



HAL
open science

Estimation de paramètres et planification d'expériences adaptée aux problèmes de cinétique - Application à la dépollution des fumées en sortie des moteurs

Matthieu Canaud

► **To cite this version:**

Matthieu Canaud. Estimation de paramètres et planification d'expériences adaptée aux problèmes de cinétique - Application à la dépollution des fumées en sortie des moteurs. Statistiques [math.ST]. Ecole Nationale Supérieure des Mines de Saint-Etienne, 2011. Français. NNT : 2011 EMSE 0619 . tel-00677758v1

HAL Id: tel-00677758

<https://theses.hal.science/tel-00677758v1>

Submitted on 9 Mar 2012 (v1), last revised 12 Jul 2013 (v2)

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.

NNT : 2011 EMSE 0619

THÈSE

présentée par

Matthieu CANAUD

pour obtenir le grade de

Docteur de l'École Nationale Supérieure des Mines de Saint-Étienne

Spécialité : Mathématiques Appliquées et Statistiques

Estimation de paramètres et planification d'expériences adaptée aux
problèmes de cinétique

Application à la dépollution des fumées en sortie des moteurs

soutenue à Saint-Étienne, le 14 septembre 2011

Membres du jury

Rapporteurs :	Clémentine PRIEUR	Professeur, Université Joseph Fourier, Grenoble
	Patrick DA COSTA	Professeur, Université Pierre et Marie Curie, Paris
Examineur :	Bertrand IOOSS	Ingénieur Chercheur, EDF, Chatou
Directeur de thèse :	Laurent CARRARO	Professeur, Telecom Saint-Étienne, Saint-Étienne
Co-directeur de thèse :	Céline HELBERT	Maître de Conférences, Université Pierre Mendès France, Grenoble
Promoteur IFPEN :	François WAHL	Ingénieur Chercheur, IFP Energies Nouvelles, Lyon

Spécialités doctorales :

SCIENCES ET GENIE DES MATERIAUX
 MECANIQUE ET INGENIERIE
 GENIE DES PROCEDES
 SCIENCES DE LA TERRE
 SCIENCES ET GENIE DE L'ENVIRONNEMENT
 MATHEMATIQUES APPLIQUEES
 INFORMATIQUE
 IMAGE, VISION, SIGNAL
 GENIE INDUSTRIEL
 MICROELECTRONIQUE

Responsables :

J. DRIVER Directeur de recherche – Centre SMS
 A. VAUTRIN Professeur – Centre SMS
 G. THOMAS Professeur – Centre SPIN
 B. GUY Maître de recherche – Centre SPIN
 J. BOURGOIS Professeur – Centre SITE
 E. TOUBOUL Ingénieur – Centre G2I
 O. BOISSIER Professeur – Centre G2I
 JC. PINOLI Professeur – Centre CIS
 P. BURLAT Professeur – Centre G2I
 Ph. COLLOT Professeur – Centre CMP

Enseignants-chercheurs et chercheurs autorisés à diriger des thèses de doctorat (titulaires d'un doctorat d'État ou d'une HDR)

AVRIL	Stéphane	MA	Mécanique & Ingénierie	CIS
BATTON-HUBERT	Mireille	MA	Sciences & Génie de l'Environnement	SITE
BENABEN	Patrick	PR 1	Sciences & Génie des Matériaux	CMP
BERNACHE-ASSOLLANT	Didier	PR 0	Génie des Procédés	CIS
BIGOT	Jean-Pierre	MR	Génie des Procédés	SPIN
BILAL	Essaïd	DR	Sciences de la Terre	SPIN
BOISSIER	Olivier	PR 1	Informatique	G2I
BORBELY	Andras	MR	Sciences et Génie des Matériaux	SMS
BOUCHER	Xavier	MA	Génie Industriel	G2I
BOUDAREL	Marie-Reine	PR 2	Génie Industriel	DF
BOURGOIS	Jacques	PR 0	Sciences & Génie de l'Environnement	SITE
BRODHAG	Christian	DR	Sciences & Génie de l'Environnement	SITE
BURLAT	Patrick	PR 2	Génie industriel	G2I
COLLOT	Philippe	PR 1	Microélectronique	CMP
COURNIL	Michel	PR 0	Génie des Procédés	SPIN
DAUZERE-PERES	Stéphane	PR 1	Génie industriel	CMP
DARRIEULAT	Michel	IGM	Sciences & Génie des Matériaux	SMS
DECHOMETS	Roland	PR 1	Sciences & Génie de l'Environnement	SITE
DESRAYAUD	Christophe	MA	Mécanique & Ingénierie	SMS
DELAFOSSÉ	David	PR 1	Sciences & Génie des Matériaux	SMS
DOLGUI	Alexandre	PR 1	Génie Industriel	G2I
DRAPIER	Sylvain	PR 2	Mécanique & Ingénierie	SMS
DRIVER	Julian	DR 0	Sciences & Génie des Matériaux	SMS
FEILLET	Dominique	PR 2	Génie Industriel	CMP
FOREST	Bernard	PR 1	Sciences & Génie des Matériaux	CIS
FORMISYN	Pascal	PR 1	Sciences & Génie de l'Environnement	SITE
FRACZKIEWICZ	Anna	DR	Sciences & Génie des Matériaux	SMS
GARCIA	Daniel	MR	Génie des Procédés	SPIN
GIRARDOT	Jean-Jacques	MR	Informatique	G2I
GOEURIOT	Dominique	MR	Sciences & Génie des Matériaux	SMS
GRAILLOT	Didier	DR	Sciences & Génie de l'Environnement	SITE
GROSSEAU	Philippe	MR	Génie des Procédés	SPIN
GRUY	Frédéric	MR	Génie des Procédés	SPIN
GUY	Bernard	MR	Sciences de la Terre	SPIN
GUYONNET	René	DR	Génie des Procédés	SPIN
HERRI	Jean-Michel	PR 2	Génie des Procédés	SPIN
INAL	Karim	PR 2	Microélectronique	CMP
KLÖCKER	Helmut	DR	Sciences & Génie des Matériaux	SMS
LAFORÉST	Valérie	CR	Sciences & Génie de l'Environnement	SITE
LERICHE	Rodolphe	CR CNRS	Mécanique et Ingénierie	SMS
LI	Jean-Michel	EC (CCI MP)	Microélectronique	CMP
MALLIARAS	George Grégory	PR 1	Microélectronique	CMP
MOLIMARD	Jérôme	MA	Mécanique et Ingénierie	SMS
MONTHEILLET	Frank	DR 1 CNRS	Sciences & Génie des Matériaux	SMS
PERIER-CAMBY	Laurent	PR 2	Génie des Procédés	SPIN
PIJOLAT	Christophe	PR 1	Génie des Procédés	SPIN
PIJOLAT	Michèle	PR 1	Génie des Procédés	SPIN
PINOLI	Jean-Charles	PR 0	Image, Vision, Signal	CIS
STOLARZ	Jacques	CR	Sciences & Génie des Matériaux	SMS
SZAFNICKI	Konrad	MR	Sciences & Génie de l'Environnement	SITE
THOMAS	Gérard	PR 0	Génie des Procédés	SPIN
TRIA	Assia		Microélectronique	CMP
VALDIVIESO	François	MA	Sciences & Génie des Matériaux	SMS
VAUTRIN	Alain	PR 0	Mécanique & Ingénierie	SMS
VIRICELLE	Jean-Paul	MR	Génie des procédés	SPIN
WOLSKI	Krzysztof	DR	Sciences & Génie des Matériaux	SMS
XIE	Xiaolan	PR 1	Génie industriel	CIS

Glossaire :

PR 0	Professeur classe exceptionnelle
PR 1	Professeur 1 ^{ère} classe
PR 2	Professeur 2 ^{ème} classe
MA(MDC)	Maître assistant
DR	Directeur de recherche
Ing.	Ingénieur
MR(DR2)	Maître de recherche
CR	Chargé de recherche
EC	Enseignant-chercheur
IGM	Ingénieur général des mines

Centres :

SMS	Sciences des Matériaux et des Structures
SPIN	Sciences des Processus Industriels et Naturels
SITE	Sciences Information et Technologies pour l'Environnement
G2I	Génie Industriel et Informatique
CMP	Centre de Microélectronique de Provence
CIS	Centre Ingénierie et Santé



Remerciements

Qu'on se le dise, la thèse est une formidable aventure. Une aventure en effet, jonchée d'obstacles, de dur labeur, de remises en cause, synonymes, bien souvent, de bons moments, de fierté et d'entraide. Les pages qui vont suivre se veulent donc une synthèse des travaux menés. Mais quels seraient ces travaux, sans toutes les personnes impliquées de près ou de loin ? Les encadrants, les amis, les collègues, les rencontres d'un jour, les disparitions de certains...il apparaît donc clairement, au moment de rédiger ces remerciements, toute la difficulté et je dirais même toute l'impossibilité d'être exhaustif dans cette tâche. Je commencerai donc cette page, en remerciant toutes les personnes qui n'y seront pas citées explicitement, mais qui se sentent concernées par ce travail.

Je voudrais commencer ces remerciements en mettant en lumière les trois personnes qui m'ont encadré au cours des ces trois années. Tout d'abord, je tiens à remercier Mr Laurent CARRARO, qui malgré un emploi du temps de ministre, a toujours su prendre le temps nécessaire, mais surtout a toujours apporté son intérêt, sa connaissance et ses compétences pour ces travaux. Ses grandes capacités en font un grand monsieur, mais qui a su rester humble, toujours au service des autres, afin de faire avancer la recherche mais surtout les chercheurs dans leur développement. En deuxième lieu, je voudrais remercier Mme Céline Helbert. Sa gentillesse et sa fraîcheur sont autant d'atouts qui combinés à son savoir et sa soif de le partager, ont permis de mener à bien ces travaux dans une bonne ambiance. Ce mélange de bonne humeur, d'optimisme et de rigueur est, à mon sens, un atout primordial pour un encadrant, et m'ont à plusieurs reprises évité des périodes de doutes et d'une manière générale fait progresser. J'en arrive maintenant à Mr François Wahl. Que dire ? Un grand merci ! Ce ne fut pas toujours évident pour toi je pense, mais ma gratitude n'en est que plus grande. C'est lui qui à la suite de mon stage à l'IFP Energies Nouvelles, m'a permis d'entrevoir ce qu'est le monde de la recherche en statistique, et m'a accueilli dans ce beau projet. Il a su me transmettre son savoir scientifique aussi bien du point de vue de la réflexion que de la rédaction (et la tâche était assez grande). Mais au delà de cet aspect, il a toujours répondu présent que se soit professionnellement, scientifiquement ou bien humainement. Pour toutes ces raisons, merci à vous trois pour m'avoir formé, encadré et fait progresser. Et merci également pour l'amitié que vous m'avez accordé.

Je suis également honoré que les professeurs Clémentine Prieur et Patrick Da Costa aient accepté d'être rapporteurs de cette thèse. Je les remercie pour la lecture attentive de ce manuscrit, leurs remarques et pour l'intérêt qu'ils ont porté à ce travail. Ma gratitude va également à Bertrand Iooss qui m'a fait le plaisir d'accepter d'être examinateur.

D'autres personnes ont également contribué à la réalisation de ces travaux de manière ponctuelle, mais tout aussi précieuse. En premier lieu, merci à Mr Pascal Duchêne, dont les critiques parfois acérées mais toujours constructives, m'ont appris le sens de la rigueur mathématique. Ces nombreux conseils aussi bien mathématique, cinétique qu'informatique, ont bien souvent permis de faire avancer les travaux efficacement. Dans un second temps, merci à Mme Claire-Noëlle Millet, l'expérimentatrice à l'origine de ce sujet. Une personne très agréable, toujours souriante et disponible pour m'expliquer le fonctionnement parfois complexe de la cinétique. Également merci à Eric Jeudy, qui a mené la campagne expérimentale. Dans ce cadre, je voudrais également remercier Mr Olivier Roustant qui m'a encadré au début de cette thèse, et Mr Xavier Bay, qui a accepté de partager un peu de son grand savoir sur les ponts browniens et son secret pour dénicher les morilles. Ces cinq personnes ont gravité autour de ces travaux dans une orbite un peu plus éloignée que mes trois encadrants, mais comme dans l'espace, leur influence est loin d'être négligeable.

Toutes ces personnes ont contribué à mener à bien ces travaux. Mais l'aspect scientifique, n'est pas le seul aspect,

dont il faut tenir compte, et pour la bonne ambiance de travail que chacun a apporté, j'aimerais remercier toutes les personnes de l'IFP Energies Nouvelles. Un lieu de passage pour beaucoup, mais j'en retiens surtout des rencontres humainement superbes. Je pense notamment aux titulaires, aux thésards anciens et nouveaux, aux stagiaires.... Chacun de vous sait, je pense, que notre rencontre fut importante. Merci donc à vous (dans l'ordre chronologique, afin d'essayer de n'oublier personnes, si c'est le cas, se reporter au premier paragraphe) : les anciens thésards Sébastien, Eugenio, Rachid, Régis et Olivier ; les titulaires Yann, Paolo, Mister Pascal Roux, Daniel, Sébastien, Maxime, Eric, Bilal, Philippe, Laurent, Luc (Mojito man) et Mister Pigeon (grand merci mec) ; ceux du BE Gael (Funkalicious!), Jean-Christophe, Jean-Charles, Stéphane, Max ; les " nouveaux " Florent, Mimmo, Johan, Anthony, Dimitrius, Viet-Dung ; ceux de passage Marianne, Damy, Aimad, Fabien, Mahmoud ; un merci spécial au breton, Jeremy ; aux secrétaires, aux personnes de la section sport et également à Leïla et doudou pour les bons repas et les cafés !

Je tiens également à remercier les personnes qui forment ce magnifique labo qu'est le 3MI. Mon temps de séjour y fut court, mais les personnes y sont vraiment sympathiques et professionnelles. Merci donc à David, Delphine, Eric, Rodolphe, Nicolas, Bertrand, Christine et Victor.

Je tenais aussi à remercier trois personnes qui ont été importantes dans mon choix de parcours Mme Anne Perrut, Mr Thierry Clopeau et Mme Gabriella Ciuperca de L'UCBL1.

Ces remerciements ne seraient pas complets sans citer également mes amis, mes potes qui ont toujours été là pour moi, même si des fois je disparaissais de la circulation. Un grand merci à vous Bib, Maxou, Neuneu, Djules, Ralph, Cinoc, Cyril, Mr Pietri et Matthieu (ca fait une liste masculine, mais j'inclus les conjointes). Merci à vous les Vichyssois Elie (my wonderful Halfbro!), Taz, Sawa, Hanny, Tomtom, Koko, Halal, Brax to brax, les Cartons. Merci à vous les bretons Jo (mon petit), Seb (ma caille) et Blanblan. Enfin merci également aux potes de master, Dr R2, Thyerrix et Huabin !

J'en oublie sûrement donc en vrac, je remercie également Sab, Max, Alex, Skand, Tonyo, Nour-Edin, David, Julien, German (kiss), Ceptitchat, Samy et Ulysse et également les personnes avec qui j'ai travaillé pendant mes études Thierry, Laurent, Cyril et tous ceux de la BaO !

Enfin j'en arrive aux remerciements qui me tiennent particulièrement à coeur, et pour cause, ceux que j'adresse à ma famille. Une famille formidable, à qui je ne dis pas assez, à quel point ils sont importants. Je profite donc de cette page pour vous remercier infiniment. Ce travail vous est dédié à vous mes parents (extraordinaires!) et mes frères et soeurs : Dam, Fofi et Nana (exceptionnels!). A mon grand père, et mes grands parents de Ganillon qui souvent me manquent. A mon parrain, qui m'a initié aux festoches!! A ma marraine, qui a su être une seconde mère. A mes beaux frères et soeurs et mes petites cousines! Merci aussi à mon filleul Hugo, même si son apport reste modeste pour le moment! Je tiens également à remercier Micheline (ma troisième mère!), Joëlle et Brigitte. Un merci enfin à Alain (papanou), Lydie (mamanou), et mémé qui m'ont accueilli dans leur famille, et que je considère comme faisant partie de ce vaste ensemble de personnes très importantes pour moi.

Pour finir, je voudrais remercier celle qui m'a supporté pendant ces trois années de thèse, celle qui j'espère me supportera encore très longtemps, la lune de mes jours, le soleil de mes nuits, celle qui parfois me fait m'énerver (pas bien mon genre pourtant), celle qui me fait m'émouvoir, celle qui me fait rire, celle qui me fait craquer.....enfin bref celle que j'aime, ma moitié : Célia.



Table des matières

Remerciements	1
Introduction	13
1 Phénomènes étudiés	19
1.1 Piège à NOx	22
1.1.1 Fonctionnement du piège à NOx	22
1.1.2 Modèle cinétique	25
1.1.3 Modèle numérique	30
1.1.4 Comparaison résultats expérimentaux et résultats numériques	34
1.1.5 Modification du modèle numérique	35
1.2 Cas maîtrisé	37
1.2.1 Présentation du cas maîtrisé et motivations	37
1.2.2 Simulations	38
1.2.3 Pertinence du cas maîtrisé	39
1.3 En bref	40
2 Outils de modélisation	41
2.1 Krigeage	43
2.1.1 Krigeage	43
2.1.2 Choix du noyau	49
2.1.2.1 Familles classiques de covariances	49
2.1.2.2 Noyaux de covariance non stationnaires	50
2.1.3 Krigeage avec bruit	52
2.1.4 Cokrigeage	54
2.2 Données fonctionnelles	58
2.2.1 Définition et état de l'art	59
2.2.2 Modèles à coefficients variables	60
2.2.2.1 Présentation	61
2.2.2.2 Méthodes d'estimations	63
2.2.3 Modèles non paramétriques additifs	69
2.3 Conclusions	73
3 Approche non-fonctionnelle	75
3.1 Krigeage à tendance non linéaire	79
3.1.1 Prédicteur de krigeage	79
3.1.2 Estimation des paramètres	80
3.1.3 Algorithme mis en place	80
3.1.4 Applications	82
3.1.4.1 Comparaison krigeage classique et à tendance non linéaire sur le cas maîtrisé	82

3.1.4.2	Application du krigeage à tendance non linéaire sur le cas d'étude	84
3.1.5	Conclusions et limites	88
3.2	Conditionnement par la dérivée	89
3.2.1	Présentation de la démarche	89
3.2.2	Application sur le cas maîtrisé	91
3.2.3	Application sur le cas d'étude	93
3.2.4	Conclusions et limites	96
3.3	Modélisation de la dérivée par rapport à l'avancement	97
3.3.1	Présentation de la démarche	98
3.3.2	Résultats sur le cas d'étude	100
3.3.3	Intégration des résultats	102
3.3.4	Calcul de la variance de prédiction	106
3.3.5	Conclusions et limites	108
3.4	Conclusions	109
4	Approche fonctionnelle	111
4.1	Modèles à coefficients variables	113
4.1.1	Introduction des modèles à coefficients variables	114
4.1.1.1	Motivations	114
4.1.1.2	Choix des variables et étude de la dépendance linéaire	115
4.1.2	Méthode 1 : Estimation des fonctions coefficients par P-splines	119
4.1.2.1	Première étape : Estimation point à point des fonctions coefficients par moindres carrés	120
4.1.2.2	Seconde étape : Lissage des fonctions coefficients par P-spline	122
4.1.3	Méthode 2 : Estimation des fonctions coefficients par polynômes locaux	125
4.1.3.1	Présentation de la méthode	125
4.1.3.2	Applications et résultats	129
4.1.4	Conclusions	130
4.2	Estimation des paramètres cinétiques	131
4.2.1	Modifications du modèle numérique et estimations des paramètres cinétiques	132
4.2.2	Modélisation des paramètres cinétiques en fonction des entrées	135
4.2.2.1	Estimation des coefficients d'une relation linéaire entre prédicteurs et paramètres cinétiques	136
4.2.2.2	Applications et résultats de la KRR	137
4.2.2.3	Récapitulatif des résultats des différentes approches et conclusion	139
4.2.3	Calcul des régions de confiance	140
4.2.3.1	Bootstrap fonctionnel	141
4.2.3.2	Estimation de la loi des paramètres de la modélisation par KRR	143
4.2.3.3	Calcul de la région de confiance	144
4.2.4	Conclusions	145
4.3	Conclusions et perspectives	145
5	Plans d'expériences exploratoires	149
5.1	Motivations	151
5.2	Plans exploratoires et critères de comparaison de plans	153
5.2.1	Plans maximin et minimax	153
5.2.2	Plans à entropie maximale	155
5.2.3	Critères de comparaison de plans d'expériences	157
5.3	Génération des plans exploratoires et comparaison	159
5.3.1	Plan à entropie maximale et maximin	159

5.3.2	Génération du plan d'expériences final	163
5.4	Ouvertures/Perspectives	165
Conclusions et perspectives		167
A Méthode de lissage des courbes expérimentales		171
A.1	Filtre moyen et de Savitzky-Golay	174
A.1.1	Filtre moyen (FM)	175
A.1.2	Filtre de Savitzky-Golay (FSG)	176
A.2	Lissage des HC	177
A.3	Lissage du CO	179
B Processus de Ornstein-Uhlenbeck conditionné aux bords		183
B.1	Mouvement brownien et pont brownien	183
B.2	Le processus d'Ornstein-Uhlenbeck conditionné aux bords	186
B.2.1	Représentation du processus d'Ornstein-Uhlenbeck	186
B.2.2	Processus d'Ornstein-Uhlenbeck conditionné aux bords	188
C Vieillessement de la carotte		191
Bibliographie		193

Liste des tableaux

1.1	Facteurs préexponentiels et énergies d'activation des constantes d'adsorption	27
1.2	Plages de variation des variables "espèces" et du débit massique	28
1.3	Numérotation et composition de HC , CO et O_2 des 20 essais réalisés sur BGS	33
2.1	Valeurs des paramètres et des ERMSPE de chaque cas	58
4.1	Récapitulatif des valeurs des ISE de la première et second étape sur les essais présentés	124
4.2	Récapitulatif des méthodes d'estimation des paramètres cinétiques et des résultats	139
5.1	Critères pour la comparaison des plans pour le plan maximin généré	161
5.2	Critères pour la comparaison des plans pour le plan à entropie maximale généré	161
5.3	Numérotation, composition de HC , CO et O_2 et richesse des 20 essais réalisés sur BGS	163

Table des figures

1.1	Informations relatives aux polluants, Pinchon [128]	22
1.2	Localisation du piège à NOx	23
1.3	Constitution d'un piège à NOx	23
1.4	Dispositif expérimental : Banc Gaz Synthétique (BGS)	24
1.5	Photo du BGS d'IFP Energies nouvelles, sur lequel les expériences ont été réalisées	24
1.6	Fonctionnement échange matériaux/gaz	25
1.7	Schéma de notre problème : modèle cinétique	28
1.8	Sorties expérimentales et lissées de l'essai 1	30
1.9	Schéma de notre problème : modèle numérique	32
1.10	Représentation du plan d'expériences dans le domaine	33
1.11	Sorties expérimentales lissées, en rouge, et du modèle numérique, en noir, de l'essai 2	34
1.12	Résidus de CO, en pointillés, et HC, en trait plein, entre les résultats expérimentaux et du modèle numérique pour l'essai 2	35
1.13	Simulation du système (1.20) en traits continus, et du système (1.22) en pointillés, pour 3 concentrations initiales différentes en fonction du temps, pour T=20°C	38
1.14	Simulation du système (1.20) en traits continus, et du système (1.22) en pointillés, pour 3 concentrations initiales différentes en fonction de la température, pour t=80sec	39
2.1	Surface de réponse du modèle $y(x_1, x_2) = 2x_1^3x_2^2$ sur $[-1, 1]^2$ et prédiction par krigeage universel avec corrélation gaussienne anisotropique à partir de 14 points d'expériences (points noirs)	48
2.2	Simulations d'un noyau de covariance non stationnaire en 1D	51
2.3	Simulations d'un noyau de covariance non stationnaire en 2D	52
2.4	Prédiction par krigeage universel conditionné par une et deux dérivées partielles à partir de 14 points d'expériences (points noirs)	57
3.1	Résultats du krigeage tendance linéaire et non-linéaire pour un temps de 100 secondes	83
3.2	Résultats de prédiction du krigeage tendance linéaire et non-linéaire pour un temps de 80 secondes et $A_0 = 2.5$	83
3.3	Résultats de prédiction pour un essai inclus dans le domaine d'apprentissage	85
3.4	Valeurs des EMQ obtenues pour chaque jeu de données à disposition selon chacune des deux sorties par krigeage à tendance non-linéaire	87
3.5	Résultats de la prédiction par krigeage à tendance non-linéaire pour la sortie CO : pire et meilleur cas au sens de l'EMQ	88
3.6	Résultats de la prédiction par krigeage à tendance non-linéaire pour la sortie HC : pire et meilleur cas au sens de l'EMQ	88
3.7	Prédictions du cas maîtrisé par krigeage à tendance non-linéaire sans dérivée (courbe pointillée rouge), et avec dérivée (courbe pleine rouge) comparées aux résultats expérimentaux (courbe pleine bleue)	92
3.8	Prédiction de la dérivée pour le cas maîtrisé (courbe rouge) comparées à la dérivée expérimentale (courbe bleue)	92

3.9	Résultats de prédiction pour un essai inclus dans le domaine d'apprentissage	93
3.10	Valeurs des EMQ obtenues pour chaque jeu de données à disposition selon chacune des deux sorties par krigeage à tendance non-linéaire conditionné par la dérivée	94
3.11	Résultats de la prédiction par krigeage non-linéaire conditionné par la dérivée pour les deux essais exposés dans la section 3.1.4.2	95
3.12	Contre-exemple du bon comportement du KNLD sur notre cas d'étude	96
3.13	Représentations des sorties CO et HC en avancement en fonction de la température pour un essai (essai 11)	98
3.14	Représentations des dérivées de l'avancement des sorties CO et HC par rapport à la températures en fonction de l'avancement pour un essai (essai 11)	99
3.15	Représentations des résidus des dérivées entre résultats expérimentaux et numériques pour les 20 essais	99
3.16	Prédictions des dérivées en température selon l'avancement des sorties CO et HC de l'essai 11	101
3.17	Prédictions des dérivées en température selon l'avancement des sorties CO et HC pour les meilleurs cas après intégration présentée dans la section 3.3.3	102
3.18	Prédictions après intégration des dérivées pour l'essai 14 et apparition du décalage	103
3.19	TLO de CO et HC en fonction de la richesse du mélange et prédiction par krigeage ordinaire avec effet pépite	104
3.20	Prédictions après intégration des dérivées et correction du décalage pour l'essai 14	104
3.21	Comparaison des valeurs des EMQ obtenues pour chaque jeu de données à disposition selon chacune des deux sorties après intégration de la dérivée prédite (points bleus) avec les EMQ calculées par KNLD (points rouges)	105
3.22	Prédictions après intégration des dérivées et correction du décalage pour les pires cas au sens de l'EQM	106
3.23	Intervalles de prédiction des dérivées par krigeage à tendance non linéaire (essai 14)	106
3.24	Limite inférieure de l'intervalle de prédiction des dérivées intégrées par le schéma d'Euler (cf. section 3.3.3)	107
3.25	Limite inférieure de l'intervalle de prédiction après intégration des dérivées par le schéma d'Euler et correction du décalage par KU avec pépite	108
4.1	Nuage de points de la température de light-off pour CO en fonction de la richesse	115
4.2	Nuages de points de la sortie CO normalisée selon la richesse pour différentes températures	117
4.3	Nuages de points de la sortie CO normalisée selon $[CO]_0$ pour différentes températures	117
4.4	Nuages de points de la sortie CO normalisée selon $[HC]_0$ pour différentes températures	117
4.5	Nuages de points de la sortie HC normalisée selon la richesse pour différentes températures	117
4.6	Nuages de points de la sortie HC normalisée selon $[CO]_0$ pour différentes températures	117
4.7	Nuages de points de la sortie HC normalisée selon $[HC]_0$ pour différentes températures	117
4.8	Représentation des fonctions s_i du modèle GAM 4.5 pour CO à différentes températures	118
4.9	Représentation des fonctions s_i du modèle GAM 4.5 pour HC à différentes températures	119
4.10	Fonctions coefficients du modèle 4.6 pour la sortie CO et HC	121
4.11	Comparaison des résultats expérimentaux (courbes bleues) et prédits (courbes rouges) avec le modèle 4.6, pour la sortie CO	121
4.12	Comparaison des résultats expérimentaux (courbes bleues) et prédits (courbes rouges) avec le modèle 4.6, pour la sortie HC	122
4.13	Résidus de prédiction pour les deux sorties et pour chaque expérience	122
4.14	Fonctions coefficients du modèle 4.6 lissés par P-splines pour la sortie CO et HC	123
4.15	Comparaison des résultats expérimentaux (courbes bleues) et prédits (courbes rouges) avec le modèle 4.6 et lissage des fonctions coefficients par P-splines, pour la sortie CO	123
4.16	Comparaison des résultats expérimentaux (courbes bleues) et prédits (courbes rouges) avec le modèle 4.6 et lissage des fonctions coefficients par P-splines, pour la sortie HC	124

4.17	Fonctions coefficients pour la sortie <i>CO</i> et <i>HC</i> estimées par la seconde méthode en traits pleins, et celles de la première méthode en pointillée	129
4.18	Comparaison des résultats expérimentaux (courbes bleues) et prédits (courbes rouges) avec le modèle 4.6, pour la sortie <i>CO</i>	129
4.19	Comparaison des résultats expérimentaux (courbes bleues) et prédits (courbes rouges) avec le modèle 4.6, pour la sortie <i>HC</i>	130
4.20	Comparaison des résultats expérimentaux (courbes bleues) et prédits (courbes rouges) avec le modèle numérique après estimation des paramètres pour la sortie <i>CO</i> de l'essai 1, et pour la sortie <i>HC</i> de l'essai 20	133
4.21	Répartition des paramètres cinétiques relatifs à la réaction d'oxydation de <i>CO</i> fonction des variables normalisées	134
4.22	Répartition des paramètres cinétiques relatifs à la réaction d'oxydation de <i>HC</i> et de la constante d'adsorption de <i>HC</i> fonction des variables normalisées	135
4.23	Comparaison des résultats expérimentaux (courbes bleues) et prédits (courbes rouges) avec le modèle numérique après estimation des coefficients de la relation linéaire des paramètres par rapport aux variables de l'expérience pour la sortie <i>CO</i> de l'essai 1, et pour la sortie <i>HC</i> de l'essai 20	137
4.24	Résidus de prédiction pour les deux sorties et pour chaque expérience	138
4.25	Comparaison des résultats expérimentaux (courbes bleues) et prédits (courbes rouges) avec le modèle numérique après estimation par KRR des paramètres cinétiques fonction des variables de l'expérience pour la sortie <i>CO</i> de l'essai 1, et pour la sortie <i>HC</i> de l'essai 20	139
4.26	Résidus entre résultats expérimentaux et prédits	143
4.27	Région de confiance de la prédiction relative à l'essai 14	144
5.1	Représentation du plan d'expériences dans l'espace des entrées deux à deux	152
5.2	Représentation d'une suite de $n=30$ points en dimension $d=2$ et du pavé $P = [0, t_1[\times [0, t_2[$	158
5.3	Représentation du plan d'expériences généré selon un critère maximin. Les points rouges correspondent au plan initial, et les points noirs au plan généré	160
5.4	Critère <i>DisL2</i> (courbe pleine bleue) et <i>Rec</i> (courbe pointillée verte) en fonction de la portée	161
5.5	Représentation du plan d'expériences généré par maximum d'entropie. Les points rouges correspondent au plan initial, et les points noirs au plan généré	162
5.6	Représentation du plan d'expériences dans l'espace des entrées	164
5.7	Relation entre TLO et richesse et vérification du bon comportement du nouvel échantillon	165
A.1	Résultats expérimentaux de l'essai 1 pour <i>CO</i> et <i>HC</i>	173
A.2	Illustration du filtrage d'une valeur par la méthode de Savitzky-Golay	176
A.3	Étape 1 du lissage du résultat expérimental pour <i>HC</i>	177
A.4	Étape 2 du lissage du résultat expérimental pour <i>HC</i>	178
A.5	Étape 3 du lissage du résultat expérimental pour <i>HC</i>	178
A.6	Résultat final du lissage du résultat expérimental pour <i>HC</i>	178
A.7	Trois exemples de lissage des résultats expérimentaux	179
A.8	Étape 1 du lissage du résultat expérimental pour <i>CO</i>	180
A.9	Étape 2 du lissage du résultat expérimental pour <i>CO</i>	180
A.10	Détermination des quatre paliers	181
A.11	Courbes discontinues et résultat final	182
A.12	Trois exemples de lissage des résultats expérimentaux	182
B.1	Représentations des résidus des dérivées entre résultats expérimentaux et numériques pour les 20 essais	183
B.2	Réalisation d'un mouvement brownien	184

B.3 Réalisation d'un pont brownien	186
C.1 Comparaison des essais de contrôle à l'essai de référence	191



Introduction

Contexte

Les phénomènes physico-chimiques dans les systèmes de dépollution catalytique sont représentés par des modèles cinétiques qui dépendent de paramètres (facteurs préexponentiels, énergies d'activation, constantes d'adsorption, et des chaleurs d'adsorption) qu'on ne peut déduire de considérations théoriques. Pour les déterminer, des expériences sont réalisées dans des installations pilotes. Le résultat de ces expériences permet d'ajuster les modèles cinétiques puis d'utiliser les modèles pour :

1. améliorer le rendement des procédés, par exemple pour les systèmes de dépollution, de traiter plus de polluants et plus efficacement avec les mêmes proportions de catalyseur, en étudiant, à l'aide de ces modèles, un grand nombre de conditions expérimentales, et obtenir ainsi une bonne compréhension du système (utilisation en prédiction),
2. améliorer la qualité des produits, notamment l'allongement de la durée de vie des systèmes, les modèles étant utilisés pour des études de survie, de comportement dans le temps (dégradation) mais également pour des études d'incertitude et de sûreté (analyse de risque),
3. réduire le temps de développement et les coûts de manière générale, les expériences à réaliser étant généralement longues et très coûteuses, cet aspect est primordial pour mener à bien le développement de nouveaux systèmes dans un contexte où les délais (et bien souvent les budgets) sont de plus en plus restreints.

Les plans d'expériences apparaissent ainsi essentiels afin de guider les chimistes dans cette étape primordiale, pour des raisons de temps, économiques et écologiques. Pour le moment, ces expériences sont réalisées à partir de connaissances métiers, et donc, le plus souvent, sans stratégie pertinente. Or, le nombre croissant de paramètres à ajuster, sur des domaines de plus en plus importants, ne permet plus de se satisfaire de cette façon de procéder.

Les concepts de plan d'expériences ont été introduits au début du XX^{ème} en Angleterre par Ronald A. Fischer, [61], dans le secteur agricole puis dans l'industrie et dans le domaine biologique et médical. Les plans d'expériences permettent d'organiser au mieux les essais qui accompagnent une recherche scientifique ou des études industrielles et correspondent à des suites d'essais rigoureusement organisés.

Depuis cette époque, les plans d'expériences n'ont eu de cesse de prendre une importance de plus en plus grande, pour faire, aujourd'hui, de cet outil un élément incontournable dans de nombreux domaines, notamment en cinétique. Les principales raisons de cet engouement sont économique, écologique, de temps et de faisabilité de développement de système. L'essor de ce concept va de pair avec les moyens d'acquisition et de traitement des données, et la complexité des problèmes étudiés, qui bien souvent obligent à redéfinir des théories plus adaptées à ces problèmes.

En cinétique, le but des plans d'expériences est d'accompagner les expérimentateurs dans leur démarche de mise au point des modèles et les aider à choisir les expériences à réaliser qui répondront au mieux aux objectifs fixés.

Trois situations peuvent survenir quand le modèle est posé a priori :

1. Si aucun point expérimental n'a encore été effectué, il s'agit de proposer une stratégie expérimentale complète.
2. Si des essais ont déjà été menés, on cherche les quelques points complémentaires susceptibles d'étendre l'application du modèle à de nouvelles conditions de fonctionnement (par exemple, un changement de formulation du catalyseur).
3. Si d'autres points que ceux proposés par la méthode des plans d'expériences doivent finalement être effectués parce que l'ingénieur responsable en a décidé ainsi, il s'agit de quantifier cette différence de démarche par un critère pertinent.

La planification d'expériences correspond au choix des combinaisons des valeurs des facteurs d'entrée, les points ou essais du domaine expérimental auxquels on va effectuer les mesures. Un choix judicieux s'avère souvent important en pratique, car le nombre d'observations est fréquemment assez réduit pour des raisons de coût ou de durée de mise en œuvre des expériences.

Quand le modèle est non linéaire ou que le domaine expérimental n'est pas d'une géométrie simple (des contraintes lient les variables expérimentales entre elles par exemple), on utilise des plans d'expériences 'optimaux', dont la construction est par essence itérative. On part des points expérimentaux déjà effectués et on retient parmi l'ensemble des points expérimentaux possibles celui qui quand on le rajoute, optimise un certain critère mathématique qui caractérise la qualité du plan. On incorpore le point trouvé aux points déjà effectués, puis on recommence avec un nouveau point, jusqu'à proposer le nombre de nouvelles expériences décidé a priori avec l'utilisateur. L'algorithme de Fedorov, [59], à la base de cette méthode, permet aussi de concevoir un plan initial, sans démarche itérative.

Toute la démarche repose donc sur la pertinence du critère d'optimalité retenu. Il en existe un grand nombre. Usuellement, parce qu'ils sont plus simples à traiter mathématiquement que d'autres, on a recours aux plans d'expériences D-optimaux. Ceux-ci ont pour défaut de proposer des points aux bornes du domaine expérimental, ce qui conduit à des essais inhabituels, hors normes, voire même délicats.

L'intérêt de cette thèse est donc de proposer des plans d'expériences ayant un sens pour l'utilisateur c'est-à-dire dont les expériences soient pertinentes, adaptés aux problèmes cinétiques rencontrés, du point de vue des modèles, et du point de vue de l'expérimentation. Les modèles comprennent quelques entrées, moins de vingt paramètres et plusieurs sorties. Ils se calculent facilement. L'expérimentation est partiellement automatisée et se déroule en quelques heures.

Phénomènes étudiés

Les modèles mathématiques représentant des systèmes physiques sont des outils importants dans de nombreux domaines de recherche scientifique. Néanmoins, le besoin croissant d'une meilleure compréhension de ces systèmes et l'augmentation de la précision voulue augmentent la complexité des modèles. C'est dans ce cadre que notre étude se situe.

La réduction "à la source" des émissions polluantes du moteur Diesel passe par un compromis NO_x/particules de suies (les oxydes d'azote, NO_x, sont des polluants de l'air toxiques à faible dose et les particules est une appellation pour les fines particules de suie issues de la combustion contenues dans les gaz et réputées cancérigènes). L'orientation prise par les constructeurs automobiles pour atteindre les niveaux réglementaires des NO_x est un ajustement de la combustion. Ces réglages moteur se font au détriment de la production de suies qui sont alors traitées par la mise en place de filtres à particules (FAP) à l'échappement du moteur. Ce choix principalement

dicté par la maturité plus élevée du FAP est généralisé avec l'entrée en vigueur de nouvelles réglementations plus sévères. D'ores et déjà, certains véhicules du segment supérieur doivent se doter également d'un système pour traiter les NOx.

Deux technologies sont identifiées pour l'élimination des NOx à l'échappement d'un moteur à "mélange pauvre", il s'agit du piège à NOx ("*NOx trap*") et de la réduction catalytique sélective à l'urée ("*SCR urée*"). Ces deux systèmes se concurrencent dans la mesure où ils présentent chacun des avantages et des inconvénients. En effet, de nombreuses questions restent en suspens pour pouvoir améliorer leur efficacité et leur durabilité, nécessitant d'avoir une bonne compréhension des phénomènes physico-chimiques se déroulant au sein de ces systèmes catalytiques.

Cette compréhension passe par la mise au point de mécanismes réactionnels, et des modèles cinétiques associés. Pour chacun des systèmes de traitement des NOx, une dizaine de réactions doit être considérée. Les vitesses de réactions mettent en jeu 20 à 30 constantes cinétiques et autres capacités de stockage/adsorption. Ces paramètres sont calibrés en se basant sur une campagne d'essais expérimentaux.

Les essais sont effectués au banc gaz synthétique, un système expérimental reproduisant au mieux les conditions d'utilisations réelles du piège à NOx. L'échantillon est une carotte de catalyseur (1 pouce de diamètre, 2 pouces de long), placée dans un réacteur en quartz, lui même monté dans un four. L'échantillon est traversé par un flux de gaz de composition, débit et composition maîtrisés, et choisit de façon à représenter les conditions existant à l'échappement d'un moteur Diesel. Les espèces ayant une influence prépondérante sur l'efficacité de la déNOx sont le monoxyde d'azote *NO*, le dioxyde d'azote *NO₂*, l'oxygène, et l'ammoniac pour la SCR. Des interactions et des inhibitions ont lieu en fonction également des teneurs en *CO₂*, en *CO*, et en hydrocarbures imbrûlés (*H_C*).

Le piège à NOx permettant de traiter les polluants est un système aux enjeux économiques et écologiques importants. Ces polluants ont des incidences directes sur notre environnement et notre santé. Par exemple, les oxydes d'azote, NOx, dont il est question dans cette application, sont responsables des pics de pollution aux abords des grandes agglomérations ce qui entraîne, notamment, des problèmes respiratoires et oculaires. Le piège à NOx traite également le monoxyde de carbone et les hydrocarbures imbrûlés, deux espèces également nocives pour la santé.

Mais au-delà de ce système choisi comme support de cette thèse, l'enjeu est de pouvoir fournir une démarche méthodologique permettant l'analyse et l'exploitation des modèles numériques utilisés en cinétique pour représenter les systèmes complexes réels. Les modèles numériques de simulation sont très courants. Ces modèles dépendent généralement de paramètres cinétiques que l'on ne peut déduire de considérations théoriques, et qui sont ajustés sur des résultats expérimentaux. Pour cela, la question de la planification expérimentale est incontournable afin de mener à bien les projets de création de ces systèmes de post-traitement, dans un temps raisonnable et financièrement viable.

L'objectif de cette thèse étant d'apporter des briques méthodologiques innovantes pour ajuster les modèles cinétiques, nous avons basé notre travail sur le piège à NOx, mais également sur un cas plus simple, proche et plus léger qui nous a servi à élaborer notre démarche, comme outil de validation des méthodes. Ce second cas consiste en l'étude de l'évolution d'une espèce chimique en présence d'un catalyseur. L'avantage de ce dernier est que nous le maîtrisons totalement permettant ainsi une analyse pertinente des solutions proposées.

Axes de recherches

Il est courant lors d'études expérimentales, que le modèle posé pour représenter un système physique ou chimique ne convienne que partiellement. C'est le cas de notre système. Ce premier modèle est fortement non linéaire, et peut s'exprimer par $\mathbf{y} = f(\mathbf{x}, T, \boldsymbol{\beta})$, où \mathbf{y} est la réponse étudiée (par exemple la teneur en hydrocarbures imbrûlés), \mathbf{x} les conditions expérimentales, T la température et $\boldsymbol{\beta}$ les paramètres cinétiques du modèle noté f .

Pour ajuster ce modèle aux n résultats expérimentaux disponibles et estimer les paramètres du vecteur, la première idée est de déterminer par moindres carrés un vecteur de paramètres globaux, valables quelle que soit l'expérience considérée.

$$\hat{\boldsymbol{\beta}} = \arg \min_{\boldsymbol{\beta}} \sum_{i=1}^n \int \|y^i - f(x^i, T, \boldsymbol{\beta})\|^2 dT$$

où y^i correspond aux résultats expérimentaux de l'expérience i , et x^i à ces conditions opératoires.

Malheureusement, tous les essais basés sur cette méthode se sont révélés infructueux, montrant que le modèle proposé est insuffisant pour représenter toute la réalité expérimentale, même si les grandes tendances sont respectées. La confrontation aux premiers résultats expérimentaux a montré son inadéquation.

La première étape est donc d'obtenir une modélisation de notre phénomène qui soit proche des résultats expérimentaux. L'objectif alors proposé est de déterminer une modélisation qui puisse représenter le comportement de notre système de dépollution. Cette première étape qui peut apparaître comme triviale, est loin de l'être, et constitue la majeure partie de nos travaux. Deux raisons font que cette étape est difficile. Premièrement, sa forte non linéarité et deuxièmement, le fait que la modélisation est soumise à des contraintes cinétiques qu'il est difficilement possible d'incorporer dans les modélisations classiques. La deuxième étape est de tenir compte de la modélisation ainsi établie pour proposer une démarche de planification expérimentale appropriée.

Une difficulté supplémentaire à notre problème s'ajoute, liée à la complexité des objets manipulés, qui sont des données fonctionnelles en sortie du simulateur. La modélisation et le traitement statistique des données ouvrent un champ très vaste de recherche visant des aspects théoriques et appliqués.

Afin d'aborder le problème, nous sommes restés, dans un premier temps, sur une modélisation d'une sortie scalaire, le temps étant considéré comme un facteur d'entrée supplémentaire. Au final, trois axes de recherche ont été explorés. Tout d'abord une étude du krigeage, métamodèle le plus utilisé pour représenter la sortie scalaire d'un code de calcul, en l'adaptant au contexte d'une tendance non linéaire. La difficulté dans ce cas étant l'estimation des paramètres et le calcul des inférences liées à cette démarche. Ensuite, nous avons pris en compte la dimension fonctionnelle, avec l'application et l'extension des modèles à coefficients variables. L'enjeu est de répondre pleinement au cadre d'application. Enfin nous avons abordé le problème des plans d'expériences, avec l'application d'un critère de maximum d'entropie, pour mener une campagne expérimentale.

Organisation du manuscrit

Nous proposons dans ce manuscrit plusieurs approches novatrices pour les problèmes de cinétique, qui dans certains cas ont montré leurs limites sur notre cas d'étude, mais qui sont tout à fait envisageables sur d'autres systèmes. Ces méthodes de modélisation sont réparties en deux catégories, les approches non fonctionnelles et les approches fonctionnelles.

Nous avons essayé d'adopter un discours pédagogique, ce manuscrit étant destiné aux expérimentateurs. Chacune

des méthodes est appliquée sur notre cas d'étude, les conditions d'applications sont détaillées ainsi que les considérations pratiques. Les résultats sont commentés en essayant à chaque fois de se rapporter aux conditions expérimentales et en faisant le parallèle avec le côté applicatif pour guider les cinéticiens dans leur démarche. Nous avons dans cette optique essayé de faire en sorte que ce manuscrit soit le plus complet possible sur le détail des méthodes, accessible aux expérimentateurs, tout en gardant un formalisme mathématique rigoureux. Certaines méthodes trop complexes ne sont pas totalement décrites, néanmoins, nous fournissons un grand nombre de références auxquelles les lecteurs désirant approfondir le sujet pourront se référer. Nous insistons donc sur le fait que les travaux présentés s'inscrivent dans le cadre d'une thèse appliquée, ceci justifiant l'organisation de ce manuscrit, qui comporte cinq chapitres.

Le **chapitre 1** est consacré à la présentation des deux phénomènes étudiés. Tout d'abord notre cas d'application est grandement détaillé en trois temps : le fonctionnement du piège à NO_x, le modèle cinétique et le modèle numérique. Les résultats de ce dernier sont enfin confrontés aux résultats expérimentaux afin de mettre en exergue son inadéquation. Au cours de cette thèse, nous nous sommes penchés sur les raisons de cette inadéquation et avons apporté des modifications par rapport au modèle numérique initial. Ces modifications sont introduites dans ce chapitre, mais il est important de souligner qu'elles n'interviennent qu'à partir du quatrième chapitre. Dans un second temps, nous présentons le deuxième cas utilisé, construit de manière à être proche du cas d'étude, mais totalement maîtrisé. La raison d'être de ce second cas est de pouvoir tester et vérifier les diverses solutions envisagées pour répondre à notre problématique, étape préalable à l'application sur notre cas d'étude, pour les méthodes non fonctionnelles. L'**annexe A**, relative à ce chapitre, présente les méthodes de lissage des données expérimentales mises en place.

Le **chapitre 2** est construit comme une "boîte à outils" des méthodes utilisées. Les diverses approches envisagées sont soit des extensions de méthodes classiques soit des utilisations particulières. Nous avons donc décidé de faire de ce second chapitre une vitrine de ces méthodes, à laquelle nous nous référons dans la suite. Cela permet de ne pas surcharger les chapitres suivants, en ne détaillant que les extensions opérées et les conditions particulières d'application. Ce chapitre se compose de deux parties, la première concerne les méthodes utilisées dans le cadre non fonctionnel avec notamment la théorie de krigeage et de cokrigeage, la seconde concerne les méthodes fonctionnelles, avec notamment l'introduction des données fonctionnelles et des outils classiques dans ce cadre.

Le premier type d'approche, approche non fonctionnelle, est abordé dans le **chapitre 3**, où nous détaillons trois méthodes de modélisation, toutes trois basées sur la théorie de krigeage. La première technique est une extension de cette théorie au cas d'une tendance non linéaire, la seconde propose une amélioration de cette dernière par un conditionnement par la dérivée, la troisième propose une modélisation directe de la dérivée en fonction de l'avancement. L'originalité de cette dernière provient de l'équivalence entre pont brownien et processus de Ornstein-Uhlenbeck conditionné aux bords, équivalence exhibée dans l'**annexe B**. Nous soulignons les limites et avantages de chacune des approches et concluons sur l'utilisation de méthodes non fonctionnelles sur des données fonctionnelles.

L'utilisation des méthodes fonctionnelles est quant à elle présentée dans le **chapitre 4** par le développement de deux types de méthodes. La première est basée sur les modèles à coefficients variables (VCM) et l'omission de l'expertise métier (modèle numérique). La seconde est une méthode analogue aux VCM étendue au cadre non linéaire. Dans cette seconde méthode, nous proposons de modéliser les paramètres cinétiques expérience par expérience par kernel ridge regression. Au final, le développement et l'application d'un bootstrap fonctionnel nous permettent d'estimer une région de confiance pour les courbes prédites, information envisagée pour définir un critère de plan d'expériences.

Enfin, le **chapitre 5** propose d'introduire une démarche de mise en place d'un plan d'expériences exploratoire. Le but principal est de fournir un outil pour les expérimentateurs. Les aspects applicatifs et pédagogiques

sont privilégiés et certaines considérations théoriques ne sont pas donc pas abordées. Au final, nous montrons l'intérêt des plans à entropie maximale et appuyons sur les perspectives d'application pour de futurs essais, et sa complémentarité avec la méthode de planification envisagée dans le chapitre précédent.

Phénomènes étudiés

1.1	Piège à NOx	22
1.1.1	Fonctionnement du piège à NOx	22
1.1.2	Modèle cinétique	25
1.1.3	Modèle numérique	30
1.1.4	Comparaison résultats expérimentaux et résultats numériques	34
1.1.5	Modification du modèle numérique	35
1.2	Cas maîtrisé	37
1.2.1	Présentation du cas maîtrisé et motivations	37
1.2.2	Simulations	38
1.2.3	Pertinence du cas maîtrisé	39
1.3	En bref	40

Phénomènes étudiés

Les modèles mathématiques représentant des systèmes physiques sont des outils importants dans de nombreux domaines de recherche scientifique. Néanmoins, le besoin croissant d'une meilleure compréhension de ces systèmes et l'augmentation de la précision voulue augmentent la complexité des modèles, qui dans certains cas ne sont pas cependant suffisants pour répondre aux attentes des expérimentateurs. C'est dans ce cadre que notre étude se situe.

Notre travail consiste en l'étude d'un système de post-traitement catalytique à la sortie du moteur Diesel : le piège à NOx. Ce système a pour but de traiter les polluants. Il s'agit là d'une application aux enjeux économiques et écologiques importants. En effet, ces polluants ont des incidences directes sur notre environnement et notre santé. Par exemple, les oxydes d'azote, NOx, dont il est question dans cette application sont responsables des pics de pollution aux abords des grandes agglomérations ce qui entraîne, notamment, des problèmes respiratoires et oculaires. En outre, le piège à NOx traite également le monoxyde de carbone et les hydrocarbures imbrûlés, deux espèces également nocives pour la santé.

Mais au-delà de ce cas d'application, l'enjeu est de pouvoir fournir une méthodologie de planification d'expériences appliquée aux problèmes rencontrés dans le domaine de la cinétique. En effet, en remplacement du système réel, le recours à des modèles numériques de simulation est très courant. Ces modèles dépendent généralement de paramètres cinétiques que l'on ne peut déduire de considérations théoriques, et qui sont ajustés sur des résultats expérimentaux. Pour cela, la planification d'expériences, qui permet de choisir les meilleures expériences à réaliser en diminuant leur nombre, devient un outil incontournable afin de mener à bien les projets de création de ces systèmes de post-traitement, dans un temps raisonnable et financièrement viable.

Néanmoins, les modèles censés représenter la réalité expérimentale ne sont pas toujours assez pertinents pour appliquer des méthodes de planification d'expériences classiques, comme dans notre cas d'étude où le modèle est inadéquat. De plus, il s'agit d'un cas très complexe. Pour cette raison, nous avons construit un second cas plus simple et pourtant similaire, nommé cas maîtrisé. Ce dernier nous a servi d'outil de validation des méthodes que nous présentons dans ce manuscrit. Ce second cas consiste en l'étude de l'évolution d'une espèce chimique en présence d'un catalyseur, et a été construit de manière à être proche dans l'esprit de notre cas d'étude. L'avantage de ce dernier est que nous le maîtrisons totalement, ce qui n'est pas le cas de notre cas d'étude, permettant ainsi une analyse pertinente des solutions proposées.

Ce chapitre se compose donc en deux parties, chacune présentant un des phénomènes étudiés. Nous commençons tout d'abord par notre cas d'étude, le piège à NOx, avec une présentation en quatre temps : fonctionnement du piège, modèle physico-chimique, modèle numérique, et nous comparons les résultats expérimentaux et numériques. Dans la deuxième partie, nous introduisons le cas maîtrisé avec un descriptif de la construction de ce cas, puis des exemples de simulations pour, au final, faire l'analogie entre les deux cas étudiés.

1.1 Piège à NOx

De nos jours, le moteur à combustion interne ne peut plus être envisagé sans son système de post-traitement et les stratégies de contrôle associées. Les dispositifs de post-traitement sont adaptés au type de moteur, diesel dans notre cas. Les catalyseurs d'oxydation diesel (DOC) sont une technologie clé pour les moteurs diesel. Ils transforment le *CO* et les *HC* et diminuent également la masse de particules diesel en oxydant certains hydrocarbures.

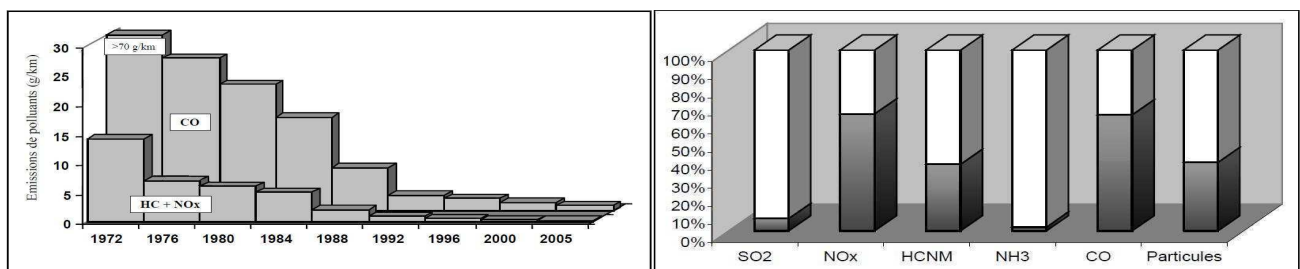
IFP Energies nouvelles développe des modèles de simulation de post-traitement, et on s'intéresse dans notre cas, au modèle du piège à NOx qui traite ces gaz toxiques à la sortie du moteur par des réactions catalytiques, en plus des NOx.

Dans un premier temps, nous présentons le principe général de fonctionnement du piège à NOx ainsi que les raisons d'être d'un tel système et nous détaillons également le protocole expérimental mis en place. Puis, dans un deuxième temps, nous exhibons le modèle physique, en introduisant le principe d'échange matériaux/gaz et le modèle cinétique retenu. Ceci nous permet de présenter plus précisément le fonctionnement du piège à NOx, les entrées et sorties et les paramètres qui nous intéressent. Enfin, dans une troisième partie, nous mettons en avant le modèle numérique utilisé, construit à partir du modèle cinétique, et introduisons le plan d'expériences à notre disposition. Les résultats de ce dernier nous permettent de confronter les résultats expérimentaux et numériques et de mettre en avant l'incapacité du modèle à représenter la réalité expérimentale.

1.1.1 Fonctionnement du piège à NOx

Le moteur Diesel émet un taux d'oxydes d'azote NOx plus élevé que le moteur essence en raison de son principe de combustion en excès d'air. En effet, en théorie, la combustion effectuée dans le moteur est complète, ce qui engendre une production du dioxyde de carbone, *CO*₂ et d'eau, *H*₂*O*. Mais en réalité, ce n'est pas le cas, ce qui provoque une émission de gaz toxiques : d'une part la combustion est incomplète d'où l'apparition des hydrocarbures imbrûlés, *HC*, et d'autre part la présence de réactions parasites entraîne la présence en sortie des oxydes d'azotes, *NO*, *NO*₂ que l'on note NOx et du monoxyde de carbone, *CO*.

Malgré les efforts déployés pour réduire à la source les émissions polluantes des moteurs à combustion interne, c'est-à-dire en optimisant le processus de combustion et les carburants, les normes de pollution de plus en plus sévères (cf figure 1.1a) imposent de recourir à des dispositifs de post-traitement des gaz de combustion de plus en plus perfectionnés. Les seuils de rejet de NOx que les normes du futur imposeront, seront inférieurs à ceux que peut garantir un pot catalytique conventionnel. Une des solutions envisagées pour répondre à leurs exigences est de compléter son action par un nouveau filtre baptisé piège à NOx, ou *NOx trap* en anglais.



(a) Évolution des normes anti-pollution européennes

(b) En grisé, contribution des transports aux émissions globales de polluants en Europe (2001)

Figure 1.1 – Informations relatives aux polluants, Pinchon [128]

Le piège à NOx retient les oxydes d'azote et les stocke dans une structure microporeuse, ce qui impose des phases de régénérations régulières. Le piège à NOx remplit le rôle à la fois de catalyseur d'oxydation traditionnel (oxydation HC/CO) et piège à NOx. Contrairement à sa fonction d'oxydation qui est continue, la fonction de piège à NOx est discontinue. Plus de détails quant à son fonctionnement précis sont donnés dans la section 1.1.2.

Pourquoi réduire les NOx ? Il faut tout d'abord rappeler le fait que le transport (véhicules lourds et véhicules légers confondus) représente la première source d'émissions de NOx (environ 70% en Europe en 2001, voir figure 1.1b), il est donc logique que les normes soient sévères car les effets de ce polluant sur la santé et l'environnement sont préoccupants. Les NOx sont responsables de l'augmentation de la teneur en ozone (O₃) dans la basse atmosphère, entraînant des pics de pollution estivaux se manifestant en général aux abords des grandes agglomérations, provoquant des irritations respiratoires et oculaires.

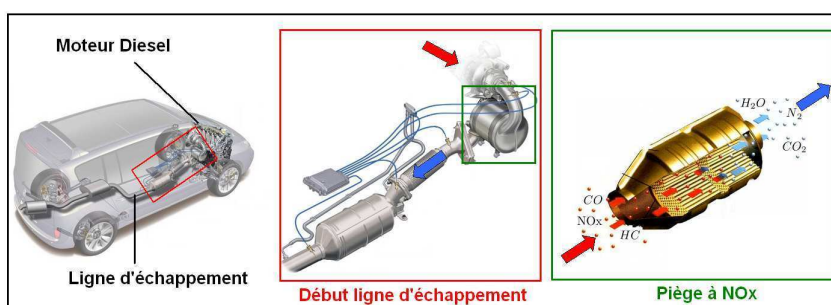


Figure 1.2 – Localisation du piège à NOx

Le piège à NOx est un système de post-traitement catalytique qui se situe dans la ligne d'échappement diesel (voir figure 1.2). Il s'agit d'une structure alvéolaire avec un support poreux imprégné de produits chimiques. Son principe consiste à faire passer les gaz d'échappement à travers cette structure, dans laquelle une réaction de catalyse se produit au contact de métaux précieux et transforme ainsi les polluants. Plus précisément, il s'agit d'adsorber puis d'oxyder des polluants grâce à l'action catalytique de métaux précieux, généralement du platine (Pt) ou plus rarement du palladium (Pd). Ensuite, les NOx résultant de cette oxydation sont stockés grâce à un autre composant, du baryum (Ba). Les NOx sont ensuite réduits et libérés par l'action d'un troisième métal, le rhodium (Rh), lors des périodes de régénération du filtre. Tout ceci étant contenu dans un *canning*, pour l'insertion dans la ligne d'échappement (voir figure 1.3).

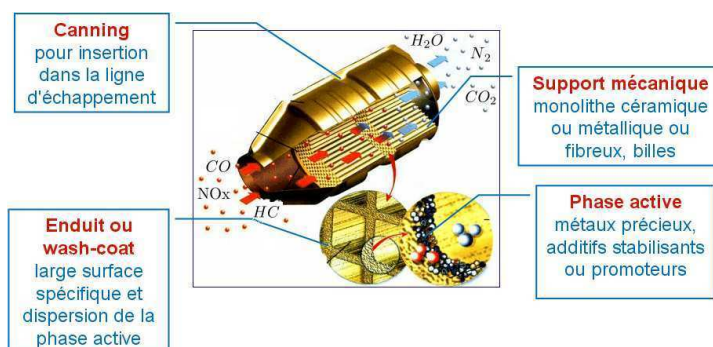


Figure 1.3 – Constitution d'un piège à NOx

Néanmoins, ces filtres ont des problèmes de fiabilité dans le temps et ne peuvent fonctionner de façon durable tant que les carburants contiennent du soufre, car le soufre conduit à l'empoisonnement du piège. De plus une des difficultés majeures est de gérer efficacement le passage entre la phase de stockage et de déstockage du piège, qui, pour cela, nécessite une bonne compréhension du système cinétique c'est-à-dire une bonne modélisation du fonctionnement du piège.

Pour des raisons de simplicité et pour contrôler précisément les paramètres qui entrent en jeu, les expériences effectuées afin d'étudier le comportement du piège à NOx, ne sont pas directement opérées sur un véhicule, mais en laboratoire sur un échantillon de catalyseur. Plus précisément, celles-ci sont menées sur banc gaz synthétique (BGS), un système expérimental reproduisant au mieux les conditions d'utilisations réelles. Chacun des essais correspond à des cas où la composition varie, de manière à couvrir le domaine d'utilisation du piège à NOx. Dans chacune des expériences réalisées, on applique une rampe de température afin de balayer l'intervalle de température dans lequel le piège à NOx fonctionne en aval du moteur. La figure 1.4 est un schéma de fonctionnement du BGS. Une photo du BGS d'IFP Energies nouvelles sur lequel nos expériences ont été effectuées est également donnée (figure 1.5).

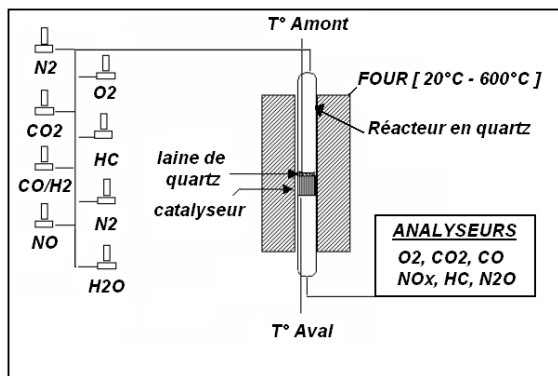


Figure 1.4 – Dispositif expérimental : Banc Gaz Synthétique (BGS)

La réalisation d'une expérience se déroule de la manière suivante :

- le dosage des proportions de chaque espèce que l'on souhaite dans notre gaz s'effectue au moyen de débitmètres placés après les vannes d'arrivées,
- le mélange obtenu est ensuite envoyé dans un réacteur en quartz, contenant, au centre, un échantillon du catalyseur de dimensions 2,5cm de diamètre et 5cm de longueur,
- ce réacteur est placé dans un four reproduisant la chaleur dégagée par le moteur, qui permet la mise en action du piège à NOx, les caractéristiques de chauffe étant choisies (température initiale, évolution de la température),
- le gaz en sortie du réacteur passe par les analyseurs qui mesurent et enregistrent la proportion de chaque espèce.

L'expérience se déroule sur 4000 secondes, et toutes les demi-secondes, nous enregistrons les différentes mesures : composition du HC, du CO et des NOx en sortie, et température amont/aval.



Figure 1.5 – Photo du BGS d'IFP Energies nouvelles, sur lequel les expériences ont été réalisées

On peut voir sur cette photo du BGS, les vannes d'arrivée des espèces que l'on souhaite dans notre gaz en entrée (cercle n°1), les débitmètres contrôlant la proportion de chacune des espèces (cercle n°2), le four contenant le réacteur en quartz (cercle n°3) et enfin les analyseurs et enregistreurs de la composition des gaz en sortie du four (cercle n°4).

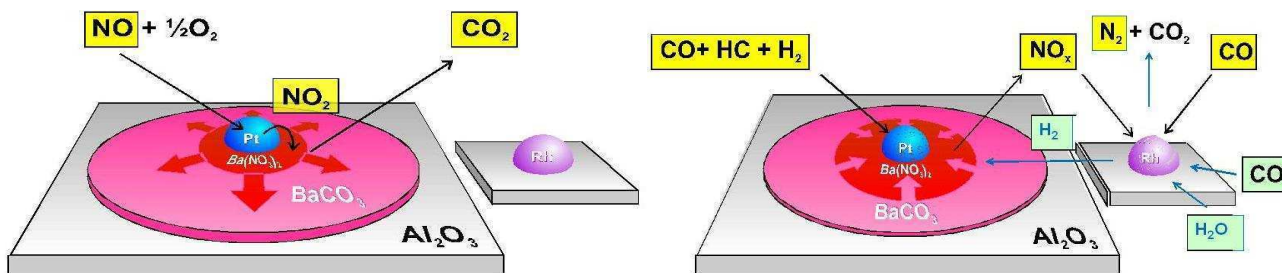
1.1.2 Modèle cinétique

Nous allons maintenant présenter plus en détails le fonctionnement du piège à NOx, en mettant en avant le processus d'échange matériaux/gaz, le modèle cinétique et en développant les entrées, sorties et paramètres de notre système. Le but principal de cette partie est de fournir aux experts métiers une description complète du système afin de replacer le problème dans son contexte.

Comme nous l'avons vu précédemment, le moteur Diesel émet un taux d'oxydes d'azote plus élevé que le moteur essence en raison de son principe de combustion en excès d'air. Leurs traitements demandent un processus chimique élaboré. Celui-ci fonctionne en deux temps explicités ci-dessous. En entrée du post-traitement nous avons les gaz d'échappement du moteur :

1. En mélange pauvre, les effluents du moteur sont en faible quantité relativement à l'oxygène qui est en excès, le piège à NOx est en phase de stockage des NOx. Le CO et les HC sont traités en continu par le piège qui fait aussi office de catalyseur d'oxydation : l'excès d'oxygène permet de les oxyder.
2. Avant que le piège commence à saturer, le système d'injection passe momentanément en dosage riche (manque d'oxygène). L'afflux de réducteurs (CO, HC) qui en résulte vient réagir avec les NOx qui ont été stockés, et les transforme en eau H₂O, azote N₂, gaz carbonique, CO₂. Il faut noter, malgré tout, que les NOx ne sont pas traités en totalité. Une partie a le temps de se déstocker et de sortir du piège avant d'avoir été réduite. Les stratégies de contrôle moteur doivent être optimisées de façon à limiter au maximum ces "bouffées" de NOx.

De manière simple, en fonctionnement nominal, mélange pauvre, le catalyseur piège les NOx mais ne les traite pas. Pour régénérer le catalyseur, le moteur doit fonctionner en mélange riche afin que les hydrocarbures imbrûlés et le monoxyde de carbone alors émis en grande quantité réduisent les NOx stockés. La première phase, nommée fonctionnement en mélange pauvre, a une durée d'environ une minute, et la seconde phase, fonctionnement en mélange riche, est de l'ordre de quelques secondes.

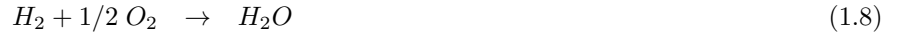
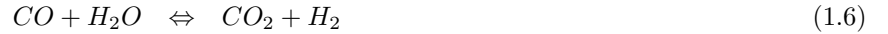
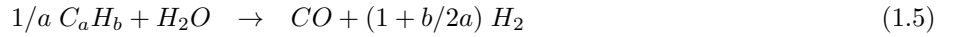
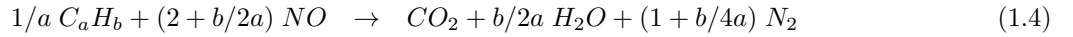
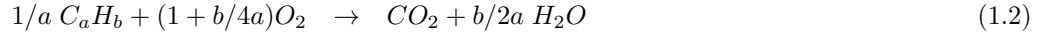


(a) Fonctionnement en mélange pauvre (b) Fonctionnement en mélange riche
Figure 1.6 – Fonctionnement échange matériaux/gaz

Le pain catalytique utilisé est particulier car, en plus du platine et du rhodium destinés à la catalyse, il contient du baryum qui va être capable de réaliser un stockage des NOx sous forme de nitrates de baryum, qui seront extraits et réduits en phase riche. Les figures 1.6a et 1.6b illustrent le fonctionnement de l'échange entre les matériaux et les gaz. Le support utilisé est de l'alumine, Al₂O₃, sur lequel on a déposé une couche d'imprégnation de baryum, ou plus exactement de withérite, BaCO₃, ainsi que des sites de platine et de rhodium. Le platine sur le catalyseur transforme le monoxyde d'azote en dioxyde d'azote (NO₂). L'oxyde de baryum fixe les molécules de NO₂ pour en faire une chaîne Ba(NO₃)₂. Lors de la phase de purge, fonctionnement en mélange riche, les NOx sont désorbés puis réagissent avec les hydrocarbures imbrûlés sur les sites de rhodium. Cette réduction entraîne

la formation d'azote, de CO_2 , de dihydrogène et d'eau. Le piège à NOx est alors nettoyé et prêt à reprendre son rôle de piège à oxydes d'azote.

Les réactions considérées sont les suivantes :



Les variables a et b représentent respectivement le nombre de carbone et d'hydrogène caractérisant notre carburant. Dans notre cas, il s'agit de propylène, soit $a = 3$ et $b = 6$.

Le système étant complexe, seul le fonctionnement du piège en mélange pauvre a été considéré dans nos travaux. Cela permet de réduire le nombre de réactions à prendre en compte, puisque, l'oxygène étant en excès, les réactions d'oxydation du CO , du NO , des HC et du H_2 deviennent prépondérantes, respectivement les équations (1.1), (1.2), (1.7) et (1.8).

Soit R la constante des gaz parfaits, T_s la température du solide et x_i la fraction molaire de l'espèce i . Les vitesses de réactions associées aux huit réactions précédentes sont définies ci-dessous. Ainsi R_i correspond à la vitesse de réaction de la $i^{\text{ème}}$ réaction. L'intérêt de présenter ces vitesses est d'introduire les k_i , que nous détaillons par la suite, équation (1.12), et qui contiennent les paramètres cinétiques étudiés. Pour plus de détails les constantes utilisées et certains paramètres voir Millet [112].

$$R_1 = \frac{k_1 \cdot x_{CO} \cdot x_{O_2}}{F},$$

$$R_2 = \frac{k_2 \cdot a \cdot x_{C_aH_b} \cdot x_{O_2}}{F},$$

$$R_3 = \frac{k_3 \cdot x_{CO} \cdot x_{NO}}{(1 + 53 \cdot \exp(\frac{1200}{T_s}) \cdot x_{CO})^2 \cdot (1 + 53 \cdot \exp(\frac{1500}{T_s}) \cdot x_{NO})^2},$$

$$R_4 = \frac{k_4 \cdot a \cdot x_{C_aH_b} \cdot x_{NO}}{F},$$

$$R_5 = \frac{k_5 \cdot a \cdot x_{C_aH_b} \cdot x_{H_2O}}{F},$$

$$R_6 = k_6 \left(x_{H_2O} \cdot x_{CO} - \frac{x_{CO_2} \cdot x_{H_2}}{\exp(\frac{-x_{gibbs}}{T_s})} \right) / F,$$

$$R_7 = k_7 \cdot \left(x_{NO}^\alpha \cdot x_{O_2}^\beta - \frac{x_{NO_2}^\alpha}{K_{eq}} \right),$$

$$R_8 = \frac{k_8 \cdot x_{H_2} \cdot x_{O_2}}{F}$$

Avec,

$$F = T_s (1 + K_0 x_{CO} + K_1 x_{HC})^2 (1 + K_2 x_{CO}^2 x_{HC}^2) (1 + K_3 x_{NO}^{0.7}) \quad (1.9)$$

$$x_{gibbs} = \left(-41169 + 42.012 \cdot T_s + 6.6 \cdot \left(T_s - 298 - T_s \cdot \log \left(\frac{T_s}{298} \right) \right) \right) / R \quad (1.10)$$

$$K_{eq} = A_{eq} \cdot \exp \left(\frac{E_{a_{eq}}}{R \cdot T_s} \right) \quad (1.11)$$

A_{eq} et Ea_{eq} étant respectivement le facteur préexponentiel et l'énergie d'activation de la constante d'équilibre de NO/NO_2 . Les constantes cinétiques, k_i , et les constantes d'adsorption, K_i suivent une loi d'Arrhenius [7] :

$$k_i = A_i \cdot \exp\left(\frac{-Ea_i}{R \cdot T_s}\right) \text{ en } mol.s^{-1}.m_{cata}^{-2} \quad (1.12)$$

$$K_i = A_i^a \cdot \exp\left(\frac{Ea_i^a}{T_s}\right), \text{ adimensionnelle} \quad (1.13)$$

Les facteurs préexponentiels et énergies d'activation A_i^a et Ea_i^a des constantes d'adsorption, tirés de la littérature, Voltz *et al* [157], sont donnés dans le tableau (1.1) :

	K_0	K_1	K_2	K_3
A^a	65.5	2080.0	3.98	479000
Ea^a	961.0	361.0	11611.0	-3733

Tableau 1.1 – Facteurs préexponentiels et énergies d'activation des constantes d'adsorption

En ce qui concerne les constantes d'adsorption, il faut noter que l'expression F qui les emploie, est empirique et couramment utilisée dans la littérature [26], [109], [110], [137]. Elle est conservée telle quelle, avec les constantes d'adsorption associées.

Néanmoins, nous avons été amenés au cours de cette thèse, à remettre en cause cette expression du dénominateur. La section 1.1.5 apporte un éclairage précis sur les motivations de cette remise en cause, la démarche mise en place, et dans quel contexte elle intervient. Pour des raisons évidentes de clarté du manuscrit, ces modifications sont introduites dans ce premier chapitre, tout en soulignant qu'elles ne sont utilisées qu'à partir du chapitre 4.

Les concentrations de toutes les espèces gazeuses intervenant dans le mécanisme réactionnel, sauf l'eau, peuvent être mesurées. Le taux de réaction molaire ω de chaque espèce est la somme des vitesses des réactions impliquant l'espèce, pondérées par le coefficient stœchiométrique (en $mol/m^3/s$) :

$$\omega_{CaH_b} = -(1/a) \cdot (R_2 + R_4 + R_5) \quad (1.14a)$$

$$\omega_{NO} = -R_3 - (2 + b/2a) \cdot R_4 - R_7 \quad (1.14b)$$

$$\omega_{O_2} = -1/2 \cdot R_1 - (1 + b/4a) \cdot R_2 - 1/2 \cdot R_7 - 1/2 \cdot R_8 \quad (1.14c)$$

$$\omega_{CO} = -R_1 - R_3 + R_5 - R_6 \quad (1.14d)$$

$$\omega_{CO_2} = R_1 + R_2 + R_3 + R_4 + R_6 \quad (1.14e)$$

$$\omega_{H_2O} = b/2a \cdot R_2 + b/2a \cdot R_4 - R_5 - R_6 + R_8 \quad (1.14f)$$

$$\omega_{H_2} = (1 + b/2a) \cdot R_5 + R_6 - R_8 \quad (1.14g)$$

$$\omega_{NO_2} = R_7 \quad (1.14h)$$

$$\omega_{N_2} = 1/2 \cdot R_3 + (1 + b/4a) \cdot R_4 \quad (1.14i)$$

En assimilant le catalyseur à un réacteur parfaitement mélangé, les équations de conservation des espèces et de l'énergie déterminent le système d'équations différentielles gouvernant notre modèle numérique, présenté dans la section suivante.

Entrées/sorties du système physique et paramètres

Entrée : En entrée de notre système, nous avons :

1. La composition massique des 9 espèces chimiques de notre gaz en entrée du catalyseur :

$$C_a H_b \text{ (notée HC), NO, O}_2, CO, CO_2, H_2O, H_2, NO_2 \text{ et N}_2.$$

2. Le débit massique de ce gaz,
3. Le profil de température de ce gaz.

Certaines des espèces présentes dans le gaz, ne sont pas considérées comme des variables du fait de considérations physiques. Entre autres, on sait expérimentalement que H_2O est constant à 4%, ou encore, H_2 est non-mesuré par les analyseurs, mais sa composition en entrée est connue et égale à un tiers de celle de CO (voir la figure 1.7).

Au final, il nous reste 5 variables "espèces". Les plages de variations de chacune, en % massique, ainsi que celle du débit massique Q , sont données dans le tableau 1.2.

	Q (kg/h)	O_2	CO_2	CO	HC	NO
min	50	2.24	4	0.07	0.09	0.025
max	200	16.5	12	0.6	0.4	0.08

Tableau 1.2 – Plages de variation des variables "espèces" et du débit massique

Ces plages de variation sont représentatives de ce qui est observé à l'échappement d'un moteur diesel. Il est tenu compte également des limitations inhérentes au montage expérimental, le BGS, sur lequel le piège à NOx est étudié. En particulier, le NO ne peut pas y être injecté dans des proportions aussi faibles que celles observées en sortie moteur : sa plage de variation a donc été rehaussée (plage de variation banc moteur, entre 50 et 190 ppm de NOx avec $1\text{ppm} = 10^{-4} \%$).

À noter, que l'un des objectifs des expérimentateurs est de pouvoir déterminer précisément la température appelée température de "light-off", TLO, qui correspond à la température à laquelle 50% du polluant considéré est converti. Cette température est importante puisque c'est autour de cette dernière que l'on observe le phénomène que l'on cherche à comprendre et expliquer. Néanmoins, des travaux ont montré (Chédotal [26]) que quelle que soit la caractéristique de chauffe du four ($5^\circ\text{C}/\text{min}$, $10^\circ\text{C}/\text{min}$,...), la TLO reste inchangée. On en déduit qu'il est possible de fixer un profil moyen pour toutes nos expériences. Dans ces conditions, notre profil de température n'est plus une entrée de notre problème, mais une condition opératoire fixe.

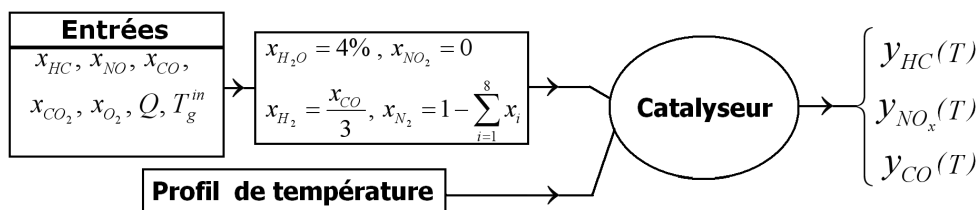


Figure 1.7 – Schéma de notre problème : modèle cinétique

Sorties : Concernant les sorties de notre système, nous en avons trois, mesurées par des analyseurs à la sortie de notre piège à NOx : la composition du gaz où seules les trois espèces suivantes sont considérées : CO , HC , et NOx (NO et NO_2 groupés). Il s'agit donc de courbes cinétiques en fonction de la température voir figures 1.8a, 1.8b et 1.8c. La figure 1.7 fournit une schématisation claire de notre système.

Paramètres : Au regard de la description du modèle cinétique, voici une liste de tous les paramètres qui doivent *a priori* être ajustés. Cette liste initiale est diminuée par des considérations développées dans un second temps :

- Pour chaque réaction, le facteur préexponentiel et l'énergie d'activation de la constante cinétique, soit $8 \times 2 = 16$ paramètres
- Le facteur préexponentiel et l'énergie d'activation des constantes d'adsorption : A^a et Ea^a détaillées précédemment
- Le facteur préexponentiel et l'énergie d'activation de la constante d'équilibre de NO/NO_2 , A_{eq} et Ea_{eq}
- Les exposants α et β présents dans le calcul de R_7
- Les constantes dans le dénominateur du calcul de R_3 , ainsi que la variable x_{gibbs} utilisée dans le calcul de R_6

Soit un total de 32 paramètres. Néanmoins, à l'aide des connaissances métiers du problème, il est possible de réduire ce nombre de paramètres :

- Tout d'abord, dans notre cas (mélange pauvre), seules les équations (1.1), (1.2) et (1.7), correspondant aux réactions d'oxydation du CO , HC et NO , sont prépondérantes. Les facteurs préexponentiels de toutes les réactions (excepté pour ces trois réactions) sont donc fixés à zéro. Ceci entraîne donc 14 paramètres en moins, à savoir les A_i , Ea_i , pour $i = 3, 4, 5, 6, 8$, soit 10 paramètres, la variables x_{gibbs} et les trois constantes présentes dans le dénominateur du calcul de R_3 , sujets à controverse (cf thèse de Matthess [109] et [110])
- Le facteur préexponentiel et l'énergie d'activation des constantes d'adsorption : A^a et Ea^a sont conservés
- De plus, les deux paramètres associés à l'oxydation du NO , A_7 et Ea_7 , peuvent être considérés comme connus. En effet, la réaction (1.7) intervenant après les deux autres réactions d'oxydation, ils ont pu être calibrés indépendamment
- Le facteur préexponentiel et l'énergie d'activation de la constante d'équilibre de NO/NO_2 , A_{eq} et Ea_{eq} , peuvent être obtenus grâce aux données thermodynamiques, en les reliant aux énergies de Gibbs, voir Millet, [112] pour plus de détails. Ils ne sont donc pas à calibrer
- Les exposants de NO et O_2 , à savoir α et β , sont supposés égaux respectivement à 1 et 0.5, de façon à respecter la stœchiométrie de la réaction d'oxydation de NO/NO_2 , et ne sont donc pas à ajuster

Au final, il ne nous reste plus que quatre paramètres (objectif d'un travail déjà effectué, voir El Habachi, [48]) : les facteurs préexponentiels et énergies d'activation des réactions d'oxydation de CO et HC .

Exemple de résultats expérimentaux

Nous nous intéressons donc principalement à la composition des gaz en sortie du piège à NOx, qui sont des sorties fonctionnelles, des courbes en fonction de la température. Ces courbes ont des allures similaires pour chaque expérience.

Pour ce qui est du CO et HC , nous avons un premier palier, nommé palier initial, symbolisant le fait qu'au début de l'essai rien ne se passe, les polluants ne sont pas traités car la température n'est pas assez élevée, puis une décroissance très forte pour le HC et plus lente pour le CO , et enfin un palier final, où tous les polluants ont été convertis. On voit donc tout l'intérêt de la TLO, qui correspond de fait à la température autour de laquelle le système est actif.

En ce qui concerne les NOx, nous avons un palier initial, puis une chute correspond au stockage des NOx et enfin une remontée au dessus du niveau initial due au relargage des NOx, dont le stockage est rendu instable du fait de la température élevée. Les figures 1.8a, 1.8b et 1.8c illustrent l'allure type de chacune des trois espèces. Les courbes rouges correspondent aux résultats expérimentaux lissés.

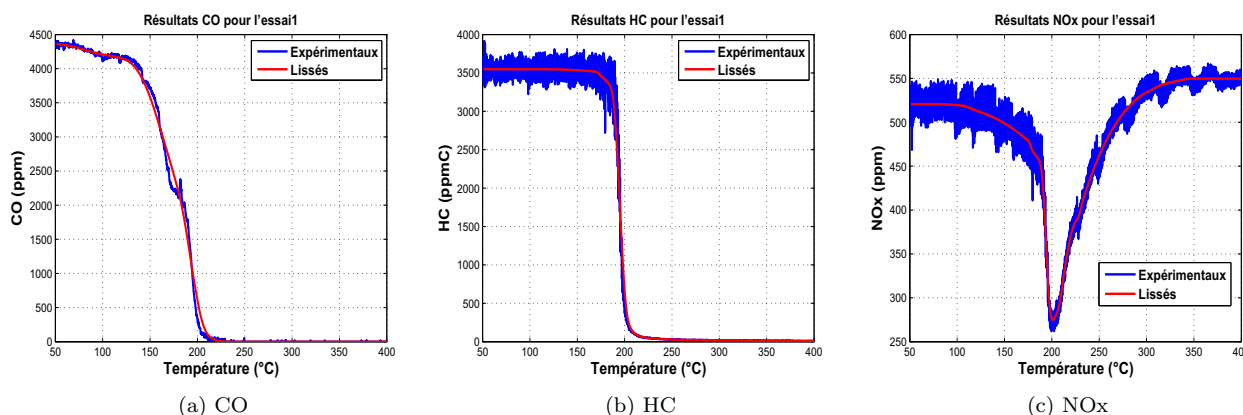


Figure 1.8 – Sorties expérimentales et lissées de l’essai 1

L’obtention des courbes lissées a été effectuée à l’aide d’un filtre adaptatif, voir Haykin [83], que nous avons développé, tenant compte de la forme des courbes. En effet, l’utilisation d’une seule spline de lissage (Wahba [158]) sur tout le domaine n’est pas suffisante du fait de la forte non linéarité, et des oscillations très fortes sur le palier initial. Ainsi, le domaine a été découpé en trois zones.

Pour HC et CO , nous avons le palier initial, la décroissance et le palier final. Pour les NOx , nous observons un palier initial, une décroissance/croissance et un palier final pour les NOx légèrement supérieur au palier initial.

Sur chacune de ces zones, des splines de lissage avec des fenêtres adaptées ont été appliquées (Pintore *et al* [129]). Néanmoins, ce découpage provoquant des cassures au niveau des jointures, un filtre médian (Hamza [10]) a dû être en plus mis en place. À noter que, de plus, des lissages locaux ont été employés, avec détection automatique par changement de valeur de la dérivée, pour contraindre la décroissance monotone pour HC et CO . Le détail complet sur la procédure mise en place est présenté dans l’annexe A, et l’on peut voir sur l’exemple exposé de la validité de notre méthode. En effet, les courbes lissées sont représentatives des résultats obtenus du point de vue des expérimentateurs.

Des méthodes de lissage sous contraintes de monotonie ont également été testées telles que la méthode de Racine [130], Bigot [14] ou Dette [41], donnant des résultats moins satisfaisants.

Dorénavant, les courbes expérimentales brutes ou lissées sont utilisées indistinctement tout au long de ce manuscrit, tout en gardant à l’esprit que ce sont sur ces dernières que les méthodes sont implémentées.

Ce choix est motivé par au moins deux raisons. D’une part du fait que les fortes oscillations sur le palier initial ne nous permettent pas de déterminer son niveau réel qui est pourtant contraint par la composition en entrée, et d’autre part car elles entraînent des problèmes numériques d’optimisation.

1.1.3 Modèle numérique

Le but de cette partie est d’introduire le modèle numérique utilisé pour représenter le système physico-chimique se produisant dans le piège à NOx qui est construit à partir du modèle cinétique. Nous donnons également des précisions quant à son implémentation, les entrées, sorties et paramètres de ce modèle numérique. Enfin, nous exhibons la matrice de plan d’expériences à notre disposition, expériences réalisées sur BGS. Cela nous permet dans la section suivante de comparer les résultats expérimentaux aux résultats numériques.

Système d'équations différentielles

À partir du modèle cinétique présenté dans la section précédente, et en assimilant le catalyseur à un réacteur parfaitement mélangé, les équations de conservation des espèces et de l'énergie forment le système d'équations différentielles suivant :

$$\left\{ \begin{array}{l} \frac{dY_i}{dt} = \frac{1}{Vol} \left(\frac{Y_i^{in}}{\sum_j Y_j^{in}} \cdot Q^{in} - \frac{Y_i}{\sum_j Y_j} \cdot Q + W_i \cdot S_{cata} \cdot \omega_i \right), \text{ pour l'ensemble des espèces gazeuses } i. \\ \frac{dT_g}{dt} = \frac{1}{m_g \cdot \sum_i X_i \cdot cv_i} \left(Q^{in} \cdot T_g^{in} \cdot \sum_i X_i^{in} \cdot cp_i - Q \cdot T_g \cdot \sum_i X_i \cdot cp_i - \frac{dm_g}{dt} \cdot T_g \cdot \sum_i X_i \cdot cv_i \right. \\ \quad \left. - \frac{2,89 \cdot \lambda_g}{L_{canaux}} \cdot S_{cata} \cdot (T_g - T_s) - m_g \cdot T_g \cdot \sum_i \frac{dX_i}{dt} \cdot \left(cp_i - \frac{R}{W_i} \right) \right). \\ \frac{dT_s}{dt} = \frac{S_{cata}}{Vol_{tot} \cdot \rho_{app} \cdot CpCC} \cdot \left(\frac{2,89 \cdot \lambda_g}{L_{canaux}} \cdot (T_g - T_s) - \sum_i \omega_i \cdot h_i \cdot R \right). \end{array} \right.$$

Avec,

- Y_i : masse volumique de l'espèce i (kg/m^3),
- Vol : volume ouvert du monolithe (m^3),
- Q : débit massique du gaz (kg/s),
- W_i : masse molaire de l'espèce i (kg/mol),
- S_{cata} : surface géométrique du catalyseur (m^2),
- ω_i : taux de réaction molaire de l'espèce i ($mol/m^3/s$),
- T : température (K),
- X_i : fraction massique de l'espèce i ($-$),
- cp_i : chaleur spécifique à pression constante de l'espèce i ($J/kg/K$),
- m : masse total (kg),
- cv_i : chaleur spécifique à volume constant de l'espèce i ($J/kg/K$),
- λ : conductivité thermique ($W/m/K$),
- L_{canaux} : largeur des canaux (m),
- h_i : enthalpie de l'espèce i (J/kg),
- ρ_{app} : densité apparente du monolithe (kg/m^3),
- Vol_{tot} : volume total du monolithe (m^3),
- R : constante des gaz parfaits ($J/mol/K$),
- $CpCC$: capacité calorifique du catalyseur ($J/kg/K$).

Indices,

- g : relatif au gaz,
- s : relatif au solide (monolithe),
- i ou j : relatif aux espèces.

Exposant,

- in : conditions en entrée.

La résolution de ce système est réalisée par une méthode de résolution d'équation différentielle basée sur la méthode de Gear, [70] (fonction LSODE de Fortran).

Dans un premier temps, ce modèle était implémenté dans le logiciel métier AMESim (voir Lynn *et al* [102] pour un exemple d'application). Il s'agit d'un environnement de simulation permettant de relier des systèmes de différents domaines physiques (hydraulique, pneumatique, mécanique, électrique, ...). Mais diverses raisons (lenteur d'exécution, plus grande facilité de modification, vérification plus facile, accès aux bibliothèques scientifiques standards), nous ont conduit à réimplémenter ce modèle en Fortran.

Entrées, sorties du modèle numérique et paramètres

En ce qui concerne les sorties seules les compositions de HC et CO du piège à NOx sont analysées, puisque nous le rappelons, dans la première phase de son fonctionnement (fonctionnement en mélange pauvre), seules ces deux espèces sont traitées.

Pour les entrées du modèle, il s'agit des mêmes que celles définies pour le modèle cinétique, à savoir le débit massique, la température du gaz en entrée, ainsi que la composition massique pour cinq espèces : HC , CO , O_2 , CO_2 et NO . Les quatre autres espèces présentes dans le gaz en entrée ainsi que le profil de température ne sont pas considérés comme entrées pour les mêmes raisons que celles présentées dans la section 1.1.2.

De plus, comme nous nous plaçons en fonctionnement en mélange pauvre, dans lequel seules les sorties HC et CO sont analysées, et après discussions avec les expérimentateurs, les entrées prépondérantes de notre système qui nous intéressent sont la composition massique de HC et de CO les effluents du moteurs, auxquelles nous rajoutons la composition massique du O_2 .

En effet, pour cette dernière, comme nous nous plaçons dans le cas d'un mélange pauvre, c'est-à-dire en excès d'oxygène, la variable O_2 apparaît également prépondérante dans l'étude du fonctionnement du piège à NOx. Par ailleurs, les compositions de CO_2 et NO , ont été prises égales à la valeur médiane de leurs plages de variations respectives (voir tableau 1.2), soit 8% et 525 ppm.

De plus, concernant le débit massique, bien qu'il ait un effet sur la conversion des polluants, il a été choisi de ne pas le considérer afin de simplifier le système. Il est donc, comme la température initiale du gaz, fixé à une valeur identique pour tous les essais. Ce qui nous conduit, au final, à trois entrées, les compositions initiales de HC , CO et O_2 , les autres composants ou conditions opératoires étant soit déduites de ces dernières, soit fixées.

Pour ce qui est des paramètres, nous retrouvons les mêmes paramètres que pour le modèle physique, à savoir les facteurs préexponentiels et énergies d'activation des réactions d'oxydation de CO et HC (A_{CO} , E_{aCO} , A_{HC} et E_{aCO} , voir équation 1.12).

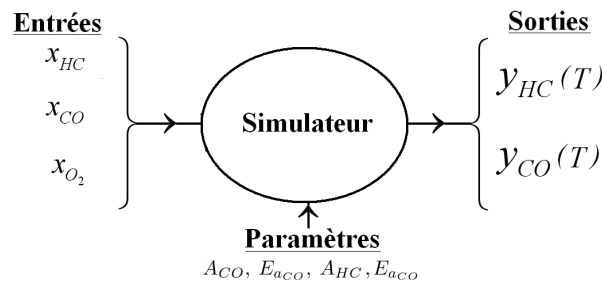


Figure 1.9 – Schéma de notre problème : modèle numérique

Nous avons donc au final trois entrées (schéma 1.9), la composition initiale de CO , HC et O_2 , deux sorties qui sont les courbes en sortie du moteur de CO et HC et quatre paramètres.

Plan d'expériences

Le tableau 1.3 présente la matrice d'expérience réalisée sur BGS, avec le numéro d'expérience affecté à chacune, ainsi que la composition de *HC*, *CO* et *O₂*, la démarche précise d'obtention de la matrice d'expérience étant présentée dans le chapitre 5. À noter tout de même que, le domaine de chaque espèce à été échantillonné en 11 nœuds du fait des appareillages expérimentaux.

n°	1	2	3	4	5	6	7	8	9	10
<i>HC</i> (ppmC)	3380	2140	3380	900	900	1520	2760	3380	3690	1520
<i>CO</i> (ppm)	4410	1230	4940	1230	3350	1760	2290	1230	2820	6000
<i>O₂</i> (%)	16.50	6.52	10.80	3.67	16.50	15.07	15.07	7.94	13.65	13.65
n°	11	12	13	14	15	16	17	18	19	20
<i>HC</i> (ppmC)	2760	900	1210	3070	2140	2140	4000	3070	4000	2140
<i>CO</i> (ppm)	700	2290	4940	4410	3350	4940	3350	2820	4940	2290
<i>O₂</i> (%)	10.80	12.22	5.09	6.52	12.22	16.50	7.94	5.09	3.67	2.24

Tableau 1.3 – Numérotation et composition de *HC*, *CO* et *O₂* des 20 essais réalisés sur BGS

La représentation graphique du plan d'expériences dans le domaine défini est donnée par la figure 1.10, les valeurs ayant été normalisées selon les plages de variations données pour chaque espèce (0 étant la borne inférieure et 1 la borne supérieure). Notre domaine ainsi défini correspond à un plan factoriel complet à 11 niveaux par facteurs. La construction de ce plan est donnée dans le chapitre 5.

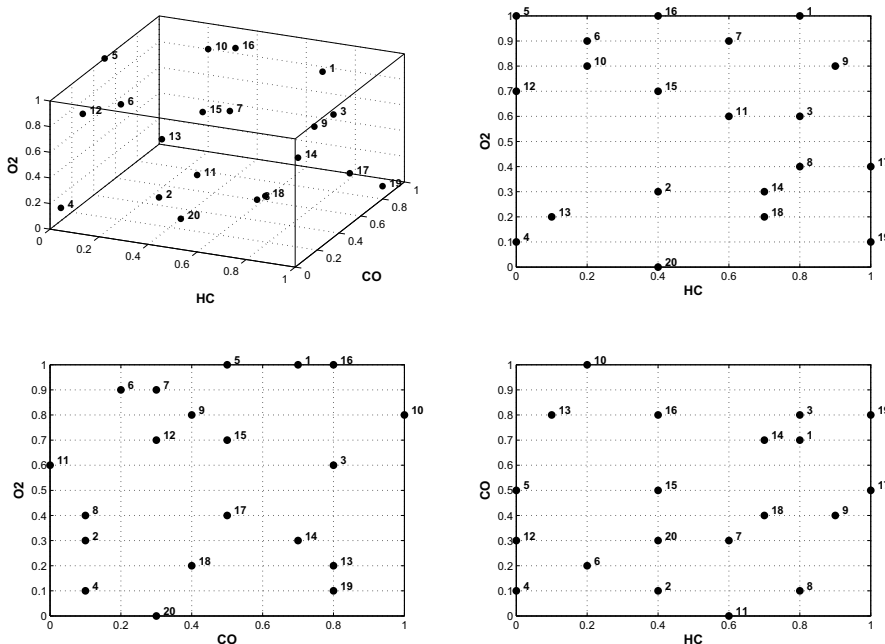


Figure 1.10 – Représentation du plan d'expériences dans le domaine

1.1.4 Comparaison résultats expérimentaux et résultats numériques

Dans cette partie, nous comparons maintenant les résultats expérimentaux obtenus sur BGS avec les résultats du modèle numérique.

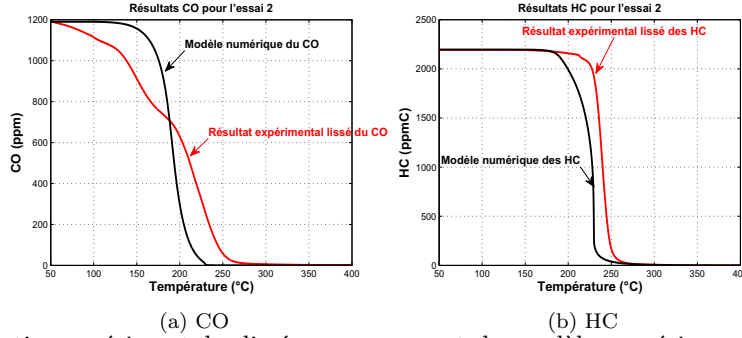


Figure 1.11 – Sorties expérimentales lissées, en rouge, et du modèle numérique, en noir, de l'essai 2

Les figures 1.11a et 1.11b montrent les résultats expérimentaux lissés de l'essai 2 en rouge et les résultats du modèle numérique en noir, dont les paramètres cinétiques sont déterminés par moindres carrés. Ce calage des paramètres cinétiques est fait sur tous les essais à la fois, ce qui peut expliquer que certains peuvent être "sacrifiés".

Il se dégage clairement de ces deux figures que, malgré un calage des paramètres par une méthode de type moindres carrés, le modèle n'arrive pas à représenter la réalité expérimentale de manière satisfaisante. Tout d'abord, pour la sortie CO , la forme de la courbe issue du modèle numérique ne correspond pas à la sortie expérimentale, cette dernière ayant une décroissance légère, alors que le modèle a une décroissance beaucoup plus forte avec un palier initial beaucoup plus marqué. En revanche, il est à noter que la température de "light-off" est relativement bien déterminée par le modèle. Pour ce qui est de la sortie HC , le modèle respecte la forme attendue, à savoir un palier initial, une forte décroissance, et un palier final, mais est décalé par rapport aux résultats expérimentaux.

Au travers de ces deux figures, il apparaît clairement que le modèle numérique est inadéquat à représenter la réalité expérimentale. Afin de mieux mettre en évidence cette différence, la figure 1.12 illustre la différence entre les résultats expérimentaux et du modèle numérique pour l'essai 2. De plus, cette figure nous permet également de constater le caractère non-stationnaire de notre problème.

En effet, notre système est un système non-stationnaire, dans le sens où ce n'est pas le temps d'expérience, c'est la température du gaz en entrée du catalyseur qui définit la progression de notre système.

La stationnarité en covariance est définie de la manière suivante, en prenant un processus temporel à valeurs réelles et en temps discret Z_1, Z_2, \dots, Z_t :

$$\mathbb{E}[Z_i] = \mu, \forall i = 1, \dots, t \quad (1.15)$$

$$\text{Cov}[Z_i, Z_{i-k}] = \rho_k \forall i = 1, \dots, t; \forall k = 1, \dots, t \quad (1.16)$$

Autrement dit, la valeur de la covariance ρ_k entre deux variables Z_t et Z_{t-k} dépend seulement de l'ampleur k . Or, pour notre cas d'étude, bien que l'intégration du système se fasse selon le temps, c'est la température qui détermine le degré d'avancement du processus. On a une relation liant la température au temps (profil de température).

Ceci entraîne le phénomène que nous voyons sur la figure 1.12, à savoir que sur le palier il n’y a pas d’erreur de modélisation et ainsi les fortes erreurs sont très locales, et donc la covariance entre deux points n’est pas indépendante de la température. Il s’agit là d’une réelle difficulté que les méthodes envisagées pour ce travail n’ont pas toutes pu intégrer complètement.

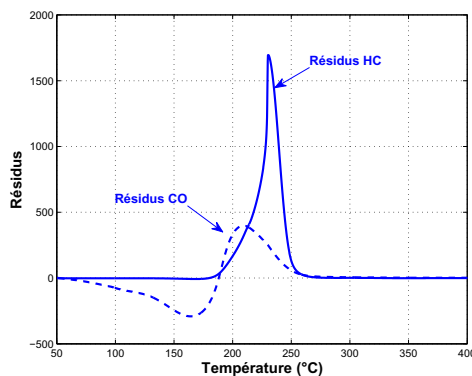


Figure 1.12 – Résidus de CO, en pointillés, et HC, en trait plein, entre les résultats expérimentaux et du modèle numérique pour l’essai 2

1.1.5 Modification du modèle numérique

Au cours de la thèse, il a été décidé, en concertation avec les expérimentateurs, de modifier le dénominateur utilisé dans le calcul des vitesses de réactions, équation 1.9. Nous soulignons le fait que le choix de cette modification est intervenu tardivement, car notre objectif premier était le suivant : étant donné un modèle certes inadéquat, mais qui apporte de l’information, quelle(s) méthode(s) peut-on mettre en place pour réaliser un plan d’expériences ?

Cependant, pour différentes raisons évoquées dans le chapitre 4, nous nous sommes penchés sur les raisons de l’inadéquation du modèle numérique. Après plusieurs études, nous sommes arrivés à déterminer une nouvelle expression pour ce dénominateur permettant de mieux ajuster les résultats du modèle aux résultats expérimentaux. Cette nouvelle expression nécessite également l’ajout d’un paramètre cinétique à estimer, la constante d’adsorption de la réaction de *HC*.

Une fois ce nouveau modèle mis en place, nous avons estimé le vecteur des cinq paramètres cinétiques constant pour tous les points du domaine. De manière classique, cette estimation est effectuée, comme précédemment, par moindres carrés, soit en minimisant la somme des résidus au carrés entre modèle numérique et résultats expérimentaux. Nous ne présentons pas les résultats obtenus par cette méthode, puisque malgré l’amélioration constatée, nous ne sommes pas capables de déterminer un vecteur optimal pour tous les points. En revanche, il est possible de déterminer un jeu de paramètres cinétiques expérience par expérience.

Cette particularité constatée, nous a permis de développer une méthode pertinente de modélisation, introduite dans la section 4.2. Nous insistons donc sur le fait que les modifications du modèle présentées ici, n’interviennent que dans le chapitre 4. Le choix d’introduire ces modifications dans ce premier chapitre est naturelle et a pour objectif de ne pas surcharger le chapitre 4 de considérations physiques et ainsi perdre le lecteur.

On rappelle que le modèle numérique correspond à l’intégration d’un système d’équations aux dérivées ordinaires (voir section 1.1.3). Ce système a été obtenu à partir des équations de conservation des espèces et de l’énergie relativement au modèle cinétique retenu.

Lors de l'intégration de ce système, il est notamment nécessaire de calculer le taux de réaction molaire ω de chaque espèce défini comme étant la somme des vitesses des réactions impliquant l'espèce, pondérées par le coefficient stœchiométrique du modèle cinétique. On rappelle que la vitesse de réaction pour l'oxydation de l'espèce générique *esp* retenue est définie par,

$$R_i = \frac{k_i \cdot x_{esp} \cdot x_{O_2}}{F}, \quad (1.17)$$

où k_i correspond à la constante cinétique, x_{esp} et x_{O_2} aux fractions molaires respectives des espèces *esp* et O_2 , et F au dénominateur utilisé dont l'expression est rappelée ci-après, équation 1.18.

Ce sont ces vitesses de réactions qui dépendent, au travers des constantes cinétiques k_i (équation 1.12), des paramètres cinétiques que l'on cherche à ajuster. Au final, nous avons déterminé quatre paramètres cinétiques à ajuster, les facteurs préexponentiels A_i et énergies d'activation Ea_i pour les réactions d'oxydation de *CO* et *HC*.

En ce qui concerne les constantes d'adsorption (équation 1.13), nous avons notamment signalé le fait que l'expression du dénominateur F (équation 1.18) qui les emploie, est empirique et couramment utilisée dans la littérature. Elle est issue des travaux de Voltz *et al* [157]. Néanmoins, dans ces travaux, l'obtention de cette expression a été réalisée dans un contexte différent du nôtre, puisqu'il s'agit d'une étude sur un moteur essence, avec une richesse de 1. La forme de ce dénominateur est la suivante :

$$F = T_s(1 + K_0x_{CO} + K_1x_{HC})^2(1 + K_2x_{CO}^2x_{HC}^2)(1 + K_3x_{NO}^{0.7}) \quad (1.18)$$

Les auteurs ne décrivent pas l'obtention de ce dénominateur mais donne une explication sur chacun des termes. Le premier terme, $(1 + K_0x_{CO} + K_1x_{HC})^2$, correspond à une fonction du second ordre prenant en compte l'effet d'inhibition due à la chimisorption de *CO* et *HC*. Le second terme, $(1 + K_2x_{CO}^2x_{HC}^2)$, est nécessaire pour l'ajustement aux données expérimentales (du cas présenté) pour des concentrations élevées de *CO* et *HC*. Et enfin le terme d'adsorption de *NO*, $(1 + K_3x_{NO}^{0.7})$ s'adapte bien à l'effet d'inhibition de *NO* pour les deux réactions d'oxydation, l'exposant 0.7 étant lui aussi déterminé de manière empirique.

Notre démarche a été alors de remettre en cause ce terme, et les constantes d'adsorption associées (tableau 1.1). Ainsi, nous sommes partis de l'expression la plus simple possible avec des effets de premier ordre pour chacune des variables, en gardant les valeurs des constantes d'adsorption puis nous avons ajusté les paramètres cinétiques par moindres carrés. Nous avons réitéré ce procédé jusqu'à l'obtention du meilleur résultat au sens de l'ajustement aux résultats expérimentaux, en modifiant soit la forme de F , soit en rajoutant dans l'estimation des paramètres cinétiques les constantes d'adsorption.

Au final, le dénominateur retenu est le suivant :

$$F = T_s(1 + K_{ads}x_{HC})(1 + K_3x_{NO}) \quad (1.19)$$

On voit donc que nous avons pris un terme d'adsorption du premier ordre pour *HC* dans lequel la constante d'adsorption K_{ads} est rajoutée au vecteur des paramètres cinétiques à calibrer, et un terme du premier ordre pour *NO*, en conservant la valeur de la constante d'adsorption K_3 associée issue de la littérature. À noter, que l'exposant 0.7 a été supprimé, car ce dernier n'avait aucune valeur cinétique, mais n'était qu'un artefact numérique.

Ainsi, nous avons au final un modèle numérique dépendant d'un vecteur β de cinq paramètres cinétiques à ajuster, soit respectivement, β_1 et β_2 le facteur préexponentiel et énergie d'activation pour la réaction de *CO*, β_3 et β_4 ceux pour la réaction de *HC*, et β_5 la constante d'adsorption de *HC*.

1.2 Cas maîtrisé

Afin de tester et vérifier les diverses solutions envisagées pour répondre à notre problématique, nous voulons partir d'un cas similaire au cas d'étude, mais que nous maîtrisons.

Ce cas maîtrisé comporte deux éléments : tout d'abord un modèle cinétique simulant le comportement expérimental, et dont les réponses sont très différentes, à forte et faible température ; ensuite, un modèle cinétique qui nous sert de simulateur. Ce modèle est choisi de telle sorte qu'il ne représente pas adéquatement les résultats expérimentaux.

On s'intéresse à l'évolution de la concentration d'une espèce, en présence d'un catalyseur. Ce système est régi par une formalisation de type Langmuir-Hinshelwood (Gadsby *et al* [68]). Le tutoriel de McCann et Maeder [111] pour la modélisation de processus chimique pourra éclairer de manière accessible le lecteur non averti, ainsi que l'article de Brauner et Shacham [15] pour les questions liées aux constantes d'Arrhenius.

Nous présentons donc dans un premier temps ces deux modèles qui vont jouer le rôle, par analogie avec notre cas d'étude, des résultats expérimentaux et numériques, et rappelons les motivations qui nous ont conduit à mettre en place ce cas maîtrisé. Puis nous introduisons les conditions de simulation, les considérations pratiques liées aux choix des paramètres et du domaine, et exposons les résultats de simulation des deux modèles pour plusieurs cas. Enfin, nous comparons ce cas avec le cas d'étude, afin de bien mettre en exergue la similitude des deux phénomènes que nous utilisons tout au long de ce manuscrit.

1.2.1 Présentation du cas maîtrisé et motivations

Le cadre de notre cas maîtrisé, est proche de notre cas d'étude puisque nous nous intéressons à l'évolution de la concentration d'une espèce en présence d'un catalyseur. Considérons un système cinétique simple de la forme,



où A et B correspondent à des espèces chimiques, qui réagissent en présence d'un catalyseur. Ce système est régi par le système d'équations (1.20), selon la formalisation de Langmuir-Hinshelwood.

$$\begin{cases} \frac{d[A]}{dt} = -\frac{k_0 \exp\{-\frac{E}{RT}\}[A]}{1 + b_0 \exp\{-\frac{\Delta H}{RT}\}[A]}, \\ [A]_0 = A_0, \end{cases} \quad (1.20)$$

où $[A]$ correspond à la concentration de l'espèce A , $[A]_0$ sa valeur initiale, T la température, R constante des gaz parfaits, et $\gamma = \{k_0, E, b_0, \Delta H\}$ est le vecteur des paramètres cinétiques.

Ces paramètres sont choisis pour conduire à des comportements de l'évolution de la concentration de l'espèce A très différents à faible et forte températures.

L'intégration du système se fait sur l'intervalle $[0, 200]$ secondes. Trois concentrations initiales sont posées pour l'espèce A , $[A]_0 = \{1, 2, 3\}$. La plage de température est fixée entre 20°C et 60°C. On choisit

$$\begin{aligned} &\text{pour } 20^\circ\text{C}, k_0 \exp\{-\frac{E}{RT}\} = 0.1 \text{ et } b_0 \exp\{-\frac{\Delta H}{RT}\} = 3, \\ &\text{pour } 60^\circ\text{C}, k_0 \exp\{-\frac{E}{RT}\} = 2 \text{ et } b_0 \exp\{-\frac{\Delta H}{RT}\} = 0.2. \end{aligned}$$

À partir de ces connaissances, il est possible de déterminer les paramètres cinétiques. On a,

$$\gamma = \{6.8541e9, 6.0815e4, 4.8056e - 10, -5.4974e4\}. \quad (1.21)$$

On note $g_\gamma(x)$, le modèle régi par le système (1.20), où γ est le vecteur de paramètre cinétique, et x correspond aux entrées de notre système : la température, le temps et la concentration initiale. En sortie, nous regardons l'évolution de la concentration de l'espèce A en fonction soit du temps, soit de la température.

On se donne le système cinétique simple suivant (terme d'adsorption non pris en compte) :

$$\begin{cases} \frac{d[A]}{dt} = -k'_0 \exp\{-\frac{E'}{RT}\}[A], \\ [A]_0 = A_0. \end{cases} \quad (1.22)$$

On note $f(x, \beta)$, le modèle régi par le système (1.22), où β est le vecteur de paramètres cinétiques, $\{k'_0, E'\}$.

Le but de ces deux systèmes est de disposer d'un cas similaire à notre cas d'étude mais plus simple et surtout totalement maîtrisé, sur lequel, il nous sera plus commode d'expérimenter diverses méthodes.

1.2.2 Simulations

Trois simulations ont été réalisées et sont présentées dans la figure 1.13. Cette figure représente l'évolution de la concentration de l'espèce A pour les trois concentrations initiales, à la température de 20°C. La simulation de la tendance (système approché (1.22)) est également tracée pour chacune des concentrations initiales, en pointillés. k'_0 et E' sont respectivement fixés à 3.6075e4 et 9.1641e4. Ces valeurs ont été obtenues par moindres carrés, sur une base de résultats expérimentaux déterminés par le système 1.20 et les résultats numériques correspondants calculés par le système 1.22.

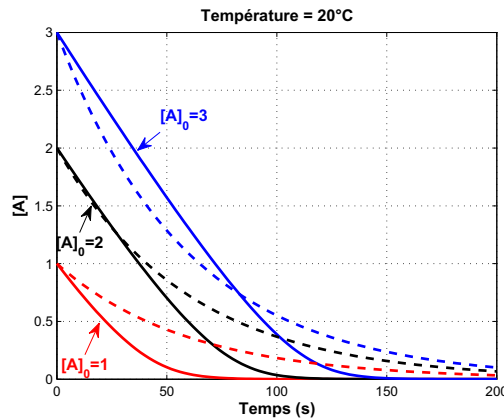


Figure 1.13 – Simulation du système (1.20) en traits continus, et du système (1.22) en pointillés, pour 3 concentrations initiales différentes en fonction du temps, pour $T=20^\circ\text{C}$

Afin de se placer dans la même disposition que le cas de déNOx, il est possible de représenter l'évolution de la concentration de A , en fonction de la température, à temps fixé (figure 1.14). Le modèle (1.22) n'arrive pas à approcher correctement le vrai système (1.20).

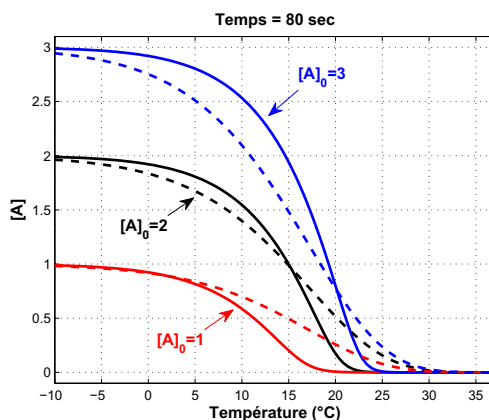


Figure 1.14 – Simulation du système (1.20) en traits continus, et du système (1.22) en pointillés, pour 3 concentrations initiales différentes en fonction de la température, pour $t=80\text{sec}$

1.2.3 Pertinence du cas maîtrisé

Le but de cette partie est de bien faire le lien entre notre cas d'étude et notre cas maîtrisé, et d'exposer les similitudes entre ces deux cas. L'objectif était de mettre en place un cas proche de notre cas d'étude mais que nous maîtrisions totalement afin de pouvoir dans un premier temps développer les méthodes implémentées sur ce cas, afin de pouvoir juger de la pertinence de ces méthodes.

Tout d'abord, l'exemple même de notre cas maîtrisé, est proche de notre cas d'étude, dans le sens où il s'agit dans les deux cas de regarder l'évolution d'une espèce qui réagit en présence d'un catalyseur, en fonction de la température. Cependant, la complexité du cas est réduite, le nombre d'entrées, de sorties et de paramètres étant beaucoup plus faible (trois entrées, quatre paramètres et deux sorties pour le cas d'étude contre trois entrées, deux paramètres et une sortie pour le cas maîtrisé). De plus nous disposons également d'une infinité de résultats expérimentaux pour le cas maîtrisé contre vingt pour le cas d'étude.

Dans un premier temps, nous nous sommes dotés d'un premier modèle cinétique régi par le système d'équations (1.20) selon la formalisation de Langmuir-Hinshelwood. Les résultats de ce premier système font office de "résultats expérimentaux". Au regard de la figure 1.14, on peut voir tout d'abord que la forme même de la courbe en sortie est proche des résultats expérimentaux de notre cas d'étude. En effet, on distingue les trois grandes zones qui caractérisent notre cas d'étude : un palier initial, une décroissance plus ou moins forte, et un palier final. De plus, comme dans notre cas d'étude, le comportement de ce système est très différent selon les conditions opératoires choisies.

Une des hypothèses avancées pour expliquer l'inadéquation du modèle numérique à représenter la réalité expérimentale pour notre cas d'étude est que le modèle cinétique retenu n'est pas complet (réaction manquante, expression du terme d'adsorption erroné, ...). Partant de ce constat, nous nous sommes dotés, pour le cas maîtrisé, d'un second système, proche du premier, mais incomplet. En effet, le terme d'adsorption n'est pas présent. Ce second système, par analogie avec le cas d'étude, joue le rôle de modèle numérique.

Après avoir calibré les paramètres de ce second système de la même manière que dans le cas d'étude, soit sur une base d'expériences, il nous est possible de simuler pour différentes conditions opératoires les résultats obtenus par ce second système. Ainsi, comme on peut le voir sur la figure 1.14, il y a incapacité du "modèle numérique" à représenter les résultats "expérimentaux". En effet, soit le décrochage de la courbe, moment où la décroissance

débute, est décalé, comme pour la sortie HC dans le cas d'étude, soit la forme de la sortie n'est pas pleinement satisfaisante, comme pour le CO .

Nous avons donc répondu à l'objectif que nous nous étions fixé : avoir un outil plus simple et totalement maîtrisé sur lequel il sera possible de jauger la pertinence des différentes méthodes que nous présentons dans ce manuscrit. La validation de ces méthodes sur ce cas maîtrisé est essentielle préalablement à l'application sur notre cas d'étude.

1.3 En bref

Nous rappelons en bref dans cette dernière section les grandes lignes de ce chapitre, afin de fournir les éléments clés de cette mise en place de notre problématique : la planification d'expériences adaptée aux problèmes de cinétique. Dans ce premier chapitre, nous présentons les phénomènes étudiés : notre cas d'étude, le piège à NO_x et un second cas fabriqué afin d'être similaire au cas d'étude mais totalement maîtrisé.

Dans un premier temps, nous avons mis en avant les enjeux économiques et écologiques qui obligent à développer des systèmes de post-traitement de plus en plus perfectionnés. Les travaux que nous présentons se situent dans ce cadre puisque notre cas d'étude est le système de post-traitement des polluants à l'échappement des moteurs Diesel, et plus exactement le piège à NO_x . Après en avoir explicité le principe général et son schéma de fonctionnement, ainsi que le protocole expérimental mis en place, nous avons pu rentrer plus en détails dans le fonctionnement précis, en introduisant le modèle physico-chimique le régissant, mais également les entrées, sorties et paramètres qui vont nous intéresser tout au long de cette thèse. Il est à noter que notre cas d'étude, se place dans un contexte fonctionnel, puisque les sorties considérées sont des courbes en fonction de la température. Enfin, le modèle numérique développé a été amplement détaillé, ce qui nous a permis de mettre en exergue son incapacité à représenter le fonctionnement réel du piège à NO_x .

Ce premier constat d'inadéquation, nous a obligé à travailler sur la modélisation du phénomène étudié, en partant de ce premier modèle numérique. Cette tâche étant loin d'être triviale, comme nous pourrions le voir, nous nous sommes dotés d'un second cas, appelé cas maîtrisé, que nous avons fabriqué afin d'être similaire au cas d'étude dans sa forme, mais que l'on pourra utiliser comme outil au développement de méthode répondant à notre problématique.

Ce second phénomène a pour but d'étudier l'évolution d'une espèce chimique en présence d'un catalyseur. Après en avoir expliqué la construction, et présenté quelques simulations, nous avons pu le comparer au cas d'étude, et nous avons constaté que ces deux cas sont bien similaires, aussi bien dans l'approche que dans sa construction, ainsi que la forme des sorties fonctionnelles étudiées.

Ces deux phénomènes maintenant présentés, nous sommes à même de traiter le problème qui nous est soumis, à savoir développer une méthodologie de planification d'expériences appliquée aux problèmes de cinétique. Cependant, pour en arriver là, il nous faut dans un premier temps résoudre le problème d'inadéquation du modèle numérique, car les méthodes envisagées pour la planification d'expériences partent du postulat que nous avons à disposition un modèle pertinent. Cette première étape a constitué la majeure partie des travaux durant cette thèse du fait de la complexité du problème, de sa forte non linéarité, mais surtout de son caractère non-stationnaire, comme nous avons pu le noter.

Outils de modélisation

2.1	Krigeage	43
2.1.1	Krigeage	43
2.1.2	Choix du noyau	49
2.1.2.1	Familles classiques de covariances	49
2.1.2.2	Noyaux de covariance non stationnaires	50
2.1.3	Krigeage avec bruit	52
2.1.4	Cokrigeage	54
2.2	Données fonctionnelles	58
2.2.1	Définition et état de l'art	59
2.2.2	Modèles à coefficients variables	60
2.2.2.1	Présentation	61
2.2.2.2	Méthodes d'estimations	63
2.2.3	Modèles non paramétriques additifs	69
2.3	Conclusions	73

Outils de modélisation

Ce chapitre présente les outils de modélisation utilisés au cours de la thèse et est scindé en deux parties. La première partie concerne les méthodes utilisées dans le cadre non fonctionnel, basées sur la théorie de krigeage, la seconde concerne les méthodes fonctionnelles, avec notamment l'introduction des données fonctionnelles et des modèles à coefficients variables.

2.1 Krigeage

Le krigeage est une méthode stochastique d'interpolation spatiale qui prévoit la valeur d'un phénomène en des sites non échantillonnés par une combinaison linéaire sans biais et à variance minimale des observations du phénomène en des sites voisins. Un des intérêts de cette méthode est qu'elle fournit, à faible coût, une estimation de l'erreur de prédiction en tout point du domaine d'étude, ce qui donne une idée de la précision du modèle aux points non échantillonnés.

Le krigeage est la première méthode d'interpolation spatiale à prendre en considération la structure de dépendance spatiale des données. Les travaux de l'ingénieur minier sud-africain Krige (1951) sont précurseurs de la méthode.

Cependant, le terme krigeage et le formalisme de cette méthode sont dus au français Matheron ([105], [106], [107]), qui en a aussi assuré le développement au Centre de Géostatistique de l'École des Mines de Paris. Les fondements de la méthode ont été développés parallèlement par d'autres chercheurs, notamment le météorologue Gandin de l'ex-URSS, mais c'est aujourd'hui sous la terminologie proposée par Matheron qu'elle est la plus connue (Cressie [35]).

Nous commençons par introduire la base de la théorie du krigeage, puis, nous présentons et discutons les noyaux de covariance les plus usuels et introduisons un exemple de noyau non stationnaire. Ensuite, deux extensions de cette théorie sont présentées. Tout d'abord, la théorie de krigeage avec bruit, puis la notion de cokrigeage.

2.1.1 Krigeage

Généralités

Le krigeage est une méthode qui prévoit la valeur d'une réponse par une combinaison linéaire sans biais et à variance minimale des données. Son modèle de base a la même forme que le modèle de régression classique ou locale, mais les erreurs sont maintenant supposées dépendantes spatialement. Ce modèle est défini par :

$$Y(s) = \mu(s) + z(s), \quad s \in D \quad (2.1)$$

où $\mu(\cdot)$ est la structure déterministe pour l'espérance de $Y(\cdot)$ et $z(\cdot)$ une fonction aléatoire stationnaire, d'espérance nulle et de structure de dépendance connue. Pour formuler complètement le modèle, il faut spécifier la forme de la tendance $\mu(\cdot)$. C'est en fait cette tendance qui précise le type de krigeage effectué.

Les trois types classiques de krigeage sont :

- Le krigeage simple : $\mu(s) = m$ est une constante connue.
- Le krigeage ordinaire : $\mu(s) = \mu$ est une constante inconnue.
- Le krigeage universel : $\mu(s) = \sum_{j=0}^p f_j(s)\beta_j$ est une combinaison linéaire de fonctions de la position de s , p étant une dimension quelconque.

De plus, la structure de dépendance de la fonction aléatoire $z(\cdot)$ doit être précisée. Si elle n'est pas connue préalablement, ce qui est presque toujours le cas en pratique, elle est déterminée à partir des données lors de l'"analyse variographique", étape qui permet de décrire la variabilité spatiale de phénomènes régionalisés.

Le krigeage est une méthode d'interpolation très souple, qui peut être globale ou locale dépendamment du voisinage choisi. De plus, il s'agit d'une méthode d'interpolation exacte. Il restitue donc les valeurs régionalisées mesurées aux sites d'observation. Cependant, il est aussi possible d'effectuer un krigeage dit avec erreurs de mesure qui constitue alors une approximation (voir section 2.1.3).

Finalement, comme en régression classique ou locale, des variables régionalisées auxiliaires peuvent être intégrées au krigeage en les ajoutant à la tendance générale $\mu(\cdot)$ du modèle. Le krigeage incorporant une telle tendance est nommé "krigeage avec dérive externe". Le krigeage possède aussi d'autres extensions multivariées, notamment le "cokrigeage" (voir section 2.1.4).

Le modèle de krigeage, équation (2.1), a la même forme que le modèle de régression classique, à ceci près que $z(s)$ n'est pas indépendamment et identiquement distribué, mais on suppose qu'il s'agit d'un processus gaussien de moyenne nulle, de variance σ^2 , et de fonction de corrélation,

$$r(\boldsymbol{\theta}; s, t) = \text{Corr}(z(s), z(t)), \quad (2.2)$$

où $r(\boldsymbol{\theta}; s, t)$ est une fonction définie positive spécifiée de $z(s)$ et $z(t)$. Étant donné que cette fonction de corrélation est censée représenter la corrélation spatiale, une classe naturelle est la famille stationnaire définie par :

$$r(\boldsymbol{\theta}; s, t) \equiv r(\boldsymbol{\theta}; |s - t|). \quad (2.3)$$

La fonction de corrélation la plus répandue dans la littérature est la fonction de corrélation exponentielle, définie par :

$$r(\boldsymbol{\theta}; s, t) = e^{-\sum_{k=1}^p \theta_k |s_k - t_k|^q}, \quad \text{pour } 0 < q \leq 2 \quad (2.4)$$

dans le cas anisotrope, et

$$r(\boldsymbol{\theta}; s, t) = e^{-\theta \sum_{k=1}^p |s_k - t_k|^q}, \quad \text{pour } 0 < q \leq 2 \quad (2.5)$$

dans le cas isotrope. La section 2.1.2 introduit quelques noyaux classiques.

Supposons que l'on connaisse les valeurs de la fonction objectif, équation (2.1), aux points x_1, \dots, x_n , et que l'on souhaite prédire $Y(x_0)$. Le modèle (2.1), implique que $Y_0 = Y(x_0)$ et $Y^n = (Y(x_1), \dots, Y(x_n))$ possèdent la loi multivariée,

$$\begin{pmatrix} Y_0 \\ Y^n \end{pmatrix} \sim \mathcal{N}_{1+n} \left[\begin{pmatrix} f_0^t \\ F \end{pmatrix} \boldsymbol{\beta}, \sigma^2 \begin{pmatrix} 1 & r_0^t \\ r_0^t & R \end{pmatrix} \right], \quad (2.6)$$

où

- $f_0 = f(x_0)$ est le vecteur $p \times 1$ des fonctions de régression pour $Y(x_0)$,
- $F = (f_j(x_i))$ est la matrice $n \times p$ des fonctions de régression pour les points du plan courant,
- $r_0 = (R(x_0 - x_1), \dots, R(x_0 - x_n))^t$ est le vecteur $n \times 1$ des corrélations de Y^n avec $Y(x_0)$,
- $R = (R(x_i - x_j))$ est la matrice $n \times n$ des corrélations de Y^n .

Les paramètres $\beta \in \mathbb{R}^p$ et $\sigma^2 > 0$, sont inconnus.

Formules de prédicteur et de variance de prédiction

Soit un ensemble de n points d'expériences, $\mathbf{S} = [\mathbf{s}_1 \dots \mathbf{s}_n]^t$ où $\mathbf{s}_i \in \mathbb{R}^p$ et le vecteur des réponses associées, $\mathbf{Y} = [\mathbf{y}_1 \dots \mathbf{y}_n]^t$. On cherche à ajuster les réponses par le modèle suivant,

$$\mathbf{y}(\mathbf{x}) = f(\mathbf{x})^t \beta + z_{\sigma^2, \theta}(\mathbf{x}). \quad (2.7)$$

Le processus aléatoire z est de moyenne nulle et la covariance entre $z(\mathbf{w})$ et $z(\mathbf{x})$ s'exprime par,

$$\mathbb{E}[z(\mathbf{w})z(\mathbf{x})] = \sigma^2 \mathcal{R}_\theta(\mathbf{w} - \mathbf{x}), \quad (2.8)$$

où σ^2 est la variance du processus et $\mathcal{R}_\theta(\mathbf{w} - \mathbf{x})$ est le modèle de corrélation, de paramètre θ . Par la suite, on note $\mathbf{R}_{ij} = \mathcal{R}_\theta(\mathbf{s}_i - \mathbf{s}_j)$, $i, j = 1, \dots, n$, et $r(\mathbf{x})$ le vecteur de corrélation entre les points d'expériences et un nouveau point \mathbf{x} .

Soit le prédicteur linéaire, $\hat{\mathbf{y}}(\mathbf{x}) = c^t \mathbf{Y}$, où $c = c(\mathbf{x}) \in \mathbb{R}^n$. L'erreur de prédiction est :

$$\begin{aligned} \hat{\mathbf{y}}(\mathbf{x}) - \mathbf{y}(\mathbf{x}) &= c^t \mathbf{Y} - \mathbf{y}(\mathbf{x}), \\ &= c^t (f(\mathbf{S})\beta + \mathbf{Z}) - (f(\mathbf{x})^t \beta + z), \\ &= c^t \mathbf{Z} - z + (f(\mathbf{S})^t c - f(\mathbf{x})^t) \beta. \end{aligned} \quad (2.9)$$

où $\mathbf{Z} = [z_1 \dots z_n]^t$ est le vecteur des erreurs aux points d'expériences.

Afin de conserver un prédicteur non biaisé, quel que soit c , il faut que $f(\mathbf{S})^t c(\mathbf{x}) = f(\mathbf{x})$. Sous cette condition, la variance de prédiction est donnée par :

$$\varphi(\mathbf{x}) = \sigma^2 (1 + c^t \mathbf{R}c - 2c^t r). \quad (2.10)$$

La fonction de Lagrange du problème de minimisation de φ selon c et son gradient sont définis par,

$$\mathcal{L}(c, \lambda) = \sigma^2 (1 + c^t \mathbf{R}c - 2c^t r(\mathbf{x})) - \lambda^t (f(\mathbf{S})^t c - f(\mathbf{x})), \quad (2.11)$$

$$\frac{\partial \mathcal{L}(c, \lambda)}{\partial c} = 2\sigma^2 (\mathbf{R}c - r(\mathbf{x})) - f(\mathbf{S})\lambda. \quad (2.12)$$

À l'optimum, le gradient s'annule et on obtient le système d'équations suivant :

$$\begin{bmatrix} \mathbf{R} & f(\mathbf{S}) \\ f(\mathbf{S}) & 0 \end{bmatrix} \begin{bmatrix} c \\ \tilde{\lambda} \end{bmatrix} = \begin{bmatrix} r(\mathbf{x}) \\ f(\mathbf{x}) \end{bmatrix} \quad \text{avec } \tilde{\lambda} = -\frac{\lambda}{2\sigma^2}, \quad (2.13)$$

dont la solution est :

$$\tilde{\lambda} = (f(\mathbf{S})^t \mathbf{R}^{-1} f(\mathbf{S}))^{-1} (f(\mathbf{S})^t \mathbf{R}^{-1} r(\mathbf{x}) - f(\mathbf{x})), \quad (2.14)$$

$$c = \mathbf{R}^{-1} (r(\mathbf{x}) - f(\mathbf{S})\tilde{\lambda}). \quad (2.15)$$

Compte tenu du fait que \mathbf{R} est symétrique, $(\mathbf{R}^{-1})^t = \mathbf{R}^{-1}$, on obtient,

$$\begin{aligned} \hat{\mathbf{y}}(\mathbf{x}) &= (r(\mathbf{x}) - f(\mathbf{S})\tilde{\lambda})^t \mathbf{R}^{-1} \mathbf{Y}, \\ &= r(\mathbf{x})^t \mathbf{R}^{-1} \mathbf{Y} - (f(\mathbf{S})^t \mathbf{R}^{-1} r(\mathbf{x}) - f(\mathbf{x}))^t (f(\mathbf{S})^t \mathbf{R}^{-1} f(\mathbf{S}))^{-1} f(\mathbf{S})^t \mathbf{R}^{-1} \mathbf{Y}. \end{aligned} \quad (2.16)$$

Or, on sait que

$$\boldsymbol{\beta}^* = (f(\mathbf{S})^t \mathbf{R}^{-1} f(\mathbf{S}))^{-1} f(\mathbf{S})^t \mathbf{R}^{-1} \mathbf{Y}. \quad (2.17)$$

On a donc, en posant $\boldsymbol{\gamma}^* = \mathbf{R}^{-1}(\mathbf{Y} - f(\mathbf{S})\boldsymbol{\beta}^*)$,

$$\hat{\mathbf{y}}(\mathbf{x}) = f(\mathbf{x})\boldsymbol{\beta}^* + r(\mathbf{x})^t \boldsymbol{\gamma}^*. \quad (2.18)$$

Avec (2.14) et (2.15), la variance de prédiction de notre prédicteur devient,

$$\varphi(\mathbf{x}) = \sigma^2 (1 + \mathbf{u}^t (f(\mathbf{S})^t \mathbf{R}^{-1} f(\mathbf{S}))^{-1} \mathbf{u} - r(\mathbf{x})^t \mathbf{R}^{-1} r(\mathbf{x})), \quad (2.19)$$

où $\mathbf{u} = f(\mathbf{S})^t \mathbf{R}^{-1} r(\mathbf{x}) - f(\mathbf{x})$ et σ^2 est obtenu par (2.26). La généralisation au cas multiréponse est immédiate, il suffit de remplacer σ par σ_l , la variance du processus pour la $l^{\text{ème}}$ réponse, pour obtenir la variance de prédiction du krigeage pour la $l^{\text{ème}}$ réponse.

Soit $\mathbf{x} = \mathbf{s}_i$, le $i^{\text{ème}}$ point d'expérience. Ainsi, $r(\mathbf{x}) = \mathbf{R}_{:,i}$, la $i^{\text{ème}}$ colonne de \mathbf{R} , et $\mathbf{R}^{-1} r(\mathbf{x}) = \mathbf{e}_i$, la $i^{\text{ème}}$ colonne de la matrice unité. En reprenant (2.16), on a alors :

$$\begin{aligned} \hat{\mathbf{y}}(\mathbf{s}_i) &= f(\mathbf{s}_i)\boldsymbol{\beta}^* + r(\mathbf{s}_i)^t \mathbf{R}^{-1}(\mathbf{Y} - f(\mathbf{S})\boldsymbol{\beta}^*), \\ &= f(\mathbf{s}_i)\boldsymbol{\beta}^* + \mathbf{e}_i^t (\mathbf{Y} - f(\mathbf{S})\boldsymbol{\beta}^*), \\ &= f(\mathbf{s}_i)\boldsymbol{\beta}^* + \mathbf{y}_i - f(\mathbf{s}_i), \\ &= \mathbf{y}_i. \end{aligned}$$

Notre prédicteur de krigeage est interpolant. On vérifie que si l'on reprend le calcul de la variance de prédiction pour $\mathbf{x} = \mathbf{s}_i$, on trouve $\mathbf{u} = f(\mathbf{S})^t \mathbf{e}_i - f(\mathbf{s}_i) = 0$, et on a,

$$\begin{aligned} \varphi(\mathbf{x}) &= \sigma^2 (1 - r(\mathbf{s}_i)^t \mathbf{R}^{-1} r(\mathbf{s}_i)), \\ &= \sigma^2 (1 - \mathbf{e}_i^t \mathbf{e}_i), \\ &= \sigma^2 (1 - \mathbf{R}_{ii}), \\ &= 0. \end{aligned}$$

Il s'agit là de la présentation historique mais, en fait, on peut montrer que tout cela découle de l'équation 2.6 (calcul d'espérance et variance conditionnelles).

Estimation des paramètres de krigeage

L'estimation des paramètres de covariance passe par la maximisation selon $\boldsymbol{\theta}$ d'un critère d'adéquation entre les données observées et celles d'un processus de noyau de covariance. En pratique, on trouve deux grandes méthodes d'estimation dans la littérature :

- L'estimation par maximum de vraisemblance (EMV) consiste à déterminer le paramètre $\hat{\boldsymbol{\theta}}$ maximisant la vraisemblance, définie comme étant la densité de probabilités des observations. Cette méthode repose sur l'hypothèse que ces observations sont une réalisation d'un vecteur gaussien de matrice de covariance construite à partir de $r(\boldsymbol{\theta}, \cdot)$. Il est préférable de travailler en pratique avec la log-vraisemblance, la méthode consistant alors à trouver,

$$\hat{\boldsymbol{\theta}}_{EMV} = \max_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta} | \mathbf{y}), \quad (2.20)$$

où $\mathcal{L}(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta} | \mathbf{y})$ est la log-vraisemblance des observations \mathbf{y} en fonction des paramètres $\boldsymbol{\beta}$, σ^2 et $\boldsymbol{\theta}$. Nous détaillons plus précisément la méthode ci-après.

– La validation croisée (CV pour *Cross validation*) a pour objectif de quantifier la capacité de généralisation d'un modèle en se basant uniquement sur les données d'apprentissage. La démarche consiste à retirer une partie des observations, à prédire, pour ces observations retirées, les réponses à l'aide du modèle de krigeage sélectionné, et à mesurer la distance entre ces observations et les prédictions (généralement en utilisant la somme des écarts au carré). On réitère cette démarche en changeant à chaque fois les observations retirées. En pratique, on utilise la technique du *leave-one-out*, dans laquelle une seule observation est retirée et on effectue la démarche sur chacune des observations. L'estimation de $\hat{\boldsymbol{\theta}}$ correspond à la valeur de $\boldsymbol{\theta}$ qui minimise l'erreur globale de validation croisée, soit

$$\hat{\boldsymbol{\theta}}_{CV} = \arg \min_{\boldsymbol{\theta}} \sum_i |\mathbf{y}_i - \hat{\mathbf{y}}_i^{-i}|^2, \quad (2.21)$$

où $\hat{\mathbf{y}}_i^{-i}$ est la prédiction par la méthode de krigeage, à partir des observations auxquelles on retire la $i^{\text{ème}}$ composante.

L'estimation par maximum de vraisemblance est la méthode que nous avons retenue. Cette méthode est moins coûteuse en temps de calcul et, de plus, les routines utilisées en pratique (par exemple la routine Dace, Lophaven *et al.* [97] [98]), ont recours à ce type d'estimation.

L'hypothèse de normalité du modèle de krigeage gaussien, entraîne la densité de \mathbf{y} suivante :

$$(2\pi\sigma^2)^{-n/2} |\mathbf{R}|^{1/2} \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{y} - f(\mathbf{x})^t \boldsymbol{\beta})^t \mathbf{R}^{-1} (\mathbf{y} - f(\mathbf{x})^t \boldsymbol{\beta}) \right\}. \quad (2.22)$$

La log-vraisemblance est :

$$\mathcal{L}(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta} | \mathbf{y}) = -\frac{n}{2} \log(\sigma^2) - \frac{1}{2} \log(|\mathbf{R}|) - \frac{1}{2\sigma^2} (\mathbf{y} - f(\mathbf{x})^t \boldsymbol{\beta})^t \mathbf{R}^{-1} (\mathbf{y} - f(\mathbf{x})^t \boldsymbol{\beta}). \quad (2.23)$$

Or $\boldsymbol{\beta}$ et σ^2 sont inconnus. $\boldsymbol{\beta}$ est estimé en minimisant,

$$(\mathbf{y} - f(\mathbf{x})^t \boldsymbol{\beta})^t \mathbf{R}^{-1} (\mathbf{y} - f(\mathbf{x})^t \boldsymbol{\beta}), \quad (2.24)$$

par rapport à $\boldsymbol{\beta}$. Cette formule correspond au critère des moindres carrés généralisés classique puisque,

$$\mathbb{E}[\mathbf{y}] = f(\mathbf{x})^t \boldsymbol{\beta} \quad \text{Cov}[\mathbf{y}] = \sigma^2 \mathbf{R}. \quad (2.25)$$

Ensuite, on estime σ^2 en résolvant $\frac{\partial \mathcal{L}(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta} | \mathbf{y})}{\partial \sigma^2} = 0$, et on a :

$$\hat{\sigma}^2 = n^{-1} (\mathbf{y} - f(\mathbf{x})^t \hat{\boldsymbol{\beta}})^t \mathbf{R}^{-1} (\mathbf{y} - f(\mathbf{x})^t \hat{\boldsymbol{\beta}}). \quad (2.26)$$

La log-vraisemblance devient alors,

$$\mathcal{L}(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta} | \mathbf{y}) = -\frac{n}{2} \left(\log(\sigma^2 \cdot |\mathbf{R}|^{\frac{1}{n}}) + 1 \right). \quad (2.27)$$

D'après (2.27) et (2.20), on cherche $\boldsymbol{\theta}^*$ tel que :

$$\boldsymbol{\theta}^* = \min_{\boldsymbol{\theta}} \{ \psi(\boldsymbol{\theta}) = \sigma^2 \cdot |\mathbf{R}|^{\frac{1}{n}} \}. \quad (2.28)$$

D'un point de vue pratique, pour éviter les problèmes d'instabilité numérique, on décompose \mathbf{R} par la factorisation de Cholesky, $\mathbf{R} = \mathbf{C}\mathbf{C}^t$, et (2.26) devient,

$$\hat{\sigma}^2 = n^{-1} \| \mathbf{C}^{-1} (\mathbf{y} - f(\mathbf{x})^t \hat{\boldsymbol{\beta}}) \|_2^2, \quad (2.29)$$

et l'estimation de β s'obtient en minimisant,

$$\|C^{-1}(\mathbf{y} - f(\mathbf{x})^t \beta)\|_2^2. \quad (2.30)$$

Dépendant du choix de la fonction de corrélation choisie et du paramètre θ , la matrice \mathbf{R} peut être mal conditionnée. Afin de réduire l'effet des erreurs d'arrondis, on utilise une factorisation de Cholesky modifiée,

$$CC^t = \mathbf{R} + \nu \mathbf{I}_d \text{ où } \nu = (10 + n)\varepsilon_m, \quad (2.31)$$

où ε_m est la précision machine ($\varepsilon_m = 2^{-52} \simeq 2,22 \cdot 10^{-16}$).

Exemple d'application

Afin d'illustrer la méthode, nous utilisons un des cas présentés dans l'ouvrage de Santner *et al.* ([138], § 4.2.4). Cet exemple est intéressant, car sa simplicité nous permet dans la section 2.1.4 de mesurer l'apport de l'information des dérivées.

On pose,

$$y(x_1, x_2) = 2x_1^3 x_2^2, \quad (2.32)$$

sur $[-1, 1]^2$.

La surface de réponse correspondante est donnée sur la figure 2.1a. Les points noirs représentent les points d'expériences qui sont utilisés pour modéliser la réponse par un modèle de krigeage universel, modèle 2.1, avec une corrélation gaussienne anisotropique (équation 2.4 avec $q = 2$). La prédiction obtenue est donnée sur la figure 2.1b.

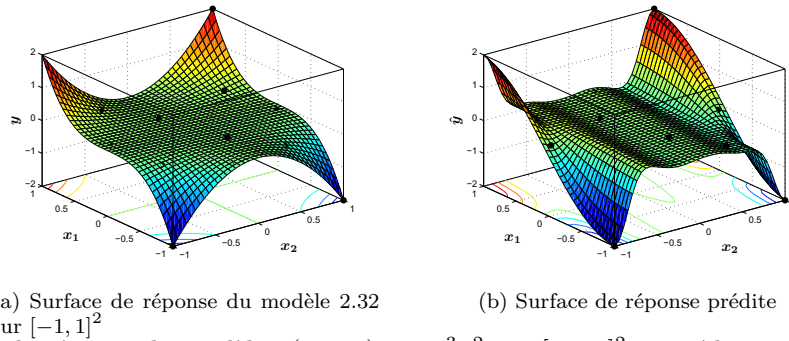


Figure 2.1 – Surface de réponse du modèle $y(x_1, x_2) = 2x_1^3 x_2^2$ sur $[-1, 1]^2$ et prédiction par krigeage universel avec corrélation gaussienne anisotropique à partir de 14 points d'expériences (points noirs)

On vérifie bien sur cet exemple que la prédiction est interpolante. Les portées obtenues pour chaque direction sont également intéressantes à analyser. Nous avons estimé $\theta_1 = 0.4$ et $\theta_2 = 12.8$. La faible valeur de θ_1 explique pourquoi selon x_1 la prédiction est moins bonne, puisque le champ d'influence des points est plus réduite. En revanche, la forme de la prédiction si l'on se place selon x_2 est plus proche de la surface de réponse, du fait de la valeur de θ plus élevée.

En bref

La prédiction quand la fonction de covariance est inconnue, mais paramétrique, s'écrit :

$$\hat{\mathbf{Y}}(\mathbf{x}) = f(\mathbf{x})^t \hat{\beta} + r^t \mathbf{R}^{-1}(\mathbf{Y} - f((S))^t \hat{\beta}), \quad (2.33)$$

où $\hat{\beta} = (f((S))^t \mathbf{R}^{-1} f((S)))^{-1} f((S))^t \mathbf{R}^{-1} \mathbf{Y}$ est l'estimateur des moindres carrés pondérés de β , et les prédictions \mathbf{R} et r sont données par un prédicteur de la fonction de corrélation $R(\cdot)$ (Santner *et al.* [138]).

L'erreur quadratique moyenne, notée EQM (ou MSE pour *Mean Square Error*), est définie par,

$$\sigma^2 (1 + (f((S))^t \mathbf{R}^{-1} r - f(\mathbf{x}))^t (f((S))^t \mathbf{R}^{-1} f((S))) (f((S))^t \mathbf{R}^{-1} r - f(\mathbf{x})) - r^t \mathbf{R}^{-1} r) \quad (2.34)$$

Pour plus de détails sur la théorie du krigeage, ses différentes variantes et les applications, de nombreux ouvrages existent. Nous citons les ouvrages complets de Fang *et al.* [57] et Santner *et al.* [138].

Le krigeage est une méthode intermédiaire entre des méthodes rigides telles que la régression, où l'on impose des fonctions, et des méthodes plus souples telles que les modèles additifs, reposant sur l'estimation de splines dans chaque direction.

2.1.2 Choix du noyau

Dans cette section, nous présentons les familles de covariances paramétrées classiques que nous avons été amenés à utiliser. Pour plus de détails, le lecteur pourra se reporter aux nombreux ouvrages traitant de ce sujet dans le domaine des statistiques notamment Yaglom [167], Matern [108], Ginsbourger [72] et Vazquez [156] et en géostatistique Cressie [35], Chilès et Delfiner [29] et Stein [147].

Nous considérons tout d'abord les covariances stationnaires usuelles, avant d'offrir une brève vision de l'utilisation des noyaux non stationnaires.

2.1.2.1 Familles classiques de covariances

Fonctions sphériques

Les expressions des fonctions de corrélation pour $d = 1, 2, 3$ et 5 sont respectivement, pour $h \leq a$,

$$r(\theta, h) = 1 - \frac{h}{\theta}, \text{ pour } d = 1, \quad (2.35)$$

$$r(\theta, h) = 1 - \frac{2}{\pi} \left(\frac{h}{\theta} \sqrt{1 - \left(\frac{h}{\theta}\right)^2} + \arcsin\left(\frac{h}{\theta}\right) \right), \text{ pour } d = 2, \quad (2.36)$$

$$r(\theta, h) = 1 - \frac{3h}{2\theta} + \frac{1}{2} \left(\frac{h}{\theta}\right)^3, \text{ pour } d = 3, \quad (2.37)$$

$$r(\theta, h) = 1 - \frac{15h}{8\theta} + \frac{5}{4} \left(\frac{h}{\theta}\right)^3 - \frac{3}{8} \left(\frac{h}{\theta}\right)^5, \text{ pour } d = 5. \quad (2.38)$$

Fonctions cubiques

Le modèle cubique possède un terme principal irrégulier en $|h|^3$. Une covariance cubique peut donc être utilisée pour modéliser un processus aléatoire différentiable en moyenne quadratique, mais n'est admissible qu'en dimension au plus égale à trois. Comme pour les covariances sphériques, la covariance cubique est nulle pour $h > \theta$. Pour $h \leq \theta$, la fonction de corrélation cubique est

$$r(\theta, h) = 1 - 7 \left(\frac{h}{\theta}\right)^2 + \frac{35}{4} \left(\frac{h}{\theta}\right)^3 - \frac{7}{2} \left(\frac{h}{\theta}\right)^5 + \frac{3}{4} \left(\frac{h}{\theta}\right)^7. \quad (2.39)$$

On peut construire sur ces modèles des covariances dérivables aux ordres supérieurs avec des termes principaux irréguliers en $|h|^{2p+1}$, $p \in \mathbb{N}$. Nous n'utilisons pas cette covariance en pratique car elle n'est utilisable qu'en dimension trois au plus.

Fonctions exponentielles

Les covariances exponentielles s'écrivent

$$r(\theta, h) = \sigma^2 \exp(-(\theta |h|)^\alpha), \text{ pour } 0 < \alpha \leq 2. \quad (2.40)$$

Un processus gaussien avec un noyau de covariance exponentiel, avec $\alpha = 1$, est également appelé processus d'Ornstein-Uhlenbeck. Pour $\alpha = 2$, on retrouve la covariance la plus utilisée en pratique, la covariance gaussienne, dont nous avons eu principalement recours lors de l'application des méthodes.

Fonctions de Matérn

Le modèle de Matérn (Matérn [108]) est défini par

$$r(\theta, h) = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{|h|}{\theta} \right)^\nu K_\nu \frac{|h|}{\theta}, \quad (2.41)$$

où $\nu \in]-1, +\infty]$ et K_ν est la fonction de Bessel modifiée de deuxième espèce de paramètre ν (Abramowitz et Stegun [2]). Ce modèle est apprécié pour la flexibilité apportée par le paramètre ν , qui permet de régler finement la régularité de r en 0, et donc la dérivabilité en moyenne quadratique des processus associés (Gaetan et Guyon [69]).

Citons enfin l'existence d'autres fonctions de covariance, telles que les noyaux circulaire ou puissance. Tous ces modèles sont détaillés et étudiés en détail dans le cadre de la géostatistique, notamment dans l'ouvrage de Cressie [35].

2.1.2.2 Noyaux de covariance non stationnaires

L'utilisation de noyaux stationnaires est relativement courante. Pourtant cette hypothèse forte, n'est pas forcément vérifiée en pratique. Dans ce cas, afin de respecter la structure des processus l'utilisation de noyaux non stationnaires semble plus indiquée.

Cependant, même si l'utilisation de noyaux de covariance non stationnaires ne pose pas de problème théorique majeur, il est rare d'en trouver dans les applications du krigeage. Les raisons de ce désintérêt sont principalement d'ordre numérique et d'optimisation. On citera tout de même les travaux de thèse de Ginsbourger [72] qui s'intéressent aux limites liées à l'hypothèse de stationnarité et l'article de Xiong et al. [166] qui propose un exemple d'application dans le domaine des plans d'expériences.

Nous nous bornons dans cette section à présenter deux types de noyaux de covariances, en dimension 1 et 2. Ces deux exemples permettent d'apprécier l'intérêt de ce type de noyau et les difficultés liées à son utilisation. L'objectif est ici de fournir un bref aperçu de ces noyaux, puisque nous avons été amenés à définir un tel type de noyau pour notre cas d'étude (section 3.2.4).

Dans les deux exemples, on se place dans le cas simple où l'on observe des résidus qui ne sont pas de variance constante en fonction de la, ou d'une des, variable(s). Le principe du noyau repose sur la volonté de modéliser les résidus localisés sur une zone bien précise.

Simulation en 1D

On se place dans le cas très simple, en une dimension, dans lequel on observe des résidus fortement localisés selon la température. Pour prendre en compte les observations, il est possible de définir un grand nombre de noyaux. Nous en présentons un en particulier pour illustrer la forme du noyau.

L'exemple proposé se rapproche de la forme des résidus que l'on a observée pour notre cas d'application. Ce choix se justifie puisque c'est ce type de noyau dont nous sommes partis lors de nos travaux, quand nous nous sommes penchés sur le choix d'un noyau non stationnaire.

Par conséquent, on se place dans cet exemple en 1D, dans le cas où l'on observe deux intervalles de température où les résidus sont peu variables, et un sur lequel les résidus sont fortement variables.

Quelques essais ont été réalisés en dimension sur $[-1, 1]$ avec le noyau suivant :

$$r(\theta, T, T') = \sigma^2 \exp\{-\theta \|T - T'\|^2\} f(T) f(T') \quad (2.42)$$

où $f(T) = \exp\{-\frac{\|T-m\|^2}{\delta}\}$ traduit la non stationnarité de la variance. Ce noyau relativement simpliste implique cependant l'ajout de deux paramètres δ et m , dont l'estimation est un problème difficile à traiter, et d'autant plus problématique, que la dimension du problème est grande.

La simulation d'un tel processus conduit aux réalisations exposées sur la figure 2.2, avec $m = 0$ et les autres paramètres fixés arbitrairement. On note que les simulations obtenues ont la même forme que les résidus observés dans le chapitre 1.

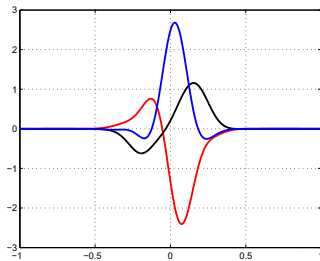


Figure 2.2 – Simulations d'un noyau de covariance non stationnaire en 1D

Évolution du noyau en 2D

En reprenant la même démarche, nous avons étendu le noyau au cas de dimension 2. On a donc deux variables, notées T et C , et la non-stationnarité n'est vérifiée que selon T .

L'extension directe en 2D du noyau précédent est

$$r(\theta, (T, C), (T', C')) = \sigma^2 \exp\{-\theta_T \|T - T'\|^2 - \theta_C \|C - C'\|^2\} f(T, C) f(T', C') \quad (2.43)$$

où

$$f(T, C) = \exp\left\{-\frac{\|T - m(T)\|^2}{\delta(T)}\right\} \quad (2.44)$$

avec $m(T) = \alpha_1 T + \alpha_2$ et $\delta(T) = \gamma_1 T + \gamma_2$. On se rend compte que pour ce noyau, on rajoute quatre paramètres qui sont à ajuster, les paramètres α et γ , qui caractérisent la non stationnarité.

La figure 2.3 illustre une simulation d'un noyau non stationnaire en T . Il est important de noter dans ce cas, que l'augmentation d'une dimension, entraîne l'ajout de deux paramètres supplémentaires.

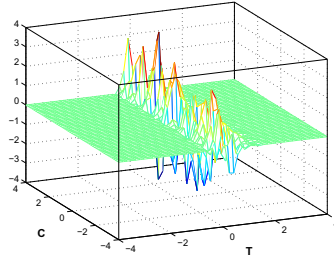


Figure 2.3 – Simulations d'un noyau de covariance non stationnaire en 2D

Ces deux exemples illustrent bien le type de noyau non stationnaire qu'il est possible de mettre en place. L'intérêt principal est de pouvoir être très représentatif des résidus que l'on observe. Dans la plupart des cas, ces noyaux rajoutent un certain nombre de paramètres dont l'estimation est un problème à part entière, ce qui peut expliquer que l'on ne trouve que très rarement des travaux utilisant ce type de noyau.

2.1.3 Krigeage avec bruit

Cette section présente les équations du krigeage lorsque l'on ajoute un bruit de mesure. On se place dans le cas du krigeage universel, dont on modifie le modèle (equation 2.1) en y ajoutant un terme modélisant le bruit de mesure,

$$Y(s) = f(s)^t \beta + Z(s) + \varepsilon, \quad (2.45)$$

où ε est un bruit gaussien, de moyenne nulle, d'écart-type σ_ε , indépendant de $Z(s)$. Pour faire la distinction, on note σ_z l'écart-type du processus Z .

Dans le cas où l'on souhaite modéliser un phénomène physique, les observations sont entachées d'un bruit de mesure dont il faut tenir compte pour éviter le phénomène de sur-ajustement (overfitting), quand le modèle a des variations trop fortes par rapport à la réalité. De plus, construire un interpolateur basé sur des observations bruitées n'aurait pas de sens, car répéter les observations en un même point donne des valeurs différentes. On va donc faire de l'approximation en les points de données, et non plus de l'interpolation comme précédemment.

Dans un premier temps, on suppose que les caractéristiques du processus sont connues. Le but est de présenter l'obtention des formules de prédiction et de variance de prédiction. Dans un second temps, nous exhibons la méthode d'estimation des paramètres par maximum de vraisemblance.

Formules de prédiction et de variance de prédiction : paramètres du processus connus

Les notations utilisées sont tirées de l'ouvrage de Santner *et al.* [138]. Soit x_1, \dots, x_n , n points en lesquels on observe la sortie y , que l'on note y_1, \dots, y_n . Comme dans le cas du krigeage (voir section 2.1.1), on cherche à construire un prédicteur linéaire de la forme

$$\hat{y}(x_0) = a_0 + \mathbf{a}^t \mathbf{Y} \quad (2.46)$$

où a_0 et \mathbf{a} sont à déterminer.

On pose les notations suivantes :

$$\mathbf{Y}^n = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \mathbf{Z}^n = \begin{pmatrix} Z(x_1) \\ Z(x_2) \\ \vdots \\ Z(x_n) \end{pmatrix}, \mathbf{F} = \begin{pmatrix} f(x_1) \\ f(x_2) \\ \vdots \\ f(x_n) \end{pmatrix}, \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}, \quad (2.47)$$

et \mathbf{R} la matrice de corrélation de \mathbf{Z}^n .

On sait que le prédicteur du krigeage est un prédicteur linéaire sans biais si, $\forall \boldsymbol{\beta}$,

$$f_0^t \boldsymbol{\beta} = a_0 + \mathbf{a}^t \mathbf{F} \boldsymbol{\beta}, \quad (2.48)$$

avec $f_0^t = f(x_0)$. Autrement dit, $a_0 = 0$ et $\mathbf{a}^t \mathbf{F} = f_0^t$.

La variance de prédicteur est donnée par

$$\begin{aligned} \mathbb{E} [(\mathbf{a}^t \mathbf{Y}^n - y_0)^2] &= \mathbb{E} [(\mathbf{a}^t (\mathbf{F} \boldsymbol{\beta} + \mathbf{Z}^n + \boldsymbol{\varepsilon}) - (f_0^t \boldsymbol{\beta} + Z_0 + \varepsilon_0))^2], \\ &= \mathbb{E} [((\mathbf{a}^t \mathbf{F} - f_0^t) \boldsymbol{\beta} + \mathbf{a}^t \mathbf{Z}^n - Z_0 + \mathbf{a}^t \boldsymbol{\varepsilon} - \varepsilon_0)^2], \\ &= \mathbb{E} [(\mathbf{a}^t \mathbf{Z}^n - Z_0 + \mathbf{a}^t \boldsymbol{\varepsilon} - \varepsilon_0)^2], \\ &= \sigma_z^2 (\mathbf{a}^t \mathbf{R} \mathbf{a} - 2\mathbf{a}^t \mathbf{r}_0 + 1) + \sigma_\varepsilon^2 (\mathbf{a}^t \mathbb{I}_n \mathbf{a} + 1), \\ &= \sigma_z^2 \left(\mathbf{a}^t \left[\mathbf{R} + \frac{\sigma_\varepsilon^2}{\sigma_z^2} \mathbb{I}_n \right] \mathbf{a} - 2\mathbf{a}^t \mathbf{r}_0 + 1 \right) + \sigma_\varepsilon^2. \end{aligned} \quad (2.49)$$

Les coefficients \mathbf{a} sont ceux qui minimisent la variance du prédicteur sous la contrainte que le prédicteur est non biaisé,

$$\mathbf{a}^t \left[\mathbf{R} + \frac{\sigma_\varepsilon^2}{\sigma_z^2} \mathbb{I}_n \right] \mathbf{a} - 2\mathbf{a}^t \mathbf{r}_0 \quad (2.50)$$

$$\text{s.c. } \mathbf{a}^t \mathbf{F} = f_0^t \quad (2.51)$$

Si l'on note \mathcal{R} la matrice définie par

$$\mathcal{R} = \mathbf{R} + \frac{\sigma_\varepsilon^2}{\sigma_z^2} \mathbb{I}_n, \quad (2.52)$$

alors le problème d'optimisation est le même que dans le cas de krigeage sans bruit (avec \mathcal{R} remplaçant \mathbf{R}).

Le prédicteur de krigeage est donc

$$\hat{y}(x_0) = f_0^t \boldsymbol{\beta} + \mathbf{r}_0^t \mathcal{R}^{-1} (\mathbf{Y}^n - \mathbf{F} \hat{\boldsymbol{\beta}}) \quad (2.53)$$

avec $\hat{\boldsymbol{\beta}}$ l'estimation par moindres carrés de $\boldsymbol{\beta}$ donnée par,

$$\hat{\boldsymbol{\beta}} = (\mathbf{F}^t \mathcal{R}^{-1} \mathbf{F})^{-1} \mathbf{F}^t \mathcal{R}^{-1} \mathbf{Y}^n. \quad (2.54)$$

En revanche, on remarque que la variance de prédiction, définie par

$$\sigma_z^2 (\mathbf{a}^t \mathcal{R} \mathbf{a} - 2\mathbf{a}^t \mathbf{r}_0 + 1) + \sigma_\varepsilon^2, \quad (2.55)$$

se distingue de celle calculée sans bruit, équation 2.10, par l'ajout de σ_ε^2 (avec le remplacement de \mathbf{R} par \mathcal{R}).

Estimation des paramètres du processus par maximum de vraisemblance

La log-vraisemblance s'écrit à une constante près (Forrester *et al.* [63])

$$\mathcal{L}(\boldsymbol{\beta}, \sigma_z^2, \boldsymbol{\theta}, \sigma_\varepsilon^2 | \mathbf{Y}^n) = -\frac{1}{2} \left[\log(|\sigma_z^2 \mathbf{R} + \sigma_\varepsilon^2 \mathbb{I}_n|) + (\mathbf{Y}^n - \mathbf{F}\boldsymbol{\beta})^t (\sigma_z^2 \mathbf{R} + \sigma_\varepsilon^2 \mathbb{I}_n)^{-1} (\mathbf{Y}^n - \mathbf{F}\boldsymbol{\beta}) \right]. \quad (2.56)$$

Il est à noter que le paramètre σ_ε^2 s'apparente à un terme de régularisation, la matrice $\sigma_z^2 \mathbf{R} + \sigma_\varepsilon^2 \mathbb{I}_n$ étant mieux conditionnée que la matrice $\sigma_z^2 \mathbf{R}$ apparaissant dans l'équation de la log-vraisemblance du krigeage sans bruit. D'ailleurs, pour améliorer le conditionnement des matrices de corrélation, les routines de krigeage telles que DACE (Lophaven *et al.* [97] [98]), utilisent cette technique (voir section 2.2.1, équation 2.31).

On note par la suite $\lambda = \sigma_z^2 / \sigma_\varepsilon^2$. En introduisant la matrice \mathcal{R} , la log-vraisemblance se réécrit

$$\mathcal{L}(\boldsymbol{\beta}, \sigma_z^2, \boldsymbol{\theta}, \lambda | \mathbf{Y}^n) = -\frac{1}{2} \left[n \log(\sigma_z^2) + \log(|\mathcal{R}|) + \frac{(\mathbf{Y}^n - \mathbf{F}\boldsymbol{\beta})^t \mathcal{R}^{-1} (\mathbf{Y}^n - \mathbf{F}\boldsymbol{\beta})}{\sigma_z^2} \right], \quad (2.57)$$

qui a une forme similaire à ce que l'on obtient dans le cas du krigeage sans bruit (équation 2.23).

En annulant les dérivées partielles par rapport à $\boldsymbol{\beta}$ et σ_z^2 , on obtient

$$\hat{\boldsymbol{\beta}} = (\mathbf{F}^t \mathcal{R}^{-1} \mathbf{F})^{-1} \mathbf{F}^t \mathcal{R}^{-1} \mathbf{Y}^n, \quad (2.58)$$

et

$$\hat{\sigma}_z^2 = \frac{1}{n} (\mathbf{Y}^n - \mathbf{F}\hat{\boldsymbol{\beta}})^t \mathcal{R}^{-1} (\mathbf{Y}^n - \mathbf{F}\hat{\boldsymbol{\beta}}). \quad (2.59)$$

En substituant ces deux estimations dans l'équation de log-vraisemblance, on a

$$\mathcal{L}(\boldsymbol{\beta}, \sigma_z^2, \boldsymbol{\theta}, \lambda | \mathbf{Y}^n) = -\frac{1}{2} \left[n \log(\hat{\sigma}_z^2) + \log(|\mathcal{R}|) + n \right], \quad (2.60)$$

qui ne dépend que de $\boldsymbol{\theta}$ et λ au travers de \mathcal{R} .

L'estimation par maximum de vraisemblance revient donc à

$$(\hat{\boldsymbol{\theta}}, \hat{\lambda}) = \arg \min_{\boldsymbol{\theta}, \lambda} \left[n \log(\hat{\sigma}_z^2) + \log(|\mathcal{R}|) \right]. \quad (2.61)$$

Dans ces travaux de thèse, Bettinger, [12], expose la construction d'un prédicteur qui n'est pas interpolant, mais dont l'erreur de prédiction est nulle aux points d'observation. Dans cette méthode, le prédicteur est obtenu par krigeage avec bruit de mesure. La variance de prédiction est calculée à partir des valeurs des paramètres de moyenne et de corrélation obtenues avec le modèle bruité, mais en utilisant la formule de variance de prédiction empirique sans bruit de mesure.

2.1.4 Cokrigeage

Nous nous intéressons dans ce paragraphe à la modélisation par krigeage dans le cas où l'on observe simultanément plusieurs réponses d'un même système, et l'on souhaite prendre en compte le fait que les réponses sont corrélées. Ce type de modélisation, appelée cokrigeage (Chilès et Delfiner, [29], Santner *et al.* [138]), utilise des processus vectoriels. Nous présentons dans un premier temps le modèle de cokrigeage avec dérivées, qui est un cas particulier du cokrigeage, ainsi que les formules de prédicteur en dimension 1 puis en dimension p , et illustrons la méthode sur un exemple.

La méthode appelée cokrigage permet d'effectuer une prédiction lorsque l'on observe plusieurs processus aléatoires $F_\alpha(\mathbf{x})$, $\alpha = 1, \dots, q$. On peut, par exemple, vouloir prédire $F_1(\mathbf{x})$ en fonction d'observations $F_\alpha(\mathbf{x}_i)$, $i = 1, \dots, n$. Le principe reste le même que dans le cas d'un seul processus aléatoire.

Soit $(\mathbf{Y}(\mathbf{x}))_{\mathbf{x} \in D}$ un processus gaussien de moyenne $\mathbb{E}(\mathbf{Y}(\mathbf{x})) = m(\mathbf{x})$ et de noyau (fonction de covariance) $\text{Cov}(\mathbf{Y}(\mathbf{x}_1), \mathbf{Y}(\mathbf{x}_2)) = r(\boldsymbol{\theta}, \mathbf{x}_1 - \mathbf{x}_2)$, où r est une forme bilinéaire symétrique positive.

Soit L un opérateur linéaire, alors $(L\mathbf{Y}(\mathbf{x}))_{\mathbf{x} \in D}$ est également un processus gaussien (sous certaines conditions d'existence). Par conséquent, si l'on prend l'opérateur de dérivée en dimension 1, alors

$$\left(\frac{dY}{dx}(\mathbf{x}) \right)_{\mathbf{x} \in D} \quad (2.62)$$

est un processus gaussien de moyenne $\mathbb{E} \left[\frac{dY}{dx}(\mathbf{x}) \right] = m'(\mathbf{x})$ (existe si m est dérivable) et de noyau (fonction de covariance) $\text{Cov} \left(\frac{dY}{du}(\mathbf{x}_1), \frac{dY}{dv}(\mathbf{x}_2) \right) = \frac{d^2 r}{dudv}(\boldsymbol{\theta}, \mathbf{x}_1 - \mathbf{x}_2)$ (existe si le noyau r est dérivable deux fois).

Soient $\mathbf{x}_1, \dots, \mathbf{x}_n$, n points du domaine où on connaît la valeur de la réponse, et $\mathbf{w}_1, \dots, \mathbf{w}_m$, m points du domaine où on connaît la valeur de la dérivée.

Le vecteur

$$\mathbf{Y}_{nm} = \begin{pmatrix} Y(\mathbf{x}_1) \\ \vdots \\ Y(\mathbf{x}_n) \\ \frac{dY}{dx}(\mathbf{w}_1) \\ \vdots \\ \frac{dY}{dx}(\mathbf{w}_m) \end{pmatrix}_{(n+m \times 1)}, \quad \text{de moyenne } \mathbf{M} = \begin{pmatrix} m(\mathbf{x}_1) \\ \vdots \\ m(\mathbf{x}_n) \\ m'(\mathbf{w}_1) \\ \vdots \\ m'(\mathbf{w}_m) \end{pmatrix}_{(n+m \times 1)}, \quad (2.63)$$

est aussi gaussien, de matrice de variance-covariance

$$\mathbf{R} = \begin{pmatrix} r(\boldsymbol{\theta}, \mathbf{x}_1 - \mathbf{x}_1) & \cdots & r(\boldsymbol{\theta}, \mathbf{x}_1 - \mathbf{x}_n) & \frac{dr}{dv}(\boldsymbol{\theta}, \mathbf{x}_1 - \mathbf{w}_1) & \cdots & \frac{dr}{dv}(\boldsymbol{\theta}, \mathbf{x}_1 - \mathbf{w}_m) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ r(\boldsymbol{\theta}, \mathbf{x}_n - \mathbf{x}_1) & \cdots & r(\boldsymbol{\theta}, \mathbf{x}_n - \mathbf{x}_n) & \frac{dr}{dv}(\boldsymbol{\theta}, \mathbf{x}_n - \mathbf{w}_1) & \cdots & \frac{dr}{dv}(\boldsymbol{\theta}, \mathbf{x}_n - \mathbf{w}_m) \\ \frac{dr}{du}(\boldsymbol{\theta}, \mathbf{w}_1 - \mathbf{x}_1) & \cdots & \frac{dr}{du}(\boldsymbol{\theta}, \mathbf{w}_1 - \mathbf{x}_n) & \frac{d^2 r}{dudv}(\boldsymbol{\theta}, \mathbf{w}_1 - \mathbf{w}_1) & \cdots & \frac{d^2 r}{dudv}(\boldsymbol{\theta}, \mathbf{w}_1 - \mathbf{w}_m) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{dr}{du}(\boldsymbol{\theta}, \mathbf{w}_m - \mathbf{x}_1) & \cdots & \frac{dr}{du}(\boldsymbol{\theta}, \mathbf{w}_m - \mathbf{x}_n) & \frac{d^2 r}{dudv}(\boldsymbol{\theta}, \mathbf{w}_m - \mathbf{w}_1) & \cdots & \frac{d^2 r}{dudv}(\boldsymbol{\theta}, \mathbf{w}_m - \mathbf{w}_m) \end{pmatrix}_{(n+m \times n+m)} \quad (2.64)$$

Plaçons-nous dans le cas d'un krigeage en dimension 1. On désire connaître la loi de la réponse au point \mathbf{x}_0 , c'est-à-dire la loi de $\mathbf{Y}(\mathbf{x}_0) | \mathbf{Y}_{nm}$. Il s'agit d'une loi normale de moyenne et de variance "habituelle", c'est-à-dire données par le krigeage simple :

$$\mathbb{E}(\mathbf{Y}(\mathbf{x}_0) | \mathbf{Y}_{nm}) = m(\mathbf{x}_0) - \mathbf{r}(\mathbf{x}_0)^t \mathbf{R}^{-1} (\mathbf{Y}_{nm} - \mathbf{M}) \quad (2.65)$$

$$\text{Var}(\mathbf{Y}(\mathbf{x}_0) | \mathbf{Y}_{nm}) = \mathbf{r}(\boldsymbol{\theta}, \mathbf{x}_0 - \mathbf{x}_0) - \mathbf{r}(\mathbf{x}_0)^t \mathbf{R}^{-1} \mathbf{r}(\mathbf{x}_0) \quad (2.66)$$

avec $\mathbf{r}(\mathbf{x}) = (r(\boldsymbol{\theta}, \mathbf{x}_1 - \mathbf{x}_0), \dots, r(\boldsymbol{\theta}, \mathbf{x}_n - \mathbf{x}_0), \frac{dr}{du}(\boldsymbol{\theta}, \mathbf{w}_1 - \mathbf{x}_0), \dots, \frac{dr}{du}(\boldsymbol{\theta}, \mathbf{w}_m - \mathbf{x}_0))^t$.

Plaçons maintenant dans un contexte élargi, i.e. en dimension p . La connaissance de la dérivée en un point implique la connaissance de p dérivées partielles. Le krigeage peut également se conduire avec une connaissance

de quelques dérivées partielles, c'est-à-dire ne connaissant pas forcément les p dérivées partielles pour chaque point.

On pose alors les notations suivantes :

$$\mathbf{R} = \begin{pmatrix} \mathbf{R}_{00} & \mathbf{R}_{01} & \cdots & \mathbf{R}_{0p} \\ \mathbf{R}_{01}^t & \mathbf{R}_{11} & \cdots & \mathbf{R}_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}_{0p}^t & \mathbf{R}_{1p}^t & \cdots & \mathbf{R}_{pp} \end{pmatrix} \text{ où } \mathbf{R}_{00} = \begin{pmatrix} r(\boldsymbol{\theta}, \mathbf{x}_1 - \mathbf{x}_1) & \cdots & r(\boldsymbol{\theta}, \mathbf{x}_1 - \mathbf{x}_n) \\ \vdots & \ddots & \vdots \\ r(\boldsymbol{\theta}, \mathbf{x}_n - \mathbf{x}_1) & \cdots & r(\boldsymbol{\theta}, \mathbf{x}_n - \mathbf{x}_n) \end{pmatrix} \quad (2.67)$$

avec

$$\forall j \in [1, p], \mathbf{R}_{0j} = \begin{pmatrix} \frac{\partial r}{\partial v_j}(\boldsymbol{\theta}, \mathbf{x}_1 - \mathbf{w}_1) & \cdots & \frac{\partial r}{\partial v_j}(\boldsymbol{\theta}, \mathbf{x}_1 - \mathbf{w}_m) \\ \vdots & \ddots & \vdots \\ \frac{\partial r}{\partial v_j}(\boldsymbol{\theta}, \mathbf{x}_n - \mathbf{w}_1) & \cdots & \frac{\partial r}{\partial v_j}(\boldsymbol{\theta}, \mathbf{x}_n - \mathbf{w}_m) \end{pmatrix} \quad (2.68)$$

et

$$\mathbf{R}_{ij} = \begin{pmatrix} \frac{\partial^2 r}{\partial u_i \partial v_j}(\boldsymbol{\theta}, \mathbf{w}_1 - \mathbf{w}_1) & \cdots & \frac{\partial^2 r}{\partial u_i \partial v_j}(\boldsymbol{\theta}, \mathbf{w}_1 - \mathbf{w}_m) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 r}{\partial u_i \partial v_j}(\boldsymbol{\theta}, \mathbf{w}_n - \mathbf{w}_1) & \cdots & \frac{\partial^2 r}{\partial u_i \partial v_j}(\boldsymbol{\theta}, \mathbf{w}_n - \mathbf{w}_m) \end{pmatrix}. \quad (2.69)$$

On pose également

$$\mathbf{Y}_{nmp} = \begin{pmatrix} Y(\mathbf{x}_1) \\ \vdots \\ Y(\mathbf{x}_n) \\ \frac{\partial Y}{\partial u_1}(\mathbf{w}_1) \\ \vdots \\ \frac{\partial Y}{\partial u_1}(\mathbf{w}_m) \\ \vdots \\ \frac{\partial Y}{\partial u_p}(\mathbf{w}_1) \\ \vdots \\ \frac{\partial Y}{\partial u_p}(\mathbf{w}_m) \end{pmatrix}, \mathbf{r}(\mathbf{x}) = \begin{pmatrix} r(\boldsymbol{\theta}, \mathbf{x}_1 - \mathbf{x}) \\ \vdots \\ r(\boldsymbol{\theta}, \mathbf{x}_n - \mathbf{x}) \\ \frac{\partial r}{\partial u_1}(\boldsymbol{\theta}, \mathbf{w}_1 - \mathbf{x}) \\ \vdots \\ \frac{\partial r}{\partial u_1}(\boldsymbol{\theta}, \mathbf{w}_m - \mathbf{x}) \\ \vdots \\ \frac{\partial r}{\partial u_p}(\boldsymbol{\theta}, \mathbf{w}_1 - \mathbf{x}) \\ \vdots \\ \frac{\partial r}{\partial u_p}(\boldsymbol{\theta}, \mathbf{w}_m - \mathbf{x}) \end{pmatrix} \text{ et } \mathbf{M} = \begin{pmatrix} m(\mathbf{x}_1) \\ \vdots \\ m(\mathbf{x}_n) \\ \frac{\partial m}{\partial u_1}(\mathbf{w}_1) \\ \vdots \\ \frac{\partial m}{\partial u_1}(\mathbf{w}_m) \\ \vdots \\ \frac{\partial m}{\partial u_p}(\mathbf{w}_1) \\ \vdots \\ \frac{\partial m}{\partial u_p}(\mathbf{w}_m) \end{pmatrix}. \quad (2.70)$$

On obtient alors les formules suivantes :

$$\mathbb{E}(\mathbf{Y}(\mathbf{x}_0) | \mathbf{Y}_{nmp}) = m(\mathbf{x}_0) - \mathbf{r}(\mathbf{x}_0)^t \mathbf{R}^{-1} (\mathbf{Y}_{nmp} - \mathbf{M}) \quad (2.71)$$

$$\text{Var}(\mathbf{Y}(\mathbf{x}_0) | \mathbf{Y}_{nmp}) = r(\boldsymbol{\theta}, \mathbf{x}_0 - \mathbf{x}_0) - \mathbf{r}(\mathbf{x}_0)^t \mathbf{R}^{-1} \mathbf{r}(\mathbf{x}_0) \quad (2.72)$$

Exemple d'application

Pour illustrer la méthode nous reprenons le cas présenté dans la section 2.1.1, tiré de l'ouvrage de Santner *et al.* [138]. On rappelle que la surface de réponse est définie sur $[-1, 1]^2$ (figure 2.1) par,

$$y(x_1, x_2) = 2x_1^3 x_2^2. \quad (2.73)$$

Les dérivées premières partielles de $y(\cdot)$ sont

$$y(x_1, x_2)^{(1)} = \frac{\partial y(x_1, x_2)}{\partial x_1} = 6x_1^2 x_2^2, \quad (2.74)$$

$$y(x_1, x_2)^{(2)} = \frac{\partial y(x_1, x_2)}{\partial x_2} = 4x_1^3 x_2. \quad (2.75)$$

La fonction de corrélation utilisée est un noyau de covariance gaussien anisotropique défini par,

$$r(\theta, h) = \exp\{-\theta_1 h_1^2 - \theta_2 h_2^2\}. \tag{2.76}$$

Cette première relation définit les corrélations entre les éléments de y . Afin de calculer les fonctions de corrélations entre les éléments de $y(\cdot)$ et $y^{(j)}(\cdot), j = 1, 2$, deux à deux, nous partons de la corrélation 2.76 que nous dérivons par rapport à chacune des composantes.

Au final, on obtient les formules de covariances suivantes :

$$\text{Cov} \left(y(x_1), y^{(j)}(x_2) \right) = -2\theta_j(x_{j,1} - x_{j,2})\sigma^2 r(\theta, x_1 - x_2) \tag{2.77}$$

$$\text{Cov} \left(y^{(j)}(x_1), y^{(j)}(x_2) \right) = (2\theta_j - 4\theta_j^2(x_{j,1} - x_{j,2})^2)\sigma^2 r(\theta, x_1 - x_2) \text{ pour } j = 1, 2. \tag{2.78}$$

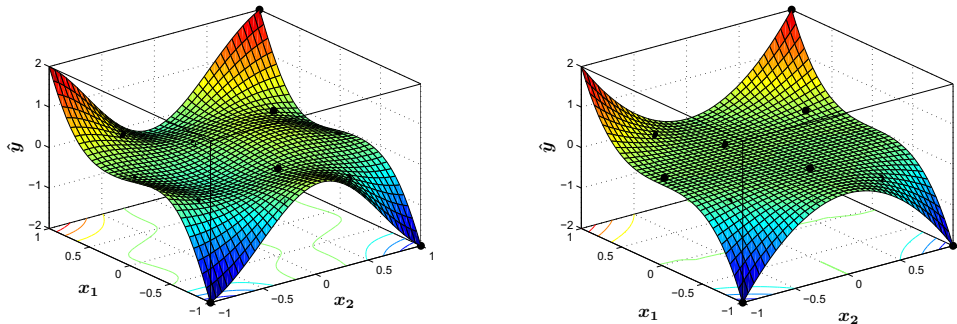
$$\text{Cov} \left(y^{(1)}(x_1), y^{(2)}(x_2) \right) = 4\theta_1\theta_2(x_{1,1} - x_{1,2})(x_{2,1} - x_{2,2})\sigma^2 r(\theta, x_1 - x_2) \tag{2.79}$$

Deux configurations sont testées. Premièrement, nous conditionnons notre krigeage par l’apport des dérivées partielles selon la première variable uniquement. Deuxièmement, nous conditionnons par les dérivées partielles des deux variables.

Pour montrer l’apport des dérivées, nous calculons et comparons pour chacune des trois prédictions (sans conditionnement, avec une seule dérivée, avec les deux dérivées) la racine de l’erreur quadratique moyenne empirique (ERMSPE pour *Empirical Root Mean Squared Prediction Error*), définie par

$$ERMSPE(\hat{y}) = \sqrt{\frac{1}{n_p} \sum_{i=1}^{n_p} (y(\mathbf{x}_i) - \hat{y}(\mathbf{x}_i))^2} \tag{2.80}$$

où n_p est le nombre de points en lesquels les prédictions sont effectuées (les mêmes pour les trois applications).



(a) Surface de réponse prédite avec conditionnement par $y^{(1)}(\cdot)$

(b) Surface de réponse prédite avec conditionnement par $y^{(1)}(\cdot)$ et $y^{(2)}(\cdot)$

Figure 2.4 – Prédiction par krigeage universel conditionné par une et deux dérivées partielles à partir de 14 points d’expériences (points noirs)

Les figures 2.4 correspondent aux prédictions obtenues. La figure 2.4a correspond au cas où seule la dérivée partielle selon la première variable est prise en compte, la figure 2.4b à celui où les deux dérivées partielles sont exploitées. Les points noirs représentent le plan d’expériences utilisé pour l’estimation des paramètres. Ce plan est le même que celui utilisé dans la section 2.1.1 (prédiction par krigeage universel).

Le tableau 2.1 résume les valeurs des $ERMSPE$ de chaque cas. La prédiction dans le cas d'un conditionnement par la dérivée partielle selon la première composante a un $ERMSPE$ plus faible, égale à 0.13 que sans conditionnement 0.35. La prédiction avec les deux dérivées partielles obtient une valeur encore plus faible de $2.7e-4$. Les prédictions obtenues sont d'une part meilleures visuellement parlant que la prédiction sans conditionnement (la seconde étant semblable à la vraie surface de réponse) et d'autre part, l'information selon une dérivée a permis de diminuer $ERMSPE$ de 63% et selon les deux dérivées de 99.9%. Cet exemple reflète bien l'intérêt de l'information des dérivées comme conditionnement du modèle de krigeage.

	Prédictions basées sur		
	$y(\cdot)$	$(y(\cdot), y^{(1)}(\cdot))$	$(y(\cdot), y^{(1)}(\cdot), y^{(2)}(\cdot))$
$ERMSPE$	0.35	0.13	$2.7e-4$

Tableau 2.1 – Valeurs des paramètres et des ERMSPE de chaque cas

En bref

Il peut être intéressant dans certaines situations de supposer des dérivées partielles (ou le gradient) connues en des points de l'espace des facteurs. Par exemple, on peut vouloir prendre en compte des connaissances a priori sur la physique du système, comme des conditions aux limites. Dans ce cas, nous pouvons considérer ces connaissances comme les observations des vecteurs aléatoires $\nabla F(y_j), j = 1, \dots, m$.

Dans les figures 2.4, nous utilisons à la fois des observations de la fonction et de sa dérivée. Ceci permet d'améliorer la prédiction. Il s'agit donc d'une autre possibilité pour incorporer de l'information connue *a priori*. Dans Morris et al., [118], ce problème est abordé du point de vue de la modélisation de résultats de simulation et de la planification d'expériences. On cite également les articles de Nerini *et al.*, [122] [123], qui proposent une application de cokrigeage dans un contexte similaire au nôtre.

Enfin, soulignons également que, dans ses travaux de thèse, Vazquez [156] présente plusieurs situations d'utilisation de cokrigeage avec l'information de la dérivée. La méthode permet de dériver des signaux bruités, ce qui est généralement considéré comme un problème délicat de traitement du signal, ou inversement d'intégrer les dérivées pour obtenir une approximation d'une fonction.

2.2 Données fonctionnelles

L'informatisation omniprésente de notre société est à l'origine d'un volume de données très important dont l'analyse et l'exploitation posent des problèmes complexes (volume de données, redondance dans les observations, structure des données). Dans l'optique de répondre à ces problèmes, l'analyse de données fonctionnelles devient un outil incontournable dans l'étude des cas rencontrés, par exemple, en météorologie, en médecine, en imagerie satellite et dans de nombreux autres domaines d'études, pour lesquels les objets manipulés sont des courbes ou des images. La modélisation et le traitement statistique de ces données ouvrent un champ très vaste de recherche visant des aspects théoriques et appliqués. L'intérêt porté à ces données au cours des dernières décennies n'a cessé de croître, offrant aujourd'hui, une littérature fournie sur ce domaine (Ramsay et Silverman [131]; Ferraty et Vieu [60]; Mizuta et Kato [114]; Hall et Horowitz [76]).

Le principe de base de la théorie et de l'analyse de données Fonctionnelles, (Ramsay et Silverman [131]; Ferraty et Vieu [60]), est de considérer que les phénomènes étudiés sont décrits par des fonctions plutôt que par des vecteurs de \mathbb{R}^n . En d'autres termes, chaque variable observée est à valeurs fonctionnelles plutôt que réelles.

Après avoir présenté ce que sont des données fonctionnelles, et donné un état de l'art des ouvrages de référence sur l'analyse de données fonctionnelles, nous introduisons le modèle que nous avons retenu pour répondre à notre problématique qui est un modèle à coefficients variables (VCM). Deux méthodes d'estimation des fonctions coefficients sont exhibées, premièrement les polynômes locaux, puis deuxièmement les P-splines. Enfin, le modèle de base retenu suppose une relation linéaire, dont l'hypothèse n'est pas forcément avérée. Pour vérifier le type de relation qui peut exister, le modèle additif est un outil intéressant, puisque sa structure même est proche des VCM. Nous introduisons donc cette théorie à laquelle nous avons eu recours.

2.2.1 Définition et état de l'art

Les évolutions en terme de puissance de calcul informatique ont ouvert la voie à des données de grande taille et de complexité accrue. Ces changements ont motivé le développement de méthodologies opérationnelles, dont le but est d'identifier des structures de données sous-jacentes, à moindres coûts. Les techniques d'analyse de données développées à partir de la théorie des statistiques montrent de bons comportements pour l'exploitation de structures cachées en grande dimension.

Parmi les nombreuses méthodes, on peut citer les modèles additifs (Breiman et Friedman, [17]; Hastie et Tibshirani, [80]), les modèles d'interaction sous-jacents (Friedman, [67]; Gu et Wahba, [74]; Stone *et al.*, [149]), les modèles partiellement linéaires (Green et Silverman, [73]), les modèles linéaires à coefficients variables (Cleveland *et al.*, [31]; Hastie et Tibshirani, [81]) et leurs hybrides (Carroll *et al.*, [24]; Fan *et al.*, [53]). L'intérêt de ces modèles est d'atténuer ce que l'on nomme le "fléau de la dimension" (appelé "*curse of dimension*" par les anglo-saxons) en explorant des structures sous-jacentes. De plus, ces différents modèles explorent différents aspects des données de grande dimension et incorporent des connaissances *a priori*. Par conséquent, en fonction des particularités des jeux de données étudiés certaines méthodes fonctionnent mieux que d'autres, et donc sont plus appropriées, mais aucune n'est globalement supérieure. Pour des connaissances générales sur les techniques de modélisation semi ou non paramétrique, nous renvoyons le lecteur aux ouvrages de Hastie et Tibshirani, [80], de Wahba, [158], de Green et Silverman, [73], et de Fan et Gijbels [52].

Dans leur ouvrage, Ferraty et Vieu, [60] définissent les données fonctionnelles de la façon suivante. Une variable aléatoire χ est appelée fonctionnelle si elle prend ses valeurs dans un espace de dimension infinie (ou espace fonctionnel). Une observation χ de χ est appelée une donnée fonctionnelle.

Pour mieux appréhender, l'émergence de ce type de données, considérons la situation dans laquelle une variable aléatoire peut être observée à différents instants, dans l'intervalle (t_{min}, t_{max}) . Une observation peut s'exprimer par la famille aléatoire $\{X(t_j)\}_{j=1, \dots, J}$. L'automatisation et l'informatisation des méthodes de mesure entraînent des grilles d'observations de plus en plus fines, indiquant que les observations sont de plus en plus proches. Une façon de tenir compte de ce phénomène est de considérer que les données sont des observations de la famille continue $\chi = \{X(t_j); t \in (t_{min}, t_{max})\}$. Avec la démocratisation et l'amélioration des puissances de calculs et de traitement des ordinateurs, on voit que ce type de données intervient dans quasiment tous les domaines de recherche, fournissant ainsi un volume de données de plus en plus grand.

Nous illustrons quelques cas, par les exemples suivants :

- l'étude du climat est un domaine particulièrement adapté à la modélisation fonctionnelle. Les objets observés présentent systématiquement une variabilité temporelle et/ou spatiale. Par exemple, les prédictions climatiques, compte tenu des variations de la température mensuelle moyenne au cours d'une année en différentes localisations. Chaque station météo est un "individu", décrit par une fonction, associant au moins la température moyenne observée (Ramsay et Silverman [131]). Une extension de cet exemple est l'observation d'une zone géographique avec deux échelles temporelles (Besse *et al.* [11]);

- les cas dans lesquels les individus étudiés présentent une variabilité temporelle. Par exemple, l'étude de la croissance d'enfants. Chaque individu (enfant) est modélisé par une fonction liant l'âge à la taille (Ramsay et Silverman, [131]). On cite également l'exemple présenté dans Abraham *et al.*, [1] qui étudie l'évolution du pH lors du processus d'acidification des fromages (individus) au cours du temps. On peut aussi citer Ferraty et Vieu, [60], (entre autres), ouvrage qui présente le problème de la consommation d'électricité mensuelle au Etats-Unis entre janvier 1973 et février 2001. Le nombre d'exemples disponibles dans la littérature, montre que la variabilité temporelle est extrêmement fréquente dans les données réelles. L'approche fonctionnelle permet, dans ces cas, de ne pas avoir recours à des simplifications souvent abusives. Elle permet aussi de traiter des données pour lesquelles les instants d'observation sont différents pour chaque individu ;
- les cas dans lesquels les individus étudiés présentent des variabilités spatiales. Par exemple, l'étude de l'évolution au cours du temps d'une ou plusieurs grandeurs observées en plusieurs emplacements, comme l'étude de l'évolution de la pluviométrie sur une zone donnée : un individu est décrit par la fonction associant des coordonnées géographiques au niveau de pluie reçu sur un temps donné ;
- enfin, l'ADF permet de répondre aux cas de tous les domaines dans lesquels on sait représenter une variabilité dans les observations caractérisant un individu sous la forme d'une ou plusieurs fonctions. Par exemple, des mesures spectroscopiques (des sources lumineuses caractérisées par leur spectre d'émission, Ferraty et Vieu, [60]) ou des courbes de réponse (en cinétique, comme notre cas). Autre exemple intéressant, celui des images, que l'on peut considérer comme une fonction liant des coordonnées dans \mathbb{R}^2 à un niveau de couleur (Morel et Solimini, [115]).

Bien souvent, l'ADF permet de résumer les données en remplaçant les observations réelles par une modélisation fonctionnelle simple, c'est-à-dire d'approcher les fonctions qui décrivent chaque individu par une représentation régulière, par exemple en projetant sur une base de splines ou d'ondelettes. Cette méthode permet, premièrement, de réduire le volume des données et, deuxièmement, de prendre en compte leur structure. Par exemple, il est possible de tenir compte d'informations *a priori* sur les individus pendant la phase de modélisation. La redondance dans les données, et plus généralement les liens entre les observations peuvent souvent être pris en compte sous forme de contraintes de régularité sur les fonctions manipulées. La phase de modélisation permet justement d'introduire ce type de contrainte : on peut par exemple représenter les individus par des fonctions C^2 .

Le livre de Ramsay et Silverman, [131], présente de façon assez synthétique et complète les méthodes les plus classiques en ADF. Cependant, cet ouvrage est avant tout orienté vers la mise en œuvre pratique des algorithmes et contient très peu de résultats théoriques. De plus, il ne contient bien entendu pas les développements les plus récents de l'ADF. En complément de Ramsay et Silverman, nous proposons les travaux de thèse de Conan-Guez, [32], qui offrent un panorama des méthodes de l'ADF, en mentionnant les résultats théoriques importants. Bien entendu, l'ADF forme un domaine trop large et trop actif pour prétendre à une quelconque exhaustivité.

En bref, cette bibliographie forcément restreinte nous permet de mieux appréhender le domaine de l'analyse fonctionnelle, et fournit surtout un éventail de méthodes relativement important. L'abondance de ce type de données mais surtout leur grande variété de caractéristique et de complexité font que chacune de ces méthodes a ses propres avantages (et donc ses propres limites), souvent directement hérités du modèle sous-jacent pour lequel la méthode a été construite. La section suivante s'attarde plus en détails sur les modèles à coefficients variables, puisque ce sont ces derniers que nous utilisons dans le chapitre 4.

2.2.2 Modèles à coefficients variables

Les modèles à coefficients variables (VCM pour *varying coefficient model*), appelés également modèle fonctionnel linéaire, ont été introduits par Hastie et Tibshirani, [81]. L'idée principale des VCM est de permettre aux coefficients de régression de varier (interagir) avec d'autres variables, générant ainsi des courbes de coefficients.

Ces fonctions coefficients peuvent, par exemple, refléter des changements de comportement en temps, en espace ou en température. L'estimation de ces fonctions coefficients nécessite généralement des algorithmes itératifs, c'est-à-dire des algorithmes estimant chacune des fonctions une à une, les lissant successivement, jusqu'à la convergence.

Dans notre cas, nous avons retenu deux méthodes d'estimation, l'utilisation des polynômes locaux cubiques, inspirée des travaux de Fan et Zhang [56], et le recours aux P-splines de Ruppert et Carroll [135], basées sur l'article de Fang *et al.* [57]. Nous avons eu recours également à la régression ridge à noyau (KRR pour *Kernel Ridge Regression*), introduite par Hastie *et al.* [82]. Mais avant de présenter ces trois outils, nous présentons plus en détails les VCM, en offrant une vue d'ensemble de cette théorie. Nous introduisons dans la section suivante, les aspects relatifs à l'estimation de ces fonctions coefficients.

2.2.2.1 Présentation

Le premier article abordant les VCM est celui de Cleveland *et al.* [31], qui propose une extension des techniques de régression locale monodimensionnelle au cas multidimensionnel. Mais la paternité de ces modèles est unanimement accordée à Hastie et Tibshirani [81], qui ont clairement défini ces modèles et leurs aspects théoriques.

Depuis, les VCM ont été appliqués avec succès à la régression non paramétrique multidimensionnelle, aux modèles linéaires généralisés, aux modèles non-linéaires de séries chronologiques, l'analyse de données longitudinales et fonctionnelles, pour le calcul des taux d'intérêt en finance, pour l'étude des conflits internationaux en sciences politiques et autres.

Soient une variable aléatoire Y et les prédicteurs X_1, \dots, X_p et U_1, \dots, U_p . Le modèle de base proposé par les auteurs est

$$Y = a_0 + X_1 a_1(U_1) + \dots + X_p a_p(U_p) + \varepsilon. \quad (2.81)$$

Les variables U_1, \dots, U_p changent les coefficients de X_1, \dots, X_p , au travers des fonctions (inconnues) a_0, \dots, a_p . L'idée de base consiste à approcher une fonction de régression inconnue multidimensionnelle par un modèle linéaire (conditionnel) avec des coefficients fonctions d'une des variables, appelée en anglais "*index*". Dans la plupart des travaux cette covariable, ou variable longitudinale, est le temps, l'espace ou la température. Cependant, dans certains cas d'application, cette dernière est difficilement identifiable. Des méthodes telles que les VCM adaptatifs permettent d'identifier ces dernières (Fan *et al.*, [54]). Nous ne présentons pas cet aspect, et nous bornons au cas où la variable longitudinale est clairement identifiée.

Si la fonction $a_j(U_j) = a_j$, c'est-à-dire si a_j est une fonction constante, alors ce terme est linéaire en X_j . Si tout les termes sont linéaires, alors le modèle (2.81) est le modèle linéaire classique. Par contre, si X_j est une constante (par exemple $X_j = 1$), alors le $j^{\text{ème}}$ terme correspond à $a_j(U_j)$, une fonction inconnue de U_j . Si tous les termes ont cette forme, ou sont linéaires, alors le modèle (2.81) correspond un modèle additif, et plus largement aux modèles additifs généralisés (GAM). Les GAM sont donc un cas particulier des VCM, comme étant le modèle linéaire dynamique généralisé (West *et al.* [160]). Cependant, l'estimation et l'inférence sont différentes dans ce dernier.

Dans la plupart des cas (comme pour nous), les U_j correspondent aux mêmes variables, un facteur tel que l'âge, le temps, l'espace (la température pour nous), dont on suppose qu'il modifie les effets de X_1, \dots, X_p . Le modèle ainsi obtenu, et auquel nous avons eu recours est

$$Y = \sum_{j=1}^p a_j(U) X_j + \varepsilon, \quad (2.82)$$

pour les covariables $(U, X_1, \dots, X_p)^T$ et la réponse Y avec

$$\mathbb{E}[\varepsilon|U, X_1, \dots, X_p] = 0 \text{ et } \text{Var}[\varepsilon|U, X_1, \dots, X_p] = \sigma^2(U). \quad (2.83)$$

D'autres cas de figure, dérivés du modèle (2.81), sont introduits dans l'article de Hastie et Tibshirani, [81], mais nous nous bornons au modèle (2.82), qui correspond à notre cas d'étude, et à la majeure partie des cas rencontrés en pratique. Il est toutefois intéressant de préciser que les U_j peuvent être des scalaires ou des vecteurs.

Bien que la forme de ce modèle semble restrictive, son domaine d'application est relativement vaste. Par exemple, on peut citer les travaux de Hoover *et al.*, [86], Brumback et Rice, [18], Wu *et al.*, [165] et Fan et Zhang, [56] pour des applications sur des données longitudinales, et Chen et Tsay, [27], et Cai *et al.*, [20], pour des applications sur des séries temporelles non linéaires.

En supposant, que les fonctions coefficients, $a_j(\cdot)$, possèdent le même degré de régularité, Hastie et Tibshirani proposent une estimation de ces fonctions à l'aide des moindres carrés pénalisés (Wahba [158]) sur un modèle linéaire dynamique (West *et al.* [160]). Par la suite, une méthode en deux étapes a été proposée par Fan et Zhang, [56], pour les cas où ces fonctions coefficients admettent différents degrés de régularité. C'est une des deux approches que nous avons utilisées dans le chapitre 4.

Une fois les fonctions coefficients estimées, $\hat{a}_j(\cdot)$, la prédiction en un nouveau point du domaine x^0 , est directe et s'obtient par

$$\hat{y}(U) = \sum_{j=1}^p \hat{a}_j(U) x_j^0. \quad (2.84)$$

Il est important de préciser que plusieurs cas de figure peuvent se rencontrer concernant la structure des différentes variables. Premièrement, le cas le plus fréquent, les variables X_j peuvent être des données fonctionnelles (fonction de la variable U). Par exemple, dans l'article de Fan et Zhang [56], l'étude porte sur le nombre d'entrées au service des problèmes respiratoires de l'hôpital de Hong-Kong, sur une période de un an, à partir des taux de certains polluants présents dans l'air mesurés quotidiennement sur la même période. Cependant, comme on le voit dans le modèle (2.82), il est également possible d'avoir des scalaires comme valeurs. C'est notre cas d'étude, qui, pour des conditions opératoires (concentration, débit, température initiale) fixes pendant une expérience, consiste à étudier la concentration des polluants en sortie du système.

Une des questions primordiales lors de l'application de ce type de modèle est de déterminer si certaines des fonctions coefficients ne sont pas constantes. En effet, dans des cas d'études où le nombre de variables est élevé, la détection des variables importantes peut être une tâche fastidieuse. L'article de Leng, [94] propose une méthode de vraisemblance pénalisée dans le cadre des modèles de spline de lissage ANOVA, avec une pénalité conçue pour distinguer automatiquement les fonctions coefficients variables et celles qui ne le sont pas. Contrairement à la procédure par étapes introduite par la suite, la méthode quantifie et estime les fonctions coefficients. Un algorithme efficace est introduit et la technique de choix des paramètres de lissage est largement discutée.

Le biais et la variance conditionnels des estimateurs sont dérivés dans les articles de Carroll *et al.*, [25], et Fan et Zhang, [56]. Comme attendu, le biais dépend uniquement des erreurs d'approximation locales, et la variance dépend du nombre effectif de points de données locaux, de la variance conditionnelle locale et de la matrice de corrélation locale des covariables \mathbf{X} . Nous ne détaillons pas plus cet aspect de la méthode des VCM, notre démarche ne nous ayant pas conduit à avoir besoin de ces spécificités.

2.2.2.2 Méthodes d'estimations

La régression analytique est une des techniques les plus utilisées en statistiques. Quand la dimension des variables indépendantes est grande, il est très difficile de conduire une analyse non-paramétrique efficace à partir des données. Alternative intéressante aux modèles additifs et autres modèles non paramétrique, les VCM peuvent réduire le biais de modélisation et prévenir, ce que l'on appelle, le "fléau de la dimension".

Il existe différentes méthodes pour l'estimation des fonctions coefficients des VCM, telles que, les méthodes à noyau (Wu *et al.*, [165]), les polynômes locaux (Cai *et al.*, [19] [20]), l'approximation par des fonctions de base (Huang *et al.*, [88], Lu et Mao, [99]), les splines de lissage (Chiang *et al.*, [28], Hoover *et al.*, [86]), la kernel ridge regression (Hastie *et al.*, [82]) et ainsi de suite. Bien que toutes ces approches pour l'ajustement des VCM aient démontré de bons comportements, certains problèmes persistent. Par exemple, le calcul de l'approximation de fonction de base est facile, mais la normalité asymptotique est difficile à établir. Pour les splines de lissage, les multiples paramètres de lissage peuvent être difficiles à estimer. Pour les méthodes à noyau et locales linéaires, toutes les fonctions coefficients sont estimées en utilisant une seule taille de fenêtre.

L'intérêt des VCM est que les fonctions coefficients peuvent être facilement estimées via une simple régression locale. Dans ce contexte, Fan et Zhang [56], présentent un algorithme efficace en deux étapes via l'utilisation des polynômes locaux cubiques. Cette méthode permet d'autoriser des largeurs de fenêtres différentes selon les directions, mais également des régularités différentes pour chacune des fonctions coefficients. Nous avons également fait le choix de nous inspirer de l'article de Fang *et al.*, [57], comme seconde méthode d'estimation. La variable longitudinale est discrétisée. Ainsi, en chacun de ces points, on obtient un modèle de régression linéaire, dont les coefficients sont estimés par moindres carrés. Des travaux ont montré, Li *et al* [95], qu'une deuxième étape de lissage peut être pertinente pour améliorer la modélisation, surtout lorsque les fonctions coefficients obtenues présentent des fluctuations.

Polynômes locaux

L'idée principale de l'estimation par polynômes locaux est d'approcher les fonctions coefficients, a , par des polynômes d'ordre q localement autour de chaque point x , à savoir

$$a(z) \approx \sum_{j=1}^q b_j(z - x)^j \tag{2.85}$$

pour z dans un voisinage de x . Les coefficients de ce polynôme sont alors estimés à partir des observations $(X_i, Y_i)_{i=1, \dots, n}$ en résolvant le problème de moindres carrés pondérés

$$\min_b \sum_{i=1}^n \left(Y_i - \sum_{j=0}^q b_j(X_i - x)^j \right)^2 K_h(X_i - x) \tag{2.86}$$

où $K_h(\cdot)$ est une fonction appelée noyau et h un paramètre d'échelle. On peut remarquer la différence avec la régression polynomiale classique due au rajout du terme $K_h(X_i - x)$, terme correctif qui permet de sélectionner les observations X_i autour de x qui vont intervenir dans le calcul du polynôme (d'où le terme local).

Les travaux basés sur ce type d'estimation des fonctions coefficients sont nombreux. Citons tout d'abord l'article de Fan et Zhang [56], dont nous nous inspirons. La procédure proposée par les auteurs est en deux étapes, dans laquelle chaque fonction coefficient est estimée par un polynôme local d'ordre 3. Dans cette approche, le choix des fenêtres de recherche h est crucial. Pour cette question, les articles de Zhang et Lee, [170] [169], permettent de mettre en place une méthode d'estimation pertinente et complémentaire à la méthode envisagée.

Première étape : estimation initiale des fonctions coefficients

Soit un échantillon aléatoire $\{(T_i, X_{i1}, X_{i2}, \dots, X_{ip}, Y_i)\}_{i=1}^n$. Une approche simple pour estimer les fonctions coefficients $(a_j(\cdot))_{j=1}^p$ est d'utiliser un modèle linéaire local, ce qui nous conduit au problème local des moindres carrés local : Minimiser

$$\sum_{i=1}^n \left[Y(T_i) - \sum_{j=1}^p \{a_j + b_j(T_i - t_0)\} X_{ij} \right]^2 K_h(T_i - t_0) \quad (2.87)$$

pour un noyau donné K et une fenêtre h , où $K_h(\cdot) = K(\cdot/h)/h$. Mais cette approche implique implicitement que les fonctions coefficients possèdent le même ordre de régularité, et qu'elles peuvent être estimées également dans le même intervalle, ce qui peut ne pas être le cas, entraînant ainsi des estimateurs sous optimaux.

Si l'on suppose que $a_p(\cdot)$ est plus lisse que les autres fonctions, plus précisément une dérivée quatrième bornée, il est possible de l'approcher localement par une fonction cubique. Cela nous conduit au problème des moindres carrés pondérés suivant :

$$\sum_{i=1}^n \left[Y_i - \sum_{j=1}^{p-1} \{a_j + b_j(T_i - t_0)\} X_{ij} - \{a_p + b_p(T_i - t_0) + c_p(T_i - t_0)^2 + d_p(T_i - t_0)^3\} X_{ip} \right]^2 K_{h_1}(T_i - t_0) \quad (2.88)$$

Seconde étape : estimation des fonctions coefficients par un polynôme local cubique

Le premier estimateur obtenu est appelé l'estimateur en une étape (one-step estimator) et il est possible de montrer que la vitesse de convergence n'est pas optimale avec cette procédure. Une procédure en deux étapes est nécessaire (two-step estimator). L'idée est que dans une première étape, on utilise une régression linéaire locale pour obtenir une estimation initiale des fonctions coefficients, on obtient les $\hat{a}_{1,1}(t_0), \dots, \hat{a}_{p,1}(t_0)$. Ensuite, dans une seconde étape, on remplace les $p - 1$ premières fonctions coefficients par leurs estimations pour estimer par un polynôme local cubique, la fonction coefficient $a_p(t_0)$, c'est-à-dire minimiser

$$\sum_{i=1}^n \left[Y_i - \sum_{j=1}^{p-1} \hat{a}_{j,1}(T_i) X_{ij} - \{a_p + b_p(T_i - t_0) + c_p(T_i - t_0)^2 + d_p(T_i - t_0)^3\} X_{ip} \right]^2 K_{h_2}(T_i - t_0) \quad (2.89)$$

selon a_p, b_p, c_p, d_p , où h_2 est la fenêtre déterminée dans la deuxième étape.

D'un point de vue pratique, le choix de la fenêtre initiale h_1 n'est pas primordial dans la procédure en deux étapes. Nous reviendrons sur ce point par la suite, mais il est à noter que la seule condition à remplir pour cette fenêtre initiale est qu'elle soit suffisamment petite pour que le biais dans la première étape soit négligeable.

Un avantage de cette procédure est que dans la deuxième étape, le problème est réellement un problème de lissage univarié, de sorte que le problème de sélection de fenêtre devient relativement simple.

Notations utilisées dans le cas classique

Avant d'introduire les formules pour l'estimateur en deux étapes, il est nécessaire d'introduire quelques notations.

$$X_1 = \begin{pmatrix} X_{11} & X_{11}(T_1 - t_0) & \cdots & X_{1p} & X_{1p}(T_1 - t_0) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ X_{n1} & X_{n1}(T_n - t_0) & \cdots & X_{np} & X_{np}(T_n - t_0) \end{pmatrix}_{(n,2p)}, \tag{2.90}$$

$$Y = (Y_1, \dots, Y_n)^t, \quad \text{et} \quad W_1 = \text{diag}(K_{h_1}(T_1 - t_0), \dots, K_{h_1}(T_n - t_0)). \tag{2.91}$$

Ainsi, la solution du problème de moindres carrés pour l'obtention des estimations initiales des fonctions coefficients est donnée par

$$\hat{a}_{j,1}(t_0) = e_{2j-1,2p}^t (X_1^T W_1 X_1)^{-1} X_1^t W_1 Y, \quad j = 1, \dots, p \tag{2.92}$$

où $e_{k,m}$ dénote le vecteur unitaire de longueur m avec 1 à la $k^{\text{ème}}$ position.

On pose également,

$$X_2 = \begin{pmatrix} X_{1p} & X_{1p}(T_1 - t_0) & X_{1p}(T_1 - t_0)^2 & X_{1p}(T_1 - t_0)^3 \\ \vdots & \vdots & \vdots & \vdots \\ X_{np} & X_{np}(T_n - t_0) & X_{np}(T_n - t_0)^2 & X_{np}(T_n - t_0)^3 \end{pmatrix}_{(n,4)}, \tag{2.93}$$

ainsi que,

$$W_2 = \text{diag}(K_{h_2}(T_1 - t_0), \dots, K_{h_2}(T_n - t_0)), \quad \text{et} \quad V = (V_1, \dots, V_n)^t, \tag{2.94}$$

où $V_i = \sum_{j=1}^{p-1} \hat{a}_{j,1}(T_i) X_{ij}$.

Ainsi, l'estimateur en deux étapes est donné par

$$\hat{a}_{p,2}(t_0) = (1 \ 0 \ 0 \ 0)(X_2^t W_2 X_2)^{-1} X_2^t W_2 (Y - V) \tag{2.95}$$

Les auteurs de cette méthode exposent également dans leur article les propriétés asymptotiques de l'estimateur. Ils ont signalé que cette méthode était plus performante (en terme de calcul) que les fonctions splines de lissage et plus souple que les méthodes à noyau classique. Pour la comparaison, il est nécessaire de comparer les résultats présentés avec le biais et la variance asymptotique de la méthode à noyau, étudiés par Hoover *et al.* [86]. En outre, Wu *et al.* [165] ont proposé des approches pour la construction des régions de confiance pour les méthodes à noyau. En ce qui concerne les intervalles de confiance pour la méthode que nous venons de présenter, ces derniers sont calculés lors de la deuxième étape de régression locale cubique. Nous renvoyons le lecteur à l'ouvrage de Fan et Gijbels, [52], (section 4.3), pour la méthode d'estimation des erreurs standards dans le contexte de la régression locale polynômiale.

P-splines

La seconde méthode d'estimation utilisée est issue de la théorie des données longitudinales (Singer et Willet [146]). La variable longitudinale est discrétisée. Ainsi, en chacun de ces points, on obtient un modèle de régression linéaire, dont les coefficients sont estimés par moindres carrés. Les travaux de Li *et al* [95], montrent l'intérêt d'une deuxième étape de lissage. Cet intérêt est d'autant plus important que les fonctions coefficients obtenues présentent des fluctuations. En supposant un nombre très élevé de nœuds fixes, il est montré que les estimateurs obtenus sont consistants et asymptotiquement normaux. Plus récemment, l'article de Lu *et al.*, [100], reprend cette démarche, en développant un algorithme d'optimisation systématique pour la sélection des multiples paramètres de lissage.

Nous proposons dans cette section de donner un aperçu des P-splines, outil que nous avons utilisé pour l'étape de lissage. La procédure mise en œuvre dans le chapitre 4 est inspirée de l'article de Li, Sudjianto et Zhang, [95]. Le principe de la méthode implémentée comporte deux étapes, avec tout d'abord, l'estimation des fonctions coefficients point à point selon la variable longitudinale, puis une étape de lissage à l'aide de P-splines.

Le lissage des fonctions coefficients est opéré par des splines de lissage, et plus particulièrement des P-splines. Eilers et Marx, [45] introduisent les P-splines comme une méthode de lissage pour les modèles linéaires généralisés. Cette méthode a plusieurs intérêts dont sa flexibilité, sa facilité d'application ainsi que le lien entre spline de lissage et régression polynômiale.

Divers articles sur différentes utilisations des P-splines ont été publiés par ces auteurs. Par exemple, Eilers et Marx, [46] et Marx et Eilers, [104] dans lesquels sont présentés une estimation d'une surface bidimensionnelle basée sur le produit tensoriel de P-splines. D'autres exemples d'application sont également disponibles dans la littérature tels que Currie *et al*, [36], Muggeo et Ferrara, [119], et Li *et al*, [95].

La théorie des P-spline de Eilers et Marx se compose d'un grand nombre de splines équidistantes, plus exactement des B-splines, où la variabilité des coefficients associés est contrôlée par une pénalité. Cet ensemble de B-splines est nommé base d'expansion. Une série de B-splines est placée sur une grille à intervalles réguliers, les nœuds, entre les valeurs minimales et maximales de la variable. Du fait de l'estimation d'un grand nombre de coefficients, bien souvent les données sont sur-ajustées. Pour palier à cela, un paramètre de pénalité est utilisé. Ce paramètre, choisi par l'utilisateur, contrôle la régularité de la forme fonctionnelle.

Les coefficients β des P-splines sont calculés en utilisant l'équation de régression modifiée

$$\beta = (\mathbf{B}^t \mathbf{B} + \lambda \mathbf{D}^t \mathbf{D})^{-1} \mathbf{B}^t \mathbf{y} \quad (2.96)$$

où \mathbf{B} est la matrice des bases de B-splines, λ est le paramètre de pénalité et \mathbf{D} est une matrice différentielle des coefficients, généralement de premier ou second ordre. Pour ce qui est des B-splines, il s'agit de bases locales de fonction, constituées de segments polynômiaux de faible degré, généralement quadratique ou cubique.

Pour une pénalité trop faible, on a sur-ajustement des données. En revanche, une trop grande pénalité ne permet pas un bon ajustement aux données. On s'aperçoit donc que dans cette approche, l'estimation du paramètre de lissage est cruciale.

Les travaux de Ruppert et Carroll [135] autorisent la pénalité à varier spatialement afin de s'adapter à l'hétérogénéité de la fonction de régression. Cette particularité semblant plus pertinente dans notre cas, nous avons décidé d'avoir recours à la théorie des P-splines introduite par ces auteurs. Contrairement, aux P-splines de Eilers et Marx dans lesquelles les nœuds sont uniformément répartis et des bases de B-splines sont utilisées, Ruppert et Carroll proposent d'utiliser les quantiles de la distribution comme nœuds et la base de fonction puissance tronquée.

Pour illustrer la méthode, plaçons nous dans le cas du lissage de la fonction coefficients pour la $j^{\text{ème}}$ direction estimée. Soient $\mathbf{A} = (a_j(t_1), \dots, a_j(t_m))^t$ et \mathbf{B} la "matrice d'expériences" pour la spline de régression telle que la $i^{\text{ème}}$ ligne de \mathbf{B} est définie par

$$\mathbf{B}_i = [1, t_i, \dots, t_i^r, (t_i - \kappa_1)_+^r, \dots, (t_i - \kappa_K)_+^r], \quad (2.97)$$

où $\kappa = (\kappa_1, \dots, \kappa_K)$ correspond aux K nœuds, et r est la puissance choisie. En pratique, une puissance de $r = 3$ est relativement bonne.

Soit $D(\lambda^*)$ une matrice diagonale pour laquelle les $(r + 1)$ premiers éléments de la diagonale sont nuls, et les autres restants sont $\lambda^*(\kappa_1), \dots, \lambda^*(\kappa_K)$. On a donc les coefficients $\beta(\lambda^*)$ des P-splines qui sont donnés par

$$\beta(\lambda^*) = (B^t B + D(\lambda^*)) B^t A. \quad (2.98)$$

Traditionnellement, la sélection des paramètres de lissage λ^* est basée sur un critère de type validation croisée généralisée (GCV) (Craven et Wahba [34]) ou critère d'information de Akaike (AIC) (Akaike [3]) et la minimisation du critère choisi est effectuée sur une grille de recherche.

On a ainsi

$$\hat{A} = B(B^t B + D(\lambda^*)) B^t A. \quad (2.99)$$

Soit $P_\lambda = B(B^t B + D(\lambda)) B^t$. Cette matrice peut être vue comme une matrice de projection et

$$e(\lambda) = \text{tr}(P_\lambda), \quad (2.100)$$

comme le degré de liberté (ou le nombre effectif de paramètres) du modèle ajusté.

Dans notre cas, la sélection du paramètre de lissage a été effectuée par minimisation de GCV. Soit

$$\lambda^* = \min_{\lambda} GCV(\lambda) \quad (2.101)$$

avec

$$GCV(\lambda) = \frac{1}{m} \frac{\|A - B\beta(\lambda)\|^2}{(1 - e(\lambda)/m)^2} \quad (2.102)$$

Kernel Ridge Regression

La Kernel Ridge Regression (KRR) introduite par Hastie *et al.*, [82], est largement utilisée notamment dans le domaine de la reconnaissance automatique d'image. On peut citer entre autres les travaux de An *et al* [5], Georgiades *et al* [71], ou encore Everingham et Zisserman [50]. Cette méthode appartenant au domaine de la classification (Breiman *et al* [17]), est une méthode de régression par moindres carrés régularisés. La version linéaire est similaire à la méthode du discriminant pour la classification (Zhang et Ma [168]) et dans le cas non-linéaire, au Support Vector Machine (SVM) (Vapnik [155]).

Afin de présenter cette méthode, dans un premier temps, nous introduisons la ridge regression comme une alternative à la méthode des moindres carrés lorsque les données sont mal conditionnées.

L'estimateur des moindres carrés, $\hat{\beta}_{MC}$, est défini par $\hat{\beta}_{MC} = \arg \min_{\beta} \|y - X\beta\|^2$ et s'obtient en résolvant l'équation normale $(X^t X)^{-1} X^t y$. Classiquement, on mesure la précision d'un estimateur $\hat{\beta}$ de β par son erreur quadratique moyenne (EQM), qui se compose de deux termes. Le premier terme correspond au carré du biais de l'estimateur tandis que le second terme correspond au terme de variance de l'estimateur. D'après le théorème de Gauss-Markov, l'estimateur des moindres carrés, $\hat{\beta}_{MC}$, est de tous les estimateurs sans biais celui de variance minimale. Il s'ensuit qu'on ne peut diminuer la variance de l'estimateur qu'au détriment d'estimateur biaisé. C'est l'objectif central des estimateurs régularisés.

Dans le cas de p variables, l'erreur quadratique moyenne de $\hat{\beta}_{MC}$ est définie par :

$$EQM(\hat{\beta}_{MC}) = \sum_{i=1}^p \mathbb{E} [(\hat{\beta}_i - \beta_i)^2] = \sum_{i=1}^p \text{Var}(\hat{\beta}_i) = \sigma^2 \text{trace}(X^t X)^{-1} = \sigma^2 \sum_{i=1}^p \frac{1}{\phi_i}. \quad (2.103)$$

Si la matrice $X^t X$ est mal conditionnée, au moins une des valeurs propres $\{\phi_i\}_{i=1}^p$ sera petite, impliquant alors une large EQM. $X^t X$ est mal conditionnée soit parce que les variables explicatives sont fortement corrélées soit parce que le nombre de variables excède le nombre d'observations. Dans de telles situations, plutôt que de chercher l'estimateur de norme minimale solution de l'équation normale, on peut chercher un compromis entre minimisation de la fonction de coût et faible valeur de la norme $\|\beta\|$: cette approche est connue sous le nom de Ridge Regression (Hoerl and Kennard, [85]). L'estimateur de la ridge regression est solution du problème d'optimisation suivant :

$$\hat{\beta}_{RR} = \arg \min_{\beta \in \mathbb{R}^p} (y - X\beta)^t (y - X\beta) + \lambda \|\beta\|^2 \quad (2.104)$$

où λ est une constante positive à valeur réelle pondérant le compromis entre faible norme et minimisation de la fonction de coût.

En annulant la dérivée par rapport à β on obtient :

$$X^t X \hat{\beta}_{RR} + \lambda \hat{\beta}_{RR} = (X^t X + \lambda \mathbb{1}_p) \hat{\beta}_{RR} = X^t y, \quad (2.105)$$

avec $\mathbb{1}_p$ la matrice identité de dimension p et dont la solution est définie par

$$\hat{\beta}_{RR} = (X^t X + \lambda \mathbb{1}_p)^{-1} X^t y. \quad (2.106)$$

Notons que $(X^t X + \lambda \mathbb{1}_p)$ est systématiquement inversible lorsque $\lambda > 0$. Le modèle final est de la forme :

$$\langle \hat{\beta}_{RR}, x \rangle = y^t X (X^t X + \lambda \mathbb{1}_p)^{-1} x \quad (2.107)$$

Cette méthode est conceptuellement intéressante car elle permet de gérer les données corrélées. Cependant, la solution $\hat{\beta}_{RR}$ est obtenue en inversant une matrice de dimension $p \times p$ rendant pour le moment sa mise en œuvre difficile dans les espaces de grande dimension. Il est possible de contourner cette difficulté en remarquant que $\hat{\beta}_{RR}$ peut s'exprimer sous sa représentation dite duale, c'est-à-dire non plus comme combinaison linéaires des variables, mais des observations (Saunders *et al.*, [139]).

$$\hat{\beta}_{RR} = \lambda^{-1} X^t (y - x \hat{\beta}_{RR}) = X^t \alpha, \quad (2.108)$$

où $\hat{\alpha} = \lambda^{-1} (y - X \hat{\beta}_{RR})$, ce qui nous conduit à

$$\hat{\alpha} = (X^t X + \lambda \mathbb{1}_n)^{-1} y, \quad (2.109)$$

et donc le modèle final est donné par,

$$\langle \hat{\beta}_{RR}, x \rangle = \sum_{i=1}^n \hat{\alpha}_i \langle x_i, x \rangle. \quad (2.110)$$

Les modèles 2.107 et 2.110 sont bien évidemment équivalents. Soulignons deux propriétés fondamentales :

1. L'expression duale de $\hat{\beta}_{RR}$ ne s'exprime qu'au travers de produits scalaires entre observations. De la même manière, par le modèle 2.110, la prédiction d'un nouvel individu x ne nécessite que l'information fournie par le produit scalaire entre x et chacun des x_i . Ainsi, il est possible d'exprimer la solution de la ridge regression uniquement par les valeurs des produits scalaires entre individus.
2. La solution $\hat{\beta}$ nécessite l'inversion de la matrice $(X^t X + \lambda \mathbb{1}_n)$ de dimension $n \times n$. Ceci est particulièrement intéressant lorsque $p \gg n$

Bien souvent, les relations entre la variable à expliquer y et les variables explicatives sont supposées linéaires. Ces relations peuvent s'avérer insuffisantes lorsque l'on cherche à maximiser le pouvoir prédictif d'un modèle. On peut alors s'interroger sur la possibilité d'étendre la ridge regression au cadre non linéaire. Pour ce faire,

nous allons nous placer dans le contexte des Espaces de Hilbert à Noyau Reproduisant (RKHS pour *Reproducing Kernel Hilbert Space*). Puisque la solution 2.110 ne s'exprime que par le produit scalaire entre observations, l'application d'un noyau au contexte de la ridge regression prend alors tout son sens.

Notons $K = (k(x_i, x_j))_{i,j=1,\dots,n}$, la matrice de Gram associée à la fonction noyau k . Cette matrice, de dimension $n \times n$ fournit les valeurs des produits scalaires des n individus dans un RKHS que l'on ne connaît pas explicitement et dans lequel $\hat{\alpha}$ est défini par

$$\hat{\alpha} = (K + \lambda \mathbb{1}_n)^{-1}y, \quad (2.111)$$

et la solution finale est fournie par :

$$\langle \hat{\beta}_{KRR}, x \rangle = \sum_{i=1}^n \hat{\alpha}_i k(x_i, x). \quad (2.112)$$

Notons que $(K + \lambda \mathbb{1}_n)$ est systématiquement inversible quand $\lambda > 0$ et que l'on retrouve alors l'équivalence souhaitée en posant $K = X^t X$.

2.2.3 Modèles non paramétriques additifs

La régression non paramétrique a été étendue au cas multidimensionnel, par Hastie et Tibshirani, [80], c'est-à-dire à l'étude d'une fonction de régression de \mathbb{R}^p , $f(x) = \mathbb{E}[Y|X = x]$ p -dimensionnelle, avec Y la variable réponse et $X = (X_1, \dots, X_p)$ un vecteur de variables explicatives.

La regression dite non paramétrique additive (Stone [148]) regroupe une gamme de modèles qui proposent, dans les cas où l'on ne dispose pas d'échantillons de taille suffisante pour envisager une approche p -dimensionnelle directe, une approximation additive pour f , du type :

$$f(x_1, \dots, x_p) = a + \sum_{j=1}^p f_j(x_j). \quad (2.113)$$

Les fonctions f_j , $1 \leq j \leq p$, sont des fonctions réelles d'une seule variable réelle, de type fonctions lisses définies point par point et non de forme analytique fixée *a priori*. Ce modèle constitue une extension naturelle de différents modèles paramétriques, prenant en compte trois aspects fondamentaux d'un modèle de régression, aspects liés aux deux qualités requises pour un modèle, qui sont les qualités explicatives et prédictives.

Ces trois aspects sont :

- la flexibilité du modèle, qui permet de s'adapter à des situations très variées. En effet, il est possible de faire le lien entre l'absence de flexibilité et le risque possible de biais du modèle estimé, risque d'autant plus grand que l'information sur la fonction de régression est faible. Cette flexibilité doit cependant être mesurée, pour ne pas détériorer les qualités prédictives ;
- la dimension du problème, lié à la variance des estimations, qui croît rapidement pour n fixé, si p augmente (fléau de la dimension) et entraîne donc l'instabilité du modèle estimé. La structure additive du modèle, avec estimation de fonctions univariées, offre une échappatoire à ce problème. La flexibilité et la dimension, sont étroitement liées et la recherche d'un modèle correct peut se baser sur un compromis de ces deux aspects ;
- l'interprétabilité du modèle, importante pour la compréhension de la structure sous jacente. La représentation des fonctions f_j , favorise une telle interprétation, en révélant la relation entre Y et un régresseur particulier, conditionnellement à la présence des autres régresseurs.

Différentes extensions de ce modèle de base ont été développées, toutes basées sur une hypothèse d'additivité de l'approximation.

On peut citer

- les modèles de Régression par Directions Révélatrices (Freedman [66]) :

$$\mathbb{E}[Y|X = x] = f(x_1, \dots, x_p) = \sum_{m=1}^M f_m \left(\sum_{j=1}^p b_{jm} x_j \right), \quad (2.114)$$

qui supposent que les fonctions f_m sont des fonctions réelles d'une variable réelle, combinaison linéaire des variables X_j . L'idée est donc d'étudier une fonction de p variables au travers de son approximation par p fonctions d'une seule variable ;

- la méthode "Alternating Conditional Expectation" (ACE) (Breiman et Friedman, [16]) :

$$\mathbb{E}[\phi(Y)|X = x] = \sum_{j=1}^p f_j(x_j) \quad (2.115)$$

qui propose une double transformation, de la variable réponse et des régresseurs ;

- les Modèles Additifs Généralisés (GAM) (Hastie et Tibshirani, [79] [80]) :

$$g(\mathbb{E}[Y|X = x]) = g(f(x_1, \dots, x_p)) = \sum_{j=1}^p f_j(x_j), \quad (2.116)$$

qui constituent une extension des modèles linéaires généralisés, avec étude de f au travers d'une fonction lien g . C'est ce type de modèles que nous avons utilisé dans nos travaux.

Afin de calibrer au mieux les modèles à coefficients variables (VCM) présentés précédemment et utilisés dans le chapitre 4, l'influence de chaque variable sur la sortie a été testée indépendamment par un modèle additif. Ce modèle permet de savoir quel type de relation (linéaire, quadratique ou autre) existe entre variables réponses et explicatives. Ces modèles ont été calculés grâce au logiciel R avec la fonction "GAM" de la librairie "MGCV" (Wood [163] [164]).

Les modèles additifs généralisés (GAM) sont des modèles linéaires généralisés dans lesquels la relation entre une variable-réponse et des prédicteurs multiples est décrite en partie par une somme de fonctions continues des prédicteurs. Les GAM offrent un bon compromis entre flexibilité et structure et sont, depuis leur introduction par Hastie et Tibshirani, [79] [80], largement appliqués en biostatistique (Cans et Lavergne, [22]), en écologie (Wood et Augustin, [162]), en épidémiologie (Richardson, [133]) et en finances (Coussement *et al.*, [33]). Par la suite, le cadre des GAM a été renforcé par la mise au point de méthodes fiables permettant de sélectionner le degré de continuité des composantes du modèle, par l'élaboration de nouveaux moyens efficaces pour représenter les fonctions continues de plusieurs prédicteurs.

On rencontre les modèles additifs généralisés quand on essaye de décrire une situation non linéaire avec un nombre de variables élevé (si le modèle était linéaire, ce nombre de variables serait tout à fait raisonnable, avec la non linéarité, le nombre de paramètres à estimer explose). La version simplifiée est le modèle additif décrit par :

$$\mathbb{E}[Y] = a + f_1(x_1) + f_2(x_2) + \dots + f_p(x_p), \quad (2.117)$$

où les f_i sont des fonctions quelconques qui peuvent être paramétriques (polynômiale, trigonométriques, splines de régression, noyau, etc.) ou non paramétriques (splines de lissage, régression locale pondérée ou non, loess). L'important est que ces fonctions dépendent d'une seule variable et ne soient pas trop complexes.

Le modèle additif généralisé (GAM) repose sur deux conditions. La première est que la fonction de distribution de Y appartient à la famille exponentielle, et la seconde est que la fonction lien g est monotone et différentiable.

Le GAM se présente donc sous la forme :

$$Y \sim L_{\text{exp}}, \mu = \mathbb{E}[Y] \text{ et } \eta = g(\mu) = a + \sum_{j=1}^p f_j(x_j). \quad (2.118)$$

Bien sûr, toutes les fonctions \mathbb{R}^p dans \mathbb{R} ne peuvent se mettre sous cette forme. De plus, dans cette représentation, on oublie complètement les éventuelles interactions entre les variables. Toutefois, on peut généraliser ce modèle en ajoutant des termes d'ordre supérieur (si nécessaire), comme par exemple, $f_{ij}(x_i, x_j)$.

Les fonctions les plus utilisées dans les modélisations de séries temporelles (proche de notre cas) sont les splines et les fonctions loess (pour *locally weighted scatterplot smoothing*, Cleveland et Devlin [30]). Nous ne rentrons pas dans les détails de ces outils, et renvoyons le lecteur à des ouvrages spécialisés, tel que l'ouvrage de Hastie et Tibshirani [80], qui traite par ailleurs d'autres fonctions de ce type. Ces fonctions sont des outils très utiles car elles permettent de réaliser des ajustements souples aux données et d'exercer un lissage de celle-ci.

Donnons une expression formelle et donc générale de ces fonctions intervenant dans les modèles additifs. Pour cela, il faut supposer qu'il existe une relation réelle mais non connue entre la variable expliquée et les variables explicatives. On peut écrire

$$Y = f(X) + \varepsilon, \quad (2.119)$$

avec ε indépendant, $\mathbb{E}[\varepsilon] = 0$, $\text{Var}[\varepsilon] = \sigma^2$ (de variance constante).

Nous savons que la partie mathématique (explicative) de la modélisation représente l'espérance de Y , conditionnellement aux variables explicatives. Donc, $\mathbb{E}[Y|X = x] = f(x)$, c'est-à-dire que quand X prend les valeurs de x , alors l'espérance conditionnelle de Y vaut $f(x)$.

Le lissage estime, pour une valeur de x donnée, x_0 par exemple, des valeurs $\hat{f}(x_0)$, proches des vraies valeurs $f(x_0)$, en général, en moyennant les valeurs de y proches de $f(x_0)$.

L'algorithme utilisé pour trouver ces fonctions est itératif. Pour illustrer la démarche, reprenons l'écriture simplifiée du modèle additif,

$$Y = a + \sum_{j=1}^p f_j(x_j) + \varepsilon. \quad (2.120)$$

On en déduit, en isolant $f_k(X_k)$, que :

$$Y - a + \sum_{\substack{j=1 \\ j \neq k}}^p f_j(x_j) = f_k(X_k) + \varepsilon. \quad (2.121)$$

D'où

$$\mathbb{E} \left[Y - a + \sum_{\substack{j=1 \\ j \neq k}}^p f_j(x_j) | X_k \right] = f_k(X_k). \quad (2.122)$$

Ceci signifie que la valeur de $f_k(X_k)$ est exprimable à partir des fonctions des autres variables.

Le principe de l'algorithme est qu'à chaque étape, la valeur prise par la fonction d'une variable donnée est calculée à partir des valeurs prises par les fonctions des autres variables à l'étape précédente.

L'algorithme d'"ajustement arrière" (*backfitting* en anglais) repose sur les étapes suivantes :

1. On choisit des valeurs initiales pour les $f_j(X_j)$:

$$a = \frac{1}{n} \sum_{i=1}^p y_i, \quad (2.123)$$

en d'autres terme, on prend la moyenne des y_i pour valeur initiale de a . Pour les autres, on choisit p valeurs de départ $f_1^0, f_2^0, \dots, f_p^0$. Comme on n'a pas d'*a priori*, on peut choisir des valeurs déduites d'une régression de y sur les x_j ;

2. On estime f_k^1 , pour $k = 1, \dots, p$, tel que

$$f_k^1 = S \left(y - a - \sum_{\substack{j=1 \\ j \neq k}}^p f_j^0 | X_k \right), \quad (2.124)$$

avec S une fonction de lissage prédéterminée ;

3. On estime ensuite f_k^2 , pour $k = 1, \dots, p$, de la même façon, mais à partir des f_k^1 et ainsi de suite ;
4. On s'arrête quand les fonctions ne diffèrent plus trop d'une étape à l'autre. Il est donc nécessaire de définir un seuil.

Dans leur ouvrage, Hastie et Tibshirani, [80], montrent la convergence de cet algorithme dans le cas de l'utilisation de splines de lissage. En revanche, cette convergence n'est pas assurée, avec les fonctions loess.

Le principe de l'algorithme reste le même dans le cas de l'ajustement du GAM, mais diffère quelque peu en pratique. Cet algorithme, appelé *local scoring algorithm*, reprend, en le généralisant, celui de la méthode des scores de Fischer (Fischer [62]). On y retrouve donc l'expression de Taylor mais avec la notion de lissage.

L'estimation de a et des f_j est effectuée par les étapes suivantes :

1. On démarre l'algorithme en donnant une valeur arbitraire à a et aux f_j :

$$a^0 = g\left(\sum_{j=1}^n \frac{y_j}{n}\right) \text{ et } f_1^0 = f_2^0 = \dots = f_p^0 = 0; \quad (2.125)$$

On sait alors calculer $\eta_i^0 = a^0 + \sum_{j=1}^p f_j^0(x_{ij})$ et $\mu_i^0 = g^{-1}(\eta_i^0)$, avec g connue (fonction ln par exemple) ;

2. On construit une nouvelle variable dépendante que l'on pondère. Soit

$$z_i = \eta_i^0 + (y_i - \mu_i^0) \left(\frac{\partial \eta_i}{\mu_i} \right)_0. \quad (2.126)$$

Soit w_i une pondération de z_i , dépendante de la dérivée de g^{-1} et de l'inverse de la variance ;

3. On ajuste un modèle additif pondéré sur les valeurs de la nouvelle variable dépendante z_i . Cet ajustement se fait à l'aide de l'algorithme précédemment introduit, d'ajustement arrière. On obtient ainsi une estimation des fonctions f_j^1 , la valeur des η_i^1 et des μ_i^1 . Enfin, on calcule le critère de convergence basé sur la somme relative des valeurs absolues des différences entre f_j^1 et f_j^0 .
4. On revient à l'étape 2, en remplaçant les valeurs initiales par les nouvelles valeurs et on itère jusqu'à ce que le critère de convergence soit suffisamment proche de 0.

Au final, il est possible de mesurer la qualité de l'ajustement en calculant la déviance, qui repose sur la log-vraisemblance ou à l'aide du critère d'Akaike (AIC pour *Akaike information criterion*) qui correspond à la déviance pénalisée par un terme dépendant du nombre de paramètres du modèle (Akaike [3] [4]).

2.3 Conclusions

Nous avons donné dans ce chapitre un aperçu des différents outils de modélisation que nous utilisons par la suite. Ces outils sont répartis en deux parties. La première concerne les méthodes basées sur la théorie du krigeage, méthodes que nous utilisons, étendons, modifions dans le chapitre 3, relatif aux méthodes de modélisation non fonctionnelle. Le second groupe s'attarde sur la définition des données fonctionnelles, des outils relatifs et plus particulièrement des VCM. Ces modèles sont utilisés dans le chapitre 4 relatif aux méthodes fonctionnelles mise en place.

L'extrême simplicité du krigeage est un de ses avantages majeurs. Le krigeage permet en effet de modéliser avec la même simplicité des systèmes observés avec ou sans bruit, des systèmes comportant des sorties corrélées, d'estimer des dérivées (ce qui est souvent considéré comme un problème difficile) ou de prendre en compte des connaissances *a priori*, telles que des comportements particuliers aux bords du domaine (exemple des gradients nuls). L'utilisation du krigeage pour la modélisation comportementale de systèmes nous semble pertinente tant du point de vue théorique que pratique.

Les différents aspects de cette théorie sont utilisés ou étendus dans le chapitre 3. Ces extensions concernent principalement notre choix de modifier la tendance déterministe par notre modèle numérique, qui est non linéaire et le caractère fonctionnel des données.

La notion de données fonctionnelles est primordiale pour notre cas d'étude. Cette caractéristique doit obligatoirement être prise en compte. Il est important de souligner que dans notre étude, contrairement aux cas rencontrés dans les nombreux travaux sur le sujet, les variables sont des scalaires et seules les sorties sont des données fonctionnelles.

Une étude plus attentive de nos données (section 4.1.1.1) fait apparaître la pertinence des modèles à coefficients variables pour notre cas d'étude. Ces modèles qui supposent une relation linéaire présentent de bonnes propriétés d'ajustement et de prédiction dans de nombreux cas, et les méthodes d'estimation des fonctions coefficients sont relativement simple à mettre en œuvre. Enfin, nous avons donné un bref aperçu des modèles additifs généralisés, un domaine vaste et très usité dans la littérature. Les modèles GAM permettent de modéliser de façon flexible et interprétable une classe très large de problème concernant la relation entre une variable réponse et plusieurs variables explicatives. C'est dans ce cadre que nous les avons appliqués dans le chapitre 4, afin de visualiser si la relation linéaire supposée dans les VCM est valable, pour leur application.

Approche non-fonctionnelle

3.1	Krigeage à tendance non linéaire	79
3.1.1	Prédicteur de krigeage	79
3.1.2	Estimation des paramètres	80
3.1.3	Algorithme mis en place	80
3.1.4	Applications	82
3.1.4.1	Comparaison krigeage classique et à tendance non linéaire sur le cas maîtrisé	82
3.1.4.2	Application du krigeage à tendance non linéaire sur le cas d'étude	84
3.1.5	Conclusions et limites	88
3.2	Conditionnement par la dérivée	89
3.2.1	Présentation de la démarche	89
3.2.2	Application sur le cas maîtrisé	91
3.2.3	Application sur le cas d'étude	93
3.2.4	Conclusions et limites	96
3.3	Modélisation de la dérivée par rapport à l'avancement	97
3.3.1	Présentation de la démarche	98
3.3.2	Résultats sur le cas d'étude	100
3.3.3	Intégration des résultats	102
3.3.4	Calcul de la variance de prédiction	106
3.3.5	Conclusions et limites	108
3.4	Conclusions	109

Approche non-fonctionnelle

Il est courant, lors d'études expérimentales, que le modèle postulé pour représenter un système physique ou chimique ne convienne que partiellement. C'est le cas de notre système, introduit dans le chapitre 1.

Un schéma réactionnel a été élaboré et un modèle pour le mettre en œuvre a été construit (simulateur). Ce premier modèle est fortement non linéaire, et peut s'exprimer par $\mathbf{y} = f(\mathbf{x}, \boldsymbol{\beta})$, où \mathbf{y} est la réponse étudiée (par exemple la teneur en hydrocarbures imbrûlés), \mathbf{x} les conditions expérimentales et $\boldsymbol{\beta}$ les paramètres cinétiques du modèle f .

Pour ajuster ce modèle aux $n = 20$ résultats expérimentaux disponibles et estimer les paramètres du vecteur, la première idée est de déterminer par moindres carrés un vecteur de paramètres globaux, valable quelle que soit l'expérience considérée.

$$\hat{\boldsymbol{\beta}} = \arg \min_{\boldsymbol{\beta}} \sum_{i=1}^n \int \|y^i(T) - f(x^i, T, \boldsymbol{\beta})\|^2 dT \quad (3.1)$$

où y^i correspond aux résultats expérimentaux de l'expérience i ($[CO]$ et $[HC]$), et x^i à ces conditions opératoires ($[CO]_0, [HC]_0, [O_2]_0$).

Malheureusement, tous les essais basés sur l'équation (3.1) se sont révélés infructueux, montrant que le modèle proposé est insuffisant pour représenter toute la réalité expérimentale, même si les grandes tendances sont respectées. La confrontation aux premiers résultats expérimentaux a montré son inadéquation (section 1.1.4).

Une deuxième idée pour exploiter au mieux le modèle est de considérer que celui-ci rend compte d'un effet moyen, qui doit être corrigé d'un effet modélisé comme un processus gaussien $z_{\sigma^2, \boldsymbol{\theta}}(\mathbf{x})$ de moyenne nulle et dont la matrice de variance-covariance est paramétrée par $\boldsymbol{\theta}$.

Notre démarche est donc la suivante : nous proposons de considérer que les différences observées entre ce premier modèle et la réalité expérimentale peuvent se représenter par un processus gaussien. Dès lors, nous posons

$$\mathbf{y} = f(\mathbf{x}, \boldsymbol{\beta}) + z_{\sigma^2, \boldsymbol{\theta}}(\mathbf{x}) \quad (3.2)$$

où la tendance linéaire d'un modèle de krigeage classique est remplacée par le modèle initial, et où $z_{\sigma^2, \boldsymbol{\theta}}(\mathbf{x})$ est un processus gaussien centré de variance σ^2 et de paramètre de portée $\boldsymbol{\theta}$. Classiquement, on a recours à un noyau gaussien.

Les principaux intérêts du krigeage sont de :

1. Corriger les écarts entre expériences et simulations
2. Calculer la variance de prédiction en tout point, à faible coût

Nous sommes partis de la théorie du krigeage, présentée dans le chapitre 2, afin de développer la théorie du krigeage étendue à une tendance non-linéaire. Deux motivations expliquent l'utilisation du krigeage dans notre cas :

1. Nécessité d'avoir un bon modèle pour faire un plan d'expériences pertinent
2. Existence d'une variance de krigeage dépendant de l'espace et donc permettant l'existence de critères et d'outils de plan d'expériences pour des modèles de krigeage

Dans cette approche, et toutes celles présentées dans ce chapitre, on considère que la sortie d'une expérience n'est pas une courbe mais une succession de points. Le caractère fonctionnel des sorties n'est pas pris en considération.

Les résultats obtenus font apparaître des comportements atypiques. Par exemple, la stricte positivité de la sortie dans certains cas n'est pas respectée, du fait de l'apparition d'oscillations sur le palier final. Les travaux de Vazquez [156] nous ont guidés vers une amélioration de cette modélisation, en conditionnant le processus gaussien par les dérivées selon la température. Ce conditionnement par la dérivée est motivé par le fait que sur les deux paliers identifiés sur les courbes, ces dérivées sont nulles. L'objectif est de prévenir les oscillations qui perturbent certaines prédictions obtenues. L'ouvrage de Santner *et al.* [138] expose précisément la théorie dans le cas du krigeage universel.

Comme nous l'exposons dans la section 3.2.2, bien que le conditionnement par la dérivée apporte une amélioration de nos prédictions, celles-ci présentent néanmoins, pour des cas extrêmes des comportements singuliers. Les limites de cette seconde approche seront clairement explicitées dans la section 3.2.3, avec notamment la remise en cause du noyau stationnaire afin de modéliser un processus non stationnaire. Si l'utilisation de noyaux de covariances non-stationnaires ne pose pas de problème conceptuel majeur dans le cadre d'une modélisation par processus gaussiens (voir Ginsbourger [72]), il est plutôt rare d'en trouver dans les applications classiques du krigeage, du fait des problèmes d'estimation qu'ils posent.

Nous avons alors décidé de déterminer une modélisation pour les dérivées en fonction de l'avancement (le taux de conversion du polluant). Notamment, l'écart entre dérivée expérimentale en fonction de l'avancement et simulée correspond au cas classique de modélisation par un processus gaussien conditionné aux bords. Il est possible de montrer que ce type de processus, appelé pont brownien (voir Ginsbourger [72]), est équivalent à prendre un noyau exponentiel conditionné aux bords. Cette particularité est à l'origine de la troisième approche présentée dans ce chapitre : la modélisation directe de la dérivée en fonction de l'avancement.

Cependant, cette approche entraîne là encore quelques obstacles que nous avons en partie su contourner (intégration de la dérivée, apparition d'un écart entre prédictions et résultats expérimentaux), et que nous présentons dans la section 3.3.2.

La construction de ce chapitre respecte donc la démarche chronologique mise en place. Dans un premier temps, nous présentons la théorie du krigeage à tendance non linéaire. Cette première méthode, validée sur le cas maîtrisé, montre ses limites que nous exposons de manière claire à partir de résultats de prédiction. Dans un second temps, nous introduisons le conditionnement par la dérivée. Encore une fois la validation sur le cas maîtrisé, nous permet de motiver l'application sur le cas d'étude. Les résultats sont notamment comparés aux résultats précédents afin de bien mettre en avant l'amélioration apportée par l'information de la dérivée, et exhibent les limites de cette seconde approche. Enfin, nous terminons par la modélisation directe de la dérivée selon l'avancement appliquée directement sur le cas d'étude.

3.1 Krigeage à tendance non linéaire

L'inadéquation du modèle à représenter la réalité expérimentale nous a conduits à développer une méthode afin de corriger les écarts entre expériences et simulations. La forme des résidus (différence entre expériences et simulations) nous permet d'envisager l'utilisation du krigeage pour ajuster le modèle aux réponses expérimentales.

Pour notre cas d'étude nous avons dû étendre la théorie de krigeage au cas d'une tendance non-linéaire. En effet, on suppose dans notre cas que les résidus peuvent se représenter par un processus gaussien $z_{\sigma^2, \theta}$ de moyenne nulle et de variance-covariance donné par un noyau d'hyperparamètres σ^2 et θ . On remplace donc la tendance d'un modèle de krigeage par notre modèle numérique soit,

$$\mathbf{y} = f(\mathbf{x}, \boldsymbol{\beta}) + z_{\sigma^2, \theta}(\mathbf{x}) \quad (3.3)$$

où \mathbf{y} est la réponse, \mathbf{x} les conditions expérimentales et $\boldsymbol{\beta}$ les paramètres cinétiques du modèle noté f .

La première difficulté de cette approche est d'estimer les paramètres $\boldsymbol{\beta}$ de la tendance, qui dans notre cas correspondent aux paramètres cinétiques, considérant le caractère non-linéaire de notre modèle et les hyperparamètres du noyau σ^2 et θ . La deuxième difficulté est de déterminer un critère pour la sélection des nouveaux points d'expériences à réaliser.

Concernant l'estimation des paramètres, de manière classique nous utilisons une estimation par maximum de vraisemblance. Cependant, de manière similaire à la théorie de régression non-linéaire, la formule analytique traditionnelle pour $\boldsymbol{\beta}$ est remplacée par une procédure de minimisation.

Pour la détermination des nouveaux points d'expériences à réaliser, l'idée est de choisir les points du domaine où notre connaissance du système au travers du modèle numérique est la plus incertaine, soit aux points de plus forte variance de prédiction. Ce premier critère de plan d'expériences n'est pas abordé dans cette partie du fait des limites de la méthode d'estimation exposées dans la section 3.1.5.

3.1.1 Prédicteur de krigeage

Soit un ensemble de n points d'expériences, $\mathbf{S} = [\mathbf{s}_1 \dots \mathbf{s}_n]^T$ où $\mathbf{s}_i \in \mathbb{R}^p$ et le vecteur des réponses associées, $\mathbf{Y} = [\mathbf{y}_1 \dots \mathbf{y}_n]^T$. On cherche à ajuster les réponses par le modèle (3.3), où $f(\mathbf{x}, \boldsymbol{\beta})$ est une fonction non-linéaire de $\mathbf{x} \in \mathbb{R}^p$ et $\boldsymbol{\beta} \in \mathbb{R}^q$.

L'approche proposée correspond à un krigeage avec dérive externe puisque $\boldsymbol{\beta}$ n'est pas estimé en même temps (Ginsbourger, [72]). On peut donc enlever la moyenne et la rajouter à la fin. En reprenant les notations de la section 2.1.1, on obtient la formule de prédicteur de krigeage, définie par :

$$\begin{aligned} \hat{\mathbf{y}}(\mathbf{x}) &= (r(\mathbf{x}) - f(\mathbf{S}, \boldsymbol{\beta}) \cdot \tilde{\boldsymbol{\lambda}})^T \mathbf{R}^{-1} \mathbf{Y}, \\ &= r(\mathbf{x})^T \mathbf{R}^{-1} \mathbf{Y} - (f(\mathbf{S}, \boldsymbol{\beta})^T \mathbf{R}^{-1} r(\mathbf{x}) - f(\mathbf{x}, \boldsymbol{\beta}))^T \\ &\quad (f(\mathbf{S}, \boldsymbol{\beta})^T \mathbf{R}^{-1} f(\mathbf{S}, \boldsymbol{\beta}))^{-1} f(\mathbf{S}, \boldsymbol{\beta})^T \mathbf{R}^{-1} \mathbf{Y}. \end{aligned} \quad (3.4)$$

où $f(\mathbf{S}, \boldsymbol{\beta})$ est la tendance définie par $f(\mathbf{S}, \boldsymbol{\beta}) = [f(\mathbf{s}_1, \boldsymbol{\beta}) \dots f(\mathbf{s}_n, \boldsymbol{\beta})]^T$, et la variance de prédiction de notre prédicteur devient,

$$\varphi(\mathbf{x}) = \sigma^2 (1 + \mathbf{u}^T (f(\mathbf{S}, \boldsymbol{\beta})^T \mathbf{R}^{-1} f(\mathbf{S}, \boldsymbol{\beta}))^{-1} \mathbf{u} - r(\mathbf{x})^T \mathbf{R}^{-1} r(\mathbf{x})), \quad (3.5)$$

où $\mathbf{u} = f(\mathbf{S}, \boldsymbol{\beta})^T \mathbf{R}^{-1} r(\mathbf{x}) - f(\mathbf{x}, \boldsymbol{\beta})$ et σ^2 est obtenu par (3.15). La généralisation au cas multiréponse est immédiate, il suffit de remplacer σ par σ_l , la variance du processus pour la $l^{\text{ème}}$ réponse, pour obtenir la variance de prédiction du krigeage pour la $l^{\text{ème}}$ réponse.

3.1.2 Estimation des paramètres

Fang *et al.*, ([57],§5.4), montrent dans le cas du krigeage classique, qu'il est possible d'estimer séparément et itérativement β et (σ^2, θ) . Nous montrons, dans le cas où la tendance est non-linéaire, que la procédure reste valable.

L'hypothèse de normalité du modèle de krigeage gaussien, entraîne la densité de \mathbf{y} suivante :

$$(2\pi\sigma^2)^{-n/2} |\mathbf{R}|^{1/2} \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{y} - f(\mathbf{x}, \beta))^T \mathbf{R}^{-1} (\mathbf{y} - f(\mathbf{x}, \beta)) \right\}. \quad (3.6)$$

La log-vraisemblance est :

$$l(\beta, \sigma^2, \theta) = -\frac{n}{2} \log(\sigma^2) - \frac{1}{2} \log(|\mathbf{R}|) - \frac{1}{2\sigma^2} (\mathbf{y} - f(\mathbf{x}, \beta))^T \mathbf{R}^{-1} (\mathbf{y} - f(\mathbf{x}, \beta)). \quad (3.7)$$

Maximiser la log-vraisemblance revient à maximiser l'estimateur de vraisemblance de $(\beta, \sigma^2, \theta)$. En pratique, maximiser simultanément $(\beta, \sigma^2, \theta)$ est instable numériquement à cause du déterminant de la matrice \mathbf{R} , noté $|\mathbf{R}|$, car \mathbf{R} peut être presque singulière et σ^2 peut être très petit. De plus, les paramètres β et θ jouent des rôles différents ; β est utilisé pour modéliser la tendance moyenne, alors que θ est un vecteur de paramètres de portée. Il est préférable de les estimer séparément. Or, on observe que,

$$\frac{\partial^2 l(\beta, \sigma^2, \theta)}{\partial \beta \partial \sigma^2} = -\frac{1}{\sigma^4} (\mathbf{y} - f(\mathbf{x}, \beta))^T \mathbf{R}^{-1} \frac{\partial f(\mathbf{x}, \beta)}{\partial \beta}, \quad (3.8)$$

et

$$\frac{\partial^2 l(\beta, \sigma^2, \theta)}{\partial \beta \partial \theta} = -\frac{1}{\sigma^2} (\mathbf{y} - f(\mathbf{x}, \beta))^T \frac{\partial \mathbf{R}^{-1}}{\partial \theta} \frac{\partial f(\mathbf{x}, \beta)}{\partial \beta}. \quad (3.9)$$

Comme $\mathbb{E}[\mathbf{y}] = f(\mathbf{x}, \beta)$,

$$\mathbb{E} \left[\frac{\partial^2 l(\beta, \sigma^2, \theta)}{\partial \beta \partial \sigma^2} \right] = 0 = \mathbb{E} \left[\frac{\partial^2 l(\beta, \sigma^2, \theta)}{\partial \beta \partial \theta} \right]. \quad (3.10)$$

Ceci implique que l'estimateur du maximum de vraisemblance de β est asymptotiquement indépendant de celui de (σ^2, θ) , car la matrice d'information de Fisher,

$$\mathbb{E} \left[\frac{\partial^2 l(\beta, \sigma^2, \theta)}{\partial \beta \partial (\sigma^2, \theta)} \right] = \mathbb{E} \begin{bmatrix} \frac{\partial^2 l(\beta, \sigma^2, \theta)}{\partial \beta^2} & \frac{\partial^2 l(\beta, \sigma^2, \theta)}{\partial \beta \partial (\sigma^2, \theta)} \\ \frac{\partial^2 l(\beta, \sigma^2, \theta)}{\partial \beta \partial (\sigma^2, \theta)} & \frac{\partial^2 l(\beta, \sigma^2, \theta)}{\partial (\sigma^2, \theta)^2} \end{bmatrix} = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}, \quad (3.11)$$

est une matrice bloc diagonale. Ceci nous permet d'estimer β et (σ^2, θ) séparément et itérativement.

3.1.3 Algorithme mis en place

Soit le vecteur de paramètres θ , de dimension d . Dans le cas isotrope, $d = 1$ et $d = p$, la dimension de \mathbf{x} , pour le cas anisotrope (à titre informatif, les différentes méthodes présentées dans ce manuscrit qui utilisent un noyau, sont appliquées dans le cas anisotropique, c'est pourquoi nous noterons θ de dimension p). On cherche (une approximation de) θ^* dans la région $0 < l_j \leq \theta_j \leq u_j$, $j = 1, \dots, q$, où l_j et u_j sont respectivement la borne inférieure et supérieure du $j^{\text{ème}}$ composant de θ .

Le choix optimal de $\boldsymbol{\theta}^*$ est défini comme l'estimateur de maximum de vraisemblance. On cherche $\boldsymbol{\theta}^*$, qui maximise la log-vraisemblance, soit

$$\boldsymbol{\theta}^* = \max_{\boldsymbol{\theta}} \{l(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta})\}. \quad (3.12)$$

Or $\boldsymbol{\beta}$ et σ^2 sont inconnus. $\boldsymbol{\beta}$ est estimé en minimisant,

$$(\mathbf{y} - f(\mathbf{x}, \boldsymbol{\beta}))^T \mathbf{R}^{-1} (\mathbf{y} - f(\mathbf{x}, \boldsymbol{\beta})), \quad (3.13)$$

par rapport à $\boldsymbol{\beta}$. Cette formule correspond au critère des moindres carrés généralisés classique puisque,

$$\mathbb{E}[\mathbf{y}] = f(\mathbf{x}, \boldsymbol{\beta}) \quad \text{Cov}[\mathbf{y}] = \sigma^2 \mathbf{R}. \quad (3.14)$$

Ensuite, on estime σ^2 en résolvant $\frac{\partial l(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta})}{\partial \sigma^2} = 0$, et on a :

$$\hat{\sigma}^2 = n^{-1} (\mathbf{y} - f(\mathbf{x}, \hat{\boldsymbol{\beta}}))^T \mathbf{R}^{-1} (\mathbf{y} - f(\mathbf{x}, \hat{\boldsymbol{\beta}})). \quad (3.15)$$

On remplace alors σ^2 par (3.15) dans (3.7), la log-vraisemblance devient :

$$l(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta}) = -\frac{n}{2} \left(\log(\sigma^2 \cdot |\mathbf{R}|^{\frac{1}{n}}) + 1 \right). \quad (3.16)$$

D'après (3.16) et (3.12), on cherche $\boldsymbol{\theta}^*$ tel que :

$$\boldsymbol{\theta}^* = \min_{\boldsymbol{\theta}} \{ \psi(\boldsymbol{\theta}) = \sigma^2 \cdot |\mathbf{R}|^{\frac{1}{n}} \}. \quad (3.17)$$

L'algorithme consiste à chercher $\boldsymbol{\theta}^*$, en minimisant la fonction objectif (3.17) à l'aide d'une méthode de *Hooke et Jeeves modifiée*, décrite dans Kowalik et Osborne,[92], Section 2.4. Pour chaque appel à la fonction objectif, on procède à une mise à jour de $\boldsymbol{\beta}$ à l'aide de (3.13) puis de σ^2 avec (3.15).

Pour plus de renseignements Lophaven *et al.*, ([97], Section 6) décrivent précisément les étapes de l'algorithme, dans le cas linéaire. La seule modification apportée porte sur le calcul de $\boldsymbol{\beta}$.

D'un point de vue pratique, pour éviter les problèmes d'instabilité numérique, on décompose \mathbf{R} par la factorisation de Cholesky,

$$\mathbf{R} = \mathbf{C}\mathbf{C}^T. \quad (3.18)$$

Ainsi, (3.15) devient,

$$\begin{aligned} \hat{\sigma}^2 &= n^{-1} (\mathbf{y} - f(\mathbf{x}, \hat{\boldsymbol{\beta}}))^T \mathbf{R}^{-1} (\mathbf{y} - f(\mathbf{x}, \hat{\boldsymbol{\beta}})), \\ &= n^{-1} (\mathbf{y} - f(\mathbf{x}, \hat{\boldsymbol{\beta}}))^T (\mathbf{C}\mathbf{C}^T)^{-1} (\mathbf{y} - f(\mathbf{x}, \hat{\boldsymbol{\beta}})), \\ &= n^{-1} (\mathbf{y} - f(\mathbf{x}, \hat{\boldsymbol{\beta}}))^T \mathbf{C}^{-T} \mathbf{C}^{-1} (\mathbf{y} - f(\mathbf{x}, \hat{\boldsymbol{\beta}})), \\ &= n^{-1} \|\mathbf{C}^{-1} (\mathbf{y} - f(\mathbf{x}, \hat{\boldsymbol{\beta}}))\|_2^2. \end{aligned} \quad (3.19)$$

De même, l'estimation de $\boldsymbol{\beta}$ s'obtient en minimisant,

$$\|\mathbf{C}^{-1} (\mathbf{y} - f(\mathbf{x}, \boldsymbol{\beta}))\|_2^2. \quad (3.20)$$

Dépendant du choix de la fonction de corrélation choisie et du paramètre $\boldsymbol{\theta}$, la matrice \mathbf{R} peut être mal conditionnée. Afin de réduire l'effet des erreurs d'arrondis, on utilise une factorisation de Cholesky modifiée, où (3.18) est remplacée par

$$\mathbf{C}\mathbf{C}^T = \mathbf{R} + \nu\mathbf{I}_d \text{ où } \nu = (10 + n)\varepsilon_m, \quad (3.21)$$

où ε_m est la précision machine ($\varepsilon_m = 2^{-52} \simeq 2,22 \cdot 10^{-16}$ pour nous).

3.1.4 Applications

Nous présentons dans cette section deux applications de notre méthode d'estimation. Tout d'abord sur le cas maîtrisé, puis sur le cas d'étude.

La première étape, sur le cas maîtrisé, nous permet de confronter les résultats obtenus par notre méthode aux résultats obtenus par un krigeage classique, et ainsi de pouvoir conclure quant à la pertinence de notre modèle. On rappelle que ce cas a été construit de manière à correspondre à notre cas d'étude, mais plus simple et totalement maîtrisé. L'idée est de valider dans un premier temps notre méthodologie sur ce cas avant de passer à notre cas d'étude.

Cette deuxième étape, l'application sur notre cas d'étude, est ensuite présentée. Des résultats de prédiction sont exhibés et commentés. À partir de ces résultats, nous dressons, dans la dernière section, les limites de notre méthode et les améliorations envisagées.

3.1.4.1 Comparaison krigeage classique et à tendance non linéaire sur le cas maîtrisé

Pour comparer le *krigeage à tendance non-linéaire* (KNL), que nous avons mis au point, au krigeage classique, nous avons appliqué ces deux méthodes sur notre cas maîtrisé. Nous montrons que le KNL s'ajuste mieux aux résultats expérimentaux que le krigeage à tendance linéaire, en dehors du domaine d'échantillonnage (zones d'extrapolation). Mais surtout, nous montrons que la qualité de prédiction est bien meilleure avec le krigeage à tendance non-linéaire.

Le modèle de krigeage à tendance linéaire choisi est le suivant :

$$\mathbf{Y}(\mathbf{x}) = f(\mathbf{x})^T \boldsymbol{\beta} + z_{\sigma^2, \boldsymbol{\theta}}(\mathbf{x}) \quad (3.22)$$

où :

- \mathbf{Y} est la réponse expérimentale,
- \mathbf{x} est le vecteur des conditions expérimentales : la concentration initiale de l'espèce A , la température et le temps,
- $f(\mathbf{x})$ est une fonction linéaire des entrées,
- $\boldsymbol{\beta}$ est le vecteur des coefficients de régression
- $z_{\sigma^2, \boldsymbol{\theta}}(\mathbf{x})$ est un processus stochastique de moyenne nulle, de paramètres σ^2 et $\boldsymbol{\theta}$.

Le krigeage non-linéaire choisi est défini par le modèle suivant :

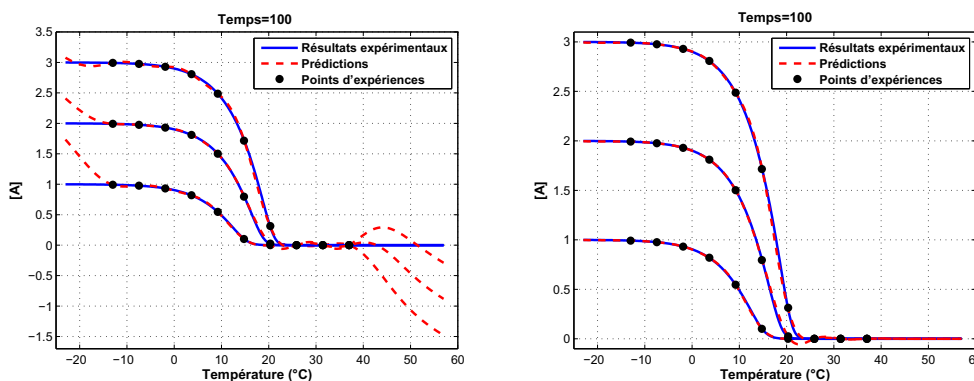
$$\mathbf{Y}(\mathbf{x}) = f(\mathbf{x}, \boldsymbol{\beta}) + z_{\sigma^2, \boldsymbol{\theta}}(\mathbf{x}) \quad (3.23)$$

où $f(\mathbf{x}, \boldsymbol{\beta})$ correspond à la tendance, dans notre cas le modèle régi par le système cinétique, et $\boldsymbol{\beta}$ est le vecteur des paramètres cinétiques du modèle.

Dans les deux cas, la corrélation utilisée est la corrélation gaussienne anisotrope. Les résultats obtenus sont présentés dans les figures 3.1a et 3.1b, correspondant respectivement au krigeage à tendance linéaire et au KNL. À temps fixé à 100 secondes, nous avons choisi 30 points (10 températures pour 3 $[A]_0$ différents) représentés par les points noirs sur les figures. La courbe bleue correspond aux résultats expérimentaux, et les courbes pointillées à la prédiction de krigeage.

Pour le krigeage à tendance linéaire, l'ajustement du modèle n'est pas bon en dehors du domaine d'échantillonnage. De plus, les β calculés n'ont pas de sens physiques, puisque ces derniers correspondent aux coefficients du modèle 3.22.

Pour le krigeage à tendance non-linéaire, l'ajustement est bien meilleur dans les zones d'extrapolation, car le simulateur est bon dans cette région. La prédiction est interpolante, et de plus, les β calculés correspondent aux paramètres cinétiques optimaux du système (1.22).



(a) Résultat du krigeage à tendance linéaire (b) Résultat du krigeage à tendance non-linéaire
 Figure 3.1 – Résultats du krigeage tendance linéaire et non-linéaire pour un temps de 100 secondes

On peut voir que la courbe de krigeage, dans le cas non-linéaire, prend des valeurs négatives autour de la température de 25°C, pour une concentration initiale de 2. Ce phénomène est dû aux oscillations induites par le krigeage. En effet, on a interpolation des points d'expériences, et entre on a une gaussienne. L'importance du phénomène est à minimiser dans un premier temps, même si rappelons-le cela n'a pas de sens physique. Il est possible de répondre en partie au problème en rajoutant des points d'expériences.

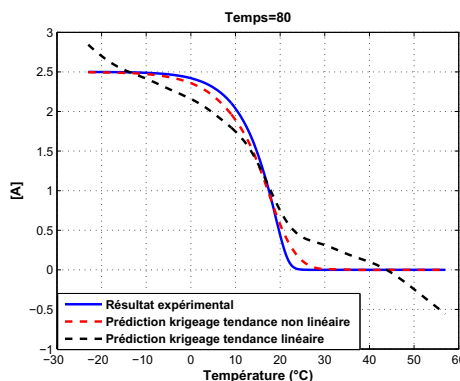


Figure 3.2 – Résultats de prédiction du krigeage tendance linéaire et non-linéaire pour un temps de 80 secondes et $A_0 = 2.5$

À partir des modèles déterminés par le krigeage à tendance linéaire et le krigeage non-linéaire, nous prédisons l'évolution de notre système pour un temps et une concentration initiale, différents de ceux pris dans le plan d'expériences. La figure 3.2 présente la prédiction pour un temps de 80 secondes et une concentration initiale, $A_0 = 2.5$, à l'aide de nos deux modèles déterminés.

La courbe bleue correspond aux résultats expérimentaux, celle en rouge à la prédiction par KNL, et celle en noir à la prédiction par krigeage tendance linéaire. Dans le cas du KNL, la prédiction est bien meilleure que celle obtenue avec le krigeage tendance linéaire. En effet, on constate que la prédiction par krigeage à tendance non-linéaire est très proche des résultats expérimentaux, alors que la prédiction par krigeage tendance linéaire non seulement est éloignée, mais la forme même de la courbe ne correspond pas à celles des résultats expérimentaux.

3.1.4.2 Application du krigeage à tendance non linéaire sur le cas d'étude

L'application du krigeage à tendance non-linéaire sur le cas d'étude a été faite à partir des expériences à notre disposition. Les points d'expériences sont choisis uniformément sur les courbes lissées par rapport à la température. On définit ainsi notre matrice d'expériences et les réponses associées.

Avant de présenter les résultats, il apparaît judicieux de présenter précisément la construction de la matrice d'expériences \mathbf{S} et du vecteur des sorties associées \mathbf{Y} . Les approches présentées dans ce chapitre considèrent les courbes en sorties non pas comme des courbes, mais comme une succession de points. Ainsi, on sélectionne un nombre n_T de températures pour lesquelles on récupère les valeurs en sortie.

La matrice \mathbf{S} est donc une matrice bloc de taille $n \cdot n_t \times p$, avec n le nombre d'expériences et p (=4 dans notre cas) le nombre de variables. Chacun des n blocs est construit de la manière suivante :

1. la première colonne correspond aux températures associées à chacun des points : discrétisation uniforme de taille n_T entre la température minimale commune à tous les essais, et la température maximale commune,
2. les trois autres colonnes correspondent respectivement aux concentrations initiales de HC , CO et O_2 de l'essai considéré. Par conséquent, dans une colonne les éléments sont égaux.

Nous donnons ci-dessous une définition de la matrice d'expériences \mathbf{S} et du vecteur des sorties associées \mathbf{Y} à titre informatif. Cette définition est un peu abusive quant aux notations utilisées, mais n'a uniquement pour but que de clarifier la construction de ces matrices.

$$\mathbf{S} = \begin{pmatrix} T_1 & [HC]_0^1 & [CO]_0^1 & [O_2]_0^1 \\ \vdots & \vdots & \vdots & \vdots \\ T_{n_T} & [HC]_0^1 & [CO]_0^1 & [O_2]_0^1 \\ T_1 & [HC]_0^2 & [CO]_0^2 & [O_2]_0^2 \\ \vdots & \vdots & \vdots & \vdots \\ T_{n_T} & [HC]_0^2 & [CO]_0^2 & [O_2]_0^2 \\ \vdots & \vdots & \vdots & \vdots \\ T_1 & [HC]_0^n & [CO]_0^n & [O_2]_0^n \\ \vdots & \vdots & \vdots & \vdots \\ T_{n_T} & [HC]_0^n & [CO]_0^n & [O_2]_0^n \end{pmatrix}, \mathbf{Y} = \begin{pmatrix} Y^1(T_1) \\ \vdots \\ Y^1(T_{n_T}) \\ Y^2(T_1) \\ \vdots \\ Y^2(T_{n_T}) \\ \vdots \\ Y^n(T_1) \\ \vdots \\ Y^n(T_{n_T}) \end{pmatrix} \quad (3.24)$$

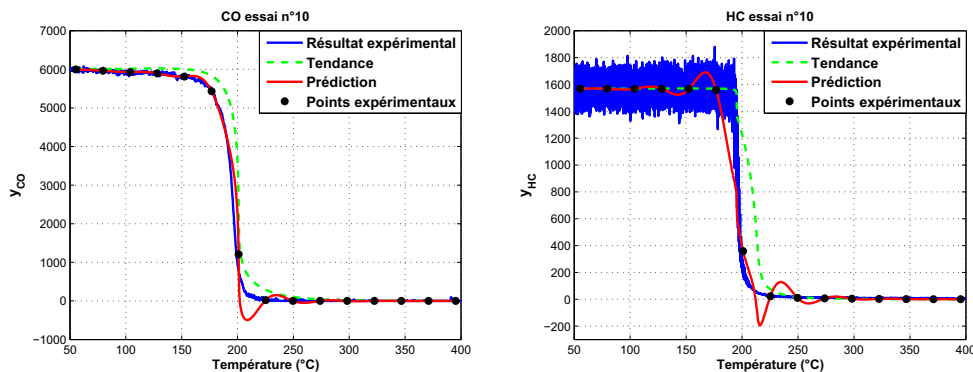
où $\mathbf{T} = [T_1, \dots, T_{n_T}]$ est le vecteur des températures discrétisées (le même pour tous les essais), $[esp]_0^i$ correspond à la concentration initiale de l'espèce esp pour l'expérience i et $Y^i(T_j)$ est la sortie considérée (soit HC ou CO) de l'expérience i à la température T_j .

Les deux sorties sont traitées indépendamment. La matrice d'expériences ainsi définie avec les réponses associées permet l'estimation de β , ainsi que σ^2 et θ , définissant le processus stochastique $z_{\sigma^2, \theta}$. À l'aide de la formule (3.4), il est alors possible de prédire l'évolution de CO et HC en n'importe quel point du domaine.

La démarche mise en place afin d'éprouver notre méthode reprend les deux mêmes étapes que pour le cas maîtrisé : dans un premier temps, nous utilisons les $n = 20$ expériences à disposition pour estimer les paramètres et étudions la prédiction pour un de ces essais afin de vérifier que nos résultats sont interpolants et de bonne qualité, puis dans un second temps, nous n'utilisons que 19 des 20 essais, et comparons le résultat de prédiction avec le résultat expérimental pour l'essai exclu du domaine d'apprentissage. Cette seconde étape est réalisée pour chaque essai et nous permet de mieux observer la qualité de prédiction.

Prédictions pour un essai inclus dans le domaine d'apprentissage

D'un point de vue pratique, nous avons pris $n_T = 15$ points uniformément répartis selon la température sur chacune des $n = 20$ expériences, soit 300 points. À partir de ces points, nous avons estimé les paramètres de notre modèle, et prédisons les sorties HC et CO pour un de ces essais. Les figures 3.3a et 3.3b illustrent respectivement pour la sortie CO et HC , les résultats de prédiction pour un essai (essai 10) compris dans le domaine d'apprentissage et utilisés pour l'estimation des paramètres.



(a) Prédiction pour la sortie CO (b) Prédiction pour la sortie HC
 Figure 3.3 – Résultats de prédiction pour un essai inclus dans le domaine d'apprentissage

Sur les deux figures, la courbe rouge correspond à la prédiction obtenue, la bleue au résultat expérimental, la verte à la tendance (le simulateur), et les points noirs aux points d'expériences relatifs à cet essai qui ont été utilisés pour l'estimation des paramètres.

La première chose que l'on vérifie à l'aide de ces premières figures est l'interpolation des prédictions aux points d'expériences utilisés. De plus, on retrouve, comme dans le cas maîtrisé, des oscillations sur le palier final mais également sur le palier initial pour HC . Deuxième remarque, on constate que pour le CO le simulateur est relativement proche des résultats ce qui n'est pas le cas pour le HC . Néanmoins, dans les deux cas, mis à part les oscillations, les prédictions obtenues sont relativement bonnes, puisque ces dernières sont proches des résultats expérimentaux.

Nous avons étudié, pour tous les jeux, les résultats de prédiction. Bien que le modèle numérique ne représente pas les résultats expérimentaux de manière satisfaisante, les prédictions obtenues par le krigeage à tendance non linéaire, pour chaque sortie, sont proches des résultats expérimentaux, c'est-à-dire que $z_{\sigma^2, \theta}$ joue un rôle complémentaire au modèle numérique.

En regardant pour les autres essais les prédictions, dont nous n'exposons pas les résultats, on observe que les oscillations, qui apparaissent sur les deux paliers, se situent là où la forme de la courbe change brusquement et fortement, soit avant et après la décroissance, et sont plus marquées pour le *HC* du fait de sa plus forte non linéarité. En plus de ne pas respecter la contrainte de décroissance du système (puisque le piège traite les polluants mais n'en fabrique pas), ces oscillations entraînent des concentrations négatives, ce qui est physiquement impossible.

Pour limiter ce phénomène, une possibilité est d'ajouter des points dans l'ensemble d'apprentissage, par une plus grande discrétisation. En prenant 17 points au lieu de 15 uniformément répartis selon la température, il est possible d'observer que les oscillations sont moins fortes mais toujours présentes. Cependant, l'ajout de points est une alternative peu pertinente, puisque d'une part le temps de calcul évolue rapidement selon le nombre de points ajoutés et d'autre part, un trop grand nombre de points peut conduire à des problèmes numériques.

Une solution envisagée pour prévenir ces oscillations est d'utiliser l'information apportée par les dérivées selon la température, puisque sur les paliers ces dérivées sont nulles. Cette solution est l'objet de la section 3.2.

Prédictions pour un essai exclu du domaine d'apprentissage

L'étape suivante a été de vérifier les qualités de prédictions de notre méthode pour des points du domaine non inclus dans l'ensemble d'apprentissage. Pour cela, nous excluons un des jeux de données de l'ensemble utilisé pour l'estimation et comparons les résultats de prédiction aux résultats expérimentaux dont nous disposons.

En pratique, nous avons choisi $n_T = 15$ points uniformément répartis selon la température sur 19 des 20 jeux d'expériences à disposition, on dispose donc de 285 points d'expériences à partir desquels les estimations des paramètres sont effectuées. Au final, nous prédisons la sortie pour l'essai non utilisé, et comparons le résultat obtenu au résultat expérimental. Nous réitérons cela pour chaque essai et chaque sortie.

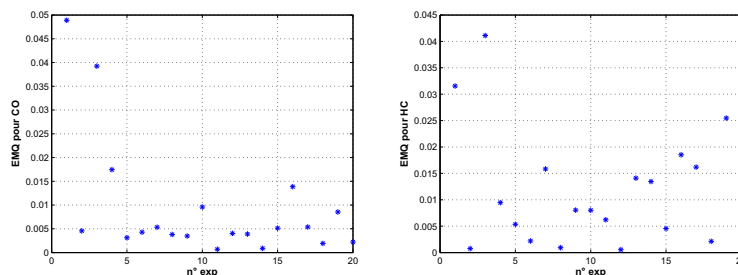
Afin de mesurer la qualité de prédiction pour chacun des jeux, nous calculons pour chacun d'eux, l'erreur moyenne quadratique (EMQ) entre le résultat expérimental (normalisé) et la prédiction obtenue, soit

$$EMQ = \frac{1}{m} \sum_{i=1}^m (y_{exp}(T_i) - \hat{y}(T_i))^2, \quad (3.25)$$

où m est le nombre de températures où la prédiction est effectuée, y_{exp} est la courbe expérimentale et \hat{y} la courbe prédite. Il est important de préciser que les calculs ont été effectués sur les valeurs normalisées (entre 1 et 0) des courbes, afin de pouvoir comparer les résultats entre eux et avec d'autres approches présentées par la suite. De plus, pour les valeurs expérimentales nous utilisons les courbes lissées (voir annexe A pour la méthode de lissage appliquée).

Les résultats de ces EMQ sont exposés sur la figure 3.4a, pour la sortie *CO* et 3.4b pour le *HC*. La première chose que l'on observe est que, d'une manière générale, la méthode de krigeage à tendance non-linéaire développée donne de meilleurs résultats pour la sortie *CO*. Ce constat s'explique par le fait que la non linéarité que l'on observe pour *CO* est moins forte que pour la sortie *HC*.

En ce qui concerne les EMQ obtenues pour *CO*, ceux-ci sont relativement faibles impliquant des résultats de prédiction proche des résultats expérimentaux. Néanmoins, pour trois de ces essais, on remarque des valeurs beaucoup plus élevées : les essais 1, 3 et 4. Pour comprendre pourquoi, pour ces essais, la méthode est moins performante, il est nécessaire de se reporter à leur localisation dans le domaine expérimental (voir figure 1.10).



(a) Valeurs des EMQ pour *CO* (b) Valeurs des EMQ pour *HC*

Figure 3.4 – Valeurs des EMQ obtenues pour chaque jeu de données à disposition selon chacune des deux sorties par krigeage à tendance non-linéaire

L'étude de la répartition de ces points, nous montre que ces trois essais correspondent à des essais extrêmes et isolés. Le fait que ces trois essais soient isolés est la raison de l'échec de la méthode, puisque, on le rappelle, à chaque fois le jeu est exclu de l'ensemble d'apprentissage pour l'estimation des paramètres. Or, on sait que les méthodes de krigeage sont d'autant plus mauvaises dans les zones d'extrapolation que l'on se situe loin de points d'expériences (voir Vasquez [156]).

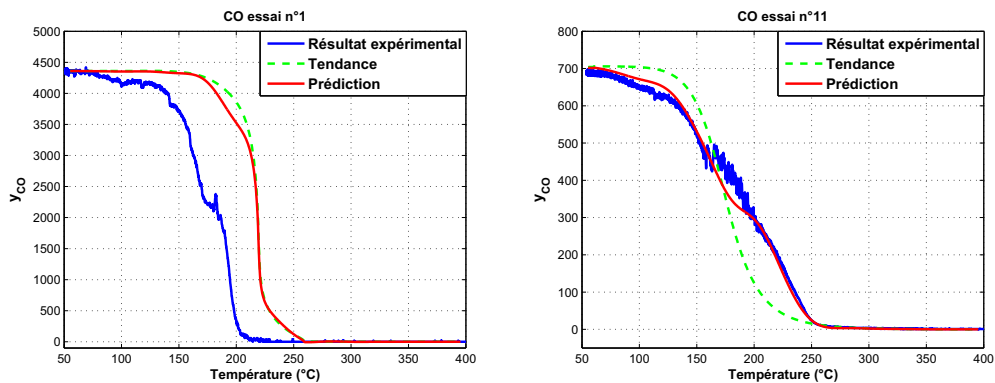
De plus, pour les essais 1 et 3, ce sont des essais à fortes teneurs en polluants, à l'inverse l'essai 4 correspond à un cas de très faible teneur en polluants. Cette faible teneur en polluants explique que l'essai 4 bien que mal prédit, soit meilleur que les deux autres. En effet, il est possible de montrer que le comportement du simulateur est meilleur dans ce cas de figure.

Pour ce qui est de *HC*, comme nous l'avons mentionné sa forte non linéarité est à l'origine des résultats de prédiction moins proches des résultats expérimentaux. On retrouve les essais 1 et 3 qui ont des EMQ très élevées, et ce pour les mêmes raisons qu'évoquées précédemment (essais isolés et forte teneur en polluants), cependant pour l'essai 4 le constat est différent. Pour cet essai, le simulateur est proche des résultats observés expérimentalement (du fait de la faible teneur en *HC*), donc bien qu'isolé, la prédiction est bonne.

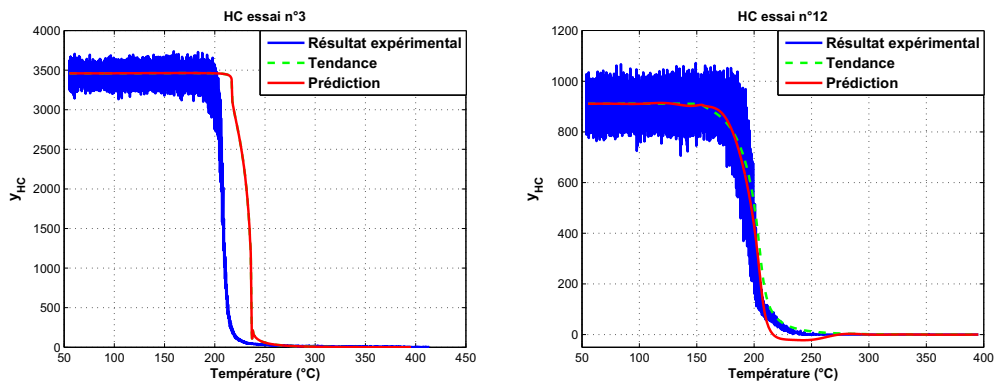
On retrouve ce constat pour les autres essais. En effet, on observe que les essais les plus mauvais au sens de l'EMQ, mis à part les essais 1 et 3, correspondent à des essais où les teneurs en polluants *CO*, *HC* ou les deux, sont élevées (essais 19, 16, 17, 9). On retrouve donc le fait que plus la teneur en polluants est élevée, plus le simulateur est inadéquat à représenter la sortie *HC*. Le complément de modèle joue bien son rôle mais ne suffit pas à compenser l'écart observé de manière satisfaisante.

Afin d'illustrer les prédictions que l'on obtient, nous exposons les pires et meilleurs résultats pour chaque sortie selon l'EMQ. Les figures 3.5a et 3.5b correspondent respectivement aux pire et meilleur cas pour *CO*, et les figures 3.6a et 3.6b celles respectivement pour *HC*. Pour chacune de ces figures, la courbe rouge correspond à la prédiction obtenue par krigeage à tendance non-linéaire, la courbe bleue aux résultats expérimentaux et la courbe en pointillé verte représente la tendance donnée par le simulateur.

L'étude de ces courbes vient conforter les premières conclusions émises. En effet, pour ce qui est des pires cas, on observe que la prédiction correspond à la tendance donnée par le simulateur. L'absence de points dans le domaine expérimental proche de ces essais, conjuguée à des petites portées est à l'origine de ce phénomène. Le complément de modèle $z_{\sigma^2, \theta}$ étant nul dans ce cas (ou presque en tout point), la prédiction correspond aux résultats du simulateur.



(a) Pire résultat selon l'EMQ pour CO : essai 1 (b) Meilleur résultat selon l'EMQ pour CO : essai 11
 Figure 3.5 – Résultats de la prédiction par krigeage à tendance non-linéaire pour la sortie CO : pire et meilleur cas au sens de l'EMQ



(a) Pire résultat selon l'EMQ pour HC : essai 3 (b) Meilleur résultat selon l'EMQ pour HC : essai 12
 Figure 3.6 – Résultats de la prédiction par krigeage à tendance non-linéaire pour la sortie HC : pire et meilleur cas au sens de l'EMQ

En revanche, pour les meilleurs cas, les prédictions des deux sorties ont des comportements différents. Pour ce qui est de CO , la forme du modèle numérique est différente de la forme des résultats expérimentaux, cependant la prédiction est bonne puisqu'elle arrive à s'ajuster aux données observées expérimentalement. Le complément de modèle joue bien son rôle.

Pour ce qui est du cas présenté pour HC , le simulateur est proche de la courbe bleue. La prédiction est donc bonne, cependant comme sur la figure 3.3b, on constate une oscillation en fin de décroissance conduisant à des concentrations négatives. Ce qui d'un point de vue expérimentateur est inacceptable.

3.1.5 Conclusions et limites

La méthode d'estimation des paramètres introduite et appliquée a montré sur notre cas maîtrisé de bons résultats de prédiction. De plus, nous avons également pu confronter ces résultats à ceux obtenus par un krigeage classique, et nous en avons conclu que ce type de méthode est pertinent pour des cas où l'on remplace la tendance linéaire par un modèle non-linéaire dont il nous faut estimer les paramètres.

Cependant, l'application sur notre cas d'étude a mis en évidence les limites de cette méthode. En effet, nous avons constaté que bien que les prédictions soient meilleures au sens de la proximité et de la forme de la courbe que le simulateur initial, on observe l'apparition d'oscillations, notamment sur les paliers initiaux et finaux. Ce comportement n'a aucun sens physique, et peut conduire à des concentrations négatives.

Afin de répondre à ce problème, la solution envisagée est de conditionner le krigeage par la dérivée. En effet, sur les deux paliers, on a une dérivée nulle. Cette méthode est connue sous le nom de cokrigeage.

L'intérêt de cette méthode est de forcer la prédiction sur ces paliers à une valeur correspondant à la concentration initiale de l'espèce concernée, pour le premier palier, et à 0 sur le palier final. De plus, cette méthode permet une amélioration de la prédiction sur des zones sous-échantillonnées et en extrapolation (Tapsoba *et al.*, [151]). C'est l'objet de la prochaine partie.

3.2 Conditionnement par la dérivée

Les oscillations apparues dans les prédictions par notre modèle de krigeage à tendance non-linéaire, surviennent principalement quand la forme de la réponse change fortement. Ces brusques variations sont expliquées par le fait que le processus n'est pas stationnaire. Les hypothèses de modélisation ne sont pas vérifiées et donc la prédiction n'est pas adéquate. Pour contraindre la forme de la courbe, nous suggérons dans cette partie d'inclure la connaissance des dérivées dans la procédure d'estimation.

En outre, puisque la dérivée est un opérateur linéaire, la loi jointe entre réponses et dérivées reste gaussienne. Par conséquent, l'inférence et la prédiction sont menées de la même manière que précédemment. Le chapitre 2 présente dans le cas du krigeage universel, l'équivalence des formules de prédicteurs et de variance, avec et sans conditionnement par la dérivée.

Partant du fait que les deux paliers initiaux et finaux ont des dérivées nulles, conditionner le krigeage par les dérivées peut être une bonne solution afin de remédier au problème d'oscillations. Tapsoba *et al.* [151] montrent que ce conditionnement par la dérivée améliore la prédiction sur des zones sous échantillonnées et en extrapolation.

Nous nous sommes inspirés de leurs travaux, et les avons adaptés à notre cas d'étude. Notamment, le conditionnement est effectué uniquement sur les paliers. Le principal intérêt de ce conditionnement par la dérivée est de conserver ces paliers observés expérimentalement. Les dérivées selon la température en chaque point d'expériences sur ces paliers sont nulles.

Comme précédemment, nous appliquons tout d'abord la méthode sur notre cas maîtrisé, et comparons les résultats à ceux par krigeage à tendance non-linéaire sans conditionnement. Les résultats probants sur ce cas, nous permettent dans un second temps d'appliquer la méthode sur notre cas d'étude. Enfin nous concluons et présentons, grâce aux résultats exposés, les avantages et inconvénients de notre méthode.

3.2.1 Présentation de la démarche

Le conditionnement par la dérivée selon la température correspond à un krigeage multisortie appelé également cokrigeage (Chilés et Delfiner [29], Santner *et al.* [138]). Une des sorties est la dérivée selon la température, et on sait exprimer la covariance entre la réponse et sa dérivée à travers une matrice \mathbf{C} . Dans ses travaux de thèse, Vazquez [156] introduit le cokrigeage et présente plusieurs applications liées à la prédiction de la dérivée. Il montre notamment que la prédiction d'une sortie avec informations sur la dérivée permet d'améliorer les prédictions,

par exemple, lorsque l'on connaît des comportements du système aux limites de son domaine de définition, tels que des gradients nuls.

Dans notre cas, on sait que les courbes ont une forme générale composée d'un premier palier où rien ne se passe (le piège n'est pas actif, la concentration est constante), une décroissance plus ou moins brutale (le piège entre en action) et un palier final où la concentration est également constante (tous les polluants sont convertis). On souhaite donc incorporer l'information des gradients nuls selon la température (une des entrées) dans notre modèle de krigeage à tendance non-linéaire.

Soient $\mathbf{s} = (s_1, \dots, s_{n_1})$ le vecteur des n_1 points d'expériences de taille p où la sortie est connue, $\mathbf{v} = (v_1, \dots, v_{n_2})$ le vecteur des n_2 points d'expériences de taille p où l'on connaît la dérivée partielle selon la première direction (la température).

On pose :

$$Y = \left(Y_1, \dots, Y_{n_1}, Y_1^{(1)}, \dots, Y_{n_2}^{(1)} \right)^T, \quad (3.26)$$

et

$$F = \left(f(s_1, \boldsymbol{\beta}), \dots, f(s_{n_1}, \boldsymbol{\beta}), f^{(1)}(v_1, \boldsymbol{\beta}), \dots, f^{(1)}(v_{n_2}, \boldsymbol{\beta}) \right) \quad (3.27)$$

où l'exposant (1) fait référence aux sorties relatives aux dérivées partielles selon la première composante (la température dans notre cas), par exemple $f^{(1)}(v_1, \boldsymbol{\beta}) = \partial f(v_1, \boldsymbol{\beta}) / \partial v_{1,1}$ avec $v_{1,1}$ correspondant à la température. La matrice d'expériences utilisée pour l'estimation est donc définie par $\mathbf{X} = [\mathbf{s} \ \mathbf{v}]^T$, avec \mathbf{s} et \mathbf{v} définie comme dans la section 3.1.4.2 (voir équation 3.24).

Concernant le calcul des dérivées expérimentales partielles selon la température, celles-ci ne peuvent être calculées à partir des résultats expérimentaux bruts, elles sont donc calculées par différences finies à partir des données expérimentales lissées. L'annexe A décrit l'étape de lissage que nous avons mise au point. Cette dernière assure notamment la contrainte de décroissance de la sortie.

En nous basant sur les travaux de Santner *et al*, [138], nous avons fait le choix d'utiliser une corrélation spatiale gaussienne anisotropique, définie par,

$$\forall h \in \mathbb{R}^p, R_{\boldsymbol{\theta}}(h) = \exp \left\{ - \sum_{i=1}^p \theta_i h_i^2 \right\}. \quad (3.28)$$

Cette première relation définit les corrélations entre les éléments de Y_i , $1 \leq i \leq n_1$. Afin de calculer les fonctions de corrélations entre les éléments de $Y_i(\cdot)$ et $Y_j^{(1)}(\cdot)$ deux à deux, nous partons de la corrélation (3.28) que nous dérivons par rapport à la température, soit la première composante des points s_i (ou v_j), que l'on note par $s_{j,1}$ (resp. $v_{j,1}$).

Au final, on obtient les formules de covariances suivantes :

$$\text{Cov} \left(Y(s_i), Y^{(1)}(v_j) \right) = \sigma^2 2\theta_1 (s_{i,1} - v_{j,1}) R_{\boldsymbol{\theta}}(s_i - v_j) \quad (3.29)$$

$$\text{Cov} \left(Y^{(1)}(v_i), Y^{(1)}(v_j) \right) = \sigma^2 (2\theta_1 - 4\theta_1^2 (v_{i,1} - v_{j,1})^2) R_{\boldsymbol{\theta}}(v_i - v_j) \quad (3.30)$$

On a donc,

$$\mathbf{C} = \begin{pmatrix} C_{00} & C_{01} \\ C_{01}^T & C_{11} \end{pmatrix}, \quad (3.31)$$

où C_{00} est la $n_1 \times n_1$ matrice des corrélations parmi les éléments de $Y_i, 1 \leq i \leq n_1$, calculée par la corrélation spatiale, équation 3.28, C_{01} est la $n_1 \times n_2$ matrice des corrélations entre Y_i et $Y_j^{(1)}, 1 \leq j \leq n_2$, donnée par l'équation 3.29, et C_{11} est la $n_2 \times n_2$ matrice des corrélations parmi les éléments de $Y_j^{(1)}$, évaluée par la relation 3.30.

Enfin, le vecteur de corrélation, $\mathbf{c} = (c_0, c_1)^T$, entre un nouveau point x_0 et les points d'expériences est défini par,

$$c_0 = (R\boldsymbol{\theta}(x_0 - s_i))_{1 \leq i \leq n_1} \quad (3.32)$$

$$c_1 = (2\theta_1(x_{0,1} - v_{j,1})R\boldsymbol{\theta}(x_0 - v_j))_{1 \leq j \leq n_2} \quad (3.33)$$

Les diverses notations ayant été définies, les formules de prédicteur de krigeage et l'algorithme d'estimation des paramètres restent les mêmes que celles introduites dans la section 3.1, mise à part le fait que la matrice \mathbf{C} se substitue à la matrice \mathbf{R} , et le vecteur \mathbf{c} au vecteur \mathbf{r} . Par conséquent, le prédicteur de krigeage est défini par,

$$\hat{\mathbf{y}}(\mathbf{x}) = \mathbf{c}(\mathbf{x})^T \mathbf{C}^{-1} \mathbf{Y} - (f(\mathbf{S}, \boldsymbol{\beta})^T \mathbf{C}^{-1} \mathbf{c}(\mathbf{x}) - f(\mathbf{x}, \boldsymbol{\beta}))^T (f(\mathbf{S}, \boldsymbol{\beta})^T \mathbf{C}^{-1} f(\mathbf{S}, \boldsymbol{\beta}))^{-1} f(\mathbf{S}, \boldsymbol{\beta})^T \mathbf{C}^{-1} \mathbf{Y}. \quad (3.34)$$

3.2.2 Application sur le cas maîtrisé

Comme précédemment, afin d'évaluer la méthode, nous l'appliquons sur notre cas maîtrisé. Les résultats et conclusions nous guideront vers l'application sur notre cas d'étude.

Le plan d'expériences est construit de la manière suivante : trois concentrations initiales, huit temps fixes sont choisis et dix températures sont prises uniformément sur chaque courbe. Ainsi, le plan contient 240 points.

Dans le cas présenté ici, le choix de la prise en compte de l'information de la dérivée s'est porté en chacun des points d'expériences, et non pas uniquement sur les paliers. Autrement dit, $\mathbf{s} = \mathbf{v}$ dans ce cas. La théorie de krigeage à tendance non-linéaire est d'abord appliquée, puis l'information de la dérivée est ajoutée, afin de conclure sur l'apport de cette information.

Il est également intéressant de préciser que pour ce cas maîtrisé, la valeur de la dérivée pour le "simulateur", est calculée par une formule analytique, déterminée à partir du système cinétique introduit dans le chapitre 1, section 1.2.1.

La figure 3.7 présente les résultats obtenus. La courbe pleine bleue correspond à la courbe expérimentale. La courbe pleine rouge représente la prédiction obtenue par krigeage à tendance non-linéaire avec les dérivées, et celle en pointillée, par krigeage à tendance non-linéaire, sans dérivée. Les points noirs correspondent aux points expérimentaux utilisés dans le cas présenté.

On observe que les deux courbes pleines sont quasiment confondues. De plus, comme nous l'espérons, le problème d'oscillations sur le palier final conduisant à des concentrations négatives a disparu.

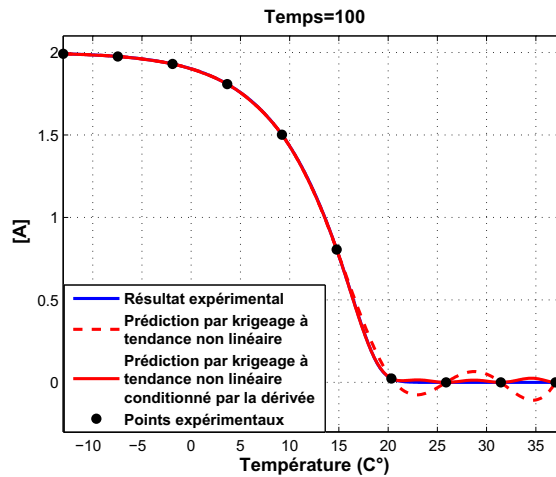


Figure 3.7 – Prédiction du cas maîtrisé par krigeage à tendance non-linéaire sans dérivée (courbe pointillée rouge), et avec dérivée (courbe pleine rouge) comparées aux résultats expérimentaux (courbe pleine bleue)

On en conclut que cette méthode corrige bien le problème d'oscillations comme attendu et par conséquent la prédiction est très bonne. Ainsi, l'application sur notre cas d'étude apparaît judicieuse afin d'obtenir une bonne représentation de notre système, sans comportements atypiques tels que des concentrations négatives.

Cette méthode nous permet également de prédire la valeur de la dérivée. En effet, le conditionnement est équivalent à un krigeage multi-sorties pour lequel la dérivée est choisie comme seconde sortie. La figure 3.8 expose le résultat de prédiction pour la dérivée.

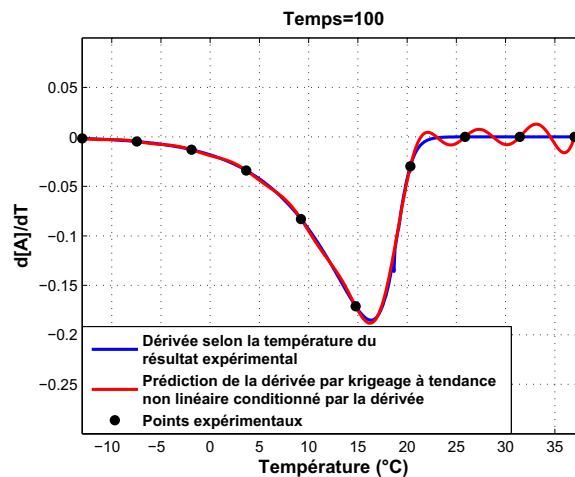


Figure 3.8 – Prédiction de la dérivée pour le cas maîtrisé (courbe rouge) comparées à la dérivée expérimentale (courbe bleue)

On s'aperçoit que la prédiction de la dérivée est bonne, mis à part pour des températures élevées, où l'on voit apparaître des oscillations. Celles-ci sont à l'origine des oscillations très petites observées sur le palier final de la figure 3.7. Néanmoins, l'importance de ces oscillations est à minimiser car d'une part celles-ci sont faibles, et d'autre part elles n'entraînent pas de concentrations négatives.

3.2.3 Application sur le cas d'étude

Les résultats obtenus sur notre cas maîtrisé justifient l'application du krigeage à tendance non-linéaire conditionné par la dérivée (KNLD) sur notre cas d'étude. Le conditionnement par la dérivée n'est effectué que sur les paliers initiaux et finaux, où les valeurs de dérivées sont en dessous d'un certain seuil, proche de zéro.

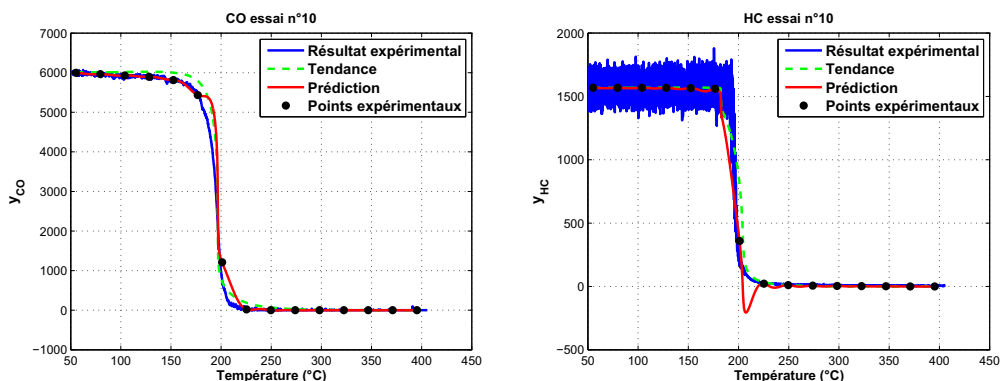
Dans un premier temps, nous avons essayé, comme sur le cas maîtrisé, de conditionner par les dérivées en chacun des points d'expériences, mais les résultats nous ont montré que le processus était trop contraint pour s'ajuster efficacement et nous rappelons de plus, que les oscillations surviennent sur les deux paliers à proximité de la décroissance, d'où notre choix de ne tenir compte de l'information de la dérivée uniquement sur les paliers.

La démarche d'application sur notre cas d'étude reprend le cheminement suivi dans la section 3.1.4.2, c'est-à-dire dans un premier temps, tous les essais sont pris pour l'estimation des paramètres et nous prédisons la sortie pour l'essai 10, puis dans un second temps, nous réalisons la prédiction pour tous les essais, avec chaque fois l'exclusion du jeu prédit de l'ensemble d'apprentissage.

La première étape nous permet de constater l'amélioration apportée par le conditionnement, la seconde de conclure quant à la pertinence de notre méthode.

Prédictions pour un essai inclus dans le domaine d'apprentissage

Comme précédemment, nous avons pris $n_T = 15$ points uniformément répartis selon la température sur chacune des $n = 20$ expériences, soit $n_1 = 300$ points contenus dans le vecteur \mathbf{s} . Parmi ces points, ceux dont la dérivée selon la température est inférieure à un certain seuil (proche de 0), sont sélectionnés pour former le vecteur \mathbf{v} . À partir de ces deux vecteurs, nous construisons la matrice d'expériences $\mathbf{X} = [\mathbf{s} \ \mathbf{v}]^T$ et le vecteur des sorties \mathbf{Y} associées, et estimons les paramètres de notre modèle, afin de prédire les sorties HC et CO pour un de ces essais. Les figures 3.9a et 3.9b illustrent respectivement pour la sortie CO et HC , les résultats de prédiction pour l'essai 10, le même que celui présenté dans la section 3.1.4.2, pour pouvoir comparer les résultats obtenus par les deux méthodes.



(a) Prédiction pour la sortie CO (b) Prédiction pour la sortie HC
 Figure 3.9 – Résultats de prédiction pour un essai inclus dans le domaine d'apprentissage

Première remarque, le simulateur (courbe en pointillé verte) est plus proche des résultats expérimentaux que ce que l'on observe sur la courbe 3.3. Cependant, les résultats de prédiction sont relativement proches des résultats précédents.

Lors de l'application du krigeage à tendance non linéaire, on rappelle que nous avons observé des oscillations en fin de décroissance pour CO et HC et également au début de la décroissance pour HC .

On en conclut que le conditionnement par la dérivée mis en place permet de répondre en partie à ce problème. En effet, pour la sortie CO , on remarque que non seulement la prédiction est très bonne, mais de plus, les oscillations ont disparus. Pour la sortie HC , les oscillations sur le palier initial ont été éliminées, mais, en revanche, il subsiste sur le palier final, à la base de la décroissance une oscillation.

Cette constatation met en lumière la plus grande difficulté d'estimation des paramètres pour la sortie HC . En effet, on a pu voir que la non linéarité pour cette sortie est très forte, la non-stationnarité du phénomène est d'autant plus marquée, alors que le noyau utilisé dans notre approche est un noyau stationnaire. Autrement dit, les observations du complément de modèle z , ne peuvent être considérées comme stationnaires en T . Nous revenons sur ce point dans la section conclusions et limites (section 3.2.4).

Prédictions pour un essai exclu du domaine d'apprentissage

La méthode d'échantillonnage utilisée est identique à celle des précédentes applications, c'est-à-dire $n_T = 15$ points uniformément répartis sur 19 des 20 expériences, et la prédiction est effectuée sur la vingtième. Les valeurs des EMQ (équation 3.25) pour chacun des essais sont données sur les figures 3.10a et 3.10b.

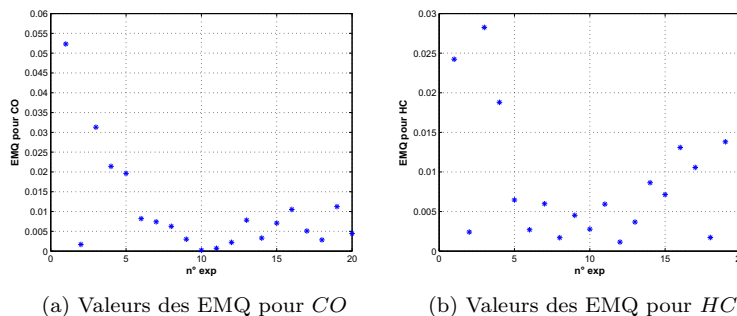


Figure 3.10 – Valeurs des EMQ obtenues pour chaque jeu de données à disposition selon chacune des deux sorties par krigeage à tendance non-linéaire conditionné par la dérivée

D'une manière générale, pour les deux sorties, les valeurs des EMQ sont inférieures aux valeurs observées des EMQ pour le krigeage à tendance non-linéaire sans conditionnement. On constate donc que l'information apportée par les dérivées entraîne une réelle amélioration de nos prédictions.

Pour ce qui est de CO , on retrouve nos trois essais à forte EMQ (essais 1, 3 et 4). Dans les autres cas, les valeurs des EMQ sont soit inférieures soit très proches des valeurs précédentes.

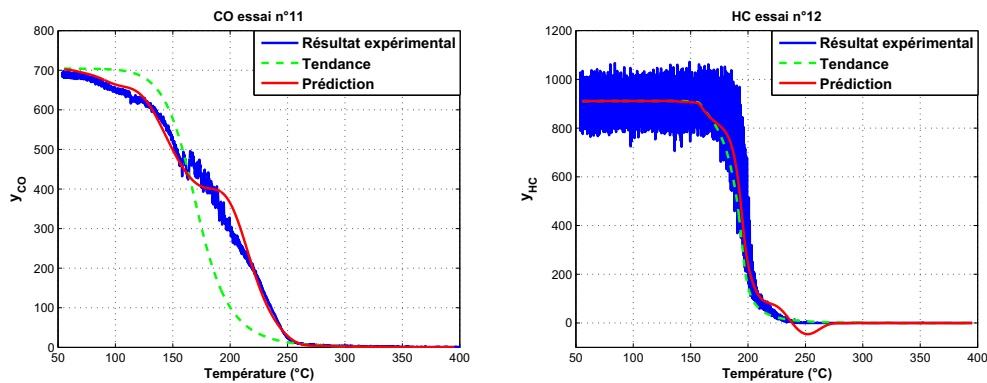
On note tout de même que seul un cas est radicalement modifié, l'essai 5, dont l'EMQ est plus élevée (valeur multipliée par 4). L'essai 5 est un essai à très forte teneur en O_2 et très faible en HC , l'oxydation du CO est donc rapide et le comportement en sortie du CO est atypique (pas de points proches). Cette particularité peut expliquer que pour ce cas le conditionnement par la dérivée dégrade la prédiction.

En ce qui concerne les valeurs des EMQ pour HC , on remarque une très nette amélioration, par rapport aux valeurs obtenues par KNL, d'une manière générale. Si l'on étudie de plus près ces valeurs une à une, et qu'on les compare à celles de la figure 3.4b, on constate que des valeurs très faibles précédemment (essai 2, 15, 20) sont

légèrement moins bonnes et inversement les valeurs très fortes sont meilleures (essais 1, 3, 7, 10, 19). On assiste à un nivellement de ces valeurs, qui d'une manière générale se traduit par une amélioration des prédictions.

Pour illustrer ces améliorations de prédiction, nous exposons les résultats pour la sortie CO de l'essai 11, figure 3.11a et pour la sortie HC de l'essai 12, figure 3.11b, et les comparons à ceux présentés précédemment (figures 3.5b et 3.6b). Ces deux essais correspondent aux meilleurs essais au sens de l'EMQ pour le KNL.

Les résultats pour les pires cas (essais 1 pour CO et essai 3 pour HC) ne sont pas exposés car comme précédemment, on observe que la prédiction correspond au simulateur, du fait du manque d'information pour ces essais (zone d'extrapolation éloignée). On notera toutefois que pour le HC , le simulateur est de meilleure qualité, ceci explique la valeur plus faible de l'EMQ.



(a) Résultat du KNLD pour l'essai 11, sortie CO (b) Résultat du KNLD pour l'essai 12, sortie HC

Figure 3.11 – Résultats de la prédiction par krigeage non-linéaire conditionné par la dérivée pour les deux essais exposés dans la section 3.1.4.2

Tout d'abord, pour les deux sorties, les prédictions sont très bonnes. La prédiction de CO pour l'essai 11 arrive à représenter le palier intermédiaire et est très proche du résultat expérimental alors que le modèle numérique (simulateur) ne respecte pas la forme même de la courbe.

Sur la figure 3.11b, la prédiction de la sortie HC pour l'essai 12 est également très proche du résultat expérimental. On observe une amélioration de cette prédiction comparée à celle obtenue par KNL (figure 3.6b). En revanche, bien que l'oscillation sur le palier final soit plus localisée, cette dernière est toujours présente, et entraîne des prédictions de concentrations négatives.

Là encore, la non stationnarité des résidus du phénomène observé est à l'origine de ce constat. On en conclut donc que pour le CO , la méthode donne de bons résultats, en revanche pour HC , ce n'est pas le cas.

Afin de mieux visualiser les limites de la méthode pour la sortie, nous donnons pour l'essai 1, la prédiction de la sortie HC sur la figure 3.12.

Ce contre-exemple du bon comportement du KNLD illustre bien le problème de non-stationnarité. En effet, on constate que dans le cas où le simulateur est très mauvais, la prédiction arrive à être meilleure sur la décroissance, à ne plus avoir d'oscillations sur le palier initial, mais en revanche, on observe une forte oscillation, en fin de décroissance, un comportement inacceptable.

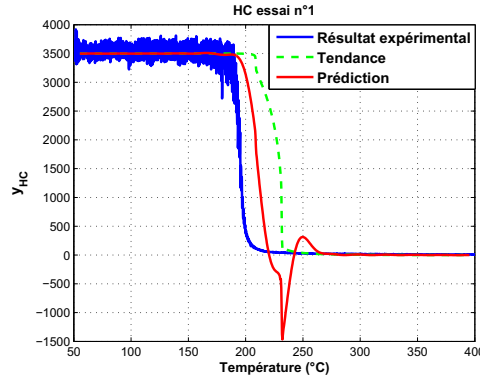


Figure 3.12 – Contre-exemple du bon comportement du KNLD sur notre cas d'étude

3.2.4 Conclusions et limites

La mise en place de la méthode de krigeage à tendance non-linéaire conditionné par la dérivée a montré de bons résultats sur notre cas maîtrisé, mais également sur notre cas d'étude. Cette méthode a également montré ses limites sur des cas extrêmes.

Deux raisons expliquent l'échec de la méthode sur notre cas d'étude : la non stationnarité des résidus selon la température et la non prise en compte du caractère fonctionnel des courbes (contraintes non respectées). Cette méthode est l'aboutissement de diverses améliorations successives : le krigeage à tendance non-linéaire (fortes oscillations), conditionnement par la dérivée en tout point (oscillations moins fortes mais processus trop contraint) et enfin conditionnement uniquement sur les paliers (résultats meilleurs sauf dans certains cas).

Divers autre points ont été pris également en considération, mais avec des résultats équivalents. Notamment, nous nous sommes attardés sur le choix du noyau. On rappelle que le noyau choisi au cours des différentes méthodes est un noyau gaussien. Par conséquent, nous avons exploré différents noyaux, qui ne nous ont pas permis de résoudre les problèmes rencontrés. Tout d'abord, nous avons testé d'autres noyaux stationnaires classiques tels que le noyau exponentiel, de Matérn, sphérique,... mais qui ont donné des résultats équivalents. La non stationnarité du problème est à l'origine de la non pertinence des noyaux stationnaires utilisés.

Les résidus observés entre le simulateur et les résultats expérimentaux ne sont pas de variance constante en fonction de la température, on observe deux intervalles de température où les résidus sont peu variables (les deux paliers sur lesquels le conditionnement a été effectué) et un intervalle sur lequel la réaction chimique se déroule, et où les résidus sont fortement variables. On pense donc à utiliser un noyau de covariance qui prenne ces observations en compte. La fonction choisie pour traduire la non stationnarité doit être fonction de la température, or on sait que la variance est la plus forte à la température de "light-off", qui est corrélée aux conditions opératoires. Partant de ce constat, nous avons mis en place un noyau non stationnaire, défini par,

$$R_{\theta}(\mathbf{u}, \mathbf{v}) = \exp \left\{ - \sum_{i=1}^p \theta_i \|u_i - v_i\|^2 \right\} g(\mathbf{u})g(\mathbf{v}), \text{ où } g(\mathbf{u}) = \exp \left\{ - \|u_T - m(\mathbf{u})\|^2 \right\}, \quad (3.35)$$

avec $m(\mathbf{u}) = \alpha_1 u_{CO} + \alpha_2 u_{HC}$ et u_i , la composante de u relative à l'entrée i . Ce noyau non-stationnaire en T est fonction des concentrations initiales de CO et HC les deux espèces étudiées en sortie. Les paramètres de non stationnarité α sont maintenant contenus dans le vecteur des hyper paramètres à estimer. Mais dans ce cas, nous avons pu observer que la procédure de minimisation se concentre uniquement sur les paramètres de non stationnarité, ce qui a pour conséquence de ne pas modifier, ou très peu, les hyperparamètres de notre modèle.

Les résultats obtenus à l'aide de ce noyau donnent des valeurs aberrantes, c'est la raison pour laquelle nous n'exposons pas les résultats et ne détaillons pas plus l'étude que nous avons effectuée sur les noyaux utilisés.

On voit donc que l'idée de considérer que le modèle numérique rend compte d'un effet moyen qui doit être corrigé d'un effet modélisé par un processus gaussien $z_{\sigma^2, \theta}$ n'est pas un succès, pour deux raisons : la première est que les réalisations observées de z ne peuvent être considérées comme stationnaire en T , la non stationnarité étant non trivialement modélisable, et la seconde que les courbes obtenues doivent respecter impérativement certaines contraintes, ce qui n'est pas le cas si on ne les considère pas comme des courbes. En effet, les sorties, une fois normalisées, doivent être strictement décroissantes toujours inférieures à 1 et positives car elles représentent des concentrations.

L'idée est alors de se ramener à une modélisation plus classique, c'est-à-dire se ramener à une modélisation dans laquelle le problème de non stationnarité est contourné. Plus exactement, une modélisation où seule la non stationnarité en variance demeure. Le travail effectué sur les dérivées, nous a offert une piste. Si l'on regarde, le résultat de prédiction de la dérivée dans le cas maîtrisé, figure 3.8, on s'aperçoit que celle-ci est relativement bonne. L'idée alors explorée est de modéliser la dérivée selon l'avancement. L'intérêt d'une telle approche est qu'aux bords du domaine (en 0 et 1) la dérivée est nulle, et par conséquent, dans cette représentation la non-stationnarité est contournée. La difficulté réside dans la reconstruction de la fonction à partir de la dérivée. Cette approche est l'objet de la prochaine partie.

3.3 Modélisation de la dérivée par rapport à l'avancement

Dans cette section, nous cherchons à trouver une autre représentation de notre problème pour laquelle la non-stationnarité est contournée. Dans l'article de Favergeon *et al.* [58], les auteurs proposent de remplacer l'étude de l'avancement de la réaction en fonction du temps par l'étude de la vitesse de réaction (dérivée de l'avancement en fonction du temps) en fonction de l'avancement.

L'avancement en cinétique, plus particulièrement en catalyse, correspond au taux de conversion d'une espèce. Autrement dit pour notre cas, un avancement de 0, symbolise le fait que le polluant n'est pas encore traité (palier initial) et un avancement de 1, représente le fait que tout le polluant est converti. Cette notion, notée α est définie par

$$\alpha(T) = \frac{[A]_0 - [A]_T}{[A]_0}. \quad (3.36)$$

avec $[A]_0$ la concentration initiale de l'espèce A , et $[A]_T$ celle relative à la température T .

Nous proposons dans cette partie de modéliser la dérivée (selon la température) des sorties étudiées (CO et HC) en fonction de l'avancement. Cette modélisation comporte un avantage majeur, à savoir que les résidus dans ce cas sont stationnaires. En effet, les deux paliers correspondent à une infinité de points en lesquels la dérivée selon T est nulle. Les résidus entre le simulateur et le résultat expérimental ne sont plus localisés. La figure 3.15 illustre la forme des résidus que l'on observe.

Après avoir présenté la démarche mise en place dans une première section, nous réalisons une étude préliminaire, dans laquelle nous définissons clairement ce qu'est l'avancement, la forme des dérivées et quelques considérations pratiques quant à l'application sur notre cas d'étude. Ensuite, nous appliquons la méthode et présentons, dans un premier temps les résultats obtenus pour la prédiction de la dérivée puis, dans un second temps, ceux après intégration. Cette intégration est réalisée en deux étapes qui nous permettent au final de présenter le calcul de l'intervalle de confiance. Enfin, nous détaillons les conclusions et les limites de cette approche, et évoquons les perspectives d'application pour d'autres cas.

3.3.1 Présentation de la démarche

Notre volonté de déterminer une représentation dans laquelle les résidus sont stationnaires, nous a guidés vers la modélisation directe de la dérivée en fonction de la température, selon l'avancement. Cette représentation (Favergeon *et al.*, [58]) est peu courante dans la littérature. Classiquement en cinétique, les sorties sont étudiées en fonction de la température ou du temps.

Dans cette représentation, on observe des courbes de résidus entre le modèle numérique et la réalité expérimentale, plus classique (voir figure 3.15), dans le sens où ces résidus sont stationnaires et ont un comportement permettant une approche couramment utilisée, à savoir le krigeage avec un noyau stationnaire. On rappelle que le problème de non stationnarité est la cause principale de l'échec sur notre cas d'application des méthodes précédemment introduites et appliquées.

Avant d'introduire plus précisément la forme des courbes et la méthode appliquée dans cette section, il est nécessaire de définir à quoi correspond, en catalyse, la notion d'avancement. D'une manière générale, ce terme, noté α , correspond à la quantité de matière d'un réactif à un instant donné. En catalyse, ce concept se ramène donc au taux de conversion d'une espèce. L'avancement, défini par (3.36), correspond au contraire de la sortie normalisée (1 sur le palier initial et 0 sur le palier final). Les figures 3.13 illustrent, pour un essai, les sorties CO et HC en avancement en fonction de la température. Les courbes bleues représentent les résultats expérimentaux (lissés) et les courbes vertes les résultats du modèle numérique.

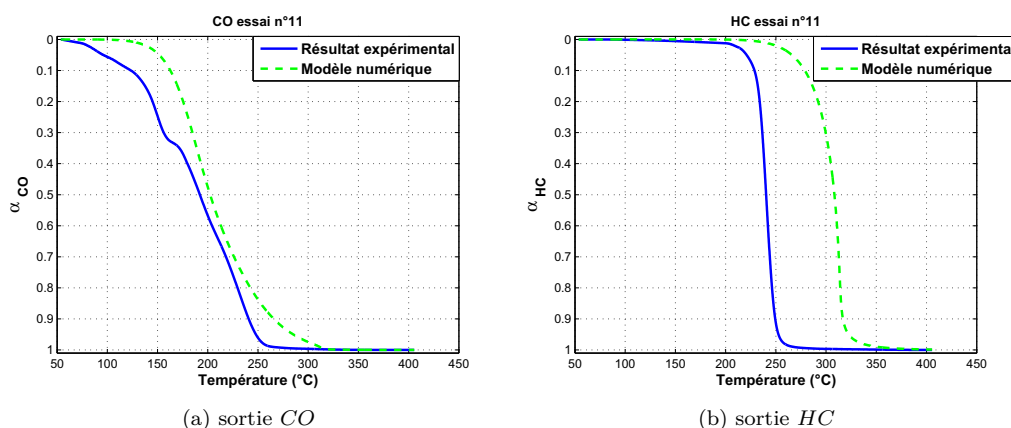


Figure 3.13 – Représentations des sorties CO et HC en avancement en fonction de la température pour un essai (essai 11)

Les figures 3.14 présentent les dérivées de l'avancement en fonction de la température selon l'avancement. La figure 3.14a est respectivement à la sortie CO et la figure 3.14b à la sortie HC . Sur ces deux figures, les courbes bleues sont les dérivées expérimentales par rapport à la température et les courbes en pointillés vertes, celles relatives au modèle numérique. Pour illustrer la forme des dérivées expérimentales et simulées, nous avons fixé un vecteur de paramètre cinétique de manière arbitraire. On rappelle (voir section 3.2.1) que les dérivées selon la température sont calculées par différences finies.

La dérivée est nulle pour des avancements de 0 et 1 et une étude plus attentive de la forme de ces résidus, nous fait penser à un pont brownien. La loi de ce dernier est définie comme étant la loi d'un mouvement brownien partant de 0 à l'instant 0 et contraint de passer par 0 en $t = 1$.

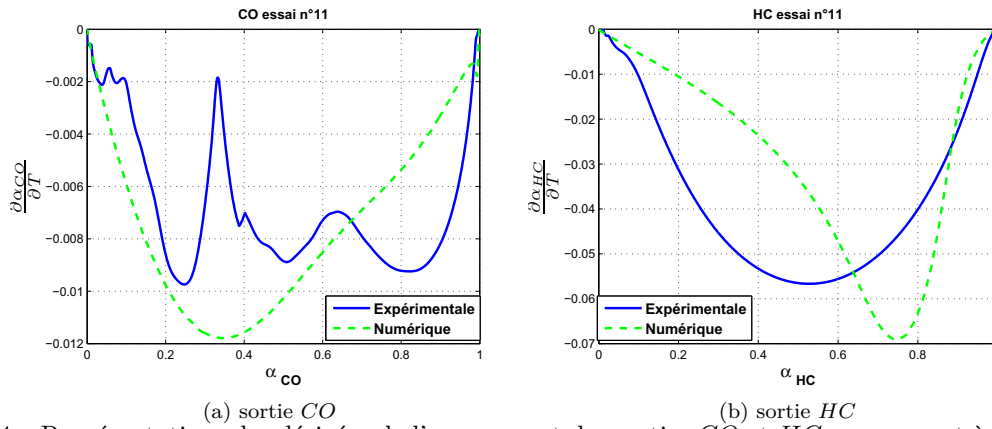


Figure 3.14 – Représentations des dérivées de l’avancement des sorties *CO* et *HC* par rapport à la températures en fonction de l’avancement pour un essai (essai 11)

Afin de bien visualiser la forme des résidus obtenus dans cette configuration, nous donnons sur les figures 3.15, les résidus entre résultats expérimentaux et numériques pour *CO* (figure 3.15a) et pour *HC* (figure 3.15b). Il est important de souligner que dans cette représentation, les résidus sont stationnaires, et de plus, deux valeurs sont conditionnées, à savoir les deux bords du domaine pour lesquels les valeurs des dérivées sont nulles (palier initial et final).

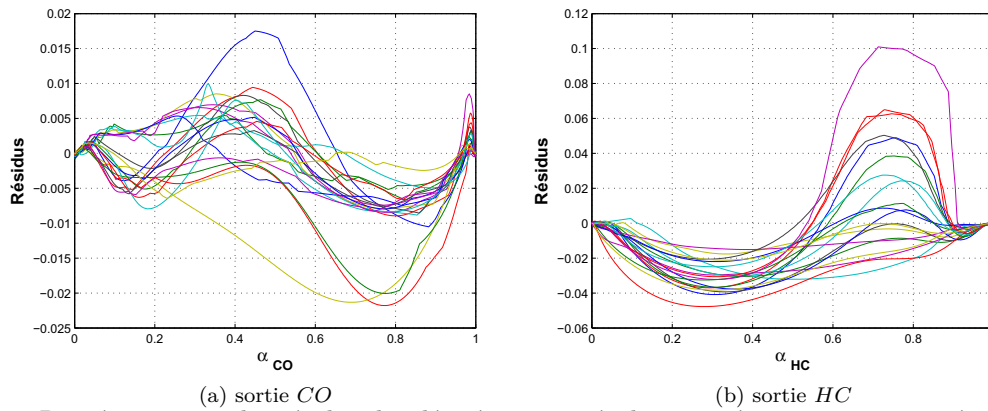


Figure 3.15 – Représentations des résidus des dérivées entre résultats expérimentaux et numériques pour les 20 essais

Nous proposons donc dans cette partie que les écarts observés peuvent se représenter par un processus gaussien, soit

$$\frac{\partial \alpha}{\partial T} = \frac{\partial g(\mathbf{x}, \beta)}{\partial T} + z_{\sigma^2, \theta}(\mathbf{x}) \tag{3.37}$$

avec α la sortie expérimentale convertie en avancement, $g(\mathbf{x}, \beta)$ le résultat du modèle numérique converti en avancement et où $z_{\sigma^2, \theta}$ est un processus gaussien. On se ramène donc à la théorie de krigeage à tendance non-linéaire que nous avons détaillée dans la section 3.1.

L’assimilation des résidus à un pont brownien est une caractérisation intéressante, puisqu’elle nous a guidés dans le choix du modèle posé et du noyau à utiliser. Le choix de ce noyau est important. Or, il est possible de montrer

que l'utilisation d'un processus gaussien avec un noyau de covariance exponentiel défini par l'équation (3.38), processus encore appelé processus d'Ornstein-Uhlenbeck, conditionné aux bords est équivalent à un pont brownien, dont on sait caractériser la moyenne et la covariance. L'équivalence entre processus d'Ornstein-Uhlenbeck conditionné aux bords et pont brownien est donnée en annexe B.

$$\forall h \in \mathbb{R}^p, R_{\theta}(h) = \exp \left\{ - \sum_{i=1}^p \theta_i |h_i| \right\}, \quad (3.38)$$

Les formules et inférences relatives à cette méthode sont introduites dans la section 3.1. Cette méthode est appliquée directement sur notre cas d'application, sans, au préalable, la tester sur le cas maîtrisé. Ce choix se justifie pour deux raisons. Tout d'abord la méthode a déjà été testée sur le cas maîtrisé (section 3.1.4.1). Ensuite, la limite de cette méthode sur notre cas provient principalement de la non stationnarité, qui dans cette nouvelle représentation est contournée.

Deux difficultés sont envisagées dans cette approche. La première concerne l'intégration des prédictions des dérivées, et la seconde provient de l'impossibilité de prédire le début de démarrage de la réaction (fin du palier initial), ce qui peut entraîner un écart avec le résultat expérimental.

3.3.2 Résultats sur le cas d'étude

Avant de présenter les résultats d'application, il est nécessaire d'introduire la construction du vecteur des réponses et de la matrice d'expériences, ainsi que de donner quelques détails sur les conditions d'applications pratiques mises en place.

Nous sommes, dans cette approche, sur une modélisation d'une sortie scalaire, l'avancement étant considéré comme un facteur d'entrée supplémentaire aux conditions expérimentales (pour rappel $[CO]_0$, $[HC]_0$ et $[O_2]_0$). Par conséquent, si l'on fait le parallèle avec la section 3.1, l'avancement prend la place de la température dans la matrice d'expériences, et le vecteur des sorties est composé des valeurs des dérivées en températures.

Afin de clarifier la construction de ces matrices, nous reprenons les notations (abusives) utilisées dans l'équation (3.24). La matrice \mathbf{S} est la matrice d'expériences utilisée et \mathbf{Y} le vecteur des réponses associé. On note n_a la taille de discrétisation de l'avancement.

$$\mathbf{S} = \begin{pmatrix} \alpha_1 & [HC]_0^1 & [CO]_0^1 & [O_2]_0^1 \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_{n_a} & [HC]_0^1 & [CO]_0^1 & [O_2]_0^1 \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_1 & [HC]_0^n & [CO]_0^n & [O_2]_0^n \\ \vdots & \vdots & \vdots & \vdots \\ \alpha_{n_a} & [HC]_0^n & [CO]_0^n & [O_2]_0^n \end{pmatrix}, \mathbf{Y} = \begin{pmatrix} \frac{\partial \alpha^{(1)}}{\partial T}(\alpha_1) \\ \vdots \\ \frac{\partial \alpha^{(1)}}{\partial T}(\alpha_{n_a}) \\ \vdots \\ \frac{\partial \alpha^{(n)}}{\partial T}(\alpha_1) \\ \vdots \\ \frac{\partial \alpha^{(n)}}{\partial T}(\alpha_{n_a}) \end{pmatrix} \quad (3.39)$$

où $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_{n_a}]$ est le vecteur des avancements discrétisés, uniformément entre 0 et 1 (le même pour tous les essais, de taille n_a), $[esp]_0^i$ correspond à la concentration initiale de l'espèce esp pour l'expérience i et $\frac{\partial \alpha^{(i)}}{\partial T}(\alpha_j)$ est la dérivée en température de la sortie considérée en avancement (soit HC ou CO) de l'expérience i à l'avancement α_j .

Les deux sorties sont traitées indépendamment. La matrice d'expériences ainsi définie avec les réponses associées permet l'estimation de $\boldsymbol{\beta}$, ainsi que σ^2 et $\boldsymbol{\theta}$, définissant le processus stochastique $z_{\sigma^2, \boldsymbol{\theta}}$. À l'aide de la formule

(3.4), il est alors possible de prédire l'évolution des dérivées en température de l'avancement de CO et HC en n'importe quel point du domaine.

La démarche mise en place afin d'éprouver notre méthode reprend uniquement la seconde étape mise en place dans les deux approches précédentes (KNL et KNLD), soit l'utilisation de 19 des 20 essais, et la comparaison du résultat de prédiction avec le résultat expérimental pour l'essai exclu du domaine d'apprentissage. Cette seconde étape est réalisée pour chaque essai et nous permet de mieux observer la qualité de prédiction.

D'un point de vue pratique, nous avons pris $n_a = 31$ points uniformément répartis selon l'avancement sur chacune des $n = 20$ expériences, soit 620 points, auxquels sont rajoutées les bornes de l'essai exclu (en 0 et 1) dont les valeurs des dérivées sont connues et égales à 0. Au final, nous prédisons la sortie pour l'essai non utilisé, et comparons le résultat obtenu au résultat expérimental. Nous réitérons cela pour chaque essai et chaque sortie. Les EQM de chacune des prédictions sont calculées après intégration de la dérivée et sont données dans la section suivante.

Afin d'illustrer les résultats de prédiction des dérivées, nous exhibons sur les figures 3.16a et 3.16b, les prédictions obtenues des dérivées respectivement pour CO et HC de l'essai 11, présenté précédemment (voir figure 3.14). Les courbes bleues correspondent aux résultats expérimentaux et les courbes rouges aux prédictions.

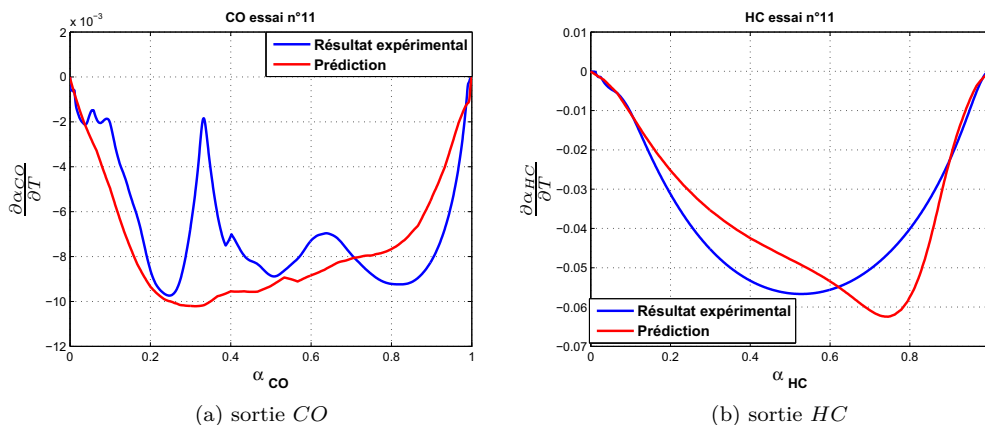


Figure 3.16 – Prédiction des dérivées en température selon l'avancement des sorties CO et HC de l'essai 11

Il est assez difficile de juger ces prédictions. En effet, dans cette approche, l'important n'est pas tant la valeur de la dérivée mais la forme de la courbe. En cela, nous pouvons dire que ces deux prédictions sont relativement bonnes, bien que certaines zones ne soient pas bien prédites. On pense par exemple au pic observé entre l'avancement de 0.3 et 0.4 pour la prédiction pour la sortie CO .

Pour cet essai 11, le comportement ponctuellement très irrégulier de la dérivée ne peut pas être reproduit par un processus de type processus gaussien. Une modélisation par ondelettes (Antoniadis [6]), par exemple, peut être une solution pour tenir compte des irrégularités locales de ce type.

Nous donnons ci-dessous les courbes prédites pour le meilleur cas après intégration pour CO et HC . Nous revenons sur la définition de meilleur cas dans la section suivante (section 3.3.3), mais il nous semble intéressant de présenter ces deux prédictions de dérivées ici. Pour le CO et le HC , il s'agit de l'essai 14, illustré respectivement par les figures 3.17a et 3.17b.

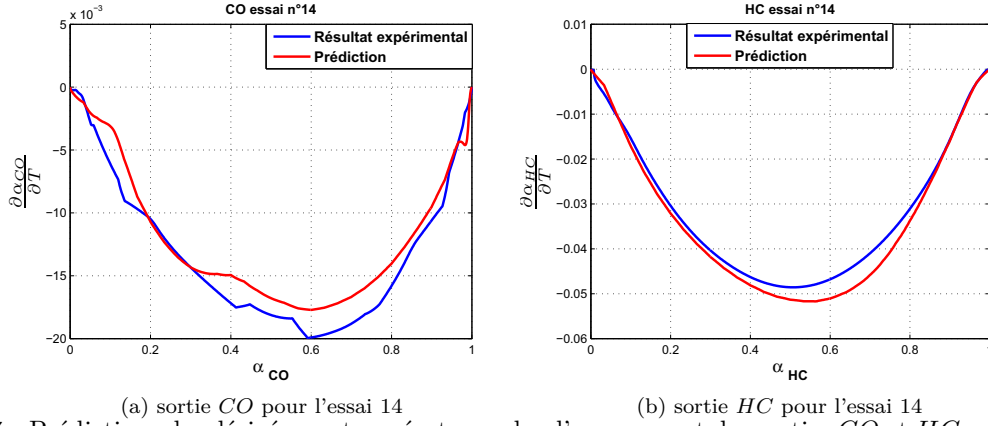


Figure 3.17 – Prédiction des dérivées en température selon l'avancement des sorties CO et HC pour les meilleurs cas après intégration présentée dans la section 3.3.3

Dans le cas du CO , la valeur de la dérivée n'est pas respectée, cependant la forme de la courbe de prédiction obtenue est similaire à celle que l'on observe expérimentalement. Ainsi, après intégration (cf. section 3.3.3) la forme de la courbe prédite est très proche du résultat expérimental. En ce qui concerne la sortie HC , aussi bien la valeur que la forme des dérivées sont très proches des dérivées expérimentales, ce qui nous conduit à de très bons résultats de prédiction du comportement du système étudié.

D'une manière générale, on en conclut que la méthode de modélisation par krigeage à tendance non linéaire, est une méthode performante pour la modélisation de la dérivée en température selon l'avancement.

3.3.3 Intégration des résultats

L'étape d'intégration des dérivées prédites se déroule en deux étapes. Tout d'abord, à l'aide d'un schéma d'Euler explicite, on reconstruit la fonction, puis dans un deuxième temps, on ajuste la courbe obtenue en prédisant la température de light-off (TLO) théorique de l'essai par krigeage universel avec effet de pépite.

Intégration des dérivées

L'approche mise en place nous permet de prédire pour chaque avancement la valeur de la dérivée selon la température, soit

$$\alpha \mapsto \frac{\partial \alpha}{\partial T}. \quad (3.40)$$

Ceci implique, que pour reconstruire la sortie à partir des dérivées, la seule solution possible est l'utilisation d'un schéma d'Euler explicite. En effet, nous souhaitons pour chaque avancement déterminer la température correspondante, sachant que pour un avancement de 0 (palier initial), la température est égale à T_0 .

On rappelle que les dérivées sont calculées par différences finies, soit

$$\frac{\partial \alpha(T_i)}{\partial T} = \frac{\alpha(T_i + h) - \alpha(T_i)}{h} \quad (3.41)$$

or $h = T_{i+1} - T_i$ en prenant un pas constant, donc

$$T_{i+1} = T_i + \frac{\alpha(T_{i+1}) - \alpha(T_i)}{\partial\alpha(T_i)/\partial T} \quad (3.42)$$

avec T_0 connu, et $\frac{\partial\alpha(T_0)}{\partial T} = 0$.

Cette méthodologie soulève néanmoins un problème, il s'agit du calcul de T_1 . Selon le schéma défini ci-dessus, le calcul de T_1 utilise comme dénominateur la valeur de la dérivée en T_0 , dont on sait qu'elle est nulle.

Cette température n'est pas forcément primordiale puisqu'elle correspond à un point sur le palier. Nous avons donc fixé arbitrairement sa valeur égale à $T_0 + 1$, et commençons l'intégration par le schéma d'Euler à partir de T_2 .

Après intégration, on obtient donc les sorties exposées sur les figures 3.18a et 3.18b pour l'essai 14 respectivement pour CO et HC .

Comme on pouvait s'y attendre, les formes des prédictions sont très proches de ce que l'on observe expérimentalement. Cependant, comme il ne nous est pas possible de définir le début de la réaction, soit la fin du palier initial au moment où la réaction commence à être effective, on a apparition d'un décalage de la courbe prédite par rapport aux résultats expérimentaux.

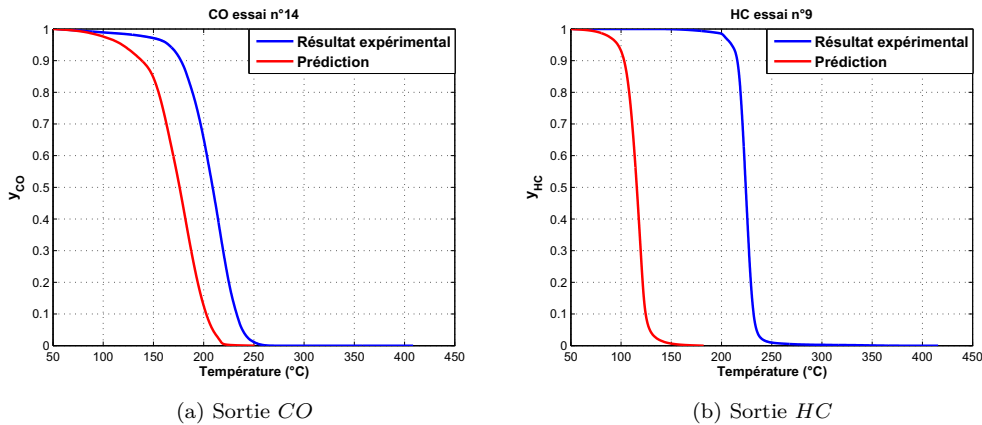


Figure 3.18 – Prédiction après intégration des dérivées pour l'essai 14 et apparition du décalage

Correction du décalage

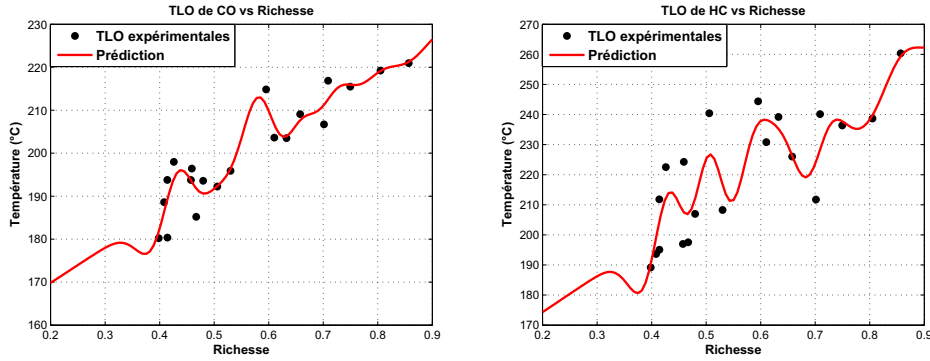
Il nous est impossible de définir la température de la fin du palier initial, ce qui nous a contraint à intégrer la dérivée à partir de T_0 , et a conduit à l'apparition d'un décalage entre prédiction et résultat expérimental. En revanche, nous avons été amenés à constater que la température de "light-off" (TLO) dépend linéairement de la richesse du mélange. Les figures 3.19a et 3.19b exposent le nuage de points des TLO observés sur nos essais respectivement pour CO et HC . On rappelle que la TLO est la température correspondant à un avancement de 0.5.

Nous proposons donc, pour la correction du décalage, d'estimer la TLO théorique à l'aide d'un modèle de krigeage universel avec pépite. La théorie est introduite dans le chapitre 2. Deux raisons ont motivé ce choix, premièrement car la relation linéaire est quelque peu abusive, et deuxièmement car cette méthode nous permet une estimation de la variance de prédiction que nous utilisons par la suite pour définir la variance de prédiction de nos sorties.

On pose donc le modèle,

$$T_{theo}^{TLO} = \gamma_0 + r \gamma_1 + z_{\sigma^2, \theta} + \epsilon \quad (3.43)$$

avec r la richesse du mélange, $z_{\sigma^2, \theta}$ un processus gaussien (différent de celui utilisé précédemment) régi par un processus gaussien, et ϵ un bruit blanc de variance choisie arbitrairement. La courbe rouge sur les figures 3.19a et 3.19b représente les prédictions en tout point du domaine par ce modèle.



(a) TLO de CO en fonction de la richesse

(b) TLO de HC en fonction de la richesse

Figure 3.19 – TLO de CO et HC en fonction de la richesse du mélange et prédiction par krigeage ordinaire avec effet pépite

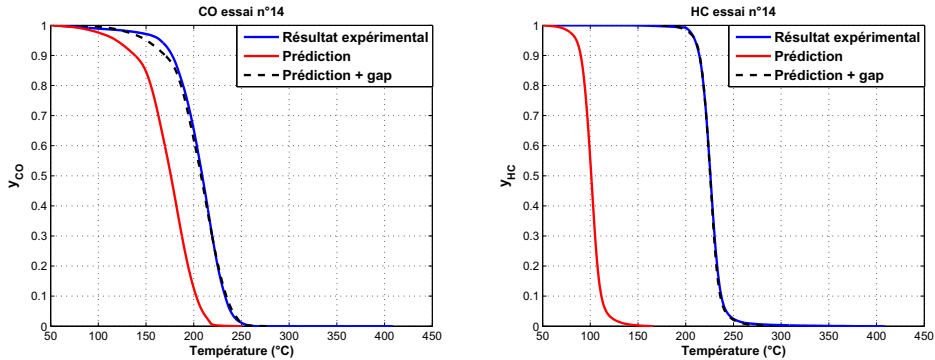
Soit ΔT^{TLO} , la correction appliquée sur notre prédiction,

$$\Delta T^{TLO} = T_{theo}^{TLO} - T_{pred}^{TLO}, \quad (3.44)$$

avec T_{theo}^{TLO} , la TLO théorique estimée par KU avec pépite, et T_{pred}^{TLO} la TLO de la prédiction. Notons \tilde{T} les températures respectives à la prédiction après intégration, \hat{T} correspondant à ces températures après ajustement. La procédure consiste à déplacer notre prédiction de ΔT^{TLO} , soit

$$\hat{T} = \tilde{T} + \Delta T^{TLO}. \quad (3.45)$$

Ainsi après correction du décalage, on obtient les courbes en pointillé noires sur les figures 3.20.



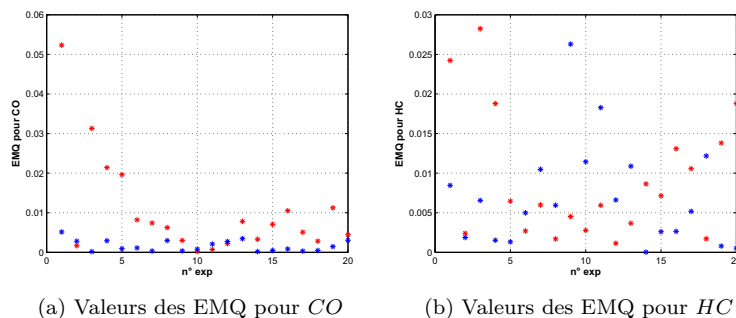
(a) Sortie CO

(b) sortie HC

Figure 3.20 – Prédiction après intégration des dérivées et correction du décalage pour l'essai 14

Calcul des EQM et comparaison avec les précédentes méthodes

Les valeurs des EMQ (équation 3.25) pour chacun des essais sont données sur les figures 3.21a et 3.21b. Il est important de souligner que pour le calcul de ces EQM, nous avons utilisé exactement la même méthodologie que pour les approches précédentes (KNL et KNLD) de manière à pouvoir comparer ces résultats entre eux. Par conséquent le nombre de points utilisé et leurs localisations sont les mêmes que précédemment, soit 1001 points répartis uniformément selon la température.



(a) Valeurs des EMQ pour *CO* (b) Valeurs des EMQ pour *HC*
 Figure 3.21 – Comparaison des valeurs des EMQ obtenues pour chaque jeu de données à disposition selon chacune des deux sorties après intégration de la dérivée prédite (points bleus) avec les EMQ calculées par KNLD (points rouges)

Ces valeurs sont très faibles. Pour *CO*, elles sont toutes inférieures à ce que nous avons obtenu précédemment, pour *HC* elles sont du même ordre, mais comme nous le montrons par la suite, nous n’observons pas de comportement atypique tel que des oscillations pouvant conduire à des concentrations négatives.

Pour les deux sorties, le meilleur cas au sens de ce critère est l’essai 14. Cet essai a été utilisé afin d’illustrer les deux étapes d’intégration de nos prédictions des dérivées. Les deux pires cas correspondent à l’essai 1 pour le *CO* et à l’essai 9 pour le *HC*.

Conclusions

Comme on peut le constater, les prédictions obtenues au final pour l’essai 14 (figures 3.20) sont très bonnes puisque quasiment confondues avec les résultats expérimentaux. D’une manière générale, les résultats obtenus sont très satisfaisants. Pour bien visualiser la pertinence de notre méthode, on expose les pires cas pour *CO* et *HC* sur les figures respectivement 3.22a et 3.22b.

Les résultats sont également satisfaisants. On notera que l’erreur pour *HC* provient principalement de la mauvaise correction du décalage.

En conclusion, on peut dire que l’on a développé un outil performant pour la prédiction. Cette approche tient compte du modèle numérique au travers de sa dérivée, et également de considération théorique telle que la relation liant richesse et TLO. Afin de parfaire cette méthode, il est nécessaire de quantifier la valeur des prédictions, au travers de l’estimation des variances de prédictions, puisqu’on le rappelle, le critère de planification d’expériences envisagé est basé sur cette information.

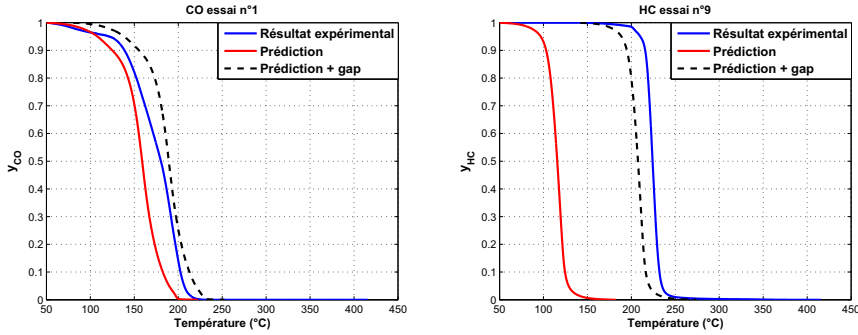
(a) Sortie CO pour l'essai 1(b) Sortie HC pour l'essai 9

Figure 3.22 – Prédiction après intégration des dérivées et correction du décalage pour les pires cas au sens de l'EQM

3.3.4 Calcul de la variance de prédiction

Les limites de notre méthode apparaissent lors du calcul de la variance de prédiction. Deux problèmes sont identifiés. Le premier provient des limites de l'intervalle de confiance pour la prédiction des dérivées, qui sur certaines zones sont positives. Le second problème est la surestimation des intervalles de confiance finaux (composés de l'intervalle des prédictions plus l'intervalle lié à l'ajustement de la courbe après intégration).

Pour illustrer ce calcul, nous utilisons l'essai 14 (figures 3.20). Tout d'abord, nous donnons l'intervalle de prédiction de la dérivée, puis celui pour l'ajustement de la courbe. Cette double estimation nous permet au final de définir l'intervalle de prédiction total.

Estimation de la variance de prédiction des dérivées

La formule de la variance de prédiction dans le cas du krigeage à tendance non linéaire (voir section 3.1 pour plus de détails) est donnée par

$$\varphi(\mathbf{x}) = \sigma^2 \left(1 + \mathbf{u}^T (f(\mathbf{S}, \beta)^T \mathbf{R}^{-1} f(\mathbf{S}, \beta))^{-1} \mathbf{u} - r(\mathbf{x})^T \mathbf{R}^{-1} r(\mathbf{x}) \right), \quad (3.46)$$

où $\mathbf{u} = f(\mathbf{S}, \beta)^T \mathbf{R}^{-1} r(\mathbf{x}) - f(\mathbf{x}, \beta)$ et σ^2 est obtenu par (3.15).

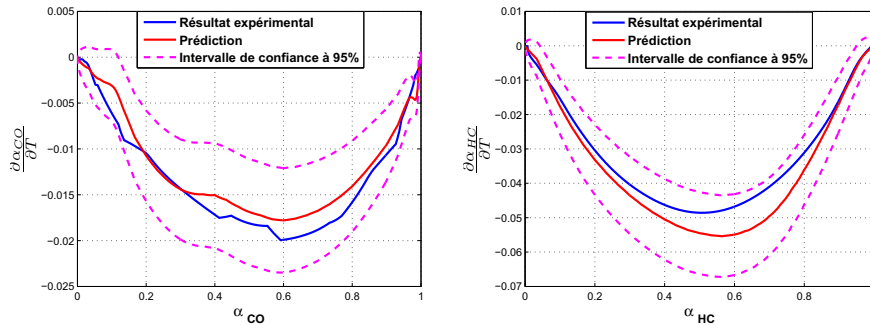
(a) Sortie CO (b) Sortie HC

Figure 3.23 – Intervalles de prédiction des dérivées par krigeage à tendance non linéaire (essai 14)

Les figures 3.23 présentent pour les dérivées prédites les limites de l'intervalle de confiance pour les deux sorties de l'essai 14, en courbes pointillées magenta. Les courbes bleues et rouges correspondent toujours respectivement aux résultats expérimentaux et prédits.

On observe pour la limite supérieure de l'intervalle de confiance des zones positives. Ce phénomène est très problématique puisque le schéma d'intégration utilisé ne permet pas de déterminer les températures relatives à chaque essai. En effet, le schéma conduit à définir des températures de plus en plus faibles pour des avancements de plus en plus petits, ce qui est physiquement impossible. On perd la surjectivité de la fonction.

On constate également la surestimation de l'intervalle. Le cas présenté correspond au meilleur cas de prédiction obtenu. Cependant, on voit que l'intervalle autour de cette prédiction est très grand. Ce phénomène est propagé et amplifié après intégration.

Nous avançons l'hypothèse que ces bandes de confiance si grandes proviennent du "plug-in" du vecteur β dans la formule 3.46. Il ne nous est pas possible, comme dans le cas du krigeage universel, de déterminer une formule indépendante de ce vecteur. Ne connaissant pas les "vraies" valeurs de ces paramètres cinétiques, nous sommes contraints à utiliser l'estimation obtenue dans la formule de variance de prédiction, ce qui peut être à l'origine de ces intervalles de confiance si larges.

Estimation de la variance de prédiction de l'ajustement du décalage

Notre but ici est d'exposer la seconde limite de notre méthode. Même si l'intégration de l'intervalle de confiance des prédictions était possible pour les deux bornes, la seconde étape (ajustement du décalage) entraîne une seconde surestimation de l'intervalle de confiance final.

Pour illustrer nos propos, nous intégrons la borne inférieure, toujours négative. L'intégration de cette courbe définit la borne de l'intervalle de prédiction finale. Les figures 3.24 illustrent, après intégration mais avant ajustement, la limite inférieure de l'intervalle de confiance.

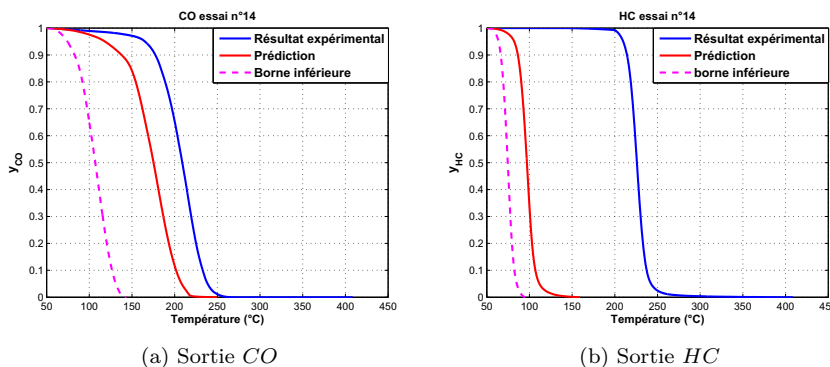


Figure 3.24 – Limite inférieure de l'intervalle de prédiction des dérivées intégrées par le schéma d'Euler (cf. section 3.3.3)

La seconde étape consiste à ajuster la prédiction après intégration par krigeage universel (KU) avec pépite. Cette méthode fournit également la variance de prédiction de T_{theo}^{TLO} . L'obtention de l'intervalle de confiance de la prédiction est effectuée en déplaçant notre borne inférieure, sur le même principe que pour le calcul de \hat{T} , mais par rapport à la variance de prédiction de la TLO théorique.

La variance de prédiction du KU avec pépite est donnée par

$$\varphi(x_0) = \sigma_z^2 \left(1 + \gamma^T \left[R + \frac{\sigma_\epsilon^2}{\sigma_z^2} I_n \right] \gamma - 2\gamma^T r_0 \right) + \sigma_\epsilon^2 \quad (3.47)$$

où I_n est la matrice identité de dimension n , σ_z la variance du processus gaussien, σ_ϵ celle de l'erreur et γ les coefficients du modèle. Pour plus de détails, nous renvoyons le lecteur au chapitre 2. À partir de φ , on calcule la borne inférieure de notre estimation de T_{theo}^{TLO} , notée T_{bf}^{TLO} .

Soient \tilde{T}_{bf} les températures respectives à la borne inférieure de l'intervalle de confiance après intégration, et \hat{T}_{bf} celles après ajustement, on a

$$\hat{T}_{bf} = \tilde{T}_{bf} + (T_{bf}^{TLO} - T_{pred}^{TLO}). \quad (3.48)$$

Les figures 3.25 exposent les limites inférieures de l'intervalle de confiance final (courbes en pointillées magenta).

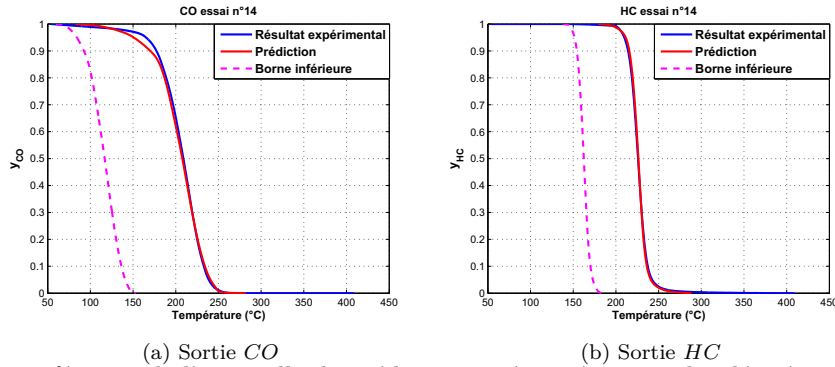


Figure 3.25 – Limite inférieure de l'intervalle de prédiction après intégration des dérivées par le schéma d'Euler et correction du décalage par KU avec pépite

On se rend compte sur cet exemple, pourtant bien prédit, que l'intervalle de prédiction obtenu est très grand. Par conséquent, l'utilisation de ce critère pour la planification est inenvisageable, la surestimation de la variance de prédiction rendant inopérant un tel critère.

3.3.5 Conclusions et limites

La modélisation directe de la dérivée en fonction de l'avancement, motivée par le contournement du problème de non stationnarité, est une méthode pertinente de prédiction pour représenter le comportement du système du piège à NOx. Cette méthode repose d'une part sur la prédiction des dérivées en températures en fonction de l'avancement et d'autre part sur l'intégration des prédictions.

Bien que les prédictions des dérivées ne soient pas pour tous les essais, exactement égales aux dérivées expérimentales, les formes de ces courbes sont relativement bien respectées. Ceci conduit après intégration à de bons résultats. Cette méthode d'intégration en deux étapes (Schéma d'Euler et correction du décalage) est judicieuse, et reste cohérente avec la connaissance des experts.

Les limites de cette troisième démarche, proviennent du calcul de la variance de prédiction que nous proposons d'utiliser comme critère de plan d'expériences. Tout d'abord, la détermination d'intervalles de confiance des dérivées positifs, empêche l'intégration par notre schéma. De plus, les différentes étapes, conduisent à la surestimation de l'intervalle de prédiction, rendant inutilisable cette information comme critère de plan d'expériences.

On conclut donc cette section en insistant sur le fait que cette méthode bien qu'inadaptée pour les objectifs de notre thèse, reste néanmoins une méthode pertinente pour les cas rencontrés en cinétique pour lesquels seule la prédiction du phénomène est importante.

3.4 Conclusions

Trois points ont été abordés dans ce chapitre. Premièrement le krigeage étendu au cas d'une tendance non-linéaire, deuxièmement, une amélioration possible de cette première approche, par le conditionnement par la dérivée et troisièmement la modélisation directe de la dérivée par rapport à l'avancement.

Le caractère fonctionnel est omis, puisque les sorties étudiées sont considérées non pas comme des courbes mais comme une succession de points. Ces trois méthodes, appelées donc approches non fonctionnelles, sont basées sur la théorie du krigeage et sont l'aboutissement de diverses améliorations successives.

1. Tout d'abord, nous avons étendu la théorie de krigeage au cas d'une tendance non linéaire qui correspond au modèle numérique. Cette méthode souffre d'un certain nombre de lacunes, dont les principales raisons sont la non stationnarité du problème et sa forte non linéarité.
2. Nous avons alors apporté une amélioration de cette première méthode, en conditionnant le krigeage par les dérivées. En effet, les premiers résultats montrent des phénomènes non désirables sur les paliers, or on sait que les dérivées sont nulles sur ces zones. Néanmoins, là encore les résultats obtenus dans certains cas sont non concluants, ce qui nous a poussés à changer de système de représentation pour contourner la non stationnarité. En effet, si l'on s'intéresse au cas maîtrisé, les résultats que nous avons exposés montrent un très bon comportement de la méthode. De même, pour les résultats de *CO* qui sont certes dans certains cas extrêmes, très mauvais du fait du manque d'information mais restent dans l'ensemble plutôt intéressants, la non stationnarité et la non linéarité étant moins forte. Ce comportement nous pousse à penser que cette méthode bien qu'inadaptée pour notre cas d'étude, reste intéressante pour d'autres cas d'application sous certaines conditions.
3. Enfin, la troisième méthode reprend la théorie de krigeage à tendance non linéaire que nous avons développée, mais cette fois-ci l'application est effectuée pour modéliser la dérivée des sorties selon la température en fonction de l'avancement. Les résultats de prédiction de la dérivée par cette approche sont très intéressants. Cependant, lors de l'application sur notre cas d'étude, la difficulté réside dans la reconstruction de la courbe cinétique à partir des dérivées. Le seul choix possible est l'utilisation d'un schéma d'Euler explicite. Cette intégration a deux conséquences principales. Premièrement, on observe l'apparition d'un écart entre la prédiction et le résultat expérimental, deuxièmement, une petite erreur au début de la prédiction de la dérivée peut entraîner de grandes erreurs une fois l'intégration effectuée. L'ajustement est opéré tenant compte de la possibilité de prédire la TLO en fonction de la richesse. Les limites de cette troisième démarche proviennent de la surestimation des intervalles de confiance.

En conclusion sur ces méthodes, nous pouvons dire qu'elles ne sont certes pas adaptées pour notre cas d'étude, mais qu'elles pourront, sous certaines conditions, convenir à d'autres cas d'application. Nous avons dans ce chapitre essayé de présenter pour chacune les limites d'applications. Principalement, il s'agit de la non stationnarité des résidus pour les deux premières méthodes et de la surestimation de la variance de prédiction pour la troisième. Concernant cette dernière, les résultats probants font de cette méthode une alternative efficace si l'on s'intéresse uniquement à la prédiction d'un système, et peut facilement s'adapter aux nombreux cas rencontrés en pratique par les expérimentateurs.

Au final, ces différentes méthodes introduites mettent en lumière l'importance de la prise en compte de la dimension fonctionnelle de nos données afin de prévenir des comportements atypiques. De plus, ces méthodes plus adaptées à notre cas, offrent la possibilité d'estimer des intervalles de confiance plus pertinents.

Approche fonctionnelle

4.1	Modèles à coefficients variables	113
4.1.1	Introduction des modèles à coefficients variables	114
4.1.1.1	Motivations	114
4.1.1.2	Choix des variables et étude de la dépendance linéaire	115
4.1.2	Méthode 1 : Estimation des fonctions coefficients par P-splines	119
4.1.2.1	Première étape : Estimation point à point des fonctions coefficients par moindres carrés	120
4.1.2.2	Seconde étape : Lissage des fonctions coefficients par P-spline	122
4.1.3	Méthode 2 : Estimation des fonctions coefficients par polynômes locaux	125
4.1.3.1	Présentation de la méthode	125
4.1.3.2	Applications et résultats	129
4.1.4	Conclusions	130
4.2	Estimation des paramètres cinétiques	131
4.2.1	Modifications du modèle numérique et estimations des paramètres cinétiques	132
4.2.2	Modélisation des paramètres cinétiques en fonction des entrées	135
4.2.2.1	Estimation des coefficients d'une relation linéaire entre prédicteurs et paramètres cinétiques	136
4.2.2.2	Applications et résultats de la KRR	137
4.2.2.3	Récapitulatif des résultats des différentes approches et conclusion	139
4.2.3	Calcul des régions de confiance	140
4.2.3.1	Bootstrap fonctionnel	141
4.2.3.2	Estimation de la loi des paramètres de la modélisation par KRR	143
4.2.3.3	Calcul de la région de confiance	144
4.2.4	Conclusions	145
4.3	Conclusions et perspectives	145

Approche fonctionnelle

Dans ce chapitre, nous détaillons deux méthodes de modélisation tenant compte du caractère fonctionnel de notre cas d'étude : l'utilisation de modèles à coefficients variables et une modélisation des paramètres cinétiques par kernel ridge regression avec calcul des régions de confiance par bootstrap fonctionnel.

Dans de nombreux domaines, les quantités mesurées sont des éléments de plus en plus complexes : courbes, images... donc mathématiquement des fonctions. C'est la raison pour laquelle on parle alors de données fonctionnelles (Ramsay and Silverman [131]). Le développement de ce thème de recherche est motivé par l'abondance de données de plus en plus fines du fait de l'automatisation et l'informatisation des procédures de mesures : c'est par exemple le cas en météorologie, en médecine, en imagerie satellitaire et dans de nombreux autres domaines d'études. La modélisation et le traitement statistique de ces données ouvrent un champ très vaste de recherche visant des aspects théoriques et appliqués.

La première partie de ce chapitre est consacrée aux modèles à coefficients variables (Hastie et Tibshirani [81], Ferraty et Vieu [60]), nous y montrons que si d'un point de vue général il est possible d'utiliser cette méthode, l'omission de l'expertise métier entraîne dans notre cas des comportements qui n'ont pas de sens physique. Deux méthodes d'estimation des fonctions coefficients sont présentées.

Dans la seconde partie, nous décrivons à partir des modifications apportées au modèle cinétique (voir chapitre 1 section 1.1.5), les approches successives qui nous ont guidés vers l'utilisation des kernel ridge regression (Hastie *et al*, [82]) pour la modélisation des paramètres cinétiques, en fonction des conditions initiales des polluants. Une extension de la technique de bootstrap (Efron, [43]) aux données fonctionnelles, nous permet dans un second temps, de déterminer la région de confiance pour les prédictions obtenues.

4.1 Modèles à coefficients variables

Les évolutions technologiques nous offrent la possibilité de collecter de plus en plus de données, avec une grande précision mais également de traiter ces données. Par conséquent, l'analyse de données fonctionnelles devient un outil incontournable dans l'étude de ces cas, afin de résoudre aux mieux les problèmes rencontrés. En général, il s'agit d'étudier des courbes en sortie d'un modèle en fonction d'une variable longitudinale généralement le temps, l'espace ou la température. De nombreux exemples sont présents dans la littérature : Ferraty et Vieu [60], Mizuta et Kato [114], Hall et Horowitz [76], et également dans l'ouvrage de Ramsay et Silvermann [131], ainsi que des modèles variés pour les traiter.

Dans le domaine de la cinétique, les études portant sur des données fonctionnelles sont devenues de plus en plus courantes (jiang et Martin [89], Shi et Wang [145]), notamment grâce aux moyens d'acquisition et de traitement de données actuels. C'est le cas pour notre application dans laquelle nous étudions des courbes d'émissions de polluants en sortie d'un système de post-traitement en fonction de la température. De plus, les niveaux de performance demandés en termes de précision entraînent le besoin de méthodes spécifiques adaptées à ce type d'étude. De nombreux travaux sont disponibles dans la littérature (Fan et Gijbels [52], Li *et al*, [95]). Deux

méthodes sont présentées et appliquées dans les parties suivantes, toutes deux basées sur un modèle fonctionnel linéaire, ou modèle à coefficients variables (voir chapitre 2) de la forme suivante :

$$Y(T) = \sum_{j=1}^p a_j(T)X_j + \epsilon, \text{ avec } \mathbb{E}[\epsilon] = 0 \text{ et } \text{Var}[\epsilon] = \sigma^2, \quad (4.1)$$

mais dont l'estimation des fonctions coefficients a_j diffère, et inspirés respectivement des travaux de Fang *et al* [57] et Fan et Zhang [56]. Ce modèle peut se voir comme un modèle de régression. Si $t = T$ on tombe sur la définition du modèle classique. Ce type de modèle peut également être vu comme un cas particulier des modèles additifs (Hastie et Tibshirani [80]), dans lequel on suppose que la sortie peut s'expliquer par une somme de fonctions chacune dépendant d'une des variables (voir equation 4.5). Dans ce cas là, on applique une relation linéaire.

Dans la première méthode proposée, il s'agit d'une estimation point à point réalisée sur un échantillonnage de la variable longitudinale, les fonctions coefficients étant lissées dans une seconde étape. La seconde méthode estime les fonctions coefficients par des polynômes locaux cubiques (voir Fan et Zang [56]).

Pour chacune de ces deux méthodes, il est nécessaire, au préalable, de définir les variables explicatives du modèle et d'étudier leur structure de dépendance par rapport aux sorties. Cette réflexion constitue la première partie. La seconde partie traite de la première approche avec estimation point à point des fonctions coefficients et lissage par P-splines (Ruppert et Carroll [135]) et la troisième partie avec estimation par polynômes locaux. De nombreux résultats sont présentés et comparés. Enfin, nous concluons sur les méthodes et résultats obtenus et donnons des pistes de recherche sur les améliorations envisageables.

4.1.1 Introduction des modèles à coefficients variables

Dans cette section, nous commençons par motiver l'étude des VCM, avant de définir le modèle retenu et de réaliser une étude sur la faisabilité et la pertinence de ces approches. Les différents outils ont été introduits dans la section 2.2, nous nous attardons en particulier sur une étude de la dépendance entre les variables explicatives choisies et les sorties considérées.

4.1.1.1 Motivations

Lors du développement des méthodes afin de résoudre le problème de modélisation du phénomène étudié, le piège à NOx, nous nous sommes rendu compte que son caractère fonctionnel rendait caduque les approches non fonctionnelles testées dans le chapitre précédent.

L'idée de l'utilisation d'un modèle fonctionnel linéaire est issue du phénomène observé dans la section 3.3.4 où nous avons montré qu'il est possible de prédire efficacement la température de "light-off" (TLO) en fonction de la richesse du mélange considéré par l'intermédiaire d'un modèle linéaire. Pour rappel, la température de light-off est la température à laquelle 50% des polluants considérés sont convertis, soit la température correspondante à un avancement de 0.5. Cette constatation innovante d'un point de vue cinétique, nous permet d'entrevoir une première approche, basée sur le principe que, s'il est possible de prédire précisément la TLO c'est-à-dire la température pour un avancement de 0.5, il est également envisageable de procéder de même pour toutes les températures.

Le modèle postulé dans ce cas est le suivant,

$$T(y) = \sum_{i=1}^p a_i(y)X_i + \epsilon, \text{ avec } \mathbb{E}[\epsilon] = 0 \text{ et } \text{Var}[\epsilon] = \sigma^2, \quad (4.2)$$

où T est la température, y la concentration normalisée des polluants CO ou HC en sortie du système et les X_i sont les variables explicatives du modèle. Les a_i sont les fonctions coefficients, qu'il nous faut estimer.

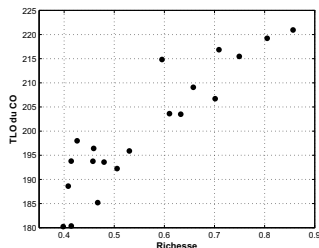


Figure 4.1 – Nuage de points de la température de light-off pour CO en fonction de la richesse

La figure 4.1 illustre la relation linéaire que l'on observe entre la TLO du CO et la richesse du mélange. Ce constat nous a permis de corriger l'écart apparu lors de l'intégration de la modélisation directe de la dérivée dans la section 3.3.4.

Cependant, la prise en compte de la richesse comme seule variable pour prédire la température correspondante à la sortie, n'est pas suffisante, pour tous les avancements. Il semble naturel, d'utiliser, en plus de la richesse, d'autres variables explicatives, à savoir les espèces présentes dans le mélange à l'entrée du catalyseur, que nous avons définies comme étant les espèces prédominantes dans le fonctionnement du piège à NO_x , soit la composition de HC , CO et O_2 .

Néanmoins, cette approche contredit l'intuition, car dans le modèle 4.2 on prédit la température en fonction de la concentration. Or, un cinéticien raisonne de manière inverse. Par conséquent, le modèle 4.1 est choisi pour tester les méthodes d'estimation des fonctions coefficients introduites par la suite.

4.1.1.2 Choix des variables et étude de la dépendance linéaire

Dans cette section nous présentons le choix des variables explicatives et nous étudions la dépendance entre ces variables et les sorties étudiées.

À l'aide des vingt expériences à disposition, dans un premier temps, nous donnons une représentation des nuages de points des sorties, fonction des variables explicatives, en certaines valeurs de température représentatives, que nous examinons graphiquement. Comme les points étudiés figurent dans un espace à quatre dimensions, dans un deuxième temps, nous appliquons les modèles additifs sur notre problème, afin de mieux pouvoir interpréter la dépendance existante et essayer de dégager la structure la plus adaptée.

Choix des variables

Nous avons quatre variables explicatives possibles pour notre cas : la richesse, $[HC]_0$, $[CO]_0$ et $[O_2]_0$ correspondant respectivement à la concentration initiale de HC , de CO et de O_2 .

On rappelle tout d'abord que l'écriture classique des modèles à coefficients variables, suppose la forme suivante :

$$Y(T) = \sum_{j=1}^p a_j(T)X_j + \epsilon, \text{ avec } \mathbb{E}[\epsilon] = 0 \text{ et } \text{Var}[\epsilon] = \sigma^2, \quad (4.3)$$

où T est la variable longitudinale, X_j la $j^{\text{ème}}$ variable explicative et Y la sortie considérée.

Concernant, les variables explicatives, afin d'avoir un maximum d'information, il semble logique de prendre les quatre variables à disposition. Néanmoins, l'information apportée par $[O_2]_0$ est très faible, puisque déjà en partie contenue dans la richesse, pour des raisons cinétiques. En effet, la richesse caractérise, de par son expression, le taux d'oxygène du mélange. De plus, la richesse est défini par,

$$\text{richesse} = \frac{2[CO_2]_0 Kb + [H_2O]_0 + (2 + 1/3)[CO]_0 Kb + 9[HC]_0}{2[CO_2]_0 Kb + [H_2O]_0 + 2[O_2]_0 Kb + [CO]_0 Kb + [NO]_0}, \quad (4.4)$$

où Kb correspond à une constante donnée. Étant donné également, que dans notre plan d'expériences les concentrations initiales des espèces CO_2 , H_2O et NO sont fixes, la richesse dépend seulement des trois espèces HC , CO et O_2 . On constate donc que l'une des deux variables, la richesse ou $[O_2]_0$ doit être retirée du modèle à postuler.

Compte tenu du fait que l'observation de la relation linéaire observée lie la richesse et la TLO, nous avons fait le choix de garder la richesse comme variable explicative. De plus, le fait de n'avoir que vingt expériences pour calibrer nos fonctions coefficients, nous incite à limiter le nombre de variables. Pour ces raisons, les variables explicatives retenues pour l'approche fonctionnelle sont la richesse, et les concentrations initiales de CO et HC .

La question désormais est de savoir si l'hypothèse d'un modèle de type VCM est envisageable, c'est à dire est-ce que la sortie dépend linéairement des variables d'entrée pour une température fixée. Nous avons voulu regarder si une telle relation est vérifiée sur les données disponibles. Ainsi, nous donnons dans un premier temps une représentation des nuages de points température après température puis dans un second temps une modélisation par modèles additifs température par température.

Représentation des nuages de points

Afin de représenter le nuage de points, nous récupérons sur les courbes expérimentales la valeur de sortie pour six températures normalisées représentatives (0, 0.25, 0.4, 0.5, 0.6, et 1). Les températures sont normalisées entre 0 et 1 (0 plus grande valeur minimale commune de degré d'avancement nul et 1 plus petite valeur maximale commune). Ainsi, selon chaque expérience, les températures choisies sont les mêmes.

Les ensembles de figures 4.2, 4.3 et 4.4 présentent les nuages de points obtenus pour la sortie CO aux différentes températures en fonction respectivement de la richesse, de $[CO]_0$ et de $[HC]_0$. Ces variables sont normalisées selon les bornes fournies par les expérimentateurs.

Si l'on s'intéresse aux figures 4.2 relatives à la variable richesse, on retrouve une relation entre la sortie CO et la richesse. Le nuage de points se déplace avec les températures étudiées et sa dispersion reste assez limitée. En revanche, sur les figures 4.3 et 4.4, relatives à $[CO]_0$ et $[HC]_0$, cette dispersion est d'autant plus importante que l'on est proche de la température de 0.4, qui se situe dans l'échelle normalisée sur la partie décroissante des courbes.

Pour ce qui est de la sortie HC , les figures 4.5, 4.6 et 4.7 exposent les nuages de points obtenus respectivement pour les variables richesse, $[CO]_0$ et $[HC]_0$. Ces nuages de points ne nous permettent pas de pouvoir envisager une structure de dépendance linéaire entre les variables et la sortie. En effet, même dans le cas favorable où l'on étudie la sortie HC en fonction de la richesse, nous n'observons pas de relation visible.

La structure de dépendance entre variables et sorties que nous voulons tester, est une structure linéaire. Cependant, bien que pour la sortie CO cette hypothèse puisse être envisagée, il n'en est pas de même pour HC , au vu des nuages de points présentés.

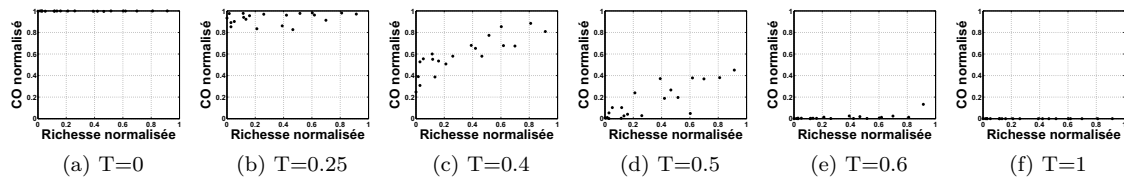


Figure 4.2 – Nuages de points de la sortie CO normalisée selon la richesse pour différentes températures

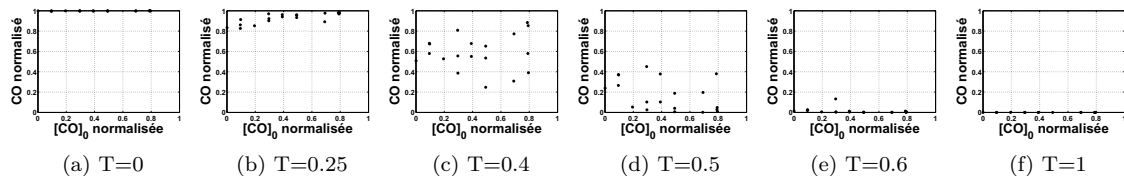


Figure 4.3 – Nuages de points de la sortie CO normalisée selon $[CO]_0$ pour différentes températures

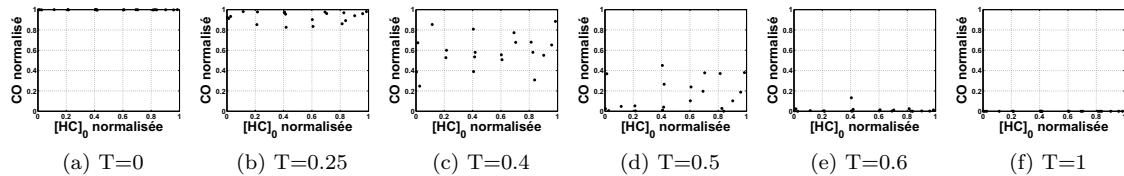


Figure 4.4 – Nuages de points de la sortie CO normalisée selon $[HC]_0$ pour différentes températures

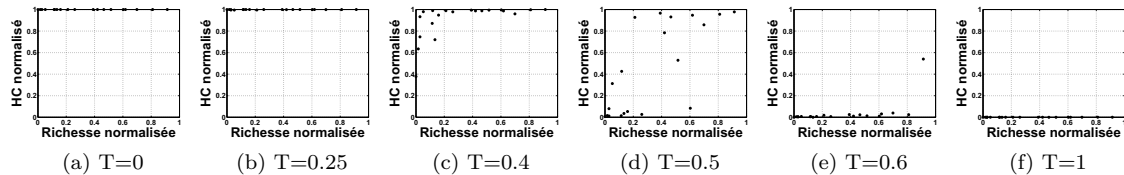


Figure 4.5 – Nuages de points de la sortie HC normalisée selon la richesse pour différentes températures

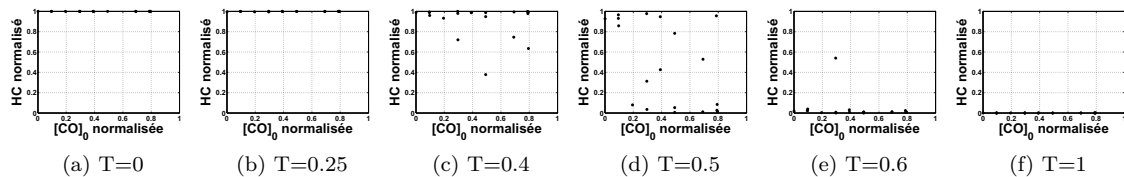


Figure 4.6 – Nuages de points de la sortie HC normalisée selon $[CO]_0$ pour différentes températures

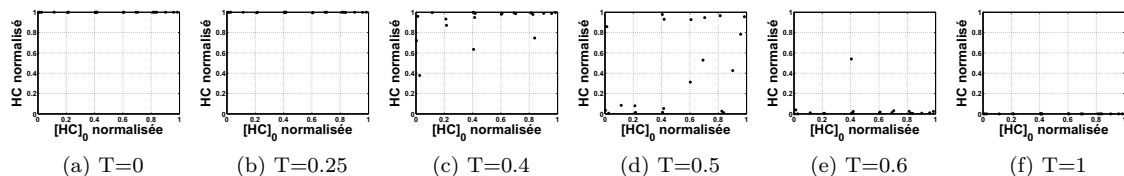


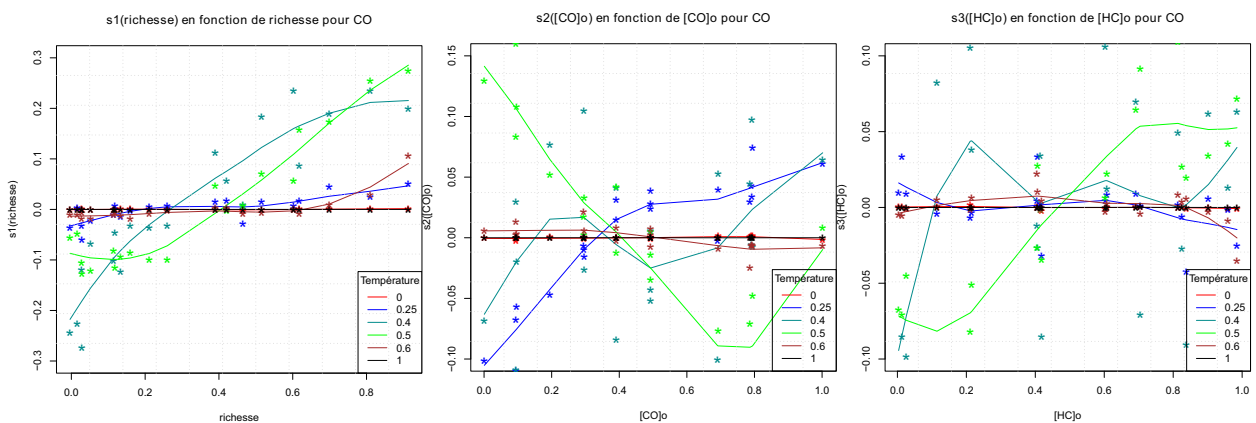
Figure 4.7 – Nuages de points de la sortie HC normalisée selon $[HC]_0$ pour différentes températures

Modèles additifs

Pour examiner de façon plus large la structure de dépendance entre la température à un avancement fixé et les variables choisies pour notre problème, nous avons recours aux modèles additifs. Les figures précédentes sont des projections du nuage sur des plans et l'on souhaite ici une meilleure représentation afin de vérifier s'il est possible de postuler une relation linéaire entre les variables explicatives et les sorties, notamment pour HC . La méthode a été introduite dans le chapitre 2. On rappelle brièvement, que cette méthode se propose de décomposer la sortie étudiée en somme de fonctions, chacune étant dépendante d'une des variables explicatives soit

$$y^T = s_0^T + s_1^T(x_1) + s_2^T(x_2) + s_3^T(x_3), \quad (4.5)$$

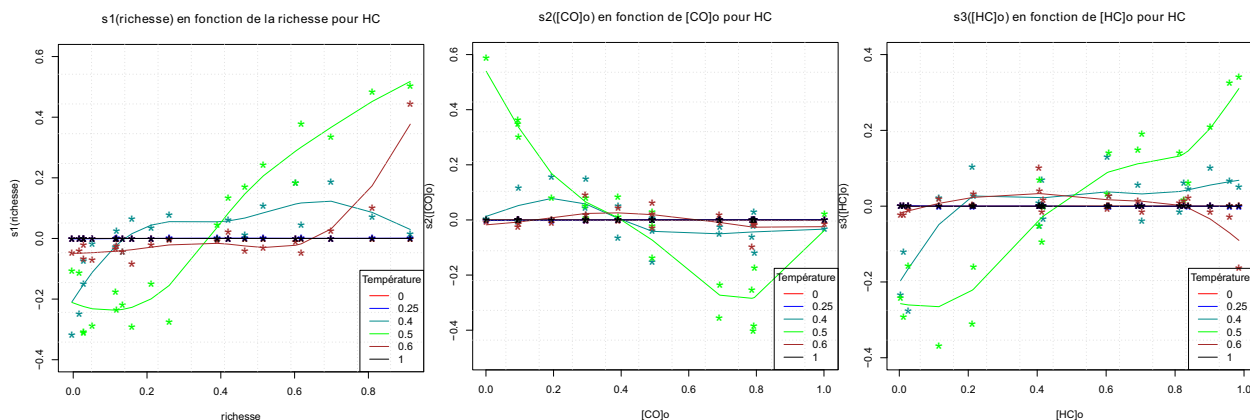
où s_0^T est une constante. On cherche à vérifier si pour T fixé, les fonctions $s_i^T(x_i)$ sont linéaires. Pour cela, nous avons eu recours au package 'GAM' (pour modèles additifs généralisés) disponible sous R (Hastie, [78]). Les figures 4.8 (resp. 4.9) exposent les $s_i, i = 1, 2, 3$ pour la sortie CO (resp. HC), relativement à la richesse, $[CO]_0$ et $[HC]_0$. Dans cette représentation, les six courbes obtenues pour les six températures d'échantillonnage sont présentées sur un même graphique. Les étoiles représentent les observations après retrait de la partie dépendant des autres variables.



(a) s_1 : GAM relatif à la richesse (b) s_2 : GAM relatif à la $[CO]_0$ (c) s_3 : GAM relatif à la $[HC]_0$
Figure 4.8 – Représentation des fonctions s_i du modèle GAM 4.5 pour CO à différentes températures

Intéressons nous tout d'abord, aux résultats obtenus pour la sortie CO (figures 4.8). On peut voir sur la figure 4.8a, comme sur les figures 4.2, que la procédure estime une relation quasi-linéaire entre la variable richesse et la sortie CO . En revanche, les estimations des fonctions s_i selon $[CO]_0$ et $[HC]_0$ sont beaucoup plus difficiles à interpréter. La figure 4.8b laisse tout de même la possibilité d'appliquer une structure linéaire, mais l'on note toutefois, qu'à la température de 0.5, s_2 a un comportement atypique. En effet, il s'agit d'une fonction affine par morceaux, avec un changement aux alentours de $[CO]_0 = 0.7$. Pour ce qui est de la relation liant $[HC]_0$ à la sortie CO mis à part l'estimation obtenue pour une température de 0.4 une relation linéaire ne semble pas totalement inappropriée.

En ce qui concerne la sortie HC , les figures 4.9 nous présentent les résultats d'estimation des s_i . On voit sur la figure 4.9a, que la relation linéaire envisagée n'est pas vérifiée pour deux températures, $T = 0.4$ et $T = 0.6$, ceci s'expliquant par la forte non linéarité de la sortie HC . Pour la fonction s_2 , fonction selon $[CO]_0$, la relation linéaire semble cohérente, sauf pour $T = 0.5$ où l'on observe sur la figure 4.9b une estimation affine par morceaux, avec une cassure pour $[CO]_0 = 0.7$, comme pour la sortie CO . Enfin, la figure 4.9c vient conforter la structure envisagée, puisque si l'on omet le début de l'estimation de s_3 à une température de 0.4, la forme des courbes obtenues montre une relation quasi-linéaire.



(a) s_1 : GAM relatif à la richesse (b) s_2 : GAM relatif à la $[CO]_0$ (c) s_3 : GAM relatif à la $[HC]_0$
 Figure 4.9 – Représentation des fonctions s_i du modèle GAM 4.5 pour HC à différentes températures

On en conclut que bien que la structure de dépendance linéaire envisagée pour le modèle à coefficients variables ne soit pas la structure la plus adaptée, celle-ci ne semble pas trop inadéquate et les estimations que nous exhibons pour les différentes températures, laissent néanmoins présager de bons résultats de prédiction avec cette structure. De plus, devant le faible nombre de données expérimentales à notre disposition (20 essais) dans un espace à trois dimensions, la valeur à accorder à ces résultats est restreinte. En effet, si nous avions observé une structure claire apparaître avec si peu de points, il nous aurait été possible de conclure précisément, ce qui n'est pas le cas. Pour autant, cela ne nous permet pas d'infirmier ou confirmer la structure envisagée.

La structure utilisée par la suite est donc définie par :

$$y(T) = a_0(T) + a_1(T)riche + a_2(T)[CO]_0 + a_3(T)[HC]_0 + \epsilon \tag{4.6}$$

avec ϵ de moyenne nulle et de variance σ^2 .

4.1.2 Méthode 1 : Estimation des fonctions coefficients par P-splines

L'objectif de cette partie est de présenter la première méthode fonctionnelle testée. Il s'agit d'une méthode basée sur un modèle à coefficients variables, issue de la théorie des données longitudinales (Singer et Willet [146]). La variable longitudinale est discrétisée. Ainsi, en chacun de ces points, on obtient un modèle de régression linéaire, dont les coefficients sont estimés par moindres carrés. Des travaux ont montré, Li *et al* [95], qu'une deuxième étape de lissage peut être pertinente pour améliorer la modélisation, surtout lorsque les fonctions coefficients obtenues présentent des fluctuations.

La procédure mise en œuvre s'est inspirée de l'article de Li, Sudjianto et Zhang, [95]. Dans leurs travaux, la méthode implémentée comporte deux étapes, avec tout d'abord, l'estimation des fonctions coefficients point à point selon la variable longitudinale, puis une étape de lissage à l'aide de P-splines. Nous présentons donc dans les sections suivantes, les résultats après une première étape dans laquelle aucun lissage n'est effectué puis les résultats à l'issue de la seconde étape.

4.1.2.1 Première étape : Estimation point à point des fonctions coefficients par moindres carrés

Soit n le nombre d'observations à disposition, p le nombre de variables explicatives et m la dimension de la variable longitudinale T . On pose,

$$\mathbf{X} = \begin{pmatrix} x_0^1 & \cdots & x_p^1 \\ \vdots & \ddots & \vdots \\ x_0^n & \cdots & x_p^n \end{pmatrix}_{(n,p+1)}, \mathbf{A}^t = \begin{pmatrix} a_0(t_1) & \cdots & a_p(t_1) \\ \vdots & \ddots & \vdots \\ a_0(t_m) & \cdots & a_p(t_m) \end{pmatrix}_{(m,p+1)}, \mathbf{Y} = \begin{pmatrix} y^1(t_1) & \cdots & y^1(t_m) \\ \vdots & \ddots & \vdots \\ y^n(t_1) & \cdots & y^n(t_m) \end{pmatrix}_{(n,m)} \quad (4.7)$$

où \mathbf{X} correspond aux variables d'entrées avec $x_0^i = 1$, pour $i = 1, \dots, n$, $\mathbf{T} = (t_1, \dots, t_m)$ aux températures normalisées et \mathbf{Y} à la sortie normalisée considérée. À noter que les sorties *CO* et *HC* sont traitées indépendamment. On a alors :

$$\mathbf{Y} = \mathbf{X}\mathbf{A} + \Sigma. \quad (4.8)$$

On pose $\mathbf{y}(t_i) = (y^1(t_i), \dots, y^n(t_i))^t$ le vecteur des réponses en t_i pour les n jeux de données, soit la $i^{\text{ème}}$ colonne de \mathbf{Y} , Σ étant la matrice des erreurs.

Ainsi, pour t_i fixé, on estime $\mathbf{a}(t_i) = (a_0(t_i), a_1(t_i), a_2(t_i), a_3(t_i))^t$ par moindres carrés,

$$\hat{\mathbf{a}}(t_i) = (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{y}(t_i), \quad (4.9)$$

et on réitère pour $i = 1, \dots, m$. La prédiction pour un nouveau jeu de données \mathbf{x}^0 s'obtient alors par

$$y(x^0, t_i) = \sum_{j=0}^p \hat{a}_j(t_i) x_j^0, i = 1, \dots, m. \quad (4.10)$$

Le résultat de cette première étape est bruité (cf. figures 4.11 et 4.12).

Ici, nous avons échantillonné T uniformément en $m = 1001$ points, de manière arbitraire. En chacun de ces points, on obtient un modèle de régression linéaire, que l'on estime par moindres carrés, à partir des $n = 20$ expériences à dispositions. À noter que, contrairement au chapitre 3, les données brutes sont utilisées dans cette approche, afin de ne pas biaiser l'estimation que l'on pourrait obtenir du fait du prétraitement des données.

Nous donnons tout d'abord, le résultat des fonctions coefficients obtenues (figures 4.10a et 4.10b) pour la sortie *CO* et *HC*, qui est utilisé pour commenter les résultats de prédiction présentés.

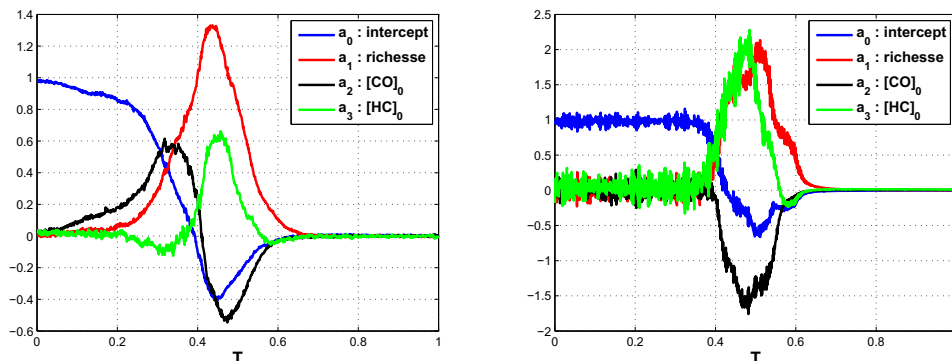
Les figures 4.11 et 4.12 illustrent quelques résultats obtenus à l'aide de cette méthode. En abscisse, nous avons la température normalisée, et y en ordonnée. Les courbes bleues correspondent aux résultats expérimentaux et les courbes rouges aux prédictions obtenues. Les figures 4.11a et 4.11b exemplifient les résultats pour la sortie *CO* puisqu'il s'agit du meilleur et pire cas (voir figure 4.13), selon le Integrated Square Error (ISE) (Scott [141]) définie par :

$$ISE = \int (y_{exp}(T) - \hat{y}(T))^2 dT, \quad (4.11)$$

où y_{exp} est la courbe expérimentale et \hat{y} la courbe prédite. De même pour les figures 4.12a et 4.12b concernant *HC*. En pratique, la valeur de l'ISE est calculée par la formule suivante :

$$ISE = \frac{1}{n_T} \sum_{i=1}^{n_T} (y_{exp}(T_i) - \hat{y}(T_i))^2 \quad (4.12)$$

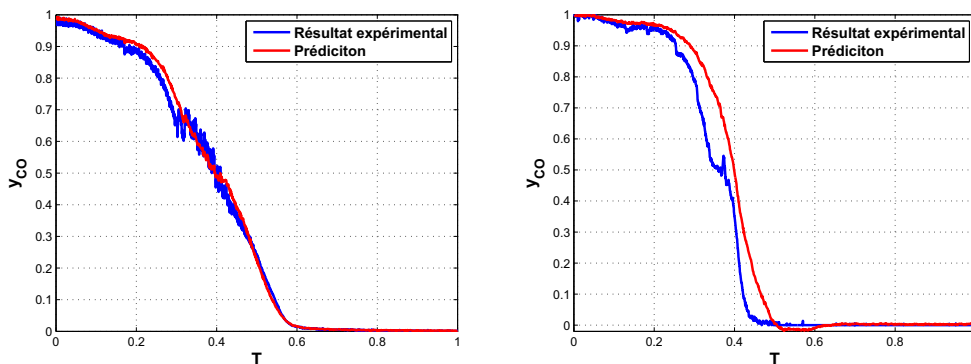
avec n_T le nombre de températures en lesquelles on calcule l'erreur entre expériences et prédictions. Ces points sont les mêmes quelles que soient les méthodes que nous présentons dans ce chapitre, de manière à pouvoir comparer les valeurs entre elles, et ainsi comparer les résultats des méthodes.



(a) Fonctions coefficients pour CO (b) Fonctions coefficients pour HC
 Figure 4.10 – Fonctions coefficients du modèle 4.6 pour la sortie CO et HC

En ce qui concerne la sortie CO , avec le modèle 4.6, pour le meilleur cas, on observe une prédiction très proche du résultat expérimental, et pour le pire cas, la forme de la courbe est proche mais on a une translation de la prédiction par rapport au résultat expérimental. En revanche, on observe une zone de prédiction de concentration négative, ce qui est inacceptable.

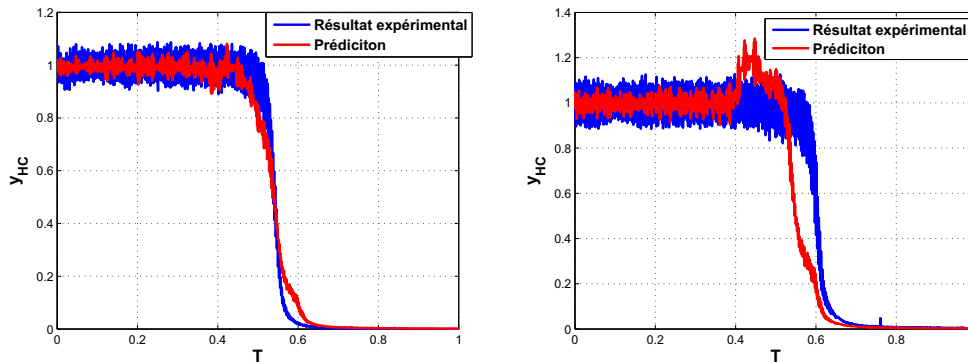
Pour ce qui est de la sortie HC , on voit que dans le pire cas, figure 4.12b, cette procédure entraîne une oscillation, juste avant le début de la décroissance. Cette oscillation, là encore, n’a aucun sens physique. De plus, pour certains cas, non présentés ici, on a également des prédictions pouvant atteindre, en fin de décroissance, des concentrations négatives.



(a) Meilleur cas (Essai 11) : ISE=4.4955e-4 (b) Pire cas (Essai 1) : ISE= 0.0055

Figure 4.11 – Comparaison des résultats expérimentaux (courbes bleues) et prédits (courbes rouges) avec le modèle 4.6, pour la sortie CO

Sur la figure 4.11b, la zone où la concentration devient négative, se situe aux alentours de $T = 0.55$. Cette figure correspond à l’essai 1 pour lequel on a une concentration initiale de CO et de HC très élevée et une richesse très faible (concentration de O_2 très forte). Si l’on observe les fonctions coefficients, on voit que la zone proche de la température de 0.55, est une zone dans laquelle a_2 , fonction correspondant à $[CO]_0$ à quasiment la même influence que a_1 relative à la richesse, qui elle, est très faible, les deux autres fonctions coefficients se compensant mutuellement. Ceci explique que dans cette zone, des cas extrêmes puissent conduire à des concentrations négatives.

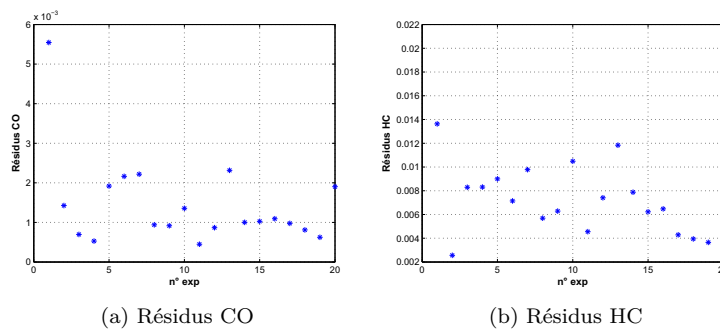


(a) Meilleur cas (Essai 2) : ISE=0.0026

(b) Pire cas (Essai 20) : ISE=0.0213

Figure 4.12 – Comparaison des résultats expérimentaux (courbes bleues) et prédits (courbes rouges) avec le modèle 4.6, pour la sortie HC

Pour la sortie HC , le pire cas présenté, essai 20, correspond au cas où la richesse est la plus forte parmi tous nos essais, avec des valeurs initiales de CO et HC de même ordre. En regardant attentivement la figure 4.10b, on voit que dans la zone de température normalisée comprise entre 0.4 et 0.5, zone où l'oscillation est présente dans la prédiction, les fonctions a_2 et a_3 , relatives respectivement à $[CO]_0$ et $[HC]_0$ ont des influences contraires, pratiquement de même ordre, mais légèrement plus forte pour a_3 . Pour l'essai 20, ces deux variables sont de même ordre de grandeur, cela conduit à un premier pic. Ceci est prolongé par le fait que la richesse est très élevée et que la fonction a_1 atteint son maximum pour une température de 0.5, second pic que l'on retrouve dans la prédiction, en $T=0.5$. Il s'agit là encore d'un cas extrême.



(a) Résidus CO

(b) Résidus HC

Figure 4.13 – Résidus de prédiction pour les deux sorties et pour chaque expérience

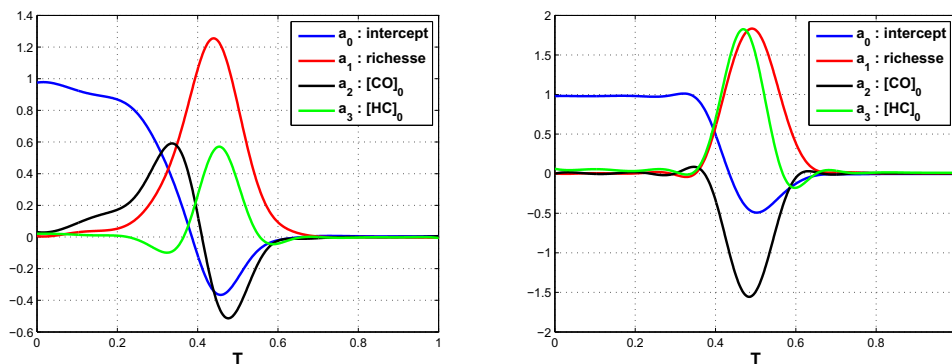
En conclusion, cette approche naturelle d'un point de vue physique, conduit à des comportements inacceptables. Regardons maintenant si les résultats sont améliorés si l'on effectue un lissage sur les estimations ponctuelles obtenues. En ce basant sur les travaux de Li *et al* [95], l'utilisation d'une seconde étape de lissage peut apparaître pertinente pour corriger les phénomènes décrits précédemment. D'autant plus que les fonctions coefficients sont très oscillantes.

4.1.2.2 Seconde étape : Lissage des fonctions coefficients par P-spline

La procédure basée sur une variable longitudinale, n'est pas suffisante pour aboutir à des résultats pertinents d'un point de vue physique. Afin d'y remédier, nous avons mis en place une seconde étape de lissage à l'aide de splines pénalisées (P-splines). L'idée est de tenir compte en un point des estimations précédentes et suivantes. Cette

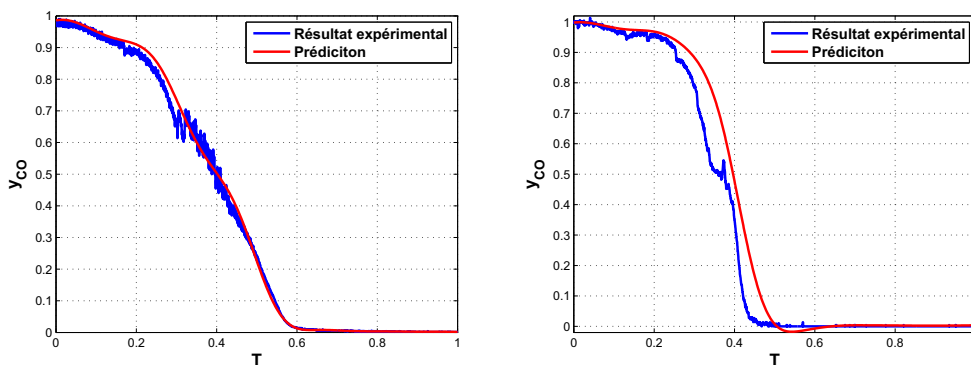
approche a été inspirée des travaux de Li *et al* 2005 [95]. Cette étape de lissage est d’autant plus intéressante que les fonctions coefficients oscillent.

La théorie des P-splines est présentée dans le chapitre 2. Pour notre cas, nous utilisons les P-splines introduits par Ruppert et Carroll, [135], avec des bases de fonctions d’ordre $r = 3$ et les nœuds choisis sont les points de la discrétisation de la variable longitudinale T . Ces choix ont été motivés par les considérations pratiques présentées dans l’article de Li *et al* [95]. Nous présentons tout d’abord le résultat des fonctions coefficients, figures 4.14a et 4.14b avant de présenter les résultats de prédiction.



(a) Fonctions coefficients lissées pour CO (b) Fonctions coefficients lissées pour HC
 Figure 4.14 – Fonctions coefficients du modèle 4.6 lissés par P-splines pour la sortie CO et HC

On peut voir que la procédure de lissage a permis d’obtenir des fonctions coefficients régulières. Cette étape a également entraîné un rabotage au sommet des différentes bosses. Notamment, pour les fonctions coefficients relatives à HC , on observe que pour a_1 , ce lissage élimine le petit palier juste avant le sommet, de même pour a_0 aux alentours de $T = 0.5$. Il est remarquable que les zones où le lissage modifie les fonctions coefficients coïncident avec les zones où l’on a des problèmes de prédiction. Étudions les courbes de prédictions des essais pris dans la première étape.



(a) Essai 11 : ISE=2.1788e-4 (b) Essai 1 : ISE=0.0049
 Figure 4.15 – Comparaison des résultats expérimentaux (courbes bleues) et prédits (courbes rouges) avec le modèle 4.6 et lissage des fonctions coefficients par P-splines, pour la sortie CO

Si l’on compare ces résultats aux résultats obtenus après la première étape, on remarque que les prédictions obtenues sont régulières. Ceci provient de la régularité estimée des fonctions coefficients. Néanmoins, on observe

toujours des cas où la prédiction de concentration est négative, comme sur la figure 4.15b. Le lissage ne traite pas le problème de fond d'inadéquation du modèle linéaire.

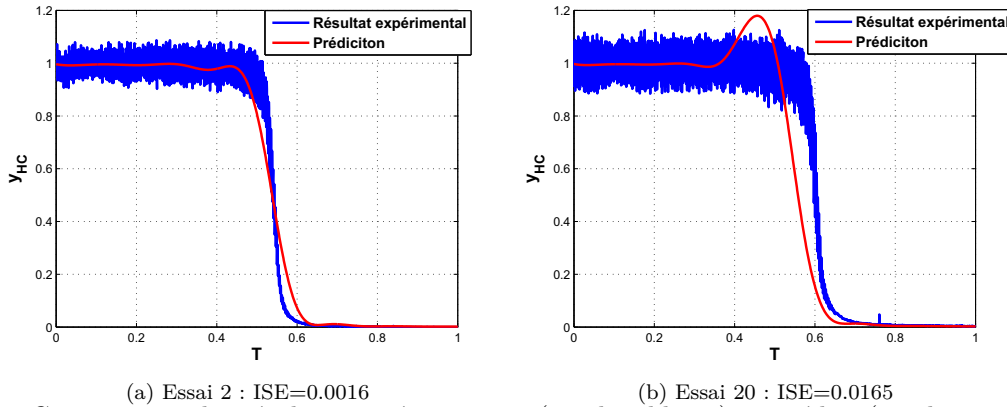


Figure 4.16 – Comparaison des résultats expérimentaux (courbes bleues) et prédits (courbes rouges) avec le modèle 4.6 et lissage des fonctions coefficients par P-splines, pour la sortie HC

Pour la sortie HC , figures 4.16a et 4.16b, là encore, les prédictions sont régulières du fait de la régularité des fonctions coefficients. Ceci n'a néanmoins pas permis de supprimer la bosse observée juste avant la décroissance, comme sur la figure 4.12b. On note toutefois que les valeurs des ISE pour les deux sorties ont été réduites après lissage des fonctions coefficients, indiquant de meilleures prédictions. Le tableau 4.1 résume les valeurs obtenues.

	ISE CO		ISE HC	
	Pire cas	Meilleur cas	Pire cas	Meilleur cas
Première étape	0.0055	4.4955e-4	0.0213	0.0026
Seconde étape	0.0049	2.1788e-4	0.0165	0.0016

Tableau 4.1 – Récapitulatif des valeurs des ISE de la première et second étape sur les essais présentés

L'étape de lissage est d'autant plus intéressante que les fonctions coefficients présentent des oscillations. Comme dans notre cas, la méthode de lissage a conduit à des courbes régulières. De plus, les zones où l'étape de lissage a principalement agi correspondent aux zones où l'on observe des phénomènes non physiques lors des prévisions en utilisant la première approche. Malgré ce constat, cette étape de lissage, n'a pas pu résoudre les problèmes rencontrés (oscillations avant la décroissance, et valeurs de concentration négatives).

Afin de nous rendre compte si les problèmes de prédictions peuvent venir de la structure de dépendance choisie (structure linéaire dont nous savons, voir section 4.1.1.2, que pour certains cas, ce choix n'est pas forcément le plus adapté), étudions plus en détails la figure 4.16b, correspondant au pire cas pour HC , et interprétons-le à l'aide notamment des courbes de GAM introduites dans la section 4.1.2.

D'un point de vue expérimental, ce cas correspond à un cas où la richesse est très élevée. On note que pour ce cas l'oscillation avant la décroissance intervient entre $T = 0.4$ et $T = 0.5$. Or, si l'on regarde la courbe GAM obtenue pour HC selon la richesse, figure 4.9a, on s'aperçoit que pour $T = 0.4$ une relation linéaire n'était pas des plus appropriées, et le fait que l'on soit sur un extremum (richesse normalisée égale à 1) peut expliquer en partie le phénomène rencontré. Il convient donc d'étudier, ce que donne une autre estimation du modèle VCM, avant de conclure sur la validité du modèle.

Cette seconde méthode d'estimation des fonctions coefficients est l'objet de la prochaine section. Dans cette seconde méthode, l'estimation est effectuée par polynômes locaux. L'idée est d'estimer les fonctions coefficients à l'aide d'une méthode directe, adaptée aux données fonctionnelles. En outre, on espère ainsi conclure quant à la mise en cause de la structure de dépendance dans les problèmes de prédiction rencontrés.

4.1.3 Méthode 2 : Estimation des fonctions coefficients par polynômes locaux

Nous nous proposons dans cette partie, d'introduire et appliquer la seconde méthode d'estimation des fonctions coefficients testée. Dans cette méthode l'estimation est effectuée par des polynômes locaux, l'ordre des polynômes étant fixé à 3.

Les modèles fonctionnels linéaires sont très utiles pour l'étude des données longitudinales. Ces modèles incluent entre autres les modèles additifs, Hastie et Tibshirani [80], et les modèles à coefficients variables, Hastie et Tibshirani [81]. L'estimation des fonctions coefficients a donné lieu à de nombreux travaux, Hoover *et al* [86], Brumback et Rice [18], Cardot, Ferraty et Sarda [23]. Parmi ces méthodes d'estimation, les splines de lissage et les méthodes à noyaux ont été proposées dans le but de fournir une estimation non paramétrique. Néanmoins, ces méthodes d'estimation sont soit très lourdes algorithmiquement, soit inefficaces en terme de performance de prédiction. Pour corriger ces problèmes, Fan et Zhang, [56], [55], proposent une estimation simple en deux étapes, basée sur une méthode de régression non paramétrique : les polynômes locaux d'ordre 3.

Le principe général de la méthode, dans le cas classique est introduit dans le chapitre 2. Dans un premier temps, nous présentons des considérations pratiques quant à l'application de notre méthode, et dans un second temps, nous exhibons les résultats de prédiction. Enfin, nous concluons et donnons quelques perspectives pour cette méthode d'analyse de données fonctionnelles.

4.1.3.1 Présentation de la méthode

Le principe de la méthode est le suivant : dans une première étape, on utilise une régression linéaire locale pour obtenir une estimation initiale des fonctions coefficients, puis, dans une seconde étape, on estime par un polynôme local cubique une des fonctions coefficients en substituant aux autres fonctions leurs estimations initiales. Cette procédure est répétée selon toutes les directions. Lors de l'application de cette méthode de régression non paramétrique, il est notamment nécessaire de déterminer les fenêtres de recherche optimales : fenêtre initiale globale et fenêtres dans chaque direction. Plusieurs travaux traitent de ce point tels que Ruppert *et al* [136] ou Fan et Gijbels [51]. La procédure choisie est une règle empirique dite du pouce décrite dans Fan et Gijbels [52].

Avant de présenter les résultats d'application de cette méthode sur notre cas d'étude, plusieurs points pratiques doivent être explicités. Tout d'abord le choix du noyau. Nous avons fait le choix de prendre le noyau d'Epanechnikov, défini par

$$K(t) = 0.75(1 - t^2)\mathbb{1}_{|t| \leq 1} \quad (4.13)$$

La formule de la fenêtre optimale peut être déterminée en minimisant l'erreur moyenne quadratique asymptotique intégrée. Cette dérivation peut être trouvée dans l'ouvrage de Fan et Gijbels, [52]. La partie ambigu de ce processus repose sur la méthode à utiliser pour estimer les quantités inconnues telles que la variance, certaines fonctions dérivées et de densité.

Fan et Gijbels, [52], suggèrent l'ajustement d'un polynôme de degré quatre par la méthode des moindres carrés ordinaires et en utilisant les dérivés de ce polynôme et la somme des carrés des résidus standardisés comme des estimations de leurs analogues inconnus. Ruppert *et al.*, [136], suggèrent d'utiliser une technique similaire, sauf que leur polynôme est ajusté par blocs. Le nombre de blocs est choisi en fonction de la statistique de Mallows [103].

Nous avons retenu la règle dite du pouce (ROT pour *Rule Of Thumb*) décrite dans l'ouvrage de Fan et Gijbels,[52] section 4.2. Le principal avantage de cette méthode est sa rapidité. En reprenant les notations utilisées par les auteurs, nous présentons brièvement la méthode ci-dessous.

En pratique, on détermine des estimations de la fonction de densité, $f(\cdot)$, de la variance conditionnelle, $\sigma^2(\cdot)$ et de $m^{(p+1)}(\cdot)$, la dérivée $p + 1^{\text{ème}}$ de l'espérance conditionnelle $m(x)$, en ajustant globalement un polynôme d'ordre $p + 3$ à cette dernière, soit

$$\check{m}(x) = \check{\alpha}_0 + \dots + \check{\alpha}_{p+3}x^{p+3}. \quad (4.14)$$

On s'intéresse à l'estimation de $m^{(\nu)}(\cdot)$ et on pose comme fonction poids $w(x) = f(x)w_0(x)$ pour des fonctions spécifiques de w_0 . En substituant ces estimations dans la formule de fenêtre constante optimale de Fan et Gijbels,[52], on obtient

$$h_{ROT} = C_{\nu,p}(K) \left[\frac{\check{\sigma}^2 \int w_0(x) dx}{\sum_{i=1}^n [\check{m}^{(p+1)}(X_i)]^2 w_0(X_i)} \right]^{1/(2p+3)}, \quad (4.15)$$

où $C_{nu,p}(K) = \left[\frac{(p+1)!^2 (2\nu+1) \int K_\nu^{*2}(t) dt}{2^{(p+1-\nu)} \left[\int t^{p+1} K_\nu^{*2}(t) dt \right]^2} \right]^{-1/(2p+3)}$ est une constante (tabulée pour différents noyaux K et différentes valeurs de ν et p), et $\check{\sigma}^2$ est la somme des carrés des résidus standardisés de l'ajustement (4.14).

Cette selection de fenêtre simple est dérivée sous certaines conditions. Cependant, même dans des situations où ces conditions ne sont pas strictement vérifiées, cette formule peut être appliquée afin d'avoir une idée du degré de lissage à utiliser.

L'estimateur en deux étapes est calculé par la formule 2.95 dans laquelle interviennent les estimations initiales des fonctions coefficients au travers des V_i . Contrairement aux exemples d'application de cette méthode, dans notre cas, il est nécessaire de réaliser l'estimation selon toutes les directions, et dans ce contexte, il apparaît plus judicieux de remplacer à chaque itération, c'est-à-dire pour chaque direction, les fonctions coefficients obtenues initialement par les estimateurs en deux étapes déjà calculés. Le pseudo code (algorithme 4.1) permet une meilleure compréhension de cette démarche.

En revanche, dans ce cadre, on s'aperçoit que l'ordre dans lequel les directions sont traitées peut avoir une influence. Afin de contourner ce problème, nous réalisons une boucle sur l'estimation des p fonctions coefficients, jusqu'au moment où ces estimations se stabilisent. En pratique, nous avons pu constater qu'à partir de trois itérations, ces coefficients ne variaient plus. De plus, nous avons effectué des tests en changeant l'ordre des variables d'entrée pour en mesurer l'influence, et nous avons pu constater que grâce à cette boucle supplémentaire, l'ordre n'intervient plus..

Notations dans notre cas

Le cadre classique dans lequel est appliquée la méthode VCM, est celui d'entrées fonctionnelles et d'une sortie scalaire. Dans notre cas, nous avons des sorties fonctionnelles et des entrées scalaires. La méthode reste valide dans cette configuration mais nécessite de redéfinir les notations précédemment introduites (section 2.2.2.2).

Soit p le nombre de variables explicatives envisagées, contenant le terme constant. Nous avons à disposition n expériences qui correspondent à différentes conditions opératoires (richesse, $[HC]_0$, $[CO]_0$). Sur chacune de ces courbes expérimentales, nous réalisons une discrétisation selon la température en m points.

Ainsi, on définit la matrice X_1 , de taille $(nm) \times 2p$ de la manière suivante

$$X_1 = \begin{pmatrix} X_{11}^1 & X_{11}^1(T_1 - t_0) & \cdots & X_{1p}^1 & X_{1p}^1(T_1 - t_0) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ X_{m1}^1 & X_{m1}^1(T_m - t_0) & \cdots & X_{mp}^1 & X_{mp}^1(T_m - t_0) \\ X_{11}^2 & X_{11}^2(T_1 - t_0) & \cdots & X_{1p}^2 & X_{1p}^2(T_1 - t_0) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ X_{m1}^n & X_{m1}^n(T_m - t_0) & \cdots & X_{mp}^n & X_{mp}^n(T_m - t_0) \end{pmatrix}_{(nm, 2p)}, \quad (4.16)$$

et le vecteur des réponses associées Y de taille (nm) par

$$Y = (Y_1^1, \dots, Y_m^1, Y_1^2, \dots, Y_m^2, \dots, Y_m^n)^T, \quad (4.17)$$

où l'exposant i , $i = 1, \dots, n$ est relatif à l'expérience i . La notation X_{kj}^i correspond donc à l'entrée j à la température k pour l'expérience i .

On pose également

$$W_1 = \text{diag}(K_{h_1}^1(T_1 - t_0), \dots, K_{h_1}^1(T_m - t_0), \dots, K_{h_1}^n(T_m - t_0)), \quad (4.18)$$

où $K_h(T_i - t_0) = \frac{3}{4h} \left(1 - \frac{(T_i - t_0)^2}{h^2}\right) \mathbb{1}_{|T_i - t_0| \leq h}$, $i = 1, \dots, m$, et h_1 est la fenêtre initiale choisie.

Quelques remarques concernant ces notations. Dans notre cas, les entrées sont constantes au cours d'une expérience, autrement dit, nous avons

$$\forall j = 1, \dots, p \quad X_{kj}^i = X_{lj}^i, \quad i = 1, \dots, n \quad k, l = 1, \dots, m. \quad (4.19)$$

On dénote donc par X^i les entrées relatives à l'expérience i .

De plus, nous avons fait le choix de discrétiser la courbe en m points uniformément répartis selon la température. Cette température est normalisée entre 0 et 1, 0 correspondant à la plus grande valeur minimale commune entre les expériences et 1 à la plus petite valeur maximale commune. Par conséquent, selon chaque expérience, les températures choisies sont les mêmes, et correspondent également aux nœuds utilisés pour le calcul des estimateurs. On s'aperçoit donc que

$$K_h^i(T_k - t_0) = K_h^j(T_k - t_0) \quad \forall k = 1, \dots, m \quad \forall j = 1, \dots, n \quad (4.20)$$

Et enfin, il nous faut encore définir les notations nécessaires pour la deuxième étape. Soit

$$X_{2,k} = \begin{pmatrix} X_{1k}^1 & X_{1k}^1(T_1 - t_0) & X_{1k}^1(T_1 - t_0)^2 & X_{1k}^1(T_1 - t_0)^3 \\ \vdots & \vdots & \vdots & \vdots \\ X_{mk}^n & X_{mk}^n(T_m - t_0) & X_{mk}^n(T_m - t_0)^2 & X_{mk}^n(T_m - t_0)^3 \end{pmatrix}_{(nm, 2p)}, \quad \forall k = 1, \dots, p, \quad (4.21)$$

$$W_{2,k} = \text{diag}(K_{h_{2,k}}^1(T_1 - t_0), \dots, K_{h_{2,k}}^n(T_m - t_0)), \quad \forall k = 1, \dots, p \quad (4.22)$$

où $h_{2,k}$ correspond à la fenêtre optimale calculée par ROT pour la direction k et enfin

$$V_{2,k} = (V_{2,k}(T_1, X^1), \dots, V_{2,k}(T_m, X^1), V_{2,k}(T_1, X^2), \dots, V_{2,k}(T_m, X^n))^T \quad (4.23)$$

où

$$V_{2,k}(T_j, X^i) = \sum_{\substack{l=1 \\ l \neq k}}^p \hat{a}_{l,1}(T_i) X_{jl}^i. \quad (4.24)$$

Avec ces nouvelles notations, on a ainsi l'estimation initiale des fonctions coefficients qui est donnée par

$$\hat{a}_{j,1}(t_0) = e_{2j-1,2p}^T (X_1^T W_1 X_1)^{-1} X_1^T W_1 Y, \quad j = 1, \dots, p \quad (4.25)$$

et l'estimateur en deux étapes est donné par

$$\hat{a}_{k,2}(t_0) = (1 \ 0 \ 0 \ 0) (X_{2,k}^T W_{2,k} X_{2,k})^{-1} X_{2,k}^T W_{2,k} (Y - V_{2,k}), \quad k = 1, \dots, p \quad (4.26)$$

Voici le pseudo-code de la méthode implémentée :

Algorithme 4.1 Estimation des fonctions coefficients par polynômes locaux

Initialisation de h_1

Calcul des $\hat{a}_{j,1}(\cdot)$ par la formule 4.25, pour $j = 1, \dots, p$

Pour $i = 1$ à **3**

1. **Pour** $k = 1$ à p

(a) Calcul de $h_{2,k}$ par ROT

(b) Calcul des $\hat{a}_{k,2}(\cdot)$ par la formule 4.26

(c) On remplace $\hat{a}_{k,1}(\cdot)$ par $\hat{a}_{k,2}(\cdot)$

2. **Fin pour**

Fin pour

Tout d'abord mesurons l'influence de la fenêtre initiale h_1 . Le choix de cette fenêtre initiale n'est pas crucial, dans le sens où il suffit qu'elle soit suffisamment petite. D'autant plus que pour chaque itération, nous remplaçons la valeur initiale des fonctions coefficients par l'estimation obtenue à la deuxième étape, et que nous réitérons cette étape. Pour s'en convaincre, nous avons testé des fenêtres initiales de 0.011, 1 et 20, puis nous comparons le critère IMSE calculé dans chaque cas.

En réalisant cette vérification, on peut s'apercevoir que le choix de h_1 n'est pas primordial. En effet, dans le cas d'une fenêtre initiale de 0.01, le critère obtenu pour la sortie *CO* est le même. En revanche, lorsque l'on prend une fenêtre initiale supérieure, ce critère a tendance à augmenter. Néanmoins, si dans les deux cas, on augmente le nombre de boucles réalisées, on retrouve les mêmes valeurs de fonctions coefficients et donc la même valeur pour le critère.

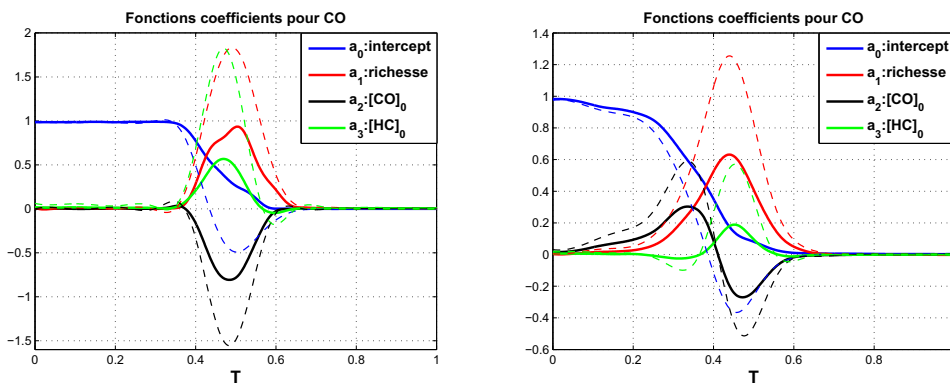
En conclusion, on peut voir que si la fenêtre initiale est mal choisie, c'est-à-dire trop grande, cela nécessitera davantage d'itérations pour arriver au résultat.

Concernant l'échantillonnage selon la température, nous avons pris les $m = 2652$ points d'échantillonnage des résultats expérimentaux, soit la totalité des points disponibles.

En revanche, dans le modèle que nous postulons, le fait de prendre un terme constant, entraîne un problème quant au calcul par ROT de la fenêtre (fenêtre nulle). Pour remédier à cela, la valeur de la fenêtre pour le terme constant a été déterminé par minimisation du ISE réalisé sur un échantillonnage fin entre 0 et 1 de $h_{2,1}$. Ainsi, la valeur de la fenêtre pour le terme constant est fixée à une valeur basse de 0.06.

4.1.3.2 Applications et résultats

Les figures 4.17a et 4.17b exhibent les fonctions coefficients, respectivement pour la sortie CO et HC . Les courbes pleines correspondent aux fonctions coefficients estimées par polynômes locaux, les courbes pointillées celles obtenues par la première méthode d'estimation (cf. section 4.1.2). Si l'on compare les deux méthodes au travers de ces fonctions coefficients, on peut voir que l'influence de chaque variable reste la même, plus précisément la forme des courbes relativement à chaque variable est proche, en revanche les amplitudes et donc leurs importances diffèrent. Malgré cette constatation, on obtient au final des résultats de prédictions comparables.

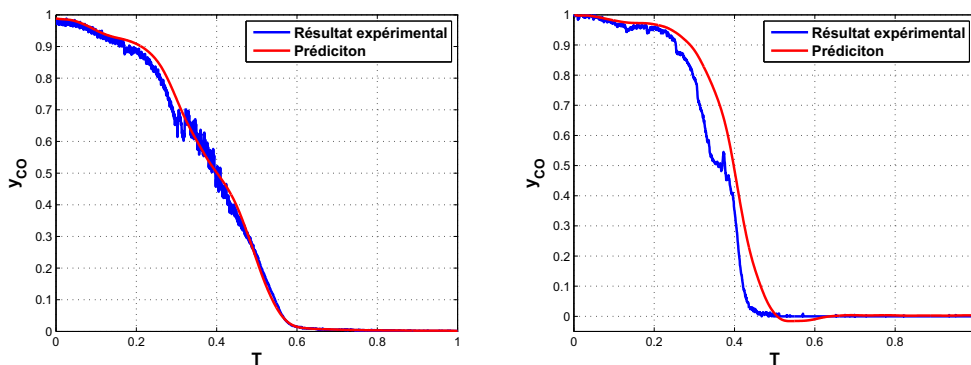


(a) Fonctions coefficients pour CO

(b) Fonctions coefficients pour HC

Figure 4.17 – Fonctions coefficients pour la sortie CO et HC estimées par la seconde méthode en traits pleins, et celles de la première méthode en pointillée

Les résultats présentés, figures 4.18 et 4.19, font références aux mêmes essais que ceux introduits dans la section 4.1.2.



(a) Essai 11 : ISE=4.6697e-4

(b) Essai 1 : ISE=0.0056

Figure 4.18 – Comparaison des résultats expérimentaux (courbes bleues) et prédits (courbes rouges) avec le modèle 4.6, pour la sortie CO

Les résultats exhibés sont intéressants puisque proches des résultats expérimentaux, tout en respectant la forme des courbes cinétiques étudiées. De plus, nous avons obtenu des valeurs d'ISE relativement faibles.

Cependant, bien que cette seconde méthode soit différente de celle présentée dans la section précédente, et donc

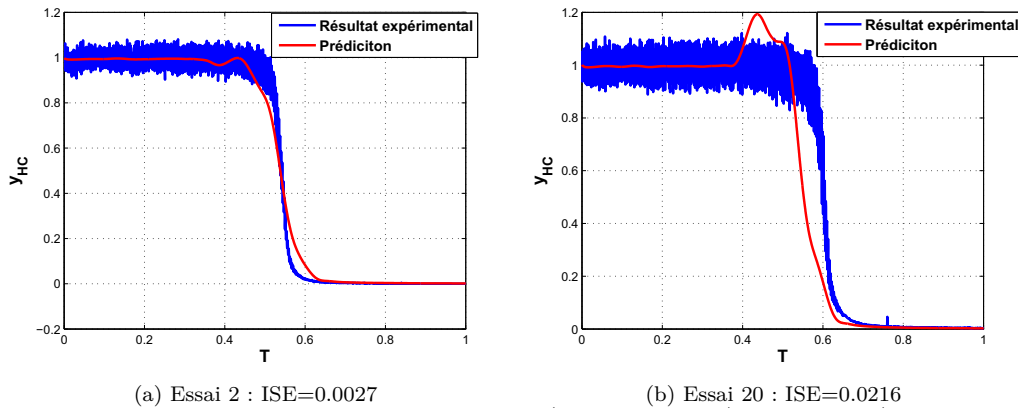


Figure 4.19 – Comparaison des résultats expérimentaux (courbes bleues) et prédits (courbes rouges) avec le modèle 4.6, pour la sortie HC

les fonctions coefficients estimées aussi (figures 4.17), les résultats nous conduisent à des prédictions semblables. Ainsi, on continue d’observer les mêmes phénomènes physiques inacceptables (concentrations négatives, bosses avant la décroissance) que ceux obtenus dans la section 4.1.2.

Plusieurs points ont également été vérifiés afin de tenter d’expliquer les problèmes de concentrations négatives, ou les oscillations avant la décroissance pour la sortie HC , notamment l’utilisation des données lissées, ou le calcul des fenêtres optimales. Cependant, aucune de ces vérifications ne nous a permis de mettre en défaut la méthode utilisée, mise à part le choix de la structure de dépendance linéaire.

En conclusion, on constate que cette méthode semble plus naturelle afin de traiter les données fonctionnelles comme dans notre cas. D’une part, par son approche plus adaptée à ce type de données, d’autre part par les résultats légèrement meilleurs obtenus. Cependant, il apparaît que dans cette méthode encore une fois le choix de la structure de dépendance est primordial.

Les diverses améliorations et vérifications apportées n’ont pas permis de corriger ces phénomènes. Il devient alors évident que la méthode mise en place n’est pas à mettre en cause, mais c’est bien la structure de dépendance linéaire, qui est à l’origine de ces résultats non pertinents.

4.1.4 Conclusions

Le but de cette première partie est l’application de méthodes fonctionnelles afin de modéliser le système physico-chimique étudié, le piège à NO_x . Il s’agit d’étudier des courbes cinétiques en fonction de la température. Dans ce contexte, plusieurs méthodes peuvent être envisagées pour répondre à nos besoins.

Du fait de la possibilité de prédire efficacement la TLO en fonction de la richesse, notre choix s’est légitimement porté sur des modèles à coefficients variables. Il s’agit de modèles de régression dans lesquels les coefficients dépendent d’une certaine variable longitudinale, généralement le temps, l’espace ou la température. Le cœur de ces méthodes réside dans l’estimation des fonctions coefficients.

Nous avons présenté deux méthodes d’estimation. La première consiste en une estimation point à point de ces fonctions coefficients par moindres carrés puis en un lissage des courbes obtenues par des P-splines. La seconde

méthode d'estimation propose une estimation de chacune des fonctions coefficients par des polynômes locaux cubiques.

À partir des variables explicatives déterminées, à savoir la richesse, $[CO]_0$ et $[HC]_0$, nous avons appliqué ces deux méthodes. Les résultats que nous avons obtenus, nous permettent de conclure que ces deux méthodes, bien que pertinentes, ne répondent pas totalement à nos attentes. En effet, certains comportements physiquement impossibles sont apparus en certains points expérimentaux.

Néanmoins, après avoir regardé plus précisément les raisons de ces défaillances, nous avons pu mettre en avant le fait que ces problèmes sont essentiellement dus à la structure de dépendance linéaire par rapport aux variables explicatives choisies.

Malgré l'étude de la dépendance entre les variables explicatives et les sorties, nous n'avons pu dégager de structure précise. Ceci provient essentiellement du nombre restreint d'essais expérimentaux à notre disposition. Ainsi, cette incapacité à définir la façon dont chaque variable influence les sorties est le principal problème des méthodes présentées. Il est important de souligner que l'on aurait pu adapter le modèle VCM si on avait pu identifier une structure de dépendance comme une structure linéaire par morceaux.

On en conclut que les deux méthodes mises en place dans ce chapitre bien qu'adaptées à notre cas d'étude, sont très influencées par la structure de dépendance localement (à température constante) utilisée. Dans notre cas, la dépendance linéaire n'est pas satisfaisante. Des travaux récents dans d'autres domaines peuvent peut être apporter une solution. Par exemple, les travaux de Chen et Tsay [27], dans lesquels les auteurs proposent dans le cadre de l'analyse de série temporelle non linéaire, une méthode d'estimation des fonctions liens autorégressives, ou encore dans le domaine de la classification avec le boosting, Tutz et Binder [153] [154], dans lesquels les variables d'intérêts sont sélectionnées et les fonctions liens sont estimées.

4.2 Estimation des paramètres cinétiques

Dans la procédure de développement de système de dépollution catalytique, et d'une manière générale, de système cinétique, bien souvent, la première étape consiste à déterminer un modèle numérique censé représenter la réalité expérimentale. Néanmoins, dans la plupart des cas, la complexité de ces systèmes imposant des simplifications, ou bien la connaissance imparfaite de ce qui se passe en réalité, conduisent à des modèles numériques inadaptés. C'est le cas de notre modèle numérique (voir chapitre 1).

L'incapacité à déterminer un jeu de paramètres cinétiques pour toutes les expériences, nous a amenés dans un premier temps à ne pas tenir compte de ce modèle dans les approches fonctionnelles testées. Cependant, l'utilisation de méthodes fonctionnelles dans lesquelles l'expertise métier est totalement ignorée, a pour conséquence l'apparition de comportements inacceptables dans certains cas, comme par exemple des concentrations négatives. Ce constat démontre l'utilité du modèle numérique. En effet, bien que ce modèle soit inadapté pour représenter ce que nous observons expérimentalement, il apporte toutefois une information cruciale sur la nature des sorties étudiées, et donc sur le lien physico-chimique entre les variables d'entrée (concentrations initiales) et la concentration de sortie.

Dans ce contexte, nous nous sommes alors penchés sur les raisons de l'inadéquation du modèle numérique. Pour cela, nous avons remis en cause le dénominateur utilisé dans le calcul des vitesses de réactions pour des raisons expertises métiers que nous détaillons dans le chapitre 1 section 1.1.5. La modification de ce dénominateur a entraîné l'ajout d'un paramètre cinétique à ajuster, soit cinq paramètres au total, et a pour conséquence une amélioration de la pertinence du modèle.

La démarche d'estimation du vecteur des cinq paramètres cinétiques mise en place a été la suivante : premièrement, une estimation constante sur tout le domaine, deuxièmement, une estimation en chaque point à disposition (vingt points) du domaine, troisièmement, une estimation en fonction des prédicteurs. Pour résoudre cette troisième étape, nous avons commencé par la seconde afin de nous guider vers la fonctionnelle la plus adéquate.

Ce plan en trois étapes se justifie puisque l'on constate que le modèle est inadéquat si on prend toutes les expériences pour estimer les paramètres cinétiques, mais convient si on considère les expériences une à une. D'où l'idée d'estimer les paramètres expérience par expérience pour trouver un modèle qui relie les paramètres aux entrées.

Cette partie se compose de trois sections. Tout d'abord, nous rappelons brièvement les modifications apportées au modèle numérique, le nouveau vecteur de paramètre cinétique et la méthode d'estimation de ces paramètres expérience par expérience. Dans une seconde section, nous introduisons trois fonctionnelles reliant paramètres cinétiques et prédicteurs. Enfin dans une troisième section, nous introduisons la méthode mise en place pour le calcul des régions de confiance, et exhibons les résultats.

4.2.1 Modifications du modèle numérique et estimations des paramètres cinétiques

Les résultats de prédiction obtenus à l'aide des méthodes fonctionnelles basées sur des modèles à coefficients variables, ont fait apparaître que l'omission de l'expertise métier dans les modèles peut conduire à des comportements impossibles.

Dans cette optique, l'utilisation du modèle numérique des experts présenté dans le chapitre 1 semble être une solution pour éviter ce type de comportements. En effet, bien qu'inadapté, ce modèle numérique apporte une certaine connaissance sur le comportement du système. L'idée étant de tirer un maximum d'information de ce modèle.

Modification du modèle numérique et nouveau vecteurs

Comme nous l'avons précisé dans le premier chapitre, le modèle numérique postulé est inadéquat pour représenter efficacement la réalité expérimentale. Malgré des tentatives d'estimations des paramètres cinétiques, les résultats obtenus ne sont pas satisfaisants. Plus précisément, nous sommes incapables de déterminer un jeu de paramètres cinétiques valable pour toutes les expériences.

Nous avons décidé, en concertation avec les expérimentateurs, de modifier le dénominateur utilisé dans le calcul des vitesses de réactions. Après plusieurs études, nous sommes arrivés à déterminer une nouvelle expression pour ce dénominateur permettant de mieux ajuster les résultats du modèle aux résultats expérimentaux. Cette nouvelle expression nécessite également l'ajout d'un paramètre cinétique à estimer, la constante d'adsorption de la réaction de *HC*. Pour plus de détails, nous renvoyons le lecteur au chapitre 1, section 1.1.5.

Ainsi, nous avons au final un modèle numérique dépendant d'un vecteur β de cinq paramètres cinétiques à ajuster, soit respectivement, β_1 et β_2 le facteur pré exponentiel et énergie d'activation pour la réaction de *CO*, β_3 et β_4 ceux pour la réaction de *HC*, et β_5 la constante d'adsorption de *HC*.

Une fois ce nouveau modèle mis en place, nous avons estimé le vecteur des cinq paramètres cinétiques constant pour tous les points du domaine. De manière classique, cette estimation est effectuée par moindres carrés, soit

$$\hat{\beta} = \arg \min_{\beta} \sum_{i=1}^n \int \|y_i(T) - f(x_i, T, \beta)\|^2 dT \quad (4.27)$$

Malheureusement, tous les essais basés sur l'équation (4.27) se sont révélés infructueux, montrant bien que le modèle proposé est insuffisant pour représenter toute la réalité expérimentale, même si les grandes tendances sont respectées. Néanmoins, on constate tout de même une amélioration par rapport à ce que l'on avait précédemment (avant les modifications). Notamment, il est possible de déterminer un jeu de paramètres cinétiques expérience par expérience, ce qui n'était pas le cas avec le modèle précédent.

Estimation des paramètres cinétiques

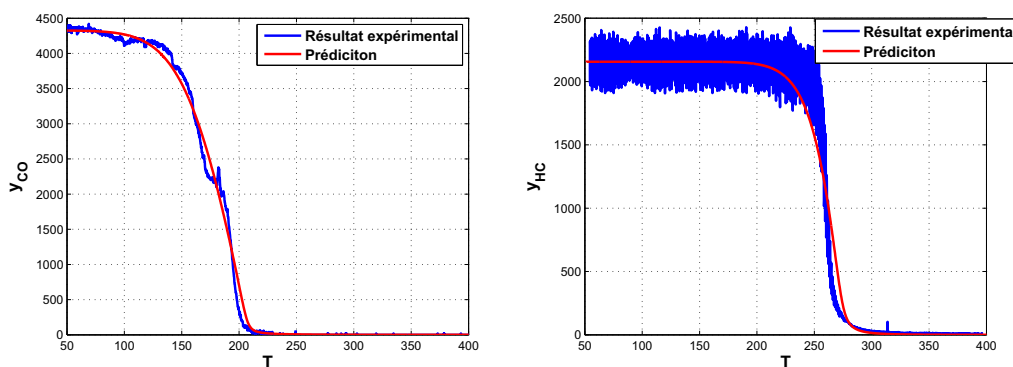
Malgré l'amélioration du modèle numérique, nous ne sommes toujours pas à même d'estimer un vecteur de paramètres cinétiques pertinents constants pour tous les points du domaine.

En revanche, il est possible de déterminer expérience par expérience un vecteur de paramètres cinétiques. Dans ce cas,

$$\hat{\beta}_i = \arg \min_{\beta} \int \|y_i(T) - f(x_i, T, \beta)\|^2 dT, \text{ pour } i = 1, \dots, n \tag{4.28}$$

où n est le nombre d'expériences à disposition, $f(\cdot)$ est le modèle numérique, $\beta_i = (\beta_{i,j})_{j=1,\dots,5}$ est le vecteur des cinq paramètres cinétiques de l'expérience i , T est la température et x_i représente les concentrations initiales de CO , HC et O_2 de l'expérience i . Cette approche revient à faire varier les paramètres en fonction des x_i . Par conséquent, on a, au final, cinq paramètres cinétiques à ajuster pour chacune des vingt expériences, soit cent paramètres.

Les figures 4.20a et 4.20b illustrent le type de prédictions que l'on obtient, et correspondent aux pires cas, respectivement pour CO et HC , étudiés dans la section précédente. Les courbes bleues représentent les résultats expérimentaux et les courbes rouges les prédictions obtenues à l'aide du modèle après estimation des paramètres cinétiques.



(a) Résultat pour CO de l'essai 1 : $ISE=4.9051e-4$ (b) Résultat pour HC de l'essai 20 : $ISE=0.0047$

Figure 4.20 – Comparaison des résultats expérimentaux (courbes bleues) et prédits (courbes rouges) avec le modèle numérique après estimation des paramètres pour la sortie CO de l'essai 1, et pour la sortie HC de l'essai 20

On peut voir que les prédictions obtenues dans ces cas présentés, sont relativement proches des résultats expérimentaux. De plus, nous n'avons pas de comportement atypique de ces prédictions. On observe également que les ISE de chacune de ces prédictions, sont largement inférieures à celles obtenues précédemment ($4.9e-4 < 5.5e-3$ pour CO et $4.7e-3 < 2.13e-2$ pour HC).

On rappelle que l'ISE est définie par l'équation

$$ISE = \frac{1}{n_T} \sum_{i=1}^{n_T} (y_{exp}(T_i) - \hat{y}(T_i))^2 \quad (4.29)$$

avec n_T le nombre de températures en lesquelles on calcule l'erreur entre expériences et prédictions (les mêmes que pour les méthodes introduites précédemment), et avec les valeurs de y_{exp} et \hat{y} normalisées (contrairement aux courbes présentées sur les figures 4.20).

On définit l'IMSE pour HC et CO , comme étant la somme des ISE de chaque essai, soit

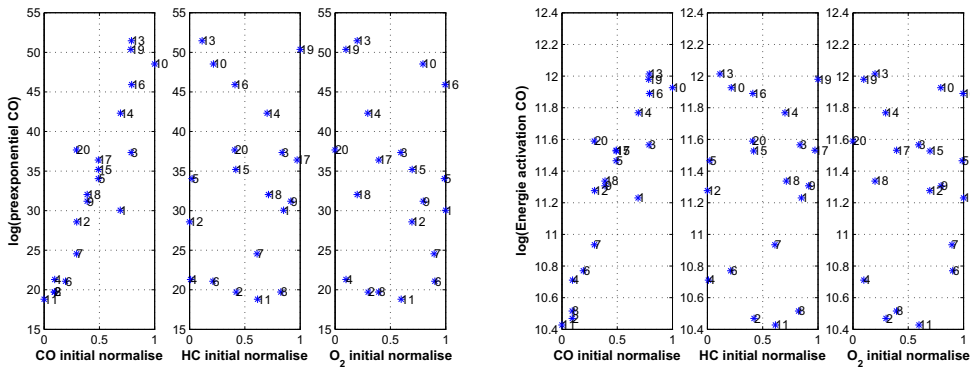
$$IMSE = \sum_{j=1}^n \frac{1}{n_T} \sum_{i=1}^{n_T} (y_{exp,j}(T_i) - \hat{y}_j(T_i))^2, \quad (4.30)$$

où $y_{exp,j}$ est la courbe expérimentale normalisée de l'expérience j et \hat{y}_j sa courbe normalisée prédite.

Pour ce qui est de la sortie HC , nous avons obtenu un IMSE de 0.0958, et pour CO de 0.0111, qui sont relativement faibles, et illustrent bien la proximité des prédictions par rapport aux résultats expérimentaux.

Nous allons maintenant étudier plus attentivement les valeurs de paramètres cinétiques estimés pour chacune des expériences, en fonction des prédicteurs. Le but est de pouvoir dégager une relation, si elle existe, reliant les valeurs de paramètres cinétiques à ces prédicteurs.

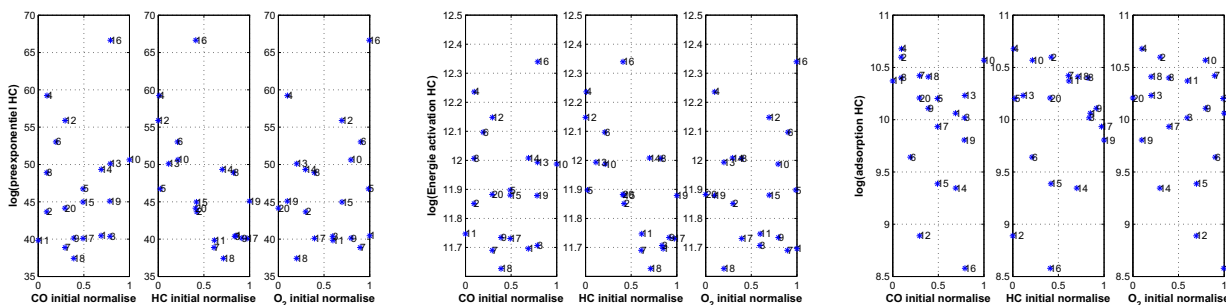
Les figures 4.21 et 4.22 exposent les valeurs des paramètres cinétiques (en log) obtenues fonction de $[CO]_0$, $[HC]_0$ et $[O_2]_0$. Les figures 4.21a et 4.21b exhibent les résultats des paramètres relatifs à CO , le facteur pré exponentiel et l'énergie d'activation de la réaction d'oxydation de CO . Les figures 4.22a, 4.22b et 4.22c représentent respectivement le facteur pré exponentiel et l'énergie d'activation de la réaction d'oxydation du HC et la constante d'adsorption de HC .



(a) Facteur pré exponentiel pour CO selon chacune des variables normalisées (b) Énergie d'activation pour CO selon chacune des variables normalisées

Figure 4.21 – Répartition des paramètres cinétiques relatifs à la réaction d'oxydation de CO fonction des variables normalisées

Les $\hat{\beta}_i$ obtenus sont dispersés, deux expériences proches pouvant conduire à des $\hat{\beta}_i$ très différents. L'exemple de l'essai 16, sur les figures 4.22 illustre bien ce phénomène, directement lié aux problèmes d'optimisation rencontrés.



(a) Facteur pré exponentiel pour HC selon chacune des variables normalisées (b) Énergie d'activation pour HC selon chacune des variables normalisées (c) Constante d'adsorption de HC fonction des variables normalisées
 Figure 4.22 – Répartition des paramètres cinétiques relatifs à la réaction d'oxydation de HC et de la constante d'adsorption de HC fonction des variables normalisées

Bien que les prédictions soient bonnes, nous souhaitons que les valeurs des paramètres cinétiques aient un sens du point de vue des expérimentateurs.

On remarque que pour certains paramètres, nous observons presque une relation linéaire entre les paramètres et une des variables. C'est le cas, par exemple, pour la valeur du paramètre pré exponentiel et de l'énergie d'activation de CO en fonction de [CO]₀. En revanche, ces relations observées ne concernent pas la variable O₂. Ce constat peut s'expliquer par le fait que les paramètres cinétiques sont relatifs à HC et CO. Par conséquent, nous choisissons d'utiliser uniquement [HC]₀ et [CO]₀ comme variable de la relation que nous mettons en place dans la section suivante.

4.2.2 Modélisation des paramètres cinétiques en fonction des entrées

Le but de cette section est de déterminer une modélisation pertinente des paramètres cinétiques en fonction des prédicteurs, définis comme étant les concentrations initiales de HC et CO. On cherche donc une fonction β(x) qui à x associe un vecteur de paramètres cinétiques tel que le modèle numérique soit proche des résultats expérimentaux.

Nous avons mis en avant dans la section précédente qu'il est possible de déterminer pour chaque essai à disposition un vecteur de paramètres cinétiques tel que le modèle numérique s'ajuste bien aux résultats expérimentaux. En étudiant la répartition de ces paramètres dans l'espace des entrées, nous avons pu observer une relation linéaire dans certains cas.

À partir de ce constat, nous avons mis en place une relation linéaire en les prédicteurs pour déterminer les paramètres cinétiques. On pose comme première approche, la fonction suivante :

$$\beta(x) = x\gamma \tag{4.31}$$

où γ représente les coefficients de la relation linéaire. La méthode repose donc sur l'estimation de ces coefficients. Les résultats de cette première fonction liant les paramètres cinétiques aux prédicteurs ne sont pas totalement satisfaisants, mais nous ont guidés vers deux autres méthodes d'estimation.

La première de ces deux méthodes consiste à résoudre un problème aux moindres carrés pondérés et à poser :

$$\hat{\beta}(x) = \arg \min_{\beta} \sum_{i=1}^n \int \|y_i(T) - f(x_i, T, \beta)\|^2 K_h(x_i, x) dT. \quad (4.32)$$

Dans cette première méthode la fonction noyau dépend des x . Dans le cas d'un noyau gaussien, le paramètre h est un vecteur de même dimension p que x .

La seconde méthode, dans la lignée de Hastie et Tibshirani [81], propose de remplacer les $\hat{\beta}_i$ du modèle 4.28 par des coefficients dont les estimations sont lissées. Pour ce faire, Hastie et Tibshirani utilisent des B-splines. Nous avons choisi d'employer la méthode "Kernel Ridge Regression" (KRR) (voir chapitre 2), exposée par exemple dans Saunders *et al.* [139] particulièrement simple à mettre en œuvre.

Les résultats obtenus par ces deux méthodes sont comparables. Cependant, la seconde offre l'avantage de déterminer l'équation d'une surface de réponse pour représenter l'évolution des paramètres, qu'il suffit d'évaluer pour déterminer un nouveau point, alors que dans la première méthode une optimisation est nécessaire à chaque fois que l'on veut déterminer $\hat{\beta}$ pour un nouveau point.

Par conséquent, nous n'introduisons, par la suite, que cette seconde méthode, qui est plus intéressante et pertinente pour notre cas. Il est à souligner que ces deux méthodes ne nécessitent que d'interfacer le calcul du modèle numérique, et sont donc non intrusives. De plus, ces méthodes sont des méthodes analogues à celles des VCM, et en étendent l'application à un cadre non-linéaire.

Nous présentons donc dans un premier temps, l'estimation des coefficients d'une relation linéaire entre prédicteurs et paramètres cinétiques, modèle 4.31, et les résultats obtenus. Dans un second temps, la théorie des KRR ayant été introduite dans le chapitre 2, nous l'appliquons sur notre cas d'étude.

4.2.2.1 Estimation des coefficients d'une relation linéaire entre prédicteurs et paramètres cinétiques

Partant des observations de relation linéaire entre les concentrations initiales de HC et CO , la première approche testée consiste à estimer les coefficients d'une relation linéaire entre les variables et les paramètres cinétiques.

Soient p le nombre de variables et q le nombre de paramètres cinétiques, on pose

$$\gamma = \begin{pmatrix} \gamma_{1,0} & \cdots & \gamma_{q,0} \\ \vdots & \ddots & \vdots \\ \gamma_{1,p} & \cdots & \gamma_{q,p} \end{pmatrix}_{(p+1,q)}. \quad (4.33)$$

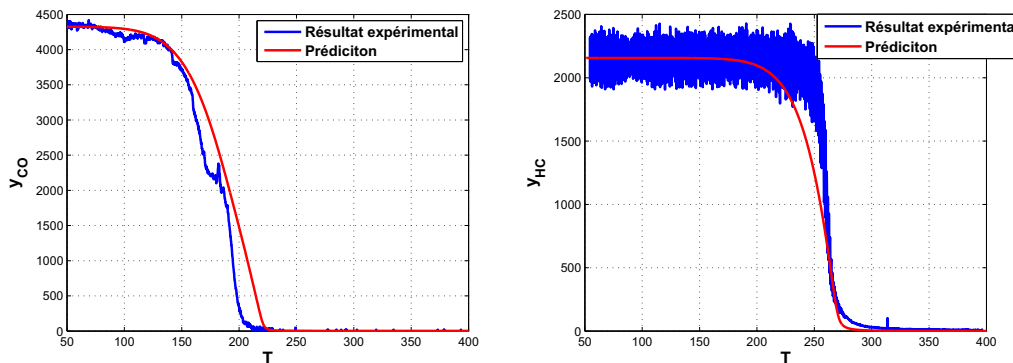
On note $x_i = (x_{i,j})_{j=1,\dots,p}$, les entrées relatives à l'expérience i , $i = 1, \dots, n$, avec n le nombre d'expériences.

L'approche consiste à estimer $\hat{\gamma}$, par moindres carrés, soit

$$\hat{\gamma} = \arg \min_{\gamma} \sum_{i=1}^n \int \|y_i(T) - f(x_i, T, \beta(x_i))\|^2 dT, \quad \text{avec } \beta(x_i) = (1, x_i) \gamma. \quad (4.34)$$

Deux variables sont retenues (i.e $p = 2$), les concentrations initiales de CO et de HC , et nous avons $q = 5$ paramètres. La méthode consiste donc à estimer les $q \times (p + 1) = 15$ paramètres de la matrice $\hat{\gamma}$.

Comme précédemment, nous présentons les résultats de prédictions de CO pour l'essai 1, et de HC pour l'essai 20, respectivement les figures 4.23a et 4.23b. Les courbes rouges correspondent aux prédictions et les courbes bleues aux résultats expérimentaux.



(a) Résultat pour CO de l'essai 1 : $ISE=0.0041$ (b) Résultat pour HC de l'essai 20 : $ISE=0.0072$

Figure 4.23 – Comparaison des résultats expérimentaux (courbes bleues) et prédits (courbes rouges) avec le modèle numérique après estimation des coefficients de la relation linéaire des paramètres par rapport aux variables de l'expérience pour la sortie CO de l'essai 1, et pour la sortie HC de l'essai 20

On peut voir sur ces deux figures, que les prédictions obtenues, sont encore une fois très proches des résultats expérimentaux. Afin de comparer les résultats obtenus à ceux précédemment présentés, nous avons estimé la matrice γ à partir de 19 des 20 expériences et prédits la vingtième, et avons réitéré cette démarche pour chaque essais à disposition. Ceci nous permet de calculer les ISE de chaque cas ainsi que les IMSE pour les sorties CO et HC .

En ce qui concerne les ISE, on voit que par rapport à la première approche, les prédictions sont moins bonnes ($4.1e-3 > 4.9e-4$ pour CO et $7.2e-3 > 4.7e-3$ pour HC). De plus, si l'on s'intéresse à l'IMSE pour cette approche, on obtient une valeur pour HC égale à 0.1453 et pour CO égale à 0.0432, soit des valeurs plus élevées que pour l'approche consistant à estimer les paramètres expérience par expérience (0.0958 et 0.0111 respectivement pour HC et CO). Néanmoins, les valeurs des paramètres cinétiques estimés semblent acceptables du point de vue des experts, et de plus, deux expériences proches conduisent à des paramètres cinétiques proches.

Ce constat peut s'expliquer par le fait que la relation linéaire que nous avons postulée, ne s'observe pas selon tous les prédicteurs (voir figures 4.21 et 4.22), par conséquent cette première modélisation est trop restrictive. Nous avons alors décidé dans un deuxième temps, de la relaxer en ne tenant compte pour l'estimation en un point que des points qui sont proches dans l'espace des entrées, proximité déterminée par un noyau. On retrouve donc le principe des kernel ridge regression.

4.2.2.2 Applications et résultats de la KRR

La théorie des KRR est présentée dans le chapitre 2. Nous présentons brièvement dans notre cas, comment nous avons utilisé cet outil.

Cette méthode dépend de deux hyperparamètres : un paramètre de régularisation λ et un paramètre pour la portée du noyau h . De manière classique, on utilise un noyau gaussien. Ces deux paramètres sont ajoutés aux paramètres cinétiques à estimer dans la procédure de minimisation.

L'avantage de la KRR pour notre cas, est de combiner une modélisation linéaire en les prédicteurs et un noyau gaussien pour ne tenir compte que des points proches dans l'espace des entrées. La simplicité d'application de cette méthode est un argument supplémentaire qui a guidé notre choix. Si pour chaque expérience i , le vecteur

b_i des "vrais" paramètres non lissés du modèle était connu, les coefficients estimés et lissés $\hat{\beta}_i$ se déduiraient de ceux-ci par la relation linéaire en b_i par

$$\hat{\beta}_i = k^t(\mathbf{K} + \lambda\mathbf{I})^{-1}b_i, \quad (4.35)$$

où $K_h(u, v)$ est une fonction noyau, k désigne le vecteur de dimension n dont chaque composante vaut $K(x_i, x)$ et \mathbf{K} est la matrice de composantes $K_h(x_i, x_j)$, \mathbf{I} est l'identité d'ordre n .

Cependant, les b_i sont inconnus et nous les collectons dans un vecteur \hat{b} qui contient donc $n \times q$ composantes pour n expériences et q paramètres cinétiques par expérience. Le critère à estimer s'apparente au critère 4.27, sauf que la minimisation porte maintenant sur b , à partir duquel on reconstruit chaque β_i :

$$\hat{b} = \arg \min_b \sum_{i=1}^n \int \|y_i(T) - f(x_i, T, \beta_i)\|^2 dT \quad (4.36)$$

Dans le cas de notre application, $q = 5$ et $n = 20$ et la minimisation de 4.36 porte donc sur $n \times q = 100$ variables au lieu des 5 initiales.

Nous présentons ici deux résultats de prédiction obtenus, un pour chaque sortie, effectués sur deux essais distincts. Comme précédemment, nous avons choisi les essais 1 et 20 respectivement pour la sortie CO et HC . Les courbes bleues correspondent aux résultats expérimentaux et les courbes rouges aux prédictions obtenues. Mais avant cela nous donnons les ISE de chaque essai sur les figures 4.24.

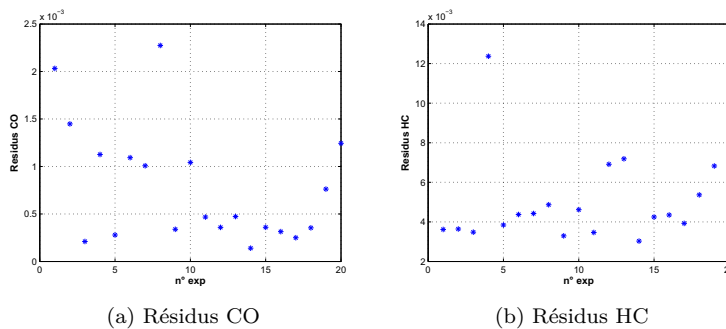
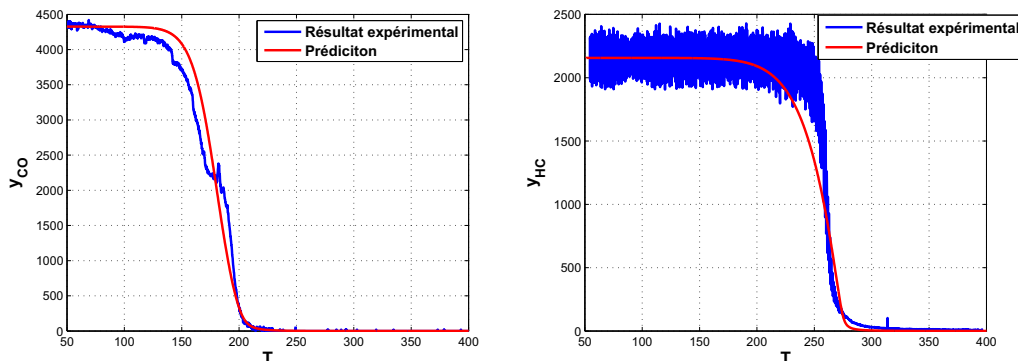


Figure 4.24 – Résidus de prédiction pour les deux sorties et pour chaque expérience

Comme les deux précédentes approches (détermination d'un vecteur global expérience par expérience et par une relation linéaire en les prédicteurs), les ISE sont relativement faibles. Les valeurs obtenues sont plus petites que celles déterminées par la seconde approche, et du même ordre que celles de l'approche 1.

Ce constat est également valable pour les autres essais (voir figure 4.24). Cela se traduit par de bonnes prédictions de nos essais, comme on peut le constater sur les figures 4.25. Ces prédictions sont proches des résultats expérimentaux. En outre, dans cette approche, les valeurs des paramètres cinétiques que nous avons estimés, ont un sens du point de vue expérimentateur.



(a) Résultat pour CO de l'essai 1 : 0.0020

(b) Résultat pour HC de l'essai 20 : 0.0068

Figure 4.25 – Comparaison des résultats expérimentaux (courbes bleues) et prédits (courbes rouges) avec le modèle numérique après estimation par KRR des paramètres cinétiques fonction des variables de l'expérience pour la sortie CO de l'essai 1, et pour la sortie HC de l'essai 20

4.2.2.3 Récapitulatif des résultats des différentes approches et conclusion

Nous avons présenté différentes approches d'estimation des paramètres que nous avons pu comparer grâce notamment aux ISE sur les essais 1 et 20 respectivement pour CO et HC , et aux IMSE. Dans ces différentes approches, nous nous sommes également intéressés aux valeurs des paramètres cinétiques et à leur répartition dans l'espace des entrées. Le tableau 4.2 fournit un récapitulatif des résultats de ces méthodes en précisant sur quelle variable porte la minimisation, sa taille, sa relation la liant aux prédicteurs x_i et enfin la valeur de l'IMSE pour chaque sortie.

Approche	Variable	Taille	Relation avec les x_i	IMSE	
				CO	HC
1	β	100	aucune	0.0958	0.0111
2	γ	15	linéaire	0.1453	0.0432
3	b	100	KRR	0.1007	0.0156

Tableau 4.2 – Récapitulatif des méthodes d'estimation des paramètres cinétiques et des résultats

La première approche consiste en l'estimation par moindres carrés des paramètres cinétiques expérience par expérience. On a $q = 5$ paramètres cinétiques à ajuster pour chacune des $n = 20$ expériences, soit 100 paramètres.

La seconde approche consiste à estimer les coefficients γ d'une relation linéaire liant les paramètres cinétiques aux prédicteurs. Deux variables sont retenues (i.e $p = 2$), les concentrations initiales de CO et de HC . La méthode consiste donc à estimer les $q \times (p + 1) = 15$ paramètres de la matrice $\hat{\gamma}$.

Enfin la troisième méthode est équivalente à estimer les paramètres d'une méthode des moindres carrés pénalisés, en tenant compte de toutes les expériences (poids suivant la distance au point courant).

L'approche 2 a l'avantage d'avoir un nombre de paramètres à estimer beaucoup plus faible que les deux autres approches, mais la relation linéaire postulée est trop restrictive, et l'on s'aperçoit que les résultats de prédiction sont moins bons.

Les prédictions de l'approche 3 sont proches de celles obtenues par l'approche 1, puisqu'on obtient un IMSE de 0.1007 pour HC et de 0.0156 pour CO , proches de ceux obtenus par l'approche 1 (respectivement de 0.0958 et 0.0111).

Ces faibles valeurs sont représentatives des bonnes prédictions obtenues. En revanche, les valeurs des paramètres cinétiques que nous avons estimés dans l'approche 3, ont un sens du point de vue expérimentateur. En cela, cette méthode est très pertinente pour notre cas puisqu'elle allie performance de prédiction et validité des valeurs estimées.

Nous avons développé un outil qui est performant en terme de prédiction et censé d'un point de vue expérimentateur. Nous sommes donc maintenant à même de prédire pour n'importe quel nouvel essai, la sortie des polluants. La prise en compte du modèle cinétique apporte une légitimité des résultats, et la méthode d'estimation des paramètres cinétiques par KRR en fonction de $[HC]_0$ et $[CO]_0$ contraint leur valeur dans un domaine raisonnable.

4.2.3 Calcul des régions de confiance

La modélisation par KRR des paramètres cinétiques, nous permet désormais de prédire pour n'importe quel point du domaine, un vecteur de paramètres cinétiques tel que le modèle numérique s'ajuste correctement aux résultats expérimentaux. L'objectif maintenant est de pouvoir quantifier la qualité et la robustesse de la méthode d'optimisation que nous avons mise en place. On propose pour cela, de calculer les régions de confiance.

L'idée est d'une part de pouvoir fournir un outil de diagnostic qualitatif sur les prédictions, et d'autre part, à l'instar de ce que nous envisagions dans le chapitre 3, de définir un critère de plan d'expériences selon ces régions de confiance. Par exemple, on peut choisir de placer les nouveaux points d'expériences là où l'aire de cette région de confiance est la plus grande.

Comme nous l'avons expliqué, la méthode de modélisation des paramètres cinétiques par KRR repose sur l'estimation d'une matrice b (voir équation 4.36) contenant $n \times q = 100$ paramètres. Pour calculer la région de confiance relative à la prédiction pour un point du domaine x^0 , la démarche consiste premièrement à générer un grand nombre de \hat{b} par bootstrap et deuxièmement, à utiliser ces estimations pour reconstruire un grand nombre de vecteurs de paramètres cinétiques à partir desquels nous prédisons la sortie du modèle en x^0 . L'enveloppe de ces prédictions déterminant la région de confiance.

Dans les cas où l'échantillon à disposition est trop petit mais également lorsque la loi de l'estimateur *a priori* est inconnue, les méthodes statistiques classiques ne peuvent être utilisées pour l'estimation de la loi des paramètres. La méthode de bootstrap (Efron [43]) est couramment utilisée (Da Veiga [38], Hertog [40]) dans ce cadre. Par conséquent, nous avons décidé de mettre en place une méthode de bootstrap des résidus, adaptée aux données fonctionnelles, afin de générer un grand nombre de matrices b , utilisées dans l'approche considérée, pour reconstruire les paramètres cinétiques.

Nous présentons dans un premier temps la méthode mise en place, appelée bootstrap fonctionnel, consécutivement à une introduction générale sur le bootstrap. Dans un second temps, nous appliquons la méthode sur notre cas d'étude, et discutons les résultats, notamment au travers du calcul de la région de confiance pour un des essais à disposition afin de pouvoir conclure quant à la pertinence de notre démarche.

4.2.3.1 Bootstrap fonctionnel

Le but de cette partie est d'introduire la méthode de bootstrap fonctionnel développée afin d'estimer la loi des paramètres de la modélisation par KRR mise en place. Cette méthode s'apparente à un bootstrap des résidus mais dans un cadre fonctionnel avec un modèle non-linéaire. Afin de mieux comprendre la méthode mise en place, nous présentons dans un premier temps une introduction générale du bootstrap avant de présenter en détail dans le cas d'un modèle de régression la procédure de bootstrap des résidus. Enfin, nous terminons par la présentation de l'application de la méthode dans notre cas pratique.

Principe du bootstrap

L'article de Efron, [43] présente la méthode dans le cas de variables aléatoires identiquement distribuées (i.i.d). D'importants travaux ont par la suite vu le jour de sorte qu'aujourd'hui les applications comme les ouvrages sur le bootstrap sont nombreux, nous pouvons citer entre autres ceux de Efron et Tibshirani [44], Shao et Tu [143] ou encore Davison et Hinkley [39].

Depuis, la méthode a été étendue aux cas des variables non i.i.d. dans divers contextes comme dans les modèles ARMA (Hardle *et al* [77]), le block bootstrap (Lahiri, [93]) ou encore le bootstrap autorégressif non-paramétrique (Franke *et al* [65]). Dans le cadre fonctionnel, cette approche peut se justifier grâce aux résultats de Bickel et Freedman [13] et Freedman [66].

Le principe de la méthode du bootstrap est de générer une série d'échantillons aléatoires et simples avec remise de n observations dans l'échantillon initial. Ces échantillons successifs seront notés :

$$\mathbf{x}_1^*, \mathbf{x}_2^*, \dots, \mathbf{x}_B^*, \quad (4.37)$$

où B est le nombre de rééchantillonnages effectués.

On considère $\hat{\theta}$ un estimateur d'un paramètre θ_0 de la distribution de l'échantillon par exemple la moyenne, la médiane ou la variance. On s'intéresse alors à la distribution de $\hat{\theta} - \theta_0$. Pour cela, on calcule pour chaque échantillon généré \mathbf{x}_i^* , l'estimateur $\hat{\theta}_i^*$ de la même manière que l'on a estimé $\hat{\theta}$. Au final, on s'attend à ce que la distribution de $\hat{\theta}_i^* - \hat{\theta}$ approche la distribution de $\hat{\theta} - \theta_0$.

Les deux cas les plus classiques reposent sur la connaissance d'appartenance ou non de la distribution à une famille de lois paramétriques. Si par exemple, on sait que la loi de l'échantillon initiale suit une loi paramétrique, on parle alors de bootstrap paramétrique, sinon on parle de bootstrap non-paramétrique ou empirique.

D'autres choix existent, l'article de Horowitz [87] en donne quelques exemples. Il est à noter que si le bootstrap paramétrique fournit une meilleure approximation que le bootstrap empirique lorsque l'on est assuré que la loi de l'échantillon appartient à une famille paramétrique, ce n'est plus le cas si l'on n'a aucune connaissance *a priori* sur cette loi (auquel cas l'hypothèse d'appartenance à une famille paramétrique peut donner de mauvais résultats si l'on choisit cette famille éloignée de la loi de l'échantillon).

La méthode dans le cas non paramétrique peut se résumer de la manière suivante :

1. On génère un échantillon $\mathbf{x}^* = (x_1^*, \dots, x_n^*)$, en tirant aléatoirement avec remise dans l'échantillon initiale $\mathbf{x} = (x_1, \dots, x_n)$
2. On calcule $\hat{\theta}^*$ à partir de \mathbf{x}^*

3. On répète cette opération B fois, avec B grand, afin d'obtenir un échantillon $(\hat{\theta}_1^*, \dots, \hat{\theta}_B^*)$ de l'estimateur bootstrap

Bootstrap des résidus

La technique de rééchantillonnage présentée ci-dessus est la plus simple et la plus courante. Des méthodes un peu différentes sont utilisées pour des applications particulières. Ainsi, dans les problèmes de régression lorsque les valeurs des variables explicatives sont fixées a priori par l'utilisateur, le rééchantillonnage d'individus peut difficilement se justifier. Dans une telle situation, on peut remplacer le bootstrap des individus par le bootstrap des résidus.

Soit \mathbf{y} le vecteur de la variable à expliquer et \mathbf{X} la matrice des variables explicatives. Soit $\hat{\boldsymbol{\theta}}$ le vecteur des coefficients estimés par une méthode donnée d'ajustement. On peut calculer le vecteur des valeurs estimées de la variable à expliquer et en déduire le vecteur des résidus :

$$\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}. \quad (4.38)$$

Dans le cas du modèle linéaire, on a :

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\theta}}, \quad (4.39)$$

mais, de manière plus générale, \mathbf{y}_i peut être une fonction quelconque des valeurs observées des variables explicatives et des paramètres estimés :

$$\hat{\mathbf{y}} = f(\mathbf{X}, \hat{\boldsymbol{\theta}}). \quad (4.40)$$

Soit \mathbf{e}_k^* un échantillon aléatoire prélevé avec remise dans le vecteur \mathbf{e} . En additionnant, \mathbf{e}_k^* à la partie déterministe du modèle $f(\mathbf{X}, \hat{\boldsymbol{\theta}})$, on obtient le vecteur \mathbf{y}_k^* :

$$\mathbf{y}_k^* = f(\mathbf{X}, \hat{\boldsymbol{\theta}}) + \mathbf{e}_k^*. \quad (4.41)$$

Ce vecteur est un rééchantillonnage à partir duquel on estime $\hat{\boldsymbol{\theta}}_k^*$, par la même méthode utilisée pour estimer $\hat{\boldsymbol{\theta}}$. Comme précédemment, on réitère cette étape B fois.

La méthode de bootstrap des résidus peut donc se résumer de la manière suivante :

1. On génère un échantillon $\mathbf{e}^* = (e_1^*, \dots, e_n^*)$, en tirant aléatoirement avec remise dans l'échantillon initiale $\mathbf{e} = (e_1, \dots, e_n)$
2. On calcule $\mathbf{y}^* = f(\mathbf{X}, \hat{\boldsymbol{\theta}}) + \mathbf{e}^*$
3. On estime $\hat{\boldsymbol{\theta}}^*$ à partir de \mathbf{y}^*
4. On répète cette opérations B fois, avec B grand, afin d'obtenir un échantillon $(\hat{\boldsymbol{\theta}}_1^*, \dots, \hat{\boldsymbol{\theta}}_B^*)$ de l'estimateur bootstrap

Bootstrap fonctionnel

Dans notre cas d'étude, les observations sont des courbes. De plus, les valeurs des variables explicatives sont fixées a priori par l'utilisateur. Ainsi, nous avons dû développer une méthode de bootstrap des résidus applicable lorsque les sorties étudiées sont des courbes.

Soit $\mathbf{y} = (y_1(T), \dots, y_n(T))$ la matrice des variables étudiées, où T est la variable fonctionnelle et $y_i(T)$ un vecteur de taille m , correspondant à l'expérience i . On pose le modèle :

$$\hat{\mathbf{y}}_i(T) = f(\mathbf{x}_i, T, \hat{\boldsymbol{\beta}}(x_i)), i = 1, \dots, n. \quad (4.42)$$

avec $\hat{\boldsymbol{\beta}}$ obtenu par KRR.

Le vecteur des résidus $\mathbf{e}_i(T)$ pour l'expérience i est donc donné par $\hat{\mathbf{y}}_i(T) - \mathbf{y}_i(T)$. Pour chaque expérience, on obtient donc une courbe de résidus.

Basée sur le bootstrap des résidus dans le cas de modèle de régression, notre méthode consiste à répéter B fois les étapes suivantes :

1. On génère un échantillon $\mathbf{e}^* = (e_1^*(T), \dots, e_n^*(T))$, en tirant aléatoirement avec remise dans l'échantillon initiale $\mathbf{e} = (e_1(T), \dots, e_n(T))$
2. On calcule $\mathbf{y}_i^*(T) = f(\mathbf{x}_i, T, \hat{\boldsymbol{\beta}}(x_i)) + \mathbf{e}_i^*(T)$ pour $i = 1, \dots, n$
3. On estime $\hat{\boldsymbol{\beta}}^*$ à partir de $\mathbf{y}^* = (y_1^*(T), \dots, y_n^*(T))$

Au final, on obtient un échantillon $(\hat{\boldsymbol{\beta}}_1^*, \dots, \hat{\boldsymbol{\beta}}_B^*)$ de l'estimateur bootstrap.

4.2.3.2 Estimation de la loi des paramètres de la modélisation par KRR

Nous avons $n = 20$ expériences à disposition. À partir de ces essais, la procédure d'optimisation nous permet d'estimer la matrice \hat{b} de la modélisation par KRR des paramètres cinétiques. Lors de cette estimation nous avons également déterminé le paramètre de lissage du KRR, λ et le paramètre de portée du noyau h . Dans la suite, λ et h sont fixes.

Il est également à noter que dans notre cas, nous avons deux sorties pour chaque expérience, la sortie CO et HC en fonction de la température T . Ainsi, dans la procédure de bootstrap fonctionnel que nous voulons tester, deux choix s'offrent à nous quant à la génération à chaque étape des nouveaux $\mathbf{y}_i^*(T)$, pour $i = 1, \dots, n$. Soit nous déterminons aléatoirement et indépendamment le vecteur des résidus pour chaque sortie et chaque expérience, soit nous fonctionnons avec des paires de courbes de résidus lors du tirage aléatoire avec remise. C'est cette option qui a été choisie, puisque les résidus des deux sorties pour une expérience donnée sont corrélés.

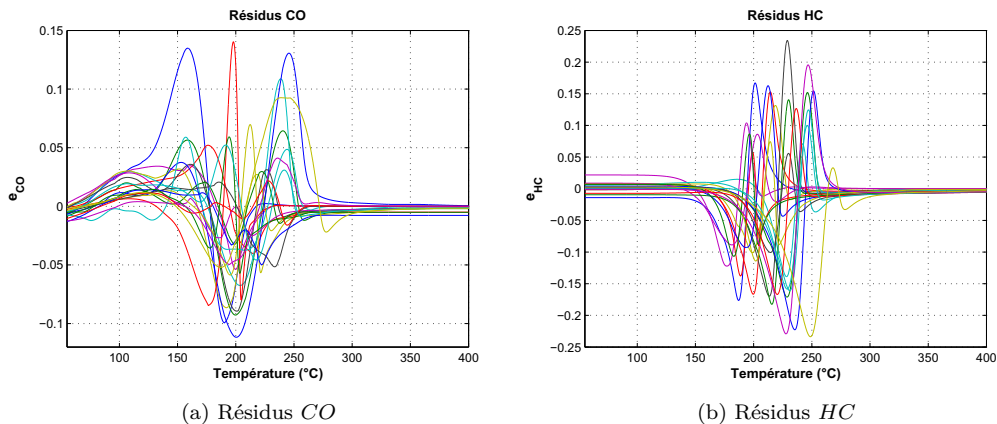


Figure 4.26 – Résidus entre résultats expérimentaux et prédits

Les figures 4.26 exposent les résidus pour CO et HC que nous avons obtenus suite à la méthode d'estimation avec utilisation de la KRR. Ce sont ces résidus dans lesquels nous tirons aléatoirement pour générer à chaque étape $\mathbf{y}_i^*(T)$.

Nous avons réalisé $B = 1000$ boucles de bootstrap. À chaque étape, nous avons donc une base de calibration de $n = 20$ expériences obtenues en rééchantillonnant les résidus par paires (CO et HC) aléatoirement avec remise.

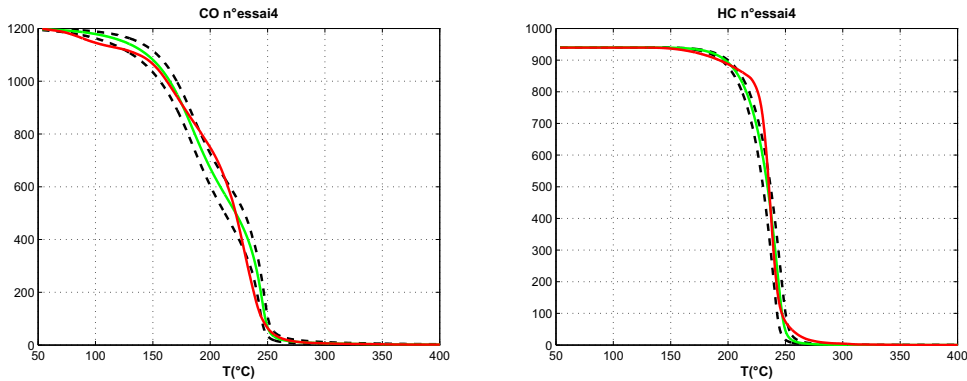
La démarche se résume de la façon suivante :

1. On estime la matrice \hat{b} de la modélisation par KRR et des hyperparamètres associés à partir des $n = 20$ essais à disposition
2. On calcule les paramètres cinétiques optimaux $\hat{\beta}_i$ pour $i = 1, \dots, n$
3. On calcule les paires de courbes de résidus $e_i(T)$ (sorties CO et HC) entre modèles numériques et résultats expérimentaux, pour chaque essai avec les paramètres cinétiques optimaux estimés.
4. On réalise B boucles de bootstrap fonctionnel des résidus de la manière suivante
 - (a) On génère un échantillon $\mathbf{e}^* = (e_1^*(T), \dots, e_n^*(T))$, en tirant aléatoirement avec remise dans l'échantillon initial $\mathbf{e} = (e_1(T), \dots, e_n(T))$
 - (b) On calcule $\mathbf{y}_i^*(T) = f(\mathbf{x}_i, T, \hat{\beta}(x_i)) + \mathbf{e}_i^*(T)$ pour $i = 1, \dots, n$
 - (c) On estime la matrice \hat{b}^* par notre procédure d'optimisation à partir de $\mathbf{y}^* = (y_1^*(T), \dots, y_n^*(T))$
 - (d) On construit $\hat{\beta}^*$ par KRR avec \hat{b}^*

4.2.3.3 Calcul de la région de confiance

Afin d'illustrer le calcul de la région de confiance, nous avons généré 1000 jeux de paramètres cinétiques, à partir de 100 générations de \hat{b} , pour un de nos essais à disposition. Les 1000 prédictions ainsi obtenues, délimitent la région de confiance.

La figure 4.27 expose les résultats pour les deux sorties HC et CO . Les courbes rouges correspondent aux résultats expérimentaux lissés (choix pour une meilleure visibilité), les courbes vertes représentent les prédictions obtenues à l'aide de notre méthode d'estimation et les courbes pointillées délimitent la région de confiance à 95%.



(a) Région de confiance pour CO

(b) Région de confiance pour HC

Figure 4.27 – Région de confiance de la prédiction relative à l'essai 14

On voit donc sur cet exemple que la région de confiance, qui est centrée sur la prédiction, est relativement resserrée. On constate que pour la sortie *CO*, le résultat expérimental est compris dans cette région, ce qui est très intéressant. En revanche, pour le *HC*, cette observation n'est pas totalement vérifiée.

Nous avons donc pu voir que l'approche mise en place dans son ensemble, c'est-à-dire la modélisation par kernel ridge regression des paramètres cinétiques et le calcul des régions de confiance à l'aide du bootstrap fonctionnel est une approche pertinente pour la modélisation du phénomène physique étudiée. Les résultats de prédiction sont proches des résultats expérimentaux comme nous avons pu le constater sur les essais à disposition.

Il est important de signaler, que la procédure de calcul de la région de confiance mise en place, fournit de très bons résultats et permet donc d'avoir un critère de qualité et de robustesse des prédictions. Du point de vue des expérimentateurs, ces résultats sont relativement pertinents, et laissent entrevoir des possibilités d'application sur d'autres cas, du fait du caractère général de la méthode.

4.2.4 Conclusions

Nous nous sommes rendu compte qu'une modification dans le calcul des vitesses de réaction et l'ajout d'un paramètre cinétique nous permet un meilleur ajustement. Ainsi, nous avons pu mettre en avant la possibilité d'estimer pour chaque expérience un jeu de paramètres cinétiques efficace.

Les valeurs de paramètres obtenues nous permettent de prédire précisément chacune des expériences, mais dans certains cas, ces valeurs n'ont aucun sens d'un point de vue expérimentateur. L'idée est alors de déterminer ces valeurs en fonction des variables d'entrées afin de contraindre les estimations obtenues dans un cadre pertinent. La méthode retenue est une estimation par Kernel Ridge Regression. Le choix de cette méthode s'est fait consécutivement à plusieurs méthodes successives.

Une fois les paramètres cinétiques estimés, à partir des variables d'entrée, il nous est possible de prédire les sorties en n'importe quel nouveau point d'essai. De plus, les sorties *CO* et *HC* sont traitées simultanément dans cette approche. Toutefois, il est nécessaire de quantifier la valeur des prédictions obtenues. Pour cela, nous avons mis en place une méthode de bootstrap fonctionnel afin d'estimer la loi des estimations obtenues.

Cette méthode de bootstrap fonctionnel nous a permis de calculer de la région de confiance pour chacune des prédictions. L'idée étant, comme auparavant, de fournir un outil efficace de prédiction pour les expérimentateurs, de quantifier cette efficacité, et au travers de cette valeur, de définir la méthodologie de plan d'expériences la plus adaptée, une des idées envisagées étant de choisir les points du domaine dont l'aire de la région de confiance est la plus élevée, méthodologie que nous n'avons pas eu le temps de développer.

4.3 Conclusions et perspectives

Conclusions

Nous avons exposé dans ce chapitre deux types de méthodes tenant compte du caractère fonctionnel du problème, afin de modéliser les sorties expérimentales. La première est basée sur un modèle à coefficients variables (VCM), en omettant l'expertise métier (le modèle numérique), la seconde consiste en l'estimation des paramètres cinétiques du modèle numérique par une relation à partir des conditions expérimentales.

L'hypothèse forte sur laquelle est basée la première méthode est que, à température constante, les sorties en fonctions des conditions expérimentales peuvent se modéliser par une relation linéaire. L'examen de ces points à

différentes températures permet de valider l'idée d'une dépendance longitudinale en température pour la sortie CO mais pas pour HC . Comme on pouvait s'y attendre, les résultats obtenus par les deux méthodes d'estimation des fonctions coefficients (par P-splines ou par polynômes locaux) pour la première réponse sont acceptables, mais ne sont pas satisfaisants pour la seconde, les courbes pouvant présenter au moment du démarrage effectif de la réaction un sursaut choquant pour l'œil expert. Cette première approche a le grave inconvénient de ne pas tenir compte de la connaissance des experts en ignorant le modèle numérique et de forcer un caractère linéaire quelle que soit la température considérée.

Pour remédier à cela, nous avons décidé de réincorporer l'expertise métier dans notre modélisation. La seconde approche est une méthode non-intrusive qui offre l'avantage majeur de déterminer l'équation d'une surface de réponse pour représenter l'évolution des paramètres, qu'il suffit d'évaluer pour déterminer un nouveau point. Cette approche emploie une méthode intrinsèquement analogue à celle des VCM, mais étendue au cadre non-linéaire. En effet, il est proposé de modéliser les paramètres cinétiques par kernel ridge regression à partir des conditions expérimentales. Les résultats de prédiction ont permis de conclure que cette méthode est très pertinente pour notre cas puisqu'elle allie performance de prédiction et validité des valeurs estimées.

Nous avons développé un outil qui est performant en terme de prédiction et censé d'un point de vue expérimentateur. Nous sommes donc maintenant à même de prédire pour n'importe quel nouvel essai, la sortie des polluants. La prise en compte du modèle cinétique apporte une légitimité des résultats, et la méthode d'estimation des paramètres cinétiques par KRR en fonction de $[HC]_0$ et $[CO]_0$ contraint leur valeur dans un domaine raisonnable.

À partir de là, nous avons réussi à développer une méthodologie judicieuse pour quantifier la qualité de ces prédictions. Pour cela, nous avons mis en place une méthode de bootstrap fonctionnel afin d'estimer la loi des estimations obtenues, et ainsi déterminer les régions de confiance des prédictions. Les résultats présentés sont convenables et représentatifs de ce que l'on est en droit d'attendre.

L'approche ainsi mise en place dans son ensemble, c'est-à-dire la modélisation par kernel ridge regression des paramètres cinétiques et le calcul des régions de confiance à l'aide du bootstrap fonctionnel est une approche pertinente pour la modélisation du phénomène physique étudié et la planification d'expériences que nous proposons de baser sur l'aire des régions de confiance obtenues.

Perspectives

Les perspectives sont nombreuses. D'une part nous avons développé une première approche (VCM) intéressante sous certaines conditions, qui certes ne sont pas respectées dans notre cas d'étude, mais qui peuvent être rencontrées par les expérimentateurs. D'autre part, la méthodologie d'estimation par KRR complétée par le calcul des régions de confiance par bootstrap fonctionnel, laisse entrevoir un critère efficace pour répondre pleinement à la problématique posée initialement par cette thèse, et plus particulièrement pour notre cas d'étude.

Il est également intéressant de souligner que cette méthodologie nécessite une base d'expériences convenablement définie. Notre choix s'est porté sur un critère de maximum d'entropie, motivé par plusieurs avantages et caractéristiques. C'est l'objet du chapitre suivant. Cette première étape vient achever la méthodologie envisagée, que l'on peut résumer de la manière suivante :

1. Génération d'un premier plan d'expériences par maximum d'entropie (cf. chapitre 5),
2. Estimation de la matrice b utilisée pour le calcul des paramètres cinétiques par KRR,
3. Calcul de la région de confiance (définie comme l'aire comprise entre la borne supérieure et inférieure de l'intervalle de confiance) en chacun des points du domaine par bootstrap fonctionnel

4. Définition du nouveau point d'expérience là où l'aire de cette région de confiance est maximale,
5. Réalisation de l'expérience et retour à l'étape d'estimation de la matrice b

Cette méthodologie complète doit maintenant être éprouvée en l'appliquant pour de futurs essais et d'autres cas d'étude proches, correspondant à de nombreux cas rencontrés en pratiques.

Plans d'expériences exploratoires

5.1	Motivations	151
5.2	Plans exploratoires et critères de comparaison de plans	153
5.2.1	Plans maximin et minimax	153
5.2.2	Plans à entropie maximale	155
5.2.3	Critères de comparaison de plans d'expériences	157
5.3	Génération des plans exploratoires et comparaison	159
5.3.1	Plan à entropie maximale et maximin	159
5.3.2	Génération du plan d'expériences final	163
5.4	Ouvertures/Perspectives	165

Plans d'expériences exploratoires

Le travail de ma thèse est de déterminer le meilleur métamodèle représentant le comportement du système physico-chimique étudié. Deux grands types de modèles ont été explorés, les premiers basés sur des modèles de krigeage (cf. chapitre 3), les seconds sur des modèles à coefficients variables (cf. chapitre 4).

L'objectif de la thèse étant de comparer différents métamodèles, les points expérimentaux doivent être bien répartis dans l'espace (Morris et Mitchell [117]). Or, le plan initial à disposition au début de la thèse présente une mauvaise répartition. Donc une des tâches de la thèse a été de refaire l'expérimentation afin de pouvoir comparer sereinement les métamodèles envisagés.

Ce chapitre présente dans un premier temps les motivations qui nous ont conduits à définir un plan exploratoire, puis dans un second temps deux types de plans que nous avons confrontés pour cet objectif. Premièrement, les plans maximin (Johnson *et al.*, [90]) et, deuxièmement, les plans à entropie maximale (Currin *et al.* [37]). Ensuite, nous présentons plus en détail l'application et les résultats du plan retenu.

Les critères de plans d'expériences classiques basés sur les métamodèles développés dans cette thèse, sont des plans itératifs. Le principe étant de rajouter un à un des points du plan en se basant sur l'incertitude des prédictions obtenues. Cette incertitude est obtenue au travers de la variance de prédiction. Pour les modèles non fonctionnels, la variance de prédiction est donnée par une formule analytique (un des avantages du krigeage). Pour les méthodes fonctionnelles, nous avons mis en place un bootstrap fonctionnel afin d'estimer les régions de confiance. Dans les deux cas, l'idée du plan itératif envisagée, est de placer les points un à un aux points du domaine où l'incertitude (l'aire de l'intervalle ou région de confiance selon le cas) est la plus forte.

5.1 Motivations

Le modèle numérique est inadéquat pour représenter de manière efficace la réalité expérimentale. Cette configuration n'est pas le reflet d'un cas isolé, puisqu'il est de plus en plus fréquent que les modèles postulés ne conviennent que partiellement à représenter le comportement d'un système (Canaud *et al.* [21], Ayed [8], Naar [120]), et ce pour deux raisons principales. D'une part, des systèmes de plus en plus complexes, d'autre part, des précisions attendues de plus en plus fines.

Nos travaux se sont donc orientés vers la détermination du métamodèle le plus pertinent. Malgré les simplifications du modèle postulées dans le chapitre 1, cette étape de modélisation est loin d'être triviale. Les différentes approches qu'elles soient non fonctionnelles (chapitre 3) ou fonctionnelles (chapitre 4), nécessitent un certain nombre d'expériences pour la calibration des paramètres des métamodèles. Il est donc nécessaire de mettre en place un premier plan d'expériences explorant le domaine le plus efficacement possible.

Pour le moment, ces expériences sont réalisées à partir de connaissances métiers, et donc, le plus souvent, sans stratégie pertinente. Or, le nombre croissant de paramètres à ajuster, sur des domaines de plus en plus importants, ne permet plus de se satisfaire de cette façon de procéder. Pour illustrer cela, nous présentons le

plan initial à disposition. Ce premier plan est issu de travaux de thèse menés sur le fonctionnement du piège à NOx, par Chedotal, [26] et présente une répartition dans l'espace des entrées, insuffisante, engendrant des problèmes d'optimisation numériques et des prédictions aberrantes.

Les figures 5.1 montrent la répartition du plan d'expériences initial sur le domaine. Nous avons étudié la répartition du plan pour chaque variable indépendamment des autres variables puis deux à deux. On code pour chacune des variables espèces par 1 la valeur maximum, et par 0 la valeur minimum.

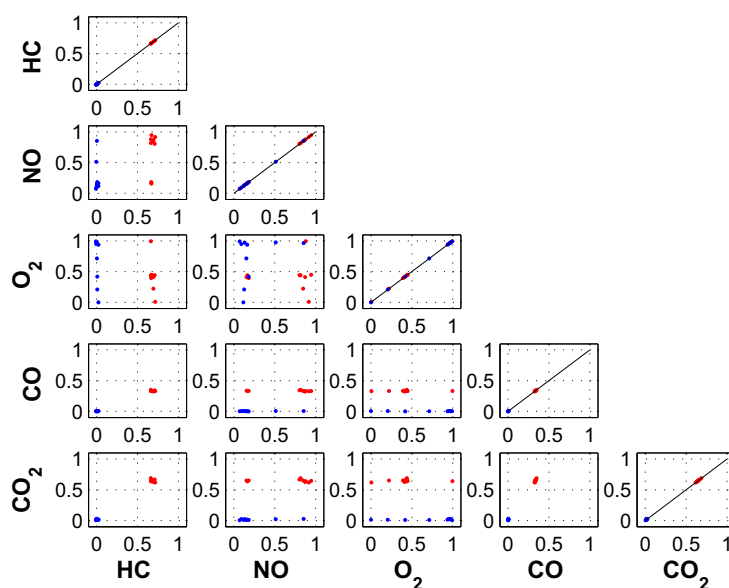


Figure 5.1 – Représentation du plan d'expériences dans l'espace des entrées deux à deux

Cette figure nous permet de distinguer plusieurs choses. Tout d'abord, on remarque que pour les variables O_2 et NO , le domaine est relativement bien couvert. En revanche, pour les trois autres espèces, HC , CO et CO_2 , seuls deux points du domaine sont explorés. De plus, pour le CO , ces deux points sont contenus dans une seule moitié du domaine (entre 0 et 0.5). Ce constat est d'autant plus dramatique que nous avons, en accord avec les expérimentateurs, déterminé que les variables prépondérantes gouvernant le système sont les espèces HC , CO et O_2 . Or, la répartition des points dans l'espace CO/HC montre uniquement deux points explorés. Nous avons ainsi défini deux types de jeux de données, symbolisés par les points rouges et bleus.

Les métamodèles explorés dans nos travaux ne peuvent être performants en prédiction, avec l'utilisation d'un tel type de plan d'expériences. On sait, par exemple, que pour les modèles de krigeage sur lesquels sont basées nos premières approches, la variance de krigeage grandit quand on s'éloigne des points, il est donc souhaitable que ceux-ci soient éparpillés (Vazquez [156]).

Nous avons alors décidé de mener une campagne expérimentale au cours de la thèse, afin de répondre à cette problématique de bonne répartition des points dans l'espace et surtout pour pouvoir baser notre étude comparative des métamodèles sur un plan d'expériences propre.

L'objectif de cette seconde campagne est double. D'une part, le but est de fournir des jeux de données de meilleure qualité pour l'estimation des paramètres cinétiques et des métamodèles. D'autre part, la méthode de planification mise en place, doit pouvoir être un premier outil pertinent pour les expérimentateurs.

5.2 Plans exploratoires et critères de comparaison de plans

Comment choisir un bon plan d'expériences? Tout dépend de l'objectif visé par l'étude expérimentale. Par exemple pour des études d'impact ou de sélection des variables, les plans factoriels à deux niveaux, sont préconisés, pour des études de surface de réponse type régression linéaire, on a recours aux plans optimaux (plus particulièrement D-optimaux), et dans le cas non-linéaire, aux plans D-optimaux non linéaires. Dans notre cas, les modèles visés sont des modèles de type splines, ce qui nécessite d'avoir des points dans tout le domaine.

Les plans remplissant l'espace, appelés plans "space-filling", ont pour objectif de répartir les points aussi uniformément que possible dans la région expérimentale et sont construits sans faire d'hypothèse sur le modèle reliant les facteurs et les réponses. Ils peuvent être générés aléatoirement, mais nous évitons cette approche car dans le cas de petits échantillons en grande dimension, les points générés aléatoirement ont tendance à se concentrer à certains endroits, phénomène appelé clustering (Santner *et al.*, [138]).

Parmi tous les plans envisageables, nous avons fait le choix d'en retenir deux. Le premier est un plan, basé sur un critère usuellement utilisé, le critère maximin (Johnson *et al.*, [90]). Le second est peu présent dans la littérature, mais est particulièrement indiqué lorsque la surface de réponse est ajustée par krigeage, ce qui constitue l'idée de base des approches non fonctionnelles explorées. L'avantage pour notre étude de ce critère, est d'obtenir un plan tel que l'information apportée soit maximale pour des méthodes telles que celles présentées dans le chapitre 3. L'idée étant d'avoir une base d'expériences, renseignant au maximum le complément de modèle. Pour cela, la méthode consiste à maximiser le déterminant de la matrice de corrélation.

Nous conseillons la lecture des travaux de Koehler et Owen [91] et Franco [64] pour un état de l'art concernant les plans space filling.

La question de la comparaison des plans a donc été abordée. Pour cela, trois critères sont utilisés, dont l'objectif est de juger du bon remplissage de l'espace ainsi que de la distribution uniforme. Ces trois critères sont la mesure de recouvrement (Gunzburger *et al.*, [75]), le critère maximin (Johnson *et al.*, [90]) et la discrédance en norme L_2 -étoile (Fang *et al.*, [57]). Nous citons le package DiceDesign de R qui permet l'évaluation de ces critères et la génération des plans à entropie maximale (Dupuy *et al.* [42]).

Cette partie est donc construite en trois sections. Premièrement nous introduisons les plans maximin/minimax, deuxièmement les plans à entropie maximale, troisièmement, les critères de comparaison de plans d'expériences.

5.2.1 Plans maximin et minimax

Les plans basés sur un critère de distance sont relativement courants. Ce critère permet de quantifier la répartition d'un ensemble de points. Parmi eux, les plus célèbres sont les plans maximin et minimax introduits par Johnson *et al.*, [90].

Introduisons, tout d'abord, quelques notations : soit D le domaine expérimental, et $S \subset D \subset \mathbb{R}^d$ un plan consistant en n points distincts (x_1, \dots, x_n) . On se donne une distance ρ sur D associée à la norme L_p :

$$\rho_p(w, x) = \left(\sum_{i=1}^d \|w_i - x_i\|^p \right)^{\frac{1}{p}} \quad (5.1)$$

D'autres critères de distance existent, notamment le critère de distance moyenne, pour la présentation duquel nous renvoyons le lecteur à l'ouvrage de Santner *et al.*, [138].

Plans maximin

Pour les plans maximin, l'approche est la suivante : on considère que les points d'un plan sont "bien répartis", si les points sont aussi éloignés les uns des autres que possible.

Une façon de quantifier la proximité des points de l'ensemble S , est de prendre la plus petite distance entre deux points quelconques de S , soit

$$\min_{w,x \in S} \rho_p(w,x). \quad (5.2)$$

Un plan qui maximise cette quantité est appelé plan de distance maximin, et est noté S^* , et

$$\min_{w,x \in S^*} \rho_p(w,x) = \max_{S \subset D} \min_{w,x \in S} \rho_p(w,x). \quad (5.3)$$

Ce type de plan garantit que deux points quelconques ne sont pas trop proches, et donc que les points du plan sont bien répartis dans D .

Typiquement, ces plans ont tendance à contenir des points sur le bord du domaine expérimental. Pour des illustrations de ce phénomène, nous renvoyons le lecteur à l'article de base de Johnson *et al.*, [90].

Une façon pratique de construire un plan maximin approché est de se focaliser sur les hypercubes latins. Chaque arête du domaine expérimental est divisé en n segments de même longueur de façon à obtenir un maillage du domaine. Un hypercube latin sélectionne alors n points parmi les n^d points de la grille de façon à ce que les n niveaux des variables d'entrée soient testés une fois par les simulations. Les hypercubes latins présentent beaucoup d'avantages :

1. Ils sont simples à construire. En effet, chaque colonne d'un hypercube est une permutation de $\{1, \dots, n\}$ ou de n symboles quelconques.
2. Les points sont uniformément distribués sur chaque axe du domaine.

La distribution uniforme sur chaque axe n'assure pas l'uniformité sur le domaine expérimental. Cependant, pour n fixé, ils existent $(n!)^d$ hypercubes latins possibles. La démarche est donc la suivante : on part de l'ensemble des hypercubes latins à n points et on choisit celui (ou un de ceux) qui satisfait le critère maximin, appelé hypercube latin maximin. Park [126] propose également un algorithme d'échange pour déterminer un hypercube optimal (localement) pour un critère donné.

Enfin, il est intéressant de préciser que le problème de construction d'un plan maximin à n points dans $[0, 1]^d$ est équivalent au problème de placer dans $[0, 1]^d$ les centres de n sphères, disjointes ou d'intersection égale à un singleton, de même rayon r , de façon à ce que r soit le plus grand possible (Bettinger, [12]).

Cette vision permet de faire le lien avec les plans de Strauss (Strauss [150], Ripley et Kelly [134], Franco [64]). Ces derniers sont similaires aux plans maximin, puisqu'ils reposent sur le processus ponctuel de Strauss, et dont le principe est de remplir l'espace de particules (points dans le domaine représentant les expériences) en minimisant les interactions entre elles, chacune ayant une sphère d'influence de rayon $r/2$.

Cependant, ces plans nécessitent la mise en place de méthodes de Monte-Carlo par chaîne de Markov, qui peuvent être coûteuses en temps de calcul. De plus, l'intérêt principal de ces plans est de pouvoir incorporer une connaissance *a priori*, par exemple sur les liens entre les entrées ou sur des zones de densité plus forte, connaissance dont nous ne disposons pas pour notre cas.

Plans minimax

Pour les plans minimax, l'approche est différente : on considère qu'un plan S est bien réparti dans D , si tout point de l'espace D est suffisamment proche d'un point de S .

Formellement, définissons la distance entre un point quelconque $x \in D$, et le plan $S \subset D$, par :

$$\rho_p(x, S) = \min_{x_i \in S} \rho_p(x, x_i). \quad (5.4)$$

Un plan est appelé plan de distance minimax, et noté S^* , si la distance maximale, entre un point $x \in D$ et le plan S^* , est minimale parmi tous les plans (à n points) $S \subset D$:

$$\max_{x \in D} \rho_p(x, S^*) = \min_{S \subset D} \max_{x \in D} \rho_p(x, S). \quad (5.5)$$

L'inconvénient de ce type de plan est qu'ils sont difficiles à générer. En effet, la fonction utilisée pour la construction de ces plans, utilise non seulement l'ensemble des points candidats pour le plan S , mais aussi l'ensemble des points de tout le domaine expérimental D .

De plus, les points sont rarement aux bords du domaine, ce qui peut amener à faire de l'extrapolation, bien moins performante que l'interpolation quand on utilise le krigeage. En effet, le krigeage (voir chapitre 2) donne de bons résultats de prédiction, à condition de se situer à l'intérieur de l'enveloppe convexe des points du plan (Vazquez, [156]).

Dans ce second type de plan, le problème de construction d'un plan minimax à n points dans $[0, 1]^d$ est équivalent au problème de placer dans $[0, 1]^d$ les centres de n sphères de même rayon r qui recouvrent D , de façon à ce que r soit le plus petit possible (Bettinger, [12]).

5.2.2 Plans à entropie maximale

Boltzmann a été le premier à formaliser de manière probabiliste l'entropie qui fut utilisée par Shannon en 1948, [142], dans le contexte de la théorie de l'information. Celui-là même a défini, dans un cadre probabiliste, une notion d'entropie, qui a des interprétations diverses en informatique, en thermodynamique et en probabilités, pourtant toutes liées par des intuitions communes. Shewry et Wynn, [144], l'ont décrit comme étant "la quantité d'information dans une expérience". Plus généralement, l'entropie mesure la quantité d'information contenue dans une distribution.

Par exemple, si l'on s'intéresse à la quantité d'incertitude liée à un événement aléatoire, prenons un événement dont la loi est uniforme, alors cette incertitude est la plus grande, son entropie est donc maximale. À l'inverse, si la densité de probabilités est très concentrée, l'incertitude est toute petite, et l'entropie minimale.

Afin de bien comprendre cette notion d'un point de vue mathématique, considérons une variable aléatoire discrète X dont la distribution est donnée par $P(X = x_i) = p_i$.

L'entropie de X est alors définie par (Shannon, [142]) :

$$H(X) = - \sum_{i=1}^n p_i \ln(p_i). \quad (5.6)$$

où $H(X)$ correspond à l'entropie associée à la variable (ou la distribution) X .

Dans le cas d'une distribution régulière des valeurs de X i.e $p_i = 1/n \forall i = 1, \dots, n$, alors cette dernière est la plus imprévisible possible et $H(X) = \ln(n)$ atteint sa valeur maximale. À l'inverse, si X prend une seule valeur x_0 , i.e. $P(X = x_0) = 1$, alors la valeur de X est complètement prévisible et l'entropie associée est minimale, $H(X) = 0$.

Lindley, [96], utilise par la suite cette notion de quantité d'information introduite par Shannon, afin de mesurer l'information apportée par une expérience. Cette vision lui permet de définir le concept de gain d'une expérience. L'article de Shewry et Wynn [144] consiste justement à établir le lien entre gain et entropie.

Construire un plan à entropie maximale a donc pour but de maximiser la quantité d'information apportée par les expériences. On voit donc que la définition générale de ces plans ne fait pas de ce critère un critère intrinsèquement space-filling, puisqu'il nécessite d'avoir la réponse aux points d'expériences.

Cependant, en s'intéressant au calcul de l'entropie de la loi d'un processus gaussien, Shewry et Wynn, [144], démontrent (voir également Koehler et Owen, [91]) que $H(X)$ dépend simplement de $\ln(\det(C(X)))$ où $C(X)$ est la matrice de corrélation associée au plan. Par conséquent, générer un plan à entropie maximale revient à maximiser le déterminant de la matrice de corrélation. Les auteurs précisent toutefois que cette équivalence existe uniquement si les réponses aux points suivent une loi normale multidimensionnelle, mais sans aucune condition sur la structure de covariance utilisée.

En revanche, le choix de la portée est déterminant. Nous avons opté, en se basant sur les noyaux les plus courants de la littérature, pour un noyau gaussien. Concernant le choix de la portée, nous avons testé en pratique plusieurs valeurs (voir section 5.3.1), et étudié les valeurs des critères de comparaison obtenues dans chacun des cas.

Algorithme 5.1 Génération d'un plan à entropie maximale

Initialisation : choisir un vecteur de n points $X_{(0)}$ au hasard et fixer un variogramme

Calcul du déterminant de la matrice de corrélation spatiale associée $\Rightarrow \det(C(X_{(0)}))$

Pour $k = 1$ à $Nmax$

1. Choisir une expérience i au hasard dans $\{1, \dots, n\}$
2. Simuler z^i uniformément dans $[0, 1]^d$
3. Accepter de remplacer x^i par z^i si $\det(C(X_{(k)})) > \det(C(X_{(k-1)}))$

Fin pour

La génération d'un plan à entropie maximale consiste donc à maximiser le déterminant de $C(X)$ par un algorithme d'échange du type Fedorov (Fedorov [59]), ou de Mitchell (Mitchell [113]). Le pseudo-code 5.1 présente l'algorithme classique dans le cas de plan d'expériences numériques. La valeur $Nmax$ est une valeur choisie arbitrairement. Il est à noter que plutôt que de fixer un nombre $Nmax$ de boucles, nous aurions pu mettre un critère d'arrêt plus pertinent, basé sur les valeurs successives du déterminant. Cependant, empiriquement, on constate qu'après quelques itérations (cinq dans notre cas), les points du plan ne sont plus modifiés.

Ce premier code général a été quelque peu modifié pour les besoins de notre cas, et d'une manière générale, pour s'adapter à différentes applications. Tout d'abord, lorsque des essais ont déjà été effectués (m essais), ces points sont inclus dans le plan d'expériences, et ne peuvent pas être modifiés. Ensuite, plutôt que de simuler uniformément z^i , le nouveau point à tester, nous balayons tous les points du domaine, et ne gardons que celui dont le déterminant de la matrice de corrélation est maximal. Et enfin, nous balayons également tous les points du plan initialement déterminé, l'ordre étant aléatoire, et nous bouclons $Nmax$ fois cette démarche. Le pseudo-code 5.2 présente l'algorithme mis en place et utilisé pour générer les plans à entropie maximale.

Algorithme 5.2 Génération d'un plan à entropie maximale

Initialisation : Rajouter aux m expériences disponibles un vecteur de n points aléatoirement dans le domaine $D \Rightarrow X_{(0)}$

Calcul du déterminant de la matrice de corrélation spatiale associée : $\det(C(X_{(0)}))$

Pour $k = 1$ à $Nmax$

1. Créer le vecteur V des indices des points modifiables dans $X_{(0)}$
2. **Pour** $i = 1$ à n
 - (a) Extraction d'un indice, ind , parmi V , aléatoirement
 - (b) **Pour** $j = 1$ à $Card(D)$
 - i. Calcul de $\det(C(X_{(j)}))$ où x_{ind} est remplacé par d_j
 - (c) **Fin pour**
 - (d) Calcul de max_j tel que $\det(C(X_{(j)}))$ est maximum pour $j = max_j \Rightarrow \det(C(X_{(i)}))$
 - (e) Accepter de remplacer x_{ind} par d_{max_j} si $\det(C(X_{(i)})) > \det(C(X_{(i-1)}))$
3. **Fin pour**

Fin pour

Cet algorithme génère des plans ayant de bonnes propriétés de remplissage, et permet de s'adapter à des cas où un certain nombre d'essais a déjà été effectué ou au cas où certains essais sont imposés, comme cela est souvent le cas en cinétique. Il s'applique également dans les cas où aucune contrainte d'essais n'est posée, et pour des domaines aux formes géométriques particulières (toutefois dans des limites raisonnables), autrement dit, l'algorithme reste valable si le domaine, pour des raisons expérimentales, est non hypercubique.

5.2.3 Critères de comparaison de plans d'expériences

Afin de comparer les plans générés et faire notre choix sur les expériences à réaliser pour la campagne, il est nécessaire de faire appel à des critères usuels permettant de juger du bon remplissage de l'espace ainsi que de la distribution uniforme.

La qualité de la répartition spatiale est mesurée soit à l'aide de critères déterministes comme les distances minimax ou maximin (Johnson *et al.*, [90]), soit à l'aide de critères statistiques comme la discrédance (Niederreiter [124], Hickernell [84] et Fang *et al.* [57]). Les trois critères retenus sont la mesure de recouvrement, le critère maximin et la discrédance en norme L_2 -étoile.

Mesure de recouvrement : Rec

La mesure de recouvrement permet de mesurer l'écart entre les points du plan et ceux d'une grille régulière (Gunzburger *et al.*, [75]), et est définie par

$$Rec = \frac{1}{\gamma} \left(\frac{1}{n} \sum_{i=1}^n (\gamma_i - \bar{\gamma})^2 \right)^{1/2}, \text{ avec } \gamma_i = \min_{k \neq i} \rho_2(x_i, x_k) \text{ et } \bar{\gamma} = \frac{1}{n} \sum_{i=1}^n \gamma_i \quad (5.7)$$

où ρ_2 est la distance euclidienne, définie par l'équation 5.1 avec $p = 2$. γ_i désigne donc la distance (euclidienne) minimale entre le point x_i et les autres points du plan et $\bar{\gamma}$, la moyenne des γ_i . Cette expression fait apparaître explicitement le coefficient de variation de l'échantillon des γ_i , c'est-à-dire le rapport de l'écart-type avec la moyenne.

Ce critère est nul pour une grille régulière ($\gamma_1 = \dots = \gamma_n = \bar{\gamma}$). L'objectif est donc de le minimiser pour se rapprocher d'une grille régulière, et ainsi assurer le remplissage de l'espace, sans toutefois l'atteindre pour respecter une distribution uniforme notamment en projection sur les axes.

Critère maximin : *Mindist*

Johnson *et al.*, [90], ont introduit les distances maximin et minimax afin de construire des plans répondant à la question de remplissage de l'espace. Le critère maximin consiste à maximiser la distance minimale entre deux points du plan.

Nous renvoyons le lecteur à la section 5.2.1 pour plus de détails.

Discrédance en norme L_2 -étoile : *DisL2*

Avant de présenter le critère utilisé, nous introduisons la notion de discrédance. La discrédance est une mesure de l'écart existant entre une situation de référence (l'uniformité théorique) et une configuration donnée. Soient n points $x = (x_1, \dots, x_n)$ dans $[0, 1]^d$, $x_i = (x_{i1}, \dots, x_{id})$ pour $i = 1, \dots, n$ et $t = (t_1, \dots, t_s) \in [0, 1]^d$. La discrédance locale, notée *disc*, est définie par

$$disc(t) = \frac{A(t, x)}{n} - \prod_{j=1}^d t_j, \quad (5.8)$$

où $A(t, x)$ est le nombre de points de x qui appartiennent au pavé $P = [0, t_1[\times \dots \times [0, t_d[$.

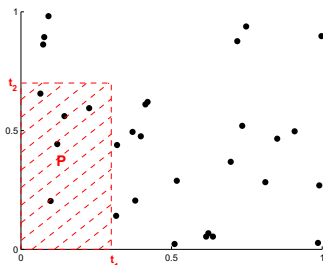


Figure 5.2 – Représentation d'une suite de $n=30$ points en dimension $d=2$ et du pavé $P = [0, t_1[\times [0, t_2[$

La figure 5.2 nous permet d'illustrer la notion de discrédance en dimension 2. Pour que la répartition de notre ensemble x se rapproche de l'uniformité, on s'attend à ce que la proportion de points situés dans le pavé P soit proche de son aire $\lambda(P) = \prod_{j=1}^d t_j$. Par conséquent, il semble naturel de juger peu uniforme une suite finie pour laquelle il est possible d'exhiber un pavé avec une grande différence entre les deux, c'est-à-dire une importante discrédance locale. Cette idée exprime la notion d'équirépartition.

Une suite est équirépartie lorsque sa fonction de répartition empirique converge vers la fonction de répartition uniforme. On voit ainsi qu'une suite de variables aléatoires uniformes indépendantes est équirépartie. La définition de discrédance découle directement de cette notion.

Soient F_n , la fonction de répartition empirique des points du plan et U celle de la loi uniforme, on définit la discrédance par,

$$\sup_{t \in [0, 1]^d} |F_n(t) - U(t)| \text{ avec } F_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{x_i \leq t} \quad (5.9)$$

Contrairement aux deux critères précédents, la discrédance n'est pas basée sur la distance entre les points. Il existe différentes mesures de discrédance (Fang *et al.*, [57] et Thiemard [152]). Nous retenons la discrédance (extrême) en norme L_2 -étoile, appelée aussi discrédance carrée moyenne.

En suivant la même approche que Warnock (Warnock [159]), Morokoff and Caflish [116] définissent la formule analytique de la discrédance extrême en norme L_2 -étoile, que nous utilisons,

$$DisL2^2 = 12^{-d} - \frac{2^{1-d}}{n} \sum_{k=1}^n \prod_{l=1}^d (1 - x_{kl}^2) + \frac{1}{n^2} \sum_{k=1}^n \sum_{j=1}^n \prod_{i=1}^d [1 - \max(x_{ki}, x_{ji})] \min(x_{ki}, x_{ji}), \quad (5.10)$$

où $x_k = (x_{k1}, \dots, x_{kd})$, $k = 1, \dots, n$, avec n le nombre de points de dimension d du plan.

5.3 Génération des plans exploratoires et comparaison

Ce chapitre s'adressant également aux chimistes et dans un souci pédagogique, nous détaillons de manière précise la démarche mise en place au cours de la campagne, les obstacles pratiques rencontrés, les solutions apportées, et les avantages de notre procédure.

Notre démarche s'est faite en deux étapes. Tout d'abord, nous avons voulu rajouter vingt points expérimentaux aux vingt expériences déjà disponibles. Cependant, nous avons rapidement constaté que l'échantillon de catalyseur utilisé précédemment, avait évolué et que son comportement ne correspondait plus à ce qui était observé. Nous avons alors décidé toujours en accord avec les expérimentateurs d'avoir recours à un nouvel échantillon, et donc d'oublier les jeux initiaux.

Nous détaillons chacune de ces étapes. Ainsi, nous présentons et comparons, tout d'abord, les plans maximin et à entropie maximale générés tenant compte des vingt expériences initiales. Nous montrons, dans l'annexe C, comment nous avons observé le changement de comportement de la carotte. Enfin, nous exhibons le plan final qui a été mené et que nous avons utilisé tout au long de ce manuscrit pour illustrer les résultats.

La première étape a donc pour objectif d'expliquer la démarche lorsque l'on a des essais déjà effectués, la seconde de montrer le changement de comportement de la carotte et ses implications, et la troisième étape, la méthode de génération du plan lorsqu'aucun essai n'a été réalisé.

5.3.1 Plan à entropie maximale et maximin

Considérations pratiques et définition du domaine

Pour les deux plans, quelques considérations pratiques ont été faites, principalement motivées par le nombre restreint d'expériences réalisables pendant cette campagne expérimentale.

Le plan initial contient majoritairement des essais ayant la même valeur de débit massique. Bien que cette variable ait un effet sur la conversion des polluants, il a été choisi de ne pas la considérer afin de simplifier le système. Il est donc, comme la température initiale du gaz et le profil de température (voir chapitre 1) fixé à une valeur identique pour tous les essais (0.948 kg/h).

Ensuite, comme nous nous plaçons en fonctionnement en mélange pauvre, dans lequel seules les sorties HC et CO sont analysées, et après discussion avec les expérimentateurs, les entrées prépondérantes de notre système qui nous intéressent sont la composition massique des HC et du CO , les effluents du moteur, auxquelles nous

rajoutons la composition massique du O_2 (voir chapitre 1). Par ailleurs, les compositions de CO_2 et NO , ont été prises égales à la valeur médiane de leurs plages de variations respectives (voir tableau 1.2), soit 8% et 525 ppm.

Notre domaine est donc de dimension $d = 3$. Pour chaque direction, les plages sont discrétisées en $n_d = 11$ points, ce qui correspond aux limites de précision expérimentale (précision des débitmètres contrôlant l'entrée de chaque espèce). On a donc au final $\text{Card}(D) = 11^3 - 20 = 1311$, dont on souhaite extraire un plan d'expériences de taille $n = 20$.

Plan maximin

La méthode de construction des plans maximin a été donnée dans la section 5.2.1. Johnson *et al.* [90] précisent que le choix optimal du plan, sous le critère de maximin, réside dans l'utilisation d'un hypercube latin. Cependant, comme nous partons d'une base de vingt expériences, la construction de ce premier plan est différente et consiste à rajouter un à un, vingt points appartenant au domaine, tels qu'à chaque itération la distance maximale soit minimale entre les points du plan et le nouveau point. Le pseudo-code 3 présente la méthode mise en place.

Algorithme 5.3 Génération d'un plan maximin

Initialisation : On dispose d'un premier plan $S \subset D$

Pour $k = 1$ à n

1. Pour chaque point du plan S on calcule la distance maximale avec un point du domaine $D \Rightarrow V$
2. On récupère la valeur minimale de V , qui correspond au point x^* du domaine D tel que la distance maximale entre ce point et le plan S soit minimale
3. On rajoute le x^* au plan S

Fin pour

La figure 5.3 illustre la répartition dans l'espace des entrées de ce premier plan généré par un critère maximin.

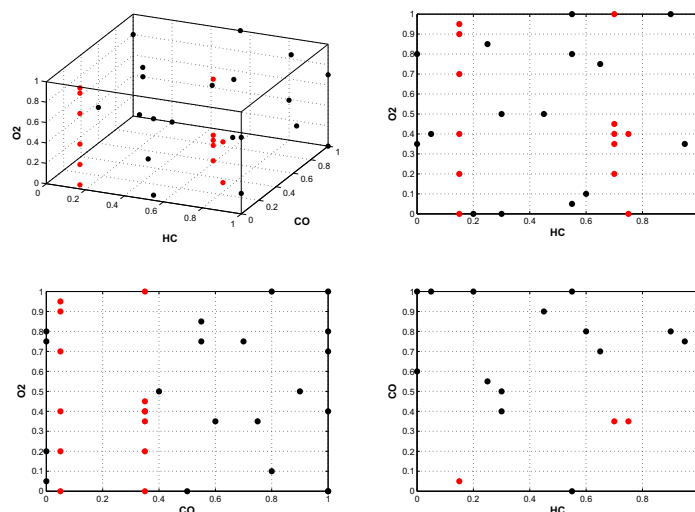


Figure 5.3 – Représentation du plan d'expériences généré selon un critère maximin. Les points rouges correspondent au plan initial, et les points noirs au plan généré

Les valeurs des trois mesures utilisées pour la comparaison des plans sont données dans le tableau 5.1. On note Rec la valeur de la mesure de recouvrement, $Mindist$ la valeur du critère maximin et $DisL2$ la valeur de la discrédance en norme L_2 -étoile.

Rec	$Mindist$	$DisL2$
0.4817	0.3674	0.0143

Tableau 5.1 – Critères pour la comparaison des plans pour le plan maximin généré

Plan à entropie maximale

La section 5.2.2 présente le principe et le pseudo-code pour la génération des plans à entropie maximale. Il est notamment précisé que le choix du noyau n'est pas essentiel. Nous avons opté arbitrairement pour un noyau gaussien.

Nous avons également étudié l'influence de la portée. Pour cela, nous avons testé un grand nombre de valeurs de portée et ne retenons que la meilleure au sens des critères de comparaison $DisL2$ et Rec . Une portée de 0.1, la taille de discrétisation choisie selon chaque axe, semble être le meilleur choix. La figure 5.4 représente les valeurs du critère $DisL2$ et Rec en fonction de la portée.

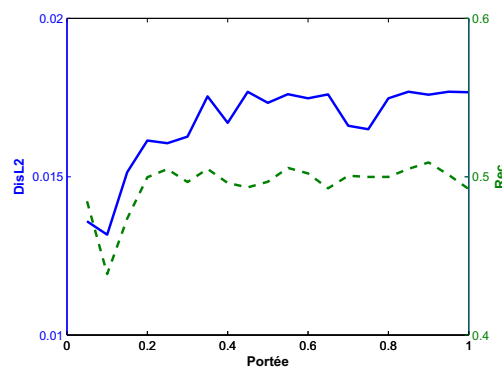


Figure 5.4 – Critère $DisL2$ (courbe pleine bleue) et Rec (courbe pointillée verte) en fonction de la portée

La figure 5.5 illustre la répartition dans l'espace des entrées de ce second plan généré par un critère d'entropie maximale.

Les valeurs des trois mesures utilisées pour la comparaison des plans sont données dans le tableau 5.2 (Rec = mesure de recouvrement, $Mindist$ = critère maximin, $DisL2$ = discrédance en norme L_2 -étoile).

Rec	$Mindist$	$DisL2$
0.4567	0.255	0,0137

Tableau 5.2 – Critères pour la comparaison des plans pour le plan à entropie maximale généré

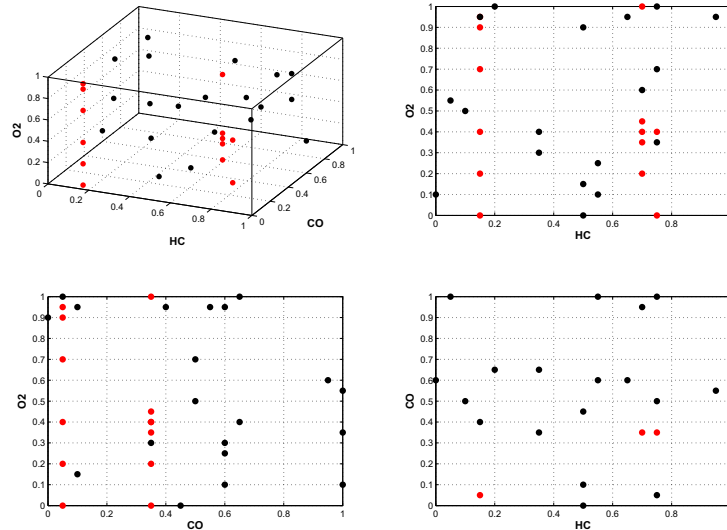


Figure 5.5 – Représentation du plan d’expériences généré par maximum d’entropie. Les points rouges correspondent au plan initial, et les points noirs au plan généré

Comparaison et conclusion

Après avoir généré ces deux plans, il a fallu déterminer celui que nous souhaitons mettre en place. Pour cela, nous avons basé notre choix selon trois critères usuels de comparaison de plan et également sur la représentation visuelle du plan puisque cette dernière est disponible. On précise également que les caractéristiques typiques de ces plans sont entrées en considérations dans notre décision.

Tout d’abord, étudions la valeur des trois critères retenus pour chaque plan. On rappelle que ces trois critères sont la mesure de recouvrement, Rec , qui permet de mesurer l’écart entre les points du plan et ceux d’une grille régulière, le critère maximin, $Mindist$ qui consiste à calculer parmi les distances minimales entre les points du plan, celle qui est la plus faible, et enfin la discrédance en norme L_2 -étoile, $DisL2$, qui mesure l’écart entre la fonction de répartition empirique des points du plan et celle de la loi uniforme.

Pour ce qui est du critère $DisL2$ et Rec , on souhaite que la valeur soit la plus petite possible. En revanche, pour $Mindist$, le critère maximin consistant à maximiser la distance minimale entre deux points du plan, l’objectif est d’obtenir la valeur la plus grande possible.

Le plan à entropie maximale montre des valeurs de critères, toutes trois inférieures à celles obtenues pour le plan maximin. Autrement dit, au sens du critère de recouvrement et de la discrédance, le plan à entropie maximale présente une meilleure répartition. En revanche, pour le critère $Mindist$, le plan maximin est meilleur. Ce constat est totalement normal puisque le principe même de ce plan est de maximiser la valeur de ce critère.

Les figures 5.3 et 5.5 nous permettent, dans un second temps, de comparer la répartition des plans dans le domaine de manière visuelle, et ainsi confirmer l’information apportée par les critères.

D’une manière générale, la répartition du plan à entropie maximale, figure 5.5, est meilleure. Pour s’en convaincre, il est nécessaire de s’attarder sur la figure présentant la répartition dans l’espace CO/HC (figure en bas à droite).

Bien que des zones, pour le plan à entropie maximale restent inexplorées, ces zones sont moins importantes et moins nombreuses que pour le plan maximin. De plus, ce dernier place un grand nombre de points aux bords du domaine et notamment aux coins. Cela se traduit entre autre, par un point du domaine dont la valeur de richesse est très proche de 0.9 (0.898).

Cette répartition n'est pas étonnante puisque, comme nous l'avons introduit dans la section 5.2.1, les plans maximin ont tendance à explorer les bords du domaine, ce qui, rappelons-le, n'est pas souhaitable pour les cas rencontrés par les expérimentateurs.

En revanche, l'utilisation du critère du maximum d'entropie est particulièrement indiquée si l'on se sert d'une modélisation par krigeage, ce qui, si l'on replace la campagne expérimentale dans le contexte de la thèse, est le principe des méthodes envisagées.

En conclusion, l'utilisation des plans à entropie maximale semble plus pertinente et plus adaptée aux cas rencontrés en pratique par les expérimentateurs. Le plan à entropie maximale fournit un premier outil facilement utilisable, pouvant inclure des contraintes (sur le domaine ou d'essais), et donnant des plans de bonne qualité.

Nous ne rentrons pas plus dans les détails concernant la composition de ce plan, puisque comme nous le montrons dans l'annexe C, des considérations expérimentales pratiques, nous ont obligés à ne plus tenir compte des expériences initiales et redéfinir un nouveau plan.

5.3.2 Génération du plan d'expériences final

Nous présentons dans cette section le nouveau plan d'expériences à effectuer, basé sur le critère du maximum d'entropie, sur un nouvel échantillon de catalyseur.

Pour déterminer, ce nouveau plan, aucun point du domaine n'a été imposé. En effet, aucune expérience n'a été réalisée au préalable sur cette carotte, et les bornes expérimentales étant choisies arbitrairement, il n'apparaît pas obligatoire de prendre les coins du domaine ou le point central dans la matrice d'expériences, comme cela est souvent le cas en cinétique.

n°	1	2	3	4	5	6	7	8	9	10
<i>HC</i> (ppmC)	3380	2140	3380	900	900	1520	2760	3380	3690	1520
<i>CO</i> (ppm)	4410	1230	4940	1230	3350	1760	2290	1230	2820	6000
<i>O₂</i> (%)	16.50	6.52	10.80	3.67	16.50	15.07	15.07	7.94	13.65	13.65
richesse	0.41	0.63	0.53	0.75	0.40	0.41	0.43	0.60	0.46	0.46
n°	11	12	13	14	15	16	17	18	19	20
<i>HC</i> (ppmC)	2760	900	1210	3070	2140	2140	4000	3070	4000	2140
<i>CO</i> (ppm)	700	2290	4940	4410	3350	4940	3350	2820	4940	2290
<i>O₂</i> (%)	10.80	12.22	5.09	6.52	12.22	16.50	7.94	5.09	3.67	2.24
richesse	0.51	0.47	0.70	0.66	0.48	0.41	0.61	0.71	0.81	0.86

Tableau 5.3 – Numérotation, composition de *HC*, *CO* et *O₂* et richesse des 20 essais réalisés sur BGS

Le tableau 5.3 présente la nouvelle matrice d'expériences. On peut voir que deux expériences ont une richesse élevée, les expériences 19 et 20. Elles ont été volontairement placées en dernier dans la matrice, afin de ne pas risquer de reproduire ce qui s'est déroulé avec la première carotte (vieillesse rapide et/ou cokage).

On remarque également que l'expérience n°4 à une richesse de 0.75, et peut éventuellement entraîner des problèmes, cependant les niveaux des *HC* et *CO* sont pratiquement au plus bas.

L'essai de référence (voir annexe C) n'apparaît pas dans cette matrice, mais des essais de contrôle ont été réalisés au cours de la démarche, afin de contrôler que l'échantillon n'était pas altéré (cf. figure 5.7).

La représentation graphique du plan d'expériences dans l'espace des entrées est donnée par la figure 5.6. Cette représentation du plan défini montre une répartition dans le domaine correcte et suffisante, et sauf comportement extrême sur ces coins, il n'est pas nécessaire de les contraindre.

On note tout de même que des zones du domaine sont non explorées, ce qui n'est pas surprenant puisque le nombre d'essais est restreint.

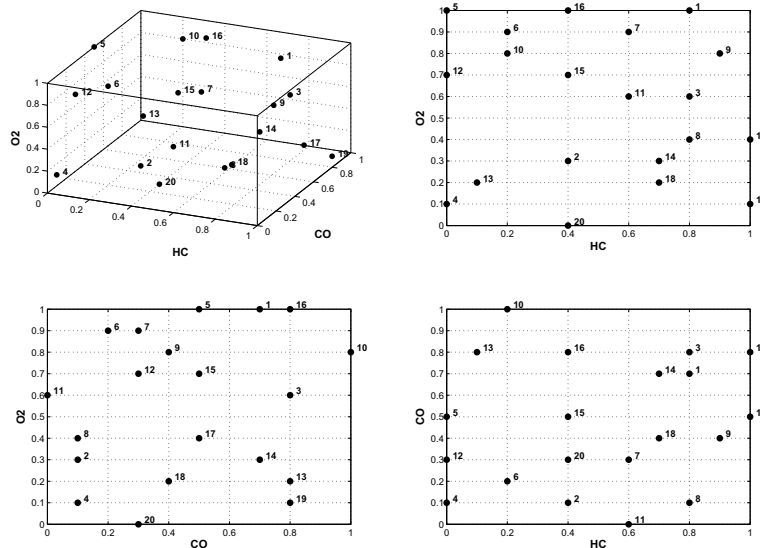


Figure 5.6 – Représentation du plan d'expériences dans l'espace des entrées

La figure 5.7 reprend le résultat des TLO du CO et HC pour chaque expériences. On peut voir que l'on a une relation linéaire entre richesse et TLO.

Trois essais de référence ont été effectués au cours de la campagne d'essais. Un au début, "ref1", un à mi-parcours, "ref2", et un à la fin, "ref3". Le point de recouplement réalisé environ à mi parcour a montré une évolution de 20°C des TLO par rapport au premier essai de référence. Cette différence, que l'on considère minime, provient du changement de comportement de l'échantillon de catalyseur. Le point de recouplement effectué à la fin de la prestation ne montre pas d'évolution des TLO par rapport à celui effectué à mi parcour. On peut donc d'une manière générale conclure que les résultats ne montrent pas de changement significatif de comportement de la carotte.

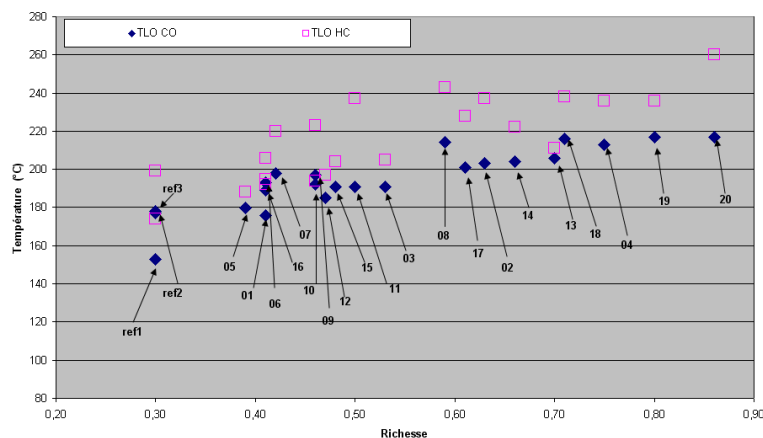


Figure 5.7 – Relation entre TLO et richesse et vérification du bon comportement du nouvel échantillon

5.4 Ouvertures/Perspectives

L'utilisation des plans à entropie maximale dans la planification d'expériences appliqués aux cas de cinétique, permet de construire des plans de bonne qualité, dont la répartition dans l'espace des entrées est pertinente relativement au faible nombre d'expériences. Ces plans sont intéressants d'un point de vue théorique et d'un point de vue pratique.

D'un point de vue théorique d'une part, puisqu'ils ne se contraignent pas à choisir les bords du domaine (même si certains sont tout de même sélectionnés), et qu'ils sont particulièrement indiqués lorsqu'une modélisation du phénomène est effectuée par krigeage. D'un point de vue pratique d'autre part, car les résultats de cette méthode d'application simple sont meilleurs que ce qui est fait actuellement et plus pertinents que des plans basés sur des critères de distance. Par ailleurs, nous avons constaté une amélioration pour l'estimation des paramètres cinétiques, bien que cette dernière ne soit pas suffisante. Cela nous a conduit à une amélioration des prédictions, et également à déterminer une méthode de modélisation satisfaisante et prometteuse (chapitre 4, section 4.2).

De surcroît, la méthodologie des plans à entropie maximale ne nécessite pas de connaissance *a priori* du système, et peut s'appliquer aux nombreux cas rencontrés par les expérimentateurs, constituant donc une alternative appropriée pour ces cas pratiques et/ou comme plan initial.

Cependant, cette approche ne répond pas entièrement à la problématique de planification d'expériences adaptée au cas de cinétique, dans le sens où l'expertise métier n'est pas prise en compte. Toutefois, nous avons introduit, dans le chapitre 4, une méthode de modélisation encourageante en vue de répondre totalement à la problématique.

Cette méthodologie propose d'estimer les paramètres cinétiques expérience par expérience en estimant ces paramètres par une méthode à noyau (*Kernel Ridge Regression*), offrant l'avantage de déterminer une surface de réponse pour représenter l'évolution des paramètres et ainsi il suffit de l'évaluer pour déterminer un nouveau point. Il s'agit donc d'une méthode non-intrusive, qui combine l'avantage de tenir compte de l'expertise métier, et de pouvoir calculer en tout point du domaine l'aire de la région de confiance de la prédiction, à l'aide d'un bootstrap fonctionnel (section 4.2.3). Cette démarche emploie une méthode analogue à celle des VCM (section 4.1), et en étend l'application à un cadre non linéaire, c'est-à-dire représentant la majeure partie des cas rencontrés en pratique par les expérimentateurs.

L'idée alors avancée et à explorer est la suivante : les nouveaux essais à réaliser (points du domaine à explorer) sont les essais dont l'aire de la région de confiance de la prédiction est maximale. Cette perspective nécessite néanmoins une série d'essais initiale, pour laquelle l'entropie maximale est un critère intéressant. On voit donc que la conjugaison de ces deux outils semble appropriée pour les cas de cinétique, pertinente à chaque étape, et laisse entrevoir une méthodologie complète répondant totalement à la problématique.

Cet axe de recherche est envisagé pour de prochaines études menées au sein de IFP Energies Nouvelles, sur d'autres cas pratiques. De plus, il est intéressant de souligner le fait que la méthodologie de plan à entropie maximale sera réutilisée, puisqu'elle offre un outil simple pour les expérimentateurs, s'adaptant aux spécificités et caractéristiques recherchées (points contraints, essais déjà réalisés ou non, domaine non hypercubique), et donnant de meilleurs résultats que ce qui est fait actuellement.



Conclusions et perspectives

Conclusions

Le travail que nous avons présenté dans ce mémoire s'articule principalement autour de la modélisation du phénomène physico-chimique se déroulant au sein du piège à NO_x, avec en arrière plan la définition d'un critère de planification d'expériences adapté. Les différentes méthodes développées pour atteindre ces objectifs ont montré leurs limites, pour la plupart, sur notre cas d'étude, mais nous ont conduits au final à répondre à la problématique posée, ou tout au moins à définir une piste pertinente et prometteuse pour y répondre pleinement.

Les premières approches, appelées approches non fonctionnelles, sont basées sur la théorie du krigeage. Le caractère fonctionnel est omis, puisque les sorties étudiées sont considérées non pas comme des courbes mais comme une succession de points. Trois méthodes ont ainsi été développées et appliquées premièrement sur le cas maîtrisé puis dans un second temps sur le cas d'étude. Tout d'abord, nous avons étendu la théorie de krigeage au cas d'une tendance non linéaire qui correspond au modèle numérique. Cette méthode souffre d'un certain nombre de lacunes, dont les principales raisons sont la non stationnarité du problème et sa forte non linéarité. Les résultats montrant des phénomènes non désirables sur les paliers, nous avons alors apporté une amélioration de cette première méthode, en conditionnant le krigeage par les dérivées, dont on sait qu'elles sont nulles sur ces zones. Néanmoins, là encore les résultats obtenus dans certains cas sont non concluants, ce qui nous a poussés à changer de système de représentation pour contourner la non stationnarité. En effet, si l'on s'intéresse au cas maîtrisé, les résultats que nous avons exposés montrent un très bon comportement de la méthode. De même, pour les résultats de *CO* qui sont certes dans certains cas extrêmes, très mauvais du fait du manque d'information mais restent dans l'ensemble plutôt intéressants, la non stationnarité et la non linéarité étant moins forte. Ce comportement nous pousse à penser que cette méthode bien qu'inadaptée pour notre cas d'étude, reste intéressante pour d'autres cas d'application sous certaines conditions. Enfin, la troisième méthode non fonctionnelle présentée, reprend la théorie de krigeage à tendance non linéaire que nous avons développée, mais cette fois-ci l'application est effectuée pour modéliser la dérivée des sorties selon la température en fonction de l'avancement. Le principe étant de se ramener à un cadre dans lequel la non stationnarité est contournée. Couplée à différents artifices pour l'intégration de la prédiction, les résultats obtenus montrent qu'il s'agit d'une méthode très pertinente pour modéliser le phénomène. En revanche, dans cette approche, le calcul de la variance de prédiction entraîne une impossibilité de définir un critère de plan d'expériences pertinent. Néanmoins, les résultats probants font de cette méthode une alternative efficace si l'on s'intéresse uniquement à la prédiction d'un système, et peut facilement s'adapter aux nombreux cas rencontrés en pratique par les expérimentateurs.

En conclusion sur ces méthodes, nous pouvons dire qu'elles ne sont certes pas adaptées pour notre cas d'étude, mais qu'elles pourront, sous certaines conditions, convenir à d'autres cas d'application. Nous avons dans le chapitre relatif à ces méthodes, le chapitre 3, essayé de présenter pour chacune les limites d'applications. Principalement, il s'agit de la non stationnarité des résidus.

Dans un deuxième temps, nous avons eu recours à des méthodes qui sont adaptées aux données fonctionnelles. La première approche consiste à utiliser un modèle à coefficients variables, approche motivée par la relation linéaire que nous avons observée entre TLO et richesse. Deux méthodes d'estimation des fonctions coefficients ont été utilisées : par P-spline et par polynômes locaux. Cette première approche a le défaut de supposer un lien linéaire entre les concentrations initiales et la concentration des polluants, hypothèse non validée en pratique. Mais surtout, cette approche ne permet pas de tenir compte de la connaissance des experts en ignorant le modèle numérique. Comme on pouvait s'y attendre, les résultats obtenus par les deux méthodes d'estimation des fonctions coefficients pour la première réponse (*CO*) sont acceptables, mais ne sont pas satisfaisants pour la seconde (*HC*), les courbes pouvant présenter au moment du démarrage effectif de la réaction un sursaut choquant pour l'œil expert. La seconde méthode propose elle de réincorporer cette expertise métier, en estimant les paramètres cinétiques utilisés dans le modèle numérique par kernel ridge regression. Cette seconde méthode permet d'obtenir des prédictions de très bonne qualité. De plus, la mise au point d'un bootstrap fonctionnel pour l'estimation de la loi des paramètres de la kernel ridge regression nous a permis de déterminer des régions de confiance satisfaisantes, dont il est possible d'avoir recours pour définir un critère de plan d'expériences. Cette méthodologie complète n'a pu être testée en pratique par manque de temps, mais sera prochainement étudiée pour de futures applications.

En conclusion de ces deux méthodes, nous pouvons dire que bien que la première méthode, basée sur les VCM, soit là encore inadaptée pour l'application, elle a déjà été réutilisée sur un autre cas d'étude et a montré un bon comportement vis à vis de la prédiction. En ce qui concerne la seconde méthode, les résultats de prédiction pour notre application sont très bons. De plus, la méthode pour l'estimation des régions de confiance permet d'entrevoir un critère répondant pleinement à la problématique posée lors de cette thèse.

Enfin, il est à noter que pour mener à bien notre méthodologie, et du fait de la possibilité de mener une campagne expérimentale au cours de ces trois ans, nous avons été amenés à réfléchir à la définition d'une méthode de planification initiale qui soit la plus pertinente pour les méthodes explorées tout au long de ce manuscrit. Notre choix s'est porté sur un critère de maximum d'entropie dont les caractéristiques sont très avantageuses d'une part par sa facilité d'application, d'autre part, par la bonne répartition dans le domaine, entraînant une information maximale pour les méthodes envisagées et couramment utilisées telles que le krigeage. Les résultats sont satisfaisants du point de vue des expérimentateurs, ce qui leur permet de disposer d'ores et déjà d'un premier outil. De plus, il s'intègre judicieusement dans la méthodologie complète que nous avons dégagée, en fournissant un plan initial.

Perspectives

Les perspectives de ces travaux sont nombreuses. Tout d'abord comme nous l'avons mentionné les trois approches non fonctionnelles pourront être réutilisées pour d'autres cas pour lesquels certaines conditions telles que la non stationnarité ou la forte non linéarité seront plus faibles. Notamment, la troisième méthode, basée sur un changement de représentation, dont le principe novateur est peu présent dans la littérature, et est particulièrement adaptée aux nombreux cas de cinétiques rencontrés en pratique.

La première approche fonctionnelle basée sur les VCM a déjà pu être réutilisée avec succès sur un autre cas étudié à IFP Énergies Nouvelles, et à ce titre mérite une étude plus approfondie notamment sur les diverses questions liées à l'estimation des fonctions coefficients, et les questions de convergence de l'algorithme développé.

Enfin, il reste encore à tester la méthodologie complète proposée sur des nouveaux échantillons de catalyseurs et à étendre son domaine d'action en prenant le modèle cinétique dans son ensemble, puisqu'on le rappelle, des simplifications ont été opérées dès le début de thèse et juste une partie du fonctionnement du piège à NOx a été

considérée.

Cette méthodologie complète réside en 6 étapes :

1. Réalisation d'un premier plan par maximum d'entropie,
 2. Estimation des paramètres cinétiques par KRR,
 3. Calcul des régions de confiance en tout point du domaine par bootstrap fonctionnel,
 4. Détection du point où l'aire de cette région de confiance est la plus forte,
 5. Prédiction par la modélisation directe de la dérivée selon l'avancement, que l'on considère comme la vraie sortie du système,
 6. Retour à l'étape 2, jusqu'à l'obtention du nombre d'expériences défini au préalable.
-

ANNEXES

A	Méthode de lissage des courbes expérimentales	171
A.1	Filtre moyen et de Savitzky-Golay	174
A.1.1	Filtre moyen (FM)	175
A.1.2	Filtre de Savitzky-Golay (FSG)	176
A.2	Lissage des HC	177
A.3	Lissage du CO	179
B	Processus de Ornstein-Uhlenbeck conditionné aux bords	183
B.1	Mouvement brownien et pont brownien	183
B.2	Le processus d'Ornstein-Uhlenbeck conditionné aux bords	186
B.2.1	Représentation du processus d'Ornstein-Uhlenbeck	186
B.2.2	Processus d'Ornstein-Uhlenbeck conditionné aux bords	188
C	Viellissement de la carotte	191

ANNEXE

A

Méthode de lissage des courbes expérimentales

La campagne expérimentale qui a été menée au cours de cette thèse, a été réalisée sur un dispositif expérimental appelé "banc gaz synthétique" (BGS). Ce dispositif et le plan d'expériences appliqué sont introduits dans le chapitre 1. Pour construire ce plan, un critère de maximum d'entropie a été utilisé, dont une description précise est donnée dans le chapitre 5.

Les courbes expérimentales obtenues lors de cette campagne, font apparaître des bruits de mesure. Ces courbes, nommées également courbes brutes pour faire le distinguo avec les courbes lissées, présentent des caractéristiques générales propres à chaque sortie. Les oscillations dues aux bruits de mesure, localisées différemment selon la sortie CO , HC ou NOx , sont plus ou moins importantes selon les conditions opératoires.

Nous proposons dans cet annexe de présenter la méthode de lissage que nous avons développée et appliquée sur ces courbes expérimentales. Nous nous sommes bornés tout au long de ce manuscrit à l'étude des sorties CO et HC , et pour cette raison, nous avons fait le choix de n'introduire cette méthode que pour ces deux cas.

L'étude des courbes brutes pour les 20 essais révèle que pour chaque sortie, il est possible de caractériser la forme générale des courbes de manière spécifique. L'idée de base de notre méthode consiste à tenir compte de ces formes particulières, en l'adaptant à la sortie considérée. Les figures A.1a et A.1b illustrent respectivement pour CO et HC les deux cas de figures constatés.

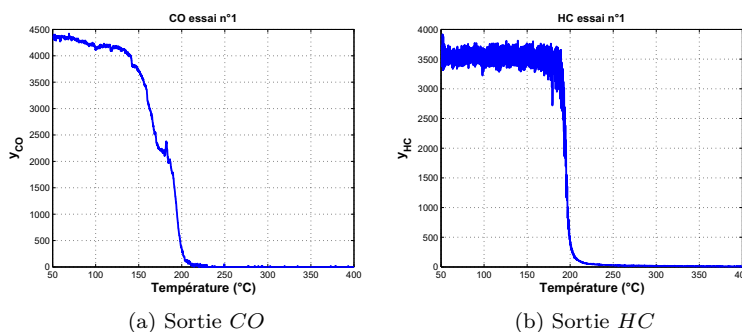


Figure A.1 – Résultats expérimentaux de l'essai 1 pour CO et HC

Pour les deux sorties, une allure générale se dégage :

- un premier palier, nommé palier initial, qui correspond au fait que tant que la température adéquate n'est pas atteinte, le piège n'est pas opérationnel et le polluant n'est pas converti (oxydé),
- une décroissance très forte pour les HC et plus lente pour le CO , les polluants sont traités, puisque la température d'activation du piège a été atteinte,
- un palier final symbolisant le fait que tous les polluants sont traités.

Cependant, on distingue tout de même des différences notables pour chaque sortie. Pour CO , le palier initial peut, pour certaines conditions expérimentales, être plus ou moins plat. Le piège fonctionne, mais la température n'est pas assez élevée pour que l'oxydation soit totale. On observe également, dans certains cas, un palier intermédiaire (entre $170^{\circ}C$ et $190^{\circ}C$ sur la figure A.1a), de longueur très variable, dont l'origine n'est pas clairement définie. Pour HC , il est à noter que le palier initial dans la majorité des cas, est très oscillant, et la décroissance brutale.

Les différences de comportement entre les deux sorties, nous ont contraints à utiliser deux méthodes, toutes deux adaptées spécifiquement selon les propriétés observées.

Les raisons du lissage des courbes brutes sont multiples. Premièrement, ce lissage permet d'obtenir des courbes représentatives de la connaissance du fonctionnement du système, comme par exemple la décroissance de la courbe (le piège ne fabrique pas de polluants), ou encore que la valeur initiale de la courbe soit la valeur initiale du polluant contenu dans le mélange. Deuxièmement, pour certaines des méthodes de modélisation que nous avons explorées, nous avons eu recours au calcul des dérivées selon la température, information disponible uniquement par différences finies sur les courbes lissées. Troisièmement, le résultat du simulateur étant une courbe C^1 , le calcul des résidus à partir des courbes lissées, dans certains cas, a été préféré. Notamment, pour le calcul de l'erreur moyenne quadratique, équation 3.25, pour une meilleure interprétabilité des résultats de prédiction.

Afin de comprendre, pour chacune des deux sorties, le fonctionnement de la méthode de lissage, nous avons fait le choix de les présenter séparément au travers d'un exemple, l'essai 1 (figures A.1). La description étape par étape décrit précisément le déroulement du lissage, qui bien qu'adapté aux spécificités citées précédemment, peut être considéré comme une boîte noire et utilisable afin de lisser les futurs essais réalisés.

Les deux principaux outils utilisés sont deux filtres, que nous appliquons sur des zones spécifiques des courbes et avec des paramétrages différents. Ces deux filtres sont :

- le filtre moyen : variante du filtre médian, un outil couramment utilisé en traitement de l'image. L'ouvrage de Najim [121] offre un descriptif des filtres utilisés dans ce domaine,
- le filtre de Savitzky-Golay : méthode introduite en 1964 par Savitzky et Golay [140], et appartenant au domaine du traitement du signal.

Nous consacrons le début de cet annexe à une description de ces deux outils, puis nous présentons la méthode de lissage étape par étape sur un exemple pour la sortie CO et enfin pour HC .

A.1 Filtre moyen et de Savitzky-Golay

Nous présentons succinctement le principe de fonctionnement de chacun des deux filtres : le filtre moyen et le filtre de Savitzky-Golay.

Le filtre moyen est une variante du filtre médian, un outil largement utilisé en traitement de l'image. Son principe est simple, il consiste à remplacer la valeur d'un pixel par la valeur médiane de l'ensemble des pixels de son voisinage. Pour nos besoins, nous avons remplacé la valeur médiane par la valeur moyenne (empirique), sur des fenêtres de températures variables, ce changement répondant mieux à nos besoins.

Savitzky et Golay ont introduit leur filtre dans leur article "*Smoothing and Differentiation of Data by Simplified Least Squares Procedures*", [140], de 1964. La méthode repose sur une régression polynomiale locale sur une distribution afin de déterminer la valeur lissée en chaque point. Elle est également utilisée pour calculer les dérivées d'ordre 1 à 5., dans des cas où ces dérivées ne sont pas disponibles.

La combinaison des avantages de chacune de ces deux méthodes, à des étapes précises, nous permet d'obtenir des résultats de lissage satisfaisants, dont certains sont présentés dans les sections A.2 et A.3.

A.1.1 Filtre moyen (FM)

Le filtre moyen est un outil classique en traitement de l'image. Cependant, dans ce domaine d'application, le filtre médian lui est souvent préféré. Une des propriétés fondamentales du filtre médian, est qu'il ne crée pas de nouvelles valeurs de niveaux de pixels dans l'image, et son principe de fonctionnement est parallélisable.

Son fonctionnement consiste à remplacer la valeur d'un pixel par la valeur médiane de l'ensemble des pixels de son voisinage. Néanmoins, dans notre cas, nous décidons d'utiliser ce filtre pour effectuer un lissage sur une courbe bruitée. Les données sont triées en ordre croissant selon la température, et nous ne souhaitons pas le paralléliser. Le choix de prendre la moyenne empirique plutôt que la médiane apparaît plus adapté. Ce choix arbitraire est fait en accord avec les résultats obtenus.

Plus précisément, le principe consiste à prédire en chaque température, la valeur lissée à l'aide d'un ensemble de points, centré sur la valeur que l'on souhaite. Ce lissage utilise une fenêtre coulissante dans le sens croissant des températures, et remplace la valeur à chaque température par la moyenne empirique de l'ensemble de points.

Soit $y_{exp}(T_i)$ la valeur de la courbe brute à la température T_i , $i = 1, \dots, m$, où m est le nombre de température discrétisées. La valeur $y_{lisse}(T_i)$ de la courbe lissée à la température T_i , $i = 1, \dots, m$ avec une fenêtre de taille h (h impaire!) est donnée par

$$y_{lisse}(T_i) = \frac{1}{h} \left(\sum_{j=i-\frac{h-1}{2}}^{i-1} y_{lisse}(T_j) + \sum_{k=i}^{i+\frac{h-1}{2}} y_{exp}(T_k) \right). \quad (\text{A.1})$$

Le lissage s'effectue pour les températures triées dans l'ordre croissant.

Cette notation met en évidence ce que l'on appelle l'effet de bord. Par exemple, le lissage de la première valeur (respectivement de la dernière) utilise des valeurs non disponibles, les y_{lisse} pour des températures plus faible (resp. des y_{exp} pour des températures plus fortes). La littérature fournit plusieurs choix pour traiter cet effet de bord : soit on fixe par convention que ces valeurs sont égales à la valeur du point le plus proche (le premier ou le dernier point), soit on fixe arbitrairement ces valeurs à 0, soit les $\frac{h-1}{2}$ premières valeurs (resp. $\frac{h+1}{2}$ dernières valeurs) ne sont pas modifiées. Ces trois choix sont les plus couramment appliqués, et nous avons décidé de retenir la première de ces trois options.

Autrement dit, pour le calcul des y_{lisse} des $\frac{h-1}{2}$ premières valeurs (resp. $\frac{h-1}{2}$ dernières valeurs), les valeurs manquantes pour le calcul de la moyenne sont fixées égales à la première valeur de y_{exp} (resp. à la dernière valeur de y_{exp}).

Ce type de filtrage local est qualifié de filtrage linéaire puisque la valeur du nouveau pixel est une combinaison linéaire des valeurs des pixels du voisinage. Son principal avantage est de pouvoir uniformiser la courbe en éliminant les discontinuités, puisqu'en prenant une fenêtre adéquate autour d'un point de discontinuité, les fortes transitions sont éliminées. En revanche, cette méthode n'élimine pas des pics importants mais ne fait que les diffuser.

Pour nos essais, nous disposons de 2652 températures pour lesquelles nous connaissons la valeur de la courbe brute. Pour l'élimination des discontinuités, la taille de fenêtre comprend environ 101 points, pour des lissages

globaux d'uniformisation, des fenêtres plus petites d'environ 30 points, et dans le cas d'un premier lissage, des fenêtres plus grande (environ 200 points).

A.1.2 Filtre de Savitzky-Golay (FSG)

Les filtres lissant de Savitzky-Golay, également appelés filtres lissant des moindres carrés, sont typiquement utilisés pour lisser un signal bruité dont l'amplitude de la fréquence (sans bruit) est large. Dans ce type d'application, ces filtres fonctionnent significativement mieux que des filtres moyens standards, notamment pour conserver la pertinence des hautes fréquences (Orfanidis [125]).

Le filtre de lissage de Savitzky-Golay utilise une régression locale au lieu du calcul habituel de la moyenne mobile. L'utilisation de ce filtre est particulièrement intéressante pour lisser des signaux dont les pics doivent être préservés aussi intégralement que possible. Dans notre cas, nous utilisons cette méthode pour fournir une première estimation de la courbe lissée du fait de la forme générale de nos sorties, mais également pour le palier initial des *HC*.

L'approche utilisée pour lisser une série temporelle est de remplacer chaque valeur de la série par une nouvelle valeur, obtenue par l'ajustement d'un polynôme basé sur les $2n + 1$ points voisins (incluant le point à lisser), avec n au moins égal à l'ordre du polynôme.

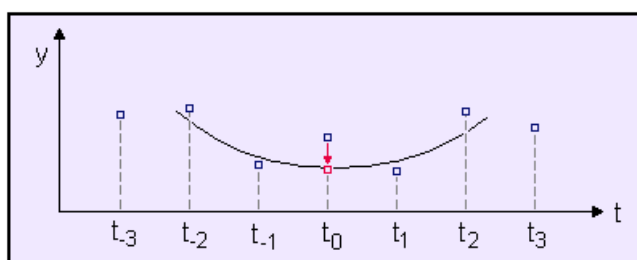


Figure A.2 – Illustration du filtrage d'une valeur par la méthode de Savitzky-Golay

La figure A.2 illustre la méthode pour le lissage d'une valeur en t_0 , à l'aide d'un polynôme local du second ordre, avec 7 points.

Dans leur article original [140], les auteurs ont montré qu'un ajustement polynomial coulissant peut être numériquement traité exactement de la même manière qu'une moyenne coulissante pondérée, puisque les coefficients de la procédure de lissage sont constants pour toutes les valeurs y . Ainsi, le lissage de Savitzky-Golay est très facile à appliquer. De plus, le même algorithme peut être utilisé pour calculer les dérivées lissées de premier et second ordre d'un signal bruité.

Nous avons fait le choix de prendre des polynômes d'ordre 3, et des tailles de fenêtre qui sont relativement grandes, afin de bien respecter ce que l'on observe expérimentalement. Comme nous l'avons évoqué précédemment, nous disposons en tout de 2652 points pour chaque essai. Ainsi, nous avons opté pour une fenêtre de taille maximale lorsque nous l'utilisons pour lisser le palier initiale de *HC* qui est très bruitée, pour obtenir un lissage le plus plat possible, tout en respectant la valeur moyenne du palier. Et une fenêtre conséquente (environ 700) pour obtenir la première courbe lissée pour *CO*, qui d'une manière générale est peu bruitée.

A.2 Lissage des HC

Les courbes pour HC ont une forme beaucoup plus simple et caractéristique que pour le CO : un palier initial très bruité, une décroissance brutale (bruité ou non) et un palier final très peu (voir pas du tout) bruité. Pour cette raison, le lissage de ces résultats est beaucoup plus simple que pour le CO , qui nécessite quelques adaptations supplémentaires (voir section A.3).

Un lissage complet unique, que se soit par filtre moyen ou de Savitzky-Golay, n'est pas satisfaisant car soit la fenêtre est trop petite et dans ce cas on a toujours des oscillations sur le palier initial, soit il est trop grand, et dans ce cas, la courbe lissée présente un écart sur la décroissance et le palier final, pouvant même entraîner des concentrations négatives.

Ainsi la méthode consiste en un premier lissage moyen sur la courbe et un second plus petit sur la première courbe lissée pour réduire les oscillations sur le palier initial. Ces deux filtrages successifs garantissent un bon lissage sur la décroissance et le palier final. Il reste néanmoins à traiter le palier initial qui, comme on le voit sur la figure A.3b, ne représente pas la connaissance du système (décroissance, palier à la valeur de $[HC]_0$).

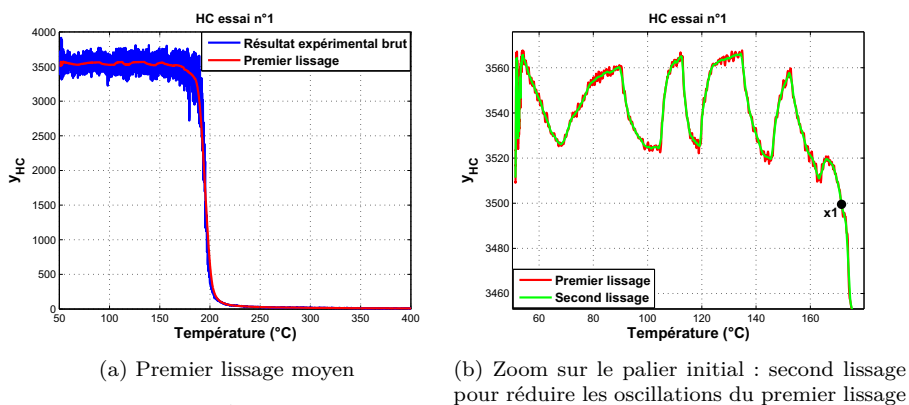


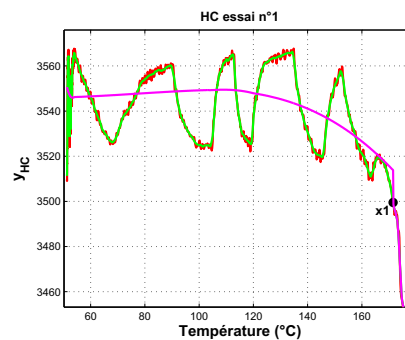
Figure A.3 – Étape 1 du lissage du résultat expérimental pour HC

La figure A.3a montre le résultat du premier lissage effectué. À l'aide d'un filtre moyen (FM), avec une taille de fenêtre comportant 81 points, on constate que le premier lissage obtenu capte bien le début de la décroissance et respecte bien les propriétés déjà évoquées à partir de cette zone.

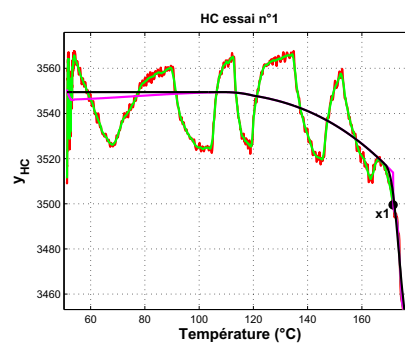
Un zoom sur le palier initial, figure A.3b, illustre les petites oscillations sur cette zone. Le second lissage par FM avec une fenêtre de 11 points sur la première courbe obtenue (courbe rouge) vient corriger ces petites oscillations. C'est une étape importante, puisqu'elle fournit une courbe plus lisse (courbe verte) nous permettant de délimiter la fin du palier initial efficacement.

La détermination de la fin du palier est opérée de manière très simple en choisissant un point suffisamment éloigné de la dernière oscillation restante. Le point noir x_1 sur la figure A.3b illustre sur l'exemple le résultat. On délimite ainsi notre palier initial entre la température initiale et le point x_1 .

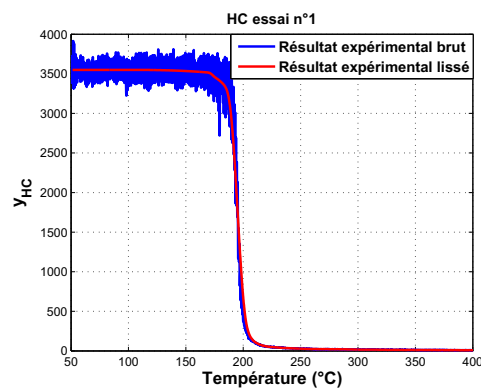
L'étape suivante consiste à lisser uniquement ce palier en appliquant un filtre de Savitzky-Golay (FSG), avec l'ensemble des points compris sur ce palier comme fenêtre. Ce choix très large de fenêtre pour le lissage par FSG permet d'obtenir une courbe ondulante de faible amplitude et de longueur d'onde élevée (courbe magenta sur la figure A.4). On notera également la discontinuité en x_1 .

Figure A.4 – Étape 2 du lissage du résultat expérimental pour HC

Afin de respecter un palier initial plat de valeur égale à $[HC]_0$, on remplace la courbe lissée entre le début de la courbe et le point maximal de l'ondulation par une droite horizontale au niveau de $[HC]_0$. Le maximum de l'ondulation est obtenu à l'aide des dérivées et se situe aux alentours de 110°C dans le cas présenté (voir figure A.5). Le tout est une nouvelle fois lissé par FM, avec une taille de fenêtre très petite (21 points) afin de lisser la discontinuité en x_1 et uniformiser la courbe générale.

Figure A.5 – Étape 3 du lissage du résultat expérimental pour HC

On conserve donc les propriétés de continuité et de décroissance, un lissage très proche du résultat expérimental, une TLO correcte et un palier initiale plat, à la valeur de $[HC]_0$ et qui représente bien ce qu'il se passe expérimentalement. La figure A.6 présente le résultat final obtenu.

Figure A.6 – Résultat final du lissage du résultat expérimental pour HC

Voici quelques exemples caractéristiques illustrant les résultats obtenus :

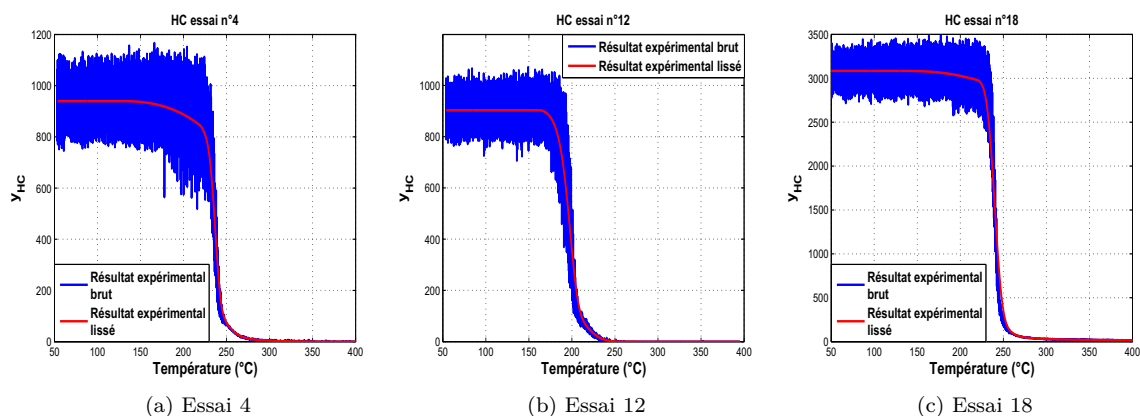


Figure A.7 – Trois exemples de lissage des résultats expérimentaux

L'essai 4 illustre un cas où le palier initial est très oscillant (figure A.7a), l'essai 12 un cas où les oscillations se poursuivent sur la décroissance (figure A.7b) et l'essai 18, un essai dans lequel le palier initial est très long, figure A.7c. Les courbes lissées obtenues pour ces trois essais représentatifs sont très bonnes dans le sens du respect des contraintes et de la concordance avec les courbes brutes.

A.3 Lissage du CO

Pour ce qui est de la méthode de lissage pour la sortie CO , le principe général s'apparente à la méthode que nous venons de présenter dans le cas des HC . Cependant, les différences de comportement notables des courbes brutes, expliquent la mise en place d'étapes supplémentaires, et des paramétrages de fenêtre différents.

Les courbes brutes des 20 essais montrent des comportements plus complexes et variés, et une décroissance pas forcément brusque selon les cas. On observe également pour la plupart des essais, un palier intermédiaire dont l'importance peut être très différente. Par exemple, ce palier est faiblement marqué sur notre exemple, figure A.8, se situant à peu près entre 160°C et 180°C. Ce palier peut être relativement plat, de longueurs variables et éventuellement présenter des oscillations.

D'une manière générale, il est possible de découper la courbe en quatre zones, chacune étant lissées d'une manière différente. Un lissage complet final permet d'uniformiser le lissage en supprimant les discontinuités aux points de jonction de chaque zone. Nous détaillons par la suite précisément comment l'identification et le découpage de ces zones sont effectués. D'une manière simple, ces quatre zones sont le palier initial (du début de la courbe au début de la décroissance plus marquée), le début de la décroissance (de la fin du palier initial à la TLO et qui contient le palier intermédiaire s'il existe), la fin de la décroissance (de la TLO à la température où la concentration est en dessous d'un seuil) et le palier final.

La première étape consiste à appliquer un FSG avec une taille de fenêtre conséquente (701 points). L'avantage de cette méthode est de bien représenter le début de la courbe brute jusqu'à la TLO. En effet, au delà, comme on peut le voir sur la figure A.8, la courbe lissée présente un certain écart avec la courbe brute et surtout décroît jusqu'à des valeurs négatives. On délimite donc deux premières zones : avant et après la TLO. On note que le palier initial n'étant pas très bruité, la valeur initiale de la courbe correspond bien à la valeur initiale $[CO]_0$ du mélange.

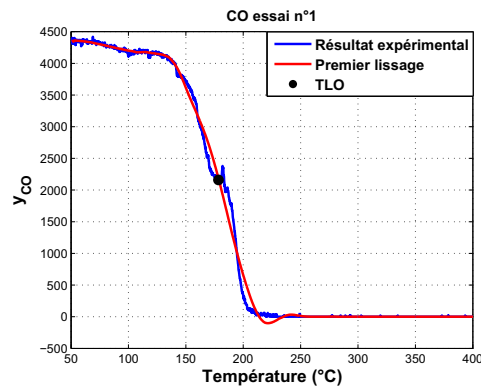


Figure A.8 – Étape 1 du lissage du résultat expérimental pour CO

En revanche, une étude plus attentive sur la première partie fait apparaître de petites oscillations, ou zone de croissance, que nous souhaitons prévenir. On détermine alors, à l'aide de la valeur de la dérivée (selon la température par différences finies), ces zones où la courbe est croissante. On parcourt la courbe lissée du début jusque la TLO, en remplaçant une à une ces croissances par une droite horizontale, de manière à "étêter" l'oscillation. Ainsi, on obtient certes plusieurs points de discontinuités, mais la décroissance de la courbe est respectée.

La figure A.9 exhibe l'étêtage d'une oscillation sur notre exemple. La courbe bleue représente la courbe brute, la rouge, la première courbe lissée que l'on parcourt, et la courbe verte, le résultat après étêtage.

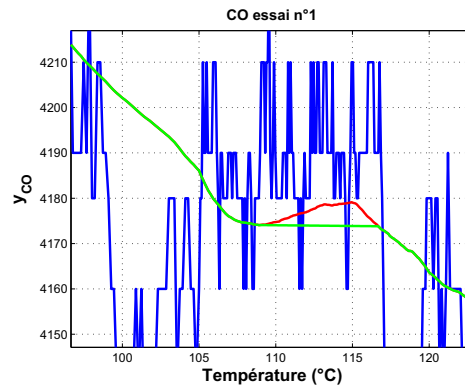


Figure A.9 – Étape 2 du lissage du résultat expérimental pour CO

Cet exemple est également intéressant car il correspond à la dernière croissance corrigée sur la première partie de la courbe. C'est précisément à cette endroit (point x_1 sur les prochaines figures) que nous délimitons la fin du palier initial.

En effet, ces croissances trouvent leur origine dans le fait que la température adéquat n'étant pas atteinte, le fonctionnement du piège n'est pas optimal, et ainsi la décroissance est faible. En revanche, dès que le piège entre pleinement en action, la décroissance s'accélère, et ces petites oscillations disparaissent. On définit ainsi la limite du palier initial.

Enfin, on détermine le premier moment où la courbe brute atteint son minimum (proche ou égal à 0). Ce point marque la fin de la décroissance et le début du palier final (point x_3).

Notre courbe est ainsi décomposée en quatre zones :

- palier initial : du début de la courbe à x_1 ,
- début de la décroissance : de x_1 à x_2 , le point correspondant à la TLO (c'est cette zone qui contient le palier intermédiaire s'il existe),
- fin de la décroissance : de x_2 à x_3
- palier final : de x_3 à la fin de la courbe.

La figure A.10 offre une visualisation claire de ces quatre zones. Sur chacune, un lissage particulier est opéré que nous détaillons un à un par la suite. Auparavant, nous insistons sur le fait que l'étêtage des croissances et la détermination des zones est automatique pour chaque essai, compte tenu des connaissances du système.

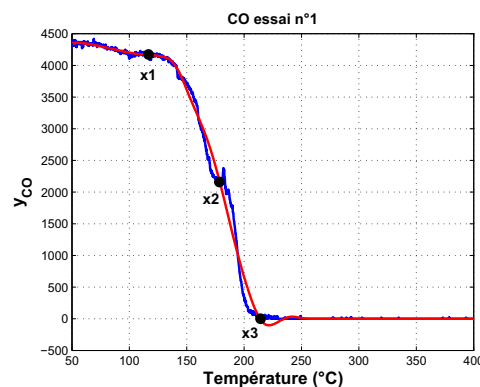


Figure A.10 – Détermination des quatre paliers

Premier palier : 0 à x_1

La valeur initiale correspond à $[CO]_0$ et, de plus, la décroissance est respectée car chacune des croissances, ou oscillations, ont été remplacées par une droite (voir figure A.9). Cependant, des discontinuités dues à l'étêtage sont présentes sur ce palier, et sont corrigées par la suite.

Second palier : x_1 à x_2

Sur le second palier, nous faisons le choix d'appliquer un FM sur la courbe lissée. La taille de la fenêtre n'est pas fixe, puisque selon les cas, cette zone est plus ou moins grande. L'idée est de choisir une fenêtre relativement grande, correspondant à un quart de la zone, pour respecter mieux la courbure brusque observée expérimentalement. Dans notre cas, une fenêtre comportant 201 points a été utilisée sur la courbe lissée.

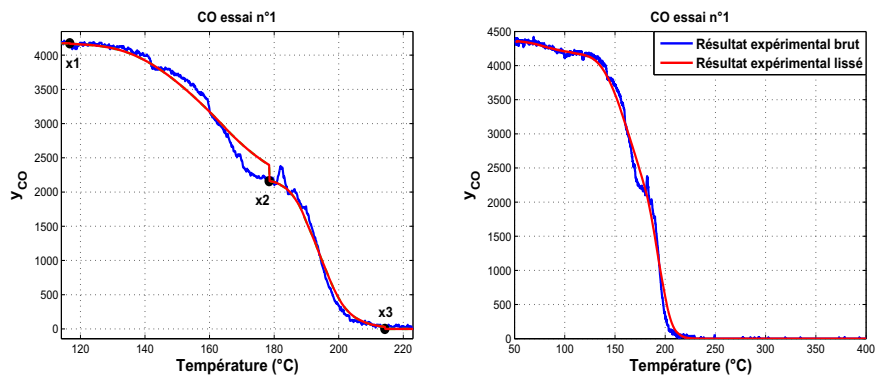
Troisième palier : x_2 à x_3

Pour la fin de la décroissance, nous avons pu voir que la courbe lissée obtenue initialement n'est pas satisfaisante, car celle-ci n'est pas proche du résultat expérimentale, mais également car la décroissance descend jusqu'à atteindre des valeurs négatives. Pour cette raison et le fait que la courbe brute soit peu chahutée, nous utilisons un FM de petite taille (101 points) appliqué sur les données expérimentales de cette zone.

Quatrième palier : x_3 à la fin

On fixe la valeur de ce palier constante et égale à la valeur lissée de x_3 (proche ou égal à 0).

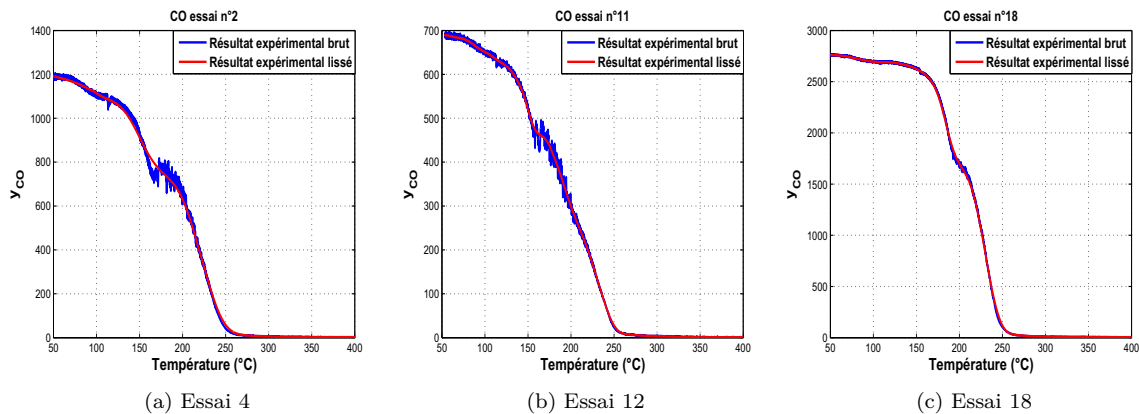
On obtient donc une fonction discontinue en les points de jonction des paliers x_1 , x_2 , x_3 . Mais qui respecte la décroissance et les paliers, en étant proche du résultat exp. L'étape finale consiste donc à uniformiser notre courbe. Cette uniformisation se déroule en deux temps : Premièrement, on lisse chacune des jonctions par un FM en prenant une zone centrée sur chacun des points x_1 , x_2 , x_3 de taille raisonnable (51, 101, 101). Cette étape permet une atténuation des discontinuités. Deuxièmement, on réalise un lissage par FM sur l'ensemble, plus grossier (fenêtre de 151 points). Ceci afin d'obtenir un lissage parfait qui respecte les contraintes de décroissance, la valeur initiale et le palier final ainsi que la TLO.



(a) Courbe discontinue aux points de jonction (b) Résultat final du lissage du résultat expérimental pour CO

Figure A.11 – Courbes discontinues et résultat final

Cette méthode automatique de lissage qui tient compte des connaissances que l'on a des courbes en sorties, donnent de bons résultats sur tous les essais. Les figures A.12a, A.12b et A.12c illustrent bien sur des exemples très différents, le bon comportement de la méthode.



(a) Essai 4

(b) Essai 12

(c) Essai 18

Figure A.12 – Trois exemples de lissage des résultats expérimentaux

L'essai 4 correspond à un cas où le palier intermédiaire est important est oscillant (figure A.12a), l'essai 12 à un essai de faible valeur initiale de CO avec une décroissance bruitée (figure A.12b), et enfin l'essai 18 illustre un cas de forte valeur de CO , avec un palier initial très plat et long, et un palier intermédiaire peu marqué (figure A.12c). Là encore, les résultats de lissage sur ces essais représentatifs, confortent le bon fonctionnement de notre méthode.

Processus de Ornstein-Uhlenbeck conditionné aux bords

Cet annexe a pour but de montrer que le noyau d'Ornstein-Uhlenbeck, ou encore noyau exponentiel, défini par :

$$\forall h \in \mathbb{R}, k(h, \sigma^2, \alpha) = \sigma^2 \left(e^{-\alpha \|h\|} \right), \quad (\text{B.1})$$

dans le cas où il est conditionné aux bords du domaine est équivalent à un pont brownien.

Notre motivation ici est de justifier l'emploi du noyau exponentiel dans l'approche de la section 3.3, dans laquelle on modélise la dérivée des sorties en fonction de l'avancement. Dans cette modélisation, les résidus entre résultats expérimentaux et simulés sont proches d'un pont brownien, et on espère ainsi parvenir à une modélisation efficace.

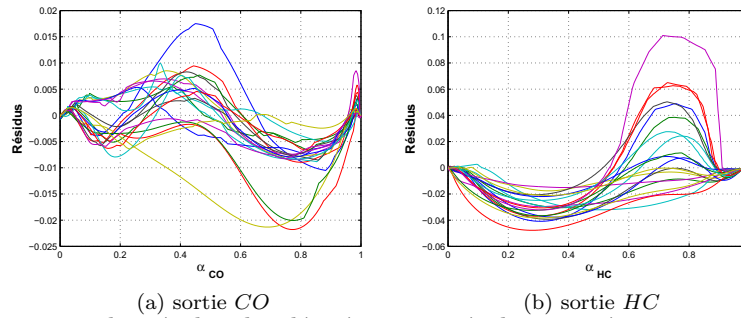


Figure B.1 – Représentations des résidus des dérivées entre résultats expérimentaux et numériques pour les 20 essais

L'objectif de cet annexe est de trouver un noyau de krigeage qui reproduise la forme d'un pont brownien.

Dans un premier temps, nous donnons un rapide aperçu historique des mouvements browniens, puis nous détaillons la définition d'un mouvement brownien que l'on conditionne en une valeur de temps. Un cas particulier est le pont brownien, dont nous détaillons quelques caractéristiques. Dans un second temps, nous introduisons la définition d'un processus d'Ornstein-Uhlenbeck, et montrons que ce dernier est, à une transformation près, un mouvement brownien, et que le fait de le conditionner aux bords se résume donc, à une transformation près, à un pont brownien, dont nous détaillons l'espérance et la covariance.

B.1 Mouvement brownien et pont brownien

Historiquement, le mouvement brownien est associé à l'analyse de particules qui évoluent au cours du temps de manière si désordonnée qu'il semble difficile de prévoir leur position, même dans un intervalle de temps très court. Il joue un rôle central dans la théorie des processus aléatoires comme le modèle de bruit blanc gaussien, parce que dans de nombreux problèmes théoriques ou appliqués, le mouvement brownien ou les diffusions que l'on en déduit fournissent des modèles limites simples sur lesquels de nombreux calculs peuvent être faits.

C'est le botaniste anglais Robert Brown qui le premier décrit en 1827 le mouvement erratique de fines particules organiques en suspension dans un gaz ou un fluide. Au XIX^{ème} siècle, après lui, plusieurs physiciens reconnaissent que ce mouvement est très irrégulier et ne semble pas admettre de tangente ; on ne pourrait donc pas parler de sa vitesse, ni *a fortiori* lui appliquer les lois de la mécanique.

Mais ce sont les travaux de Paul Lévy [101] qui ont fixé l'usage de nommer "mouvement brownien", les processus introduits mathématiquement par Wiener. Auparavant, Louis Bachelier [9] introduit le mouvement brownien pour modéliser la dynamique des prix des actions à la Bourse. Cependant, ses travaux n'ont connu une notoriété importante, notamment en mathématiques financières que bien après. En fait, ce sont les travaux de Einstein, [47], et de Perrin, [127], au début du siècle dernier qui sont à l'origine du grand intérêt porté à ce processus. Et en 1923, Norbert Wiener, [161] est le premier à construire rigoureusement "la fonction aléatoire" du mouvement brownien, il établit en particulier que les trajectoires sont continues.

D'autres liens ou sources sont établis par plusieurs auteurs. Les retombées de ces différentes approches sont très nombreuses, et notre but n'est pas ici de définir de manière précise l'historique et les applications de ce type de processus. Le récent ouvrage de El Karoui et Gobet [49] fournit une description historique et précise de son origine à nos jours, et les implications dans les recherches récentes et dans notre quotidien. On citera également les divers travaux de Kahane qui fournissent un éclairage accessible, notamment historique, et captivant de cet objet.

Le mouvement brownien joue un rôle clef en théorie des processus aléatoires et en calcul stochastique. On le définit comme étant un processus stochastique $\{B_t\}_{t \in \mathbb{R}^+}$ tel que $B_0 = 0$ (départ à l'origine), $\forall t \in \mathbb{R}^+$, $B_t \sim \mathcal{N}(0, t)$, et $\forall t_1, t_2, t_3, t_4 \in \mathbb{R}^+$ tel que $t_1 < t_2 < t_3 < t_4$, $B_{t_4} - B_{t_3}$ est indépendant de $B_{t_2} - B_{t_1}$. On parle alors d'accroissements indépendants et stationnaires.

Le mouvement brownien possède de nombreuses propriétés remarquables, parmi lesquelles la propriété de Markov : si $0 \leq s < t$, la loi de B_t ne dépend des valeurs $\{B_u\}_{u \leq s}$ prises par le processus B avant le temps s que via la valeur B_s .

Nous donnons ci-dessous une autre définition du mouvement brownien B , utilisée par la suite. La figure B.2 présente une réalisation d'un mouvement brownien.

Définition B.1.1 *Un mouvement brownien est défini comme étant le processus gaussien B tel que*

$$\mathbb{E}(B_t) = 0 \quad \text{et} \quad \text{Cov}(B_s, B_t) = s \wedge t \quad (\text{B.2})$$

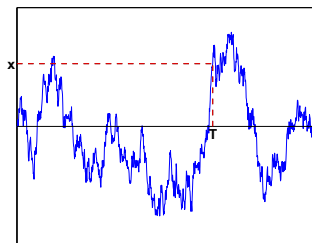


Figure B.2 – Réalisation d'un mouvement brownien

On s'intéresse maintenant à un mouvement brownien conditionné par la valeur x au temps T (cf. figure B.2). Ce type de problème est connu sous l'appellation de conditionnement d'un processus gaussien. Il s'agit d'un sujet

central en probabilités et constitue le fondement théorique de nombreuses méthodes. Les travaux de Revuz et Yor, [132], introduisent ces objets de manière claire, et fournissent notamment un descriptif du conditionnement.

Déterminons maintenant le processus conditionnel $(B_t, t \in [0, T] | B_T = x)$, qui pour des raisons théoriques est supposé gaussien, soit le mouvement brownien tel que la valeur en T est fixée à x .

On écrit la projection orthogonale (dans $L^2(P)$) du processus B_t :

$$B_t = \alpha_t B_T + \epsilon_t; \quad 0 \leq t \leq T \quad (\text{B.3})$$

où donc α_t est solution de

$$\mathbb{E}(B_t B_T) = \alpha_t \mathbb{E}(B_T^2). \quad (\text{B.4})$$

On en déduit

$$B_t = \frac{t}{T} B_T + \epsilon_t; \quad 0 \leq t \leq T \quad (\text{B.5})$$

On a donc

$$\mathbb{E}(B_t | B_T = x) = \frac{t}{T} x \quad \text{et} \quad \text{Cov}(B_t, B_s) = s \wedge t = \frac{t}{T} \frac{s}{T} T + \text{Cov}(\epsilon_t, \epsilon_s). \quad (\text{B.6})$$

On en tire que :

$$\text{Cov}(\epsilon_t, \epsilon_s) = t \wedge s - \frac{ts}{T} = (t \wedge s) \left[1 - \frac{t \vee s}{T} \right]. \quad (\text{B.7})$$

D'où le résultat :

Définition B.1.2 *Le processus gaussien $(B_t, t \in [0, T] | B_T = x)$ est*

$$\text{de moyenne} \quad t \rightarrow \frac{t}{T} x \quad (\text{B.8})$$

$$\text{de covariance} \quad (t \wedge s) \left[1 - \frac{t \vee s}{T} \right] \quad (\text{B.9})$$

C'est aussi le processus

$$t \rightarrow \frac{t}{T} x + B_t - \frac{t}{T} B_T \quad (\text{B.10})$$

où $(B_t)_t$ est un mouvement brownien (standard).

On vérifie bien que $\text{Cov}(X_t, X_s) = (t \wedge s) \left[1 - \frac{t \vee s}{T} \right]$ en notant $X_t = \frac{t}{T} x + B_t - \frac{t}{T} B_T$.

On retient donc :

Définition B.1.3 *Soit B_t un mouvement brownien standard (i.e. $\text{Cov}(B_t, B_s) = t \wedge s$).*

Alors $(B_t, t \in [0, T] | B_T = x)$ est le processus gaussien

$$X_t = \frac{t}{T} x + B_t - \frac{t}{T} B_T, \quad 0 \leq t \leq T \quad (\text{B.11})$$

de moyenne $t \rightarrow \frac{t}{T} x$, et de covariance $(t \wedge s) \left[1 - \frac{t \vee s}{T} \right]$

Ce type de processus gaussien conditionné en un point est l'exemple du pont brownien.

On appelle loi du pont brownien, la loi d'un mouvement brownien B_t sur $[0, 1]$ contraint à passer par 0 en $t = 1$. En prenant, la définition B.1.3, et en l'appliquant pour le cas $T = 1$, il est possible de définir le pont brownien (standard) comme étant le processus $B_t - tB_1$, $0 \leq t \leq 1$. Ce processus est un processus gaussien centré et de covariance $(t \wedge s) [1 - t \vee s]$.

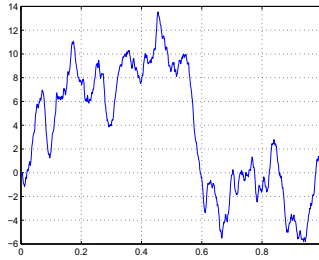


Figure B.3 – Réalisation d'un pont brownien

B.2 Le processus d'Ornstein-Uhlenbeck conditionné aux bords

Cette section présente dans un premier temps le processus d'Ornstein-Uhlenbeck, et nous montrons qu'il s'agit, à une transformation près, d'un mouvement brownien. Ainsi, comme nous venons de le montrer, le conditionnement aux bords du domaine, avec comme cas particulier $T = 1$, de ce processus entraîne l'équivalence, à une transformation près, à un pont brownien dont nous détaillons dans un second temps, les caractéristiques (moyenne et covariance).

B.2.1 Représentation du processus d'Ornstein-Uhlenbeck

L'origine de ce processus introduit par Ornstein et Uhlenbeck appartient là encore au domaine de la physique, et peut représenter, par exemple, la vitesse d'une particule dans un milieu visqueux, soumise à des variations désordonnées dues aux chocs des molécules.

En géostatistique, le variogramme (ou noyau) associé est plus connu sous le nom de noyau exponentiel, et est défini par l'équation B.1. Le paramètre $1/\alpha$ correspond à la portée, et σ^2 est appelé de manière générale, palier, et représente la valeur limite de k lorsque $\|h\|$ tend vers l'infini. Cette valeur correspondant à la distance $|h|$ à partir de laquelle $k = 0$. La portée donne justement un ordre de grandeur de la distance critique à partir de laquelle le variogramme atteint ce palier, c'est-à-dire à partir de laquelle deux réalisations cessent d'avoir de l'influence l'une sur l'autre.

Pour démontrer que le processus d'Ornstein-Uhlenbeck est un mouvement brownien (ou plutôt un cousin), on considère l'équation différentielle stochastique suivante :

$$\begin{cases} dX_t &= -\alpha X_t dt + \sigma_0 dB_t \\ X_0 &\sim \mathcal{N}\left(0, \frac{\sigma_0^2}{2\alpha}\right) \text{ indépendant de } B \end{cases} \quad (\text{B.12})$$

où B_t est un mouvement brownien standard.

En multipliant par $e^{\alpha t}$, les deux membres de l'équation B.12, on obtient

$$e^{\alpha t} (dX_t + \alpha X_t dt) = \sigma_0 dB_t e^{\alpha t}, \quad (\text{B.13})$$

soit

$$d[e^{\alpha t} X_t] = \sigma_0 dB_t e^{\alpha t}. \quad (\text{B.14})$$

Cette équation est équivalente à

$$X_t = X_0 e^{-\alpha t} + \sigma_0 \int_0^t e^{-\alpha(t-u)} dB_u, \quad (\text{B.15})$$

ce qui constitue une expression du processus d'Ornstein-Uhlenbeck, d'où la définition suivante :

Définition B.2.1 Soit X_t le processus défini par :

$$X_t = X_0 e^{-\alpha t} + e^{-\alpha t} \int_0^t \sigma_0 e^{\alpha u} dB_u \quad (\text{B.16})$$

où X_0 est de loi normale, et B est un mouvement brownien. Alors X_t est le processus gaussien centré de covariance :

$$\mathbb{C}ov(X_t, X_s) = \frac{\sigma_0^2}{2\alpha} e^{-\alpha|t-s|} = \sigma^2 e^{-\alpha|t-s|} \quad (\text{B.17})$$

en posant $\sigma^2 = \frac{\sigma_0^2}{2\alpha}$.

L'équation B.16 peut également s'exprimer sous la forme, où la lettre B désigne à chaque fois un mouvement brownien différent,

$$X_t = X_0 e^{-\alpha t} + e^{-\alpha t} B_{\int_0^t \sigma_0^2 e^{2\alpha u} du} \quad (\text{B.18})$$

Soit,

$$\begin{aligned} X_t &= X_0 e^{-\alpha t} + e^{-\alpha t} B_{\sigma_0^2 \frac{e^{2\alpha t} - 1}{2\alpha}}, \\ &= X_0 e^{-\alpha t} + e^{-\alpha t} B_{\sigma^2 (e^{2\alpha t} - 1)}, \\ &= X_0 e^{-\alpha t} + \sigma e^{-\alpha t} B_{e^{2\alpha t} - 1}. \end{aligned} \quad (\text{B.19})$$

On a donc :

$$X_t = X_0 e^{-\alpha t} + \sigma e^{-\alpha t} B_{e^{2\alpha t} - 1}. \quad (\text{B.20})$$

On vérifie bien que, X_t bien gaussien centré ($\mathbb{E}(X_t) = 0$) et de covariance :

$$\mathbb{C}ov(X_t, X_s) = \sigma^2 e^{-2\alpha t} + \sigma^2 e^{-\alpha t} e^{-\alpha s} (e^{2\alpha t} - 1) \wedge (e^{2\alpha s} - 1), \quad (\text{B.21})$$

si donc $s \leq t$, on a :

$$\begin{aligned} \mathbb{C}ov(X_t, X_s) &= \sigma^2 e^{-2\alpha t} + \sigma^2 e^{-\alpha t} e^{-\alpha s} (e^{2\alpha t} - 1) \wedge (e^{2\alpha s} - 1), \\ &= \sigma^2 e^{-2\alpha t} + \sigma^2 e^{-\alpha(t+s)} (e^{2\alpha s} - 1), \\ &= \sigma^2 e^{-\alpha(t-s)}, \\ &= \sigma^2 e^{-\alpha|t-s|}. \end{aligned} \quad (\text{B.22})$$

Cette dernière représentation et ses caractéristiques, nous permet de conclure que le processus d'Ornstein-Uhlenbeck est, à un changement de temps près et une transformation physique près, un mouvement brownien !

B.2.2 Processus d'Ornstein-Uhlenbeck conditionné aux bords

Nous avons donc défini le processus de Ornstein-Uhlenbeck par :

$$X_t = X_0 e^{-\alpha t} + \sigma e^{-\alpha t} B_{e^{2\alpha t}-1}. \quad (\text{B.23})$$

Le conditionnement de ce processus aux bords du domaine est donc

$$(X_t | X_0 = x_0, X_1 = x_1) = x_0 e^{-\alpha t} + \sigma e^{-\alpha t} \underbrace{B_{e^{2\alpha t}-1} | X_1 = x_1}_{B_{e^{2\alpha t}-1} | B_{e^{2\alpha}-1} = \frac{1}{\sigma}(e^\alpha x_1 - x_0)}. \quad (\text{B.24})$$

D'après l'équation B.10, on a

$$\left(B_{e^{2\alpha t}-1} | B_{e^{2\alpha}-1} = \frac{1}{\sigma}(e^\alpha x_1 - x_0) \right) = B_{e^{2\alpha t}-1} - \frac{e^{2\alpha t}-1}{e^{2\alpha}-1} B_{e^{2\alpha}-1} + \frac{e^{2\alpha t}-1}{e^{2\alpha}-1} \frac{1}{\sigma}(e^\alpha x_1 - x_0). \quad (\text{B.25})$$

Ainsi, on conclut que $(X_t | X_0 = x_0, X_1 = x_1)$ est le processus gaussien :

$$Z_t = x_0 e^{-\alpha t} + \sigma e^{-\alpha t} \left[B_{e^{2\alpha t}-1} - \frac{e^{2\alpha t}-1}{e^{2\alpha}-1} B_{e^{2\alpha}-1} + \frac{e^{2\alpha t}-1}{e^{2\alpha}-1} \frac{1}{\sigma}(e^\alpha x_1 - x_0) \right] \quad (\text{B.26})$$

et donc :

$$Z_t = x_0 e^{-\alpha t} + e^{-\alpha t}(e^\alpha x_1 - x_0) \left(\frac{e^{2\alpha t}-1}{e^{2\alpha}-1} \right) + \sigma e^{-\alpha t} \left[B_{e^{2\alpha t}-1} - \frac{e^{2\alpha t}-1}{e^{2\alpha}-1} B_{e^{2\alpha}-1} \right] \quad (\text{B.27})$$

Il nous faut maintenant déterminer l'espérance et la covariance de ce processus.

Sa moyenne est donc :

$$\begin{aligned} m(t) &= x_0 e^{-\alpha t} + e^{-\alpha t}(e^\alpha x_1 - x_0) \frac{e^{2\alpha t}-1}{e^{2\alpha}-1}, \\ &= \frac{x_0 e^{-\alpha t}(e^{2\alpha}-1)}{e^{2\alpha}-1} + e^{-\alpha t}(e^\alpha x_1 - x_0) \frac{e^{2\alpha t}-1}{e^{2\alpha}-1}, \\ &= (e^{2\alpha}-1)^{-1} [x_0 e^{-\alpha t}(e^{2\alpha}-1) + e^{-\alpha t}(e^\alpha x_1 - x_0)(e^{2\alpha t}-1)], \\ &= (e^{2\alpha}-1)^{-1} [x_0 e^{-\alpha t+2\alpha} - x_0 e^{-\alpha t} + (e^\alpha x_1 - x_0)(e^{\alpha t} - e^{-\alpha t})], \\ &= \frac{e^\alpha}{e^{2\alpha}-1} [x_0 e^{-\alpha(t-1)} - x_0 e^{-\alpha(t+1)} + (x_1 - x_0 e^{-\alpha})(e^{\alpha t} - e^{-\alpha t})], \\ &= (e^\alpha - e^{-\alpha})^{-1} [x_0 (e^{-\alpha(t-1)} - e^{-\alpha(t+1)} - e^{-\alpha+\alpha t} + e^{-\alpha-\alpha t}) + x_1 (e^{\alpha t} - e^{-\alpha t})], \\ &= (e^\alpha - e^{-\alpha})^{-1} [x_0 (e^{\alpha(1-t)} - e^{-\alpha(1-t)}) + x_1 (e^{\alpha t} - e^{-\alpha t})], \\ &= x_0 \frac{\text{sh}\alpha(1-t)}{\text{sh}\alpha} + x_1 \frac{\text{sh}\alpha t}{\text{sh}\alpha}. \end{aligned} \quad (\text{B.28})$$

où sh correspond au sinus hyperbolique. On vérifie bien que $m(t=0) = x_0$ et $m(t=1) = x_1$.

Pour le calcul de la covariance, on sait que $\text{Cov}(Z_t, Z_s) = \mathbb{E}(Z_t Z_s) - \mathbb{E}(Z_t)\mathbb{E}(Z_s)$.

Or on a :

$$\begin{aligned} \mathbb{E}(Z_t Z_s) &= x_0^2 e^{-\alpha(t+s)} + x_0 e^{-\alpha(t+s)} (e^\alpha x_1 - x_0) \left[\frac{e^{2\alpha t} - 1}{e^{2\alpha} - 1} + \frac{e^{2\alpha s} - 1}{e^{2\alpha} - 1} \right] + e^{-\alpha(t+s)} (e^\alpha x_1 - x_0) \frac{e^{2\alpha t} - 1}{e^{2\alpha} - 1} \frac{e^{2\alpha s} - 1}{e^{2\alpha} - 1} \\ &\quad + \sigma^2 e^{-\alpha(t+s)} \left[((e^{2\alpha t} - 1) \wedge e^{2\alpha s} - 1) - \frac{(e^{2\alpha t} - 1)(e^{2\alpha s} - 1)}{e^{2\alpha} - 1} \right], \end{aligned} \quad (\text{B.29})$$

et

$$\mathbb{E}(Z_t) \mathbb{E}(Z_s) = x_0^2 e^{-\alpha(t+s)} + x_0 e^{-\alpha(t+s)} (e^\alpha x_1 - x_0) \left[\frac{e^{2\alpha t} - 1}{e^{2\alpha} - 1} + \frac{e^{2\alpha s} - 1}{e^{2\alpha} - 1} \right] + e^{-\alpha(t+s)} (e^\alpha x_1 - x_0) \frac{e^{2\alpha t} - 1}{e^{2\alpha} - 1} \frac{e^{2\alpha s} - 1}{e^{2\alpha} - 1}. \quad (\text{B.30})$$

Donc,

$$\begin{aligned} \text{Cov}(Z_t, Z_s) &= \sigma^2 e^{-\alpha(t+s)} \left[((e^{2\alpha t} - 1) \wedge e^{2\alpha s} - 1) - \frac{(e^{2\alpha t} - 1)(e^{2\alpha s} - 1)}{e^{2\alpha} - 1} \right], \\ &= \sigma^2 e^{-\alpha(t+s)} \left[((e^{2\alpha t} - 1) \wedge e^{2\alpha s} - 1) - \frac{((e^{2\alpha t} - 1) \wedge (e^{2\alpha s} - 1))((e^{2\alpha t} - 1) \vee (e^{2\alpha s} - 1))}{e^{2\alpha} - 1} \right], \quad (\text{B.31}) \\ &= \sigma^2 e^{-\alpha(t+s)} [(e^{2\alpha t} - 1) \wedge (e^{2\alpha s} - 1)] \left[1 - \frac{(e^{2\alpha t} - 1) \vee (e^{2\alpha s} - 1)}{e^{2\alpha} - 1} \right]. \end{aligned}$$

En prenant de nouveau $s \leq t$, on obtient,

$$\begin{aligned} \text{Cov}(Z_s, Z_t) &= \sigma^2 e^{-\alpha(t+s)} (e^{2\alpha s} - 1) \left(1 - \frac{e^{2\alpha t} - 1}{e^{2\alpha} - 1} \right), \\ &= \sigma^2 e^{-\alpha(t+s)} (e^{2\alpha s} - 1) (e^{2\alpha} - e^{2\alpha t}), \\ &= \frac{\sigma^2}{e^{2\alpha} - 1} e^{-\alpha t} (e^{\alpha s} - e^{-\alpha s}) (e^{2\alpha} - e^{2\alpha t}), \quad (\text{B.32}) \\ &= \frac{\sigma^2}{e^{2\alpha} - 1} e^{-\alpha t} (e^{\alpha s} - e^{-\alpha s}) (e^{2\alpha} - e^{2\alpha t}), \\ &= \frac{\sigma^2}{e^\alpha - e^{-\alpha}} (e^{\alpha s} - e^{-\alpha s}) (e^{\alpha(1-t)} - e^{-\alpha(1-t)}) \end{aligned}$$

D'où,

$$\text{Cov}(Z_s, Z_t) = \sigma^2 \frac{\text{sh}\alpha(s \wedge t) \text{sh}\alpha(1 - t \vee s)}{\text{sh}\alpha} \quad (\text{B.33})$$

et, si $s = t$,

$$\text{Var}(Z_t) = \sigma^2 \frac{\text{sh}\alpha t \text{sh}\alpha(1 - t)}{\text{sh}\alpha}. \quad (\text{B.34})$$

On vérifie bien $\text{Var}(Z_t) = 0$ pour $t = 0$ et $t = 1$.

Définition B.2.2 *Le processus d'Ornstein-Uhlenbeck conditionné aux bords est le processus gaussien*

$$Z_t = m(t) + \sigma e^{-\alpha t} \left[B_{e^{2\alpha t} - 1} - \frac{e^{2\alpha t} - 1}{e^{2\alpha} - 1} B_{e^{2\alpha} - 1} \right], \quad (\text{B.35})$$

où

$$m(t) = x_0 \frac{\text{sh}\alpha(1-t)}{\text{sh}\alpha} + x_1 \frac{\text{sh}\alpha t}{\text{sh}\alpha} \quad (\text{B.36})$$

de covariance,

$$\mathbb{C}ov(Z_s, Z_t) = \sigma^2 \frac{\text{sh}\alpha(s \wedge t) \text{sh}\alpha(1 - t \vee s)}{\text{sh}\alpha} \quad (\text{B.37})$$

Dans le cas particulier où $x_0 = x_1 = 0$, alors on a

$$Z(t) \sim \mathcal{PG} \left(0, \sigma^2 \frac{\text{sh}\alpha(s \wedge t) \text{sh}\alpha(1 - t \vee s)}{\text{sh}\alpha} \right). \quad (\text{B.38})$$

On voit donc que le processus d'Ornstein-Uhlenbeck est, à un changement de temps près et une transformation physique près un pont brownien !

Vieillessement de la carotte

Ce manuscrit s'adressant également aux expérimentateurs afin de les guider dans leurs démarches de planification d'expériences, nous avons fait le choix de présenter dans cette annexe les raisons qui nous ont poussés à mettre de coté les expériences initiales. Nous détaillons comment nous avons observé le changement de comportement de l'échantillon utilisé historiquement, et avançons quelques explications quant à ce changement.

D'un point de vue expérimental, pour chaque nouvel échantillon de catalyseur, appelé "carotte", il est réalisé un premier essai après la stabilisation du système, nommé essai de référence. Cet essai permet aux expérimentateurs de vérifier si le comportement du catalyseur a évolué au cours des expériences (vieillessement, cokage,...).

Il a donc été tout d'abord vérifié que la carotte n'avait pas évolué, afin de pouvoir réutiliser les expériences déjà effectuées. Les figures C.1 présentent les résultats pour CO et HC de l'essai de référence initial et du premier essai effectué. On peut voir que le comportement est proche avec néanmoins une différence sur la décroissance pour le CO. Cependant, il apparaît qu'il est possible de réutiliser cet échantillon pour la nouvelle matrice d'expériences à réaliser.

Malheureusement, après le cinquième essai, l'essai de référence a été reconduit et il est clairement apparu que la carotte avait subi un vieillissement rapide et important. Ce vieillissement peut s'expliquer par les cinq premiers essais réalisés, qui ont des richesses très élevées (proche de 0.9), dont on sait qu'elles favorisent le cokage des catalyseurs, surtout ceux ayant subi un grand nombre d'essais.

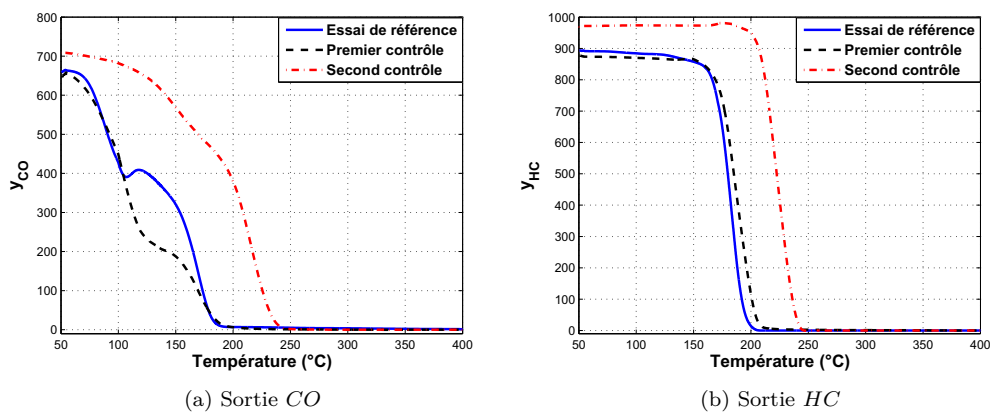


Figure C.1 – Comparaison des essais de contrôle à l'essai de référence

Il s'est alors posé la question de la suite à donner quant aux expériences à réaliser. En effet, deux choix s'offraient à nous. Soit on continuait sur le même échantillon, en sachant que nous ne pouvions pas réutiliser les essais déjà réalisés, et que la carotte était susceptible d'avoir un comportement particulier du fait du grand nombre d'essai

menés dessus. Soit on partait sur une nouvelle carotte sur laquelle nous n'avions pas d'expériences réalisées, ayant un comportement proche de la carotte initiale, et relativement jeune. Nous avons décidé, pour des raisons évidentes, de choisir cette seconde option.

Il a fallu alors redéfinir un plan d'expériences, puisque celui que nous avions déterminé, tenait compte du fait que des expériences étaient disponibles.

Bibliographie

- [1] C. ABRAHAM, P.A. CORNILLON, E. MATZNER-LOBER et N. MOLINARI : Unsupervised curve clustering using b-splines. Rapport technique, ENSAM-INRA-UM II-Montpellier, juin 2002.
- [2] M. ABRAMOWITZ et I. STEGUN : *Handbook of Mathematical Functions*. Dover, 1965. New York.
- [3] H. AKAIKE : A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19:719–722, 1974.
- [4] H. AKAIKE : On the likelihood of a time series model. *Statistician*, 27:215–235, 1978.
- [5] S. AN, W. LIU et S. VENKATESH : Face recognition using kernel ridge regression. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 0:1–7, 2007.
- [6] A. ANTONIADIS : Wavelets in statistics : a review. Rapport technique, Université Joseph Fourier, Grenoble, 1997.
- [7] S. ARRHENIUS : On the reaction velocity of the inversion of cane sugar by acids. *Zeitschrift für Physikalische Chemie* 4, 226, 1889.
- [8] H. AYED : *Analyse expérimentale et modélisation du transfert de matière et du mélange dans une couche cisailée à bulles*. Thèse de doctorat, Institut National Polytechnique de Toulouse, 2007.
- [9] L. BACHELIER : *Théorie de la spéculation*. Thèse de doctorat, Annales scientifiques de l'École normale supérieure, 1900.
- [10] A. BEN HAMZA : Some properties of relaxed median filters. *Digital Signal Processing Proceedings*, 2:957–960, 1997.
- [11] P. BESSE, H. CARDOT et D. STEPHENSON : Autoregressive forecasting of some functional climatic variations. *Scandinavian Journal of Statistics*, 4:673–688, 2000.
- [12] R. BETTINGER : *Inversion d'un système par krigeage - Application à la synthèse de catalyseurs à haut débit*. Thèse de doctorat, Université de Nice-Sophia Antipolis, 2009.
- [13] P.J. BICKEL et D.A. FREEDMAN : Some asymptotic theory for the bootstrap. *Ann. Statist.*, 9(6):1196–1217, 1981.
- [14] J. BIGOT et S. GADAT : Homeomorphic smoothing splines : a new tool for monotonizing an unconstrained estimator in nonparametric regression. *Institut de Mathématiques de Toulouse. Université Paul Sabatier, Toulouse*, 2007.
- [15] N. BRAUNER et M. SHACHAM : Statistical analysis of linear and nonlinear correlation of the arrhenius equation constants. *Chemical Engineering and Processing*, 36(3):243–249, 1997.
- [16] L. BREIMAN et J.H. FRIEDMAN : Estimating optimal transformation for multiple regression and correlation. *Journal of the American Statistical Association*, 80:580–619, 1985.
- [17] L. BREIMAN, J.H. FRIEDMAN, R.A. OLSHEN et C.J. STONE : *Classification and Regression Trees*. Chapman and Hall, 1984. New York.
- [18] B. BRUMBACK et J. RICE : Smoothing spline models for the analysis of nested and crossed sample curves. *Journal of the American Statistical Association*, 93:961–976, 1998.

-
- [19] Z. CAI, J. FAN et R. LI : Efficient estimation and inferences for varying-coefficient models. *Journal of the American Statistical Association*, 95(451):888–902, 2000.
- [20] Z. CAI, J. FAN et Q. YAO : Functional-coefficient regression model for nonlinear time series. *Journal of the American Statistical Association*, 95:941–956, 2000.
- [21] M. CANAUD, F. WAHL, C. HELBERT et L. CARRARO : Design of experiments for smoke depollution from the output of diesel engine. In *ENBIS-EMSE*, 2009.
- [22] C. CANS et C. LAVERGNE : De la régression logistique vers un modèle additif généralisé : un exemple d'application. *Revue de statistique appliquée*, 43(2):77–90, 1995.
- [23] Hervé CARDOT, Frédéric FERRATY et Pascal SARDA : Functional linear model. *Statistics and Probability Letters*, 45(1):11 – 22, 1999.
- [24] R.J. CARROLL, J. FAN, I. GIJBELS et M.P. WAND : Generalized partially linear single-index models. *Journal of the American Statistical Association*, 92:477–489, 1997.
- [25] R.J. CARROLL, D. RUPPERT et A.H. WELSH : Local estimating equations. *Journal of the American Statistical Association*, 93:214–227, 1998.
- [26] R. CHÉDOTAL : *Modélisation du catalyseur 4-voies*. Thèse de doctorat, Université Pierre et Marie Curie, Paris, 2008.
- [27] R. CHEN et R.S. TSAY : Functional-coefficient autoregressive models. *Journal of the American Statistical Association*, 88(421):298–308, 1993.
- [28] C.T. CHIANG, J.A. RICE et C.O. WU : Smoothing spline estimation for varying coefficient models with repeatedly measure dependent variables. *J. Amer. Statist. Assoc.*, 96:605–619, 2001.
- [29] J-P. CHILÉS et P. DELFINER : *Geostatistics : modeling spatial uncertainty*. Wiley, 1999. New York.
- [30] W.S. CLEVELAND et S.J. DEVLIN : Locally-weighted regression : An approach to regression analysis by local fitting. *Journal of the American Statistical Association*, 83(403):596–610, 1988.
- [31] W.S. CLEVELAND, E. GROSSE et M.J. SHYU : Local regression models. In J.M. CHAMBERS et T. HASTIE, éditeurs : *Statistical Models in S*, pages 309–376. Chapman and Hall, New York, 1992.
- [32] B. CONAN-GUEZ : *Modélisation supervisée de données fonctionnelles par perceptron multi-couches*. Thèse de doctorat, Université Paris IX Dauphine, 2002.
- [33] K. COUSSEMENT, D.F. BENOIT et D. Van den POEL : Improved marketing decision making in a customer churn prediction context using generalized additive models. *Expert Systems with Applications*, 37(3):2132–2143, 2010.
- [34] P. CRAVEN et G. WHABA : Smoothing noisy data with spline functions : estimating the correct degree of smoothing by the methods of generalized cross-validation. *Numerische Mathematik*, 31:377–403, 1979.
- [35] N. CRESSIE : The origins of kriging. *Mathematical Geology*, 22(3):239–252, 1990.
- [36] I.D. CURRIE, M. DURBAN et P.H.C. EILERS : Using p-splines to extrapolate two-dimensional poisson data. *Proceedings of 18th International Workshop on Statistical Modelling*, pages 97–102, 2003. Leuven, Belgium.
- [37] C. CURRIN, T. MITCHELL, M. MORRIS et D. YLVISAKER : Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments. *Journal of the American Statistical Association*, 86:953–963, 1991.
- [38] Sébastien DA VEIGA : *Analyse d'incertitudes et de sensibilité - Application aux modèles de cinétique chimique*. Thèse de doctorat, Université Toulouse III, 2007.
- [39] A.C. DAVISON et D.V. HINKLEY : *Bootstrap Methods and their Application*. Cambridge University Press, 1997. Cambridge.
- [40] D. den HERTOEG, J.P.C. KLEIJNEN et A.Y.D. SIEM : The correct kriging variance estimated by bootstrapping. *Center Discussion Paper No. 2004-46*, 2004. Available at SSRN : <http://ssrn.com/abstract=557862>.
-

-
- [41] H. DETTE et R. SCHEDER : Strictly monotone and smooth nonparametric regression for two or more variables. *The Canadian Journal of Statistics*, 34(44):535–561, 2006.
- [42] D. DUPUY, C. HELBERT et J. FRANCO : Dicedesign and diceeval : new r packages for design and analysis of computer experiments. *In preparation*, 2010.
- [43] B. EFRON : Bootstrap methods : Another look at the jackknife. *Annals of statistics*, 7(1):1–26, 1979.
- [44] B. EFRON et R.J. TIBSHIRANI : *An Introduction to the Bootstrap*. Chapman and Hall, 1993. New York.
- [45] P.H.C. EILERS et B.D. MARX : Flexible smoothing with b-splines and penalties. *Statist. Sci.*, 25(1):89–121, 1996.
- [46] P.H.C. EILERS et B.D. MARX : Multivariate calibration with temperature interaction using two-dimensional penalized signal regression. *hemometrics and Intelligent Laboratory Systems*, 66:159–174, 2003.
- [47] A. EINSTEIN : Ueber die von der molekularkinetischen theorie der wärme geforderte bewegung... *Annalen der Physik*, 17:549–560, 1905.
- [48] A. EL HABACHI : Planification d'expériences adaptée aux problèmes de cinétique. applications à la dépollution des fumées en sortie de moteurs. Rapport de stage de l'université Jean Monnet - Saint-Étienne, 2007.
- [49] N. EL KAROUI et E. GOBET : *Les outils stochastiques des marchés financiers : Une visite guidée de Einstein à Black-Scholes*. Editions de l'École Polytechnique, 2011.
- [50] M. EVERINGHAM et A. ZISSERMAN : Regression and classification approaches to eye localization in face images. *Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition*, pages 441–448, 2006.
- [51] J. FAN et I. GIJBELS : Data-driven bandwidth selection in local polynomial fitting : Variable bandwidth and spatial adaptation. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(2):371–394, 1995.
- [52] J. FAN et I. GIJBELS : *Local polynomial modelling and its applications*. Chapman and Hall, 1996. London.
- [53] J. FAN, W. HARDLE et E. MAMMEN : Direct estimation of additive and linear components for high dimensionnal data. *The Annals of statistics*, 26:943–971, 1998.
- [54] J. FAN, Q. YAO et Z. CAI : Adaptive varying-coefficient linear models. *Journal Of The Royal Statistical Society Series B*, 65(1):57–80, 2003.
- [55] J. FAN et J.T. ZHANG : Two-step estimation of functional linear models with applications to longitudinal data. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 62(2):303–322, 2000.
- [56] J. FAN et W. ZHANG : Statistical estimation in varying-coefficient models. *The Annals of Statistics*, 27(5):1491–1518, 1999.
- [57] K.T. FANG, R. LI et A. SUDJINATO : *Design and Modeling for computer experiments*. Chapman and Hall, 2005. Computer Science And Data Analysis Series.
- [58] L. FAVERGEON, M. PIJOLAT, F. VALDIVIESO et C. HELBERT : Experimental rate and monte carlo simulation of the nucleation and growth processes during the dehydration of $\text{li}_2\text{so}_4 \cdot \text{h}_2\text{o}$ single crystals. *Physical Chemistry Chemical Physics*, 7(21):3723–3727, 2005.
- [59] V.V. FEDOROV : *Theory of optimal experiments*. New York Academic Press. Inc., 1972.
- [60] F. FERRATY et P. VIEU : *Nonparametric functional data analysis : methods, theory, applications and implementations*. Springer-Verlag, 2006. London.
- [61] R.A. FISCHER : The arrangement of field experiments. *Journal of the Ministry of Agriculture of Great Britain*, 33:503–513, 1926.
- [62] R.A. FISCHER : Ta system of scoring linkage data, with special reference to the pied factors in mice. *Amer. Nat.*, 80:568–578, 1946.
-

-
- [63] A.I.J. FORRESTER, A.J. KEANE et N.W. BRESSLOFF : Design and analysis of noisy computer experiments. *AIAA Journal*, 44:2331–2339, 2006.
- [64] J. FRANCO : *Planification d'expériences numériques en phase exploratoire pour des codes de calculs simulant des phénomènes complexes*. Thèse de doctorat, École Nationale Supérieure des Mines de Saint-Etienne, 2008.
- [65] J. FRANKE, J.P. KREISS et E. MAMMEN : Bootstrap of kernel smoothing in nonlinear time series. *Bernoulli*, 8(1):1–37, 2002.
- [66] D.A. FREEDMAN : Bootstrapping regression models. *The Annals of Statistics*, 9(6):1218–1228, 1981.
- [67] J.H. FRIEDMAN : Multivariate adaptive regression splines (with discussion). *The Annals of Statistics*, 19:1–141, 1991.
- [68] J. GADSBY, F. J. LONG, P. SLEIGHTHOLM et K. W. SYKES : The mechanism of the carbon dioxide-carbon reaction. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 193(1034):357–376, 1948.
- [69] C. GAETAN et X. GUYON : *Modélisation et statistique spatiales*. Springer, 2008.
- [70] C.W. GEAR : Simultaneous numerical solution of differential-algebraic equations. *IEEE Transactions on Circuit Theory*, 18(1):89–95, Jan. 1971.
- [71] A. GEORGHIADES, P. BELHUMEUR et D. KRIEGMAN : From few to many : illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 23(6):643–660, 2005.
- [72] D. GINSBOURGER : *Métamodèles Multiples pour l'Approximation et l'Optimisation de Fonctions Numériques Multivariées*. Thèse de doctorat, École Nationale Supérieure des Mines de Saint-Etienne, 2009.
- [73] P.J. GREEN et B.W. SILVERMAN : *Nonparametric Regression and Generalized Linear Models : a Roughness Penalty Approach*. Chapman and Hall, 1994. London.
- [74] C. GU et G. WAHBA : Smoothing spline anova with component-wise bayesian confidence intervals. *J. Comput. Graph. Statist.*, 2:97–117, 1993.
- [75] M. GUNZBURGER et J. BURKARDT : Uniformity measures for point sample on hypercubes, 2004. <http://people.sc.fsu.edu/~jburkardt/pdf/ptmeas.pdf>.
- [76] P. HALL et J.L. HOROWITZ : Methodology and convergence rates for functional linear regression. *Annals of statistics*, 35(1):70–91, 2007.
- [77] W. HARDLE, J.L. HOROWITZ et J.P. KREISS : Bootstrap methods for time series. *International Statistical Review*, 71(2):435–459, 2003.
- [78] T. HASTIE : *GAM : Generalized Additive Models*, 2009. R package version 1.01, <http://CRAN.R-project.org/package=gam>.
- [79] T. HASTIE et R. TIBSHIRANI : Generalized additive models (with discussion). *Statist. Sci.*, 1:297–318, 1986.
- [80] T. HASTIE et R. TIBSHIRANI : *Generalized Additive Models*. Chapman and Hall, London, 1990.
- [81] T. HASTIE et R. TIBSHIRANI : Varying-coefficient models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 55(4):757–796, 1993.
- [82] T. HASTIE et R. TIBSHIRANI : *The elements of statistical learning : data mining, inference and prediction*. Springer, 2008.
- [83] S. HAYKIN : *Adaptive Filter Theory*. Prentice Hall, 2002.
- [84] F. HICKERNELL : A generalized discrepancy and quadrature error bound. *Mathematics of computation*, 67:299–322, 1998.
-

-
- [85] A.E. HOERL et R.W. KENNARD : Ridge regression : biased estimation for nonorthogonal problems. *Technometrics*, 12:55–77, 1970.
- [86] D.R. HOOVER, C.O. RICE, J.A. WU et L.P. YANG : Nonparametric smoothing estimates of time varying coefficient models with longitudinal data. *Biometrika*, 85:809–822, 1998.
- [87] J. L. HOROWITZ : Bootstrap methods in econometrics : theory and numerical performance. *Advances in Economics and Econometrics : Theory and Application*, 3:188–222, 1997. Cambridge, Cambridge University Press.
- [88] J.Z. HUANG, C.O. WU et L. ZHOU : Varying-coefficient models and basis functions approximations for the analysis of repeated measurements. *Biometrika*, 89:111–128, 2002.
- [89] C. JIANG et E.B. MARTIN : Functional data analysis for the development of a calibration model for near-infrared data. *18th European Symposium on Computer Aided Process Engineering*, 25:683–688, 2008.
- [90] M.E. JOHNSON, L.M. MOORE et D. YLVIKAKER : Minimax and maximin distance design. *J. Statist. Plann. Inf.*, 26:131–148, 1990.
- [91] J.R. KOEHLER et A.B. OWEN : Computer experiments. *Handbook of statistics*, 13:261–308, 1996.
- [92] J. KOWALIK et M.R. OSBORNE : *Methods for unconstrained optimization problems*. American Elsevier Pub. Co., 1968. New York, USA.
- [93] S.N. LAHIRI : *Resampling Methods for Dependent data*. Springer-Verlag, 2003. New York.
- [94] C. LENG : A simple approach for varying-coefficient models selection. *J. Stat. Planning et Inference*, 139:2138–2146, 2009.
- [95] R. LI, A. SUDJANTO et Z. ZHANG : Modeling computer experiments with functional response. *SAE SP*, 114(6):1661–1666, 2005.
- [96] D.V. LINDLEY : On a measures of the information provided by an experiment. *The Annals of Mathematical Statistics*, 27:986–1005, 1956.
- [97] S.N. LOPHAVEN, H.S. NIELSEN et J. SØNDERGAARD : Aspect of the matlab toolbox dace. Rapport IMM-REP-2002-13, Informatics and mathematical modelling, DTU, 2002. 44 pages.
- [98] S.N. LOPHAVEN, H.S. NIELSEN et J. SØNDERGAARD : A matlab kriging toolbox, version 2.0. Rapport IMM-REP-2002-12, Informatics and mathematical modelling, DTU, 2002. 34 pages.
- [99] Y. LU et S. MAO : Local asymptotics for b-spline estimators of the varying-coefficient model. *Commun. Statist. Theor. Meth.*, 33:1119–1138, 2004.
- [100] Y. LU, R. ZHANG et L. ZHU : Penalized spline estimation for varying-coefficient models. *Commun. Statist. Theor. Meth.*, 37:2249–2261, 2008.
- [101] P. LÉVY : *Processus stochastiques et mouvement brownien*. Gauthier-Villars, 1948. Paris.
- [102] A. LYNN, E. SMID, M. ESHRAGI, N. CALDWELL et D. WOODY : Modeling hydraulic regenerative hybrid vehicles using amesim and matlab/simulink. *Proc. SPIE*, 5805(24), 2005.
- [103] C.L. MALLOWS : Some comments on cp. *Technometrics*, 15:661–675, 1973.
- [104] B.D. MARX et P.H.C EILERS : Multidimensional penalized signal regression. *Technometrics*, 47(1):13–22, 2005.
- [105] G. MATHERON : *Traité de géostatistique appliquée, Tome I*. Numéro 14 de Mémoires du Bureau de Recherches Géologiques et Minières. Technip, 1962. Paris.
- [106] G. MATHERON : Principles of geostatistics. *Economic Geology*, 58:1246–1266, 1963.
- [107] G. MATHERON : *Traité de géostatistique appliquée, Tome II : Le krigeage*. Numéro 24 de Mémoires du Bureau de Recherches Géologiques et Minières. B.R.G.M., 1963. Paris.
- [108] B. MATÈRN : *Spatial Variation*. Springer ; 2nd ed. edition, 1986.
-

-
- [109] N. MATTHESS : *Détermination des lois cinétiques d'épuration des gaz d'échappement automobile par interprétation des courbes de light-off*. Thèse de doctorat, Université Claude Bernard - Lyon1, 2000.
- [110] N. MATTHESS, D. SCHWEICH, B. MARTIN et F. CASTAGNA : From light-off curves to kinetic rate expressions for three-way catalysts. *Plenum Publishing Corporation*, 16/17(1-4), 2001.
- [111] N. McCANN et M. MAEDER : Tutorial : The modelling of chemical processes. *Analytica Chimica Acta*, 647(1):31–39, 2009.
- [112] C.N. MILLET : Développement d'un piège à nox pour la bibliothèque ifp-exhaust. Rapport technique, IFP Energies nouvelles, 2006. D060-QSEF20-rév3.
- [113] T.J. MITCHELL : An algorithm for the construction of d-optimal experimental designs. *Technometrics*, 16:203–210, 1974.
- [114] M. MIZUTA et J. KATO : *Functional Data Analysis and Its Application*. Springer Berlin / Heidelberg, 2007.
- [115] J.M. MOREL et S. SOLIMINI : Variational methods in image segmentation. *Progress in Nonlinear Differential Equations and Their Applications*, 14, 1995.
- [116] W.J. MOROKOFF et E. CAFLISH : Quasi-random sequences and their discrepancies. *SIAM Journal on Scientific Computing*, 15(6):1251–1279, 1994.
- [117] M.D. MORRIS et T.J. MITCHELL : Exploratory designs for computational experiments. *Journal of Statistical Planning and Inference*, 43(3):381–402, 1995.
- [118] M.D. MORRIS, T.J. MITCHELL et D. YLVIKAKER : Bayesian design and analysis of computer experiments : Use of derivatives in surface prediction. *Technometrics*, 35(3):243–255, 1993.
- [119] V.M.R. MUGGIO et G. FERRARA : Fitting generalized linear models with unspecified link function : A p-spline approach. *Computational Statistics and Data Analysis*, 52(5), 2008.
- [120] R. NAAR : *Modélisation du comportement mécanique du béton par approche multi-physique (couplage chimie-mécanique) : application à la réaction alcali-silice*. Thèse de doctorat, École Nationale Supérieure des Mines de Paris, 2009.
- [121] M. NAJIM : *Modélisation, estimation et filtrage optimal en traitement du signal*. Hermès - Lavoisier, 2006.
- [122] D. NERINI et P. MONESTIEZ : A cokriging method for spatial functional data with applications in oceanology. *First International Workshop on Functional and Operatorial Statistics*, 2008.
- [123] D. NERINI, P. MONESTIEZ et C. MANTÉ : Cokriging for spatial functional data. *Journal of Multivariate Analysis*, 101:409–418, 2010.
- [124] H. NIEDERREITER : Low-discrepancy and low-dispersion sequences. *Journal of number theory*, 30:51–70, 1987.
- [125] S.J. ORFANIDIS : *Introduction to Signal Processing*. Prentice-Hall, 1996. Englewood Cliffs, NJ.
- [126] J.S. PARK : Optimal latin hypercube designs for computer experiments. *J. of Statist. Planning and Inference*, 39:95–111, 1994.
- [127] J. PERRIN : *Les atomes*. Félix Alcan, 1913. Paris.
- [128] P. PINCHON : Futures évolutions des motorisations dans l'automobile. *Annales des mines*, novembre 2003.
- [129] A. PINTORE, P. SPECKMAN et C.C. HOLMES : Spatially adaptive smoothing splines. *Biometrika*, 93(1):113–125, 2006.
- [130] J.S. RACINE et C.F. PARMETER : Constrained nonparametric kernel regression : estimation and inference. *disponible sur <http://web.uvic.ca/econ/research/papers/pdfs/racine.pdf>*, 2008.
- [131] J.O. RAMSAY et B.W. SILVERMAN : *Functional data analysis*. New York : Springer, 1997.
- [132] D. REVUZ et M. YOR : *Continuous martingales and Brownian motion*. Grundlehren der mathematischen Wissenschaften. Springer, 3. ed édition, 1999. Paris.
-

-
- [133] S. RICHARDSON : Problèmes méthodologiques dans les études écologiques santé-environnement. *C.R. Acad. ScI Paris, Sciences de la vie/Life Sciences*, 323:611–616, 2000.
- [134] B.D. RIPLEY et F.P. KELLY : Markov point processes. *Journal of The London Mathematical Society*, 15:188–192, 1977.
- [135] D. RUPPERT et R. J. CARROLL : Spatially-adaptive penalties for spline fitting. *Aust.N.Z.J.Statist.*, 42(2): 205–223, 2000.
- [136] D. RUPPERT, S. J. SHEATHER et M. P. WAND : An effective bandwidth selector for local least squares regression. *Journal of the American Statistical Association*, 90(432):1257–1270, 1995.
- [137] C.S. SAMPARA, E.J. BISSET, M. CHMIELEWSKI et D. ASSANIS : Global kinetics for platinum diesel oxidation catalysts. *Ind. Eng. Chem. Res.*, 46(24):7993–8003, 2007.
- [138] T.J. SANTNER, B.J. WILLIAMS et W.I. NOTZ : *The design and analysis of computer experiments*. Springer-Verlag, 2003. New York.
- [139] C. SAUNDERS, A. GAMMERMAN et V. VOVK : Ridge regression learning algorithm in dual variables. In *Proceedings of the 15th International Conference on Machine Learning*, pages 515–521, 1998.
- [140] A. SAVITZKY et J.E. GOLAY : Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36:1627–1639, 1964.
- [141] D.W. SCOTT : *Multivariate Density Estimation : Theory, Practice ans Visualization*. Wiley, 1992. New York.
- [142] C.E. SHANNON : A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423,623–656, 1948.
- [143] J. SHAO et D. TU : *The Jackknife and Bootstrap*. Springer-Verlag, 1995. New York.
- [144] M.C. SHEWRY et H.P. WYNN : Maximum entropy sampling. *Journal of Applied Statistics*, 14:165–170, 1987.
- [145] J.Q. SHI et B. WANG : Curve prediction and clustering with mixtures of gaussian process functional regression models. *Statistics and Computing*, 18:267–283, 2008.
- [146] J.D. SINGER et J.B. WILLET : *Applied longitudinal data analysis*. Oxford University Press, mars 2003. New York.
- [147] M.L. STEIN : *Interpolation of Spatial Data ; Some Theory for Kriging*. Springer, 1999. New York.
- [148] C.J. STONE : Additive regression and other non parametric model. *Annales of Statistics*, 13:689–705, 1985.
- [149] C.J. STONE, M. HANSEN, C. KOOPERBERG et Y.K. TRUONG : Polynomial splines and their tensor products in extended linear modeling. *Ann. Statist.*, 25:1371–1470, 1997.
- [150] D.J. STRAUSS : A model for clustering. *Biometrika*, 62:467–475, 1975.
- [151] D. TAPSOBA, V. FORTIN, F. ANCTIL et M. HACHE : Apport de la technique du krigeage avec dérive externe pour une cartographie raisonnée de l'équivalent en eau de la neige. *Canadian journal of civil engineering*, 32(1):289–297, 2005.
- [152] E. THIÉMARD : *Sur le calcul et la majoration de la discrédance à l'origine*. Thèse de doctorat, École polytechnique fédérale de Lausanne, 2000.
- [153] G. TUTZ et H. BINDER : Generalized additive modeling with implicit variable selection by likelihood-based boosting. *Biometrics*, 62(4):2961–971, 2006.
- [154] G. TUTZ et S. PETRY : Nonparametric estimation of the link function including variable selection. Rapport technique, Department of Statistics, University of Munich, 2010.
- [155] V. VAPNIK : *The Nature of Statistical Learning*. Springer, 1995. New York.
- [156] E. VAZQUEZ : *Modélisation comportementale de systèmes non-linéaires multivariés par méthodes à noyaux et applications*. Thèse de doctorat, Université Paris XI, 2005.
-

-
- [157] S.E. VOLTZ, C.R. MORGAN, D. LIEDERMANN et S.M. JACOB : Kinetic study of carbon monoxide and propylene oxidation on platinum catalysts. *Ind. Eng. Chem. Prod. Res. Dev.*, 12(4):294, 1973.
- [158] G. WAHBA : *Spline Models for observational data*. SIAM, 1990. Philadelphie.
- [159] T.T. WARNOCK : Computational investigations of low discrepancy point sets. In S. K. ZAREMBA, éditeur : *Applications of Number Theory to Numerical Analysis*, pages 319–343. Academic Press, New-York, 1972.
- [160] M. WEST, P.J. HARRISON et H.S. MIGON : Dynamic generalized linear models and bayesian forecasting (with discussion). *J. Amer. Statist. Assoc.*, 80:73–97, 1985.
- [161] N. WIENER : Differential space. *J. Math. Phys.*, 2:131–174, 1923.
- [162] S. WOOD et N. AUGUSTIN : Gams with integrated model selection using penalized regression splines and applications to environmental modelling. *Ecological Modelling*, 157(2-3):157–177, 2002.
- [163] S.N. WOOD : mgcv : Gams and generalized ridge regression for r. *R News*, 1(2):20–25, 2001.
- [164] S.N. WOOD : *Generalized Additive Models : an Introduction with R*. Chapman and Hall-CRC, 2006. Boca Raton.
- [165] C.O. WU, C.T. CHIANG et D.R. HOOVER : Asymptotic confidence regions for smoothing of a varying-coefficient model with longitudinal data. *J. Amer. Statist. Assoc.*, 93:1388–1402, 1998.
- [166] Y. XIONG, W. CHEN, D. APLEY et X. DING : A non-stationary covariance-based kriging method for metamodelling in engineering. *International Journal for Numerical Methods in Engineering*, 71:733–756, 2007.
- [167] A.M. YAGLOM : *An introduction to the theory of stationnary random Functions*. Dover Publications, 2004.
- [168] J. ZHANG et K. MA : Kernel fisher discriminant for texture classification, septembre 2004.
- [169] W. ZHANG et S.Y. LEE : On local polynomial fitting of varying coefficient models. *manuscript submitted for publication*, 1998.
- [170] W. ZHANG et S.Y. LEE : Variable bandwidth selection in varying coefficient models. *J. Multivariate Anal.*, 74(1):116–134, 2000.
-

École Nationale Supérieure des Mines
de Saint-Étienne

NNT : 2011 EMSE 0619

Matthieu CANAUD

Parameter estimation and design of experiments adapted to kinetics problems -
Application for depollution of exhaust smoke from the output of engines

Speciality : Applied mathematics and Statistics

Keywords : Fonctionnal responses, Design of experiments, Kriging, Non parametrics models, Varying coefficient models, Bootstrap, Maximum of entropy

Abstract :

Physico-chemical models designed to represent experimental reality may prove to be inadequate. This is the case of nitrogen oxide trap, used as an application support of our thesis, which is a catalyst system treating the emissions of the diesel engine. The outputs are the curves of concentrations of pollutants, which are functional data, depending on scalar initial concentrations.

The initial objective of this thesis is to propose experiential design that are meaningful to the user. However, the experimental design relying on models, most of the work has led us to propose a statistical representation taking into account the expert knowledge, and allows to build this plan.

Three lines of research were explored. We first considered a non-functional modeling with the use of kriging theory. Then, we took into account the functional dimension of the responses, with the application and extension of varying coefficient models. Finally, starting again from the original model, we developed a model depending on the kinetic parameters of the inputs (scalar) using a nonparametric representation.

To compare the methods, it was necessary to conduct an experimental campaign, and we propose an exploratory design approach, based on maximum entropy.

École Nationale Supérieure des Mines
de Saint-Étienne

NNT : 2011 EMSE 0619

Matthieu CANAUD

Estimation de paramètres et planification d'expériences adaptée aux problèmes de cinétique - Application à la dépollution des fumées en sortie des moteurs

Spécialité: Mathématiques appliquées et statistiques

Mots clefs : Réponses fonctionnelles, Plan d'expériences, Krigeage, Modèles non paramétriques, Modèles à coefficients variables, Bootstrap, Maximum d'entropie

Résumé :

Les modèles physico-chimiques destinés à représenter la réalité expérimentale peuvent se révéler inadéquats. C'est le cas du piège à oxyde d'azote, utilisé comme support applicatif de notre thèse, qui est un système catalytique traitant les émissions polluantes du moteur Diesel. Les sorties sont des courbes de concentrations des polluants, qui sont des données fonctionnelles, dépendant de concentrations initiales scalaires.

L'objectif initial de cette thèse est de proposer des plans d'expériences ayant un sens pour l'utilisateur. Cependant les plans d'expérience s'appuyant sur des modèles, l'essentiel du travail a conduit à proposer une représentation statistique tenant compte des connaissances des experts, et qui permette de construire ce plan.

Trois axes de recherches ont été explorés. Nous avons d'abord considéré une modélisation non fonctionnelle avec le recours à la théorie du krigeage. Puis, nous avons pris en compte la dimension fonctionnelle des réponses, avec l'application et l'extension des modèles à coefficients variables. Enfin en repartant du modèle initial, nous avons fait dépendre les paramètres cinétiques des entrées (scalaires) à l'aide d'une représentation non paramétrique.

Afin de comparer les méthodes, il a été nécessaire de mener une campagne expérimentale, et nous proposons une démarche de plan exploratoire, basée sur l'entropie maximale.