(x_i, x')\phi_j(x')dx' \right) = 0 \tag{4.63}$$

Pour qu'un tel système possède une unique solution, une condition nécessaire et suffisante est que le nombre m de points de collocation soit égal au nombre n d'éléments dans la famille $(\phi_1, \ldots, \phi_n)$, et que de plus le déterminant de la matrice dont les coefficients sont $\phi_j(t_i)$ soit non nulle. La solution $\hat{u}^{(n)}$ ainsi définie sera la fonction de l'espace Φ_n vérifiant :

$$\forall 1 \leq i \leq n, \hat{u}^{(n)}(x_i) = u(x_i) \tag{4.64}$$

Cette fonction $\hat{u}^{(n)}$ peut être vu comme la projection de u sur Φ_n par l'opérateur de projection $P_c^{(n)}$ qui associe à chaque fonction continue u de $L_2(D)$ la fonction de Φ_n prenant les mêmes valeurs que u aux points x_i. Des résultats sur la convergence de la solution approchée $\hat{u}^{(n)}$ vers la vraie solution u sont présentés dans [10], la solution $\hat{u}^{(n)}$ convergeant vers u aussi rapidement que la projection $P_c^{(n)}(u)$ converge vers u. Le système linéaire (4.63) n'est cependant pas exploitable directement puisqu'il implique une intégrale qu'il convient d'éva-

luer. Pour évaluer cette intégrale, une formule de quadrature peut être utilisée, qui dans le cas général impliquera p poids w_k *et p points* x'_k *à priori différents des points d'interpolation* x_i*. Cette discrétisation des intégrales présentes dans* (4.63) *conduit à un nouveau système linéaire :*

$$\forall 1 \leq i \leq m, \sum_{j=1}^{n} d_j \left(\hat{\lambda}^{(n)} \phi_j(x_i) - \sum_{k=1}^{p} w_k C_{UU}(x_i, x'_k) \phi_j(x'_k) \right) = 0 \quad (4.65)$$

La solution de ce nouveau système linéaire est différente de la solution au système (4.63) *invoquant les intégrales exactes. La vitesse convergence de la solution approximée vers la vraie solution peut être impacté dans ce cas si la formule de quadrature utilisée n'est pas assez précise par rapport à la méthode de projection utilisée, et inversement [10].*

Tout comme pour le cas de la méthode de noyaux dégénérés, le système (4.65) *peut se mettre sous forme matriciel :*

$$C^{(n)} A^{(n)} D^{(n)} = \lambda^{(n)} B^{(n)} D^{(n)} \quad (4.66)$$

L'expression (4.66) *implique le vecteur colonne* $D^{(n)}$ *de taille n, ainsi que les matrices suivantes :*

- $C^{(n)}$ *de coefficients* $C_{UU}(x_i, x'_k)$ *de taille* $n \times p$
- $A^{(n)}$ *de coefficients* $w_k \phi_j(x'_k)$ *de taille* $p \times n$
- $B^{(n)}$ *de coefficients* $\phi_j(x_i)$ *de taille* $n \times n$

Le produit des matrices $C^{(n)}$ $A^{(n)}$ *donnant une matrice carré d'ordre n, le système matriciel à résoudre est donc un problème aux valeurs propres généralisé de taille n.*

4.3.3.2.2 Méthodes de Galerkin *Les méthodes de Galerkin sont elles aussi des méthodes projectives, et l'objectif est également de rendre le résidu* $r^{(n)}$ *présent dans* (4.62) *aussi proche de* 0 *que possible. Dans ces méthodes, on cherche à rendre le résidu* $r^{(n)}$ *orthogonale à l'espace* Φ_n*, de sorte que :*

$$\forall 1 \leq i \leq n, (r^{(n)}, \phi_i) = 0 \quad (4.67)$$

En combinant les relations (4.62) *et* (4.67)*, on trouve que la condition d'orthogonalité du résidu avec les éléments de* Φ_n *se traduit par le système linéaire suivant :*

$$\forall 1 \leq i \leq n, \sum_{j=1}^{n} d_j^{(n)} \int_D \left(\int_D C_{UU}(x, x') \phi_i(x) \phi_j(x') dx' - \hat{\lambda}^{(n)} \phi_i(x) \phi_j(x) \right) dx = 0$$

(4.68)

Les mêmes résultats de convergence que pour les méthodes de colocation existent pour les méthodes de Galerkin, à savoir que la vitesse de convergence de la solution $\hat{u}^{(n)}$ *de* (4.68) *vers la solution* u *de* (4.41) *sera aussi rapide que la convergence de la projection orthogonale de* u *sur* Φ_n *vers* u*. L'expression* (4.68) *n'est pas directement exploitable, car impliquant des intégrales qu'il convient de calculer. Suivant le choix des fonctions* ϕ_j*, l'intégrale impliquant le produit* $\phi_i\phi_j$ *peut être déterminé analytiquement, dans le cas de polynômes par exemple, mais il n'est généralement pas de même pour l'autre intégrale impliquant le noyau* C_{UU}*. Encore une fois, des méthodes de quadrature utilisant les poids* w_k *et les points* x_k *peuvent être utilisées pour approximer ces intégrales. Le système linéaire* (4.68) *après utilisation de ces méthodes de quadrature est alors changé en un nouveau système linéaire pouvant se mettre sous la forme matricielle suivante :*

$$\forall 1 \leq i \leq n, \left(C^{(n)} - \lambda^{(n)} B^{(n)}\right) D^{(n)} = 0 \tag{4.69}$$

Ce système matriciel implique le vecteur colonne $D^{(n)}$ *ainsi que les matrices suivantes :*

- $C^{(n)}$ *de coefficients* $\sum_{k=1}^{p}\sum_{l=1}^{p} w_k w_l C_{UU}(x_k, x_l)\phi_i(x_k)\phi_j(x_l)$ *de taille* $n \times n$
- $B^{(n)}$ *de coefficients* $\sum_{k=1}^{p} w_k \phi_i(x_k)\phi_j(x_k)$ *de taille* $n \times n$

Le système matricielle (4.69) *est différent du système linéaire* (4.68)*, et les solutions de ces deux systèmes sont donc à priori différentes. Néanmoins, comme montré dans [10], la solution du système* (4.69) *converge bien vers la vraie solution du système* (4.41)*.*

4.3.3.3 Méthode de Nyström

La méthode de Nyström consiste à résoudre directement l'équation (4.41) *en discrétisant l'intégrale à l'aide d'une méthode de quadrature afin d'obtenir la solution sur l'ensemble des points de quadrature, puis à obtenir la solution en dehors de ces points à l'aide d'une formule d'interpolation donnée par la méthode. En considérant une formule de quadrature utilisant* q_n *poids* $w_k^{(n)}$ *et* q_n *points* $x_k^{(n)}$*, on peut donc discrétiser l'intégrale de l'équation* (4.41) *pour obtenir la relation suivante pour la solution approchée* $u^{(n)}$ *à l'équation de Fredholm discrétisée :*

$$\forall x \in D, \sum_{k=1}^{q_n} w_k^{(n)} C_{UU}(x, x_k^{(n)}) u^{(n)}(x_k^{(n)}) = \lambda^{(n)} u^{(n)}(x) \tag{4.70}$$

La résolution de l'équation (4.70) *permet d'obtenir une solution approchée à notre problème initiale pour l'ensemble des valeurs x de D. La méthode de Nyström consiste à n'obtenir les valeurs de cette solution approchée que pour les valeurs $x_k^{(n)}$ correspondant aux points de quadrature, ce qui donne le système linéaire suivant :*

$$\forall 1 \leq l \leq q_n, \sum_{k=1}^{q_n} w_k C_{UU}(x_l^{(n)}, x_k^{(n)}) u^{(n)}(x_k^{(n)}) = \lambda^{(n)} u^{(n)}(x_l^{(n)}) \tag{4.71}$$

Dans le cas où l'on se place sur l'ensemble des fonctions continues muni de la norme infini, on peut montrer que la solution approchée $u^{(n)}$ obtenue par la résolution du système linéaire (4.71) *converge vers la solution u du système* (4.41) *pour cette norme infinie [10]. Une condition nécessaire et suffisante pour assurer cette convergence en norme infini dans l'espace des fonctions continues est que l'ensemble des poids $w_k^{(n)}$ vérifient la relation* (4.72) *[10].*

$$\lim_{n \to +\infty} \sum_{k=1}^{q_n} |w_k^{(n)}| < +\infty \tag{4.72}$$

Lorsque les poids considérés sont positifs, cette relation est vérifiée, comme on peut s'en convaincre pour l'intégration de la fonction constante et égale à 1 *sur D qui est borné. Des problèmes peuvent survenir uniquement si certains poids présents dans la formule de quadrature sont négatifs. L'utilisation de méthode de cubature à grille creuse comme par exemple la méthode de Smolyak font intervenir des poids négatifs, et possèdent de plus la particularité de ne pas vérifier systématiquement la condition* (4.72)*, la méthode de Smolyak en dimension d appliquée avec la méthode de quadrature de Clenshaw-Curtis vérifiant pour la somme de la valeur absolue de ses poids la relation* (4.73)*, l étant le niveau de la méthode de Smolyak [41].*

$$\sum |w| = O(\log(2^l l^{d-1})^{d-1}) \tag{4.73}$$

L'utilisation de grille creuse avec la méthode de Nyström, qui est faite dans le chapitre 8, ne garantit donc pas que la convergence de la solution aura lieu de manière uniforme, mais la convergence peut tout de même avoir lieu suivant d'autres normes, le formalisme permettant de s'intéresser à cette convergence étant détaillé dans [10]. On supposera donc dans la suite de cette section que l'ensemble des poids sont positifs.

La dernière étape de la méthode de Nyström consiste en une formule d'interpolation, nommée formule d'interpolation de Nyström, donnée par l'expression (4.74)*, permettant d'étendre la solution discrétisée obtenue à l'ensemble*

des points de D.

$$\forall x \in D, u^{(n)}(x) = \frac{1}{\lambda^{(n)}} \sum_{k=1}^{q_n} w_k^{(n)} C_{UU}(x, x_k^{(n)}) u^{(n)}(x_k^{(n)}) \tag{4.74}$$

Cette formule d'interpolation permet en fait d'obtenir la solution pour tout x de D de l'équation (4.70) comme on peut le vérifier, et est généralement considérée comme une bonne formule d'interpolation [10]. L'utilisation de cette formule d'interpolation plutôt qu'une formule d'interpolation plus classique comme une interpolation linéaire ou utilisant des splines permet d'obtenir une erreur sur l'ensemble des points de D du même ordre que celle obtenue sur les points $x_k^{(n)}$. L'utilisation d'une quadrature de Gauss avec une interpolation linéaire pour l'obtention de la solution ne serait pas meilleur que l'utilisation d'une méthode des trapèzes, l'interpolation linéaire ramenant l'erreur à un niveau similaire à celui obtenu avec la méthode des trapèzes. En pratique, l'utilisation de l'interpolation de Nyström suppose que l'on est capable d'évaluer l'auto-covariance en n'importe quel point x du domaine D, sans utilisation d'une interpolation qui encore une fois pénaliserait la solution. Cette formule d'interpolation n'est donc pas toujours utilisable. L'utilisation de la méthode permet d'obtenir une erreur en norme infinie sur la solution du même ordre que l'erreur obtenue sur le calcul numérique de l'intégrale [10], et est donc d'autant meilleure que la cubature utilisée est efficace, sous réserve de bon comportement des fonctions à intégrer.

Le système linéaire (4.71) peut à nouveau se mettre sous forme matricielle, comme fait dans l'expression (4.75), où $U^{(n)}$ est la matrice colonne de taille q_n contenant les valeurs $u^{(n)}(x_k^{(n)})$, $W^{(n)}$ est la matrice diagonale de taille $q_n \times q_n$ avec l'ensemble des poids $w_k^{(n)}$ sur sa diagonale et $C^{(n)}$ est la matrice de taille $q_n \times q_n$ dont les coefficients sont les valeurs $C_{UU}(x_k^{(n)}, x_l^{(n)})$.

$$C^{(n)} W^{(n)} U^{(n)} = \lambda^{(n)} U^{(n)} \tag{4.75}$$

Le problème aux valeurs propres (4.75) implique la matrice carré $C^{(n)}W^{(n)}$ de taille q_n et pourrait donc être résolu directement ainsi. Cependant, la matrice $C^{(n)}$ étant une matrice réelle symétrique, positive puisque c'est une matrice de covariance, et l'hypothèse faite sur la positivité des poids $w_k^{(n)}$ permettant de définir la matrice diagonale $W_{sqrt}^{(n)}$ dont la diagonale est l'ensemble des racines carrées des poids, il est possible de transformer le système matriciel (4.75) de la manière suivante :

$$\left(W_{sqrt}^{(n)} C^{(n)} W_{sqrt}^{(n)}\right) \left(W_{sqrt}^{(n)} U^{(n)})\right) = \lambda^{(n)} \left(W_{sqrt}^{(n)} U^{(n)}\right) \tag{4.76}$$

En notant $B^{(n)}$ *la matrice* $W^{(n)}_{sqrt}C^{(n)}W^{(n)}_{sqrt}$, *et* $V^{(n)}$ *le vecteur colonne* $W^{(n)}_{sqrt}U^{(n)}$, *l'expression* (4.76) *peut se réécrire :*

$$B^{(n)}V^{(n)} = \lambda^{(n)}V^{(n)} \tag{4.77}$$

La résolution numérique revient donc à résoudre un problème aux valeurs propres de taille q_n, *impliquant une matrice symétrique réelle qu'est la matrice* $B^{(n)}$. *Il est plus intéressant de résoudre ce problème aux valeurs propres impliquant une matrice réelle symétrique que le problème* (4.75) *impliquant une simple matrice carrée car les algorithmes de résolutions sont plus rapides dans le cas symétrique réel comme le montre la figure 4.2 qui compare le temps d'exécution nécessaire à la résolution d'un système aux valeurs propres impliquant une matrice symétrique réelle avec la routine DSYEVD de la librairie LAPACK et celle impliquant une simple matrice carré avec la routine DGEEV de la librairie LAPACK, en fonction de la taille des matrices.*

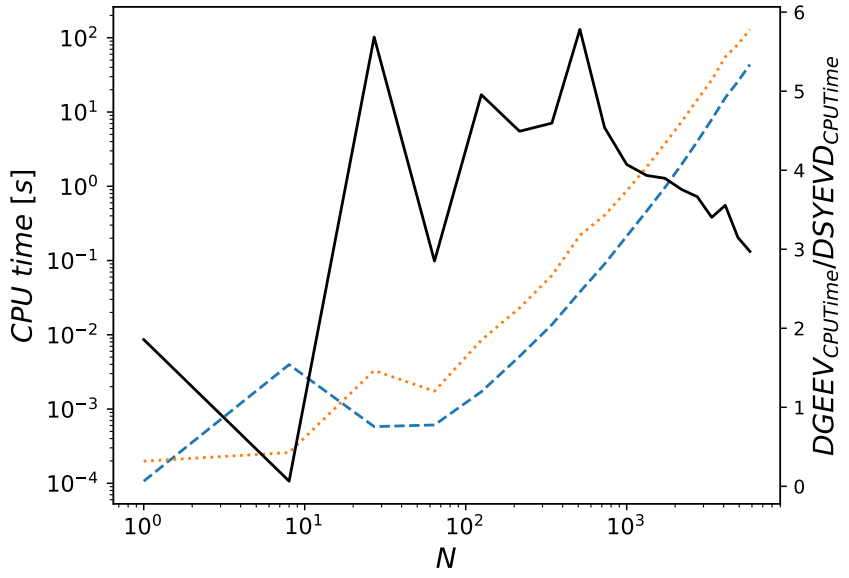

FIGURE 4.2 – *Temps CPU en secondes nécessaire à la résolution d'un système aux valeurs propres avec l'utilisation de la routine DGEEV (pointillés) et avec la routine DSYEVD (tirets) en fonction de la taille* N *du système, ainsi que rapport de ces deux temps (ligne pleine).*

La solution $U^{(n)}$ *à l'équation* (4.76) *peut alors être simplement retrouvée*

comme :

$$U^{(n)} = (W_{sqrt}^{(n)})^{-1} V^{(n)} \tag{4.78}$$

4.3.3.4 Choix de la méthode utilisée

L'étape finale permettant l'obtention de la solution approchée dans l'ensemble des méthodes présentées consiste en la résolution d'un système linéaire, dont la dimension est le nombre de degrés de liberté introduit pour cette résolution, qui sont :

- *Le nombre de termes dans la somme permettant la construction du noyau dégénéré dans le cas de la méthode à noyaux dégénérés*
- *La dimension de l'espace sur lequel projeter la solution, dans le cas des méthodes projectives*
- *Le nombre de points de quadratures où seront déterminées les valeurs de la solution dans le cas de la méthode de Nyström*

Une idée de la solution à priori peut permettre des bons choix de fonctions dans le cas des méthodes projectives ou la méthode à noyaux dégénérés, permettant de réduire significativement le nombre de degré de liberté, et donc la taille du système à résoudre. Un autre avantage de ces méthodes est qu'elles peuvent appréhender efficacement des comportements oscillants de la solution par un bon choix des fonctions de base. De nombreuses études sont faites dans l'optique de chercher des méthodes projectives impliquant de nouvelles fonctions et permettant de répondre à des spécificités de certains problèmes [20, 113]. Néanmoins, sans informations à priori sur la forme qu'aura la solution, la méthode de Nyström reste un choix particulièrement intéressant, de par sa simplicité d'implémentation et de par sa puissance, notamment en terme de vitesse de convergence, qui peut être très rapide via l'utilisation de méthodes de quadrature efficaces. Seule la méthode de Nyström sera donc utilisée dans la suite.

4.3.4 Etude numérique

Une étude numérique a été réalisée afin de valider l'implémentation de la méthode de Nyström et de retrouver certains résultats théoriques, notamment sur la convergence. Afin de réaliser cette validation, le processus gaussien d'Ornstein-Uhlenbeck [137] a été considéré, offrant l'avantage d'avoir une solution analytique à l'équation (4.41).

4.3.4.1 Processus d'Ornstein-Uhlenbeck

Le processus d'Ornstein-Uhlenbeck est un processus stochastique gaussien permettant la modélisation de la vitesse d'une particule en suspension dans un liquide visqueux [46]. On le considère ici défini sur un intervalle de temps $D = [0, 1]$ *et sur un espace probabilisé* $(\Theta, \mathbb{F}, P)$. *Le processus d'Ornstein-*

Uhlenbeck considéré ici est le processus d'Ornstein-Uhlenbeck stationnaire U_{OU}, de paramètres α et σ, défini par la relation suivante [46] :

$$U_{OU}(t,\omega) = e^{-\alpha t} N_{0,\sigma}(\omega) + \sigma\sqrt{2\alpha} \int_0^t e^{-\alpha(t-t')} dB(t',\omega) \tag{4.79}$$

Dans l'expression (4.79), $N_{0,\sigma}$ est une variable aléatoire gaussienne de moyenne nulle et d'écart-type σ, et B est le mouvement brownien unidimensionnel [25]. Le processus stochastique U_{OU} ainsi construit possède une moyenne nulle, c'est à dire :

$$\forall t \in [0,1], E\left[U_{OU}(t,.)\right] = 0 \tag{4.80}$$

De plus, sa fonction d'auto-covariance C_{UU} a pour expression :

$$\forall t,t' \in [0,1], C_{UU}(t,t') = \sigma^2 e^{-\alpha|t-t'|} \tag{4.81}$$

Le paramètre α présent dans la fonction d'auto-covariance peut être interprété comme l'inverse d'un temps caractéristique de corrélation, tandis que σ correspond à l'écart-type à tout instant du processus. Des réalisations du processus d'Ornstein-Uhlenbeck peuvent être obtenues à l'aide de l'équation (4.79), en discrétisant l'intégrale présente dans l'expression (4.79) suivant des pas de temps $0 = t_0 < t_1 < \cdots < t_n = 1$, de sorte que :

$$U_{OU}(t,\omega) \approx e^{-\alpha t} N_{0,\sigma}(\omega) + \sigma\sqrt{2\alpha} \sum_{i=0}^{n-1} e^{-\alpha(t-t_i)} \left(B(t_{i+1},\omega) - B(t_i,\omega)\right) \tag{4.82}$$

B étant un mouvement Brownien, pour tout i, $B(t_{i+1},.) - B(t_i,.)$ suit une loi normale centrée de variance $v = t_{i+1} - t_i$ indépendante des différences $B(t_{j+1},.) - B(t_j,.)$ pour $j < i$. Pour la discrétisation temporelle présente, $n+1$ nombres aléatoires indépendants doivent être générés pour obtenir une réalisation : 1 pour la variable gaussienne initiale, et n pour chacun des différences $B(t_{i+1},.) - B(t_i,.)$, faisant du problème discrétisé un problème de dimension stochastique $n+1$. Des réalisations du processus sont présentées sur la figure 4.3, pour une valeur des paramètres α et σ égale à 1.

La forme de la fonction d'auto-covariance permet la résolution analytique de l'équation (4.41) qui est réalisée dans [42], et qui donne les expressions suivantes pour les valeurs propres λ_i et les fonctions propres u_i associées. Les

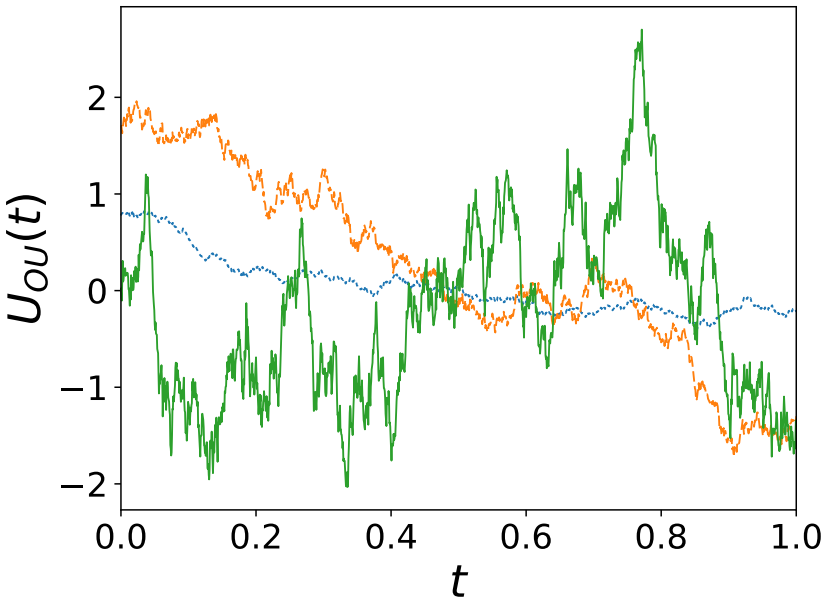

FIGURE 4.3 – *Réalisations du processus d'Ornstein-Uhlenbeck pour* $\sigma = 1$ *et* α *prenant les valeurs* 0.1, 1 *et* 10. α *étant l'inverse d'un temps de corrélation, la courbe la plus oscillante correspond à* $\alpha = 10$, *la moyennement oscillante à* $\alpha = 1.0$ *et la plus oscillante à* $\alpha = 0.1$, $n = 1000$ *intervalles de temps de même taille ayant été utilisés.*

valeurs propres sont données par l'expression (4.83).

$$\forall i \geq 1, \lambda_i = \sigma^2 \frac{2\alpha}{\alpha^2 + \beta_i^2} \tag{4.83}$$

Dans l'expression précédente, β_i *est la i-ème solution positive de l'équation :*

$$\left(\alpha - \beta \tan\left(\frac{\beta}{2}\right)\right)\left(\beta + \alpha \tan\left(\frac{\beta}{2}\right)\right) = 0 \tag{4.84}$$

Les fonctions propres quant à elle sont données, dans le cas où i est impair, par :

$$u_i(\omega) = \frac{cos\left(\beta_i\left(\omega - \frac{1}{2}\right)\right)}{\sqrt{\alpha + \frac{sin(2\beta_i\alpha)}{2\beta_i}}} \tag{4.85}$$

Et dans le cas où i est pair, par :

$$u_i(\omega) = \frac{sin\left(\beta_i\left(\omega - \frac{1}{2}\right)\right)}{\sqrt{\alpha - \frac{sin(2\beta_i\alpha)}{2\beta_i}}} \tag{4.86}$$

Les premières valeurs propres ainsi que les premières fonctions propres sont représentées sur la figure 4.4.

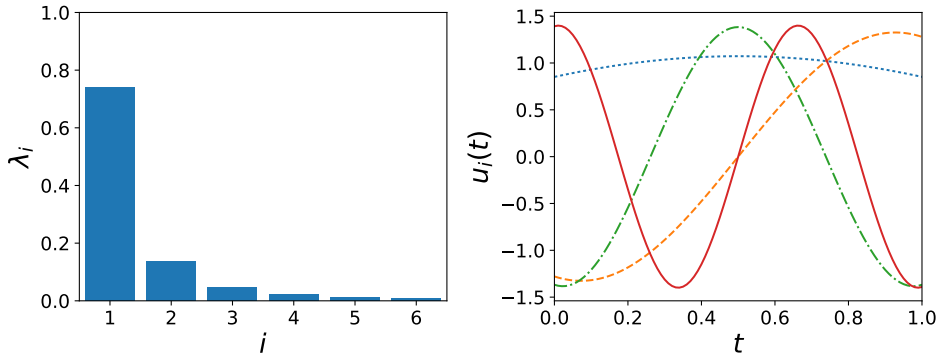

FIGURE 4.4 – *Gauche : 6 premières valeurs propres de l'expansion de Karhunen-Loève du processus d'Ornstein-Uhlenbeck pour $\sigma = 1$ et $\alpha = 1$. Droite : 4 premières fonctions propres de l'expansion de Karhunen-Loève du processus d'Ornstein-Uhlenbeck pour $\sigma = 1$ et $\alpha = 1$ (point : première, tiret : seconde, tiret-point ; troisième, ligne pleine : quatrième).*

Connaissant les valeurs exactes des valeurs et fonctions propres, il est maintenant possible d'étudier numériquement les comportements de convergence de la méthode de Nyström. Cette étude sera faite en deux parties. Dans un premier temps, l'étude se focalisera sur l'impact de la méthode de quadrature utilisée ainsi que de la méthode d'interpolation utilisée pour la reconstruction du signal, en utilisant l'expression analytique de la fonction d'auto-covariance. Dans un second temps, cette fonction d'auto-covariance sera évaluée numériquement à l'aide d'une méthode de Monte Carlo ou de Quasi-Monte Carlo utilisant des réalisations du processus d'Ornstein-Uhlenbeck.

4.3.4.2 Etude de l'impact de la méthode de quadrature utilisée

Pour étudier l'impact de la méthode de quadrature utilisée, les solutions obtenues à partir de la résolution de l'équation matricielle (4.77) pour différentes quadratures, et retransformée au moyen de l'expression (4.78) vont être comparée à la solution analytique, la matrice d'auto-covariance utilisée ayant été obtenue à l'aide de la formule analytique (4.81). La résolution numérique du système matriciel (4.77) est réalisée ici à l'aide de la routine DSYEVD de la librairie LAPACK [8], conçue pour résoudre des problèmes aux valeurs propres

impliquant une matrice symétrique réelle. Sur la figure 4.5 sont présentés le maximum des écarts absolus aux points de quadrature entre les quatre premiers modes propres analytiques et les quatre premiers modes propres numériques pour différentes méthodes de quadrature, en fonction du nombre n de points de quadratures impliqués. La vitesse de convergence observée est identique pour l'ensemble des méthodes de quadrature utilisées et des modes propres, comme le témoignent les pentes identiques de l'ensemble des courbes sur la figure 4.5.

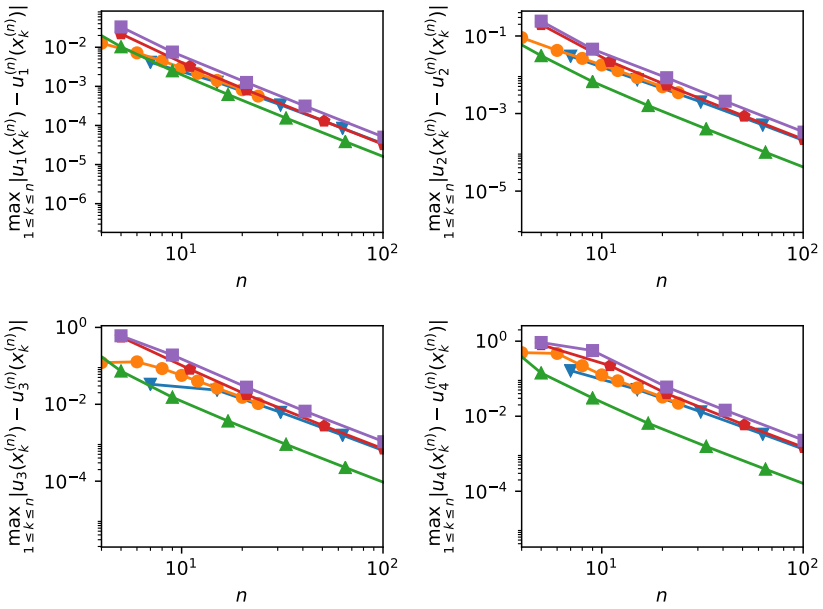

FIGURE 4.5 – *Maximum des écarts absolus sur l'ensemble des points de quadratures entre les modes propres analytiques et numériques, en fonction du nombre n de points de quadratures utilisés. Les méthodes de quadratures présentes sont : la méthode des trapèzes (triangles hauts), la méthode de Simpson (pentagones), la méthode de Newton-Cotes fermée composite d'ordre 4 (carrés), la seconde méthode de Fejér (triangles bas) et la méthode de Gauss-Legendre (ronds).*

La vitesse de convergence est en théorie la même que la convergence de l'approximation numérique de l'intégrale au premier membre de l'équation (4.41), qui se trouve dans ce cas précis être l'intégrale d'une fonction non régulière, la covariance du processus d'Ornstein-Uhlenbeck (4.81) n'étant pas différentiable sur la diagonale du carré unité. Cette non différentiabilité de la covariance limite la convergence des puissantes méthodes de quadratures telle la méthode de Gauss-Legendre.

Il est cependant possible d'exprimer la covariance en fonction des valeurs propres et des fonctions propres de l'expansion de Karhunen-Loève dans le cas d'un processus stochastique U de moyenne nulle, comme décrite dans l'énoncé

des propriétés de l'opérateur K, ce qui donne pour C_{UU} l'expression suivante :

$$C_{UU}(x,x') = \sum_{i=1}^{+\infty} \lambda_i u_i(x) u_i(x') \tag{4.87}$$

Connaissant l'expression analytique des valeurs propres λ_i et des fonctions propres u_i pour le processus d'Ornstein-Uhlenbeck, il est possible d'utiliser une version tronquée de l'auto-covariance, ne possédant pas de problèmes de non différentiabilité sur la diagonale du carré, en ne conservant que les N premiers termes de l'expression (4.87), ce qui donne la nouvelle fonction d'auto-covariance suivante :

$$C_{UU}^{(N)}(x,x') = \sum_{i=1}^{N} \lambda_i u_i(x) u_i(x') \tag{4.88}$$

Cette nouvelle fonction d'auto-covariance $C_{UU}^{(N)}$ ne correspond plus à celle du processus stochastique U, mais seuls les valeurs propres et modes propres nous intéressent ici, et il est immédiat que les N premières valeurs propres et fonctions propres obtenues en résolvant l'équation (4.41) en utilisant la covariance complète donnée par (4.87) ou tronquée à N termes donnée par (4.88) seront les mêmes. Sur la figure 4.6 sont représentés les mêmes courbes que sur la figure 4.5, mais cette fois-ci en utilisant la covariance tronquée à 4 termes, n'étant intéressé que par les quatre premières fonctions propres.

La quantité sous l'intégrale dans le terme de gauche de l'équation (4.41) est maintenant indéfiniment différentiable, assurant un bon comportement pour l'ensemble des méthodes de quadratures utilisées. On observe bien sur les courbes les pentes de -1, -2 et -4 associées à la convergence des méthodes des trapèzes, et des méthodes de Newton-Cotes fermées d'ordre 2 et 4 respectivement. De plus, on observe la convergence forte des méthodes de Fejér et de Gauss-Legendre, ne nécessitant que quelques dizaines de points de quadratures pour atteindre une erreur de l'ordre de la précision machine.

Les résultats de la figure 4.5 et 4.6 ne concernent que les erreurs aux points de quadratures, et non pas sur l'ensemble du domaine de définition des modes. Pour obtenir la solution sur l'ensemble du domaine, il est nécessaire d'utiliser une méthode d'interpolation. Sur la figure 4.7 sont comparées trois méthodes d'interpolation, permettant la reconstruction des modes sur l'ensemble du domaine, lorsque la méthode de quadrature utilisée est la méthode de Fejér, et la covariance tronquée est utilisée. La norme infinie considérée est obtenue en pratique en considérant 10000 points réparties uniformément sur le segment $[0,1]$ et en prenant le maximum de l'écart en valeur absolue entre le mode obtenu analytiquement et le mode obtenu par une interpolation.

La figure 4.7 montre que la méthode d'interpolation peut "dégrader" l'er-

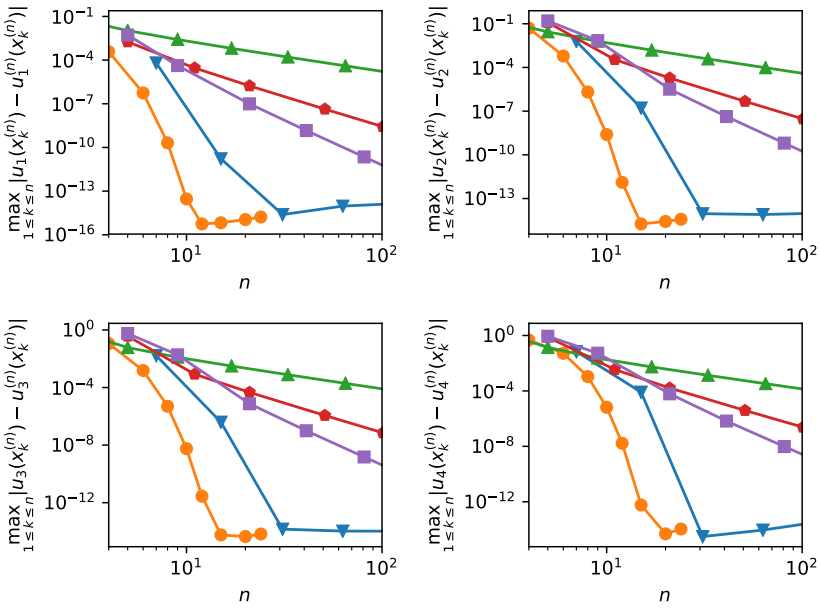

FIGURE 4.6 – *Maximum des écarts absolus sur l'ensemble des points de quadratures entre les modes propres analytiques et numériques, en fonction du nombre n de points de quadratures utilisés, avec utilisation d'une covariance tronquée à 4 termes. Les méthodes de quadratures présentes sont : la méthode des trapèzes (triangles hauts), la méthode de Simpson 2 (pentagones), la méthode de Newton-Cotes fermée composite d'ordre 4 (carrés), la seconde méthode de Fejér (triangles bas) et la méthode de Gauss-Legendre (ronds).*

reur faite par la méthode de quadrature. En effet, dans le cas de l'interpolation linéaire, la pente de la courbe pour l'ensemble des fonctions propres est -1, ce qui correspond à l'erreur commise par une interpolation linéaire d'une fonction régulière. Pour les splines cubiques naturelles, cette pente est de -3 qui correspond encore une fois à l'erreur d'interpolation pour des splines cubiques. En revanche, la méthode d'interpolation de Nyström conserve l'erreur qui avait été obtenue par la méthode de Fejér, et en fait donc en théorie le candidat d'interpolation idéal. Néanmoins, cette dernière formule d'interpolation nécessite d'être capable d'estimer exactement la covariance en tout point du domaine, ce qui n'est pas toujours nécessairement le cas en pratique, et des méthodes d'interpolation plus classiques telles que les deux autres présentées doivent alors être utilisées.

Cette dernière remarque soulève également la question de l'impact de l'estimation de la covariance, nécessaire à la résolution de l'équation de Fredholm (4.41). En pratique, celle-ci nécessite parfois d'être estimée à l'aide d'une méthode statistique telle la méthode de Monte Carlo ou de Quasi-Monte Carlo randomisée, au travers du calcul de trajectoires du processus stochastique étu-

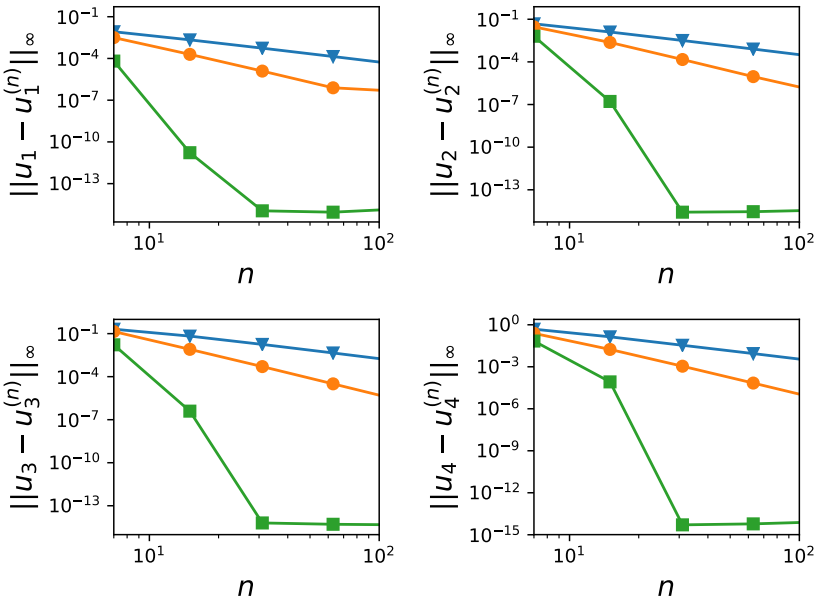

FIGURE 4.7 – *Normes infinies de la différence entre les modes propres analytiques et numériques, en fonction du nombre n de points de quadrature utilisés, avec utilisation d'une covariance tronquée à 4 termes et différentes méthodes d'interpolations. Les méthodes d'interpolation présentes sont : interpolation linéaire (triangles), spline cubique naturelle (ronds) et interpolation de Nyström (carrés).*

dié, et cette estimation va elle aussi avoir un impact sur les fonctions propres et valeurs propres de la décomposition de Karhunen-Loève.

4.3.4.3 Etude de l'impact de l'estimation de la fonction d'autocovariance

L'étude précédente présupposait la connaissance parfaite de la fonction d'auto-covariance. En pratique, la fonction d'auto-covariance est inconnue, et il est nécessaire de l'estimer. Dans les cas qui seront traités dans les chapitres suivants, cette estimation se fera à partir d'une méthode de Monte Carlo ou de Quasi-Monte Carlo randomisé, au cours de laquelle des réalisations du processus stochastique d'intérêt auront été calculées. Concernant le processus d'Ornstein-Uhlenbeck stationnaire, de tels réalisations peuvent être obtenues par une discrétisation de l'intégrale stochastique comme exposé au travers de la formule (4.82)*. L'objectif de cette section est l'étude de l'impact de l'estimation de la matrice de covariance sur la solution à l'équation* (4.41)*. Le calcul des réalisations du processus d'Ornstein-Uhlenbeck au moyen d'une discrétisation comme dans l'expression* (4.82) *induit une erreur sur le calcul de ces réalisations qu'il sera difficile voir impossible de distinguer de l'erreur d'estimation dû à la méthode de Monte-Carlo ou de Quasi-Monte Carlo. Cette difficulté peut dans le cas*

du processus d'Ornstein-Uhlenbeck stationnaire être contournée, puisque ce dernier peut également s'exprimer uniquement à l'aide du mouvement Brownien, au travers de l'expression (4.89)*, comme montré dans [98].*

$$\forall t \in [0,1], U_{OU}(t) = \sigma e^{-\alpha t} B(e^{2\alpha t}) \tag{4.89}$$

Il est facile de vérifier que le processus ainsi défini est bien de moyenne nulle, de fonction d'auto-covariance $e^{-\alpha|t-t'|}$*, et que c'est également un processus gaussien tout comme le processus d'Ornstein-Uhlenbeck, ce qui en fait le même par caractérisation des processus gaussiens par leurs propriétés d'ordre* 2 *[17]. En utilisant la forme définie par* (4.89)*, il est possible de déterminer facilement la valeur du processus aux points d'intérêts, qui sont ici les points de quadratures* t_k*, puisqu'on a alors :*

$$\begin{cases} B(e^{2\alpha t_1}) \sim \mathcal{N}(0, e^{2\alpha t_1}) \\ \forall 2 \leq k \leq q_n, B(e^{2\alpha t_k}) - B(e^{2\alpha t_{k-1}}) \sim \mathcal{N}(0, e^{2\alpha t_k} - e^{2\alpha t_{k-1}}) \end{cases} \tag{4.90}$$

La génération des valeurs du processus d'Ornstein-Uhlenbeck aux n *points de quadrature ne nécessite que la génération de* n *nombres aléatoires. En plus d'être exacte, elle permet donc de réduire la dimension stochastique du problème étudié ici au strict minimum, ce qui est préférable pour l'utilisation de méthodes de Quasi-Monte Carlo qui souffrent d'une dégradation de leurs performances avec l'augmentation de la dimension stochastique.*

L'estimation de la matrice d'auto-covariance est effectuée de manière différente selon que l'on utilise une méthode de Monte Carlo ou de Quasi-Monte Carlo randomisé, afin de s'assurer d'avoir dans les deux cas une estimation sans biais. Dans le cas de la méthode de Monte-Carlo, on considère que l'on a N *réalisations indépendantes* $(u_{OU}^{(i)})_{1 \leq i \leq N}$ *du processus d'Ornstein-Uhlenbeck, à partir desquelles est estimée la covariance entre* $U_{OU}(t_k)$ *et* $U_{OU}(t_l)$ *de la manière suivante :*

$$\begin{aligned} Cov_{MC}\left[U_{OU}(t_k), U_{OU}(t_l)\right] \approx & \frac{1}{N-1} \sum_{i=1}^{N} u_{OU}^{(i)}(t_k) u_{OU}^{(i)}(t_l) \\ & - \frac{N}{N-1} \left(\frac{1}{N} \sum_{i=1}^{N} u_{OU}^{(i)}(t_k) \right) \left(\frac{1}{N} \sum_{i=1}^{N} u_{OU}^{(i)}(t_l) \right) \end{aligned} \tag{4.91}$$

Dans le cas de la méthode de Quasi-Monte Carlo randomisé, on considère que l'on a Q *séquences à discrépance faible randomisées et indépendantes entre elles, comportant chacune* M *points et permettant donc de générer un total de*

$N = MQ$ réalisations $(u_{OU}^{(m,q)})_{1\leq m\leq M, 1\leq q\leq Q}$ du processus d'Ornstein-Uhlenbeck, à partir desquelles est estimée la covariance entre $U_{OU}(t_k)$ et $U_{OU}(t_l)$ de la manière suivante :

$$Cov_{QMCR}\left[U_{OU}(t_k), U_{OU}(t_l)\right] \approx \frac{1}{MQ}\sum_{q=1}^{Q}\sum_{m=1}^{M} u_{OU}^{(m,q)}(t_k)u_{OU}^{(m,q)}(t_l) - \frac{1}{Q(Q-1)}\sum_{\substack{1\leq q,q'\leq Q \\ q\neq q'}}\left(\frac{1}{M}\sum_{m=1}^{M} u_{OU}^{(m,q)}(t_k)\right)\left(\frac{1}{M}\sum_{m=1}^{M} u_{OU}^{(m,q')}(t_l)\right) \tag{4.92}$$

Afin de dissocier l'erreur liée à la méthode de quadrature de celle liée à l'estimation de la matrice de covariance, on fixe pour la suite une méthode de quadrature et un nombre n de points de quadratures, en l'occurrence la seconde méthode de Fejér à 63 points. La solution de référence est donc constituée des vecteurs propres u_k de l'expansion de Karhunen-Loève qui sont les solutions de l'équation (4.77) et transformées par l'équation (4.78) lorsque la matrice d'auto-covariance est obtenue analytiquement. Les solutions numériques sont quant à elles constituées des vecteurs propres $\hat{u}_k$ de l'expansion de Karhunen-Loève obtenus avec une matrice d'auto-covariance estimée statistiquement à l'aide d'une méthode de Monte Carlo ou de Quasi-Monte Carlo randomisé. La comparaison des deux méthodes se fait grâce à l'erreur quadratique moyenne $MSE_k(t_l)$ du k-ième vecteur propre au point t_l, à partir de l'expression (4.93), où $\hat{U}_k$ est une solution numérique obtenu à l'aide d'une méthode statistique impliquant N échantillons.

$$MSE_k(t_l) = E\left[\left(\hat{U}_k(t_l) - u_k(t_l)\right)^2\right] \tag{4.93}$$

Afin d'estimer l'erreur quadratique moyenne de l'expression (4.93), 100 simulations de Monte Carlo indépendantes ainsi que 100 simulations de Quasi-Monte Carlo randomisé indépendantes ont été utilisées. Les résultats de cette étude sont montrés sur la figure 4.8 où sont présentées les normes infinies des erreurs quadratiques moyennes pour les 4 premiers vecteurs propres, les résultats de Quasi-Monte Carlo randomisé impliquant 10 séquences de Sobol randomisés avec une méthode de Full-Scrambling. La comparaison est effectuée à nombre de réalisations total égal.

La méthode de Monte Carlo permet d'obtenir une convergence de l'ordre de $1/N$ pour la norme infinie de l'erreur quadratique de l'ensemble des vecteurs propres, qui est la convergence théorique de cette méthode. La méthode de Quasi-Monte Carlo randomisé permet d'obtenir une convergence plus rapide

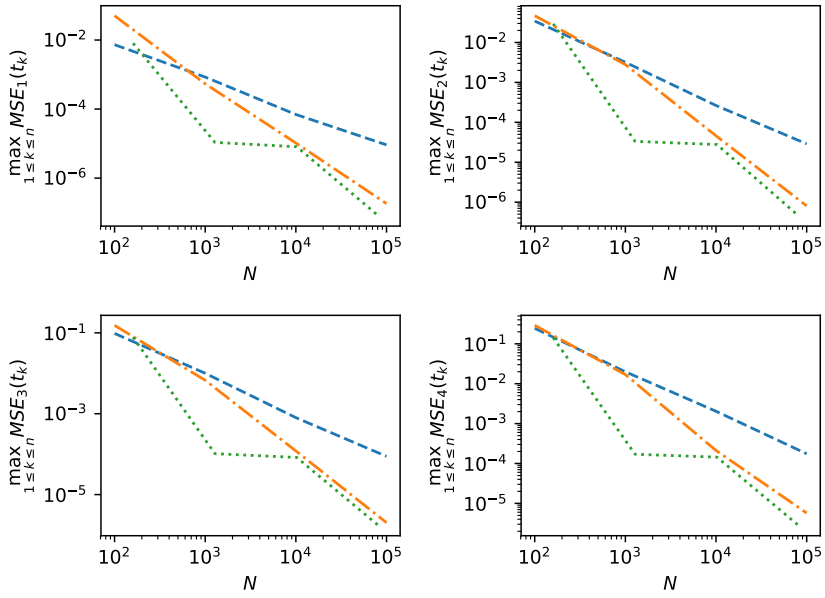

FIGURE 4.8 – *Normes infinies de l'erreur quadratique moyenne des* 4 *premiers vecteurs propres en fonction du nombre total* N *de réalisations pour la méthode de Monte-Carlo (tirets), la méthode de Quasi-Monte Carlo randomisé avec un nombre de points par séquence de Sobol qui n'est pas une puissance de* 2 *(tirets-points) et la méthode de Quasi-Monte Carlo randomisé avec un nombre de points par séquence de Sobol qui est une puissance de* 2 *(pointillés).*

que la méthode de Monte Carlo, et plus rapide encore si le nombre de points par séquence est un multiple de 2 *comme prévu théoriquement pour une randomisation utilisant la méthode de Full-Scrambling. Un autre point, touchant la convergence des deux méthodes, est que la convergence absolue est meilleure pour les premiers vecteurs propres que pour les suivants, fait qui se retrouvait également pour les méthodes de quadratures.*

4.3.4.4 Enseignements de l'étude numérique

L'utilisation de la méthode de Nyström pour le calcul numérique de l'expansion de Karhunen-Loève implique différentes étapes, chacune nécessitant le choix d'une méthode, introduisant des erreurs sur le résultat, qui sont les suivantes :

- *l'erreur due à la discrétisation de l'équation de Fredholm, nécessitant le choix d'une méthode de cubature*
- *l'erreur d'interpolation pour la reconstruction des fonctions propres*
- *l'erreur due à l'estimation de la matrice d'auto-covariance*

La première erreur peut être relativement faible avec l'utilisation de méthodes de cubature puissantes, à la condition que la fonction d'auto-covariance

soit suffisamment régulière. L'augmentation du nombre de points de cubature permet alors de diminuer cette erreur. Afin de ne pas dégrader l'erreur obtenue grâce à une méthode de cubature, il est nécessaire que la méthode d'interpolation permette une erreur du même ordre de grandeur. En pratique, on se limitera à l'utilisation de spline cubique naturelle, la méthode d'interpolation de la méthode de Nyström étant inaccessible, ce qui permet une erreur d'interpolation plus faible qu'avec une simple interpolation linéaire. Enfin, l'erreur d'estimation de la matrice d'auto-covariance introduit une erreur dans l'estimation des valeurs et modes propres, qu'il est nécessaire de limiter. Les derniers résultats montrent un avantage clair des méthodes de Quasi-Monte Carlo randomisées sur la méthode de Monte Carlo pour l'estimation de la matrice d'auto-covariance, et motive leur utilisation dans le reste de cette thèse.

L'estimation de la matrice d'auto-covariance nécessite le calcul de réalisations du processus stochastique étudié. Ces réalisations du processus stochastique sont utiles en pratique au calcul de réalisations des variables aléatoires η_k introduites par l'expansion de Karhunen-Loève dont l'expression est donnée par (4.46). Il est nécessaire de discrétiser l'expression (4.46) pour obtenir ces réalisations. Étant donné une trajectoire $u^{(j)}$ du processus stochastique U étudié, les réalisations des variables aléatoires $\eta_k^{(j)}$ associé à cette trajectoire de U sont obtenues comme :

$$\eta_k^{(j)} = \frac{1}{\sqrt{\lambda_k^{(n)}}} \sum_{i=1}^{q_n} \omega_i u_k^{(n)}(x_i) u^{(j)}(x_i) \tag{4.94}$$

En pratique, les trajectoires $u^{(j)}$ du processus stochastique sont obtenues à l'aide d'une méthode de Monte Carlo ou de Quasi-Monte Carlo randomisée. Dans le cas de l'utilisation d'une méthode de Quasi-Monte Carlo randomisée, les échantillons des nouvelles variables aléatoires $\eta_k^{(j)}$ obtenues à l'aide de l'expression (4.94) qui est la discrétisation de l'expression (4.94) ne seront pas indépendants, et hériteront de dépendance mutuelle inhérente à la méthode de Quasi-Monte Carlo randomisé.

4.4 Conclusion

Dans ce chapitre, différentes méthodes ont été présentées permettant d'approximer un processus stochastique en le représentant comme dépendant d'un nombre donné et réduit de variables aléatoires. Les méthodes proposées diffèrent notamment par les variables aléatoires impliquées dans la reproduction du processus stochastique.

La première approche ne construit pas de nouvelles variables aléatoires, et consiste à remplacer le processus stochastique par son expansion en polynômes du Chaos en un jeu de variables aléatoires déjà connues. En particulier, il est

possible d'utiliser les paramètres incertains ayant servi à paramétrer l'incertitude du système. Cette dernière utilisation des polynômes du chaos est très utilisée pour la propagation d'incertitudes dans la littérature. A travers cette expansion en polynômes du Chaos, il est également possible d'avoir accès à des informations relatives à la sensibilité du processus stochastique en certains des paramètres initiaux, permettant de donner des informations susceptibles de réduire le nombre de paramètres incertains initiaux, simplifiant de fait le système étudié.

La seconde approche consiste à représenter le processus stochastique à l'aide d'une troncature de son expansion de Karhunen-Loève. Cette dernière permet de construire un jeu de nouvelles variables aléatoires, qui ont la particularité d'être rangées par ordre d'importance quant à leur impact sur la variance du processus stochastique. Ces nouvelles variables aléatoires présentent la particularité d'être décorrélées deux à deux, mais elles ne sont en général pas indépendantes. Des dépendances importantes entre elles peuvent obliger à une modélisation de celles-ci, afin de pouvoir avoir une représentation correcte du processus stochastique d'intérêt. Dans la présente thèse, de telles dépendances "problématiques" ont été observées, ce qui a motivé l'étude des méthodes présentées dans le chapitre suivant, permettant de modéliser la loi de probabilité jointe d'un ensemble de variables aléatoires dépendantes dont on possède des réalisations.

Les résultats de ce chapitre ont permis de réaliser une communication au sein de l'ICMF présentée en annexe C. Le travail présenté consistait en la construction d'une surface de réponse de la vitesse axiale en fonction du diamètre de goutte au sein d'une simulation monodisperse d'un écoulement diphasique à l'aide d'une expansion en polynôme du chaos. La construction de cette expansion en polynôme du chaos a été réalisée grâce à une méthode non-intrusive projective utilisant une méthode de quadrature de Clenshaw-Curtis, ayant nécessité 9 *simulations aux grandes échelles. Les résultats ont permis d'obtenir des informations sur le diamètre de goutte optimal à utiliser dans le cas de simulations monodisperses.*

Chapitre 5

Estimation de densité de probabilité et génération de vecteurs aléatoires à composantes dépendantes

Prérequis :

- *Méthode d'inversion généralisée pour la génération de vecteurs aléatoires*

Notions clés et apports du chapitre :

- *Histogrammes pour l'estimation de densité de variables et vecteurs aléatoires*
- *Méthodes à noyaux pour l'estimation de densité de variables et vecteurs aléatoires*
- *Présentation des leviers possibles pour l'amélioration des méthodes à noyaux*
- *Utilisation de la méthode d'inversion généralisée dans le cadre de l'utilisation de méthodes à noyaux gaussiens*
- *Comparaison de variantes de méthodes à noyaux gaussiens*

La décomposition d'un processus stochastique à l'aide de sa décomposition de Karhunen-Loève introduit de nouvelles variables aléatoires, qu'il convient de modéliser. Dans le cas de processus gaussiens, ces nouvelles variables aléatoires présentent l'avantage de toutes être gaussiennes et indépendantes [65]. Dans le cas où le processus stochastique étudié s'éloigne significativement d'un processus gaussien, il est nécessaire de modéliser suffisamment finement le vecteur aléatoire composé des nouvelles variables aléatoires retenues, et notamment les dépendances présentes entre les différentes composantes de ce vecteur aléatoire. La modélisation s'entend ici comme l'attribution d'une loi au vecteur aléatoire à partir de réalisations de celui-ci, loi que l'on espère estimer correctement tandis que la vraie loi de celui-ci est inconnue. Deux grandes familles de méthodes existent afin de modéliser la loi d'un vecteur aléatoire : les méthodes paramétriques et les méthodes non-paramétriques. Dans le cas des premières, une hypothèse est faite sur le type de loi suivie par le vecteur aléatoire à modéliser, introduisant des paramètres, qu'il convient ensuite d'estimer à l'aide des échantillons du vecteur aléatoire. Le nombre de paramètres à estimer est fixé à priori suivant la complexité des hypothèses utilisées, et ne varie pas avec la taille de l'échantillon utilisé pour estimer ces paramètres. Les secondes méthodes, dites non-paramétriques ne font pas d'hypothèses à priori sur la loi du vecteur aléatoire, et n'introduisent donc pas de paramètres à estimer devant caractériser cette loi. En ce sens, elles sont plus flexibles mais payent cette flexibilité au prix d'une convergence plus faible [117]. Le terme non-paramétrique est légèrement trompeur, des paramètres pouvant être utilisés, leur nombre variant cependant avec le nombre d'échantillons utilisé pour l'estimation. Dans un souci de flexibilité, ne connaissant pas a priori les lois des vecteurs aléatoires rencontrés dans cette thèse, une méthode non-paramétrique, l'estimation par noyau [117], est présentée dans ce chapitre, visant à estimer la densité du vecteur aléatoire à modéliser. Ce chapitre ne se veut pas exhaustif, des livres entiers étant consacrés à cette méthode, mais présente plus l'aspect pratique de ces méthodes pour l'utilisation faite dans cette thèse.

5.1 Histogrammes

5.1.1 Présentation de la notion d'histogramme

La visualisation de la distribution de réalisations $(x_i)_{1\leq i\leq N}$ d'une suite de variables aléatoires réelles $(X_i)_{1\leq i\leq N}$ de même loi de probabilité se fait classiquement à l'aide d'un histogramme. Afin de construire un histogramme, des sous intervalles $([t_k, t_{k+1}[)_{k\in\mathbb{Z}}$ que l'on considérera tous de longueur h dans la suite sont définis. L'histogramme correspond alors à la fonction constante par

morceaux $\hat{h}$ définie par la relation suivante :

$$\forall k \in \mathbb{Z}, \forall x \in [t_k, t_{k+1}[, \hat{h}(x) = \frac{1}{Nh} \sum_{i=1}^{N} \mathbb{1}_{[t_k, t_{k+1}[}(x_i) \tag{5.1}$$

Un tel histogramme peut se caractériser à l'aide de deux paramètres, que sont la longueur h des intervalles ainsi que la valeur t_0, la modification de cette dernière permettant de translater l'ensemble de ces intervalles. Sur la figure 5.1 sont présentés plusieurs histogrammes d'un mélange de deux variables aléatoires gaussiennes construits à l'aide de $N = 10000$ réalisations, avec différentes valeurs de h et une valeur de $t_0 = 0$.

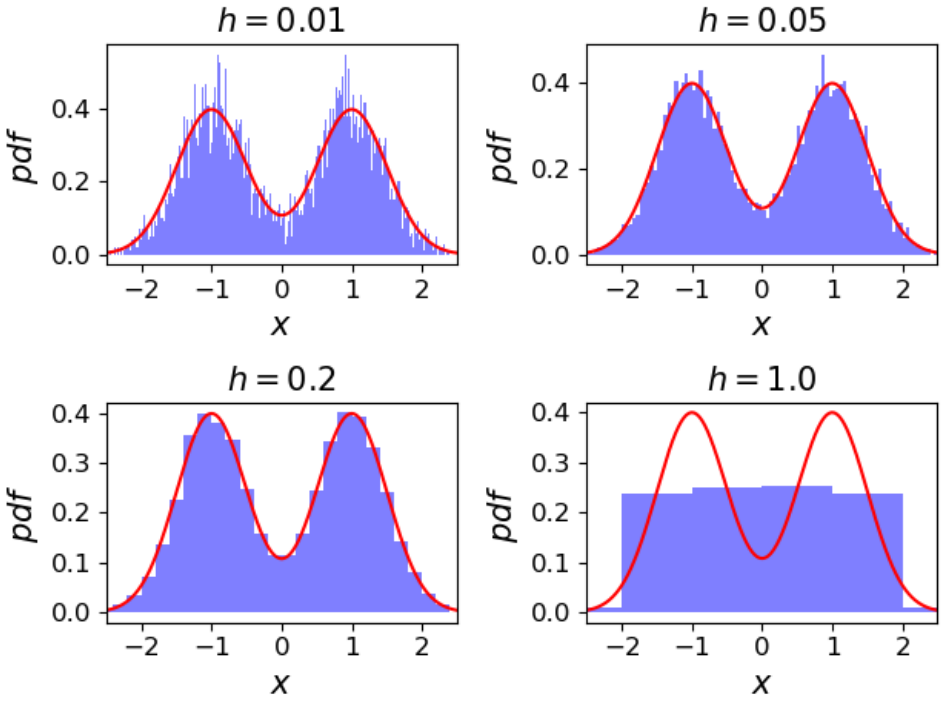

FIGURE 5.1 – *Histogrammes d'un mélange de deux variables aléatoires gaussiennes obtenues avec $N = 10000$ réalisations pour différentes valeurs du paramètre h et une valeur de $t_0 = 0$, la courbe rouge correspond à la densité de probabilité de ce mélange de gaussiennes.*

La figure 5.1 montre qu'un histogramme permet d'approcher la densité de probabilité à l'aide d'une fonction constante par morceaux. Le choix du paramètre h définissant la largeur des intervalles $[x_k, x_{k+1}[$ impacte directement l'allure de l'histogramme, et sa capacité à approcher la vraie densité de probabilité. Une valeur trop importante de celui-ci empêche l'histogramme de s'approcher de cette densité de probabilité de par la nature constante par morceaux de celui-ci,

alors qu'une valeur trop faible implique un histogramme "bruité" du fait que le nombre N_k d'échantillons par intervalle, qui est une variable aléatoire, fluctue significativement. Un bon choix de h dépend en fait du nombre d'échantillons N, et il est possible de choisir le paramètre h en fonction de N en imposant que l'écart entre l'histogramme et la densité de probabilité soit minimal en un certain sens [117].

Dans le cas $h = 1$, on peut voir sur la figure 5.1 que l'histogramme n'est pas en mesure de capturer le double pic présent dans la densité de probabilité. Il est cependant possible de capturer ce double pic avec cette valeur du paramètre h en jouant sur le second paramètre qu'est t_0, comme visible sur la figure 5.2.

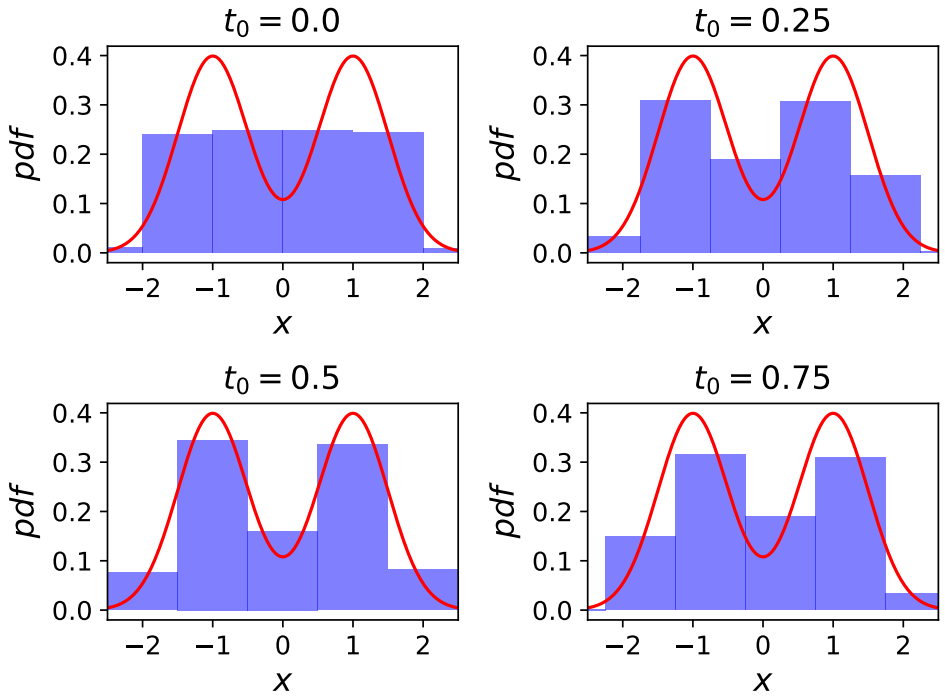

FIGURE 5.2 – *Histogrammes d'un mélange de deux variables aléatoires gaussiennes obtenues avec $N = 10000$ réalisations pour différentes valeurs du paramètre t_0 et une valeur de $h = 1$, la courbe rouge correspond à la densité de probabilité de ce mélange de gaussiennes.*

Le choix de t_0 peut donc jouer un rôle dans la capacité de l'histogramme à correctement approcher la densité de probabilité. Il est cependant possible de s'affranchir de cette dépendance en t_0 de l'histogramme, retirant de fait un degré de liberté à l'estimation par histogramme.

5.1.2 Histogrammes décalés moyennés

La figure 5.2 montre l'impact du paramètre t_0 sur la capacité de l'histogramme à reproduire ou non les pics de la densité de probabilité par exemple. Afin de ne plus dépendre de cette dépendance au paramètre t_0, certains auteurs ont proposé [117] de moyenner les histogrammes obtenus avec plusieurs valeurs de t_0. Pour cela, m histogrammes $(\hat{h}_j)_{1\leq j\leq m}$ sont construits, avec pour l'histogramme j une valeur de $t_0 = jh/m$, permettant de construire le nouvel estimateur $\hat{h}^{(m)}$ pour la densité de probabilité comme :

$$\hat{h}^{(m)}(x) = \frac{1}{m}\sum_{j=1}^{m}\hat{h}_j(x) \tag{5.2}$$

Une telle construction donne encore un histogramme pour l'approximation de la densité de probabilité, la fonction $\hat{h}^{(m)}$ étant également constante par morceaux sur les sous-intervalles de la forme $([t_k+jh/m, t_k+(j+1)h/m[)_{1\leq j<m, k\in\mathbb{Z}}$. De tels histogrammes sont visibles sur la figure 5.3, et ne peuvent être obtenus à l'aide de la définition classique donnée par la formule (5.1). L'augmentation de la valeur du paramètre m tend à réduire la base des rectangles présents dans l'histogramme, et cela sans augmenter le "bruit" de l'histogramme comme c'était le cas avec la définition précédente des histogrammes. En fait, le "bruit" présent dans les histogrammes ne dépend que du paramètre h, qui correspondait justement à la base des rectangles dans la définition des histogrammes données par (5.1). Le paramètre m permet donc de diminuer la taille de la base des rectangles de l'histogramme, sans introduire de bruit supplémentaire. Cependant, l'histogramme ainsi construit bien que possédant peu de "bruit" est incapable de reproduire correctement les pics comme visible sur la figure 5.3 où les pics sont sous estimés et trop large. Il est nécessaire de jouer sur le paramètre h pour capturer mieux ces pics.

Le paramètre m permet néanmoins de lisser l'histogramme, et il est montré dans [117] qu'en faisant tendre m vers l'infini dans l'expression (5.2), l'estimateur $\hat{h}^{(\infty)}$ obtenu a pour expression l'expression suivante :

$$\hat{h}^{(\infty)}(x) = \frac{1}{Nh}\sum_{i=1}^{N}\left(1-\frac{|x-x_i|}{h}\right)\mathbb{1}_{[-1,1]}\left(\frac{x-x_i}{h}\right) \tag{5.3}$$

La courbe en ligne pleine sur la figure 5.3 correspond à l'estimateur $\hat{h}^{(\infty)}$, où l'on peut observer que les histogrammes $\hat{h}^{(m)}$ tendent bien vers l'estimateur $\hat{h}^{(\infty)}$ lorsque m tend vers l'infini.

L'estimateur $\hat{h}^{(\infty)}$ peut en fait s'écrire sous la forme plus générale suivante,

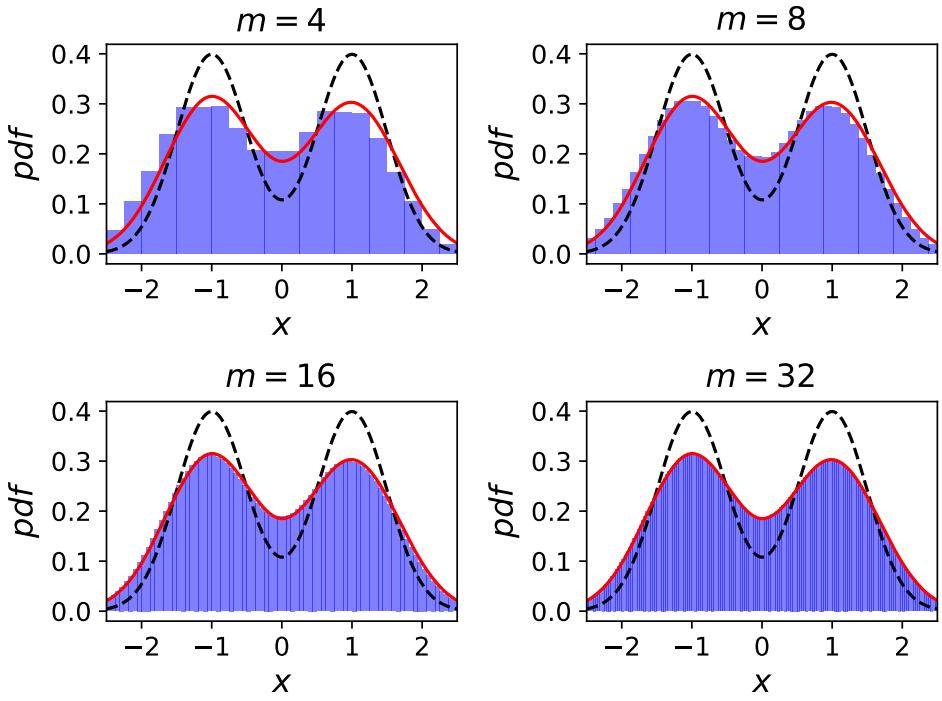

FIGURE 5.3 – *Moyennes de m histogrammes obtenues par l'expression (5.2) d'un mélange de deux variables aléatoires gaussiennes obtenues avec $N = 10,000$ réalisations pour une valeur de $h = 1$, la courbe en trait plein correspond à la limite des histogrammes pour m tendant vers l'infini alors que la courbe pointillée correspond à la densité de probabilité de ce mélange de gaussiennes.*

où K est la fonction triangle d'expression $K(t) = (1 - |t|)\mathbb{1}_{[-1,1]}(t)$.

$$\hat{h}^{(\infty)}(x) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{h}K\left(\frac{|x - x_i|}{h}\right) \tag{5.4}$$

L'expression (5.4) peut en fait être utilisée avec d'autres fonctions K que la fonction triangle précédente, qui sont dénommées noyaux, et introduisant une classe d'estimateur de densité de probabilité non paramétrique que l'on nomme les estimateurs à noyau. Ces estimateurs peuvent être plus coûteux à calculer, la complexité dépendant du coût d'évaluation de la fonction K. Néanmoins,de tels estimateurs présentent l'avantage par rapport aux histogrammes de ne pas introduire le paramètre x_0 qui peut conduire à de fausses conclusions quant à la forme prise par la densité. Ils permettent également d'obtenir des estimateurs avec une régularité plus forte que les histogrammes, ceux-ci héritant de la régularité du noyau K utilisé. Enfin, en dimension plus grande que 1, les histogrammes présentent très rapidement une limitation du fait du fléau de la dimension. En

effet, une très grande taille d'échantillon est nécessaire afin d'avoir un nombre significatif de réalisations dans les pavés droits $[t_{k_1}^{(1)}, t_{k_1+1}^{(1)}] \times \cdots \times [t_{k_d}^{(d)}, t_{k_d+1}^{(d)}]$ *et donc un histogramme qui ne soit pas trop bruité. Ce problème spécifique n'apparaît pas avec les méthodes à noyaux bien que le fléau de la dimension s'applique à ces méthodes.*

5.2 Estimation non-paramétrique à noyaux

L'objectif de l'estimation non-paramétrique à noyau est d'estimer la densité π *d'une variable ou d'un vecteur aléatoire, étant donné une réalisation* $(x^{(i)})_{1\leq i\leq N}$ *d'un échantillon* $(X^{(i)})_{1\leq i\leq N}$ *de taille* N *de variables suivant la même loi que cette variable ou de ce vecteur aléatoire. La qualité de cette estimation est bien entendu liée au nombre* N *de réalisations utilisées, ce qui sera brièvement explicité dans cette section.*

5.2.1 Méthode en dimension 1

5.2.1.1 Principe de la méthode

Comme introduit précédemment, l'ingrédient de base des méthodes d'estimation à noyau est le noyau K *utilisé, qui est dans le cas monodimensionnel une fonction d'une variable réelle, que l'on choisit souvent vérifiant les propriétés suivantes :*

- K *est positive sur* $\mathbb{R}$
- K *est intégrable d'intégrale* 1
- K *est symétrique*
- K *est de variance finie* σ_K^2

Les deux premières propriétés font de K *une densité de probabilité d'une variable aléatoire réelle et la quatrième propriété lui assure l'existence d'un moment d'ordre* 2*, alors que la troisième propriété assure que la variable aléatoire associée est de moyenne nulle. Ces propriétés ne sont pas obligatoires [117], mais nous les considérerons vérifiées dans la suite. Tout comme pour les histogrammes, un paramètre* h *nommé fenêtre est généralement introduit, permettant de considérer une famille de noyaux* $(K_h)_{h\in\mathbb{R}^+}$ *obtenus à partir du noyau* K*, à l'aide de l'expression suivante :*

$$K_h(x) = \frac{1}{h} K\left(\frac{x}{h}\right) \tag{5.5}$$

Les noyaux K_h *conservent l'ensemble des propriétés précédemment citées du noyau* K*, la variance du noyau se retrouvant quant à elle multipliée par un facteur* h^2*. En conséquence, les noyaux* K_h *se voient plus ou moins piqués suivant que la valeur de* h *est plus ou moins grande. Un exemple de l'influence du paramètre* h *sur la forme du noyau est visible sur la figure 5.4 pour un noyau*

gaussien.

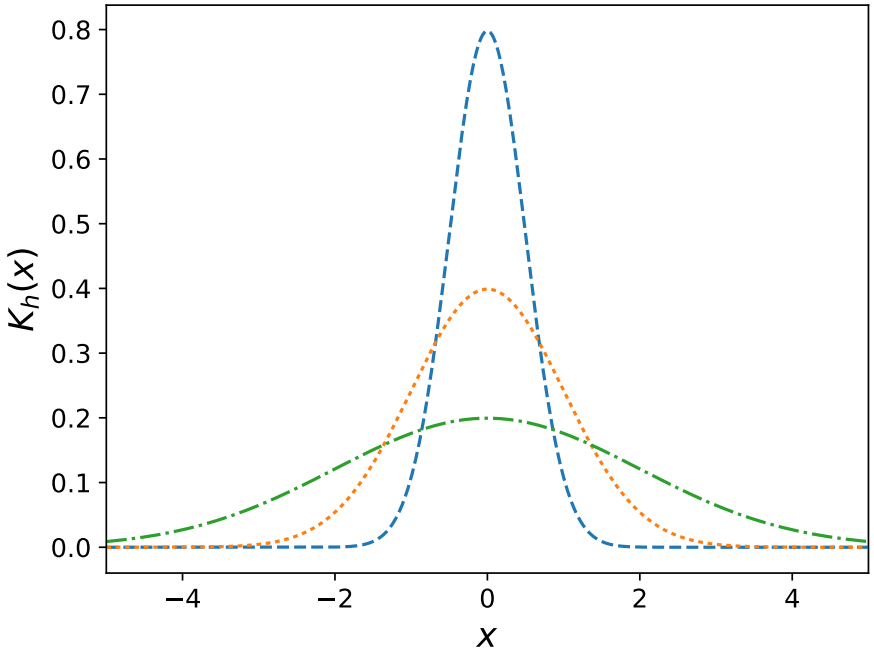

FIGURE 5.4 – *Profils de noyaux gaussiens avec $h = 2$ (tirets), $h = 1$ (pointillés) et $h = 0.5$ (tiret-points).*

Étant donné un échantillon de variables aléatoires $(X^{(i)})_{1\leq i\leq N}$ toutes de densité de probabilité π et mutuellement indépendantes, on estime la densité de probabilité π à l'aide de l'estimateur $\hat{\pi}$, dont l'expression est la suivante :

$$\hat{\pi}(x) = \frac{1}{N}\sum_{i=1}^{N} K_h(x - x^{(i)}) \tag{5.6}$$

Plusieurs choses peuvent être dites de l'expression (5.6) de l'estimateur $\hat{\pi}(x)$. K_h étant une densité de probabilité, l'ensemble des réalisations de $\hat{\pi}$ sont également des densités de probabilité du fait de la normalisation. La dépendance de $\hat{\pi}(x)$ en l'échantillon $(X^{(i)})_{1\leq i\leq N}$ est bien visible dans l'expression (5.6). En fait, $\hat{\pi}(x)$, en tant qu'estimateur, peut être vu comme une variable aléatoire dépendant de la famille de variables aléatoires $(X^{(i)})_{1\leq i\leq N}$, dont on utilisera une réalisation $(x^{(i)})_{1\leq i\leq N}$ de l'échantillon pour une utilisation pratique.

La présence de la fenêtre h permet à chaque $X^{(i)}$ d'influencer un voisinage plus ou moins grand autout de lui, et assurer que la valeur locale de la densité de probabilité n'est dû qu'aux $X^{(i)}$ "proches" de ce point. Pour rappel, la valeur

de la densité de probabilité π en un point x peut être informellement définie comme vérifiant la relation suivante, pour une quantité dx infinitésimale :

$$\pi(x)dx \approx P(x < X < x + dx) \tag{5.7}$$

De fait, la valeur de la densité de probabilité π en x est une quantité dépendant du voisinage immédiat de x, la relation précédente étant d'autant plus vraie que la quantité dx est petite. Cette dernière définition informelle de la densité de probabilité permet de comprendre que le choix de la valeur de la fenêtre h est crucial en ce sens qu'il permet de spécifier le voisinage influencé par chacune des réalisations, et donc les réalisations qui influenceront la valeur en un point x. Ce choix est déterminant afin d'obtenir une bonne estimation de la densité de probabilité, et on peut facilement se convaincre que cette fenêtre doit avoir une certaine dépendance avec la taille N de l'échantillon. En effet, une fenêtre h ne dépendant pas de N, et donc constante, ne permettra par exemple pas de reproduire un pic de la densité de probabilité dont la largeur est inférieure à la fenêtre, et la densité de probabilité estimée présentera alors un pic de largeur supérieure à h. Cette dernière affirmation peut s'expliquer par le fait que l'estimateur à noyau est en fait la convolution du noyau K par la densité de probabilité empirique π_{emp} définie par la relation (5.8), où δ_x fait référence à la distribution de Dirac au point x, qui n'est en pratique pas utilisé sauf pour les études par ré-échantillonnage (bootstrap).

$$\pi_{emp}(x) = \frac{1}{N}\sum_{i=1}^{N} \delta_{X^{(i)}}(x) \tag{5.8}$$

Cette opération de convolution implique que chacune des distributions de Dirac présente dans l'expression (5.8) se retrouve élargie avec une largeur de l'ordre de h, ne permettant pas la reproduction des pics trop étroits. On retrouve ici le fait que la densité de probabilité en un point x est dépendante du voisinage local de ce point, le caractère local de ce voisinage dépendant du point considéré. Dans le cas où la seule information disponible correspond à la réalisation de l'échantillon dont les composantes suivent la loi donnée par la densité de probabilité, ce voisinage local peut être défini simplement en regroupant les réalisations les plus proches du point x considéré. Bien entendu, à mesure que la taille de l'échantillon augmente, de plus en plus d'entre elles se situeront proche de x, ce qui implique que ce voisinage local devra être raffiné à mesure que le nombre N augmente, afin de permettre de ne prendre en compte que les réalisations les plus proches de x. Cela revient donc à demander que la fenêtre h décroissent avec la taille N de l'échantillon, pour ultimement tendre vers 0. Cette décroissance ne doit cependant pas être trop rapide. En effet, cela reviendrait à "isoler" chaque réalisation, offrant systématiquement

une densité de probabilité estimée avec un nombre de pics de l'ordre de la taille de l'échantillon.

La qualité de l'estimation de la densité de probabilité dépend donc de la fenêtre choisie, mais également de la taille de l'échantillon utilisé bien entendu. Il convient de mesurer cette qualité, ce qui peut être réalisé de différentes manières, qui sont brièvement passées en revue dans la suite avec quelques résultats théoriques de convergence permettant d'introduire les leviers possibles d'amélioration de ces méthodes.

5.2.1.2 Critère d'évaluation de la qualité de l'estimation

La qualité de l'estimation de la densité de probabilité, entendue ici comme l'erreur commise par l'estimation, peut être définie de différentes manières, souvent en lien avec une ou plusieurs caractéristique de l'estimation que l'on souhaite être la meilleure possible et pour laquelle des résultats de convergences sont recherchés.

Il est ainsi possible de s'intéresser à une mesure de qualité ponctuelle, s'intéressant à un point x du domaine particulier. Une mesure de qualité ponctuelle classique est l'erreur quadratique moyenne (MSE pour Mean Squarred Error en anglais) [117], définie par :

$$MSE(x) = E\left[\left(\hat{\pi}(x) - \pi(x)\right)^2\right] = Bias\left[\hat{\pi}(x)\right]^2 + Var\left[\hat{\pi}(x)\right] \tag{5.9}$$

L'erreur quadratique moyenne se décompose comme la somme d'un terme associé au biais de l'estimateur $\hat{\pi}(x)$, dont l'expression est donnée par (5.10), et un terme associé à la variance de cet estimateur dont l'expression est donnée par (5.11).

$$Bias\left[\hat{\pi}(x)\right] = E\left[\hat{\pi}(x) - \pi(x)\right] = E\left[\hat{\pi}(x)\right] - \pi(x) \tag{5.10}$$

$$Var\left[\hat{\pi}(x)\right] = E\left[\left(\hat{\pi}(x) - E\left[\hat{\pi}(x)\right]\right)^2\right] \tag{5.11}$$

Cette décomposition de l'erreur quadratique, parfois nommée compromis biais-variance, est un résultat classique en statistiques pour l'estimation d'une quantité par un estimateur. $\hat{\pi}(x)$ est en effet une variable aléatoire comme précédemment expliqué, et peut donc être vu comme un estimateur statistique de $\pi(x)$. L'espérance mathématiques dans l'expression (5.9) est à ce titre prise selon la loi de la variable aléatoire $\hat{\pi}(x)$, qui est directement liée à la loi de l'échantillon de variables aléatoires $(X^{(i)})_{1\leq i\leq N}$ utilisé dans la construction de $\hat{\pi}(x)$.

Optimiser une mesure de qualité en un ou plusieurs points comme l'er-

reur quadratique moyenne permet d'assurer une bonne qualité de l'estimation au niveau de ces points, mais peut mener à une dégradation de cet estimateur ailleurs. Le besoin peut cependant être d'assurer une bonne qualité de l'estimateur sur l'ensemble du domaine. Des mesures de qualité plus globales peuvent être envisagées à cette fin. Parmi ces mesures de qualité se trouve l'erreur quadratique intégrée moyenne $MISE$ (Mean Integrated Squarred Error en anglais), définie par :

$$MISE(\hat{\pi}) = E\left[\int_{\mathbb{R}} [\hat{\pi}(x) - \pi(x)]^2 \, dx\right] \tag{5.12}$$

Dans le terme de droite de l'expression (5.12), il est possible d'intervertir l'opérateur d'espérance et d'intégration, ce qui permet d'affirmer que l'erreur quadratique intégrée moyenne est égale à l'erreur quadratique moyenne intégrée $IMSE$ (Integrated Mean Squarred Error en anglais), qui correspond à l'intégrale de l'erreur quadratique moyenne précédemment présentée sur l'ensemble du domaine. Cette mesure de qualité est basée sur la norme L_2, offrant une étude théorique plus simple que pour d'autres normes, telles que la norme L_1 ou la norme L_∞ par exemple. Une brève étude théorique de cette mesure de qualité est effectuée dans la section suivante, afin de rendre compte des leviers d'amélioration d'une estimation par noyau.

5.2.1.3 Quelques résultats théoriques de convergence

L'objet de cette section est d'expliciter brièvement un résultat de convergence et de montrer l'importance cruciale du paramètre de fenêtre h dans la convergence d'un estimateur à noyau. Pour cette étude, les variables aléatoires $(X^{(i)})_{1\leq i\leq N}$ sont supposées indépendantes et uniformément distribuées selon la loi donnée par la densité de probabilité π à estimer. Comme le montre l'expression (5.9), l'erreur quadratique moyenne est la somme du carré du biais de l'estimateur et de la variance de cet estimateur. Il convient donc dans un premier temps d'exprimer chacun de ces deux termes. Tout d'abord, en utilisant l'indépendance des variables aléatoires $X^{(i)}$ ainsi que le fait qu'elles suivent la même loi, on peut exprimer l'espérance de $\hat{\pi}(x)$, nécessaire au calcul du biais, par l'expression (5.13) où X est une variable aléatoire de même loi que les $X^{(i)}$.

$$E\left[\hat{\pi}(x)\right] = \frac{1}{N}\sum_{i=1}^{N} E\left[K_h\left(x - X^{(i)}\right)\right] = E\left[K_h\left(x - X\right)\right] \tag{5.13}$$

Afin d'obtenir une expression pour le MSE, on considère la densité π comme étant suffisamment régulière. Ainsi, en considérant une expansion en série de Taylor de π au point x, et en utilisant l'expression (5.13), il est possible

d'obtenir l'expression de l'espérance de $\hat{\pi}(x)$ suivante :

$$\begin{aligned} E\left[\hat{\pi}(x)\right] &= \int_{\mathbb{R}} K_h\left(x-t\right)\pi(t)dt = \int_{\mathbb{R}} \frac{1}{h}K\left(\frac{x-t}{h}\right)\pi(t)dt \\ &= \int_{\mathbb{R}} K(w)\pi(x-hw)dw \\ &= \int_{\mathbb{R}} K(w)\left[\pi(x) - hw\pi'(x) + \frac{1}{2}h^2w^2\pi''(x) + o(h^2)\right]dw \\ &= \pi(x)\int_{\mathbb{R}} K(w)dw - h\pi'(x)\int_{\mathbb{R}} wK(w)dw \\ &+ \frac{1}{2}h^2\pi''(x)\int_{\mathbb{R}} w^2K(w)dw + o(h^3) \end{aligned} \tag{5.14}$$

L'expression (5.14) permet de justifier certaines des hypothèses faites sur le noyau K utilisé. En effet, considérer un noyau K qui est une densité de probabilité permet au premier terme de l'expression (5.14) de se simplifier pour ne laisser que $\pi(x)$. La symétrie du noyau K permet quant à elle de s'affranchir du second terme, l'intégrale s'annulant. Finalement, le biais de $\hat{\pi}(x)$ a pour expression :

$$E\left[\hat{\pi}(x) - \pi(x)\right] = \frac{1}{2}h^2\sigma_K^2\pi''(x) + o(h^3) \tag{5.15}$$

L'expression du biais (5.15) permet d'affirmer que l'estimateur à noyau $\hat{\pi}(x)$ converge vers $\pi(x)$ si la fenêtre h tend vers 0 lorsque le nombre d'échantillons N tend vers l'infini, comme suggéré précédemment. Le second terme intervenant dans l'expression du MSE est la variance de l'estimateur $\hat{\pi}(x)$, dont une expression peut être obtenue de manière similaire sous les mêmes conditions [117], donnant pour celle-ci l'expression suivante :

$$Var\left[\hat{\pi}(x)\right] = \frac{\pi(x)\int_{\mathbb{R}} K(t)^2dt}{Nh} - \frac{\pi(x)^2}{N} + O\left(\frac{h}{N}\right) \tag{5.16}$$

Le premier terme de l'expression (5.16) ne peut tendre vers 0 que si h ne tend pas trop vite vers 0, plus précisément si le produit hN tend vers l'infini lorsque N tend vers l'infini. Cette dernière affirmation justifie la nécessité qu'à la fenêtre h de ne pas décroître trop rapidement avec N, comme il avait été suggéré intuitivement plus tôt. En notant dans la suite $R(f) = \int_{\mathbb{R}} f(t)^2dt$, on peut à partir des expressions (5.15) et (5.16) obtenir une expression de $MISE(\hat{\pi})$,

qui est donnée par :

$$MISE(\hat{\pi}) = \frac{1}{4}\sigma_K^4 h^4 R(\pi'') + \frac{R(K)}{Nh} - \frac{R(\pi)}{N} + o(h^6) + O\left(\frac{h}{N}\right) \tag{5.17}$$

Afin de faciliter l'étude, le $AMISE$ est étudié, qui n'est autre que le $MISE$ asymptotique, dans lequel les termes négligeables sont omis, de sorte que le $AMISE$ de $\hat{\pi}$ est donné par l'expression (5.18).

$$AMISE(\hat{\pi}) = \frac{1}{4}\sigma_K^4 h^4 R(\pi'') + \frac{R(K)}{Nh} - \frac{R(\pi)}{N} \tag{5.18}$$

Le $AMISE$, en plus de dépendre de caractéristiques intrinsèques de la densité de probabilité π à estimer et du noyau K utilisé, dépend de la fenêtre h ainsi que de la taille N de l'échantillon. En tant que fonction de la fenêtre h, il possède un minimum qu'il est possible de trouver en annulant la dérivé par rapport à h de l'expression (5.18), minimum se trouvant pour la valeur h^* de la fenêtre donnée par l'expression (5.19).

$$h^* = \left(\frac{R(K)}{\sigma_K^4 R(\pi'')}\right)^{\frac{1}{5}} N^{-\frac{1}{5}} \tag{5.19}$$

Ainsi, la valeur de la fenêtre h optimale à utiliser pour le $AMISE$ tend bien vers 0 lorsque N tend vers l'infini, et vérifie bien également que le produit hN tend vers l'infini lorsque N tend vers l'infini. En injectant l'expression de h^* dans l'expression (5.18), il est possible d'obtenir le $AMISE$ optimal $AMISE^*$, qui est une fonction du nombre d'échantillons.

$$AMISE^*(\hat{\pi}) = \frac{4}{5}\left(R(K)\sigma_K\right)^{\frac{4}{5}} R(\pi'')^{\frac{1}{5}} N^{-\frac{4}{5}} \tag{5.20}$$

Le dernier terme présent dans l'expression (5.18) a ici été omis, puisque négligeable devant l'unique terme de l'expression (5.20). Cette dernière expression montre que l'on peut jouer sur la qualité asymptotique de notre estimateur de deux façons différentes. La première consiste à choisir le noyau, celui-ci ayant une influence sur l'$AMISE^*$, au travers du produit $R(K)\sigma_K$. Ce produit permet d'introduire la notion d'efficacité pour un noyau [117], le noyau le plus efficace étant celui d'Epanechnikov [29]. Cependant, d'autres critères que l'efficacité sont à prendre en compte dans le choix du noyau, qui peuvent être d'ordre pratique ou théorique. La seconde façon de jouer sur la qualité est d'augmenter la taille de l'échantillon utilisé. Cette dernière étude permet de donner un ordre de grandeur de la convergence que l'on peut atteindre d'un estimateur issu d'une méthode à noyau, qui est de l'ordre de $N^{-\frac{4}{5}}$ pour l'erreur quadratique moyenne

intégrée. Cette convergence est plus lente que dans le cas de méthodes paramétriques, où la vitesse de convergence de l'AMISE est plutôt de l'ordre de N^{-1} *[117]. La vitesse de convergence n'est cependant pas le seul critère à prendre en compte, et la flexibilité des méthodes à noyau reste un atout essentiel pour la reproduction de densités de probabilité pour lesquelles aucune information n'est disponible à priori.*

Cette rapide description des critères de qualité et de leur convergence a permis de montrer brièvement ce que l'on peut attendre des méthodes à noyau d'un point de vue théorique en dimension 1.

5.2.2 Cas général en dimension supérieure à 1

Pour l'utilisation de méthodes à noyau en dimension strictement supérieure à 1*, il convient d'utiliser des noyaux multidimensionnels* K*. Dans la suite, on supposera que les noyaux multidimensionnels utilisés vérifient les propriétés de positivité et de symétrie, et également qu'ils sont intégrables d'intégrale* 1 *sur* $\mathbb{R}^d$*. Ces propriétés étaient déjà demandées pour les noyaux en dimension* 1*. La propriété de variance finie se doit d'être remplacée par l'existence d'une matrice de covariance, mais est dans un premier temps remplacée par la propriété donnée par l'expression* (5.21) *dans laquelle* $\mathbf{I}_d$ *est la matrice identité.*

$$\int_{\mathbb{R}^d} \mathbf{x}\mathbf{x}^T K(\mathbf{x}) d\mathbf{x} = \mathbf{I}_d \tag{5.21}$$

Dans le cas unidimensionnel, l'introduction de la fenêtre h *permettait de modifier la variance des noyaux considérés, à l'aide d'une composition par une transformation linéaire caractérisé par ce paramètre* h*. Cela permettait, étant donné un noyau* K*, de construire un noyau* K_h *avec une variance quelconque. Le noyau utilisé dans le cas multidimensionnel peut également être composé préalablement à une transformation linéaire afin d'obtenir un noyau avec une matrice de covariance quelconque. Une telle transformation linéaire est caractérisée par une matrice carré* H *de dimension la dimension* d *de l'espace, de sorte que le noyau* K_H *utilisé est donné par :*

$$K_H(\mathbf{x}) = \frac{1}{|H|} K\left(H^{-1}\mathbf{x}\right) \tag{5.22}$$

Dans l'expression (5.22)*, le déterminant de la matrice* H *apparaissant au dénominateur permet la normalisation du noyau afin d'avoir son intégrale égale à* 1 *sur* $\mathbb{R}^d$*. La matrice* H *se doit donc d'être inversible pour que le noyau* K_H *puisse être défini. La matrice de covariance du noyau* K_H *est, quant à elle,*

donnée par :

$$\int_{\mathbb{R}^d} \mathbf{x}\mathbf{x}^T K_H(\mathbf{x}) d\mathbf{x} = \int_{\mathbb{R}^d} \mathbf{x}\mathbf{x}^T K(H^{-1}\mathbf{x}) \frac{d\mathbf{x}}{|H|} = \int_{\mathbb{R}^d} (H\mathbf{u})(H\mathbf{u})^T K(\mathbf{u}) d\mathbf{u} \quad (5.23)$$

En s'intéressant au coefficient $c_{i,j}$ de la ligne i et de la colonne j de la matrice de covariance du noyau K_H à l'aide du membre de gauche de l'expression (5.23), et en utilisant la covariance du noyau K donnée par l'expression (5.21), on trouve l'expression suivante pour ce coefficient :

$$\begin{aligned} c_{i,j} &= \int_{\mathbb{R}^d} \left(\sum_{k=1}^{d} h_{i,k} u_k \right) \left(\sum_{l=1}^{d} h_{i,l} u_l \right) K(\mathbf{u}) d\mathbf{u} \\ &= \sum_{k=1}^{d} \sum_{l=1}^{d} h_{i,k} h_{j,l} \int_{\mathbb{R}^d} u_k u_l K(\mathbf{u}) d\mathbf{u} \\ &= \sum_{k=1}^{d} h_{i,k} h_{j,k} \end{aligned} \quad (5.24)$$

Le coefficient $c_{i,j}$ n'est autre que le coefficient de la ligne i et de la colonne j de la matrice HH^T. La matrice de covariance du noyau K_H est donc la matrice HH^T. Il est donc bien possible de construire à partir d'un noyau K vérifiant les propriétés décrites plus haut un noyau K_H avec une matrice de covariance quelconque Σ, sous réserve que celle-ci soit non dégénérée. Il suffit pour cela de prendre une matrice H vérifiant $HH^T = \Sigma$, la racine carré $\Sigma^{\frac{1}{2}}$ de la matrice Σ faisant l'affaire, et existant de par la nature symétrique et définie positive de cette dernière matrice. La matrice H introduite peut en fait être écrite sous la forme $H = hA$, avec h un nombre réel positif et A une matrice de carré de dimension d et de déterminant 1. Il suffit pour cela de prendre $h = |H|^{\frac{1}{d}}$ et l'expression de la matrice A est alors donnée par l'expression suivante :

$$A = \frac{H}{h} = \frac{H}{|H|^{\frac{1}{d}}} \quad (5.25)$$

Avec cette dernière écriture, la matrice A contrôle la forme du noyau, alors que le paramètre h contrôle sa taille.

L'estimateur $\hat{\pi}$ d'une densité de probabilité π en un point $\mathbf{x}$ de $\mathbb{R}^d$, obtenu à partir de N vecteurs aléatoires $(\mathbf{X}^{(i)})_{1 \leq i \leq N}$ indépendants de densité de probabilité π et d'un noyau $K_{h,A}$ est, similairement au cas de la dimension 1, donné

par l'expression suivante :

$$\hat{\pi}(\mathbf{x}) = \frac{1}{Nh^d} \sum_{i=1}^{N} K_{h,A}(\mathbf{x} - \mathbf{X}^{(i)}) \tag{5.26}$$

Les critères de qualité précédemment présentés pour la dimension 1 *peuvent également être définis pour le cas multidimensionnel, et notamment le* $AMISE$ *de l'estimateur* $\hat{\pi}$ *qui est donnée dans [117] par l'expression suivante, en supposant des conditions de régularité et d'intégrabilité suffisante pour la densité* π *:*

$$AMISE(\hat{\pi}) = \frac{1}{4}h^4 \int_{\mathbb{R}^d} \left(Tr\left[AA^T \nabla^2 \pi(\mathbf{x})\right]\right)^2 d\mathbf{x} + \frac{R(K)}{Nh^d} \tag{5.27}$$

On peut montrer à partir de l'expression (5.27) *que pour une matrice* A *fixée, la valeur de* h *permettant de minimiser le* $AMISE$ *est proportionnelle à* $N^{-\frac{1}{4+d}}$*, et que le* $AMISE$ *optimal est alors proportionnel à* $N^{-\frac{4}{4+d}}$*. Ce résultat est bien en accord avec le résultat obtenu lors de l'analyse en dimension* 1*, et montre que la vitesse de convergence de l'estimateur dépend de la dimension* d *du problème, et diminue à mesure que la dimension* d *augmente. Cette vitesse de convergence qui diminue avec la dimension est une conséquence du "fléau de la dimension", comme expliqué dans le chapitre* 7 *de [117]. Le premier terme dans l'expression* (5.27) *est lié au biais de l'estimateur, alors que le second terme est lié à la variance de celui-ci. Le biais de l'estimateur est d'autant plus faible que les propriétés locales sont bien capturées, ce qui est d'autant plus vrai lorsque la valeur de* h *est petite comme le traduit la dépendance du terme de biais en* h^4*. Le second terme, correspondant à la variance, est au contraire pénalisé lorsque le voisinage considéré autour des échantillons est "trop" local. En effet, dans une telle situation où le voisinage est trop local, plusieurs possibilités se présenterait pour un point du domaine probable :*

- *Ou bien il est présent dans le voisinage d'un ou plusieurs des réalisations, entraînant une valeur conséquente de l'estimateur, d'autant plus importante que le voisinage est local (la valeur de l'estimateur est inversement proportionnel au volume du voisinage).*
- *Ou bien il n'est présent dans aucun voisinage d'échantillons, entraînant une valeur de l'estimateur nulle ou proche de* 0*.*

Considérer des voisinages autour des échantillons "trop" locaux correspond en fait à avoir une probabilité significative de se retrouver dans la seconde situation où la valeur de l'estimateur est proche de 0*, entraînant de fait une importante variance de celui-ci localement, pouvant prendre avec des probabilités significatives des valeurs faibles ou importantes. Afin de réduire la contribution de la variance de l'estimateur dans l'*$AMISE$*, il est donc nécessaire*

d'avoir des voisinages autour des échantillons suffisamment étendus, afin d'assurer une bonne "couverture" de l'ensemble des points probables du domaine. Un compromis entre le biais et la variance est donc à trouver afin de minimiser l'AMISE. Cependant, à mesure que la dimension d de l'espace augmente, la "couverture" de l'espace par les voisinages des échantillons devient de plus en plus difficile du fait du fléau de la dimension, nécessitant de considérer une décroissance du paramètre h avec le nombre d'échantillons N de plus en plus faible à mesure que la dimension augmente, décroissance qui est comme précisé précédemment de l'ordre de $N^{-\frac{1}{4+d}}$*, et qui entraîne une vitesse de convergence vers* 0 *de l'AMISE de plus en plus faible à mesure que la dimension augmente.*

Dans la présente thèse, l'utilisation des méthodes à noyau a pour but d'être capable d'échantillonner des vecteurs aléatoires de faible dimension (typiquement inférieure à 5*), obtenus après réduction de la dimension stochastique du système considéré. De ce fait, les méthodes à noyau restent tout de même envisageables malgré la diminution de la vitesse de convergence de celles-ci. Les résultats théoriques présentés jusqu'à présent permettaient de donner une idée de ce que l'on pouvait attendre comme convergence de ces méthodes. En pratique, des différences notables par rapport à ce qui a été présenté seront cependant présentes, notamment parce que les résultats théoriques présentés ne peuvent être utilisés du fait de la non connaissance de la densité de probabilité que l'on cherche à estimer. Le paramètre h était dans cette section considéré uniforme sur l'ensemble du domaine, ce qui ne sera pas nécessairement le cas en pratique. Les résultats théoriques obtenus reposaient sur l'hypothèse que les échantillons étaient issus de vecteurs aléatoires indépendants entre eux, ce qui ne sera pas non plus le cas dans cette thèse, les échantillons étant issus de méthodes de Quasi-Monte Carlo randomisé. Enfin, diverses méthodes qui seront présentées par la suite, inspirées de considérations théoriques, permettent d'améliorer en pratique la qualité de l'estimation de la densité de probabilité.*

5.2.3 Utilisation de noyaux gaussiens pour la génération de vecteurs aléatoires

Cette partie est destinée à montrer qu'il est possible d'utiliser simplement la méthode d'inversion généralisée présentée dans le chapitre 3 pour la génération de vecteurs aléatoires lorsque des noyaux gaussiens sont utilisés pour l'estimation de la densité de probabilité de la loi inconnue. Pour rappel, cette méthode nécessite d'avoir accès à l'ensemble des lois de probabilité conditionnelles, plus précisément les fonctions de répartition de ces lois conditionnelles. Les noyaux gaussiens en particulier présentent de nombreuses propriétés permettant d'accéder à ces fonctions de répartition conditionnelles. On suppose donc dans la suite que le noyau K considéré est un noyau gaussien de matrice de covariance

Σ, *de sorte que sa densité de probabilité est donnée par :*

$$K^{\mathbf{\Sigma}}(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{d}{2}}|\mathbf{\Sigma}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}\mathbf{x}^T\mathbf{\Sigma}^{-1}\mathbf{x}\right) \tag{5.28}$$

Un tel noyau peut bien entendu être vu comme la composition du noyau gaussien qui a pour matrice de covariance la matrice identité, avec une transformation linéaire de matrice $H = \Sigma^{\frac{1}{2}}$. *Pour la construction de l'estimateur* $\hat{\pi}$ *de la densité de probabilité* π, *on étend la définition donnée par l'expression* (5.26) *en autorisant également que la matrice de covariance du noyau dépende de l'échantillon considéré [117], de sorte que l'expression de l'estimateur* $\hat{\pi}$ *en un point* $\mathbf{x}$ *est donnée par :*

$$\hat{\pi}(\mathbf{x}) = \frac{1}{N}\sum_{i=1}^{N} K^{\mathbf{\Sigma}^{(i)}}(\mathbf{x} - \mathbf{X}^{(i)}) = \frac{1}{N}\sum_{i=1}^{N} K^{\mathbf{X}^{(i)},\mathbf{\Sigma}^{(i)}}(\mathbf{x}) \tag{5.29}$$

La densité de probabilité $\hat{\pi}$ *étant désormais connue, la méthode d'inversion généralisée peut être utilisée afin de :*

- *générer des échantillons du vecteur aléatoire* $\hat{\mathbf{X}} = (\hat{X}_1, \ldots, \hat{X}_d)$ *de densité de probabilité* $\hat{\pi}$ *à partir d'un vecteur aléatoire* $\mathbf{U} = (U_1, \ldots, U_d)$ *de loi uniforme sur* $[0,1]^d$.
- *générer des échantillons* $\hat{\mathbf{U}} = (\hat{U}_1, \ldots, \hat{U}_d)$ *de loi "presque" uniforme à partir d'un vecteur aléatoire* $\mathbf{X} = (X_1, \ldots, X_d)$ *de densité de probabilité* π.

Dans les deux cas, le choix est fait de construire les composantes des vecteurs aléatoires dans l'ordre naturel. Les deux relations reliant les premières composantes $\hat{X}_1$ *et* U_1, *ainsi que* X_1 *et* $\hat{U}_1$, *sont données par les relations* (5.30) *:*

$$\begin{cases} \hat{X}_1 = F^{-1}_{\hat{\pi}_{(1)}}(U_1) \\ \hat{U}_1 = F_{\hat{\pi}_{(1)}}(X_1) \end{cases} \tag{5.30}$$

Les deux relations (5.30) *font intervenir la fonction de répartition marginale* $F_{\hat{\pi}_1}$, *dont l'expression peut être facilement obtenue à partir des fonctions de répartitions marginales* $F_{K^{\mathbf{X}^{(i)},\mathbf{\Sigma}^{(i)}}_{(1)}}$ *des noyaux gaussiens, et est donnée par* (5.31).

$$F_{\hat{\pi}_1}(x_1) = \frac{1}{N}\sum_{i=1}^{N} F_{K^{\mathbf{X}^{(i)},\mathbf{\Sigma}^{(i)}}_{(1)}}(x_1) \tag{5.31}$$

Les noyaux marginaux $K^{\mathbf{X}^{(i)},\mathbf{\Sigma}^{(i)}}_{(1)}$ *sont des noyaux gaussiens [26] d'une*

seule variable, et les fonctions de répartitions $F_{K_1^{\mathbf{X}^{(i)},\mathbf{\Sigma}^{(i)}}}$ *ont donc une expression analytique rendant leur évaluation simple, l'expression analytique de la fonction de répartition d'une variable aléatoire gaussienne de moyenne* μ *de de variance* σ^2 *étant donnée par l'expression suivante :*

$$F_{\mu,\sigma^2}(x) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right] \tag{5.32}$$

Le calcul de $\hat{U}_1$ *est alors direct, alors que le calcul de* $\hat{X}_1$ *nécessite l'inversion de la fonction de répartition* $F_{\hat{\pi}_1}$*, pouvant être réalisée à l'aide d'une recherche de zéro d'une fonction continue, tâche pour laquelle de nombreuses méthodes existent [104] parmi lesquelles une recherche par dichotomie ou encore une méthode de Newton.*

Une fois la première composante du vecteur aléatoire calculée, il convient de calculer les suivantes. On suppose désormais que les j *premières composantes sont calculées avec* $j < d$*. La* $(j+1)$*-ème composante peut être obtenue par la méthode d'inversion généralisée, à travers les relations suivantes :*

$$\begin{cases} \hat{X}_{j+1} = F^{-1}_{\hat{\pi}_{(j+1|1,\dots,j)}}(U_{j+1}|U_1,\dots,U_j) \\ \hat{U}_{j+1} = F_{\hat{\pi}_{(j+1|1,\dots,j)}}(X_{j+1}|X_1,\dots,x_j) \end{cases} \tag{5.33}$$

Les relations (5.33) font intervenir la fonction de répartition associée à la densité de probabilité conditionnelle $\hat{\pi}_{(j+1|1,\dots,j)}$*, dont la définition est donnée par l'expression suivante :*

$$\hat{\pi}_{(j+1|1,\dots,j)}(x_{j+1}|x_1,\dots,x_j) = \frac{\hat{\pi}_{(1,\dots,j+1)}(x_1,\dots,x_{j+1})}{\hat{\pi}_{(1,\dots,j)}(x_1,\dots,x_j)} \tag{5.34}$$

Le dénominateur de l'expression (5.34) correspond à la densité de probabilité marginale de $\hat{\pi}$ *correspondant aux* j *premières composantes, et s'exprime*

de la façon suivante :

$$\begin{aligned}\hat{\pi}_{(1,\dots,j)}(x_1,\dots,x_j) &= \int_{\mathbb{R}^{d-j}} \hat{\pi}(x_1,\dots,x_d)dx_{j+1}\dots dx_d \\ &= \int_{\mathbb{R}^{d-j}} \left[\frac{1}{N}\sum_{i=1}^{N} K^{\mathbf{X}^{(i)},\mathbf{\Sigma}^{(i)}}(x_1,\dots,x_d)\right] dx_{j+1}\dots dx_d \\ &= \frac{1}{N}\sum_{i=1}^{N}\int_{\mathbb{R}^{d-j}} K^{\mathbf{X}^{(i)},\mathbf{\Sigma}^{(i)}}(x_1,\dots,x_d)dx_{j+1}\dots dx_d \\ &= \frac{1}{N}\sum_{i=1}^{N} K^{\mathbf{X}^{(i)},\mathbf{\Sigma}^{(i)}}_{(1,\dots,j)}(x_1,\dots,x_j)\end{aligned} \tag{5.35}$$

Ainsi, le calcul du dénominateur nécessite la connaissance de la densité de probabilité marginale $K^{\mathbf{X}^{(i)},\mathbf{\Sigma}^{(i)}}_{(1,\dots,j)}$ *des* j *premières composantes pour l'ensemble des noyaux* $K^{\mathbf{X}^{(i)},\mathbf{\Sigma}^{(i)}}$*, qui dans le cas d'un noyau gaussien sont également des noyaux gaussiens et sont donc connus. Le numérateur est lui aussi une densité de probabilité marginale de* $\hat{\pi}$*, qu'il est préférable d'exprimer différemment afin d'avoir un accès simple à la fonction de répartition de la densité de probabilité conditionnelle* $\hat{\pi}_{(j+1|1,\dots,j)}$*. En se servant de l'expression (5.35) pour l'expression d'une densité de probabilité marginale de* $\hat{\pi}$*, il vient :*

$$\begin{aligned}\hat{\pi}_{(1,\dots,j+1)}(x_1,\dots,x_{j+1}) &= \frac{1}{N}\sum_{i=1}^{N} K^{\mathbf{X}^{(i)},\mathbf{\Sigma}^{(i)}}_{(1,\dots,j+1)}(x_1,\dots,x_{j+1}) \\ = \frac{1}{N}\sum_{i=1}^{N} K^{\mathbf{X}^{(i)},\mathbf{\Sigma}^{(i)}}_{(1,\dots,j)}(x_1,\dots,x_j) K^{\mathbf{X}^{(i)},\mathbf{\Sigma}^{(i)}}_{(j+1|1,\dots,j)}(x_{j+1}|x_1,\dots,x_j)\end{aligned} \tag{5.36}$$

Les expressions (5.35) et (5.36) font toutes deux apparaître les densités de probabilité marginales $K^{\mathbf{X}^{(i)},\mathbf{\Sigma}^{(i)}}_{(1,\dots,j)}$ *des noyaux gaussiens, elles mêmes des noyaux gaussiens [26]. Dans l'expression (5.36) apparaissent également les densités marginales conditionnelles* $K^{\mathbf{X}^{(i)},\mathbf{\Sigma}^{(i)}}_{(j+1|1,\dots,j)}$ *des noyaux gaussiens, qui sont également des noyaux gaussiens [26]. La densité de probabilité conditionnelle* $\hat{\pi}_{(j+1|1,\dots,j)}$ *peut donc s'exprimer comme une somme pondérée présentée dans l'expression (5.37).*

$$\hat{\pi}_{(j+1|1,\dots,j)}(x_{j+1}|x_1,\dots,x_j) = \sum_{i=1}^{N}\alpha_i(x_1,\dots,x_j)K^{\mathbf{X}^{(i)},\mathbf{\Sigma}^{(i)}}_{(j+1|1,\dots,j)}(x_{j+1}|x_1,\dots,x_j) \tag{5.37}$$

Les poids α_i de cette somme pondérée sont quant à eux donnés par l'expression (5.38).

$$\alpha_i(x_1,\dots,x_j) = \frac{K^{\mathbf{X}^{(i)},\mathbf{\Sigma}^{(i)}}_{(1,\dots,j)}(x_1,\dots,x_j)}{\sum_{k=1}^{N} K^{\mathbf{X}^{(k)},\mathbf{\Sigma}^{(k)}}_{(1,\dots,j)}(x_1,\dots,x_j)} \tag{5.38}$$

L'expression (5.37) de la densité de probabilité conditionnelle $\hat{\pi}_{(j+1|1,\dots,j)}$ étant désormais disponible, il est possible d'en déduire une expression pour la fonction de répartition $F_{\hat{\pi}_{(j+1|1,\dots,j)}}$ nécessaire au calcul de la $(j+1)$-ème composante des vecteurs aléatoires. L'expression de la fonction de répartition $F_{\hat{\pi}_{(j+1|1,\dots,j)}}$ s'obtient par l'intégration de la densité de probabilité $\hat{\pi}_{(j+1|1,\dots,j)}$, ce qui donne :

$$\begin{aligned}
F_{\hat{\pi}_{(j+1|1,\dots,j)}}(x_{j+1}|x_1,\dots,x_j) &= \int_{-\infty}^{x_{j+1}} \hat{\pi}_{(j+1|1,\dots,j)}(t|x_1,\dots,x_j)dt \\
&= \int_{-\infty}^{x_{j+1}} \left[\sum_{i=1}^{N} \alpha_i(x_1,\dots,x_j) K^{\mathbf{X}^{(i)},\mathbf{\Sigma}^{(i)}}_{(j+1|1,\dots,j)}(t|x_1,\dots,x_j)\right] dt \\
&= \sum_{i=1}^{N} \alpha_i(x_1,\dots,x_j) \int_{-\infty}^{x_{j+1}} K^{\mathbf{X}^{(i)},\mathbf{\Sigma}^{(i)}}_{(j+1|1,\dots,j)}(t|x_1,\dots,x_j)dt \\
&= \sum_{i=1}^{N} \alpha_i(x_1,\dots,x_j) F_{K^{\mathbf{X}^{(i)},\mathbf{\Sigma}^{(i)}}_{(j+1|1,\dots,j)}}(x_{j+1}|x_1,\dots,x_j)
\end{aligned} \tag{5.39}$$

Les noyaux conditionnels $K^{\mathbf{X}^{(i)},\mathbf{\Sigma}^{(i)}}_{(j+1|1,\dots,j)}$ étant des noyaux gaussiens, les fonctions de répartitions associées ont une expression analytique, et la fonction de répartition $F_{\hat{\pi}_{(j+1|1,\dots,j)}}$ peut donc être simplement évaluer. Le calcul de la composante $\hat{U}_{j+1}$ est donc possible grâce à cette dernière expression, et le calcul de la composante $\hat{X}_{j+1}$ quant à lui nécessite l'inversion de la fonction de répartition $F_{\hat{\pi}_{(j+1|1,\dots,j)}}$ qui a été effectuée dans cette thèse grâce à une recherche dichotomique.

L'utilisation de noyau gaussien pour l'estimation de densité de probabilité permet donc, grâce à leurs propriétés, d'utiliser simplement la méthode d'inversion généralisée avec l'estimateur à noyau obtenu. Il reste désormais à expliciter les propriétés des noyaux gaussiens utilisés, qui sont caractérisés par la matrice de covariance $\mathbf{\Sigma}$ utilisée. La caractérisation de ces noyaux gaussiens en pratique est l'objet de la prochaine section.

5.3 Détermination des paramètres de l'estimateur à noyau

Dans la première section de ce chapitre ont été présentés des résultats théoriques concernant les méthodes à noyaux permettant d'avoir une idée de la vitesse de convergence de ces méthodes. Ces résultats théoriques concernent des résultats asymptotiques, et sont inutilisables en pratiques de par le fait qu'ils nécessitent d'avoir la connaissance de la densité de probabilité que l'on cherche à estimer. Les résultats présentés avaient également le point commun d'utiliser la même paramétrisation du noyau pour l'ensemble de l'échantillon. Il est cependant possible de paramétrer le noyau de chaque élément de l'échantillon différemment, ce qui peut permettre d'obtenir de meilleures estimations dans certaines situations [117, 130]. La première approche est globale, et sera dans la suite qualifiée comme telle, alors que la seconde approche présente un caractère local. Cette section présente les choix réalisés dans cette thèse afin de déterminer les propriétés des noyaux gaussiens utilisés, à savoir les fenêtres $h^{(i)}$ ainsi que les transformations linéaires caractérisées par les matrices de covariances $\mathbf{A}^{(i)}$, qui définissent à eux deux les propriétés du voisinage considéré autour de chaque élément de l'échantillon.

5.3.1 Méthodes adaptatives locales

Les résultats théoriques présentés précédemment, qui sont des résultats globaux, ne sont pas directement exploitables en pratique car ils impliquent de déterminer les paramètres que sont la fenêtre h et la matrice A à l'aide des caractéristiques de la densité de probabilité à estimer, qui est inconnue. Ce problème se retrouve généralement, que ce soit dans le cas global ou local, la détermination des paramètres optimaux nécessitant la connaissance de caractéristiques globales ou locales de la densité de probabilité à estimer. L'idée des méthodes adaptatives est d'estimer les caractéristiques nécessaires de la densité de probabilité afin d'obtenir une valeur pour les paramètres optimaux. Différentes méthodes ont été étudiées, qui diffèrent selon que l'on s'intéresse à une optimisation globale ou locale. Pour le cas globale, plusieurs méthodes sont disponibles parmi lesquelles les méthodes de validations croisées (cross-validation en anglais) [117] ou encore les méthodes de sur-lissage (oversmoothing en anglais) [117].

Les méthodes locales peuvent offrir des avantages comparées aux méthodes globales, et cela d'autant plus que la dimension d de l'espace est importante pour certaines méthodes [130]. Dans le cas le plus général, la fenêtre $h^{(i)}$ ainsi que la matrice $\mathbf{A}^{(i)}$ sont des fonctions du point d'évaluation $\mathbf{x}$, de l'élément de l'échantillon $\mathbf{x}^{(i)}$, ainsi que de la densité de probabilité à estimer π, dont les informations disponibles sont contenues dans l'échantillon $(\mathbf{x}^{(j)})_{1\leq j\leq N}$, de

sorte que l'on peut écrire :

$$\begin{cases} h^{(i)} = h(\mathbf{x}, \mathbf{x}^{(i)}, \pi) \approx h(\mathbf{x}, \mathbf{x}^{(i)}, (\mathbf{x}^{(j)})_{1 \leq j \leq N}) \\ \mathbf{A}^{(i)} = \mathbf{A}(\mathbf{x}, \mathbf{x}^{(i)}, \pi) \approx \mathbf{A}(\mathbf{x}, \mathbf{x}^{(i)}, (\mathbf{x}^{(j)})_{1 \leq j \leq N}) \end{cases} \tag{5.40}$$

Le choix est fait ici de ne pas considérer de dépendance avec le point d'évaluation $\mathbf{x}$ *pour des raisons pratiques. Dans une telle situation, les fenêtres* $h^{(i)}$ *et les matrices* $A^{(i)}$ *peuvent donc être calculées à priori et stockées, ne nécessitant pas de nouveaux calculs à chaque évaluation. Un tel choix assure également que l'estimateur de la densité de probabilité est lui même bien une densité de probabilité, ce qui n'est pas le cas avec une dépendance en* $\mathbf{x}$ *[117]. Pour chaque élément de l'échantillon, il convient donc de définir le "voisinage" pris en compte autour du noyau, au travers du paramètre* $h^{(i)}$ *déterminant la taille de ce voisinage, ainsi qu'au travers de la matrice* $\mathbf{A}^{(i)}$ *permettant de déterminer la forme de ce "voisinage".*

5.3.1.1 Détermination de la taille d'un "bon" voisinage

Intuitivement, la taille du voisinage à considérer autour du i*-ème élément de l'échantillon, donnée par le paramètre* $h^{(i)}$*, doit être faible si l'échantillon* $\mathbf{x}^{(i)}$ *se trouve dans une zone présentant un pic de la densité à estimer, afin de ne pas trop lisser celui-ci, et plus importante dans une zone où la densité à estimer ne présente pas de pic afin d'éviter de faire apparaître un pic dans l'estimateur qui n'existe pas. Cette dernière affirmation se traduit en pratique par le fait que la taille du voisinage varie inversement avec la valeur de la densité de probabilité* π *que l'on cherche à estimer. Ainsi, il est proposé dans [5] l'expression* (5.41) *pour* $h^{(i)}$*, vérifiant bien la variation en sens inverse de la valeur de la densité de probabilité.*

$$h^{(i)} = \frac{h}{\sqrt{\pi(\mathbf{x}^{(i)})}} \tag{5.41}$$

Cette dernière formule n'est bien entendu pas utilisable en pratique, la densité de probabilité π *étant inconnue. Il est possible de l'évaluer à l'aide d'une première estimation de celle-ci [4], pour ensuite utiliser cette première estimation. Il est cependant possible de se passer d'une estimation préalable de la densité de probabilité en utilisant seulement les informations disponibles, à savoir l'ensemble des éléments de l'échantillon. Cela est réalisé au travers de la méthode des* k *plus proches voisins (*k*-NN pour* k *Nearest Neighbours en anglais) [77] qui consiste à choisir le paramètre* h *comme la distance, notée* $h_k^{(i)}$ *dans la suite, séparant l'élément* i *de l'échantillon de son* k*-ème plus proche voisin parmi les autres éléments de l'échantillon. La méthode des* k *plus proches voisins nécessite, pour donner un estimateur consistant, que le nombre* k *soit*

fonction du nombre d'échantillons N et vérifie les propriétés suivantes :

$$\begin{cases} \lim\limits_{N\to\infty} k = \infty \\ \lim\limits_{N\to\infty} \frac{k}{N} = 0 \end{cases} \tag{5.42}$$

La détermination du nombre k optimal est possible dans certains cas [36], comme par exemple dans le cas d'une loi normale en dimension d où k est donné par la relation (5.43) pour un échantillon de taille N impliquant la fonction Γ, mais pas dans le cas général, ce qui demande un choix judicieux de ce paramètre en pratique.

$$k = \left[\frac{(d+2)^2 (d-2)^{2+d/2}}{\Gamma^{4/d}\left(\frac{d+2}{2}\right) d^{1+d/2}(d^2-6d+16)} \right]^{\frac{d}{d+4}} N^{\frac{4}{4+d}} \tag{5.43}$$

L'utilisation de cette dernière méthode suppose de plus l'indépendance des vecteurs aléatoires dans l'échantillon. Dans cette thèse, les éléments de l'échantillon sont obtenus à partir d'une méthode de Quasi-Monte Carlo randomisé, cette méthode permettant une meilleur convergence pour l'estimation des statistiques nécessaires au calcul de l'expansion de Karhunen-Loève, ce qui remet en cause l'hypothèse d'indépendance des éléments de l'échantillon. Seul un récent papier dans la littérature [2] présente des résultats théoriques et pratiques quant à l'utilisation des méthodes à noyau pour l'estimation de densité avec un échantillonnage de Quasi-Monte Carlo randomisé. Dans ce papier est montré que l'échantillonnage ne change pas le terme lié au biais mais seulement celui lié à la variance dans l'erreur quadratique moyenne intégrée. Leurs résultats théoriques suggèrent que la réduction de ce terme de variance grâce à un échantillonnage de Quasi-Monte Carlo randomisé est d'autant meilleur que la dimension effective est faible. Leurs résultats pratiques présentent quant à eux une réduction significative de cette variance dans la majorité des cas et une faible augmentation de celle-ci dans de rares cas. L'ensemble de leurs résultats conforte donc l'utilisation faite dans cette thèse de l'estimation par noyau avec un échantillonnage issu d'une méthode de Quasi-Monte Carlo randomisé. La détermination des plus proches voisins dans un espace de dimension arbitraire est un problème algorithmique classique qui a été ici traité efficacement à l'aide d'un arbre $k-d$ [12].

L'idée de la taille d'un bon voisinage ayant été définie dans le cas d'une méthode locale, il convient maintenant de s'intéresser à la forme de celui-ci.

5.3.1.2 Détermination de la forme d'un "bon" voisinage

En dimension supérieur à 1, la paramétrisation du noyau devient plus complexe puisqu'il fait intervenir la matrice A, permettant de considérer différentes

formes pour le voisinage de l'échantillon. Ce voisinage aura en fait une forme ellipsoïdale dépendant de la matrice A, celui-ci étant sphérique si cette dernière matrice est la matrice identité. En fait, dans le cadre d'une paramétrisation globale, on peut montrer [117] que le biais asymptotique $AB(\mathbf{x})$ de l'estimateur en un point $\mathbf{x}$ est donné par l'expression (5.44), dans laquelle la matrice $\mathbf{S_x}$ correspond à la matrice hessienne de la densité de probabilité π au point $\mathbf{x}$.

$$AB(\mathbf{x}) = \frac{1}{2}h^2 \operatorname{Tr}\left(\mathbf{A}^T\mathbf{S_x}\mathbf{A}\right) \tag{5.44}$$

On peut alors montrer [130] qu'un choix judicieux de la matrice $\mathbf{A}$ permet de minimiser voire d'annuler ce biais asymptotique suivant que la matrice hessienne est définie positive ou négative, ou qu'elle ne l'est pas. Ce résultat théorique permet de montrer que le choix de la matrice $\mathbf{A}$ peut avoir un impact sur la qualité de l'estimation.

Dans le cas présent, l'objectif est de se tourner vers une méthode locale où les matrices $\mathbf{A}^{(i)}$ possèdent chacune leur propre valeur. Une idée intéressante présente dans [18] suggère de construire directement la matrice de covariance $\mathbf{\Sigma}^{(i)}$ à l'aide de l'expression (5.45), où K_h représente un noyau gaussien isotrope de fenêtre h fixée.

$$\mathbf{\Sigma}^{(i)} = \alpha\mathbf{I}_d + \sum_{j=1}^{N} K_h\left(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}\right)\left(\mathbf{x}^{(i)} - \mathbf{x}^{(j)}\right)\left(\mathbf{x}^{(i)} - \mathbf{x}^{(j)}\right)^T \tag{5.45}$$

Le second terme de l'expression permet d'orienter le voisinage en accord avec les autres éléments de l'échantillon, dont la répartition est dictée par la densité de probabilité à estimer, et cela d'autant plus que ceux-ci sont proches, la notion de proximité étant donnée par la fenêtre h du noyau gaussien isotrope K_h, alors que le premier terme est un terme de régularisation selon les auteurs. Dans la littérature, les paramètres α et h sont globaux, et une méthode d'optimisation est utilisée afin de déterminer des candidats intéressants. Leur optimisation les amène à relier α et h à l'aide de la relation $\alpha = h/5$ qui sera utilisée dans la suite, donnant pour $\mathbf{\Sigma}^{(i)}$ l'expression suivante :

$$\mathbf{\Sigma}^{(i)} = \frac{h\mathbf{I}_d}{5} + \sum_{j=1}^{N} K_h\left(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}\right)\left(\mathbf{x}^{(i)} - \mathbf{x}^{(j)}\right)\left(\mathbf{x}^{(i)} - \mathbf{x}^{(j)}\right)^T \tag{5.46}$$

Dans l'expression (5.46), le paramètre h est global et donc commun à l'ensemble des éléments de l'échantillon. Une construction locale de $\mathbf{\Sigma}^{(i)}$ peut également être envisagée, en considérant le paramètre h comme local à l'aide de la distance au k-ème voisin par exemple. La matrice de covariance $\mathbf{\Sigma}^{(i)}$ est alors

donnée par l'expression suivante :

$$\mathbf{\Sigma}^{(i)} = \frac{h_k^{(i)}\mathbf{I}_d}{5} + \sum_{j=1}^{N} K_{h_k^{(i)}}\left(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}\right)\left(\mathbf{x}^{(i)} - \mathbf{x}^{(j)}\right)\left(\mathbf{x}^{(i)} - \mathbf{x}^{(j)}\right)^T \tag{5.47}$$

Le choix de k va bien entendu avoir une influence sur la matrice de covariance $\mathbf{\Sigma}^{(i)}$*, et est dans le cas présent le seul paramètre à choisir.*

Ces méthodes adaptatives locales sont relativement coûteuses du fait qu'il est nécessaire pour le calcul de chaque élément de l'échantillon de prendre en compte l'ensemble des autres éléments éléments de l'échantillon, ce qui implique pour la construction de l'estimateur une complexité en N^2 *où* N *est la taille de l'échantillon utilisé pour la construction de l'estimateur.*

5.3.2 Comparaison de différentes méthodes

5.3.2.1 Critères de qualité utilisés

Les critères de qualité présentés précédemment, tel le $MISE$ *renseigne sur la convergence moyenne, à la fois spatiale et probabiliste, de l'estimateur de densité vers la vraie densité. Dans cette thèse, le recours à ces méthodes à noyau vise à être capable d'échantillonner suivant une loi inconnue lorsque l'on dispose d'une réalisation d'un échantillon de vecteurs aléatoires suivant cette loi inconnue. Il est donc intéressant de s'intéresser à un critère de qualité concernant l'échantillonnage réalisé à l'aide de la densité de probabilité estimée plutôt qu'un critère de qualité se concentrant sur la convergence de la densité de probabilité comme c'est le cas pour le* $MISE$*.*

On considère un échantillon $(\mathbf{X}_1, \ldots, \mathbf{X}_N)$ *de vecteurs aléatoires identiquement distribuées suivant une loi de densité* π *et pouvant posséder des dépendances entre elles afin de ne pas exclure un échantillonnage par méthode de Quasi-Monte Carlo randomisé. On construit à partir de cet échantillon un estimateur à noyau gaussien* $\hat{\pi}_N$ *de la densité* π*. Une fois la densité estimée* $\hat{\pi}_N$ *construite, celle-ci peut être utilisée pour l'échantillonnage comme présentée précédemment. Si on se donne un vecteur aléatoire* $\mathbf{U}$ *de loi uniforme sur* $[0,1]^d$*, il est alors possible de construire deux vecteurs aléatoires* $\mathbf{X}$ *et* $\hat{\mathbf{X}}$ *vérifiant la relation suivante,* T *correspondant à la transformation (3.12) :*

$$\begin{cases} \mathbf{X} = T_\pi(\mathbf{U}) \\ \hat{\mathbf{X}} = T_{\hat{\pi}_N}(\mathbf{U}) \end{cases} \tag{5.48}$$

De même, en se donnant un vecteur aléatoire $\mathbf{X}$ *suivant la loi donnée par la densité* π*, il est possible de construire deux vecteurs aléatoires* $\mathbf{U}$ *et* $\hat{\mathbf{U}}$ *vérifiant*

la relation suivante :

$$\begin{cases} \mathbf{U} = T_{\pi}^{-1}(\mathbf{X}) \\ \hat{\mathbf{U}} = T_{\hat{\pi}_N}^{-1}(\mathbf{X}) \end{cases} \tag{5.49}$$

La qualité de l'estimateur est ici mesurée de deux façons, chacune correspondant à un des deux usages précédents de la densité de probabilité estimée.

La première façon consiste en l'erreur quadratique moyenne $Err_{\mathbf{X}}$ *suivante :*

$$Err_{\mathbf{X}} = E_{(\mathbf{X}_1,\ldots,\mathbf{X}_N)}\left[E_{\mathbf{U}}\left[(T_{\hat{\pi}_N}(\mathbf{U}) - T_{\pi}(\mathbf{U}))^2\right]\right] \tag{5.50}$$

Dans cette dernière expression, $E_{(\mathbf{X}_1,\ldots,\mathbf{X}_N)}$ *correspond à une espérance sur l'échantillon* $(\mathbf{X}_1,\ldots,\mathbf{X}_N)$ *alors que* $E_{\mathbf{U}}$ *correspond à une espérance sur le vecteur aléatoire* $\mathbf{U}$*. Le vecteur aléatoire* $\mathbf{U}$ *et l'échantillon* $(\mathbf{X}_1,\ldots,\mathbf{X}_N)$ *sont bien entendu considérés indépendants dans ce qui précède. En pratique, cette erreur quadratique moyenne peut être estimée à l'aide d'une méthode de Monte Carlo pour le calcul de ces deux espérances, ce qui revient à estimer* $Err_{\mathbf{X}}$ *de la façon suivante :*

$$Err_X \approx \frac{1}{Q}\sum_{i=1}^{Q}\frac{1}{R}\sum_{j=1}^{R}\left(T_{\hat{\pi}_N^{(i)}}(\mathbf{u}^{(\mathbf{j})}) - T_{\pi}(\mathbf{u}^{(\mathbf{j})})\right)^2 \tag{5.51}$$

Dans cette dernière expression, les $\hat{\pi}_N^{(i)}$ *sont des densités de probabilités estimées à l'aide de réalisations* $(\mathbf{x}_1^{(i)},\ldots,\mathbf{x}_N^{(i)})$ *de l'échantillon* $(\mathbf{X}_1,\ldots,\mathbf{X}_N)$ *indépendantes entre elles, alors que les* $\mathbf{u}^{(j)}$ *correspondent à une réalisation d'un échantillon* $(\mathbf{U}^{(1)},\ldots,\mathbf{U}^{(R)})$ *de vecteurs aléatoires indépendants et uniformément distribuées sur* $[0,1]^d$*.*

La seconde façon d'évaluer la qualité de l'estimation consiste en le calcul de l'erreur quadratique $Err_{\mathbf{U}}$ *donnée par l'expression suivante :*

$$Err_{\mathbf{U}} = E_{(\mathbf{X}_1,\ldots,\mathbf{X}_N)}\left[E_{\mathbf{X}}\left[(T_{\hat{\pi}_N}^{-1}(\mathbf{X}) - T_{\pi}^{-1}(\mathbf{X}))^2\right]\right] \tag{5.52}$$

Cette dernière expression est similaire à l'expression (5.50), une des deux espérances étant cependant cette fois calculée suivant le vecteur aléatoire $\mathbf{X}$ *plutôt que sur le vecteur aléatoire* $\mathbf{U}$*. Il est à nouveau possible d'estimer cette dernière erreur à l'aide d'une méthode de Monte Carlo, ce qui donne pour*

l'estimation l'expression suivante :

$$Err_{\mathbf{U}} \approx \frac{1}{Q}\sum_{i=1}^{Q}\frac{1}{R}\sum_{j=1}^{R}\left(T^{-1}_{\hat{\pi}^{(i)}_N}(\mathbf{x^{(j)}}) - T^{-1}_{\pi}(\mathbf{x^{(j)}})\right)^2 \tag{5.53}$$

Cette dernière expression est similaire à l'expression (5.51)*,* $(\mathbf{x^{(1)}}, \ldots, \mathbf{x^{(R)}})$ *correspondant à une réalisation d'une famille de vecteurs aléatoires indépendants entre eux et distribués suivant la loi de* $\mathbf{X}$*.*

Les deux erreurs ainsi que la manière de les estimer étant définies, il convient de définir les méthodes qui seront comparées.

5.3.2.2 Méthodes comparées

La construction du noyau de chaque échantillon peut se faire en utilisant des paramètres globaux ou locaux. Afin de se faire une idée de la meilleure stratégie à adopter, différentes méthodes vont être comparées utilisant toutes un noyau gaussien.

5.3.2.2.1 Estimateur non orienté avec fenêtre globale (NOG) *Le premier estimateur considéré est un estimateur pour lequel la fenêtre* h *est globale. La valeur de la fenêtre* h *donnée par l'expression* (5.54) *qui correspond à la valeur optimale de la fenêtre pour l'estimation de la densité de probabilité d'un vecteur aléatoire gaussien [117].*

$$h = \frac{1}{\sqrt{d}}\left[\frac{4}{(2+d)N}\right]^{\frac{1}{4+d}} \tag{5.54}$$

Les noyaux considérés pour cet estimateur, bien que non orientés, ne sont pas isotropes résultant en une matrice $\mathbf{A}$*, elle aussi globale, différente de l'identité, mais tout de même diagonale, avec les écart-types* σ_i *des composantes du vecteur aléatoire à estimer placés sur cette diagonale. De fait, si une grande variance est présente selon un axe de coordonnée, le noyau présentera lui aussi une grande variance suivant cet axe de coordonnée. Le noyau résultant d'une telle paramétrisation avec une matrice* $\mathbf{A}$ *est parfois appelé noyau produit, car pouvant se réécrire comme le produit de noyaux gaussiens unidimensionnels.*

5.3.2.2.2 Estimateur non orienté utilisant les plus proches voisins (NOL) *Le second estimateur considéré est un estimateur pour lequel la fenêtre* h *est locale, et est la distance au* k*-ième plus proche voisin. Comme dit précédemment, il existe un choix optimal pour cette valeur de* k *pour la minimisation de l'AMISE, donné dans le cas où la densité à estimer est celle d'un vecteur aléatoire gaussien par l'expression* (5.43)*. Dans la présente étude, pour*

les critères de qualité utilisés et pour les tailles d'échantillon N investiguées, il s'avère qu'un choix de k égal à 1 permet la minimisation de $Err_{\mathbf{X}}$, tout en étant pas trop éloigné de la valeur optimale de k pour $Err_{\mathbf{U}}$ comme visible sur la figure 5.5. Dans la suite, la valeur de $k = 1$ sera donc choisie pour l'ensemble des comparaisons pour cette méthode.

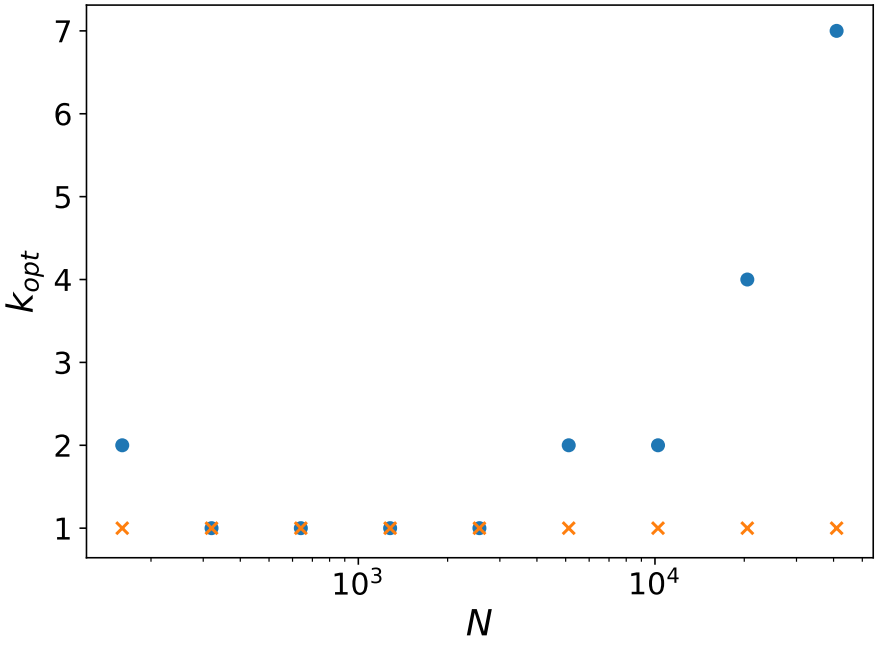

FIGURE 5.5 – *Valeurs de k optimales trouvées numériquement pour l'estimateur NOL en fonction de la taille de l'échantillon N pour l'erreur $Err_{\mathbf{U}}$ (ronds) et pour l'erreur $Err_{\mathbf{X}}$ (croix).*

Tout comme pour l'estimateur précédent, les noyaux considérés pour cet estimateur sont également des noyaux produits, la matrice $\mathbf{A}$ étant globale et identique à celle de l'estimateur précédent.

5.3.2.2.3 Estimateur orienté avec fenêtre globale (OG) *Le troisième estimateur considéré est un estimateur pour lequel la paramétrisation de chacun des noyaux est donnée par l'expression (5.46), la valeur de h dans cette expression étant la valeur donnée par l'expression (5.54) utilisée pour le premier estimateur considéré.*

5.3.2.2.4 Estimateur orienté avec fenêtre utilisant les plus proches voisins (OL) *Le quatrième estimateur considéré est un estimateur pour lequel la paramétrisation de chacun des noyaux est donnée par l'expression (5.47). Les investigations numériques ont permis de choisir la valeur $k = 1$ pour cet*

estimateur, les deux erreurs étant minimales avec cette valeurs de k, comme visible sur la figure 5.6. Les valeurs optimales de k trouvées numériquement et leur évolution avec la taille N de l'échantillon ne semble pas

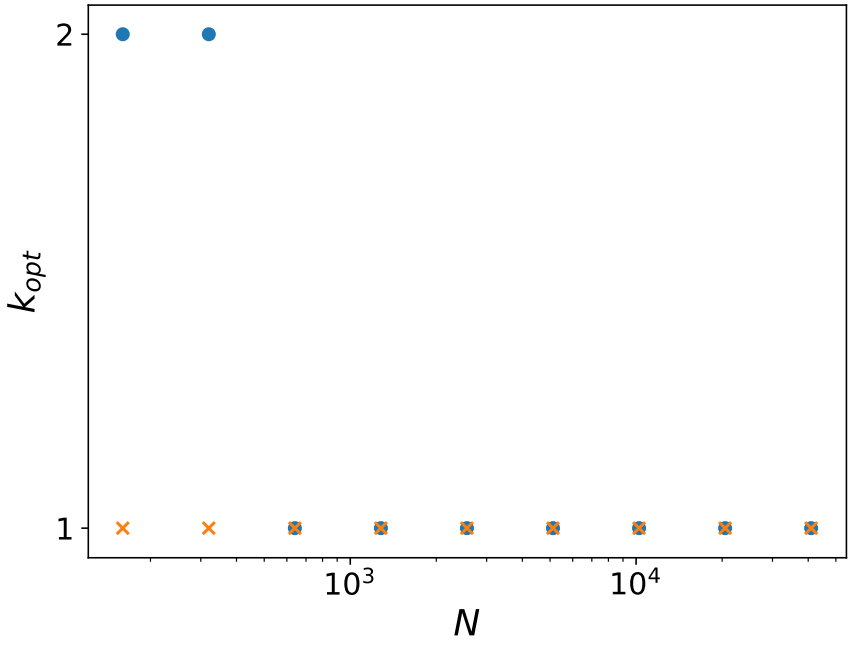

FIGURE 5.6 – *Valeurs de k optimales trouvées numériquement pour l'estimateur OL en fonction de la taille de l'échantillon N pour l'erreur $Err_{\mathbf{U}}$ (ronds) et pour l'erreur $Err_{\mathbf{X}}$ (croix).*

5.3.2.3 Résultats des comparaisons

Les méthodes précédentes sont comparées à l'aide d'une loi de probabilité qui est une mixture de 25 gaussiennes ayant chacune le même poids, et étant telle qu'elles sont centrées en un point de l'hypercube $[-1, 1]^d$ avec $d = 4$ choisi aléatoirement indépendamment et uniformément sur celui-ci, et la matrice de covariance $\mathbf{\Sigma}$ est construite à partir de sa décomposition spectrale $\mathbf{\Sigma} = \mathbf{ODO}^T$, $\mathbf{D}$ et $\mathbf{O}$ étant construit de la manière suivante :

- *les coefficients diagonaux de $\mathbf{D}$ sont choisis aléatoirement et indépendamment de manière uniforme dans $[0, 1]$*
- *la matrice O est issue de l'orthogonalisation de d vecteurs dont les composantes sont tirées aléatoirement et indépendamment de manière uniforme $[0, 1]$*

Pour chacune des méthodes présentées, une réalisation de l'estimateur est construite sur la base d'une réalisation d'un échantillon de taille totale N obtenu à l'aide de 10 séquences de Sobol randomisées. Concernant l'estimation des

erreurs $Err_{\mathbf{U}}$ *et* $Err_{\mathbf{X}}$*, en reprenant les notations des expressions* (5.51) *et* (5.53)*, les valeurs de* $Q = 100$ *et de* $R = 1000$ *ont été choisies afin d'avoir des résultats relativement convergés pour un coût de calcul restant raisonnable.*

Sur la figure 5.7 sont visibles les évolutions des erreurs $Err_{\mathbf{U}}$ *et* $Err_{\mathbf{X}}$ *pour les différentes méthodes en fonction de la taille* N *de l'échantillon utilisé. L'augmentation de la taille* N *de l'échantillon utilisé permet dans l'ensemble des cas une diminution de l'erreur comme attendu. Les deux méthodes utilisant des noyaux orientés présentent une diminution de l'erreur avec* N *plus importantes que les méthodes possédant des noyaux parallèles aux axes, justifiant en pratique l'intérêt de considérer des noyaux pouvant s'adapter à la distribution de l'échantillon. Parmi les méthodes utilisant des noyaux orientés, la méthode OL présente une diminution de l'erreur plus importante que la méthode OG pour les deux erreurs. Cependant, elle ne surpasse la méthode OG que pour l'erreur* $Err_{\mathbf{U}}$ *pour les tailles d'échantillons investiguées, la méthode OG présentant une erreur* $Err_{\mathbf{X}}$ *plus faible pour les tailles* N *d'échantillons faibles.*

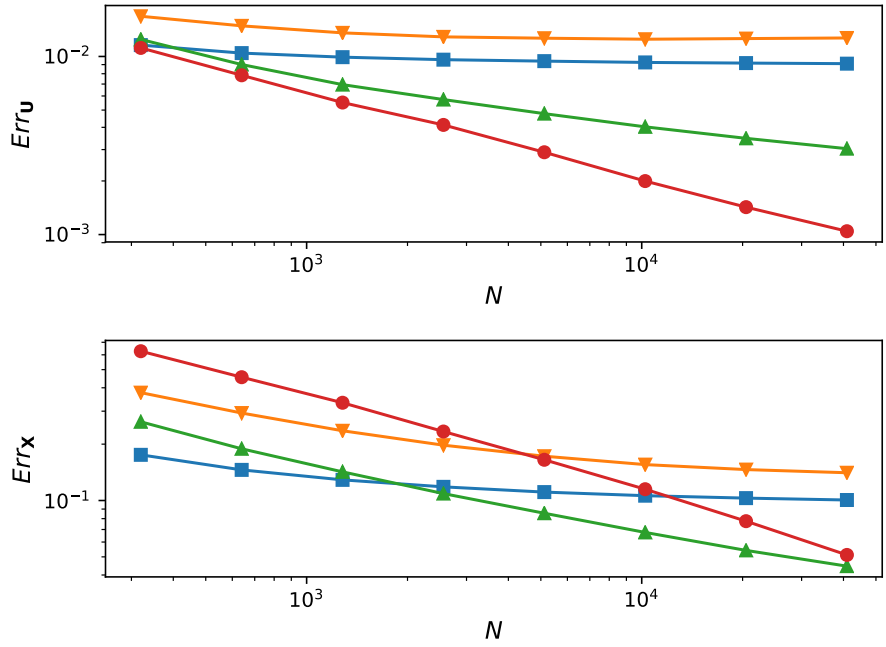

FIGURE 5.7 – *Évolutions des erreurs* $Err_{\mathbf{U}}$ *et* $Err_{\mathbf{X}}$ *en fonction de la taille* N *d'échantillon utilisée pour les différentes méthodes comparées. Les carrés correspondent à la méthode NOG, les triangles bas à la méthode NOL, les triangles hauts à la méthode OG et les ronds à la méthodes OL.*

Les erreurs $Err_{\mathbf{U}}$ *et* $Err_{\mathbf{X}}$ *donnent une information concernant la reproduction de l'ensemble des composantes des vecteurs aléatoires. Il peut également être intéressant de regarder la reproduction de chacune des composantes séparément. Cela peut être réalisé à l'aide des mêmes métriques* $Err_{\mathbf{U}}$ *et* $Err_{\mathbf{X}}$ *qui*

cette fois-ci sont appliquées à chacune des composantes séparément.

Sur la figure 5.8 sont représentées les racines carrées des erreurs Err_U pour les 4 composantes du vecteur aléatoire étudié et pour les différentes méthodes comparées. Pour l'ensemble des méthodes, la composante 1 est celle présentant l'erreur la plus faible, suivie de la seconde, de la troisième et de la quatrième. Cela est directement lié au fait que la génération des composantes dépend des composantes précédentes avec les choix faits précédemment, ce qui entraîne une accumulation des erreurs sur les composantes générées en dernière. Cette augmentation de l'erreur est bien plus présente pour les méthodes utilisant des noyaux non orientés que pour les autres, pour lesquelles la dégradation de l'erreur sur les dernières composantes est moins marquée. Dans le cas d'une modélisation des nouvelles variables aléatoires d'une expansion de Karhunen-Loève, il est à priori préférable que les premières composantes soient mieux reproduites que les suivantes, étant celles ayant à priori le plus d'impacts. Le fait d'avoir pris la racine carré de l'erreur Err_U permet de se ramener à une erreur qui a la même dimension que les composantes de $\mathbf{U}$ qui évoluent dans l'intervalle $[0, 1]$, ce qui permet de conclure que l'erreur commise pour les deux premières composantes est en moyenne de quelques millièmes, alors qu'elle est de quelques centièmes sur les deux dernières composantes.

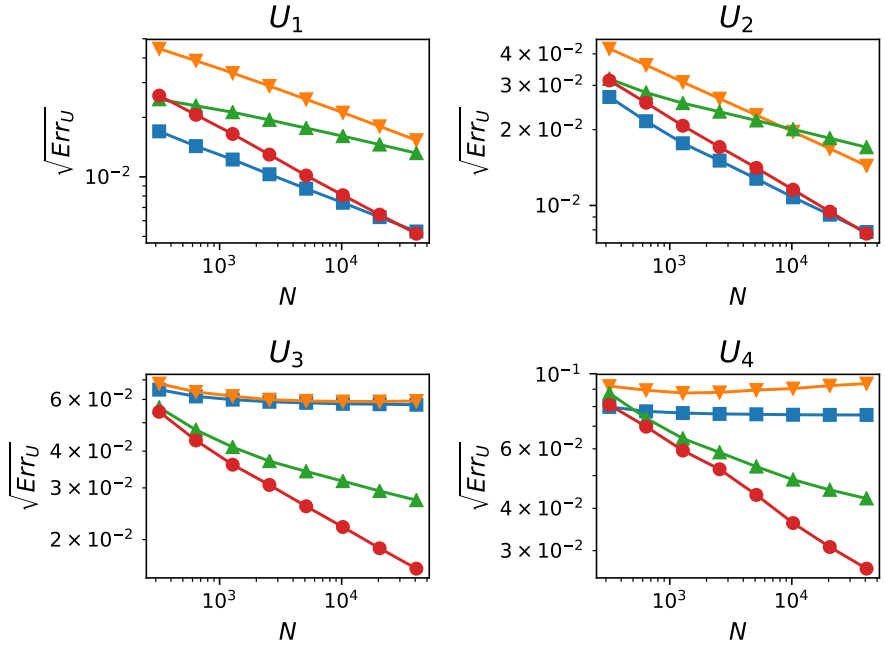

FIGURE 5.8 – *Évolutions des racines carrés des erreurs Err_U pour chacune des composantes en fonction de la taille N d'échantillon utilisée pour les différentes méthodes comparées. Les carrés correspondent à la méthode NOG, les triangles bas à la méthode NOL, les triangles hauts à la méthode OG et les ronds à la méthodes OL.*

L'étude par composante est également réalisée pour l'erreur Err_X, comme présenté sur la figure 5.9 où sont cette fois représentées les racines carrés des erreurs Err_X sur chacune des composantes divisées par l'écart-type de la composante considérée. Les mêmes remarques que pour l'erreur Err_U peuvent être faites concernant la diminution de l'erreur des différentes composantes pour les différentes méthodes. Avec les méthodes orientées qui se comportent mieux sur l'ensemble des composantes, l'erreur moyenne commise se trouve être de l'ordre de 10% de l'écart-type de la composante considérée avec des échantillons de taille 40960 pour l'exemple considéré. La méthode NOG offre des erreurs plus faibles pour les premières composantes, qui sont de l'ordre de 1% impliquant que cette méthode n'est pas à exclure surtout dans les cas où le problème étudié est surtout sensible aux premières composantes du vecteur aléatoire à modéliser.

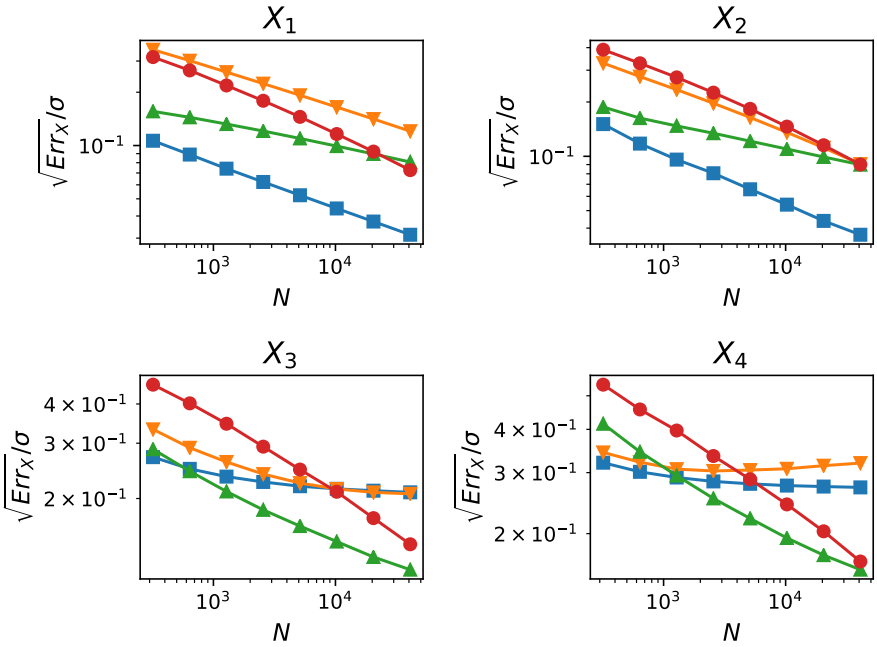

FIGURE 5.9 – *Évolutions des racines carrés des erreurs Err_X divisé par les écart-types de la composante impliquée pour chacune des composantes en fonction de la taille N d'échantillon utilisée pour les différentes méthodes comparées. Les carrés correspondent à la méthode NOG, les triangles bas à la méthode NOL, les triangles hauts à la méthode OG et les ronds à la méthodes OL.*

5.4 Conclusion

Ce chapitre a permis de présenter rapidement les méthodes à noyaux permettant l'estimation de la densité de probabilité d'un vecteur aléatoire à partir

d'échantillons de celui-ci. Ces méthodes sont raisonnables à utiliser lorsque la dimension stochastique reste relativement faible, ce qui convient parfaitement au cadre de cette thèse où l'intérêt est la réduction de la dimension stochastique d'un système pour ne garder que 4 *ou* 5 *paramètres incertains au plus. L'utilisation de noyaux gaussiens pour l'estimation de la densité de probabilité du vecteur aléatoire permet une utilisation directe de la méthode d'inversion généralisée afin de générer des échantillons de ce vecteur aléatoire.*

Les méthodes à noyaux sont encore aujourd'hui un domaine de recherche actif offrant une grande richesse et une grande diversité de résultats dans la littérature. Ce chapitre n'a pas pour but d'être exhaustif sur l'état de l'art de ces méthodes, mais de présenter rapidement certains des concepts liés à ces méthodes, notamment pour les améliorer. Les résultats pratiques présentés reposaient sur une expérience numérique qui n'est pas forcément représentative de ce qui sera rencontré dans les études envisagées. Néanmoins, les conclusions de cette étude suggèrent que la stratégie à envisager dépend du besoin.

Dans la suite de cette thèse, deux usages majeurs seront fait de l'estimation de densité de probabilité par des méthodes à noyaux. La première concerne la génération d'échantillons du vecteur aléatoire choisi comme nouveau paramétrage stochastique du problème, pour lequel seuls des échantillons de ce vecteur aléatoire seront disponibles. Pour la propagation d'incertitudes, notamment à l'aide de méthodes de cubatures, il sera nécessaire d'être capable de transformer les points de cubatures de l'hypercube $[0,1]^d$ *en des points de* $\mathbb{R}^d$ *suivant la loi de ce vecteur aléatoire. La méthode NOG semble la plus adaptée à cette usage, surtout dans le cas où la reproduction des premières composantes est plus importantes que les autres. Si tel n'est pas le cas, l'utilisation de la méthode OL semble la plus appropriée.*

La seconde utilisation sera la transformation des échantillons du vecteur aléatoire en des échantillons d'un vecteur aléatoire théoriquement de loi uniforme sur $[0,1]^d$ *qui sera alors le nouveau vecteur aléatoire servant à la paramétrisation stochastique du système. Il peut être plus intéressant de se ramener à un tel vecteur dans certains cas qui seront détaillés dans la suite. L'expérience numérique réalisée ici suggère que la méthode OL semble la plus adaptée pour cet usage.*

Deuxième partie

Propagation d'incertitudes utilisant la chimie tabulée

Chapitre 6

Tabulation de cinétique chimique incertaine

Prérequis :

- Calcul d'intégrales multiples (Cubature, Monte Carlo et Quasi-Monte Carlo randomisé)

Notions clés et apports du chapitre :

- Description des incertitudes d'un système chimique canonique (réacteur adiabatique homogène à pression constante) en présence de paramètres de cinétique chimique incertains
- Description des incertitudes des grandeurs tabulées d'un système chimique canonique (réacteur adiabatique homogène à pression constante) en présence de paramètres de cinétique chimique incertains
- Proposition de caractériser les incertitudes des grandeurs tabulées actives grâce aux incertitudes d'une variable d'avancement de la réaction uniquement
- Proposition pour retrouver la moyenne et la variance de l'ensemble des grandeurs tabulées dans un calcul en ajoutant les variances des grandeurs tabulées dans la table chimique

Cette thèse s'inscrit dans le développement de méthodes permettant de

propager les incertitudes présentes dans la cinétique chimique au sein de simulations aux grandes échelles d'écoulements réactifs, et cela via l'utilisation de méthodes non-intrusives afin de pouvoir utiliser les méthodes et les codes déjà existants. Plusieurs méthodes décrites dans le chapitre 1 permettent de réaliser des simulations aux grandes échelles d'écoulements réactifs, et seules les méthodes de chimie tabulée sont explorées dans cette thèse.

Une méthode est déjà présente dans la littérature permettant de réaliser une telle propagation dans le cas où la combustion est modélisée à l'aide d'un modèle de flammelette stationnaire et qui a été discuté dans le chapitre 1. Il existe cependant des cas où un tel modèle de flammelette est inapproprié, et où d'autres modèles sont nécessaires. Les modélisations de type FPI [43] sont l'objet d'étude de cette thèse, pour lesquelles l'évolution d'une variable de progrès permettant de caractériser l'avancement du processus de combustion est centrale.

Ce chapitre se focalise sur une construction de la table chimique permettant la propagation d'incertitudes de la cinétique chimique au sein d'une simulation aux grandes échelles, et s'attache à définir les conditions dans lesquelles l'utilisation de la table proposée permet une bonne propagation des incertitudes. L'étude menée s'est intéressée uniquement à une configuration de réacteur homogène adiabatique isobare, qui a déjà été utilisé afin de modéliser la combustion dans une simulation aux grandes échelles [37]. La caractérisation d'un tel système chimique est réalisée dans la section suivante, pour un mélange d'air-hydrogène, à la fois pour une cinétique chimique déterministe et pour une cinétique chimique incertaine.

6.1 Caractérisation de la configuration chimique 0D étudiée

Le système chimique $0D$ étudié correspond à un réacteur homogène adiabatique isobare contenant un mélange d'air et d'hydrogène. Le mécanisme réactionnel utilisé pour décrire la cinétique chimique est le mécanisme de Konnov [63], qui offre l'avantage d'avoir les facteurs d'incertitudes documentés pour chacune des réactions élémentaires. Les informations concernant le mécanisme de Konnov sont présentées sur la figure 6.1.

Un tel système évolue spontanément de son état initial vers son état d'équilibre thermodynamique, et cela plus ou moins vite en fonction de l'état thermodynamique initial, mais également de la cinétique chimique. Cette évolution spontanée vers l'état d'équilibre thermodynamique implique physiquement un processus de combustion, qui est qualifié d'auto-allumage, le système considéré étant fermé. La caractérisation de ce système chimique est faite dans les cas d'une cinétique chimique classique ainsi que dans le cas d'une cinétique chimique incertaine.

H/O kinetic mechanism: units are cm^3 mol s cal K, $k = AT^n \exp(-E_a/RT)$, UF = uncertainty factor

No.	Reaction	A	n	E_a	Temperatures	UF	Source
1a	$H + H + M = H_2 + M$[a]	7.00E+17	−1.0	0	77–5000	2	[33][d]
	Enhanced third-body efficiencies (relative to Ar):						
	$H_2 = 0$, $N_2 = 0$, H = 0, $H_2O = 14.3$				300–2000	5	[33][d]
1b	$H + H + H_2 = H_2 + H_2$	1.00E+17	−0.6	0	50–5000	2.5	[33][d]
1c	$H + H + N_2 = H_2 + N_2$	5.40E+18	−1.3	0	77–2000	3.2	[33][d]
1d	$H + H + H = H_2 + H$	3.20E+15	0	0	50–5000	3.2	[33][d]
2	$O + O + M = O_2 + M$[a]	1.00E+17	−1.0	0	300–5000	2	[73][d]
	Enhanced third-body efficiencies (relative to Ar):						
	$O = 28.8$, $O_2 = 8$, NO = 2, N = 2, $N_2 = 2$,					2	[3,106][d]
	$H_2O = 5$					3	[107][e]
3	$O + H + M = OH + M$[a]	6.75E+18	−1.0	0	2950–3700	3	[36][f]
	Enhanced third-body efficiency:						
	$H_2O = 5$						[104][e]
4a	$H_2O + M = H + OH + M$[a]	6.06E+27	−3.312	120,770	300–3400	2	[35][f]
	Enhanced third-body efficiencies (relative to Ar):						[35][f]
	$H_2O = 0$, $H_2 = 3$, $N_2 = 2$, $O_2 = 1.5$						
4b	$H_2O + H_2O = H + OH + H_2O$	1.00E+26	−2.44	120,160	300–3400	2	[35][f]
5a	$H + O_2 (+M) = HO_2 (+M)$[a,b]	4.66E+12	0.44	0	300–2000	1.2	[38][d,g]
	Low-pressure limit:	5.70E+19	−1.4	0	300–2000		[38][d,g]
	$F_{cent} = 0.5$						
	Enhanced third-body efficiencies (relative to N_2):						
	Ar = 0, $H_2O = 0$, $O_2 = 0$, $H_2 = 1.5$, He = 0.57						[42][d,g,f]
5b	$H + O_2 (+Ar) = HO_2 (+Ar)$[b]	4.66E+12	0.44	0	300–2000	1.2	[38][d,g]
	Low-pressure limit:	7.43E+18	−1.2	0	300–2000		[38][d,g]
	$F_{cent} = 0.5$						
5c	$H + O_2 (+O_2) = HO_2 (+O_2)$[b]	4.66E+12	0.44	0	300–2000	1.2	[38][d,g]
	Low-pressure limit:	5.69E+18	−1.094	0	300–700	1.3	[42][d,g,f]
	$F_{cent} = 0.5$						
5d	$H + O_2 (+H_2O) = HO_2 (+H_2O)$[b]	9.06E+12	0.2	0	1050–1250	1.4	[40][d,g,f]
	Low-pressure limit:	3.67E+19	−1.0	0	1050–1250		[40][d,g,f]
	$F_{cent} = 0.8$						
6a	$OH + OH (+M) = H_2O_2 (+M)$[a,b]	1.00E+14	−0.37	0	200–1500	2.5	see text
	Low-pressure limit:	2.38E+19	−0.8	0	250–1400	2.5	[12][d]
	$F_{cent} = 0.5$						
	Enhanced third-body efficiency:						
	$H_2O = 0$						
6b	$OH + OH (+H_2O) = H_2O_2 (+H_2O)$[b]	1.00E+14	−0.37	0	200–1500	2.5	see text
	Low-pressure limit:	1.45E+18	0	0	300–400	2.5	[12,46][d]
	$F_{cent} = 0.5$						
7	$O + H_2 = OH + H$	5.06E+04	2.67	6290	297–2495	1.3	[47][d,f]
8	$H + O_2 = OH + O$	2.06E+14	−0.097	15,022	800–3500	1.5	[12][d]
9	$H_2 + OH = H_2O + H$	2.14E+08	1.52	3450	300–2500	2	[12][d]
10	$OH + OH = H_2O + O$	3.34E+04	2.42	−1930	250–2400	1.5	[12][d]
11	$HO_2 + O = OH + O_2$	1.63E+13	0	−445	220–400	1.2	[62][d]
12	$H + HO_2 = OH + OH$	1.90E+14	0	875	300–1000	2	see text
13	$H + HO_2 = H_2O + O$	1.45E+12	0	0	300	3	[12][d]
14	$H + HO_2 = H_2 + O_2$	1.05E+14	0	2047	250–1000	2	[12][d]
15	$H_2 + O_2 = OH + OH$	2.04E+12	0.44	69,155	298–1000	3	[65][g]
16	$HO_2 + OH = H_2O + O_2$[c]	2.89E+13	0	−500	250–2000	3	[12][d], see text
		+ 9.27E+15	0	17,500		3	[12][d]
17a	$HO_2 + HO_2 = H_2O_2 + O_2$[c]	1.03E+14	0	11,040	300–1250	2.5	[45][f]
		+ 1.94E+11	0	−1409		1.4	
17b	$HO_2 + HO_2 + M = H_2O_2 + O_2 + M$	6.84E+14	0	−1950	230–420	1.4	[26][d]
18	$H_2O_2 + H = HO_2 + H_2$	1.70E+12	0	3755	300–1000	3	[12][d]
19	$H_2O_2 + H = H_2O + OH$	1.00E+13	0	3575	300–1000	2	[12][d]
20	$H_2O_2 + O = HO_2 + OH$	9.55E+6	2	3970	300–2500	3	[52][d]
21	$H_2O_2 + OH = HO_2 + H_2O$[c]	2.00E+12	0	427	240–1700	2	[68][f]
		+ 1.70E+18	0	29,400			

[a] All other species have efficiencies equal to unity.
[b] The fall-off behavior of this reaction is expressed in the form as used by Baulch et al. [12] and others.
[c] Rate constant is the sum of two expressions.
[d] Review.
[e] Estimate.
[f] Measurements.
[g] Theoretical calculations.

FIGURE 6.1 – *Mécanisme de Konnov [63].*

6.1.1 Cas d'une cinétique chimique déterministe

Le comportement d'un réacteur homogène adiabatique isobare peut être décrit à l'aide du système d'équations différentielles ordinaires suivant :

$$\begin{cases} \forall 1 \leq k \leq N_s, \dfrac{dY_k}{dt}(t) = \dot{\omega}_k(t) \\ \dfrac{dh}{dt}(t) = 0 \\ \dfrac{dP}{dt}(t) = 0 \end{cases} \tag{6.1}$$

La première ligne du système correspond à la conservation de la masse et des éléments chimiques. Elle comporte N_s équations décrivant l'évolution temporelle de la composition chimique du système, comportant N_s espèces chimiques différentes. La composition chimique du système est ici caractérisée par les fractions massiques Y_k des espèces chimiques, et l'évolution de celles-ci est donnée par leurs termes sources chimiques respectifs $\dot{\omega}_k$, qui dépendent directement de la cinétique chimique du mécanisme réactionnel. La seconde ligne correspond quant à elle à la conservation de l'enthalpie totale du système, qui se conserve car le système est adiabatique et à pression constante, et la dernière ligne traduit le fait que le système évolue à pression constante. Ce système d'équations différentielles ordinaires est couplé. En effet, les termes sources des différentes espèces chimiques dépendent de l'état thermodynamique du système, donc de l'ensemble des Y_k, mais aussi de la pression et de l'enthalpie au travers de la température. De même, l'enthalpie et la pression dépendront elles aussi de l'état thermodynamique du système. Un tel système d'équations différentielles ordinaires est de plus souvent raide, provenant du fait que des temps chimiques différents existent au sein d'un tel système [80]. La raideur du système d'équations différentielles ordinaires nécessite l'utilisation de solveurs numériques robustes et efficaces.

6.1.1.1 Résolution numérique de l'ODE

Compte tenu de la conservation de la pression et de l'enthalpie, le système d'équations différentielles ordinaires à résoudre peut se ramener au système de N_s équations différentielles ordinaires (6.2), qui est le système effectivement résolu numériquement.

$$\begin{cases} \forall 1 \leq k \leq N_s, \dfrac{dY_k}{dt}(t) = \dot{\omega}_k(t) \\ h(t) = h(0) \\ P(t) = P(0) \end{cases} \tag{6.2}$$

Le solveur d'équations différentielles ordinaires utilisé ici est le solveur RADAU5 [47], qui utilise une méthode de Runge-Kutta implicite (Radau IIA) d'ordre 5 avec une adaptation automatique du pas de temps, et permettant la résolution efficace de systèmes d'équations différentielles ordinaires raides.

La précision du résultat obtenu à l'aide d'un tel solveur se fait par la spécification de tolérances, l'une, tol_{rel}, étant dénommée relative et l'autre, tol_{abs}, étant dénommée absolue. Le solveur RADAU5 permet la spécification de tolérances relatives et absolues propres à chaque ligne du système d'équations différentielles ordinaires, mais cette fonctionnalité n'est pas utilisée, la tolérance relative et la tolérance absolue étant utilisées pour l'ensemble des lignes du système. En considérant la résolution du système (6.2), ces tolérances assurent "grossièrement" que l'erreur locale commise par l'estimation $Y_k^{est}(t)$ de $Y_k(t)$ vérifie la relation suivante :

$$|Y_k^{est}(t) - Y_k(t)| \leq tol_{abs} + tol_{rel}|Y_k^{est}(t)| \tag{6.3}$$

La tolérance relative permet d'assurer que l'erreur relative commise sur $Y_k(t)$ est grossièrement inférieure à cette tolérance, lorsque la valeur de $Y_k(t)$ est grande devant la valeur de la tolérance absolue tol_{abs}, plus précisément lorsque $tol_{rel}|Y_k^{est}(t)| > tol_{abs}$. Pour de telles valeurs de $Y_k(t)$, le nombre de chiffres significatifs attendus de la solution est grossièrement de l'ordre de $\log_{10}(tol_{rel})$. Dans le cas du solveur RADAU5, la valeur minimale pour tol_{rel} est de 10 fois la valeur retournée par la fonction $EPSILON$ de Fortran, de sorte que le nombre de chiffres significatifs maximum pouvant être demandé est environ 14 en double précision. La tolérance absolue permet de s'affranchir d'une bonne précision pour les valeurs de $Y_k(t)$ qui sont jugées négligeables par l'utilisateur pour l'application étudiée. Ainsi, si $Y_k(t) \leq tol_{abs}$, la valeur estimée $Y_k^{est}(t)$ peut ne posséder aucun chiffre significatif. Cette valeur pour le solveur RADAU5 doit simplement être positive, et la valeur minimale possible est donc celle renvoyée par la fonction $TINY$ de Fortran, renvoyant le plus petit nombre positif représentable dans l'arithmétique flottante utilisée. La spécification de ces deux tolérances permet donc de jouer sur la précision du résultat obtenu, mais impacte également le temps de calcul nécessaire à la résolution du système (6.2). En effet, l'obtention d'une plus grande précision se fait par l'utilisation de pas de temps plus faibles, ce qui implique un nombre d'itérations plus important et donc un coût de calcul plus important.

L'impact du choix des tolérances peut être étudié numériquement sur le système d'équations différentielles ordinaires (6.2). La solution de référence au système d'équations différentielles ordinaires (6.2) est obtenue en considérant les valeurs minimales possibles pour les tolérances relatives et absolues, qui ont été précédemment explicitées. Afin de s'assurer de la robustesse du solveur RADAU5 dans les cas pratiques qui peuvent être rencontrés, l'état initial du système est choisi aléatoirement en imposant une température initiale entre

700 K et 1200 K et une richesse initiale entre 0.2 et 2.0, le système étant considéré à pression atmosphérique. La cinétique chimique est de plus perturbée aléatoirement comme mentionnée plus tard dans ce chapitre. Un paramètre important du processus d'auto-allumage est le délai d'auto-allumage, correspondant à la durée nécessaire pour que le système s'auto-allume. L'étude de l'impact du choix des tolérances est effectuée sur le délai d'auto-allumage $\tau^{c=0.1}$, qui sera défini ultérieurement.

Sur la figure 6.2 sont tracés les nuages de points correspondant à l'erreur relative (par rapport à un cas de référence pour lequel $tol_{rel} = 10^{-14}$) sur le délai d'auto-allumage $\tau^{c=0.1}$ obtenu à l'aide de différentes valeurs pour tol_{rel}, la valeur minimale pour tol_{abs} ayant été fixée. La diminution de la valeur de tol_{rel} permet bien une augmentation de la précision sur la détermination du délai d'auto-allumage $\tau^{c=0.1}$, mais cette amélioration n'est pas linéaire avec la valeur de tol_{rel}. Pour les valeurs de la tolérance relative les plus faibles, l'erreur relative d'estimation du délai d'auto-allumage $\tau^{c=0.1}$ est plus faible que la tolérance relative. Ceci peut s'expliquer par le fait que l'erreur relative contrôle l'erreur sur les $Y_k(t)$, et non sur le délai d'auto-allumage $\tau^{c=0.1}$.

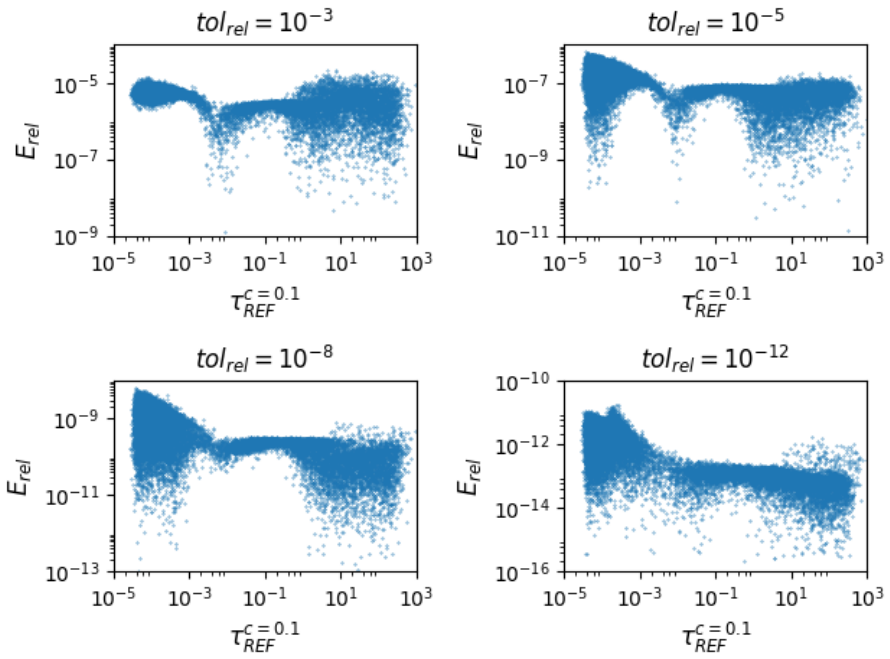

FIGURE 6.2 – *Erreur relative E_{rel} sur le calcul du délai d'auto-allumage $\tau^{c=0.1}$ obtenu avec différentes valeurs pour la tolérance relative tol_{rel}, et une tolérance absolue tol_{abs} fixée à la plus petite valeur possible, en fonction du délai d'auto-allumage de référence $\tau^{c=0.1}_{REF}$.*

Sur la figure 6.3 sont tracés les nuages de points correspondant à l'erreur relative sur le délai d'auto-allumage $\tau^{c=0.1}$ obtenu à l'aide de différentes valeurs

pour tol_{abs}, la valeur de 10^{-8} pour tol_{rel} ayant été choisie. Pour la valeur la plus grande de tol_{abs}, qui est 10^{-15}, l'erreur commise sur le délai d'auto-allumage $\tau^{c=0.1}$ est plus importante pour certains points que l'erreur d'auto-allumage avec les autres valeurs de tol_{abs}, surtout pour les points présentant un court délai d'auto-allumage. En revanche, l'erreur relative est sensiblement la même pour l'ensemble des valeurs de tol_{abs} inférieure à 10^{-20}. Le comportement différent observé pour la plus grande valeur de tol_{abs} peut sans doute s'expliquer par la mauvaise reproduction de la très faible concentration d'espèces radicalaires dans les instants initiaux, entraînant ensuite une erreur sur le délai d'auto-allumage.

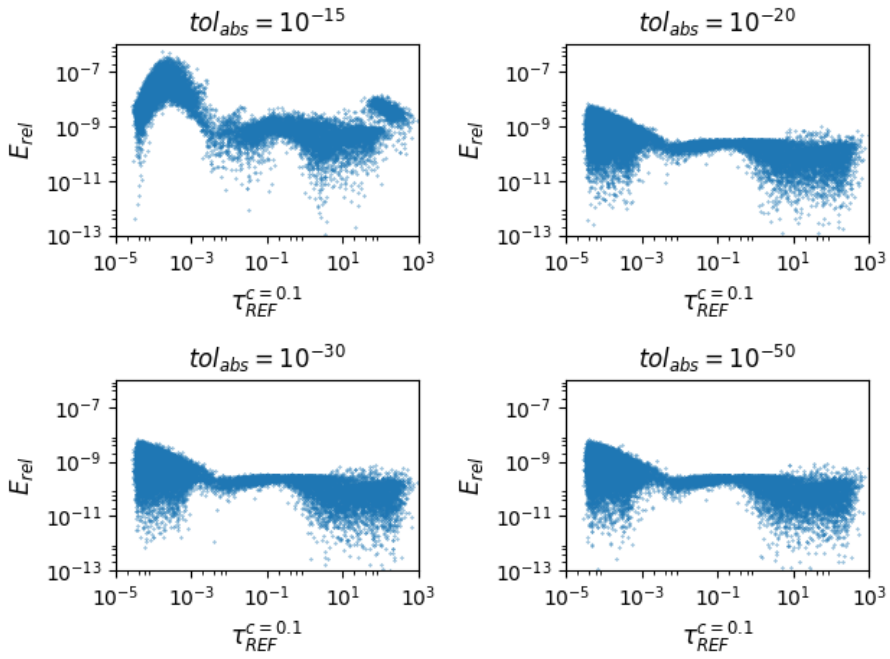

FIGURE 6.3 – *Erreur relative E_{rel} sur le calcul du délai d'auto-allumage $\tau^{c=0.1}$ obtenu avec différentes valeurs pour la tolérance absolue tol_{abs}, et une tolérance relative tol_{rel} fixée à la valeur 10^{-8}, en fonction du délai d'auto-allumage de référence $\tau_{REF}^{c=0.1}$.*

Sur la figure 6.4 sont tracées les évolutions du temps de calcul moyen, sur un cœur de Intel Xeon CPU E5-2670 v3, par résolution du système d'équations différentielles ordinaires (6.2), en fonction de tol_{rel} avec tol_{abs} fixée, et de tol_{abs} avec tol_{rel} fixée. La réduction des tolérances entraîne bien une augmentation du temps de calcul comme attendu par le nombre d'itérations plus importantes nécessaires. L'impact d'une diminution de la tolérance relative, qui augmente en pratique significativement la précision du résultat, est plus important que l'impact d'une diminution de la tolérance absolue.

Dans la suite, afin d'avoir une grande précision ainsi qu'un coût de calcul raisonnable, la tolérance relative sera toujours fixée à 10^{-8} et la tolérance absolue tol_{abs} à 10^{-30}.

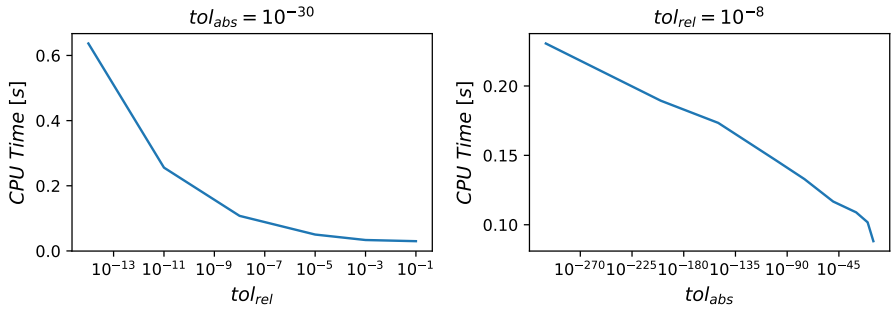

FIGURE 6.4 – *Temps de calcul moyen pour la résolution du système d'équations différentielles ordinaires (6.2), en fonction de la tolérance relative tol_{rel} avec $tol_{abs} = 10^{-30}$ (gauche) et de la tolérance absolue tol_{abs} avec $tol_{rel} = 10^{-8}$ (droite).*

6.1.1.2 Caractérisation du système chimique

La résolution du système d'équations différentielles ordinaires (6.2) permet d'obtenir l'évolution temporelle du système. Cette évolution du système peut se caractériser à travers l'évolution temporelle de différentes grandeurs physiques. L'évolution de la composition chimique du système par exemple est caractérisée par l'évolution temporelle des fractions massiques Y_k des différentes espèces présentes. D'autres grandeurs thermodynamiques permettent également de décrire l'évolution du système comme la température T, la masse volumique ρ. En plus de ces grandeurs, l'évolution de la capacité calorifique massique à pression constante c_p par exemple, ou la viscosité dynamique μ peuvent également être obtenues. Toutes ces grandeurs peuvent être nécessaires lors de calculs d'écoulements réactifs et sont tabulées lorsqu'une méthode de tabulation est utilisée.

L'évolution de ces différentes grandeurs est présentée sur la figure 6.5, pour une température initiale de 1200 K et un mélange air-hydrogène stœchiométrique à pression atmosphérique.

On observe sur la figure 6.5 que l'ensemble des grandeurs évoluent à partir de leur valeur initiale vers leur valeur finale, qui est leur valeur d'équilibre thermodynamique. L'évolution temporelle des grandeurs est liée à la cinétique chimique définie alors que l'état final est uniquement caractérisé par la thermodynamique. La majorité des profils présentent initialement un plateau, le mélange ne s'étant pas encore auto-allumé, suivi d'une forte pente de la courbe correspondant aux instants du processus de combustion. La ligne verticale rouge est placé au temps correspondant au délai d'auto-allumage, marqué par cette forte pente pour la majorité des profils. L'espèce HO_2 voit sa concentration augmenter avant que le mélange s'auto-allume, son pic de concentration permettant d'amorcer l'auto-allumage. Celle-ci se voit ensuite consommée, tout comme les autres espèces intermédiaires tels que H, O, H_2O_2 ou encore OH présentant un pic de concentration durant l'auto-allumage marquant la pré-

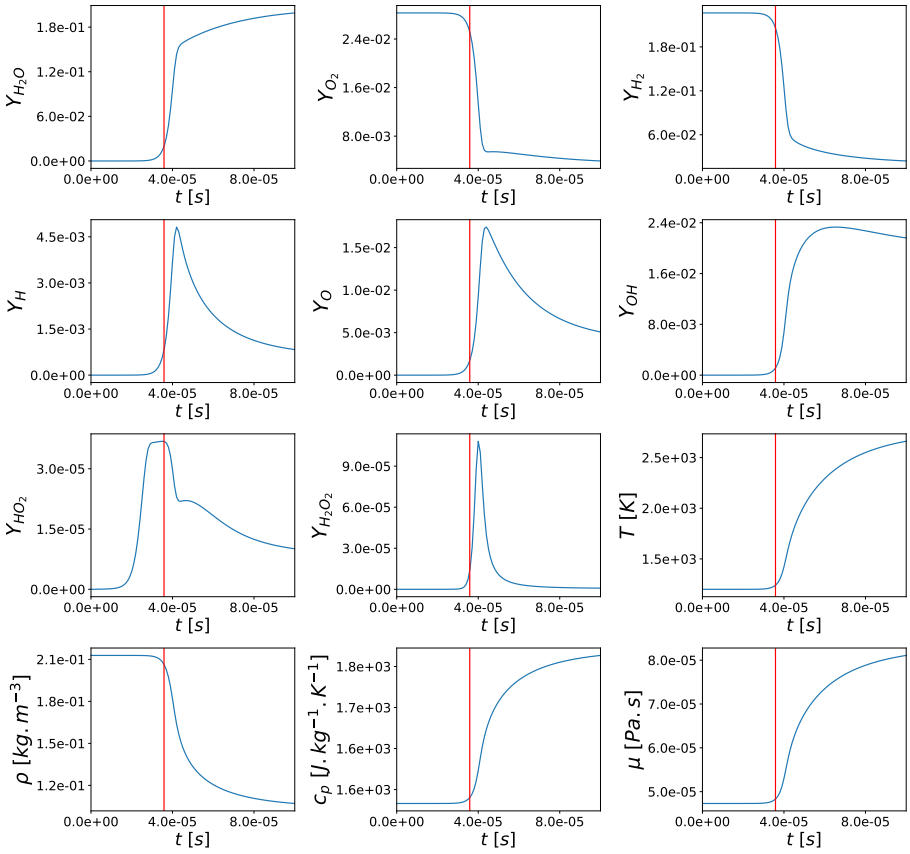

Figure 6.5 – *Évolutions temporelles de grandeurs pour un réacteur homogène adiabatique à pression constante et atmosphérique d'un mélange stœchiométrique air-hydrogène initialement à $T = 1200\ K$. La ligne verticale rouge a pour abscisse le délai d'auto-allumage.*

sence de la flamme, qui sont présentes à l'état final avec des concentrations négligeables. Les réactifs initiaux que sont le dihydrogène H_2 et le dioxygène O_2 sont consommés alors que le produit principal qu'est l'eau est produit en grande quantité. Du fait du mélange stœchiométrique, les quantités finales des réactifs sont faibles. La combustion entraîne une hausse de la température du mélange qui s'accompagne d'une dilatation des gaz entraînant une baisse de la masse volumique. La capacité calorifique massique à pression constante du mélange évolue elle aussi, du fait de l'augmentation de la température mais également du changement de composition chimique du système. Une évolution de la viscosité dynamique du gaz est également visible pour les mêmes raisons de changement de composition chimique et de température.

Le système évolue spontanément de l'état initial vers l'équilibre thermodynamique, la trajectoire suivie dans l'espace des phases étant naturellement paramétrée par le temps. Il est intéressant pour la tabulation de cette trajectoire

de paramétrer différemment cette trajectoire, à l'aide d'une variable de progrès traduisant l'état d'avancement de la réaction de combustion. Ce changement de paramétrage peut simplement être réalisé à l'aide d'une fonction monotone du temps. Il est classique de définir une variable d'avancement de la réaction Y_c comme combinaison linéaire des fractions massiques Y_k comme dans la relation (6.4), des méthodes ayant été développées pour déterminer pratiquement une "bonne" combinaison linéaire à considérer [92].

$$Y_c = \sum_{k=1}^{N_s} \alpha_k Y_k \tag{6.4}$$

L'avantage d'une telle définition de la variable d'avancement de la réaction est que celle-ci, sous certaines hypothèses suivant les cas, suit une équation de transport de même nature que les fractions massiques Y_k. Dans le cas d'un réacteur homogène adiabatique à pression constante, il est immédiat que la variable d'avancement Y_c suit l'équation différentielle ordinaire suivante :

$$\frac{dY_c}{dt} = \sum_{k=1}^{N_s} \alpha_k \dot{\omega}_k(t) = \dot{\omega}_{Y_c}(t) \tag{6.5}$$

Il est courant de normaliser cette variable d'avancement de la réaction afin d'obtenir une nouvelle variable d'avancement c qui évolue entre 0, dans les gaz frais, et 1 dans les gaz brûlés. En notant Y_c^f la valeur de Y_c dans les gaz frais, et Y_c^b sa valeur dans les gaz brûlés, une expression simple pour la variable d'avancement c est la suivante :

$$c = \frac{Y_c - Y_c^f}{Y_c^b - Y_c^f} \tag{6.6}$$

Dans le cas présent, la variable d'avancement de la réaction Y_c est choisie comme :

$$Y_c = Y_{H_2O} - Y_{H2} - Y_{O2} \tag{6.7}$$

La monotonie de cette variable d'avancement de la réaction a été vérifiée en pratique dans l'ensemble des cas explorés. Cette monotonie permet donc de décrire la trajectoire de la solution de (6.2) en fonction de la variable d'avancement de la réaction normalisée c. L'ensemble des profils des grandeurs présentes dans la figure 6.5 sont tracés en fonction c sur la figure 6.6.

Sur l'ensemble des courbes, la valeur en $c = 0$ correspond à la valeur initiale de la grandeur alors que la valeur en $c = 1$ correspond à la valeur d'équilibre

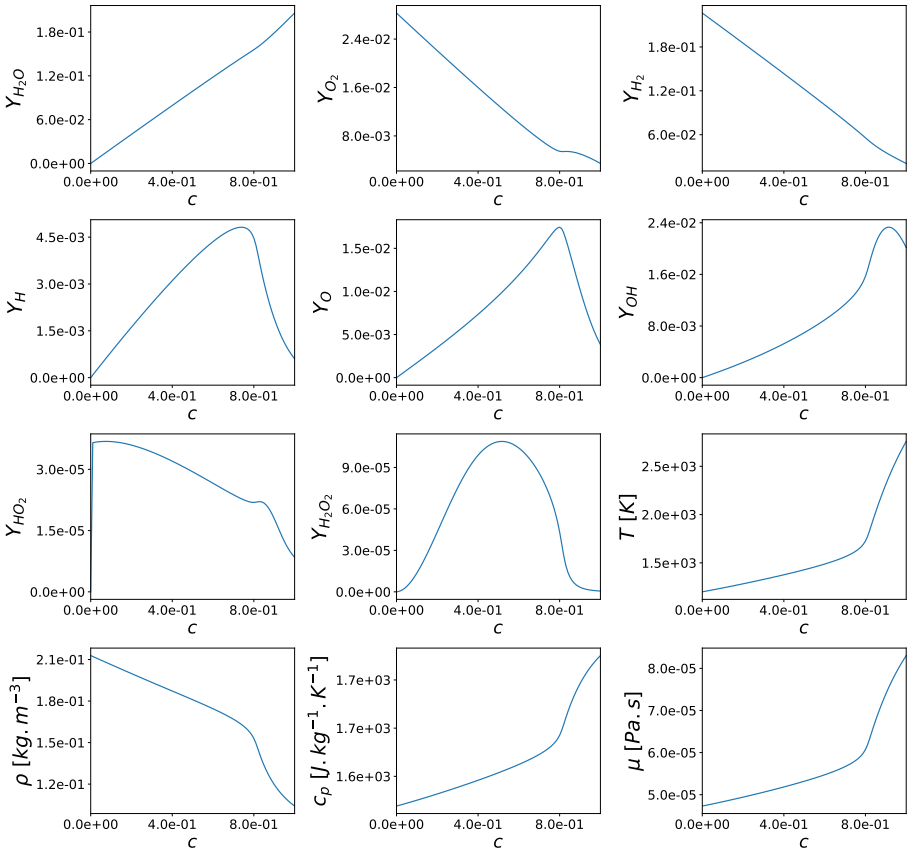

Figure 6.6 – *Grandeurs en fonction de c pour un réacteur homogène adiabatique à pression constante et atmosphérique d'un mélange stœchiométrique air-hydrogène initialement à* $T = 1200K$.

thermodynamique de la grandeur. Les profils des réactifs H_2 et O_2 décroissent avec l'avancement de la réaction du fait de leur consommation, alors que le profil de H_2O croît du fait de sa production. Concernant la température, la masse volumique, la capacité calorifique massique à pression constante ou encore la viscosité dynamique du mélange, les profils ont la même monotonie qu'ils ont en fonction du temps. Les autres espèces chimiques, qui sont des espèces chimiques très réactives et présentes essentiellement durant le processus de combustion, présentent toutes un pic correspondant à leur présence au sein du processus de combustion. En particulier, l'espèce HO_2 qui présentait un pic dans les instants antérieurs à l'instant d'auto-allumage présente un pic pour des valeurs de la variable d'avancement de la réaction très proche de 0.

Cette description de la trajectoire suivie par la solution de l'équation (6.2) permet en pratique la tabulation de l'état thermodynamique du système en fonction de cette variable d'avancement c. La chimie tabulée en pratique consiste donc à tabuler, entre autres, les courbes présentées sur la figure 6.6.

L'utilisation pratique de la variable d'avancement repose sur la résolution de l'équation différentielle ordinaire (6.5) pour laquelle le membre de droite a été tabulé en fonction de Y_c ou de c indifféremment. Dans la construction de la variable d'avancement de la réaction Y_c proposée, les réactifs H_2 et O_2 sont également utilisés en plus du produit H_2O alors même que le profil temporel de H_2O est monotone, permettant de définir une variable d'avancement de la réaction uniquement avec H_2O. En fait, un état initial contenant uniquement H_2 et O_2 implique, compte tenu du schéma réactionnel, que le terme source $\dot{\omega}_{H_2O}$ est nul pour cet état initial. De fait, choisir $Y_c = Y_{H_2O}$ implique dans un tel cas que l'état initial est un point d'équilibre pour (6.5), conduisant le système à rester dans son état initial. Afin de ne pas être dans un tel cas, Y_c est défini en faisant également intervenir les réactifs H_2 et O_2, permettant d'avoir $\dot{\omega}_{Y_c}$ non nul initialement, et donc de ne pas avoir le point initial comme point d'équilibre du système, permettant à celui-ci d'évoluer.

6.1.2 Cas d'une cinétique chimique incertaine

6.1.2.1 Modèle statistique des constantes de cinétique chimique

Le modèle d'incertitude pour les constantes de cinétiques chimiques k_{fr} a été choisi comme étant une loi log-normale pour le facteur pré-exponentiel A_r de chacune des réactions élémentaires, assurant que A_r soit dans l'intervalle $[A_r^0/f_r, A_r^0 f_r]$ (A^0 correspond à A et f correspond à UF dans 6.1) avec une probabilité de $99,7\%$, correspondant à un intervalle de 3σ. De plus, les incertitudes entre les différentes constantes chimiques sont considérées indépendantes entre elles. Dans les cas incertains qui suivent, chaque facteur pré-exponentiel A_r est une variable aléatoire suivant une loi log-normale caractérisée par les paramètres $\mu = \ln(A_r^0)$ et $\sigma = \ln(f_r)/3$, de sorte que la fonction de répartition pour A_r est :

$$F_r(x) = P(A_r \leq x) = \frac{1}{2} + \frac{1}{2}\,\mathrm{erf}\left[3\frac{\ln(x) - \ln(A_r^0)}{\ln(f_r)\sqrt{2}}\right] \tag{6.8}$$

Le choix a été fait ici de prendre un modèle relativement simple, qui permet d'introduire des incertitudes au sein de notre système. L'ensemble des résultats qui suivent pourraient être obtenus avec un modèle d'incertitude différents pour les k_{fr}, et les mêmes conclusions pourraient être obtenues.

Dans le cas incertain, les grandeurs du système sont désormais incertaines et dépendent donc du vecteur aléatoire $\mathbf{A} = (A_1, \ldots, A_{N_r})$ de $\mathbb{R}^{N_r}$, dont la densité de probabilité sera désormais notée $\pi_{\mathbf{A}}$.

6.1.2.2 Caractérisation du système chimique incertain

La propagation d'incertitudes pour les réacteurs homogènes adiabatiques à pression constante est réalisée dans la suite à l'aide d'une méthode non intrusive,

basée sur une méthode de Quasi-Monte Carlo randomisée utilisant 10 séquences de Sobol.

Dans le cas déterministe, une grandeur peut se caractériser par son évolution temporelle, et est donc représentée à l'aide d'une fonction x dépendant du temps.

$$\begin{aligned} x : \quad & \mathbb{R}^+ \to \mathbb{R} \\ & t \mapsto x(t) \end{aligned} \tag{6.9}$$

Dans le cas d'une cinétique chimique incertaine, cette grandeur n'est pas une simple fonction du temps, mais une fonction du temps et des paramètres incertains $\mathbf{A}$, ce qui en fait un processus stochastique que l'on notera X dans la suite.

$$\begin{aligned} X : \quad & \mathbb{R}^+ \times \mathbb{R}^{N_r} \to \mathbb{R} \\ & (t, \mathbf{A}) \mapsto X(t, \mathbf{A}) \end{aligned} \tag{6.10}$$

Désormais, pour un instant t donné, $X(t, .)$ n'est pas une valeur mais une variable aléatoire qui sera notée X_t dans la suite. Pour une valeur $\mathbf{a}$ des paramètres incertains, $X(., \mathbf{a})$ est une fonction du temps, aussi appelé trajectoire de X et qui sera notée $X_.(\mathbf{a})$ dans la suite. Des trajectoires temporelles des différentes grandeurs présentées dans le cas déterministe sur la figure 6.5 sont présentes sur la figure 6.7.

Les trajectoires temporelles d'une même grandeur ont des profils similaires et ont toutes en communs le point initial, l'état initial étant le même pour l'ensemble d'entre elles, ainsi que le point final correspondant à l'équilibre thermodynamique (qui n'est pas encore atteint et donc pas visible sur la figure) du fait que l'équilibre thermodynamique ne dépend pas de la cinétique chimique. Les différentes trajectoires présentent une forte pente correspondant à l'auto-allumage à différents instants, qui est le résultat d'une cinétique chimique plus ou moins rapide. Un autre impact est visible sur la concentration maximale des espèces intermédiaires qui varient avec la trajectoire considérée, du fait que la compétition entre la production et la consommation de ces espèces se voit modifiée également par la modification des vitesses de réaction.

L'ensemble des trajectoires possibles caractérise un processus stochastique, mais étant infini, il est nécessaire en pratique de décrire le processus stochastique à l'aide d'autres informations utilisables en pratique. Deux statistiques, que sont la moyenne m_X et l'écart-type σ_X (ou la variance σ_X^2 indifféremment), permettent de donner des informations intéressantes et souvent utiles en pratique sur le processus stochastique X. La moyenne m_X du processus stochastique X est une fonction du temps définie par la relation suivante pour un

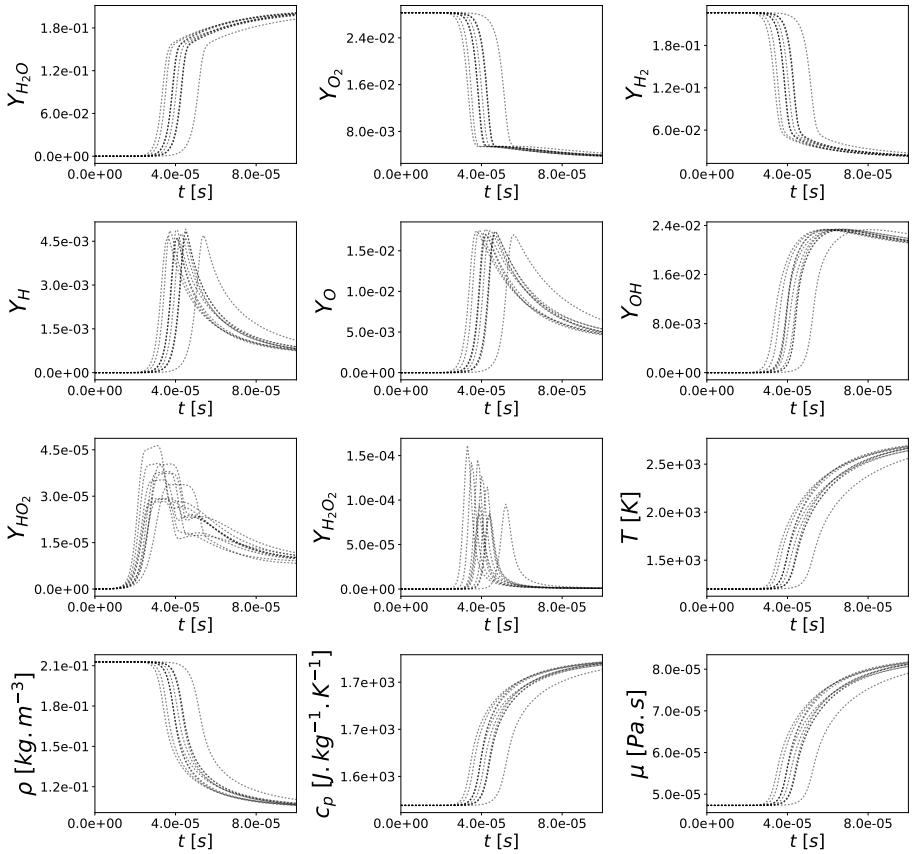

FIGURE 6.7 – *Trajectoires de processus stochastiques pour un réacteur homogène adiabatique à pression constante et atmosphérique d'un mélange stœchiométrique air-hydrogène initialement à* $T = 1200\ K$.

instant t :

$$m_X(t) = E\left[X_t\right] = \int_{\mathbb{R}^{N_r}} X_t(\mathbf{A})\pi_{\mathbf{A}}(\mathbf{A})d\mathbf{A} \tag{6.11}$$

La variance σ_X^2 quant à elle est également une fonction du temps qui est définie par la relation suivante pour un instant t :

$$\sigma_X^2 = E\left[\left(X_t - E\left[X_t\right]\right)^2\right] \tag{6.12}$$

Ces deux statistiques pour certaines grandeurs sont tracées en fonction du temps sur la figure 6.8 pour un réacteur homogène adiabatique à pression constante et atmosphérique d'un mélange stœchiométrique air-hydrogène initialement à 1200 K. Les moyennes temporelles pour l'ensemble des grandeurs

possèdent des profils d'aspect proches de ceux des trajectoires, et vont toutes de leur valeur dans l'état initial à leur valeur dans l'état d'équilibre thermodynamique. L'écart-type pour l'ensemble des valeurs est nul initialement et tend vers 0 à mesure que l'on se rapproche de l'état d'équilibre thermodynamique, ces deux états étant certains pour le système. Pour les autres instants, cet écart-type est non nul pour l'ensemble des grandeurs, présentant un pic pour l'ensemble des grandeurs excepté pour la fraction massique de HO_2 qui présente deux pics.

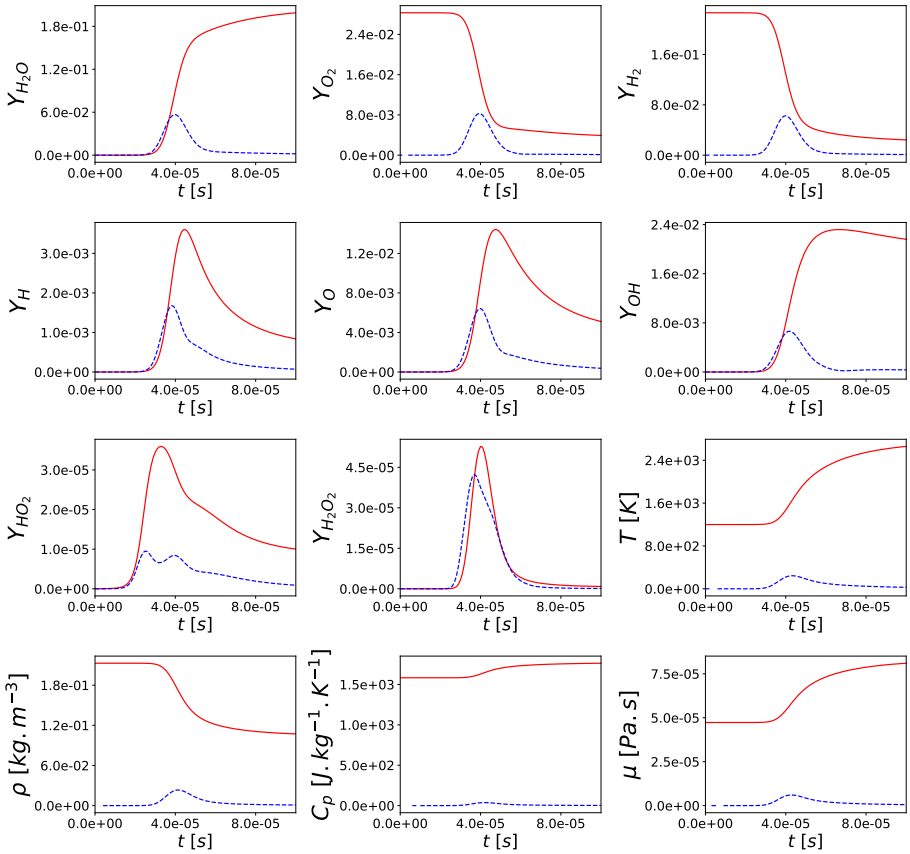

Figure 6.8 – *Moyennes m_X (lignes pleines rouges) et écart-types σ_X (pointillés bleus) temporels de processus stochastiques X pour un réacteur homogène adiabatique à pression constante et atmosphérique d'un mélange stœchiométrique air-hydrogène initialement à $T = 1200K$.*

Les statistiques que sont la moyenne et la variance ne suffisent pas à caractériser les processus stochastiques, mais ce sont en pratique les premières statistiques d'intérêt, et l'objectif est donc de reproduire en priorité ces deux statistiques.

Tout comme dans le cas déterministe, il est possible de définir une variable d'avancement de la réaction Y_c qui sera maintenant un processus stochastique,

que l'on pourra ensuite normaliser pour donner une variable d'avancement de la réaction normalisée C, elle même étant désormais un processus stochastique. La même définition (6.7) a été choisie pour Y_c, en s'assurant qu'elle était toujours monotone quelle que soit les valeurs prises par le vecteur aléatoire $\mathbf{A}$ en pratique. Des trajectoires temporelles du processus stochastique C, ainsi que sa moyenne et son écart-type sont présentés sur la gauche de la figure 6.9, où l'on observe bien la croissance de l'ensemble des trajectoires. La moyenne du processus stochastique C passe de la valeur 0 à la valeur 1 de façon monotone comme l'ensemble des trajectoires. La pente maximale de celle-ci est globalement plus faible que l'ensemble des trajectoires, ce qui est plus marqué encore pour des conditions initiales plus froides notamment. L'écart-type du processus stochastique C est initialement nul, et tend également vers une valeur nulle pour un temps infini correspondant à l'équilibre thermodynamique. La zone pour laquelle celui-ci a une valeur significativement différentes de 0 est la zone pour laquelle la moyenne temporelle est significativement éloignée de ses valeurs extrêmes que sont 0 et 1. Sur la droite de la figure 6.9 est présenté une estimation par une méthode à noyau de la densité de probabilité du délai d'auto-allumage $\tau^{C=0.1}$. Ce délai d'auto-allumage est défini comme l'instant pour lequel la variable d'avancement normalisée C atteint la valeur 0.1, ce qui définit bien une variable aléatoire. Cette densité de probabilité permet de visualiser globalement l'impact de l'incertitude de la cinétique chimique sur la vitesse du processus de combustion.

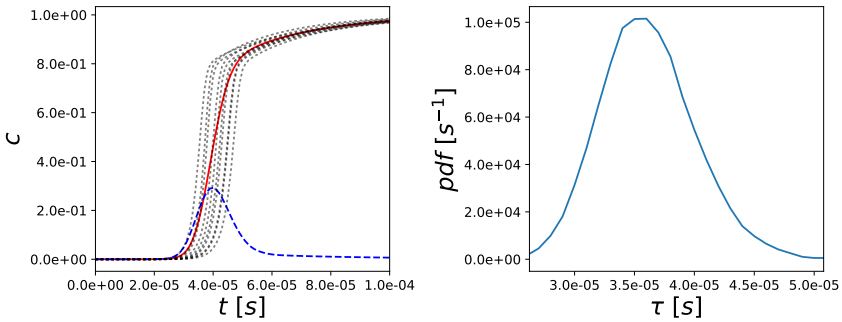

FIGURE 6.9 – *Droite : moyenne temporelle m_C de ce processus stochastique (ligne pleine rouge) et écart-type temporel σ_C de ce processus stochastique (tirets bleus) et trajectoires (lignes pointillés) du processus stochastique C pour un réacteur homogène adiabatique à pression constante et atmosphérique d'un mélange stœchiométrique air-hydrogène initialement à $T = 1200\ K$. Droite : densité de probabilité du délai d'auto-allumage $\tau^{C=0.1}$ pour un réacteur homogène adiabatique à pression constante et atmosphérique d'un mélange stœchiométrique air-hydrogène initialement à $T = 1200\ K$.*

Chacune des trajectoires $C_\cdot(\mathbf{a})$ du processus stochastique C est strictement croissante, ce qui assure l'existence d'une bijection réciproque $T_\cdot(\mathbf{a})$ de cette trajectoire. Cela permet de définir un processus stochastique $T = (T_c)_{c\in[0,1]}$ dont les trajectoires sont les bijections réciproques des trajectoires $C_\cdot(\mathbf{a})$. Les rela-

tions suivantes existent alors par construction entre les processus stochastiques C et T :

$$\begin{cases} \forall t \geq 0, \forall \mathbf{a} \in \mathbb{R}^{N_r}, T(C_t(t, \mathbf{a}), \mathbf{a}) = t \\ \forall 0 \leq c \leq 1, \forall \mathbf{a} \in \mathbb{R}^{N_r}, C(T(c, \mathbf{a}), \mathbf{a}) = c \end{cases} \tag{6.13}$$

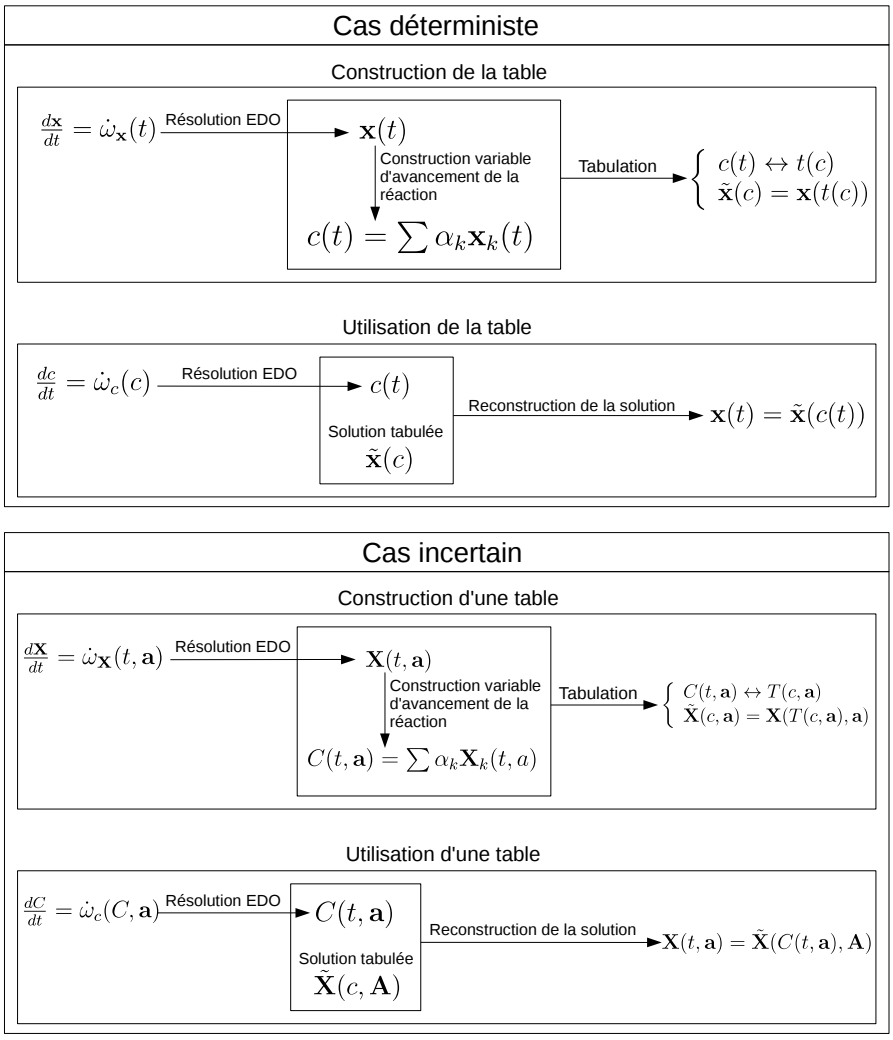

Figure 6.10 – *Construction de la table et utilisation de celle-ci dans un cas déterministe et incertain.*

Le processus stochastique T représente le temps nécessaire pour que le processus stochastique C représentant la variable d'avancement de la réaction ait atteint une certaine valeur. Il est, avec cette définition du processus sto-

chastique T, possible pour n'importe quel processus stochastique temporel X de définir un processus stochastique $\tilde{X}$ de la manière suivante :

$$\forall 0 \leq c \leq 1, \tilde{X}(c, \mathbf{A}) = X(T(c, \mathbf{A}), \mathbf{A}) \tag{6.14}$$

Une relation symétrique existe entre X et $\tilde{X}$:

$$\forall t \geq 0, X(t, \mathbf{A}) = \tilde{X}(C(t, \mathbf{A}), \mathbf{A}) \tag{6.15}$$

Ces deux dernières relations peuvent également s'écrire $\tilde{X}_c = X_{T_c}$ et $X_t = \tilde{X}_{C_t}$. Ce formalisme dans le cas incertain fait écho à ce qui est fait dans le cas déterministe dans la construction et l'utilisation directe d'une table chimique. La construction et l'utilisation directe d'une table chimique sont brièvement décrite par le schéma de la figure 6.10, pour le cas déterministe, et également dans un cas incertain où une méthode non intrusive est utilisé, impliquant la construction d'une table pour chaque échantillon $\mathbf{a}$ du vecteur aléatoire $\mathbf{A}$ des paramètres incertains. Alors que dans le cas déterministe, la tabulation consiste à tabuler des fonctions $\tilde{x}$, dans le cas incertain des processus stochastiques $\tilde{X}$ sont désormais à considérer à la place de ces fonctions. La tabulation d'une trajectoire d'un processus stochastique $\tilde{X}$ est identique à la tabulation dans le cas déterministe, mais la tabulation complète du processus stochastique $\tilde{X}$ est pratiquement non souhaitée, puisque cela nécessite l'augmentation de la dimension de la table de la dimension stochastique N_r à priori.

La figure 6.11 présente des trajectoires de processus $\tilde{X}$, leurs moyennes $m_{\tilde{X}}$ ainsi que leurs écart-types $\sigma_{\tilde{X}}$. Les trajectoires présentées correspondent aux trajectoires temporelles présentées dans la figure 6.7, le passage des trajectoires temporelles aux trajectoires en fonction de la variable d'avancement de la réaction se faisant grâce à l'expression (6.14). Alors que pour les espèces majoritaires que sont O_2, H_2 et H_2O, les trajectoires de $X_.(\mathbf{a})$ sont bien distinctes, ce n'est pas le cas pour les trajectoires de $\tilde{X}_.(\mathbf{a})$ qui sont presque toutes confondues avec la valeur moyenne $m_{\tilde{X}}$, comme en témoigne également l'écart-type $\sigma_{\tilde{X}}$ qui a une valeur négligeable devant la valeur moyenne sur l'ensemble du segment $[0, 1]$. Un comportement similaire est présent pour les grandeurs que sont la température, la masse volumique, la capacité calorifique massique à pression constante ainsi que la viscosité dynamique du gaz. Concernant les autres espèces chimiques, deux groupes peuvent être distingués. Le premier groupe comprend les espèces O, H et OH pour lesquelles les trajectoires $\tilde{X}_.(\mathbf{a})$ restent très proches de la moyenne $m_{\tilde{X}}$, ce qui se traduit pas un écart-type $\sigma_{\tilde{X}}$ relativement faible sur l'ensemble du segment $[0, 1]$. Les espèces HO_2 et H_2O_2 quant à elle présentent des trajectoires $\tilde{X}_.(\mathbf{a})$ qui sont plus dispersées autour de la moyenne $m_{\tilde{X}}$, comme le traduit l'écart-type $\sigma_{\tilde{X}}$ qui a une valeur non négligeable devant $m_{\tilde{X}}$ pour certaines valeurs de $[0, 1]$. Pour ces deux espèces, les

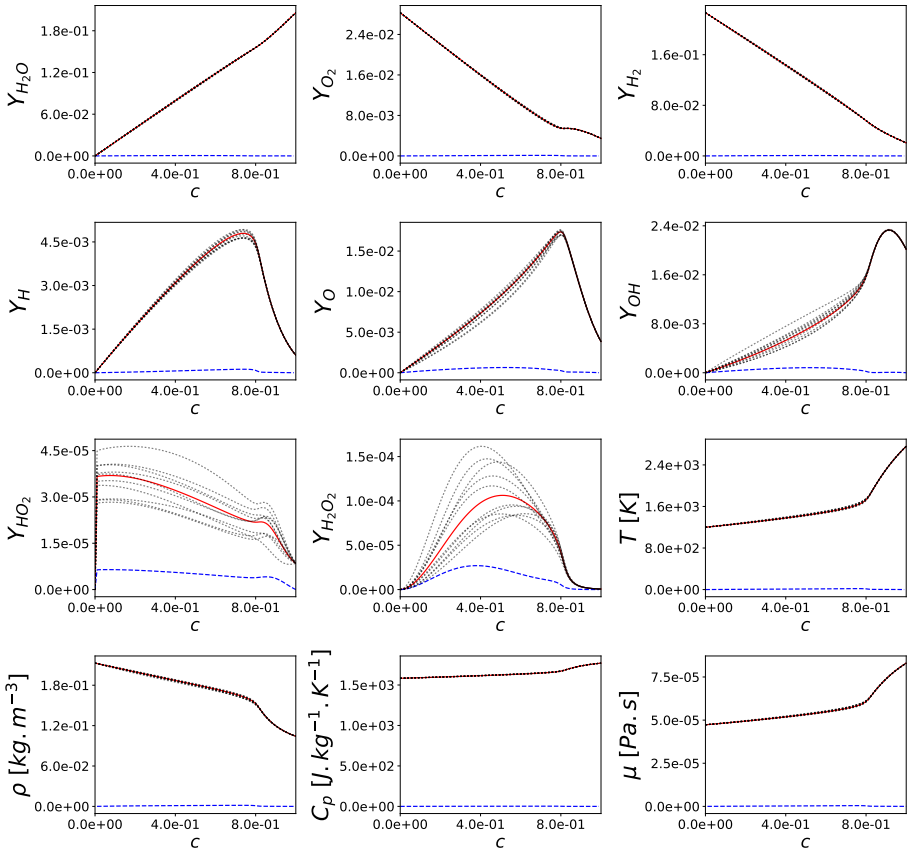

Figure 6.11 – *Moyennes $m_{\tilde{X}}$ (lignes pleines rouges), écart-types $\sigma_{\tilde{X}}$ (pointillés bleus) et trajectoires (pointillés) de processus stochastiques $\tilde{X}$ pour un réacteur homogène adiabatique à pression constante et atmosphérique d'un mélange stœchiométrique air-hydrogène initialement à $T = 1200K$.*

trajectoires $X_{\cdot}(\mathbf{a})$ présentent des valeurs maximales différentes de la moyenne m_X, contrairement aux autres espèces et grandeurs.

La considération d'une cinétique chimique incertaine amène à une complexification du problème, avec notamment les fonctions de tabulation des grandeurs $\tilde{x}$ devenant des processus stochastiques $\tilde{X}$. Une méthode non intrusive de force brute pour la propagation d'incertitude dans une simulation aux grandes échelles avec une chimie tabulée nécessiterait de créer un échantillonnage de table chimique, chacune d'elle tabulant les trajectoires des processus stochastiques $\tilde{X}$. Cependant, une telle méthode est trop coûteuse en pratique du fait qu'elle ne réduit pas la dimension stochastique. Il est donc nécessaire de s'intéresser à une paramétrisation de faible dimension stochastique de tables chimiques permettant de propager efficacement les incertitudes de la cinétique chimique dans une simulation aux grandes échelles via l'utilisation de méthodes non intrusives.

6.2 Introduction d'incertitudes dans la chimie tabulée

6.2.1 Chimie tabulée dans le cas déterministe

Dans le cas de l'utilisation d'un réacteur homogène adiabatique à pression constante pour décrire la combustion dans une simulation aux grandes échelles, la variable d'avancement de la réaction c fait office de paramètre de contrôle, c'est à dire que l'ensemble des grandeurs tabulées le sont entre autres en fonction de la variable d'avancement de la réaction. L'obtention de la valeur de cette variable d'avancement de la réaction au sein du calcul se fait à l'aide de son équation de transport, qui nécessite la tabulation du terme source $\dot{\omega}_c$ (ou indifféremment $\dot{\omega}_{Y_c}$) de celle-ci.

Dans une telle simulation aux grandes échelles, il faut distinguer deux types de grandeurs tabulées, comme fait dans [87] : les grandeurs actives, qui interviennent dans une ou plusieurs équations de l'écoulement, comme la masse volumique ou des coefficients de diffusions par exemple, et les grandeurs passives qui n'interviennent pas dans les équations de l'écoulement, et qui n'influencent donc pas celui-ci, qui sont par exemple l'ensemble des espèces chimiques.

Dans le cas incertain, ces deux types de grandeurs ne sont pas à traiter de la même manière, les incertitudes des grandeurs actives devant être impérativement correctement reproduites afin de pouvoir propager les incertitudes de la cinétique chimique au sein de l'écoulement.

6.2.2 Chimie tabulée dans le cas d'une cinétique chimique incertaine

L'utilisation de méthodes de propagation non intrusive pour la propagation d'incertitude en simulations aux grandes échelles d'un écoulement réactif utilisant une chimie tabulée implique la réalisation de plusieurs simulations aux grandes échelles, chacune possédant sa propre table chimique. Comme vu précédemment, les tables chimiques dans le cas de l'utilisation d'une méthode non intrusive de Monte Carlo contiendrait des trajectoires des processus stochastiques $\tilde{X}$. De telles tables chimiques seraient à même de reproduire les incertitudes sur la cinétique chimique, mais sont paramétrées par un nombre trop important de paramètres incertains en pratique pour l'utilisation de méthodes de propagation efficaces. L'utilisation de méthodes de propagation efficaces nécessite d'être capable de construire des tables chimiques paramétrées par un nombre restreint de paramètres incertains.

Dans ce chapitre, une construction de table chimique est proposée, capable de reproduire la cinétique chimique incertaine, et pouvant donc être utilisée en simulation aux grandes échelles d'écoulements réactifs. L'ensemble des difficultés et contraintes venant d'être mentionnées nous ont conduit à envisager de considérer l'utilisation d'une chimie tabulée reposant sur une table similaire au

cas déterministe, c'est à dire uniquement paramétrée par des dimensions déterministes, c dans le cas étudié dans cette section. Pour une grandeur X, on suppose donc donnée une fonction $\tilde{x}_{tab}$. L'introduction d'incertitudes dans une telle chimie tabulée ne passera que par les incertitudes introduites par le terme source de la variable d'avancement non normalisée Y_c, qui ne sera pas obtenu à travers la table, mais qui proviendra dans cette partie de calculs de chimie détaillée. Le terme source de Y_c considéré incertain, dépend de l'ensemble des constantes de cinétiques chimiques incertaines données par le vecteur aléatoire $\mathbf{A}$. L'ODE régissant l'évolution de Y_c dans le cas d'un réacteur adiabatique homogène à pression constante avec une cinétique chimique incertaine est désormais :

$$\frac{dY_c}{dt} = \dot{\omega_{Y_c}}(c, \mathbf{A}) \tag{6.16}$$

Une fois une réalisation $\mathbf{a}$ du vecteur aléatoire $\mathbf{A}$ donnée, cette ODE est résolue de la même manière que dans le cas déterministe, et une trajectoire de $Y_c(t, \mathbf{a})$ peut ainsi être obtenue, qui après normalisation donne une trajectoire $C_{\cdot}(\mathbf{a})$ de la variable d'avancement normalisée. Cette trajectoire de C est dans ce cas, pour une résolution parfaite de l'équation différentielle ordinaire (6.16), la même que celle obtenue avec une chimie détaillée dont les paramètres cinétiques sont données par le vecteur $\mathbf{a}$. A partir de là, il est possible d'obtenir la valeur temporelle de n'importe quelle grandeur d'intérêt X comme :

$$X_{tab}(t, \mathbf{a}) = \tilde{x}_{tab}(C_t(t, \mathbf{a})) \tag{6.17}$$

A travers ce procédé, on est donc capable de construire pour n'importe quelle grandeur d'intérêt X un processus stochastique X_{tab}, étant donnée un processus stochastique C pour la variable d'avancement, à travers la relation :

$$X_{tab}(t) = \tilde{x}_{tab}(C(t)) \tag{6.18}$$

Une question importante est maintenant de comparer le processus stochastique X obtenu à l'aide de la chimie détaillée avec le processus stochastique X_{tab} correspondant obtenue à l'aide du processus stochastique C, obtenu lui avec la chimie détaillée, notamment en terme de moyenne et de variance. Dans la suite, le processus X_{tab} correspondra toujours à $\tilde{x}_{tab}(C(t))$.

6.3 Reproduction de la valeur moyenne temporelle

Pour commencer la comparaison, il convient de s'intéresser au moment d'ordre 1 des processus X et X_{tab}, à savoir l'espérance de ceux-ci. L'espérance

du processus X_{tab} à un instant t quelconque peut s'écrire sous la forme suivante :

$$E\left[X_{tab}(t)\right] = E\left[\tilde{x}_{tab}(C_t)\right] = \int_0^1 \tilde{x}_{tab}(c)\pi_{C_t}(c)dc \tag{6.19}$$

Dans l'expression (6.19), π_{C_t} correspond à la densité de probabilité de la variable aléatoire C_t. Il est possible d'exprimer l'espérance de X_t sous une forme similaire, en utilisant l'espérance conditionnelle de X_t par rapport à C_t :

$$E\left[X_t\right] = E\left[E\left[X_t|C_t\right]\right] = \int_0^1 E\left[X_t|C_t = c\right]\pi_{C_t}(c)dc \tag{6.20}$$

Cette dernière expression présente la même forme que l'expression (6.19). En fait, l'espérance conditionnelle $E\left[X_t|C_t\right]$ peut s'exprimer sous la forme suivante :

$$E\left[X_t|C_t\right] = \psi_X^{(t)}\left(C_t\right) \tag{6.21}$$

Dans l'expression (6.21), $\psi_X^{(t)}$ est une fonction mesurable, qui dépend à priori de l'instant t. En combinant les expressions (6.20) et (6.21), l'espérance de X_t peut désormais s'exprimer comme :

$$E\left[X_t\right] = E\left[\psi_X^{(t)}\left(C_t\right)\right] = \int_0^1 \psi_X^{(t)}\left(c\right)\pi_{C_t}(c)dc \tag{6.22}$$

La comparaison des expressions (6.19) et (6.20) permet de considérer comme candidat pour $\tilde{x}_{tab}$ la fonction $\psi_X^{(t)}$ pour la bonne reproduction de l'espérance de X_t. Cependant, la fonction $\tilde{x}_{tab}$ est une fonction indépendante de l'instant t, dont le choix doit permettre la bonne reproduction de la moyenne m_X du processus stochastique X à tout instant t, et non pas seulement à un instant t donné. Il est donc nécessaire de choisir la fonction $\tilde{x}_{tab}$ au mieux pour qu'elle soit la plus "proche" possible de l'ensemble des fonctions $\psi_X^{(t)}$.

Dans le cas où la variable aléatoire X_t possède un moment d'ordre 2, comme c'est le cas dans le cas étudié, l'espérance conditionnelle de X_t par rapport à C_t est caractérisée par le fait qu'elle est la projection orthogonale pour le produit scalaire canonique de X_t sur l'ensemble des fonctions C_t-mesurables, c'est à dire s'écrivant sous la forme $\phi(C_t)$, où ϕ est une fonction borélienne. Comme projection orthogonale, et puisque l'ensemble des fonctions C_t-mesurables est un sous-espace vectoriel fermé, on peut aussi la caractériser comme fonction C_t-mesurables la plus proche de X_t au sens de la norme associée au produit scalaire. La fonction $\psi_X^{(t)}$ recherchée est alors la solution du

problème d'optimisation suivant :

$$\psi_X^{(t)} = \underset{\phi}{\operatorname{argmin}}\ E\left[(X_t - \phi(C_t))^2\right] \tag{6.23}$$

Cette dernière expression permet de calculer en pratique l'espérance conditionnelle $E[X_t|C_t]$. En effet, si l'on possède des réalisations $(c_i, x_i)_{i=1,N}$ des variables aléatoires (C_t, X_t), et une fonction ϕ, la distance entre X_t et $\phi(C_t)$ peut être estimée par :

$$E\left[(X_t - \phi(C_t))^2\right] \approx \frac{1}{N}\sum_{i=1}^{N}(x_i - \phi(c_i))^2 \tag{6.24}$$

Le calcul de l'espérance conditionnelle revient donc à déterminer la fonction $\psi_X^{(t)}$ qui minimise la quantité de droite dans l'expression (6.24), ce qui revient à résoudre un problème d'optimisation des moindres carrés non linéaire. Ce problème d'optimisation a été résolu en considérant pour l'espace des fonctions ϕ des splines cubiques naturelles, et en utilisant l'algorithme de Marquardt [81]. Sur la figure 6.12 sont présentées les solutions $\psi_X^{(t)}$ du problème d'optimisation pour les différentes grandeurs précédemment présentées pour l'instant $t = 40\mu s$, pour un réacteur homogène initialement à $T = 1200K$ et un mélange air-hydrogène stœchiométrique, ainsi que les nuages de points des points (c_i, x_i) ayant servi à l'optimisation.

L'ensemble des fonctions $\psi_X^{(t)}$ présentées sur la figure 6.12 se situent bien à l'intérieur du nuage de points, étant au plus proche de l'ensemble des points au sens du carré de la norme associée au produit scalaire canonique. Les profils des fonctions $\psi_X^{(t)}$ sont similaires à ceux des moyennes $m_{\tilde{X}}$ des processus stochastiques $\tilde{X}$ présentées sur la figure 6.11. La dispersion des nuages de points autour de la fonction $\psi_X^{(t)}$ suit également les résultats de la figure 6.11, étant beaucoup plus importante pour les grandeurs présentant un écart-type $\sigma_{\tilde{X}}$ important.

Comme dit auparavant, l'espérance conditionnelle $E[X_t|C_t]$ dépend de l'instant t considéré, tout comme la variable aléatoire C_t dépend de l'instant considéré comme le montre la figure 6.13 où sont représentées des estimations par noyau gaussien de la densité de probabilité π_{C_t} de variables aléatoires C_t pour différents instants t. On peut voir sur cette figure qu'il existe des intervalles pour lesquels la valeur de la densité de probabilité de C_t est quasiment nulle.

Pour les valeurs c présentant une densité de probabilité π_{C_t} nulle, l'espérance conditionnelle $E[X_t|C_t]$ peut en théorie prendre n'importe quelle valeur. En pratique, les zones pour lesquelles la densité de probabilité π_{C_t} est faible présentent moins de points c_i utilisés pour l'estimation par la méthode d'optimisation, ce qui explique les oscillations de certaines estimations de $E[X_t|C_t]$

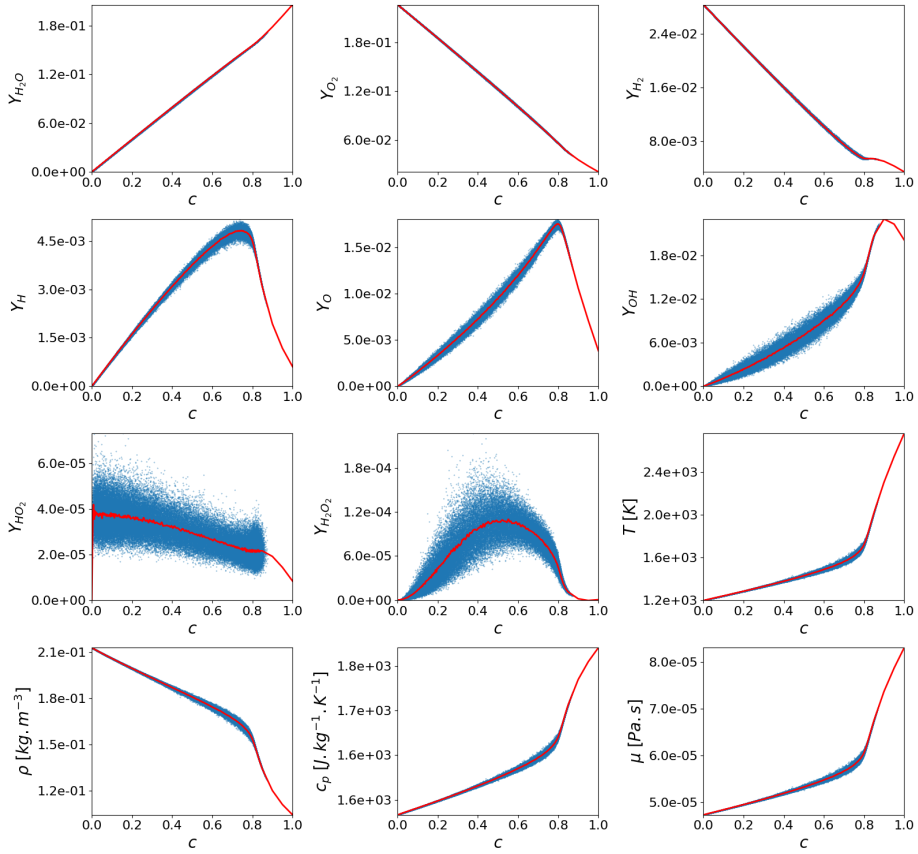

FIGURE 6.12 – *Nuages de points de réalisations indépendantes des couple de variables aléatoires* (C_t, X_t) *et solutions* $\psi_X^{(t)}$ *du problème d'optimisation* (6.24) *(ligne pleine rouge), calculés à l'aide d'une chimie détaillée pour un réacteur homogène adiabatique à pression constante et atmosphérique d'un mélange stœchiométrique air-hydrogène initialement à* $T = 1200K$*, et pour l'instant* $t = 40\mu s$*.*

qui sont mal convergées pour ces zones.

La figure 6.14 montre que pour les espèces H_2O, H_2, O_2, les fonctions $\psi_X^{(t)}$ pour différents instants t coïncident, les différences relatives entre elles n'étant pas significatives. Pour le cas de telles espèces, le choix de la valeur tabulée à utiliser pour la reproduction de la valeur moyenne temporelle de la concentration de ces espèces est donc cette valeur commune des fonctions $\psi_X^{(t)}$. On peut en fait relier cette invariance de la fonction $\psi_X^{(t)}$ avec le temps t à l'espérance du processus $\tilde{X}$. En effet, supposons que la relation (6.25) soient vérifiée, où ψ_X est une fonction ne dépendant pas de l'instant t.

$$\forall t, E\left[X_t | C_t\right] = \psi_X(C_t) \tag{6.25}$$

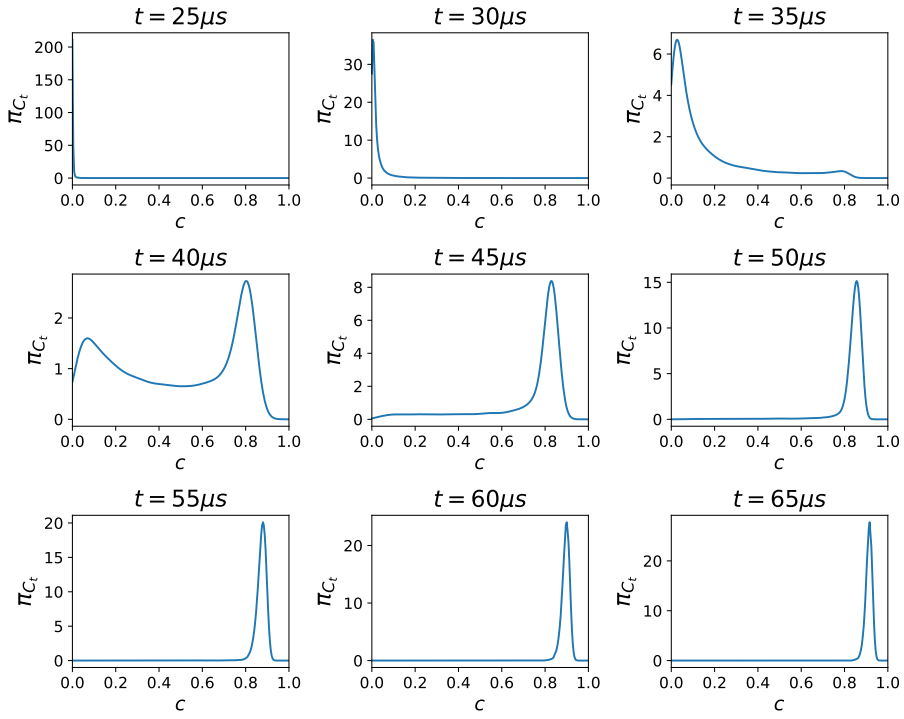

Figure 6.13 – *Densité de probabilité π_{C_t}, obtenues à l'aide d'une méthode à noyau gaussien, des variables aléatoires C_t à différents instants t dans le cas d'un réacteur homogène et adiabatique à pression constante et atmosphérique d'un mélange stœchiométrique air-hydrogène, initialement à $T = 1200K$.*

Il est alors possible de montrer que dans un tel cas, l'espérance de la variable aléatoire $\tilde{X}_c$ pour c dans $[0, 1]$ est égale à $\psi_X(c)$, en remarquant que la variable aléatoire C_{T_c} est presque sûrement égale à c :

$$\begin{aligned}
E\left[\tilde{X}_c\right] &= E\left[X_{T_c}\right] = E\left[E\left[X_{T_c}|C_{T_c}\right]\right] = E\left[E\left[E\left[X_{T_c}|C_{T_c}\right]|T_c\right]\right] \\
&= \int_{\mathbb{R}^+} E\left[E\left[X_{T_c}|C_{T_c}\right]|T_c = t\right]\pi_{T_c}(t)dt \\
&= \int_{\mathbb{R}^+} E\left[E\left[X_t|C_t\right]|T_c = t\right]\pi_{T_c}(t)dt \\
&= \int_{\mathbb{R}^+} E\left[\psi_X(C_t)|T_c = t\right]\pi_{T_c}(t)dt \\
&= \int_{\mathbb{R}^+} E\left[\psi_X(C_{T_c})|T_c = t\right]\pi_{T_c}(t)dt \\
&= \int_{\mathbb{R}^+} E\left[\psi_X(c)|T_c = t\right]\pi_{T_c}(t)dt \\
&= E\left[E\left[\psi_X(c)|T_c\right]\right] = \psi_X(c)
\end{aligned} \tag{6.26}$$

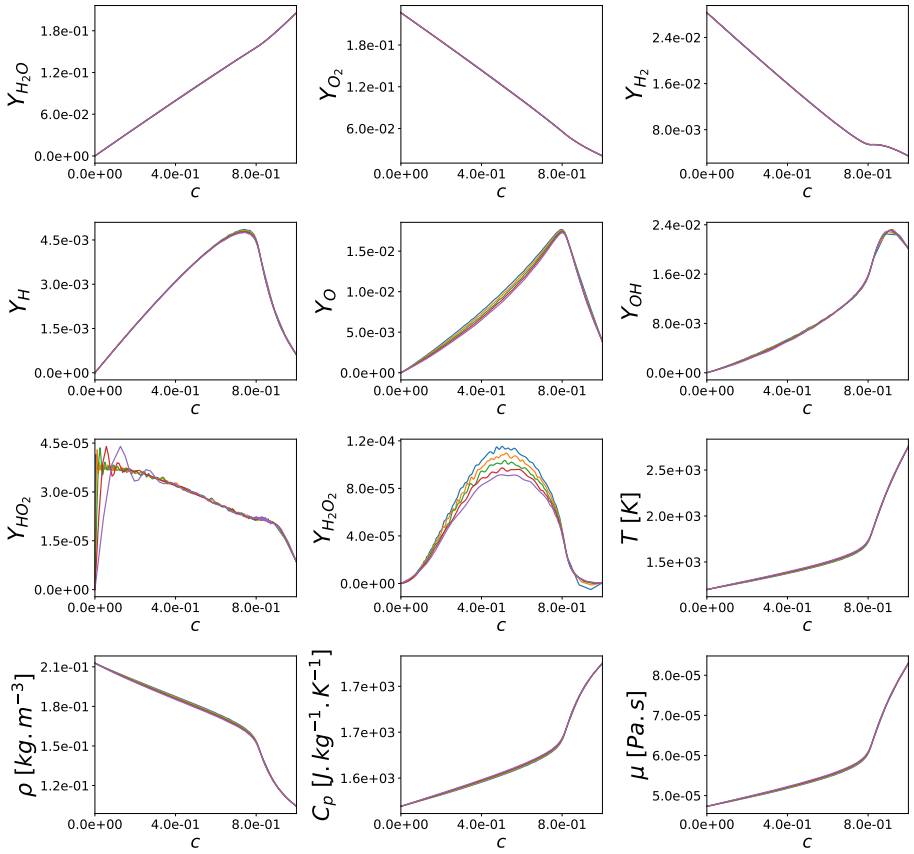

FIGURE 6.14 – *Fonctions $\psi_X^{(t)}$ des différentes grandeurs pour des instants entre $t = 40\mu s$ et $t = 45\mu s$ dans le cas d'un réacteur homogène et adiabatique à pression constante et atmosphérique d'un mélange stœchiométrique air-hydrogène, initialement à $T = 1200K$.*

Il suffit donc de mettre dans la table la valeur moyenne du processus $\tilde{X}$ pour reproduire la moyenne de ces espèces chimiques pour les espèces H_2O, H_2 et O_2. Les espèces H, OH, et dans une moindre mesure l'espèce O, présentent elles aussi une certaine invariance concernant les fonctions $\psi_X^{(t)}$, c'est pourquoi il paraît également raisonnable de tabuler également la valeur moyenne du processus $\tilde{X}$ pour celles-ci. L'espèce HO_2 présentent également une relativement bonne invariance de la fonction $\psi_X^{(t)}$ avec le temps t, les différences présentes étant essentiellement dû à la méthode numérique employée, qui donne une bonne estimation dans les régions où la densité de nuage de points est suffisante uniquement. L'espèce H_2O_2 quant à elle présente des fonctions $\psi_X^{(t)}$ avec des variations relatives entre elles plus importantes pour les valeurs de c intermédiaires, entre 0.2 et 0.8 essentiellement. Néanmoins, il est également nécessaire de faire un choix pour cette espèce, et compte tenu de la contrainte sur la somme des Y_k qui doit être présente dans la table, et en remarquant que la

prise de moyenne commune sur les processus $\tilde{X}$ conservent cette contrainte, il est naturel de mettre dans la table pour cette espèce la moyenne du processus $\tilde{X}$. Concernant les autres grandeurs thermodynamiques nécessaires au calcul de l'écoulement, on peut voir que dans leur ensemble, elles ont un comportement similaire à celui des espèces majoritaires que sont H_2O, H_2 et O_2, c'est à dire que les fonctions $\psi_X^{(t)}$ à différents instants coïncident entre elles. On peut donc pour ces grandeurs aussi prendre $m_{\tilde{X}}$ comme fonction à tabuler. Un tel choix implique, du fait de non linéarité, que la table construite ne respectera pas nécessairement les mêmes lois thermodynamiques que celles utilisées classiquement, comme la loi des gaz parfaits par exemple. Dans le cas présent, la différence entre $E\left[\tilde{X}\right]$ et la valeur "physique" obtenue grâce aux compositions moyennes $E\left[\tilde{Y_k}\right]$, l'enthalpie et la pression étant fixée, ne diffèrent que très peu comme le montre la figure 6.15, où sont tracées les différences relatives. On peut donc considérer dans ce cas que la table construite respecte bien les lois thermodynamiques habituelles, et est donc "physique".

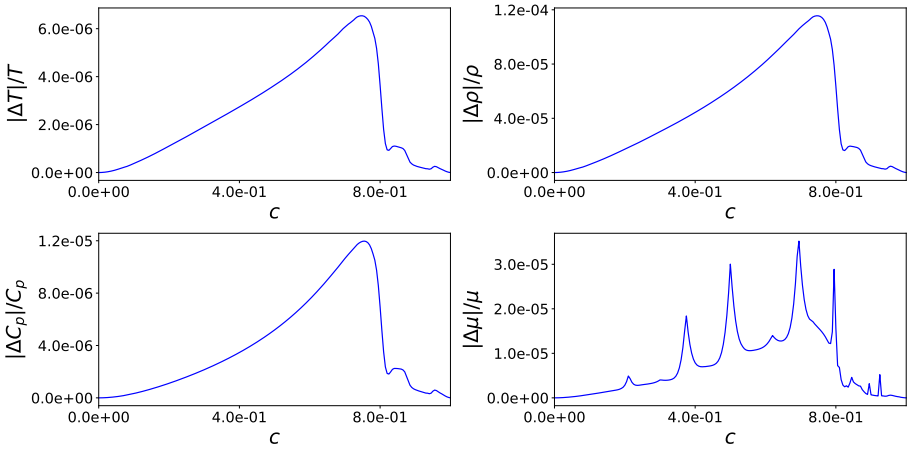

Figure 6.15 – *Valeur absolue des différences relatives entre $m_{\tilde{X}}$ et la grandeur x calculée à partir des fractions massiques $m_{\tilde{Y_k}}$, de l'enthalpie et de la pression pour des variables thermodynamiques pouvant influencer l'écoulement.*

La table ainsi construite vérifie $\tilde{x}_{tab} = m_{\tilde{X}}$ pour chacune des grandeurs tabulées. Il devient possible de comparer les évolutions temporelles des moyennes des variables aléatoires X_t, qui sont obtenus à l'aide de la chimie détaillée, ainsi que les évolutions temporelles des valeurs moyenne des variables aléatoires $m_{\tilde{X}}(C_t)$, qui correspondent au résultat obtenus grâce à la chimie tabulée introduite, ce qui est fait sur la figure 6.16 pour le même ensemble de grandeurs que précédemment.

Comme le montre la figure 6.16, le choix fait concernant la valeur tabulée des grandeurs permet de correctement reproduire les moyennes temporelles des concentrations massiques de l'ensemble des espèces chimiques. Les moyennes

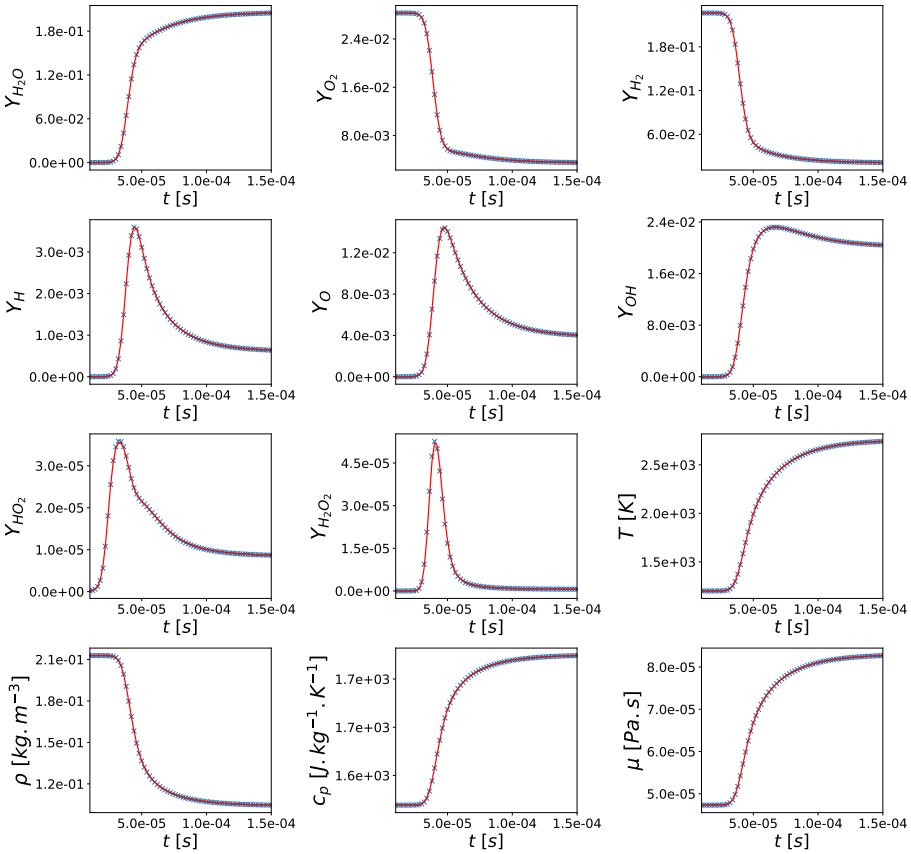

FIGURE 6.16 – *Évolutions temporelles des moyennes des X_t (croix) et des $m_{\tilde{X}}(C_t)$ (ligne pleine) pour un réacteur homogène adiabatique à pression constante et atmosphérique d'un mélange stœchiométrique air-hydrogène, initialement à $T = 1200\ K$.*

temporelles des autres grandeurs, que sont la température, la masse volumique, la capacité calorifique massique à pression constante et la viscosité dynamique, qui sont des grandeurs "actives" influençant l'écoulement, sont également très bien reproduites. Pour des conditions initiales différentes de température et de richesse, la bonne reproduction de l'ensemble des moyennes temporelles a également été observée pour l'ensemble des grandeurs.

Il convient désormais de s'intéresser à d'autres statistiques que la seule moyenne temporelle.

6.4 Reproduction de la variance temporelle

La variance, ou indifféremment l'écart-type, est en plus de la moyenne une statistique souvent étudiée, car elle caractérise la dispersion autour de la moyenne. La table étant maintenant fixée afin de reproduire les valeurs moyennes temporelles m_X des processus stochastiques X, on peut directement

observer et comparer les variances à différents instants t des variables aléatoires X_t et $m_{\tilde{X}}(C_t)$. Cette comparaison est effectuée sur la figure 6.17, où sont tracés les écart-types et non les variances.

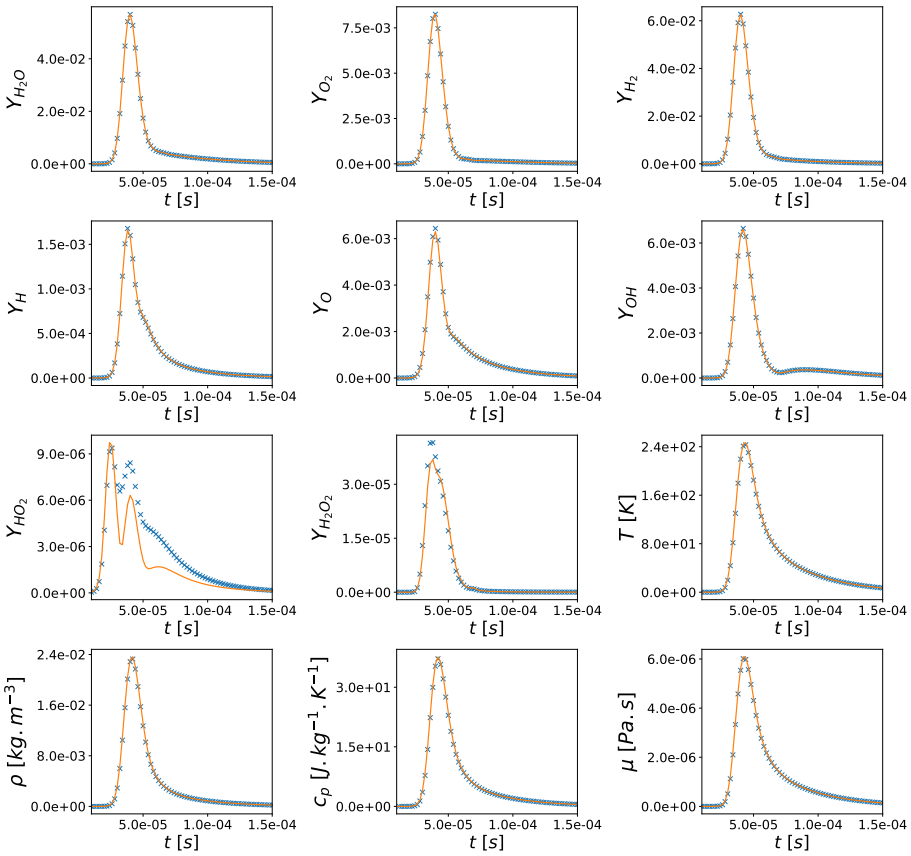

Figure 6.17 – *Évolutions temporelles des écart-types des variables aléatoires X_t (croix) et de $m_{\tilde{X}}(C_t)$ (ligne pleine) pour un réacteur homogène adiabatique à pression constante et atmosphérique d'un mélange stœchiométrique air-hydrogène, initialement à $T = 1200\ K$.*

La figure 6.17 montre que l'évolution temporelle de l'écart-type est correctement reproduite pour l'ensemble des grandeurs à l'exception des espèces HO_2 et H_2O_2. Pour des conditions initiales différentes de température et de richesse, le même constat a été fait pour l'ensemble des grandeurs, à savoir une bonne reproduction de l'évolution temporelle de l'écart-type, excepté pour les fractions massiques de HO_2 et H_2O_2 qui ne sont pas toujours correctement reproduites.

Afin d'expliquer la différence entre la variance de X_t et de $m_{\tilde{X}}(C_t)$ pour HO_2 et H_2O_2, il est intéressant de décomposer la variance de X_t à l'aide de la formule de décomposition de la variance [144] présentée dans l'expression

(6.27).

$$Var\left[X_t\right] = Var\left[E\left[X_t|C_t\right]\right] + E\left[Var\left[X_t|C_t\right]\right] \tag{6.27}$$

Le premier terme dans le membre de droite de l'expression (6.27) correspond à ce que certains statisticiens appellent la "composante expliquée" par C_t de la variance de X_t, puisque cette composante correspond à la variance de l'espérance conditionnelle de X_t selon C_t, qui est la variable aléatoire C_t-mesurables, pouvant donc s'exprimer comme une fonction de C_t, la plus proche de X_t. Dans le cas où X_t est indépendant de C_t, l'espérance conditionnelle $E\left[X_t|C_t\right]$ est alors presque sûrement égale à l'espérance $E\left[X_t\right]$ de X_t, et sa variance est donc nulle, ce qui entraîne que la composante expliquée de la variance de X_t par C_t est nulle, ce qui est en accord avec le fait que X_t est indépendant de C_t. L'autre cas extrême correspond au cas où X_t est C_t-mesurable, ce qui entraîne que presque sûrement, $E\left[X_t|C_t\right] = X_t$, et donc la composante expliquée de la variance de X_t par C_t explique l'intégralité de la variance.

Le second terme correspond à la "composante non expliquée" ou "intrinsèque" de X_t, qui ne peut s'expliquer par C_t. Celui-ci est l'espérance de la variance conditionnelle de X_t sachant C_t, qui, tout comme l'espérance conditionnelle de X_t sachant C_t, est une variable aléatoire définit par l'expression (6.28).

$$Var\left[X_t|C_t\right] = E\left[\left(X_t - E\left[X_t|C_t\right]\right)^2|C_t\right] \tag{6.28}$$

La variance conditionnelle de X_t sachant C_t est donc l'espérance conditionnelle du carré de la différence de X_t à son espérance conditionnée par C_t sachant C_t. La variance conditionnelle de X_t sachant C_t est donc par construction une variable aléatoire positive, ce qui implique que la composante inexpliquée ne peut être nulle que si la variance conditionnelle est presque sûrement nulle.

Tout comme il a été possible d'estimer l'espérance conditionnelle dans la section précédente, il est possible d'estimer la variance conditionnelle de la même manière, la variance conditionnelle de X_t sachant C_t s'écrivant comme une fonction de C_t :

$$Var\left[X_t|C_t\right] = \eta_X^{(t)}(C_t) \tag{6.29}$$

Sur la figure 6.18 sont présentés des nuages de points de réalisations de $(C_t, (X_t - E[X_t|C_t])^2)$ ainsi que les courbes des fonctions $\eta_X^{(t)}$ pour l'instant $t = 40\mu s$, obtenues à partir de la résolution d'un problème d'optimisation similaire au problème (6.24) pour la variance conditionnelle, résolu à l'aide de l'algorithme de Marquardt encore une fois en considérant des splines cubiques.

Avec la donnée des espérances conditionnelles et des variances condition-

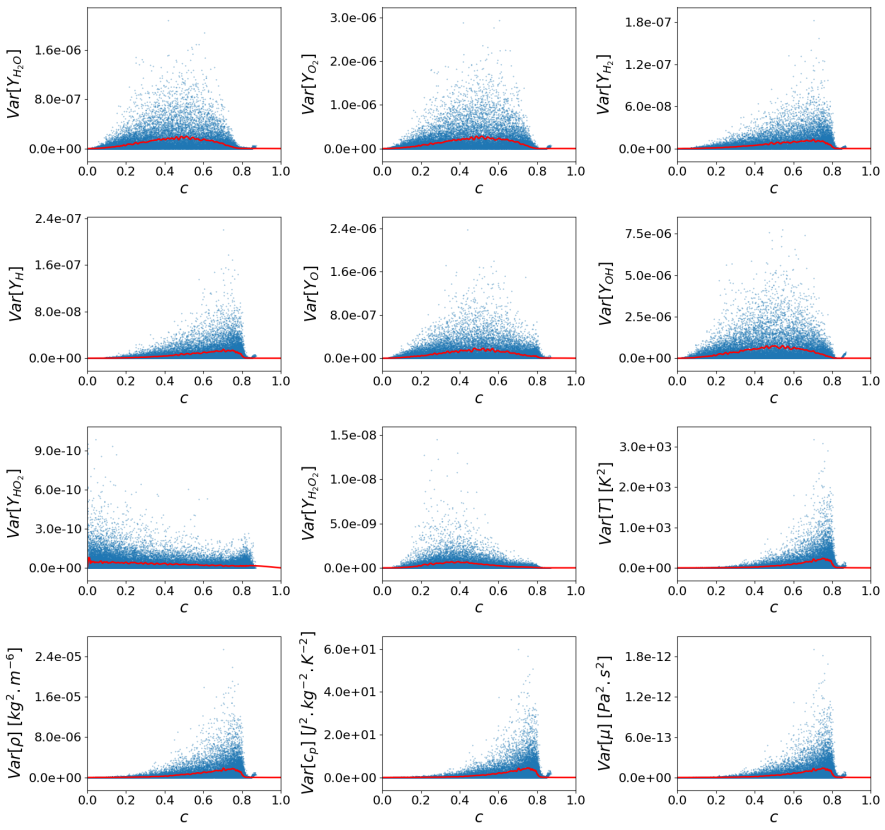

Figure 6.18 – *Nuages de points de réalisations des couple de variables aléatoires* $(C_t, (X_t - E[X_t|C_t])^2)$ *et fonctions* $\eta_X^{(t)}$ *(lignes pleines rouges) calculés à l'aide d'une chimie détaillée pour un réacteur homogène adiabatique à pression constante et atmosphérique d'un mélange stœchiométrique air-hydrogène initialement à* $T = 1200\ K$*, et pour l'instant* $t = 40\mu s$*.*

nelles sachant C_t pour chaque instant, il est possible d'obtenir la contribution expliquée et non-expliquée de la variance en chaque instant. Ces différentes contributions sont présentées sur la figure 6.19, pour les écart-types et non la variance, et l'on retrouve bien que la somme (en terme de variance) de ces deux contributions donne l'écart-type total pour l'ensemble des grandeurs et des instants, montrant que les estimations des espérances ainsi que des variances conditionnelles sont correctement réalisées. Pour toutes les espèces exceptées HO_2 et H_2O_2, la contribution non expliquée est négligeable de sorte que la partie expliquée permet à elle seule de reproduire l'écart-type total. Cela suggère que pour ces grandeurs, X_t est C_t-mesurables et peut donc s'écrire comme une fonction de la variable aléatoire C_t.

Ce résultat est particulièrement intéressant pour les grandeurs "actives" influençant l'écoulement, telles que la température, la masse volumique ou encore la viscosité dynamique. La quasi-nullité de la composante inexpliquée de

la variance pour ces grandeurs signifie que celles-ci peuvent s'exprimer comme des fonctions de la variable aléatoire C_t, et donc que l'ensemble des incertitudes sur ces grandeurs peuvent être expliquées par C_t, ce qui suggère la possibilité de ne considérer que C_t comme incertain pour la propagation d'incertitude.

Pour les espèces H_2O_2, et plus particulièrement pour HO_2, la contribution de la partie non expliquée est non négligeable dans l'écart-type total, et la seule prise en compte de la partie expliquée ne saurait rendre compte de l'écart-type des concentrations de ces espèces au cours du temps. Il est donc nécessaire pour ces deux grandeurs de considérer la contribution de la partie non expliquée de la variance.

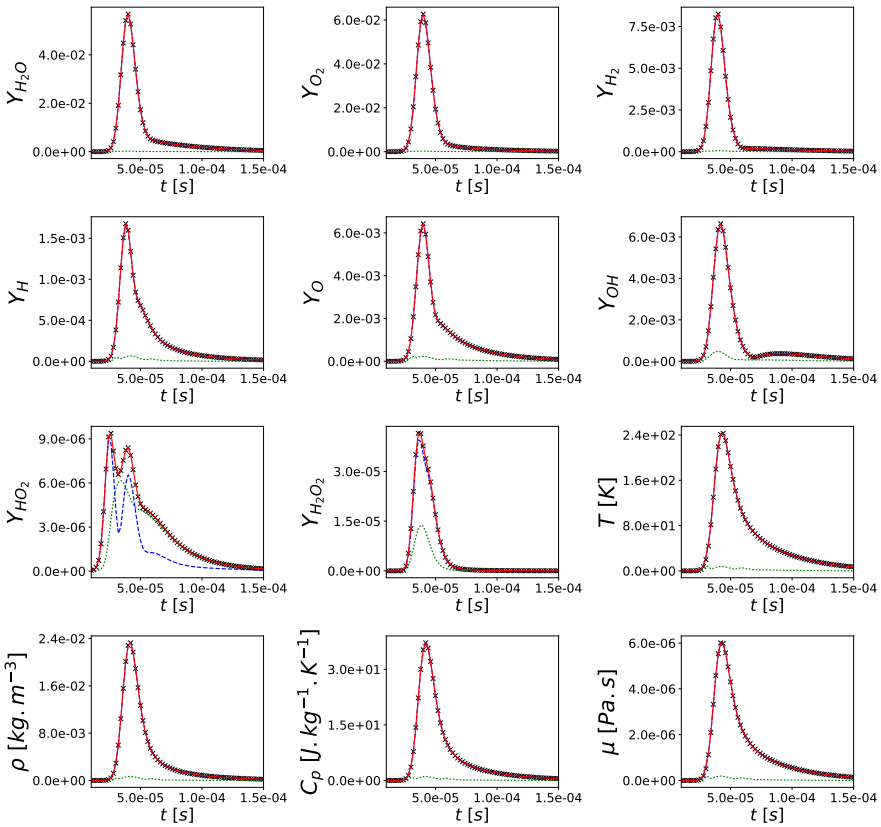

FIGURE 6.19 – *Évolutions temporelles des écart-types des X_t (croix), des contributions expliquées de l'écart-types de X_t sachant C_t (pointillés), des contributions non expliquées de l'écart-type de X_t sachant C_t (points) et de la somme des deux contributions (ligne pleine) pour un réacteur homogène et adiabatique à pression constante et atmosphérique d'un mélange stœchiométrique air-hydrogène, initialement à $T = 1200\ K$.*

En comparant les contributions expliquées des écart-types sur la figure 6.19 avec les évolutions temporelles des écart-types des variables aléatoires $m_{\tilde{X}}(C_t)$ présentés sur la figure 6.17, on observe que les deux ont un comportement similaire. Ces deux écart-types sont en fait tout deux les écart-types

de fonctions de la variable aléatoire C_t, qui sont de plus deux fonctions ne présentant pas de grands écarts entre elles, le choix de la fonction tabulée $m_{\tilde{X}}$ ayant été fait pour reproduire correctement l'espérance conditionnelle $E[X_t|C_t]$ pour les différents instants t. Il n'est donc pas étonnant d'observer ce comportement similaire, et il peut être intéressant d'avoir un raisonnement similaire concernant la contribution inexpliquée, afin de pouvoir reproduire également celle-ci. En effet, tout comme pour l'espérance conditionnelle de X_t sachant C_t, les variances conditionnelles dépendent encore une fois de l'instant t considéré. Sur la figure 6.20 sont présentées les estimations des fonctions $\eta_X^{(t)}$ pour les différentes grandeurs et pour différents instants t. Les oscillations présentes sont liées à la méthode d'estimation employée à l'aide des splines cubiques. La réduction de telles oscillations pourrait être faite en imposant un terme pénalisant les fonctions trop oscillantes dans la méthode d'optimisation, comme pour les méthodes de lissage par splines [120]. En dehors des parties oscillantes qui sont pour l'essentielle localisées dans les zones où la densité de probabilité π_{C_t} est presque nulle, les fonctions $\eta_X^{(t)}$ ne présentent pas de différences significatives pour des instants t différents, tout comme c'était le cas pour les fonctions $\psi_X^{(t)}$ pour l'espérance conditionnelle sachant C_t.

En accord avec les résultats de la figure 6.17, on peut supposer pour les variances conditionnelles $Var[X_t|C_t]$ l'hypothèse (6.30), similaire à l'hypothèse (6.25) pour l'espérance conditionnelle, η_X étant une fonction indépendante du temps t.

$$\forall t, Var[X_t|C_t = c] = \eta_X(c) \tag{6.30}$$

En utilisant les relations (6.25), (6.26), (6.28) et (6.30), et en notant que

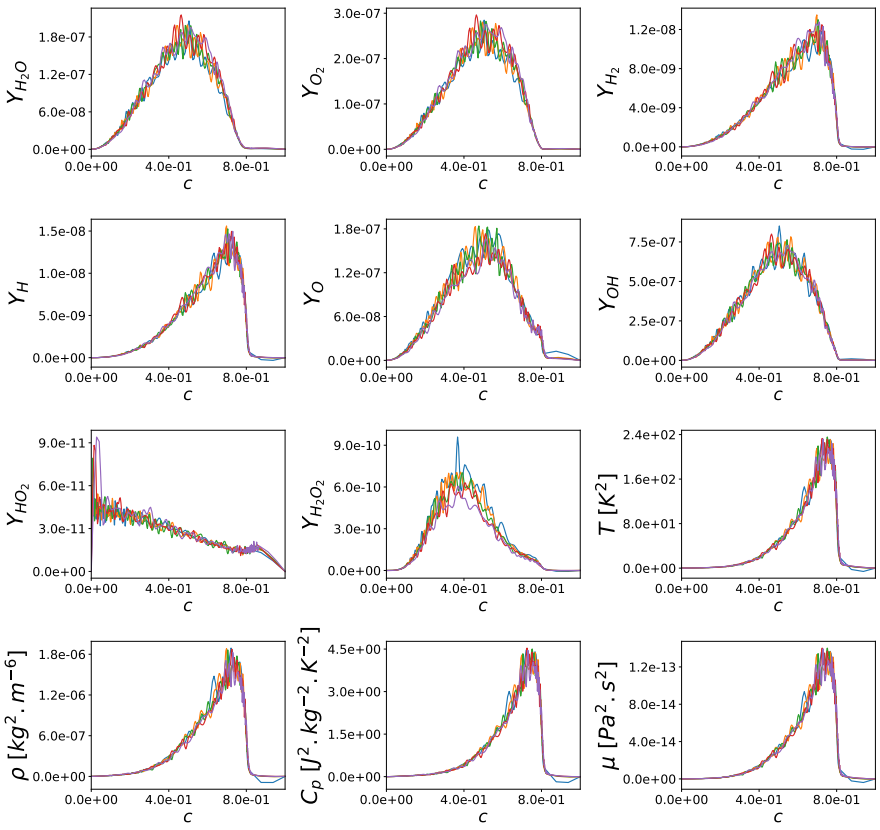

FIGURE 6.20 – *Fonctions $\eta_X^{(t)}$ des différentes grandeurs pour les instants $t = 36\mu s$, $t = 38\mu s$, $t = 40\mu s$, $t = 42\mu s$, et $t = 44\mu s$ dans le cas d'un réacteur homogène et adiabatique à pression constante et atmosphérique d'un mélange stœchiométrique air-hydrogène, initialement à $T = 1200\ K$.*

C_{T_c} est presque sûrement égale à c, il vient pour la variance du processus $\tilde{X}_c$:

$$
\begin{aligned}
Var\left[\tilde{X}_c\right] &= Var\left[X_{T_c}\right] = E\left[X_{T_c}^2\right] - E\left[\tilde{X}_c\right]^2 = E\left[E\left[X_{T_c}^2|C_{T_c}\right]\right] - \psi_X(c)^2 \\
&= E\left[E\left[E\left[X_{T_c}^2|C_{T_c}\right]|T_c\right]\right] - \psi_X(c)^2 \\
&= \int_{\mathbb{R}^+} E\left[E\left[X_{T_c}^2|C_{T_c}\right]|T_c = t\right]\pi_{T_c}(t)dt - \psi_X(c)^2 \\
&= \int_{\mathbb{R}^+} E\left[E\left[X_t^2|C_t\right]|T_c = t\right]\pi_{T_c}(t)dt - \psi_X(c)^2 \\
&= \int_{\mathbb{R}^+} E\left[Var\left[X_t|C_t\right] + E\left[X_t|C_t\right]^2|T_c = t\right]\pi_{T_c}(t)dt - \psi_X(c)^2 \\
&= \int_{\mathbb{R}^+} E\left[\eta_X(C_t) + \psi_X(C_t)^2|T_c = t\right]\pi_{T_c}(t)dt - \psi_X(c)^2 \\
&= \int_{\mathbb{R}^+} E\left[\eta_X(C_{T_c}) + \psi_X(C_{T_c})^2|T_c = t\right]\pi_{T_c}(t)dt - \psi_X(c)^2 \\
&= \int_{\mathbb{R}^+} E\left[\eta_X(c) + \psi_X(c)^2|T_c = t\right]\pi_{T_c}(t)dt - \psi_X(c)^2 \\
&= E\left[E\left[\eta_X(c) + \psi_X(c)^2|T_c\right]\right] - \psi_X(c)^2 \\
&= E\left[\eta_X(c) + \psi_X(c)^2\right] - \psi_X(c)^2 = \eta_X(c)
\end{aligned}
$$

(6.31)

Ainsi, sous les hypothèses d'existence de fonctions ψ_X et η_X indépendantes du temps t, il s'avère que la variance de la variable aléatoire $\tilde{X}_c$ est donné par la valeur $\eta_X(c)$. Ainsi, la variance conditionnelle de X_t sachant C_t peut s'exprimer sous ces hypothèses comme $\sigma^2_{\tilde{X}}(C_t)$.

En rajoutant à la table la variance $\sigma^2_{\tilde{X}}$ de chacun des $\tilde{X}$, on peut espérer pouvoir reproduire la partie non expliquée de la variance, et donc la totalité de la variance des grandeurs X au cours du temps. Sur la figure 6.21 sont représentées les variances des concentrations des différentes grandeurs, ainsi que les contributions expliquées, inexpliquées et leur somme pour la même configuration que précédemment.

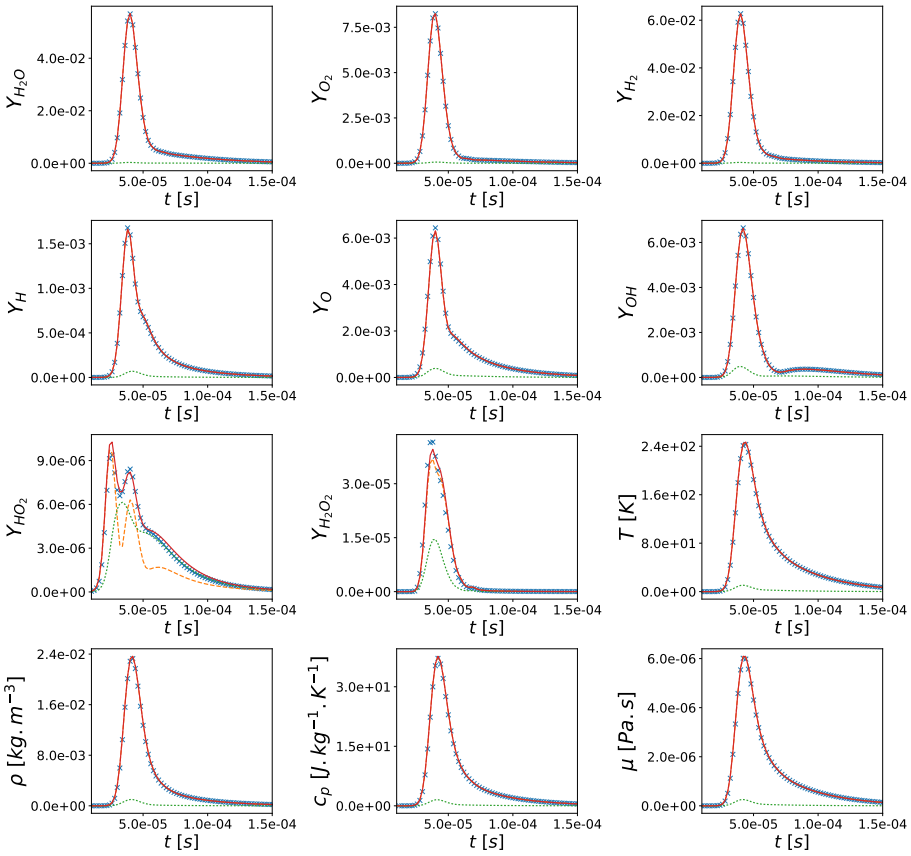

Figure 6.21 – *Évolutions temporelles des écart-types des variables aléatoires X_t (croix) et de $m_{\tilde{X}}(C_t)$ (tirets), des racines carrés des espérances de $\sigma^2_{\tilde{X}}(C_t)$ (pointillés), et racine carré de la somme des carrés des deux dernières courbes (ligne pleine) pour un réacteur homogène et adiabatique à pression constante et atmosphérique d'un mélange stœchiométrique air-hydrogène, initialement à $T = 1200$ K.*

Sur la figure 6.21, on peut voir que l'incorporation de la variance de $\tilde{X}$

dans la table permet une bonne reproduction de la contribution inexpliquée de la variance de X_t présente sur la figure 6.19, permettant d'améliorer significativement la reproduction de l'écart-type pour les fractions massiques des espèces HO_2 et H_2O_2. La prise en compte des deux contributions obtenues permet une bonne reproduction de l'écart-type pour l'ensemble des grandeurs.

6.5 Conclusion

L'utilisation de la chimie tabulée n'est pas immédiate lorsque la cinétique chimique est incertaine. Les résultats présentés dans ce chapitre suggèrent qu'il est possible de reproduire les incertitudes sur de nombreuses grandeurs, notamment les grandeurs "actives" influençant l'écoulement, pour lesquelles la reproduction des incertitudes est primordiale afin qu'elles se répercutent sur l'écoulement, en considérant une tabulation de ces grandeurs indépendantes des paramètres incertains, et pouvant donc être considérée comme déterministe. L'utilisation d'une telle table déterministe est possible par l'introduction des incertitudes via le processus stochastique C, lui même découlant de la résolution de l'équation différentielle ordinaire (6.5), suggérant de ne considérer que le terme source $\dot{\omega}_{Y_c}$ incertain, celui-ci contenant l'ensemble des informations des incertitudes de la cinétique chimique influençant l'écoulement.

Pour un processus stochastique X associé à une grandeur, deux cas de figures ont été observées dans ce chapitre :

- Le processus stochastique peut s'exprimer comme une fonction du processus stochastique C (ou indifféremment Y_c). Cela s'est traduit par des espérances conditionnelles de X_t par rapport à C_t coïncidant pour différents instants t, ainsi que des variances conditionnelles de X_t sachant C_t négligeables. En pratique, ces conditions sont nécessairement vérifiées lorsque le processus stochastique $\tilde{X}$ présente une variance négligeable (pour la grandeur considérée) pour toutes les valeurs de c.
- Le processus stochastique ne peut être exprimée comme une fonction du processus stochastique C (ou indifféremment Y_c). Pour être dans cette situation, il suffisait que la variance du processus stochastique $\tilde{X}$ ne soit pas négligeable (pour la grandeur considérée).

Les résultats présentés suggèrent de tabuler la moyenne $m_{\tilde{X}}$ des processus stochastiques $\tilde{X}$ pour chaque grandeur, de nombreuses variables aléatoires X_t, notamment les variables aléatoires associées aux grandeurs "actives" influençant l'écoulement, pouvant presque s'écrire $m_{\tilde{X}}(C_t)$. Les courbes des figures 6.16 et 6.17 montrent en effet que les moments d'ordre 1 et 2 coïncident entre les variables aléatoires X_t et $m_{\tilde{X}}(C_t)$, mais en fait la quasi-nullité de la composante inexpliquée de la variance ainsi que la coïncidence de l'ensemble des fonctions $\psi_X(t)$ pour ces grandeurs permettent de suggérer la relation plus forte $X = m_{\tilde{X}}(C)$ entre les processus stochastiques X et C. Cette forte relation semble notamment s'appliquer aux grandeurs actives, permettant la propaga-

tion d'incertitudes au sein de la cinétique chimique au travers d'une simulation aux grandes échelles d'écoulement réactifs. Concernant les variables de post-traitement, comme les fractions massiques d'espèces par exemple, il n'est pas nécessaire qu'elles se trouvent dans le premier cas de figure pour reproduire correctement l'impact des incertitudes sur l'écoulement. Si une variable de post-traitement se trouve être dans le premier cas de figure, son comportement incertain pourra être correctement caractérisé, puisque la relation $X = m_{\tilde{X}}(C)$ sera vérifiée. Si elle se trouve dans le second cas de figure, sa moyenne pourra être obtenue, ainsi que sa variance au prix de rajouter une variable tabulée : la variance $\sigma_{\tilde{X}}2$ du processus $\tilde{X}$.

Les résultats ont été illustrés dans ce chapitre à l'aide d'une unique condition initiale de température et de richesse. Des comportements similaires ont toutefois été observés pour d'autres températures et richesses initiales. De plus, l'ensemble de la démarche adoptée, ainsi que les hypothèses réalisées peuvent être généralisées sans difficultés à une autre configuration canonique de flamme qu'un réacteur homogène adiabatique à pression constante.

L'utilisation d'une telle chimie tabulée dans un contexte de propagation d'incertitudes de cinétique chimique en simulation aux grandes échelles d'écoulement réactifs requiert encore un élément, qui est la réduction de la dimension stochastique. La méthode présentée dans ce chapitre propose de ne considérer que le processus stochastique Y_c comme étant incertain. La reproduction de ce processus stochastique avec peu de paramètres incertains est l'objet du prochain chapitre.

Chapitre 7

Représentation du terme source incertain de la variable d'avancement à l'aide des variables aléatoires initiales

Prérequis :

- Calcul d'intégrales multiples (Cubature, Monte Carlo et Quasi-Monte Carlo randomisé)
- Expansion en Polynôme du Chaos pour des processus stochastiques
- Analyse de sensibilité globale

Notions clés et apports du chapitre :

- Modélisation du terme source incertain d'une variable d'avancement à l'aide d'une expansion en polynômes du chaos
- Utilisation et validation de cette modélisation du terme source incertain d'une variable d'avancement sur le système chimique canonique étudié

7.1 Introduction

L'objectif de ce chapitre et du chapitre suivant, suivant les conclusions du chapitre 6, est la reproduction du processus stochastique C avec un nombre de variables aléatoires limité. Classiquement en chimie tabulée, l'évolution temporelle de la variable d'avancement se fait à l'aide de son terme source $\dot{\omega}_c$, préalablement tabulé. En effet, dans le cas déterministe, l'évolution temporelle de la variable d'avancement c peut être obtenue à partir du terme source $\dot{\omega}_c$ au travers de la résolution de l'équation différentielle ordinaire (7.1).

$$\frac{dc}{dt} = \dot{\omega}_c \tag{7.1}$$

Dans le cas où la cinétique chimique est incertaine, l'équation différentielle ordinaire s'applique à chacune des réalisations de C, de sorte que si $\mathbf{a} = (a_1, \ldots, a_n)$ est une réalisation du vecteur aléatoire $\mathbf{A} = (A_1, \ldots, A_n)$ représentant les paramètres cinétiques incertains, la réalisation de C correspondant aux $\mathbf{a} = (a_1, \ldots, a_n)$ est donnée par la relation (7.4).

$$\frac{dC(., \mathbf{a})}{dt} = \dot{\omega}_c(c, \mathbf{a}) \tag{7.2}$$

En fait, dans la suite, le terme source de la variable d'avancement $\dot{\omega}_{Y_c}$ de la variable d'avancement de la réaction non normalisé Y_c sera étudié. Ce dernier est simplement le terme source $\dot{\omega}_c$ multiplié par un facteur indépendant de c, et la relation les liant est la suivante, Y_c^0 et Y_c^∞ correspondant aux valeurs initiales et finales de Y_c :

$$\dot{\omega}_c = \frac{\dot{\omega}_{Y_c}}{Y_c^\infty - Y_c^0} \tag{7.3}$$

La relation (7.2) peut se réécrire en considérant $\dot{\omega}_{Y_c}$ plutôt que $\dot{\omega}_c$, donnant :

$$\frac{dC(., \mathbf{a})}{dt} = \frac{\dot{\omega}_{Y_c}(c, \mathbf{a})}{Y_c^\infty - Y_c^0} \tag{7.4}$$

C'est donc ce terme source que l'on va chercher à exprimer en fonction des variables aléatoires permettant la reproduction des incertitudes de C, les incertitudes du processus stochastique C se retrouvant immédiatement à partir de la résolution de l'équation différentielle ordinaire (7.4). L'ensemble de l'étude se base sur une méthode non intrusive, de sorte que les propriétés du processus stochastique C seront déterminées à l'aide de réalisations de ce processus sto-

chastique, lesquelles réalisations seront obtenues par l'intégration de l'équation différentielle ordinaire (7.4).

Différents jeux de variables aléatoires peuvent être utilisées pour la paramétrisation des incertitudes, et la construction de tels jeux peut reposer sur différentes considérations. Dans une première partie, cette construction est faite en considérant le délai d'auto-allumage du réacteur, celui-ci étant un paramètre macroscopique dont la bonne reproduction est primordiale pour des simulations complexes notamment de moteur à combustion interne[135], et qui est relié au processus stochastique $\dot{\omega}_{Y_c}$ au travers de l'équation suivante :

$$\tau^{C=c^*} = (Y_c^{\infty} - Y_c^0) \int_0^{c^*} \frac{dc}{\dot{\omega}_{Y_c}(c)} \tag{7.5}$$

Dans un second temps, la construction des variables aléatoires se fait en ne considérant pas le délai d'auto-allumage, mais directement le processus stochastique C.

7.2 Utilisation de la chimie tabulée

La chimie tabulée nécessite l'utilisation d'un maillage de l'espace des paramètres de contrôle, où la valeur de chaque grandeur d'intérêt est stockée. Pour n'importe quelles valeurs des paramètres de contrôle, il devient alors possible d'accéder à la valeur d'une grandeur d'intérêt via une interpolation utilisant les valeurs stockées dans la table. La précision des résultats obtenus dépend donc des interpolations réalisées, dont la qualité dépend fortement du maillage de l'espace des paramètres de contrôles utilisé. Le maillage doit être adapté au cas d'étude, et suffisamment fin pour être à même de reproduire les grandeurs d'intérêts de manière acceptable.

Le présent chapitre a pour objectif la reproduction du processus stochastique C à l'aide d'une modélisation du terme source incertain $\dot{\omega}_{Y_c}$, dont les réalisations sont régies par l'équation différentielle ordinaire (7.4). Le terme de droite de cette dernière équation différentielle ordinaire est en pratique tabulé en fonction de la variable d'avancement de la réaction normalisée c, et il est donc nécessaire de s'assurer que le maillage en c est adapté à la résolution de cette équation différentielle ordinaire, et cela pour l'ensemble des valeurs possibles que peuvent prendre les paramètres incertains. Afin de valider le maillage utilisé, les étapes suivantes sont réalisées pour un grand nombre de jeu de paramètres incertains choisis aléatoirement ou quasi-aléatoirement (c'est à dire construit à partir d'une séquence à discrépance faible) :

- Le calcul du réacteur est réalisé en utilisant la chimie détaillée avec un jeu de paramètres incertains $\mathbf{a}$
- La réalisation $C(., \mathbf{a})$ du processus stochastique C obtenue est conservée

— Le terme source $\dot{\omega}_{Y_c}(., \mathbf{a})$ obtenu est tabulé suivant c en utilisant le maillage
— L'équation différentielle ordinaire (7.4) est résolue en utilisant le terme source précédemment tabulée
— La réalisation du processus stochastique $C_{tab}(., \mathbf{a})$ obtenue est conservée

Ces étapes permettent donc d'obtenir un ensemble de couples de réalisations des processus stochastique C et C_{tab}, sur lesquels il est possible de vérifier si la chimie tabulée est fidèle à la chimie détaillée, ce qui signifiera que le maillage est adapté.

Le maillage en c utilisé dans la suite du chapitre possède les caractéristiques suivantes :

— La longueur maximale de maille est de 0.005, et concerne les mailles à droite du segment $[0, 1]$ qui sont toutes de cette longueur
— Les mailles partant de 0 du segment $[0, 1]$ ont une longueur croissante suivant une série géométrique de raison 1.2, jusqu'à ce que cette taille soit égale à 0.005
— La première maille a une longueur inférieure à 10^{-10}, et c'est la seule maille dont la longueur est inférieure à cette valeur

Ce maillage a donc été validé en suivant les étapes précédemment décrite, la résolution numérique des équations différentielles ordinaires ayant été effectuée à l'aide du solveur RADAU5, et le choix a été fait ici d'effectuer la validation par la comparaison des délais d'auto-allumage $\tau^{C=c^*}$. Pour chaque couple de réalisations (c, c_{tab}), il est possible de déterminer un couple de délai d'auto-allumage $(\tau^{c=c^*}, \tau_{tab}^{c=c^*})$ associé. Il est alors possible de tracer les nuages de points correspondants à ces couples de délai d'auto-allumage pour l'ensemble des réalisations effectuées.

Ces nuages de points sont visibles sur la figure 7.1 pour une valeur de $c^* = 0.1$, et pour quatre conditions initiales qui seront explicitées plus tard dans ce chapitre. Les différents nuages de points se placent pour l'ensemble des conditions initiales sur une droite, ce qui indique une forte corrélation linéaire entre les variables aléatoires $\tau^{C=0.1}$ et $\tau_{tab}^{C=0.1}$. Il est possible d'effectuer une régression linéaire sur ces nuages de points, afin d'obtenir des informations sur les droites par lesquelles passent ces nuages de points.

La régression linaire a été effectuée pour les quatre conditions initiales, et pour différentes valeurs de c^*, et les résultats obtenus sont présentés dans le tableau 7.1. Trois informations sont présentes pour chaque nuage de points, qui sont :

— Le coefficient directeur a de la droite de régression
— Le ratio $b/\bar{\tau}$ entre l'ordonnée à l'origine b de la droite de régression et le délai d'auto-allumage moyen $\bar{\tau} = E\left[\tau^{C=c^*}\right]$
— Le coefficient de corrélation linéaire r entre $\tau^{C=c^*}$ et $\tau_{tab}^{C=c^*}$

Les coefficients directeurs a sont légèrement inférieur à 1, compris entre 0.99383 et 0.99999 et les ratios $b/\bar{\tau}$ sont tous proches de 0, compris entre

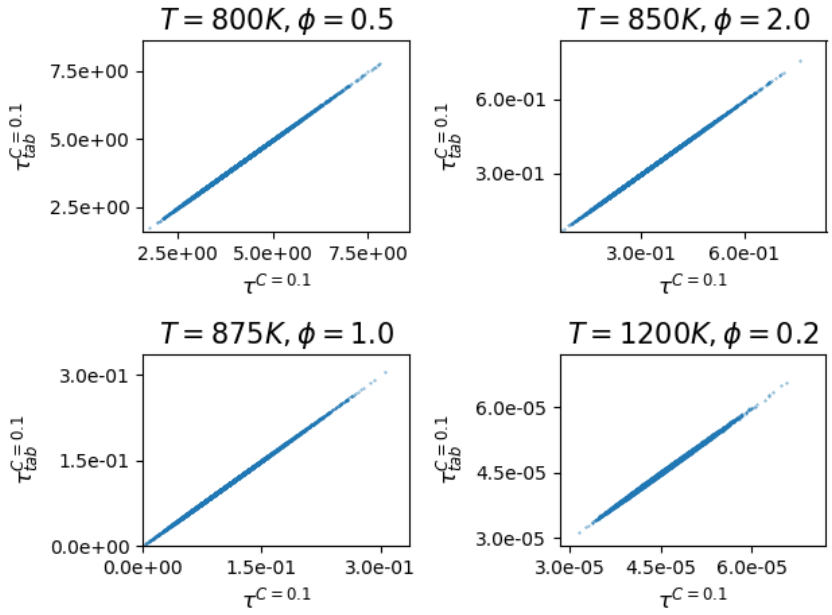

FIGURE 7.1 – *Nuages de points des délais d'auto-allumage $\tau^{C=0.1}$ calculés avec une chimie détaillée, et des délais d'auto-allumage $\tau_{tab}^{C=0.1}$ calculés avec une chimie tabulée pour les quatre points de température et de richesse identifiés par l'analyse de sensibilité globale de la section suivante.*

-0.0095 et 0.0016. L'ordonnée à l'origine seule n'a pas de réelle signification, et il est plus intéressant de la comparer à un ordre de grandeur de la grandeur étudiée, ce qui est fait ici en la rapportant au délai d'auto-allumage moyen dans chacun des cas. Les droites de régressions sont donc toutes légèrement en dessous de la droite d'équation $\tau^{C=c^*} = \tau_{tab}^{C=c^*}$, ce qui indique que la chimie détaillée est légèrement plus "rapide" que la chimie tabulée en moyenne. La valeur du coefficient de corrélation linéaire r revêt également une importance primordiale dans cette validation. Pour rappel, le coefficient de corrélation linéaire r entre deux variables aléatoires X et Y est donné par la formule (7.6).

$$r = \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}} \tag{7.6}$$

Ce coefficient est compris entre -1 et 1, et vaut 1 lorsque la corrélation entre les variables aléatoires est parfaite. En l'occurrence, les valeurs présentes dans le tableau sont toutes très proches de 1, puisqu'elles sont comprises entre 0.9906 et 1.00000. Cela traduit le fait que les points sont très peu dispersés et sont donc très proches de la droite de régression.

L'ensemble de ces trois informations permet de conclure que, pour le maillage considéré, on peut considérer que les variables aléatoires $\tau^{C=c^*}$ et

	$T = 800\ K$ $\phi = 0.5$	$T = 850\ K$ $\phi = 2$	$T = 875\ K$ $\phi = 1$	$T = 1200\ K$ $\phi = 0.2$
$\tau^{c=0.01}$	$a = 0.9938$ $b/\bar{\tau} = -0.0006$ $r = 0.9996$	$a = 0.9953$ $b/\bar{\tau} = -0.0008$ $r = 1.0000$	$a = 0.9970$ $b/\bar{\tau} = -0.0000$ $r = 1.0000$	$a = 0.9942$ $b/\bar{\tau} = -0.0095$ $r = 0.9906$
$\tau^{c=0.1}$	$a = 0.9955$ $b/\bar{\tau} = -0.0009$ $r = 1.0000$	$a = 0.9943$ $b/\bar{\tau} = -0.0002$ $r = 1.0000$	$a = 0.9970$ $b/\bar{\tau} = -0.0000$ $r = 1.0000$	$a = 0.9986$ $b/\bar{\tau} = -0.0001$ $r = 0.9995$
$\tau^{c=0.2}$	$a = 0.9955$ $b/\bar{\tau} = -0.0009$ $r = 1.0000$	$a = 0.9943$ $b/\bar{\tau} = -0.0002$ $r = 1.0000$	$a = 0.9970$ $b/\bar{\tau} = -0.0000$ $r = 1.0000$	$a = 0.9991$ $b/\bar{\tau} = 0.0004$ $r = 0.9997$
$\tau^{c=0.3}$	$a = 0.9955$ $b/\bar{\tau} = -0.0009$ $r = 1.0000$	$a = 0.9943$ $b/\bar{\tau} = -0.0002$ $r = 1.0000$	$a = 0.9970$ $b/\bar{\tau} = -0.0000$ $r = 1.0000$	$a = 0.9991$ $b/\bar{\tau} = 0.0009$ $r = 0.9998$
$\tau^{c=0.4}$	$a = 0.9955$ $b/\bar{\tau} = -0.0009$ $r = 1.0000$	$a = 0.9943$ $b/\bar{\tau} = -0.0002$ $r = 1.0000$	$a = 0.9970$ $b/\bar{\tau} = -0.0000$ $r = 1.0000$	$a = 0.9982$ $b/\bar{\tau} = 0.0010$ $r = 0.9999$
$\tau^{c=0.5}$	$a = 0.9955$ $b/\bar{\tau} = -0.0009$ $r = 1.0000$	$a = 0.9943$ $b/\bar{\tau} = -0.0002$ $r = 1.0000$	$a = 0.9970$ $b/\bar{\tau} = -0.0000$ $r = 1.0000$	$a = 0.9984$ $b/\bar{\tau} = 0.0006$ $r = 0.9999$
$\tau^{c=0.6}$	$a = 0.9955$ $b/\bar{\tau} = -0.0009$ $r = 1.0000$	$a = 0.9943$ $b/\bar{\tau} = -0.0002$ $r = 1.0000$	$a = 0.9970$ $b/\bar{\tau} = -0.0000$ $r = 1.0000$	$a = 0.9987$ $b/\bar{\tau} = 0.0002$ $r = 0.9999$
$\tau^{c=0.7}$	$a = 0.9955$ $b/\bar{\tau} = -0.0009$ $r = 1.0000$	$a = 0.9943$ $b/\bar{\tau} = -0.0002$ $r = 1.0000$	$a = 0.9970$ $b/\bar{\tau} = -0.0000$ $r = 1.0000$	$a = 0.9981$ $b/\bar{\tau} = 0.0016$ $r = 0.9998$
$\tau^{c=0.8}$	$a = 0.9955$ $b/\bar{\tau} = -0.0009$ $r = 1.0000$	$a = 0.9943$ $b/\bar{\tau} = -0.0002$ $r = 1.0000$	$a = 0.9970$ $b/\bar{\tau} = -0.0000$ $r = 1.0000$	$a = 0.9994$ $b/\bar{\tau} = 0.0011$ $r = 0.9997$
$\tau^{c=0.9}$	$a = 0.9955$ $b/\bar{\tau} = -0.0009$ $r = 1.0000$	$a = 0.9943$ $b/\bar{\tau} = -0.0002$ $r = 1.0000$	$a = 0.9970$ $b/\bar{\tau} = 0.0000$ $r = 1.0000$	$a = 1.0000$ $b/\bar{\tau} = -0.0004$ $r = 0.9995$
$\tau^{c=0.99}$	$a = 0.9955$ $b/\bar{\tau} = -0.0009$ $r = 1.0000$	$a = 0.9943$ $b/\bar{\tau} = -0.0000$ $r = 1.0000$	$a = 0.9970$ $b/\bar{\tau} = 0.0000$ $r = 1.0000$	$a = 0.9957$ $b/\bar{\tau} = 0.0019$ $r = 0.9930$

TABLE 7.1 – *Coefficients directeur a, ordonnée à l'origine b de la droite de régression linéaire, et coefficient de corrélations r calculés à partir de* 10×1024 *échantillons de Quasi-Monte Carlo randomisé de* $\tau^{C=c^*}$ *et* $\tau_{tab}^{C=c^*}$ *pour les quatre points de température et de richesse identifiés par l'analyse de sensibilité globale de la section suivante.*

$\tau_{tab}^{C=c^*}$ sont égales. Cela étant vrai pour une large plage de valeur de c^*, cela traduit le fait que les réalisations des processus stochastiques C et C_{tab} sont presque identiques, signifiant que le maillage choisi est bien adapté au problème.

Dans la suite du chapitre, les résultats obtenus sont issus d'une utilisation d'une tabulation de modélisation du terme source incertain $\dot{\omega}_{Y_c}$. Afin de ne pas introduire d'erreur liée à l'utilisation d'un maillage, il eut été préférable de comparer les différents résultats au processus stochastique C_{tab} plutôt qu'à C. Cependant, compte tenu des résultats précédents montrant que le processus stochastique C est très bien reproduit par le processus stochastique C_{tab}, les comparaisons seront directement faites par rapport au processus stochastique C.

7.3 Analyse de sensibilité globale

7.3.1 Délai d'auto-allumage

Le système constitué du réacteur homogène et adiabatique à pression constante évolue spontanément d'un état initial composé d'un mélange de gaz frais vers un état final composé d'un mélange de gaz brûlés. Entre ces deux états, le système passe par une continuité d'états, qu'il est possible de caractériser à l'aide de la valeur de la variable d'avancement c. Dans le présent chapitre, le délai d'auto-allumage considéré correspond au temps nécessaire pour atteindre l'état caractérisé par une valeur c^* de la variable d'avancement, et sera noté $\tau^{C=c^*}$. Dans un cas incertain, ce délai d'auto-allumage correspond à $T(c^*, .)$, T étant le processus stochastique introduit dans le chapitre 6, ce qui fait que ce délai d'auto-allumage est une variable aléatoire. Avec la définition présente du délai d'auto-allumage, la bonne reproduction des incertitudes du processus stochastique C sont suffisantes à la reproduction des incertitudes du délai d'auto-allumage incertain $\tau^{C=c^*}$. En conséquence, la bonne reproduction de ce délai d'auto-allumage incertain $\tau^{C=c^*}$ est une condition nécessaire à la bonne reproduction du processus stochastique C, qui est l'objectif fixé par le précédent chapitre.

Afin de construire des variables aléatoires permettant de reproduire ce délai d'auto-allumage incertain, une première approche proposée ici consiste à ne retenir que les paramètres incertains initiaux influençant significativement ce délai d'auto-allumage incertain. Pour pouvoir identifier ces paramètres influents, une étude de sensibilité globale est utilisée. Une fois ces paramètres influents identifiés, une expansion en polynôme du chaos est utilisée afin de reconstruire le terme source incertain $\dot{\omega}_{Y_c}$.

Une seconde approche consiste à construire des variables aléatoires directement à partir du processus stochastique $\dot{\omega}_{Y_c}$, au travers d'une expansion de Karhunen-Loève de celui-ci, le délai d'auto-allumage étant relié à $\dot{\omega}_{Y_c}$ au travers de l'expression (7.5).

7.3.2 Analyse de sensibilité globale

7.3.2.1 Analyse de sensibilité globale du délai d'auto-allumage $\tau^{C=0.1}$

L'analyse de sensibilité globale correspond au calcul des indices de Sobol présenté dans le chapitre 4, qui permettent de rendre compte de l'importance des différents paramètres sur une quantité d'intérêt. Le calcul des indices de Sobol a été réalisé grâce à une approximation des termes du premiers ordres de la RS-HDMR du délai d'auto-allumage à l'aide de polynômes de Hermite de degré maximum 5, les variables aléatoires considérées étant les variables aléatoires gaussiennes centrées et réduites (ξ_r) reliées aux facteurs pré-exponentiels A_r de la loi d'Arrhénius par l'expression suivante, les notations étant les mêmes que dans l'expression (6.8) :

$$\xi_r = 3\frac{\ln(A_r) - \ln(A_r^0)}{\ln(f_r)} \tag{7.7}$$

Les coefficients de l'expansion en polynôme du chaos ont été obtenus grâce à une méthode projective utilisant 10 séquences de Sobol randomisées par une méthode de Full-Scramblig de 8192 points chacune.

Un calcul des indices de Sobol du premier ordre pour $\tau^{C=0.1}$ de chacune des composantes indépendantes de $\mathbf{A}$ a été réalisé pour une richesse ϕ variant dans l'intervalle $[0.2, 2.0]$ et une température initiale variant dans l'intervalle $[800K, 1200K]$. Ces indices de Sobol sont présentés sur la figure 7.3 en fonction de la température et de la richesse. Les indices de Sobol présentés sont ceux présentant une valeur supérieure à 5% sur au moins un point du domaine de température et de richesse considéré.

L'erreur de convergence relative des indices de Sobol présentés sur la figure 7.2 est présentée sur la figure 7.3. Cette erreur de convergence a été obtenue par une méthode de Jack-knife présentée dans le chapitre 3. L'erreur relative commise pouvant devenir très grande pour des petites valeurs des indices de Sobol, l'erreur est présentée uniquement lorsque la valeur de l'indice de Sobol est plus grande que 0.1%, la valeur 0 étant considérée autrement et représentée par la couleur blanche sur les cartes.

Pour l'ensemble des indices de Sobol du premier ordre, l'erreur commise est globalement un ordre de grandeur en dessous de la valeur de ceux-ci lorsque ceux-ci sont non négligeable (c'est à dire supérieur à 0.1%), le maximum de l'erreur relative étant de 0.15 sur une petite partie du domaine pour la réaction $O + H_2 = OH + H$. En regardant simultanément les cartes des figures 7.2 et 7.3, il apparaît que l'erreur relative de convergence est la plus importante dans les zones où les indices de Sobol ont une valeur faible. Dans les zones pour lesquelles l'indice de Sobol considéré est significatif, l'erreur de convergence est plutôt de deux ordres de grandeurs inférieure à la valeur de l'indice de Sobol, permettant d'avoir confiance dans les résultats de l'analyse de sensibilité globale présentés.

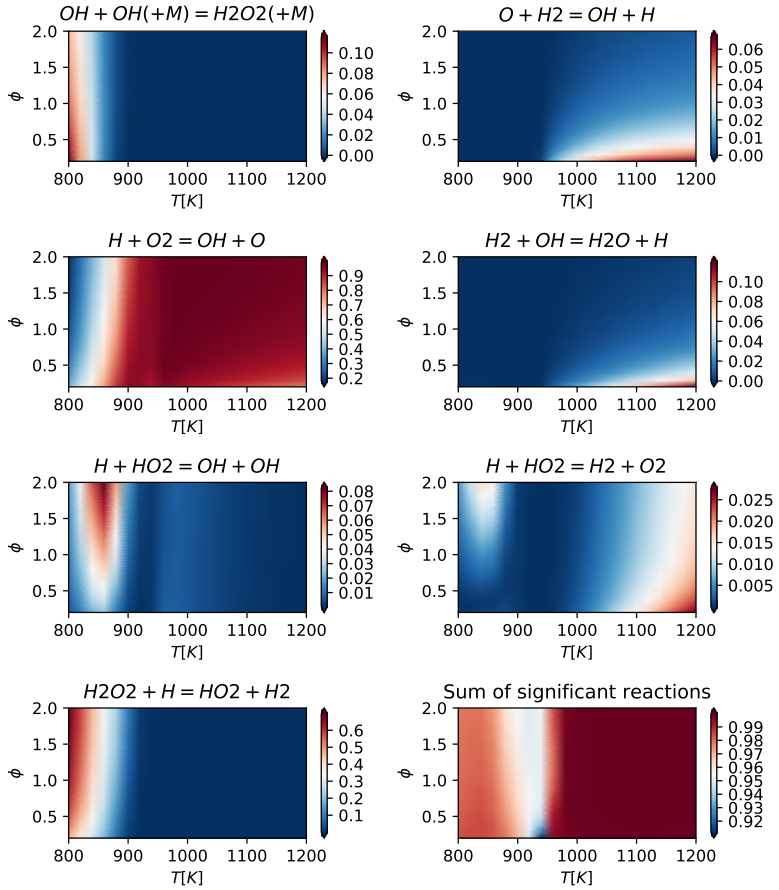

Figure 7.2 – *Indices de Sobol du premier ordre pour le délai d'auto-allumage* $\tau^{C=0.1}$ *pour une température variant entre* $800K$ *et* $1200K$*, et une richesse variant entre* 0.2 *et* 2*. La dernière carte représente la somme de ces indices de Sobol du premier ordre.*

Cette analyse de sensibilité globale permet d'observer que seules quelques réactions ont une influence significative sur le délai d'auto-allumage étudié, et que les domaines d'influence des différentes réactions sont propres à chacune de ces réactions. Parmi ces réactions, on peut noter deux réactions majoritaires :

— La réaction $H + O_2 = OH + O$ à haute température.

— La réaction $H_2O_2 + H = HO_2 + H_2$ à basse température.

Les sept réactions identifiées ont un impact au moins supérieur à 92% sur le délai d'auto-allumage $\tau^{C=0.1}$ pour l'ensemble du domaine de température et de richesse étudié. Dans la suite, quatre conditions initiales d'intérêt seront retenues, caractérisées par des températures et richesses différentes, et présentant des réactions impactantes différentes. Ces conditions initiales sont les suivantes :

— Le point de température $800K$ et de richesse 0.5, étant impacté par

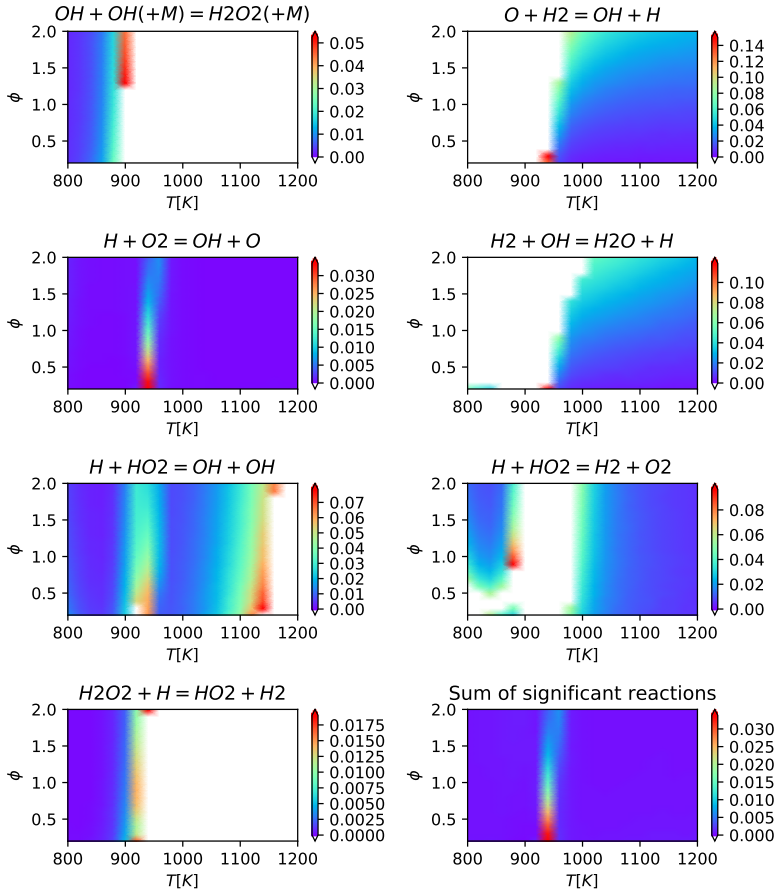

FIGURE 7.3 – *Erreur relative de l'estimateur des indices de Sobol du premier ordre pour le délai d'auto-allumage* $\tau^{C=0.1}$ *pour une température variant entre* $800K$ *et* $1200K$*, et une richesse variant entre* 0.2 *et* 2*. L'erreur est montrée uniquement sur la partie du domaine pour laquelle l'indice de Sobol considéré a une valeur supérieure à* 0.1%*, la couleur blanche étant présente sur la partie du domaine où cette condition n'est pas vérifiée.*

les réactions $H_2O_2 + H = HO_2 + H_2$ ($S = 57\%$), $H + O_2 = OH + O$ ($S = 30\%$) et $OH + OH(+M) = H_2O_2(+M)$ ($S = 10\%$).

— Le point de température $850K$ et de richesse 2, étant impacté par les réactions $H + O_2 = OH + O$ ($S = 46\%$), $H_2O_2 + H = HO_2 + H_2$ ($S = 40\%$), $H + HO_2 = OH + OH$ ($S = 8\%$), $OH + OH(+M) = H_2O_2(+M)$ ($S = 3\%$) et $H + HO_2 = H_2 + O_2$ ($S = 2\%$).

— Le point de température $875K$ et de richesse 1, étant impacté par les réactions $H + O_2 = OH + O$ ($S = 70\%$), $H_2O_2 + H = HO_2 + H_2$ ($S = 21\%$), $H + HO_2 = OH + OH$ ($S = 3\%$).

— Le point de température $1200K$ et de richesse 0.2, étant impacté par les réactions $H + O_2 = OH + O$ ($S = 78\%$), $H_2 + OH = H_2O + H$

($S = 12\%$), $O + H_2 = OH + H$ ($S = 7\%$) et $H + HO_2 = OH + OH$ ($S = 3\%$).

Suivant le seuil que l'on fixe, on peut déduire de cette analyse de sensibilité globale que le nombre de paramètres incertains à considérer peut être drastiquement réduit si l'on souhaite reproduire correctement le délai d'auto-allumage $\tau^{C=0.1}$. Autrement dit, seul un sous-ensemble des composantes du vecteur $\mathbf{A}$ peuvent être conservées, les autres n'ayant pas d'influence sur la quantité d'intérêt qu'est le délai d'auto-allumage.

Dans la suite, pour chacun des points de fonctionnements choisis, lorsque l'on considérera un nombre n de variables aléatoires dans l'expansion en polynôme du chaos, ces variables aléatoires correspondront aux n paramètres incertains associés aux réactions ayant un impact significatif sur le délai d'auto-allumage, par ordre d'importance.

En contexte de chimie tabulée, ce délai d'auto-allumage peut se retrouver via l'expression (7.5), faisant intervenir le terme source $\dot{\omega}_{Y_c}$, qu'il convient donc également d'étudier via une analyse de sensibilité.

7.3.2.2 Analyse de sensibilité globale du terme source $\dot{\omega}_{Y_c}$

Le processus stochastique qu'est le terme source $\dot{\omega}_{Y_c}$ dépend du vecteur aléatoire $\mathbf{A}$, et pour l'ensemble des valeurs de la variable d'avancement c, les indices de Sobol pour la variable aléatoire $\dot{\omega}_{Y_c}(c, .)$ peuvent être calculés de la même manière que cela a été fait pour le délai d'auto-allumage $\tau^{C=0.1}$. Ces indices de Sobol du premier ordre ont été obtenus en considérant des polynômes de Hermite de degré maximal 10, et à l'aide d'une méthode de Quasi-Monte Carlo randomisée impliquant 10 séquences de Sobol de 8192 points chacune. Il a été observé que la prise en compte des polynômes de degré inférieur à 2 donnait significativement les mêmes résultats que l'utilisation d'un degré plus élevé.

Sur la figure 7.4 sont représentés les indices de Sobol cumulés en fonction de c pour le terme source $\dot{\omega}_{Y_c}$ pour les quatre conditions initiales étudiées. Chaque bande de couleur représente la valeur prise par un indice de Sobol, les bandes de gris se superposant de sorte que la somme des indices de Sobol ressort. Pour chacun des points, les indices de Sobol considérés sont ceux présentés précédemment influençant significativemet le délai d'auto-allumage $\tau^{C=0.1}$, l'ordre d'apparition des réactions présentés précédemment correspondant ici avec le niveau de gris, la réaction apparaissant la première étant la plus claire et la dernière la plus foncée.

Les réactions ayant un impact significatif sur le délai d'auto-allumage $\tau^{C=0.1}$ ont également un impact significatif sur le terme source de la variable d'avancement pour des valeurs de c inférieures à 0.1, comme visible sur la figure 7.4. Cela peut s'expliquer par le fait que le délai d'auto-allumage peut s'exprimer en fonction du terme source $\dot{\omega}_{Y_c}$ comme dans la relation (7.5).

Les différentes variables aléatoires initiales ayant un impact significatif

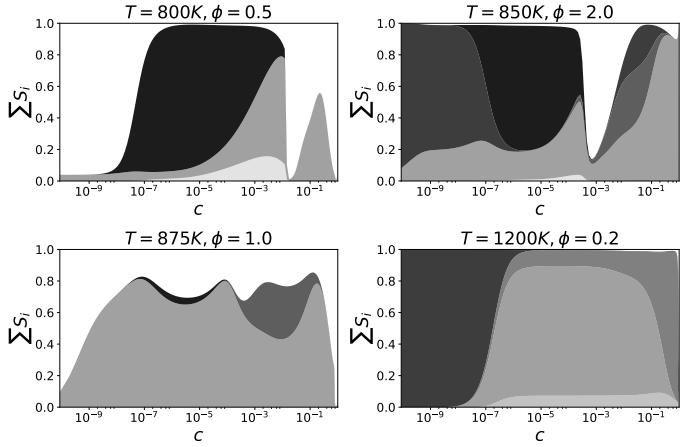

FIGURE 7.4 – *Principaux indices de Sobol du premier ordre pour le terme source de la variable de progrès $\dot{\omega}_{Y_c}$ en fonction de la variable d'avancement normalisée c, pour les quatre conditions initiales étudiées.*

étant identifiées, l'objectif est maintenant de vérifier que ne garder que celles-ci permet bien la reproduction du délai d'auto-allumage via une chimie tabulée où le terme source $\dot{\omega}_{Y_c}$ est modélisé et ne dépend que des variables aléatoires influentes.

7.4 Expansion en polynômes du Chaos du terme source $\dot{\omega}_{Y_c}$

L'analyse de sensibilité précédemment effectuée a permis d'identifier les réactions ayant un impact significatif sur l'incertitude du délai d'auto-allumage $\tau^{C=0.1}$, ainsi que sur le terme source incertain $\dot{\omega}_{Y_c}$ de la variable d'avancement de la réaction, pour des valeurs de c inférieures à 0.1 au moins. Il est donc à priori possible de modéliser le terme source incertain $\dot{\omega}_{Y_c}$ en ne prenant en compte que les paramètres incertains de ces réactions.

7.4.1 Rappels sur l'HDMR d'une fonctions dépendant d'un vecteur aléatoire

Grâce à l'analyse de sensibilité précédemment effectuée, on peut désormais réduire le jeu de paramètres incertains initiaux donnés par le vecteur aléatoire $\mathbf{A}$ à un jeu de paramètres incertains $\mathbf{A}_{I_{red}}$ donnés par un vecteur aléatoire construit en ne gardant que certaines composantes de $\mathbf{A}$. I_{red} est en fait un sous ensemble de l'ensemble $[\![1, n]\!]$ correspondant aux indices des paramètres ayant une importance significative. Les composantes de $\mathbf{A}$ étant toutes supposées indépendantes, il n'y a aucune difficulté à ne considérer qu'une partie de ses composantes, et la densité du vecteur aléatoire $\mathbf{A}_{I_{red}}$ est obtenue facilement,

comme le produit des densités de ses composantes.

Comme expliqué dans la partie du chapitre 4 sur l'analyse de sensibilité globale, une quantité d'intérêt Q dépendant du vecteur aléatoire $\mathbf{A}$ à n composantes peut se décomposer sous la forme RS-HDMR suivante :

$$Q(\mathbf{A}) = \sum_{I \subset [\![1,n]\!]} Q_I(\mathbf{A}_I) \tag{7.8}$$

Le fait que seulement les composantes contenues dans $\mathbf{A}_{I_{red}}$ de $\mathbf{A}$ aient un impact significatif selon l'analyse de sensibilité globale signifie simplement que la somme des variances des termes $Q_I(\mathbf{A}_I)$ pour $I \subset I_{red}$ est presque égale à la variance totale de $Q(\mathbf{A})$. En ce sens, on peut donc écrire la relation (7.9).

$$Q(\mathbf{A}) = \sum_{I \in [\![1,n]\!]} Q_I(\mathbf{A}_I) \approx \sum_{I \subset I_{red}} Q_I(\mathbf{A}_I) \tag{7.9}$$

La réduction du nombres de paramètres se traduit donc ici par la détermination des termes $Q_I(\mathbf{A}_I)$ pour $I \subset I_{red}$, qui sont des fonctions de plusieurs variables. Par ailleurs, il n'est pas vrai en général qu'on a la relation (7.10), où $(\mathbf{A}_{red}, \overline{\mathbf{a}}_{red})$ est le vecteur aléatoire, dont les composantes d'indices dans I_{red} suivent la même loi que dans $\mathbf{A}$, et les autres composantes sont chacune presque sûrement égale à une constante a_i.

$$\sum_{I \subset I_{red}} Q_I(\mathbf{A}_I) \neq Q(\mathbf{A}_{red}, \overline{\mathbf{a}}_{red}) \tag{7.10}$$

Pour se convaincre que l'égalité précédente n'est pas toujours vérifié, on peut considérer l'espérance de chacun des deux membres. L'espérance du membre de gauche est simplement $E\left[Q(\mathbf{A})\right]$ de par les propriétés de la RS-HDMR. Concernant le membre de droite, il existe à priori une fonction g dépendant de $\overline{\mathbf{a}}_{red}$ telle que :

$$E\left[Q(\mathbf{A}_{red}, \overline{\mathbf{a}}_{red})\right] = g(\overline{\mathbf{a}}_{red}) \tag{7.11}$$

Cette dernière fonction n'est à priori pas constante, et de ce fait ne peut pas être tout le temps égale à l'espérance $E\left[Q(\mathbf{A})\right]$. Pour cette raison, et sachant que l'analyse de sensibilité globale précédente se traduit par l'approximation (7.9) et non par la considération de $Q(\mathbf{A}_{red}, \overline{\mathbf{a}}_{red})$ pour une certaine valeur de $\overline{\mathbf{a}}_{red}$, on s'attachera dans la suite à déterminer les termes de la RS-HDMR afin de modéliser le terme source $\dot{\omega}_{Y_c}$, ce qui est fait dans la section suivante.

7.4.2 Construction de l'Expansion en Polynômes du Chaos

Comme rappelé par la dernière remarque, les coefficients de l'expansion en polynômes du chaos sont calculés en projetant le terme source incertain dépendant de l'ensemble des indices sur une base de polynômes orthonormaux, ne dépendant eux que du jeu réduit de paramètres incertains. Étant données la propriété de positivité du terme source de la variable d'avancement, le choix a été fait ici de modéliser le logarithme de celui-ci pour la construction d'une expansion en polynôme du chaos, que l'on notera désormais $\dot{\omega}_{Y_c}^{log}$. Cela permet d'assurer la positivité du terme source qui sera utilisé ensuite, étant l'exponentiel du modèle construit.

Étant donnée une base de polynômes orthonormaux (P_k), l'expression des coefficients α_k associés aux polynômes P_k est donc donnée par :

$$\alpha_k(c) = \int_{\mathbb{R}^d} \dot{\omega}_{Y_c}^{log}(c, \mathbf{A}) P_k(\mathbf{A}_{red}) \pi(\mathbf{A}) d\mathbf{A} \tag{7.12}$$

Le choix a été fait ici d'utiliser des polynômes de Hermite pour l'expansion en polynômes du chaos. De ce fait, il a été nécessaire d'effectuer un changement de variable sur chacune des composantes du vecteur aléatoire $\mathbf{A}$ afin de se ramener à un vecteur aléatoire gaussien ξ dont les composantes sont indépendantes et centrée réduite. L'expression (7.13) pour les coefficients $\alpha_k^{Hermite}$ de l'expansion en polynômes du chaos utilisant les polynômes de Hermite est similaire à l'expression (7.12).

$$\alpha_k^{Hermite}(c) = \int_{\mathbb{R}^d} \dot{\omega}_{Y_c}^{log}(c, \xi) P_k(\xi_{red}) \pi(\xi) d\xi \tag{7.13}$$

Le calcul de l'intégrale dans l'expression (7.13) a été réalisée à l'aide d'une méthode de Quasi-Monte Carlo randomisée utilisant 10 séquences de Sobol randomisés à l'aide d'une méthode de Full-Scrambling. Une étude de sensibilité similaire à celle réalisée pour $\dot{\omega}_{Y_c}$ a été réalisée pour $\dot{\omega}_{Y_c}^{log}$, laquelle montrait que seuls les polynômes de degré inférieurs à 2 pour l'approximation de l'HDMR avaient un impact significatif sur les indices de Sobol du premier ordre calculés pour les réactions conservées, c'est à dire celle impactant significativement le délai d'auto-allumage $\tau^{C=0.1}$ ayant été identifiées précédemment. De ce fait, le degré maximal des polynômes considérés pour l'expansion en polynômes du chaos ici est de 2, et l'erreur de convergence des valeurs des coefficients est représentées sur la figure 7.5 pour des séquences de Sobol comportant 4096 points, les coefficients ayant été calculés pour l'ensemble des valeurs c_i du maillage de la variable d'avancement permettant la tabulation.

La convergence des coefficients est telle que l'erreur statistique commise n'excède jamais 10^{-2} en valeur absolue. La figure 7.5 montre qu'il existe des coefficients estimés $\hat{\alpha}_k(c_i)$ présentant une erreur d'estimation plus grande que leur

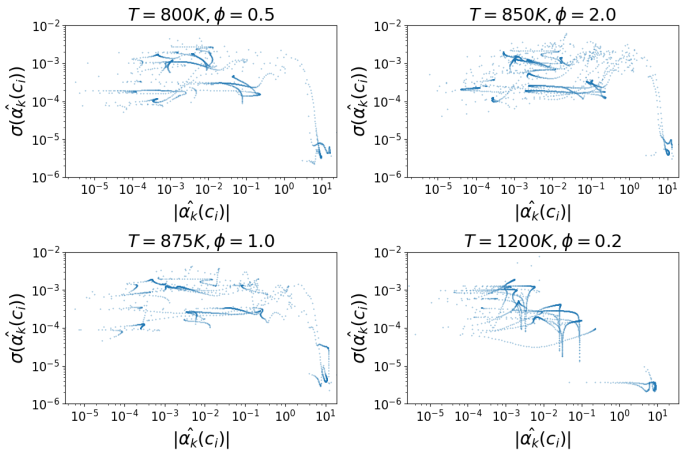

FIGURE 7.5 – *Nuages des points* $(|\hat{\alpha}_k(c_i)|, \sigma(\hat{\alpha}_k(c_i)))$ *pour les quatre points de fonctionnements,* $\hat{\alpha}_k(c_i)$ *étant l'estimation de Quasi-Monte Carlo obtenue pour le coefficient* $\alpha_k(c_i)$*, et* $\sigma(\hat{\alpha}_k(c_i))$ *étant l'écart-type sur l'estimation de Quasi-Monte Carlo randomisé associée.*

valeur absolue estimée. Ces estimations peuvent donc être considérées à priori comme problématiques. Néanmoins, cela n'arrive que pour des coefficients qui sont nécessairement proche de 0 en valeur absolue, et l'impact de ces coefficients est en fait négligeable pour cette raison, comme cela est montré dans la section suivante où les expansions en polynômes du chaos calculées sont utilisées pour la propagation d'incertitudes.

7.4.3 Validation du modèle pour le terme source $\dot{\omega}_{Y_c}$

Le terme source modélisé $\dot{\omega}_{Y_c}^{PCE}$ de la variable d'avancement est dans cette section modélisé comme :

$$\dot{\omega}_{Y_c}^{PCE}(c, \mathbf{A}_{red}) = \exp\left(\sum_{k=0}^{P} \alpha_k(c) P_k(\mathbf{A}_{red})\right) \tag{7.14}$$

Ce terme source incertain modélisé est ensuite utilisé afin d'obtenir des trajectoires temporelles de la variable d'avancement c, par intégration de l'équation différentielle ordinaire :

$$\frac{dc}{dt}(c, \mathbf{a}_{red}) = \frac{\dot{\omega}_{Y_c}^{PCE}(c, \mathbf{a}_{red})}{Y_c^{\infty} - Y_c^{0}} \tag{7.15}$$

A partir de ces trajectoires temporelles de la variable d'avancement c, il est possible d'obtenir des réalisations du délai d'auto-allumage $\tau_{PCE}^{C=0.1}$. Pour les

quatre points de fonctionnements, 10,000 réalisations aléatoires du délai d'auto-allumage ont été calculées, les unes en utilisant la chimie détaillée ($\tau_{REF}^{C=0.1}$) et les autres en utilisant le terme source modélisé $\dot{\omega}_{Y_c}^{PCE}$ basé sur une expansion en polynôme du chaos de degré 2 impliquant uniquement une ou plusieurs des réactions les plus influentes. A l'aide de l'ensemble de ces réalisations, les diagrammes quantile-quantile des variables aléatoires $\tau_{REF}^{C=0.1}$ et $\tau_{PCE}^{C=0.1}$ ont pu être construits, et sont représentés sur la figure 7.6.

Etant données deux variables aléatoires X et Y de fonctions de répartition respectives F_X et F_Y, le diagramme quantile-quantile de X et Y consiste en la courbe constituées par les points de coordonnées $(F_X^{-1}(p), F_Y^{-1}(p))$ pour p variant dans l'intervalle $[0, 1]$. Dans le cas où X et Y suivent la même loi, cette courbe se situe sur la première bissectrice. Un écart de la courbe du diagramme quantile-quantile de la première bissectrice sur la figure 7.6 traduit donc une mauvaise reproduction de la loi de la variable aléatoire $\tau^{C=0.1}$ par $\tau_{PCE}^{C=0.1}$.

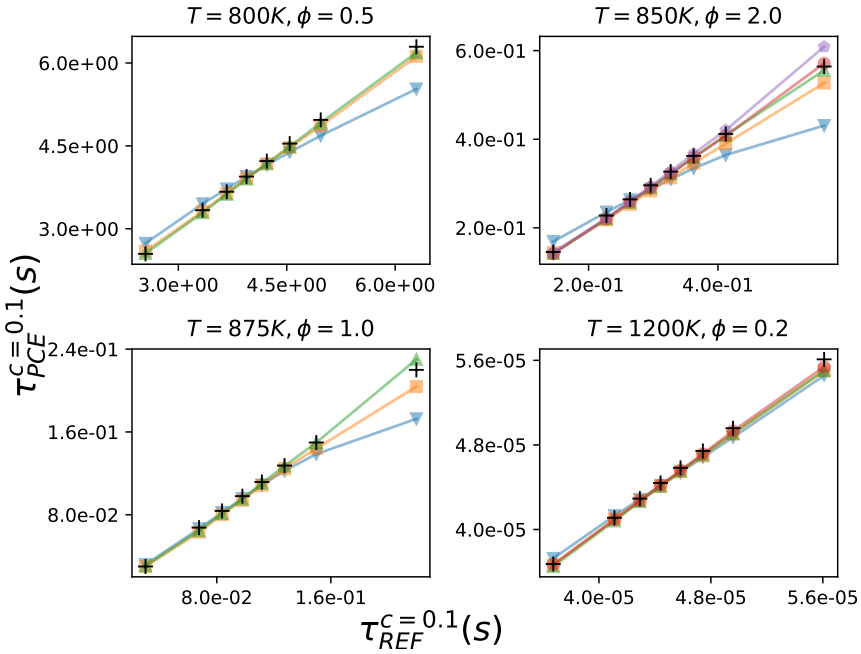

FIGURE 7.6 – *Diagramme quantile-quantile du délai d'auto-allumage $\tau^{C=0.1}$ pour les différents points de fonctionnements, calculés à l'aide de la chimie détaillée incertaine ($\tau_{REF}^{C=0.1}$) et à l'aide du terme source modélisé $\dot{\omega}_{Y_c}^{PCE}$ ($\tau_{PCE}^{C=0.1}$), impliquant un nombre variables des paramètres incertains impactant le plus $\tau^{C=0.1}$: 1 paramètre incertain (triangle bas), 2 paramètres incertains (carrés), 3 paramètres incertains (triangle haut), 4 paramètres incertains (ronds) et 5 paramètres incertains (pentagones), la référence correspondant aux croix noires. Pour chaque courbe, la position des symboles correspond aux q-quantiles avec q prenant ses valeurs dans* $\{0.01, 0.15, 0.29, 0.43, 0.57, 0.71, 0.85, 0.99\}$.

La figure 7.6 montre une bonne reproduction de la loi de la variable aléatoire $\tau^{C=0.1}$ par la variable aléatoire $\tau_{PCE}^{C=0.1}$, les courbes passant par les points

$(\tau_{REF}^{C=0.1}, \tau_{PCE}^{C=0.1})$ étant quasiment confondu avec la première bissectrice à mesure que le nombre de paramètres . On observe une concordance entre la somme des indices de Sobol des variables impliquées et la bonne reproduction de la loi de $\tau^{C=0.1}$. En effet, pour le point de température $T = 1200\ K$ et de richesse $\phi = 0.2$, l'indice de Sobol correspondant à la réaction la plus influente est de 78%, et on observe une reproduction de la loi de $\tau^{C=0.1}$ correcte avec cette seule variable aléatoire dans $\dot{\omega}_{Y_c}^{PCE}$, alors qu'avec une seule variable aléatoire pour les autres points de fonctionnements, la reproduction est moins bonne. La reproduction de la loi de $\tau^{C=0.1}$ est la plus difficile pour les zones où les variables impliquées sont les plus nombreuses avec des ordres de grandeurs similaires, correspondant dans le cas présenté aux points à plus basse température, particulièrement les points de température $T = 850\ K$ et $T = 875\ K$. De plus, la reproduction des longs délais d'auto-allumage semble être le plus difficile pour l'ensemble des conditions initiales, comme l'atteste le fait que les courbes s'éloignent de la première bissectrice pour les grandes valeurs de $\tau^{C=0.1}$.

Les conclusions du précédent chapitre portaient sur la nécessité de la bonne reproduction du processus stochastique C et non celle du délai d'auto-allumage $\tau^{C=0.1}$. Il est intéressant de savoir si ce processus stochastique C est correctement reproduit lui aussi avec l'utilisation du terme source $\dot{\omega}_{Y_c}^{PCE}$. La bonne reproduction ici sera jugée à travers les statistiques que sont la moyenne et l'écart-type du processus C. Sur la figure 7.7 sont présentés les moyennes temporelles du processus stochastique C obtenues par une méthode de Quasi-Monte Carlo randomisé. La reproduction de cette moyenne pour le point à haute température est excellente avec une seule variable aléatoire, alors que pour les trois points à basse température, la reproduction de cette moyenne nécessite de retenir deux, voire trois variables aléatoires pour être aussi bonne.

Sur la figure 7.8 sont présentés les écart-types temporels du processus stochastique C. L'écart-type semble être plus difficile à reproduire que la moyenne du processus C, puisque deux, voire trois variables aléatoires sont à considérer pour l'ensemble des conditions initiales afin d'avoir une bonne reproduction de celui-ci sur l'ensemble de la fenêtre temporelle présentée.

Le choix du délai d'auto-allumage comme critère pour décider quelles sont les variables aléatoires à conserver est un bon choix dans le cas présent, puisque ces mêmes variables aléatoires permettent bien de reproduire la moyenne ainsi que l'écart-type du processus stochastique C, qui est l'objectif donné par la conclusion du chapitre précédent.

7.5 Conclusion

Ce chapitre a introduit l'analyse de sensibilité globale comme moyen de réduire le nombre de paramètres incertains à considérer. L'analyse de sensibilité a été ici réalisée sur le délai d'auto-allumage, qui est un paramètre macroscopique important pour les réacteurs adiabatiques homogène et à pression constante.

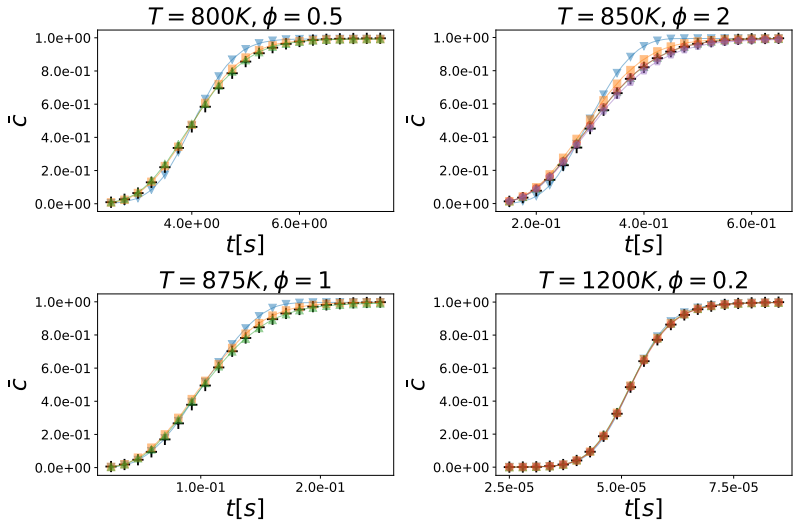

FIGURE 7.7 – *Moyennes temporelles $\hat{C}$ du processus C pour les quatre points de fonctionnements identifiés, calculé à l'aide de la chimie détaillée (croix), et à l'aide du terme source $\dot{\omega}_{Y_c}^{PCE}$ impliquant un nombre variable de paramètres incertains :*1 *paramètre (triangle bas),* 2 *paramètres (carrés),* 3 *paramètres (triangle haut),* 4 *paramètres (ronds) et* 5 *paramètres (pentagones).*

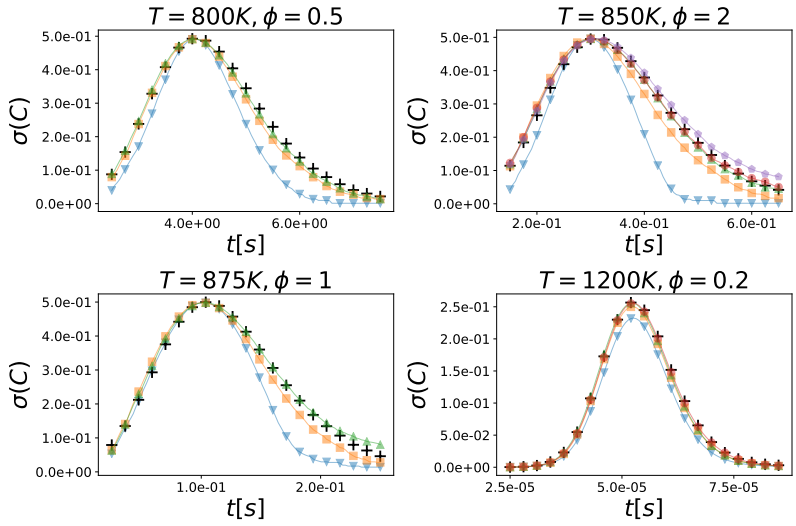

FIGURE 7.8 – *Écart-types temporels $\sigma(C)$ du processus C pour les quatre points de fonctionnements identifiés, calculé à l'aide de la chimie détaillée (croix), et à l'aide du terme source $\dot{\omega}_{Y_c}^{PCE}$ impliquant un nombre variable de paramètres incertains :*1 *paramètre (triangle bas),* 2 *paramètres (carrés),* 3 *paramètres (triangle haut),* 4 *paramètres (ronds) et* 5 *paramètres (pentagones).*

Les variables aléatoires impactant significativement le délai d'auto-allumage se trouve être les mêmes que celles impactant significativement le terme source de la variable de progrès dont l'objectif fixé par le chapitre 6 est la reproduction des incertitudes. Ces observations ont permis de considérer une expansion en polynôme du chaos pour modéliser le terme source incertain de la variable de progrès, ce qui a permis une bonne reproduction des incertitudes sur la variable de progrès, d'autant meilleur que le nombre de variables aléatoires considérées pour la PCE étant important.

La limitation pouvant vite arriver avec cette méthode est le nombre de variables aléatoires à prendre en compte. Dans le présent exemple, les variables aléatoires à prendre en compte changent avec les conditions initiales considérées. Ainsi, l'augmentation de la taille du domaine des conditions initiales implique l'augmentation du nombres de variables aléatoires à considérer. Ce fait se retrouvera a priori pour n'importe quel mécanisme réactionnel, de nouvelles réactions s'activant à mesure que le domaine des conditions initiales croît. Dans le cas de l'hydrogène, le nombre de réactions ayant un impact significatif reste restreint, mais il est certain que ce nombre croîtra avec la complexité du mécanisme réactionnel. Dans un tel cas, le nombre limité de variables aléatoires pourrait excéder 5, le nombre de réactions total retenu étant ici de 7, et il sera alors nécessaire de faire un choix au niveau des variables aléatoires, en fonction de la précision souhaité ainsi que des ressources en calcul disponibles.

D'autres méthodes sont présentées dans le chapitre suivant, qui ne font pas intervenir les variables aléatoires initiales.

Chapitre 8

Représentation du terme source incertain de la variable d'avancement à l'aide de nouvelles variables aléatoires

Prérequis :

- Calcul d'intégrales multiples (Cubature, Monte Carlo et Quasi-Monte Carlo randomisé)
- Expansion en Polynôme du Chaos et Expansion de Karhunen-Loève pour des processus stochastiques
- Méthodes à noyaux pour la génération de vecteurs aléatoires étant donnée une réalisation d'un échantillon de vecteurs aléatoires suivant la même loi

Notions clés et apports du chapitre :

- Modélisation du terme source incertain d'une variable d'avancement à l'aide de différentes représentations spectrales
- Utilisation et validation de cette modélisation du terme source incertain d'une variable d'avancement sur le système chimique canonique étudié

Les limitations de la méthode du chapitre précédent ont poussé au développement d'autres méthodes, permettant de s'affranchir de l'utilisation directe des paramètres incertains initiaux en définissant de nouvelles variables aléatoires. Le point clé de ces méthodes est que ces nouvelles variables aléatoires sont définies de telle sorte qu'elles sont rangées par ordre d'impact sur une quantité d'intérêt, ce qui permet de sélectionner facilement les variables aléatoires à conserver en fonction du besoin et des contraintes de coût.

8.1 Expansion de Karhunen-Loève pour le terme source $\dot{\omega}_{Y_c}$

8.1.1 Température et richesse fixées

8.1.1.1 Prise en compte des spécificités du cas

Le terme source incertain $\dot{\omega}_{Y_c}$ étant un processus stochastique, il est possible de le représenter à partir de son expansion de Karhunen-Loève. Cependant, tout comme pour l'expansion en polynômes du chaos, plutôt que de s'intéresser au terme source directement, le logarithme de celui-ci, $\dot{\omega}_{Y_c}^{log}$, sera considéré. Comme vu précédemment, l'utilisation de la tabulation du terme source pour l'intégration de l'équation différentielle (7.4) nécessite d'avoir un maillage en la variable d'avancement c permettant de réaliser correctement cette intégration. La reproduction du terme source incertain à l'aide d'une expansion en polynômes du chaos ne pose pas de problèmes, une expansion de la variable aléatoire $\dot{\omega}_{Y_c}(c)$ pouvant être réalisée en chacun des points de ce maillage. Le calcul de l'expansion de Karhunen-Loève à l'aide d'une méthode de Nyström (voir chapitre 4) impose pour sa part des valeurs de la variable d'avancement c, qui sont directement liées à la méthode de quadrature utilisée pour la méthode de Nyström. De ce fait, utiliser directement la méthode de Nyström ne permet pas d'avoir une discrétisation suivant le maillage nécessaire à la bonne intégration de l'équation différentielle (7.4). En particulier, dans le cas de l'auto-allumage d'un réacteur homogène, adiabatique et à pression constante, le maillage doit être fin pour les faibles valeurs de c, et peut être relâché pour les plus grandes valeurs de c. L'utilisation directe des méthodes de quadratures présentées dans le chapitre 2 n'est donc pas envisageable. En effet, l'utilisation d'une méthode de quadrature basée sur les points d'abscisses $x_k = (1 + cos(-k\pi/n))/2$, pour $k = 1, \ldots, n-1$, comme la méthode de Fejér par exemple, avec le premier point du maillage situé en c_{min}, impose au point x_1 d'être de l'ordre de c_{min}, soit en utilisant un développement limité pour l'expression de x_1, $\pi^2/(2n^2) \sim c_{min}$, c'est à dire $n \sim \pi/\sqrt{2c_{min}}$. Pour une valeur de c_{min} de l'ordre de 10^{-10}, on a donc une valeur de n de l'ordre de 10^5, ce qui implique une méthode de quadrature avec $100,000$ points seulement pour la variable d'avancement c, ce qui n'est pas envisageable compte tenu du problème aux valeurs propres de taille n à résoudre. Une première approche pour éviter ce problème serait d'utiliser une

méthode de quadrature directement basée sur les points du maillage à utiliser pour tabuler le terme source. Cela implique de construire une telle méthode au cas par cas, et de déterminer les poids de celle-ci afin d'avoir les meilleures propriétés pour celle-ci. Une seconde méthode, beaucoup plus simple à mettre en place, consiste à effectuer un changement de variables pour la variable d'avancement. Ce changement de variable consiste à considérer une nouvelle variable ξ telle que fonction $c = \psi(\xi)$, avec ψ un difféomorphisme de classe C^∞. De cette manière, plutôt que de s'intéresser au processus stochastique $\dot{\omega}_{Y_c}^{log}$, on s'intéressera au processus stochastique $\dot{\omega}_{Y_c}^{log,\psi}$, défini par la relation (8.1).

$$\dot{\omega}_{Y_c}^{log,\psi}(\xi, \mathbf{A}) = \dot{\omega}_{Y_c}^{log}(\psi(\xi), \mathbf{A}) \tag{8.1}$$

Afin d'avoir un changement de variable avec de bonnes propriétés, le maillage de points $(c_k)_{k=0,\dots,n}$ est considéré, et la fonction ψ est construite telle que $\psi(k/n) = c_k$, la reconstruction en dehors de ces points étant faite par une interpolation linéaire. La fonction ψ correspondante est présentée sur la figure 8.1.

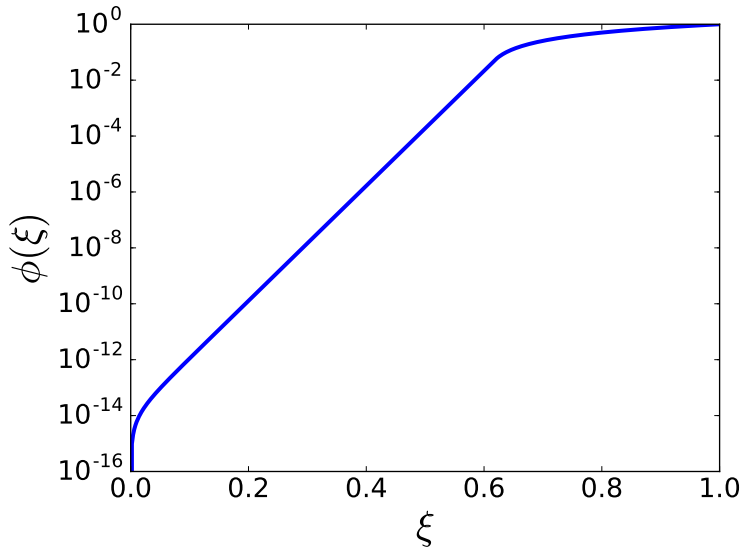

FIGURE 8.1 – *Changement de variable utilisé au niveau de la variable d'avancement pour calculer l'expansion de Karhunen-Loève.*

Une fois ce changement de variable effectué, il est possible de construire une expansion de Karhunen-Loève pour $\dot{\omega}_{Y_c}^{log,\psi}$ et d'obtenir une expansion pour $\dot{\omega}_{Y_c}^{log}$ à l'aide de la relation (8.2) utilisant la bijection réciproque ψ^{-1} de ψ.

$$\dot{\omega}_{Y_c}^{log}(c, \mathbf{A}) = \dot{\omega}_{Y_c}^{log,\psi}(\psi^{-1}(c), \mathbf{A}) \tag{8.2}$$

Différents aspects interviennent pour le calcul et l'utilisation de l'expansion de Karhunen-Loève du processus $\dot{\omega}_{Y_c}^{log,\psi}$. En effet, concernant le calcul, notamment la convergence de celui-ci, il faut s'assurer que les estimations de Monte Carlo ou de Quasi-Monte Carlo sont suffisantes, mais également que le nombre de points utilisés dans la quadrature est suffisant. Un autre aspect qui sera traité plus tard concerne l'utilisation de l'expansion de Karhunen-Loève, et notamment la modélisation qu'il est nécessaire d'effectuer pour les nouvelles variables aléatoires.

8.1.1.2 Convergence statistiques des estimations

Le calcul de l'expansion de Karhunen-Loève par la méthode de Nyström nécessite l'estimation de la matrice d'auto-covariance du processus stochastique aux points de quadratures. Les valeurs et modes propres calculés à partir de cette matrice de covariance sont alors des estimations des vrais valeurs et modes propres, dont il convient d'analyser l'erreur d'estimation comme effectué dans le chapitre 4. Le calcul de cette erreur d'estimation par une méthode directe comme dans le chapitre 4 est coûteuse, une évaluation du processus nécessitant la résolution d'un réacteur homogène adiabatique à pression constante. Cependant, comme vu au chapitre 3, l'utilisation d'une méthode de Quasi-Monte Carlo randomisé se prête particulièrement bien à une estimation par ré-échantillonnage bootstrap. Une méthode de ré-échantillonnage bootstrap a donc été utilisée afin d'obtenir une estimation de l'écart-type de l'estimateur des valeurs propres de l'expansion de Karhunen-Loève. Cette estimation a été obtenue à partir de 12000 ré-échantillonnage des $m = 10$ séquences de Sobol randomisées de n points chacune à l'aide d'une méthode de Full-Scrambling utilisées pour la méthode de Quasi-Monte Carlo randomisée. Les résultats de cette étude sont montrés sur la figure 8.2 où sont tracés les convergences relatives de ces estimateurs, à savoir le rapport de l'écart-type de l'estimateur sur la valeur moyenne estimée, en fonction du nombre total de points utilisés $N = mn$ pour les cinq premières valeurs propres. Les calculs ont été effectués pour les quatre points de fonctionnements, et la quadrature utilisée pour ces calculs était la seconde quadrature de Fejér impliquant 511 points. La résolution du problème aux valeurs propres pour chacun des cas est effectuée grâce à la routine DSYEV de la librairie LAPACK [7].

La convergence observée sur la figure 8.2 est globalement supérieure à $1/\sqrt{N}$ et proche de $1/N$, qui est la convergence attendue d'une méthode de Quasi-Monte Carlo. On observe également que la convergence relative est globalement la même pour l'ensemble des valeurs propres et des conditions initiales, et que celle-ci est inférieure à 1% pour la quasi totalité des valeurs propres présentées et la quasi totalité des conditions initiales quand le nombre de points par séquence de Sobol est supérieur à 1024. Dans la suite, compte tenu de ce comportement similaire pour l'ensemble des valeurs propres des différentes conditions initiales, on supposera une convergence similaire pour les autres cas

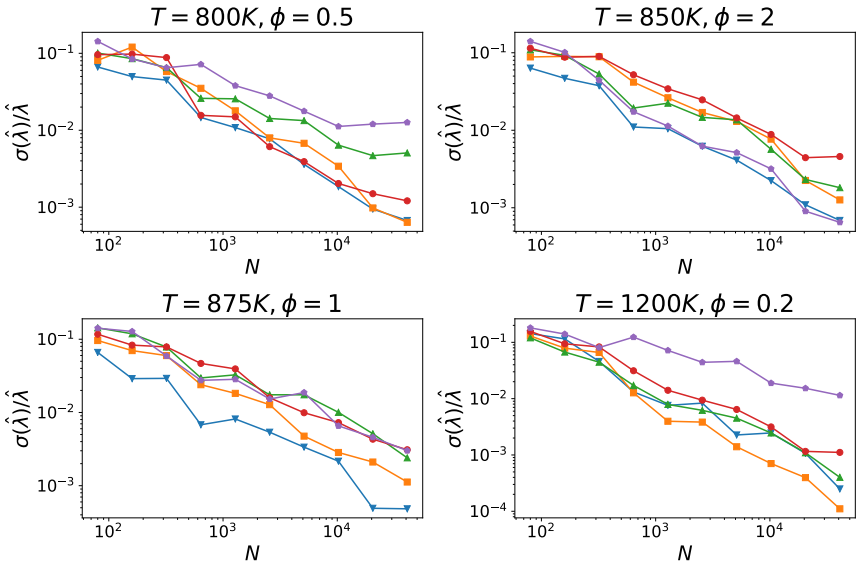

FIGURE 8.2 – *Convergence relative des estimations des* 5 *premières valeurs propres de l'expansion de Karhunen-Loève de* $\dot{\omega}_{Y_c}^{log,\psi}$ *pour les quatre conditions initiales, en fonction du nombre total* N *de points utilisés pour l'estimation. Les triangles bas correspondent à la première valeur propre, les carrés à la seconde, les triangles hauts à la troisième, les ronds à la quatrième et les pentagones à la cinquième.*

qui seront étudiés sans vérifier cette convergence explicitement.

8.1.1.3 Choix du niveau de quadrature

Un autre aspect rentrant en compte pour le calcul de l'expansion de Karhunen-Loève concerne la quadrature utilisée pour la méthode de Nyström, et plus particulièrement le nombre de points à utiliser dans cette quadrature. Deux aspects peuvent être distingués derrière ce choix du nombre de points. Le premier concerne la convergence des valeurs propres et des modes propres. Pour rappel, les valeurs et modes propres obtenus dans le cas où la matrice d'auto-covariance est parfaitement connue sont ceux du système discrétisé, pas du système continu. L'augmentation du nombre de points dans la quadrature permet aux valeurs et modes propres obtenus de converger vers les valeurs et modes propres du système continu. Le second aspect concerne la reconstruction des modes propres sur l'ensemble du domaine par des méthodes d'interpolation. Si le nombre de points est insuffisant, la qualité de la reconstruction peut ne pas être suffisante pour que l'utilisation du terme source modélisé permette de retrouver la bonne évolution temporelle pour le processus C.

Pour s'intéresser au premier aspect, le calcul de l'expansion de Karhunen-Loève a été effectué avec différents niveaux de quadratures. L'estimation de la matrice d'auto-covariance a été réalisée dans tous les cas à l'aide d'une méthode

de Quasi-Monte Carlo impliquant 10 séquences de Sobol randomisées à l'aide d'une méthode de Full-Scrambling et contenant chacune 4096 points. Les séquences de Sobol randomisées utilisées sont de plus les mêmes pour l'ensemble des cas, assurant que seul le niveau de quadrature varie entre les différents cas. Encore une fois, les calculs ont été réalisés pour l'ensemble des conditions initiales identifiées, et la quadrature utilisée est la seconde quadrature de Fejér impliquant $n = 2^l - 1$ points pour l variant de 4 à 10. Sur la figure 8.3 sont représentés les rapports, pour les cinq premières valeurs propres, de l'estimation obtenue avec $n = 2^l - 1$ points dans la quadrature sur l'estimation dite de référence, obtenus à l'aide de 1023 points dans la quadrature de Fejér.

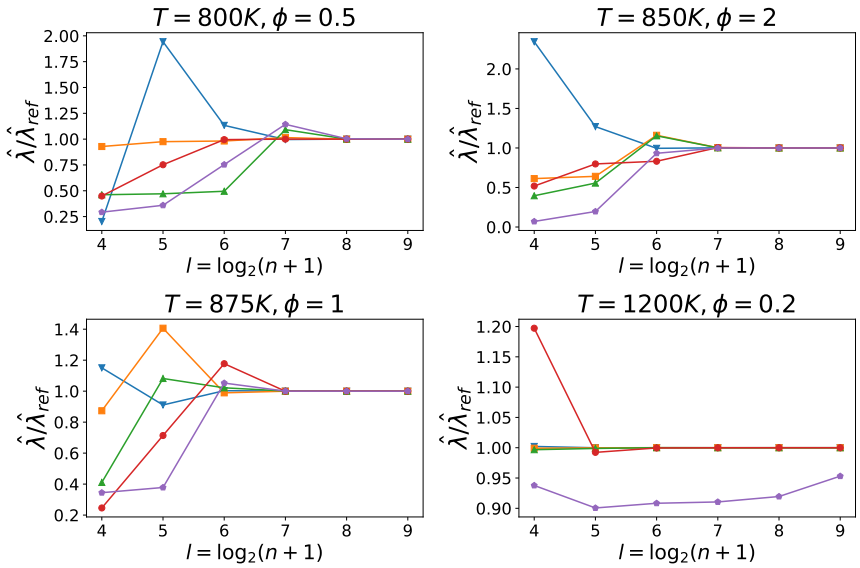

FIGURE 8.3 – *Rapports des estimations des* 5 *premières valeurs propres de l'expansion de Karhunen-Loève de* $\dot{\omega}_{Y_c}^{log,\psi}$ *pour les quatre conditions initiales sur l'estimation de référence obtenue avec* 1023 *points de quadrature, en fonction du nombre* $n = 2^l - 1$ *de points utilisés dans la quadrature de Fejér. Les triangles bas correspondent à la première valeur propre, les carrés à la seconde, les triangles hauts à la troisième, les ronds à la quatrième et les pentagones à la cinquième.*

Les rapports des estimations des valeurs propres sur l'estimation de référence tendent bien tous vers 1. Le niveau de quadrature l est à choisir suivant le nombre de valeurs propres à conserver, ainsi que le seuil de tolérance souhaité sur les valeurs propres. Un tel seuil de tolérance ne doit cependant pas être le même pour l'ensemble des valeurs propres, celles-ci étant ordonnées par ordre décroissant. Il est donc surtout important de reproduire le mieux possible les premières valeurs propres, qui sont celles qui auront le plus d'impact. Cette convergence des valeurs propres n'est pas la seule à prendre en compte, comme le suggère la figure 8.4, où le premier mode de l'expansion de Karhunen-Loève de $\dot{\omega}_{Y_c}^{log,\psi}$ est tracé pour l'ensemble des conditions initiales, et obtenu avec dif-

férents nombres de points dans la quadrature.

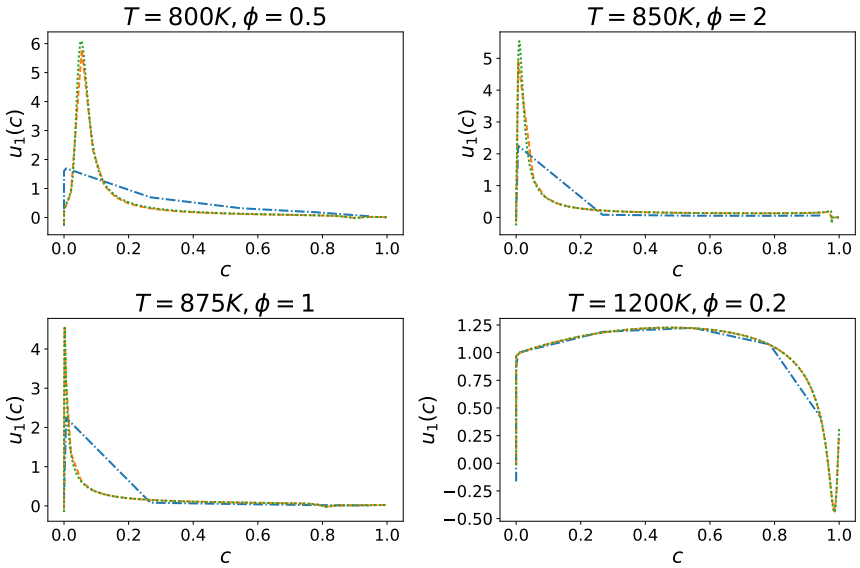

FIGURE 8.4 – *Premier mode de l'expansion de Karhunen-Loève de* $\dot{\omega}_{Y_c}^{log,\psi}$ *pour les points de fonctionnements choisis, calculés à partir d'une méthode de Nyström impliquant une interpolation linéaire et une quadrature de Fejér de* 7 *points (pointillés-tirets), de* 63 *points (tirets) et de* 511 *points (pointillés).*

Il est clair sur la figure 8.4 qu'un nombre trop limité de points dans la quadrature entraînera des erreurs dû à l'interpolation notamment, qui ne permettront pas une reproduction de $\dot{\omega}_{Y_c}^{log,\psi}$ assurant une bonne reproduction des incertitudes du processus stochastique C. Afin de déterminer le niveau de quadrature à choisir, une solution est d'observer l'impact du niveau de la quadrature directement sur certaines statistiques du processus stochastique reconstruit C, comme sa moyenne temporelle ou son écart-type temporel qui sont obtenus grâce à de nombreuses réalisations de celui-ci obtenues par résolution de l'équation différentielle ordinaire suivante :

$$\frac{dc}{dt} = \sum_{k=1}^{M} \sqrt{\lambda_k} u_k(c) \eta_k^{(i)} \tag{8.3}$$

La résolution de l'équation différentielle ordinaire précédente implique l'utilisation de réalisations $\eta_k^{(i)}$ des variables aléatoires η_k (4.46) introduites par l'expansion de Karhunen-Loève, pour lesquelles sont prises ici les réalisations obtenues à partir des réalisations ayant servi à la construction de la matrice d'auto-covariance du processus stochastique. Les figures 8.5 et 8.6 présentent respectivement les moyennes et les écart-types temporels du processus stochastique C obtenu par intégration de la troncature à un terme ($M = 1$) du

terme source modélisé, celui-ci ayant été calculé à l'aide de différents niveaux de quadratures.

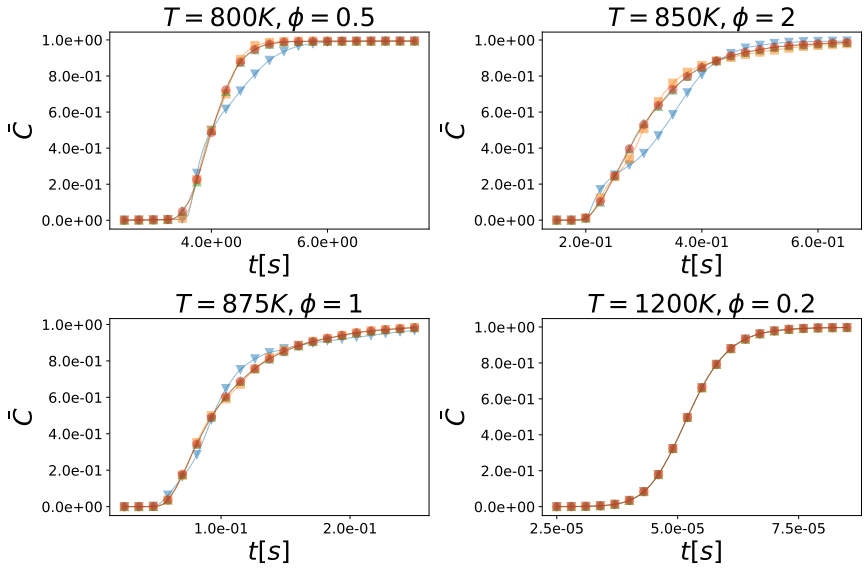

FIGURE 8.5 – *Moyennes temporelles du processus stochastique C obtenues par intégration de la troncation à un terme du terme source modélisé calculé avec différents niveaux de quadratures de la seconde quadrature de Fejér, pour les quatre conditions initiales. Les nombres de points utilisés pour les quadratures sont les suivants :* 31 *points (triangle hauts),* 63 *points (carrés),* 127 *points (triangle bas) et* 255 *points (ronds).*

On observe que les résultats sont similaires pour les résultats où le terme source a été calculé à l'aide de quadratures comportant plus de 127 points, alors qu'avec des quadratures contenant moins de points, des différences sont présentes, aussi bien au niveau de la moyenne que de l'écart-type. Ces résultats indiquent qu'utiliser 127 points dans la quadrature n'engendre pas d'erreurs trop importantes au niveau de la reconstruction du terme source par interpolation, qui impacteraient le processus stochastique C.

Dans la suite, sauf mention contraire, la seconde quadrature de Fejér utilisant 127 points sera utilisée pour le calcul des expansions de Karhunen-Loève de $\dot{\omega}_{Y_c}^{log,\psi}$.

8.1.1.4 Validation de l'expansion de Karhunen-Loève

Vu les résultats de convergence précédents, l'expansion de Karhunen-Loève pour $\dot{\omega}_{Y_c}^{log,\psi}$ a été calculée pour l'ensemble des quatre conditions initiales identifiées à l'aide de la seconde quadrature de Fejér impliquant 127 points, et une méthode de Quasi-Monte Carlo randomisée utilisant 10 séquences de Sobol de 4096 points chacune, randomisées à l'aide d'une méthode de Full-Scrambling. Les sommes des premières valeurs propres de l'expansion de Karhunen-Loève

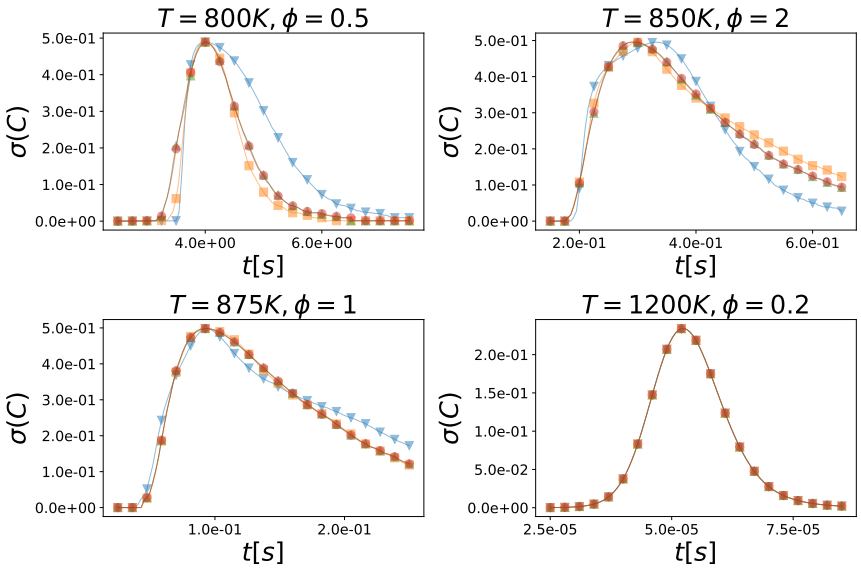

FIGURE 8.6 – *Écart-types temporels du processus stochastique C obtenus par intégration de la troncation à un terme du terme source modélisé calculé avec différents niveaux de quadratures de la seconde quadrature de Fejér, pour les quatre conditions initiales. Les nombres de points utilisés pour les quadratures sont les suivants :* 31 *points (triangle hauts),* 63 *points (carrés),* 127 *points (triangle bas) et* 255 *points (ronds).*

normalisées par la somme de l'ensemble des valeurs propres sont présentées sur la figure 8.7. Suivant les conditions initiales considérées, la somme cumulée des valeurs propres n'évolue pas de la même façon, et le nombre de modes à considérer pour reproduire une certaine quantité de la variance totale change donc avec les conditions initiales.

A partir de troncatures de cette expansion, il est possible de modéliser le terme source $\dot{\omega}_{Y_c}$. Pour caractériser complètement ce terme source modélisé, il est nécessaire de définir une loi de probabilité jointe pour les nouvelles variables aléatoires construites lors de cette expansion. En effet, l'expansion de Karhunen-Loève nous assure que ces variables aléatoires sont par construction centrées et réduites, et également décorrélées deux à deux. Les histogrammes pour les quatre premières nouvelles variables aléatoires η_1, η_2, η_3 et η_4, ainsi que des nuages de points de ces variables deux à deux, construits à partir des réalisations de la simulation de Quasi-Monte Carlo ayant servi au calcul de l'expansion de Karhunen-Loève sont représentés sur les figure 8.8 et 8.9, l'ensemble correspondant respectivement à la condition initiale à $T = 1200\ K$ et $\phi = 0.2$ et à la condition initiale à $T = 850K$ et $\phi = 2$.

Les histogrammes des quatre premières variables aléatoires introduites par l'expansion de Karhunen-Loève de $\dot{\omega}_{Y_c}^{\log,\psi}$ présentes sur la figure 8.8 sont quasiment symétriques, et présentent un aspect proche de gaussiennes centrées réduites. De plus les nuages de points ne laissent pas paraître de dépendances

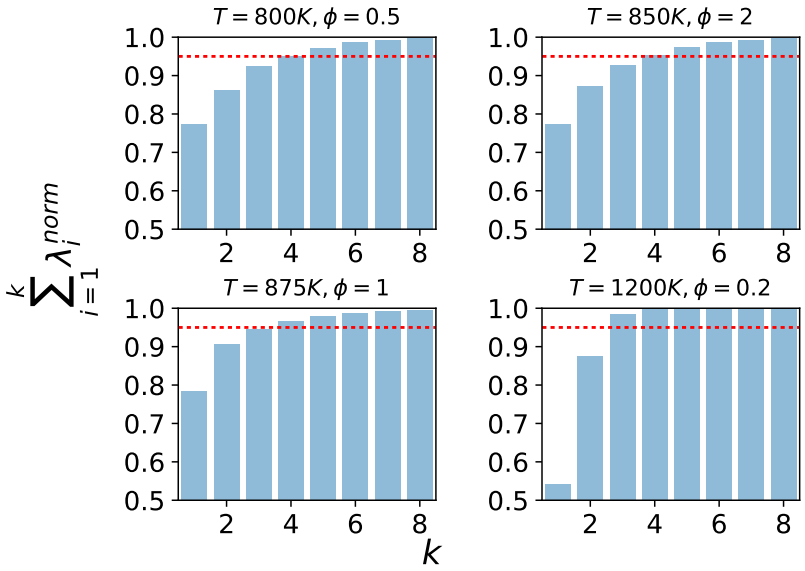

FIGURE 8.7 – *Sommes cumulées des premières valeurs propres normalisées des expansions de Karhunen-Loève du processus $\dot{\omega}_{Y_c}^{log,\psi}$ pour les quatre points de fonctionnements choisis. La ligne pointillée correspond à une valeur de* 95%.

entre les couples de variables aléatoires, et présentent eux aussi un aspect typique de gaussiennes bivariées centrées réduites. Il n'en est pas de même pour la figure 8.9 où les variables aléatoires ne sont clairement pas symétriques. De plus, elles présentent de fortes dépendances deux à deux comme le suggèrent les nuages de points.

Ces deux exemples nous montrent que suivant le cas considéré, la modélisation des nouvelles variables aléatoires introduites par l'expansion de Karhunen-Loève peut être plus ou moins complexe. Dans un premier temps, afin de ne pas introduire d'erreurs liées à la modélisation de ces variables aléatoires, les échantillons de ces variables aléatoires construits lors du calcul de l'expansion de Karhunen-Loève, qui par construction suivent la bonne loi de probabilité, seront utilisés. Sur la figure 8.10 sont présentés les diagrammes quantile-quantile pour le délai d'auto-allumage $\tau^{C=0.1}$ obtenu à l'aide de la chimie détaillée et celui obtenu à l'aide de différentes troncatures de l'expansion de Karhunen-Loève de $\dot{\omega}_{Y_c}^{log,\psi}$ en résolvant (8.3) pour les différentes conditions initiales.

La figure 8.10 montre que pour l'ensemble des conditions initiales étudiées, une troncature de l'expansion de Karhunen-Loève de $\dot{\omega}_{Y_c}^{log,\psi}$ comportant trois termes permet la bonne reproduction de la loi du délai d'auto-allumage. Le nombre de termes, et donc de variables aléatoires, à conserver dans la troncature de l'expansion n'est pas, pour l'ensemble des conditions initiales, le même que le nombre de variables à conserver dans le cadre d'une expansion en polynômes

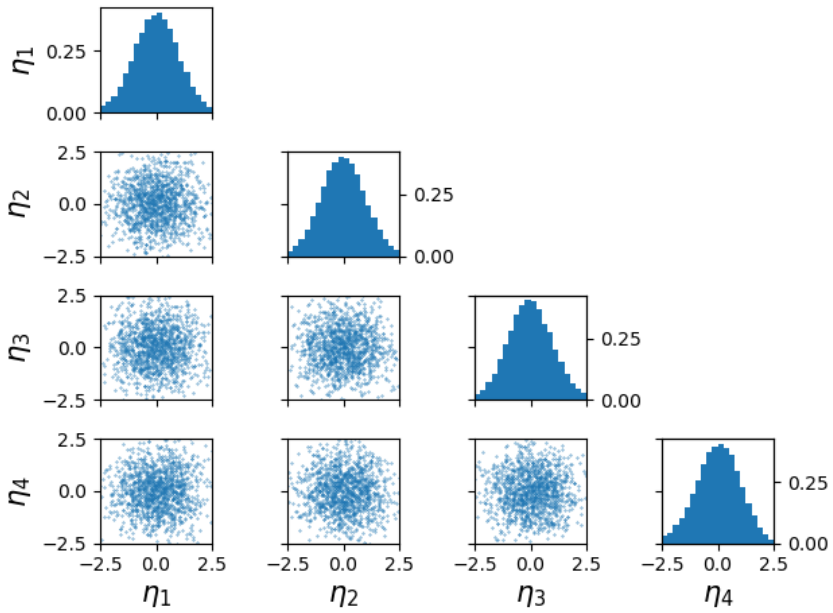

FIGURE 8.8 – *Sur la diagonale sont présents les histogrammes des quatre variables aléatoires η_1, η_2, η_3 et η_4 introduites par l'expansion de Karhunen-Loève de $\dot{\omega}_{Y_c}^{\log,\psi}$. Sous la diagonale sont représentés les nuages de points de ces variables aléatoires deux à deux obtenus à partir des échantillons de Quasi-Monte Carlo ayant été utilisés pour estimer la fonction d'auto-covariance. L'ensemble correspond à la condition initiale à $T = 1200\ K$ et $\phi = 0.2$.*

du chaos. Ainsi, pour le point à la température de 1200 K, conserver une seule variable pour l'expansion en polynômes du chaos permettait d'obtenir une meilleur reproduction de la loi de $\tau^{C=0.1}$ qu'avec une troncation à deux termes dans l'expansion de Karhunen-Loève. La bonne reproduction de l'incertitude sur le délai d'auto-allumage est bien entendu liée à la reproduction de $\dot{\omega}_{Y_c}^{log,\psi}$ par les troncatures de son expansion de Karhunen-Loève. En fait, pour le délai d'auto-allumage $\tau^{C=0.1}$, seules comptent les valeurs de $\dot{\omega}_{Y_c}$ pour les valeurs de c inférieures à 0.1, comme le montre la relation (7.5). Comme vu précédemment au chapitre 4, l'expansion de Karhunen-Loève cherche à maximiser l'intégrale de la variance des troncatures de celle-ci. Sur la figure 8.11 sont représentés pour les différentes conditions initiales la variance du processus $\dot{\omega}_{Y_c}^{log,\psi}$, ainsi que la proportion de la variance reproduite par les différentes troncatures en fonction de la variable d'avancement c, définie comme :

$$\frac{Var_{KL}^{(m)}(c)}{Var_{Tot}(c)} = \frac{\sum_{k=1}^{m} \lambda_k u_k(c)^2}{\sum_{k=1}^{+\infty} \lambda_k u_k(c)^2} \tag{8.4}$$

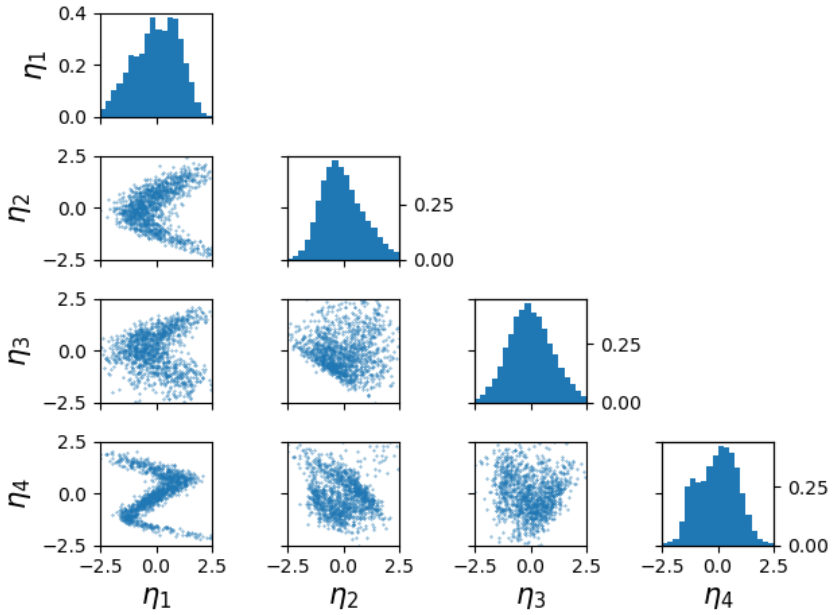

FIGURE 8.9 – *Sur la diagonale sont présents les histogrammes des quatre variables aléatoires η_1, η_2, η_3 et η_4 introduites par l'expansion de Karhunen-Loève de $\dot{\omega}_{Y_c}^{\log,\psi}$. Sous la diagonale sont représentés les nuages de points de ces variables aléatoires deux à deux obtenus à partir des échantillons de Quasi-Monte Carlo ayant été utilisés pour estimer la fonction d'auto-covariance. L'ensemble correspond à la condition initiale à $T = 850\ K$ et $\phi = 2$.*

La figure 8.11 montre que les premiers modes de l'expansion de Karhunen-Loève de $\dot{\omega}_{Y_c}^{\log,\psi}$ tendent à maximiser l'intégrale de la variance des troncatures, reproduisant préférablement la variance du processus stochastique là où celle-ci est importante sur un intervalle suffisamment large. La reproduction préférentielle de la variance du processus $\dot{\omega}_{Y_c}^{log,\psi}$ par les troncatures de l'expansion de Karhunen-Loève n'est cependant pas nécessairement adaptée à la bonne reproduction des incertitudes du délai d'auto-allumage $\tau^{C=0.1}$. Par exemple, pour la condition initiale à $T = 1200K$ et $\phi = 0.2$, les deux premiers modes de l'expansion de Karhunen-Loève reproduisent une grande partie de la variance pour des valeurs de c supérieure à 0.1, valeurs pour lesquelles la valeur de $\dot{\omega}_{Y_c}^{log}$ n'impacte pas $\tau^{C=0.1}$, mais un creux dans cette reproduction de variance existe pour des valeurs de c entre environ 10^{-7} et 10^{-1}, qui se trouve comblé par l'ajout du troisième mode et qui explique la bonne reproduction de l'incertitude du délai d'auto-allumage avec la troncature à trois termes et pas avec celle à deux termes. Cet aspect de reproduction de la variance du processus $\dot{\omega}_{Y_c}^{log}$ préférentiellement pour certaines valeurs de c est une voie d'amélioration possible pour une bonne reproduction des incertitudes sur le délai d'auto-allumage $\tau^{C=0.1}$ qui sera abordée dans la section suivante.

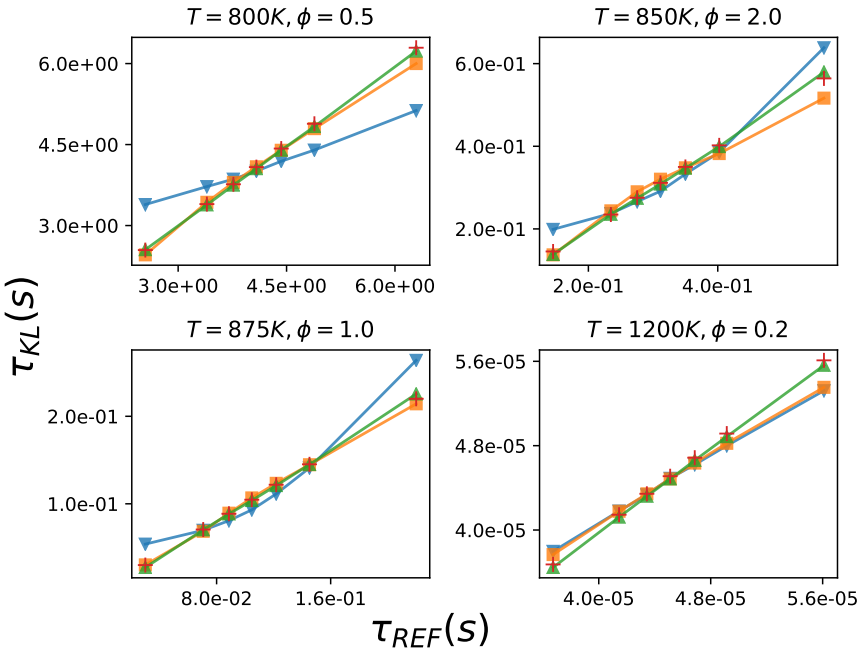

FIGURE 8.10 – *Diagramme quantile-quantile du délai d'auto-allumage $\tau^{C=0.1}$ pour les différentes conditions initiales, calculés à l'aide de la chimie détaillée incertaine et à l'aide d'une expansion de Karhunen-Loève de $\dot{\omega}_{Y_c}^{log,\psi}$, avec différentes troncatures de cette expansion : 1 paramètre incertain (triangles bas), 2 paramètres incertains (carrés), 3 paramètres incertains (triangles hauts).*

Une information intéressante à mettre en parallèle de cette reproduction de variance en fonction de c et à mettre en relation avec l'expansion en polynôme du chaos est la contribution des variables aléatoires initiales aux nouvelles variables aléatoires introduite par l'expansion de Karhunen-Loève. Cela peut être fait encore une fois à l'aide des indices de Sobol des variables aléatoires initiales pour les nouvelles variables aléatoires. Les principaux indices de Sobol du premier ordre sont représentés pour les différents points de fonctionnements sur la figure 8.12.

Les réactions préalablement identifiées comme impactant le terme source $\dot{\omega}_{Y_c}$ impactent également les nouvelles variables aléatoires η_k. Chacune des nouvelles variables aléatoires est associée à un mode de l'expansion de Karhunen-Loève, lequel mode contribue à la reproduction de la variance du processus pour une partie des valeurs de c. En s'intéressant au point de fonctionnement à $1200K$, il est possible d'illustrer le lien existant entre les informations des figures 8.11 et 8.12 avec les informations de la figure 7.4. En effet, le premier mode possède un pic de reproduction de la variance du processus stochastique pour des valeurs de c comprise entre 10^{-2} et 1, là où la réaction $H_2+OH = H_2O+H$ est prépondérante pour expliquer la variance de $\dot{\omega}_{Y_c}^{log}$, ce qui se retrouve dans le fait que cette même réaction est la principale contributrice à la variance de

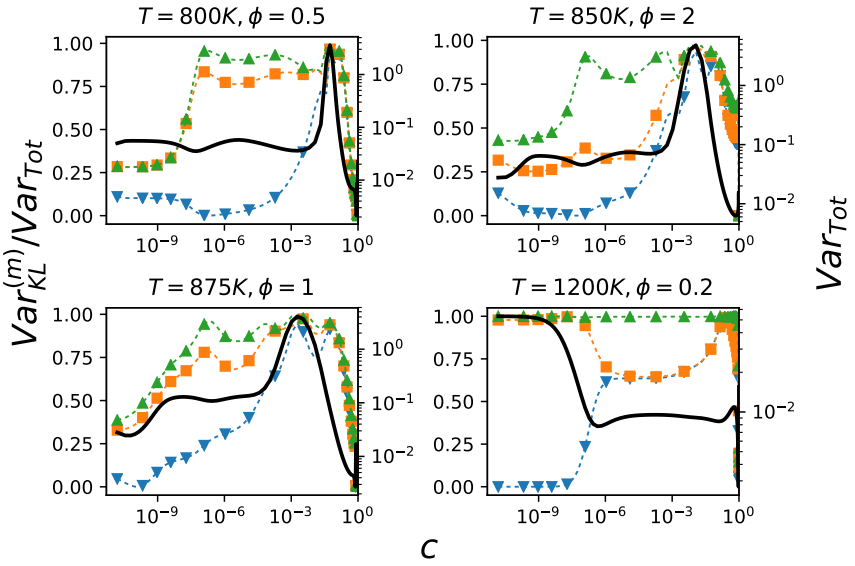

FIGURE 8.11 – *Proportion de la variance totale reproduite par les troncatures de l'expansion de Karhunen-Loève de $\dot{\omega}_{Y_c}^{log,\psi}$ (échelle de gauche), pour un ordre de troncature m valant 1 (triangles bas), 2 (carrés) et 3 (triangles hauts), et variance du processus $\dot{\omega}_{Y_c}^{log}$ (échelle de droite) en fonction de la variable d'avancement c (ligne pleine noire) pour les différentes conditions initiales.*

la variable aléatoire η_1. Le second mode tend lui à reproduire la variance du processus stochastique pour des valeurs de c inférieure à 10^{-7}, où la réaction $H + HO_2 = H_2 + O_2$ a une importance prépondérante vis à vis de $\dot{\omega}_{Y_c}^{log}$, ce qui se retrouve encore une fois au niveau des indices de Sobol de la variable aléatoire η_2. Enfin, le troisième mode apporte une contribution à la variance du processus stochastique pour des valeurs de c comprises entre 10^{-7} et 10^{-2}, où la réaction $H + O_2 = OH + O$ joue un rôle important ce qui se retrouve encore une fois au niveau des indices de Sobol de η_3. Les nouvelles variables aléatoires η_k peuvent donc s'expliquer à l'aide des variables aléatoires initiales au travers des modes de l'expansion de Karhunen-Loève et de la contribution des variables aléatoires initiales à la variance du processus stochastique pour les différentes valeurs de c. On observe que ces nouvelles variables aléatoires sont en fait un mix des variables aléatoires initiales.

L'objectif du chapitre précédent était la bonne reproduction du processus stochastique C, et non celle du délai d'auto-allumage $\tau^{C=0.1}$. Il est possible de vérifier que la bonne reproduction des incertitudes du délai d'auto-allumage permet encore une fois la bonne reproduction du processus stochastiques C. Sur les figures 8.13 et 8.14 sont présentés respectivement les moyennes et les écart-types temporels de C pour différentes troncatures de l'expansion de Karhunen-

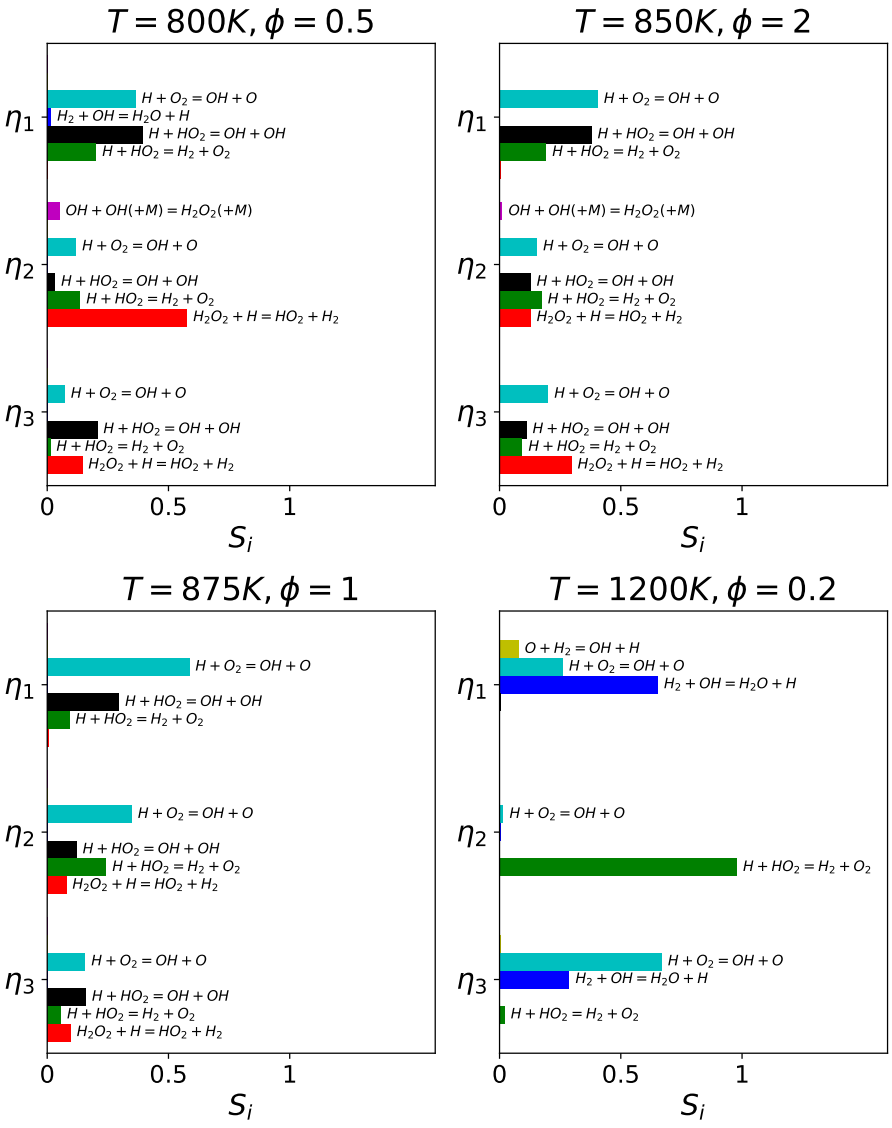

FIGURE 8.12 – *Indices de Sobol du premier ordre des principales réactions pour les trois premières nouvelles variables aléatoires obtenues par l'expansion de Karhunen-Loève de* $\dot{\omega}_{Y_c}^{log,\psi}$ *pour les différentes conditions initiales.*

Loève, ainsi que la référence obtenue à l'aide de la chimie détaillée.

Les résultats de ces deux figures sont en accord avec ceux sur le délai d'auto-allumage de la figure 8.10 pour l'ensemble des conditions initiales, une reproduction convenable des statistiques du processus stochastique C étant obtenue à partir d'une troncature comportant 3 termes en utilisant pour les η_i les échantillons de Quasi-Monte Carlo.

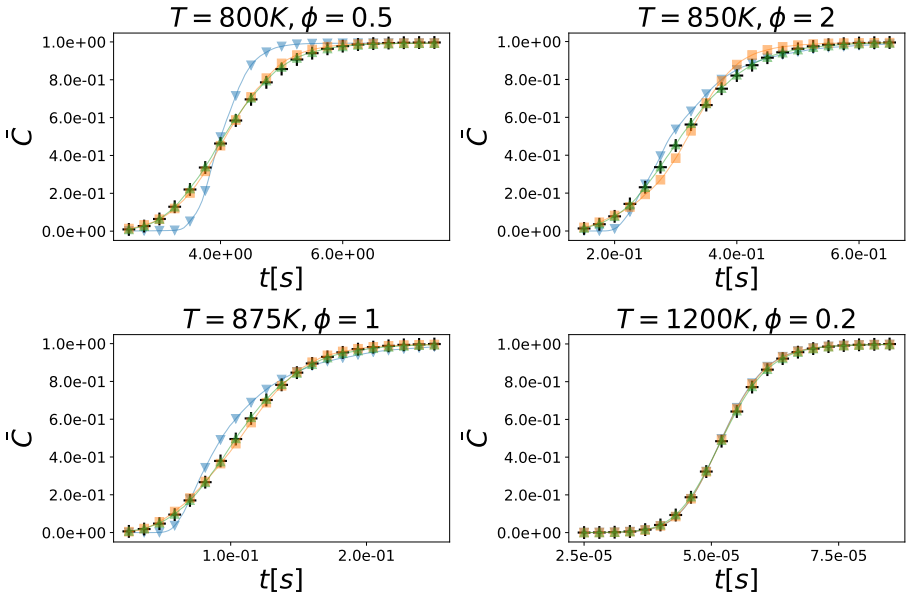

FIGURE 8.13 – *Moyennes temporelles du processus stochastique C pour les quatre conditions initiales, obtenue à l'aide de la chimie détaillée (croix), ainsi que de troncatures de l'expansion de Karhunen-Loève de $\dot{\omega}_{Y_c}^{\log,\psi}$ à l'ordre 1 (tirets), à l'ordre 2 (tirets-pointillés) et à l'ordre 3 (pointillés).*

8.1.1.5 Amélioration de l'expansion de Karhunen-Loève

Il ressort des constatations faites précédemment qu'une allure différente du profil de variance du processus stochastique à modéliser en fonction de la variable d'avancement c mènerait à une expansion de Karhunen-Loève dont les troncatures donnerait une reproduction différente de sa variance en fonction de c. Suivant cette idée, on considère un nouveau processus $\dot{\omega}_{Y_c}^{log,\psi,var}$, résultant d'une modification simple du processus $\dot{\omega}_{Y_c}^{log,\psi}$ donnée par l'expression (8.5) et impliquant une fonction f_{Var}.

$$\dot{\omega}_{Y_c}^{log,\psi,var}(c,\mathbf{A}) = f_{Var}(c)\dot{\omega}_{Y_c}^{log,\psi}(c,\mathbf{A}) \tag{8.5}$$

L'objectif est maintenant de déterminer un bon candidat pour la fonction f_{Var} permettant de modifier le profil de variance du processus stochastique, avec pour objectif la bonne reproduction du délai d'auto-allumage $\tau^{C=0.1}$ avec le plus petit nombre possible de termes dans la troncature. En fait, il est essentiel de reproduire correctement le processus T, dont le délai d'auto-allumage n'est que sa valeur prise en $c = 0.1$, et donc l'ensemble des délais d'auto-allumage s'exprimant comme une variable aléatoire $T(c)$ pour une valeur de la variable d'avancement c fixée. Pour rappel, cette variable aléatoire $T(c)$ peut s'exprimer

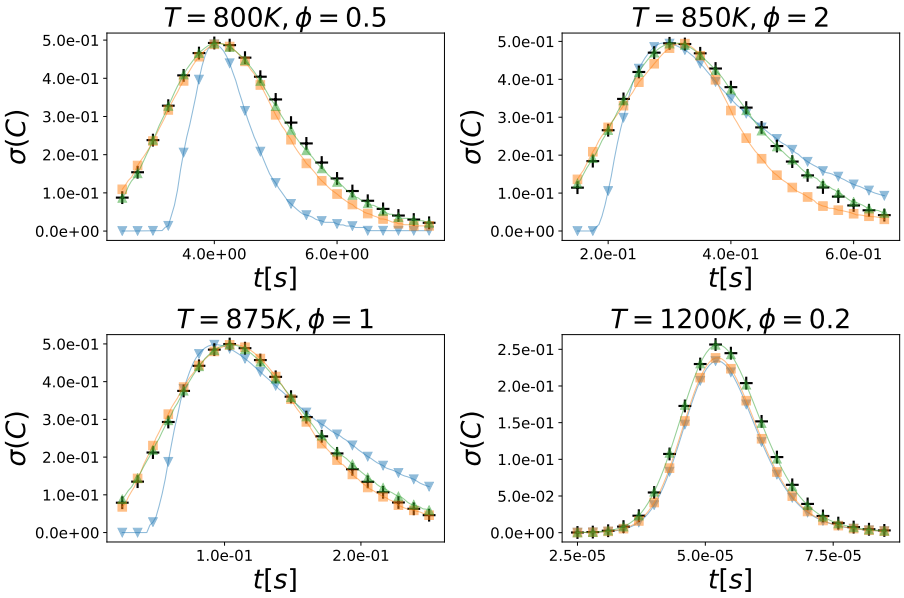

FIGURE 8.14 – *Écart-types temporels du processus stochastique C pour les quatres points de fonctionnements choisis, obtenu à l'aide de la chimie détaillée (croix), ainsi que de troncatures de l'expansion de Karhunen-Loève à l'ordre 1 (tirets), à l'ordre 2 (tirets-pointillés) et à l'ordre 3 (pointillés).*

en fonction du processus $\dot{\omega}_c$ (qui est proportionnel à $\dot{\omega}_{Y_c}$), suivant l'expression (8.6).

$$T(c, \mathbf{A}) = \int_0^c \frac{dc'}{\dot{\omega}_c(c')} \tag{8.6}$$

A partir de l'expression (8.6), il est possible d'obtenir une expression pour la variance de $T(c)$, dont le détail est donné par (8.7).

$$\begin{aligned} Var\left[T(c)\right] &= E\left[\left(\int_0^c \frac{dc'}{\dot{\omega}_c(c')}\right)^2\right] - E\left[\int_0^c \frac{dc'}{\dot{\omega}_c(c')}\right]^2 \\ &= \int_0^c \int_0^c E\left[\frac{1}{\dot{\omega}_c(c')\dot{\omega}_c(c'')}\right] dc'dc'' - \int_0^c \int_0^c E\left[\frac{1}{\dot{\omega}_c(c')}\right] E\left[\frac{1}{\dot{\omega}_c(c'')}\right] dc'dc'' \\ &= \int_0^c \int_0^c \left(E\left[\frac{1}{\dot{\omega}_c(c')\dot{\omega}_c(c'')}\right] - E\left[\frac{1}{\dot{\omega}_c(c')}\right] E\left[\frac{1}{\dot{\omega}_c(c'')}\right]\right) dc'dc'' \\ &= \int_0^c \int_0^c Cov\left[\frac{1}{\dot{\omega}_c(c')}, \frac{1}{\dot{\omega}_c(c'')}\right] dc'dc'' \end{aligned} \tag{8.7}$$

L'expression (8.7) permet d'obtenir la variance de la variable aléatoire $T(c)$ en fonction de la covariance du processus stochastique $\dot{\omega}_c^{inv} = 1/\dot{\omega}_c$. Sous l'hypothèse que les valeurs du processus stochastiques $\dot{\omega}_c$ ne sont pas trop éloignées de sa moyenne, on peut exprimer le processus stochastique $\dot{\omega}_c^{inv}$ comme suit, $\Delta\dot{\omega}_c$ correspondant au processus $\dot{\omega}_c$ centré :

$$\dot{\omega}_c^{inv}(c, \mathbf{a}) \approx \frac{1}{E\left[\dot{\omega}_c(c)\right]} \left(1 - \frac{\Delta\dot{\omega}_c(c, \mathbf{a})}{E\left[\dot{\omega}_c(c)\right]}\right) \tag{8.8}$$

En réinjectant l'expression approchée de $\dot{\omega}_c^{inv}$ obtenue dans (8.8) dans l'expression (8.7) de la variance de $T(c)$, on obtient :

$$\begin{aligned} Var\left[T(c)\right] &\approx \int_0^c \int_0^c Cov\left[\frac{\Delta\dot{\omega}_c(c')}{E\left[\dot{\omega}_c(c')\right]^2}, \frac{\Delta\dot{\omega}_c(c'')}{E\left[\dot{\omega}_c(c'')\right]^2}\right] dc'dc'' \\ &= \int_0^c \int_0^c Cov\left[\frac{\dot{\omega}_c(c')}{E\left[\dot{\omega}_c(c')\right]^2}, \frac{\dot{\omega}_c(c'')}{E\left[\dot{\omega}_c(c'')\right]^2}\right] dc'dc'' \end{aligned} \tag{8.9}$$

L'expression approchée (8.9) pour la variance de $T(c)$ montre que celle-ci peut être reproduite correctement à l'aide de la connaissance de l'auto-covariance du processus stochastique $\dot{\omega}_c/E\left[\dot{\omega}_c\right]^2$, ou indifféremment à l'aide de la connaissance de l'auto-covariance du processus stochastique $\dot{\omega}_{Y_c}/E\left[\dot{\omega}_{Y_c}\right]^2$, les deux étant égaux. Ces derniers résultats conduisent à envisager de considérer l'expansion de Karhunen-Loève du processus $\dot{\omega}_{Y_c}/E\left[\dot{\omega}_{Y_c}\right]^2$ pour reproduire correctement les délais d'auto-allumage $T(c)$. Cependant, afin d'assurer la positivité, le logarithme du terme source est considéré, et cela amène à considérer $\dot{\omega}_{Y_c}^{log}$. Or, sous la même hypothèse que précédemment concernant les variations de $\dot{\omega}_{Y_c}$ qui restent petite devant $E\left[\dot{\omega}_{Y_c}\right]$, on peut approximer $\dot{\omega}_{Y_c}^{log}$ comme :

$$\dot{\omega}_{Y_c}^{log}(c, \mathbf{a}) \approx \ln\left(E\left[\dot{\omega}_{Y_c}(c)\right]\right) + \frac{\Delta\dot{\omega}_{Y_c}(c, \mathbf{a})}{E\left[\dot{\omega}_{Y_c}(c)\right]} \tag{8.10}$$

Compte tenu des propriétés de la covariance, et en se servant des expressions (8.10) et (8.9), on peut approximer l'expression de la variance de $T(c)$ comme :

$$Var\left[T(c)\right] \approx \int_0^c \int_0^c Cov\left[\frac{\dot{\omega}_{Y_c}^{log}(c')}{E\left[\dot{\omega}_{Y_c}(c')\right]}, \frac{\dot{\omega}_{Y_c}^{log}(c'')}{E\left[\dot{\omega}_{Y_c}(c'')\right]}\right] dc'dc'' \tag{8.11}$$

L'expression (8.11) permet d'exprimer la variance de $T(c)$ en fonction de l'auto-covariance du processus $\dot{\omega}_{Y_c}^{log}/E\left[\dot{\omega}_{Y_c}\right]$, s'exprimant simplement en fonction de $\dot{\omega}_{Y_c}^{log}$. En réalité, pour les raisons évoquées précédemment concernant le

maillage nécessaire à la capture de l'auto-allumage du mélange, un changement de variable ψ est utilisé, et le processus stochastique initialement utilisé est $\dot{\omega}_{Y_c}^{log,\psi}$. En partant du membre de droite dans l'expression (8.11), et en réalisant un changement de variable dans les intégrales, il est possible d'exprimer la variance de $T(c)$ comme suit :

$$\begin{aligned} Var\left[T(c)\right] &\approx \int_0^c \int_0^c Cov\left[\frac{\dot{\omega}_{Y_c}^{log}(c')}{E\left[\dot{\omega}_{Y_c}(c')\right]}, \frac{\dot{\omega}_{Y_c}^{log}(c'')}{E\left[\dot{\omega}_{Y_c}(c'')\right]}\right] dc'dc'' \\ &= \int_0^{\psi^{-1}(c)} \int_0^{\psi^{-1}(c)} Cov\left[\frac{\dot{\omega}_{Y_c}^{log,\psi}(\xi')}{E\left[\dot{\omega}_{Y_c}(\psi(\xi'))\right]}, \frac{\dot{\omega}_{Y_c}^{log,\psi}(\xi'')}{E\left[\dot{\omega}_{Y_c}(\psi(\xi''))\right]}\right] \psi^{'}(\xi')\psi^{'}(\xi'')d\xi'd\xi'' \\ &= \int_0^{\psi^{-1}(c)} \int_0^{\psi^{-1}(c)} Cov\left[\frac{\psi^{'}(\xi')\dot{\omega}_{Y_c}^{log,\psi}(\xi')}{E\left[\dot{\omega}_{Y_c}(\psi(\xi'))\right]}, \frac{\psi^{'}(\xi'')\dot{\omega}_{Y_c}^{log,\psi}(\xi'')}{E\left[\dot{\omega}_{Y_c}(\psi(\xi''))\right]}\right] d\xi'd\xi'' \end{aligned} \tag{8.12}$$

La dernière expression permet de considérer un nouveau processus stochastique $\dot{\omega}_{Y_c}^{\log,\psi,opt_1}$ dont l'expression est donnée par l'expression (8.13), et pour lequel l'expansion de Karhunen-Loève sera construite et utilisée afin d'approximer le terme source $\dot{\omega}_{Y_c}$ de la variable d'avancement.

$$\dot{\omega}_{Y_c}^{\log,\psi,opt_1} = \frac{\psi^{'}(\xi')\dot{\omega}_{Y_c}^{\log,\psi}(\xi')}{E\left[\dot{\omega}_{Y_c}(\psi(\xi'))\right]} \tag{8.13}$$

En pratique, pour des raisons de simplification de l'implémentation, le dénominateur de l'expression (8.13) a été approximé par l'exponentiel de l'espérance du logarithme de $\dot{\omega}_{Y_c}$, approximation d'autant plus vraie que les fluctuations de $\dot{\omega}_{Y_c}$ autour de sa moyenne sont faibles. Cela revient à finalement considérer le processus $\dot{\omega}_{Y_c}^{\log,\psi,opt_2}$ donné par (8.14) pour lequel l'expansion de Karhunen-Loève sera construite et utilisée afin d'approximer le terme source $\dot{\omega}_{Y_c}$ de la variable d'avancement.

$$\dot{\omega}_{Y_c}^{\log,\phi,opt_2} = \frac{\psi^{'}(\xi')\dot{\omega}_{Y_c}^{\log,\psi}(\xi')}{\exp\left(E\left[\dot{\omega}_{Y_c}^{\log,\psi}(\xi')\right]\right)} \tag{8.14}$$

Le processus stochastique $\dot{\omega}_{Y_c}^{\log,\psi,opt_2}$ consiste en fait en le processus $\dot{\omega}_{Y_c}^{\log}$ précédemment utilisé pour le calcul de l'utilisation de Karhunen-Loève, que l'on a multiplié par une fonction f_{scal}, dont l'expression est donnée par (8.15), dépendant de ξ, ou de manière équivalente de c, afin de présenter un profil de

variance différent en fonction de c.

$$f_{scal}(\xi) = \frac{\psi'(\xi')}{\exp\left(E\left[\dot{\omega}_{Y_c}^{\log,\psi}(\xi')\right]\right)} \tag{8.15}$$

Le profil de cette fonction f_{scal} est présenté sur la figure 8.15 en fonction de c et non de ξ, pour les différents points de fonctionnements. Pour l'ensemble des points, on peut observer un premier pic pour des faibles valeurs de c. Ce pic correspond aux valeurs de c impactant le plus les délais d'auto-allumages $T(c)$, et qui seront donc ici privilégier par rapport aux autres valeurs de c. Une autre particularité que présente cette fonction est sa divergence pour $c = 1$. Cette divergence s'explique par le fait que le dénominateur dans l'expression de la fonction f_{scal} tend vers 0 pour la valeur $c = 1$ correspondant à l'équilibre thermodynamique. Cette divergence peut également se relier à la variance du délai d'auto-allumage $T(c)$ pour c tendant vers 1. La variable aléatoire $T(1)$ n'a en théorie pas de moment d'ordre 1, et de surcroît pas de moments d'ordre 2 non plus.

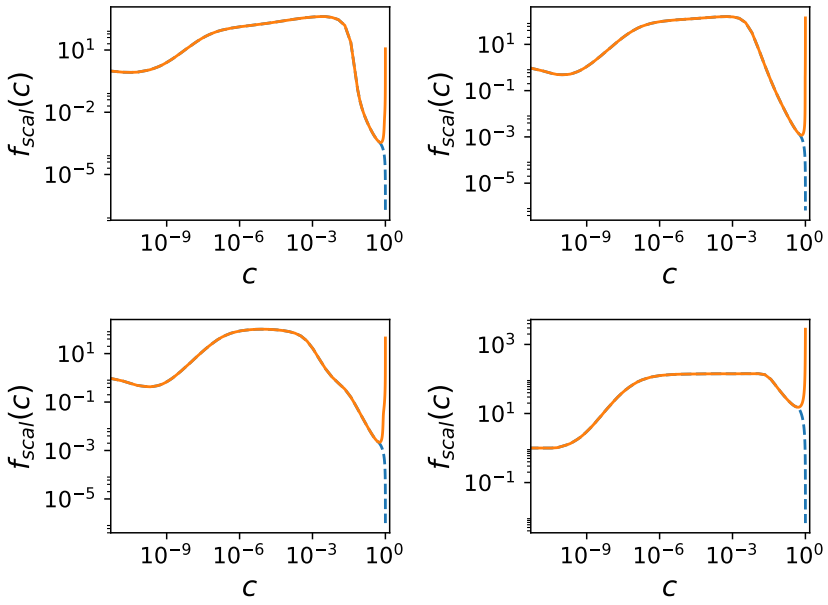

FIGURE 8.15 – *Profils des fonctions f_{scal} en fonction de c pour les différents points de fonctionnements. La ligne pleine correspond à la fonction f_{scal} non modifiée, alors que la ligne pointillé correspond à la fonction f_{scal} pour laquelle les valeurs de c supérieure à* 0.5 *ont été atténuées.*

Cette divergence de la fonction f_{scal} est problématique, puisqu'elle donne une grande importance aux valeurs de c proche de 1, n'impactant pas le début

du processus d'allumage et donc le délai d'auto-allumage $\tau^{C=0.1}$. Pour remédier à ce problème, le choix d'atténuer la valeur de f_{scal} pour les valeurs de c proche de 1 est fait. Les profils pointillés sur la figure 8.15 correspond aux profils de f_{scal} pour lesquels l'importance des valeurs de c proche de 1 ont été atténuées, en choisissant un point c_{cut}, ici égal à 0.5, et en appliquant l'expression (8.16).

$$\forall c \geq c_{cut}, f_{scal}(c) = f_{scal}(c_{cut})\frac{1-c}{1-c_{cut}} \tag{8.16}$$

Les processus stochastiques $\dot{\omega}_{Y_c}^{\log,\psi}$ et $\dot{\omega}_{Y_c}^{\log,\psi,opt_2}$ étant différents, ils n'ont pas la même expansion de Karhunen-Loève. La différence des expansions peut se voir à travers les valeurs propres λ_k. Sur la figure 8.16 sont représentées les sommes cumulées normalisées des valeurs propres de $\dot{\omega}_{Y_c}^{\log,\psi,opt_2}$, à comparer avec la figure 8.7. Cette comparaison des deux figures permet d'observer que la convergence de la somme cumulée normalisée des valeurs propres vers 1 est plus rapide pour le processus $\dot{\omega}_{Y_c}^{\log,\psi,opt_2}$ que pour le processus $\dot{\omega}_{Y_c}^{\log,\psi}$, signifiant que la portion de variance totale du processus stochastique reproduite par les troncatures est plus importante pour un nombre de termes donné présents dans la troncature.

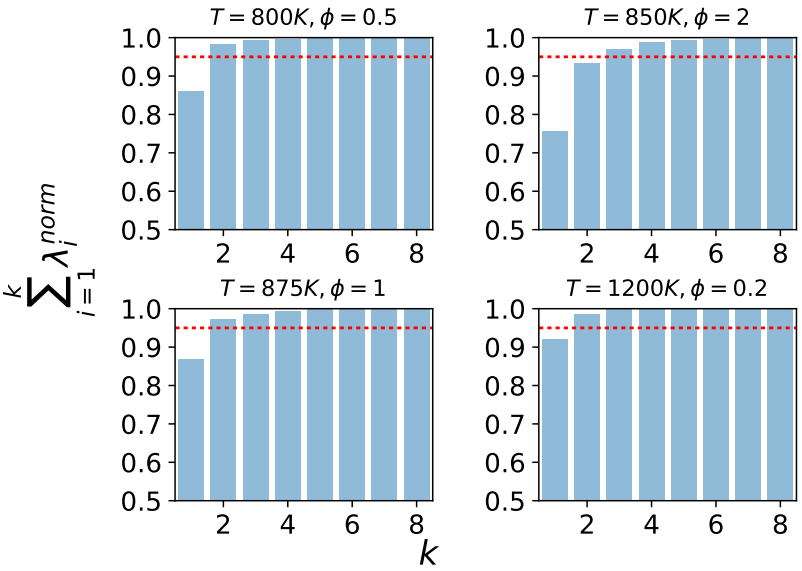

Figure 8.16 – *Sommes cumulées des premières valeurs propres normalisées des expansions de Karhunen-Loève du processus $\dot{\omega}_{Y_c}^{log,\psi,opt_2}$ pour les quatre conditions initiales. La ligne pointillée correspond à une valeur de 95%.*

La somme cumulée normalisée des valeurs propres donne une information globale de la reproduction de la variance totale du processus stochastique, que

l'on peut détailler en s'intéressant à la reproduction de la variance par les troncatures en fonction de c, qui est faite sur la figure 8.17, à comparer avec la figure 8.11. Encore une fois, les troncatures tendent à reproduire en premier la variance du processus stochastique pour les intervalles de c suffisamment large et présentant une variance importante, afin de reproduire le plus possible de l'intégrale de la variance du processus stochastique. Le profil de variance du processus stochastique $\dot{\omega}_{Y_c}^{\log,\psi,opt_2}$ étant différent de celui de $\dot{\omega}_{Y_c}^{\log,\psi}$, cette reproduction de la variance des troncatures ne se fait pas aux mêmes valeurs de c, en particulier elle a tendance à se faire pour des faibles valeurs de c où la variance du processus stochastique $\dot{\omega}_{Y_c}^{\log,\psi,opt_2}$ y est plus importante que pour le processus stochastique $\dot{\omega}_{Y_c}^{\log,\psi}$.

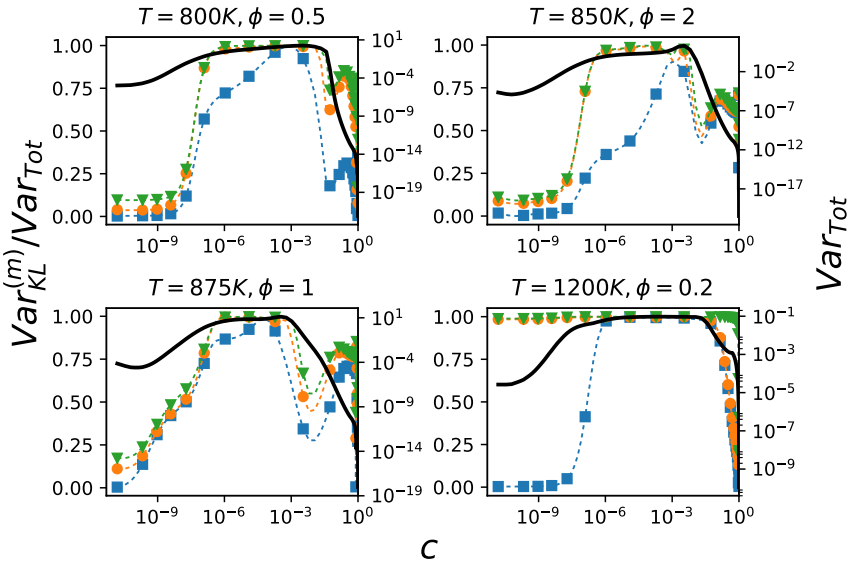

FIGURE 8.17 – *Proportion de la variance reproduite par les troncations de l'expansion de Karhunen-Loève de $\dot{\omega}_{Y_c}^{log,opt_2}$, pour un ordre de troncature m valant 1 (carrés), 2 (cercles) et 3 (triangles bas) et variance du processus $\dot{\omega}_{Y_c}^{log,opt_2}$ en fonction de la variable d'avancement c pour les différentes conditions initiales.*

L'expansion de Karhunen-Loève du processus stochastique $\dot{\omega}_{Y_c}^{\log,\psi,opt_2}$ ayant été construite, il est possible de l'utiliser afin d'obtenir un modèle pour le terme source incertain $\dot{\omega}_{Y_c}$ de la variable d'avancement. Encore une fois, ce modèle permet d'obtenir des réalisations du terme source $\dot{\omega}_{Y_c}$ qu'il est possible d'utiliser afin de résoudre l'équation différentielle (7.4) servant à obtenir l'évolution temporelle de la variable d'avancement normalisée c. A partir de ces réalisations, il est possible d'obtenir des statistiques du délai d'auto-allumage $\tau^{C=0.1}$, qu'il est possible de comparer avec le délai d'auto-allumage obtenu à partir de

la chimie détaillée incertaine.

Sur la figure 8.18 sont présentés, pour les différentes conditions initiales, les diagrammes quantile-quantile du délai d'auto-allumage $\tau^{C=0.1}$ obtenu à l'aide de la chimie détaillé et à l'aide du terme source incertain modélisé grâce à une expansion de Karhunen-Loève tronquée de 1, 2 et 3 termes de $\dot{\omega}_{Y_c}^{\log,\psi,opt}$. On observe que pour l'ensemble des points de fonctionnements, la reproduction du délai d'auto-allumage incertain $\tau^{C=0.1}$, à ordre de troncature fixé, est meilleure que dans le cas où l'expansion de Karhunen-Loève du processus stochastique $\dot{\omega}_{Y_c}^{\log,\psi}$ avait été utilisé. Ce nouveau processus permet d'avoir des résultats satisfaisant avec un ordre de troncature de 1 pour l'ensemble des conditions initiales considérées, alors qu'il en fallait 3 lors de l'utilisation du processus stochastique $\dot{\omega}_{Y_c}^{\log,\psi}$.

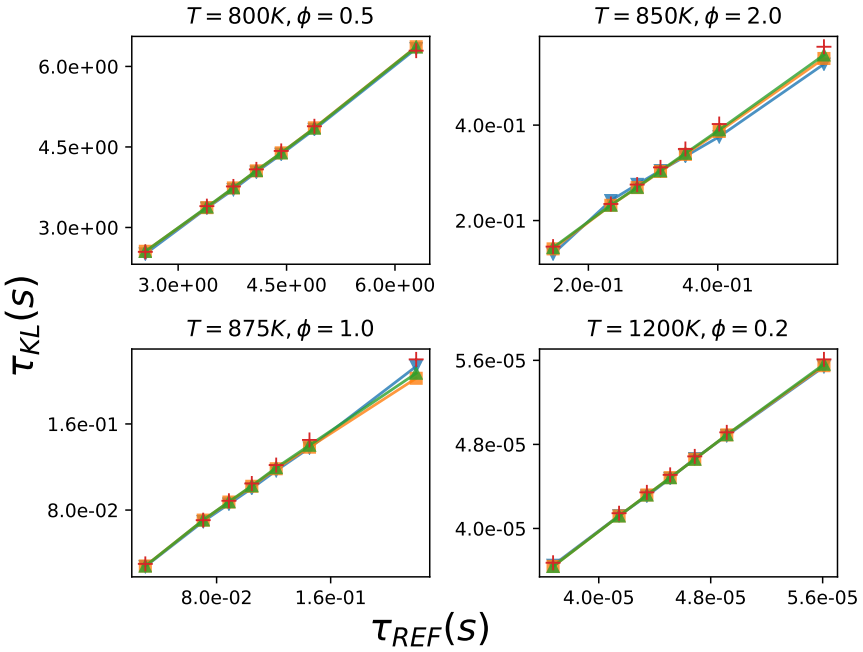

FIGURE 8.18 – *Diagramme quantile-quantile du délai d'auto-allumage $\tau^{C=0.1}$ pour les différentes conditions initiales, calculés à l'aide de la chimie détaillée incertaine et à l'aide d'une expansion de Karhunen-Loève du terme source $\dot{\omega}_{Y_c}$, avec différentes troncatures de cette expansion : 1 paramètre incertain (triangles bas), 2 paramètres incertains (carrés), 3 paramètres incertains (triangles hauts).*

Il est possible de vérifier que la bonne reproduction des incertitudes du délai d'auto-allumage permet encore une fois la bonne reproduction du processus stochastiques C. Sur les figures 8.19 et 8.20 sont présentés respectivement les moyennes et les écart-types temporelles de C pour différents ordres de troncatures de l'expansion de Karhunen-Loève ainsi que la référence obtenue à l'aide de la chimie détaillée.

Les résultats de ces deux figures sont en accord avec ceux sur le délai

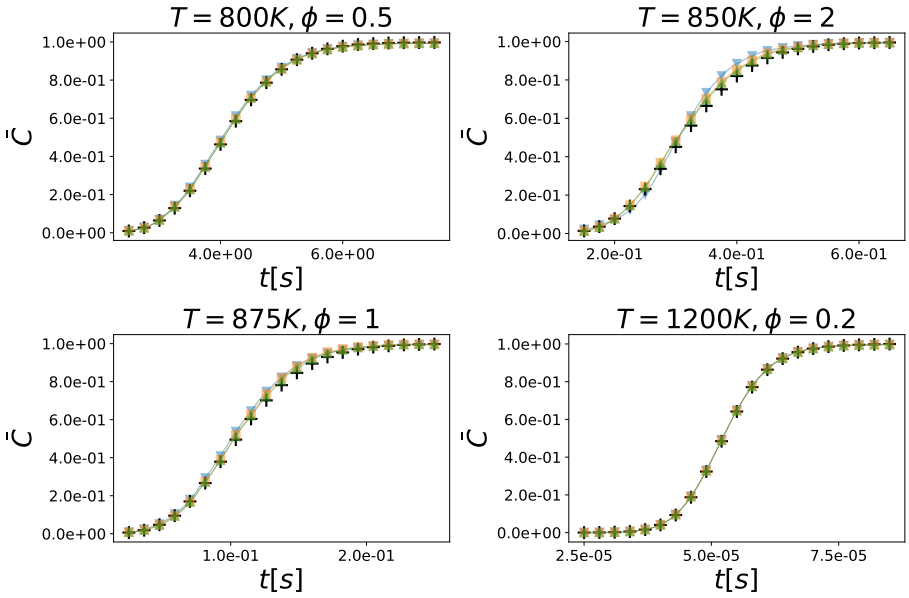

FIGURE 8.19 – *Moyennes temporelles du processus stochastique C pour les quatre conditions initiales, obtenue à l'aide de la chimie détaillée (croix), ainsi que de troncatures de l'expansion de Karhunen-Loève contenant* 1 *terme (triangles bas),* 2 *termes (carrés) et* 3 *termes (triangles hauts).*

d'auto-allumage de la figure 8.18 pour l'ensemble des conditions initiales, et une reproduction convenable des statistiques du processus stochastique C peut être obtenue ici à partir d'une troncature à 1 terme.

Les utilisations des expansions de Karhunen-Loève ont jusqu'à présent été réalisées en utilisant les échantillons des nouvelles variables aléatoires η_k calculés à l'aide des réalisations du processus stochastique étudié nécessaires à l'estimation de la matrice de covariance de ce même processus stochastique. Les validations faites revenait donc à obtenir les statistiques suivant une méthode de Monte Carlo ou de Quasi-Monte Carlo randomisé, suivant que les échantillons utilisés avaient été obtenues à l'aide de réalisations de Monte Carlo ou de Quasi-Monte Carlo randomisé du processus. L'objectif à terme est d'obtenir des statistiques à l'aide de méthodes de quadratures efficaces, et il faut donc être en mesure d'échantillonner suivant ces nouvelles variables aléatoires, ce qui nécessite de définir une loi de probabilité jointe pour l'ensemble de ces nouvelles variables aléatoires.

8.1.1.6 Modélisation des nouvelles variables aléatoires η

Différents modèles pour la loi de probabilité jointe des variables aléatoires η_k peuvent être envisagés, chaque modèle reposant sur des hypothèses entraînant une modélisation plus ou moins complexe. Sans être exhaustif, trois mo-

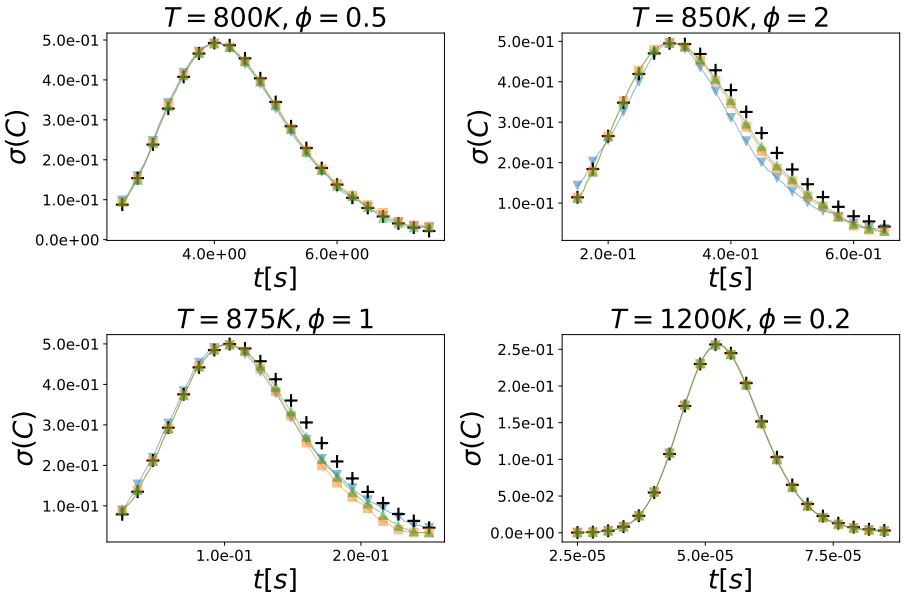

FIGURE 8.20 – *Écart-types temporels du processus stochastique C pour les quatre conditions initiales, obtenu à l'aide de la chimie détaillée (croix), ainsi que de troncatures de l'expansion de Karhunen-Loève contenant* 1 *terme (triangles bas),* 2 *termes (carrés) et* 3 *termes (triangles hauts).*

dèles différents ont été envisagés ici, qui sont du plus simple au plus complexe :

- Modélisation à l'aide d'un vecteur aléatoire gaussien centré réduit, dont les composantes sont donc indépendantes
- Modélisation à l'aide d'un vecteur aléatoire à composantes indépendantes, la loi de chaque composante étant donnée par la fonction de répartition empirique obtenue à partir des échantillons
- Modélisation de la densité à l'aide d'une méthode à noyau, obtenue à partir des échantillons

Il est important de noter que dans un cas où une seule variable aléatoire est retenue, les deux derniers points sont presque équivalents. Afin de ne pas se retrouver dans une telle situation où la première variable aléatoire joue un rôle prépondérant, les résultats présentés ici seront ceux concernant le terme source modélisé à l'aide d'une expansion de Karhunen-Loève de $\dot{\omega}_{Y_c}^{\log,\psi}$, c'est à dire sans modification de la variance du processus. Les exemples de nuages de points des figures 8.9 et 8.8 correspondent aux échantillons des variables aléatoires η_k obtenus avec cette modélisation du terme source. Bien que la figure 8.9 présentée précédemment suggère que des hypothèses trop faibles ne permettront pas d'échantillonner correctement ces nouvelles variables aléatoires, il convient de s'assurer qu'un mauvais échantillonnage impactera conséquemment les quantités d'intérêt que l'on souhaite reproduire. En effet, par construction de l'expansion de Karhunen-Loève, la première variable aléatoire η_1 a plus d'importance

que les autres, et on peut espérer qu'une bonne reproduction de celle-ci, même couplée à une mauvaise reproduction des variables aléatoires suivantes, tant au niveau de leurs lois que de leurs dépendances mutuelles, permettra une suffisamment bonne reproduction des statistiques des variables aléatoires d'intérêts. La figure 8.21 présente, pour les différentes conditions initiales, les diagrammes quantile-quantile du délai d'auto-allumage $\tau^{C=0.1}$ obtenus à l'aide d'une modélisation du terme source ne retenant qu'une troncature de trois termes de l'expansion de Karhunen-Loève, la figure 8.10 montrant que cela suffisait à une bonne reproduction du délai d'auto-allumage incertain. Les trois différentes modélisations pour les nouvelles variables aléatoires (τ_{KL}^{RAND}) sont comparées aux références correspondant à l'utilisation des échantillons des nouvelles variables aléatoires (τ_{KL}^{REF}), obtenues lors de la construction de l'expansion de Karhunen-Loève. La construction de la densité de probabilité jointe est réalisée à l'aide d'une méthode à noyau gaussien décrite dans le chapitre 5 et désigné sous l'acronyme *NOG* dans ce même chapitre.

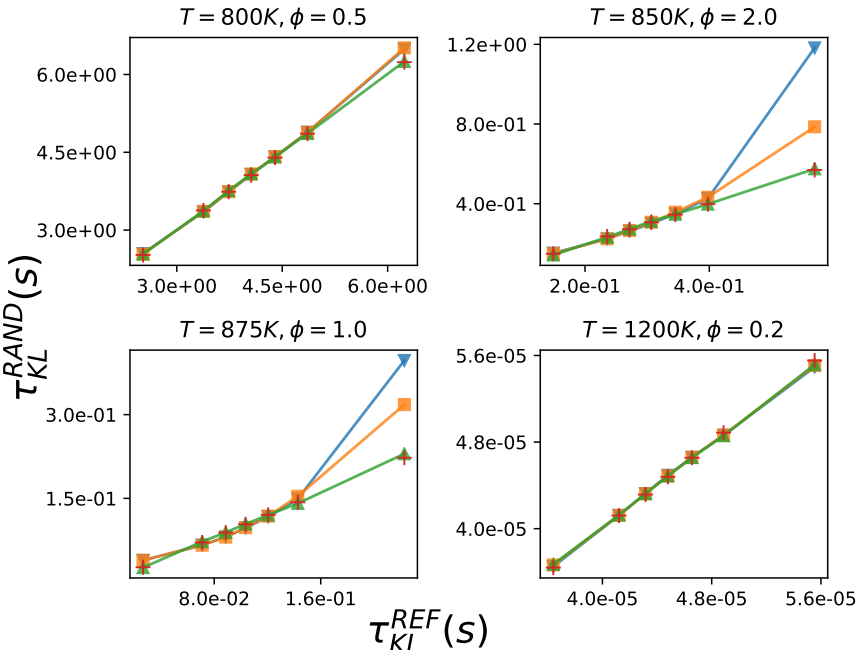

FIGURE 8.21 – *Diagrammes quantile-quantile du délai d'auto-allumage $\tau^{C=0.1}$ obtenu à l'aide d'une chimie tabulée impliquant un terme source incertain $\dot{\omega}_{Y_c}$ obtenu à l'aide d'une troncature à 3 termes de l'expansion de Karhunen-Loève avec les échantillons des nouvelles variables aléatoires η_k et avec les trois différentes modélisations proposées pour ces variables aléatoires η_k. Les nouvelles variables aléatoires sont modélisées par des gaussiennes indépendantes (triangle bas), des variables indépendantes avec densité marginale empirique (carrés) et des variables dépendantes obtenue par une densité jointe obtenue par une méthode à noyau (triangle haut), la référence correspondant aux croix. Les symboles sont placés aux quantiles $q_{0.01}$, $q_{0.15}$, $q_{0.29}$, $q_{0.43}$, $q_{0.57}$, $q_{0.71}$, $q_{0.85}$ et $q_{0.99}$.*

Les résultats présentés sur la figure 8.21 montrent que suivant la condition initiale considérée, une modélisation plus ou moins complexe des nouvelles variables aléatoires doit être envisagée. Pour la condition initiale à $T = 1200K$ et $\phi = 0.2$, les trois modélisations donnent des résultats équivalents qui sont en accord avec la référence. La figure 8.8 laissait entrevoir ce résultat, puisque les nuages de points ne présentaient pas de dépendances et les histogrammes des densités de probabilités marginales présentaient des allures de gaussiennes centrées réduites. La condition initiale à $T = 800K$ et $\phi = 0.5$ présente également une bonne reproduction pour l'ensemble des modélisations des nouvelles variables aléatoires, bien que la prise en compte des dépendances offrent une légère amélioration sur les deux modélisations où les dépendances entre variables aléatoires ne sont pas prises en compte. En revanche, les conditions initiales à $T = 850K$ et $T = 875K$ nécessitent clairement la prise en compte des dépendances pour la bonne reproduction du délai d'auto-allumage, spécialement pour les longs délais d'auto-allumage, correspondant aux quantiles $q_{0.85}$ et $q_{0.99}$, mais également pour les courts délais d'auto-allumage correspondant au quantile $q_{0.01}$ pour la condition initiale à $875K$. L'utilisation de gaussiennes indépendantes introduit de plus un écart plus important que l'utilisation des lois empiriques indépendantes, les lois marginales des différentes variables aléatoires s'éloignant de gaussiennes centrées réduites comme visible sur les histogrammes de la figure 8.9. La modélisation de la densité de probabilité jointe à l'aide de la méthode à noyau gaussien permet quant à elle de reproduire le délai d'auto-allumage incertain pour l'ensemble des conditions initiales.

Il est également intéressant de regarder l'impact du choix de la modélisation des nouvelles variables aléatoires η_k sur les statistiques du processus stochastique C, sa bonne reproduction étant l'objectif fixé. Sur la figure 8.22 sont représentées les moyennes temporelles obtenues à l'aide des différentes modélisations possibles pour les nouvelles variables aléatoires η_k. Les résultats sont en accord avec ceux obtenus pour le délai d'auto-allumage présentés sur la figure 8.21, à savoir que l'ensemble des modélisations parvient à reproduire correctement la moyenne temporelle de C pour les conditions initiales à $800K$ et $1200K$, alors que des différences sont observés pour les conditions initiales à $850K$ et $875K$ pour les modélisations considérant les variables aléatoires η_k comme indépendantes.

Une deuxième statistique d'importance est l'écart-type temporel du processus C, qui est montré sur la figure 8.23 pour l'ensemble des quatre conditions initiales et pour l'ensemble des modélisations des nouvelles variables aléatoires η_k. Les mêmes comportements que pour la moyenne et pour le délai d'auto-allumage sont encore une fois observés, l'indépendance des variables aléatoires ne permettant pas de reproduire correctement l'écart-type pour les conditions initiales à 850 K et 875 K, et une prise en compte de leurs dépendances mutuelles est nécessaire.

Une modélisation trop simpliste pour les nouvelles variables aléatoires η_k peut entraîner une mauvaise reproduction des statistiques du processus stochas-

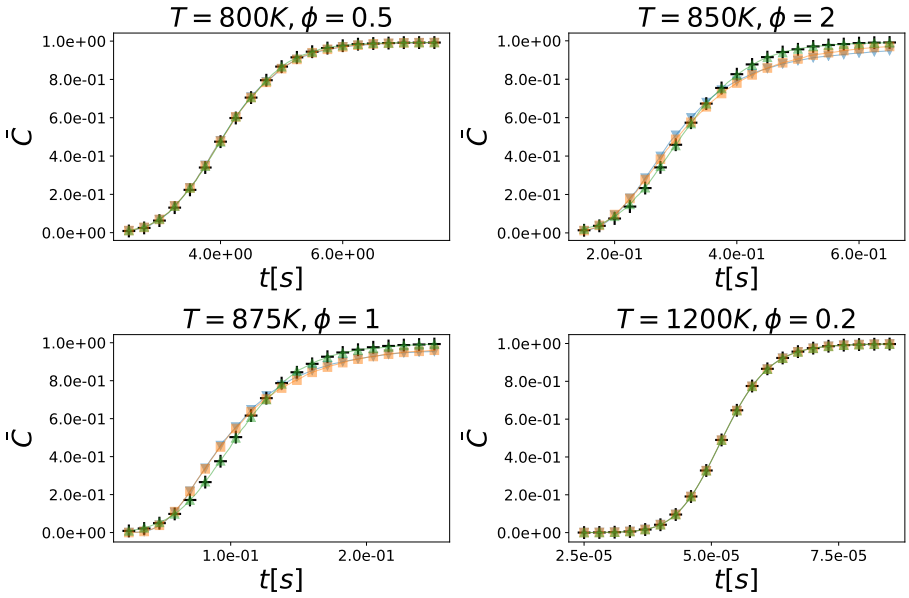

FIGURE 8.22 – *Moyennes temporelles du processus stochastique C pour les quatre conditions initiales, obtenues à l'aide d'une modélisation du terme source à l'aide d'une troncature à trois termes de l'expansion de Karhunen-Loève de $\dot{\omega}_{Y_c}^{\log,\psi}$, les variables aléatoires η_k étant obtenues à l'aide des échantillons (croix), de gaussiennes centrées réduites indépendantes (triangle bas), des marginales empiriques indépendantes (carrés) et d'une densité de probabilité jointe obtenue par méthode à noyau gaussien (triangle haut).*

tique C dans certains cas. En particulier, la prise en compte des dépendances entre ces différentes variables aléatoires peut être nécessaire, ce qui peut s'effectuer en construisant un estimateur de la densité de probabilité jointe de ces variables aléatoires à l'aide d'une méthode à noyau gaussien et des échantillons de ces variables aléatoires obtenus lors du calcul de l'expansion de Karhunen-Loève de $\dot{\omega}_{Y_c}^{\log,\psi}$. L'effort de modélisation des nouvelles variables aléatoires à entreprendre dépend bien entendu du cas considéré, ainsi que des exigences de précision fixées.

8.1.2 Extension du nombre de paramètres déterministes : température et richesse variables

L'ensemble de l'étude précédente se restreignait à la reproduction des incertitudes à travers le terme source $\dot{\omega}_{Y_c}$ où celui-ci ne dépendait que de la variable d'avancement c. En pratique, la tabulation de la cinétique chimique pour des configurations complexes ne se fait pas seulement selon la variable d'avancement c, mais implique d'autres variables de contrôle. Les processus stochastiques à considérer sont alors indexés par l'ensemble de ces variables de contrôles, ce qui complexifie le problème. Cette complexification était vi-

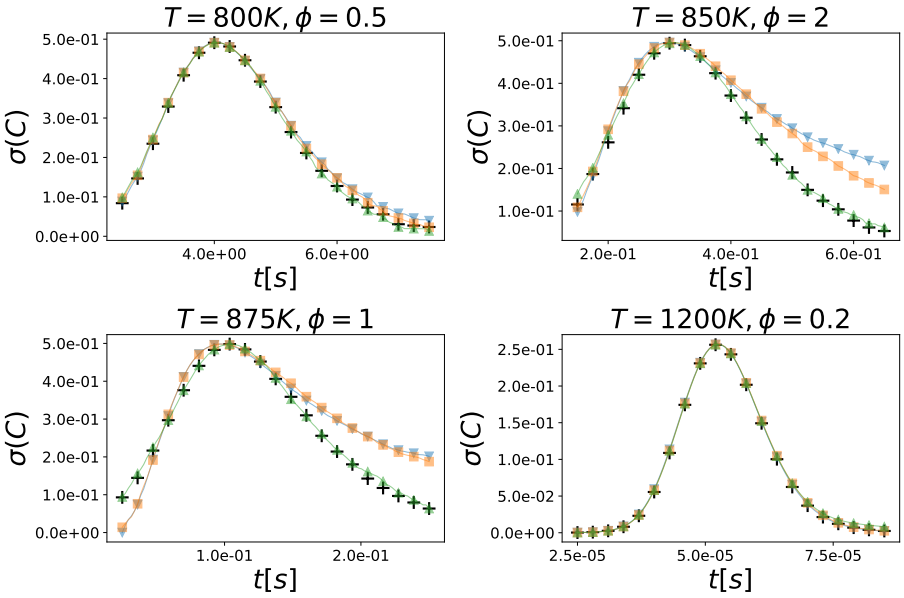

FIGURE 8.23 – *Écart-types temporels du processus stochastique C pour les quatre conditions initiales, obtenus à l'aide d'une modélisation du terme source à l'aide d'une troncature à trois termes de l'expansion de Karhunen-Loève de $\dot{\omega}_{Y_c}^{\log,\psi}$, les variables aléatoires η_k étant obtenues à l'aide des échantillons (croix), de gaussiennes centrées réduites indépendantes (triangle bas), des marginales empiriques indépendantes (carrés) et d'une densité de probabilité jointe obtenue par méthode à noyau gaussien (triangle haut).*

sible lors de l'analyse de sensibilité par le fait que les réactions importantes ne sont pas les mêmes suivant les conditions initiales. L'agrandissement de l'espace déterministe à une plage de température et de richesse implique que le processus stochastique $\dot{\omega}_{Y_c}$ va être influencé par un plus grand nombre de réactions, complexifiant celui-ci.

L'objectif est dans cette partie la reproduction des incertitudes pour le domaine de température et de richesse précédemment introduit pour l'analyse de sensibilité globale et la reproduction des incertitudes à l'aide de polynômes du chaos. L'indexation du processus stochastique n'est donc plus seulement faite par $c \in [0,1]$, mais par $(T,\phi,c) \in [800,1200] \times [0.2,2] \times [0,1]$ dans le cas considéré. La principale difficulté introduite par cette augmentation du nombre de paramètres indexant le processus stochastique vient de l'explosion du nombre de points présents dans la cubature utilisée pour la méthode de Nyström.

8.1.2.1 Cubature issue d'une tensorisation pleine

Afin de prendre en compte les nouvelles variables de contrôle, la méthode de Nyström impose l'utilisation d'une cubature qui est ici issue d'une tensorisation pleine de quadrature. Étant donnés les résultats de convergence obtenus

précédemment à température et richesse fixées, la quadrature utilisée dans la dimension de la variable d'avancement c est la seconde quadrature de Féjer impliquant 127 points. Le changement de variable suivant la variable d'avancement c est bien entendu toujours utilisé. Concernant la température et la richesse, il s'avère qu'il existe de plus grandes variations du terme source suivant la température que suivant la richesse dans la zone considérée, obligeant à plus de points de quadratures dans la dimension de la température que de la richesse. Le choix a donc été fait d'utiliser une quadrature de Clenshaw-Curtis pour la température et la richesse, avec 33 points de quadrature en température et 9 points de quadrature en richesse. Plusieurs complications se présentent par rapport au cas à température et richesse fixées.

La première provient du calcul des réalisations dans le but d'obtenir des statistiques, et notamment la matrice d'auto-covariance introduite par la méthode de Nyström. Un jeu de valeurs pour les paramètres incertains amène au calcul de $33 \times 9 = 297$ réacteurs adiabatiques à pression constantes, chacun prenant de l'ordre de $1s$ à résoudre sur les processeurs utilisés. En supposant une méthode de Quasi-Monte Carlo randomisée avec 10 séquences de Sobol de 1024 points chacune, cela revient à un temps de calcul de l'ordre de $10 \times 1024 \times 297 \approx 3000000s$, soit environ 30 jours de calcul sur un seul processeur. La parallélisation est donc nécessaire, d'autant plus que le cas étudié ici est le plus simple envisageable, à savoir un calcul $0D$ impliquant l'hydrogène qui est le carburant le plus simple. Un cas plus complexe mènerait à une explosion du temps de calcul où la parallélisation deviendrait encore plus indispensable.

Ensuite, le nombre total est donc de $37,719$ points dans la cubature. La taille du problème aux valeurs propres à résoudre est donc de $N = 37,719$, ce qui pose deux difficultés principales.

- La première provient de la place nécessaire pour stocker une telle matrice, qui est en double précision (8 octets par flottants) de plus de $10Go$, nécessitant donc une architecture de calculateur présentant suffisamment de mémoire vive.
- La seconde provient du temps de calcul nécessaire pour la résolution du système aux valeurs propres, la complexité algorithmique du problème au valeur propre étant en $O(N^3)$ [7] pour une matrice pleine, comme c'est le cas ici.

Le problème de mémoire peut difficilement être contourné, seule l'augmentation de la capacité des machines permettant d'envisager des problèmes plus grands. Les calculs sont réalisés sur la machine FUSION, possédant 64 Go de mémoire par nœud. La taille maximale d'une matrice que l'on souhaiterait juste stocker dans la mémoire d'un nœud est de $N = 92,681$, aucune place n'étant laissée pour réaliser des calculs. La valeur de N envisagée ici est donc proche de cette limite, et très peu de marge de manœuvre est possible de ce point de vue. Concernant le temps de calcul, il est possible de diviser celui-ci en parallélisant la résolution du problème aux valeurs propres. Pour cela, la librairie ScaLAPACK (Scalable LAPACK) [14], et plus particulièrement la routine PDSYEV

permettant la résolution d'un problème aux valeurs propres pour une matrice réelle symétrique, a ici été utilisée, permettant de diviser le temps de calcul par le nombre de processeurs utilisés.

Compte tenu du temps de calcul nécessaire, une étude de convergence à la fois au niveau du nombre de points de quadratures utilisé, et au niveau du nombre d'échantillons utilisé dans la méthode de Quasi-Monte Carlo n'a pas été réalisée, et l'assurance que les résultats sont suffisamment convergés se fait donc à posteriori, en vérifiant que les quantités d'intérêts peuvent bien être reproduites.

8.1.2.2 Utilisation directe du terme source

Dans un premier temps, le processus stochastique $\dot{\omega}_{Y_c}^{\log,\psi}$ est considéré, le changement de variable ψ étant conservé et ne dépendant pas de la température ni de la richesse, comme précédemment. L'expansion de Karhunen-Loève de $\dot{\omega}_{Y_c}^{\log,\psi}$ a été calculée à l'aide d'une méthode de Quasi-Monte Carlo randomisée impliquant 10 séquences de Sobol randomisées à l'aide d'une méthode de Full-Scrambling, chacune comportant 1024 points. Les sommes cumulées normalisées des premières valeurs propres sont présentées sur la figure 8.24. En comparant cette figure à la figure 8.7, on observe que la convergence de la somme des valeurs propres normalisées vers 1 est plus lente que la convergence pour chacune des conditions initiales prise séparément. La vitesse de convergence étant reliée à l'auto-corrélation du processus stochastique, la plus lente convergence provient du fait que désormais, des points éloignées en température et en richesse sont présents au sein du processus stochastique, introduisant moins d'auto-corrélation au sein du processus stochastique. Néanmoins, la convergence reste relativement rapide puisque 90% de la variance totale de $\dot{\omega}_{Y_c}^{\log,\psi}$ peut être représentée à l'aide de seulement 4 modes de son expansion de Karhunen-Loève.

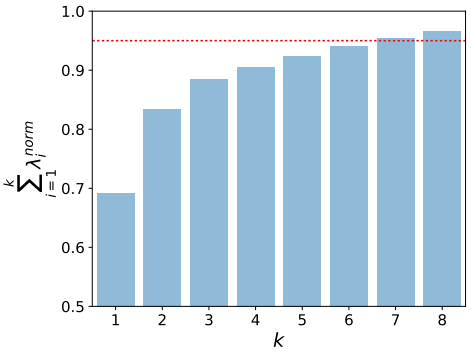

Figure 8.24 – *Sommes cumulées des premières valeurs propres normalisées de l'expansion de Karhunen-Loève du processus* $\dot{\omega}_{Y_c}^{log,\psi}$. *La ligne pointillée correspond à une valeur de* 95%.

Les histogrammes et des nuages de points des premières nouvelles variables aléatoires introduites par l'expansion de Karhunen-Loève de $\dot{\omega}_{Y_c}^{\log,\psi}$ sont présentées sur la figure 8.25. Les histogrammes montrent que ces variables aléatoires ne se rapprochent pas toutes de gaussiennes centrées réduites, et les nuages de points présentent des dépendances deux à deux de ces variables aléatoires. Comme il sera montré ultérieurement, une modélisation trop simpliste de ces variables aléatoires entraînera une reproduction approximative des quantités d'intérêts.

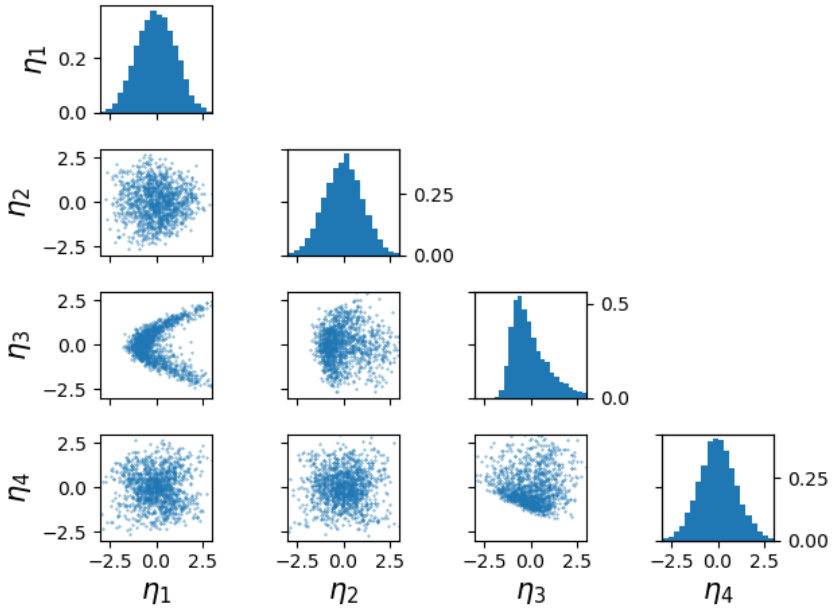

FIGURE 8.25 – *Sur la diagonale sont présents les histogrammes normalisés des quatre variables aléatoires η_1, η_2, η_3 et η_4 introduites par l'expansion de Karhunen-Loève de $\dot{\omega}_{Y_c}^{log,\psi}$. Sous la diagonale sont représentés les nuages de points de ces variables aléatoires deux à deux obtenus à partir des échantillons de Quasi-Monte Carlo ayant été utilisés pour estimer la matrice d'auto-covariance.*

Le test de la capacité de reproduire des quantités d'intérêts comme le délai d'auto-allumage à l'aide de la modélisation du terme source $\dot{\omega}_{Y_c}$ par l'expansion de Karhunen-Loève précédemment construite est encore une fois réalisé pour les quatre conditions initiales précédemment choisies, les températures et les richesses de ces points ne faisant pas partie de l'ensemble des points de cubature. La reconstruction du processus stochastique en n'importe quel point du domaine est réalisée via une interpolation impliquant des splines cubiques naturelles, l'interpolation de la méthode de Nyström n'étant pas possible du fait que l'auto-covariance n'est pas connu en tous les points. Les diagrammes quantile-quantile du délai d'auto-allumage $\tau^{C=0.1}$ obtenus par une chimie détaillée incertaine et par une chimie tabulée utilisant les troncatures de l'expan-

sion de Karhunen-Loève de $\dot{\omega}_{Y_c}^{\log,\psi}$ contenant 1 à 5 termes sont présentés sur la figure 8.26, les échantillons des nouvelles variables aléatoires étant utilisés afin de ne pas introduire une erreur liée à leur modélisation. Pour l'ensemble des conditions initiales, l'augmentation du nombre de termes dans la troncature permet une meilleure reproduction du délai d'auto-allumage incertain, bien que la différence entre la troncature à 4 termes et à 5 termes soit peu perceptible. Cette reproduction du délai d'auto-allumage est cependant plus ou moins bonne suivant la condition initiale considérée. Ainsi, avec une température initiale de $800K$ et une richesse de 0.5, une troncature à 5 termes ne reproduit pas correctement les courts et longs délai d'auto-allumage, les sous estimant. Ce phénomène de non reproduction des délais extrêmes est également présent dans une moindre mesure pour les conditions initiales de température initiale $850K$ et de richesse 2.0 et de température initiale $1200K$ et de richesse 1.0, alors que les délais d'auto-allumage intermédiaires ont tendance à être mieux reproduits, et pour des ordres de troncatures plus faibles.

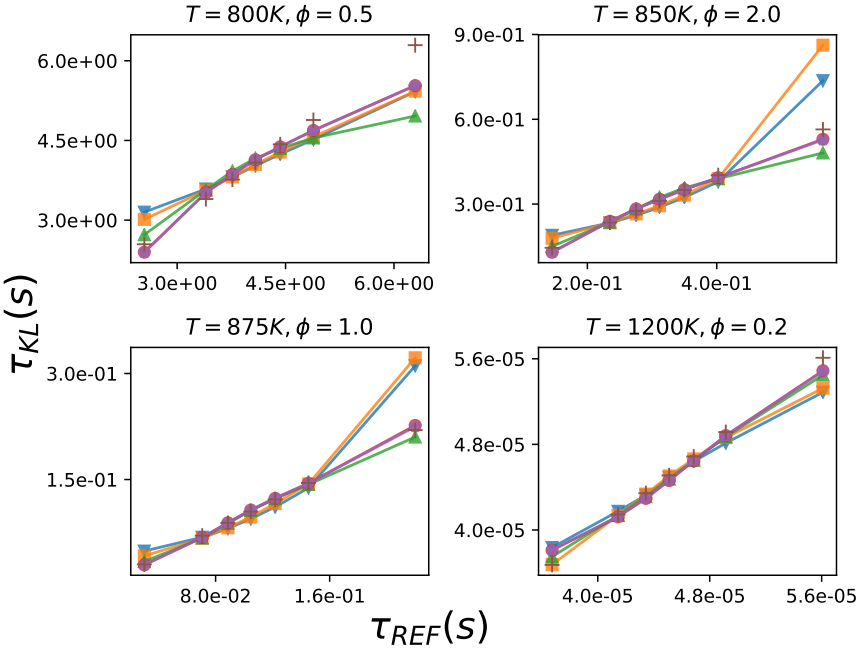

FIGURE 8.26 – *Diagrammes quantile-quantile du délai d'auto-allumage $\tau^{C=0.1}$ pour les différentes conditions initiales, calculés à l'aide de la chimie détaillée incertaine et à l'aide d'une expansion de Karhunen-Loève de $\dot{\omega}_{Y_c}^{log,\psi}$, avec différentes troncatures de cette expansion : 1 paramètre incertain (triangles bas), 2 paramètres incertains (carrés), 3 paramètres incertains (triangles hauts), 4 paramètres incertains (ronds) et 5 paramètres incertains (pentagones).*

Sur les figures 8.27 et 8.28 sont représentés respectivement les moyennes temporelles ainsi que les écart-types temporels du processus stochastique C obtenue avec une chimie détaillée et avec une chimie tabulée utilisant des tron-

catures contenant 1 à 5 termes de l'expansion de Karhunen-Loève de $\dot{\omega}_{Y_c}^{log,\psi}$.

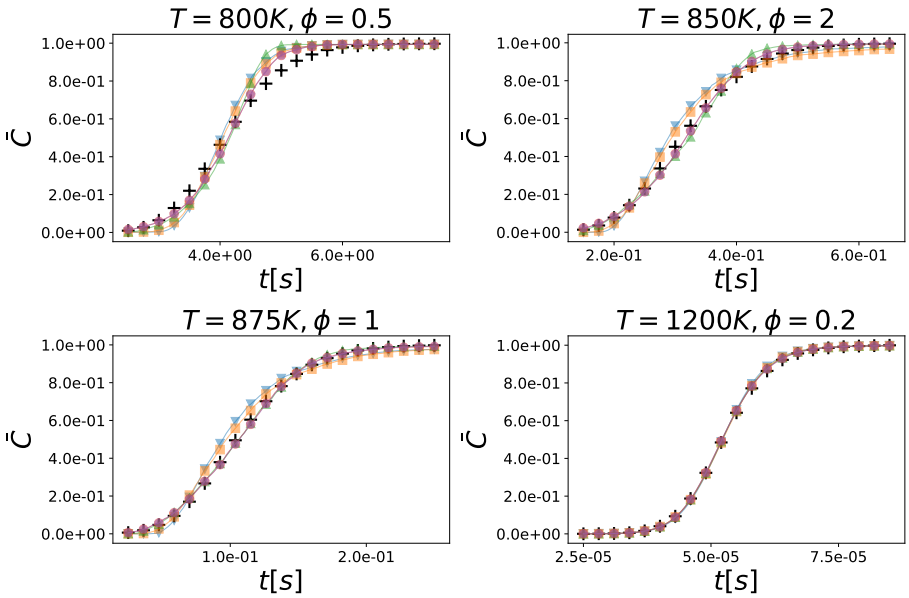

FIGURE 8.27 – *Moyennes temporelles du processus stochastique C pour les quatre conditions initiales choisies, obtenues à l'aide de la chimie détaillée (croix), ainsi que de troncatures de l'expansion de Karhunen-Loève de $\dot{\omega}_{Y_c}^{\log,\psi}$ à 1 terme (triangles bas), à 2 termes (carrés), à 3 termes (triangles hauts), à 4 termes (ronds) et à 5 termes (pentagones).*

Les résultats de ces deux figures sont en accord avec ceux sur le délai d'auto-allumage de la figure 8.26 pour l'ensemble des conditions initiales. En effet, tout comme pour le délai d'auto-allumage, il n'existe pas de différences visibles sur les courbes entre les troncatures à 4 et 5 termes pour la moyenne et l'écart-type temporels de C. L'utilisation de ces troncatures à 4 et 5 termes permet de reproduire correctement la moyenne et l'écart-type temporels pour l'ensemble des conditions initiales exceptées celle de température initiale $800K$ et de richesse 0.5, où la moyenne atteint trop rapidement sa valeur limite de 1, et par voie de conséquence l'écart-type se retrouve à 0 trop rapidement. L'augmentation du nombre de termes considérés tendant bien entendu à améliorer la reproduction de la moyenne et de l'écart-type temporels de C, il s'avère que 5 termes dans la troncature ne permettent pas une reproduction correcte des quantités d'intérêts pour l'ensemble des conditions initiales investiguées.

Dans les cas à température et richesse initiales fixées, une amélioration était possible en considérant l'expansion de Karhunen-Loève d'un autre processus stochastique que $\dot{\omega}_{Y_c}^{\log,\psi}$, obtenu par une modification simple de celui-ci. La prochaine section s'intéresse à une telle modification dans le cas où le domaine des conditions thermodynamiques initiales n'est pas réduit à un point.

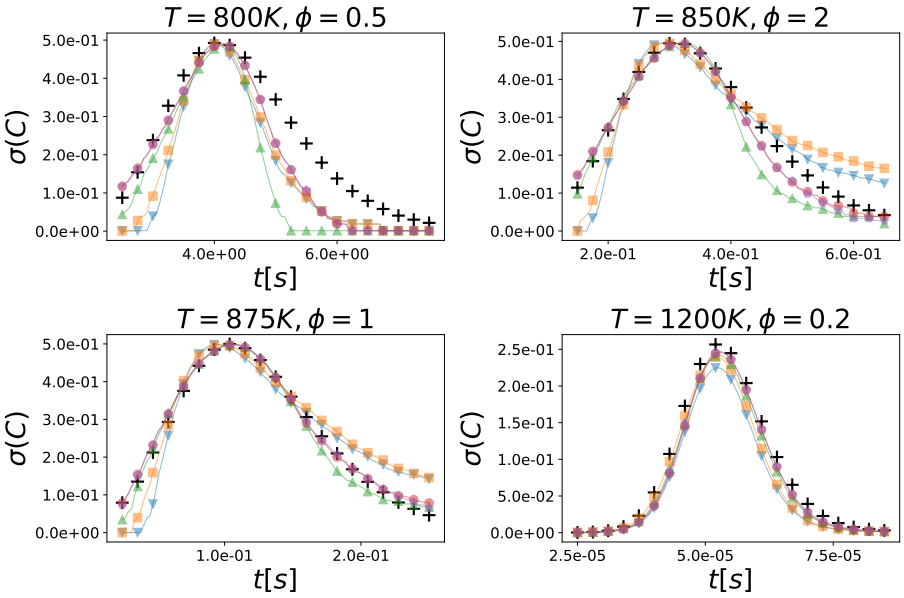

FIGURE 8.28 – *Écart-type temporels du processus stochastique C pour les quatre conditions initiales choisies, obtenus à l'aide de la chimie détaillée (croix), ainsi que de troncatures de l'expansion de Karhunen-Loève à 1 terme (triangles bas), à 2 termes (carrés), à 3 termes (triangles hauts), à 4 termes (ronds) et à 5 termes (pentagones).*

8.1.2.3 Modification du terme source

Tout comme pour le cas où les conditions initiales étaient réduites à un point, il est possible dans le cas où le domaine des conditions initiales n'est pas réduit à un point de multiplier le processus stochastique $\dot{\omega}_{Y_c}^{log,\psi}$ par une fonction f_{scal} dépendant de la variable d'avancement normalisée c mais également des autres paramètres d'indexation du processus stochastique, que sont la température T et la richesse ϕ dans le cas présent. Une difficulté supplémentaire est cependant présente, provenant du fait que la fonction f_{scal} dépend désormais également de la température et de la richesse. Dans un premier temps, on définit une fonction $f_{scal}^{T,\phi}$ pour chaque valeur de T et de ϕ comme décrit par la relation (8.16). Une fois définie $f_{scal}^{T,\phi}$ pour chacune des valeurs de température et de richesse, la fonction finale f_{scal} est construite en normalisant $f_{scal}^{T,\phi}$ en utilisant (8.17), afin d'assurer que chacun des points de température et de richesse fixées aient la même importance.

$$f_{scal}(T,\phi,c) = \frac{f_{scal}^{T,\phi}(c)}{\int_0^1 f_{scal}^{T,\phi}(c)dc} \tag{8.17}$$

Le nouveau processus stochastique $\dot{\omega}_{Y_c}^{\log,\psi,opt}$ est finalement défini :

$$\dot{\omega}_{Y_c}^{\log,\psi,opt}(T,\phi,c) = \dot{\omega}_{Y_c}^{\log,\psi}(T,\phi,c) f_{scal}(T,\phi,c) \tag{8.18}$$

Le calcul de l'expansion de Karhunen-Loève de $\dot{\omega}_{Y_c}^{\log,\psi,opt}$ a été réalisé dans les mêmes conditions que pour $\dot{\omega}_{Y_c}^{\log,\psi}$, à savoir la même quadrature et le même nombre d'échantillons dans la méthode de Quasi-Monte Carlo. Les sommes cumulées normalisées des valeurs propres de l'expansion de Karhunen-Loève pour ce processus stochastique sont présentées sur la figure 8.29. La convergence y est plus rapide que pour le processus stochastique $\dot{\omega}_{Y_c}^{\log,\psi}$, seulement 4 valeurs propres permettant de reproduire 95% de la variance totale du processus stochastique.

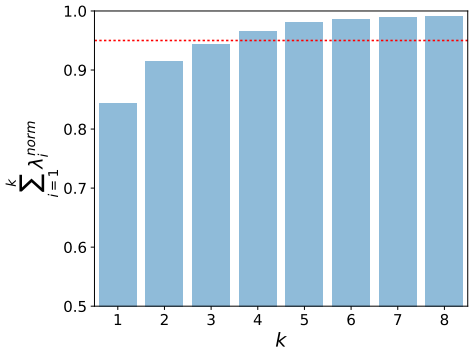

FIGURE 8.29 – *Sommes cumulées des premières valeurs propres normalisées de l'expansion de Karhunen-Loève du processus $\dot{\omega}_{Y_c}^{log,\psi,opt}$. La ligne pointillée correspond à une valeur de* 95%.

Les histogrammes des premières nouvelles variables aléatoires ainsi que des nuages de points de ces variables deux à deux sont présentés sur la figure 8.30. Encore une fois, des dépendances apparaissent entre les différentes variables aléatoires, et aucune d'elles ne présente un profil de gaussienne centrée réduite.

Sur la figure 8.31 sont représentés les diagrammes quantile-quantile pour les délai d'auto-allumage $\tau^{C=0.1}$ calculés à l'aide d'une chimie détaillée et à l'aide d'une chimie tabulée utilisant les troncatures de 1 à 5 termes de l'expansion de Karhunen-Loève de $\dot{\omega}_{Y_c}^{log,\psi,opt}$, les échantillons des nouvelles variables aléatoires η_k obtenus lors de la construction de l'expansion de Karhunen-Loève ayant été utilisés afin de ne pas introduire une source d'erreur supplémentaire. Avec 5 termes dans la troncature, la reproduction du délai d'auto-allumage $\tau^{C=0.1}$ est meilleure que celle obtenue en considérant l'expansion de Karhunen-Loève de $\dot{\omega}_{Y_c}^{log,\psi}$, présentant une légère sous estimation des délais d'auto-allumage longs pour les conditions initiales de température initiale $800K$ et $1200K$ et de richesse 0.5 et 0.2 respectivement. Pour l'ensemble des conditions initiales, considérer l'expansion de Karhunen-Loève du processus stochas-

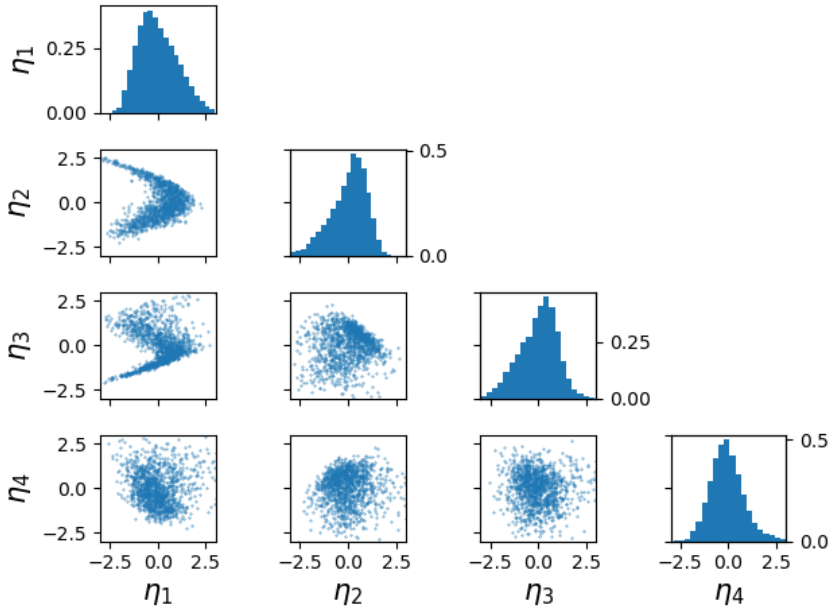

FIGURE 8.30 – *Sur la diagonale sont présents les histogrammes normalisés des quatre variables aléatoires η_1, η_2, η_3 et η_4 introduites par l'expansion de Karhunen-Loève de $\dot{\omega}_{Y_c}^{log,\psi,opt}$. Sous la diagonale sont représentés les nuages de points de ces variables aléatoires deux à deux obtenus à partir des échantillons de Quasi-Monte Carlo ayant été utilisés pour estimer la matrice d'auto-covariance.*

tique $\dot{\omega}_{Y_c}^{log,\psi,opt}$ plutôt que celle de $\dot{\omega}_{Y_c}^{log,\psi}$ semble offrir une légère amélioration visible par l'écartement moins important des courbes à la situation idéale, mais cette amélioration est bien moins marquée que dans le cas où le domaine des conditions initiales était réduit à un point.

Cette amélioration est plus visible en regardant les moyennes et écart-types temporels du processus stochastique C, visible respectivement sur les figures 8.32 et 8.33.

Cette fois-ci en utilisant 5 termes dans la troncature de l'expansion de Karhunen-Loève, la moyenne et l'écart-type temporels pour l'ensemble des conditions initiales sont reproduits correctement, notamment pour la condition initiale de température $800K$ et de richesse 0.5, où la moyenne temporelle n'a plus la tendance à atteindre trop tôt sa valeur de 1, ce qui permet la bonne reproduction de l'écart-type temporel. Le point de température initiale $1200K$ et de richesse 0.2 présente cependant un écart-type légèrement sous estimé, surtout pour les temps avancés, qui peut être mis en relation avec la sous estimation des délais d'auto-allumage longs que l'on peut observer sur la figure 8.26.

L'augmentation du nombre de termes dans la troncature de l'expansion de Karhunen-Loève permet encore une fois d'améliorer la reproduction des

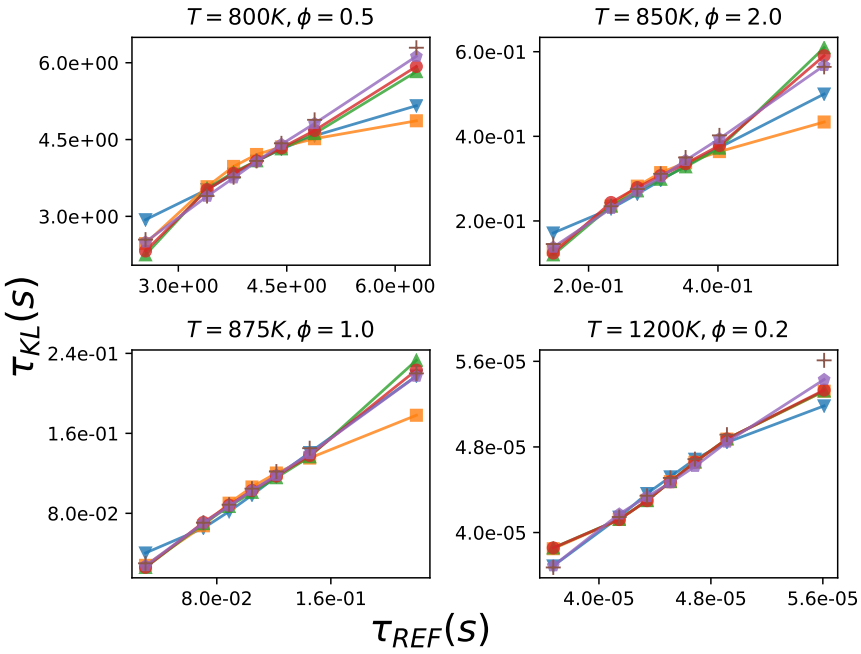

FIGURE 8.31 – *Diagramme quantile-quantile du délai d'auto-allumage $\tau^{C=0.1}$ pour les différentes conditions initiales, calculés à l'aide de la chimie détaillée incertaine et à l'aide d'une expansion de Karhunen-Loève de $\dot{\omega}_{Y_c}^{log,\psi,opt}$, avec différentes troncatures de cette expansion : 1 paramètre incertain (triangles bas), 2 paramètres incertains (carrés), 3 paramètres incertains (triangles hauts), 4 paramètres incertains (ronds) et 5 paramètres incertains (pentagones).*

quantités d'intérêts. L'utilisation de 3 termes permet déjà une reproduction des tendances des moyennes et des écarts-type temporels pour l'ensemble des conditions initiales.

La dernière étape pour l'utilisation d'une telle modélisation du terme source incertain $\dot{\omega}_{Y_c}$ est la modélisation des nouvelles variables aléatoires η_k introduites par l'expansion de Karhunen-Loève de $\dot{\omega}_{Y_c}^{\log,\psi,opt}$. La figure 8.30 laisse entrevoir que des dépendances existent entre celles-ci. Les trois modélisations pour les nouvelles variables aléatoires η_k présentées précédemment ont été comparées sur la figure 8.34 pour le délai d'auto-allumage, à savoir des gaussiennes centrées réduites indépendantes, des variables indépendantes dont la loi est obtenue empiriquement à partir des échantillons de chacune des variables aléatoires, et enfin une modélisation à l'aide d'une méthode à noyau gaussien de la densité de probabilité jointe. Encore une fois, les dépendances entre les variables aléatoires ne peuvent pas être ignorées pour obtenir une bonne reproduction des quantités d'intérêts pour l'ensemble des conditions initiales, et la densité de probabilité jointe estimée à l'aide des échantillons des variables aléatoires et de la méthode à noyau gaussien assure une bonne reproduction de la loi du délai d'auto-allumage.

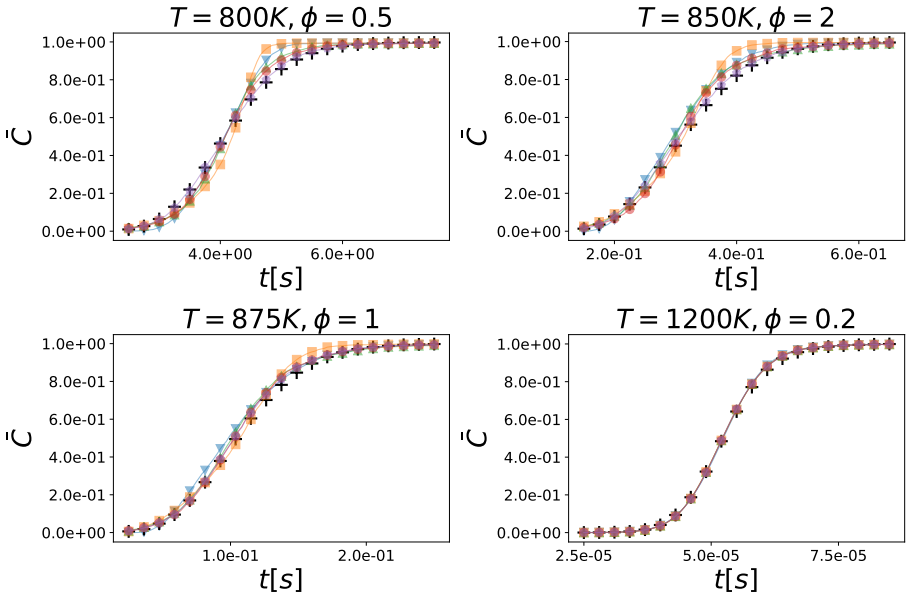

FIGURE 8.32 – *Moyennes temporelles du processus stochastique C pour les quatre conditions initiales choisies, obtenues à l'aide de la chimie détaillée (croix), ainsi que de troncatures de l'expansion de Karhunen-Loève à 1 terme (triangles bas), à 2 termes (carrés), à 3 termes (triangles hauts), à 4 termes (ronds) et à 5 termes (pentagones).*

8.1.3 Critique de la modélisation de $\dot{\omega}_{Y_c}$ par une expansion de Karhunen-Loève.

La modélisation présentée dans cette section pour le terme source incertain $\dot{\omega}_{Y_c}$ introduit de nouvelles variables aléatoires, et n'utilise donc pas les variables aléatoires paramétrant initialement le système. Les nouvelles variables aléatoires construites sont un mélange des variables aléatoires initiales, comme le montre les indices de Sobol ayant été calculés présentés sur la figure 8.12. Même si plusieurs variables aléatoires initiales devaient être prise en compte pour la reproduction d'une quantité d'intérêt, cette quantité d'intérêt peut être reproduite correctement avec une seule de ces nouvelles variables aléatoires mélangeant ces variables aléatoires initiales, comme la comparaison des figures 7.6 et 8.31 le montrent. Le nombre de nouvelles variables aléatoires à conserver peut ainsi être plus faible que le nombre de variables aléatoires initiales ayant un impact significatif sur les quantités d'intérêt à reproduire. Pour des situations impliquant un carburant plus complexe impliquant potentiellement un plus grand nombre de réactions impactant les quantités d'intérêt par exemple, on peut espérer que cette méthode puisse reproduire suffisamment bien les quantités d'intérêts même avec un nombre faible de variables aléatoires prises en compte. D'autant plus qu'il est possible de transformer simplement le processus stochastique dont on souhaite utiliser l'expansion de Karhunen-Loève

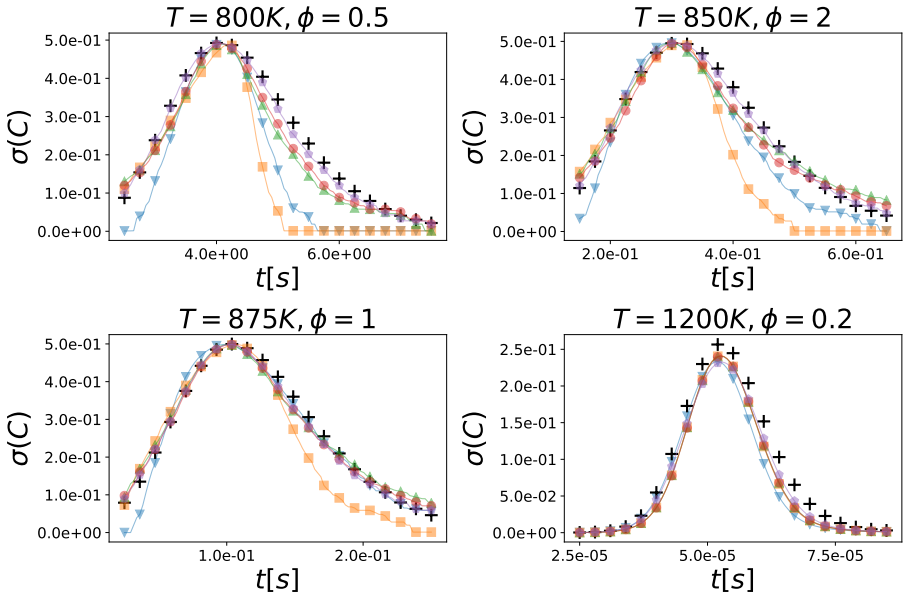

FIGURE 8.33 – *Écarts-type temporels du processus stochastique C pour les quatre conditions initiales choisies, obtenus à l'aide de la chimie détaillée (croix), ainsi que de troncatures de l'expansion de Karhunen-Loève à 1 terme (triangles bas), à 2 termes (carrés), à 3 termes (triangles hauts), à 4 termes (ronds) et à 5 termes (pentagones).*

afin de reproduire plus efficacement les quantités d'intérêt comme cela a été fait ici pour la reproduction du délai d'auto-allumage.

Il existe cependant des limitations à l'approche présentée. Tout d'abord, le processus stochastique dont on doit calculer l'expansion de Karhunen-Loève impacte fortement la reproduction de la quantité d'intérêt. Une première modification a été faite en réalisant un changement de variable en la variable d'avancement, qui était nécessaire pour avoir une bonne résolution du terme source pour les faibles valeurs de c afin d'être capable d'intégrer l'équation différentielle (7.2). La seconde modification provenait de l'approche proposée ici pour améliorer la reproduction du délai d'auto-allumage en apportant une modification simple au processus stochastique duquel on calcule l'expansion de Karhunen-Loève, mais la définition de cette modification nécessite une analyse du problème, et celle-ci peut nécessiter un choix arbitraire comme il a été fait ici en définissant une valeur c_{cut} afin d'éviter la divergence de la normalisation proposée en la valeur $c = 1$. L'autre limitation provient de la taille du système matriciel à résoudre qui est limitée par les moyens de calculs à disposition, en particulier en terme de mémoire vive. De ce fait, la dimension des tables chimiques considérées ne peut guère être supérieure à 2 voire 3 actuellement. En effet, en considérant un nombre faible de 20 points par dimension, le nombre total de points présents dans la table est alors de $160,000$, nombre trop important pour espérer pouvoir résoudre le système matriciel sur les machines actuelles.

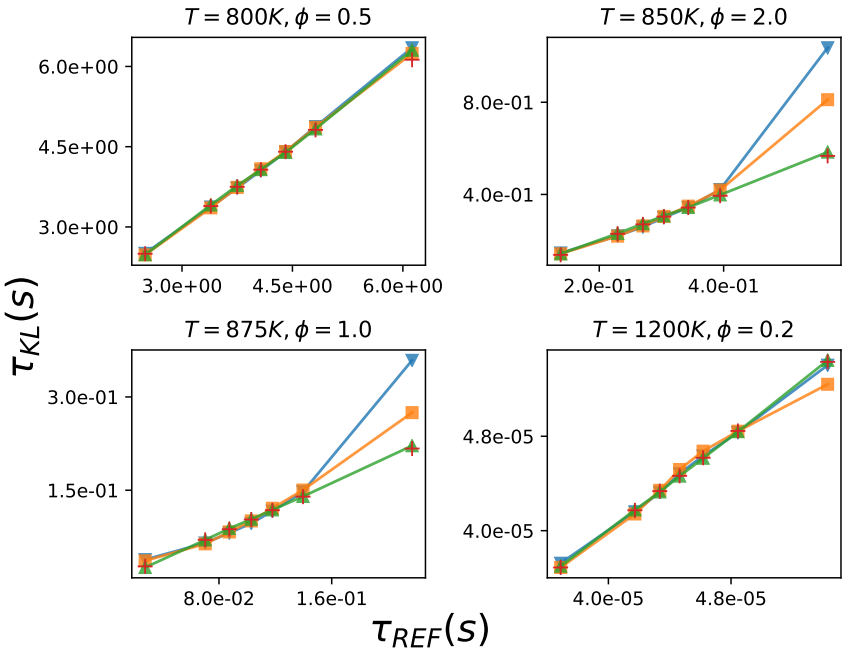

FIGURE 8.34 – *Diagrammes quantile-quantile du délai d'auto-allumage $\tau^{C=0.1}$ obtenus à l'aide d'une chimie tabulée impliquant un terme source incertain $\dot{\omega}_{Y_c}$ obtenu à l'aide d'une troncature de 3 termes de l'expansion de Karhunen-Loève avec les échantillons des nouvelles variables aléatoires η_k et avec les trois différentes modélisations proposées pour ces variables aléatoires η_k. Les nouvelles variables aléatoires sont modélisées par des gaussiennes indépendantes (triangle bas), des variables indépendantes avec densité marginale empirique (carrés) et des variables dépendantes obtenue par une densité jointe empirique (triangle haut), la référence correspondant aux croix. Les symboles sont placés aux quantiles $q_{0.01}$, $q_{0.15}$, $q_{0.29}$, $q_{0.43}$, $q_{0.57}$, $q_{0.71}$, $q_{0.85}$ et $q_{0.99}$.*

La considération de 20 points par dimension utilisée ici est de surcroît faible, certaines dimensions nécessitant au moins une centaine de points comme dans le cas présent pour la dimension correspondant à la variable d'avancement c. La construction de tensorisation creuse pour la construction de la table chimique a été envisagée afin de réduire le nombre de points total dans la table, mais la reconstruction par interpolation du terme source tabulée ne semblait pas être de qualité suffisante pour le problème étudié à moins d'avoir un nombre de points du même ordre de grandeur que le nombre de points avec la tensorisation pleine, ce qui faisait perdre tout intérêt à la méthode.

L'ensemble de ces limitations a conduit à envisager une dernière méthode. Contrairement aux méthodes présentées jusqu'alors qui se servait du délai d'auto-allumage afin de sélectionner les variables aléatoires, ou afin de définir une modification du processus stochastique dont on calculait l'expansion de Karhunen-Loève, la méthode suivante ne s'appuie que sur des informations apportées par le processus stochastique C que l'on souhaite reproduire.

8.2 Utilisation d'une expansion de Karhunen-Loève du processus stochastique C.

Dans le chapitre 7 ainsi que la section précédente, le délai d'auto-allumage jouait un rôle majeur, que ce soit pour la sélection des variables aléatoires initiales à conserver comme dans le cas de la PCE, ou dans la définition d'une modification du processus stochastique $\dot{\omega}_{Y_c}$ duquel était obtenue l'expansion de Karhunen-Loève permettant de le modéliser. Dans les deux cas, il s'agissait d'exprimer le terme source $\dot{\omega}_{Y_c}$ comme une fonction des variables aléatoires identifiées ou construite afin d'avoir une bonne reproduction du délai d'auto-allumage incertain $\tau^{C=0.1}$. Or, le chapitre 6 donnait pour objectif la bonne reproduction du processus stochastique C, dont la bonne reproduction implique bien entendu la bonne reproduction du délai d'auto-allumage $\tau^{C=0.1}$, mais dont la bonne reproduction n'est pas nécessairement assurée par la bonne reproduction du délai d'auto-allumage $\tau^{C=0.1}$.

L'idée est ici de s'intéresser directement au processus stochastique C, que l'on se doit de reproduire au mieux, plutôt qu'au délai d'auto-allumage $\tau^{C=0.1}$ pour l'obtention d'un jeu de variables aléatoires à même de caractériser l'incertitude présente dans ce processus stochastique.

8.2.1 Construction de nouvelles variables aléatoires

Compte tenu de l'équation différentielle (7.4) régissant le comportement du système étudié, l'objectif est toujours de représenter le terme source incertain $\dot{\omega}_{Y_c}$ comme une fonction d'un vecteur aléatoire ayant un nombre de composantes le plus petit possible. Précédemment, ce vecteur aléatoire a été construit en ne gardant qu'une partie des composantes du vecteur aléatoire paramétrant le système, et également en considérant le vecteur aléatoire comportant les premières variables aléatoires de l'expansion de Karhunen-Loève d'une modification du processus stochastique $\dot{\omega}_{Y_c}$. Cette idée de l'expansion de Karhunen-Loève peut être reprise pour s'appliquer directement au processus stochastique que l'on souhaite reproduire au mieux, à savoir C. En fait, il n'est pas possible directement de construire une telle expansion de Karhunen-Loève, ce processus possédant une indexation par l'ensemble des instants $t \in [0, \infty[$ qui n'est pas borné. Il est donc nécessaire de s'arranger pour changer l'indexation de ce processus stochastique afin qu'elle soit bornée. Plusieurs solutions sont envisageables pour répondre à ce problème, parmi lesquelles :

- Restreindre l'intervalle de temps afin de le borner, en fixant une valeur de temps maximum par exemple
- Utiliser un changement de variable envoyant l'intervalle temporel $[0, +\infty[$ sur un segment, qui sera de fait borné

Dans le premier cas, le processus stochastique est toujours indexé par le temps mais seule une partie des instants est pris en compte, alors que dans le second cas tous les instants sont pris en compte, mais l'indexation du processus

stochastique n'est plus temporelle ce qui se traduira en fait par des instants ayant une importance plus grande que d'autres.

Le choix est fait ici d'utiliser la seconde méthode, basée sur un changement de variable. Le changement de variable utilisé envoie l'intervalle $[0, +\infty[$ sur le segment $[0, 1]$ par le biais de la fonction de répartition d'une loi log-normale dont la moyenne et la variance sont prises égale à la moyenne et à la variance du délai d'auto-allumage $\tau^{C=0.5}$. Le nouvel index u du processus stochastique est ainsi donné par la relation (8.19).

$$u = \frac{1}{2}\left(1 + \text{erf}\left[\frac{\ln(t) - \mu}{\sigma\sqrt{2}}\right]\right) \tag{8.19}$$

Dans l'expression (8.19), les paramètres μ et σ sont donnés respectivement par les expressions (8.20) et (8.21).

$$\mu = \ln\left(E\left[\tau^{C=0.5}\right]\right) - \frac{1}{2}\ln\left(1 + \frac{Var\left[\tau^{C=0.5}\right]}{E\left[\tau^{C=0.5}\right]^2}\right) \tag{8.20}$$

$$\sigma = \sqrt{\ln\left(1 + \frac{Var\left[\tau^{C=0.5}\right]}{E\left[\tau^{C=0.5}\right]^2}\right)} \tag{8.21}$$

Avec ce changement d'indexation, le processus stochastique désormais étudié n'est plus celui indexé par le temps t, mais par la variable u et n'est donc plus le même, mais la notation C pour celui-ci sera gardée dans la suite. L'existence de l'expansion de Karhunen-Loève de ce nouveau processus stochastique C est donc assurée et il est possible de la calculer. Les résultats présentés ici concernent directement le processus stochastique indexé par le jeu de variables (T, ϕ, u) appartenant au domaine $[800K, 1200K] \times [0.2, 2] \times [0, 1]$, le domaine de température et de richesse étudié étant le même que précédemment.

Le calcul de l'expansion de Karhunen-Loève du processus C a été réalisée à l'aide d'une tensorisation creuse impliquant la quadrature de Clenshaw-Curtis pour la température et la richesse, et la seconde quadrature de Fejér pour la variable u, comportant au total 5455 points de cubatures. L'estimation de la matrice de covariance a quant à elle été réalisée à l'aide d'une méthode de Quasi-Monte Carlo Randomisées impliquant 10 séquences de Sobol de 4096 points chacune et randomisées à l'aide d'une méthode de Full-Scrambling.

Les sommes cumulées normalisées des valeurs propres de la décomposition de Karhunen-Loève du processus stochastique C sont présentées sur la figure 8.35. La convergence de la somme vers 1 n'est pas aussi rapide que pour les processus stochastiques présentés précédemment, et atteindre 95% de variance totale reproduite nécessite la prise en compte de 24 modes de l'expansion. Ce-

pendant, considérer seulement 3 modes permet la reproduction de presque 80% de la variance totale du processus stochastique, les premières valeurs propres sont donc tout de même prépondérantes.

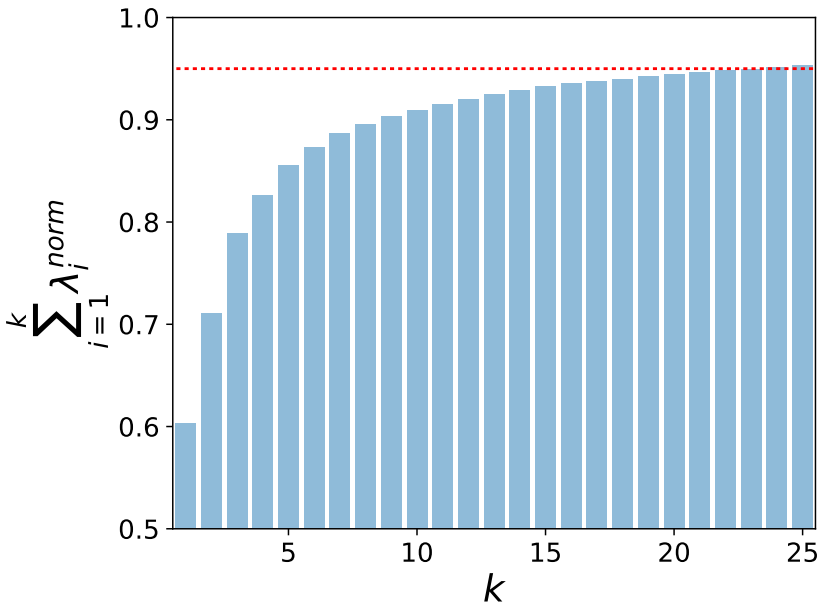

FIGURE 8.35 – *Sommes cumulées normalisées des* 25 *premières valeurs propres de la décomposition de Karhunen-Loève du processus stochastique* C*. La ligne pointillée correspond à* 95%*.*

L'objectif de l'expansion de Karhunen-Loève est ici l'obtention d'un jeu de variables aléatoires dont dépend significativement le processus stochastique C et à partir desquelles exprimer le terme source incertain $\dot{\omega}_{Y_c}$, en considérant justement les premières variables aléatoires de cette expansion de Karhunen-Loève. Des histogrammes ainsi que des nuages de points deux à deux des 4 premières variables aléatoires η_k obtenues via l'expansion de Karhunen-Loève de C sont présentés sur la figure 8.36. Les variables aléatoires η_k n'ont clairement pas un profil gaussien, et bien que décorrélées par construction, il existe de fortes dépendances entre elles.

L'objet de la section suivante est l'utilisation de ces variables aléatoires présentant de fortes dépendances et présentant des lois variées pour modéliser le terme source incertain $\dot{\omega}_{Y_c}$.

8.2.2 Expansion en Polynômes du Chaos du terme source $\dot{\omega}_c$

Une expansion en Polynômes du Chaos a été utilisée précédemment pour modéliser le terme source incertain $\dot{\omega}_{Y_c}$ avec les variables aléatoires initiales, considérées indépendantes. Les variables aléatoires construites à l'aide de l'ex-

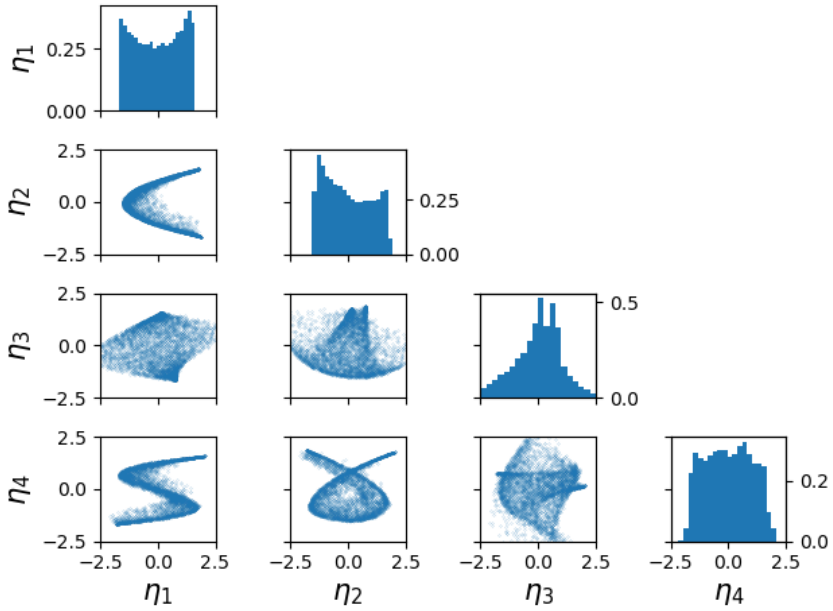

FIGURE 8.36 – *Sur la diagonale sont présents les histogrammes normalisés des quatre variables aléatoires η_1, η_2, η_3 et η_4 introduites par l'expansion de Karhunen-Loève de C. Sous la diagonale sont représentés les nuages de points de ces variables aléatoires deux à deux obtenus à partir des échantillons de Quasi-Monte Carlo ayant été utilisés pour estimer la fonction d'auto-covariance.*

pansion de Karhunen-Loève présentent de fortes dépendances, et de ce fait il n'est pas possible d'utiliser directement une expansion en Polynômes du Chaos construite sur des variables aléatoires indépendantes comme précédemment. Deux solutions sont proposées ici, l'une consistant à construire une base de polynômes du chaos orthogonaux pour les variables aléatoires, et une autre consistant à construire des nouvelles variables aléatoires indépendantes à partir des variables aléatoires dépendantes η_k, et une expansion en Polynômes du Chaos sera alors construites en utilisant ces variables aléatoires indépendantes.

8.2.2.1 Utilisation directe des variables dépendantes

La construction de la base de Polynômes du Chaos est réalisée à l'aide des échantillons de Quasi-Monte Carlo des variables aléatoires η_k. Le produit scalaire entre deux fonctions f et g étant estimé à l'aide des échantillons $\eta^{(i,q)}$ de la méthode de Quasi-Monte Carlo, suivant la relation (8.22), Q étant le nombre de séquences de Sobol randomisées utilisées, et N le nombre de points

par séquence de Sobol.

$$\langle f, g \rangle = \int f(\eta) g(\eta) \pi_\eta(\eta) d\eta \approx \sum_{q=1}^{Q} \sum_{i=1}^{N} f(\eta^{(i,q)}) g(\eta^{(i,q)}) \tag{8.22}$$

Comme expliqué dans le chapitre 4, le résultat de l'algorithme de Gram-Schmidt ne dépend que de l'ordre choisi pour les éléments de la base. L'ordre choisi ici pour les éléments de base suit les règles suivantes, les règles énoncées en première étant prioritaires sur les règles suivantes.

- Pour deux polynômes, le polynôme de plus faible degré total apparaît en premier dans la base
- Pour deux polynômes de même degré total, le polynôme dépendant du moins de variables apparaît en premier dans la base
- Pour deux polynômes de même degré total et dépendant d'un même nombre de variables, les polynômes sont classés suivant l'ordre lexicographique pour leur jeu de variables (η_1 avant η_2 ...)

Cet ordre assure que le premier polynôme est le polynôme constant égale 1, et les polynômes de degré 1 sont les monômes η_k du fait de la normalisation de ceux-ci et de leur décorrélation deux à deux. L'algorithme de Gram-Schmidt basique a ici été utilisé. Une fois la base de polynômes du chaos généralisés construite, il est possible de projeter le terme source incertain $\dot{\omega}_{Y_c}$ sur celle-ci, plus précisément son logarithme comme précédemment. Cette projection du logarithme du terme source a été réalisée pour l'ensemble des quatre conditions initiales, et la modélisation ainsi obtenue du terme source a été utilisée afin d'intégrer l'équation différentielle (7.4). Une fois cette équation différentielle intégrée, le délai d'auto-allumage $\tau^{C=0.1}$ a ainsi pu être obtenu et comparé à celui obtenu grâce à une chimie détaillée incertaine à travers le diagramme quantile-quantile de la figure 8.37. Les diagrammes quantile-quantile pour chacune des conditions initiales comportent les résultats obtenus à l'aide d'une expansion en polynôme du chaos impliquant 1 à 3 des variables η_k de degré maximal 5, les échantillons des variables aléatoires η_k du Quasi-Monte Carlo ayant été utilisés.

Comme attendu, la prise en compte d'un nombre croissant de variables aléatoires permet une amélioration de la reproduction du délai d'auto-allumage incertain $\tau^{C=0.1}$. La prise en compte des 3 variables aléatoires η_1, η_2 et η_3 permet de reproduire correctement ce délai, avec une légère sous estimation des délais d'auto-allumages longs pour l'ensemble des conditions initiales, et une sur estimation des délais d'auto-allumage courts pour le point de fonctionnement à $800K$. Il est également possible de regarder les statistiques du processus stochastique C que sont sa moyenne et son écart-type pour les différentes conditions initiales, ce qui est fait sur les figures 8.38 et 8.39 respectivement.

Des conclusions similaires à celle du délai d'auto-allumage peuvent être faites pour la reproduction de la moyenne et de l'écart-type temporels du pro-

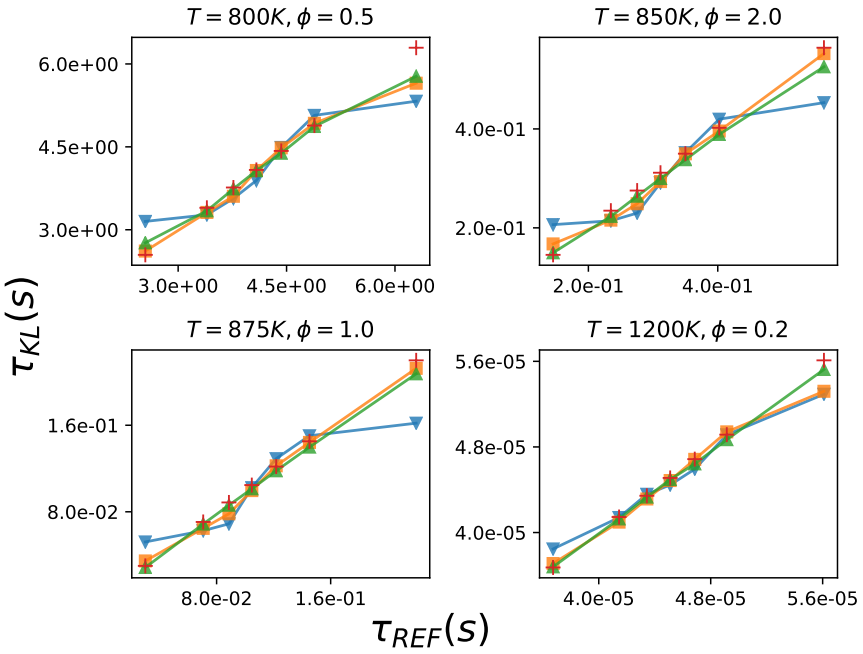

FIGURE 8.37 – *Diagramme quantile-quantile du délai d'auto-allumage $\tau^{C=0.1}$ pour les différents points de fonctionnements, calculés à l'aide de la chimie détaillée incertaine et à l'aide d'une expansion en Polynôme du Chaos du terme source $\dot{\omega}_{Y_c}$, impliquant les η_k : 1 paramètre incertain (triangles bas), 2 paramètres incertains (carrés), 3 paramètres incertains (triangles hauts).*

cessus stochastique C. L'augmentation du nombres de variables aléatoires prises en compte permet pour l'ensemble des conditions initiales une amélioration de la reproduction de la moyenne et de l'écart-type temporels du processus stochastique. De plus, pour les instants les plus avancés, l'écart-type est sous estimé pour l'ensemble des conditions initiales, fait pouvant être mis en relation avec la sous estimation des délais d'auto-allumage longs pour l'ensemble des conditions initiales.

Les résultats obtenus à l'aide d'expansion en Polynômes du Chaos de $\dot{\omega}_{Y_c}$ dépendant des η_k obtenus via l'expansion de Karhunen-Loève du processus stochastique C sont proches de ceux obtenus avec l'utilisation de l'expansion de Karhunen-Loève du processus stochastique $\dot{\omega}_{Y_c}^{\log,\phi,opt}$. Cependant, le degré de l'expansion en polynômes du chaos a ici été limité à 5, en partie par le fait que la valeur absolue des coefficients de l'expansion polynomiale impliqués explosait, entraînant ainsi de l'instabilité numérique. Cette explosion de la valeur absolue des coefficients n'a pas été investiguée, et en particulier il n'est pas clair si elle est un artefact numérique lié à l'utilisation de l'algorithme de Gram-Schmidt basique qui est connu pour être instable numériquement [13], ou s'il est intrinsèque aux variables aléatoires η_k. Une autre approche a cependant été considérée, consistant à se ramener à des variables aléatoires indépendantes.

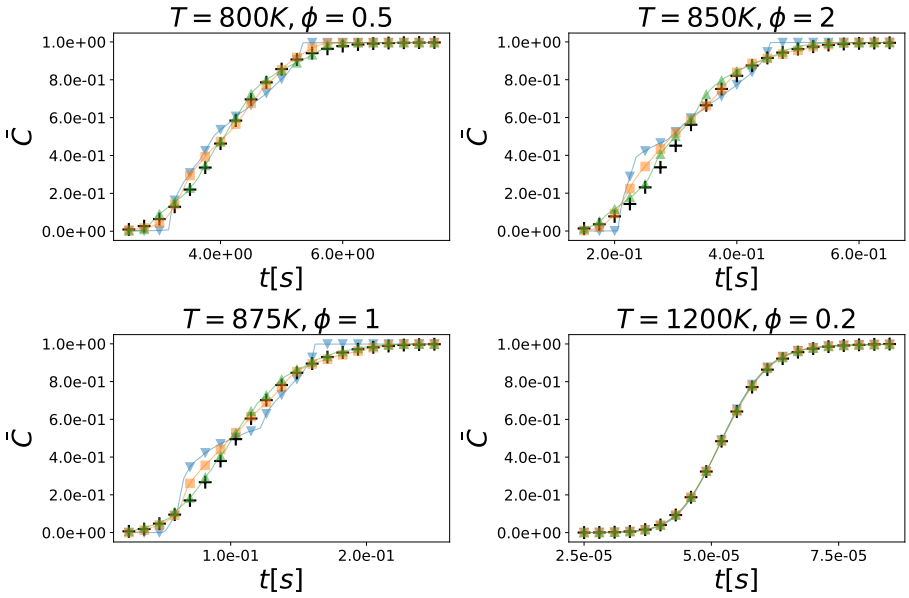

FIGURE 8.38 – *Moyenne temporelle du processus stochastique C pour les quatre conditions initiales choisies, obtenue à l'aide de la chimie détaillée (croix), ainsi que d'expansion en Polynôme du Chaos impliquant juste η_1 (triangles bas), impliquant η_1 et η_2 (carrés), et impliquant η_1, η_2 et η_3 (triangles hauts).*

8.2.2.2 Utilisation de variables indépendantes

L'ensemble des échantillons des variables aléatoires η_k obtenu, dont une présentation graphique partielle est faite avec la figure 8.36, permet d'obtenir une estimation $\tilde{\pi}$ de la densité jointe π de celles-ci à l'aide des méthodes à noyaux gaussiens présentées dans le chapitre 5. Dans ce même chapitre a été présenté comment passer d'un vecteur aléatoire uniforme $\mathbf{U}$ sur l'hypercube $[0,1]^d$ à un vecteur aléatoire de densité π, et inversement à l'aide de la méthode d'inversion généralisée. En fait, pour un vecteur aléatoire à d composantes, il existe $d!$ façons d'utiliser la méthode d'inversion généralisée, suivant l'ordre de construction des composantes du vecteur aléatoire. Dans le cas présent, l'objectif est de construire à partir du vecteur aléatoire $(\eta_1, \ldots, \eta_d)$ un vecteur aléatoire uniforme $(U_1, \ldots, U_d)$ sur l'hypercube $[0,1]^d$. Compte tenu du fait que les η_k sont rangés par leur ordre d'importance dans leur contribution à l'expansion de Karhunen-Loève, l'ordre naturel va ici être conservé dans l'utilisation de la méthode d'inversion généralisée. En effet, la densité jointe utilisée n'est qu'une estimation de la vraie densité, et comme présenté dans le chapitre 5, les premières composantes du vecteur aléatoire $\mathbf{U}$ sont mieux estimées que les suivantes. Dans la suite, on distinguera le vecteur aléatoire théorique $\mathbf{U}$ qu'il est possible de construire avec la connaissance de la densité jointe π, du vecteur aléatoire $\tilde{\mathbf{U}}$ construit à partir de la densité jointe estimée $\tilde{\pi}$, qui n'est

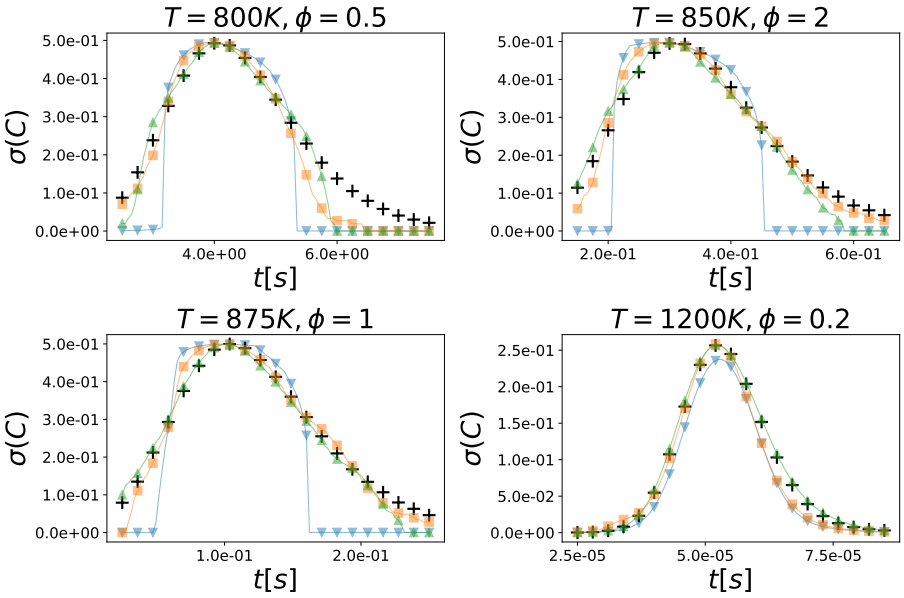

FIGURE 8.39 – *Écart-type temporelle du processus stochastique C pour les quatre conditions initiales choisies, obtenue à l'aide de la chimie détaillée (croix), ainsi que d'expansion en polynômes du chaos impliquant juste η_1 (triangles bas), impliquant η_1 et η_2 (carrés), et impliquant η_1, η_2 et η_3 (triangles hauts).*

pas rigoureusement uniforme.

Une fois le vecteur aléatoire $\tilde{\mathbf{U}}$ construit, il est possible grâce à un changement de variable de se ramener à un vecteur aléatoire dont les composantes sont presque indépendantes et suivent chacune une loi quelconque. Afin d'utiliser des polynômes de Hermite, un vecteur aléatoire $\tilde{\xi}$ est obtenu à partir de $\tilde{\mathbf{U}}$ dont chaque composante est construite suivant la relation (8.23).

$$\tilde{\xi}_i = \sqrt{2}\,\mathrm{erf}^{-1}(2\tilde{U}_i - 1) \tag{8.23}$$

Le vecteur aléatoire $\tilde{\xi}$ ainsi construit est alors proche d'un vecteur aléatoire dont les composantes sont des gaussiennes centrées réduites indépendantes. Bien que les variables aléatoires $\tilde{\xi}_k$ ne suivent pas rigoureusement une loi gaussienne, les polynômes de Hermite P_j^{Herm} sont tout de même utilisés afin d'obtenir une expansion en polynômes du chaos.

On suppose dans un premier temps qu'à partir des échantillons de Quasi-Monte Carlo $\eta^{(i,r)}$ du vecteur aléatoire η sont construits des échantillons $\xi^{(i,r)}$ du vecteur aléatoire ξ. On peut alors les coefficients α_j de l'expansion en polynômes

du chaos d'une variable aléatoire X à l'aide de l'expression (8.24).

$$\alpha_j = E\left[X(\xi)P_j^{Herm}(\xi)\right] \approx \sum_{r=1}^{R}\sum_{i=1}^{N} X(\xi^{(i,r)})P_j^{Herm}(\xi^{(i,r)}) \tag{8.24}$$

Cependant, en pratique les échantillons du vecteur aléatoire ξ ne sont pas accessibles car la densité π est inconnue, et seule des échantillons de $\tilde{\xi}$ peuvent être construits à partir des échantillons du vecteur aléatoire η et de la densité estimée $\tilde{\pi}$. Les coefficients de l'expansion en Polynômes du Chaos $\tilde{\alpha_j}$ utilisant les échantillons du vecteur aléatoire $\tilde{\xi}$ sont calculées similairement aux coefficients α_j, au travers de l'expression (8.25).

$$\tilde{\alpha_j} \approx \sum_{r=1}^{R}\sum_{i=1}^{N} X(\tilde{\xi}^{(i,r)})P_j^{Herm}(\tilde{\xi}^{(i,r)}) \tag{8.25}$$

En utilisant l'expression (8.25), les coefficients de l'expansion en Polynômes du Chaos pour le logarithme du terme source incertain $\dot{\omega}_{Y_c}$ ont été calculés pour les quatre conditions initiales, en utilisant le même maillage en la variable d'avancement c que pour l'expansion en polynômes du chaos utilisant les variables aléatoires initiales. Une fois ces expressions en polynômes du chaos construites, il devient possible de les utiliser dans l'équation différentielle (7.4) pour obtenir des trajectoires de C, permettant d'obtenir des statistiques du processus stochastique C ou du délai d'auto-allumage $\tau^{C=0.1}$ par exemple. L'utilisation de ces expansions en polynômes du chaos se fait avec des variables gaussiennes centrées réduites indépendantes, celles-ci ayant été construites avec cette hypothèse. Les diagrammes quantile-quantile pour le délai d'auto-allumage $\tau^{C=0.1}$ pour les quatre conditions initiales sont visibles sur la figure 8.40, le degré maximale de l'expansion en polynômes du chaos considéré étant 2 et celles-ci impliquant de 1 à 3 variables indépendantes.

Sur la figure 8.40, l'augmentation du nombres de variables prises en compte dégrade la qualité de la reproduction du délai d'auto-allumage incertain $\tau^{C=0.1}$, et cette dégradation est plus importante en passant de 2 à 3 variables qu'en passant de 1 à 2 variables. De plus, la reproduction en ne considérant qu'une seule variable est bien plus acceptable que les autres. Une explication possible à ce phénomène est le fait que la première variable aléatoire est plus proche d'une gaussienne centrée réduite que la seconde variable aléatoire, étant elle même plus proche d'une variable centrée réduite que la troisième et ainsi de suite, et de la même façon des dépendances de plus en plus fortes apparaissent entre les variables aléatoires d'indice important. Cet écart du vecteur aléatoire $\tilde{\xi}$ par rapport au vecteur aléatoire ξ se manifeste notamment lorsque l'on estime numériquement les espérances $E\left[P_i^{Herm}(\tilde{\xi})\right]$ des polynômes de Hermite

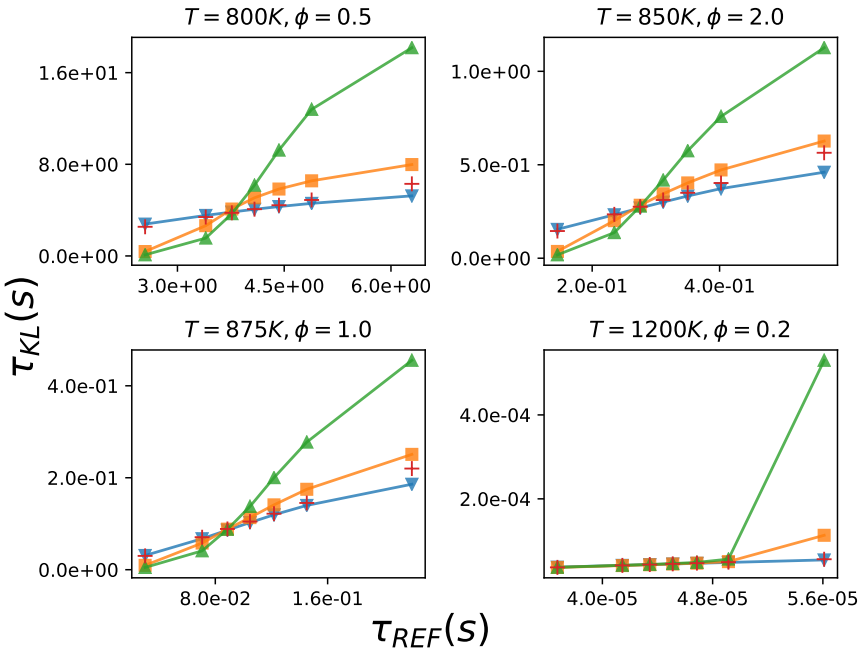

FIGURE 8.40 – *Diagramme quantile-quantile du délai d'auto-allumage $\tau^{C=0.1}$ pour les différents points de fonctionnements, calculés à l'aide de la chimie détaillée incertaine et à l'aide d'une expansion en Polynôme du Chaos du terme source $\dot{\omega}_{Y_c}$, obtenue par l'expression (8.25) : 1 paramètre incertain (triangles bas), 2 paramètres incertains (carrés), 3 paramètres incertains (triangles hauts).*

appliqués au vecteur aléatoire $\tilde{\xi}$.

En effet, l'ensemble de ces espérances devraient être nul excepté pour le polynôme constant, mais ne le sont pas en pratique, les polynômes de Hermite appliqués au vecteur aléatoire $\tilde{\xi}$ possédant un biais. Ce biais peut être estimé à l'aide des échantillons de Quasi-Monte Carlo et est donc accessible. L'idée est alors de prendre en compte ce biais afin de centrer les polynômes du chaos $P_i^{Herm}(\tilde{\xi})$ qui sont centrés dans le cas idéal, lors de leur utilisation pour le calcul des coefficients de l'expansion en polynôme du chaos. Cette dernière affirmation se traduit par l'expression suivante pour l'estimation des coefficients $\tilde{\alpha_j}$ de l'expansion en polynômes du chaos :

$$\tilde{\alpha_j} \approx \sum_{r=1}^{R} \sum_{i=1}^{N} X(\tilde{\xi}^{(i,r)}) \left(P_j^{Herm}(\tilde{\xi}^{(i,r)}) - \sum_{p=1}^{R} \sum_{l=1}^{N} P_j^{Herm}(\tilde{\xi}^{(l,p)}) \right) \tag{8.26}$$

La figure 8.41 présente les diagrammes quantile-quantile pour le délai d'auto-allumage $\tau^{C=0.1}$, comparant le délai d'auto-allumage obtenu à l'aide d'une chimie détaillée incertaine et celui obtenu à l'aide d'une expansion en polynômes du chaos dont les coefficients ont été obtenus via l'expression (8.26).

Les quatre conditions initiales y sont représentées, et les expansions en polynômes du chaos impliquent de 1 à 3 variables aléatoires, et sont toutes d'un degré maximal de 5. Pour l'ensemble des conditions initiales, l'augmentation du nombre de paramètres incertains utilisés dans l'expansion en polynômes du chaos permet une amélioration de la reproduction du délai d'auto-allumage incertain $\tau^{C=0.1}$. De plus, la reproduction du délai d'auto-allumage avec 3 variables ne présente que quelques sous estimations des délais courts et longs pour certaines conditions initiales, mais est globalement de qualité similaire à la reproduction avec 5 paramètres incertains présentées sur la figure 8.31, impliquant l'expansion de Karhunen-Loève de $\dot{\omega}_{Y_c}^{\log,\psi,opt}$.

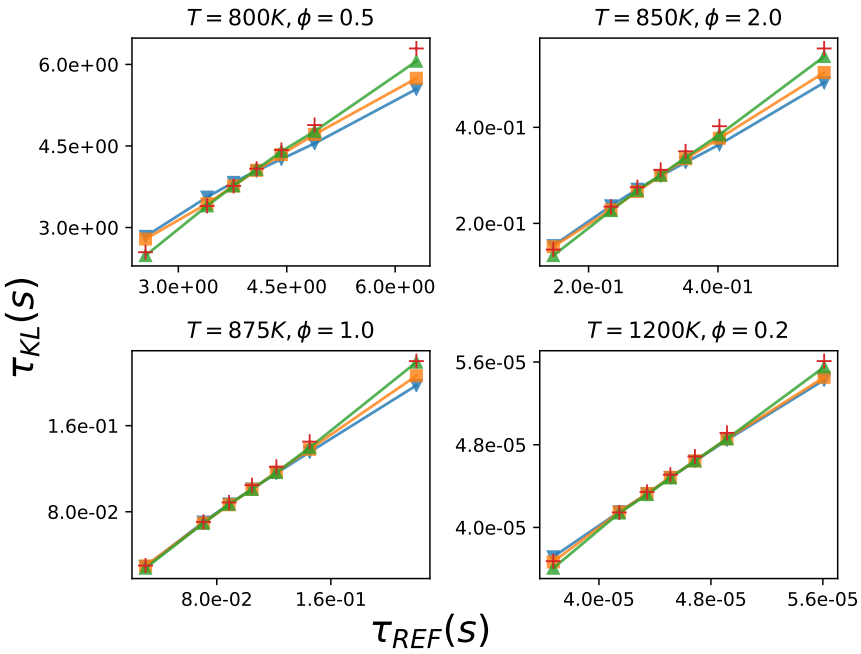

FIGURE 8.41 – *Diagramme quantile-quantile du délai d'auto-allumage $\tau^{C=0.1}$ pour les différentes conditions initiales, calculés à l'aide de la chimie détaillée incertaine et à l'aide d'une expansion en polynôme du chaos du terme source $\dot{\omega}_{Y_c}$, obtenue par l'expression (8.26) : 1 paramètre incertain (triangles bas), 2 paramètres incertains (carrés), 3 paramètres incertains (triangles hauts).*

La prise en compte du biais pour corriger l'estimation des coefficients de l'expansion en polynômes du chaos permet dans la situation présentée d'obtenir une modélisation du terme source à même de reproduire le délai d'auto-allumage incertain avec peu de paramètres incertains. Le délai d'auto-allumage $\tau^{C=0.1}$ n'est cependant pas l'objectif visé, et encore une fois il est intéressant de regarder des statistiques du processus stochastique C. Sur les figures 8.42 et 8.43 sont présentés respectivement les moyennes temporelles et les écart-types temporels du processus stochastique C.

Encore une fois, la bonne reproduction du délai d'auto-allumage et la

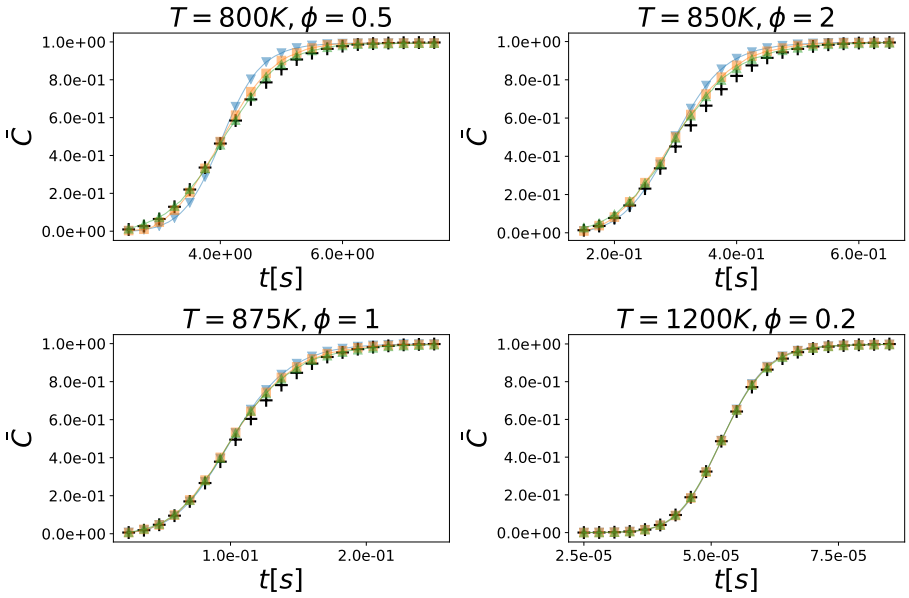

FIGURE 8.42 – *Moyennes temporelles du processus stochastique C pour les quatre conditions initiales choisies, obtenue à l'aide de la chimie détaillée (croix), ainsi que d'expansion en polynôme du chaos impliquant* 1 *variables aléatoires (triangles bas), impliquant* 2 *variables aléatoires (carrés), et impliquant* 3 *variables aléatoires (triangles hauts).*

bonne reproduction des moyennes et des écart-types temporels coïncident. L'augmentation du nombres de variables aléatoires prises en compte permet une amélioration de la reproduction de ces deux statistiques. La moyenne temporelle est correctement reproduite pour l'ensemble des points de fonctionnements, alors que l'écart-type temporel est pour sa part légèrement sous estimé pour les instants avancés de l'ensemble des conditions initiales, et légèrement sur estimé pour les premiers instants montrés du point à la température initiale de $850K$ et à la richesse 2.

La construction de variables aléatoires indépendantes à partir des variables aléatoires construites via l'expansion de Karhunen-Loève du processus stochastique C plutôt que l'utilisation directe de ces variables aléatoires pour une modélisation du terme source incertain $\dot{\omega}_{Y_c}$ à l'aide d'une expansion en polynômes du chaos semble permettre une meilleure reproduction des statistiques d'intérêts du processus stochastique C, du moins pour une expansion en polynômes du chaos avec un degré maximum raisonnable, ici de 5. Cependant, la raison de la présence de cet avantage est encore inconnue, pouvant être liée à une mauvaise construction de la base de polynômes orthonormaux dans le cas des variables dépendantes, ou à l'utilisation d'un degré trop faible. Néanmoins, il est également important de noter que le bon fonctionnement de l'utilisation de variables indépendantes repose sur la qualité de leur construction, qui

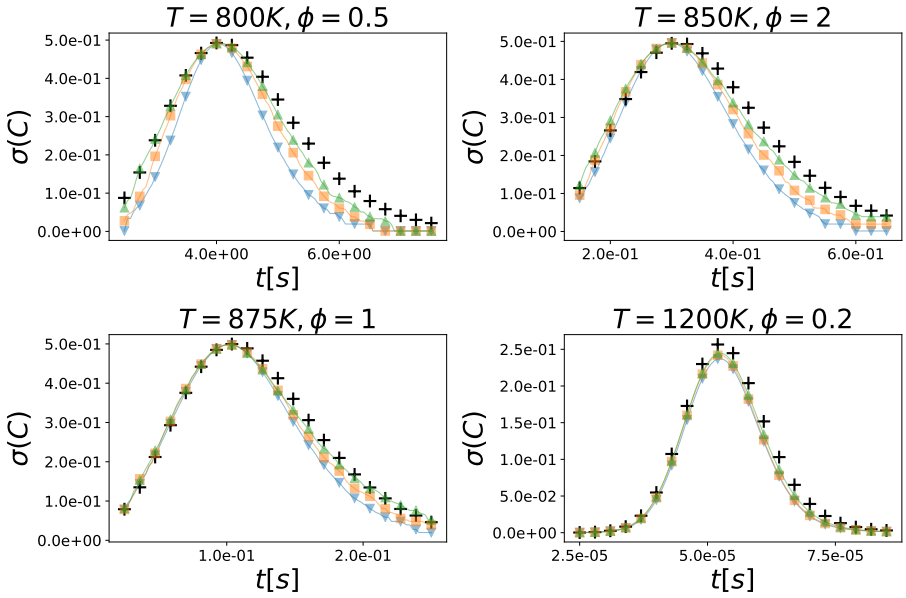

FIGURE 8.43 – *Écart-types temporels du processus stochastique C pour les quatre conditions initiales choisies, obtenue à l'aide de la chimie détaillée (croix), ainsi que d'expansion en polynôme du chaos impliquant* 1 *variables aléatoires (triangles bas), impliquant* 2 *variables aléatoires (carrés), et impliquant* 3 *variables aléatoires (triangles hauts).*

s'avère d'autant plus difficile que ce nombre de variables est important, mais également semblerait-il que les échantillons des variables issues de l'expansion de Karhunen-Loève reposent sur une variété de dimension plus faible que la dimension de l'espace dans laquelle ils résident, ce qui est en partie le cas ici comme visible sur la figure 8.44.

8.2.3 Critique de la modélisation par expansion en polynômes du chaos utilisant les variables aléatoires de l'expansion de Karhunen-Loève de C.

Les reproductions du délai d'auto-allumage incertain $\tau^{C=0.1}$, mais également de statistiques du processus stochastique C ont été possibles, aussi bien avec la méthode consistant à utiliser directement l'expansion de Karhunen-Loève de $\dot{\omega}_{Y_c}^{\log,\psi,opt}$, qu'avec la dernière méthode présentée dans cette section. Cependant, la dernière méthode présente l'avantage d'offrir une bonne reproduction avec un nombre de variables aléatoires plus restreint, se limitant à 3 plutôt qu'à 5. De plus, l'utilisation d'une ou deux variables aléatoires avec cette dernière méthode permet des résultats d'une qualité globalement comparable sur l'ensemble des conditions initiales avec l'utilisation de 3 voire 4 variables aléatoires avec la méthode utilisant directement l'expansion de Karhunen-Loève

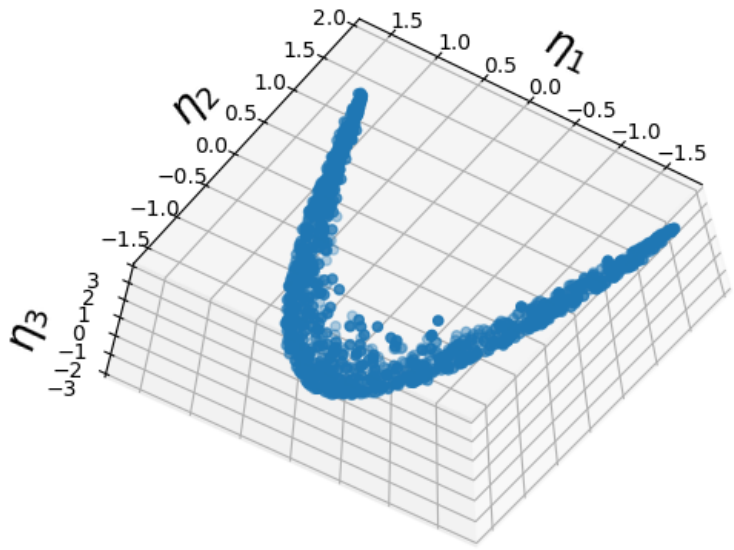

Figure 8.44 – *Nuages de points constitués des* 10×128 *premiers échantillons de Quasi-Monte Carlo du vecteur aléatoire* (η_1, η_2, η_3)*. Une variété de dimension* 2 *se dessine à travers ce nuage de points.*

de $\dot{\omega}_{Y_c}^{\log,\psi,opt}$. Cette dernière méthode semble donc supérieure aux précédentes en ce sens.

Un autre avantage de cette dernière méthode provient du fait que seules les nouvelles variables aléatoires introduites par l'expansion de Karhunen-Loève du processus stochastique C sont utiles. De ce fait, aucun problème d'interpolation lors de la reconstruction de la partie modélisée du terme source incertain $\dot{\omega}_{Y_c}$ ne pose de problèmes. Le nombre de points de cubature utilisé dans la méthode de Nyström n'est donc critique que pour la convergence du calcul de l'expansion de Karhunen-Loève afin d'obtenir des échantillons correctes des variables aléatoires η_k. De plus, l'utilisation de tensorisation creuse semble adaptée dans cette situation, offrant la possibilité de réduire le nombre de points de cubatures. Ainsi, l'étude réalisée ici impliquait 5455 points de cubature, mais les η_k obtenus avec 3231 points de cubatures étaient les mêmes comme le montre la figure 8.45, et aurait donc mené à des résultats similaires. Le nombre de points de cubatures utilisés ici laisse une marge de manœuvre, permettant d'envisager une table avec au moins 4 dimensions, voire peut être 5 dimensions. Cela permet d'envisager la possibilité de l'étude de cas plus complexes où la dimension de la table est supérieure à 3. Qui plus est, le maillage utilisé ensuite pour

la tabulation n'est plus imposé, et peut être utilisé directement en calculant l'expansion en polynômes du chaos pour chacun des points de ce maillage.

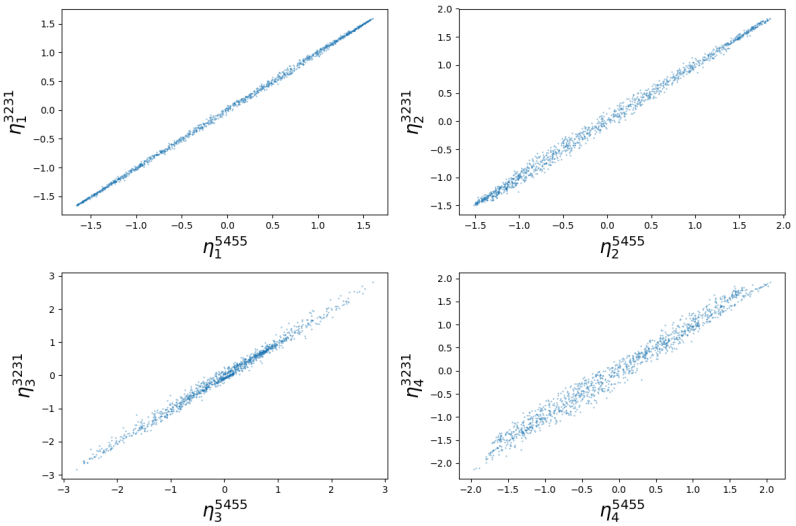

FIGURE 8.45 – *Nuages de points des variables aléatoires* η_k^{3231} *et* η_k^{5455} *pour* k *allant de* 1 *à* 4, *constitués des* 10×128 *premiers échantillons de Quasi-Monte Carlo.*

La figure 8.44 montre que les échantillons du vecteur aléatoire (η_1, η_2, η_3) sont proches d'une variété de dimension 2 dans l'espace de dimension 3. Afin de pouvoir profiter de cette caractéristique de l'ensemble de points, deux opérations essentielles sont à réaliser :

- Envoyer la sous-variété de dimension d dans l'espace euclidien de dimension d.
- Associer une densité de probabilité à l'espace euclidien de dimension d à même de reproduire la densité de probabilité sur la variété.

De nombreuses méthodes permettent d'adresser le premier point [71], c'est à dire de caractériser une sous-variété de dimension d sur laquelle sont situés les échantillons du vecteur aléatoire en la paramétrant avec un nombre de paramètres restreint. Une fois cette étape réalisée, il est possible d'associer à chacun des échantillons du vecteur aléatoire un point dans l'espace euclidien de dimension d. Il est alors possible de construire une densité jointe pour ces nouveaux points, que l'on peut ensuite utiliser pour la propagation d'incertitude.

8.3 Conclusion

Dans ce chapitre, différentes méthodes ont été présentées afin de modéliser les incertitudes du processus stochastique C à l'aide d'un vecteur aléatoire de faible dimension, typiquement inférieure à 5. L'ensemble de ces méthodes propose la propagation des incertitudes en modélisant le terme source incertain $\dot{\omega}_{Y_c}$

comme une fonction de ce vecteur aléatoire. La définition du vecteur aléatoire et son usage varient selon les méthodes proposées.

La première méthode proposée permet de définir de nouvelles variables aléatoires via l'expansion de Karhunen-Loève d'une modification du terme source incertain $\dot{\omega}_{Y_c}$. Chacune des nouvelles variables aléatoires ainsi construite dépend des différentes variables aléatoires initiales, offrant la possibilité de couvrir partiellement les effets de plusieurs des variables aléatoires initiales au sein d'une seule variable aléatoire. Cette méthode offre de plus l'avantage d'offrir un moyen simple pour le choix des variables aléatoires à conserver, celles-ci étant rangés par ordre d'importance. Cependant, la définition d'une bonne modification du processus stochastique $\dot{\omega}_{Y_c}$ n'est pas immédiate. La principale limitation provient cependant du nombre de points de cubatures qui explosent rapidement avec la dimension de la table chimique à considérer, interdisant la résolution pratique du problème aux valeurs propres donné par la méthode de Nyström du fait de sa taille. Dans l'application étudiée, cette méthode a néanmoins permis la bonne reproduction des incertitudes grâce à un vecteur aléatoire de dimension 5. Le vecteur aléatoire obtenu présentait des dépendances entre ses composantes, devant nécessairement être prise en compte. La loi jointe du vecteur aléatoire a pu être estimée à l'aide des échantillons de ce vecteur aléatoire, permettant d'échantillonner selon ce vecteur aléatoire.

La seconde méthode dans l'application étudiée est la méthode ayant permis d'obtenir les meilleurs résultats. Le vecteur aléatoire permettant la paramétrisation du système est construit à partir de l'expansion de Karhunen-Loève du processus stochastique C que l'on cherche à reproduire, après un changement d'indexation de celui-ci afin d'assurer l'existence de cette expansion. L'avantage de cette dernière méthode sur les autres et qu'elle s'intéresse directement au processus stochastique C que l'on cherche à reproduire, plutôt qu'au terme source de celui-ci. Partant de ce vecteur aléatoire, deux stratégies ont été envisagées, toutes deux s'appuyant sur une expansion en polynômes du chaos pour le terme source de la variable d'avancement. La première stratégie a consisté à construire une base de polynômes du chaos adaptée au vecteur aléatoire obtenu, alors même que les composantes de ce vecteur aléatoire présentent des dépendances. Les résultats obtenus montraient des difficultés à reproduire correctement le délai d'auto-allumage incertain $\tau^{C=0.1}$ avec la considération d'un vecteur aléatoire possédant 3 composantes. La raison de cela n'est cependant pas claire, et une investigation plus poussée reste nécessaire, mais la reproduction correcte de ce même délai d'auto-allumage avec un vecteur aléatoire possédant 3 composantes également grâce à la seconde stratégie présentée après laisse penser que la construction de la base de polynômes du chaos a souffert d'instabilité numérique. Cette seconde stratégie repose sur l'estimation de la densité de probabilité du vecteur aléatoire permettant d'exprimer celui-ci en fonction d'un vecteur aléatoire de même dimension mais possédant une indépendance de ses composantes. La base de polynômes du chaos utilisée est alors construite pour ce dernier vecteur aléatoire à composantes indépendantes, per-

mettant une expansion en polynômes du chaos du terme source incertain de la variable d'avancement, le calcul de cette expansion nécessitant une correction due à l'erreur introduite par l'estimation imparfaite de la densité de probabilité du vecteur aléatoire initial. Des pistes d'améliorations sont apparues avec la dernière méthode, les échantillons du vecteur aléatoire issu de l'expansion de Karhunen-Loève du processus stochastique C semblant se trouver sur une variété de plus faible dimension que la dimension de l'espace.

L'ensemble des méthodes a permis une reproduction correcte des deux premiers moments du processus stochastique C pour un réacteur homogène à pression constante et adiabatique d'un mélange air-hydrogène. Cette configuration offrait l'avantage d'avoir le plus faible coût de calcul parmi les configurations utilisables dans des simulations haute fidélité de systèmes complexes, et offrait donc un objet d'étude idéal pour le développement de ces méthodes du fait de temps de retour raisonnable. Il convient de vérifier si les résultats obtenus peuvent s'étendre à d'autres configurations canoniques de systèmes chimique, mais également avec des mécanismes réactionnels impliquant des carburants plus complexes que l'hydrogène. L'utilisation de configurations différentes impliquera de choisir :

- une modification intéressante du terme source incertain, permettant de ne garder qu'un faible nombre de terme dans l'expansion de Karhunen-Loève de celui-ci pour la première méthode
- de choisir, ou de ne pas choisir suivant le cas, un nouveau paramétrage du processus stochastique C (ou Y_c) pour la seconde méthode

Les différentes méthodes peuvent à priori s'appliquer pour d'autres configurations, et sans retour d'expérience les concernant, il n'est pas possible de s'avancer concernant la possibilité de réduire significativement la dimension stochastique suffisamment dans ces cas, notamment en cas d'utilisation d'un mécanisme réactionnel plus complexe.

Conclusion

Travail réalisé

Cette thèse s'inscrit dans le développement du sujet de la propagation d'incertitudes pour la simulation numérique de la combustion au sein du laboratoire EM2C. Deux objectifs principaux étaient présents au sein de cette thèse.
Le premier objectif consistait en l'appropriation d'outils du domaine de la propagation d'incertitudes au sein du laboratoire. Un état de l'art de l'ensemble de ces outils a permis de ne considérer que les méthodes non intrusives pour la propagation d'incertitudes, le coût associé au développement et à l'utilisation de méthodes intrusives étant trop important. L'étude de ces méthodes non intrusive a permis le développement d'un code de calcul parallèle utilisable sur des supercalculateurs pour les méthodes de Monte Carlo et de Quasi-Monte Carlo randomisées, permettant l'obtention de statistiques sur des systèmes de combustion canonique efficacement. Cela a également permis l'utilisation de la méthode de Quasi-Monte Carlo randomisée pour la résolution de l'équation de transfert radiatif dans des simulations $3D$ couplées complexes permettant un gain significatif en terme de coût de calcul pour de telles simulations. En plus de cela, un code de calcul a également été écrit pour l'estimation numérique d'intégrales multiples en dimension faible utilisant des méthodes de cubature, ainsi qu'un code de calcul permettant l'utilisation de méthodes spectrales pour la représentation de processus stochastiques, parmi lesquelles les expansions en polynômes du chaos généralisés et l'expansion de Karhunen-Loève.
Le second objectif était le développement d'une méthodologie pour la propagation d'incertitudes sur les paramètres de cinétique chimique au sein de simulations aux grandes échelles. Cette méthodologie vise uniquement les simulations aux grandes échelles pour lesquelles une tabulation de la cinétique chimique est utilisée avec une description de l'évolution de la réaction de combustion prise en compte au travers de l'évolution d'une variable d'avancement de la réaction. Il a été montré dans cette thèse que dans le cas où la modélisation du processus de combustion est réalisée à l'aide d'un réacteur adiabatique à pression constante d'un mélange air-hydrogène homogène, les incertitudes sur l'ensemble des variables pouvaient être séparées en deux composantes : la première composante correspond aux incertitudes induites par l'incertitude sur la variable

d'avancement, alors que la seconde composante correspond à une incertitude intrinsèque à la variable. Il se trouve que pour les variables influençant l'écoulement, la seconde composante des incertitudes est négligeable devant l'autre, ce qui couplé à une certaine invariance temporelle des incertitudes implique que les incertitudes sur ces variables sont très bien expliquées par les incertitudes de la variable d'avancement, rendant possible de ne considérer que cette dernière comme incertaine pour propager les incertitudes de la cinétique chimique au sein de l'écoulement. En fait, pour l'ensemble des variables possédant la seconde composante des incertitudes négligeable, la seule considération des incertitudes de la variable d'avancement permet de reproduire les incertitudes sur ces variables. Pour les variables possédant leur seconde composante des incertitudes non négligeable, comme c'est le cas pour les espèces HO_2 et H_2O_2, il est tout de même possible de retrouver la variance de ces espèces au prix de rajouter une variable à la table correspondant à la variance intrinsèque de ces espèces. La reproduction des incertitudes sur la variable d'avancement a été réalisée à l'aide de la reproduction du terme source incertain de celle-ci à l'aide de différentes méthodes spectrales pour la reproduction de processus stochastiques. Le but était d'avoir une représentation de ce processus stochastique dépendant du plus petit nombre de variables aléatoires possible. Deux choix ont alors été identifiés pour l'obtention de ce jeu restreint de variables aléatoires :

- Le premier choix consiste à ne garder que les variables aléatoires initiales influençant le plus le système considéré, nécessitant de spécifier un critère permettant de caractériser l'influence de ces variables aléatoires, qui a été le délai d'auto-allumage dans le cas étudié
- Le second choix consiste à construire de nouvelles variables aléatoires en considérant une représentation spectrale optimale d'un processus stochastique du système. Différents choix ont été explorés pour le choix de ce processus stochastique, parmi lesquelles des versions modifiées du terme source de la variable d'avancement, ou encore une version modifiée de la variable d'avancement.

Perspectives

Ce travail de thèse a permis de définir l'ensemble des ingrédients nécessaires à une propagation d'incertitudes au sein d'une simulation aux grandes échelles basée sur une chimie tabulée utilisant une variable d'avancement pour caractériser l'état d'avancement du processus de combustion, pour un coût de calcul restant abordable actuellement. Ce travail n'est pas achevé, et différents résultats sont encore à obtenir, en plus de possibles voies d'améliorations de la méthode actuelle :

- Le premier résultat à obtenir est l'application des résultats de cette thèse à une simulation aux grandes échelles pour laquelle la chimie est modélisée à l'aide d'un réacteur adiabatique à pression constante d'un

mélange air-hydrogène

— Une autre partie du travail restant consistera en l'étude d'autres carburants que l'hydrogène, ainsi qu'à d'autres configurations canoniques. Il sera alors nécessaire de vérifier si les mêmes comportement sont observées, notamment considérant la caractérisation des incertitudes des variables influençant l'écoulement à l'aide des seules incertitudes sur la variable d'avancement, qui est une condition nécessaire à la propagation des incertitudes au sein d'une simulation aux grandes échelles à l'aide de la méthodologie proposée

— Investiguer d'autres méthodes de représentation de processus stochastique, en particulier des méthodes non linéaires qui pourraient sans doute permettre une réduction plus importante de la dimension comme suggéré par la figure 8.44 sur laquelle une variété de dimension 2 se dessine

Annexe A

Fonctions test pour l'intégration numérique

Cet annexe présente les fonctions de Genz [40] dont le but est de tester les différentes méthodes d'intégration numérique.

A.1 Fonctions test de Genz en dimension 1

A.1.1 Présentation des fonctions de Genz en dimension 1

Les fonctions de Genz d'une seule variable sont toutes définies sur le segment $[0, 1]$ et dépendent toutes de deux paramètres, à l'exception de la fonction de type "Corner Peak" :

- un paramètre a positif, traduisant la difficulté d'intégration
- un paramètre u appartenant à $[0, 1]$, permettant de déterminer un aspect de la fonction (déphasage, localisation du minimum, localisation du maximum ou localisation d'une irrégularité)

Chacune des fonctions possède une allure différente, et le nom choisit par Genz traduit l'allure générale de chacune des fonctions. L'expression de ces fonctions est donné dans le tableau A.1.

Sur la figure A.1 sont représentées les allures des différentes fonctions de

Dénomination	Expression
Oscillatory	$g(x) = \cos(2\pi u + ax)$
Continuous	$g(x) = \exp(-a\lvert x - u\rvert)$
Discontinuous	$g(x) = 0$ si $x > u, \exp(ax)$ sinon.
Product Peak	$g(x) = \frac{1}{a^{-2}+(x-u)^{-2}}$
Corner Peak	$g(x) = (1 + ax)^{-2}$
Gaussian	$g(x) = \exp\left(-a^2(x-u)^2\right)$

Table A.1 – *Expression des fonctions de Genz dépendant d'une seule variable.*

Genz pour différentes valeurs du paramètre a représentant la difficulté d'intégration. L'augmentation de la valeur du paramètre a permet d'accentuer la caractéristique de chacune des fonctions, rendant la fonction "oscillatory" plus oscillante, ou piquant plus les fonctions présentant un pic par exemple.

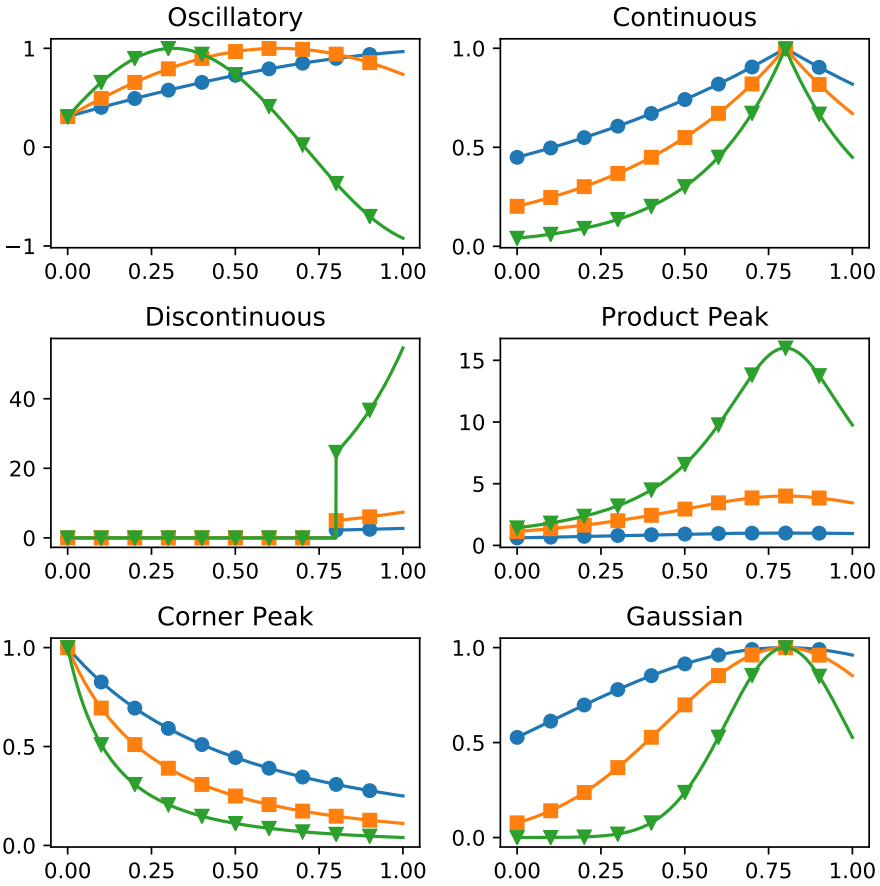

FIGURE A.1 – *Allure des différentes fonctions test de Genz d'une variable, avec une valeur de $u = 0.8$ et différentes valeurs de a : $a = 1$ (ronds), $a = 2$ (carrés) et $a = 4$ (triangles).*

A.1.2 Valeur analytique de l'intégrale des fonctions de Genz

Les fonctions de Genz présente l'avantage que leur intégrale peut se calculer analytiquement, ce qui permet de comparer les résultats d'intégration numérique avec la valeur exacte de l'intégrale. Les valeurs des intégrales sur le segment $[0, 1]$ des différentes fonctions dans le cas monovariable en fonction des paramètres a et u sont répertoriées dans le tableau A.2.

Dénomination	Expression analytique de l'intégrale
Oscillatory	$\frac{\sin(2\pi u+a)-\sin(2\pi u)}{a}$
Continuous	$\frac{2-\exp(-au)-\exp(a(u-1))}{a}$
Discontinuous	$\frac{\exp(au)-1}{a}$
Product Peak	$a\left(\arctan\left[a(1-u)\right]+\arctan(au)\right)$
Corner Peak	$\frac{1}{a}\left(1-\frac{1}{1+a}\right)$
Gaussian	$\frac{\sqrt{\pi}}{2a}\left(\operatorname{erf}\left[a(1-u)\right]+\operatorname{erf}(au)\right)$

TABLE A.2 – *Expressions analytiques des intégrales des différentes fonctions de Genz en fonction des paramètres a et u.*

A.2 Fonctions test de Genz en dimension supérieure à 2

A.2.1 Présentation des fonctions de Genz en dimension supérieure à 2

Les fonctions de Genz de plusieurs variables sont définies sur l'hypercube unité $[0,1]^d$ et sont définies à l'aide de deux vecteurs de paramètres $\mathbf{a}$ et $\mathbf{u}$ de dimension d :

- un vecteur de paramètres $\mathbf{a}$ dont les éléments sont tous positifs, traduisant la difficulté d'intégration
- un vecteur de paramètres $\mathbf{u}$ dont les éléments appartiennent à $[0,1]$, permettant de déterminer un aspect de la fonction (déphasage, localisation du minimum, localisation du maximum ou localisation d'une irrégularité)

Les valeurs présentes dans le vecteur $\mathbf{a}$ permettent de jouer sur la difficulté d'intégration de la fonction. Il est ainsi possible de définir une difficulté anisotrope, certaines dimensions étant plus difficile à intégrer que d'autres. Sur la figure A.2 sont représentées les allures des différentes fonctions de Genz dépendant de deux variables pour un vecteur de paramètre $\mathbf{a} = (2, 10)$ et un vecteur de paramètres $\mathbf{u} = (0.8, 0.2)$.

A.2.2 Valeur analytique de l'intégrale des fonctions de Genz

L'ensemble des fonctions de Genz de plusieurs variables possèdent une expression analytique pour leur intégrale sur l'hypercube $[0,1]^d$, en fonctions des vecteurs de paramètres $\mathbf{a}$ et $\mathbf{u}$. Cependant, pour les fonctions "Oscillatory" et "Corner Peak", cette expression analytique est complexe à trouver et seule une

Dénomination	Expression
Oscillatory	$g(x) = \cos\left(2\pi u_1 + \sum_{i=1}^{d} a_i x_i\right)$
Continuous	$g(x) = \exp\left(-\sum_{i=1}^{d} a_i \lvert x_i - u_i\rvert\right)$
Discontinuous	$g(x) = 0$ si $x_1 > u_1$ ou $x_2 > u_2, \exp\left(\sum_{i=1}^{d} a_i x_i\right)$ sinon.
Product Peak	$g(x) = \prod_{i=1}^{d} \frac{1}{a_i^{-2} + (x_i - u_i)^{-2}}$
Corner Peak	$g(x) = \left(1 + \sum_{i=1}^{d} a_i x_i\right)^{-d-1}$
Gaussian	$g(x) = \exp\left(-\sum_{i=1}^{d} a_i^2 (x_i - u_i)^2\right)$

TABLE A.3 – *Expression des fonctions de Genz dépendant de d variables.*

expression récurrente sera donnée, qu'il est possible d'implémenter simplement à l'aide d'une fonction récursive. L'expression analytique des intégrales pour les 4 autres fonctions sont répertoriées dans le tableau A.4.

Concernant l'intégration des fonctions "Oscillatory", il est possible de trouver une formule récursive la définissant. Afin d'expliciter cette formule, les grandeurs $I_{\mathbf{a},\mathbf{u}}^{\cos,k}$ et $I_{\mathbf{a},\mathbf{u}}^{\sin,k}$ sont définies, respectivement par les expressions (A.1) et (A.2).

$$I_{\mathbf{a},\mathbf{u}}^{\cos,k} = \int_{[0,1]^k} \cos\left(2\pi u_1 + \sum_{i=1}^{k} a_i x_i\right) d\mathbf{x} \tag{A.1}$$

$$I_{\mathbf{a},\mathbf{u}}^{\sin,k} = \int_{[0,1]^k} \sin\left(2\pi u_1 + \sum_{i=1}^{k} a_i x_i\right) d\mathbf{x} \tag{A.2}$$

L'intégrale de la fonction "Oscillatory" correspond à la valeur de $I_{\mathbf{a},\mathbf{u}}^{\cos,k}$. En utilisant les formules d'addition trigonométrique et en intégrant par rapport à la dernière variable, il vient la relation de récurrence suivante :

$$\forall 1 \leq k \leq d, \begin{cases} I_{\mathbf{a},\mathbf{u}}^{\cos,k} = \dfrac{\sin(a_k)}{a_k} I_{\mathbf{a},\mathbf{u}}^{\cos,k-1} - \dfrac{1-\cos(a_d)}{a_d} I_{\mathbf{a},\mathbf{u}}^{\sin,k-1} \\ I_{\mathbf{a},\mathbf{u}}^{\sin,k} = \dfrac{\sin(a_k)}{a_k} I_{\mathbf{a},\mathbf{u}}^{\sin,k-1} + \dfrac{1-\cos(a_d)}{a_d} I_{\mathbf{a},\mathbf{u}}^{\cos,k-1} \end{cases} \tag{A.3}$$

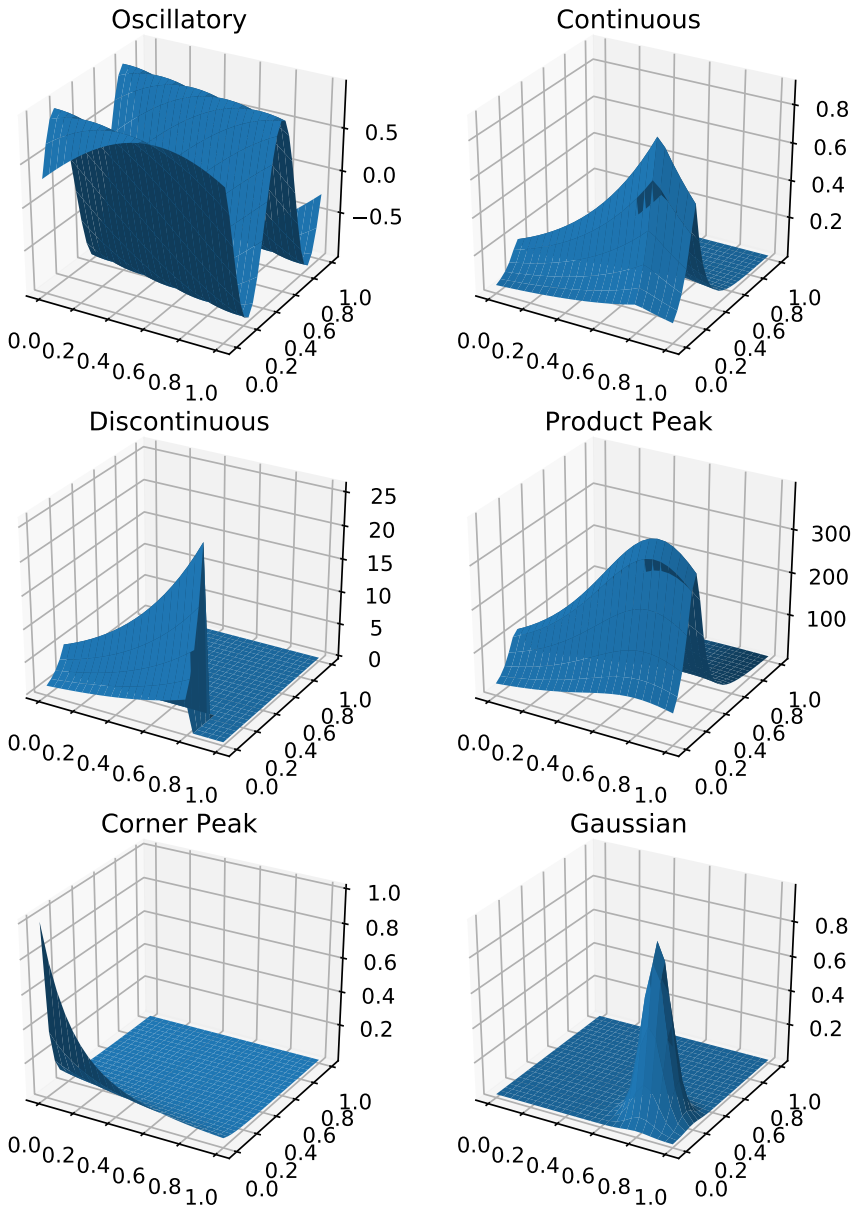

FIGURE A.2 – *Allure des différentes fonctions test de Genz de deux variables, avec un valeur de* $\mathbf{a} = (2, 10)$ *et une valeur de* $\mathbf{u} = (0.8, 0.2)$.

Les valeurs $I^{\cos,0}_{\mathbf{a},\mathbf{u}}$ et $I^{\cos,0}_{\mathbf{a},\mathbf{u}}$ sont quant à elles données par les expressions suivantes :

$$\begin{cases} I^{\cos,0}_{\mathbf{a},\mathbf{u}} = \cos(2\pi u_1) \\ I^{\sin,0}_{\mathbf{a},\mathbf{u}} = \sin(2\pi u_1) \end{cases} \tag{A.4}$$

Dénomination	Expression analytique de l'intégrale
Continuous	$\prod_{i=1}^{d} \frac{2-\exp(-a_i u_i)-\exp(a_i(u_i-1))}{a}$
Discontinuous	$\prod_{i=1}^{2} \frac{\exp(a_i u_i)-1}{a_i} \prod_{i=2}^{d} \frac{\exp(a_i)-1}{a_i}$
Product Peak	$\prod_{i=1}^{d} a_i \left(\arctan\left[a_i(1-u_i)\right] + \arctan(a_i u_i)\right)$
Gaussian	$\prod_{i=1}^{d} \frac{\sqrt{\pi}}{2a_i} \left(\operatorname{erf}\left[a_i(1-u_i)\right] + \operatorname{erf}(a_i u_i)\right)$

TABLE A.4 – *Expressions analytiques des intégrales des différentes fonctions de Genz en fonction des vecteurs de paramètres* $\mathbf{a}$ *et* $\mathbf{u}$.

Il est donc possible de calculer la valeur exacte de l'intégrale des fonctions "Oscillatory" en un temps linéaire en la dimension d.
Concernant les fonctions de type "Corner Peak", une relation de récurrence peut également être obtenue. Afin d'expliciter cette relation de récurrence, on définit la grandeur $I^{(k)}_{\mathbf{a},\alpha}$ par l'expression suivante :

$$I^{(k)}_{\mathbf{a},\alpha} = \int_{[0,1]^k} \frac{dx_1 \dots dx_k}{\left(\alpha + \sum_{i=1}^{k} a_i x_i\right)^{k+1}} \tag{A.5}$$

Ainsi définie, la valeur exacte de l'intégrale est donc donnée par $I^{(d)}_{\mathbf{a},1}$. La relation de récurrence suivante, obtenue en intégrant simplement par rapport à la dernière variable, permet de calculer cette grandeur :

$$\begin{cases} I^{(0)}_{\mathbf{a},\alpha} = \dfrac{1}{\alpha} \\ \forall 1 \leq k \leq d, I^{(k)}_{\mathbf{a},\alpha} = \dfrac{I^{(k-1)}_{\mathbf{a},\alpha} - I^{(k-1)}_{\mathbf{a},\alpha+a_k}}{k a_k} \end{cases} \tag{A.6}$$

Cette relation de récurrence implique un algorithme de complexité exponentielle en $O(2^d)$, ce qui rend vite le calcul irréalisable pour des grandes dimensions, $d = 30$ impliquant de l'ordre de 10^9 évaluations de quantités $I^{(k)}_{\mathbf{a},\alpha}$. En pratique, le calcul exact de la valeur de l'intégrale n'est réalisé que pour des valeurs de d inférieure à 30.

Annexe B

Communication à l'ASME

Proceedings of ASME 2017: Turbomachinery Technical Conference and Exposition
GT2017
June 26-30, 2017, Charlotte, NC, USA

GT2017-64179

DRAFT: COMPARISON OF MONTE CARLO METHODS EFFICIENCY TO SOLVE RADIATIVE ENERGY TRANSFER IN HIGH FIDELITY UNSTEADY 3D SIMULATIONS

Lorella Palluotto[§*], Nicolas Dumont[§], Pedro Rodrigues[§], Chai Koren[‡§], Ronan Vicquelin[§], Olivier Gicquel[§]
[§]Laboratoire EM2C, CNRS
CentraleSupélec
Université Paris-Saclay
Grande Voie des Vignes, 92295
Chatenay-Malabry cedex, France

[‡]Air Liquide
Centre de Recherche Paris-Saclay
1 Chemin de la Porte des Loges, 78350
Les-Loges-en-Josas, France

ABSTRACT

The present work assesses different Monte Carlo methods in radiative heat transfer problems, in terms of accuracy and computational cost. Achieving a high scalability on numerous CPUs with the conventional forward Monte Carlo method is not straightforward. The Emission-based Reciprocity Monte Carlo Method (ERM) allows to treat each mesh point independently from the others with a local monitoring of the statistical error, becoming a perfect candidate for high-scalability. ERM is however penalized by a slow statistical convergence in cold absorbing regions. This limitation has been overcome by an Optimized ERM (OERM) using a frequency distribution function based on the emission distribution at the maximum temperature of the system. Another approach to enhance the convergence is the use of low-discrepancy sampling. The obtained Quasi-Monte Carlo method is combined with OERM. The efficiency of the considered Monte-Carlo methods are compared.

*Address all correspondence to this author: lorella.palluotto@centralesupelec.fr

NOMENCLATURE

DNS	Direct Numerical Simulation
FM	Forward Method
I	Radiative intensity [$W\ sr^{-1}\ m^{-2}$]
LES	Large Eddy Simulation
MCM	Monte Carlo Method
N,n	Number [-]
$QMCM$	Quasi Monte Carlo method
ERM	Emission-based Reciprocity Method
$OERM$	Optimized Emission-based Reciprocity Method
P	Radiative power per unit volume [$W\ m^{-3}$]
PDF	Probability Density Function
$RANS$	Reynolds-Averaged Navier-Stokes equations
T	Temperature [K]
T_{CPU}	Computational time [s]
TRI	Turbulence-Radiation Interaction
f	Probability density function [-]
rms	root mean square
Δ	Direction of photon bundle [m]
δ	Channel half-width [m]
η	Efficiency
θ	Polar angle [sr]

κ	Absorption coefficient $[m^{-1}]$
ν	Radiation Wave number $[cm^{-1}]$
Ω	Solid angle $[sr]$
σ	Standard Deviation
σ^2	Variance
ϕ	Azimuthal angle $[sr]$
^{exch}	Exchanged quantity
^{e}	Emitted quantity
$^{\circ}$	Equilibrium quantity

INTRODUCTION

Conductive heat fluxes and radiative energy fluxes at walls greatly affect the design stage and the material choice of combustion systems. Incorporating these different contributions in numerical simulations is therefore a great challenge that is widely investigated. In the context of gas turbines, the efficient mitigation of conduction from burnt gases with film and effusion cooling leaves radiation as the main contributor to wall heat fluxes. Radiative heat transfer is however difficult to account for in turbulent flows. Local radiative intensity is indeed strongly correlated to the instantaneous medium distribution in the spatial domain. Furthermore it also shows a highly non-linear response to temperature and species concentrations. Therefore accurate calculation of radiative transfer requires an instantaneous spatially resolved information regarding the temperature and species composition fields. Carrying out RANS simulations does not provide such information as only average quantities are calculated. Then accounting for Turbulence-Radiation Interaction (TRI) [1, 2] in such configuration requires TRI modelling. While deriving such models is still an ongoing research domain, another approach to alleviate significantly this modeling issue is to couple the radiative solver to direct numerical simulations (DNS) as in [3–5], that fully resolves in time and space the flow field, but these simulations remain not accessible for use in large-scale applications. Therefore a intermediate choice is to use large-eddy simulation (LES) instead of DNS, providing time resolved solution and a good estimation of the spatial correlation in the simulation domain. The subgrid-scale TRI effects are nonetheless strictly not negligible and modeling efforts are ongoing [6, 7].

As regarding the methods to solve the radiative transfer equation, Monte Carlo methods are the more interesting for their straightforward accounting for spectral gas radiative properties and for complex geometries. A Monte Carlo method (MCM) is a statistical method where a large number of stochastic events is simulated. In radiative transfer a stochastic event is represented by an optical path of photons bundles whose departure point, propagation direction and spectral frequency are independently and randomly chosen according to given distribution functions. The average of all the stochastic events contributions give/constitutes the solution of the problem, *i.e.* the local values of radiative power and wall radiative fluxes. In the conventional Forward Method a large number of photon bundles are emitted in the whole system and their history is traced until the carried energy is absorbed by the participative medium, at the wall, or until it exits the system.

Such methods provide an estimation of the statistical error for the computed radiative power and wall fluxes, commonly represented by the standard deviation. The standard deviation tends to be proportional to $1/\sqrt{N}$ (Howell 1998), where N is the total number of bundles. One of the main drawbacks is the need of a large number of rays to obtain statistically and physically meaningful results, and this handicap becomes stronger in optically thick media, where most of photons are absorbed in the vicinity of their emission source. Although these methods are deemed to be computationally expensive, all the more when coupled with unsteady 3D simulations, but the increase in computational resources has nowadays made such computations possible. Nevertheless, it is still necessary to reduce the cost of these coupled simulations to make them more and more affordable.

For this purpose, different strategies have been proposed in the last years. One alternative to reduce conventional Monte Carlo convergence time and large memory requirement is the Reciprocal Monte Carlo approach proposed by Walters and Buckius [8], where the net power exchanged between two cells is directly calculated, fulfilling the reciprocity principle. The main interest of such a reciprocal approach is that the net power exchanged between two cells at the same temperature is rigorously null. This property is only statistically verified by the FM [9]. Cherkaoui et al. [10] reported that the reciprocal method converges at least two orders faster than the conventional Monte-Carlo method and was much less sensitive to optical thickness.

But a complete Monte Carlo Reciprocity Method, based on complete calculation of exchange powers between all the couples of cells of the discretization, is not realistic for system involving participating gases characterized by spectral radiative properties in complex geometrical configurations.

Among the reciprocal Monte Carlo methods, the Emission Reciprocity Method (ERM) developed by Tesse et al. [9] proposes a deterministic estimation of the local emissive power while the local absorbption is estimated with the reciprocal principle. Zhang et al [11] proposed a method to improve the efficiency of ERM, through an approach of importance sampling based on a new frequency distribution function that aims to reduce the Monte Carlo variance, accelerating its convergence (Optimized Emission Reciprocity Method OERM).

Another approach, alternative to the variance reduction techniques, is to use a sampling mechanism whose error has a better convergence rate than classical MCM. Using alternative sampling mechanisms for numerical integration is usually referred to as 'Quasi-Monte Carlo' integration [12]. While considered in semi-conductor applications [13], such methods have not been investigated for participating media such as the ones met in combustors.

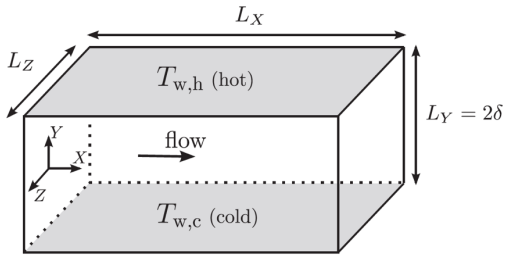

FIGURE 1. Computational domain of channel flow case. x, y and z are, respectively, the streamwise, wall normal and spanwise directions. Periodic boundary conditions are applied along x and z. δ is the channel half-width, equals to 0.01 m and the dimensions of the channel case L_x,L_y and L_z are $2\pi\delta$, 2δ and $\pi\delta$. The lower wall is at 950 k and the upper wall is at 2050 K.

This present study focuses on convergence acceleration of MC simulations: first the interest of ERM will be highlighted, then it will be compared to its optimized version (OERM). OERM is then combined with the Quasi-Monte Carlo method. MC and QMC methods will be assessed in terms of accuracy and computational cost in two configurations. The first configuration is a turbulent channel flow DNS (case C3R1 from [5]) characterized by a simple geometry that allows to perform simulations on a structured grid. The channel characteristics are showed in the Fig. 1: a homogeneous non-reacting $CO_2-H_2O-N_2$ gaseous mixture, at 40 bars, flowing between two walls with imposed temperature values (Fig. 2) and its computational domain is made of 4.2 millions of grid points. The second configuration is a laboratory scale burner [14, 15] computed in LES [16, 17] with an unstructured grid of 8 millions cells and 1.26 millions points. The burner hosts a turbulent premixed flame of a methane-air mixture injected through a swirl injector and confined by cold walls. An instantaneous field of temperature into the chamber is showed in Fig. 3 For both configurations, instantaneous snapshots of unsteady 3d simulations (DNS for the first one, LES for the second one) are used to asses the computational efficiency of the considered Monte Carlo methods.

RADIATION SIMULATIONS WITH RECIPROCAL MONTE CARLO ERM

The general organization of the radiation model, based on a reciprocal Monte Carlo approach, has been detailed by Tess et al. [9]. The principles of this method are briefly summarized here; in this approach the radiation computational domain is discretized into N_v and N_f isothermal finite cells of volume V_i and faces of area S_i, respectively. The radiative power of the node

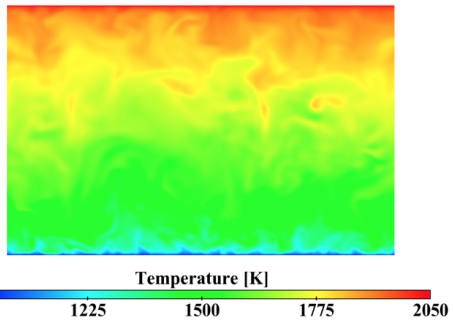

FIGURE 2. Instantaneous fields of temperature on a longitudinal section of the channel.

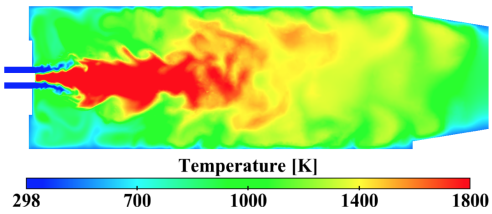

FIGURE 3. 2D slice of the instantaneous 3D field of temperature of the studied burner.

i per unit volume is written as the sum of the exchange powers P_{ij}^{exch} between the node i and all the other cells j, i.e.

$$P_i = \sum_{j=1}^{N_v+N_f} P_{ij}^{exch} = -\sum_{j=1}^{N_v+N_f} P_{ji}^{exch}. \quad (1)$$

where P_{ij}^{exch} is given by

$$P_{ij}^{exch} = \int_0^{+\infty} \kappa_\nu(T_i)\left[I_\nu^\circ(T_j) - I_\nu^\circ(T_i)\right] \int_{4\pi} A_{ij\nu} d\Omega_i d\nu, \quad (2)$$

where $I_\nu^\circ(T)$ is the equilibrium spectral intensity and $\kappa_\nu(T_i)$ the spectral absorption coefficient relative to the cell i. $d\Omega$ is an elementary solid angle. $A_{ij\nu}$ accounts for all the paths between emission from the node i and absorption in any point of the cell j, after transmission, scattering and possible wall reflections along the paths. Its expression is detailed in [9].
As in a Monte Carlo method propagation direction $\Delta(\theta,\phi)$ and

wave-number ν of the photon bundles emitted are determinated randomly according to a Probability Density Function (PDF) $f_i(\Delta(\theta,\phi),\nu)$, that will be written as $f_i(\Delta,\nu)$, introducing the emitted power $P_i^e(T_i)$ per unit volume, Eq. (2) can be written as

$$P_{ij}^{exch} = P_i^e(T_i)\int_0^{+\infty}\left[\frac{I_\nu^\circ(T_j)}{I_\nu^\circ(T_i)}-1\right]\int_{4\pi}A_{ij\nu}f_i(\Delta,\nu)d\Omega_i d\nu, \quad (3)$$

where the PDF is expressed as

$$\begin{aligned} f_i(\Delta,\nu)d\Omega_i d\nu &= f_{\Delta i}(\Delta)d\Omega_i f_{\nu i}(\nu)d\nu \\ &= \frac{1}{4\pi}d\Omega_i\frac{\kappa_\nu(T_i)I_\nu^\circ(T_i)}{\int_0^{+\infty}\kappa_\nu(T_i)I_\nu^\circ(T_i)}d\nu. \end{aligned} \quad (4)$$

As in this method the emitted energy is calculated in a deterministic way while the absorbed one is computed by using a statistical approach, the accuracy of computed emitted energy will be more accurate than the absorbed energy and hence ERM is more adapted to the zone where emission is dominant than absorption, i.e. high temperature zone [11].
As in the ERM only the bundles leaving the node *i* are needed to estimate the local radiative power. It is possible to estimate the radiative power at one point without performing such estimation in all other points of the domain. This main feature allows an estimation of the radiative power in reduced parts of the domain and it gives the possibility to have a control on the local accuracy.

Scalability

Scalability becomes a very challenging problem in large-scale simulations involving radiative transfer. Fluid mechanics and most other phenomena in combustion physics are short range phenomena, so the energy balance equations can be solved over infinitesimal volumes, making them amenable to domain decomposition. Conversely, radiation is a long-distance phenomenon and corresponding equations must be solved over the entire considered domain, thus creating difficulties for domain decomposition. Each node of the domain needs information about all other nodes, so each processor shares radiation field variables with all other processors. Achieving a high level of scalability with the conventional forward Monte Carlo method is not straightforward. Moreover scalability in massively-parallel computing is difficult to obtain due to load imbalancing and interprocessor communication demands. The feature of the ERM method to treat each mesh point independently from the others with a local monitoring of the statistical error insures a high degree of scalability. The RAINIER code used for the simulations presented in this paper solves the radiative transfer equation in order to determine the fields of radiative power and radiative heat fluxes to walls. It is characterized by a master/slave framework. The master process assigns work to all of the other processes, called slaves and the exchange of information occurs through MPI commands. The master, then, collects and saves the results as they are returned from the slaves. As each slave process completes the assigned work, it requests additional work to the master process. To exhibit the computational demand of the ERM method for different cores counts, a scalability analysis was performed on a Bull cluster equipped with Intel E5-2690 processors. The case retained for the scalability test is the laboratory scale burner whose computational domain is made of 8 millions cells. Tests have been conducted on a range of cores, from 120 up to 1920. Two tests have been performed with a fixed number of rays emitted in each point of the domain (200 for the first case, 1000 for the second one), and no convergence criteria have been imposed. The test characterized by a lower number of emitted rays presents some disadvantageous conditions to scalability, as a huge number of communications between slaves and master is required. The results of the scalability analysis are summarized in Fig. 4 where it can be noticed a perfect ability of the method to require less wall-clock time as the number of processors is increased up to 1000 cores. When the cores number is higher than 1000, the time to exchange the informations between the master process and the slaves improves. Consequently the case with a lower number of rays prevents to achieve good scalability at larger process counts because of an overload of the master, which increases in proportion to the number of processes used. On the contrary, when the load of the slaves grows, a linear scalability is accomplished up to 1920 cores, with no deviation from the ideal scalability curve, meaning that the scalability limit is not reached. This trend let us expect that strong scalability will continue further, for a larger number of processors.

The efficiency plot in fig. 5 confirms the code performance. In the case characterized by the overloading of the master, the efficiency decreases to 70 % for a large number of cores, while, for the second case, it remains close to 100% whatever the number of cores.

Local Convergence

As already mentioned, one of the main interests of the ERM method is the possibility to control the local convergence. To show this feature, instantaneous snapshots of unsteady 3D DNS simulations of the turbulent channel flow, defined in fig. 1, are used to solve the radiation field.

To estimate the local standard deviation the actual total number of optical paths (N) is divided into n packages with N/n optical paths for each package. In order to evaluate the convergence of a Monte Carlo solution, the control is done on the relative and the absolute value of the standard deviation. The relative standard deviation is the ratio of the local standard deviation to the local radiative power. However, this parameter

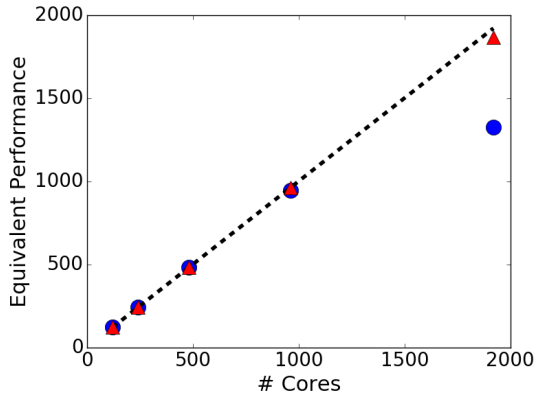

FIGURE 4. Scalability plot. Blue circles: test performed with 200 rays; red triangles: test performed with 1000 rays; dashed line: ideal curve.

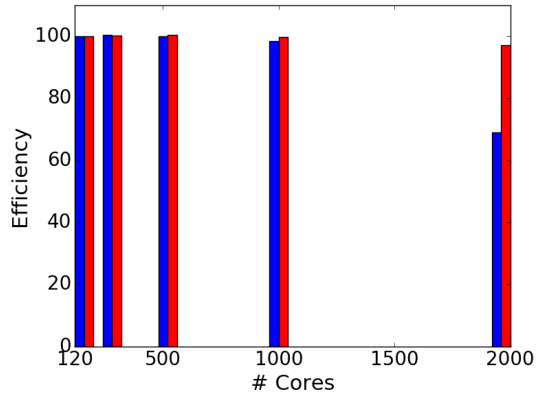

FIGURE 5. Efficiency bar chart. Blue: test performed with 200 rays; red: test performed with 1000 rays.

is not enough as there can be some regions, such as the injector of a combustion chamber, where there are no participating gases, and the radiative power is zero. Therefore the absolute value of the local standard deviation, is checked to be lower than a prescribed maximum.

The ERM method is simulated in two different cases: in the first one a given number of realizations, or optical paths, is imposed to be the same for all the nodes of the domain; in the second one a local convergence criterion is imposed. For all the simulations

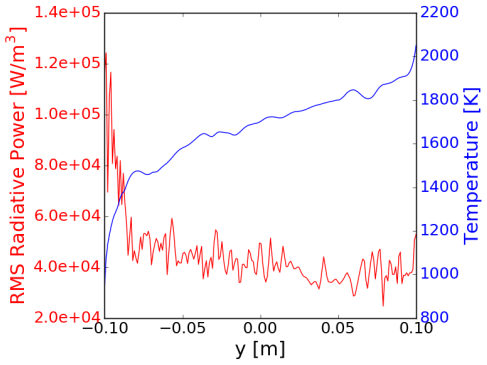

FIGURE 6. Field of RMS of radiative power on a transversal section of the channel (top). Plot of RMS of radiative power (red) and temperature (blue) on the same section obtained with the Monte Carlo ERM in fixed rays number tests (bottom).

the gases spectral properties are computed using the correlated κ-distribution [18].

Case 1: Simulations with a fixed rays number In this test, the rays number is imposed to 10,000 for all the computed points. To evaluate the achieved level of convergence, it can be interesting to take a look at the standard deviation of the radiative power. This variable, together with the temperature, is plotted over the y-axis of the channel in Fig. 6, showing that a better accuracy is reached in the region near the hot wall of the channel, while high values of rms radiative power are encountered in the colder regions of the channel.

Case 2: Simulations with imposed convergence criteria This test case is set-up in such a way that calculations are performed until the relative criterion is lower than 5% or the locale absolute value of the standard deviation is lower than 10% of the maximum value of the mean radiative power. Here the number of rays generated from each cell is not anymore set a priori, but it varies spatially according to the local standard deviation. The local convergence controlling algorithm makes possible to relate the local standard deviation to the local number of optical paths: the fig. 7 shows that in regions where the convergence is difficult to achieve, more optical paths are provided, or equivalently that the regions characterized by a number of shots lower then the maximum, have achieved the convergence.

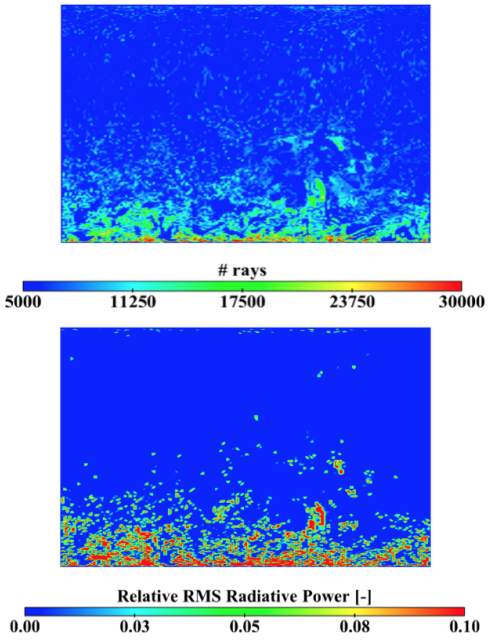

FIGURE 7. Number of rays (top) and relative standard deviation (bottom) obtained with the Monte Carlo ERM in controlled convergence.

To conclude it can be confirmed that the radiative power field predicted by ERM is less hard to converge in high temperature regions where the accuracy is bigger. The reason of the different behavior in hot and cold zones lies in the frequency distribution function used in ERM, as it is based on the spectral emitted power. Consequently the optical paths issued from colder cells are characterized by low frequencies. But the radiative power absorbed by a cold cell has mainly be emitted by hot regions, emitting at much higher frequencies. The absorbed radiative power is then strongly underestimated in cold regions. This phenomenon does not appear for hot cells as the emitted radiation spectrum is very close to the absorbed one [11]. These considerations clearly show that the distribution function used in the ERM method may not be optimized for fast convergence in the cold regions, leading to excessive CPU time.

MONTE CARLO OERM

To alleviate the mentioned problem different methods exist, one of the most important ones is the so-called importance sampling: a variance reduction method to accelerate Monte Carlo convergence. This is the core of the Optimized Emission-based Reciprocity Method (OERM) [11], where the frequency distribution function is chosen in such a way as to correct the ERM drawback and decrease the variance.
In the OERM method the frequency distribution function, $f_\nu(\nu, T_{max})$, is based on the emission distribution at the maximum temperature encountered in the system and it is expressed as

$$f_\nu(\nu, T_{max}) = \frac{\kappa_\nu(T_{max}) I^\circ_\nu(T_{max})}{\int_0^{+\infty} \kappa_\nu(T_{max}) I^\circ_\nu(T_{max}) d\nu}. \tag{5}$$

In these conditions, the radiative exchange power for unit volume between i and j, given by (2), can be expressed as

$$P^{exch}_{ij} = P^e_i(T_{max}) \int_0^{+\infty} \frac{I^\circ_\nu(T_i)}{I^\circ_\nu(T_{max})} \frac{\kappa_\nu(T_i)}{\kappa_\nu(T_{Imax})} \left[\frac{I^\circ_\nu(T_j)}{I^\circ_\nu(T_i)} - 1\right] f_\nu(\nu, T_{max}) d\nu f_{\Omega_i} d_{\Omega_i} \tag{6}$$

The use of the pdf (5) allows to eliminate the disadvantage of the classical approaches of ERM in the cold regions. To illustrate the advantages of the OERM method, computations of the radiative transfer in the channel flow are performed. In a first step solutions of radiative field obtained with a OERM approach are obtained with the same computation conditions of the case 1, at imposed number of rays, and they are compared to the solutions obtained with the ERM method. The standard deviation for both of the methods is exhibited in Fig. 8: on the hot wall results of ERM and OERM overlap as the two frequency distribution functions are practically identical, therefore OERM turns into ERM. Focalizing on the colder regions on the bottom of the section, the same figure shows that the standard deviation in the OERM case is much lower than in the ERM case, meaning that with the same number of realizations, the frequency distribution function of OERM allows the absorption by the cold regions to be more accurately computed, contrary to the case of ERM.
Consequently if a convergence criterion is fixed, calculations conducted with an OERM method need a lower number of realizations to satisfy the same criterion as it can be seen in fig. 9, leading to a less expensive computational cost.

The OERM method is now investigated on a semi-industrial configuration, the burner of fig. 3. The temperature, pressure, CO_2 and H_2O molar fractions used in OERM simulations are instantaneous values extracted from unsteady 3D Large Eddy Simulations of the flow. As seen in Fig. 10 most of the domain emits energy through radiative heat transfer (negative radiative power); the regions where energy absorption dominates (positive radiative power) are the coldest gas pockets mainly located in thin layers near the walls.
In the set-up of this case relative and absolute values of standard

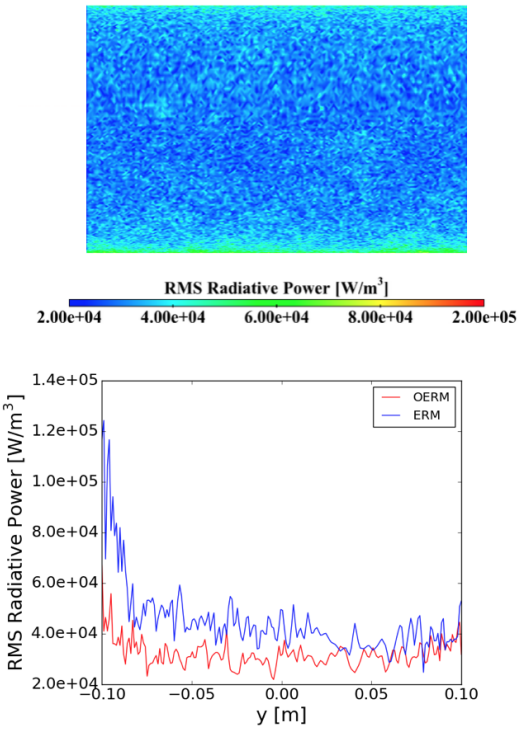

FIGURE 8. Instantaneous field of rms of radiative power obtained with Monte Carlo OERM (top). Plot of the rms of radiative power for ERM (blue line) and OERM (red line) in test with fixed rays number.

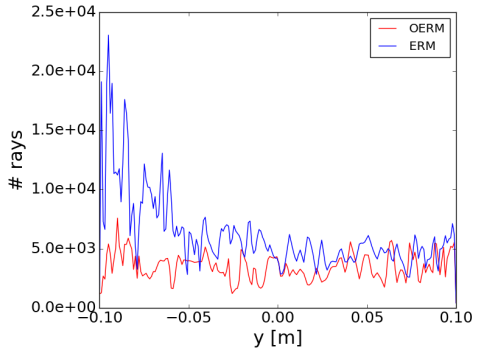

FIGURE 9. Plot of the number of rays needed for the ERM (blue line) and OERM (red line) in controlled convergence.

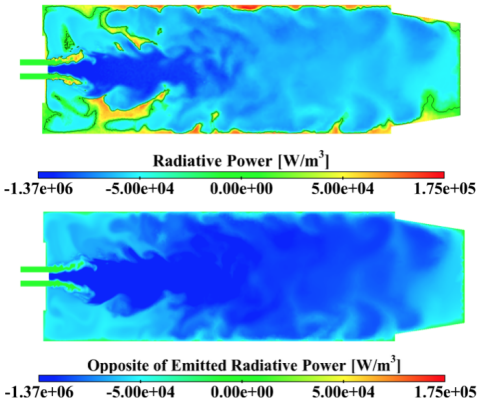

FIGURE 10. Instantaneous fields of Radiative Power (top) and the opposite of the emitted power (bottom). Black line is the iso-contour for radiative power = 0.

deviation are controlled in order to insure that the simulation ends up in a limited CPU time, so a maximum number of rays emitted per point is imposed. If the simulation is locally stopped because of this criterion, the convergence is not achieved in these points. Tests are performed limiting the maximum number of possible optical paths departing from the nodes to 10 000 and 20 packages of 500 realizations each are taken into account for the error estimation. The convergence condition of the Monte Carlo algorithm is that of an rms lower than 3 % of the mean value; while in the regions where the criterion of relative rms is never satisfied, a control on the absolute value of the rms, whose value is imposed at 3 % of the maximum value of the mean radiative power, is done. In the fig. 11 gray zones are the ones where the absolute criterion is respected, keeping in mind that in these zones the rms of the radiative power is close to zero, while in the remaining part of the chamber a control on the relative error is done. It can be seen that zones where it is most difficult attain the established convergence criterion are characterized by a larger number of realizations.

QUASI MONTE CARLO

If the technique used in the OERM method is aimed to reduce the variance through importance sampling; another approach to improve the Monte Carlo error is to replace the pure random sampling with a quasi-random (also called low-discrepancy) sampling, without modifying the frequency distribution function. Lower error and improved convergence may be

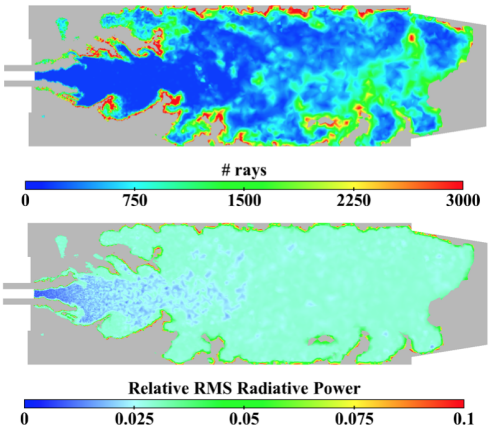

FIGURE 11. Number of rays (top) and relative rms of radiative power (bottom) obtained using the Monte Carlo OERM method.

attained by replacing the pseudo-random sequences using low-discrepancy sequences, whose points are distributed in a way to provide greater uniformity. For this study a Sobol sequence has been used and its construction uses results from [19]. Using this alternative sampling method in the context of multivariate integration is usually referred to as Quasi Monte-Carlo, that can be seen like a deterministic version of Monte Carlo method.
Its advantage lies in enhancing the convergence rate [20]. It is possible to asses the error using a Randomized Quasi-Monte Carlo [12]. In the context of radiation simulations, as for the Monte Carlo, n packages are considered; within each of this package, a low discrepancy sequence of N/n points is used, while the n sequences of the packages are randomized using an I-binomial scrambling [21]. This approach allows to benefit from the faster convergence rate of Quasi-Monte Carlo within each package and to have an estimation of the error using the variance between the packages, as it is done for the Monte Carlo method. The obtained Quasi-Monte Carlo method can be combined with both ERM and OERM methods, so that a comparison with the Monte Carlo simulations, previously presented, can be conducted. Only OERM results are considered in the following.

Quasi Monte Carlo combined with OERM method

Simulations with a Quasi-Monte Carlo method in its OERM version have been conducted on snapshots of 3D LES of the laboratory scale burner. In a first step, simulations have been carried out setting the same number of optical paths departing from all the nodes of the domain, without imposing convergence criteria. Such an analysis allows to evaluate the accuracy of both methods.

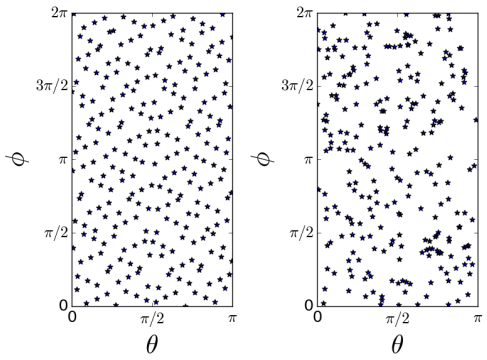

FIGURE 12. Sampling of polar (θ) and azimuthal angle (ϕ) using a Sobol sequence (left) and a random sequence (right).

In fig. 13 the relative standard deviation for both the methods is shown on the whole longitudinal section of the chamber. It can be seen that with the same number of realizations, QMC simulations are more accurate than MC ones, as the relative error is much lower on the whole domain, even in the zones more difficult to converge, such as the ones closed to the cold walls of the chamber.
In order to compare the convergence rate of Monte Carlo and Quasi Monte Carlo simulations, in a second step tests of convergence are performed. Their set-up is the same of OERM simulations of the previous chapter, in terms of maximum number of rays and packages, and parameters for the control error. Tests withs local convergence control allow to highlight the advantage of QMC in terms of computational cost. As expected, the number of realizations necessary to respect the convergence criterion is much lower in the case of QMC simulations as showed in fig. 14.

CPU efficiency of Monte Carlo and Quasi-Monte Carlo methods

A more complete comparison can be done evaluating the efficiency of both Monte Carlo and Quasi Monte Carlo methods. The local efficiency of both the methods has been compared and evaluated as

$$\eta_i = \frac{1}{\sigma_i^2 \cdot nb_{int,i} \cdot (T_{CPU}/nb_{int,tot})} \tag{7}$$

where i represents the considered point, $nb_{int,i}$ is the number of the intersections of the point i, $T_{CPU}/nb_{int,tot}$ is the cost of an intersection. In the fig. 15 the ratio of the local efficiencies of quasi Monte Carlo algorithm and Monte Carlo is showed on a longitudinal plane of the chamber: the ratio is bigger than 1 on almost

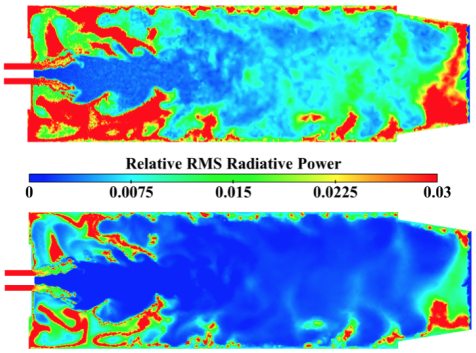

FIGURE 13. Instantaneous field of rms of radiative power obtained with Monte Carlo OERM (top) and Quasi-Monte Carlo OERM (bottom) at imposed number of rays.

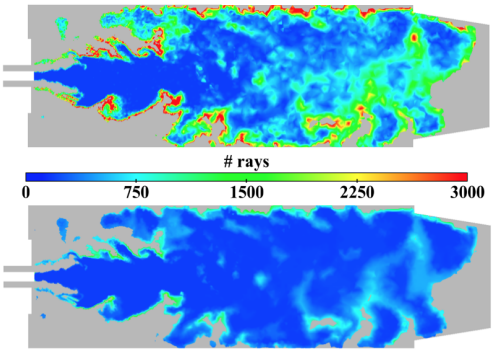

FIGURE 14. Number of rays necessary for the convergence used by Monte Carlo OERM (top) and Quasi-Monte Carlo OERM (bottom) in controlled convergence.

the whole domain, meaning that the QMC method improves the efficiency of the MC, by a value that can be greater than 5, depending on the considered points of the domain.

In order to localize the regions where Quasi Monte Carlo becomes more efficient, it is interesting to look at the scatter plot of the efficiency ratio in relation to the temperature for all the domain points and it is shown in Fig. 16. It is worth noting that this ratio is high in the cold pockets of the chamber near the walls, where normally the convergence is hard to be achieved, and that the regions characterized by a higher efficiency ratio are the ones at intermediate temperature (around 1000 K), which cover most of the domain.

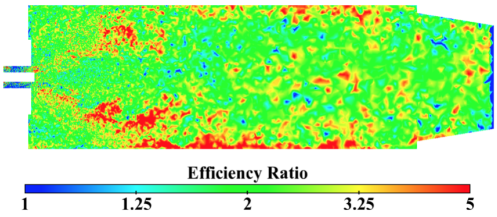

FIGURE 15. 2D map of the ratio between efficiency of Quasi-Monte Carlo and Monte Carlo methods.

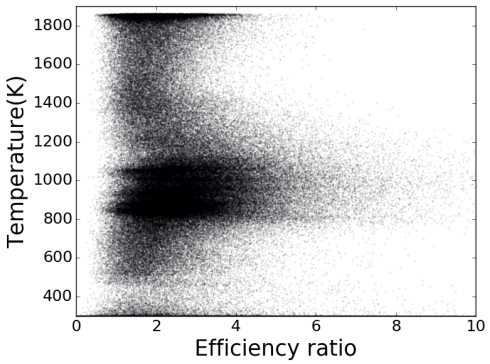

FIGURE 16. Scatter plot of temperature in relation to the efficiency ratio between Quasi-Monte Carlo and Monte Carlo for all the points of the domain.

CONCLUSION

Monte Carlo methods applied to radiative heat transfer problems are known for being computationally expensive. In order to afford coupled 3D simulations of reactive flows, it is necessary to reduce the computational cost. Different strategies have been proposed to face this limit, some of them, like the ERM or the OERM methods, have been used in this study. Finally a technique to further improve the efficiency of Monte Carlo method, based on a low-discrepancy sampling, has been applied and the obtained quasi-Monte Carlo method has been combined with OERM and compared to the Monte Carlo in a complex configuration. Simulations results have shown a significant improvement from the quasi-Monte Carlo in terms of computational efficiency, introducing them as an excellent candidate for coupled high-fidelity simulations.

ACKNOWLEDGMENT

This project has received funding from the European Unions Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 643134. It was also granted access to the HPC resources of CINES under the allocation 2016-020164 made by GENCI.

REFERENCES

[1] Coelho, P. J., 2007. "Numerical simulation of the interaction between turbulence and radiation in reactive flows". *Progress in Energy and Combustion Science,* ***33***, pp. 311–383.

[2] Coelho, P. J., 2012. "Turbulence-Radiation Interaction: From Theory to Application in Numerical Simulations". *Journal of Heat Transfer-Transactions of the ASME,* ***134***(3).

[3] Deshmukh, K. V., Modest, M. F., and Haworth, D. C., 2008. "Direct numerical simulation of turbulence-radiation interactions in a statistically one-dimensional nonpremixed system". *JOURNAL OF QUANTITATIVE SPECTROSCOPY & RADIATIVE TRANSFER,* ***109***(14), SEP, pp. 2391–2400.

[4] Deshmukh, K. V., Haworth, D. C., and Modest, M. F., 2007. "Direct numerical simulation of turbulence-radiation interactions in homogeneous nonpremixed combustion systems". *Proceedings of the Combustion Institute,* ***31***(1), pp. 1641–1648.

[5] Zhang, Y., Vicquelin, R., Gicquel, O., and Taine, J., 2013. "Physical study of radiation effects on the boundary layer structure in a turbulent channel flow". *International Journal of Heat and Mass Transfer,* ***61***, pp. 654–666.

[6] Soucasse, L., Riviere, P., and Soufiani, A., 2014. "Subgrid-scale model for radiative transfer in turbulent participating media". *Journal of Computational Physics,* ***257***(A), pp. 442–459.

[7] Gupta, A., Haworth, D., and Modest, M., 2013. "Turbulence-radiation interactions in large-eddy simulations of luminous and nonluminous nonpremixed flames". *Proceedings of the Combustion Institute,* ***34***(1), pp. 1281 – 1288.

[8] Walters, D. V., and Buckius, R. O., 1992. "Rigorous development for radiation heat transfer in nonhomogeneous absorbing, emitting and scattering media". *International Journal of Heat and Mass Transfer,* ***35***(12), pp. 3323 – 3333.

[9] Tessé, L., Dupoirieux, F., Zamuner, B., and Taine, J., 2002. "Radiative transfer in real gases using reciprocal and forward monte carlo methods and a correlated-k approach". *International Journal of Heat and Mass Transfer,* ***45***(13), pp. 2797–2814.

[10] Cherkaoui, M., Dufresne, J.-L., Fournier, R., Grandpeix, J.-Y., and Lahellec, A., 1996. "Monte carlo simulation of radiation in gases with a narrow-band model and a net-exchange formulation". *Journal of Heat Transfer,* ***118***, pp. 401–407.

[11] Zhang, Y., Gicquel, O., and Taine, J., 2012. "Optimized emission-based reciprocity monte carlo method to speed up computation in complex systems". *International Journal of Heat and Mass Transfer,* ***55***(25–26), pp. 8172 – 8177.

[12] Lemieux, C., 2009. *Monte carlo and quasi-monte carlo sampling*. Springer Science & Business Media.

[13] Kersch, A., Morokoff, W., and Schuster, A., 1994. "Radiative heat transfer with quasi-monte carlo methods". *Transport Theory and Statistical Physics,* ***23***(7), 09, pp. 1001–1021.

[14] Guiberti, T. F., Durox, D., Zimmer, L., and Schuller, T., 2015. "Analysis of topology transitions of swirl flames interacting with the combustor side wall". *Combustion and Flame,* ***162***(11), pp. 4342–4357.

[15] Guiberti, T., Durox, D., Scouflaire, P., and Schuller, T., 2015. "Impact of heat loss and hydrogen enrichment on the shape of confined swirling flames". *Proceedings of the Combustion Institute,* ***35***(2), pp. 1385–1392.

[16] Mercier, R., Guiberti, T., Chatelier, A., Durox, D., Gicquel, O., Darabiha, N., Schuller, T., and Fiorina, B., 2016. "Experimental and numerical investigation of the influence of thermal boundary conditions on premixed swirling flame stabilization". *Combustion and Flame,* ***171***, pp. 42–58.

[17] Koren, C., Vicquelin, R., and Gicquel, O., 2017. "High-fidelity multiphysics simulation of a confined premixed swirling flame combining large-eddy simulation, wall heat conduction and radiative energy transfer". *ASME Turbo EXPO 2017 (Submitted)*.

[18] Taine, J., and Soufiani, A., 1999. "Gas radiative properties: From spectroscopic data to approximate models". Vol. 33 of *Advances in Heat Transfer*. Elsevier, pp. 295 – 414.

[19] Joe, S., and Kuo, F. Y., 2008. "Constructing sobol sequences with better two-dimensional projections". *SIAM Journal on Scientific Computing,* ***30***(5), pp. 2635–2654.

[20] Hlawka, E., 1961. "Funktionen von beschränkter variatiou in der theorie der gleichverteilung". *Annali di Matematica Pura ed Applicata,* ***54***(1), pp. 325–333.

[21] Tezuka, S., and Faure, H., 2003. "I-binomial scrambling of digital nets and sequences". *Journal of complexity,* ***19***(6), pp. 744–757.

Annexe C

Communication à l'ICMF

Uncertainty quantification of injected droplet size in mono-dispersed Eulerian simulations

Théa Lancien[1], Nicolas Dumont[1], Kevin Prieur[1,2], Daniel Durox[1], Sébastien Candel[1], Olivier Gicquel[1] and Ronan Vicquelin[1] *

[1] *Laboratoire EM2C, CNRS, CentraleSupélec, Université Paris-Saclay, Grande Voie des Vignes, 92295 Chatenay-Malabry cedex, France*

[2] *Safran Tech, E&P, Rue des Jeunes Bois, Chateaufort, CS 80112, 78772 Magny-Les-Hameaux, France*

Abstract

Large-eddy simulations (LES) of a laboratory-scale two-phase burner are considered by describing the disperse liquid spray with a mono-disperse Eulerian approach. In this simplified framework, the choice of the size of the injected droplets becomes a critical issue. The impact of this key parameter upon the numerical results is carefully assessed though uncertainty quantification tools. Using Polynomial Chaos Expansion and Clenshaw-Curtis nested quadrature rule, several LES are performed for different injected droplet sizes in order to obtain a response surface of velocity and diameter fields at any point in the computational domain as a function of the injected one. Post-treatment of the response surface gives access to the precise impact of the chosen injected droplet size on the results. It is shown that information obtained from different mono-disperse simulations enables to answer a couple of practical questions in such two-phase flow simulations: How can the mono-disperse simulations be compared to the poly-disperse experimental results and their accuracy evaluated? More importantly, if only one simulation is to be carried out for a larger case, which value of the injected droplet size is the best?

Keywords: Two-phase flows, Polynomial Chaos Expansion, Large Eddy Simulation, Eulerian-Eulerian approach

1. Introduction

Reliable and easy ignition constitutes a central issue in the design of practical combustion systems and is specifically important in the case of gas turbines and aero-engines. In the last devices, one complication results from the presence of multiple injection units and a successful ignition process follows then three main stages [14]: (1) A spark is produced by an igniter and leads to the formation of a kernel of hot gases; (2) If the kernel's size and temperature enable its spreading, the corresponding volume increases in a second phase until it reaches the closest fuel injector, thus establishing an initial flame; (3) In the final stage, the flame propagates from burner to burner until a flame is stabilized around each injector. Due to the geometrical complexity of the chamber with its many injectors, this last propagation step, also called light-round, has been less well investigated in the literature than the first two stages. Clearly, a detailed understanding of the ignition process in such systems may be gained by combining well controlled experiments with calculations based on advanced large eddy simulations. One pioneering demonstration of the feasibility of full scale calculations of the light-round in a generic helicopter gas turbine combustor is reported by Boileau et al. [4] but without detailed comparisons with experiments. Detailed ignition experiments in a laboratory-cscale annular combustor (MICCA) operating under premixed conditions [5] have recently provided high speed visualizations of the flame spreading process together with systematic measurements of the ignition delay. Large eddy simulations of this configuration have been successfully compared with further experimental data in [21, 20]. The effect of spacing between injectors was studied in [3] on a linear five-burner configuration. Further experiments under perfectly premixed conditions and gaseous non-premixed conditions are also reported in [16].

However, with the exception of [4], all previous studies were carried out with gaseous flows while aero-engine combustors operate with liquid fuel injected as a spray. Accounting for the presence of the spray of fuel droplets is clearly necessary to be truly representative of ignition in aeronautical gas turbines. It is also worth noting that new data are available for liquid spray injection in the MICCA annular configuration [23] and that these data could be used in comparisons with detailed simulations.

The final objective of this research is to develop such simulations and compare results with these data. The present article is intended to define modeling aspects and it specifically considers questions linked to the modeling of the disperse phase.

Envisioning such large scale simulations with liquid spray injection gives rise to several modeling issues: (i) One is first faced with the problem of describing the spray atomization process. This cannot be included in such calculations and the disperse liquid phase injected in the simulations needs to be modeled; (ii) Second, a compromise has to be found for the description of the polydisperse droplet mist to be consistent with the available computational resources. This issue is considered in what follows. Among the many possible representations, two approaches have been explored in large eddy simulation of two-phase reacting flows. Both rely on a mesoscopic point-particle approximation [9]: The Eulerian-Eulerian approach where moments (see also quadrature-based methods [8]) of the number density function (NDF) are transported [27, 25] and the Eulerian-Lagrangian in which a large ensemble of particles is transported [15, 6, 11]. While accounting for polydispersity in the Eulerian-Lagrangian framework is straightforward, the Eulerian-Eulerian approach requires additional transport equations for moments and/or classes of particle sizes [17, 28]. On the other hand, one of the drawbacks of the Lagrangian methods is the complex handling of computational load balancing on parallel machines for large scale simulations [10]. This issue is expected to become even more problematic in the envisioned light-round simulations where the balance of computational load between ignited and non-ignited injectors needs to be treated dynamically. In order to find the best com-

* This work was supported by grant ANR14-CE23-0009-01 of the French Agence Nationale de la Recherche.

promise between cost and accuracy, under constraints of limited computational resources, the present study explores the possible use of a mono-disperse Eulerian-Eulerian representation of a polydisperse spray. An original feature of the method proposed in this article is that the optimum value of the mono-disperse injected droplets diameter that best represents the evolution of the spray is deduced by computing a surface response on mono-disperse Eulerian simulations thanks to uncertainty quantification (UQ) methodology. This analysis is carried out on a single injector configuration and its result will enable future light-round simulations with a controlled accuracy of the injected droplet diameter.

The single burner investigated experimentally and numerically is presented is section 2. This section also discusses mono-disperse simulation using the spray's Sauter Mean Diameter (SMD). The study of the influence of the injected diameter is then carried out in the uncertainty quantification perspective as detailed in section 3.

2. Mono-disperse Eulerian simulation with the spray Sauter Mean Diameter

2.1. Experimental setup

The experimental burner studied is displayed in Fig. 1. The air injected in the system flows through a swirl injector ("G" arrows on the figure) before meeting the n-heptane liquid injector, a simplex atomizer located with a 6 mm recess from the convergent exhaust. The two-phase flow exits the burner into the atmosphere through a diameter $d = 8$ mm with a measured swirl number of 0.68. A Phase Doppler Anemometry system (PDA) is used to measure the gas and droplets velocity profiles as well as the droplets diameter. Measurement were carried out at different distances from the injector exhaust section. The studied operating conditions correspond to air mass flow rate 1.84 g/s and liquid n-heptane mass flow rate 0.11 g/s.

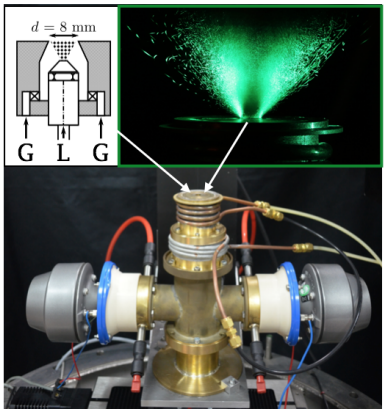

Figure 1: Experimental burner. A sketch of the swirler appears in the top left-hand corner and a vertical tomography of the spray is shown in the top right-hand corner.

Top-right corner Fig.1 shows a tomographic slice of the droplet spray, visualized by means of an argon-ion laser at 514.5 nm. The hollow cone shape of the flow appears clearly, with an inner recirculation zone where few droplets are present. The droplet diameter repartition measured at one point located in the spray (at the radius $r = 4.5$ mm) 2.5 mm above the exhaust plane is presented in Fig. 2. The number distribution spans from $d_l = 0.5$ μm to $d_l = 35$ μm, which shows the polydisperse nature of the spray. The diameter interval corresponds to a range of Stokes number $\tau_p d/U$ of [0, 0.45] where $\tau_p = (1 + 0.15Re_p^{0.687})\frac{\rho_l d_l^2}{18\mu}$ is the droplet's drag characteristic time given the droplet diameter d_l. The evaporation time τ_e expressed in terms of $\tau_e d/U$ corresponds to the range [0.08; 300], which shows that evaporation is reduced for most droplets.

The mean diameter $D_{10} = (\sum_N d)/N$ and the Sauter Mean Diameter $D_{32} = (\sum_N d^3)/(\sum_N d^2)$ are indicated for the considered point ($r = 4.5$ mm, $z = 2.5$ mm) in Fig. 2. With $D_{10} = 8$ μm, the spray is mainly formed by small droplets. In the perspective of combustion, the evaporation rate of the droplets in the polydisperse spray is determined by the spray repartitions in mass and surface [14]. For reactive flow simulations, the Sauter Mean Diameter is usually considered for an equivalent mono-disperse spray simulation that has the same volume to surface ratio as the polydisperse spray. In the present case, $D_{32} = 20$ μm.

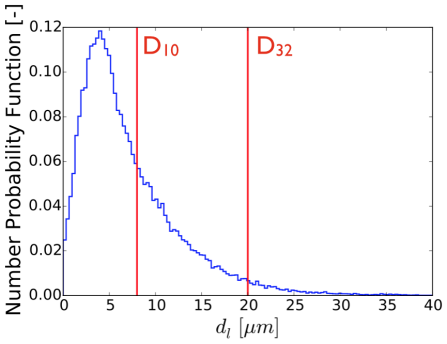

Figure 2: Distribution of droplet diameter in the spray at $x = 2.5$ mm from the exhaust plane and $r = 4.5$ mm.

2.2. Numerical setup

Simulations are carried out with the three-dimensional compressible Navier-Stokes solver AVBP [26], jointly developed by CERFACS and IFP Energies Nouvelles. It is based on a centered scheme and uses a two-step Taylor-Galerkin weighted residual central distribution scheme, third order in time and space (TTGC [7]) for both gaseous and liquid phases. The Wall Adapting Local Eddy model (WALE) [19] describes the sub grid scale turbulence. As motivated in the introduction, the numerical description of the liquid disperse phase is a mono-disperse Eulerian approach. The evaporation of the droplets is represented by using the Abramzon-Sirignano model [1]. The computational domain is displayed in Fig. 3. The air guides are included and the ambient air above the burner exhaust plane is taken into account through a large meshed volume. The boundary conditions are standard Navier-Stokes characteristic boundary conditions (NSCBC) [22] and are set according to the experimental parameters. The gas and liquid temperatures are set to 298 K. Initial results are presented for an injected droplet diameter $d_l^{inj} = D_{32} = 20$ μm. All the walls are considered to be adiabatic in the simulations. The mesh counts 17.5 million cells, which corresponds to 3 million nodes. The regions where the highest velocity gradients are found are refined, as well as the area around the liquid injection to deal with high gradients of volume fraction (see Fig. 3). The atmospheric domain where the spray is observed experimentally is also refined in order to capture its dynamics. The mesh is then progressively coarsened until the limits of the domain. A mesh convergence study has demonstrated the adequacy of the retained discretiza-

tion.

The simulation is first run by only injecting the air to establish the gaseous flow. Mean velocity fields are calculated and validated against experimental data. Fuel droplets are then injected in the flow. It is worth noting that the gaseous solution used to initialize the two-phase flow simulations is the same throughout the whole study. After computing a transient physical time equal to several flow-through times, statistical mean fields are computed. Several consecutive average fields are compared to ensure statistical convergence of the velocity field.

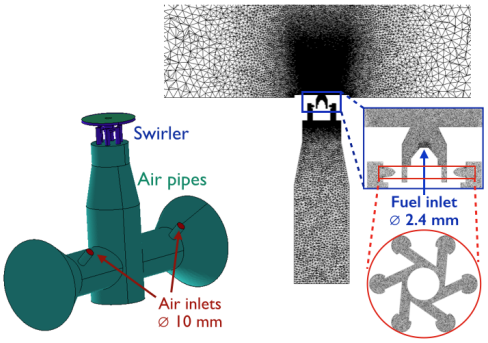

Figure 3: Computational domain for the simulations showing a slice in the mesh. The outer atmospheric domain is not shown.

2.3. *Results*

Gaseous phase velocity

Profiles of the axial and azimutal components of the mean velocity of the gaseous phase are displayed in Fig. 4. The symbols represent the experimental data and the full lines show the numerical results. Data from the purely gaseous case (in black) as well as from the two-phase flow case (in red) are presented. The experimental gaseous flow is nearly identical with or without droplets, which indicates that the diluted disperse liquid phase has little influence on the gaseous phase. For both cases, the simulation is able to retrieve the experimental data with a good accuracy. In particular, the radial position of the recirculation zone is well predicted, and the velocity peak levels are obtained with great accuracy.

Liquid phase velocity

Figure 5 shows the profiles of the mean axial velocity of the liquid phase at an axial distance of $x = 2.5$ mm from the burner exhaust plane. Two different experimental averaged velocities are plotted. The triangles correspond to the arithmetic mean of the droplet's velocity, meaning that every droplet has the same weight in the average, whatever its mass is. The predicted numerical velocity profile does not agree with this type of measurement, which can be explained. On the one hand, as the spray is mainly populated (in number) by small droplets in the experimental setup (see Fig 2), the experimental mean velocity is governed by small droplets dynamics which, having a smaller Stokes number, tend to follow the air flow. On the other hand, the simulated disperse liquid phase corresponds to droplets whose diameter at the injection is equal to the spray's Sauter Mean Diameter $D_{32} = 20\ \mu m$. Because of this larger injected diameter than the majority of droplets, the simulated droplets will have a more ballistic type of trajectory and their velocity relaxes to the gas velocity with a larger characteristic time, which explains the discrepancy between the results.

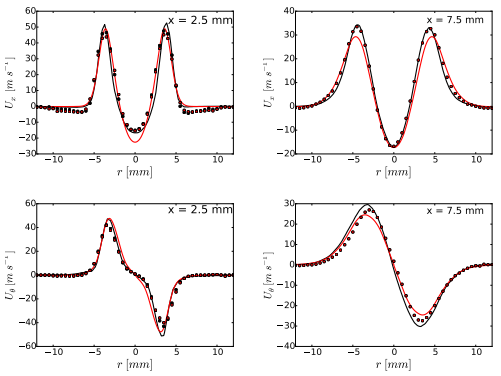

Figure 4: Mean velocity profiles for the gas phase at $x = 2.5$ mm (left) and $x = 7.5$ mm (right) from the exhaust plane: Axial velocity (top) and azimutal velocity (bottom). Black curves represent the results from the gaseous simulation and the red curves the gaseous fields in the two-phase flow simulation. $-$: Numerical results; $\bullet$: Experimental data.

In light of these considerations, the numerical fields are compared to the experimental velocities weighted by each particle's mass. Indeed, the final objective of this study is to represent a burning polydisperse spray with a mono-disperse approach, which means describing accurately the spray mass distribution and momentum. Weighting the droplets velocities by their mass seems therefore appropriate. This field is plotted in red in Fig. 5. The numerical results are then much closer to the experimental mass-averaged velocity than they were to the arithmetic average. As with the gaseous fields, the radial position of the spray is well recovered, as well as the magnitude of the peaks. The central recirculation zone that appears in the axial velocity profile is also satisfactorily captured.

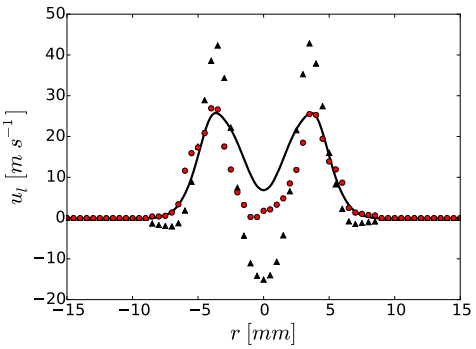

Figure 5: Mean axial velocity for the liquid phase at $x = 2.5$ mm. $-$: Numerical results; ▲: Experimental arithmetic average; •: Experimental mass-weighted average.

Figure 6 shows the radial profiles for the mean liquid axial velocity, included size-conditioned statistics, at $x = 7.5$ mm in the flame stabilization zone. The full line represents the numerical result, while the experimental mass-averaged velocity is represented in symbols. The dotted lines show the experimental

velocities for several classes of droplets. To ensure experimental convergence, only points where enough data (more than 100 droplets) have been collected are plotted. Comparing the different experimental fields, it appears that the different droplet classes behave quite similarly with a spreading due to the different drag relaxation time. This similarity leads to the mass-averaged field being quite close to the arithmetic average (not shown), which was not the case at $x = 2.5$ mm. Indeed, the measurement plane $x = 7.5$ mm being further away from the burner exhaust plane, the bigger droplets have had more time to relax towards a more uniform flow. Figure 4 showed that the gaseous velocity field was very well predicted. However, some discrepancies appear for the liquid phase fields. First of all, only very small droplets are present in the center recirculation zone in the experiment, whereas the numerical 20 μm droplets are not sensitive to the same entrainment effect in the center recirculation zone, which is not retrieved then by the simulation. Looking at the two high velocity peaks indicates that, though the simulations is very accurate in predicting the gaseous peaks as well as the radial position of the spray, it is not able to retrieve the correct magnitude of the average liquid velocity. The larger droplets that are numerically resolved are not sufficiently entrained by the air flow to be representative of the mass-weighted liquid velocity. The numerical profile is however closer to the experimental 20 μm-class velocity (in red), indicating that, while the injected droplets do not behave like the average of the spray, their dynamics matches the corresponding experimental class.

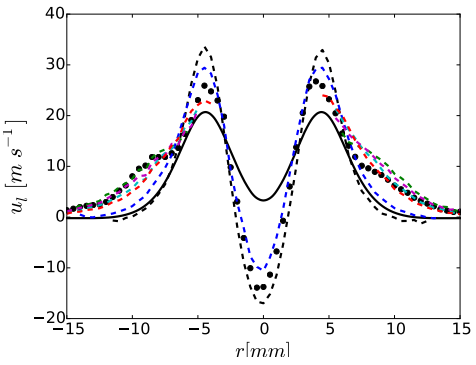

Figure 6: Mean axial velocity field for the liquid phase at $x = 7.5$ mm. $-$: Numerical results; •: Mass-averaged experimental field; Experimental profiles: - -: $d_l = 2-3\ \mu m$; - -: $d_l = 10-12\ \mu m$; - -: $d_l = 20-23\ \mu m$; - -: $d_l = 23-36\ \mu m$; - -: $d_l = 26-30\ \mu m$. - -: $d_l = 30-34\ \mu m$;

Consequently, although the general behavior of the two-phase flow is retrieved qualitatively, the single simulation of a mono-disperse spray with $d_l^{inj} = D_{32}$ is not able to accurately predict the magnitude of the mean velocity fields of a polydisperse spray, which is not surprising. However, if one focuses on the corresponding experimental class of droplet, the simulation is in fair agreement with its evolution. In light of these observations, considering the mono-disperse description as a surrogate model of the polydisperse spray, an optimal injected diameter should be determined in order to retrieve a spray dynamics representative of the experimental one.

3. Uncertainty quantification of the injected droplet size

3.1. Surface response computation

An interesting approach in order to evaluate the impact of one or several parameters on a complex system is uncertainty quantification and the Polynomial Chaos Expansions (PCE) [29, 24]. Indeed, PCE allow to approach uncertain fields that depend on both deterministic and uncertain parameters. A given field u can then be written as $u(x, \omega)$ where x represents the deterministic parameters and ω the uncertain ones. In the present study the injected diameter d_l^{inj} is considered to be the unique uncertain parameter. Through PCE, one is able to estimate any given field with the polynomial decompositon:

$$u(x_j, d_l^{inj}) \approx \sum_{k=0}^{N} a_k(x_j) P_k(d_l^{inj}) \tag{1}$$

For any given point x_j, knowing the value of the coefficients $a_k(x_j)$, $u(x_j, d_l^{inj})$ becomes a continuous function of the uncertain parameter d_l^{inj}, whose study is then straightforward. Using non-intrusive methods, the computation of the coefficients $a_k(x_j)$, which are defined by integrals, is carried out with nested quadrature rules: $M = 2^l + 1$ evaluations of $u(x_j, d_l^{inj})$ are required for the l-quadrature level. In the context of LES, several simulations, corresponding to different values of d_l^{inj} are then performed. The retained Clenshaw-Curtis nested quadrature rule enables to limit the number of evaluations for several quadrature levels [12], which is a great benefit given the computational cost of carrying out several large-eddy simulations. The considered injected diameter distribution is considered uniform between 0.5 μm and 35 μm. Due to the cost of each simulation, the maximum quadrature level was limited to 3 for this study. According to the Clenshaw-Curtis quadrature rule, the different values of d_l^{inj} to simulate are presented in table 1: nine large-eddy simulations have been performed in total.

The PCE is here used to build a response surface of LES results in terms of d_l^{inj}. Each field can therefore be estimated by the polynomial approximation for any value of the injected diameter, even one that was not simulated. This provides a way to determine an optimal diameter more efficiently than by carrying out a parametric study with a finite set of values.

3.2. Analysis of results

Once all the LES statistics have converged and the averaged fields have been recovered, response surfaces can be reconstructed. At a given point in space and for a given physical field, the polynomial reconstruction yields the variation of this field according to the injection diameter. An example of response surface is given in Fig. 7 for the liquid velocity. Figure 8 gives the response curves of the axial liquid velocity at $x = 7.5$ mm for four different values of the radial coordinate r. Polynomial chaos expansions with different truncation level corresponding to different polynomial degrees are shown along with the LES values (symbols) for the nine considered simulations. Good approximation is achieved from polynomial degrees of four and above. Some oscillations appear at the degree 6 (in magenta), which leads to selecting the degree 5 (in cyan) for further analysis. All PCE results correspond to the 3rd quadrature level. The other quadrature levels (lines #1 and #2 in Tab. 1) enable to check the numerical convergence of the presented results.

Table 1: Values of the injected diameter

Quadrature level	Injected diameter [μm]								
1	0.5	-	-	-	17.75	-	-	-	35.0
2	0.5	-	5.55	-	17.75	-	29.95	-	35.0
3	0.5	1.81	5.55	11.15	17.75	24.35	29.95	33.69	35.0

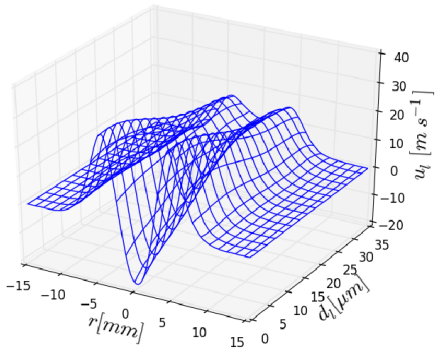

Figure 7: Response surface for the axial velocity at $x = 7.5$ mm and some of the numerical fields used for the expansion.

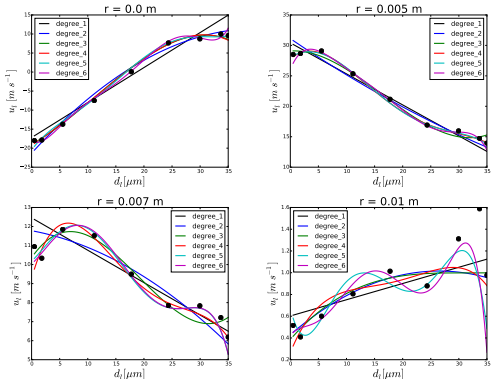

Figure 8: PCE approximation for the axial velocity at $r = 0$ mm (top left), $r = 5$ mm (top right), $r = 7$ mm (bottom left) and $r = 10$ mm (bottom right). $-$: PCE expansions; $\bullet$: Numerical values from mono-disperse simulations.

3.3. Optimization of the injected diameter

The response surfaces can be built for any quantity of interest. In order to determine a best value for the injection diameter, it is necessary to define one or several criteria. A first optimization criterion can be the minimization of numerical error on the mass-weighted axial liquid velocity u_{exp}^{poly}. This error criterion, denoted by $||u_l^{mono}(d_l^{inj}) - u_{l,exp}^{poly}||_2$, is defined here for a given height x as

$$\sqrt{\frac{1}{N_{exp}} \sum_{N_{exp}} (u_l^{pce}(r_j, d_l^{inj}) - u_{exp}^{poly}(r_j))^2}, \tag{2}$$

where u_l^{pce} is the response surface of axial liquid velocity obtained in mono-disperse simulations and r_j corresponds to the N_{exp} experimental points at the considered height. The black curve in Fig. 9 shows the relative error norm according to the selected injection diameter. An optimal value for the injected diameter clearly appears at $d_l^{inj} = 9.7\mu$m for this criterion.

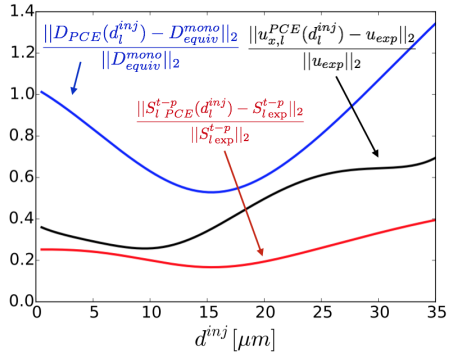

Figure 9: Normalised L^2-norm of both criteria at $x = 7.5$ mm.

The final objective being reactive simulations of flame propagation in a two-phase flow, a key quantity is the laminar burning speed, to consider here in the presence of droplets. Ballal and Lefebvre [2] gave an expression for such a two-phase laminar flame speed S_l^{t-p}, whose validity has been investigated numerically in [18]:

$$S_l^{t-p} = \alpha_g \left[\frac{C_3^3 \rho_l D_{32}^2}{8 C_1 \rho_g \ln(1+B)} + \frac{\alpha_g^2}{S_L^2} \right]^{-0.5} \tag{3}$$

with α_g the thermal conductivity, ρ_l and ρ_g the liquid and gaseous densities, B the Spalding number and S_L the gaseous laminar flame speed. This formula is valid for a polydisperse spray, with $C_1 = D_{20}/D_{32}$ and $C_3 = D_{30}/D_{32}$. In the case of a mono-disperse spray, the coefficients C_1 and C_3 are both equal to unity, which yields the equivalent mono-disperse diameter that conserves the two-phase laminar burning speed:

$$D_{equiv}^{mono} = D_{32}^{poly} \times \sqrt{\frac{C_3^3}{C_1}} \tag{4}$$

The predicted diameter can then be compared to the equivalent one with respect to the flame speed defined in Eq. 4, which is displayed in blue in Fig. 9, again using the relative error norm. The corresponding error on S_l^{t-p} is given in red. It appears that the optimal diameter varies strongly depending on the criterion used. In order to retrieve the liquid velocity fields, injecting small droplets, with a diameter around 9.5 μm, which is close to the spray's mean diameter D_{10}, seems optimal. However, to reproduce the flame speed, a better injection diameter would be around 15.3 μm. For this diameter, the relative error on the flame speed is of 16%, which remains noticeable. These results are consistent with the definition of the diameters given by Lefebvre in [13],

who states that among a set of available representative diameters, the D_{10} represents the velocity fields while the D_{32} is more suited for combustion. However, in order to reduce even more the error on the flame speed, a finer optimization would be required, for example on the actual position of the flame front, which is unknown in the present case.

4. Conclusion

The present investigation is carried out as a first stage preparation of a large scale simulation of the light-round ignition process in a full annular combustor. The objective is to complete the calculation of such a system comprising multiple injectors fed with liquid fuel. One central issue in such simulations is that of the modeling of the liquid droplet spray. It is considered that such a large scale simulation cannot accommodate models which account for the spray polydispersity and that it is interesting to use a mono-disperse Eulerian framework to reduce the computational intensity and comply with limitations in CPU resources. It is next indicated that the droplet size of the spray becomes a key parameter that is here adjusted by making use of an uncertainty quantification framework. This novel method is explored in the case of a single injector investigated experimentally and numerically. Several mono-disperse Eulerian simulations are performed in order to study the impact of the mono-disperse simplification when modeling a polydisperse spray. It is shown that this approach provides a reasonable description of the spray formed in the single injector configuration. The optimal choice of a droplet diameter is then based on the response surface obtained by polynomial chaos expansion and varies with the purpose of the simulation, and therefore the accuracy criterion.

This study relies on the hypothesis that the classes behave in an independent manner. This hypothesis was not validated, which would require a reference simulation using the polydisperse Lagrangian formalism. The inter-class interactions would then be quantified. In order to further validate the optimal injection diameter as well as the criterion used to select it, a simulation with combustion would also be necessary. This will be included in some future work.

References

[1] B. Abramzon and W.A. Sirignano. Droplet vaporization model for spray combustion calculations. *Int. Journal of Heat and Mass Transfer*, 32(9):1605 – 1618, 1989.

[2] D.R. Ballal and A.H. Lefebvre. Flame propagation in heterogeneous mixtures of fuel droplets, fuel vapor and air. *Int. Symposium on Combustion*, 18(1):321 – 328, 1981.

[3] D. Barré, L. Esclapez, M. Cordier, E. Riber, B. Cuenot, G. Staffelbach, B. Renou, A. Vandel, L.Y.M. Gicquel, and G. Cabot. Flame propagation in aeronautical swirled multi-burners: Experimental and numerical investigation. *Combustion and Flame*, 161(9):2387 – 2405, 2014.

[4] M. Boileau, G. Staffelbach, B. Cuenot, T. Poinsot, and C. Bérat. Les of an ignition sequence in a gas turbine engine. *Combustion and Flame*, 154(1–2):2 – 22, 2008.

[5] J.-F. Bourgouin, D. Durox, T. Schuller, J. Beaunier, and S. Candel. Ignition dynamics of an annular combustor equipped with multiple swirling injectors. *Combustion and Flame*, 160(8):1398 – 1413, 2013.

[6] M. Chrigui, J. Gounder, A. Sadiki, A. R. Masri, and J. Janicka. Partially premixed reacting acetone spray using les and fgm tabulated chemistry. *Combustion and Flame*, 159(8):2718–2741, 8 2012.

[7] O. Colin and M. Rudgyard. Development of high-order taylor–galerkin schemes for les. *Journal of Computational Physics*, 162(2):338 – 371, 2000.

[8] R. Fan, D. L. Marchisio, and R. O. Fox. Application of the direct quadrature method of moments to polydisperse gas–solid fluidized beds. *Powder Technology*, 139(1):7–20, 1 2004.

[9] R. O. Fox. Large-Eddy-Simulation Tools for Multiphase Flows. In Davis, SH and Moin, P, editor, *Annual Review of Fluid Mechanics*, volume 44, pages 47–76. Davis, SH and Moin, P, 2012.

[10] M. Garcia. *Development and validation of the Euler-Lagrange formulation on a parallel and unstructured solver for large-eddy simulation*. PhD thesis, Institut National Polytechnique de Toulouse, 2009.

[11] W. P. Jones, S. Lyra, and S. Navarro-Martinez. Numerical investigation of swirling kerosene spray flames using large eddy simulation. *Combustion and Flame*, 159(4):1539–1561, 4 2012.

[12] M. Khalil, G. Lacaze, J. C. Oefelein, and H. N. Najm. Uncertainty quantification in les of a turbulent bluff-body stabilized flame. *Proceedings of the Combustion Institute*, 35(2):1147–1156, 2015.

[13] A.H. Lefebvre. *Atomization and Sprays*. Taylor and Francis, 1989.

[14] A.H. Lefebvre and D. R. Ballal. *Gas Turbine Combustion*. Taylor and Francis, 2010.

[15] K. Luo, H. Pitsch, M. G. Pai, and O. Desjardins. Direct numerical simulations and analysis of three-dimensional n-heptane spray flames in a model swirl combustor. *Proceedings of the Combustion Institute*, 33(2):2143–2152, 2011.

[16] E. Machover and E. Mastorakos. Spark ignition of annular non-premixed combustors. *Experimental Thermal and Fluid Science*, 73:64–70, 5 2016.

[17] M. Massot. *Multiphase Reacting Flows: Modelling and Simulation*, chapter Eulerian Multi-Fluid Models for Polydisperse Evaporating Sprays, pages 79–123. Springer Vienna, 2007.

[18] A. Neophytou and E. Mastorakos. Simulations of laminar flame propagation in droplet mists. *Combustion and Flame*, 156(8):1627 – 1640, 2009.

[19] F. Nicoud and F. Ducros. Subgrid-scale stress modelling based on the square of the velocity gradient tensor. *Flow, Turbulence and Combustion*, 62(3):183–200, 1999.

[20] M. Philip, M. Boileau, R. Vicquelin, E. Riber, T. Schmitt, B. Cuenot, D. Durox, and S. Candel. Large eddy simulations of the ignition sequence of an annular multiple-injector combustor. *Proceedings of the Combustion Institute*, 35(3):3159 – 3166, 2015.

[21] M. Philip, M. Boileau, R. Vicquelin, T. Schmitt, D. Durox, J.-F. Bourgouin, and S. Candel. Simulation of the ignition process in an annular multiple-injector combustor and comparison with experiments. *Journal of Engineering for Gas Turbines and Power*, 137(3):031501–031501, 09 2014.

[22] T.J Poinsot and S.K Lele. Boundary conditions for direct simulations of compressible viscous flows. *Journal of Computational Physics*, 101(1):104 – 129, 1992.

[23] K. Prieur, D. Durox, J. Beaunier, T. Schuller, and S. Candel. Ignition dynamics in an annular combustor for liquid spray and premixed gaseous injection. *Submitted, Proceedings of the Combustion Institute*, 2016.

[24] M. T. Reagana, H. N. Najm, R. G. Ghanem, and O. M. Knio. Uncertainty quantification in reacting-flow simulations through non-intrusive spectral projection. *Combustion and Flame*, 132(3):545–555, 2003.

[25] M. Sanjosé, J. M. Senoner, F. Jaegle, B. Cuenot, S. Moreau, and T. Poinsot. Fuel injection model for euler–euler and euler–lagrange large-eddy simulations of an evaporating spray inside an aeronautical combustor. *Int. Journal of Multiphase Flow*, 37(5):514–529, 6 2011.

[26] T. Schönfeld and M. Rudgyard. Steady and unsteady flow simulations using the hybrid flow solver avbp. *AIAA Journal*, 37(11):1378–1385, 2016/02/17 1999.

[27] J. M. Senoner, M. Sanjosé, T. Lederlin, F. Jaegle, M. García, E. Riber, B. Cuenot, L. Gicquel, H. Pitsch, and T. Poinsot. Eulerian and lagrangian large-eddy simulations of an evaporating two-phase flow. *Comptes Rendus Mécanique*, 337(6–7):458–468, 2009.

[28] A. Vié, F. Laurent, and M. Massot. Size-velocity correlations in hybrid high order moment/multi-fluid methods for polydisperse evaporating sprays: Modeling and numerical issues. *Journal of Computational Physics*, 237:177–210, 3 2013.

[29] D. Xiu and G. E. Karniadakis. The wiener–askey polynomial chaos for stochastic differential equations. *SIAM journal on scientific computing*, 24(2):619–644, 2002.

Bibliographie

[1] *Transport energy futures : long-term oil supply trends and projections. Report 117*. Bureau of Infrastructure, Transport and Regional Economics, 2009. p. xv, 2, 3

[2] Amal Ben Abdellah, Pierre L'Ecuyer, Art B Owen, and Florian Puchhammer. Density estimation by randomized quasi-monte carlo. *arXiv preprint arXiv :1807.06133*, 2018. p. 210

[3] Milton Abramowitz and Irene A Stegun. *Handbook of mathematical functions : with formulas, graphs, and mathematical tables*, volume 55. Courier Corporation, 1964. p. 146

[4] Ian S Abramson. Arbitrariness of the pilot estimator in adaptive kernel methods. *Journal of Multivariate Analysis*, 12(4) :562–567, 1982. p. 209

[5] Ian S Abramson. On bandwidth variation in kernel estimates-a square root law. *The annals of Statistics*, pages 1217–1223, 1982. p. 209

[6] Alen Alexanderian. A brief note on the karhunen-lo\eve expansion. *arXiv preprint arXiv :1509.07526*, 2015. p. 161, 162

[7] E. Anderson, Z. Bai, C. Bischof, L. S. Blackford, J. Demmel, Jack J. Dongarra, J. Du Croz, S. Hammarling, A. Greenbaum, A. McKenney, and D. Sorensen. *LAPACK Users' Guide (Third Ed.)*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1999. p. 284, 310

[8] Edward Anderson, Zhaojun Bai, Christian Bischof, L Susan Blackford, James Demmel, Jack Dongarra, Jeremy Du Croz, Anne Greenbaum, Sven Hammarling, Alan McKenney, et al. *LAPACK Users' guide*. SIAM, 1999. p. 176

[9] Richard Askey and James Arthur Wilson. *Some basic hypergeometric orthogonal polynomials that generalize Jacobi polynomials*, volume 319. American Mathematical Soc., 1985. p. 147

[10] Kendall Atkinson and Weimin Han. *Numerical Solution of Fredholm Integral Equations of the Second Kind*. Springer, 2009. p. 165, 167, 168, 169, 170, 171

[11] RS Barlow and JH Frank. Effects of turbulence on species mass fractions in methane/air jet flames. In *Symposium (International) on Combustion*, volume 27, pages 1087–1095. Elsevier, 1998. p. 17

[12] Jon Louis Bentley. Multidimensional binary search trees used for associative searching. *Communications of the ACM*, 18(9) :509–517, 1975. p. 210

[13] Åke Björck. Solving linear least squares problems by gram-schmidt orthogonalization. *BIT Numerical Mathematics*, 7(1) :1–21, 1967. p. 327

[14] L Susan Blackford, Jaeyoung Choi, Andy Cleary, Eduardo D'Azevedo, James Demmel, Inderjit Dhillon, Jack Dongarra, Sven Hammarling, Greg Henry, Antoine Petitet, et al. *ScaLAPACK users' guide*. SIAM, 1997. p. 310

[15] Thomas Bradley, Jacques du Toit, Robert Tong, Mike Giles, and Paul Woodhams. Parallelization techniques for random number generators. In *GPU Computing Gems Emerald Edition*, pages 231–246. Elsevier, 2011. p. 82

[16] Paul Bratley and Bennett L Fox. Algorithm 659 : Implementing sobol's quasirandom sequence generator. *ACM Transactions on Mathematical Software (TOMS)*, 14(1) :88–100, 1988. p. 117

[17] Jean-Christophe Breton. Processus gaussiens. *Université de La Rochelle*, 2006. p. 181

[18] Thomas Brox, Bodo Rosenhahn, Daniel Cremers, and Hans-Peter Seidel. Nonparametric density estimation with adaptive, anisotropic kernels for human motion tracking. In *Human Motion–Understanding, Modeling, Capture and Animation*, pages 152–165. Springer, 2007. p. 211

[19] Robert H Cameron and William T Martin. The orthogonal development of non-linear functionals in series of fourier-hermite functionals. *Annals of Mathematics*, pages 385–392, 1947. p. 145

[20] Vasile Carutasu. Numerical solution of two-dimensional nonlinear fredholm integral equations of the second kind by spline functions. *Gen. Mathematics*, 9 :31–48, 2001. p. 173

[21] Su Chen, Josef Dick, and Art B Owen. Consistency of markov chain quasi-monte carlo on continuous state spaces. *The Annals of Statistics*, pages 673–701, 2011. p. 88

[22] Charles W Clenshaw and Alan R Curtis. A method for numerical integration on an automatic computer. *Numerische Mathematik*, 2(1) :197–205, 1960. p. 47

[23] Devroye. *Non-uniform random variate generation*. Springers, 1986. p. 88

[24] Luc Devroye. Nonuniform random variate generation. *Handbooks in operations research and management science*, 13 :83–121, 2006. p. 84

[25] Rick Durrett. *Probability : theory and examples*. Cambridge university press, 2010. p. 174

[26] Morris L. Eaton. Multivariate statistics : A vector space approach. 2007. p. 204, 206

[27] Bradley Efron. Bootstrap methods : another look at the jackknife. In *Breakthroughs in statistics*, pages 569–593. Springer, 1992. p. 134

[28] Bradley Efron and Robert J Tibshirani. *An introduction to the bootstrap*. CRC press, 1994. p. 134

[29] Vassiliy A Epanechnikov. Non-parametric estimation of a multivariate probability density. *Theory of Probability & Its Applications*, 14(1) :153–158, 1969. p. 199

[30] James F Epperson. On the runge example. *Amer. Math. Monthly*, 94(4) :329–341, 1987. p. 37

[31] Oliver G Ernst, Antje Mugler, Hans-Jörg Starkloff, and Elisabeth Ullmann. On the convergence of generalized polynomial chaos expansions. *ESAIM : Mathematical Modelling and Numerical Analysis*, 46(2) :317–339, 2012. p. 148, 149

[32] Leopold Fejér. Mechanische quadraturen mit positiven cotesschen zahlen. *Mathematische Zeitschrift*, 37(1) :287–309, 1933. p. 47

[33] Benoit Fiorina, Denis Veynante, and Sébastien Candel. Modeling combustion chemistry in large eddy simulation of turbulent flames. In *TSFP DIGITAL LIBRARY ONLINE*. Begel House Inc., 2013. p. 15

[34] B Franzelli, E Riber, M Sanjosé, and Thierry Poinsot. A two-step chemical scheme for kerosene–air premixed flames. *Combustion and Flame*, 157(7) :1364–1373, 2010. p. 14

[35] Michael Frenklach, Hai Wang, and Martin J Rabinowitz. Optimization and analysis of large chemical kinetic mechanisms using the solution mapping method—combustion of methane. *Progress in Energy and Combustion Science*, 18(1) :47–73, 1992. p. 10

[36] Keinosuke Fukunaga. *Introduction to statistical pattern recognition*. Academic press, 2013. p. 210

[37] J Galpin, C Angelberger, A Naudin, and L Vervisch. Large-eddy simulation of h2–air auto-ignition using tabulated detailed chemistry. *Journal of Turbulence*, (9) :N3, 2008. p. 224

[38] Carl Friedrich Gauss. *Methodus nova integralium valores per approximationem inveniendi*. apvd Henricvm Dieterich, 1815. p. 45

[39] Stuart Geman and Donald Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. In *Readings in Computer Vision*, pages 564–584. Elsevier, 1987. p. 88

[40] Alan Genz. Testing multidimensional integration routines. In *Proc. of international conference on Tools, methods and languages for scientific and engineering computation*, pages 81–94. Elsevier North-Holland, Inc., 1984. p. 50, 64, 126, 136, 343

[41] Thomas Gerstner and Michael Griebel. Numerical integration using sparse grids. *Numerical algorithms*, 18(3-4) :209–232, 1998. p. 62, 64, 170

[42] Roger G Ghanem and Pol D Spanos. *Stochastic finite elements : a spectral approach*. Courier Corporation, 2003. p. 174

[43] Olivier Gicquel, Nasser Darabiha, and Dominique Thévenin. Liminar premixed hydrogen/air counterflow flame simulations using flame prolongation of ildm with differential diffusion. *Proceedings of the Combustion Institute*, 28(2) :1901–1908, 2000. p. 224

[44] RG Gilbert, K_ Luther, and J Troe. Theory of thermal unimolecular reactions in the fall-off range. ii. weak collision rate constants. *Berichte der Bunsengesellschaft für physikalische Chemie*, 87(2) :169–177, 1983. p. 9

[45] Gene H Golub and John H Welsch. Calculation of gauss quadrature rules. *Mathematics of computation*, 23(106) :221–230, 1969. p. 45, 46, 50

[46] Piotr Graczyk and Tomasz Jakubowski. A survey. p. 173, 174

[47] E Hairer et al. Solving ordinary differential equations stiff and differentialxalgebraic problems springer series in comput, 1996. p. 227

[48] Philip Hall. The distribution of means for samples of size n drawn from a population in which the variate takes values between 0 and 1, all such values being equally probable. *Biometrika*, pages 240–245, 1927. p. 97

[49] John H Halton. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numerische Mathematik*, 2(1) :84–90, 1960. p. 115

[50] Hans Hamburger. Über eine erweiterung des stieltjesschen momentenproblems. *Mathematische Annalen*, 81(2-4) :235–319, 1920. p. 148

[51] W Keith Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1) :97–109, 1970. p. 88

[52] Edmund Hlawka. Funktionen von beschränkter variatiou in der theorie der gleichverteilung. *Annali di Matematica Pura ed Applicata*, 54(1) :325–333, 1961. p. 113

[53] Masao Iri, Sigeiti Moriguti, and Yoshimitsu Takasawa. On a certain quadrature formula. *Journal of computational and applied mathematics*, 17(1-2) :3–20, 1987. p. 44

[54] Stephen Joe and Frances Y Kuo. Constructing sobol sequences with better two-dimensional projections. *SIAM Journal on Scientific Computing*, 30(5) :2635–2654, 2008. p. 118

[55] Steven G Johnson. Notes on the convergence of trapezoidal-rule quadrature, 2010. p. 47

[56] Galin L Jones et al. On the markov chain central limit theorem. *Probability surveys*, 1(299-320) :5–1, 2004. p. 88

[57] WP Jones and RP Lindstedt. Global reaction schemes for hydrocarbon combustion. *Combustion and flame*, 73(3) :233–249, 1988. p. 14

[58] Vesa Kaarnioja et al. Smolyak quadrature. 2013. p. 58

[59] Kari Karhunen. *Über lineare Methoden in der Wahrscheinlichkeitsrechnung*, volume 37. Universitat Helsinki, 1947. p. 159

[60] Andrew P Kelley, Wei Liu, YX Xin, AJ Smallbone, and CK Law. Laminar flame speeds, non-premixed stagnation ignition, and reduced mechanisms in the oxidation of iso-octane. *Proceedings of the Combustion Institute*, 33(1) :501–508, 2011. p. 13

[61] Mohammad Khalil, Guilhem Lacaze, Joseph C Oefelein, and Habib N Najm. Uncertainty quantification in les of a turbulent bluff-body stabilized flame. *Proceedings of the Combustion Institute*, 35(2) :1147–1156, 2015. p. 16

[62] Andrei Nikolaevich Kolmogorov. Grundbegriffe der wahrscheinlichkeitsrechnung, berlin, 1933. *English translation, Chelsea, New York*, 1950. p. 92

[63] Alexander A Konnov. Remaining uncertainties in the kinetic mechanism of hydrogen combustion. *Combustion and flame*, 152(4) :507–528, 2008. p. xxii, 224, 225

[64] Dirk Laurie. Calculation of gauss-kronrod quadrature rules. *Mathematics of Computation of the American Mathematical Society*, 66(219) :1133–1145, 1997. p. 46

[65] Olivier Le Maître and Omar M Knio. *Spectral methods for uncertainty quantification : with applications to computational fluid dynamics*. Springer Science & Business Media, 2010. p. 7, 150, 151, 152, 153, 163, 188

[66] Pierre L'Ecuyer. Uniform random number generators : a review. In *Proceedings of the 29th conference on Winter simulation*, pages 127–134. IEEE Computer Society, 1997. p. 77

[67] Pierre L'ecuyer. Good parameters and implementations for combined multiple recursive random number generators. *Operations Research*, 47(1) :159–164, 1999. p. 78, 80

[68] Pierre L'Ecuyer and Peter Hellekalek. Random number generators : Selection criteria and testing. In *Random and Quasi-Random Point Sets*, pages 223–265. Springer, 1998. p. 77

[69] Pierre L'Ecuyer, Christian Lécot, and Bruno Tuffin. A randomized quasi-monte carlo simulation method for markov chains. *Operations Research*, 56(4) :958–975, 2008. p. 88

[70] Pierre L'Ecuyer and Richard Simard. On the performance of birthday spacings tests with certain families of random number generators. *Mathematics and Computers in Simulation*, 55(1) :131–137, 2001. p. 77

[71] John A Lee and Michel Verleysen. *Nonlinear dimensionality reduction*. Springer Science & Business Media, 2007. p. 336

[72] Christiane Lemieux. *Monte carlo and quasi-monte carlo sampling*. Springer Science & Business Media, 2009. p. 105, 112, 113, 114, 118, 120, 122, 128

[73] Christiane Lemieux and Art B Owen. Quasi-regression and the relative importance of the anova components of a function. *Monte Carlo and Quasi-Monte Carlo Methods*, pages 331–344, 2002. p. 158

[74] Genyuan Li, Sheng-Wei Wang, and Herschel Rabitz. Practical approaches to construct rs-hdmr component functions. *The Journal of Physical Chemistry A*, 106(37) :8721–8733, 2002. p. 156, 157

[75] Genyuan Li, Sheng-Wei Wang, Carey Rosenthal, and Herschel Rabitz. High dimensional model representations generated from low dimensional data samples. i. mp-cut-hdmr. *Journal of Mathematical Chemistry*, 30(1) :1–30, 2001. p. 156

[76] FA Lindemann, Svante Arrhenius, Irving Langmuir, NR Dhar, J Perrin, and WC McC Lewis. Discussion on "the radiation theory of chemical action". *Transactions of the Faraday Society*, 17 :598–606, 1922. p. 9

[77] Don O Loftsgaarden and Charles P Quesenberry. A nonparametric estimate of a multivariate density function. *The Annals of Mathematical Statistics*, pages 1049–1051, 1965. p. 209

[78] Tianfeng Lu and Chung K Law. A directed relation graph method for mechanism reduction. *Proceedings of the Combustion Institute*, 30(1) :1333–1341, 2005. p. 13

[79] Tianfeng Lu and Chung K Law. A criterion based on computational singular perturbation for the identification of quasi steady state species : A reduced mechanism for methane oxidation with no chemistry. *Combustion and Flame*, 154(4) :761–774, 2008. p. 13

[80] Ulrich Maas and Stephen B Pope. Simplifying chemical kinetics : intrinsic low-dimensional manifolds in composition space. *Combustion and flame*, 88(3-4) :239–264, 1992. p. 15, 226

[81] Donald W Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the society for Industrial and Applied Mathematics*, 11(2) :431–441, 1963. p. 245

[82] AR Masri, JB Kelman, and BB Dally. The instantaneous spatial structure of the recirculation zone in bluff-body stabilized flames. In *Symposium (International) on Combustion*, volume 27, pages 1031–1038. Elsevier, 1998. p. 16

[83] A Massias, D Diamantis, E Mastorakos, and DA Goussis. An algorithm for the construction of global reduced mechanisms with csp data. *Combustion and Flame*, 117(4) :685–708, 1999. p. 13

[84] Jiří Matoušek. On the l2-discrepancy for anchored boxes. *J. Complex.*, 14(4) :527–556, December 1998. p. 121, 122, 124, 125

[85] Michael D McKay, Richard J Beckman, and William J Conover. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 42(1) :55–61, 2000. p. 110, 111

[86] Bradley Moskowitz and Russel E Caflisch. Smoothness and dimension reduction in quasi-monte carlo methods. *Mathematical and Computer Modelling*, 23(8-9) :37–54, 1996. p. 86

[87] Michael E Mueller, Gianluca Iaccarino, and Heinz Pitsch. Chemical kinetic uncertainty quantification for large eddy simulation of turbulent nonpremixed combustion. *Proceedings of the Combustion Institute*, 34(1) :1299–1306, 2013. p. 17, 242

[88] Michael E Mueller and Venkat Raman. Effects of turbulent combustion modeling errors on soot evolution in a turbulent nonpremixed jet flame. *Combustion and Flame*, 161(7) :1842–1848, 2014. p. 16

[89] Tibor Nagy and Tamas Turanyi. Determination of the uncertainty domain of the arrhenius parameters needed for the investigation of combustion kinetic models. *Reliability Engineering & System Safety*, 107 :29–34, 2012. p. 11, 12

[90] Habib N Najm, Bert J Debusschere, Youssef M Marzouk, Steve Widmer, and OP Le Maître. Uncertainty quantification in chemical systems. *International journal for numerical methods in engineering*, 80(6-7) :789–814, 2009. p. 12

[91] H Niederreiter. Random number generation and quasi-monte carlo methods, siam cbms-nsf regional conference series in applied mathematics, vol. 63. *SIAM, Philadelphia, PA*, 1992. p. 57

[92] Yi-Shuai Niu, Luc Vervisch, and Pham Dinh Tao. An optimization-based approach to detailed chemistry tabulation : Automated progress variable definition. *Combustion and Flame*, 160(4) :776–785, 2013. p. 232

[93] Carsten Olm, Tamás Varga, Éva Valkó, Henry J Curran, and Tamás Turányi. Uncertainty quantification of a newly optimized methanol and formaldehyde combustion mechanism. *Combustion and Flame*, 186 :45–64, 2017. p. 12

[94] AB Owen. Variance and discrepancy with alternative scrambling. *ACM Trans Model Comput Simul*, 13 :363–378, 2003. p. 125

[95] Art B Owen. Randomly permuted (t, m, s)-nets and (t, s)-sequences. In *Monte Carlo and quasi-Monte Carlo methods in scientific computing*, pages 299–317. Springer, 1995. p. 122

[96] Art B Owen. Scrambled net variance for integrals of smooth functions. *The Annals of Statistics*, pages 1541–1562, 1997. p. 122

[97] Athanasios Papoulis. Probability, random variables, and stochastic processes. 1965. p. 94

[98] Chull Park. Representations of gaussian processes by wiener processes. *Pacific Journal of Mathematics*, 94(2) :407–415, 1981. p. 181

[99] N Peters. Numerical and asymptotic analysis of systematically reduced reaction schemes for hydrocarbon flames. In *Numerical simulation of combustion phenomena*, pages 90–109. Springer, 1985. p. 13

[100] Norbert Peters. Laminar diffusion flamelet models in non-premixed turbulent combustion. *Progress in energy and combustion science*, 10(3) :319–339, 1984. p. 17

[101] Gaël Poëtte and Didier Lucor. Non intrusive iterative stochastic spectral representation with application to compressible gas dynamics. *Journal of Computational Physics*, 231(9) :3587–3609, 2012. p. 147, 148

[102] Thierry Poinsot and Denis Veynante. *Theoretical and numerical combustion*. RT Edwards, Inc., 2005. p. 14

[103] Georg Pólya. Über den zentralen grenzwertsatz der wahrscheinlichkeitsrechnung und das momentenproblem. *Mathematische Zeitschrift*, 8(3) :171–181, 1920. p. 96

[104] William H Press, Saul A Teukolsky, William T Vetterling, and Brian P Flannery. *Numerical recipes 3rd edition : The art of scientific computing*. Cambridge university press, 2007. p. 87, 205

[105] Alfio Maria Quarteroni, Riccardo Sacco, and Fausto Saleri. *Méthodes numériques : algorithmes, analyse et applications*, chapter 8. Springer Science & Business Media, 2008. p. 41

[106] Maurice H Quenouille. Notes on bias in estimation. *Biometrika*, 43(3/4) :353–360, 1956. p. 132

[107] Maurice H Quenouille et al. Problems in plane sampling. *The Annals of Mathematical Statistics*, 20(3) :355–375, 1949. p. 132

[108] Lewis Fry Richardson. The approximate arithmetical solution by finite differences of physical problems involving differential equations, with an application to the stresses in a masonry dam. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 210 :307–357, 1911. p. 42

[109] Theodore-J Rivlin. Chebyshev polynomials. 1990. p. 49

[110] Jean-Étienne Rombaldi. *Interpolation & approximation : analyse pour l'agrégation : cours & exercices résolus*, page 330. Vuibert, 2005. p. 42

[111] Werner Romberg. Vereinfachte numerische integration. *Det Kongelige Norske Videnskabers Selskab Forhandlinger*, 28(7) :30–36, 1955. p. 42

[112] Murray Rosenblatt. Remarks on a multivariate transformation. *The annals of mathematical statistics*, 23(3) :470–472, 1952. p. 90

[113] Prakash Kumar Sahu and S Saha Ray. Numerical solutions for the system of fredholm integral equations of second kind by a new approach involving semiorthogonal b-spline wavelet collocation method. *Applied Mathematics and Computation*, 234 :368–379, 2014. p. 173

[114] Mohamed Sallak, Felipe Aguirre, and Walter Schon. Incertitudes aléatoires et épistémiques, comment les distinguer et les manipuler dans les études de fiabilité ? In *QUALITA2013*, 2013. p. 4, 5

[115] Andrea Saltelli, Stefano Tarantola, and KP-S Chan. A quantitative model-independent method for global sensitivity analysis of model output. *Technometrics*, 41(1) :39–56, 1999. p. 156

[116] Charles Schwartz. Numerical integration of analytic functions. *Journal of Computational Physics*, 4(1) :19–29, 1969. p. 44

[117] David W Scott. *Multivariate density estimation : theory, practice, and visualization*. John Wiley & Sons, 2015. p. 188, 190, 191, 193, 196, 198, 199, 200, 202, 204, 208, 209, 211, 214

[118] Claude Elwood Shannon. A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, 5(1) :3–55, 2001. p. 47

[119] J Shao and D Tu. The jackknife and bootstrap. 1995, 1995. p. 133, 135

[120] Bernard W Silverman and PJ Green. Nonparametric regression and generalized linear models : A roughness penalty approach. 1993. p. 255

[121] Yakov G Sinai. *Probability theory : an introductory course*. Springer Science & Business Media, 2013. p. 6

[122] Evgeny Slutsky. Uber stochastische asymptoten und grenzwerte. *Metron*, 5(3) :3–89, 1925. p. 102

[123] Sergey A Smolyak. Quadrature and interpolation formulas for tensor products of certain classes of functions. In *Dokl. Akad. Nauk SSSR*, volume 4, page 123, 1963. p. 60

[124] Ilya M Sobol. Sensitivity estimates for nonlinear mathematical models. *Mathematical modelling and computational experiments*, 1(4) :407–414, 1993. p. 153

[125] Ilya M Sobol. Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates. *Mathematics and computers in simulation*, 55(1-3) :271–280, 2001. p. 155

[126] Il'ya Meerovich Sobol'. On the distribution of points in a cube and the approximate evaluation of integrals. *Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki*, 7(4) :784–802, 1967. p. 119

[127] IM Sobol and Yu L Levitan. The production of points uniformly distributed in a multidimensional cube. *Preprint IPM Akad. Nauk SSSR*, 40(3), 1976. p. 117

[128] Christian Soize and Roger Ghanem. Physical systems with random uncertainties : chaos representations with arbitrary probability measure. *SIAM Journal on Scientific Computing*, 26(2) :395–410, 2004. p. 147, 149

[129] Bruno Sudret. Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering & System Safety*, 93(7) :964–979, 2008. p. 153

[130] George R Terrell and David W Scott. Variable kernel density estimation. *The Annals of Statistics*, pages 1236–1265, 1992. p. 208, 211

[131] Shu Tezuka and Henri Faure. I-binomial scrambling of digital nets and sequences. *Journal of complexity*, 19(6) :744–757, 2003. p. 122, 125

[132] Alison S Tomlin. The role of sensitivity and uncertainty analysis in combustion modelling. *Proceedings of the Combustion Institute*, 34(1) :159–176, 2013. p. 11

[133] AS Tomlin, T Turanyi, and MJ Pilling. Mathematical tools for the construction, investigation and reduction of combustion mechanisms. *ChemInform*, 29(32), 1998. p. 13

[134] Lloyd N Trefethen. Is gauss quadrature better than clenshaw-curtis ? *SIAM review*, 50(1) :67–87, 2008. p. 47

[135] Diana Elena Tudorache. *Tabulation de la cinétique chimique pour la prédiction des polluants dans les moteurs à combustion interne*. PhD thesis, Châtenay-Malabry, Ecole centrale de Paris, 2013. p. 263

[136] John W Tukey. Bias and confidence in not-quite large samples. *Ann. Math. Statist.*, 29 :614, 1958. p. 132, 133

[137] George E Uhlenbeck and Leonard S Ornstein. On the theory of the brownian motion. *Physical review*, 36(5) :823–841, 1930. p. 173

[138] Bart Vandewoestyne and Ronald Cools. Good permutations for deterministic scrambled halton sequences in terms of l2-discrepancy. *Journal of computational and applied mathematics*, 189(1) :341–361, 2006. p. 122

[139] Erik Vanmarcke. *Random fields : analysis and synthesis*. World Scientific, 2010. p. 161

[140] Ronan Vicquelin. *Tabulation de la cinétique chimique pour la modélisation et la simulation de la combustion turbulente*. PhD thesis, Châtenay-Malabry, Ecole centrale de Paris, 2010. p. 15

[141] Jörg Waldvogel. Fast construction of the fejér and clenshaw–curtis quadrature rules. *BIT Numerical Mathematics*, 46(1) :195–202, 2006. p. 48, 49

[142] Hai Wang and David A Sheen. Combustion kinetic model uncertainty quantification, propagation and minimization. *Progress in Energy and Combustion Science*, 47 :1–31, 2015. p. 11

[143] J Warnatz, U Maas, and RW Dibble. Combustion : physical and chemical fundamentals, modeling and simulation, experiments, pollutant formation, berlin, 2006. p. 13

[144] Neil A Weiss. *A course in probability*. Addison-Wesley, 2006. p. 251

[145] Norbert Wiener. The homogeneous chaos. *American Journal of Mathematics*, 60(4) :897–936, 1938. p. 146

[146] Dongbin Xiu and George Em Karniadakis. The wiener–askey polynomial chaos for stochastic differential equations. *SIAM journal on scientific computing*, 24(2) :619–644, 2002. p. 147

[147] Yuan Xu. On orthogonal polynomials in several variables. *Special functions, q-series and related topics, The Fields Institute for Research in Mathematical Sciences, Communications Series*, 14 :247–270, 1997. p. 147

université PARIS-SACLAY

ÉCOLE DOCTORALE
Sciences mécaniques et énergétiques, matériaux et géosciences (SMEMAG)

Titre : Méthodes numériques et modèle réduit de chimie tabulée pour la propagation d'incertitudes de cinétique chimique.

Mots clés : incertitudes, chimie tabulée, cinétique chimique

Résumé : La simulation numérique joue aujourd'hui un rôle majeur dans le domaine de la combustion, que ce soit au niveau de la recherche en offrant la possibilité de mieux comprendre les phénomènes ayant lieu au sein des écoulements réactifs ou au niveau du développement de nouveaux systèmes industriels par une diminution des coûts liés à la conception de ces systèmes. A l'heure actuelle, la simulation aux grandes échelles est l'outil le mieux adapté à la simulation numérique d'écoulements réactifs turbulents. Cette simulation aux grandes échelles d'écoulements réactifs n'est en pratique possible que grâce à une modélisation des différents phénomènes:

- la turbulence est modélisée pour les plus petites structures permettant de n'avoir à résoudre que les plus grandes structures de l'écoulement et ainsi réduire le coût de calcul
- la chimie des différentes espèces réactives est modélisée à l'aide de méthodes de réduction permettant de considérablement réduire le coût de calcul

La maturité de la simulation aux grandes échelles d'écoulements réactifs en fait aujourd'hui un outil fiable, prédictif et prometteur. Il fait désormais sens de s'intéresser à l'impact des paramètres impliqués dans les différents modèles sur le résultat de la simulation. Cette étude de l'impact des paramètres de modélisation peut être vue sous l'angle de la propagation d'incertitudes, et peut donner des informations intéressantes à la fois d'un côté pratique pour la conception robuste de systèmes mais également d'un côté théorique afin d'améliorer les modèles utilisées et d'orienter les mesures expérimentales à réaliser afin d'améliorer la fiabilité de ces modèles.
Le contexte de cette thèse est le développement de méthodes efficaces permettant la propagation d'incertitudes présentes dans les paramètres de cinétique chimique des mécanismes réactionnels au sein de simulation aux grandes échelles, ces méthodes devant être non intrusive afin de profiter de l'existence des différents codes de calcul qui sont des outils nécessitant de lourds moyens pour leur développement. Une telle propagation d'incertitude à l'aide d'une méthode de force brute souffre du "fléau de la dimension" du fait du grand nombre de paramètres de cinétique chimique, impliquant une impossibilité pratique avec les moyens de calculs actuels et justifiant le développement de méthodes efficaces.
L'objectif de la thèse est donc le développement d'un modèle réduit utilisable pour la propagation d'incertitudes dans la simulation aux grandes échelles. La prise en main et l'implémentation de différents outils issus de la propagation d'incertitudes a été un travail préliminaire indispensable dans cette thèse afin d'amener ces connaissances et compétences au sein du laboratoire EM2C.
La méthode développée dans cette thèse pour la propagation d'incertitudes des paramètres de cinétique chimique se restreint aux cas d'une modélisation de la chimie dans laquelle l'avancement du processus de combustion est résumé par l'évolution d'une variable d'avancement donnée par une équation de transport, l'accès aux autres informations se faisant grâce à l'utilisation d'une table. Au travers de l'étude de l'évolution d'un réacteur adiabatique à pression constante contenant un mélange homogène d'air et de dihydrogène, il est montré qu'une grande partie des incertitudes d'un tel système peuvent être expliquées grâce aux incertitudes de la variable d'avancement. Cela permet de définir une table chimique utilisable pour la propagation d'incertitudes des paramètres de cinétique chimique dans les simulations aux grandes échelles. L'introduction des incertitudes se fait alors uniquement par la modélisation du terme source présent dans l'équation de transport de la variable d'avancement, lequel peut être paramétré à l'aide de quelques paramètres incertains évitant ainsi le "fléau de la dimension".

Title : Numerical methods and reduced model of tabulated chemistry for uncertainties propagation of chemical kinetic.

Keywords : uncertainties, tabulated chemistry, chemical kinetic

Abstract : Numerical simulation plays a key role in the field of combustion today, either in the research area by permitting a better understanding of phenomenons taking place inside reactive flows or in the development of industrial application by reducing designing cost of systems. Large Eddy Simulation is at the time the most suited tool for the simulation of reactive flows. Large Eddy Simulation of reactive flows is in practice only possible thanks to a modeling of different phenomenons:

- turbulence is modeled for small structures allowing to resolve only big structures which results in lower computational cost
- chemistry is modeled using reduction methods which allows to drastically reduce computational cost

The maturity of Large Eddy Simulation of reactive flows makes it today a reliable, predictive and promising tool. It now makes sense to focus on the impact of the parameters involved in the different models on the simulation results. This study of the impact of the modeling parameters can be seen from the perspective of uncertainties propagation, and can give interesting informations both from a practical side for the robust design of systems but also on the theoretical side in order to improve the models used and guide the experimental measurements to be made for the reliability improvement of these models.
The context of this thesis is the development of efficient methods allowing the propagation of uncertainties present in the chemical kinetic parameters of the reaction mechanisms within Large Eddy Simulation, these methods having to be non-intrusive in order to take advantage of the existence of the different computation codes which are tools requiring heavy means for their development. Such a propagation of uncertainties using a brute-force method suffers from the "curse of dimensionality" because of the large number of chemical kinetic parameters, implying a practical impossibility with the current means of computation which justifies the development of efficient methods.
The objective of the thesis is the development of a reduced model that can be used for uncertainties propagation in Large Eddy simulations. The handling and implementation of various tools resulting from the uncertainties propagation framework has been an essential preliminary work in this thesis in order to bring this knowledge and skills into the EM2C laboratory.
The method developed in this thesis for the propagation of chemical kinetic parameters uncertainties is limited to chemistry models in which the advancement of the combustion process is summarized by the evolution of a progress variable given by a transport equation, the access to other informations being made through the use of a table. Through the study of the evolution of a constant pressure adiabatic reactor containing a homogeneous mixture of air and dihydrogen, it is shown that a large part of the uncertainties of such a system can be explained by the uncertainties of the progress variable. This makes it possible to define a chemical table that can be used to propagate uncertainties of chemical kinetic parameters in Large Eddy Simulations. The introduction of the uncertainties is then done only by the modeling of the source term present in the transport equation of the progress variable, which can be parameterized with the help of few uncertain parameters thus avoiding the "curse of dimensionality".

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France