



HAL
open science

Planification d'expériences numériques en multi-fidélité : Application à un simulateur d'incendies

Rémi Stroh

► **To cite this version:**

Rémi Stroh. Planification d'expériences numériques en multi-fidélité : Application à un simulateur d'incendies. Autre. Université Paris Saclay (COMUE), 2018. Français. NNT : 2018SACL049 . tel-01936763

HAL Id: tel-01936763

<https://theses.hal.science/tel-01936763>

Submitted on 27 Nov 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.

Planification d'expériences numériques en multi-fidélité. Application à un simulateur d'incendies.

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

École doctorale n°580 Sciences et technologies de l'information et de la
communication (STIC)
Spécialité de doctorat : traitement du signal

Thèse présentée et soutenue à Gif-sur-Yvette, le 26 juin 2018, par

Rémi STROH

Composition du Jury :

Julien BECT Maître de conférences, CentraleSupélec (L2S)	Directeur de thèse
Séverine DEMEYER Docteur-ingénieur, LNE (DMS)	Encadrant
David GINSBOURGER Chercheur senior, Institut de Recherche Idiap (UQOD) Professeur titulaire, Université de Berne (IMSV)	Rapporteur
Amandine MARREL Ingénieur de recherche, CEA (DER/SESI)	Rapporteur
Hervé MONOD Directeur de recherche, INRA (MaIAGE)	Examineur
Bernard PORTERIE Professeur, Université d'Aix-Marseille (IUSTI)	Examineur
Nicolas FISCHER Docteur-ingénieur, LNE (DMS)	Invité et encadrant
Bertrand IOOSS Chercheur senior, EDF R&D	Invité
Damien MARQUIS Docteur-ingénieur, LNE (DCFSI)	Invité et encadrant
Emmanuel VAZQUEZ Maître de conférences, CentraleSupélec (L2S)	Invité et encadrant



École doctorale des Sciences et Technologies de l'Information et de la
Communication (ED-STIC)
Laboratoire des Signaux et Systèmes (L2S), CentraleSupélec, Univ. Paris-Sud,
CNRS, Université Paris-Saclay
Thèse réalisée dans le cadre d'un Contrat Industriel de Formation par la
REcherche (CIFRE) au Laboratoire National de métrologie et d'Essais (LNE)



Remerciements

Avant toute chose, j'aimerais remercier mes directeurs de thèse, Prof Julien BECT et Prof Emmanuel VAZQUEZ, qui m'ont été de précieux conseils durant ces trois ans de thèse. J'aimerais aussi remercier mes encadrants du Laboratoire National de métrologie et d'Essais (LNE), Mme Séverine DEMEYER et M Nicolas FISCHER pour m'avoir guidé cette thèse et me permettre d'aboutir à la situation où je suis actuellement. Je remercie aussi M Damien MARQUIS pour son expertise en sécurité incendie, qui a permis de mettre en place une application pratique des développements de cette thèse.

Je remercie aussi les différents membres du jury pour leurs remarques et suggestions pertinentes. Je remercie Mme Amandine MARREL et M David GINSBOURGER pour avoir relu le manuscrit de thèse et rapporté des suggestions constructives sur le travail. Je remercie M Hervé MONOD et M Bertrand PORTERIE d'être venus examiner mes travaux de recherche pour ma soutenance, et M Bertrand IOOSS d'avoir répondu à l'invitation. Merci de vos critiques, propositions, commentaires et autres remarques.

Je remercie les différents membres du LNE de m'avoir accueilli pendant ces trois ans, et de m'avoir présenté les différentes facettes de leurs métiers, Catherine, Géraldine, Alexandre et Loïc du département mathématiques et statistiques ; Pascal du service qualité ; Gilles du service informatique ; et Cédric, Eric et Simon du département incendie.

Je remercie les différentes personnes travaillant à Supélec : Élisabeth, Arthur, Gilles, José, Laurent, Luc, Jérôme, Laurent et Morgan. Je remercie aussi les doctorant que j'ai pu côtoyer, Antoine et Arnaud en premier lieu.

Je remercie ma famille de m'avoir toujours encouragé.

A tous, merci beaucoup.

Table des matières

Remerciements	v
Table des matières	vii
Table des figures	xi
Liste des tableaux	xv
Liste des abréviations	xvii
Introduction	1
1 Contexte	1
2 Évaluation de la conformité et incertitudes	2
3 Modélisations des simulateurs coûteux	6
4 Planification d’expériences numériques	9
5 Organisation du manuscrit	10
Publications	13
1 Modélisation bayésienne de la multi-fidélité	15
1.1 Introduction	15
1.1.1 Notions de multi-fidélité	15
1.1.2 Approches non-bayésiennes	16
1.2 Processus gaussiens modélisant la multi-fidélité	18
1.2.1 Modèle auto-régressif	18
1.2.2 Modèle non-stationnaire	31
1.2.3 Discussions	42
1.3 Extension aux simulateurs stochastiques multi-fidèles	44
1.3.1 Modèles de la littérature	45

1.3.2	Modèle simplifié dédié à la multi-fidélité	47
1.3.3	Illustrations	50
1.4	Conclusion	60
2	Planification d'expériences en multi-fidélité	63
2.1	Introduction	63
2.2	Plans d'expériences non-séquentiels	65
2.2.1	Plans d'expériences sur un niveau	65
2.2.2	Plans non-séquentiels multi-fidèles de la littérature	66
2.2.3	Plans non-séquentiels emboîtés maximin	69
2.3	Plan séquentiel par réduction progressive de l'incertitude	71
2.3.1	Principe des stratégies par réduction progressive de l'incertitude	71
2.3.2	Critères pour estimer des probabilités de dépassement de seuil	73
2.3.3	Un exemple de stratégie SUR sans multi-fidélité	83
2.3.4	Intégration selon les hyper-paramètres	84
2.4	Planification séquentielle en multi-fidélité	89
2.4.1	Revue de la littérature	89
2.4.2	Taux maximal de réduction de l'incertitude	95
2.4.3	Illustrations	98
2.4.4	Extensions pour la parallélisation	114
2.5	Conclusion	120
3	Application en sécurité incendie	123
3.1	Ingénierie de la sécurité incendie	124
3.2	Description et formalisation du problème	125
3.2.1	Présentation du cas d'étude	125
3.2.2	Propriétés remarquables du modèle numérique	132
3.2.3	Quantités d'intérêt	140
3.3	Illustration de la méthodologie proposée	141
3.3.1	Plan d'expériences initial	141
3.3.2	Échantillonnage des hyper-paramètres	143
3.3.3	Prédiction des sorties	146
3.3.4	Distribution a posteriori d'une probabilité de défaillance	149
3.3.5	Courbes de probabilités de défaillance	152
3.3.6	Planification séquentielle d'expériences	152
3.4	Validation de la méthodologie proposée	158

3.4.1	Comparaison des estimations des hyper-paramètres	158
3.4.2	Validation des prédictions des sorties	159
3.4.3	Validation des estimations de probabilités de références	163
3.4.4	Répétition de la planification séquentielle d'expériences	163
3.4.5	Résultats complémentaires sur la réduction d'incertitude	170
3.5	Conclusion	173
Conclusion		175
1	Résumé et contributions	175
2	Perspectives	177
A Régression par processus gaussien		181
A.1	Formalisation mathématiques	181
A.2	Fonction de covariance de Matérn	185
A.3	Intégration des hyper-paramètres	186
A.3.1	Échantillon initial des hyper-paramètres	187
A.3.2	Mise à jour de l'échantillon des hyper-paramètres	189
B Théorèmes et preuves		191
B.1	Construction de modèles multi-fidèles	191
B.1.1	Principal théorème de construction	191
B.1.2	Modèle auto-régressif	193
B.1.3	Modèle non-stationnaire	196
B.2	Inversion des niveaux	199
B.3	Espérance de probabilités de dépassement de seuil	200
C Méthode d'estimation de la probabilité de défaillance		203
C.1	Hypothèses	203
C.2	Moyenne et variance	204
C.3	Échantillonnage selon la loi a posteriori	206
D Distribution a priori des hyper-paramètres		209
E Figures complémentaires sur la validation de la méthode sur FDS		213
E.1	Comparaison des prédictions et des observations	213
E.2	Évaluation des mesures d'incertitudes en fonction du coût	213
Bibliographie		219

Table des figures

1	Les sorties de FDS en fonction d'une entrée	3
2	Diagramme d'une expérience numérique	4
3	Un exemple de méta-modèles sur les sorties de FDS	7
4	Deux images de FDS	8
5	Des trajectoires conditionnelles issues d'un modèle auto-régressif	22
6	Extrapolation en haute-fidélité avec un modèle auto-régressif	23
7	Deux problèmes jouets en multi-fidélité	25
8	Échantillonnage des paramètres sur un modèle auto-régressif à deux niveaux	27
9	Un exemple de prédiction sur deux niveaux avec le modèle auto-régressif .	28
10	Échantillonnage des paramètres sur un modèle auto-régressif à cinq niveaux	29
11	Un exemple d'extrapolation en haute-fidélité avec le modèle auto-régressif	30
12	Deux régressions selon les niveaux avec deux covariances différentes	34
13	Des trajectoires conditionnelles issues d'un modèle non-stationnaire	34
14	Extrapolation en haute-fidélité avec un modèle non-stationnaire	36
15	Échantillons des paramètres d'un modèle non-stationnaire à deux niveaux	38
16	Un exemple de prédiction sur deux niveaux avec le modèle non-stationnaire	39
17	Échantillons des paramètres d'un modèle non-stationnaire à cinq niveaux	40
18	Un exemple d'extrapolation en haute-fidélité avec le modèle non-stationnaire	41
19	Échantillons des variances du bruit sur un modèle à deux niveaux	51
20	Distributions des variances du bruit d'un simulateur bruité à deux niveaux	52
21	Prédictions sur un simulateur bruité multi-fidèle à deux niveaux	52
22	Exemples de trajectoires du simulateur	54
23	Histogrammes des sorties du simulateur en un point	55
24	Moyennes et variances du simulateur stochastique par oscillations amorties	56
25	Échantillons des variances du bruit sur l'oscillateur aléatoire	57
26	Loi a priori et a posteriori des variances du bruit de l'oscillateur	58
27	Estimations de toutes les variances du bruit	58

28	Prédiction en haute-fidélité de l'oscillateur aléatoire	59
29	Un hypercube latin	66
30	Un exemple de plan d'expériences emboîté	67
31	Méthode de génération de plans emboîtés de Le Gratiet	68
32	Méthode de génération de plans emboîtés de Qian	69
33	Comparaison niveau par niveau des méthodes d'optimisation de plan NLHS	71
34	Un exemple de planification séquentielle d'expériences (première partie) .	85
35	Un exemple de planification séquentielle d'expériences (seconde partie) . .	86
36	Exemple en simulation de la réduction d'incertitude en multi-fidélité . . .	97
37	Exemple de plan séquentiel d'expériences en multi-fidélité	98
38	Un exemple de planification séquentielle sur deux niveaux (première partie)	101
39	Un exemple de planification séquentielle sur deux niveaux (seconde partie)	102
40	Comparaison entre SUR et MRUR sur un exemple à deux niveaux	103
41	Analyse des plans obtenus avec MRUR sur l'exemple à deux niveaux . . .	105
42	Un exemple de planification séquentielle sur dix niveaux (première partie)	107
43	Un exemple de planification séquentielle sur dix niveaux (deuxième partie)	108
44	Un exemple de planification séquentielle sur dix niveaux (troisième partie)	109
45	Un exemple de planification séquentielle sur dix niveaux (quatrième partie)	110
46	Un exemple de planification séquentielle sur dix niveaux (cinquième partie)	111
47	Un exemple de planification séquentielle sur dix niveaux (sixième partie) .	112
48	Comparaison entre SUR et MRUR sur l'oscillateur aléatoire	113
49	Analyse des plans obtenus avec MRUR sur l'oscillateur aléatoire	115
50	Comparaison entre des stratégies à un point et par paquet de points . . .	118
51	Illustration du problème de la planification asynchrone	120
52	Bâtiment étudié	127
53	Taux de Dégagement de Chaleur	129
54	Des sorties de FDS	133
55	Temps de calcul de Fire Dynamics Simulator	134
56	Sorties de FDS en fonction du niveau	136
57	Estimation de l'écart-type de l'effet stochastique en fonction du niveau . .	137
58	Histogramme des résidus normalisées	138
59	Évolutions des différences en fonction du niveau	139
60	Plan initial d'expériences sur FDS	142
61	Des lois a posteriori d'hyper-paramètres sur FDS	144
62	Des chaînes ayant engendré les échantillons d'hyper-paramètres sur FDS .	145
63	Prédictions médianes de FDS en selon deux entrées à 20 cm	147

64	Prédictions médianes de FDS en selon une entrée	148
65	Prédictions médianes de FDS en selon la fidélité	150
66	Densité a posteriori de probabilités de défaillance en un scénario	151
67	Estimations des probabilités de défaillance des sorties de FDS	153
68	Densité a posteriori de probabilités de défaillance après planification	155
69	Front de Pareto successifs des critères sur une expérience	156
70	Comparaison des densités a posteriori avant et après planification	157
71	Estimation des hyper-paramètres sur FDS	160
72	Comparaison entre des prédictions et des observations de FDS	161
73	Comparaison de deux méthodes d'estimations de probabilité de défaillance	164
74	Répétitions d'une planification séquentielle d'expériences sur la température	167
75	Comparaison des méthodes d'estimation de la probabilité sur le plan 3	168
76	Comparaison des méthodes d'estimation de la probabilité sur le plan 4	169
77	Niveaux observés en fonction du coût du plan	171
78	Répétitions d'une planification séquentielle d'expériences sur la visibilité	172
79	Evolution des différentes parties de l'incertitude en fonction du coût	172
80	Un exemple de régression par processus gaussien	184
81	Les fonctions de covariance de Matérn	186
82	Estimation avec intégration des hyper-paramètres	189
83	Comparaison entre des prédictions et des observations de FDS	214
84	Comparaison entre des prédictions et des observations de FDS	215
85	Comparaison entre des prédictions et des observations de FDS	216
86	Comparaison des méthodes d'estimation de la probabilité sur le plan 1	217
87	Comparaison des méthodes d'estimation de la probabilité sur le plan 2	218

Liste des tableaux

1	Présentation des niveaux de fidélité de l'oscillateur aléatoire	106
2	Comparaison de plans d'expériences dans une planification asynchrone . .	119
3	Description des niveaux de fidélité de FDS	130
4	Liste des entrées de FDS	130
5	Variables étudiées et seuils de dangerosité	131
6	Plan d'expériences initial	142
7	Les points de référence pour les prédictions	146
8	Comparaison des estimations entre le plan initial et le plan final	154
9	Erreur de prédiction du modèle sur les trois sorties	162
10	Estimation d'une probabilité de défaillance en un scénario de référence . .	165
11	Nombre de nouvelles observations sur les plans séquentiels d'expériences .	170
12	Amplitude approximative des sorties	209

Liste des abréviations

Laboratoires

LCPP	Laboratoire Centrale de Préfecture de Police
LNE	Laboratoire National de métrologie et d'Essais
NIST	National Institute of Standards and Technology

Logiciels

FDS	Fire Dynamics Simulator
STK	Small Toolbox for Kriging

Sécurité incendie

CFD	Mécanique des Fluides Numérique <i>Computational Fluid Dynamics</i>
CFL	Courant-Friedrichs-Lewy
ERP	Établissement Recevant du Public
HRR	Taux de Dégagement de Chaleur <i>Heat Release Rate</i>
ISI	Ingénierie de la Sécurité Incendie

Statistiques

BF	Basse-Fidélité
CRPS	Score de Probabilité Classée Continue <i>Continuous Ranked Probability Score</i>
DoE	Plan d'Expériences <i>Design of Experiments</i>
HF	Haute-Fidélité
LHS	Hypercube Latin <i>Latin Hypercube Sampling</i>

LOO	Leave-One-Out
MAP	Maximum A Posteriori
MCMC	Monte-Carlo par Chaînes de Markov <i>Markov Chain Monte Carlo</i>
NLHS	Hypercube Latin Emboîté <i>Nested Latin Hypercube Sampling</i>
QoI	Quantité d'Intérêt <i>Quantity of Interest</i>
SMC	Monte-Carlo Séquentiel <i>Sequential Monte-Carlo</i>

Critères d'échantillonnage

AKG	Gradient de Connaissance Approximatif <i>Approximate Knowledge Gradient</i>
EHVI	Augmentation Espérée de l'Hyper-Volume <i>Expected Hyper-Volume Improvement</i>
EI	Amélioration Espérée <i>Expected Improvement</i>
EQI	Amélioration Espérée des Quantiles <i>Expected Quantile Improvement</i>
GP-UCB	Borne Supérieure de Confiance d'un Processus Gaussien <i>Gaussian Process Upper Confidence Bound</i>
IAGO	Approche Informative de l'Optimisation Globale <i>Informational Approach to Global Optimization</i>
IMSE	Erreur Quadratique Moyenne Intégrée <i>Integrated Mean Square Error</i>
MF-GP-UCB	Borne Supérieure de Confiance d'un Processus Gaussien en Multi-Fidélité <i>Multi-Fidelity Gaussian Process Upper Confidence Bound</i>
MRUR	Taux Maximal de Réduction de l'Incertainitude <i>Maximal Rate of Uncertainty Reduction</i>
SUR	Réduction Progressive de l'Incertainitude <i>Stepwise Uncertainty Reduction</i>

Affiliations des membres du jury

CEA	Commissariat à l'Énergie Atomique et aux énergies alternatives
DCFSI	Département du Comportement au Feu et Sécurité Incendie

DER/SESI	Département d'Études des Réacteurs/Service d'Études des Systèmes Innovants
DMS	Département Mathématiques et Statistiques
EDF	Électricité De France
Idiap	Institut d'Intelligence Artificielle Perceptive
IMSV	Institut de Statistique Mathématique et de science Actuarielle <i>Institut für Mathematische Statistik und Versicherungslehre</i>
INRA	Institut national de la recherche agronomique
IUSTI	Institut Universitaire des Systèmes Thermiques Industriels
L2S	Laboratoire des Signaux et Systèmes
MaIAGE	MATHématiques et Informatique Appliquées du Génome à l'Environnement
UQOD	groupe de Quantification de l'Incertitude et de Planification Optimale <i>Uncertainty Quantification and Optimal Design group</i>

Introduction

Sommaire

1	Contexte	1
2	Évaluation de la conformité et incertitudes	2
3	Modélisations des simulateurs coûteux	6
4	Planification d'expériences numériques	9
5	Organisation du manuscrit	10

1 Contexte

Selon la réglementation française, les bâtiments, en particulier les Établissements Recevants du Public (ERP), doivent disposer de systèmes de désenfumage suffisamment performants pour pouvoir extraire les fumées en début d'incendie de sorte à limiter la propagation du feu et à assurer l'évacuation du public. Le dimensionnement d'un système de désenfumage d'un ERP est dit conforme lorsqu'il suit des instructions techniques, ou à défaut, lorsqu'une étude d'Ingénierie de la Sécurité Incendie (ISI) montre que le système respecte des objectifs de sécurité (voir par exemple Thiry-Muller [2017]). Ce second cas s'applique quand la réglementation descriptive est difficile d'application, par exemple dans des bâtiments non conventionnels comme des monuments historiques.

Le Laboratoire National de métrologie et d'Essais (LNE) fait partie des organismes reconnus compétents en ingénierie du désenfumage en France. La principale fonction de ces organismes lors des études est d'évaluer les performances des systèmes vis-à-vis des objectifs de sécurité, et éventuellement, dans le cas où les performances ne sont pas satisfaisantes, de proposer des solutions alternatives. Pour évaluer les performances, ces organismes doivent, entre autres choses, proposer, puis simuler numériquement un certain nombre de scénarios d'incendie. L'analyse des résultats de la simulation permet de vérifier si les objectifs sont atteints.

Dans une étude d'ISI, les objectifs de sécurité sont principalement liés à la protection des individus. En pratique, les études ISI se focalisent sur l'évolution de quantités physiques, comme par exemple, la visibilité, la température et le flux radiatif des fumées, évaluées à 1,8 m du sol, à hauteur d'homme. Ces quantités sont comparées à des seuils de tenabilité définis dans la norme ISO 13571 [2012], qui correspondent aux limites physiques que peut supporter un corps adulte avant de subir des dommages importants. Ainsi, la visibilité est liée à la perte d'orientation lors de l'évacuation et aux intoxications, le flux radiatif aux brûlures de la peau et la température aux brûlures internes, notamment de l'appareil respiratoire [Guillaume, 2006]. En d'autres termes, une étude ISI cherche à s'assurer que les quantités ne dépassent pas leurs seuils de tenabilités dans les différents scénarios considérés, et si oui, en quel instant et en quel endroit.

Les études de conformité d'un système de désenfumage s'effectuent généralement à partir de modèles numériques adaptés pour traiter les problématiques incendies. L'un des simulateurs numériques utilisés par la communauté scientifique est Fire Dynamics Simulator (FDS) [McGrattan et al., 2014]. Ce simulateur s'appuie sur un modèle dit de champs, utilisant la Mécanique des Fluides Numérique (CFD). Il découpe l'espace en mailles parallélépipédiques, et résout à chaque pas de temps une forme des équations de Navier-Stokes, dite à bas nombre de Mach, par des méthodes de différences finies. Pour ce type de modèles, les coûts en temps de calcul sont extrêmement élevés. En fonction du scénario, des tailles des mailles et du nombre de processeurs utilisés, les temps de simulation peuvent varier de quelques jours à quelques semaines.

Une simulation réalisée avec FDS permet d'obtenir l'évolution des champs de variables physiques en fonction du temps et de l'espace. Ceci permet deux types d'analyses : les analyses qualitatives, qui concernent l'identification d'événements survenant lors de l'incendie (écoulements, stagnation, tourbillons, hausse de pression...); et les analyses quantitatives, qui chiffrent les performances du système de désenfumage. Afin de simplifier l'analyse quantitative, les résultats spatio-temporels sont post-traités de façon à en extraire quelques variables d'intérêt que, dans cette thèse, nous appelons par abus de langage « les sorties de FDS ». La conformité finale du système de désenfumage est ainsi évaluée par comparaison de ces sorties à leurs seuils de tenabilité respectifs.

2 Évaluation de la conformité et incertitudes

Les scénarios sont définis par un vecteur de variables x en entrée, le reste de la configuration étant fixé. À partir d'un vecteur de variables en entrée, le modèle numérique considéré construit la configuration associée, simule le développement de l'incendie à

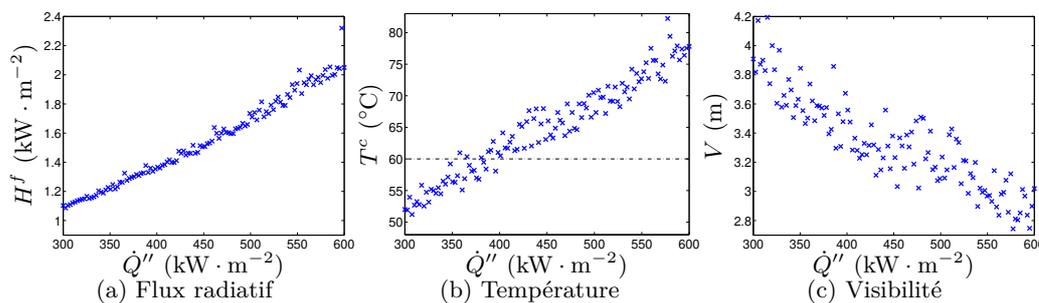


FIGURE 1 – Les trois sorties de FDS en fonction de la puissance surfacique du feu \dot{Q}'' , l'une des entrées de FDS. Les croix bleues sont des résultats issus de FDS.

partir de FDS, exécute le post-traitement, et renvoie les sorties de FDS. En pratique, la conformité s'évalue par le fait que les sorties de FDS restent en-deçà de leur seuils critiques. Dans le domaine de la sécurité incendie, les incertitudes ne sont généralement pas prises en compte dans les évaluations de conformité. Pourtant, comme le soulignent par exemple McGrattan et Toman [2011], cette approche n'est pas satisfaisante car, à cause de deux sources d'incertitudes, une en entrée et une en sortie, la notion de dépassement de seuil n'est pas forcément une réponse booléenne.

En entrée, certaines variables dans la simulation sont généralement fixées, alors que leurs valeurs sont incertaines, comme les variables météorologiques (température, pression, humidité...). Ces variables dites environnementales possèdent une certaine plage de variabilité qui entraîne une incertitude sur le dépassement ou non du seuil de tenabilité en un scénario donné. Pour modéliser les incertitudes en entrées, une solution utilisée par exemple dans Au et Beck [2001], Le Maître et Knio [2010] et Pasanisi et Dutfoy [2012] est de définir une loi de probabilité f^{env} sur l'espace des entrées, dont l'élicitation est basée sur l'avis des experts. Les sorties de FDS sont, quant à elles, considérées comme « stochastiques », dans le sens où deux simulations avec des variables d'entrées proches peuvent aboutir à des résultats très différents. Un tel comportement est parfois appelé « effet de pépite ». Ce comportement est visible figure 1, où des simulations ont été réalisées en faisant varier la puissance surfacique du terme source \dot{Q}'' entre 300 et 600 $\text{kW} \cdot \text{m}^{-2}$. Une méthode pour prendre en compte cette incertitude en sortie dans les évaluations est d'ajouter une loi de probabilité f_x^{sim} en sortie dépendant des entrées (voir, par exemple, Ankenman et al. [2010], Dubourg et al. [2011] ou Oakley et Youngman [2017]).

Dans ces conditions, une solution pour décrire la conformité en présence d'incertitude

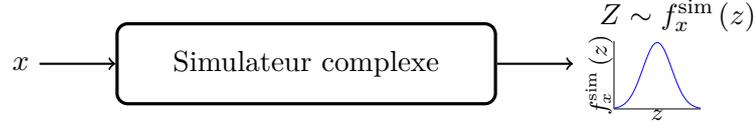


FIGURE 2 – Diagramme d’une expérience numérique.

est d’utiliser une probabilité de dépassement de seuil de tenabilité (voir, par exemple, Kennedy et al. [1980], Schueremans [2001] ou Zio [2013]). Pour définir la probabilité d’intérêt, le vecteur des entrées x est décomposé en deux parties : les variables environnementales x_e aléatoires, qui sont distribuées une distribution d’entrée f^{env} , et les variables de scénario x_s , qui sont des variables d’intérêt décrivant l’incendie. La séparation entre les variables environnementales et les variables de scénario est déterminée par l’utilisateur et dépend exclusivement de l’étude ISI. Le résultat d’une simulation du modèle numérique est un vecteur de sorties scalaires. Une sortie Z est supposée être tirée selon une loi de sortie dépendant des entrées f_x^{sim} (voir figure 2). Soit z^{crit} le seuil de tenabilité associé à la sortie Z . Nous supposons que la défaillance se produit quand la sortie Z dépasse son seuil critique z^{crit} . Dans le cas où la défaillance se produit quand la sortie Z est plus petite que son seuil, il est possible de revenir au cas prétendant en calculant la probabilité que la valeur $-Z$ excède le seuil $-z^{\text{crit}}$. Nous définissons alors la probabilité de dépassement de seuil locale comme la probabilité que la sortie Z associé à l’entrée $x = (x_e, x_s)$ dépasse son seuil :

$$p(x_e, x_s) = \mathbb{P}\left(Z > z^{\text{crit}} \mid x_e, x_s\right) = \int_{z^{\text{crit}}}^{+\infty} f_{x_e, x_s}^{\text{sim}}(z) dz. \quad (1)$$

Notre quantité d’intérêt est la probabilité de dépassement de seuil intégrée selon les variables environnementales :

$$P(x_s) = \mathbb{P}\left(Z > z^{\text{crit}} \mid x_s\right) = \int_{\mathbb{X}_e} p(x_e, x_s) f^{\text{env}}(x_e) dx_e, \quad (2)$$

qui prend en compte deux sources d’aléas : celle des entrées environnementales, et celle de la sortie, considérée comme stochastique. Ces quantités sont des fonctions, au nombre de une par sortie, définies sur l’espace des variables de scénario. Par concision, nous nommons ces fonctions « probabilités de défaillance ». Nous cherchons à estimer ces fonctions de probabilités de défaillance.

Pour estimer des probabilités de dépassement de seuil, plusieurs solutions existent

[Li, 2012]. Une méthode répandue en ingénierie de la sécurité incendie est l'estimation par Monte-Carlo. Par exemple, Kong et al. [2012] utilisent cette méthode pour obtenir la fonction de répartition des temps d'atteinte des seuils critiques dans un bâtiment. Dans un contexte différent, Hostikka et al. [2005] appliquent cette méthode pour mesurer les risques incendies en fonction du temps. Des méthodes plus sophistiquées sont également utilisées pour limiter le nombre de simulations nécessaire. Par exemple, Au et al. [2007] utilisent une méthode de Monte-Carlo fondée sur de l'échantillonnage préférentiel pour estimer des probabilités d'évènements rares. Demeyer et al. [2017] estiment une fonction de probabilité de défaillance en fonction du scénario étudié en combinant une modélisation par processus gaussien et une estimation des probabilités par Monte-Carlo.

En effet, compte-tenu des temps de calcul des modèles CFD tels que FDS et de la puissance informatique nécessaire, la méthode de Monte-Carlo n'est actuellement appliquée que sur des simulateurs d'incendie plus simples, dits de zone, dont les temps de calculs sont de l'ordre d'une fraction de seconde. En effet, la méthode de Monte-Carlo nécessite un grand nombre de simulations. Avec cette méthode, quand une estimation \hat{p} d'une probabilité p est obtenue, le coefficient de variation c vaut

$$c = \frac{\sqrt{\text{var}[\hat{p}]}}{p} = \sqrt{\frac{p(1-p)}{np^2}} = \sqrt{\frac{1-p}{np}}. \quad (3)$$

Si ce coefficient de variation c est fixé, il est possible d'obtenir le nombre n théorique de simulations nécessaires :

$$n = (1-p)/(pc^2). \quad (4)$$

Pour évaluer la probabilité en M scénarios, le nombre total de simulations est de l'ordre de $\sum_{k=1}^M (1-p_k)/(p_k c^2)$, où p_k est la probabilité de défaillance du k -ième scénario étudié. Par exemple, pour $M = 19$ scénarios de probabilités respectives $p_k = k \cdot 5\% = 5\%, 10\%, \dots, 95\%$ et une précision relative de $c = 10\%$, le nombre total de simulations nécessaires est d'environ 5200. Si cela est réalisable sur des modèles numériques de zone, cela est difficilement applicable sur des modèles numériques de champs. Par exemple, en supposant qu'il faille deux jours par simulation, et que celles-ci puissent être exécutées en parallèle sur cent cœurs, le temps nécessaire serait de $5200/100 \times 2$ jours = 104 jours, soit un peu plus de trois mois de calculs ininterrompus. Pour cette raison, l'utilisation d'une telle méthode est inadaptée et le développement de méthodes particulières dédiées à l'estimation de probabilité sur des simulateurs coûteux est nécessaire.

3 Modélisations des simulateurs coûteux

En raison des coûts de calcul du simulateur, une alternative à l’approche par Monte-Carlo consiste à utiliser un modèle de substitution ou méta-modèle. Le principe est de remplacer le simulateur par une approximation moins coûteuse et de mener les analyses sur cette approximation. Différentes méthodes existent pour construire ces approximations, comme par exemple, les régressions polynomiales, les fonctions à base radiale, le krigeage, les réseaux de neurones. . . Des listes non-exhaustives de modèles de substitution peuvent être trouvées dans Forrester et Keane [2009] et dans Koziel et al. [2011]. Dans le domaine incendie, l’usage de méta-modèles peut se trouver par exemple dans Van Weyenberge et al. [2016], qui développent une méthode d’estimation des dangers fondées sur des polynômes du chaos, ou dans Arnold et Arnold [2017] utilisent une régression linéaire pour évaluer les capacités de désenfumage d’une pièce. Ces méthodes ont comme intérêt de permettre l’analyse des simulateurs coûteux. Leur inconvénient est que leur utilisation rajoute une incertitude due à la différence entre le simulateur et le modèle de substitution. Dans la liste des sources d’incertitudes établie par Kennedy et O’Hagan [2001], cette nouvelle source d’incertitude est qualifiée « d’incertitude du code » et est liée à la méconnaissance du modèle numérique. Cette incertitude du code signifie que le résultat exact du programme est inconnu du fait de sa complexité, et exclue l’incertitude qui apparaît du fait de l’écart entre le résultat du simulateur et le résultat d’une expérience réelle (qui ne sera pas traitée dans cette thèse).

Dans cette thèse, la méthode de substitution utilisée est fondée sur des méthodes bayésiennes, qui ont le principal intérêt de simplifier l’estimation des incertitudes dues à la méconnaissance du modèle numérique. Dans ce type d’approche, les quantités inconnues sont associées à des lois de probabilité dites a priori, représentant les valeurs crédibles qu’elles peuvent prendre. Puis, en utilisant la formule de Bayes, ces lois sont mises à jour quand des observations, dans notre cas, des couples entrées-sorties du simulateur $(x_i, z_i)_{1 \leq i \leq n}$, sont effectuées. Ces lois dites a posteriori représentent les valeurs possibles compte-tenu des hypothèses et des observations. De ces lois a posteriori, il est possible d’en extraire des estimateurs et des mesures d’incertitudes. Plus d’informations sur les méthodes bayésiennes peuvent être trouvées par exemple dans Gelman et al. [2013]. Dans le cas où la quantité inconnue est une fonction, un modèle bayésien classique est le processus gaussien [Currin et al., 1991; Rasmussen et Williams, 2006; Santner et al., 2003; Stein, 1999; Welch et al., 1992]. Ce processus a priori permet de modéliser simplement les valeurs que peut prendre la fonction ainsi que les relations entre les évolutions des entrées et celle des sorties. De plus, les processus gaussiens simplifient

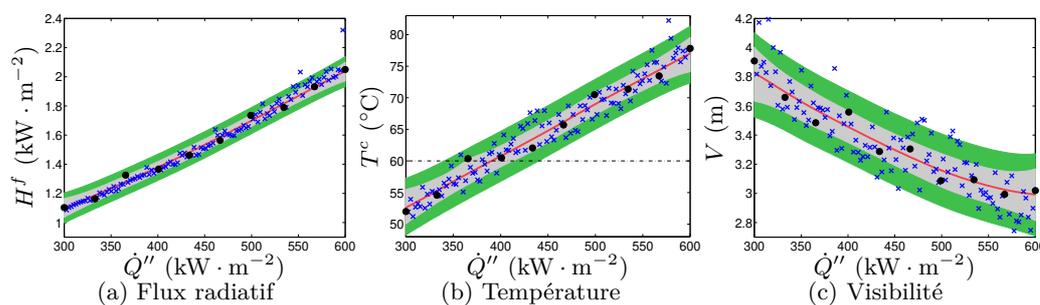


FIGURE 3 – Un exemple de prédictions par processus gaussiens sur les trois sorties de FDS. Les sorties sont représentées en fonction de la puissance surfacique du feu \dot{Q}'' . Les croix bleues sont des vrais résultats de FDS. Les points noirs sont les observations utilisées pour construire la prédiction. Le trait continu rouge est la médiane a posteriori, la zone intérieure grise est l'intervalle de crédibilité à 95 % sur cette prédiction et la zone extérieure verte est l'intervalle de crédibilité à 95 % de prédiction sur les observations en tenant compte de la partie stochastique du simulateur.

les calculs des lois a posteriori dans le cas où la loi de sortie f_x^{sim} est une loi normale, puisque la loi a posteriori est alors un processus gaussien dont les formules sont données par les formules du krigeage (voir Sacks et al. [1989] pour les formules dans un cadre déterministe et Ankenman et al. [2010]; Huang et al. [2006] dans un cadre stochastique). Un exemple de résultats de régression par processus gaussiens sur les trois sorties de FDS est affiché figure 3. Pour chaque sortie, le processus a posteriori est calculé à partir des dix observations, et permet de déduire une estimation par la médiane a posteriori et une mesure d'incertitude par l'intervalle de crédibilité à 95 %. Ce type de modèle est dit boîte noire, car il ne prend en compte que les entrées et les sorties du simulateur, en ignorant l'implémentation interne.

Pour débiter une régression par processus gaussien sur un simulateur déterministe (sans effet stochastique en sortie), une recommandation classique de Loepky et al. [2009] est d'observer environ $10d$ fois le simulateur, où d est la dimension du vecteur d'entrée x . Cependant, comme FDS est vu comme un simulateur stochastique et que les simulations sont particulièrement coûteuses, il est difficile d'appliquer directement cette recommandation. Pour économiser du temps de calcul, une solution consiste à utiliser une approche multi-fidèle. La multi-fidélité est une solution qui combine des simulations basse-fidélité (rapides mais grossières) à des simulations haute-fidélité (fiables mais coûteuses). Dans le cas de FDS, qui procède par différences finies, la fidélité peut être évaluée par l'intermédiaire de la taille de maille δ . En général, pour des simulateurs

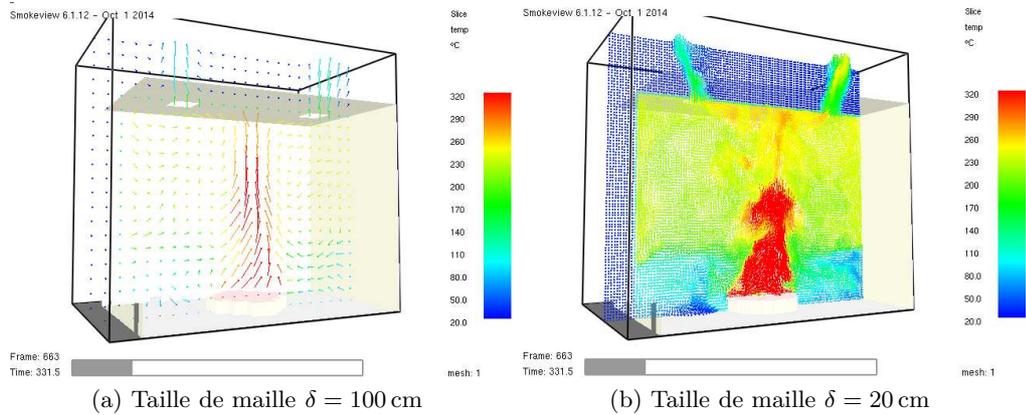


FIGURE 4 – Deux images de la même expérience au même instant de FDS, à deux tailles de maille différentes.

de champs, une simulation est plus fidèle lorsque la taille de maille δ est plus petite. La figure 4 illustre l’aspect multi-fidèle de FDS en montrant deux images prises au même instant d’une simulation de FDS avec les mêmes paramètres d’entrées x , à deux tailles de maille différentes.

Plusieurs auteurs ont déjà combiné l’approche multi-fidèle avec des modèles fondés sur des processus gaussiens. Le modèle de processus gaussiens multi-fidèle de référence est le modèle de Kennedy et O’Hagan [2000], qui relie deux niveaux de fidélité successifs par une relation auto-régressive. Ce modèle connaît beaucoup d’applications [Forrester et al., 2007; Perdikaris et Karniadakis, 2016; Wankhede, 2012]. Il existe également plusieurs développements qui se fondent sur ce modèle. En particulier, Qian et Wu [2008] construisent une approche complètement bayésienne, et Le Gratiet [2013] étend ce modèle à un nombre quelconque de niveaux de fidélité avec des relations plus complexes entre les niveaux. Il existe un autre modèle de processus gaussiens dédié à multi-fidélité, développé indépendamment par Picheny et Ginsbourger [2013] et Tuo et al. [2014]. Ce second modèle décrit le simulateur comme un processus idéal, correspondant à une observation obtenue après un temps de simulation infini, plus une erreur numérique indépendante qui s’ajoute. Les deux modèles relient différemment les niveaux de fidélité. Il convient alors de connaître les propriétés de chacun pour pouvoir choisir celui qui est le mieux adapté à notre situation. Il est à noter aussi qu’il n’existe pas dans la littérature en multi-fidélité de modèle bayésien dédié à la modélisation de la partie stochastique.

4 Planification d'expériences numériques

Le modèle bayésien multi-fidèle permet d'obtenir des estimations de quantités d'intérêt, comme des probabilités de défaillance, avec une incertitude associée à partir d'un nombre n d'observations $\chi_n = (x_i, \delta_i, z_i)_{1 \leq i \leq n}$. Cette incertitude finale associée à l'estimation de la quantité d'intérêt trouve son origine dans trois sources d'incertitude fondamentale :

- l'incertitude en entrée du simulateur ;
- l'incertitude en sortie du simulateur ;
- l'incertitude due à l'utilisation d'une approximation à la place du simulateur.

Les deux premières sont irréductibles, puisqu'elles dépendent du problème. La dernière dépend quant à elle du nombre et du choix des points d'observation $(x_i, \delta_i)_{1 \leq i \leq n}$. Afin de réduire l'incertitude sur la quantité d'intérêt en limitant le temps de simulation nécessaire, il convient de choisir judicieusement les points d'observation. Les méthodes permettant de choisir ces points entrent dans le domaine de la planification d'expériences numériques. Cette thèse distingue deux types de plans d'expériences : les plans d'expériences séquentiels et les plans d'expériences non-séquentiels. Le principe des plans d'expériences séquentiels est de choisir le nouveau point d'observation (x_{n+1}, δ_{n+1}) en utilisant les n observations déjà disponibles. Pour cela, une méthode est de définir un critère d'échantillonnage $J(x, \delta; \chi_n) = J_n(x, \delta)$, et de choisir le point d'observation qui minimise ce critère d'échantillonnage

$$(x_{n+1}, \delta_{n+1}) = \arg \min_{x, \delta} J_n(x, \delta). \quad (5)$$

Le critère d'échantillonnage définit l'intérêt d'observer le point (x, δ) . Une façon de construire ce critère d'échantillonnage est d'utiliser le principe de la Réduction Progressive de l'Incertitude (SUR) [Fleuret et Geman, 1999; MacKay, 1992]. Le principe d'une méthode SUR est de sélectionner le point d'observation qui réduit au plus une mesure d'incertitude, cette dernière étant liée à l'objectif qui est recherché. Villemonteix et al. [2008] et Vazquez et Bect [2009] proposent d'appliquer la méthodologie SUR sur des modèles fondés sur des processus gaussiens pour construire des critères d'échantillonnages dédiés à des objectifs précis, bien que des articles antérieurs proposent des critères d'échantillonnage, qui peuvent être déduits de ce principe [Frazier et al., 2009; Jones et al., 1998]. Dans la littérature, des critères d'échantillonnage sont construits à partir de cette méthodologie [Chevalier et al., 2014; Picheny et al., 2010]. Une solution simple pour construire une planification séquentielle d'expérience est de définir de nouveaux

critères SUR dédiés à l'évaluation des probabilités de défaillance dans notre contexte.

Cependant, la méthode SUR n'est pas recommandable en multi-fidélité. En effet, l'intérêt d'observer un niveau de fidélité augmente avec le coût calculatoire. Il y a un compromis à trouver entre le coût calculatoire à observer un niveau de fidélité et le gain d'information que l'observation apporte. Pour résoudre ce dilemme, Picheny et al. [2013] et Zhou et al. [2017] proposent de ne comparer que des plans d'expériences ayant le même coût calculatoire. Ainsi, si le niveau haute-fidélité est q fois plus cher que le niveau basse-fidélité, ils comparent un plan d'expériences avec q points basse-fidélité et un plan d'expérience avec un point haute-fidélité. Une autre méthode consiste à incorporer le coût dans le critère d'échantillonnage. Une façon de définir un critère d'échantillonnage prenant en compte les coûts d'observation est de considérer un rapport entre l'intérêt d'observer le niveau et le coût d'observation, le point d'observation choisi étant celui qui maximise ce rapport. Huang et al. [2006], Swersky et al. [2013], Le Gratiet et Cannamela [2015] et He et al. [2017] adaptent ainsi des critères simples niveaux à des configurations où le coût d'observation dépend du vecteur des entrées. Cependant, ces méthodes de construction restent très heuristiques, et sont difficilement généralisables à des situations autres que celles des articles où elles sont développées. D'une part, les critères d'échantillonnage développés sont dédiés soit à la recherche du minimum d'une fonction, soit à la construction d'une approximation d'une fonction coûteuse, ce qui ne correspond pas à notre situation. Et d'autre part, les algorithmes de planification séquentielle proposés par Le Gratiet et Cannamela [2015] et par He et al. [2017] reposent sur les propriétés particulières du modèle bayésien utilisé. Comme nous cherchons à planifier les expériences sur différents types de simulateurs, modélisés avec différents modèles de processus gaussiens, nous souhaitons disposer d'une méthodologie générale pour construire des critères d'échantillonnage dans le cas où le coût d'observation varie avec les entrées.

5 Organisation du manuscrit

La suite du mémoire est organisée comme suit.

Le chapitre 1 présente la modélisation bayésienne de la multi-fidélité. Le chapitre présente d'abord les deux principaux modèles de processus gaussiens dédiés à la multi-fidélité. Les définitions ainsi que les principales propriétés des deux modélisations sont présentées et illustrées sur des exemples académiques. Ensuite, ce chapitre continue par l'une des contribution de cette thèse : une extension des modèles de processus gaussiens aux simulateurs stochastiques. La modélisation bayésienne complète d'un simulateur multi-fidèle stochastique est illustrée sur deux exemples académiques. Ce chapitre

reprend des travaux présentés dans [Stroh et al., 2016a, comm1], [Stroh et al., 2016, conf1], [Stroh et al., 2017a, conf2].

Le chapitre 2 se focalise sur les plans d'expériences, avec en particulier la prise en compte du coût de calcul des observations. Ce chapitre présente rapidement les plans multi-niveaux emboîtés non-séquentiels, qui sont utilisés comme plans initiaux d'expériences. En particulier, cette section présente un nouveau résultat sur la façon de construire ces plans. Puis, il continue sur un rappel sur les stratégies de SUR, et présente des nouveaux critères d'échantillonnage qui sont des extensions aux cas stochastiques de critères SUR dédiés à l'estimation de probabilités de dépassement de seuil. La dernière partie présente le principal apport de cette thèse : une méthode générale pour construire un stratégie séquentielle de plans d'expérience sous l'hypothèse d'un coût d'observation variable. Cette stratégie est illustrée sur deux exemples académiques. Ces résultats ont déjà été présentés dans [Stroh et al., 2017, comm3], [Stroh et al., 2017b, conf3] et font l'objet d'un article de journal en préparation [Stroh et al., 2018, arti2].

Le chapitre 3 illustre les méthodes présentées dans les chapitres précédents sur un exemple d'étude ISI utilisant FDS. L'exemple consiste à évaluer la conformité d'un système de désenfumage d'un bâtiment existant. La première partie de ce chapitre commence par décrire le problème et présente les caractéristiques de FDS. La seconde partie applique le modèle et les méthodes décrites dans les chapitres 1 et 2. En particulier, cette partie présente l'application d'un modèle multi-fidèle pour estimer les sorties et les probabilités de dépassement de seuil. Il poursuit avec l'illustration de la méthode de planification séquentielle développée dans le cas de coût d'observation variable. Ce chapitre reprend la présentation de [Stroh et al., 2016b, comm2] et la publication de [Stroh et al., 2017, arti1].

Finalement, le dernier chapitre récapitule les conclusions majeures et les résultats, puis trace les perspectives de ces travaux.

Publications

Articles de journaux

- [artil] Rémi Stroh, Julien Bect, Séverine Demeyer, Nicolas Fischer, Damien Marquis, et Emmanuel Vazquez. Assessing fire safety using complex numerical models with a Bayesian multi-fidelity approach. *Fire Safety Journal*, 91 :1016–1025, 2017. doi : <https://doi.org/10.1016/j.firesaf.2017.03.059>. Fire Safety Science : Proceedings of the 12th International Symposium.
- [arti2] Rémi Stroh, Julien Bect, Séverine Demeyer, Nicolas Fischer, et Emmanuel Vazquez. Bayesian sequential estimations supervised by the maximum rate of uncertainty reduction for multi-fidelity simulators. *en écriture*, 2018.

Articles de conférences

- [conf1] Rémi Stroh, Julien Bect, Séverine Demeyer, Nicolas Fischer, et Emmanuel Vazquez. Gaussian process modeling for stochastic multi-fidelity simulators, with application to fire safety. In 48^{emes} *Journées de Statistique de la SFdS (JdS 2016)*, Montpellier, France, mai 2016.
- [conf2] Rémi Stroh, Julien Bect, Séverine Demeyer, Nicolas Fischer, et Emmanuel Vazquez. Integrating hyper-parameter uncertainties in a multi-fidelity Bayesian model for the estimation of a probability of failure. In 11th *Advanced Mathematical and Computational Tools in Metrology and Testing Conference (AMCTM 2017)*, Glasgow, Écosse, août 2017a.
- [conf3] Rémi Stroh, Séverine Demeyer, Nicolas Fischer, Julien Bect, et Emmanuel Vazquez. Sequential design of experiments to estimate a probability of exceeding a threshold in a multi-fidelity stochastic simulator. In 61st *World Statistics Congress (ISI 2017)*, Marrakech, Maroc, juillet 2017b.

Autres communications

- [comm1] Rémi Stroh, Julien Bect, Séverine Demeyer, Nicolas Fischer, et Emmanuel Vazquez. Design and analysis of multi-level numerical experiments, with application to fire safety. Journées annuelles du GdR MASCOT NUM (MASCOT NUM 2016), mars 2016a. Poster.
- [comm2] Rémi Stroh, Julien Bect, Emmanuel Vazquez, Séverine Demeyer, Nicolas Fischer, et Damien Marquis. Planification et analyse d'expériences numériques, appliquées à la sécurité en ingénierie incendie. In *21^{èmes} journées du GdR Feux*, Cadarache, France, janvier 2016b.
- [comm3] Rémi Stroh, Julien Bect, Emmanuel Vazquez, Séverine Demeyer, et Nicolas Fischer. Sequential design of experiment on a stochastic multi-fidelity simulator to estimate a probability of exceeding a threshold. In *Journées annuelles du GdR MASCOT NUM (MASCOT NUM 2017)*, Paris, France, mars 2017.

Chapitre 1

Modélisation bayésienne de la multi-fidélité

Sommaire

1.1	Introduction	15
1.1.1	Notions de multi-fidélité	15
1.1.2	Approches non-bayésiennes	16
1.2	Processus gaussiens modélisant la multi-fidélité	18
1.2.1	Modèle auto-régressif	18
1.2.2	Modèle non-stationnaire	31
1.2.3	Discussions	42
1.3	Extension aux simulateurs stochastiques multi-fidèles	44
1.3.1	Modèles de la littérature	45
1.3.2	Modèle simplifié dédié à la multi-fidélité	47
1.3.3	Illustrations	50
1.4	Conclusion	60

1.1 Introduction

1.1.1 Notions de multi-fidélité

La multi-fidélité est un concept qui décrit des simulateurs sur lesquels le compromis entre fidélité d'une simulation (dans le sens, qualité de la simulation vis-à-vis du système que le simulateur est sensé décrire) et coût calculatoire est réglable. Une simulation en

haute-fidélité donnera un résultat précis, mais consommera également beaucoup de ressources calculatoires, alors qu'une simulation en basse-fidélité fournira plus rapidement un résultat, mais celui-ci sera bien moins fiable.

Il existe deux configurations courantes dans lesquelles il y a une structure multi-fidèle. D'abord, le cas le plus classique concerne les simulateurs avec un paramètre d'entrée, noté δ , réglant fidélité de la simulation. Cela peut être, par exemple, un simulateur fonctionnant par différences finies [Tuo et al., 2014] ou par éléments finis, ou bien un simulateur fondé sur une méthode de Monte-Carlo [Picheny et Ginsbourger, 2013]. Ensuite, la multi-fidélité peut englober des cas où il y a plusieurs types d'expériences (classiques ou numériques) différentes, modélisant le même phénomène et dont le niveau de précision est connu. L'objectif est de relier ces différents niveaux de fidélité entre eux, pour limiter le nombre d'expériences du niveau de plus haute fidélité coûteux. La configuration la plus courante est celle de plusieurs simulateurs ayant des précisions différentes [Forrester et al., 2007; Qian et Wu, 2008]. Goh et al. [2013] et Ezzat et al. [2017] parlent également de multi-fidélité dans un problème combinant des résultats d'un simulateur et ceux d'une expérience réelle.

Remarque 1. Dans un contexte multi-fidèle, le nombre et le type de facteurs d'entrées peut varier en fonction du niveau de fidélité. Par exemple, Le Gratiet et Cannamela [2015] traite un cas avec trois entrées sur simulateur basse-fidélité et huit sur celui de haute-fidélité.

Pour pouvoir analyser un modèle numérique composé de plusieurs niveaux de fidélité, nous utilisons des modèles fondés sur des processus gaussiens. Pour une description plus complète de la modélisation par processus gaussiens sans notion de multi-fidélité, nous renvoyons par exemple à Santner et al. [2003]. Dans cette thèse, nous décrivons des modèles de processus gaussiens adaptés aux simulateurs multi-fidèles. Pour cela, la relation entre les niveaux de multi-fidélité est représentée dans les processus a priori. Ces processus permettent, après conditionnement par les observations, d'estimer les valeurs des sorties du simulateur à tous les niveaux de fidélité, y compris et surtout, au niveau de plus haute fidélité. Nous proposons également une extension de ces modèles aux cas stochastiques.

1.1.2 Approches non-bayésiennes

Cette sous-section décrit des méthodes de la littérature associées traitant la multi-fidélité, qui ne relèvent pas de l'approche bayésienne développée dans ce travail. Pour plus d'informations, Giselle Fernández-Godino et al. [2016] propose une revue des méthodes

permettant de prendre en compte la multi-fidélité, et De Lozzo [2015] se focalise sur les méthodes fondées sur modèles de substitution, adaptés à la multi-fidélité.

Une façon simple d'exploiter une structure multi-fidélité, consiste à relier deux niveaux de fidélité par une fonction dite « de transfert ». L'algorithme consiste alors à alterner des phases d'amélioration de la fonction de transfert, et des phases de résolution du problème sur la fonction basse-fidélité. Alexandrov et al. [1999] et Leifsson et al. [2012] résument la méthodologie dans le cas où l'objectif est de trouver l'argument du minimum d'une fonction haute-fidélité f^{HF} coûteuse. Ils supposent disponible une fonction basse-fidélité f^{BF} , qui est une approximation grossière de la fonction haute-fidélité, et dont l'observation d'un point est instantanée. Dans ce cas, une approximation de la fonction haute-fidélité est obtenue par application d'une fonction de transfert s appliquée sur le niveau basse-fidélité, $\widehat{f}^{\text{HF}}(x; \theta) = s(x, f^{\text{BF}}; \theta)$, θ étant des hyper-paramètres de l'approximation. La fonction de transfert s peut prendre différentes formes, les plus simples étant une relation affine ou un changement de variable. L'algorithme d'optimisation applique alors successivement les étapes suivantes :

1. optimisation de la fonction de substitution : $x_{n+1} = \arg \max_{x \in \mathbb{X}} \{ \widehat{f}^{\text{HF}}(x; \theta_n) \}$;
2. observation de la vraie fonction haute-fidélité : $y_{n+1} = f^{\text{HF}}(x_{n+1})$;
3. amélioration de l'approximation selon un critère, par exemple, l'erreur quadratique moyenne : $\theta_{n+1} = \arg \min_{\theta \in \Theta} \left\{ \frac{1}{n+1} \sum_{k=1}^{n+1} \left(y_k - \widehat{f}^{\text{HF}}(x_k; \theta) \right)^2 \right\}$.

Suivant cette approche, l'efficacité de l'algorithme dépend alors de la qualité de la fonction basse-fidélité et du choix de la fonction de transfert. Alexandrov et al. [1999] appliquent cette méthodologie pour optimiser une aile d'avion, et soulignent déjà les problèmes de combinaison des niveaux de fidélité et des répartitions des temps de calcul entre les niveaux. Pour créer des relations plus complexes entre les niveaux, il est possible d'utiliser différents modèles de régression. Ainsi, Leary et al. [2003] et Kim et al. [2007] utilisent des réseaux de neurones et Hage Hassan et al. [2014] des fonctions à base radiale pour construire des modèles multi-fidèles.

La notion de multi-fidélité peut également renvoyer à des méthodes de Monte-Carlo multi-niveaux. Ces techniques permettent, entre autres, d'échantillonner plus rapidement des trajectoires de solutions d'équations différentielles stochastiques en jouant sur le pas de temps de la discrétisation. La méthode a été proposée par Heinrich [2001] et étudiée notamment par Giles [2008]. Dans leur problème, les auteurs cherchent à calculer $Y = \mathbb{E}[f(S_T)]$, où f est une fonction connue, T un réel et S la solution d'une équation différentielle stochastique de la forme : $dS_t = a(S_t, t)dt + b(S_t, t)dW_t$. La solution classique consiste à combiner différence finie et Monte-Carlo. Un nombre N de trajectoires

aléatoires discrétisées selon une simple méthode d'Euler avec un pas de temps h sont tracées, $\widehat{S}_n^{(i)} \approx S_{nh}$, $0 \leq n \leq T/h$, $1 \leq i \leq N$. L'estimation de Y est la moyenne empirique par Monte-Carlo des résultats sur les n trajectoires. En choisissant correctement les valeurs de N et de h , le coût calculatoire C est proportionnel à l'inverse-cube de l'incertitude finale de l'estimation ϵ , $C = \mathcal{O}(\epsilon^{-3})$ [Duffie et Glynn, 1995].

La méthode de Monte-Carlo multi-niveaux de Giles [2008] propose de considérer $L + 1$ pas de temps $h_l = T/M^l$, où $l \in \llbracket 0; L \rrbracket$ et M est un entier. Pour chacun de ces pas de temps, N_l trajectoires sont tracés pour obtenir $\widehat{S}_{l,T/h_l}^{(i)}$, puis une estimation $\widehat{P}_l^{(i)} = f\left(\widehat{S}_{l,T/h_l}^{(i)}\right)$ est calculée. L'algorithme calcule alors les moyennes empiriques des différences et l'estimation finale par la somme de toutes les différences :

$$\begin{cases} \widehat{Y}_0 = (1/N_0) \sum_{i=1}^{N_0} \widehat{P}_0^{(i)} \\ \widehat{Y}_l = (1/N_l) \sum_{i=1}^{N_l} \left(\widehat{P}_l^{(i)} - \widehat{P}_{l-1}^{(i)} \right), \quad l \geq 1. \\ \widehat{Y} = \sum_{l=0}^L \widehat{Y}_l \end{cases} \quad (1.1)$$

Cet estimateur \widehat{Y} est non-biaisé, et pour une incertitude ϵ requiert un coût calculatoire $C = \mathcal{O}(\epsilon^{-2} \log(\epsilon)^2)$, à condition de choisir correctement les valeurs de L , M et N_l . Ces méthodes d'échantillonnage par Monte-Carlo multi-niveaux ne seront pas étudiées plus en profondeur dans cette thèse.

1.2 Processus gaussiens modélisant la multi-fidélité

Cette partie présente les modèles multi-fidèles fondés sur des processus gaussiens. Nous supposons dans cette section que les simulateurs considérés sont déterministes.

1.2.1 Modèle auto-régressif

Définitions

Un modèle multi-fidèle par processus gaussiens est le modèle auto-régressif, développé par Kennedy et O'Hagan [2000]. Ce modèle est conçu pour un nombre fini S de niveaux. Le niveau le plus fidèle est indexé par S et le moins fidèle par 1. Chaque niveau de fidélité est défini sur le même espace des entrées \mathbb{X} . Kennedy et O'Hagan [2000] proposent de

modéliser les réponses aux différents niveaux de fidélité par S processus conjointement gaussiens $(\xi_s)_{1 \leq s \leq S}$.

Dans le cas où il n'y a que deux niveaux, $S = 2$, l'hypothèse fondamentale du modèle [O'Hagan, 1998], est que, connaissant la valeur de $\xi_1(x)$, aucune information supplémentaire sur la valeur de $\xi_2(x)$ ne peut être obtenue en observant ξ_1 . Plus mathématiquement,

$$\forall (x, x') \in \mathbb{X}^2, \text{cov} [\xi_2(x), \xi_1(x') | \xi_1(x)] = 0. \quad (1.2)$$

Par application du théorème situé dans Le Gratiet [2013, Annexe A.1], qui est un cas particulier du théorème B.1.1, (1.2) implique une relation entre les niveaux

$$\begin{cases} \xi_1(x) = \eta_1(x) \\ \xi_2(x) = \gamma_1(x)\xi_1(x) + \eta_2(x) \end{cases}, \quad (1.3)$$

où γ_1 est une fonction connue de relation entre les niveaux (selon le théorème B.1.1, $\gamma_1(x) = \frac{\text{cov}[\xi_2(x), \xi_1(x)]}{\text{var}[\xi_1(x)]}$), et $(\eta_s)_{1 \leq s \leq 2}$ sont deux processus gaussiens et indépendants, qui forment la base du modèle. Le Gratiet [2013] propose également une extension dans le cas où il y a un nombre S quelconque de niveaux de fidélité. Le modèle étendu s'écrit :

$$\begin{cases} \xi_1(x) = \eta_1(x) \\ \xi_{s+1}(x) = \gamma_s(x)\xi_s(x) + \eta_{s+1}(x), \quad s \in \llbracket 1; S-1 \rrbracket \end{cases}, \quad (1.4)$$

où les $(\gamma_s)_{1 \leq s \leq S-1}$ sont $S-1$ fonctions connues de x et $(\eta_s)_{1 \leq s \leq S}$ sont S processus gaussiens et indépendants. Pour obtenir cette décomposition, l'hypothèse d'indépendance conditionnelle doit être plus forte. Pour assurer l'indépendance de tous les η_s entre eux, il faut et suffit que, pour tout triplet de niveau $s^{\text{HF}} \geq s \geq s^{\text{BF}}$

$$\forall (x, x') \in \mathbb{X}^2, \text{cov} [\xi_{s^{\text{HF}}}(x), \xi_{s^{\text{BF}}}(x') | \xi_s(x)] = 0, \quad (1.5)$$

si la sortie au point x au niveau de fidélité s est connue, alors aucune information supplémentaire ne peut être extraite des niveaux de fidélité inférieurs pour estimer la valeur en x d'un niveau de fidélité supérieur (voir démonstration en annexe B.1.2). Bien que Le Gratiet [2013] et Le Gratiet et Garnier [2014] traitent le cas où les fonctions γ_s peuvent varier en fonction de x , nous supposons désormais que les fonctions γ_s sont constantes.

A partir de (1.4), il est possible de déduire une expression générale du processus ξ_s

en fonction des processus indépendants

$$\xi_s(x) = \sum_{t=1}^s \left(\prod_{k=t}^{s-1} \gamma_k \right) \eta_t(x), \quad (1.6)$$

et d'en déduire une expression des moyennes et des covariances des processus ξ_s en fonction des moyennes et des covariances des processus η_s

$$\mathbb{E} [\xi_s(x)] = \sum_{t=1}^s \left(\prod_{k=t}^{s-1} \gamma_k \right) \mathbb{E} [\eta_t(x)] \quad (1.7a)$$

$$\text{cov} [\xi_s(x), \xi_{s'}(x')] = \left(\prod_{l=\min\{s,s'\}}^{\max\{s,s'\}-1} \gamma_l \right) \cdot \sum_{t=1}^{\min\{s,s'\}} \left(\prod_{k=t}^{\min\{s,s'\}-1} \gamma_k^2 \right) \text{cov} [\eta_t(x), \eta_t(x')]. \quad (1.7b)$$

En pratique, le modèle est défini à partir des fonctions de covariance des processus η_t . Ces fonctions de covariance sont choisies dans une famille de fonctions, paramétrées par des hyper-paramètres. L'une de ces familles de fonctions de covariance est la famille des fonctions de covariance de Matérn, dont une présentation succincte est disponible annexe A.2, et une présentation plus détaillée dans Santner et al. [2003, Section 2.3.4]. Les hyper-paramètres du modèle peuvent être soit estimés, puis substitués directement dans le modèle qui les traite comme des quantités connues, soit être intégrés par une méthode complètement bayésienne. Dans cette thèse, nous privilégions cette dernière méthode de traitement des hyper-paramètres, afin de prendre en compte leurs incertitudes.

Le modèle auto-régressif est très apprécié dans la littérature et est utilisé dans de nombreuses applications, dont par exemple :

- Forrester et al. [2007] et Brooks et al. [2011] pour mettre au point une aile d'avion ;
- Wankhede [2012] pour concevoir une chambre de combustion à mélange pauvre ;
- Goh et al. [2013] pour calibrer un simulateur pour mieux représenter des expériences de chocs radiatifs ;
- Elsayed [2015] pour optimiser un cyclone de séparation ;
- Le Gratiet et Cannamela [2015] pour mener une étude des tensions dans un réservoir nucléaire ;
- Thenon et al. [2016] pour tracer la carte des gisement en géo-statistiques ;
- Perdikaris et Karniadakis [2016] pour résoudre un problème inverse sur un simulateur de circulation sanguine ;
- Demeyer et al. [2017] pour évaluer la sécurité incendie d'un bâtiment.

Dans ces exemples, le nombre de niveaux de fidélité S est toujours inférieur à trois, en

général égal à deux. Il existe des améliorations de ce modèle : Le Gratiet et Garnier [2014] étendent le modèle pour prendre en compte le cas où les fonctions γ_s dépendent des entrées x , Qian et Wu [2008] développent une approche complètement bayésienne du modèle, Le Gratiet [2013] propose une implémentation rapide de celui-ci dans le cas d'un plan d'expériences emboîtés, qui est reprise et améliorée par Perdikaris et al. [2015]

Remarque 2. Il est possible de considérer une extension au cas continu de ce modèle. Considérons un continuum de niveaux de fidélité indexé par $\delta \in [0; 1]$, où $\delta = 0$ indexe le niveau de plus haute fidélité et $\delta = 1$ indexe le niveau de plus basse fidélité. La transposition de (1.5) pour des modèles continus permet d'aboutir à l'expression

$$\xi(x, \delta) = r(x, \delta) \cdot \xi_1(x) + n(x, \delta), \quad (1.8)$$

où $\xi_1(x) = \xi(x, 1)$ correspond au niveau de plus basse fidélité, r est une fonction connue, et n est un processus gaussien indépendant de ξ_1 , qui respecte une certaine propriété.

Dans un souci de simplification, la fonction r est supposée être égale à un, $r(x, \delta) = 1$, ce qui, dans le cas d'un nombre fini de niveaux, correspond au cas où tous les coefficients γ_s valent un. Dans ces conditions, (1.8) s'écrit

$$\xi(x, \delta) = \xi_1(x) + n(x, \delta), \quad (1.9)$$

et la propriété que doit respecter le processus n est

$$\delta^{\text{BF}} \geq \delta^{\text{HF}} \Rightarrow \text{cov} \left[n(x, \delta^{\text{BF}}), n(x', \delta^{\text{HF}}) - n(x', \delta^{\text{BF}}) \right] = 0. \quad (1.10)$$

En particulier, cela implique que en un point x_0 donné, le processus $\delta \mapsto n(x_0, \delta)$ possède des accroissements indépendants, propriété que possède aussi le processus de Wiener.

L'équation (1.9) ressemble à (1.16), qui définit le modèle multi-fidèle non-stationnaire décrit section 1.2.2. La différence est que le niveau de référence du processus ξ est le niveau de plus basse-fidélité, à l'opposé de ce que suggère le modèle non-stationnaire. Ce modèle avec un paramètre de fidélité continu a été proposé par Kennedy et O'Hagan [2000], avec la fonction r fixée à 1. Cependant, celui-ci semble n'avoir jamais été utilisé à ce jour.

Propriétés et prédictions des niveaux de haute-fidélité

L'équation (1.4) décrit une relation « corrective » entre deux niveaux de fidélité successifs. En effet, le processus du niveau de fidélité $s + 1$ est égal au processus du

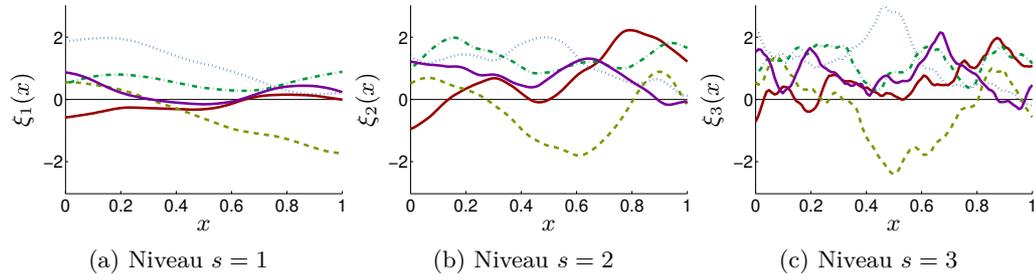


FIGURE 5 – Des trajectoires conditionnelles issues d’un modèle auto-régressif. Chaque image correspond à un niveau, chaque style de courbe à une trajectoire. Les hyper-paramètres de la covariance sont fixés. Les coefficients de régression entre les niveaux sont $\gamma_1 = \gamma_2 = 1$. Les processus η_s sont centrés et ont des fonctions de covariance de Matérn, dont les hyper-paramètres sont $\sigma_1 = 1$, $\nu_1 = 2.5$, $\rho_1 = 0.8$; $\sigma_2 = 0.8$, $\nu_2 = 2.5$, $\rho_2 = 0.3$; et $\sigma_3 = 0.4$, $\nu_3 = 1.5$, $\rho_3 = 0.1$.

niveau s (à un facteur multiplicatif près) plus un terme η_s , qui vient corriger le niveau de plus basse-fidélité. La figure 5 présente un exemple de trajectoires a priori issues d’un modèle auto-régressif. L’aspect correctif du modèle est visible sur un triplet de trajectoires issues du même tirage (même style). Le second niveau est égal au premier niveau plus quelques variations, et le troisième niveau est égal au second niveau plus des variations plus rapide.

L’aspect correctif du modèle entraîne plusieurs propriétés. La première est que la régularité des processus ne peut que diminuer avec le niveau de fidélité. Ainsi, si $S \geq s^{\text{HF}} \geq s^{\text{BF}} \geq 1$, la régularité du processus $\xi_{s^{\text{HF}}}$, définie dans [Santner et al., 2003, Section 2.3.4] comme la dérivabilité en moyenne quadratique, sera inférieure ou égale à celle de $\xi_{s^{\text{BF}}}$. Cela est dû au fait que, à chaque nouveau niveau de fidélité, un nouveau terme correctif s’ajoute au processus, ce qui entraîne une diminution de la régularité des trajectoires.

Une autre propriété du modèle est liée à sa moyenne a posteriori, utilisée comme prédiction. Soit des observations à différents niveaux de fidélité, $\chi_n = (x_i, s_i; z_i)$, $1 \leq i \leq n$, où x_i est le $i^{\text{ème}}$ vecteur d’entrées, s_i le $i^{\text{ème}}$ niveau de fidélité, et z_i la sortie obtenue au niveau s_i au point x_i . Supposons qu’il n’y ait d’aucune observation au niveau de plus haute fidélité, $\max\{s_i\} < S$, et que toutes les moyennes et covariances des processus η_s soient connues. Alors la moyenne a posteriori du niveau S est égale à γ_{S-1} fois la moyenne a posteriori du niveau $S-1$, plus la moyenne a priori de η_S . Une relation

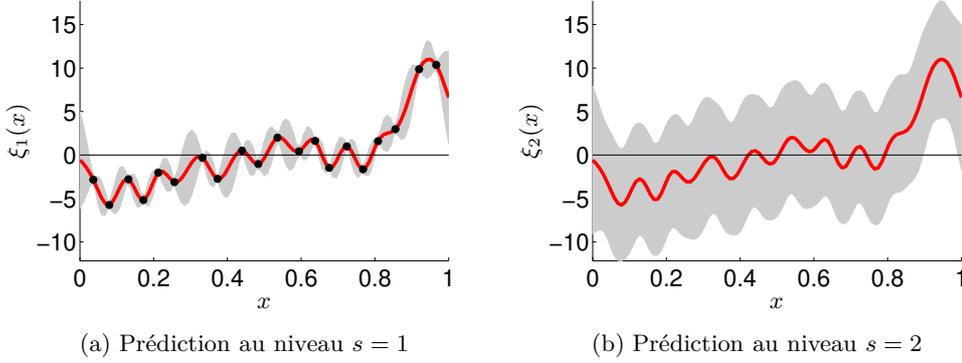


FIGURE 6 – Un exemple d’extrapolation en haute-fidélité avec un modèle auto-régressif. Les points sont les observations, les lignes continues sont les prédictions des différents niveaux de fidélité et les bandes correspondent aux intervalles de crédibilité à 95 %. Les hyper-paramètres de la covariance sont fixés. Le coefficient de régression entre les niveaux est $\gamma_1 = 1$. Les processus η_s sont de moyennes constantes et ont des fonctions de covariance de Matérn, dont les hyper-paramètres sont $\sigma_1 = 4.02$, $\nu_1 = 2.5$, $\rho_1 = 0.07$; et $\sigma_2 = 3.27$, $\nu_2 = 2.5$, $\rho_2 = 0.06$. Toutes les observations sont sur le premier niveau.

affine du même ordre peut être trouvée entre la variance a posteriori du niveau de plus haute-fidélité et celle du niveau précédent. Ainsi, si $\max\{s_i\} < S$, alors

$$\mathbb{E}[\xi_S(x)|\chi_n] = \gamma_{S-1}\mathbb{E}[\xi_{S-1}(x)|\chi_n] + \mathbb{E}[\eta_S(x)] \quad (1.11a)$$

$$\text{var}[\xi_S(x)|\chi_n] = \gamma_{S-1}^2 \text{var}[\xi_{S-1}(x)|\chi_n] + \text{var}[\eta_S(x)], \quad (1.11b)$$

où les processus sont supposés gaussiens (hyper-paramètres fixés). La figure 6 montre un exemple de ce principe avec $S = 2$, $\gamma_{S-1} = 1$ et $\mathbb{E}[\eta_S(x)] = 0$. Il n’y a aucune observation au niveau de plus haute fidélité. En conséquence, la prédiction du dernier niveau de fidélité est égale à la prédiction du niveau précédent. L’augmentation de l’intervalle de crédibilité sur le niveau $s = 2$ est liée à la variance a priori du processus η_2 .

Enfin, la propriété la plus appréciable de ce modèle multi-fidélité est la possibilité de le décomposer lorsque les plans d’expériences sont emboîtés. Le plan d’expériences χ_n est emboîté, lorsque tout point x observé à un niveau s est également observé aux niveaux de plus basse fidélité, $s' < s$. Dans ce cas, grâce à la définition du modèle (1.5), l’estimation des hyper-paramètres maximisant la vraisemblance et la prédiction du processus multi-fidèle avec S niveaux peut se simplifier en S problèmes récursifs d’estimation

d'hyper-paramètres et de prédiction d'un niveau de fidélité. La clef de la méthode, qui est présentée en détails dans Le Gratiet [2013, Chapitre 4], est de considérer les différences entre les sorties deux à deux, $\xi_{s+1}(x) - \gamma_s \xi_s(x)$. L'algorithme part du niveau de plus basse fidélité, $s = 1$. Comme le plan est emboîté, et que les processus $(\eta_s)_{s \geq 2}$ sont indépendants de η_1 , les observations situées aux niveaux de fidélité supérieurs à un n n'apportent aucune information sur η_1 . Les méthodes dédiées au cas d'un simple niveau s'appliquent pour estimer les hyper-paramètres associés à η_1 et pour prédire $\eta_1 = \xi_1$ en ne considérant que les observations du premier niveau. Une fois cela achevé, l'algorithme passe au deuxième niveau. Le paramètre $\widehat{\gamma}_1$ maximisant la vraisemblance est obtenu par une méthode des moindres carrés entre les observations du deuxième niveau et les observations du premier niveau associées. Dans ces conditions, les valeurs du processus η_2 aux points d'observations sont estimées par $\eta_2(x) = \xi_2(x) - \widehat{\gamma}_1 \xi_1(x)$. Comme η_2 est indépendant du premier niveau, les observations basse fidélité n'apportent aucune information supplémentaire. De plus, comme le plan est emboîté, et que les processus $(\eta_s)_{s \geq 3}$ sont indépendants de η_2 , les observations situées aux niveaux de fidélité supérieurs à deux n'apportent aucune information sur η_2 . Les méthodes dédiées au cas d'un simple niveau s'appliquent pour estimer les hyper-paramètres correspondant au processus η_2 et pour prédire η_2 en ne considérant que le deuxième niveau. Les prédictions du second niveau ξ_2 se calculent en utilisant les formules de reconstruction (1.11). Les niveaux suivants s'obtiennent récursivement de la même façon. Ainsi, dans le cas d'un plan emboîté, il est possible d'utiliser un modèle multi-fidèle sans avoir à développer d'outils particuliers en plus de la régression par processus gaussien sur un niveau de fidélité.

Cependant, ce modèle souffre de défauts. En particulier, il est conçu et utilisé dans le cas où il n'y a que peu de niveaux de fidélité (deux ou trois dans les applications). S'il y a plus de niveaux, typiquement dans un simulateur avec un paramètre de maillage, la méthode utilisée dans la littérature consiste à ne choisir que quelques niveaux parmi ceux possibles. De plus, si l'indépendance entre les processus η_s permet des simplifications, elle implique également d'avoir des observations à tous les niveaux de fidélité pour pouvoir estimer tous les hyper-paramètres, y compris aux niveaux les plus coûteux où le nombre d'observations est limité. Enfin, chaque nouveau niveau implique un nouveau processus η_s , et donc un nouveau vecteur d'hyper-paramètres. Dans le cas où les fonctions de covariances des processus η_s appartiennent toutes à la même famille paramétrique, cela implique que le nombre d'hyper-paramètres évolue linéairement avec le nombre de niveaux, ce qui devient problématique quand le nombre de niveaux augmente. Une solution pour pallier à cette augmentation du nombre d'hyper-paramètres peut être d'ajouter des relations a priori entre les hyper-paramètres de différents niveaux. Cepen-

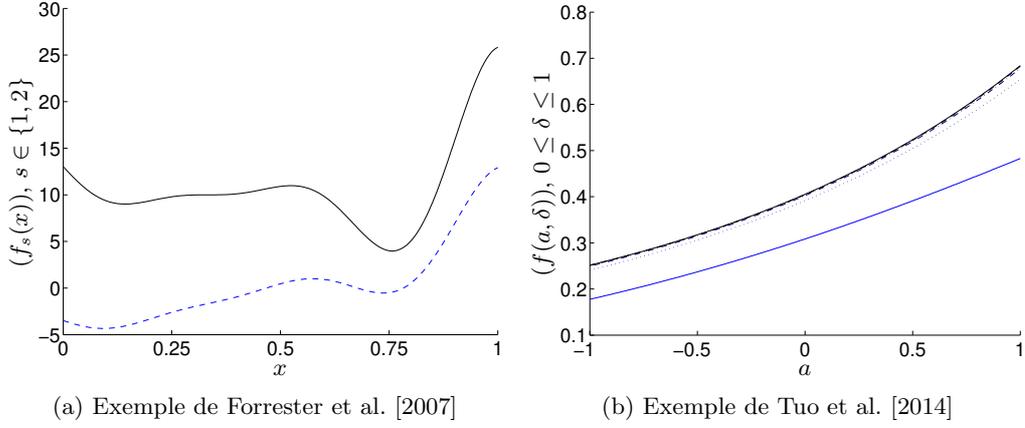


FIGURE 7 – Deux problèmes jouets en multi-fidélité. Toutes les fonctions sont déterministes. Sur les deux figures, chaque style de dessin correspond à un niveau de fidélité. Le niveau de plus haute-fidélité est en haut et le niveau de plus basse-fidélité est en bas. Toutes les fonctions sont déterministes. Sur l'exemple de Tuo et al. [2014], cinq niveaux sont représentés, correspondant aux valeurs $\delta = 1/2, 1/5, 1/11, 1/21$ et 0 . Les niveaux $\delta = 1/21$ et $\delta = 0$ sont quasiment identiques et leurs courbes sont confondues.

dant, faire cela empêche de pouvoir estimer les hyper-paramètres du modèle niveaux par niveaux, qui est l'un des atouts de ce modèle.

Illustrations

Dans cette section, le modèle auto-régressif est illustré à travers deux exemples de la littérature.

Le premier exemple a été proposé par Forrester et al. [2007] pour illustrer les modèles multi-fidèles auto-régressifs (section 1.2.1). Le modèle numérique est constitué de deux niveaux, basse-fidélité et haute-fidélité, définis sur $\mathbb{X} = [0; 1] \subset \mathbb{R}^1$ comme

$$f_1(x) = 0.5 \cdot (6x - 2)^2 \sin(12x - 4) + 10x - 5 \quad (1.12a)$$

$$\begin{aligned} f_2(x) &= (6x - 2)^2 \sin(12x - 4) + 10 & (1.12b) \\ &= 2f_1(x) + 20 - 20x. \end{aligned}$$

La relation entre les deux fonctions est très proche de la structure multi-fidélité (1.4). Les deux fonctions sont représentées figure 7a.

Le second exemple a été proposé par Tuo et al. [2014] pour illustrer les modèles multi-

fidèles représentant les simulateurs ayant des tailles de mailles variables (section 1.2.2). Le simulateur est défini pour $a \in \mathbb{X} = [-1; 1] \subset \mathbb{R}^1$, et pour $\delta = \frac{1}{n+1}$, $n \in \mathbb{N}^*$. Il procède en deux étapes. Dans un premier temps, il résout une équation de Poisson bi-dimensionnelle sur $[0; 1]^2$. Dans un second temps, il retourne l'intégrale de la solution sur son ensemble de définition. La résolution de l'équation de Poisson se fait par une approximation par différences finies, dans laquelle la taille de maille est fixée à δ , et l'intégrale s'approche par une méthode des trapèzes sur l'ensemble des mailles. Dans ce problème, l'équation de Poisson à résoudre est

$$\begin{cases} \frac{\partial^2 u_a}{\partial x^2}(x, y) + \frac{\partial^2 u_a}{\partial y^2}(x, y) = [(a^2 - 2\pi^2) \sin(\pi x) + 2a\pi \cos(\pi x)] e^{ax} \sin(\pi y) \\ u_a(0, y) = u_a(1, y) = u_a(x, 0) = u_a(x, 1) = 0 \end{cases}. \quad (1.13)$$

La solution exacte de cette équation est

$$u_a(x, y) = e^{ax} \sin(\pi x) \sin(\pi y). \quad (1.14)$$

Le simulateur exact est

$$f(a, 0) = \int_{x=0}^1 \int_{y=0}^1 u_a(x, y) dx dy = 2 \frac{e^a + 1}{a^2 + \pi^2}. \quad (1.15)$$

Quelques niveaux de fidélité de ce simulateur sont représentés figure 7b.

Sur chacun des exemples, le même type de modèle a priori est considéré. Les niveaux de fidélité sont liés par le modèle auto-régressif (1.4). Les processus $(\eta_s)_{1 \leq s \leq S}$ sont des processus gaussiens stationnaires de moyennes constantes inconnues suivants une loi uniforme impropre sur \mathbb{R} . Les covariances des processus η_s sont des covariances de Matérn de régularité $5/2$, $\sigma_s^2 \mathcal{M}_{5/2}(|x - x'| / \rho_s)$. L'expression de la fonction $\mathcal{M}_{5/2}$ est donnée (A.16). Les hyper-paramètres de ce modèle sont $\theta = \{(\sigma_s^2, \rho_s)_{1 \leq s \leq S}; (\gamma_s)_{1 \leq s \leq S-1}\}$. Une loi a priori est ajoutée sur θ , formée de produit de lois indépendantes, $p(\theta) = \prod_{s=1}^{S-1} p(\gamma_s) \prod_{s=1}^S p(\sigma_s^2) p(\rho_s)$. Les lois a priori des coefficients de régression γ_s sont des lois normales indépendantes, $\gamma_s \sim \mathbf{N}(1, 2^2)$. Les lois a priori des autres hyper-paramètres sont définies par la méthode décrite dans l'annexe D. Sur les deux exemples, le plan d'expériences est emboîté (voir section 2.2).

Pour mener une méthode complètement bayésienne, nous avons besoin d'échantillonner la loi a posteriori des hyper-paramètres. Pour cela, nous utilisons un algorithme de Monte-Carlo par Chaînes de Markov (MCMC)¹ pour engendrer un échantillon d'hyper-

1. Markov Chain Monte-Carlo

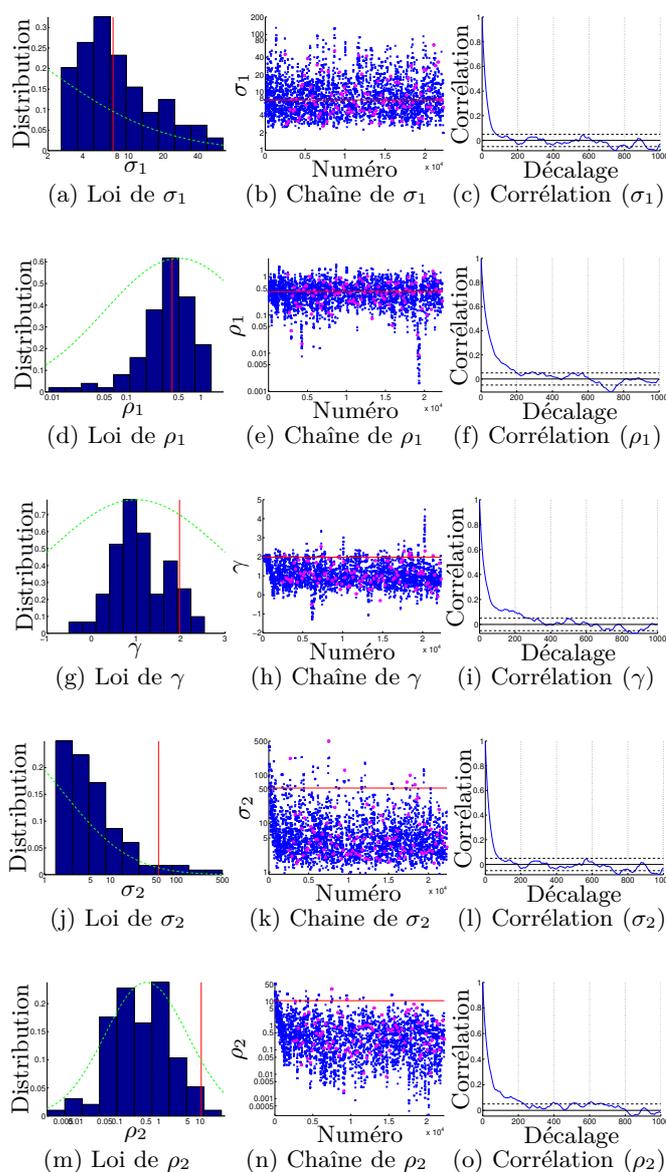


FIGURE 8 – Un échantillonnage de tous les hyper-paramètres sur un modèle auto-régressif appliqué à l'exemple à deux niveaux de Forrester et al. [2007]. Chaque ligne correspond à un hyper-paramètre. Gauche : histogramme des échantillons des hyper-paramètres distribués selon leur loi a posteriori. Le trait pointillé est la loi a priori normalisée et le trait rouge vertical est le MAP. Milieu : la chaîne de Markov issue de l'échantillonnage. Le trait horizontal rouge est le MAP et les points mauves sont le sous-échantillonnage de la chaîne. Droite : auto-corrélation de la chaîne. Les traits pointillés horizontaux indiquent le décalage pris entre chaque sous-échantillon de la chaîne Markov.

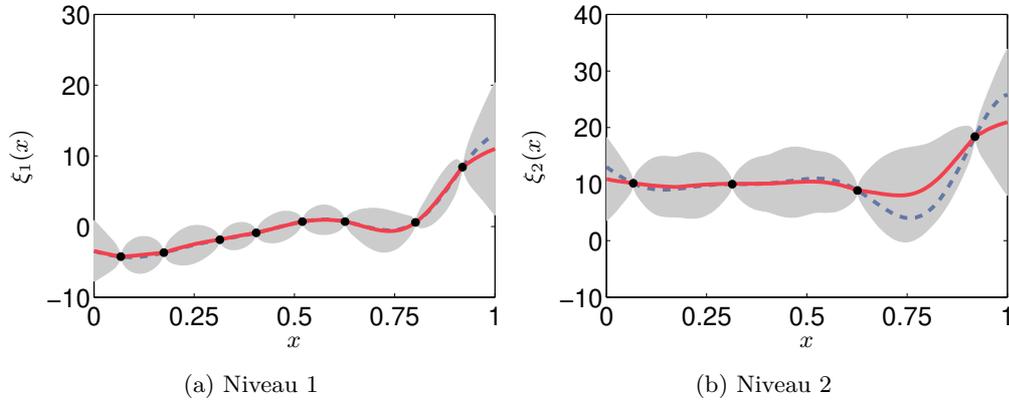


FIGURE 9 – Un exemple de prédiction sur deux niveaux avec le modèle auto-régressif. Les points correspondent aux observations, les traits pleins aux prédictions, les aires autour aux intervalles à deux écarts-types. Les traits pointillés sont les vraies fonctions.

paramètres. Puis, nous calculons les moyennes, les variances et les lois a posteriori en remplaçant la loi a posteriori des hyper-paramètres par l'échantillon obtenu. Pour échantillonner les hyper-paramètres, nous utilisons une méthode de Métropolis adaptative proposée par Haario et al. [2001]. Cette méthode consiste à adapter progressivement la loi propositionnelle en fonction des échantillons de la chaîne de Markov obtenue. La méthode d'échantillonnage et d'intégration des hyper-paramètres est présentée en annexe A.3.

Le premier exemple possède deux niveaux de fidélité. Huit observations sont effectuées au premier niveau et quatre au second niveau. Pour construire les lois a priori selon la méthode de l'annexe D, les valeurs de l'amplitude approximative de la sortie r^{out} et de la taille de l'intervalle selon chaque dimension b^{in} sont utilisées. Sur cet exemple, l'amplitude approximative de la sortie est $r^{\text{out}} = 20$ et la taille de l'intervalle est $b^{\text{in}} = 1$.

La figure 8 montre les lois a posteriori des cinq hyper-paramètres comparés à leurs lois a priori, ainsi que les chaînes de Markov ayant générées les échantillons. Les figures de gauche montrent que les lois a posteriori des paramètres du premier niveau sont plus resserrées que leurs lois équivalentes a priori. Cependant, la loi a posteriori du paramètre de portée du second niveau ρ_2 est très proche de la loi a priori. Cela est cohérent, puisque ce paramètre n'est estimé qu'avec les observations haute-fidélité qui sont peu nombreuses. La figure 9 montre les prédictions obtenues par ce modèle sur les deux niveaux de fidélité. La prédiction haute-fidélité prédit correctement la fonction vraie en haute-fidélité.

Sur le second exemple, cinq niveaux de fidélité sont considérés. Les quatre niveaux de

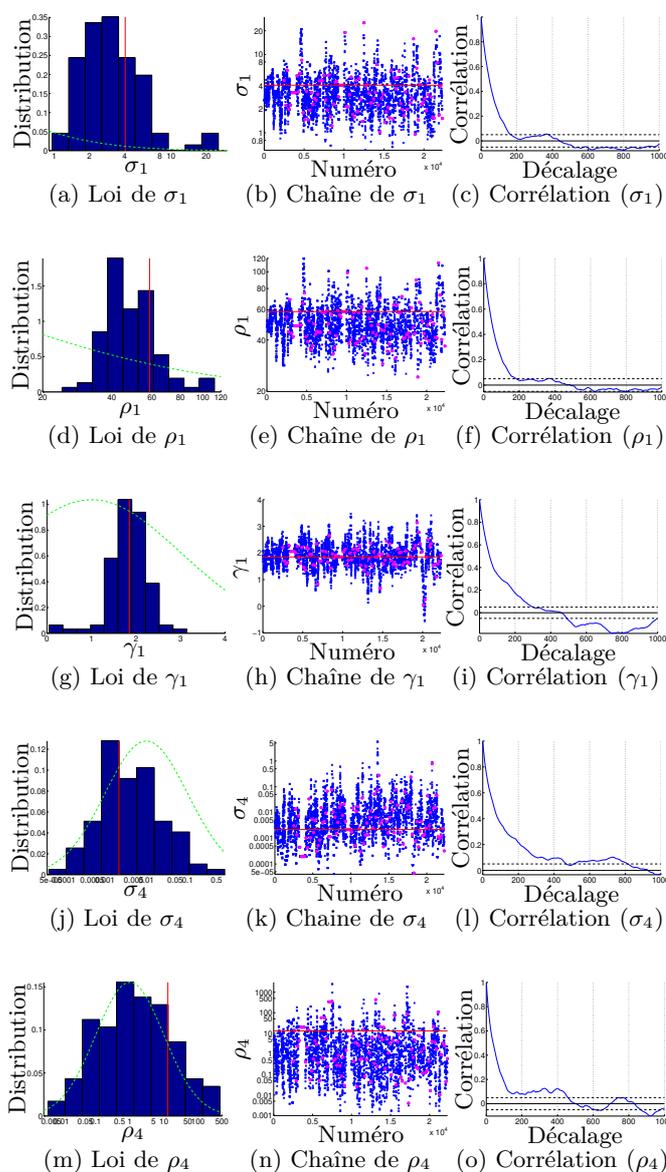


FIGURE 10 – Un échantillonnage de cinq des quatorze hyper-paramètres sur un modèle auto-régressif appliqué à l'exemple à cinq niveaux de Tuo et al. [2014]. Chaque ligne correspond à un hyper-paramètre. Gauche : histogramme des échantillons des hyper-paramètres distribués selon leur loi a posteriori. Le trait pointillé est la loi a priori normalisée et le trait rouge vertical est le MAP. Milieu : la chaîne de Markov issue de l'échantillonnage. Le trait horizontal rouge est le MAP et les points mauves sont le sous-échantillonnage de la chaîne. Droite : auto-corrélation de la chaîne. Les traits pointillés horizontaux indiquent le décalage pris entre chaque sous-échantillon de la chaîne Markov.

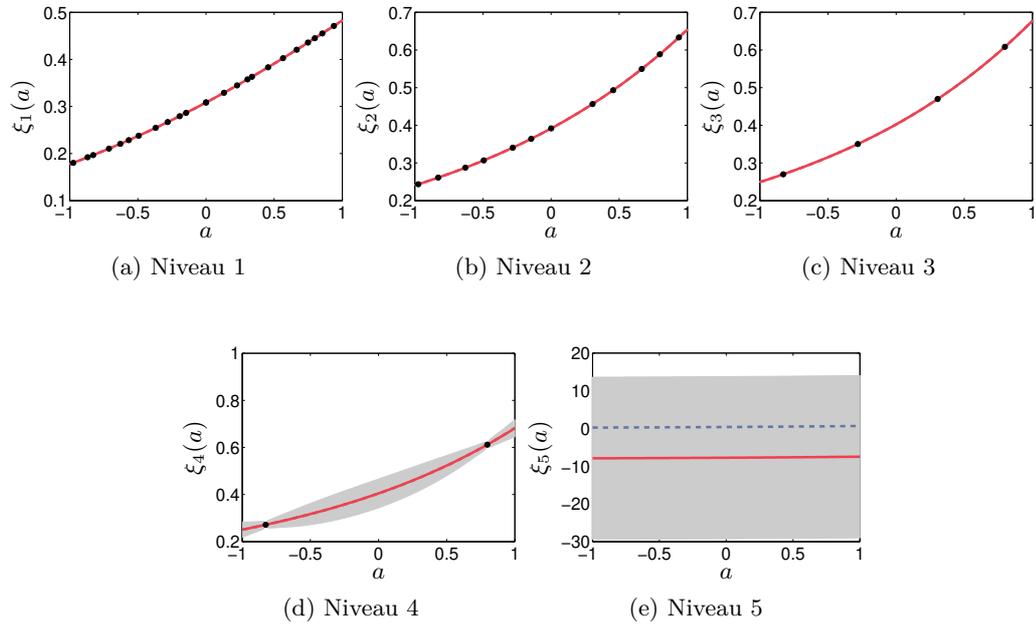


FIGURE 11 – Un exemple d’extrapolation en haute-fidélité avec le modèle auto-régressif. Les points correspondent aux observations, les traits pleins aux prédictions, les aires autour aux intervalles à deux écarts-types. Les traits pointillés sont les vraies fonctions.

fidélité observés correspondent à $\delta = 1/2, 1/5, 1/11$ et $1/21$, numérotés respectivement $s = 1, 2, 3$ et 4 . Le niveau inobservé à prédire correspond à $\delta = 0$ et est numéroté $s = 5$. Il y a respectivement à chaque niveau, 24, 12, 4 et 2 observations. L’amplitude approximative de la sortie est $r^{\text{out}} = 1$ et la taille de l’intervalle est $b^{\text{in}} = 2$.

La figure 10 montre les lois a posteriori de certains hyper-paramètres du modèle, ainsi que les chaînes ayant générées ces hyper-paramètres. Les hyper-paramètres représentés correspondent à ceux des niveaux observés de plus basse et de plus haute-fidélité. Les figures de gauche montrent que les lois a posteriori des hyper-paramètres du niveau 4 (plus haut niveau de fidélité observé) sont proches de leurs lois a priori respectives. Cela est cohérent, car il y a peu d’observations sur le niveau 4, et les processus de chaque niveau sont mutuellement indépendants. La figure montre aussi que la valeur de γ_1 le coefficient de régression entre le premier et le second niveau est estimé à environ deux, ce qui est cohérent avec les vraies fonctions (figure 7b).

La figure 11 montre les prédictions obtenues par ce modèle sur les cinq niveaux de fidélité. Sur tous les niveaux observés, les prédictions sont assez justes et trouvent la

fonction vraie correspondant au niveau. Cependant, le modèle est complètement incertain sur le niveau où il n'y a aucune observation. La zone visible sur le niveau cinq correspond à l'incertitude a priori de η_5 . Comme le modèle a priori est peu informatif, et que η_5 ne peut être estimé que par des estimations haute-fidélité, l'incertitude est très importante. Pour pouvoir estimer correctement les niveaux haute-fidélité inobservés, il faudrait imposer des lois a priori plus informatives. Par exemple, la valeurs des variances a priori des hyper-paramètres σ_s^2 pourrait décroître avec le niveau.

1.2.2 Modèle non-stationnaire

Définitions

Le modèle non-stationnaire est une autre façon de modéliser la multi-fidélité. Proposé indépendamment par Picheny et Ginsbourger [2013] et Tuo et al. [2014], il est conçu pour des modèles dont les niveaux de fidélité sont très nombreux ou forment un continuum de niveaux. Typiquement, cela englobe les simulateurs dont la fidélité est réglée par un temps de simulation, un nombre de pas de Monte-Carlo, ou une taille de maille dans un simulateur par différences/éléments finis. Le paramètre d'entrée contrôlant la fidélité de la simulation est noté δ à valeur dans \mathbb{R}^+ . Dans cette thèse, nous posons que le plus haut niveau de fidélité correspond à $\delta \rightarrow 0$. Le niveau $\delta = 0$ n'est pas observable en pratique, mais nous supposons qu'une limite existe, notée $\xi_0(x) = \lim_{\delta \rightarrow 0} \xi(x, \delta)$.

Le modèle décompose le processus ξ comme somme de deux processus gaussiens indépendants

$$\xi(x, \delta) = \xi_0(x) + \epsilon(x, \delta), \quad (1.16)$$

où $\xi_0 = \xi(\cdot, 0)$ représente le simulateur idéal si un temps de calcul infini était disponible, et ϵ est un processus centré, $\mathbb{E}[\epsilon(x, \delta)] = 0$, s'annulant en $\delta = 0$, et qui représente l'erreur numérique au niveau δ . En utilisant la proposition B.1.2, il est possible de montrer que ce modèle de processus gaussien est équivalent à

$$\mathbb{E}[\xi(x, \delta) | \xi(\cdot, 0)] = \xi(x, 0). \quad (1.17)$$

Cette équation signifie que si le processus idéal décrit par le simulateur était parfaitement connu, alors la meilleur prédiction du comportement du simulateur serait le processus idéal lui-même. Ce modèle a été utilisé par He et al. [2017] pour optimiser un processus de moulage d'alliage.

Ce modèle s'inspire des modèles de processus gaussiens spatio-temporels, où δ remplace le paramètre de temps. En particulier, comme la valeur de ϵ en $\delta = 0$ est nulle, la

fonction de covariance de ϵ ne peut pas être stationnaire en δ . Deux types de modèles ont été considérés dans les articles de Picheny et Ginsbourger [2013] et de Tuo et al. [2014], donnant deux types de fonctions de covariance pour ϵ .

Un premier type de covariance peut être envisagé si une hypothèse supplémentaire est imposée au modèle. Si $\delta^{\text{BF}} > \delta^{\text{HF}}$ sont deux niveaux de fidélité, respectivement basse et haute, une définition plus forte du modèle non-stationnaire donne

$$\mathbb{E}[\xi(x, \delta^{\text{BF}}) | \xi(\cdot, \delta^{\text{HF}})] = \xi(x, \delta^{\text{HF}}), \quad (1.18)$$

(1.17) devenant un cas particulier où $\delta^{\text{HF}} = 0$. Dans ces conditions, le processus ϵ possède des accroissements indépendants entre deux niveaux de fidélité

$$\text{cov} [\epsilon(x, \delta^{\text{BF}}) - \epsilon(x, \delta^{\text{HF}}), \epsilon(x', \delta^{\text{HF}})] = 0 \quad (1.19)$$

(voir proposition B.1.3). L'équation (1.19) est la duale de (1.10) : dans le modèle auto-régressif, la covariance du processus dépendant de la fidélité n est fonction du niveau de plus basse fidélité ; dans le modèle non-stationnaire, la covariance du processus dépendant de la fidélité ϵ est fonction du niveau de plus haute fidélité. Pour aller plus loin, si un nombre S de niveaux de fidélité sont choisis tel que $0 = \delta_1 < \delta_2 < \dots < \delta_S$, et que $\xi_s(x) = \xi(x, \delta_s)$, alors le modèle ainsi défini correspond au modèle auto-régressif (1.4) avec trois différences notables : les coefficients γ_s sont tous fixés à un ; les moyennes des processus η_s sont nulles ; et surtout, le niveau de plus haute fidélité correspond au premier niveau, et celui de plus basse fidélité au niveau S , inversement à la définition du modèle auto-régressif. Un modèle respectant cette propriété a été proposé dans Tuo et al. [2014] et utilisée dans He et al. [2017]. La fonction de covariance du processus ϵ est une covariance séparable en δ et en x de la forme

$$\text{cov} [\epsilon(x, \delta), \epsilon(x', \delta')] = \min\{\delta, \delta'\}^L \cdot k_\epsilon(x, x'), \quad (1.20)$$

où k_ϵ est une fonction de covariance sur \mathbb{X} et L est le degré de décroissance de l'erreur. La fonction $\min\{\delta, \delta'\}^L = \min\{\delta^L, \delta'^L\}$ est la fonction de covariance du processus $Y_t = W_{tL}$ où W_t est un mouvement brownien standard.

Un autre type de covariance peut être considéré pour modéliser l'erreur numérique ϵ . L'hypothèse consiste à supposer qu'il existe une fonction σ tel que $(\xi(x, \delta) - \xi_0(x))/\sigma(\delta)$ soit un processus gaussien stationnaire en δ quelque soit $x \in \mathbb{X}$. Dans ces conditions, la fonction de variance de ϵ devrait être proportionnelle à $\sigma(\delta)^2$. Une forme possible de

covariance pour ϵ est

$$\text{cov} [\epsilon(x, \delta), \epsilon(x', \delta')] = \sigma(\delta)\sigma(\delta') \cdot r(\delta - \delta') \cdot k_\epsilon(x, x'), \quad (1.21)$$

où σ est la fonction de l'écart-type de l'erreur entre le niveau de plus haute fidélité et le niveau courant, r est une fonction de corrélation, et k_ϵ est une fonction de corrélation sur \mathbb{X} . Ce type de covariance est proposé par Picheny et Ginsbourger [2013], qui proposent une fonction $\sigma(\delta) = \exp(-\alpha\delta_0/\delta)$, et par Tuo et al. [2014], qui suggèrent une fonction $\sigma(\delta) = (\delta/\delta_0)^L$, où α et L sont des hyper-paramètres de la covariance. Pour la fonction r , les deux articles suggèrent une fonction de corrélation de Matérn mono-dimensionnelle.

La fonction de corrélation en δ implique des comportements différents de la régression en fonction des niveaux. Ainsi, si la covariance brownienne (1.20) est utilisée, la prédiction par moyenne a posteriori sera constante hors de l'intervalle défini par les niveaux observés ; alors que si la covariance produit (1.21) est utilisée, la prédiction aura des variations lisses hors de l'intervalle d'observation. La figure 12 illustre les deux possibilités sur un exemple de régression multi-fidèle où le vecteur des entrées x est fixé à x_0 . Il y a quatre observations de $\xi(x_0, \cdot)$ à quatre niveaux de fidélité différents. Tous les hyper-paramètres sont fixés et la valeur de $\xi_0(x_0)$ est supposée inconnue. La figure 12a présente une prédiction, où la fonction de covariance du processus ϵ est une fonction de covariance brownienne déformée (1.20). Les prédictions hors des points d'observations sont des fonction constantes, car il est supposé impossible de prévoir le sens de variation général de l'erreur. La figure 12b est une prédiction, où la fonction de covariance du processus ϵ est celle décrite (1.21), avec $\sigma(\delta) = \delta$. L'extrapolation hors des points d'observations est variable. Dans le cas général, une covariance brownienne semble mieux adaptée en l'absence d'information complémentaire.

Propriétés et prédictions des niveaux de haute-fidélité

Le modèle multi-fidélité (1.16) décrit une relation de type « déconvolutive » entre un niveau quelconque et le niveau de plus haute-fidélité $\delta = 0$. Le simulateur est décrit comme un simulateur idéal à estimer, plus une erreur numérique qui vient perturber l'observation. La figure 13 présente des trajectoires a priori d'un modèle multi-fidèle non-stationnaire sur trois niveaux de fidélité ($\delta = 1$, $\delta = 0.5$ et $\delta = 0$). Un style de courbe sur les trois sous-figures correspond a une trajectoire. Pour chaque trajectoire, correspondant à un style de courbe, le simulateur apparait comme la courbe en $\delta = 0$ plus une erreur qui s'amplifie quand la fidélité diminue. Comme le modèle non-stationnaire est l'opposé du modèle auto-régressif, toutes les propriétés d'évolution avec la fidélité

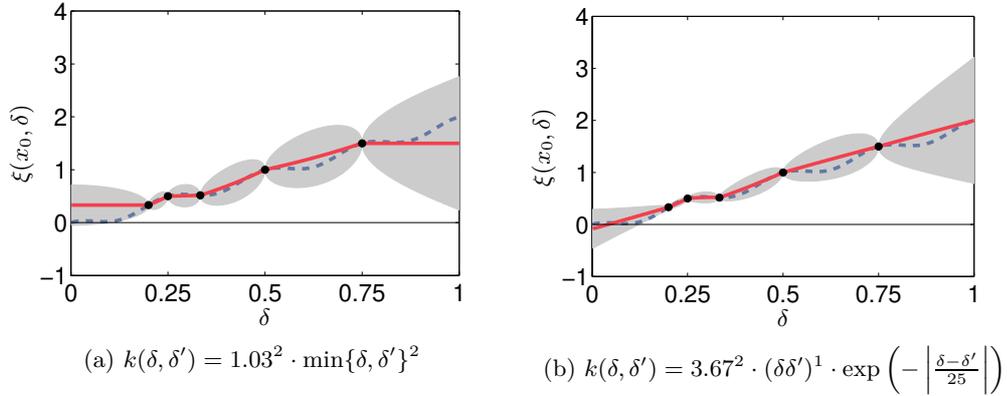


FIGURE 12 – Deux exemples de régressions selon les niveaux avec deux fonctions de covariances différentes. Celle de gauche utilise une fonction de covariance brownienne, celle de droite une fonction de covariance qui est un produit entre une fonction de covariance linéaire et une covariance exponentielle avec une très grande portée. Le trait pointillé est la vraie fonction d’erreur, le trait plein est la prédiction à partir des cinq observations et la zone autour est l’intervalle à 95 % de la prédiction. Dans les fonctions de covariance, l’exposant est fixé pour avoir $k(\delta, \delta) \propto \delta^2$, la portée est fixée et les variances sont calculées par maximum de vraisemblance.

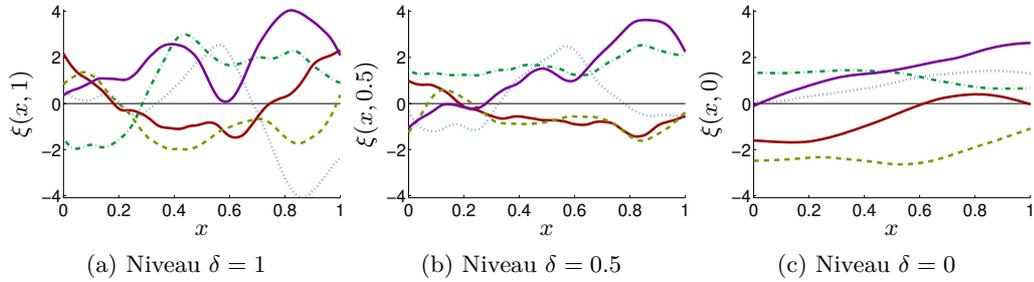


FIGURE 13 – Des trajectoires conditionnelles issues d’un modèle non-stationnaire. Chaque image correspond à un niveau, chaque style de courbe à une trajectoire. Les hyper-paramètres de la covariance sont fixés. Le processus idéal possède une fonction de covariance de Matérn dont les hyper-paramètres sont $\sigma_0 = 1$, $\nu_0 = 2.5$ et $\rho_0 = 0.8$. Le processus d’erreur à une covariance de la forme $\min\{\delta, \delta'\}^L \cdot k(x, x')$ avec $L = 1$, et k est une fonction de covariance de Matérn, dont les hyper-paramètres sont $\sigma_\epsilon = 2$, $\nu_\epsilon = 2.5$, $\rho_\epsilon = 0.3$.

sont inversées. Ainsi, quand la fidélité augmente, la régularité tend à augmenter.

La relation entre la prédiction en haute-fidélité et les observations est moins évidente, car la relation entre les niveaux de fidélité est inversée. Il existe un cas simple dans lequel l'expression de la prédiction sur un niveau de plus haute-fidélité inobservé peut être obtenue. Supposons pour simplifier que n observations soient toutes effectuées au même niveau $\delta^{\text{obs}} > 0$, $\chi_n = (x_i, \delta^{\text{obs}}; z_i)_{1 \leq i \leq n}$, et que le modèle soit fondé sur un processus gaussien multi-fidèle non-stationnaire, dont la covariance du processus idéal est stationnaire et la covariance du processus d'erreur est un produit d'une fonction de covariance dépendant de δ pour modéliser l'évolution en fonction de la fidélité, et d'une fonction de covariance stationnaire sur l'espace des entrées. La covariance totale du processus est alors de la forme

$$k((x, \delta), (x', \delta')) = \sigma_0^2 r_0(x - x') + \sigma_\epsilon^2 u(\delta, \delta') r_\epsilon(x - x'), \quad (1.22)$$

où $\sigma_0^2 r_0(x - x')$ est la covariance du processus idéal ($r_0(0) = 1$) et $\sigma_\epsilon^2 u(\delta, \delta') r_\epsilon(x - x')$ celle du processus d'erreur ($r_\epsilon(0) = 1$). Les hyper-paramètres des processus gaussiens sont fixés. Soit $u(\delta^{\text{obs}}) = u(\delta^{\text{obs}}, \delta^{\text{obs}})$. En utilisant la proposition B.2.1, la prédiction du niveau de plus haute fidélité est exprimable en fonction de la prédiction au niveau δ^{obs}

$$\begin{aligned} \mathbb{E}[\xi(x, 0) | \chi_n] &= (g * m_n(\cdot, \delta^{\text{obs}}))(x) + \{m_0(x) - (g * m_0)(x)\} \\ &= m_0(x) + \left[g * \left(m_n(\cdot, \delta^{\text{obs}}) - m_0 \right) \right] (x) \end{aligned} \quad (1.23)$$

où la fonction m_0 est la fonction de moyenne a priori du processus idéal, $m_0(x) = \mathbb{E}[\xi(x, 0)]$, la fonction $m_n(x, \delta)$ est la fonction de moyenne a posteriori du niveau δ en x , $m_n(x, \delta) = \mathbb{E}[\xi(x, \delta) | \chi_n]$ et la fonction g est définie par sa transformée de Fourier

$$\text{TF}[g](w) = \frac{\sigma_0^2 R_0(w)}{\sigma_0^2 R_0(w) + \sigma_\epsilon^2 u(\delta^{\text{obs}}) R_\epsilon(w)} = \frac{1}{1 + \frac{\sigma_\epsilon^2 u(\delta^{\text{obs}})}{\sigma_0^2} \frac{R_\epsilon}{R_0}(w)} \quad (1.24)$$

avec R_0 et R_ϵ les transformées de Fourier respectives de r_0 et r_ϵ . Par exemple, les transformées de Fourier de la fonction de covariance de Matérn sont données dans la annexe A.2. Ainsi, la prédiction au niveau de plus haute fidélité est un filtrage de la prédiction au niveau δ^{obs} par le filtre g . Ce filtre g correspond au filtre de Wiener. Ce filtre a deux effets principaux : un effet d'arbitrage entre la prédiction au niveau δ^{obs} et la moyenne a priori ; et un effet de lissage de la prédiction entre les deux niveaux.

L'effet d'arbitrage est visible dans le rapport des variances. Pour simplifier le raisonnement, supposons que les deux fonctions de corrélation r_0 et r_ϵ soient égales. Leurs

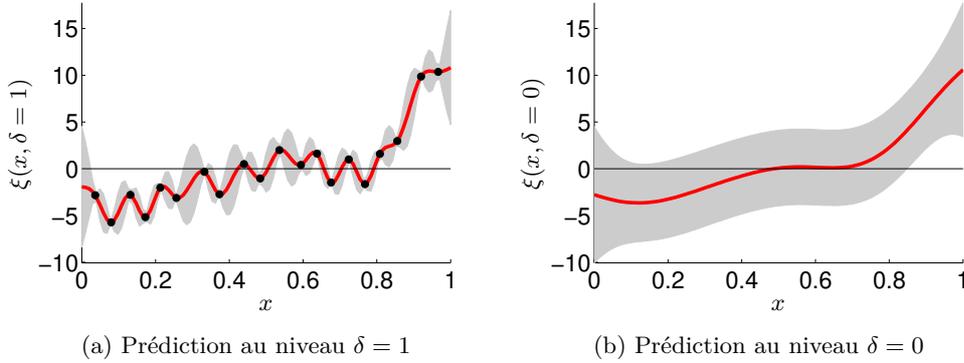


FIGURE 14 – Un exemple d’extrapolation en haute-fidélité avec un modèle non-stationnaire. Les points sont les observations, les traits pleins sont les prédictions des différents niveaux de fidélité et les bandes correspondent aux intervalles à 95 %. Les hyper-paramètres de la covariance sont fixés. La covariance du processus idéal ξ_0 est une fonction de covariance de Matérn dont les hyper-paramètres sont $\sigma_0 = 21.17$, $\nu_0 = 2.5$, $\rho_0 = 0.68$. La covariance du processus d’erreur est de la forme $\min\{\delta, \delta'\}^L \cdot k(x, x')$, où $L = 1$ et k est une covariance de Matérn, dont les hyper-paramètres sont $\sigma_\epsilon = 3.27$, $\nu_\epsilon = 2.5$, $\rho_\epsilon = 0.06$. Toutes les observations sont sur le premier niveau.

transformées de Fourier sont alors égales, et le filtre g devient un terme multiplicatif de la forme $1/(1 + \frac{\sigma_\epsilon^2 u(\delta^{\text{obs}})}{\sigma_0^2})$. La fraction $\frac{\sigma_\epsilon^2 u(\delta^{\text{obs}})}{\sigma_0^2}$ correspond au rapport de la variance de l’erreur au niveau δ^{obs} et la variance du processus idéal. Si ce rapport est faible, c’est-à-dire si l’erreur est petite, le coefficient g va tendre vers un, et donc la prédiction au niveau haute-fidélité sera égale à celle du niveau δ^{obs} . A l’inverse, si ce rapport est grand, c’est-à-dire si l’erreur est grande, le coefficient g va tendre vers zéro, et donc la prédiction au niveau haute-fidélité va tendre vers sa moyenne a priori. En régression par processus gaussien stationnaire, cet effet apparaît quand les observations ne permettent pas de déduire des informations au point considéré. Cela est vrai dans le cas où l’erreur est grande : les observations sont trop noyées dans l’erreur numérique pour que des informations puissent en être déduites sur le processus idéal. Le facteur g tend vers un quand $u(\delta^{\text{obs}})$ tend vers zéro, par exemple quand δ^{obs} s’approche de zéro. Cela signifie que, quand l’observation de plus haute fidélité disponible s’approche du niveau de fidélité idéal, la prédiction au niveau haute-fidélité se rapproche de la prédiction au niveau δ^{obs} , le modèle a de plus en plus confiance dans l’approximation.

L’effet de lissage est visible dans le rapport R_ϵ/R_0 . L’effet de ce filtre dépend des

longueurs de corrélation des deux covariances. Si la longueur de corrélation de r_0 est supérieur à celle de r_ϵ , $\rho_0 > \rho_\epsilon$, cela signifie qu'a priori, le processus idéal $\xi(x, 0)$ varie moins vite que le processus réel $\xi(x, \delta^{\text{obs}})$. Dans ce cas, le filtre g va se comporter comme un filtre passe-bas : il élimine les variations rapides observables sur la prédiction au niveau δ^{obs} pour en extraire les variations moins brusques. La figure 14 présente un exemple de régression en haute-fidélité sans observation à ce niveau, avec $\rho_0 > \rho_\epsilon$. La prédiction en haute-fidélité est la courbe basse-fidélité ôtée de ses variations rapides : il y a eu un lissage entre les deux prédictions. A l'inverse, si la portée de corrélation de r_0 est inférieur à celle de r_ϵ , $\rho_0 < \rho_\epsilon$, cela signifie qu'a priori, le processus idéal varie plus vite que l'erreur. L'erreur sera alors considérée comme une fonction constante. Dans ce cas, le filtre g est un passe-haut : il conserve les variations rapides, et élimine les variations plus lentes.

Illustrations

Dans cette section, le modèle non-stationnaire est illustré à travers les deux exemples jouets présentés dans la section 1.2.1.

Sur chacun des exemples, un modèle multi-fidèle non-stationnaire (1.16) est utilisé. Le processus total s'écrit comme un processus idéal plus un processus d'erreur, $\xi(x, \delta) = \xi_0(x) + \epsilon(x, \delta)$. La moyenne du processus ξ est une constante indépendante de x ou de δ suivant une loi uniforme impropre sur \mathbb{R} (krigeage ordinaire). Le processus ξ_0 est stationnaire, et possède une fonction de covariance de Matérn 5/2, $\sigma_0^2 \mathcal{M}_{5/2}(|x - x'| / \rho_0)$. Le processus ϵ a une fonction de covariance séparable en x et δ , de la forme d'un produit entre une covariance brownienne déformée $G\sigma_0^2 \min\{t, t'\}^L$ et une covariance de Matérn 5/2 $\mathcal{M}_{5/2}(|x - x'| / \rho_\epsilon)$, où G est le rapport bruit sur signal, $G = \sigma_\epsilon^2 / \sigma_0^2$. Les hyper-paramètres de ce modèle sont $\theta = \{\sigma_0^2, \rho_0, G, L, \rho_\epsilon\}$. La loi a priori des hyper-paramètres θ , est formée de produit de lois indépendantes, $p(\theta) = p(\sigma_0^2)p(\rho_0)p(G)p(L)p(\rho_\epsilon)$. Les lois a priori des hyper-paramètres sont des lois log-normales, dont les expressions sont données dans l'annexe D. Les prédictions sont obtenues après application d'une méthode complètement bayésienne. Pour cela, comme dans la section 1.2.1, cent hyper-paramètres sont échantillonnés selon leur loi a posteriori grâce à l'algorithme de MCMC adaptatif de Haario et al. [2001]. Puis la moyenne et la variance a posteriori sont calculées suivant la méthode présentée en annexe A.3.

Le premier exemple possède deux niveaux de fidélité. Il y a huit observations au niveau un et quatre au niveau deux. L'amplitude approximative de la sortie est $r^{\text{out}} = 20$ et la taille de l'intervalle est $b^{\text{in}} = 1$, qui sont encore une fois utilisées pour construire les

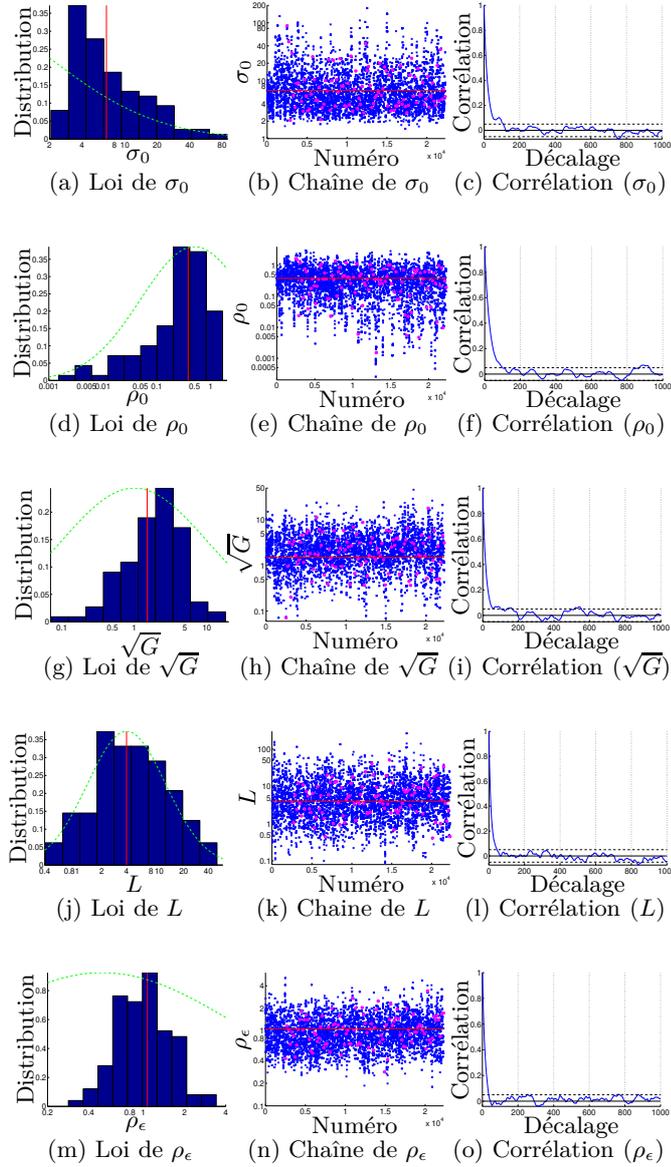


FIGURE 15 – Un échantillonnage de tous les hyper-paramètres sur un modèle non-stationnaire appliqué à l'exemple à deux niveaux de Forrester et al. [2007]. Chaque ligne correspond à un hyper-paramètre. Gauche : histogramme des échantillons des hyper-paramètres distribués selon leur loi a posteriori. Le trait pointillé est la loi a priori normalisée et le trait rouge vertical est le MAP. Milieu : la chaîne de Markov issue de l'échantillonnage. Le trait horizontal rouge est le MAP et les points mauves sont le sous-échantillonnage de la chaîne. Droite : auto-corrélation de la chaîne. Les traits pointillés horizontaux indiquent le décalage pris entre chaque sous-échantillon de la chaîne Markov.

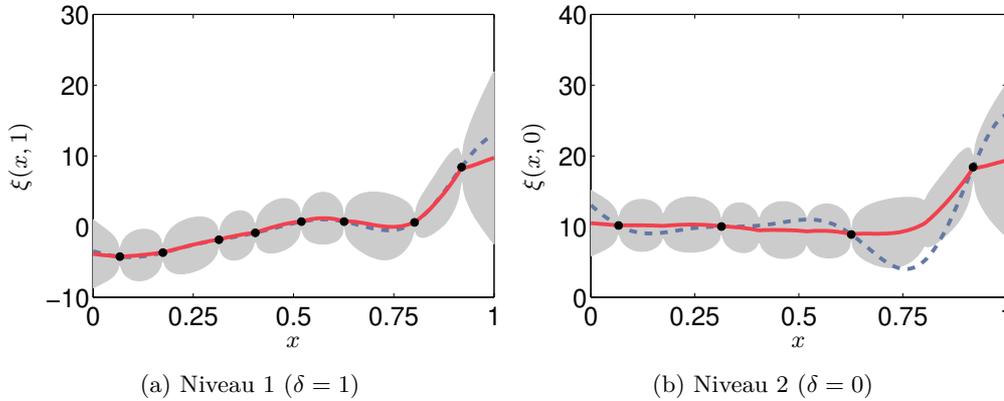


FIGURE 16 – Un exemple de prédiction sur deux niveaux avec le modèle non-stationnaire. Les points correspondent aux observations, les traits pleins aux prédictions, les aires autour aux intervalles à deux écarts-types. Les traits pointillés sont les vraies fonctions.

densités a priori des hyper-paramètres. Les valeurs de δ sont fixées arbitrairement à un pour la fonction basse-fidélité et zéro pour la fonction haute-fidélité.

La figure 15 présente les échantillons des hyper-paramètres distribués selon leur loi a posteriori, ainsi que les chaînes de Markov qui les ont engendrés. La figure 15j montre la loi a posteriori du paramètre L suit sa loi a priori. Ce résultat est cohérent, car le triplet d’hyper-paramètres (σ_0^2, G, L) nécessite au moins trois niveaux de fidélité pour pouvoir être estimé correctement, alors qu’il n’y a que deux niveaux sur cet exemple. Le modèle est non-identifiable sur cet exemple.

La figure 16 montre les prédictions obtenues par ce modèle sur les deux niveaux de fidélité. Cette figure est à comparer à la figure 9. En particulier, malgré le fait que le niveau basse-fidélité soit bien prédit, la prédiction en haute-fidélité ne correspond pas parfaitement à la vraie fonction.

Sur le second exemple, quatre niveaux de fidélité sont observés, et l’objectif est de prédire le niveau de plus haute-fidélité. Les quatre niveaux de fidélité observé correspondent à $\delta = 1/2, 1/5, 1/11$ et $1/21$. Le niveau inobservé à prédire correspond à $\delta = 0$. Les observations disponibles à chaque niveau sont au nombre de 24, 12, 4 et 2 respectivement aux niveaux $\delta = 1/2, 1/5, 1/11$ et $1/21$. L’amplitude approximative de la sortie est $r^{\text{out}} = 1$ et la taille de l’intervalle est $b^{\text{in}} = 2$.

La figure 17 présente les résultats d’échantillonnage des cinq hyper-paramètres. Les figures de gauche montrent toutes les lois a posteriori sont différentes de leurs lois a

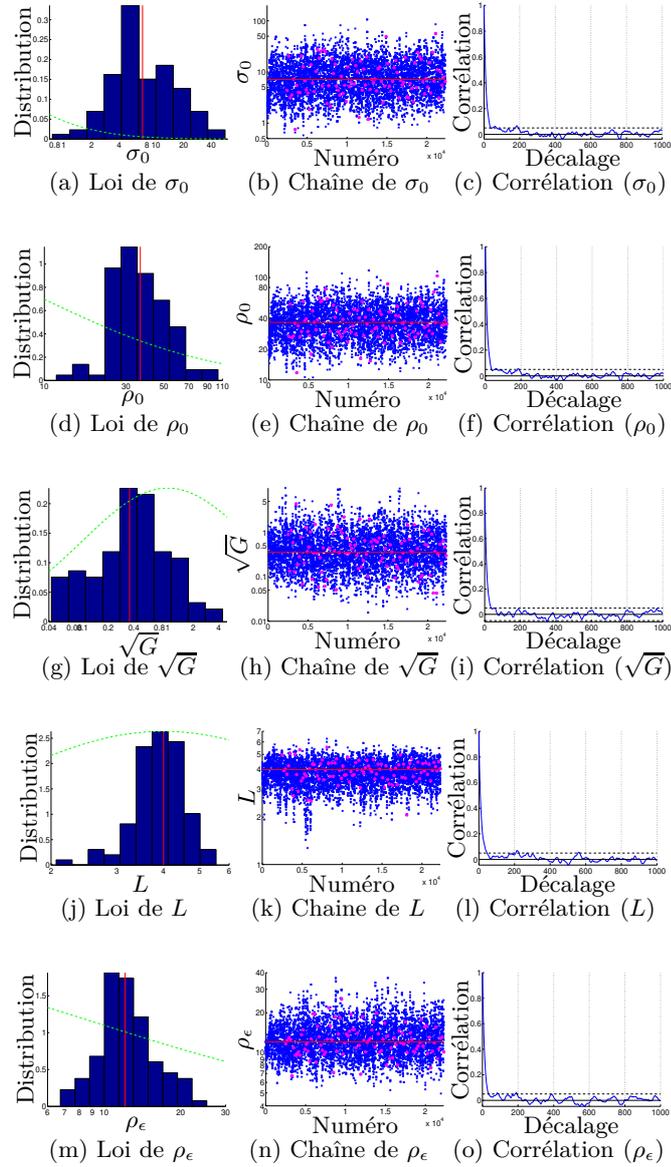


FIGURE 17 – Un échantillonnage de tous les hyper-paramètres sur un modèle non-stationnaire appliqué à l'exemple à cinq niveaux de Tuo et al. [2014]. Chaque ligne correspond à un hyper-paramètre. Gauche : histogramme des échantillons des hyper-paramètres distribués selon leur loi a posteriori. Le trait pointillé est la loi a priori normalisée et le trait rouge vertical est le MAP. Milieu : la chaîne de Markov issue de l'échantillonnage. Le trait horizontal rouge est le MAP et les points mauves sont le sous-échantillonnage de la chaîne. Droite : auto-corrélation de la chaîne. Les traits pointillés horizontaux indiquent le décalage pris entre chaque sous-échantillon de la chaîne Markov.

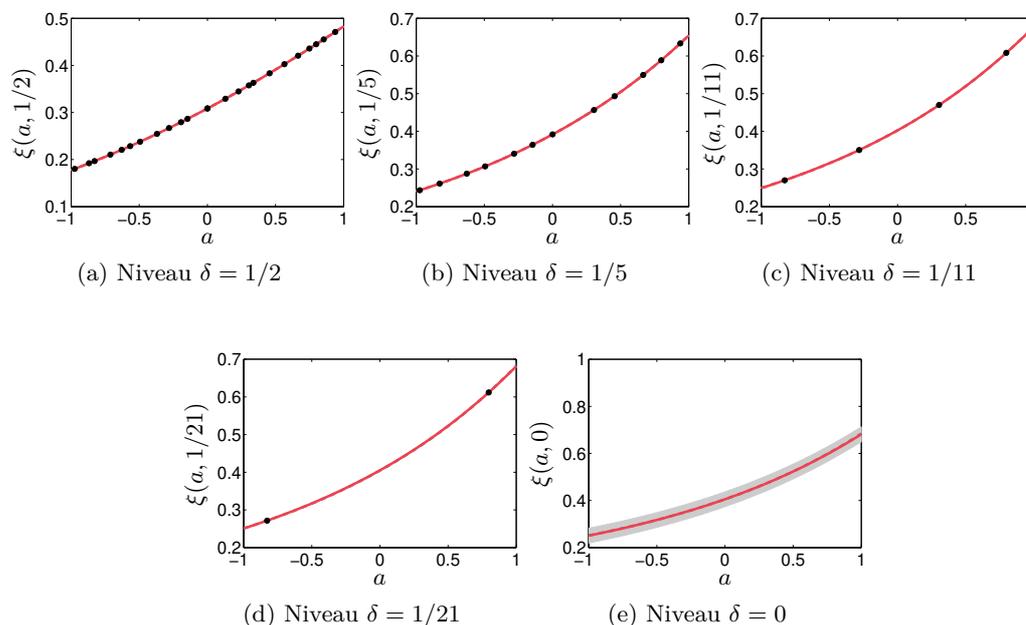


FIGURE 18 – Un exemple d’extrapolation en haute-fidélité avec le modèle non-stationnaire. Les points correspondent aux observations, les lignes pleines aux prédictions, les aires aux intervalles à deux écarts-types et les traits discontinus aux vraies fonctions.

priori, y compris les hyper-paramètres réglant le processus d’erreur. En particulier, les lois a posteriori des processus L et G sont réduites par rapport à leurs lois a priori.

La figure 18 montre les prédictions obtenues par ce modèle sur les cinq niveaux de fidélité. Cette figure est à comparer avec la figure 11. Comme dans la section 1.2.1, les prédictions sont justes sur tous les niveaux observés et trouvent les fonction correspondantes au niveau. La différence entre les deux modèles apparaît sur le dernier niveau de prédiction en $\delta = 0$. La régression en haute fidélité devient possible avec le modèle non-stationnaire, alors qu’elle était impossible avec le modèle auto-régressif. La prédiction est assez juste et les intervalles à deux écarts-types restent raisonnables. Le modèle permet donc d’envisager des régressions en haute-fidélité, et est adapté pour décrire des modèles numériques réglés par un paramètre de fidélité. Dans leurs formes actuelles, le modèle non-stationnaire est mieux adapté aux cas où le paramètre de fidélité δ est continu que le modèle auto-régressif.

1.2.3 Discussions

Comparaison entre les deux modèles

Finalement, deux modélisations par processus gaussiens sont envisageables pour prendre en compte la multi-fidélité. Dans le modèle auto-régressif, le niveau haute-fidélité est proportionnel au niveau basse-fidélité plus un terme correctif. Dans le modèle non-stationnaire, le niveau basse-fidélité est égal au niveau haute-fidélité plus un terme d'erreur.

La différence la plus visible entre les deux modèles est le type de fidélité qu'ils représentent. Dans le modèle auto-régressif, les niveaux de fidélité sont représentés de façon discrète. Dans le modèle non-stationnaire, ils sont représentés de façon continue. Même si il est possible de transformer l'un ou l'autre modèle en son équivalent continu ou discret, le modèle auto-régressif a été conçu pour des modèles dans lequel il y a peu de niveaux, alors que le modèle non-stationnaire dans le cas où le nombre de niveaux tend vers l'infini.

Une conséquence de cette différence est le nombre total d'hyper-paramètres de chaque modèle. Dans un modèle avec un nombre discret de niveau, il y a un processus gaussien par niveau, et donc un jeu d'hyper-paramètres par niveau. Le nombre total d'hyper-paramètres augmente avec le nombre de niveaux, ce qui devient problématique quand des niveaux d'observations s'ajoutent. A l'opposé, dans un modèle avec un continuum de niveaux, il n'y a que deux processus. Le nombre d'hyper-paramètres reste constant, même si de nouveaux niveaux d'observations s'ajoutent.

Une autre différence importante apparaît quand une prédiction est effectuée sur un niveau haute-fidélité inobservé. Sur un modèle auto-régressif, la prédiction sur un niveau haute-fidélité inobservé est proportionnelle à celle du dernier niveau observé plus la moyenne a priori du nouveau processus indépendant. De même, la variance a posteriori, qui est proportionnelle à celle du dernier niveau observé plus la variance a priori du nouveau processus indépendant. En pratique, cela signifie que sur un modèle avec des lois a priori peu informatives, la variance des prédictions sur des niveaux de haute-fidélité inobservés est extrêmement grande. Pour réduire la variance des prédictions des niveaux inobservés, la solution est de choisir des lois a priori informatives. A l'inverse, avec le modèle non-stationnaire, la prédiction à un niveau haute-fidélité inobservé est une version filtrée de la prédiction du dernier niveau observée. La variance a posteriori du niveau haute-fidélité inobservé est un arbitrage entre la variance a posteriori du dernier niveau observé et la variance a priori du niveau où la prédiction est effectuée. En d'autres termes, la variance a posteriori suit deux tendances opposées : d'un côté, l'absence d'observation

sur le niveau tend à faire augmenter la variance a posteriori ; de l'autre, comme la variance a priori du modèle décroît avec la fidélité, la variance a posteriori tend à diminuer. Ces deux effets opposés vont faire que la variance a posteriori d'un niveau haute-fidélité inobservé reste modérée. Il est donc possible de faire une extrapolation en haute-fidélité avec le modèle non-stationnaire. Cela sous-entend également que la structure du modèle non-stationnaire est plus informative que celle du modèle auto-régressif.

Enfin, le modèle auto-régressif présente un avantage sur le modèle non-stationnaire : beaucoup plus de littérature théorique et applicative est disponible, ce qui en fait un modèle plus simple à mettre en œuvre. D'autant plus que, dans le cas où le plan d'expérience est emboîté, le modèle auto-régressif permet des simplifications algorithmiques pour pouvoir estimer les hyper-paramètres et faire les prédictions.

En résumé :

Modèle auto-régressif	Modèle non-stationnaire
Peu de niveaux (S valant deux ou trois) d^θ croissant avec S	Continuum de niveaux d^θ constant selon S
Relation corrective : $\xi^{\text{HF}} = \xi^{\text{BF}} + \eta$	Relation déconvolutive : $\xi^{\text{BF}} = \xi^{\text{HF}} + \epsilon$
Régularité diminuant avec la fidélité	Régularité augmentant avec la fidélité
Variation rapide en haute-fidélité	Variation rapide en basse-fidélité
Régression en haute-fidélité incertaine	Régression en haute-fidélité possible
Simplifications algorithmiques	Peu de simplifications algorithmiques
Beaucoup de littérature	Peu de littérature

où d^θ est le nombre d'hyper-paramètres.

Améliorations potentielles et perspectives

Cette section a présenté les modélisation de la multi-fidélité par processus gaussiens. L'objectif est de représenter les relations entre les niveaux, notamment via la fonction de covariance. Cependant, des améliorations de ces modèles sont envisageables.

Il ressort que la covariance modélisant la relation entre les niveaux peut être améliorée. Les covariances proposées sont séparables entre x et δ , c'est-à-dire que la fonction de covariance totale s'exprime par un produit de covariances : $k((x, \delta), (x', \delta')) = r(\delta, \delta') \cdot \tilde{k}(x, x')$. Or, ce genre de covariance peut avoir des propriétés d'interpolation étranges. Par exemple, Stein [2011] montre que les covariances séparables ne respectent pas la propriété d'écrantage de l'interpolation, ce qui peut aboutir à des résultats contre-intuitifs. Des relations différentes entre entrées x et fidélité δ sont envisageables, par exemple en s'inspirant des différentes covariances spatio-temporelles proposées dans la

littérature, où le paramètre de fidélité remplacerait le paramètre de temps. Ainsi, Stein [2005] étudie les propriétés souhaitables sur de telles covariances, Gneiting [2002] propose quelques fonctions de covariance inséparables simples à implémenter, et Subba Rao et Terdik [2013] et Horrell et Stein [2017] proposent d'autres covariances en se basant sur la transformée de Fourier. De manière plus générale, les covariances non-stationnaires pourraient être mieux adaptées pour modéliser l'évolution en fonction des niveaux. De telles covariances peuvent être obtenues soit par transformation de l'espace des entrées [Picheny et Ginsbourger, 2013], soit par convolution de noyaux gaussiens [Mateu et al., 2013; Paciorek et Schervish, 2006].

Un autre aspect qui pourrait être exploré serait la prise en compte de plusieurs sorties en même temps. Jusqu'ici, le modèle numérique multi-fidèle considéré prenait un vecteur d'entrée (x, δ) , et renvoyait une sortie scalaire Z . Or, dans le cas où ce modèle numérique renverrait plusieurs sorties en même temps, nous pourrions prendre compte les éventuelles corrélations entre les sorties pour améliorer la prédiction. Une solution simple serait de considérer que le modèle numérique prend un troisième type d'entrée, noté a , qui règle le type de sortie renvoyé, et qu'à chaque fois que le simulateur est exécuté pour un vecteur d'entrée (x, δ) , il renvoie toutes les sorties correspondantes à toutes les valeurs de a possibles. Une telle entrée serait une variable quantitative. Qian et al. [2008] proposent une façon de modéliser un simulateur ayant une variable quantitative. Leur approche décrit la covariance totale comme une covariance séparable entre a et (x, δ) , dont les hyper-paramètres respectent certaines conditions. Une autre solution peut être de considérer le modèle de Gneiting et al. [2010] fondé sur des Matérn croisées, qui permet de prendre en compte plusieurs sorties en même temps.

1.3 Extension aux simulateurs stochastiques multi-fidèles

Jusqu'ici les modèles traités supposaient que la sortie du simulateur considéré était déterministe, c'est-à-dire que la réponse Z du simulateur en un point x au niveau de fidélité δ était de la forme $Z = \xi(x, \delta)$, où ξ est modélisé par un processus gaussien. Cependant, nous souhaitons également pouvoir modéliser des simulateurs multi-fidèles stochastiques, comme le modèle numérique fondé sur Fire Dynamics Simulator (FDS) (voir chapitre 3). Dans un simulateur stochastique, la réponse Z est tirée aléatoirement selon une loi de probabilité qui dépend des entrées (x, δ) . La question devient alors de modéliser cette loi de probabilité en tout point d'entrée. Dans cette section, la réponse Z est supposée scalaire.

Une hypothèse courante sur les modèles stochastiques est que la distribution de la

sortie en une entrée (x, δ) donnée est une loi normale dont la moyenne et la variance dépendent des entrées :

$$Z|x, \delta; \xi, \lambda \sim \mathbf{N}(\xi(x, \delta), \lambda(x, \delta)). \quad (1.25)$$

Cette hypothèse est faite par exemple par Ankenman et al. [2010] sur un modèle de queues de clientèles, par Picheny et al. [2013] sur un simulateur de système nucléaire, et par Andrianakis et al. [2015] sur un simulateur étudiant la transmission du Virus de l'Immuno-déficience Humaine (VIH). Son principal intérêt est qu'elle simplifie les calculs de la loi a posteriori : si les sorties Z_i en des points (x_i, δ_i) suivent des lois normales indépendantes dont la variance est connue et si la loi a priori de ξ est un processus gaussien, alors les lois a posteriori de Z et de ξ en un point (x, δ) sont des lois normales dont les moyennes et les variances sont calculables avec les formules du krigeage [Stein, 1999]. Cette hypothèse n'est cependant pas systématique dans l'étude de simulateur stochastique. Ainsi, Oakley et Youngman [2017] calibre un simulateur modélisant le risque de cancer de l'intestin en fonction du patient. La sortie de ce simulateur suit une loi binomiale dont les paramètres dépendent des caractéristiques du patient. De même, Moutoussamy et al. [2015], dans un contexte de sécurité nucléaire, traitent le cas où la distribution de sortie est inconnue et doit être estimée. Des simulateurs stochastiques multi-fidèles ont déjà été considérés par Raissi et al. [2017] et Pang et al. [2017]. Les deux articles considèrent un modèle multi-fidèle à deux niveaux et ajoutent une loi normale en sortie, dont la moyenne est modélisée par un processus gaussien multi-fidèle auto-régressif (section 1.2.1) et dont la variance est constante sur un niveau donné et différente sur chaque niveau.

Pour la modélisation de la fonction moyenne ξ , nous réutilisons les processus gaussiens multi-fidèles décrit dans la section 1.2. Dans cette section, nous allons définir un a priori sur la fonction de variance de la partie stochastique λ dans un cadre multi-fidèle. Nous allons d'abord présenter les modèles de simulateurs stochastiques hétéroscédastiques, puis nous présentons notre modèle de λ dans un cadre multi-fidèle, que nous illustrerons sur deux exemples.

1.3.1 Modèles de la littérature

Dans cette section, nous présentons les modèles de la littérature traitant de sorties stochastiques hétéroscédastiques dans un contexte avec un seul niveau de fidélité. Par simplicité d'écriture, le vecteur des entrées est noté x .

Dans un cadre bayésien, les fonctions inconnues sont modélisées avec des processus

aléatoires. Parmi tous les processus envisageables, le plus simple et le plus connu est le processus gaussien. Ainsi, Ankenman et al. [2010] et Marrel et al. [2012] proposent de placer un processus gaussien a priori sur la variance λ . Cependant cette approche est limitée car un processus gaussien suppose que la fonction modélisée peut prendre n'importe quelle valeur réelle, ce qui n'est pas le cas pour une fonction variance, qui doit rester positive. Ankenman et al. [2010] et Marrel et al. [2012] résolvent le problème en ajoutant une moyenne constante positive et en supposant que la fonction de variance reste loin de zéro.

Une autre idée de la littérature consiste à placer un processus gaussien sur le logarithme de la fonction de variance, $l(x) = \log(\lambda(x))$. Dans cette approche, deux processus gaussiens indépendants sont définis : un pour la moyenne ξ , et un pour le logarithme de la variance l . Il y a alors deux vecteurs d'hyper-paramètres à estimer, un pour le processus gaussien de ξ , θ_ξ et un pour celui de l , θ_l . Boukouvalas [2010] propose de les estimer successivement, D'abord, une estimation initiale $\hat{\theta}_{\xi_0}$ des hyper-paramètres du processus ξ peut être obtenue en supposant la log-variance l être constante. Ensuite, en supposant $\hat{\theta}_{\xi_0}$ fixé, le processus ξ est un processus gaussien dont les fonctions de moyenne et de covariance sont connues. Dans ces conditions, une approximation de l aux points d'observations est calculée, puis les hyper-paramètres du processus de l sont estimés, $\hat{\theta}_{l_1}$. A partir de là, la log-variance l est estimée aux points d'observations, qui est supposée ensuite connue, pour pouvoir ré-estimer les hyper-paramètres du processus ξ , $\hat{\theta}_{\xi_1}$. Les étapes sont réitérés jusqu'à convergence. La principale difficulté réside dans l'estimation de l aux points d'observations. En effet, contrairement au processus ξ , il n'y a pas d'observations (même bruitée) de la fonction de log-variance l . Conséquemment, il devient impératif de pouvoir approcher la fonction de log-variance l pour pouvoir appliquer cet algorithme. Quand il y a plusieurs observations $(z_i^{(j)})_{1 \leq j \leq m}$ pour le même point x_i , une approximation naturelle de la fonction de log-variance $l(x_i)$ est le logarithme de la variance empirique $\log\left(\frac{1}{m-1} \sum_{j=1}^m \left[z_i^{(j)} - \frac{1}{m} \sum_{j'=1}^m z_i^{(j')}\right]^2\right)$, comme suggéré par Kersting et al. [2007]. Cependant, cette méthode est inapplicable quand il n'y a qu'une observation par point. Boukouvalas et Cornford [2009] proposent une méthode pour obtenir une estimation dans ce cas, fondée sur les simulations conditionnelles.

D'autres modèles existent pour prendre en compte l'hétéroscédasticité, comme par exemple celui Gramacy et Lee [2008] fondée sur une décomposition l'espace des entrées selon un arbre. Cependant, tous ces modèles restent relativement complexes. Nous souhaitons un modèle relativement simple pour prendre en compte les variations de la variance de l'effet stochastique.

1.3.2 Modèle simplifié dédié à la multi-fidélité

Nous cherchons à modéliser une log-variance dépendant des entrées (x, δ) , $l(x, \delta) = \log(\lambda(x, \delta))$ pour pouvoir notamment obtenir des estimations de la variance en des points et des niveaux non-observés. De plus, nous voulons mener une approche complètement bayésienne en intégrant l'incertitude sur λ . Pour cela, un modèle a priori de processus gaussien est ajouté sur l . Ce processus gaussien est supposé avoir une moyenne constante, dont la constante suit une loi a priori gaussienne. Ainsi, si $X = (x_i, \delta_i)_{1 \leq i \leq n}$ est un ensemble de n points,

$$l(X)|X, k_\lambda, l_0, \varsigma^2; l_m \sim \mathbf{N}(l_m \mathbf{1}_n, K_\lambda(X, X)) \quad (1.26a)$$

$$l_m|X, k_\lambda, l_0, \varsigma^2 \sim \mathbf{N}(l_0, \varsigma^2), \quad (1.26b)$$

où $\mathbf{1}_n$ est le vecteur rempli de un de longueur n , l_m est la moyenne du processus gaussien, l_0 est sa valeur à priori, ς^2 sa variance a priori, k_λ est la fonction de covariance du processus l , et $K_\lambda(X, X) = (k_\lambda((x_i, \delta_i), (x_j, \delta_j)))_{1 \leq i, j \leq n}$ est la matrice de covariance du processus aux points d'observation.

A partir de là, le modèle est simplifié pour pouvoir inférer la variance de l'effet stochastique. Le choix des simplifications est guidé par l'application FDS que nous cherchons à modéliser. Une première simplification consiste à supposer que la fonction de covariance du processus l est stationnaire. Celle-ci peut alors se réécrire

$$k_\lambda((x, \delta), (x', \delta')) = \sigma_\lambda^2 r_\lambda \left(\left(\frac{x_k - x'_k}{\rho_k} \right)_{1 \leq k \leq d}, \frac{\delta - \delta'}{\rho_\delta}; \widetilde{\theta}_\lambda \right), \quad (1.27)$$

où d est la dimension du vecteur x , σ_λ^2 est la variance de la covariance, r_λ est une fonction de corrélation, $(\rho_k)_{1 \leq k \leq d}, \rho_\delta$ sont $d + 1$ paramètres de portées, et $\widetilde{\theta}_\lambda$ sont d'éventuels autres hyper-paramètres de la covariance.

Ensuite, une hypothèse forte est ajoutée sur les hyper-paramètres de la fonction de corrélation r_λ . D'abord la fonction de variance λ est supposée dépendre uniquement du niveau de fidélité δ , et donc ne plus varier en fonction des entrées x . Ensuite, nous supposons qu'entre deux niveaux de fidélité observés, la différence de niveaux est beaucoup plus grande que la portée de corrélation. En d'autres termes, si (x, δ) et (x', δ') sont deux points d'observation, les portées de la fonction de covariance r_λ de la log-variance sont

supposées respecter les inégalités suivantes

$$\forall k \in \llbracket 1; d \rrbracket, \quad |x_k - x'_k| \ll \rho_k, \quad (1.28a)$$

$$|\delta - \delta'| \gg \rho_\delta, \quad (1.28b)$$

c'est-à-dire que les variations selon x sont insuffisantes pour voir varier la fonction de variance λ , et que la moindre variance selon δ entraîne des modifications importantes. En conséquence, d'une part, (1.28a) donne $l(x, \delta) = l(\delta)$, la variance de la partie stochastique ne dépend que du niveau de fidélité. Cette simplification est également faite sur d'autres modèles stochastiques multi-fidèles de la littérature [Pang et al., 2017; Raissi et al., 2017].

D'autre part, si $(\delta_s)_{1 \leq s \leq S}$ sont S niveaux de fidélité différents, alors

$$\begin{aligned} (l(\delta_s))_{1 \leq s \leq S} \mid (\delta_s)_{1 \leq s \leq S}, \sigma_\lambda^2, l_0, \varsigma^2; l_m &\sim \mathbf{N} \left(l_m \mathbf{1}_S, \sigma_\lambda^2 I_S \right) \\ l_m \mid (\delta_s)_{1 \leq s \leq S}, \sigma_\lambda^2, l_0, \varsigma^2 &\sim \mathbf{N} \left(l_0, \varsigma^2 \right) \end{aligned}, \quad (1.29)$$

avec I_S la matrice identité de taille S . La matrice de covariance $K_\lambda(X, X)$ devient proportionnelle à la matrice identité, car $k_\lambda((x, \delta), (x', \delta')) = k_\lambda(\delta, \delta')$ vaut σ_λ^2 si $\delta = \delta'$ (termes diagonaux), et zéro si $\delta \neq \delta'$ (termes hors diagonaux, (1.28b)). Après intégration de l_m , et en écrivant $\varsigma^2 = a\sigma_\lambda^2$, le modèle a priori de la log-variance devient

$$(l(\delta_s))_{1 \leq s \leq S} \mid (\delta_s)_{1 \leq s \leq S}, \sigma_\lambda^2, l_0, a \sim \mathbf{N} \left(l_0 \mathbf{1}_S, \sigma_\lambda^2 (I_S + aU_S) \right), \quad (1.30)$$

où U_S est la matrice de taille S remplie de un. L'équation (1.30) est le modèle a priori que nous utilisons dans toute la thèse pour modéliser la variance λ d'un simulateur stochastique multi-fidèle.

Ce modèle dépend de trois hyper-paramètres qui doivent être fixés par l'utilisateur, l_0 , σ_λ^2 et a . La moyenne l_0 est la moyenne a priori de la log-variance du simulateur. Le paramètre σ_λ^2 est un paramètre positif qui règle la variance a priori sur la valeur de la log-variance. Enfin, le paramètre positif a règle la corrélation des valeurs de l entre les niveaux. Quand $a = 0$, les valeurs de $l(\delta)$ sont a priori indépendantes entre elles. Et quand a tend vers l'infini, les valeurs de $l(\delta)$ tendent à être toutes égales. Pour mieux comprendre l'impact de σ_λ^2 et de a , remarquons que si δ et δ' sont deux paramètres de

fidélité différents, alors, a priori,

$$\text{var} [\log (\lambda(\delta))] = \sigma_{\lambda}^2(1+a) \quad (1.31a)$$

$$\text{var} [\log (\lambda(\delta)) - \log (\lambda(\delta'))] = \text{var} \left[\log \left(\frac{\lambda(\delta)}{\lambda(\delta')} \right) \right] = 2\sigma_{\lambda}^2. \quad (1.31b)$$

Nous cherchons à dire qu'a priori les valeurs de $\lambda(\delta)$ sont peu connues, mais à peu près égales entre elles. Cela se traduit par une valeur de $\sigma_{\lambda}^2(1+a)$ très élevée, et par une valeur de $2\sigma_{\lambda}^2$ relativement plus faible. Une méthode permettant d'obtenir des valeurs est proposée dans l'annexe D.

L'intérêt de ce modèle est que, malgré sa simplicité, il permet de pouvoir inférer sur les valeurs des variances λ sur les niveaux peu ou pas observés. Si tous les niveaux étaient observés avec suffisamment de points, les données seraient suffisantes pour pouvoir obtenir une estimation des λ , et le modèle a priori pourrait supposer que les variances à chaque niveau sont indépendantes, $a = 0$. Cependant, certains niveaux d'intérêt en haute-fidélité, disposent de peu, voire pas, de points d'observations. Afin d'estimer la valeur de la variance de l'effet stochastique sur ces niveaux, une corrélation importante est ajoutée entre toutes les valeurs des variances λ . A noter que, dans le cas où l'utilisation d'une transformation logarithmique n'est pas souhaitable, il est possible de considérer un modèle similaire de processus gaussien sur la variance directement, en imposant une contrainte de positivité (de manière similaire à la méthode de Marrel et al. [2012]).

Ce modèle a une particularité intéressante : il ne propose pas de tendance particulière sur l'évolution de la variance du bruit en fonction de la fidélité. Ceci est justifié par nos applications sur des cas concrets. En effet, sur ces simulateurs réellement coûteux, il ne nous semblait pas que la variance suive une tendance particulière selon la fidélité (ni croissance, ni décroissance). Par conséquent, nous n'avons pas rajouté de tendance dans notre modèle de la variance. En particulier, il est possible que la variance d'un niveau de plus haute-fidélité soit plus importante que celle d'un niveau de plus-basse fidélité.

Remarque 3. Le déterminant et l'inverse de la matrice $\sigma_{\lambda}^2 I_S + a\sigma_{\lambda}^2 U_S$ peuvent être explicités :

$$\det \left(\sigma_{\lambda}^2 I_S + a\sigma_{\lambda}^2 U_S \right) = \left(\sigma_{\lambda}^2 \right)^S \cdot (1+aS) \quad (1.32a)$$

$$\left(\sigma_{\lambda}^2 I_S + a\sigma_{\lambda}^2 U_S \right)^{-1} = \frac{1}{\sigma_{\lambda}^2} \left(I_S - \frac{a}{1+aS} U_S \right). \quad (1.32b)$$

Par conséquent, la loi a priori des log-variances sur S niveaux distincts peut s'écrire

simplement :

$$p\left(\left(l(\delta_s)\right)_{1 \leq s \leq S} \mid (\delta_s)_{1 \leq s \leq S}, \sigma_\lambda^2, l_0, a\right) = \frac{\exp\left(-\frac{S}{2\sigma_\lambda^2} \left[\frac{(\bar{l}-l_0)^2}{1+aS} + \overline{(l-\bar{l})^2}\right]\right)}{\sqrt{(2\pi\sigma_\lambda^2)^S (1+aS)}}, \quad (1.33)$$

où $\bar{l} = (1/S) \sum_{s=1}^S l(\delta_s)$ et $\overline{(l-\bar{l})^2} = (1/S) \sum_{s=1}^S (l(\delta_s) - \bar{l})^2$. Le terme dans l'exponentielle est constitué de deux parties. La première partie tend à favoriser des vecteurs de log-variances dont la moyenne est proche de la valeur a priori. La seconde à favoriser des vecteurs de log-variances ayant des valeurs proches entre eux.

Il est possible de considérer une limite de cet a priori. Dans le cas où aucune information n'est disponible sur la moyenne a priori l_0 , le coefficient a tend vers $+\infty$, en supposant la variance σ_λ^2 fixée. Dans ce cas, l'a priori impropre devient une loi exponentielle en la variance empirique du vecteur des log-variances

$$p\left(\left(l(\delta_s)\right)_{1 \leq s \leq S} \mid (\delta_s)_{1 \leq s \leq S}, \sigma_\lambda^2, l_0, a\right) \propto \exp\left(-\frac{S}{2\sigma_\lambda^2} \overline{(l-\bar{l})^2}\right). \quad (1.34)$$

Comme cet a priori est impropre, il faudrait chercher des conditions qui assurent que les loi a posteriori soient bien définies. Nous recommandons plutôt d'utiliser la loi a priori (1.33) avec une valeur de a élevée.

1.3.3 Illustrations

Dans cette section, nous illustrons le modèle a priori de la variance λ (1.30) à travers deux exemples simples.

D'abord, considérons l'exemple de Forrester et al. [2007], développé dans la section 1.2. Le simulateur possède deux niveaux de fidélité $f_1(x) = 0.5(6x - 2)^2 \sin(12x - 4) + 10x - 5$ et $f_2(x) = (6x - 2)^2 \sin(12x - 4) + 10$. A la différence de la section 1.2, nous supposons maintenant que les observations de ces deux niveaux sont bruitées, avec un bruit gaussien de variance $\lambda(1) = \lambda_1 = 1.1^2$ pour le niveau de basse-fidélité et $\lambda(2) = \lambda_2 = 1.4^2$ pour le niveau de haute-fidélité. Dans cet exemple, la variance du bruit du niveau de haute-fidélité est plus importante que celle du niveau de basse-fidélité.

La sortie du simulateur en un point (x, s) suit une loi normale $\mathbf{N}(\xi_s(x), \lambda_s)$ où la fonction de moyenne et les valeurs de la variance sont inconnues. Comme il n'y a que deux niveaux de fidélité, la moyenne $\xi_s(x)$ est modélisé par processus gaussien multi-fidèle auto-régressif (voir section 1.2.1). Les covariances individuelles sont des covariance de Matérn 5/2. Les lois a priori sur les hyper-paramètres des fonctions de covariances

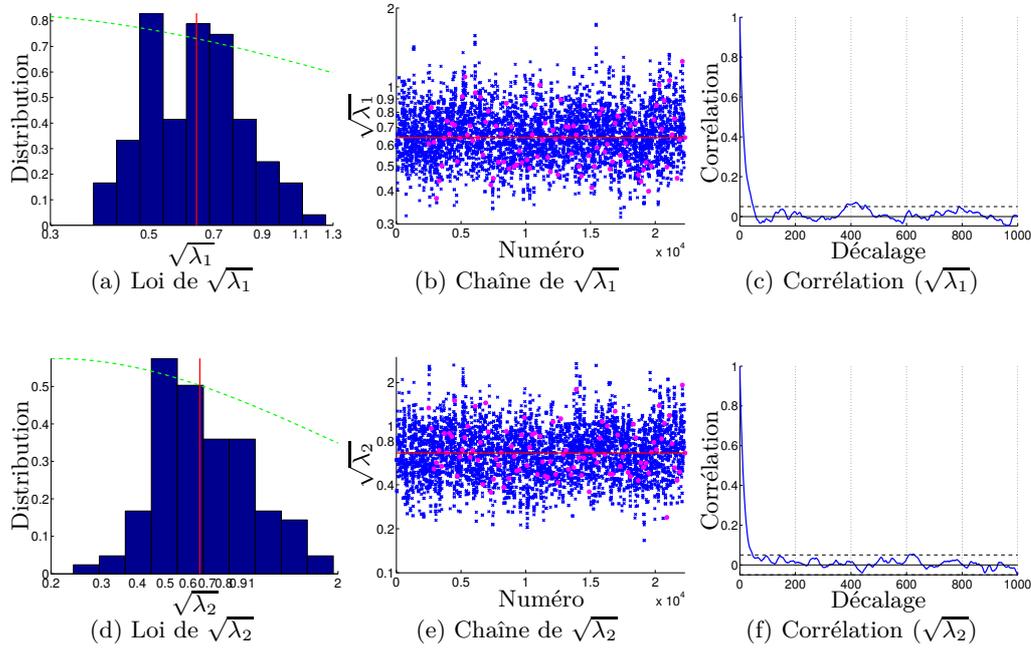


FIGURE 19 – Un échantillonnage des deux variances du bruit appliqué à l'exemple à deux niveaux bruités de Forrester et al. [2007]. Chaque ligne correspond à une variance d'un niveau. Gauche : histogramme des échantillons des variances de bruit distribués selon leur loi a posteriori. Le trait pointillé est la loi a priori normalisée et le trait rouge vertical est le MAP. Milieu : la chaîne de Markov issue de l'échantillonnage. Le trait horizontal rouge est le MAP et les points mauves forment le sous-échantillonnage de la chaîne. Droite : auto-corrélation de la chaîne. Les traits pointillés horizontaux indiquent le décalage pris entre chaque sous-échantillon de la chaîne Markov.

sont des lois indépendantes. Ainsi, pour les hyper-paramètres de variances, $\log(\sigma_s^2) \sim \mathcal{N}(\log(0.2^2), \log(100)^2)$, pour les hyper-paramètres de portées, $\log(\rho_s) \sim \mathcal{N}(\log(0.5), \log(10)^2)$ et pour le paramètre de corrélation entre les niveaux, $\gamma_1 \sim \mathcal{N}(1, 2^2)$.

La variance du bruit est modélisée par deux paramètres à estimer, λ_1 et λ_2 , une variance de bruit par niveau. Ces deux paramètres reçoivent le modèle a priori décrit dans la section 1.3. Il y a trois hyper-paramètres à fixer : l_0 , σ_λ^2 et a . Nous fixons $l_0 = \log(0.2^2)$. La valeur de $\exp(l_0)$ est supposée très incertaine, de l'ordre d'un facteur cent. Ainsi, la variance totale a priori des valeurs de $\log(\lambda_s)$ est de $\log(100)^2$, soit $(1+a)\sigma_\lambda^2 = \log(100)^2$. Nous supposons que 99% de cette variance vient de la variance sur la valeur moyenne, c'est à dire que $a\sigma_\lambda^2 / ((1+a)\sigma_\lambda^2) = 0.99$. Les 1% restants des variations de la variance du bruit autour de cette moyenne. Ce rapport donne un facteur $a = 99$. Dans ces conditions,

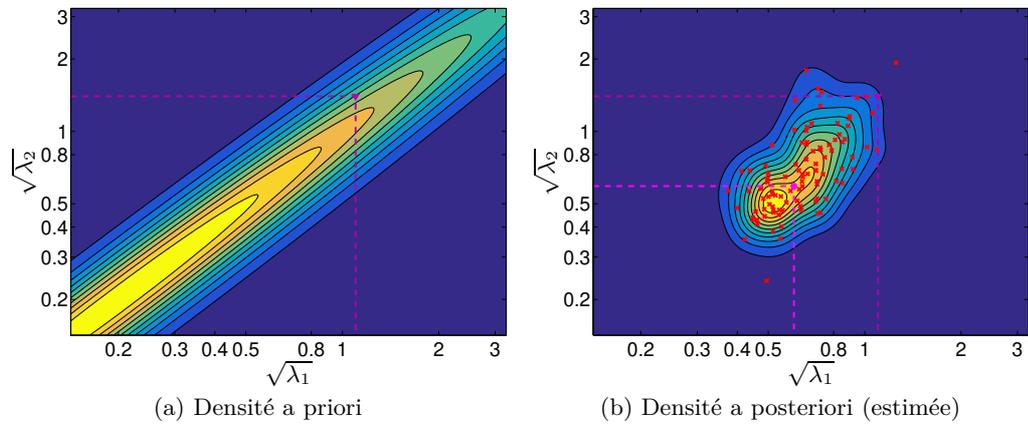


FIGURE 20 – Distribution a priori et a posteriori des écarts-types du bruit (en échelle logarithmique). Le triangle violet correspond aux vraies valeurs des écart-type du bruit (pointillés figure (a) et figure (b)), le rond rose est le maximum a posteriori (pointillé figure (b)), les croix sont des valeurs issues de l'échantillonnage. La densité a posteriori est estimée par un estimateur à noyau.

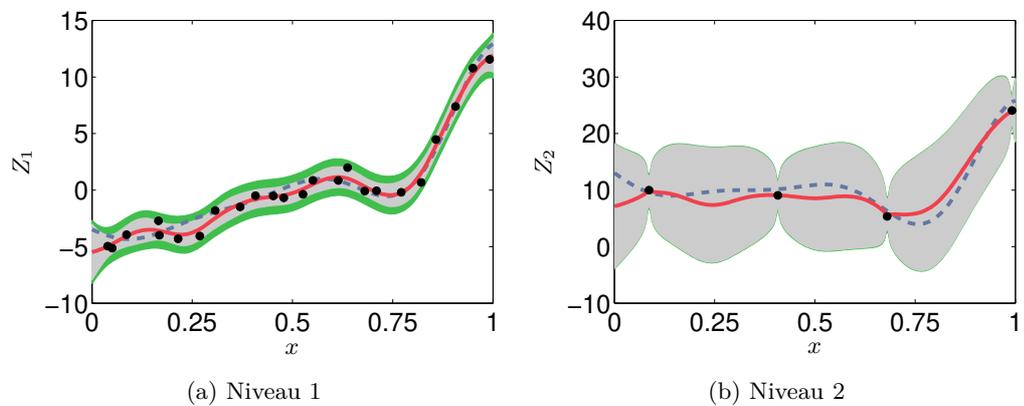


FIGURE 21 – Des prédictions sur un simulateur bruité multi-fidèle à deux niveaux. Chaque image correspond à un niveau de fidélité. Les points sont les observations bruitées, les traits pointillés sont les vraies fonctions moyennes du simulateur, les traits pleins sont les moyennes a posteriori, la première zone est l'intervalle à deux écarts-types de l'estimation de la fonction moyenne, la seconde zone est l'intervalle à deux écarts-types des observations futures.

$\sigma_\lambda^2 = \log(100)^2 / (1 + a) = \log(100)^2 / 100$. La loi a priori du couple des log-variances est

$$\begin{pmatrix} \log(\lambda_1) \\ \log(\lambda_2) \end{pmatrix} \sim \mathbf{N} \left(\begin{pmatrix} \log(0.2^2) \\ \log(0.2^2) \end{pmatrix}, \begin{pmatrix} \log(100)^2 & 0.99 \log(100)^2 \\ 0.99 \log(100)^2 & \log(100)^2 \end{pmatrix} \right). \quad (1.35)$$

Il y a quatre observations au niveau haute-fidélité et vingt-quatre au niveau basse-fidélité. Une méthode complètement bayésienne est appliquée, avec intégration des hyper-paramètres de la covariance et des variances du bruit selon leurs lois a posteriori. Pour cela, nous utilisons comme précédemment le MCMC adaptatif proposé par Haario et al. [2001]. Les résultats d'échantillonnage sur les variances λ sont présentés figure 19. La comparaison des densités a priori et a posteriori des deux variances permet de voir que les lois a posteriori sont plus resserrées que leurs a priori. La loi de λ_2 est plus large que celle de λ_1 , ce qui est cohérent puisqu'il y a moins d'observations sur le second niveau que sur le premier. La figure 20 vient en complément pour afficher les corrélations. La figure 20a présente la densité a priori des variances du bruit. La figure 20b présente une estimation à noyau de la densité a posteriori des variances des bruits à partir des échantillons de la loi a posteriori. Les points basse-fidélité permettent d'obtenir une estimation de λ_1 , puis l'a priori permet d'en déduire une valeur de λ_2 . Enfin, la figure 21 présente les prédictions avec intégration des hyper-paramètres en basse et haute-fidélité. Les prédictions sont correctes et correspondent à ce qui peut être attendu d'une régression par processus gaussien dans un contexte bruité.

Le second exemple a été proposé par Au et Beck [2001]. Un pseudo-simulateur simule un tremblement de terre sur une structure. La réponse temporelle de la structure est notée X , et suit l'équation différentielle ordinaire stochastique

$$\ddot{X}(t) + 2\omega_0\zeta\dot{X}(t) + \omega_0^2X(t) = W(t), \quad \dot{X}(0) = 0, \quad X(0) = 0 \quad (1.36)$$

où ω_0 et ζ sont respectivement la pulsation propre et l'amortissement, qui dépendent de propriétés physiques de la structure, et W est un bruit blanc gaussien dont l'intensité spectrale est de un, et qui représente le séisme. La quantité d'intérêt est la réponse du bâtiment en fonction de ses paramètres physiques. Le tremblement dure $t^{\text{end}} = 30$ s et la valeur d'intérêt est le maximum en valeur absolue de X sur cette période. Les sorties sont étudiées en échelle logarithmique. Le simulateur idéal est donc

$$\begin{array}{ccc} \mathbb{X} & \rightarrow & \mathbb{R} \\ (\omega_0, \zeta) & \mapsto & \max \left\{ \log(|X(t)|), t \in [0; t^{\text{end}}] \right\} \end{array} \quad (1.37)$$

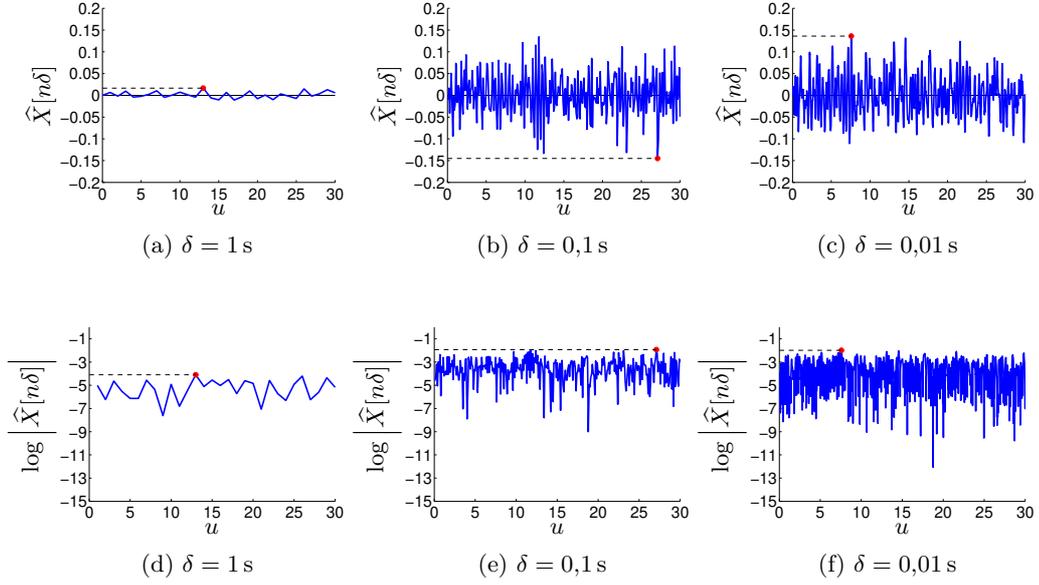


FIGURE 22 – Des exemples de trajectoires du simulateur \widehat{X}_n à différents pas de temps. En haut, trajectoire réelle, en bas, en échelle logarithmique. La pulsation propre et l’amortissement sont $\omega_0 = 15 \text{ rad} \cdot \text{s}^{-1}$ et $\zeta = 0.2$. Le point indique la valeur renvoyée par le simulateur.

où $\mathbb{X} = [0; 30] \text{ rad} \cdot \text{s}^{-1} \times [0; 1]$.

Une méthode par différence finie, le schéma d’Euler explicite exponentiel proposé par Jentzen et Kloeden [2009], est utilisée pour résoudre cette équation. Un pas de temps δ est fixé, et la solution est approchée par une approximation $\widehat{X}_n \approx X(n\delta)$. Nous choisissons cette méthode pour obtenir un simulateur-jouet multi-fidèle stochastique, et nous ne prétendons pas que ce schéma de discrétisation est la meilleure façon de résoudre (1.36). Notre simulateur multi-fidèle est alors

$$\begin{aligned} \mathbb{X} \times [0; 1] \text{ s} &\rightarrow \mathbb{R} \\ (\omega_0, \zeta, \delta) &\mapsto \max \left\{ \log \left(\left| \widehat{X}_n \right| \right), 0 \leq n \leq \left\lceil \frac{t_{\text{end}}}{\delta} \right\rceil \right\}. \end{aligned} \quad (1.38)$$

Ce simulateur est stochastique, puisque X dépend d’un bruit blanc gaussien. La figure 22 montre un exemple de trajectoire X possible pour une entrée $x = (\omega_0, \zeta)$ et trois niveaux de fidélité δ . La figure 23 trace les histogrammes des sorties obtenues en lançant mille simulations au même point d’entrée. Ces histogrammes suivent la loi normale dont les paramètres sont calculés empiriquement, ce qui signifie que la sortie Z en un point donné

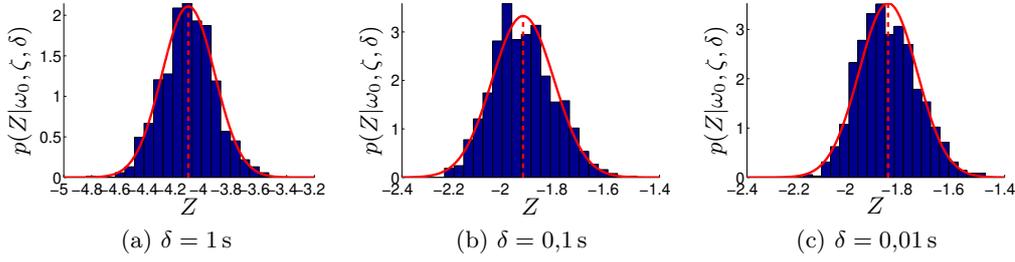


FIGURE 23 – Des histogrammes des sorties du simulateur en un point et trois niveaux de fidélité. Chaque histogramme est tracé avec mille simulations effectuées au même point. La pulsation propre et l’amortissement valent $\omega_0 = 15 \text{ rad} \cdot \text{s}^{-1}$ et $\zeta = 0.2$. Le trait plein est la densité de probabilité d’une loi normale dont la moyenne et la variance sont la moyenne et la variance empiriques sans biais.

peut être approchée par une loi normale, $Z|x, \delta \sim \mathcal{N}(\xi(x, \delta), \lambda(x, \delta))$, où ξ est la fonction de moyenne du simulateur et λ est la fonction de variance du simulateur.

Enfin, la figure 24 présente les fonction de moyenne et de variance du simulateur. La fonction de moyenne évolue beaucoup entre les niveaux de basse-fidélité et les niveaux de haute-fidélité, mais qu’elle reste assez régulière. La fonction de variance est quasiment constante à un niveau δ donné. Cela signifie que les valeurs de la variance λ peuvent être supposées constantes à un niveau δ donné et ne pas dépendre des autres entrées x . De plus, les valeurs de $\lambda(\delta)$, bien que différentes à chaque niveau, restent relativement proche les unes des autres.

Les observations du simulateur sont au nombres respectifs de 96, 48, 24, 12 et 6 aux niveaux $\delta = 1 \text{ s}$, $0,5 \text{ s}$, $0,25 \text{ s}$, $0,2 \text{ s}$ et $0,1 \text{ s}$. Il n’y a aucune observation au niveau $\delta = 0,01 \text{ s}$.

La sortie en un point (x, δ) est supposée suivre une loi normale de moyenne $\xi(x, \delta)$ et de variance $\lambda(\delta)$. Un a priori de processus gaussien est placé sur ξ . Comme le simulateur est fondé sur une équation différentielle qui est approchée avec différents pas de temps, le modèle utilisé pour la moyenne est le modèle par processus gaussien multi-fidèle non-stationnaire décrit section 1.2.2 :

$$\xi|m, \theta \sim \text{GP}(m, \sigma_0^2 \mathcal{M}_{5/2}(x, x'|\rho_0) + \sigma_0^2 G \min\{\delta, \delta'\}^L \mathcal{M}_{5/2}(x, x'|\rho_\epsilon)), \quad (1.39)$$

où $\theta = \{\sigma_0^2, \rho_0^{\omega_0}, \rho_0^\zeta, G, L, \rho_\epsilon^{\omega_0}, \rho_\epsilon^\zeta\}$ est le vecteur des $3 + 2d = 7$ hyper-paramètres. La moyenne m du processus gaussien est intégrée selon une loi uniforme impropre sur \mathbb{R} .

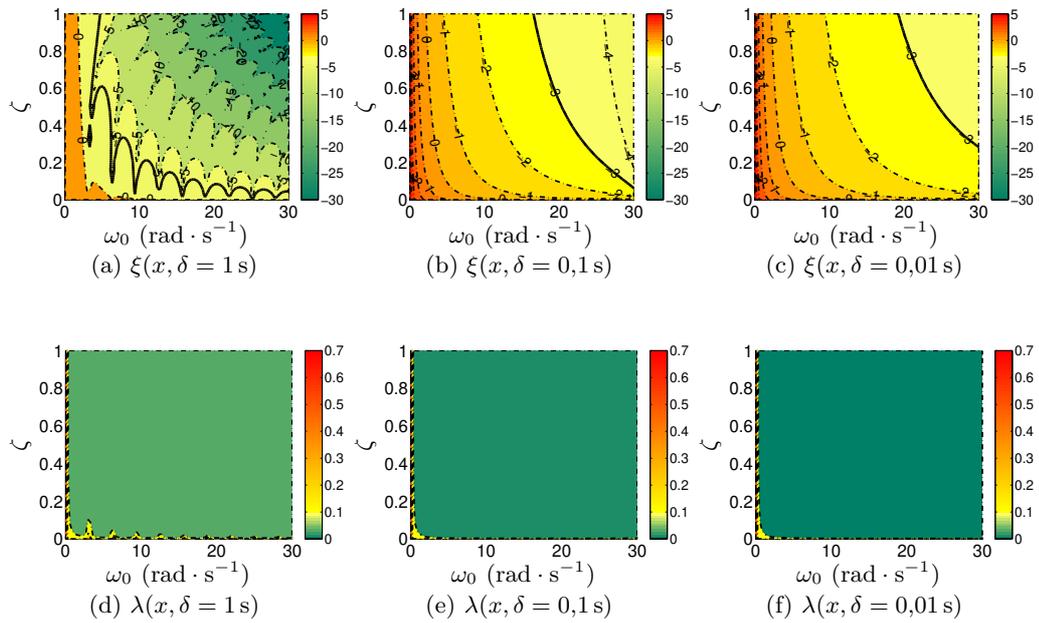


FIGURE 24 – Les fonctions de moyennes et de variances empiriques (dix mille simulations) calculées sur une grille cent par cent du simulateur stochastique sur trois niveaux de fidélité. Sur les figures représentant la moyenne, le trait plein est la ligne de niveau $\xi(x, \delta) = -3$.

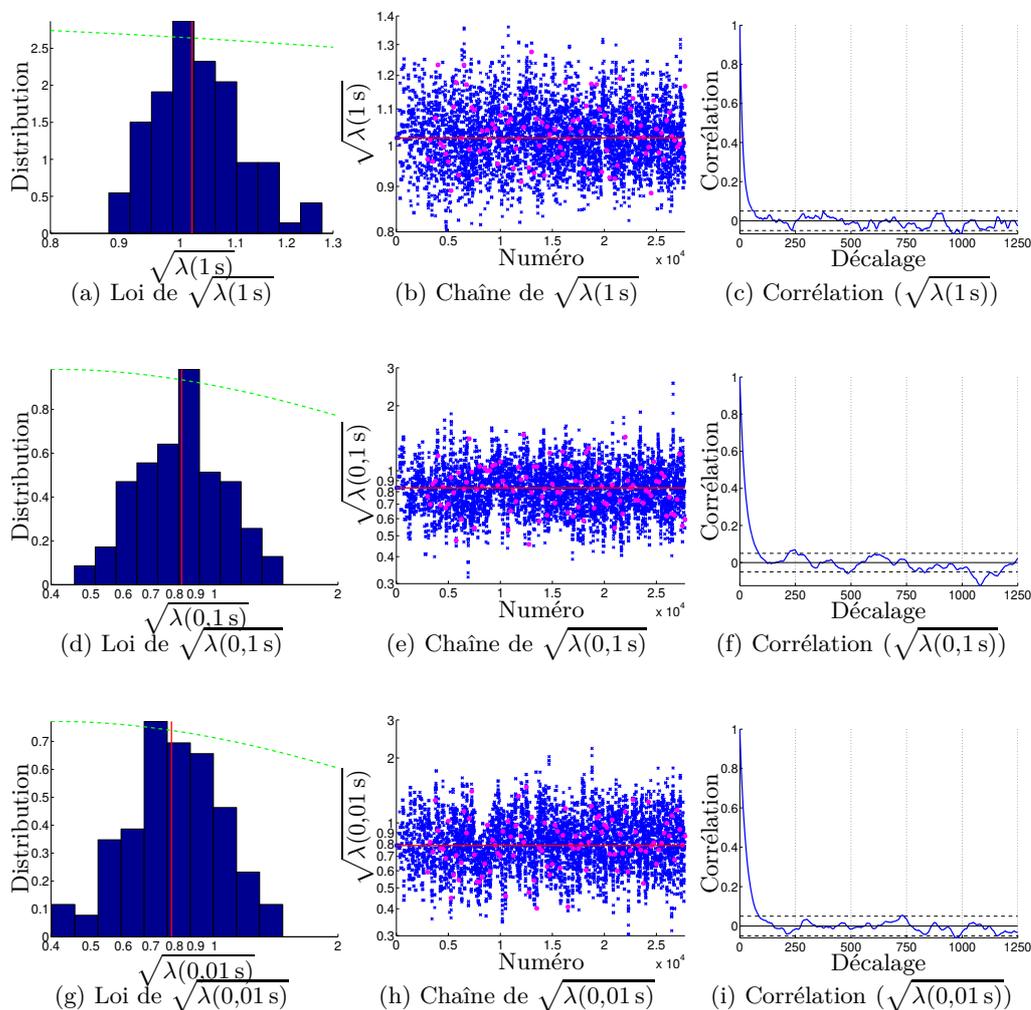


FIGURE 25 – Un échantillonnage de trois des variances du bruit appliqué à l'oscillateur aléatoire amorti. Chaque ligne correspond à une variance. Gauche : histogramme des échantillons des variances de bruit distribuées selon leur loi a posteriori. Le trait pointillé est la loi a priori normalisée et le trait rouge vertical est le MAP. Milieu : la chaîne de Markov issue de l'échantillonnage. Le trait horizontal rouge est le MAP et les points mauves correspondent au sous-échantillonnage de la chaîne. Droite : auto-corrélation de la chaîne. Les traits pointillés horizontaux indiquent le décalage pris entre chaque sous-échantillon de la chaîne Markov.

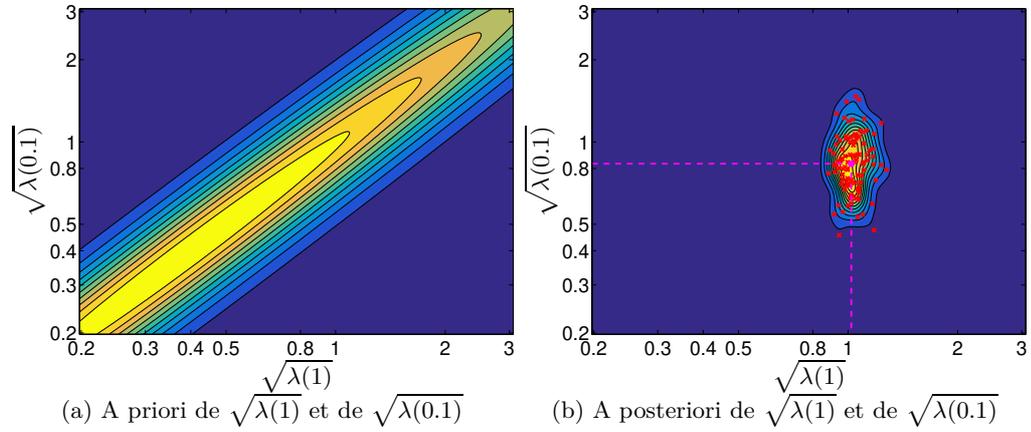


FIGURE 26 – Loi a priori et a posteriori des variances du bruit de l’oscillateur aléatoire. La loi a posteriori est estimée avec un estimateur à noyau, à partir des échantillons simulés selon leur loi a posteriori (les croix sur la figure de droite). Le point souligné par les traits pointillés (figure b) est le maximum a posteriori.

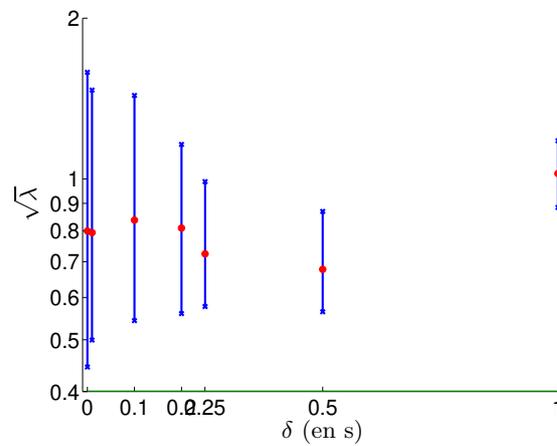


FIGURE 27 – Estimations des valeurs des variances des effets stochastiques en fonction du niveau. Les points rouges sont les médianes a posteriori, les lignes bleues sont les intervalles de crédibilité à 95 % et le trait vert est la valeur l_0 utilisée comme moyenne a priori.

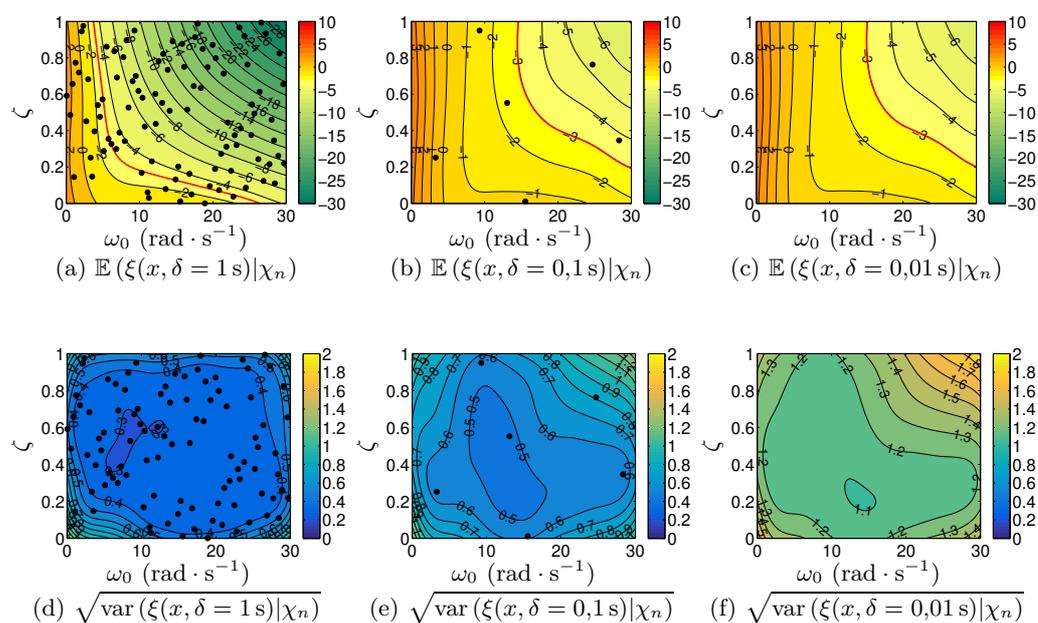


FIGURE 28 – Prédiction aux niveaux $\delta = 1\text{ s}$, $0,1\text{ s}$ et $0,01\text{ s}$ de l'oscillateur aléatoire. En haut les moyennes a posteriori, en bas les écarts-types a posteriori. Les points sur les figures de gauche sont des observations, le trait en évidence est la ligne de niveau à -3 .

Pour appliquer une approche complètement bayésienne (voir annexe A.3), une loi a priori est ajoutée sur les hyper-paramètres. Ces lois a priori sur les hyper-paramètres sont indépendantes, et sont des lois log-normales dont les expressions sont explicitées annexe D. Les amplitudes approximatives des entrées et des sorties sont fixées à $r^{\text{out}} = 40$, $b^{\text{in}\omega_0} = 30$ et $b^{\text{in}\zeta} = 1$.

Pour la variance du bruit λ , le modèle a priori est le processus log-gaussien décrit dans la section 1.3. Ce modèle possède trois hyper-paramètres, la moyenne $l_0 = \log((r^{\text{out}}/100)^2)$, la variance des effets de pépite $\sigma_\lambda^2 = \log(100)^2/100$ et le coefficient $a = 99$ (voir annexe D). Ces hyper-paramètres sont supposés connus. Les hyper-paramètres de la fonction de covariance, ainsi que les variances du bruit sont intégrés selon leur loi a posteriori. Pour cela, nous échantillons leur loi avec l'algorithme de Métropolis adaptatif de Haario et al. [2001] avant d'approcher les moyennes et les variances a posteriori des réponses du simulateur par une somme sur les échantillons.

La figure 25 présente des lois a posteriori de quelques variances du bruit. La figure 26 présente les lois a priori et a posteriori du couple $\lambda(1), \lambda(0.1)$, le niveau 0.1 étant le niveau observé de plus haute-fidélité. Sur les deux figures, la loi a posteriori est beaucoup moins étalée que la loi a priori : les observations ont apportés de l'information. La figure 27 présente les estimations de la variance du bruit sur tous les niveaux observés, ainsi que sur des niveaux de plus haute-fidélité. Tous les intervalles de crédibilité sont réduits comparativement aux intervalles de crédibilité a priori, mais surtout que ces intervalles de crédibilité sont plus petits sur les niveaux où les observations sont plus nombreuses. Sur les niveaux où il n'y a pas d'observation, les intervalles de crédibilité sont plus larges. Enfin, la figure 28 présente les prédictions de ce modèle aux niveaux 1 s, 0,1 s et 0,01 s, ainsi que les écart-types a posteriori lié à la prédiction. Cette figure est à comparer aux valeurs de référence (figure 24).

1.4 Conclusion

Dans ce chapitre, nous avons vu comment modéliser un simulateur dans un cadre multi-fidélité, avec une approche bayésienne fondée sur les processus gaussiens. Nous avons notamment décrit les deux principaux modèles de processus gaussiens pour les simulateurs déterministes : le modèle auto-régressif correctif, proposé par Kennedy et O'Hagan [2000] ; et le modèle non-stationnaire déconvolutif, développé par Picheny et Ginsbourger [2013] et Tuo et al. [2014].

Les deux modèles ont leurs qualités et leurs défauts. Le modèle auto-régressif doit être privilégié lorsque le nombre de niveaux est faible, et que les niveaux de fidélité

n'ont pas de relation évidente entre eux (cas de deux simulateurs différents). Le modèle non-stationnaire est préférable lorsque le nombre de niveaux devient élevé, et que le simulateur peut se décomposer entre un modèle idéal et une erreur numérique qui décroît avec la fidélité (cas de simulateur par différences finies).

Les modèles de la littérature sont conçus pour des simulateurs déterministes. Dans le cas de simulateurs stochastiques, la sortie en un point est supposée suivre une certaine loi dont les paramètres dépendent des entrées. Des lois a priori sont ajoutées sur les fonctions de ces paramètres. Dans le cas d'une régression par processus gaussien, la sortie est supposée suivre une loi normale, ce qui permet de simplifier les calculs. La moyenne de cette loi normale est modélisée par l'un des processus gaussien cités juste avant.

Pour la fonction de variance du simulateur, une approche classique consiste à utiliser un modèle de processus log-gaussien sur la variance. Comme un processus log-gaussien quelconque complexifie considérablement les inférences bayésiennes et la prise en compte de l'incertitude de la variance du bruit sur l'incertitude finale, nous proposons un modèle log-gaussien simplifié sur la variance. Ce modèle a pour principales hypothèses que la variance du bruit ne dépend que du niveau, et que les valeurs de cette variance sont relativement proches entre elles. Ce modèle permet notamment de pouvoir estimer les variances du bruit sur des niveaux peu ou pas observés, et donc de pouvoir faire des prédictions sur les sorties en haute-fidélité du comportement du simulateur multi-fidèle stochastique.

Tout cela permet de pouvoir estimer les densités prédictives du simulateur et de n'importe quelle quantité associée. Cependant, jusqu'ici, les observations $\chi_n = (x_i, \delta_i; z_i)_{1 \leq i \leq n}$ étaient supposées fournies à l'avance. Dans le prochain chapitre, nous verrons comment choisir les points d'observations afin de réduire l'incertitude finale d'une quantité d'intérêt.

Chapitre 2

Planification d'expériences en multi-fidélité

Sommaire

2.1	Introduction	63
2.2	Plans d'expériences non-séquentiels	65
2.2.1	Plans d'expériences sur un niveau	65
2.2.2	Plans non-séquentiels multi-fidèles de la littérature	66
2.2.3	Plans non-séquentiels emboîtés maximin	69
2.3	Plan séquentiel par réduction progressive de l'incertitude	71
2.3.1	Principe des stratégies par réduction progressive de l'incertitude	71
2.3.2	Critères pour estimer des probabilités de dépassement de seuil	73
2.3.3	Un exemple de stratégie SUR sans multi-fidélité	83
2.3.4	Intégration selon les hyper-paramètres	84
2.4	Planification séquentielle en multi-fidélité	89
2.4.1	Revue de la littérature	89
2.4.2	Taux maximal de réduction de l'incertitude	95
2.4.3	Illustrations	98
2.4.4	Extensions pour la parallélisation	114
2.5	Conclusion	120

2.1 Introduction

Dans le chapitre 1, nous sommes partis du principe que les observations $\chi_n = (x_i, \delta_i; z_i)_{1 \leq i \leq n}$ étaient fournies à l'avance, et qu'il fallait seulement les utiliser pour

pouvoir estimer les Quantités d'Intérêt (QoI)¹. Cependant, dans de nombreuses situations, il est possible de choisir les points à observer. Le choix est important, d'une part, car le Plan d'Expériences (DoE)² influe fortement sur les estimations, et d'autre part, car le nombre d'observations est limité par le coût calculatoire du simulateur.

En général, l'utilisation d'un simulateur se fait afin d'atteindre un objectif précis. Cet objectif peut être l'analyse complète d'un phénomène (approximation d'une fonction coûteuse), la configuration optimale d'un système (recherche d'un minimum/maximum) ou, par exemple en sécurité incendie, l'estimation de la probabilité de défaillance d'un système dans un scénario donné. En fonction de l'objectif ciblé, les points observés ne seront pas placés au même endroit, mais seront optimisés en vue d'atteindre le plus rapidement possible l'objectif.

Dans les situations où toutes les observations ont le même coût calculatoire, la notion de rapidité se traduit en terme de nombre d'observations. Dans le cas de la multi-fidélité, comme les observations n'ont pas le même coût, la notion de rapidité se traduit par le coût total des observations (les temps d'analyse et de choix de l'observation étant négligés). L'idée derrière la multi-fidélité est qu'une observation haute-fidélité plus coûteuse en temps de calcul apporte (ou du moins est censée apporter) plus d'information qu'une simulation basse-fidélité, moins chère. Au problème du choix de l'entrée x qui permet d'atteindre au mieux l'objectif visé, s'ajoute donc le problème du choix du niveau de fidélité δ qui optimise le compromis entre temps de calcul et gain d'information.

Deux types de DoE peuvent être distingués : les DoE non-séquentiels et les DoE séquentiels. Les DoE séquentiels sont des méthodes de choix des points d'observation utilisant les observations déjà disponibles. A l'inverse, les plans non-séquentiels ne pré-supposent pas l'existence de résultats du simulateur. Une méthode non-séquentielle permet d'obtenir un DoE initial, puis une stratégie de planification séquentielle permet de compléter successivement les observations.

Un façon courante de construire un DoE séquentiel est de choisir les nouvelles observations en optimisant un critère. Plus exactement, l'idée est de définir un critère d'échantillonnage $J_n : \mathbb{X} \rightarrow \mathbb{R}$, dépendant des n observations déjà effectuées. Puis, la $(n + 1)$ -ième observation est choisie par minimisation du critère sur l'espace des entrées :

$$x_{n+1} = \arg \min_{x \in \mathbb{X}} \{J_n(x)\}. \quad (2.1)$$

Le comportement de la planification dépend alors du critère d'échantillonnage choisi.

1. Quantities of Interest
2. Design of Experiments

Ce critère d'échantillonnage doit donc quantifier l'intérêt d'observer le point x pour atteindre l'objectif. Dans le cas de la multi-fidélité, il doit également prendre en compte le coût d'observation du point.

Ce chapitre est organisé de la manière suivante. La section 2.2 présente les méthodes permettant d'engendrer des DoE non-séquentiels en multi-fidélité. La section 2.3 présente des stratégies de Réduction Progressive de l'Incertitude (SUR), une méthodologie permettant de créer des critères d'échantillonnage pour planification séquentielle en l'absence de multi-fidélité. La section présente également de nouveaux critères d'échantillonnage dédiés à l'estimation de probabilité de dépassement de seuil sur un simulateur stochastique. La section 2.4 focalise sur les plans séquentiels en multi-fidélité. Après une présentation des méthodes de la littérature, nous proposons un principe pour créer des critères d'échantillonnage dans ce contexte, et nous illustrons ce principe sur deux exemples.

2.2 Plans d'expériences non-séquentiels

2.2.1 Plans d'expériences sur un niveau

Dans le cadre d'une régression par processus gaussien, l'objectif des DoE initiaux est de fournir des observations dans l'espace des entrées pour permettre : d'une part d'estimer les hyper-paramètres θ du processus, constitués des hyper-paramètres de la fonction de covariance et des variances des effets stochastiques en sortie; et d'autre part, d'obtenir une première prédiction du simulateur étudié en ayant des observations dispersées dans l'espace des entrées.

Dans le cadre standard avec un seul niveau de fidélité, il existe une vaste littérature sur la construction de DoE initiaux, dont un résumé peut être trouvé dans Pronzato et Müller [2012]. Nous cherchons à obtenir des DoE initiaux qui représentent bien l'ensemble du processus que nous étudions. Les DoE initiaux à un niveau de fidélité que nous recherchons doivent posséder deux propriétés principales. D'abord, nous souhaitons que les points soient dispersés dans l'espace des entrées, et ensuite, que leurs projections selon les sous-espaces de dimension 1 soient régulièrement espacées.

Pour mesurer la dispersion des points d'observations, plusieurs critères existent. Par exemple, Morris et Mitchell [1995] proposent des critères ϕ_p pour mesurer le remplissage de l'espace, Fang et al. [2000] utilisent la discrèpance et Currin et al. [1991] se fondent sur l'entropie. Dans notre cas, nous utilisons le critère maximin [Johnson et al., 1990], qui est simple à calculer, mesurant la distance entre les deux points les plus proches du

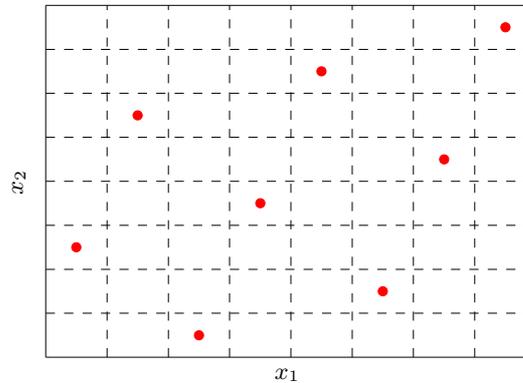


FIGURE 29 – Un hypercube latin maximin en dimension deux avec huit points d’observations.

DoE. Ce critère, que nous souhaitons maximiser, s’écrit $\min_{1 \leq i < j \leq n} \{\|x_i - x_j\|\}$, où les x_i sont à valeurs dans $[0; 1]$.

Le type de plan d’expérience assurant que les projections selon les sous-espaces de dimension 1 soient régulièrement réparties se nomme les Hypercubes Latins (LHS)³. Les LHS sont un type de DoE initialement proposés par McKay et al. [1979] pour améliorer les méthodes de Monte-Carlo. Cette propriété permet notamment d’assurer une bonne représentation de chacune des variables d’entrées. Il est possible de considérer des plans avec des meilleurs propriétés de projection, comme les LHS orthogonaux, développés par Tang [1993].

Les types de plans simple niveau que nous considérons sont des DoE qui sont à la fois LHS et qui maximisent le critère maximin. La figure 29 présente un exemple d’un tel plan en dimension $d = 2$ avec $n = 8$ points, où les deux variables d’entrées x_1 et x_2 suivent des lois uniformes.

2.2.2 Plans non-séquentiels multi-fidèles de la littérature

Cependant, toutes ces approches ne peuvent pas s’appliquer directement en multi-fidélité, car les niveaux de fidélité ne sont pas tous équivalents. En effet, en multi-fidélité, les niveaux de haute-fidélités sont supposés plus coûteux que les niveaux de basse-fidélité. Dans un souci d’économie de temps de calcul, les niveaux peu coûteux sont observés

3. Latin Hypercube Sampling

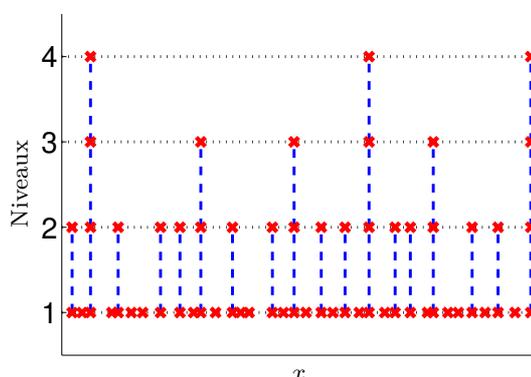


FIGURE 30 – Un exemple de plan d'expériences emboîté.

préférentiellement par rapport aux niveaux chers.

En multi-fidélité, une propriété souvent souhaitée des DoE est l'emboîtement. Un DoE multi-fidèle est emboîté lorsque tout point x observé à un niveau de fidélité δ est également observé aux niveaux de fidélité inférieurs présents dans le DoE. Un exemple de DoE emboîté à quatre niveaux est présenté figure 30. Cette propriété permet de simplifier certaines analyses, par exemple les estimations des hyper-paramètres dans le cas du modèle auto-régressif (voir section 1.2.1). La génération de DoE non-séquentiels emboîtés présente donc un intérêt particulier en multi-fidélité.

Nous supposons dans la suite de la section que nous souhaitons engendrer un DoE emboîté sur S niveaux, indexés par $s \in \llbracket 1; S \rrbracket$. Le niveau S correspond au niveau de plus haute fidélité, et le niveau 1 au niveau le plus bas. L'ensemble des points d'observations du DoE situés au niveau s est noté X_s et comprend n_s points. Le nombre total de points dans le DoE multi-fidèle est $\sum_{s=1}^S n_s$. La suite n_s est évidemment décroissante, puisque le DoE est emboîté.

Le Gratiet et Cannamela [2015] proposent un algorithme pour engendrer des DoE emboîtés. Celui-ci fonctionne par génération et déplacements successifs à chaque niveau. L'algorithme commence par engendrer un DoE au niveau de plus haute-fidélité X_S . Le Gratiet et Cannamela [2015] utilisent un générateur de LHS maximisant le critère maximin. Ensuite, récursivement sur le niveau s , ils engendrent un DoE Y_s ayant n_s points et calculent les distances entre les points de Y_s et de X_{s+1} . Les n_{s+1} points de Y_s les plus proches des points de X_{s+1} sont retirés et remplacés ceux-ci. Le DoE X_s est alors constitué des $n_s - n_{s+1}$ points restants et des points de X_{s+1} , assurant ainsi l'em-

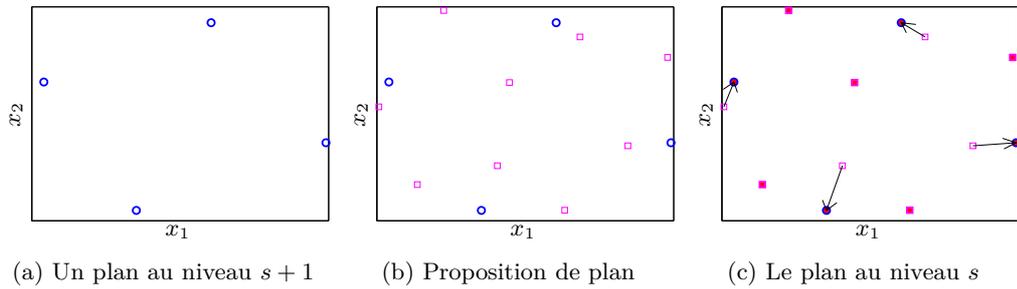


FIGURE 31 – Les trois étapes de la génération du plan d’expériences au niveau s à partir d’un plan au niveau $s + 1$. (a) Le plan au niveau $s + 1$, maximin LHS avec quatre points. (b) Une proposition de plan pour le niveau s , un maximin LHS avec neuf points. (c) Retrait des observations les plus proches du plan $s + 1$.

boîtement. La figure 31 illustre ce processus. L’intérêt de cette méthode est sa simplicité, puisqu’elle ne nécessite que d’avoir une méthode de génération de DoE non-séquentiels à un niveau de fidélité. Cette méthode garantit juste l’emboîtement. Elle n’assure pas, par exemple, que les plans d’expériences obtenus à chaque niveau sont des LHS.

Qian [2009] propose une méthode pour engendrer des Hypercubes Latins Emboîtés (NLHS)⁴. Un NLHS est un DoE emboîté dans lequel chaque DoE d’un niveau de fidélité forme un LHS. Cette méthode de génération suppose toutefois que le nombre de points au niveau $s + 1$ divise le nombre de points au niveau s . Pour engendrer ce DoE multi-fidèle, l’algorithme initialise avec un LHS au niveau S . Puis entre les niveaux $s + 1$ et s , comme le DoE du niveau $s + 1$, X_{s+1} , est un LHS et que n_{s+1} divise n_s , le plan X_{s+1} peut être vu comme le début d’un LHS à n_s points. En retirant les zones interdites selon chacune des dimensions, l’espace autorisé restant forme un nouvel hypercube. Cet hypercube peut être complété avec un LHS de taille $n_s - n_{s+1}$. La concaténation des deux LHS donne un LHS global avec n_s points. La figure 32 illustre ce principe. Cette approche permet notamment de conserver les bonnes propriétés d’échantillonnage aléatoire des LHS en multi-fidélité, puisqu’elle est conçue pour pouvoir appliquer un Monte-Carlo sur plusieurs niveaux de fidélité. Kong et al. [2016], Xu et al. [2017], Guo et al. [2017] et Garg et Stogner [2017] proposent des méthodes plus générales pour engendrer des NLHS aléatoires pour des nombres quelconques de points à chaque niveau. De plus, He et Qian [2011] et Yang et al. [2014] étendent la notion de LHS orthogonaux aux DoE emboîtés. Cependant, leurs méthodes sont trop restrictives dans les relations sur le nombre de

4. Nested Latin Hypercube Sampling

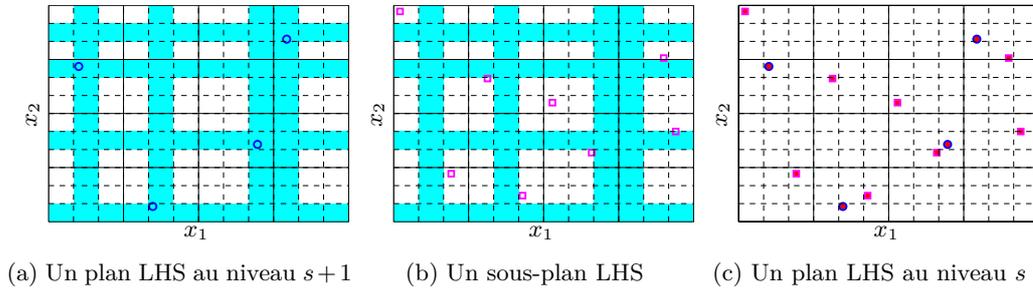


FIGURE 32 – Les trois étapes de la génération du plan d’expériences LHS au niveau s à partir d’un plan LHS au niveau $s + 1$. (a) Le plan LHS au niveau $s + 1$ avec quatre points. Les lignes et colonnes colorées sont interdites lors de l’ajout des nouveaux points au niveau de plus basse fidélité. (b) Les carrés restants sont complétés avec un sous-LHS. (c) Un plan LHS au niveau s après complétion, constitué de 4×3 points.

points, et difficilement extensibles au-delà de deux niveaux de fidélité.

2.2.3 Plans non-séquentiels emboîtés maximin

Dans notre contexte, nous souhaitons utiliser la méthode de Qian [2009]. Cependant, nous cherchons aussi à répartir les points d’observations dans l’espace des entrées. Pour mesurer l’espacement entre les points en multi-fidélité, Rennen [2009] et Rennen et al. [2010] proposent le critère suivant, dédié aux DoE multi-fidèles :

$$\min_{1 \leq s \leq S} \left\{ \frac{J(X_s)}{\sqrt[d]{n_s - 1}} \right\} \quad (2.2)$$

où $J(X_s)$ est le critère maximin au niveau s et d est le nombre de facteur(s) autre que le paramètre de fidélité. Le dénominateur $\sqrt[d]{n_s - 1}$ est la distance caractéristique entre deux points. Sa présence permet de comparer des DoE ayant des nombres d’observations différents. Nous cherchons alors à engendrer des NLHS en optimisant le critère (2.2).

La méthode de Qian engendre des NLHS aléatoirement. Pour obtenir une amélioration du critère de remplissage, nous modifions leur façon de générer des NLHS. Dans cette thèse, nous avons comparé deux possibilités. La première méthode est de tirer un nombre N de NLHS et de choisir celui qui maximise le critère maximin multi-fidèle (2.2). La seconde méthode consiste à modifier le tirage des plans LHS, que ce soit le plan LHS du niveau de plus haute fidélité, ou les compléments. Sur le niveau de plus haute fidélité, N plans LHS sont tirés aléatoirement, et celui avec le meilleur critère maximin est

sélectionné. Ensuite, à chaque complétion du LHS du niveau $s + 1$, N compléments sont tirés aléatoirement. Le complément choisi est celui qui maximise le critère maximin du DoE X_s . En terme de coût calculatoire, ces deux méthodes sont équivalentes. Nous les avons comparées pour savoir si l'une d'elles se distinguait.

Nous expérimentons ces deux méthodes sur $S = 4$ niveaux, où les facteurs x sont de dimension $d = 8$. Le nombre de points par niveau est $n_1 = 270$, $n_2 = 90$, $n_3 = 30$ et $n_4 = 10$ (voir section 3.3). Le nombre de tirages aléatoires par génération est fixé à $N = 1000$, et $p = 1000$ DoE sont engendrés pour chacune des deux méthodes. Nous engendrons également p NLHS complètement aléatoirement pour avoir une référence. Pour comparer ces NLHS, le critère maximin est calculé à chaque niveau.

La figure 33 présente les valeurs du critère maximin obtenues à chaque niveau pour chaque expérience. La position des marques indique que la seconde méthode, qui choisit les meilleurs plans niveau par niveau, construit des plans d'expériences ayant de meilleures valeurs du critère maximin sur tous les niveaux. De plus, les plus hauts niveaux de fidélité (niveau 3 et 4), la première méthode, choisissant le NLHS à la fin ne fait pas mieux qu'un tirage aléatoire. Cela pourrait suggérer que le critère (2.2) n'est pas idéal, et qu'il ne tend à favoriser la dispersion des observations qu'aux niveaux de fidélité les plus bas.

Bien que non présenté dans cette thèse, la même expérience a été refaite en changeant le nombre de points et de niveaux dans DoE, ainsi qu'en modifiant le critère (2.2) (remplacement du critère maximin par un critère ϕ_p de Morris et Mitchell [1995], minimum sur les niveaux remplacé par une somme sur les niveaux, changement dans la pondération). A chaque fois, le même résultat est obtenu : la seconde méthode fournit de meilleurs DoE multi-fidèles que la première, aussi bien selon le critère multi-fidèle agrégé, que selon les critères appliqués à chaque niveau.

En résumé, pour engendrer un plan initial multi-fidèle, nous recommandons d'utiliser la méthode de Qian [2009] qui engendre des NLHS, en appliquant la seconde méthode de choix des plans maximin à chaque niveau. Cette méthode est appliquée dans tout le manuscrit pour engendrer des plans d'expériences initiaux multi-fidèles. Notons la méthode utilise une optimisation grossière des plans non-séquentiels. Utiliser un algorithme d'optimisation plus efficace sur ces plans d'expériences pourraient être une façon simple d'améliorer la répartition des points dans l'espace. Dans cette thèse, la méthode de Qian [2009] est appliquée avec le critère maximin, mais d'autres critères, comme la discrédance, pourraient être envisagés à la place.

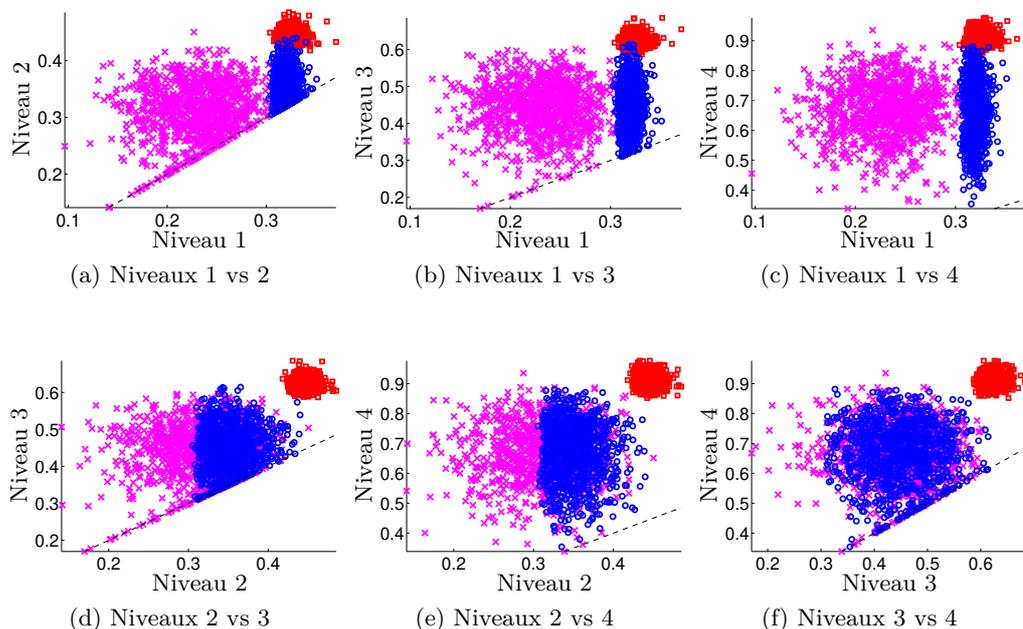


FIGURE 33 – Valeurs des critères maximin à chaque niveau, en fonction des méthodes. Une figure correspond à une paire de niveaux. Les croix roses sont des NLHS complètement aléatoires, les cercles bleus sont des NLHS engendrés avec la première méthode (choix à la fin de la génération), et les carrés rouges sont des NLHS engendrés avec la seconde méthode, où les plans sont choisis niveau par niveau.

2.3 Plan séquentiel par réduction progressive de l'incertitude

Dans cette section, par simplicité d'écriture, le vecteur des entrées est noté x . Pour passer au cas multi-fidèle, il suffit de remplacer x par (x, δ) dans toutes les équations, le discours restant le même.

2.3.1 Principe des stratégies par réduction progressive de l'incertitude

La méthodologie de Réduction Progressive de l'Incertainitude (SUR)⁵ est une façon de construire des critères d'échantillonnage dédiés à la planification séquentielle d'expériences. Le nom a été proposé par Fleuret et Geman [1999], bien que d'autres articles, comme par exemple MacKay [1992] ou Cohn et al. [1996] développent des méthodes de

5. Stepwise Uncertainty Reduction

planifications similaires. La méthodologie a été appliquée par Villemonteix et al. [2008] et Vazquez et Bect [2009] sur des modèles bayésiens fondés sur des processus gaussiens. L'idée est de définir une mesure d'incertitude H_n aussi appelée risque, dépendant des n observations χ_n . Par exemple, dans le cas où l'objectif est d'estimer une certaine QoI salaire, une mesure d'incertitude peut-être la variance a posteriori de cette QoI. Le point d'observation à choisir selon le principe des stratégies SUR est celui qui minimise l'incertitude future espérée

$$x_{n+1} = \arg \min_{x \in \mathbb{X}} \{\mathbb{E}_n [H_{n+1} | X_{n+1} = x]\}, \quad (2.3)$$

l'indice n indiquant que l'espérance est prise conditionnellement aux n premières observations. La fonction

$$J_n(x) = \mathbb{E}_n [H_{n+1} | X_{n+1} = x] \quad (2.4)$$

forme le critère d'échantillonnage, et correspond à l'incertitude attendue avec $n+1$ points, si le $(n+1)$ ^{ième} point observé est le point x . Nous notons également $G_n(x) = H_n - J_n(x)$ la réduction d'incertitude. Remarquez que minimiser le critère d'échantillonnage J_n est équivalent à maximiser le gain G_n . L'espérance définie dans le critère J_n est une intégrale sur l'ensemble des valeurs possibles de la sortie du simulateur en x . Dans le cas où il faudrait choisir q observations, le critère d'échantillonnage SUR s'écrit

$$J_n(x_{n+1}, \dots, x_{n+q}) = \mathbb{E}_n [H_{n+q} | X_{n+1} = x_{n+1}, \dots, X_{n+q} = x_{n+q}]. \quad (2.5)$$

Dans la littérature, il existe un certain nombre de critères SUR appliqués à des processus gaussiens. Li [2012] propose des critères SUR pour estimer des probabilités de dépassement de seuil sur un simulateur déterministe, puis Chevalier et al. [2014] développent un algorithme de calcul rapide de ces critères. Azzimonti et al. [2016b] proposent un critère SUR pour l'estimation conservatrice de sous-ensemble de l'espace des entrées \mathbb{X} . De même, Labopin-Richard et Picheny [2016] proposent un critère pour l'estimation de quantile des lois des sorties. Le critère d'Amélioration Espérée (EI)⁶ proposé par Jones et al. [1998], utilisé pour l'optimisation de fonction déterministe, peut également être vu comme un critère SUR. Toujours dans un objectif d'optimisation, l'Approche Informatrice de l'Optimisation Globale (IAGO)⁷ est un critère proposé par Villemonteix et al. [2008] dont la mesure d'incertitude est fondé sur l'entropie de l'argument du minimum,

6. Expected Improvement

7. Informational Approach to Global Optimization

le Gradient de Connaissance Approximatif (AKG)⁸ de Scott et al. [2011] est un critère SUR dédié à l'optimisation d'un simulateur stochastique et le critère d'Augmentation Espérée de l'Hyper-Volume (EHVI)⁹ proposé par Emmerich et al. [2006] est conçu pour l'optimisation multi-objectif. Enfin, Bect et al. [2017] proposent des conditions suffisantes que doit suivre la mesure d'incertitude pour garantir la convergence de la stratégie SUR associée.

2.3.2 Critères pour estimer des probabilités de dépassement de seuil

Dans cette section, nous présentons de nouveaux critères SUR pour l'estimation de probabilité de dépassement de seuil dans un cadre stochastique. Ces critères sont motivés par l'application incendie présentée au chapitre 3. Les critères développés sont calculables pour un nombre q quelconque de nouvelles observations.

La réponse Z du simulateur en un point x est supposée suivre une loi normale de moyenne $\xi(x)$ et de variance $\lambda(x)$,

$$Z|x \sim \mathbf{N}(\xi(x), \lambda(x)). \quad (2.6)$$

Le modèle bayésien de ξ est un modèle de processus gaussien. Ainsi, après n observations du simulateur, le processus modélisant ξ est un processus gaussien de fonction moyenne m_n et de fonction de covariance k_n . Les hyper-paramètres de la fonction de covariance ainsi que la fonction de variance λ sont supposés connus et fixés. Soit z^{crit} un seuil critique. Les QoI considérées dans les différents critères sont liées à la fonction de probabilité de dépassement de seuil

$$p(x) = \mathbb{P}(Z \geq z^{\text{crit}}|x) = \Phi \left(\frac{\xi(x) - z^{\text{crit}}}{\sqrt{\lambda(x)}} \right), \quad (2.7)$$

où Φ est la fonction de répartition de la loi normale centrée réduite. Nous notons $\Phi_2(\cdot, \cdot; r)$ la fonction de répartition de la loi normale centrée réduite bivariée de corrélation r .

8. Approximate Knowledge Gradient

9. Expected Hyper-Volume Improvement

Critère SUR 1 : estimation de la fonction de probabilité en norme L2

Pour construire un critère SUR, nous avons besoin d'une mesure d'incertitude. Considérons par exemple l'espérance de la perte quadratique :

$$H_n = \mathbb{E}_n \left[\|p - \hat{p}_n\|_{\mathbb{L}_X^2(\mu)}^2 \right] = \int_{\mathbb{X}} \text{var}_n \left[\Phi \left(\frac{\xi(x) - z^{\text{crit}}}{\sqrt{\lambda(x)}} \right) \right] \mu(dx). \quad (2.8)$$

où μ est une mesure positive bornée (par exemple une mesure de probabilité ou la mesure de Lebesgue sur un domaine borné), et $\hat{p}_n(x)$ est l'espérance a posteriori de $p(x)$ utilisée comme estimateur,

$$\hat{p}_n(x) = \mathbb{E}_n \left[\Phi \left(\frac{\xi(x) - z^{\text{crit}}}{\sqrt{\lambda(x)}} \right) \right] = \Phi(u_n(x)) \quad (2.9)$$

où $u_n(x) = \frac{m_n(x) - z^{\text{crit}}}{\sqrt{k_n(x,x) + \lambda(x)}}$ est la distance normalisée entre la moyenne et le seuil critique. L'espérance de la probabilité de dépassement de seuil (2.9) se calcule avec (B.4). Or, $\text{var}_n[p(x)] = \mathbb{E}_n[p(x)^2] - \mathbb{E}_n[p(x)]^2$. Le second terme est le carré de l'estimation. Quand au premier terme, il se calcule en utilisant (B.5)

$$\mathbb{E}_n \left[\Phi \left(\frac{\xi(x) - z^{\text{crit}}}{\sqrt{\lambda(x)}} \right)^2 \right] = \Phi_2(u_n(x), u_n(x); r_n(x, x)), \quad (2.10)$$

avec $r_n(x, x) = k_n(x, x)/V_n(x)$, et $V_n(x) = k_n(x, x) + \lambda(x)$ la variance totale de la sortie Z en x . La corrélation $r_n(x, x)$ est le rapport entre la variance de la fonction ξ en x et celle de la sortie Z en x . La mesure d'incertitude se ré-écrit donc

$$H_n = \int_{\mathbb{X}} \left[\Phi_2(u_n(x), u_n(x); r_n(x, x)) - \Phi(u_n(x))^2 \right] \mu(dx). \quad (2.11)$$

Calculons alors le critère SUR ainsi défini. Soit $\tilde{x} = (\tilde{x}_l)_{1 \leq l \leq q}$, q propositions de points d'observations. Conditionnellement à ces nouvelles observations, le processus $\xi|_{\mathcal{X}_{n+q}}$ reste un processus gaussien, de fonctions de moyenne m_{n+q} et de covariance k_{n+q} . Or, une propriété importante des processus gaussiens est qu'il est possible de calculer la fonction de covariance k_{n+q} juste en connaissant les points d'observations, sans avoir à connaître les résultats de ces observations, grâce à la formule des mises à jour de la covariance :

$$k_{n+q}(x, x') = k_n(x, x') - k_n(x, \tilde{x})K_n(\tilde{x}, \tilde{x})^{-1}k_n(\tilde{x}, x') = k_n(x, x') - \nu_n(x, x'; \tilde{x}), \quad (2.12)$$

où $K_n(\tilde{x}, \tilde{x}) = (k_n(\tilde{x}_l, \tilde{x}_m) + \lambda(\tilde{x}_l)\mathbb{1}_{l=m})_{1 \leq l, m \leq q}$ est la matrice de covariance des q propo-

sitions conditionnellement aux n premières observations et $k_n(\tilde{x}, x') = (k_n(\tilde{x}_l, x'))_{1 \leq l \leq q}$ est le vecteur de covariance entre \tilde{x} et x' . La fonction $\nu_n(x, x'; \tilde{x})$ est la variation de covariance si les propositions \tilde{x} sont observées. De plus, il est possible d'exprimer la loi de la fonction moyenne. En effet, $m_{n+q}(x) = m_n(x) + k_n(x, \tilde{x})K_n(\tilde{x}, \tilde{x})^{-1}(Z - m_n(\tilde{x}))$, où Z sont les sorties du simulateur aux propositions \tilde{x} . Conséquemment, en utilisant le modèle bayésien des sorties conditionnellement aux observations, il est possible d'exprimer la loi de la fonction moyenne après $n + q$ observations,

$$m_{n+q} | \chi_n, \tilde{x} \sim \text{GP}(m_n, \nu_n(\cdot, \cdot; \tilde{x})). \quad (2.13)$$

Le processus m_{n+q} conditionnellement à χ_n est un processus gaussien, de fonction de moyenne m_n et de fonction de covariance $\nu_n(\cdot, \cdot; \tilde{x})$. Le critère SUR est, selon la définition (2.3),

$$J_n(\tilde{x}) = \mathbb{E}_n[H_{n+q} | \tilde{x}] = \int_{\mathbb{X}} \mathbb{E}_n[\Phi_2(u_{n+q}(x), u_{n+q}(x); r_{n+q}(x, x)) - \Phi(u_{n+q}(x))^2] \mu(dx). \quad (2.14)$$

En utilisant (2.13) et (B.4), le critère se calcule comme

$$J_n(\tilde{x}) = \int_{\mathbb{X}} [\Phi_2(u_n(x), u_n(x); r_n(x, x)) - \Phi_2(u_n(x), u_n(x); \tilde{r}_n(x, x; \tilde{x}))] \mu(dx), \quad (2.15)$$

avec $\tilde{r}_n(x, x'; \tilde{x}) = \nu_n(x, x'; \tilde{x}) / \sqrt{V_n(x)V_n(x')}$. Enfin, il est possible d'exprimer la fonction de gain, définie comme $G_n(\tilde{x}) = H_n - J_n(\tilde{x})$,

$$G_n(\tilde{x}) = \int_{\mathbb{X}} [\Phi_2(u_n(x), u_n(x); \tilde{r}_n(x, x; \tilde{x})) - \Phi(u_n(x))^2] \mu(dx) \quad (2.16)$$

Ce critère est résumé dans l'encadré critère SUR 1.

Ce critère est conçu pour les simulateurs stochastiques, dont les observations suivent une loi normale dont la variance est donnée par la fonction λ . Néanmoins, il est possible d'étendre ce critère au cas déterministe, en faisant tendre λ vers zéro. Dans, ce cas, la QoI devient la fonction de sélection, $p(x) = \mathbb{1}_{\{\xi(x) \geq z^{\text{crit}}\}}$, et la mesure d'incertitude devient $H_n = \int_{\mathbb{X}} \hat{p}_n(x) (1 - \hat{p}_n(x)) dx$. Le critère correspond alors à l'un des critères proposés par Li [2012] et Chevalier et al. [2014].

Critère SUR 2 : estimation de la probabilité intégrée en norme L2

Il est possible d'appliquer la même méthodologie pour estimer une intégrale de la fonction de probabilité. L'espace des entrées est alors séparé en deux parties, $\mathbb{X} = \mathbb{X}_e \times \mathbb{X}_s$.

QoI : la fonction de probabilité

$$p(x) = \mathbb{P} \left(Z \geq z^{\text{crit}} | x \right) = \Phi \left(\frac{\xi(x) - z^{\text{crit}}}{\sqrt{\lambda(x)}} \right),$$

Estimateur : moyenne a posteriori

$$\hat{p}_n(x) = \Phi(u_n(x)),$$

Mesure d'incertitude : intégrale de la variance

$$\begin{aligned} H_n &= \mathbb{E}_n \left[\|p - \hat{p}_n\|_{\mathbb{L}_X^2(\mu)}^2 \right] \\ &= \int_{\mathbb{X}} \left[\Phi_2(u_n(x), u_n(x); r_n(x, x)) - \Phi(u_n(x))^2 \right] \mu(dx) \end{aligned}$$

Critère d'échantillonnage :

$$J_n(\tilde{x}) = \int_{\mathbb{X}} \left[\Phi_2(u_n(x), u_n(x); r_n(x, x)) - \Phi_2(u_n(x), u_n(x); \tilde{r}_n(x, x; \tilde{x})) \right] \mu(dx)$$

Fonction de gain :

$$G_n(\tilde{x}) = \int_{\mathbb{X}} \left[\Phi_2(u_n(x), u_n(x); \tilde{r}_n(x, x; \tilde{x})) - \Phi(u_n(x))^2 \right] \mu(dx)$$

Critère SUR 1 – Estimation de la fonction de probabilité de dépassement de seuil

QoI : la fonction de probabilité intégrée,

$$P(x_s) = \mathbb{P}(Z \geq z^{\text{crit}} | x_s) = \int_{\mathbb{X}_e} \Phi \left(\frac{\xi(x_e, x_s) - z^{\text{crit}}}{\sqrt{\lambda(x_e, x_s)}} \right) f^{\text{env}}(x_e) dx_e,$$

Estimateur : moyenne a posteriori

$$\widehat{P}_n(x_s) = \int_{\mathbb{X}_e} \Phi(u_n(x_e, x_s)) f^{\text{env}}(x_e) dx_e,$$

Mesure d'incertitude : intégrale de la variance,

$$\begin{aligned} H_n &= \mathbb{E}_n \left[\left\| P - \widehat{P}_n \right\|_{\mathbb{L}_{\mathbb{X}_s}^2(\mu)}^2 \right] \\ &= \int_{\mathbb{X}_s} \int_{\mathbb{X}_e^2} [\Phi_2(u_n(x_e, x_s), u_n(x'_e, x_s); r_n((x_e, x_s), (x'_e, x_s)))) \\ &\quad - \Phi(u_n(x_e, x_s)) \Phi(u_n(x'_e, x_s))] f^{\text{env}}(x_e) dx_e f^{\text{env}}(x'_e) dx'_e \mu(dx_s) \end{aligned}$$

Critère d'échantillonnage :

$$\begin{aligned} J_n(\tilde{x}) &= \int_{\mathbb{X}_s} \int_{\mathbb{X}_e^2} [\Phi_2(u_n(x_e, x_s), u_n(x'_e, x_s); r_n((x_e, x_s), (x'_e, x_s)))) \\ &\quad - \Phi_2(u_n(x_e, x_s), u_n(x'_e, x_s); \widetilde{r}_n((x_e, x_s), (x'_e, x_s); \tilde{x}))] \\ &\quad f^{\text{env}}(x_e) dx_e f^{\text{env}}(x'_e) dx'_e \mu(dx_s) \end{aligned}$$

Fonction de gain :

$$\begin{aligned} G_n(\tilde{x}) &= \int_{\mathbb{X}_s \times \mathbb{X}_e^2} [\Phi_2(u_n(x_e, x_s), u_n(x'_e, x_s); \widetilde{r}_n((x_e, x_s), (x'_e, x_s); \tilde{x})) \\ &\quad - \Phi(u_n(x_e, x_s)) \Phi(u_n(x'_e, x_s))] f^{\text{env}}(x_e) dx_e f^{\text{env}}(x'_e) dx'_e \mu(dx_s) \end{aligned}$$

Critère SUR 2 – Estimation de la probabilité de défaillance intégrée selon les variables environnementales

D'un coté, les variables environnementales $x_e \in \mathbb{X}_e$ sont des variables à intégrer dans le calcul de la probabilité. De l'autre, les variables de scénario $x_s \in \mathbb{X}_s$ sont des variables dont nous souhaiterions étudier l'impact. Soit une mesure de probabilité f^{env} définie sur les variables environnementales. La QoI se définit comme

$$P(x_s) = \int_{\mathbb{X}_e} p(x_e, x_s) f^{\text{env}}(x_e) dx_e. \quad (2.17)$$

En choisissant le même type de mesure d'incertitude et appliquant des calculs similaires, il est possible d'obtenir le critère présenté dans l'encadré critère SUR 2.

Ce critère peut être étendu au cas où $\mathbb{X} = \mathbb{X}_e$, il suffit alors d'enlever les intégrales selon \mathbb{X}_s dans les formules. Si en plus, pour revenir au cas déterministe, λ tend vers zéro, le critère obtenu correspond au critère de variance calculé dans Chevalier et al. [2014]. Intuitivement, une planification séquentielle contrôlée par le critère SUR 2 devrait permettre d'estimer la fonction de probabilité intégrée plus rapidement qu'avec le critère SUR 1. Cependant, étant donné que ce critère est plus complexe à calculer en pratique, et que la convergence de la mesure d'incertitude définissant le critère SUR 1 implique la convergence de la mesure d'incertitude définissant le critère SUR 2, nous recommandons plutôt d'utiliser le critère d'échantillonnage du critère SUR 1, de manière similaire aux recommandations de Chevalier et al. [2014].

Critère SUR 3 : estimation de la fonction de probabilité en norme L1

Toujours dans l'objectif d'estimer la fonction $p(x)$, la mesure d'incertitude peut être fondée sur une perte \mathbb{L}^1

$$H_n = \mathbb{E}_n \left[\|p - \hat{p}_n\|_{\mathbb{L}^1_{\mathbb{X}}(\mu)} \right] = \int_{\mathbb{X}} \mathbb{E}_n [|p(x) - \hat{p}_n(x)|] \mu(dx). \quad (2.18)$$

où $\hat{p}_n(x)$ est la médiane a posteriori de $p(x)$ qui peut être aussi utilisée comme estimateur

$$\hat{p}_n(x) = \Phi \left(\frac{m_n(x) - z^{\text{crit}}}{\sqrt{\lambda(x)}} \right). \quad (2.19)$$

Notez que la médiane de $p(x)$ n'est pas égale à l'espérance de $p(x)$ (2.9). La mesure d'incertitude devient l'intégrale de la déviation absolue

$$H_n = \int_{\mathbb{X}} \mathbb{E}_n \left[\left| \Phi \left(\frac{\xi(x) - z^{\text{crit}}}{\sqrt{\lambda(x)}} \right) - \Phi \left(\frac{m_n(x) - z^{\text{crit}}}{\sqrt{\lambda(x)}} \right) \right| \right] \mu(dx). \quad (2.20)$$

QoI : la fonction de probabilité

$$p(x) = \mathbb{P}(Z \geq z^{\text{crit}} | x) = \Phi \left(\frac{\xi(x) - z^{\text{crit}}}{\sqrt{\lambda(x)}} \right),$$

Estimateur : médiane a posteriori

$$\hat{p}_n(x) = \Phi \left(\frac{m_n(x) - z^{\text{crit}}}{\sqrt{\lambda(x)}} \right),$$

Mesure d'incertitude : intégrale de la déviation absolue

$$\begin{aligned} H_n &= \mathbb{E}_n \left[\|p - \hat{p}_n\|_{\mathbb{L}_X^1(\mu)} \right] \\ &= \int_{\mathbb{X}} \left[\Phi(u_n(x)) - 2\Phi_2 \left(u_n(x), 0; -\sqrt{r_n(x, x)} \right) \right] \mu(dx) \end{aligned}$$

Critère d'échantillonnage :

$$J_n(\tilde{x}) = \int_{\mathbb{X}} \left[\Phi(u_n(x)) - 2\Phi_2 \left(u_n(x), 0; -\sqrt{r_n(x, x) - \tilde{r}_n(x, x; \tilde{x})} \right) \right] \mu(dx)$$

Fonction de gain :

$$\begin{aligned} G_n(\tilde{x}) &= 2 \int_{\mathbb{X}} \left[\Phi_2 \left(u_n(x), 0; -\sqrt{r_n(x, x) - \tilde{r}_n(x, x; \tilde{x})} \right) \right. \\ &\quad \left. - \Phi_2 \left(u_n(x), 0; -\sqrt{r_n(x, x)} \right) \right] \mu(dx) \end{aligned}$$

Critère SUR 3 – Estimation de la fonction de probabilité de défaillance (norme \mathbb{L}^1)

En utilisant (B.6), l'espérance peut être exprimée à l'aide de fonction de répartition de lois normales,

$$H_n = \int_{\mathbb{X}} \left[\Phi(u_n(x)) - 2\Phi_2\left(u_n(x), 0; -\sqrt{r_n(x, x)}\right) \right] \mu(dx). \quad (2.21)$$

Finalement, le critère est toujours défini selon le principe des approches SUR (2.3). La future fonction de covariance est toujours exprimable comme $k_{n+q}(x, x') = k_n(x, x') - \nu_n(x, x'; \tilde{x})$, et la future fonction de moyenne suit toujours la loi d'un processus gaussien de fonction de moyenne m_n et de fonction de covariance ν_n . En utilisant (B.4), il est possible de calculer l'incertitude espérée après q observations. Le gain se calcule par différence entre la mesure d'incertitude courante et la mesure d'incertitude espérée. Le tout est résumé dans l'encadré critère SUR 3.

Une fois encore, ce critère est conçu pour les simulateurs stochastiques, mais il est possible de l'étendre au cas déterministe, en faisant tendre λ vers 0. Dans, ce cas, la QoI devient la fonction de sélection, $p(x) = \mathbb{1}_{\{\xi(x) \geq z^{\text{crit}}\}}$, et la mesure d'incertitude devient $H_n = \int_{\mathbb{X}} \min\{\hat{p}_n(x), 1 - \hat{p}_n(x)\} dx$, ce qui correspond alors à un autre critère proposé par Li [2012].

Critère SUR 4 : estimation de probabilités dépendant de plusieurs sorties

Enfin, il est possible d'appliquer la même méthodologie pour des simulateurs ayant plusieurs sorties. Supposons que le simulateur renvoie b sorties, $(Z^{(a)})_{1 \leq a \leq b}$. Supposons que la distribution de ces sorties au point x soient des lois normales, telles que la sortie $Z^{(a)}$ associée à l'entrée x suivent une loi de moyenne $\xi^{(a)}(x)$ inconnue et de variance $\lambda^{(a)}(x)$ connue. Définissons b seuils critiques, $(z^{\text{crit}(a)})_{1 \leq a \leq b}$. Dans ce cadre, deux types de conditions de défaillance peuvent être définis : soit la défaillance se produit lorsque l'une des sorties dépasse son seuil critique (système en série), auquel cas la probabilité d'intérêt est

$$\mathbb{P}\left(\exists a \in \llbracket 1; b \rrbracket, Z^{(a)} \geq z^{\text{crit}(a)}\right); \quad (2.22)$$

soit la défaillance se produit lorsque toutes les sorties dépassent leurs seuils critiques (système en parallèle), auquel cas la probabilité d'intérêt est

$$\mathbb{P}\left(\forall a \in \llbracket 1; b \rrbracket, Z^{(a)} \geq z^{\text{crit}(a)}\right). \quad (2.23)$$

Dans le cas d'une étude en sécurité incendie, la situation de défaillance se produit quand l'un des sorties dépasse son seuil. La probabilité d'intérêt est donc celle (2.22). Cepen-

QoI : la fonction de probabilité

$$p(x) = \mathbb{P} \left(\forall a \in \llbracket 1; b \rrbracket, Z^{(a)} \geq z^{\text{crit}(a)} \mid x \right) = \prod_{a=1}^b \Phi \left(\frac{\xi^{(a)}(x) - z^{\text{crit}(a)}}{\sqrt{\lambda^{(a)}(x)}} \right),$$

Estimateur : moyenne a posteriori

$$\hat{p}_n(x) = \prod_{a=1}^b \Phi \left(u_n^{(a)}(x) \right),$$

Mesure d'incertitude : intégrale de la variance

$$\begin{aligned} H_n &= \mathbb{E}_n \left[\|p - \hat{p}_n\|_{\mathbb{L}_{\mathbb{X}}^2(\mu)}^2 \right] \\ &= \int_{\mathbb{X}} \left[\left\{ \prod_{a=1}^b \Phi_2 \left(u_n^{(a)}(x), u_n^{(a)}(x); r_n^{(a)}(x, x) \right) \right\} \right. \\ &\quad \left. - \left\{ \prod_{a'=1}^b \Phi \left(u_n^{(a')}(x) \right)^2 \right\} \right] \mu(dx) \end{aligned}$$

Critère d'échantillonnage :

$$\begin{aligned} J_n(\tilde{x}) &= \int_{\mathbb{X}} \left[\left\{ \prod_{a=1}^b \Phi_2 \left(u_n^{(a)}(x), u_n^{(a)}(x); r_n^{(a)}(x, x) \right) \right\} \right. \\ &\quad \left. - \left\{ \prod_{a'=1}^b \Phi_2 \left(u_n^{(a')}(x), u_n^{(a')}(x); \tilde{r}_n^{(a')}(x, x; \tilde{x}) \right) \right\} \right] \mu(dx) \end{aligned}$$

Fonction de gain :

$$\begin{aligned} G_n(\tilde{x}) &= \int_{\mathbb{X}} \left[\left\{ \prod_{a=1}^b \Phi_2 \left(u_n^{(a)}(x), u_n^{(a)}(x); \tilde{r}_n^{(a)}(x, x; \tilde{x}) \right) \right\} \right. \\ &\quad \left. - \left\{ \prod_{a'=1}^b \Phi \left(u_n^{(a')}(x) \right)^2 \right\} \right] \mu(dx) \end{aligned}$$

Critère SUR 4 – Estimation de probabilité de dépassements de seuils simultanés

dant, comme la formule des complémentaires permet de passer d'une situation à l'autre

$$\mathbb{P}\left(\exists a \in \llbracket 1; b \rrbracket, Z^{(a)} \geq z^{\text{crit}(a)}\right) = 1 - \mathbb{P}\left(\forall a \in \llbracket 1; b \rrbracket, Z^{(a)} \leq z^{\text{crit}(a)}\right) \quad (2.24)$$

et que la probabilité de défaillance d'un système en parallèle (2.23) est plus simple à étudier, le critère est conçu pour estimer des probabilités de systèmes en parallèle.

Considérons donc, quitte à utiliser la formule des complémentaires, que la quantité d'intérêt est la probabilité que toutes les sorties dépassent leurs seuils respectifs,

$$p(x) = \mathbb{P}\left(\forall a \in \llbracket 1; b \rrbracket, Z^{(a)} \geq z^{\text{crit}(a)} \mid x\right). \quad (2.25)$$

Dans le cas où les sorties d'une même simulation sont mutuellement indépendantes entre elles, cette probabilité s'exprime comme un produit de probabilités indépendantes. De plus, si les processus gaussiens a priori modélisant $(\xi^{(a)})$ sont indépendants, alors il devient possible d'exprimer les fonctions de moyenne et de covariance de $p(x)$ à l'aide de produit de fonctions. Ceci permet d'aboutir au critère présenté dans l'encadré critère SUR 4.

Notons qu'il est théoriquement possible de calculer le critère SUR 4 dans le cas où les sorties sont corrélées. Dans ce cas, il faut remplacer les produits de b fonctions Φ_2 par une unique fonction Φ_{2b} prenant en compte les corrélations nécessaires. De même, il est possible d'étendre le critère SUR 4 dans le cas où la quantité d'intérêt est une intégrale de la fonction de probabilité (comme dans le critère SUR 2). Le nouveau critère d'échantillonnage dans le cas de plusieurs sorties est similaire à celui où il n'y a qu'une sortie, en plaçant les produits de fonctions de répartition aux bons endroits.

Il est également possible de construire un autre critère d'échantillonnage quand il y a plusieurs sorties. Par exemple, si au lieu de s'intéresser à la probabilité conjointe (2.25), l'objectif est d'estimer toutes les fonctions de probabilité de défaillance de toutes sorties, $p^{(a)}(x)$, alors une mesure d'incertitude possible est l'adaptation du critère SUR 1 dans le cas où il y a plusieurs sorties

$$H_n = \sum_{a=1}^b c_a \int_{\mathbb{X}} \left(p^{(a)}(x) - \hat{p}_n^{(a)}(x)\right)^2 \mu_{\mathbb{X}}(dx), \quad (2.26)$$

où $\hat{p}_n^{(a)}(x)$ est la fonction de moyenne a posteriori de p en x pour la sortie a , $\mu_{\mathbb{X}}$ est une mesure positive bornée sur \mathbb{X} , et $(c_a)_{1 \leq a \leq b}$ sont des coefficients positifs représentant l'importance de la sortie a dans la mesure d'incertitude. Le critère SUR associée à la mesure d'incertitude (2.26) est une somme pondérée de critères similaires au critère

SUR 1 adaptés à chacune des sorties.

Remarque 4. Contrairement aux critères SUR 1 à 3, le critère SUR 4 n'est pas symétrique. Sur les autres critères, le critère pour estimer la probabilité d'être au-dessus du seuil, $\mathbb{P}(Z \geq z^{\text{crit}}|x)$, est aussi le critère pour estimer la probabilité d'être en-dessous du seuil, $\mathbb{P}(Z \leq z^{\text{crit}}|x)$. Dans ces cas, seul l'estimation doit être modifiée. Par contre, si il y a plusieurs sorties, les critères obtenus en changeant l'orientation d'un signe dans l'objectif ne sont pas identiques.

2.3.3 Un exemple de stratégie SUR sans multi-fidélité

Dans cette section, nous allons illustrer le critère SUR 1 sur un exemple simple niveau. Soit la fonction du niveau haute-fidélité dans l'exemple de Forrester et al. [2007], déjà présentée section 1.2.1, $f(x) = (6x - 2)^2 \sin(12x - 4) + 10$. Elle est définie sur $\mathbb{X} = [0; 1] \subset \mathbb{R}^1$. Cette fonction est observée avec un bruit gaussien de variance constante $\lambda = 1.4^2$. Le seuil critique est fixé à $z^{\text{crit}} = 10$. L'objectif est d'estimer la probabilité de dépassement de seuil, $p(x) = \mathbb{P}(Z \geq z^{\text{crit}}|x) = \Phi\left(\frac{f(x) - z^{\text{crit}}}{\sqrt{\lambda}}\right)$. Pour ce faire, un plan initial de $n = 5$ points est utilisé. Puis, à chaque boucle, la fonction de probabilité de dépassement de seuil est estimée, et un nouveau point est choisi par minimisation du critère SUR 1. Le nombre de point total à la fin de la planification est fixé à trente points.

Cette fonction est modélisée par un processus gaussien ayant une moyenne constante et une fonction de covariance de Matérn 5/2. La loi a priori de la moyenne est une loi uniforme impropre sur \mathbb{R} . Les deux hyper-paramètres de la fonction de Matérn suivent a priori des lois log-normales indépendantes, $\log(\sigma^2) \sim \mathbf{N}(2 \log(0.2), \log(100)^2)$ et $\log(\rho) \sim \mathbf{N}(\log(1/2), \log(10)^2)$, où σ^2 et ρ sont respectivement la variance et la portée de la covariance. Le paramètre de variance du bruit suit a priori une loi log-normale indépendante des autres hyper-paramètres, $\log(\lambda) \sim \mathbf{N}(2 \log(0.2), \log(100)^2)$. Les hyper-paramètres sont échantillonnés à l'initialisation selon leur loi a posteriori en utilisant l'algorithme de Metropolis adaptatif de Haario et al. [2001], puis mis à jour avec le Monte-Carlo Séquentiel (SMC)¹⁰ proposé par Chopin [2002] après ajout de chaque nouvelle observation.

Les prédictions des sorties et les estimations de la probabilité de dépassement de seuil sont estimées de manière complètement bayésienne, où l'intégration sur les hyper-paramètres est remplacée par une méthode de Monte-Carlo. Le critère est calculé avec les hyper-paramètres fixés à leur Maximum A Posteriori (MAP) et l'intégrale sur l'espace

10. Sequential Monte-Carlo

des entrées est remplacée par une moyenne sur les points d'une grille régulière. Les points d'observation sont ajoutés un par un.

Les figures 34 et 35 présentent les prédictions des sorties du simulateur, les estimations de la probabilité et la valeur du critère en chaque point des entrées à différentes étapes de la planification. D'observation en observation, les intervalles de crédibilité sur l'estimation de la probabilité tend à se réduire, et la moyenne a posteriori de l'estimation de la probabilité se rapproche de la fonction vraie. Les observations tendent à être choisies en des endroits où les sorties du simulateur sont proches du seuil. Le critère d'échantillonnage diminue quand l'estimation de la fonction se rapproche du seuil critique. Enfin, certaines observations sont faites quasiment au même point. Comme le cadre est stochastique, le critère recommande parfois de refaire des observations en des points déjà observés (visible par exemple figure 34a).

2.3.4 Intégration selon les hyper-paramètres

Jusqu'ici, les critères d'échantillonnage proposés supposent que le facteur d'incertitude réductible est lié au processus ξ qui est supposé être un processus gaussien. Ces critères ne prennent pas en compte la réduction des incertitudes sur les hyper-paramètres θ du processus gaussien. Formellement, la même méthode pourrait être appliquée pour définir des critères SUR. Cependant, cela est généralement impossible car il est impossible d'obtenir des formules analytiques simples dans des cas généraux. La principale raison est que, dans le cas général non gaussien, la fonction de covariance future k_{n+q} dépend des résultats des observations aux points d'observations. Nous ne pouvons plus calculer les critères SUR par une intégrale sur les valeurs possibles des q valeurs aux nouvelles observations.

Pour résoudre ce problème, une solution aisément applicable consiste à appliquer une méthode de substitution : les hyper-paramètres θ sont fixés à une valeur $\hat{\theta}_n$, par exemple le maximum de vraisemblance ou le MAP. Le calcul du critère d'échantillonnage se fait alors avec le processus gaussien $\xi|\chi_n, \hat{\theta}_n$ en négligeant l'incertitude sur les hyper-paramètres θ . Cette solution a été appliquée dans la section 2.3.3 sur l'exemple illustratif.

Dans le cas où la variabilité des hyper-paramètres doit être prise en compte dans le critère d'échantillonnage, il est possible d'obtenir une approximation du critère idéal. Dans un cadre plus général [Bect et al., 2017], une mesure d'incertitude est définie comme $H_n = \inf_{d \in \mathbb{D}} \{\mathbb{E}_n [\mathcal{L}(d, \xi)]\}$, où \mathbb{D} est un espace de décision et \mathcal{L} une fonction de perte. Si les hyper-paramètres θ sont supposées connus, la mesure d'incertitude peut être définie conditionnellement à θ , en conservant la même fonction de perte :

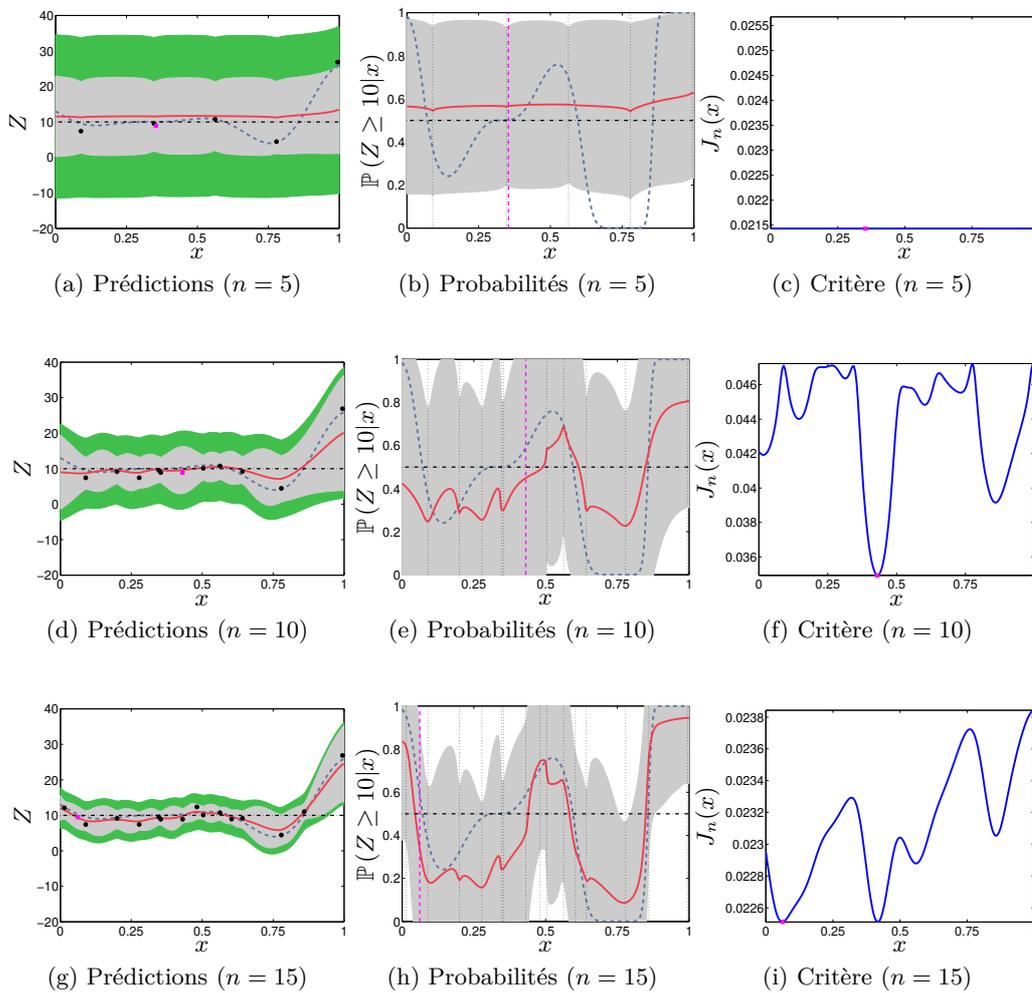


FIGURE 34 – Première partie d’une planification séquentielle d’expériences (seconde partie figure 35) en simple fidélité en utilisant un critère SUR. Chaque ligne correspond à une étape, chaque colonne à un résultat. (a, d, g) Prédictions des sorties. Le trait plein est la moyenne a posteriori, l’aire intérieure grise est l’intervalle à deux écarts-types sur la moyenne ξ , l’aire extérieure verte est l’intervalle à deux écarts-types sur les sorties Z , le trait discontinu est la vraie fonction, le trait discontinu horizontal est le seuil critique, les points noirs sont les observations et le carré mauve est la prochaine observation. (b, e, h) Estimations des probabilités. Le trait plein est la moyenne a posteriori, l’aire est l’intervalle à deux écarts-types, le trait discontinu est la vraie probabilité de défaillance, les traits verticaux sont les observations (courantes et futures). (c, f, i) Critère d’échantillonnage. Le trait plein est la valeur du critère en chaque point et le carré mauve est le minimum.

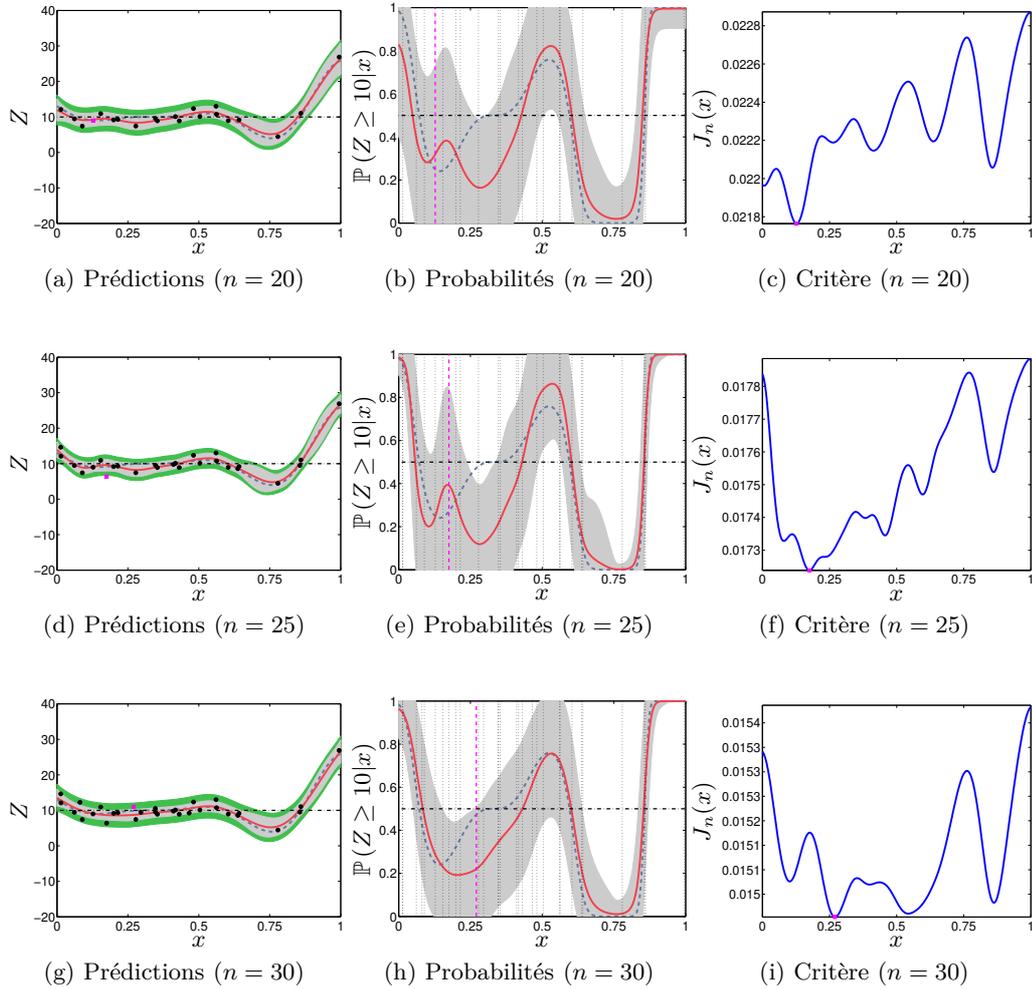


FIGURE 35 – Seconde partie d’une planification séquentielle d’expériences (première partie figure 34) en simple fidélité en utilisant un critère SUR. Chaque ligne correspond à une étape, chaque colonne à un résultat. (a, d, g) Prédictions des sorties. Le trait plein est la moyenne a posteriori, l’aire intérieure grise est l’intervalle à deux écarts-types sur la moyenne ξ , l’aire extérieure verte est l’intervalle à deux écarts-types sur les sorties Z , le trait discontinu est la vraie fonction, le trait discontinu horizontal est le seuil critique, les points noirs sont les observations et le carré mauve est la prochaine observation. (b, e, h) Estimations des probabilités. Le trait plein est la moyenne a posteriori, l’aire est l’intervalle à deux écarts-types, le trait discontinu est la vraie probabilité de défaillance, les traits verticaux sont les observations (courantes et futures). (c, f, i) Critère d’échantillonnage. Le trait plein est la valeur du critère en chaque point et le carré mauve est le minimum.

$H_n(\theta) = \inf_{d \in \mathbb{D}} \{\mathbb{E}_{n,\theta} [\mathcal{L}(d, \xi)]\}$, où le θ en indice indique que l'espérance est calculée conditionnellement aux hyper-paramètres. Dans ces conditions,

$$H_n = \mathbb{E}_n [H_n(\theta)] + (H_n - \mathbb{E}_n [H_n(\theta)]). \quad (2.27)$$

La quantité $(H_n - \mathbb{E}_n [H_n(\theta)])$ représente l'incertitude supplémentaire qui s'ajoute du fait de l'incertitude sur les hyper-paramètres. Cette quantité est positive. En effet,

$$\begin{aligned} H_n - \mathbb{E}_n [H_n(\theta)] &= \inf_{d \in \mathbb{D}} \{\mathbb{E}_n [\mathcal{L}(d, \xi)]\} - \mathbb{E}_n \left[\inf_{d \in \mathbb{D}} \{\mathbb{E}_{n,\theta} [\mathcal{L}(d, \xi)]\} \right] \\ &= \inf_{d \in \mathbb{D}} \{\mathbb{E}_n [\mathbb{E}_{n,\theta} [\mathcal{L}(d, \xi)]]\} - \mathbb{E}_n \left[\inf_{d \in \mathbb{D}} \{\mathbb{E}_{n,\theta} [\mathcal{L}(d, \xi)]\} \right]. \end{aligned}$$

Comme $\inf_{d \in \mathbb{D}} \{\mathbb{E}_n [\mathbb{E}_{n,\theta} [\mathcal{L}(d, \xi)]]\} \geq \mathbb{E}_n [\inf_{d \in \mathbb{D}} \{\mathbb{E}_{n,\theta} [\mathcal{L}(d, \xi)]\}]$, le nouveau terme qui s'ajoute est positif.

Dans le cas particulier où les minima sont atteints, il est possible de réécrire (2.27) sous une forme proche de la loi de la variance totale. Soit $d_n = \arg \min_{d \in \mathbb{D}} \{\mathbb{E}_n [\mathcal{L}(d, \xi)]\}$ et $d_n(\theta) = \arg \min_{d \in \mathbb{D}} \{\mathbb{E}_{n,\theta} [\mathcal{L}(d, \xi)]\}$. Nous définissons alors

$$I_n(\theta) = \mathbb{E}_{n,\theta} [\mathcal{L}(d_n, \xi) - \mathcal{L}(d_n(\theta), \xi)] \quad (2.28)$$

une fonction positive en θ . Ceci nous permet d'écrire :

$$H_n = \mathbb{E}_n [H_n(\theta)] + \mathbb{E}_n [I_n(\theta)]. \quad (2.29)$$

Par exemple, si l'objectif est d'estimer une QoI scalaire Q , une fonction de perte possible est la perte quadratique. L'estimation optimale est alors $\hat{Q}_n = \arg \min_{q \in \mathbb{R}} \{\mathbb{E}_n [(Q - q)^2]\} = \mathbb{E}_n [Q]$ et la mesure d'incertitude est la variance $H_n = \text{var}_n [Q]$. L'estimateur conditionnellement aux hyper-paramètres est l'espérance conditionnellement aux hyper-paramètres $\hat{Q}_n(\theta) = \mathbb{E}_{n,\theta} [Q]$. Dans ces conditions,

$$I_n(\theta) = \mathbb{E}_{n,\theta} \left[(Q - \mathbb{E}_n [Q])^2 - (Q - \mathbb{E}_{n,\theta} [Q])^2 \right] = (\mathbb{E}_n [Q] - \mathbb{E}_{n,\theta} [Q])^2. \quad (2.30)$$

La décomposition (2.29) donne ainsi

$$H_n = \mathbb{E}_n [\text{var}_{n,\theta} [Q]] + \mathbb{E}_n \left[(\mathbb{E}_{n,\theta} [Q] - \mathbb{E}_n [Q])^2 \right] = \mathbb{E}_n [\text{var}_{n,\theta} [Q]] + \text{var}_n [\mathbb{E}_{n,\theta} [Q]], \quad (2.31)$$

qui est la loi de la variance totale.

Le critère d'échantillonnage définie dans l'approche SUR s'écrit :

$$J_n(\tilde{x}) = \mathbb{E}_n [H_{n+q}|\tilde{x}] = \mathbb{E}_n [H_{n+q}(\theta)|\tilde{x}] + \mathbb{E}_n [I_{n+q}(\theta)|\tilde{x}]. \quad (2.32)$$

La première quantité est facile à exprimer

$$\mathbb{E}_n [H_{n+q}(\theta)|\tilde{x}] = \mathbb{E}_n [\mathbb{E}_n [H_{n+q}(\theta)|\tilde{x}, \theta] | \tilde{x}] = \mathbb{E}_n [J_n(\tilde{x}, \theta)] = \int_{\Theta} J_n(\tilde{x}, \theta) \pi_{\Theta}(\theta|\chi_n) d\theta, \quad (2.33)$$

où $J_n(\tilde{x}, \theta)$ est le critère calculé sous l'hypothèse que les hyper-paramètres sont fixées. L'intégrale sur les hyper-paramètres peut être approchée par une méthode de Monte-Carlo. Par contre, l'autre quantité $\mathbb{E}_n [I_{n+q}(\theta)|\tilde{x}]$ est très difficilement estimable. Même en supposant que la loi des hyper-paramètres reste inchangée quand de nouvelles observations sont ajoutées, $\pi_{\Theta}(\theta|\chi_{n+q}) = \pi_{\Theta}(\theta|\chi_n)$, cette quantité peut nécessiter plusieurs intégration sur l'espace des hyper-paramètres.

Une façon d'approcher ce critère est d'utiliser l'approximation

$$\mathbb{E}_n [I_{n+q}(\theta)|\tilde{x}] \approx \mathbb{E}_n [I_n(\theta)], \quad (2.34)$$

ce qui revient à négliger l'évolution de l'incertitude liée à la méconnaissance des hyper-paramètres. Sous cette approximation, les critères d'échantillonnage et de gain s'écrivent

$$J_n(\tilde{x}) \approx \mathbb{E}_n [J_n(\tilde{x}, \theta)] + \mathbb{E}_n [I_n(\theta)] \quad (2.35a)$$

$$G_n(\tilde{x}) \approx \mathbb{E}_n [G_n(\tilde{x}, \theta)], \quad (2.35b)$$

où $G_n(\tilde{x}) = H_n - J_n(\tilde{x})$ est le gain total et $G_n(\tilde{x}, \theta) = H_n(\theta) - J_n(\tilde{x}, \theta)$ est le gain conditionnellement aux hyper-paramètres θ . Comme la quantité $\mathbb{E}_n [I_n(\theta)]$ ne dépend pas des points proposées \tilde{x} , il n'est pas nécessaire de la calculer pour choisir le(s) prochain(s) point(s) d'observation. Dans ce cas, le critère et le gain deviennent des moyennes sur l'espace des hyper-paramètres de leurs équivalents respectifs conditionnement aux hyper-paramètres. Ces moyennes peuvent être approchées par des méthodes de Monte-Carlo.

Remarque 5. Dans le cas particulier de l'EI, le nouveau terme d'incertitude due aux hyper-paramètres $\mathbb{E}_n [I_n(\theta)]$ est nul. Cela est du au fait que la décision optimale ne dépend pas des hyper-paramètres [Bect et al., 2017, Section 4.4], $d_n = d_n(\theta)$, ce qui implique la nullité de $I_n(\theta)$ pour tout θ (voir (2.28)). Il est donc possible de créer un critère d'échantillonnage complètement bayésien. Benassi et al. [2011] développent cet aspect et proposent un algorithme de calcul et d'optimisation de ce critère.

2.4 Planification séquentielle en multi-fidélité

Dans la section 2.3, nous avons présenté une méthode de construction de stratégie de planification séquentielle d'expériences dans le cas où toutes les observations ont le même coût, la méthode SUR. Dans cette section, nous revenons sur les plans multi-fidèles, dont la fidélité est indexée par un paramètre d'entrée δ . La principale différence entre les deux situations est que l'observation d'un point x à un niveau δ nécessite un temps de calcul dépendant des entrées. Comme la quantité d'intérêt est à évaluer sur le niveau de plus haute fidélité, plus une simulation a une fidélité élevée, plus il est certain qu'elle sera coûteuse en temps de calcul, mais en même temps, plus il est probable qu'elle apportera une plus grande quantité d'information. L'objectif est donc de trouver le niveau de fidélité réalisant le meilleur compromis entre gain d'information et coût de simulation. Le coût de simulation prenant les entrées x et δ est noté $C(x, \delta)$.

2.4.1 Revue de la littérature

Méthodes ne prenant pas en compte le coût

La littérature propose plusieurs méthodes pour planifier des expériences dans un contexte multi-fidèle. Certaines méthodes n'exploitent pas le coût des observations. La plus simple des méthodes de planifications séquentielles en multi-fidélité consiste à choisir un point x avec une méthode dédiée à la planification sur un niveau de fidélité (voir section 2.3), puis à observer tous les niveaux de fidélité disponibles. L'intérêt de la multi-fidélité dans cette approche réside dans l'utilisation d'un plan initial multi-fidèle qui permet de trouver plus rapidement les zones d'intérêts. Cet algorithme est appliqué par Forrester et al. [2007] dans un objectif d'optimisation de la fonction haute-fidélité. Les auteurs utilisent l'EI sur le plus haut niveau de fidélité pour choisir le point x puis observent tous les niveaux de fidélité (dans l'article, il y a deux niveaux). Le même algorithme est appliqué à l'identique par Pang et al. [2017] dans un cadre bruité, et également par Courrier et al. [2014], à la nuance près que ces derniers alternent des phases d'améliorations du méta-modèle (critère : erreur quadratique) et des phases de recherche de l'optimum (critère : EI). L'intérêt de cette approche est d'abord sa simplicité et ensuite qu'elle permet de conserver une structure emboîtée.

Une autre méthode simple, utilisée par Koziel et al. [2013] toujours dans un problème d'optimisation, consiste à utiliser le modèle de basse-fidélité pour trouver un point initial, puis à appliquer un algorithme d'optimisation au niveau haute-fidélité en partant de ce point initial. De manière assez similaire, Kuya et al. [2011] propose un plan d'ex-

périences multi-fidèles à deux niveaux en deux étapes : d'abord, construire un modèle de régression par processus gaussien au niveau de plus basse-fidélité, en ajoutant progressivement des observations jusqu'à ce que le niveau soit suffisamment connu, en utilisant un critère dédié à l'approximation d'une fonction coûteuse. Ensuite, utiliser un critère soit d'approximation, soit de recherche d'optimum sur le niveau haute-fidélité en fonction de l'objectif. Il est remarquable que Kuya et al. [2011] n'impose pas que le plan d'expériences soit emboîté, ce qui leur permet d'observer des points haute-fidélité qui ne sont pas observés en basse-fidélité.

Pour obtenir une régression d'une fonction coûteuse haute-fidélité à partir d'une approximation basse-fidélité, Xiong et al. [2013] propose de construire des NLHS successifs. L'algorithme se fonde sur l'algorithme de Qian [2009] présenté section 2.2. Le plan initial est un plan NLHS avec S niveaux, $\{X_s\}_{1 \leq s \leq S}$, où $X_{s+1} \subset X_s$ et les X_s sont des LHS. Le nombre de points au niveau s est le double du nombre de points au niveau $s + 1$, $n_s = 2n_{s+1}$. Un plan d'expérience X_0 composé de $n_0 = 2n_1$, est tiré aléatoirement avec la méthode de Qian [2009], de sorte que $\{X_s\}_{0 \leq s \leq S}$ forme un NLHS. Ensuite, quand un nouveau point doit être ajouté au niveau s , ils choisissent le point parmi l'ensemble $X_{s-1} \setminus X_s$ qui minimise l'erreur de validation croisée future. Chaque niveau s est observé deux fois plus souvent que le niveau $s + 1$: ainsi, pour une nouvelle observation au niveau S , il y a deux observations au niveau $S - 1$, quatre au niveau $S - 2$, ..., 2^{S-1} au premier niveau. Toutes les $2^S - 1$ observations, le nombre total d'observations à tous les niveaux aura doublé, et les plans d'expériences se sont décalés d'un niveau (le plan du premier niveau sera X_0 , celui du deuxième niveau sera X_1 ...). Dans ce cas, un nouveau LHS emboîté sur X_0 est tiré et l'algorithme recommence. L'intérêt de l'algorithme est de conserver un plan emboîté avec une structure en LHS. En particulier, quand les plans d'expériences se sont décalés d'un niveau, le plan ainsi obtenu est un NLHS. Son inconvénient majeur est la rigidité algorithmique imposée sur les points d'observation potentiels ainsi que sur le rapport entre les nombres de points à chaque niveau. Cet algorithme est appliqué par Ezzat et al. [2017] pour résoudre un problème de calibration entre un simulateur basse-fidélité et une expérimentation réelle haute-fidélité, avec de légères modifications. Le simulateur prend deux type d'entrées, des entrées physiques x_p et des entrées de calibrations x_c , alors que les expériences réelles ne prennent que des entrées physiques x_p , l'objectif étant de trouver les entrées x_c qui réduisent l'erreur entre les deux modèles. De ce fait, quand un plan LHS dans le domaine $\mathbb{X}_p \times \mathbb{X}_c$ est proposé pour le niveau haute-fidélité, seule les entrées physiques dans \mathbb{X}_p sont conservées.

Chen et al. [2013] propose un autre algorithme dans le cas où l'objectif est d'évaluer un contour du niveau haute-fidélité, $\{x \in \mathbb{X}, \xi_S(x) = z^{\text{crit}}\}$, où ξ_S est la fonction du

niveau haute-fidélité, et z^{crit} est le seuil critique. Leur problème ne possède que deux niveau de fidélité, $s = 1$ ou 2 . Pour sélectionner les observations (x_{n+1}, s_{n+1}) , ils choisissent d'abord le point d'observation x_{n+1} en minimisant un critère dédié au problème, calculé au niveau de plus haute-fidélité (en l'occurrence, le critère de Ranjan et al. [2008]). Puis, ils appliquent une heuristique pour choisir le niveau de fidélité. Ils comparent la distance entre la prédiction haute-fidélité et le seuil critique. Si la prédiction est suffisamment proche du seuil, l'observation est faite au niveau haute-fidélité ; si par contre, la prédiction est jugée trop lointaine du seuil, l'observation est faite au niveau basse-fidélité. Les auteurs justifient cet algorithme en supposant que le modèle cherche à observer un point proche du seuil pour raffiner d'une solution déjà existante, et à observer un point loin du seuil lors des phases exploratoires. Si cet algorithme est relativement simple à mettre en œuvre, l'hypothèse de base est fortement discutable. Une observation en haute-fidélité peut avoir pour objectif de raffiner la relation entre les niveaux. Elle peut aussi se justifier dans le cas où le niveau basse-fidélité représente mal le niveau haute-fidélité, auquel cas la meilleure solution serait de n'observer que le niveau haute-fidélité. A l'inverse, dans le cas où le niveau basse-fidélité représente parfaitement le niveau haute-fidélité ($\xi_1(x) = \xi_2(x)$), il est inutile de gaspiller des ressources calculatoires sur le niveau haute-fidélité. De ce fait, une observation basse-fidélité peut se justifier même pour raffiner une solution.

Kandasamy et al. [2016] propose un plan d'expérience multi-fidèle plus sophistiqué. Dans un contexte d'un problème de bandit manchot multi-fidèle, les auteurs cherchent à minimiser le regret cumulé lié à la fonction haute-fidélité ξ_S . Plusieurs approximations $(\xi_s)_{1 \leq s \leq S-1}$ sont disponibles et l'erreur maximale entre un niveau quelconque s et le niveau de plus haute-fidélité est supposée connue, $\zeta_s = \|\xi_s - \xi_S\|_\infty$. Dans ces conditions, leur nouveau point (x_{n+1}, s_{n+1}) est choisi en deux étapes, suivant un algorithme nommé Borne Supérieure de Confiance d'un Processus Gaussien en Multi-Fidélité (MF-GP-UCB)¹¹. D'abord, le nouveau point d'observation x_{n+1} est choisi par maximisation d'un critère d'échantillonnage

$$x_{n+1} = \arg \max_{x \in \mathbb{X}} \{J(x)\}, \quad J(x) = \min_{1 \leq s \leq S} \left\{ m_n(x, s) + \beta \sqrt{k_n(x, s)} + \zeta_s \right\}. \quad (2.36)$$

Ce critère d'échantillonnage représente la borne de confiance supérieure de la fonction $\xi_S(x)$. Cet algorithme est directement adapté de l'algorithme de la Borne Supérieure de Confiance d'un Processus Gaussien (GP-UCB)¹² proposé par Cox et John [1997] et

11. Multi-Fidelity Gaussian Process Upper Confidence Bound

12. Gaussian Process Upper Confidence Bound

présenté par Srinivas et al. [2009], qui permet d'optimiser des fonctions coûteuses. L'idée est d'optimiser la borne de confiance supérieure plutôt que la moyenne directement afin de prendre en compte les incertitudes sur les prédictions. Le terme ζ_s permet de prendre en compte l'écart entre le niveau de fidélité et le niveau d'intérêt. Une fois le point x_{n+1} choisi, ils sélectionnent le premier niveau de fidélité incertain en partant du niveau de plus basse-fidélité tel que l'écart-type dépasse un seuil critique :

$$s_{n+1} = \min \left\{ t \in \llbracket 1; S \rrbracket, \beta \sqrt{k_n(x, t)} > \gamma^{\text{crit}} \right\}, \quad (2.37)$$

où γ^{crit} est un seuil maximal d'incertitude fixé par l'utilisateur. L'idée est que si l'écart-type du niveau s est plus grand que $\gamma^{\text{crit}}/\beta$, alors il est possible de gagner beaucoup d'information en observant le niveau s . Ceci garantit qu'avant d'observer un niveau s au point x , les niveaux de plus-basse fidélité sont bien connus en ce point. Cette approche se retrouve également dans Le Gratiet [2013]. Dans le cas où tous les écart-types sont inférieures à γ^{crit} , le niveau de plus haute-fidélité est choisi automatiquement, $s_{n+1} = S$. L'inconvénient majeur de cette méthode est qu'elle requiert de connaître l'erreur absolue entre deux niveaux de fidélité, $\zeta_s = \|\xi_s - \xi_S\|_\infty$, ce qui n'est pas le cas dans les applications en sécurité incendie.

Méthodes utilisant les coûts d'observations

Certaines méthodes de planifications séquentielles d'expériences utilisent le coût d'observation d'un niveau de fidélité $C(x, \delta)$ pour choisir le niveau observé. Généralement dans les applications en multi-fidélité, ce coût ne dépend, au moins en première approximation, que du niveau de fidélité δ , $C(x, \delta) = C(\delta)$. Zhou et al. [2017] suggèrent de ne comparer que des plans d'expériences ayant des coûts égaux. Leur objectif est d'obtenir une estimation de la fonction haute-fidélité, ξ_S . Pour cela, ils définissent comme mesure d'incertitude l'intégrale de la racine carrée de l'écart-type,

$$H_n = \int_{\mathbb{X}} \sqrt[4]{k_n(x, x)} dx, \quad (2.38)$$

de manière similaire à l'Erreur Quadratique Moyenne Intégrée (IMSE)¹³. L'intégrale est approchée par une méthode de Monte-Carlo. L'idée est de comparer les réductions d'incertitude apportées par q_s observations au niveau s avec $q_s = C(S)/C(s)$. Ainsi, pour chaque niveau de fidélité s , ils sélectionnent les q_s points du niveau s ayant la plus

13. Integrated Mean Square Error

grande incertitude de prédiction en tenant compte des nouveaux points de prédiction,

$$X_s = \left(x_{n+1}^{(s)}, \dots, x_{n+q_s}^{(s)} \right), \quad x_{n+l}^{(s)} = \arg \max_{x \in \mathbb{X}} \{ k_{n+l-1}((x, s), (x, s)) \}, \quad (2.39)$$

où k_{n+l-1} est la variance a posteriori conditionnellement aux χ_n points d'observations ainsi qu'aux $(l-1)$ points proposées $(x_{n+1}^{(s)}, \dots, x_{n+l-1}^{(s)})$. Puis ils choisissent le plan d'expériences associé au niveau s qui fournit la plus grande réduction d'incertitude

$$X_s = \arg \min_{1 \leq s \leq S} \{ H_{n+q_s} \}. \quad (2.40)$$

Cet algorithme présente l'inconvénient que le rapport des coûts doit être un entier. Pour l'étendre à d'autres types de problèmes, la principale limitation est la capacité de calculer la réduction d'incertitude après q observations.

La plupart des algorithmes de planification séquentielle utilisant la fonction de coût procèdent généralement autrement. Les articles redéfinissent un critère d'échantillonnage sur l'espace complet des entrées $\mathbb{X} \times \Delta$, physiques et fidélité. Comme il est souhaitable de minimiser le coût dépensé, et que seule des quantités à coût égal sont comparables, les critères généralement proposés sont de la forme

$$(x_{n+1}, \delta_{n+1}) = \arg \max_{x \in \mathbb{X}, \delta \in \Delta} \left\{ \frac{G_n^{\text{multi}}(x, \delta)}{C(x, \delta)} \right\}, \quad (2.41)$$

où $G_n^{\text{multi}}(x, \delta)$ mesure l'intérêt à observer le niveau δ au point x . Cette forme de critère se retrouve chez Huang et al. [2006], Swersky et al. [2013], Le Gratiet et Cannamela [2015] et He et al. [2017]. Chacun d'entre eux partent d'un critère défini à un niveau de fidélité pour définir leur critère d'échantillonnage multi-fidélité.

Ainsi, en notant $\delta = 0$ le niveau de plus haute fidélité, Huang et al. [2006] utilise une variante de l'EI pour optimiser un modèle numérique haute-fidélité

$$G_n^{\text{multi}}(x, \delta) = EI_n(x, 0) \cdot \frac{k_n((x, \delta), (x, 0))}{\sigma_n(x, 0)\sigma_n(x, \delta)} \cdot \left(1 - \sqrt{\frac{\lambda(x, \delta)}{\sigma_n^2(x, \delta) + \lambda(x, \delta)}} \right), \quad (2.42)$$

où $\sigma_n^2(x, \delta) = k_n((x, \delta), (x, \delta))$, et $\lambda(x, \delta)$ est la variance de la sortie au point (x, δ) . Le premier terme est l'EI du point x au plus haut niveau de fidélité, le deuxième terme est la corrélation a posteriori entre le niveau de plus haute-fidélité et le niveau δ , et le troisième terme permet d'adapter l'EI aux simulateurs stochastiques. Toujours dans un problème d'optimisation, He et al. [2017] utilise une variante de l'Amélioration Espérée

des Quantiles (EQI)¹⁴ en considérant la différence $\xi(x, \delta) - \xi(x, 0)$ comme un bruit perturbant l'observation de la fonction idéale.

Le Gratiet et Cannamela [2015] proposent un algorithme de planification séquentielle dans un problème d'approximation de la fonction coûteuse. Pour cela, il commence par définir une mesure d'incertitude qui est l'intégrale de la variance prédictive multipliée par un coefficient d'ajustement, (inspiré de l'IMSE). Puis, les auteurs comparent successivement deux rapports : le rapport entre la réduction d'incertitude apportée par le niveau s et celle du niveau $s + 1$; et le rapport des coûts d'observations des deux niveaux de fidélité. Dès que le rapport des réductions est plus grand que le rapport des coûts, le niveau s est observé. Cet algorithme peut se ré-écrire de sous la forme (2.41) avec

$$G_n^{\text{multi}}(x, \delta) = \int_{\mathbb{X}} \left(\tilde{\sigma}_n^2(y, 0) - \tilde{\sigma}_{n+1}^2((y, 0)|(x, \delta)) \right) dy \quad (2.43)$$

où $\tilde{\sigma}_n$ est la fonction de variance ajustée, et $\tilde{\sigma}_{n+1}^2((y, 0)|(x, \delta))$ est la variance ajustée a posteriori future au point $(y, 0)$ si le point (x, δ) est observé. La principale différence est que les auteurs choisissent le maximum local du rapport G^{multi}/C qui est le plus proche du niveau de basse-fidélité. Diverses approximations sont utilisées pour choisir le point x et pour le calcul d'intégrale.

Enfin, cette forme de planification séquentielle peut aussi être trouvée dans un contexte différent de la multi-fidélité. Dans un problème d'optimisation d'une fonction prenant des entrées x ayant des coûts d'observation $C(x)$, Swersky et al. [2013] définissent un critère de réduction d'entropie comme fonction de gain

$$G_n^{\text{multi}}(x, \delta) = \mathbb{E}_n \left[\mathbb{H}(P_n^{\min}) - \mathbb{H}(P_{n+1}^{\min}) | X_{n+1} = x \right], \quad (2.44)$$

où P_n^{\min} est la loi de probabilité de l'argument du minimum conditionnellement aux observations χ_n , \mathbb{H} désigne l'entropie, et P_{n+1}^{\min} est la loi de probabilité de l'argument du minimum conditionnellement aux observations χ_n et à la nouvelle observation en x . Ce critère est la réduction d'incertitude du critère IAGO [Villemonteix et al., 2008]. par conséquent, le critère proposé par Swersky et al. [2013] est une réduction d'incertitude au sens des stratégies SUR. La loi de l'argument du minimum est approchée par Monte-Carlo. Dans le même cadre que Swersky et al. [2013], Snoek et al. [2012] propose de considérer directement l'EI comme mesure d'intérêt à observer le point x .

14. Expected Quantile Improvement

2.4.2 Taux maximal de réduction de l'incertitude

Nous proposons une méthode générale pour construire un critère d'échantillonnage dans le cas où le coût d'observation dépend du point (x, δ) . Ce coût $C(x, \delta)$ est supposé pouvoir dépendre de toutes les entrées. Ceci inclut en particulier le cas de la multi-fidélité, où le coût ne dépend que du paramètre de fidélité δ .

L'objectif est de minimiser une mesure d'incertitude H_n en limitant le coût de calcul. Même dans le cas où le coût de calcul varie avec le point observé $C(x, \delta)$, il reste possible de calculer le critère SUR associé en n'importe quel point d'observation, $J_n(x, \delta) = \mathbb{E}_n [H_{n+1} | X_{n+1} = x, D_{n+1} = \delta]$. Il y a donc deux objectifs à minimiser simultanément : l'incertitude future espérée et le coût. De fait, quelle que soit la façon dont le point d'observation est choisi, il paraît raisonnable de choisir un point d'observations qui soit sur le front de Pareto de l'espace (J, C) . Si pour une proposition d'observation (x, δ) , il est possible de trouver un autre point, soit qui possède la même incertitude future pour un coût moindre, soit qui possède une incertitude espérée inférieure pour un coût similaire, alors il faudrait préférer ce nouveau point.

La figure 36 présente un exemple en simulation du type de situation considérée. Soit un processus gaussien multi-fidèle sur $\mathbb{X} \times \Delta$, $\xi(x, \delta)$ où δ est le paramètre de fidélité. L'espace des entrées x est $\mathbb{X} = [-0.5, 0.5] \subset \mathbb{R}^1$, le paramètre de fidélité δ varie dans $\Delta = [0, 1]$, où le niveau 0 est le niveau de plus haute-fidélité. Le processus ξ est un processus gaussien de la forme décrite dans la section 1.2.2,

$$\xi \sim \text{GP} \left(0, (1.0)^2 \cdot \mathcal{M}_{5/2} \left(\frac{|x - x'|}{0.3} \right) + (4.0)^2 \cdot \min \{ \delta, \delta' \}^2 \cdot \mathcal{M}_{5/2} \left(\frac{|x - x'|}{0.1} \right) \right). \quad (2.45)$$

Les points sont observées avec un bruit homoscédastique de variance $\lambda = 0.4^2$. Un DoE initial de douze, six, six et trois points est tiré selon la méthode décrite section 2.2.2 aux niveaux $\delta = 1, 1/2, 1/4$ et $1/10$. Les sorties associées à ces entrées sont obtenues par tirage d'une trajectoire non-conditionnelle du modèle. Un seuil critique z^{crit} est fixé à zéro, et le critère SUR 1 est utilisé pour calculer l'incertitude future sur la probabilité de dépassement de seuil au niveau de plus haute fidélité, les hyper-paramètres de la fonction de covariance étant supposés connus. La valeur de la réduction d'incertitude est tracée pour chaque observation potentielle (x, δ) en figure 36a. Évidemment, plus l'observation est proche du niveau haute-fidélité, plus elle tend à apporter de l'information sur le dépassement de seuil en haute-fidélité. La fonction de coût de cet exemple est $C(x, \delta) = 1/\delta$. La figure 36b montre le rapport entre la réduction d'incertitude et le coût d'observation dans l'espace des entrées. Ceci fait apparaître des optimum locaux

à l'intérieur du domaine. Les figures 36c et 36d montrent ces mêmes critères mais dans les espaces des objectifs. Le trait plein est le front de Pareto de l'espace coût-réduction d'incertitude. Ce front correspond à l'ensemble des points d'intérêt pour une future observation. Sur ce front, le point de plus bas coût est le point $(0, H_n)$, le point où aucune observation n'est faite et le point de plus haut coût correspond au point renvoyé par le critère SUR. Le front de Pareto commence par descendre, puis finit par atteindre un plateau à partir d'un certain coût, à droite de la figure 36c. Tout l'objectif est d'essayer de choisir le point du front se trouvant juste avant ce plateau, juste au niveau où l'augmentation du coût d'observation n'apporte plus d'information sur l'objectif.

Pour ce faire, nous proposons la stratégie du Taux Maximal de Réduction de l'Incertitude (MRUR)¹⁵, qui consiste à choisir le point qui maximise le rapport entre la réduction d'incertitude espérée et le coût :

$$\begin{aligned} (x_{n+1}, \delta_{n+1}) &= \arg \max_{x \in \mathbb{X}, \delta \in \Delta} \left\{ \frac{H_n - \mathbb{E}_n [H_{n+1} | X_{n+1} = x, D_{n+1} = \delta]}{C(x, \delta)} \right\} \\ &= \arg \max_{x \in \mathbb{X}, \delta \in \Delta} \left\{ \frac{G_n(x, \delta)}{C(x, \delta)} \right\}. \end{aligned} \quad (2.46)$$

Dans le cas où tous les coûts d'observation sont égaux, c'est-à-dire, dans la situation classique, ce critère devient équivalent au critère SUR. Les critères proposés par Snoek et al. [2012] et Swersky et al. [2013] se trouvent être des cas particuliers de critères MRUR, mais proposés dans un cadre autre que la multi-fidélité. Swersky et al. [2013] propose comme critère la réduction d'entropie de la loi de l'argument du minimum (2.44) et Snoek et al. [2012] l'expérience d'amélioration. Ces deux fonctions correspondent aux fonctions de gain dans leurs critères SUR respectifs. Les stratégies proposées sont donc des cas particulier de MRUR. De même, la stratégie de Le Gratiet et Cannamela [2015] peut se ré-écrire sous la forme d'un critère MRUR fondé sur une variante de l'IMSE.

La figure 37 présente le résultat obtenu en poursuivant l'exemple illustratif et en choisissant à chaque fois le point maximisant le critère MRUR. La figure 37a montre que la réduction d'incertitude obtenue par MRUR est obtenue plus rapidement que si le choix proposé par la stratégie SUR, sans tenir compte du coût d'observation, avait été observé. Ainsi, l'observation du point choisi par SUR aurait coûté environ 100 alors que au bout de 25, la stratégie MRUR a obtenue une réduction d'incertitude plus importante. La figure 37b montre que la valeur maximale du critère MRUR tend à se réduire avec le nombre d'observations.

15. Maximal Rate of Uncertainty Reduction

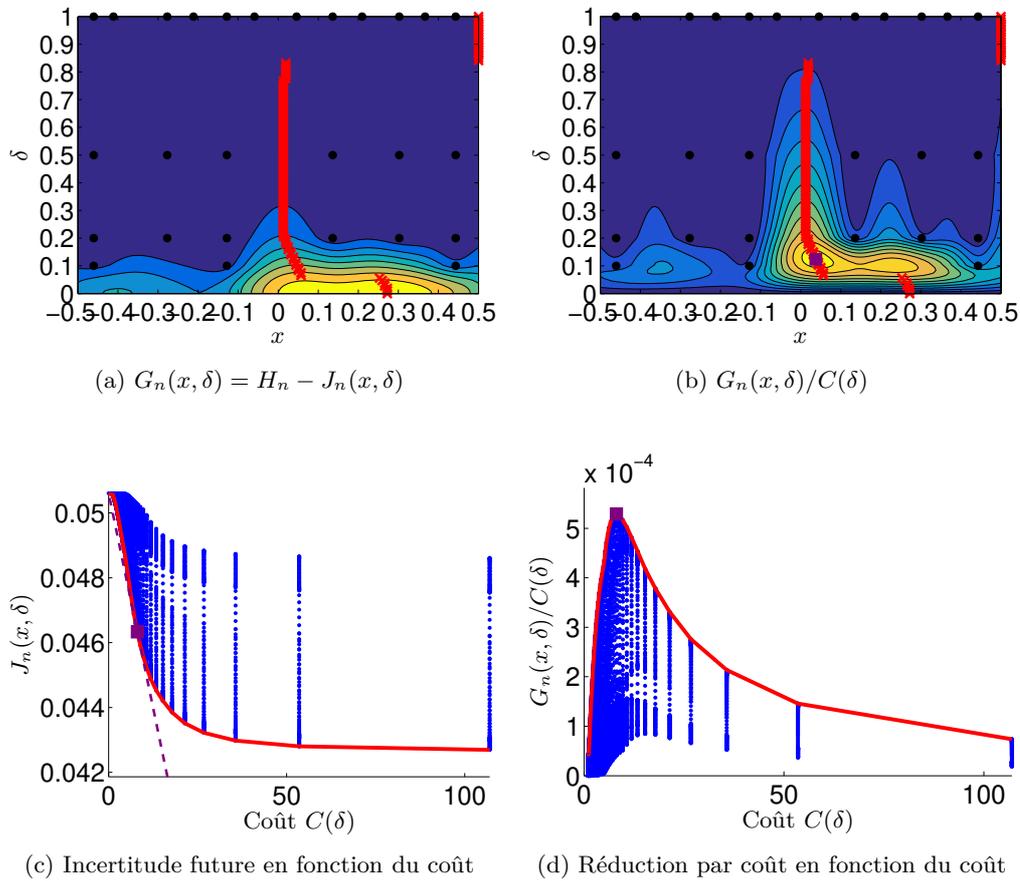


FIGURE 36 – Un exemple en simulation de réduction d’incertitude en multi-fidélité. (a, b) Valeurs des critères dans l’espace des entrées $\mathbb{X} \times \Delta$. Le niveau $\delta = 0$ est le plus haut niveau de fidélité. Les points sont des observations, les croix sont des observations Pareto optimal dans l’espace des objectifs (C, J) , le carré violet indique le point avec le meilleur rapport réduction d’incertitude sur coût. (c, d) Valeurs des critères dans l’espace des objectifs. Les points sont les valeurs des critères sur les différents points de la grille, le trait rouge est le front de Pareto de l’espace (C, J) , le carré violet est le point avec le meilleur rapport réduction d’incertitude sur coût.

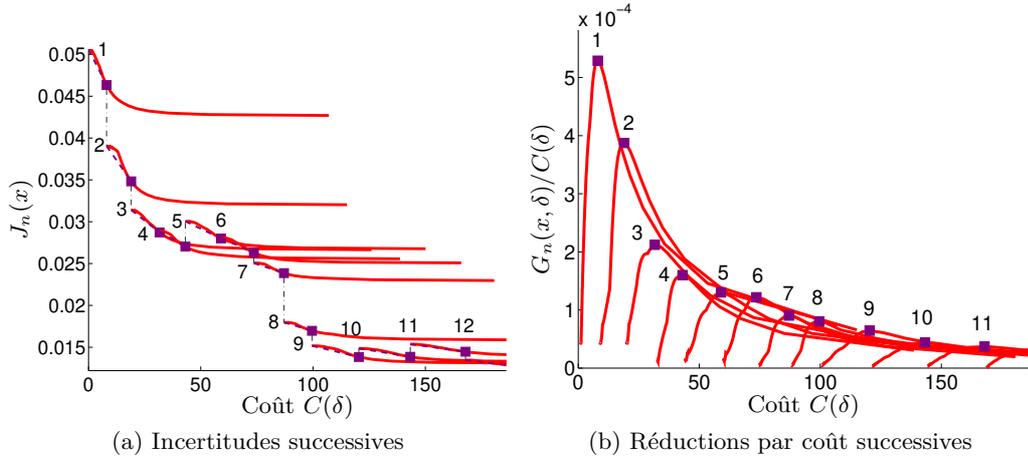


FIGURE 37 – Un exemple en simulation de plan séquentiel d'expériences en multi-fidélité. Les numéros indiquent l'ordre des expériences et les carrés mauves les points sélectionnés par MRUR (a) Fronts successifs de Pareto dans l'espace des objectifs (C, J) . (b) Fronts successifs de Pareto dans l'espace $(C, G/C)$.

Remarque 6. Dans le cas où le coût d'observation est aléatoire, la solution la plus naturelle serait de maximiser l'espérance du rapport $\mathbb{E}_n [(H_n - H_{n+1}) / C(x, \delta)]$. Dans le cas où $H_n - H_{n+1}$ et C sont indépendants, il est possible de réécrire l'espérance de comme $\mathbb{E}_n [H_n - H_{n+1}] \cdot \mathbb{E}_n [1/C(x, \delta)]$. Si, de plus, la variance du coût est négligeable devant le carré de l'espérance, $\text{var}_n [C(x, \delta)] \ll \mathbb{E}_n [C(x, \delta)]^2$ alors le critère MRUR peut être approché par le rapport des espérances

$$(x_{n+1}, \delta_{n+1}) = \arg \max_{x \in \mathbb{X}, \delta \in \Delta} \left\{ \frac{H_n - \mathbb{E}_n [H_{n+1} | X_{n+1} = x, D_{n+1} = \delta]}{\mathbb{E}_n [C(x, \delta) | X_{n+1} = x, D_{n+1} = \delta]} \right\}. \quad (2.47)$$

L'aléa du coût C peut provenir soit d'un aspect stochastique du coût, c'est-à-dire que deux observations au même point n'auront pas le même coût, soit car le coût C est inconnu et modélisé par un processus aléatoire. Consultez par exemple Snoek et al. [2012].

2.4.3 Illustrations

Dans cette section, nous illustrons la stratégie MRUR à travers deux exemples académiques. Sur chaque exemple, la même méthodologie est appliquée. Chaque simulateur multi-fidèle possède S niveaux de fidélité observables. Le niveau de plus haute-fidélité

est indexé par $S > 1$, celui de plus bas par un. Le DoE initial est un plan maximin NLHS tiré avec la méthode décrite à la fin de la section 2.2.2. Nous effectuons les prédictions des sorties, ainsi que l'estimation des probabilités de dépassement de seuil grâce à une approche complètement bayésienne avec intégration des hyper-paramètres. Les intégrales sont approchées avec un Monte-Carlo par Chaînes de Markov (MCMC) sur cent hyper-paramètres. L'échantillonnage initial des hyper-paramètres est obtenu par l'algorithme Métropolis adaptatif de Haario et al. [2001] Les observations sont ajoutées une par une, et les hyper-paramètres sont mis à jour avec un SMC [Chopin, 2002]. Cependant, pour des raisons de temps de calcul, les critères SUR et MRUR sont calculés en supposant les hyper-paramètres connus et fixés à leur MAP. Sur les deux exemples, un seuil critique z^{crit} et une fonction de coût C sont donnés sur chaque exemple. La fonction de coût ne dépend que du niveau δ , elle est indépendante du point x . Finalement, le critère utilisé est le critère SUR 1.

Sur chaque expérience, nous comparons des stratégies SUR à la stratégie MRUR. Quand une stratégie SUR est utilisée sur un niveau s donné, tous les nouveaux points sont ajoutés sur le niveau en question, la position x étant choisie par minimisation du critère d'échantillonnage :

$$\delta_{n+1} = s, \quad x_{n+1} = \arg \min_{x \in \mathbb{X}} \{\mathbb{E}_n [H_{n+1} | X_{n+1} = x, D_{n+1} = s, \theta = \theta^{\text{map}}]\} \quad (2.48)$$

La stratégie MRUR choisit le couple point-niveau maximisant le critère (2.46). Chaque expérience est répétée un certain nombre de fois. Quand une courbe est présentée, elle représente la médiane des résultats obtenus sur toutes les expériences.

Exemple à deux niveaux

La première illustration reprend l'exemple multi-fidèle à deux niveaux proposé par Forrester et al. [2007], déjà présenté section 1.2.1. Ce simulateur est un simulateur déterministe à deux niveaux, $S = 2$, définis sur $\mathbb{X} = [0; 1] \subset \mathbb{R}^1$. Nous fixons le coût d'observation à un quart pour le premier niveau et à un pour le second : $C(1) = 1/4$, $C(2) = 1$. De même, nous fixons un seuil critique à $z^{\text{crit}} = 10$. Notre objectif est d'estimer la zone de dépassement de seuil, $\{x \in \mathbb{X}, \xi_S(x) \geq z^{\text{crit}}\}$. Le modèle bayésien est le même que dans la section 1.2.1. Le plan d'expériences initial est un NLHS maximin constitué de six points sur le premier niveau et trois sur le second, pour un coût initial de $6 \times 1/4 + 3 \times 1 = 4.5$.

Pour estimer la zone de dépassement de seuil, la fonction cible (quantité d'intérêt) est l'indicatrice $\mathbb{1}_{\xi_S(x) \geq z^{\text{crit}}}$. L'estimateur est $p_n(x) = \mathbb{E}_n [\mathbb{1}_{\xi_S(x) \geq z^{\text{crit}}}]$. Les stratégies de

planification se fondent sur le critère SUR 1 avec $\lambda = 0$ car le simulateur est déterministe.

Les figures 38 et 39 présentent un exemple de planification séquentielle sur le budget donné. Les points d'observation sont choisis successivement par le critère MRUR. Les figures de gauche (prédictions du niveau de haute-fidélité et observations), montrent que les observations s'accumulent là où le niveau de haute-fidélité est proche du seuil, particulièrement entre 0.2 et 0.4. Sur cet exemple, le nombre d'observations supplémentaires en basse-fidélité est de $26 - 6 = 20$, celui en haute-fidélité est de $7 - 3 = 4$. Il y a donc cinq fois plus de nouveaux points en basse-fidélité qu'en haute-fidélité. Les figures du milieu montrent que l'estimation de la fonction de probabilité de défaillance tend vers sa valeur vraie, et que les intervalles de crédibilité tendent à se réduire. Les figures de droite montrent la valeur du critère MRUR en chaque point (x, s) . Même si les maximums des deux courbes respectives à chaque niveau ne sont pas au même point x , les valeurs des deux critères tendent à avoir des allures similaires. Cela indique que si un point x est intéressant à observer à un niveau donné selon le critère MRUR, alors ce point x sera aussi intéressant à observer sur l'autre niveau de fidélité.

La figure 40 compare trois stratégies de planification : la stratégie SUR au niveau de basse-fidélité, la stratégie SUR sur le niveau de haute-fidélité et la stratégie MRUR. L'expérience est répétée pour 60 plans initiaux différents. Sur une répétition donnée, le même plan initial multi-fidèle est utilisé pour les trois stratégies séquentielles.

La figure 40 montre que la stratégie MRUR est meilleure que les deux autres SUR. La figure 40a compare l'erreur entre la vraie fonction indicatrice et son estimation par une méthode complètement bayésienne. Cette figure montre d'abord que la stratégie SUR au niveau basse-fidélité est inefficace et n'apporte aucun gain d'information. Ensuite, elle montre surtout que à coût égal, la stratégie MRUR est meilleure que la stratégie SUR au niveau haute-fidélité. Ceci est confirmé avec la figure 40b qui montre la proportion d'expériences où MRUR est meilleure que SUR au niveau haute-fidélité. Cette figure montre que, dans plus de 85 % des cas, la stratégie MRUR est plus efficace que la stratégie SUR au niveau haute-fidélité. De plus, à partir de la deuxième observation haute-fidélité, coût total supérieur à 6,5, la stratégie MRUR est strictement meilleure que la stratégie SUR au dernier niveau pour plus de 50 % des expériences. Par conséquent, même si observer uniquement le niveau de plus basse-fidélité est inefficace, il est possible d'extraire de l'information utile de ce niveau.

La figure 41 représente la répartition entre les niveaux des observations choisies par la stratégie MRUR. La figure 41a représente la proportion de stratégies MRUR choisissant d'observer n_2 fois le plus haut niveau de fidélité en fonction de n_2 . La stratégie SUR premier niveau correspond à $n_2 = 0$, la stratégie SUR second niveau à $n_2 = 9$ et la

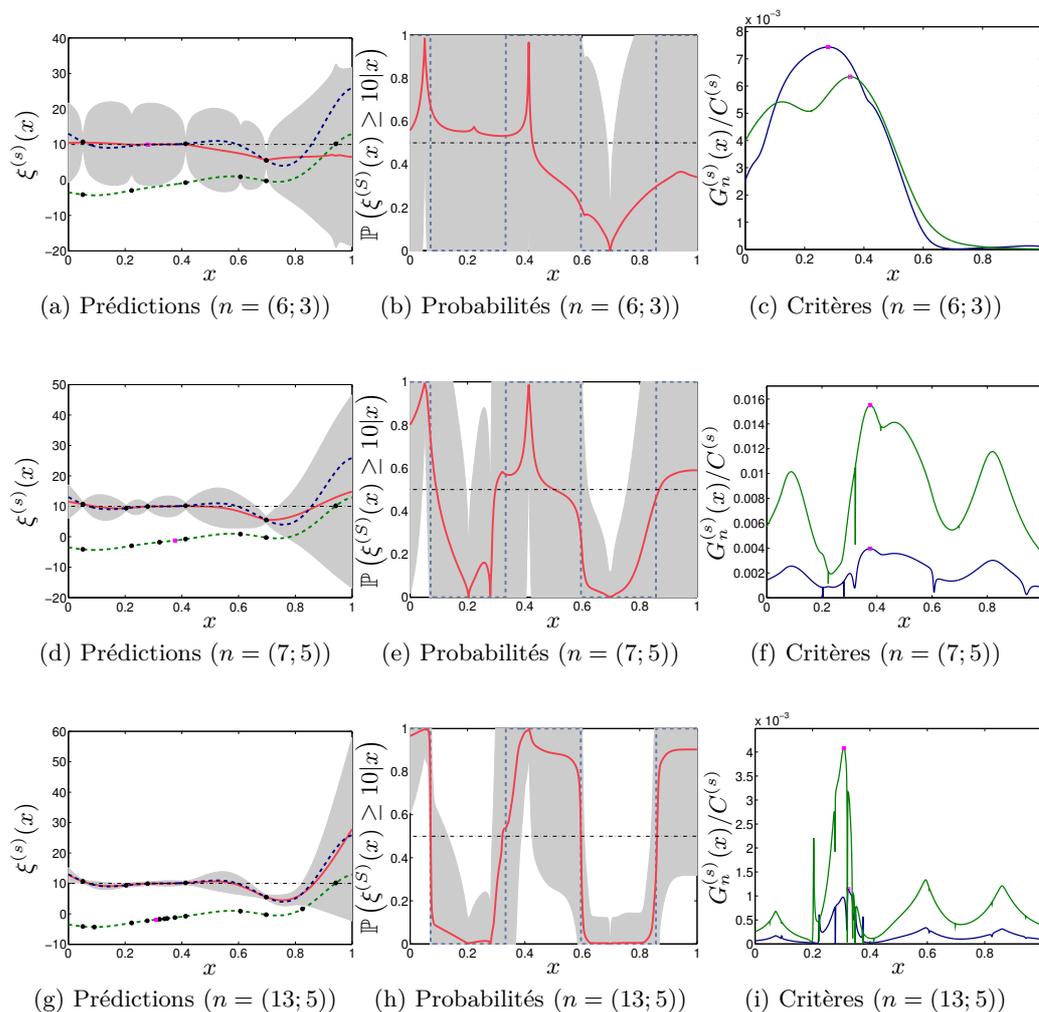


FIGURE 38 – Première partie d’une planification séquentielle d’expérience sur deux niveaux de fidélité (seconde partie figure 39). Le niveau haute-fidélité est quatre fois plus coûteux que le niveau basse-fidélité. Chaque ligne correspond à un instant différent (les nombres sous les figures sont les nombres d’observations sur chaque niveau), chaque colonne à un résultat. (a, d, g) Estimation du niveau haute-fidélité. Le trait plein rouge est la moyenne a posteriori, la zone grise est l’intervalle de crédibilité à deux écarts-types. Les traits pointillés sont les deux niveaux de fidélité (bleu en haut, le haut-niveau ; vert en bas, le bas niveau), le trait pointillé vertical correspond au seuil critique. Les points noirs sont les différentes observations sur les deux niveaux, le carré mauve est la prochaine observation. (b, e, h) Estimation de la probabilité que le niveau de haute fidélité dépasse son seuil critique en fonction de l’entrée x . Le trait plein rouge est la moyenne a posteriori, la zone grise est l’intervalle de crédibilité à deux écarts-types, le trait pointillé bleu est la valeur vraie. (c, f, i) Valeur du critère MRUR sur les deux niveaux en fonction de l’entrée x . Le trait bleu correspond à la valeur critère sur le haut-niveau de fidélité et le trait vert au critère sur le bas-niveau. Les carrés mauves sont les points qui maximisent le critère sur chaque niveau de fidélité.

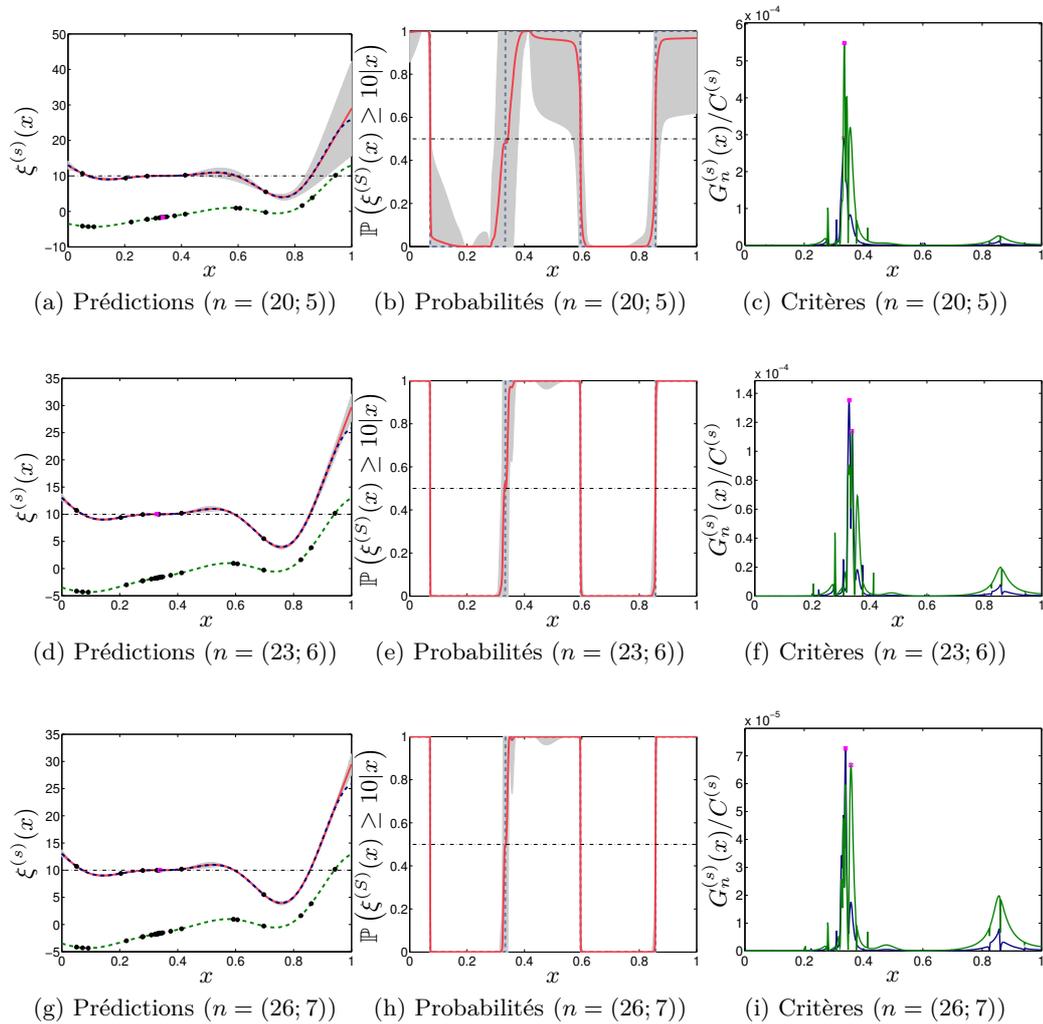


FIGURE 39 – Seconde partie d’une planification séquentielle d’expérience sur deux niveaux de fidélité (première partie figure 38). Le niveau haute-fidélité est quatre fois plus coûteux que le niveau basse-fidélité. Chaque ligne correspond à un instant différent (les nombres sous les figures sont les nombres d’observations sur chaque niveau), chaque colonne à un résultat. (a, d, g) Estimation du niveau haute-fidélité. Le trait plein rouge est la moyenne a posteriori, la zone grise est l’intervalle de crédibilité à deux écarts-types. Les traits pointillés sont les deux niveaux de fidélité (bleu en haut, le haut-niveau ; vert en bas, le bas niveau), le trait pointillé vertical correspond au seuil critique. Les points noirs sont les différentes observations sur les deux niveaux, le carré mauve est la prochaine observation. (b, e, h) Estimation de la probabilité que le niveau de haute fidélité dépasse son seuil critique en fonction de l’entrée x . Le trait plein rouge est la moyenne a posteriori, la zone grise est l’intervalle de crédibilité à deux écarts-types, le trait pointillé bleu est la valeur vraie. (c, f, i) Valeur du critère MRUR sur les deux niveaux en fonction de l’entrée x . Le trait bleu correspond à la valeur critère sur le haut-niveau de fidélité et le trait vert au critère sur le bas-niveau. Les carrés mauves sont les points qui maximisent le critère sur chaque niveau de fidélité.

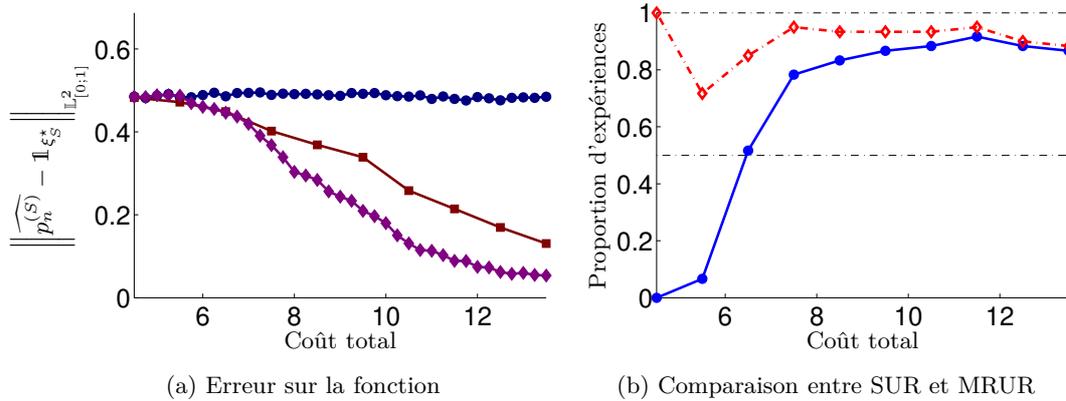


FIGURE 40 – Comparaison entre des stratégies SUR et la stratégie MRUR sur l'exemple de Forrester et al. [2007]. (a) Médiane sur les expériences des erreurs obtenues avec chacune des stratégies en fonction du coût dépensé. La stratégie SUR au niveau de basse-fidélité correspond à la courbe bleue marquée par des cercles, la stratégie SUR au niveau de haute-fidélité à la courbe rouge marquée par des carrés et la stratégie MRUR à la courbe violette marquée par des losanges. (b) Proportion d'expériences où la stratégie MRUR est meilleure que la stratégie SUR au niveau haute-fidélité en fonction du coût. La courbe en rouge du haut est la proportion d'expérience où la stratégie MRUR est supérieure ou égale à la stratégie SUR, celle en bleu du bas est la proportion où la stratégie MRUR est strictement meilleure que la stratégie SUR.

méthode de Xiong et al. [2013], qui recommandent d'observer deux fois plus le premier niveau que le second, à $n_2 = 6$. En moyenne, la stratégie MRUR observe $n_2 = 3,9 \pm 1,8$ fois le niveau de plus haute-fidélité. De plus, la stratégie MRUR observe au moins deux fois le niveau haute-fidélité, et il n'y a qu'une seule expérience qui observe uniquement le niveau haute-fidélité. La stratégie MRUR tend à trouver un compromis entre les deux stratégies SUR.

La figure 41b est une représentation des soixante expériences en fonction du coût total. Chaque ligne correspond à une expérience. La couleur à un coût donnée représente le niveau qui est en train d'être observé. Cette figure montre que la stratégie MRUR tend à commencer par observer le second niveau une ou deux fois, puis à observer principalement le premier niveau. Ceci est confirmé par les figures 41c et 41d. La planification est décomposée en phases de durée égale, une phase durant le temps nécessaire pour faire une observation au plus haut niveau de fidélité. Sur les deux premières phases de la planification, les coûts d'observations sont principalement dépensés sur le niveau haute-fidélité, alors que sur les sept suivantes, le niveau principalement observé est le niveau basse-fidélité. Cependant, durant chacune des phases, il y a toujours quelques répétitions d'expériences qui choisissent d'observer le niveau de plus haute-fidélité. Ce comportement peut s'expliquer par le fait que les deux fonctions sont liées par une relation simple, $f_2(x) = 2f_1(x) + (20 - 20x)$. L'algorithme commence par placer des points en haute-fidélité pour bien apprendre la relation entre les deux niveaux. Puis, quand la relation est bien connue, toutes les nouvelles observations sont placées sur le niveau basse-fidélité moins coûteux, car les deux niveaux apportent la même quantité d'information.

Oscillateur harmonique aléatoire amorti

Le second exemple étudié est l'oscillateur harmonique aléatoire amorti présenté section 1.3.3. Ce simulateur est un simulateur stochastique comprenant un paramètre de fidélité δ compris entre 0 et 1 s, où 0 s est le niveau de plus haute fidélité et 1 s le plus bas. L'entrée x est de dimension deux, et est à valeur dans $\mathbb{X} = [0; 30] \text{ rad} \cdot \text{s}^{-1} \times [0; 1]$. Dans un souci de comparaison avec une valeur de référence, le niveau étudié est $\delta^{\text{ref}} = 0,01 \text{ s}$. L'objectif est d'estimer la fonction de probabilité de défaillance $p(x, \delta^{\text{ref}}) = \mathbb{P}(Z > z^{\text{crit}} | x, \delta^{\text{ref}})$, où le seuil critique est $z^{\text{crit}} = -3$. En observant la figure 24, le contour $\xi(x, \delta) = z^{\text{crit}}$ évolue beaucoup entre le bas et le haut niveau. La fonction de coût est une fonction affine en l'inverse du paramètre de fidélité : $C(\delta) = (a/\delta) + b$, où $a = 0.0098$ et $b = 0.0208$. Les coefficients sont normalisées pour avoir $C(\delta^{\text{ref}}) = 1$. Le modèle bayésien est le même que dans la section 1.3.3.

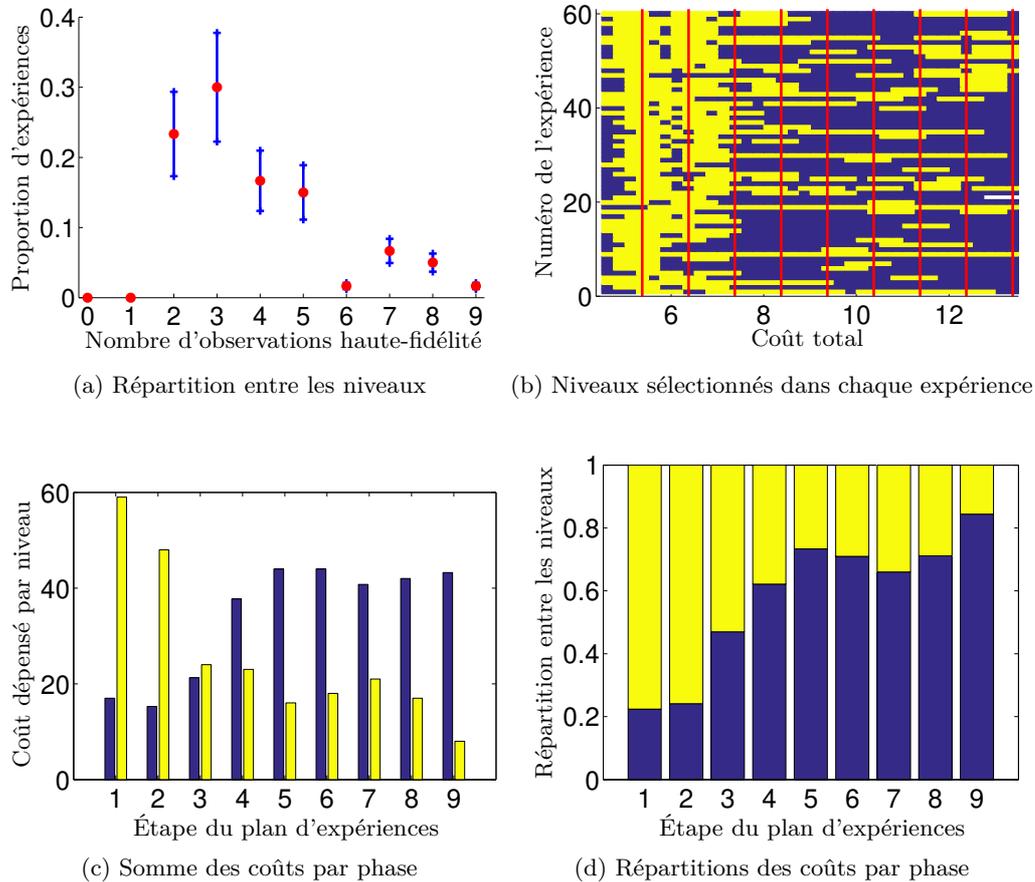


FIGURE 41 – Analyse des plans d’expériences obtenus par la stratégie MRUR sur l’exemple à deux niveaux. (a) Histogramme de la proportion d’expériences observant n fois le niveau de plus haute fidélité. Les points sont les proportions obtenues, les barres sont les intervalles à deux écarts-types. Sur (b), (c) et (d), le bleu correspond au bas niveau de fidélité et le jaune au haut niveau de fidélité. (b) Représentation des niveaux sélectionnés en fonction du coût. Chaque ligne correspond à une expérience, chaque couleur à un niveau. Les barres verticales indiquent les différentes phases du plan. (c) Somme sur toutes les expériences des coûts dépensés sur chaque niveau durant chaque phase. Les sommes affichées sont les sommes sur toutes les expériences. (d) Proportions des coûts dépensés sur chaque niveau durant chaque phase.

Niveau δ	1 s	0,5 s	1/3 s	0,25 s	0,2 s	1/6 s	0,1 s	0,05 s	0,02 s	0,01 s
$C(\delta)^{-1}$	32.7	24.8	19.9	16.7	14.3	12.6	8.4	4.6	2	1
Plan initial	180	60	20	10	5	0	0	0	0	0

TABLE 1 – Les différents niveaux de fidélité, avec leurs coûts inverses et le nombre d'observation à chaque niveau dans le plan initial

Dans cette expérience, nous ne considérons que dix niveaux de fidélité. Le plan initial est un NLHS maximin, dont le nombre de points par niveau est présenté tableau 1. Le coût total du plan initial est de 9,88. Le budget accordé est de vingt, ce qui laisse un budget supplémentaire pour le plan séquentiel de 10,12.

Les figures 42 à 47 présentent une planification séquentielle d'expérience sur cet exemple. Chaque figure correspond à un instant de la planification. La somme des coûts des observations de chaque figure est indiqué dans sa légende, la différence de budget dépensé entre chaque figure étant environ égale à deux. Les figures (c) montrent que d'observations en observations, le niveau de contour à 50 % tend vers sa valeur de référence, mais que par contre, les niveaux de contours à 25 % et à 75 % restent loin de leurs valeurs de référence. Cela est dû au fait que la moyenne ξ est bien estimée, mais que par contre, la valeur de la variance de l'effet stochastique λ est sur-estimée. Les figures (d) montrent que les écart-types tendent à se réduire sur tous les points de l'espace. Les figures (e) montrent que sur cet exemple, les observations s'accumulent sur les niveaux de fidélité correspondant à $\delta = 0,05$ s et $\delta = 0,02$ s. Enfin, les figures (f) représentent l'évolution de la réduction d'incertitude en fonction du coût d'observation du niveau.

Étrangement, sur certaines figures, la réduction d'incertitude attendue après une observation au dernier niveau est inférieure à celle du niveau précédent. Ce résultat est peut-être causé par deux phénomènes. D'une part, cela peut provenir des approximations nécessaires pour calculer le critère SUR (Monte-Carlo, algorithme optimisation) amplifié par le fait que les valeurs optimales de réduction d'incertitude sur les deux derniers niveaux sont assez proches l'une de l'autre. D'autre part, cela peut être dû à la fonction de covariance utilisée pour modéliser la fonction d'erreur ϵ (voir section 1.2.2), à savoir un fonction de covariance séparable entre x et δ .

Dans cette section, cinq stratégies SUR sont comparées à la stratégie MRUR. Chaque stratégie SUR est associée à un niveau d'observation donné. L'expérience est répétée 48 fois. A chaque répétition, un nouveau plan initial multi-fidélité est tiré. A une répétition donnée, le même plan initial est utilisé à l'initialisation de toutes les planifications séquentielles. Le critère de comparaison est l'erreur moyenne quadratique entre l'estima-

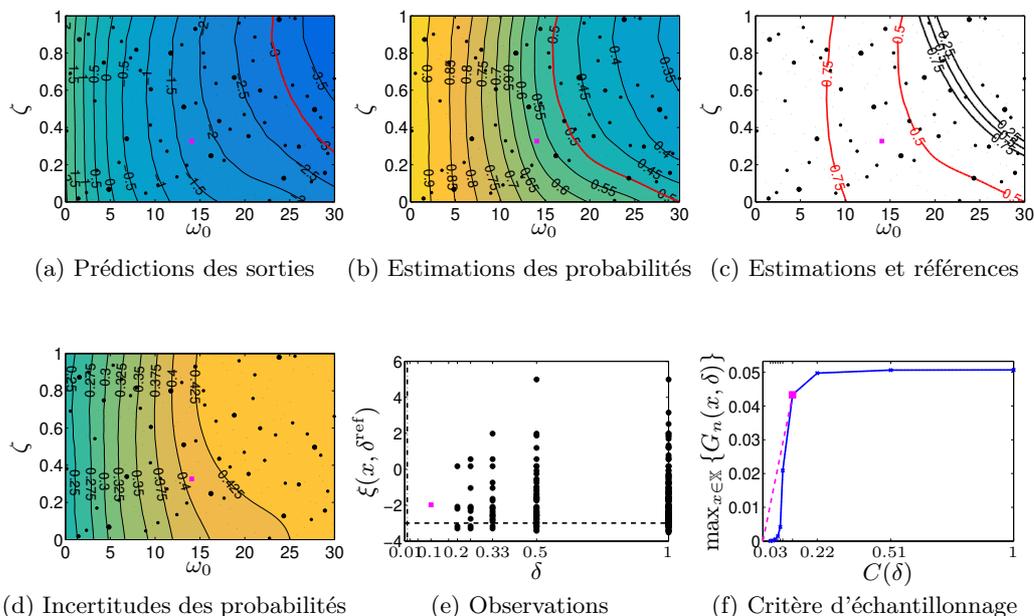


FIGURE 42 – Une étape d’une planification séquentielle sur dix niveaux (partie 1/6, suivant figure 43). Le coût des observations est de $C = 9.88$. Sur toutes les figures, les points noirs correspondent aux observations courantes et le carré mauve est la future observations. Sur les figures représentant des contours dans l’espace \mathbb{X} , la taille des points indique le niveau d’observation : un point plus gros correspond à un plus haut niveau de fidélité. (a) Prédictions des sorties Z par moyenne a posteriori. Le trait rouge est le seuil critique z^{crit} . (b) Estimations des probabilités $\hat{p}_n(\cdot, \delta^{\text{ref}})$ par moyenne a posteriori. Le trait rouge est le seuil de 50%. (c) Comparaison entre l’estimation $\hat{p}_n(\cdot, \delta^{\text{ref}})$ (contours rouges) et la valeur de référence $p^*(\cdot, \delta^{\text{ref}})$ (contours noirs). (d) Écart-type a posteriori des probabilités $p(x)$. (e) Positionnement des observations $(x_i, \delta_i)_{1 \leq i \leq n+1}$. L’abscisse est le niveau de fidélité du point d’observation δ_i , l’ordonnée est la valeur de la partie moyenne associée à l’entrée x_i au niveau de référence δ^{ref} . Le trait pointillé vertical est le seuil critique. (f) Réduction d’incertitude en fonction du coût.

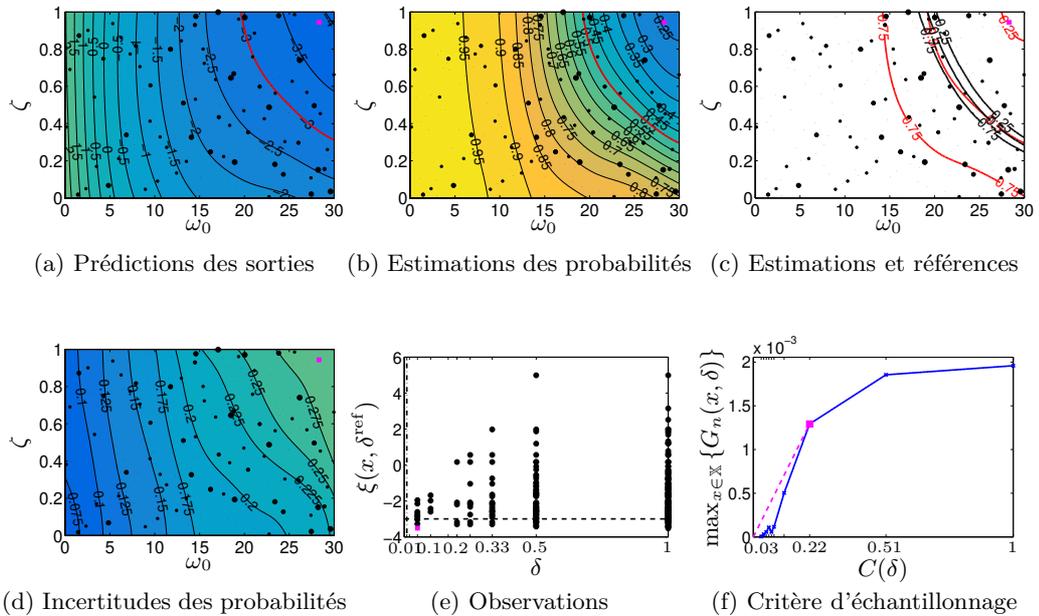


FIGURE 43 – Une étape d'une planification séquentielle sur dix niveaux (partie 2/6, précédent figure 42, suivant figure 44). Le coût des observations est de $C = 12.08$. Sur toutes les figures, les points noirs correspondent aux observations courantes et le carré mauve est la future observations. Sur les figures représentant des contours dans l'espace \mathbb{X} , la taille des points indique le niveau d'observation : un point plus gros correspond à un plus haut niveau de fidélité. (a) Prédictions des sorties Z par moyenne a posteriori. Le trait rouge est le seuil critique z^{crit} . (b) Estimations des probabilités $\hat{p}_n(\cdot, \delta^{\text{ref}})$ par moyenne a posteriori. Le trait rouge est le seuil de 50%. (c) Comparaison entre l'estimation $\hat{p}_n(\cdot, \delta^{\text{ref}})$ (contours rouges) et la valeur de référence $p^*(\cdot, \delta^{\text{ref}})$ (contours noirs). (d) Écart-type a posteriori des probabilités $p(x)$. (e) Positionnement des observations $(x_i, \delta_i)_{1 \leq i \leq n+1}$. L'abscisse est le niveau de fidélité du point d'observation δ_i , l'ordonnée est la valeur de la partie moyenne associée à l'entrée x_i au niveau de référence δ^{ref} . Le trait pointillé vertical est le seuil critique. (f) Réduction d'incertitude en fonction du coût.

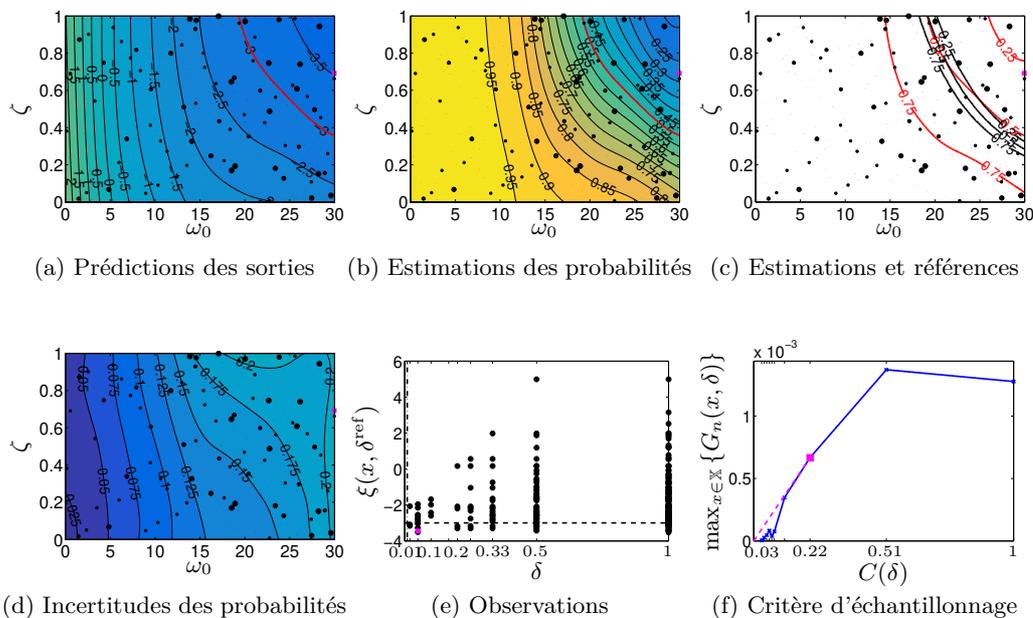


FIGURE 44 – Une étape d'une planification séquentielle sur dix niveaux (partie 3/6, précédent figure 43, suivant figure 45). Le coût des observations est de $C = 14.04$. Sur toutes les figures, les points noirs correspondent aux observations courantes et le carré mauve est la future observations. Sur les figures représentant des contours dans l'espace \mathbb{X} , la taille des points indique le niveau d'observation : un point plus gros correspond à un plus haut niveau de fidélité. (a) Prédictions des sorties Z par moyenne a posteriori. Le trait rouge est le seuil critique z^{crit} . (b) Estimations des probabilités $\hat{p}_n(\cdot, \delta^{\text{ref}})$ par moyenne a posteriori. Le trait rouge est le seuil de 50%. (c) Comparaison entre l'estimation $\hat{p}_n(\cdot, \delta^{\text{ref}})$ (contours rouges) et la valeur de référence $p^*(\cdot, \delta^{\text{ref}})$ (contours noirs). (d) Écart-type a posteriori des probabilités $p(x)$. (e) Positionnement des observations $(x_i, \delta_i)_{1 \leq i \leq n+1}$. L'abscisse est le niveau de fidélité du point d'observation δ_i , l'ordonnée est la valeur de la partie moyenne associée à l'entrée x_i au niveau de référence δ^{ref} . Le trait pointillé vertical est le seuil critique. (f) Réduction d'incertitude en fonction du coût.

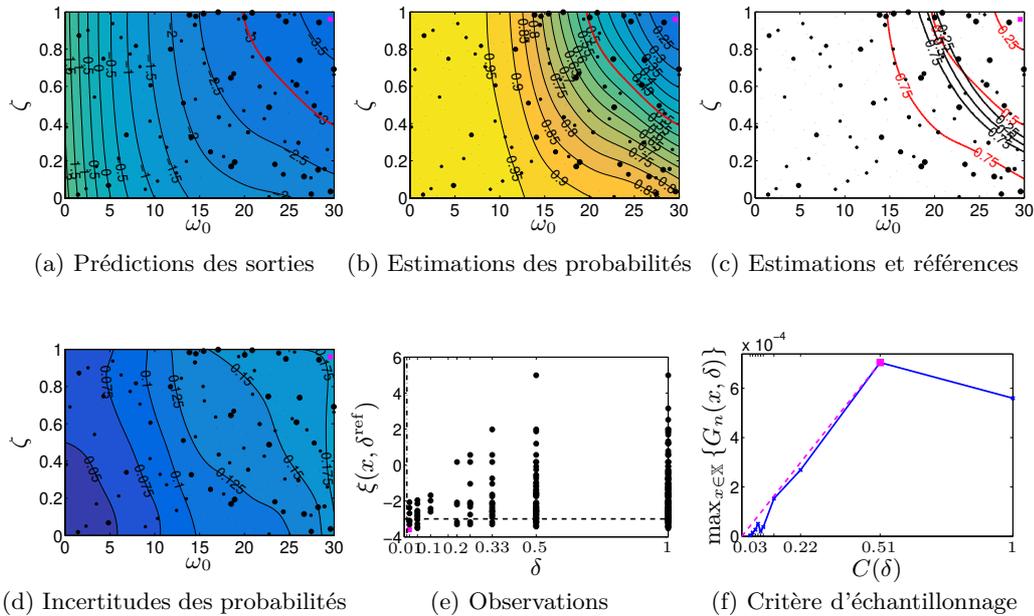


FIGURE 45 – Une étape d'une planification séquentielle sur dix niveaux (partie 4/6, précédent figure 44, suivant figure 46). Le coût des observations est de $C = 16.3$. Sur toutes les figures, les points noirs correspondent aux observations courantes et le carré mauve est la future observations. Sur les figures représentant des contours dans l'espace \mathbb{X} , la taille des points indique le niveau d'observation : un point plus gros correspond à un plus haut niveau de fidélité. (a) Prédictions des sorties Z par moyenne a posteriori. Le trait rouge est le seuil critique z^{crit} . (b) Estimations des probabilités $\hat{p}_n(\cdot, \delta^{\text{ref}})$ par moyenne a posteriori. Le trait rouge est le seuil de 50%. (c) Comparaison entre l'estimation $\hat{p}_n(\cdot, \delta^{\text{ref}})$ (contours rouges) et la valeur de référence $p^*(\cdot, \delta^{\text{ref}})$ (contours noirs). (d) Écart-type a posteriori des probabilités $p(x)$. (e) Positionnement des observations $(x_i, \delta_i)_{1 \leq i \leq n+1}$. L'abscisse est le niveau de fidélité du point d'observation δ_i , l'ordonnée est la valeur de la partie moyenne associée à l'entrée x_i au niveau de référence δ^{ref} . Le trait pointillé vertical est le seuil critique. (f) Réduction d'incertitude en fonction du coût.

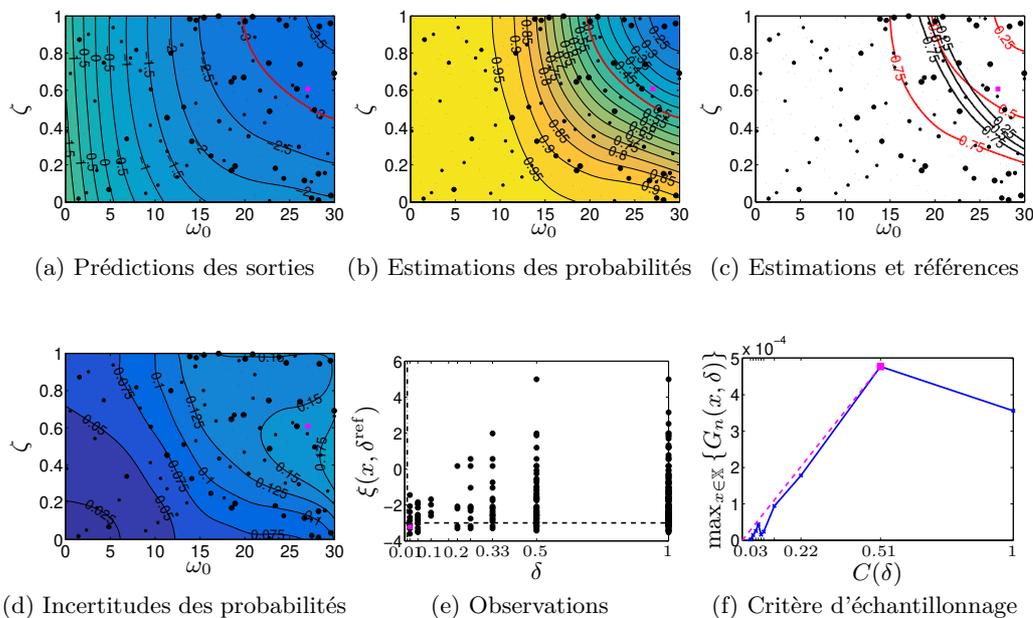


FIGURE 46 – Une étape d'une planification séquentielle sur dix niveaux (partie 5/6, précédent figure 45, suivant figure 47). Le coût des observations est de $C = 18.06$. Sur toutes les figures, les points noirs correspondent aux observations courantes et le carré mauve est la future observations. Sur les figures représentant des contours dans l'espace \mathbb{X} , la taille des points indique le niveau d'observation : un point plus gros correspond à un plus haut niveau de fidélité. (a) Prédictions des sorties Z par moyenne a posteriori. Le trait rouge est le seuil critique z^{crit} . (b) Estimations des probabilités $\hat{p}_n(\cdot, \delta^{\text{ref}})$ par moyenne a posteriori. Le trait rouge est le seuil de 50%. (c) Comparaison entre l'estimation $\hat{p}_n(\cdot, \delta^{\text{ref}})$ (contours rouges) et la valeur de référence $p^*(\cdot, \delta^{\text{ref}})$ (contours noirs). (d) Écart-type a posteriori des probabilités $p(x)$. (e) Positionnement des observations $(x_i, \delta_i)_{1 \leq i \leq n+1}$. L'abscisse est le niveau de fidélité du point d'observation δ_i , l'ordonnée est la valeur de la partie moyenne associée à l'entrée x_i au niveau de référence δ^{ref} . Le trait pointillé vertical est le seuil critique. (f) Réduction d'incertitude en fonction du coût.

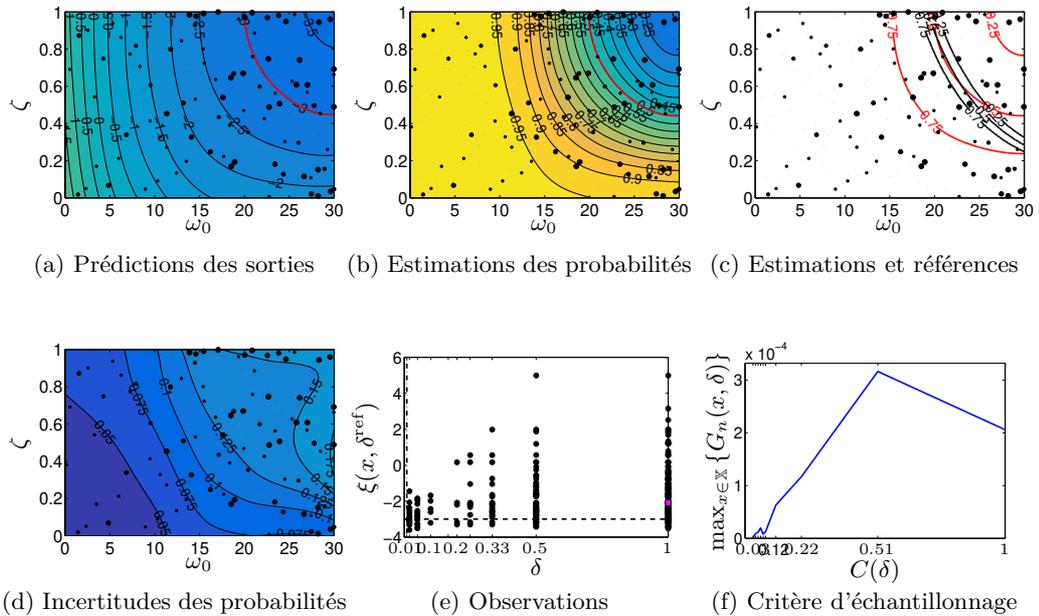


FIGURE 47 – Une étape d'une planification séquentielle sur dix niveaux (partie 6/6, précédent figure 46). Le coût des observations est de $C = 19.98$. Sur toutes les figures, les points noirs correspondent aux observations courantes et le carré mauve est la future observations. Sur les figures représentant des contours dans l'espace \mathbb{X} , la taille des points indique le niveau d'observation : un point plus gros correspond à un plus haut niveau de fidélité. (a) Prédictions des sorties Z par moyenne a posteriori. Le trait rouge est le seuil critique z^{crit} . (b) Estimations des probabilités $\hat{p}_n(\cdot, \delta^{\text{ref}})$ par moyenne a posteriori. Le trait rouge est le seuil de 50%. (c) Comparaison entre l'estimation $\hat{p}_n(\cdot, \delta^{\text{ref}})$ (contours rouges) et la valeur de référence $p^*(\cdot, \delta^{\text{ref}})$ (contours noirs). (d) Écart-type a posteriori des probabilités $p(x)$. (e) Positionnement des observations $(x_i, \delta_i)_{1 \leq i \leq n+1}$. L'abscisse est le niveau de fidélité du point d'observation δ_i , l'ordonnée est la valeur de la partie moyenne associée à l'entrée x_i au niveau de référence δ^{ref} . Le trait pointillé vertical est le seuil critique. (f) Réduction d'incertitude en fonction du coût.

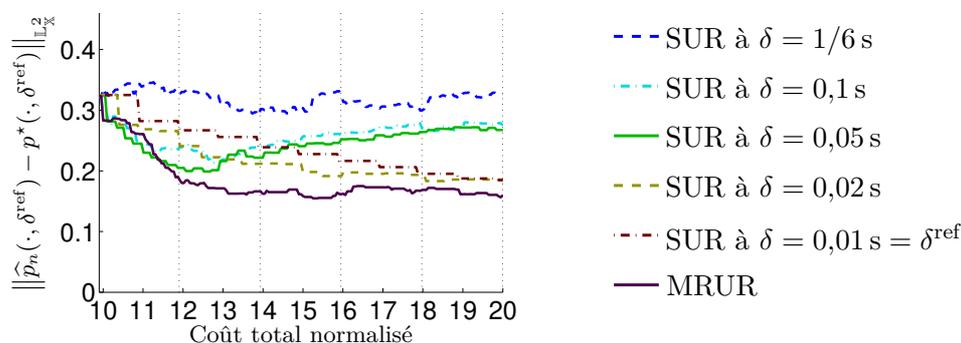


FIGURE 48 – Comparaison entre des stratégies SUR et la stratégie MRUR sur l’oscillateur amorti aléatoire : erreur entre la prédiction par le modèle et la vraie valeur de la fonction de probabilité. Chaque courbe correspond à une stratégie. Le coût est normalisé par le coût d’une observation haute-fidélité à $\delta = 0,01$ s.

tion de la fonction de probabilité \hat{p}_n et la fonction vraie p^* , au niveau δ^{ref} . Les résultats sont présentés figure 48. La figure montre que, dès que le plan d’expériences séquentiel dépense un coût équivalent au coût de deux observations haute-fidélité, la meilleure stratégie est la stratégie MRUR. Parmi les stratégies SUR, la meilleure stratégie est d’abord la stratégie SUR à $\delta = 0,05$ s (huitième niveau considéré) pour les coûts compris entre dix et treize, puis, la stratégie SUR à $\delta = 0,02$ s (neuvième niveau considéré) pour les coûts entre treize et vingt. Il est aussi remarquable sur la figure 48 sur certaines stratégies, l’erreur de prédiction commence par diminuer quand des observations s’ajoutent puis augmente au fur et à mesure. C’est notamment le cas de la stratégie SUR à $\delta = 0,05$ s. A notre avis, cela est dû à un compromis biais-variance qui devient nécessaire si un seul niveau est observé. Au début, l’ajout des observations réduit la variance de la QoI, puis, quand celle-ci devient faible, la prédiction donnée par le modèle tend vers le niveau observé. Le biais qui apparaît alors est lié à la différence entre le niveau réellement observé ($\delta = 0,05$ s) et le niveau étudié ($\delta = 0,01$ s).

La figure 49 permet de représenter les niveaux d’observations choisis par les stratégies MRUR. Ces quatre figures montrent qu’au début sur la première phase, la stratégie MRUR observe surtout le septième et le huitième niveau (respectivement $\delta = 0,1$ s et $0,05$ s), puis, durant les phases suivantes, principalement le neuvième niveau ($\delta = 0,02$ s). Ce résultat suggère que la stratégie MRUR tend à observer le niveau le plus efficace durant la bonne phase (voir figure 48), et donc qu’il arrive à trouver le bon compromis

entre coût et réduction d'incertitude. Les deux premiers niveaux ($\delta = 1$ s et $0,5$ s) sont beaucoup observés mais comme ils sont peu coûteux, ils n'apparaissent pas dans les bilans de coûts. Le dernier niveau ($\delta = 0,01$ s), qui est notre niveau de référence est finalement peu observé. De façon surprenante, les niveaux intermédiaires (3, 4, 5 et 6 ; respectivement $\delta = 1/3$ s, $0,25$ s, $0,2$ s et $1/6$ s) ne sont que peu, voire pas du tout choisis par la stratégie MRUR.

D'autres résultats sont visibles après une observation plus attentive de la figure 49a. Ainsi, la large majorité des expériences commencent par observer le niveau 7 ($\delta = 0,1$ s), puis poursuivent avec le niveau 8 ($\delta = 0,05$ s), puis finissent avec le niveau 9 ($\delta = 0,02$ s). A noter que le plan initial n'observe que les cinq premiers niveaux et qu'il n'y a aucune observation sur le sixième niveau ($\delta = 1/6$ s, voir tableau 1). Sur cette expérience, il y a trois types de comportement qui se dégagent. Le comportement majoritaire (34 expériences sur 48, soit 71 %) consiste à poursuivre la stratégie initiale, c'est-à-dire à observer un niveau, puis progressivement passer au suivant. C'est ce comportement qui donne la prédominance des observations du neuvième niveau ($\delta = 0,02$ s) sur les figures 49c et 49d. Le comportement alternatif (12 expériences sur 48, soit 25 %) consiste à alterner des observations sur le niveau référence ($\delta = 0,01$ s) et des niveaux de très basses fidélité ($\delta = 1$ s ou $0,5$ s). Ainsi, ces expériences se focalisent directement sur le niveau de plus haute-fidélité, en observant de temps en temps les deux premiers niveaux. Les dernières expériences restantes (2 expériences sur 48, soit 4 %) observent le dernier niveau ($\delta = 0,01$ s), puis à partir de la troisième phase, alternent des observations sur les niveaux basse-fidélités ($\delta = 1$ s et $0,5$ s) et le huitième niveau ($\delta = 0,05$ s).

Finalement, ces deux expériences montrent que la stratégie MRUR réussit à trouver un bon compromis entre coût de simulation et réduction d'incertitude.

2.4.4 Extensions pour la parallélisation

Dans certaines situations, il peut être souhaitable de choisir plusieurs points en même temps, par exemple, si les simulations peuvent être lancées en parallèle. Dans ce cas, il devient nécessaire d'adapter la stratégie MRUR au choix de $q > 1$ points d'observations. Dans cette section, le critère SUR utilisé est supposé calculable pour n'importe quelle taille q de proposition de plans d'expérience. C'est le cas pour les critères SUR 1 à 4.

En planification parallèle, Girdziusas et al. [2012] distinguent deux types de plans séquentiels d'expériences : les plans synchrones et les plans asynchrones. Dans un plan synchrone à q cœurs, les points sont sélectionnés par lot de q . Une fois qu'un paquet est lancé, l'algorithme attend que toutes les simulations soient terminées avant d'en choisir

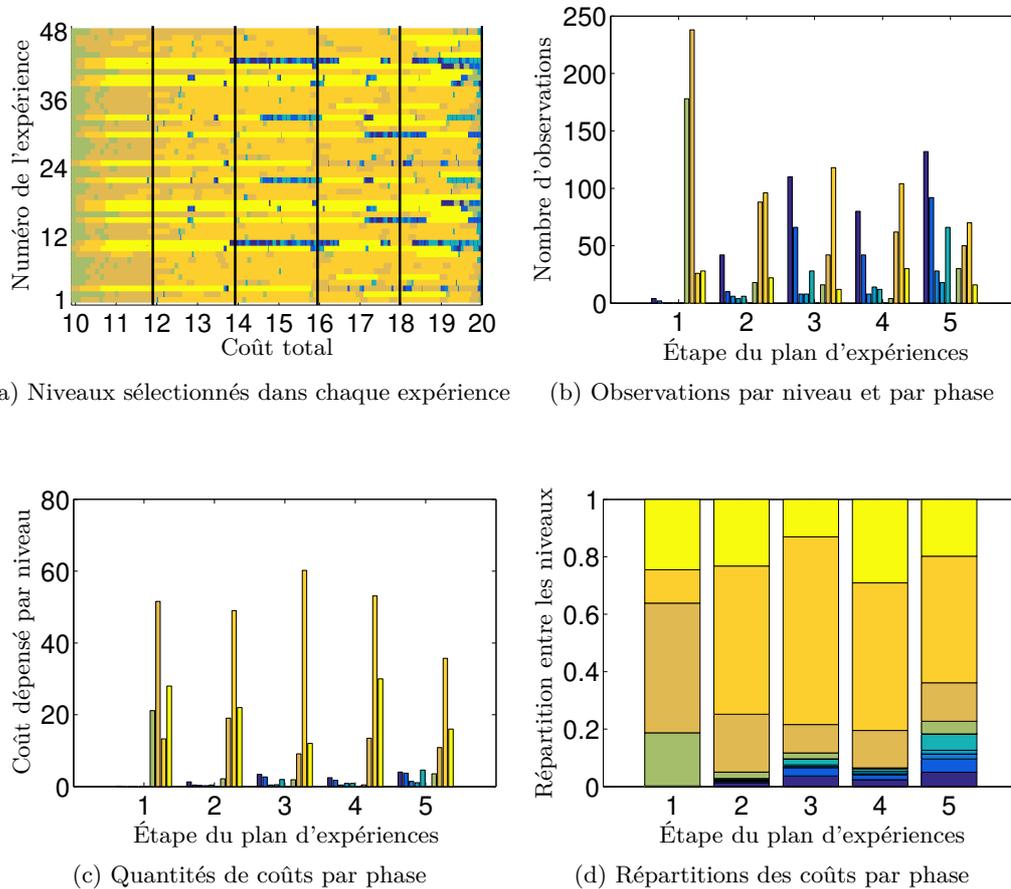


FIGURE 49 – Analyse des plans d’expériences obtenus par une stratégie MRUR sur l’oscillateur amorti aléatoire. Les données sont obtenues avec quarante-huit expériences. Sur toutes les figures, les couleurs sombres correspondent aux bas niveaux de fidélité et les couleurs claires aux hauts niveaux. (a) Représentation des niveaux sélectionnés en fonction du coût. Chaque ligne correspond à une expérience, chaque couleur à un niveau. Les traits verticaux indiquent les différentes phases du plan. (b) Nombre brut d’observation sur chaque niveau durant chaque phase. (c) Quantités de coûts dépensés sur chaque niveau durant chaque phase. Sur (b) et (c), les sommes affichées sont les sommes sur toutes les expériences. (d) Répartitions des coûts dépensés sur chaque niveau durant chaque phase.

des nouvelles. La conséquence est que les q simulations à choisir doivent avoir le même temps de calcul, pour éviter d'avoir un cœur au repos. Les planifications séquentielles d'expériences parallèles synchrones sont plus simples à implémenter. Dans un contexte de parallélisation des expériences numériques, Ginsbourger et al. [2010b] développent un critère d'échantillonnage nommé q -EI pour trouver le minimum d'une fonction. Ils proposent également des approximations de ce critère, tels que le « constant liar » ou le « kriging believer ». Williams et al. [2011] présentent des algorithmes de planifications séquentielles ainsi que des critères d'échantillonnage dédiés à la planification par lot. De la même façon, les critères SUR de Chevalier et al. [2014] dédiés à l'estimation de dépassement de seuil sur des fonctions déterministes ont développé pour des lot de q nouvelles observations. Toujours dans la planification par lot, González et al. [2016] proposent un algorithme simplifiant l'optimisation du critère d'échantillonnage sur plusieurs observations futures. Dans un plan séquentiel asynchrone, chaque fois qu'une simulation s'achève, une autre est lancée immédiatement à la place, sans attendre d'autres résultats. L'intérêt est de pouvoir lancer en même temps plusieurs simulations dont les temps de calcul sont différents, ce qui permet de limiter le temps calculatoire total. L'inconvénient est qu'un algorithme de planification séquentielle d'expériences parallèles asynchrones est plus difficile à implémenter. En particulier, il requiert de pouvoir créer un critère d'échantillonnage J^{busy} qui prend en compte le fait que certains points sont en cours d'observation. De plus, dans le cas où les coûts sont variables, ce critère d'échantillonnage doit aussi prendre en compte les différents coûts de chaque simulation. Ginsbourger et al. [2010a] et Snoek et al. [2012] obtiennent une version de J^{busy} de l'EI en intégrant par Monte-Carlo le critère standard sur les valeurs de sorties attendues. La même procédure est appliquée par Desautels et al. [2014] sur un critère de borne de confiance supérieure.

Il est assez simple d'adapter la stratégie MRUR dans le cas d'une planification synchrone. Pour simplifier, supposons que le coût d'une observation ne dépende que du niveau de fidélité, $C(x, \delta) = C(\delta)$. Dans ce cas, comme toutes les observations doivent avoir le même coût, les q observations doivent être effectuées au même niveau de fidélité, $\delta_{n+1} = \dots = \delta_{n+q}$. Dans ces conditions, le critère MRUR qui vient naturellement consiste à maximiser le rapport entre la réduction d'incertitude après q observations, et

le coût d'observation du niveau

$$G_n^{(q)} \left(\left(x^{(l)}, \delta^{(l)} \right)_{1 \leq l \leq q} \right) = H_n - \mathbb{E}_n \left[H_{n+q} \mid \left(X_{n+l} = x^{(l)}, D_{n+l} = \delta^{(l)} \right)_{1 \leq l \leq q} \right]$$

$$\left(\left(x_{n+l} \right)_{1 \leq l \leq q}, \delta \right) = \arg \max_{\left(x^{(l)} \right)_{1 \leq l \leq q} \in \mathbb{X}^q, \delta \in \Delta} \left\{ \frac{G_n^{(q)} \left(\left(x^{(l)}, \delta \right)_{1 \leq l \leq q} \right)}{C(\delta)} \right\} . \quad (2.49)$$

$$\delta_{n+l} = \delta, \quad 1 \leq l \leq q$$

Afin d'expérimenter la stratégie MRUR en planification parallèle synchrone, nous reprenons l'exemple de l'oscillateur amorti aléatoire présenté section 2.4.3. Le même type d'expérience est considéré avec les mêmes niveaux de fidélité, la même quantité d'intérêt, le même modèle bayésien, le même algorithme de génération du plan initial et le même budget. Quatre stratégies de planification séquentielle sont comparées : la stratégie SUR au niveau de référence avec les points choisis un par un (stratégie SUR classique), la stratégie MRUR avec une planification par paquet de un point (stratégie MRUR illustrée dans la section 2.4.3), une stratégie SUR avec une planification par paquet de $q = 5$ points, et une stratégie MRUR synchrone par paquet de $q = 5$ points appliquant la méthode (2.49). Comme l'optimisation du critère à q points est difficile, une optimisation point par point est appliquée comme dans Chevalier et al. [2014]. Plus précisément, à un niveau donné, nous commençons par choisir le meilleur point du niveau, puis pour chacun des point restants, nous maximisons la réduction d'incertitude totale en laissant figés les points précédents :

$$x^{(l)}(\delta) = \arg \max_{x \in \mathbb{X}} \left\{ G_n^{(l)} \left(\left(x^{(j)}(\delta), \delta \right)_{1 \leq j \leq l-1}, (x, \delta) \right) \right\} . \quad (2.50)$$

Le niveau est choisi une fois que tous les niveaux de fidélité ont fourni une proposition d'observations. L'expérience est répétée 24 fois. Sur chaque expérience, le même plan initial est utilisé pour toutes les stratégies séquentielles.

Les résultats sont présentés figure 50. Les courbes représentent les erreurs pour un coût donné. Sur la figure 50a, le coût affiché en abscisse est le coût total des points individuels : si q points d'observations sont choisis au niveau δ , le coût considéré est $qC(\delta)$. Sur cette échelle de comparaison, les stratégies par lot sont moins efficaces que leurs équivalents planifiant les points un par un. Ce résultat est naturel puisque les plans par lot disposent de moins de « marge de manœuvre » pour s'adapter au simulateur. Cependant, la stratégie MRUR par paquet obtient des résultats tout à fait satisfaisants, comparables à ceux de la stratégie SUR au dernier niveau point par point. A la fin de

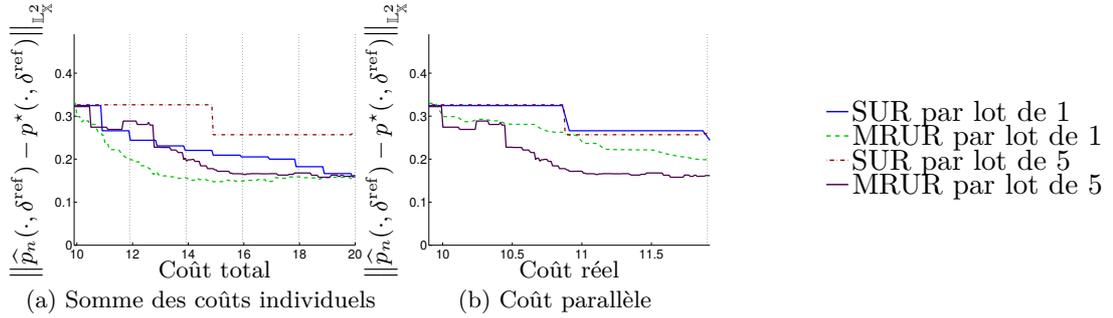


FIGURE 50 – Comparaison entre des stratégies où les observations sont choisies une par une et celles où les observations sont choisies par paquet de cinq. Les courbes représentent l’erreur entre la prédiction par le modèle et la vraie valeur de la fonction de probabilité. Chaque courbe correspond à une stratégie. (a) Erreur en fonction du coût total des observations, c’est-à-dire de la somme de tous les coûts individuels. (b) Erreur en fonction du coût réel après parallélisation.

la planification, les deux stratégies MRUR et la stratégie SUR par paquet de un point donnent des résultats comparables. Sur la figure 50b, les coûts affichés sont en « wall-clock time », c’est-à-dire en temps réel après parallélisation : si q points d’observations sont choisis au niveau δ , le coût considéré est $C(\delta)$. Avec cette échelle de comparaison, à temps de calcul réel identique, la stratégie MRUR par paquet de q points obtient une erreur plus faible que la stratégie MRUR point par point. En observant la situation finale sur la figure 50a et en regardant les échelles de temps sur la figure 50b, il sort sur cet exemple que la planification par MRUR synchrone par paquet de q permet de diviser le temps de calcul parallèle par q pour atteindre le même niveau d’erreur. Il est envisageable d’adapter ce genre de méthode de parallélisation dans le cas où le coût des niveaux sont des multiples entiers. Ainsi, si un niveau basse-fidélité est deux fois moins cher qu’un niveau haute-fidélité, il faudrait comparer la réduction relative de : quatre points basse-fidélité ; deux points basse-fidélité et un point haute-fidélité ; et deux points haute-fidélité. Cette idée est d’ailleurs proposée par Zhou et al. [2017].

Il est par contre relativement délicat d’adapter directement la stratégie MRUR pour des planifications parallèle asynchrones. Une solution simple serait de maximiser le rapport entre la réduction d’incertitude $G_n^{(q)}$ d’un plan d’expériences à q points et la somme des coût des q points du plan. Cependant, nous ne recommandons pas cette stratégie, puisqu’elle peut amener à des comportements indésirables. Supposons une situation où il n’y a que trois observations disponibles, x_1 , x_2 et x_3 . Les gains d’ob-

Plans d'expérience(s)	x_1	x_2	x_3	(x_1, x_2)	(x_1, x_3)	(x_2, x_3)
Gain	G	G	0	$2G$	G	G
Coût	C_1	C_2	C_3	$C_1 + C_2$	$C_1 + C_3$	$C_2 + C_3$
Gain par somme des coûts	$\frac{G}{C_1}$	$\frac{G}{C_2}$	0	$\frac{2G}{C_1+C_2}$	$\frac{G}{C_1+C_3}$	$\frac{G}{C_2+C_3}$
Gain par maximum des coûts	$\frac{G}{C_1}$	$\frac{G}{C_2}$	0	$\frac{2G}{C_2}$	$\frac{G}{C_1}$	$\frac{G}{C_2}$

TABLE 2 – Comparaison de différents plans d'expériences dans une planification asynchrone.

servations des points sont $G_n^{(1)}(x_1) = G_n^{(1)}(x_2) = G$ et $G_n^{(1)}(x_3) = 0$. Pour simplifier $G^{(2)}(x_i, x_j) = G_n^{(1)}(x_i) + G_n^{(1)}(x_j)$, le gain total d'une planification sur deux points est la somme des gains indépendants. Dans ce scénario, observer le point x_3 est alors inutile. Les coût d'observations sont $C(x_1) = C_1$, $C(x_2) = C_2 > C_1$ et $C(x_3) = C_3$ avec $0 < C_3 < (C_2 - C_1)/2$, de sorte que $C_3 + C_1 < (C_1 + C_2)/2$. Les différents plans d'expériences sont résumés tableau 2.

Si tous les plans d'expériences à un et deux points sont comparés, le critère choisit le plan d'expérience composé uniquement de x_1 , ce qui n'est pas forcément mauvais, mais dépend de comment les coûts sont calculés. Par contre, si seuls les plans à deux points sont comparés, alors le plan d'expériences optimal à choisir est (x_1, x_2) , car x_3 n'apporte aucune information. Or, le plan d'expériences à deux points choisi avec le critère MRUR de gain par somme des coûts est (x_1, x_3) : la stratégie recommande de lancer une observation inutile. Ce problème est visible figure 51. Le critère MRUR par somme n'est pas adapté pour la planification parallèle asynchrone, et n'est pas forcément adapté pour comparer des plans d'expériences avec des nombres différents d'observations.

Une solution pourrait être de construire un critère MRUR de planification asynchrone, en considérant le rapport entre la réduction d'incertitude $G_n^{(q)}$ et le *maximum* des coûts d'observations des q points. Cependant, cette stratégie n'est pas efficace non plus. Reconsidérons les trois mêmes plans d'expériences (x_1, x_2, x_3) de gains respectifs $(G, G, 0)$ et de coûts respectifs (C_1, C_2, C_3) (voir tableau 2). Supposons cette fois-ci que $C_3 < C_1 < C_2/2 \Leftrightarrow 2C_1 < C_2$. Dans ces conditions, le plan d'expérience optimal reste le plan (x_1, x_2) . Or, le critère de gain par maximum des coûts indique que les plans d'expériences x_1 et (x_1, x_3) sont préférables au plan (x_1, x_2) . Ce critère ne semble pas adapté non plus à la planification asynchrone sous hypothèse de coûts variables. Dans la pratique, si ce critère était utilisé, il est probable que les plans d'expériences choisis seraient synchrones. En effet, comme seul le maximum des coûts des points est pris en compte, une fois le point de plus haute-fidélité choisi, il est inutile selon ce critère

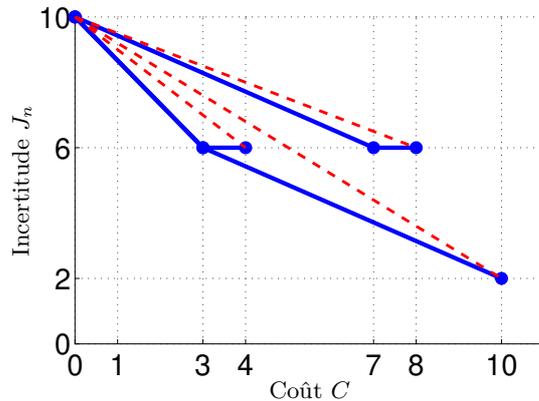


FIGURE 51 – Illustration du problème de la planification asynchrone d'expériences en parallèle. Les valeurs sont $H_n = 10$, $G = 4$, $C_1 = 3$, $C_2 = 7$ et $C_3 = 1$.

d'économiser du temps de calcul. Par exemple, considérons deux niveaux de fidélité δ^{HF} en haute-fidélité et δ^{BF} en basse-fidélité. Dans ces conditions, les plans d'expériences $\{(x_1, \delta^{\text{HF}}), (x_2, \delta^{\text{BF}})\}$ et $\{(x_1, \delta^{\text{HF}}), (x_2, \delta^{\text{HF}})\}$ possèdent le même maximum des coûts $C(\delta^{\text{HF}})$. En supposant qu'observer un point plus coûteux réduit plus l'incertitude, cela pousserait à choisir des observations dont les niveaux de fidélité sont tous égaux. Les plans obtenus seraient donc synchrones.

La question de la planification parallèle asynchrone sous hypothèse de coût d'observation variable reste une question ouverte.

2.5 Conclusion

Dans ce chapitre, nous avons présenté des méthodes de planification d'expériences dans un cadre multi-fidèle.

Dans la première section, nous avons présenté les méthodes de planifications non-séquentielles, dont nous nous servons pour obtenir un plan initial d'expériences. Nous recommandons l'usage de la méthode de Qian [2009] avec une optimisation maximin à chaque niveau pour engendrer un plan d'expériences initial destiné à étudier un simulateur complexe multi-fidèle. Le plan d'expériences ainsi obtenu possède alors trois bonnes propriétés : l'emboîtement, les projections sur les espaces de dimension un et un remplissage d'espace supérieur à celui d'un plan aléatoire (selon le critère de Rennen et al. [2010]).

Dans la seconde section, nous avons rappelé le principe des stratégies de planifications séquentielles SUR. Puis, nous avons proposé différents critères SUR pour pouvoir estimer une fonction de probabilité de défaillance sur un simulateur stochastique. Nous recommandons dans le cas général d'utiliser le critère SUR 1, qui est l'un des plus simples à implémenter, et dont la convergence de la mesure d'incertitude associée assure la convergence de la mesure d'incertitude associée au critère SUR 2.

Enfin, dans la troisième section, nous avons proposé une nouvelle méthode générique de planification séquentielle d'expériences dédiée au cas où le coût d'observation dépend du point choisi. Cette méthode, intitulée MRUR, consiste à choisir le point qui maximise le rapport entre la réduction d'incertitude et le coût d'observation. Cette méthode est une généralisation d'algorithmes déjà existants dans la littérature pour résoudre des problèmes plus spécifiques. Cette stratégie a été testée sur deux exemples académiques multi-fidèles. Sur ces deux expériences, la stratégie MRUR était plus efficace que n'importe quelle stratégie SUR à un niveau donné. De plus, le niveau d'observation choisi par la stratégie MRUR correspond généralement au niveau optimal à observer. Ceci laisse à penser que la stratégie MRUR arrive à trouver le bon compromis entre coût et réduction d'incertitude. Enfin, une dernière expérience montre que la stratégie MRUR peut aisément être étendue dans le cas d'une planification séquentielle d'expériences multi-fidèles parallélisées synchrones.

Cependant, de nombreuses questions restent en suspens. D'abord, il existe peu de recommandations générales sur les caractéristiques des plans initiaux multi-fidèles, telles que, le nombre initial de niveaux d'observation ou le nombre de points par niveau. Une analyse des influences des plans initiaux d'expériences multi-fidèle sur la qualité des prédictions et le comportement des stratégies MRUR pourra être envisagée. Ensuite, le calcul des critères SUR ne prend pas en compte les incertitudes sur les hyper-paramètres. Nous proposons bien une méthode pour tenir compte de ces incertitudes, mais celle-ci n'est qu'une approximation, et n'a pas été testée dans les exemples pour cause de temps de calcul trop élevé.

Mais surtout, beaucoup d'améliorations peuvent être considérées sur la stratégie MRUR. Ainsi, comme mentionné section 2.4.4, la stratégie MRUR ne peut pas être étendue simplement dans le cadre d'une parallélisation asynchrone. Dans le même ordre d'idée, il est difficile d'étendre simplement le critère dans le cas où il y aurait plusieurs types de coûts à prendre en compte (par exemple, un coût monétaire et un coût en temps de calcul). Enfin, les exemples en simulation montrent des situations où la réduction d'incertitude sur un niveau déjà observé est inférieure à celle d'un niveau légèrement moins coûteux. Nous supposons que cela est dû aux fonctions de covariance actuellement

utilisées. Le lien entre le type de processus gaussien et la forme du front de Pareto demanderait donc à être étudié plus en profondeur.

Chapitre 3

Application en sécurité incendie

Sommaire

3.1	Ingénierie de la sécurité incendie	124
3.2	Description et formalisation du problème	125
3.2.1	Présentation du cas d'étude	125
3.2.2	Propriétés remarquables du modèle numérique	132
3.2.3	Quantités d'intérêt	140
3.3	Illustration de la méthodologie proposée	141
3.3.1	Plan d'expériences initial	141
3.3.2	Échantillonnage des hyper-paramètres	143
3.3.3	Prédiction des sorties	146
3.3.4	Distribution a posteriori d'une probabilité de défaillance	149
3.3.5	Courbes de probabilités de défaillance	152
3.3.6	Planification séquentielle d'expériences	152
3.4	Validation de la méthodologie proposée	158
3.4.1	Comparaison des estimations des hyper-paramètres	158
3.4.2	Validation des prédictions des sorties	159
3.4.3	Validation des estimations de probabilités de références	163
3.4.4	Répétition de la planification séquentielle d'expériences	163
3.4.5	Résultats complémentaires sur la réduction d'incertitude	170
3.5	Conclusion	173

3.1 Ingénierie de la sécurité incendie

Dans ce chapitre, nous appliquons les méthodes présentées dans les chapitres 1 et 2 sur une étude d'évaluation des performances du système de désenfumage d'un bâtiment.

L'évaluation de la conformité de bâtiments tels que les Établissements Recevants du Public (ERP), vis-à-vis des normes de sécurité incendie, vise à évaluer la dangerosité des incendies dans des scénarios d'apparition et de propagations des feux. L'approche réglementaire classique, dite prescriptive, décrit les moyens à mettre en œuvre afin de réduire le risque incendie, par exemple, en réduisant la charge de combustibles ou en utilisant des systèmes d'extinction [Carlotti et Lamalle, 2013].

Lorsque la réglementation prescriptive est difficile d'application dans des bâtiments non conventionnels (travaux onéreux, architecture complexe, monument classé...), une alternative reconnue depuis plus de dix ans est le recours à l'Ingénierie de la Sécurité Incendie (ISI), en particulier sur le désenfumage. Cette alternative se fonde sur l'utilisation de codes de calcul de Mécanique des Fluides Numérique (CFD)¹ appliqués à l'incendie pour évaluer les dangers en fixant des objectifs de résultats.

En ISI, le principal objectif est de protéger la vie et la santé des individus. En pratique, le bâtiment doit être en mesure de conserver un environnement supportable (en terme de température, de visibilité...) le temps que les occupants puissent évacuer vers un endroit sûr. Les critères de tenabilité utilisés dans cette thèse proviennent de la norme ISO 13571 [2012] et sont fondés sur le jugement scientifique disponible sur les conséquences de l'exposition humaine aux incendies (voir par exemple Guillaume [2006]). L'objectif d'une étude ISI est de déterminer dans différents scénarios d'incendie si les conditions environnantes restent supportables le temps de l'évacuation, selon les modèles numériques modélisant l'incendie.

Les modèles CFD appliqués aux incendies permettent de prédire les différentes phases de développement du feu. Cependant, ces modèles dits « de champs » sont coûteux en temps de calcul, ce qui limite le nombre d'appels possibles au code. Pour mener les évaluations de conformité, la méthode actuelle consiste à choisir quelques scénarios représentatifs, de simuler en détails ces scénarios, et d'analyser les résultats ainsi obtenus. La déclaration de conformité ou de non-conformité est laissée au jugement de l'expert en sécurité incendie. Cette méthode présente le défaut d'être très dépendante de l'opérateur en charge de l'étude. De plus, cette méthode ne fournit aucune indication sur la fiabilité

1. Computational Fluid Dynamics

de la décision de conformité ou non du bâtiment.

Pour résoudre ces problèmes, nous proposons une méthode d'analyse de la conformité sur des simulateurs coûteux en temps de calcul, destinée à être un outil d'aide à la décision. Cette méthode d'analyse est fondée sur une approche multi-fidèle bayésienne, définie chapitre 1, qui permet de pouvoir prédire les résultats d'une simulation à partir d'un plan d'expérience faible en coût d'observation. De plus, nous proposons une méthode de choix des simulations, fondée sur les principes décrits dans le chapitre 2, dans le but d'améliorer une évaluation initiale de la conformité.

La section 3.2 décrit le problème étudié, ainsi que le modèle numérique qui permet d'obtenir des observations. La section 3.3 présente un exemple réaliste de la méthodologie dans lequel l'utilisateur ne dispose que d'un coût limité en observation. Enfin, la section 3.4 valide la méthodologie en présentant des répétitions de la méthode et en validant les résultats sur des données de références.

3.2 Description et formalisation du problème

3.2.1 Présentation du cas d'étude

Présentation de Fire Dynamics Simulator

L'utilisation de modèles CFD dédiés à l'incendie est de plus en plus courante dans les analyses de conformité de systèmes de désenfumage. Un des simulateurs les plus utilisés par la communauté scientifique en sécurité incendie est Fire Dynamics Simulator (FDS). Écrit en Fortran et développé par le National Institute of Standards and Technology (NIST) [McGrattan et al., 2010], ce simulateur résout les équations de Navier-Stokes (équations de la mécanique des fluides : conservations de la masse, de la quantité de mouvement et de l'énergie) sous l'hypothèse d'un faible nombre de Mach (vitesse d'écoulement de l'air très petite devant la vitesse du son, ce qui implique notamment que la compressibilité de l'air peut être négligée). Ce simulateur est en particulier dédié à la description et l'étude du transport des fumées et de la chaleur.

Le simulateur FDS procède par différences finies : il découpe l'espace en petits parallélépipèdes et le temps en petits intervalles, puis résout les équations de la mécanique des fluides discrétisées en espace et en temps. Dans cette thèse, seules des mailles cubiques de longueur δ_{xyz} sont considérées. Soit δ_t le pas de temps. Pour des raisons de stabilité numérique, le pas de temps est lié à la taille de maille, notamment par la contrainte de

Courant-Friedrichs-Lewy (CFL) qui impose

$$\delta_t \leq \frac{\delta_{xyz}}{\|\vec{u}\|_{+\infty}} \quad (3.1)$$

où \vec{u} est la vitesse du fluide [McGrattan et al., 2014, Section 6.4.7]. Le pas de temps est réglé automatiquement par FDS en utilisant la taille de maille δ_{xyz} . Dans la pratique, cela implique qu’une simulation avec une plus petite taille de maille utilise aussi un pas de temps plus petit, ce qui augmente d’autant le temps de calcul (voir section 3.2.2).

Puisque le pas de temps est réglé automatiquement, le seul paramètre réglé par l’utilisateur qui contrôle la qualité de la simulation est la taille de maille δ_{xyz} . Ce paramètre joue le rôle de paramètre de fidélité : une petite taille de maille permet d’obtenir une simulation plus fidèle mais plus coûteuse, tandis qu’une grande taille de maille permet d’obtenir un résultat plus rapidement, mais qui est plus grossier.

Le simulateur FDS impose une contrainte sur le nombre de maille dans une simulation. En effet, FDS utilise une transformée de Fourier rapide pour résoudre l’équation de la chaleur, ce qui impose que le nombre de mailles dans une simulation doit être de la forme $2^p 3^q 5^r$, où (p, q, r) sont des entiers naturels [McGrattan et al., 2014, Section 6.3.1]. Cette condition impose une restriction sur la taille de maille, qui dépend du bâtiment simulé.

Dans cette thèse, nous utilisons la version 5.5.3. de FDS, qui est la version utilisée au Laboratoire National de métrologie et d’Essais (LNE), même si la version 6.6.0. de FDS est aussi disponible. Cela ne change en rien la méthodologie mise en œuvre. Nous précisons toutefois qu’il est recommandé que les résultats analysés soient tous issus d’une même version d’un simulateur.

Cas d’application

L’exemple étudié dans cette thèse est l’étude du désenfumage d’un bâtiment existant au LNE, présenté figure 52. Le bâtiment est un hall d’essais mesurant $20 \text{ m} \times 12 \text{ m} \times 16 \text{ m}$. Il dispose de quatre ouvertures, deux portes au sol de taille $2 \text{ m} \times 1 \text{ m}$ et deux exutoires au plafond de taille $2 \text{ m} \times 2 \text{ m}$ permettant l’évacuation des fumées en cas d’incendie. La ventilation est supposée naturelle. Les épaisseurs et les propriétés thermiques des parois sont définies en accord avec la structure réelle. Les dimensions et les propriétés des différents composants du bâtiment sont parfaitement connues.

À l’état initial, la température dans le domaine de calcul est réglée par la température ambiante. Sur les bords libres, une condition aux limites est imposée sur la pression

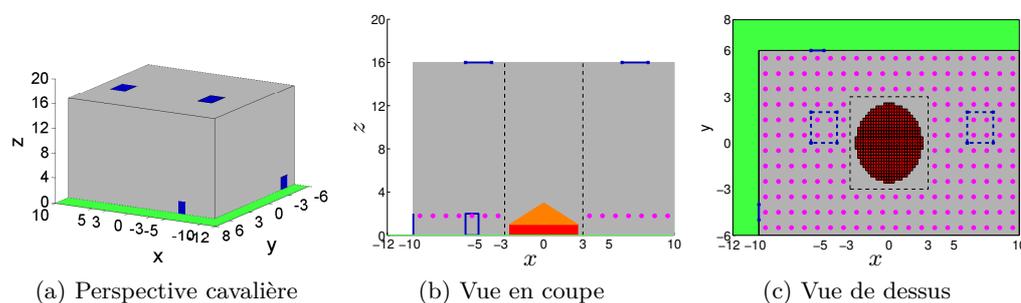


FIGURE 52 – Vues du bâtiment étudié. Les ouvertures sont en bleues, le feu est en rouge au centre du bâtiment et les capteurs sont affichés par les points mauves.

statique. Dans le scénario considéré, le feu est placé au centre de la pièce. Le combustible est un uréthane, un produit souvent utilisé en industrie, par exemple pour fabriquer la mousse de rembourrage dans les canapés ou des mousses d'isolation.

Définition des variables

Hormis les caractéristiques du bâtiment, supposées connues dans notre étude, les variables définissant une simulation peuvent être classées en deux catégories : les variables environnementales x_e et les variables de scénario x_s .

Les variables environnementales sont des variables qui doivent être fixées dans une simulation de FDS, mais dont la valeur réelle est incertaine dans un scénario donné. Les variables environnementales considérées dans cette thèse sont la température extérieure T_{ext} , la pression atmosphérique P_{atm} et la température ambiante T_{amb} . Dans un souci de simplification, les effets du vent (dont la vitesse est supposée nulle) et de l'humidité de l'air (fixée à 50%) ne sont pas pris en compte. L'incertitude des variables environnementales est représentée par une distribution de probabilité f^{env} . La distribution de probabilité des variables environnementales représente l'ensemble des conditions climatiques rencontrées sur une année au bâtiment étudié. Le bâtiment est situé à Trappes (France). Selon des rapports météorologiques, la distribution des températures extérieures peut être modélisée par une loi normale dont l'intervalle à trois écarts-types est $[-10; 30]$ °C, celle de la pression atmosphérique par une loi normale dont l'intervalle à trois écarts-types est $[98; 102]$ kPa. Selon une étude interne au LNE, la distribution de

la température ambiante connaissant la température extérieure est

$$T_{amb} (\text{°C})|T_{ext} (\text{°C}) \sim N\left(22.5 + 2.5 \times 0.8 \times \frac{T_{ext} - 10}{20/3}, \left(2.5\sqrt{1 - 0.8^2}\right)^2\right) \quad (3.2)$$

Conséquemment, la distribution des variables environnementales f^{env} est une loi gaussienne dont le vecteur moyenne et la matrice de covariance sont respectivement

$$\begin{pmatrix} 10 \text{ °C} \\ 100 \text{ kPa} \\ 22,5 \text{ °C} \end{pmatrix}, \quad \begin{pmatrix} (20/3 \text{ °C})^2 & 0 & 0.8 \times 20/3 \text{ °C} \times 2,5 \text{ °C} \\ 0 & (2/3 \text{ kPa})^2 & 0 \\ 0.8 \times 20/3 \text{ °C} \times 2,5 \text{ °C} & 0 & (2,5 \text{ °C})^2 \end{pmatrix}. \quad (3.3)$$

Les variables de scénario x_s sont des variables ayant une influence sur le comportement de l'incendie. Ces variables définissent le scénario, et sont donc fixées dans un scénario donné. Certaines de ces variables décrivent l'évolution de la source de l'incendie en fonction du temps, représentée dans FDS par le Taux de Dégagement de Chaleur (HRR)². Le HRR, exprimé en kW, représente la puissance calorifique dégagée par le foyer en fonction du temps. Quatre variables sont nécessaires pour décrire l'évolution du HRR : le taux de croissance α en $\text{kW} \cdot \text{s}^{-2}$, la surface du feu A_f en m^2 , la chaleur surfacique \dot{Q}'' en $\text{kW} \cdot \text{m}^{-2}$ et l'énergie surfacique totale q_{fd} en $\text{MJ} \cdot \text{m}^{-2}$. L'évolution du HRR est décomposée en trois phases, correspondant aux trois phases de l'incendie :

1. l'allumage : le HRR suit une évolution parabolique [Alpert, 2016] ;
2. l'état stable : le HRR est constant, $HRR(t) = A_f \cdot \dot{Q}''$, où la constante est le produit entre la surface du feu A_f et la chaleur surfacique maximale \dot{Q}'' ;
3. l'extinction : le HRR décroît linéairement, $HRR(t) = A_f \dot{Q}'' \cdot \frac{t^{\text{feu}} - t}{t^{\text{feu}} - t_i}$ où t^{feu} est l'instant d'extinction du feu, et t_i est l'instant où le feu commence à s'éteindre. L'instant d'extinction du feu et le début de la phase d'extinction sont défini par l'énergie dégagée totale $\int_0^{t^{\text{feu}}} HRR(t)dt = q_{fd}$, et $\int_0^{t_i} HRR(t)dt = 70\% \cdot q_{fd}$.

La forme générale du HRR en fonction du temps est présentée figure 53.

Selon Drysdale [2011], le taux de croissance α peut varier entre $0,011\,338 \text{ kW} \cdot \text{m}^{-2}$ pour un feu se développant lentement et $0,2 \text{ kW} \cdot \text{m}^{-2}$ pour un feu se développant rapidement. Ces valeurs sont utilisées comme bornes pour la variable α . Dans la réglementation française concernant les systèmes de désenfumage, la surface du feu dans les scénarios doit valoir 9 m^2 , 18 m^2 ou 36 m^2 en fonction des bâtiments étudiés. Dans cette étude, la valeur de 36 m^2 est écartée et le domaine d'étude de la surface du feu A_f est $[1; 20] \text{ m}^2$.

2. Heat Release Rate

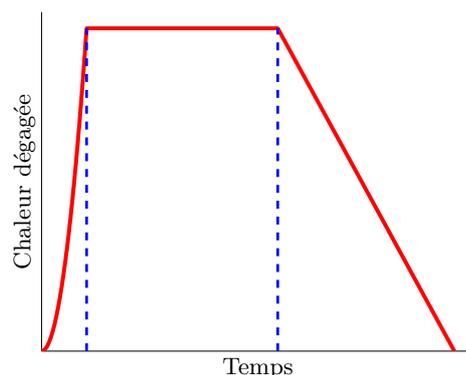


FIGURE 53 – Évolution de la chaleur dégagée par le feu en fonction du temps

Enfin, l'usage en ISI est de considérer que \dot{Q}'' varie entre $300 \text{ kW} \cdot \text{m}^{-2}$ et $600 \text{ kW} \cdot \text{m}^{-2}$ et q_{fd} entre $300 \text{ MJ} \cdot \text{m}^{-2}$ et $600 \text{ MJ} \cdot \text{m}^{-2}$.

En plus de ces sept variables d'entrées que sont $(T_{ext}, P_{atm}, T_{amb}; \alpha, A_f, \dot{Q}'', q_{fd})$, le taux de production de fumée Y_{soot} , en $\text{kg} \cdot \text{kg}^{-1}$, est également considéré, car il influence grandement la production de fumée, et donc la visibilité. Il est admis dans cette thèse que le taux de production de fumée varie entre $0,01227 \text{ kg} \cdot \text{kg}^{-1}$ et $0,1227 \text{ kg} \cdot \text{kg}^{-1}$.

En pratique, la taille de maille δ_{xyz} idéale est calculable à partir du diamètre du feu. Lin et al. [2009] recommandent d'utiliser une taille de maille inférieure ou égale au dixième du diamètre caractéristique de l'incendie, $\delta_{xyz} \leq D^*/10$, où le diamètre du feu D^* se calcule de la façon suivante :

$$D^* = \left(\frac{\dot{Q}'' A_f}{\rho^{\text{air}} C_p^{\text{air}} T^{\text{air}} \sqrt{g}} \right)^{\frac{2}{5}}, \quad (3.4)$$

où $\dot{Q}'' A_f$ est le HRR maximal de la source de l'incendie, ρ^{air} est la masse volumique de l'air, C_p^{air} sa capacité thermique isobare, T^{air} sa température et g l'accélération de pesanteur terrestre. Pour $\dot{Q}'' = 450 \text{ kW} \cdot \text{m}^{-2}$, $A_f = 10 \text{ m}^2$, $\rho^{\text{air}} = 1,2 \text{ kg} \cdot \text{m}^{-3}$, $C_p^{\text{air}} = 1000 \text{ J} \cdot \text{kg}^{-1} \cdot \text{K}$, $T^{\text{air}} = 300 \text{ K}$ et $g = 9,8 \text{ m} \cdot \text{s}^{-2}$, le diamètre est de $D^* = 1,7 \text{ m}$ et donc la taille de maille idéale de 17 cm . Compte-tenu des variations possibles des différents facteurs, la taille de maille idéale pour traiter le problème est de l'ordre de 10 cm pour cette expérience.

Du fait des restrictions de FDS sur le nombre de maille et des dimensions du bâtiment étudié, la taille de maille δ_{xyz} ne peut pas prendre n'importe quelle valeur. La géométrie

Niveau	1	2	3	4	5	6	8	9	10
δ_{xyz} (cm)	100	50	33.3	25	20	17.7	12.5	11.1	10
Durée	4,1 min	69,3 min	6,05 h	19,6 h	48,6 h	4,26 j	13,8 j	22,3 j	34,3 j

TABLE 3 – Liste des niveaux possibles de l’expérience de FDS. Les temps affichés sont des temps moyens obtenus par une régression logarithmique (figure 55).

Nom	Unité	Variable	Domaine d’étude
Température extérieure	°C	T_{ext}	$[-10; 30]$
Pression atmosphérique	kPa	P_{atm}	$[98; 102]$
Température ambiante	°C	T_{amb}	$[10; 30]$
Accélération d’allumage	$\text{kW} \cdot \text{s}^{-2}$	α	$[0.011338; 0.2]$
Surface du feu	m^2	A_f	$[1; 20]$
Puissance surfacique maximale	$\text{kW} \cdot \text{m}^{-2}$	\dot{Q}''	$[300; 600]$
Énergie surfacique dégagée totale	$\text{MJ} \cdot \text{m}^{-2}$	q_{fd}	$[300; 600]$
Taux de production de fumée	$\text{kg} \cdot \text{kg}^{-1}$	Y_{soot}	$[0.01227; 0.1227]$
Taille de maille	cm	δ_{xyz}	$[10; 100]$

TABLE 4 – Liste des variables d’entrée pour le cas d’étude considéré.

du bâtiment simulé et la nature du maillage imposent que la taille de maille doit être une fraction entière de 100 cm. En combinant cette caractéristique à la limitation sur le nombre de mailles, la taille de maille ne peut être qu’une fraction entière de la taille maximale, $\delta_{xyz} = 100 \text{ cm} / (2^p 3^q 5^r)$, où le dénominateur est un produit d’une puissance de deux, de trois et de cinq. La combinaison de toutes ces caractéristiques donne neuf niveaux de fidélité possibles présentés tableau 3. Enfin, il est à noter qu’à un niveau δ_{xyz} fixé, les surfaces de feu A_f doivent être un multiple entier de l’aire d’une maille, soit de δ_{xyz}^2 , ce qui limite d’autant l’espace des entrées.

Les huit entrées physiques sont résumées en un vecteur x . A ces huit variables, s’ajoute le paramètre de fidélité, la taille de maille δ_{xyz} , qui définit la fidélité de la simulation. Toutes les entrées sont présentées dans le tableau 4.

Données de sortie et grandeurs

Une étude ISI cherche à évaluer si les conditions sur le chemin d’évacuation restent supportables par les occupants. La notion de tenabilité est liée à l’évolution de variables physiques représentant la dangerosité de l’incendie. Dans cette thèse, trois types de menaces liées aux incendies sont considérés :

Nom	Variable	Sécurité
Flux thermique radiatif	H^f	$H^f \leq 2,5 \text{ kW} \cdot \text{m}^{-2}$
Température à 1,8 m	T^c	$T^c \leq 60 \text{ °C}$
Visibilité	V	$V \geq 5 \text{ m}$

TABLE 5 – Variables étudiées et seuils de dangerosité

- la perte de visibilité qui impacte l'évacuation des personnes (désorientation, exposition plus longue aux dangers. . .). Cette menace est évaluée par la visibilité V ;
- les brûlures générées par le feu et la fumée, qui peuvent conduire à l'incapacité ou à la mort des individus. Ces brûlures sont causées par le flux de chaleur H^f associé au rayonnement de la fumée et à la température de l'atmosphère. De plus, des brûlures thermiques de la peau et des voies respiratoires peuvent se produire lorsque la température T^c atteint des valeurs élevées.
- l'intoxication par les fumées. Compte tenu de la difficulté pour prédire numériquement les phénomènes chimiques et biologiques, l'intoxication est évaluée au travers du terme de visibilité V .

Les données sont évaluées à une hauteur de 1,8 m (correspondant à la hauteur de la tête d'un individu). Pour chacune des trois variables d'intérêt (flux radiatif, température et visibilité), un seuil critique de dangerosité a été choisi à partir de la norme ISO 13571 [2012]. Ces seuils sont présentés tableau 5. A noter que le guide du désenfumage du Laboratoire Centrale de Préfecture de Police (LCPP) [Thiry-Muller, 2017] a récemment fixé de nouveaux seuils, légèrement différents de ceux retenus dans cette thèse, issues d'un consensus entre les organismes français reconnus compétents en désenfumage. Ces travaux de thèse étant antérieurs à ce document, ces critères n'ont pas été retenus, mais la méthodologie mise en place reste applicable.

Une simulation de FDS ne renvoie pas un résultat scalaire par variable physique d'intérêt, mais un ensemble de séries temporelles, au nombre de une par capteur présent dans la simulation. En effet, pour chaque simulation, des capteurs sont répartis dans la pièce. Chaque capteur est à une position fixe, situé à un 1,8 m du sol. Les capteurs sont espacés d'une distance de 1 m, pour une densité de un capteur par m^2 . Cependant, il n'y a aucun capteur situé dans le panache au-dessus du feu pour éviter de perturber les résultats avec des valeurs extrêmes. Cette zone sans capteur est un carrée de 6 m de coté centré sur le centre du feu. La position des capteurs est représentée figure 52. Le nombre total de capteurs est de $(20 \text{ m} \times 12 \text{ m} - (6 \text{ m})^2) \times 1 \text{ capteur} \cdot \text{m}^{-2} = 204 = n^{\text{sens}}$ capteurs. Ainsi, après une simulation, n^{sens} sorties temporelles par quantité étudiée sont

récupérées, $\left(H_i^f(t), T_i^c(t), V_i(t)\right)_{1 \leq i \leq n^{\text{sens}}, t \in [0, t^{\text{feu}}]}$. Les figures 54a à 54c présentent des sorties de FDS par capteur pour une expérience et les figures 54d à 54f tracent des quantiles de ces sorties en fonction du temps.

Il est délicat de définir directement à partir des résultats bruts issus des capteurs (figures 54a à 54c) si les variables de sortie restent dans leur zone de conformité. Afin de simplifier l'analyse, un post-traitement est effectué pour chaque variable de sortie à partir de ces courbes temporelles pour obtenir une quantité scalaire. Le résultat de ce post-traitement est comparée au seuil critique correspondant (tableau 5). Le post-traitement que nous proposons est antérieur au guide du désenfumage du LCPP [Thiry-Muller, 2017], d'où de légères différences, et il semble qu'il n'existe pas d'autres approches similaires pour évaluer la conformité dans la littérature sur la sécurité incendie. Cependant, les méthodes développées durant cette thèse peuvent parfaitement s'appliquer si un autre post-traitement, d'autres sorties (salaires ou vectorielles) ou un autre simulateur étaient utilisés.

Notre post-traitement effectue d'abord une moyenne spatiale sur l'ensemble des capteurs, puis retourne la valeur extrême de cette moyenne sur l'ensemble de la courbe (maximum ou minimum sur le temps, selon le cas). Ainsi, nous appelons par abus de langage « les sorties de FDS », les valeurs suivantes

$$H^f = \max_{t \in [0, t^{\text{feu}}]} \left\{ \frac{1}{n^{\text{sens}}} \sum_{i=1}^{n^{\text{sens}}} H_i^f(t) \right\}, \quad (3.5a)$$

$$T^c = \max_{t \in [0, t^{\text{feu}}]} \left\{ \frac{1}{n^{\text{sens}}} \sum_{i=1}^{n^{\text{sens}}} T_i^c(t) \right\}, \quad (3.5b)$$

$$V = \min_{t \in [0, t^{\text{feu}}]} \left\{ \frac{1}{n^{\text{sens}}} \sum_{i=1}^{n^{\text{sens}}} V_i(t) \right\}, \quad (3.5c)$$

où t^{feu} est l'instant d'extinction du feu, défini à partir des variables d'entrée. Les figures 54g à 54i illustrent ce post-traitement. Thiry-Muller [2017] propose de lisser les sorties au préalable en effectuant une moyenne glissante avec une fenêtre de trente secondes avant d'appliquer les mêmes opérations.

3.2.2 Propriétés remarquables du modèle numérique

Le modèle numérique est vu comme une fonction définie sur un espace de dimension neuf (huit entrées physiques, plus un paramètre de fidélité) à valeur dans un espace de dimension trois. Le code de calcul FDS est l'une des composantes de ce modèle numérique. Dans cette section, nous présentons quelques propriétés de ce modèle numérique.

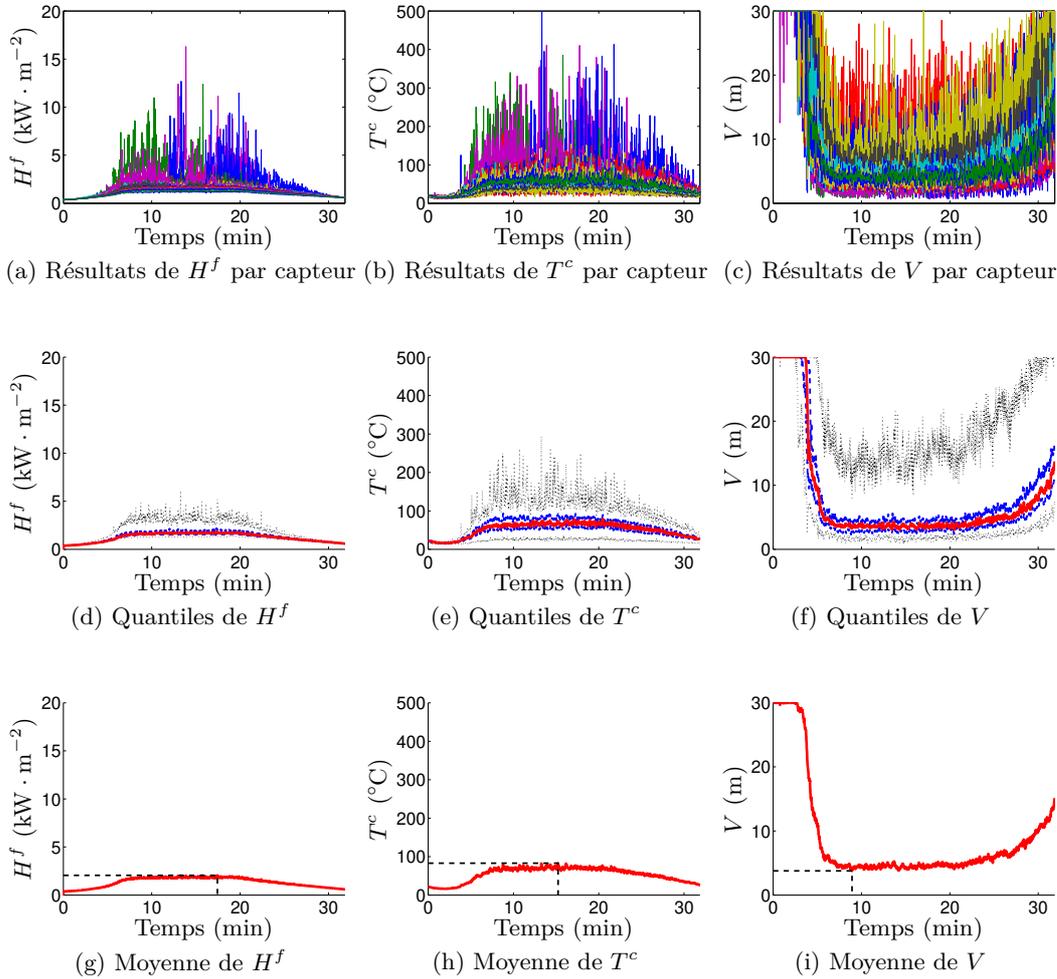


FIGURE 54 – Sorties de FDS, pour les entrées $T_{ext} = 11,855 \text{ }^\circ\text{C}$, $P_{atm} = 99,439 \text{ kPa}$, $T_{amb} = 21,531 \text{ }^\circ\text{C}$, $\alpha = 0,0555 \text{ kW} \cdot \text{s}^{-2}$, $A_f = 21 \text{ m}^2$, $\dot{Q}'' = 412,22 \text{ kW} \cdot \text{m}^{-2}$, $q_{fd} = 523,08 \text{ MJ} \cdot \text{m}^{-2}$, $Y_{soot} = 0,030256 \text{ kg} \cdot \text{kg}^{-1}$ et $\delta_{xyz} = 20 \text{ cm}$. (a, b, c) Résultats bruts de trente capteurs $(H_i^f(t), T_i^c(t), V_i(t))_{1 \leq i \leq 30}$. (d, e, f) Médianes (traits continus rouges), quartiles (traits discontinus bleus) et premier et dernier quadragintiles (quantiles à 2,5 % et 97,5 %, traits pointillés noirs) des différentes sorties sur l'ensemble des capteurs en fonction du temps. (g, h, i) Moyennes des différentes sorties sur l'ensemble des capteurs en fonction du temps. Les traits pointillés indiquent les valeurs considérées comme les sorties du modèle numérique, ainsi que l'instant où l'extremum est atteint.

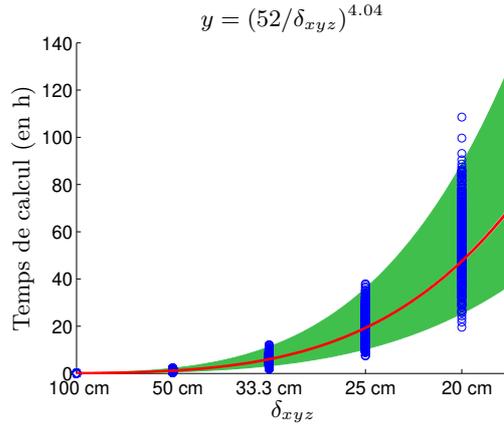


FIGURE 55 – Temps de calcul de FDS en fonction de la taille de maille. Les points correspondent à des temps de calcul renvoyés par FDS, la courbe continue est une régression logarithmique, et la zone pleine est l'intervalle de confiance à 95 %.

Coût d'observation

Le coût de calcul C est le temps nécessaire pour mener une simulation sur FDS, les temps de calcul des autres morceaux du modèle numérique étant négligés. Dans cette thèse, nous considérons que le coût ne dépend que de la taille de maille, les variations résiduelles étant considérées comme aléatoires. Le temps de calcul C est modélisé par une variable aléatoire loi log-gaussienne, dont les paramètres dépendent de la taille de maille. Une régression logarithmique sur 11 225 simulations réalisées durant la thèse, réparties sur les cinq niveaux de plus basse-fidélité, permet d'obtenir les paramètres de la loi,

$$\log_{10} \left(\frac{C}{c_0} \right) | \delta_{xyz} \sim \mathcal{N} \left(\kappa \log_{10} \left(\frac{\delta_0}{\delta_{xyz}} \right), s_c^2 \right) \quad (3.6)$$

avec $\delta_0 = 52$ cm pour $c_0 = 1$ h, $\kappa = 4.04$ et $s_c = 0.138$, ce qui correspond à une incertitude d'un facteur $10^{2s_c} = 1.9$. Les temps de calcul ainsi estimés sont affichés tableau 3, et l'évolution en fonction de la fidélité est représentée figure 55.

Variance de l'effet stochastique en sortie

La figure 56 présente l'évolution des trois sorties de FDS en fonction de deux entrées : la puissance surfacique du feu \dot{Q}'' et la taille de maille δ_{xyz} , les autres entrées étant fixées. Une première chose à voir sur cette figure est que les sorties possèdent des variations très rapides : deux entrées proches peuvent renvoyer des sorties différentes. Cela laisse

penser à un comportement stochastique de la part du modèle numérique.

Le modèle numérique n'est pas, à proprement parler, stochastique, puisque deux simulations de ce modèle numérique avec les mêmes entrées obtiennent exactement les mêmes sorties. Cependant, nous considérons malgré tout ce modèle numérique comme stochastique en prenant en compte le fait que deux simulations ayant deux entrées aussi proches que possibles aboutissent à des valeurs en sortie sensiblement différentes. Ce phénomène est parfois appelé « effet de pépite ».

Ce comportement remarquable semble lié à l'initialisation dans FDS. En effet, afin d'éviter les feux symétriques, le champ de pression initial est initialisé avec une fonction génératrice pseudo-aléatoire. Or, la graine de cette fonction génératrice est fixée lors de la compilation. De ce fait, deux scripts d'entrées identiques créent la même expérience avec le même champ de pression initial et donc produisent des résultats identiques (le reste du simulateur étant déterministe). Par contre, comme le comportement du simulateur est « chaotique », au sens où deux simulations engendrées à partir de deux vecteurs d'entrées proches mais différents produisent des incendies légèrement différents à l'état initial. Cette différence s'amplifie au cours des simulations, de sorte que le modèle numérique aboutit à des valeurs en sorties notablement différentes entre les deux simulations.

Une valeur grossière de la variance de cet effet stochastique peut être obtenue en appliquant un modèle de régression. Sur chacun des cinq niveaux de plus basse fidélité, nous utilisons un certain nombre d'observations, respectivement 1309 au niveau 100 cm, 1309 au niveau 50 cm, 1309 au niveau 33 cm, 909 au niveau 25 cm et 529 au niveau 20 cm. Chaque plan d'expériences associé à un niveau de fidélité voit ses observations réparties dans l'espace des entrées de dimension huit. Sur chacun des cinq plans d'expériences, deux modèles de régression sont appliqués : un modèle de régression linéaire, et un modèle de régression par processus gaussien. Le processus gaussien est un processus stationnaire de moyenne constante suivant une loi uniforme impropre sur \mathbb{R} et de fonction de covariance une covariance de Matérn anisotrope. Sur chacun de ces modèles, l'effet stochastique est supposé homoscédastique. Sur dix modèles (cinq niveaux fois deux types de modèles), les hyper-paramètres, dont la variance de l'effet stochastique, sont estimés par maximum de vraisemblance.

La figure 57 montre les estimations des écarts-types des effets stochastiques de chaque sortie selon deux modèles, comparés aux écarts-types totaux des valeurs des sorties à un niveau donné. Les résultats sont tracés en fonction du niveau. La figure indique que les variances des effets stochastiques varient d'un niveau à l'autre, mais de manière faible comparées aux variations totales des sorties à un niveau donné. De plus, il n'y a pas de tendance générale qui apparaît sur la variance de l'effet stochastique en fonction des

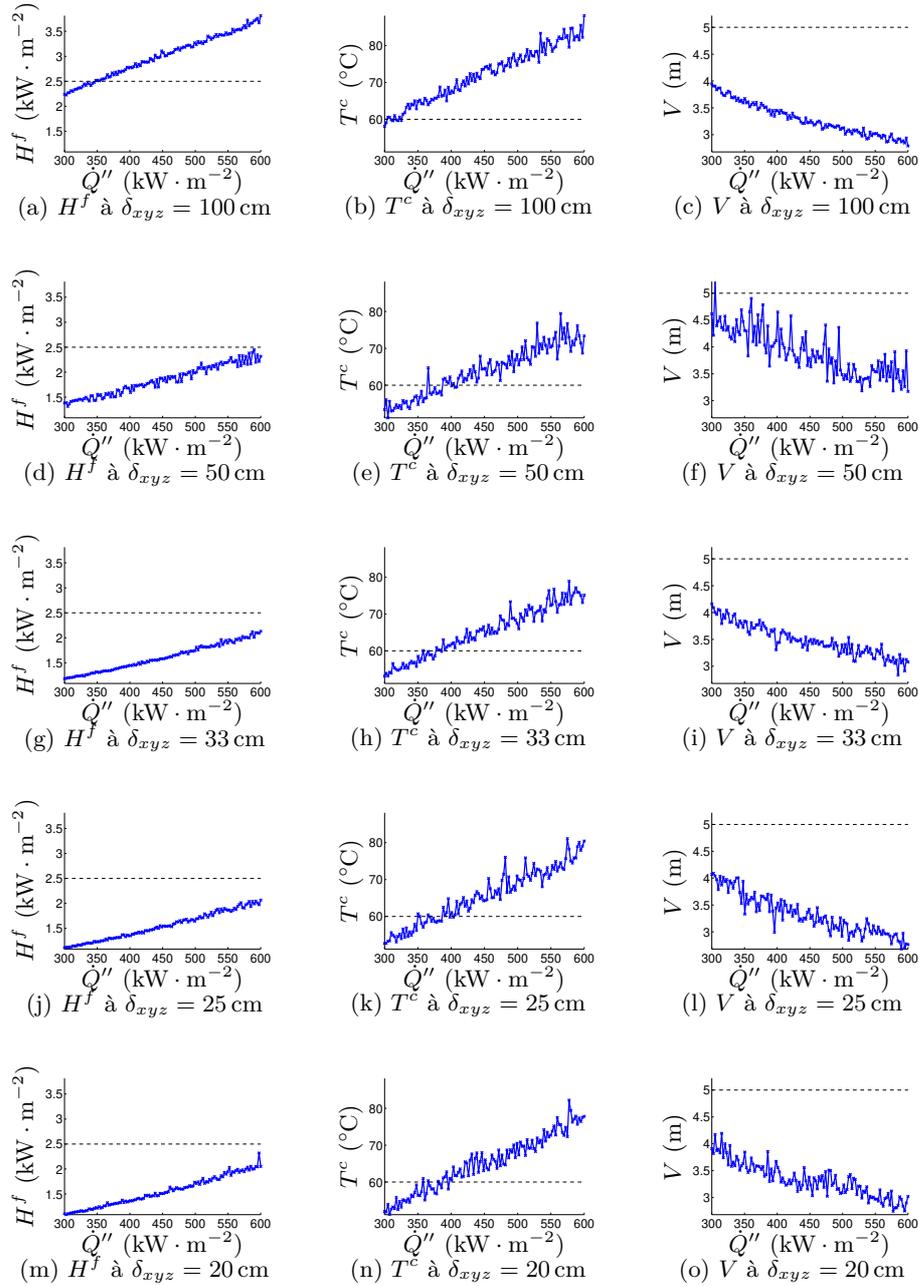


FIGURE 56 – Les trois sorties étudiées en fonction de \dot{Q}'' sur les cinq premiers niveaux. Chaque colonne correspond à une sortie (respectivement, de gauche à droite, H^f , T^c et V), chaque ligne à un niveau (respectivement, de haut en bas, $\delta_{xyz} = 100, 50, 33, 25$ et 20 cm). Les autres entrées sont fixées à $T_{ext} = 10$ °C, $P_{atm} = 100$ kPa, $T_{amb} = 22,5$ °C, $\alpha = 0,1057$ kW · s⁻², $A_f = 16$ m², $q_{fd} = 450$ MJ · m⁻², $Y_{soot} = 0,0409$ kg · kg⁻¹

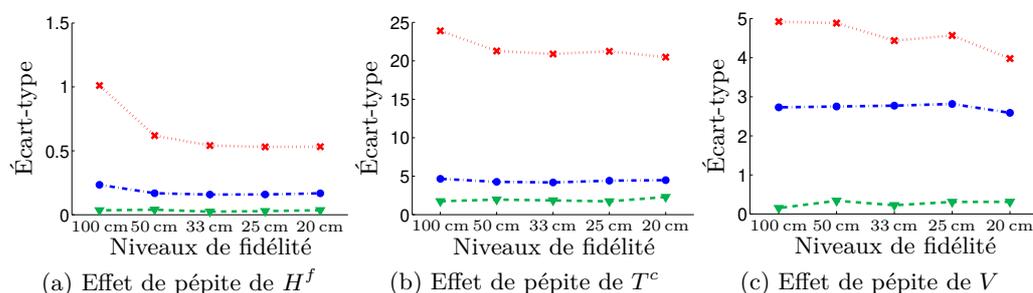


FIGURE 57 – Estimation de l'écart-type de l'effet stochastique en fonction du niveau pour les trois sorties. Les croix rouges en haut correspondent aux écarts-types totaux de la sortie en fonction du niveau de fidélité, les cercles bleus au milieu aux estimations issues des régressions linéaires et les triangles verts en bas aux estimations issues des régressions par processus gaussiens.

niveaux. La figure met aussi en évidence que l'effet de pépite résiduel est plus faible dans le modèle de régression par processus gaussien que dans le modèle de régression linéaire. Cela laisse penser que le modèle numérique possède des sorties non-linéaires, particulièrement pour la visibilité V .

Dans un modèle de régression linéaire ou de régression par processus gaussien, la valeur de sortie est supposée suivre une loi gaussienne dont la variance est la variance de l'effet stochastique. Une façon de vérifier si les résultats sont compatibles avec cette hypothèse, est de regarder la distribution des résidus. Nous nous focalisons sur le niveau 20 cm et sur son plan d'expérience associé, constitué de 529 observations. Les hyper-paramètres, dont la variance de l'effet stochastique, sont estimées par maximum de vraisemblance. En chaque point x_i du plan d'expérience à 20 cm, une estimation de la valeur moyenne de la sortie $\hat{\xi}_{-i}(x_i)$ est obtenue en excluant la valeur (x_i, z_i) des données d'apprentissage (Leave-One-Out (LOO)), en supposant les hyper-paramètres fixées. Cette estimation de la valeur moyenne est associée à un écart-type $\sigma_{-i}(x_i)$. La figure 58 tracent les histogrammes des résidus normalisés des estimations par validation croisée, $(z_i - \hat{\xi}_{-i}(x_i)) / \sqrt{\sigma_{-i}^2(x_i) + \lambda}$, pour chaque modèle et pour chaque sortie au niveau, où z_i est l'observation de FDS au point x_i au niveau 20 cm. La figure montre que les résidus ainsi obtenus suivent approximativement la loi normale centrée réduite. Ceci laisse à penser que la distribution des sorties peut être modélisée par une loi normale.

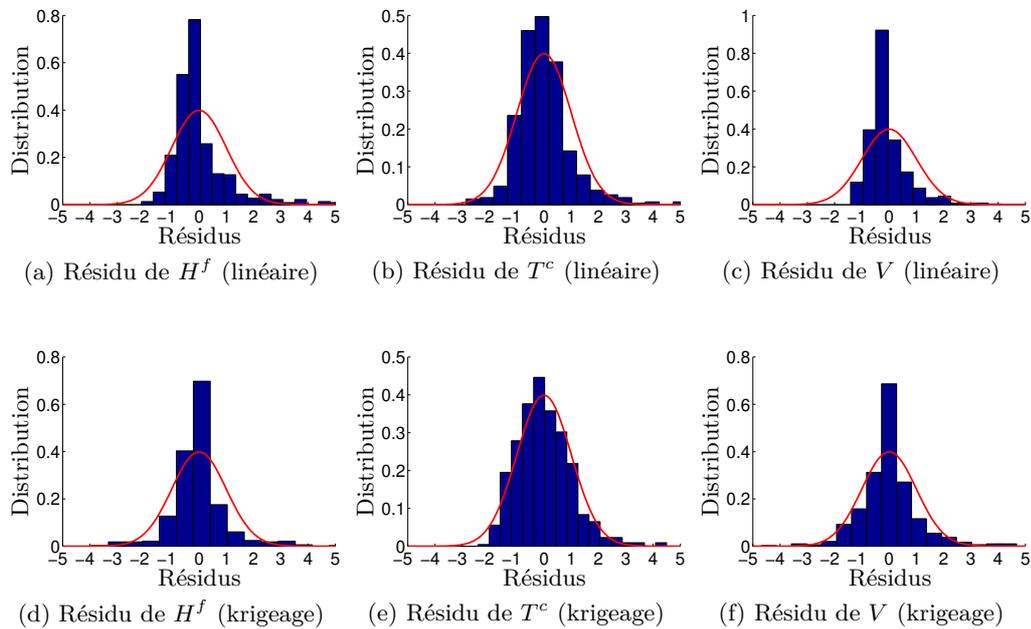


FIGURE 58 – Histogramme des résidus normalisés des estimations de chaque sortie, par régression linéaire (en haut) et par krigeage (en bas). La ligne continue représente la densité de probabilité de la loi normale centrée réduite.

Évolution en fonction du niveau de fidélité

Un autre comportement visible sur la figure 56 est l'évolution de la moyenne du modèle numérique en fonction du niveau. Il apparaît que, même si la moyenne évolue de niveau en niveau, la forme générale de cette fonction reste la même : même monotonie, même vitesse de croissance. . . Il n'y a apparemment pas de changement majeur de comportement d'un niveau de fidélité à un autre.

L'évolution des sorties de FDS en fonction de la taille de maille δ_{xyz} peut être observée en considérant des plans d'expériences emboîtés, c'est-à-dire des plans d'expériences où les mêmes entrées x sont observés à plusieurs niveau de fidélité δ .

Nous disposons de 529 points x observés sur les cinq niveaux de plus basse fidélité, répartis sur l'espace des entrées de dimension huit. En chacun de ces points, les différences des sorties $Z_{x,\delta} - Z_{x,\delta^{\text{ref}}}$ sont calculées, où δ^{ref} correspond ici au niveau 20 cm. La figure 59 présente l'évolution des différences pour chaque sortie en fonction du niveau. Ainsi, les figures 59a à 59c présentent l'évolution de la différence des moyennes empiriques des différences en fonction du niveau, ainsi que les intervalles à deux écarts-types des

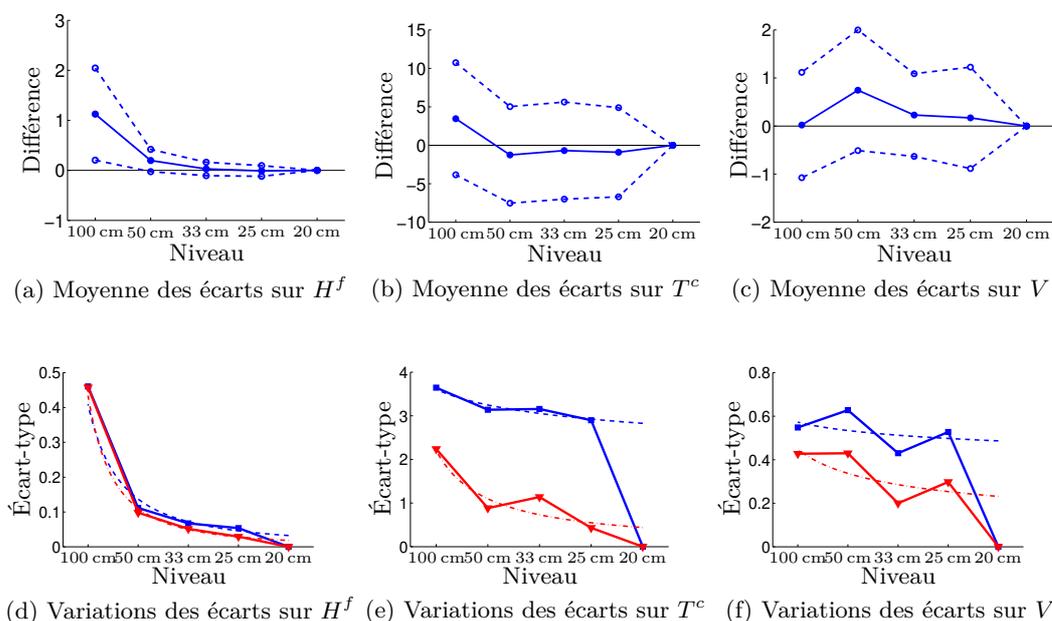


FIGURE 59 – Évolutions des différences entre les sorties aux différents niveaux et des sorties au niveau 20 cm. En haut, moyennes et intervalles à deux écarts-types des différences. En bas, écarts-types en fonction du niveau. Sur chaque figure du bas, la ligne du haut en bleu correspond aux écarts-types totaux des différences et celle du bas en rouge aux écarts-types ôtés des écarts-types de l'effet stochastique estimés par des modèles de processus gaussiens stationnaires. Les courbes discontinues sont des régressions logarithmiques sur les quatre points des quatre niveaux de plus basse fidélité.

différences. Une tendance sur l'évolution de la moyenne des différences apparaît sur H^f . Par contre, aucune tendance particulière ne semble apparaître sur les deux autres sorties.

Afin de mieux apprécier l'évolution des variations de l'écart-type des différences en fonction des niveaux, celles-ci sont représentées figures 59d à 59f. Deux courbes sont représentées : l'écart-type total de la différence, $\sqrt{\text{var}[Z_{x_i,\delta} - Z_{x_i,\delta^{\text{ref}}}]}$; et l'écart-type résiduel ôté de la variance de la partie stochastique, $\sqrt{\text{var}[Z_{x_i,\delta} - Z_{x_i,\delta^{\text{ref}}}] - \lambda_\delta - \lambda_{\delta^{\text{ref}}}}$. L'écart-type résiduel est une mesure grossière des variations dues à la variation du terme de moyenne en fonction des entrées x . Les variances des parties stochastiques sont constantes à un niveau donné. Elles sont estimées par maximum de vraisemblance sur les modèles de processus gaussiens stationnaires utilisés sur la figure 57. Les courbes en pointillées sont des régressions logarithmiques sur les quatre premiers niveaux, destinées à afficher la tendance. Il apparaît que, sur les trois sorties, l'écart-type des différences

tend à se réduire avec le niveau. Cependant, la vitesse de décroissance de l'écart-type des différences diffère pour chaque sortie.

3.2.3 Quantités d'intérêt

Dans une étude ISI, l'objectif est d'évaluer la conformité du système de désenfumage dans différents scénarios d'incendie. La conformité s'estime par comparaison entre les trois sorties du modèle numérique $(Z_a)_{1 \leq a \leq 3}$ avec leurs seuils critiques respectifs $(z_a^{\text{crit}})_{1 \leq a \leq 3}$ (voir tableau 5). Cependant, à cause de sources d'incertitudes en entrées et en sorties, la notion de conformité en un scénario donné ne peut pas être binaire.

Dans cette thèse, la façon d'évaluer la conformité qui a été retenue est d'estimer des probabilités de dépassement de seuil. D'une part, comme les sorties en un point (x, δ) donné sont considérées comme stochastiques (voir section 3.2.2), il est possible de calculer une probabilité locale de dépassement de seuil pour chaque sortie, $p_a(x, \delta)$. Sur notre modèle numérique fondé sur FDS, les fonctions de probabilité de défaillance sont

$$p_{H^f}(x, \delta) = \mathbb{P}(H^f \geq 2,5 \text{ kW} \cdot \text{m}^{-2} | x, \delta); \quad (3.7a)$$

$$p_{T^c}(x, \delta) = \mathbb{P}(T^c \geq 60 \text{ °C} | x, \delta); \quad (3.7b)$$

$$p_V(x, \delta) = \mathbb{P}(V \leq 5 \text{ m} | x, \delta). \quad (3.7c)$$

D'autre part, certaines des entrées du modèle numérique, les entrées environnementales x_e , sont considérées comme aléatoires par opposition aux variables de scénario x_s qui sont connues dans un scénario donné. Ces variables environnementales sont supposées suivre une distribution environnementales f^{env} , représentant l'incertitude sur ces valeurs d'entrées. Dans notre cas d'étude, la distribution environnementale f^{env} est une loi normale dont les paramètres sont donnés (3.3). Dans ces conditions, la probabilité de dépassement de seuil à mesurer est la probabilité conditionnellement aux variables de scénario. Les quantités d'intérêt sont alors

$$P_a(x_s, \delta_{xyz}) = \int_{\mathbb{X}_e} p_a(x_e, x_s, \delta_{xyz}) f^{\text{env}}(x_e) dx_e, \quad (3.8)$$

où $a \in \{H^f, T^c, V\}$ indique la sortie considérée. Cependant, les probabilités de défaillance n'ont d'intérêt que sur des niveaux de suffisamment haute-fidélité. Il est inutile de mesurer la probabilité de défaillance en un niveau jugé trop grossier pour bien modéliser l'incendie. En pratique, nous chercherons à estimer les probabilités de défaillance particulièrement en un niveau de fidélité de référence δ^{ref} , $P_a(x_s, \delta = \delta^{\text{ref}})$.

Enfin, dans le cas où les sorties du simulateur sont supposées suivre une loi normale, $Z_a|x, \delta \sim \mathbf{N}(\xi^{(a)}(x, \delta), \lambda^{(a)}(x, \delta))$, les probabilités locale peuvent se réécrire en fonction de la fonction de répartition de la loi normale centrée réduite Φ . Nos fonctions d'intérêt deviennent alors

$$\forall a \in \{H^f, T^c, V\}, \quad P_a(x_s) = \int_{\mathbb{X}_e} \Phi\left(\frac{\xi^{(a)}(x_e, x_s, \delta^{\text{ref}}) - z_a^{\text{crit}}}{\sqrt{\lambda^{(a)}(x_e, x_s, \delta^{\text{ref}})}}\right) f^{\text{env}}(x_e) dx_e. \quad (3.9)$$

L'objectif est d'évaluer ces fonctions de probabilité de défaillance en fonction des variables de scénario x_s , en proposant un estimateur et une mesure d'incertitude, tout en minimisant le coût en temps de calcul du simulateur.

3.3 Illustration de la méthodologie proposée

Dans cette section, nous illustrons la méthodologie sur FDS. Nous partons d'un plan initial d'expériences, et nous présentons différents résultats qu'il est possible d'obtenir. Par opposition, dans la section 3.4, nous répéterons plusieurs fois la méthodologie à partir de plusieurs plans initiaux, et comparerons les résultats à des résultats issus de données de validation.

Dans les sections 3.3 et 3.4, toutes les prédictions sont des médianes a posteriori de la quantité étudiée, et tous les intervalles de crédibilité sont des intervalles de plus hautes densités, c'est-à-dire les intervalles minimisant la distance entre les deux bornes parmi les intervalles de même probabilité.

3.3.1 Plan d'expériences initial

Dans cette section, nous présentons différents résultats qu'il est possible d'obtenir à partir d'un plan d'expériences. Le plan d'expériences initial décrit tableau 6 est constitué de 130 observations réparties sur trois niveaux de fidélité. Son coût total d'observation est de $90 \times 69,3 \text{ min} + 30 \times 6,05 \text{ h} + 10 \times 19,6 \text{ h} = 20 \text{ j}$, soit l'équivalent du temps nécessaire pour faire dix observations sur le niveau 20 cm. Ce plan d'expériences initial est un Hypercube Latin Emboîté (NLHS) optimisant le remplissage d'espace (voir section 2.2), à la nuance près que les surfaces du feu A_f sont arrondies au mètre carré (à cause des contraintes que FDS impose à ses entrées, voir section 3.2.1). Pour chacune des entrées, les trois sorties de FDS sont observées. La figure 60 présente les valeurs des sorties en fonction de la taille de maille.

Chacune des trois sorties est analysée indépendamment. Pour chacune des sorties,

Niveaux	100 cm	50 cm	33 cm	25 cm	20 cm	Total
Nombre de points	0	90	30	10	0	130
Temps calculatoire	4,1 min	69,3 min	6,05 h	19,6 h	48,6 h	20 j
Temps comparés à 20 cm	1/712	1/42	1/8.0	1/2.5	1	9.9

TABLE 6 – Plan d’expériences initial utilisé pour l’application FDS.

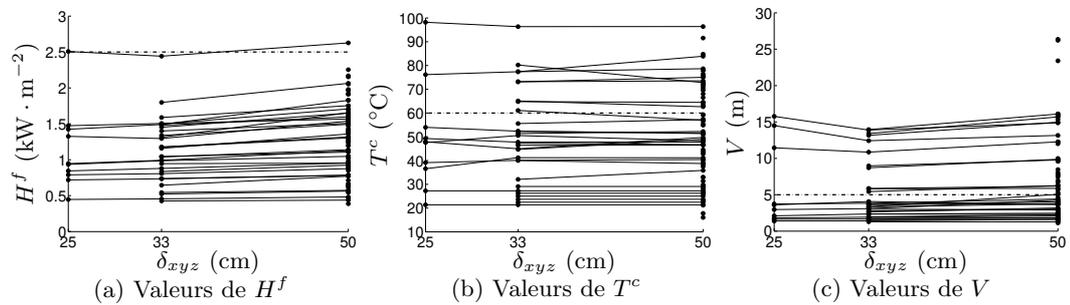


FIGURE 60 – Plan initial d’expériences sur FDS. Les points noirs sont les valeurs des sorties en fonction du niveau de fidélité. Les traits horizontaux pointillés sont les seuils critiques. Deux points de niveaux de fidélité différents sont reliés quand ils sont observés aux mêmes entrées x .

un modèle bayésien fondé sur des processus gaussiens est utilisé. La partie moyenne de la sortie ξ est modélisée par un processus gaussien multi-fidèle non-stationnaire décrit section 1.2.2, la partie variance de l'effet stochastique en sortie est modélisée par le modèle de variance hétéroscédastique présenté section 1.3. Les fonctions de covariance utilisées dans le modèle multi-fidèle non-stationnaire sont des fonctions de covariance de Matérn de régularité $5/2$. La fonction de covariance selon le paramètre de fidélité δ_{xyz} est une fonction de covariance brownienne déformée (1.20). Les hyper-paramètres θ du modèle sont constitués des hyper-paramètres de la fonction de covariance et des variances du bruit à chaque niveau. Ils sont au nombre de $2d+S+3 = 25$, où $d = 8$ est la dimension du vecteur x en entrée, et $S = 6$ est le nombre de niveaux de fidélité considéré (les cinq niveaux du tableau 6, plus le niveau 0 cm). Les lois a priori sur les hyper-paramètres sont présentées annexe D.

3.3.2 Échantillonnage des hyper-paramètres

Pour mener une étude complètement bayésienne, nous échantillonnons 100 hyper-paramètres selon leur loi a posteriori, en utilisant l'algorithme de Métropolis adaptatif de Haario et al. [2001]. Afin de vérifier si l'échantillonnage se déroule correctement, celui-ci est lancé deux fois depuis la valeur Maximum A Posteriori (MAP) qui sert de graine aux chaînes de Markov.

La figure 61 affiche les lois a posteriori de cinq des vingt-cinq hyper-paramètres, estimées à partir des cent échantillons et comparées à leurs lois a priori. La figure dévoile que les lois des hyper-paramètres a posteriori sont moins dispersées que leurs équivalents a priori, et que les deux échantillonnages donnent à peu près la même loi a posteriori. La figure 62 présente les chaînes après le temps de chauffe qui ont permis d'obtenir les échantillons. La figure affiche des chaînes qui semblent bien mélangées, sauf sur les hyper-paramètres G et L , qui semblent plus difficiles à échantillonner. La figure montre également que les deux chaînes de Markov semblent similaires. Ceci laisse à penser que la méthode d'échantillonnage est robuste. Notons que certaines valeurs du MAP de la loi jointe sont positionnée à l'extrémité de leur loi marginale associée, par exemple sur les figures 61i, 61k et 61o. Ceci laisse à penser que les lois a posteriori dans l'espace des hyper-paramètres sont multi-modales.

Dans la suite, seul le premier échantillon est utilisé pour calculer les prédictions et évaluer la conformité. Remarquez que seuls cinq des vingt-cinq hyper-paramètres sont présentés dans cette section, mais que cette analyse a été menée pour tous les hyper-paramètres échantillonnés.

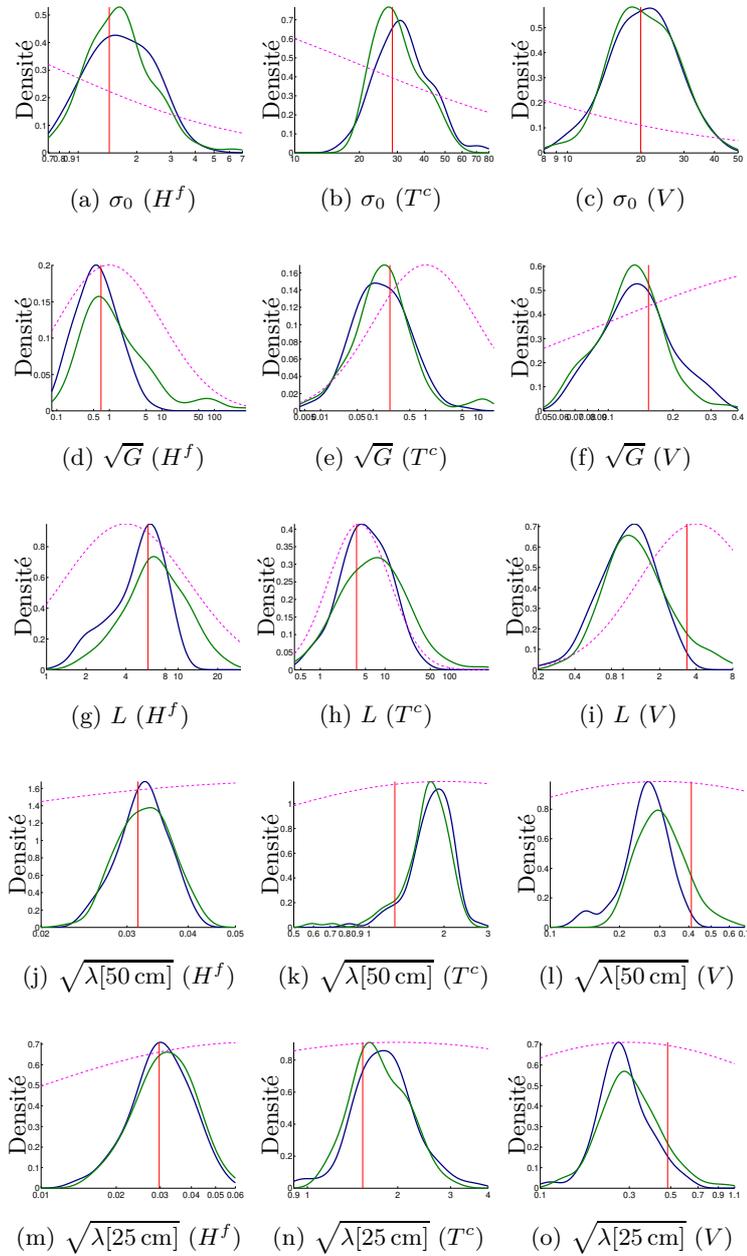


FIGURE 61 – Les lois marginales de quelques-uns des hyper-paramètres des modèles des sorties de FDS. Chaque ligne correspond à un hyper-paramètre, chaque colonne à une sortie. Les courbes continues bleues et vertes sont des estimations des densités a posteriori des hyper-paramètres, issues de deux échantillonnages indépendants. Les courbes pointillées mauves sont les densités a priori, normalisées pour que le maximum de la distribution a priori soit égal au maximum de la distribution a posteriori. Les traits continus verticaux rouges indiquent le MAP de la loi jointe.

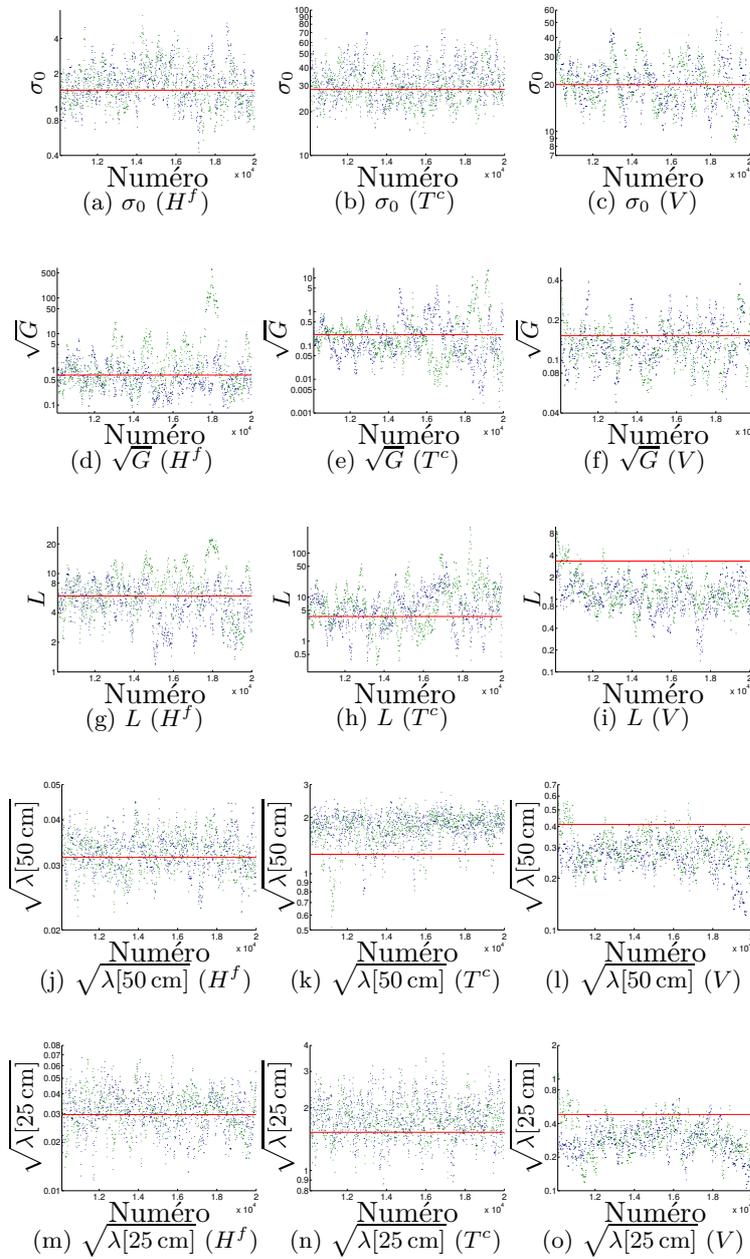


FIGURE 62 – Quelques-unes des chaînes ayant engendré les échantillons des hyper-paramètres des modèles associés aux sorties de FDS. Chaque ligne correspond à un hyper-paramètre, chaque colonne à une sortie. Chaque couleur correspond à une chaîne indépendante. Les traits continus horizontaux rouges indiquent le MAP de la loi jointe.

Entrée	T_{ext}	P_{atm}	T_{amb}	α	A_f	\dot{Q}''	q_{fd}	Y_{soot}
Unité	°C	kPa	°C	$\text{kW} \cdot \text{s}^{-2}$	m^2	$\text{kW} \cdot \text{m}^{-2}$	$\text{MJ} \cdot \text{m}^{-2}$	$\text{kg} \cdot \text{kg}^{-1}$
x^{ref}	10	100	22.5	0.1057	14	460	450	0.027

(a) Scénario de référence

Entrée	T_{ext}	P_{atm}	T_{amb}	α	A_f	\dot{Q}''	q_{fd}	Y_{soot}
Unité	°C	kPa	°C	$\text{kW} \cdot \text{s}^{-2}$	m^2	$\text{kW} \cdot \text{m}^{-2}$	$\text{MJ} \cdot \text{m}^{-2}$	$\text{kg} \cdot \text{kg}^{-1}$
x_1	22.5253	98.3838	17.5000	0.199047	1	422.727	359.091	0.051869
x_2	-2.5253	100.9290	29.7727	0.016102	18	440.909	340.909	0.118796
x_3	4.7475	100.8080	19.0152	0.107575	17	392.424	440.909	0.044061

(b) Points de régression utilisés sur la figure 65

TABLE 7 – Différentes points de références utilisés pour les prédictions.

3.3.3 Prédiction des sorties

Une fois des échantillons des hyper-paramètres obtenus, il est possible de prédire différentes quantités d'intérêt. L'une de ces quantités d'intérêt est la valeur de la fonction moyenne ξ_a et des sorties Z_a , $a \in \{H^f, T^c, V\}$, en fonction des entrées x .

La figure 63 présente une prédiction de la fonction moyenne pour les trois sorties en fonction de deux entrées, la surface du feu A_f et la puissance surfacique du feu \dot{Q}'' , à 20 cm, les autres entrées étant fixées à la valeur de référence affichée dans le tableau 7a. Les prédictions affichées sont calculées sur une grille de 77×76 points (une prédiction tous les $0,25 \text{ m}^2$ pour la surface et les $4 \text{ kW} \cdot \text{m}^{-2}$ pour la puissance). La figure montre des résultats cohérents, puisque plus le feu est grand et puissant, plus le hall devient dangereux (le flux radiatif et la température croissent et la visibilité décroît). Les sorties de FDS semblent monotones en fonction de la taille du feu A_f et de la puissance surfacique \dot{Q}'' , à une nuance près pour la visibilité, qui semble changer de sens de croissance pour $A_f = 12 \text{ m}^2$ et $\dot{Q}'' = 350 \text{ kW} \cdot \text{m}^{-2}$ sur la figure 63c. A ce stade, l'origine la plus probable de ce résultat surprenant est l'incertitude sur l'estimation de la médiane du fait du nombre limité d'observations.

La figure 64 affiche des résultats similaires, mais selon une seule entrée, la surface du feu. La puissance du feu est fixée à $\dot{Q}'' = 460 \text{ kW} \cdot \text{m}^{-2}$. Les autres entrées sont fixées à leur valeur de référence du tableau 7a. La figure représente la prédiction des trois sorties selon la surface du feu à deux niveaux de fidélité. Elle montre les prédictions des fonctions de moyenne ξ_a , les intervalles de crédibilité à 95 % sur la prédiction de ξ_a , et l'intervalle de crédibilité à 95 % sur les sorties de FDS. Ce dernier intervalle prend en compte l'incertitude due à l'effet stochastique λ_a , l'incertitude sur la valeur de la

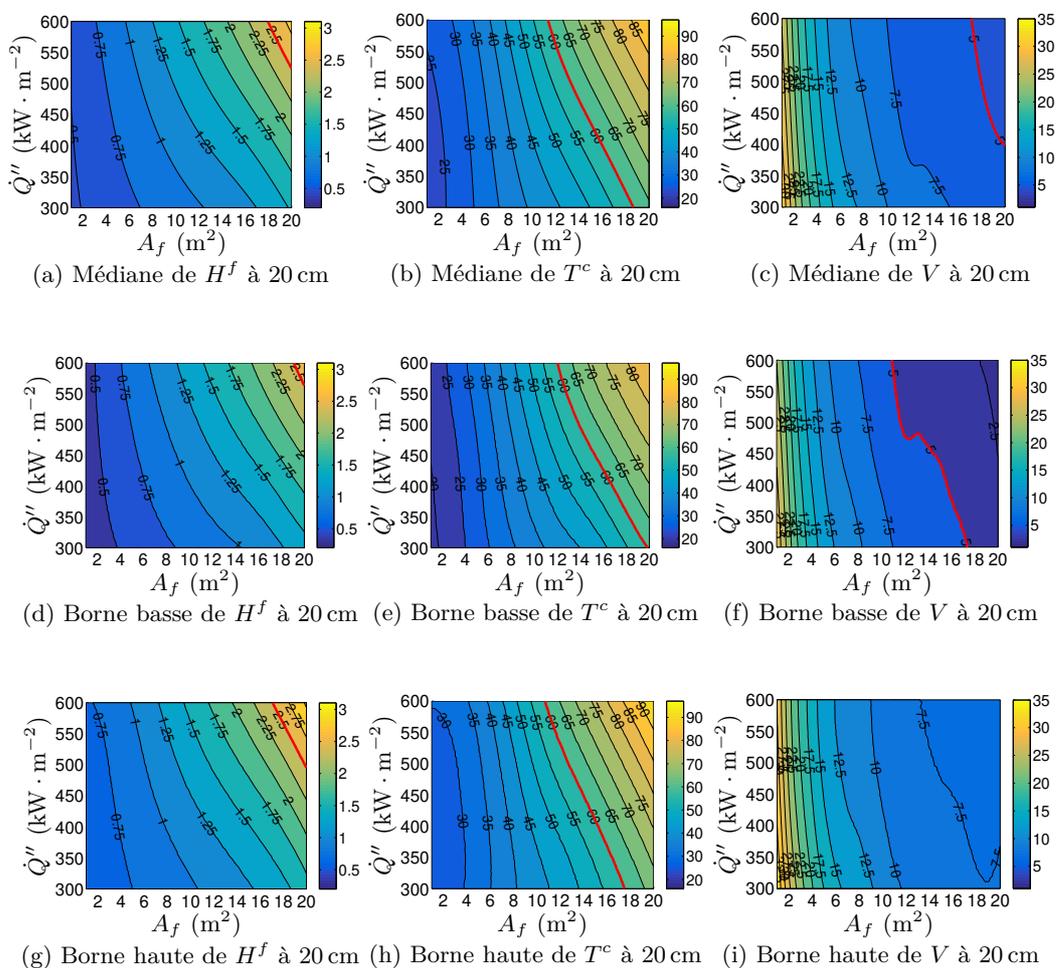


FIGURE 63 – Prédictions des sorties de FDS selon la surface du feu A_f et la puissance surfacique \dot{Q}'' . Chaque colonne correspond à une sortie, chaque ligne à un type de résultat. La première ligne présente les médianes à posteriori, la deuxième les bornes de crédibilité inférieure à 95 % sur l'estimation de la moyenne, la troisième les bornes de crédibilité supérieure à 95 % sur l'estimation de la moyenne. Les intervalles de crédibilité sont les intervalles de plus haute densité a posteriori. Les traits épais rouges correspondent aux seuils critiques. Le niveau de fidélité correspond à $\delta_{xyz} = 20$ cm.

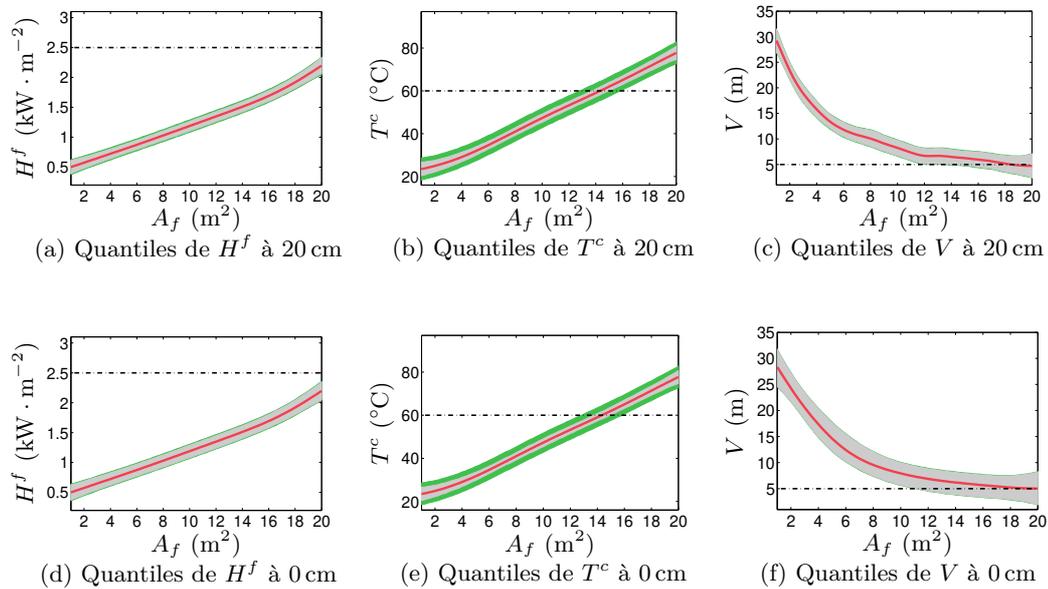


FIGURE 64 – Prédictions des sorties de FDS selon la surface du feu A_f . Chaque colonne correspond à une sortie, chaque ligne à un niveau de fidélité. Les traits pleins sont les médianes à posteriori, et les intervalles sont les bornes de crédibilité à 95 %. La zone interne grise est l'intervalle de crédibilité sur la prédiction de la valeur moyenne, la zone externe verte est l'intervalle de crédibilité sur les futures observations, en rajoutant la partie stochastique du simulateur. Les traits pointillés horizontaux correspondent aux seuils critiques.

fonction moyenne ξ_a , l'incertitude sur la valeur de la variance de l'effet stochastique λ_a et l'incertitude sur les hyper-paramètres θ . De plus, la figure confirme encore notre intuition sur le comportement des sorties. Enfin, en comparant les prédictions au niveau 20 cm et celles du niveau 0 cm, la figure met en évidence que les prédictions restent relativement similaires, avec une augmentation de la taille des intervalles de crédibilité entre le niveau 20 cm et le niveau 0 cm. Ce résultat est cohérent avec le modèle bayésien utilisé. Sur la figure 64, le flux radiatif n'atteint jamais son seuil critique, la température dépasse son seuil vers pour des feux de surface supérieur à 14 m^2 et la visibilité prédite pour des feux très grand est proche de son seuil critique. Ces différentes prédictions se traduisent par des comportements différents dans l'estimation des courbes de probabilités (figure 67).

Enfin, il est possible de tracer les prédictions des sorties de FDS en fonction de la fidélité δ_{xyz} . Trois vecteurs d'entrées, présentés tableau 7b sont ainsi choisis, et nous prédisons les sorties en ces entrées pour une fidélité comprise entre 0 cm et 50 cm. Nous

disposons en ces points sur quatre niveaux de fidélité de vraies sorties de FDS qui n'ont pas été utilisées pour construire le modèle. Les prédictions sont affichées figure 65. Les intervalles de crédibilité incluant la partie stochastique (zone extérieure verte) sont calculées avec la valeur estimée de la variance de l'effet stochastique en 0 cm. Les figures montrent que les prédictions arrivent à bien suivre les vraies valeurs, car les croix sont à l'intérieur des intervalles de crédibilités. Les figures révèlent également que, pour le flux radiatif H^f et la visibilité V , les modèles arrivent à trouver une tendance en fonction de la fidélité pour deux des trois points affichés, mais que par contre, il n'y a pas de tendance générale sur l'évolution en fonction de la fidélité pour la température T^c . La figure 65i illustre un cas où la tendance trouvée n'est pas forcément strictement monotone.

3.3.4 Distribution a posteriori d'une probabilité de défaillance

Une autre quantité d'intérêt est la probabilité de défaillance en un scénario donné (3.9). Pour cela, nous échantillons la loi a posteriori des trois probabilités de défaillance en un scénario (une par sortie), selon la méthode décrite annexe C, avec les 100 hyper-paramètres, 1000 simulations conditionnelles par hyper-paramètres et 400 points environnementaux pour l'intégration. Nous obtenons ainsi $100 \times 1000 = 10^5$ échantillons des trois probabilités de défaillance en un scénario. La distribution des variables environnementales est une loi gaussienne multivariée dont les paramètres sont données (3.3). Les variables de scénario sont fixés à leur valeur décrite dans le tableau 7a.

La figure 66 présente les lois a posteriori des probabilités de défaillance sur deux niveaux de fidélité. Pour ce scénario, les échantillons de probabilités de défaillance de la sortie H^f sont tous à 0 % ; la loi a posteriori de la probabilité de défaillance de la sortie T^c est uni-modale, et estimée entre 40 % et 55 % ; et la loi a posteriori de la probabilité de défaillance de la sortie V est estimée inférieure à 20 %. Remarquez que, sur la figure 66f, l'intervalle de crédibilité à 95 % fait tout l'intervalle, mais que les valeurs échantillonnées sont pour la majorité proche de 0 % ou de 100 %. Cela correspond à un cas où l'incertitude sur la valeur moyenne est très grande devant l'incertitude due à l'effet stochastique en sortie. En effet, les figures 64c et 64f montrent que pour le scénario considéré que la prédiction de la sortie est proche du seuil critique, avec une incertitude sur la valeur moyenne ξ_V très importante devant la partie stochastique. Cela signifie que, parmi les trajectoires a posteriori du processus modélisant V , beaucoup de ces trajectoires sont au dessus ou en dessous du seuil à plusieurs fois $\sqrt{\lambda_V}$, ce qui donne des estimations de probabilités à zéro ou un. Dans ces conditions, la valeur estimée (ici à 0 %) n'a plus grand sens.

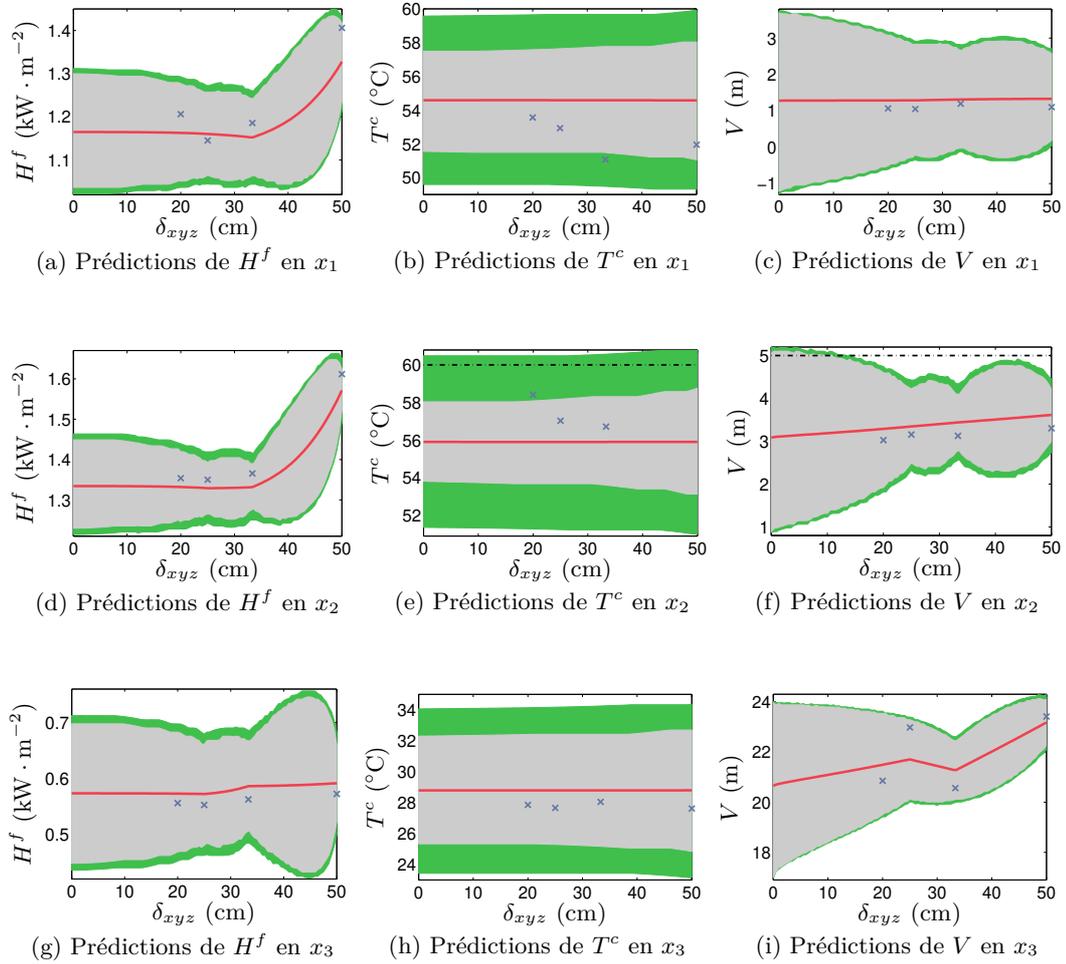


FIGURE 65 – Prédictions des sorties de FDS selon la taille de maille δ_{xyz} . Chaque colonne correspond à une sortie, chaque ligne à un vecteur d'entrées. Les traits pleins sont les médianes à posteriori, et les intervalles sont les bornes de crédibilité à 95 %. La zone interne grise est l'intervalle de crédibilité à 95 % sur la prédiction de la valeur moyenne, la zone externe verte est l'intervalle de crédibilité à 95 % sur les futures observations, en rajoutant la partie stochastique du simulateur. Les croix bleues sont des vraies sorties, qui n'ont pas été utilisées pour construire le modèle. Les traits pointillés horizontaux correspondent aux seuils critiques. Les valeurs des points x_1 , x_2 et x_3 sont affichées tableau 7b.

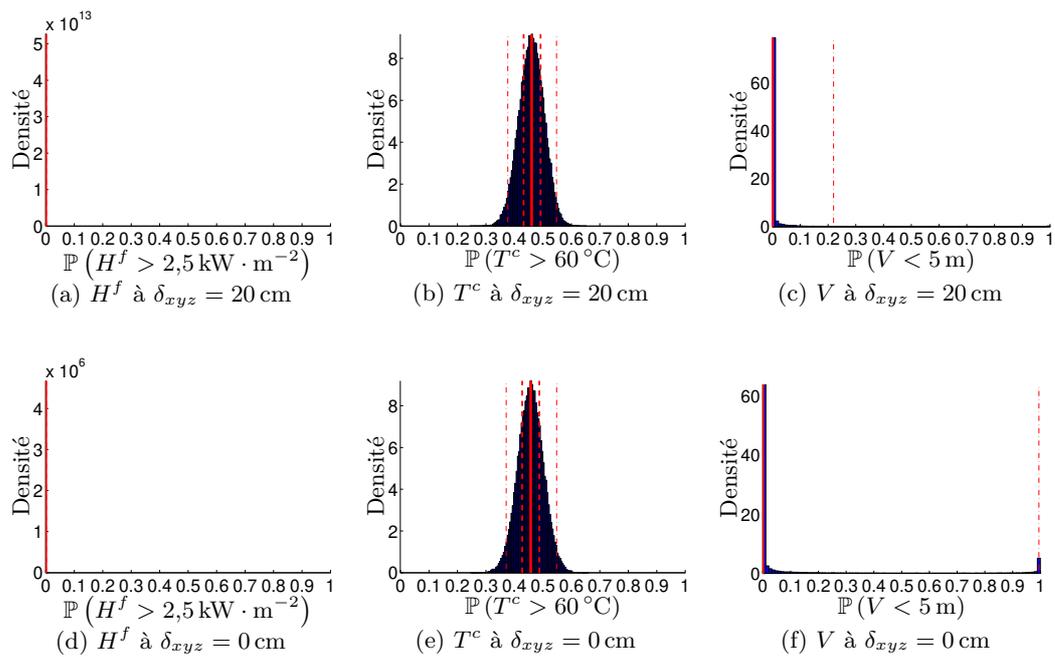


FIGURE 66 – Densité a posteriori de probabilités de défaillance en un scénario. Chaque colonne correspond à une sortie, chaque ligne à un niveau de fidélité. Les histogrammes sont des estimations de la loi a posteriori à partir de 100×1000 échantillons. Les traits pleins verticaux rouges sont les médianes, les traits discontinus verticaux rouges correspondent aux intervalles de crédibilité à 50 %, les traits pointillés verticaux rouges aux intervalles de crédibilité à 95 %.

3.3.5 Courbes de probabilités de défaillance

Toujours dans l'optique d'évaluer la conformité nous pouvons chercher à estimer des probabilités de défaillance en fonction d'une des variables de scénario. Dans cette section, nous estimons des probabilités de défaillance pour 96 valeurs de la surface du feu A_f (une valeur toutes les $0,2\text{ m}^2$) sur deux niveaux de fidélité. Les autres variables de scénario sont fixées à leurs valeurs de référence (tableau 7a). Les entrées environnementales sont intégrées selon leur distribution. Pour chacun de ces scénarios, la méthodologie de l'annexe C est appliquée.

La figure 67 présente les courbes de probabilité de défaillance ainsi obtenues, pour les trois sorties, sur les deux niveaux de fidélité. L'estimateur est la médiane a posteriori et les intervalles de crédibilité à 95 % et 99 % sont les intervalles de plus haute densité a posteriori. Les résultats sont similaires à ceux obtenues pour l'évaluation des sorties de FDS (figure 64). D'abord, plus le feu est grand, plus le scénario devient critique. Ensuite, les intervalles de crédibilité au niveau 0 cm sont plus grandes que celle au niveau 20 cm. Un des intérêts de ces graphes est qu'ils permettent de voir le scénario critique pour chaque sortie. Ainsi, les autres variables de scénario étant fixées, la surface de feu critique est aux alentours de 19 m^2 pour le flux radiatif H^f , 10 m^2 pour la température T^c et 10 m^2 pour la visibilité V . Sur les courbes de probabilités de défaillance associées à la visibilité, les situations où l'intervalle de crédibilité représente la totalité de l'intervalle $[0; 1]$ correspondent à la situation de la figure 66f : les probabilités échantillonnées sont majoritairement autour de 0 % ou 100 %, car la prédiction de la visibilité est proche du seuil critique avec une grande incertitude sur la valeur de la moyenne ξ_V (voir figures 64c et 64f).

3.3.6 Planification séquentielle d'expériences

Enfin, il est possible d'améliorer les estimations de probabilités de défaillance en ajoutant des observations. Dans cette section, nous cherchons à améliorer l'estimation de la probabilité de défaillance T^c en un scénario de référence donné. Les variables de scénario de référence sont fixées à leur valeur de référence (tableau 7a). Le niveau de fidélité où la probabilité doit être évaluée est $\delta_{xyz} = 20\text{ cm}$.

Pour cela, nous ajoutons les observations une par une, en partant du plan initial. Chaque nouveau point est choisi en utilisant la stratégie du Taux Maximal de Réduction de l'Incertitude (MRUR) présentée dans la section 2.4.2. Tous les nouveaux points sont ajoutés pour la variable de scénario de référence, la stratégie MRUR choisissant les entrées environnementales et le niveau de fidélité δ_{xyz} . Le critère Réduction Progressive

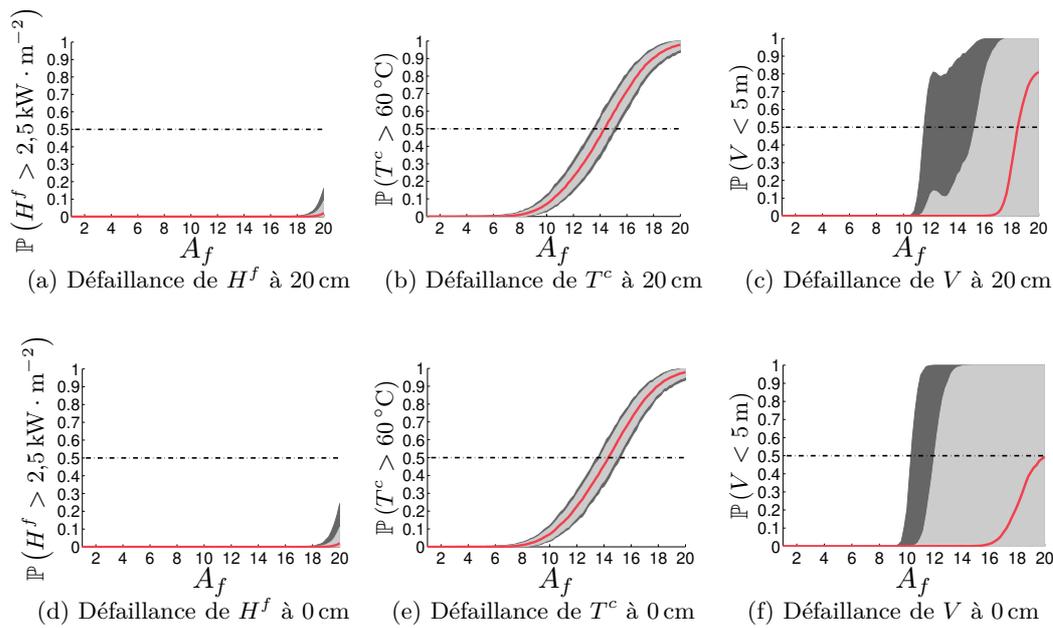


FIGURE 67 – Estimations des probabilités de défaillance des trois sorties de FDS en fonction de la surface du feu A_f . Chaque colonne correspond à une sortie, chaque ligne à un niveau de fidélité. Les courbes pleines sont les médianes a posteriori, les zones internes les intervalles de crédibilité à 95 % et les zones externes les intervalles de crédibilité à 99 %.

Nombre d'observations	Coût	Médiane	Intervalle à 50 %	Intervalle à 95 %
130 (plan initial)	20 j	46,1 %	[43,2 % ; 49,2 %]	[37,7 % ; 54,8 %]
169 (plan final - bayésien)	29,6 j	50,7 %	[49,2 % ; 52,6 %]	[45,5 % ; 55,5 %]
169 (plan final - MAP)	29,6 j	49,1 %	[47,8 % ; 50,2 %]	[45,7 % ; 52,8 %]

TABLE 8 – Comparaison de l'estimation de la probabilité de défaillance de la température T^c pour les variables de scénario de référence (tableau 7a) entre un plan initial multi-fidèle à 130 observations et un plan séquentiel multi-fidèle à 169 observations. Les deux lignes correspondant à l'estimation avec 169 observations avec deux méthodes différentes : la ligne « bayésien » correspond à des résultats où les hyper-paramètres sont intégrés selon leur loi a posteriori, alors que la ligne « MAP » correspond à des résultats où les hyper-paramètres sont fixés à leur valeur MAP.

de l'Incertitude (SUR) utilisé dans la stratégie MRUR est le critère SUR 1. Lors du calcul du critère, le vecteur d'hyper-paramètres θ est fixé à sa valeur MAP. Pour assurer que le MAP est bien trouvé, les échantillons des hyper-paramètres sont mis à jour à chaque nouveau point observé en utilisant un algorithme de Monte-Carlo Séquentiel (SMC). Afin de limiter les coûts de calcul, l'intégrale du critère SUR 1 est calculée après élagage des points d'intégration, selon la méthode de Bect et al. [2012] : d'abord, la variance de la probabilité locale est calculée sur 10^4 points environnementaux distribués selon leur distribution environnementale ; puis l'intégrale sur l'espace des points environnementaux est remplacée par une somme sur les 1500 points d'intégration ayant la plus grande variance locale. La planification est poursuivie sur 8,1 j de simulation sur FDS.

La figure 68a présente la loi a posteriori de probabilité de défaillance du scénario ciblé, au niveau 20 cm après planification séquentielle. Cette figure est à comparer avec la figure 66b, qui présente la même figure, mais au plan initial. La comparaison de ces deux figures permet de voir que l'incertitude s'est réduite. Ceci est confirmé par le tableau 8, qui compare les estimations de la probabilités de défaillance en ce scénario entre le plan initial et le plan final. Ce tableau montre notamment que la taille de l'intervalle de crédibilité à 95 % est passée de $54,8\% - 37,7\% = 17,1\%$ à $55,5\% - 45,5\% = 10,0\%$. L'intervalle de crédibilité final reste cependant relativement important, il serait nécessaire de continuer pour obtenir une meilleure estimation de la probabilité de défaillance de T^c en ce scénario.

La comparaison des figures 68a et 68b permet de constater l'impact de l'incertitude sur les hyper-paramètres. Les valeurs sont affichées dans le tableau 8. Ainsi, l'intervalle de crédibilité à 95 % est de longueur 10,0 % pour une estimation complètement bayésienne, alors que l'intervalle de crédibilité à 95 % est de longueur $52,8\% - 45,7\% = 7,1\%$ pour

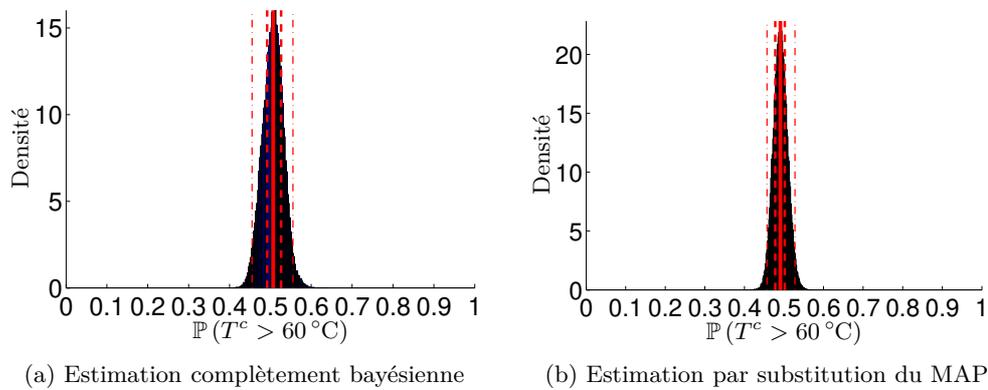


FIGURE 68 – Densité a posteriori de la probabilité de défaillance de la sortie T^c en un scénario, après planification séquentielle. Le nombre d'observations est de 169. Le trait plein vertical rouge est la médiane, les traits discontinues verticaux rouges correspondent à l'intervalle de crédibilité à 50 %, les traits pointillés verticaux rouges à l'intervalle de crédibilité à 95 %. Le niveau de fidélité est $\delta_{xyz} = 20$ cm. (a) Histogramme des échantillons de la loi a posteriori avec intégration des hyper-paramètres. (b) Histogramme des échantillons de la loi a posteriori dans le cas où les hyper-paramètres sont fixés à leur valeur MAP.

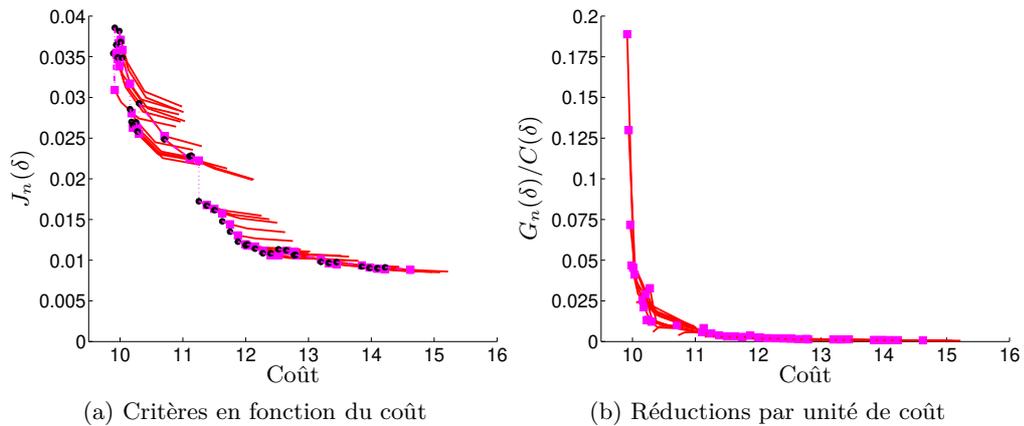


FIGURE 69 – Fronts de Pareto successifs des réductions d’incertitudes sur la planification d’expériences. Les traits rouges sont l’ensemble des valeurs optimales à un niveau donné, les carrés mauves indiquent le niveau qui est choisi. les cercles noirs indiquent les valeurs de H_n utilisées par le critère d’échantillonnage. Le niveau de fidélité ciblé est $\delta_{xyz} = 20$ cm. (a) Fronts de Pareto dans l’espace (C, J) . (b) Fronts de Pareto dans l’espace $(C, G/C)$.

une estimation avec une substitution des hyper-paramètres par leur MAP. Il y a donc environ un tiers de l’incertitude sur l’estimation de cette probabilité de défaillance qui est due à l’incertitude sur les hyper-paramètres.

La figure 69 présente les fronts de Pareto successifs à chaque itération dans l’espace (C, J) et dans l’espace $(C, G/C)$. La figure permet de voir que la mesure d’incertitude se réduit progressivement d’observation en observation, et que la valeur maximale du taux de réduction tend elle aussi à se réduire.

Enfin, la figure 70 présente une comparaison des densité de probabilité a posteriori de la probabilité de défaillance de la visibilité, au scénario de référence (tableau 7a), avec intégration des variables environnementales, obtenue avant et après la planification séquentielle. La figure 70a présente les résultats obtenues avec le plan initial, 130 observations pour un coût de 20 jours. La figure 70b présente les résultats avec le plan amélioré, qui comprend 165 observations pour un coût total de 48.4 jours. Il apparaît que les formes des densités sont différentes : la densité a posteriori avant planification est bimodale en 0 et 1, alors que la densité a posteriori après planification est unimodale. Ce changement de forme est expliqué à la réduction d’incertitude due au méta-modèle, alors que l’incertitude due à l’aspect stochastique du simulateur reste constant.

Ainsi, avant la planification, l’incertitude totale est dominée par l’incertitude due

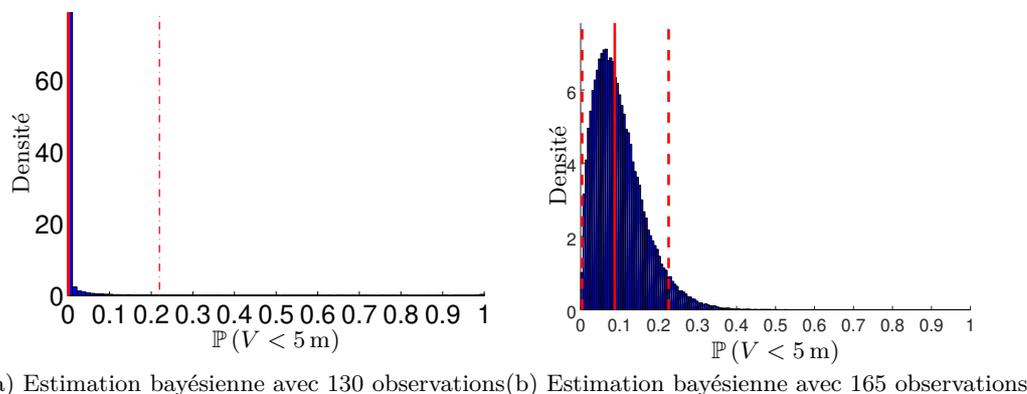


FIGURE 70 – Densité a posteriori de la probabilité de défaillance de la sortie V en un scénario, avant et après planification séquentielle. Le trait plein vertical rouge est la médiane, les traits discontinues verticaux rouges correspondent à l'intervalle de crédibilité à 50 %, les traits pointillés verticaux rouges à l'intervalle de crédibilité à 95 %. Le niveau de fidélité est $\delta_{xyz} = 20$ cm. (a) Histogramme des échantillons de la loi a posteriori avant planification séquentielle, avec 130 observations. (b) Histogramme des échantillons de la loi a posteriori avant planification séquentielle, avec 165 observations.

au modèle de processus gaussien, signifiant que la fonction moyenne ξ est peu connu. Quand elle s'approche du seuil critique, les trajectoires conditionnelles peuvent alors passer soit au-dessus, soit en-dessous du seuil. Ceci donne des modes principalement concentrés en 0 (correspondant aux trajectoires conditionnelles au dessus du seuil), ou en 1 (correspondant aux trajectoires conditionnelles en dessous du seuil). A l'inverse, après la planification, l'incertitude totale est dominée par l'incertitude due au la variance stochastique λ , signifiant que la fonction moyenne ξ est bien connu, comparativement à la variance du bruit. Quand la fonction moyenne s'approche du seuil critique, les trajectoires conditionnelles passent au même endroit. La variabilité qui reste est dû aux observations, qui peuvent être au dessus ou en dessous du seuil critique. Ceci fait apparaître une distribution unimodale sur la densité finale.

Ces figures montrent que la méthode MRUR arrive à réduire l'incertitude liée à la modélisation par processus gaussien.

3.4 Validation de la méthodologie proposée

Dans la section précédente, nous montrons des résultats qu'il était possible d'obtenir à partir d'un unique plan initial d'expériences. Dans cette section, nous vérifions la méthodologie en répétant la méthode et en vérifiant les résultats. Pour cela, nous disposons de plusieurs types de plans d'expériences :

- dix plans d'expériences initiaux suivant la même répartition que dans le tableau 6, qui nous permettent de répéter la méthodologie ;
- un plan Hypercube Latin (LHS) de 199 points de dimension huit au niveau 20 cm, dont les entrées couvrent l'ensemble du domaine d'étude. Ce plan nous sert de données de validation afin de vérifier la qualité de prédiction des sorties ;
- un plan de 150×4 points. Les variables de scénario $(\alpha, A_f, \dot{Q}'' , q_{fd}, Y_{soot})$ sont fixées à leurs valeurs de référence (tableau 7a). Les variables environnementales $(T_{ext}, P_{atm}, T_{amb})$ ont été tirées 150 fois de manière indépendante et identiquement distribuées selon leur distribution environnementale. Les 150 points d'observation ainsi obtenus sont observés sur quatre niveaux de fidélité. Ce dernier plan nous permet d'avoir une estimation par Monte-Carlo de la probabilité de défaillance en ce scénario, pour les trois sorties et quatre niveaux de fidélité.

3.4.1 Comparaison des estimations des hyper-paramètres

Les premières quantités qui sont comparables d'une répétition de la méthodologie à l'autre sont les lois a posteriori des hyper-paramètres sur les dix plans d'expériences initiaux. Les échantillons des hyper-paramètres sont obtenus avec l'algorithme de Métropolis adaptatif de Haario et al. [2001]. Pour chacun de ces plans, le MAP est estimé, et utilisé comme graine de deux chaînes de Markov indépendantes desquelles sont extraits les échantillons.

La figure 71 présente les résultats sur cinq des vingt-cinq hyper-paramètres des modèles. Les résultats sont présentés pour les trois sorties et les dix plans d'expériences. Le plan d'expérience utilisé dans la section 3.3 est le plan numéroté 1. La figure illustre d'abord la robustesse de la méthode, puisque sur tous les plans et toutes les sorties, les deux échantillonnages aboutissent approximativement à la même loi a posteriori. Ensuite, la figure montre que les lois a posteriori des hyper-paramètres sont très étalées. Par exemple, le paramètre L associé à la température T^c (figure 71h) prend des valeurs entre 1 et 50. Il est aussi visible que les lois a posteriori des hyper-paramètres sont différents d'un plan initial à un autre, par exemple sur la variance de l'effet stochastique à 50 cm du flux radiatif H^f (figure 71j) ou sur le paramètre L de la visibilité V (figure 71i). Une

façon d'atténuer ces effets serait d'utiliser des lois a priori plus informatives.

3.4.2 Validation des prédictions des sorties

Une façon de vérifier la qualité du modèle est de comparer les prédictions en certains points d'observations avec des sorties de FDS obtenues en ces points. Nous disposons ainsi d'un LHS de 199 points au niveau 20 cm sur lequel nous avons des valeurs des sorties de FDS, qui nous sert de plan de validation. Ce plan n'a pas été utilisé pour construire les modèles bayésiens.

La figure 72 présente une comparaison entre les prédictions obtenues avec le modèle bayésien multi-fidèle issu du plan d'expérience numéro 1 (utilisé dans la section 3.3) aux points de validations et les sorties de FDS en ces points. Sur la figure, les prédictions obtenues avec ce modèle semblent suivre les observations, car les intervalles de crédibilité coupent la première bissectrice. Les figures équivalentes pour les neuf autres plans d'expériences initiaux sont présentées dans la annexe E.1. Les conclusions restent identiques.

Une façon de mieux quantifier la capacité prédictive des modèles est de calculer des scores de prédiction. Pour cela, nous comparons les prédictions obtenues pour les trois sorties et les dix plans d'expériences aux vraies sorties de FDS. Pour comparer les résultats, nous utilisons deux critères. D'abord, pour vérifier les valeurs des prédictions, nous calculons l'erreur moyenne quadratique

$$\text{EQM}^{1/2} = \sqrt{\frac{1}{n} \sum_{i=1}^n (z_i - \hat{z}_i)^2}, \quad (3.10)$$

où z_i est une donnée de validation et \hat{z}_i est la prédiction au vecteur d'entrées correspondant. Ensuite, pour évaluer la loi prédictive a posteriori F_i , nous utilisons le Score de Probabilité Classée Continue (CRPS)³ présenté par Gneiting et Raftery [2007] et défini comme

$$\text{CRPS} = \frac{1}{n} \sum_{i=1}^n \text{BS}(z_i, F_i), \quad (3.11)$$

où $\text{BS}(z_i, F_i)$ est le score de Brier de la fonction de répartition F_i sur l'observation z_i défini comme

$$\text{BS}(z_i, F_i) = \int_{y=-\infty}^{+\infty} (F_i(y) - \mathbf{1}_{y > z_i})^2 dy. \quad (3.12)$$

Dans le cas où la loi prédictive F_i est une loi normale $N(m_i, \sigma_i^2)$, le score de Brier s'écrit

3. Continuous Probability Ranked Score

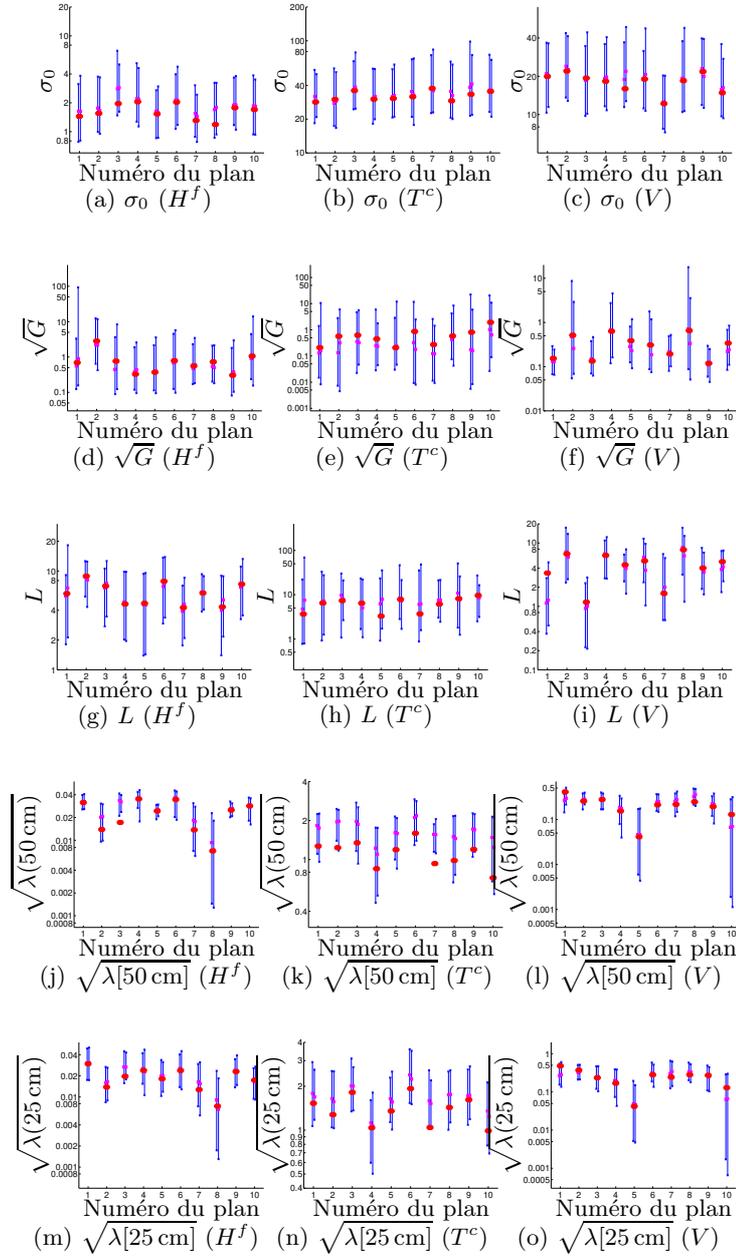


FIGURE 71 – Estimations des hyper-paramètres des modèles associés aux sorties de FDS sur les dix plans d'expériences initiaux. Chaque ligne correspond à un hyper-paramètre, chaque colonne à une sortie. Les points rouges sont les MAP de la loi jointe, les carrés mauves sont les médianes des cent échantillons d'hyper-paramètres et les traits bleus verticaux sont les intervalles de crédibilité a posteriori à 95 % sur les cent échantillons d'hyper-paramètres. Les deux traits pour chaque plan correspondent aux deux échantillonnages indépendants.

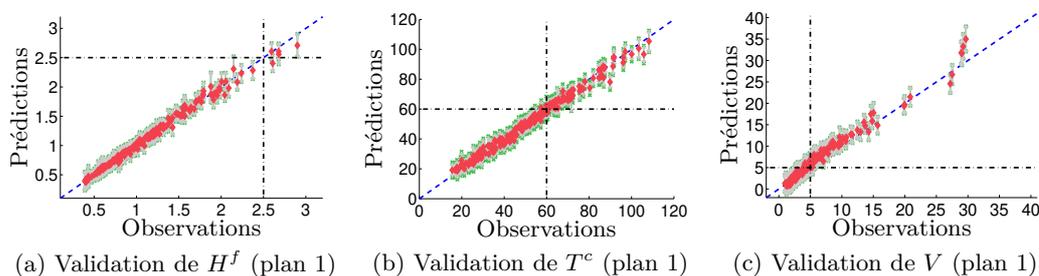


FIGURE 72 – Comparaison entre les observations et les prédictions obtenues par le plan d’expérience numéro 1, à 20 cm. Les points sont les prédictions par médiane à posteriori, les barres correspondent aux intervalles de crédibilité à 95 %, en gris à l’intérieur sur l’estimation de la moyenne en vert à l’extérieur sur les futures observations. Les traits pointillés verticaux et horizontaux correspondent aux seuils critiques. Le trait pointillé oblique bleu est la première bissectrice.

[Gneiting et Raftery, 2007] :

$$\text{BS} \left(z_i, \mathcal{N}(m_i, \sigma_i^2) \right) = (z_i - m_i) \cdot \left[2\Phi \left(\frac{z_i - m_i}{\sigma_i} \right) - 1 \right] + 2\sigma_i \phi \left(\frac{z_i - m_i}{\sigma_i} \right) - \frac{\sigma_i}{\sqrt{\pi}}. \quad (3.13)$$

L’unité du CRPS est celle de la sortie z étudiée.

Nous calculons ces deux scores pour les modèles associés aux dix plans initiaux et aux trois sorties. Afin de comparer à des valeurs de référence, nous calculons ces scores pour trois autres types de modèles : un modèle linéaire en dimension huit (sans la fidélité), fondé sur les 199 données de validation ; un modèle linéaire en dimension huit (sans la fidélité), fondé sur les 90 données du plan initial au niveau 50 cm ; et un modèle linéaire en dimension huit (sans la fidélité), fondé sur les 30 données du plan initial au niveau 33 cm.

Les résultats numériques sont affichés tableau 9. Le tableau met en évidence que les prédictions obtenues par le modèle multi-fidèle bayésien sont bien meilleures que les prédictions par des modèles linéaires. Cela dénote la non-linéarité des sorties de FDS et la capacité du modèle bayésien à capturer ces non-linéarités.

Afin de vérifier la qualité des intervalles de crédibilité à 95 %, nous calculons aussi le taux de couverture empirique des intervalles de crédibilité, c’est-à-dire le nombre fois où les observations se trouvent dans leurs intervalles. Les résultats sont affichés dans la dernière colonne du tableau 9. La colonne montre que les taux de couverture ne sont ni à 100 %, les intervalles ne sont ni trop longs, ni trop courts, les intervalles ne sont pas trop

Modèle	EQM ^{1/2}			CRPS			% Couv. MF-GP	
	L[50]	L[33]	MF-GP	L[50]	L[33]	MF-GP		
<i>H^f</i>	Réf.		0,153			0,0828	95,0 %	
	Plan 1	0,213	0,174	0,0488	0,123	0,0921	0,0254	99,0 %
	Plan 2	0,224	0,169	0,0564	0,130	0,0936	0,0266	91,5 %
	Plan 3	0,230	0,186	0,0530	0,133	0,100	0,0263	95,5 %
	Plan 4	0,234	0,180	0,0535	0,136	0,0997	0,0265	97,0 %
	Plan 5	0,229	0,178	0,0581	0,132	0,0972	0,0285	94,0 %
	Plan 6	0,212	0,163	0,0497	0,123	0,0906	0,0254	93,0 %
	Plan 7	0,208	0,169	0,0740	0,119	0,0934	0,0360	89,9 %
	Plan 8	0,219	0,195	0,0526	0,125	0,104	0,0271	87,4 %
	Plan 9	0,203	0,184	0,0545	0,117	0,103	0,0279	95,5 %
	Plan 10	0,222	0,167	0,0748	0,128	0,0916	0,0352	87,9 %
<i>T^c</i>	Réf.		5,28			2,85	95,0 %	
	Plan 1	5,77	6,15	2,47	3,16	3,42	1,35	97,5 %
	Plan 2	5,65	6,41	2,61	3,07	3,58	1,40	96,0 %
	Plan 3	5,89	6,66	3,00	3,19	3,69	1,57	92,0 %
	Plan 4	5,54	5,74	2,72	3,01	3,18	1,53	83,9 %
	Plan 5	5,81	5,78	2,81	3,19	3,20	1,52	93,5 %
	Plan 6	5,59	5,48	2,61	3,03	2,96	1,43	98,0 %
	Plan 7	5,97	6,16	3,01	3,18	3,20	1,61	91,0 %
	Plan 8	6,02	6,70	2,91	3,21	3,65	1,54	92,5 %
	Plan 9	6,01	7,06	3,02	3,25	3,97	1,60	93,5 %
	Plan 10	5,75	5,92	2,87	3,10	3,29	1,56	87,4 %
<i>V</i>	Réf.		3,39			1,75	95,0 %	
	Plan 1	3,82	3,59	1,10	2,12	1,87	0,577	89,4 %
	Plan 2	3,59	3,97	0,835	1,99	2,25	0,423	89,4 %
	Plan 3	3,50	3,68	1,10	1,88	1,94	0,545	92,0 %
	Plan 4	3,72	3,80	0,806	2,06	2,01	0,369	91,5 %
	Plan 5	3,62	3,49	0,937	1,98	1,88	0,451	85,4 %
	Plan 6	3,62	4,07	0,906	1,97	2,26	0,471	81,9 %
	Plan 7	3,62	3,56	0,922	1,98	1,83	0,457	93,5 %
	Plan 8	3,63	4,00	1,02	1,98	2,02	0,497	88,9 %
	Plan 9	3,66	5,00	1,12	2,04	2,84	0,568	77,4 %
	Plan 10	3,76	3,65	0,907	2,09	1,96	0,422	88,9 %

TABLE 9 – Gauche du tableau : racine carrée de l’erreur quadratique moyenne, et CRPS issues des dix plans d’expériences, sur le plan de validation à 20 cm, sur les trois sorties. La valeur de référence est obtenue avec une régression linéaire sur les données de validation. Les trois colonnes correspondent au modèle linéaire à 50 cm (L[50]), au modèle linéaire à 33 cm (L[33]), et au modèle multi-fidèle fondé sur un processus gaussien (MF-GP). Droite du tableau : taux de couverture des intervalles de crédibilité à 95 % obtenus avec le modèle bayésien multi-fidèle.

courts, à une nuance près pour la visibilité, qui a tendance à sous-évaluer ses intervalles de crédibilité.

3.4.3 Validation des estimations de probabilités de références

Une autre façon de vérifier la méthodologie est de comparer les estimations de probabilités qu'elle obtient avec une autre méthode d'estimation. Nous disposons ainsi d'estimations par une méthode de Monte-Carlo de probabilités de défaillance en un scénario de référence sur les trois sorties sur quatre niveaux de fidélité. Les variables de scénario sont à leur valeur de référence (tableau 7a), et les variables environnementales sont intégrées selon leurs distributions. L'estimateur de Monte-Carlo est calculé avec 150 observations, dont les variables scénarios sont fixées au scénarios de référence et les variables environnementales sont tirées selon leur distribution. Le temps de simulation de FDS pour obtenir cet échantillon est de $150 \times 48,6 \text{ h} = 304 \text{ j}$, soit $304/20 = 15$ fois plus que pour obtenir un plan d'expériences initial. Ces valeurs de référence sont comparées aux estimations obtenues avec les modèles construits à partir des dix plans d'expériences.

La figure 73 illustre les estimations des différentes probabilités de défaillance aux différents niveaux. Le tableau 10 présente les valeurs numériques pour le niveau 20 cm. Pour ce scénario, la probabilité de défaillance du flux radiatif H^f est toujours estimée à 0 %, la probabilité de défaillance de la température T^c est bien estimée, mais la probabilité de défaillance de la visibilité V est assez mal évaluée. En effet, pour la visibilité, il y a trois modèles bayésiens dont les intervalles de crédibilité à 95 % ne recourent pas l'intervalle de confiance à 95 % de la méthode de Monte-Carlo. Ceci est peut-être lié à la fonction de covariance utilisée pour modéliser l'évolution des sorties en fonction de la fidélité, en l'occurrence une fonction de covariance brownienne déformée (1.20). En effet, cette covariance ne cherche pas à prédire le sens de variation des sorties en fonction de la fidélité, ce qui signifie que l'estimation de la sortie en haute-fidélité est fortement liée à celle du dernier niveau observé. Par conséquent, un changement dans la sortie entre deux niveaux de fidélité, par exemple pour la visibilité entre le niveau 25 cm (figure 73i) et le niveau 20 cm (figure 73l), sera difficile à capter avec ce modèle.

3.4.4 Répétition de la planification séquentielle d'expériences

Enfin, la méthode de planification séquentielle d'expériences est répétée sur quatre plans afin de voir le comportement de la stratégie MRUR. La probabilité ciblée est la même que dans la section 3.3 : probabilité que la température T^c dépasse son seuil critique en un scénario de référence, à $\delta^{\text{ref}} = 20 \text{ cm}$. Les variables du scénario de référence

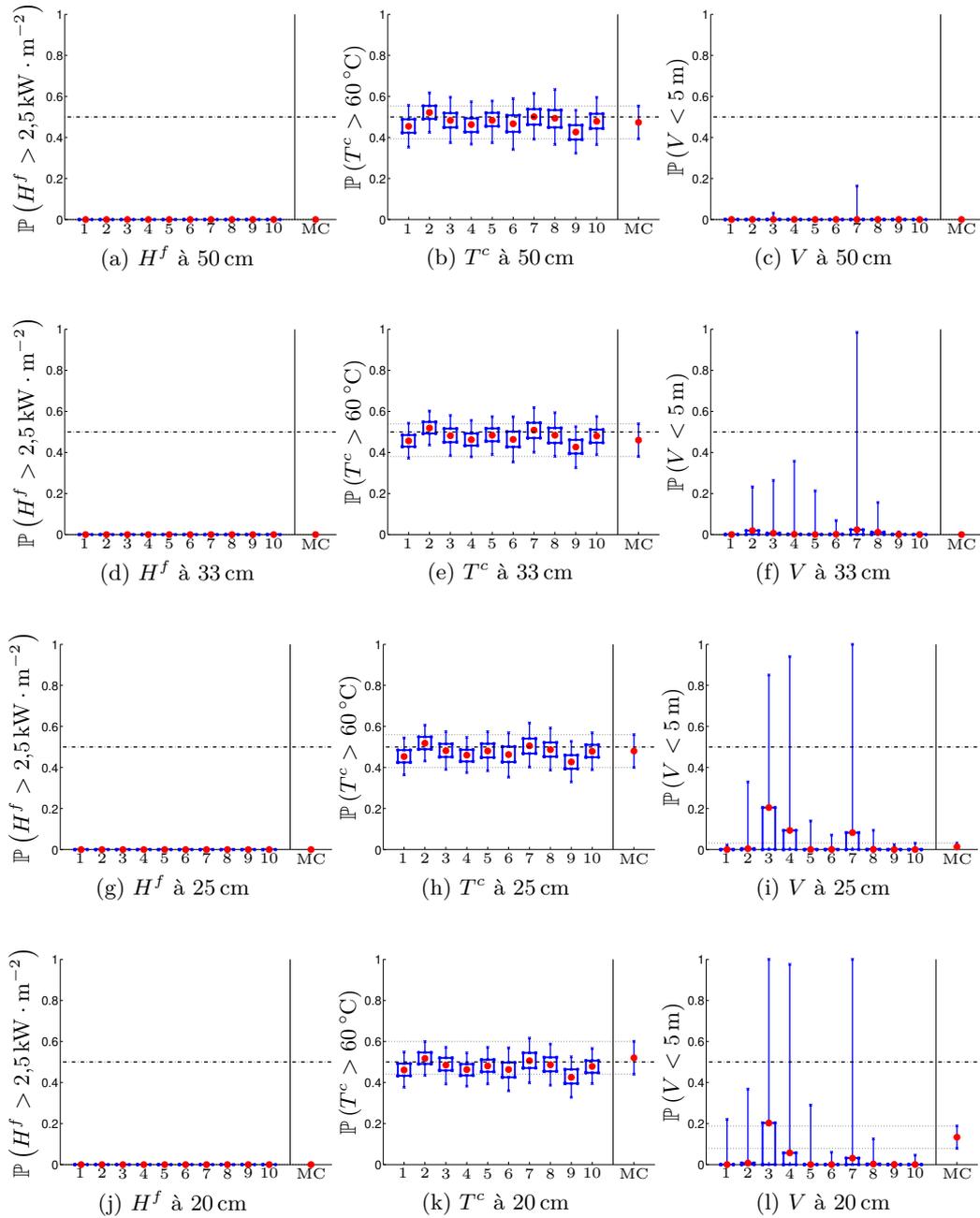


FIGURE 73 – Comparaison des estimations de la probabilité de dépassement de seuil pour les trois sorties par les dix plans d'expériences sur quatre de niveau de fidélité avec des valeurs de référence. Pour les dix plans d'expériences, le point rouge correspond à la médiane a posteriori, le rectangle bleu à l'intervalle de crédibilité à 50 % et le trait bleu à l'intervalle de crédibilité à 95 %. Pour les résultats obtenues par Monte-Carlo sur 150 points, le point rouge correspond à la moyenne, le rectangle bleu à l'intervalle de confiance à 0.67 écart-type et le trait bleu à l'intervalle de confiance à 1.96 écarts-types.

Numéro du plan	Médiane	Intervalle à 50 %	Intervalle à 95 %	
H^f	Plan 1	0,0 %	[0,0 % ; 0,0 %]	[0,0 % ; 0,0 %]
	Plan 2	0,0 %	[0,0 % ; 0,0 %]	[0,0 % ; 0,0 %]
	Plan 3	0,0 %	[0,0 % ; 0,0 %]	[0,0 % ; 0,0 %]
	Plan 4	0,0 %	[0,0 % ; 0,0 %]	[0,0 % ; 0,0 %]
	Plan 5	0,0 %	[0,0 % ; 0,0 %]	[0,0 % ; 0,0 %]
	Plan 6	0,0 %	[0,0 % ; 0,0 %]	[0,0 % ; 0,0 %]
	Plan 7	0,0 %	[0,0 % ; 0,0 %]	[0,0 % ; 0,0 %]
	Plan 8	0,0 %	[0,0 % ; 0,0 %]	[0,0 % ; 0,0 %]
	Plan 9	0,0 %	[0,0 % ; 0,0 %]	[0,0 % ; 0,0 %]
	Plan 10	0,0 %	[0,0 % ; 0,0 %]	[0,0 % ; 0,0 %]
	Monte-Carlo	0,0 %	[0,0 % ; 0,0 %]	[0,0 % ; 0,0 %]
T^c	Plan 1	46,1 %	[43,2 % ; 49,2 %]	[37,7 % ; 54,8 %]
	Plan 2	51,7 %	[49,0 % ; 54,6 %]	[43,5 % ; 59,9 %]
	Plan 3	48,5 %	[45,9 % ; 52,0 %]	[39,3 % ; 57,1 %]
	Plan 4	46,3 %	[43,4 % ; 49,0 %]	[38,2 % ; 54,4 %]
	Plan 5	48,1 %	[45,2 % ; 51,1 %]	[39,3 % ; 57,1 %]
	Plan 6	46,3 %	[42,5 % ; 49,7 %]	[36,0 % ; 56,9 %]
	Plan 7	50,6 %	[47,1 % ; 54,4 %]	[39,9 % ; 61,6 %]
	Plan 8	48,6 %	[45,4 % ; 52,2 %]	[38,7 % ; 58,7 %]
	Plan 9	42,5 %	[39,5 % ; 46,4 %]	[32,8 % ; 52,5 %]
	Plan 10	47,8 %	[44,8 % ; 50,7 %]	[39,5 % ; 56,6 %]
	Monte-Carlo	52,0 %	[49,2 % ; 54,8 %]	[44,0 % ; 60,0 %]
V	Plan 1	0,0 %	[0,0 % ; 0,0 %]	[0,0 % ; 22,0 %]
	Plan 2	0,7 %	[0,0 % ; 0,7 %]	[0,0 % ; 36,8 %]
	Plan 3	20,2 %	[0,0 % ; 20,3 %]	[0,0 % ; 100,0 %]
	Plan 4	5,7 %	[0,0 % ; 5,7 %]	[0,0 % ; 97,5 %]
	Plan 5	0,0 %	[0,0 % ; 0,0 %]	[0,0 % ; 29,0 %]
	Plan 6	0,0 %	[0,0 % ; 0,0 %]	[0,0 % ; 6,0 %]
	Plan 7	3,1 %	[0,0 % ; 3,1 %]	[0,0 % ; 100,0 %]
	Plan 8	0,3 %	[0,0 % ; 0,3 %]	[0,0 % ; 12,5 %]
	Plan 9	0,0 %	[0,0 % ; 0,0 %]	[0,0 % ; 0,9 %]
	Plan 10	0,0 %	[0,0 % ; 0,0 %]	[0,0 % ; 4,6 %]
	Monte-Carlo	13,3 %	[11,5 % ; 15,2 %]	[7,9 % ; 18,8 %]

TABLE 10 – Estimation et intervalle de crédibilité a posteriori des probabilités de défaillance en un scénario de référence, pour les trois sorties, selon les dix plans d'expériences.

sont fixées à leur valeur de référence (tableau 7a). La même méthodologie est appliquée (SMC sur les échantillons d'hyper-paramètres, estimations des probabilités complètement bayésienne, critère d'échantillonnage évaluée avec les hyper-paramètres fixés à leur MAP), en changeant à chaque fois le plan initial d'expériences.

La figure 74 montre les estimations de la probabilité de défaillance et de la mesure d'incertitude en fonction du coût du plan d'expériences. Les estimations affichées sont obtenues après intégration des hyper-paramètres et intégration sur l'espace des entrées environnementales. La mesure d'incertitude est l'intégrale de la variance, décomposable par la formule de la variance totale :

$$H_n = \int_{\mathbb{X}_e} \text{var}_n [p(x_e, x_s, \delta^{\text{ref}})] f^{\text{env}}(x_e) dx_e \quad (3.14a)$$

$$= \int_{\mathbb{X}_e} \mathbb{E}_n [\text{var}_{n,\theta} [p(x_e, x_s, \delta^{\text{ref}})]] f^{\text{env}}(x_e) dx_e \quad (3.14b)$$

$$+ \int_{\mathbb{X}_e} \text{var}_n [\mathbb{E}_{n,\theta} [p(x_e, x_s, \delta^{\text{ref}})]] f^{\text{env}}(x_e) dx_e \quad (3.14c)$$

Les intégrales sont approchées par des sommes sur 100 hyper-paramètres et 10^4 points environnementaux. La figure met en évidence que les estimations de la probabilité de défaillance se rapprochent de la valeur de référence obtenue par Monte-Carlo, et que sur trois expériences, la mesure d'incertitude tend à se réduire. Par contre, il est aussi visible sur la figure 74b, que la mesure d'incertitude associée au plan d'expérience numéro 4 n'est pas décroissante.

Afin de mieux cerner le comportement de ces mesures d'incertitudes, une analyse est menée sur les résultats. Pour chaque plan d'expériences, les suites d'observations obtenues par la stratégie MRUR sont reprises. A chaque nouveau point d'observation, les estimations de probabilités et les mesures d'incertitude sont recalculées suivant trois méthodes : une méthode où les hyper-paramètres sont fixés pour toutes les observations ; une méthode où les hyper-paramètres sont estimés à leur MAP et substitués dans les calculs ; et une méthode où les hyper-paramètres sont intégrés selon leur loi a posteriori. L'expérience est répétée six fois par plan d'expériences.

Les figures 75 et 76 présentent les résultats obtenus en fonction du coût du plan d'expériences, pour les plans numéro 3 et 4, à raison de une figure par plan d'expériences. Les figures correspondantes pour les plans numérotés 1 et 2 sont en annexe E.2. La première chose à remarquer est que les estimations de probabilité sont similaires pour toutes les répétitions, sur tous les plans d'expériences. Ensuite, la mesure d'incertitude calculée quand les hyper-paramètres sont fixées est toujours décroissante. Les différences

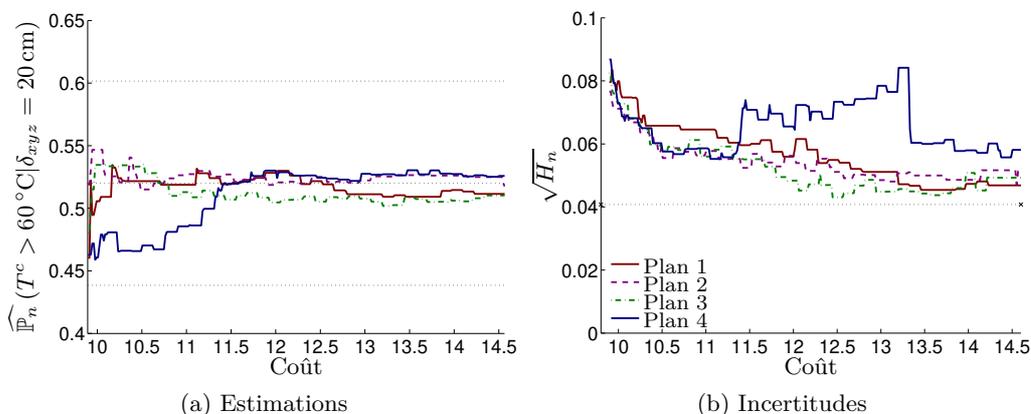


FIGURE 74 – Quatre répétitions d’un plan séquentiel d’expérience destiné à estimer la probabilité que la température T^c dépasse son seuil critique en un scénario de référence. Les courbes colorés sont les évolutions de l’estimation par moyenne et de la mesure d’incertitude en fonction du coût dépensé. Les courbes en pointillé correspondent à la référence par Monte-Carlo sur 150 points. Les coûts sont en nombre de points équivalents à 20 cm.

apparaissent sur les mesures d’incertitude quand les hyper-paramètres varient.

Sur la figure 75, les six courbes de mesures d’incertitude sont homogènes et décroissantes. La planification séquentielle d’expériences se passe bien. A l’inverse sur la figure 76, les courbes sont moins homogènes et tendent à sauter sur différents paliers. La figure 76f dévoile que ces variations dans la mesure d’incertitude totale sont dues à la partie moyenne de la variance (3.14b), la partie variance de la moyenne (3.14c) restant plutôt constante. Cela laisse à penser que les hyper-paramètres sont mal estimés et que le nombre d’échantillons d’hyper-paramètres est trop faible pour obtenir une bonne estimation de la mesure d’incertitude.

Enfin, les figures f de chaque plan d’expériences affichent la décomposition de la mesure d’incertitude. Il est visible que la partie moyenne de la mesure d’incertitude, notée $\mathbb{E}_n[H_n(\theta)]$ dans la section 2.3.4, tend à se réduire, conformément à la définition de la stratégie. Par contre, la partie variance de la moyenne, notée $\mathbb{E}_n[I_n(\theta)]$ dans la section 2.3.4, non seulement ne décroît dans aucune expérience, mais ne semble pas négligeable dans le calcul de la variance totale. Ceci est particulièrement visible sur la figure 75f où la moyenne de la mesure d’incertitude $\mathbb{E}_n[H_n(\theta)]$ devient inférieur à la variance de la moyenne $\mathbb{E}_n[I_n(\theta)]$ sur la fin de la planification.

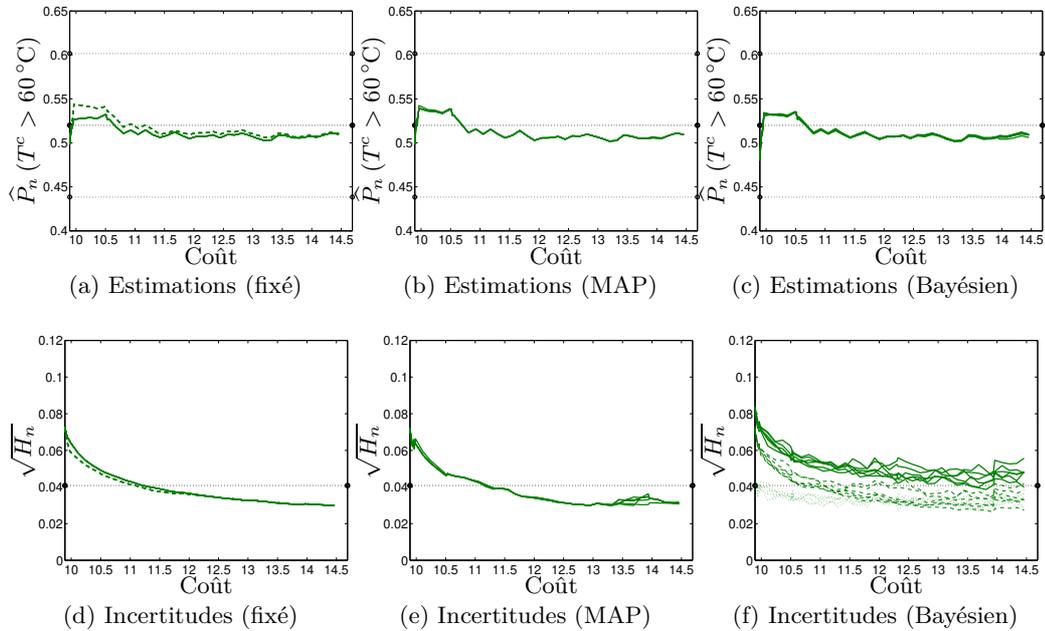


FIGURE 75 – Comparaison des méthodes d’estimation de la probabilité en fonction du coût du plan d’expériences sur le plan numéro 3. La ligne du haut représente les estimations de la probabilité, la ligne du bas les mesures d’incertitude, une borne majorante de l’écart-type. (a, d) Estimations dans le cas où les hyper-paramètres sont fixés. Les lignes pleines correspondent au cas où les hyper-paramètres sont fixés à leur valeurs finales (MAP avec toutes les observations), les lignes discontinues au cas où les hyper-paramètres sont fixés à leur valeurs initiales (MAP du plan initial). (b, e) Estimations dans le cas où les hyper-paramètres sont estimés à leur valeur MAP à chaque nouveau point d’observation et substitués dans le modèle. Il y a six courbes par figure, car la méthode est répétée six fois. (c, f) Estimations par des méthodes complètement bayésiennes. Sur la figure f, les traits pleins sont les mesures d’incertitude totales, les traits discontinus sont la racine carré de la moyenne de la variance, et les traits pointillés sont les écart-types des moyennes. Il y a six jeux de courbes, car la méthode est répétée six fois.

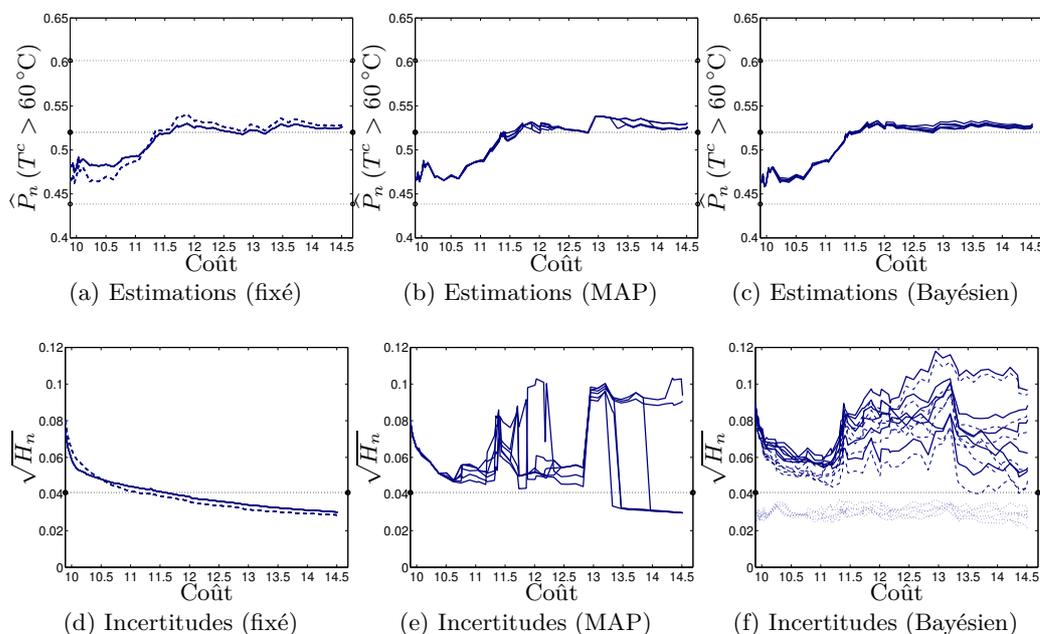


FIGURE 76 – Comparaison des méthodes d’estimation de la probabilité en fonction du coût du plan d’expériences sur le plan numéro 4. La ligne du haut représente les estimations de la probabilité, la ligne du bas les mesures d’incertitude, une borne majorante de l’écart-type. (a, d) Estimations dans le cas où les hyper-paramètres sont fixés. Les lignes pleines correspondent au cas où les hyper-paramètres sont fixés à leur valeurs finales (MAP avec toutes les observations), les lignes discontinues au cas où les hyper-paramètres sont fixés à leur valeurs initiales (MAP du plan initial). (b, e) Estimations dans le cas où les hyper-paramètres sont estimés à leur valeur MAP à chaque nouveau point d’observation et substitués dans le modèle. Il y a six courbes par figure, car la méthode est répétée six fois. (c, f) Estimations par des méthodes complètement bayésiennes. Sur la figure f, les traits pleins sont les mesures d’incertitude totales, les traits discontinus sont la racine carré de la moyenne de la variance, et les traits pointillés sont les écart-types des moyennes. Il y a six jeux de courbes, car la méthode est répétée six fois.

Niveau	50 cm	33 cm	25 cm	20 cm	Total
Plan 1	15	19	5	0	39
	17,3 h	4,79 j	4,07 j		9,58 j
Plan 2	15	26	3	0	44
	17,3 h	6,55 j	2,44 j		9,72 j
Plan 3	9	35	0	0	44
	10,4 h	8,82 j			9,25 j
Plan 4	27	32	0	0	59
	31,2 h	8,06 j			9,36 j

TABLE 11 – Nouvelles observations par niveau de fidélité sur les quatre répétitions d’expériences. Chaque ligne correspond à un plan initial. Pour chaque plan, les chiffres du haut indiquent le nombre d’observation, les chiffres du bas indiquent le temps de calcul total dépensé sur le niveau.

Enfin, nous pouvons regarder quel niveau de fidélité a été observé dans cette expérience. La figure 77 montre les niveaux observés en fonction du coût de calcul. Le tableau 11 vient en complément en affichant le nombre d’observations et les temps de simulation de FDS dépensés sur chaque niveau. Ils révèlent que la planification séquentielle d’expérience a majoritairement dépensé le temps de calcul sur le niveau 33 cm. Plus précisément, il est remarquable sur la figure 77 que les quatre plans d’expériences ont tous commencé par observer plusieurs fois le niveau 50 cm avant de s’attarder sur le niveau 33 cm. Ceci semble cohérent avec le comportement observé dans la section 2.4.3 : la stratégie MRUR choisit des observations en priorité un niveau de fidélité, puis passe au niveau de plus haute fidélité juste au dessus au bout d’un certain moment. Dans ce cas, la figure afficherait le début de ce comportement sur les deux premiers niveaux de fidélité, ainsi que la phase transitoire entre le niveau 33 cm et 25 cm pour les plans d’expériences numéro 1 et 2.

3.4.5 Résultats complémentaires sur la réduction d’incertitude

Nous avons également répété la méthode de planification séquentielle en vue d’améliorer la probabilité de défaillance de la visibilité. Nous sommes partis de quatre plans initiaux d’expériences suivant la méthode de la section 3.3.1. Sur chacun de ces quatre plans, nous avons appliqué la méthode MRUR en vue d’améliorer l’estimation initiale de la probabilité de défaillance de la visibilité sur le niveau 20 cm. Chacune des planifications séquentielle a duré environ 48.4 jours, soit le temps équivalent pour faire 23.89 observations au niveau 20 cm. En ajoutant le coût du plan initial, le coût total du plan

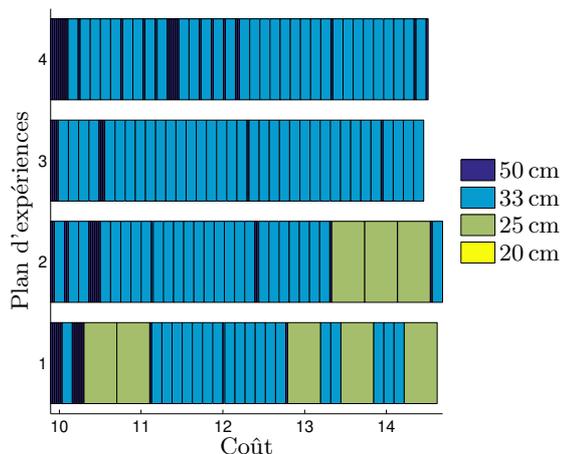


FIGURE 77 – Représentation des niveaux sélectionnés en fonction du coût du plan. Chaque ligne correspond à un plan d'expériences, chaque couleur à un niveau.

final est donc de $10j + 48,4j = 58,4j$.

La figure 78 présente les résultats d'estimation en fonction du coût dépensé. La figure 78a présente l'évolution de l'espérance de la probabilité de défaillance, la figure 78b l'évolution de la racine carrée de la mesure d'incertitude $\sqrt{H_n}$, qui est une borne supérieure de l'écart-type. Les figures montrent que sur toutes les planifications, les estimations par la méthode MRUR tendent vers l'estimation obtenue par Monte-Carlo et que la mesure d'incertitude tend à se réduire progressivement. Ceci montre que la méthode MRUR fonctionne sur cet exemple.

Il est cependant observable sur les figures 74 et 78 que les mesures d'incertitudes semblent tendre vers un minimum sans jamais s'annuler. Pour voir la raison de cette convergence vers un minimum, nous avons décomposé la mesure d'incertitude de l'une des expériences de planification séquentielle sur la visibilité, selon la formule équation (3.14) : $H_n = \mathbb{E}_n[H_n(\theta)] + \mathbb{E}_n[I_n(\theta)]$. L'expérience reprise correspond aux résultats représentés en vert pointillé sur la figure 78. La figure 79 représente cette décomposition sous différents angles. La figure 79a montre les deux parties en échelle absolues. Il apparaît que l'incertitude totale H_n se réduit beaucoup durant l'expérience, et que au début, la principale source d'incertitude est la partie $\mathbb{E}_n[H_n(\theta)]$. Afin de mieux voir la répartition des deux termes au cours de la planification, la seconde figure 79b montre la répartition des deux sources. Il est alors visible que le terme $\mathbb{E}_n[I_n(\theta)]$ est négligeable au début de la planification, mais qu'à la fin de la planification, il correspond à 25 % de l'incerti-

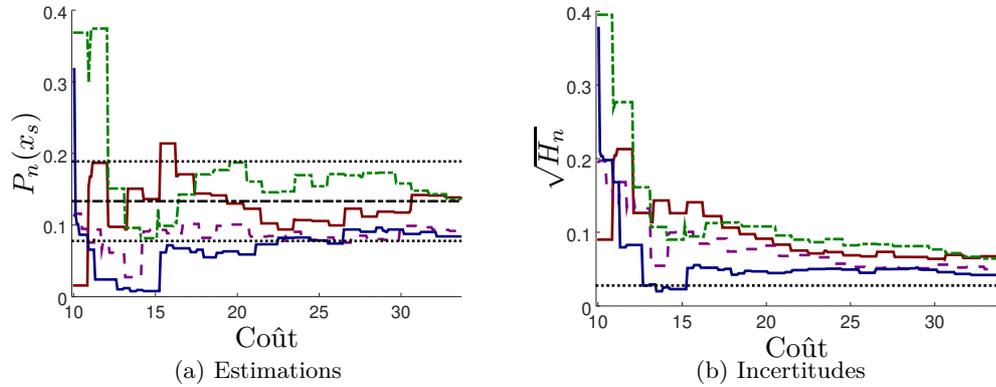


FIGURE 78 – Quatre répétitions d’un plan séquentiel d’expérience destiné à estimer la probabilité que la température V dépasse son seuil critique en un scénario de référence. Les courbes colorés sont les évolutions de l’estimation par moyenne et de la mesure d’incertitude en fonction du coût dépensé. Les courbes en pointillé correspondent à la référence par Monte-Carlo sur 150 points. Les coûts sont en nombre d’observations équivalentes à 20 cm.

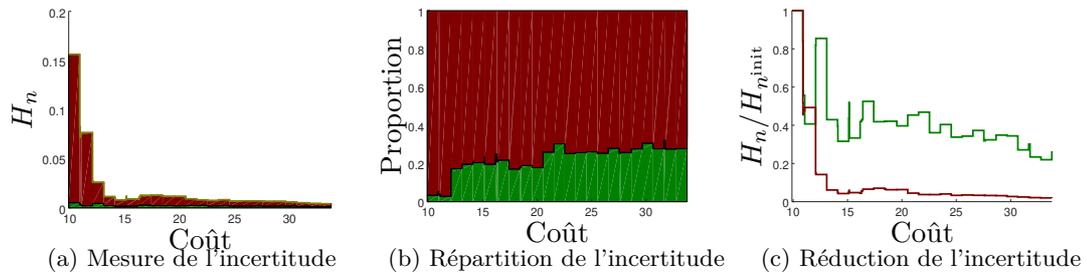


FIGURE 79 – Réduction de l’incertitude en fonction du coût dépensé dans le plan d’expériences séquentiel. Le rouge correspond à la partie $\mathbb{E}_n[H_n(\theta)]$, le vert à la partie $\mathbb{E}_n[I_n(\theta)]$. (a) Incertitude en échelle absolue. (b) Incertitude en échelle relative (divisée par l’incertitude totale H_n). (c) Réduction d’incertitude : la courbe rouge représente $\mathbb{E}_n[H_n(\theta)]/\mathbb{E}_{n,init}[H_{n,init}(\theta)]$, la courbe verte $\mathbb{E}_n[I_n(\theta)]/\mathbb{E}_{n,init}[I_{n,init}(\theta)]$

tude. Ceci montre que la partie $\mathbb{E}_n[I_n(\theta)]$ devient de plus en plus importante au fur et à mesure de la planification séquentielle. Ceci est d'ailleurs confirmé par la figure 79c qui montre les deux sources d'incertitudes normalisées par leur valeurs initiales. Il est visible que les deux termes $\mathbb{E}_n[H_n(\theta)]$ et $\mathbb{E}_n[I_n(\theta)]$ se réduisent, mais le premier beaucoup plus vite que le second. Ceci est parfaitement explicable, puisque l'algorithme cherche spécifiquement à réduire le terme $\mathbb{E}_n[H_n(\theta)]$. De cette façon, nous espérons ainsi que le terme $\mathbb{E}_n[I_n(\theta)]$ se réduise de lui-même quand de nouvelles observations s'ajoutent. Cette réduction a d'ailleurs lieu, mais pas aussi rapidement que souhaitée. Ceci explique pourquoi les courbes de mesures d'incertitude semblent converger vers une valeur non-nulle de l'incertitude. Cela suggère qu'il faudrait développer des algorithmes pour chercher à réduire spécifiquement le terme $\mathbb{E}_n[I_n(\theta)]$.

3.5 Conclusion

Dans ce chapitre, nous avons illustré la méthodologie développée sur un exemple applicatif en sécurité incendie. Cet exemple consiste en l'évaluation la conformité d'un système de désenfumage d'un bâtiment du LNE sur un simulateur d'incendies intitulé FDS. La sécurité se mesure par la probabilité que certaines sorties du simulateur dépassent des seuils réglementaires. Ces probabilités sont vues comme des fonctions qui dépendent du scénario incendie et de la distribution aléatoire des conditions environnementales.

Nous avons d'abord posé le problème incendie et présenté succinctement des propriétés remarquables de FDS. Ensuite, nous avons présenté et illustré la méthodologie d'estimation de probabilité de défaillance sur simulateur multi-fidèle. En premier lieu, nous avons montré un exemple réaliste d'étude utilisant la méthodologie pour obtenir différents résultats, notamment des évaluations de probabilité de défaillance. Puis, nous avons validé la méthode en la répétant sur plusieurs plans d'expériences. Les estimations obtenues avec cette méthode étaient globalement justes pour le flux radiatif et la température, mais assez moyennes pour la visibilité. Nous avons également illustré la stratégie MRUR sur l'estimation d'une probabilité de défaillance sur une sortie en un scénario de référence. Les deux intérêts principaux de la méthode est d'abord qu'elle permet de limiter le coût d'observation nécessaire de FDS, et ensuite, qu'elle donne une mesure d'incertitude sur les estimations.

Plusieurs perspectives peuvent se dégager à partir de ces travaux. D'abord, comme la visibilité est la variable la moins bien estimée à partir du plan initial d'expérience, nous aimerions appliquer la stratégie MRUR sur la visibilité pour voir si l'estimation

de probabilité de défaillance s'améliore. Concernant les hyper-paramètres, nous avons constaté que leur loi a posteriori était probablement multi-modale, ce qui complique leur estimation et leur intégration. Une façon de revenir dans un cas uni-modal serait d'avoir des lois a priori plus informatives. La question de l'impact de l'incertitude des hyper-paramètres sur l'incertitude des quantités d'intérêt reste aussi une question ouverte. En particulier, il semblerait que l'incertitude supplémentaire due aux hyper-paramètres dans les mesures d'incertitudes ne soit pas négligeable. Ceci implique qu'il faudrait adapter les stratégies de planification séquentielle pour tenir compte de ce terme. Dans la question de la modélisation, il n'y a pas de méthode bien définie pour choisir la fonction de covariance modélisant l'évolution de la fidélité. Ainsi, il faudrait comparer les résultats obtenus sur les sorties de FDS avec les deux types de covariance proposés pour le modèle multi-fidèle non-stationnaires afin de déterminer si l'un des deux est plus efficient. Enfin, les modèles multi-fidèles ne disposent pas de méthodes de validation croisée adaptée, qui permettraient de vérifier la qualité des modèles sans données de validation. Des travaux futurs pourraient se concentrer sur la construction de telles méthodes.

Conclusion

Sommaire

1	Résumé et contributions	175
2	Perspectives	177

1 Résumé et contributions

Cette thèse se focalise sur l'évaluation de quantités d'intérêt issues de simulateurs multi-fidèles, et en particulier, de l'évaluation de probabilité de défaillance de systèmes décrits par un modèle numérique. Dans un cadre réglementaire ou de prise de décision, ces estimations doivent s'accompagner de mesures d'incertitude. Pour mener ces évaluations, la méthode proposée s'appuie sur une approche multi-fidèle bayésienne. Le modèle numérique est décrit comme une application inconnue dont les réponses suivent une loi de probabilité. La distribution en sortie possède des paramètres qui sont des fonctions inconnues dépendantes des entrées. Afin d'estimer ces fonctions, un modèle a priori fondé sur des processus gaussiens est associée à chacune de ces fonctions. Les lois a priori de ces processus transcrivent en particulier l'aspect multi-fidèle du modèle numérique étudié. En utilisant la formule de Bayes, il est possible d'obtenir les lois a posteriori des processus à ces fonctions, et en déduire une évaluation des quantités d'intérêt, en utilisant par exemple une méthode de Monte-Carlo.

Cette thèse apporte des contributions sur deux problèmes. D'abord, elle étudie les différentes modélisations par processus gaussiens existant dans la littérature de modèles numériques multi-fidèles. Elle propose une extension de ces modèles pour modéliser la variance d'un effet stochastique en sortie du modèle numérique. Ensuite, elle propose une méthodologie pour créer un algorithme de planification séquentielle destinée à proposer de nouvelles observations du simulateur multi-fidèle, en vue de réduire l'incertitude sur l'estimation de la quantité d'intérêt.

Le chapitre 1 porte sur la modélisation par processus gaussien de la multi-fidélité. Deux modèles sont étudiés : le modèle auto-régressif et le modèle non-stationnaire. Ces deux modèles sont fondés sur un principe concernant la relation entre les niveaux de fidélité. Pour chacun de ces modèles, la relation entre les niveaux de fidélité se traduit dans le modèle de processus gaussien a priori, à travers les fonctions de moyenne et de covariance. Dans cette thèse, les différentes propriétés des deux modèles sont étudiées (aspect « correctif » pour le modèle auto-régressif, aspect « déconvolutif » pour le modèle non-stationnaire) puis illustrées sur des exemples simples. Cependant, ces modèles sont conçus pour des simulateurs déterministes. Ainsi, la fin du chapitre 1 se concentre sur l'extension de modèle multi-fidèle où la sortie est stochastique suivant une loi gaussienne. La thèse propose ainsi un modèle log-gaussien pour modéliser la variance de l'effet stochastique du simulateur. Plus exactement, ce modèle décrit la variance comme une fonction ne dépendant que de la fidélité. Cette fonction est égale à une valeur moyenne inconnue, plus des variations faibles et indépendantes à chaque niveau. Ce modèle permet de mener une analyse complètement bayésienne en incluant la variance dans les hyper-paramètres.

Le chapitre 2 présente des méthodes de planification séquentielle d'expériences dans le cadre multi-fidèle. Le chapitre débute par une présentation des méthodes de génération de plan Hypercube Latin Emboîté (NLHS). Nous proposons une amélioration simple pour engendrer des plans NLHS maximisant le remplissage d'espace. Ces plans d'expériences sont utilisés comme plans d'expériences initiaux dans une démarche séquentielle. Ensuite, nous rappelons les principes de la Réduction Progressive de l'Incertitude (SUR), qui est une méthodologie générale pour construire des algorithmes de planification séquentielle dans un cadre où les coûts d'observation sont tous égaux. Nous adaptons plusieurs critères SUR de la littérature dédiés à l'estimation de probabilité de dépassement de seuil au cas stochastique dont la distribution des sorties est gaussienne. Enfin, nous présentons la contribution principale du chapitre, la stratégie du Taux Maximal de Réduction de l'Incertitude (MRUR). Celle-ci propose de choisir l'observation maximisant le rapport entre le coût d'observation et la réduction espérée d'incertitude. Le coût d'observation est une fonction supposée connue. La réduction d'incertitude attendue est mesurée par la réduction du critère de SUR entre l'état actuel et l'état futur. La stratégie MRUR est testée sur des exemples académiques multi-fidèles, où elle affiche des résultats très satisfaisants. En particulier, les expériences montrent que la stratégie MRUR choisit prioritairement des niveaux de fidélité qui réalisent un bon compromis entre coût et réduction d'incertitude. Nous avons aussi illustré sur un exemple académique une stratégie MRUR étendue à un cadre de planification parallélisée synchrone.

Enfin, la méthodologie est testée dans le chapitre 3 sur un exemple issu de la sécurité incendie. L'exemple présenté est l'évaluation de la conformité d'un système de désenfumage d'un bâtiment existant par rapport aux normes sur la sécurité incendie. En particulier, l'exemple se focalise sur la comparaison entre des variables physiques (flux radiatif, température, visibilité) simulées sur Fire Dynamics Simulator (FDS) à leurs seuils de dangerosité définis dans la norme ISO 13571 [2012]. Le chapitre présente la méthodologie, utilisée dans cet exemple pour évaluer des indicateurs de conformité, à savoir des courbes de probabilités de défaillance, évaluées comme des probabilités de dépassement des seuils normatifs en fonction de variables de scénario. Ces courbes de probabilités de défaillance et les intervalles de crédibilité associées prennent en compte trois sources d'incertitudes :

- l'incertitude des entrées, liée à la variabilité de certaines entrées qualifiées d'environnementales ;
- l'incertitude des sorties, décrivant le fait que, même si le modèle numérique était parfaitement connu, la sortie pour un vecteur d'entrées donné resterait imprévisible (par exemple, dans le cas d'un simulateur stochastique) ;
- l'incertitude due à la méconnaissance du modèle numérique, qui représente le manque de connaissance disponible sur le simulateur, notamment du fait que le nombre de simulations possibles reste limité par les temps de calcul du simulateur.

Cette dernière source est prise en compte grâce au modèle bayésien.

Ce type de résultats est prévu pour venir en complément des analyses traditionnelles effectuées par les ingénieurs en sécurité incendie.

2 Perspectives

Cette thèse propose une méthodologie pour estimer des quantités d'intérêt sur un simulateur multi-fidèle. Cependant, de nombreux points pourraient être développés plus en détails.

Dans cette thèse, la méthode utilisée pour estimer les probabilités de dépassement de seuil en fonction de variables de scénario est une méthode d'estimation ponctuelle par Monte-Carlo. Cependant, cette méthode est utilisée pour sa simplicité. D'autres méthodes d'estimation de probabilité pourraient être utilisées sans que cela ne change rien à la méthode générale. Par exemple, les méthodes d'échantillonnage préférentiel [Au et al., 2007] pourraient être utilisées pour estimer des probabilités d'événements rares. De plus, l'estimation d'une courbe de probabilité passe par une estimation point par point de la valeur de la courbe, ce qui amène à des estimations coûteuses en temps de

calcul, comparativement aux autres parties de l'analyse. Un algorithme d'estimation de courbe complète, exploitant les régularités des courbes, pourrait être conçu. La principale difficulté à l'estimation directe de courbes de probabilités de défaillance est de pouvoir estimer les lois conjointes des probabilités de défaillance en différents scénarios. La méthode d'estimation des probabilités intégrées de dépassement de seuil repose sur la génération de trajectoires conditionnelles sur un grand nombre de points, et le calcul de la moyenne de probabilité de défaillance en chacun de ces points. Or, cette méthode restreint le nombre maximum de points sur lequel faire des trajectoires conditionnelles. L'utilisation de méthodes de génération de trajectoires conditionnelles sur un très grand nombre de points, comme la méthode de Azzimonti et al. [2016a], permettrait de résoudre ce problème. Dans le cadre de la multi-fidélité, un problème supplémentaire peut apparaître lors à l'utilisation de modèle de processus gaussiens dédiés à la multi-fidélité. Comme les niveaux de plus basse-fidélité sont moins coûteux, le nombre d'observations risque de devenir rapidement assez élevé car beaucoup d'observations sont faites au niveau de plus basse-fidélité. Or, une augmentation du nombre de points ralentit considérablement les méthodes d'analyses, que ce soit les prédictions des sorties ou les estimations des probabilités de dépassement de seuil. Des développements pourraient être envisagés pour estimer les probabilités quand le nombre d'observations basse-fidélité augmente.

Cette thèse repose sur l'utilisation de processus gaussien multi-fidèle pour combiner des résultats à différents niveaux. Par conséquent, toute amélioration dans la construction de ces modèles améliorera l'estimation des quantités d'intérêt finales. Ainsi, des travaux pourraient se concentrer sur l'estimation des hyper-paramètres dans un modèle multi-fidèle en général. Actuellement, le seul cas où l'estimation est simple est celui du modèle auto-régressif observé sur un plan emboîté. Dans cette thèse, nous utilisons une approche complètement bayésienne avec intégration des hyper-paramètres selon leur loi a posteriori. Nous choisissons une graine par maximisation de la loi a posteriori puis nous échantillonnons la loi par une méthode de Monte-Carlo par Chaînes de Markov (MCMC). Cependant, l'estimation du Maximum A Posteriori (MAP) reste assez heuristique. Des développements pourraient être envisagés pour améliorer l'estimation du MAP, par exemple en utilisant la méthode de prédiction de Qian et Wu [2008], dans un cas où le plan d'expériences n'est pas emboîté. D'autres recherches peuvent également être menées sur l'impact des différents niveaux sur les prédictions en haute-fidélité, la question étant de savoir si la prédiction haute-fidélité en un point dépend plus des autres observations de même niveau ou des observations à différents niveaux faites au même point. L'un des manques dans la méthodologie actuelle est l'absence de validation a posteriori du modèle et notamment de la relation entre les niveaux de fidélité. Une

validation croisée sur les points pris individuellement nous paraît être une mauvaise idée. En effet, comme il y a plus de d'observations sur les niveaux de basse-fidélité que sur les niveaux de haute-fidélité, une validation croisée favoriserait les modèles apprenant bien le niveau-basse fidélité au détriment des niveaux de plus hautes-fidélité. Pour appliquer ce genre de méthode, il faudrait a minima pondérer les erreurs de validation croisée en fonction du niveau de fidélité (il est plus grave que le modèle se trompe en haute-fidélité qu'en basse-fidélité). Une autre solution pourrait se trouver dans la décomposition que font les modèles multi-fidèles du simulateur entre : une partie commune à tous les niveaux (le niveau de plus basse fidélité pour le modèle correctif, le niveau de plus haute fidélité pour le modèle non-stationnaire) ; et une partie variable avec la fidélité. Un algorithme de validation pourrait recomposer chacune des deux parties du modèle notamment en procédant par différence entre les prédictions au niveau de référence et les sorties aux différents niveaux, et valider chaque morceau du modèle indépendamment. Ceci permettrait notamment de pouvoir choisir la covariance appropriée pour modéliser la relation entre les niveaux de fidélité. A terme cela permettrait de faire de la sélection automatique de modèle. Enfin, sur la question de la modélisation, l'aspect stochastique reste une question ouverte. Dans cette thèse, un modèle simplifié a été proposé pour modéliser la variance de l'effet stochastique. Cependant, les hypothèses de ce modèle sont assez fortes : les variances ne dépendent que du niveau de fidélité et la valeur d'une sortie en un point suit une loi gaussienne. Des développements futurs pourraient se focaliser sur une intégration complètement bayésienne d'une fonction de variance hétéroscédastique selon toutes les variables d'entrée.

La thèse se concentre sur la planification d'expériences dans un cadre multi-fidèle. Dans le cas des plans d'expériences non-séquentiel, même si quelques développements ont été effectués, de nombreuses questions sur les plans d'expériences initiaux sont encore sans réponse, comme le nombre de niveaux d'observations initiales ou le nombre d'observations par niveaux. Dans cette thèse, la stratégie de planification séquentielle proposée repose sur la comparaison entre la réduction d'une mesure d'incertitude et du coût calculatoire. Or, dans cette thèse, les hyper-paramètres n'ont jamais été intégrés dans le calcul des critère SUR. Même si nous proposons une approximation pour prendre en compte la loi des hyper-paramètres, celle-ci n'a pas été utilisée pour cause de temps de calcul. Sur ce sujet, les travaux de Benassi et al. [2011] présente un algorithme résolvant ce problème dans le cadre du critère de l'Amélioration Espérée (EI). De plus, l'approximation que nous proposons se contente d'intégrer le critère d'échantillonnage sur l'espace des hyper-paramètres. En particulier, elle n'estime pas l'incertitude due directement aux hyper-paramètres, ce qui fait que la réduction totale d'incertitude est

supérieure à l'estimation. Ceci pourrait poser problème dans le cas où l'incertitude sur le critère d'échantillonnage due aux hyper-paramètres n'est pas négligeable. Des stratégies évaluant la réduction d'incertitude due aux hyper-paramètres dans le calcul du critère d'échantillonnage pourraient être envisagées.

Enfin, le principal axe des développements futurs concerne la stratégie MRUR, et de manière plus générale, la planification séquentielle avec des coûts d'observation variables. La thèse montre sur des exemples que cette stratégie est très efficace dans le contexte de la multi-fidélité. Toutefois, le comportement de la stratégie MRUR n'est pas toujours bien compris. Nous constatons notamment une propension de la stratégie MRUR à observer les niveaux de très basse fidélité. A ce jour, nous ne savons pas si ce comportement est souhaitable ou non. Des recherches plus approfondies sur le comportement des stratégies MRUR, en lien avec la nature du modèle utilisé, pourraient être menées. Si la stratégie MRUR est efficace en multi-fidélité, il est incertain que cela reste le cas dans un cadre plus général où le coût d'observation dépendrait de toutes les variables d'entrée, car l'aspect multi-fidèle induit que le gain d'information augmente systématiquement avec le coût d'observation. Il serait judicieux de tester la stratégie MRUR hors du cadre multi-fidèle, dans le cas où le coût des observations varie en fonction des entrées, sans qu'il y ait de relations a priori entre le coût d'observation et le gain d'information. Une autre hypothèse de la stratégie actuellement proposée est que le coût d'observation est déterministe et connu. Des développements pourraient être imaginés dans le cas où le coût est stochastique et/ou doit être modélisé par un processus aléatoire. Par exemple, Snoek et al. [2012] modélisent la fonction de coût par un processus log-gaussien, qu'ils utilisent pour calculer l'espérance de l'inverse du coût. Enfin, si la stratégie MRUR peut être utilisée en planification d'expériences parallèle synchrone, il n'existe pas d'extensions simples de la stratégie MRUR en planification asynchrone. Des développements dans cette configuration pourraient être envisagés.

Annexe A

Régression par processus gaussien

A.1 Formalisation mathématiques

Dans cette partie, nous présentons le modèle bayésien classique de la régression par processus gaussien. Plus de détails sur la modélisation par processus gaussien, alias le krigeage, peuvent être trouvés dans Stein [1999], Santner et al. [2003], Rasmussen et Williams [2006] ou Forrester et al. [2008].

Soit une application prenant des entrées $x \in \mathbb{X} \subset \mathbb{R}^d$, où d est la dimension de x , et renvoyant en sortie une variable aléatoire réelle Z . Si $X = (x_i)_{1 \leq i \leq n}$ sont n points de l'espace des entrées, les sorties $Z = (Z_i)_{1 \leq i \leq n}$ correspondants sont supposées suivre des lois normales indépendantes

$$Z|X; \xi, \lambda \sim \mathbf{N}(\xi(X), \text{diag}(\lambda(X))), \quad (\text{A.1})$$

où ξ est la fonction moyenne, qui représente sa partie déterministe ; λ est la fonction variance, qui règle l'amplitude de la partie stochastique ; $\xi(X)$ et $\lambda(X)$ sont respectivement les vecteurs de ces fonctions aux points X , $(\xi(x_i))_{1 \leq i \leq n}$ et $(\lambda(x_i))_{1 \leq i \leq n}$; et diag est l'opérateur qui à un vecteur renvoie la matrice diagonale des éléments de ces vecteurs. Le fait que la matrice de variance du vecteur Z soit diagonale représente l'indépendance des observations.

Ensuite, une loi a priori est ajoutée sur la fonction ξ . Ainsi, la moyenne ξ est supposée être un processus gaussien de fonction moyenne m et de fonction de covariance k . La fonction moyenne s'écrit comme une combinaison linéaire de fonctions connues, $m(x) = \sum_{l=1}^{p_\beta} h_l(x)\beta_l = h(x)\beta$. Dans ces conditions,

$$\xi(X)|X; \lambda; h, \beta, k \sim \mathbf{N}(H(X)\beta, K(X, X)) \quad (\text{A.2})$$

où β est le vecteur $(\beta_l)_{1 \leq l \leq p_\beta}$ de longueur p_β , $h = (x \mapsto h_l(x))_{1 \leq l \leq p_\beta}$ est une famille de p_β fonction(s), $H(X)$ est la matrice $(h_l(x_i))_{1 \leq i \leq n, 1 \leq l \leq p_\beta}$ de taille $n \times p_\beta$, et $K(X, X)$ est la matrice de covariance valant $(k(x_i, x_j))_{1 \leq i, j \leq n}$.

Généralement, les fonctions de variance λ et de covariance k ne sont pas fixées a priori. L'idée est plutôt de choisir deux familles de fonctions (une pour la variance, l'autre pour la covariance), indexées par un vecteur θ appelé vecteur des hyper-paramètres. L'utilisateur choisit alors ses familles de fonctions λ_θ et k_θ et des algorithmes estiment ou intègrent les hyper-paramètres par rapport aux observations. Le modèle se ré-écrit alors $\{X; \lambda_\theta; h, \beta, k_\theta, \theta\}$, où λ_θ et k_θ sont deux familles de fonctions, et θ est le vecteur d'hyper-paramètres qui a été choisi.

Enfin, on complète ce modèle hiérarchique avec une loi normale a priori sur les coefficients de la partie linéaire du modèle :

$$\beta | X; \lambda_\theta; h, k_\theta, \theta; \beta_0, Q \sim \mathbf{N}(\beta_0, Q), \quad (\text{A.3})$$

où β_0 est un vecteur de longueur p_β qui représente la moyenne a priori de β et Q est une matrice symétrique positive de taille p_β qui est la variance à priori du vecteur β . Le modèle $\mathcal{M} = \{X; \lambda_\theta; h, k_\theta, \theta; \beta_0, Q\}$ est l'ensemble des quantités à fournir pour pouvoir utiliser ce modèle bayésien.

Théorème A.1.1 (Krigage). *Soit $X = (x_i)_{1 \leq i \leq n} \in \mathbb{X}^n$ des points d'observations, $Y = (y_i)_{1 \leq i \leq N} \in \mathbb{X}^N$, N point(s). L'ensemble des observations est noté $\chi_n = (x_i, Z_i)$. Alors, les lois a posteriori sont des lois normales*

$$\beta | \mathcal{M}, \chi_n \sim \mathbf{N}(\beta_{n,\theta}, Q_{n,\theta}) \quad (\text{A.4})$$

$$\xi(Y) | \mathcal{M}, \chi_n \sim \mathbf{N}(m_{n,\theta}(Y), K_{n,\theta}(Y, Y)) \quad (\text{A.5})$$

$$Z_Y | \mathcal{M}, \chi_n \sim \mathbf{N}(m_{n,\theta}(Y), K_{n,\theta}(Y, Y) + \text{diag}\{\lambda_\theta(Y)\}) \quad (\text{A.6})$$

où

$$V_\theta(X) = K_\theta(X, X) + \text{diag}(\lambda_\theta(X)) \quad (\text{A.7})$$

$$Q_{n,\theta} = [Q^{-1} + H(X)^T V_\theta(X)^{-1} H(X)]^{-1} \quad (\text{A.8})$$

$$\beta_{n,\theta} = Q_{n,\theta} (Q^{-1} \beta_0 + H(X)^T V_\theta(X)^{-1} Z_X) \quad (\text{A.9})$$

$$\mu_{n,\theta}(Y) = V_\theta(X)^{-1} K_\theta(X, Y) \quad (\text{A.10})$$

$$m_{n,\theta}(Y) = H(Y)\beta_{n,\theta} + \mu_{n,\theta}(Y)^T (Z_X - H(X)\beta_{n,\theta}) \quad (\text{A.11})$$

$$K_{n,\theta}(Y, Y) = K_\theta(Y, Y) - \mu_{n,\theta}(Y)^T K_\theta(X, Y) \quad (\text{A.12})$$

$$+ \left(H(Y) - \mu_{n,\theta}(Y)^T H(X) \right) Q_{n,\theta} \left(H(Y)^T - H(X)^T \mu_{n,\theta}(Y) \right).$$

Comme ces équations sont vraies pour n'importe quels ensembles de points de prédiction Y , on parle alors des processus à posteriori $\xi|\chi_n$ et $Z|\chi_n$, le modèle étant sous-entendu.

Par défaut, dans cette thèse nous nous placerons toujours en krigeage universel et presque toujours en krigeage ordinaire, suivant la recommandation générale de la littérature [Zimmerman et al., 1999]. Le krigeage universel correspond au cas où $Q^{-1} = 0$. La loi de β s'écrit alors de manière impropre, $\beta \sim \mathcal{U}(\mathbb{R}^{p_\beta})$ et la valeur de β_0 n'intervient plus dans les lois a posteriori. Le krigeage ordinaire suppose en plus que $p_\beta = 1$ et $h_1(x) = 1$ la moyenne est une fonction constante inconnue indépendante de x .

Trois cas peuvent arriver pour la variance de l'effet stochastique. Si $\lambda_\theta = 0$, la fonction de variance est nulle, alors Z_X et $\xi(X)$ sont confondus, le simulateur est dit déterministe. Si la variance du bruit λ_θ est une constante non nulle indépendante de x , l'effet stochastique est dit homoscédastique. Sinon, si la variance du bruit λ_θ est non nulle et dépend des entrées x , il est dit hétéroscédastique.

Le théorème A.1.1 est illustré sur un exemple simple. Soit un simulateur inconnu défini sur $\mathbb{X} = [0; 1] \subset \mathbb{R}^1$. Supposons que $n = 8$ observations (z_i) aux points d'entrées (x_i) soient disponibles. Nous cherchons à prédire la valeur de cette fonction aux points de prédiction $y_j = j/(N - 1)$, où $N = 100$ est le nombre points de prédictions. Dans cet exemple, l'effet stochastique est supposé homoscédastique, $\lambda_\theta(x) = \lambda$. La fonction de covariance choisie est une fonction de covariance de Matérn de régularité $5/2$, $k_\theta(x, x') = \sigma^2 \left(1 + \sqrt{10} \frac{|x-x'|}{\rho} + \frac{10}{3} \left(\frac{|x-x'|}{\rho} \right)^2 \right) e^{-\sqrt{10} \frac{|x-x'|}{\rho}}$ (voir annexe A.2). Le vecteur $\theta = (\lambda, \sigma^2, \rho)$ est le vecteur des hyper-paramètres du processus gaussien. Dans cet exemple, les hyper-paramètres seront fixés à $\lambda = 0$ (pas d'aléa en sortie, le simulateur est déterministe), $\sigma^2 = 20^2$ et $\rho = 0.5$. Pour le terme de moyenne $m(x) = h(x)\beta$, nous nous plaçons en krigeage ordinaire : une seule fonction de régression, $p_\beta = 1$, qui est une fonction constante, $h(x) = h_1(x) = 1$, $\beta = \beta_1$. La variance a priori de β est supposée infinie, $Q^{-1} = 0$. Comme Q^{-1} est nul, la valeur de la moyenne a priori des coefficients linéaires de la moyenne β_0 n'a plus d'importance. La valeur choisie par défaut est $\beta_0 = 0$.

Une fois le modèle établi, le théorème A.1.1 peut être appliqué. Le résultat est la loi a posteriori du processus aux points de prédiction voulus, conditionnellement au modèle a priori et aux observations. Cette loi est une loi gaussienne (A.5), dont la fonc-

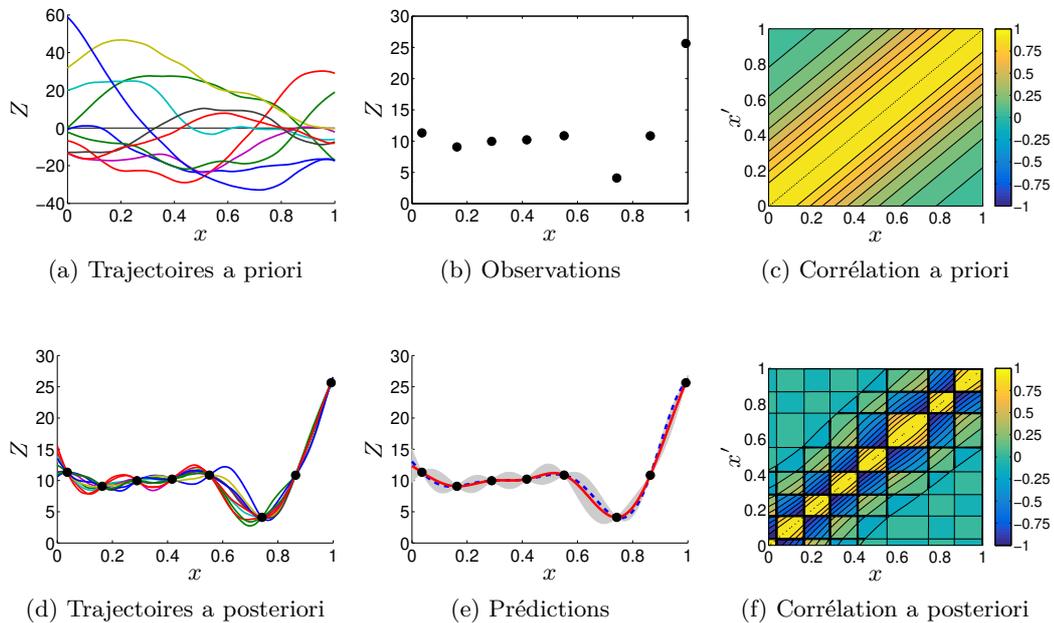


FIGURE 80 – Les étapes d’une régression par processus gaussien. (a) Des trajectoires du modèle a priori. (b) Des observations. (c) La fonction de corrélation a priori. (d) Des trajectoire a posteriori. (e) Représentation classique de la prédiction par processus gaussien. La courbe continue est la fonction de moyenne a posteriori. La zone autour correspond à l’intervalle de crédibilité à 95 %. La courbe discontinue est la vraie fonction dont sont issues les observations. (f) La fonction de corrélation a posteriori.

tion moyenne (A.11) donne un estimateur de la fonction vraie, et la fonction de covariance (A.12) une mesure de son incertitude.

La figure 80 présente les différentes étapes de la régression par processus gaussien. Les figures 80a et 80b présentent les informations à fournir avant la régression. La figure 80a illustre le modèle a priori par des trajectoires issues du processus a priori. La figure 80b présente les observations. La figure 80d illustre le modèle a posteriori par des trajectoires conditionnelles aux points d’observations. Le simulateur étant déterministe, toutes les trajectoires du processus a posteriori passe par les points d’observations. La figure 80e est un résumé de la figure 80d, sur laquelle est représentée la fonction moyenne a posteriori et la zone de crédibilité à 95 %. Cette courbe n’est qu’un résumé, notamment parce qu’elle n’affiche pas les corrélations entre les valeurs des sorties en plusieurs points. Enfin, les figures 80c et 80f tracent les fonctions de corrélation a priori et a posteriori. La même échelle de couleurs est utilisée sur les deux figures. Sur la corrélation a posteriori

(figure 80f), les traits horizontaux et verticaux correspondent aux observations. Sur la figure 80c, la valeur de la corrélation des points proches de la diagonale, tend vers un, signifiant que une valeur en un point est connue, alors la valeur en un point proche est connue elle-aussi. Sur la figure 80f les valeurs des corrélations de part et d'autre d'un point d'observation tendent vers -1. Cela signifie que de part et d'autre d'un point d'observation, si la trajectoire avant le point d'observation est au-dessus de la moyenne, alors la même trajectoire après le point d'observation est en-dessous de la moyenne (et réciproquement).

A.2 Fonction de covariance de Matérn

Les fonction de covariance anisotrope de Matérn sont une famille de fonctions de covariances stationnaires

$$\forall \nu \in]0; +\infty], \quad k(x, x') = \sigma^2 \mathcal{M}_\nu \left(\sqrt{\sum_{k=1}^d \left(\frac{x_k - x'_k}{\rho_k} \right)^2} \right) \quad (\text{A.13})$$

où $(\sigma^2, \nu, (\rho_k)_{1 \leq k \leq d})$ sont les $d + 2$ hyper-paramètres de la fonction de covariance, respectivement de variance, de régularité et de portées, et $h \mapsto \mathcal{M}_\nu(h)$ est la fonction de Matérn de régularité ν , défini par sa transformée de Fourier

$$\text{TF}[\mathcal{M}_\nu](\omega) = \int_{-\infty}^{+\infty} \mathcal{M}_\nu(h) e^{-i\omega h} dh = \frac{\sqrt{\pi} \Gamma(\nu + 1/2)}{\sqrt{\nu} \Gamma(\nu)} \frac{1}{\left(\left(\frac{\omega}{2\sqrt{\nu}} \right)^2 + 1 \right)^{\nu+1/2}} \quad (\text{A.14})$$

Les fonctions de covariance de Matérn possèdent une expression générale dans l'espace des entrées :

$$\mathcal{M}_\nu(h) = \frac{2}{\Gamma(\nu)} \cdot (\sqrt{\nu} |h|)^\nu \cdot \mathcal{K}_\nu(2\sqrt{\nu} |h|), \quad (\text{A.15})$$

où \mathcal{K}_ν est la fonction modifiée de Bessel du second type de régularité ν , et h est la distance pondérée. Il existe des formules plus simples dans le cas où $\nu = p + \frac{1}{2}$, avec p

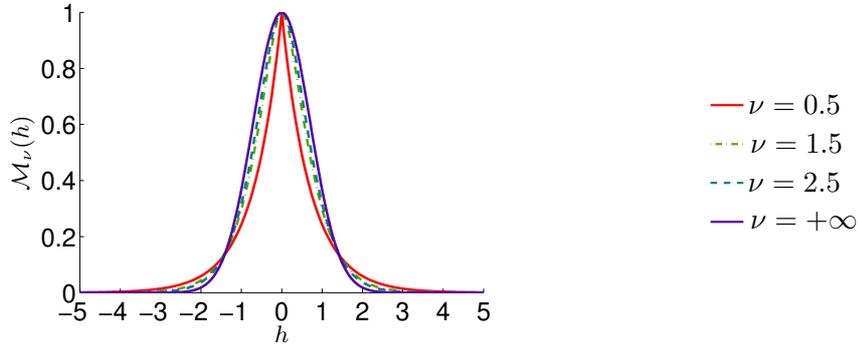


FIGURE 81 – Quelques fonctions de covariance de Matérn.

entier. Ainsi, pour $p = 0, 1, 2$ et 3 ,

$$\mathcal{M}_{\frac{1}{2}}(h) = e^{-\sqrt{2}|h|} \quad (\text{A.16a})$$

$$\mathcal{M}_{\frac{3}{2}}(h) = \left(1 + \sqrt{6}|h|\right) \cdot e^{-\sqrt{6}|h|} \quad (\text{A.16b})$$

$$\mathcal{M}_{\frac{5}{2}}(h) = \left(1 + \sqrt{10}|h| + \frac{10}{3}|h|^2\right) \cdot e^{-\sqrt{10}|h|} \quad (\text{A.16c})$$

$$\mathcal{M}_{\frac{7}{2}}(h) = \left(1 + \sqrt{14}|h| + \frac{28}{5}|h|^2 + \frac{14}{15}\sqrt{14}|h|^3\right) \cdot e^{-\sqrt{14}|h|} \quad (\text{A.16d})$$

De plus, il est possible de considérer la limite de cette fonction quand la régularité tend vers l'infini. On définit alors :

$$\mathcal{M}_{+\infty}(h) = e^{-h^2}, \quad (\text{A.17})$$

ce qui inclue dans la famille la fonction de covariance exponentielle au carrée. Ces quatre fonctions de covariance de Matérn (équations (A.16) et (A.17)) sont représentées figure 81.

A.3 Intégration des hyper-paramètres

Nous utilisons une approche complètement bayésienne pour mener les analyses. Connaissant les observations χ_n , et la loi de probabilité a priori sur les hyper-paramètres $\pi(\theta)$, la loi de probabilité a posteriori des hyper-paramètres θ est toujours calculable, $\pi(\theta|\chi_n)$. Soit Z les sorties aux points Y . En utilisant le théorème A.1.1, il est possible de calculer pour chaque valeur de θ la loi a posteriori des sorties, $\pi(Z|\theta, \chi_n)$, qui est une loi normale

de vecteur moyenne $m_{n,\theta}(Y)$ et de matrice de covariance $k_{n,\theta}(Y, Y)$. Dans ces conditions, il est possible d'utiliser la formule des probabilités totales pour obtenir la loi a posteriori intégrée des sorties

$$\pi(Z|\chi_n) = \int_{\theta \in \Theta} \pi(Z|\theta, \chi_n) \pi(\theta|\chi_n) d\theta. \quad (\text{A.18})$$

Dans le cas où seules la moyenne et la variance a posteriori des sorties sont nécessaires, les formules suivantes sont applicables

$$\mathbb{E}[Z|\chi_n] = \mathbb{E}[m_{n,\theta}(Y)|\chi_n] = \int_{\Theta} m_{n,\theta}(Y) \pi(\theta|\chi_n) d\theta \quad (\text{A.19a})$$

$$\begin{aligned} \text{var}[Z|\chi_n] &= \mathbb{E}[k_{n,\theta}(Y, Y)|\chi_n] + \text{var}[m_{n,\theta}(Y)|\chi_n] \\ &= \int_{\Theta} k_{n,\theta}(Y, Y) \pi(\theta|\chi_n) d\theta + \int_{\Theta} (m_{n,\theta}(Y) - \mathbb{E}[Z|\chi_n])^2 \pi(\theta|\chi_n) d\theta. \end{aligned} \quad (\text{A.19b})$$

Les intégrales des équations (A.18) et (A.19) sont alors approchées par une méthode de Monte-Carlo. Pour plus de détails, nous vous recommandons Robert [2014].

A.3.1 Échantillon initial des hyper-paramètres

Nous commençons par simuler un échantillon de p vecteurs d'hyper-paramètres (θ_j) selon leur lois a posteriori $\pi(\theta|\chi_n)$ grâce à l'algorithme de Haario et al. [2001] présenté algorithme 1. Puis, nous remplaçons les intégrales par des sommes finies, et obtenir les estimations voulues

$$\hat{\pi}(Z|\chi_n) = \frac{1}{p} \sum_{j=1}^p \pi(Z|\theta_j, \chi_n) \quad (\text{A.20a})$$

$$\widehat{m}_n(Y) = \frac{1}{p} \sum_{j=1}^p m_{n,\theta_j}(Y) \quad (\text{A.20b})$$

$$\widehat{k}_n(Y, Y) = \frac{1}{p} \sum_{j=1}^p k_{n,\theta_j}(Y, Y) + \left(1 + \frac{1}{p}\right) \frac{1}{p-1} \sum_{j=1}^p (m_{n,\theta_j}(Y) - \widehat{m}_n(Y))^2. \quad (\text{A.20c})$$

Le facteur $1 + \frac{1}{p}$ correspond à la variance sur la moyenne, $\text{var}[m_{n,\theta}(Y)|\chi_n]$, plus la variance qui se rajoute avec la méthode Monte-Carlo, $\text{var}[m_{n,\theta}(Y)|\chi_n]/p$.

La figure 82 présente un exemple en dimension $d = 1$, avec $n = 4$ observations. Les hyper-paramètres sont la variance σ^2 et la portée ρ . Une loi a priori log-normale est ajoutée sur les hyper-paramètres. Les hyper-paramètres sont intégrés selon leurs lois a posteriori. Les $p = 200$ hyper-paramètres ainsi obtenues sont représentées figure 82b.

Algorithme 1 Algorithme de Métropolis adaptatif [Haario et al., 2001]

Require: — π la loi cible, $\pi(\theta) = \pi(\chi_n|\theta)\pi(\theta)$;
 — $\theta_0 \in \mathbb{R}^{d^\theta}$ la graine, un vecteur d'hyper-paramètre de longueur d^θ ;
 — C_0 une covariance initiale pour la propositionnelle ;
 — $\epsilon > 0$;
 — $p \in \mathbb{N}^*$ le nombre d'hyper-paramètres souhaités ;
 — $t_0 \in \mathbb{N}^*$ le temps d'adaptation de la covariance ;
 — $t_1 \in \mathbb{N}^*$ le temps de chauffe supplémentaire ;
 — $l \in \mathbb{N}^*$ un décalage.

- 1: $a \leftarrow 0$, nombre d'acceptations
- 2: **for** $k = 1$ à $t_0 + t_1 + l \cdot (p - 1)$ **do**
- 3: {Algorithme de Métropolis-Hastings standard}
- 4: $y \sim \mathbf{N}(\theta_{k-1}, C_{k-1})$, tire un vecteur d'hyper-paramètres selon la propositionnelle
- 5: $u \sim \mathbf{U}([0; 1])$
- 6: **if** $u \leq \frac{\pi(y)}{\pi(\theta_k)}$ **then**
- 7: $\theta_k \leftarrow y$, accepte y avec une probabilité $\min \left\{ \frac{\pi(y)}{\pi(\theta_k)}, 1 \right\}$
- 8: $a \leftarrow a + 1$, incrémente le nombre d'acceptations
- 9: **else**
- 10: $\theta_k \leftarrow \theta_{k-1}$, rejette y
- 11: **end if**
- 12: {Phase d'adaptation}
- 13: $C_k \leftarrow \frac{t_0}{k+t_0}C_0 + \frac{k}{k+t_0}s_{d^\theta} \{\text{cov}[\theta_0, \dots, \theta_k] + \epsilon I_d\}$, moyenne pondérée entre C_0 et la covariance empirique des tirages. $s_{d^\theta} = \frac{2 \cdot 4^2}{d^\theta}$ et ϵ permet d'éviter les dégénérescences.
- 14: **end for**
- 15: **return** $\left(\theta_{t_0+t_1+(k-1) \cdot l} \right)_{1 \leq k \leq p}$
- 16: **return** $a / (t_0 + t_1 + l \cdot (p - 1))$, taux d'acceptation

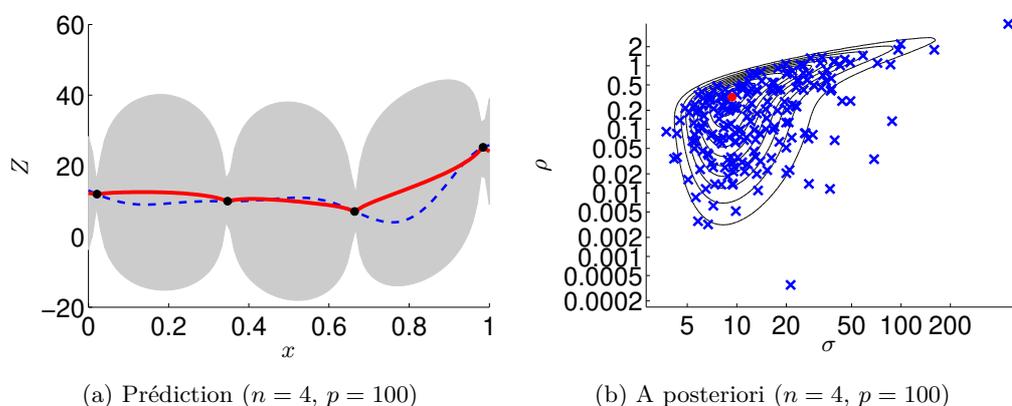


FIGURE 82 – Une régression par processus gaussien, avec intégration des hyper-paramètres selon leurs lois a posteriori. (a) Les prédictions. Le trait plein est la prédiction, la zone est l'intervalle à deux sigmas, le trait pointillé est la fonction dont sont issues les observations, les points sont les observations. (b) La loi de probabilité a posteriori des hyper-paramètres. Le point rouge indique l'argument du maximum. Les 200 croix indiquent les tirages obtenus par la méthode d'Haario, qui correspondent aux hyper-paramètres utilisés.

A.3.2 Mise à jour de l'échantillon des hyper-paramètres

Pour pouvoir appliquer une méthode complètement bayésienne, il faut disposer d'une liste d'hyper-paramètres $(\theta_j)_{1 \leq j \leq p}$ distribués selon la loi a posteriori $\pi(\theta|\chi_n)$. Or, dans une approche séquentielle, de nouvelles observations s'ajoutent régulièrement. Pour pouvoir appliquer une approche complètement bayésienne, il faut donc mettre à jour régulièrement la liste des hyper-paramètres. Pour cela, nous utilisons un algorithme de Monte-Carlo séquentielle développé par Chopin [2002] et présenté algorithme 2.

Les formules de calcul de moyenne et de variance sous pondération sont

$$W = \sum_{j=1}^p w_j, \quad \Omega = \sum_{j=1}^p w_j^2, \quad \hat{m} = \frac{1}{W} \sum_{j=1}^p w_j \theta_j,$$

$$\hat{v} = \frac{W}{W^2 - \Omega} \sum_{j=1}^p w_j (\theta_j - \hat{m})^2 = \frac{W}{W^2 - \Omega} \sum_{j=1}^p w_j \theta_j^2 - \frac{W^2}{W^2 - \Omega} \hat{m}^2,$$

où \hat{m} est d'estimateur de la moyenne des hyper-paramètres θ et \hat{v} d'estimateur de la variance des hyper-paramètres θ .

Algorithme 2 Algorithme de Monte-Carlo Séquentiel [Chopin, 2002]

Require: — $(\pi_n, \pi_{n+m}) \in \{\mathbb{R}^{d^\theta} \rightarrow \mathbb{R}^+\}^2$ les loi a posteriori avec respectivement n et $n + m$ points (pas forcément normalisées) ;

- $(\theta_j)_{1 \leq j \leq p} \in (\mathbb{R}^{d^\theta})^p$ un échantillon de p hyper-paramètres de dimension d^θ ;
- $(w_j)_{1 \leq j \leq p} \in (\mathbb{R}^+)^p$ les pondérations associées aux hyper-paramètres ;
- p_{eff}^{crit} taille effective minimale de l'échantillon ;
- MCMC un algorithme engendrant une chaîne de Markov selon une loi cible à partir d'une graine (exemples : Metropolis-Hastings, Metropolis adaptatif, ...).

- 1: $\forall j \in [1; p], w_j \leftarrow w_j \frac{\pi_{n+m}(\theta_j)}{\pi_n(\theta_j)}$ {Repondération selon la nouvelle loi a posteriori}
- 2: $\forall j \in [1; p], w_j \leftarrow w_j / \sum_{j=1}^p w_j$ {Normalisation}
- 3: $p_{eff} \leftarrow \frac{(\sum_{j=1}^p w_j)^2}{(\sum_{j=1}^p w_j^2)}$ {Calcul de la taille effective de l'échantillon}
- 4: **if** $p_{eff} < p_{eff}^{crit}$ **then**
- 5: {Si la taille réelle de l'échantillon est trop faible}
- 6: $\forall j \in [1; p], x_j \leftarrow \lfloor pw_j \rfloor$
- 7: $X \leftarrow \sum_{j=1}^p x_j$ {Nombre d'éléments qui sont choisis directement}
- 8: Création une nouvelle liste $(\tilde{\theta}_k)_{1 \leq k \leq X}$ où chaque élément θ_j apparait x_j fois.
- 9: $\forall j \in [1; p], z_j \leftarrow w_j - x_j/p$
- 10: $z_j \leftarrow z_j / \sum_{j=1}^p z_j$ {Loi de probabilité des résidus}
- 11: Complétion la nouvelle liste en tirant $p - X$ élément(s) selon la loi $z : \forall k \in [X+1; p], \mathbb{P}(\tilde{\theta}_k = \theta_j) = z_j$. {Phase de ré-échantillonnage}
- 12: $\forall j \in [1; p], w_j \leftarrow 1/p$ {Ré-équilibrer les poids.}
- 13: $\forall j \in [1; p], \theta_j \leftarrow \text{MCMC}(\tilde{\theta}_j, \pi_{n+m})$ {Propagation des hyper-paramètres selon la nouvelle loi a posteriori}
- 14: **end if**
- 15: **return** $(\theta_j)_{1 \leq j \leq p}, (w_j)_{1 \leq j \leq p}$

Annexe B

Théorèmes et preuves

B.1 Construction de modèles multi-fidèles

B.1.1 Principal théorème de construction

Théorème B.1.1 (Relation de fidélité entre des vecteurs aléatoires gaussiens).
Soit A , B et C trois vecteurs aléatoires conjointement gaussiens. Si $\text{var}[C]$ est une matrice inversible, alors les quatre propositions suivantes sont équivalentes

1. $\text{cov}[A, B|C] = 0$;
2. $\text{cov}[A, B] = \text{cov}[A, C]\text{var}[C]^{-1}\text{cov}[C, B]$;
3. *il existe une matrice réelle γ et un vecteur aléatoire η conjointement gaussien avec A , B et C tel que $A = \gamma C + \eta$, avec η indépendant de B et de C ;*
4. $\mathbb{E}[A|B, C] = \mathbb{E}[A|C]$.

Et alors, $\gamma = \text{cov}[A, C]\text{var}[C]^{-1}$, et $\mathbb{E}[A|B, C] = \gamma C + \mathbb{E}[\eta]$.

Démonstration. La démonstration est cyclique.

Supposons que $\text{cov}[A, B|C] = 0$. Notons $A_c = A - \mathbb{E}[A]$ et $B_c = B - \mathbb{E}[B]$ les variables centrées respectives de A et B . Dans ces conditions, $\text{cov}[A, B|C] = \text{cov}[A_c, B_c|C] = 0 \Leftrightarrow \mathbb{E}[A_c B_c^T | C] = \mathbb{E}[A_c | C]\mathbb{E}[B_c | C]^T$. Alors

$$\begin{aligned}\text{cov}[A, B] &= \text{cov}[A_c, B_c] \\ &= \mathbb{E}[A_c B_c^T] \\ &= \mathbb{E} \left[\mathbb{E}[A_c B_c^T | C] \right]\end{aligned}$$

$$\begin{aligned}
&= \mathbb{E} \left[\mathbb{E}[A_c|C] \mathbb{E}[B_c|C]^T \right] \\
&= \mathbb{E} \left[\left(\text{cov}[A_c, C] \text{var}[C]^{-1} (C - \mathbb{E}[C]) \right) \cdot \left((C - \mathbb{E}[C])^T \text{var}[C]^{-1} \text{cov}[C, B_c] \right) \right] \\
&= \mathbb{E}[\text{cov}[A_c, C] \text{var}[C]^{-1} (C - \mathbb{E}[C]) (C - \mathbb{E}[C])^T \text{var}[C]^{-1} \text{cov}[C, B_c]] \\
&= \text{cov}[A_c, C] \text{var}[C]^{-1} \mathbb{E}[(C - \mathbb{E}[C]) (C - \mathbb{E}[C])^T] \text{var}[C]^{-1} \text{cov}[C, B_c] \\
&= \text{cov}[A_c, C] \text{var}[C]^{-1} \text{var}[C] \text{var}[C]^{-1} \text{cov}[C, B_c] \\
&= \text{cov}[A_c, C] \text{var}[C]^{-1} \text{cov}[C, B_c] \\
&= \text{cov}[A, C] \text{var}[C]^{-1} \text{cov}[C, B].
\end{aligned}$$

Supposons que $\text{cov}[A, B] = \text{cov}[A, C] \text{var}[C]^{-1} \text{cov}[C, B]$, notons $\gamma = \text{cov}[A, C] \text{var}[C]^{-1}$ et $\eta = A - \gamma C$. Trivialement, $A = \gamma C + \eta$, et η est un vecteur conjointement gaussien avec A , B et C . Or,

$$\begin{aligned}
\text{cov}[\eta, C] &= \text{cov}[A - \gamma C, C] \\
&= \text{cov}[A, C] - \gamma \text{var}[C] \\
&= \text{cov}[A, C] - \text{cov}[A, C] \text{var}[C]^{-1} \text{var}[C] \\
&= 0,
\end{aligned}$$

η et C sont décorrélés, donc indépendants (car gaussiens). De plus,

$$\begin{aligned}
\text{cov}[\eta, B] &= \text{cov}[A - \gamma C, B] \\
&= \text{cov}[A, B] - \gamma \text{cov}[C, B] \\
&= \text{cov}[A, C] \text{var}[C]^{-1} \text{cov}[C, B] - \text{cov}[A, C] \text{var}[C]^{-1} \text{cov}[C, B] \\
&= 0
\end{aligned}$$

η et C sont eux-aussi décorrélés, donc indépendants.

Supposons que $A = \gamma C + \eta$, avec γ une matrice connue, et η un vecteur aléatoire gaussien indépendant de B et C , alors

$$\begin{aligned}
\mathbb{E}[A|B, C] &= \mathbb{E}[\gamma C + \eta|B, C] \\
&= \gamma \mathbb{E}[C|B, C] + \mathbb{E}[\eta|B, C] \\
&= \gamma C + \mathbb{E}[\eta] \\
&= \gamma \mathbb{E}[C|C] + \mathbb{E}[\eta|C] \\
&= \mathbb{E}[\gamma C + \eta|C]
\end{aligned}$$

$$= \mathbb{E}[A|C].$$

Enfin, si $\mathbb{E}[A|B, C] = \mathbb{E}[A|C]$, alors

$$\begin{aligned} \mathbb{E}[AB^T|C] &= \mathbb{E}[\mathbb{E}[AB^T|B, C]|C] \\ &= \mathbb{E}[\mathbb{E}[A|B, C]B^T|C] \\ &= \mathbb{E}[\mathbb{E}[A|C]B^T|C] \\ &= \mathbb{E}[A|C]\mathbb{E}[B^T|C] \\ \Leftrightarrow \text{cov}[A, B|C] &= 0. \end{aligned}$$

Ce qu'il fallait démontrer.

B.1.2 Modèle auto-régressif

Proposition B.1.1 (Construction du modèle auto-régressif à S niveaux). *Soit ξ un processus gaussien défini sur $\mathbb{X} \times \llbracket 1; S \rrbracket$. Alors,*

$$\forall (s, s', s''; x, x') \in \llbracket 1; S \rrbracket^3 \times \mathbb{X}^2, \quad s \geq s' \geq s'' \Rightarrow \text{cov} [\xi(x, s), \xi(x', s'') | \xi(x, s')] = 0,$$

si et seulement si, il existe $S - 1$ fonctions $(\gamma_s)_{1 \leq s \leq S-1}$ et S processus $(\eta_s)_{1 \leq s \leq S}$ conjointement gaussiens et indépendants tels que

$$\begin{cases} \xi(x, 1) = \eta_1(x) \\ \xi(x, s+1) = \gamma_s(x)\xi(x, s) + \eta_{s+1}(x), \quad s \in \llbracket 1; S-1 \rrbracket \end{cases}$$

Démonstration. Cette preuve est en deux parties : d'abord la preuve de l'implication directe, puis celle de l'implication réciproque.

D'abord, supposons que

$$\forall (s, s', s''; x, x') \in \llbracket 1; S \rrbracket^3 \times \mathbb{X}^2, \quad s \geq s' \geq s'' \Rightarrow \text{cov} [\xi(x, s), \xi(x', s'') | \xi(x, s')] = 0,$$

alors en particulier pour $s = s' + 1$ et $s' = s''$, $\text{cov} [\xi(x, s' + 1), \xi(x', s') | \xi(x, s')] = 0$. Le théorème B.1.1 donne la relation suivante

$$\xi(x, s' + 1) = \gamma_{s'}(x)\xi(x, s') + \eta_{s'+1}(x),$$

où $\gamma_{s'}(x) = \text{cov}[\xi(x, s' + 1), \xi(x, s')]/\text{var}[\xi(x, s')]$ et $\eta_{s'+1}(x) = \xi(x, s' + 1) - \gamma_{s'}(x)\xi(x, s')$ est indépendant de $\xi(\cdot, s')$. Notons $\eta_1 = \xi(\cdot, 1)$. Comme tous les η_s sont des combinaisons linéaires du processus gaussien ξ , ils forment des processus conjointement gaussiens.

Montrons que les η_s sont mutuellement indépendants. Pour cela, il suffit de montrer qu'ils sont décorrélés. Montrons déjà que $\forall s \in \llbracket 1; S-1 \rrbracket$, η_{s+1} est décorrélé avec $(\xi(\cdot, t))_{1 \leq t \leq s}$. En effet, soit $t < s + 1$, alors

$$\text{cov}[\eta_{s+1}(x), \xi(x', t)] = \text{cov}[\xi(x, s + 1), \xi(x', t)] - \gamma_s(x)\text{cov}[\xi(x, s), \xi(x', t)].$$

Or, selon l'hypothèse de base,

$$\text{cov}[\xi(x, s + 1), \xi(x', t)|\xi(x, s)] = 0.$$

Selon le théorème B.1.1, cela est équivalent à

$$\begin{aligned} \text{cov}[\xi(x, s + 1), \xi(x', t)] &= \left(\frac{\text{cov}[\xi(x, s + 1), \xi(x, s)]}{\text{var}[\xi(x, s)]} \right) \text{cov}[\xi(x, s), \xi(x', t)] \\ &= \gamma_s(x)\text{cov}[\xi(x, s), \xi(x', t)] \\ \Leftrightarrow \text{cov}[\eta_{s+1}(x), \xi(x', t)] &= 0. \end{aligned}$$

Finalement, si $s > 1$, alors

$$\text{cov}[\eta_s(x), \eta_1(x')] = \text{cov}[\eta_s(x), \xi(x', 1)] = 0,$$

et si $s > s' > 1$, alors

$$\text{cov}[\eta_s(x), \eta_{s'}(x')] = \text{cov}[\eta_s(x), \xi(x', s')] - \gamma_{s'-1}(x')\text{cov}[\eta_s(x), \xi(x', s' - 1)] = 0.$$

Donc, les processus $(\eta_s)_{1 \leq s \leq S}$ sont décorrélés. Donc, comme ils sont conjointement gaussiens, ils sont conjointement indépendants.

Montrons, maintenant l'implication réciproque, supposons le modèle

$$\begin{cases} \xi(x, 1) = \eta_1(x) \\ \xi(x, s + 1) = \gamma_s(x)\xi(x, s) + \eta_{s+1}(x), \quad s \in \llbracket 1; S-1 \rrbracket \end{cases}$$

où les $(\gamma_s)_{1 \leq s \leq S-1}$ sont $S - 1$ fonctions connues, et les $(\eta_s)_{1 \leq s \leq S}$ S processus conjointement gaussiens et indépendants. Dans ces conditions, on peut aisément montrer par

réurrence que

$$\xi(x, s) = \sum_{k=1}^s \left[\prod_{j=k}^{s-1} \gamma_j(x) \right] \eta_k(x),$$

avec $\prod_{j=s}^{s-1} \gamma_j(x) = 1$. Soit $s \geq s' \geq s''$, trois niveaux de fidélité. Dans ces conditions,

$$\begin{aligned} \xi(x, s) &= \sum_{k=1}^s \left[\prod_{j=k}^{s-1} \gamma_j(x) \right] \eta_k(x) \\ &= \sum_{k=1}^{s'} \left[\prod_{j=k}^{s-1} \gamma_j(x) \right] \eta_k(x) + \sum_{k=s'+1}^s \left[\prod_{j=k}^{s-1} \gamma_j(x) \right] \eta_k(x) \\ &= \left[\prod_{j=s'}^{s-1} \gamma_j(x) \right] \sum_{k=1}^{s'} \left[\prod_{j=k}^{s'-1} \gamma_j(x) \right] \eta_k(x) + \sum_{k=s'+1}^s \left[\prod_{j=k}^{s-1} \gamma_j(x) \right] \eta_k(x) \\ &= \left[\prod_{j=s'}^{s-1} \gamma_j(x) \right] \xi(x, s') + \sum_{k=s'+1}^s \left[\prod_{j=k}^{s-1} \gamma_j(x) \right] \eta_k(x). \end{aligned}$$

Comme $\xi(\cdot, s')$ s'exprime comme une combinaison linéaire des $(\eta_t)_{1 \leq t \leq s'}$, il est indépendant des $(\eta_t)_{s'+1 \leq t \leq s}$. Par conséquent,

$$\text{cov}[\xi(x, s), \xi(x', s')] = \left[\prod_{j=s'}^{s-1} \gamma_j(x) \right] \text{cov}[\xi(x, s'), \xi(x', s')].$$

De la même façon,

$$\begin{aligned} \text{cov}[\xi(x, s), \xi(x', s'')] &= \left[\prod_{j=s''}^{s-1} \gamma_j(x) \right] \text{cov}[\xi(x, s''), \xi(x', s'')] \\ \text{cov}[\xi(x, s'), \xi(x', s'')] &= \left[\prod_{j=s''}^{s'-1} \gamma_j(x) \right] \text{cov}[\xi(x, s''), \xi(x', s'')]. \end{aligned}$$

Dans ces conditions,

$$\begin{aligned}
& \text{cov}[\xi(x, s), \xi(x, s')] \text{cov}[\xi(x, s'), \xi(x', s'')] \\
&= \left[\prod_{j=s'}^{s-1} \gamma_j(x) \right] \text{var}[\xi(x, s')] \left[\prod_{j=s''}^{s'-1} \gamma_j(x) \right] \text{cov}[\xi(x, s''), \xi(x', s'')] \\
&= \text{var}[\xi(x, s')] \left[\prod_{j=s''}^{s-1} \gamma_j(x) \right] \text{cov}[\xi(x, s''), \xi(x', s'')] \\
&= \text{var}[\xi(x, s')] \text{cov}[\xi(x, s), \xi(x', s'')] \\
\Leftrightarrow \text{cov}[\xi(x, s), \xi(x', s'')] &= \frac{\text{cov}[\xi(x, s), \xi(x, s')]}{\text{var}[\xi(x, s')]} \text{cov}[\xi(x, s'), \xi(x', s'')].
\end{aligned}$$

Or, selon le théorème B.1.1, cette égalité est équivalente à

$$\text{cov}[\xi(x, s), \xi(x', s'') | \xi(x, s')] = 0.$$

Ce qu'il fallait démontrer.

B.1.3 Modèle non-stationnaire

Proposition B.1.2 (Construction du modèle non-stationnaire). *Soit ξ un processus gaussien défini sur $\mathbb{X} \times \mathbb{R}^+$. Alors,*

$$\forall (x, \delta) \in \mathbb{X} \times \mathbb{R}^+, \mathbb{E}[\xi(x, \delta) | \xi(\cdot, 0)] = \xi(x, 0),$$

si et seulement si, il existe ϵ un processus centré conjointement gaussien et indépendant de $\xi(\cdot, 0)$, avec $\epsilon(x, 0) = 0$ tel que

$$\xi(x, \delta) = \xi(x, 0) + \epsilon(x, \delta).$$

Démonstration. Ce théorème est une application directe du théorème B.1.1 avec $A = \xi(x, \delta)$, $B = \xi(\cdot, 0)$ et $C = \xi(x, 0)$. Cette preuve est en deux parties : d'abord la preuve de l'implication directe, puis celle de l'implication réciproque.

Supposons que

$$\forall (x, \delta) \in \mathbb{X} \times \mathbb{R}^+, \mathbb{E}[\xi(x, \delta) | \xi(\cdot, 0)] = \xi(x, 0),$$

Dans ces conditions,

$$\begin{aligned}\mathbb{E}[\xi(x, \delta)|\xi(x, 0)] &= \mathbb{E}[\mathbb{E}[\xi(x, \delta)|\xi(\cdot, 0)]|\xi(x, 0)] \\ &= \mathbb{E}[\xi(x, 0)|\xi(x, 0)] \\ &= \xi(x, 0) \\ &= \mathbb{E}[\xi(x, \delta)|\xi(\cdot, 0)].\end{aligned}$$

Donc, $\mathbb{E}[\xi(x, \delta)|\xi(x, 0)] = \mathbb{E}[\xi(x, \delta)|\xi(\cdot, 0)]$. Donc, selon le théorème B.1.1,

$$\xi(x, \delta) = \gamma(x, \delta)\xi(x, 0) + \epsilon(x, \delta),$$

avec ϵ un processus gaussien indépendant de $\xi(\cdot, 0)$. De plus, $\mathbb{E}[\xi(x, \delta)|\xi(x, 0)] = \xi(x, 0)$, ce qui implique que $\gamma(x, \delta) = 1$ et que $\mathbb{E}[\epsilon(x, \delta)] = 0$. Donc,

$$\xi(x, \delta) = \xi(x, 0) + \epsilon(x, \delta).$$

Enfin, en posant $\delta = 0$ dans la précédente équation, $\epsilon(x, 0) = 0$.

Supposons maintenant le modèle,

$$\xi(x, \delta) = \xi(x, 0) + \epsilon(x, \delta).$$

avec ϵ un processus gaussien centré indépendant de $\xi(\cdot, 0)$, avec $\epsilon(x, 0) = 0$. Le modèle s'écrit alors sous la forme du théorème B.1.1. Donc,

$$\mathbb{E}[\xi(x, \delta)|\xi(x, \cdot)] = \mathbb{E}[\xi(x, \delta)|\xi(x, 0)] = \xi(x, 0).$$

Ce qu'il fallait démontrer.

Proposition B.1.3 (Modèle non-stationnaire avec relation brownienne). *Soit ξ un processus gaussien défini sur $\mathbb{X} \times \mathbb{R}^+$. Alors,*

$$\delta \geq \delta' \Rightarrow \forall x \in \mathbb{X}, \mathbb{E}[\xi(x, \delta)|\xi(\cdot, \delta')] = \xi(x, \delta'),$$

si et seulement si, il existe ϵ un processus centré conjointement gaussien et indé-

pendant de $\xi(\cdot, 0)$, avec $\epsilon(x, 0) = 0$ tel que

$$\xi(x, \delta) = \xi(x, 0) + \epsilon(x, \delta),$$

et

$$\text{cov}[\epsilon(x, \delta), \epsilon(x', \delta')] = \text{cov}[\epsilon(x, \min\{\delta, \delta'\}), \epsilon(x', \min\{\delta, \delta'\})].$$

Démonstration. D'abord, si

$$\delta \geq \delta' \Rightarrow \forall x \in \mathbb{X}, \mathbb{E} [\xi(x, \delta) | \xi(\cdot, \delta')] = \xi(x, \delta'),$$

alors en vertu du proposition B.1.2, $\forall \delta \geq \delta' \geq 0$, les trois formules suivantes peuvent être écrites

$$\begin{aligned} \xi(x, \delta) &= \xi(x, 0) + \eta(x, \delta; 0) \\ &= \xi(x, \delta') + \eta(x, \delta; \delta') \\ \xi(x, \delta') &= \xi(x, 0) + \eta(x, \delta'; 0), \end{aligned}$$

où $\eta(\cdot, \cdot; 0) = \epsilon$ est indépendant de $\xi(\cdot, 0)$, et $\eta(\cdot, \cdot; \delta')$ est indépendant de $\xi(\cdot, \delta')$. Dans ces conditions, $\eta(x, \delta; \delta') = \eta(x, \delta; 0) - \eta(x, \delta'; 0) = \epsilon(x, \delta) - \epsilon(x, \delta')$. Or,

$$\begin{aligned} &\text{cov}[\eta(x, \delta; \delta'), \xi(x', \delta')] = 0 \\ \Leftrightarrow &\text{cov}[\epsilon(x, \delta) - \epsilon(x, \delta'), \xi(x', 0) + \epsilon(x', \delta')] = 0 \\ \Leftrightarrow &\text{cov}[\epsilon(x, \delta) - \epsilon(x, \delta'), \epsilon(x', \delta')] = 0 \\ \Leftrightarrow &\text{cov}[\epsilon(x, \delta), \epsilon(x', \delta')] = \text{cov}[\epsilon(x, \delta'), \epsilon(x', \delta')] \\ &= \text{cov}[\epsilon(x, \min\{\delta, \delta'\}), \epsilon(x', \min\{\delta, \delta'\})]. \end{aligned}$$

Réciproquement, supposons que

$$\xi(x, \delta) = \xi(x, 0) + \epsilon(x, \delta),$$

avec ϵ un processus centré conjointement gaussien et indépendant de $\xi(\cdot, 0)$, avec $\epsilon(x, 0) = 0$ et $\text{cov}[\epsilon(x, \delta), \epsilon(x', \delta')] = \text{cov}[\epsilon(x, \min\{\delta, \delta'\}), \epsilon(x', \min\{\delta, \delta'\})]$. Dans ces conditions, nous pouvons écrire pour $\delta \geq \delta'$

$$\xi(x, \delta) = \xi(x, \delta') + \epsilon(x, \delta) - \epsilon(x, \delta') = \xi(x, \delta') + \eta(x, \delta; \delta')$$

Dans ces conditions, $\eta(\cdot, \cdot; \delta')$ est un processus gaussien centré qui vaut 0 quand $\delta = \delta'$. De plus,

$$\begin{aligned} \text{cov}[\xi(x, \delta'), \eta(x', \delta; \delta')] &= \text{cov}[\xi(x, 0) + \epsilon(x, \delta'), \epsilon(x, \delta) - \epsilon(x, \delta')] \\ &= \text{cov}[\epsilon(x, \delta'), \epsilon(x, \delta)] - \text{cov}[\epsilon(x, \delta'), \epsilon(x, \delta')] \\ &= \text{cov}[\epsilon(x, \delta'), \epsilon(x, \delta')] - \text{cov}[\epsilon(x, \delta'), \epsilon(x, \delta')] \\ &= 0, \end{aligned}$$

les processus $\xi(\cdot, \delta')$ et $\eta(\cdot, \cdot; \delta')$ sont conjointement gaussiens et décorrélés. Ils sont donc indépendant. Donc, en vertu du proposition B.1.2,

$$\mathbb{E} [\xi(x, \delta) | \xi(\cdot, \delta')] = \xi(x, \delta').$$

Ce qu'il fallait démontrer.

B.2 Inversion des niveaux

Proposition B.2.1 (Inversion des niveaux). *Soit Z_1, Z_2 et η trois processus conjointement gaussiens, stationnaires, définis sur \mathbb{X} . Supposons qu'il existe une fonction f connue telle que*

$$Z_2(x) = (f * Z_1)(x) + \eta(x),$$

et que Z_1 et η soient indépendants. Soit $c_1(x-x') = \text{cov}[Z_1(x), Z_1(x')]$ et $c_\eta(x-x') = \text{cov}[\eta(x), \eta(x')]$. Supposons que c_1, c_η et f admettent des transformées de Fourier, respectivement C_1, C_η et F . Alors, il existe une fonction g et un processus gaussien ϵ indépendant de Z_2 , tel que

$$Z_1(x) = (g * Z_2)(x) + \epsilon(x). \tag{B.1}$$

Dans ces conditions,

$$\begin{aligned} g &= \text{TF}^{-1} \left[\frac{F^* C_1}{|F|^2 C_1 + C_\eta} \right] \\ \epsilon &= \left(\text{TF}^{-1} \left[\frac{C_\eta}{|F|^2 C_1 + C_\eta} \right] * Z_1 \right) - (g * \eta). \end{aligned}$$

Démonstration. Déjà, remarquons qu'en vertu du théorème de Bochner, C_1 et C_δ sont réels et positifs. En particulier, $C_1^* = C_1$ et $C_\delta^* = C_\delta$.

Notons $g = \text{TF}^{-1}[G]$, avec $G = \frac{F^*C_1}{|F|^2C_1 + C_\eta}$. Évidemment, $Z_1 = (g * Z_2) + (Z_1 - (g * Z_2))$. Soit $\epsilon = Z_1 - (g * Z_2)$, et montrons que Z_2 et ϵ sont indépendants. Déjà, ϵ est un processus gaussien, car résultat d'une application linéaire appliquée à des processus conjointement gaussiens. Calculons alors la covariance entre $\epsilon(x)$ et $Z_2(x')$:

$$\begin{aligned} \text{cov}[\epsilon(x), Z_2(x')] &= \text{cov}[Z_1(x) - (g * Z_2)(x), Z_2(x')] \\ &= \text{cov}[Z_1(x) - (g * f * Z_1)(x) - (g * \eta)(x), (f * Z_1)(x') + \eta(x')] \\ &= \text{cov}[Z_1(x), (f * Z_1)(x')] - \text{cov}[(g * f * Z_1)(x), (f * Z_1)(x')] \\ &\quad - \text{cov}[(g * \eta)(x), (f * Z_1)(x')] + \text{cov}[Z_1(x), \eta(x')] \\ &\quad - \text{cov}[(g * f * Z_1)(x), \eta(x')] - \text{cov}[(g * \eta)(x), \eta(x')]. \end{aligned}$$

Or, par la formule des inférences, la covariance entre ϵ et Z_2 est stationnaire, et en appliquant la transformée de Fourier,

$$\begin{aligned} \text{TF}[h \mapsto \text{cov}[\epsilon(0), Z_2(0+h)]] &= C_1 F^* - G F C_1 F^* - G \cdot 0 \cdot F^* + 0 - G F \cdot 0 - G C_\eta \\ &= F^* C_1 - G(|F|^2 C_1 + C_\eta) \\ &= 0 \\ \Leftrightarrow \text{cov}[\epsilon(0), Z_2(0+h)] &= \text{cov}[\epsilon(x), Z_2(x')] = 0. \end{aligned}$$

Donc, ϵ et Z_2 sont décorrélés et comme ils sont conjointement gaussiens, ils sont indépendants. *Ce qu'il fallait démontrer.*

Conséquence B.2.1 (Inversion d'un modèle auto-régressif). Dans le cas particulier où $f = \gamma\delta$, δ étant l'impulsion de Dirac et γ un réel, la relation devient

$$Z_2(x) = \gamma Z_1(x) + \eta(x),$$

et la fonction g devient

$$g = \text{TF}^{-1} \left[\frac{\gamma C_1}{\gamma^2 C_1 + C_\eta} \right]. \quad (\text{B.2})$$

B.3 Espérance de probabilités de dépassement de seuil

Proposition B.3.1 (La formule de Chevalier et al. [2014]). *Soit $d \in \mathbb{N}^*$ une dimension, m et T deux vecteurs de longueur d , K et R deux matrices symétriques positives de taille d , \mathcal{A} un sous-ensemble de \mathbb{R}^d et $\Phi_d(\cdot, \mu; \Sigma)$ la fonction de répartition de la loi normale de moyenne μ et de variance Σ de dimension d . Si $X \sim \mathbf{N}(m, K)$, alors*

$$\mathbb{E}[\Phi_d(X, T; R)\mathbf{1}_{X \in \mathcal{A}}] = \mathbb{P}\left(\begin{pmatrix} Y \leq m \\ X \in \mathcal{A} \end{pmatrix}\right) \text{ où } \begin{pmatrix} Y \\ X \end{pmatrix} \sim \mathbf{N}\left(\begin{pmatrix} T \\ m \end{pmatrix}, \begin{pmatrix} K+R & -K \\ -K & K \end{pmatrix}\right). \quad (\text{B.3})$$

Démonstration. En effet,

$$\begin{aligned} \mathbb{E}[\Phi_d(X, T; R)\mathbf{1}_{X \in \mathcal{A}}] &= \int_{x \in \mathcal{A}} \Phi_d(x, T; R)\phi_d(x; m, K)dx \\ &= \int_{x \in \mathcal{A}} \mathbb{P}(Z \leq X | X = x)p_X(x)dx \\ &= \mathbb{P}(Z \leq X \wedge X \in \mathcal{A}) \end{aligned}$$

où $\phi_d(\cdot, \mu; \Sigma)$ est la densité de probabilité d'une loi normale de moyenne μ et de variance Σ , et Z est une variable aléatoire mutuellement gaussienne et indépendante de X , $Z \sim \mathbf{N}(T, R)$.

Notons $Y = Z - X + m$. Alors,

$$\mathbb{E}[\Phi_d(X, T; R)\mathbf{1}_{X \in \mathcal{A}}] = \mathbb{P}(Y \leq m \wedge X \in \mathcal{A})$$

et

$$\begin{pmatrix} Y \\ X \end{pmatrix} \sim \mathbf{N}\left(\begin{pmatrix} T \\ m \end{pmatrix}, \begin{pmatrix} R+K & -K \\ -K & K \end{pmatrix}\right).$$

Ce qu'il fallait démontrer.

Conséquences B.3.1 (Cas particuliers). En particulier,

1. Si $\mathcal{A} = \mathbb{R}^d$,

$$\mathbb{E}[\Phi_d(X, T; R)] = \Phi_d(m, T, K + R); \quad (\text{B.4})$$

2. Si $\mathcal{A} = \mathbb{R}^d$, et $R = \text{diag} \{r_k\}_{1 \leq k \leq d}$ est une matrice diagonale,

$$\mathbb{E} \left[\prod_{k=1}^d \Phi_1(x_k, T_k; r_k) \right] = \Phi_d(m, T, K + \text{diag} \{r_k\}_{1 \leq k \leq d}); \quad (\text{B.5})$$

3. Si $d = 1$ et $\mathcal{A} =]a; b[\subset \mathbb{R}$,

$$\begin{aligned} \mathbb{E} [\Phi_1(X, T; R) \mathbf{1}_{a < X < b}] &= \Phi_2 \left(\begin{pmatrix} m \\ b \end{pmatrix}, \begin{pmatrix} T \\ m \end{pmatrix}; \begin{pmatrix} K + R & -K \\ -K & K \end{pmatrix} \right) \\ &\quad - \Phi_2 \left(\begin{pmatrix} m \\ a \end{pmatrix}, \begin{pmatrix} T \\ m \end{pmatrix}; \begin{pmatrix} K + R & -K \\ -K & K \end{pmatrix} \right). \end{aligned} \quad (\text{B.6})$$

Annexe C

Méthode d'estimation de la probabilité de défaillance

C.1 Hypothèses

Cette section décrit comment estimer la probabilité de dépassement de seuil d'une sortie à partir d'un modèle bayésien fondé sur un processus gaussien. Le simulateur est supposé stochastique. Il prend un vecteur d'entrée $x = (x_e, x_s)$ et un paramètre de fidélité δ et renvoie une sortie Z suivant une loi aléatoire dépendant de x et de δ . Le vecteur $x_e \in \mathbb{X}_e$ est le vecteur des entrées environnementales qui doivent être intégrées selon leur distribution f^{env} et le vecteur x_s est le vecteur des entrées scénario dont on souhaite évaluer l'impact. La Quantité d'Intérêt (QoI) à estimer est la fonction de probabilité de dépassement de seuil intégrée sur l'espace des entrées environnementales au niveau de référence δ^{ref} ,

$$P(x_s, \delta^{\text{ref}}) = \mathbb{P}(Z \geq z^{\text{crit}} | x_s, \delta = \delta^{\text{ref}}) = \int_{\mathbb{X}_e} \mathbb{P}(Z \geq z^{\text{crit}} | x_e, x_s, \delta = \delta^{\text{ref}}) f^{\text{env}}(x_e) dx_e, \quad (\text{C.1})$$

où z^{crit} est le seuil critique. On note $p(x_e, x_s, \delta) = \mathbb{P}(Z \geq z^{\text{crit}} | x_e, x_s, \delta)$. On dispose de n observations du simulateur $\chi_n = (x_i, \delta_i; z_i)_{1 \leq i \leq n}$.

Dans cette section, on suppose que la sortie du simulateur au point (x, δ) suit une loi normale de moyenne $\xi(x, \delta)$ et de variance $\lambda(x, \delta)$

$$Z | x, \delta \sim \text{N}(\xi(x, \delta), \lambda(x, \delta)). \quad (\text{C.2})$$

Deux simulations différentes sont supposées indépendantes. Le processus ξ est modélisé

par un processus gaussien conditionnement à des hyper-paramètres, et le processus λ est supposé connu conditionnement à d'autres hyper-paramètres. On note θ l'ensemble des hyper-paramètres. La loi a posteriori de ξ conditionnellement aux hyper-paramètres et aux observations est un processus gaussien de fonction de moyenne $m_{n,\theta}$ et de fonction de covariance $k_{n,\theta}$

$$\xi|\theta, \chi_n \sim \text{GP}(m_{n,\theta}, k_{n,\theta}). \quad (\text{C.3})$$

Enfin, on suppose que p échantillons indépendants d'hyper-paramètres distribuées selon leur lois a posteriori sont disponibles

$$\theta_j \sim p(\theta|\chi_n), \quad 1 \leq j \leq p, \quad (\text{C.4})$$

ainsi que q points environnementaux indépendants et identiquement distribuées selon leur distribution environnementale

$$x_e^{(l)} \sim f^{\text{env}}, \quad 1 \leq l \leq q. \quad (\text{C.5})$$

C.2 Moyenne et variance

Dans ces conditions, il est possible de calculer l'espérance et la covariance des probabilités locales $p(x_e, x_s, \delta)$ conditionnellement aux hyper-paramètres et aux observations

$$\mathbb{E}_{n,\theta} [p(y)] = \Phi(u_{n,\theta}(y)) \quad (\text{C.6a})$$

$$\text{cov}_{n,\theta} [p(y), p(y')] = \Phi 2(u_{n,\theta}(y), u_{n,\theta}(y'); r_{n,\theta}(y, y')) - \Phi(u_{n,\theta}(y)) \Phi(u_{n,\theta}(y')) \quad (\text{C.6b})$$

où $y = (x_e, x_s, \delta)$, $y' = (x'_e, x'_s, \delta')$, $\mathbb{E}_{n,\theta}[A] = \mathbb{E}[A|\chi_n, \theta]$, $\text{cov}_{n,\theta}[A, B] = \mathbb{E}_{n,\theta}[AB] - \mathbb{E}_{n,\theta}[A]\mathbb{E}_{n,\theta}[B]$, Φ est la fonction de répartition de la loi normale mono-dimensionnelle, $\Phi 2(\cdot, \cdot; r)$ celle de la loi normale bi-dimensionnelle de corrélation r et

$$u_{n,\theta}(y) = \frac{m_{n,\theta}(y) - z^{\text{crit}}}{\sqrt{k_{n,\theta}(y, y) + \lambda_\theta(y)}} \quad (\text{C.7a})$$

$$r_{n,\theta}(y, y') = \frac{k_{n,\theta}(y, y')}{\sqrt{k_{n,\theta}(y, y) + \lambda_\theta(y)} \sqrt{k_{n,\theta}(y', y') + \lambda_\theta(y')}}. \quad (\text{C.7b})$$

On note $\mu_{n,\theta}$ la fonction de moyenne a posteriori de p , $\kappa_{n,\theta}$ sa fonction de covariance et $v_{n,\theta}$ sa fonction de variance.

Ensuite, comme la probabilité globale $P(x_s, \delta^{\text{ref}})$ est l'intégrale de la probabilité locale $p(x_e, x_s, \delta^{\text{ref}})$, il est possible de calculer l'espérance et la variance de la probabilité global conditionnellement aux observations et aux hyper-paramètres

$$\mathbb{E}_{n,\theta} [P(x_s, \delta^{\text{ref}})] = \int_{\mathbb{X}_e} \mu_{n,\theta}(x_e, x_s, \delta^{\text{ref}}) f^{\text{env}}(x_e) dx_e \quad (\text{C.8a})$$

$$\text{var}_{n,\theta} [P(x_s, \delta^{\text{ref}})] = \int_{\mathbb{X}_e^2} \kappa_{n,\theta}((x_e, x_s, \delta^{\text{ref}}), (x'_e, x_s, \delta^{\text{ref}})) f^{\text{env}}(x_e) dx_e f^{\text{env}}(x'_e) dx'_e \quad (\text{C.8b})$$

où $\text{var}_{n,\theta}[A] = \text{cov}_{n,\theta}[A, A]$. Comme la variance nécessite une double intégrale selon \mathbb{X}_e ainsi que les covariances a posteriori, on considère également la mesure d'incertitude

$$H_{n,\theta} [P(x_s, \delta^{\text{ref}})] = \int_{\mathbb{X}_e} v_{n,\theta}(x_e, x_s, \delta^{\text{ref}}) f^{\text{env}}(x_e) dx_e, \quad (\text{C.9})$$

qui ne nécessite qu'une intégrale simple de la variance et un calcul des variances a posteriori. Alors $\text{var}_{n,\theta} [P(x_s, \delta^{\text{ref}})] \leq H_{n,\theta} [P(x_s, \delta^{\text{ref}})]$.

En utilisant les formules des lois totales, il est possible de calculer l'espérance, la variance et la mesure d'incertitude conditionnellement aux observations

$$\mathbb{E}_n [P(x_s, \delta^{\text{ref}})] = \int_{\Theta} \mathbb{E}_{n,\theta} [P(x_s, \delta^{\text{ref}})] p(\theta|\chi_n) d\theta \quad (\text{C.10a})$$

$$\begin{aligned} \text{var}_n [P(x_s, \delta^{\text{ref}})] &= \int_{\Theta} \text{var}_{n,\theta} [P(x_s, \delta^{\text{ref}})] p(\theta|\chi_n) d\theta \quad (\text{C.10b}) \\ &+ \int_{\Theta} \left(\mathbb{E}_{n,\theta} [P(x_s, \delta^{\text{ref}})] - \mathbb{E}_n [P(x_s, \delta^{\text{ref}})] \right)^2 p(\theta|\chi_n) d\theta \end{aligned}$$

$$\begin{aligned} H_n [P(x_s, \delta^{\text{ref}})] &= \int_{\Theta} H_{n,\theta} [P(x_s, \delta^{\text{ref}})] p(\theta|\chi_n) d\theta \quad (\text{C.10c}) \\ &+ \int_{\Theta} \left(\mathbb{E}_{n,\theta} [P(x_s, \delta^{\text{ref}})] - \mathbb{E}_n [P(x_s, \delta^{\text{ref}})] \right)^2 p(\theta|\chi_n) d\theta \end{aligned}$$

où Θ est l'espace des hyper-paramètres, $\mathbb{E}_n[A] = \mathbb{E}[A|\chi_n]$ et $\text{var}_n[A] = \mathbb{E}_n[A^2] - \mathbb{E}_n[A]^2$. On a $\text{var}_n [P(x_s, \delta^{\text{ref}})] \leq H_n [P(x_s, \delta^{\text{ref}})]$.

Les différentes intégrales sont approchées par des approximations de Monte-Carlo. Ainsi, l'estimateur de $P(x_s, \delta^{\text{ref}})$ est

$$\hat{P}(x_s, \delta^{\text{ref}}) = \frac{1}{pq} \sum_{j=1}^p \sum_{l=1}^q \mu_{n,\theta_j} (x_e^{(l)}, x_s, \delta^{\text{ref}}), \quad (\text{C.11})$$

et l'incertitude de l'estimateur est

$$\begin{aligned}
u^2 \left(\widehat{P}(x_s, \delta^{\text{ref}}) \right) &= \left(1 + \frac{1}{q} \right) \frac{1}{p} \sum_{j=1}^p \frac{2}{q(q-1)} \sum_{1 \leq l < l' \leq q} \kappa_{n, \theta_j} \left((x_e^{(l)}, x_s, \delta^{\text{ref}}), (x_e^{(l')}, x_s, \delta^{\text{ref}}) \right) \\
&+ \left(1 + \frac{1}{p} \right) \frac{1}{p-1} \sum_{j=1}^p \left(\left\{ \frac{1}{q} \sum_{l=1}^q \mu_{n, \theta_j} \left(x_e^{(l)}, x_s, \delta^{\text{ref}} \right) \right\} - \widehat{P}(x_s, \delta^{\text{ref}}) \right)^2.
\end{aligned} \tag{C.12}$$

Les termes en $1 + 1/q$ et $1 + 1/p$ correspondent aux variances de bases plus les variances qui s'ajoutent à cause des approximations par Monte-Carlo. L'incertitude de l'estimateur peut être majorée par

$$\begin{aligned}
u^2 \left(\widehat{P}(x_s, \delta^{\text{ref}}) \right) &\leq \left(1 + \frac{1}{q} \right) \frac{1}{p} \sum_{j=1}^p \frac{1}{q} \sum_{l=1}^q v_{n, \theta_j} \left(x_e^{(l)}, x_s, \delta^{\text{ref}} \right) \\
&+ \left(1 + \frac{1}{p} \right) \frac{1}{p-1} \sum_{j=1}^p \left(\left\{ \frac{1}{q} \sum_{l=1}^q \mu_{n, \theta_j} \left(x_e^{(l)}, x_s, \delta^{\text{ref}} \right) \right\} - \widehat{P}(x_s, \delta^{\text{ref}}) \right)^2.
\end{aligned} \tag{C.13}$$

Enfin, dans le cas où l'incertitude sur les hyper-paramètres est négligeable, la valeur des hyper-paramètres est fixée à leur Maximum A Posteriori (MAP). Dans ce cas, l'estimateur devient

$$\widehat{P}(x_s, \delta^{\text{ref}}) = \frac{1}{q} \sum_{l=1}^q \mu_{n, \theta^{\text{MAP}}} \left(x_e^{(l)}, x_s, \delta^{\text{ref}} \right), \tag{C.14}$$

et l'incertitude de l'estimateur

$$u^2 \left(\widehat{P}(x_s, \delta^{\text{ref}}) \right) = \left(1 + \frac{1}{q} \right) \frac{2}{q(q-1)} \sum_{1 \leq l < l' \leq q} \kappa_{n, \theta^{\text{MAP}}} \left((x_e^{(l)}, x_s, \delta^{\text{ref}}), (x_e^{(l')}, x_s, \delta^{\text{ref}}) \right) \tag{C.15a}$$

$$\leq \left(1 + \frac{1}{q} \right) \frac{1}{q} \sum_{l=1}^q v_{n, \theta^{\text{MAP}}} \left(x_e^{(l)}, x_s, \delta^{\text{ref}} \right). \tag{C.15b}$$

C.3 Échantillonnage selon la loi a posteriori

Dans le cas où le calcul de la moyenne et de la variance de $P(x_s, \delta^{\text{ref}})$ est insuffisant, en particulier si on souhaite obtenir les quantiles à posteriori voire la loi a posteriori toute

entière, il n'y a plus de formule explicite. Dans ce cas, une solution consiste à échantillonner aléatoirement la loi a posteriori de la probabilité de défaillance. L'algorithme 3 est l'algorithme d'échantillonnage par défaut. Notons toutefois que cet algorithme est extrêmement lent (deux à trois minutes par scénario pour $p = 100$ hyper-paramètres et $q = 400$ entrées environnementales). La limitation vient du tirage de trajectoires à posteriori (Ligne 6) qui nécessite une décomposition de Cholesky coûteuse en temps de calcul. Par contre, le nombre de trajectoires tirées n'a que peu d'impact sur le temps total de l'algorithme.

Algorithme 3 Échantillonnage de $P(x_s, \delta^{\text{ref}})$ selon sa loi a posteriori

Require: — $m_{n,\theta}$ et $k_{n,\theta}$ les fonctions de moyennes et de covariance a posteriori de ξ conditionnellement aux hyper-paramètres θ ;
 — λ_θ la variance de l'effet stochastique conditionnellement aux hyper-paramètres ;
 — z^{crit} le seuil critique ;
 — x_s, δ^{ref} le scénario d'intérêt ;
 — $(\theta_j)_{1 \leq j \leq p}$ p hyper-paramètres indépendants et identiquement distribués selon leurs lois a posteriori $p(\theta|\chi_n)$;
 — f^{env} la distribution des entrées environnementales ;
 — q, r deux nombres entiers, représentant respectivement le nombre d'entrées environnementales et le nombre de trajectoires par hyper-paramètres

- 1: **for** $j \in \llbracket 1; p \rrbracket$ **do**
- 2: {Pour chaque hyper-paramètre. . .}
- 3: $(x_e^{(l)})_{1 \leq l \leq q} \sim_{\text{iid}} f^{\text{env}}$, tirez aléatoirement q points environnementales de manières indépendantes
- 4: $M^{(j)} \leftarrow (m_{n,\theta_j}(x_e^{(l)}, x_s, \delta^{\text{ref}}))_{1 \leq l \leq q}$, $M^{(j)}$ est le vecteur de longueur q des moyennes a posteriori aux points d'intérêt conditionnellement à θ_j
- 5: $K^{(j)} \leftarrow (k_{n,\theta_j}((x_e^{(l)}, x_s, \delta^{\text{ref}}), (x_e^{(l')}, x_s, \delta^{\text{ref}})))_{1 \leq l, l' \leq q}$, $K^{(j)}$ est la matrice de taille q des covariances a posteriori aux points d'intérêt conditionnellement à θ_j
- 6: $(\xi_{lk}^{(j)})_{1 \leq l \leq q, 1 \leq k \leq r} \sim_{\text{iid}} \mathbf{N}(M^{(j)}, K^{(j)})$, tirez aléatoirement r trajectoires a posteriori aux points d'intérêt conditionnellement à θ_j
- 7: $\forall (l, k) \in \llbracket 1; q \rrbracket \times \llbracket 1; r \rrbracket$, $p_{lk}^{(j)} \leftarrow \Phi\left(\frac{\xi_{lk}^{(j)} - z^{\text{crit}}}{\sqrt{\lambda_{\theta_j}(x_e^{(l)}, x_s, \delta^{\text{ref}})}}\right)$, calculez la probabilité de dépassement de seuil locale
- 8: $\forall k \in \llbracket 1; r \rrbracket$, $P_k^{(j)} \leftarrow \frac{1}{q} \sum_{l=1}^q p_{lk}^{(j)}$, calculez la probabilité de dépassement de seuil globale, en approchant l'intégrale sur les entrées environnementales par une approximation de Monte-Carlo. Il y a un résultat par trajectoire
- 9: **end for**
- 10: **return** $(P_k^{(j)})_{1 \leq j \leq p, 1 \leq k \leq r}$

Annexe D

Distribution a priori des hyper-paramètres

Dans cette section, nous décrivons comment les lois a priori sur les hyper-paramètres sont choisies tout au long de la thèse.

Déjà, les hyper-paramètres sont supposés a priori indépendants. Par exemple, pour un modèle non-stationnaire (section 1.2.2) la loi s'écrit :

$$\begin{aligned}
 & p\left(l_{\sigma}^0, l_{\nu}^0, \left(l_{\rho_k}^0\right)_{1 \leq k \leq d}, l_G, l_L, l_{\nu}^{\epsilon}, \left(l_{\rho_k}^{\epsilon}\right)_{1 \leq k \leq d}, l_0, \sigma_{\lambda}^2, a\right) \\
 & = p\left(l_{\sigma}^0\right) p\left(l_{\nu}^0\right) p\left(l_G\right) p\left(l_L\right) p\left(l_{\nu}^{\epsilon}\right) p\left(l_m\right) p\left(s^2\right) p\left(\varsigma^2\right) \prod_{k=1}^d p\left(l_{\rho_k}^0\right) p\left(l_{\rho_k}^{\epsilon}\right).
 \end{aligned} \tag{D.1}$$

Des corrélations a priori entre $l_{\rho_k}^0$ et $l_{\rho_k}^{\epsilon}$ pourraient être considérées, comme suggérées par le modèle de Picheny et Ginsbourger [2013]. Cependant, le besoin ne s'en est jamais senti. Toutes les lois a priori sont des lois gaussiennes mono-variables.

On suppose connue l'amplitude approximative de la sortie, noté r^{out} , ainsi que l'amplitude totale des entrées, noté b_k^{in} . Si la $k^{\text{ième}}$ variable varie entre les bornes a_k et b_k , $b_k^{\text{in}} = b_k - a_k$. Pour les amplitudes approximatives des sorties de Fire Dynamics Simulator

Sorties	Amplitude r^{out}
Flux radiatif H^f	7 kW · m ⁻²
Température T^c	200 °C
Visibilité V	30 m

TABLE 12 – Valeurs des amplitudes approximatives des différentes sorties r^{out} .

(FDS), j'utilisai les valeurs présentées tableau 12.

Distribution a priori des variances des covariance, l_σ^0 . La valeur de σ_0 est supposée valoir environ un centième de l'amplitude totale de la sortie

$$\sigma_0 \approx \frac{r^{\text{out}}}{100} \Leftrightarrow \sigma_0^2 \approx \left(\frac{r^{\text{out}}}{100}\right)^2 \Leftrightarrow l_\sigma^0 \approx 2 \log\left(\frac{r^{\text{out}}}{100}\right). \quad (\text{D.2})$$

Pour la variance a priori, comme peu d'information sur la valeur exacte de σ_0 sont disponibles, un écart-type de $\log(100)$ sur les variations logarithmiques de σ_0^2 est ajouté. Ainsi, celui-ci peut varier de huit ordres de grandeur.

$$l_\sigma^0 | r^{\text{out}} \sim \text{N}\left(2 \log\left(\frac{r^{\text{out}}}{100}\right), \log(100)^2\right). \quad (\text{D.3})$$

Distribution a priori des régularité, l_ν^0 et l_ν^ϵ . Les régularités des covariances de Matérn sont fixées. Principalement pour des raisons calculatoire, Small Toolbox for Kriging (STK) calcule plus rapidement les covariances de Matérn lorsque la régularité de celle-ci vaut $1/2$, $3/2$, $5/2$ ou $+\infty$. La régularité $5/2$ permet de considérer des trajectoires continues et dérivables une fois, ce qui est assez courant pour les simulateurs.

$$l_\nu^0 = l_\nu^\epsilon = \log\left(\frac{5}{2}\right). \quad (\text{D.4})$$

Dans le cas où on souhaite vraiment prendre en compte l'incertitude sur la régularité, on peut supposer que la régularité n'est jamais inférieur à $1/2$. Conséquemment,

$$\nu \geq \frac{1}{2} \Leftrightarrow l_\nu \geq \log\left(\frac{1}{2}\right) \Rightarrow m - 3\sigma = \log\left(\frac{1}{2}\right) \Rightarrow \sigma = \frac{m - \log\left(\frac{1}{2}\right)}{3} = \frac{\log(5)}{3}.$$

La loi a priori devient alors

$$l_\nu^0, l_\nu^\epsilon \sim \text{N}\left(\log\left(\frac{5}{2}\right), \left(\frac{\log(5)}{3}\right)^2\right)$$

Remarquons que la régularité à $+3\sigma$ a priori est de $\log(25/2)$, ce qui signifie que la régularité varie a priori plutôt entre $1/2$ et $25/2$.

Distribution a priori des portées, $l_{\rho_k}^0$ et $l_{\rho_k}^\epsilon$. Le $k^{\text{ième}}$ facteur des entrées varie d'une amplitude b_k^{in} entre sa valeur minimale et sa valeur maximale. On suppose que sur cet

intervalle, le simulateur peut changer environ une fois de sens, ce que se traduit par

$$\rho_k \approx \frac{b_k^{\text{in}}}{2} \Leftrightarrow l_{\rho_k} \approx \log\left(\frac{b_k^{\text{in}}}{2}\right).$$

Comme cette valeur est assez incertaine, un écart-type de $\log(10)$ en échelle logarithmique est ajouté, ce qui autorise des variations de quatre ordre de magnitudes, entre le cas d'un bruit (valeur de portée très faible) et le cas d'un simulateur constant (valeur de portée très grande).

$$l_{\rho_k}^0, l_{\rho_k}^\epsilon | b_k^{\text{in}} \sim \mathbf{N}\left(\log\left(\frac{b_k^{\text{in}}}{2}\right), \log(10)^2\right). \quad (\text{D.5})$$

Distribution a priori du rapport des variances, l_G . Le facteur G est le rapport entre σ_ϵ^2 et σ_0^2 . Comme ces deux coefficients doivent être du même ordre de grandeur, la valeur par défaut de G est un

$$G \approx 1 \Leftrightarrow l_G \approx 0.$$

De plus, comme ce rapport est très incertain, et qu'il s'agit de l'un des hyper-paramètres les plus important du modèle, un écart-type de huit ordre de grandeur est ajouté

$$l_G \sim \mathbf{N}\left(0, \log(100)^2\right). \quad (\text{D.6})$$

Au passage comme $l_\sigma^\epsilon = l_G + l_\sigma^0$, la loi a priori de l_σ^ϵ est une loi normale $\mathbf{N}\left(2 \log\left(\frac{r^{\text{out}}}{100}\right), \log(10000)^2\right)$.

Distribution a priori du degré de l'erreur, L . Pour des raisons analytiques, Tuo et al. [2014] recommandent les valeurs $L = 2$ et $L = 4$. La valeur quatre est alors prise comme moyenne, et un écart-type se rajoute pour essayer de prendre en compte toutes les valeurs intéressantes possibles. En ajoutant une écart-type de $\log(3)$ sur $\log(L)$, l'intervalle a priori à deux sigmas est $[4/9, 36]$. Une valeur en deçà de $4/9$ ou au-delà de 36 sera de toute façon inestimable avec le peu de niveaux disponible (décroissance trop rapide ou trop lente). La loi a priori de L s'écrit

$$l_L \sim \mathbf{N}\left(\log(4), \log(3)^2\right) \quad (\text{D.7})$$

Hyper-paramètres de la loi a priori de la variance de la sortie, l_0 , σ_λ^2 et a . Tout comme pour les paramètres de variance de la covariance, que la variance de la sortie vaut a priori environ $r^{\text{out}}/100$ avec un écart-type de $\log(100)$ en échelle logarithmique. Dans la variance, 99 % de celle-ci est supposée venir de la variance sur la moyenne globale,

et le 1% restant de l'effet de pépite sur les variance. Ceci donne $a = 99\%/1\% = 99$.
Donc,

$$\left. \begin{array}{l} l_0 = 2 \log \left(\frac{r^{\text{out}}}{100} \right) \\ a = 99 \\ \sigma_\lambda^2 = \frac{\log(100)^2}{100} \end{array} \right\} \Rightarrow l_\lambda(\delta) | r^{\text{out}} \sim \text{GP} \left(2 \log \left(\frac{r^{\text{out}}}{100} \right), \frac{\log(100)^2}{100} (\mathbb{1}_{\delta=\delta'} + 99) \right). \quad (\text{D.8})$$

Ces hyper-paramètres sont supposés connus, et leur incertitudes ne sont pas intégrés dans l'incertitude finale.

Annexe E

Figures complémentaires sur la validation de la méthode sur FDS

E.1 Comparaison des prédictions et des observations

Les figures 83 à 85 présentent les comparaisons entre les prédictions par les modèles bayésiens multi-fidèles associés aux sorties de FDS et des vraies observations. Les modèles sont construits à partir des plans numérotés 2 à 10 dans la section 3.4. Ces figures viennent en complément de la figure 72.

E.2 Évaluation des mesures d'incertitudes en fonction du coût

Les figures 86 et 87 montrent les évaluations de différentes mesures d'incertitudes en fonction du coût du plan d'expériences. Chaque figure correspond à un plan d'expériences. Ces figures viennent en complément des figures 75 et 76.

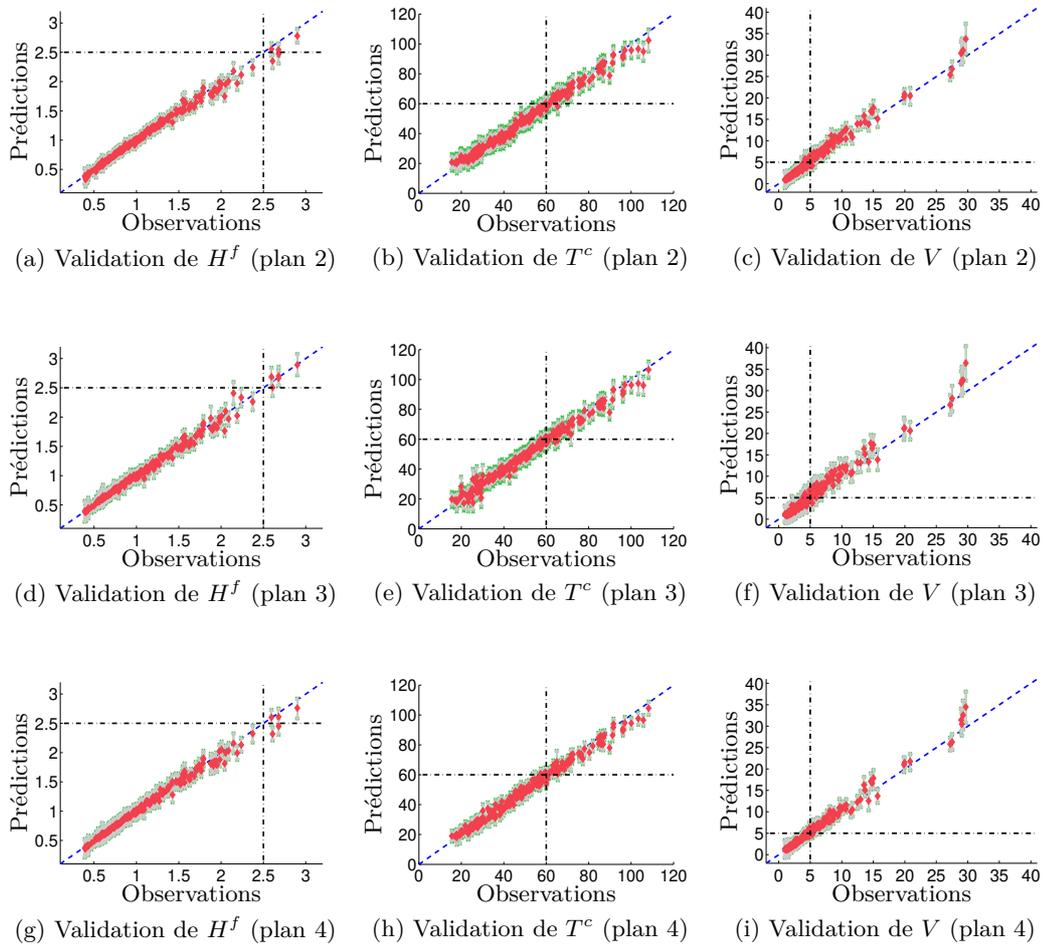


FIGURE 83 – Comparaison entre les observations et les prédictions obtenues par les plans d’expérience numéro 2, 3 et 4, à 20 cm. Les points sont les prédictions par médiane à posteriori, les barres correspondent aux intervalles de crédibilité à 95 %, en gris à l’intérieur sur l’estimation de la moyenne en vert à l’extérieur sur les futures observations. Les traits pointillés verticaux et horizontaux correspondent aux seuils critiques. Le trait pointillé oblique bleu est la première bissectrice.

E.2. ÉVALUATION DES MESURES D'INCERTITUDES EN FONCTION DU COÛT²¹⁵

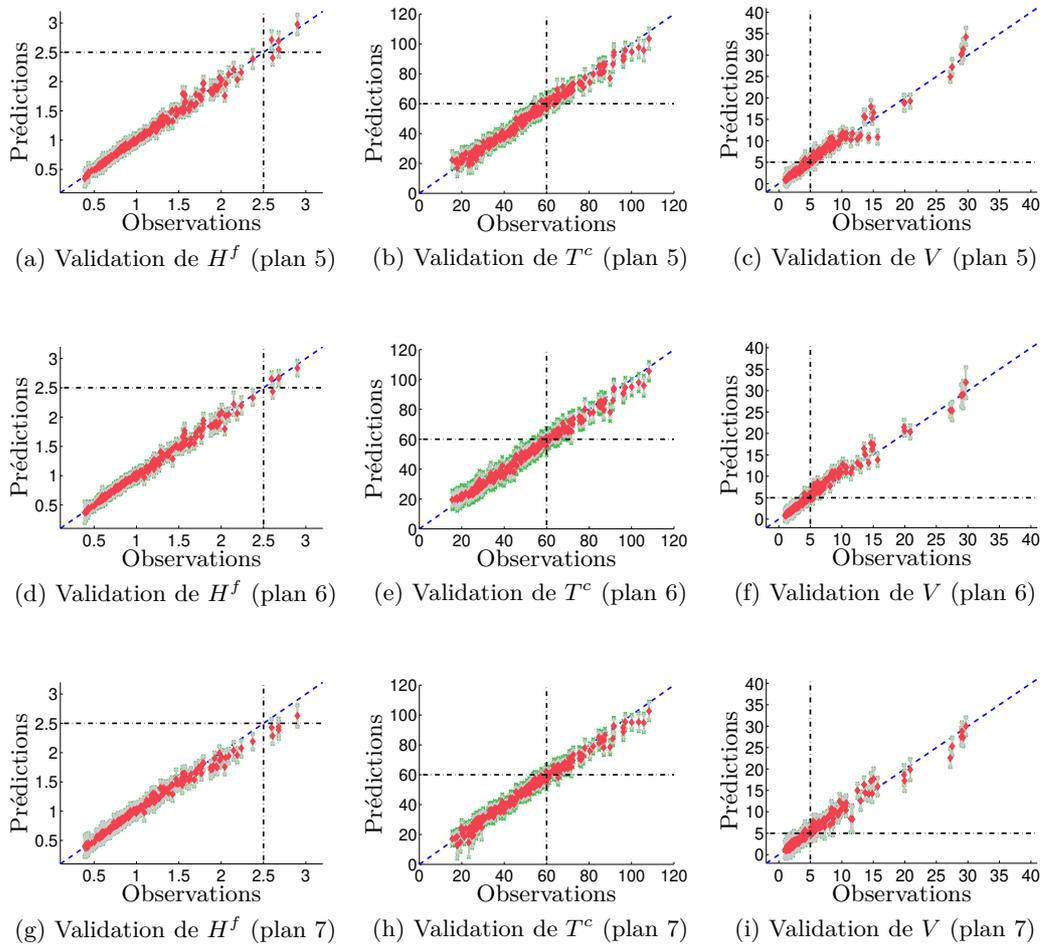


FIGURE 84 – Comparaison entre les observations et les prédictions obtenues par les plans d'expérience numéro 5, 6 et 7, à 20 cm. Les points sont les prédictions par médiane à posteriori, les barres correspondent aux intervalles de crédibilité à 95 %, en gris à l'intérieur sur l'estimation de la moyenne en vert à l'extérieur sur les futures observations. Les traits pointillés verticaux et horizontaux correspondent aux seuils critiques. Le trait pointillé oblique bleu est la première bissectrice.

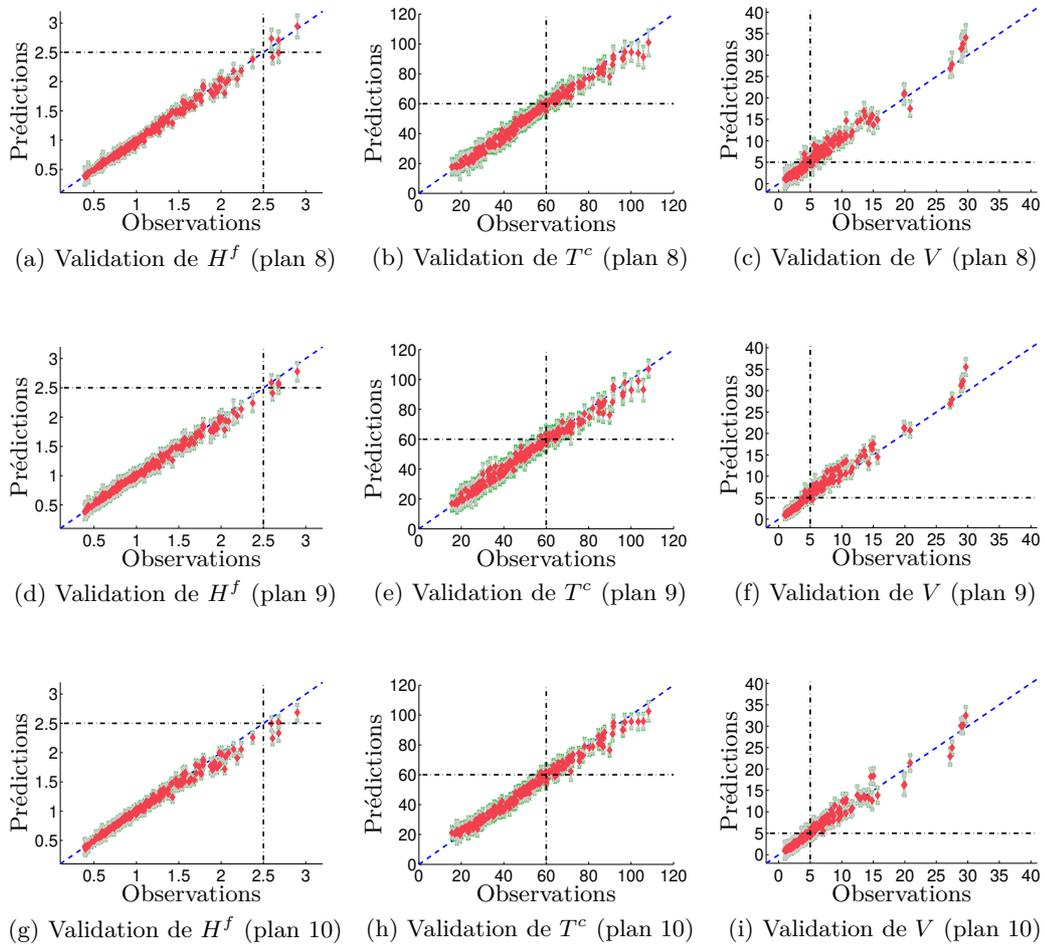


FIGURE 85 – Comparaison entre les observations et les prédictions obtenues par les plans d'expérience numéro 8, 9 et 10, à 20 cm. Les points sont les prédictions par médiane à posteriori, les barres correspondent aux intervalles de crédibilité à 95 %, en gris à l'intérieur sur l'estimation de la moyenne en vert à l'extérieur sur les futures observations. Les traits pointillés verticaux et horizontaux correspondent aux seuils critiques. Le trait pointillé oblique bleu est la première bissectrice.

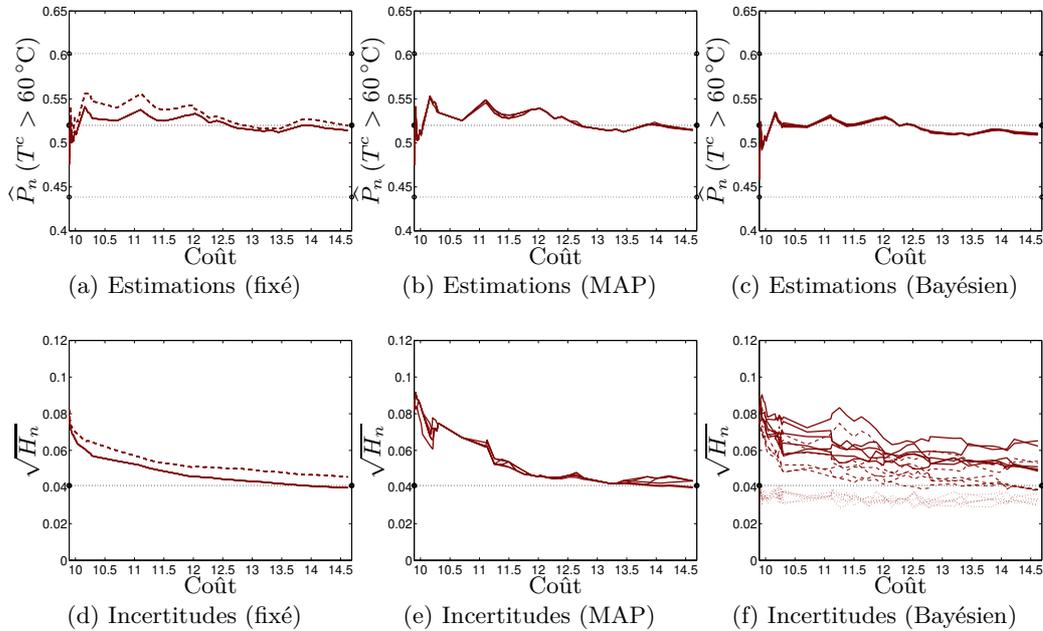


FIGURE 86 – Comparaison des méthodes d’estimation de la probabilité en fonction du coût du plan d’expériences sur le plan numéro 1. La ligne du haut représente les estimations de la probabilité, la ligne du bas les mesures d’incertitude, une borne majorante de l’écart-type. (a, d) Estimations dans le cas où les hyper-paramètres sont fixés. Les lignes pleines correspondent au cas où les hyper-paramètres sont fixés à leur valeurs finales (MAP avec toutes les observations), les lignes discontinues au cas où les hyper-paramètres sont fixés à leur valeurs initiales (MAP du plan initial). (b, e) Estimations dans le cas où les hyper-paramètres sont estimés à leur valeur MAP à chaque nouveau point d’observation et substitués dans le modèle. Il y a six courbes par figure, car la méthode est répétée six fois. (c, f) Estimations par des méthodes complètement bayésiennes. Sur la figure f, les traits pleins sont les mesures d’incertitude totales, les traits discontinus sont la racine carré de la moyenne de la variance, et les traits pointillés sont les écart-types des moyennes. Il y a six jeux de courbes, car la méthode est répétée six fois.

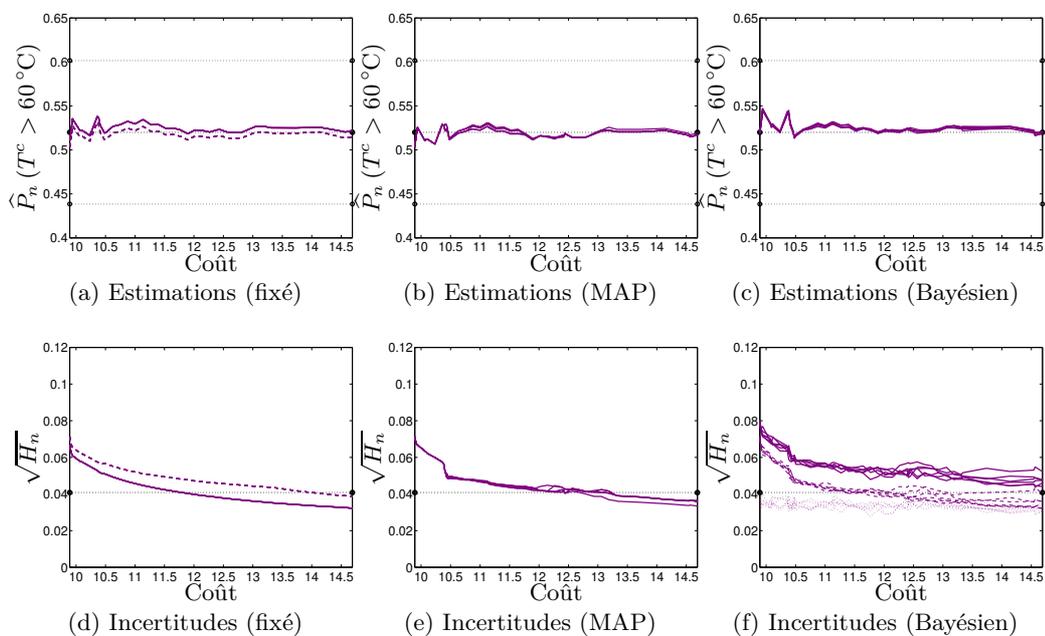


FIGURE 87 – Comparaison des méthodes d’estimation de la probabilité en fonction du coût du plan d’expériences sur le plan numéro 2. La ligne du haut représente les estimations de la probabilité, la ligne du bas les mesures d’incertitude, une borne majorante de l’écart-type. (a, d) Estimations dans le cas où les hyper-paramètres sont fixés. Les lignes pleines correspondent au cas où les hyper-paramètres sont fixés à leur valeurs finales (MAP avec toutes les observations), les lignes discontinues au cas où les hyper-paramètres sont fixés à leur valeurs initiales (MAP du plan initial). (b, e) Estimations dans le cas où les hyper-paramètres sont estimés à leur valeur MAP à chaque nouveau point d’observation et substitués dans le modèle. Il y a six courbes par figure, car la méthode est répétée six fois. (c, f) Estimations par des méthodes complètement bayésiennes. Sur la figure f, les traits pleins sont les mesures d’incertitude totales, les traits discontinus sont la racine carré de la moyenne de la variance, et les traits pointillés sont les écart-types des moyennes. Il y a six jeux de courbes, car la méthode est répétée six fois.

Bibliographie

- Natalia M. Alexandrov, Robert Michael Lewis, Clyde R. Gumbert, Larry L. Green, et Perry A. Newman. Optimization with variable-fidelity models applied to wing design. Technical report, Defense Technical Information Center, Hampton VA, december 1999.
- Ronald L. Alpert. *Ceiling Jet Flows*, pages 429–454. Springer, New York, NY, 2016. doi : https://doi.org/10.1007/978-1-4939-2565-0_14.
- Ioannis Andrianakis, Ian R. Vernon, Nicky McCreesh, Trevelyan J. McKinley, Jeremy E. Oakley, Rebecca N. Nsubuga, Michael Goldstein, et Richard G. White. Bayesian history matching of complex infectious disease models using emulation : a tutorial and a case study on hiv in uganda. *PLoS computational biology*, 11(1) :e1003968, 2015. doi : <https://doi.org/10.1371/journal.pcbi.1003968>.
- Bruce Ankenman, Barry L. Nelson, et Jeremy Staum. Stochastic kriging for simulation metamodeling. *Operations Research*, 58(2) :371–382, 2010. doi : <http://dx.doi.org/10.1287/opre.1090.0754>.
- Lukas Arnold et Simone Arnold. Multi-factorial analysis of the smoke layer height. *SFPE Europe*, 8 :4, 2017. <http://www.sfpe.org/?page=Issue8Feature4>.
- Siu-Kui Au et James L. Beck. Estimation of small failure probabilities in high dimensions by subset simulation. *Probabilistic Engineering Mechanics*, 16(4) :263–277, 2001. doi : [http://dx.doi.org/10.1016/S0266-8920\(01\)00019-4](http://dx.doi.org/10.1016/S0266-8920(01)00019-4).
- Siu Kui Au, Zhi-Hua Wang, et Siu-Ming Lo. Compartment fire risk analysis by advanced Monte Carlo simulation. *Engineering Structures*, 29(9) :2381–2390, 2007. doi : <http://dx.doi.org/10.1016/j.engstruct.2006.11.024>.
- Dario Azzimonti, Julien Bect, Clément Chevalier, et David Ginsbourger. Quantifying uncertainties on excursion sets under a Gaussian random field prior. *SIAM/ASA Journal on Uncertainty Quantification*, 4(1) :850–874, 2016a.

- Dario Azzimonti, David Ginsbourger, Clément Chevalier, Julien Bect, et Yann Richet. Adaptive design of experiments for conservative estimation of excursion sets. In *2016 SIAM Conference on Uncertainty Quantification*, Lausanne, Switzerland, April 2016b.
- Julien Bect, David Ginsbourger, Ling Li, Victor Picheny, et Emmanuel Vazquez. Sequential design of computer experiments for the estimation of a probability of failure. *Statistics and Computing*, 22(3) :773–793, 2012. doi : <http://dx.doi.org/10.1007/s11222-011-9241-4>.
- Julien Bect, François Bachoc, et David Ginsbourger. A supermartingale approach to Gaussian process based sequential design of experiments. working paper or preprint, July 2017.
- Romain Benassi, Julien Bect, et Emmanuel Vazquez. Robust Gaussian process-based global optimization using a fully Bayesian expected improvement criterion. In Carlos A. Coello Coello, éditeur, *Learning and Intelligent Optimization*, pages 176–190, Berlin, Heidelberg, 2011. Springer, Springer Berlin Heidelberg.
- Alexis Boukouvalas. *Emulation of random output simulators*. PhD thesis, Aston University, 2010.
- Alexis Boukouvalas et Dan Cornford. Learning heteroscedastic Gaussian processes for complex datasets. Technical report, Neural Computing Research Group, Aston University, april 2009.
- Christopher James Brooks, A. I. J. Forrester, A. J. Keane, et S. Shahpar. Multi-fidelity design optimisation of a transonic compressor rotor. In *9th European Conf. Turbomachinery Fluid Dynamics and Thermodynamics*, Istanbul, Turkey, march 2011.
- Pierre Carlotti et Damien Lamalle. Sensitivity to boundary conditions for simulation of fire plume in enclosure. In Association Française de la Mécanique (AFM), éditeur, *21^{emes} congrés français de mécanique (CFM 2013)*, Bordeaux, France, août 2013.
- Ray-Bing Chen, Ying-Chao Hung, Weichung Wang, et Sung-Wei Yen. Contour estimation via two fidelity computer simulators under limited resources. *Computational Statistics*, 28(4) :1813–1834, 2013. doi : <http://dx.doi.org/10.1007/s00180-012-0380-7>.
- Clément Chevalier, Julien Bect, David Ginsbourger, Emmanuel Vazquez, Victor Picheny, et Yann Richet. Fast parallel kriging-based stepwise uncertainty reduction with application to the identification of an excursion set. *Technometrics*, 56(4) :455–465, 2014. doi : <http://dx.doi.org/10.1080/00401706.2013.860918>.

- Nicolas Chopin. A sequential particle filter method for static models. *Biometrika*, 89 (3) :539–552, 2002. doi : <http://dx.doi.org/10.1093/biomet/89.3.539>.
- David A. Cohn, Zoubin Ghahramani, et Michael I. Jordan. Active learning with statistical models. *Journal of artificial intelligence research*, 4 :129—145, 1996.
- Nicolas Courrier, Pierre-Alain Boucard, et Bruno Soulier. The use of partially converged simulations in building surrogate models. *Advances in Engineering Software*, 67 :186–197, 2014. doi : <http://dx.doi.org/10.1016/j.advengsoft.2013.09.008>.
- Dennis D. Cox et Susan John. Sdo : A statistical method for global optimization. In *Multidisciplinary Design Optimization : State-of-the-Art*, pages 315–329, 1997.
- Carla Currin, Toby Mitchell, Max Morris, et Don Ylvisaker. Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments. *Journal of the American Statistical Association*, 86(416) :953–963, 1991. doi : <http://dx.doi.org/10.1080/01621459.1991.10475138>.
- Matthias De Lozzo. Substitution de modèle et approche multifidélité en expérimentation numérique. *Journal de la Société Française de Statistique*, 156(3) :21–55, 2015.
- Séverine Demeyer, Nicolas Fischer, et Damien Marquis. Surrogate model based sequential sampling estimation of conformance probability for computationally expensive systems : application to fire safety science. *Journal de la Société Française de Statistique*, 158(1) :111–138, 2017.
- Thomas Desautels, Andreas Krause, et Joel W. Burdick. Parallelizing exploration-exploitation tradeoffs in Gaussian process bandit optimization. *The Journal of Machine Learning Research*, 15(1) :4053–4103, 2014.
- Dougal Drysdale. *An introduction to fire dynamics*. John Wiley & Sons, 2011.
- Vincent Dubourg, Bruno Sudret, et Jean-Marc Bourinet. Reliability-based design optimization using kriging surrogates and subset simulation. *Structural and Multidisciplinary Optimization*, 44(5) :673–690, 2011. doi : <http://dx.doi.org/10.1007/s00158-011-0653-8>.
- Darrell Duffie et Peter Glynn. Efficient Monte-Carlo simulation of security prices. *The Annals of Applied Probability*, 5(4) :897–905, 1995.

- Khairy Elsayed. Optimization of the cyclone separator geometry for minimum pressure drop using co-kriging. *Powder Technology*, 269 :409–424, 2015. doi : <http://dx.doi.org/10.1016/j.powtec.2014.09.038>.
- Michael T. M. Emmerich, Kyriakos C. Giannakoglou, et Boris Naujoks. Single- and multiobjective evolutionary optimization assisted by gaussian random field metamodels. *IEEE Transactions on Evolutionary Computation*, 10(4) :421–439, August 2006.
- Ahmed Aziz Ezzat, Arash Pourhabib, et Yu Ding. Sequential design for functional calibration of computer models. *Technometrics*, 0(just-accepted) :0–0, 2017. doi : <https://doi.org/10.1080/00401706.2017.1377638>.
- Kai-Tai Fang, Dennis K.J. Lin, Peter Winker, et Yong Zhang. Uniform design : theory and application. *Technometrics*, 42(3) :237–248, 2000. doi : [10.1080/00401706.2000.10486045](https://doi.org/10.1080/00401706.2000.10486045).
- Francois Fleuret et Donald Geman. Graded learning for object detection. In *Proceedings of the IEEE workshop on Statistical and Computational Theories of Vision*, volume 2, pages 544–549. Citeseer, 1999.
- Alexander I. J. Forrester et Andy J. Keane. Recent advances in surrogate-based optimization. *Progress in Aerospace Sciences*, 45(1–3) :50–79, 2009. doi : <http://dx.doi.org/10.1016/j.paerosci.2008.11.001>.
- Alexander I. J. Forrester, András Sóbester, et Andy J. Keane. Multi-fidelity optimization via surrogate modelling. *Proceedings of the Royal Society of London A : Mathematical, Physical and Engineering Sciences*, 463(2088) :3251–3269, 2007. doi : <http://dx.doi.org/10.1098/rspa.2007.1900>.
- Alexander I.J. Forrester, András Sóbester, et Andy J. Keane. *Engineering Design via Surrogate Modelling : A Practical Guide*. Wiley, 2008.
- Peter Frazier, Warren Powell, et Savas Dayanik. The knowledge-gradient policy for correlated normal beliefs. *INFORMS Journal on Computing*, 21(4) :599–613, 2009. doi : <https://doi.org/10.1287/ijoc.1080.0314>.
- Vikram V. Garg et Roy H. Stogner. Hierarchical latin hypercube sampling. *Journal of the American Statistical Association*, 112(518) :673–682, 2017. doi : <https://doi.org/10.1080/01621459.2016.1158717>.

- Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, et Donald B. Rubin. *Bayesian data analysis*. Chapman & Hall/CRC, third edition, 2013.
- Michael B. Giles. Multilevel Monte Carlo path simulation. *Operations Research*, 56(3) : 607–617, 2008. doi : <http://dx.doi.org/10.1287/opre.1070.0496>.
- David Ginsbourger, Janis Janusevskis, et Rodolphe Le Riche. Dealing with asynchronicity in parallel Gaussian process based global optimization. In *4th International Conference of the ERCIM WG on computing & statistics (ERCIM'11)*, 2010a.
- David Ginsbourger, Rodolphe Le Riche, et Laurent Carraro. *Kriging Is Well-Suited to Parallelize Optimization*, pages 131–162. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010b. doi : https://doi.org/10.1007/978-3-642-10701-6_6.
- Ramunas Girdziusas, Rodolphe Le Riche, Fabien Viale, et David Ginsbourger. Parallel budgeted optimization applied to the design of an air duct. Technical report, september 2012.
- María Giselle Fernández-Godino, Chanyoung Park, Nam-Ho Kim, et Raphael T. Haftka. Review of multi-fidelity models. *ArXiv e-prints*, september 2016.
- Tilmann Gneiting. Nonseparable, stationary covariance functions for space–time data. *Journal of the American Statistical Association*, 97(458) :590–600, 2002. doi : <http://dx.doi.org/10.1198/016214502760047113>.
- Tilmann Gneiting et Adrian E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477) :359–378, 2007. doi : <http://dx.doi.org/10.1198/016214506000001437>.
- Tilmann Gneiting, William Kleiber, et Martin Schlather. Matérn cross-covariance functions for multivariate random fields. *Journal of the American Statistical Association*, 105(491) :1167–1177, 2010. doi : <http://dx.doi.org/10.1198/jasa.2010.tm09420>.
- Joslin Goh, Derek Bingham, James Paul Holloway, Michael J. Grosskopf, Carolyn C. Kuranz, et Erica Rutter. Prediction and computer model calibration using outputs from multifidelity simulators. *Technometrics*, 55(4) :501–512, 2013. doi : <http://dx.doi.org/10.1080/00401706.2013.838910>.
- Javier González, Zhenwen Dai, Philipp Hennig, et Neil Lawrence. Batch Bayesian optimization via local penalization. In Arthur Gretton et Christian C. Robert, éditeurs,

- Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*, volume 51 of *Proceedings of Machine Learning Research*, pages 648–657, Cadiz, Spain, May 2016. PMLR.
- Robert B. Gramacy et Herbert K. H. Lee. Bayesian treed Gaussian process models with an application to computer modeling. *Journal of the American Statistical Association*, 103(483) :1119–1130, 2008. doi : <http://dx.doi.org/10.1198/016214508000000689>.
- Eric Guillaume. Effets du feu sur les personnes. Bibliographic Synthesis G020284/C672X01/CEMATE/1, Laboratoire National de métrologie et d'Essais, Trappes, France, july 2006. fre.
- Bing Guo, Xue-Ping Chen, et Min-Qian Liu. Construction of latin hypercube designs with nested and sliced structures. *Statistical Papers*, pages 1–14, October 2017. doi : <https://doi.org/10.1007/s00362-017-0959-8>.
- Heikki Haario, Eero Saksman, et Johanna Tamminen. An adaptive Metropolis algorithm. *Bernoulli*, 7(2) :223–242, 2001.
- Maya Hage Hassan, Ghislain Remy, Guillaume Krebs, Claude Marchand, Luc Dupré, et Professor Dr Guillaume Crevecoeur. Radial output space mapping for electromechanical systems design. *COMPEL - The International Journal for Computation and Mathematics in Electrical and Electronic Engineering*, 33(3) :965–975, 2014. doi : <http://dx.doi.org/10.1108/COMPEL-05-2013-0192>.
- Xu He et Peter Z. G. Qian. Nested orthogonal array-based latin hypercube designs. *Biometrika*, 98(3) :721–731, 2011. doi : <http://dx.doi.org/10.1093/biomet/asr028>.
- Xu He, Rui Tuo, et CF Jeff Wu. Optimization of multi-fidelity computer experiments via the EQIE criterion. *Technometrics*, 59(1) :58–68, 2017. doi : 10.1080/00401706.2016.1142902.
- Stefan Heinrich. Multilevel Monte Carlo methods. In Svetozar Margenov, Jerzy Waśniewski, et Plamen Yalamov, éditeurs, *Large-Scale Scientific Computing : 3rd International Conference*, pages 58–67, Berlin, Heidelberg, june 2001. Springer. doi : http://dx.doi.org/10.1007/3-540-45346-6_5.
- Michael T. Horrell et Michael L. Stein. Half-spectral space-time covariance models. *Spatial Statistics*, 19(Supplement C) :90–100, 2017. doi : <https://doi.org/10.1016/j.spasta.2016.12.002>.

- Simo Hostikka, Timo Korhonen, et Olavi Keski-Rahkonen. Two-model monte carlo simulation of fire scenarios. *Fire Safety Science*, 8 :1241–1252, 2005. doi : <http://dx.doi.org/10.3801/IAFSS.FSS.8-1241>.
- Deng Huang, Theodore T. Allen, William I. Notz, et R. A. Miller. Sequential kriging optimization using multiple-fidelity evaluations. *Structural and Multidisciplinary Optimization*, 32(5) :369–382, 2006. doi : <http://dx.doi.org/10.1007/s00158-005-0587-0>.
- Arnulf Jentzen et Peter E. Kloeden. Overcoming the order barrier in the numerical approximation of stochastic partial differential equations with additive space–time noise. *Proceedings of the Royal Society of London A : Mathematical, Physical and Engineering Sciences*, 465(2102) :649–667, 2009. doi : <http://dx.doi.org/10.1098/rspa.2008.0325>.
- Mark E. Johnson, Leslie M. Moore, et Donald Ylvisaker. Minimax and maximin distance designs. *Journal of Statistical Planning and Inference*, 26(2) :131–148, 1990. doi : [https://doi.org/10.1016/0378-3758\(90\)90122-B](https://doi.org/10.1016/0378-3758(90)90122-B).
- Donald R. Jones, Matthias Schonlau, et William J. Welch. Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, 13(4) :455–492, december 1998. doi : <https://doi.org/10.1023/A:1008306431147>.
- Kirthevasan Kandasamy, Gautam Dasarathy, Junier B. Oliva, Jeff Schneider, et Barnabas Poczos. Multi-fidelity Gaussian process bandit optimisation. *ArXiv e-prints*, march 2016.
- Marc C. Kennedy et Anthony O’Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)*, 63(3) :425–464, 2001. doi : <http://dx.doi.org/10.1111/1467-9868.00294>.
- Marc C. Kennedy et Anthony O’Hagan. Predicting the output from a complex computer code when fast approximations are available. *Biometrika*, 87(1) :1–13, 2000. doi : <http://dx.doi.org/10.1093/biomet/87.1.1>.
- Robert P. Kennedy, C. Allin Cornell, R. D. Campbell, S. Kaplan, et H. F. Perla. Probabilistic seismic safety study of an existing nuclear power plant. *Nuclear Engineering and Design*, 59(2) :315–338, 1980. doi : [https://doi.org/10.1016/0029-5493\(80\)90203-4](https://doi.org/10.1016/0029-5493(80)90203-4).
- Kristian Kersting, Christian Plagemann, Patrick Pfaff, et Wolfram Burgard. Most likely heteroscedastic Gaussian process regression. In *Proceedings of the 24th international*

- conference on Machine learning*, ICML '07, pages 393–400, New York, NY, USA, 2007. ACM, ACM. doi : 10.1145/1273496.1273546.
- Hong Seok Kim, Muammer Koc, et Jun Ni. A hybrid multi-fidelity approach to the optimal design of warm forming processes using a knowledge-based artificial neural network. *International Journal of Machine Tools and Manufacture*, 47(2) :211–222, 2007. doi : <https://doi.org/10.1016/j.ijmachtools.2006.04.007>.
- Depeng Kong, Nils Johansson, Patrick van Hees, Shouxiang Lu, et Siuming Lo. A Monte Carlo analysis of the effect of heat release rate uncertainty on available safe egress time. *Journal of Fire Protection Engineering*, 23(1) :5–29, 2012. doi : <http://dx.doi.org/10.1177/1042391512452676>.
- Xiangshun Kong, Mingyao Ai, et Kwok Leung Tsui. Design for sequential follow-up experiments in computer emulations. *Technometrics*, 0(0) :1–9, 2016. doi : <https://doi.org/10.1080/00401706.2016.1258010>.
- Slawomir Koziel, David Echeverría Ciaurri, et Leifur Leifsson. *Surrogate-based methods*, pages 33–59. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011. doi : http://dx.doi.org/10.1007/978-3-642-20859-1_3.
- Slawomir Koziel, Leifur Leifsson, Ivo Couckuyt, et Tom Dhaene. Robust variable-fidelity optimization of microwave filters using co-kriging and trust regions. *Microwave and Optical Technology Letters*, 55(4) :765–769, 2013. doi : <http://dx.doi.org/10.1002/mop.27447>.
- Yuichi Kuya, Kenji Takeda, Xin Zhang, et Alexander I. J. Forrester. Multifidelity surrogate modeling of experimental and computational aerodynamic data sets. *AIAA journal*, 49(2) :289–298, february 2011. doi : <https://doi.org/10.2514/1.J050384>.
- Tatiana Labopin-Richard et Victor Picheny. Sequential design of experiments for estimating percentiles of black-box functions. *ArXiv e-prints*, 2016.
- Loic Le Gratiet. *Multi-fidelity Gaussian process regression for computer experiments*. PhD thesis, Université Paris-Diderot-Paris VII, october 2013.
- Loic Le Gratiet et Claire Cannamela. Cokriging-based sequential design strategies using fast cross-validation techniques for multi-fidelity computer codes. *Technometrics*, 57(3) :418–427, 2015. doi : <http://dx.doi.org/10.1080/00401706.2014.928233>.

- Loic Le Gratiet et Josselin Garnier. Recursive co-kriging model for design of computer experiments with multiple levels of fidelity. *International Journal for Uncertainty Quantification*, 4(5) :365–386, 2014.
- Olivier Le Maître et Omar M Knio. *Spectral methods for uncertainty quantification : with applications to computational fluid dynamics*. Springer Science & Business Media, 2010. doi : <http://dx.doi.org/10.1007/978-90-481-3520-2>.
- Stephen J. Leary, Atul Bhaskar, et Andy J. Keane. A knowledge-based approach to response surface modelling in multifidelity optimization. *Journal of Global Optimization*, 26(3) :297–319, 2003. doi : <http://dx.doi.org/10.1023/A:1023283917997>.
- Leifur Leifsson, Slawomir Koziel, et Stanislav Ogurtsov. Low-fidelity model mesh density and the performance of variable-resolution shape optimization algorithms. *Procedia Computer Science*, 9 :842–851, 2012. doi : <http://dx.doi.org/10.1016/j.procs.2012.04.090>.
- Ling Li. *Sequential Design of Experiments to Estimate a Probability of Failure*. PhD thesis, Supélec, may 2012.
- Chih-Hung Lin, Yuh-Ming Ferng, et Wen-Shieng Hsu. Investigating the effect of computational grid sizes on the predicted characteristics of thermal radiation for a fire. *Applied Thermal Engineering*, 29(11–12) :2243–2250, 2009. doi : <http://dx.doi.org/10.1016/j.applthermaleng.2008.11.010>.
- Jason L. Loeppky, Jerome Sacks, et William J. Welch. Choosing the sample size of a computer experiment : A practical guide. *Technometrics*, 51(4) :366–376, 2009. doi : <https://doi.org/10.1198/TECH.2009.08040>.
- David J. C. MacKay. Information-based objective functions for active data selection. *Neural Computation*, 4(4) :590–604, 1992. doi : <https://doi.org/10.1162/neco.1992.4.4.590>.
- Amandine Marrel, Bertrand Iooss, Sébastien Da Veiga, et Mathieu Ribatet. Global sensitivity analysis of stochastic computer models with joint metamodels. *Statistics and Computing*, 22(3) :833–847, may 2012. doi : [10.1007/s11222-011-9274-8](https://doi.org/10.1007/s11222-011-9274-8).
- Jorge Mateu, Gema Fernández-Avilés, et José María Montero. On a class of non-stationary, compactly supported spatial covariance functions. *Stochastic Environmental Research and Risk Assessment*, 27(2) :297–309, 2013. doi : <http://dx.doi.org/10.1007/s00477-011-0510-8>.

- Kevin McGrattan et Blaza Toman. Quantifying the predictive uncertainty of complex numerical models. *Metrologia*, 48(3) :173, 2011.
- Kevin McGrattan, Randall McDermottand, Simo Hostikka, et Jason E. Floyd. *Fire dynamics simulator (version 5), user's guide*. National Institute of Standards and Technology (NIST), 2010.
- Kevin McGrattan, Simo Hostikka, Randall McDermott, Jason Floyd, Craig Weinschenk, et Kristopher Overholt. *Fire Dynamics Simulator - Technical Reference Guide - Volume 1 : Mathematical Model*. National Institute of Standards and Technology (NIST), Gaithersburg, Maryland, USA, NIST special publication 1018 edition, september 2014.
- Michael D. McKay, Richard J. Beckman, et William J. Conover. Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2) :239–245, 1979. doi : <https://doi.org/10.1080/00401706.1979.10489755>.
- Max D. Morris et Toby J. Mitchell. Exploratory designs for computational experiments. *Journal of Statistical Planning and Inference*, 43(3) :381–402, 1995. doi : [http://dx.doi.org/10.1016/0378-3758\(94\)00035-T](http://dx.doi.org/10.1016/0378-3758(94)00035-T).
- Vincent Moutoussamy, Simon Nanty, et Benoît Pauwels. Emulators for stochastic simulation codes. *ESAIM : Proceedings and Surveys*, 48 :116–155, 2015. doi : <https://doi.org/10.1051/proc/201448005>.
- norme ISO 13571. *Life-threatening components of fire – Guidelines for the estimation of time to compromised tenability in fires*. International Organization for Standardization (ISO), Geneva, Switzerland, 2012.
- Jeremy E. Oakley et Benjamin D. Youngman. Calibration of stochastic computer simulators using likelihood emulation. *Technometrics*, 59(1) :80–92, 2017. doi : [10.1080/00401706.2015.1125391](https://doi.org/10.1080/00401706.2015.1125391).
- Anthony O'Hagan. A Markov property for covariance structures. Statistics Research Report, september 1998.
- Christopher J. Paciorek et Mark J. Schervish. Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics*, 17(5) :483–506, 2006. doi : <http://dx.doi.org/10.1002/env.785>.

- Guofei Pang, Paris Perdikaris, Wei Cai, et George Em Karniadakis. Discovering variable fractional orders of advection–dispersion equations from field data using multi-fidelity Bayesian optimization. *Journal of Computational Physics*, 348(Supplement C) :694–714, 2017. doi : <https://doi.org/10.1016/j.jcp.2017.07.052>.
- Alberto Pasanisi et Anne Dutoy. An industrial viewpoint on uncertainty quantification in simulation : Stakes, methods, tools, examples. In Andrew M. Dienstfrey et Ronald F. Boisvert, éditeurs, *Uncertainty Quantification in Scientific Computing*, pages 27–45, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.
- Paris Perdikaris et George Em Karniadakis. Model inversion via multi-fidelity Bayesian optimization : a new paradigm for parameter estimation in haemodynamics, and beyond. *Journal of The Royal Society Interface*, 13(118), 2016. doi : <http://dx.doi.org/10.1098/rsif.2015.1107>.
- Paris Perdikaris, Daniele Venturi, Johannes O. Royset, et George Em Karniadakis. Multi-fidelity modelling via recursive co-kriging and Gaussian–Markov random fields. *Proceedings of the Royal Society of London A : Mathematical, Physical and Engineering Sciences*, 471(2179), 2015. doi : <http://dx.doi.org/10.1098/rspa.2015.0018>.
- Victor Picheny et David Ginsbourger. A nonstationary space-time Gaussian process model for partially converged simulations. *SIAM/ASA Journal on Uncertainty Quantification*, 1(1) :57–78, 2013. doi : <http://dx.doi.org/10.1137/120882834>.
- Victor Picheny, David Ginsbourger, Olivier Roustant, Raphael T. Haftka, et Nam-Ho Kim. Adaptive designs of experiments for accurate approximation of a target region. *Journal of Mechanical Design*, 132(7) :071008, June 2010. doi : <http://dx.doi.org/10.1115/1.4001873>.
- Victor Picheny, David Ginsbourger, Yann Richet, et Gregory Caplin. Quantile-based optimization of noisy computer experiments with tunable precision. *Technometrics*, 55(1) :2–13, 2013. doi : <http://dx.doi.org/10.1080/00401706.2012.707580>.
- Luc Pronzato et Werner G. Müller. Design of computer experiments : space filling and beyond. *Statistics and Computing*, 22(3) :681–701, 2012. doi : <http://dx.doi.org/10.1007/s11222-011-9242-3>.
- Peter Z. G. Qian. Nested latin hypercube designs. *Biometrika*, 96(4) :957–970, 2009. doi : <http://dx.doi.org/10.1093/biomet/asp045>.

- Peter Z. G. Qian et C. F. Jeff Wu. Bayesian hierarchical modeling for integrating low-accuracy and high-accuracy experiments. *Technometrics*, 50(2) :192–204, 2008. doi : <http://dx.doi.org/10.1198/004017008000000082>.
- Peter Z. G. Qian, Huaiqing Wu, et C. F. Jeff Wu. Gaussian process models for computer experiments with qualitative and quantitative factors. *Technometrics*, 50(3) :383–396, 2008. doi : <http://dx.doi.org/10.1198/004017008000000262>.
- Maziar Raissi, Paris Perdikaris, et George Em Karniadakis. Inferring solutions of differential equations using noisy multi-fidelity data. *Journal of Computational Physics*, 335(Supplement C) :736–746, 2017. doi : <https://doi.org/10.1016/j.jcp.2017.01.060>.
- Pritam Ranjan, Derek Bingham, et George Michailidis. Sequential experiment design for contour estimation from complex computer codes. *Technometrics*, 50(4) :527–541, 2008. doi : <https://doi.org/10.1198/004017008000000541>.
- Carl Edward Rasmussen et Christopher K.I. Williams. *Gaussian processes for machine learning*. The MIT Press, Cambridge, MA, USA, 2006.
- Gijs Rennen. *Efficient approximation of black-box functions and Pareto sets*. PhD thesis, Tilburg University, november 2009.
- Gijs Rennen, Bart Husslage, Edwin R. Van Dam, et Dick Den Hertog. Nested maximin latin hypercube designs. *Structural and Multidisciplinary Optimization*, 41(3) :371–395, 2010. doi : <http://dx.doi.org/10.1007/s00158-009-0432-y>.
- Christian P. Robert. *Monte Carlo Methods*. John Wiley & Sons, Ltd, 2014. doi : <http://dx.doi.org/10.1002/9781118445112.stat03876.pub2>.
- Jerome Sacks, William J. Welch, Toby J. Mitchell, et Henry P. Wynn. Design and analysis of computer experiments. *Statistical science*, 4(4) :409–423, 1989.
- Thomas J. Santner, Brian J. Williams, et William I. Notz. *The Design and Analysis of Computer Experiments*. Springer Series in Statistics. Springer, New York, 2003. doi : <http://dx.doi.org/10.1007/978-1-4757-3799-8>.
- Luc Schueremans. *Probabilistic evaluation of structural unreinforced masonry*. PhD thesis, KU Leuven, 2001.
- Warren Scott, Peter Frazier, et Warren Powell. The correlated knowledge gradient for simulation optimization of continuous parameters using Gaussian process regression.

- SIAM Journal on Optimization*, 21(3) :996–1026, 2011. doi : <https://doi.org/10.1137/100801275>.
- Jasper Snoek, Hugo Larochelle, et Ryan P. Adams. Practical Bayesian optimization of machine learning algorithms. In F. Pereira, C. J. C. Burges, L. Bottou, et K. Q. Weinberger, éditeurs, *Advances in Neural Information Processing Systems*, volume 25, pages 2951–2959. Curran Associates, Inc., 2012.
- Niranjan Srinivas, Andreas Krause, Sham M. Kakade, et Matthias W. Seeger. Gaussian process optimization in the bandit setting : No regret and experimental design. *arXiv*, December 2009.
- Michael L. Stein. *Interpolation of Spatial Data : Some Theory for Kriging*. Springer Series in Statistics. Springer, New York, 1999. doi : <http://dx.doi.org/10.1007/978-1-4612-1494-6>.
- Michael L. Stein. Space–time covariance functions. *Journal of the American Statistical Association*, 100(469) :310–321, 2005. doi : <http://dx.doi.org/10.1198/016214504000000854>.
- Michael L. Stein. 2010 Rietz lecture : When does the screening effect hold? *The Annals of Statistics*, 39(6) :2795–2819, december 2011. doi : <https://doi.org/10.1214/11-AOS909>.
- Tata Subba Rao et Gy Terdik. A space-time covariance function for spatio-temporal random processes and spatio-temporal prediction (kriging). *ArXiv e-prints*, november 2013.
- Kevin Swersky, Jasper Snoek, et Ryan P. Adams. Multi-task Bayesian optimization. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, et K. Q. Weinberger, éditeurs, *Advances in Neural Information Processing Systems*, volume 26, pages 2004–2012. Curran Associates, Inc., 2013.
- Boxin Tang. Orthogonal array-based latin hypercubes. *Journal of the American statistical association*, 88(424) :1392–1397, 1993. doi : <http://dx.doi.org/10.1080/01621459.1993.10476423>.
- Arthur Thenon, Véronique Gervais, et Mickaële Le Ravalec. Multi-fidelity meta-modeling for reservoir engineering-application to history matching. *Computational Geosciences*, 20(6) :1231–1250, december 2016. doi : [10.1007/s10596-016-9587-y](https://doi.org/10.1007/s10596-016-9587-y).

- Anne Thiry-Muller. Guide de bonnes pratiques pour les études d'ingénierie du désenfumage dans les établissements recevant du public. Technical Report 14/2811/RG1, Laboratoire Central de la Préfecture de Police, Paris, France, November 2017.
- Rui Tuo, C. F. Jeff Wu, et Dan Yu. Surrogate modeling of computer experiments with different mesh densities. *Technometrics*, 56(3) :372–380, 2014. doi : <http://dx.doi.org/10.1080/00401706.2013.842935>.
- Bart Van Weyenberge, Xavier Deckers, Robby Caspeele, et Bart Merci. Development of a full probabilistic qra method for quantifying the life safety risk in complex building designs. In *Performance based codes and fire safety design methods*, pages 1–12. Society of Fire Protection Engineers (SFPE), may 2016.
- Emmanuel Vazquez et Julien Bect. A sequential Bayesian algorithm to estimate a probability of failure. *IFAC Proceedings Volumes*, 42(10) :546–550, 2009.
- Julien Villemonteix, Emmanuel Vazquez, et Eric Walter. An informational approach to the global optimization of expensive-to-evaluate functions. *Journal of Global Optimization*, 44(4) :509, September 2008. doi : <https://doi.org/10.1007/s10898-008-9354-2>.
- Mores J. Wankhede. *Multi-fidelity strategies for lean burn combustor design*. PhD thesis, University of Southampton, january 2012.
- William J. Welch, Robert J. Buck, Jerome Sacks, Henry P. Wynn, Toby J. Mitchell, et Max D. Morris. Screening, predicting, and computer experiments. *Technometrics*, 34(1) :15–25, 1992. doi : <https://doi.org/10.1080/00401706.1992.10485229>.
- Brian J. Williams, Jason L. Loepky, Leslie M. Moore, et Mason S. Macklem. Batch sequential design to achieve predictive maturity with calibrated computer models. *Reliability Engineering & System Safety*, 96(9) :1208–1219, 2011. doi : <https://doi.org/10.1016/j.res.2010.04.017>. Quantification of Margins and Uncertainties.
- Shifeng Xiong, Peter Z. G. Qian, et C. F. Jeff Wu. Sequential design and analysis of high-accuracy and low-accuracy computer codes. *Technometrics*, 55(1) :37–46, 2013. doi : <http://dx.doi.org/10.1080/00401706.2012.723572>.
- Jin Xu, Xiaojun Duan, Zhengming Wang, et Liang Yan. A general construction for nested latin hypercube designs. *Statistics & Probability Letters*, 2017. doi : <https://doi.org/10.1016/j.spl.2017.10.022>.

- Jinyu Yang, Min-Qian Liu, et Dennis K. J. Lin. Construction of nested orthogonal latin hypercube designs. *Statistica Sinica*, 24(1) :211–219, 2014. doi : <http://dx.doi.org/dx.doi.org/10.5705/ss.2012.139>.
- Qi Zhou, Yan Wang, Seung-Kyum Choi, Ping Jiang, Xinyu Shao, et Jiexiang Hu. A sequential multi-fidelity metamodeling approach for data regression. *Knowledge-Based Systems*, 134(Supplement C) :199–212, 2017. doi : <https://doi.org/10.1016/j.knosys.2017.07.033>.
- Dale Zimmerman, Claire Pavlik, Amy Ruggles, et Marc P. Armstrong. An experimental comparison of ordinary and universal kriging and inverse distance weighting. *Mathematical Geology*, 31(4) :375–390, 1999. doi : <http://dx.doi.org/10.1023/A:1007586507433>.
- Enrico Zio. *The Monte-Carlo simulation method for system reliability and risk analysis*, volume 39. Springer, London, United Kingdom, 2013. doi : <https://doi.org/10.1007/978-1-4471-4588-2>.

Titre : Planification d'expériences numériques en multi-fidélité. Application à un simulateur d'incendies.

Mots clés : Expériences numériques, Co-krigeage, Planification séquentielle, Multi-fidélité, Probabilité de défaillance, Sécurité incendie

Résumé : Les travaux présentés portent sur l'étude de modèles numériques multi-fidèles, déterministes ou stochastiques. Plus précisément, les modèles considérés disposent d'un paramètre réglant la qualité de la simulation, comme une taille de maille dans un modèle par différences finies, ou un nombre d'échantillons dans un modèle de Monte-Carlo. Dans ce cas, il est possible de lancer des simulations basse fidélité, rapides mais grossières, et des simulations haute fidélité, fiables mais coûteuses. L'intérêt d'une approche multi-fidèle est de combiner les résultats obtenus aux différents niveaux de fidélité afin d'économiser du temps de simulation. La méthode considérée est fondée sur une approche bayésienne. Le simulateur est décrit par un modèle de processus gaussiens multi-niveaux développé dans la littérature que nous adaptons aux cas stochastiques dans une ap-

proche complètement bayésienne. Ce méta-modèle du simulateur permet d'obtenir des estimations de quantités d'intérêt, accompagnés d'une mesure de l'incertitude associée. L'objectif est alors de choisir de nouvelles expériences à lancer afin d'améliorer les estimations. En particulier, la planification doit sélectionner le niveau de fidélité réalisant le meilleur compromis entre coût d'observation et gain d'information. Pour cela, nous proposons une stratégie séquentielle adaptée au cas où les coûts d'observation sont variables. Cette stratégie, intitulée « Maximal Rate of Uncertainty Reduction » (MRUR), consiste à choisir le point d'observation maximisant le rapport entre la réduction d'incertitude et le coût. La méthodologie est illustrée en sécurité incendie, où nous cherchons à estimer des probabilités de défaillance d'un système de désenfumage.

Title : Sequential design of numerical experiments in multi-fidelity. Application to a fire simulator.

Keywords : Numerical experiments, Co-kriging, Sequential design, Multi-fidelity, Probability of failure, Fire safety

Abstract : The presented works focus on the study of multi-fidelity numerical models, deterministic or stochastic. More precisely, the considered models have a parameter which rules the quality of the simulation, as a mesh size in a finite difference model or a number of samples in a Monte-Carlo model. In that case, the numerical model can run low-fidelity simulations, fast but coarse, or high-fidelity simulations, accurate but expensive. A multi-fidelity approach aims to combine results coming from different levels of fidelity in order to save computational time. The considered method is based on a Bayesian approach. The simulator is described by a state-of-art multilevel Gaussian process model which we adapt to stochastic cases in a fully-Bayesian approach. This

meta-model of the simulator allows estimating any quantity of interest with a measure of uncertainty. The goal is to choose new experiments to run in order to improve the estimations. In particular, the design must select the level of fidelity meeting the best trade-off between cost of observation and information gain. To do this, we propose a sequential strategy dedicated to the cases of variable costs, called Maximum Rate of Uncertainty Reduction (MRUR), which consists of choosing the input point maximizing the ratio between the uncertainty reduction and the cost. The methodology is illustrated in fire safety science, where we estimate probabilities of failure of a fire protection system.

