



**HAL**  
open science

# Métamodélisation et optimisation de dispositifs photoniques

Cédric Durantin

► **To cite this version:**

Cédric Durantin. Métamodélisation et optimisation de dispositifs photoniques. Modélisation et simulation. COMUE Université Côte d'Azur (2015 - 2019), 2018. Français. NNT : 2018AZUR4036 . tel-01868323

**HAL Id: tel-01868323**

**<https://theses.hal.science/tel-01868323v1>**

Submitted on 5 Sep 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# THÈSE DE DOCTORAT

## Métamodélisation et optimisation de dispositifs photoniques

Présentée par

**Cédric DURANTIN**

Thèse préparée au sein du Laboratoire des capteurs optiques du CEA LETI.

Présentée en vue de l'obtention du grade de docteur en mathématiques appliquées de  
l'Université Côte d'Azur

**Soutenue le 28 mai 2018 , devant le jury composé de :**

Régis Orobtcouk,	Maître de Conférence HDR,	INSA,	Rapporteur
Pierre Alain Boucard,	Professeur,	Université Paris Sud 11,	Rapporteur
Delphine Sinoquet,	Ingénieur-chercheur,	IFPEN,	Examinatrice
Rodolphe Le Riche,	Directeur de recherche,	CNRS,	Président du jury
Jean Antoine Désidéri,	Directeur de recherche,	INRIA,	Directeur
Alain Glière,	Ingénieur-chercheur HDR,	CEA,	Co-directeur



**Titre : Métamodélisation et optimisation de dispositifs photoniques**

La simulation numérique est couramment utilisée pour étudier le comportement d'un composant et optimiser sa conception. Pour autant, chaque calcul est souvent coûteux en termes de temps et l'optimisation nécessite de résoudre un grand nombre de fois le modèle numérique pour différentes configurations du composant. Une solution actuelle pour réduire le temps de calcul consiste à remplacer la simulation coûteuse par un métamodèle (également appelé modèle de substitution). Des stratégies sont ensuite mises en place pour réaliser l'optimisation du composant à partir de l'information fournie par le métamodèle. Dans le cadre de cette thèse, trois dispositifs représentatifs des applications pouvant être traitées au sein du département d'optique et de photonique du CEA LETI sont identifiés. L'étude de ces cas permet d'établir deux problématiques à résoudre. La première concerne la métamodélisation multi-fidélité, qui consiste à construire un métamodèle à partir de deux simulations du même composant ayant une précision différente. Les simulations sont obtenues à partir de différentes approximations du phénomène physique et aboutissent à un modèle appelé haute fidélité (précis et coûteux) et un modèle basse fidélité (grossier et rapide à évaluer). Le travail sur cette méthode pour le cas de la cellule photoacoustique a amené au développement d'un nouveau métamodèle multifidélité basé sur les fonctions à base radiale. La deuxième problématique concerne la prise en compte des incertitudes de fabrication dans la conception de dispositifs photoniques. L'optimisation des performances de composants en tenant compte des écarts observés entre la géométrie désirée et la géométrie obtenue en fabrication a nécessité le développement d'une méthode spécifique pour le cas du coupleur adiabatique. L'ensemble du travail de cette thèse est capitalisé dans une boîte à outils logiciel.

**Mots clés :** métamodèle, krigeage, RBF, multifidélité, stratégie adaptative, optimisation, robuste, multi-objectif, composants photoniques.

**Title : Metamodeling and optimization of photonics devices**

Numerical simulation is widely employed in engineering to study the behavior of a device and optimize its design. Nevertheless, each computation is often time consuming and, during an optimization sequence, the simulation code is evaluated a large number of times. An interesting way to reduce the computational burden is to build a metamodel (or surrogate model) of the simulation code. Adaptive strategies are then set up for the optimization of the component using the metamodel prediction. In the context of this thesis, three representative devices are identified for applications that can be encountered within the CEA LETI optics and photonics department. The study of these cases resulted in two problems to be treated. The first one concerns multifidelity metamodeling, which consists of constructing a metamodel from two simulations of the same component that can be hierarchically ranked in accuracy. The simulations are obtained from different approximations of the physical phenomenon. The work on this method for the case of the photoacoustic cell has generated the development of a new multifidelity surrogate model based on radial basis function. The second problem relate to the consideration of manufacturing uncertainties in the design of photonic devices. Taking into account the differences observed between the desired geometry and the geometry obtained in manufacturing for the optimization of the component efficiency requires the development of a particular method for the case of the adiabatic coupler. The entire work of this thesis is capitalized in a software toolbox.

**Keywords :** metamodel, kriging, RBF, multifidelity, adaptive strategy, optimization, robust, multi-objective, photonic devices.



---

# Remerciements

---

Ce manuscrit est le résultat d'un travail personnel mais il est également issu de corrections, de soutiens et d'apports (directs ou indirects) de plusieurs personnes.

Je tiens tout d'abord à remercier mon directeur de thèse, Jean-Antoine Désidéri, d'avoir accepté de suivre mon travail depuis la magnifique garrigue de Sophia Antipolis. Mes déplacements se sont toujours soldés par une liste d'éléments à creuser qui me redonnaient de la motivation.

Je souhaite remercier Pierre Alain Boucard pour le temps accordé à la rédaction d'un rapport détaillé et pertinent sur ce travail. Les remarques m'ont apporté du courage pour la préparation de la soutenance. Je remercie Régis Orobtschouk, également rapporteur de ce travail, pour s'être intéressé aux aspects mathématiques de cette thèse malgré sa spécialisation en photonique.

Mes remerciements vont ensuite à Delphine Sinoquet et Rodolphe Le Riche qui ont accepté de faire le déplacement à Grenoble afin d'examiner mes travaux.

Pour terminer sur les membres de mon jury, je remercie tout particulièrement mon co-directeur de thèse et encadrant au sein du CEA LETI de Grenoble, Alain Glière, pour m'avoir supporté sur l'ensemble de mon doctorat. Tu as grandement participé à l'amélioration de ma culture générale en sciences physiques et en modélisation numérique. Au terme de cette formation, je peux également affirmer avoir appris auprès de toi la rigueur scientifique et l'importance de fournir des résultats reproductibles. Sur un plan plus personnel, je me souviendrai longtemps de ce jour où nous nous sommes envoyés en l'air, pour effectuer un vol de planeur au-dessus du Vercors, de la Chartreuse et de Belledonne. Merci également pour l'agréable invitation annuelle à un délicieux repas chez toi. Je suis heureux de pouvoir dire que nous avons réussi à maintenir une bonne entente entre nous tout au long de cette expérience malgré des périodes difficiles. C'est une preuve de tes qualités humaines.

Durant mon doctorat, j'ai eu l'occasion de superviser le stage de Master recherche de Thomas Braure. J'ai apprécié la qualité de son travail et le remercie sincèrement pour les apports mentionnés dans ce manuscrit.

Je souhaite remercier les membres de mon laboratoire d'accueil, le laboratoire des capteurs optiques, pour l'ensemble des discussions techniques qui m'ont fait découvrir le milieu de la photonique. J'ai pu observer de nombreux projets et la volonté des personnes responsables de mener à terme leurs recherches.

Bien que mes travaux de doctorat aient été initiés en Octobre 2014, j'ai commencé à travailler sur ce sujet en 2013. Pendant deux stages de six mois aux Etats-Unis et en Allemagne, j'ai pu découvrir la métamodélisation grâce au suivi académique de Jean-Marc Bourinet. Je profite donc de ce chapitre pour le remercier car sans lui je n'aurais certainement pas travaillé dans la recherche. Je tiens également à remercier Mathieu Balesdent, qui a encadré mon stage de fin d'études d'ingénieur à l'Onera et permis l'obtention de ma première publication sur le domaine.

Je vais maintenant consacrer la seconde partie de ces remerciements à mes proches, hors contexte professionnel.

Je commence par les « anciens » doctorants grenoblois, pour m'avoir intégré rapidement dans le groupe. Merci Anthony, Hélène, Julien, Justin, Shayma (ordre alphabétique, sinon tout le monde sait que j'aurai mis le vilain A. en dernier !), vous êtes responsables de beaucoup de moments agréables. Pour Adrien (Rrruuuu), on se souviendra du Top gun. Viennent ensuite les affreux Boris et Nicolas, avec qui j'ai passé de magnifiques instants dans un nouveau bureau de qualité. Ils ont été d'une patience avec moi-même inversement proportionnelle à leur beauté physique.

A mon arrivée à Grenoble, j'ai passé un an en collocation dans une ferme équestre avec Hervé et Cathy que j'inclus avec plaisir dans ces remerciements.

Mention spéciale à Rudy Chocat, mon tantôt adversaire tantôt partenaire de travail à mes débuts. Je suis content de te retrouver en région parisienne pour de prochaines discussions animées. J'en profite pour remercier les participants de notre conversation des thésards, qui permet aux jeunes ifmaliens de se soutenir durant leurs travaux de recherche.

Toute une équipe de mes premières amitiés à Clermont-Ferrand a plus ou moins suivi mon parcours scolaire. J'ai passé de bon moment à me défouler sur les jeux vidéo avec eux ou à diverses activités sportives. Mention spécifique à cette fine bande de « feedeur », qui se reconnaîtront si leur curiosité les amène ici.

Pour terminer, il faut remercier mes parents sans qui j'aurais certainement intégré la Gendarmerie Nationale après mon baccalauréat. Merci pour le soutien physique, morale, financier et ... scolaire (blague). Sans vous, je serais peut-être en train de mesurer des excès de vitesse au lieu d'avoir la chance de travailler sur les thématiques scientifiques qu'offre mon nouvel emploi. Merci à mes sœurs pour me laisser être le plus diplômé de la famille. Cécile, je t'excuserais si tu ne pleures pas de joie. Emilie, on retiendra que l'éducation nationale ne te laisse pas avoir le droit de venir assister à la soutenance de ton frère malgré l'excellente qualité de ton travail... Et enfin mention particulière à Blandine, ma compagne qui a réussi à commencer son doctorat après moi et à le finir avant moi !





---

---

# Table des Matières

---

---

Remerciements.....	5
Table des Matières.....	8
Introduction.....	12
I. Cas d'application .....	16
A. La Photonique.....	16
1. Historique.....	16
2. Guidage optique.....	17
3. Composant d'optique intégrée et fabrication.....	18
B. Contexte de l'étude.....	21
1. Du phénomène physique au modèle numérique.....	21
2. Définition de l'optimisation numérique en conception.....	22
3. Diversité des problèmes rencontrés.....	22
C. Cas d'étude 1 : la microsource infrarouge.....	22
1. Fonctionnement du capteur optique.....	23
2. Description de la source.....	24
3. Modélisation de la source.....	24
4. Problématique de conception / optimisation.....	26
D. Cas d'étude 2 : la cellule photoacoustique.....	26
1. Principe de fonctionnement du capteur.....	27
2. Description de la cellule photoacoustique étudiée.....	27
3. Modélisation du comportement acoustique.....	29
4. Problématique de conception / optimisation.....	31
E. Cas d'étude 3 : le coupleur directionnel adiabatique.....	31
1. Description du composant.....	31
2. Modèle numérique.....	32
3. Problématique d'optimisation.....	32
F. Motivations.....	33
II. Modèles de substitution .....	35
A. Définitions.....	35
B. Métamodèles pour la simulation à un niveau donné de fidélité.....	36
1. Plan d'expérience.....	37
2. Krigeage.....	39
3. Fonctions à base radiale (RBF).....	51

4.	Critères de validation du métamodèle.....	56
C.	Métamodèles multifidélité.....	57
1.	Co-krigeage par modèle autorégressif.....	58
D.	Etude comparative dans le cadre des d'applications considérées.....	60
1.	Microsource infrarouge.....	60
2.	La cellule photoacoustique.....	66
3.	Le coupleur adiabatique.....	72
E.	Conclusion.....	75
III.	Stratégies adaptatives pour l'optimisation avec métamodèle.....	76
A.	Réduction de l'erreur de prédiction.....	77
1.	MSE (Mean Squared Error).....	77
2.	IMSE (Integrated Mean Squared Error).....	78
B.	Optimisation globale à un objectif.....	78
1.	Algorithmes d'optimisation du sous-problème.....	79
2.	Critères pour le krigeage.....	82
3.	Critères pour les RBF.....	84
4.	Comparaison des critères d'optimisation non contrainte sur un exemple analytique.....	87
5.	Optimisation sous contraintes.....	89
C.	Optimisation multi-objectif.....	92
1.	Algorithmes d'optimisation.....	94
2.	Critères pour stratégie adaptative.....	99
D.	Optimisation robuste.....	103
1.	Introduction.....	104
2.	L'optimisation robuste par stratégie adaptative.....	112
E.	Conclusion.....	113
IV.	Contribution méthodologique pour traiter les cas d'application.....	114
A.	Optimisation de la microsource infrarouge.....	114
1.	Optimisation de la puissance rayonnée.....	114
2.	Optimisation de la température moyenne.....	116
3.	Optimisation de l'efficacité.....	117
4.	Bilan.....	119
B.	Métamodélisation multifidélité pour la cellule photoacoustique.....	120
1.	RBF multifidélité (co-RBF).....	121
2.	Comparaison des métamodèles multifidélité sur modèles analytiques.....	122
3.	Stratégie adaptative pour la RBF multifidélité.....	126
4.	Application au cas de la cellule photoacoustique.....	126
5.	Discussion.....	130

C.	Optimisation robuste du coupleur selon la longueur d'onde.....	131
1.	Problématique initiale .....	132
2.	Métamodèle pour paramètres discrets.....	135
3.	Optimisation du gradient de l'amélioration espérée .....	141
4.	Application de la stratégie générale.....	143
5.	Conclusion.....	145
	Conclusion .....	146
	Bibliographie .....	149
V.	Annexe boîte à outil « SBDO ».....	154

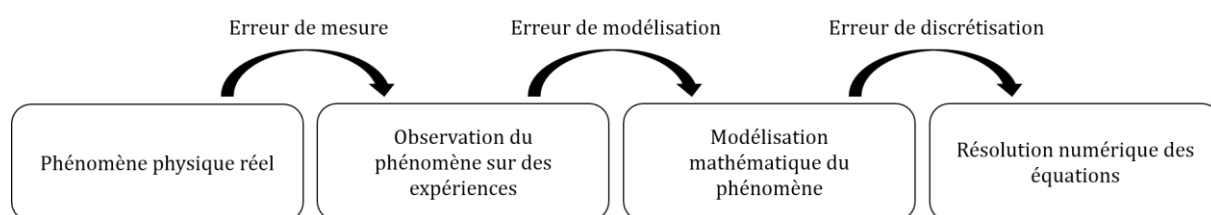


---

# Introduction

---

La modélisation numérique constitue actuellement un outil indispensable pour concevoir des systèmes complexes. Cette démarche scientifique consiste à représenter des phénomènes physiques réels par des équations mathématiques. Ces équations sont ensuite résolues pas à pas, dans le temps et l'espace, par des moyens de calcul informatique. Les résultats obtenus sont ensuite confrontés avec des mesures effectuées sur des expériences mettant en jeu les phénomènes modélisés. Les paramètres physiques intervenant dans le modèle numérique peuvent alors être modifiés afin de faire correspondre les données simulées à celles obtenues par expérimentation. Une fois que cette étape de calibration est réalisée, le modèle est complètement exploitable pour prédire le comportement du système étudié. Le processus de modélisation numérique, qui est entaché de sources d'erreurs qui peuvent s'additionner entre le phénomène physique réel et la solution obtenue par simulation, est résumé sur le graphique suivant :



**Figure : Processus de modélisation numérique. L'erreur de mesure est liée à l'expérimentation qui ne permet parfois que d'observer une partie de la réalité du phénomène physique. L'erreur de modélisation peut être réduite par calibration. L'erreur de discrétisation est liée aux méthodes numériques nécessaires à la résolution informatique des équations.**

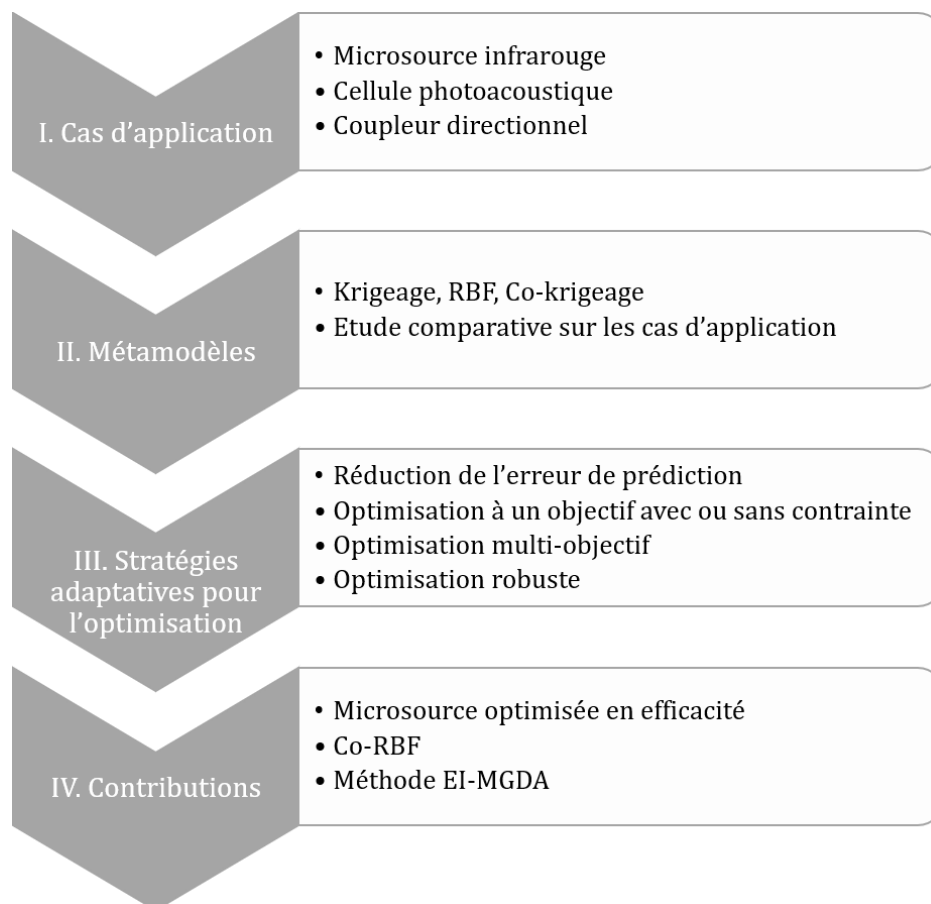
Dans le cadre d'une démarche de conception, la modélisation numérique présente plusieurs avantages. Le premier concerne la réduction des coûts de conception. En effet, la conception d'un système implique la réalisation de prototypes afin de mener des essais expérimentaux. Selon les matériaux et les procédés de fabrication mis en jeu, le prototype peut être très coûteux. Les expériences de tests peuvent également être complexes à mettre en place. Il s'agit là d'un autre avantage de la simulation : pouvoir réaliser des expériences difficilement praticables en réalité. Enfin le modèle numérique constitue également un bilan du savoir disponible sur un phénomène physique particulier. Lorsque la capacité de prédiction de la simulation est attestée sur un système particulier, le modèle peut être réutilisé pour d'autres conceptions impliquant le même type de physique. Le principal désavantage des simulations est le temps de calcul. L'accroissement exponentiel de la puissance des moyens informatiques au cours des dernières décennies a laissé penser que le coût de calcul des simulations diminuerait en termes de temps. Au contraire, les simulations deviennent de plus en plus complexes (avec des modèles multi-physiques ou des domaines plus étendus) et exploitent totalement les ressources informatiques disponibles. Par exemple, la résolution des équations de Maxwell par différences finies dans le domaine temporel peut nécessiter une journée complète de calcul selon la taille du domaine simulé. De nombreux appels à ce type de simulation ne peuvent être viables à l'échelle de projets industriels de conception. Pour pallier ce problème, une des solutions actuelles consiste à remplacer le modèle numérique par un modèle de substitution, appelé également métamodèle. Ces modèles de modèles sont capables de prédire la réponse de la simulation instantanément sur de nouvelles expériences à partir de quelques évaluations préalables. Leur utilisation permet de réduire considérablement l'impact du temps du calcul de la conception par simulation coûteuse, et ainsi de rendre abordables les algorithmes d'optimisation numérique.

Dans le cadre de cette thèse, se déroulant au sein du département d'optique et de photonique du CEA - LETI, les activités de recherche ciblent des marchés industriels majeurs tels que l'imagerie, l'éclairage, la transmission de données et les capteurs environnementaux. La conception de composants innovants dans ces secteurs d'activités est liée à l'intégration de fonctionnalité optique sur des puces électroniques. Dans

ce contexte de recherche technologique, de nombreux phénomènes physiques sont mis en jeu au travers des simulations numériques coûteuses. Ces modèles doivent être exploités afin d'analyser le comportement du système à concevoir, d'optimiser ses performances et de garantir son fonctionnement une fois le composant fabriqué. Le métamodèle permet de définir des stratégies d'évaluations parcimonieuses des simulations applicables à ces différentes problématiques. Pour l'optimisation, la spécificité du problème (le nombre d'objectif, la présence ou non de contrainte, etc.) peut nécessiter l'utilisation de stratégies adaptées.

Trois cas d'études représentatifs de la grande diversité des applications traitées au sein du département ont été identifiés afin de cerner la problématique de ce travail et de guider les développements méthodologiques effectués. L'analyse de ces problèmes de conception a permis dans un premier temps d'identifier la métamodélisation multifidélité comme un moyen permettant de réduire le coût de l'exploitation des simulations. Cette méthodologie correspond au rassemblement de l'information fournie par deux modèles numériques de fidélité différente sur le même dispositif. L'intérêt est de faire appel le plus possible au modèle le moins coûteux des deux (également le moins précis) tout en garantissant la précision de la prédiction globale avec un nombre réduit d'évaluations du modèle haute fidélité. Suite à ces travaux, un nouveau métamodèle utilisant les fonctions à base radiale (RBF) dans un cadre multifidélité est proposé. La deuxième problématique extraite des cas d'application est la prise en compte des incertitudes de fabrication dans la conception de dispositifs photoniques. L'optimisation des performances de composants en tenant compte des écarts observés entre la géométrie désirée et la géométrie obtenue en fabrication a nécessité le développement d'une méthode spécifique.

Le travail de cette thèse, inscrit dans une démarche de recherche reproductible [1], a également été capitalisé dans une boîte à outil logicielle, décrite en annexe, permettant notamment de répéter les résultats obtenus et présentés dans la suite de ce mémoire. Le plan du manuscrit est résumé sous forme graphique sur la figure ci-dessous, avant d'être précisé dans les paragraphes suivant.



Le premier chapitre du manuscrit expose le contexte de ce travail. Les principes physiques, les composants de base et les méthodes de fabrication de l'optique intégrée sont d'abord introduits. La diversité des problèmes rencontrés est ensuite spécifiée avec la présentation des trois cas représentatifs. La microsource infrarouge est le premier composant, il est utilisé dans un capteur de gaz bas coût. Son comportement est modélisé par un couplage thermoélectrique. Le deuxième dispositif concerne également la détection de gaz mais adapté à des concentrations plus faibles. Ce dispositif est une cellule photoacoustique pour laquelle nous disposons de deux modélisations, l'une reposant sur la mécanique des fluides et l'autre sur une analyse simplifiée de type acoustique. Enfin le dernier cas est celui d'un coupleur directionnel adiabatique, un composant réalisant une fonction fréquemment utilisée dans les systèmes d'optique intégrée. Son modèle est basé sur une résolution de l'équation de propagation d'Helmholtz, obtenue à partir des équations de Maxwell régissant les phénomènes optiques.

Le second chapitre se concentre sur les modèles de substitution. La méthodologie générale concernant leur construction est présentée puis deux méthodes particulières sont détaillées, à savoir le krigeage et les fonctions à base radiale. Un métamodèle construit à partir de deux modélisations de fidélité différentes pour une même conception est également exposé (le co-krigeage). Une étude comparative sur la précision des métamodèles pour chaque cas d'application est ensuite menée. Les différentes possibilités de réglage de chaque méthode sont envisagées pour déterminer le meilleur choix possible de substitution des différents modèles numériques.

Le troisième chapitre traite des stratégies adaptatives. Ces procédures itératives permettent d'améliorer la prédiction du métamodèle sur des zones particulières de l'espace de variation des paramètres de conception. Quatre types de stratégies sont présentés : la réduction de l'erreur globale de prédiction, l'optimisation à un objectif avec ou sans contraintes, l'optimisation multi-objectif et l'optimisation robuste. Ces différentes problématiques sont bien adaptées à la conception par simulation des composants constituant nos cas d'études.

Le quatrième et dernier chapitre, qui synthétise les apports principaux de ce travail, décrit la mise en place de différentes méthodes permettant d'optimiser la conception des trois dispositifs d'étude identifiés dans le premier chapitre. Pour la microsource infrarouge, les géométries résultantes de l'optimisation sous contraintes de trois objectifs différents sont comparées. En ce qui concerne la cellule photoacoustique, un métamodèle multifidélité basé sur les fonctions à base radiale est proposé. Ses performances sont confrontées à celles du métamodèle multifidélité de référence, le co-krigeage, avant de réaliser l'optimisation du composant. L'optimisation robuste du coupleur vis-à-vis des incertitudes de fabrication liées à sa géométrie est ensuite étudiée. La stratégie identifiée au troisième chapitre pour ce type de problème ne permet pas d'obtenir un résultat satisfaisant. Suite à la réduction de dimension du problème par discrétisation d'une variable, deux axes de développement sont alors envisagés. Le premier repose sur la construction d'un métamodèle adapté aux problèmes à variables mixtes. Le deuxième concerne la mise en place d'une méthode multi-objectif pour garantir l'obtention d'un minimum robuste.

Ce travail a donné lieu aux publications suivantes :

- C. Durantin, J. Rouxel, J. A Désidéri, A. Glière. Multifidelity surrogate modeling based on radial basis functions. *Structural and Multidisciplinary Optimization*, 2017, pp 1-15.
- K. Hassan, C. Durantin, V. Hugues, B. Szelag, A. Glière. Robust silicon-on-insulator adiabatic splitter optimized by metamodeling. *Applied Optics*, 2017, 56(8), pp 2047-2052.
- E. Lesmanne, R. Espiau de Lamaestre, S. Boutami, C. Durantin, L. Dussopt, G. Badano. Multispectral Detection with Metal-Dielectric Filters: An Investigation in Several Wavelength Bands with Temporal Coupled-Mode Theory, *Journal of Electronic Materials*, 2016, 45(9), pp 4603-4606.
- C. Durantin, J. Marzat, M. Balesdent. Analysis of multi-objective Kriging-based methods for constrained global optimization, *Computational Optimization and Applications*, April 2016, 63(3), pp 903-926.

et aux communications orales suivantes (orateur souligné) :

- C. Durantin, K. Hassan, A. Glière. Robust optimization of adiabatic power splitter. WCSMO 12, Braunschweig, Germany, 5-9 June 2017.
- C. Durantin, J. Rouxel, J.-A. Désidéri, A. Glière. Optimization of a photoacoustic gas sensor using multifidelity RBF metamodeling, MASCOTNUM 2017, Paris, France, 22-24 Mars 2017.
- C. Durantin, J. Rouxel, J.-A. Désidéri, A. Glière. Optimization of a photoacoustic gas sensor using multifidelity RBF metamodeling, ECCOMAS Congress 2016, Crete Island, Greece, 5-10 June 2016.
- E. Lesmanne, R. Espiau de Lamaestre, S. Boutami, C. Durantin, L. Dusopt, G. Badano. Infrared multispectral detection with metal-dielectric filters: a model base on the temporal coupled mode theory, OPTRO 2016, Paris, France, 2-4February 2016.
- J. Favreau, C. Durantin, J.-M. Fédéli, S. Boutami, G.-H. Duan. Suspended mid-infrared fiber-to-chip grating couplers for SiGe waveguides, Optical Interconnects XVI, San Francisco, United States, 13-17 February 2016.



---

# I. Cas d'application

---

Le développement d'un composant photonique se déroule en trois phases principales : la conception, la fabrication et la caractérisation. La conception consiste à dimensionner le produit en définissant la géométrie, les matériaux et d'autres paramètres physiques assurant la fonctionnalité du composant vis-à-vis d'un cahier des charges. La fabrication du composant correspond à une réalisation d'étapes successives de procédés technologiques jusqu'à l'obtention de la géométrie désirée. Une fois le composant produit, la caractérisation permet d'analyser sa fonctionnalité. Les données obtenues en caractérisation sont ensuite confrontées aux performances attendues lors de la conception, permettant de recalibrer les modèles de conception avant de lancer si nécessaire un second cycle de développement. Une interaction entre les acteurs des trois phases est nécessaire avant de finaliser la conception, afin de valider les choix des paramètres physiques pour simplifier la fabrication du produit et faciliter sa caractérisation. Cette étude se déroule au niveau de la phase de conception du composant. Pour aborder les cas d'application choisis, une introduction au domaine de la photonique est d'abord proposée.

## A. La Photonique

### 1. Historique

Le développement de l'industrie de la microélectronique a permis le lancement d'une nouvelle révolution industrielle. Celle-ci commence avec l'invention du transistor en 1947 et la réalisation du premier circuit intégré en 1958. Les transistors sont des interrupteurs permettant la réalisation de circuits de traitement numérique d'une information binaire. Plus le nombre de transistors sur un circuit est important, plus sa complexité et donc ses fonctionnalités sont grandes. Ainsi la réduction de la taille du transistor va permettre de limiter l'encombrement des circuits, assurer un meilleur rendement de production et augmenter les capacités techniques des systèmes microélectroniques. La course à la miniaturisation a donc été lancée afin d'améliorer les performances des circuits intégrés. L'augmentation du nombre de transistors par circuit a été formalisée en 1965 par G. Moore [2] donnant lieu à la fameuse loi de Moore. Cette loi part de la constatation du doublement de la complexité des circuits tous les ans pour un coût constant et extrapole cette tendance pour les dix années à suivre. En 1975, cette tendance est réévaluée pour atteindre un doublement du nombre de transistors sur un même circuit tous les 18 mois. Cette loi empirique a été vérifiée jusqu'en 2016, où l'arrêt de la miniaturisation rapide est proclamé avec des transistors comportant une longueur de grille de 14 nm [3]. Les objectifs atteignent des limites physiques et technologiques, comme par exemple la température générée par tous les systèmes du circuit concentré dans un volume réduit. L'échelle actuelle implique également l'apparition d'effets quantiques aléatoires rendant le composant peu fiable.

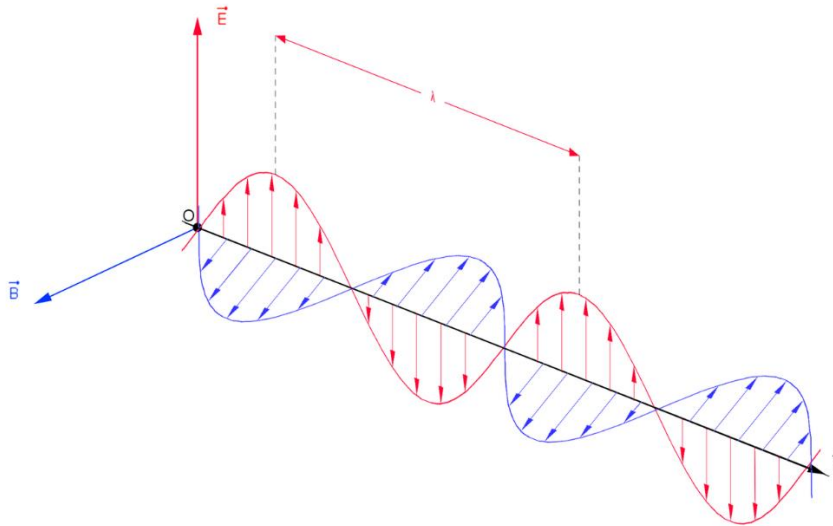
En parallèle, l'optique moderne a commencé en 1960 avec l'invention du premier laser par Maiman [4]. On parle d'optique intégrée en 1969, où Miller [5] prédit la miniaturisation des systèmes optiques grâce aux techniques de fabrication de la microélectronique. Les fonctions optiques essentielles pour créer des systèmes seraient toutes intégrées sur un circuit et reliées entre elles par des guides d'ondes. Le dispositif deviendrait alors moins sensible à certains effets thermiques ou mécaniques rencontrés sur les bancs d'optiques et réduirait fortement son coût de production. L'apparition de la première fibre optique avec de faibles pertes [6] permet d'envisager des systèmes de communications très rapides. L'intégration sur circuit sera alors une solution pour réduire la taille des systèmes liée aux fibres, tel que les répéteurs définis par Miller.

Si le développement des composants d'optique intégrée a été fortement lié au domaine des télécommunications, cette avancée autorise la combinaison de composants électroniques, électro-optiques

et optiques pour diverses applications. Ce mélange de disciplines lié à l'optique s'appelle désormais la photonique (pour le contrôle du photon, similaire à l'électronique avec l'électron).

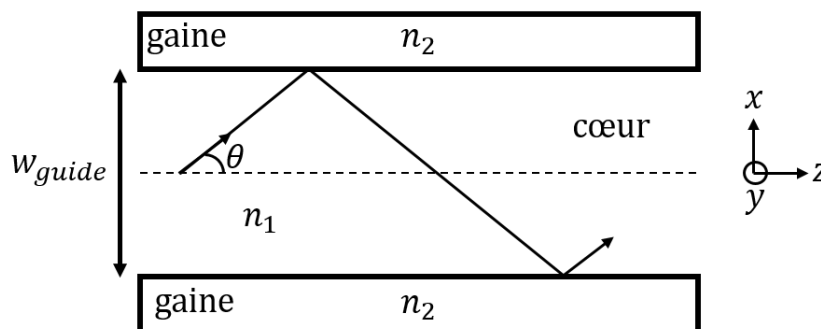
## 2. Guidage optique

Les principes physiques de base permettant d'appréhender le domaine de la photonique sont ici brièvement présentés. Le premier élément à aborder concerne la lumière, décrite comme une onde électromagnétique caractérisée par un champ électrique  $\vec{E}$  et un champ magnétique  $\vec{B}$ . Sur la Figure I-1, cette onde est représentée avec  $\lambda$  la longueur d'onde de la lumière correspondant à la période des champs  $\vec{E}$  et  $\vec{B}$ . Le vecteur d'onde  $\vec{k}$  suit la direction de propagation de l'onde et sa norme est reliée à l'indice optique du matériau dans lequel l'onde se propage.



**Figure I-1 : Représentation d'une onde électromagnétique. Source : Wikiversity.**

En optique intégrée, le but est de maîtriser la propagation de la lumière en utilisant des guides d'ondes. Ce composant de base est composé d'un matériau formant le cœur, où la lumière est confinée, et entouré d'une gaine. Ainsi la lumière est propagée dans le guide grâce au principe de la réflexion totale interne. Dans le cadre de l'optique géométrique, la condition pour satisfaire ce principe s'obtient à partir de la loi de Snell-Descartes en utilisant l'angle d'incidence  $\theta$  dans le guide plan de la Figure I-2. Cet angle ne doit pas dépasser l'angle critique  $\theta_c = \arccos(n_2/n_1)$ , relation existant seulement si l'indice de réfraction  $n_1$  du cœur est supérieur à celui de la gaine  $n_2$ . Dans cet exemple, les parties supérieure et inférieure de la gaine sont composées du même matériau mais il est également possible d'avoir un empilement de couches de différents matériaux.



**Figure I-2 : Rayon réfracté dans un guide plan.**

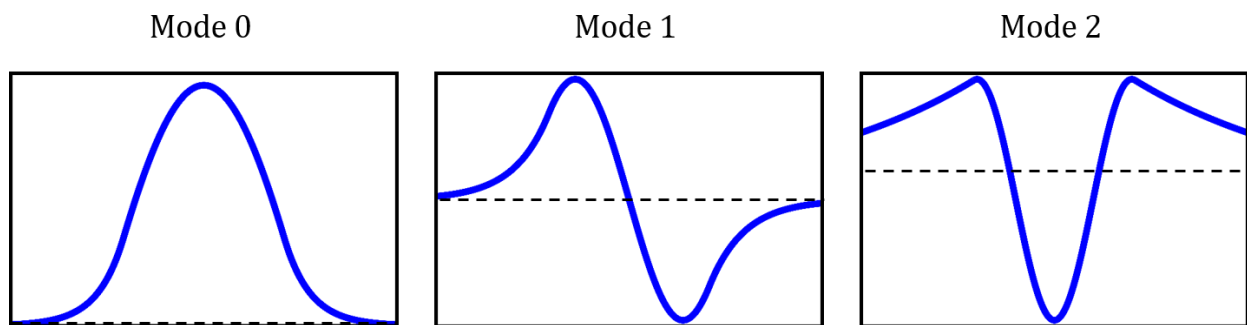
Quand bien même la condition sur l'angle critique serait respectée, cela ne serait pas suffisant pour que la

lumière ayant un angle admissible soit guidée. Il faut également que les différents chemins optiques empruntés respectent la condition d'accord de phase et forment des interférences constructives. L'équation de dispersion [I-1] est obtenue à partir de cette condition et montre que la propagation dans le guide n'est possible que pour des valeurs discrètes de l'angle d'incidence  $\theta$  :

$$\frac{2\pi}{\lambda} w_{guide} \sqrt{n_1^2 - (n_1 \cos \theta)^2} - 2 \arctan \left( g \frac{\sqrt{(n_1 \cos \theta)^2 - n_2^2}}{\sqrt{n_1^2 - (n_1 \cos \theta)^2}} \right) = m_\theta \pi, \quad [I-1]$$

$$m_\theta = 0, 1, 2, \dots$$

avec  $w_{guide}$  la largeur du guide,  $g$  une constante dépendant de la polarisation de l'onde. On appelle alors mode guidé toute onde ayant un angle d'incidence solution de cette équation. On parle de guide monomode lorsque ses caractéristiques (indices du cœur et de la gaine, largeur) sont telles qu'il n'admet qu'un seul mode guidé. Le profil des trois premiers modes d'un guide plan quelconque est représenté sur la Figure I-3.



**Figure I-3 : Composante  $E_y$  du champ électromagnétique calculée pour les trois premiers modes d'un guide plan arbitraire. Le mode 0 est appelé mode fondamental et la ligne pointillée représente  $E_y = 0$ .**

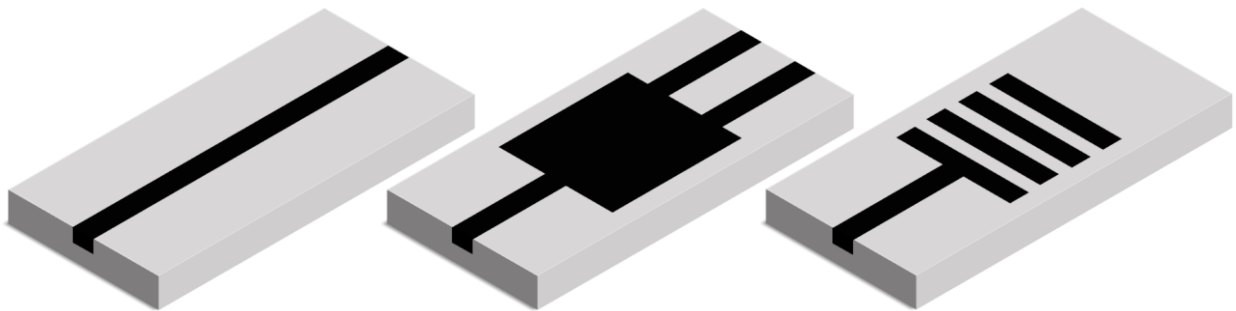
Les modes guidés dans un guide plan peuvent être décrits comme la combinaison linéaire de deux modes transverses. Le mode TE (transverse électrique) se décrit par une absence de champ électrique dans le plan de propagation de l'onde (ici le plan  $xz$ ) et ne fait donc intervenir que les composantes  $E_y$ ,  $H_x$  et  $H_z$  du champ électromagnétique. Le mode TM (transverse magnétique) correspond à une absence de champ magnétique dans le plan de propagation, ne faisant intervenir que les composantes  $H_y$ ,  $E_x$  et  $E_z$  du champ électromagnétique. L'indépendance des composantes du mode TE avec celles du mode TM n'est plus vraie dans le cas général (géométrie 3D). On peut tout de même se ramener à cette formulation en négligeant les composantes mineures du champ. On parle alors de mode quasi-TE ou quasi-TM.

Pour expliquer les phénomènes de guidage, l'optique géométrique n'est pas l'unique vision possible. La théorie de l'électromagnétique basée sur les équations de Maxwell permet d'appréhender la propagation de la lumière. Une explication plus détaillée de cette théorie appliquée aux guides d'ondes est disponible dans le chapitre 1 et 2 du livre d'Okamoto [7]. La compréhension de l'existence de modes guidés est suffisante pour aborder la réalisation des composants d'optique intégrée.

### 3. Composant d'optique intégrée et fabrication

Les procédés de la microélectronique sont utilisés pour réaliser les composants photoniques. Ainsi les dispositifs sont intégrés sur du silicium (Si) pouvant être traité par les technologies de fabrication des puces électroniques. L'élément sur lequel la fabrication commence est une plaque circulaire appelée « wafer ». Ces wafers de base sont généralement, dans le cadre de l'optique intégrée, de type silicium-sur-isolant (SOI pour silicon-on-insulator) correspondant à un empilement de silicium/dioxyde de silicium ( $\text{SiO}_2$ ) /silicium. Un exemple de dimension standard d'épaisseur de SOI correspond à 220 nm de Si sur la couche supérieure avec 2  $\mu\text{m}$  de box (buried oxyde). Les wafers ont un diamètre allant de 1 à 12 pouces (25 à 304 mm). A partir de cette base, la structure guidant la lumière est construite sur la couche supérieure du wafer par différentes opérations de fabrication impliquant l'ajout ou la mise en forme de matériaux. Comme pour l'électronique, il existe un certain nombre de composants réalisant des fonctions de base, combinées entre

elles dans de nombreux circuits photoniques. Par exemple, trois composants sont visibles sur la Figure I-4. Ils ont pour but de réaliser des fonctions optiques, comme la connexion des éléments du circuit pour le guide droit, la division de puissance optique pour le MMI ou la réflexion d'onde pour le réseau de Bragg. D'autres exemples sont présentés dans le chapitre 1 du livre de Lifante [8].



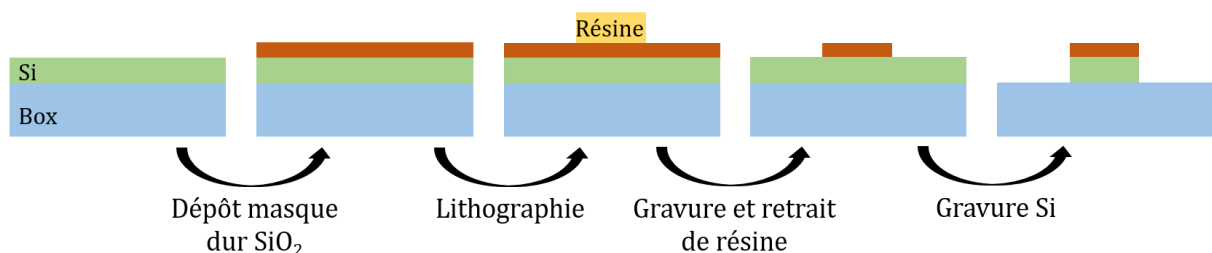
**Figure I-4 : Trois composants classiques pour la photonique. De gauche à droite, il s'agit d'un guide droit, d'un MMI (multimode interference coupler) et d'un réseau.**

Ces différentes fonctions sont assemblées sur le même circuit pour former un système optique complet. L'un des intérêts de l'intégration sur silicium est l'utilisation conjointe de composants électroniques dans le même circuit, permettant ainsi de complexifier la fonctionnalité du système pour un encombrement et un coût réduit. Par contre, le silicium ne permet pas de créer un composant actif capable de générer de la lumière. Celle-ci est généralement apportée depuis une source lumineuse (tel qu'un laser) par fibre optique et transmise au circuit par des réseaux de couplage. Un des enjeux actuels est l'intégration de matériaux actifs directement sur les circuits par collage [9].

Afin de réaliser les composants et circuits photoniques, on peut mettre en avant trois opérations principales impliquées dans les procédés de fabrication :

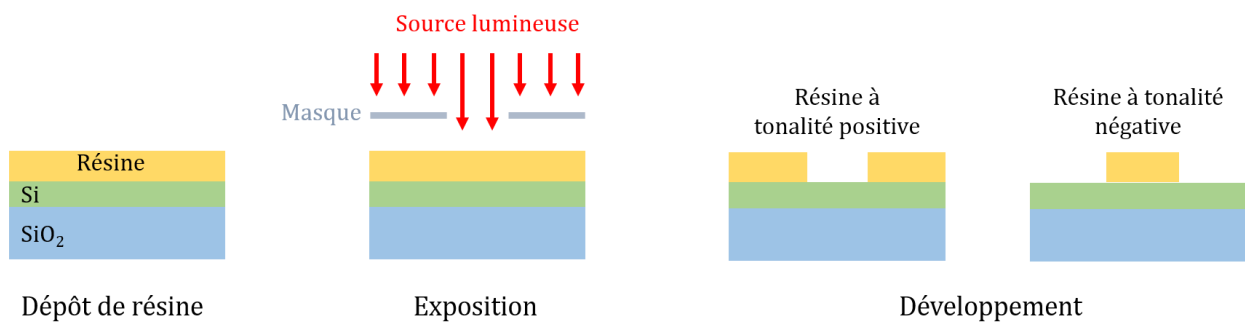
- Le dépôt est l'ajout de matière à la surface du substrat en un film de même épaisseur sur tout le wafer par croissance cristalline ou dépôt de composé amorphe.
- La lithographie est un principe d'impression de motifs sur la surface du wafer par le biais d'une résine photosensible.
- La gravure permet l'élimination de matière en surface lorsque celle-ci n'est pas protégée.

Il existe d'autres étapes possibles d'opérations sur le wafer, le chapitre 3 du livre de Senturia [10] les aborde avec plus de détails sur des problématiques de microélectronique. Pour mieux visualiser ces trois opérations, prenons en exemple la réalisation d'un guide rectangulaire schématisée à partir d'un wafer SOI sur la Figure I-5 (vue en coupe d'un plan orthogonal à l'axe de propagation de la lumière dans le guide). La première étape consiste à déposer une couche de 80 nm de  $\text{SiO}_2$  appelée masque dur. Ensuite l'étape de lithographie permet d'imprimer le motif du guide dans une résine et la gravure d'enlever la partie du masque dur non protégée par la résine. Après la gravure, la résine protectrice est également retirée. Une deuxième gravure est appliquée à partir du gabarit du masque dur pour former le guide dans le silicium. Une dernière étape non représentée ici permet à travers une oxydation thermique à 1100 °C d'encapsuler les parois verticales du guide par une couche de 10 nm de  $\text{SiO}_2$ .



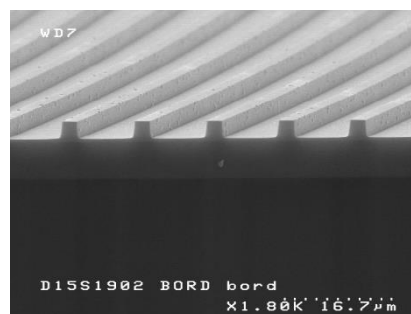
**Figure I-5 : Exemple de succession d'étapes pour fabriquer un guide d'onde silicium.**

Du fait des aléas de fabrication, la géométrie désirée initialement peut ne pas être obtenue exactement. Lors du dépôt, le temps pendant lequel le matériau est déposé influe sur la hauteur de la couche finale. Selon l'opération demandée, la vitesse de dépôt est plus ou moins maîtrisée et donc la couche déposée peut être plus épaisse qu'attendue (ou plus mince). On peut également se retrouver dans notre exemple avec une forme de guide trapézoïdale à cause de l'étape de gravure. Cette dernière peut ne pas être unidirectionnelle et donner au flanc du guide une légère pente par rapport à la verticale. Pour discuter de l'aléa de fabrication lié à la lithographie, il convient de détailler un peu plus cette étape. Les différentes phases de ce procédé sont disponibles sur la Figure I-6. Une résine photosensible est déposée sur la couche de silicium. On vient ensuite l'éclairer avec une source lumineuse les zones d'intérêt délimitées par le biais d'un masque (équivalent à un pochoir pour de la peinture). Si la résine est à tonalité positive, la partie illuminée est retirée lors d'une phase de développement. Avec une tonalité négative, la résine exposée devient insoluble et reste après le développement. La qualité de la résine, la forme du motif sur le masque et la précision de la machine employée sont des éléments importants pour la maîtrise du procédé. Selon la position du masque, le motif imprimé peut être dilaté ou rétréci par rapport à la taille souhaitée.



**Figure I-6 : Procédé de lithographie optique.**

Les différents composants photoniques réalisés par ces procédés peuvent atteindre des tailles de l'ordre du micromètre, et comportent des motifs atteignant la cinquantaine de nanomètres. La plupart des opérations sont réalisées à l'échelle du wafer, permettant de réaliser de nombreuses puces photoniques en même temps et ainsi réduire le coût final du composant. Durant la fabrication, différentes méthodes de métrologie permettent de vérifier les côtes de la géométrie obtenue. On peut prendre en exemple la microscopie électronique à balayage (MEB) qui permet de produire une image en coupe du wafer avec de hautes résolutions. La Figure I-7 montre le résultat obtenu après réalisation de guides courbés.



**Figure I-7 : Image MEB de guides rectangulaires courbés, de largeur 1  $\mu\text{m}$  [11].**

Une fois le composant fabriqué, les puces à analyser sont alors découpées sur le wafer puis amenées sur un banc de caractérisation. Celui-ci est composé d'éléments permettant de réaliser un chemin optique passant par la puce, comportant une source lumineuse et un détecteur. La lumière de la source est couplée au composant en entrée de la puce et suit le parcours imposé par le circuit photonique. En sortie de la puce la lumière est à nouveau dirigée par le banc de mesure vers un détecteur. Celui-ci permet de vérifier ainsi que la lumière a bien traversé le circuit et de mesurer les pertes de puissances optiques liées aux composants du circuit. Il existe également des moyens de mesure automatisés qui peuvent être mis en place selon les

composants utilisés en entrée et en sortie du circuit. La découpe du wafer n'est plus nécessaire dans ces cas et l'alignement de la source lumineuse avec le circuit est automatisé.

La suite de cette étude se situe au niveau de la conception de ces composants. Pourtant, la compréhension du processus de fabrication est nécessaire pour éviter les propositions de géométries irréalisables. Il est également important d'être capable d'analyser des résultats de caractérisations afin de comparer les performances d'un composant estimées lors de la conception et ce qui a été mesuré. Un recalage des modèles de conception peut alors être effectué.

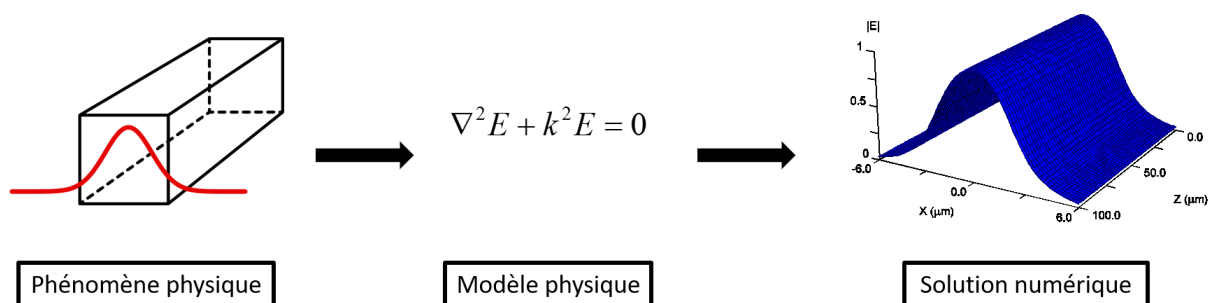
## B. Contexte de l'étude

### 1. Du phénomène physique au modèle numérique

Pour dimensionner un composant, une bonne pratique consiste à utiliser une modélisation de son comportement. L'objectif est de fournir une représentation approchée du fonctionnement de l'objet une fois réalisé. La conception commence donc par l'étude d'un phénomène physique (on prendra ici en exemple la propagation d'une onde dans un milieu isotrope). Ce phénomène est simplifié en un modèle physique (par exemple les équations de Maxwell) pouvant être basé sur une structure mathématique (telle que les équations aux dérivées partielles). Dans la plupart des cas, ces équations n'ont pas de solutions analytiques. Il faut alors recourir à la simulation numérique pour obtenir une solution, par exemple par la méthode des différences finies.

Historiquement, avant l'arrivée de moyens de calcul numérique, le dimensionnement d'un produit impliquant une physique complexe pouvait passer par un apprentissage lié à un cycle d'essai-erreur. Les modèles de conception étaient très simplifiés afin d'obtenir une solution rapidement. Si le composant fabriqué ne fonctionnait pas, un changement de conception s'effectuait par analyse de la défaillance observée en caractérisation. Ces cycles peuvent mener à un grand nombre de composants fabriqués avant d'atteindre un résultat satisfaisant. Le temps nécessaire pour le développement ainsi que le coût d'une telle opération rendent cette méthode peu rentable.

Avec les circuits intégrés, les ordinateurs personnels apparaissent dans les années 60. La programmation est facilitée par l'utilisation de langages évolués (par exemple le Fortran). Les méthodes numériques, telles que la méthode des éléments finis [12], sont alors en plein essor. Le modèle numérique apparaît alors comme l'approximation numérique et l'implémentation d'algorithmes permet de résoudre les équations du modèle physique. Les étapes du phénomène physique à la solution du modèle numérique sont représentées sur la Figure I-8.



**Figure I-8 : Du phénomène physique à la solution numérique (exemple de la propagation unidirectionnelle d'une onde lumineuse dans un guide). Le modèle physique utilisé est l'équation d'Helmholtz. L'image de la solution numérique correspond au champ électrique normalisé dans le guide calculé par « Beam Propagation Method ».**

Deux types d'erreurs interviennent durant ces étapes. Les erreurs de modélisation correspondent à l'écart entre la réalité du phénomène physique et de sa simplification choisie. Les erreurs numériques sont liées aux approximations mathématiques effectuées et aux algorithmes employés pour la résolution des équations. Il est souvent difficile de supprimer ces erreurs mais le bon comportement consiste à les estimer et les contrôler.

## 2. Définition de l'optimisation numérique en conception

Une fois le modèle numérique obtenu, il reste à exploiter l'information fournie par celui-ci. Ce modèle est composé de deux types de paramètres, le premier étant les paramètres de conception (par exemple des paramètres géométriques) sur lesquels le concepteur peut agir. Le deuxième correspond aux paramètres environnementaux, qui sont nécessaires à la modélisation mais qui ne peuvent être modifiés par le concepteur (la température de l'environnement par exemple). La réalisation du système est généralement liée à un besoin fonctionnel, tel que concevoir une « source de rayonnement infrarouge ». Ce besoin fonctionnel peut être divisé en plusieurs sous-fonctions, jusqu'à une définition plus précise des objectifs, parfois liés à des choix techniques. Du point de vue du concepteur, une grandeur mesurable doit être extraite du modèle afin de déterminer si la solution choisie correspond aux besoins fonctionnels. Ici il s'agit d'une fonction scalaire dépendant des paramètres de conception. Lorsque celle-ci détermine un critère à améliorer pour obtenir une meilleure fonctionnalité du système, elle est nommée fonction objectif (« puissance optique disponible en sortie du composant »). Lorsqu'elle permet de valider l'admissibilité de la solution technique vis-à-vis du cahier des charges (CdC) de la conception, elle est nommée fonction contrainte (« longueur totale du composant inférieure à 10  $\mu\text{m}$  »). Le problème qui se pose alors au concepteur est de modifier les paramètres de conception afin d'optimiser les objectifs sous les contraintes du CdC. Des méthodes numériques d'optimisation existent pour répondre à cette problématique et le but de ce travail est de les valider ou les adapter aux problèmes de photonique.

## 3. Diversité des problèmes rencontrés

Cette étude se déroule au sein du CEA LETI, un établissement public à caractère industriel et commercial (EPIC). Cet institut est une interface entre le milieu industriel et la recherche académique. Il a pour but de développer et transférer des technologies innovantes dans des secteurs liés aux micro et nanotechnologies. Ainsi, si l'on se focalise sur le département d'optique, on trouve des applications allant de l'étude mécanique de micro matériaux précontraints à l'analyse du transfert de puissance optique entre un réseau de couplage intégré et une fibre. Pour chaque phénomène physique rencontré, des méthodes numériques spécifiques sont mises en place pour assurer une bonne modélisation. En considérant l'exemple du phénomène de propagation unidirectionnelle d'une onde de lumière dans des guides, la « Beam Propagation Method » (BPM) [13] permet d'obtenir un modèle assez fiable et rapide à évaluer. Mais on peut tout aussi bien utiliser la méthode « Finite Difference Time Domain » (FDTD) [14] qui est plus précise mais aussi plus coûteuse. Ce choix de modèle dépend de l'expertise du concepteur et de l'erreur de modélisation tolérée pour l'étude.

Les méthodes d'optimisation numériques développées dans ce document ont pour but de permettre l'optimisation des différents problèmes qui peuvent être rencontrés dans l'environnement du CEA LETI. La diversité des phénomènes physiques impose une utilisation de méthodes non-intrusives. Ce type de méthode permet de considérer la simulation comme un modèle entrée sortie (une boîte noire). Ainsi les spécificités physiques du modèle ne sont pas directement prises en compte et la méthodologie peut s'appliquer sur différents problèmes de conception. Afin d'imager les différents types de problèmes qui devront être résolus, trois cas d'études qui servent de références sont maintenant présentés. Ces cas sont retenus parmi l'ensemble des études traitées car ils comportent des problématiques de conception nécessitant une adaptation méthodologique pour être solutionnés.

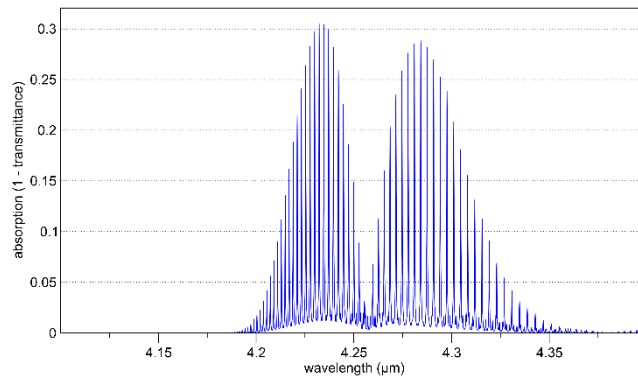
### C. Cas d'étude 1 : la microsource infrarouge

Ce cas d'étude est issu du projet de l'Agence Nationale de la Recherche (ANR) intitulé Infrared Directional EmittEr (IDEE). L'objectif du projet est de réaliser un capteur de gaz permettant de contrôler le taux de dioxyde de carbone ( $\text{CO}_2$ ) dans une pièce afin d'activer si nécessaire la ventilation. Le rôle du CEA dans ce projet est de concevoir une microsource infrarouge à intégrer dans le capteur. Dans le cadre d'une utilisation en autonomie tel que définie pour le projet IDEE, la consommation électrique du capteur, doit être faible. La puissance électrique étant principalement utilisée par la source, le travail de conception va être principalement orienté sur cette problématique tout en améliorant la puissance optique fournie par la source. Une étude complète sur la conception de cette source infrarouge est disponible dans la thèse

d'Anthony Lefebvre [15].

## 1. Fonctionnement du capteur optique

La plupart des gaz possèdent un spectre d'absorption unique et c'est sur ce principe que sont conçus les capteurs optiques. Lorsque certaines longueurs d'ondes sont absorbées par les molécules qui constituent le gaz, on peut détecter sa présence. Celui du CO<sub>2</sub> est présenté Figure I-9, où il a été calculé à partir de la base de données HITRAN [16] pour une cellule de gaz de longueur 0.01 cm et un rapport de mélange en volume de 0.1. La bande d'absorption se situe dans l'infrarouge moyen, ce qui justifie la nécessité d'une source lumineuse infrarouge.

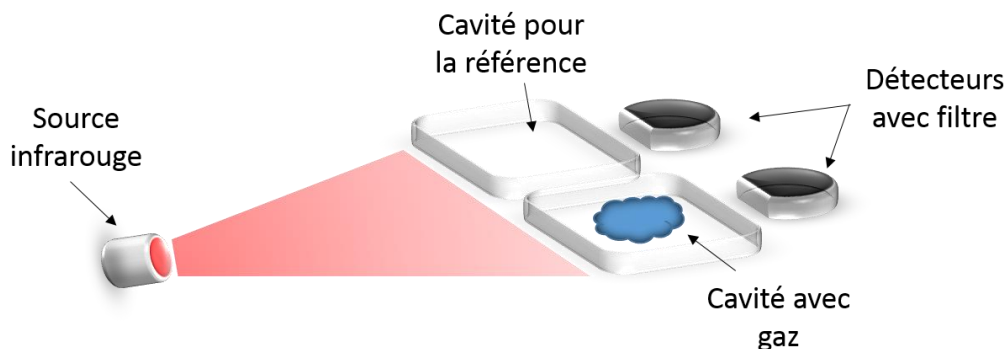


**Figure I-9 : Spectre d'absorption du dioxyde de carbone.**

Le capteur de gaz infrarouge est constitué de trois éléments principaux, à savoir une source, une cavité et un détecteur. Le rayonnement émis par la source traverse la cavité, où le gaz se situe, pour finalement arriver sur le détecteur. La loi de Beer-Lambert permet de déterminer la concentration du gaz d'intérêt  $C_{gaz}$  en fonction de l'intensité  $I$  mesurée par le détecteur :

$$I = I_0 10^{-\varepsilon_{gaz} L_{cav} C_{gaz}} \quad [I-2]$$

avec  $\varepsilon_{gaz}$  le coefficient d'absorptivité du gaz,  $L_{cav}$  la longueur de la cavité et  $I_0$  l'intensité de référence (mesure sans gaz ou intensité du faisceau incident). La totalité du rayonnement lumineux provenant de la source arrive directement sur le gaz sans rencontrer d'élément dispersif (tel qu'un prisme) auparavant, ce qui permettrait une dispersion en longueur d'onde. Cette technique s'appelle Non Dispersive InfraRed (NDIR). Pour assurer une détection sélective sur le type de gaz, un filtre est positionné sur le détecteur pour ne laisser passer que la bande spectrale désirée. Un deuxième détecteur comportant d'une cavité sans gaz peut servir de voie de référence afin de supprimer certains aléas durant la mesure (dérive de la source, variations de température). Le schéma de principe du capteur NDIR est visible sur la Figure I-10.



**Figure I-10 : Schéma de principe d'un capteur de gaz NDIR.**

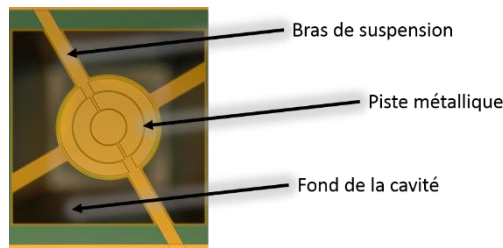
L'avantage principal de ce type de capteurs est l'isolation des éléments optiques contre le gaz, ce qui permet de travailler avec différents types de molécules sans altérer le fonctionnement du capteur. Son point faible



se situe dans la longueur de cavité qui doit être assez grande pour que le signal détecté soit suffisant (voir équation [I-2]), ce qui a pour effet d'augmenter l'encombrement du capteur.

## 2. Description de la source

La solution technique choisie pour la source de rayonnement infrarouge est une microsource incandescente. Il s'agit d'un microsystème électromécanique (MEMS, MicroElectroMechanical Systems) dont la réalisation a été rendue possible sur silicium grâce aux progrès de la microélectronique. La Figure I-11 représente la membrane fabriquée par le LETI [17], dans une version avec quatre bras de suspension. La structure est composée de plusieurs pistes métalliques permettant de chauffer la membrane par effet Joule. Pour réduire les pertes thermiques par conduction dans le substrat, la membrane est suspendue. Les bras permettent de maintenir la membrane lors de sa libération. Initialement conçue pour fonctionner à l'air libre, la source est maintenant encapsulée sous vide afin de supprimer les pertes par conduction thermique dans l'air. Entre les pistes métalliques, des homogénéiseurs métalliques sont introduit pour réduire les variations de température dans la membrane. Ces homogénéiseurs sont isolés des pistes conduisant le courant électrique afin d'éviter tout court-circuit.



**Figure I-11 : Image au microscope optique d'une membrane suspendue réalisée au LETI.**

L'empilement de matériaux formant la membrane est composé de bas en haut de : nitrure (SiN) 100 nm, nitrure de titane (TiN) 10 nm, platine (Pt) 30 nm, nitrure de titane 10 nm, silice (SiO<sub>2</sub>) 100 nm et nitrure 100 nm. L'ensemble TiN/Pt/TiN correspond aux pistes métalliques. Le diamètre du disque central actuellement utilisé est de 250 μm, relié au substrat par des bras rectangulaire de 40 μm de large et 125μm de long.

## 3. Modélisation de la source

Le comportement de la source peut être obtenu par un modèle thermoélectrique. En combinant l'équation de Maxwell-Gauss et la loi d'Ohm, la puissance dissipée sous forme de chaleur (effet Joule) peut être calculée :

$$P_{joule} = \int_{volume} \rho_E \mathbf{J}^2 dv \quad [I-3]$$

avec  $P_{joule}$  la puissance dissipée sur le volume du conducteur,  $\rho_E$  la résistivité électrique du matériau et  $\mathbf{J}$  la densité de courant.

Le modèle est utilisé en régime permanent, on peut donc écrire l'équation de la chaleur dans le cas stationnaire :

$$-k_T \Delta \mathbf{T} = P_{joule} \quad [I-4]$$

où  $k_T$  représente la conductivité thermique du matériau et  $\mathbf{T}$  le champ de température sur le domaine. La conductivité thermique et la résistivité électrique sont deux grandeurs physiques liées à la température du matériau. Une approximation linéaire de la dépendance peut être effectuée selon la formule suivante :

$$\rho_E(T) = \rho_{E_0} \left( 1 + c_p (T - T_0) \right) \quad [I-5]$$

$$k_T(T) = k_{T_0}(1 + c_k(T - T_0))$$

avec les valeurs  $\rho_{E_0}, k_{T_0}, c_\rho, c_k$  définies pour chaque matériaux dans le Tableau 1.

Matériau	$\rho_{E_0} [\Omega.m]$	$c_\rho [K^{-1}]$	$k_{T_0} [W.m^{-1}.K^{-1}]$	$c_k [K^{-1}]$
SiN			2.6	1.3e-3
TiN	57e-8	1.37e-3	19.2	9.35e-6
Pt	1.06e-7	3.92e-3	58	0.7e-3
SiO2			1.2	1e-3

**Tableau 1 : Conductivité thermique et résistivité électrique des matériaux (données extraites de [18], [19]).**

Le flux thermique radiatif émis par la surface de la membrane  $S_m$  d'émissivité  $\varepsilon_m$  vers un environnement à la température  $T_0$  s'écrit :

$$\phi_{rad}(T) = \frac{\varepsilon_m S_m}{\rho_E} (T^4 - T_0^4) \quad [I-6]$$

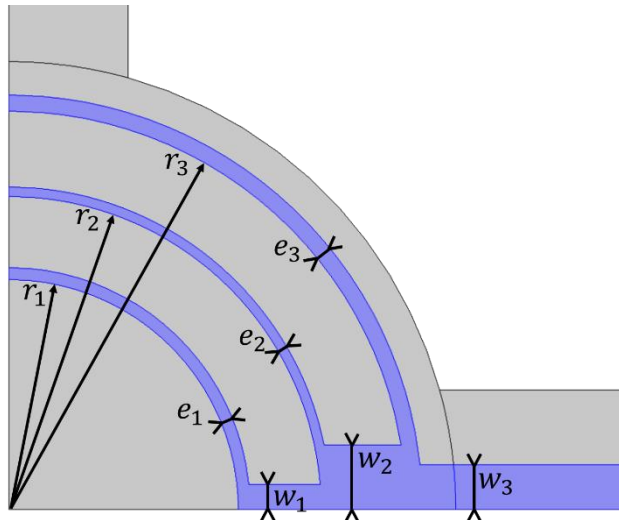
La puissance électrique consommée par la membrane est calculée à partir de l'intégrale de la densité de courant au bout des bras de suspension avec  $\vec{n}$  le vecteur normal à la section métallique et  $U_m$  la tension appliquée aux bornes de la membrane :

$$P_{elec} = U_m \iint \vec{j} \cdot \vec{n} dS \quad [I-7]$$

Le rendement de la source  $\eta_{source}$  peut donc être défini comme le rapport entre la puissance optique utile et la puissance électrique consommée. La puissance optique utile correspond dans notre cas au flux thermique radiatif émis vers le haut, donc la moitié du flux total :

$$\eta_{source} = \frac{\phi_{rad}/2}{P_{elec}} \quad [I-8]$$

La solution de l'équation [I-8] est obtenue par une simulation par éléments finis sous COMSOL (COMSOL AB, Sweden). Afin de réduire le temps de calcul, les symétries du dispositif sont prises en compte, ce qui permet de réduire le domaine d'étude. Le rayon de la membrane étant très grand devant son épaisseur ( $150 \mu m \gg 350 nm$ ), les variations verticales de température sont négligées pour se ramener à un modèle 2D. L'empilement de matériaux est alors transformé en un seul matériau équivalent (démarche d'homogénéisation). Chacune de ses propriétés physiques est une moyenne de celles des différentes couches pondérées par leurs épaisseurs. Une condition de Dirichlet  $T = T_0 = 20^\circ C$  est imposée aux bouts des bras de suspensions, représentant la liaison avec le wafer à température ambiante. La géométrie utilisée pour modèle numérique est visible sur la Figure I-12, ainsi que les paramètres de conception qui seront utilisés pour l'optimisation.



**Figure I-12 : Géométrie de la membrane chauffante pour le modèle numérique. Les pistes électriques sont représentées en bleu et sont définies par les paramètres correspondant au rayon, épaisseur de la partie circulaire et épaisseur de la partie d'arrivée par les bras. Les bras sont tronqués pour réduire la taille de l'image.**

Le temps nécessaire pour obtenir la solution du modèle numérique est de 9 secondes avec un nœud de calcul de 10 processeurs cadencés à 3 GHz.

#### 4. Problématique de conception / optimisation

L'optimisation de la position des pistes électriques est un des chemins d'amélioration possible de la source infrarouge étudiée ici. Le modèle numérique permet d'établir le profil de température sur toute la membrane et de calculer ensuite les grandeurs physiques d'intérêt. La température moyenne étant directement liée au flux radiatif, sa maximisation paraît être un objectif intéressant. Par contre, il faut éviter tout point chaud dépassant les 650 °C pour des raisons de durabilité du système. Le rendement de la source peut aussi être utilisé comme critère car il inclut une notion de consommation électrique. Si la puissance électrique consommée n'est pas une limitation, la puissance rayonnée par la microsource peut aussi être optimisée. Il y a donc trois d'objectifs possibles à optimiser indépendamment sous une contrainte de température maximale. Une autre contrainte s'applique sur les paramètres de conception pour éviter le chevauchement des trois pistes. Afin de pouvoir traiter le plus grand nombre de situation possible, les paramètres des pistes les plus à l'extérieur de la microsource influent sur les bornes des paramètres des pistes les plus au centre. L'espace dans lequel les paramètres évoluent n'est donc plus un hypercube.

Concernant la réalisation de la membrane, les procédés impliquant le dépôt des couches successives et la réalisation des motifs des pistes électriques sont soumis à des variations. En effet le choix des opérations nécessaires pour obtenir la membrane n'est pas encore standardisé, générant des différences entre les générations de deux séries différentes. La combinaison des aléas de fabrication peut amener à produire des dispositifs avec des géométries assez différentes de celles désirées. La quantification des incertitudes par une loi de probabilité estimée à partir des séries de mesures n'est pas disponible. Seules des approximations définissant un intervalle de valeurs possibles des paramètres est donné.

#### D. Cas d'étude 2 : la cellule photoacoustique

La production à grande échelle de système ainsi que la réduction des coûts de production passent par une intégration sur puce de nouveaux composants. Dans le contexte de l'optique intégrée, le laboratoire sur puce est un des axes principaux de recherche en conception de système. Ce type de circuit dispose d'une ou plusieurs fonctions permettant le recueil et le traitement d'une donnée. L'intérêt est de diminuer le coût de fabrication du dispositif d'analyse par de la fabrication de masse. La réduction de dimension permet également d'inclure le système dans différentes technologies (tel que le smartphone). Actuellement, les

capteurs optiques de précisions sont des composants manquants pour l'analyse de compositions de gaz intégrée sur puce. Dans le cadre de projets institutionnels, le CEA LETI et du Groupe de Spectrométrie Moléculaire et Atmosphérique (GSMA) ont travaillé sur la miniaturisation de capteurs de gaz fondés sur l'effet photoacoustique. Une partie de ces travaux est accessible dans la thèse de Justin Rouxel [20].

## 1. Principe de fonctionnement du capteur

Le capteur repose sur le principe d'absorption de lumière du gaz déjà détaillé au début du paragraphe I.C.1. Le principe de la photoacoustique commence par la modulation, en intensité ou en longueur d'onde, d'une source lumineuse (un laser dans notre cas). L'éclairage est alors dirigé vers le gaz d'étude contenu dans une cuve photoacoustique. L'absorption du rayonnement par les molécules de gaz produit un échauffement local et donc une variation de pression. Lorsque la fréquence de modulation de la source est faible par rapport aux fréquences de relaxation des molécules du gaz, la modulation de la température est assurée. La variation de la pression à une fréquence correspondant à la fréquence de modulation s'obtient à partir de la loi des gaz dans un volume fermé. Cette dernière est mesurée grâce à un microphone pour finalement obtenir un signal électrique. L'intérêt de ce capteur par rapport à la mesure NDIR réside dans sa performance pour mesurer des faibles traces du gaz d'étude.

Au premier ordre, le signal  $S_{PA}$  peut être dérivé de la loi de Beer-Lambert (équation [I-2]), linéarisée dans le cas des faibles concentrations :

$$S_{PA} = R_{cel} R_M P_{opt} \epsilon_{gaz} C_{gaz} \quad [I-9]$$

où  $R_{cel}$  [ $\text{Pa} \cdot \text{W}^{-1} \cdot \text{cm}$ ] est la constante de cuve (représentative du fonctionnement de la cellule vide sans gaz),  $R_M$  [ $\text{mV} \cdot \text{Pa}^{-1}$ ] la sensibilité de détection des microphones et  $P_{opt}$  [W] la puissance optique injectée dans la cellule. La constante de cuve est proportionnelle à l'inverse du volume de la cuve, ainsi lorsque la taille de la cellule est réduite, le signal mesuré augmente et permet la mesure de concentration plus faible de gaz. Ce type de capteur se prête donc bien à la miniaturisation pour une intégration sur circuit.

Le schéma de principe de la Figure I-13 présente le système nécessaire pour assurer une mesure de concentration d'un gaz à l'intérieur de la cuve photoacoustique. La détection synchrone sert à extraire du bruit le très faible signal fourni par les microphones à la fréquence du modulateur. Le puissance-mètre permet d'obtenir la puissance optique moyenne injectée dans la cellule.

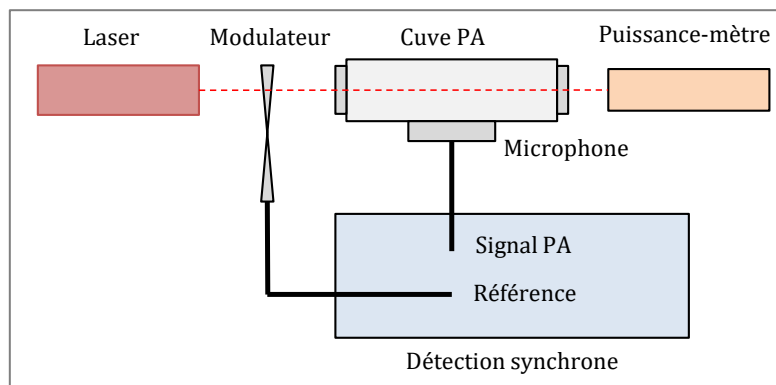
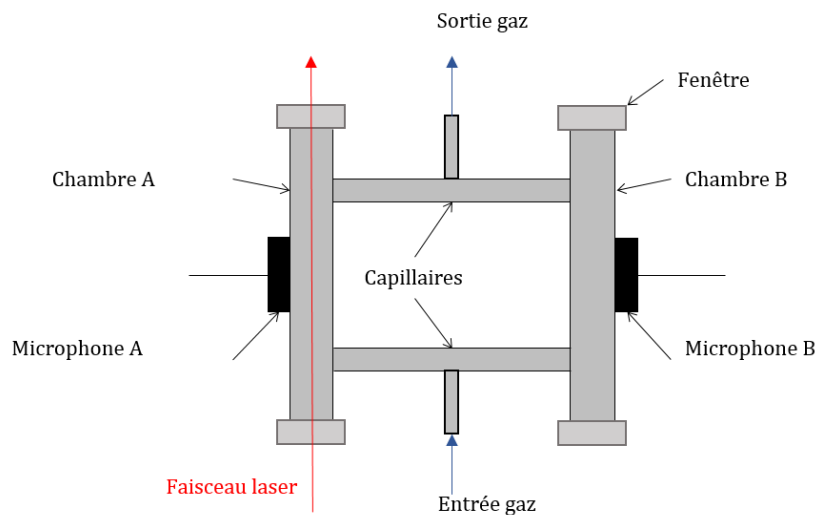


Figure I-13 : Principe de mesure de concentration de gaz avec un capteur photoacoustique [20].

## 2. Description de la cellule photoacoustique étudiée

Plusieurs types de cellules existent, utilisant différents modes de résonance. Pour cette étude, le résonateur différentiel d'Helmholtz [21] (DHR) a été choisi. La géométrie de la cavité est représentée sur la Figure I-14. La cellule est composée de deux chambres, dont une est illuminée par le laser modulé (flèche rouge). Les chambres sont reliées entre elles par deux capillaires. Un système pour l'entrée et la sortie du gaz présent dans la cuve est mis en place sur chacun des capillaires. Des fenêtres transparentes à la longueur d'onde du laser sont disposées à chaque extrémité des chambres afin de rendre la cuve étanche aux gaz environnants

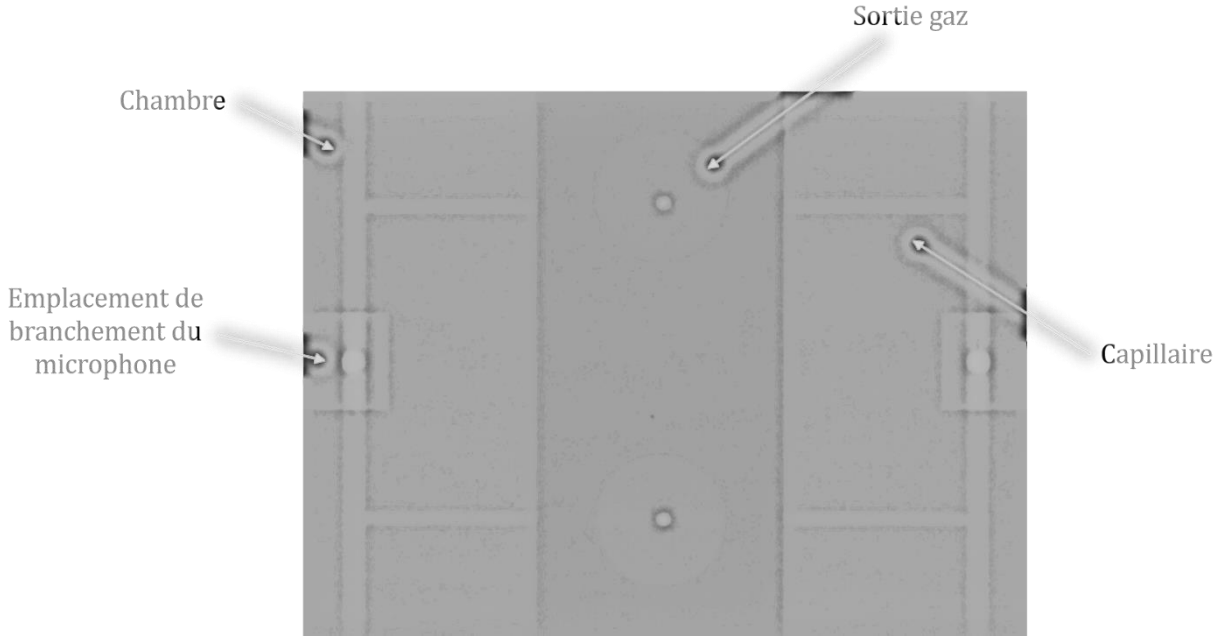
tout en laissant passer le rayonnement du laser. Un microphone est placé au centre de chaque chambre.



**Figure I-14 : Géométrie de la cavité d'une cellule photoacoustique de type DHR.**

Cette géométrie permet une résonance en opposition de phase dans les deux chambres, qui est obtenue à relativement basse fréquence. Ainsi lorsque le gaz dans la chambre A est dilaté, celui de la chambre B est comprimé. L'amplitude du champ de pression est mesurée localement par les microphones. La différence entre les signaux permet de s'affranchir du bruit extérieur et d'obtenir une meilleure sensibilité. Une détection de concentration de l'ordre du ppb (partie par milliard) est possible avec ce type de dispositif.

La fabrication de cette cellule sur circuit silicium a été lancée et comporte de nombreuses étapes. Le procédé n'est pas encore totalement au point, nécessitant un temps de réalisation du système important. Afin d'attester des performances du capteur à une échelle millimétrique, la cellule a été fabriquée par procédé d'impression 3D. Le contour de la cuve est formé par une fusion de poudres métalliques. La radiographie de la cellule sur la Figure I-15 permet d'observer la cavité ainsi réalisée.



**Figure I-15 : Radiographie de la cellule photoacoustique réalisée en impression 3D métallique. Les éléments éclaircis correspondent aux endroits du bloc métallique où il n'y a pas de matière, c'est-à-dire la cavité.**

### 3. Modélisation du comportement acoustique

Deux modélisations du comportement acoustique de la cellule sont disponibles et détaillées ci-dessous. La première correspond à une formulation complète du problème physique limité au premier ordre. La deuxième est une approximation du premier avec des pertes estimées par facteurs de qualités. Dans les deux cas, la résolution des équations analytiques se fait par la méthode des éléments finis.

#### a. Modèle Full Linearized Navier-Stokes (FLNS)

Cette modélisation part des équations de la mécanique des fluides, à savoir les équations de Navier-Stokes. Celles-ci sont linéarisées sous l'hypothèse de petites variations de pression  $p_{gaz}$ , température  $T$ , vitesse  $u_{gaz}$  et masse volumique du gaz  $\rho_{gaz}$ . Ces équations s'écrivent alors dans le domaine fréquentiel :

$$\begin{aligned}
 i\omega\tilde{\rho}_{gaz} + \nabla \cdot \tilde{\mathbf{u}}_{gaz} &= 0 \\
 i\omega\rho_0\tilde{\mathbf{u}}_{gaz} &= -\nabla \cdot \tilde{\mathbf{p}}_{gaz} + \nabla \cdot \left( \mu(\nabla \cdot \tilde{\mathbf{u}}_{gaz} + \nabla \cdot \tilde{\mathbf{u}}_{gaz}^T) + \left( \eta - \frac{2\mu}{3} \right) (\nabla \cdot \tilde{\mathbf{u}}_{gaz}) \mathbf{I} \right) \\
 i\omega\rho_0 C_p \tilde{T} &= -\nabla \cdot (-k_T \nabla \cdot \tilde{T}) + i\omega\tilde{p}_{gaz} + Q_h
 \end{aligned} \tag{I-10}$$

où les variables tildes représentent les composantes d'ordre 1 de la linéarisation, les variables d'indices 0 représentent les composantes moyennes,  $\mu$  la viscosité dynamique du fluide,  $\eta$  la viscosité dilatationnelle du fluide,  $\omega$  la fréquence de modulation,  $Q_h$  la puissance liée à la source de chaleur et  $\mathbf{I}$  le tenseur identité. Les parois de la cavité sont supposées isothermes et parfaitement rigides, définissant ainsi les conditions aux limites suivantes, où  $n$  est le vecteur unitaire normal à la paroi :

$$\begin{aligned}
 \mathbf{u} &= \mathbf{0} \\
 T_{paroi} &= 20 \text{ °C} \\
 \frac{\partial p}{\partial n} &= 0
 \end{aligned} \tag{I-11}$$

#### b. Modèle Kreuzer

Le modèle précédent permet de prendre complètement en compte les pertes surfaciques et volumiques de la cellule. Mais ces pertes ont une influence négligeable sur le signal quand l'échelle de la cuve est

macroscopique. Le processus dans la cellule peut alors être supposé adiabatique (pas d'échange thermique avec l'extérieur) et les effets visqueux négligés. Dans ces conditions, la simplification des équations [I-10] permet de se ramener à une équation d'Helmholtz inhomogène :

$$\Delta p_{gaz} + k_{onde}^2 p_{gaz} = i\omega \frac{\gamma_{gaz} - 1}{c_{onde}^2} \mathcal{H} \quad [I-12]$$

Avec  $k_{onde}$  le nombre d'onde,  $\gamma_{gaz}$  le coefficient de Laplace,  $c_{onde}$  la vitesse de l'onde acoustique et  $\mathcal{H}$  la transformée de Fourier de la densité de puissance liée à la source de chaleur.

A une échelle intermédiaire, lorsque la taille de la cellule ne permet plus de négliger les pertes, il existe différentes manières de les prendre en compte. L'utilisation de facteurs de qualité exprimant les pertes volumiques et surfaciques, issue des travaux de Kreuzer [22], est celle utilisée ici. Ce dernier décompose la solution de l'équation [I-12] sur une base de modes propres :

$$p_{gaz} = \sum_j A_j p_{jgaz} \quad [I-13]$$

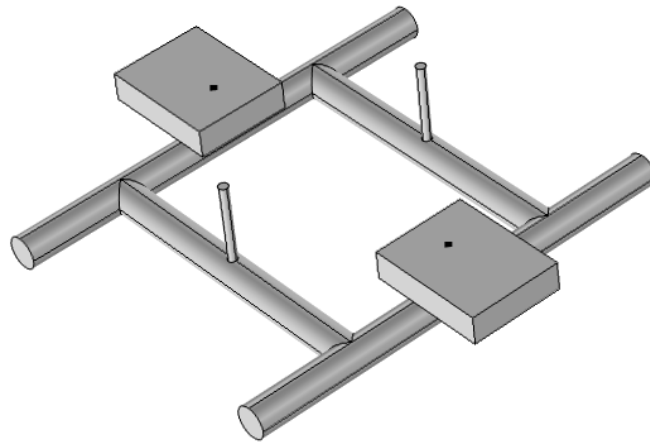
où  $p_{jgaz}$  représente les modes obtenus par résolution de l'équation d'Helmholtz homogène (cas où le terme de droite de l'équation [I-12] est égal à 0). L'amplitude  $A_j$  du signal des modes propres intègre les effets visqueux et thermiques par le biais d'un facteur de qualité  $Q_j$  :

$$A_j = i \frac{\mathcal{A}_j \omega}{\omega^2 - \omega_j^2 + \frac{i\omega\omega_j}{Q_j}} \quad [I-14]$$

avec  $\mathcal{A}_j$  l'amplitude d'excitation calculée à partir du conjugué complexe du mode propre  $p_{jgaz}$  et  $\omega_j$  correspondant aux fréquences propres. Le facteur  $Q_j$  est une somme de différents facteurs liés aux pertes. Les expressions de ces facteurs de pertes (volumique, visqueuse et thermique) sont disponibles dans la section 2.3.4 de la thèse de J. Rouxel [20].

### c. Géométrie et données physiques des modèles

La géométrie de la cavité utilisée pour les deux modèles est visible sur la Figure I-16. Les microphones sont représentés car leur volume actif conditionne la résonance de la cellule et donc le signal détecté. Les chambres, les capillaires et l'entrée/sortie du gaz sont cylindriques. Le diamètre des chambres est fixé à 1  $\mu\text{m}$ . Les paramètres de conception sont la longueur des chambres, la longueur des capillaires et leur diamètre. Le produit  $\varepsilon_{gaz} C_{gaz}$  utilisé pour calculer le signal photoacoustique est fixé à 0.1  $\text{cm}^{-1}$  et la puissance optique  $P_{opt}$  à 1 mW. Les données extraites des modèles sont la fréquence de résonance de la cellule ainsi que le signal calculé à cette dernière.



**Figure I-16 : Géométrie de la cellule photoacoustique utilisée pour les modèles numériques.**

Le temps de calcul nécessaire pour obtenir le signal par le modèle FLNS est d'environ 1h30 sur un nœud d'un cluster de calcul avec 20 processeurs cadencés à 3 GHz. Pour le modèle Kreuzer sur le même nœud, le temps est de 3 minutes. Les deux modèles fournissent une fréquence de résonance légèrement différente, celle du modèle Kreuzer étant supérieure à celle du modèle FLNS. Le signal est également plus élevé sur le modèle simplifié, dans les deux cas à cause d'une sous-estimation des pertes.

#### 4. Problématique de conception / optimisation

Une première génération de cellule a été produite avec un diamètre de chambre de 750 nm. Suite aux difficultés d'injection rencontrées sur cette génération, de nouvelles cellules avec un diamètre de chambre à 1  $\mu\text{m}$  doivent être fabriquées. L'optimisation du signal doit être effectuée en agissant sur les paramètres de conception. Le modèle FLNS est plus proche des résultats expérimentaux que celui de Kreuzer mais il est bien plus coûteux en temps de calcul. L'utilisation d'une combinaison des deux modèles pourrait permettre d'obtenir une géométrie optimale pour un coût de calcul modéré.

### E. Cas d'étude 3 : le coupleur directionnel adiabatique

Pour réaliser un circuit photonique, il est nécessaire d'avoir des composants avec des fonctionnalités basiques tels que le guide d'onde ou le réseau de couplage. Ici on s'intéresse à la division de puissance optique d'une entrée vers deux sorties obtenue grâce à un coupleur directionnel adiabatique. Le but est d'obtenir une division de puissance robuste face aux incertitudes de fabrication avec un dispositif le plus compact possible. En effet ce type de composant peut être utilisé en cascade dans des modulateurs et sa longueur influe fortement sur la longueur totale du système.

#### 1. Description du composant

Cette étude reprend la géométrie utilisée par J. Xing et al. [23], représentée sur la Figure I-17. La lumière arrive dans le guide supérieur pour être transférée en partie vers le guide inférieur par phénomène de couplage. Ce phénomène intervient lorsque deux guides d'ondes se rapprochent. Il est dû au chevauchement physique de ce qui correspond en temps normal au champ évanescent de l'onde lumineuse propagée dans un seul guide. Ce chevauchement permet le transfert de puissance optique d'un guide à l'autre. Si on prend l'exemple de deux guides droits très proches, la lumière est transférée périodiquement d'un guide à l'autre. En écartant ensuite les deux guides, le couplage n'est plus possible. Cette configuration permet de créer un diviseur de puissance optique si on maîtrise le moment où les deux guides se séparent.



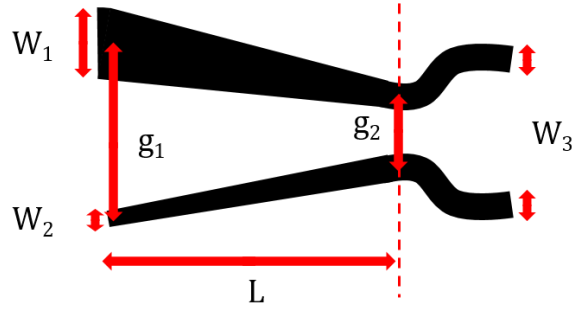


Figure I-17 : Géométrie et paramètres du coupleur directionnel adiabatique.

Les guides du coupleur sont partiellement gravés dans du silicium : seul 150 nm est gravé pour former les bords du guide sur une hauteur de couche de Si de 300 nm. Les couches supérieures et inférieures sont composées de dioxyde de silicium. Les deux guides subissent une variation de largeur linéaire de  $W_1$  à  $W_3$  pour le supérieur et de  $W_2$  à  $W_3$  pour l'inférieur. La distance séparant les milieux des deux guides initialement est définie par  $g_1$  et au plus étroit (à la fin de la zone de couplage) par  $g_2$ . Cette zone de couplage est de longueur  $L$  et doit être la plus courte possible. Après cette zone, les deux guides s'écartent par le biais de guides en S, optimisés pour réduire les pertes optiques dans les virages (voir [24] pour une description de la méthode).

## 2. Modèle numérique

A partir des équations de Maxwell simplifiées au cas des ondes continues dans un milieu diélectrique linéaire, homogène et dépourvu de charge, l'équation d'Helmholtz suivante est obtenue :

$$\Delta \mathbf{E} + k_c^2 n_c^2 \mathbf{E} = 0 \quad [\text{I-15}]$$

avec  $\mathbf{E}$  les composantes du champ électrique,  $n_c$  l'indice de réfraction et  $k_c$  le nombre d'onde. Si  $n_c$  ne varie que très peu selon la direction de propagation  $z$ , alors la solution de l'équation d'Helmholtz est en  $\mathbf{E}(x, y, z)e^{j\bar{k}_c z}$ , avec  $\bar{k}_c = k_0 \bar{n}_0$  où  $\bar{n}_0$  est un indice de réfraction de référence. Cette approximation d'enveloppe lentement variable (SVEA pour slowly varying envelope approximation) permet de réécrire l'équation [I-16] pour une composante  $E_i$  du champ :

$$\Delta \xi + 2j\bar{k}_c \partial_z \xi + (k^2 - \bar{k}_c^2) \xi = 0, \text{ avec } E_i = \xi e^{-j\bar{k}_c z} \quad [\text{I-16}]$$

La faible variation du champ selon  $z$  permet de négliger le terme  $\partial_{zz} \xi$  (approximation paraxiale). Cette modélisation de la propagation de la lumière correspond à la beam propagation method [13]. Les différentes approximations permettent une résolution par plans successifs dans la direction du vecteur d'onde à partir de valeurs initiales dans la section d'entrée du guide. La résolution peut être effectuée par éléments finis ou différence finie. Le module BeamPROP du logiciel commercial RSoft (Synopsis, USA) est utilisé ici pour obtenir la solution de l'équation [I-16]. Pour autant, notre cas d'étude est dans les limites d'utilisation de la BPM car la direction de propagation de la lumière subit des variations assez importantes. Cependant l'étude de J. Xing et al [23] a validé la même méthode de simulation pour ce type de coupleur en obtenant des résultats similaires en expérimentation.

Les indices de réfraction utilisés pour le cœur et la gaine des guides sont respectivement de 3,506 et 1,447 pour une longueur d'onde d'étude de 1.31  $\mu\text{m}$ . Le mode injecté est polarisé en quasi-TM. La taille des éléments de discrétisation est de 20 nm dans la direction de propagation et 10 nm dans les deux autres. La puissance normalisée en sortie des guides correspond à la donnée extraite en sortie du modèle. Son calcul pour une longueur d'onde donnée est d'environ 5 minutes sur 20 cœurs cadencé à 3 GHz.

## 3. Problématique d'optimisation

La conception de ce composant nécessite normalement plusieurs étapes de calcul numériques avant

d'obtenir une division de puissance sans pertes optiques. L'utilisation d'une géométrie paramétrée et d'algorithmes d'optimisation peut permettre d'obtenir le même résultat avec un coût de calcul réduit. Les objectifs à minimiser sont :

- La longueur du dispositif (avoir la plus petite valeur possible du paramètre  $L$ ).
- La différence entre la puissance normalisée en sortie de chaque guide et la valeur cible, à savoir 0.5.

En considérant  $P_1$  la puissance normalisée dans le guide supérieur et  $P_2$  celle dans le guide inférieur de la Figure I-17, l'objectif  $T_c$  à optimiser est défini par l'écart au taux de couplage souhaité :

$$T_c = \frac{P_1 - 0.5}{P_1 + P_2} \quad [I.17]$$

Ce critère est à minimiser. Les puissances normalisées  $P_1$  et  $P_2$  varient entre 0 et 1. Si le dispositif n'entraîne pas de perte de puissance optique, la somme de ces puissances doit être égale à 1. Ainsi dans le cas où la somme est inférieure à 1, le critère  $T_c$  renvoi une valeur plus élevée et pénalise les solutions comportant des pertes.

L'optimisation du critère  $T_c$  doit s'effectuer en prenant en compte les écarts possibles entre la géométrie souhaitée et la géométrie réellement fabriquée. On parle alors d'optimisation robuste selon des incertitudes quantifiées sur les paramètres concernés. Les phases de lithographie ou de gravure sont les plus impactantes sur la géométrie du coupleur. Le motif général peut se retrouver plus gros ou plus petit, avec des effets sur les largeurs de guides et les gaps. Il n'y a malheureusement aucune donnée précise à partir de mesures sur les côtes obtenus après fabrication d'un composant cible. Des règles liées à l'expérience indiquent que les paramètres incertains peuvent varier de 10 % de la dimension critique du plus petit élément.

## F. Motivations

Les différentes descriptions des cas d'études ne permettent pas toutes de mettre immédiatement en évidence des problématiques bien précises. Avec la microsource infrarouge, les paramètres évoluent dans un domaine non hypercubique régi par des contraintes sur la géométrie. Les objectifs identifiés sont eux aussi soumis à une contrainte lors de leur optimisation liée à la température maximale du dispositif. La cellule photoacoustique dispose de deux modèles numériques. L'un très précis mais trop coûteux pour l'optimisation, l'autre très rapide à évaluer mais ses résultats ne correspondent pas suffisamment à l'expérimentation. Exploiter ces deux modèles en minimisant le coût numérique de l'opération est une solution à étudier dans la suite. Pour le coupleur, les données sur la quantification des incertitudes sont manquantes. L'optimisation robuste selon les écarts en fabrication doit être définie en fonction des données disponibles.

Il faut rappeler que ces cas d'application sont un extrait représentatif de la grande diversité des problèmes de conception rencontrés dans le domaine de l'optique et de la photonique. Ils sont suffisants pour définir différents enjeux globaux tel que l'optimisation sous contrainte, l'optimisation multi-objectif et l'optimisation robuste. Les différents verrous pour résoudre ces problèmes sont précisés au fil des études menées dans les chapitres. Cet ordre permet de respecter la chronologie entre l'identification de la difficulté et la solution apportée.

Une première caractéristique commune à tous les cas étudiés peut être malgré tout identifiée dès à présent : les modèles numériques s'avèrent trop coûteux pour être directement utilisables par un algorithme d'optimisation numérique. En utilisant des algorithmes génétiques, l'optimisation des cas cités ici pourrait prendre plusieurs jours de calculs. De plus, une seule optimisation est rarement suffisante, le travail consiste généralement à mener plusieurs essais avec différents objectifs ou plusieurs gammes de jeux de paramètres. Une longue phase de conception n'est pas toujours acceptable à l'échelle de projets industriels devant également inclure une phase de fabrication et de caractérisation. Dans le Tableau 2, une estimation du temps de calcul est fournie en comptant 1000 appels au simulateur par nombre de paramètre. La prise

en compte d'une contrainte multiplié par 2 le nombre d'appels, et la robuste par 10 (valeurs arbitrairement choisies pour donner une estimation).

Cas d'application	Microsource infrarouge	Cellule photoacoustique	Coupleur directionnel
Nombre de paramètres	9	3	4
Temps de simulation	9 sec	1h30 / 3 min	5 min
Nombre d'objectifs	3	2	1
Nombre de contraintes	1	0	0
Cadre de l'optimisation	Déterministe	Déterministe	Robuste
Estimation temps d'optimisation	45 h	4500 h	3333 h

**Tableau 2 : Synthèse des cas étudiés**

Une des solutions actuellement utilisables consiste à remplacer la simulation couteuse par un modèle de substitution (appelé également métamodèle). Ce dernier est une approximation de la fonction objectif dépendant des paramètres de conception et son coût d'évaluation est négligeable par rapport au temps de calcul du modèle numérique. Les notions liées aux modèles de substitution sont présentées dans le chapitre II, et illustrées par l'application des techniques numériques aux grandeurs d'intérêt des trois cas d'études considérés.

---

## II. Modèles de substitution

---

Les modèles numériques utilisés pour concevoir des composants photoniques sont souvent coûteux en termes de temps de calcul. Une première solution envisageable consiste à simplifier la formulation physique du problème. Des approximations sont en effet souvent possibles, réduisant ainsi la complexité des équations tout en représentant correctement le phénomène étudié. Lors de la résolution numérique des équations qui s'appuient souvent sur un découpage de l'espace et du temps, la finesse de la discrétisation peut également être diminuée. Ces modifications permettent d'obtenir diverses fidélités du comportement d'un même composant. Malheureusement, il reste des cas où les temps de calcul restent élevés après ces étapes.

Une autre solution consiste à remplacer la simulation par des modèles de substitution, souvent appelés métamodèles ou surface de réponse. Le métamodèle émule la relation statistique existant entre les données d'entrée de la simulation et les données fournies en sortie. Son apprentissage est donc fait à partir d'un jeu de données issues de l'évaluation du modèle numérique et consiste à estimer des paramètres d'un modèle mathématique. Le métamodèle peut ensuite prédire quasi-instantanément la sortie du simulateur pour un nouveau jeu de données d'entrées. Cette solution est celle qui est étudiée dans ce chapitre.

### A. Définitions

Le modèle numérique  $\mathcal{M}$  (équation [II.1]) est défini comme une observation de valeurs de sorties dépendant de paramètres d'entrées. Cette formulation s'appelle la modélisation boîte noire (ou modèle entrées-sorties). La partie physique de la simulation est complètement opaque pour l'utilisateur. Dans le cadre d'utilisation de logiciels commerciales ou propriétaires, les méthodes numériques mises en place pour résoudre les équations mises en jeu ne sont également pas visibles par l'utilisateur. Cela permet de définir des méthodes d'approximations qui pourront s'appliquer indépendamment du type de modélisation physique.

$$\mathcal{M}: \begin{cases} \mathbb{X} \subset \mathbb{R}^d \rightarrow \mathbb{R}^m \\ \mathbf{x} \mapsto y = \mathcal{M}(\mathbf{x}) \end{cases} \quad [\text{II.1}]$$

La dimension de l'espace d'entrée est  $d$ , définissant le nombre de paramètres étudiés. Lorsque la simulation renvoie plusieurs sorties ( $m > 1$ ), elles sont approximées une par une. Les données d'apprentissage sont constituées d'un ensemble de points d'entraînement  $\mathcal{X}_{tr} = \{\mathbf{x}_{tr}^{(1)}, \dots, \mathbf{x}_{tr}^{(n)}\}$  et de valeurs observées associées  $\mathcal{Y}_{tr} = \{y_{tr}^{(1)}, \dots, y_{tr}^{(n)}\}$ . Le couple d'ensembles  $\mathcal{D} = \{\mathcal{X}_{tr}, \mathcal{Y}_{tr}\}$  forme l'ensemble d'entraînement, appelé également plan d'expérience de telle sorte que :

$$y_{tr}^{(i)} = \mathcal{M}(\mathbf{x}_{tr}^{(i)}), \quad i = 1, \dots, n \quad [\text{II.2}]$$

On note également  $\mathbf{X}_{tr}$  la matrice contenant tous les points d'entraînements  $\mathcal{X}_{tr}$  et  $\mathbf{Y}_{tr}$  la matrice (une seule colonne lorsque  $m = 1$ ) contenant toutes les observations  $\mathcal{Y}_{tr}$ . Le modèle de substitution  $\widehat{\mathcal{M}}$  est construit à partir du plan d'expérience et permet de prédire de nouvelles valeurs  $\hat{y}$  de la sortie :

$$\hat{y} = \widehat{\mathcal{M}}(\mathbf{x}) \quad [\text{II.3}]$$

Dans l'apprentissage statistique, la classification et la régression sont deux disciplines distinctes. Leur différence se trouve au niveau de l'espace des valeurs prises par la sortie de la simulation. Pour les problèmes de régression, la sortie  $y$  est continue. Pour la classification, cette variable prend des valeurs discrètes correspondant aux étiquettes des différentes classes, par exemple  $y \in \{-1, 1\}$ . Cette discipline est

très répandue dans le domaine de la fiabilité mécanique, où deux labels sont utilisés : le système est défaillant ou fonctionnel. Ce type de problèmes ne sera pas traité ici, les codes de calculs des problèmes identifiés au chapitre I ont des sorties continues. De plus, le résultat de ces codes n'est pas considéré comme dépendant d'un aléa (maîtrisé ou non) ou d'un bruit numérique. Les fonctions à approximer sont donc lisses et les métamodèles utilisés interpolent les données d'entraînement (contrairement à la régression pure où la courbe ne passe pas forcément par les données d'apprentissage) :

$$\widehat{\mathcal{M}}(\mathbf{x}_{tr}^{(i)}) = y_{tr}^{(i)}, \quad i = 1, \dots, n \quad [\text{II.4}]$$

Lors de la construction du métamodèle, la normalisation des données d'entrée permet d'éviter certains effets numériques. En effet, les différences d'unités et les tailles d'intervalles de variations des paramètres peuvent perturber le processus d'apprentissage. On peut trouver dans nos cas d'étude un élément de  $\mathbf{x}$  correspondant à la longueur d'onde de la lumière de l'ordre du  $\mu\text{m}$  et un autre élément correspondant à une tension en V. Il est alors recommandé de remettre à la même échelle chaque paramètre, par exemple dans l'intervalle  $[0,1]$ . Centrer-réduire les données est également une possibilité (c'est le choix fait par les auteurs de la boîte à outil DACE [25]).

Dans la suite de ce chapitre, deux cadres d'utilisation des métamodèles sont présentés : tout d'abord lorsqu'un seul modèle  $\mathcal{M}$  du composant est disponible, puis lorsque plusieurs modélisations avec différentes fidélités sont possibles.

## B. Métamodèles pour la simulation à un niveau donné de fidélité

Lorsqu'un seul modèle numérique du comportement d'un composant est disponible, il existe plusieurs métamodèles permettant de le substituer. La revue de Wang [26] présente un bon aperçu des méthodes disponibles. Outre la classique régression polynomiale, on peut citer les Multivariate Adaptive Regression Splines (MARS) [27] qui correspondent à une décomposition de la fonction d'intérêt en des sommes pondérées de fonctions charnières. Cette méthode peut être vue comme une procédure géométrique de découpage du domaine par nœuds et de régression par splines dans chaque sous-domaine. Elle ne convient donc pas pour l'interpolation des données d'entraînement.

Parmi les autres méthodes évoquées, les Support Vector Machine (SVM) [28] sont très utilisées en classification. Elles se caractérisent par la recherche du meilleur séparateur linéaire des différentes classes du problème. Lorsque les données ne peuvent être séparées linéairement, il est possible d'autoriser une erreur de classification lors de la construction du modèle. Pour limiter cette erreur, une transformation de l'espace des paramètres d'entrée vers un espace de plus grande dimension est utilisée pour trouver un séparateur linéaire. Cette technique, dite Kernel trick, permet d'obtenir un bon séparateur pour des données non-linéaires. Les SVM ont été adaptées à la régression grâce à l'introduction d'une fonction de perte [29], définissant comment pénaliser les forts écarts entre les valeurs prédites et les valeurs observées. La courbe résultante est une minimisation de la marge entre la prédiction et les données d'entraînement. Selon le type de fonction de perte, l'interpolation des données est possible en pénalisant fortement les écarts de prédiction. Les SVM ne sont pas utilisées par la suite car la formulation est plus intéressante pour la régression que pour l'interpolation et le réglage de ses paramètres n'est pas simple.

Pour finir sur les métamodèles non explorés dans ces travaux, les polynômes de chaos [30] sont utilisés pour la propagation d'incertitude. Le domaine d'application de ce type de métamodèle ne correspond donc pas à celui de cette étude mais c'est un sujet très prisé actuellement. La réponse du modèle numérique est projetée sur une base de polynômes orthogonaux. L'approximation est construite en tenant compte de la loi de distribution de chaque paramètre d'entrée. L'intérêt de cette méthode réside dans l'obtention des moments statistiques et des indices de sensibilités liés aux données en entrée. Ces valeurs sont calculées analytiquement à partir des coefficients de la décomposition (chapitre 4.2 de [31]).

Dans la suite du document, nous commençons par présenter les plans d'expérience en section II.B.1. Le choix est ensuite fait de se limiter à deux méthodes de métamodélisation représentatives des méthodes

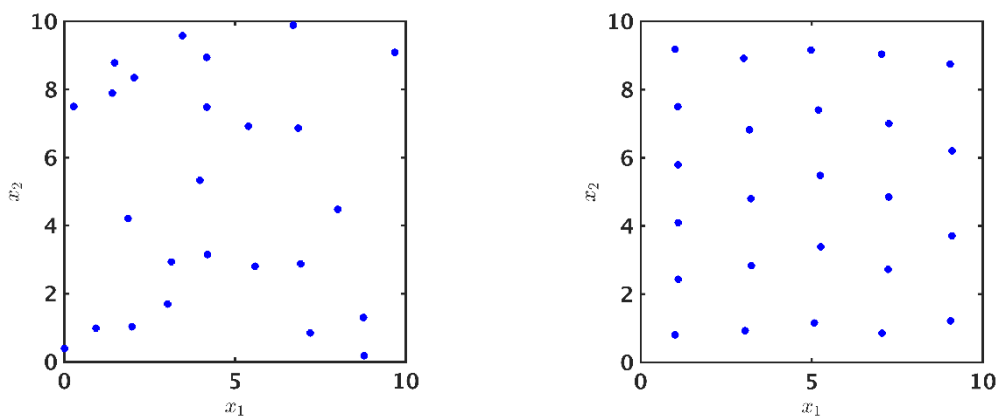
d'interpolation utilisées actuellement : le krigeage en section II.B.2 et les fonctions à base radiale en section II.B.3.

## 1. Plan d'expérience

Pour entraîner un modèle de substitution, il est nécessaire de disposer d'un plan d'expérience fournissant une information partielle sur le modèle d'intérêt. Plusieurs types de plans d'expériences existent et sont adaptés à différents contextes d'utilisation. Dans le cadre des métamodèles, le but est de remplir au mieux l'espace des paramètres d'entrée afin de récolter le plus d'information sur le modèle numérique à substituer. De plus, la simulation étant coûteuse en temps d'évaluation, le nombre de points dans le plan d'expérience doit être le plus réduit possible. Dans cette section, quelques plans d'expériences utilisés dans le cadre des modèles de substitution sont présentés. Il est nécessaire de rappeler que les stratégies d'échantillonnage forment à elles seules un véritable champ de prospection scientifique qui ne être pleinement retranscrits dans ce travail. Les éléments abordés ici sont principalement issus du livre de Niederreiter [32] et de la thèse de Franco [33]. En pratique, on utilise le terme DOE (Design Of Experiments) pour faire référence au plan d'expérience.

### a. Plan aléatoire uniforme et quasi-aléatoire

Les plans aléatoires sont très rapides à mettre en place. Il s'agit d'effectuer un tirage aléatoire pour chaque paramètre d'une loi uniforme multivariée dans le cas déterministe, c'est-à-dire lorsqu'aucune loi de probabilité régit le choix d'une valeur du paramètre. Dans le cas où les paramètres suivent des lois probabilistes particulières, il suffit simplement de générer les points en utilisant ces lois. Par contre, il est nécessaire de tirer un grand nombre de points pour bien couvrir l'espace d'entrée. Niederreiter [32] a introduit le concept de « discrédance », définie comme la déviation d'un ensemble de points  $\mathcal{X}$  par rapport à la distribution uniforme sur  $[0, 1]^d$ . Lorsque le nombre  $n$  de points dans  $\mathcal{X}$  tend vers l'infini, on considère que les points sont uniformément distribués si la discrédance tend vers zéro. Par exemple, la discrédance d'un plan aléatoire est égale à  $1/\sqrt{n}$ . Ces travaux ont motivé l'utilisation de suites à faible discrédance tel que les suites de Halton, Faure et Sobol'. Malheureusement, ces suites sont construites de sortes à maintenir une faible discrédance lors d'ajout de points, alors que l'intérêt ici est d'avoir une discrédance la plus faible possible pour un nombre donné de points. Une stratégie possible consiste à générer un grand nombre de points avec ces suites puis d'utiliser un algorithme de partitionnement de données (clustering) tel que les k-moyennes pour obtenir un DOE avec  $n$  points bien répartis [34]. Sur la Figure II-1, un plan aléatoire uniforme et un plan généré avec une suite de Sobol' agrégée par k-moyenne sont représentés.



**Figure II-1 : Plan aléatoire uniforme de 25 points (à gauche). Suite de Sobol' avec  $10^5$  points regroupés en 25 points par k-moyenne (à droite).**

### b. Plan factoriel complet

Ce plan est certainement le plus connu, quel que soit le type d'expérience menée. Il est composé des nœuds d'une grille formée par  $L$  niveaux des  $d$  paramètres composant l'espace d'entrée [35]. Dans la pratique, ce

type de plan est très utilisé car il est rapide à mettre en place (Figure II-2) car il ne demande aucune approche théorique. Il suffit de déterminer le nombre de valeurs à évaluer pour chaque paramètre, représentant le nombre de niveaux de la grille. Lorsque le nombre de niveaux et de variables est important, le nombre de points dans le plan d'expérience croît très rapidement ( $n = L^d$ ). De plus, sa discrétion, qui varie comme  $1/n$ , est plus élevée que celle d'un plan aléatoire uniforme. Ce type de plan ne convient donc pas pour l'entraînement d'un métamodèle dans ces travaux. Une alternative consiste à utiliser les plans factoriels fractionnaires, qui correspondent à un sous-ensemble d'un plan factoriel complet. Cependant ces plans sont utilisés avec une hypothèse de faible interaction entre les variables du modèle, ce qui ne correspond toujours pas au cadre de cette étude.

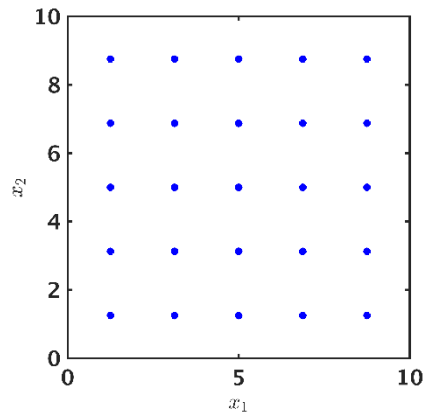


Figure II-2 : Plan factoriel complet de 25 points

c. Plan hypercube latin

Les plans latins hypercubes (LHS pour Latin Hypercube Sampling) sont très régulièrement choisis pour l'entraînement de métamodèles (voir [36] pour une vue d'ensemble récente). Ils peuvent être décrits facilement en dimension 2 : pour une grille de  $n \times n$  cases, il ne peut y avoir plus d'une évaluation pour chaque colonne et chaque ligne. Le nombre de LHS possible est alors de  $n!^{d-1}$ . Une des propriétés intéressantes se trouve dans la projection de ce type de DOE sur un sous-espace, qui est elle aussi un LHS. Toutefois, la couverture de l'espace n'est pas forcément bien assurée avec ce type de plan, qui permet simplement d'obtenir d'une bonne distribution des valeurs prises par chaque paramètre (exemple sur la Figure II-3).

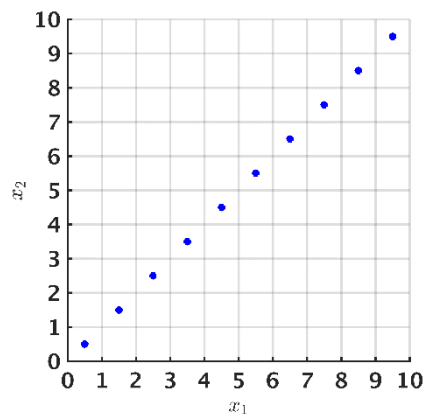


Figure II-3 : Plan latin hypercube avec une mauvaise couverture de l'espace d'entrée.

Il convient alors d'utiliser des plans latins hypercubes optimisés (OLHS), par exemple avec un critère tel que la maximisation de la distance minimale entre les points du plan [37]. Un exemple de ce type de plan est présenté sur la Figure II-4. Il correspond au type de DOE privilégié dans cette étude. D'autres choix de critère sont possibles, tel que la minimisation de la distance maximale entre les points du plan.

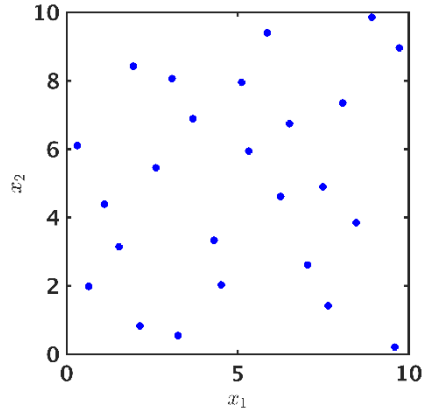


Figure II-4 : OLHS avec critère maximin de 25 points

## 2. Krigeage

La prédiction par krigeage est basée sur la théorie des processus gaussiens. Cette méthode a été nommée suite aux travaux en géostatistiques de Krige [38]. Actuellement, il s'agit de l'une des méthodes les plus utilisées pour l'interpolation de données de fonctions coûteuses. Cet intérêt s'explique par sa qualité de prédiction générale et de la possibilité d'estimer un intervalle de confiance sur les valeurs prédites. L'interpolation fournie par le krigeage est la moyenne des réalisations d'un processus gaussien conditionné par les points d'entraînement. Il convient de commencer par définir quelques notions associées.

### a. Quelques notions de modélisation probabiliste

Considérons un espace de probabilité  $(\Omega, \mathcal{A}, \mathcal{P})$ , où  $\Omega$  est un univers lié à l'expérience,  $\mathcal{A}$  une tribu de  $\Omega$  et  $\mathcal{P}$  une mesure de probabilité associée tel que  $\mathcal{P}(\Omega) = 1$ . Une variable aléatoire est une application définie sur un espace de probabilité à valeur dans un espace  $\mathbb{A}$  :

$$Z: \begin{cases} (\Omega, \mathcal{A}, \mathcal{P}) \rightarrow \mathbb{A} \\ \omega \mapsto z \end{cases} \quad [\text{II.5}]$$

où  $\omega$  est un évènement de  $\Omega$ . Une réalisation de cette variable est représentée par  $z = Z(\omega)$ . Selon la nature de l'espace  $\mathbb{Z}$ ,  $Z$  peut être continue ou discrète. Dans les deux cas, elle est complètement définie par sa fonction de répartition :

$$F_Z(z) = \mathcal{P}[Z \leq z] \quad [\text{II.6}]$$

Les variables aléatoires peuvent aussi admettre une densité de probabilité définie comme suit dans le cas continu :

$$f_Z(z) = \lim_{\substack{h \rightarrow \infty \\ h > 0}} \frac{\mathcal{P}[z \leq Z \leq z + h]}{h} = \frac{dF_Z(z)}{dz} \quad [\text{II.7}]$$

La moyenne de la variable aléatoire est définie en prenant  $\mathbb{E}$  comme notation de l'espérance mathématique :

$$\mu_Z = \mathbb{E}[Z] = \int_{\mathbb{Z}} z f_Z(z) dz \quad [\text{II.8}]$$

La variance (définie comme le moment centré d'ordre 2) est obtenue à partir de la moyenne :

$$\sigma_Z^2 = \mathbb{E}[(Z - \mu_Z)^2] \quad [\text{II.9}]$$



(a) Exemple : la loi normale.

La loi normale, notée  $\mathcal{N}(\mu_z, \sigma_z)$ , est utilisée dans la suite de ce document et constitue un bon exemple d'illustration des concepts évoqués précédemment. Une variable aléatoire suit une loi normale si elle admet une densité de probabilité de la forme :

$$\varphi(z) = \frac{1}{\sigma_z \sqrt{2\pi}} \exp \left[ -\frac{1}{2} \left( \frac{z - \mu_z}{\sigma_z} \right)^2 \right] \quad [\text{II.10}]$$

La fonction de répartition de la loi normale ne peut être évaluée que numériquement en utilisant la fonction d'erreur complémentaire :

$$\Phi(z) = -\frac{1}{2} \operatorname{erfc} \left( \frac{z - \mu_z}{\sigma_z \sqrt{2}} \right) \quad [\text{II.11}]$$

Sur la Figure II-5, ces deux fonctions sont tracées pour  $\mu_z = 2$  et  $\sigma_z = 0,5$ . La densité de probabilité indique les valeurs que va prendre le plus vraisemblablement la variable aléatoire (ici, 2 est la valeur la plus probable). La fonction de répartition donne une information probabiliste avant la réalisation de la variable (il y a 50 % de chances dans notre exemple d'avoir une réalisation inférieure à 2).

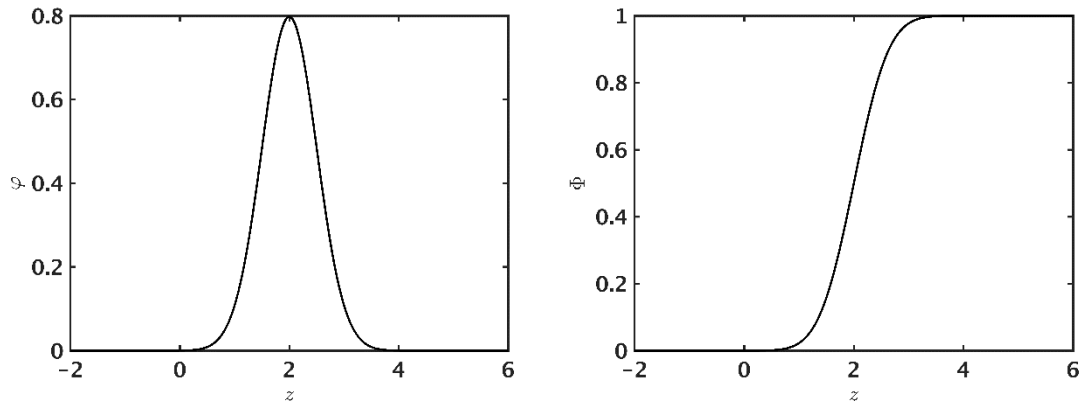


Figure II-5 : Densité de probabilité et fonction de répartition de la loi normale pour  $\mu_z = 2$  et  $\sigma_z = 0,5$ .

### b. Processus gaussien

Un processus stochastique (ou aléatoire) est une application définie sur un espace de probabilité indexé sur un espace  $\mathbb{X}$  à valeur dans un espace  $\mathbb{A}$  :

$$Z : \begin{cases} (\Omega, \mathcal{A}, \mathcal{P}) \times \mathbb{X} \rightarrow \mathbb{A} \\ (\mathbf{x}, \omega) \mapsto z \end{cases} \quad [\text{II.12}]$$

Une réalisation (ou trajectoire) de ce processus est définie par  $z \equiv Z(\mathbf{x}, \omega)$ . Pour un certain évènement  $\omega_0 \in \Omega$ ,  $Z(\mathbf{x}, \omega_0)$  est une fonction. Tandis que pour un élément  $\mathbf{x}_0 \in \mathbb{X}$ ,  $Z(\mathbf{x}_0, \omega)$  est une variable aléatoire. Dans la suite, l'évènement  $\omega$  est implicite et omis pour faciliter la lecture. Le processus aléatoire est dit gaussien si n'importe quel nombre fini de ses variables aléatoires admet une distribution normale multivariée [39]. On précise qu'un vecteur aléatoire  $\mathbf{z}$  de taille  $n$  suit une distribution normale multivariée s'il s'exprime comme la combinaison linéaire de  $r$  variables aléatoires indépendantes normales centrées réduites :

$$\begin{aligned} \mathbf{z} &= \mathbf{L} \times \boldsymbol{\Theta} + \mathbf{m}, \\ \boldsymbol{\Theta} &= (\Theta_i \sim \mathcal{N}(0,1), i = 1, \dots, r), \end{aligned} \quad [\text{II.13}]$$

où  $\mathbf{L}$  est une matrice réelle de  $n \times r$  et  $\mathbf{m}$  un vecteur à  $n$  composant réels. Tout processus gaussien est défini par sa moyenne  $\mu_{GP}$  et sa fonction d'autocovariance  $C_{GP}$  :

$$\mu_{GP}(\mathbf{x}) \equiv \mathbb{E}[Z(\mathbf{x})]. \quad [\text{II.14}]$$

$$C_{GP}(\mathbf{x}, \mathbf{x}') \equiv \mathbb{E}[(Z(\mathbf{x}) - \mu_{GP}(\mathbf{x}))(Z(\mathbf{x}') - \mu_{GP}(\mathbf{x}'))], \quad (\mathbf{x}, \mathbf{x}') \in \mathbb{X} \times \mathbb{X}.$$

La fonction d'autocovariance doit être symétrique définie positive, et donc satisfaire les deux conditions suivantes :

$$C_{GP}(\mathbf{x}, \mathbf{x}') = C_{GP}(\mathbf{x}', \mathbf{x}), \quad \forall (\mathbf{x}, \mathbf{x}') \in \mathbb{X} \times \mathbb{X}.$$

$$\sum_{i=1}^n \sum_{j=1}^n w_i w_j C_{GP}(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) \geq 0, \quad [\text{II.15}]$$

pour tout sous ensemble  $\{\mathbf{x}^{(i)}, i = 1, \dots, n\}$  de  $\mathbb{X}$  et tout vecteur  $\mathbf{w} = (w_i, i = 1, \dots, n)^T \in \mathbb{R}^n$ . La suite de l'étude se limite à l'utilisation de processus gaussiens *stationnaires* (invariants par translation). Ceci signifie que leur fonction d'autocovariance s'écrit :

$$C_{GP}(\mathbf{x}, \mathbf{x}') = \sigma_{GP}^2 R(\mathbf{x} - \mathbf{x}'), \quad (\mathbf{x}, \mathbf{x}') \in \mathbb{X} \times \mathbb{X} \quad [\text{II.16}]$$

où  $\sigma_{GP}^2$  est la variance du processus (constante ici) et  $R$  la fonction d'autocorrélation qui dépend de seulement de la différence  $\mathbf{x} - \mathbf{x}'$ . Plusieurs fonctions d'autocorrélation utilisées habituellement dans la prédiction par processus gaussiens ([39] - section 4.2) sont présentées dans la section suivante.

### c. Fonctions d'autocorrélation dans le cadre de l'étude

Parmi les fonctions d'autocorrélation, deux classes peuvent être distinguées : isotropes et anisotropes. Tout d'abord, notons les fonctions d'autocorrélation dépendent d'une certaine norme permettant le calcul de la distance entre les points :

$$R(\mathbf{x}, \mathbf{x}') = R(\|\mathbf{x} - \mathbf{x}'\|_A) \quad [\text{II.17}]$$

où  $\|\mathbf{x} - \mathbf{x}'\|_A = \sqrt{(\mathbf{x} - \mathbf{x}')^T \mathbf{A} (\mathbf{x} - \mathbf{x}')}$  avec  $\mathbf{A}$  une matrice semi-définie positive. Les fonctions d'autocorrélation isotrope correspondent au cas où la distance n'est pas pondérée selon chaque dimension, soit  $\mathbf{A} = \mathbf{I}$  avec  $\mathbf{I}$  la matrice identité. Dans notre cadre d'application, l'hypothèse d'isotropie ne se justifie pas car les différentes dimensions de l'espace d'entrée représentent différentes variables physiques. Il ne s'agit pas de problèmes d'homogénéité, qui sont réglés par la mise à l'échelle des paramètres d'entrée du modèle. Ici, il est question de l'influence de chaque variable physique sur la sortie d'intérêt du modèle  $\mathcal{M}$ . Si un des éléments de  $\mathbf{x}$  a une faible influence sur la sortie  $y$ , il est intéressant de considérer une fonction d'autocorrélation de longue portée pour généraliser la tendance du modèle selon cette dimension. Au contraire, une fonction d'autocorrélation de courte portée peut être utilisée dans une direction où  $y$  va beaucoup varier. Cette prise en compte de l'importance de chaque dimension de l'espace des paramètres sur la sortie d'intérêt est possible en utilisant des fonctions d'autocorrélation dites anisotropes géométriques ([40] - section 3.2.1). Dans ce manuscrit, seul le cas  $\mathbf{A} = \text{diag}(\boldsymbol{\theta}^2)$  est considéré où  $\boldsymbol{\theta}$  est un vecteur positif de  $d$  paramètres d'échelles représentant l'influence spatiale de chaque dimension de l'espace  $\mathbb{X} \subset \mathbb{R}^d$ . La notation  $\|\cdot\|_{\boldsymbol{\theta}}$  est alors employée pour définir le cas où  $\mathbf{A} = \text{diag}(\boldsymbol{\theta}^2)$ . Pour finir, on parle d'autocorrélation anisotrope lorsque la matrice  $\mathbf{A}$  ne correspond pas à la matrice identité. L'anisotropie géométrique est donc un cas particulier des fonctions d'autocorrélation anisotrope.

Les processus aléatoires illustrés dans la suite de cette section sont tous stationnaires, de moyenne  $\mu_{GP} = 2$  et de variance  $\sigma_{GP}^2 = 0,5$ . Les fonctions d'autocorrélation sont les seuls éléments différant entre chaque représentation. Les trajectoires de processus gaussiens sont obtenues par la méthode par décomposition de Cholesky ([41] - section A.5.1). En définissant la matrice de covariance  $\mathbf{C}$  telle que :

$$\mathbf{C}_{ij} = C_{GP}(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}), \quad i, j = 1, \dots, n. \quad [\text{II.18}]$$

La décomposition de Cholesky de cette matrice permet d'obtenir une matrice triangulaire inférieure  $\mathbf{L}_C$  telle

que :

$$\mathbf{C} = \mathbf{L}_C \mathbf{L}_C^T \quad [\text{II.19}]$$

Les trajectoires sont simulées à partir de la définition de l'équation [II.13], avec  $r = n$  :

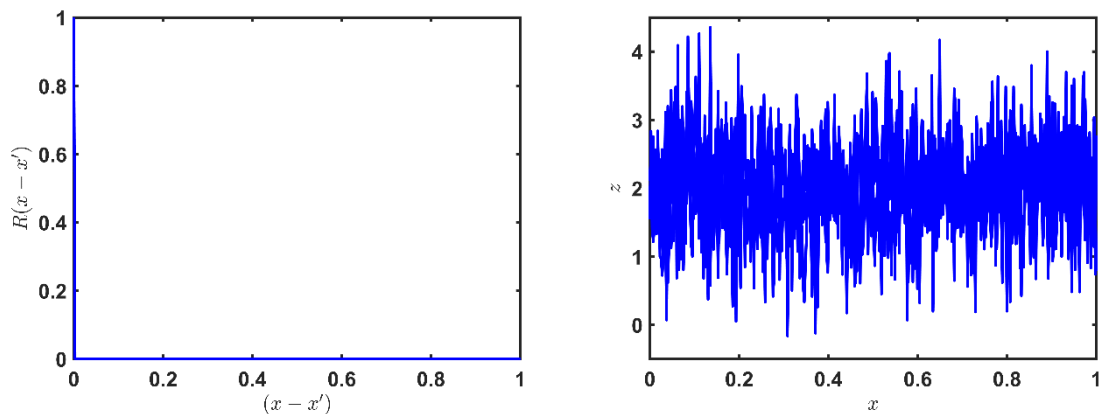
$$\mathbf{Z} = \mathbf{L}_C \times \boldsymbol{\Theta} + \mu_{GP}. \quad [\text{II.20}]$$

### 1. Fonction d'autocorrélation de type bruit blanc

Cette fonction est définie à partir de la fonction de Dirac  $\delta_D(x)$ , qui vaut 1 lorsque  $x$  est nul et 0 sinon :

$$R(\mathbf{x}, \mathbf{x}') = \delta_D(\mathbf{x} - \mathbf{x}'). \quad [\text{II.21}]$$

Elle représente l'absence de corrélation entre les différentes observations du processus aléatoire. Ce type de processus gaussien est connu sous le nom de bruit blanc et ses réalisations sont discontinues (voir Figure II-6). Il est souvent utilisé pour modéliser un bruit dans les données. Cette fonction est également appelée autocorrélation pépitique de par son utilisation en géostatistique pour représenter une pépite d'or dans un bloc par rapport à d'autres blocs n'en possédant pas ([40] - section 3.2.1).



**Figure II-6 : Fonction d'autocorrélation bruit blanc (à gauche) et 3 trajectoires de processus gaussiens simulées avec cette fonction (à droite).**

### 2. Fonction d'autocorrélation exponentielle

Cette fonction est une fonction d'autocorrélation de type anisotrope géométrique, définie par :

$$R(\mathbf{x}, \mathbf{x}', \boldsymbol{\theta}) = \exp(-\|\mathbf{x} - \mathbf{x}'\|_{\boldsymbol{\theta}}), \quad [\text{II.22}]$$

où  $\boldsymbol{\theta}$  est le paramètre d'échelle représentant l'influence spatiale de chaque dimension de l'espace  $\mathbb{X} \subset \mathbb{R}^d$ . Ce paramètre est nommé *hyperparamètre* (ou longueur de corrélation, ou portée). Les propriétés algébriques de l'exponentielle permettent également de se ramener à une expression de forme anisotrope (similaire à l'équation [II.19]). Quelques trajectoires en dimension 1 sont simulées sur la Figure II-7 avec différentes valeurs de l'hyperparamètre. Ces trajectoires sont  $\mathcal{C}^0$  (continues et non différentiables). Lorsque la valeur de l'hyperparamètre est petite, l'influence du paramètre correspondant sur la trajectoire du processus est faible.

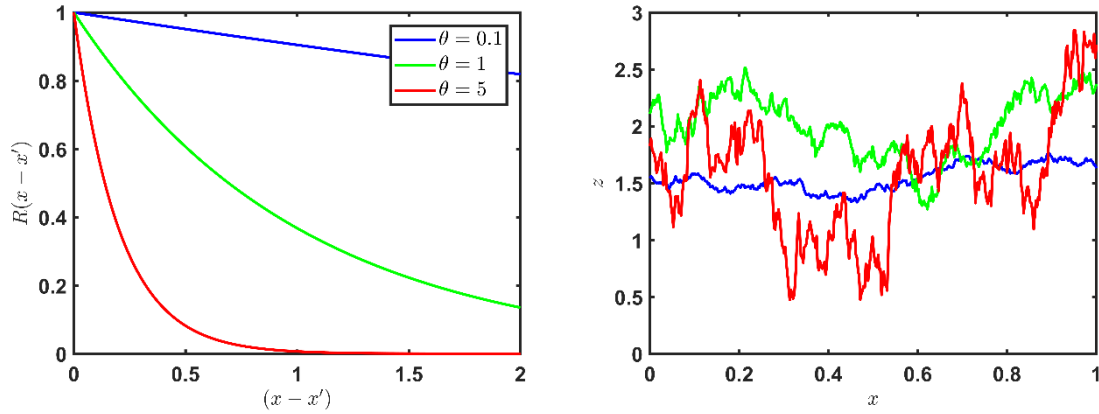


Figure II-7 : Fonction d'autocorrélation exponentielle (à gauche) et 3 trajectoires de processus gaussiens simulées avec cette fonction (à droite).

### 3. Fonction d'autocorrélation gaussienne

Cette fonction est une fonction d'autocorrélation de type anisotrope géométrique, définie par :

$$R(\mathbf{x}, \mathbf{x}', \boldsymbol{\theta}) = \exp(-\|\mathbf{x} - \mathbf{x}'\|_{\boldsymbol{\theta}}^2). \quad [\text{II.23}]$$

Les trajectoires générées avec cette fonction sont  $C^\infty$ . Les réalisations simulées sur la Figure II-8 sont plus « lisses » que celles qui sont obtenues avec une fonction d'autocorrélation exponentielle. La régularité de l'interpolation s'explique par l'exposant de la norme des écarts entre les observations. Le comportement des trajectoires est plus lisse lorsque cet exposant est proche de 2 (cas gaussien) que lorsqu'il est proche de 1 (cas exponentielle).

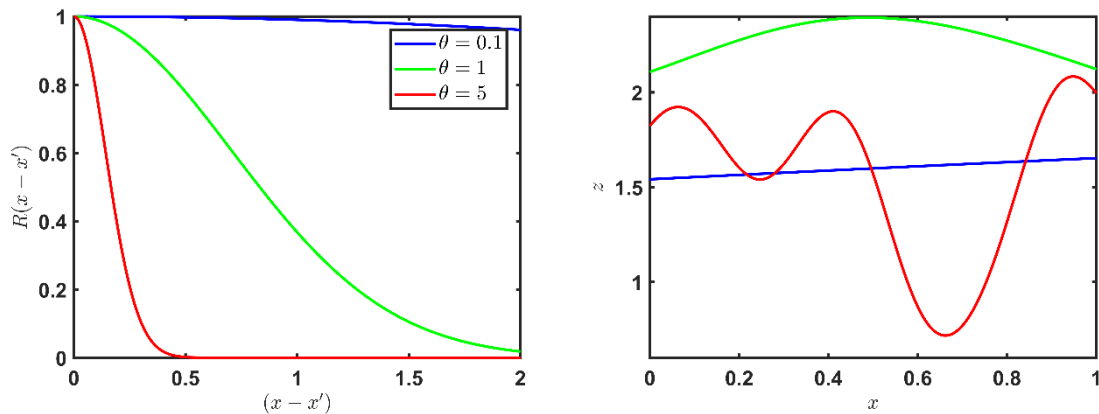


Figure II-8 : Fonction d'autocorrélation gaussienne (à gauche) et 3 trajectoires de processus gaussiens simulées avec cette fonction (à droite).

### 4. Fonction d'autocorrélation Matérn

Cette fonction est une fonction d'autocorrélation de type anisotrope géométrique, définie par :

$$R(\mathbf{x}, \mathbf{x}', \boldsymbol{\theta}, \nu) = \frac{2^{1-\nu}}{\Gamma(\nu)} (\sqrt{2\nu}\|\mathbf{x} - \mathbf{x}'\|_{\boldsymbol{\theta}})^\nu K_\nu(\sqrt{2\nu}\|\mathbf{x} - \mathbf{x}'\|_{\boldsymbol{\theta}}), \quad [\text{II.24}]$$

où  $\Gamma$  représente la fonction Gamma et  $K_\nu$  est la fonction de Bessel modifiée de deuxième espèce. L'expression [II.24] se simplifie lorsque  $\nu$  s'écrit comme la somme d'un entier et de  $1/2$ . Elle devient alors le produit d'un polynôme et d'une exponentielle. Lorsque  $\nu = 1/2$ , la fonction d'autocorrélation Matérn correspond à l'autocorrélation exponentielle. Quand  $\nu$  tend vers l'infini, elle correspond à l'autocorrélation gaussienne (plus le paramètre  $\nu$  est grand, plus l'interpolation devient régulière). Dans ce document, nous limitons l'étude aux cas  $\nu = 3/2$  et  $\nu = 5/2$  :

$$R_{3/2}(\mathbf{x}, \mathbf{x}', \boldsymbol{\theta}) = (1 + \sqrt{3}\|\mathbf{x} - \mathbf{x}'\|_{\boldsymbol{\theta}}) \times \exp(-\sqrt{3}\|\mathbf{x} - \mathbf{x}'\|_{\boldsymbol{\theta}}), \quad [\text{II.25}]$$

$$R_{5/2}(\mathbf{x}, \mathbf{x}', \boldsymbol{\theta}) = \left(1 + \sqrt{5}\|\mathbf{x} - \mathbf{x}'\|_{\boldsymbol{\theta}} + \frac{5}{3}\|\mathbf{x} - \mathbf{x}'\|_{\boldsymbol{\theta}}^2\right) \times \exp(-\sqrt{5}\|\mathbf{x} - \mathbf{x}'\|_{\boldsymbol{\theta}}).$$

Les trajectoires sont visibles sur la Figure II-9 pour le cas Matérn 3/2 et sur la Figure II-10 pour le cas Matérn 5/2. Une légère différence peut être observée sur la forme de la fonction de corrélation. Les trajectoires sont légèrement plus régulières pour le cas Matérn 5/2.

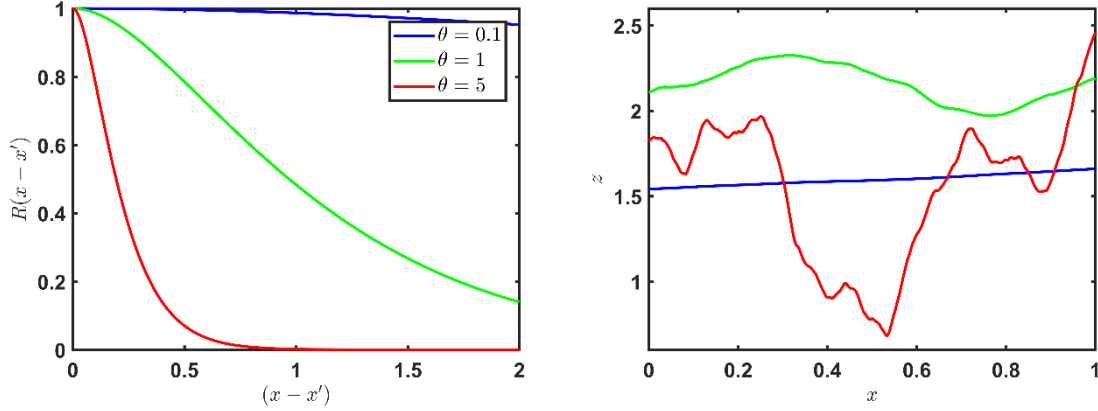


Figure II-9 : Fonction d'autocorrélation Matérn 3/2 (à gauche) et 3 trajectoires de processus gaussiens simulées avec cette fonction (à droite).

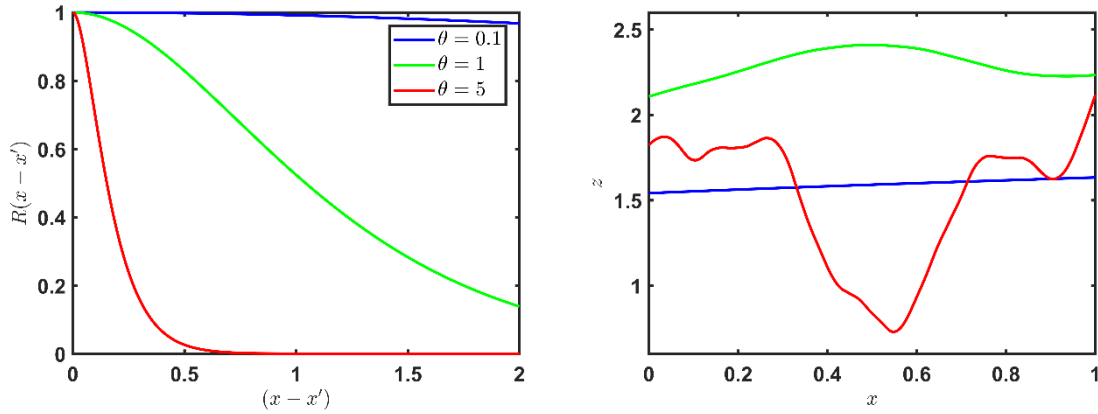


Figure II-10 : Fonction d'autocorrélation Matérn 5/2 (à gauche) et 3 trajectoires de processus gaussiens simulées avec cette fonction (à droite).

#### d. Prédiction par processus gaussien

Le modèle  $\mathcal{M}$ , qui exprime la relation entre les données d'entrées  $\mathcal{X}_{tr}$  et de sortie  $\mathcal{Y}_{tr}$ , est ici défini comme une réalisation d'un processus gaussien. On suppose que la fonction d'autocovariance et ses hyperparamètres sont connus. L'estimation des hyperparamètres est abordée plus tard dans la section II.B.2.e. Les données de sorties peuvent être considérées comme une réalisation d'un vecteur aléatoire  $\mathbf{Z}_{tr}$  tel que :

$$\mathbf{Z}_{tr} \sim \mathcal{N}(\boldsymbol{\mu}_{tr}, \mathbf{C}_{tr}), \quad [\text{II.26}]$$

$$\boldsymbol{\mu}_{tr} = [m(\mathbf{x}_{tr}^{(1)}), \dots, m(\mathbf{x}_{tr}^{(n)})]^T,$$

avec  $m$  une fonction qui sera précisée dans la suite et  $\mathbf{C}_{tr} = \sigma_{GP} \mathbf{R}_{tr}$  la matrice de covariance, calculée sur les données  $\mathcal{X}_{tr}$ . La réalisation d'un jeu de données de test  $\mathcal{X}_{te}$  de taille  $n_{te}$  est donnée par un autre vecteur aléatoire  $\mathbf{Z}_{te}$ . Ce dernier correspond aux nouvelles valeurs à prédire. Il faut donc connaître la distribution

conditionnelle du vecteur  $\mathbf{Z}_{te}$  connaissant les données d'entraînement  $\mathcal{D}$ . Cette distribution s'exprime comme suit ([39] - section A.2) :

$$\mathbf{Z}_{te} | \mathbf{Z}_{tr} \sim \mathcal{N}(\boldsymbol{\mu}_{te} + (\mathbf{Y}_{tr} - \boldsymbol{\mu}_{tr}) \mathbf{R}_{tr}^{-1} \boldsymbol{\Sigma}^T, \mathbf{C}_{te} - \boldsymbol{\Sigma} \mathbf{C}_{tr}^{-1} \boldsymbol{\Sigma}^T), \quad [\text{II.27}]$$

où  $\boldsymbol{\mu}_{te}$  et  $\mathbf{C}_{te}$  se définissent comme  $\boldsymbol{\mu}_{tr}$  et  $\mathbf{C}_{tr}$ , mais sur les données  $\mathcal{X}_{te}$ . La matrice  $\boldsymbol{\Sigma}$  est la matrice de corrélation entre les données d'entraînement  $\mathcal{X}_{tr}$  et les points de test  $\mathcal{X}_{te}$  :

$$\Sigma_{ij} = R(\mathbf{x}_{te}^{(i)}, \mathbf{x}_{tr}^{(j)}), i = 1, \dots, n_{te}, j = 1, \dots, n. \quad [\text{II.28}]$$

La prédiction du krigeage est alors la moyenne de la distribution conditionnelle, qui s'exprime sous la forme :

$$\hat{y}(\mathbf{x}) = \mathbb{E}[\mathbf{Z}_{te} | \mathbf{Z}_{tr}] = m(\mathbf{x}) + (\mathbf{Y}_{tr} - \boldsymbol{\mu}_{tr}) \mathbf{R}_{tr}^{-1} \boldsymbol{\Sigma}_x^T, \quad [\text{II.29}]$$

avec  $\boldsymbol{\Sigma}_x$  la matrice de corrélation entre les données  $\mathcal{X}_{tr}$  et les points à prédire  $\mathbf{x}$ . La fonction  $m$  correspond à une tendance globale de la sortie du modèle  $\mathcal{M}$ . On parle de krigeage *universel* lorsque cette tendance correspond à une régression polynomiale sur les données observées. Dans le cas où la régression se limite à un polynôme d'ordre 0 (une constante), le terme krigeage *ordinaire* est employé. Ce dernier est utilisé dans ce travail car aucune information permettant de choisir le degré d'une tendance polynomiale n'est disponible a priori sur les modèles à approximer. L'expression [II.29] se simplifie en :

$$\hat{y}(\mathbf{x}) = \mu_0 + \boldsymbol{\Sigma}_x \mathbf{R}_{tr}^{-1} (\mathbf{Y}_{tr} - \mu_0), \quad [\text{II.30}]$$

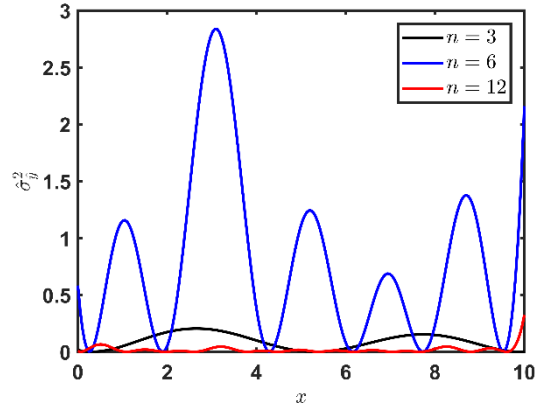
avec  $\mu_0$  la tendance moyenne à estimer. La distribution conditionnelle de l'équation [II.27] donne également la variance de prédiction du krigeage, s'exprimant par :

$$\hat{\sigma}_y^2(\mathbf{x}) = \sigma_{GP}^2 \left( \mathbf{1} - \boldsymbol{\Sigma}_x \mathbf{R}_{tr}^{-1} \boldsymbol{\Sigma}_x^T + \frac{(\mathbf{1} - \mathbf{1}^T \mathbf{R}_{tr}^{-1} \boldsymbol{\Sigma}_x^T)^2}{\mathbf{1}^T \mathbf{R}_{tr}^{-1} \mathbf{1}} \right), \quad [\text{II.31}]$$

avec  $\mathbf{1}$  le vecteur de taille  $n$  ne contenant que des 1. Des précisions supplémentaires sur le développement des expressions de la moyenne et de la variance du krigeage sont disponibles dans la thèse de Vincent Dubourg ([41] - section 1.4.3).

### 1. Quelques propriétés notables de la prédiction par krigeage

Une première propriété démontrée par Vazquez ([42] - section 5.3) concerne la convergence asymptotique de la prédiction par krigeage. Lorsque le nombre de points d'entraînement devient dense dans  $\mathbb{X}$ , la variance de prédiction du krigeage tend vers zéro sur tout le domaine si la régularité de la covariance est bien estimée. Les utilisations de covariances de type Matérn sont privilégiées dans son étude. Il est toutefois important de préciser qu'une matrice de covariance construite avec des points d'entraînement très proches les uns des autres a de fortes chances d'induire des erreurs numériques dues à son mauvais conditionnement. Un exemple de convergence asymptotique est visible sur la Figure II-11, où la variance de prédiction tend vers de faibles valeurs lorsqu'une douzaine de points d'entraînement sont utilisés pour l'entraînement.

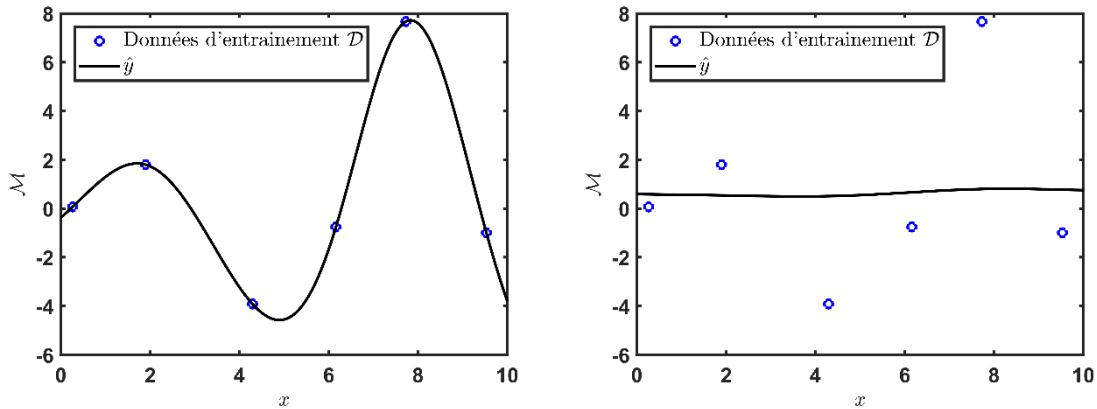


**Figure II-11 : Convergence asymptotique de la prédiction pour une fonction  $\mathcal{M}(x) = x \sin(x)$ . La fonction d'autocovariance est la fonction Matérn 5/2. Les hyperparamètres sont optimisés pour chaque cas.**

Une deuxième propriété concerne le caractère interpolant du krigeage :

$$\hat{y}(\mathbf{X}_{tr}) = \mu_0 + \mathbf{R}_{tr} \mathbf{R}_{tr}^{-1} (\mathbf{Y}_{tr} - \mu_0) = \mu_0 + \mathbf{I} (\mathbf{Y}_{tr} - \mu_0) = \mathbf{Y}_{tr}, \quad [\text{II.32}]$$

avec  $\mathbf{I}$  la matrice identité. Ce caractère peut être mis à défaut par le fameux *effet pépité* lié à la fonction d'autocovariance de bruit blanc. En ajoutant une valeur  $\epsilon_p$  à la diagonale de la matrice de covariance, le conditionnement de la matrice est amélioré. Si cette valeur est suffisamment faible devant la variance du processus gaussien, la condition d'interpolation peut être considérée comme conservée. Le paramètre  $\epsilon_p$  peut être appelé variance pépitique. Dans le cadres de données d'entraînements bruitées, obtenues à partir d'un code non-déterministe ou incluant un bruit de mesure par exemple, une forte valeur de la variance pépitique peut être choisie afin d'obtenir une régression des données. Ce paramètre peut même être optimisé avec les méthodes de la section II.B.2.e. Une interpolation et une régression, effectuées à partir du même jeu de données, sont visibles sur la Figure II-12. Dans le second cas, la variance pépitique est volontairement fixée à une forte valeur.



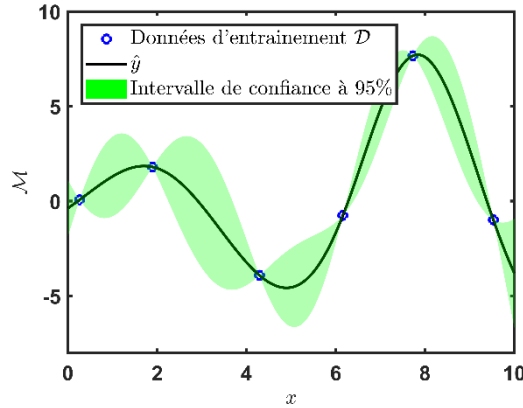
**Figure II-12 : Interpolation des données d'entraînement sur une fonction  $\mathcal{M}(x) = x \sin(x)$  (gauche). Régression des mêmes données avec  $\epsilon_p = 15$  (droite).**

Pour terminer, la troisième propriété concerne l'hypothèse de distribution normale des données observées qui est transmise à la prédiction. Cette propriété indiquée dès l'expression de la distribution conditionnelle de l'équation [II.27] permet l'estimation d'un intervalle de confiance sur la prédiction à partir de sa moyenne et de son écart-type :

$$\mathbf{Z}_{te} | \mathbf{Z}_{tr} \in \left[ \hat{y} - \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right) \hat{\sigma}_{\hat{y}}; \hat{y} + \Phi^{-1} \left( 1 - \frac{\alpha}{2} \right) \hat{\sigma}_{\hat{y}} \right]. \quad [\text{II.33}]$$

Pour rappel,  $\Phi$  désigne la fonction de répartition de la loi normale centrée réduite et donc  $\Phi^{-1}$  sa fonction

inverse. En prenant la valeur  $\alpha = 5\%$  et  $\Phi^{-1}(1 - 0.05/2) = 1.96$ , on obtient l'intervalle de confiance à 95% de la prédiction (exemple visible sur la Figure II-13).



**Figure II-13 : Intervalle de confiance de la prédiction sur la fonction  $\mathcal{M}(x) = x \sin(x)$  avec 6 points d'entraînement.**

#### e. Estimation des paramètres du processus gaussien

Dans la section précédente, la fonction d'autocovariance ainsi que ses hyperparamètres étaient supposés connus. Or ce n'est pas le cas dans la pratique. Il faut, dans un premier temps, choisir la fonction d'autocorrélation à utiliser puis estimer ses paramètres à partir des données d'entraînement  $\mathcal{D}$ . L'estimation par maximum de vraisemblance est la méthode la plus employée pour obtenir les paramètres du krigeage. La vraisemblance des observations  $\mathbf{Y}_{tr}$  selon la loi normale multivariée est :

$$L(\mu_0, \sigma_{GP}^2, \boldsymbol{\theta} | \mathbf{Y}_{tr}) = \frac{1}{(2\pi\sigma_{GP}^2)^{n/2} \sqrt{\det(\mathbf{R}_{tr})}} \exp \left[ -\frac{(\mathbf{Y}_{tr} - \mu_0)^T \mathbf{R}_{tr}^{-1} (\mathbf{Y}_{tr} - \mu_0)}{2\sigma_{GP}^2} \right] \quad [\text{II.34}]$$

La maximisation de  $L$  est équivalente à la minimisation de l'opposé de son logarithme. Avec les conditions d'optimalité de premier ordre, on obtient les estimations suivantes :

$$\hat{\mu}_0 = \frac{\mathbf{1}^T \mathbf{R}_{tr}^{-1} \mathbf{Y}_{tr}}{\mathbf{1}^T \mathbf{R}_{tr}^{-1} \mathbf{1}}, \hat{\sigma}_{GP}^2 = \frac{(\mathbf{Y}_{tr} - \hat{\mu}_0)^T \mathbf{R}_{tr}^{-1} (\mathbf{Y}_{tr} - \hat{\mu}_0)}{n} \quad [\text{II.35}]$$

En substituant ces deux solutions dans l'équation [II.34], la fonction de vraisemblance réduite apparait et permet d'estimer les hyperparamètres :

$$\boldsymbol{\theta} = \arg \min \frac{n}{2} \ln(\hat{\sigma}_{GP}^2) + \frac{1}{2} \ln(\det(\mathbf{R}_{tr})) \quad [\text{II.36}]$$

Cette fonction est souvent multimodale et son optimisation globale peut être complexe lorsque le nombre d'hyperparamètres devient important.

Une autre méthode d'estimation possible est la validation croisée. Cette dernière est un outil très couramment utilisé pour la sélection des paramètres d'un modèle. Son principe consiste en premier lieu à séparer les données d'entraînement  $\mathcal{D}$  en  $K$  sous-ensembles disjoints  $\mathcal{D}_k$  tel que :

$$\forall (i, j) \in \{1, 2, \dots, K\}^2, i \neq j, \mathcal{D}_i \cap \mathcal{D}_j = \emptyset \text{ et } \bigcup_{k=1}^K \mathcal{D}_k = \mathcal{D} \quad [\text{II.37}]$$

Dans un second temps, le métamodèle est construit pour chaque sous-ensemble  $k$  en utilisant toutes les données sauf celles de  $\mathcal{D}_k$ . Les valeurs prédites par le métamodèle aux points de  $\mathcal{D}_k$  sont comparées avec les vraies valeurs. Le métamodèle est alors construit  $K$  fois pour obtenir une erreur de prédiction pour chaque sous-ensemble. Cette erreur peut être calculée en utilisant des mesures décrites dans la section II.B.4. Au lieu d'optimiser la vraisemblance réduite, il s'agit de diminuer l'erreur de prédiction donnée par



validation croisée pour sélectionner les hyperparamètres. Dans le cas où  $K = n$ , cette méthode prend le nom de « leave-one-out » (LOO). Une formule développée par Dubrule [43] permet d'obtenir la prédiction des points de chaque sous-ensemble dans le cadre du LOO en construisant un seul krigeage sur l'ensemble des données. En notant  $\mathbf{S}$  la matrice :

$$\mathbf{S} = \begin{bmatrix} \sigma_{GP}^2 \mathbf{R}_{tr} & \mathbf{1} \\ \mathbf{1}^T & 0 \end{bmatrix}. \quad [\text{II.38}]$$

La prédiction  $\hat{y}_{-k}$  du sous-ensemble  $k$  s'il est retiré des données d'entraînement ainsi que la variance de prédiction  $\hat{\sigma}_{\hat{y}_{-k}}^2$  sont données par :

$$\hat{y}_{-k}(\mathbf{x}^{(k)}) = \sum_{\substack{i=1 \\ i \neq k}}^n \frac{[\mathbf{S}^{-1}]_{ki}}{[\mathbf{S}^{-1}]_{kk}} y_{tr}^{(i)}, \quad \hat{\sigma}_{\hat{y}_{-k}}^2 = \frac{1}{[\mathbf{S}^{-1}]_{kk}} \quad [\text{II.39}]$$

A partir des valeurs prédites aux points retirés, il est possible de calculer le log de la pseudovraisemblance :

$$\log p_{LOO}(k, \boldsymbol{\theta}) = -\frac{1}{2} \log \hat{\sigma}_{\hat{y}_{-k}}^2 - \frac{\hat{y}_{-k}^2}{2\hat{\sigma}_{\hat{y}_{-k}}^2} - \frac{1}{2} \log 2\pi \quad [\text{II.40}]$$

Cette mesure d'erreur permet de tenir compte de la variance de prédiction contrairement à d'autres, comme par exemple l'erreur au sens des moindres carrés ([39] – section 5.4.2). L'erreur au sens du leave-one-out est finalement la somme des erreurs de chaque sous-ensemble :

$$L_{LOO}(\boldsymbol{\theta}) = \sum_{i=1}^n \log p_{LOO}(k, \boldsymbol{\theta}). \quad [\text{II.41}]$$

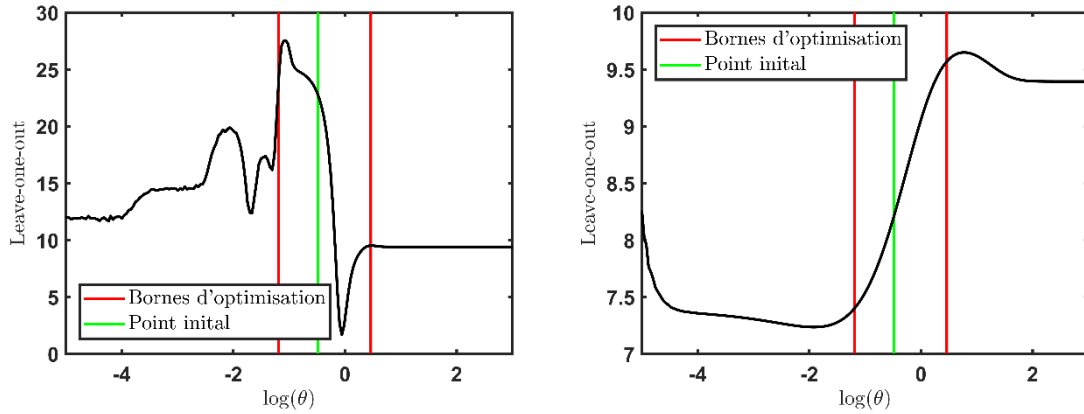
Des exemples de différence entre les deux méthodes d'estimation des hyperparamètres sont présentés dans la section suivante.

#### f. Utilisation dans le cadre de la thèse

Dans ce manuscrit, le krigeage utilisé repose sur la boîte à outil ooDACE [44]. Dans ce cadre, l'optimisation des hyperparamètres peut être réalisée par maximum de vraisemblance ou leave-one-out avec un algorithme de quasi-Newton, utilisable grâce aux formules analytiques des gradients. Afin d'assurer une bonne estimation des hyperparamètres, les bornes et le point initial de l'optimisation doivent être correctement choisis pour éviter des phénomènes de surentrainement. Ce phénomène apparaît lorsqu'on souhaite trop reproduire les données observées, ce qui tend à créer un métamodèle instable et peu général. Le surentrainement peut se traduire par des valeurs extrêmes des hyperparamètres. Ici la distance entre les points d'entraînement est utilisée pour cadrer l'optimisation :

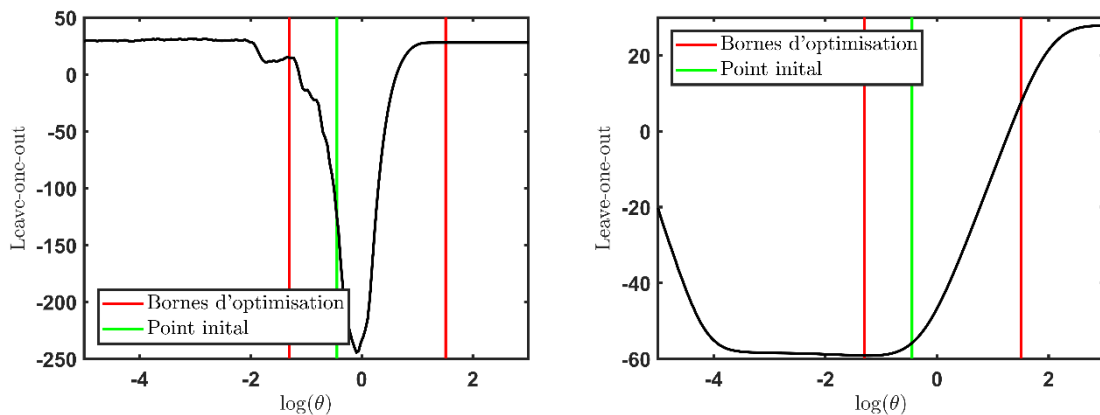
$$\theta_i \in \left[ \left( \frac{1}{\sqrt{2} \times \max \|\mathbf{x}_{tr}^{(i)} - \mathbf{x}_{tr}^i\|_1} \right)^2, \left( \frac{1}{\sqrt{2} \times \min \|\mathbf{x}_{tr}^{(i)} - \mathbf{x}_{tr}^{(i)}\|_1} \right)^2 \right], \quad i = 1, \dots, d. \quad [\text{II.42}]$$

Ce choix, de nature heuristique, est fondé sur une interprétation de la mise à l'échelle des données dans le cas gaussien. Le point initial de l'optimisation correspond à  $\theta_0 = 1/(\sqrt{2} \times \text{moy} \|\mathbf{X}_{tr} - \mathbf{X}_{tr}\|_1)^2$ , avec *moy* correspondant à la moyenne. Le résultat de ce choix est visible sur la Figure II-14, où le leave-one-out est utilisé pour l'estimation. Dans le cas où la fonction d'autocovariance est gaussienne, la fonction à optimiser comporte de nombreux minimums locaux ou plateaux. L'utilisation des bornes permet de cibler la zone du minimum global. Avec une fonction d'autocorrélation de type Matérn 5/2, la fonction est plus régulière mais les bornes sont un peu trop restrictives car elles ne permettent pas d'atteindre le minimum global.



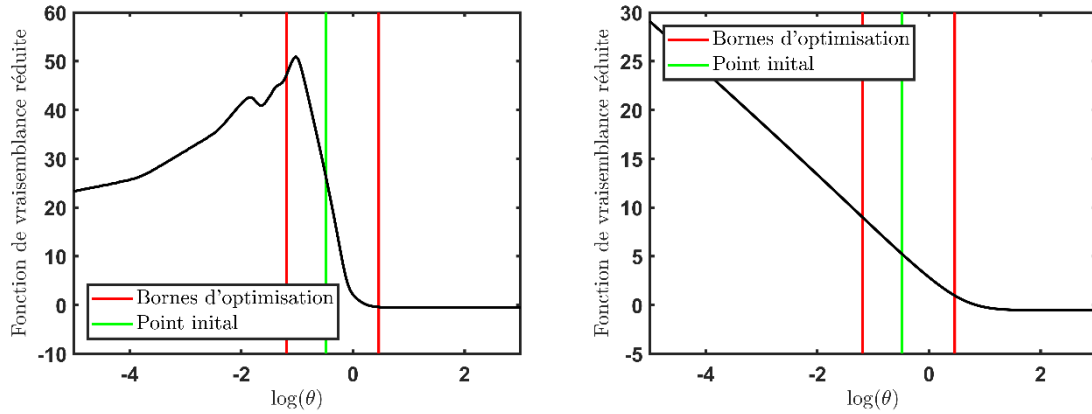
**Figure II-14 : Bornes d'optimisation des hyperparamètres avec estimation par leave-one-out et une fonction d'autocorrélation gaussienne (gauche) et Matérn 5/2 (droite). Exemple sur  $\mathcal{M}(x) = x \sin(x)$  avec 7 points d'entraînement.**

Lorsque le nombre de point d'entraînement augmente, les distances entre les points d'entraînement changent et permettent d'élargir les bornes d'optimisations (Figure II-15). La fonction du leave-one-out est moins oscillante dans le cas gaussien. Pour les deux fonctions d'autocovariance, le point initial de l'optimisation est très proche du minimum global.

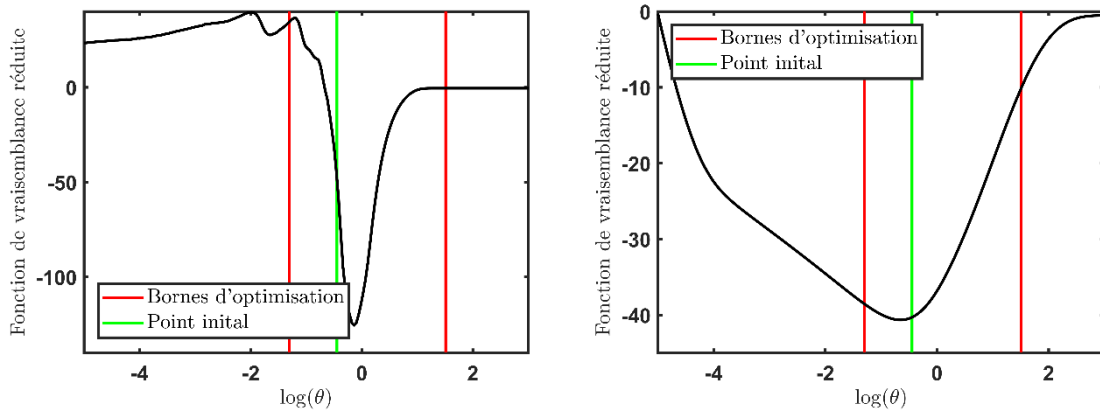


**Figure II-15 : Bornes d'optimisation des hyperparamètres avec estimation par leave-one-out et une fonction d'autocorrélation gaussienne (gauche) et Matérn 5/2 (droite). Exemple sur  $\mathcal{M}(x) = x \sin(x)$  avec 20 points d'entraînement.**

Avec le maximum de vraisemblance, la fonction à optimiser n'admet pas de minimum global net lorsque le nombre de points d'entraînement est insuffisant (Figure II-16). Il s'agit plutôt d'un plateau, et l'optimum est alors déterminé par la borne supérieure des hyperparamètres. En augmentant le nombre de points d'entraînement, un optimum global se dessine et le point initial de l'optimisation est à nouveau très proche de celui-ci (Figure II-17).

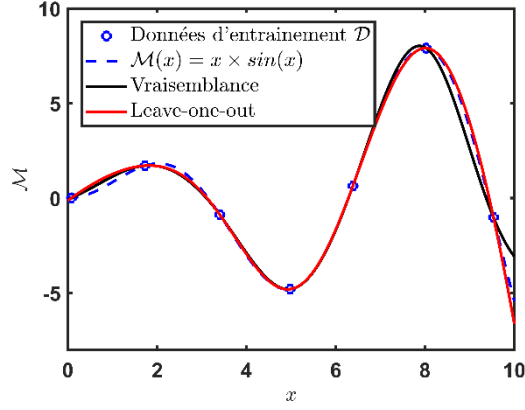


**Figure II-16 : Bornes d'optimisation des hyperparamètres avec estimation par maximum de vraisemblance et une fonction d'autocorrélation gaussienne (gauche) et Matérn 5/2 (droite). Exemple sur  $\mathcal{M}(x) = x \sin(x)$  avec 7 points d'entraînement.**



**Figure II-17 : Bornes d'optimisation des hyperparamètres avec estimation par maximum de vraisemblance et une fonction d'autocorrélation gaussienne (gauche) et Matérn 5/2 (droite). Exemple sur  $\mathcal{M}(x) = x \sin(x)$  avec 20 points d'entraînement.**

Le résultat de l'interpolation selon la méthode d'estimation des paramètres est visible sur la Figure II-18 avec la fonction d'autocovariance Matérn 5/2. Les deux krigeages sont assez proches et fournissent une bonne approximation. Celui obtenu à partir du leave-one-out est tout de même plus proche de la vraie courbe. Il ne faut cependant pas établir de généralités à partir de cet exemple. Une étude comparative est présentée en section II.D afin de mieux distinguer les caractéristiques des approximations selon les paramètres du krigeage.



**Figure II-18 : Krigeage avec autocovariance Matérn 5/2. Optimisation des hyperparamètres par maximum de vraisemblance et leave-one-out sur  $\mathcal{M}(x) = x \sin(x)$  avec 7 points d'entraînement.**

### 3. Fonctions à base radiale (RBF)

Les premiers travaux liés aux RBF sont apparus à la fin des années 1970 [45]. Depuis, elles n'ont cessé d'être étudiées pour comprendre leurs propriétés et elles restent un sujet de recherche actif au 21<sup>ème</sup> siècle. Les éléments disponibles dans cette section sont un extrait des propriétés décrites dans les livres de Buhman, Wendland et Schaback [46]–[48].

#### a. Définitions

Les RBF sont définies comme la combinaison linéaire de fonctions radiales  $\varphi_R$ . Pour rappel, une fonction est radiale si elle est de la forme :

$$\varphi_R: \begin{cases} \mathbb{R}^d \rightarrow \mathbb{R} \\ \mathbf{x} \mapsto \varphi_R(\|\mathbf{x}\|) \end{cases} \quad [\text{II.43}]$$

avec  $\|\cdot\|$  une norme usuelle sur  $\mathbb{R}^d$  (généralement la norme euclidienne). L'interpolation par RBF peut alors être décrite par l'équation suivante :

$$\hat{y}(\mathbf{x}) = \sum_{i=1}^n \beta_i \varphi_R(\mathbf{x} - \mathbf{x}_{tr}^{(i)}), \quad [\text{II.44}]$$

où les coefficients  $\beta_i$  sont réels. Pour construire l'approximation par RBF, il faut choisir une famille de fonctions de base puis estimer les coefficients  $\boldsymbol{\beta} = [\beta_1, \dots, \beta_n]$ . Une première condition à satisfaire est l'interpolation des données d'entraînement,  $\hat{y}(\mathbf{X}_{tr}) = \mathbf{Y}_{tr}$ . Cette condition amène à résoudre le système d'équation linéaires  $\mathbf{Y}_{tr} = \boldsymbol{\Phi}_{tr} \boldsymbol{\beta}$ , où :

$$\boldsymbol{\Phi}_{tr_{ij}} = \varphi_R(\mathbf{x}_{tr}^{(i)} - \mathbf{x}_{tr}^{(j)}). \quad [\text{II.45}]$$

Mais la résolution n'est pas toujours possible car, selon la fonction de base utilisé, la matrice  $\boldsymbol{\Phi}_{tr}$  peut être singulière. Afin de résoudre le système, une solution proposée par Micchelli [49] est de s'assurer que la matrice  $\boldsymbol{\Phi}_{tr}$  de la fonction de base est conditionnellement définie positive. Ceci permet de garantir l'unicité de la solution obtenue. On dit que la fonction radiale  $\varphi_R$  est conditionnellement définie positive d'ordre  $l$  si pour tous les points de  $\mathbf{X}_{tr}$  et pour tout  $\boldsymbol{\gamma} \in \mathbb{R}^n$  non nul satisfaisant l'équation [II.46], la forme quadratique de l'équation [II.47] est définie positive.

$$\sum_{i=1}^n \gamma_i p_j(\mathbf{x}_{tr}^{(i)}) = 0, \quad j = 1, \dots, l - 1. \quad [\text{II.46}]$$

$$\sum_{i=1}^n \sum_{j=1}^n \gamma_i \gamma_j \varphi_R(\mathbf{x}_{tr}^{(i)} - \mathbf{x}_{tr}^{(j)}). \quad [\text{II.47}]$$

Dans l'équation [II.46], les  $p_j$  sont les monômes de degré égal à  $j$ . Pour la suite, on note  $\mathcal{V}_l \subset \mathbb{R}^n$  l'espace vectoriel contenant tous les  $\boldsymbol{\gamma} \in \mathbb{R}^n$  satisfaisant l'équation [II.46]. Une fonction conditionnellement définie positive d'ordre  $l$  l'est également pour un ordre  $l' \geq l$ . L'ordre minimal est généralement celui qui est indiqué lorsqu'on parle de cette propriété. Si la fonction  $\varphi_R$  est conditionnellement définie positive d'ordre  $l$ , alors pour tout  $\boldsymbol{\gamma} \in \mathcal{V}_l \setminus \{\mathbf{0}\}$  :

$$\sum_{i=1}^n \sum_{j=1}^n \gamma_i \gamma_j \Phi_{tr\,ij} > 0. \quad [\text{II.48}]$$

Pour revenir au cas des RBF, l'inégalité suivante est obtenue avec l'ordre polynomial  $l_0$  dépendant de la fonction de base choisie :

$$(-1)^{l_0} \boldsymbol{\beta}^T \Phi_{tr} \boldsymbol{\beta} > 0, \forall \boldsymbol{\beta} \in \mathcal{V}_{l_0} \setminus \{\mathbf{0}\}. \quad [\text{II.49}]$$

L'interpolation par RBF dans le cas étendu aux fonctions conditionnellement définies positives s'écrit alors :

$$\hat{y}(\mathbf{x}) = \sum_{i=1}^n \beta_i \varphi_R(\mathbf{x} - \mathbf{x}_{tr}^{(i)}) + Q(\mathbf{x}), \quad [\text{II.50}]$$

avec le polynôme  $Q$  défini par la formulation suivante où  $l_0 - 1$  est son degré,  $\boldsymbol{\alpha}$  est un vecteur de coefficients réels :

$$Q(\mathbf{x}) = \sum_{k=1}^{l_0} \alpha_k p_{k-1}(\mathbf{x}). \quad [\text{II.51}]$$

Les coefficients  $\boldsymbol{\beta}$  et  $\boldsymbol{\alpha}$  de la RBF sont déterminés par la résolution du système suivant, représentant la condition d'interpolation et la condition de positivité de la matrice  $\Phi_{tr}$  :

$$\begin{pmatrix} \Phi_{tr} & \mathbf{F}_{tr} \\ \mathbf{F}_{tr}^T & \mathbf{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\alpha} \end{pmatrix} = \begin{pmatrix} \mathbf{Y}_{tr} \\ \mathbf{0} \end{pmatrix}, \quad [\text{II.52}]$$

où  $\mathbf{F}_{tr} = (p_0(\mathbf{X}_{tr}), \dots, p_{l_0-1}(\mathbf{X}_{tr}))$ . En utilisant les coefficients obtenus avec ce système dans l'équation [II.49], il est possible d'évaluer le modèle  $\mathcal{M}$  en de nouveaux points. Une mesure de l'erreur de prédiction de la RBF peut également être fournie. Cette mesure est référencée comme la fonction puissance  $P_\varphi$  (*power function dans [50] - section 14.4*) :

$$P_\varphi(\mathbf{x}) = \sqrt{(-1)^{l_0} \times \left( \varphi_R(0) - \begin{pmatrix} \mathbf{u}_x^T & \mathbf{F}_x^T \end{pmatrix} \begin{pmatrix} \Phi_{tr} & \mathbf{F}_{tr} \\ \mathbf{F}_{tr}^T & \mathbf{0} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{u}_x \\ \mathbf{F}_x \end{pmatrix} \right)} \quad [\text{II.53}]$$

avec  $\mathbf{u}_x = [\varphi_R(\mathbf{x} - \mathbf{x}_{tr}^{(1)}), \dots, \varphi_R(\mathbf{x} - \mathbf{x}_{tr}^{(n)})]^T$  et  $\mathbf{F}_x = [p_0(\mathbf{x}), \dots, p_{l_0-1}(\mathbf{x})]^T$ . La fonction puissance permet de quantifier l'erreur dans l'interpolation due au choix de la fonction radiale, de ses paramètres et de la position des points d'entraînement.

## b. Fonctions radiales dans le cadre de l'étude

Pour réaliser une interpolation de données avec les RBF, il ne reste plus qu'à choisir la catégorie de fonctions de base. Une partie des fonctions potentiellement disponibles est présentée dans la section suivante. Comme dans le cadre du krigeage avec les fonctions d'autocorrélation (section II.B.2.c), l'influence de

chaque variable sur l'interpolation du modèle  $\mathcal{M}$  peut être prise en compte séparément dans les fonctions radiales par anisotropie géométrique. Cette pratique est aussi mise en oeuvre dans les travaux de Mackman et al. [51]. Au lieu d'utiliser simplement la norme euclidienne dans la définition de  $\varphi_R$ , on utilisera la norme  $\|\mathbf{x} - \mathbf{x}'\|_{\boldsymbol{\theta}} = \sqrt{(\mathbf{x} - \mathbf{x}')^T \mathbf{A} (\mathbf{x} - \mathbf{x}' )}$  avec  $\mathbf{A} = \text{diag}(\boldsymbol{\theta})$ . Pour rappel,  $\boldsymbol{\theta}$  est vecteur positif de  $d$  paramètres d'échelles représentant l'influence spatiale de chaque dimension de l'espace  $\mathbb{X} \subset \mathbb{R}^d$ . Dans le cadre des RBF,  $\boldsymbol{\theta}$  est appelé paramètre de forme. Par soucis de cohérence avec le krigeage, nous continuerons d'utiliser le terme hyperparamètre pour faire référence à  $\boldsymbol{\theta}$ .

Les fonctions radiales utilisées dans le cadre de cette étude sont présentées dans le Tableau 3, conjointement avec l'ordre  $l_0$  du polynôme  $Q$ . Si l'ordre est égal à 0, par convention cela signifie que  $Q \in \Pi_{-1} \equiv \emptyset$ , c'est-à-dire que la fonction  $\varphi_R$  est strictement définie positive et l'interpolation par RBF se ramène dans ce cas à l'équation [II.44]. Par ailleurs, la liste de fonctions radiales présente ici n'est pas exhaustive. Des généralisations de certaines expressions existent (voir [50] - section 4 et 8) et on peut également citer les fonctions radiales à support compact ([47] - section 9.4). Pour correspondre au cadre défini par Hardy [45], les fonctions multiquadrique et inverse multiquadrique sont redéfinies à partir de la norme euclidienne. L'hyperparamètre est dans ce cas un facteur à effet par addition.

	$\varphi_R(\mathbf{x} - \mathbf{x}')$	Ordre $l_0$
Linéaire	$\ \mathbf{x} - \mathbf{x}'\ _{\boldsymbol{\theta}}$	1
Cubique	$(\ \mathbf{x} - \mathbf{x}'\ _{\boldsymbol{\theta}})^3$	2
Spline en plaque mince	$\ \mathbf{x} - \mathbf{x}'\ _{\boldsymbol{\theta}}^2 \ln(\ \mathbf{x} - \mathbf{x}'\ _{\boldsymbol{\theta}})$	2
Multiquadrique	$\sqrt{\sum_{i=1}^d \left( \frac{1}{\theta_i^2} + \ x_i - x'_i\ _2^2 \right)}$	1
Inverse multiquadrique	$\left( \sum_{i=1}^d \left( \frac{1}{\theta_i^2} + \ x_i - x'_i\ _2^2 \right) \right)^{-1/2}$	0
Matérn 3/2	$(1 + \sqrt{3}\ \mathbf{x} - \mathbf{x}'\ _{\boldsymbol{\theta}}) \times \exp(-\sqrt{3}\ \mathbf{x} - \mathbf{x}'\ _{\boldsymbol{\theta}})$	0
Matérn 5/2	$\left( 1 + \sqrt{5}\ \mathbf{x} - \mathbf{x}'\ _{\boldsymbol{\theta}} + \frac{5}{3}\ \mathbf{x} - \mathbf{x}'\ _{\boldsymbol{\theta}}^2 \right) \times \exp(-\sqrt{5}\ \mathbf{x} - \mathbf{x}'\ _{\boldsymbol{\theta}})$	0
Gaussienne	$\exp(-\ \mathbf{x} - \mathbf{x}'\ _{\boldsymbol{\theta}}^2)$	0

**Tableau 3 : Fonctions radiales utilisées dans le cadre de cette étude et leur ordre  $l_0$ .**

Plusieurs fonctions radiales du Tableau 3 sont déjà présentées en section II.B.2.c dans le cadre des fonctions d'autocorrélations du krigeage. Seules les nouvelles fonctions sont illustrées sur la Figure II-19, pour différentes valeurs d'hyperparamètres. A part la fonction inverse multiquadrique, ces fonctions tendent toutes vers des valeurs élevées lorsque la distance entre les points augmente. Les valeurs de la fonction spline en plaque mince sont négatives pour de faibles valeurs de distance entre les points. Le choix des valeurs des hyperparamètres est abordé dans la section suivante.

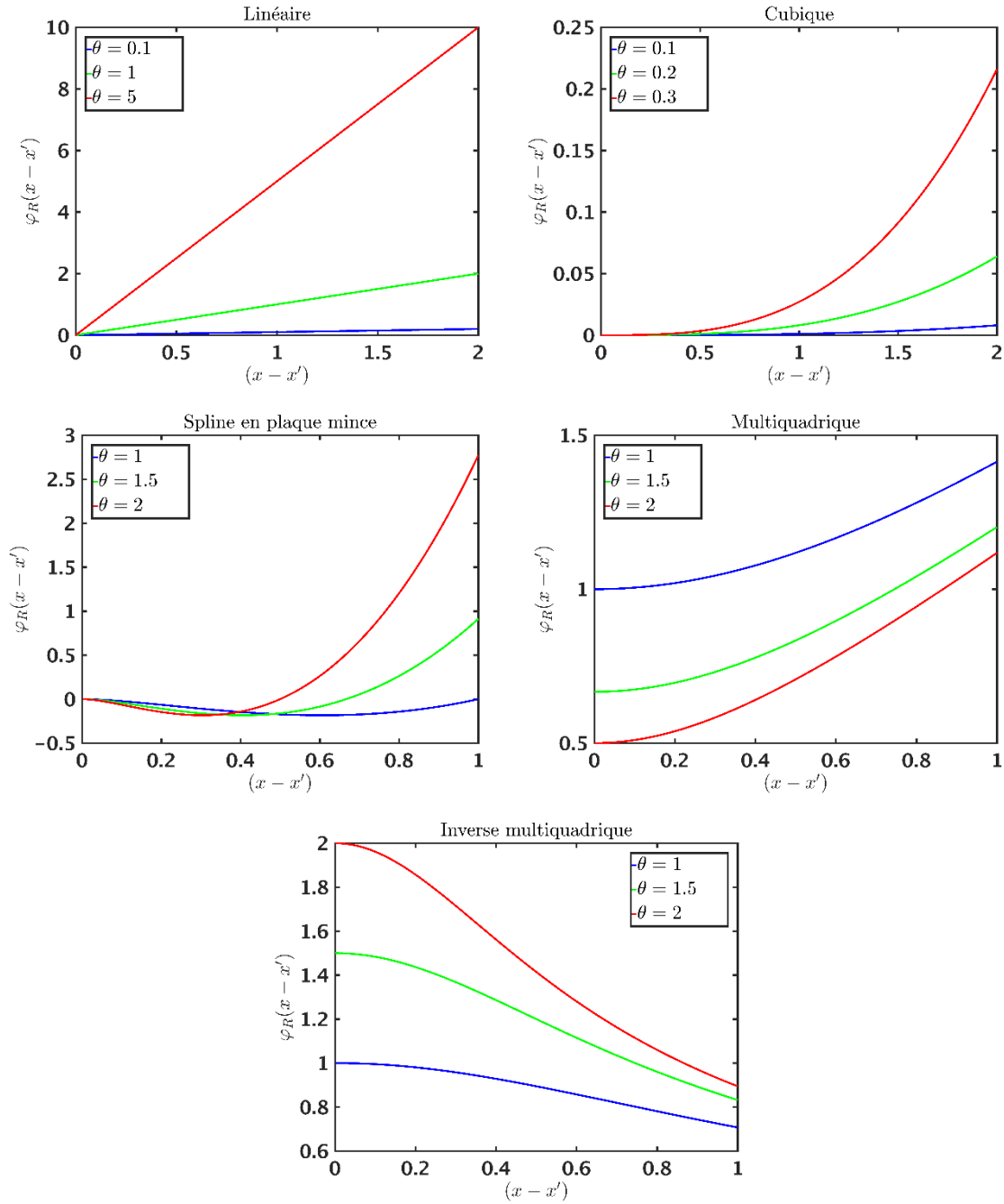


Figure II-19 : Illustrations des fonctions radiales de type linéaire, cubique, spline en plaque en mince, multiquadrique et inverse multiquadrique.

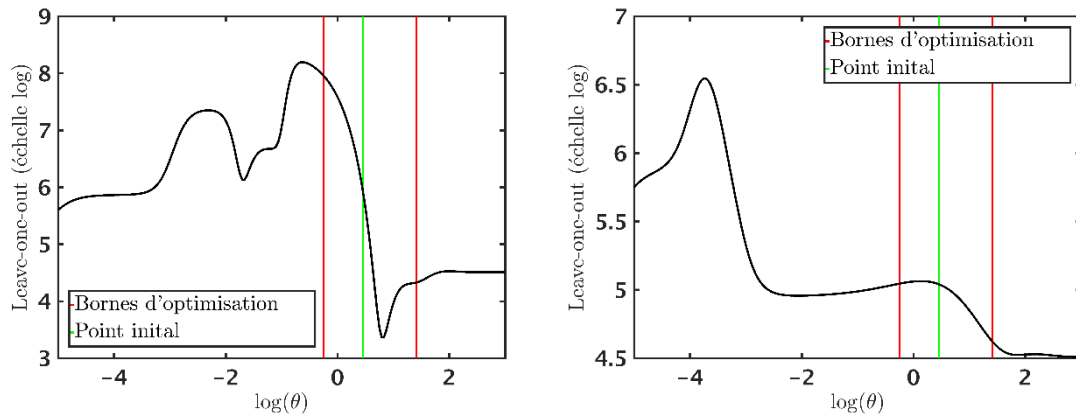
### c. Estimation des hyperparamètres par leave-one-out

Lors de la création d'une RBF, en plus de l'obtention des coefficients  $\beta$  et  $\alpha$ , l'estimation des hyperparamètres  $\theta$  doit être réalisée. Ripplà [52] développe dans ses travaux un critère d'erreur basé sur la méthode du leave-one-out (principe précédemment expliqué en section II.B.2.e comme un cas particulier de la validation croisée). Ce critère permet l'estimation de l'erreur LOO en une seule résolution du système à l'équation [II.52] :

$$LOO_{RBF} = \sum_{i=1}^n \left( \frac{\beta_i}{(\Phi_{tr}^{-1})_{ii}} \right)^2 \quad [II.54]$$

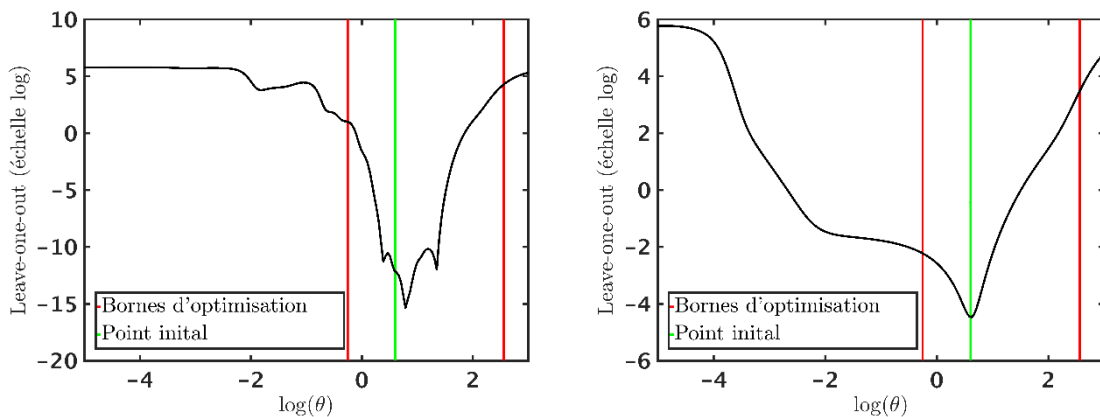
Le critère  $LOO_{RBF}$  est représenté en fonction des hyperparamètres sur la Figure II-20 pour deux fonctions

radiales différentes, lors de l'apprentissage d'une fonction exemple avec seulement 7 points d'entraînement. Ce critère présente plusieurs minimums locaux, justifiant l'utilisation de bornes pour faciliter son optimisation. Dans ce manuscrit, ces bornes sont calculées à partir de la distance entre les points d'entraînements, de même que pour le krigeage (voir équation [II.42]). Pour l'exemple traité avec la fonction radiale Matérn 5/2, la présence de bornes permet d'éviter que la valeur de l'hyperparamètre ne soit trop élevée, ce qui conduirait à un surentraînement.



**Figure II-20 : Bornes d'optimisation des hyperparamètres avec estimation par leave-one-out et une fonction radiale gaussienne (gauche) et Matérn 5/2 (droite). Exemple sur  $\mathcal{M}(x) = x \sin(x)$  avec 7 points d'entraînement.**

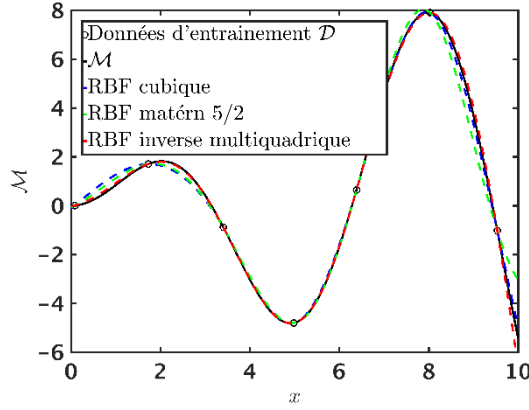
Lorsqu'on augmente le nombre de points d'entraînement, un minimum global se dessine nettement pour le cas Matérn 5/2 (Figure II-21). Avec le cas gaussien, le bassin du minimum global présente quelques minimums locaux. Toutefois, l'utilisation de la distance moyenne entre les points d'entraînement permet de fournir dans les deux cas un point de départ proche du minimum global.



**Figure II-21 : Bornes d'optimisation des hyperparamètres avec estimation par leave-one-out et une fonction radiale gaussienne (gauche) et Matérn 5/2 (droite). Exemple sur  $\mathcal{M}(x) = x \sin(x)$  avec 20 points d'entraînement.**

L'interpolation obtenue en estimant les hyperparamètres par leave-one-out est visible sur la Figure II-22 pour différentes fonctions à base radiales, toujours sur l'exemple  $\mathcal{M}(x) = x \sin(x)$ . Dans cet exemple assez simple, les prédictions des RBF sont très proches de la réalité avec peu de points d'entraînement. Des comparaisons plus poussées entre les différentes bases radiales sur des exemples physiques sont présentées en section II.D.





**Figure II-22 : RBF avec fonction radiale cubique, matérn 5/2 et inverse multiquadratique. Optimisation des hyperparamètres par leave-one-out sur  $\mathcal{M}(x) = x \sin(x)$  avec 7 points d'entraînement.**

d. Note sur la différence entre krigeage et RBF

A l'issue de la présentation du krigeage et des RBF, des analogies apparaissent de manière évidente entre les deux métamodèles. Leur relation est explicitement formulée dans les travaux de Costa et al. [53]. En fait, la différence entre les prédictions réside seulement dans le choix de fonction de corrélation (ou radiale) et dans l'estimation des hyperparamètres. Par contre, les hypothèses nécessaires pour construire ces métamodèles et les conventions issues de différentes branches des mathématiques entraînent des propriétés différentes. Ainsi l'hypothèse définissant les évaluations du modèle numérique comme la réalisation d'un processus gaussien permet l'obtention d'un intervalle de confiance pour la prédiction par krigeage, ce qui n'est pas le cas des RBF. Pour autant si la même fonction de corrélation est choisie avec les mêmes hyperparamètres, les deux métamodèles fournissent exactement la même réponse.

4. Critères de validation du métamodèle

Pour attester de la précision d'un métamodèle, un jeu de données n'ayant pas été utilisé pour sa construction est nécessaire. On parlera alors des données de test  $\mathcal{D}_{test} = \{\mathcal{X}_{test}, \mathcal{Y}_{test}\}$  représentant les  $n_{test}$  couples points d'évaluations et valeurs observées sur le modèle numérique d'intérêt. La validation du métamodèle a donc un coût supplémentaire. Pour éviter d'avoir à calculer le jeu de données  $\mathcal{D}_{test}$ , il est possible de recourir à la validation croisée. Toutefois son usage n'est pas recommandé car la validation croisée permet simplement d'attester la sensibilité du métamodèle à la perte d'information [26]. Il est donc possible de considérer que le métamodèle est valide alors qu'il possède une mauvaise précision générale. Ou au contraire de rejeter un métamodèle car il est sensible à la perte d'information alors qu'il possède une bonne qualité de prédiction globale. Des nouvelles données de test sont donc utilisées dans ce travail pour valider la qualité du métamodèle. Parmi les choix possibles de critères de validation, quatre sont sélectionnés ici pour mesurer les erreurs locales et globales du modèle de substitution. Pour quantifier l'erreur globale, la moyenne de l'erreur absolue divisée par l'écart type (RAAE, pour Relative Average Absolute Error) est calculée :

$$RAAE(\mathbf{x}_{test}) = \frac{\sum_{i=1}^{n_{test}} |y_{test}^{(i)} - \hat{y}(\mathbf{x}_{test}^{(i)})|}{n_{test} \times \sigma_{y_{test}}} \quad [II.55]$$

où  $\sigma_{y_{test}}$  est l'écart type des valeurs observées sur les points de tests. La valeur fournie est liée à l'échelle de l'espace de sortie du modèle. Plus elle est faible, plus le métamodèle est précis. L'erreur moyenne quadratique est également très utilisée pour quantifier l'erreur globale. Pour l'erreur locale, le maximum de l'erreur absolue divisée par l'écart type (RMAE, pour Relative Maximum Absolute Error) est choisi :

$$\text{RMAE}(\mathbf{x}_{test}) = \frac{\max_{i=1, \dots, n_{test}} |y_{test}^{(i)} - \hat{y}(\mathbf{x}_{test}^{(i)})|}{\sigma_{y_{test}}} \quad [\text{II-56}]$$

Ce critère complète la RAAE. Il traduit le plus grand écart entre valeurs prédites et valeurs observées sur les points tests. Parmi les autres critères couramment utilisés, on retrouve la RMSE (pour Root Mean Squared Error) :

$$\text{RMSE}(\mathbf{x}_{test}) = \sqrt{\frac{1}{n_{test}} \sum_{i=1}^{n_{test}} (y_{test}^{(i)} - \hat{y}(\mathbf{x}_{test}^{(i)}))^2} \quad [\text{II-57}]$$

Plus la valeur de la RMSE est faible et plus le métamodèle a une capacité de prédiction générale qui est bonne. Il y a également le critère du  $R^2$  qui est très utilisé pour quantifier l'erreur globale :

$$R^2 = 1 - \frac{\sum_{i=1}^{n_{test}} (y_{test}^{(i)} - \hat{y}(\mathbf{x}_{test}^{(i)}))^2}{\sigma_{y_{test}}^2} \quad [\text{II-58}]$$

Plus le critère  $R^2$  est proche de 1 et plus la qualité de prédiction du métamodèle est intéressante.

### C. Métamodèles multifidélité

Il est parfois possible d'obtenir plusieurs modèles numériques de fidélité différente pour le même phénomène physique. Le modèle le plus précis est généralement celui qui est le plus coûteux en temps de calcul. La métamodélisation multifidélité consiste alors à utiliser l'information provenant des différents modèles de fidélité différente pour ne former qu'une seule réponse. Parmi les cas d'études rencontrés, les modèles de la cellule photoacoustique correspondent au cadre de la multifidélité. Le modèle FLNS est précis mais coûteux en temps de calcul. Le modèle Kreuzer est rapide à évaluer mais correspond moins aux données recueillies en expérimentation. Il existe également d'autres exemples possibles dans le cadre de la résolution des équations de Maxwell avec la combinaison d'un modèle BPM avec un modèle FDTD (différences finies dans le domaine temporel).

La construction d'un métamodèle multifidélité permet de prédire le comportement du modèle le plus fidèle. L'approche consiste à recaller la tendance obtenue sur le modèle basse fidélité en utilisant quelques points de données du modèle haute fidélité. Le plan d'expérience d'apprentissage est alors constitué d'un grand nombre d'évaluations sur les modèles peu coûteux et de peu de points pour les modèles de fidélité plus élevées. Ainsi le coût de construction du métamodèle multifidélité est plus faible comparé à la construction d'un métamodèle uniquement sur un grand nombre de données haute fidélité. Plusieurs travaux ont été effectués sur les différentes manières de regrouper les modèles de fidélité différentes et sont présentés dans une récente revue [54]. Le krigeage est le métamodèle le plus présent dans les travaux analysés après les surfaces de réponses. Sa version multifidélité s'appelle le co-krigeage. Cette dénomination est également rencontrée lorsque le krigeage est construit en utilisant à la fois les valeurs des grandeurs physiques à approcher et leurs gradients [55], ce qui peut être une source d'ambiguïté.

Il existe des cas où une seule modélisation du phénomène physique est disponible et l'obtention d'une version approchée moins coûteuse consiste à dégrader le maillage dans une simulation numérique d'équation aux dérivées partielles, ou d'utiliser des solutions partiellement convergées en arrêtant prématurément la résolution. La covariance utilisée pour le krigeage est alors modifiée afin de refléter la structure de l'erreur qui dépend du niveau de convergence de la simulation [56]. Dans le cadre de ce manuscrit, différentes modélisations du même phénomène sont disponibles grâce à une réduction de la complexité par des approximations physiques. L'approche du co-krigeage par modèle autorégressif correspond à ce cadre.

## 1. Co-krigeage par modèle autorégressif

Le modèle autorégressif a été utilisé dans le cadre de la métamodélisation multifidélité par Kennedy et O'Hagan [57]. Une généralisation de cette formulation a été proposée par la suite par Qian et Wu [58]. Cette méthode s'applique pour un grand nombre de modélisations de différente fidélité mais seul le cas à deux niveaux de fidélité est traité ici par souci de lisibilité. Le modèle de plus haute fidélité (également le plus coûteux) correspond à  $\mathcal{M}_{HF}$  et celui de plus basse fidélité à  $\mathcal{M}_{LF}$ . De même, dans la suite du document, l'indice HF indique les variables liées au modèle haute fidélité et l'indice LF celles du modèle basse fidélité (pour « low fidelity »). Si les grandeurs physiques en sortie de ces modèles n'ont aucune corrélation ou caractéristique en commun, l'apport de l'approche multifidélité est nul [59].

Comme pour le krigeage, les sorties des modèles sont supposées être issues de la réalisation de processus gaussiens, notés  $Z_{HF}$  et  $Z_{LF}$ . Deux plans d'expérience,  $\mathcal{D}_{HF} = \{\mathcal{X}_{HF}, \mathcal{Y}_{HF}\}$  et  $\mathcal{D}_{LF} = \{\mathcal{X}_{LF}, \mathcal{Y}_{LF}\}$ , contenant des données évaluées à l'aide des deux modèles sont nécessaires pour l'entraînement du co-krigeage. On note également  $\mathbf{Y}_{HF}$  et  $\mathbf{Y}_{LF}$  les matrices contenant toutes les observations  $\mathcal{Y}_{HF}$  et  $\mathcal{Y}_{LF}$ , ainsi que  $\mathbf{X}_{HF}$  et  $\mathbf{X}_{LF}$  les matrices contenant tous les points d'entraînements  $\mathcal{X}_{HF}$  et  $\mathcal{X}_{LF}$ . On suppose ici que les deux plans sont imbriqués tel que  $\mathcal{X}_{HF} \subset \mathcal{X}_{LF}$ . Ils sont tous les deux définis sur  $\mathbb{X} \subset \mathbb{R}^d$ .

Le but est maintenant de déterminer la distribution conditionnelle du processus gaussien  $Z_{HF}$  connaissant  $\mathbf{Y}_{LF}$ , soit  $[Z_{HF}(\mathbf{x}) | Z_{HF}(\mathbf{X}_{HF}) = \mathbf{Y}_{HF}, Z_{LF}(\mathbf{X}_{LF}) = \mathbf{Y}_{LF}]$  pour  $\mathbf{x} \in \mathbb{X}$ . L'hypothèse suivante a été introduite par Kennedy et O'Hagan [57] :

$$\text{Cov}(Z_{HF}(\mathbf{x}), Z_{LF}(\mathbf{x}') | Z_{LF}(\mathbf{x})) = 0, \forall \mathbf{x} \neq \mathbf{x}' \quad [\text{II.59}]$$

Elle signifie que si  $Z_{LF}(\mathbf{x})$  est connu, plus rien ne peut être appris sur  $Z_{HF}(\mathbf{x})$  avec une autre évaluation de  $Z_{LF}$ . Cette hypothèse permet de définir le modèle autorégressif suivant (voir l'annexe A.1 de [60] pour la preuve) :

$$Z_{HF}(\mathbf{x}) = \rho_{LF}(\mathbf{x}) \times Z_{LF}(\mathbf{x}) + Z_D(\mathbf{x}) \quad [\text{II.60}]$$

Le modèle autorégressif définit le modèle HF comme le produit d'un facteur d'échelle  $\rho_{LF}$  avec le modèle LF, additionné à un processus gaussien  $Z_D$  représentant la différence entre les deux modèles. Ici le facteur d'échelle est considéré comme étant constant :  $\rho_{LF}(\mathbf{x}) = \rho_{LF}$ . Dans la méthode d'évofusion de Forrester et al. [61], seul un krigeage de la différence est nécessaire pour obtenir le metamodèle multifidélité. Pour prédire la réponse HF, il faut d'abord évaluer le modèle LF puis appliquer la correction fournie par la prédiction de  $Z_D$ . Pour le modèle autorégressif, la construction du co-krigeage consiste à estimer les paramètres des processus  $Z_{LF}$  et  $Z_D$  pour ensuite utiliser l'équation [II.60] et prédire la réponse du modèle HF. Le processus gaussien  $Z_{LF}$  est supposé être indépendant de  $Z_D$ . Son entraînement se ramène donc au cas classique du krigeage décrit en section II.B.2, permettant d'obtenir :

$$\hat{y}_{LF}(\mathbf{x}) = \mathbb{E}[Z_{LF}(\mathbf{x}) | Z_{LF}(\mathbf{X}_{LF}) = \mathbf{Y}_{LF}] = \mu_{LF} + \boldsymbol{\Sigma}_{LF} \mathbf{R}_{LF}^{-1} (\mathbf{Y}_{LF} - \mu_{LF}) \quad [\text{II.61}]$$

Avec  $\mu_{LF}$  la tendance moyenne,  $\boldsymbol{\Sigma}_{LF}$  la matrice covariance entre les données d'entraînement et les points à prédire et  $\mathbf{C}_{LF}$  la matrice de covariance des données d'entraînement. D'autres hypothèses sont possibles pour réduire le nombre de paramètres à estimer, telle que la corrélation croisée de Zimmermann et al. [62]. Dans ses travaux, la structure de covariance est supposée être égale entre les données HF et LF. Il n'y a alors qu'un seul paramètre supplémentaire représentant la corrélation entre les deux modèles en plus des longueurs de corrélations à estimer durant la phase d'entraînement. Dans le cadre du modèle autorégressif, de nouvelles données d'entraînement  $\mathcal{D}_D = \{\mathcal{X}_{HF}, \mathcal{Y}_D\}$  sont définies pour estimer les paramètres de  $Z_D$  :

$$\mathbf{Y}_D = \mathbf{Y}_{HF} - \rho_{LF} \hat{y}_{LF}(\mathbf{X}_{HF}) \quad [\text{II.62}]$$

Si l'hypothèse  $\mathcal{X}_{HF} \subset \mathcal{X}_{LF}$  a été respectée,  $\hat{y}_{LF}(\mathbf{X}_{HF}) = \mathcal{M}(\mathbf{X}_{HF})$  et l'estimation des paramètres de  $Z_D$  ne dépendra pas de l'estimation de ceux de  $Z_{LF}$ . Si les plans d'expériences d'entraînement ne sont pas imbriqués, une mauvaise estimation des paramètres de  $Z_{LF}$  peut perturber l'estimation de ceux de  $Z_D$ . Dans

tous les cas, le paramètre  $\rho_{LF}$  doit être estimé lors de l'entraînement de  $Z_D$ . Forrester et al. [63] proposent de l'inclure dans les variables d'optimisation de la fonction de vraisemblance ou de la validation croisée (voir section II.B.2.e) car ces critères dépendent de  $\mathbf{Y}_D$ . Après avoir choisi les fonctions de corrélation  $R_{LF}$  et  $R_D$  pour  $Z_{LF}$  et  $Z_D$ , l'application de l'ensemble de la procédure décrite ci-dessus permet d'obtenir les paramètres  $\mu_{LF}, \sigma_{LF}^2, \boldsymbol{\theta}_{LF}, \mu_D, \sigma_D, \boldsymbol{\theta}_D$  correspondant respectivement à la tendance, la variance et aux hyperparamètres des deux processus gaussiens.

En reprenant les résultats sur la prédiction par processus gaussien en section II.B.2.d, les données d'entraînement sont la réalisation d'un vecteur aléatoire  $[\mathbf{Z}_{HF}, \mathbf{Z}_{LF}] \sim \mathcal{N}[(\mu_{LF}, \mu_{HF}), \mathbf{C}_{co}]$  avec :

$$\mathbf{C}_{co} = \begin{pmatrix} \mathbf{C}_{LF} & \mathbf{C}_{LF/HF} \\ \mathbf{C}_{LF/HF} & \mathbf{C}_{HF} \end{pmatrix}, \quad [\text{II.63}]$$

et

$$\begin{aligned} \mathbf{C}_{LFij} &= \sigma_{LF}^2 R_{LF}(\mathbf{X}_{LF}^{(i)}, \mathbf{X}_{LF}^{(j)}), \quad i, j = 1, \dots, n_{LF} \\ \mathbf{C}_{LF/HFij} &= \rho_{LF} \sigma_{LF}^2 R_{LF}(\mathbf{X}_{LF}^{(i)}, \mathbf{X}_{HF}^{(j)}), \quad i = 1, \dots, n_{LF}; j = 1, \dots, n_{HF} \\ \mathbf{C}_{HFij} &= \rho_{LF}^2 \sigma_{LF}^2 R_{LF}(\mathbf{X}_{HF}^{(i)}, \mathbf{X}_{HF}^{(j)}) + \sigma_D^2 R_D(\mathbf{X}_{HF}^{(i)}, \mathbf{X}_{HF}^{(j)}), \quad i, j = 1, \dots, n_{HF} \end{aligned} \quad [\text{II.64}]$$

L'espérance de la distribution conditionnelle  $[Z_{HF}(\mathbf{x}) | Z_{HF}(\mathbf{X}_{HF}) = \mathbf{Y}_{HF}, Z_{LF}(\mathbf{X}_{LF}) = \mathbf{Y}_{LF}]$  peut alors être définie comme :

$$\hat{y}_{HF}(\mathbf{x}) = \mu_{HF} + \boldsymbol{\Sigma}_{co} \mathbf{C}_{co}^{-1} \mathbf{Y}_{co}, \quad [\text{II.65}]$$

où  $\mu_{HF} = \rho_{LF} \mu_{LF} + \mu_D$ ,  $\mathbf{Y}_{co} = [\mathbf{Y}_{LF} - \mu_{LF}, \mathbf{Y}_{HF} - \mu_{HF}]^T$  et  $\boldsymbol{\Sigma}_{co} = [\boldsymbol{\Sigma}_{LF/HF}, \boldsymbol{\Sigma}_{HF}]^T$  avec :

$$\begin{aligned} \boldsymbol{\Sigma}_{LF/HF}(\mathbf{x}) &= \rho_{LF} \sigma_{LF}^2 R_{LF}(\mathbf{X}_{LF}, \mathbf{x}) \\ \boldsymbol{\Sigma}_{HF}(\mathbf{x}) &= \rho_{LF}^2 \sigma_{LF}^2 R_{LF}(\mathbf{X}_{HF}, \mathbf{x}) + \sigma_D^2 R_D(\mathbf{X}_{HF}, \mathbf{x}) \end{aligned} \quad [\text{II.66}]$$

La variance de prédiction s'obtient également à partir de la distribution conditionnelle et s'exprime sous la forme :

$$\hat{\sigma}_{\hat{y}_{HF}}^2(\mathbf{x}) = \rho_{LF}^2 \sigma_{LF}^2 + \sigma_D^2 - \boldsymbol{\Sigma}_{co} \mathbf{C}_{co}^{-1} \boldsymbol{\Sigma}_{co}^T + \frac{(1 - \mathbf{1}^T \mathbf{C}_{co}^{-1} \boldsymbol{\Sigma}_{co}^T)^2}{\mathbf{1}^T \mathbf{C}_{co}^{-1} \mathbf{1}} \quad [\text{II.67}]$$

Les équations [II.65] et [II.67] sont ici exprimées pour le cas du co-krigeage ordinaire par modèle autorégressif. Le développement de ces formules est disponible dans l'annexe A de [63]. Une définition plus générale du co-krigeage est décrite dans la thèse de Le Gratiet ([60] - chapitre 2). Pour illustrer les capacités de la multifidélité, le krigeage et le co-krigeage sont comparés sur la Figure II-23 avec l'exemple 1D de Forrester ([63] - section 3.a). Les quatre points d'entraînement sur le modèle haute fidélité ne permettent pas au krigeage de bien substituer le modèle  $\mathcal{M}$ . En rajoutant une information à partir de données d'un modèle basse fidélité, la prédiction par co-krigeage est plus proche de la réalité.

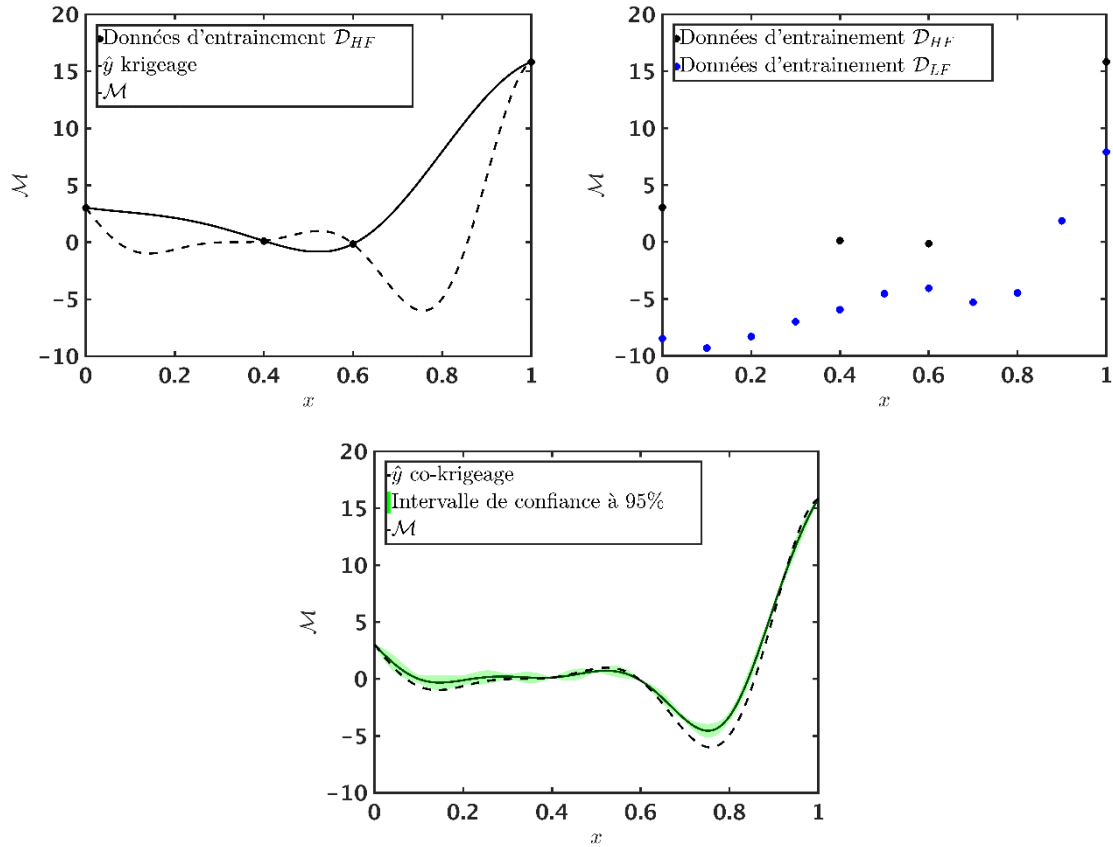


Figure II-23 : Comparaison krigeage et co-krigeage sur la fonction exemple de Forrester.

## D. Etude comparative dans le cadre des d'applications considérées

Dans cette section, les modèles numériques des différents cas d'études présentés au chapitre I sont substitués par des métamodèles. Le but est de déterminer les paramètres du krigeage ou des RBF qui prédisent au mieux les grandeurs physiques à approximer. L'ensemble des fonctions de corrélation et fonctions radiales sont testées et leur influence sur l'erreur de prédiction est analysée. Comme précisé en section II.B.4 ,l'erreur de prédiction est calculée à partir d'un jeu de données indépendant, comportant un grand nombre de points par rapport au nombre de points d'entraînement.

### 1. Microsource infrarouge

Le cas d'étude de la microsource a été présenté en section I.C. Son modèle comporte 9 paramètres en entrée : 8 paramètres géométriques décrits sur la Figure I-12 ( $e_1$  est fixé à  $1 \mu\text{m}$ ) et la tension aux bornes des pistes électrique  $U_{source}$ . Les bornes de variations des paramètres pour cette étude sont disponibles dans le Tableau 4.

	$r_1$	$r_2$	$r_3$	$e_2$	$e_3$	$w_1$	$w_2$	$w_3$	$U_{source}$
Borne inférieure	10	10	10	1	1	4	4	4	1.3
Borne supérieure	117	117	117	3	3	35	35	35	1.9

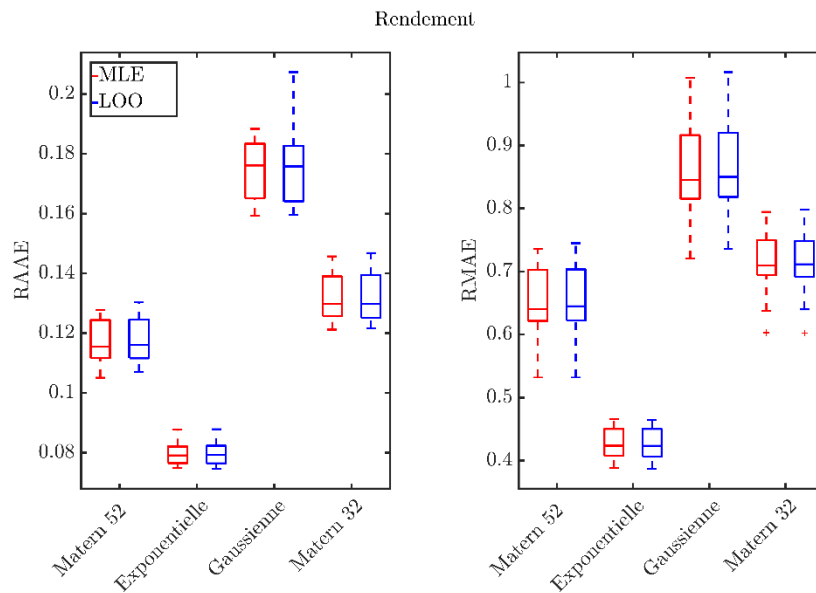
Tableau 4 : Bornes des paramètres du modèle de la microsource infrarouge (en  $\mu\text{m}$  pour la géométrie et en V pour la tension).

Les points d'évaluation doivent respecter certaines contraintes linéaires issues de tolérances géométriques

(espacement minimal d'un micromètre entre pistes circulaires et tailles décroissantes des pistes d'alimentation). L'espace admissible est ainsi défini par :

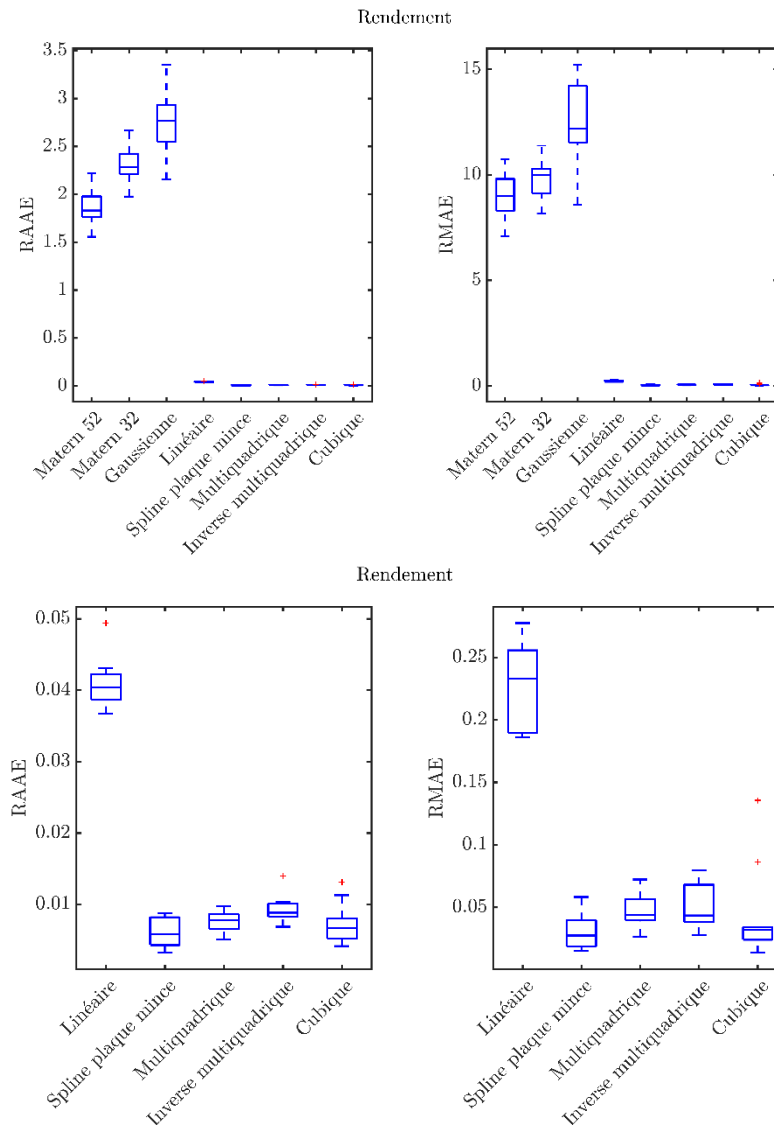
$$\begin{aligned}
 r_1 + 1 &\leq r_2 \\
 r_2 + 1 &\leq r_3 \\
 w_1 &\leq w_2 \\
 w_2 &\leq w_3
 \end{aligned}
 \tag{II.68}$$

Pour générer un plan d'expérience satisfaisant ces contraintes sur les paramètres, les suites de Sobol' agrégées par k-moyenne sont utilisées (voir section II.B.1.a). Un grand nombre de points est généré avec les suites, puis les points non admissibles sont retirés du plan et finalement l'algorithme k-moyenne est appliqué. Après évaluation, les grandeurs physiques étudiées sont le rendement de la source, la température moyenne et la température maximum du disque central. Un plan d'expérience de 300 points est utilisé pour calculer l'erreur de prédiction. Pour l'entraînement des métamodèles, 10 plans d'expériences de 50 points sont constitués pour moyenner les résultats sur les différentes réalisations des plans d'expérience. Sur la Figure II-24, l'erreur de prédiction du krigeage est représentée en fonction de la fonction de corrélation et de la méthode d'estimation des hyperparamètres. MLE correspond au maximum de vraisemblance et LOO correspond à la pseudovraisemblance. Ici, les résultats ne varient pas ou peu selon la méthode d'estimation utilisée. Par contre, il apparaît clairement que la fonction de corrélation exponentielle permet d'obtenir une erreur de prédiction globale et locale la plus faible.



**Figure II-24 : Erreur de prédiction du rendement calculée sur 10 plans d'expériences pour le krigeage.**

Avec les RBF, le seul estimateur disponible est le critère LOO. Les fonctions radiales Matérn et gaussienne amènent à de fortes erreurs de prédiction (Figure II-25). Pour les autres fonctions radiales, les capacités de prédiction sont quasi équivalentes, sauf pour le cas linéaire où l'erreur est plus élevée. Les valeurs de RAAE et RMAE obtenues avec les quatre types de RBF les mieux adaptés sont plus faibles que celles obtenues avec le krigeage.



**Figure II-25 : Erreur de prédiction du rendement calculée sur 10 plans d'expériences pour les RBF.**

L'analyse des valeurs des hyperparamètres permet d'expliquer en partie la différence de précision entre les métamodèles. Cette observation est réalisée sur la métamodélisation de la température maximale, les résultats sont similaires pour les autres données physiques observables. Les valeurs des hyperparamètres obtenus par optimisation de la fonction de vraisemblance pour le krigeage sont visibles sur la Figure II-26. Pour les paramètres géométriques (1 à 8), quelle que soit la fonction de corrélation, l'estimation donne les mêmes valeurs. Il s'agit en fait de la borne inférieure du problème d'optimisation, qui dépend de la distance entre les points du plan d'expérience d'entraînement. Pour la tension (paramètre 9), les valeurs d'hyperparamètres sont plus élevées et varient selon la fonction de corrélation. Avec les RBF (Figure II-27, valeurs en log), les bornes inférieures sont également atteintes pour les paramètres 1 à 8 mais seulement dans les cas des fonctions Matérn et gaussiennes. Pour les autres fonctions radiales, l'optimisation ne converge pas vers les bornes du domaine. Par ailleurs, les cas spline en plaque mince et cubique donnent des valeurs d'hyperparamètres assez proches. Ces deux fonctions radiales sont aussi celles qui produisent les métamodèles les plus précis.

La régularité des fonctions Matérn et gaussienne n'est apparemment pas compatible avec le modèle à approximer. Leur utilisation amène à une mauvaise estimation des hyperparamètres. Les bornes du domaine d'optimisation des hyperparamètres peuvent alors apparaître comme trop restrictives. En les élargissant, les métamodèles obtenus n'ont pas donné de meilleurs résultats, ils étaient même parfois moins

bons. Ici les bornes ont donc bien un rôle de contrôle. Une comparaison des valeurs d'hyperparamètres entre krigeage et RBF n'est pas directement possible, car les données d'entraînement du krigeage sont centrées-réduites alors que celles des RBF sont simplement normalisées. Il y a donc une légère différence dans les distances entre points qui servent au calcul des hyperparamètres.



Température maximum

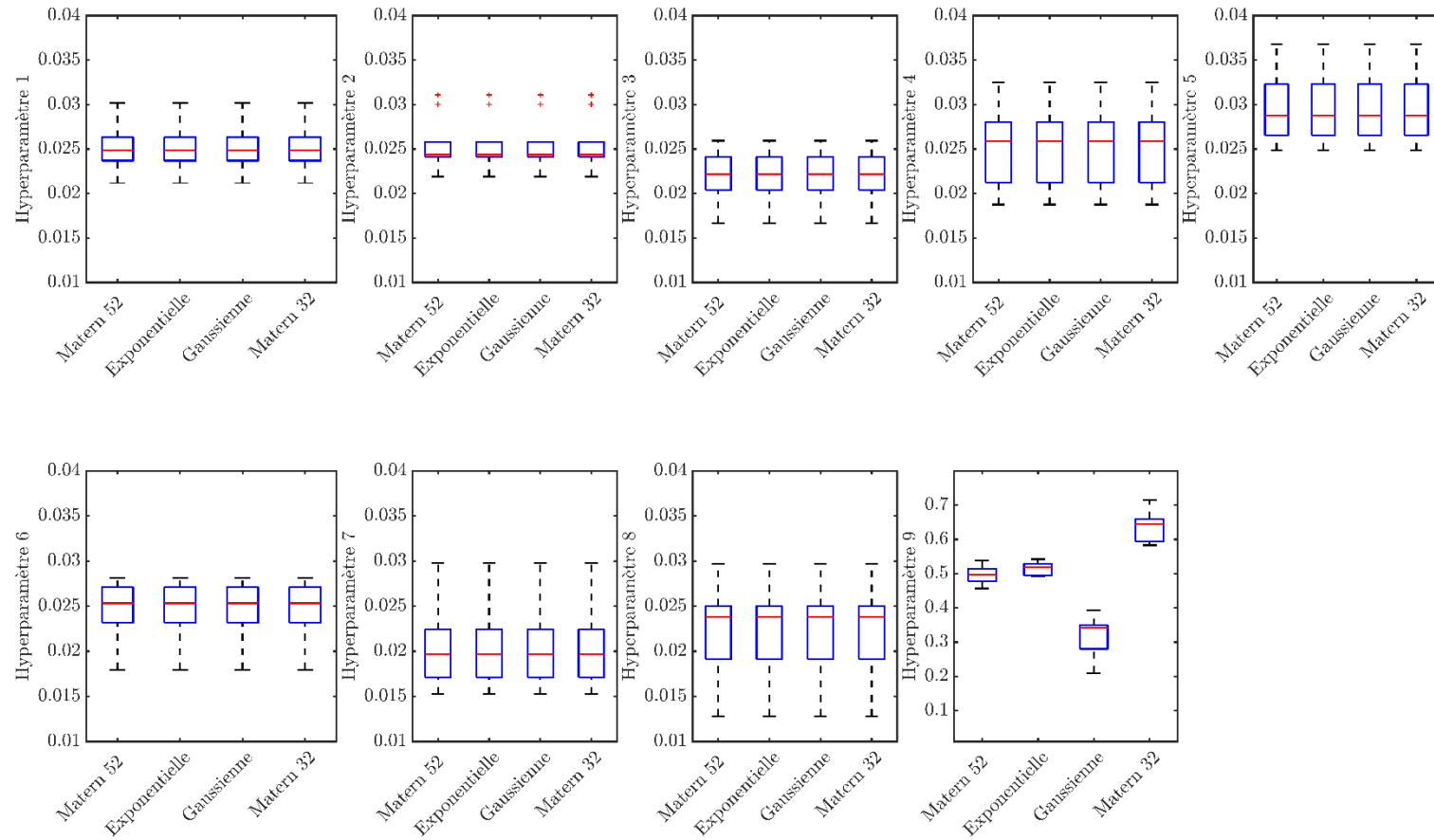


Figure II-26 : Valeurs des hyperparamètres du krigeage obtenu à partir de 10 plans d'expériences pour la température maximum.

Température maximum

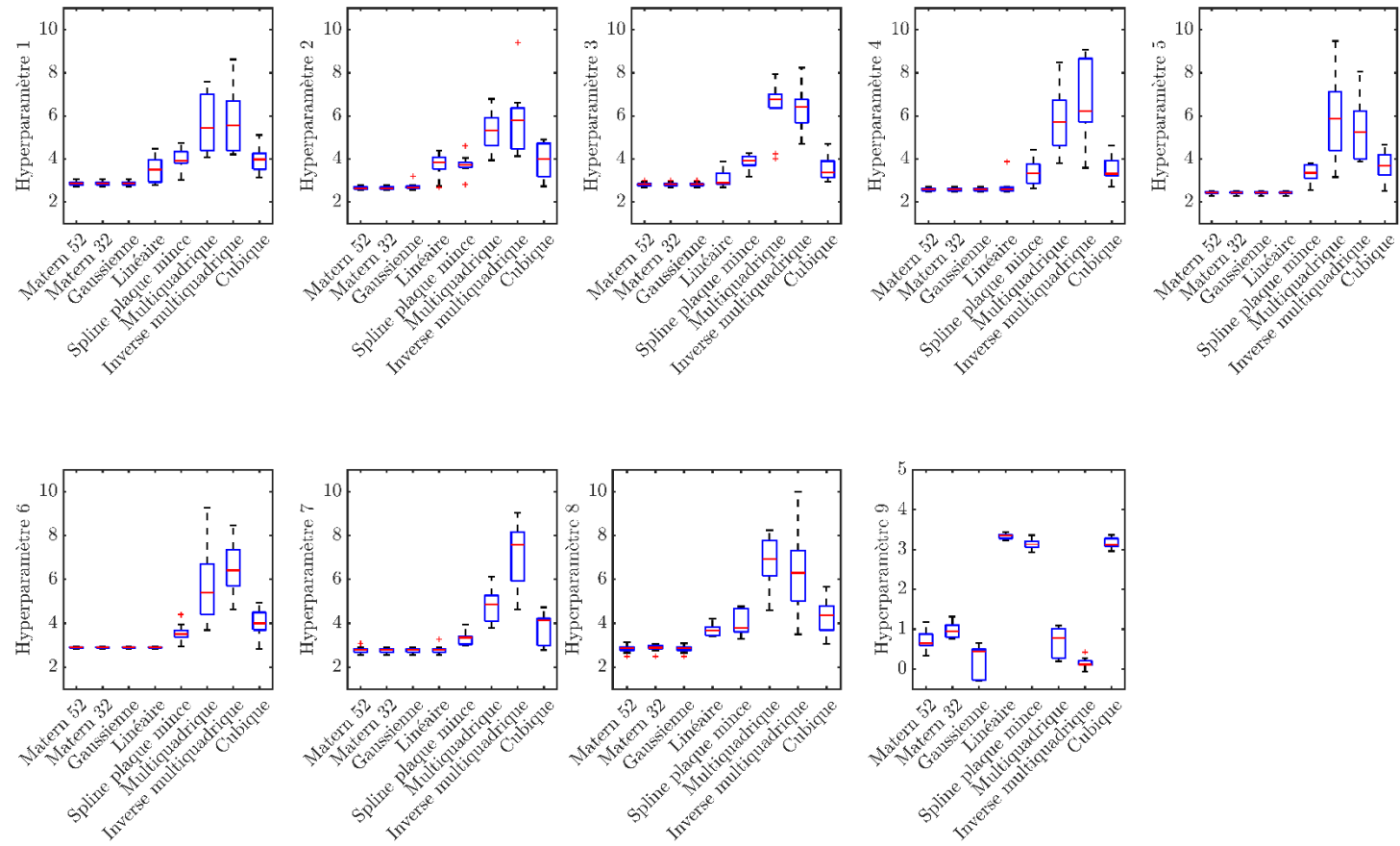


Figure II-27 : Valeurs (log) des hyperparamètres des RBF obtenues à partir de 10 plans d'expériences pour la température maximum.

Les résultats obtenus pour les autres grandeurs physiques sont similaires au cas du rendement et de la température maximale. En augmentant le nombre de points dans les plans d'expériences d'entraînement à 100, les tendances des résultats restent également les mêmes. Pour la suite du benchmark, la fonction de corrélation exponentielle et l'estimation par MLE sont choisies pour le krigeage. Pour les RBF, la fonction cubique est employée. Afin d'observer les différences entre les deux métamodèles, la Figure II-28 montre la comparaison entre les valeurs prédites et les valeurs réelles pour le krigeage et la RBF sur la température moyenne. Ici le krigeage présente des d'écart entre les valeurs prédites et les valeurs attendues tandis que les RBF reproduisent quasiment parfaitement la réponse du modèle numérique.

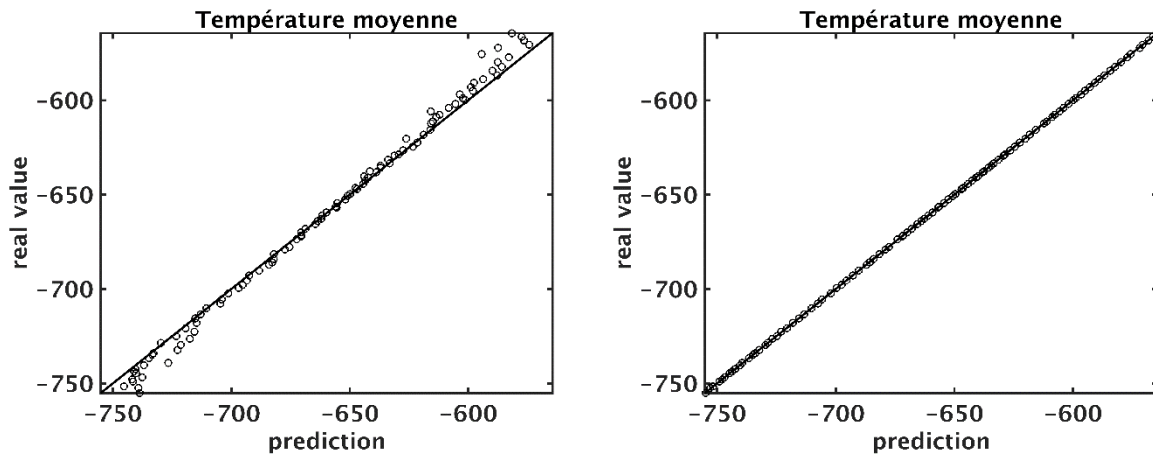


Figure II-28 : Graphique de comparaison entre les valeurs prédites et les valeurs réelles pour le krigeage (à gauche) et pour les RBF (à droite) sur la prédiction de l'opposée de la température moyenne.

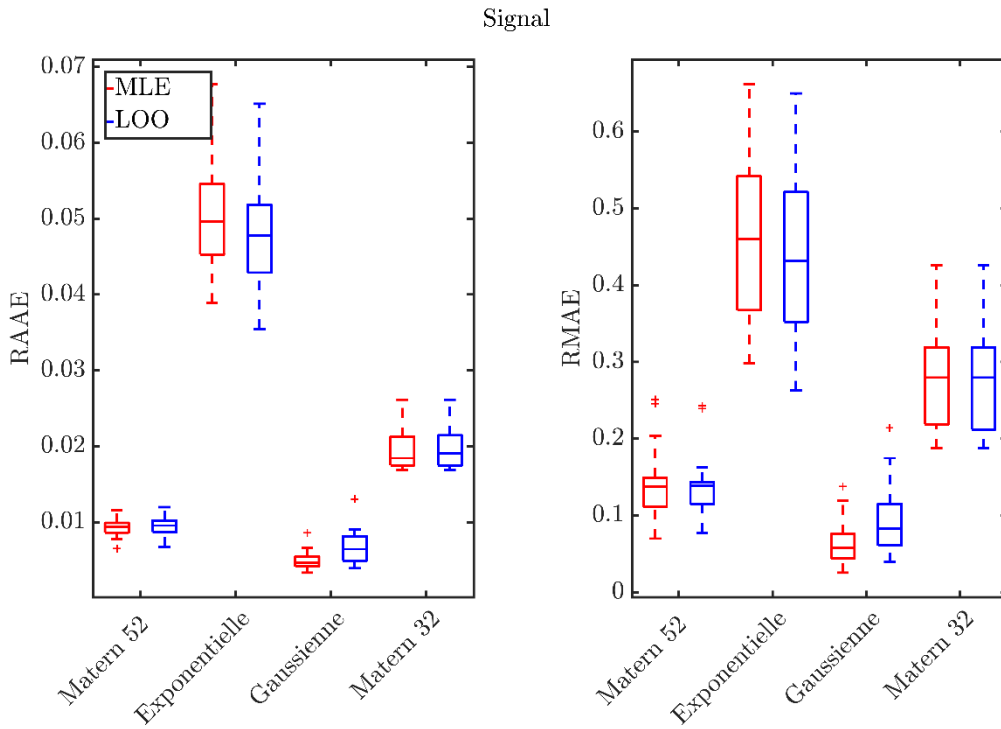
## 2. La cellule photoacoustique

Le cas d'étude de la cellule photoacoustique a été présenté section I.D. Son modèle comporte 3 paramètres géométriques d'entrée décrits à partir de la Figure I-14: le diamètre des capillaires qu'on note  $D_{cap}$  ainsi que les longueurs des chambres et des capillaires notées  $L_{cap}$  et  $L_{ch}$ . Les bornes de variation des paramètres pour cette étude sont disponibles dans le Tableau 5 (en mm). Pour rappel, deux modèles du comportement de la cellule sont disponibles. Le modèle FLNS est celui de plus haute fidélité et le modèle Kreuzer correspond à la version basse fidélité. Pour chaque modèle, deux sorties de la simulation sont étudiées : le signal de la cellule et sa fréquence de résonance. Les plans d'expériences qui sont utilisés par la suite sont de type OLHS.

	$D_{cap}$	$L_{cap}$	$L_{ch}$
Borne inférieure	0.15	10	5
Borne supérieure	0.5	20	20

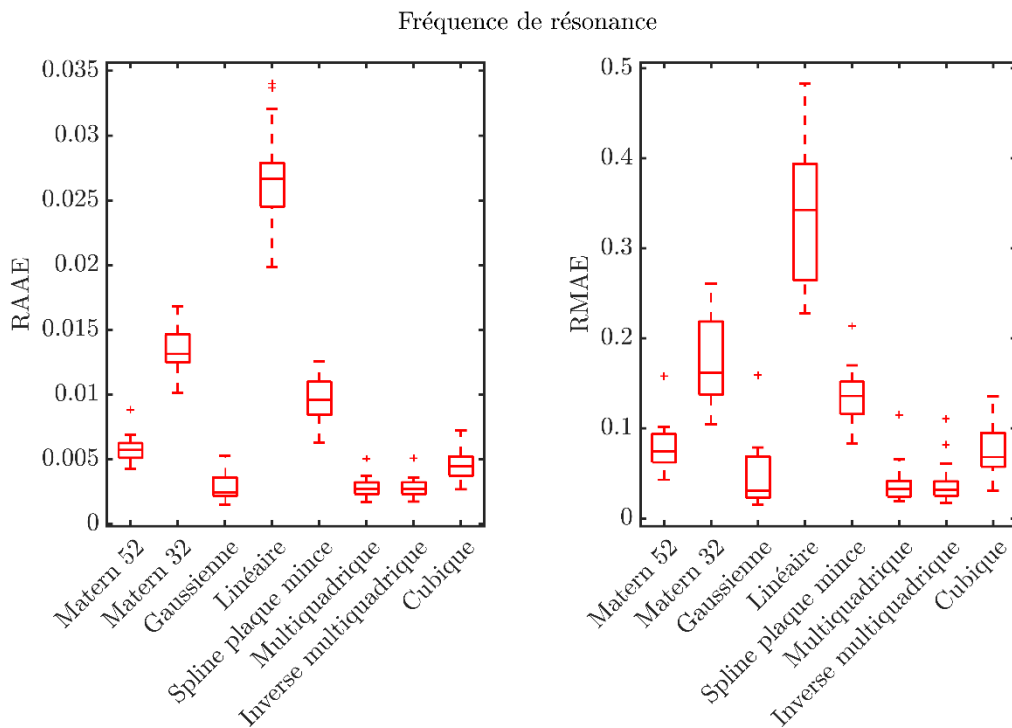
Tableau 5 : Bornes des paramètres du modèle de la cellule photoacoustique (en mm).

La comparaison des deux métamodèles du modèle Kreuzer est étudiée en premier. Pour l'entraînement, 20 plans d'expériences de 80 points sont utilisés. Les capacités de prédiction sont mesurées avec 300 points de test. Sur la Figure II-29, l'erreur de prédiction du krigeage du signal est représentée en fonction de la fonction de corrélation et de la méthode d'estimation des hyperparamètres. Les résultats sont similaires pour le cas de la fréquence de résonance. La fonction de corrélation gaussienne, avec estimation par maximum de vraisemblance, correspond à la combinaison la plus intéressante.



**Figure II-29 : Erreur de prédiction du signal calculée sur 20 plans d'expériences pour le krigeage.**

Pour les RBF, l'erreur de prédiction sur la fréquence de résonance en fonction des fonctions radiales utilisées est visible sur la Figure II-30. Les résultats sont similaires pour le signal. C'est aussi la corrélation gaussienne qui donne les meilleurs résultats. Les fonctions de type multiquadrique sont également intéressantes.



**Figure II-30 : Erreur de prédiction de la fréquence de résonance calculée sur 20 plans d'expériences pour les RBF.**

Les deux métamodèles sont très précis avec ce nombre de points d'entraînement. Afin d'observer des écarts notables, le diagramme comparant les valeurs prédites et les valeurs réelles pour le krigeage et la RBF

(Figure II-31) sur le signal est construit en utilisant seulement 20 points pour l'entraînement. La fonction de corrélation gaussienne est utilisée pour le krigeage et la multiquadrique pour les RBF. Sur cet exemple, le krigeage comporte des erreurs légèrement plus importantes que les RBF. Mais cette différence disparaît vite avec l'augmentation du nombre de points d'entraînement.

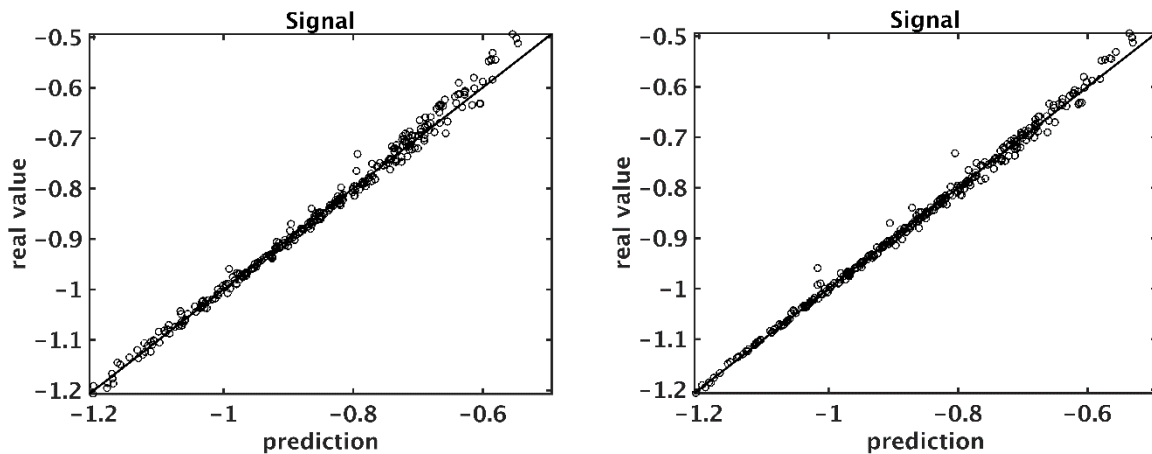


Figure II-31 : Graphique de comparaison entre les valeurs prédites et les valeurs réelles pour le krigeage (à gauche) et pour les RBF (à droite) sur la prédiction de l'opposé du signal.

La substitution du modèle FLNS, plus coûteux, est maintenant étudiée. 10 plans d'expériences de 20 points sont utilisés pour l'entraînement. Les capacités de prédiction sont mesurées avec 80 points de test. Sur la Figure II-32, l'erreur de prédiction du krigeage sur le signal est représentée selon la fonction corrélation et la méthode d'estimation des hyperparamètres. Les résultats pour la fréquence de résonance sont présentés sur la Figure II-33. Pour l'approximation du signal, les différences de résultats entre les fonctions de corrélation ne sont pas très importantes. Seule l'exponentielle donne en moyenne des erreurs plus grandes. Pour la fréquence de résonance, on retrouve les résultats obtenus avec le modèle Kreuzer : la fonction de corrélation gaussienne apporte le moins d'erreur de prédiction.

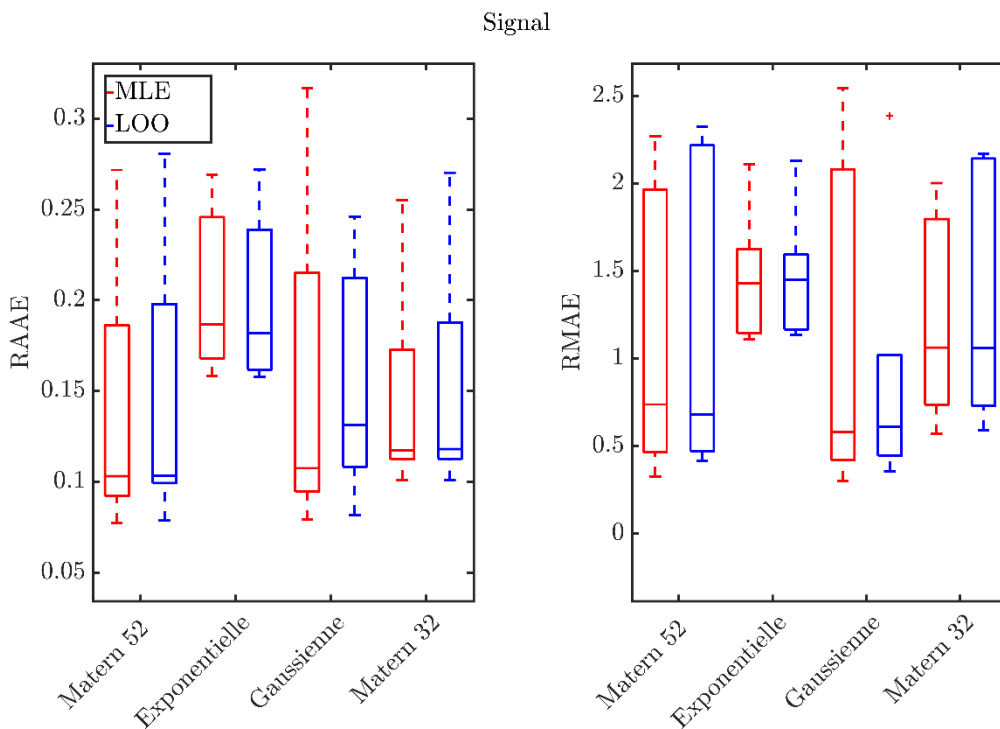
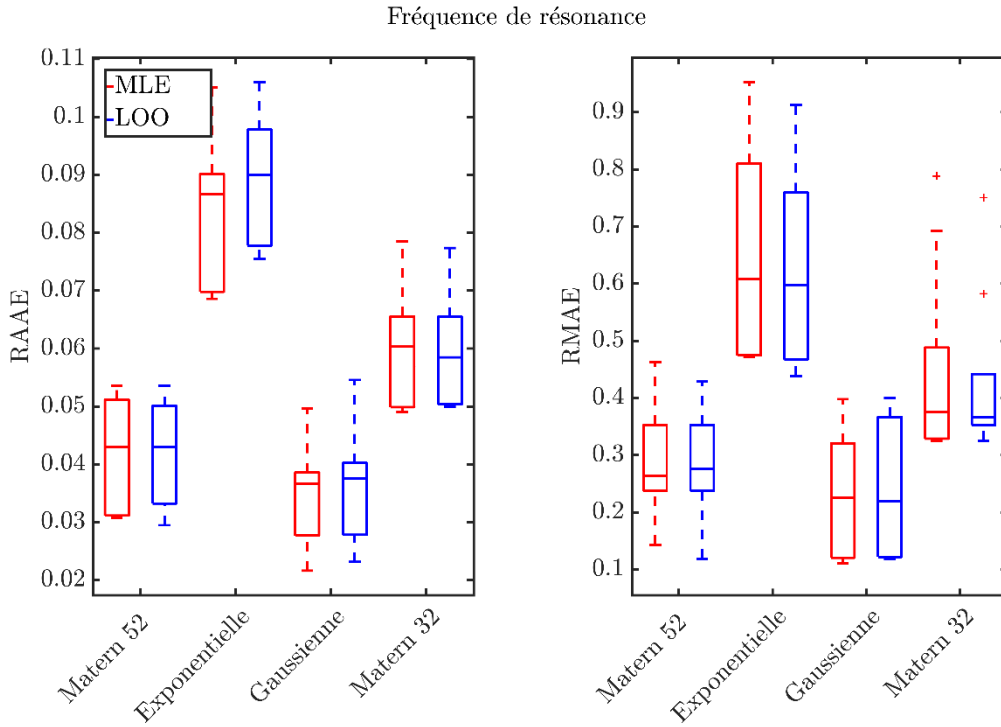
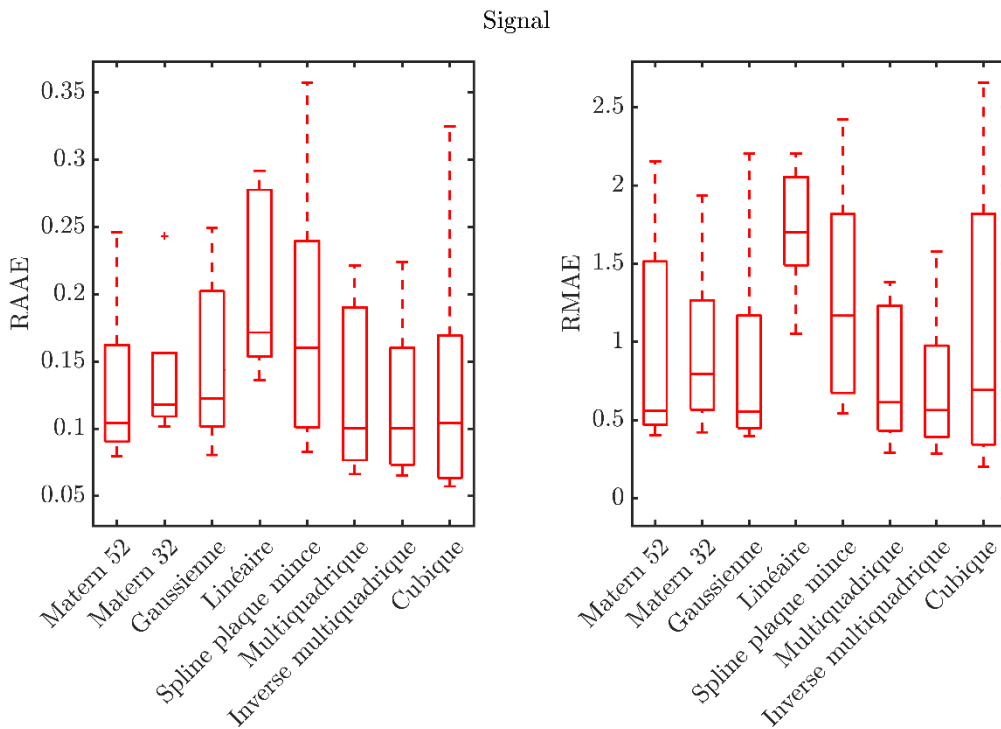


Figure II-32 : Erreur de prédiction du signal calculée sur 10 plans d'expériences pour le krigeage.



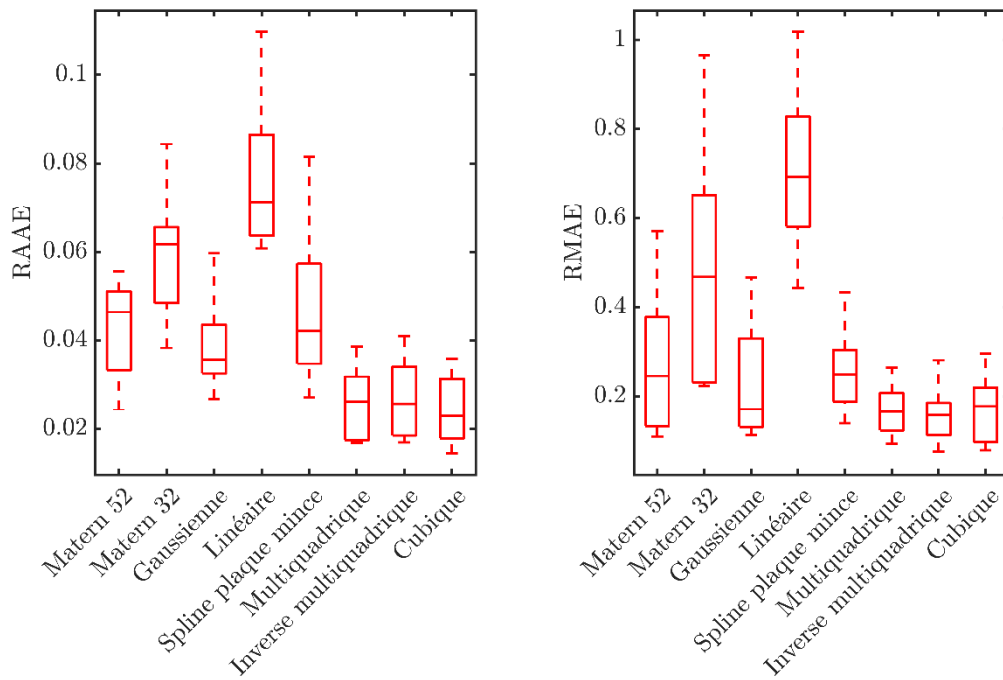
**Figure II-33 : Erreur de prédiction de la fréquence de résonance calculée sur 10 plans d'expériences pour le krigeage.**

Pour les RBF, l'erreur de prédiction selon la base de fonctions radiales choisie est visible pour le signal sur la Figure II-34 et pour la fréquence de résonance sur la Figure II-35. Comme pour le krigeage, les résultats sur le signal sont assez similaires. Pour les deux sorties, les fonctions multiquadrique et cubique donnent les erreurs de prédiction les plus faibles.



**Figure II-34 : Erreur de prédiction du signal calculée sur 10 plans d'expériences pour les RBF.**

Fréquence de résonance



**Figure II-35 : Erreur de prédiction de la fréquence de résonance calculée sur 10 plans d'expériences pour les RBF.**

Pour vérifier les capacités de prédiction du krigeage et des RBF sur le modèle FLNS, le diagramme des valeurs prédites en fonction des valeurs réelles est tracé pour chaque sortie sur la Figure II-36. La fonction de corrélation Matérn 5/2 est utilisée pour le krigeage et la base radiale multiquadrique pour les RBF. Pour le signal les erreurs de prédiction sont réparties de la même manière pour les deux métamodèles. Sur la fréquence de résonance, les RBF sont légèrement plus précises.

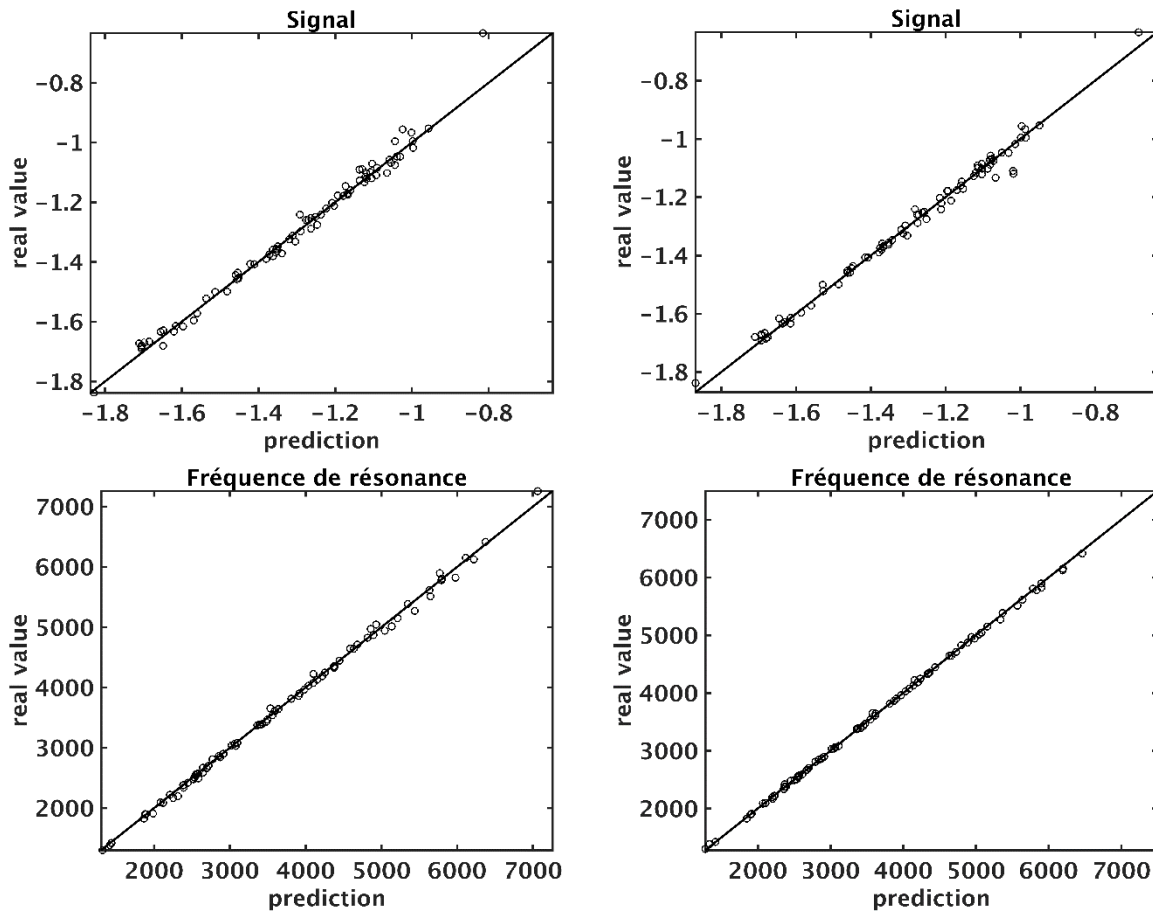


Figure II-36 : Graphique de comparaison entre les valeurs prédites et les valeurs réelles pour le krigeage (colonne de gauche) et pour les RBF (colonne de droite) sur la prédiction de l'opposée du signal et la fréquence de résonance.

Pour synthétiser cette étude, le Tableau 6 récapitule les meilleurs résultats pour chaque sortie sur les deux modèles. La moyenne et l'écart de type de la RAAE sont choisies pour mesurer la fidélité des métamodèles. Les fonctions de corrélation et bases radiales permettant d'obtenir ces résultats sont précisées à chaque cas. Les capacités des deux métamodèles sont très proches et il ne convient pas dans ce cas d'affirmer la supériorité de l'un par rapport à l'autre. On remarque par contre que le krigeage apporte les meilleurs résultats sur l'approximation du modèle Kreuzer et les RBF pour celle du modèles FLNS.



	Sorties	Krigage		RBF	
Kreuzer	Signal	0.0050±0.0012	Gaussienne	0.0055±0.0018	Gaussienne
	Fréquence	0.0023±0.0006	Gaussienne	0.0029±0.0012	Gaussienne
FLNS	Signal	0.1425±0.0754	Matérn 52	0.1222±0.0603	Inverse multi
	Fréquence	0.0351±0.0086	Gaussienne	0.0241±0.0077	Cubique

**Tableau 6 : Récapitulatif des résultats sur la substitution des modèles de la cellule photoacoustique. Pour chaque cas, la moyenne et l'écart type de la RAAE sont donnés. Ces valeurs sont issues de la combinaison fonction de corrélation et base radiale amenant à la plus faible erreur de prédiction.**

Puisque deux modèles du même phénomène physique sont disponibles, une approche multifidélité peut être envisagée ici. Comme précisé dans la section II.C.1, il existe une version multifidélité du krigage nommée co-krigage. Par contre rien de tel n'existe pour les RBF. Au vu des résultats présentés dans cette étude, il paraît intéressant de proposer une version multifidélité des RBF comme alternative au co-krigage. Ce travail est présenté au chapitre IV, où les performances de la co-RBF, version multifidélité des RBF développée durant la thèse, sont comparées au co-krigage.

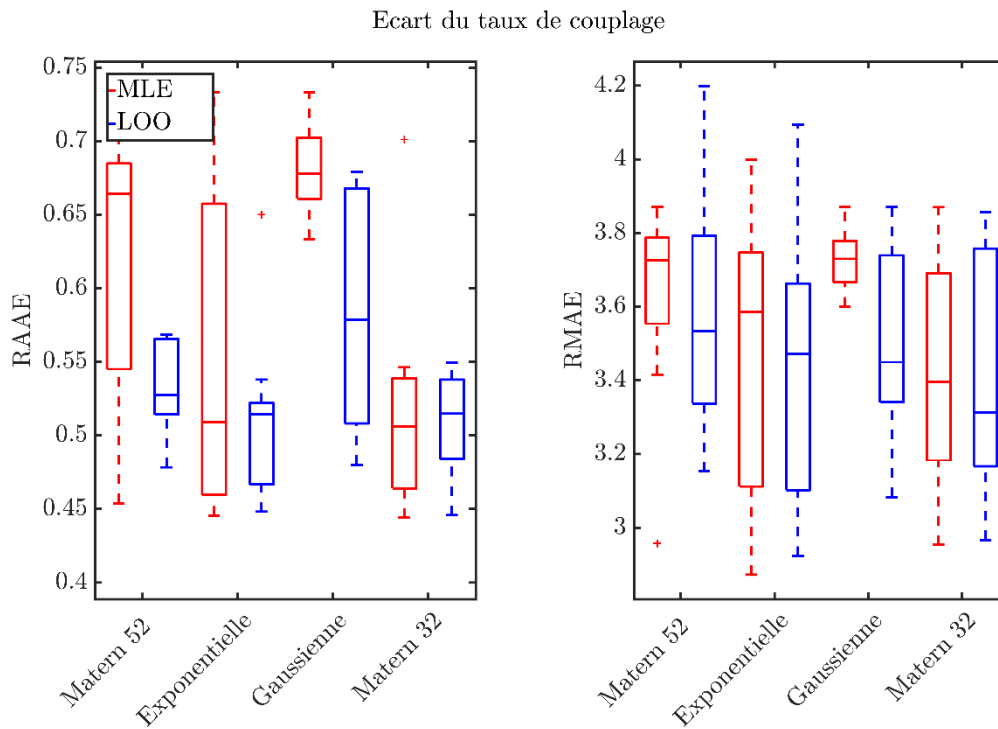
### 3. Le coupleur adiabatique

Le cas d'étude du coupleur adiabatique a été présenté en section I.E. Son modèle comporte 4 paramètres géométriques en entrée et un paramètre environnemental correspondant à la longueur d'onde. Suite à une étude non fournie ici, la longueur de la zone de couplage  $L$  est fixée ici à 160  $\mu\text{m}$ . Ces paramètres sont décrits à la figure I-18. Les bornes de variations pour cette étude sont disponibles dans le Tableau 7. La sortie de la simulation étudiée correspond à l'écart au taux de couplage souhaité, qui est de 50 % ici. Les plans d'expériences qui sont utilisés par la suite sont de type OLHS.

	$W_1$	$W_2$	$g_1$	$g_2$	$\lambda_c$
Borne inférieure	0.35	0.35	1	0.5	1.26
Borne supérieure	0.55	0.55	1.5	0.7	1.36

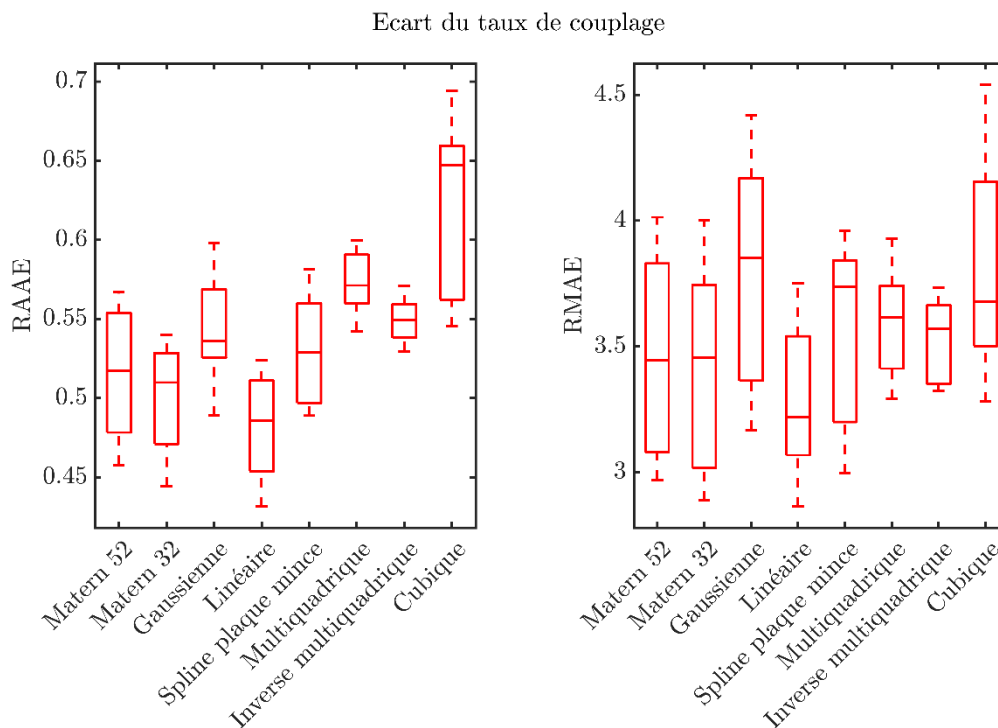
**Tableau 7 : Bornes des paramètres du modèle du coupleur adiabatique (en  $\mu\text{m}$ ).**

Neuf plans d'expériences de 200 points sont utilisés pour l'entraînement. Les capacités de prédiction sont mesurées sur 200 points de test. Sur la Figure II-37, l'erreur de prédiction du krigage est représentée en fonction de la fonction de corrélation et de la méthode d'estimation des hyperparamètres. Par rapport aux cas précédents, nous constatons un écart plus important entre les résultats selon la méthode d'estimation des hyperparamètres. La pseudovraisemblance (LOO) produit des résultats plus stables en fonction des plans d'expériences utilisés pour l'entraînement (moins de variances sur les valeurs de RAAE). Les fonctions de corrélation exponentielle ou Matérn 3/2 permettent d'obtenir une erreur de prédiction plus faible.



**Figure II-37 : Erreur de prédiction de l'écart au taux de couplage calculée sur 9 plans d'expériences pour le krigeage.**

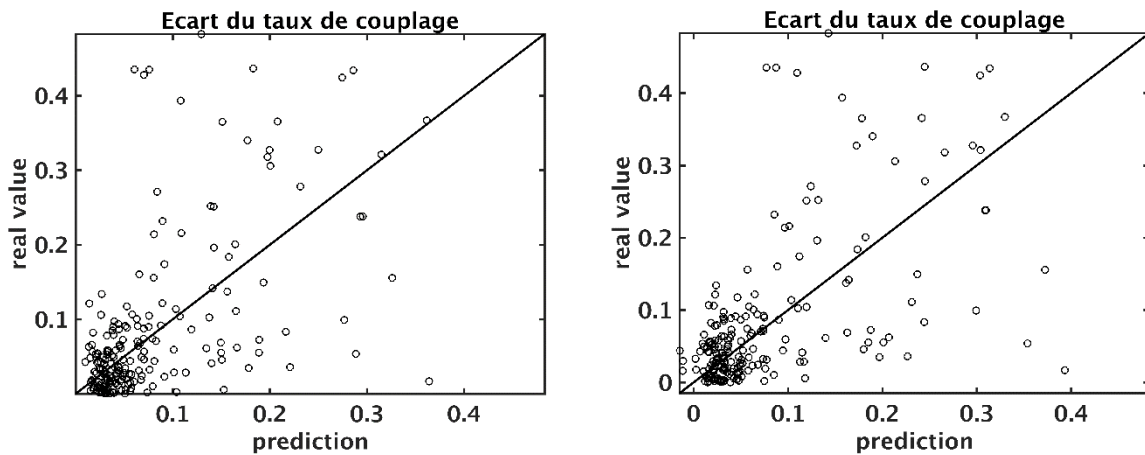
Les résultats obtenus avec les RBF sont visibles sur la Figure II-38. La fonction radiale linéaire apporte une RAAE la plus faible. Ce résultat est intrigant car la relation qui lie les variations géométriques du coupleur et la puissance en sortie des guides, analysée en quelques points, n'est pas linéaire.



**Figure II-38 : Erreur de prédiction de l'écart au taux de couplage calculée sur 9 plans d'expériences pour les RBF.**

Avec les graphiques comparant la valeur réelle des points tests et celle prédite (Figure II-39), on remarque que les deux métamodèles prédisent très mal la réponse du modèle numérique. Ces graphiques ont été

obtenus avec une fonction de corrélation exponentielle pour le krigeage et linéaire pour les RBF. Il est clair que l'information contenue dans les plans d'entraînement n'est pas suffisante pour bien appréhender la complexité de la simulation.



**Figure II-39 : Graphique de comparaison entre les valeurs prédites et les valeurs réelles pour le krigeage (à gauche) et pour les RBF (à droite) sur la prédiction de l'écart au taux de couplage.**

Pour mieux comprendre l'origine de ces fortes erreurs de prédiction, la surface de réponse obtenue avec le krigeage est affichée pour certains jeux de paramètres sur la Figure II-40. Les vues en coupe permettent de voir si le comportement de la sortie du modèle numérique est fortement non-linéaire. En faisant varier les paramètres  $g_1$ ,  $g_2$  et  $\lambda_c$ , la surface de réponse obtenue est assez lisse. Par contre, la variation des paramètres  $W_1$  et  $W_2$  amène à une évolution de  $T_c$  très erratique. Pour limiter l'impact de ces paramètres sur la réponse du métamodèle, l'intervalle de variation a été réduit le plus possible.

Afin d'obtenir un métamodèle plus précis, il faut donc soit ajouter plus de points au plan d'entraînement, soit enlever des paramètres. Augmenter le nombre de points peut permettre de mieux approcher les phénomènes locaux de la réponse du modèle numérique. Enlever des paramètres de conception réduit le nombre d'hyperparamètre à estimer lors de l'entraînement du métamodèle. Actuellement le nombre de points utilisé pour l'entraînement est déjà important. Il a donc été choisi de d'abord fixer le paramètre de longueur d'onde à  $\lambda_c = 1.31 \mu\text{m}$  pour voir l'impact sur la précision du modèle de substitution. Malheureusement, la valeur de RAAE pour le problème réduit est du même ordre de grandeur que celle obtenue en faisant varier  $\lambda_c$ . Par contre la précision générale de la prédiction du métamodèle n'est pas l'élément recherché. Le but final est de trouver le minimum de l'écart au taux de couplage désiré, si possible pour une gamme importante de longueurs d'ondes. Si rajouter des points aux données d'entraînement est réalisable, il n'est pas nécessaire de les évaluer sur tout l'espace de variation des paramètres. Il existe des méthodes permettant de sélectionner des points à ajouter au plan d'entraînement seulement dans les zones d'intérêt. Elles portent le nom de stratégies adaptatives et sont étudiées dans le prochain chapitre. La zone ciblée pour cette étude correspond au minimum global de la surface de réponse.

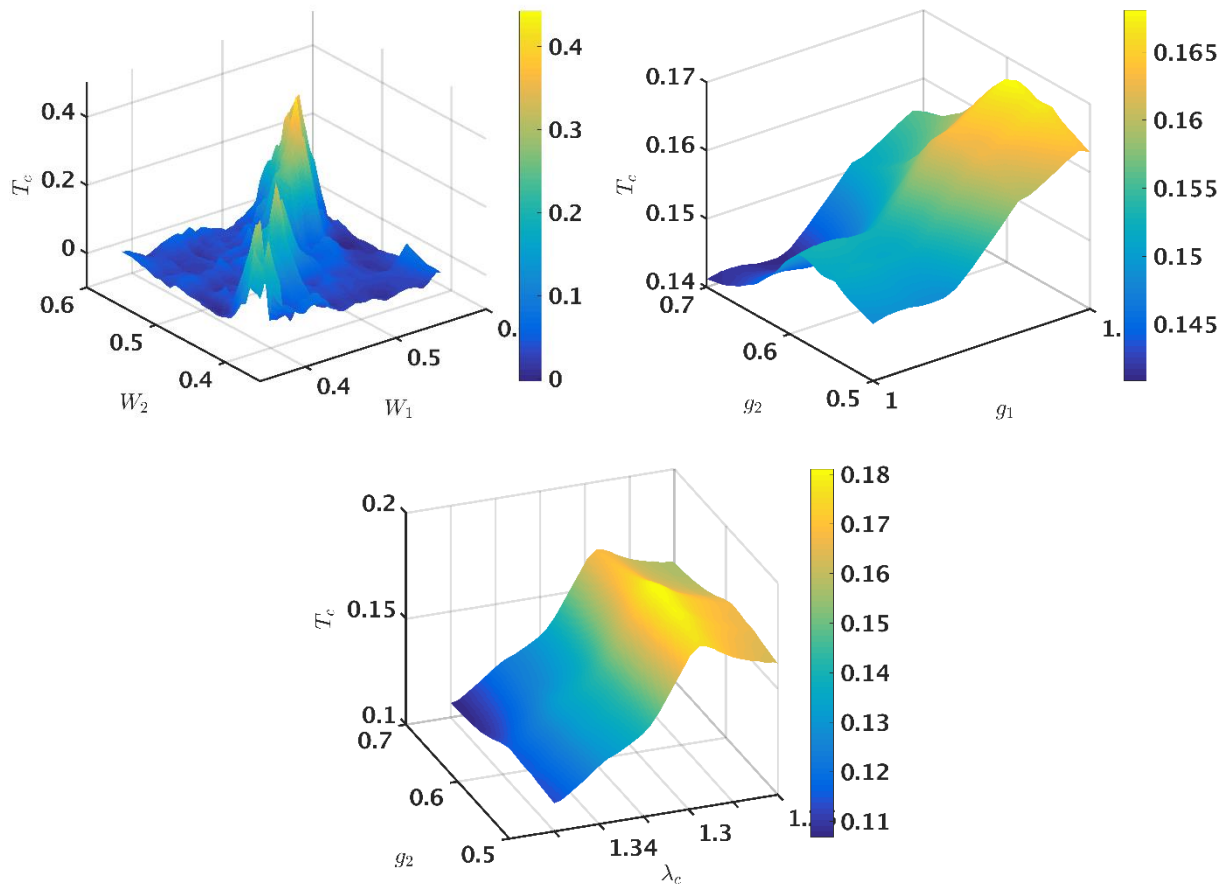


Figure II-40 : Vue en coupe de la surface de réponse prédite par le krigeage. Pour chaque coupe, les valeurs centrales des intervalles de variation sont utilisées pour les paramètres non affichés.

## E. Conclusion

La métamodélisation permet de remplacer un simulateur numérique coûteux par un modèle très rapide à évaluer. Ce type de méthode consiste en fait en une interpolation (ou régression) à partir de données pré-évaluées sur le simulateur, sans avoir besoin de définir *a priori* la forme globale de l'interpolant. Le krigeage et les RBF sont les deux métamodèles utilisés dans le cadre de cette étude. Leurs performances ont été comparées sur les cas d'applications définis au chapitre I. Les résultats obtenus montrent que les deux métamodèles ont des capacités de prédiction relativement similaire. Selon les exemples, les fonctions de corrélations ou radiales préconisées ne sont pas toujours les mêmes. L'étude comparative a permis de mettre en évidence qu'il n'existe pas de type de métamodèle toujours supérieur quel que soit le problème.

Une fois le métamodèle construit, il peut être exploité comme outil de conception par simulation. Sur les cas d'études proposés, un des principaux problèmes à résoudre consiste à optimiser la réponse du modèle numérique. Ceci correspond à rechercher des zones très particulières de la surface de réponse, par exemple localiser la plus petite ou la plus grande valeur. Si le métamodèle n'est pas très précis dans ces zones d'intérêt, il faut alors rajouter de l'information pour améliorer la prédiction. Des méthodes appelées stratégies adaptatives permettent de raffiner localement la précision du métamodèle en fonction du type de problème posé (optimisation sous contrainte, optimisation robuste, optimisation multi-objectif, etc...). Ce type de stratégies est étudié dans le chapitre III.

---

### III. Stratégies adaptatives pour l'optimisation avec métamodèle

---

Dans le chapitre précédent, deux méthodes ont été présentées permettant d'émuler la relation statistique entre les entrées et les sorties d'un modèle coûteux  $\mathcal{M}$  à partir d'un jeu de données  $\mathcal{D}$ . Cette approche permet de prédire les valeurs renvoyées par  $\mathcal{M}$  en de nouveaux points pour un coût quasi nul. Une fois que le métamodèle est disponible, il est possible de l'exploiter pour répondre à des problématiques de conception. Selon l'application visée (par exemple l'optimisation ou la fiabilité), la précision du métamodèle doit être améliorée sur tout ou une partie de l'espace d'entrée. Le plan d'expérience initial est alors complété avec de nouveaux points pour améliorer la capacité de prédiction du métamodèle sur des zones d'intérêt. Une stratégie naïve consisterait à résoudre par exemple le problème d'optimisation directement sur le métamodèle. Ce type de méthode n'est pas productive car elle ne tient pas compte de la précision du métamodèle. La valeur de l'optimum obtenu sur le métamodèle peut être très éloigné de celle fournie par la simulation. Pour mieux choisir le lieu du prochain point à évaluer, un critère  $\mathcal{C}$  correspondant au type d'étude à mener est à optimiser. Par convention, l'optimisation correspond à la minimisation d'une fonction, donc si le critère est à maximiser, il faut prendre son opposé. Un optimum local correspond au point  $\mathbf{x}^*$ , tel que pour tout point  $\mathbf{x}$  appartenant au voisinage de  $\mathbf{x}^*$ ,  $\mathcal{C}(\mathbf{x}^*) \leq \mathcal{C}(\mathbf{x})$ . On parle d'optimum global lorsque  $\forall \mathbf{x} \in \mathbb{X}, \mathcal{C}(\mathbf{x}^*) \leq \mathcal{C}(\mathbf{x})$ . En pratique, il n'est pas aisé de garantir l'obtention d'un optimum global. Le critère  $\mathcal{C}$  est calculé à partir de la prédiction fournie par le métamodèle et son optimum global est le nouveau point à évaluer. Ce processus, défini sous le nom de stratégie adaptative, est itéré jusqu'à ce que la précision du métamodèle requise par l'utilisateur soit atteinte. Il est résumé sur la Figure III-1. Pour éviter de confondre avec l'optimisation de la réponse du modèle  $\mathcal{M}$ , on utilise par la suite le terme de sous-problème d'optimisation pour définir la recherche de l'optimum du critère  $\mathcal{C}$ .

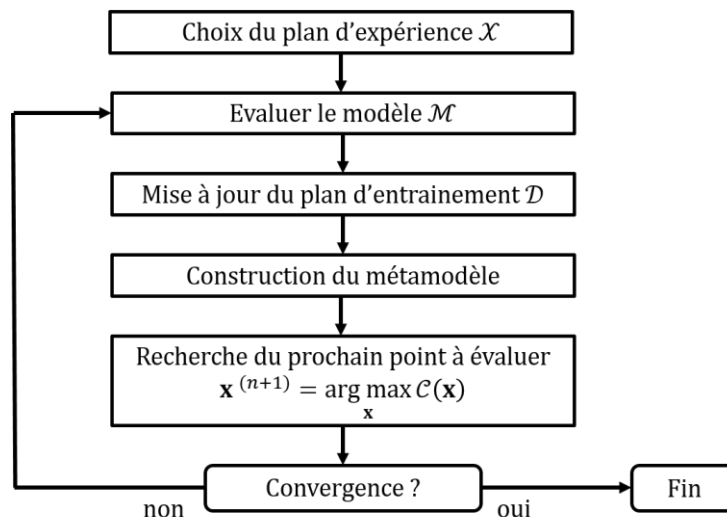


Figure III-1 : Processus de la stratégie adaptative d'amélioration du métamodèle dans le cadre général.

Dans ce chapitre quatre catégories de critères de stratégies adaptatives sont abordés. La première concerne la réduction générale de l'erreur de prédiction, en section III.A. Ce type de critère permet d'obtenir un métamodèle précis sur l'ensemble du domaine. La deuxième catégorie, en section III.B, concerne l'optimisation globale d'un objectif, qui est complétée pour le cas où des contraintes doivent être satisfaites en section III.B.5. Lorsque plusieurs objectifs sont à optimiser en même temps, d'autres critères sont

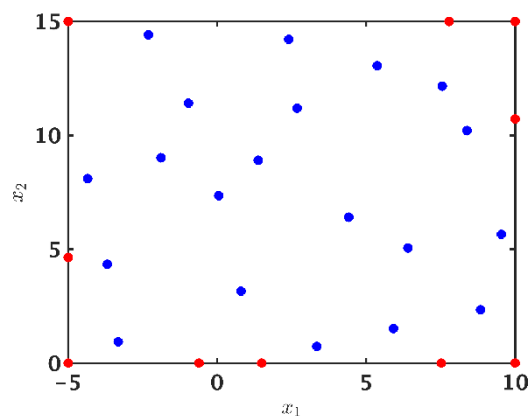
utilisables et traités en section III.C. Pour terminer, les critères de l'optimisation robuste, présentés en section III.D, permettent de tenir compte des incertitudes liées aux tolérances de fabrication sur les paramètres du modèle numérique.

## A. Réduction de l'erreur de prédiction

Dans le cadre de la conception en microélectronique, la surface de réponse du métamodèle peut être intégrée dans des Process Design Kit (PDK) en tant que boîte noire. Les PDK sont des bibliothèques contenant plusieurs composants à assembler dans le but de construire un système fonctionnel. Les composants peuvent avoir des spécifications fixées par le fabricant, tel que des réseaux de couplage pour des longueurs d'ondes spécifiques. Dans certains cas, l'utilisateur peut choisir les propriétés du composant. Par exemple, il est possible de choisir le ratio de la division de puissance en sortie pour un diviseur de puissance optique à deux sorties. Si le concepteur souhaite simuler la réponse finale de son système, la réponse de chaque composant doit être directement disponible sans repasser par des étapes de simulations coûteuses. Avec un métamodèle, le problème inverse permettant de trouver quels sont les paramètres du composant pour obtenir les spécifications souhaitées peut être résolu très rapidement. Le métamodèle peut également fournir la réponse du composant en fonction de paramètres environnementaux tel que la longueur d'onde. Il faut toutefois vérifier que la précision de la prédiction globale du métamodèle est suffisante pour son utilisation. Deux critères pour une stratégie adaptative sont présentés dans la suite permettant la réduction de l'erreur de prédiction.

### 1. MSE (Mean Squared Error)

Pour améliorer la prédiction du métamodèle, il faut ajouter des informations manquantes dans le plan d'expérience utilisé initialement. Le premier critère utilise directement la mesure d'erreur de prédiction du métamodèle. Pour le krigeage et les RBF, ce critère est proche d'une maximisation de la distance minimale entre les points d'entraînement et le point à ajouter. Au lieu d'utiliser directement la distance euclidienne, les erreurs de prédiction sont basées sur une distance tenant compte de la covariance du modèle. Selon le métamodèle choisi,  $MSE = \hat{\sigma}_y^2$  pour le krigeage et  $MSE = P_\phi$  (voir équations II.32 et II.54) pour les RBF. Un exemple de stratégie adaptative par MSE est visible sur la Figure III-2 pour le krigeage. Les points ajoutés se situent tous sur le bord du domaine. Ce phénomène s'explique par la mauvaise capacité d'extrapolation des métamodèles présentés. L'erreur de prédiction est toujours importante sur les limites du domaine lorsqu'il n'y a pas de points. De plus, les chances d'ajouter des points au centre du domaine diminuent fortement en plus grande dimension.



**Figure III-2 : Stratégie adaptative par MSE ajoutant 10 points (en rouge) à partir de 20 points d'entraînement (en bleu). Cette stratégie a été appliquée avec la fonction analytique Branin (voir III.B.4) pour le krigeage.**

Une des propriétés intéressantes du critère MSE est qu'il ne dépend pas des valeurs de sortie, mais seulement de la distance entre les points du plan d'expérience d'entraînement. Ainsi, lorsqu'un premier point  $\mathbf{x}^{(n+1)}$  est obtenu par maximisation de la MSE, il est possible de recommencer une optimisation de la

MSE en ayant ajouté ce nouveau point sans l'avoir évalué. Si plusieurs points peuvent être évalués en même temps par le biais de calculs distribués, il est intéressant de sélectionner un ensemble de plusieurs points par optimisation itérative de la MSE avec ajout virtuel de l'optimum dans le plan d'expérience.

Lorsque ce critère est utilisé sur des fonctions présentant des zones de fortes variations, un certain nombre d'itérations est nécessaire avant d'aller échantillonner dans ces zones. Ceci vient du fait que le critère ne dépend pas des réponses évaluées sur le modèle, mais simplement de la configuration du plan d'expérience initial. Pour pallier ce problème, Marmin et al. [64] proposent d'utiliser un critère se basant sur le gradient dans le cadre des processus gaussiens.

## 2. IMSE (Integrated Mean Squared Error)

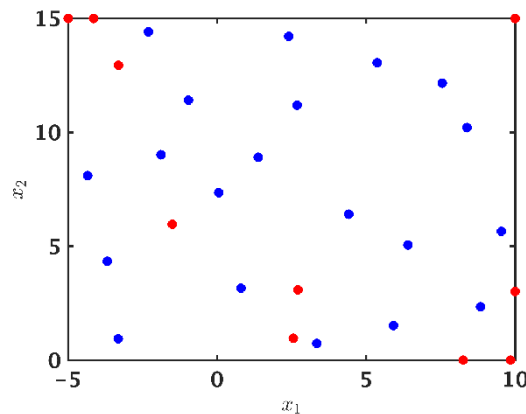
Pour mieux tenir compte de l'apport d'un nouveau point à la réduction de variance, Sacks et al. [65] définissent le critère IMSE (« integrated mean-squared error »). Il consiste à choisir le point qui, une fois ajouté virtuellement au plan d'entraînement (comme décrit en fin de section précédente), réduit le plus la variance sur tout le domaine. Pour mesurer cette réduction de variance, l'intégrale sur tout le domaine de la MSE est utilisée. Le critère se définit donc comme :

$$\text{IMSE}(\mathcal{X}) = \int_{\mathbf{x}} \text{MSE} \, d\mathbf{x} \quad [\text{III.1}]$$

Et le point  $\mathbf{x}^{(n+1)}$  est celui qui minimise cette quantité une fois qu'il a été ajouté virtuellement au plan d'expérience :

$$\mathbf{x}^{(n+1)} = \arg \min_{\mathbf{x}} \text{IMSE}(\{\mathcal{X}, \mathbf{x}\}) \quad [\text{III.2}]$$

En reprenant le même exemple que pour la MSE, on remarque sur la Figure III-3 que les points obtenus par stratégie IMSE ne sont pas tous placés sur les bords du domaine.



**Figure III-3 : Stratégie adaptative par IMSE ajoutant 10 points (en rouge) à partir de 20 points d'entraînement (en bleu). Cette stratégie a été appliquée avec la fonction analytique Branin (voir III.B.4) pour le krigeage.**

Le calcul du critère IMSE est plus coûteux, du fait de l'intégration, que le critère MSE. En pratique, ces critères sont fortement non-linéaires et comportent de nombreux minimums locaux. L'algorithme CMA-ES (voir section III.B.1.b) est utilisé ici pour optimiser ces critères mais l'obtention de l'optimum global n'est pas garantie au vu de la difficulté du problème. Dans le cas où l'optimum de la sortie physique du modèle numérique doit être trouvé, la réduction de la variance de prédiction n'est pas le critère le plus judicieux à utiliser. Des critères permettant d'obtenir le minimum global à partir de la surface de réponse sont présentés dans la section suivante.

## B. Optimisation globale à un objectif

La notion d'optimisation apparaît lorsqu'on souhaite améliorer la conception d'un composant vis-à-vis

d'une certaine mesure de performance. Dans cette section, une seule mesure de performance est prise en compte, donc un seul objectif à optimiser. Lorsqu'il n'y a pas de contraintes de conception, le problème d'optimisation globale suivant est à résoudre :

$$\min_{\mathbf{x} \in \mathbb{X}} \mathcal{M}(\mathbf{x}) \quad [\text{III-3}]$$

Une approche naïve de l'optimisation avec métamodèle consisterait à chercher l'optimum directement sur la surface de réponse. Si l'erreur de prédiction est élevée, l'optimum obtenu ne serait pas le bon. Il est alors possible de passer par une première étape de réduction de la variance de prédiction, par exemple avec un critère MSE, puis de trouver l'optimum à partir du métamodèle plus précis. Cette approche n'est pas efficace car il est possible de raffiner la précision du métamodèle directement dans le bassin du minimum global. Ces critères sont présentés en section III.B.2 pour le krigeage et en section III.B.3 pour les RBF. Nous commençons par décrire en section III.B.1 deux exemples d'algorithmes permettant de les optimiser.

## 1. Algorithmes d'optimisation du sous-problème

Les algorithmes présentés ici peuvent résoudre des problèmes d'optimisations à un objectif. Ils sont donc utilisables sur certains critères décrits dans ce chapitre. Le premier est une méthode locale et le deuxième est un algorithme évolutionnaire pour obtention du minimum global.

### a. Méthodes de gradient

Le contexte des méthodes de gradient est la minimisation d'une fonction  $f: \mathbb{R}^d \rightarrow \mathbb{R}$  différentiable. A partir d'un point quelconque  $\mathbf{x}^{(0)} \in \mathbb{R}^d$ , le but de ces méthodes est de faire converger la suite définie par l'équation [III.4] vers un point stationnaire de  $f$  tel que  $f(\mathbf{x}^{(k+1)}) < f(\mathbf{x}^{(k)})$  à l'itération  $k$ :

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \mathbf{g}_k \quad [\text{III.4}]$$

Avec  $\mathbf{g}_k$  une quantité à déterminer à chaque itération. Les méthodes de gradient utilisent la direction de plus forte pente  $\mathbf{d}_k$  pour déterminer  $\mathbf{g}_k$ . Celle-ci se définit par  $\mathbf{d}_k = -\nabla f(\mathbf{x}^{(k)})$ . Si la fonction  $f$  n'est pas connue explicitement, le calcul du gradient peut s'effectuer par différences finies : la différence calculée vers l'avant est définie par l'équation [III.5], avec  $h$  le pas et  $\mathbf{e}_i$  le vecteur unitaire sur la direction  $\mathbf{x}_i$  :

$$\frac{\partial f}{\partial x_i}(\mathbf{x}) = \frac{f(\mathbf{x} + h\mathbf{e}_i) - f(\mathbf{x})}{h} \quad [\text{III.5}]$$

Une fois la direction de plus forte pente obtenue, la recherche linéaire d'un pas optimum doit être effectuée afin de s'assurer de la diminution de la valeur de la fonction objectif à l'itération suivante. Pour cela il faut trouver le minimum  $\alpha^*$  de la fonction  $g$  définie à l'équation [III.6] par le biais d'une méthode telle que celle de Brent :

$$g(\alpha) = f(\mathbf{x}^{(k)} - \alpha \nabla f(\mathbf{x}^{(k)})) \quad [\text{III.6}]$$

Le déplacement à l'itération  $k$  pour un pas optimal est donc :

$$\mathbf{g}_k = -\alpha^* \nabla f(\mathbf{x}^{(k)}) \quad [\text{III.7}]$$

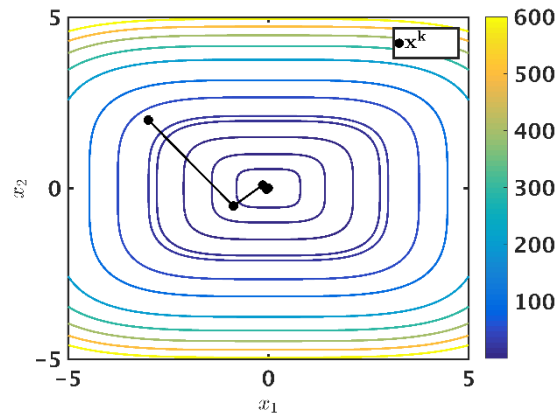
Une fois initialisée, la méthode de la plus forte pente répète le calcul du nouveau point  $\mathbf{x}^{(k+1)}$  jusqu'à convergence. Cette dernière peut être définie à partir des valeurs de  $f$  ou de la distance entre les points itérés, avec  $\epsilon$  un scalaire qui tend vers 0 :

$$|f(\mathbf{x}^{(k+1)}) - f(\mathbf{x}^{(k)})| < \epsilon \text{ ou } |\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}| < \epsilon \quad [\text{III.8}]$$

Un exemple d'optimisation sur la fonction  $f(\mathbf{x}) = x_1^4/4 + x_2^4$  est visible sur la Figure III-4. Le point de départ de l'algorithme est  $\mathbf{x}^{(0)} = [-3, 2]$  et le critère de convergence est  $\epsilon = 1e^{-4}$ . La méthode nécessite 5 itérations avant de converger et 170 appels à la fonction objectif pour le calcul de la dérivée par différences



finies.



**Figure III-4 : Exemple d'optimisation d'une fonction convexe en 5 itérations par méthode de gradient à pas optimal.**

### b. CMA-ES

L'algorithme CMA-ES [66] pour Covariance Matrix Adaptation Evolution Strategy contient seulement une procédure d'adaptation et de sélection de la population contrairement à un algorithme génétique classique. Les nouvelles populations sont générées en utilisant une distribution multi-normale qui est calculée et adaptée à partir des itérations précédentes. Cette méthode permet de traiter des problèmes non-linéaires ou non-convexes avec une faible taille de population à chaque génération.

CMA-ES repose sur la génération d'une population de  $\lambda_C$  individus selon une distribution normale multivariée  $\mathcal{N}(m_C, \mathbf{C}_C)$ . Cette distribution est déterminée par sa moyenne  $m_C$  et sa matrice de covariance  $\mathbf{C}_C$ . Le support de cette distribution est infini mais ses contours isoprobabilistes peuvent être identifiés comme des ellipses. On considère généralement une seule ellipse, par exemple celle où la probabilité qu'un point soit échantillonné dans le volume défini soit de 0.99. Cette ellipse délimite alors un hypervolume de recherche dans lequel des individus sont générés. Le principe de CMA-ES consiste à déplacer et mettre à jour cet hypervolume pour qu'il se resserre sur l'optimum global de la fonction à optimiser.

Les axes principaux de l'ellipse correspondent aux vecteurs propres associés à la matrice de covariance et leur longueur correspond à la racine carrée des valeurs propres. La décomposition en valeur propre de la matrice de covariance, qui est symétrique définie positive, s'écrit  $\mathbf{C}_C = \mathbf{B}_C \mathbf{D}_C^2 \mathbf{B}_C^T$  avec  $\mathbf{D}_C$  la matrice diagonale contenant les valeurs propres et  $\mathbf{B}_C$  la matrice de passage contenant les vecteurs propres. Il est alors possible d'écrire la distribution de plusieurs façons, avec  $\mathbf{I}$  correspondant à la matrice identité :

$$\begin{aligned} \mathcal{N}(m_C, \mathbf{C}_C) &\sim m_C + \mathcal{N}(0, \mathbf{C}_C) \\ &\sim m_C + \mathbf{C}_C^{\frac{1}{2}} \mathcal{N}(0, \mathbf{I}) \\ &\sim m_C + \mathbf{B}_C \mathbf{D}_C \mathbf{B}_C^T \mathcal{N}(0, \mathbf{I}) \\ &\sim m_C + \mathbf{B}_C \mathbf{D}_C \mathcal{N}(0, \mathbf{I}) \end{aligned}$$

Si on souhaite déplacer le centre de l'ellipse vers l'optimum, il faut actualiser la moyenne. Pour cela, une sélection des  $\mu_C$  meilleurs individus  $\mathbf{x}$  de la génération  $k$  est effectuée. Puis la moyenne est recalculée selon :

$$m_C^{(k+1)} = \sum_{i=1}^{\mu_C} \omega_i \mathbf{x}_{i:\lambda}^{(k+1)} \text{ avec } \sum_{i=1}^{\mu_C} \omega_i = 1 \quad [\text{III.9}]$$

Déplacer le centre de l'ellipse ne suffit pas à mieux diriger l'optimisation, il faut également orienter la population vers l'optimum en actualisant la matrice de covariance. Cette matrice met en relation les informations provenant des individus sélectionnés à chaque génération et le chemin parcouru durant l'historique d'optimisation. Un vecteur d'évolution  $p_c^{(k+1)}$  faisant intervenir la distance entre les points des

génération successive est calculé :

$$p_c^{(k+1)} = (1 - c_c)p_c^{(k)} + \sqrt{c_c(2 - c_c)}\mu_{eff} \frac{m_c^{(k+1)} - m_c^{(k)}}{\sigma_c^{(k)}} \quad [\text{III.10}]$$

où  $\mu_{eff}$  est la variance des poids sélectionnés,  $c_c \leq 1$  est le taux d'apprentissage d'accumulation et  $\sigma_c$  la taille du pas. La matrice de covariance s'exprime alors de la façon suivante :

$$\mathbf{C}_c^{(k+1)} = (1 - c_1 - c_\mu)\mathbf{C}_c^{(k)} + c_1 \underbrace{p_c^{(k+1)} p_c^{(k+1)T}}_{\text{mise à jour de rang 1}} + c_\mu \sum_{i=1}^{\mu_C} \omega_i \underbrace{l_{i,\lambda}^{(k+1)} l_{i,\lambda}^{(k+1)T}}_{\text{mise à jour de rang } \mu} \quad [\text{III.11}]$$

où  $c_1$  et  $c_\mu$  sont des taux d'apprentissage pour la mise à jour de la matrice de covariance et  $l_{i,\lambda}^{(k+1)} = (\mathbf{x}_{i,\lambda}^{(k+1)} - m_c^{(k)})/\sigma_c^{(k)}$ . Cette formulation permet d'utiliser l'information sur la population d'une génération avec la mise à jour de rang  $\mu_C$  d'un côté (importante avec les populations nombreuses), et la corrélation entre les générations de la mise à jour de rang 1 (importante avec les faibles populations). Par contre la matrice de covariance, telle qu'elle est présentée précédemment, ne permet pas de contrôler explicitement la dispersion de la génération. Pour cela, il faut ajuster le paramètre de taille du pas  $\sigma_c$  en utilisant la méthode Cumulative Step length Adaptation qui fait intervenir le paramètre d'évolution  $p_\sigma^{(k+1)}$  :

$$p_\sigma^{(k+1)} = (1 - c_\sigma)p_\sigma^{(k)} + \mathbf{C}_c^{(k)-1/2} \sqrt{c_\sigma(2 - c_\sigma)}\mu_{eff} \left( \frac{m_c^{(k+1)} - m_c^{(k)}}{\sigma_c^{(k)}} \right) \quad [\text{III.12}]$$

où  $c_\sigma < 1$  est le taux d'apprentissage du paramètre d'évolution. Quand  $p_\sigma^{(k+1)}$  a été actualisé, la taille du pas peut être mise à jour :

$$\sigma_c^{(k+1)} = \sigma_c^{(k)} \exp \left( \frac{c_\sigma}{d_\sigma} \left( \frac{\|p_\sigma^{(k+1)}\|}{\mathbb{E}\|\mathcal{N}(0, \mathbf{I})\|} \right) \right) \quad [\text{III.13}]$$

où  $d_\sigma > 0$  est le paramètre de contrôle de la magnitude de  $\sigma$  et  $\mathbb{E}\|\mathcal{N}(0, \mathbf{I})\|$  est la norme espérée d'une variable normale distribuée selon  $\mathcal{N}(0, \mathbf{I})$  qui est le critère de comparaison pour définir si  $p_\sigma$  est trop petit ou trop grand. Pour simplifier les étapes précédemment décrites, l'algorithme CMA-ES peut se résumer par les étapes suivantes :

- Initialisation de la matrice de covariance  $\mathbf{C}_c^{(0)} = \mathbf{I}$ , de la moyenne  $m_c^{(0)}$ , de la taille du pas  $\sigma_c^{(0)}$  et des paramètres de sélection  $(\lambda_c, \mu_c, \omega, c_c, c_1, c_\mu, c_\sigma, d_\sigma)$ .
- $k \leftarrow 0$
- **Tant que** la convergence du CMA-ES n'est pas atteinte, **faire** :
  - Générer un nouvel échantillon de  $\lambda_c$  individus selon  $\mathcal{N}(m_c^{(k)}, \mathbf{C}_c^{(k)})$  et sélectionner les  $\mu_c$  meilleurs individus.
  - Mettre à jour la moyenne  $m_c^{(k+1)}$ , les paramètres de chemin et de pas  $(p_c^{(k+1)}, p_\sigma^{(k+1)}, \sigma_c^{(k+1)})$
  - Mettre à jour la matrice de covariance  $\mathbf{C}_c^{(k+1)}$
- Retourner l'optimum obtenu.

Les critères de convergence numérique du CMA-ES sont basés sur le nombre d'appels à la fonction à optimiser, la valeur de l'optimum obtenu ou l'écart type de la population échantillonnée (lié la longueur des axes principaux de l'ellipse).

Cet algorithme est celui qui est généralement utilisé pour réaliser l'optimisation des sous-problèmes dans la suite du document. Il nécessite un nombre d'appels à la fonction objectif plus élevé qu'une méthode de gradient mais permet beaucoup plus sûrement d'atteindre le minimum global des critères des stratégies adaptatives.

## 2. Critères pour le krigeage

Une grande variété de critères a été développée pour obtenir le minimum global d'une fonction métamodélisée par le krigeage. Le chapitre 4 du mémoire de thèse de Sasena [67] présente un aperçu des critères existants. Il effectue également une comparaison de ces critères sur plusieurs exemples d'optimisation à partir de krigeage. Dans cette section, deux critères seulement sont présentés. Le premier critère, développé par Kushner [68], est à l'origine des stratégies d'enrichissement des données d'entraînement pour l'optimisation. Le deuxième critère nommé amélioration espérée est le plus utilisé actuellement de par son efficacité.

### a. Probabilité d'amélioration

A partir de la prédiction du krigeage, ce critère quantifie la probabilité de trouver un meilleur minimum que celui qui est actuellement présent dans les données d'entraînement. Il convient de définir dans un premier temps la notion d'amélioration  $I$  à partir du minimum actuel des données  $y_{\min}$  et de la réalisation d'un processus gaussien  $Z$  :

$$I(\mathbf{x}) = \begin{cases} y_{\min} - Z(\mathbf{x}), & \text{si } Z(\mathbf{x}) < y_{\min} \\ 0, & \text{sinon} \end{cases} \quad \text{[III.14]}$$

L'amélioration  $I$  dépend de la réalisation d'un processus gaussien. Elle correspond donc à une variable gaussienne, ici tronquée aux valeurs positives. Il est alors possible de calculer la probabilité que l'amélioration soit non nulle, ce qui équivaut à estimer  $\mathcal{P}(Z(\mathbf{x}) < y_{\min})$ . Cette quantité correspond à la probabilité d'amélioration (PI pour Probability of Improvement) :

$$\text{PI}(\mathbf{x}) = \mathcal{P}(Z(\mathbf{x}) < y_{\min}) = \Phi\left(\frac{y_{\min} - \hat{y}(\mathbf{x})}{\hat{\sigma}_y(\mathbf{x})}\right) \quad \text{[III.15]}$$

avec pour rappel  $\Phi$  correspondant à la fonction de répartition de la loi normale,  $\hat{y}$  la moyenne de prédiction du krigeage et  $\hat{\sigma}_y$  son écart type. Dans ce critère, l'erreur de prédiction n'est que très peu prise en compte. Ainsi le raffinement de la surface de réponse est essentiellement local, autour du minimum actuel. Kushner [68] a défini sa probabilité d'amélioration  $\text{PI}_K$  en utilisant un poids  $\epsilon_K$  :

$$\text{PI}_K(\mathbf{x}, \epsilon_K) = \mathcal{P}(Z(\mathbf{x}) < y_{\min} - \epsilon_K) \quad \text{[III.16]}$$

Lorsque la valeur du poids est élevée, le critère  $\text{PI}_K$  favorise un peu plus l'exploration. Lorsque le poids est faible, le critère  $\text{PI}_K$  tend vers le comportement du critère PI. L'inconvénient de ce critère réside dans le choix du poids lors du processus de stratégie adaptative. Lors de l'enrichissement, il est intéressant d'avoir un poids plus grand au début et de le réduire algorithmiquement au fur et à mesure. Mais il est difficile de déterminer exactement quand changer de valeurs afin de ne pas trop explorer ou ne pas raffiner localement trop tôt. Le critère d'amélioration espérée décrit dans la section suivante a été développé en partie pour éviter le problème du poids de Kushner.

En pratique, le critère PI amène à ajouter beaucoup de points proches entre eux dans le jeu de données d'entraînement. Ceci a pour conséquence un mauvais conditionnement de la matrice de corrélation  $\mathbf{C}_{tr}$  du krigeage, qui peut amener à stopper le processus de la stratégie adaptative pour conserver une certaine stabilité numérique. Une condition sur la distance entre le point à ajouter et les points déjà présents dans les données d'entraînement est donc souvent mise en œuvre pour éviter ces problèmes de conditionnement. Cette condition d'arrêt a d'ailleurs été souvent atteinte avant celle du nombre maximal d'appels autorisé au modèle numérique lors de nos expérimentations. Sur la Figure III-5, le critère PI a été calculé à partir d'un exemple quelconque. L'échelle des valeurs du critère a été modifiée pour faciliter la visualisation par rapport aux valeurs prédites par le krigeage. On observe bien ici que le maximum du critère se situe très près d'un point déjà évalué et que l'erreur de prédiction fournie par  $\hat{\sigma}_y$  n'est pas assez prise en compte.

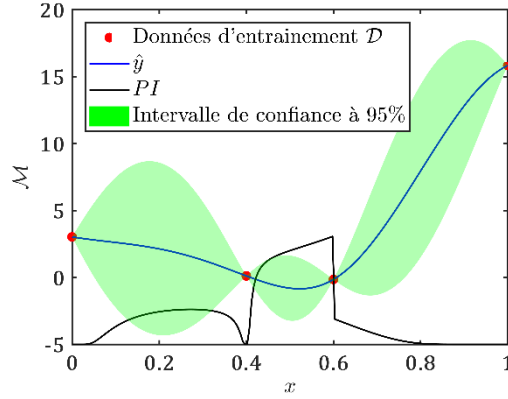


Figure III-5 : Critère de probabilité d'amélioration sur un exemple 1D. L'échelle de valeurs du critère est modifiée pour faciliter la lecture.

### b. Amélioration espérée

La définition du critère d'amélioration espérée est fondée sur la notion d'amélioration comme le critère PI. Cependant, au lieu de provenir de la probabilité de l'évènement, c'est à partir de l'espérance mathématique de l'amélioration que ce critère est calculé. Selon l'amélioration définie à l'équation [III.14], l'amélioration espérée (EI pour Expected Improvement) se formule :

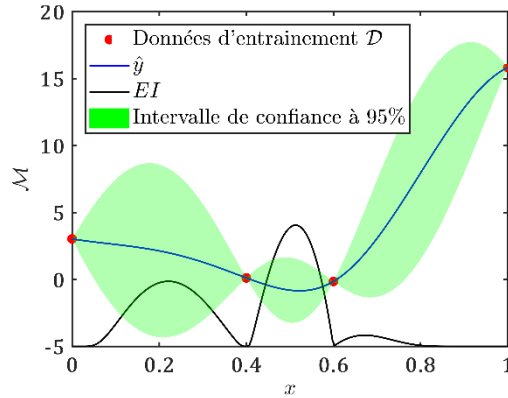
$$EI(\mathbf{x}) \equiv \mathbb{E}[I(\mathbf{x})] = \int_{-\infty}^{y_{\min}} (y_{\min} - z) \varphi\left(\frac{z - \hat{y}(\mathbf{x})}{\hat{\sigma}_{\hat{y}}(\mathbf{x})}\right) dz \quad [III.17]$$

Après un développement et une intégration par partie (voir Bichon [69] p. 143-145), l'équation [III.17] se ramène à la forme suivante :

$$EI(\mathbf{x}) = (y_{\min} - \hat{y}(\mathbf{x}))\Phi\left(\frac{y_{\min} - \hat{y}(\mathbf{x})}{\hat{\sigma}_{\hat{y}}(\mathbf{x})}\right) + \hat{\sigma}_{\hat{y}}(\mathbf{x})\varphi\left(\frac{y_{\min} - \hat{y}(\mathbf{x})}{\hat{\sigma}_{\hat{y}}(\mathbf{x})}\right) \quad [III.18]$$

Cette expression se compose de deux éléments additionnés. Dans le terme de gauche, on retrouve la probabilité d'amélioration multipliée par l'amélioration non limitée aux valeurs positives. Ce terme permet de mettre en avant les zones où le métamodèle prédit une valeur plus faible que le minimum actuel des données d'entraînement. Il correspond donc à une exploitation directe des valeurs fournies par la prédiction. Le terme de droite est un terme d'exploration, qui fait intervenir l'écart type des valeurs prédites. L'ensemble de l'expression permet de réaliser un compromis entre l'ajout de points dans les zones où un minimum est observé et les zones de l'espace qui sont mal prédites par le métamodèle.

Le critère tel qu'il est présenté à l'équation [III.18] doit être maximisé. Le mauvais conditionnement de la matrice de corrélation du krigeage est évité car le terme de droite devient proche de 0 lorsque le point à évaluer est proche d'un point d'entraînement. Une condition d'arrêt sur la valeur même du critère EI est possible car cette valeur est homogène à l'écart entre le minimum actuel et celui qu'on peut obtenir en ajoutant un nouveau point. Dans cette étude, si la valeur de l'EI à son optimum est inférieure à  $10^{-4}$  durant trois itérations successives alors l'algorithme adaptatif est stoppé. La validation du critère sur plusieurs itérations successives permet d'éviter une convergence trop rapide vers un minimum local. Un exemple de la forme du critère est visible sur la Figure III-6. L'échelle des valeurs du critère a été modifiée pour faciliter la visualisation par rapport aux valeurs prédites par le krigeage. Deux pics de maximums locaux très nets sont observés, l'un pour le raffinement local, où le minimum global de la prédiction se situe, l'autre pour l'exploration d'une zone éloignée de deux points d'entraînement, où les valeurs prédites sont relativement faibles.



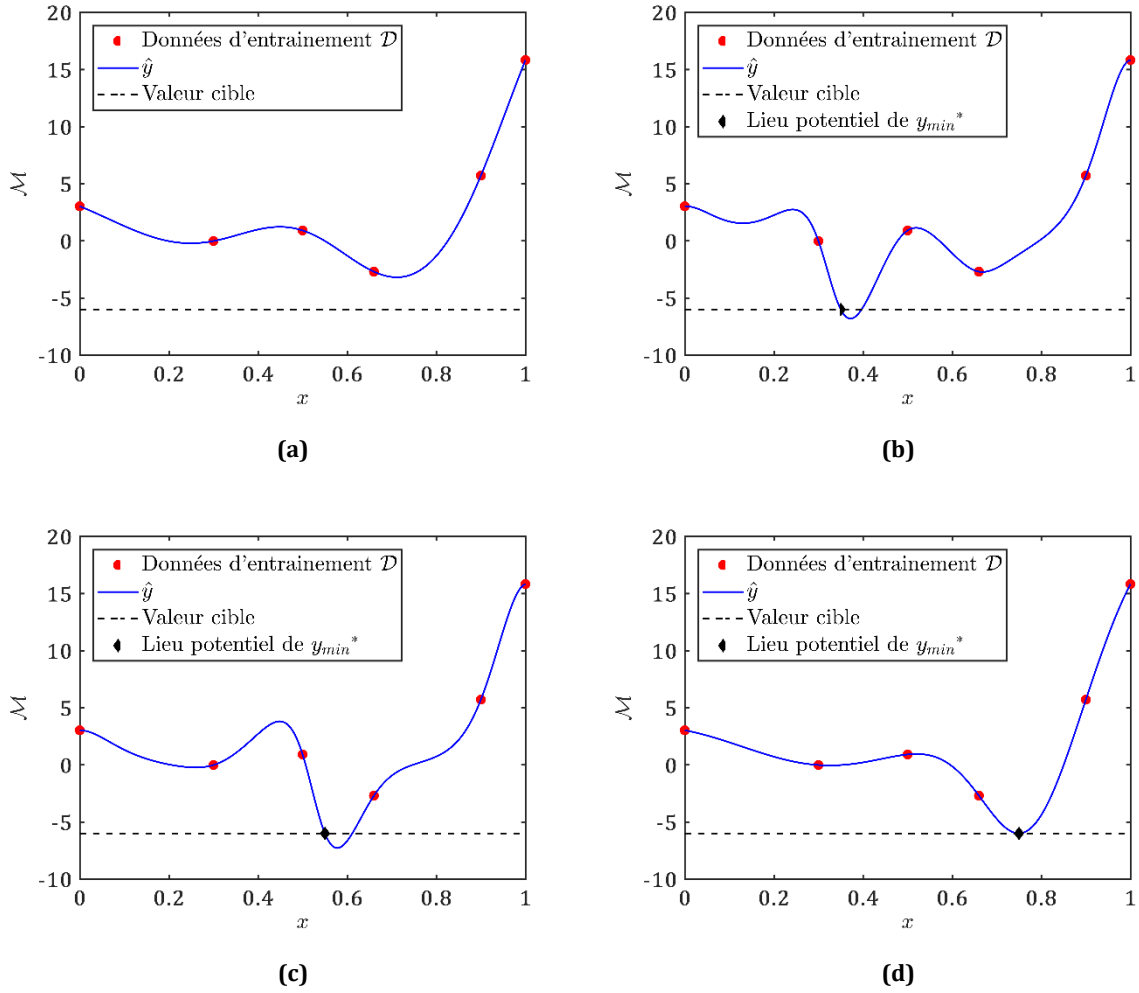
**Figure III-6 : Critère d'amélioration espérée sur un exemple 1D. L'échelle de valeurs du critère est modifiée pour faciliter la lecture.**

### 3. Critères pour les RBF

Comparativement au cas du krigeage, les stratégies adaptatives d'optimisation de métamodèles fondés sur les RBF comportent moins de choix de critères. Les travaux de Gutmann [70] sont à l'origine du premier critère en 2001. Un deuxième critère a également été développé par Regis et Shoemaker [71].

#### a. Gutmann-RBF

La méthode définie par Gutmann se base sur la méthode itérative de Jones [72]. On considère qu'une estimation du minimum global  $y_{\min}^*$  de la fonction à optimiser est disponible. Cette dernière est appelée valeur cible de l'optimisation. L'idée initiale de Jones [72] est d'ajouter des points au métamodèle là où le minimum global doit raisonnablement se situer, sous réserve que la valeur du minimum global corresponde bien à la valeur cible  $y_{\min}^*$ . En pratique, il s'agit de trouver le point  $t \notin \mathcal{X}_{tr}$  tel que le métamodèle entraîné sur les données  $\mathcal{D}_t = \{\mathcal{X}_{tr} \cup \{t\}, \mathcal{Y}_{tr} \cup \{y_{\min}^*\}\}$  donne une interpolation la plus « raisonnable ». Un exemple est présenté sur la Figure III-7 où, à partir d'un jeu de données d'entraînement  $\mathcal{D}$  et de la RBF construite dessus, trois points de différentes localisations  $t$  sont ajoutés avec la valeur  $y_{\min}^*$ . La sous-figure (a) représente l'état initial et les sous-figures (b), (c) et (d) correspondent à l'ajout de points. Parmi ces trois exemples de localisation d'ajout de points, l'interpolant le plus raisonnable correspond à la sous-figure (d) car la RBF résultante est la plus proche de l'état initial.



**Figure III-7 : Exemple de recherche de lieu le plus raisonnable d'un point correspondant à la valeur cible  $y_{\min}^*$ . La sous-figure (a) représente l'état initial de la prédiction de la RBF à partir des données d'entraînement. Les sous-figures (b), (c) et (d) représentent l'ajout d'un point aux données d'entraînement en différents lieux avec la valeur  $y_{\min}^*$  et la RBF résultant de ces nouvelles données.**

Gutmann [70] interprète la notion de raisonnable comme étant l'interpolant le moins « bosselé », cette dernière caractéristique étant mesurée par une semi-norme de l'interpolant  $\widehat{\mathcal{M}}$  :

$$I_G(\widehat{\mathcal{M}}) = \int [\widehat{\mathcal{M}}''(\mathbf{x})]^2 d\mathbf{x} \quad \text{[III-19]}$$

Gutmann montre que l'interpolation obtenue avec les RBF minimise cette semi-norme. A partir de cette expression, il dérive le critère suivant à maximiser pour obtenir le point à évaluer le plus « raisonnable » pour une valeur cible choisie :

$$\text{GUT}(\mathbf{x}) = \left( \frac{P_\varphi(\mathbf{x})}{\widehat{y}(\mathbf{x}) - f_{\min}^*} \right)^2 \quad \text{[III-20]}$$

avec  $f_{\min}^*$  la valeur cible à choisir et pour rappel  $P_\varphi$  la mesure de l'erreur de prédiction de la RBF. En pratique, la quantité  $-\log(\text{GUT})$  est minimisée pour condenser la gamme de variation, et ainsi éviter des difficultés numériques [73]. La valeur cible  $f_{\min}^*$  est choisie dans l'intervalle  $[-\infty, y_{\min}^*]$  sur des cycles de  $N_{\text{GUT}} + 1$  itérations. Chaque cycle commence avec une valeur cible très faible pour finir vers des valeurs proches de  $y_{\min}^*$ . Ce changement de valeur cible permet d'alterner entre l'exploration globale de l'espace (pour des faibles valeurs cibles) et l'exploitation locale du métamodèle (pour les valeurs plus élevées) [70].

La procédure de choix de la valeur cible au cours des itérations du cycle est maintenant détaillée. Tout d'abord, définissons  $\alpha_g$  une permutation de  $\{1, \dots, n\}$  tel que :

$$y\left(\mathbf{x}_{tr}^{(\alpha_g(1))}\right) \leq \dots \leq y\left(\mathbf{x}_{tr}^{(\alpha_g(n))}\right) \quad \text{[III-21]}$$

La valeur cible se définit alors comme :

$$f_{\min}^* = y_{\min}^* - W_n \times \left( y\left(\mathbf{x}_{tr}^{(\alpha_g(k_n))}\right) - y_{\min}^* \right) \quad \text{[III-22]}$$

où :

$$W_n = \left( \frac{N_{GUT} - (n - n_0) \bmod N_{GUT} + 1}{N_{GUT}} \right)^2 \quad \text{[III-23]}$$

$$k_n = \begin{cases} n, & \text{si } (n - n_0) \bmod N_{GUT} + 1 = 0 \\ k_{n-1} - \lfloor (n - n_0) / N_{GUT} \rfloor, & \text{sinon} \end{cases}$$

Des précautions doivent être prises dans le cas où  $W_n = 0$ , car il implique  $f_{\min}^* = y_{\min}^*$ . Par ailleurs, si la valeur  $y_{\min}^*$  est proche d'un point déjà évalué, il convient de redéfinir  $f_{\min}^*$  pour qu'il soit légèrement inférieur à cette valeur. Le critère d'arrêt de l'algorithme est l'atteinte d'un nombre maximal d'itérations défini. Des problèmes de convergence de cet algorithme ont été observés par Regis et Shoemaker [74] dans certains cas lorsque un point du jeu de données d'entraînement est proche du minimum global. Ils observent que la recherche locale peine à ajouter des points dans le bassin du minimum global. Ce type de problème apparaît pour des minimums où la fonction présente de forts gradients. Ils proposent alors de restreindre dans ce cas la zone d'optimisation du critère GUT au bassin du minimum global pour les itérations où  $W_n$  est proche de 0. Cette extension de l'algorithme de Gutmann n'est pas traitée ici car les fonctions à optimiser dans les cas d'application qui nous concernent ne présentent pas ces caractéristiques.

#### b. CORS (Constrained Optimization Using Response Surfaces)

Une alternative au critère de Gutmann a été proposée par Regis et Shoemaker [71]. Ces auteurs définissent le critère  $\mathcal{C}$  à minimiser comme étant directement la réponse fournie par la RBF. Cette procédure peut amener à obtenir un point déjà évalué comme optimum de  $\mathcal{C}$ . Pour éviter ce problème, ils ajoutent une contrainte au problème d'optimisation basée sur la distance avec les points du jeu de données d'entraînement. La distance « maximin » est utilisée ici pour mesurer la limite haute possible :

$$\Delta_{Mm} = \max_{\mathbf{x} \in \mathbb{X}} \min_{1 \leq i \leq n} \|\mathbf{x} - \mathbf{x}_{tr}^{(i)}\| \quad \text{[III-24]}$$

Le problème d'optimisation est alors défini de la manière suivante :

$$\begin{aligned} & \min_{\mathbf{x} \in \mathbb{X}} \hat{y}(\mathbf{x}) \\ & \text{sous } \|\mathbf{x} - \mathbf{x}_{tr}^{(i)}\| \geq \beta_N \Delta_{Mm}, i = 1, \dots, n \end{aligned} \quad \text{[III-25]}$$

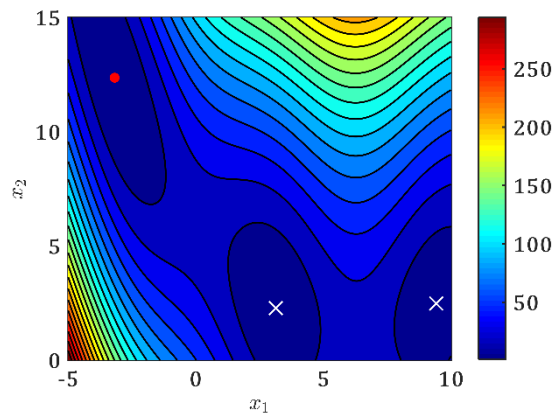
avec  $\beta_N$  le facteur de distance compris entre 0 et 1. Pour passer de l'exploration à l'exploitation locale du métamodèle, des cycles de  $N_{CORS} + 1$  itérations sont effectués dans lesquels le paramètre  $\beta_N$  décroît de 1 (exploration) vers 0 (exploitation). Lorsque  $\beta_N = 0$  et que le résultat du problème d'optimisation [I-16] est un point déjà évalué, l'itération est recommencée en diminuant la valeur du facteur de distance. Pour la suite de l'étude, les valeurs choisies pour le facteur de distance sont  $\beta_N = \{0.90, 0.75, 0.25, 0.05, 0.03, 10^{-3}\}$ . Ces valeurs permettent un bon compromis entre exploration et raffinement local [71]. Le critère d'arrêt de cette méthode est, comme dans la méthode de Gutmann, basé sur le nombre maximal d'itérations ou, autrement dit, le nombre d'appel à la fonction coûteuse.

#### 4. Comparaison des critères d'optimisation non contrainte sur un exemple analytique

Dans cette section, les critères décrits précédemment sont comparés lors de la recherche du minimum de la fonction de Branin. Cette fonction est dérivée d'un exemple analytique de référence pour la comparaison de métamodèles ou de méthodes d'optimisation [75]. Elle est définie selon l'équation suivante :

$$\mathcal{M}_{Branin}(x_1, x_2) = \left(x_2 - \frac{5.1}{4\pi^2}x_2 + \frac{5}{\pi}x_1 - 6\right)^2 + 10\left(\left(1 - \frac{1}{8\pi}\right)\cos(x_1) + 1\right) + 5x_1 \quad \text{[III-26]}$$

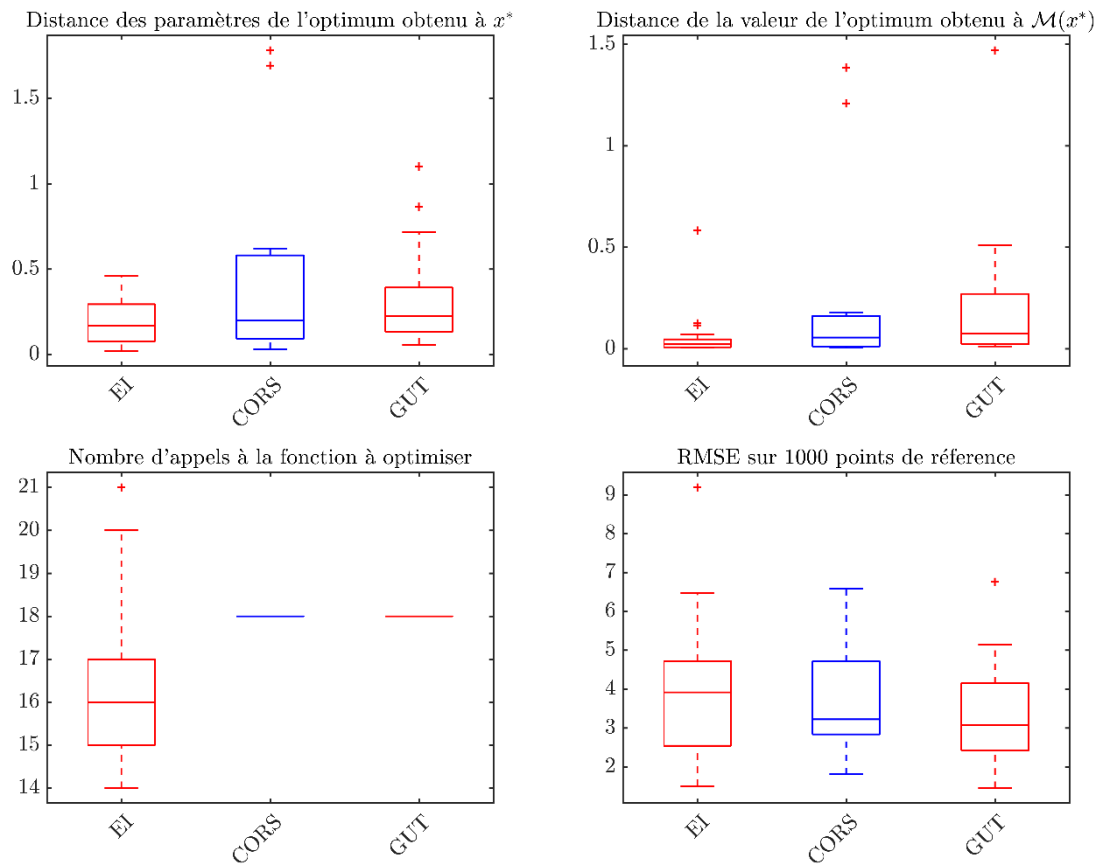
avec  $x_1 \in [-5, 10]$  et  $x_2 \in [0, 15]$ . La fonction de Branin possède un minimum global  $x^*$ , situé approximativement en  $[-3.18, 12.36]$ , dont la valeur est environ 1.01. Deux minimums locaux sont également présents ce qui rend l'optimisation un peu plus complexe. La Figure III-8 représente les courbes de niveaux de la fonction de Branin avec la localisation du minimum global au point rouge et des minimums locaux aux croix blanches.



**Figure III-8 : Courbes de niveaux de la fonction de Branin. Le point rouge représente le minimum global et les croix blanches les minimums locaux.**

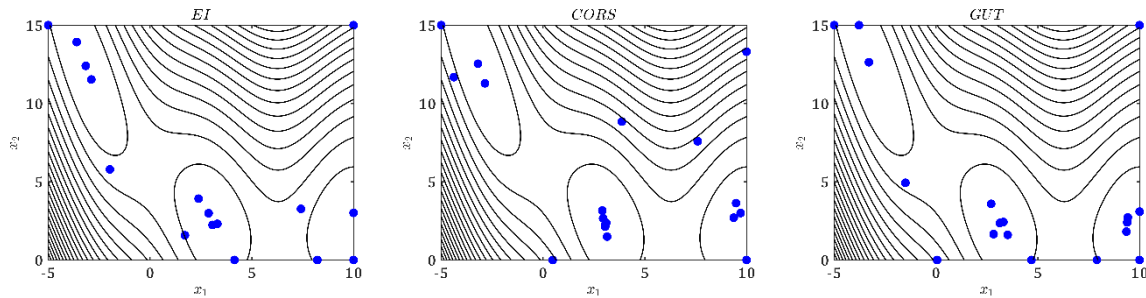
Pour l'initialisation de l'optimisation, 20 points d'entraînement sont utilisés et la fonction de corrélation ou base radiale utilisée dans les métamodèles est le noyau Matérn 5/2. L'optimisation par critère EI, CORS ou GUT est limitée à 18 itérations. Ce choix a été effectué après analyse des résultats de l'optimisation par critère EI. En effet, pour rappel, l'amélioration espérée possède une condition d'arrêt liée à la valeur de son critère. Ici cela a permis d'estimer le nombre raisonnable d'appels à la fonction de Branin permettant de trouver un optimum. Afin de limiter l'impact du plan d'expérience initial sur les résultats, une vingtaine d'expériences ont été réalisées pour chaque méthode, à partir d'un plan d'expérience initial différent à chaque fois. Les résultats obtenus sont présentés sur la Figure III-9. Sur ces résultats, la méthode utilisant le critère PI n'est pas présentée car elle présente un très mauvais bilan. La condition d'arrêt atteinte pour ce critère est toujours celle concernant l'ajout de points trop proches avec ceux déjà présent dans le plan d'entraînement. Ce phénomène est lié à une convergence du critère PI vers le même minimum local sur plusieurs itérations. Ce critère a donc été mis de côté pour faciliter l'interprétation visuelle des résultats des autres méthodes.





**Figure III-9 : Résultats de l'optimisation de la fonction de Branin à partir de trois des quatre critères présentés précédemment sur 20 expériences différentes. Les critères observés sont les distances à l'optimum (paramètres obtenus et valeurs), le nombre d'appels à la fonction Branin et la précision globale du métamodèle en fin d'optimisation.**

Sur la Figure III-9, plusieurs mesures sont choisies pour étudier les résultats de l'optimisation fournie par chaque méthode. Tout d'abord la distance à l'optimum est analysée. La localisation de l'optimum renvoyé à la fin des différents algorithmes et la valeur de la fonction en ce point sont comparées aux données du minimum global de fonction Branin. Sur ce point, le critère EI obtient le meilleur résultat. Le critère CORS obtient en moyenne une meilleure valeur de fonction que le critère de Gutmann mais l'optimum est localisé un peu moins précisément. Ensuite c'est le nombre d'appels à la fonction Branin qui est compté. Sur ce point le critère EI a dépassé trois fois les 18 appels par rapport aux méthodes à base de RBF. Pour autant, il a souvent eu moins d'appels que les méthodes CORS et Gutmann, tout en restant performant sur la précision. En dernière mesure, la précision globale du métamodèle à la fin de la stratégie adaptative est observée, en utilisant 1000 points de référence pour calculer la RMSE. Ici les méthodes à base de RBF sont meilleures avec un avantage pour le critère de Gutmann. Pour terminer la comparaison, la Figure III-10 permet de visualiser les points ajoutés à l'issue de la stratégie adaptative, pour chaque méthode et sur un même plan d'expérience de départ.



**Figure III-10 : Localisation des points ajoutés suite à la stratégie adaptative pour chaque méthode sur la même expérience.**

On constate que chaque méthode a dépensé plusieurs appels sur le minimum local proche de  $x_1 = 3$  et finalement, peu de points sont localisés dans le bassin du minimum global. Pour autant, il n'y a pas de fortes différences entre les points évalués par chaque méthode. Sur cet exemple numérique, on peut même conclure que les trois stratégies adaptatives basées sur les critères EI, CORS et Gutmann produisent des résultats relativement proches.

Ces méthodes peuvent être étendues au cas où des fonctions contraintes doivent être respectées. Cet aspect est traité dans la section suivante.

## 5. Optimisation sous contraintes

En conception, la notion de contrainte dans le dimensionnement d'un composant est souvent liée à une condition du cahier des charges à satisfaire obligatoirement sous peine de futur dysfonctionnement. Dans l'exemple du cas d'application de la microsource infrarouge, cela se traduit par le fait que la température maximale en tout point du composant ne doit pas dépasser 650 °C. Dans cette section, le problème défini à l'équation [III.4] doit maintenant satisfaire une ou plusieurs contraintes :

$$\begin{aligned} \min_{\mathbf{x} \in \mathbb{X}} \mathcal{M}(\mathbf{x}) \\ \text{sous } \mathbf{g}(\mathbf{x}) \leq 0 \end{aligned} \quad \text{[III-27]}$$

Ces fonctions contraintes  $\mathbf{g}$  peuvent être de deux natures dans le cadre de cette étude : elles peuvent être issues de la résolution du modèle numérique coûteux  $\mathcal{M}$  (contraintes fonctionnelles) ou provenir d'une expression formelle.

Le cas où les contraintes sont directement disponibles sans passer par la simulation est le plus simple. L'algorithme d'optimisation du sous-problème doit alors intégrer ces contraintes dans la recherche de l'optimum. Dans le cadre de l'utilisation du CMA-ES, deux méthodes de prise en compte des contraintes sont possibles. La première consiste à régénérer un second individu lorsque le premier n'est pas en zone admissible vis-à-vis de la contrainte. Les individus peuvent être régénérés un grand nombre de fois avant d'en trouver un en zone admissible, ce qui constitue l'inconvénient principale de cette méthode. La valeur de la contrainte évaluée pour chaque individu rejeté n'est pas prise en compte. Une autre méthode a donc été développée par Chocat et al. [76] permettant de redéfinir l'hypervolume de génération des individus selon les valeurs de la fonction contrainte. Grâce à ce type de méthodes, la nouvelle formulation du sous-problème d'optimisation peut être résolue simplement.

Dans l'autre cas, les contraintes sont issues du modèle numérique et sont alors coûteuses à évaluer. Il faut donc utiliser un métamodèle afin de pouvoir accéder à une prédiction de la zone admissible pour un coût quasi-nul. A partir du métamodèle de la contrainte, deux options sont possibles. La prédiction des valeurs de la contrainte peut directement être utilisée comme contrainte du sous problème. On se ramène alors au cas décrit au paragraphe précédent. L'autre option consiste à utiliser complètement l'information fournie par le métamodèle, c'est-à-dire utiliser le cadre probabiliste du krigeage ou l'information sur l'erreur de prédiction des RBF. Dans la suite de cette étude, seuls les critères utilisant le krigeage ont été analysés et implémentés. Quelques possibilités sont détaillées dans la section suivante. Pour autant, il existe des

méthodes d'optimisation sous contrainte à partir des RBF. Par exemple, l'algorithme COBRA [77] propose un schéma en deux phases successives permettant de trouver d'abord un point en zone admissible puis de chercher à trouver une meilleure solution au sens de la fonction objectif à parti de ce point. Ici, les contraintes de l'optimisation du sous-problème correspondent à la prédiction fournie par les RBF.

a. Méthodes de prise en compte de la contrainte dans les stratégies adaptatives par krigeage

Les éléments décrits ci-dessous reprennent une étude effectuée par Durantin et al. [78] sur les différents critères permettant de prendre en compte les contraintes lors de l'optimisation par krigeage. Dans cette étude, la notion centrale concerne la probabilité de faisabilité PF. Cette mesure est analogue au critère PI comme le montre l'équation suivante où  $\hat{g}_i(\mathbf{x})$  correspond à la prédiction du krigeage pour la  $i$ -ème contrainte et  $\hat{\sigma}_{\hat{g}_i}(\mathbf{x})$  son écart type :

$$PF_i(\mathbf{x}) = \Phi\left(\frac{0 - \hat{g}_i(\mathbf{x})}{\hat{\sigma}_{\hat{g}_i}(\mathbf{x})}\right) \quad \text{[III-28]}$$

Ce critère permet de transformer le sous-problème d'optimisation en un problème non contraint en multipliant le critère EI utilisé pour l'optimisation par la probabilité de présence dans la zone admissible [79]. Le critère adaptatif à maximiser dans le sous problème devient alors :

$$\mathcal{C}(\mathbf{x}) = EI(\mathbf{x}) \times PF(\mathbf{x}) = EI(\mathbf{x}) \times \prod_{i=1}^{m_g} PF_i(\mathbf{x}) \quad \text{[III-29]}$$

où  $m_g$  est le nombre de fonctions contraintes à prendre en compte. Grâce à la probabilité de faisabilité, l'EI est conduit à tendre vers 0 dans les zones où il y a peu de chance de sélectionner un point admissible vis-à-vis de l'ensemble des contraintes. Par contre, il est à noter que ce critère confère à l'optimisation un comportement conservateur, proposant peu de points à proximité des bornes définies par les contraintes. Lors de tests numériques [78], des difficultés sont également observées sur le maintien d'un équilibre au cours de la stratégie adaptative entre une bonne estimation de l'optimum, l'admissibilité de la solution vis-à-vis des contraintes et une bonne approximation des contraintes. Pour pallier ces problèmes, Parr et al. [80] proposent de traiter le critère EI et le critère PF comme deux objectifs individuels à optimiser en même temps. Le sous-problème d'optimisation devient alors un problème d'optimisation multi-objectif :

$$\max_{\mathbf{x} \in \mathbb{X}} \left\{ EI(\mathbf{x}), \prod_{i=1}^{m_g} PF_i(\mathbf{x}) \right\} \quad \text{[III-30]}$$

La résolution de ce type de problème permet d'obtenir plusieurs solutions optimales révélatrices de différents compromis entre les deux objectifs. Plus de détails sur les méthodes de résolution des problèmes multi-objectif sont disponibles dans la section III.C. Parmi les solutions obtenues pour le problème [I-16], seule la solution maximisant le produit  $EI(\mathbf{x}) \times PF(\mathbf{x})$  est retenue à la fin. Il est important de noter que la solution dérivée à partir du problème multi-objectif peut être différente de celle obtenue par la maximisation directe du produit des critères. L'algorithme NSGA-II [81], décrit en section III.C.1.a, est utilisé ici pour résoudre le problème multi-objectif. Pour l'optimisation du critère  $EI \times PF$  (équation [III-29]) ou celle du critère EI versus PF (équation [III-30]), la condition d'arrêt utilisée pour le critère EI peut s'appliquer. Ici cela correspond à 3 itérations successives où la valeur de  $EI(\mathbf{x}) \times PF(\mathbf{x})$  est inférieure à  $10^{-4}$ .

L'étude de Durantin et al [78] présente d'autres critères pour les deux formulations du sous-problème (un ou plusieurs objectifs). Lors des comparaisons de plusieurs exemples numériques, le sous problème de l'équation [I-16] est le plus performant en moyenne. Il est surtout avantageux lorsque la zone admissible par les contraintes est disjointe. La formulation multi-objectif permet alors une meilleure exploration. Dans la section suivante, une comparaison sur un exemple analytique est effectuée entre les formulations des équations [III-29] et [I-16] pour le krigeage et la prise en compte directe de la prédiction des contraintes

pour les RBF.

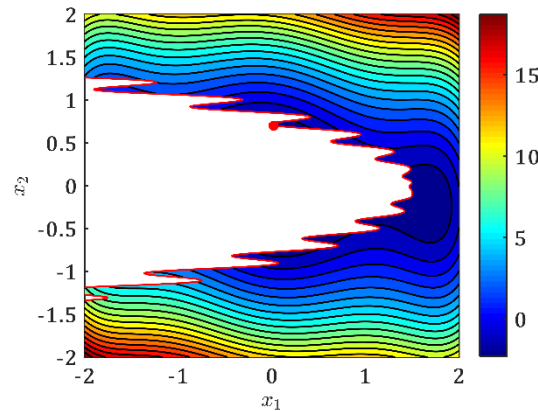
b. Comparaisons des critères d'optimisation sous contraintes sur un exemple analytique

Un nouvel exemple analytique est choisi ici pour comparer une partie des différents critères pour l'optimisation sous contraintes. La fonction à optimiser est la « six-hump Camel » [82], avec une contrainte multimodale définie sur le carré  $[-2, 2]^2$  :

$$\mathcal{M}_{Camel}(x_1, x_2) = \left(4 - 2.1x_1^2 + \frac{x_1^4}{3}\right)x_1^2 + x_1x_2 + (4x_2^2 - 4)x_2^2 \quad \text{[III-31]}$$

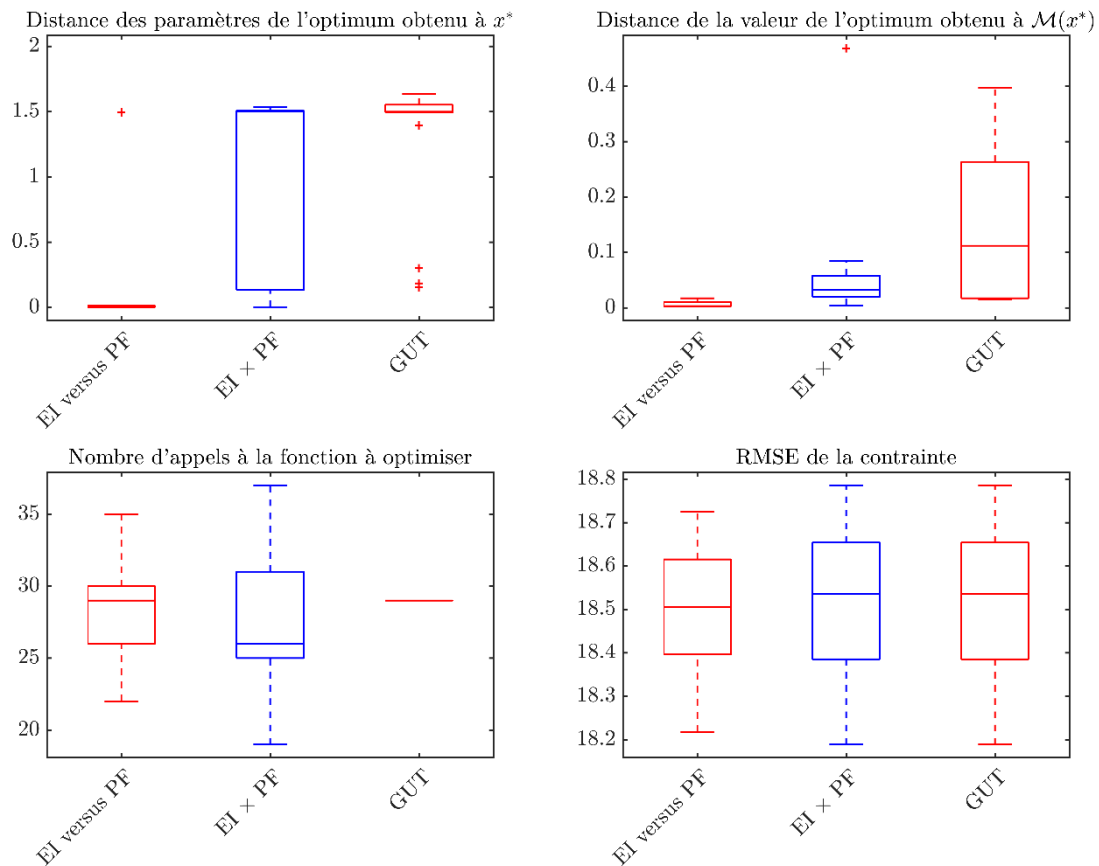
$$\text{sous } g_{Camel}(x_1, x_2) = 1.5 - \left(1.5x_2 - \frac{\cos(31x_2)}{6}\right)^2 - x_1 \leq 0$$

Le minimum global de cette fonction, de valeur -1.02, est situé à proximité du point  $[-0.03, 0.72]$ . La Figure III-11 représente les lignes de niveau de la fonction objectif ainsi que la limite de la zone admissible définie par la contrainte ( $g_{Camel} = 0$ ).



**Figure III-11 : Fonction « six-hump Camel » avec la contrainte définie à l'équation [I-16]. La zone blanche représente la partie non admissible. Le minimum global est localisé au point rouge.**

Trois critères sont comparés sur cet exemple : EI  $\times$  PF, EI versus PF et Gutmann utilisant directement les valeurs prédites de la contrainte. Pour l'initialisation de l'optimisation, 20 points d'entraînement sont utilisés et la fonction de corrélation ou base radiale utilisée dans les métamodèles est la fonction Matérn 5/2. L'optimisation suivant le critère de Gutmann est limitée à 29 itérations. Comme pour l'étude en section III.B.4, ce choix a été effectué après analyse des résultats de l'optimisation par critère EI versus PF permettant d'estimer un nombre raisonnable d'appels à la fonction pour trouver un optimum. Afin de limiter l'impact du plan d'expérience initial sur les résultats, une vingtaine d'expériences ont été réalisées pour chaque méthode avec un plan d'expérience initial différent à chaque fois. Les résultats sont visibles sur la Figure III-12, avec une comparaison sur la distance entre l'optimum obtenu par stratégie adaptative et le vrai optimum, sur le nombre d'appels à la fonction et sur la précision du métamodèle de la contrainte en fin d'optimisation.



**Figure III-12 : Résultats de l'optimisation sous contrainte de la fonction Camel à partir de trois méthodes sur 20 expériences différentes. Les critères observés sont les distances à l'optimum (paramètres obtenus et valeurs), le nombre d'appels à la fonction et la précision globale du métamodèle de la contrainte en fin d'optimisation.**

Comme attendu, la méthode la plus sophistiquée pour le krigeage (EI versus PF) est celle qui apporte les meilleurs résultats. L'optimisation du critère EI × PF a alterné dans les résultats entre le minimum global et un optimum local de la fonction Camel sous contrainte situé en  $[0.07, -0.77]$ . Avec les RBF et le critère de Gutmann, ce même minimum local a majoritairement été atteint. Il ne faut pas oublier que cette méthode est la moins évoluée pour la prise en compte de la contrainte dans le sous-problème d'optimisation. Au niveau des appels à la fonction objectif, le critère EI × PF a tendance à s'arrêter un peu plus vite que l'EI versus PF, ce qui peut expliquer le fait qu'il ait atteint un minimum local à plusieurs reprises. Concernant la précision de prédiction de la contrainte, les trois critères sont similaires, avec un léger avantage pour l'EI versus PF qui a été plus précis dans la zone de l'optimum global. Evidemment, l'exemple traité ici comporte une contrainte assez difficile à métamodéliser, ce qui a permis de mettre en difficulté les critères testés. On constate cependant qu'il est possible de trouver la solution au problème d'optimisation sous contrainte avec la méthode EI versus PF.

La problématique d'optimisation multi-objectif a ensuite été abordée avec l'optimisation du sous problème pour le critère EI versus PF. Cette formulation peut aussi être utilisée pour le problème général de conception, lorsque plusieurs objectifs sont à améliorer en même temps. La définition de ce type de problème, un aperçu de méthodes existantes pour le résoudre ainsi que les critères de stratégies adaptatives liés sont traités dans la section suivante.

### C. Optimisation multi-objectif

Lorsque plusieurs critères de performance sont identifiés, on peut choisir de les optimiser simultanément. Certains objectifs peuvent également être décomposés en plusieurs « sous-objectifs ». Par exemple, l'efficacité de la microsource infrarouge correspond en fait conjointement à la maximisation de la puissance

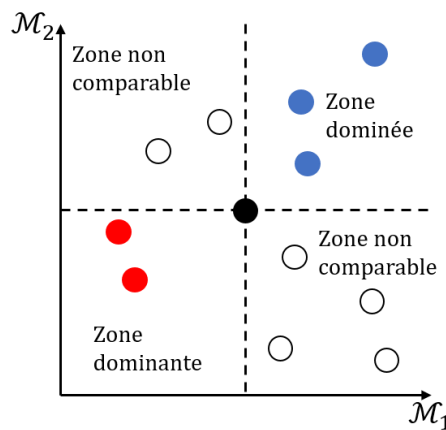
optique rayonnée et la minimisation de la puissance électrique consommée. On se retrouve alors dans le cadre de l'optimisation multi-objectif, avec une définition du problème suivant :

$$\min_{\mathbf{x} \in \mathbb{X}} \{\mathcal{M}_1(\mathbf{x}), \dots, \mathcal{M}_m(\mathbf{x})\} \quad [\text{III-32}]$$

avec, pour rappel,  $m$  le nombre d'objectifs en sortie du modèle numérique. Ce type de problèmes ne comporte pas de solution unique mais un ensemble de solutions représentatives de différents compromis entre les objectifs à optimiser. Pour définir les solutions optimales, la notion de dominance de Pareto est utilisée ici. On dit qu'une solution  $\mathbf{x}^{(1)}$  domine « en efficacité » une solution  $\mathbf{x}^{(2)}$  si toutes les valeurs des objectifs au point  $\mathbf{x}^{(1)}$  sont inférieures ou égales à celles au point  $\mathbf{x}^{(2)}$ , avec au moins une des inégalités strictes :

$$\mathbf{x}^{(1)} < \mathbf{x}^{(2)} \Leftrightarrow \begin{cases} \forall i \in \{1, \dots, m\}, \mathcal{M}_i(\mathbf{x}^{(1)}) \leq \mathcal{M}_i(\mathbf{x}^{(2)}) \\ \exists j \in \{1, \dots, m\}, \mathcal{M}_j(\mathbf{x}^{(1)}) < \mathcal{M}_j(\mathbf{x}^{(2)}) \end{cases} \quad [\text{III-33}]$$

Cette relation de dominance au sens de Pareto, notée  $<$ , ne permet pas de comparer toutes les solutions. Lorsque deux solutions ont des valeurs d'objectifs alternativement supérieure et inférieure, elles sont considérées comme équivalentes. La notion de dominance au sens de Pareto est illustrée sur la Figure III-13. Le point noir domine les points bleus, il est dominé par les points rouges et il est équivalent aux points blancs.

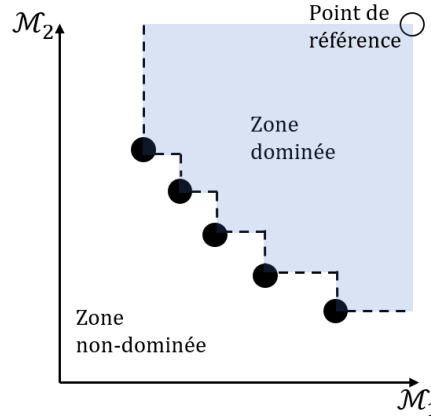


**Figure III-13 : Représentation de la dominance au sens de Pareto. Le point noir domine les points bleus, il est dominé par les rouges et les points blancs lui sont équivalents.**

En conservant l'ensemble des points non dominés, on obtient les points Pareto-optimaux. Il n'existe pas de solution unique au problème d'optimisation multi-objectif car l'ensemble des points non dominés sont équivalents entre eux. On appelle alors front de Pareto l'ensemble  $\mathbf{X}^{PF}$  regroupant ces solutions, défini par :

$$\mathbf{X}^{PF*} = \{\mathbf{x} \in \mathbb{X} : \nexists \mathbf{x}' \in \mathbb{X}, \mathbf{x}' < \mathbf{x}\} \quad [\text{III-34}]$$

Un front de Pareto est représenté sur la Figure III-14 par l'ensemble  $\mathbf{Y}^{PF}$  correspond aux réalisations des points de  $\mathbf{X}^{PF}$ . Dans le cas de deux fonctions objectifs, on retrouve souvent le front de Pareto représenté par un arc continu reliant les points Pareto-optimaux. Il faut rester critique vis-à-vis de cette représentation car la notion de continuité dans l'espace de réalisation des objectifs ne l'est pas forcément dans l'espace  $\mathbb{X}$  des paramètres : il est possible que deux points très proches dans l'espace des objectifs soient en réalité très éloignés dans l'espace des paramètres. Sur la Figure III-14, la ligne pointillée représente la limite de la zone dominée par le front de Pareto. Le point de référence nommé  $\mathbf{y}_{ref}$  (appelé également point NADIR) correspond au point ayant la pire valeur possible de tous les objectifs, un point qui est toujours dominé. Ce point est utile pour calculer l'hypervolume de domination d'un front de Pareto, une mesure permettant de comparer deux fronts de Pareto.



**Figure III-14 : Front de Pareto, représentation de l'ensemble  $Y^{PF}$ . Le point de référence  $y_{ref}$  est défini comme la pire valeur possible des objectifs  $\mathcal{M}_1$  et  $\mathcal{M}_2$ .**

Les notions liées aux solutions du problème d'optimisation multi-objectif ayant été brièvement introduites, les méthodes utilisées pour résoudre ce type de problème sont présentées dans la section suivante et les critères pour stratégie adaptative sont abordés ensuite.

### 1. Algorithmes d'optimisation

Pour résoudre les problèmes d'optimisation multi-objectif, une première stratégie consiste à reformuler le problème en un problème mono-objectif avec ou sans contraintes. Cette reformulation est possible lorsque le concepteur arrive à définir une hiérarchie entre les différents objectifs. Ce choix intervient généralement lorsqu'on souhaite obtenir une unique solution au problème multi-objectif afin de ne pas devoir choisir parmi les solutions constituant le front de Pareto. Le rapport de Marler et Aora [83] fournit plusieurs méthodes de ce type. Par exemple la combinaison des objectifs par des réels positifs  $a_i$  tels que  $\sum_{i=1}^m a_i = 1$  constitue un critère global de type somme pondérée :

$$f(\mathbf{x}) = \sum_{i=1}^m a_i \mathcal{M}_i(\mathbf{x}) \quad \text{[III-35]}$$

Ce type de critère exige de remettre à la même échelles les différents objectifs et d'accorder plus de poids aux plus importants par le biais des coefficients  $a_i$ . Le désavantage de cette méthode est lié aux choix des valeurs des coefficients, qui dépend de l'expertise du concepteur. De plus, si on souhaite changer les pondérations, il est nécessaire de recommencer une optimisation. Une autre possibilité sans coefficients réside dans la méthode lexicographique. Elle consiste à résoudre itérativement des problèmes d'optimisation sous contrainte de type :

$$\min_{\mathbf{x} \in X} \mathcal{M}_i(\mathbf{x}) \text{ sous } \mathcal{M}_j(\mathbf{x}) < \mathcal{M}_j(\mathbf{x}^*), j \neq i \quad \text{[III-36]}$$

où  $\mathbf{x}^*$  est l'optimum du problème précédent lié au critère  $\mathcal{M}_j$ . Lors de la première itération, l'objectif le plus important est optimisé sans contrainte. Ensuite la formulation de l'équation [III-56] est utilisée à partir du résultat de la première itération. La difficulté de cette méthode est issue des contraintes qui deviennent très vite restrictives au fil des itérations. En revanche, cette procédure permet d'obtenir des solutions qui respectent la condition de Pareto-optimalité.

Les méthodes précédentes nécessitent des connaissances a priori sur les objectifs. Si le concepteur ne dispose pas de ces informations ou souhaite obtenir un ensemble de solutions constituant différents compromis, il est préférable de passer aux algorithmes multi-objectif globaux. Deux méthodes de ce type sont présentées dans la suite, la première étant NSGA-II, un algorithme génétique, très utilisé pour la résolution de problèmes multi-objectif et pouvant même comporter des contraintes. La deuxième méthode, MGDA, permet d'atteindre des points du front de Pareto à partir du gradient des objectifs.

a. NSGA-II

L'algorithme NSGA-II [81] (pour nondominated sorting genetic algorithm II) a été choisi ici car il est parmi les algorithmes génétiques les plus utilisés pour les problèmes multi-objectif. Comme toutes les méthodes fondées sur une évolution de population, il s'adapte plutôt bien aux différents problèmes à résoudre et permet d'obtenir un front de Pareto explorant correctement l'espace des objectifs. Par contre, il peut nécessiter beaucoup d'appels aux fonctions à optimiser, en fonction du nombre de points de sa population et du nombre de générations calculées. Son principe repose sur deux actions principales : la génération des enfants de la population initiale, suivie du classement avec sélection des meilleurs individus. L'algorithme est initialisé avec une population initiale de  $N_{MO}$  individus dans  $\mathbb{X}$ . Cette population sert de base pour générer une population dite « enfant » à partir d'opérations de croisement et de mutation. A partir de la combinaison de la population initiale et de la population enfant, un classement entre les individus est effectué selon la notion de dominance au sens de Pareto. Un nombre  $N_{MO}$  d'individus est alors sélectionné parmi les meilleurs. Ce processus est résumé sur la Figure III-15. Le processus génétique est ici appliqué dans une représentation « réelle » sur les valeurs des paramètres (contrairement à la représentation « binaire » parfois employée).

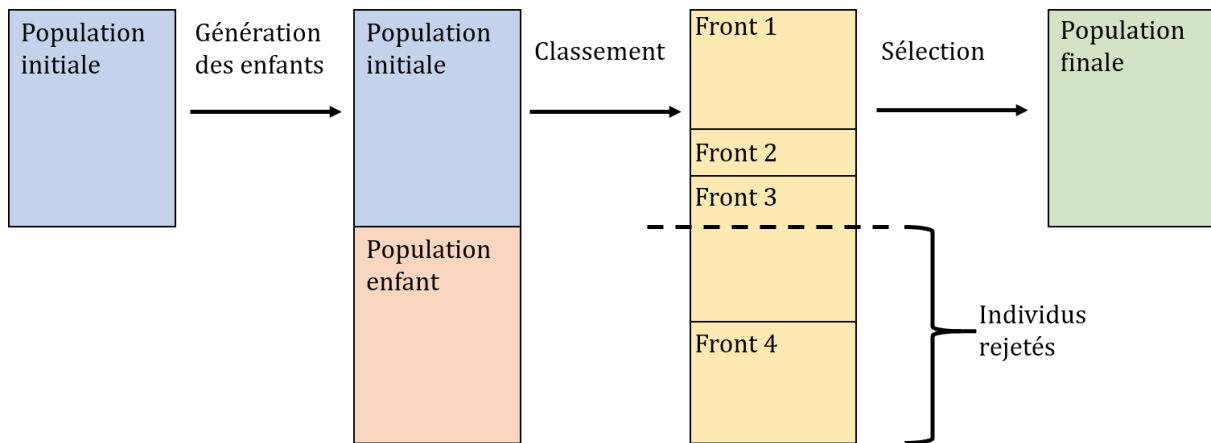


Figure III-15 : Schéma de principe de l'algorithme NSGA-II

Les différentes étapes sont maintenant détaillées. Tout d'abord la génération de la première population peut être réalisée à l'aide d'un plan d'expérience de type space-filling. Ici c'est un plan OLHS qui est utilisé. La génération des enfants commence par une opération de croisement, où certaines valeurs de paramètres entre deux individus sont échangées. Ici le croisement intermédiaire [84] est la méthode choisie et définit les valeurs des paramètres  $\mathbf{x}_e$  des enfants en fonction des paramètres  $\mathbf{x}_p$  des parents :

$$\begin{aligned} \mathbf{x}_{e_1} &= \mathbf{x}_{p_1} + r_c(\mathbf{x}_{p_1} - \mathbf{x}_{p_2}) \\ \mathbf{x}_{e_2} &= \mathbf{x}_{p_2} - r_c(\mathbf{x}_{p_1} - \mathbf{x}_{p_2}) \end{aligned} \quad \text{[III-37]}$$

où  $r_c$  correspond au ratio de croisement, tiré aléatoirement dans  $[0, 1.2]$ . Pour deux individus parents, il y a donc deux individus enfants générés. Sur l'ensemble des duos de parents, un tirage aléatoire est préalablement effectué avec  $1/d$  chances de réussite pour activer ou non l'opération de croisement (pour rappel,  $d$  est la dimension de l'espace des paramètres). Après le croisement, la population résultante est composée de parents inchangés et de nouveaux enfants. Cette population passe alors par une opération de mutation pouvant modifier aléatoirement la valeur des paramètres des individus. La mutation gaussienne [85] est utilisée ici et correspond à l'obtention de nouveaux individus  $\mathbf{x}_m$  à partir des individus  $\mathbf{x}_c$  issus du croisement :

$$\mathbf{x}_m = \mathbf{x}_c + \delta_g \quad \text{[III-38]}$$

où  $\delta_g$  correspond ici au tirage aléatoire d'un vecteur aléatoire gaussien de taille 1 par  $d$ , de moyenne nulle



et de variance correspondant à 10 % de l'intervalle de variation des paramètres. Comme pour le croisement, un tirage aléatoire est préalablement effectué avec  $1/d$  chances de réussite pour activer ou non l'opération de mutation. La population résultante constitue la population appelée enfant sur la Figure III-15. L'action de classement s'effectuant par la suite est la spécificité de NSGA-II. Elle consiste à classer la population en front de Pareto de rangs successifs. L'ensemble des premiers points non-dominés est classé dans le front 1 puis les individus le constituant sont retirés de la population. Un deuxième front de Pareto peut alors être obtenu car des individus qui étaient dominés avant le retrait du front 1 ne le sont plus. Cette procédure est réalisée jusqu'à ce que tous les points aient été classés par front de Pareto. Pour rappel, la population totale est constituée de  $2N_{MO}$  est seulement  $N_{MO}$  individus doivent être sélectionnés. Sur la Figure III-15, c'est à partir du front 3 que le nombre d'individus total dépasse le nombre de  $N_{MO}$ . Les individus des fronts de rang précédent sont alors automatiquement sélectionnés. Pour les individus du front 3, la distance de nichage [81] est utilisée pour les différencier. Cette mesure permet de repérer les points d'un même front qui sont les plus éloignés des autres points. Sélectionner ces points en priorité permet de conserver une forte diversité parmi les solutions du front de Pareto. Dans notre exemple, les points avec la distance de nichage la plus élevée dans le front 3 sont transférés dans la population finale jusqu'à obtenir un nombre total d'individus sélectionnés égal à  $N_{MO}$ . Tout ce processus représente une seule génération de NSGA-II. Il est itéré plusieurs fois, en utilisant la population finale de chaque génération comme la population initiale de la suivante. Le critère d'arrêt de l'algorithme est basé sur le maximal autorisé de génération.

### b. MGDA

L'algorithme MGDA (pour Multiple-Gradient Descent Algorithm) est apparu suite à la généralisation par Désidéri [86] de la méthode du gradient au cas de l'optimisation multi-objectif. Le principe est d'améliorer simultanément chacun des objectifs lors d'une même descente de gradient. Il repose sur la recherche d'une direction de descente commune aux objectifs permettant de réaliser cette optimisation simultanée. Par hypothèse, les fonctions objectifs sont localement convexes autour d'un point  $\mathbf{x}_0$ . En notant  $u_0^i = \nabla \mathcal{M}_i(\mathbf{x}_0)$ ,  $i = 1, \dots, m$  les gradients des fonctions objectif, l'ensemble des vecteurs suivants permet de définir une enveloppe convexe :

$$\bar{U} = \left\{ \mathbf{w} \in \mathbb{R}^n, \mathbf{w} = \sum_{i=1}^m \alpha_{g_i} u_0^i, \forall i \alpha_{g_i} \geq 0, \sum_i \alpha_{g_i} = 1 \right\} \quad \text{[III-39]}$$

$\bar{U}$  est fermée, bornée et convexe. De ce fait,  $\bar{U}$  admet un élément unique de norme minimale appelé ici  $\omega$  dont l'opposé est une direction de descente [86] (lemme 1 et 2). Si  $\omega$  est nul alors le point  $\mathbf{x}_0$  est un point Pareto stationnaire, une condition nécessaire de la Pareto-optimalité.

Dans le cas particulier où  $m = 2$ , l'élément de plus petite norme est connu analytiquement. Il s'obtient par la projection orthogonale du vecteur nul sur l'enveloppe convexe [87] :

$$\langle \alpha u_0^1 + (1 - \alpha) u_0^2, u_0^1 - u_0^2 \rangle = 0 \quad \text{[III-40]}$$

$$\text{D'où } \alpha = \frac{(u_0^2)^2 - u_0^1 u_0^2}{\|u_0^1 - u_0^2\|_2^2}$$

En tenant compte du cas où la projection orthogonale du vecteur nul est en dehors de l'enveloppe convexe, la solution générale pour deux objectifs est :

$$\alpha = \begin{cases} \frac{(u_0^2)^2 - u_0^1 u_0^2}{\|u_0^1 - u_0^2\|_2} & \text{si } u_0^1 u_0^2 < \min(\|u_0^1\|_2, \|u_0^2\|_2) \\ 0 \text{ ou } 1 & \text{selon si } \min(\|u_0^1\|_2, \|u_0^2\|_2) = \|u_0^1\|_2 \text{ ou } \|u_0^2\|_2 \end{cases} \quad \text{[III-41]}$$

Le cas à deux objectifs n'est pas représentatif de la méthode dans le cadre général. Lorsqu'un plus grand nombre de critères sont optimisés, Désidéri [86] propose de s'appuyer sur le processus d'orthogonalisation de Gram-Schmidt permettant de créer une base orthogonale du sous-espace engendré par les gradients. A l'issue de ce processus, les gradients forment une famille liée à la base construite, le but étant que le secteur

conique associé à l'enveloppe convexe de la nouvelle base soit le plus large possible. Dans le cas le plus favorable, une direction de descente est calculée explicitement par projection orthogonale de l'élément nul sur l'enveloppe convexe (si celle-ci appartient à l'enveloppe, sinon c'est le vecteur de la base ayant la plus petite norme). Si la direction obtenue n'est pas commune à tous les objectifs, il faut résoudre un problème de programmation quadratique formulé à partir de différentes simplifications. En effet une identification des gradients appartenant déjà à l'enveloppe convexe de la nouvelle base est possible, permettant de les exclure de l'analyse et de réduire la dimension du problème. Dans un second temps, les gradients résiduels en dehors l'enveloppe sont intégrés itérativement dans le problème, élargissant la famille de la base. La forme quadratique est alors minimisée dans des successions de domaines emboîtés qui deviennent de plus en plus grands permettant de trouver une direction commune à tous les gradients.

Une fois  $\omega$  obtenu, il suffit d'appliquer la méthode du gradient (voir section III.B.1.a) pour tous les objectifs. Les nombreux détails techniques de cet algorithme sont fournis dans le rapport de Désidéri [86], notamment la procédure de sélection des gradients pour l'orthogonalisation de Gram-Schmidt.

Cette méthode permet à partir d'un point de départ quelconque d'obtenir un point Pareto-optimal sous réserve de convexité locale des objectifs. Le front de Pareto global n'est pas directement fourni. Il faut donc effectuer plusieurs optimisations avec MGDA à partir de différents points pour obtenir un front. Des travaux sont en cours pour l'utilisation du continuum d'équilibre de Nash pour permettre l'exploration du front de Pareto à partir d'une solution de MGDA. Cette méthode a pour but de fournir un nouveau point de départ à MGDA sans s'écarter trop fortement du front de Pareto. L'information des points Pareto-optimaux déjà obtenus est ainsi utilisée plutôt que de relancer le calcul à partir de points initiaux quelconques.

### c. Extraire de l'information du front de Pareto

Des mesures de performances sont disponibles pour comparer deux fronts de Pareto obtenus sur le même problème multi-objectif avec des méthodes différentes. Cette section reprend les quatre mesures utilisées par Shimoyama et al. [88], avec tout d'abord la distance aux valeurs des points  $\mathbf{Y}^{PF^*}$  du front théorique :

$$Distance^* = \frac{1}{N_{PF^*}} \sum_{i=1}^{N_{PF^*}} \min_j \|\mathbf{Y}_i^{PF^*} - \mathbf{Y}_j^{PF}\|_2 \quad [\text{III-42}]$$

où  $N_{PF^*}$  est le nombre de points dans le front de Pareto théorique. Cette mesure impose donc d'avoir accès à un ensemble de solutions optimales du problème multi-objectif.

La deuxième mesure représente l'uniformité des solutions obtenues dans le front :

$$Espacement = \sqrt{\frac{1}{N_{PF} - 1} \sum_{i=1}^{N_{PF}} (\bar{d} - d_i)} \quad [\text{III-43}]$$

avec  $\bar{d} = \sum_{i=1}^{N_{PF}} d_i / N_{PF}$

où  $d_i$  représente la distance minimale de la solution  $i$  aux autres solutions du front de Pareto. Plus l'espacement est petit, plus le front est intéressant car il apporte une bonne continuité dans les différents compromis entre les objectifs.

La troisième mesure correspond à l'amplitude des valeurs de chaque objectif sur le front de Pareto :

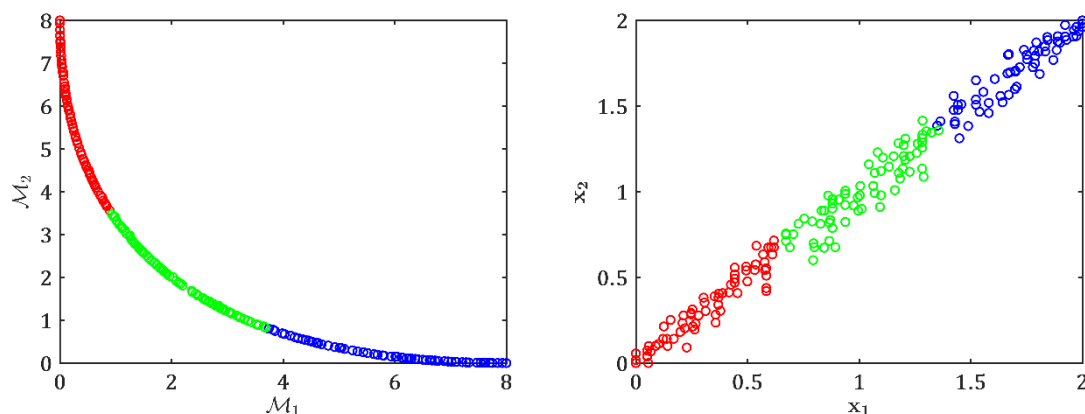
$$Etalement = \sum_{i=1}^m \left( \max_j \mathcal{M}_i(\mathbf{X}_j^{PF^*}) - \min_j \mathcal{M}_i(\mathbf{X}_j^{PF^*}) \right) \quad [\text{III-44}]$$

Plus l'étalement est grand, plus le front de Pareto est large et couvre l'espace des objectifs. Un petit espacement cumulé avec un grand étalement indique une très bonne couverture du front de Pareto et permet l'identification des solution « extrêmes ».

La dernière mesure consiste à mesurer l'hypervolume situé entre le front de Pareto et un point de référence, par exemple le point nadir de la Figure III-14. Si on dispose d'un front théorique, il est également possible de mesurer le ratio entre l'hypervolume du front obtenu et l'hypervolume du front théorique. Cette dernière mesure a des valeurs comprises entre 0 et 1, permettant de quantifier la précision des solutions obtenues (plus le ratio est proche de 1, plus le front est proche du résultat théorique).

Selon la méthode utilisée et la difficulté du problème multi-objectif, le front de Pareto peut contenir beaucoup de points ou avoir une forme difficilement interprétable. Il peut être intéressant dans ce cas d'utiliser des outils permettant de simplifier le résultat et d'extraire une information exploitable par le concepteur. Les méthodes de classification de données sont très utilisées pour analyser les fronts de Pareto. Elles consistent à regrouper les points qui contiennent des similitudes, souvent basées sur des notions de distance entre deux données. L'un des algorithmes les plus utilisés, car assez simple d'utilisation, est le partitionnement en k-moyennes [89]. Il consiste à découper l'ensemble des solutions en  $k_M$  classes. Lors de l'initialisation de l'algorithme, on tire aléatoirement  $k_M$  points, appelés centroïdes, représentant le centre de chaque classe. Dans l'espace des paramètres, les distances entre les points  $X^{PF}$  et tous les centroïdes sont calculées. Chaque point de  $X^{PF}$  est alors associé au centroïde le plus proche, formant ainsi  $k_M$  classes de données. Le calcul du barycentre de chaque classe définit la nouvelle position des centroïdes. Le processus est alors réitéré avec les opérations de classement et de redéfinition des centroïdes jusqu'à convergence. Pour s'affranchir du caractère aléatoire du résultat, l'algorithme peut être relancé plusieurs fois sur les mêmes données pour avoir un classement moyenné des données. On peut alors résumer le front de Pareto aux  $k_M$  points de  $X^{PF}$  les plus proches de chaque centroïde. L'analyse des classes permet d'établir quels sont les paramètres qui conduisent à des solutions proches en termes de performance. On remarque bien sûr que le résultat du classement dépend fortement du nombre  $k_M$  de centroïdes recherchés, ce qui conduit naturellement à effectuer une succession de partition des mêmes données en augmentant  $k_M$  à chaque fois. Le but est alors de détecter le nombre de centroïdes à partir duquel la distance moyenne entre les points  $X^{PF}$  et les centroïdes n'augmente plus significativement.

Un exemple de classement de données issues d'une optimisation multi-objectif est visible sur la Figure III-16. Le choix de trois classes pour la séparation des données a été choisi arbitrairement. On observe bien une similitude dans les valeurs de paramètres selon les zones du front de Pareto dans l'espace des objectifs. La classe rouge par exemple contient les points avec une petite valeur de  $x_1$  et  $x_2$ .



**Figure III-16 : Front de Pareto solution d'un problème exemple à deux objectifs (à gauche). Les couleurs représentent le résultat du partitionnement par k-moyenne en trois classes. A droite, le résultat est présenté dans l'espace des paramètres, ici de dimension 2.**

Les méthodes d'optimisations multi-objectif décrites précédemment peuvent s'appliquer sur le modèle numérique ou le métamodèle. Le coût de NSGA-II en termes d'appels est relativement élevé. Comme expliqué en introduction, ce type de méthode utilisée directement sur le métamodèle peut être imprécis. Il est plus que conseillé d'utiliser les stratégies adaptatives permettant d'améliorer la précision du métamodèle autour des solutions Pareto-optimales. Dans le cas de la méthode MGDA, il peut être

intéressant de travailler directement sur le modèle numérique si ce dernier fournit conjointement le gradient des objectifs, par exemple par l'intermédiaire d'un calcul d'adjoint. Ce n'est pas le cas des simulateurs « boîte noire » utilisés dans cette étude.

## 2. Critères pour stratégie adaptative

Dans le cadre de la stratégie adaptative, le front de Pareto est calculé à partir des points, évalués sur le modèle numérique, qui composent les données d'entraînement des métamodèles. Le but est donc de trouver des points non-dominés par ceux qui ont déjà été évalués. Des méthodes existent pour l'optimisation multi-objectif assistée par RBF. Mais celles-ci permettent en fait d'utiliser n'importe quel type de métamodèle et ne sont pas uniquement utilisables avec les RBF. Dans l'algorithme GOMORS de Akhtar et Shoemaker [90], un algorithme évolutionnaire est utilisé pour fournir un premier ensemble de solutions à partir du métamodèle. La solution la plus éloignée du front actuel est alors utilisée dans une nouvelle optimisation multi-objectif mais sur une zone plus restreinte et centrée sur le point retenu précédemment. L'ensemble des solutions globales et locales sont ensuite regroupées pour choisir les points à évaluer sur le modèle numérique. Plusieurs procédures de sélection sont utilisées simultanément comme la méthode du tirage aléatoire, l'amélioration de l'hypervolume, l'augmentation de la distance euclidienne par rapport au front actuel dans l'espace des paramètres et des objectifs. Chaque procédure permet de retirer un certain nombre de candidats pour l'évaluation du modèle numérique. Le processus est itéré jusqu'à ce que le nombre d'évaluations maximal autorisé soit atteint. Pour tenir compte de contraintes dans le problème multi-objectif, Regis [91] propose une autre stratégie basée sur le tirage aléatoire de points sur le métamodèle selon une certaine loi de probabilité puis de sélection en deux étapes. Les points violant le moins possible la contrainte sont retenus. Parmi ces points, seuls les points non-dominés au sens des objectifs sont conservés. Une opération de sélection s'effectue ensuite, à partir de la distance entre les points déjà présents dans les données d'entraînement faisant partie du front de Pareto et les points précédemment conservés. Ces deux méthodes de stratégies adaptatives multi-objectif pour les RBF ne tiennent pas compte de la mesure de l'erreur de prédiction et ne sont pas spécifiquement développées pour ce type de métamodèle. Dans cette section, le choix a donc été fait de se concentrer sur les critères pour le krigeage car ils utilisent toute l'information fournie par la prédiction et son intervalle de confiance.

### a. Critères adaptatifs pour le krigeage

Les critères adaptatifs pour l'optimisation multi-objectif avec krigeage reposent sur la notion d'amélioration espérée (voir section III.B.2.b). Dans les premières études réalisées sur cette problématique, le principe de combinaison des objectifs, permettant de se ramener à un problème mono-objectif et ainsi d'utiliser directement la formulation du critère EI, est exploité. Par exemple, l'algorithme ParEGO [92] utilise la pondération de Tchebycheff avec une normalisation préalable des objectifs :

$$f_{\lambda_T}(\mathbf{x}) = \max_{i=1,\dots,m} (\lambda_{T_i} \mathcal{M}_i(\mathbf{x})) + 0.05 \sum_{i=1}^m \lambda_{T_i} \mathcal{M}_i(\mathbf{x}) \quad \text{[III-45]}$$

où le vecteur des poids  $\lambda_T$  est tiré aléatoirement parmi l'ensemble des vecteurs suivant :

$$\Lambda = \left\{ \lambda_T = (\lambda_{T_1}, \dots, \lambda_{T_m}), \sum_{i=1}^m \lambda_{T_i} = 1, \lambda_{T_i} = \frac{l_T}{s_T}, l_T \in \{0, \dots, s_T\} \right\} \quad \text{[III-46]}$$

Le paramètre  $s_T$  conditionne le nombre de vecteurs possible dans  $\Lambda$ . En prenant par exemple  $s_T = 5$  dans un problème à deux objectifs,  $\Lambda = \{(0,1), (0.2,0.8), (0.4,0.6), (0.6,0.4), (0.8,0.2), (1,0)\}$ . La méthode consiste alors à tirer aléatoirement à chaque itération un vecteur dans  $\Lambda$ , de métamodéliser la fonction correspondant à l'équation [III-45], puis d'appliquer la stratégie adaptive d'optimisation classique avec le critère EI. Le changement de poids selon les itérations assure d'obtenir différents compromis entre les objectifs.

Des études successives se sont intéressées à l'extension directe de la notion d'amélioration espérée au

problème multi-objectif. Considérons l'ensemble  $\mathbb{Y} = \{\mathbf{y} \in \mathbb{R}^m, \mathbf{y} \leq \mathbf{y}_{ref}\}$  des réalisations des objectifs limités par un point de référence tel que le sous-ensemble  $H_{x_{tr}} = \{\mathbf{y} \in \mathbb{Y}, \exists i \in n, \mathbf{y}_{tr}^{(i)} < \mathbf{y}\}$  correspond à un espace fini et dominé par les données d'entraînement. La notion d'amélioration multi-objectif peut alors naturellement être associée au gain du volume de  $H_{x_{tr}}$  lors de l'ajout d'un point aux données d'entraînement :

$$I_{MO}(\mathbf{x}) = |H_{\{x_{tr} \cup \mathbf{x}\}}| - |H_{x_{tr}}| \quad [\text{III-47}]$$

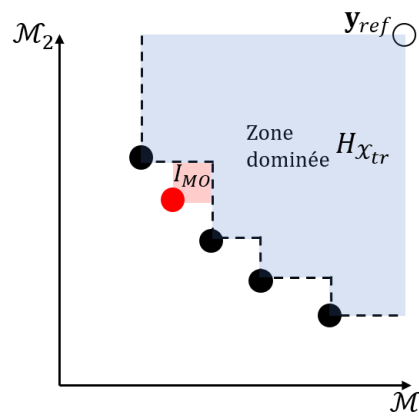
en notant  $|\cdot|$  la mesure de volume dans  $\mathbb{R}^m$ . Le critère EI pour le multi-objectif peut alors être décrit par l'intégrale suivante :

$$\text{EHVI}(\mathbf{x}) = \mathbb{E}[I_{MO}(\mathbf{x})] = \int_{\mathbb{Y} \setminus H_{x_{tr}}} \mathcal{P}(\hat{\mathbf{y}}(\mathbf{x}) < \mathbf{h}) d\mathbf{h} \quad [\text{III-48}]$$

Ce critère nommé « expected hypervolume improvement » est issu des travaux de Emmerich et al. [93], [94]. Pour simplifier ce critère, Svenson [95] montre que la corrélation entre les processus gaussien des différents objectifs peut être négligée, permettant de formuler une hypothèse d'indépendance entre les métamodèles. Le critère EHVI se reformule alors :

$$\text{EHVI}(\mathbf{x}) = \int_{\mathbb{Y} \setminus H_{x_{tr}}} \prod_{i=1}^m \Phi\left(\frac{h_i - \hat{y}_i(\mathbf{x})}{\hat{\sigma}_{y_i}(\mathbf{x})}\right) d\mathbf{h} \quad [\text{III-49}]$$

Le calcul de cette intégrale reste un problème difficile. La décomposition de domaine permettant de calculer analytiquement l'intégration de l'équation [I-16] est souvent choisie [96]. Elle présente le désavantage de devenir numériquement complexe lorsque le nombre d'objectifs à optimiser est important. Féliot et al. [97] proposent d'utiliser des méthodes de type Monte-Carlo séquentiel pour permettre de gérer plus d'objectifs. Par ailleurs, ils proposent l'utilisation de la boîte à outil STK [98], qui contient une relativement rapide implémentation du critère EHVI. Elle met en œuvre l'algorithme WFG [99], permettant d'obtenir une décomposition du domaine dominé par le front de Pareto qui facilite le calcul de l'intégrale. Cette implémentation est celle qui est retenue ici pour l'optimisation multi-objectif par critère EHVI. Sur la Figure III-17, une illustration du critère est disponible. La zone rouge représente l'amélioration espérée de l'hypervolume  $|H_{x_{tr}}|$  si le point rouge est ajouté aux données d'entraînement.



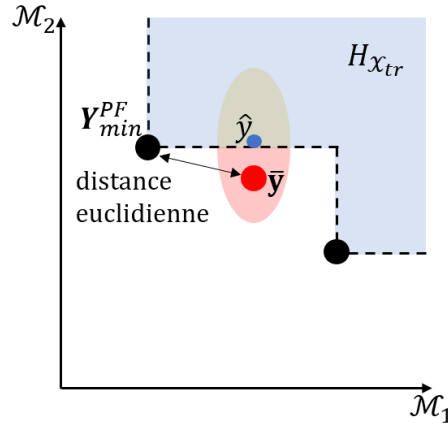
**Figure III-17 : Critère EHVI. Le nouveau point rouge permet une augmentation potentielle de  $|H_{x_{tr}}|$  d'un volume de volume  $I_{MO}$ , défini par la zone rouge.**

On trouve également d'autres définitions de l'amélioration que celle qui est définie par l'équation [III-47]. Par exemple, Forrester et Keane [100] cherchent à maximiser l'espérance de la distance euclidienne du point à ajouter avec le point le plus proche du front de Pareto actuel. Ils définissent leur critère comme :

$$EI_{euclid}(\mathbf{x}) = \mathcal{P}(\hat{\mathbf{y}}(\mathbf{x}) < \mathbf{Y}^{PF}) \times \|\bar{\mathbf{y}}(\mathbf{x}) - \mathbf{Y}_{min}^{PF}\|_2$$

$$\bar{\mathbf{y}}_i(\mathbf{x}) = \frac{\mathbb{E}[\hat{y}_i(\mathbf{x}) \leq \mathbf{Y}_i^{PF}]}{\mathcal{P}(\hat{\mathbf{y}}(\mathbf{x}) < \mathbf{Y}^{PF})}$$
[III-50]

où  $\mathbf{Y}_{min}^{PF}$  représente le point du front de Pareto le plus proche de  $\bar{\mathbf{y}}$ . Le critère  $EI_{euclid}$  est illustré sur la Figure III-18. La prédiction  $\hat{\mathbf{y}}$  (point bleu) est à la limite de la zone dominée. Son intervalle de confiance à 95 % est représenté par l'espace à fond coloré l'entourant, la partie jaune étant dominée. La zone rouge représente la partie non dominée, dont on calcule le centroïde  $\bar{\mathbf{y}}$  par l'équation [III-50]. Le point  $\mathbf{Y}_{min}^{PF}$  est alors déterminé comme le plus proche du centroïde et la distance euclidienne entre ces deux points correspond au critère à maximiser.



**Figure III-18 : Critère  $EI_{euclid}$ .** Le point bleu représente la valeur prédite par le krigeage en un point quelconque et l'ellipse l'entourant correspond à son intervalle de confiance à 95 %. La zone rouge représente alors la partie non dominée permettant d'obtenir le centroïde  $\bar{\mathbf{y}}$ . La distance euclidienne est calculée entre le point du front de Pareto le plus proche ( $\mathbf{Y}_{min}^{PF}$ ) et le centroïde  $\bar{\mathbf{y}}$ .

Les critères présentés ici permettent d'ajouter un seul point à chaque itération de la stratégie adaptative. Pourtant la complexité du problème multi-objectif peut amener plusieurs points distants à être intéressants à évaluer lors d'une même itération. De plus, selon la puissance des machines de calcul utilisées pour résoudre le modèle numérique, il est parfois possible de distribuer les calculs permettant d'avoir un coût en terme de temps qui est égal entre une ou plusieurs évaluations simultanées du modèle. Guerra ([101] – section 6.4) propose dans son mémoire de thèse une stratégie appelée MOEGO NSGA-II pour réaliser des optimisations multi-objectifs sur krigeage à partir des critères définis ci-dessus et permettant d'ajouter plusieurs points simultanément. La première étape consiste à utiliser l'algorithme NSGA-II sur les prédictions fournies par le krigeage pour chaque objectif. Le front de Pareto obtenu permet d'avoir une bonne première sélection de candidats. Ces solutions sont ensuite séparées en un nombre  $n_{MO}$  de classes correspondant au nombre d'évaluations simultanées autorisées au modèle numérique. L'algorithme des k-moyennes est utilisé ici pour réaliser cette étape. Les centroïdes des classes sont alors utilisés comme points de départ d'une optimisation locale du critère EHVI ou  $EI_{euclid}$ . Les  $n_{MO}$  points obtenus sont ensuite traités par un filtre de redondance. En effet, dans les premières itérations de l'algorithme, les critères d'amélioration espérée peuvent avoir un nombre de minimums locaux inférieur à  $n_{MO}$ . La variance de prédiction  $\hat{\sigma}_y$  du krigeage est alors utile pour vérifier si un point à ajouter est trop proche d'un autre, qui doit lui aussi être évalué. Tout d'abord, il est nécessaire de trier les points selon la valeur du critère adaptatif à la fin de l'optimisation locale. Le premier point est alors ajouté virtuellement au plan d'expérience. La variance de prédiction ne dépend pas de la valeur observée aux points d'entraînement, on peut donc leur attribuer virtuellement une valeur quelconque. La différence entre la variance initiale et la variance issue du plan d'entraînement avec les données virtuelles est calculée dans l'ordre des candidats précédemment triés. Si la variance initiale a été réduite de 90 %, le point n'est pas ajouté au plan d'entraînement virtuel et ne sera pas évalué. Si la variance n'a pas été trop réduite, le point est ajouté et la variance donnée à partir du plan d'entraînement virtuel est recalculée avant de faire l'analyse du prochain candidat de la liste.

L'ensemble des étapes du processus est répété jusqu'à ce que le nombre d'appels maximal au modèle numérique ait été atteint. Cette stratégie est celle qui est utilisée dans la suite de cette étude pour l'optimisation multi-objectif avec krigeage.

### b. Exemple analytique pour comparer les deux critères

A partir d'un problème d'optimisation à deux objectifs comportant deux paramètres, une partie des mesures présentées en section III.C.1.c sont utilisées pour comparer les critères EHVI,  $EI_{euclid}$  et l'optimisation par NSGA-II. Les fonctions analytiques d'étude sont :

$$\begin{aligned} \mathcal{M}_1(x_1, x_2) &= x_1^2 + x_2^2 \\ \mathcal{M}_2(x_1, x_2) &= (x_1 - 2)^2 + (x_2 - 2)^2 \end{aligned} \quad \text{[III-51]}$$

Pour l'optimisation directe du problème multi-objectif par NSGA-II, une population de 50 individus suivis au cours de 100 générations a été utilisée, soit 5000 appels aux deux fonctions. Les critères adaptatifs sont initialisés avec un plan d'expérience contenant 20 points et le critère d'arrêt est fixé à 40 appels additionnels aux deux fonctions. Le nombre de points évalués au cours d'une même itération de la procédure MOEGO est fixé arbitrairement à 5. L'ensemble des expériences est répété une vingtaine de fois avec des initialisations différentes pour moyenner les résultats, présentés sur la Figure III-19. Sur l'ensemble des essais, l'algorithme NSGA-II permet d'obtenir le meilleur front : la totalité de sa population finale comporte des individus non-dominés et bien répartis. Le critère EHVI obtient des résultats très proches sur presque toutes les mesures de performances. Le critère  $EI_{euclid}$  apporte globalement de moins bons résultats où seul l'étalement des valeurs qu'il produit est au niveau du critère EHVI. Ceci est dû au faible nombre de points composant en moyenne son front de Pareto final, qui ne permet pas d'avoir un hypervolume dominé intéressant et un faible espacement entre les solutions. Il faut rappeler que les résultats fournis par les critères EHVI et  $EI_{euclid}$  n'ont nécessité que 60 appels aux fonctions par rapport aux 5000 de NSGA-II.

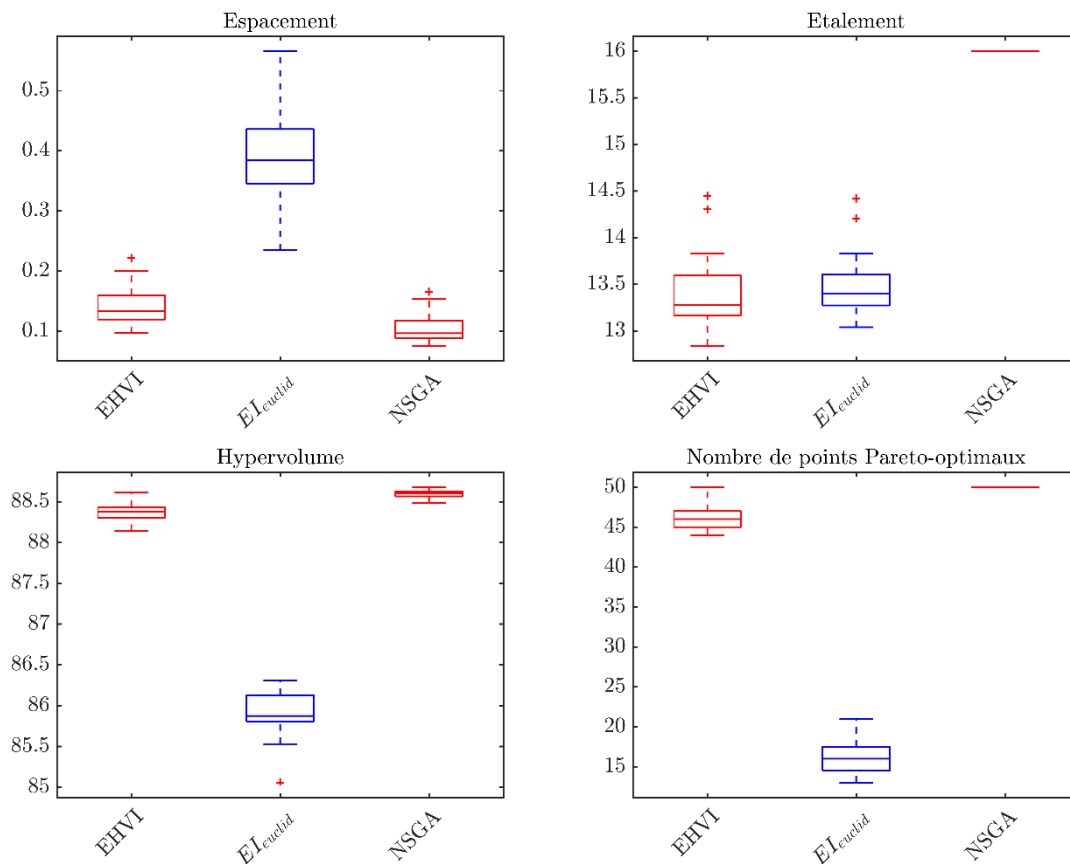


Figure III-19 : Résultat de l'optimisation multi-objectif des fonctions à l'équation [I-16] par critère EHVI ou  $EI_{euclid}$  et par l'algorithme NSGA-II. Les mesures d'étalement, d'hypervolume et le nombre de points Pareto-

**optimaux doivent avoir des valeurs les plus grandes possibles, contrairement à l'espacement qui doit être faible.**

Toutes les particularités du processus de conception liées au cahier des charges ne peuvent être prises seulement par des formulations de problèmes contraints ou multi-objectifs. En effet, une fois passée l'étape de dimensionnement par simulation numérique, le produit doit être réalisé et on retrouve parfois des écarts significatifs pour les performances du composant entre la géométrie préconisée par le modèle numérique et celle qui est fabriquée. La gestion de ces paramètres de conceptions associés à leurs incertitudes entre dans le cadre de l'optimisation robuste, étudiée dans la section suivante.

## D. Optimisation robuste

La simulation numérique des composants à concevoir permet d'obtenir des résultats reproductibles. Ces modèles peuvent comporter des variables qualifiées d'environnementales qui ont une influence sur le modèle (par exemple la pression ou la température lors de l'étude) mais qui ne font pas partie des variables à optimiser. Par rapport à une expérience réelle, ce type de paramètres est totalement maîtrisé en modélisation, d'où la possibilité de reproduire exactement les résultats obtenus. Même les modèles considérés comme « aléatoires » (par exemple le déplacement d'une particule dans un fluide) peuvent être contrôlés grâce à des méthodes de génération de nombres pseudo-aléatoires tel que Mersenne twister [102]. Pour autant, tous les modèles numériques ne sont qu'une approximation de la réalité et on ne peut pas être sûr que l'optimum obtenu est le vrai optimum si on cherche à reproduire l'expérience simulée. De plus, des incertitudes de fabrication peuvent empêcher de produire la solution définie comme étant optimale. La précision du procédé de fabrication nécessaire pour concevoir l'objet se trouve parfois être trop coûteuse pour être réalisable ou simplement irréalisable techniquement. Durant une expérimentation, certains paramètres environnementaux peuvent évoluer dynamiquement mais c'est une valeur statique qui leur est attribué dans le modèle numérique (toujours l'exemple de la pression ou de la température). Afin de trouver le véritable optimum de toutes ces situations, on doit alors recourir à l'optimisation robuste.

Différentes méthodes existent pour résoudre les problèmes d'optimisation robuste selon leur type. Une analyse des différents cas d'étude présentés au chapitre I permet de cibler les besoins et d'adapter le développement des méthodes de résolution. La fabrication des composants nanophotoniques repose sur différentes étapes technologiques permettant la mise en forme du wafer (principalement de sa face supérieur). Sur une plaque, le même composant peut être reproduit un grand nombre de fois. Une des étapes principales, la photolithographie (voir section I.A.3), permet l'impression de forme sur la plaque à partir d'un masque décrivant les motifs souhaités. Une résine, déposée initialement sur la plaque, est insolée par un rayonnement ultraviolet à travers le masque qui peut être vu comme une forme de pochoir. Selon le temps d'exposition, le motif imprimé peut se retrouver dilaté ou réduit par rapport à la forme souhaitée. Un mauvais alignement du masque peut également conduire à un décalage dans l'impression des motifs. Parmi les autres étapes clés, le dépôt de couches de différents matériaux peut également être soumis à des aléas en fabrication. L'épaisseur des couches dépend en partie du temps accordé au dépôt qui peut varier selon les conditions de température et de pression. L'impact des incertitudes en lithogravure et lors du dépôt dépend de la position du composant sur une même plaque. On observe généralement une plus grande variation des procédés sur le bord de la plaque. Le centre du wafer a plus de chance de correspondre aux spécifications souhaitées.

Ces deux situations révèlent des incertitudes qui peuvent s'appliquer sur des variables de conception ou environnementales. La géométrie désirée doit être celle qui assurera des performances les plus stables en tenant compte des aléas de fabrication. Pour un procédé mature, la densité de probabilité décrivant l'aléa de fabrication est connue. Dans le cadre d'un laboratoire de recherche, les composants ne sont pas produits suffisamment de fois pour obtenir une estimation fiable de l'incertitude sur la géométrie demandée. Notre expérience montre que l'ordre de grandeur de l'écart entre les dimensions cibles et celles qui sont effectivement obtenues à l'issue des procédés de lithogravure est d'environ plus ou moins 10 %. Le choix des méthodes d'optimisations robustes est orienté par les problématiques identifiées sur nos cas d'études. Tout d'abord, les premiers concepts de l'optimisation robuste sont détaillés avant d'aborder les méthodes



choisies pour l'optimisation robuste de composants photoniques.

## 1. Introduction

Le problème d'optimisation sous contraintes dans un cadre déterministe classique se présente de la forme suivante :

$$\min_{\mathbf{x} \in \mathbb{R}^n} \mathcal{M}(\mathbf{x}, \mathbf{z}) \text{ sous } \mathbf{g}(\mathbf{x}, \mathbf{z}) \leq 0 \quad [\text{III-52}]$$

où  $\mathcal{M}$  représente la fonction objectif à optimiser sous un nombre  $m_g$  de contraintes  $g$  à satisfaire. Ces fonctions dépendent de variables de conceptions  $\mathbf{x}$  et de variables environnementales  $\mathbf{z}$ . Les variables de conceptions sont celles dont les valeurs doivent être déterminées par l'algorithme d'optimisation, elles sont choisies initialement lors de la formulation du problème. Les variables environnementales définissent les conditions de fonctionnement du système, elles sont utiles à la modélisation du système mais ne sont pas optimisées. Les premières étapes de l'analyse de la robustesse d'une conception sont l'identification de l'origine des incertitudes et la définition d'une mesure de la robustesse.

### a. Les types d'incertitudes

Quatre types d'incertitudes sont distingués dans l'article de synthèse de Beyer et Sendhoff [103]. Ils sont détaillés ci-dessous, avec la perturbation liée aux incertitudes représentée par une variable aléatoire  $\delta$  :

- Type A – Changement des conditions d'opérations et environnementales. L'incertitude se situe au niveau des variables non-optimisée mais qui influent sur le système, à savoir les variables environnementales.

$$\mathcal{M}(\mathbf{x}, \mathbf{z}(\delta))$$

- Type B – Erreurs de fabrication. La réalisation du système est soumise à des incertitudes de fabrication qui ne permettent pas forcément d'obtenir la géométrie souhaitée. Ce sont donc les variables d'optimisations qui sont perturbées.

$$\mathcal{M}(\mathbf{x}(\delta), \mathbf{z})$$

- Type C – Mesure de la sortie bruitée. Ce type d'incertitude permet de prendre en compte un bruit lié à la mesure de l'objectif à optimiser ou des erreurs ou approximations concernant le modèle du système.

$$\tilde{\mathcal{M}}(\mathcal{M}(\mathbf{x}(\delta), \mathbf{z}))$$

- Type D – Satisfaction des contraintes sous incertitudes. Ce dernier type correspond à l'application des précédents sur les contraintes au lieu des objectifs. Il ne s'agit donc pas vraiment d'un nouveau type.

Au vu des cas d'études disponibles, les types A et B sont ceux qui seront le plus rencontrés. Ces différents types d'incertitudes peuvent être modélisés de différentes façons :

- Le point de vue déterministe définit un intervalle dans lequel le paramètre soumis aux incertitudes varie, sans disposer d'informations statistiques sur son comportement.
- Le point de vue probabiliste attribue une loi de variation à l'aléa, telle qu'une densité de probabilité.
- La vision possibiliste utilise la logique floue pour savoir si un évènement est possible ou non, par exemple à partir de d'informations partielles sur le comportement de l'aléa (connaissance de la variance mais pas de la distribution).

La méthode déterministe avec la théorie des intervalles [104] permet de traiter les incertitudes lorsque l'utilisation de méthode probabiliste est risquée (pas assez d'information sur la distribution de l'aléa). Historiquement, les premiers éléments permettant de traiter l'incertitude ont été développés par Kolmogorov [105]. Avec ces outils, la méthode probabiliste reste la plus utilisée traditionnellement pour modéliser l'aléa. Dans son mémoire de thèse, Chen [106] indique qu'il est préférable de n'utiliser des méthodes probabilistes que lorsque l'information sur la loi de distribution de l'incertitude est certaine. Ses

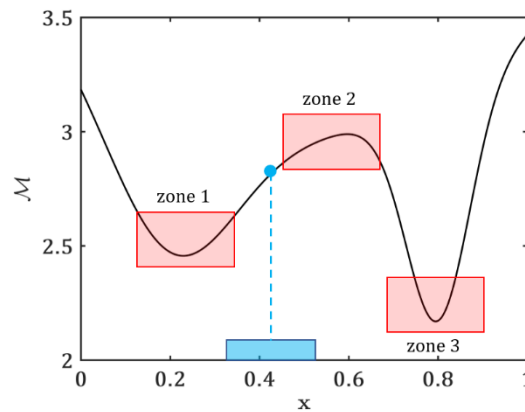
résultats sont issus d'une comparaison entre les méthodes possibilistes et les méthodes probabilistes. Les méthodes possibilistes sont apparues avec la théorie de l'évidence [107], [108]. Si l'information est disponible, le bon compromis entre représentation rapide et qualité de la modélisation se trouve dans l'utilisation de lois de distributions. Dans ce document, les lois de variations de l'aléa ne sont pas toujours connues et la méthode à utiliser dépendra des données disponibles sur le cas d'étude.

### b. Mesures de robustesse

Les incertitudes étant quantifiées, il faut maintenant définir les mesures permettant de comparer la robustesse de différentes solutions. Ainsi l'optimisation de la sortie du code de calcul est remplacée par l'optimisation de la mesure de robustesse. La distribution de la sortie peut être caractérisée et ses éléments peuvent être utilisés pour la définition de mesures  $\rho$ . Différentes mesures existent et sont listées dans l'article de Beyer et Sendhoff [103], mais seules quatre d'entre elles sont détaillées ici. Cette liste s'adapte au cas de quantification déterministe et probabiliste. La fonction exemple à l'équation [I-16] est utilisée pour illustrer l'impact de chacune des mesures sur l'optimum du problème robuste, avec  $\mathbf{x} \in [0,1]$ . L'aléa pour cet exemple est de type B, uniforme avec une variance  $\sigma_\delta$  égale à 0.0033 et s'additionne aux valeurs du paramètre de la fonction.

$$\mathcal{M}_{robuste}(\mathbf{x}) = 5 - 0.7e^{-\frac{(x-0.2)^2}{0.04}} - 2e^{-(x-0.5)^2} - e^{-\frac{(x-0.8)^2}{0.01}} \quad \text{[III-53]}$$

La réponse de cette fonction est visible sur la Figure III-20, où l'intervalle de variation de  $\mathbf{x}$ , selon une loi uniforme, est représentée en un point quelconque par le rectangle bleu. Trois zones d'intérêt sont remarquables pour une future analyse : la zone 1 est un minimum local assez large, la zone 2 une partie de la fonction avec de faibles variations et la zone 3 correspond au minimum global du problème déterministe. Les mesures étudiées dans la suite ont leur optimum dans l'une de ces trois zones.



**Figure III-20 : Fonction exemple pour le calcul de mesure de robustesse. L'aléa uniforme en un point quelconque est représenté par le rectangle bleu indiquant l'intervalle des valeurs que peut prendre le paramètre incertain. Les trois zones d'intérêt concernant la robustesse sont délimitées par les rectangles rouges.**

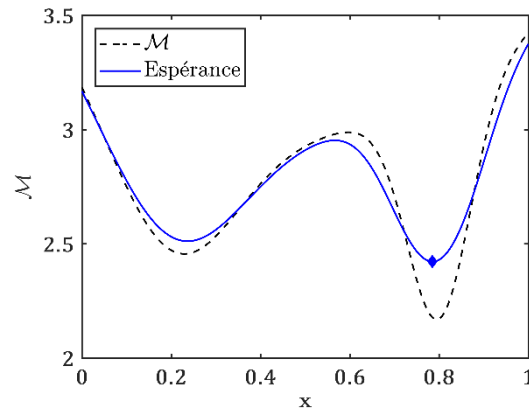
#### 1. *Espérance*

L'espérance est à la valeur moyenne d'un échantillon et correspond à une caractéristique de tendance centrale. Cette mesure de robustesse s'écrit :

$$\rho = \mathbb{E}[\mathcal{M}(\mathbf{x}, \mathbf{z}, \delta)] = \int \mathcal{M}(\mathbf{x}, \mathbf{z}) p_\delta d\delta \quad \text{[III-54]}$$

avec  $p_\delta$  la densité de probabilité de la variable aléatoire représentant l'incertitude. Cette mesure de la robustesse tend à réduire les fortes variations de la fonction objectif. L'espérance calculée sur la fonction exemple est visible sur la Figure III-21. L'optimum est légèrement décalé par rapport au problème déterministe, mais il reste dans la zone 3. Pour autant la valeur de l'espérance dans la zone 1 est proche de

celle de la zone 3.



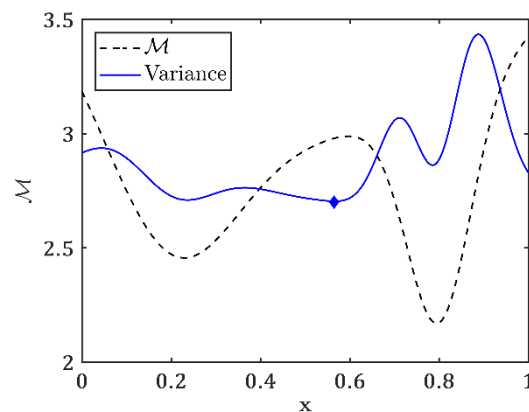
**Figure III-21 : Espérance de la fonction exemple pour une incertitude uniforme sur  $x$ . L'optimum de l'espérance est représenté par le losange bleu.**

### 1. Variance

La variance est l'écart quadratique moyen entre les valeurs d'un échantillon et la moyenne. Cette mesure correspond à une caractéristique de dispersion, c'est-à-dire l'étalement des valeurs observées. Elle se définit par :

$$\rho = \text{Var}[\mathcal{M}(\mathbf{x}, \mathbf{z}, \delta)] = \mathbb{E}[\mathcal{M}^2(\mathbf{x}, \mathbf{z}, \delta)] - \mathbb{E}^2[\mathcal{M}(\mathbf{x}, \mathbf{z}, \delta)] \quad [\text{III-55}]$$

Cette mesure prend des valeurs élevées lorsque des variations locales de la fonction objectif sont grandes. Notons que les performances d'une solution de forte variance peuvent être fortement dégradées. La variance calculée sur la fonction exemple est visible sur la Figure III-22. L'optimum se trouve dans la zone 2 qui avait été repérée par avance comme une zone où les variations de la fonction sont plus faibles. La zone 1 présente une plus faible variance que la zone 3. Au niveau de la performance, l'optimum de la variance est un maximum local de la fonction objectif et n'est donc pas un résultat de conception très intéressant au sens de l'objectif.



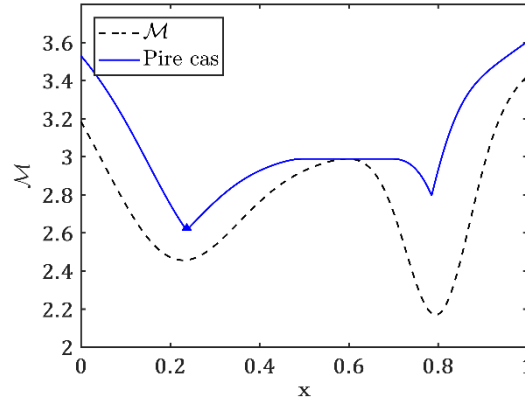
**Figure III-22 : Variance de la fonction exemple pour une incertitude uniforme sur  $x$ . L'optimum de la variance est représenté par le losange bleu.**

### 2. Pire-scénario

La mesure du pire scénario (ou pire cas, worst-case en anglais) est la valeur la plus défavorable observée au voisinage du point de calcul. C'est une mesure de performance qui est considérée comme très conservatrice car elle pénalise fortement les zones où quelques observations n'ont pas des valeurs intéressantes pour l'objectif. Cette mesure s'écrit :

$$\rho = \sup[\mathcal{M}(\mathbf{x}, \mathbf{z}, \delta)] = \max_{\delta} \mathcal{M}(\mathbf{x}, \mathbf{z}) \quad \text{[III-56]}$$

L'optimisation de ce critère revient à minimiser la valeur maximale locale de l'objectif soumis à l'incertitude. Le calcul de la mesure sur la fonction exemple est visible sur la Figure III-23. On observe que cette mesure n'est pas lisse car l'utilisation du max réduit la régularité. L'optimum se situe dans la zone 1 du fait de la faible variance des observations locale. La zone 3 est fortement pénalisée par les plus faibles valeurs de l'objectif.



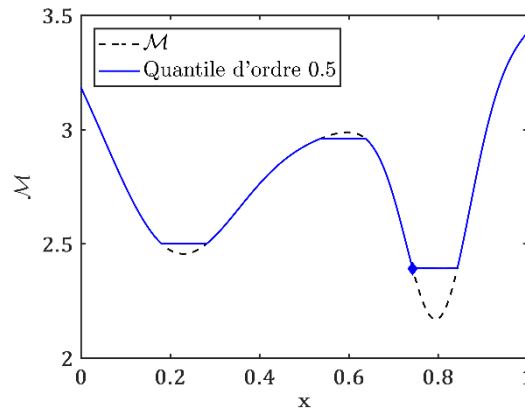
**Figure III-23 : Mesure du pire cas sur la fonction exemple pour une incertitude uniforme sur  $x$ . L'optimum du pire cas est représenté par le losange bleu.**

### 3. Quantiles

La mesure du pire scénario est en fait un cas particulier, issu de la définition générale des quantiles. Ces derniers sont des mesures de performances assurant qu'un certain pourcentage des observations est meilleur que la valeur du quantile. On définit le quantile d'ordre  $k \in [0,1]$  comme :

$$\rho = Q_k[\mathcal{M}] = \inf\{q \in \mathbb{R} : P(\mathcal{M}(\mathbf{x}, \mathbf{z}, \delta) \leq q) \geq k\} \quad \text{[III-57]}$$

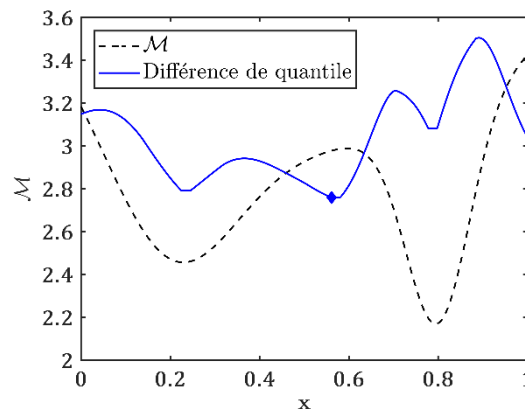
La mesure du pire scénario se retrouve pour  $k = 1$ . Pour avoir un exemple visuel, le quantile représentant la médiane ( $k = 0.5$ ) est visible sur la Figure III-24. L'utilisation de la médiane a pour effet de tronquer la fonction objectif sur certains intervalles : il assure que 50 % des observations ont une meilleure valeur que la sienne. Les zones de plats créées rendent difficiles l'obtention d'un minimum précis. L'intérêt des quantiles se situe dans le choix du degré de robustesse par l'intermédiaire de l'ordre. Par contre l'optimisation de cette mesure peut vite devenir impraticable à cause des zones de plats.



**Figure III-24 : Quantile d'ordre 0.5 sur la fonction exemple pour une incertitude uniforme sur  $x$ . Un des optimums du quantile est représenté par le losange bleu.**

Pour obtenir une mesure de dispersion à partir des quantiles, il est possible de calculer la différence entre

deux quantiles. Un exemple est présenté sur la Figure III-25, correspondant à la différence entre les quantiles 0.9 et 0.1. Le résultat obtenu est très proche de la variance, les optimums des deux mesures sont dans la zone 2. Ce critère reste moins lisse que la variance et présente de petites zones plates.



**Figure III-25 : Différences des quantiles 0.9 et 0.1 sur la fonction exemple pour une incertitude uniforme sur  $x$ . L'optimum de cette mesure de robustesse est représenté par le losange bleu.**

#### 4. Synthèse

Parmi les mesures présentées, aucune ne fait consensus. Selon la formulation de la robustesse choisie ou les propriétés désirées, la mesure la plus adaptée est sélectionnée. Comme guide lors de la sélection, on peut distinguer deux groupes catégorisant les mesures de robustesses. Il y a d'abord les critères concernant la performance des solutions obtenues où on retrouve l'espérance et les quantiles. Puis les critères mesurant la dispersion des solutions, avec la variance et la différence de quantiles. A partir des observations sur la fonction exemple, les mesures d'espérance et de variance offrent des critères plus lisses, donc plus faciles à optimiser. L'utilisation de ces deux dernières est d'ailleurs plus présente dans les problèmes d'optimisations robuste tandis que les mesures à base quantiles se retrouvent plus souvent dans les problèmes d'optimisations fiabilistes [109]. L'étude du pire scénario reste intéressante dans la mesure où l'information sur l'aléa se limite à un intervalle de variation. On peut ainsi effectuer une optimisation certes conservatrice mais qui justement permet d'assurer une certaine performance de la conception. La suite de cette étude utilise donc l'espérance, la variance et le pire scénario comme mesures de la robustesse.

#### c. Méthodes d'estimations des mesures

Les mesures de la robustesse ayant été définies, il faut maintenant établir un moyen de les calculer. Les méthodes d'estimation peuvent encore une fois être classées entre approches déterministes et approches stochastiques, selon les approximations qui sont effectuées lors du calcul. Dans leur analyse, Beyer et Sendhoff [103] définissent l'approche déterministe comme celle qui tend à ramener le problème d'optimisation robuste vers un problème d'optimisation déterministe classique (comme celui en section III.B). L'approche stochastique étant celle qui travaille directement sur des fonctions bruitées. Dans cette dernière approche, on retrouve l'utilisation d'algorithmes de gradient stochastique, évolutionnaires (similaire au CMA-ES en section III.B.1.b), ou de type pattern search qui ne seront pas plus détaillés ici. Le retour à un problème d'optimisation déterministe classique est une approche intéressante car elle permet de se ramener à l'utilisation d'outils déjà mis en place précédemment dans ce chapitre. Par contre cette méthodologie implique d'estimer précisément les mesures de robustesse choisies. Quelques méthodes d'estimations sont listées dans la suite, permettant d'obtenir une approximation des mesures.

##### 1. Monte-Carlo Brut (CMC pour Crude Monte Carlo)

Cette méthode présente l'avantage d'être rapide à mettre en place et se retrouve couramment lors du calcul d'une intégrale. Dans le cas où  $\mathcal{M}$  est un modèle boîte noire, l'intégrale de l'équation [III-54] ne peut être évaluée analytiquement. Un nombre  $M$  d'échantillons  $\mathbf{x}_{(1)} \mathbf{z}_{(1)}, \dots, \mathbf{x}_{(M)} \mathbf{z}_{(M)}$  indépendants sont générés

suivant la densité de probabilité de chaque variable. Ils sont ensuite évalués puis l'espérance et la variance sont approchées numériquement par :

$$\begin{aligned}\mathbb{E}[\mathcal{M}] &= \frac{1}{M} \sum_{i=1}^M \mathcal{M}(\mathbf{x}_{(i)}, \mathbf{z}_{(i)}) \\ \text{Var}[\mathcal{M}] &= \frac{1}{M-1} \sum_{i=1}^M (\mathcal{M}(\mathbf{x}_{(i)}, \mathbf{z}_{(i)}) - \mathbb{E}[f])^2\end{aligned}\tag{III-58}$$

D'après la loi faible des grands nombres, on sait que ces deux estimations convergent en probabilité vers leurs vraies valeurs. L'inconvénient de cette méthode est qu'elle requiert un grand nombre d'échantillons pour fournir une bonne estimation. Une règle communément utilisée en fiabilité sur l'estimation d'une probabilité  $P_f$  par CMC indique que la taille de l'échantillon doit au moins être  $N = 10^{2-\log(P_f)}$ . Pour connaître la précision de l'estimation, l'erreur relative  $RE$  peut être calculée en utilisant la variance de l'estimateur  $\sigma_{\mathbb{E}[f]}^2$  :

$$RE(\mathbb{E}[\mathcal{M}]) = \frac{\sigma_{\mathbb{E}[\mathcal{M}]}^2}{\mathbb{E}[\mathcal{M}]^2} = \frac{\sqrt{\mathbb{E}[\mathcal{M}] - \mathbb{E}[\mathcal{M}]^2}}{\mathbb{E}[\mathcal{M}]\sqrt{M}}\tag{III-59}$$

Le calcul d'intervalle de confiance peut également apporter une information sur la précision de l'estimation. A partir de la loi de probabilité de la variable aléatoire, des tables statistiques permettent de remonter à une valeur  $u_{\alpha/2}$  telle que pour une variable aléatoire  $U$ :

$$P(U \leq u_{\alpha/2}) = 1 - \frac{\alpha}{2} \Rightarrow P\left(-u_{\alpha/2} \leq U \leq u_{\alpha/2}\right) = 1 - \alpha\tag{III-60}$$

Par exemple, dans le cadre de la loi normale centrée réduite, pour un intervalle de confiance à 95 %,  $u_{\alpha/2} = 1.96$  ( $1 - \alpha = 0.95$ ). La formule générale de l'intervalle de confiance  $IC$  pour l'estimation de la moyenne revient alors à :

$$IC_{1-\alpha}(\mathbb{E}[\mathcal{M}]) = \left[ \mathbb{E}[\mathcal{M}] - u_{\alpha/2} \frac{\sigma_{\mathbb{E}[\mathcal{M}]}^2}{\sqrt{M}}; \mathbb{E}[\mathcal{M}] + u_{\alpha/2} \frac{\sigma_{\mathbb{E}[\mathcal{M}]}^2}{\sqrt{M}} \right]\tag{III-61}$$

## 2. Développement en série de Taylor

L'approximation locale en série de Taylor permet d'obtenir une estimation des moments statistiques. Cette méthode repose sur la connaissance du gradient de la fonction  $\mathcal{M}$ , voire du Hessien si on souhaite utiliser une approximation à l'ordre 2. Avec les métamodèles, cette information peut s'obtenir analytiquement ou par des méthodes numériques, telle que les différences finies. Par contre, l'approximation du premier ordre correspond au cas des fonctions presque linéaires. Pour les fonctions non-linéaires il faut augmenter l'ordre ce qui réduit également la précision de l'estimation [110]. Une dérivation détaillée du calcul aux deux premiers ordres est disponible dans l'étude de Beyer et Sendhoff [103], cette méthode ne sera pas utilisée ici car les cas d'applications sont complexes et ne satisfont pas toujours l'hypothèse de linéarité.

## 3. Polynômes de chaos

Le développement en polynômes de chaos (PCE, *Polynomial chaos expansion*), développé par Wiener [111], permet de construire un métamodèle local. Son principe repose sur une décomposition de la fonction d'intérêt sur une base polynomiale orthogonale  $\Psi$  :

$$\widehat{\mathcal{M}}(\mathbf{u}) = \sum_{j=0}^{\infty} a_j \Psi(\mathbf{u})\tag{III-62}$$

avec  $a_j$  les coefficients de la décomposition et  $\mathbf{u}$  une variable aléatoire. Il faut noter qu'en pratique le

développement est tronqué à un certain ordre. Cette représentation est très utilisée pour la propagation d'incertitudes [112] où la famille de polynôme utilisée est choisie en fonction de la distribution des variables aléatoires. Une synthèse sur les méthodes de construction de ce type de métamodèle est disponible dans le chapitre 3 du mémoire de thèse de Blatman [113]. L'intérêt de la PCE se situe au niveau de l'estimation des mesures de robustesse ou d'indices de sensibilités, qui sont directement fournis analytiquement à partir des composantes de l'expansion. Ne connaissant pas précisément les distributions des variables aléatoires de nos problèmes, cette méthode ne sera pas utilisée par la suite.

#### 4. *Calculs analytiques sur surface de réponse*

Le calcul des mesures de robustesses tel que défini pour la PCE peut s'appliquer en fait à d'autres types de métamodèles. Les formules analytiques peuvent être calculées lorsque la surface de réponse se décompose selon la forme suivante :

$$\widehat{\mathcal{M}} = a_0 + \sum_{i=1}^m a_i \prod_{j=1}^n h'_{ij} \quad \text{[III-63]}$$

où  $a_i$  ( $i = 0, \dots, m$ ) représente les différents coefficients de la décomposition et  $h'_{ij}$  des fonctions monovariées. Le krigeage, les RBF et les polynômes correspondent donc à ce cadre. On retrouve dans le mémoire de thèse de Baudouin [109] la formule analytique de l'espérance et de la variance calculées à partir de ces métamodèles pour différentes distributions. Les calculs qu'il a détaillés en annexe permettent même de prendre en compte l'erreur de prédiction dans le calcul des mesures à partir du krigeage. Par rapport à un calcul de type CMC effectué sur le métamodèle, ces formules analytiques sont plus rapides à évaluer. En revanche, l'hypothèse d'indépendance entre les variables aléatoires doit être admise.

#### 5. *Choix retenus pour cette étude*

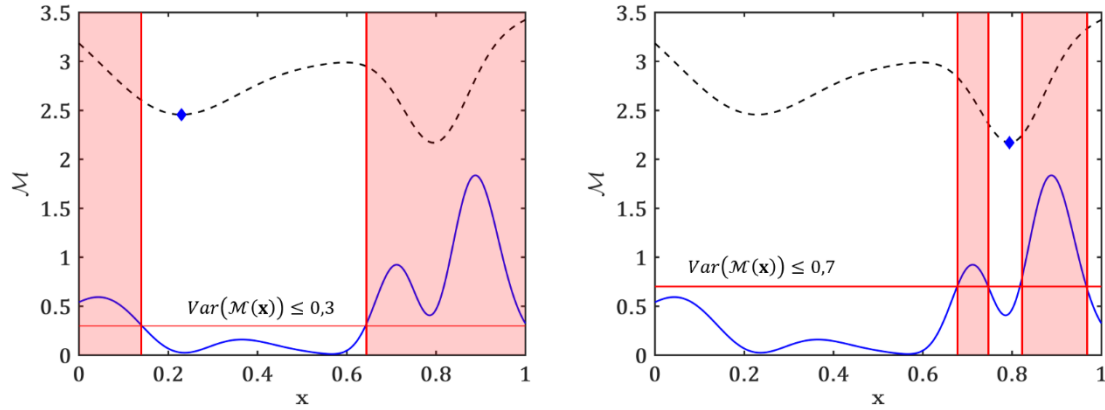
Dans les problèmes applicatifs liés à la photonique, l'incertitude entre plusieurs paramètres géométriques peut être corrélée. La fabrication par lithogravure peut produire un effet quasi uniforme de dilatation/rétrécissement du motif désiré. Si les paramètres d'optimisation appartiennent au même motif, leurs variations sont corrélées entre elles et le calcul analytique par surface de réponse des mesures de robustesse est impossible. Ainsi un calcul des mesures de robustesses à partir d'un calcul de Monte-Carlo, effectué sur le métamodèle est la méthode retenue. La surface de réponse est construite à partir de données incluant des variables de conception et des variables d'environnement incertaines. Le domaine de variation des paramètres d'environnement est défini en fonction des perturbations possibles du domaine de fonctionnement du dispositif.

##### d. Définition du problème d'optimisation

Les mesures permettant de qualifier la robustesse d'une solution sont maintenant connues et une méthode permettant de les estimer a été choisie. Il ne reste plus qu'à définir le problème d'optimisation afin d'obtenir une conception optimale au sens de la mesure de robustesse utilisée. Le but étant de se ramener à la même formulation que dans un cadre déterministe. On peut ainsi réutiliser les stratégies adaptatives développées pour les problèmes déterministes.

##### 1. *Utilisation directe des mesures pour l'objectif et/ou la contrainte*

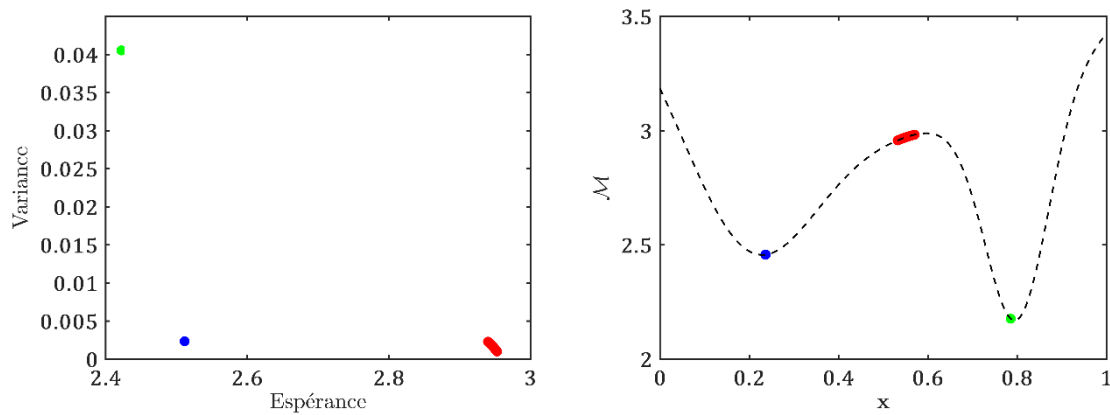
En première possibilité, la mesure de robustesse peut être directement optimisée, en la choisissant comme objectif. Si la mesure sert à vérifier la validité d'un seuil, on peut plutôt l'utiliser en tant que contrainte, avec le seuil comme limite de la zone admissible. Dans ce cas, l'objectif de la boîte noire est optimisé sous contrainte de robustesse. Un exemple de ce type de formulation est présenté sur la Figure III-26, où l'objectif du problème déterministe est optimisé sous contrainte que la variance de ce même objectif soit inférieure à 0.3 ou à 0.7. On remarque que selon le seuil choisi, l'optimum peut se situer dans la zone 1 ou la zone 3. Par ailleurs, le seuil choisi sur la variance peut conduire à des zones admissibles disjointes, ce qui rend le problème un peu plus délicat à résoudre.



**Figure III-26 : Optimisation de l'objectif (ligne pointillée) sous contrainte de robustesse (critère sur la variance en bleu). Le seuil de la contrainte est représenté par la ligne rouge horizontale, la zone admissible du domaine est la partie non colorée.**

### 2. Multi-objectif

Lorsque plusieurs critères sont intéressants à optimiser, on peut se placer dans le cas de l'optimisation multi-objectif de mesures de robustesse. On obtient alors un ensemble de solutions révélatrices de différents compromis entre les mesures choisies pour la formulation du problème. Cela peut permettre au concepteur de tester plusieurs solutions et d'analyser l'impact du compromis entre les critères sur la solution optimale. Le front de Pareto obtenu sur la fonction exemple est visible sur la Figure III-27 avec l'espérance et la variance comme critères. Les trois zones identifiées initialement sont atteintes selon le compromis entre les deux objectifs, on a donc bien une diversité dans les solutions obtenues. La solution traduisant le compromis le plus équilibré entre la variance et l'espérance correspond au point bleu, optimum de la zone 1.

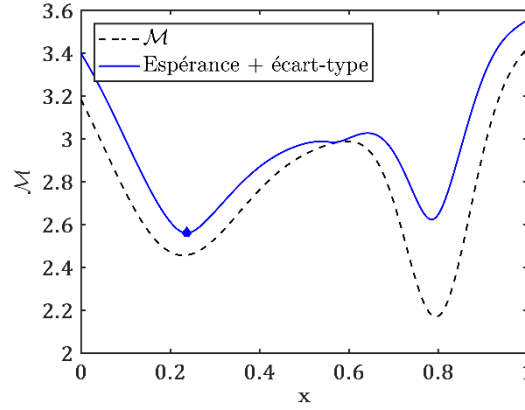


**Figure III-27 : Front de Pareto obtenu sur la fonction exemple avec l'espérance et la variance comme critères à minimiser (figure de gauche). Représentation des solutions de Pareto sur la fonction objectif (figure de droite).**

### 3. Agrégation de critères

En dernier lieu, nous considérons la possibilité d'agréger les critères, ce qui permet de simplifier le cas multi-objectif. Les objectifs sont regroupés en un seul par une sommation pondérée des critères. La pondération permet d'ajuster l'importance de chacun ou de les remettre à la même échelle. Cette formulation impose une connaissance préalable du compromis voulu entre les objectifs pour la solution optimale. Sur la fonction exemple, une agrégation possible est présentée sur la Figure III-28 où le combiné espérance et écart-type amène à un optimum dans la zone 1.





**Figure III-28 : Optimisation de critères de robustesse agrégés sur le cas de la fonction exemple. L'optimum de cette agrégation, ici la somme de l'espérance et de l'écart-type, est représenté par le losange bleu.**

## 2. L'optimisation robuste par stratégie adaptative

Dans le cadre de l'optimisation à base de surface de réponse, les stratégies adaptatives pour les problèmes déterministes s'appliquent aussi au cas de mesures de robustesse. Selon la définition du problème d'optimisation robuste (voir section III.D.1.d), il est nécessaire d'utiliser des stratégies adaptatives pour l'optimisation classique à un objectif, l'optimisation sous contraintes ou l'optimisation multi-objectif. En se focalisant sur les stratégies adaptatives pour le krigeage, le critère EI au centre de plusieurs méthodes se définit comme suit :

$$EI(\mathbf{x}) = (\rho_{\min} - \hat{\rho}(\mathbf{x}))\Phi\left(\frac{\rho_{\min} - \hat{\rho}(\mathbf{x})}{\hat{\sigma}_{\hat{\rho}}(\mathbf{x})}\right) + \hat{\sigma}_{\hat{\rho}}(\mathbf{x})\varphi\left(\frac{\rho_{\min} - \hat{\rho}(\mathbf{x})}{\hat{\sigma}_{\hat{\rho}}(\mathbf{x})}\right) \quad \text{[III-64]}$$

avec  $\hat{\rho}$  la mesure de robustesse à minimiser estimée sur métamodèle par CMC,  $\hat{\sigma}_{\hat{\rho}}$  l'écart type de prédiction de la mesure de robustesse et  $\rho_{\min}$  le minimum de la mesure de robustesse calculé à partir de la prédiction du métamodèle. Cette formulation a d'abord été utilisée pour une optimisation de l'espérance [114]. En choisissant le pire cas comme mesure de robustesse, Rehman et al. [115] proposent d'utiliser la variance du krigeage au point désigné comme le pire cas sur l'intervalle de variation lié aux incertitudes. Jurecka [116] utilise le calcul de la moyenne de la variance de prédiction du krigeage sur le même intervalle comme critère pour l'erreur de prédiction de la mesure de robustesse. Il propose également d'aller chercher la pire valeur de la variance sur l'intervalle mais cette méthode est jugée extrêmement conservatrice et peut amener à une convergence très lente du critère EI au fil des itérations. La condition de convergence du critère EI pour la robustesse est la même que pour l'optimisation classique. Lorsque celle-ci est atteinte, il convient de ne pas oublier que la valeur  $\rho_{\min}$  qui est retournée comme minimum de la mesure de robustesse est une valeur estimée par Monte-Carlo et dépend de la précision locale du métamodèle.

Lorsque des variables d'environnement sont également présentes dans le problème d'optimisation, une étape supplémentaire est nécessaire à l'algorithme. Ces variables ont une valeur nominale utilisée par le modèle numérique du problème. Si on enrichit la base d'apprentissage du métamodèle seulement pour ces valeurs, la précision de la prédiction pour d'autres valeurs sur l'intervalle dû à l'incertitude est mauvaise. Les mesures de robustesses calculées à partir du métamodèle sont donc peu précises. Il faut alors veiller à évaluer les variables environnementales sur tout l'intervalle lié à leur aléa. Après avoir choisi les valeurs des variables des conceptions  $x_{opt}$  par optimisation de l'EI, Williams et al. [114] proposent de maximiser la variance du krigeage par rapport aux variables d'environnement en ce point :

$$z_{opt} = \underset{\mathbf{z}}{\operatorname{argmax}} \hat{\sigma}_{\hat{y}}(x_{opt}, \mathbf{z}) \quad \text{[III-65]}$$

Dans leur cas les variables environnementales sont discrètes, mais on peut étendre la méthode aux variables continues. Le critère de distance par rapport aux points déjà présents dans les données d'entraînement est

proposé par Lehman et al [117] comme une alternative à l'équation [III-65]. Dans ce travail, la prise en compte d'une contrainte durant l'optimisation est effectuée par multiplication du critère EI par la probabilité de faisabilité du point d'évaluation. Dans la section 6.2.3 de son mémoire de thèse, Jurecka [116] propose un critère « d'expected worseness » qui peut être adapté lorsqu'on optimise la mesure du pire cas. Il s'agit d'une application du critère EI, pour la recherche d'un point qui permet de réduire l'erreur de prédiction de la mesure de robustesse mais également d'explorer les zones où le modèle numérique est peu évalué.

Pour la suite de ce travail, dans le cas où la mesure de robustesse est basée sur la variance ou l'espérance, la moyenne de la variance de krigeage est utilisée comme mesure d'incertitude de prédiction dans le critère EI. La valeur des variables environnementales pour le point d'évaluation est obtenue par minimisation de la variance du krigeage. Pour la mesure du pire cas, la différence concerne l'évaluation de la variance de prédiction de la mesure de robustesse : la méthode proposée par Rehman et. al [115], à savoir l'incertitude de prédiction du krigeage au point du pire cas, est employée.

## E. Conclusion

La formulation générale de la stratégie adaptative permet de répondre à différents problèmes liés à l'utilisation des métamodèles dans un outil de conception. Parmi les applications possibles, la réduction de l'erreur de prédiction, l'optimisation monocritère avec ou sans contrainte, l'optimisation multi-objectif et l'optimisation robuste ont été abordées. Pour chaque problème, des critères issus de développements mathématiques permettent d'utiliser en totalité ou en partie l'information apportée par le métamodèle. La majorité des critères étudiés sont liés au krigeage, même si des méthodes ont été développées ces dernières années pour les RBF. Les critères pour le krigeage prennent souvent l'avantage car ils utilisent l'intervalle de confiance de la prédiction pour favoriser l'exploration de l'espace des paramètres. Les méthodes RBF reposent plutôt sur des contraintes de distance entre les points du plan d'entraînement et les points à ajouter. Dans tous les cas, les stratégies adaptatives sont dépendantes de la qualité de la prédiction fournie par les métamodèles. Si les fonctions sont difficiles à estimer, les critères ont tendance à plutôt explorer et améliorer la précision globale de prédiction. La convergence peut alors être très lente et mener à un résultat insatisfaisant. Pour ce type de situation, le concepteur doit alors reformuler le problème ou le simplifier. Des paramètres peuvent être fixés ou des objectifs décomposés en plusieurs sous-objectifs. Dans ces cas particuliers, des critères spécifiques au problème peuvent être mis en place permettant d'apporter une solution plus satisfaisante ou avec un nombre d'appels plus faible que les critères « classiques » sur lesquels la nouvelle formulation s'appuie.

La suite de cette étude est dédiée à l'application des stratégies adaptatives dédiées à la conception optimale des composants présentés au chapitre I. Le premier problème concerne l'optimisation de l'efficacité sous contrainte de température maximale de la microsource de rayonnement infrarouge. Le deuxième problème aborde la proposition d'un métamodèle multifidélité pour mieux décrire la réponse de la cellule photoacoustique avant de l'optimiser. Le troisième et dernier problème concerne la mise en place d'une stratégie spécifique pour l'optimisation robuste du coupleur adiabatique en longueur d'onde.

---

## IV. Contribution méthodologique pour traiter les cas d'application

---

Dans les chapitres précédents, des méthodes permettant de réduire le coût des simulations en terme de temps de calcul ont été présentées. La construction d'un métamodèle interpolant les données issues des modèles numériques est une solution intéressante car elle permet d'obtenir une approximation de la réponse du composant en fonction de plusieurs paramètres de conception. Lorsqu'on utilise une stratégie adaptive spécifique pour rajouter des informations dans les données entraînant le métamodèle, il devient possible de réaliser différents types d'optimisation des grandeurs physiques mesurées en sortie de la simulation. Les métamodèles et stratégies adaptives décrites précédemment sont maintenant utilisés pour résoudre les problèmes de conception identifiés au chapitre I. Le traitement de ces cas d'application a nécessité deux développements méthodologiques originaux. Le premier concerne l'utilisation d'un métamodèle multifidélité RBF pour substituer les deux modèles numériques de la cellule photoacoustique. Le deuxième correspond à la mise en œuvre d'une stratégie adaptative spécifique permettant d'obtenir un coupleur robuste en longueur d'onde. Dans ce chapitre, l'optimisation de la microsource infrarouge est également présentée car elle a permis de distinguer une nouvelle géométrie non identifiée précédemment par les chercheurs du laboratoire et présentant un bilan énergétique intéressant.

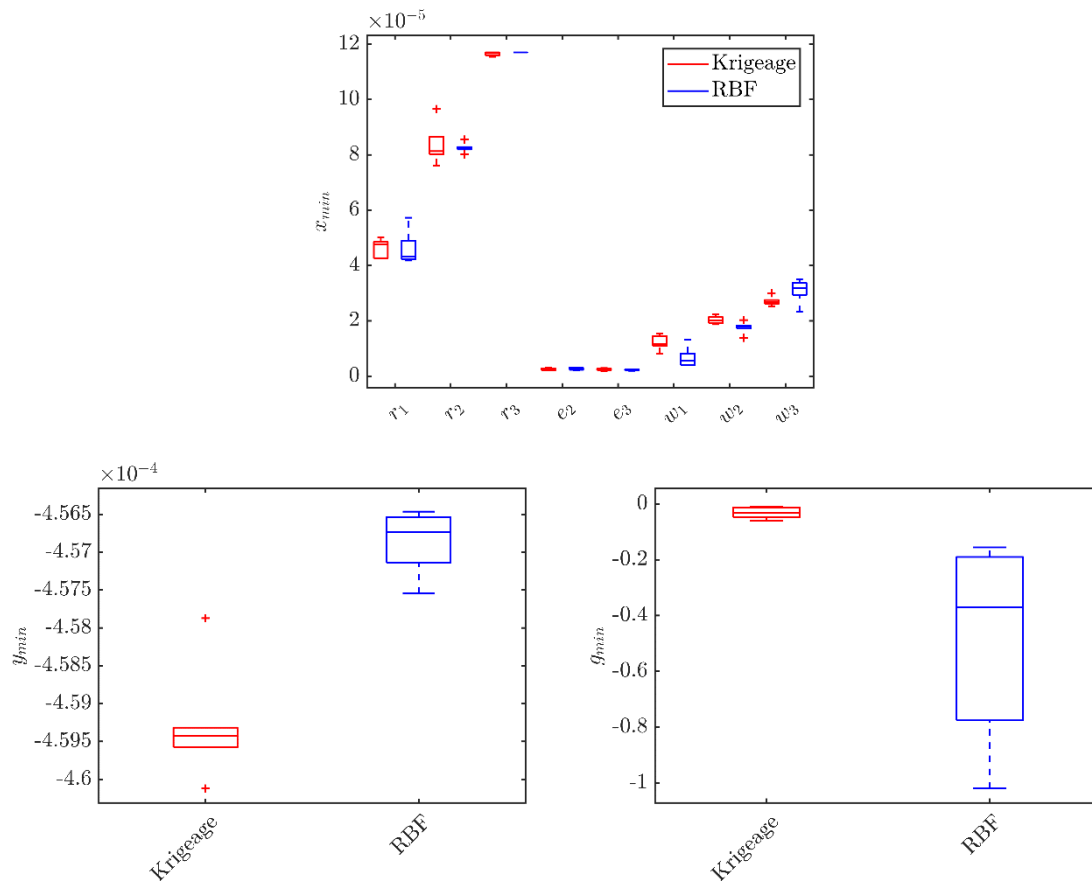
### A. Optimisation de la microsource infrarouge

Les spécificités liées à la conception de la microsource infrarouge sont détaillées en section I.C. Le but est de déterminer la position des pistes électriques et la tension pour obtenir une source optimale, soit 9 paramètres à régler. La source est considérée comme optimale si elle améliore les grandeurs physiques mesurées sur le composant tel que la température moyenne, la puissance optique rayonnée ou directement l'efficacité (traduite par le ratio entre la puissance rayonnée et la puissance électrique consommée). Dans la section II.D.1, une étude sur l'utilisation des métamodèles pour substituer les modèles numériques de la microsource montre qu'il est intéressant d'utiliser les RBF avec une fonction de base cubique, ou le krigeage avec une corrélation exponentielle. Ces métamodèles permettent de mieux prédire la réponse du composant en fonction des paramètres, avec une erreur de prédiction plus faible pour les RBF que pour le krigeage d'après les observations. Les deux métamodèles sont utilisés dans la suite pour l'optimisation sous contrainte par stratégie adaptive des différents objectifs. En effet, les trois objectifs cités précédemment doivent être optimisés mais tout en conservant une température maximale en dessous de 650 °C. A cette contrainte s'ajoute les limites du domaine admissible des paramètres décrites à l'équation [II.67], représentant des tolérances géométriques. Les stratégies adaptatives pour résoudre ce type de problème ont été décrites en section III.B.5, à savoir une optimisation multi-objectif des critères EI et PF pour le krigeage et l'optimisation sous contrainte du critère Gutmann pour les RBF. Ces deux méthodes sont employées pour résoudre l'optimisation de la microsource sur chacun des trois objectifs et les résultats sont comparés. Les mesures de comparaison sont les valeurs des paramètres, de l'objectif et de la contrainte à l'optimum ( $x_{min}$ ,  $y_{min}$  et  $g_{min}$ ) ainsi que la précision des métamodèles à la fin de l'optimisation. Le nombre maximal d'appels au modèle numérique est de 300.

#### 1. Optimisation de la puissance rayonnée

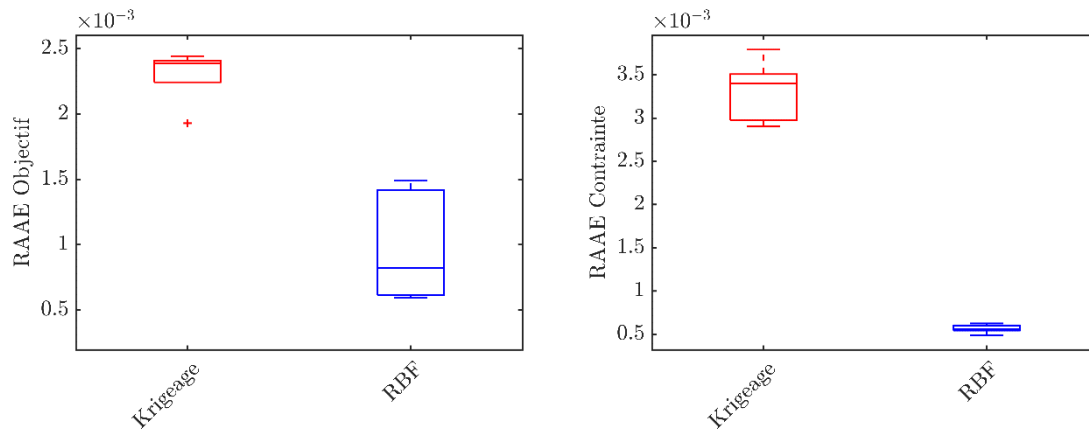
A partir de 6 plans d'entraînement composés de 100 points, des optimisations sous contraintes utilisant les deux métamodèles ont été lancées. Les éléments de comparaison des optimums obtenus selon les deux méthodes sont disponibles sur la Figure IV-1. Le krigeage permet d'obtenir une meilleure puissance rayonnée avec une médiane à  $4.59 \times 10^{-4}$  W contre  $4.57 \times 10^{-4}$  W pour la RBF. Ce résultat est lié au fait

que les points rajoutés sont plus proches de la limite de la contrainte correspondant à  $g(\mathbf{x}) = 0$ . La température maximale de la source est alors un peu plus élevée. La puissance rayonnée dépendant de la température à la puissance 4, l'écart de température est suffisant pour obtenir une valeur de l'objectif légèrement supérieure. Concernant la valeur des paramètres à l'optimum, les solutions apportées par les deux méthodes sont en moyenne assez similaires. Les rayons intérieurs des pistes sont quasiment les mêmes ainsi que leurs largeurs. La différence se situe au niveau de la largeur des pistes dans les bras de maintien. La tension appliquée aux bornes de la source n'est pas représentée par soucis d'échelle mais les valeurs sont du même ordre de grandeur pour l'ensemble des expériences.



**Figure IV-1 : Comparaison des optimums obtenus entre le krigeage et les RBF pour l'optimisation de la puissance radiée. Les valeurs de l'objectif sont négatives pour correspondre à une minimisation.**

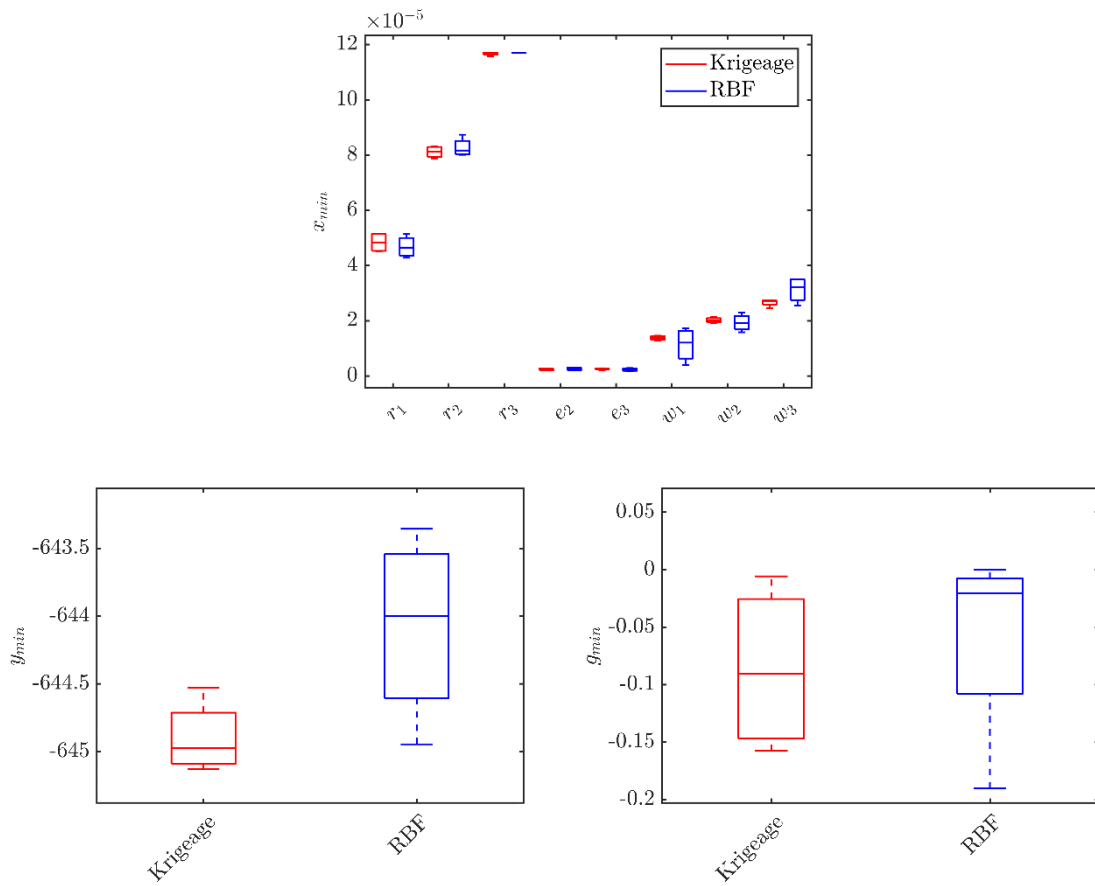
Si on s'intéresse à la précision des métamodèles, la Figure IV-2 montre que les résultats sont analogues à ceux de l'étude comparative du chapitre II. Les RBF sont plus précises que le krigeage sur l'ensemble des expériences, avec une valeur de RAAE plus faible sur les mêmes points de test à la fois pour la contrainte et l'objectif. Pour les deux métamodèles et sur tous les résultats, l'arrêt de la procédure des stratégies adaptatives s'est effectué car le nombre maximal d'appels au modèle numérique a été atteint. Au final, les plans d'entraînement sont composés de 400 points.



**Figure IV-2 : Capacité de prédiction des métamodèles une fois l'optimisation de la puissance rayonnée terminée.**

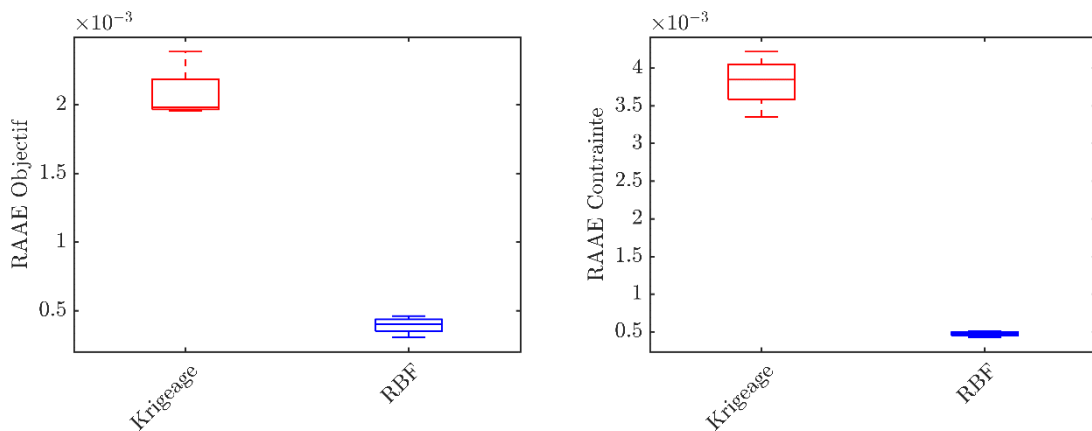
## 2. Optimisation de la température moyenne

Pour l'optimisation de la température moyenne, quatre plans d'expérience différents sont utilisés. Les éléments de comparaison des optimums obtenus selon les deux méthodes sont disponibles sur la Figure IV-3. Les positions des pistes électriques par krigeage ou RBF sont très proches, avec une différence plus marquée sur la largeur au niveau des bras de maintien (variables  $w_1$ ,  $w_2$  et  $w_3$ ). La valeur de l'objectif atteint avec le krigeage est par contre plus intéressante avec une valeur de 645 °C en médiane contre 644 °C pour les RBF. Pour autant, les valeurs de la contrainte à l'optimum sont proches pour les deux métamodèles.



**Figure IV-3 : Comparaison des optimums obtenus entre le krigeage et les RBF pour l'optimisation de la température moyenne sur le disque central. Les valeurs de l'objectif sont négatives pour correspondre à une minimisation.**

La Figure IV-4 compare la qualité de prédiction globale du krigeage et des RBF pour l'objectif et la contrainte à la fin de l'optimisation. Encore une fois, les RBF restent plus précises que le krigeage. Il faut toutefois noter que seuls 263 points en moyenne sont rajoutés par stratégie adaptive avec le krigeage contre 300 pour les RBF. Ces résultats sont en accord avec la comparaison effectuée en section II.D.1.

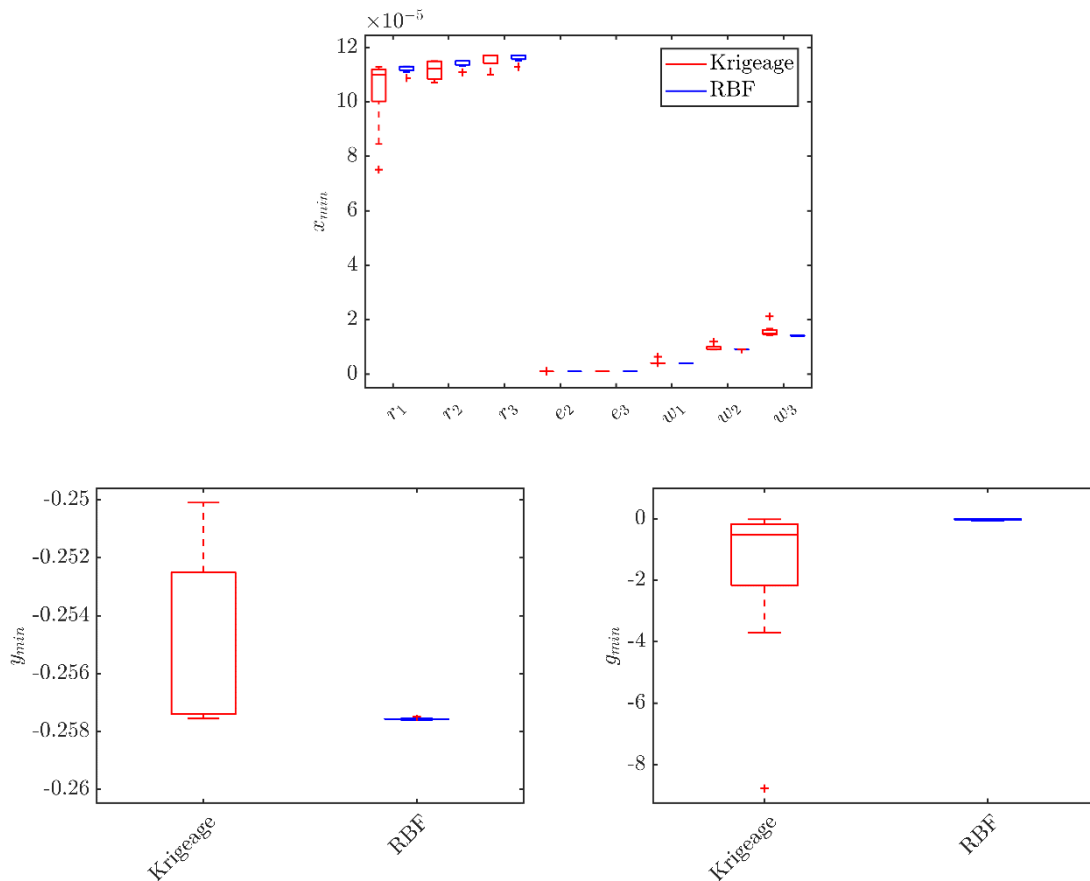


**Figure IV-4 : Capacité de prédiction des métamodèles une fois l'optimisation de la température moyenne terminée.**

### 3. Optimisation de l'efficacité

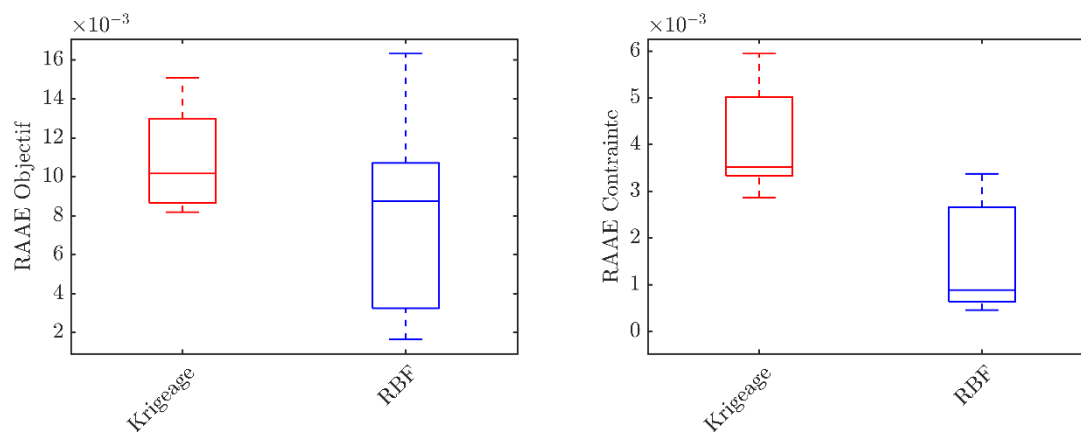
La dernière optimisation concerne l'efficacité de la microsource. Pour cet objectif, neuf plans d'entraînement différents sont utilisés. En observant les résultats présentés à la Figure IV-5, on remarque

que les paramètres optimaux ont bien plus de variabilité avec le krigeage qu'avec les RBF. Ceci s'observe également pour les valeurs optimales de l'objectif et de la contrainte. La stratégie adaptive par krigeage a convergé pour chaque expérience à cause de la valeur du critère. Le nombre d'appels au modèle numérique est en moyenne de 127, ce qui est faible comparé aux 300 points ajoutés avec les RBF. Pour autant, la valeur de l'objectif à l'optimum est la même en médiane pour les deux métamodèles, à savoir de 0.257. Deux expériences seulement ont convergé trop vite pour le krigeage. Sur le plan de la conception, les rayons intérieurs des trois pistes sont très proches (paramètres  $r_1, r_2$  et  $r_3$ ). Une microsource avec une seule piste électrique peut alors s'envisager comme une solution avec une meilleure efficacité. En effectuant ce changement de conception et en réalisant une optimisation de la position de l'unique piste, il est possible d'obtenir une efficacité proche de 0.31.



**Figure IV-5 : Comparaison des optimums obtenus entre le krigeage et les RBF pour l'optimisation de l'efficacité. Les valeurs de l'objectif sont négatives pour correspondre à une minimisation.**

Au niveau de la prédiction globale des métamodèles après optimisation, la Figure IV-6 montre que les RBF restent toujours plus précis mais il y a moins de différence avec le krigeage qu'auparavant.



**Figure IV-6 : Capacité de prédiction des métamodèles une fois l'optimisation de l'efficacité terminée.**

#### 4. Bilan

L'optimisation de la microsource selon trois différentes mesures physiques sous contrainte de température maximale a été réalisée. Deux stratégies adaptatives sont comparées, l'une utilisant le krigeage et l'autre les RBF. Les résultats montrent qu'il est possible d'obtenir un bon optimum par le biais des deux méthodes, avec un certain avantage pour le krigeage. Après optimisation, c'est pourtant la RBF qui permet de prédire la réponse du modèle numérique avec le plus de précision. Les valeurs de RAAE une fois l'optimisation terminée ont toutes diminué par rapport à celles calculées lors de l'étude en section II.D.1. On note ici que l'intérêt de la stratégie adaptative par krigeage réside dans sa condition d'arrêt qui a permis d'éviter des appels n'améliorant pas la recherche de l'optimum global.

D'un point de vue physique, il n'est pas possible de comparer directement les solutions obtenues ici avec une version existante de la microsource. Certaines données physiques telles que les propriétés matériaux diffèrent avec la précédente conception et rendent la comparaison impossible. Pour autant il est possible de comparer les solutions optimales entre elles. Le meilleur résultat entre les deux stratégies adaptatives pour chaque optimisation est disponible dans le Tableau 8.

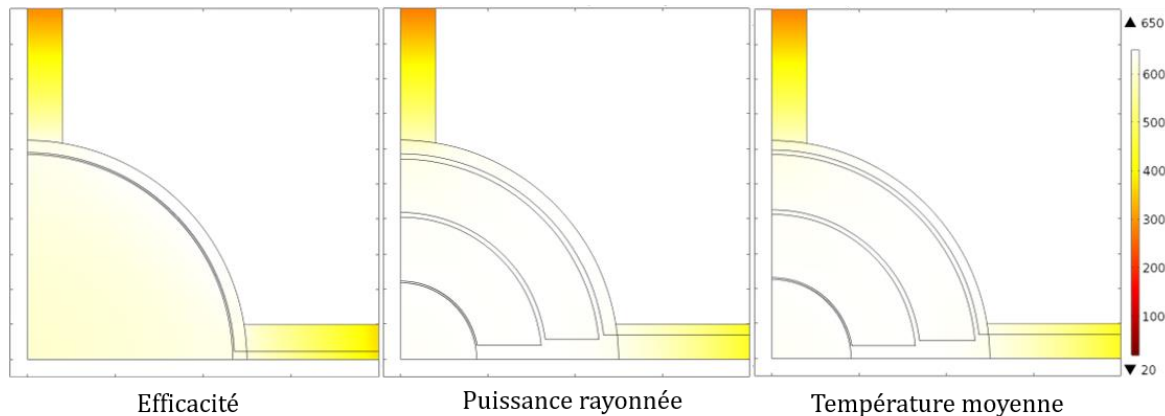
	Température moyenne (°C)	Puissance rayonnée (W)	Puissance électrique consommée (W)	Efficacité (%)
Puissance rayonnée optimisée	645.1	$4.6 \times 10^{-4}$	$2.6 \times 10^{-3}$	17.57
Température moyenne optimisée	645.1	$4.6 \times 10^{-4}$	$2.5 \times 10^{-3}$	18.09
Efficacité Optimisée (3 pistes)	628	$4.3 \times 10^{-4}$	$1.7 \times 10^{-3}$	25.74
Efficacité Optimisée (1 piste)	625	$4.2 \times 10^{-4}$	$1.3 \times 10^{-3}$	31.56

**Tableau 8 : Comparaison des différentes conceptions selon le critère physique optimisé.**

Les conceptions par puissance rayonnée ou température moyenne optimisée amènent à un résultat similaire. Les valeurs de température moyenne sont les mêmes avec une légère différence sur l'efficacité. Comme la puissance rayonnée dépend de la température à la puissance 4, il est compréhensible que son



optimisation tend à faire augmenter le plus possible la température moyenne. La conception optimale selon l'efficacité amène à réduire la puissance électrique consommée en réduisant légèrement la puissance rayonnée. Les trois pistes électriques sont positionnées très proches les unes des autres sur la géométrie optimale. Une version à une piste est alors envisagée et le résultat de son optimisation permet de gagner 6 % environ d'efficacité, c'est-à-dire de réduire davantage encore la consommation électrique. Les positions des pistes électriques apportant les meilleurs résultats pour chaque objectif sont visibles sur la Figure IV-7. Comme annoncé précédemment les optimums pour la température moyenne et la puissance rayonnée sont très proches.



**Figure IV-7 : Positions des pistes optimales selon chaque critère avec la cartographie de la température (°C) en fonctionnement.**

La nouvelle géométrie comportant une seule piste a été obtenue grâce à la gestion des contraintes sur la position des pistes (espace des paramètres non hypercubique) et de la contrainte de température maximale. La méthodologie employée sur les conceptions précédentes se basait sur un réglage automatique de la tension pour atteindre la température maximale en un point précis et ainsi éviter d'avoir cette contrainte lors de l'optimisation. De plus, chaque piste ne pouvait se déplacer que dans une partie restreinte du disque central pour éviter la prise en compte du chevauchement lors de l'exploration des différents jeux de paramètres possibles.

D'autres études ont été réalisées à partir de ce cas d'application mais ne sont pas détaillées ici. Plusieurs objectifs étant disponibles, différents scénarios d'optimisation multi-objectif ont été conduits. Les géométries optimales n'ont pas eu de grande différence avec les versions précédentes. Une optimisation robuste a également été effectuée pour voir l'impact des aléas de fabrication sur la géométrie. Le résultat montre que c'est essentiellement le paramètre de tension aux bornes de la piste qui est impacté pour s'assurer que même dans le pire des cas, la température maximale ne dépasse pas les 650 °C.

## **B. Métamodélisation multifidélité pour la cellule photoacoustique**

Pour modéliser le phénomène physique lié au fonctionnement de la cellule photoacoustique, deux simulations numériques sont disponibles. La première est une modélisation précise utilisant les équations de Navier-Stokes (modèle FLNS). La deuxième néglige les pertes thermiques et visqueuses aux parois de la cellule, pour les remplacer par des facteurs de qualité et se contenter de résoudre une équation d'Helmholtz (modèle Kreuzer). L'information provenant de ces deux modèles de fidélité différents (avec un coût en temps de calcul plus faible pour celui qui est le moins précis) peut être rassemblée en une seule information par l'intermédiaire de la métamodélisation multifidélité. Le principe est de construire un unique métamodèle à partir de nombreuses données du modèle LF et de quelques points du modèle HF. La prédiction est ainsi plus précise et moins coûteuse à construire que si seules des données HF avait été utilisées. Suite à la comparaison du krigeage et de la RBF appliqués aux modèles de la cellule en section II.D.2, le krigeage a été en moyenne plus précis pour le modèle LF et la RBF plus précise en moyenne pour le modèle HF. Dans cette section, il s'agit maintenant de comparer les performances de métamodèle

multifidélité. Pour le krigeage, la version multifidélité correspond au co-krigeage, détaillé en section II.C.1. Pour la RBF, il n'existe pas à notre connaissance de version multifidélité, à part les travaux de Sun et al. [118] qui utilisent un modèle RBF pour approximer un des deux niveaux de multifidélité. L'un des apports principaux de ce travail a donc consisté à proposer une méthode de RBF multifidélité avec une méthodologie similaire au co-krigeage. Cette nouvelle méthode est donc présentée ci-après, avant d'être comparée dans un second temps sur les données de la cellule photoacoustique.

### 1. RBF multifidélité (co-RBF)

L'approche multifidélité qui est décrite dans cette section peut s'appliquer à de multiples niveaux de fidélité mais pour faciliter la lecture, deux niveaux seulement sont utilisés ce qui correspond également au cas d'application traité. Les notations de la section II.C.1 sont réemployées ici, à savoir  $\mathcal{D}_{HF} = \{\mathcal{X}_{HF}, \mathcal{Y}_{HF}\}$  et  $\mathcal{D}_{LF} = \{\mathcal{X}_{LF}, \mathcal{Y}_{LF}\}$  se réfèrent aux plans d'entraînement des données LF et HF. L'hypothèse est faite que les points du modèle HF sont aussi évalués et présents dans le plan du modèle LF. Ceci permet l'utilisation du modèle autorégressif de Kennedy et O'Hagan [119] rappelé dans l'équation suivante :

$$y_{HF}(\mathbf{x}) = \rho_{LF}(\mathbf{x}) \times y_{LF}(\mathbf{x}) + y_D(\mathbf{x}) \quad [\text{IV-1}]$$

où  $\rho_{LF}$  est un facteur d'échelle et  $y_D$  un modèle de la différence entre les deux niveaux de fidélité après la mise à l'échelle du modèle LF. Dans le cadre de la RBF multifidélité et grâce à cette formulation, les modèles  $y_{LF}$  et  $y_D$  sont des RBF pouvant utiliser une fonction de base différente (par exemple cubique pour le modèle LF et gaussienne pour le modèle de différence). Pour estimer les paramètres du modèle RBF basse fidélité, les données LF sont d'abord considérées comme indépendantes des données HF, et conduit à l'estimation des paramètres classiques de la RBF avec la résolution du système décrit à l'équation [II.53]. Une fois les paramètres  $(\boldsymbol{\beta}_{LF} \ \boldsymbol{\alpha}_{LF})^T$  obtenus, il faut estimer les paramètres du modèle de différence. Ces paramètres sont calculés à partir des valeurs du modèle LF aux points du modèle HF, c'est-à-dire  $y_{LF}(\mathcal{X}_{HF})$ . Si les points  $\mathcal{X}_{HF}$  ne sont pas inclus dans  $\mathcal{X}_{LF}$ , la prédiction de la RBF du modèle LF venant juste d'être entraînée peut s'utiliser pour obtenir les valeurs  $\hat{y}_{LF}(\mathcal{X}_{HF})$ . Grâce à ces valeurs, le système linéaire suivant est alors à résoudre pour obtenir les paramètres  $(\boldsymbol{\beta}_D \ \boldsymbol{\alpha}_D)^T$  du modèle de différence :

$$\begin{pmatrix} \boldsymbol{\Phi}_{D \ tr} & \mathbf{F}_{D \ tr} \\ \mathbf{F}_{D \ tr}^T & \mathbf{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_D \\ \boldsymbol{\alpha}_D \end{pmatrix} = \begin{pmatrix} \mathbf{Y}_{HF} - \rho_{LF} \times \hat{y}_{LF}(\mathcal{X}_{HF}) \\ \mathbf{0} \end{pmatrix} \quad [\text{IV-2}]$$

où les matrices  $\boldsymbol{\Phi}_{D \ tr}$  et  $\mathbf{F}_{D \ tr}$  sont calculées à partir des données HF (voir section II.B.3). Ce système dépend du facteur d'échelle  $\rho_{LF}$ . Celui-ci est en fait estimé durant l'obtention des paramètres du modèle de différence. L'erreur de leave-one-out est utilisée ici (voir section II.B.3.c), ce qui revient à estimer ce paramètre en même temps que les hyperparamètres de la fonction de base du modèle de différence. Une fois ces paramètres obtenus et le métamodèle de différence construit, la prédiction du métamodèle multifidélité est calculée à partir de l'équation [IV-1]. Ce métamodèle est dénommé co-RBF dans la suite du document. Sur la Figure IV-8, la capacité de prédiction de la co-RBF est illustrée sur l'exemple de Forrester et al. ([120] – section 3.a). Le comportement du modèle  $\mathcal{M}$  est parfaitement reproduit grâce à l'utilisation de données basse et haute fidélité. L'estimation des paramètres de la co-RBF est à l'origine de ce résultat. Sur la Figure IV-8 de droite, les lignes de niveau de l'erreur LOO et son optimum sont représentées. Le facteur d'échelle estimé a pour valeur 2.07, ce qui est très proche de la valeur réelle de 2.

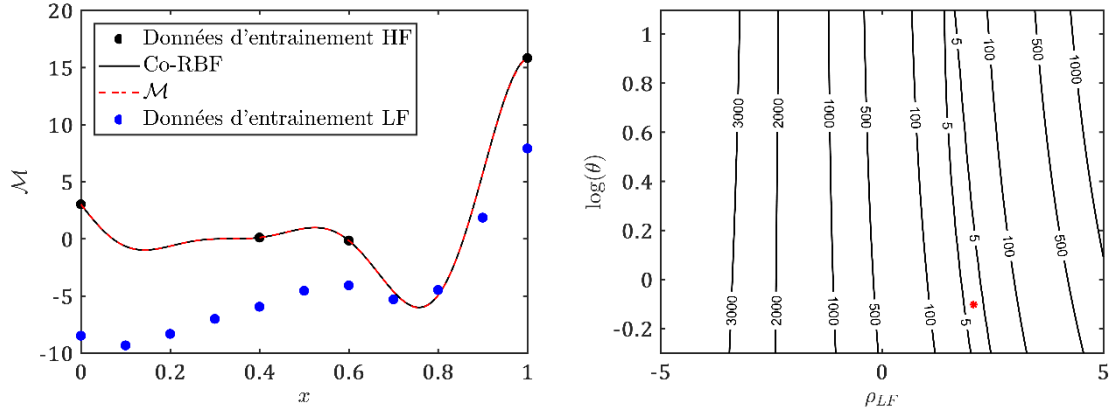


Figure IV-8 : Capacité de prédiction de la co-RBF sur l'exemple analytique de Forrester et al. (à gauche). Ligne de niveau de l'erreur LOO en fonction de l'hyperparamètre et du facteur d'échelle (à droite). L'optimum est représenté par l'étoile rouge.

## 2. Comparaison des métamodèles multifidélité sur modèles analytiques

Une première comparaison de la prédiction fournie par la co-RBF et celle du co-krigeage est effectuée sur deux exemples analytiques différents. Les problèmes sont issus des travaux de Xiong et al. [121], qu'ils utilisent pour illustrer l'utilisation de plans d'expérience séquentiels pour l'entraînement de métamodèles multifidélités. Pour chaque problème, la capacité de prédiction des métamodèles est comparée en calculant la RAAE sur des points de validation. Cette erreur est évaluée pour différentes combinaisons de nombre de points d'entraînement haute fidélité et basse fidélité. L'intérêt est de pouvoir observer si l'augmentation du nombre de points d'entraînement permet une amélioration de la prédiction. Chaque résultat est moyenné sur une centaine de plans d'expérience différents.

Les premières fonctions définissent un problème 2D pour avoir des premiers éléments de comparaison. Les deux niveaux de fidélité sont donnés dans l'équation [IV-3] où les deux paramètres sont bornés entre 0 et 1 :

$$\begin{aligned} \mathcal{M}_{HF}(x_1, x_2) &= \left(1 - \exp\left(-\frac{1}{2x_2}\right)\right) \frac{2300x_1^3 + 1900x_1^2 + 2092x_1 + 60}{100x_1^3 + 500x_1^2 + 4x_1 + 20} \\ \mathcal{M}_{LF}(x_1, x_2) &= (1/4) \\ &\times \left[ \mathcal{M}_{HF}\left(x_1 + \frac{1}{20}, x_2 + \frac{1}{20}\right) + \mathcal{M}_{HF}\left(x_1 + \frac{1}{20}, \max\left(0, x_2 - \frac{1}{20}\right)\right) \right. \\ &\left. + \mathcal{M}_{HF}\left(x_1 - \frac{1}{20}, x_2 + \frac{1}{20}\right) + \mathcal{M}_{HF}\left(x_1 - \frac{1}{20}, \max\left(0, x_2 - \frac{1}{20}\right)\right) \right] \end{aligned} \quad [IV-3]$$

Une étude préliminaire a permis de déterminer la fonction de corrélation la plus intéressante pour chaque métamodèle. La fonction Matérn 5/2 est associée au co-krigeage et la fonction de base cubique est utilisée pour la co-RBF. Les diagrammes de la Figure IV-9 indiquent la moyenne de l'erreur de prédiction pour chaque combinaison de points HF et LF, et ceux de la Figure IV-10 correspondent à l'écart type. Cette expérience sur un problème relativement simple permet de mettre en évidence que l'augmentation du nombre de points HF et LF améliore la précision des deux métamodèles. Le co-krigeage a cependant une erreur légèrement plus faible avec un peu moins de dispersion dans les résultats.

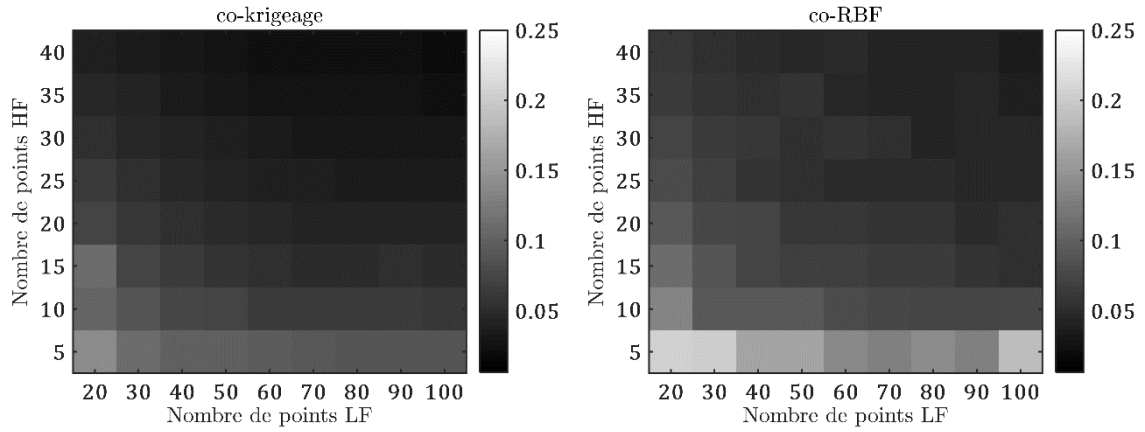


Figure IV-9 : Moyenne de l'erreur de prédiction (RAAE) du problème de l'équation [IV-3] sur 100 plans d'expériences différents.

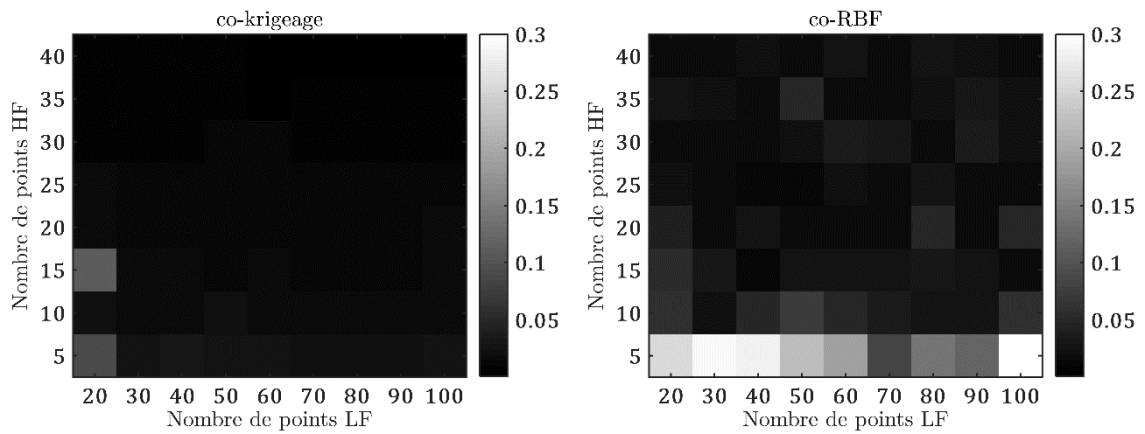


Figure IV-10 : Ecart type de l'erreur de prédiction (RAAE) du problème de l'équation [IV-3] sur 100 plans d'expériences différents.

Le second problème comporte huit paramètres afin d'étudier l'impact sur la prédiction d'une montée en dimension. Les fonctions sont données à l'équation [I-16] et les bornes de chaque paramètre sont indiquées dans le Tableau 9.

$$\begin{aligned}
 \mathcal{M}_{HF} &= \frac{2\pi T_u (H_u - H_l)}{\log(r/r_w) \left[ 1 + \frac{2LT_u}{\log(r/r_w) r_w^2 K_w} + \frac{T_u}{T_l} \right]} \\
 \mathcal{M}_{LF} &= \frac{5\pi T_u (H_u - H_l)}{\log(r/r_w) \left[ 1.5 + \frac{2LT_u}{\log(r/r_w) r_w^2 K_w} + \frac{T_u}{T_l} \right]}
 \end{aligned}
 \tag{IV-4}$$

Paramètres	Bornes
$r_w$	[0.05 , 0.15]
$r$	[100 , 50000]
$T_u$	[63070 , 115600]
$H_u$	[990 , 1110]
$T_l$	[63.1 , 116]
$H_l$	[700 , 820 ]
$L$	[1120 , 1680]
$K_w$	[9855 , 12045]

**Tableau 9 : Bornes des paramètres du problème de l'équation [I-16].**

Une étude préliminaire désigne la fonction de corrélation gaussienne comme étant la plus intéressante pour ce problème avec les deux métamodèles multifidélités. Les diagrammes de la Figure IV-11 indiquent la moyenne de l'erreur de prédiction pour chaque combinaison de points HF et LF, et ceux de la Figure IV-12 correspondent à l'écart type. Ici la co-RBF conserve une augmentation de la précision avec l'augmentation du nombre de points HF et LF. Pour le co-krigeage, les résultats ne sont pas aussi réguliers. L'estimation des hyperparamètres est réalisée par maximum de vraisemblance. En observant les valeurs du facteur d'échelles  $\rho_{LF}$  en fonction du nombre de points HF, certaines valeurs sont très éloignées de la valeur médiane estimée par la co-RBF (proche de 0.4).

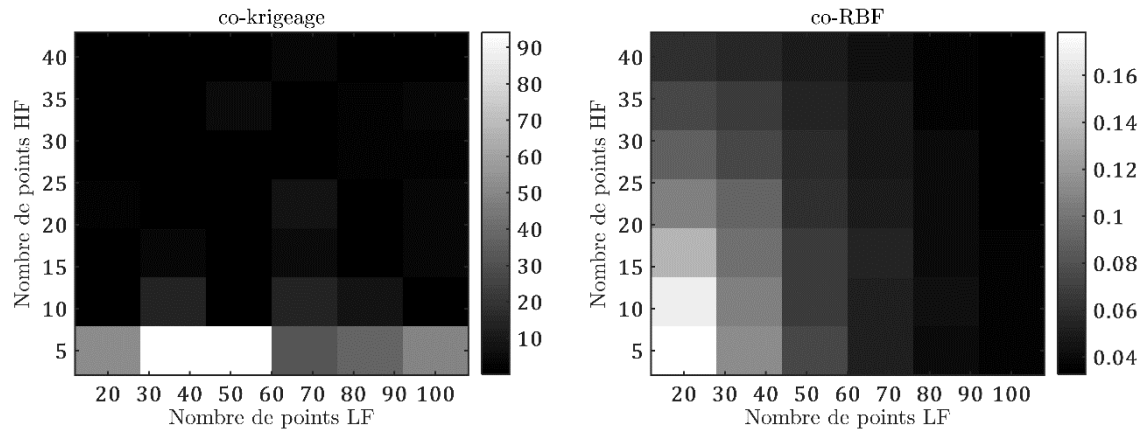


Figure IV-11 : Moyenne de l'erreur de prédiction (RAAE) du problème de l'équation [I-16] sur 10 plans d'expériences différents.

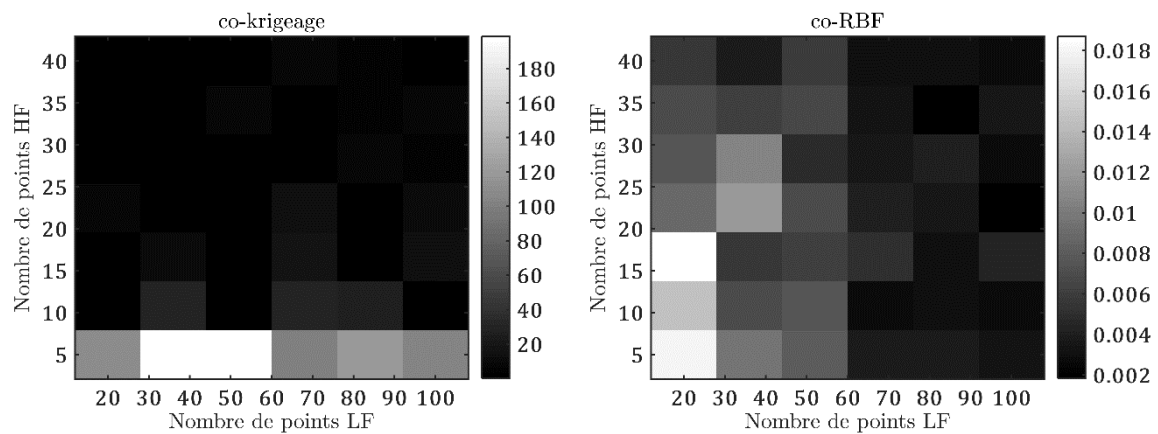


Figure IV-12 : Ecart type de l'erreur de prédiction (RAAE) du problème de l'équation [I-16] sur 10 plans d'expériences différents.

En utilisant la pseudo-vraisemblance et le calcul d'une erreur LOO pour estimer les hyperparamètres du co-krigeage, les valeurs du facteur d'échelle observées sont toutes uniformes avec très peu de variations. La capacité de prédiction du co-krigeage avec ses hyperparamètres estimés par LOO est visible sur la Figure IV-13. Les résultats sont plus intéressants avec une erreur de prédiction bien plus faible que la co-RBF.

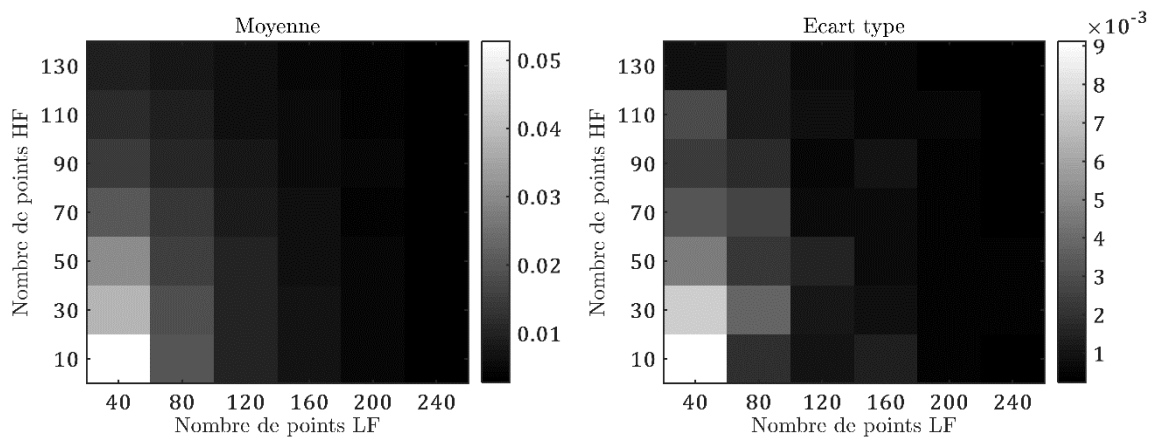


Figure IV-13 : Moyenne et écart type de l'erreur de prédiction (RAAE) du problème de l'équation [I-16] sur 10 plans d'expériences différents pour le co-krigeage avec optimisation des hyperparamètres par LOO.

Avant d'appliquer les métamodèles multifidélité comparés précédemment au problème de la cellule photoacoustique, un critère pour l'optimisation par co-RBF est proposé. Ce critère permet par la suite de

réaliser l'optimisation de la cellule.

### 3. Stratégie adaptative pour la RBF multifidélité

Afin de pouvoir réaliser une optimisation avec la co-RBF par stratégie adaptative, le critère défini par Gutmann est adapté au problème multifidélité. Pour rappel, l'approche de Gutmann consiste à trouver le point qui perturbe le moins possible l'interpolation lorsqu'il est ajouté aux données d'entraînement. La valeur du point ajouté correspond à une estimation du minimum global de la fonction. En faisant varier d'une manière spécifique l'estimation du minimum global, les points ajoutés permettent à la fois d'explorer l'espace des paramètres et de raffiner la précision de la prédiction autour du minimum global. En pratique, la minimisation de la quantité  $-\log \text{GUT}$  défini à l'équation [III-20] permet de trouver le prochain point à évaluer qui a le plus de chance d'être le minimum global de la fonction à optimiser. Pour adapter ce critère à la co-RBF, il est tout d'abord considéré que le coût de l'évaluation du modèle LF est bien plus faible que celui du modèle HF. Ainsi il est possible de construire un métamodèle précis du modèle LF car beaucoup d'évaluations peuvent être effectuées. Cette hypothèse permet de n'impliquer que les données HF dans le processus de la stratégie adaptative multifidélité. Pour améliorer la capacité de prédiction de la co-RBF, seul l'ajout de points du modèle HF est intéressant si le métamodèle LF est déjà précis. L'erreur de prédiction utilisée dans le critère GUT pour la multifidélité ( $\text{GUT}_{\text{MF}}$ ) est alors  $P_{\varphi_d}$  (celle du métamodèle de la différence) car elle dépend des données HF :

$$\text{GUT}_{\text{MF}}(\mathbf{x}) = \left( \frac{P_{\varphi_d}(\mathbf{x})}{\hat{y}_{\text{HF}}(\mathbf{x}) - f_{\text{min}}^*} \right) \quad [\text{IV-5}]$$

Dans ses travaux, Le Gratiet ([122] – section 5.2.2) utilise la MSE fournie par le krigeage sur les différents niveaux de fidélité afin de déterminer si l'évaluation du modèle LF peut permettre une plus grande réduction de l'erreur de prédiction que l'évaluation du modèle HF. Transposer cette stratégie pour le cadre de la co-RBF est une piste intéressante mais non explorée ici. Le choix du critère de stratégie adaptative a été porté sur celui de Gutmann mais une adaptation de CORS avec une contrainte de distance à la fois sur le modèle HF et LF est également envisageable. Pour autant, cette adaptation ne permet toujours pas d'indiquer le modèle le plus intéressant à évaluer pour réduire l'erreur de prédiction générale à moindre coût de calcul.

La prochaine étape consiste maintenant à utiliser la co-RBF sur les modèles numériques de la cellule photoacoustique et de comparer ses performances de prédiction avec le co-krigeage. Une optimisation de la conception est également possible grâce à l'adaptation du critère de Gutmann.

### 4. Application au cas de la cellule photoacoustique

Les deux données physiques d'études, à savoir le signal mesuré aux microphones de la cellule et la fréquence de résonance de la cavité sont à approximer par métamodèle. Pour construire les métamodèles multifidélité, des plans d'expériences contenant entre 5 et 20 points évalués sur le modèle FLNS (HF) et entre 10 et 100 points évalués sur le modèle Kreuzer (LF) sont utilisés. Pour chaque combinaison de points HF et LF évalués sur les modèles, 5 plans d'expérience différents sont construits pour moyenniser les résultats. Une mesure d'erreur de prédiction (la RAAE) est calculée à partir de 80 points de validation obtenus sur le modèle FLNS. Les fonctions de corrélation et les fonctions de base retenues pour chaque expérience correspondent à celles apportant les meilleurs résultats.

Les résultats obtenus pour l'approximation du signal de la cellule sont d'abord étudiés sur la Figure IV-14. Pour cette expérience, la fonction de corrélation Matérn 5/2 est utilisée pour le co-krigeage et la fonction de base cubique est associée à la co-RBF. Les diagrammes de la Figure IV-14 indiquent la moyenne de l'erreur de prédiction pour chaque combinaison de points HF et LF, et ceux de la Figure IV-15 correspondent à l'écart type. Le co-krigeage a une précision légèrement plus intéressante que la co-RBF sur l'ensemble des résultats. Il est également intéressant de remarquer que seule l'augmentation du nombre de points HF améliore de manière significative la capacité de prédiction des métamodèles.

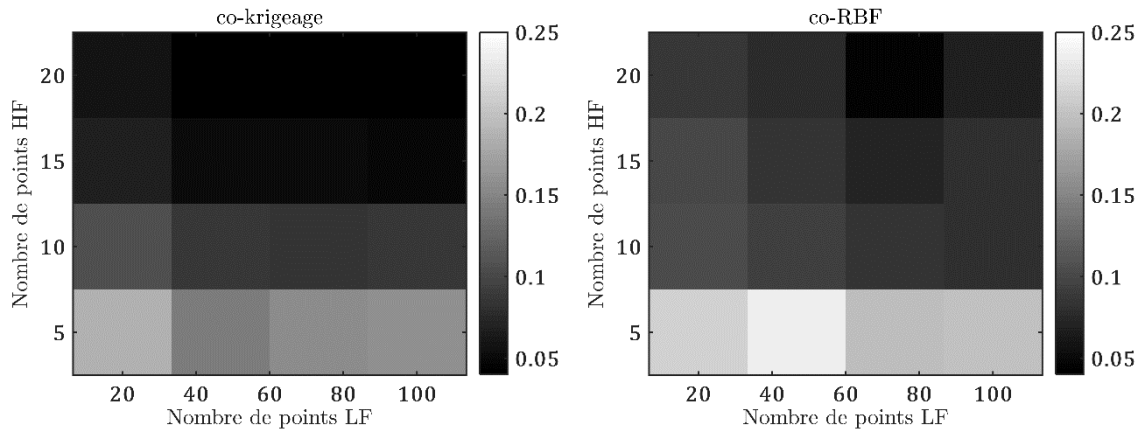


Figure IV-14 : Moyenne de l'erreur de prédiction (RAAE) du signal sur 5 plans d'expérience différents.

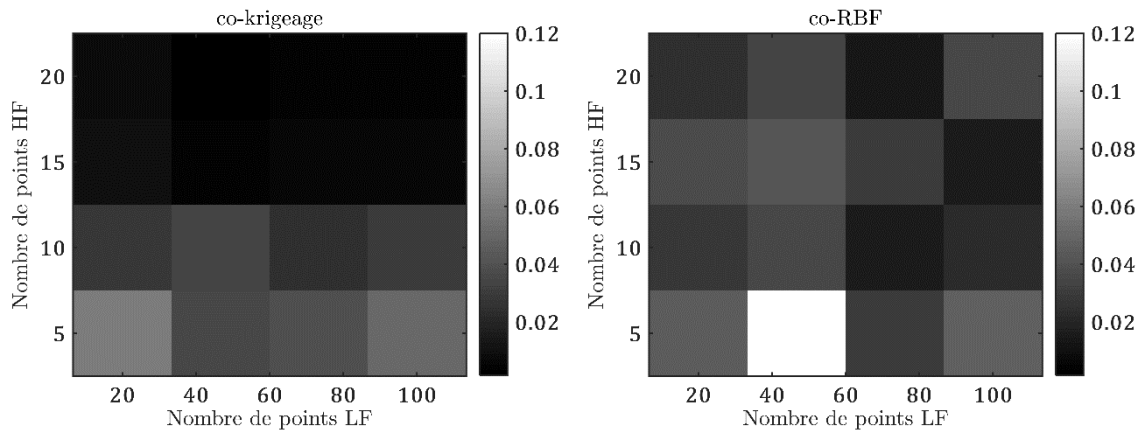


Figure IV-15 : Ecart type de l'erreur de prédiction (RAAE) du signal sur 5 plans d'expérience différents.

Un des éléments pouvant justifier la différence entre les deux métamodèles, hormis le choix de la fonction de corrélation, est l'estimation des hyperparamètres. Les valeurs des longueurs de corrélation ne peuvent être directement comparées entre les deux métamodèles car les échelles de distance entre les points ne sont pas les mêmes (données centrées-réduites pour le co-krigeage). Par contre le facteur d'échelle  $\rho_{LF}$  obtenu pour chaque métamodèle est comparable. La Figure IV-16 représente son évolution en fonction du nombre de points HF. Sur ces données, les variations de l'estimation du facteur d'échelle sont liées au nombre de points LF utilisés pour entraîner le métamodèle et le plan d'expérience utilisé (5 plans différents par combinaisons de points HF et LF). Le nombre de points LF a un impact limité sur cette estimation par rapport à celui des points HF, d'où le choix d'afficher ces résultats en fonction des points HF pour plus de lisibilité. Le co-krigeage comporte ici beaucoup moins de variation dans son estimation que la co-RBF. Lorsque le nombre de point HF augmente, les deux métamodèles fournissent une valeur estimée assez proche ( $\rho_{LF} \approx 1.2$ ). D'un point de vue physique, cela signifie que le signal fourni par le modèle Kreuzer est en tendance générale légèrement inférieur à celui obtenu par la modèle FLNS.



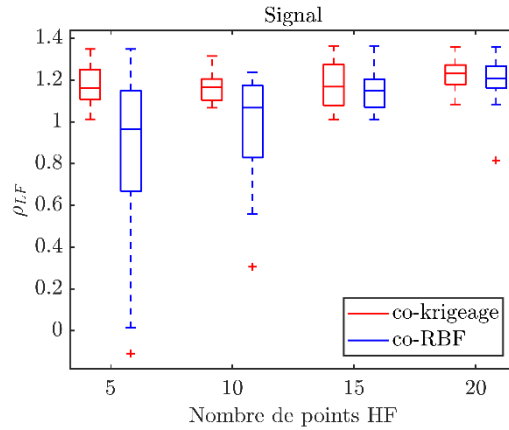


Figure IV-16 : Facteur d'échelle pour l'approximation du signal en fonction du nombre de points HF.

L'approximation de la deuxième donnée physique d'étude correspondant à la fréquence de résonance de la cavité est maintenant réalisée. La fonction de corrélation Matérn 5/2 est à nouveau utilisée pour le co-krigeage et la fonction de base gaussienne est associée à la co-RBF. Les diagrammes de la Figure IV-17 indiquent la moyenne de l'erreur de prédiction pour chaque combinaison de points HF et LF, et ceux de la Figure IV-18 correspondent à l'écart type. Les moyennes et écart-types des erreurs de chaque expérience sont très faibles pour le co-krigeage. Sur ces données, il est indéniable que le co-krigeage est plus précis que la co-RBF, même avec un faible nombre de points LF et HF.

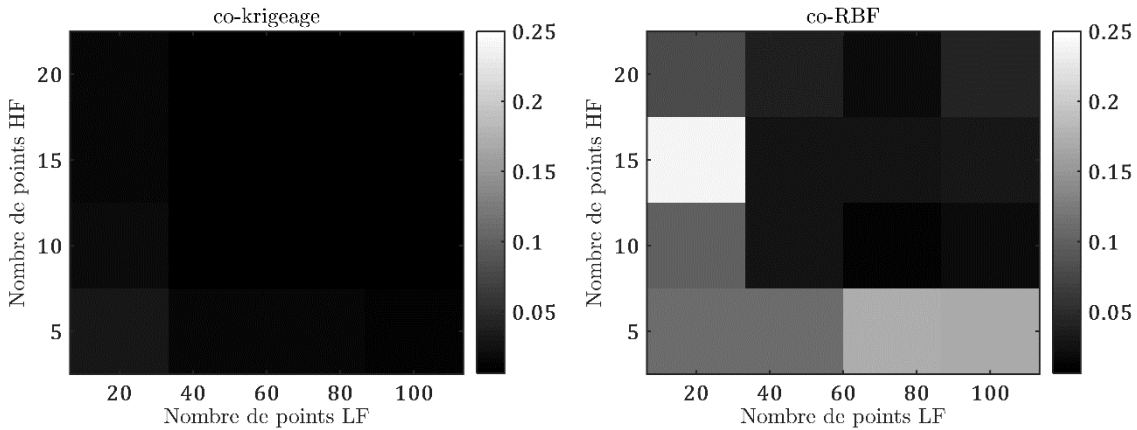


Figure IV-17 : Moyenne de l'erreur de prédiction (RAAE) de la fréquence de résonance sur 5 plans d'expériences différents.

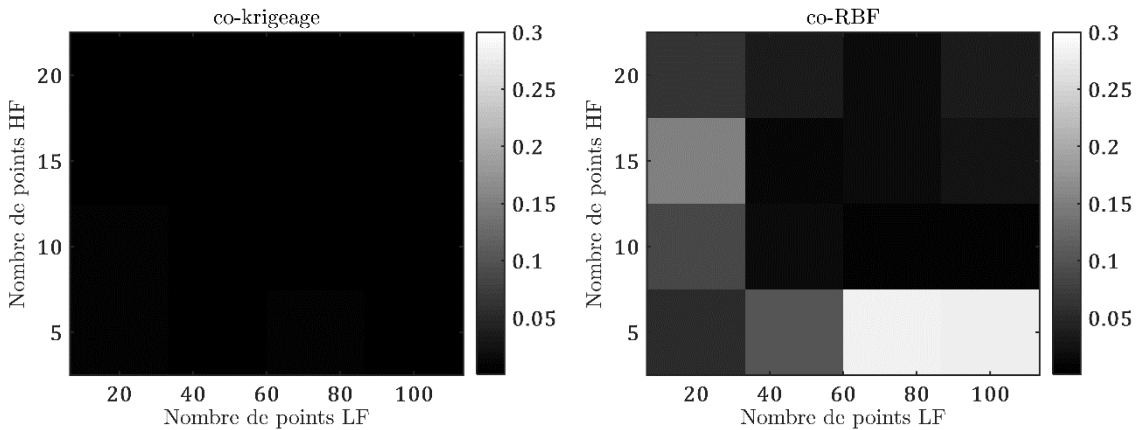
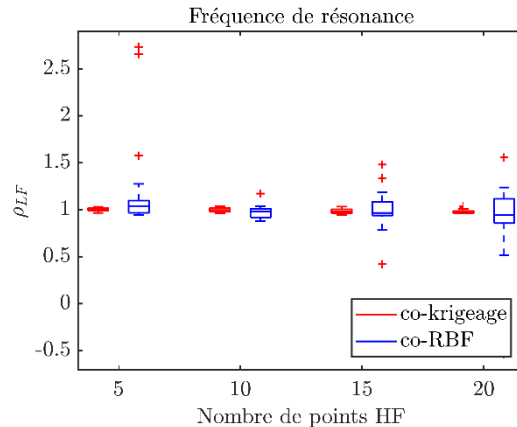


Figure IV-18 : Ecart type de l'erreur de prédiction (RAAE) de la fréquence de résonance sur 5 plans d'expériences différents.

L'analyse du facteur d'échelle sur la Figure IV-19 indique que la co-RBF comporte plus de variations dans son estimation que pour le cas de l'approximation du signal. L'estimation du co-krigeage reste toujours stable, avec une valeur pour 20 points HF en données d'entraînement qui est légèrement inférieure à 1. La fréquence de résonance fournie par les deux modèles numériques est donc relativement proche sur la plage de variation des paramètres de conception.



**Figure IV-19 : Facteur d'échelle pour l'approximation de la fréquence de résonance en fonction du nombre de points HF.**

Pour vérifier le bénéfice de l'approche multifidélité sur l'approximation des deux grandeurs physiques d'étudiées, les résultats des RBF, de la co-RBF et du co-krigeage sont comparés. Le Tableau 10 résume les précédentes expériences. Les données récoltées pour les RBF sur le modèle FLNS sont construites à partir des mêmes plans d'expériences que pour les métamodèle multifidélité. Le tableau donne également un rappel des données obtenues pour la co-RBF et le co-krigeage avec 100 points LF. Pour chaque nombre de points HF utilisés, les résultats sont toujours moyennés sur 5 plans d'expérience différents. Lorsque le nombre de points HF augmente, le bénéfice de l'approche multifidélité est plus faible. Le co-krigeage est le métamodèle le plus précis sur l'ensemble des résultats.

	Métamodèle	5 points HF	10 points HF	15 points HF	20 points HF
Signal	RBF	0.37±0.03	0.17±0.04	0.13±0.05	0.08±0.03
	co-RBF	0.20±0.05	0.08±0.02	0.08±0.02	0.07±0.04
	co-krigeage	0.16±0.05	0.09±0.03	0.05±7×10 <sup>-3</sup>	0.04±4×10 <sup>-3</sup>
Fréquence de résonance	RBF	0.38±0.17	0.12±0.03	0.06±0.01	0.04±5×10 <sup>-3</sup>
	co-RBF	0.17±0.28	0.02±0.01	0.03±0.03	0.04±0.04
	co-krigeage	0.01±2×10 <sup>-3</sup>	7×10 <sup>-3</sup> ±7×10 <sup>-4</sup>	7×10 <sup>-3</sup> ±6×10 <sup>-4</sup>	7×10 <sup>-3</sup> ±1×10 <sup>-3</sup>

**Tableau 10 : Comparaison des résultats entre RBF, co-RBF et co-krigeage. Les données multifidélités sont obtenues avec 100 points LF.**

L'amélioration de la prédiction de la fréquence de résonance avec peu de points HF peut être intéressante pour réduire le temps du calcul du modèle FLNS. Le résultat de cette modélisation est issu d'un balayage en fréquence. La réduction de la plage des fréquences étudiées est possible grâce à une estimation de la fréquence de résonance, celle où le signal extrait de la cellule est le plus élevé. Actuellement, 50 valeurs

entre 1000 Hz et 10000 Hz sont utilisées pour calculer le signal, puis une interpolation par spline des données est effectuée. En utilisant la valeur de fréquence de résonance fournie par le co-krigeage, il est possible de réduire le balayage à 5 valeurs, correspondant à une réduction de temps de calcul de 40 minutes.

Maintenant qu'un métamodèle du signal apporté par la cellule est disponible, une optimisation par stratégie adaptative est réalisée. Pour la co-RBF, le critère adapté à partir de celui de Gutmann est utilisé. Le résultat est comparé avec celui obtenu par critère EI pour le co-krigeage. Les plans d'entraînements utilisés au départ de l'optimisation sont composés de 5 points HF et 100 points LF. Il y a cinq plans différents pour observer les différences de résultats selon les conditions initiales et le nombre d'appels maximal autorisé au modèle FLNS est de 16 points. Les résultats sont visibles dans le Tableau 11 où les minimums obtenus et leur localisation sont comparés.

Les deux stratégies adaptatives conduisent à des résultats similaires. Deux optimums sont trouvés, avec le premier correspondant à une longueur de capillaire égale à celle des chambres et le deuxième où ce sont les diamètres des capillaires et des chambres qui sont égaux. Les deux solutions sont aux bornes de l'espace des paramètres de conception. Il n'est pas possible de modifier ces limites car elles correspondent à des choix d'intégration du composant dans un système global. Les performances du capteur sur ces deux solutions optimales ne sont pas significativement différentes. Il est possible d'approximer la cellule photoacoustique par un système équivalent masse-ressort où les capillaires sont la masse et les chambres correspondent aux ressorts. Les deux géométries optimales correspondent à une masse équivalente similaire, ce qui peut expliquer que le comportement du capteur est le même pour les deux solutions.

Plan d'expérience initial	Valeur du minimum (Pa)		Localisation du minimum	
	co-krigeage	co-RBF	co-krigeage	co-RBF
1	-1.88	-1.91	[0.02 , $4.18 \times 10^{-4}$ , 0.0179]	[0.02 , $5.00 \times 10^{-4}$ , 0.0177]
2	-1.87	-1.90	[0.02 , $5.00 \times 10^{-4}$ , 0.0159]	[0.02 , $4.97 \times 10^{-4}$ , 0.0173]
3	-1.91	-1.90	[0.02 , $4.86 \times 10^{-4}$ , 0.0188]	[0.02 , $4.73 \times 10^{-4}$ , 0.0200]
4	-1.90	-1.88	[0.02 , $4.54 \times 10^{-4}$ , 0.0200]	[0.02 , $5.00 \times 10^{-4}$ , 0.0200]
5	-1.87	-1.91	[0.02 , $4.58 \times 10^{-4}$ , 0.0200]	[0.02 , $4.46 \times 10^{-4}$ , 0.0200]

**Tableau 11 : Comparaison des résultats de l'optimum du signal de la cellule photoacoustique entre le co-krigeage et la co-RBF.**

## 5. Discussion

Une méthodologie pour la métamodélisation multifidélité à base de RBF a été proposée. Le métamodèle est construit à partir d'un modèle autorégressif utilisé au départ pour le co-krigeage. La comparaison sur deux problèmes analytiques montre que le co-krigeage et la co-RBF apportent des résultats similaires sur le problème à faible dimension. Lorsque le nombre de paramètres augmente, la co-RBF apporte des résultats présentant une erreur de prédiction moins élevée que le co-krigeage sur une des deux méthodes d'estimation des hyperparamètres (le maximum de vraisemblance apporte plus de variations dans les résultats que l'erreur LOO pour le co-krigeage sur notre exemple). L'intérêt de la co-RBF a également été validé sur un vrai problème de conception : la cellule photoacoustique pour la détection de gaz. Une extension du critère initialement proposé par Gutmann a permis de réaliser l'optimisation de la conception avec la co-RBF. Bien que moins performant sur sa capacité de prédiction par rapport au co-krigeage, les solutions de l'optimisation par co-RBF sont équivalentes à celles obtenues par critère EI. Ces résultats sont dans la continuité des nouvelles possibilités de stratégies adaptatives pour les méthodes à base de RBF dont le but est d'offrir une solution à différents types de problèmes, telle que l'optimisation multi-objectif [90],

[91].

### C. Optimisation robuste du coupleur selon la longueur d'onde

La fabrication de composants photoniques comporte de nombreuses sources d'incertitudes qui rendent difficile d'assurer la parfaite réalisation de la géométrie désirée. Le coupleur adiabatique doit assurer une division de puissance à une longueur d'onde très précise. Si la géométrie conçue n'est pas fabriquée selon des tolérances strictes, le risque est élevé que le ratio de division de puissance souhaité soit décalé sur une autre longueur d'onde, impliquant un futur dysfonctionnement du système optique complet. Pour cette étude, les incertitudes de fabrication ne sont pas suffisamment connues pour réaliser une optimisation robuste sur la géométrie. D'après les travaux de Lu et Vučković [123], la recherche de composants larges bandes est une heuristique valide pour l'obtention de composants tolérants aux erreurs de fabrication. Ce résultat a tout d'abord été observé en pratique, avec une géométrie optimale du coupleur obtenue par une méthodologie moins évoluée que celle qui est décrite par la suite dans cette section [124]. Deux coupleurs adiabatiques ont été fabriqués, le premier est optimisé pour la longueur d'onde centrale de l'étude  $1.31 \mu\text{m}$  et le deuxième correspond à un optimum robuste en longueur d'onde sur la bande d'intérêt. Le but est de comparer leur réponse est de vérifier si celle-ci est stable pour l'ensemble des puces produites sur le wafer. Ce travail a été réalisé en collaboration avec Karim Hassan pour le suivi de fabrication en salle blanche et l'interprétation des résultats après les mesures expérimentales.

Les composants sont fabriqués sur la ligne pilote 200 mm du CEA LETI à partir d'un substrat SOI et d'une couche supérieure de silicium de 300 nm d'épaisseur. Une lithographie à rayon ultraviolet profond et une gravure ionique réactive permettent de dessiner la géométrie dans le silicium. Une encapsulation de  $1 \mu\text{m}$  d'épaisseur par un plasma de haute densité est réalisée pour protéger les dispositifs. La mesure des caractéristiques de réponse des composants est effectuée grâce à un prober fibre-à-fibre semi-automatisé. Ces données sont normalisées par rapport à la réponse des réseaux de couplage et d'un guide droit de référence. La mesure de la réponse des coupleurs de chaque type est visible sur la Figure IV-20. Les spectres de réponses sont légèrement décalés vers les plus basses longueurs d'ondes par rapport aux simulations. Ces écarts peuvent être dus à une sous gravure de la nervure des guides de 5 à 10 nm ou de l'approximation paraxiale de la simulation BPM qui est trop forte. Pour autant, une division quasi parfaite de 1270 à 1310 nm est observée pour le coupleur optimisé sur toute la bande d'intérêt. La division de puissance parfaite pour le coupleur optimisé à  $1310 \text{ nm}$  est plutôt située à  $1340 \text{ nm}$  sur cette mesure. Les franges présentes aux extrémités des spectres peuvent être attribuées à la normalisation par la réponse des réseaux de couplages car ces derniers sont conçus pour  $\lambda_c = 1310 \text{ nm}$ .

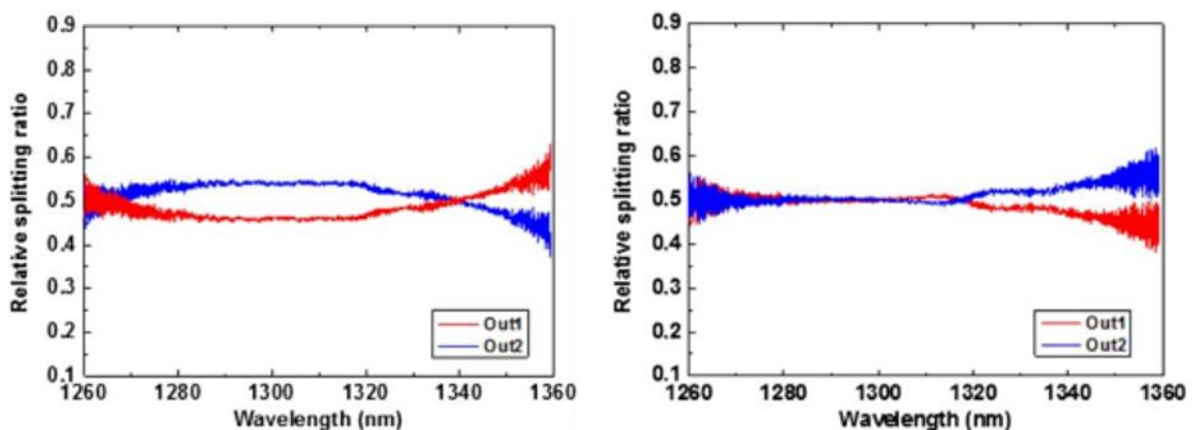
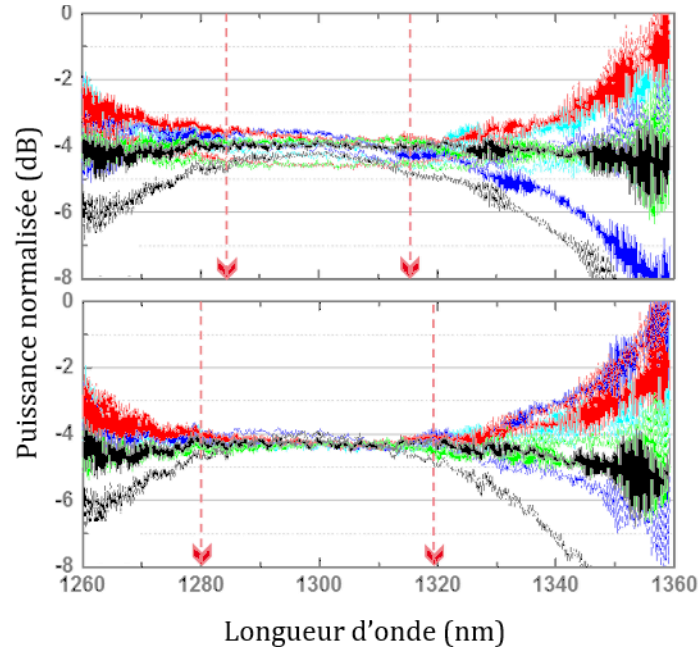


Figure IV-20 : Mesure de puissance en sortie des coupleurs sur une puce. La réponse à gauche correspond à un coupleur optimisé à la longueur d'onde  $1.31 \mu\text{m}$  et celle de droite à un coupleur optimisé sur toute la bande d'étude. [124]

La robustesse des coupleurs est maintenant analysée en effectuant des mesures sur cinq puces situées à différents lieux bien distincts sur toute la surface du wafer. Sur la Figure IV-21, le spectre supérieur montre une réponse plutôt plate mais avec léger décalage entre chaque puce. Le comportement des coupleurs est

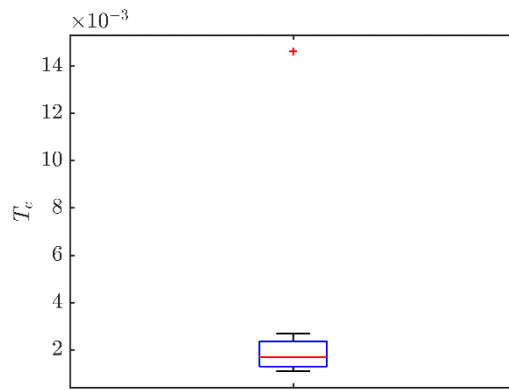
stable sur une bande de 32.7 nm mais avec une variance importante entre les puces ( $\pm 0.75$  dB). Ceci peut être dû aux variations des procédés de lithographie ou de gravure sur l'ensemble du wafer, voir même de l'épaisseur du SOI. Il est à noter qu'au -3 dB attendu sur le composant il faut ajouter des pertes d'insertion de 1 dB à cause de virages en S non-optimisés qui sont utilisés pour la connexion des coupleurs aux réseaux de couplage. Pour les coupleurs optimisés en robustesse dans la bande d'intérêt (Figure IV-21, spectre inférieur), une bonne stabilité est observée avec une fenêtre de 38.3 nm avec un écart de  $\pm 0.5$  dB entre les puces confirmant un transfert adiabatique de la lumière. Dans la suite de cette étude, l'optimisation robuste est alors appliquée sur la longueur d'onde.



**Figure IV-21 : Spectres des transmissions normalisées de chaque sortie du coupleur sur 5 puces différentes. La figure du haut correspond à un coupleur optimisé pour une seule longueur d'onde et celle du bas pour le coupleur robuste en longueur d'onde. Chaque couleur représente une mesure sur une puce différente. [124]**

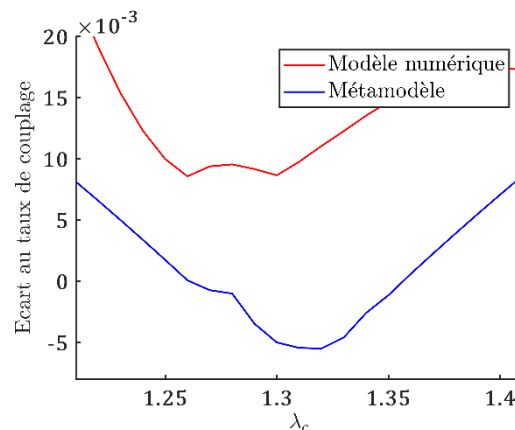
## 1. Problématique initiale

Lors de l'étude en section II.D.3, les métamodèles de l'écart au taux de couplage n'ont pas une bonne capacité de prédiction. La variation des paramètres de largeur des guides rend la réponse du modèle numérique assez erratique, d'où la difficulté de construire un métamodèle précis. Le but est maintenant d'utiliser l'information fournie par le métamodèle pour trouver la géométrie autorisant un écart au taux de couplage le plus faible possible sur une certaine plage de longueurs d'ondes. Une stratégie adaptative d'optimisation robuste au sens du pire scénario sur le paramètre de longueur d'onde est la méthode choisie ici. Ce type de stratégie a été décrit en section III.C et correspond aux travaux de Rehman et al. [115]. Le métamodèle utilisé est un krigeage avec une fonction de corrélation exponentielle. L'optimisation est lancée à partir de 8 plans d'expérience différents comportant chacun 200 points afin de moyennner les résultats. Le nombre d'itérations maximal a été atteint à chaque expérience et correspond à 450 appels au modèle numérique. Sur la Figure IV-22, les différentes valeurs des optimums obtenus sont représentées.



**Figure IV-22 : Minimums au sens du pire cas de l'optimisation robuste de l'écart au taux de couplage sur 8 expériences.**

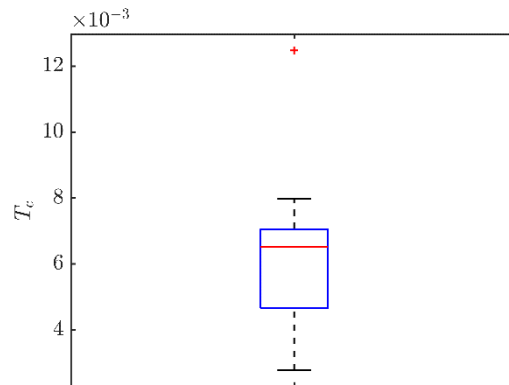
Les valeurs obtenues du pire écart au taux de couplage à la fin de l'optimisation sont calculées à partir de la prédiction du métamodèle. La longueur d'onde étant considérée comme une variable environnementale dans le cadre de cette étude, chaque appel au modèle numérique ne se fait que pour une seule valeur de longueur d'onde. Dans le cadre du pire scénario, la valeur d'une variable environnementale choisie pour l'évaluation correspond à celle ayant la plus grosse erreur de prédiction (comme indiqué à l'équation [III-65]). Pour la géométrie optimale, il n'y a eu donc qu'une seule valeur de longueur d'onde sur toute la bande étudiée dont le modèle numérique a fourni la réponse. Les taux de couplages pour d'autres longueurs d'ondes sont obtenus grâce à la prédiction du krigeage. D'après les résultats, le pire écart au taux de couplage souhaité est proche de  $2 \times 10^{-3}$  en moyenne sur l'ensemble des expériences. Or si le métamodèle n'est pas suffisamment précis, ces résultats peuvent être faux. Sur la Figure IV-23, la réponse du meilleur optimum des 8 expériences est complètement évaluée avec le modèle numérique et comparée avec celle fournie par le métamodèle. Des valeurs négatives sont prédites même si physiquement cela n'a aucune signification, il s'agit simplement d'une zone où le métamodèle comporte une forte erreur de prédiction.



**Figure IV-23 : Comparaison de la réponse de la géométrie optimale fournie par le métamodèle et celle fournie par évaluation du modèle numérique sur toute la bande d'étude.**

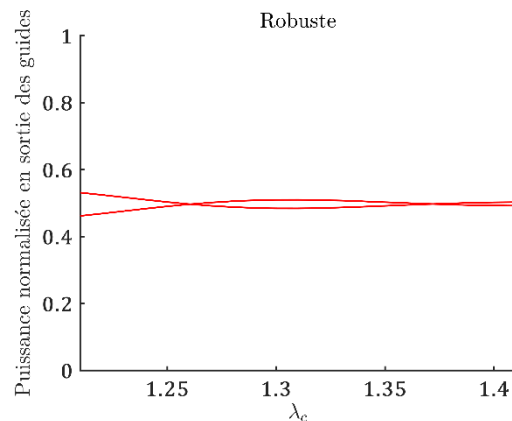
L'écart entre la réponse fournie par le métamodèle et celle du modèle numérique est très important. La méthode choisie ici pour l'optimisation robuste du coupleur ne permet pas d'obtenir le résultat attendu. Le problème ciblé concerne la capacité de prédiction du métamodèle qui met en échec la stratégie adaptative. La réponse fournie par le modèle numérique est certainement trop complexe pour que la construction d'un modèle de substitution soit possible à partir d'un nombre limité de points. Une des pistes envisagées pour améliorer la prédiction consiste à discrétiser la bande d'étude en un nombre limité de longueur d'onde, avec un métamodèle construit pour chaque longueur d'onde. L'ensemble des prédictions des métamodèles est alors utilisé dans la stratégie adaptative et lorsqu'un appel doit être effectué au modèle numérique, chaque

longueur d'onde initialement choisie est évaluée. Ainsi, le pire écart au taux de couplage correspond exactement à la réalité du modèle numérique, limité cependant à la discrétisation initiale de la bande d'étude. Pour cette étude, trois longueurs d'ondes ont été choisies dans la bande d'intérêt, à savoir  $\lambda_c = \{1.26, 1.31, 1.36\}$ . Sur la Figure IV-24, les valeurs obtenues du pire écart au taux de couplage sur les trois longueurs d'ondes choisies sont représentées pour 10 optimisations à partir de plans d'expérience différents.



**Figure IV-24 : Minimums au sens du pire cas de l'optimisation robuste de l'écart au taux de couplage avec discrétisation en longueur d'onde sur 10 expériences.**

Ces résultats avec une médiane proche de  $6.5 \times 10^{-3}$  sont moins bons que ceux annoncés sur la Figure IV-22 mais sont censés mieux représenter la véritable réponse du modèle numérique. La réponse complète du meilleur optimum obtenu avec la discrétisation en longueur d'onde est disponible sur la Figure IV-25. La puissance normalisée en sortie de chaque guide est tracée selon la longueur d'onde. Les valeurs de puissance doivent être le plus proche de 0.5 pour obtenir la division de puissance souhaitée. Ici le résultat est très bon pour les longueurs d'ondes 1.26  $\mu\text{m}$  et 1.36  $\mu\text{m}$  mais reste perfectible pour la valeur centrale d'intérêt de 1.31  $\mu\text{m}$ .



**Figure IV-25 : Comparaison de la réponse de la géométrie optimale avec discrétisation de la bande d'intérêt et celle fournie par évaluation du modèle numérique sur toute la bande d'étude.**

A partir des résultats précédents, deux axes de travail peuvent être construits. Le premier concerne l'amélioration de la métamodélisation du modèle numérique du coupleur. Le paramètre de longueur d'onde a été discrétisé en seulement trois valeurs et un métamodèle est construit pour chaque valeur. Il est certainement possible de tenir compte de l'information contenue dans le métamodèle pour une longueur d'onde et de l'utiliser dans la construction du métamodèle des deux autres longueurs d'ondes. Cette idée est développée dans la section suivante. Le deuxième axe possible concerne la stratégie adaptative pour l'optimisation robuste. Il est certainement possible d'obtenir des résultats plus intéressants en adaptant le critère d'optimisation à la nouvelle formulation du problème. La discrétisation de la longueur d'onde réduit

l'évaluation de la mesure de robustesse à un calcul sur trois valeurs au lieu d'un intervalle continu. Dans le cadre du pire scénario comme mesure de robustesse, l'optimisation simultanée des trois valeurs d'intérêt revient à optimiser la mesure de robustesse. En considérant la valeur de l'écart au taux de couplage pour les trois longueurs d'ondes choisies comme des objectifs indépendants, l'optimisation du pire scénario revient à réaliser une optimisation multi-objectif de ces trois objectifs. Un nouveau critère de stratégie adaptive multi-objectif pour le krigeage est proposé en section IV.C.3 comme une alternative au critère EHVI (présenté en section III.C.2.a) pour répondre à ce problème spécifique.

## 2. Métamodèle pour paramètres discrets

Avec la nouvelle formulation du problème de conception du coupleur, les quatre paramètres géométriques sont continus et la longueur d'onde est un paramètre discret (à trois valeurs). On appelle ce type de situation un problème à variables mixtes, au sens où la fonction à métamodéliser dépend de variables continues et discrètes. La construction de plusieurs métamodèles, un pour chaque combinaison de valeurs des variables discrètes, est contre-productive pour ce type de problème. En effet, si le nombre de niveaux d'une variable discrète est important ou s'il y a beaucoup de variables discrètes, un grand nombre de métamodèles est à construire. De plus la corrélation entre les différents niveaux d'une même variable discrète n'est pas prise en compte. Dans le rapport SANDIA de Swiler et al. [125], trois méthodes sont référencées pour générer des métamodèles dans le cas de problème mixte. La première méthode nommée ACOSSO [126] est basée sur les splines pour décomposer la fonction d'intérêt en plusieurs composantes moins complexes. La deuxième méthode consiste à définir une fonction de covariance pour les processus gaussiens entre les valeurs des paramètres discrets [127]. La dernière méthode nommée treed gaussian process [128] correspond à une construction de différents processus gaussiens sur des partitions de l'espace des paramètres. D'après les recommandations du rapport, l'utilisation du krigeage pour variables mixtes avec l'utilisation de fonction de corrélation particulière semble être la méthode la plus intéressante. De plus elle permet de continuer à utiliser les stratégies adaptatives développées pour le krigeage. Ces travaux ont été réalisés dans le cadre de l'encadrement du stage de Master recherche de Thomas Braure (Ecole des Mines de Saint-Etienne) qui a réalisé l'implémentation de la méthode et les expériences numériques.

### a. Krigeage pour variables mixtes

Dans le cadre des problèmes mixtes, les points de l'espace des paramètres sont désormais définis par  $\mathbf{w} = (\mathbf{x}, \mathbf{u})$  où  $\mathbf{x}$  correspond aux paramètres continus et  $\mathbf{u}$  aux paramètres discrets. Les niveaux sont les différentes valeurs que peuvent prendre un paramètre discret. Une surface de réponse fournie par le modèle numérique correspond à une seule combinaison des niveaux des différentes variables discrètes. Le changement de valeur d'un seul paramètre discret implique une nouvelle combinaison et donc une nouvelle surface de réponse qui est disjointe de la première. On note par la suite  $m_k$  le nombre de niveau d'une variable discrète. Pour rappel dans le cas de variables continues, la matrice de corrélation du krigeage se calcule :

$$\Sigma_{ij} = C_{GP}(\mathbf{x}_{te}^{(i)}, \mathbf{x}_{tr}^{(j)}) = \sigma_{GP}^2 R(\mathbf{x}_{te}^{(i)}, \mathbf{x}_{tr}^{(j)}, \boldsymbol{\theta}),$$

$$i = 1, \dots, n_{te} \text{ et } j = 1, \dots, n \quad \text{[IV-6]}$$

Avec l'indice  $tr$  désignant les données d'entraînement et l'indice  $te$  les points de test où prédire la réponse du métamodèle. Dans le cas du krigeage pour variables mixtes, la nouveauté consiste à ajouter un terme de corrélation entre les niveaux d'une même variable discrète [127] :

$$C_{GP}(\mathbf{w}_{te}^{(i)}, \mathbf{w}_{tr}^{(j)}) = \sigma_{GP}^2 \prod_k^K \tau_{k_{\mathbf{u}_{te}, \mathbf{u}_{tr}}} R_k(\mathbf{x}_{te}^{(i)}, \mathbf{x}_{tr}^{(j)}, \boldsymbol{\theta}) \quad \text{[IV-7]}$$

avec  $K$  le nombre de variables discrètes. La variable  $\tau_{k_{\mathbf{u}_{te}, \mathbf{u}_{tr}}}$  correspond à la corrélation entre les niveaux



des données d'entraînement et de test de la  $k$ -ième variable. L'ensemble des corrélations inter-niveaux peuvent être regroupées en une matrice  $\mathbf{T}$  diagonale par bloc :

$$\mathbf{T} = \begin{bmatrix} \mathbf{T}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{T}_K \end{bmatrix} \quad [\text{IV-8}]$$

Où les  $\mathbf{T}_k, k = 1, \dots, K$  correspondent aux matrices de corrélation inter-niveaux de chaque paramètre discret. Dans un exemple où le second paramètre discret contient 3 niveaux, on obtient la matrice de corrélation inter-niveaux suivante :

$$\mathbf{T}_2 = \begin{bmatrix} 1 & \tau_{2_{12}} & \tau_{2_{13}} \\ \tau_{2_{12}} & 1 & \tau_{2_{23}} \\ \tau_{2_{13}} & \tau_{2_{23}} & 1 \end{bmatrix} \quad [\text{IV-9}]$$

Pour que  $C_{GP}$  soit une fonction de corrélation valide, il est nécessaire que  $\mathbf{T}$  soit définie positive à diagonale unitaire (propriété nommée PDUDE pour positive definite matrix with unit diagonal elements). Les corrélations inter-niveaux font parties des hyperparamètres et sont donc estimées lors de l'optimisation de la MLE ou du critère LOO. Il faut donc que la condition PDUDE sur  $\mathbf{T}$  soit en contrainte de ce problème d'optimisation. Une optimisation avec ce type de contrainte peut être résolue par des algorithmes spécifiques appelés SDP (pour SemiDefinite Programming). L'implémentation de ce type de méthode dans le cadre du krigeage pour variables mixtes est précisée par Qian et al. [127]. Le principal défaut de cette approche est la complexité numérique de l'optimisation SDP qui peut augmenter grandement le temps de calcul nécessaire à la construction du métamodèle. De plus, le nombre de paramètres à optimiser augmente rapidement avec un grand nombre de variables discrètes ou de niveaux. Afin d'éviter cette opération coûteuse, il existe des hypothèses sur les corrélations inter-niveaux permettant de simplifier leur optimisation. Quatre simplifications sont présentées ici, à savoir le cas isotrope, le cas corrélations groupées, le cas Cholesky et le cas hétéroscedastique.

La simplification par hypothèse isotrope suppose que les corrélations inter-niveaux sont les mêmes pour un même paramètre discret. Les matrices de corrélations inter-niveaux deviennent donc :

$$\mathbf{T}_k = \begin{bmatrix} 1 & c_k & \cdots & c_k \\ c_k & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & c_k \\ c_k & \cdots & c_k & 1 \end{bmatrix} \quad [\text{IV-10}]$$

avec  $c_k$  la corrélation inter-niveaux pour la  $k$ -ième variable discrète. Pour assurer la propriété PDUDE à cette matrice, il suffit de borner la corrélation  $c_k$  dans l'intervalle  $]1/(m_k - 1), 1[$  lors de l'optimisation des hyperparamètres [127]. Il faut préciser ici que la matrice  $\mathbf{T}$  est PDUDE si toutes les matrices de corrélation inter-niveaux sont PDUDE. Avec l'hypothèse isotrope, il n'y a que  $K$  hyperparamètres supplémentaires à optimiser par rapport au krigeage pour variables continues.

L'hypothèse des corrélations groupées est un peu moins restrictive que l'isotropie. Elle consiste à classer les niveaux d'un même paramètre discret en différents groupes. Pour les niveaux d'un même groupe, une hypothèse de corrélation isotrope est appliquée. Entre chaque groupe de niveaux, une hypothèse de corrélation isotrope est également choisie. La matrice de corrélation inter-niveaux suivante est alors obtenue :

$$\mathbf{T}_k = \begin{bmatrix} \mathbf{U}_1 & g_k & \cdots & g_k \\ g_k & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & g_k \\ g_k & \cdots & g_k & \mathbf{U}_{n_g} \end{bmatrix} \text{ avec } \mathbf{U}_i = \begin{bmatrix} 1 & c_{k_i} & \cdots & c_{k_i} \\ c_{k_i} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & c_{k_i} \\ c_{k_i} & \cdots & c_{k_i} & 1 \end{bmatrix} \quad [\text{IV-11}]$$

où  $g_k$  correspond la corrélation inter-groupe d'une même variable,  $c_{k_i}$  la corrélation du  $i$ -ième groupe et  $n_g$  le nombre de groupe total. Pour satisfaire la propriété PDUDE sur les matrice  $\mathbf{U}_i$ , les paramètres  $c_{k_i}$  doivent

être dans l'intervalle  $]1/(m_{k_i} - 1), 1[$  où  $m_{k_i}$  correspond au nombre de niveaux dans le groupe  $i$ , avec  $i = 1, \dots, n_g$ . Pour obtenir les limites des corrélations  $g_k$ , une transformation de la matrice  $\mathbf{T}_k$  est nécessaire :

$$\overline{\mathbf{T}}_k = \begin{bmatrix} v_1 & g_k & \cdots & g_k \\ g_k & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & g_k \\ g_k & \cdots & g_k & v_{n_k} \end{bmatrix} \text{ avec } v_i = \frac{1}{m_{k_i}} + \frac{m_{k_i}-1}{m_{k_i}} c_{k_i}, \quad i = 1, \dots, n_k \quad [\text{IV-12}]$$

D'après les récents travaux de Roustant et al. [129], les racines positives et négatives les plus proches de 0 du déterminant de la matrice  $\overline{\mathbf{T}}_k$  donnent les bornes de l'optimisation des paramètres  $g_k$  pour assurer la propriété PDUDE aux matrices  $\mathbf{T}_k$ . L'inconvénient de l'hypothèse des corrélations groupées réside dans le choix des groupes. Il est nécessaire de connaître les niveaux à rassembler grâce à une connaissance de l'impact des paramètres sur la réponse du modèle numérique. De plus il faut donc recourir à une optimisation imbriquée pour obtenir les paramètres  $g_k$  : une fois l'estimation des paramètres  $c_{k_i}$  effectuée, les bornes de  $g_k$  sont disponibles et ces paramètres peuvent être optimisés vis-à-vis de l'estimateur (LOO ou MLE). A chaque itération de l'optimisation de l'estimateur par rapport aux hyperparamètres, les  $c_{k_i}$  changent et il est nécessaire de réoptimiser les  $g_k$ . Le processus de construction du métamodèle est donc coûteux à cause de la procédure d'optimisation des hyperparamètres.

La troisième hypothèse étudiée ici est basée sur la décomposition de Cholesky des matrices de corrélations inter-niveaux. La méthode consiste à obtenir le résultat suivant :  $\forall k \in \{1, \dots, K\}, \mathbf{T}_k = \mathbf{L}_k \mathbf{L}_k^t$ , où les matrices  $\mathbf{L}_k$  sont triangulaires inférieures à diagonale strictement positive et dont les lignes sont des vecteurs unitaires [130]. Chaque ligne des matrices  $\mathbf{L}_k$  est alors modélisée comme les coordonnées d'un point sur une hypersphère unitaire de dimension  $m_k$ . Les éléments de la  $i$ -ième ligne et  $j$ -ième colonne de la matrice  $\mathbf{L}_k = \{l_{k,i,j}\}$  sont alors exprimés à partir d'angles  $\{\delta_{k,i,j}\} \in [0, \pi]$  :

$$l_{k,1,1} = 1 \text{ et } \forall k \in \llbracket 1, K \rrbracket, \forall i \in \llbracket 2, m_k \rrbracket, \quad \begin{cases} l_{k,i,1} = \cos(\delta_{k,i,1}) \\ \forall j \in \llbracket 2, i-1 \rrbracket, l_{k,i,j} = \sin(\delta_{k,i,1}) \dots \sin(\delta_{k,i,j-1}) \cos(\delta_{k,i,j}) \\ l_{k,i,i} = \sin(\delta_{k,i,1}) \dots \sin(\delta_{k,i,i-2}) \sin(\delta_{k,i,i-1}) \end{cases} \quad [\text{IV-13}]$$

La limitation des angles à l'intervalle  $[0, \pi]$  permet d'assurer que les éléments diagonaux des matrices  $\mathbf{L}_k$  soient strictement positifs, et donc que les matrices  $\mathbf{T}_k$  soient PDUDE. Ce type de paramétrisation est introduit initialement dans les travaux de Pinheiro et al. [131]. Les corrélations inter-niveaux peuvent prendre des valeurs négatives grâce à cette méthode, permettant ainsi une plus grande adaptation aux données d'entraînement durant l'estimation des paramètres. Les angles sont déterminés comme des hyperparamètres lors de l'optimisation des critères MLE ou LOO. La simplification par décomposition de Cholesky est actuellement la méthode ayant le meilleur compromis entre flexibilité et rapidité de construction du métamodèle.

L'hypothèse d'hétéroscédasticité est une extension à la décomposition de Cholesky. Il s'agit d'associer une variance de processus gaussien différente pour chaque niveau des variables discrètes. La décomposition se définit alors comme :

$$\forall k \in \{1, \dots, K\}, \mathbf{T}_k = \mathbf{L}_k \mathbf{D}_k \mathbf{L}_k^t, \quad [\text{IV-14}]$$

avec  $\mathbf{D}_k$  une matrice diagonale à termes strictement positifs. Cette hypothèse rajoute donc  $K \times m_k$  termes à estimer lors de l'optimisation des hyperparamètres. Il est important de noter que dans l'équation [II.32] concernant la variance de prédiction du krigeage, le terme  $\sigma_{GP}^2$  n'est plus factorisé. En effet la matrice  $\mathbf{T}$  est une matrice de covariance dans le cas hétéroscédastique. La variance de prédiction pour une combinaison quelconque de niveaux des paramètres discrets devient alors :

$$\hat{\sigma}_y^2(\mathbf{w}) = \prod_{k=1}^K \tau_{k_u} - \boldsymbol{\Sigma}_x \mathbf{C}_{tr}^{-1} \boldsymbol{\Sigma}_x^T + \frac{(1 - \mathbf{1}^T \mathbf{C}_{tr}^{-1} \boldsymbol{\Sigma}_x^T)^2}{\mathbf{1}^T \mathbf{C}_{tr}^{-1} \mathbf{1}} \quad [\text{IV-15}]$$

Le terme  $\tau_{k_u}$  correspond à la corrélation du  $k$ -ième paramètre discret pour le niveau où la prédiction est souhaitée. Il s'agit d'un des éléments de la diagonale de la matrice  $\mathbf{T}_k$ . Dans le cas hétéroscédastique, ces éléments ne sont pas égaux à 1 car ils incluent la variance. Cette simplification augmente la flexibilité de la décomposition de Cholesky à s'adapter aux données d'entraînement mais le nombre de paramètres à estimer devient plus important.

Sur les quatre méthodes présentées ici, l'hypothèse des corrélations groupées n'a pas été retenue dans la suite de cette étude car elle ne correspondait pas à une utilisation dans le cadre de notre application. Avant de passer à la construction métamodèle, il est nécessaire de spécifier les types de plans d'expérience intéressants pour les problèmes à variables mixtes. Tout d'abord il est possible de construire des plans de type OLHS pour chaque combinaison de niveaux des paramètres discrets. Il existe cependant le risque que les mêmes paramètres continus soient associés à deux combinaisons différentes de paramètres discrets. Dans ce cas la matrice de corrélation  $\boldsymbol{\Sigma}$  possède deux lignes identiques et devient non-inversible. Les plans de type « sliced LHS » (SLHS) permettent d'éviter cette situation car ils possèdent des propriétés LHS pour chaque combinaison de paramètres discrets tout en conservant une propriété LHS sur l'ensemble regroupé des points des paramètres continus. L'algorithme permettant la construction de ce type de plan est formalisé par Qian [132]. Une optimisation de ce type de plan est possible selon le critère de distance maximin, comme pour les plans OLHS. Yang et al. [133] proposent d'utiliser un algorithme de seuil en perturbant itérativement le plan SLHS par permutation jusqu'à l'amélioration du critère maximin d'une valeur prédéterminée. Ce type de plan est noté OSLHS dans la suite de cette étude.

Tous les éléments nécessaires à la construction du krigeage pour variable mixte sont maintenant disponibles. L'implémentation de la méthode a été validée par comparaison sur des cas analytiques avec les résultats du paquet R « mixgp » développé dans le cadre de la chaire OQUAIDO. Dans la section suivante, les performances de ce métamodèle sont directement comparées avec la construction de krigeages indépendants pour chaque combinaison de niveaux sur le modèle du coupleur.

#### b. Application à la substitution du modèle numérique du coupleur

Pour la comparaison du krigeage pour paramètres mixtes et la construction de multiples krigeages indépendants, un plus grand nombre de longueurs d'ondes est utilisé dans la bande d'intérêt, à savoir  $\lambda_c = \{1.26, 1.276, 1.293, 1.31, 1.326, 1.343, 1.36\}$ . Le but est de montrer l'intérêt du métamodèle pour paramètres mixtes lorsque le nombre de niveaux est important. Les plans d'expérience sont composés de 30 points par niveau pour le krigeage pour variables mixtes (OLHS, SLHS, OSLHS). Au total, il y a donc 210 points qui composent les différents plans. Afin d'évaluer la précision de la prédiction, 400 points pour chaque niveau sont générés par suite de Sobol' agrégés pour être utilisés comme références. Sur la Figure IV-26, les résultats de la prédiction des métamodèles construits sur dix plans d'expérience différents sont disponibles. Le critère pour mesurer la précision du métamodèle utilisé ici est la RMSE normalisée. Le terme « Ind » correspond aux krigeages indépendants construits sur chaque niveau, « iso » à la simplification isotrope pour le krigeage pour variables mixtes, « chol » pour la décomposition de Cholesky et « Het » pour le cas hétéroscédastique. La fonction de corrélation employée est la Matérn 5/2.

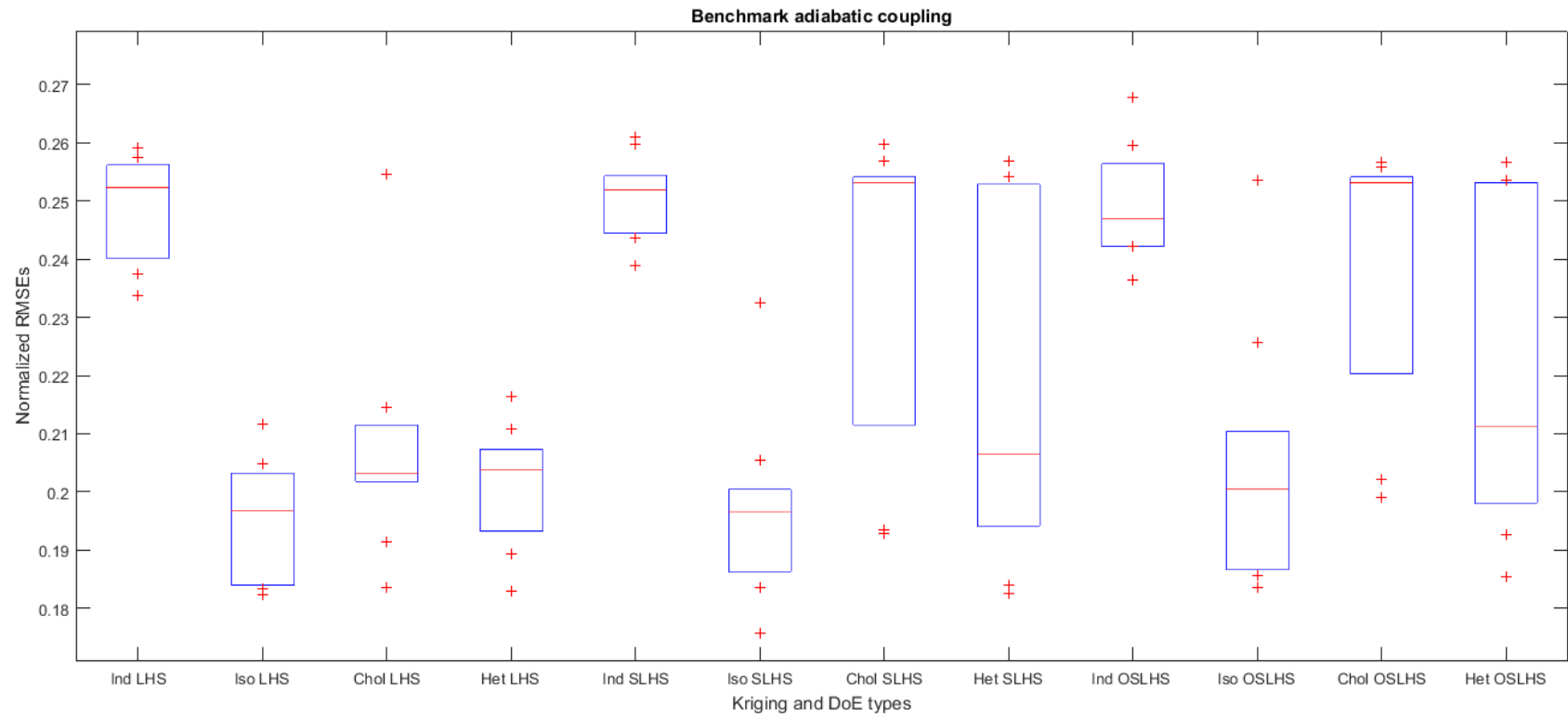
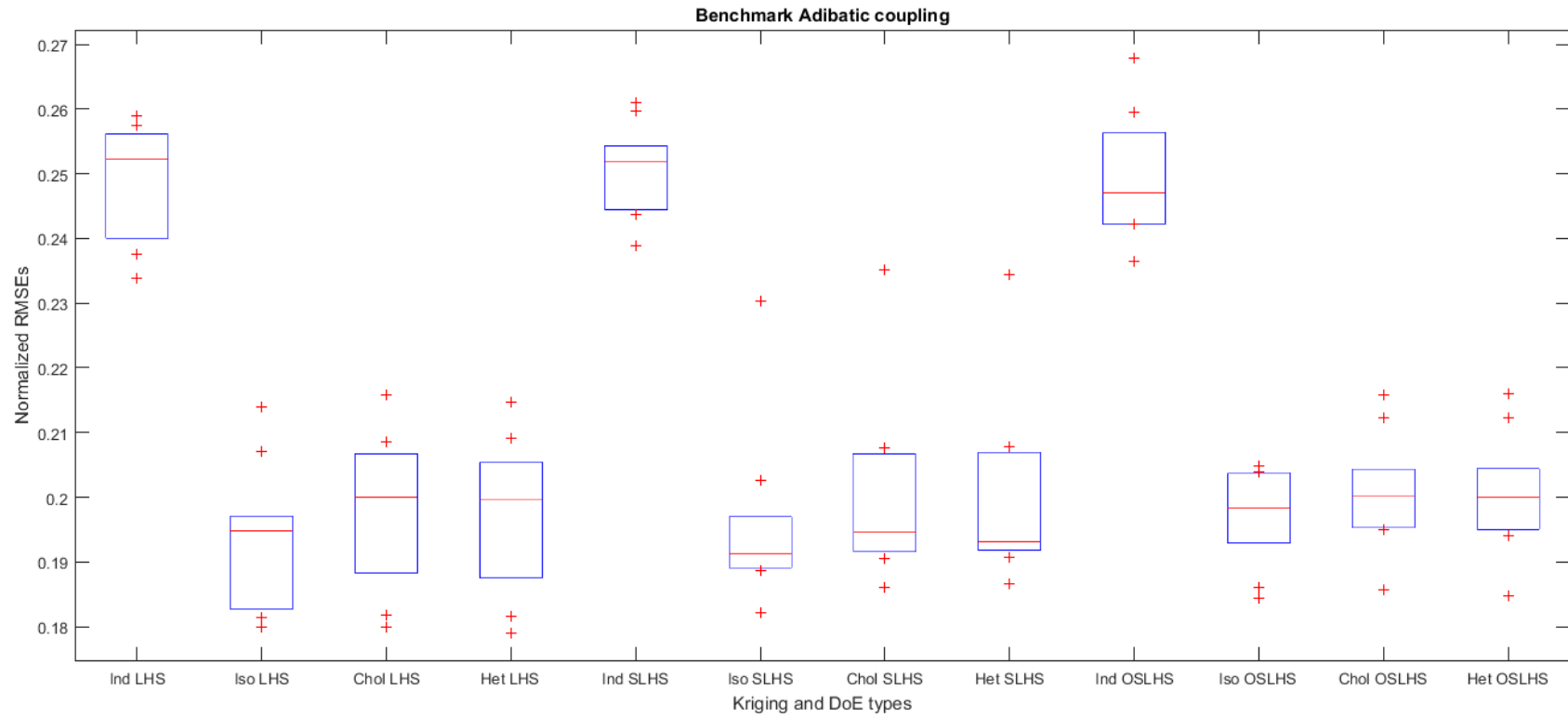
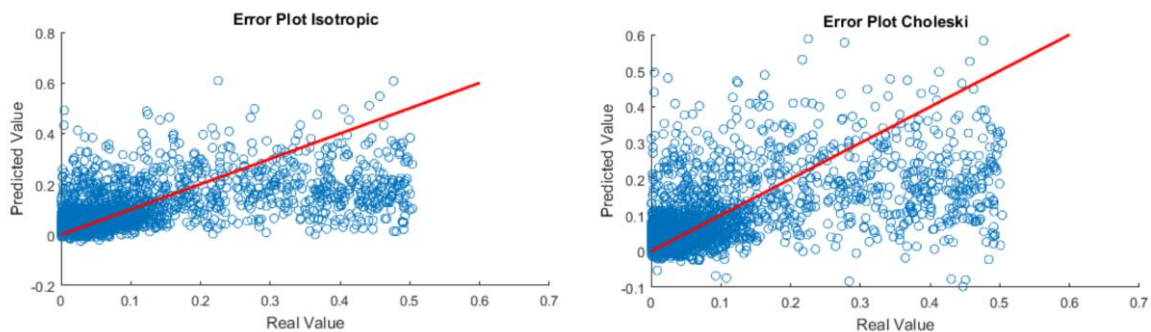


Figure IV-26 : Comparaison du krigeage indépendant et du krigeage pour variables mixtes sur le modèle du coupleur. La fonction de corrélation utilisée est la Matérn 5/2. [134]



**Figure IV-27 : Comparaison du krigeage indépendant et du krigeage pour variables mixtes sur le modèle du coupleur. La fonction de corrélation utilisée est la fonction exponentielle. [134]**

De grandes variances dans les résultats s’observent sur les plans d’expériences SLHS et OSLHS avec les simplifications par décomposition de Cholesky et le cas hétéroscédastique. Le krigeage indépendant est moins précis sur la majorité des résultats par rapport au krigeage pour variables mixtes. La Figure IV-27 présente la même expérience mais en utilisant la fonction de corrélation exponentielle qui est plus appropriée au cas du coupleur d’après l’étude en section II.D.3. Les erreurs des krigeages pour variables mixtes de type Cholesky et hétéroscédastique sont alors plus précises qu’avec une corrélation Matérn 5/2. La différence avec les krigeages indépendants est toujours très nette. La simplification par isotropie est celle qui a apporté le krigeage le plus précis sur l’ensemble des résultats. Pour autant les erreurs de prédiction sont toujours très importantes. La Figure IV-28 montre un exemple de la prédiction fournie par le krigeage pour variables mixtes de type isotrope ou Cholesky. Le résultat ne permet pas de se fier aux valeurs prédites par le métamodèle.



**Figure IV-28 : Graphique de comparaisons entre valeurs prédites et valeurs réelles pour le krigeage pour variables mixtes sur la prédiction de l’écart au taux de couplage. [134]**

Pour que ce type de krigeage soit intéressant dans le cadre de l’optimisation du coupleur, il est nécessaire que l’erreur de prédiction soit très faible. Si ce n’est pas le cas, l’ensemble des longueurs d’ondes lors de la stratégie adaptative doit être évalué pour la même géométrie. Or cette action n’est plus possible avec le krigeage pour variables mixtes sous peine de ne plus pouvoir inverser la matrice de corrélation. Dans la suite de cette étude, un krigeage indépendant par niveau du paramètre de longueur d’onde est donc construit avant de garantir que la solution obtenue en fin d’optimisation correspond réellement à un minimum intéressant.

### 3. Optimisation du gradient de l’amélioration espérée

Le deuxième axe de travail pour améliorer les résultats de l’optimisation robuste du coupleur concerne le choix de la stratégie adaptative. La nouvelle définition du problème permet la transformation de l’optimisation de la mesure du pire scénario en une optimisation multi-objectif. Le critère EHVI (voir section III.C.2.a) permet d’apporter des solutions à ce type de problème. Malheureusement, le coût de calcul de ce critère devient non négligeable lorsque le nombre d’objectifs est important. Afin de proposer une alternative si le nombre de longueurs d’ondes sélectionnées dans la bande d’intérêt est grand, une méthodologie originale basée sur l’optimisation multi-objectif du gradient du critère EI est proposée dans la suite.

#### a. Méthode EI-MGDA

Dans la section III.C.2.A, trois critères de stratégie adaptative pour l’optimisation multi-objectif par krigeage sont détaillés. Il existe cependant plusieurs alternatives telles que celles utilisées dans le benchmark de Horn et al [135]. Ces méthodes peuvent être grossièrement classées comme des techniques basées sur la notion de domination d’une solution par rapport à une autre. Feliot et al. [97] considère ces approches comme des extensions heuristiques du critère EI car elles ne sont pas issues d’une formulation bayésienne. Pour autant elles permettent d’obtenir des solutions pour un coût numérique assez faible comparé au calcul d’intégrale nécessaire pour le critère EHVI. Parmi ces méthodes, Jeong et al. [136] proposent d’utiliser les critères EI calculés pour chaque objectif directement dans une optimisation multi-objectif par algorithme

génétique. Leur cas d'application comporte deux objectifs et trois points parmi les solutions obtenues sont ensuite évalués sur le modèle numérique. Il s'agit des deux points ayant la valeur maximale pour chaque critère EI puis du point au milieu de l'ensemble des solutions. Cette méthode de sélection des points à évaluer est arbitraire et privilégie les valeurs extrêmes du front de Pareto.

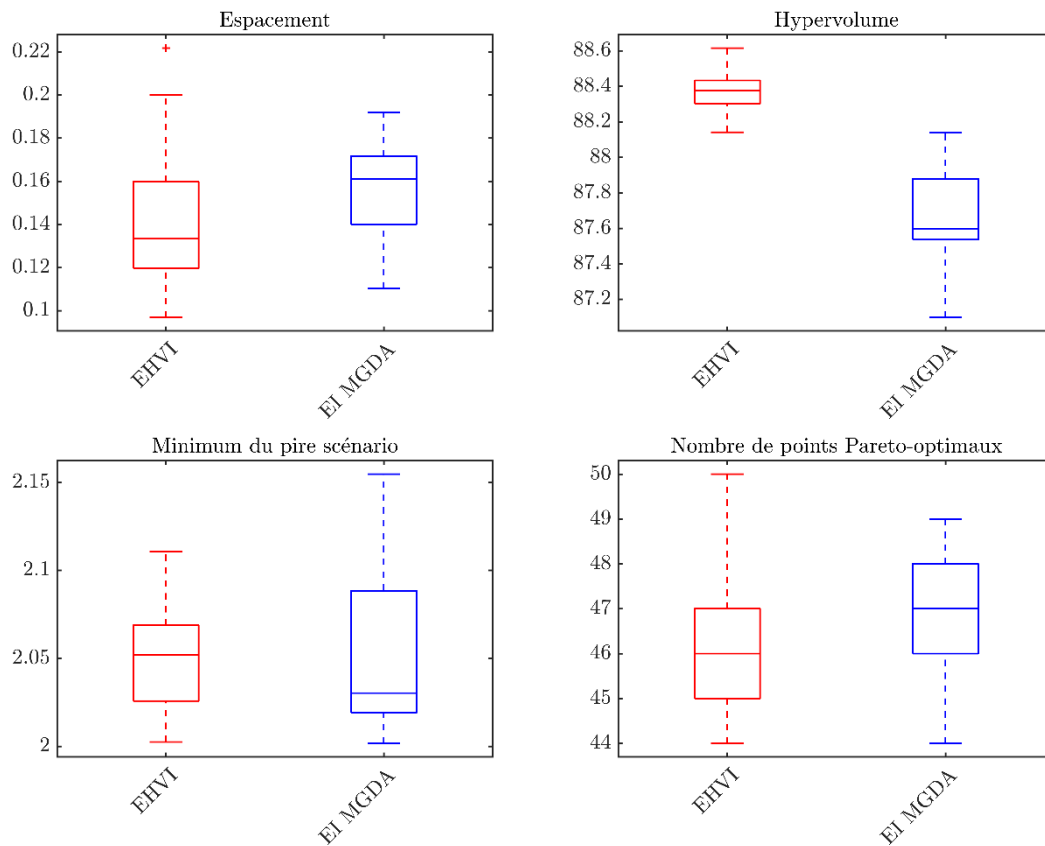
Pour notre cas d'application particulier, le but n'est pas d'améliorer une mesure spécifique du front de Pareto mais simplement de s'approcher du point idéal, qui représente le meilleur compromis entre les objectifs au sens du pire scénario. Par ailleurs, l'optimisation multi-objectif de l'EI permet de s'affranchir de la complexité numérique du critère EHVI lorsque le nombre d'objectifs augmente. La stratégie MOEGO de Guerra [101] peut permettre de mieux sélectionner les points à évaluer que le choix arbitraire effectué par Jeong et al. Par contre il est nécessaire de trouver un algorithme multi-objectif autorisant l'optimisation des critères à partir d'un point spécifique. MGDA apparaît comme une option intéressante ici car le calcul de la direction de descente reste très rapide à obtenir même lorsque le nombre d'objectifs est important. De plus, ce calcul est facilité par le gradient du critère EI qui est disponible analytiquement (voir [137] – section 4.3.1 pour le développement) :

$$\nabla EI(\mathbf{x}) = \left( EI(\mathbf{x}) - \frac{y_{\min} - \hat{y}(\mathbf{x})}{\hat{\sigma}_{\hat{y}}(\mathbf{x})} \Phi \left( \frac{y_{\min} - \hat{y}(\mathbf{x})}{\hat{\sigma}_{\hat{y}}(\mathbf{x})} \right) \right) \frac{\nabla \hat{\sigma}_{\hat{y}}(\mathbf{x})}{\hat{\sigma}_{\hat{y}}(\mathbf{x})} - \frac{\nabla \hat{y}(\mathbf{x})}{\hat{\sigma}_{\hat{y}}(\mathbf{x})} \quad [IV-16]$$

Cette proposition de stratégie adaptative a été construite pour la résolution de l'optimisation robuste du coupleur pour un grand nombre de longueurs d'ondes. Elle est maintenant comparée au critère EHVI sur l'exemple analytique section III.C.2.A.

#### b. Comparaison sur un exemple analytique avec le critère EHVI

Les fonctions objectifs de cet exemple sont disponibles à l'équation [III.51]. Pour rappel, le plan d'expérience au départ de la stratégie adaptative est composé de 20 points et le critère d'arrêt autorise 40 appels supplémentaires aux fonctions objectifs. Le nombre de points évalués sur une même itération de MOEGO a été arbitrairement fixé à 5. L'ensemble des expériences est répété une vingtaine de fois avec un plan initial différent. Les résultats sont visibles sur la Figure IV-29. Le critère EHVI comporte un espacement entre les points plus faible que le critère EI MGDA et son hypervolume est également plus important. Par contre, le critère EI MGDA permet d'obtenir plus de points Pareto-optimaux et de meilleurs optimums pour la mesure du pire scénario.

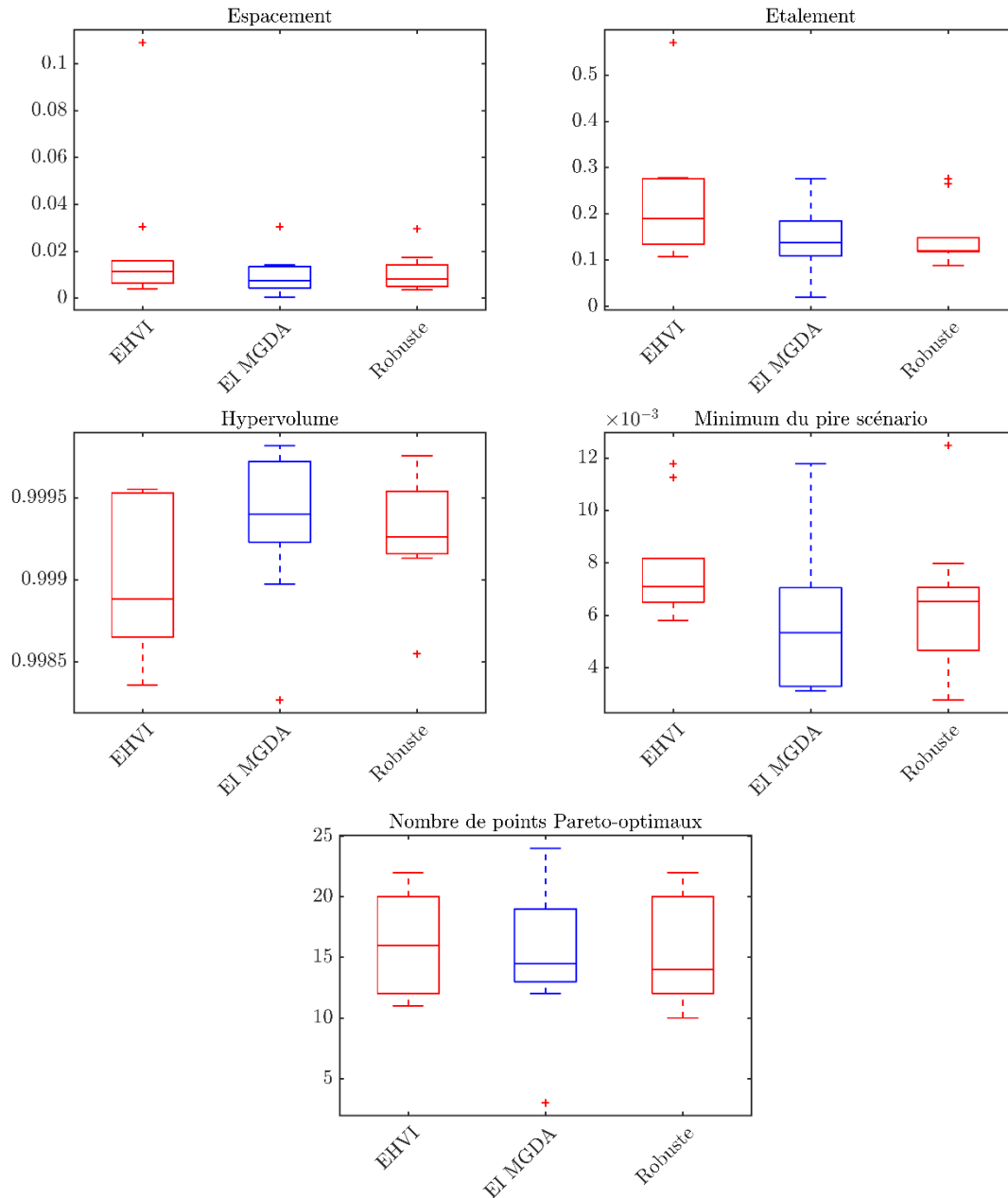


**Figure IV-29 : Résultat de l'optimisation multi-objectif des fonctions à l'équation [III-51] par critère EHVI ou EI MGDA.**

#### 4. Application de la stratégie générale

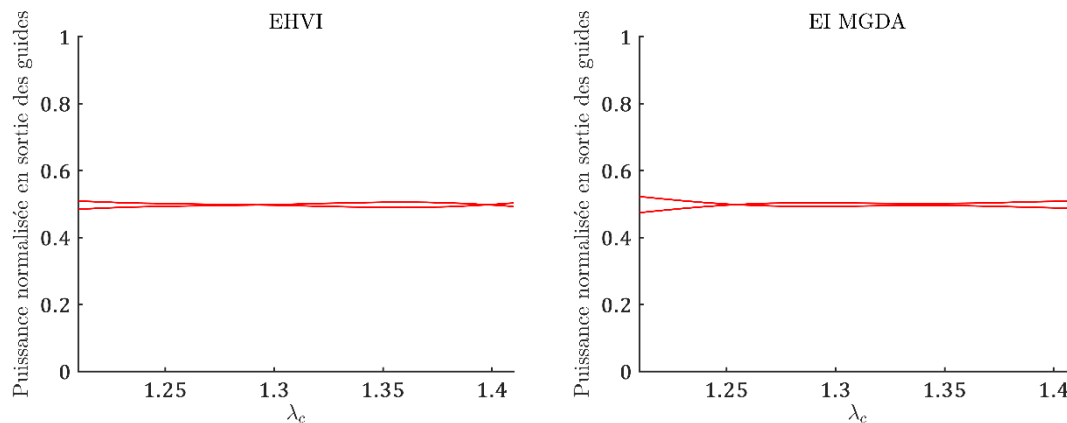
Les deux critères de stratégie adaptative sont maintenant comparés à l'optimisation robuste au sens du pire scénario sur le modèle numérique du coupleur. Les algorithmes sont initialisés à partir de 10 plans d'expérience différents comportant 200 points évalués sur trois longueurs d'ondes. Le critère d'arrêt de chaque stratégie est fixé à 50 nouveaux points ajoutés aux plans d'entraînement des métamodèles. Les résultats de l'optimisation robuste pour trois longueurs d'ondes en section IV.C.1 sont utilisés ici pour comparaison. Le calcul des mesures de comparaison s'effectue après obtention du dernier front de Pareto sur l'ensemble des données évaluées. Les résultats sont regroupés sur la Figure IV-30. L'indicateur le plus important dans notre cas correspond au minimum obtenu sur la mesure du pire scénario. En moyenne, c'est bien le critère EI MGDA qui amène les meilleurs résultats. Les autres indicateurs concernent la qualité du front de Pareto obtenu après optimisation. Les trois critères ont un nombre très proche de points Pareto-optimaux. On note toutefois que le critère EHVI n'est pas celui qui donne le meilleur hypervolume sur ces expériences alors qu'il a été développé spécifiquement dans ce but.





**Figure IV-30 : Comparaison des résultats de l'optimisation multi-objectif de l'écart au taux de couplage par critère EHV1, EI MGDA et optimisation robuste classique.**

Les puissances normalisées en sortie de chaque guide sont représentées sur la Figure IV-31 pour les meilleures solutions obtenues avec les critères EHV1 et EI MGDA. Ces résultats doivent être comparés avec la solution obtenue par optimisation du critère du pire scénario sur des longueurs d'ondes discrétisées. Les puissances en sortie de guides sont plus proches de 0.5 sur l'ensemble des longueurs d'ondes que la solution à la Figure IV-25. Il faut rappeler que les solutions présentées ici sont optimisées pour la bande  $\lambda_c \in [1.26 ; 1.36]$ . Il est donc normal d'observer un écart par rapport au taux de couplage désiré à l'extérieur de la bande d'étude.



**Figure IV-31 : Meilleurs optimums obtenus avec les critères EHVI et EI MGDA.**

Des géométries ayant été identifiées comme robustes selon les incertitudes de fabrications par simulation, il reste à vérifier si le résultat s'observe également en pratique.

## 5. Conclusion

La robustesse aux incertitudes de fabrication peut donc être assurée par une optimisation large bande des composants. Une nette différence dans les performances des composants est observée lorsque l'écart au taux de couplage souhaité est minimisé pour une ou plusieurs longueurs d'ondes. Comparée à la géométrie du coupleur initialement proposée par Xing et al. [138] pour une longueur d'onde centrale de 1.55  $\mu\text{m}$ , la longueur totale du composant a été réduite de moitié grâce à l'optimisation des paramètres géométriques. L'intérêt de la méthode proposée ici est également lié au fait que le comportement adiabatique du coupleur est assuré par le biais d'une méthode numérique automatique, permettant de se passer de l'étude proposée par Xing et al. [138] nécessitant différentes simulations physiques. La nouvelle méthodologie proposée fait appel à l'algorithme MGDA permettant l'optimisation simultanée de plusieurs critères à partir d'un point de départ. Bien que le problème soit posé comme une optimisation multi-objectif, le but n'est pas d'améliorer la qualité générale du front de Pareto mais seulement les solutions révélant le meilleur compromis au sens de la mesure du pire scénario. Comparé au critère EHVI, les résultats obtenus sont relativement similaires avec un certain avantage cependant pour le critère EI MGDA. L'utilisation du métamodèle pour paramètres mixtes n'a pas été concluante pour cette application car le modèle numérique a une réponse fortement non linéaire qui complique sa substitution. Toutefois elle a permis de gagner en précision sur la prédiction par rapport à la création de multiples métamodèles indépendants pour un même nombre de points d'entraînement.

---

## Conclusion

---

L'exploitation de simulation numérique pour la conception de dispositifs photoniques a nécessité le développement d'outils spécifiques. Tout d'abord, trois cas d'étude ont été sélectionnés et une identification des problématiques de conception a été réalisée au chapitre I. Dans le cas de la microsource infrarouge, les paramètres liés à la position des pistes électriques évoluent dans un domaine non hypercubique ce qui permet de pouvoir traiter le plus de configurations possibles. L'optimisation des objectifs doit également satisfaire une contrainte de température maximale afin de garantir la durée de vie du composant. Pour la cellule photoacoustique, l'utilisation du modèle haute fidélité est trop coûteuse en temps de calcul pour être complètement viable dans un contexte d'optimisation. Le modèle basse fidélité n'est pas assez représentatif du phénomène observé en expérimentation. Il apparaît nécessaire de combiner l'information des deux modèles pour réduire le coût de l'optimisation. Dans le cadre de l'optimisation robuste du coupleur, les sources d'incertitudes pour la réalisation de la géométrie en fabrication sont identifiées mais non quantifiées. Une alternative consiste à assurer la robustesse de fonctionnement du composant selon la longueur d'onde.

Une possibilité pour limiter l'impact du coût en temps de calcul de chaque modèle numérique lors de la conception de ces dispositifs repose sur l'utilisation de métamodèles. Dans le chapitre II, deux modèles de substitution ont été comparés sur l'ensemble des cas d'études : le krigeage et les RBF. Pour la microsource infrarouge, les RBF ont permis de prédire la réponse des sorties du modèle numérique avec plus de précision que le krigeage. Un plan d'expérience construit avec des suites de Sobol' agrégées par k-moyenne a permis de satisfaire les contraintes de l'espace de variation des paramètres. Sur le modèle basse fidélité de la cellule photoacoustique, le krigeage s'est révélé plus précis pour les deux grandeurs physiques étudiées tandis que sur le modèle haute fidélité, les RBF se sont avérées plus intéressantes. Pour le coupleur, les deux métamodèles ont une très mauvaise capacité de prédiction. L'analyse de l'évolution de la surface de réponse selon chaque paramètre a montré que la réponse de la simulation est fortement non linéaire pour de faibles variations de la largeur de chaque guide du coupleur. Cette étude comparative a permis de montrer que les deux métamodèles ont des performances similaires et qu'aucun de ces deux modèles de substitution n'est supérieur à l'autre sur l'ensemble des cas.

Les stratégies concernant l'exploitation des métamodèles, principalement pour différents types d'optimisation, ont été détaillées dans le chapitre III. Ces processus itératifs se basent sur des critères à optimiser afin de déterminer un nouveau point à évaluer par le modèle numérique, permettant d'enrichir localement le métamodèle. L'optimisation à un seul objectif a été réalisée à l'aide du critère d'amélioration espérée pour le krigeage. Lorsqu'une contrainte doit être prise en compte, l'utilisation de la probabilité de faisabilité est ajoutée au sous-problème d'optimisation pour former un problème multi-objectif. Pour les RBF, le critère de Gutmann a été choisi car il est plus efficace que la stratégie CORS. Dans le cadre des problèmes d'optimisation multi-objectif et d'optimisation robuste, le choix s'est concentré sur les critères adaptés au krigeage. L'amélioration espérée de l'hypervolume du front de Pareto (critère EHVI) a été retenue comme étant le plus intéressant dans une stratégie de type MOEGO. Pour l'optimisation robuste, un dérivé de l'amélioration espérée est adapté à toutes les mesures de robustesse envisagée. Dans le cadre de l'optimisation du coupleur, les mesures ont été estimées par Monte Carlo sur le métamodèle. Si des variables d'environnement sont présentes dans le problème, elles sont évaluées là où le krigeage possède la plus haute variance de prédiction. D'un point de vue général, l'ensemble des stratégies adaptatives sont dépendantes de la précision des métamodèles. Les méthodes à base de krigeage présentent l'intérêt d'utiliser l'information probabiliste associée à la prédiction. Dans le cas des méthodes utilisant les RBF, les stratégies reposent essentiellement sur des contraintes de distance avec les points précédemment évalués.

L'étude s'est poursuivie au chapitre IV par l'optimisation des trois dispositifs constituant nos cas d'application. L'optimisation sous contrainte de la microsource infrarouge a permis de mettre en avant une géométrie comportant une seule piste électrique, ce qui permet de réduire la consommation électrique tout en conservant une puissance rayonnée satisfaisante. Les stratégies adaptatives avec RBF ou krigeage fournissent des résultats similaires. Le krigeage conduit toutefois à des optimums légèrement plus intéressants pour un nombre d'appels total au modèle numérique plus faible qu'avec les RBF. Pour réduire le coût de l'optimisation numérique de la cellule photoacoustique, une version multifidélité des RBF a été proposée. Lors de la comparaison avec le co-krigeage, la co-RBF se trouve être moins performante pour la substitution des modèles numériques de la cellule photoacoustique. Une extension du critère Gutmann a également été définie permettant de réaliser une optimisation à partir de la co-RBF. Les optimums obtenus ont des performances comparables à ceux obtenus avec le co-krigeage et le critère d'amélioration espérée. Concernant la conception du coupleur directionnel, l'optimisation robuste a échoué avec la stratégie définie au chapitre III car le métamodèle n'est pas suffisamment précis. Une discrétisation du paramètre de longueur est la solution choisie pour réduire la dimension du problème. A partir de là, une étude sur l'utilisation du krigeage pour variable mixte a été menée dans le cadre de l'encadrement d'un stage de Master 2 recherche. Les résultats montrent un gain de précision sur la prédiction par rapport à la construction de métamodèle indépendant pour chaque longueur d'onde. Pour autant, ce type de métamodèle n'apporte toujours pas d'optimums satisfaisant avec une stratégie d'optimisation robuste. Le développement d'une méthode spécifique a alors été effectué, transformant le sous-problème d'optimisation en un problème multi-objectif. La méthode MGDA est utilisée pour trouver des points Pareto-optimaux à partir du calcul de l'amélioration espérée de chaque objectif. Cette stratégie permet d'identifier des points qui améliorent l'efficacité du coupleur sur toutes les longueurs d'ondes simultanément. Comparée au critère EHVI, cette stratégie permet d'apporter des optimums légèrement plus intéressants. Une série de coupleurs robustes en longueur a alors été fabriquée. Les résultats montrent une uniformité de la réponse du composant sur plusieurs puces malgré les aléas de fabrication et permettent ainsi de valider la méthode de conception.

A travers les travaux présentés précédemment, on remarque qu'un deuxième modèle du coupleur plus précis peut également être construit en utilisant les différences finies dans le domaine temporel. A court terme, cette simulation haute fidélité est extrêmement plus coûteuse que le modèle utilisé actuellement mais quelques évaluations pourraient permettre l'utilisation du co-krigeage et peut-être améliorer la prédiction du métamodèle. Dans tous les cas, les mesures de robustesse sont évaluées sur métamodèle par Monte Carlo car les formules analytiques disponibles sont valables uniquement si les variables sont décorréliées. Sur les applications concernant l'optique intégrée, les incertitudes sur les différentes variables géométriques sont souvent liées au même procédé de fabrication et donc corrélées. Il serait intéressant de développer à moyen terme une stratégie pour calculer des mesures de robustesse autrement que par Monte Carlo sur ce type de cas. Par ailleurs, les polynômes de chaos n'ont pas été étudiés dans ces travaux et pourraient correspondre à une solution intéressante pour cette problématique de propagation des incertitudes. Sur la même échelle de temps et concernant la problématique de substitution des simulations coûteuses, plusieurs stratégies adaptatives utilisant les RBF n'ont pas été implémentés car elles n'étaient spécifiquement développées pour ce métamodèle. Pour autant, elles peuvent être intéressantes lorsque les RBF permettent d'obtenir une prédiction plus précise qu'avec le krigeage. En possibilité de nouveau développement, un critère d'optimisation pour la co-RBF sélectionnant le niveau de fidélité du modèle à évaluer est envisageable. L'intérêt est de réduire le temps de calcul nécessaire à l'optimisation en faisant le moins d'appel possible au modèle haute fidélité si celui-ci est suffisamment corrélé avec le modèle basse fidélité mais qu'il se trouve être bien plus coûteux. A plus long terme, l'étape de calibration des paramètres physiques des modèles numériques à l'aide de données expérimentales doit être traitée. Actuellement, cette opération est effectuée à l'aide des moindres carrés. Il existe une méthode statistique basée sur le théorème de Bayes et reprenant la théorie des processus gaussien qui permet de mieux prendre en compte l'erreur de modélisation et l'erreur de mesure. Avec le grand nombre de puces photoniques et de mesures possibles sur l'ensemble d'un wafer, la calibration bayésienne sera un outil adapté pour mieux recalibrer les paramètres physiques des simulations.

Durant cette thèse, une boîte à outils MATLAB programmée en langage orientée objet a été développée. Elle permet la construction des métamodèles étudiés et la réalisation de différents types d'optimisation. Elle a d'ores et déjà été mise en œuvre dans un certain nombre de projets de recherche technologique et tous les résultats obtenus dans ce document peuvent être reproduits à l'aide de cette boîte à outils. Des détails concernant son implémentation sont fournis en annexe.

---

# Bibliographie

---

- [1] S. Fomel et J. F. Claerbout, « Guest Editors' Introduction: Reproducible Research », *Comput. Sci. Eng.*, vol. 11, n° 1, p. 5-7, févr. 2009.
- [2] G. E. Moore, « Cramming more components onto integrated circuits », *Electronics*, vol. 38, n° 8, 1965.
- [3] M. M. Waldrop, « The chips are down for Moore's law », *Nat. News*, vol. 530, n° 7589, p. 144, 2016.
- [4] T. H. Maiman, « Stimulated Optical Radiation in Ruby Masers », *Nature*, n° 187, p. 493-494, 1960.
- [5] S. E. Miller, « Integrated optics: An introduction », *Bell Syst. Tech. J.*, vol. 48, n° 7, p. 2059-2069, sept. 1969.
- [6] F. P. Kapron, « Radiation Losses In Glass Optical Waveguides », *Appl. Phys. Lett.*, vol. 17, n° 10, p. 423, 1970.
- [7] K. Okamoto, *Fundamentals of optical waveguides*. San Diego: Academic Press, 2000.
- [8] G. Lifante, *Integrated photonics: fundamentals*. Hoboken, NJ: J. Wiley, 2003.
- [9] H. Duprez, « From design to characterization of hybrid III-V on silicon lasers for photonic integrated circuits », Université de Lyon, 2016.
- [10] S. D. Senturia, *Microsystem design*. Boston: Kluwer Academic Publishers, 2001.
- [11] M. Fournier, « Image MEB ».
- [12] J. L. Meek, « A brief history of the beginning of the finite element method », *Int. J. Numer. Methods Eng.*, vol. 39, p. 3761-3774, 1996.
- [13] J. A. Fleck Jr, J. R. Morris, et M. D. Feit, « Time-dependent propagation of high energy laser beams through the atmosphere », *Appl. Phys.*, vol. 10, n° 2, p. 129-160, 1976.
- [14] K. S. Yee et others, « Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media », *IEEE Trans Antennas Propag.*, vol. 14, n° 3, p. 302-307, 1966.
- [15] A. Lefebvre, « Simulation et conception de microsources infrarouges nanophotoniques pour la détection de gaz », Université de Paris-Saclay, 2015.
- [16] L. S. Rothman *et al.*, « The HITRAN2012 molecular spectroscopic database », *J. Quant. Spectrosc. Radiat. Transf.*, vol. 130, p. 4-50, nov. 2013.
- [17] P. Barritault, M. Brun, S. Gidon, et S. Nicoletti, « Mid-IR source based on a free-standing microhotplate for autonomous CO2 sensing in indoor applications », *Sens. Actuators Phys.*, vol. 172, n° 2, p. 379-385, déc. 2011.
- [18] J. F. Creemer *et al.*, « MEMS hotplates with TiN as a heater material », in *IEEE Sensors, 2005.*, 2005, p. 4-pp.
- [19] E. Cozzani *et al.*, « Material properties measurement and numerical simulation for Characterization of ultra-low-power consumption hotplates », in *TRANSDUCERS 2007-2007 International Solid-State Sensors, Actuators and Microsystems Conference, 2007*, p. 1661-1664.
- [20] J. Rouxel, « Conception et réalisation de cellules photoacoustiques miniaturisées pour la détection de traces de gaz », Université de Reims Champagne-Ardenne, 2015.
- [21] V. Zeninari, V. A. Kapitanov, D. Courtois, et Y. N. Ponomarev, « Design and characteristics of a differential Helmholtz resonant photoacoustic cell for infrared gas detection », *Infrared Phys. Technol.*, vol. 40, n° 1, p. 1-23, 1999.
- [22] L. B. Kreuzer, « The Physics of Signal Generation and Detection », in *Optoacoustic Spectroscopy and Detection*, Y.-H. PAO, Éd. Academic Press, 1977, p. 1-25.
- [23] J. Xing *et al.*, « Silicon-on-insulator-based adiabatic splitter with simultaneous tapering of velocity and coupling », *Opt. Lett.*, vol. 38, n° 13, p. 2221, juill. 2013.
- [24] F. Ladouceur et E. Labeye, « A new general approach to optical waveguide path design », *J. Light. Technol.*, vol. 13, n° 3, p. 481-492, 1995.
- [25] H. B. Nielsen, S. N. Lophaven, et J. Søndergaard, *DACE - A Matlab Kriging Toolbox*. Informatics and Mathematical Modelling, Technical University of Denmark, DTU, 2002.
- [26] G. G. Wang et S. Shan, « Review of metamodeling techniques in support of engineering design optimization », *J. Mech. Des.*, vol. 129, n° 4, p. 370-380, 2007.
- [27] J. Friedman, « Multivariate adaptive regression splines », *Ann. Stat.*, vol. 19, n° 1, p. 1-67, 1991.
- [28] V. Vapnik, *The Nature of Statistical Learning Theory*. 1995.
- [29] A. J. Smola et others, « Regression estimation with support vector learning machines », Master's thesis, Technische Universität München, 1996.
- [30] C. Soize et R. Ghanem, « Physical Systems with Random Uncertainties: Chaos Representations with

- Arbitrary Probability Measure », *SIAM J. Sci. Comput.*, vol. 26, n° 2, p. 395-410, janv. 2004.
- [31] Y. Caniou, « Global sensitivity analysis for nested and multiscale modelling », Université Blaise Pascal-Clermont-Ferrand II, 2012.
- [32] H. Niederreiter, *Random Number Generation and Quasi-Monte Carlo Methods*. Philadelphia: Society for Industrial and Applied Mathematics, 1992.
- [33] J. Franco, « Planification d'expériences numériques en phase exploratoire pour la simulation des phénomènes complexes », Ecole Nationale Supérieure des Mines de Saint-Etienne, 2008.
- [34] L. Pronzato, « Plans d'expériences numériques », présenté à PECNUM, 2015.
- [35] D. C. Montgomery, *Design and analysis of experiments*. John Wiley & Sons, 2017.
- [36] F. A. Viana, « Things you wanted to know about the Latin hypercube design and were afraid to ask », in *10th World Congress on Structural and Multidisciplinary Optimization*, 2013, p. 1-9.
- [37] M. D. Morris et T. J. Mitchell, « Exploratory designs for computational experiments », *J. Stat. Plan. Inference*, vol. 43, n° 3, p. 381-402, févr. 1995.
- [38] D. Krige, « A Statistical Approach to Some Basic Mine Valuation Problems on the Witwatersrand », Dezember 1951.
- [39] C. E. Rasmussen et C. K. I. Williams, *Gaussian processes for machine learning*. Cambridge, Mass.: MIT Press, 2006.
- [40] D. Ginsbourger, « Multiples métamodèles pour l'approximation et l'optimisation de fonctions numériques multivariées », Ecole Nationale Supérieure des Mines de Saint-Etienne, 2009.
- [41] V. Dubourg, « Adaptive surrogate models for reliability analysis and reliability-based design optimization », Université Blaise Pascal-Clermont-Ferrand II, 2011.
- [42] E. Vazquez, « Modélisation comportementale de systèmes non-linéaires multivariées par méthodes à noyaux et applications », Université Paris XI - UFR Scientifique D'orsay, 2005.
- [43] O. Dubrule, « Cross validation of kriging in a unique neighborhood », *Math. Geol.*, vol. 15, n° 6, p. 687-699, 1983.
- [44] I. Couckuyt, T. Dhaene, et P. Demeester, « ooDACE toolbox: a flexible object-oriented kriging implementation », *J. Mach. Learn. Res.*, vol. 15, n° 1, p. 3183-3186, 2014.
- [45] R. L. Hardy, « Multiquadric equations of topography and other irregular surfaces », *J. Geophys. Res.*, vol. 76, n° 8, 1971.
- [46] M. D. Buhmann, *Radial Basis Function : Theory and implementattions*, Cambridge University Press. 2003.
- [47] H. Wendland, *Scattered Data Approximation*. Leiden: Cambridge University Press, 2004.
- [48] C. S. Chen, Y. C. Hon, et R. A. Schaback, « Scientific computing with radial basis functions », *Dep. Math. Univ. South. Miss. Hattiesburg MS*, vol. 39406, 2005.
- [49] C. A. Micchelli, « Interpolation of scattered data: Distance matrices and conditionally positive definite functions », *Constr. Approx.*, vol. 2, n° 1, p. 11-22, 1986.
- [50] G. E. Fasshauer, *Meshfree approximation methods with MATLAB*. New Jersey [u.a.: World Scientific, 2010.
- [51] T. J. Mackman, C. B. Allen, M. Ghoreyshi, et K. J. Badcock, « Comparison of Adaptive Sampling Methods for Generation of Surrogate Aerodynamic Models », *AIAA J.*, vol. 51, n° 4, p. 797-808, avr. 2013.
- [52] S. Rippa, « An algorithm for selecting a good value for the parameter c in radial basis function interpolation », *Adv. Comput. Math.*, vol. 11, n° 2-3, p. 193-210, 1999.
- [53] J.-P. Costa, L. Pronzato, et E. Thierry, « A comparison between Kriging and radial basis function networks for nonlinear prediction. », in *Nonlinear Signal and Image Processing*, 1999, p. 726-730.
- [54] G. Fernández-Godino, C. Park, N. H. Kim, et Raphael T. Haftka, « Review of multi-fidelity models », *Struct. Multidiscip. Optim.*, p. 2017.
- [55] L. Laurent, P.-A. Boucard, et B. Soulier, « Generation of a cokriging metamodel using a multiparametric strategy », *Comput. Mech.*, vol. 51, n° 2, p. 151-169, févr. 2013.
- [56] V. Picheny et D. Ginsbourger, « A Nonstationary Space-Time Gaussian Process Model for Partially Converged Simulations », *SIAMASA J. Uncertain. Quantif.*, vol. 1, n° 1, p. 57-78, janv. 2013.
- [57] M. C. Kennedy et A. O'Hagan, « Predicting the Output from a Complex Computer Code When Fast Approximation are Available », *Biometrika*, vol. 87, n° 1, p. 1-13, mars 2000.
- [58] P. Z. G. Qian et C. F. J. Wu, « Bayesian Hierarchical Modeling for Integrating Low-Accuracy and High-Accuracy Experiments », *Technometrics*, vol. 50, n° 2, p. 192-204, mai 2008.
- [59] D. J. Toal, « Some considerations regarding the use of multi-fidelity Kriging in the construction of surrogate models », *Struct. Multidiscip. Optim.*, vol. 51, n° 6, p. 1223-1245, 2015.
- [60] L. Le Gratiet, « Multi-fidelity Gaussian process regression for computer experiments », Université Paris 7, 2013.

- [61] A. I. J. Forrester, N. W. Bressloff, et A. J. Keane, « Optimization using surrogate models and partially converged computational fluid dynamics simulations », *Proc. R. Soc. Math. Phys. Eng. Sci.*, vol. 462, n° 2071, p. 2177-2204, juill. 2006.
- [62] R. ZIMMERMANN et Z.-H. HAN, « Simplified Cross-Correlation Estimation For Multi-Fidelity Surrogate Cokriging Models », *Adv. Appl. Math. Sci.*, vol. 7, n° 2, p. 181-202, 2010.
- [63] A. I. J. Forrester, A. Sóbester, et A. J. Keane, « Multi-fidelity optimization via surrogate modelling », *Proc. R. Soc. Math. Phys. Eng. Sci.*, vol. 463, n° 2088, p. 3251-3269, déc. 2007.
- [64] S. Marmin, D. Ginsbourger, J. Baccou, et J. Liandrat, « Warped Gaussian processes and derivative-based sequential design for functions with heterogeneous variations », *Hal-01507368*, 2017.
- [65] J. Sacks, S. Schiller, et W. Welch, « Designs for computer experiments », *Technometrics*, vol. 31, n° 1, p. 41-47, 1989.
- [66] N. Hansen, « The CMA evolution strategy: A tutorial », vol. 29, 2005.
- [67] M. J. Sasena, « Flexibility and efficiency enhancements for constrained global design optimization with kriging approximations », General Motors, 2002.
- [68] H. J. Kushner, « A New Method of Locating the Maximum Point of an Arbitrary Multipeak Curve in the Presence of Noise », *J. Basic Eng.*, vol. 86, n° 1, p. 97-106, mars 1964.
- [69] B. Bichon, « Efficient surrogate modeling for reliability analysis and design. », Faculty of the Graduate School of Vanderbilt University, Nashville, Tennessee, 2010.
- [70] H.-M. Gutmann, « A radial basis function method for global optimization », *J. Glob. Optim.*, vol. 19, n° 3, p. 201-227, 2001.
- [71] R. G. Regis et C. A. Shoemaker, « Constrained global optimization of expensive black box functions using radial basis functions », *J. Glob. Optim.*, vol. 31, n° 1, p. 153-171, 2005.
- [72] D. R. Jones, « Global optimization with response surfaces », présenté à Fifth SIAM Conference on Optimization, Victoria, Canada, 1996.
- [73] M. Björkman et K. Holmström, « Global optimization of costly nonconvex functions using radial basis functions », *Optim. Eng.*, vol. 1, n° 4, p. 373-397, 2000.
- [74] R. G. Regis et C. A. Shoemaker, « Improved Strategies for Radial basis Function Methods for Global Optimization », *J. Glob. Optim.*, vol. 37, n° 1, p. 113-135, déc. 2006.
- [75] A. I. J. Forrester, A. Sóbester, et A. J. Keane, *Engineering design via surrogate modelling a practical guide*. Chichester, West Sussex, England; Hoboken, NJ: J. Wiley, 2008.
- [76] R. Chocat, L. Brevault, M. Balesdent, et S. Defoort, « Modified Covariance Matrix Adaptation – Evolution Strategy algorithm for constrained optimization under uncertainty, application to rocket design », *Int. J. Simul. Multidiscip. Des. Optim.*, vol. 6, p. A1, 2015.
- [77] R. G. Regis, « Constrained optimization by radial basis function interpolation for high-dimensional expensive black-box problems with infeasible initial points », *Eng. Optim.*, vol. 46, n° 2, p. 218-243, 2014.
- [78] C. Durantin, J. Marzat, et M. Balesdent, « Analysis of multi-objective Kriging-based methods for constrained global optimization », *Comput. Optim. Appl.*, vol. 63, n° 3, p. 903-926, avr. 2016.
- [79] M. Schonlau et W. J. Welch, « Global Optimization with Nonparametric Function Fitting », *Proc. Sect. Phys. Eng. Sci.*, 1996.
- [80] J. M. Parr, A. J. Keane, A. I. J. Forrester, et C. M. E. Holden, « Infill sampling criteria for surrogate-based optimization with constraint handling », *Eng. Optim.*, vol. 44, n° 10, p. 1147-1166, oct. 2012.
- [81] K. Deb, A. Pratap, S. Agarwal, et T. Meyarivan, « A fast and elitist multiobjective genetic algorithm: NSGA-II », *Evol. Comput. IEEE Trans. On*, vol. 6, n° 2, p. 182-197, 2002.
- [82] M. Molga et C. Smutnicki, « Test functions for optimization needs », 2005. [En ligne]. Disponible sur: <http://www.zsd.ict.pwr.wroc.pl/files/docs/functions.pdf>.
- [83] R. T. Marler et J. S. Arora, « Survey of multi-objective optimization methods for engineering », *Struct. Multidiscip. Optim.*, vol. 26, n° 6, p. 369-395, avr. 2004.
- [84] D. Schlierkamp-Voosen et H. Mühlenbein, « Predictive models for the breeder genetic algorithm », *Evol. Comput.*, vol. 1, n° 1, p. 25-49, 1993.
- [85] R. Hinterding, « Gaussian mutation and self-adaption for numeric genetic algorithms », in *Proceedings of 1995 IEEE International Conference on Evolutionary Computation*, 1995, vol. 1, p. 384.
- [86] J.-A. Désidéri, « Révision de l'algorithme de descente à gradients multiples (MGDA) par orthogonalisation hiérarchique », Inria Sophia Antipolis; INRIA, 2015.
- [87] A. Zerbinati, « Algorithme à gradients multiples pour l'optimisation multiobjectif en simulation de haute fidélité: application à l'aérodynamique compressible », Université Nice Sophia Antipolis, 2013.
- [88] K. Shimoyama, K. Sato, S. Jeong, et S. Obayashi, « Comparison of the criteria for updating Kriging response surface models in multi-objective optimization », in *Evolutionary Computation (CEC), 2012 IEEE Congress on*, 2012, p. 1-8.



- [89] J. A. Hartigan et M. A. Wong, « Algorithm AS 136: A K-Means Clustering Algorithm », *J. R. Stat. Soc. Ser. C Appl. Stat.*, vol. 28, n° 1, p. 100-108, 1979.
- [90] T. Akhtar et C. A. Shoemaker, « Multi objective optimization of computationally expensive multimodal functions with RBF surrogates and multi-rule selection », *J. Glob. Optim.*, vol. 64, n° 1, p. 17-32, janv. 2016.
- [91] R. G. Regis, « Multi-objective constrained black-box optimization using radial basis function surrogates », *J. Comput. Sci.*, vol. 16, p. 140-155, sept. 2016.
- [92] J. Knowles, « ParEGO: a hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems », *IEEE Trans. Evol. Comput.*, vol. 10, n° 1, p. 50-66, févr. 2006.
- [93] M. Emmerich, N. Beume, et B. Naujoks, « An EMO algorithm using the hypervolume measure as selection criterion », in *International Conference on Evolutionary Multi-Criterion Optimization*, 2005, p. 62-76.
- [94] M. Emmerich et J. Klinkenberg, « The computation of the expected improvement in dominated hypervolume of Pareto front approximations », *Rapp. Tech. Leiden Univ.*, vol. 34, 2008.
- [95] J. D. Svenson, « Computer experiments: Multiobjective optimization and sensitivity analysis », The Ohio State University, 2011.
- [96] I. Couckuyt, D. Deschrijver, et T. Dhaene, « Fast calculation of multiobjective probability of improvement and expected improvement criteria for Pareto optimization », *J. Glob. Optim.*, vol. 60, n° 3, p. 575-594, nov. 2014.
- [97] P. Feliot, J. Bect, et E. Vazquez, « A Bayesian approach to constrained single- and multi-objective optimization », *J. Glob. Optim.*, vol. 67, n° 1-2, p. 97-133, janv. 2017.
- [98] J. Bect, E. Vazquez, et et al., « STK: a Small (Matlab/Octave) Toolbox for Kriging ». version 2.4.2, 2017.
- [99] L. While, L. Bradstreet, et L. Barone, « A Fast Way of Calculating Exact Hypervolumes », *IEEE Trans. Evol. Comput.*, vol. 16, n° 1, p. 86-95, févr. 2012.
- [100] A. I. J. Forrester et A. J. Keane, « Recent advances in surrogate-based optimization », *Prog. Aerosp. Sci.*, vol. 45, n° 1-3, p. 50-79, janv. 2009.
- [101] J. Guerra, « Optimisation multi-objectif sous incertitudes de phénomènes de thermique transitoire », Institut supérieur de l'aéronautique et de l'espace (ISAE), 2016.
- [102] M. Matsumoto et T. Nishimura, « Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator », *ACM Trans Model Comput Simul*, vol. 8, n° 1, p. 3-30, 1998.
- [103] H.-G. Beyer et B. Sendhoff, « Robust optimization – A comprehensive survey », *Comput. Methods Appl. Mech. Eng.*, vol. 196, n° 33-34, p. 3190-3218, juill. 2007.
- [104] R. E. Moore, R. B. Kearfott, et M. J. Cloud, *Introduction to Interval Analysis*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2009.
- [105] A. N. Kolmogorov, *Foundations of the theory of probability*. Oxford, England: Chelsea Publishing Co., 1950.
- [106] Q. Chen, « Comparing probabilistic and fuzzy set approaches for designing in the presence of uncertainty », Virginia Polytechnic Institute and State University, 2000.
- [107] J. Shafer, « A Mathematical Theory of Evidence. », *Math. Theory Evid.*, 1976.
- [108] A. P. Dempster, « A generalization of bayesian inference », *J. R. Stat. Soc.*, p. 1968.
- [109] V. Baudoui, « Optimisation robuste multiobjectifs par modèles de substitution », Toulouse, ISAE, 2012.
- [110] L. Brevault, « Contributions à l'optimisation multidisciplinaire sous incertitude, application à la conception de lanceurs », EMSE, 2015.
- [111] N. Wiener, « The Homogeneous Chaos », *Am. J. Math.*, vol. 60, n° 4, p. 897, oct. 1938.
- [112] B. Sudret, « Global sensitivity analysis using polynomial chaos expansions », *Reliab. Eng. Syst. Saf.*, vol. 93, n° 7, p. 964-979, juill. 2008.
- [113] G. Blatman, « Adaptive sparse polynomial chaos expansions for uncertainty propagation and sensitivity analysis », Université BLAISE PASCAL - Clermont II, 2008.
- [114] B. J. Williams, T. J. Santner, et W. I. Notz, « Sequential design of computer experiments to minimize integrated response functions », *Stat. Sin.*, p. 1133-1152, 2000.
- [115] S. ur Rehman, M. Langelaar, et F. van Keulen, « Efficient Kriging-based robust optimization of unconstrained problems », *J. Comput. Sci.*, vol. 5, n° 6, p. 872-881, nov. 2014.
- [116] F. Jurecka, « Robust design optimization based on metamodeling techniques », Technische Universität München, 2007.
- [117] J. S. Lehman, T. J. Santner, et W. I. Notz, « Designing computer experiments to determine robust control variables », *Stat. Sin.*, p. 571-590, 2004.
- [118] G. Sun, G. Li, M. Stone, et Q. Li, « A two-stage multi-fidelity optimization procedure for honeycomb-type cellular materials », *Comput. Mater. Sci.*, vol. 49, n° 3, p. 500-511, sept. 2010.

- [119] M. C. Kennedy et A. O'Hagan, « Predicting the Output from a Complex Computer Code When Fast Approximation are Available », *Biometrika*, vol. 87, n° 1, p. 1-13, mars 2000.
- [120] A. I. J. Forrester, A. Sóbester, et A. J. Keane, « Multi-fidelity optimization via surrogate modelling », *Proc. R. Soc. Math. Phys. Eng. Sci.*, vol. 463, n° 2088, p. 3251-3269, déc. 2007.
- [121] S. Xiong, P. Z. G. Qian, et C. F. J. Wu, « Sequential Design and Analysis of High-Accuracy and Low-Accuracy Computer Codes », *Technometrics*, vol. 55, n° 1, p. 37-46, févr. 2013.
- [122] L. Le Gratiet, « Multi-fidelity Gaussian process regression for computer experiments », Université Paris 7, 2013.
- [123] J. Lu et J. Vučković, « Nanophotonic computational design », *Opt. Express*, vol. 21, n° 11, p. 13351, juin 2013.
- [124] K. Hassan, C. Durantin, V. Hugues, B. Szelag, et A. Glière, « Robust silicon-on-insulator adiabatic splitter optimized by metamodeling », *Appl. Opt.*, vol. 56, n° 8, p. 2047, mars 2017.
- [125] L. P. Swiler, P. D. Hough, P. Qian, X. Xu, C. Storlie, et H. Lee, « Surrogate models for mixed discrete-continuous variables », in *Constraint Programming and Decision Making*, Springer, 2014, p. 181-202.
- [126] C. B. Storlie, H. D. Bondell, B. J. Reich, et H. H. Zhang, « Surface estimation, variable selection, and the nonparametric oracle property », *Stat. Sin.*, vol. 21, n° 2, p. 679, 2011.
- [127] P. Z. G. Qian, H. Wu, et C. F. J. Wu, « Gaussian Process Models for Computer Experiments With Qualitative and Quantitative Factors », *Technometrics*, vol. 50, n° 3, p. 383-396, août 2008.
- [128] R. B. Gramacy et H. K. H. Lee, « Bayesian Treed Gaussian Process Models With an Application to Computer Modeling », *J. Am. Stat. Assoc.*, vol. 103, n° 483, p. 1119-1130, sept. 2008.
- [129] O. Roustant *et al.*, « Group kernels for Gaussian process metamodels with categorical inputs », *ArXiv Prepr. ArXiv180202368*, 2018.
- [130] Q. Zhou, P. Z. G. Qian, et S. Zhou, « A Simple Approach to Emulation for Computer Models With Qualitative and Quantitative Factors », *Technometrics*, vol. 53, n° 3, p. 266-273, août 2011.
- [131] J. C. Pinheiro et D. M. Bates, « Unconstrained parametrizations for variance-covariance matrices », *Stat. Comput.*, vol. 6, n° 3, p. 289-296, 1996.
- [132] P. Z. G. Qian, « Sliced Latin Hypercube Designs », *J. Am. Stat. Assoc.*, vol. 107, n° 497, p. 393-399, mars 2012.
- [133] J. Yang, H. Chen, D. K. J. Lin, et M.-Q. Liu, « Construction of sliced maximin-orthogonal latin hypercube designs », *Stat. Sin.*, vol. 26, n° 2, p. 589-603, 2016.
- [134] T. Braure, « Gaussian process with both quantitative and qualitative variables », CEA LETI, Rapport de stage Master 2 recherche, 2017.
- [135] D. Horn, T. Wagner, D. Biermann, C. Weihs, et B. Bischl, « Model-Based Multi-objective Optimization: Taxonomy, Multi-Point Proposal, Toolbox and Benchmark. », in *EMO (1)*, 2015, p. 64-78.
- [136] S. Jeong et S. Obayashi, « Efficient global optimization (EGO) for multi-objective problem and data mining », in *Evolutionary Computation, 2005. The 2005 IEEE Congress on*, 2005, vol. 3, p. 2138-2145.
- [137] D. Ginsbourger, « Multiples métamodèles pour l'approximation et l'optimisation de fonctions numériques multivariées », Ecole Nationale Supérieure des Mines de Saint-Etienne, 2009.
- [138] J. Xing *et al.*, « Silicon-on-insulator-based adiabatic splitter with simultaneous tapering of velocity and coupling », *Opt. Lett.*, vol. 38, n° 13, p. 2221, juill. 2013.

---

## V. Annexe boîte à outil « SBDO »

---

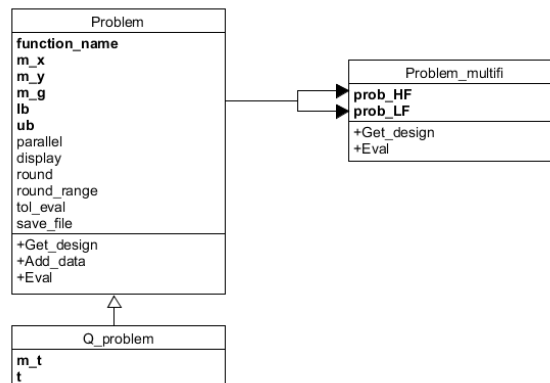
L'ensemble des travaux présentés dans ce document ont été réalisés à partir d'une boîte à outil programmée en langage orienté objet sur MATLAB. Les fichiers sources sont disponibles pour téléchargement à l'adresse suivante :

<https://github.com/freeSBDO/SBDOT>

La boîte à outil SBDO contient des méthodes permettant la construction de métamodèles et leur exploitation à travers des stratégies adaptatives (SBDO signifie Surrogate Based Design Optimization). L'optimisation à un objectif avec ou sans contrainte, l'optimisation multi-objectif et l'optimisation robuste sont réalisables. Les RBF, le krigeage, la co-RBF, le co-krigeage et le krigeage pour variables mixtes sont les métamodèles disponibles. SBDO repose actuellement sur l'implémentation de boîte à outil tel que ooDACE [44] et STK [98]. La présence de fichiers exemples donne à l'utilisateur un aperçu des possibilités. Une documentation est présente dans les fichiers de chaque objet pour comprendre les différentes propriétés et méthodes à partir de commentaires. Le fonctionnement de SBDO impose la définition du problème en premier lieu. Le but est de créer un objet contenant toutes les informations nécessaires pour la construction du métamodèle ou de l'utilisation stratégies adaptatives. A partir de l'objet problème, un plan d'expérience peut être choisi puis évalué par la simulation. Ces données sont toujours sauvegardées dans l'objet et peuvent être enrichies par la suite. Cet objet sert de propriété de base pour les autres objets liés à la construction de métamodèle ou à la réalisation d'une stratégie adaptative.

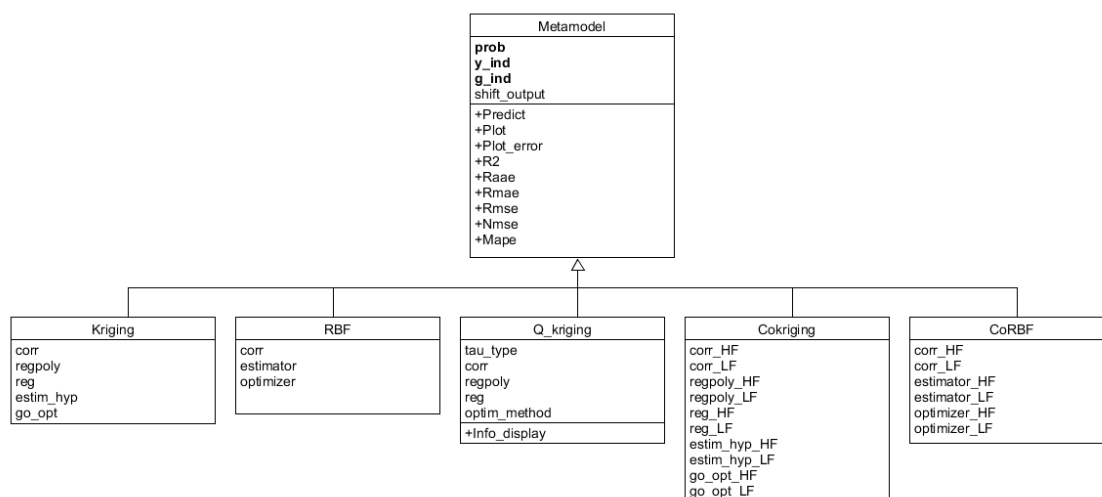
Il existe 3 groupes principaux de classes dans SBDO : les objets pour la définition du problème, ceux pour les métamodèles et ceux pour les stratégies adaptatives. Les principales méthodes et propriétés de ces groupes sont maintenant décrit. Sur la Figure V-1, le diagramme des objets liés à la définition des problèmes est représenté. Les propriétés des objets indispensables à leur construction sont en gras, les méthodes associées aux objets ont un signe + devant leur nom. La classe principale de ce groupe, celle à utiliser dans le cas le plus général, correspond à l'objet Problem. Pour définir un problème, il faut avoir une fonction MATLAB permettant de lancer la simulation numérique à partir d'un choix de paramètre de conception. Si le modèle numérique doit être résolu par un solveur spécifique, la commande « system » de MATLAB permet l'utilisation de ligne de commande UNIX ou WINDOWS qui appelle le programme souhaité. Le nom du fichier MATLAB où réside la procédure de résolution du modèle numérique correspond à la propriété **function\_name**. Le nombre de variables de conception est contenu dans **m\_x**. Le nombre de données physiques récoltées en sortie de simulation qui sont considérées comme des objectifs est contenu dans **m\_y**. Le nombre de contraintes issues de la simulation est contenu dans **m\_g**. Les bornes inférieures et supérieures de variation des paramètres sont respectivement contenues dans **lb** et **ub**. Les autres propriétés sont optionnelles. Il est possible de définir si la simulation peut évaluer plusieurs de jeux de paramètres en même temps (**parallel**). L'affichage d'information dans l'interpréteur MATLAB est défini par **display**. Lorsqu'un arrondi des paramètres à évaluer est nécessaire pour le solveur numérique, les propriétés **round** et **round\_range** permettent de le définir. La tolérance sur la distance d'évaluation entre deux points d'évaluations est liée à **tol\_eval**, cette tolérance est nécessaire pour éviter la construction d'un métamodèle avec des points trop proches entre eux. Enfin, les valeurs des objectifs et contraintes récoltés en sortie de la simulation peuvent sauvegardées à chaque appel selon le réglage de la propriété **save\_file**. Avec un objet Problem, il est possible de créer un plan d'expérience et de l'évaluer à l'aide de la méthode Get\_design. Si des données déjà évaluées par l'utilisateur sont disponibles, la méthode Add\_data permet de les ajouter à l'objet. Enfin, la méthode Eval permet d'évaluer et d'ajouter à l'objet des points spécifiques définis par l'utilisateur. Lorsqu'un problème multifidélité doit être décrit, il faut passer par l'objet Problem\_multifi. Ce dernier se construit à partir de la définition d'un problème classique avec le modèle HF et d'un avec le modèle LF. Pour terminer, l'objet Q\_problem permet la définition de problèmes pour variables mixtes. La propriété **m\_t** correspond au nombre de niveaux de chaque variable discrète et la

propriété **t** correspond aux valeurs que ces dernières peuvent prendre.



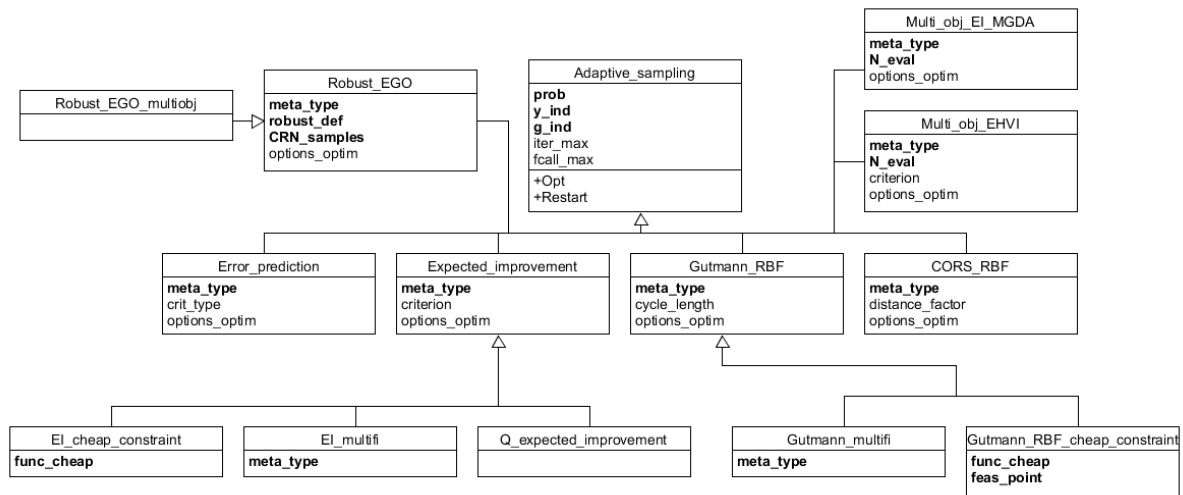
**Figure V-1 : Diagramme des classes liées à la définition du problème. Les flèches noires indiquent une instanciation de l'objet. Les flèches blanches définissent le sens de la classe vers la superclasse.**

Le diagramme du second groupe de classe, à savoir celui des métamodèles, est représenté sur la Figure V-2. La classe principale est définie par l'objet Metamodel, qui n'est jamais directement instancié. Il contient les éléments communs à la construction de tous les métamodèles. La première propriété nécessaire pour la création d'un métamodèle est un objet problème avec des données évaluées (**prob**). Avec celui-ci, il faut préciser la donnée de sortie de la simulation à substituer. Il s'agit de définir l'indice sur l'ensemble des données extraites de la simulation correspondant à l'objectif ou la contrainte d'intérêt (propriété **y\_ind** et **g\_ind**). Ces données de sorties peuvent être remises à une certaine échelle définie par l'utilisateur par l'intermédiaire de la propriété **shift\_output**. La principale méthode applicable avec un métamodèle concerne la prédiction de valeurs en de nouveaux points (Predict). Il est également possible de tracer des vues en coupe 1D ou 2D de la surface de réponse obtenue (Plot). Si des données de validation sont disponibles, des mesures d'erreurs peuvent être calculées (R2, ..., MAPE) ainsi qu'un diagramme comparant les valeurs réelles et les valeurs prédites (Plot\_error). Les objets de chaque métamodèle ont des propriétés optionnelles spécifiques à leur construction, tel que la fonction de corrélation à utiliser pour le krigeage (**corr**) ou l'hypothèse sur la matrice de corrélation inter-niveaux pour le krigeage pour variables mixtes (**tau-type**). Des précisions sur chacune de ces options sont disponibles en commentaires dans les fichiers source des objets. Les propriétés optionnelles ont des valeurs par défaut permettant la construction d'un premier métamodèle assez rapidement.



**Figure V-2 : Diagramme des classes liées à la création d'un métamodèle. Les flèches blanches définissent le sens de la classe vers la superclasse.**

Le dernier groupe de classe concernant les stratégies adaptatives est représenté sur la Figure V-3. La classe principale est défini avec l'objet Adaptive\_sampling. Dans cet objet, c'est la procédure algorithmique de la Figure III-1 qui est implémentée. Le traitement du sous-problème d'optimisation lié au critère est ensuite défini dans chaque sous-classe qui correspondent à une problématique spécifique. La propriété nécessaire à chaque sous-classe est le type de métamodèle utilisé (**meta\_type**). Certaines stratégies ne laissent pas beaucoup de choix, comme la stratégie CORS\_RBF qui est utilisable seulement avec les RBF. Les sous-classes des stratégies permettent la définition de cas particulier. Par exemple la stratégie EI\_cheap\_constraint permet de prendre en compte une contrainte peu coûteuse lors de l'optimisation de l'amélioration espérée. C'est cette stratégie qui a permis l'optimisation de la microsource à rayonnement infrarouge.



**Figure V-3 : Diagramme des classes liées au lancement d'une stratégie adaptative. Les flèches blanches définissent le sens de la classe vers la superclasse.**

Les algorithmes d'optimisations multi-objectif NSGA-II et MGDA sont également disponibles sous la forme d'objet. Ils sont utilisés dans la résolution du sous-problème d'optimisation de certaines stratégies. Une estimation des indices de Sobol' à partir d'un métamodèle est également disponible. Ces indices permettent de mesurer l'impact d'une variable de conception sur la variance de la donnée physique étudiée en sortie de simulation. Des fichiers exemples sont également disponibles sur ces méthodes afin d'appréhender leur utilisation.