

N⁰ D'ORDRE :?????

UNIVERSITE PARIS XI - UFR SCIENTIFIQUE D'ORSAY

THESE

présentée pour obtenir le grade de

DOCTEUR EN SCIENCES DE L'UNIVERSITE PARIS XI ORSAY

Spécialité : **Mathématiques**

présentée par

Sophie DONNET

**INVERSION DE DONNEES IRMf. ESTIMATION ET
SELECTION DE MODELES**

Rapporteurs : M. Jean-Claude **FORT** Université Toulouse III
M. Marc **HOFFMANN** Université Marne-la-Vallée

Soutenue publiquement le 30 novembre 2006 devant le jury composé de

M. Olivier	FAUGERAS	INRIA - Sophia Antipolis
M. Jean-Claude	FORT	Université Toulouse III
M. Marc	LAVIELLE	Université Paris V
M. Pascal	MASSART	Université Paris XI
M. Jean-Baptiste	POLINE	CEA-SHFJ

Remerciements

Avant toute chose, merci à Marc de m'avoir encadrée au cours de ces trois années, de m'avoir d'abord suivie avec attention puis de m'avoir permis de prendre mon envol. Merci de m'avoir proposé cette collaboration fructueuse avec Adeline et surtout de n'avoir jamais désespéré de voir mes études sur données réelles aboutir !

Merci à Jean-Baptiste de m'avoir ouvert la porte de la neuroimagerie. Merci d'avoir pris de ton précieux temps pour répondre à mes questions, souvent naïves, sur ce sujet compliqué qu'est l'imagerie cérébrale. J'espère que nous pourrions continuer à collaborer pendant encore longtemps.

Merci à Marc Hoffmann et Jean-Claude Fort d'avoir rapporté ce travail.

Merci à tous les membres du jury pour leur présence. Merci à Olivier Faugeras d'avoir trouvé un peu de temps pour assister à cette soutenance et merci à Pascal Massart d'être présent aujourd'hui, six ans après m'avoir dispensé mon premier cours de statistiques.

Cette thèse ne serait pas ce qu'elle est sans la collaboration avec Adeline. Merci pour ton enthousiasme débordant, tes idées bouillonnantes, nous n'avons jamais été aussi performantes que lorsque nous travaillions ensemble, toi, moi et notre éternelle tasse de thé. Il me tarde que nous reprenions notre travail là où nous l'avons laissé. Merci aussi pour tes relectures attentives et tes critiques toujours constructives.

Merci à Valentine Genon-Catalot d'avoir accepté de m'aider sur le problème des diffusions à paramètres aléatoires et de m'avoir accueillie à l'Université Paris V cette année.

Merci aux membres du laboratoire de mathématiques d'Orsay et plus particulièrement à Patrick Jakubowicz et Yves Misiti pour leur aide efficace en informatique. Merci à Valérie Lavigne pour son assistance précieuse au moment des dossiers administratifs : votre efficacité nous simplifie énormément la tâche.

On ne fait pas de thèse de statistiques appliquées sans un bon ordinateur. Je souhaitais donc remercier chaleureusement le mien pour ses bons et loyaux services, bien que "chaleureusement" ne soit pas le terme adéquat. Il a en effet succombé à une panne de ventilateur, sous le soleil vénézuélien.

Ces trois années de thèse ont été l'occasion de m'initier à l'enseignement dans le cadre de mon monitorat. Cette demie-journée d'activité radicalement différente qui ponctuait mes semaines était parfois le bol d'air salvateur. Pour cela, je remercie Sara Brofferio et Arnaud Le Ny avec qui cela a toujours été un plaisir de collaborer.

Certains enseignants ont marqué plus que d'autres mon parcours scolaire/universitaire qui –enfin– s'achève. Je souhaitais donc remercier tout particulièrement M. Bourdaud, mon professeur de mathématiques de collège avec qui le développement de polynômes est devenu l'exercice le plus passionnant du monde. Toute ma reconnaissance va à M. Roger (mon professeur de lycée) pour son enseignement de la rigueur. Enfin, je dois mon arrivée à Orsay aux encouragements de mes enseignants de classe préparatoire M. Potier et M. Dufait.

Parce que je ne suis pas une solitaire et que je n'aurais jamais pu faire une thèse sans avoir autour de moi une bande de joyeux thésards, merci aux habitants du bâtiment 430. Tout d'abord, un merci tout spécial à Ismael pour sa disponibilité permanente pour répondre à toutes sortes de questions mathématiques. Ismael, tes réponses sont d'une clarté incroyable et en m'offrant de ton temps, tu m'en as souvent fait gagner beaucoup.

Merci à Marie T., Bertrand M. et Sylvain pour ces pauses-thé dignes de ce nom : sans vous je n'aurais jamais su qu'il existait du thé au riz ou du thé fumé nauséabond. Merci à Neil et Aurélien pour leurs idées saugrenues de tri-relai, raid et autres courses d'orientation... Merci à Laurent Z. de m'avoir donné une raison de décharger mon stress pendant la rédaction de ce manuscrit (il faut dire que tu avais exagéré ! allez, sans rancune !). Merci aux anciennes doctorantes Estelle et Christine pour leurs conseils avisés en tous genres. Enfin merci à Marie S., Laurent T., Benoît, Guillemette, Nicolas et tous les autres pour ces petits services rendus, pour ces repas animés au CESFO, pour ces pauses-thé parfois à rallonge qui ont fait que finalement ces trois années sont passées très vite.

Parce qu'il n'y a pas que la thèse dans la vie, merci à Alex, Aurore, Laure-Marie et Stéphanie pour ces moments privilégiés que sont les week-end "filles". Merci à Soisick, François, Najette, Jérôme, Jean-Maxime, Adeline, Jérôme... pour ces week-end citrouille, soirée royale ou autres sorties à thème qui ont souvent servi de soupape. Merci à ceux que j'oublie ici.

Merci à la famille Perez-Duarte d'avoir accueillie une statisticienne de plus.

Merci à mes parents, grands-parents et à ma soeur de votre indéfectible soutien même lorsque je pensais être "un boulet pour la science".

Alexis, merci de me faire confiance dans ce que j'entreprends.

Table des matières

Résumé de la thèse	7
I Introduction	9
1 L'IRMf	13
1.1 L'IRMf, une technique d'imagerie cérébrale incontournable	13
1.2 Principe de l'IRMf	14
1.2.1 Fonctionnement de l'IRM en général	14
1.2.2 Application en imagerie cérébrale fonctionnelle	15
1.3 Acquisition des données IRMf	16
2 Analyse statistique des données IRMf	21
2.1 Le modèle linéaire	21
2.1.1 Variantes autour du modèle linéaire et méthodes d'estimation existantes	21
2.1.2 Les apports de cette thèse	24
2.2 Les modèles physiologiques	27
2.2.1 Modèles physiologiques de la littérature et méthodes d'estimation	27
2.2.2 Les apports de cette thèse	29
3 Modèles définis par EDS	33
3.1 État de l'art	34
3.2 Les apports de cette thèse	35
II Inversion sous hypothèse linéaire	37
4 Inversion sous hypothèse de linéarité	41
4.1 Introduction	41
4.2 Définition des modèles statistiques	42
4.2.1 Modélisation des séries temporelles observées avec un unique type de stimuli	42
4.2.2 Modélisation dans le cas de plusieurs types de stimuli	44
4.3 Estimation sous $\mathcal{M}_{\mathcal{P},0}$	45
4.3.1 Contraintes sur \mathbf{h}	45
4.3.2 Algorithme d'estimation de $\theta_{\mathcal{P},0}$, \mathbf{l} et \mathbf{h} sous le modèle $\mathcal{M}_{\mathcal{P},0}$	45
4.4 Estimation sous $\mathcal{M}_{\mathcal{P},1}$	46
4.4.1 Rappels sur l'algorithme Expectation-Maximisation	47
4.4.2 Estimation de $\theta_{\mathcal{P},1}$ par l'algorithme EM	48
4.5 Estimation sous $\mathcal{M}_{\mathcal{P},1}$ avec contrainte de positivité	49
4.5.1 Versions stochastiques de l'algorithme EM	49

4.5.2	Estimation de $\theta_{\mathcal{P},1}$ par l'algorithme SAEM-MCMC dans le cas d'une contrainte de positivité	54
4.6	Calcul de la vraisemblance des modèles	54
4.6.1	Vraisemblance du modèle $\mathcal{M}_{\mathcal{P},0}$	55
4.6.2	Vraisemblances du modèle $\mathcal{M}_{\mathcal{P},1}$ avec et sans contrainte de positivité	55
4.7	Sélection de modèles	56
4.7.1	$\mathcal{M}_{\mathcal{P},0}$ versus $\mathcal{M}_{\mathcal{P},1}$	56
4.7.2	Sélection de la famille \mathcal{P}	57
4.8	Application sur données simulées et réelles	58
4.8.1	Illustration des performances des algorithmes d'estimation sur données simulées	58
4.8.2	Application sur données réelles (1) : les amplitudes sont-elles variables au cours du temps?	62
4.8.3	Application sur données réelles (2) : sélection des types de stimuli	65
4.9	Discussion et conclusion	67

III Estimation dans des modèles définis par équations différentielles ordinaires **71**

5	Estimation in models defined by dynamical systems	75
5.1	Introduction	76
5.2	Models and notations	77
5.2.1	An incomplete data model defined by ODEs	77
5.2.2	Approximation of the regression function	77
5.2.3	An approximate incomplete data model	78
5.3	Simulation of non-observed data with the conditional distribution	79
5.3.1	The Hastings-Metropolis algorithm	79
5.3.2	Convergence	80
5.4	Estimation of parameters	80
5.4.1	Maximum likelihood approach	81
5.4.2	Bayesian approach	82
5.5	Application to nonlinear mixed-effects model	83
5.5.1	Numerical settings	84
5.5.2	Results	84
5.6	Discussion	86
5.7	Appendix A: Convergence of the H-M	88
5.8	Appendix B: The modified LL scheme	90
5.8.1	Principle	91
5.8.2	Convergence of the method	91
6	Modèles physiologiques en IRMf	95
6.1	Introduction	95
6.2	Description du modèle physiologique	97
6.2.1	La composante "Balloon"	97
6.2.2	La composante "rCBF"	99
6.2.3	Le modèle hémodynamique global	100
6.3	Propriétés du modèle physiologique	100
6.3.1	Influence des paramètres sur le signal BOLD	101
6.3.2	Influence de l'amplitude α	103
6.3.3	Mise en évidence de la non-additivité du modèle	105

6.3.4	Conclusion	107
6.4	Estimation paramétrique	108
6.4.1	Etat de l'art	108
6.4.2	Estimation par SAEM-MCMC-ODE	110
6.5	Illustration sur données simulées	111
6.6	Application sur données réelles	115
6.6.1	Description des données et problématique	116
6.6.2	Pré-traitements	118
6.6.3	Implémentation de l'algorithme SAEM-MCMC-ODE	120
6.6.4	Résultats	120
6.7	Discussion	122
 IV Estimation dans des modèles définis par équations différentielles stochastiques		125
7	Parametric inference for SDE models	129
7.1	Data and Model	132
7.1.1	Incomplete data model defined by SDEs	132
7.1.2	Diffusion model	133
7.1.3	Introduction of an approximate statistical model	133
7.2	Maximum Likelihood Estimation on the model \mathcal{M}_h	134
7.2.1	Stochastic versions of the EM algorithm	134
7.2.2	Extension of the SAEM algorithm to the model \mathcal{M}_h	135
7.2.3	Convergence of the SAEM algorithm on the model \mathcal{M}_h	136
7.2.4	Simulation of the non-observed data using a MCMC procedure	136
7.3	Bayesian estimation of the model \mathcal{M}_h	138
7.4	Survey of the error	139
7.5	Theophyllin pharmacokinetic example	141
7.5.1	Pharmacokinetics and Non-linear mixed effects models	141
7.5.2	Simulation study	142
7.5.3	A real data example	144
7.6	Discussion	145
7.7	Appendix A: Proof of theorem 3 (chapter 7)	148
8	Identifiabilité	151
8.1	Position du problème	151
8.2	Cas d'une unique observation	153
8.2.1	Le processus de Ornstein-Uhlenbeck	153
8.2.2	EDS linéaire	156
8.2.3	Modèle de Cox-Ingersoll-Ross	159
8.2.4	Processus brownien géométrique	160
8.3	Cas d'observations multiples	161
8.3.1	Cas du processus d'Ornstein-Uhlenbeck	161
8.3.2	Cas des modèles à effets mixtes	162
8.4	Conclusion et perspectives	167
8.5	Résultats techniques	168
8.5.1	Preuve du lemme 1	168
8.5.2	Croissance 1	169
8.5.3	Distribution Inverse Gamma	169

8.5.4	Correspondance $\Gamma - \chi^2$	169
V	Conclusion Générale et Bibliographie	171
9	Conclusion Générale	173
	Bibliographie	174

Cette thèse est consacrée à l'analyse de données d'Imagerie par Résonance Magnétique fonctionnelle (IRMf). Cette technique permet de suivre le décours temporel des phénomènes hémodynamiques (variations locales du débit sanguin) engendrés par l'activité cérébrale. Le modèle le plus largement utilisé pour l'analyse statistique des données IRMf est un modèle linéaire généralisé mais d'autres modèles physiologiques ont récemment vu le jour.

En se basant sur ces deux familles de modèles, nous avons proposé des modèles statistiques permettant de prendre en compte des variabilités ignorées dans les études classiques et développé des méthodes d'estimation adaptées.

Dans un premier temps, nous avons cherché à tester l'hypothèse selon laquelle l'amplitude des réponses à un type de stimuli donné aurait une variabilité inter-occurrences. Ainsi, deux stimuli de même type (deux clics droits par exemple) engendreraient des réponses d'amplitude différentes. Pour cela, nous avons proposé une variante du modèle linéaire standard permettant de prendre en compte cette variabilité. L'estimation des paramètres de ce modèle est délicate, l'expression de la vraisemblance n'étant pas explicite. Nous avons donc eu recours à l'algorithme *Expectation-Maximisation* et à l'une de ses versions stochastiques (SAEM), particulièrement adaptés dans le cadre de modèles à données non-observées. Nous avons comparé ce modèle au modèle sans variabilité inter-occurrences en proposant un test du rapport des vraisemblances. Nous avons en outre présenté une méthode graphique de sélection des types de stimuli du paradigme expérimental ayant une réelle influence sur le signal.

Pendant le modèle linéaire souffre d'un manque de fondement biologique. Afin d'y remédier, des modèles physiologiques ont été proposés, permettant de lier activité neuronale et activité hémodynamique. Dans ce cadre, le signal mesuré s'écrit comme la somme d'un terme solution d'une équation différentielle ordinaire (EDO) sans solution analytique dépendant d'un paramètre aléatoire, et d'un bruit de mesure supposé gaussien. Comme dans le cas du modèle linéaire, le but est de tester l'existence de variabilités, jusque là ignorées, sur les efficacités neuronales (paramètres quantifiant la façon dont les neurones entraînent une variation du débit sanguin).

Dans ce cadre, nous avons proposé une extension de l'algorithme SAEM adaptée aux modèles définis par EDO intégrant une méthode de résolution numérique du système dynamique. L'estimation des paramètres est réalisée sur un modèle statistique approché dont la fonction de régression est une approximation numérique de la solution de l'EDO. Nous avons montré la convergence des estimateurs des paramètres du modèle statistique approché produits par cet algorithme, tout en contrôlant l'erreur induite par l'approximation de la solution du système différentiel sur l'estimation des paramètres. Nous avons ensuite appliqué cette méthode à un modèle physiologique à la fois sur données simulées et sur données réelles.

Enfin, nous avons introduit un niveau de variabilité supplémentaire, considérant ainsi des modèles définis par équations différentielles stochastiques (EDS) dépendant d'un paramètre aléatoire. Nous avons proposé des algorithmes d'estimation par approche bayésienne et maximum de vraisemblance. Comme précédemment, nous avons montré la convergence des algorithmes proposés et contrôlé l'erreur induite sur l'estimation des paramètres par l'approximation numérique du processus en fonction du pas de ce schéma numérique. La méthode d'estimation par maximum de vraisemblance proposée a alors été appliquée dans le cadre d'un modèle à effets mixtes, sur un jeu de données réelles issu de la pharmacocinétique. Finalement, nous nous sommes intéressés à l'identifiabilité des modèles définis par équations différentielles stochastiques dépendant de paramètres aléatoires.

Première partie

Introduction

Cette partie –composée de trois chapitres– présente au lecteur les bases de l’Imagerie par Résonance Magnétique fonctionnelle (IRMf) ainsi que la façon dont les données considérées dans cette thèse sont acquises.

Au chapitre 2, nous rappelons les principales méthodes d’analyse statistique utilisées par la communauté des neurosciences pour le traitement de ces données et introduisons les résultats obtenus dans cette thèse. Les développements statistiques proposés s’intègrent d’une part dans le cadre d’un modèle linéaire et d’autre part dans le cadre de modèles définis par équations différentielles ordinaires.

Enfin, le troisième chapitre de cette partie introduit les travaux réalisés sur les modèles définis par équations différentielles stochastiques à paramètres aléatoires.

Chapitre 1

L’Imagerie par résonance magnétique fonctionnelle

Sommaire

1.1	L’IRMf, une technique d’imagerie cérébrale incontournable	13
1.2	Principe de l’IRMf	14
1.3	Acquisition des données IRMf	16

1.1 L’IRMf, une technique d’imagerie cérébrale incontournable

Afin de comprendre la structure et le rôle des différentes parties du cerveau, les chercheurs ont eu recours pendant longtemps à des méthodes indirectes. Tout d’abord, la dissection post-mortem de sujets sains, ou ayant souffert de déficits suite à une lésion cérébrale localisée, a permis de faire des déductions quant aux implications fonctionnelles de structures particulières. Ainsi, c’est en observant post-mortem le cerveau d’un patient devenu incapable de parler à la suite d’un accident vasculaire cérébral et qui présentait une zone détruite dans le lobe frontal gauche, que Paul Broca déduisit en 1862 le rôle de cette région dans les processus de langage [21]. Plus récemment, au milieu du siècle dernier, des stimulations électriques appliquées directement sur le cerveau lors de neuro-chirurgies ont permis à Wilder Penfield en premier d’établir les premières cartes cérébrales fonctionnelles chez l’humain [98].

Depuis le début des années 1990, différentes techniques d’imagerie cérébrale ont révolutionné cette quête en permettant de “voir le cerveau penser” en temps réel. Si ces techniques nous montrent ce qui se passe dans le cerveau au cours d’une tâche sans avoir à ouvrir la boîte crânienne, c’est surtout grâce aux progrès de l’informatique et de la détection des rayonnements qui ont eu lieu à la fin du XXème siècle.

On distingue généralement l’*imagerie anatomique* de l’*imagerie fonctionnelle*. La première est conçue pour mettre en valeur les structures cérébrales et tout élément pouvant venir les perturber (tumeurs, hémorragies, caillots ou autres déformations). L’imagerie fonctionnelle mesure pour sa part l’activité du cerveau durant des tâches précises (aussi appelées “événements” ou “stimuli”). Utilisée pour la recherche fondamentale qui vise à mieux comprendre le rôle de nos diverses structures cérébrales, elle est aussi employée pour diagnostiquer des foyers épileptiques ou encore avant des opérations chirurgicales, pour identifier les aires cérébrales au rôle essentiel à garder intactes. Une technique d’imagerie anatomique est cependant très souvent utilisée de pair avec une technique fonctionnelle pour mieux définir l’anatomie et la fonction d’une aire cérébrale chez un individu particulier.

Différentes méthodes d'imagerie concurrentes mais toutefois complémentaires permettent de cartographier le cerveau en activité. Nous décrivons ici rapidement les principales méthodes utilisées. Le lecteur trouvera d'excellentes introductions à l'imagerie cérébrale dans les ouvrages de Mazoyer [88] et Houdé *et al.* [64] ou dans la thèse de Flandin [44].

- La tomographie par émission de positrons (TEP) consiste à mesurer les modifications du débit sanguin au moyen d'un traceur radioactif. Le traceur est préalablement injecté par voie intraveineuse (la méthode est dite "invasive") d'où un inconvénient non-négligeable pour le sujet.
- L'électro-encéphalographie (EEG), mise au point en 1929 par le physiologiste allemand Hans Berger, fut la première méthode de neuroimagerie non invasive. Le matériel nécessaire aux mesures est assez léger puisqu'il consiste en une vingtaine d'électrodes placées sur le scalp du sujet. L'EEG fournit une mesure directe de l'activité électrique. Elle offre une résolution temporelle seulement limitée par la vitesse de l'électronique de mesure mais est relativement peu précise spatialement. L'EEG s'est particulièrement répandue dans la pratique médicale courante –en particulier dans la pratique épiléptologique– à partir des années 1950.
- La magnéto-encéphalographie (MEG) mesure les signaux de même source que l'EEG, soit le courant généré dans les neurones. Mais tandis que l'EEG s'intéresse aux champs électriques, la MEG se concentre sur les champs magnétiques accompagnant le courant. Contrairement à l'EEG, la MEG ne nécessite aucun contact direct entre le corps et les électrodes. Cependant le matériel requis est plus contraignant : en effet le champ magnétique observé étant faible (de l'ordre du femto-Tesla), le détecteur doit être placé dans une cage de Faraday de façon à atténuer les perturbations dues à l'environnement, d'où un coût financier largement supérieur à celui de l'EEG. Par conséquent, cette méthode est peu utilisée.
- L'imagerie par résonance magnétique nucléaire fonctionnelle (IRMf) consiste à mesurer le signal BOLD (Blood Oxygen Level Dependant). Ce signal (aussi appelé "*contrast BOLD*") est lié au rapport entre quantités d'hémoglobine chargée en oxygène et d'hémoglobine pauvre en oxygène (voir au paragraphe 1.2 pour plus de détails). Cette méthode a connu son essor au début des années 1990 car elle présente à la fois une bonne résolution spatiale (quelques millimètres) et une bonne résolution temporelle (quelques secondes). De plus elle offre la possibilité d'acquérir chez un même individu des images multiples et répétées de façon totalement non invasive.
- D'autres techniques d'imagerie optique existent mais demeurent marginales. En effet ces méthodes ne permettent de mesurer que des signaux à la surface du cerveau.

Après plus d'un siècle de recherche sur le sujet, la méthode la plus couramment employée pour cartographier l'activité du cerveau humain *in vivo* est l'IRMf. Dans le reste de cette thèse, nous nous intéressons à ce type de données.

1.2 Principe de l'IRMf

L'IRMf est un cas particulier d'utilisation de l'Imagerie par Résonance Magnétique Nucléaire. Dans ce paragraphe, nous décrivons rapidement son principe de fonctionnement, description nécessaire à la compréhension du signal mesuré. Pour une description plus détaillée, le lecteur se reportera à [64, 117, 124].

1.2.1 Fonctionnement de l'IRM en général

L'imagerie par résonance magnétique (nucléaire) a été inventée dans les années 1970 par Paul Lauterbur [78], co-lauréat du prix Nobel de médecine 2003.

Les images que nous considérons dans cette thèse sont des mesures du signal de résonance magnétique nucléaire des protons de l'hydrogène H de l'eau (l'eau formant plus de 80% du corps). Pour cela, le sujet est allongé sur une table et sa tête est placée dans un champ magnétique très intense (0,1 à

3 Tesla), généré par un électro-aimant formé par des matériaux supra-conducteurs (voir figure 1.1). Dans la suite, nous considérons l'espace muni d'un repère orthonormal (Ox, Oy, Oz) plongé dans



FIG. 1.1 – Sujet s'apprêtant à entrer dans un scanner IRM

un champ magnétique \vec{B}_0 dirigé selon l'axe Oz . Placés dans ce champ magnétique \vec{B}_0 , les protons d'hydrogène de l'eau acquièrent une aimantation \vec{M}_0 proportionnelle à \vec{B}_0 ainsi qu'un mouvement de précession autour de l'axe Oz (de fréquence de rotation proportionnelle à la norme de \vec{B}_0). Le but est de mesurer l'amplitude de \vec{M}_0 . Pour cela, une antenne radio-fréquence est placée autour de la tête du sujet. Celle-ci émet, pendant une durée limitée, une impulsion radio-fréquence qui fait s'écarter le champ magnétique \vec{M}_0 de l'axe Oz . Son angle de bascule dépend de la durée de l'excitation. A la fin de l'impulsion, le vecteur aimantation \vec{M}_0 revient à sa position d'origine par un mouvement de précession – ce retour à l'équilibre s'appelle la *relaxation* – en émettant un champ électromagnétique mesuré (pour plus de détails voir [43]).

1.2.2 Application en imagerie cérébrale fonctionnelle

Le couplage hémodynamique

Le fait qu'il y ait une relation entre l'activité cérébrale et l'apport sanguin –activité hémodynamique– dans les tissus est connu depuis plus de 100 ans : en 1890, Roy et Sherrington [111] conclurent que toute activité fonctionnelle augmente le débit sanguin en réponse à un besoin métabolique. En effet, ils observèrent une modification locale de la couleur du cortex des sujets soumis à des stimulations. Plus précisément, ils constatèrent un passage du rouge vif (sang artériel chargé en oxygène) au bleu violet (sang veineux appauvri en oxygène).

Ainsi, lors d'une activation cérébrale –générée par un stimulus–, la région cérébrale activée doit être alimentée en dioxygène, nécessaire à la métabolisation du glucose, unique source d'énergie du cerveau. Ce dioxygène est transporté sur le lieu de la métabolisation par le sang. De ce fait, un apport en dioxygène requiert une augmentation simultanée du débit sanguin local. Il a cependant été noté que cette significative croissance du débit sanguin local (composé de sang fortement chargé en oxygène) est en réalité de forte amplitude au regard à l'augmentation de la proportion de O_2 métabolisée. Autrement dit, la quantité d'oxygène apportée sur le lieu de l'activation (sous forme d'oxyhémoglobine) est largement supérieure à la quantité d'oxygène réellement consommée. Conséquence de ce déséquilibre, les capillaires sanguins de la zone cérébrale activée sont sensiblement plus oxygénés, d'où une nette diminution de la proportion de déoxyhémoglobine (hémoglobine non chargée en O_2).

Remarque 1. *Les conditions de ce couplage demeurent controversées et de nombreuses questions restent ouvertes sur les liens entre activités neuronales, débit sanguin, volume sanguin, concentration en oxygène et métabolisme.*

La figure 1.2 schématise le couplage hémodynamique.

Le signal BOLD

Ce couplage hémodynamique et ce déséquilibre sont à la base de l'imagerie par résonance magnétique. En effet, il a été observé que l'oxyhémoglobine (chargée en oxygène) est diamagnétique et ne produit pas de perturbation du champ magnétique. Par ailleurs, la déoxyhémoglobine (pauvre en oxygène) est paramagnétique et modifie localement faiblement le champ magnétique entraînant une perturbation du phénomène de relaxation précédemment évoqué. Au final les tissus contenant des vaisseaux riches en déoxyhémoglobine présenteront un signal moindre que les tissus contenant des vaisseaux riches en oxyhémoglobine. Le signal ainsi mesuré est le signal BOLD (*Blood Oxygen Level Dependent*) [76, 94], qui dépend des proportions d'oxyhémoglobine et de déoxyhémoglobine.

Ainsi, au cours d'une expérience IRMf, les mesures obtenues permettent de rendre compte des variations du signal BOLD donc de l'activité hémodynamique. Contrairement à l'EEG, l'IRMf ne permet pas de mesurer directement l'activité neuronale.

Nous représentons sur la figure 1.3 une réponse BOLD de forme standard à un unique stimulus de courte durée. Elle se décompose de la façon suivante. Un pic de signal a lieu environ 5 secondes après le stimulus. Il est suivi d'une décroissance du signal et d'une chute temporaire du signal au dessous du niveau de base à la fin de l'activation ("*undershoot*" final). Le pic peut être précédé d'un creux au début de la réponse ("*initial dip*"). On admet en général que la fonction HRF revient à son niveau de base au bout de 25 secondes. La compréhension de ces différents phénomènes fait encore aujourd'hui l'objet de nombreuses recherches.

1.3 Acquisition des données IRMf

On appelle *session* une série de volumes cérébraux imagés par le scanner IRM au cours du temps ; chacun des volumes cérébraux est appelé *scan*. Un scan est constitué de coupes (figure 1.4), qui sont chacune divisées en petits volumes élémentaires, dénommés *voxels*. Un cerveau complet est constitué d'environ 40000 voxels (de taille de l'ordre de 2mm de côté). Bien que les données soient acquises à intervalles de temps très courts, l'IRM met environ 2 à 3 secondes à parcourir le volume complet du cerveau. Par conséquent, nous obtenons une mesure toutes les 2 à 3 secondes pour un voxel donné.

Les données sont acquises au cours d'un paradigme expérimental. On appelle *paradigme*, une succession de stimuli séparés par des périodes de repos (aussi appelées "tâches de contrôle"). Les stimuli peuvent être de différents types :

- stimuli visuels : damiers horizontaux ou verticaux, présentation de visages connus ou inconnus, présentation de mots...
- stimuli moteurs : clic droit ou gauche sur un bouton placé dans la main du sujet...
- stimuli auditifs : écoute de phrases, calculs mentaux sur ordre oral...

Un exemple de paradigme composé de stimuli visuels est représenté sur la figure 1.5. Dans la littérature, une différence est souvent faite entre les paradigmes dits "événementiels" et les paradigmes "*blocs*". Dans les premiers, les stimuli ou tâches présentés/demandés au sujet sont brefs (de durée inférieure à 3 secondes) et séparés par de courts temps inter-stimuli. Ces paradigmes ont prouvé leur efficacité dans l'étude des relations entre l'évènement (stimulus ou tâche) et l'activité cérébrale mesurée par le signal BOLD. Au contraire, les paradigmes "blocs" sont composés de stimuli de durée supérieure à 3 secondes. Dans la plupart des expériences, les paradigmes sont asynchrones i.e. les événements n'ont pas nécessairement lieu aux instants d'acquisition des données.

Une session dure entre 5 et 10 minutes et est généralement répétée cinq fois pour un même sujet. Dans cette thèse, nous considérons les données issues d'une seule session, en choisissant par exemple celle au cours de laquelle le sujet a le moins bougé dans le scanner.

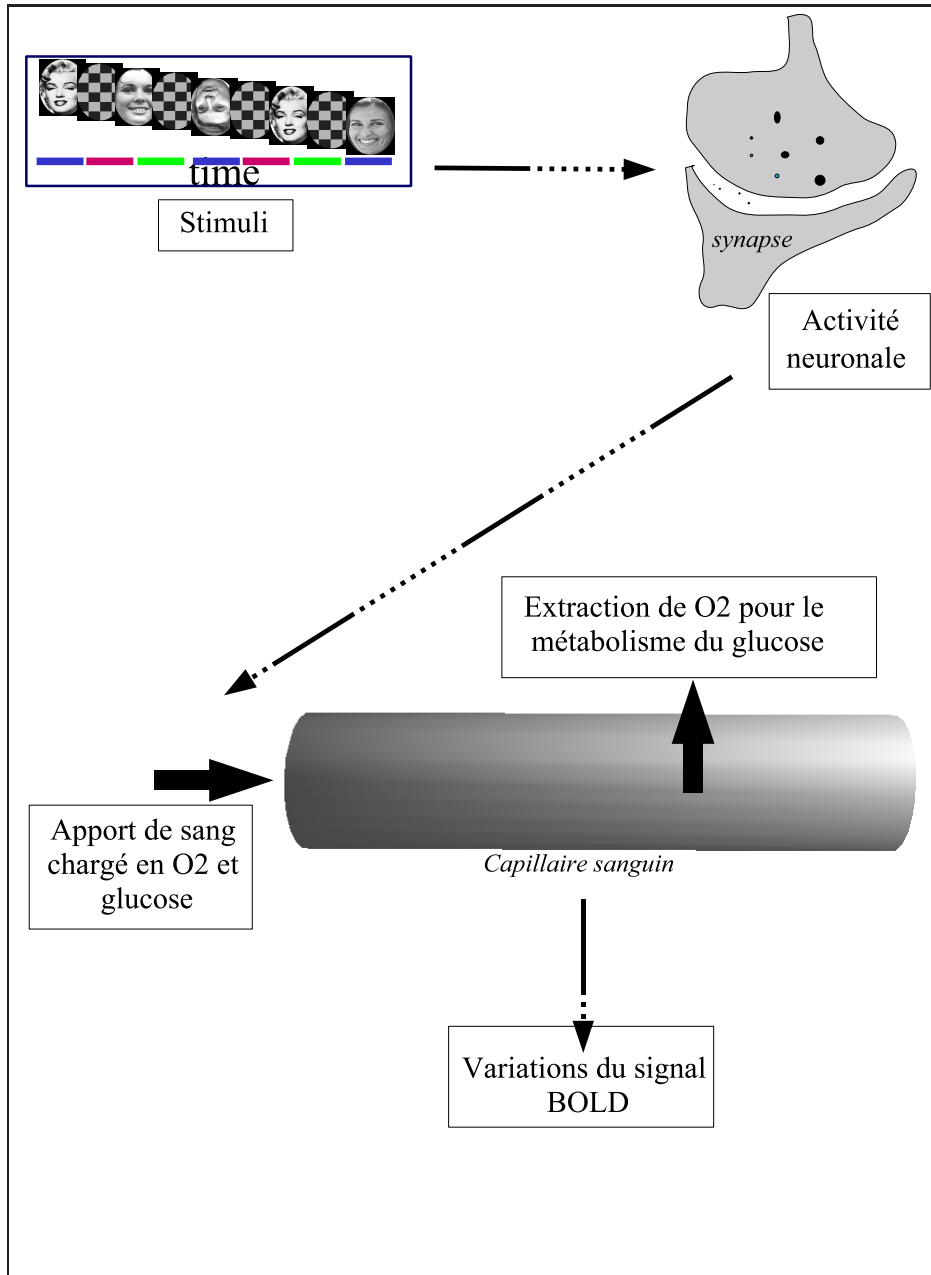


FIG. 1.2 – Schéma récapitulant le couplage activité neuronale - activité hémodynamique

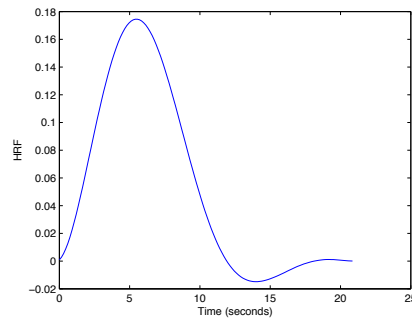


FIG. 1.3 – Réponse BOLD à un unique et très court stimulus

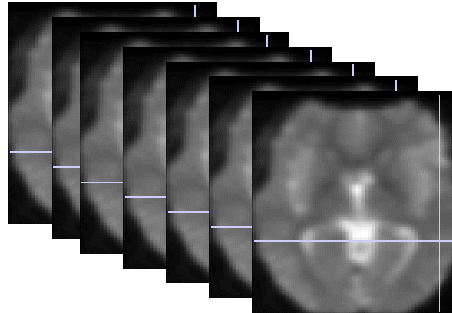


FIG. 1.4 – Acquisition des images cérébrales sous forme de coupes

Les données sur lesquelles nous avons travaillé au cours de cette thèse avaient fait l'objet au préalable de nombreux pré-traitements (pour de plus amples détails voir [44]). Les transformations effectuées ont les objectifs principaux suivants.

- Tout d'abord, il est nécessaire de rephaser les images temporellement. En effet, l'acquisition d'un volume cérébral n'est pas instantanée, elle peut prendre plusieurs secondes en fonction du nombre de coupes ; chacune étant acquise en quelques milli-secondes. Ainsi, si l'on souhaite comparer des signaux venant de coupes différentes il faut tenir compte du fait que ces décours temporels n'ont pas été échantillonnés aux mêmes instants [61, 96].
- Lors d'une acquisition d'images fonctionnelles, les sujets sont encouragés à rester immobiles (ils y sont aidés par divers moyens de contention tels que sparadraps, oreillers...). Cependant, une session dure longtemps et de petits mouvements sont inévitables, par exemple dans le cas de tâches motrices et bien évidemment à cause des mouvements respiratoires. Le but de ce pré-traitement est de réaligner chaque volume cérébral fonctionnel acquis sur le premier de la série de façon à compenser les mouvements du sujet. De nombreuses méthodes ont été développées [110, 109, 129]. Cependant, même après correction, il s'avère que les mouvements du sujet expliquent une grande part de la variance des signaux mesurés. Lorsque ces mouvements ne sont pas corrélés au paradigme, une solution consiste à enlever cette composante par filtrage. Le cas où les mouvements du sujet sont corrélés avec le paradigme est une situation plus difficile à traiter [4, 47].
- La normalisation spatiale vise à placer dans le même espace spatial les données provenant de plusieurs sujets distincts. En effet, afin de pouvoir tirer des conclusions à l'échelle d'une population et non d'un simple sujet, les expériences fonctionnelles font classiquement appel à de nombreux sujets. Pour pouvoir les comparer entre eux, il est alors nécessaire d'établir une transformation spatiale permettant une correspondance unique entre chaque position dans un cerveau et sa position dans un autre cerveau. La littérature dans ce domaine est vaste et nous n'entrerons pas dans les détails ici (voir par exemple [62] pour un état de l'art). Notons toutefois que l'approche

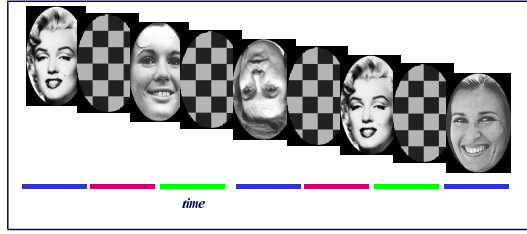


FIG. 1.5 – Paradigme expérimental ici composé de photographies de visages connus et inconnus et de damiers.

géométrique la plus ancienne mais encore couramment employée est la méthode de Talairach et Tournoux [118, 119], englobant le cerveau dans 12 parallélépipèdes auxquels sont appliqués ensuite des facteurs d'échelle. Le système de normalisation proportionnelle de Talairach a été vivement critiqué en raison de son imprécision. La plupart des méthodes proposées aujourd'hui pour placer une image dans la grille proportionnelle définie par Talairach diffèrent de l'approche originale. Elles reposent le plus souvent sur le recalage sur une image cible correspondant à un modèle (*template*) lui-même placé préalablement dans le repère proportionnel de Talairach. Le *Montreal Neurological Institute* (MNI) a à cet effet distribué plusieurs modèles qui sont depuis utilisés comme standard pour la normalisation spatiale par la communauté des neurosciences [42].

De nombreux autres pré-traitements existent. Le lecteur pourra se référer à la thèse de Guillaume Flandin [44] pour une excellente introduction à ce sujet. Dans la suite de cette thèse, nous avons travaillé sur données pré-traitées avec le logiciel SPM (*Statistical Parametric Mapping*)¹ [45] développé par le *Wellcome Department of Cognitive Neuroscience, Functional Imaging Laboratory* de Londres. Ce logiciel d'analyse statistique de données d'IRM fonctionnelles est le plus usité dans la communauté des neurosciences.

Une fois les données acquises sur la totalité du volume cérébral et pré-traitées, nous nous intéressons à un petit nombre de voxels (de l'ordre d'une dizaine) situés dans une région d'intérêt (ROI) et considérons une moyenne du signal généré dans ces voxels. Ainsi, dans la suite, nous notons $(y(t_1), \dots, y(t_n)) \in \mathbb{R}^n$ les observations aux instants (t_1, \dots, t_n) .

Dans le chapitre suivant, nous décrivons les modèles statistiques existants permettant de décrire ces données et les méthodes d'estimation associées.

¹<http://www.fil.ion.ucl.ac.uk/spm/>

Chapitre 2

Analyse statistique des données IRMf

Sommaire

2.1	Le modèle linéaire	21
2.2	Les modèles physiologiques	27

Chaque stimulus ou événement entraîne trois types de phénomènes, d’une part une activité chimique dont nous ne parlerons au cours de cette thèse, d’autre part une activité neuronale (électrique) et enfin une activité “hémodynamique”. Rappelons qu’on appelle activité “hémodynamique” les modifications de concentration en oxygène dans le sang, mesurables par l’IRMf au travers du signal BOLD. Malgré de nombreuses années de recherche, les mécanismes liant ces deux dernières activités demeurent mal connus et restent un des enjeux majeurs de la neuroimagerie. Afin de comprendre ces mécanismes, de nombreux modèles ont été proposés, du plus simple –le modèle linéaire–, aux plus complexes –les modèles physiologiques. Dans ce chapitre, nous présentons différents modèles ainsi que les méthodes d’estimation des paramètres correspondantes.

2.1 Le modèle linéaire

2.1.1 Variantes autour du modèle linéaire et méthodes d’estimation existantes

Définition des modèles

Modèle linéaire classique Le modèle le plus largement utilisé pour l’analyse statistique des données IRMf est le modèle linéaire généralisé. Ce modèle repose sur deux hypothèses essentielles i.e. linéarité et invariance dans le temps de la réponse hémodynamique. Plus précisément,

- si une stimulation entraîne une réponse BOLD dans une région, alors la même stimulation répétée plusieurs fois dans le temps aura pour réponse la somme des réponses à chacune des stimulations (*linéarité*).
- La relation d’entrée-sortie est indépendante du temps. Autrement-dit, un délai (ou retard) dans la stimulation entraîne le même retard dans la réponse (*invariance dans le temps*).

Ces hypothèses sont largement répandues dans la littérature [20, 32, 48, 54, 56, 58, 77, 103].

Les données sont alors modélisées comme le résultat du produit de convolution entre la suite des stimuli et la réponse à un unique et très court stimulus appelée la “*fonction réponse hémodynamique*” (Hemodynamic Response Function - HRF). Plus précisément,

- notons $\mathbf{y} = (y(t_1), \dots, y(t_n))' \in \mathbb{R}^n$, les observations du signal BOLD aux instants d’observation (t_1, \dots, t_n) .

- Soit $h(t) : [0, +\infty[\rightarrow \mathbb{R}$ la fonction HRF. Un exemple de réponse HRF standard a été représenté sur la figure 1.3.
- Soient $\tau_1, \dots, \tau_J \in [0, +\infty[^J$ les instants d'occurrence des stimuli (connus). Définissons alors $x(t) : [0, +\infty[\rightarrow \mathbb{R}$ la suite des stimuli par

$$x(t) = \sum_{j=1}^J \alpha_j S_j(t)$$

avec

$$(\alpha_1, \dots, \alpha_J) \in \mathbb{R}^J \quad \text{et} \quad S_j(t) = \begin{cases} 1 & \text{si } t \in \{\tau_1, \dots, \tau_j\} \\ 0 & \text{sinon} \end{cases}.$$

Les $(\alpha_j)_{j=1 \dots J}$ sont les amplitudes des réponses aux stimulations. Ces paramètres sont à estimer.

- Soit $\varepsilon = (\varepsilon_1 \dots \varepsilon_n)'$ la suite représentant le bruit d'erreur de mesure aux instants d'observation (t_1, \dots, t_n) .

Alors, $\forall i = 1 \dots n$, considérons le modèle linéaire généralisé suivant

$$y(t_i) = \sum_{j=1}^J \alpha_j S_j * h(t_i) + \varepsilon_i. \quad (2.1.1)$$

où $*$ est le produit de convolution. Nous obtenons ainsi l'écriture matricielle suivante :

$$\begin{pmatrix} y(t_1) \\ \vdots \\ y(t_n) \end{pmatrix} = \underbrace{\begin{pmatrix} S_1 * h(t_1) & \dots & S_J * h(t_1) \\ \vdots & \vdots & \vdots \\ S_1 * h(t_n) & \dots & S_J * h(t_n) \end{pmatrix}}_H \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_J \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

$$\mathbf{y} = H\boldsymbol{\alpha} + \boldsymbol{\varepsilon}$$

Dans ce modèle, le but est d'estimer $\boldsymbol{\alpha}$, h et la variance du bruit de mesure.

Remarque 2. Ce modèle est implémenté dans SPM, logiciel d'analyse statistique de données d'IRM fonctionnelles le plus usité dans la communauté des neurosciences est SPM (Statistical Parametric Mapping) [45].

Variante du modèle linéaire : introduction d'une dérive basse fréquence Lors d'expériences de neuroimagerie, les données brutes sont contaminées par une dérive basse-fréquence en grande partie due à des artefacts physiologiques : respiration, variations lentes du rythme cardiaque, dérive de la machine, etc [135]. En général, les données ont fait l'objet d'un pré-traitement –par le logiciel SPM par exemple– permettant de supprimer en partie ces artefacts. Cependant, il est aussi possible et intéressant de modéliser simultanément la réponse BOLD et la dérive basse fréquence [31, 82]. La dérive basse fréquence est alors modélisée par une famille de fonctions orthonormales $(P_q)_{q=1 \dots Q}$, par exemple $P_q(t) = \sqrt{\frac{2}{n}} \cos\left(\frac{\pi(2t+1)(q-1)}{2n}\right)$. Le nombre Q de ces fonctions dépend de la fréquence minimale f_{min} attribuable à la dérive et peut par exemple être défini par : $Q = [2nTRf_{min}] + 1$. f_{min} est généralement l'ordre de $\frac{1}{180}$ à $\frac{1}{120}$ Hertz, cependant cette question reste ouverte. Chaque fonction est alors pondérée par une constante l_q à estimer. Dans ce cas, l'expression (2.1.1) devient

$$y(t_i) = \sum_{q=1}^Q l_q P_q(t_i) + \sum_{j=1}^J \alpha_j S_j * h(t_i) + \varepsilon_i \quad (2.1.2)$$

Soit l'écriture matricielle suivante :

$$\begin{pmatrix} y(t_1) \\ \vdots \\ y(t_n) \end{pmatrix} = \begin{pmatrix} S_1 * h(t_1) & \dots & S_J * h(t_1) & P_1(t_1) & \dots & P_Q(t_1) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ S_1 * h(t_n) & \dots & S_J * h(t_n) & P_1(t_n) & \dots & P_Q(t_n) \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_J \\ l_1 \\ \vdots \\ l_q \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

Estimation des paramètres

Estimation des paramètres dans le cas où la fonction h est supposée connue Dans un premier temps, la fonction h a été supposée connue de forme standard (comme c'est le cas dans le logiciel SPM dans lequel la fonction HRF de référence est construite à partir de la différence de deux fonctions Gamma). Dans ce cas, si le bruit d'observation est supposé gaussien centré de variance $\sigma^2 : \varepsilon_i \sim_{i.i.d} \mathcal{N}(0, \sigma^2)$, $\forall i = 1 \dots n$, alors les estimateurs des moindres carrés et du maximum de vraisemblance du vecteur α sont les mêmes, solutions des équations normales : $H' \mathbf{y} = H' H \alpha$ (où X' est la matrice transposée de X). L'estimateur de σ^2 est alors : $\hat{\sigma}^2 = \|\mathbf{y} - H \hat{\alpha}\|^2$.

Cependant, il a été montré que les séries temporelles de données IRMF sont corrélées dans le temps [22], cette corrélation n'étant pas prise en compte dans le cas où $\varepsilon \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_n)$. Les origines de cette auto-corrélation ne sont que partiellement connues, certains auteurs ont évoqué le principe d'acquisition ou des bruits physiologiques, etc.

Considérons alors une matrice de variance-covariance V non-diagonale, i.e. $\varepsilon \sim \mathcal{N}(0, V)$ avec $V \in \mathcal{G}\ell_n(\mathbb{R})$. L'estimateur du maximum de vraisemblance s'écrit :

$$\hat{\alpha} = (H' V^{-1} H)^{-1} H' V^{-1} \mathbf{y}$$

où X^- représente la matrice pseudo-inverse de X . L'estimateur obtenu est non biaisé. Cependant, il nécessite la connaissance de la matrice de covariance ce qui est inenvisageable en pratique. L'estimation de cette matrice est délicate car il faudrait alors estimer plus de paramètres qu'il n'y a d'observations. Friston *et al.* [46] soulignent que la technique la plus efficace pour prendre en compte la corrélation du bruit consiste en une approche par "blanchiment" (*whitening*). Plus précisément, la matrice de variance V est estimée en imposant une contrainte forte sur sa structure (un modèle auto-régressif d'ordre 1 par exemple). Cependant, une mauvaise approximation des auto-corrélations peut engendrer un biais sur l'estimation de α [46].

Remarque 3. *Dans le cas d'un modèle incluant une dérive basse fréquence, les estimateurs des moindres carrés ou du maximum de vraisemblances peuvent être obtenus comme précédemment.*

Estimation de la forme de la réponse HRF Les méthodes d'estimation de α et de σ^2 présentées précédemment supposent la fonction HRF h connue et utilisent une fonction standard. Cependant, cette fonction peut varier fortement en forme d'un individu à l'autre, d'une région d'intérêt à l'autre ou d'un stimulus à l'autre. Une estimation précise de la HRF est essentielle pour une meilleure compréhension des activations cérébrales, c'est pourquoi elle a fait l'objet d'intenses recherches ces dernières années.

- **Estimation paramétrique**

Les premières méthodes d'estimation de la réponse HRF comme fonction de transfert du système linéaire et invariant dans le temps précédemment décrit, ont été des méthodes paramétriques [32, 48, 77, 103]. Dans ce cadre, la fonction HRF h est supposée appartenir à une famille de fonctions $\mathcal{H} = \{h_\theta, h_\theta : [0, \infty[\rightarrow \mathbb{R}, \theta \in \Theta\}$, par exemple des gaussiennes, des fonctions Gamma

ou des fonctions de Poisson... Le paramètre θ de h_θ est estimé en général par moindres carrés (par l'algorithme de Levenberg-Marquardt par exemple [32]) mais aussi dans un cadre Bayésien [77], de façon à prendre en compte le délai de la HRF (l'instant du pic). Cependant, ces méthodes sont peu satisfaisantes puisqu'elles ne peuvent prendre en compte les variations de la forme de HRF dans le cerveau.

- **Estimation non-paramétrique**

Récemment des méthodes non paramétriques ont été proposées dans un cadre bayésien [31, 86]. Plus précisément, définissons une fine grille d'échantillonnage temporel $(s_m)_{m=0,\dots,M}$ avec $s_m = m\Delta_e$ contenant à la fois les instants d'observation $(t_1 \dots t_n)$ et les instants d'occurrence des stimuli $(\tau_1 \dots \tau_J)$. Δ_e définit le pas de grille d'échantillonnage temporel. Soit \mathbf{h} le vecteur des valeurs de la fonction HRF h aux instants de la petite grille d'échantillonnage temporelle :

$$\mathbf{h} = (h(0), h(\Delta_e), h(2\Delta_e), \dots, h(L\Delta_e))'$$

où $L\Delta_e \approx 25$ secondes. Dans un cadre non paramétrique, le but est d'estimer le vecteur \mathbf{h} . Le cadre bayésien permet d'inclure une information a priori, issue du savoir biologique, notamment une condition de régularité sur la réponse HRF \mathbf{h} . Par exemple, dans Marrélec *et al.* [86], la distribution a priori est choisie la forme suivante pour le vecteur \mathbf{h} :

$$\mathbf{h} \sim \mathcal{N}(0, \tau \mathbf{R})$$

avec $\mathbf{R} = (\mathbf{D}_2 \mathbf{D}_2')^{-1}$ et

$$\mathbf{D}_2 = \begin{bmatrix} 5 & -4 & 1 & 0 & \dots & & & 0 \\ -4 & 6 & -4 & 1 & 0 & & & \\ 1 & -4 & 6 & -4 & 1 & 0 & & \\ 0 & 1 & -4 & 6 & -4 & 1 & 0 & \vdots \\ & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \\ \vdots & & 0 & 1 & -4 & 6 & -4 & 1 & 0 \\ & & & 0 & 1 & -4 & 6 & -4 & 1 \\ & & & & 0 & 1 & -4 & 6 & -4 \\ 0 & \dots & & & 0 & 1 & -4 & 5 \end{bmatrix}$$

la matrice \mathbf{D}_2 représentant la différentielle seconde discrète.

Ces méthodes initialement proposées par Goutte, Nielsen ou Marrélec [58, 85, 92] dans le cadre de paradigmes périodiques ou événementiels synchronisés, ont été étendues par Ciuciu *et al.* [31] dans un cadre de paradigmes plus généraux (paradigmes asynchrones). De plus, dans ce travail, les auteurs estiment une forme de HRF distincte par type de stimuli. Ceci est fait en utilisant un algorithme Expectation Conditional Maximisation [89], variante de l'algorithme *Expectation Maximisation* (EM). Récemment, Makni et ses collaborateurs ont émis l'hypothèse que la HRF pourrait être constante en forme au sein d'une région d'intérêt donnée mais que son amplitude pourrait varier en fonction des voxels considérés dans cette région d'intérêt [82]. Le vecteur \mathbf{h} est alors estimé dans un cadre bayésien en implémentant l'échantillonneur de Gibbs. L'algorithme de Gibbs et l'algorithme EM seront décrits en détails au chapitre 4.

2.1.2 Les apports de cette thèse

Introduction d'une variabilité inter-occurrences

Makni *et al.* [82] ou encore Ciuciu *et al.* [31] ont introduit des modèles permettant de prendre en compte une importante variabilité de la fonction HRF, par exemple d'un voxel à l'autre [82] ou encore

entre les types de stimuli [31]. Cependant, la stationnarité au cours des occurrences est le plus souvent supposée : autrement-dit, deux stimuli de même type (par exemple deux clics droits) engendrent une fonction HRF constante *en forme et en amplitude*.

Récemment, il a été suggéré par Duann *et al.* [39] que cette hypothèse pourrait être remise en cause. Dans cet article, les auteurs montrent de façon convaincante que la réponse hémodynamique observée pourrait varier dans le temps. Les travaux de Arieli *et al.* [6] –utilisant l’imagerie optique ou électro-physiologique– ont aussi suggéré que l’amplitude des réponses peut dépendre de l’activité cérébrale déjà en cours, ou d’effets variables tels que l’attention du sujet. Lu *et al.* [81] proposent un cadre de travail général permettant de prendre en compte les variations inter-occurrences de la HRF en forme et en amplitude. Ces derniers utilisent un F-test pour montrer que le modèle le plus complexe est en général le plus approprié. Cependant, la procédure d’optimisation utilisée est complexe et soumise à de fortes contraintes sous lesquelles la validité de la sélection de modèles par un F-test n’a pas encore été démontrée. De plus, leur approche se limite aux paradigmes dans lesquels les stimuli doivent être synchronisés avec les instants d’observation. Enfin, dans ces travaux, la HRF doit ressembler à la HRF standard (voir figure 1.3, chapitre 1) alors qu’il a été prouvé par Aguirre *et al.* [2] dans des travaux antérieurs que la forme de la réponse HRF peut varier fortement d’un sujet à l’autre, d’une région cérébrale à l’autre ou d’un type d’évènement à l’autre. Dans la première partie de cette thèse, nous proposons un modèle statistique simple, permettant de modéliser une variabilité inter-occurrences pour un même type de stimuli.

Plus précisément, notons h la fonction HRF. Supposons que le paradigme contienne K types de stimuli. Soit $\tau_{k,j}$ l’instant d’occurrence du j -ième stimulus de type k . Notons $\alpha_{k,j}$ l’amplitude de la réponse au j -ième stimulus de type k . Soit J_k le nombre d’occurrences des stimuli de type k . $J = \sum_{k=1}^K J_k$ est le nombre total de stimuli. Alors nous proposons de définir le signal d’entrée suivant :

$$x(t) = \sum_{k=1}^K \sum_{j=1}^{J_k} \alpha_{k,j} S_{k,j}(t)$$

avec

$$(\alpha_{1,1}, \dots, \alpha_{K,J_K}) \in \mathbb{R}^J \quad \text{et} \quad S_{k,j}(t) = \begin{cases} 1 & \text{si } t \in \{\tau_{1,1} \dots \tau_{K,J_K}\} \\ 0 & \text{sinon} \end{cases}.$$

Dans ce modèle, les amplitudes $\alpha_{k,j}$ sont supposées avoir une variabilité inter-occurrence. Le choix le plus simple en l’absence de toute autre information est de supposer leur distribution gaussienne : $\forall k \in \{1, \dots, K\}, \forall j \in \{1, \dots, J_k\}$

$$\alpha_{k,j} \sim_{i.i.d} \mathcal{N}(\mu_k, \omega_k).$$

Alors, nous considérons le modèle statistique suivant : $\forall i = 1 \dots n$,

$$\begin{cases} y(t_i) &= \sum_{k=1}^K \sum_{j=1}^{J_k} \alpha_{k,j} h(t_i - \tau_{k,j}) + \varepsilon_i, \\ \alpha_{k,j} &\sim i.i.d \mathcal{N}(\mu_k, \omega_k) \\ \varepsilon_i &\sim i.i.d \mathcal{N}(0, \sigma^2) \end{cases} \quad (\mathcal{M}_1)$$

Nous devons estimer les paramètres de la distribution des amplitudes et la variance de bruit ($\theta = (\mu_k, \omega_k, \sigma^2)_{k=1 \dots K}$) ainsi que la fonction HRF h .

Remarque 4. *Le bruit d’observation est supposé gaussien de matrice de variance-covariance $\sigma^2 \mathbf{I}_n$. Une telle hypothèse peut sembler simplificatrice puisque il est bien connu que les séries temporelles de données IRMF sont corrélées dans le temps [22]. D’autres modèles de bruit plus précis tels des modèles auto-régressifs ont été proposés [130, 131, 132]. Cependant, l’influence des différentes structures de corrélation de bruit a été étudiée [85], et les auteurs ont montré que dans le cadre d’une estimation bayésienne, la distribution du terme d’erreur n’influence pas l’estimation de la HRF de façon significative. Le même résultat a été souligné par Birn et al. [16].*

Remarque 5. *Comme précédemment, une dérive basse fréquence peut être modélisée simultanément.*

Estimation des paramètres du modèle et de la réponse HRF h

Nous cherchons à la fois, à estimer les paramètres de la distribution des amplitudes $(\mu_k, \omega_k)_{k=1\dots K}$, σ^2 et la fonction HRF h . Pour cela, comme dans [31], nous nous plaçons dans un cadre non-paramétrique et estimons le vecteur

$$\mathbf{h} = (h(0), h(\Delta_e), h(2\Delta_e), \dots, h(L\Delta_e))'$$

où $L\Delta_e \approx 25$ secondes. Afin d'estimer \mathbf{h} , nous utilisons une méthode de minimisation des moindres carrés sous contraintes, permettant de prendre en compte un a priori biologique sous la forme d'une contrainte de régularité. Les paramètres de la distribution des amplitudes $(\mu_k, \omega_k)_{k=1\dots K}$ sont estimés par maximisation de la vraisemblance au moyen de l'algorithme itératif Expectation-Maximisation (EM) proposé par Dempster *et al.* [36] (décrit au chapitre 4, paragraphe 4.4.1), en considérant les amplitudes α comme des données non-observées.

Pour certains types de stimuli, dans certaines régions d'intérêt, il peut être pertinent d'imposer une contrainte de positivité sur les amplitudes $\alpha_{k,j}$. Dans ce cas, la distribution des amplitudes est naturellement une loi log-normale dont il faut estimer les paramètres :

$$\log \alpha_{k,j} \sim_{i.i.d} \mathcal{N}(\mu_k, \omega_k)$$

Dans ce cadre, l'implémentation de l'algorithme EM n'est pas envisageable (l'étape E n'ayant pas d'expression explicite), nous avons donc recours à une version stochastique de cet algorithme suggérée par Kuhn et Lavielle [72] pour estimer les paramètres de la distribution des amplitudes. Nous rappelons le principe et les résultats de convergence de cet algorithme au chapitre 4, paragraphe 4.5.1.

Sélection de modèles

Dans cette thèse, nous nous intéressons à deux questions de sélection de modèles.

- D'une part nous cherchons à valider l'hypothèse de variabilité inter-occurrences. Pour cela, nous testons le modèle précédemment décrit contre le modèle plus simple dans lequel toutes les réponses à un même type de stimuli ont la même amplitude. Ce test d'hypothèse est fondé sur le rapport des vraisemblances.
- D'autre part, dans toutes les méthodes proposées dans la littérature, tous les types de stimuli présents dans le paradigme expérimental sont inclus dans le modèle. Dans cette thèse, nous considérons que si un certain type de stimuli ne génère pas de signal significatif, celui-ci ne doit pas être inclus dans le modèle. Ainsi nous cherchons à sélectionner les types de stimuli ayant une réelle influence sur le signal. Il est intéressant de penser ceci en terme de sélection de modèles, recherchant le meilleur modèle parmi les 2^K modèles, où K est le nombre de types de stimuli différents possibles. Nous proposons une méthode graphique fondée sur la vraisemblance des modèles permettant de sélectionner les types de stimuli influant de façon significative le signal.

Applications sur données réelles et simulées

Les méthodes d'estimation et de sélection de modèles précédentes sont appliquées sur données simulées afin d'illustrer leur validité (voir chapitre 4, paragraphe 4.8). En outre, le modèle statistique proposé (autorisant une variabilité inter-occurrences) est testé sur données réelles à grande échelle. La validation du modèle sur plusieurs sujets et dans plusieurs régions cérébrales a fait l'objet d'un article co-écrit avec Marc Lavielle¹ et Jean-Baptiste Poline^{2,3}, paru en juillet 2006 dans la revue **NeuroImage** (*Volume 31, Issue 3, 1 July 2006, Pages 1169-1176*)

¹Laboratoire de mathématiques, Université de Paris-Sud, 91405 Orsay, France

²SHFJ/CEA 4 Place du Général Leclerc, 91406 Orsay, France, IFR 49

³Institut d'Imagerie Neurofonctionnelle, Paris, France

La mise en évidence d'une nouvelle variabilité entre les types de stimuli a de fortes implications sur les futures études de données IRMF, notamment dans la comparaison de réponses entre elles. Cependant, l'origine de cette variabilité reste inexpliquée et le manque de fondement biologique du modèle linéaire ne favorise pas son explication.

2.2 Les modèles physiologiques

Bien que de nombreuses variantes du modèle de convolution aient été proposées (amplitudes ou formes variables d'un sujet à l'autre, d'un voxel à l'autre, estimation paramétrique ou non paramétrique de la HRF...), aucune d'entre elles ne prend en compte les mécanismes physiologiques à l'origine des variations du signal BOLD. Plus précisément, aucun lien avec la dynamique neuronale locale sous-jacente n'est évoqué. De plus, les modèles linéaires, comme leur nom l'indique, ne modélisent pas les non-linéarités du signal BOLD. Cependant, des preuves de l'existence de tels phénomènes non linéaires dans la réponse ont été fournies empiriquement [12, 66]. Ainsi, l'introduction de modèles physiologiques permettant de prendre en compte ces non-linéarités et liant les phénomènes neuroaux aux phénomènes hémodynamiques est nécessaire. Les enjeux de ces modèles se trouvent donc d'une part au niveau de la compréhension des mécanismes cérébraux et d'autre part au niveau des reconstructions des données.

2.2.1 Modèles physiologiques de la littérature et méthodes d'estimation

Le premier modèle physiologique liant les phénomènes neuroaux aux phénomènes vasculaires fut proposé par Buxton en 1997 et 1998 et appelé "*Balloon model*" [23, 26]. Mandeville [83] proposa lui aussi un modèle biophysique appelé "*Windkessel model*" en 1998. Ces modèles ont été validés par des expériences physiologiques. Ils décrivent comment, suite à un stimulus, le débit sanguin local $f_{in}(t)$ augmente, conduisant ainsi à l'augmentation locale du volume $v(t)$ d'un capillaire sanguin (aussi appelé "ballon veineux"). Le sang affluant dans le ballon veineux est fortement chargé en oxygène, oxygène qui n'est que partiellement consommé. Ceci entraîne une diminution locale de la concentration en déoxyhémoglobine $q(t)$ d'où une variation du signal BOLD.

Ce modèle originel a été étendu par Friston [50] de façon à inclure les effets de l'activité neuronale $u(t)$ sur un signal vaso-dilatateur auto-régulé $s(t)$, permettant finalement de lier l'activité neuronale à l'activité hémodynamique. Un schéma résumant ce modèle global est représenté sur la figure 2.1. Ce modèle est du type entrée-sortie comportant un vecteur d'état $\mathbf{x} = (s, f_{in}, v, q)'$: $[0, +\infty[\rightarrow \mathbb{R}^4$. Les interactions de ces variables d'état sont décrites au cours du temps par le système d'équations différentielles suivant :

$$\begin{cases} ds(t)/dt = \alpha \cdot u(t) - \frac{s(t)}{\tau_s} - \frac{f_{in}(t)-1}{\tau_f} \\ df_{in}(t)/dt = s(t) \\ dv(t)/dt = \frac{1}{\tau_0} (f_{in}(t) - v(t)^{1/\alpha_w}) \\ dq(t)/dt = \frac{1}{\tau_0} \left(\frac{f_{in}(t)}{E_0} (1 - (1 - E_0)^{1/f_{in}}) - q(t)v(t)^{1/\alpha_w-1} \right) \end{cases} \quad (2.2.3)$$

avec pour conditions initiales :

$$\begin{cases} s(0) = 0 \\ f_{in}(0) = 1 \\ v(0) = 1 \\ q(0) = 1 \end{cases}$$

Ce système est paramétré par cinq paramètres hémodynamiques $\mathbf{a} = (\tau_s, \tau_f, \tau_0, \alpha_w, E_0)$. Les trois premières constantes sont des constantes de temps, α_w conditionne la vitesse à laquelle le sang est expulsé du ballon et E_0 est la fraction d'oxygène extraite du sang.

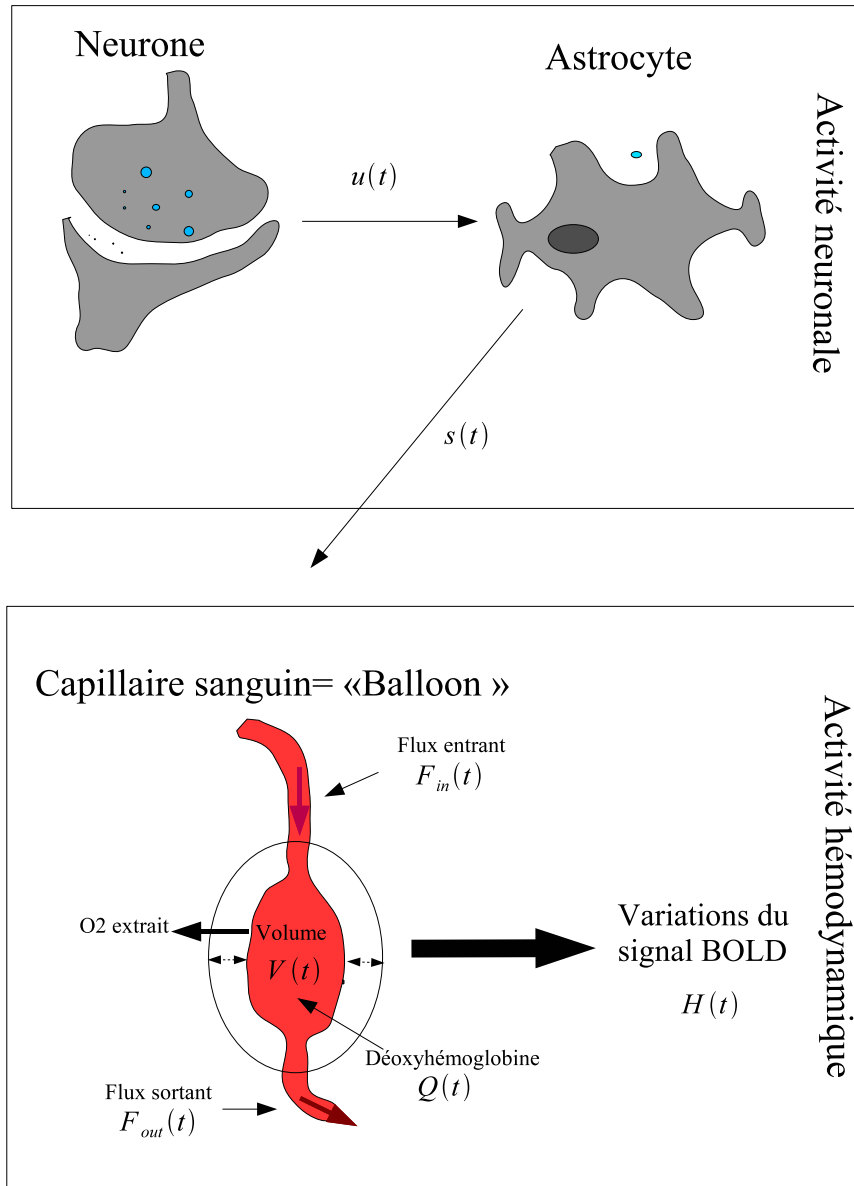


FIG. 2.1 – Schéma du “Balloon Model” [23, 50]

Une fois le système (2.2.3) résolu, le signal BOLD est alors modélisé par la fonction suivante H (2.2.4) :

$$H \circ \mathbf{x}(t, \boldsymbol{\alpha}, \mathbf{a}) = V_0 (k_1(1 - q(t)) + k_2(1 - v(t))) \quad (2.2.4)$$

où k_1 et k_2 sont des constantes. L'entrée de ce modèle est le signal vaso-dilatateur auto-régulé (s) déclenché par les réponses neuronales $u(t)$ aux stimuli du paradigme. Ces réponses neuronales sont pondérées par des paramètres "efficacités" $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_J)$ où α_j est l'amplitude de la réponse neuronale au j -ième stimulus.

Remarque 6. *Les efficacités neuronales α du "Balloon Model" n'ont pas le même sens biologique que les amplitudes du modèle linéaire. Cependant, comme précédemment, elles quantifient la façon dont le cerveau réagit à la stimulation. De plus, elles joueront le même rôle dans le modèle statistique. Par conséquent nous avons décidé d'appeler α les efficacités neuronales du modèle physiologique.*

De nombreuses variantes ont été proposées autour de ce modèle. En effet, tout d'abord, ce modèle suppose que la concentration en dioxygène dans les tissus est nulle. Par conséquent, le taux d'extraction d'oxygène des capillaires ne dépend que de l'oxygène apporté et donc du débit sanguin local. Bien que ce lien entre taux d'extraction d'oxygène et débit sanguin ait été fortement étayé par des études récentes [63], certaines expériences montrent que cette hypothèse est peut-être trop simplificatrice [120]. Des modèles d'extraction d'oxygène plus compliqués peuvent être envisagés. Par exemple Zheng *et al.* [136] ont étendu le modèle de Friston en fournissant une modélisation dynamique du transport d'oxygène dans les tissus, intégrant ainsi deux variables d'état supplémentaires : le rapport entre la concentration en oxygène dans les tissus et celle dans le plasma au bout de l'artère d'une part et la concentration moyenne en oxygène dans le capillaire sanguin d'autre part. En outre, d'autres propriétés de visco-élasticité du ballon ont été envisagées [24]. Enfin, différents modèles incluant le métabolisme du glucose ont été proposés [7]. Dans tous les cas, le système différentiel devient plus complexe (davantage de variables d'état, davantage d'équations).

Cependant, quelque soit le modèle physiologique considéré, le signal est modélisé comme solution d'un système d'équations différentielles sans solution analytique.

Dans le cadre d'un modèle statistique, les données $y(t_i)$ sont la somme du terme de régression $H \circ \mathbf{x}(t_i, \mathbf{a}, \boldsymbol{\alpha})$ et d'un terme d'erreur ε_i :

$$\begin{cases} y(t_i) & = H \circ \mathbf{x}(t_i, \mathbf{a}, \boldsymbol{\alpha}) + \varepsilon_i \\ \frac{\partial \mathbf{x}(t_i, \mathbf{a}, \boldsymbol{\alpha})}{\partial t} & = F(\mathbf{x}(t_i, \mathbf{a}, \boldsymbol{\alpha}), t_i, \mathbf{a}, \boldsymbol{\alpha}) \end{cases}$$

où $(\mathbf{a}, \boldsymbol{\alpha}) \in \mathbb{R}^k$. $H : \mathbb{R}^d \rightarrow \mathbb{R}$ et $F : \mathbb{R}^d \times [0, +\infty[\times \mathbb{R}^k \rightarrow \mathbb{R}^d$ sont des fonctions connues. $\mathbf{x}(t_i, \mathbf{a}, \boldsymbol{\alpha}) : [0, +\infty[\times \mathbb{R}^k \rightarrow \mathbb{R}^d$ représente les variables d'état, solution d'un système différentiel sans solution analytique.

Dans un cadre bayésien, Friston *et al.* [50] ont proposé d'estimer les distributions des paramètres hémodynamiques et efficacités neuronales en utilisant un développement en série de Volterra du système. Les moyennes et variances de ces distributions ont été utilisées comme a priori empiriques dans un schéma bayésien utilisant un algorithme EM, fondé sur une approximation bilinéaire du système différentiel [49]. Des méthodes d'estimation par minimisation du critère des moindres carrés ont aussi été envisagées (voir la thèse de Thomas Deneux [37]).

2.2.2 Les apports de cette thèse

La méthode de minimisation des moindres carrés proposées par Thomas Deneux dans sa thèse [37] a été appliquée pour des efficacités neuronales $\boldsymbol{\alpha}$ constantes. La méthode développée par Friston

[49], reposant sur une linéarisation du système, pourraient rencontrer des problèmes de temps de convergence dans le cas où de nombreuses efficacités neuronales devraient être estimées. Nous avons donc cherché à proposer une méthode efficace d'estimation dans des modèles à efficacités neuronales variables, ce, sans linéariser le système différentiel.

En ce fondant sur le modèle et les méthodes proposées dans la partie 2.1, nous considérons le modèle suivant. Supposons que le paradigme compte K types de stimuli. Soit $\alpha_{k,j}$ l'efficacité neuronale du j -ième stimulus de type k . Nous modélisons les distributions des efficacités neuronales par des gaussiennes, $\forall j = 1 \dots J_k, k = 1 \dots K$:

$$\alpha_{k,j} \sim \mathcal{N}(\mu_k, \omega_k)$$

Le modèle statistique est alors le suivant :

$$\left\{ \begin{array}{l} y(t_i) = H \circ x(t_i, \mathbf{a}, \boldsymbol{\alpha}) + \varepsilon_i \\ \frac{\partial x(t_i, \mathbf{a}, \boldsymbol{\alpha})}{\partial t} = F(x(t_i, \mathbf{a}, \boldsymbol{\alpha}), t_i, \mathbf{a}, \boldsymbol{\alpha}) \\ \varepsilon_i \sim \mathcal{N}(0, \sigma^2) \\ \alpha_{k,j} \sim \mathcal{N}(\mu_k, \omega_k) \end{array} \right. \quad \begin{array}{l} \forall i = 1 \dots n \\ \\ \\ \forall j = 1 \dots J_k, k = 1 \dots K \end{array}$$

Sur le même schéma que dans le cas d'un modèle de convolution, nous proposons alors une estimation des paramètres physiologiques et des paramètres de la distribution des efficacités neuronales en deux temps :

1. d'une part, une estimation des paramètres physiologiques par minimisation d'un critère des moindres carrés sous contraintes,
2. d'autre part, une estimation des paramètres $(\mu_k, \omega_k)_{k=1 \dots K}$ par maximum de vraisemblance.

Méthodes d'estimation dans des modèles à données non-observées définis par équation différentielles ordinaires (EDO)

Ce deuxième point nous amène à considérer une famille plus générale de modèles statistiques dans lesquels les observations sont la somme d'un bruit gaussien et d'un terme de régression, solution d'une équation différentielle sans solution analytique et dépendant d'un paramètre aléatoire non observé :

$$\left\{ \begin{array}{l} y(t_i) = H \circ f(t_i, \phi) + \varepsilon_i \\ \varepsilon_i \sim \mathcal{N}(0, \sigma^2) \\ \phi \sim \pi(\cdot; \beta) \end{array} \right. \quad \forall i = 1 \dots n \quad (\mathcal{M})$$

où f est solution de l'équation différentielle suivante :

$$\left\{ \begin{array}{l} \frac{\partial f(t, \phi)}{\partial t} = F(f(t, \phi), t, \phi) \\ f(t_0, \phi) = f_0(\phi) \end{array} \right. \quad (2.2.5)$$

Remarque 7. *Ces modèles ne sont pas réservés aux données IRMf, mais trouvent aussi leur utilité en pharmacocinétique (PK) ou en pharmacodynamie (PD) d'un traitement médicamenteux. La pharmacocinétique a pour but d'étudier le devenir d'un principe actif contenu dans un médicament dans l'organisme, alors que la pharmacodynamie décrit l'action du médicament sur l'organisme : c'est l'étude détaillée de l'interaction récepteur/principe actif. En général, le corps humain est assimilé à un ensemble de compartiments entre lesquels le médicament peut se mouvoir et éventuellement se transformer. Cette modélisation aboutit à ce que la quantité de médicament dans le sang soit solution d'un système différentiel plus ou moins complexe, selon le nombre de compartiments considérés par exemple. Lors d'un essai clinique, plusieurs sujets subissent plusieurs prélèvements au cours du temps. Les données sont alors généralement analysées en utilisant un modèle non linéaire à effets mixtes*

qui permet de distinguer la variabilité intra-sujet de la variabilité inter-sujet [99]. Les modèles non linéaires à effets mixtes s'écrivent de la façon suivante :

$$\begin{cases} y_{ij} &= Z(t_{ij}, \phi_i) + \varepsilon_{ij} \\ \varepsilon_{ij} &\sim \mathcal{N}(0, \sigma^2) \\ \phi_i &\sim \mathcal{N}(\mu, \Omega) \end{cases}$$

où y_{ij} est l'observation du sujet i , $i = 1, \dots, I$ au temps t_{ij} , $j = 1, \dots, J_i$ et ϕ_i est le vecteur des paramètres individuels du sujet i , Z étant solution d'une équation différentielle ordinaire.

Le système différentiel (2.2.5) n'ayant pas de solution analytique en général, l'estimation des paramètres (β, σ^2) doit être faite sur un modèle \mathcal{M}_h dit "approché" dont la fonction de régression $g_h = H \circ f_h$ est une approximation numérique de la solution de l'EDO (2.2.5), de pas h et d'ordre p : $\sup |f(t, \phi) - f_h(t, \phi)| = O(h^p)$.

Sur le modèle approché \mathcal{M}_h , l'estimateur de maximum de vraisemblance ne peut être obtenu explicitement. Nous avons donc recours à une version stochastique de l'algorithme EM [35, 36] en considérant que les données \mathbf{y} sont issues de l'observation partielle des données complètes (\mathbf{y}, ϕ) . Nous estimons alors les paramètres (β, σ^2) en adaptant l'algorithme *Stochastic Approximation Expectation Maximisation* (SAEM) couplé avec une procédure de Monte Carlo par Chaîne de Markov (MCMC) proposé par Kuhn et Lavielle [72]. De plus, nous quantifions l'erreur engendrée par l'approximation numérique sur la maximisation de la vraisemblance en fonction du pas et de l'ordre de la méthode. Le même travail est réalisé dans le cadre d'une estimation bayésienne par un algorithme de Gibbs. Le chapitre 5 de cette thèse est consacrée à ce travail sous la forme d'un article co-écrit avec Adeline Samson⁴ et accepté par *Journal of Statistical Planning and Inference* (sous réserve de modifications mineures)

Application de la méthode d'estimation par maximum de vraisemblance dans le cadre des données IRMF

Après une présentation détaillée du "*Ballon model*" et une étude précise du rôle des paramètres physiologiques ainsi que de ses propriétés de non-linéarité et d'identifiabilité (chapitre 6, sections 6.2 et 6.3), nous illustrons la précision de notre méthode d'estimation par maximum de vraisemblance dans ce cadre sur données simulées.

La même méthode est ensuite appliquée sur un jeu de données réelles fourni par T. Deneux⁵. Sur ce paradigme "blocs", composé de répétitions aléatoires de blocs de flashes visuels (blocs du type 8×1 seconde - 4×1 seconde - 2×1 seconde - 1×1 seconde - 5×200 ms - 1×200 ms), nous mettons en évidence par un test du maximum de vraisemblance l'existence d'une variabilité inter-type de stimuli. Ainsi, les blocs du type " 8×1 s" n'ont pas la même efficacité neuronale que les blocs du type " 4×1 s", etc... D'autre part, il est connu dans la littérature qu'il existe au niveau hémodynamique, un phénomène de "*suppression à la répétition*" : plus précisément, si un stimulus est précédé par d'autres stimuli, la réponse qu'il engendre est plus faible que s'il était isolé dans le temps (ce phénomène non-linéaire est une des motivations des modèles physiologiques). Nous mettons en évidence qu'il existe le même genre de phénomène au niveau neuronal. De plus, nous parvenons à quantifier ce phénomène neuronal de suppression à la répétition.

⁴INSERM U738, Département de biostatistiques, Hôpital Bichat, 46 rue Huchard, 75018 PARIS

⁵Weizmann Institute of Science - Department of Neurobiology, Israel

Chapitre 3

Estimation dans des modèles définis par des équations différentielles stochastiques à paramètres aléatoires

Sommaire

3.1 État de l'art	34
3.2 Les apports de cette thèse	35

Nous avons considéré précédemment les modèles statistiques dont la fonction de régression est solution d'une équation différentielle ordinaire (EDO). Plus précisément, $\forall i = 1 \dots n$,

$$\begin{cases} y(t_i) &= H \circ f(t_i, \phi) + \varepsilon_i \\ \varepsilon_i &\sim \text{i.i.d } \mathcal{N}(0, \sigma^2) \\ \phi &\sim \pi(\cdot; \beta) \end{cases}$$

avec f telle que :

$$\begin{cases} \frac{\partial f(t, \phi)}{\partial t} &= F(f(t, \phi), t, \phi) \\ f(t_0, \phi) &= f_0(\phi) \end{cases} .$$

Dans ce cas, seul le bruit d'observation est pris en compte : $\varepsilon_i \sim \text{i.i.d } \mathcal{N}(0, \sigma^2)$. Il est naturel d'étendre ces modèles au cas où l'équation différentielle ordinaire est remplacée par une équation différentielle stochastique (EDS) :

$$\begin{cases} dZ(t, \phi) &= F(Z(t, \phi), t, \phi) dt + \gamma(Z(t, \phi), t, \phi) dB_t \\ Z(t_0, \phi) &= Z_0(\phi) \end{cases} \tag{3.0.1}$$

où B_t est un mouvement brownien.

L'introduction d'une variabilité supplémentaire dans le modèle est une façon de prendre en compte des erreurs corrélées dans le temps ou de modéliser par exemple des variations des paramètres au cours du temps. Cette généralisation a été suggérée en IRMf par Riera [105] et en pharmacocinétique dans le cadre des modèles non linéaires à effets mixtes par Overgaard *et al.* [95] (voir aussi [71]).

Dans cette thèse, nous nous intéressons donc à l'estimation paramétrique de modèles de diffusions bruitées, dépendant de paramètres aléatoires.

3.1 État de l'art

L'estimation des paramètres des diffusions est largement traitée dans la littérature. Nous en faisons ici un rapide état de l'art.

Les diffusions observées à temps discrets

Les diffusions observées de façon continue ont été largement étudiées [75, 102]. Cependant, pour des raisons pratiques évidentes, les données biologiques (IRMf, concentration de médicament, etc) sont collectées à temps discrets. Dans ce cadre, l'inférence statistique des processus de diffusion observés à temps discrets est une question difficile.

Lorsque la probabilité de transition du processus de diffusion est explicite, Dacunha-Castelle *et al.* [34] ont construit un estimateur consistant. Des algorithmes bayésiens classiques tels que l'algorithme de Gibbs peuvent aussi être appliqués directement dans ce cas particulier. Cependant, les probabilités de transition n'ont en général pas d'expression explicite et les méthodes d'estimation doivent surmonter cette difficulté. Nous donnons ici un aperçu des méthodes existantes (voir [102, 114] pour une revue complète).

Les méthodes analytiques incluent celles de Bibby *et al.* [15], Sorensen *et al.* [115] – utilisant des fonctions d'estimation –, Poulsen *et al.* [101] – utilisant une solution numérique de l'équation de Kolmogorov – ou Aït-Sahalia [3] – basée sur une approximation analytique non gaussienne de la log-vraisemblance.

D'autres méthodes approchent la densité de transition par simulation. Dans ce cas, les instants non-observés sont considérés comme des données manquantes et les auteurs introduisent des instants intermédiaires entre chaque paire d'instant d'observation. En ces instants intermédiaires, le processus peut être approché en utilisant une approximation gaussienne de Euler-Maruyama, de façon à estimer la fonction de vraisemblance par intégration numérique ou, dans un cadre bayésien, pour évaluer la distribution a posteriori encore par intégration numérique, comme proposé par Pedersen [97] et Elerian *et al.* [40].

Dans ce contexte, à la fois pour l'estimation par maximum de vraisemblance et l'estimation Bayésienne, des méthodes de Monte Carlo par Chaîne de Markov standard sont mises en place pour générer le processus sous la distribution conditionnelle. Cependant, la vitesse de convergence décroît à mesure que le volume de données manquantes augmente. Différentes solutions sont évoquées pour surmonter cette difficulté.

Eraker [41] suggère l'échantillonnage d'un élément à la fois alors que Elerian *et al.* [40] proposent de simuler bloc par bloc par un algorithme d'échantillonneur d'importance. Roberts *et al.* [108] prennent une approche légèrement différente puisqu'ils échantillonnent une transformation du processus de diffusion.

Afin de ne pas utiliser d'approximation de Euler-Maruyama, Beskos *et al.* [14, 13] ont récemment développé une méthode de simulation exacte du processus de diffusion, applicable même en l'absence d'expression analytique de la probabilité de transition. L'algorithme peut être inclus dans une procédure de Monte Carlo de façon à approcher la fonction de vraisemblance dans un cadre classique ou dans un algorithme de Gibbs dans le cadre de l'inférence bayésienne. Cependant, cette méthode a des contraintes. En effet, elle peut rencontrer des difficultés algorithmiques puisqu'elle requiert la mise en place d'un algorithme d'acceptation-rejet qui peut parfois demander un très grand temps de calcul. Enfin, dans le cadre que nous considérons, cette méthode de simulation exacte ne nous permet pas d'appliquer l'algorithme de maximisation de la vraisemblance SAEM, le modèle considéré n'étant pas du type exponentiel dans ce cas. Par conséquent, dans cette thèse, nous considérons une approche par approximation d'Euler-Maruyama.

Les diffusions observées à temps discrets, avec bruit d'observation

Les travaux précédemment cités ne prennent pas en compte le bruit d'observation. Cette hypothèse est irréaliste dans de nombreux cas, notamment en biologie où les mesures sont soumises au manque de précision de l'appareil de mesure. Certains articles de la littérature [53, 67, 112] considèrent des diffusions observées à temps discrets avec bruit d'observation :

$$\begin{cases} y(t_i) &= Z(t_i) + \varepsilon_i \\ \varepsilon_i &\sim \mathcal{N}(0, \sigma^2) \\ dZ(t, \phi) &= F(Z(t, \phi), t, \phi) dt + \gamma(Z_t, t\phi) dB_t, Z(t_0, \phi) = Z_0(\phi) \end{cases} \quad (3.1.2)$$

Dans de tels modèles, deux types de bruits fondamentalement différents doivent être distingués :

- le bruit dynamique représenté par γ représente les fluctuations aléatoires autour d'un modèle dynamique théorique,
- le bruit de mesure paramétré par σ^2 représente la variance résiduelle associées au dosage, à l'analyse, à l'image etc.

Le problème d'estimation de paramètres dans ce cadre est peu évoqué dans la littérature et résolu de façon incomplète. Dans le cas particulier des EDS linéaires, le filtre de Kalman [112] ou encore l'algorithme EM [113] peuvent être utilisés. Lorsque le processus observé est une martingale gaussienne, Jensen *et al.* [67] et Gloter *et al.* [53] exhibent des estimateurs et étudient leurs propriétés théoriques. Cependant, le cas général n'a pas été étudié.

Les diffusions observées à temps discrets, avec bruit d'observation et paramètres aléatoires

Dans cette thèse, nous supposons de plus que l'équation différentielle stochastique dépend d'un paramètre aléatoire ϕ , réalisation d'une distribution π dépendant d'un paramètre β .

$$\begin{cases} y_i &= Z(t_i) + \varepsilon_i, & 0 \leq j \leq J \\ \varepsilon_i &\sim_{i.i.d} \mathcal{N}(0, \sigma^2), \\ dZ(t) &= F(Z, t, \phi) dt + \gamma dB(t), Z(t_0, \phi) = Z_0(\phi) \\ \phi &\sim \pi(\cdot, \beta) \end{cases} \quad (\mathcal{M})$$

Le but est d'estimer les paramètres $\theta = (\beta, \gamma^2, \sigma^2)$.

Remarque 8. *Par souci de simplicité, nous supposons $\gamma(Z(t, \phi), t, \phi) = \gamma$. De plus, la fonction H est ici prise égale à la fonction identité. Une extension de ce travail pourrait être de supposer $H \neq id$.*

Dans le cas particulier des modèles non linéaires à effets mixtes (introduits en remarque 7), Ditlevsen *et al.* [38] ont proposé une méthode d'estimation des paramètres $\theta = (\mu, \Omega, \gamma^2, \sigma^2)$ mais leur étude se restreint à un exemple où les probabilités de transition ont des expressions explicites. Overgaard et Tornøe [95, 122] ont proposé une méthode combinant une généralisation du filtre de Kalman avec un algorithme d'estimation par maximum de vraisemblance approchée, fondé sur une linéarisation du modèle. Cependant, les propriétés de convergence des algorithmes reposant sur cette linéarisation n'ont pas été prouvées.

3.2 Les apports de cette thèse

Estimation

Dans cette thèse, nous proposons des méthodes d'estimation des paramètres $\theta = (\beta, \gamma^2, \sigma^2)$ du modèle \mathcal{M} à la fois dans un cadre bayésien et dans un cadre de maximum de vraisemblance.

Nous considérons les paramètres aléatoires ϕ et la diffusion Z aux instants d'observations comme des données non-observées. Nous cherchons alors à adapter la version stochastique de l'algorithme

EM proposée par Kuhn et Lavielle [73] dans un cadre d'estimation par maximum de vraisemblance. Un algorithme de Gibbs est aussi proposé dans le cadre bayésien. Nous obtenons des résultats de convergence des algorithmes sous des hypothèses assez générales.

Notre méthode repose sur une approximation de la diffusion par le schéma d'Euler-Maruyama, il est donc intéressant de quantifier l'erreur induite par cette approximation sur la qualité des estimateurs. Notre résultat repose sur les résultats de Bally *et al.* [9] bornant l'erreur sur les probabilités de transition.

La méthode d'estimation par maximum de vraisemblance est appliquée sur un exemple issu de la pharmacocinétique, à la fois sur données simulées et données réelles. L'illustration sur données simulées montre que malgré la difficulté du problème, les paramètres sont estimés avec précision. L'analyse des données réelles de pharmacocinétique de la Théophylline (médicament anti-asthmatique) illustre la pertinence de l'approche stochastique par rapport à l'approche déterministe, les courbes individuelles étant mieux prédites par le modèle différentiel stochastique.

Ces travaux font l'objet d'un article co-écrit avec Adeline Samson et soumis à *ESAIM : Probability and Statistics* en juillet 2006.

Identifiabilité

Finalement, nous terminons cette thèse en abordant la difficile question de l'identifiabilité des modèles définis par équations différentielles stochastiques à paramètres aléatoires. La propriété d'identifiabilité est nécessaire pour obtenir par exemple la consistance des estimateurs. Cependant, cette question est peu traitée dans la littérature. Dans le cadre des modèles linéaires et non linéaires à effets mixtes standards, Nuñez et Concordet [33] ont énoncé des conditions suffisantes assurant l'identifiabilité des modèles. Cependant, dans le cas des modèles à effets mixtes définis par EDS (diffusions mixtes), aucun résultat n'existe dans la littérature.

Nous présentons au chapitre 8, différents modèles pour lesquels nous sommes capables de prouver l'identifiabilité des paramètres. Nous mettons de plus en exergue les difficultés rencontrées.

Ce travail est un travail en cours, sous la direction de Valentine Genon-Catalot¹.

¹Laboratoire MAP5, Université René Descartes- Paris 5

Deuxième partie

Inversion de données IRMf et sélection de modèles sous hypothèse de linéarité

Dans cette partie, nous proposons une variante du modèle linéaire standard présenté dans l'introduction (chapitre 2), en introduisant une variabilité inter-occurrences des amplitudes de la réponse hémodynamique. Nous développons une méthode d'estimation des paramètres de nos modèles statistiques ainsi qu'une méthode de sélection de modèles. Nous testons la validité de l'hypothèse de variabilité des amplitudes sur un grand nombre d'individus et de régions d'intérêts .

Ce chapitre est l'occasion de rappeler les principes et résultats de convergence de l'algorithme *Expectation-Maximisation* de Dempster *et al.* [36], et de ses versions stochastiques proposées par Delyon *et al.* [35] et Kuhn et Lavielle [72]. Nous réutiliserons ces résultats tout au long de cette thèse.

Chapitre 4

Inversion de données IRMf dans un cadre linéaire. Estimation et sélection de modèle.

Sommaire

4.1	Introduction	41
4.2	Définition des modèles statistiques	42
4.3	Estimation sous $\mathcal{M}_{\mathcal{P},0}$	45
4.4	Estimation sous $\mathcal{M}_{\mathcal{P},1}$	46
4.5	Estimation sous $\mathcal{M}_{\mathcal{P},1}$ avec contrainte de positivité	49
4.6	Calcul de la vraisemblance des modèles	54
4.7	Sélection de modèles	56
4.8	Application sur données simulées et réelles	58
4.9	Discussion et conclusion	67

4.1 Introduction

Une estimation précise de la fonction HRF est une première étape vers une meilleure compréhension de l'activité cérébrale et une meilleure quantification de l'activité neuronale. Récemment, des progrès ont été faits concernant l'estimation non-paramétrique de la HRF. Cependant, les modèles et les méthodes d'estimation proposés jusque là reposent sur deux hypothèses cruciales conduisant à deux problématiques que nous détaillons ci-dessous.

Problématique $\mathcal{P}b_1$: Pendant longtemps, les travaux de la littérature ont supposé que chaque stimulus d'un type donné (en d'autres termes chaque occurrence d'une condition expérimentale donnée) engendrait une réponse BOLD constante en forme et en amplitude : par exemple deux clics droits engendrent une réponse de même forme et de même amplitude. Des modèles ont certes été introduits pour modéliser une importante variabilité entre les types de stimuli mais la stationnarité au cours des occurrences est la plus souvent supposée. Récemment, il a été suggéré que cette hypothèse pourrait être remise en cause [6, 39, 81]. Dans cette thèse, nous proposons un modèle statistique simple permettant non seulement d'estimer la forme et l'amplitude de la réponse HRF mais aussi de tester l'hypothèse de variabilité inter-occurrences des amplitudes. En outre, notre approche peut être appliquée à tout paradigme expérimental événementiel. Nous

présentons des résultats sur un grand nombre de régions d'intérêt (32 régions d'intérêts répartis sur 8 sujets).

Problématique $\mathcal{P}b_2$: D'autre part, dans tous les travaux publiés dans la littérature, les différents types de stimuli présents dans le paradigme expérimental ont toujours été inclus dans le modèle. Ici, nous cherchons à sélectionner les types de stimuli ayant une réelle influence sur le signal. Par exemple, si les observations sont issues du cortex moteur gauche, il est intéressant de se demander si les clics gauches engendrent une réponse non négligeable. Nous proposons une méthode graphique fondée sur la vraisemblance des modèles permettant de sélectionner les types d'évènements influant de façon significative sur le signal.

Le chapitre est organisé comme suit. Dans une première partie (paragraphe 4.2), nous décrivons en détail les modèles statistiques utilisés en insistant sur les hypothèses sous-jacentes issues de la biologie. Dans les paragraphes 4.3, 4.4 et 4.5, nous présentons des méthodes d'estimation des paramètres des différents modèles notamment par l'algorithme *Expectation-Maximisation* développé par Dempster *et al.* [36] et une de ses versions stochastiques proposée par Kuhn et Lavielle [73]. Le paragraphe 4.6 détaille le calcul des vraisemblances de ces modèles. Nous proposons une méthode de sélection de modèles au paragraphe 4.7. Nous illustrons les performances des algorithmes proposés sur données simulées (paragraphe 4.8.1) . Enfin nous appliquons ces procédures sur 2 jeux de données réelles (4.8.2 et 4.8.3) :

- sur un premier jeu (composé de 8 sujets et 4 régions d'intérêt par sujet), nous cherchons à valider l'hypothèse de variabilité des amplitudes au cours des occurrences (problématique $\mathcal{P}b_1$). Cette validation à grande échelle a fait l'objet d'un article co-écrit avec Marc Lavielle¹ et Jean-Baptiste Poline^{2,3}, accepté en août 2005 par **Neuroimage** et paru en juillet 2006.
- Sur un deuxième jeu de données, nous appliquons la procédure de sélection des types de stimuli ayant un effet significatif sur le signal (problématique $\mathcal{P}b_2$). Cette étude a fait l'objet d'un article publié dans les **actes de la conférence ISBI 2005** et co-écrit avec Marc Lavielle¹, Philippe Ciuciu^{2,3} et Jean-Baptiste Poline^{2,3}.

Dans une dernière partie, nous discutons des conséquences des résultats obtenus.

4.2 Définition des modèles statistiques

4.2.1 Modélisation des séries temporelles observées avec un unique type de stimuli

Soit $\mathbf{y} = (y_1, \dots, y_n)' \in \mathbb{R}^n$ la série temporelle des observations IMRF extraites d'une région d'intérêt donnée, pour un sujet, pendant une session (\mathbf{v}' est la transposée de \mathbf{v}). $y_i = y(t_i)$ est la mesure à l'instant t_i , avec $t_i = i\text{TR}$, TR étant le temps inter-scans. n est le nombre de scans pendant la session. La fonction HRF est notée $h(t)$. Sous l'hypothèse d'un modèle de convolution, nous avons :

$$y_i = \sum_{m=1}^M x(s_m)h(t_i - s_m) + \sum_{q=1}^Q l_q P_q(t_i) + \varepsilon_i, \quad i = 1 \dots n \quad (4.2.1)$$

où

- $(x(s_m))_{m=1 \dots M}$ est la série des entrée avec $s_m = m\Delta_e$. Δ_e définit une grille d'échantillonnage temporel plus fine que celle définie par TR. En effet, un évènement peut avoir lieu à n'importe quel moment, non nécessairement aux instants de mesure. Typiquement TR est compris entre 2

¹Laboratoire de mathématiques, Université de Paris-Sud, 91405 Orsay, France

²SHFJ/CEA 4 Place du Général Leclerc, 91406 Orsay, France , IFR 49

³Institut d'Imagerie Neurofonctionnelle, Paris, France

et 4 secondes alors que les événements peuvent avoir lieu toutes les 100ms environ. Dans les jeux de données réelles étudiées au paragraphe 4.8, nous prendrons $\Delta_e = TR/16$.

- La somme $\sum_{q=1}^Q l_q P_q(t_i)$ représente la dérive basse fréquence. Nous modélisons les basses fréquences par une famille orthonormale de fonctions (P_1, \dots, P_Q) , de la forme $P_q(t) = \sqrt{\frac{2}{n}} \cos\left(\frac{\pi(2t+1)(q-1)}{2n}\right)$. $Q = \lceil 2nTRf_{min} \rceil + 1$ où f_{min} est la plus basse fréquence attribuable au terme de dérive (f_{min} autour de $\frac{1}{120}$ Hertz). Les quantités $(l_q)_{q=1\dots Q}$ représentent les coefficients de pondération inconnus des fonctions de la base et doivent être estimés.
- $\boldsymbol{\varepsilon} = (\varepsilon_1 \dots \varepsilon_n)$ est un vecteur gaussien de variance $\sigma^2 \mathbf{I}_n$ à estimer.

De façon standard, nous supposons que $h(u) = 0$ pour $u \leq 0$. De plus, nous tronquons h , en posant $h(u) = 0$ for $u > L\Delta_e$, avec $L\Delta_e$ typiquement autour de 20 à 25 secondes.

Soient $\tau_1, \tau_2, \dots, \tau_J$ les instants d'occurrences des stimuli. Nous avons

$$\begin{cases} x(s) &= 0 & \text{si } \tau \notin \{\tau_1, \tau_2, \dots, \tau_J\}, \\ x(\tau_j) &= \alpha_j & \text{pour } 1 \leq j \leq J. \end{cases}$$

Alors, le modèle devient :

$$y_i = \sum_{j=1}^J \alpha_j h(t_i - \tau_j) + \sum_{q=1}^Q l_q P_q(t_i) + \varepsilon_i, \quad i = 1 \dots n \quad (4.2.2)$$

Définissons les vecteurs suivants :

$$\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_J)' \in \mathbb{R}^J, \quad \mathbf{h} = (h(0), h(\Delta_e), h(2\Delta_e), \dots, h(L\Delta_e))' \in \mathbb{R}^{L+1}$$

et

$$\mathbf{l} = (l_1, \dots, l_Q)' \in \mathbb{R}^Q.$$

De plus, soient A , H et P les matrices :

$$H = (H_{ij})_{1 \leq i \leq n, 1 \leq j \leq J}, \quad A = (A_{im})_{1 \leq i \leq n, 1 \leq m \leq M} \quad \text{et} \quad P = (P_{iq})_{1 \leq i \leq n, 1 \leq q \leq Q}$$

telles que

$$H_{ij} = h(t_i - \tau_j), \quad A_{im} = x(t_i - s_m) \quad \text{et} \quad P_{iq} = P_q(t_i)$$

Puisque $\forall i \in \{1, \dots, n\}, \forall j \in \{1, \dots, J\}, t_i - \tau_j \in \{k\Delta_e, k \in \mathbb{N}\}$, H ne dépend que des composantes du vecteur \mathbf{h} . De même, A dépend du vecteur $\boldsymbol{\alpha}$. P , H et A ont autant de lignes qu'il y a d'observations (n), H a autant de colonnes qu'il y a de stimuli (J), A a autant de colonnes qu'il y a d'instantanés sur la plus fine grille d'échantillonnage temporel (nTR/Δ_e), P a autant de colonnes qu'il y a de fonctions dans la base choisie (Q). Alors nous obtenons la double écriture matricielle suivante :

$$\begin{cases} \mathbf{y} &= H\boldsymbol{\alpha} + P\mathbf{l} + \boldsymbol{\varepsilon} \\ &= A\mathbf{h} + P\mathbf{l} + \boldsymbol{\varepsilon} \end{cases}$$

Cette double écriture matricielle préfigure des problèmes à venir puisqu'il s'agit de retrouver conjointement la réponse HRF \mathbf{h} , et les amplitudes de cette réponse $\boldsymbol{\alpha}$.

Remarque 9. Dans ce modèle, nous faisons l'hypothèse que la forme de la réponse hémodynamique est constante au cours des occurrences car contrainte par l'environnement neuro-vasculaire (\mathbf{h} ne dépend pas de j). Seules les amplitudes de cette réponse ($\boldsymbol{\alpha}$) varient.

4.2.2 Modélisation dans le cas de plusieurs types de stimuli

Nous considérons maintenant K types de stimuli différents. Par exemple, les paradigmes choisis dans la partie 4.8 contiennent des événements tels que des tâches motrices (clics droit et gauche), des stimuli visuel ou auditifs, etc. Notons $\tau_{k,1}, \dots, \tau_{k,J_k}$ les instants d'occurrences des stimuli de type k ($1 \leq k \leq K$) et $\alpha_{k,j} \in \mathbb{R}$ l'amplitude du j -ième stimulus de type k . Le nombre total de stimuli est alors $\sum_{j=1}^K J_k = J$ et l'expression (4.2.2) devient

$$y_i = \sum_{k=1}^K \sum_{j=1}^{J_k} \alpha_{k,j} h(t_i - \tau_{k,j}) + \sum_{q=1}^Q l_q P_q(t_i) + \varepsilon_i, \quad i = 1 \dots n \quad (4.2.3)$$

Distribution des amplitudes $\alpha_{k,j}$

Nous cherchons à modéliser des amplitudes ayant une variabilité inter-occurrences (problématique $\mathcal{P}b_1$). Nous devons choisir une distribution de probabilité sur les amplitudes $\alpha_{k,j}$. Le choix le plus simple en l'absence de toute autre information est de considérer une distribution gaussienne. Autrement dit, l'amplitude du j -ième stimulus de type k ($1 \leq j \leq J_k, 1 \leq k \leq K$) vérifie :

$$\alpha_{k,j} \sim \mathcal{N}(\mu_k, \omega_k)$$

les amplitudes étant supposées indépendantes. Les paramètres $(\mu_k)_{k=1 \dots K}$ et $(\omega_k)_{k=1 \dots K}$ sont à estimer.

Remarque 10. *Il peut parfois être judicieux d'imposer une contrainte de positivité sur les amplitudes. La distribution la plus adaptée est alors une distribution log-normale :*

$$\log \alpha_{k,j} \sim \mathcal{N}(\mu_k, \omega_k)$$

Remarque 11. *Notons que, bien que le modèle soit une simplification des phénomènes biologiques sous-jacents (puisqu'il n'inclut aucune modélisation physiologique telle que celle fournie par le "Balloon model" par exemple), il peut cependant fournir des informations utiles sur le comportement au cours du temps de la réponse cérébrale à un stimulus donné.*

En outre, ce modèle est plus flexible que ceux proposés jusque là dans la mesure où nous ne fixons pas de forme de HRF a priori ni ne supposons de stationnarité des amplitudes dans le temps.

Définition des modèles

Soit \mathcal{P} une sous-famille de l'ensemble des types des stimuli du paradigme : $\mathcal{P} \subset \{1, 2, \dots, K\}$. Nous définissons alors les modèles $\mathcal{M}_{\mathcal{P},0}$ et $\mathcal{M}_{\mathcal{P},1}$ sur les distributions des amplitudes par :

$$\mathcal{M}_{\mathcal{P},0} : \begin{cases} \mu_k \neq 0 \text{ et } \omega_k = 0 & \text{si } k \in \mathcal{P} \\ \mu_k = 0 \text{ et } \omega_k = 0 & \text{sinon} \end{cases},$$

et

$$\mathcal{M}_{\mathcal{P},1} : \begin{cases} \mu_k \neq 0 \text{ et } \omega_k \neq 0 & \text{si } k \in \mathcal{P} \\ \mu_k = 0 \text{ et } \omega_k = 0 & \text{sinon} \end{cases}.$$

$\mathcal{M}_{\mathcal{P},0}$ est le modèle couramment utilisé (amplitudes inconnues et constantes par type de stimuli) alors que $\mathcal{M}_{\mathcal{P},1}$ est le modèle dans lequel les amplitudes sont variables d'une occurrence à l'autre pour un même type d'événements (la variance de la distribution a priori est non nulle). Dans $\mathcal{M}_{\mathcal{P},0}$, les événements sont modélisés comme des effets fixes d'amplitude μ_k alors que dans $\mathcal{M}_{\mathcal{P},1}$ les événements

sont aléatoires de moyenne μ_k et de variance ω_k . Il est important de noter que, outre la réponse HRF et les coefficients de la dérive basse fréquence, le modèle $\mathcal{M}_{\mathcal{P},0}$ comporte $\text{Card}(\mathcal{P}) + 1$ paramètres à estimer (les $\text{Card}(\mathcal{P})$ moyennes μ_k et la variance de bruit σ^2) alors que le modèle $\mathcal{M}_{\mathcal{P},1}$ en comporte $2\text{Card}(\mathcal{P}) + 1$ (moyenne et variance pour chaque type k inclu dans la famille \mathcal{P} et σ^2).

En outre, dans le modèle $\mathcal{M}_{\mathcal{P},1}$, nous ne considérons pas chaque événement individuellement mais bien leur comportement global au travers de leur moyenne et leur variance. Le nombre de paramètres à estimer est ainsi beaucoup plus petit que le nombre d'occurrences (J).

Remarque 12. *Nous ne considérons pas les modèles “composites” dans lesquels un des types de stimuli générerait des réponses à amplitudes variables et un autre des réponses à amplitudes constantes.*

Nous avons maintenant trois objectifs :

1. estimer la réponse HRF ainsi que les paramètres des modèles définis précédemment ;
2. à famille de stimuli \mathcal{P} fixée, choisir entre les modèles $\mathcal{M}_{\mathcal{P},0}$ ou $\mathcal{M}_{\mathcal{P},1}$ c'est-à-dire décider si les stimuli ont des amplitudes constantes ou non au cours du temps (problématique $\mathcal{P}b_1$) ;
3. sélectionner la famille \mathcal{P} c'est-à-dire les types de stimuli ayant une influence significative sur le signal (problématique $\mathcal{P}b_2$).

4.3 Estimation sous le modèle $\mathcal{M}_{\mathcal{P},0}$

Soit $\mathcal{P} \subset \{1, 2, \dots, K\}$ une famille de stimuli de cardinal $\text{Card}\mathcal{P}$. Notons $\theta_{\mathcal{P},0} = ((\mu_k)_{k \in \mathcal{P}}, \sigma^2)$ les paramètres du modèle $\mathcal{M}_{\mathcal{P},0}$. Nous devons procéder à l'estimation conjointe de $\theta_{\mathcal{P},0}$, \mathbf{l} les poids de la dérive basse fréquence et \mathbf{h} la réponse HRF. La HRF \mathbf{h} doit répondre à certaines contraintes.

4.3.1 Contraintes sur \mathbf{h}

Tout d'abord, le modèle s'écrivant matriciellement $\mathbf{y} = \mathbf{A}\mathbf{h} + \mathbf{P}\mathbf{l} + \boldsymbol{\varepsilon}$, la norme de \mathbf{h} est fixée égale à un de façon à rendre le modèle identifiable :

$$\|\mathbf{h}\|^2 = \sum_i h_i^2 = 1$$

Dans le cas contraire, pour tout $\lambda \in \mathbb{R}^*$, $\mathbf{A}\mathbf{h} = (\lambda\mathbf{A}) \frac{\mathbf{h}}{\lambda}$.

D'autre part, comme proposé par Ciuciu *et al.* [31], nous cherchons une fonction HRF suffisamment lisse en accord avec notre savoir biologique. Ainsi nous imposons la contrainte $\int h''^2 = C_{reg}$ où h'' est la dérivée seconde de h . Dans notre cas, nous imposons au vecteur \mathbf{h} de vérifier :

$$R(\mathbf{h}) = \sum_{l=1}^{L-1} (\mathbf{h}_{l-1} - 2\mathbf{h}_l + \mathbf{h}_{l+1})^2 = C_{reg}.$$

La constante C_{reg} est fixée arbitrairement. Cependant, nous avons fait varier C_{reg} de 10^{-6} à 10^{-5} sur un jeu de données, et nous nous sommes assurés que les fonctions HRF obtenues ne dépendaient pas trop lourdement du choix de cet hyper-paramètre.

Remarque 13. *Un prolongement de ce travail serait de chercher à estimer C_{reg} à partir des données.*

4.3.2 Algorithme d'estimation de $\theta_{\mathcal{P},0}$, \mathbf{l} et \mathbf{h} sous le modèle $\mathcal{M}_{\mathcal{P},0}$

Sous le modèle $\mathcal{M}_{\mathcal{P},0}$, pour tout $k \in \mathcal{P}$, et tout $1 \leq j \leq J_k$ nous avons $\alpha_{k,j} = \mu_k$ (amplitudes constantes au cours du temps pour un même type d'événements). Ainsi :

$$y_i = \sum_{k \in \mathcal{P}} \sum_{j=1}^{J_k} \mu_k h(t_i - \tau_{k,j}) + \sum_{q=1}^Q l_q P_q(t_i) + \varepsilon_i \quad i = 1 \dots n$$

Nous estimons $(\theta_{\mathcal{P},0}, \mathbf{l}, \mathbf{h})$ en minimisant le critère des moindres carrés suivant :

$$U_0(\boldsymbol{\mu}, \mathbf{h}, \mathbf{l}) = \sum_{i=1}^n \left(y_i - \sum_{k \in \mathcal{P}} \sum_{j=1}^{J_k} \mu_k h(t_i - \tau_{k,j}) - \sum_{q=1}^Q l_q P_q(t_i) \right)^2 \quad (4.3.4)$$

sous les contraintes $\|\mathbf{h}\|^2 = 1$ et $R(\mathbf{h}) = C_{reg}$.

Ce problème d'optimisation présente deux difficultés. Tout d'abord, \mathbf{h} , \mathbf{l} et $\boldsymbol{\mu}$ ne peuvent être obtenus conjointement. Nous avons par conséquent recours à un processus itératif dans lequel $U_0(\boldsymbol{\mu}, \mathbf{l}, \mathbf{h})$ est successivement minimisé par rapport à $(\boldsymbol{\mu}, \mathbf{l})$ puis \mathbf{h} . Deuxièmement, la minimisation de $U_0(\boldsymbol{\mu}, \mathbf{l}, \mathbf{h})$ sous la contrainte $R(\mathbf{h}) = C_{reg}$ ne peut-être réalisée de façon exacte. De façon classique, nous minimisons un critère des moindres carrés pénalisé :

$$U_0(\boldsymbol{\mu}, \mathbf{l}, \mathbf{h}) + \lambda R(\mathbf{h}).$$

Le paramètre de régularisation λ est ajusté de façon à ce que la solution du problème de minimisation $\widehat{\mathbf{h}}_{\mathcal{P},0}$ satisfasse la contrainte $R(\widehat{\mathbf{h}}_{\mathcal{P},0}) = C_{reg}$.

Ainsi nous implémentons l'algorithme suivant. A l'itération 0, nous initialisons avec une réponse HRF $\mathbf{h}^{(0)}$ de forme standard. A l'itération ℓ , l'algorithme s'écrit :

1. Calcul de $\boldsymbol{\mu}^{(\ell)}$ et $\mathbf{l}^{(\ell)}$ par minimisation de la fonction $(\boldsymbol{\mu}, \mathbf{l}) \rightarrow U_0(\boldsymbol{\mu}, \mathbf{l}, \mathbf{h}^{(\ell-1)})$,
2. Calcul de $\mathbf{h}^{(\ell)}$ par minimisation de $\mathbf{h} \rightarrow U_0(\boldsymbol{\mu}^{(\ell)}, \mathbf{l}^{(\ell)}, \mathbf{h}) + \lambda_\ell R(\mathbf{h})$ où λ_ℓ est telle que la contrainte de régularité soit satisfaite (i.e. $R(\mathbf{h}^{(\ell)}) \leq C_{reg}$). Normalisation de $\mathbf{h}^{(\ell)}$.

Cet algorithme converge vers un minimum local du critère des moindres carrés $U_0(\boldsymbol{\mu}, \mathbf{l}, \mathbf{h})$ noté $(\widehat{\boldsymbol{\mu}}_0, \widehat{\mathbf{l}}_{\mathcal{P},0}, \widehat{\mathbf{h}}_{\mathcal{P},0})$ satisfaisant les contraintes détaillées précédemment. L'estimateur du maximum de vraisemblance de σ^2 est alors obtenu par

$$\widehat{\sigma}_0^2 = \frac{1}{n} U_0(\widehat{\boldsymbol{\mu}}_0, \widehat{\mathbf{l}}_{\mathcal{P},0}, \widehat{\mathbf{h}}_{\mathcal{P},0}).$$

Nous avons vérifié graphiquement que la solution obtenue ne dépend pas de l'initialisation de l'algorithme en initialisant sur plusieurs réponses $\mathbf{h}^{(0)}$.

4.4 Estimation de $\theta_{\mathcal{P},1}$ \mathbf{h} et \mathbf{l} sous le modèle $\mathcal{M}_{\mathcal{P},1}$

Sous le modèle $\mathcal{M}_{\mathcal{P},1}$, pour tout $k \in \mathcal{P}$, les amplitudes $\alpha_{k,j}$ sont indépendantes et identiquement distribuées de loi gaussienne de moyenne μ_k et de variance ω_k :

$$\alpha_{k,j} \sim_{i.i.d} \mathcal{N}(\mu_k, \omega_k).$$

Nous cherchons à estimer les paramètres $\theta_{\mathcal{P},1} = ((\mu_k, \omega_k)_{k \in \mathcal{P}}, \sigma^2)$, \mathbf{h} et \mathbf{l} . L'écriture matricielle du modèle $\mathcal{M}_{\mathcal{P},1}$ est la suivante :

$$\begin{cases} \mathbf{y} &= \mathbf{H}\boldsymbol{\alpha} + \mathbf{P}\mathbf{l} + \boldsymbol{\varepsilon} \\ \boldsymbol{\alpha} &\sim \mathcal{N}(\boldsymbol{\varphi}, \mathbf{G}) \\ \boldsymbol{\varepsilon} &\sim \mathcal{N}(0, \sigma^2 \mathbf{I}_n) \end{cases}$$

où $\boldsymbol{\alpha}$ est le vecteur des amplitudes classées dans leur ordre chronologique (α_j est l'amplitude du j -ième stimulus), $\boldsymbol{\varphi}$ et \mathbf{G} sont les moyennes et variances de $\boldsymbol{\alpha}$ ($\boldsymbol{\varphi}$ est un vecteur de longueur J et \mathbf{G} est une matrice diagonale de taille $J \times J$ tels que $\varphi_j = \mu_k$ et $g_{jj} = \omega_k$ si le j -ième stimulus est de type k .)

Ce modèle est alors un *modèle linéaire à données non observées*, ici les amplitudes α . Dans ce cadre, la version standard de l'algorithme *Expectation-Maximisation* (EM) proposé par Dempster *et al.* [36] permet d'estimer les paramètres $\theta_{\mathcal{P},1}$ du modèle par maximum de vraisemblance.

Dans la suite, nous supposons \mathbf{h} connue. En effet, l'estimation simultanée de la réponse \mathbf{h} et des paramètres de la distribution des amplitudes requerrait un schéma de relaxation (comme décrit dans la partie 4.3 pour estimer $\theta_{\mathcal{P},0}$) combiné avec l'algorithme EM. Cependant, la convergence d'un tel algorithme n'a pas été prouvée. Par conséquent, nous avons choisi d'estimer la réponse \mathbf{h} avant d'appliquer l'algorithme EM, en utilisant le même schéma de relaxation que celui présenté pour le modèle $\mathcal{M}_{\mathcal{P},0}$ (paragraphe 4.3).

4.4.1 Rappels sur l'algorithme Expectation-Maximisation

L'algorithme EM proposé par Dempster *et al.* [36] est un algorithme d'estimation de paramètres par maximisation de vraisemblance particulièrement adapté aux modèles à données non observées dans le cas où la maximisation de la vraisemblance des données observées ne peut être faite de façon explicite.

Dans cette partie, notons $y \in \mathbb{R}^n$ le vecteur des observations et α le vecteur des données non observées appartenant à \mathbb{R}^l . Le vecteur (y, α) est appelé vecteur des données *complètes*. Soit $\theta \in \Theta \subset \mathbb{R}^p$ le vecteur des paramètres à estimer. Supposons que la vraisemblance des données complètes (y, α) appartienne à une famille paramétrique connue analytiquement, notée $\{p(\cdot; \theta), \theta \in \Theta\}$ par rapport à une mesure borélienne ν σ -finie (par exemple la mesure de Lebesgue sur l'espace de définition de α). La vraisemblance des observations est alors définie par

$$p(y; \theta) = \int_{\mathbb{R}^l} p(y, \alpha; \theta) \nu(d\alpha)$$

Définissons la quantité suivante :

$$Q(\theta|\theta') = \int \log p(y, \alpha; \theta) p(\alpha|y; \theta') \nu(d\alpha),$$

utilisant le fait que la vraisemblance des données complètes $p(y, \alpha; \theta)$ est connue analytiquement. Le succès de cet algorithme repose largement sur sa propriété de monotonie énoncée dans la proposition suivante :

Proposition 1. *Pour tout (θ, θ') dans Θ^2 , si $Q(\theta|\theta') \geq Q(\theta|\theta)$, alors $\log p(y; \theta') \geq \log p(y; \theta)$ où $p(y; \cdot)$ est la vraisemblance des données observées.*

Cette propriété implique que tout accroissement de Q engendre un accroissement de $p(y; \cdot)$. Ainsi, lorsque la maximisation de Q est plus simple que la maximisation de $p(y; \cdot)$, des maximisations successives de Q peuvent permettre d'atteindre un maximum de $p(y; \cdot)$. S'appuyant sur ce principe, Dempster *et al.* [36] proposent donc l'algorithme itératif EM. A l'itération k , cet algorithme s'écrit en deux étapes :

Étape E : calcul de l'espérance conditionnelle de la log-vraisemblance complète

$$Q(\theta|\theta_k) = \int \log p(y, \alpha; \theta) p(\alpha|y; \theta_k) \nu(d\alpha).$$

Étape M : maximisation de cette quantité en θ

$$\theta_{k+1} = \arg \max_{\theta} Q(\theta|\theta_k)$$

La suite $(p(y; \theta_k))_{k \geq 1}$ est une suite de valeurs croissantes de la vraisemblance des données observées. La convergence de la suite $(\theta_k)_{k \geq 1}$ fournie par l'algorithme EM vers le maximum de la vraisemblance

a été largement étudiée. Citons les travaux de Dempster *et al.* [36], complétés quelques années plus tard par ceux de Wu [133]. Nous présentons un résultat de convergence de l'algorithme EM proposé par Delyon *et al.* [35] qui obtiennent pour le cadre particulier des modèles de type exponentiel des hypothèses plus simples que celles de Wu.

Soient les hypothèses

(M1) L'espace des paramètres Θ est un ouvert de \mathbb{R}^p . La fonction de vraisemblance des données complètes s'écrit

$$p(y, \boldsymbol{\alpha}; \theta) = \exp \left\{ -\psi(\theta) + \left\langle \tilde{S}(\boldsymbol{\alpha}), \phi(\theta) \right\rangle \right\}, \quad (4.4.5)$$

où $\langle \cdot, \cdot \rangle$ est le produit scalaire, $\tilde{S}(\cdot)$ est une fonction sur \mathbb{R}^l prenant ses valeurs dans un ouvert \mathcal{S} de \mathbb{R}^m . Pour tout $\theta \in \Theta$, $\int \tilde{S}(\boldsymbol{\alpha}) p(\boldsymbol{\alpha}|y; \theta) d\boldsymbol{\alpha} < \infty$.

(M2) Soit la fonction $L : \mathcal{S} \times \Theta \rightarrow \mathbb{R}$ définie par $L(s; \theta) = -\psi(\theta) + \langle s, \phi(\theta) \rangle$.

Les fonctions $\psi(\theta)$ et $\phi(\theta)$ sont deux fois continuellement différentiables sur Θ .

(M3) La fonction $\bar{s} : \Theta \rightarrow \mathcal{S}$ définie par $\bar{s}(\theta) = \int \tilde{S}(\boldsymbol{\alpha}) p(\boldsymbol{\alpha}|y; \theta) d\boldsymbol{\alpha}$ est continuellement différentiable sur Θ .

(M4) La fonction $l(\theta)$ est continuellement différentiable sur Θ et $\partial_\theta \int p(y, \boldsymbol{\alpha}; \theta) d\boldsymbol{\alpha} = \int \partial_\theta p(y, \boldsymbol{\alpha}; \theta) d\boldsymbol{\alpha}$.

(M5) Il existe une fonction $\hat{\theta} : \mathcal{S} \rightarrow \Theta$ telle que $\forall \theta \in \Theta, \forall s \in \mathcal{S}, L(s, \hat{\theta}(s)) \geq L(s, \theta)$. De plus, la fonction $\hat{\theta}(s)$ est continuellement différentiable sur \mathcal{S} .

Nous avons alors le théorème suivant.

Théorème 1 (Delyon *et al.* (1999) [35]). *Soit \mathcal{L} l'ensemble des points stationnaires de la fonction $\log p(y; \theta) : \mathcal{L} = \{\theta \in \Theta; \partial_\theta \log p(y; \theta) = 0\}$. Sous les hypothèses (M1)-(M5), et en supposant que pour tout $\theta \in \Theta$, l'adhérence de $\mathcal{L}(\theta)$ est un sous-ensemble compact de Θ , pour tout point initial $\theta_0 = \theta$, la suite $(\log p(y; \theta_k))_{k \leq 1}$ est croissante et $\lim_{k \rightarrow \infty} d(\theta_k, \mathcal{L}(\theta)) = 0$.*

4.4.2 Estimation de $\theta_{\mathcal{P},1}$ par l'algorithme EM

Considérons maintenant \mathbf{h} connue (ou estimée précédemment) et décrivons l'algorithme EM adapté à ce cadre. Son implémentation requiert de connaître la distribution de données non-observées conditionnées par les données observées. Nous détaillons son calcul ci-après.

Calcul de la distribution conditionnelle $\boldsymbol{\alpha}|y$

Par linéarité du modèle $\mathcal{M}_{\mathcal{P},0}$, la distribution de $\boldsymbol{\alpha}|y$ est une gaussienne de moyenne et variance à déterminer. Par la formule de Bayes nous avons :

$$\begin{aligned} p(\boldsymbol{\alpha}|y) &\propto p(y|\boldsymbol{\alpha})p(\boldsymbol{\alpha}) \\ &\propto \exp \left[-\frac{1}{2\sigma^2} \|\mathbf{y} - P\mathbf{l} - H\boldsymbol{\alpha}\|^2 - \frac{1}{2} (\boldsymbol{\alpha} - \boldsymbol{\varphi})' G^{-1} (\boldsymbol{\alpha} - \boldsymbol{\varphi}) \right] \\ &\propto \exp \left[-\frac{1}{2} \left[\boldsymbol{\alpha}' \left(\frac{H'H}{\sigma^2} + G^{-1} \right) \boldsymbol{\alpha} - \left(\frac{H'(\mathbf{y} - P\mathbf{l})}{\sigma^2} + G^{-1} \boldsymbol{\varphi} \right)' \boldsymbol{\alpha} - \boldsymbol{\alpha}' \left(\frac{H'(\mathbf{y} - P\mathbf{l})}{\sigma^2} + G^{-1} \boldsymbol{\varphi} \right) \right] \right] \end{aligned}$$

Posons :

$$\Gamma = \left(\frac{H'H}{\sigma^2} + G^{-1} \right)^{-1} \quad \text{et} \quad \mathbf{m} = \Gamma \left(\frac{H'(\mathbf{y} - P\mathbf{l})}{\sigma^2} + G^{-1} \boldsymbol{\varphi} \right)$$

Γ et \mathbf{m} dépendent des paramètres $\theta_{\mathcal{P},1}$ et \mathbf{l} . Nous avons alors :

$$\boldsymbol{\alpha}|\mathbf{y} \sim \mathcal{N}(\mathbf{m}, \Gamma) \quad , \quad E[\boldsymbol{\alpha}_j|\mathbf{y}; \theta_{\mathcal{P},1}] = \mathbf{m}_j \quad , \quad E[\boldsymbol{\alpha}_j^2|\mathbf{y}; \theta_{\mathcal{P},1}] = \Gamma_{j,j} + \mathbf{m}_j^2$$

où α_j est l'amplitude du j -ième stimulus (tous types de stimuli confondus) .

Implémentation de l'algorithme EM dans le cadre des données IRMf

Une fois la loi conditionnelle $p(\boldsymbol{\alpha}|\mathbf{y})$ explicitée, nous pouvons décrire l'algorithme EM implémenté ici. A l'itération 0, l'algorithme est initialisé sur des paramètres arbitraires $\theta_{\mathcal{P},1}^{(0)}$. Puis, à l'itération ℓ , l'algorithme EM s'écrit :

Étape E : Calcul des espérances conditionnelles $E[\boldsymbol{\alpha}^{(\ell)}|\mathbf{y}; \theta_{\mathcal{P},1}^{(\ell-1)}]$, $E[(\boldsymbol{\alpha}^2)^{(\ell)}|\mathbf{y}; \theta_{\mathcal{P},1}^{(\ell-1)}]$ avec les formules précédentes en fonction de $\theta_{\mathcal{P},1}^{(\ell-1)}$ et $\mathbf{l}^{(\ell-1)}$

Étape M : Mise à jour des paramètres :

- Calcul de $(\mu_k^{(\ell)})$ et $(\omega_k^{(\ell)}) : \forall k \in \mathcal{P}$

$$\mu_k^{(\ell)} = \frac{1}{J_k} \sum_{j=1}^{J_k} E[\alpha_{k,j}^{(\ell)}|\mathbf{y}; \theta_{\mathcal{P},1}^{(\ell-1)}] \quad ; \quad \omega_k^{(\ell)} = \frac{1}{J_k} \sum_{j=1}^{J_k} E[(\alpha_{k,j}^2)^{\ell}|\mathbf{y}; \theta_{\mathcal{P},1}^{(\ell-1)}] - (\mu_k^{(\ell)})^2$$

- Calcul de $\mathbf{l}^{(\ell)}$:

$$\mathbf{l}^{(\ell)} = (PP')^{-1}P'[\mathbf{y} - HE[\boldsymbol{\alpha}^{(\ell)}|\mathbf{y}; \theta_{\mathcal{P},1}^{(\ell-1)}]]$$

- Calcul de $\sigma^{2(\ell)}$:

$$\sigma^{2(\ell)} = \frac{1}{n} \left\| \mathbf{y} - P\mathbf{l}^{(\ell)} - HE[\boldsymbol{\alpha}^{(\ell)}|\mathbf{y}; \theta_{\mathcal{P},1}^{(\ell-1)}] \right\|^2 + \frac{1}{n} H'\Gamma H$$

Le modèle $\mathcal{M}_{\mathcal{P},1}$ est un modèle exponentiel. Dans ce cadre, les résultats de convergence du théorème 1 de Delyon *et al.* [35] s'appliquent. Nous illustrons ses performances sur données simulées dans au paragraphe 4.8.1. L'algorithme sera ensuite appliqué sur données réelles dans les parties 4.8.2 et 4.8.3.

4.5 Estimation des paramètres $\theta_{\mathcal{P},1}$ dans le cas d'une contrainte de positivité

Nous avons souligné dans la partie 4.2 (remarque 10) qu'une contrainte de positivité pouvait être imposée sur les amplitudes. La loi sur les amplitudes est une log-normale :

$$\log(\alpha_{k,j}) \sim \mathcal{N}(\mu_k, \omega_k).$$

Dans ce cas, l'algorithme EM ne peut être appliqué puisque le calcul des espérances conditionnelles ne peut être fait de façon explicite (**Étape E**). Nous avons alors recours à une version stochastique de l'algorithme EM dont nous rappelons le principe et les conditions de convergence dans le paragraphe suivant.

4.5.1 Versions stochastiques de l'algorithme EM

Soient $y \in \mathbb{R}^n$ sont les données observées. $\boldsymbol{\alpha} \in \mathbb{R}^l$ représente les données non observées tandis que $\theta \in \Theta \subset \mathbb{R}^p$ est le paramètre à estimer.

Une version stochastique de l'algorithme EM

Lorsque l'espérance conditionnelle de la log-vraisemblance des données complètes ne peut pas être calculée à l'étape E de l'algorithme EM, Wei et Tanner [127] proposent l'algorithme MCEM (*Monte Carlo EM*), qui consiste naturellement à l'itération k de l'algorithme EM à approcher l'intégrale $Q(\theta|\theta_k)$ par une méthode de Monte Carlo. A chaque itération, un nombre T de variables aléatoires $(\alpha_k^t)_{1 \leq t \leq T}$ est simulé à partir de la loi a posteriori $p(\alpha|y; \theta_k)$ et la fonction Q est approchée par une moyenne empirique

$$Q(\theta|\theta_k) \approx \frac{1}{T} \sum_{t=1}^T \log p(y, \alpha_k^t; \theta_k).$$

Wei et Tanner [127] illustrent leur article par simulation sur deux exemples, mais aucun résultat théorique de convergence n'est présenté. Un algorithme MCEM est également proposé par Meng and Rubin [89]. Cependant l'algorithme MCEM peut avoir des problèmes numériques, tels qu'une convergence lente ou même inexistante (voir Booth and Hobert [19]).

Une alternative à la fois au problème de convergence presque sûre mais aussi au problème numérique repose sur une approximation stochastique de l'étape E et a été proposée par Delyon *et al.* et Kuhn et Lavielle [35, 73]. Dans ce cas, l'étape E de cet algorithme, appelé algorithme SAEM (*Stochastic Approximation EM*), est divisée en une étape de simulation et une étape d'approximation stochastique. A l'itération k , l'étape de simulation S consiste à simuler une réalisation α_k des données manquantes α d'après la loi conditionnelle $p(\alpha|y; \theta_k)$. L'étape d'approximation stochastique SA introduit une suite décroissante de pas positifs (γ_k) et réalise l'approximation stochastique suivante à l'itération k

$$Q_{k+1}(\theta) = Q_k(\theta) + \gamma_k(\log p(y, \alpha_k; \theta_k) - Q_k(\theta)),$$

où α_k est simulé sous la loi $p(\alpha|y, \theta_k)$. L'étape M devient

$$\theta_{k+1} = \arg \max_{\theta \in \Theta} Q_{k+1}(\theta).$$

Sous l'hypothèse **(M1)** d'un modèle du type exponentiel, il suffit de réaliser l'approximation stochastique sur une statistique exhaustive \tilde{S} du modèle.

Résultat de convergence Le résultat de convergence, obtenu par Delyon *et al.* [35], nécessite, outre les hypothèses de régularité **(M1)**-**(M5)**, des hypothèses relatives à la méthode de simulation des données manquantes α et à la suite de pas (γ_k) .

Supposons que les variables aléatoires $\alpha_1, \dots, \alpha_k$ sont définies sur le même espace de probabilité (Ω, \mathcal{A}, P) . Définissons $\mathcal{F} = \{\mathcal{F}_k\}_{k \geq 0}$ la famille de σ -algèbre croissante générée par les variables aléatoires $\alpha_1, \dots, \alpha_k$.

Soient les hypothèses :

(SAEM 1) Pour tout $k > 0$, $0 \leq \gamma_k \leq 1$, $\sum_{k=1}^{\infty} \gamma_k = \infty$ et $\sum_{k=1}^{\infty} \gamma_k^2 < \infty$

(SAEM 2) $l : \Theta \rightarrow \mathbb{R}$ et $\hat{\theta} = \mathcal{S} \rightarrow \Theta$ sont m fois différentiables,

(SAEM 3) 1. Pour toute fonction positive borélienne Φ , $E[\Phi(\alpha_{k+1})|\mathcal{F}_k] = \int \Phi(\alpha)p(\alpha|y; \theta_k)d\alpha$

2. Pour tout $\theta \in \Theta$, $\int \|\tilde{S}(z)\|^2 p(\alpha|y; \theta)d\alpha < \infty$ et la fonction $\theta \rightarrow \text{Cov}_\theta(\tilde{S}(z))$ est continue par rapport à θ .

Théorème 2 (Delyon *et al.* [35]). *Sous les hypothèses (M1)-(M5), et (SAEM1)-(SAEM3), si la suite $(s_k)_{k \geq 0}$ prend ses valeurs dans un sous-ensemble compact de \mathcal{S} , la suite $(\theta_k)_{k \geq 0}$ générée par SAEM converge vers un point stationnaire de la vraisemblance observée $p(y)$.*

Extension dans le cas où l'étape de simulation n'est pas directe Cet algorithme requiert donc la simulation d'une seule réalisation des données α à chaque itération, ce qui réduit de façon conséquente le temps de calcul par rapport aux algorithmes MCEM. Néanmoins d'un point de vue pratique, la mise en oeuvre de l'algorithme SAEM nécessite de savoir simuler des réalisations de variables aléatoires suivant la loi a posteriori $p(\alpha|y; \theta)$, ce qui n'est pas toujours possible.

Kuhn et Lavielle [72] ont donc proposé de combiner l'algorithme SAEM avec une méthode de Monte Carlo par chaînes de Markov (MCMC) dont nous rappelons le principe au paragraphe suivant pour contourner ce problème. Dans ce nouvel algorithme SAEM, seule l'étape S de l'algorithme est modifiée. A l'itération k , l'algorithme SAEM est alors réalisée à travers ces étapes :

- **Etape E :**

Etape S : Simulation de α_k par algorithme de Metropolis-Hastings d'après une probabilité de transition Π_{θ_k} correspondant à une chaîne de Markov de loi stationnaire $p(\alpha|y, \theta_k)$

Etape SA : Approximation stochastique des statistiques suffisantes du modèle

$$s_{k+1} = s_k + \gamma_k(\tilde{S}(y, \alpha_k) - s_k)$$

- **Etape M :** Maximisation de la vraisemblance complète

$$\theta_{k+1} = \hat{\theta}(s_{k+1})$$

Kuhn et Lavielle [72] étendent le résultat de convergence de Delyon *et al.* [35], sous l'hypothèse supplémentaire suivante :

(SAEM 3') 1. La chaîne $(\alpha_k)_{k \geq 0}$ prend ses valeurs dans un compact \mathcal{E} de $\mathcal{R} \subset \mathbb{R}^d$.

2. Pour tout compact V de Θ , il existe une constante réelle L telle que pour tout (θ, θ') dans V^2

$$\sup_{\{(\alpha), (\alpha')\} \in \mathcal{E}} |\Pi_{\theta}(\alpha'|\alpha) - \Pi_{\theta'}(\alpha'|\alpha)| \leq L \|\theta - \theta'\|_{\mathbb{R}^q}$$

3. La probabilité de transition Π_{θ} fournit une chaîne uniformément ergodique dont la probabilité invariante est la distribution conditionnelle $p(\cdot|y; \theta)$

$$\exists K_{\theta} \in \mathbb{R}^+, \quad \exists \rho_{\theta} \in]0, 1[\quad | \quad \forall k \in \mathbb{N} \quad \|\Pi_{\theta}^k(\cdot|\alpha) - p(\cdot|y; \theta)\|_{TV} \leq K_{\theta} \rho_{\theta}^k$$

où $\|\cdot\|_{TV}$ est la norme en variation totale, et

$$K = \sup_{\theta \in \Theta} K_{\theta} < \infty \quad \text{et} \quad \rho = \sup_{\theta \in \Theta} \rho_{\theta} < 1$$

4. La fonction S_h est bornée sur \mathcal{E} .

Théorème 3 (Kuhn et Lavielle [72]). *Sous les hypothèses (M1)-(M5), (SAEM1)-(SAEM2) et (SAEM3'), et si la suite $(s_k)_{k \geq 0}$ prend ses valeurs dans un sous-ensemble compact de \mathcal{S} . Alors avec probabilité 1, la suite $(\theta_k)_{k \geq 0}$ générée par SAEM-MCMC converge vers un maximum (local) de la vraisemblance observée $p(y; \cdot)$.*

Rappels sur les méthode MCMC

On appelle algorithme de *Monte Carlo par chaînes de Markov* (MCMC) toute méthode produisant une chaîne de Markov ergodique ayant pour loi stationnaire une loi d'intérêt $\pi(\cdot)$. Ces méthodes sont généralement utilisées lorsque la loi $\pi(\cdot)$ n'est pas simulable directement et/ou lorsqu'elle est connue à une constante de normalisation près.

Plusieurs algorithmes MCMC ont été proposés, qui sont des versions hybrides ou des généralisations de l'échantillonneur de Gibbs et de l'algorithme de Metropolis-Hastings. Robert et Casella proposent une revue de ces méthodes [106].

Nous rappelons ici leurs principes, plusieurs algorithmes hybrides ayant été développés dans cette thèse.

L'échantillonneur de Gibbs et l'algorithme de Metropolis-Hastings sont basés sur le même principe : ils assurent l'existence de noyaux de transition, noté $\Pi(\boldsymbol{\alpha}, \boldsymbol{\alpha}')d\boldsymbol{\alpha}'$, permettant de générer une chaîne de Markov de loi stationnaire $\pi(\cdot)$. Notons $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_d) \in \mathbb{R}^d$.

Échantillonneur de Gibbs L'algorithme de Gibbs repose sur la probabilité de transition suivante

$$\Pi(\boldsymbol{\alpha}, \boldsymbol{\alpha}') = \prod_{i=1}^d q_i(\boldsymbol{\alpha}'_i | \boldsymbol{\alpha}'_1, \dots, \boldsymbol{\alpha}'_{i-1}, \boldsymbol{\alpha}_{i+1}, \dots, \boldsymbol{\alpha}_d)$$

où

$$q_i(\boldsymbol{\alpha}_i | \boldsymbol{\alpha}_j, j \neq i) = \pi(\boldsymbol{\alpha}_i | \boldsymbol{\alpha}_j, j \neq i)$$

est la i -ème densité conditionnelle. A partir de la valeur $\boldsymbol{\alpha}^{(t)}$ de la chaîne obtenue à l'itération t , la valeur $\boldsymbol{\alpha}^{(t+1)}$ est obtenue après la réalisation de d étapes de simulation à l'aide des lois q_i appelées aussi lois instrumentales (*proposal distributions*) :

Étape 1 : Simulation de $\boldsymbol{\alpha}_1^{(t+1)}$ suivant $q_1(\cdot | \boldsymbol{\alpha}_2^{(t)}, \dots, \boldsymbol{\alpha}_d^{(t)})$

Étape i , $i = 2, \dots, d-1$: Simulation de $\boldsymbol{\alpha}_i^{(t+1)}$ suivant $q_i(\cdot | \boldsymbol{\alpha}_1^{(t+1)}, \dots, \boldsymbol{\alpha}_{i-1}^{(t+1)}, \boldsymbol{\alpha}_i^{(t)}, \dots, \boldsymbol{\alpha}_d^{(t)})$

Étape d : Simulation de $\boldsymbol{\alpha}_d^{(t+1)}$ suivant $q_d(\cdot | \boldsymbol{\alpha}_1^{(t+1)}, \dots, \boldsymbol{\alpha}_{d-1}^{(t+1)})$

La convergence de cet algorithme et l'ergodicité géométrique ou uniforme de la chaîne simulée sont assurées, par exemple, sous une condition de minoration proposée par Tierney [121]. Robert et Casella [106] détaillent d'autres conditions de convergence. La principale limite de l'échantillonneur de Gibbs réside dans le choix restreint de lois instrumentales possibles. En effet, il est nécessaire de connaître π analytiquement ou de savoir simuler chacune des densités conditionnelles pour implémenter cet algorithme. Lorsque ce n'est pas le cas, il est possible d'avoir recours à l'algorithme de Metropolis-Hastings.

Algorithme de Metropolis-Hastings L'algorithme de Metropolis-Hastings repose sur l'utilisation d'une loi instrumentale $q(\boldsymbol{\alpha}' | \boldsymbol{\alpha})$ ayant une densité par rapport à la mesure dominante du modèle. En pratique, cet algorithme ne peut être mis en place que si cette loi instrumentale est simulable rapidement et de densité disponible analytique (à constante près) ou symétrique, i.e. telle que $q(\boldsymbol{\alpha}' | \boldsymbol{\alpha}) = q(\boldsymbol{\alpha} | \boldsymbol{\alpha}')$.

L'algorithme de Metropolis-Hastings associé à la loi d'intérêt π et de loi instrumentale q produit une chaîne de Markov $\{\boldsymbol{\alpha}^{(t)}; t \geq 0\}$ avec $\boldsymbol{\alpha}^{(t)} = (\boldsymbol{\alpha}_1^{(t)}, \dots, \boldsymbol{\alpha}_d^{(t)})$ fondée sur la transition suivante : étant donné $\boldsymbol{\alpha}^{(t)}$,

1. Simuler un candidat $\boldsymbol{\alpha}'$ selon la loi $\boldsymbol{\alpha}' \sim q(\boldsymbol{\alpha}' | \boldsymbol{\alpha}^{(t)})$
2. Prendre

$$\boldsymbol{\alpha}^{(t+1)} = \begin{cases} \boldsymbol{\alpha}' & \text{avec probabilité } \rho(\boldsymbol{\alpha}^{(t)}, \boldsymbol{\alpha}') \\ \boldsymbol{\alpha}^{(t)} & \text{avec probabilité } 1 - \rho(\boldsymbol{\alpha}^{(t)}, \boldsymbol{\alpha}') \end{cases}$$

où

$$\rho(\boldsymbol{\alpha}^{(t)}, \boldsymbol{\alpha}') = \min \left(\frac{\pi(\boldsymbol{\alpha}')}{\pi(\boldsymbol{\alpha}^{(t)})} \frac{q(\boldsymbol{\alpha}^{(t)} | \boldsymbol{\alpha}')}{q(\boldsymbol{\alpha}' | \boldsymbol{\alpha}^{(t)})}, 1 \right)$$

Remarque 14. • Cette procédure accepte systématiquement les candidats proposés $\boldsymbol{\alpha}'$ tels que le quotient $\frac{\pi(\boldsymbol{\alpha}^{(t)})}{q(\boldsymbol{\alpha}' | \boldsymbol{\alpha}^{(t)})}$ est supérieur au précédent.

- D'autre part, la probabilité d'acceptation ne dépend que du rapport des distributions, ce qui permet de contourner le problème du calcul, souvent impossible, de la constante de normalisation.

L'universalité de cet algorithme en fait un outil très puissant. En effet, quelque soit la loi instrumentale q , par définition de l'algorithme, π est loi invariante de la chaîne de Markov produite (dès lors que les supports de π et q coïncident). Cependant, cette universalité reste théorique si la loi instrumentale q ne simule que rarement des "bons" candidats, c'est-à-dire des candidats dans la région où π a la plus grande masse. Le choix de q est donc important et dépend fortement de π . Deux lois instrumentales particulières sont en général utilisées.

- Dans le premier cas, on considère une loi instrumentale possédant la propriété suivante $q(\boldsymbol{\alpha}'|\boldsymbol{\alpha}^{(t)}) = q(\boldsymbol{\alpha}')$. On parle alors d'algorithme de Metropolis-Hastings *indépendant*, la simulation du candidat étant indépendante de la valeur précédente de la chaîne. Dans ce cas, lorsque la loi instrumentale q est proche de la loi π , la convergence est rapide. Cependant dans certains cas, cet algorithme est très sensible aux valeurs initiales et aux états absorbants de la chaîne. Il est donc recommandé de ne pas l'utiliser seul.
- La deuxième classe de lois instrumentales utilisées repose sur la propriété de symétrie $q(\boldsymbol{\alpha}'|\boldsymbol{\alpha}^{(t)}) = q(\boldsymbol{\alpha}' - \boldsymbol{\alpha}^{(t)}) = q(\boldsymbol{\alpha}^{(t)} - \boldsymbol{\alpha}')$, on parle alors d'une *marche aléatoire*. L'idée sous-jacente consiste à prendre en compte la dernière valeur simulée pour simuler la suivante. La chaîne de Markov est alors géométriquement ergodique sous des conditions générales raisonnables, par exemple lorsque la loi π appartient à la famille exponentielle et est suffisamment régulière. En général, cette marche aléatoire dépend d'un paramètre d'échelle δ dont le choix est délicat, un exemple courant consistant à simuler $\boldsymbol{\alpha}'$ selon la loi $\mathcal{N}(\boldsymbol{\alpha}^{(t)}, \delta I_d)$. Si δ est trop grand, une large proportion de candidats vont être rejetés. Inversement si δ est trop petit, la marche aléatoire n'accepte que des candidats dans un petit voisinage et se déplace très lentement dans l'espace des paramètres. Lorsque la dimension d de la chaîne de Markov est petite, différents auteurs [52, 107] recommandent d'adapter ce paramètre d'échelle afin d'assurer un taux de candidats *acceptable* proche de 30%. Cependant, dès que d augmente, aucune procédure n'est proposée.

Algorithme MCMC hybride Le point faible de l'algorithme de Metropolis-Hastings est le choix de la loi instrumentale. Cette loi peut être très différente de la loi π visée, et l'algorithme risque alors de simuler "grossièrement" π . Ce n'est pas le cas pour l'échantillonneur de Gibbs, dont la loi instrumentale est directement déduite de la loi π . Cependant, la structure composée de l'échantillonneur de Gibbs peut être une faiblesse. Par exemple, pour une loi de mélange π , l'échantillonneur de Gibbs peut n'explorer qu'une seule des deux composantes du support de π . La liberté du choix de q dans l'algorithme de Metropolis-Hastings permet parfois de remédier à ce problème, en particulier au travers de l'optimisation des paramètres d'échelle.

Pour conserver les avantages de ces deux méthodes, Tierney [121] propose une approche hybride. On appelle alors algorithme *MCMC hybride* ou *Gibbs hybride* une méthode MCMC combinant simultanément des étapes d'un échantillonneur de Gibbs avec des étapes d'algorithmes de Metropolis-Hastings. Si Π_1, \dots, Π_n sont les noyaux de transition correspondants à ces étapes, on appelle *cycle* de Π_1, \dots, Π_n l'algorithme de noyau $\Pi^* = \Pi_1 \circ \dots \circ \Pi_n$, où \circ est l'opérateur de composition de fonctions. Une composition de noyaux associés à la même loi stationnaire π correspond à un noyau Π^* de loi stationnaire π . L'ergodicité uniforme de la chaîne de Markov est assurée dès que l'un des noyaux la garantit.

Différentes versions de méthodes MCMC hybrides sont développées et proposées au cours de cette thèse, à commencer par l'estimation des paramètres de la loi des amplitudes dans le cas d'une contrainte de positivité, dans le cadre d'un algorithme SAEM-MCMC.

4.5.2 Estimation de $\theta_{\mathcal{P},1}$ par l'algorithme SAEM-MCMC dans le cas d'une contrainte de positivité

Nous décrivons l'algorithme SAEM-MCMC implémenté dans le cas du modèle à amplitudes variables positives. Soit $(\gamma_\ell)_{\ell \in \mathbb{N}}$ une suite décroissante de nombres de l'intervalle $[0, 1]$, telle que $\sum_\ell \gamma_\ell = +\infty$ et $\sum_\ell \gamma_\ell^2 < \infty$. Alors à l'itération ℓ l'algorithme SAEM-MCMC s'écrit :

Étape E :

Étape S : Générer $\alpha^{(\ell)}$ par M itérations d'une procédure MCMC de distribution invariante $p(\alpha | \mathbf{y}; \theta_{\mathcal{P},1}^{(\ell-1)})$

Étape SA : Approximation stochastique des statistiques suffisantes : pour $k = 1, \dots, P$,

$$S_{k,1} = \sum_{j=1}^{J_k} \log(\alpha_{k,j}^{(\ell)}) \quad , \quad S_{k,2} = \sum_{j=1}^{J_k} \left(\log(\alpha_{k,j}^{(\ell)}) \right)^2 \quad , \quad S_3 = \left\| \mathbf{y} - P\mathbf{l}^{(\ell-1)} - H\alpha^{(\ell)} \right\|^2$$

$$s_{k,1}^{(\ell)} = s_{k,1}^{(\ell-1)} + \gamma_\ell \left[S_{k,1} - s_{k,1}^{(\ell-1)} \right] \quad , \quad s_{k,2}^{(\ell)} = s_{k,2}^{(\ell-1)} + \gamma_\ell \left[S_{k,2} - s_{k,2}^{(\ell-1)} \right]$$

$$s_3^{(\ell)} = s_3^{(\ell-1)} + \gamma_\ell \left[S_3 - s_3^{(\ell-1)} \right]$$

Étape M : Maximisation de la vraisemblance complète : pour $k \in \mathcal{P}$,

$$\mathbf{l}^{(\ell)} = (PP')^{-1}P' \left[\mathbf{y} - H\alpha^{(\ell)} \right]$$

$$\widehat{\mu}_k^{(\ell)} = \frac{1}{J_k} s_{k,1}^{(\ell)} \quad , \quad \widehat{V}_k^{(\ell)} = \frac{1}{J_k} s_{k,2}^{(\ell)} - \left(\widehat{\mu}_k^{(\ell)} \right)^2 \quad , \quad \widehat{\sigma}^2^{(\ell)} = \frac{1}{n} s_3^{(\ell)}$$

Description de l'Étape S A l'itération ℓ de l'algorithme SAEM-MCMC, l'étape S consiste en la mise en place d'une procédure MCMC. Pour les raisons évoquées au paragraphe 4.5.1, nous mettons en place un algorithme MCMC hybride. Pour cela nous combinons les noyaux suivants :

1. D'une part, la loi instrumentale est la loi a priori, i.e. la loi log-normale de paramètre courant $\theta_{\mathcal{P},1}^{(\ell-1)}$.
2. D'autre part, nous utilisons une marche aléatoire sur une composante à la fois de α . Plus précisément, à l'itération $m \in \{1, \dots, M\}$ de l'algorithme MCMC, nous proposons un candidat α' tel que, si j est l'une de ses composantes :

$$\alpha'_j = \begin{cases} \alpha_j^{(m)} + \mathcal{N}(0, \delta) & \text{si } m - j \equiv 0[J] \\ \alpha_j^{(m)} & \text{sinon} \end{cases}$$

En pratique cette combinaison de lois instrumentales permet une bonne convergence de l'algorithme (voir résultats sur données simulées au paragraphe 4.8.1).

4.6 Calcul de la vraisemblance des modèles

La sélection de modèles nécessite le calcul des vraisemblances des modèles $\mathcal{M}_{\mathcal{P},0}$ et $\mathcal{M}_{\mathcal{P},1}$. Dans cette partie, nous détaillons les calculs de ces quantités. Considérons une famille de types de stimuli \mathcal{P} fixée.

4.6.1 Vraisemblance du modèle $\mathcal{M}_{\mathcal{P},0}$

Sous le modèle $\mathcal{M}_{\mathcal{P},0}$, le vecteur \mathbf{y} est gaussien. Par conséquent, pour toute valeur de $\theta_{\mathcal{P},0}$

$$L_0(\mathbf{y}; \theta_{\mathcal{P},0}) = \frac{1}{2} \left(-n \log(2\pi) - n \log(\sigma^2) - \frac{U_0(\boldsymbol{\mu}, \mathbf{l}, \mathbf{h})}{\sigma^2} \right),$$

$U_0(\boldsymbol{\mu}, \mathbf{l}, \mathbf{h})$ ayant été défini par l'expression (4.3.4). Au maximum de la vraisemblance, $(\hat{\theta}_{\mathcal{P},0})$, la log-vraisemblance vaut :

$$L_0(\mathbf{y}; \hat{\theta}_{\mathcal{P},0}) = \frac{n}{2} \left(-\log(2\pi) - \log(\hat{\sigma}_0^2) - 1 \right)$$

4.6.2 Vraisemblances du modèle $\mathcal{M}_{\mathcal{P},1}$ avec et sans contrainte de positivité

Sous le modèle $\mathcal{M}_{\mathcal{P},1}$, nous avons vu au chapitre précédent que la loi conditionnelle de $\boldsymbol{\alpha}|\mathbf{y}$ est une distribution gaussienne de paramètres

$$\Gamma = \left(\frac{H'H}{\sigma^2} + G^{-1} \right)^{-1} \quad \mathbf{m} = \Gamma \left(\frac{H'(\mathbf{y} - P\mathbf{l})}{\sigma^2} + G^{-1}\boldsymbol{\varphi} \right)$$

où $\boldsymbol{\varphi}$ et G sont les moyenne et variance des amplitudes a priori. Alors la log-vraisemblance du modèle s'écrit :

$$\begin{aligned} L_1(\mathbf{y}; \theta_{\mathcal{P},1}) &= \log p(\mathbf{y}|\boldsymbol{\alpha}) + \log p(\boldsymbol{\alpha}) - \log p(\boldsymbol{\alpha}|\mathbf{y}) \\ &= \frac{1}{2} \left(-n \log(2\pi) - n \log(\sigma^2) - \frac{\|\mathbf{y} - P\mathbf{l} - H\boldsymbol{\alpha}\|^2}{\sigma^2} \right) \\ &\quad + \frac{1}{2} \left(-J \log(2\pi) - \log(|G|) - (\boldsymbol{\alpha} - \boldsymbol{\varphi})' G^{-1} (\boldsymbol{\alpha} - \boldsymbol{\varphi}) \right) \\ &\quad - \frac{1}{2} \left(-J \log(2\pi) - \log(|\Gamma|) - (\boldsymbol{\alpha} - \mathbf{m})' \Gamma^{-1} (\boldsymbol{\alpha} - \mathbf{m}) \right) \\ L_1(\mathbf{y}; \theta_{\mathcal{P},1}) &= \frac{1}{2} \left(-n \log(2\pi) + \log(|\Gamma G^{-1}|) - n \log(\sigma^2) - \frac{(\mathbf{y} - P\mathbf{l})'(\mathbf{y} - P\mathbf{l})}{\sigma^2} - \boldsymbol{\varphi}' G^{-1} \boldsymbol{\varphi} + \mathbf{m}' \Gamma^{-1} \mathbf{m} \right) \end{aligned}$$

Calcul de la vraisemblance dans le cas d'une loi log-normale

Dans ce cas, le calcul de la vraisemblance ne peut être fait de façon exacte. Nous avons donc recours à une procédure de Monte Carlo pour calculer cette quantité. Plus précisément, soit $(\boldsymbol{\alpha}^{(m)})_{m=1\dots M_1}$ un échantillon des données non observées simulées sous la loi log-normale :

$$\hat{l}_{M_1} = \log \left[\frac{1}{M_1} \sum_{m=1}^{M_1} p(\mathbf{y} | \boldsymbol{\alpha}^{(m)}; \theta_{\mathcal{P},1}) \right]$$

Cet estimateur est asymptotiquement sans biais. Cependant, sa variance peut-être arbitrairement grande. Dans ce cas, cet estimateur est remplacé par l'estimateur d'échantillonnage d'importance [60] défini de la façon suivante : soit $(\boldsymbol{\alpha}^{(m)})_{m=1\dots M}$ une réalisation des données non observées sous une distribution $\tilde{\pi}$, nous estimons la log-vraisemblance par

$$\hat{l}_{M_1} = \log \left[\frac{1}{M_1} \sum_{m=1}^{M_1} p(\mathbf{y} | \boldsymbol{\alpha}^{(m)}; \theta_{\mathcal{P},1}) \frac{\pi(\boldsymbol{\alpha}^{(m)})}{\tilde{\pi}(\boldsymbol{\alpha}^{(m)})} \right]$$

Cet estimateur est asymptotiquement sans biais mais de variance moindre si la distribution $\tilde{\pi}$ est choisie de façon adéquate (voir [60]). Le but est de choisir $\tilde{\pi}$ le plus proche possible de la distribution conditionnelle $p(\boldsymbol{\alpha}|\mathbf{y}; \hat{\theta}_{\mathcal{P},1})$. Pour cela, dans les applications numériques du paragraphe 4.8, nous prenons une distribution log-normale de paramètres les moyenne et variance empiriques $\log(\boldsymbol{\alpha})$.

4.7 Sélection de modèles

Une fois l'estimation des paramètres faite et les vraisemblances calculées, nous sommes en mesure de faire de la sélection de modèles. Les objectifs sont les suivants :

- à famille de stimuli \mathcal{P} fixée, valider ou non l'hypothèse de variabilité inter-occurrences des amplitudes de la réponse BOLD i.e choisir entre les modèles $\mathcal{M}_{\mathcal{P},0}$ et $\mathcal{M}_{\mathcal{P},1}$ (problématique $\mathcal{P}b_1$).
- sélectionner les types de stimuli ayant une influence significative sur le signal observé i.e. choisir la famille \mathcal{P} à inclure dans le modèle (problématique $\mathcal{P}b_2$).

La méthode proposée pour mener à bien ces deux étapes est décrite dans cette partie.

4.7.1 $\mathcal{M}_{\mathcal{P},0}$ versus $\mathcal{M}_{\mathcal{P},1}$: problématique $\mathcal{P}b_1$

Considérons une famille de types de stimuli \mathcal{P} fixée. Nous devons résoudre le problème suivant : choisir le meilleur modèle entre $\mathcal{M}_{\mathcal{P},0}$ et $\mathcal{M}_{\mathcal{P},1}$. Ceci revient à faire un test d'hypothèses de

$$\mathcal{H}_0 : \text{"}\omega_k = 0 \forall k \in \mathcal{P}\text{"} \quad \text{contre} \quad \mathcal{H}_1 : \text{"}\omega_k > 0\text{"}$$

Nous considérons la statistique du test de rapport des vraisemblances suivante :

$$\xi_{\mathcal{P}} = 2 \left[L_1(y; \hat{\theta}_{\mathcal{P},1}) - L_0(y; \hat{\theta}_{\mathcal{P},0}) \right]$$

Comme nous l'avions remarqué au paragraphe 4.2, la différence du nombre de paramètres à estimer entre les modèles $\mathcal{M}_{\mathcal{P},0}$ et $\mathcal{M}_{\mathcal{P},1}$ est $\text{Card}(\mathcal{P})$. Dans ce paragraphe, nous notons $P = \text{Card}(\mathcal{P})$.

Déterminer la distribution asymptotique de $\xi_{\mathcal{P}}$ sous l'hypothèse \mathcal{H}_0 n'est pas une question triviale. Définissons $\theta = (\mu_k, \omega_k, \sigma^2)_{k \in \mathcal{P}}$. Soient H_0 et H_1 les deux sous-ensembles de \mathbb{R}^{2P+1} suivants :

$$H_0 = \mathbb{R}^P \times \{0\}_{\mathbb{R}^P} \times \mathbb{R}^{+*} \quad H_1 = \{\theta \in \mathbb{R}^{2P} \times \mathbb{R}^{+*} \mid \forall k \in \mathcal{P}, \omega_k \geq 0\}.$$

Alors nous cherchons la distribution asymptotique du test du rapport de vraisemblance du test de $\mathcal{H}_0 : \theta \in H_0$ contre $\mathcal{H}_1 : \theta \in H_1$. Chernoff [30] a montré que dans le cas où les ensembles H_0 et H_1 sont localement égaux à des espaces linéaires, la distribution asymptotique de la statistique de test est une distribution du khi-deux à P degrés de liberté, P étant la différence entre les dimensions locales de H_0 et H_1 . Ce résultat ne peut être appliqué ici dans la mesure où les paramètres à tester $\{\omega_k, k \in \mathcal{P}\}$ se trouvent sur le bord du domaine des paramètres H_1 ($\{\omega_k \geq 0, \forall k \in \mathcal{P}\}$) (l'espace H_1 est non localement linéaire). Dans ce cas il est possible de montrer que la distribution asymptotique de la statistique de test est un mélange de distributions de khi-deux.

En effet, ce test entre dans la famille plus large des tests du type $\mathcal{H}_0 : \{g(\theta) = 0\}$ versus $\mathcal{H}_1 : \{g(\theta) \geq 0, g(\theta) \neq 0\}$ où g est une fonction de \mathbb{R}^r dans \mathbb{R}^P de rang P [57]. Soit θ_0 la vraie valeur du paramètre. Soit $I(\theta_0)$ la matrice d'information de Fisher en θ_0 . Soit $R \in \mathcal{G}\ell_P(\mathbb{R})$ telle que :

$$R = \left[\frac{\partial g}{\partial \theta}(\theta_0) I(\theta_0)^{-1} \frac{\partial g'}{\partial \theta}(\theta_0) \right]^{-1/2}.$$

Alors, d'après Gourieroux et Monfort [57] sous l'hypothèse \mathcal{H}_0 , la statistique $\xi_{\mathcal{P}}$ suit asymptotiquement un mélange de distributions de khi-deux :

$$\xi \sim \sum_{d=0}^P m(d) \chi^2(d)$$

où $m(d)_{d=1,\dots,P}$ est une suite de poids ($m(d) \geq 0$ et $\sum_{d=0}^P m(d) = 1$) et où $\chi^2(0)$ désigne la masse ponctuelle en zéro. Plus précisément, les poids s'écrivent de la façon suivante. Notons $(R_j)_{j=1,\dots,P}$ les

vecteurs colonnes de la matrice R et $(\tilde{R}_j)_{j=1,\dots,P}$ les vecteurs colonnes de son inverse R^{-1} . Pour toute partie A de $\{1, \dots, P\}$, définissons les cônes C_A par :

$$C_A = \left\{ \sum_{j \in A} \lambda_j \tilde{R}_j + \sum_{j \in A^c} \lambda_j R_j; (\lambda_j)_{j=1,\dots,P} \in \mathbb{R}^P, \lambda_j \geq 0 \forall j \in A, \lambda_j \leq 0 \forall j \in A^c \right\}.$$

Alors si $Z \sim \mathcal{N}(0, \mathbf{I}_P)$, les poids $m(d)$ s'écrivent :

$$m(d) = \sum_{A: \text{Card} A^c = d} P[Z \in C_A]$$

La complexité du calcul des poids $m(d)$ rend inenvisageable le calcul explicite de la distribution asymptotique de $\xi_{\mathcal{P}}$. Nous proposons donc de tabuler la distribution de la statistique de test $\xi_{\mathcal{P}}$ de façon numérique. Pour cela, pour un paradigme expérimental donné, nous simulons M_2 jeux de données selon le même paradigme et sous le modèle $\mathcal{M}_{\mathcal{P},0}$. Nous calculons alors, pour ces M_2 jeux de données, les quantités $(\xi_{\mathcal{P}}^m)_{m=1..M_2}$. Le quantile q_P tel que le risque de première espèce soit de 5% est obtenu à partir de la fonction de répartition empirique de $\xi_{\mathcal{P}}$:

$$\begin{aligned} \text{Risque de première espèce} &= P_{\mathcal{H}_0}(\text{rejeter } \mathcal{H}_0) \\ &= P_{\mathcal{H}_0} \left(2 \left[L_1(y; \hat{\theta}_{\mathcal{P},1}) - L_0(y; \hat{\theta}_{\mathcal{P},0}) \right] \geq q_P \right) \\ 0.05 &= 1 - \hat{F}_{M_2, \xi_{\mathcal{P}}}(q_P). \end{aligned}$$

Exemple

Pour le paradigme du premier jeu de données réelles présenté dans la partie 4.8, nous générons 2000 jeux de données sous le modèle $\mathcal{M}_{\mathcal{P},0}$ et calculons $(\xi_{\mathcal{P}}^m)_{m=1..2000}$ pour chacun des jeux de données. Nous obtenons alors la fonction de répartition empirique suivant (figure 4.1) et $q_P = 4.74$.

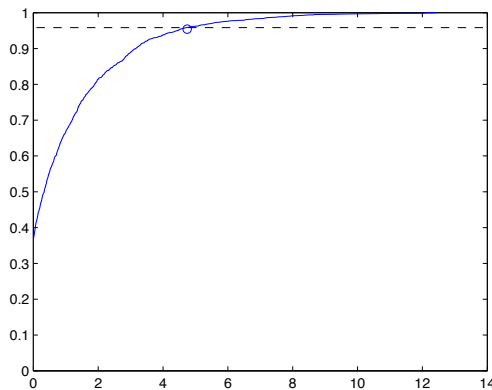


FIG. 4.1 – Fonction de répartition empirique du rapport de vraisemblance pour le paradigme du jeu de données réelles 1 (en trait plein) et quantile obtenu (cercle).

4.7.2 Sélection de la famille \mathcal{P}

Supposons d'abord que le nombre de types de stimuli, c'est-à-dire le cardinal des familles \mathcal{P} , est fixé égal à k . Pour toute famille \mathcal{P} incluant k types différents de stimuli, nous calculons les log-vraisemblances $lv_{\mathcal{P},0} = L_1(y; \hat{\theta}_{\mathcal{P},0})$ et $lv_{\mathcal{P},1} = L_1(y; \hat{\theta}_{\mathcal{P},1})$. Nous conservons les familles $\hat{\mathcal{P}}_{k,0}$ et $\hat{\mathcal{P}}_{k,1}$

maximisant ces deux quantités respectivement :

$$\widehat{\mathcal{P}}_{k,0} = \arg \max_{\mathcal{P} | \text{Card}(\mathcal{P})=k} lv_{\mathcal{P},0} \quad \text{et} \quad \widehat{\mathcal{P}}_{k,1} = \arg \max_{\mathcal{P} | \text{Card}(\mathcal{P})=k} lv_{\mathcal{P},1}$$

Notons

$$lv_0(k) = \max_{\mathcal{P} | \text{Card}(\mathcal{P})=k} lv_{\mathcal{P},0} \quad \text{et} \quad lv_1(k) = \max_{\mathcal{P} | \text{Card}(\mathcal{P})=k} lv_{\mathcal{P},1}$$

les log-vraisemblances de ces familles $\widehat{\mathcal{P}}_{k,0}$ et $\widehat{\mathcal{P}}_{k,1}$. Une fois que les meilleures familles (au sens du maximum de vraisemblance) $(\widehat{\mathcal{P}}_{k,0})_{k=1..K}$ et $(\widehat{\mathcal{P}}_{k,1})_{k=1..K}$ ont été identifiées pour tout cardinal $k = 1 \dots K$, nous traçons la suite de log-vraisemblances $(lv_0(k))_{k=1..K}$ et $(lv_1(k))_{k=1..K}$ en fonction du cardinal k . Nous avons alors recours à une méthode graphique pour sélectionner les cardinaux \widehat{k}_0 et \widehat{k}_1 , en retenant les valeur à partir desquelles les log-vraisemblances atteignent un plateau. En d'autres termes, sélectionner un plus grand nombre de types de stimuli n'engendrerait pas une augmentation significative de la log-vraisemblance.

4.8 Application sur données simulées et réelles

Dans cette partie,

- nous présentons d'une part une illustration des performances des algorithmes d'estimation EM et SAEM ainsi qu'une illustration de la procédure de sélection de modèles sur données simulées.
- D'autre part, nous présentons les résultats obtenus sur données réelles. Dans une première partie, nous cherchons à mettre en évidence la validité de l'hypothèse de variabilité inter-occurrences des amplitudes (problématique $\mathcal{P}b_1$) sur un large nombre de sujets et un large nombre de régions cérébrales d'intérêt. Dans une deuxième partie, nous nous intéressons à l'aspect "sélection de la famille des stimuli ayant un effet non négligeable sur le signal" (problématique $\mathcal{P}b_2$) sur un deuxième jeu de données.

4.8.1 Illustration des performances des algorithmes d'estimation sur données simulées

Le but de cette étude sur données simulées est d'étudier les performances des algorithmes présentés dans les parties 4.4, 4.5 et 4.7.

Illustration des performances de l'algorithme EM

Pour cela, nous générons 1500 jeux de données simulées sous le modèle $\mathcal{M}_{\mathcal{P},1}$ contenant 2 types de stimuli (comptant respectivement 26 et 25 occurrences) et en utilisant les paramètres suivants $(\mu_1, \mu_2) = (-3, 2)$, $(\omega_1, \omega_2) = (2, 1.5)$, soit :

$$\alpha_{1,j} \sim \mathcal{N}(-3, 2) \quad \alpha_{2,j} \sim \mathcal{N}(2, 1.5).$$

La variance de bruit est prise égale à $\sigma^2 = 3.10^{-2}$ (soit un rapport signal sur bruit de 2.5 environ). Nous traçons sur la figure 4.2 un exemple de jeu de données simulées. Pour chacun des jeux de données simulées, nous procédons à l'estimation des paramètres et réponse HRF par les algorithmes présentés dans les parties précédentes. Nous implémentons l'algorithme EM avec 300 itérations. Il met moins d'une minute à converger. Nous traçons sur la figure 4.3 un exemple de convergence des paramètres. La convergence est bien atteinte au terme des 300 itérations.

Plusieurs aspects nous intéressent :

- d'une part la qualité de l'estimation de la réponse HRF (en forme). En effet cette phase d'estimation précède l'estimation des autres paramètres et va logiquement influencer l'estimation des

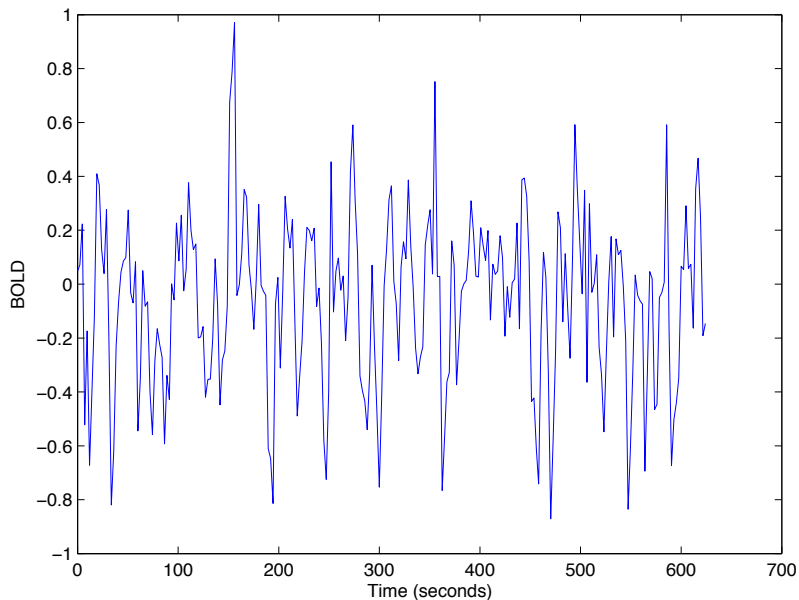


FIG. 4.2 – Exemple de jeu de données simulées dans le cas gaussien ($\alpha_{k,j} \sim \mathcal{N}(\mu_k, \omega_k)$)

paramètres de la loi des amplitudes de la HRF. Nous présentons sur la figure 4.4 deux HRF obtenues sur deux jeux de données différents. La HRF de gauche est très bien estimée, celle de droite l'est un peu moins bien. Cependant le pic de la HRF est bien estimé dans les deux cas.

- D'autre part, nous cherchons à savoir si les estimateurs des paramètres ($\sigma^2, \mu_1, \mu_2, \omega_1, \omega_2$) fournis par l'algorithme EM présenté dans la partie 4.4 sont précis. Pour cela, nous évaluons les biais relatifs moyens(%) et erreurs quadratiques relatives moyennes (RMSE) (%) pour les 1500 jeux de données (voir tableau 4.1) Les biais obtenus sont tous très faibles. Les RMSE des moyennes

Paramètres	Bias (%)	RMSE (%)
μ_1	0.11	11.28
μ_2	0.18	15.43
ω_1	4.92	22.38
ω_2	3.55	20.53
σ^2	-0.80	9.48

TAB. 4.1 – Données simulées dans le cas gaussien : biais relatif moyen(%) et erreur quadratique relative moyenne (RMSE) (%) des paramètres estimés par l'algorithme EM sur les 1500 jeux de données simulées.

μ sont tout à fait satisfaisantes. Celles des variances V sont un peu plus élevées mais demeurent acceptables.

- Enfin nous nous intéressons aussi à la qualité de reconstruction des amplitudes α . Pour cela, nous comparons les amplitudes simulées avec la quantité $E[\alpha | \mathbf{y}; \widehat{\theta}_{\mathcal{P},1}]$ obtenue à l'issue de l'algorithme EM. Nous représentons sur la figure 4.5 un exemple typique de reconstruction des amplitudes. Celle-ci nous semble satisfaisante au vu de la difficulté du problème puisque dans ce problème nous estimons à la fois la réponse HRF \mathbf{h} et les amplitudes α .

Cette étude des performances des méthodes d'estimation nous permet d'envisager une application

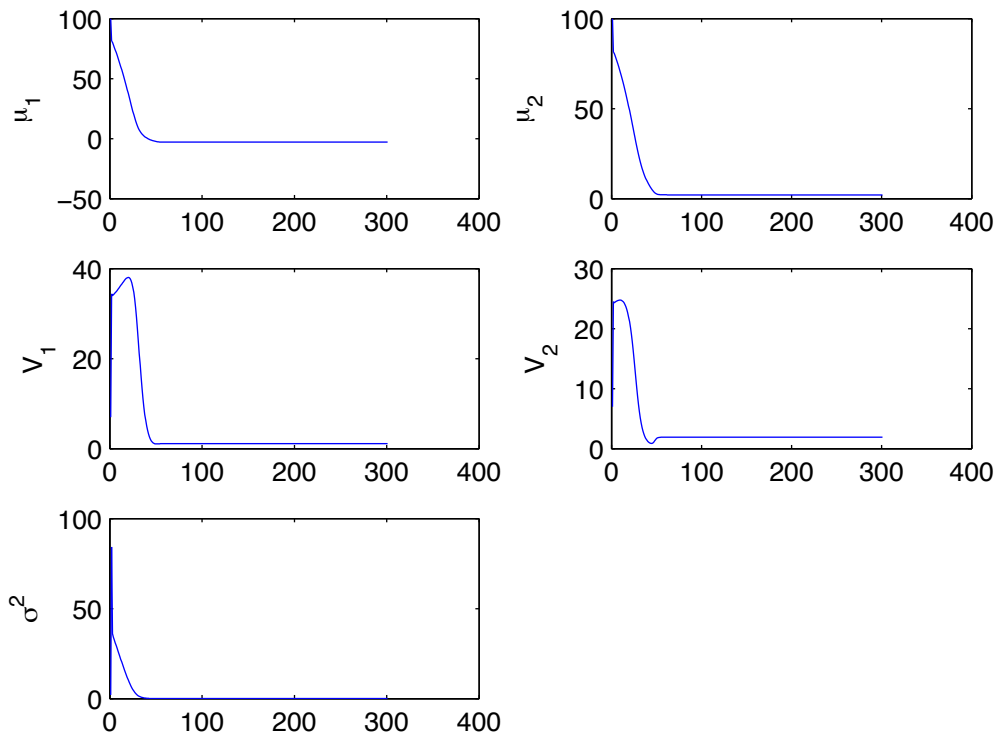


FIG. 4.3 – *Données simulées dans le cas gaussien* : évolution des estimateurs des paramètres obtenus par l’algorithme EM en fonction du nombre d’itérations

sur données réelles. Celle ci sera faite dans les parties 4.8.2 et 4.8.3

Illustration des performances de l’algorithme SAEM et de l’algorithme de sélection de modèle

Dans cette partie, nous cherchons à illustrer les performances de l’algorithme SAEM-MCMC présenté dans la partie 4.5 ainsi que la procédure de sélection de modèle. Pour cela, nous simulons un unique jeu de données simulées contenant 3 types de stimuli ($\mathcal{P} = \{1, 2, 3\}$) ayant respectivement 26, 25 et 8 occurrences chacun (le paradigme utilisé est issu d’une expérience réelle). Nous utilisons les paramètres de simulation suivants : $\sigma^2 = 10$ et

$$\begin{cases} \log(\alpha_{1,j}) \sim \mathcal{N}(\mu_1 = 5, \omega_1 = 0.3) \\ \log(\alpha_{2,j}) \sim \mathcal{N}(\mu_1 = 3, \omega_2 = 0.1) \\ \log(\alpha_{3,j}) \sim \mathcal{N}(\mu_1 = 3, \omega_3 = 0.2) \end{cases}$$

Nous estimons la réponse HRF par minimisation du critère des moindres carrés en imposant une contrainte de positivité sur les amplitudes. La réponse estimée est donnée sur la figure 4.6 (en haut à gauche). Sa qualité nous permet de l’utiliser pour estimer les paramètres de la loi des amplitudes.

Nous implémentons l’algorithme SAEM-MCMC avec 200 itérations. La suite de l’approximation stochastique γ_l vaut 1 pendant les 100 premières itérations puis $\gamma_l = \frac{1}{l-100}$ pour $l > 100$. La convergence des estimateurs est obtenue (voir figure 4.7). Sur cette figure, nous observons deux phases dans la convergence des paramètres. Dans un premier temps, ($l < 100$) les paramètres varient fortement afin de trouver le maximum de vraisemblance. La décroissance de γ_l à partir de $l = 100$ permet d’obtenir

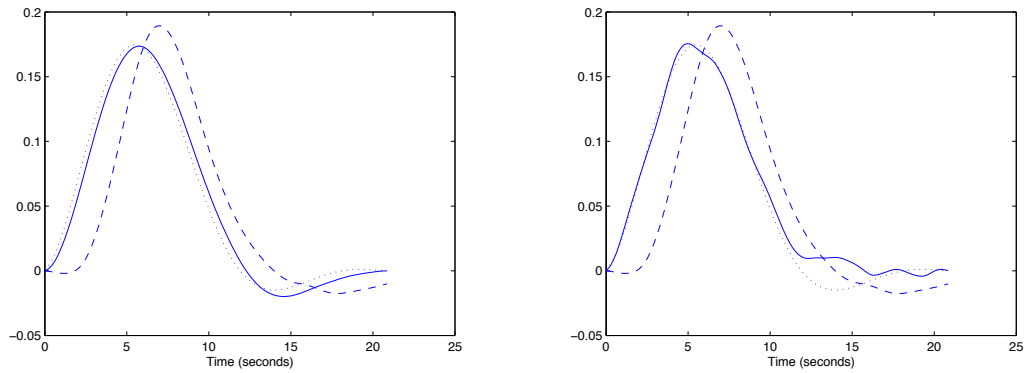


FIG. 4.4 – *Données simulées dans le cas gaussien* : estimation des HRF sur deux jeux de données simulées. HRF d’initialisation en tirets (—), HRF de simulation en pointillés (⋯), HRF estimée en trait plein. Estimation de très bonne qualité à gauche. Estimation moins bonne à droite

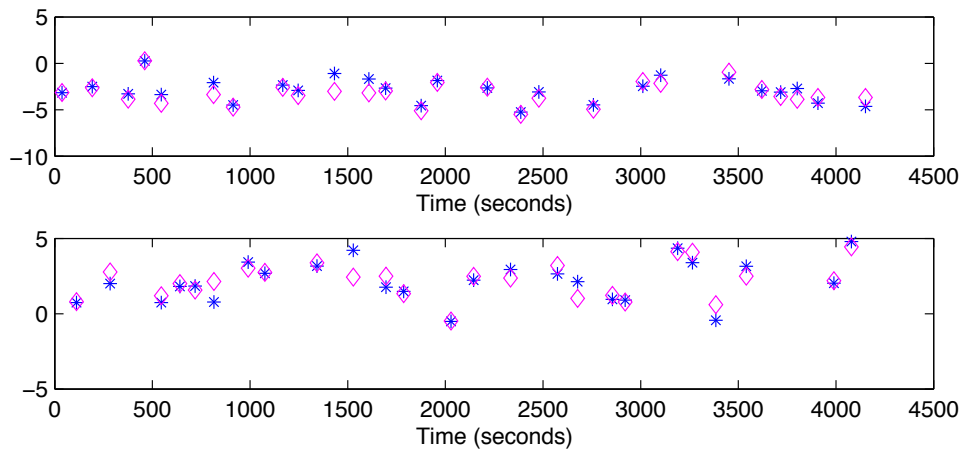


FIG. 4.5 – *Données simulées dans le cas gaussien* : Reconstruction des amplitudes α pour les stimuli de type 1 en haut (25 occurrences) et de type 2 en bas (26 occurrences). Amplitudes simulées (*) et amplitudes reconstruites (\diamond)

la convergence des paramètres. Les paramètres estimés sont reportés dans le tableau 4.2. Nous obtenons de bons estimateurs pour les paramètres des événements de type 1 et 2. Les paramètres des événements de types 3 sont estimés de façon moins précise, ce qui peut être dû à leur faible nombre d’occurrences.

Nous cherchons maintenant à tester notre procédure de sélection de modèle. Pour cela, nous conservons le jeu de données simulées précédemment et considérons un modèle plus général constitué de 6 types d’évènements ($\mathcal{P} = \{1, 2, 3, 4, 5, 6\}$). Nous appliquons la procédure décrite dans la partie 4.7 pour tenter de retrouver parmi ces six évènements les évènements de type $\{1, 2, 4\}$. Les familles successivement sélectionnées sous le modèle 1 sont les familles $\{1\}$, $\{1, 2\}$, $\{1, 2, 4\}$, $\{1, 2, 4, 5\}$, $\{1, 2, 3, 4, 6\}$, $\{1, 2, 3, 4, 5, 6\}$.

Pour chacune des familles testées, le modèle $\mathcal{M}_{\mathcal{P},1}$ est préféré au modèle $\mathcal{M}_{\mathcal{P},0}$.

Enfin, lorsque l’on trace les log-vraisemblances des familles de plus grandes vraisemblance ($(l_{v_1}(k))_{k=1\dots 6}$) (figure 4.6 en haut à droite et en bas), nous détectons très bien les évènements de type 1 et 2 : la

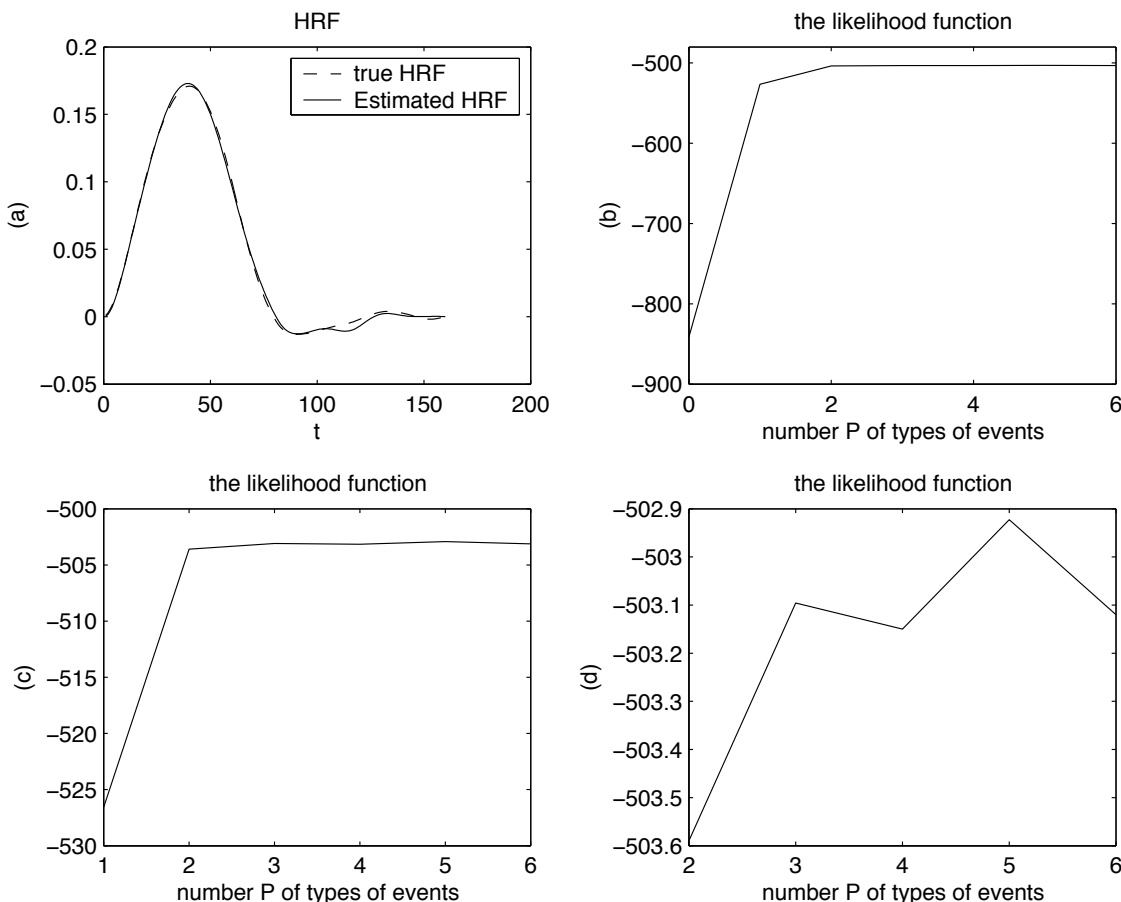


FIG. 4.6 – Données simulées dans le cas log-normal. En haut à gauche : estimation de la réponse HRF. En haut à droite : suite des $lv_1(P)$ pour $P = 0 \dots 6$. En bas à gauche : (zoom) suite des $lv_1(P)$ pour $P = 1 \dots 6$. En bas à droite : (zoom) suite des $lv_1(P)$ pour $P = 2 \dots 6$.

vraisemblance atteint un plateau net en $k = 2$. L'évènement de type 3 est plus difficilement détecté (pas de plateau de la vraisemblance évident). Ceci est sans doute du au fait que ce type de stimuli n'a que 8 occurrences (alors que les évènements de types 1 et 2 en comptent environ 25).

4.8.2 Application sur données réelles (1) : les amplitudes sont-elles variables au cours du temps ?

Acquisition des données

Nous avons utilisé un paradigme événementiel comprenant dix types d'évènements. Les sujets ont été soumis à des stimuli de types visuels (damiers horizontaux et verticaux), clics droits après instruction auditive, clics droits après instruction vidéo, clics gauches après instruction auditive, clics gauches après instruction auditive, calculs mentaux (soustraction) après instruction auditive ou visuelle, écoutes de phrases.

Les évènements arrivent aléatoirement dans le temps (temps moyen inter-stimuli : 3s), à raison d'environ dix occurrences par type de stimuli (excepté pour les clics moteurs qui ne comptent que cinq occurrences par session).

Huit sujets droitiers ont participé à l'étude. Les sujets ont donné leur consentement éclairé et le

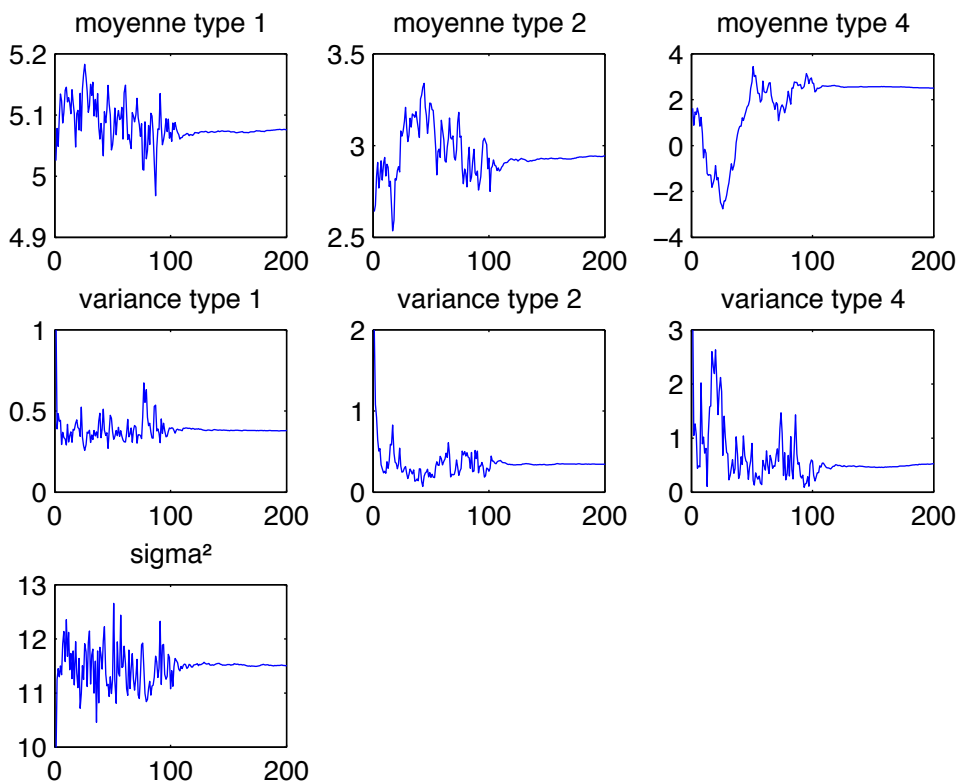


FIG. 4.7 – Données simulées dans le cas log-normal : estimateurs des paramètres fournis par l’algorithme SAEM en fonction du nombre d’itérations

	μ_1	μ_2	μ_3	ω_1	ω_2	ω_3	σ^2
Simulation	5	3	3	0.3	0.2	0.2	10
Estimation	5.07	2.94	2.50	0.38	0.34	0.53	10.5

TAB. 4.2 – Données simulées dans le cas log-normal : Estimation des paramètres avec l’algorithme SAEM-MCMC

protocole a été approuvé par le comité d’éthique local.

Les images fonctionnelles ont été acquises sur un scanner Bruker 3 Tesla (TR = 2400ms, TE = 60ms, taille de matrice=64 × 64, FOV = 24cm × 24cm). Chaque session comprenait 130 scans. Les images anatomiques ont été acquises sur le même scanner avec une résolution spatiale de 1 × 1 × 1.2 mm³.

Les données IRMf ont subi divers traitements permettant de corriger les mouvements des sujets. Pour tous les sujets, les mouvements estimés étaient inférieurs à 1 mm ou 1 degré. Enfin les données ont été normalisée spatialement de façon à pouvoir comparer les sujets entre eux. La dérive basse-fréquence a été préalablement ôtée afin d’éliminer les éventuels phénomènes de grande période polluant les données.

Sélection des régions d’intérêt (ROI)

Pour chaque sujet, quatre régions d’intérêt ont été définies de forme sphériques de 10mm de diamètres (figure 4.8). Une étude de contraste a permis de montrer une forte activité hémodynamique dans le

gyrus temporal gauche et une moins marquée dans le gyrus temporal droit. Nous avons ainsi pu tester notre modèle à amplitudes variables dans des régions montrant différents niveaux d'activité. Les régions d'intérêt étaient localisées aux coordonnées $(x=59 y=28 z=2)$, $(x=-59 y=28 z=2)$ dans le gyrus temporal supérieur et plus en arrière autour du sillon temporal supérieur.

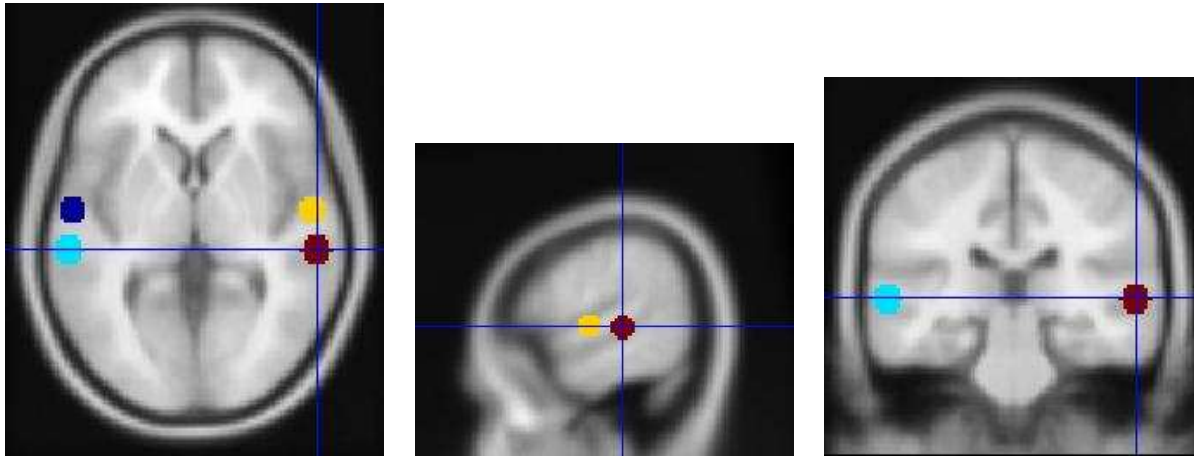


FIG. 4.8 – *Jeu de données réelles (1)* : La localisation des ROI est illustrée sur le “MNI template”. Coupes axiale (à gauche), sagittale (au centre) et coronale (à droite)

Résultats

Les algorithmes d'estimation décrits dans les paragraphes 4.3 et 4.4 ont été utilisés pour estimer la réponse HRF et les amplitudes. Tous les types de stimuli auditifs (exceptés les clics sur ordre auditif) ont été regroupés pour ne former qu'un seul type de stimuli, de même pour les stimuli visuels.

Remarque 15. *Dans cette étude, nous n'avons pas cherché à sélectionner les types de stimuli ayant une réelle influence sur le signal (problématique $\mathcal{P}b_2$) mais à mettre en évidence sur données réelles la validité de l'hypothèse de variabilité inter-occurrences des amplitudes (problématique $\mathcal{P}b_1$).*

Nous présentons tout d'abord les fonctions HRF estimées dans les régions temporales gauches (ROI 1, $x=59 y=28 z=2$) pour les huit sujets, ordonnés de 1 à 8 de gauche à droite et de haut en bas (figure 4.9). Dans la plupart des cas, la réponse estimée est proche de la réponse utilisée dans les analyses standard (voir figure 1.3). Toutes les HRF estimées atteignent un pic environ 5 secondes après le début de la réponse ; toutes les réponses, exceptées celle des sujets 2 et 6, présentent des undershoots finaux. De façon assez intéressante, les sujets 5 et 7 présentent de légers undershoots initiaux. Deux réponses ont des formes différentes de la réponse standard (sujet 3 et 6) mais cette différence n'est pas suffisante que cette forme soit biologiquement invraisemblables. Nous considérons que l'estimation des fonctions HRF est satisfaisante. La qualité des HRF estimées étaient très similaires dans les trois autres ROI.

Puisque le modèle $\mathcal{M}_{\mathcal{P},1}$ utilise plus de paramètres que le modèle $\mathcal{M}_{\mathcal{P},0}$, nous avons logiquement constaté que les prédictions sous le modèle $\mathcal{M}_{\mathcal{P},1}$ étaient toujours meilleures que celles obtenues sous le modèle 0. Sur la figure 4.10, nous représentons les amplitudes des réponses à tous les stimuli auditifs (ainsi que leur écart type estimé) pour la ROI 4 et les sujets 4 et 5. Nous constatons clairement des amplitudes d'intensités variables au cours des occurrences des stimuli auditifs.

Dans le tableau 4.3, nous avons consigné les résultats du test \mathcal{H}_0 contre \mathcal{H}_1 pour les 8 sujets et les 4 ROI (soient 32 régions au total). Pour 22 des 32 régions, le test sélectionne le modèle $\mathcal{M}_{\mathcal{P},1}$ au modèle $\mathcal{M}_{\mathcal{P},0}$. Ceci ne semble pas être propre à un sujet ou une région. En effet, les huit sujets présentent des régions pour lesquelles le modèle $\mathcal{M}_{\mathcal{P},0}$ est préféré. De la même façon, les quatre ROI présentent des

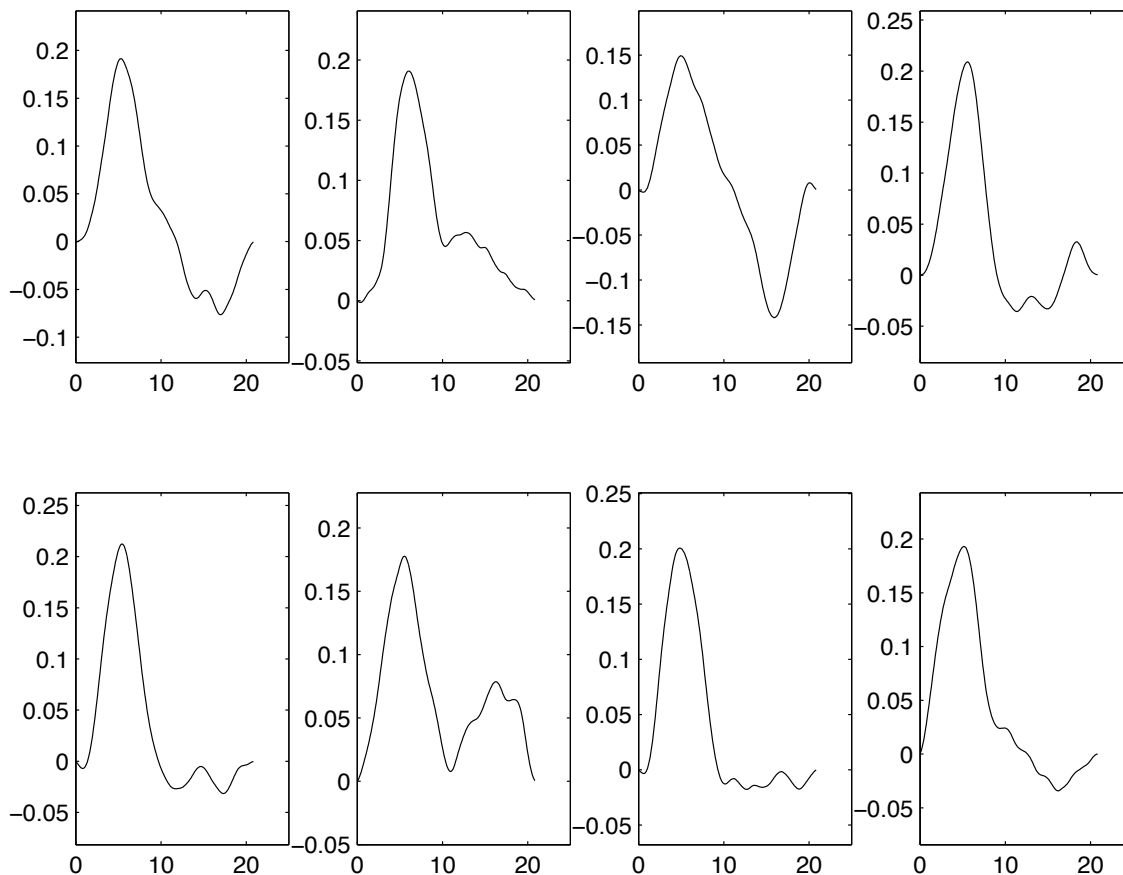


FIG. 4.9 – *Jeu de données réelles (1)* : Fonctions HRF estimées pour les sujets 1 (en haut à gauche) à 8 (en bas à droite) dans la ROI 1

sujets pour lesquels le modèle $\mathcal{M}_{\mathcal{P},0}$ est retenu. Étant donné que le risque d’erreur était fixé à 5%, il est vraisemblable que le modèle $\mathcal{M}_{\mathcal{P},1}$ soit plus approprié à ce jeu de données.

En outre, en marge de cette étude présentée ici, nous avons appliqué le même test dans deux autres ROI définies dans le cortex occipital. Nous y avons trouvé un plus petit nombre de tests significatifs (3 sur 16) mais ceci prouve tout de même que des amplitudes variables peuvent être trouvées dans d’autres cortex primaires.

La conclusion de cette étude et ses implications pour les analyses de données IRMf futures sont discutées dans la partie 4.9.

4.8.3 Application sur données réelles (2) : sélection des types de stimuli

Dans cette partie, nous considérons un autre jeu de données et cherchons à sélectionner les types de stimuli ayant un effet non négligeable sur le signal (problématique $\mathcal{P}b_2$).

Les données

Nous considérons des données acquises au cours d’un paradigme expérimental comprenant huit types de stimuli. Les événements de type 1 et 2 sont des clics droit et gauche respectivement. Les événements de types 3 à 8 sont des stimuli visuels. Deux cent soixante quatre scans ont été acquis avec un scanner Brucker 3 Tesla, toutes les 2.4 secondes. Les données présentées ici ont été extraites du cortex moteur

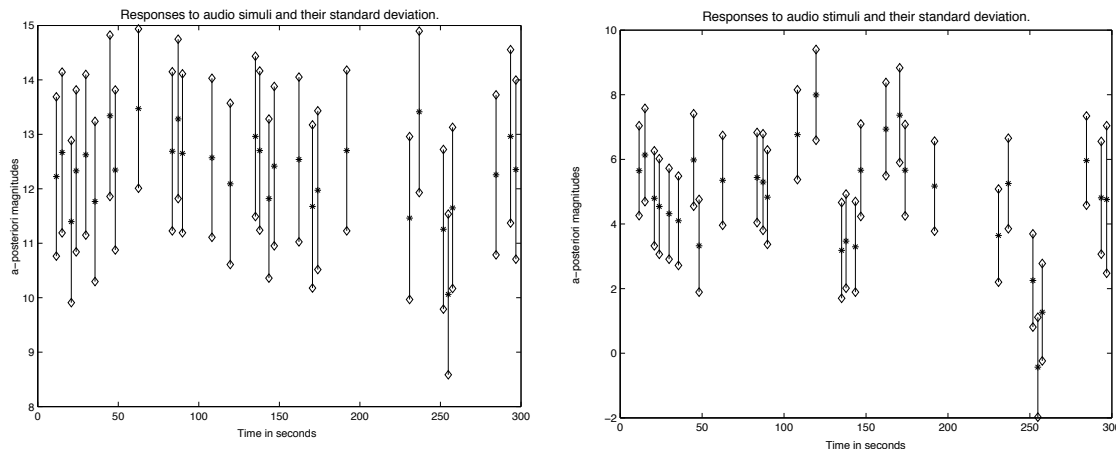


FIG. 4.10 – *Jeu de données réelles (1)* : Amplitudes et écart-types des stimuli auditifs, pour la ROI 4 des sujets 4 (à gauche) et 5 (à droite).

	ROI1	ROI2	ROI3	ROI4
sujet 1	1	1	0	0
sujet 2	1	0	1	0
sujet 3	1	0	1	1
sujet 4	0	1	0	1
sujet 5	1	1	0	1
sujet 6	1	0	1	1
sujet 7	0	0	0	1
sujet 8	0	1	0	1

TAB. 4.3 – *Jeu de données réelles (1)* : résultats du test \mathcal{M}_0 versus \mathcal{M}_1 pour tous les sujets et toutes les régions d'intérêt.

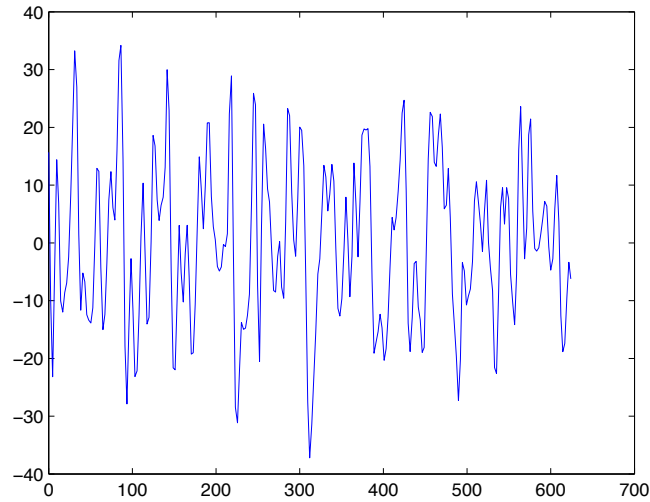
gauche et filtrées de la même façon que précédemment afin d'enlever les éventuelles dérives basses fréquences. Elles sont représentées sur la Figure 4.11.

Estimation et sélection de modèles

La fonction HRF estimée est tracée sur la figure 4.12. Sa forme (dotée d'un undershoot final) diffère assez peu des formes standards et est donc biologiquement plausible.

Nous avons ensuite appliqué la procédure de sélection de modèles, de façon à déterminer quels types stimuli ont une réelle influence sur le signal. Pour cela, comme décrit dans la partie 4.7.2, nous avons tout d'abord fixé le cardinal de la famille \mathcal{P} et sélectionné parmi ces familles celles de plus grande vraisemblance sous les modèles \mathcal{M}_0 et \mathcal{M}_1 . Puis nous avons fait varier le cardinal de cette famille entre 0 et 8. Les familles alors obtenues sont consignées dans le tableau 4.4. Les log-vraisemblances correspondantes à ces familles sont tracées sur la figure 4.13.

Les familles successivement sélectionnées sous le modèle \mathcal{M}_1 sont les familles $\{1\}$, $\{1, 6\}$, $\{1, 6, 7\}$, $\{1, 6, 5, 3\}$, $\{1, 6, 5, 3, 4\}$, $\{1, 6, 5, 3, 4, 7\}$, $\{1, 6, 5, 3, 4, 7, 8\}$ et $\{1, 6, 5, 3, 4, 7, 8, 2\}$. L'évènement 1 est clairement prépondérant. Ceci est dû au fait que les évènements de type 1 sont des clics *droits* et que ce signal est issu d'une région d'intérêt située dans le cortex moteur *gauche*. Nous nous attendions donc à ce résultat. Les familles suivantes sont emboîtées sauf la troisième. Par méthode graphique nous sélectionnons le cardinal k à partir duquel la vraisemblance cesse de croître significativement,


 FIG. 4.11 – *Jeu de données réelles (2)* : données extraites du cortex moteur gauche

k	$\widehat{\mathcal{P}}_{k,0}$	$\widehat{\mathcal{P}}_{k,1}$
1	1	1
2	1,6	1,6
3	1,6,7	1,6,4
4	1,6,5,3	1,4,2,5
5	1,6,5,3,4	1,6,4,2,5
6	1,6,5,3,4,7	1,6,4,2,5,3
7	1,6,5,3,4,7,8	1,6,4,2,5,3,7
8	1,6,5,3,4,7,8,2	1,6,4,2,5,3,7,8

 TAB. 4.4 – *Jeu de données réelles (2)* : sélection de la famille \mathcal{P} sous le modèle \mathcal{M}_0 à gauche et \mathcal{M}_1 à droite en fonction de k son cardinal.

soit $k = 2$ sous le modèle \mathcal{M}_0 et $k = 5$ pour le modèle \mathcal{M}_1 .

4.9 Discussion et conclusion

Dans ce travail nous avons présenté une méthode d'estimation en deux étapes, d'une part l'estimation de la forme de la HRF, d'autre part l'estimation des amplitudes des événements dans un modèle comprenant des amplitudes variables au cours des occurrences. Nous avons en outre proposé une méthode de sélection de modèles permettant de savoir quels types d'événements ont une réelle influence sur le signal. Cette étude est d'un intérêt certain puisque les connexions entre les aires cérébrales sont très complexes et demeurent mal connues. Cependant, cette approche doit être développée plus largement, sur un plus grand nombre de données, la méthode de sélection (reposant sur un critère graphique) doit quant à elle être perfectionnée.

Nous avons enfin montré que le test de \mathcal{H}_0 contre \mathcal{H}_1 rejette souvent le modèle ayant des amplitudes constantes au cours des occurrences d'un même type d'événements. La mise en évidence de cette nouvelle variabilité par notre méthode n'est pas sans conséquences sur l'analyse future de données IRMf.

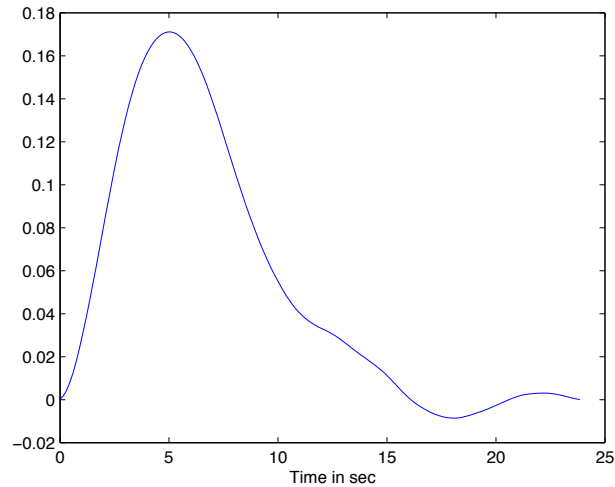


FIG. 4.12 – *Jeu de données réelles (2)* : fonction HRF estimée

Conséquence pour l'analyse de données IRMf

Ce dernier résultat a d'importantes implications pour l'analyse de données IRMf. En effet, pour comparer les réponses à différents types de stimuli, il ne serait pas suffisant de comparer l'amplitude moyenne de la fonction HRF mais il est nécessaire de prendre en compte la variance des amplitudes des fonctions. Cette étude révèle donc un terme de variance jusque là ignoré et qui devrait être pris en compte dans les futures analyses.

Origine de la variation

Cette variation au cours du temps peut avoir de nombreuses origines. La plus évidente semble être une variation des stimuli eux-mêmes (par exemple, les différentes phrases entendues peuvent engendrer des activités neuronales variables et des besoins métaboliques différents dans les cortex primaires auditifs). Cependant d'autres origines sont envisageables. L'attention du sujet peut varier au cours du temps et conduire à des variations de l'amplitude ou des retardements de la réponse. Dans la mesure où nous estimons une unique forme de HRF pour tous les événements, un décalage dans le temps de la réponse entraîne alors une variabilité de son amplitude. Enfin, cette variabilité peut être due à un phénomène de ralentissement et d'adaptation non modélisé qui régulerait la réponse vasculaire de façon inconnue mais aboutirait aux variations observées (voir [55, 79] pour plus de détails). Les contributions relatives de ces différents facteurs sur la variabilité totale ou l'implication d'autres facteurs doivent encore être explorées.

Travaux futurs

L'étude a démontré que des événements répétés pouvaient avoir une variabilité significative dans le contexte de tâches de compréhension de phrases et dans les gyri temporaux. Il reste cependant à montrer que ce type de résultats peut être étendu à d'autres paradigmes et à d'autres régions cérébrales. Il se pourrait tout à fait que des stimuli sensoriels, mieux contrôlés (stimulation visuelle par exemple) ne présentent pas de tels comportements. En effet, comme souligné précédemment, il est probable que les processus mentaux dépendant de l'attention ou de l'état du sujet soient d'amplitude variable alors que les processus sensoriels tels que les stimuli visuels aient des réponses constantes.

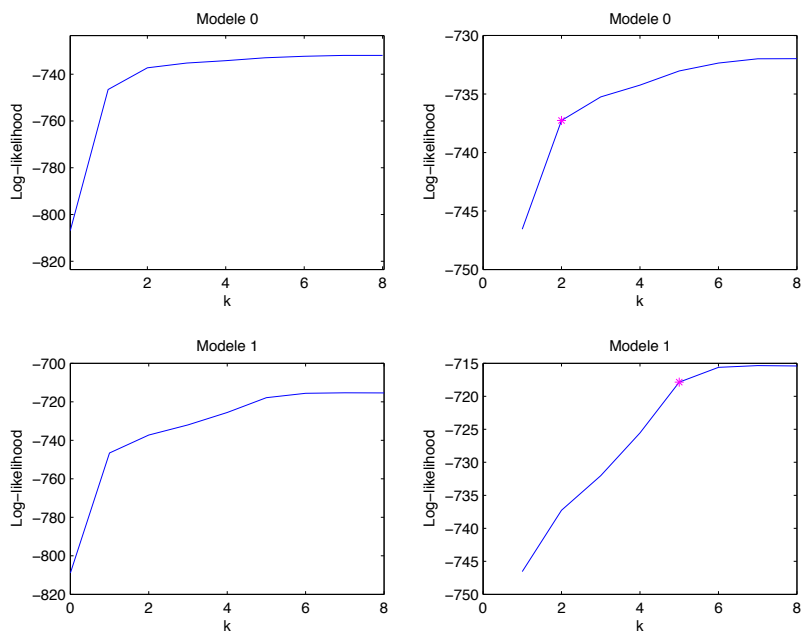


FIG. 4.13 – *Jeu de données réelles (2)* : En haut à gauche : suite des $lv_0(k)$ pour $k = 0 \dots 8$. En haut à droite (zoom) : suite des $lv_0(k)$ pour $k = 1 \dots 8$. En bas à gauche : suite des $lv_1(k)$ pour $k = 0 \dots 8$. En bas à droite (zoom) : suite des $lv_1(k)$ pour $k = 1 \dots 8$.

Récemment, des mesures simultanées EEG-IRMf ont été réalisées. Cette technique est potentiellement intéressante puisqu'elle devrait permettre de mesurer les paramètres reflétant l'activité liée aux courants électriques post-synaptiques. Nous envisageons donc de reproduire les mêmes expériences mais en utilisant conjointement l'EEG de façon à savoir si les variations du signal BOLD sont liées à l'énergie de certaines bandes de fréquence des mesures électriques.

Bien que les modèles linéaires soient attractifs pour des raisons de facilité de calcul et par conséquent les plus utilisés, il est évident qu'ils ne sont pas satisfaisants d'un point de vue biologique. Des modèles plus vraisemblables d'un point de vue biologique ont été introduits récemment ([105, 50, 7]). Les travaux futurs devraient donc étudier l'effet de ces modèles. Enfin, la variation du retardement des phénomènes hémodynamiques pourraient aussi être incluses dans ces modèles et testée dans le futur.

Conclusion

Nous avons montré que les amplitudes des événements doivent être modélisées non seulement avec une variance inter mais aussi intra-événements. Alors que les origines de ces variations doivent encore être explorées, ce résultat devrait modifier de façon significative les analyses statistiques à venir en imposant de prendre en compte une variabilité supplémentaire.

Troisième partie

Estimation dans des modèles définis par équations différentielles ordinaires

Nous avons vu qu'en IRMf, les modèles linéaires peuvent être insatisfaisants d'un point de vue biologique. Il est alors nécessaire d'introduire des modèles physiologiques tels que le "*Balloon model*" proposé par Buxton [23] et complété par Friston [50]. Les modèles physiologiques aboutissent à l'écriture du signal BOLD comme solution d'un système différentiel sans solution analytique, observé avec bruit de mesure.

Dans cette partie, nous nous intéressons tout d'abord au problème général de l'estimation paramétrique dans le cadre de modèles statistiques à données non-observées, dont la fonction de régression est définie par un système d'équations différentielles ordinaires (chapitre 5)⁴. Les méthodes proposées sont appliquées sur un modèle non-linéaire à effets mixtes issu de la pharmacocinétique.

Au chapitre 6, nous appliquons la méthode d'estimation par maximum de vraisemblance précédemment proposée aux modèles physiologiques développés en IRMf à la fois sur données simulées et sur données réelles.

⁴Ce chapitre ayant fait l'objet d'un article, celui-ci est écrit en anglais.

Chapitre 5

Estimation of parameters in incomplete data models defined by dynamical systems

Travail réalisé en collaboration avec Adeline Samson ¹, article accepté par *Journal of Statistical Planning and Inference*

ABSTRACT:

Parametric incomplete data models defined by ordinary differential equations (ODEs) are widely used in biostatistics to describe biological processes accurately. Their parameters are estimated on approximate models, whose regression functions are evaluated by a numerical integration method. Accurate and efficient estimations of these parameters are critical issues. This paper proposes parameter estimation methods involving either a stochastic approximation EM algorithm (SAEM) in the maximum likelihood estimation, or a Gibbs sampler in the Bayesian approach. Both algorithms involve the simulation of non-observed data with conditional distributions using Hastings-Metropolis (H-M) algorithms. A modified H-M algorithm, including an original Local Linearization scheme to solve the ODEs, is proposed to reduce the computational time significantly. The convergence on the approximate model of all these algorithms is proved. The errors induced by the numerical solving method on the conditional distribution, the likelihood and the posterior distribution are bounded. The Bayesian and maximum likelihood estimation methods are illustrated on a simulated pharmacokinetic nonlinear mixed-effects model defined by an ODE. Simulation results illustrate the ability of these algorithms to provide accurate estimates.

KEY-WORDS:

Bayesian estimation - Incomplete data model - Local linearization scheme - MCMC algorithm - Nonlinear mixed-effects model - ODE integration - SAEM algorithm

¹INSERM U738, Paris, France ; University Paris 7, Paris, France

5.1 Introduction

When a biological or physiological process is measured, the regression function of the statistical model corresponding to the observed data is often derived from a differential equation describing the underlying dynamic process. Difficulties arise when the differential equation has no analytical solution and/or when the parameters of the regression function are random and thus non-observed. Such example can be found in pharmacokinetics, which aims to study drug evolutions in human organism, this evolution being described by differential systems of compartment interactions. Mixed models, for which regression parameters are considered as random variable and non-observed data, are widely used for the analysis of pharmacokinetic datasets, which have classically repeated measurements in several patients.

This paper aims at providing a general answer to the estimation problem in such statistical incomplete data models.

Let y be the noised observations of a biological process measured at instants (t_1, \dots, t_J) . The biological process is described by the solution g of an ordinary differential equation (ODE), depending on a stochastic non-observed parameter ϕ :

$$y_j = g(t_j, \phi) + \varepsilon_j \quad \text{for } j = 1 \dots J.$$

We consider that the observable vector Y is part of a so-called complete vector (Y, ϕ) . We assume that both Y and (Y, ϕ) have density functions, $p_Y(y; \theta)$ and $p_{Y, \phi}(y, \phi; \theta)$ respectively, depending on a parameter θ belonging to some subset Θ of the Euclidean space \mathbb{R}^q . The estimation of the parameter θ has been widely studied when the regression function g has an explicit form. Two approaches can be followed to tackle this challenge, respectively the maximum likelihood and the Bayesian estimations.

Generally, the maximization of the likelihood of the observations cannot be done in a closed form. Dempster *et al.* [36] propose the iterative Expectation-Maximization (EM) algorithm for incomplete data problems. At the k^{th} iteration, the E-step of EM algorithm computes $Q(\theta|\theta_k) = E(\log p_Y(y; \theta)|y; \theta_k)$ while the M-step determines θ_{k+1} maximizing $Q(\theta|\theta_k)$. For cases where the E-step has no closed form, stochastic versions of EM are introduced. Celeux *et al.* [29] introduce the Stochastic EM algorithm (SEM). Wei & Tanner [127] suggest the Monte-Carlo EM (MCEM) estimating $Q(\theta|\theta_k)$ by the averaging of m Monte-Carlo replications. Recently, Wu [134] emphasizes that MCEM is computationally intensive. As an alternative, Delyon *et al.* [35] propose the Stochastic Approximation EM algorithm (SAEM) replacing the E-step by a stochastic approximation of $Q(\theta|\theta_k)$. These methods require the simulation of the non-observed data ϕ . For cases where this simulation can not be performed in a closed form, Kuhn & Lavielle [72] suggest to resort to iterative methods such as Monte Carlo Markov Chain algorithms (MCMC).

The Bayesian approach estimates the posterior distribution $p_{\theta|Y}(\cdot|y)$ of θ , a prior $p_{\theta}(\cdot)$ being given. Because of the conditional independence structure of $p_{\theta|Y} = \int p_{\theta|Y, \phi} p_{\phi|Y} d\phi$ and $p_{\phi|Y} = \int p_{\phi|Y, \theta} p_{\theta|Y} d\theta$, Gelfand & Smith [51] propose a Gibbs sampling to evaluate these two integrals simultaneously. At iteration k , ϕ_k , a realization of ϕ , is simulated with $p_{\phi|Y}(\cdot, \theta_{k-1})$ followed by θ_k , a realization of θ with $p_{\theta|Y, \phi}(\cdot, \phi_k)$. Consequently, as in maximum likelihood estimation, difficulties arise when the simulation of the conditional distribution can not be performed in a closed form. For these cases, a Hastings-Metropolis (H-M) algorithm can be included in the Gibbs sampler.

The use of the H-M algorithm in estimation algorithms requires the evaluation of the regression function g at each iteration. When g is a non-analytical solution of a dynamical system, it is evaluated using a numerical integration method. Thus a trade-off between accuracy, stability and computational cost is required. In this paper, we detail the Local Linearization scheme [see e.g. 18, 104, 68] not only because of its stability performances but also because this scheme can be extended to a so-called modified Local Linearization scheme, adapted to its inclusion in the H-M algorithm. The estimation algorithms are then applied to an approximate model whose regression function is an approximate solution of the ODE.

The objective of this research is to quantify the error induced by the numerical approximation of the regression function g . The paper is organized as follows. Section 5.2 defines the original statistical model; the Local Linearization scheme and its modified version are detailed; the approximate statistical model resulting from the numerical approximation is introduced. Section 5.3 focuses on the H-M algorithm to simulate the non-observed data ϕ with the conditional distribution. The error induced by the numerical approximation of g on the conditional distribution is quantified. Section 5.4 is dedicated to the parameter estimation algorithms. Concerning maximum likelihood and Bayesian estimations, the standard algorithms are adapted to solve the approximate model. The error induced by the use of the numerical solving method is bounded respectively on the likelihood and the posterior distribution. This error is distinct from the error on the estimates induced by the estimation algorithm which is evaluated by their standard errors. Finally, the SAEM algorithm and the Bayesian Gibbs sampler are applied on a nonlinear mixed-effects model deriving from pharmacokinetics in Section .

5.2 Models and notations

5.2.1 An incomplete data model defined by ODEs

Let $y = (y_j)_{j=1..J}$ denote the observations measured at times (t_1, \dots, t_J) . We consider the incomplete data model, called model \mathcal{M} , defined as follows:

$$\begin{aligned} y_j &= g(t_j, \phi) + \varepsilon_j & 1 \leq j \leq J \\ \varepsilon_j &\sim \mathcal{N}(0, \sigma^2) \\ \phi &\sim \pi(\cdot; \beta) \end{aligned} \tag{\mathcal{M}}$$

where $g(\cdot)$ is a nonlinear function of ϕ , ε_j represents the error of the measurement j , σ^2 is the residual variance, ϕ is a non-observed random parameter distributed with the density $\pi(\cdot, \beta)$, depending only on the parameter β . The parameter $\theta = (\beta, \sigma^2)$ belongs to some open subset $\Theta \subset \mathbb{R}^q$.

Let g be written $g = H \circ f$, where $H : \mathbb{R}^d \rightarrow \mathbb{R}$ is a known function and, $f : \mathbb{R} \times \mathbb{R}^k \rightarrow \mathbb{R}^d$ is defined as the solution of the following ODE:

$$\begin{aligned} \frac{\partial f(t, \phi)}{\partial t} &= F(f(t, \phi), t, \phi) \\ f(t_0, \phi) &= f_0(\phi) \end{aligned} \tag{5.2.1}$$

with a known function $F : \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^k \rightarrow \mathbb{R}^d$ and the initial condition $f_0(\phi) \in \mathbb{R}^d$, $t \in [t_0, T]$.

We make the following additional assumptions:

- Assumption **H1**: π has a compact support $K_1 \subset \mathbb{R}^k$, and there exist two constants a and, b such that

$$0 < a < \pi(\phi; \beta) < b \text{ for all } \phi \in K_1.$$

- Assumption **H2**: $F : \mathbb{R}^d \times \mathbb{R} \times \mathbb{R}^k \rightarrow \mathbb{R}^d$ is \mathcal{C}^2 on its definition domain, and $\phi \rightarrow f_0(\phi) \in \mathbb{R}^d$ is \mathcal{C}^1 on K_1 .
- Assumption **H3**: H is an L_H -lipschitzian function.

5.2.2 Approximation of the regression function

A great variety of numerical schemes have been proposed to solve ODEs [see e.g. 59]. The accuracy of such numerical methods is qualified by the order and the step size of these schemes. The numerical scheme is applied on sub-intervals $[t_n, t_{n+1}[$, $n = 0, \dots, N-1$, of the time interval $[t_0, T]$, with $t_N = T$. The maximal length or step size of the sub-intervals is denoted h . The order of a numerical scheme is defined as follows:

Definition 1. Let f_h be the resulting approximate function obtained by a numerical integration scheme of step size h . This scheme is of order p if there exists a constant C such that

$$\sup_{t \in [t_0, T]} |f(t, \phi) - f_h(t, \phi)| \leq Ch^p.$$

Of all the numerical schemes, the Local Linearization (LL) scheme provides a good trade-off between computational cost and numerical stability, as exposed in [18]. It derives from the local linearization of the right term of the ODE (5.2.1) with respect to time t , and the exact integration of the deduced linear differential equation. Its implementation requires matrix exponential computations using new algorithms such as Pade or Schur methods, which have proved their efficiency and stability. The LL scheme has the additional advantage of preserving stability properties on stiff systems [see e.g. 104, 69]. In cases where the ODE depends on parameter ϕ , we extend this scheme using a Taylor expansion with respect to time t and parameter ϕ . More precisely, let the solution at ϕ_0 be computed using the LL scheme. Let ϕ be in a neighborhood of ϕ_0 . On each sub-interval $[t_n, t_{n+1}[$, $n = 0, \dots, N - 1$, the solution f_{h, ϕ_0} of the following linear equation

$$\begin{aligned} \frac{\partial f(t, \phi)}{\partial t} &= F(f(t_n, \phi_0), t_n, \phi_0) + \frac{dF}{df}(f(t_n, \phi_0), t_n, \phi_0) (f(t, \phi) - f(t_n, \phi_0)) \\ &+ \frac{dF}{dt}(f(t_n, \phi_0), t_n, \phi_0) (t - t_n) + \frac{dF}{d\phi}(f(t_n, \phi_0), t_n, \phi_0) (\phi - \phi_0). \end{aligned}$$

is evaluated. Details of this scheme (called LL2 scheme) are given in appendix 5.8. This LL2 scheme does not involve any additional computation of matrix exponentials: the required matrix exponential has already been computed with the LL scheme at ϕ_0 . This reduces the computational time, which is a key issue in iterative processes. For instance, during H-M algorithm implementation, the ODE has to be integrated at ϕ contained in a neighborhood of ϕ_0 , for which the ODE has already been solved by LL. This leads us to propose a modified version of the H-M algorithm, taking advantage of these schemes (see Section 5.3).

Convergence properties of the two previous numerical schemes are given in the following lemma. Let us consider the following assumption:

- Assumption **H1'**: ϕ remains in the compact set $K_1 \in \mathbb{R}^k$.

Lemma 1. Let ϕ_0 and ϕ be in K_1 . Let $f(\cdot, \phi)$ be the exact solution of the ODE (5.2.1), $f_h(\cdot, \phi)$ the one obtained by the LL scheme with step size h , and $f_{h, \phi_0}(\cdot, \phi)$ the LL2 solution. Assume that **H1'** and **H2** hold. Then:

1. there exists a constant C independent of ϕ , such that, for any $t \in [t_0, T]$ and for any ϕ ,

$$|f(t, \phi) - f_h(t, \phi)| \leq Ch^2,$$

2. there exist constants C_1 and C_2 such that, for any $t \in [t_0, T]$ and for any ϕ ,

$$|f(t, \phi) - f_{h, \phi_0}(t, \phi)| \leq \max(C_1 h^2, C_2 \|\phi - \phi_0\|_{\mathbb{R}^k}^2).$$

Part 1 is proved in [104], part 2 is proved in appendix 5.8.

5.2.3 An approximate incomplete data model

In practice, estimation algorithms require the numerical approximation of the regression function and are thus applied to an approximate version of the model \mathcal{M} . Let f_h be the approximate solution

of the ODE (5.2.1), obtained by a numerical integration method of step size h and order p . Let the approximate statistical model \mathcal{M}_h be defined by:

$$\begin{aligned} y_j &= g_h(t_j, \phi) + \varepsilon_j & 1 \leq j \leq J \\ \varepsilon_j &\sim \mathcal{N}(0, \sigma^2) \\ \phi &\sim \pi(\cdot; \beta) \end{aligned} \tag{\mathcal{M}_h}$$

where $g_h = H \circ f_h$. Subsequently, the different distributions of the model \mathcal{M}_h are subscripted with h .

5.3 Simulation of non-observed data with the conditional distribution

In this paper, the Hastings-Metropolis (H-M) algorithm is combined successively with the SAEM algorithm in the maximum likelihood estimation, and the Gibbs sampler in the Bayesian estimation. The H-M algorithm updates ϕ in the target distribution $p(\phi|y; \theta)$. The computation of its acceptance probabilities requires an explicit expression of the regression function. As a consequence, this H-M algorithm can only be applied to the approximate statistical model \mathcal{M}_h .

The standard H-M algorithm and a modified version of the random-walk H-M algorithm including the LL2 scheme leading to computational time savings in practice, are presented in Section 5.3.1. Section 5.3.2 presents the convergence of the algorithms.

5.3.1 The Hastings-Metropolis algorithm

The iterative H-M algorithm implemented on model \mathcal{M}_h generates a Markov chain with the target distribution $p_h(\phi|y; \theta)$ as invariant distribution and using a proposal density q .

At step $r + 1$, given $\phi^{(r)}$:

- Generate a candidate ϕ^c from the proposal density $q(\cdot|\phi^{(r)})$.
- Generate $U \sim \mathcal{U}([0, 1])$. Then,

$$\phi^{(r+1)} = \begin{cases} \phi^c & \text{if } U < \rho(\phi^{(r)}, \phi^c), \\ \phi^{(r)} & \text{if } U > \rho(\phi^{(r)}, \phi^c), \end{cases}$$

where

$$\rho(\phi^{(r)}, \phi^c) = \min \left\{ 1, \frac{p_{h,\phi|Y}(\phi^c) q(\phi^{(r)}|\phi^c)}{p_{h,\phi|Y}(\phi^{(r)}) q(\phi^c|\phi^{(r)})} \right\}$$

is the acceptance probability.

The choice of the proposal density q is essentially arbitrary, although in practice a careful choice will help the algorithm to move quickly inside the parameter space. Two proposal densities are combined. First the prior density $q(\cdot|\phi^{(r)}) = \pi(\cdot; \beta)$ allows to move inside the parameters space efficiently. Second, a symmetric distribution $q(\phi^c|\phi^{(r)}) = q(\phi^{(r)}|\phi^c)$ is used resulting in the so-called random-walk H-M algorithm [see e.g. 11]. In this case, $\phi^c = \phi^{(r)} + \delta$ with δ simulated from a centered symmetric distribution such that ϕ^c remains in K_1 .

Remark 1. In practice, the random-walk H-M algorithm can be modified to reduce the computational time using the LL2 scheme. Indeed, on the model \mathcal{M}_h defined by the LL scheme, the random-walk H-M algorithm requires solving the ODE (5.2.1) at ϕ^c in a bounded neighborhood of $\phi^{(r)}$, for which the LL approximate solution has been computed.

Let the centered symmetric proposal density verify the following property, called property (5.3.2): there exists $\eta > 0$ such that, almost surely,

$$\|\phi^{(r)} - \phi^c\|_{\mathbb{R}^k} < \eta. \quad (5.3.2)$$

The modified H-M algorithm including the LL2 scheme is outlined as followed. At step $r + 1$, given $\phi^{(r)}$,

- Generate a candidate $\phi^c = \phi^{(r)} + \delta$
- Accept ϕ^c with probability $\rho^{(2)}(\phi^{(r)}, \phi^c)$ where

$$\rho^{(2)}(\phi^{(r)}, \phi^c) = \min \left\{ 1, \frac{p_{h,\phi|Y}^{(2)}(\phi^c) q(\phi^{(r)}|\phi^c)}{p_{h,\phi|Y}(\phi^{(r)}) q(\phi^c|\phi^{(r)})} \right\} = \min \left\{ 1, \frac{p_{h,Y|\phi}^{(2)}(\phi^c)\pi(\phi^c)}{p_{h,Y|\phi}(\phi^{(r)})\pi(\phi^{(r)})} \right\}$$

and $p_{h,Y|\phi}^{(2)}(\cdot)$ is evaluated using the LL2 scheme.

If the move is accepted $f_h(t, \phi^{(r+1)})$ and the $p_{h,Y|\phi}(\phi^{(r+1)})$ density are re-evaluated using the LL scheme. If it is not accepted no additional matrix exponential is computed, significantly reducing the computational time.

5.3.2 Convergence

The convergence of the standard H-M algorithm and the error induced by the numerical integration scheme are studied in Theorem 1.

Theorem 1. Let f be the exact solution of ODE (5.2.1). Let $p_{\phi|Y}$ be the conditional distribution for the model \mathcal{M} . Assume that **H1**, **H2** and **H3** hold. Let f_h be the approximate solution obtained by a numerical integration method of step size h and order p . Let $p_{h,\phi|Y}$ be the conditional distribution of the model \mathcal{M}_h .

1. Then on the model \mathcal{M}_h , the H-M algorithm converges towards its stationary distribution $p_{h,\phi|Y}$.
2. Furthermore, there exists a constant C_y such that for any small h ,

$$D(p_{\phi|Y}, p_{h,\phi|Y}) \leq C_y h^p.$$

where $D(\cdot, \cdot)$ denotes the total variation distance.

The rates of convergence of the standard H-M algorithms have been widely studied [see e.g. 121], hence are not discussed here.

Remark 2. The modified H-M algorithm based on the LL2 scheme results in a non standard form of the acceptance probability due to the two approximations of the target distribution. Thus, the proof of its convergence is complex and beyond the scope of this paper. In cases where the modified H-M algorithm converges to a stationary distribution $p_h^{(2)}$, provided the property (5.3.2) is checked and the chain $(\phi^{(r)})$ remains in K_1 , there exists a constant $C_y^{(2)}$ such that $D(p_{\phi|Y}, p_h^{(2)}) \leq C_y^{(2)} \eta + C_y h^2$.

The proofs of theorem 1 and the previous inequality are given in appendix 5.7. Numerical illustrations are given in Section 5.5.

5.4 Estimation of parameters

In the following section, we extend to incomplete data models defined by ODEs, the SAEM algorithm coupled with the H-M algorithm for the maximum likelihood estimation, and the Gibbs sampling algorithm for the Bayesian approach, to estimate the parameters $\theta = (\beta, \sigma^2)$.

5.4.1 Maximum likelihood approach

The EM algorithm proposed by Dempster *et al.* [36] maximizes the $Q(\theta|\theta') = E(\log p_{Y,\phi}(\cdot; \theta)|y; \theta')$ function in two steps. At the k^{th} iteration, the E-step is the evaluation of $Q_k(\theta) = Q(\theta | \theta_k)$ while the M-step updates θ_k by maximizing $Q_k(\theta)$. For cases where the E-step has no closed form, Delyon *et al.* [35] introduce a stochastic version SAEM of the EM algorithm. The $Q_k(\theta)$ integral is evaluated by a stochastic approximation procedure. The E-step is divided into a simulation step (S-step) of the non-observed data ϕ_k with the conditional distribution $p_{\phi|Y}(\cdot; \theta_k)$ and a stochastic approximation step (SA-step):

$$Q_{k+1}(\theta) = Q_k(\theta) + \gamma_k (\log(p_{Y,\phi}(\cdot; \theta_k)) - Q_k(\theta)),$$

where (γ_k) is a sequence of positive numbers decreasing to 0. They prove the convergence of this algorithm under general conditions in the case where $p_{Y,\phi}$ belongs to a regular curved exponential family. When the non-observed data cannot be directly simulated, Kuhn & Lavielle [72] suggest using a MCMC scheme by building a Markov chain with $p_{\phi|Y}(\cdot; \theta_k)$ as unique stationary distribution at the k^{th} iteration.

When the regression function is defined by ODE, SAEM is implemented on the model \mathcal{M}_h . Let $\pi(\cdot, \beta)$ be such that $p_{h,Y,\phi}$ belongs to the exponential family:

$$p_{h,Y,\phi}(\cdot; \theta) = \exp \{-\Psi(\theta) + \langle S_h(y, \phi), \Phi(\theta) \rangle\},$$

where ψ and Φ are two functions of θ , $\langle \cdot, \cdot \rangle$ is the scalar product and $S_h(y, \phi)$ is known as the minimal sufficient statistics of the complete model, taking its value in a subset \mathcal{S} of \mathbb{R}^m . Let $\pi(\cdot; \beta)$ be of class \mathcal{C}^m . At the k^{th} iteration, the SAEM algorithm is:

- S-STEP: the non-observed data ϕ_k is simulated by the H-M algorithm developed in section 5.3 with $p_{h,\phi|Y}(\cdot; \theta_k)$ as unique stationary distribution,
- SA-STEP: s_{k+1} is updated by the stochastic approximation:

$$s_{k+1} = s_k + \gamma_k (S_h(y, \phi_k) - s_k),$$

- M-STEP: θ_k is updated by

$$\theta_{k+1} = \arg \max_{\theta} (-\Psi(\theta) + \langle s_{k+1}, \Phi(\theta) \rangle).$$

Kuhn & Lavielle [72] propose estimates of the Fisher information matrix, using the Louis's missing information principle [80], either by importance sampling or by stochastic approximation. We adapt their estimates when the regression function is not known analytically and, as a consequence, the extended SAEM supplies the standard errors of the estimates.

The convergence of SAEM is proved on \mathcal{M}_h and the distance between the likelihoods of the two models is quantified in the following theorem.

Theorem 2. *Let us consider a numerical scheme of step size h and order p . Let **H1**, **H2** and **H3** hold. Let (γ_k) be a sequence of positive numbers decreasing to 0 such that for any k in \mathbb{N} , $\gamma_k \in [0, 1]$, $\sum_{k=1}^{\infty} \gamma_k = \infty$ and $\sum_{k=1}^{\infty} \gamma_k^2 < \infty$.*

1. *Assuming the sequence $(s_k)_{k \geq 0}$ takes its values in a compact set of \mathcal{S} , the sequence $(\theta_k)_{k \geq 0}$ obtained by the SAEM algorithm on \mathcal{M}_h , converges almost surely towards a (local) maximum of the likelihood $p_{h,Y}(y, \cdot)$.*
2. *For any $\sigma_0^2 > 0$, there exists a constant θ -independent C such that*

$$\sup_{\theta=(\beta, \sigma^2) \mid \sigma^2 > \sigma_0^2} |p_Y(y; \theta) - p_{h,Y}(y, \theta)| \leq Ch^p.$$

Hence, as a principal consequence of this theorem, and assuming regularity hypotheses on the Hessians of the likelihoods of both models \mathcal{M} and \mathcal{M}_h , the bias of the estimates induced by both the numerical approximation and the estimation algorithm, is controlled.

Proof. 1. Assumptions of convergence of the SAEM algorithm are checked by the model \mathcal{M}_h . See [72] for more details.

2. In the proof of theorem 1, we obtain the result (1b) for a fixed θ and small enough h that:

$$|p_Y(y; \theta) - p_{h,Y}(y, \theta)| \leq \frac{C_y}{(2\pi\sigma^2)^{J/2}} h^p.$$

Consequently, for any $\sigma_0^2 > 0$, for any $\theta \in \Theta$ with $\sigma^2 \geq \sigma_0^2$, there exists a constant C , independent of θ , such that

$$|p_Y(y; \theta) - p_{h,Y}(y, \theta)| \leq Ch^p.$$

□

5.4.2 Bayesian approach

The Bayesian model is defined as follows:

$$\begin{aligned} y_j &= g(t_j, \phi) + \varepsilon_j & 1 \leq j \leq J \\ \varepsilon_j &\sim \mathcal{N}(0, \sigma^2) \\ \phi &\sim \pi(\cdot; \beta) \\ \theta &\sim p_\theta(\cdot; \Gamma) \end{aligned}$$

where $\theta = (\beta, \sigma^2)$ is distributed from the prior distribution p_θ with fixed hyperparameters Γ . For example, a Gamma prior distribution can be chose for σ^{-2} . The prior distribution on β depends on the specific distribution π . The Bayesian approach consists in the evaluation of the posterior distribution $p_{\theta|Y}$. The iterative Gibbs sampling algorithm is outlined as follows [see 65, for more details]:

- STEP 1: initialize the iteration counter of the chain $k = 1$ and start with initial values $\sigma^{-2(0)}, \beta^{(0)}, \phi^{(0)}$.
- STEP 2: obtain a new value $\sigma^{-2(k)}, \beta^{(k)}, \phi^{(k)}$ from $\sigma^{-2(k-1)}, \beta^{(k-1)}, \phi^{(k-1)}$ through successive generation of values
 1. $\sigma^{-2(k)} \sim p(\sigma^{-2} | \beta^{(k-1)}, \phi^{(k-1)}, y)$ $\beta^{(k)} \sim p(\beta | \sigma^{-2(k)}, \phi^{(k-1)}, y)$
 2. $\phi^{(k)} \sim p(\phi | \sigma^{-2(k)}, \mu^{(k)}, \Omega^{(k)}, y)$
- STEP 3: change the counter from k to $k + 1$ and return to STEP 2 until convergence is reached.

For a Gamma prior distribution on σ^{-2} , the conditional distribution $p(\sigma^{-2} | \beta^{(k-1)}, \phi^{(k-1)}, y)$ is a Gamma distribution. The conditional distribution $p(\beta | \sigma^{-2(k)}, \phi^{(k-1)}, y)$ depends on the specific form of the distribution $\pi(\cdot; \beta)$. To generate a realization of ϕ , Bennet *et al.* [11] describe several approaches, such as a Rejection Gibbs, a Ratio Gibbs, independent or Random-walk H-M algorithms. Gilks *et al.* [52] recommend the use of initial iterations allowing a "burn-in" phase, followed by a large number of iterations.

For models defined by ODEs, the inclusion of such H-M algorithms requires the evaluation of $g(\phi, t_j)$ at each iteration. Hence, the Gibbs algorithm is implemented on the approximate statistical model \mathcal{M}_h .

In practice, we estimate the posterior distribution of the model \mathcal{M}_h instead of the distribution of interest $p_{\theta|Y}$. The following theorem quantifies the total variation distance between the posterior distribution $p_{h,\theta|Y}$ and this original distribution of interest, $p_{\theta|Y}$.

Theorem 3. *Let us consider a numerical scheme of step size h and order p . Let $p_{\theta|Y}$ and $p_{h,\theta|Y}$ be the posterior distributions respectively of \mathcal{M} and \mathcal{M}_h . Assume that **H1**, **H2** and **H3** hold.*

1. *The Gibbs sampling algorithm converges on the model \mathcal{M}_h .*
2. *There exists a \mathbf{y} -dependent constant C_y such that*

$$D(p_{h,\theta|Y}, p_{\theta|Y}) \leq C_y h^p.$$

Hence, as a principal consequence of this theorem, the bias on the posterior mean is controlled: under moments hypotheses on $p_{\theta}(\theta)$ and $p_{h,\theta|Y}(\theta)$, there exists a constant C'_y such that $|E_{\theta|y}(\theta) - E_{h,\theta|y}(\theta)| = |\int \theta p_{\theta|Y}(\theta) d\theta - \int \theta p_{h,\theta|Y}(\theta) d\theta| \leq C'_y h^p$ where $E_{\theta|y}(\cdot)$ and $E_{h,\theta|y}(\cdot)$ are the expectation under the posterior distributions $p_{\theta|y}$ and $p_{h,\theta|y}$ respectively. Similar result can be obtained for the bias of the posterior mode, assuming regularity hypotheses on the distributions $p_{Y|\theta}$ and $p_{h,Y|\theta}$ of both models \mathcal{M} and \mathcal{M}_h .

Proof. 1. Assumptions of convergence of the Gibbs sampling algorithm are checked on the model \mathcal{M}_h , see [28] for more details.

2. By Bayes theorem, we have

$$p_{\theta|Y} = \frac{p_{Y|\theta} p_{\theta}}{p_Y}$$

where $p_Y = \int p_{Y|\theta} p_{\theta} d\theta$, and the same equality for the model \mathcal{M}_h . From the result (1b) of the proof of theorem 1, we deduce that there exists a constant C , independent of θ , such that for any $\theta \in \Theta$ with $\sigma^2 > \sigma_0^2 > 0$, $|p_{Y|\theta}(y) - p_{h,Y|\theta}(y)| \leq Ch^p$ and $|p_Y(y) - p_{h,Y}(y)| \leq Ch^p$. We now bound $|p_{\theta|Y}(\theta) - p_{h,\theta|Y}(\theta)|$:

$$\begin{aligned} |p_{\theta|Y}(\theta) - p_{h,\theta|Y}(\theta)| &\leq \frac{p_{\theta}(\theta)}{|p_Y(y)|} \left| |p_{Y|\theta}(y) - p_{h,Y|\theta}(y)| + \frac{p_Y(y)}{p_{h,Y}(y)} |p_Y(y) - p_{h,Y}(y)| \right| \\ &\leq \frac{Ch^p}{|p_Y(y)|} p_{\theta}(\theta) \left| 1 + \frac{p_{h,Y|\theta}(y)}{p_{h,Y}(y)} \right| = \frac{Ch^p}{|p_Y(y)|} (p_{\theta}(\theta) + p_{h,\theta|Y}(\theta)). \end{aligned}$$

Thus

$$D(p_{\theta|Y}, p_{h,\theta|Y}) \leq \frac{Ch^p}{|p_Y(y)|}.$$

□

5.5 Application to nonlinear mixed-effects model

Nonlinear mixed-effects models are widely used in pharmacokinetics (PK) and pharmacodynamics (PD) to estimate PK/PD parameters. They are interesting because of their capacity to discriminate the intra- from the inter-subject variabilities and to test covariate effect on the PK/PD parameters. They are modeled by:

$$\begin{aligned} y_{ij} &= C(t_{ij}, \phi_i) + \varepsilon_{ij} \\ \varepsilon_{ij} &\sim \mathcal{N}(0, \sigma^2) \\ \phi_i &\sim \pi(\cdot; \beta), \end{aligned}$$

where y_{ij} is the observation of the drug concentration C for subject i , $i = 1, \dots, N$, at time t_{ij} , $j = 1, \dots, n_i$ and ϕ_i is the vector of individual non-observed PK/PD parameters of subject i . A Gaussian distribution $\pi(\cdot; \beta) = \mathcal{N}(\mu, \Omega)$ is classically chosen for ϕ , resulting in $\beta = (\mu, \Omega)$. In the Bayesian framework, Gamma prior distribution is chosen for σ^{-2} , Wishart or Gamma prior distribution is chosen for Ω^{-1} and a Gaussian prior distribution is used for μ .

We consider a case where the drug concentration C is defined through a differential equation without analytical solution. On a simulated dataset, we illustrate applications of the H-M, the SAEM and the Bayesian Gibbs algorithms.

5.5.1 Numerical settings

The following one-compartment pharmacokinetic system with a first order absorption and a Michaelis-Menten saturable elimination describes the concentration C of a drug:

$$\frac{dC}{dt}(t, \phi) = \frac{k_a D}{V} e^{-k_a t} - \frac{V_m C(t, \phi)}{k_m + C(t, \phi)}, \quad (5.5.3)$$

where D is the known administered dose, V the total volume of distribution, k_a the absorption constant, k_m the Michaelis-Menten constant and V_m the maximum rate of metabolism. Hence, the parameter vector is $\phi = [V, k_a, k_m, V_m] \in \mathbb{R}^4$.

We consider the pharmacokinetic parameters of hydroxurea, an anti-cancerous drug, studied by Tracewell *et al.* [123]: $V=12.2$ L, $k_a=2.72$ h⁻¹, $k_m=0.37$ mmol/L, $V_m=0.082$ mmol/h/L. One dataset of 20 patients is simulated with a dose $D = 13.8$ mmol and measurements at time $t = 0, 0.5, 1, 1.5, 2$ hours and then every hour until 12 hours. We choose a Gaussian distribution for ϕ with a diagonal variance-covariance matrix (diagonal components equal to 0.4). The residual variance σ^2 is set to 0.01. The numerical method used to simulate this dataset is the ode45 solving Matlab function, that implements a Runge-Kutta scheme of the fourth order, with a very small maximal step size of resolution h equal to 0.001. Data are plotted on figure 5.1.

5.5.2 Results

Several numerical integration methods included in the H-M algorithm are compared on the simulation of the conditional distributions. 1000 independent 200-long Markov chains are generated for each method. The first method is the ode45 function with maximal step size $h = 0.001$ in order to obtain an exact cumulative distribution function; this is the reference as this numerical method is of the fourth order with a very small step size. Then we compare the two cumulative distribution functions obtained by using at first ode45 with the default step size $h = 0.1$ and then the LL and LL2 schemes. As seen on Figure 5.2, plotting the empirical cumulative distribution functions, the three numerical methods provide similar simulated conditional distributions.

The SAEM and Bayesian Gibbs algorithms are applied to the simulated dataset. We implemented the SAEM algorithm in Matlab and the Bayesian Gibbs algorithm is the one implemented in WinBugs and its Differential Interface software [116]. These estimates are compared with those obtained by the NONMEM software, proposed by Beal & Sheiner [10] and used by 80% of the pharmacokineticists in drug companies. NONMEM is an implementation of the First Order Conditional Estimate algorithm, which is based on a first order linearization of the regression function around the conditional estimates of the parameters ϕ , the standard errors of the estimates being evaluated by linearization.

Initial values have been arbitrarily chosen and are presented in Table 5.1. The evolutions of each SAEM parameter estimate are plotted against iterations on Figure 5.3. The estimates converge rapidly to a neighborhood of the simulated values. In the Bayesian framework, a Gamma prior distribution has first been used on Ω^{-1} leading to unacceptable convergence graphs. A Wishart prior distribution

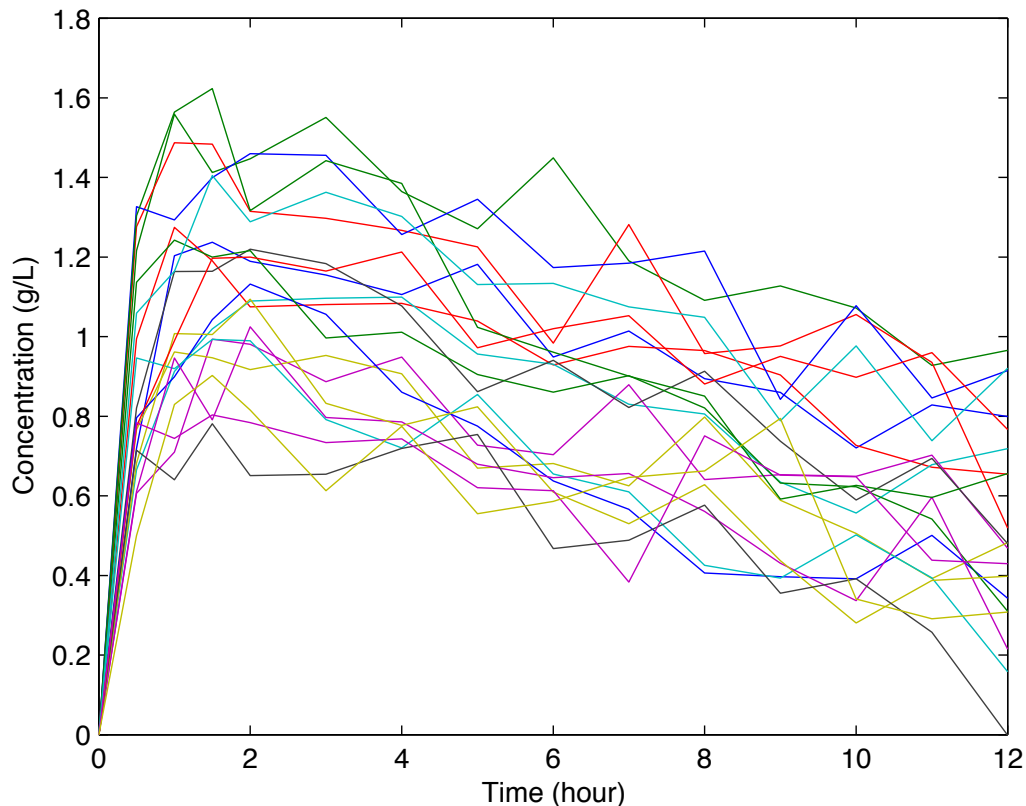


Figure 5.1: Individual concentrations of pharmacokinetic hydroxurea simulated for 20 patients.

has then been used on Ω^{-1} with satisfactory convergence and posterior density graphs after 100000 iterations.

The parameter estimates and their standard errors obtained by SAEM, NONMEM and the Bayesian Gibbs algorithms are presented in Table 5.1.

In this example, it takes about ten minutes for SAEM to converge using a conventional Intel Pentium IV 3,2 GHz workstation. All the SAEM estimates almost reach the simulated values. The SAEM algorithm achieves the evaluation of the information Fisher matrix, and almost all the standard errors of the estimates are satisfactory. Using the same computer, the NONMEM software stops after about ten minutes without convergence towards the maximum of the likelihood. The estimate of V_m does not change from its initial value, the estimate of k_m is far from its simulation value, $\text{var } k_m$ is estimated near zero while $\text{var } V_m$ and $\text{var } k_a$ are overestimated. The NONMEM software fails to evaluate the standard errors of all estimates. Using the same computer, it takes about one hour for the Winbugs software to compute 100 000 iterations of one Markov chain of the Gibbs algorithm. All the Bayesian estimates almost reach the simulated values and the standard errors of the estimates are satisfactory.

This simulated example illustrates the ability of the proposed algorithms to estimate precisely the parameters of mixed models defined by ODE.

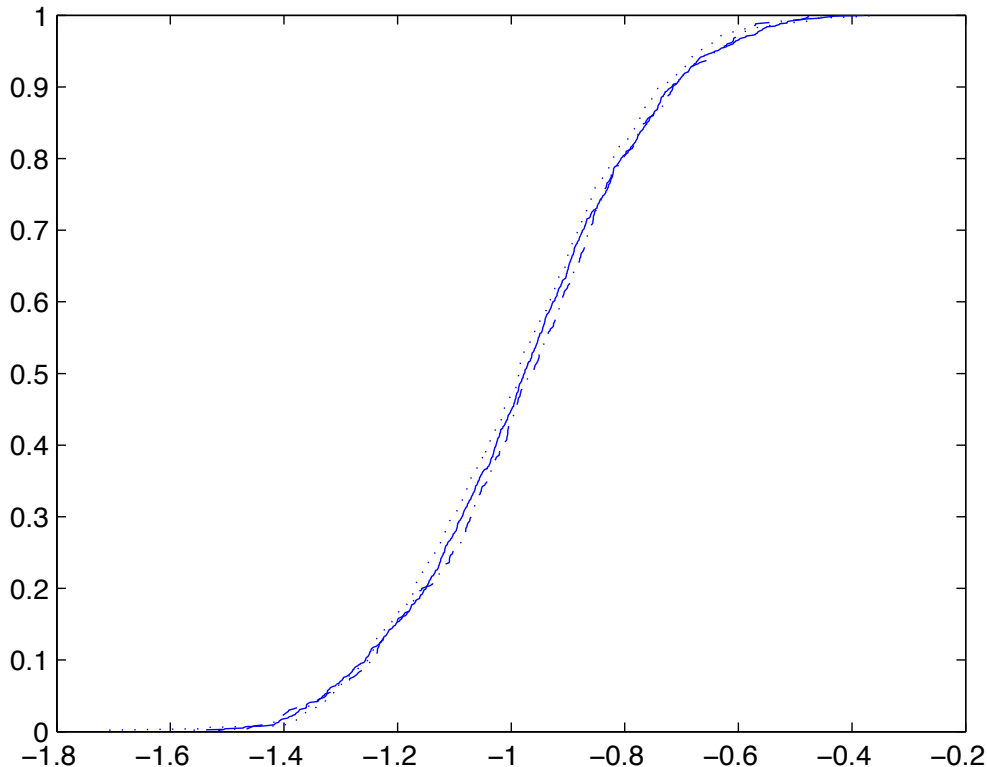


Figure 5.2: Empirical cumulative distribution functions of the k_m conditional distribution simulated by Hastings-Metropolis using a very precise Runge-Kutta solving scheme (plain line), a classical Runge-Kutta scheme (dotted line), or the LL and LL2 schemes (dashed line).

5.6 Discussion

This paper extends the statistical approaches used to estimate incomplete data model parameters to the frequent cases where such a model \mathcal{M} is defined by a dynamical system. To that purpose, an approximate model \mathcal{M}_h is introduced, of which regression function is evaluated by a numerical integration method. The standard estimation algorithms are adapted to estimate this approximate model. The convergence of the H-M, the SAEM and the Gibbs Sampling algorithms on the model \mathcal{M}_h is proved.

This paper quantifies the error induced by the use of a numerical solving method. The errors on the conditional distribution, the likelihood and the posterior distribution between the model \mathcal{M} and the approximate model \mathcal{M}_h are controlled by h^p , where h is the step size and p the order of the numerical integration method. This error is distinct from the error on the estimates induced by the estimation algorithms, which is classically controlled by the standard errors evaluated through the Fisher information matrix of the estimates.

This paper proposes an extended version of the LL scheme, using the dependence to the parameter ϕ of the ODE, which allows to reduce significantly the computational cost of the H-M algorithm in practice. The major advantage of the LL schemes is its behavior on stiff dynamical systems, providing an interesting trade-off between computational cost and numerical stability when included in an iterative stochastic algorithm [18].

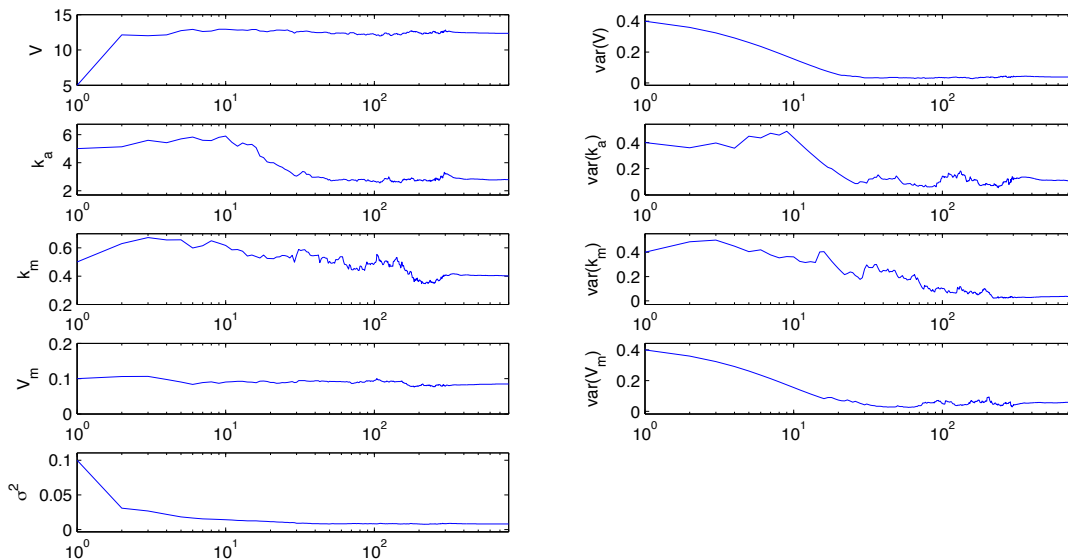


Figure 5.3: Evolution of the estimates in function of the iteration of SAEM algorithm (with a logarithm scale for the abscis axis).

Regarding the maximum likelihood, we study the SAEM algorithm instead of the Monte-Carlo EM proposed by Wei and Tanner [127] or Wu [134]. With an analytical regression function, the MCEM is computational intensive because of the large sample of non-observed data simulated at each iteration, while the Stochastic Approximation method requires the generation of only one realization of non-observed data at each iteration. Thus, due to computational time considerations, we only extend the SAEM algorithm to the case of regression function implicitly defined.

The SAEM algorithm is applied to a dataset simulated using a pharmacokinetic model defined by ODEs. The SAEM estimates are compared with those obtained by the standard estimation software NONMEM, the only available software providing estimates by maximum likelihood in nonlinear mixed models defined by ODEs. SAEM provides satisfying estimates and standard errors of the parameters, while NONMEM does not converge on this simulated example and fails to evaluate the standard errors. The estimation algorithm implemented in NONMEM is based on the linearization of the regression function. Despite the fact that Vonesh [125] highlights problems of consistence and convergence of the estimates produced by such estimation methods, NONMEM is used by 80% of the pharmacokineticists in the pharmaceutical industry. The simulation results presented in this paper point out the poor ability of this software to estimate the parameters in nonlinear mixed model defined through ODEs. Thus we recommend using the SAEM algorithm to analyze such incomplete data problems defined by ODE. The SAEM algorithm is implemented by the MONOLIX group in a free Matlab function (<http://mahery.math.u-psud.fr/~lavielle/monolix/index.html>). The extension of the monolix function to ODE models will soon be available on the same web-site.

Several methods have been suggested to simulate the non-observed data included in a Bayesian approach. Wakefield [126] *et al.* sample the non-observed data using a ratio-of-uniform while Tierney [121] proposes using a Hastings-Metropolis algorithm. Gilks *et al.* [52] summarize these methods in their book. We use the Winbugs software and its Differential Interface software [116], which implements hybrid Gibbs and Hastings-Metropolis algorithms. On the simulated dataset, the posterior distributions are well estimated while the subject number was low.

In this paper, the methodology is illustrated on a mixed model issued from pharmacocinetics. How-

Table 5.1: Parameter estimates obtained by the SAEM, NONMEM and Bayesian Gibbs algorithms on the simulated dataset.

	V	k_a	k_m	V_m	var V	var k_a	var k_m	var V_m	σ^2
initial values	5.0	5.00	0.50	0.100	0.400	0.400	0.400	0.400	0.1000
simulation values	12.2	2.72	0.37	0.082	0.040	0.040	0.040	0.040	0.0100
<i>SAEM</i>									
estimates	12.3	2.79	0.40	0.085	0.038	0.049	0.039	0.038	0.0081
SE	0.5	0.21	0.01	0.004	0.012	0.019	0.013	0.013	6.10^{-4}
<i>NONMEM</i>									
estimates	12.3	2.57	0.60	0.100	0.036	0.068	10^{-8}	0.062	0.0088
SE	-	-	-	-	-	-	-	-	-
<i>Bayesian Gibbs</i>									
mode estimates	12.4	2.64	0.39	0.085	0.040	0.040	0.048	0.052	0.0082
SE	0.6	0.23	0.62	0.044	0.014	0.046	0.028	0.029	7.58^{-4}

ever, there exist many others fields of applications. First, it can be applied in Functional Magnetic Resonance Imaging (fMRI), a neuroimaging technique using the vascular oxygenation contrast as an indirect measure of cerebral activity. Despite the extensive development of such techniques, the coupling mechanisms between neuronal activity and cerebral physiological changes -such as vascular changes- are still poorly understood. Recently, dynamical models such as the Balloon model have been introduced to explain these interactions [25], leading to define statistical models defined by ODEs. Many others applications can be found in the framework of functional data analysis in the case where the underlying process can be described by a differential system.

Diffusion models described by stochastic differential equations (SDEs) are natural extensions to the corresponding deterministic models (defined by ODEs) to account for time-dependent residual errors and to handle real life variations in model parameters occurring over time. The two estimation methods proposed in this paper can be extended to this case, opening wide perspectives of application.

5.7 Appendix A: Convergence of the H-M algorithm:

1. Proof of Theorem 1

(a) Following [121], the H-M algorithm provides a uniformly ergodic Markov chain with $p_{h,\phi|Y}$ as invariant distribution, as soon as π is one of the proposal density.

(b) By Bayes theorem, we have $p_{\phi|Y}(\phi) = \frac{p_{Y|\phi}(y|\phi)\pi(\phi)}{p_Y(y)}$, and

$$|p_{\phi|Y}(\phi) - p_{h,\phi|Y}(\phi)| = \frac{\pi(\phi)}{p_Y(y)} \left[|p_{Y|\phi}(y) - p_{h,Y|\phi}(y)| + \frac{p_{h,Y|\phi}(y)}{p_{h,Y}(y)} |p_Y(y) - p_{h,Y}(y)| \right].$$

We have $|p_{Y|\phi}(y|\phi) - p_{h,Y|\phi}(y|\phi)| = |p_{Y|\phi}(y)| \left| 1 - \frac{p_{h,Y|\phi}(y)}{p_{Y|\phi}(y)} \right|$, and

$$p_{Y|\phi}(y) = \frac{1}{(2\pi\sigma^2)^{J/2}} \exp \left[-\frac{1}{2\sigma^2} \sum_{j=1}^J (y_j - g(t_j, \phi))^2 \right].$$

Furthermore, by **H3**, we have:

$$\begin{aligned} |g_h(t, \phi) - g(t, \phi)| &= |H \circ f_h(t, \phi) - H \circ f(t, \phi)| \\ &\leq L_H \sup_{t, \phi} |f_h(t, \phi) - f(t, \phi)| \leq L_H C h^p. \end{aligned}$$

By **H1** and the assumptions on the proposal densities, (t, ϕ) remains a.s. uniformly in the compact set $[t_0, T] \times K_1$. Thus there exist M and, M_h such that

$$M = \sup_{(t, \phi) \in [t_0, T] \times K_1} |g(t, \phi)|, \quad M_h = \sup_{(t, \phi) \in [t_0, T] \times K_1} |g_h(t, \phi)|,$$

and, we can prove that $M_h \leq M + Ch^p$. Hence, there exists A_h such that

$$\left| (y_j - g_h(t_j, \phi))^2 - (y_j - g(t_j, \phi))^2 \right| \leq A_h Ch^p$$

and, we obtain a ϕ -independent bound:

$$|p_{Y|\phi}(y) - p_{h, Y|\phi}(y)| \leq \frac{1}{(2\pi\sigma^2)^{J/2}} (e^{\frac{1}{2\sigma^2} J A_h Ch^p} - 1) \leq \frac{1}{(2\pi\sigma^2)^{J/2}} (e^{B A_h h^p} - 1),$$

with $B = \frac{JC}{2\sigma^2}$. Consequently, by integrating the previous inequality,

$$|p_Y(y) - p_{h, Y}(y)| \leq \int |p_{Y|\phi}(y) - p_{h, Y|\phi}(y)| \pi(\phi) d\phi \leq \frac{1}{(2\pi\sigma^2)^{J/2}} (e^{B A_h h^p} - 1).$$

Finally, using the previous inequalities, we have:

$$\int |p_{\phi|Y}(\phi) - p_{h, \phi|Y}(\phi)| d\phi \leq \frac{2}{p_Y(y) (2\pi\sigma^2)^{J/2}} (e^{B A_h h^p} - 1).$$

Let $D(\cdot, \cdot)$ denote the total variation distance. If h is small enough, there exists a h -independent constant C_y such that:

$$D(p_{\phi|Y}, p_{h, \phi|Y}) \leq C_y h^p.$$

2. Proof of the inequality of remark 2

Let assume that the Markov Chain has a stationary distribution $p_h^{(2)}$. At step $r + 1$, let $\phi^{(r)}$ be the current value and ϕ^c be the candidate of the Markov chain. By definition of the LL and LL2 schemes, we have

$$|f_h(t, \phi^c) - f_{h, \phi^{(r)}}(t, \phi^c)| = O(\|\phi^c - \phi^{(r)}\|_{\mathbb{R}^k}). \quad (5.7.4)$$

Moreover, by definition of the acceptance probability,

$$|\rho(\phi^{(r)}, \phi^c) - \rho^{(2)}(\phi^{(r)}, \phi^c)| \leq \frac{\pi(\phi^c)}{p_{h, Y|\phi}(\phi^{(r)}) \pi(\phi^{(r)})} \left| p_{h, Y|\phi}(\phi^c) - p_{h, Y|\phi}^{(2)}(\phi^c) \right|$$

Let quote

$$M_{LL} = \sup_{(t, \phi) \in [t_0, T] \times K_1} |H \circ f_h(t, \phi)| \quad \text{and} \quad M_{LL2} = \sup_{(t, \phi) \in [t_0, T] \times K_1} |H \circ f_{h, \phi^{(r)}}(t, \phi)|.$$

These quantities exist as ϕ remains in the compact set K_1 by definition of the proposal density and, because $\phi \rightarrow f_0(\phi)$ is \mathcal{C}^1 . According to (5.7.4), there exists M such that:

$$\begin{aligned} & \left| (y_j - H \circ f_h(t_j, \phi^c))^2 - (y_j - H \circ f_{h, \phi^{(r)}}(t_j, \phi^c))^2 \right| \\ & \leq \underbrace{(2 \max |y_j| + M_{LL} + M_{LL2}) L_H M}_{\equiv A} \underbrace{\|\phi^c - \phi^{(r)}\|_{\mathbb{R}^k}}_{\leq \eta \text{ by property (2)}}. \end{aligned}$$

Consequently we have

$$|p_{h, Y|\phi}(\phi^c) - p_{h, Y|\phi}^{(2)}(\phi^c)| \leq \frac{1}{(2\pi\sigma^2)^{J/2}} (e^{\frac{1}{2\sigma^2} J A \eta} - 1).$$

As $p_{h,Y|\phi}$ is continuous in ϕ , there exists a constant c such that

$$\inf_{\phi \in K_1} p_{h,Y|\phi}(\phi) \geq \frac{c}{(2\pi\sigma^2)^{J/2}}.$$

By **H1**, and combining the previous inequalities, we obtain

$$|\rho(\phi^{(r)}, \phi^c) - \rho^{(2)}(\phi^{(r)}, \phi^c)| \leq \frac{b}{ca} \left(e^{\frac{1}{2\sigma^2} JA\eta} - 1 \right) \leq B\eta, \quad (5.7.5)$$

for a η small enough. The transition kernel of the Markov chains simulated by the standard and, modified Hastings-Metropolis algorithms are quoted \mathcal{K} and $\mathcal{K}^{(2)}$ respectively. Using the previous inequality, we have:

$$\begin{aligned} \left| \mathcal{K}(\phi; \{\phi^c\}) - \mathcal{K}^{(2)}(\phi; \{\phi^c\}) \right| &= \left| \rho(\phi, \phi^c) - \rho^{(2)}(\phi, \phi^c) \right| q(\phi^c|\phi) \\ &\leq B\eta q(\phi^c|\phi) \end{aligned}$$

As a consequence, we have

$$\sup_{\phi \in K_1} D \left(\mathcal{K}(\phi; \cdot), \mathcal{K}^{(2)}(\phi, \cdot) \right) \leq B\eta$$

The result follows from the part 1 of this theorem, applied for the Local Linearization scheme, combined with the following lemma.

Lemma 2. *Let \mathcal{K} and $\mathcal{K}^{(2)}$ be the respective transition kernels of two Markov chains defined on a space \mathcal{E} and let p and $p^{(2)}$ be their respective stationary distributions. Assume that \mathcal{K} supplies a uniformly ergodic chain and that there exists a constant C such that*

$$\sup_{\phi \in \mathcal{E}} D \left(\mathcal{K}(\phi; \cdot), \mathcal{K}^{(2)}(\phi, \cdot) \right) \leq C.$$

Then, there exists a constant α such that:

$$D \left(p, p^{(2)} \right) \leq \alpha C.$$

This lemma directly derives from the Poisson equality:

$$p \cdot f - p^{(2)} \cdot f = p^{(2)} (\mathcal{K}^{(2)} - \mathcal{K}) \cdot Vf$$

where $Vf(x) = \sum_{n=0}^{\infty} (\mathcal{K}^n f(x) - p \cdot f)$ and $p \cdot f = \int f(x)p(dx)$ for any measurable function f . □

s suggested by the referees, we underline that no hypothesis is assumed on the stationary distribution. The result is based on the transition kernel properties.

5.8 Appendix B: The modified Local Linearization scheme

The LL scheme is based on a local linearization of the second member of ODE (5.2.1) with respect to time t and f . The new LL2 scheme is deduced using a Taylor expansion of the right term with respect to t , f and ϕ .

5.8.1 Principle

Let the equation (5.2.1) be solved by the LL scheme at a given ϕ_0 . Let ϕ be in a bounded neighborhood of ϕ_0 . Let the time interval $[t_0, T]$ be divided in N sub-intervals $[t_n, t_{n+1}[$, $t_n = t_0 + nh$, $n = 0, \dots, N-1$, where h is the step size of the method. On each time interval $[t_n, t_{n+1}[$, the linearized equation deriving from the equation (5.2.1) at (t, ϕ) with $t \in [t_n, t_{n+1}[$ is:

$$\begin{aligned} F(f(t, \phi), t, \phi) &\simeq F(f(t_n, \phi_0), t_n, \phi_0) + \frac{dF}{df}(f(t_n, \phi_0), t_n, \phi_0) (f(t, \phi) - f(t_n, \phi_0)) \\ &+ \frac{dF}{dt}(f(t_n, \phi_0), t_n, \phi_0) (t - t_n) + \frac{dF}{d\phi}(f(t_n, \phi_0), t_n, \phi_0) (\phi - \phi_0). \end{aligned} \quad (5.8.6)$$

The following notations are introduced:

$$\begin{cases} f_n(\phi) &= f_h(t_n, \phi) & F'_n(\phi) &= \frac{dF}{dt}(f_n(\phi), t_n, \phi) \\ F_n(\phi) &= F(f_n(\phi), t_n, \phi) & R_k(X, h) &= \int_0^h \exp(uX)u^k du \\ DF_n(\phi) &= \frac{dF}{df}(f_n(\phi), t_n, \phi) & D_\phi F(\phi_0) &= \frac{dF}{d\phi}(f_n(\phi_0), t_n, \phi_0). \end{cases}$$

The LL2 scheme is:

$$f_{n+1}(\phi) = f_n(\phi_0) + \Lambda_n^m(\phi_0, \phi, h) \quad (5.8.7)$$

with

$$\begin{aligned} \Lambda_n^m(\phi_0, \phi, h) &= [hR_0(DF_n(\phi_0), h) - R_1(DF_n(\phi_0), h)] F'_n(\phi_0) \\ &+ R_0(DF_n(\phi_0), h) (F_n(\phi_0) + D_\phi F(\phi_0)(\phi - \phi_0)) + \exp(hDF_n(\phi_0)) (f_n(\phi) - f_n(\phi_0)). \end{aligned}$$

Remark 3. 1. The previous recursive formula taken at $\phi = \phi_0$ is the same as the one deriving from the LL scheme.

2. For autonomous dynamic system, the scheme (5.8.7) is simplified by:

$$f_{n+1}(\phi) = f_n(\phi_0) + \lambda_n^m(\phi_0, \phi, h),$$

with

$$\begin{aligned} \lambda_n^m(\phi_0, \phi, h) &= R_0(DF_n(\phi_0), h) (F_n(\phi_0) + D_\phi F(\phi_0)(\phi - \phi_0)) \\ &+ \exp(hDF_n(\phi_0)) (f_n(\phi) - f_n(\phi_0)). \end{aligned}$$

5.8.2 Convergence of the method

Biscay *et al.* [18] and Ramos & García-López [104] prove that the LL scheme is of convergence rate h^2 . We extend their results to the LL2 scheme:

Lemma 3 (Error estimation). *Let f be the exact solution of ODE (5.2.1) and f_{h, ϕ_0} be the approximate solution obtained by the LL2 scheme with the step size h , given a point ϕ_0 . Under assumptions **H1'** and **H2**, there exist two constants C_1 and C_2 , ϕ and ϕ_0 -independent such that for any $t \in [t_0, T]$ and for any ϕ ,*

$$|f(t, \phi) - f_{h, \phi_0}(t, \phi)| \leq \max(C_1 h^2, C_2 \|\phi - \phi_0\|_{\mathbb{R}^k}^2).$$

Proof. The proof, essentially the same as the [104]'s one, is presented for $d = 1$ but is easily generalizable for any d .

On the interval $[t_n, t_{n+1}[$, $0 \leq n \leq N-1$, f_{h, ϕ_0} is the exact solution of the following linear equation:

$$\begin{cases} \frac{\partial f_{h, \phi_0}(t, \phi)}{\partial t} &= F_{h, \phi_0}(f_{h, \phi_0}(t, \phi), t, \phi) \\ f_{h, \phi_0}(t_0, \phi) &= f_0(\phi), \end{cases}$$

where F_{h,ϕ_0} is the Taylor expansion of F at point (t_n, ϕ_0) defined by the equation (5.8.6). For any t in $[t_n, t_{n+1}]$, we have:

$$|f(t, \phi) - f_{h,\phi_0}(t, \phi)| = \left| \int_{t_n}^t F(f(u, \phi), u, \phi) - F_{h,\phi_0}(f_{h,\phi_0}(u, \phi), u, \phi) du + f(t_n, \phi) - f_{h,\phi_0}(t_n, \phi) \right|$$

Assuming **H2**, F is \mathcal{C}^2 on its domain of definition. Since F_{h,ϕ_0} is the second order Taylor expansion of F at (t_n, ϕ_0) , there exists $\xi \in \mathbb{R}^d \times (t_n, t_{n+1}) \times \mathbb{R}^k$ such that:

$$\begin{aligned} |F(f_{h,\phi_0}(u, \phi), u, \phi) - F_{h,\phi_0}(f_{h,\phi_0}(u, \phi), u, \phi)| &\leq \frac{1}{2} \left| \frac{\partial^2 F}{\partial f^2}(\xi, \phi_0)(f_{h,\phi_0}(u, \phi) - f_{h,\phi_0}(t_n, \phi_0))^2 \right| \\ &+ \frac{1}{2} \left| \frac{\partial^2 F}{\partial t^2}(\xi, \phi_0)(u - t_n)^2 \right| + \frac{1}{2} \left| \frac{\partial^2 F}{\partial f \partial t}(\xi, \phi_0)(f_{h,\phi_0}(u, \phi) - f_{h,\phi_0}(t_n, \phi_0))(u - t_n) \right| \\ &+ \frac{1}{2} \left| \sum_{m=1}^k \frac{\partial^2 F}{\partial f \partial \phi_m}(\xi, \phi_0)(f_{h,\phi_0}(u, \phi) - f_{h,\phi_0}(t_n, \phi_0))(\phi_m - \phi_{0_m}) \right| + \frac{1}{2} \left| \sum_{m=1}^k \frac{\partial^2 F}{\partial t \partial \phi_m}(\xi, \phi_0)(u - t_n)(\phi_m - \phi_{0_m}) \right| \\ &+ \frac{1}{2} \left| \sum_{m', m=1}^k \frac{\partial^2 F}{\partial \phi_m \partial \phi_{m'}}(\xi, \phi_0)(\phi_m - \phi_{0_m})(\phi_{m'} - \phi_{0_{m'}}) \right|. \end{aligned} \quad (5.8.8)$$

Assuming **H1'** and **H2**, there exists a constant c independent of t and ϕ , which upper-bounds every second order differential of F . Moreover, a short recursive argument together with **H1'** implies that f_{h,ϕ_0} is \mathcal{C}^1 on $[t_0, T] \times K_1$. Thus, there exist constants η and η' such that

$$|f_{h,\phi_0}(u, \phi) - f_{h,\phi_0}(t_n, \phi_0)| \leq \max(\eta|u - t_n|, \eta' \|\phi - \phi_0\|_{\mathbb{R}^k})$$

Hence, using (5.8.8), there exist two constants A and A' such that:

$$|F(f_{h,\phi_0}(u, \phi), u, \phi) - F_{h,\phi_0}(f_{h,\phi_0}(u, \phi), u, \phi)| \leq \max(A|u - t_n|, A' \|\phi - \phi_0\|_{\mathbb{R}^k})^2.$$

The F function is \mathcal{C}^2 . Thus, by quoting L_F its Lipschitz constant, we can write:

$$|F(f(u, \phi), u, \phi) - F(f_{h,\phi_0}(u, \phi), u, \phi)| \leq L_F \|f(u, \phi) - f_{h,\phi_0}(u, \phi)\|$$

Finally, by quoting $E_\phi(u) = \|f(u, \phi) - f_{h,\phi_0}(u, \phi)\|$, and combining the previous inequalities, we have:

$$E_\phi(t) \leq \int_{t_n}^t L_F E_\phi(u) du + \int_{t_n}^t \max(A|u - t_n|, A' \|\phi - \phi_0\|_{\mathbb{R}^k})^2 du + E_\phi(t_n).$$

As $|u - t_n| \leq |t - t_n| \leq h$, we have:

$$E_\phi(t) \leq \int_{t_n}^t L_F E_\phi(u) du + \max(Ah, A' \|\phi - \phi_0\|_{\mathbb{R}^k})^2 (t - t_n) + E_\phi(t_n).$$

The expected result derives from the following lemma and similar arguments as those presented by Ramos & García-López [104]:

Lemma 4. *Let u be a positive function such that*

$$u(t) \leq a + b(t - t_0) + c \int_{t_0}^t u(s) ds$$

Then we have

$$u(t) \leq ae^{c(t-t_0)} + \frac{b}{c}(e^{c(t-t_0)} - 1).$$

□

Acknowledgements

The authors are grateful to their advisor Professor Marc Lavielle for his constructive advice and help. The authors thank Professor France Mentré for her helpful comments. The authors would like to thank also Rolando Biscay and Jean-Louis Foulley for helpful discussions about respectively Local Linearisation schemes and Bayesian estimation.

Chapitre 6

Application en IRMf : Introduction d’un modèle physiologique

Sommaire

6.1	Introduction	95
6.2	Description du modèle physiologique	97
6.3	Propriétés du modèle physiologique	100
6.4	Estimation paramétrique	108
6.5	Illustration sur données simulées	111
6.6	Application sur données réelles	115
6.7	Discussion	122

6.1 Introduction

Nous avons souligné en discussion du chapitre 4, la nécessité d’introduire un modèle physiologique afin de mieux comprendre les relations entre les phénomènes hémodynamiques et la dynamique neuronale sous-jacente. En outre, l’hypothèse de linéarité du modèle de convolution est une hypothèse forte qui n’est pas toujours vérifiée dans la réalité. En effet, plusieurs phénomènes non-linéaires ont été mis en évidence [12, 17, 66] notamment un phénomène de *suppression à la répétition*. Plus précisément, quand un stimulus est précédé par un autre et qu’ils sont rapprochés dans le temps (moins de 2 secondes environ), la réponse au deuxième stimulus est plus faible que si il avait été isolé dans le temps. Nous nous intéressons donc au “*Balloon model*”– modèle physiologique originellement proposé par Buxton [23] ainsi qu’à certaines de ses extensions permettant de répondre à ces problématiques.

Le “*Balloon model*” lie le débit sanguin local au signal BOLD, assimilant le capillaire sanguin (aussi appelé “compartiment”) à un ballon se dilatant sous l’effet des variations du débit sanguin local. Nous avons vu (voir partie 1) que toute activation cérébrale entraîne une augmentation locale du débit sanguin en réponse aux besoins métaboliques du cerveau. Nous avons d’autre part souligné qu’il existe un déséquilibre entre les réels besoins du cerveau et cette augmentation du débit sanguin local, ce déséquilibre étant à l’origine même de la cartographie cérébrale par Imagerie par Résonance Magnétique. Il est expliqué par Buxton *et al.* [23] comme étant le résultat d’un étroit couplage entre débit sanguin et métabolisme en présence d’une quantité limitée d’oxygène délivrée. Ce modèle de limitation d’oxygène repose sur deux hypothèses :

- (MB1) *D'une part, la totalité de l'oxygène quittant les capillaires est métabolisée. En d'autres termes, le métabolisme de l'oxygène est parfaitement efficace.*
- (MB2) *D'autre part, l'augmentation du débit sanguin n'est pas due à un recrutement de capillaires mais à une augmentation de la vitesse de parcours du sang dans les capillaires. De nombreuses preuves ont été apportées confortant cette dernière hypothèse (voir [23] pour des références).*

Sous l'hypothèse (MB2), une augmentation du débit sanguin local conduit à une diminution du temps moyen de transit dans le capillaire –le sang parcourt plus vite le capillaire– donc une extraction moins efficace de l'oxygène. Ainsi, la quantité d'oxygène délivrée augmente beaucoup moins que le débit lui-même. Par conséquent, une faible augmentation des besoins en oxygène métabolisé requiert une forte augmentation du débit sanguin local.

Le “*Balloon model*” est un modèle du type “entrée-sortie”, à deux variables d'état (le volume sanguin et la concentration en déoxyhémoglobine). L'entrée du système est le débit sanguin local tandis que la sortie du système est le signal BOLD. Le signal BOLD résultant du “*Balloon model*” est alors une fonction non-linéaire du volume du ballon et de la concentration en déoxyhémoglobine. Friston *et al.* [50] ont élargi ce modèle en introduisant une composante supplémentaire –appelée composante “*rCBF*”– liant l'activité neuronale au débit sanguin local.

Le modèle hémodynamique global –liant finalement l'activité neuronale au signal BOLD– est présenté dans ce chapitre. Un schéma récapitulatif a été représenté sur la figure 2.1 dans la partie “introduction”.

La mise en équations des phénomènes précédemment décrits aboutit à un système d'équations différentielles à quatre équations présenté dans la section 6.2. Nous évoquons les extensions possibles de ce modèle au cours de cette même section.

Le système d'équations différentielles ordinaires (EDOs) obtenu dépend de six paramètres physiologiques, notés $\mathbf{a} = (a_1, a_2, a_3, a_4, a_5, a_6)$ dans la suite. En outre, le signal est déclenché par les réponses neuronales aux stimuli du paradigme subi. Ces réponses neuronales sont pondérées par des paramètres “efficacités” notés $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_J)$ où α_j est l'efficacité de la j -ième réponse neurale. L'influence de chaque paramètre physiologique sur le signal BOLD généré par le système d'ODEs est étudié dans la section 6.3. Nous mettons de plus en évidence la non-linéarité du modèle.

En règle générale [23, 37, 50, 105], les efficacités neuronales sont prises constantes :

$$\forall j = 1, \dots, J, \alpha_j = \alpha.$$

Dans cette thèse nous cherchons à modéliser une variabilité supplémentaire. Dans un premier temps, cette variabilité peut être inter-types de stimuli. Ainsi, nous définissons une efficacité neuronale par type de stimuli. Dans un deuxième temps, nous pouvons envisager d'introduire une variabilité inter-occurrences comme nous l'avons fait dans le cas linéaire. En l'absence de toute autre hypothèse supplémentaire, plaçons nous dans le cas le plus général et modélisons une variabilité inter-occurrences. Comme dans le cas linéaire (voir chapitre 4), supposons que le paradigme contienne K types de stimuli. Pour tout $k = 1, \dots, K$ soient $(\alpha_{k,j})_{j=1, \dots, J_k}$ les efficacités neuronales des stimuli de type k (par ordre d'occurrence) et supposons les efficacités distribuées selon une loi gaussienne de paramètres (μ_k, ω_k) : $\forall j = 1, \dots, J_k$

$$\alpha_{k,j} \sim_{\text{i.i.d}} \mathcal{N}(\mu_k, \omega_k).$$

Lorsqu'une erreur de mesure est prise en compte, les observations sont décrites de la façon suivante. Notons $(t_1, \dots, t_n) \in [0, +\infty[^n$ les instants d'observation, $(y(t_1), \dots, y(t_n)) \in \mathbb{R}^n$ les observations du signal BOLD en ces instants. Soit $x : [0, +\infty[\times \mathbb{R}^5 \times \mathbb{R}^J \rightarrow \mathbb{R}^4$ la solution du système suivant :

$$\begin{cases} \frac{\partial \mathbf{x}(t, \mathbf{a}, \boldsymbol{\alpha})}{\partial t} &= F(\mathbf{x}(t, \mathbf{a}, \boldsymbol{\alpha}), t, \mathbf{a}, \boldsymbol{\alpha}) \\ \mathbf{x}(0, \mathbf{a}, \boldsymbol{\alpha}) &= \mathbf{x}_0 \end{cases}$$

où $F : \mathbb{R}^4 \times [0, +\infty[\times \mathbb{R}^5 \times \mathbb{R}^J \rightarrow \mathbb{R}^4$ est définie par le “*Balloon model*” et $x_0 \in \mathbb{R}^4$ est la condition initiale. Soit $H : \mathbb{R}^4 \rightarrow \mathbb{R}$ une fonction connue, définie par le “*Balloon model*”. Alors nous sommes amenés à considérer le modèle statistique global suivant : $\forall i = 1, \dots, n$,

$$\begin{cases} \mathbf{y}(t_i) &= H \circ x(t_i, \mathbf{a}, \boldsymbol{\alpha}) + \varepsilon_i \\ \alpha_{k,j} &\sim \text{i.i.d.} \mathcal{N}(\mu_k, \omega_k) \\ \varepsilon_i &\sim \text{i.i.d.} \mathcal{N}(0, \sigma^2) \end{cases}$$

Nous cherchons à estimer d’une part les paramètres physiologiques \mathbf{a} et d’autre part les paramètres $(\mu_k, \omega_k, \sigma^2)_{k=1, \dots, K}$. Nous traitons de l’estimation des paramètres de ce modèle dans la section 6.4. Après avoir rappelé les méthodes existantes dans la littérature, nous proposons une méthode utilisant les résultats démontrés au chapitre 5.

Enfin nous appliquons ces méthodes sur données simulées et données réelles dans les sections 6.5 et 6.6. Sur le jeu de données réelles, nous validons par un test du maximum de vraisemblance l’hypothèse d’existence d’une variabilité inter-types de stimuli au niveau des efficacités neuronales. D’autre part, nous mettons en évidence l’existence d’un phénomène de suppression à la répétition au niveau neuronal. En outre, notre méthode nous permet de quantifier ce phénomène.

6.2 Description du modèle physiologique

6.2.1 La composante “Balloon”

Cette composante du modèle hémodynamique global lie sa sortie -le signal BOLD- à son entrée -le débit sanguin local- par un système différentiel dans lequel interviennent deux variables d’état, -le volume du ballon et la concentration en déoxyhémoglobine. Au temps $t = 0$, le sujet (et donc le système) est supposé au repos.

Dans un premier temps, introduisons les notations suivantes.

- Les deux variables d’état sont respectivement notées V pour le volume sanguin et Q pour la concentration en déoxyhémoglobine. F_{in} et F_{out} représentent les débits sanguins respectivement entrant et sortant du compartiment sanguin (autrement appelé “ballon”).
- La sortie du système -le signal BOLD- est notée H .
- Notons f_{in} , f_{out} , q et v les variables d’état renormalisées par leur état au repos :

$$q(t) = \frac{Q(t)}{Q_0} \quad v(t) = \frac{V(t)}{V_0} \quad f_{in}(t) = \frac{F_{in}(t)}{F_0} \quad f_{out}(t) = \frac{F_{out}(t)}{F_0}$$

où Q_0 , V_0 et F_0 sont respectivement les concentration en déoxyhémoglobine, volume et débits sanguins à l’état de repos c’est-à-dire l’état initial. Par définition, à l’instant initial :

$$f_{in}(0) = 1 \quad v(0) = 1 \quad q(0) = 1$$

Le “*Balloon model*” peut désormais être décrit étape par étape.

- Le changement de volume du ballon est égal à la différence entre le flux sanguin entrant dans le compartiment (F_{in}) et le flux sanguin sortant du compartiment (F_{out}). Ainsi, on obtient l’équation :

$$\frac{dV(t)}{dt} = F_{in}(t) - F_{out}(t)$$

En considérant les variables renormalisées, cette équation s’écrit :

$$\tau_0 \frac{dv(t)}{dt} = f_{in}(t) - f_{out}(t) \tag{6.2.1}$$

La constante de temps τ_0 égale à V_0/F_0 est en fait le temps de transit moyen dans le ballon ou autrement-dit le temps moyen que met le compartiment à se remplir à l'état de repos.

Le flux sortant $f_{out}(t)$ dépend du volume $v(t)$. Cette fonction liant les deux quantités modélise la capacité du compartiment veineux à expulser le sang à plus grande vitesse quand il est gonflé. Buxton *et al.* [26] modélisent cette capacité par un unique paramètre α_W , dit paramètre de "Windkessel", issu de la théorie de Windkessel :

$$f_{out}(v) = v^{1/\alpha_W} \quad (6.2.2)$$

Le paramètre α_W vérifie : $1/\alpha_W = \beta + \gamma$, les paramètres γ , égal à 2 et $\beta > 1$ étant respectivement appelés "coefficient d'écoulement régulier" ("*laminar flow*") et "coefficient de diminution du volume de la réserve à haute pression" ("*diminished volume reserve at high pressure*"). La combinaison des équations (6.2.1) et (6.2.2) aboutit à

$$\tau_0 \frac{dv(t)}{dt} = f_{in}(t) - v(t)^{1/\alpha_W} \quad (6.2.3)$$

Remarque 16. Buxton *et al.* [24] ont proposé d'ajouter un terme de résistance aux changements de volume du ballon, en introduisant un terme de viscosité τ_{visc} :

$$f_{out} = v^{1/\alpha_W} + \frac{1}{\tau_{visc}} \frac{dv}{dt}$$

De plus, comme cette résistance à la variation de vitesse peut différer si le volume augmente ou diminue, τ_{visc} prend deux valeurs distinctes, conduisant à l'identité suivante :

$$\tau_0 \frac{dv(t)}{dt} = \begin{cases} f_{in}(t) - v(t)^{1/\alpha_W} - \frac{1}{\tau_{visc}^+} v & \text{si } \frac{dv}{dt} \geq 0 \\ f_{in}(t) - v(t)^{1/\alpha_W} - \frac{1}{\tau_{visc}^-} v & \text{si } \frac{dv}{dt} \leq 0 \end{cases}$$

- Les modifications de la concentration en déoxyhémoglobine $q(t)$ traduisent l'apport de déoxyhémoglobine dans le compartiment veineux moins ce qui en sort. Ainsi nous avons :

$$\tau_0 \frac{dq(t)}{dt} = f_{in}(t) \frac{E(f_{in}(t), E_0)}{E_0} - f_{out}(t) \frac{q(t)}{v(t)}$$

où $E(f_{in}(t), E_0)$ est la fraction de dioxygène extraite du sang entrant. Cette fraction dépend de l'oxygène délivré et donc du débit sanguin. Selon Buxton *et al.* [26], pour un grand nombre de conditions de transport, une approximation raisonnable est la suivante :

$$E(f_{in}, E_0) = 1 - (1 - E_0)^{1/f_{in}}$$

où E_0 est le taux d'extraction d'oxygène au repos.

On obtient ainsi, en combinant les deux dernières équations :

$$\tau_0 \frac{dq(t)}{dt} = f_{in}(t) \frac{1 - (1 - E_0)^{1/f_{in}}}{E_0} - v(t)^{1/\alpha_W} \frac{q(t)}{v(t)} \quad (6.2.4)$$

- L'expression du signal BOLD en fonction des variables d'état est plus complexe. Elle est largement détaillée dans l'article de Buxton *et al.* [27] ou dans la thèse de Deneux [37]. Plusieurs expressions ont été proposées pour le signal BOLD. Ici, nous considérons celle proposée par Buxton *et al.* dans laquelle le signal s'écrit comme suit :

$$H(t) = V_0 [k_1(1 - q(t)) - k_2(1 - v(t))].$$

Le premier terme de la somme représente le signal extra-vasculaire, le deuxième représente le signal intra-vasculaire. V_0 est le volume de sang par unité de volume au repos. k_1 et k_2 sont

deux constantes dépendant du taux d'extraction d'oxygène au repos E_0 , du scanner IRMF et des conditions d'acquisition. Les valeurs estimées par Obata *et al.* [93] pour un champ magnétique de 1.5 Tesla avec un temps d'échantillonnage de 40 ms et $E_0 = 0.4$ sont $k_1 = 3.4$ et $k_2 = 1.0$. Dans le cas d'un scanner 3 Tesla, ces valeurs n'ont pas été calculées. Deneux [37] propose d'imposer la relation $k_1/(k_1 + k_2) = 1/10$ (au lieu de $k_1/(k_1 + k_2) = 1/4.4$ dans le cas 1.5 Tesla) et d'introduire un nouveau paramètre $b = V_0(k_1 + k_2)$ de façon à ce que le signal devienne :

$$H(t) = b[0.9(1 - q(t)) - 0.1(1 - v(t))] \quad (6.2.5)$$

Finalement, les équations (6.2.3), (6.2.4) et (6.2.5) constituent le “*Balloon model*”. Elles dépendent de quatre paramètres physiologiques : E_0 , la fraction de dioxygène extraite au repos, τ_0 le temps de transit moyen, α_W le paramètre de “Windkessel” et b . Décrivons maintenant l'extension proposée par Friston *et al.* ([50] liant l'activité neuronale au rCBF.

6.2.2 La composante “rCBF”

Il est souvent admis que, dans des conditions normales, débit sanguin et activité neuronale sont linéairement liés. Cette hypothèse a été démontrée directement sur des animaux par des études combinant des enregistrements par électrodes multiples et des mesures de flux par laser Doppler [87] et indirectement chez l'Homme [91]. Sous cette hypothèse, Friston *et al.* ([50]) proposent d'utiliser le modèle le plus simple :

$$\frac{df_{in}(t)}{dt} = s(t) \quad (6.2.6)$$

où s est un signal engendrant le flux. Ce signal s est supposé être la somme de nombreuses sous-composantes neuronales et est généré par une activité neuronale $an(t)$.

$$\frac{ds(t)}{dt} = an(t) - \frac{s(t)}{\tau_s} - \frac{f_{in}(t) - 1}{\tau_f} \quad (6.2.7)$$

où τ_s and τ_f sont deux paramètres physiologiques déterminant la dynamique de cette composante du modèle hémodynamique. τ_s est une constante de temps d'élimination du signal. τ_f est une constante de temps d'auto-régulation du système par *feed-back* à partir du débit sanguin. Le système étant initialement au repos, on a $s(0) = 0$.

A propos de l'activité neuronale Chaque stimulus engendre une activité neuronale que nous devons modéliser. En règle générale, les stimuli sont modélisés par une fonction-crêneau $u(t)$ avec $u(t) = 1$ si t est un instant de stimulation. Cette fonction-crêneau est pondérée par un terme “efficacité neuronale” –noté α – qui traduit la capacité des neurones à engendrer une augmentation du signal $s(t)$. Ainsi,

$$an(t) = \alpha u(t).$$

Considérer un modèle dans lequel tous les types stimuli auraient la même efficacité neuronale semble inapproprié au vu des conclusions faites dans le cadre du modèle linéaire. Nous considérerons donc un modèle permettant de modéliser une variabilité de l'efficacité neuronale. Plus précisément, supposons que le paradigme comprenne K types de stimuli ayant chacun J_k occurrences. Alors pour tout $k = 1, \dots, K$ et tout $j = 1, \dots, J_k$, soit $u_{k,j}(t)$ la fonction-crêneau représentant le décours temporel du j -ième stimulus de type k et $\alpha_{k,j}$ son efficacité neuronale. L'activité neuronale s'écrit :

$$an(t) = \alpha \cdot u(t) = \sum_{k=1}^K \sum_{j=1}^{J_k} \alpha_{k,j} u_{k,j}(t) \quad (6.2.8)$$

Remarque 17. *Riera et al. [105] suggèrent une autre approche dans laquelle la totalité de la séquence temporelle du signal neuronal est estimée. Il est alors modélisé par une suite de gaussiennes centrées en des instants $\tau_l = TR/2$, où TR est le temps entre deux instants de mesure, de "largeur" $\rho_l = TR/6$ et d'amplitude β_l inconnue :*

$$an(t) = \sum_l \beta_l \frac{e^{-(t-\tau_l)^2}}{2\rho_l^2}$$

6.2.3 Le modèle hémodynamique global

Finalement, le système dynamique – noté \mathcal{B} – définissant le modèle global s'écrit avec les équations suivantes :

$$\begin{cases} ds(t)/dt = \alpha \cdot u(t) - \frac{s(t)}{\tau_s} - \frac{f_{in}(t)-1}{\tau_f} \\ df_{in}(t)/dt = s(t) \\ dv(t)/dt = \frac{1}{\tau_0} (f_{in}(t) - v(t)^{1/\alpha_w}) \\ dq(t)/dt = \frac{1}{\tau_0} \left(\frac{f_{in}(t)}{E_0} (1 - (1 - E_0)^{1/f_{in}}) - q(t)v(t)^{1/\alpha_w - 1} \right) \end{cases} \quad (\mathcal{B})$$

avec les conditions initiales suivantes :

$$\begin{cases} s(0) = 0 \\ f_{in}(0) = 1 \\ v(0) = 1 \\ q(0) = 1 \end{cases}$$

Le signal BOLD est alors obtenu par l'équation (6.2.9) :

$$H(t, \alpha, \mathbf{a}) = b [0.9(1 - q(t)) - 0.1(1 - v(t))] \quad (6.2.9)$$

où

$$\mathbf{a} = (a_1, a_2, a_3, a_4, a_5, a_6) = (\tau_s, \tau_f, \tau_0, \alpha_w, E_0, b)$$

est le vecteur des six constantes physiologiques paramétrant ce modèle. Le modèle dépend aussi des efficacités neuronales α . Il est résumé sur la figure 6.1.

Modèle linéaire versus modèle hémodynamique Dans le "*Balloon model*", les paramètres liant les entrées à l'hémodynamique, i.e. les efficacités neuronales α , sont propres aux types des stimuli ou stimuli alors que les autres paramètres \mathbf{a} , purement physiologiques ceux-là, sont constants pour chacun des voxels ou petit groupe de voxels ("*cluster*"). Cette modélisation correspond à l'hypothèse que nous avons faite sous le modèle linéaire, c'est-à-dire une forme constante de la réponse hémodynamique HRF alors que les amplitudes étaient variables. Ainsi sur ce point, les deux modèles considérés sont similaires.

6.3 Propriétés du modèle physiologique

Avant d'aborder le problème de l'estimation des paramètres au chapitre 6.4, l'étude des propriétés du modèle hémodynamique est une étape essentielle qui fait l'objet de ce paragraphe. Les paramètres de ce modèle conditionnent non seulement la réponse hémodynamique à un seul stimulus mais aussi l'effet de la réponse à un stimulus sur la réponse à un autre stimulus.

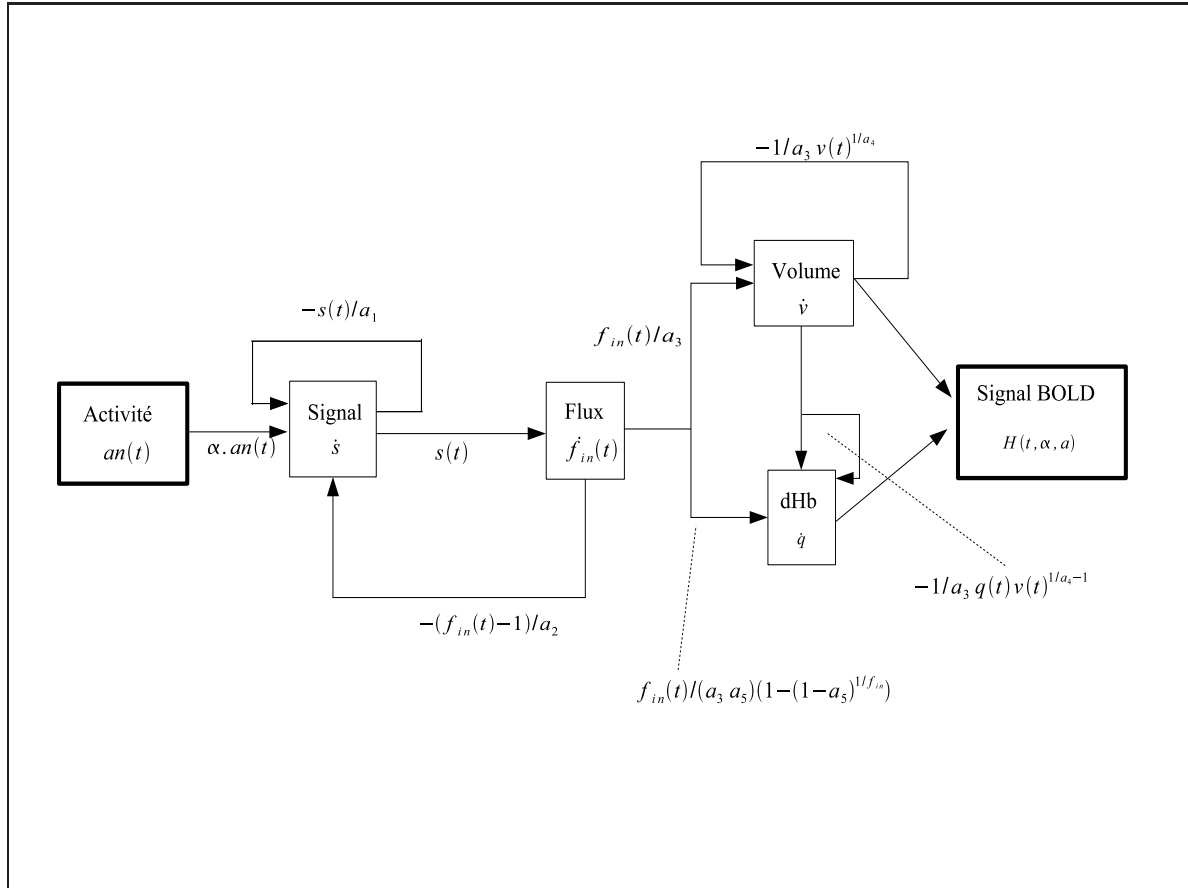


FIG. 6.1 – Schéma récapitulatif du modèle hémodynamique global

Dans un premier temps, nous cherchons à comprendre le rôle de chacun des paramètres physiologiques $\mathbf{a} = (\tau_s, \tau_f, \tau_0, \alpha_W, E_0, b)$. Dans un second temps, nous étudions l'influence de l'efficacité neuronale α sur le signal hémodynamique. Enfin, les non-linéarités du modèle en les amplitudes α sont mises en évidence.

6.3.1 Influence des paramètres sur le signal BOLD

Nous cherchons à déterminer le rôle de chacun des paramètres sur le signal hémodynamique, solution du système dynamique (\mathcal{B}). Pour cela, comme fait par Friston *et al.* ([50]), nous observons le comportement de la réponse BOLD à un unique stimulus, en fonction des paramètres physiologique \mathbf{a} et de l'efficacité neuronale α .

En pratique, un ensemble de paramètres de référence est choisi. Afin de conserver un sens physiologique aux paramètres, leurs valeurs ne peuvent être prises arbitrairement grandes ou petites mais doivent demeurer dans un domaine “réaliste”. Nous prenons les valeurs fournies dans [50] respectivement égales à 1.54s pour τ_s , 2.5s pour τ_f , 0.98s pour τ_0 , 0.33 pour α_W , 34% pour E_0 et 0.16 pour b . α est prise égale à 0.5. A ces paramètres correspond une unique solution du système différentiel (\mathcal{B}) appelée “réponse BOLD de référence”. Afin de mettre en évidence l'influence de chaque paramètre sur la réponse BOLD, ceux-ci sont successivement modifiés et la réponse BOLD obtenue est comparée à la réponse BOLD de référence. Ainsi, α est multipliée par 2, τ_s par 0.5, τ_f par 2, τ_0 par 2, α_W par 2 et E_0 par 2.3.

Les résultats sont représentés sur la figure 6.2. Sur chaque graphique, nous représentons la réponse

BOLD de référence (tirets $--$) ainsi que celle obtenue après modification du paramètre (trait plein).

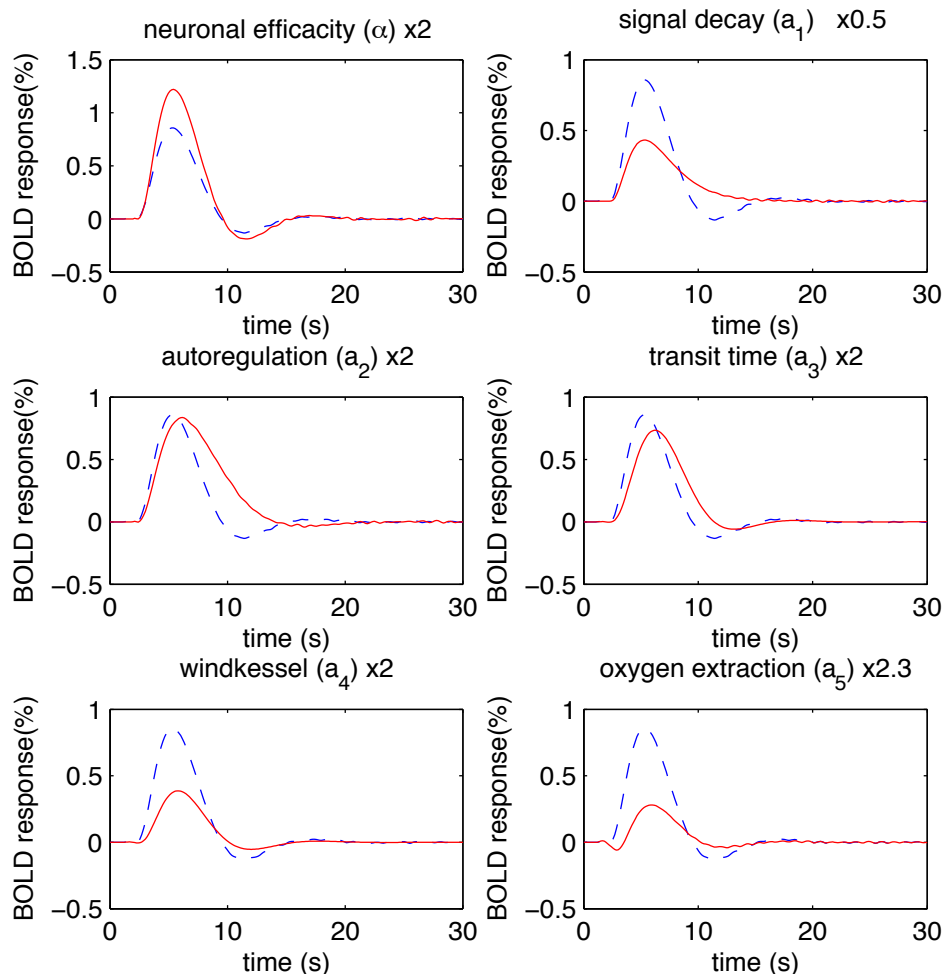


FIG. 6.2 – Influence des paramètres physiologiques (\mathbf{a}) sur la réponse BOLD à un stimulus. Les facteurs multiplicatifs sont ceux appliqués aux paramètres. La réponse de référence est en trait discontinu ($--$)

Les commentaires inspirés de ceux de Friston *et al.* ([50]) sont les suivants :

1. Influence de α (neuronal efficacy) : Ce paramètre représente l'efficacité neuronale ou plutôt la capacité des neurones à entraîner une augmentation du débit sanguin local. On constate qu'il module l'amplitude de la réponse BOLD sans en modifier la forme. Notons qu'une multiplication par 2 de ce facteur n'implique pas exactement une multiplication par 2 de l'amplitude de la réponse. L'influence de ce paramètre sera plus largement étudiée dans la partie 6.3.3.
2. Influence de $a_1 = \tau_s$ (signal decay) : Ce paramètre conditionne l'élimination du signal BOLD. Une augmentation de ce paramètre supprime l'*undershoot* final.
3. Influence de $a_2 = \tau_f$ (auto-regulation) : Ce paramètre est la constante de régulation par *feedback* du système. Sa nature physiologique demeure mal définie (voir [50]). On peut cependant constater qu'une augmentation de ce paramètre rend la réponse BOLD plus lisse et atténuée –voire élimine– l'*undershoot*.

4. Influence de $a_3 = \tau_0$ (transit time) : Ce paramètre physiologique est prépondérant puisqu'il détermine la dynamique du système. Rappelons que ce paramètre représente le temps moyen de remplissage du ballon. Une augmentation de ce temps moyen ralentit la dynamique du signal BOLD par rapport aux changements du débit sanguin local. La réponse obtenue est assez similaire à la réponse de référence mais s'exprime plus lentement.
5. Influence de $a_4 = \alpha_W$ (Windkessel parameter) : L'effet de ce paramètre sur la forme d'une seule réponse n'est pas très marqué bien qu'il semble avoir une influence sur l'amplitude de la réponse. De par sa place dans les équations du système dynamique (\mathcal{B}), on s'attend à ce qu'il ait tendance à représenter le degré de non-linéarité du système. Cet aspect sera détaillé dans le paragraphe 6.3.3.
6. Influence de $a_5 = E_0$ (oxygen extraction) : L'augmentation de la valeur de ce paramètre peut entraîner d'importants changements sur la forme de la réponse BOLD notamment sur le *dip* initial mais aussi sur son amplitude.
7. Influence de b : L'influence de ce paramètre n'est pas été illustrée graphiquement. Cependant, il est clair que b est un paramètre d'échelle, d'après l'expression (6.2.9) du signal BOLD.

Ainsi, il est assez simple de comprendre l'influence de chacun des paramètres sur la réponse BOLD à un stimulus. Cependant, dès lors que l'on veut considérer l'action conjointe de plusieurs paramètres, les choses deviennent plus compliquées. En effet, des paramètres différents peuvent influencer sur une même caractéristique : ainsi, τ_f et τ_s caractérisent tous deux l'*undershoot* final. De nombreux paramètres modulent l'amplitude de la réponse BOLD. En outre, un seul et même paramètre peut avoir des effets multiples – par exemple E_0 influe sur la forme du *dip* initial ainsi que sur l'amplitude du signal.

Le “Balloon model”, un modèle identifiable ? Nous venons d'étudier l'influence de chaque paramètre sur une réponse BOLD. Cependant, nous avons remarqué que plusieurs paramètres définissent l'amplitude de la réponse BOLD notamment l'efficacité neuronale α , ce paramètre pouvant ainsi être vu comme un facteur d'échelle. D'autre part, d'après l'expression (6.2.9)

$$H(t) = b [0.9(1 - q(t)) - 0.1(1 - v(t))]$$

il paraît clairement que b est un facteur d'échelle du signal. Ainsi se pose le problème de l'identifiabilité du modèle : il est aisé de trouver deux couples de paramètres (α, b) distincts générant des réponses BOLD semblables. Nous illustrons notre propos avec la figure 6.3. Sur cette figure, nous avons représenté les réponses BOLD à un stimulus pour des paramètres respectifs $(\alpha, b) = (0.5, 0.03)$ et $(\alpha, b) = (0.25, 0.0546)$, les autres étant fixés. Nous constatons que ces réponses sont quasiment confondues. Ainsi, le modèle physiologique présenté en tant que tel, n'est pas identifiable. En pratique, sur données réelles, il sera envisageable de fixer le paramètre b égal à une valeur courante issue de la littérature. Ceci ne rend pas le modèle identifiable pour autant. Dans le cadre d'une procédure d'estimation des paramètres, nous envisagerons donc de restreindre les intervalles d'estimation ou de fixer d'autres paramètres. Ce point reste un des problèmes majeurs des modèles physiologiques.

6.3.2 Influence de l'amplitude α

Dans cette partie, nous étudions l'influence du paramètre α –efficacité neuronale– sur l'amplitude de la réponse hémodynamique.

Bien évidemment, plus la capacité des neurones à entraîner une augmentation du débit sanguin locale est grande et plus la réponse hémodynamique est grande. Cependant, comme déjà évoqué dans la partie 6.3.1, ce rapport entre efficacité neuronale et amplitude hémodynamique n'est pas linéaire. Nous allons donc chercher à déterminer la relation liant ces deux paramètres. Une façon simple de

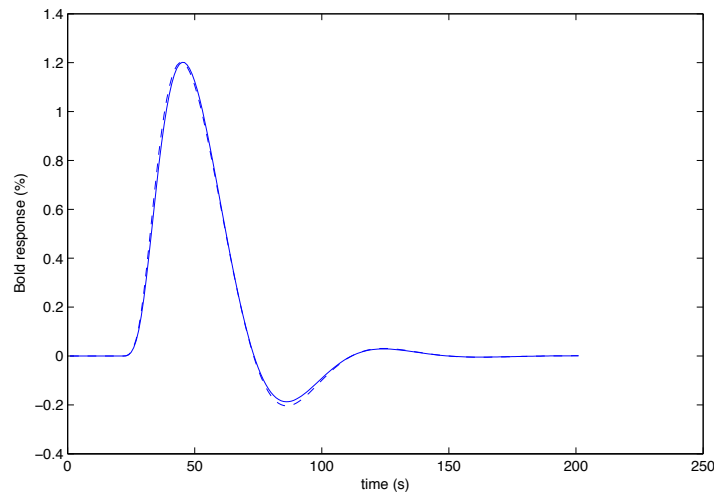


FIG. 6.3 – Mise en évidence de la non-identifiabilité du “*Balloon model*” : réponse BOLD pour un stimulus avec paramètres respectifs $(\alpha, b) = (0.5, 0.03)$ en trait plein et $(\alpha, b) = (0.25, 0.0546)$ en trait discontinu.

faire consiste à considérer un paradigme ayant un unique stimulus d’efficacité neuronale α , à faire varier ce paramètre α tout en fixant les paramètres physiologiques à $\mathbf{a} = (1.54, 2.46, 0.98, 0.33, 0.34, 0.16)$.

Nous faisons varier l’efficacité neuronale α entre 0.5 et 10. La valeur maximale de la réponse BOLD obtenue par le “*Balloon model*” est représentée sur la figure 6.4 en fonction de l’efficacité neuronale.

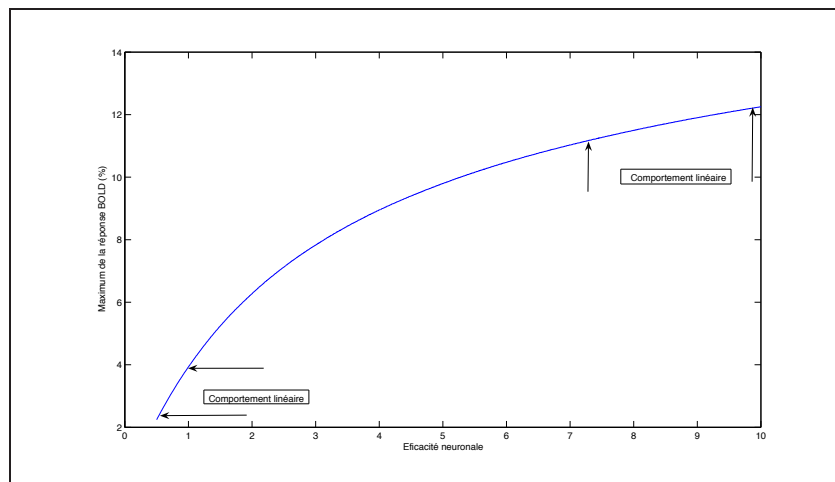


FIG. 6.4 – Influence de α sur l’amplitude hémodynamique. La courbe représente la valeur maximale de la réponse hémodynamique obtenue par le “*Balloon model*” en fonction de l’efficacité neuronale.

Comme attendu, nous constatons qu’efficacité neuronale et amplitude hémodynamique ne sont pas linéairement liées. Notons surtout que, pour de grande valeurs de l’efficacité neuronale, l’amplitude hémodynamique croît moins vite. On constate donc un phénomène de *saturation hémodynamique*. Autrement-dit, l’environnement physiologique semble limiter l’amplitude de la réponse hémodynamique.

Modèle linéaire versus “*Balloon model*” Le phénomène de saturation hémodynamique mis en évidence n’est pas sans conséquence. En effet, considérons deux stimuli consécutifs n’ayant pas d’interactions entre eux au niveau neuronal. Les réponses BOLD générées par ces stimuli peuvent avoir une interaction dans la mesure où leur amplitude conjointe est limitée par le phénomène de saturation hémodynamique. Ainsi le “*Balloon model*” permet de distinguer les interactions neuronales des interactions hémodynamiques, ce qui n’est pas le cas du modèle linéaire.

6.3.3 Mise en évidence de la non-additivité du modèle

Le modèle physiologique a été introduit afin de pouvoir prendre en compte les non-linéarités du signal BOLD. Considérons donc la fonction qui, à un vecteur d’efficacités neuronales α , associe le signal BOLD $H \circ x(t, \alpha, \mathbf{a})$ solution du système dynamique (\mathcal{B})

$$\begin{aligned} \mathbb{R}^J &\longmapsto \mathcal{C}([0, +\infty[\times \mathbb{R}^J \times \mathbb{R}^6, \mathbb{R}) \\ \alpha &\longmapsto H \circ x \end{aligned}$$

Dans cette partie nous mettons en évidence le caractère non-additif de cette fonction.

L’aspect non linéaire le plus caractéristique des signaux BOLD est appelé “*hemodynamic refractoriness*”. Plus précisément, il a été mis en évidence que les interactions entre stimuli peuvent aboutir à la suppression ou à l’augmentation de l’amplitude ou du temps de latence de la réponse d’un stimulus lorsque celui ci est précédé par d’autres stimuli. Autrement-dit, si deux stimuli sont consécutifs rapprochés dans le temps, la réponse hémodynamique au second stimuli peut être plus tardive ou d’amplitude moindre (*suppression à la répétition*) que si le stimulus avait eu lieu seul. Le “*Balloon model*” permet de prendre en compte ce phénomène.

Illustrons ceci par un exemple simple. Considérons trois séries de deux stimuli consécutifs, séries largement espacées dans le temps. Au sein de chaque série, les stimuli sont respectivement espacés de $\tau = 1$, $\tau = 3$ et $\tau = 10$ secondes. Le signal d’entrée $an(t)$ est représenté dans le graphique du haut de la figure 6.5.

Nous observons la réponse BOLD **au deuxième stimulus** uniquement puisque seule cette réponse est susceptible d’être différente dans le cas du “*Balloon model*”. Cette réponse est obtenue en soustrayant la réponse au premier stimulus à la réponse globale générée par les deux stimuli. Nous comparons cette réponse à celle obtenue sous le modèle linéaire. Sur la figure 6.5, la courbe en trait discontinu (—) est la réponse obtenue sous le modèle linéaire, celle en trait continu (—) est la réponse obtenue avec le modèle hémodynamique.

D’autre part, selon [50], le paramètre de Windkessel (a_4) contrôle le degré de non-linéarité du modèle. Il est donc intéressant d’observer l’influence de ce paramètre sur la forme de la réponse au deuxième stimulus. Ainsi les graphiques du centre et du bas de la figure 6.5 représentent les réponses BOLD au deuxième stimulus obtenues pour a_4 valant respectivement 0.4 et 0.18. Les autres paramètres sont égaux aux valeurs de référence.

Résultats

- *Suppression à la répétition* : Comme nous pouvions nous y attendre, pour deux stimuli rapprochés dans le temps ($\tau = 1$ secondes), la réponse obtenue par le modèle hémodynamique est sensiblement atténuée par rapport à la réponse linéaire. Cette différence est moins flagrante quand on augmente le temps inter-stimuli ($\tau = 3$ secondes). Cette tendance disparaît pour $\tau = 10$ secondes.
- *Influence de $a_4 = \alpha_W$* : S’agissant de l’influence du paramètre a_4 sur la non-linéarité du modèle, les choses sont moins évidentes. Le tableau 6.1 contient les pourcentages de variation de la réponse au deuxième stimulus sous le “*Balloon model*” par rapport au modèle linéaire :

$$\frac{\max(\text{réponse linéaire}) - \max(\text{réponse BOLD})}{\max(\text{réponse linéaire})} \times 100$$

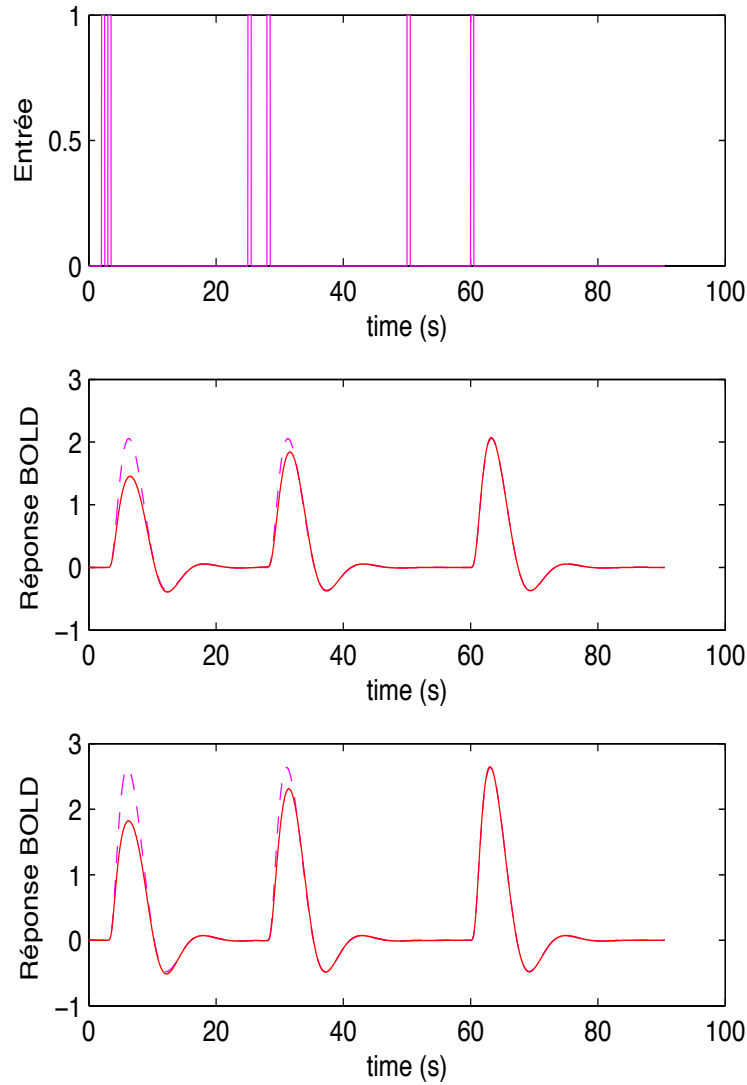


FIG. 6.5 – Mise en évidence de la non-additivité du modèle hémodynamique. Graphe du haut, entrée $an(t)$. Les stimuli sont espacés de 0.5, 2 et 10 sec. Graphes du milieu et en bas : réponse BOLD **au deuxième stimulus** avec le modèle hémodynamique en trait continu, avec le modèle linéaire en trait discontinu, pour différents a_4 : figure du milieu, $a_4 = 0.4$, figure du bas, $a_4 = 0.18$.

	$\tau = 1$	$\tau = 3$
$a_4 = 0.40$	29.98 %	15.94%
$a_4 = 0.18$	31.25%	17.56%

TAB. 6.1 – Taux de variation de la réponse BOLD au deuxième stimulus issue du “*Balloon model*” par rapport à celle issue du modèle linéaire

en fonction de a_4 et τ le temps inter-stimuli. On en conclut que, effectivement, quand a_4 diminue (s'éloigne de 1), le phénomène de non-linéarité est amplifié. Cependant, cet effet n'est pas très marqué. Nous ne pouvons faire varier plus amplement ce paramètre puisqu'il est important de rester dans un domaine de valeurs réaliste.

- *Augmentation du temps de latence* : L'augmentation du temps de latence peut aussi être observé sur cet exemple. Pour cela nous faisons un agrandissement du graphique du bas de la figure 6.5 pour $\tau = 3$ secondes (figure 6.6). On constate que le pic de la réponse sous le “*Balloon model*” est retardé par rapport au pic sous le modèle linéaire.

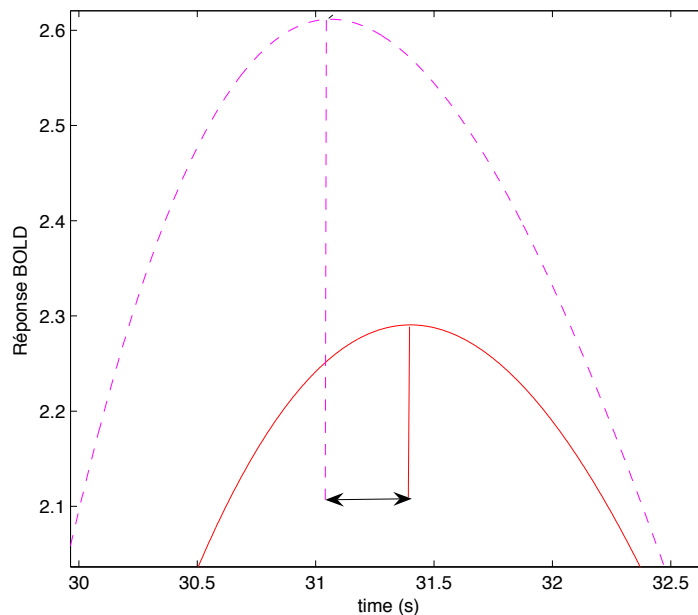


FIG. 6.6 – Mise en évidence de l'augmentation du temps de latence : zoom de la figure 6.5 pour $a_4 = 0.18$ et $\tau = 3s$. La réponse BOLD sous le “*Balloon model*” (–) atteint son maximum plus tard que sous le modèle linéaire (—).

6.3.4 Conclusion

Ainsi cette partie nous a permis d'identifier le rôle de chaque paramètre dans le “*Balloon model*” et d'explorer les propriétés du modèle. Les informations majeures que nous en retirons sont les suivantes.

1. D'une part, le “*Balloon model*” étant un modèle biophysique, chaque paramètre physiologique a un sens biologique, ces paramètres devront donc être estimés sous contraintes de façon à demeurer dans un intervalle de validité biologique.

2. D'autre part, ce modèle est clairement non identifiable. Ceci posera certainement des problèmes au moment de l'estimation de ses paramètres.
3. Enfin, le “*Balloon model*” devrait permettre de modéliser des effets non-linéaires négligés dans la modélisation linéaire.

6.4 Estimation paramétrique

Dans cette partie, nous nous intéressons à l'estimation des paramètres du modèle hémodynamique (paramètres physiologiques \mathbf{a} et efficacités neuronales α), décrit dans la section 6.2.

Soient (t_1, \dots, t_n) les instants d'observation et $(y(t_1), \dots, y(t_n))$ les observations du signal BOLD en ces instants. Considérons le modèle de régression suivant :

$$\begin{cases} y(t_i) &= H \circ x(t_i, \boldsymbol{\alpha}, \mathbf{a}) + \varepsilon_i & \text{pour } i = 1 \dots n. \\ \varepsilon_i &\sim \text{i.i.d } \mathcal{N}(0, \sigma^2) \end{cases}$$

où $H : \mathbb{R}^4 \rightarrow \mathbb{R}$ est définie en (6.2.5) et $x : [0, +\infty[\times \mathbb{R}^5 \times \mathbb{R}^J \rightarrow \mathbb{R}^4$ est solution du système d'équations différentielles (\mathcal{B}).

Remarque 18. *Ce modèle est parfois appelé déterministe dans la littérature. Les modèles dits déterministes font référence aux modèles dont la dynamique est entièrement décrite par l'état du système. Les effets aléatoires n'entrent qu'au point où le système est observé, instant de mesure. Ces modèles déterministes s'opposent aux modèles dits stochastiques, c'est-à-dire où un processus stochastique intervient au niveau des évolutions des variables d'états elles-mêmes (voir [105]).*

Plusieurs méthodes d'estimation des paramètres ont été développées dans la littérature dans les versions déterministe et stochastique du modèle. Nous donnons un rapide aperçu de ces méthodes.

Nous proposons ensuite une nouvelle méthode d'estimation par maximum de vraisemblance des paramètres en utilisant l'algorithme SAEM-MCMC adapté aux équations différentielles ordinaires développé dans le chapitre 5. Cette méthode est illustrée sur données simulées et données réelles.

6.4.1 Etat de l'art

Estimation des paramètres dans la version déterministe du modèle

Estimation par noyaux de Volterra (voir Friston *et al.* [50]) : Dans un premier temps, les paramètres \mathbf{a} ont été estimés par Friston *et al.* [50]. Les sujets devaient écouter des noms concrets monosyllabiques ou bi-syllabiques. Une fois les données recueillies, lissées et pré-traitées, les paramètres physiologiques \mathbf{a} et l'amplitude α ont été estimés en utilisant les noyaux de Volterra, cette méthode consistant en un développement de Taylor du système différentiel. Les estimations des paramètres ont ensuite été obtenues par un critère des moindres carrés entre les noyaux de Volterra empiriques et ceux issus du modèle hémodynamique. A partir des valeurs issues de différents voxels, les auteurs ont obtenu des histogrammes des paramètres, montrant des distributions gaussiennes avec pour moyennes les valeurs suivantes :

$$a_1 = 1.54s \quad a_2 = 2.5s \quad a_3 = 0.98s \quad a_4 = 0.33 \quad a_5 = 34\%$$

Dans ce modèle, une unique efficacité neuronale est estimée. La valeur moyenne trouvée est

$$\alpha = 0.54.$$

Pour plus de détails, nous invitons le lecteur à se reporter à l'article d'origine de [50]. Cette première estimation sert de base pour définir les distributions a priori dans le cadre de l'estimation bayésienne suivante.

Estimation Bayésienne (voit Friston [49]) : Dans le cas d’un modèle à K efficacités neuronales distinctes, Friston ([49]) propose une méthode bayésienne permettant d’estimer les lois conditionnelles des paramètres du modèle hémodynamique. Cette procédure est présentée sous l’intitulé “*EM implementation of a Gauss Newton search for the maximum of the conditional density*”. Notons $\eta = (\mathbf{a}, \boldsymbol{\alpha})$ l’ensemble des paramètres à estimer et écrivons le modèle sous la forme : $y(t_i) = H \circ x(t_i, \eta) + \varepsilon_i$ pour $i = 1, \dots, n$. Supposons les paramètres η de distribution a priori gaussienne. L’inférence bayésienne est basée sur la densité conditionnelle des paramètres sachant les observations $p(\eta|y)$. En supposant cette densité gaussienne, le problème se réduit à trouver ses deux premiers moments, la moyenne conditionnelle $m_{\eta|y}$ et la variance-covariance $C_{\eta|y}$. Ceci est fait en linéarisant en les paramètres η la fonction de régression $H \circ x(t, \eta)$ au voisinage de l’estimateur courant $m_{\eta|y}^{(i)}$. Cette linéarisation nécessite le calcul de la matrice Jacobienne $J = \frac{dH \circ x(t, m_{\eta|y})}{d\eta}$. Pour cela, les auteurs proposent d’utiliser un développement à l’ordre 2 du système dynamique lui-même. En supposant la probabilité conditionnelle gaussienne, la linéarisation de la fonction de régression permet d’en avoir une expression explicite. Les auteurs proposent alors un algorithme EM détaillé dans [49].

Cet algorithme repose sur des distributions a priori pour les paramètres. Leur choix conditionne la robustesse de la procédure et sa vitesse de convergence. Les valeurs initiales ne sont pas prises arbitrairement mais viennent de la première estimation par noyau de Volterra dont la méthodologie a été brièvement exposée au paragraphe précédent (voir [50] pour plus de détails). Bien que rapide en terme de temps de calcul (car reposant sur la linéarisation du système), cet algorithme a le défaut de dépendre des distributions a priori. De plus, dans le cas d’efficacités neuronales variables, le nombre de paramètres augmente rapidement, pouvant entraîner une convergence lente de l’algorithme.

Estimation par moindres carrés (voir Deneux [37]) : Deneux propose une estimation des paramètres par moindres carrés en utilisant l’algorithme itératif de Levenberg Marquardt [84] implémenté dans la fonction `lsqcurvefit` du logiciel **Matlab**. Cet algorithme nécessite le calcul à chaque itération de la Jacobienne du système $\frac{\partial H \circ x}{\partial \eta}$ par rapport aux paramètres. Deneux montre que la Jacobienne est elle-même solution d’un système différentiel. Cette méthode évite une linéarisation du système différentiel. Comme on pouvait s’y attendre, Deneux souligne dans sa thèse des problèmes d’identifiabilité du modèle. Enfin, Deneux estime une unique efficacité neuronale.

Estimation des paramètres dans la version stochastique du modèle

Notons $x(t) = (s(t), f_{in}(t), v(t), q(t))^T$ les quatre variables d’état et comme précédemment $\eta = (\boldsymbol{\alpha}, \mathbf{a})$. Le système (\mathcal{B}) est parfois considéré dans sa version stochastique, conduisant à l’écriture du modèle statistique suivant :

$$\begin{cases} dx(t) &= F(x(t), u(t), \eta)dt + dW(t) \\ y(t) &= H \circ x(t, \eta) + \sigma^2 \varepsilon(t) \end{cases} \quad (6.4.10)$$

La première équation est la version stochastique du système différentiel issu du “*Balloon model*” (\mathcal{B}). $dW(t)$ est un mouvement brownien représentant le bruit d’évolution.

Riera *et al.* ([105]) considèrent un modèle sans bruit d’observation ($\sigma^2 = 0$) et estiment les paramètres par une méthode de filtration basée sur le schéma de linéarisation locale (LL) [18]. Dans le cas où les deux bruits sont inclus dans le modèle (bruit d’évolution et bruit de mesure), Deneux propose d’utiliser une version étendue du filtre de Kalman qui consiste en un développement de Taylor au premier ordre du système différentiel à chaque itération de l’algorithme.

Remarque 19. *En l’état actuel des choses, rien ne justifie de préférer le modèle déterministe au modèle stochastique. Dans cette thèse, nous commençons par considérer pas la version déterministe du “Balloon model”. L’approche stochastique pourrait faire l’objet d’un travail futur.*

6.4.2 Estimation par SAEM-MCMC-ODE

Nous proposons une approche différente applicable sur un modèle à efficacités neuronales variables. Nous nous plaçons sur la version déterministe du modèle et développons ici une méthode d'estimation des paramètres par maximum de vraisemblance. Pour cela, nous décrivons d'abord en détail le modèle statistique considéré puis la méthodologie utilisée afin d'estimer d'une part les paramètres physiologiques et d'autre part les efficacités neurales.

Estimation des paramètres physiologiques (Etape 1)

Considérons un paradigme contenant K types de stimuli, ayant chacun J_k occurrences. Soit $\alpha_{k,j}$ l'efficacité neuronale des stimuli de type k . Notons $\boldsymbol{\alpha} = (\alpha_{k,j})_{k=1,\dots,K,j=1,\dots,J_k}$ et considérons le modèle de régression suivant :

$$\begin{cases} y_i &= H(t_i, \boldsymbol{\alpha}, \mathbf{a}) + \varepsilon_i & \text{pour } i = 1 \dots n \\ \varepsilon_i &\sim \text{i.i.d } \mathcal{N}(0, \sigma^2) \end{cases}$$

L'enjeu de cette partie est d'estimer les paramètres physiologiques \mathbf{a} et $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_K)$. La non-identifiabilité du modèle en les paramètres (mise en évidence au chapitre 6.3 de cette partie) conduit à réduire l'ensemble des paramètres à estimer. Ainsi nous supposons d'une part une unique source de variabilité $\alpha_{k,j} = \alpha_k$ pour tout $j = 1, \dots, J_k$ (la variabilité inter-occurrence est éliminée). D'autre part, comme cela est fait dans [105], il est envisageable de fixer les paramètres \mathbf{a}_4 et \mathbf{a}_5 (égaux à 0.33 et 0.34 respectivement). Les autres paramètres $(\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$ doivent être estimés en restant dans un intervalle "physiologiquement plausible". Nous proposons un estimateur par minimisation du critère des moindres carrés pondéré :

$$\Lambda_\lambda(\mathbf{a}, \boldsymbol{\mu}) = \sum_{i=1}^N [y(t_i) - H(t_i, \boldsymbol{\mu}, \mathbf{a})]^2 + \sum_{\ell=1}^3 \lambda_\ell (\mathbf{a}_\ell - \mathbf{a}_{\ell,0})^2 \quad (6.4.11)$$

où \mathbf{a}_0 est une valeur de référence à définir. Les valeurs des constantes λ_ℓ seront précisées lors de l'illustration numérique. Le critère (6.4.11) est minimisé en utilisant la méthode numérique du simplexe permettant de résoudre des problèmes d'optimisation en grande dimension (ici $K+3$). Cette méthode est implémentée sous **matlab**, dans la fonction `fminsearch`. Nous obtenons alors $\hat{\mathbf{a}}$ et $\hat{\boldsymbol{\mu}}$ estimateurs de \mathbf{a} et $\boldsymbol{\mu}$. Dorénavant, les paramètres physiologiques seront pris égaux à ces valeurs. Cependant, afin d'alléger les notations, nous continuerons à noter simplement \mathbf{a} ces paramètres.

Estimation par maximum de vraisemblance des efficacités neuronales $\boldsymbol{\alpha}$ (Etape 2)

De nombreux modèles supposent les efficacités neuronales toutes identiques (voir [37] par exemple). Nous proposons ici une approche différente dans laquelle les efficacités sont des réalisations d'une variable aléatoire gaussienne dont nous souhaitons estimer les paramètres. Une fois ces paramètres estimés, nous serons alors capable de comparer les groupes de stimuli entre eux, d'analyser leur variabilité, etc.

Plus précisément, considérons le modèle suivant. Soit $\alpha_{k,j}$ l'efficacité neuronale du j -ième stimulus du groupe k :

$$\alpha_{k,j} \sim \mathcal{N}(\mu_k, \omega_k)$$

où k est le numéro du groupe. Posons $\boldsymbol{\alpha}$ le vecteur des efficacités ordonnées dans l'ordre d'occurrence. Définissons de la même façon le vecteur des moyennes \mathbf{m} tel que $m_j = \mu_k$ si le j -ième stimulus est de type k . Soit $\boldsymbol{\omega}$ la matrice diagonale des variances correspondantes. Définissons le modèle statistique suivant (noté \mathcal{B}_1) :

$$\begin{cases} y(t_i) &= H \circ x(t_i, \boldsymbol{\alpha}, \mathbf{a}) + \varepsilon_i & \text{pour } i = 1 \dots n \\ \boldsymbol{\alpha} &\sim \mathcal{N}(\mathbf{m}, \boldsymbol{\omega}) \\ \varepsilon_i &\sim \text{i.i.d } \mathcal{N}(0, \sigma^2) \end{cases} \quad (\mathcal{B}_1)$$

Sous ce modèle, nous cherchons à estimer les paramètres $\theta = (\mu_k, \omega_k, \sigma^2)_{k=1\dots K} \in \mathbb{R}^{2K+1}$.

Le modèle statistique considéré peut être vu comme un modèle à données non-observées. En effet, les données $\mathbf{y} = (y(t_1), \dots, y(t_n))$ sont issues de l'observation partielle des données complètes $(\mathbf{y}, \boldsymbol{\alpha})$ composées des observations et des efficacités neuronales $\boldsymbol{\alpha}$. Les données complètes $(\mathbf{y}, \boldsymbol{\alpha})$ ont pour densité

$$p(\mathbf{y}, \boldsymbol{\alpha}; \theta) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (y(t_i) - H \circ x(t_i, \mathbf{a}, \boldsymbol{\alpha}))^2 \right\} \\ \times \frac{1}{(2\pi)^{J/2} \det(\boldsymbol{\omega})^{1/2}} \exp \left\{ -\frac{1}{2} (\boldsymbol{\alpha} - \mathbf{m})' \boldsymbol{\omega}^{-1} (\boldsymbol{\alpha} - \mathbf{m}) \right\}$$

Les données observées \mathbf{y} ont pour densité :

$$p(\mathbf{y}; \theta) = \int p(\mathbf{y}, \boldsymbol{\alpha}; \theta) d\boldsymbol{\alpha}$$

Maximiser la vraisemblance $p(\mathbf{y}; \theta)$ ne peut être fait de façon exacte et nécessite le recours à un procédé itératif. Par définition du modèle statistique \mathcal{B}_1 , nous sommes exactement dans le cadre d'application de l'algorithme SAEM-MCMC dans sa version adaptée aux modèles définis par équations différentielles ordinaires, présentée dans le chapitre 5. L'estimation des paramètres est faite sur le modèle approché $\mathcal{B}_{1,h}$:

$$\begin{cases} y(t_i) = H_h(t_i, \boldsymbol{\alpha}, \mathbf{a}) + \varepsilon_i & \text{pour } i = 1 \dots n \\ \boldsymbol{\alpha} \sim \mathcal{N}(\mathbf{m}, \boldsymbol{\omega}) \\ \varepsilon_i \sim \text{i.i.d } \mathcal{N}(0, \sigma^2) \end{cases} \quad (\mathcal{B}_{1,h})$$

avec h le pas de la méthode de résolution numérique d'équations différentielles retenue et x_h la solution approchée du système. Dans ce cas, l'algorithme SAEM combiné avec une simulation par MCMC se décompose en les étapes détaillées ci-dessous. Plus précisément, à l'itération ℓ de l'algorithme SAEM-MCMC :

- *Etape de simulation (S)*

Simulation des efficacités neuronales $\boldsymbol{\alpha}^\ell$ sous la loi a posteriori $p_{\boldsymbol{\alpha}|\mathbf{y}}(\cdot, \mathbf{y}; \hat{\theta}^{\ell-1})$ en faisant quelques itérations de l'algorithme de Hastings-Metropolis.

- *Calcul des statistiques suffisantes* : $\forall k \in \{1 \dots K\}$,

$$S_{k,1} = \sum_{j=1}^{J_k} \alpha_{k,j}^\ell, \quad S_{k,2} = \sum_{j=1}^{J_k} (\alpha_{k,j}^\ell)^2, \quad S_3 = \sum_{i=1}^n (y(t_i) - H \circ x_h(t_i, \boldsymbol{\alpha}^\ell, \mathbf{a}))^2$$

- *Approximation stochastique (SA)* : $\forall k \in \{1 \dots K\}$,

$$s_{k,1}^\ell = s_{k,1}^{\ell-1} + \gamma_\ell [S_{k,1} - s_{k,1}^{\ell-1}], \quad s_{k,2}^\ell = s_{k,2}^{\ell-1} + \gamma_\ell [S_{k,2} - s_{k,2}^{\ell-1}], \quad s_3^\ell = s_3^{\ell-1} + \gamma_\ell [S_3 - s_3^{\ell-1}]$$

- *Maximisation de l'espérance conditionnelle (M)* : $\forall k \in \{1 \dots K\}$,

$$\hat{\mu}_k^\ell = \frac{1}{J_k} s_{k,1}^\ell, \quad \hat{V}_k^\ell = \frac{1}{J_k} s_{k,2}^\ell - (\hat{\mu}_k^\ell)^2, \quad \widehat{\sigma^2}^\ell = \frac{1}{n} s_3^\ell$$

6.5 Illustration sur données simulées

Nous considérons deux jeux de données simulées sous le modèle hémodynamique sur lequel nous testons les performances des méthodes d'estimation présentées précédemment.

Les données

- Les observations sont faites toutes les $TR = 2.4$ secondes.
- Les deux paradigmes simulés contiennent 2 types de stimuli ayant 15 occurrences chacun. Les temps inter-stimuli sont choisis aléatoirement dans une distribution de mélange de deux lois gaussiennes de paramètres $\mathcal{N}(5, 0.1)$ et $\mathcal{N}(0.2, 0.001)$. Plus précisément, soient τ_{j-1} et τ_j les instants d'occurrence des $(j-1)$ -ième et j -ième stimuli :

$$\begin{aligned} \tau_j - \tau_{j-1} &\sim \frac{1}{2}\mathcal{N}(6, 0.5) + \frac{1}{2}\mathcal{N}(5, 0.01) && \text{dans le jeu de données 1} \\ \tau_j - \tau_{j-1} &\sim \frac{1}{2}\mathcal{N}(6, 0.5) + \frac{1}{2}\mathcal{N}(0.8, 0.01) && \text{dans le jeu de données 2} \end{aligned}$$

- Les paramètres physiologiques sont choisis égaux à $\mathbf{a} = (1.54, 2.46, 0.98, 0.33, 0.34)$. V_0 est pris égal à 0.03.
- Les efficacités neuronales sont simulées suivant les distributions suivantes :

$$\begin{aligned} \alpha_{1,j} &\sim \mathcal{N}(0.4, 0.3) & \alpha_{2,j} &\sim \mathcal{N}(0.5, 0.3) && \text{dans le jeu de données 1} \\ \alpha_{1,j} &\sim \mathcal{N}(0.4, 0.05) & \alpha_{2,j} &\sim \mathcal{N}(0.8, 0.05) && \text{dans le jeu de données 2} \end{aligned}$$

- Le bruit est simulé avec une variance .

$$\begin{aligned} \sigma^2 &= 0.1 && \text{dans le jeu de données 1} \\ \sigma^2 &= 0.2 && \text{dans le jeu de données 2} \end{aligned}$$

Les signaux BOLD obtenus pour les jeux de données 1 et 2 sont respectivement représentés sur les figures 6.7 et 6.8.

Méthode de résolution numérique du système différentiel La méthode choisie pour approcher numériquement la solution du système différentiel (\mathcal{B}) est la fonction ODE23 du logiciel **Matlab**. Cette fonction est une implémentation d'un schéma explicite de Runge-Kutta (2,3) avec pour coefficients les paires de Bogacki et Shampine. Les données sont simulées avec un pas 10 fois plus petit que le pas de résolution utilisé lors de l'estimation (égal à $1/20$).

Mise en place de l'algorithme SAEM-MCMC

- Choix des noyaux de l'algorithme MCMC

Afin d'obtenir une estimation performante, nous devons choisir soigneusement les noyaux de l'algorithme MCMC. En effet, ce choix conditionne la vitesse de convergence de l'algorithme vers la distribution d'intérêt, ici la loi conditionnelle $p_{\alpha|y}$. Les noyaux choisis sont les suivants :

1. La première distribution de proposition est la loi a priori. Ainsi, à l'itération r de l'algorithme MCMC, un candidat est proposé suivant la loi :

$$\boldsymbol{\alpha}^{(c)} \sim \mathcal{N}(\mathbf{m}^{(\ell)}, \boldsymbol{\omega}^{(\ell)})$$

où $\mathbf{m}^{(\ell)}$ et $\boldsymbol{\omega}^{(\ell)}$ sont les valeurs courantes des paramètres.

2. Le second noyau consiste en un petit déplacement. Plus précisément, le candidat est simulé dans un voisinage du vecteur $\boldsymbol{\alpha}$ courant :

$$\boldsymbol{\alpha}^{(c)} = \boldsymbol{\alpha}^{r-1} + \mathcal{N}(0, \delta)$$

Une combinaison de ces deux noyaux permet d'explorer efficacement le support de la loi d'intérêt. Le premier noyau permet de couvrir une grande "surface" du domaine tandis que le deuxième permet une exploration plus fine (voir rappels sur les méthodes MCMC au chapitre 4). Cependant, le paramètre δ influence fortement la vitesse de convergence de l'algorithme MCMC et requiert donc une calibration précise.

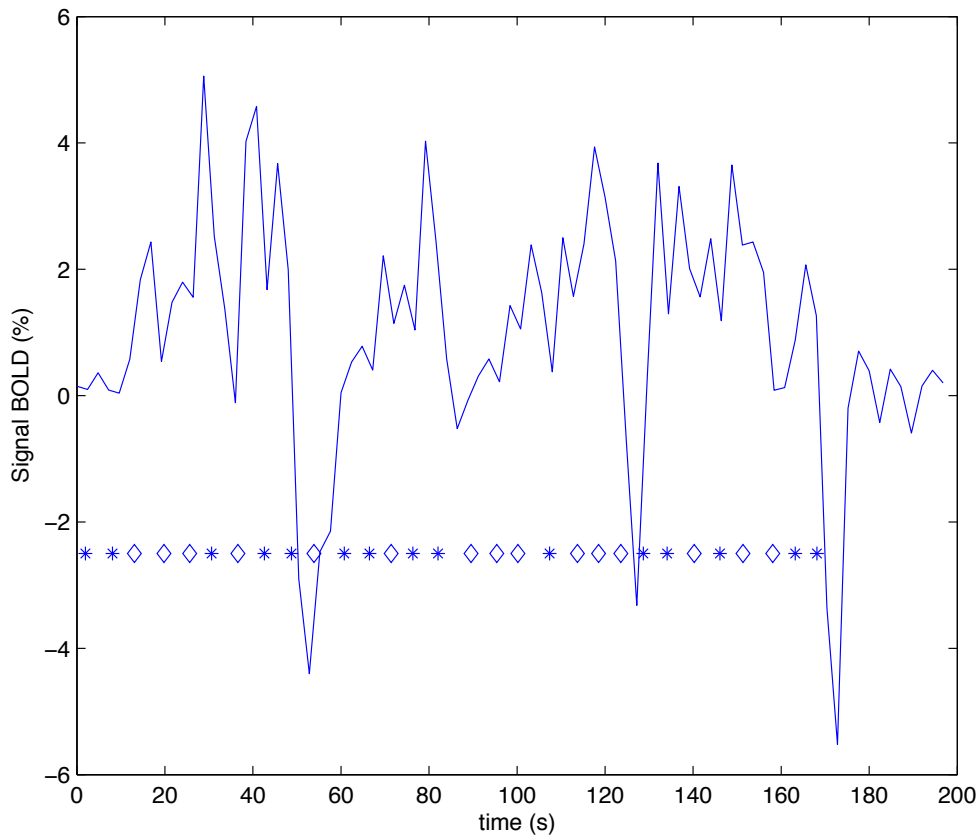


FIG. 6.7 – Jeu de données simulées 1 sous le “Balloon model” : les étoiles et les diamants représentent respectivement les instants d’occurrences des stimuli de type 1 et 2.

· *Approximation stochastique*

Le nombre d’itérations de l’algorithme SAEM est fixé à 200. La suite de pas $(\gamma_\ell)_{\ell \in \mathbb{N}}$ de l’approximation stochastique vaut

$$\gamma_\ell = \begin{cases} 1 & \text{si } \ell \leq 70 \\ \frac{1}{(\ell-70)^{0.2}} & \text{si } 70 < \ell \leq 120 \\ \frac{1}{(\ell-120)^{0.5}} & \text{si } 120 < \ell \leq 170 \\ \frac{1}{(\ell-170)} & \text{si } 170 < \ell \end{cases}$$

Résultats Les résultats se divisent en deux parties permettant de tester chaque partie du procédé d’estimation proposé.

- Dans un premier temps, nous considérons **le jeu de données 1 et travaillons à paramètres physiologiques a connus**. Nous ne cherchons qu’à estimer les paramètres des distributions des efficacités neuronales, le but étant de tester les performances de l’algorithme SAEM-MCMC sur le “Balloon model”. Les résultats obtenus sont résumés dans le tableau 6.2. L’évolution des estimées au cours des itérations est donnée sur la figure 6.9.

Nous constatons que la suite des estimateurs fournis par l’algorithme SAEM-MCMC converge vers des valeurs proches des valeurs de simulation. Les estimations des paramètres (μ_1, μ_2) et (ω_1, ω_2) étant faites sur un petit nombre d’occurrence (pour des raisons de temps de calculs), il

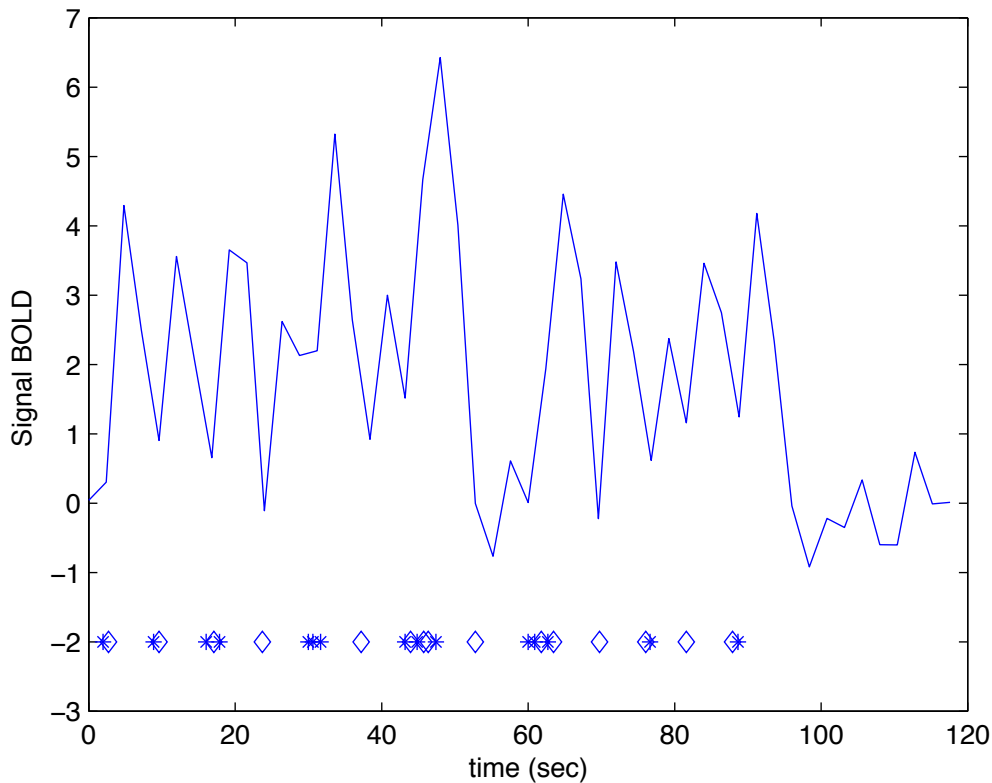


FIG. 6.8 – *Jeu de données simulées 2 sous le “Balloon model”*. Les étoiles et les diamants représentent respectivement les instants d’occurrences des stimuli de type 1 et 2.

est intéressant de comparer les moyennes et variances empiriques des efficacités neuronales simulées avec les estimations des paramètres. Plus précisément, on constate que la moyenne empirique des efficacité neuronales de type 1 et 2 valent respectivement 0.42 et 0.58, μ_1 et μ_2 étant estimées à 0.43 et 0.59. De même, les variances empiriques des efficacité neuronales de type 1 et 2 valent respectivement 0.37 et 0.37 tandis que ω_1 et ω_2 sont estimées à 0.35 et 0.35. Ainsi l’algorithme fournit des estimateurs de bonne qualité. L’algorithme mettant plus de 48 heures à converger sur un ordinateur Acer processeur Intel Pentium R, 1700 MHz, nous ne fournissons pas de biais ni de RMSE des estimateurs. Ce temps de convergence est du au fait qu’à chaque itération de l’algorithme MCMC, il faut résoudre le système différentiel (\mathcal{B}) qui comprend quatre équations sur un intervalle de temps important.

- Dans un deuxième temps, nous mettons en place la procédure complète d’estimation sur **le jeu de données simulées 2** : nous estimons dans un premier temps les paramètres physiologiques ($\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$) en supposant les efficacités constantes par groupes de stimuli, puis nous estimons les paramètres ($\mathbf{m}, \omega, \sigma^2$).

Les paramètres physiologiques sont obtenus en minimisant le critère des moindres carrés suivant :

$$\Lambda_\lambda(\mathbf{a}, \boldsymbol{\mu}) = \sum_{i=1}^N [y(t_i) - H \circ x(t_i, \boldsymbol{\mu}, \mathbf{a})]^2 + \sum_{\ell=1}^3 \lambda_\ell (\mathbf{a}_\ell - \mathbf{a}_{\ell,0})^2$$

où $\mathbf{a}_{1,0} = 1.5$, $\mathbf{a}_{2,0} = 2.4$, $\mathbf{a}_{3,0} = 1$. Les constantes λ_ℓ sont prises respectivement égales à 0.5, 0.5 et 1. Les estimations obtenues sont résumées dans le Tableau 6.3.

Nous traçons pour chaque ensemble de paramètres physiologiques (de simulation et estimés) la

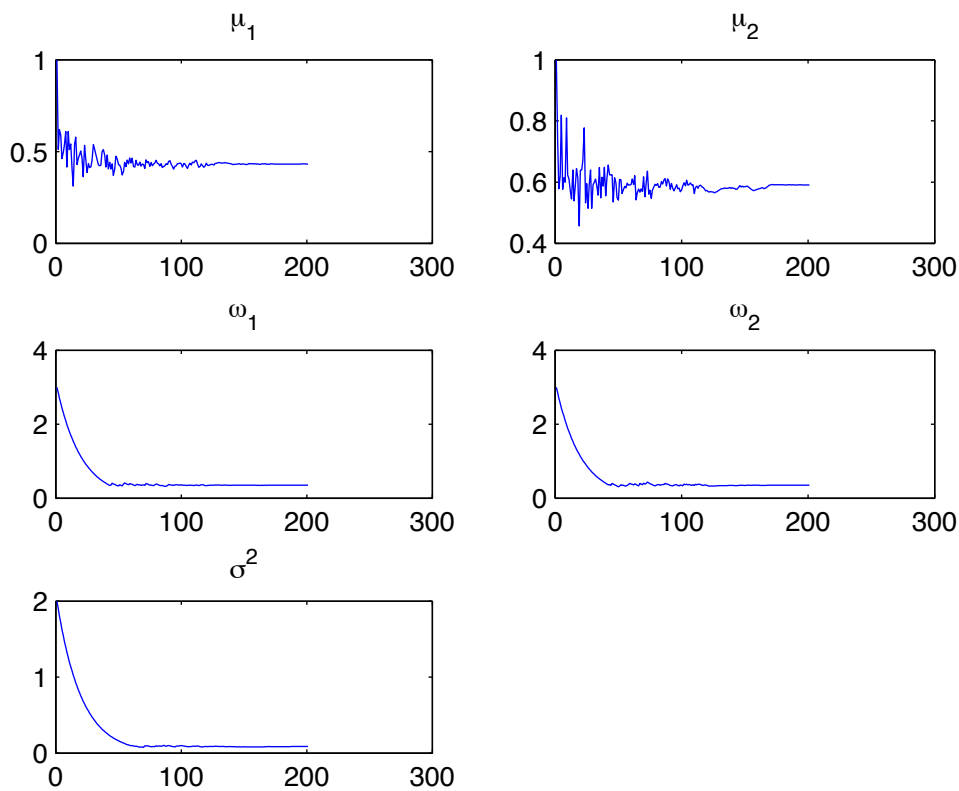


FIG. 6.9 – Jeu de données simulées 1 sous le “Balloon model” : Convergence des estimées des paramètres

réponse BOLD correspondante issu du “*Balloon model*”. Ces réponses sont tracées sur la figure 6.10. Nous constatons que chaque ensemble de paramètres bien que différent fournit une réponse BOLD similaire. Nous nous confrontons ainsi aux problèmes d’identifiabilité précédemment évoqués. Ces paramètres physiologiques sont ensuite utilisés pour estimer les paramètres de la distribution des efficacités neuronales. Pour cela nous utilisons l’algorithme SAEM-MCMC. L’évolution des estimateurs au cours des itérations est représentée sur la figure 6.11. Les résultats obtenus sont résumés dans le tableau 6.4.

Les résultats sont d’un peu moins bonne qualité comme nous pouvions nous y attendre. En effet, dans ce jeu de données, nous avons doublé la variance du bruit d’observation (0.2 contre 0.1 précédemment). De plus, ici nous avons travaillé avec des paramètres physiologiques estimés et non connus. Cependant les résultats obtenus restent tout à fait satisfaisants. Le principal handicap de cette méthode demeure son coût en terme de temps de calcul.

6.6 Application sur données réelles

Après avoir illustré notre méthode d’estimation sur jeux de données simulées, nous l’appliquons sur données réelles.

	μ_1	μ_2	ω_1	ω_2	σ^2
Simulation	0.4	0.5	0.3	0.3	0.1
Val empiriques	0.42	0.58	0.37	0.37	0.09
Initialisation	1	1	3	3	2
Estimation	0.43	0.59	0.35	0.35	0.087

 TAB. 6.2 – *Jeu de données simulées 1 sous le “Balloon model”*: résultats des estimations des paramètres

	a_1	a_2	a_3
Simulation	1.54	2.46	0.98
Estimation	1.52	2.39	0.94

 TAB. 6.3 – *Jeu de données simulées 2 sous le “Balloon model”*: résultats des estimations des paramètres physiologiques

6.6.1 Description des données et problématique

Les données Le jeu de données utilisé ici est fourni par T. Deneux. Les données ont été acquises à l’hôpital de la Timone (Marseille, France) sur un scanner 3T. Le temps inter-scans est de 825 ms. Huit volontaires ont participé à cette étude (six hommes et deux femmes, âgés de 19 à 25 ans, sans problème de vue). Les stimuli sont constitués de damiers clignotants. Nous appellerons chaque apparition du damier “flash”. Leur apparition est plus ou moins rapprochée dans le temps. L’état de repos est un écran gris.

Le paradigme considéré est composé de répétitions aléatoires des flashes arrivant par blocs et séparés d’environ 25 secondes. Ces blocs peuvent être constitués de huit flashes d’1 seconde, de quatre flashes d’1 seconde, de deux flashes d’1 seconde, d’un unique flash d’1 seconde, de cinq flashes de 200ms ou encore d’un unique flash d’une durée de 200ms ($8 \times 1s$ - $4 \times 1s$ - $2 \times 1s$ - $1 \times 1s$ - $5 \times 200ms$ - $1 \times 200ms$). La fonction-créneaux correspondant à ce paradigme est tracée sur la figure 6.12. Les données considérées sont extraites du cortex visuel primaire d’un des sujets. Le signal est moyenné sur une région d’intérêt composée de 40 voxels. Les données brutes sont représentées sur la figure 6.13.

Problématique Nous avons vu au paragraphe 6.3.1 que le modèle physiologique permet de rendre compte d’un phénomène de saturation hémodynamique (phénomène non-linéaire). Autrement dit, si le sujet est soumis à plusieurs stimuli rapprochés dans le temps, les réponses hémodynamiques aux deuxième, troisième, etc, stimuli sont de plus faibles amplitudes que la réponse hémodynamique au premier stimulus. Il est intéressant de se demander si un tel phénomène existe au niveau neuronal. Nous parlerons alors de *suppression à la répétition*.

Pour cela, considérons le paradigme expérimental et introduisons une variabilité des efficacités neuronales en fonction du type de blocs de flashes considérés. Le paradigme est constitué de 6 types de stimuli c’est-à-dire d’une succession de blocs de flashes pouvant être $8 \times 1s$, $4 \times 1s$, $2 \times 1s$, $1 \times 1s$, $5 \times 200ms$ ou $1 \times 200ms$.

type 1	$8 \times 1s$
type 2	$4 \times 1s$
type 3	$2 \times 1s$
type 4	$1 \times 1s$
type 5	$5 \times 200ms$
type 6	$1 \times 200ms$

Pour $k = 1 \dots 6$, notons α_k l’efficacité neuronale des stimuli de type k . De façon logique, nous nous attendons à ce que $\alpha_1 \geq \alpha_2 \geq \alpha_3 \geq \alpha_4$ par définition des stimuli. Si un phénomène de suppression

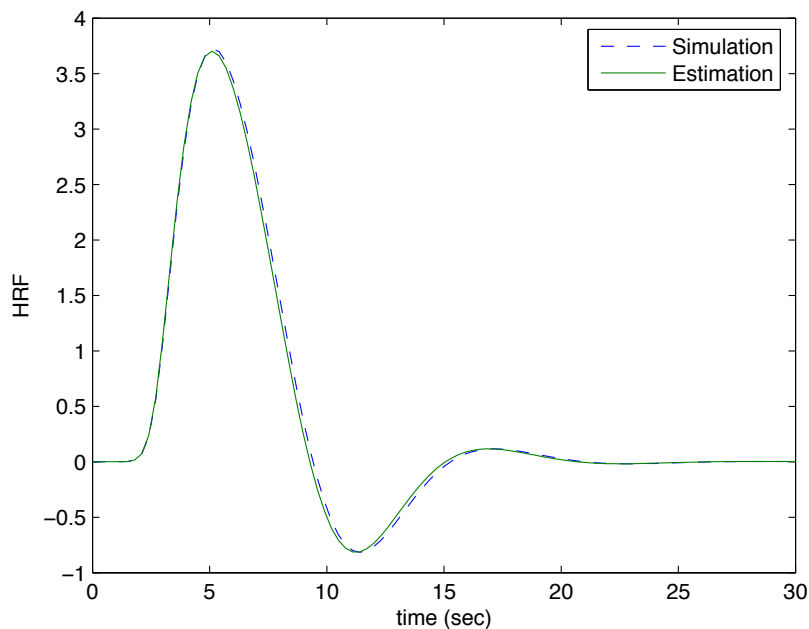


FIG. 6.10 – *Jeu de données simulées 2 sous le “Balloon model”* : reconstruction des réponses BOLD à partir des paramètres physiologiques de simulation (---) et estimés (trait plein)

	μ_1	μ_2	ω_1	ω_2	σ^2
Simul	0.40	0.80	0.224	0.224	0.20
V. emp	0.38	0.75	0.217	0.252	0.12
Estim	0.45	0.72	0.25	0.26	0.13

TAB. 6.4 – *Jeu de données simulées 2 sous le “Balloon model”* : résultats des estimations des paramètres par SAEM-MCMC-ODE

à la répétition existe, alors 8 flashes d’1 seconde ne seraient pas d’efficacité 2 fois plus grande que 4 flashes d’1 seconde, etc... Plus précisément,

$$\begin{aligned}\alpha_1 &\leq 2\alpha_2 \\ \alpha_2 &\leq 2\alpha_3 \\ \alpha_3 &\leq 2\alpha_4\end{aligned}$$

Nous avons alors deux objectifs :

- D’une part, nous cherchons à valider ou non l’hypothèse de variabilité des efficacités neuronales. Pour cela, nous utilisons la variante du “*Balloon model*” (proposée dans cette thèse) dans laquelle les efficacités sont la réalisation d’une variable aléatoire gaussienne de moyenne μ et de variance ω : $\forall k = 1 \dots 6$,

$$\alpha_k \sim_{\text{i.i.d}} \mathcal{N}(\mu, \omega)$$

L’estimation par maximum de vraisemblance des paramètres (μ, ω, σ^2) (par l’algorithme SAEM-MCMC-ODE) permet de réaliser un test du maximum de vraisemblance.

- D’autre part, nous cherchons à mettre en évidence l’existence d’un phénomène de suppression à la répétition. Pour cela, l’implémentation de l’algorithme SAEM-MCMC-ODE fournit une estimation

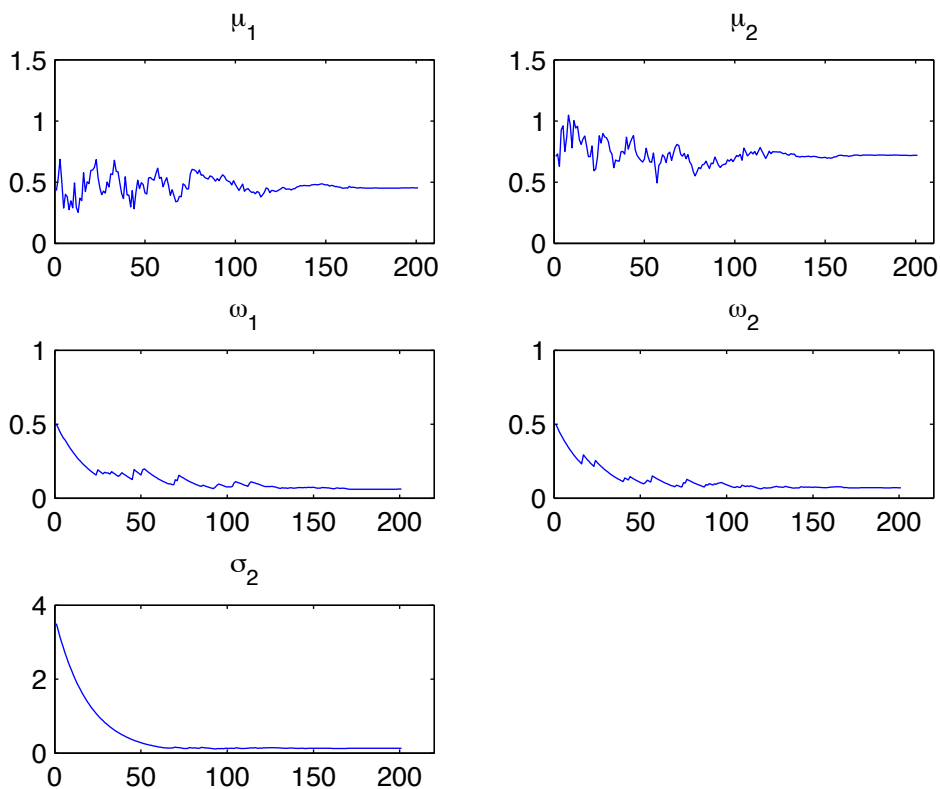


FIG. 6.11 – *Jeu de données simulées 2 sous le “Balloon model”* : Convergence des estimées des paramètres par SAEM-MCMC-ODE

des espérances conditionnelle des efficacités neuronales $(E[\alpha_k|\mathbf{y}])_{k=1\dots 6}$. Nous comparons alors ces valeurs entre elle.

Remarque 20. *Notons que ici, nous avons un niveau de variabilité en moins que dans ceux présentés dans le cadre de l’illustration sur données simulées de la partie 6.5. Dans cette partie, nous avons simulé des efficacités neuronales avec une variabilité inter-types d’évènements et inter-occurrences ($\alpha_{k,j}$ dépendait de k et j l’occurrence). Le modèle présenté ici en est un cas particulier (α_k ne dépend que de k). Cependant, l’hypothèse de variabilité inter-types d’évènements n’ayant pas été proposée sur ce jeu de données réelles, il semblait essentiel de commencer par cette étape.*

6.6.2 Pré-traitements

Estimation de la dérive basse fréquence Les données récupérées n’ont pas été filtrées au préalable et de ce fait sont polluées par des phénomènes basses fréquences tels que dérive lente de la machine, mouvements lents du sujet, etc. Nous cherchons donc dans un premier temps à filtrer les données. Comme décrit au chapitre 4, nous modélisons la dérive basse fréquence par une famille de fonctions orthonormales (P_1, \dots, P_Q) , de la forme $P_q(t) = \sqrt{\frac{2}{n}} \cos\left(\frac{\pi(2t+1)(q-1)}{2n}\right)$. Le nombre de fonctions dans cette base (noté Q) vaut $Q = [2nTRf_{min}] + 1$ où f_{min} est la plus basse fréquence attribuable au terme de dérive et vaut $\frac{1}{120}$ Hertz ici. Comme $n = 1117$, nous obtenons $Q = 16$. Les quantités $(l_q)_{q=1\dots Q}$ représentent les coefficients-poids inconnus des fonctions de la base et doivent être estimés.

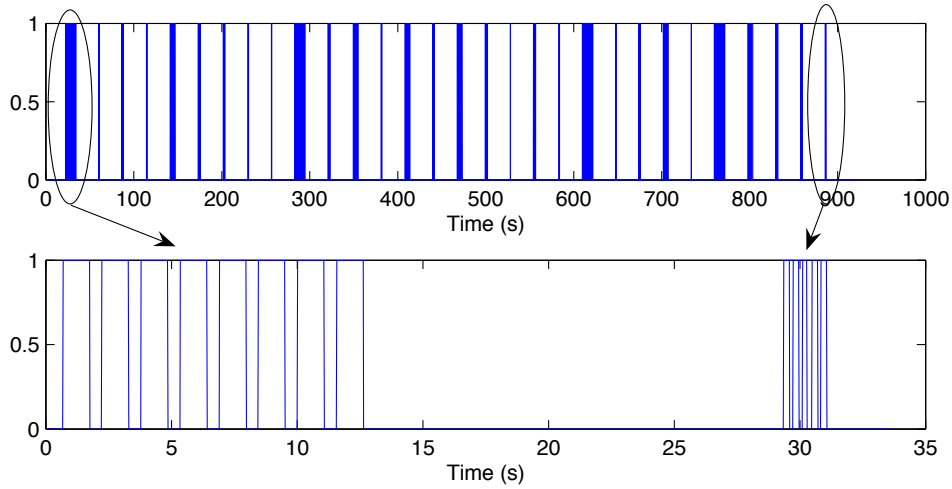


FIG. 6.12 – *Jeu de données réelles (3)* : paradigme subi par les sujets. Figure du haut : paradigme complet. Figure du bas : zoom de deux stimuli-blocs composés que flash de 8×1 secondes et 5×200 ms respectivement.

Le modèle statistique s'écrit alors :

$$\begin{pmatrix} y(t_1) \\ \vdots \\ y(t_n) \end{pmatrix} = \begin{bmatrix} P_1(t_1) & \dots & P_Q(t_1) \\ \vdots & & \vdots \\ P_1(t_n) & \dots & P_Q(t_n) \end{bmatrix} \begin{pmatrix} l_1 \\ \vdots \\ l_Q \end{pmatrix} + \begin{pmatrix} H \circ x(t_1, \mathbf{a}, \boldsymbol{\alpha}) \\ \vdots \\ H \circ x(t_n, \mathbf{a}, \boldsymbol{\alpha}) \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

$$\mathbf{y} = P\mathbf{l} + \begin{pmatrix} H \circ x(t_1, \mathbf{a}, \boldsymbol{\alpha}) \\ \vdots \\ H \circ x(t_n, \mathbf{a}, \boldsymbol{\alpha}) \end{pmatrix} + \boldsymbol{\varepsilon}$$

avec $\boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_n)$.

L'estimation simultanée des paramètres \mathbf{l} , \mathbf{a} et $\boldsymbol{\alpha}$ est beaucoup trop ambitieuse sur un jeu de données réelles. Nous considérons donc un ensemble de paramètres physiologiques \mathbf{a} et d'efficacités neuronales $\boldsymbol{\alpha}$ permettant d'avoir un ajustement convenable. Plus précisément, nous prenons $\mathbf{a} = (1.54, 2.46, 0.98, 0.33, 0.34, 0.5)$ et $\alpha_k = 1$ pour tout k . Nous les supposons alors connus et nous estimons les poids \mathbf{l} par moindres carrés. La dérive basse-fréquence ($P\hat{\mathbf{l}}$) obtenue est représentée sur la figure 6.13 en trait épais. Dorénavant, nous considérons les données filtrées $\mathbf{y}_f = \mathbf{y} - P\hat{\mathbf{l}}$.

Remarque 21. *Dans la suite, pour soucis de simplicité, nous continuerons à noter \mathbf{y} les données filtrées.*

Estimation des paramètres physiologiques Dans un premier temps, nous cherchons à estimer les paramètres physiologiques en supposons les efficacités neuronales constantes. Il s'avère que pour ce jeu de données particulier, les reconstructions obtenues avec les valeurs standards des paramètres ($\mathbf{a} = (1.54, 2.46, 0.98, 0.33, 0.34, 0.5)$ et $\alpha_k = 1$) donnent de bons ajustements. De plus, la mise en place d'un algorithme de minimisation de moindres carrés ne permet pas de diminuer la variance résiduelle de façon efficace. Par conséquent, pour ce jeu de données réelles, nous prenons ces valeurs des paramètres physiologiques. De plus, nous ne poussons pas plus loin l'étude, puisque nous nous intéressons avant tout à l'estimation des efficacités neuronales.

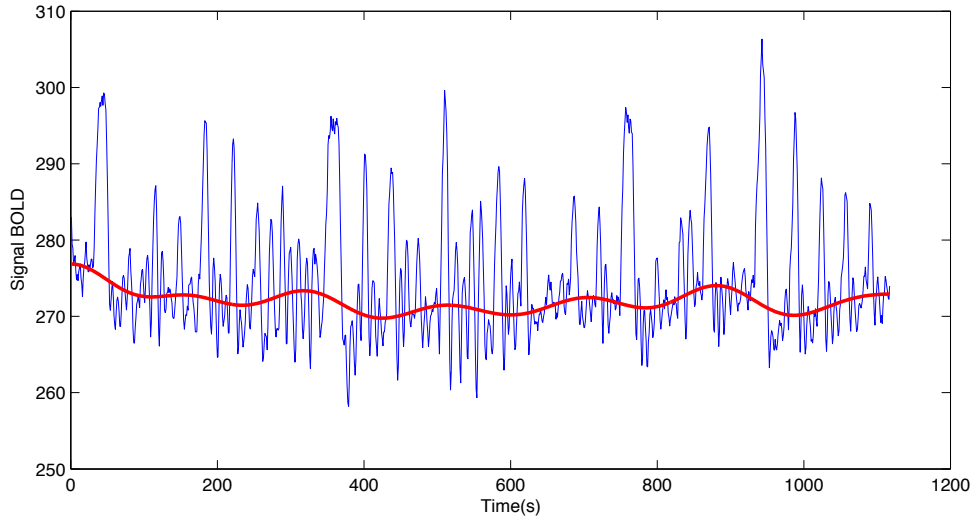


FIG. 6.13 – *Jeu de données réelles (3)* : données brutes. Estimation de la dérive basses-fréquences en trait épais.

6.6.3 Implémentation de l'algorithme SAEM-MCMC-ODE

Afin d'estimer les paramètres du modèle statistique $\theta = (\mu, \omega, \sigma^2)$, nous implémentons l'algorithme SAEM-MCMC-ODE présenté dans la section 6.4.2 avec 500 itérations. La suite des pas de l'approximation stochastique vaut :

$$\gamma_\ell = \begin{cases} 1 & \text{si } \ell \leq 100 \\ \frac{1}{(\ell-100)^{0.2}} & \text{si } 100 < \ell \leq 250 \\ \frac{1}{(\ell-250)^{0.5}} & \text{si } 250 < \ell \leq 300 \\ \frac{1}{(\ell-300)} & \text{si } 300 < \ell \end{cases}$$

L'algorithme est initialisé sur les valeurs $\mu_0 = 1$, $\omega_0 = 1$ et $\sigma_0^2 = 1$. La convergence des estimateurs aux cours des occurrences de l'algorithme est tracée sur la figure 6.14. Nous obtenons alors les valeurs suivantes : $\hat{\mu} = 0.88$, $\hat{\omega} = 0.015$ et $\hat{\sigma} = 18.4$. La prédiction des données est tracée sur la figure 6.15. Enfin nous traçons sur la figure 6.16 les espérances conditionnelles des efficacités neuronales $\hat{E}[\alpha_k | \mathbf{y}]$ pour $k = 1 \dots 6$, en fonction du type de stimuli k . Sur la même figure nous traçons les intervalles de confiance obtenus.

6.6.4 Résultats

Validation de l'hypothèse de variabilité des efficacités Comme dans le cas linéaire, nous cherchons à valider ou non l'hypothèse de variabilité des efficacité neuronales. Comme précédemment, cela revient à tester l'hypothèse \mathcal{H}_0 : " $\omega = 0$ " contre l'hypothèse \mathcal{H}_1 : " $\omega > 0$ ". Nous envisageons un test du rapport des vraisemblances. Définissons les deux modèles statistiques suivants :

$$\begin{cases} y(t_i) = H \circ x(t_i, \boldsymbol{\alpha}, \mathbf{a}) + \varepsilon_i \\ \varepsilon_i \sim \text{i.i.d } \mathcal{N}(0, \sigma^2) \\ \alpha_k \sim \text{i.i.d } \mathcal{N}(\mu, \omega) \end{cases} \quad (\mathcal{M}_1)$$

et

$$\begin{cases} y(t_i) = H \circ x(t_i, \boldsymbol{\alpha}, \mathbf{a}) + \varepsilon_i \\ \varepsilon_i \sim \text{i.i.d } \mathcal{N}(0, \sigma^2) \\ \alpha_k = \alpha \end{cases} \quad (\mathcal{M}_0)$$

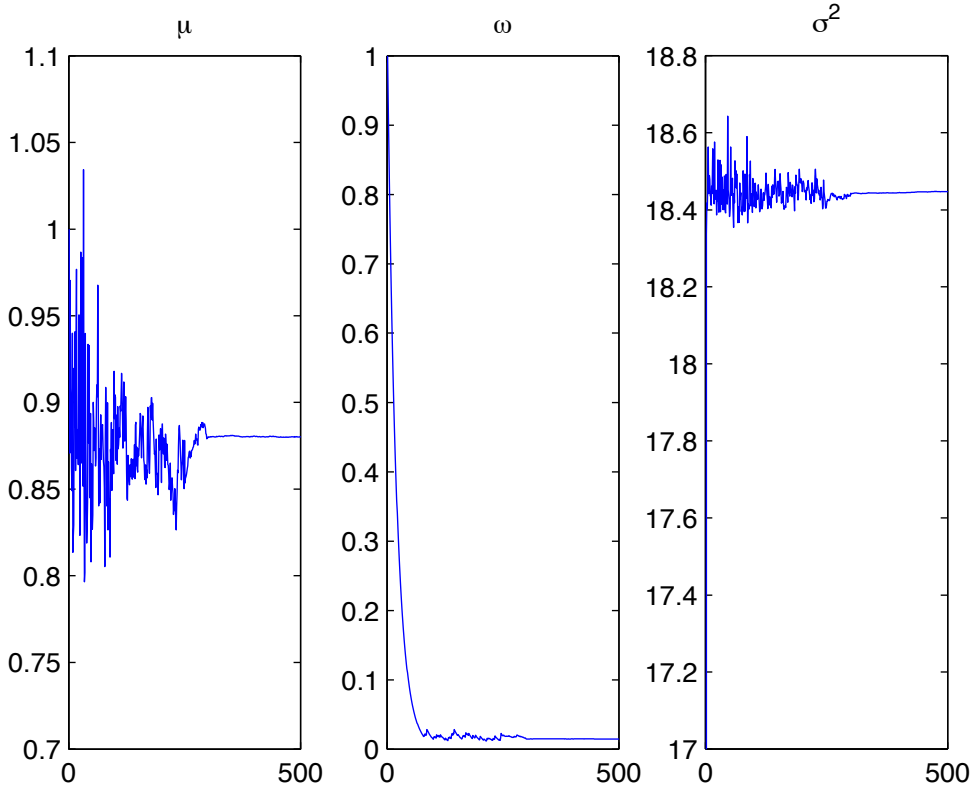


FIG. 6.14 – *Jeu de données réelles (3)* : estimation des paramètres. Convergence des paramètres au cours des itérations

Le modèle \mathcal{M}_1 est le modèle précédemment décrit dans lequel les efficacités neuronales sont variables. Le modèle \mathcal{M}_0 est le modèle standard dans la littérature dans lequel $\alpha_k = \alpha$ pour tout $k = 1, \dots, 6$.

Soient $lv_0(\mathbf{y}; \alpha, \sigma^2)$ et $lv_1(\mathbf{y}; \theta)$ les log-vraisemblances respectives des modèles \mathcal{M}_0 et \mathcal{M}_1 . Nous avons alors : $lv_0(\mathbf{y}; \alpha, \sigma^2) = -\frac{n}{2} \log 2\pi\sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (y(t_i) - H \circ x(t_i, \alpha, \mathbf{a}))^2$. Soient $\hat{\alpha}_0$ et $\hat{\sigma}_0^2$ les estimateurs du maximum de vraisemblance respectifs de α et σ^2 sous le modèle \mathcal{M}_0 :

$$\hat{\alpha}_0 = \arg \min_{\alpha \in \mathbb{R}} \sum_{i=1}^n (y(t_i) - H \circ x(t_i, \alpha, \mathbf{a}))^2 \quad \text{et} \quad \hat{\sigma}_0^2 = \frac{1}{n} \sum_{i=1}^n (y(t_i) - H \circ x(t_i, \hat{\alpha}_0, \mathbf{a}))^2.$$

Alors

$$lv_0(\mathbf{y}; \hat{\alpha}_0, \hat{\sigma}_0^2) = -\frac{n}{2} \log 2\pi\hat{\sigma}_0^2 - \frac{n}{2}.$$

Sous le modèle \mathcal{M}_1 , $\hat{\theta}$ a été obtenu para l'algorithme SAEM-MCMC-ODE. La log-vraisemblance ne peut être calculée de façon explicite :

$$lv_1(\mathbf{y}; \hat{\theta}) = \log \left[\int_{\alpha} p_{\mathbf{y}|\alpha}(\mathbf{y}|\alpha; \hat{\sigma}^2) \pi(\alpha; \hat{\mu}, \hat{\omega}) d\alpha \right]$$

Nous avons donc recours à une méthode de Monte Carlo pour calculer cette quantité. Soient $(\alpha^{(m)})_{m=1, \dots, M}$ M réalisations de la variable aléatoire α , générées sous la distribution gaussienne de moyenne $\hat{\mu}$ et de variance $\hat{\omega}$. Alors, définissons l'estimateur de la log-vraisemblance $lv_1(\mathbf{y}; \hat{\theta})$ suivant :

$$\tilde{lv}_{1,M}(\mathbf{y}; \hat{\theta}) = -\frac{n}{2} \log(2\pi\hat{\sigma}^2) + \log \left[\frac{1}{M} \sum_m \exp \left\{ -\frac{1}{2\hat{\sigma}^2} \sum_{i=1}^n (y(t_i) - H \circ x(t_i, \alpha^{(m)}, \mathbf{a}))^2 \right\} \right]$$

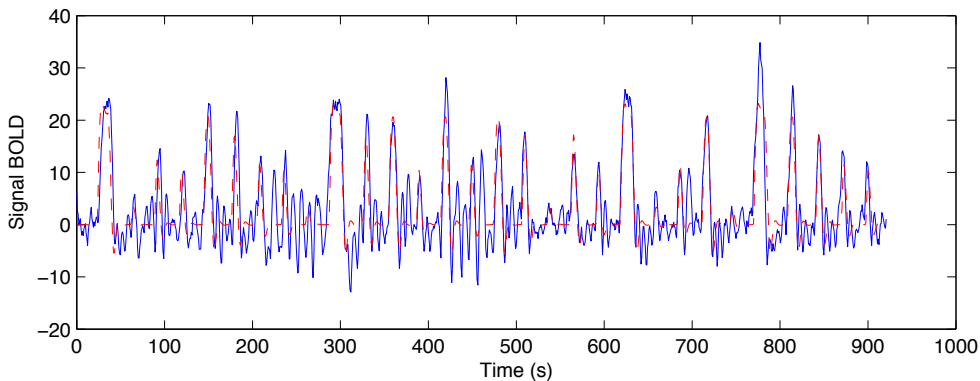


FIG. 6.15 – *Jeu de données réelles (3)* : reconstruction des données après estimation. Données observées en trait plein, reconstruction en traits pointillés.

Définissons la statistique de test suivante : $\xi_M = 2(\tilde{lv}_{1,M}(\mathbf{y}; \hat{\theta})) - lv_0(\mathbf{y}; \hat{\alpha}, \hat{\sigma}_0^2)$. Par la loi forte des grands nombres, ξ_M converge presque sûrement vers $\xi = 2(lv_1(\mathbf{y}; \hat{\theta})) - lv_0(\mathbf{y}; \hat{\alpha}, \hat{\sigma}_0^2)$, quand M tend vers l'infini. Nous avons discuté au chapitre 4 (section 4.7) de la distribution asymptotique de ξ . Dans le cas d'une unique variance, il s'agit d'un mélange entre un dirac en 0 et un khi deux à 1 degré de liberté (chacun étant pondéré du poids $1/2$) : $\frac{1}{2}\delta_0 + \frac{1}{2}\chi^2(1)$.

Dans le cadre de ce jeu de données réelles, après calculs, nous avons $lv_0(\mathbf{y}; \hat{\alpha}, \hat{\sigma}_0^2) = -3225.0$. Avec $M = 200$, nous obtenons $\tilde{lv}_{1,M}(\mathbf{y}; \hat{\theta}) = -3217.6$. Le quantile de ξ à 95% vaut 1.92. Nous rejetons l'hypothèse \mathcal{H}_0 .

Mise en évidence et quantification d'un phénomène de suppression à la répétition Comme prévu, nous constatons sur la figure 6.16 que

$$\hat{E}[\alpha_1|\mathbf{y}] \geq \hat{E}[\alpha_2|\mathbf{y}] \geq \hat{E}[\alpha_3|\mathbf{y}] \geq \hat{E}[\alpha_4|\mathbf{y}].$$

Il paraît difficile de conclure quant à l'efficacité des événements de type 6 ($1 \times 200\text{ms}$). Le fait que leur efficacité $\hat{E}[\alpha_6|\mathbf{y}]$ soit plus grande que celle de type 5 $\hat{E}[\alpha_5|\mathbf{y}]$ ($5 \times 200\text{ms}$) semble biologiquement peu plausible.

Nous cherchons à mettre en évidence un phénomène de suppression à la répétition au niveau neuronal. Nous comparons donc

$$\frac{\hat{E}[\alpha_1|\mathbf{y}]}{8}, \quad \frac{\hat{E}[\alpha_2|\mathbf{y}]}{4}, \quad \frac{\hat{E}[\alpha_3|\mathbf{y}]}{2} \quad \text{et} \quad \hat{E}[\alpha_4|\mathbf{y}]$$

En l'absence de phénomène de suppression à la répétition, ces termes devraient être égaux. Dans notre cas, ils valent respectivement :

$$0.13, \quad 0.22, \quad 0.42 \quad \text{et} \quad 0.75$$

Donc il existe clairement un phénomène de suppression à la répétition qui est d'autant plus important que les répétitions sont nombreuses.

Enfin, il est intéressant de comparer les stimuli de type 4 et 5 ($1 \times 1\text{s}$ et $5 \times 200\text{ms}$ respectivement). Nous constatons que

$$\hat{E}[\alpha_4|\mathbf{y}] = 0.8404 \leq \hat{E}[\alpha_5|\mathbf{y}] = 0.9298$$

6.7 Discussion

Dans ce travail, nous avons proposé une variante du "*Balloon model*" permettant d'inclure une variabilité inter-type de stimuli sur les efficacités neuronales et développé une méthode d'estimation

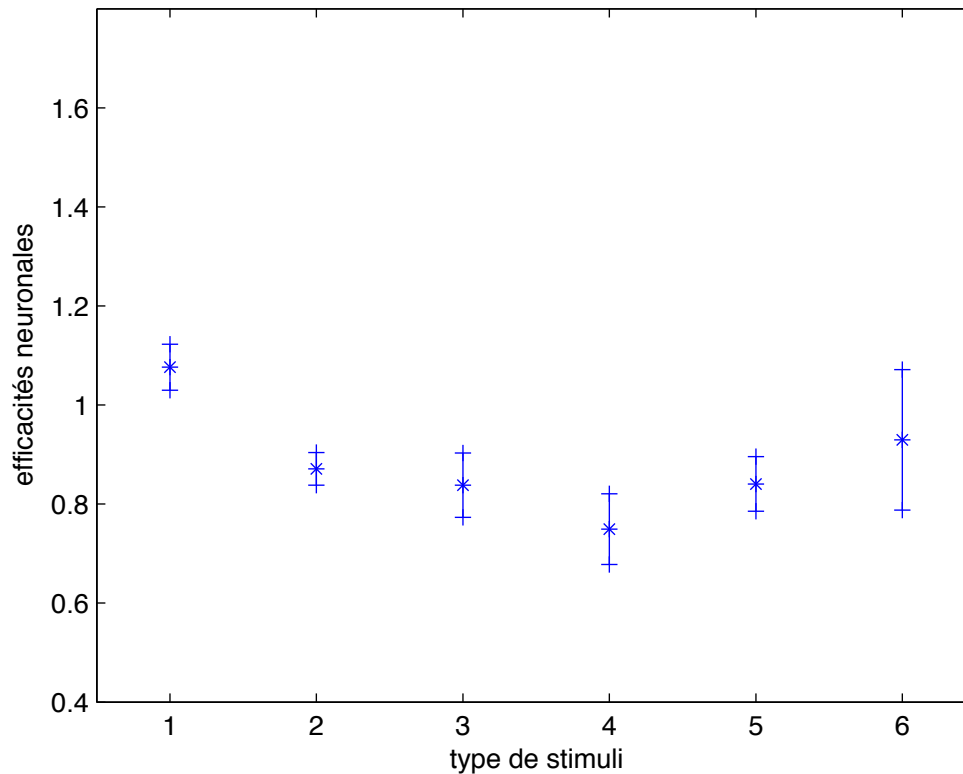


FIG. 6.16 – *Jeu de données réelles (3)* : espérances conditionnelles des efficacités neuronales en fonction de l'ordre d'occurrence dans le bloc des flashes avec intervalle de confiance

des paramètres du modèle par maximum de vraisemblance. La méthode a prouvé son efficacité sur données simulées et a été appliquée sur données réelles.

Sur le jeu de données considéré, l'hypothèse de variabilité inter-type de stimuli a été validée et un phénomène de suppression à la répétition au niveau neuronal a été mis en évidence et quantifié.

Afin de poursuivre cette étude, il serait intéressant d'ajouter une variabilité supplémentaire en considérant par exemple des efficacités dépendant de la place du flash dans le bloc de flashes : $\alpha_{k,j}$ où k est le type de blocs ($8 \times 1s$, $4 \times 1s$ etc...) et j est la place du flash dans le bloc. Ceci permettrait de quantifier plus précisément le phénomène de suppression à la répétition.

Une étude sur plusieurs sujets avec d'autres types de stimuli pourrait venir compléter ce travail pour valider à grande échelle les phénomènes observés. L'acquisition conjointe des données EEG et IRMF est une technique potentiellement intéressante dans ce cadre puisqu'elle nous permettrait d'avoir un décours simultané des activités électriques.

Quatrième partie

Estimation dans des modèles définis par équations différentielles stochastiques

Les modèles présentés au chapitre 5 ne considèrent qu'un bruit d'observation. Il est naturel de les étendre aux modèles où l'équation différentielle est en fait une équation différentielle stochastique. Au chapitre 7¹ de cette partie, nous nous intéressons donc à l'estimation paramétrique de processus de diffusion observés à temps discrets, avec bruit de mesure et dépendant de paramètres aléatoires, à la fois dans un cadre bayésien et dans un cadre d'estimation par maximum de vraisemblance.

La méthode d'estimation par maximum de vraisemblance proposée est appliquée sur un jeu de données réelles issue de la pharmacocinétique dans le cadre d'un modèle non-linéaire à effets mixtes.

Au chapitre 8, nous nous intéressons à la difficile question de l'identifiabilité de tels modèles.

¹Ce chapitre ayant fait l'objet d'un article, celui-ci est écrit en anglais.

Chapitre 7

Parametric inference for incomplete data models defined by stochastic differential equations

Travail réalisé en collaboration avec Adeline Samson ¹, article soumis à *ESAIM, Probability and Statistics*.

RÉSUMÉ :

Nous considérons des processus de diffusion (solutions d'EDS) observés à temps discrets et avec bruit de mesure, et dont les paramètres de la fonction de drift sont aléatoires. Deux méthodes d'estimation, l'une par maximum de vraisemblance et l'autre par approche bayésienne (respectivement une version stochastique de l'algorithme EM et un échantillonneur de Gibbs), sont développées pour ces modèles. Elles reposent sur une approximation par la méthode d'Euler-Maruyama du processus de diffusion, approximation obtenue en introduisant des temps auxiliaires entre les instants de mesure. La convergence de ces algorithmes est démontrée. Les erreurs induites par l'approximation d'Euler-Maruyama sur la fonction de vraisemblance et la distribution a posteriori des paramètres sont contrôlées en fonction du pas du schéma d'approximation. Une étude sur données simulées à partir d'un modèle non-linéaire à effets mixtes issu de la pharmacocinétique illustre la précision de la méthode d'estimation par maximum de vraisemblance proposée. L'analyse du jeu de données réelles de Théophylline illustre la pertinence de l'approche par SDE par rapport à l'approche déterministe (par ODE).

MOTS-CLEFS :

Bayesian estimation, Brownian bridge, Diffusion process, Euler-Maruyama approximation, Gibbs algorithm, Incomplete data model, Maximum likelihood estimation, Non-linear mixed effects model, SAEM algorithm

¹INSERM U738, Paris, France ; University Paris 7, Paris, France

Introduction

Time-dependent dynamic processes that follow the laws of finance, physics, physiology or biology are usually described by differential systems. For example, stock price dynamics or short-term interest rates can be described using a wide class of financial differential systems. As another example, in biology, pharmacokinetics consists in the study of the evolution of a drug in an organism. It is described through dynamic systems, the human body being assimilated to a set of compartments within which the drug flows. In these contexts, diffusion models described by stochastic differential equations (SDEs) are natural extensions to the corresponding deterministic models (defined by ordinary differential equations, ODEs) to account for time-dependent or serial correlated residual errors and to handle real life variations in model parameters occurring over time. This variability in the model parameters is most often not predictable, not fully understood or too complex to be modeled deterministically. Thus the SDEs consider errors associated with misspecifications and approximations in the dynamic system.

The parametric estimation of such diffusion processes is a key issue. Estimation of continuously observed diffusion processes is widely studied (see for instance [75, 102]). However, for obvious practical purposes, real longitudinal data are always gathered at discrete points in time (for example stock prices collected once a day, drug concentration measured every hour in patient blood, etc.). Within this framework, statistical inference of discretely observed diffusion processes is a critical question for both maximum likelihood and Bayesian approaches. When the transition probability of the diffusion process is explicitly known, Dacunha-Castelle and Florens-Zmirou [34] propose a consistent maximum likelihood estimator. Classical Bayesian algorithm such as Gibbs sampling can also be directly applied in this particular case.

However, this transition density has generally no closed form and the estimation methods have to sidestep this difficulty. A short summary of such estimation methods is provided below (see [102, 114] for complete reviews). Analytical methods include those of Bibby and Sorensen [15], Sorensen [115] – using estimating functions –, Poulsen [101] – using a numerical solution of the Kolmogorov equation – or Aït-Sahalia [3] – based on an analytical non-Gaussian approximation of the likelihood function. Other methods approximate the transition density via simulation. They consider the unobserved paths as missing data and introduce a set of auxiliary latent data points between every pair of observations. Along these auxiliary latent data points, the process can be finely sampled using the Gaussian Euler-Maruyama approximation to evaluate the likelihood function via numerical integration as proposed by Pedersen [97] and Elerian et al. [40], or to evaluate the posterior distribution in a Bayesian analysis again via numerical integration, as discussed by Eraker [41] and Roberts and Stramer [108]. In this context and for both maximum likelihood and Bayesian estimations, standard Markov Chain Monte-Carlo (MCMC) methods are used to sample the process with the conditional distributions. However, the convergence rate of these estimation methods decreases with the increase in number of latent data points. Different solutions are proposed to overcome this difficulty: Eraker [41] suggests the sampling of only one element at a time, while Elerian et al. [40] propose to sample block-wise with an importance sampling algorithm. Roberts and Stramer [108] take a slightly different approach as they sample transformations of the diffusion process. To sidestep the Euler-Maruyama approximation, Beskos et al. [14, 13] develop an exact simulation method of the diffusion process, applicable even without any analytical form of the transition density. This algorithm can be included in a Monte-Carlo procedure to approximate the likelihood function for a classical estimation and in a Gibbs algorithm for a Bayesian inference. However, this exact simulation method is only adapted for time-homogeneous SDEs, which is frequently not the case when studying biological dynamical systems for example. Furthermore, even under the conditions defined by Beskos et al. [14, 13], this exact method requires the inclusion of accept-reject algorithms, which are difficult to implement in the general case of non-linear SDEs and often require a large computational time. Therefore an Euler-Maruyama approximation approach is considered in this paper.

The above-cited papers do not take into account the observation noise on the collected data, which is non-realistic in many cases. For example in the financial context, the daily evolution of an asset price depends on the price fluctuations within each business day. In the biological context, endpoints such as drug concentrations are generally measured with a certain variability due to experimental limits. To reflect this observation noise, we consider the following regression statistical model \mathcal{M} : the observed data $y = (y_1, \dots, y_J)$ are a realization of a random variable Y deduced from a scalar diffusion process Z , as stated by the following equation:

$$Y_j = Z(t_j) + \varepsilon_j, \tag{\mathcal{M}}$$

where $(\varepsilon_j)_{j=1, \dots, J}$ is a sequence of i.i.d Gaussian random variables of variance σ^2 , representing the measurement errors. The diffusion process Z is defined as the solution of the SDE describing the observed dynamic process:

$$dZ(t) = F(Z, t, \phi)dt + \gamma dB(t),$$

driven by a Brownian motion $\{B_t, t_0 \leq t \leq T\}$, a drift function F depending on a parameter ϕ and a volatility coefficient γ . If the volatility coefficient γ is null, the SDE is an ODE, the SDE model parameter ϕ being evidently equivalent to the parameter of the corresponding ODE system, and therefore being interpreted in the same way. In such models, two fundamentally different types of noise have to be distinguished: the dynamic noise γ , reflecting the real random fluctuations around the corresponding theoretical dynamic model, and the measurement noise σ representing the uncorrelated part of the residual variability associated with assay, dosing and sampling errors, for instance, in a biological context. The problem of the parameter estimation of discretely observed diffusion processes with additive measurement noise is evoked in few papers and is not completely solved. In the particular case of linear SDEs, the Kalman filter [112] or the EM algorithms [113] can be used. When the observed process is a Gaussian martingale, Jensen and Petersen [67] and Gloter and Jacod [53] exhibit estimators and study their theoretical properties. Unfortunately, these explicit forms of maximum likelihood estimates are limited to the linear SDEs case.

In this paper, we assume in addition that the parameter ϕ is a realization of a random variable Φ distributed with a probability π depending on a parameter β . This is the case for example in drug pharmacokinetics studies of which use will be detailed below. Basically, in order to estimate drug pharmacokinetic parameters, the drug concentration is sampled repeatedly among several individuals, the parameter ϕ being assumed different between the subjects and thus considered as individual non-observed random data.

The main objective of this paper is to develop methods to estimate the parameters vector $\theta = (\beta, \gamma^2, \sigma^2)$ in the general case of non-linear SDEs. Such a method is proposed by Ditlevsen et al [38] in the particular case of non-linear mixed effects models but their example is restricted to the case where the transition probabilities have explicit expression. Overgaard et al. [95] and Tornøe et al. [122] have also introduced SDEs in mixed models, combining an extended Kalman filter of the diffusion process with an approximated maximum likelihood estimation algorithm based on a linearization of the model. However, the convergence properties of this estimation algorithm based on linearization are not proved. A different point of view can be taken for the parameters estimation, the random quantities Z and Φ being considered as non-observed random data. In that case, the model \mathcal{M} belongs to the framework of incomplete data models, for which several estimation methods are developed for both classical and Bayesian approaches. For classical inference, the Expectation-Maximization (EM) algorithm proposed by Dempster et al. [36] is a broadly applied approach taking advantage of the incomplete data model structure. When the E-step has no closed form, Celeux and Diebolt [29], Wei and Tanner [127] and Delyon et al. [35] propose different stochastic versions of this algorithm. These methods require the simulation of the non-observed data using Markov Chain Monte-Carlo (MCMC) algorithms, as proposed by Kuhn and Lavielle [72]. For the Bayesian approach, tuned Gibbs algorithms are developed to estimate the posterior distribution $p_{\theta|Y}(\cdot|y)$ of θ , a specified

prior distribution $p_\theta(\cdot)$ for θ being given. When the simulation under the posterior distribution cannot be done in a closed form, hybrid Gibbs sampling algorithms are proposed in the literature, including Metropolis-Hastings procedures [126, 11]. To our knowledge, these estimation methods are not yet extended to noisy discretely observed diffusion processes models considered in this paper.

Our objective is thus to propose efficient estimation methods of the vector of parameters θ for the model \mathcal{M} , together with theoretical convergence results for both classical and Bayesian inference. We consider an approximate statistical model, of which the regression term is the Euler-Maruyama discretized approximate diffusion process of the SDE. The parameter inference is then performed on this new model, using a stochastic version of the EM algorithm for the classical estimation approach, or using a hybrid version of the Gibbs sampling algorithm for the Bayesian approach.

Section 7.1 describes the setup of the problem which is considered in this paper, detailing the diffusion process and its Euler-Maruyama approximation. The estimation algorithms for the maximum likelihood and the Bayesian approaches are respectively presented in Sections 7.2 and 7.3. These sections detail a tuned MCMC procedure supplying both theoretical and computational convergence properties to these algorithms. The error on the estimation induced by the Euler-Maruyama scheme is quantified in Section 7.4. In Section 7.5, the maximum likelihood algorithm is applied to a non-linear mixed effects model issued from pharmacokinetics. Section 7.6 concludes with some discussion.

7.1 Data and Model

7.1.1 Incomplete data model defined by SDEs

Let $y = (y_j)_{j=0..J}$ denote the vector of the observations measured at times $t_0 \leq t_1 \leq \dots \leq t_J \leq T$. We consider that y is a realization of the random variable Y defined through the following statistical model \mathcal{M} :

$$\left. \begin{aligned} Y_j &= Z(t_j) + \varepsilon_j, & 0 \leq j \leq J \\ \varepsilon_j &\sim_{i.i.d} \mathcal{N}(0, \sigma^2), \\ dZ(t) &= F(Z, t, \Phi) dt + \gamma dB(t), & Z(t_0, \Phi) = Z_0(\Phi), \\ \Phi &\sim \pi(\cdot, \beta) \end{aligned} \right\} \quad (7.1.1) \quad (\mathcal{M})$$

where $\varepsilon = (\varepsilon_1, \dots, \varepsilon_J)$ represents the measurement error, with a measurement noise variance σ^2 . The regression term is a realization of the diffusion process $Z : \mathbb{R} \rightarrow \mathbb{R}$ defined by the equation (7.1.1), with B a one-dimensional Brownian motion, γ is the volatility coefficient and the function $F : \mathbb{R} \times [t_0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ is the known measurable drift function, non-linearly depending on the non-observed parameter $\Phi \in \mathbb{R}^d$. We assume that Φ is a random variable, distributed with the density π , depending on the parameter $\beta \in \mathbb{R}^p$. The initial condition Z_0 of this process is a deterministic known function of the random parameter Φ (this deterministic function can be a constant).

Our objective is to propose a classical and a Bayesian estimation methods of the parameters vector θ , where $\theta = (\beta, \gamma^2, \sigma^2)$ belongs to some open subset Θ of the Euclidean space \mathbb{R}^{p+2} . As the random parameter Φ and the random trajectory Z are not observed, this statistical problem can be viewed as an incomplete data model. The observable vector Y is thus consider as part of a so-called complete vector (Y, Z, Φ) .

Remark 1. • *This work can be extended to a statistical model with a regression function being equal to $g(Z(t))$, with g a linear or non-linear function, i.e.*

$$Y_j = g(Z(t_j)) + \varepsilon_j, \quad 0 \leq j \leq J.$$

However, for the simplicity's sake, we only consider the case $g(Z(t)) = Z(t)$ in this paper.

- *The identifiability of this model is a complex problem which is beyond the scope of this paper. However, for simple examples such as linear SDEs the parameters identifiability can be proved.*

7.1.2 Diffusion model

The diffusion process (7.1.1) is defined on a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$. Statistical inference makes sense only if the existence and uniqueness of a solution of the SDE (7.1.1) for all $Z(t_0)$, Φ and γ is ensured. Sufficient conditions of existence and uniqueness are the following globally Lipschitz, linear growth and boundedness conditions:

Assumption (A0):

1. For all $\phi \in \mathbb{R}^d$, for all $0 < R < \infty$, there exists $0 < K_R < \infty$ such that for all $t_0 \leq t \leq T$, for all $x, x' \in \mathbb{R}$ with $|x| \leq R, |x'| \leq R$

$$|F(x, t, \phi) - F(x', t, \phi)| \leq K_R |x - x'|.$$

2. For all $\phi \in \mathbb{R}^d$, for all $0 < T < \infty$, there exists a constant $0 < C_T < \infty$ such that for all $t_0 \leq t \leq T$, for all $x \in \mathbb{R}$

$$\gamma + |F(x, t, \phi)| \leq C_T(1 + |x|).$$

Under this assumption, for any $t_0 < t < T$, the distribution of $Z(t)$ conditioned by the filtration \mathcal{F}_{t-} is absolutely continuous with respect to the Lebesgue measure on \mathbb{R} (\mathcal{F}_{t-} being the filtration generated by $\{Z(s), s < t\}$). This distribution is denoted $p_{Z|\Phi}(\cdot|\phi; \gamma^2)$ in the following. As a consequence, both Y and (Y, Z, Φ) have density functions, denoted respectively $p_Y(y; \theta)$ and $p_{Y,Z,\Phi}(y, z, \phi; \theta)$ depending on the parameter θ .

7.1.3 Introduction of an approximate statistical model

For common SDEs, the diffusion density $p_{Z|\Phi}$ has generally no closed form. Consequently neither the likelihood of the observed data $p_Y(y; \theta)$ nor the likelihood of the complete data $p_{Y,Z,\Phi}(y, z, \phi; \theta)$ have analytical forms, which further complicates the parameters estimation. To overcome this difficulty, an approximate statistical model, based on the Euler-Maruyama approximation of the diffusion process is introduced.

Euler-Maruyama approximation of the diffusion process

The Euler-Maruyama scheme is one of the simplest discrete-time approximation of a diffusion process leading to Gaussian approximations of the transition densities. If the time intervals between the observation instants are too great to obtain a good approximation of the transition density, a natural approach is to introduce a set of auxiliary latent data points between every pair of observations, as first proposed by Pedersen [97]. Let $t_0 = \tau_0 < \tau_1 < \dots < \tau_n < \dots < \tau_N = t_J$ denote the deduced discretization of the time interval $[t_0, t_J]$. Let us assume that, for all $j = 0 \dots J$, there exists an integer n_j verifying $t_j = \tau_{n_j}$, with $n_0 = 0$ by definition. Let $(h_n)_{1 \leq n \leq N}$ be the sequence of the step sizes defined as $h_n = \tau_n - \tau_{n-1}$. Let $h = \max_{1 \leq n \leq N} h_n$ be the maximal step size.

Then the diffusion process denoted W and supplied by the Euler-Maruyama approximation of the SDE is described by the following iterative scheme: for a fixed ϕ , $W_0 = Z_0(\phi)$, and for $n = 1 \dots N$,

$$\begin{aligned} h_n &= \tau_n - \tau_{n-1}, \\ W_n &= W_{n-1} + h_n F(W_{n-1}, \tau_{n-1}, \phi) + \gamma \sqrt{h_n} \xi_n, \\ \xi_n &\sim_{i.i.d} \mathcal{N}(0, 1). \end{aligned}$$

Consequently, $(w_{n_0}, \dots, w_{n_J})$ is an approximation of the original diffusion process at observations instants $(z(t_0), \dots, z(t_J))$. In the following, let $w = (w_n)_{n=0 \dots N}$ denote a realization vector of the process W at the discrete times $(\tau_n)_{n=0 \dots N}$.

Approximate statistical model

Using this approximation of the diffusion process provided by the Euler-Maruyama scheme of step size h , an approximate statistical model denoted model \mathcal{M}_h is defined as:

$$\left. \begin{aligned} Y_j &= W_{n_j} + \varepsilon_j, & 0 \leq j \leq J, \\ \varepsilon_j &\sim_{i.i.d} \mathcal{N}(0, \sigma^2), \\ h_n &= \tau_n - \tau_{n-1}, \\ W_n &= W_{n-1} + h_n F(W_{n-1}, \tau_{n-1}, \Phi) + \gamma \sqrt{h_n} \xi_n, & 1 \leq n \leq N, \\ \xi_n &\sim_{i.i.d} \mathcal{N}(0, 1), \\ \Phi &\sim \pi(\cdot; \beta), \end{aligned} \right\} (\mathcal{M}_h)$$

with $W_0 = Z_0(\Phi)$. On this model \mathcal{M}_h , Y results from the partial observation of the complete data (Y, W, Φ) where W is the process at the discrete times $(\tau_n)_{n=0 \dots N}$.

Remark 2. *In this data augmentation framework, the choice of the discretization grid $(\tau_n)_{0 \leq n \leq N}$ is a central issue to guarantee the fast convergence of the estimation algorithms. Indeed, on the one hand, a small step size h ensures a fine Gaussian diffusion approximation. However, on the other hand, it increases the volume of missing data (W, Φ) , which can lead to arbitrarily poor convergence properties of the algorithms when the missing data volume widely exceeds the volume of actually observed data Y . Furthermore, the time intervals between two observations can be strongly different. Therefore, for practical purposes and to prevent unbalanced volumes of missing data, we propose to adjust the step sizes for each single time interval.*

In the following, the distributions referring to the model \mathcal{M}_h are denoted q while those referring to the model \mathcal{M} are denoted p . On \mathcal{M}_h , the observation vector y is distributed with density distribution $q_Y(y; \theta)$, which has no closed form because of the SDE non-linearity with respect to ϕ . But by enriching the observed data with the missing data, and by the Markov property of the diffusion process, the complete data likelihood is analytically known:

$$\begin{aligned} q_{Y, W, \Phi}(y, w, \phi; \theta) &= q_{Y|W}(y|w; \sigma^2) \prod_{n=1}^N q_{W|\Phi}(w_n|w_{n-1}, \phi; \gamma^2) \pi(\phi; \beta) \\ &= q_{Y|W}(y|w; \sigma^2) \prod_{n=1}^N d(w_n; w_{n-1} + h_n F(w_{n-1}, \tau_{n-1}, \phi), \gamma^2 h_n) \pi(\phi; \beta), \end{aligned}$$

where $d(\cdot; m, v)$ denotes the Gaussian density with mean m and variance v . As a consequence, the estimation of θ can be performed on the model \mathcal{M}_h , using a stochastic version of the EM algorithm for a Maximum Likelihood approach or a Gibbs algorithm for a Bayesian approach.

7.2 Maximum Likelihood Estimation on the model \mathcal{M}_h

In this section we propose a maximum likelihood estimation method, the vector of parameters θ being thus estimated as the maximizing value of the likelihood $q_Y(\cdot; \theta)$.

7.2.1 Stochastic versions of the EM algorithm

The Expectation Maximization (EM) algorithm proposed by Demspeter et al. [36] takes advantage of the incomplete data model structure. We consider that the observed data Y are the partial observations of the complete data (Y, X) with X the vector of the non-observed data. The EM algorithm is useful in situations where the direct maximization of $\theta \rightarrow q_Y(\cdot; \theta)$ is more complex than the maximization of $\theta \rightarrow Q(\theta|\theta')$, with:

$$Q(\theta|\theta') = E_{X|Y} [\log p_{Y, X}(y, x; \theta)|y; \theta'] .$$

The EM algorithm is an iterative procedure: at the k -th iteration, the E-step is the evaluation of $Q_k(\theta) = Q(\theta | \theta_{k-1})$ while the M-step updates θ_{k-1} by maximizing $Q_k(\theta)$. For cases where the E-step has no closed form, Delyon et al. [35] propose the Stochastic Approximation EM algorithm (SAEM) replacing the E-step by a stochastic approximation of $Q_k(\theta)$. The E-step is thus divided into a simulation step (S-step) of the non-observed data $x^{(k)}$ with the conditional distribution $p_{X|Y}(\cdot | y; \theta_{k-1})$ and a stochastic approximation step (SA-step):

$$Q_k(\theta) = Q_{k-1}(\theta) + \alpha_k \left[\log \left(p_{Y,X}(y, x^{(k)}; \theta_{k-1}) \right) - Q_{k-1}(\theta) \right],$$

where $(\alpha_k)_{k \in \mathbb{N}}$ is a sequence of positive numbers decreasing to zero.

The distribution $p_{X|Y}(\cdot | y; \theta_{k-1})$ is likely to be a complex distribution, as for the model \mathcal{M}_h , resulting in the impossibility of a direct simulation of the non-observed data x . For such cases, Kuhn and Lavielle [72] suggest a MCMC scheme by constructing a Markov chain with an unique stationary distribution $p_{X|Y}(\cdot | y; \theta_{k-1})$ at the k -th iteration. They prove the convergence of the estimates sequence provided by this SAEM algorithm towards a maximum of the likelihood under general conditions and in the case where $p_{Y,X}$ belongs to a regular curved exponential family.

7.2.2 Extension of the SAEM algorithm to the model \mathcal{M}_h

In the particular case of the model \mathcal{M}_h , the non-observed data vector is equal to $X = (W, \Phi)$. As the simulation under the conditional distribution $q_{W, \Phi | Y}$ can not be performed directly, the SAEM algorithm combined with a MCMC procedure is applied to the model \mathcal{M}_h to estimate the model parameter θ . To ensure the convergence of the SAEM algorithm, the model \mathcal{M}_h is assumed to fulfill some regular conditions:

Assumption (A1):

1. $\pi(\cdot; \beta)$ is such that $q_{Y,W,\Phi}$ belongs to the exponential family:

$$\log q_{Y,W,\Phi}(y, w, \phi; \theta) = -\psi(\theta) + \langle S(y, w, \phi), \nu(\theta) \rangle,$$

where ψ and ν are two functions of θ , $S(y, w, \phi)$ is known as the minimal sufficient statistics of the complete model, taking its value in a subset $\tilde{\mathcal{S}}$ of \mathbb{R}^m and $\langle \cdot, \cdot \rangle$ is the scalar product on \mathbb{R}^m .

2. $\beta \mapsto \pi(\phi; \beta)$ is of class \mathcal{C}^m for all $\phi \in \mathbb{R}^d$.

Under the assumption (A1), the SA-step of the SAEM algorithm reduces to the approximation of $E[S(y, w, \phi) | y; \theta']$. The k -th iteration of the SAEM algorithm is thus

- S-Step: a realization of the non-observed data $(w^{(k)}, \phi^{(k)})$ is generated through M iterations of a MCMC procedure providing an uniformly ergodic Markov chain with $q_{W, \Phi | Y}(\cdot | y; \theta_{k-1})$ as unique stationary distribution,
- SA-Step: s_{k-1} is updated using the following stochastic approximation scheme:

$$s_k = s_{k-1} + \alpha_k (S(y, w^{(k)}, \phi^{(k)}) - s_{k-1}),$$

- M-Step: θ_{k-1} is updated to maximize the complete log-likelihood:

$$\hat{\theta}_k = \arg \max_{\theta} (-\psi(\theta) + \langle s_k, \nu(\theta) \rangle).$$

For example, the sufficient statistics corresponding to σ^2 and γ^2 are:

$$\begin{aligned} S^{(1)}(y, w, \phi) &= \frac{1}{J+1} \sum_{j=0}^J (y_j - w_{n_j})^2, \\ S^{(2)}(y, w, \phi) &= \frac{1}{N} \sum_{n=1}^N \frac{(w_n - h_n F(w_{n-1}, \tau_{n-1}, \phi))^2}{h_n}, \end{aligned}$$

and the M-step for σ^2 and γ^2 at iteration k reduces to $\widehat{\sigma}_k^2 = s_k^{(1)}$ and $\widehat{\gamma}_k^2 = s_k^{(2)}$. The sufficient statistics for β depend on the distribution $\pi(\cdot; \beta)$.

7.2.3 Convergence of the SAEM algorithm on the model \mathcal{M}_h

Let denote Π_θ the transition probability of the Markov chain generated by the MCMC algorithm. Following [72], the convergence of the SAEM algorithm combined with MCMC is ensured under the following additional assumption:

Assumption (A2):

1. The chain $(w^{(k)}, \phi^{(k)})_{k \geq 0}$ takes its values in a compact set \mathcal{E} of $\mathbb{R}^N \times \mathbb{R}^d$.
2. For any compact subset V of Θ , there exists a real constant L such that for any (θ, θ') in V^2

$$\sup_{\{(w, \phi), (w', \phi')\} \in \mathcal{E}} |\Pi_\theta(w', \phi' | w, \phi) - \Pi_{\theta'}(w', \phi' | w, \phi)| \leq L \|\theta - \theta'\|_{\mathbb{R}^{p+2}}$$

3. The transition probability Π_θ supplies an uniformly ergodic chain whose invariant probability is the conditional distribution $q_{W, \Phi | Y}(\cdot; \theta)$, i.e.

$$\exists K_\theta \in \mathbb{R}^+, \quad \exists \rho_\theta \in]0, 1[\quad | \quad \forall k \in \mathbb{N} \quad \|\Pi_\theta^k(\cdot | w, \phi) - q_{W, \Phi | Y}(\cdot; \theta)\|_{TV} \leq K_\theta \rho_\theta^k$$

where $\|\cdot\|_{TV}$ is the total variation norm. Furthermore,

$$K = \sup_{\theta \in \Theta} K_\theta < \infty \quad \text{and} \quad \rho = \sup_{\theta \in \Theta} \rho_\theta < 1$$

4. The function S_h is bounded on \mathcal{E} .

Theorem 1. *Let assumptions (A0-A1-A2) hold. Let $q_{W, \Phi | Y}$ have finite moments of order 1 and 2. Let (α_k) be a sequence of positive numbers decreasing to 0 such that for all k in \mathbb{N} , $\alpha_k \in [0, 1]$, $\sum_{k=1}^{\infty} \alpha_k = \infty$ and $\sum_{k=1}^{\infty} \alpha_k^2 < \infty$.*

Assuming the sequence $(s_k)_{k \geq 1}$ takes its values in a compact set, the sequence $(\widehat{\theta}_k)_{k \geq 1}$ obtained by the SAEM algorithm on the model \mathcal{M}_h converges almost surely towards a (local) maximum of the likelihood q_Y .

Proof: Assuming (A1-A2) and the existence of finite moments for $q_{W, \Phi | Y}$, the assumptions of [72] are fulfilled and ensure the convergence of the estimates towards a local maximum of the likelihood function.

Remark 3. *If the compactness on $(s_k)_{k \geq 0}$ is not checked or difficult to check, the algorithm can be stabilized using a projection of the stochastic approximation sequence on varying bounds proposed by Andrieu and Moulines [5].*

A MCMC procedure fulfilling the assumption of uniform ergodicity is proposed in the following part 7.2.4.

7.2.4 Simulation of the non-observed data using a MCMC procedure

At the k -th iteration of the SAEM algorithm, given an estimate $\widehat{\theta}_{k-1}$, a realization of the non-observed data $(w^{(k)}, \phi^{(k)})$ is generated through the succession of M iterations of a MCMC procedure. MCMC procedures construct a Markov chain with $q_{W, \Phi | Y}(\cdot | y; \widehat{\theta}_{k-1})$ as the invariant distribution, by proposing candidates (ϕ^c, w^c) with any proposal density Q . However, sampling all the missing data at the same time can lead to poor convergence properties. Therefore, a hybrid Gibbs algorithm is implemented and realized successively M times, the m -th iteration being written as:

1. generation of $\phi^{(m)}$, using a Metropolis-Hastings (M-H) procedure with Q_1 as proposal density and such that $q_{\Phi|Y,W}(\cdot | y, w^{(m-1)}; \widehat{\theta}_{k-1})$ is the invariant distribution.
2. generation of $w^{(m)}$, using a M-H procedure with Q_2 as proposal distribution and such that $q_{W|Y,\Phi}(\cdot | y, \phi^{(m)}; \widehat{\theta}_{k-1})$ is the invariant distribution.

A careful choice of the proposal densities Q_1 and Q_2 will help the algorithm to quickly explore the parameters space. In the following, some proposal densities of which efficiency is proved on numerical examples are detailed. To simplify the notation, the parameter $\widehat{\theta}_{k-1}$ is omitted since this simulation is performed for a fixed $\widehat{\theta}_{k-1}$.

Proposal distributions

1. Simulation of the candidate ϕ^c can be carried out with the prior density π which allows an efficient exploration of the space of parameters. This leads to an independent M-H algorithm. An alternative consists in generating a candidate in a neighborhood of $\phi^{(m-1)}$, $\phi^c = \phi^{(m-1)} + \eta$ with $\eta \sim \mathcal{N}(0, \delta)$ and where δ is a scaling parameter on which the algorithm convergence depends. This results in the so-called random-walk M-H algorithm (see for example [11]).
2. A trajectory candidate w^c can be generated using the Euler-Maruyama scheme which corresponds to the prior distribution. An alternative to simulate w^c consists in splitting the vector w into two parts $(w_{n_0}, \dots, w_{n_J})$ and w_{aux} , the former being the process observed at times $(t_j)_{j=0 \dots J}$ and the latter being the process observed at the auxiliary latent times excluding the observation times. The simulation of $(w_{n_0}^c, \dots, w_{n_J}^c)$ can be performed with random walk distributions: $w_{n_j}^c = w_{n_j}^{(m-1)} + \eta'$ where $\eta' \sim \mathcal{N}(0, \delta')$ and δ' is a scaling parameter chosen to ensure good convergence properties. As proposed by Pedersen [97], the trajectory at the auxiliary times w_{aux}^c can be generated using an unconditioned distribution but it would have poor convergence properties. A more appropriate strategy consists in generating a candidate w_{aux}^c using Brownian bridges, conditioning the proposed bridge on the events $(w_{n_j}^c)_{j=0 \dots J}$, as suggested by Eraker [41] or Roberts and Stramer [108]. More precisely, for $n_{j-1} < n < n_j$, w_{τ_n} is simulated with:

$$w_{\tau_n}^c = w_{n_{j-1}}^c + \frac{w_{n_j}^c - w_{n_{j-1}}^c}{t_j - t_{j-1}}(\tau_n - t_{j-1}) + \overline{B}_{\tau_n}$$

where \overline{B} is a standard Brownian bridge on $[0, 1]$ equal to zero for $t = 0$ and $t = 1$, which can be easily simulated.

Uniform ergodicity of the MCMC procedure

For checking assumption **(A2-3)**, it is possible to verify some minoration condition or Doeblin's condition for the transition probability Π_θ (see Chap. 16 of [90]). Otherwise, each case has to be considered individually.

For the independent M-H algorithm, its uniform ergodicity is ensured as soon as the proposal distribution verifies:

$$\exists \lambda \in \mathbb{R}^+ \quad | \quad \forall (w, \phi) \in \mathcal{E}, \quad Q(w, \phi) \geq \lambda q_{W,\Phi|Y}(w, \phi|y),$$

(see Th 2.1 in [121] for more details). The proposal distribution equal to the prior distribution $Q(w, \phi) = q_{W,\Phi}(w, \phi)$ fulfills this condition.

Moreover, in case of a cyclic combination, the uniform ergodicity of the Markov Chain is ensured if one of the proposal distributions satisfies a minoration condition (Prop. 3 and 4 of [121]). Thus the introduction of the prior distribution as proposal distribution is sufficient to ensure the uniform ergodicity of the Markov Chain.

7.3 Bayesian estimation of the model \mathcal{M}_h

The Bayesian model is defined as follows:

$$\left. \begin{aligned} Y_j &= Z(t_j) + \varepsilon_j, & 0 \leq j \leq J \\ \varepsilon_j &\sim_{i.i.d} \mathcal{N}(0, \sigma^2), \\ dZ(t) &= F(Z, t, \Phi) dt + \gamma dB(t), & Z(t_0, \Phi) = Z_0(\Phi), \\ \Phi &\sim \pi(\cdot, \beta) \\ \theta &\sim p_\theta(\cdot; \Gamma) \end{aligned} \right\} (\mathcal{M})$$

where $\theta = (\beta, \gamma^2, \sigma^2)$ is distributed from the prior distribution p_θ with fixed hyperparameters Γ . We assume that β , γ^2 and σ^2 have continuous prior densities $p_\beta(\cdot)$, $p_\gamma(\cdot)$, $p_\sigma(\cdot)$ respectively on \mathbb{R}^p , \mathbb{R} and \mathbb{R} . For example, a Gamma prior distribution can be chose for γ^{-2} and σ^{-2} . The prior distribution on β depends on the specific distribution π .

The Bayesian approach consists in the evaluation of the posterior distribution $p_{\theta|Y}$. According to the arguments developed in Section 7.1, the estimation procedure is applied to the model \mathcal{M}_h .

A simplistic approach would be to consider a basic Gibbs algorithm which simulates the non-observed data (ϕ, w) and then updates the parameter θ . However, as emphasized and illustrated by Roberts and Stramer [108], the quadratic variation of the diffusion process satisfies, for almost surely all observation times t_j and t_{j+1} :

$$\lim_{h \rightarrow 0} \sum_{n=n_j}^{n_{j+1}-1} (w_{n+1} - w_n)^2 = (t_{j+1} - t_j)\gamma^2. \quad (7.3.2)$$

Therefore, conditional on any process satisfying (7.3.2), the posterior distribution of the volatility $q(\gamma^2|y, \phi, w, \sigma^2, \beta) \propto q_{W|\Phi}(w|\phi; \gamma^2)p_\gamma(\gamma^2)$ is just a point mass at γ^2 . Consequently this data augmentation scheme is reducible. Roberts and Stramer [108] propose a reparameterization to avoid this problem and consider the following transformation:

$$\dot{w}_n = \frac{w_n}{\gamma} \quad \text{and} \quad \dot{F}(x, t, \phi) = \frac{F(\gamma x, t, \phi)}{\gamma}.$$

Consequently, Bayesian inference is performed on the approximate model $\dot{\mathcal{M}}_h$ deduced from the model \mathcal{M}_h using the same reparameterization:

$$\left. \begin{aligned} Y_j &= \gamma \dot{W}_{n_j} + \varepsilon_j, & 0 \leq j \leq J, \\ \varepsilon_j &\sim_{i.i.d} \mathcal{N}(0, \sigma^2), \\ h_n &= \tau_n - \tau_{n-1}, \\ \dot{W}_n &= \dot{W}_{n-1} + h_n \dot{F}(\dot{W}_{n-1}, \tau_{n-1}, \Phi) + \sqrt{h_n} \xi_n, & 1 \leq n \leq N, \\ \xi_n &\sim_{i.i.d} \mathcal{N}(0, 1), \\ \Phi &\sim \pi(\cdot; \beta). \end{aligned} \right\} (\dot{\mathcal{M}}_h)$$

A Gibbs algorithm based on this reparameterization is described below. The posterior distributions can be written as:

- $q_{\Phi, \dot{W}|Y, \theta}(\phi, \dot{w}|y, \theta) \propto q_{Y|\dot{W}}(y|\dot{w}; \gamma, \sigma^2)q_{\dot{W}|\Phi}(\dot{w}|\phi, \gamma^2)p(\phi; \beta),$
- $q(\sigma^2|y, \phi, \dot{w}, \beta, \gamma^2) \propto q_{Y|\dot{W}}(y|\dot{w}; \gamma^2, \sigma^2)p_\sigma(\sigma^2),$
- $q(\gamma^2|y, \phi, \dot{w}, \sigma^2, \beta) \propto q_{Y|\dot{W}}(y|\dot{w}; \gamma^2, \sigma^2)q_{\dot{W}|\Phi}(\dot{w}|\phi, \gamma^2)p_\gamma(\gamma^2),$
- $q(\beta|y, \phi, \dot{w}, \sigma^2, \gamma^2) \propto p(\phi; \beta)p_\beta(\beta).$

These conditional distributions provide the basis for the algorithm, alternating between updating (ϕ, \dot{w}) , β , γ^2 and σ^2 according to their conditional posterior distributions. The iterative Gibbs sampling algorithm is outlined as follows

- STEP 1: initialize the iteration counter of the chain $k = 1$ and start with initial values $\sigma^{-2(0)}, \gamma^{-2(0)}, \beta^{(0)}, \phi^{(0)}, \dot{w}^{(0)}$.
- STEP 2: obtain a new value $\sigma^{-2(k)}, \gamma^{-2(k)}, \beta^{(k)}, \phi^{(k)}, \dot{w}^{(k)}$ from $\sigma^{-2(k-1)}, \gamma^{-2(k-1)}, \beta^{(k-1)}, \phi^{(k-1)}, \dot{w}^{(k-1)}$ through successive generation of values
 1. $\sigma^{-2(k)} \sim q(\sigma^{-2}|y, \phi^{(k-1)}, \dot{w}^{(k-1)}, \beta^{(k-1)}, \gamma^{-2(k-1)})$
 2. $\gamma^{-2(k)} \sim q(\gamma^{-2}|y, \phi^{(k-1)}, \dot{w}^{(k-1)}, \sigma^{-2(k)}, \beta^{(k-1)})$
 3. $\beta^{(k)} \sim q(\beta|y, \phi^{(k-1)}, \dot{w}^{(k-1)}, \sigma^{-2(k)}, \gamma^{-2(k)})$
 4. $(\phi^{(k)}, \dot{w}^{(k)}) \sim q(\phi, \dot{w}|y, \sigma^{-2(k)}, \gamma^{-2(k)}, \beta^{(k)})$
- STEP 3: change the counter from k to $k + 1$ and return to STEP 2 until convergence is reached.

Updating β , γ^2 and σ^2 can be carried out using standard M-H algorithms and is not discussed in detail here. Updating (ϕ, \dot{w}) is less straightforward and is detailed in Section 7.2.4 in the case of the basic data augmentation. This procedure is easily adjustable to the reparameterization case by using the conditional distributions detailed previously. For a practical implementation, we recommend the paper of [108] which can be adapted to the model \mathcal{M}_h .

The convergence of this Gibbs algorithm is proved in the following theorem:

Theorem 2. *Let p_θ and $q_{\theta|Y}$ be respectively the prior and the posterior distributions of θ on the model \mathcal{M}_h .*

Assuming the proposal distributions specified above, the hybrid Gibbs algorithm detailed previously converges and provides an ergodic Markov Chain generated with the posterior distribution $q_{\theta|Y}$.

Proof: As previously detailed in Section 7.2.4, the ergodicity of the Markov Chain is ensured if one of the proposal distributions of the cyclic combination fulfills a minoration condition detailed in [121], which is the case with the proposal distributions used to generate (ϕ, \dot{w}) .

7.4 Survey of the error induced by the Euler-Maruyama approximation

Both estimation methods proposed in this paper, respectively the maximum likelihood and the Bayesian schemes, generate two distinct types of errors on the parameter estimates that have to be controlled.

The first type of error is induced by the estimation method itself. For the maximum likelihood approach, the estimation algorithm produces a sequence $(\hat{\theta}_k)_{k \geq 0}$ of estimates which converges towards θ_h^* , a (local) maximum of the \mathcal{M}_h -likelihood $q_Y(y; \cdot)$ function. The limiting distribution of the estimate is a rather delicate issue, which is beyond the scope of this paper. However the variance of this estimate $\hat{\theta}_k$ is classically controlled by the standard error which is evaluated through the Fisher information matrix of the estimates. Kuhn and Lavielle [72] propose to estimate this Fisher information matrix by using the stochastic approximation procedure and the Louis' missing information principle [80]. This Fisher information matrix estimate can be adapted to the model (\mathcal{M}_h) for a fixed value of the step size h . At the k -th iteration of the SAEM algorithm, the 3 following quantities are evaluated:

$$\Delta_k = \Delta_{k-1} + \alpha_k \left[\partial_\theta \log q_{Y,W,\Phi}(y, w^{(k)}, \phi^{(k)}; \hat{\theta}_k) - \Delta_{k-1} \right] \quad (7.4.3)$$

$$G_k = G_{k-1} + \alpha_k \left[\partial_\theta^2 \log q_{Y,W,\Phi}(y, w^{(k)}, \phi^{(k)}; \hat{\theta}_k) + \partial_\theta \log q_{Y,W,\Phi}(y, w^{(k)}, \phi^{(k)}; \hat{\theta}_k) \partial_\theta \log q_{Y,W,\Phi}(y, w^{(k)}, \phi^{(k)}; \hat{\theta}_k)^t - G_{k-1} \right] \quad (7.4.4)$$

$$H_k = \Delta_k \Delta_k^t - G_k \quad (7.4.5)$$

As the sequence $(\hat{\theta}_k)_{k \geq 0}$ converges to the maximum of the likelihood, the sequence $(H_k)_{k \geq 0}$ converges to the observed Fisher information matrix. The diagonal elements of the inverse of this matrix provide estimates of the variance of the estimate θ_h^* .

For the Bayesian approach, the equivalent of the problem of obtaining the standard errors is to obtain estimated variances for the posterior mean $E(\theta|y)$. Gilks et al. [52] propose MCMC convergence diagnostics tools in their book, the simplest and more generally used one is independent parallel simulations mixed in together. Different variance estimates are proposed such as the effective sample size estimate or the batching approach, that provide 95% confidence interval for $E(\theta|y)$ (see Carlin and Louis [28] for a review of such methods). Because this type of error is not specific to the situation exposed in this paper, it is not further discussed here.

A second type of error is induced on the estimates by the Euler-Maruyama scheme. Indeed, for the reasons evoked in Section 7.1, the estimation algorithms are applied to the model \mathcal{M}_h instead of to the model \mathcal{M} . In the maximum likelihood approach, the algorithm maximizes the \mathcal{M}_h -likelihood function q_Y instead of the \mathcal{M} -likelihood function p_Y ; in the Bayesian framework, the parameters are generated under the posterior distribution $q_{\theta|Y}$ instead of the posterior distribution $p_{\theta|Y}$.

The aim of this section is to study this second type of error induced by the Euler-Maruyama scheme on the conditional distribution $q_{W,\Phi|Y}$, on the likelihood function q_Y and on the posterior distribution $q_{\theta|Y}$. In Theorem 3 we propose bounds of these three errors as functions of the maximal step size of the Euler-Maruyama scheme h . In the following, some additional assumptions hold:

Assumption (A3):

The function $F : \mathbb{R} \times [t_0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}$ is infinitely differentiable in the variable space and its partial derivatives of any order are uniformly bounded with respect to x and ϕ .

Assumption (A4):

The assumption (UH) of Bally and Talay [9] is satisfied. More precisely, let A_0 and A_1 denote the vector fields defined respectively by $A_0 = F(\cdot)\partial_z$ and $A_1 = \gamma\partial_z$. For multiindices $a = (a_1, \dots, a_\ell) \in \{0, 1\}^\ell$, let the vector fields A_1^a be defined by induction: $A_1^\emptyset := A_1$ and for $j = 0$ or 1 $A_1^{(a,j)} := [A_1^j, A_1^a]$ where $[\cdot, \cdot]$ denotes the Lie bracket. For $L \leq 1$, let define the quadratic form $V_L(\xi, \eta) := \sum_{|a| \leq L-1} \langle A_1^a(\xi), \eta \rangle^2$. We assume that

$$V_L(\xi) := 1 \wedge \inf_{\|\eta\|=1} V_L(\xi, \eta) \geq 0$$

Theorem 3. *Let the assumptions (A0-A4) hold.*

1. *Let Z and W be the diffusion processes of the models \mathcal{M} and \mathcal{M}_h respectively, at the observation times: $Z = (Z(t_0), \dots, Z(t_J))$ and $W = (W(t_0), \dots, W(t_J))$.*

Let $p_{Z,\Phi|Y}$ and $q_{W,\Phi|Y}$ be the conditional distributions on the models \mathcal{M} and \mathcal{M}_h respectively. There exists a constant $C(y)$ such that, for any $0 < h < H_0$,

$$\|p_{Z,\Phi|Y} - q_{W,\Phi|Y}\|_{TV} \leq C(y)h,$$

where $\|\cdot\|_{TV}$ denotes the total variation distance.

2. *Let p_Y and q_Y be the likelihoods of the models \mathcal{M} and \mathcal{M}_h respectively. There exists a constant $C_2(y)$ such that for all $0 < h < H_0$*

$$\sup_{\{\theta=(\beta,\gamma^2,\sigma^2),\sigma^2>\sigma_0^2,\gamma^2>\gamma_0^2\}} |p_Y(y;\theta) - q_Y(y;\theta)| \leq C_2(y)h.$$

3. *In the Bayesian approach, let p_θ denote the prior distribution. Let $p_{\theta|Y}$ and $q_{\theta|Y}$ be the posterior distributions of the models \mathcal{M} and \mathcal{M}_h respectively. There exists a constant $C_3(y)$ such that for*

all $0 < h < H_0$

$$\|q_{\theta|Y} - p_{\theta|Y}\|_{TV} \leq C_3(y)h.$$

Theorem 3 is proved in Appendix 7.7. These results are based on the convergence rate of the transition densities proposed by Bally and Talay [9].

Remark 4. Assumption (A3) requires only the derivatives of the function F to be bounded and not F itself. Assumption (A4) is easily satisfied for linear drift functions $F: F(x, t, \phi) = A(\phi, t) + xB(\phi, t)$.

Remark 5. • Hence, as a principal consequence of part (2) of this theorem, and assuming regularity hypotheses on the Hessians of the likelihoods of both models \mathcal{M} and \mathcal{M}_h , the bias of the estimates induced by both the numerical approximation and the estimation algorithm, is controlled.

- Similarly, as a principal consequence of part (3) of this theorem, in the Bayesian framework, the bias on the posterior mean is controlled: under moments hypotheses on $p_{\theta}(\theta)$ and $q_{\theta|Y}(\theta)$, there exists a constant C'_y such that $|E_{\theta|y}[\theta] - E_{h,\theta|y}[\theta]| = |\int \theta p_{\theta|Y}(\theta)d\theta - \int \theta q_{\theta|Y}(\theta)d\theta| \leq C'_y h$ where $E_{\theta|y}[\theta]$ and $E_{h,\theta|y}[\theta]$ are the expectation under the posterior distributions $p_{\theta|y}$ and $p_{h,\theta|y}$ respectively. Similar result can be obtained for the bias of the posterior mode, assuming regularity hypotheses on the distributions $p_{Y|\theta}$ and $q_{Y|\theta}$ of both models \mathcal{M} and \mathcal{M}_h .

7.5 Theophyllin pharmacokinetic example

The maximum likelihood estimation method developed in Section 7.2 is applied below to a pharmacokinetics example.

7.5.1 Pharmacokinetics and Non-linear mixed effects models

Pharmacokinetics (PK) studies the time course of drug substances in the organism. This can be described through dynamic systems, the human body being assimilated to a set of compartments within which the drug evolves with time. In general, these systems are considered in their deterministic version. However, in a recent book on PK modeling, Krishna [71] claims that the fluctuations around the theoretical pharmacokinetic dynamic model may be appropriately modeled by using SDEs rather than ODEs. Overgaard et al. [95] suggest the introduction of SDEs to consider serial correlated residual errors due for example to erroneous dosing, sampling history or structural model misspecification. This new variability is distinct from the standard measurement noise representing the experimental uncertainty such as assay error.

Generally, several patients are followed up in a clinical trial, their drug concentration being measured along time repeatedly. Longitudinal data are thus gathered at discrete times and classically analyzed using non-linear mixed-effects models. Indeed, the mixed-effects models are a means to discriminate the intra-subject variability from the inter-subject variability, the parameter ϕ being a random parameter proper to each subject. The non-linear mixed-effects model can be written as follows:

$$\left. \begin{aligned} y_{ij} &= Z(t_{ij}, \phi_i) + \varepsilon_{ij} \\ \varepsilon_{ij} &\sim \mathcal{N}(0, \sigma^2) \\ \phi_i &\sim \mathcal{N}(\mu, \Omega) \end{aligned} \right\} (\mathcal{M}_{mix})$$

where y_{ij} is the observation for subject i , $i = 1, \dots, I$ at time t_{ij} , $j = 1, \dots, J_i$ and ϕ_i is the vector of individual and non-observed parameters of subject i .

In a deterministic approach, the regression function Z is defined as the solution of a PK ordinary differential system: $dZ(t)/dt = F(Z(t), t, \phi)$ with $Z(t_0) = Z_0$, each component of the vector ϕ having

a PK meaning. For example, a classic one compartment PK model with first order absorption and first order elimination is described by the following dynamic equation: $Z_0 = 0$ and

$$\frac{dZ(t, \phi)}{dt} = \frac{Dose \cdot K_a K_e}{Cl} e^{-K_e t} - K_e Z(t, \phi), \quad (7.5.6)$$

where Z is the drug concentration, $Dose$ is the known drug oral dose received by the subject, K_e is the elimination rate constant, K_a is the absorption rate constant and Cl is the clearance of the drug. A stochastic differential system can be deduced from the ODE:

$$dZ(t, \phi) = \left(\frac{Dose \cdot K_a K_e}{Cl} e^{-K_e t} - K_e Z(t, \phi) \right) dt + \gamma dB_t \quad (7.5.7)$$

where B_t is a Brownian motion and γ is the volatility coefficient of the SDE.

In its SDE version, the non-linear mixed-effects model (\mathcal{M}_{mix}) is a particular case of the model \mathcal{M} previously presented i.e. a diffusion process is observed at discrete times with noise measurement and its drift function parameters are random.

7.5.2 Simulation study

The aim of this simulation study is to illustrate the accuracy (bias and root mean square errors) of the extended SAEM algorithm developed in Section 7.2.2 on a PK application.

We use the previous PK model to mimic the Theophyllin drug pharmacokinetic. To prevent the parameters from taking unrealistic negative values, the vector $\phi \in \mathbb{R}^3$ is classically composed of the log parameters $\phi = (\log(K_e), \log(K_a), \log(Cl))$. The individual parameters ϕ are thus simulated with Gaussian distributions $\mathcal{N}(\mu, \Omega)$, with μ equal to $(-2.52, 0.40, -3.22)$ as proposed by [99]. A diagonal variance-covariance matrix Ω is assumed for the Gaussian distribution of ϕ . Let $\omega^2 = (\omega_1^2, \omega_2^2, \omega_3^2)$ denote the vector of these variances. The inter-subject variability is set equal for the three parameters: $\omega_1^2 = \omega_2^2 = \omega_3^2 = 0.01$, corresponding to a variation coefficient of 10%. We set a volatility coefficient equal to $\gamma^2 = 0.2$ and an additive Gaussian measurement error $\sigma^2 = 0.1$. We generate 100 datasets with $I = 36$ subjects and with nine blood samples per patient ($J = 8$), taken at 15 minutes, 30 minutes, 1, 2, 3.5, 5, 7, 9, 12 hours after dosing. The drug oral dose ($Dose$) received by the subject is chosen arbitrarily between 3 and 6 mg.

To evaluate the accuracy of the estimates of $\theta = (\mu, \omega^2, \gamma^2, \sigma^2)$ produced by the SAEM algorithm, the estimation of the parameters is performed on the 100 simulated datasets using the extension of the SAEM algorithm presented in Section 7.2.2.

The Euler-Maruyama scheme included in the SAEM algorithm is implemented on a grid with auxiliary latent data points introduced between each pair of observation instants as detailed in Section 7.1.3. The number of auxiliary points has to be chosen carefully because a volume of missing data too large can induce arbitrarily poor convergence properties of the Gibbs algorithm. In this example, we divide each time interval $[t_{i,j}, t_{i,j+1}]$ into 20 sub-intervals of equal length. This choice supplies a reasonable volume of missing data, avoids unbalance between the observation-time intervals and proves its numerical efficiency in accurately approximating the solution of the SDE.

The implementation of the Gibbs procedure included in the SAEM algorithm requires subtle tuning in practice. In particular, the simulation of the diffusion process w on the auxiliary grid is highly critical. An unconditioned trajectory simulation with $q(w_{n_j} | w_{n_{j-1}}; \theta)$ as proposed by Pedersen [97] provides poor numerical results in the case of this example. Indeed, a great number of these simulated trajectories produce large jumps ($w_{\tau_{n_j}} - w_{\tau_{n_j-1}}$). The probability of such trajectories being close to zero, it induces too low an acceptance rate. As suggested by Eraker [41] or Roberts and Stramer [108] and detailed in Section 7.2.4, a conditioned trajectory simulation through Brownian bridge distributions is preferred. Moreover, we update the missing trajectories at once for each subject, as recommended by Elerian et al. [40] to avoid a high level of rejection. In this example, we obtain acceptance rates in the neighborhood of 25%.

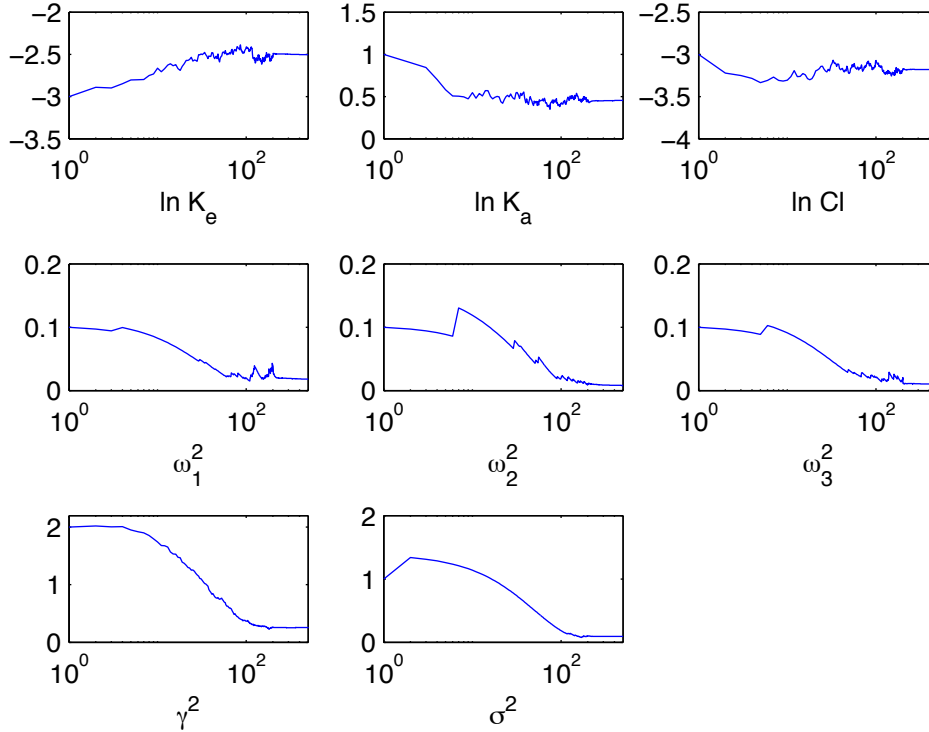


Figure 7.1: Evolution of the SAEM parameter estimates function of the iteration number in a logarithmic scale

The implementation of the SAEM algorithm requires initial values and the choice of the stochastic approximation sequence $(\alpha_k)_{k \geq 0}$. The initial values of the parameters are chosen arbitrarily and set to $\theta_0 = (-3, 1, -3, 0.1, 0.1, 0.1, 2, 1)$. The step of the stochastic approximation scheme is chosen as recommended by Kuhn and Lavielle [73]: $\alpha_k = 1$ during the first iterations $1 \leq k \leq K_1$, and $\alpha_k = (k - K_1)^{-1}$ during the subsequent iterations. Indeed, the initial guess θ_0 might be far from the maximum likelihood value and the first iterations with $\alpha_k = 1$ allow the sequence of estimates to converge to a neighborhood of the maximum likelihood estimate. Subsequently, smaller step sizes during $K - K_1$ additional iterations ensure the almost sure convergence of the algorithm to the maximum likelihood estimate. We implement the extended SAEM algorithm with $K_1 = 200$ and $K = 500$ iterations. Figure 1 illustrates the convergence of the parameter estimates provided by the extended SAEM algorithm as a function of the iteration number in a logarithmic scale. During the first iterations of SAEM, the parameter estimates fluctuate, reflecting the Markov chain construction. After 200 iterations, the curves smooth out but still continue to converge towards a neighborhood of the likelihood maximum. Convergence is obtained after 500 iterations.

The relative bias and relative root mean square error (RMSE) for each component of θ are computed and presented in Table 1.

The estimates of the mean parameter μ have very low bias ($< 5\%$). The variance parameters have small bias ($< 9\%$) except γ^2 , this variance parameter being slightly over-estimated (13%). The RMSE are very satisfactory for the mean parameter ($< 9\%$). The RMSE for the variance parameters are greater but still satisfactory ($\leq 40\%$) in comparison to the small number of subjects ($I = 36$). The RMSE of σ^2 is particularly satisfactory ($< 20\%$) considering the complexity of the variability model.

Table 7.1: Relative bias (%) and relative root mean square error (RMSE) (%) of the estimated parameters evaluated by the SAEM algorithm from 100 simulated trials with $I = 36$ subjects.

Parameters	Bias (%)	RMSE (%)
$\log K_e$	0.42	-3.19
$\log K_a$	4.14	8.95
$\log Cl$	-0.23	-2.27
ω_1^2	3.83	40.03
ω_2^2	8.49	36.76
ω_3^2	-8.81	37.52
γ^2	13.02	21.31
σ^2	-4.44	18.79

In conclusion, even if this simulation study is performed on a complex model, the convergence of the extended SAEM algorithm towards the maximum likelihood neighborhood is computationally efficient. In addition despite the fact that the number of subjects is small, the extended SAEM algorithm all in all supplies accurate estimations of the parameters. Furthermore, the accuracy is comparable to that obtained with the classic SAEM algorithm for an ODE version of a mixed model (\mathcal{M}_{mix}) i.e. for a model with one less variability level.

7.5.3 A real data example

The extended SAEM algorithm is used to estimate the PK parameters of the Theophyllin drug PK real dataset. This new analysis of the Theophyllin dataset aims at illustrating the advantage of the SDE approach over the ODE approach.

In this clinical trial, twelve subjects received a single oral dose of 3 to 6 mg of Theophyllin. Ten blood samples were taken 15 minutes, 30 minutes, 1, 2, 3.5, 5, 7, 9, 12 and 24 hours after dosing. The individual data are displayed in Figure 7.2. The Theophyllin PK is classically described by the one compartment model with first order absorption and first order elimination presented previously. We fit the Theophyllin data with the regression term successively defined as the solution (7.5.6) and then as that of the SDE (7.5.7).

In the ODE approach, the differential equation (7.5.6) has an explicit solution. Thus, the parameters estimates are obtained using the SAEM algorithm combined with a MCMC procedure proposed by Kuhn and Lavielle [72]. The individual concentration profiles are predicted by $\hat{Z}_{ij} = Z(t_{ij}, \hat{\phi}_i)$ for all i and j where Z is the solution of (7.5.6) and $\hat{\phi}_i$ is the posterior mean evaluated during the last iterations of the SAEM algorithm.

In the SDE approach, the same implementation of the extended SAEM algorithm as the one detailed for the simulation study (i.e. with an Euler-Maruyama approximation of the SDE) is used. The individual concentration predictions $E(W(t_{ij}, \phi_i)|y_i; \hat{\theta})$ for all i and j are evaluated by $\hat{W}_{ij} = 1/100 \sum_{k=K-99}^K W^{(k)}(t_{ij}, \phi_i^{(k)})$ where $W^{(k)}(t_{ij}, \phi_i^{(k)})$ is simulated under the conditional distribution $q_{W, \Phi|Y}(\cdot | y_i; \hat{\theta})$ during the 100 last iterations of the extended SAEM algorithm.

However, as the differential equation (7.5.7) is linear, an exact simulation of the diffusion process Z can also be performed and be combined with the SAEM algorithm. The individual concentration predictions $E(Z(t_{ij}, \phi_i)|y_i; \hat{\theta})$ for all i and j are evaluated by $\hat{Z}_{ij} = 1/100 \sum_{k=K-99}^K Z^{(k)}(t_{ij}, \phi_i^{(k)})$ where $Z^{(k)}(t_{ij}, \phi_i^{(k)})$ is simulated under the conditional distribution $p_{Z, \Phi|Y}(\cdot | y_i; \hat{\theta})$ during the 100 last iterations of the extended SAEM algorithm. As a consequence, on this particular example, the influence of the Euler approximation on the predictions can be estimated.

The SAEM algorithm is implemented with 500 iterations for the 3 models. The step size for the

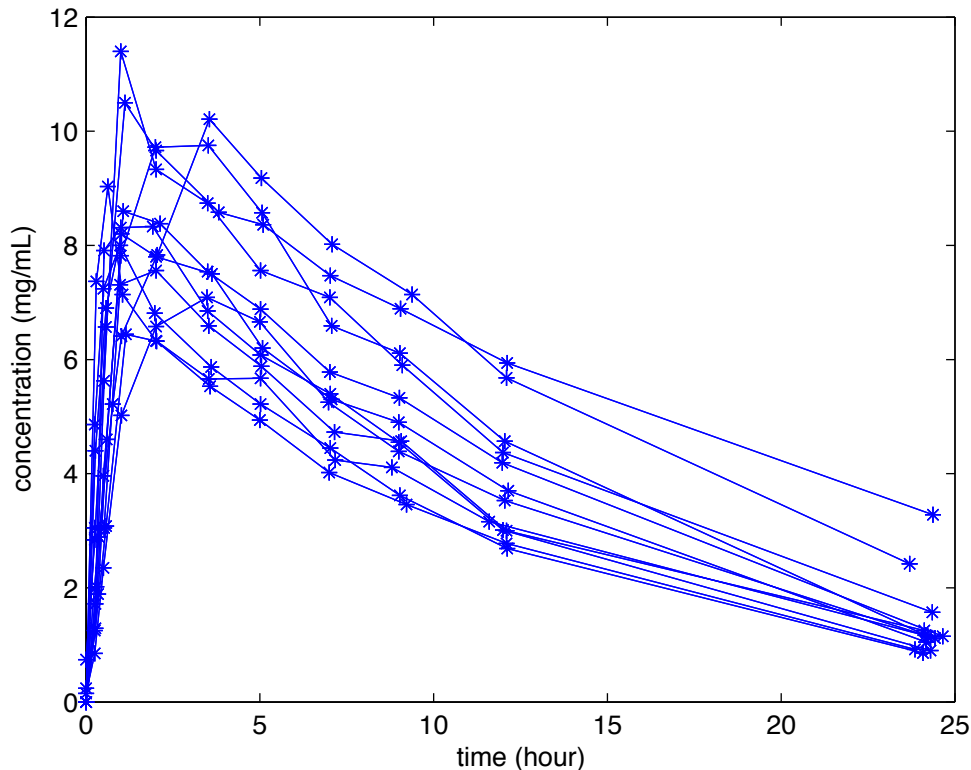


Figure 7.2: Individual concentrations for the pharmacokinetics of Theophyllin for 12 subjects.

Euler-Maruyama approximation is such that 100 latent points are introduced between each pair of observations instants.

The ODE and the two SDE predictions are overlaid on the data in Figure 7.3 for four typical subjects. Both ODE and SDE predicted curves for the other eight subjects are satisfactory and thus not presented here.

First, the results obtained by the Euler-Maruyama approximation and the exact simulation based approaches illustrate the control of the error induced by the Euler-Maruyama approximation (result of Theorem 3), as the two predicted curves are almost always very closed. Moreover this graph illustrates that this error is insignificant in comparison of the contribution of the SDE approach. Secondly, for the subject 3, the ODE predicted curve is satisfactory as well as the two SDE predicted curves. For the subjects 1, 2 and 12, the ODE predicted curves miss some of the observed data, particularly the last one and the elimination phase of the drug. The SDE predicted curves improve all of these individual profiles, especially the elimination phase.

In conclusion, in this real dataset case study, the individual predictions supplied by the SDE model fit the data better than those obtained by the ODE model. Consequently, in this case, the SDE approach has to be preferred to the ODE approach.

7.6 Discussion

This paper proposes estimation methods for models defined by a discretely observed diffusion process including additive measurement noise and with random drift function parameters. To that end, an approximate model \mathcal{M}_h is introduced, of which the regression term is evaluated using a Gaussian Euler-

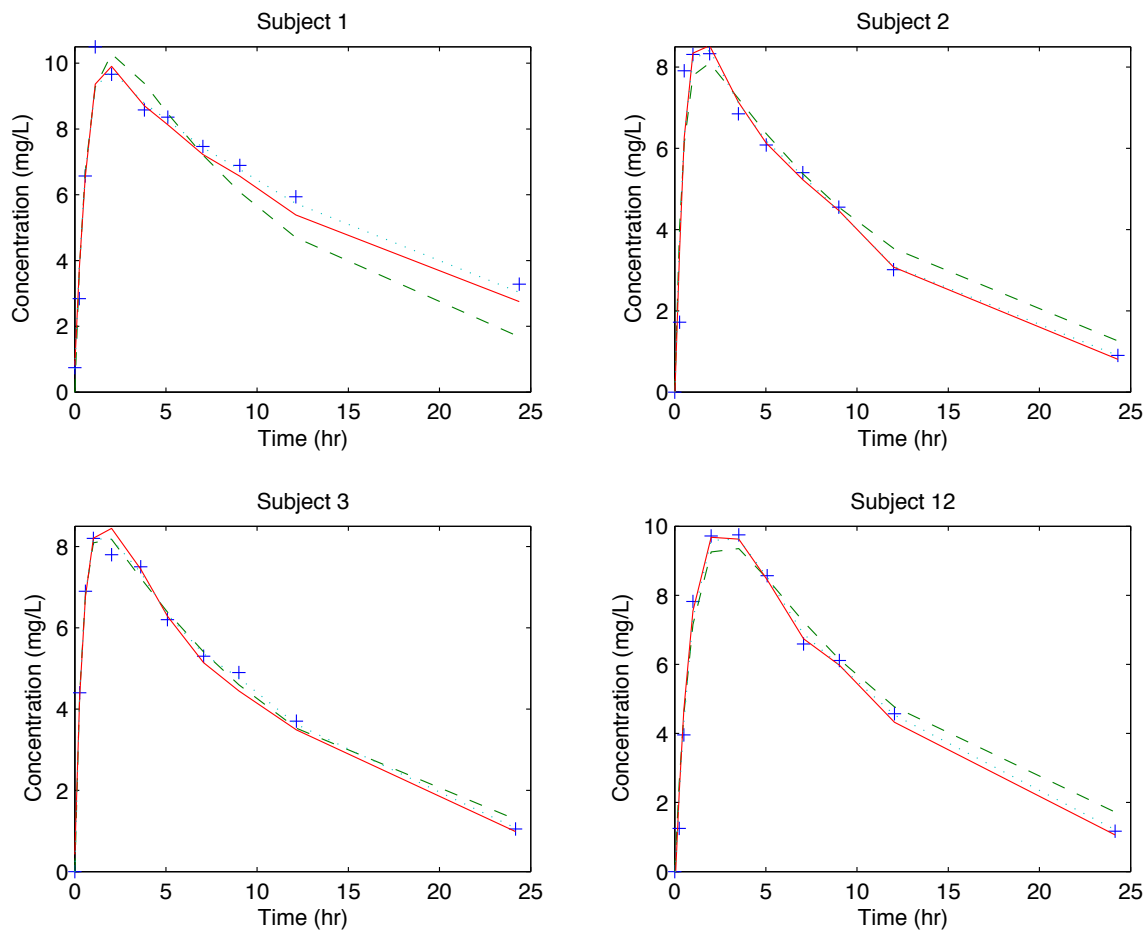


Figure 7.3: Four individual concentration curves predicted by SAEM with the ODE approach (dotted line), the SDE approach based on the Euler-Maruyama approximation (plain line) and the SDE approach based on an exact simulation of the diffusion (dashed line) overlaid on the data points for the pharmacokinetic of Theophyllin

Maruyama approximation of maximal step size h . A Gibbs sampler based on the reparameterization of the model suggested by Roberts and Stramer [108] in the Bayesian framework and the SAEM algorithm in the Maximum Likelihood approach are extended to this model. These two estimation algorithms require the simulation of the missing data (w, ϕ) with the conditional distribution $q_{W, \Phi|Y}$. The choice of the proposal distributions governs the convergence properties of the algorithm and thus is a key issue. A tuned MCMC procedure to perform this simulation is thus proposed, combining a hybrid Gibbs algorithm with independent or random walk Metropolis-Hastings schemes.

Moreover, we prove that the error induced by the Euler-Maruyama Gaussian approximation and measured in the total variation distance of the conditional distributions, the likelihoods and the posterior distributions decreases linearly when the step size h of the numerical scheme goes to zero. However, when the step size h decreases and the number of auxiliary latent times increases, the simulation of the diffusion process becomes more difficult. Therefore, a trade-off between a small discretisation error and a small simulation variance has to be found. An extension of this work would be a theoretical result to find this trade-off. As this result is beyond the scope of this paper, we used in the simulation example an empirical trade-off.

The discretisation error is distinct from the error on the estimates induced by the estimation algorithms, which is classically evaluated through the Fisher information matrix in the classical approach. This Fisher information matrix is estimated by stochastic approximation using the Louis' missing information principle [80]. This matrix provides standard errors of the estimates. The limiting distribution of the estimates has been described by Delyon et al. [35] for an averaged SAEM procedure. However this result can not be applied when the SAEM algorithm is combined with a MCMC algorithm because, the random noise in the stochastic approximation scheme is not anymore a martingale increment. In the case of SAEM-MCMC algorithm, the limiting distribution of the estimates is difficult to derive and this problem is beyond the scope of this paper.

In the maximum likelihood approach, the stochastic version of the EM algorithm SAEM proposed by Kuhn and Lavielle [72] is preferred to the Monte-Carlo EM (MCEM) developed by Wei and Tanner [127] or Wu [134] because of its computational properties. Indeed, SAEM requires the generation of only one realization of the non-observed data at each iteration. In a context where the missing data have to be simulated by a MCMC method, decreasing the size of these missing data is a key issue to ensure acceptable computational times.

The accuracy of the extended SAEM algorithm is illustrated on a pharmacokinetic simulation study using a non-linear mixed effect model defined by SDEs. The parameters are estimated with small bias and the mean square error are satisfactory given the complexity of the model. The relevance of the SDEs approach with respect to the deterministic one is exemplified on a real dataset based on a linear SDE. The comparison between the exact simulation of the SDE and the Euler-Maruyama approximation illustrates the accuracy of the discretisation approach. Therefore, this example justifies the use of this discretisation approach for the parameter estimation of more general models without explicit solution.

7.7 Appendix A: Proof of theorem 3 (chapter 7)

1. The aim is to bound

$$\|p_{Z,\Phi|Y} - q_{W,\Phi|Y}\|_{TV} = \int |p_{Z,\Phi|Y}(x, \phi|y; \theta) - q_{W,\Phi|Y}(x, \phi|y; \theta)| dx d\phi,$$

Using the fact that the conditional distributions $p_{Y|Z}(x; \sigma^2)$ and $q_{Y|W}(x; \sigma^2)$ are equal, the Bayes theorem application provides:

$$\begin{aligned} & |p_{Z,\Phi|Y}(x, \phi|y; \theta) - q_{W,\Phi|Y}(x, \phi|y; \theta)| \leq \\ & \frac{p_{Y|Z}(x; \sigma^2)\pi(\phi; \beta)}{p_Y(y; \theta)} \left[|p_{Z|\Phi}(x|\phi; \gamma^2) - q_{W|\Phi}(x|\phi; \gamma^2)| + \frac{q_{W|\Phi}(x|\phi; \gamma^2)}{q_Y(y; \theta)} |p_Y(y; \theta) - q_Y(y; \theta)| \right] \end{aligned}$$

As a consequence, the total variation distance is bounded by:

$$\begin{aligned} \|p_{Z,\Phi|Y} - q_{W,\Phi|Y}\|_{TV} & \leq \frac{\int p_{Y|Z}(x; \sigma^2)\pi(\phi; \beta) dx d\phi}{p_Y(y; \theta)} \left[\sup_{x, \phi} |p_{Z|\Phi}(x|\phi; \gamma^2) - q_{W|\Phi}(x|\phi; \gamma^2)| \right. \\ & \left. + \frac{|p_Y(y; \theta) - q_Y(y; \theta)|}{q_Y(y; \theta)} \sup_{x, \phi} q_{W|\Phi}(x|\phi; \gamma^2) \right] \end{aligned} \quad (7.7.8)$$

(a) The quantity $\sup_{x, \phi} |p_{Z|\Phi}(x|\phi; \gamma^2) - q_{W|\Phi}(x|\phi; \gamma^2)|$ is bounded using a result demonstrated by [8]. This result, based on the Malliavin Calculus, controls the density convergence rate in the case of the Euler-Maruyama scheme.

- By the assumption **(A3)** and because the volatility function is constant, the Hörmander's condition detailed in [8] is verified. Thus, there exists a constant $C(\phi, \gamma^2, t_j - t_{j-1})$ independent of h , x_j and x_{j-1} such that

$$|p_{Z|\Phi}(x_j|x_{j-1}, \phi; \gamma^2) - q_{W|\Phi}(x_j|x_{j-1}, \phi; \gamma^2)| \leq C(\phi, \gamma^2, t_j - t_{j-1})h.$$

The constant depends on the bounds of the derivatives of the drift function, independent of ϕ under assumption **(A3)**. Besides, if γ^2 is contained in $[\gamma_0, \Gamma_0]$, there exists C_1 independent of γ^2 such that, for all $j = 1 \dots J$,

$$|p_{Z|\Phi}(x_j|x_{j-1}, \phi; \gamma^2) - q_{W|\Phi}(x_j|x_{j-1}, \phi; \gamma^2)| \leq C_1 h \quad (7.7.9)$$

- Under the assumption **(A4)** and using a result of [74] based on the Malliavin calculus (corollary 3.25), for all $j = 1 \dots J$, there exists a constant $C_2(\phi, \gamma^2, t_j - t_{j-1})$ such that

$$p_{Z|\Phi}(x_j|x_{j-1}, \phi; \gamma^2) \leq C_2(\phi, \gamma^2, t_j - t_{j-1})$$

By the same arguments as before, this constant is bounded independently of ϕ and γ^2 . Hence, there exists C_2 such that, for all $j = 1 \dots J$,

$$p_{Z|\Phi}(x_j|x_{j-1}, \phi; \gamma^2) \leq C_2 \quad (7.7.10)$$

- In addition, we can write:

$$\begin{aligned} & |q_{W|\Phi}(x_j|x_{j-1}, \phi; \gamma^2)| \\ & \leq |q_{W|\Phi}(x_j|x_{j-1}, \phi; \gamma^2) - p_{Z|\Phi}(x_j|x_{j-1}, \phi; \gamma^2)| + |p_{Z|\Phi}(x_j|x_{j-1}, \phi; \gamma^2)| \\ & \leq hC_1 + C_2. \end{aligned} \quad (7.7.11)$$

- Finally, the Markov property provides:

$$|p_{Z|\Phi}(x|\phi; \gamma^2) - q_{W|\Phi}(x|\phi; \gamma^2)| = \left| \prod_{j=1}^J p_{Z|\Phi}(x_j|x_{j-1}, \phi; \gamma^2) - \prod_{j=1}^J q_{W|\Phi}(x_j|x_{j-1}, \phi; \gamma^2) \right| \quad (7.7.12)$$

for any $j = 1 \dots J$. So, by combining the (7.7.12), (7.7.9), (7.7.10) and (7.7.11), there exists a bound C_3 independent of h , j and γ^2 such that

$$\sup_{x, \phi} |p_{Z|\Phi}(x|\phi; \gamma^2) - q_{W|\Phi}(x|\phi; \gamma^2)| \leq C_3 h \quad (7.7.13)$$

- (b) By the Markov decomposition of the probability $q_{W|\Phi}(x|\phi; \gamma^2)$, and using the inequality (7.7.11), there exists C_4 such that

$$\sup_{x, \phi} q_{W|\Phi}(x|\phi; \gamma^2) \leq C_4 \quad (7.7.14)$$

By integration and using the inequality (7.7.13), we have:

$$|p_Y(y; \theta) - q_Y(y; \theta)| \leq C_3 h \int p_{Y|Z}(x; \sigma^2) \pi(\phi; \beta) dx d\phi = C_3 h \quad (7.7.15)$$

- (c) The quantity $q_{W|\Phi}(x|\phi; \gamma^2)$ can be down-bounded. Indeed,

$$\begin{aligned} q_Y(y; \theta) &\geq p_Y(y; \theta) - |p_Y(y; \theta) - q_Y(y; \theta)| \\ &\geq p_Y(y; \theta) - C_3 h \quad \text{following the inequality (7.7.15)} \\ &\geq p_Y(y; \theta) - C_3 H_0 \quad \text{for } h < H_0 \text{ and } H_0 \text{ small enough.} \end{aligned}$$

Hence there exists $C_5(y)$ such that

$$q_Y(y; \theta) \geq C_5(y) \quad (7.7.16)$$

- (d) Finally, the inequalities (7.7.8), (7.7.14), (7.7.15) and (7.7.16) provide the final result:

$$\|p_{Z, \Phi|Y} - q_{W, \Phi|Y}\|_{TV} \leq \frac{1}{p_Y(y; \theta)} \left[C_3 h + \frac{C_3 h}{C_5(y)} C_4 \right]$$

□

- The proof of the part 2 of Theorem 3 directly derives from (7.7.15).
- By Bayes theorem, we have:

$$p_{\theta|Y}(\theta|y) = \frac{p_{Y|\theta}(y|\theta)p(\theta)}{p_Y(y)}$$

where $p_Y(y) = \int p_{Y|\theta}(y|\theta)p(\theta)d\theta$. From (7.7.15), there exists a constant C_3 , independent of θ such that $|p_{Y|\theta}(y|\theta) - q_{Y|\theta}(y|\theta)| \leq hC_3$. Consequently $|p_Y(y) - q_Y(y)| \leq C_3 h$ and

$$\begin{aligned} |p_{\theta|Y}(\theta|y) - q_{\theta|Y}(\theta|y)| &\leq \frac{p(\theta)}{p_Y(y)} \left| |p_{Y|\theta}(y|\theta) - q_{Y|\theta}(y|\theta)| + \frac{q_{Y|\theta}(y|\theta)}{q_Y(y)} |q_Y(y) - p_Y(y)| \right| \\ &\leq \frac{C_3 h}{p_Y(y)} p(\theta) \left| 1 + \frac{q_{Y|\theta}(y|\theta)}{p_Y(y)} \right| = C_6(y) h [p(\theta) + q_{\theta|Y}(\theta|y)]. \end{aligned}$$

The final result can be directly deduced:

$$\begin{aligned} \|q_{\theta|Y} - p_{\theta|Y}\|_{TV} &= \int |p_{\theta|Y}(\theta|y) - q_{\theta|Y}(\theta|y)| d\theta \\ &\leq C_6(y) h \int (p(\theta) + q_{\theta|Y}(\theta|y)) d\theta \leq 2 C_6(y) h \end{aligned}$$

□

Chapitre 8

Remarques sur l'identifiabilité des modèles définis par équations différentielles stochastiques à paramètres aléatoires

Sommaire

8.1	Position du problème	151
8.2	Cas d'une unique observation	153
8.3	Cas d'observations multiples	161
8.4	Conclusion et perspectives	167
8.5	Résultats techniques	168

Au chapitre 7 de cette thèse, nous nous sommes intéressés à l'estimation paramétrique dans des modèles définis par des équations différentielles stochastiques à paramètres aléatoires. Nous avons proposé une méthode d'estimation par maximum de vraisemblance des paramètres. Cependant, nous n'y avons pas abordé la difficile question de l'identifiabilité du modèle. Dans ce chapitre, nous proposons des éléments de réponse à cette question.

8.1 Position du problème

Simplifions tout d'abord le problème en considérant que le processus de diffusion est observé sans bruit de mesure. Le modèle statistique considéré est alors le suivant. Soient (Z_1, \dots, Z_J) les observations aux instants (t_1, \dots, t_J) d'un processus de diffusion Z dépendant d'un paramètre aléatoire $\Phi \in \mathbb{R}^d$. A paramètre Φ fixé, le processus de diffusion Z est solution de l'équation différentielle stochastique générale suivante :

$$\begin{cases} dZ_t &= F(Z_t, t, \Phi, \eta) dt + G(Z_t, t, \Phi, \gamma) dB_t, \\ Z_{t_0}(\Phi) &\sim F_0(\Phi), \end{cases} \quad (8.1.1)$$

où B_t est un mouvement brownien, $(\eta, \gamma) \in O \subset \mathbb{R}^p \times \mathbb{R}^q$, $F : \mathbb{R} \times [0, +\infty[\times \mathbb{R}^d \times \mathbb{R}^p \rightarrow \mathbb{R}$ et $G : \mathbb{R} \times [0, +\infty[\times \mathbb{R}^d \times \mathbb{R}^q \rightarrow \mathbb{R}$. De plus, Φ est aléatoire de loi de distribution

$$\Phi \sim \pi(\cdot; \beta)$$

avec $\beta \in \mathbb{R}^r$. Dans la suite, nous supposons que la loi de probabilité $\pi(\cdot; \beta)$ est absolument continue par rapport à la mesure de Lebesgue sur \mathbb{R}^d et, par abus de notation, nous noterons aussi $\pi(\cdot; \beta)$ la densité associée à cette probabilité.

Notons $\theta = (\eta, \gamma, \beta) \in \Theta \subset \mathbb{R}^{p+q+r}$. Ce paramètre se scinde en deux parties : $\theta = (\theta_1, \theta_2) \in \Theta_1 \times \Theta_2 \subset \mathbb{R}^{p_1} \times \mathbb{R}^{p_2}$. θ_1 est la partie connue du vecteur θ (autrement appelée *hyper-paramètre*) tandis que θ_2 est la partie inconnue du vecteur θ (paramètre du modèle à estimer). Nous cherchons donc à étudier l'identifiabilité du paramètre θ_2 .

Hypothèses sur F et G . Dans la suite, à Φ fixé, nous nous plaçons dans les conditions standard d'existence d'une unique solution $\{Z_t, t_0 \leq t \leq T\}$ de l'équation différentielle stochastique (8.1.1). Plus précisément,

1. *Condition de Lipschitz* : $\forall \Phi \in \mathbb{R}^d, \forall (\eta, \gamma) \in O \subset \mathbb{R}^{p+q}$, il existe $K > 0$ telle que pour tout $t_0 \leq t \leq T$, pour tout $(x, x') \in \mathbb{R}^2$

$$\begin{aligned} |F(x, t, \Phi, \eta) - F(x', t, \Phi, \eta)| &\leq K|x - x'| \\ |G(x, t, \Phi, \gamma) - G(x', t, \Phi, \gamma)| &\leq K|x - x'| \end{aligned}$$

2. *Croissance linéaire* : $\forall \Phi \in \mathbb{R}^d, \forall (\eta, \gamma) \in O \subset \mathbb{R}^{p+q}$, il existe $C > 0$ telle que pour tout $t_0 \leq t \leq T$, pour tout $x \in \mathbb{R}$

$$\begin{aligned} |F(x, t, \Phi, \eta)|^2 &\leq C^2(1 + |x|^2) \\ |G(x, t, \Phi, \gamma)|^2 &\leq C^2(1 + |x|^2) \end{aligned}$$

Sous ces hypothèses, pour tout $t_0 < t < T$, à paramètre Φ fixé, la distribution Z_t conditionnée par la filtration \mathcal{F}_{t-} est absolument continue par rapport à la mesure de Lebesgue sur \mathbb{R} (\mathcal{F}_{t-} étant la filtration engendrée par $\{Z_s, s < t\}$). Dans la suite, nous notons $p_{Z_t|\Phi}(\cdot|\phi; \eta, \gamma)$ la densité de cette distribution.

Soient (Z_1, \dots, Z_J) , les J observations du processus $\{Z_t, t_0 \leq t \leq T\}$ aux instants (t_1, \dots, t_J) . $\forall j = 2, \dots, J$, soit $p_{Z_j|Z_{j-1}, \Phi}(\cdot|z_{j-1}, \phi; \eta, \gamma)$ les probabilités de transition du processus. Alors, par la propriété de Markov, la distribution marginale du J -uplet s'écrit :

$$\begin{aligned} p_{Z_1, \dots, Z_J}(z_1, \dots, z_J; \theta) &= \int_{\phi \in \mathbb{R}^d} p_{Z_1, \dots, Z_J|\Phi}(z_1, \dots, z_J|\phi; \eta, \gamma) \pi(\phi; \beta) d\phi \\ &= \int_{\phi \in \mathbb{R}^d} \prod_{j=1}^J p_{Z_j|Z_{j-1}, \Phi}(z_j|z_{j-1}, \phi; \eta, \gamma) \pi(\phi, \beta) d\phi \end{aligned} \quad (8.1.2)$$

Définition 1. $\theta_2 \in \Theta_2$ est dit *identifiable* pour $p_{Z_1, \dots, Z_J}(\cdot; \theta_1, \theta_2)$ si et seulement si, pour tout couple $(\theta_2, \theta'_2) \in \Theta_2^2$

$$\begin{aligned} p_{Z_1, \dots, Z_J}(\cdot; \theta_1, \theta_2) &= p_{Z_1, \dots, Z_J}(\cdot; \theta_1, \theta'_2) \\ \iff \theta_2 &= \theta'_2 \end{aligned}$$

Lorsque le processus $\{Z_t, t_0 \leq t \leq T\}$ est solution d'une EDS, la distribution $p_{Z_1, \dots, Z_J|\Phi}(\dots|\Phi; \eta, \gamma)$ n'a, en règle générale, pas d'expression explicite. Ceci est particulièrement vrai dans le cas d'observations multiples ($J > 1$), son expression requérant alors le calcul des probabilités de transition. Dans les cas où cette expression explicite existe, le fait que le processus dépende du paramètre aléatoire Φ complique le problème. En effet, d'après la définition du modèle, la distribution marginale s'exprime comme un mélange de distributions (voir (8.1.2)). Notre capacité à expliciter celle-ci dépend alors de la distribution du paramètre aléatoire Φ , $\pi(\dots; \beta)$.

Dans ce chapitre, nous cherchons à exhiber des couples “équation différentielle stochastique / distribution π sur Φ ” permettant d’avoir d’une part, une expression explicite de la distribution marginale des observations et d’autre part l’identifiabilité des paramètres.

Nous nous plaçons tout d’abord dans le cas où $J = 1$ (unique observation) (partie 8.2) et considérons le processus de Ornstein-Uhlenbeck (8.2.1). Plusieurs modèles découlent directement de ce premier exemple. Nous considérons ensuite des EDS linéaires (8.2.2) puis le modèle de Cox-Ingersoll-Ross (8.2.3) ainsi que le processus brownien géométrique (8.2.4).

Dans le cas d’observations multiples ($J > 1$), nous montrons que les résultats obtenus sur le processus de Ornstein-Uhlenbeck dans le cas $J = 1$ ne peuvent être généralisés (8.3.1). Nous nous intéressons alors aux modèles à effets mixtes (8.3.2) : après avoir rappelé les résultats existants concernant les modèles linéaires et non-linéaires à effets mixtes, nous traitons un exemple de diffusion mixte avec et sans bruit d’observation pour lesquels nous sommes capables de montrer l’identifiabilité des paramètres.

Nous terminons ce chapitre sur des perspectives permettant d’envisager une poursuite de ce travail en cours.

8.2 Cas d’une unique observation

Considérons dans un premier temps le cas où les observations sont constituées d’une unique observation Z_t .

8.2.1 Le processus de Ornstein-Uhlenbeck

Soit un processus de Ornstein-Uhlenbeck, c’est-à-dire solution de l’équation différentielle stochastique suivante :

$$dZ_t = \Phi Z_t + \gamma dB_t, \quad \Phi < 0 \tag{8.2.3}$$

Alors, le processus solution est $Z_t = e^{\Phi t} Z_0 + \int_0^t \gamma e^{\Phi(t-s)} dB_s$, $\forall t > 0$. Ainsi, conditionnellement à Φ , sa distribution est :

$$Z_t | \Phi \sim \mathcal{N} \left(e^{\Phi t} Z_0, e^{2\Phi t} \text{Var}[Z_0] - \frac{\gamma^2}{2\Phi} [1 - e^{2\Phi t}] \right)$$

En ce plaçant sous le régime stationnaire, i.e. $Z_0 | \Phi \sim \mathcal{N} \left(0, -\frac{\gamma^2}{2\Phi} \right)$, nous obtenons pour distribution conditionnellement à Φ :

$$Z_t | \Phi \sim \mathcal{N} \left(0, -\frac{\gamma^2}{2\Phi} \right), \quad \forall t > 0. \tag{8.2.4}$$

Nous cherchons une loi sur Φ , permettant d’identifier les paramètres de sa distribution à partir de la loi marginale de Z_t . Pour cela, utilisons le lemme suivant :

Lemme 1. *Soit Φ une variable aléatoire de loi Γ de paramètres $(\frac{n}{2}, \frac{d}{2})$ et X une variable aléatoire à valeurs dans \mathbb{R} telle que sa distribution conditionnellement à Φ soit :*

$$X | \Phi \sim \mathcal{N} \left(m, \frac{c}{\Phi} \right)$$

avec $m \in \mathbb{R}$ et $c > 0$. Alors, posons $R = cd/n$. La distribution marginale de la variable aléatoire $\frac{X-m}{R^{1/2}}$ est une Student à n degrés de liberté :

$$\frac{X - m}{R^{1/2}} \sim T_n([0, 1])$$

La preuve du lemme 1 est donnée au paragraphe 8.5.1. Nous pouvons alors déduire le résultat suivant :

Résultat 1. Si Z_{t_1} est une observation à l'instant t_1 du processus suivant :

$$\begin{cases} dZ_t &= \Phi Z_t + \gamma dB_t \quad \forall t > 0 \\ Z_0 | \Phi &\sim \mathcal{N}\left(0, -\frac{\gamma^2}{2\Phi}\right) \\ -\Phi &\sim \Gamma\left(\frac{n}{2}, \frac{d}{2}\right) \end{cases}$$

Alors posons $R = \frac{\gamma^2 d}{2}$. Nous avons $\frac{Z_{t_1}}{R^{1/2}} \sim T_n([0, 1])$. De plus, si $\theta_2 = n \in \mathbb{R}^{*+}$ et $\theta_1 = (d, \gamma) \in \mathbb{R}^{*+} \times \mathbb{R}^*$ alors θ_2 est identifiable.

Démonstration. • **Distribution marginale** La distribution marginale de Z_{t_1} est une conséquence immédiate du lemme 1 et de l'expression de la distribution conditionnelle de $Z_{t_1} | \Phi$ (8.2.4).

- **Densité de Z_{t_1}** $Z_{t_1} = R^{1/2}T$ où $T \sim T_n([0, 1])$. La densité d'une Student à n degrés de liberté s'écrit :

$$p_T(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{n\pi}} \frac{1}{\left[1 + \frac{x^2}{n}\right]^{\frac{n+1}{2}}}$$

où Γ désigne la fonction Gamma. Par conséquent, pour toute fonction f mesurable bornée,

$$\begin{aligned} E[f(Z_{t_1})] &= E\left[f\left(R^{1/2}T\right)\right] \\ &= \int_{\mathbb{R}} f\left(R^{1/2}x\right) p_T(x) dx \\ &= \int_{\mathbb{R}} f\left(R^{1/2}x\right) \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{n\pi}} \frac{1}{\left[1 + \frac{x^2}{n}\right]^{\frac{n+1}{2}}} dx \\ &= \int_{\mathbb{R}} f(z) \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{n\pi}} \frac{1}{\left[1 + \frac{z^2}{Rn}\right]^{\frac{n+1}{2}}} \frac{1}{R^{1/2}} dz, \quad \text{en posant } z = R^{1/2}x. \end{aligned}$$

Ainsi, Z_{t_1} a pour densité :

$$p_{Z_{t_1}}(z; \theta) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{n\pi}} \frac{1}{\left[1 + \frac{z^2}{\frac{\gamma^2 d}{2}}\right]^{\frac{n+1}{2}}} \frac{1}{\sqrt{\frac{\gamma^2 d}{2}}} = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{\pi}} \frac{1}{\left[1 + \frac{z^2}{\frac{\gamma^2 d}{2}}\right]^{\frac{n+1}{2}}} \frac{1}{\sqrt{\frac{\gamma^2 d}{2}}}.$$

- **Identifiabilité**

Si $\theta_1 = (d, \gamma) \in \mathbb{R}^{*+} \times \mathbb{R}^*$ et $\theta_2 = n \in \mathbb{R}^{*+}$, l'identifiabilité du paramètre $\theta_2 = n$ est démontrable. En effet, soient (n, n') tels que pour $z \in \mathbb{R}$

$$p_{Z_{t_1}}(z; \theta) = p_{Z_{t_1}}(z; \theta')$$

Alors, d'après ce qui précède,

$$\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)\sqrt{\pi}} \frac{1}{\left[1 + \frac{z^2}{\frac{\gamma^2 d}{2}}\right]^{\frac{n+1}{2}}} \frac{1}{\sqrt{\frac{\gamma^2 d}{2}}} = \frac{\Gamma\left(\frac{n'+1}{2}\right)}{\Gamma\left(\frac{n'}{2}\right)\sqrt{\pi}} \frac{1}{\left[1 + \frac{z^2}{\frac{\gamma'^2 d'}{2}}\right]^{\frac{n'+1}{2}}} \frac{1}{\sqrt{\frac{\gamma'^2 d'}{2}}}$$

Donc en $z = 0$, nous obtenons :

$$\frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} = \frac{\Gamma\left(\frac{n'+1}{2}\right)}{\Gamma\left(\frac{n'}{2}\right)}$$

Or la fonction $x \mapsto \frac{\Gamma\left(\frac{x+1}{2}\right)}{\Gamma\left(\frac{x}{2}\right)}$ est strictement croissante sur \mathbb{R}^{*+} (voir lemme 3 au paragraphe 8.5.2).

Donc $n = n'$, $\theta_2 = n$ est identifiable. □

Remarque 22. • D'après l'expression de la densité de Z_{t_1} , il est très facile d'obtenir l'identifiabilité de $\theta_2 = d$ lorsque $\theta_1 = (n, \gamma)$ ou de $\theta_2 = \gamma$ lorsque $\theta_1 = (n, d)$. Cependant, montrer l'identifiabilité de n et d semble plus difficile mais pas impossible. La démonstration d'un tel résultat pourrait prolonger ce travail.

- Il pourrait paraître plus simple de chercher à identifier les paramètres à partir des moments de Z_{t_1} . Cependant, une Student à n degrés de liberté n'admet de moment non nul fini que si $n > 2$. Par conséquent, nous n'aurions pu montrer l'identifiabilité de n que sur un intervalle du type $]2, +\infty[$.

Ainsi, sur un exemple simple –le processus de Ornstein-Uhlenbeck– dans le cas d'une unique observation, nous avons été capable d'exhiber une distribution sur le paramètre de l'équation différentielle permettant d'explicitier la distribution marginale. A partir de là, nous avons pu prouver l'identifiabilité du paramètre n .

Autres applications du lemme 1 Notre raisonnement repose essentiellement sur le lemme 1. De la même façon, il est possible d'exhiber d'autres types de modèles pour lesquels nous montrons des résultats d'identifiabilité.

Résultat 2. Soit $b : D_b \subset \mathbb{R} \rightarrow \mathbb{R}$ une fonction dérivable sur son ensemble de définition et telle que $1/b$ soit continue de primitive $h : D_h \subset \mathbb{R} \rightarrow \mathbb{R}$ admettant une fonction réciproque. Soit b' sa fonction dérivée. Si Z_{t_1} est une observation à l'instant t_1 du processus suivant :

$$\begin{cases} dZ_t &= \frac{1}{2\Phi}b(Z_t)b'(Z_t) + \frac{1}{\sqrt{\Phi}}b(Z_t)dB_t, \quad \forall t > 0 \\ Z_0 &= cste \in D_h \\ \Phi &\sim \Gamma\left(\frac{n}{2}, \frac{d}{2}\right) \end{cases} \quad (8.2.5)$$

Alors,

$$\frac{h(Z_{t_1}) - h(Z_0)}{(t_1 n/d)^{1/2}} \sim T_n([0, 1]).$$

De plus, si $\theta_1 = d \in \mathbb{R}^{*+}$ et $\theta_2 = n \in \mathbb{R}^{*+}$, alors θ_2 est identifiable.

Remarque 23. Comme précédemment, le même résultat peut être obtenu avec $\theta_1 = n \in \mathbb{R}^{*+}$ et $\theta_2 = d \in \mathbb{R}^{*+}$.

Démonstration. En effet, sous les hypothèses sur la fonction b énoncées précédemment, $h(x) = \int^x \frac{1}{b(s)} ds$ admet une fonction réciproque. Par conséquent –voir [70] chapitre 4– pour tout $\Phi > 0$, l'équation stochastique $dZ_t = \frac{1}{2\Phi}b(Z_t)b'(Z_t) + \frac{1}{\sqrt{\Phi}}b(Z_t)dB_t$ admet pour solution :

$$Z_t = h^{-1}\left(\frac{1}{\sqrt{\Phi}}B_t + h(Z_0)\right) \quad \forall t > 0.$$

Ainsi, en t_1 , $h(Z_{t_1}) = \frac{1}{\sqrt{\Phi}}B_{t_1} + h(Z_0) \sim \mathcal{N}\left(h(Z_0), \frac{t_1}{\Phi}\right)$. Et, par le lemme 1, en prenant $\Phi \sim \Gamma\left(\frac{n}{2}, \frac{d}{2}\right)$, on obtient

$$\frac{h(Z_{t_1}) - h(Z_0)}{(t_1 n/d)^{1/2}} \sim T_n([0, 1]).$$

Par le même raisonnement que précédemment, nous obtenons l'identifiabilité des paramètres n ou d . \square

Exemples

- *Exemple 1* : $dZ_t = \frac{1}{2\Phi}Z_t + \frac{1}{\sqrt{\Phi}}Z_t dB_t$ i.e. $b(Z_t) = Z_t$. Alors, conditionnellement à Φ , $Z_t = Z_0 \exp\left[\frac{1}{\sqrt{\Phi}}B_t\right]$. En prenant, $\Phi \sim \Gamma(\frac{n}{2}, \frac{d}{2})$, on obtient :

$$\frac{\ln(Z_t) - \ln(Z_0)}{(tn/d)^{1/2}} \sim T_n([0, 1])$$

- *Exemple 2* : $dZ_t = -\frac{1}{2\Phi}Z_t + \frac{1}{\sqrt{\Phi}}\sqrt{1 - Z_t^2}dB_t$ i.e. $b(Z_t) = \sqrt{1 - Z_t^2}$ Alors

$$Z_t = \sin\left[\frac{1}{\sqrt{\Phi}}B_t + \arcsin(Z_0)\right].$$

En prenant, $\Phi \sim \Gamma(\frac{n}{2}, \frac{d}{2})$, on obtient :

$$\frac{\arcsin(Z_t) - \arcsin(Z_0)}{(tn/d)^{1/2}} \sim T_n([0, 1])$$

- *Exemple 3* : $dZ_t = \frac{1}{2\Phi} + \frac{1}{\sqrt{\Phi}}\sqrt{Z_t}dB_t$ i.e. $b(Z_t) = \sqrt{Z_t}$. Alors

$$Z_t = \left[\frac{1}{\sqrt{\Phi}}B_t + \sqrt{(Z_0)}\right]^2.$$

En prenant, $\Phi \sim \Gamma(\frac{n}{2}, \frac{d}{2})$, on obtient :

$$\frac{\sqrt{Z_t} - \sqrt{Z_0}}{(tn/d)^{1/2}} \sim T_n([0, 1])$$

8.2.2 EDS linéaire

Considérons l'équation stochastique générale suivante :

$$dZ_t = \alpha(b - Z_t)dt + \gamma Z_t dB_t \quad \alpha > 0, b > 0, \gamma \in \mathbb{R}^* \quad (8.2.6)$$

Remarque 24. *Au cours de ce paragraphe, nous allons étudier l'identifiabilité du modèle dans différentes configurations, l'aléa portant successivement sur α et b .*

Cette EDS étant linéaire, nous sommes capable de donner une expression explicite de la solution :

$$Z_t = e^{(-\alpha - \frac{1}{2}\gamma^2)t + \gamma B_t} Z_0 + b\alpha \int_0^t e^{(\alpha + \frac{1}{2}\gamma^2)(s-t) + \gamma(B_t - B_s)} ds$$

Cependant, cette expression ne nous permet pas d'avoir une expression simple de la loi Z_t conditionnellement aux paramètres. Nous nous plaçons donc tout d'abord dans le cas particulier où $b = 0$ puis sous le régime stationnaire.

Cas 1 : $b = 0$

Dans le cas où $b = 0$ et où le paramètre aléatoire est α (noté alors Φ), l'équation (8.2.6) se réduit à

$$dZ_t = -\Phi Z_t dt + \gamma Z_t dB_t.$$

Nous obtenons alors le résultat suivant :

Résultat 3. *Si Z_{t_1} est une observation à l'instant t_1 du processus suivant :*

$$\begin{cases} dZ_t &= -\Phi Z_t dt + \gamma Z_t dB_t \quad \forall t > 0 \\ Z_0 &= cste > 0 \\ \Phi &\sim \mathcal{N}(m, V) \end{cases}$$

Alors

$$\ln(Z_{t_1}) \sim \mathcal{N}\left(\ln(Z_0) - \left(m + \frac{1}{2}\gamma^2\right)t_1, \gamma^2 t_1 + V\right).$$

De plus si $\theta_1 = \gamma \in \mathbb{R}^*$ et $\theta_2 = (m, V) \in \mathbb{R} \times \mathbb{R}^{*+}$, alors θ_2 est identifiable.

Démonstration. En effet, à Φ fixé, par définition de l'EDS, $Z_t = e^{(-\Phi - \frac{1}{2}\gamma^2)t + \gamma B_t} Z_0, \forall t > 0$. Puisque Z_0 est une constante strictement positive, alors, conditionnellement à Φ ,

$$\ln(Z_t)|\Phi \sim \mathcal{N}\left(\ln(Z_0) - \left(\Phi + \frac{1}{2}\gamma^2\right)t, \gamma^2 t\right).$$

Par conséquent, si $m \in \mathbb{R}$, $V > 0$ et $\Phi \sim \mathcal{N}(m, V)$, nous obtenons une expression explicite de la loi marginale de l'observation :

$$\ln(Z_{t_1}) \sim \mathcal{N}\left(\ln(Z_0) - \left(m + \frac{1}{2}\gamma^2\right)t_1, \gamma^2 t_1 + V\right)$$

Alors si γ est un hyper-paramètre ($\theta_1 = \gamma$ et $\theta_2 = (m, V)$) nous sommes capable d'identifier θ_2 . \square

Remarque 25. Cet exemple sera repris dans le cas d'observations multiples, dans le cadre de diffusions mixtes. Nous montrerons que dans ce cas, nous sommes capables d'identifier les paramètres $\theta_2 = (m, V, \gamma)$ (voir paragraphe 8.3.2).

Ainsi, dans ce cas particulier, nous avons été capables de trouver une distribution sur Φ aboutissant à une expression explicite de la distribution marginale de Z_{t_1} ainsi qu'à l'identifiabilité des paramètres. Dans le cas général, nous n'avons pas d'expression explicite de la distribution conditionnelle $Z_{t_1}|\Phi$. Nous nous plaçons sous le régime stationnaire.

Cas 2 : sous le régime stationnaire

Considérons dorénavant que le paramètre aléatoire est b (noté Φ). L'équation s'écrit alors :

$$dZ_t = \alpha(\Phi - Z_t)dt + \gamma Z_t dB_t, \alpha > 0, \Phi > 0, \gamma \in \mathbb{R}^*$$

Rappels sur la distribution stationnaire Soit (Ω, \mathcal{F}, P) un espace probabilisé, $(\mathcal{F}_t)_{t \geq 0}$, une filtration de \mathcal{F} . Soit $(X_t)_{t \geq 0}$ un processus de diffusion (\mathcal{F}_t) adapté à valeurs dans \mathbb{R} , homogène en temps, unique solution sous P de l'EDS suivante :

$$dX_t = a(X_t)dt + b(X_t)dB_t, t \geq 0$$

où $x \mapsto a(x)$ est une fonction borélienne, bornée sur tout compact de \mathbb{R} , $x \mapsto b^2(x)$ est une fonction continue, strictement positive. Posons

$$s(x) = \exp\left[-2 \int^x \frac{a(z)}{b(z)^2} dz\right]$$

et $S(x) = \int_0^x s(z)dz$. Si $\lim_{x \rightarrow \pm\infty} S(x) = \pm\infty$, la diffusion est récurrente, de mesure invariante de densité

$$\frac{1}{s(\cdot)b(\cdot)^2}.$$

Détermination de la probabilité stationnaire à Φ fixé Posons $s(x) = \exp \left[- \int^x 2 \frac{\alpha(\Phi-y)}{(\gamma y)^2} dy \right]$.

Après intégration , nous obtenons $s(x) = e^{\frac{2\alpha\Phi}{\gamma^2 x} - \frac{2\alpha}{\gamma^2}}$. Par conséquent, d'après le rappel précédent, la probabilité stationnaire s'écrit

$$\mu(x) \propto \frac{1}{s(x)(\gamma x)^2} = e^{-\frac{2\alpha\Phi}{\gamma^2} \frac{1}{x} - \frac{2\alpha}{\gamma^2} - 2}$$

On reconnaît comme distribution stationnaire conditionnellement aux paramètres, une loi inverse Γ . Plus précisément (voir lemme 4, paragraphe 8.5.3),

$$\frac{1}{Z_t} \sim \Gamma \left(\frac{2\alpha}{\gamma^2} + 1, \frac{2\alpha\Phi}{\gamma^2} \right) \quad \forall t > 0$$

Nous obtenons le résultat suivant :

Résultat 4. Si Z_{t_1} est une observation à l'instant t_1 du processus suivant :

$$\begin{cases} dZ_t = \alpha(\Phi - Z_t)dt + \gamma Z_t dB_t & t > 0, \alpha > 0, \gamma \in \mathbb{R}^* \\ \frac{1}{Z_0} \sim \Gamma \left(\frac{2\alpha}{\gamma^2} + 1, \frac{2\alpha\Phi}{\gamma^2} \right) \\ \Phi \sim \chi^2(a) \end{cases}$$

Alors, posons $R = \frac{4\alpha}{4\alpha+2\gamma^2}$, $\frac{Z_{t_1}}{Ra} \sim \mathcal{F}(a, \frac{4\alpha}{\gamma^2} + 2)$ où $\mathcal{F}(a, b)$ désigne une distribution de Fisher de paramètres (a, b) . De plus, si $\theta_1 = \alpha \in \mathbb{R}^{*+}$ et $\theta_2 = (a, \gamma) \in]2, \infty[\times \mathbb{R}^{*+}$, θ_2 est identifiable.

Démonstration. • **Distribution marginale**

Puisque nous sommes sous le régime stationnaire, nous avons pour loi conditionnelle :

$$\frac{1}{Z_t} | \Phi \sim \Gamma \left(\frac{2\alpha}{\gamma^2} + 1, \frac{2\alpha\Phi}{\gamma^2} \right)$$

Or, d'après le lemme 5 (paragraphe 8.5.4), $\forall (a, \sigma) \in \mathbb{R}^{+*2}$,

$$\Gamma \left(\frac{a}{2}, \frac{1}{2\sigma^2} \right) = \sigma^2 \chi^2(a).$$

Donc, finalement

$$\frac{1}{Z_t} | \Phi \sim \frac{\gamma^2}{4\alpha\Phi} \chi^2 \left(\frac{4\alpha}{\gamma^2} + 2 \right)$$

Supposons que $\Phi \sim \chi^2(a)$ avec $a \in \mathbb{R}^{*+}$, alors

$$\begin{aligned} Z_t &= \frac{4\alpha}{\gamma^2} \frac{\Phi}{X} && \text{où } X \sim \chi^2 \left(\frac{4\alpha}{\gamma^2} + 2 \right) \text{ et } \Phi \sim \chi^2(a) \\ &= \frac{4\alpha a}{\gamma^2 \left(\frac{4\alpha}{\gamma^2} + 2 \right)} \frac{\Phi/a}{X / \left(\frac{4\alpha}{\gamma^2} + 2 \right)} \\ &= \frac{4\alpha a}{4\alpha + 2\gamma^2} F && \text{où } F \sim \mathcal{F} \left(a, \frac{4\alpha}{\gamma^2} + 2 \right) \\ &= RaF && \text{avec } R = \frac{4\alpha}{4\alpha + 2\gamma^2} \end{aligned}$$

par définition de la distribution de Fisher comme quotient de deux χ^2 .

• **Identifiabilité**

– L'identifiabilité du paramètre a peut aisément être démontrée.

En effet $\frac{4\alpha}{\gamma^2} + 2 > 2$ donc Z_{t_1} a un moment d'ordre 1. Plus précisément,

$$\begin{aligned}
 E[Z_{t_1}] &= RaE[F] && \text{où } F \sim \mathcal{F}\left(a, \frac{4\alpha}{\gamma^2} + 2\right) \\
 &= Ra \frac{\frac{4\alpha}{\gamma^2} + 2}{\left(\frac{4\alpha}{\gamma^2} + 2\right) - 2} \\
 &= a \frac{4\alpha}{4\alpha + 2\gamma^2} \frac{\frac{4\alpha}{\gamma^2} + 2}{\frac{4\alpha}{\gamma^2}} \\
 &= a \frac{\frac{4\alpha}{\gamma^2}}{\frac{4\alpha}{\gamma^2} + 2} \frac{\frac{4\alpha}{\gamma^2} + 2}{\frac{4\alpha}{\gamma^2}} \\
 &= a
 \end{aligned}$$

Par conséquent, soient (a, α, γ) et $(a', \alpha', \gamma') \in \mathbb{R}^{*+} \times \mathbb{R}^{*+} \times \mathbb{R}^{*+}$. Supposons que, $p_{Z_{t_1}}(\cdot; a, \alpha, \gamma) = p_{Z_{t_1}}(\cdot; a', \alpha', \gamma')$. Alors,

$$\begin{aligned}
 \int_0^{+\infty} zp_{Z_{t_1}}(z; a, \alpha, \gamma) dz &= \int_0^{+\infty} zp_{Z_{t_1}}(z; a', \alpha', \gamma') dz \\
 a &= a'
 \end{aligned}$$

– Supposons de plus que $a \in]2, +\infty[$. Alors $F \sim \mathcal{F}\left(a, \frac{4\alpha}{\gamma^2} + 2\right)$ admet un mode noté \bar{m}_F .

$$\bar{m}_F = \frac{a - 2}{a} \frac{\frac{4\alpha}{\gamma^2} + 2}{\left(\frac{4\alpha}{\gamma^2} + 2\right) + 2}.$$

Notons \bar{m}_{t_1} le mode de Z_{t_1} . Nous avons :

$$\begin{aligned}
 \bar{m}_{t_1} &= Ra \bar{m}_F && \text{où } \bar{m}_F \text{ est le mode de } F \\
 &= Ra \frac{a-2}{a} \frac{\frac{4\alpha}{\gamma^2} + 2}{\frac{4\alpha}{\gamma^2} + 2 + 2} \\
 &= \frac{4\alpha}{4\alpha + 2\gamma^2} (a - 2) \frac{4\alpha + 2\gamma^2}{4\alpha + 4\gamma^2} \\
 &= (a - 2) \frac{\alpha}{\alpha + \gamma^2}
 \end{aligned}$$

Or a a déjà été identifié. Par conséquent, si α est connu, et $\gamma \in \mathbb{R}^{*+}$, alors γ est identifiable à partir du mode. En résumé, si $\theta_1 = \alpha \in \mathbb{R}^{*+}$ et $\theta_2 = (a, \gamma) \in]2, \infty[\times \mathbb{R}^{*+}$, θ_2 est identifiable. \square

8.2.3 Modèle de Cox-Ingersoll-Ross

Considérons maintenant l'équation suivante :

$$dZ_t = (2\Phi Z_t + \delta\gamma^2)dt + 2\gamma\sqrt{Z_t}dB_t, \quad \delta > 0, \Phi < 0 \tag{8.2.7}$$

Nous ne sommes pas capable de donner une solution explicite. Plaçons-nous sous le régime stationnaire.

Détermination de la loi stationnaire à Φ fixé Posons $s(x) = \exp\left[-\int^x 2\frac{2\Phi y + \delta\gamma^2}{4\gamma^2 y} dy\right] = e^{-\frac{\Phi}{\gamma^2}x} x^{-\frac{\delta}{2}}$.

Alors la mesure invariante est $\mu(x) = \frac{1}{s(x)4\gamma^2 x} \propto e^{\frac{\Phi}{\gamma^2}x} x^{\frac{\delta}{2}-1}$. Ainsi, conditionnellement aux paramètres, la loi stationnaire est une loi Γ :

$$\Gamma\left(\frac{\delta}{2}, -\frac{\Phi}{\gamma^2}\right)$$

Résultat 5. Si Z_{t_1} est une observation à l'instant t_1 du processus suivant :

$$\begin{cases} dZ_t &= (2\Phi Z_t + \delta\gamma^2)dt + 2\gamma\sqrt{Z_t}dB_t, & \delta > 0, \Phi < 0 \\ Z_0|\Phi &\sim \Gamma\left(\frac{\delta}{2}, -\frac{\Phi}{\gamma^2}\right) \\ -\Phi &\sim \chi^2(a) \end{cases}$$

Alors $Z_{t_1} \sim \frac{\gamma^2\delta}{2} \frac{F}{a}$ où $F \sim \mathcal{F}(\delta, a)$. De plus, si $\theta_1 = \delta \in]2, +\infty[$ et $\theta_2 = (a, \gamma) \in]2, +\infty[\times \mathbb{R}^*$, θ_2 est identifiable.

Démonstration. • **Distribution marginale** Nous savons que

$$Z_t|\Phi \sim \Gamma\left(\frac{\delta}{2}, -\frac{\Phi}{\gamma^2}\right)$$

Par conséquent, d'après le lemme 5 paragraphe 8.5.4,

$$Z_t|\Phi \sim -\frac{\gamma^2}{2\Phi}\chi^2(\delta).$$

En choisissant pour distribution sur $-\Phi$ une χ^2 à a degrés de liberté ($-\Phi \sim \chi^2(a)$), il s'en suit :

$$\begin{aligned} Z_t &= \frac{\gamma^2}{2} \frac{X}{-\Phi} && \text{où } X \sim \chi^2(\delta) \\ &= \frac{\gamma^2}{2} \frac{\delta}{a} \frac{X/\delta}{-\Phi/a} \\ &= \frac{\gamma^2\delta}{2} \frac{F}{a} && \text{où } F \sim \mathcal{F}(\delta, a) \end{aligned}$$

- **Identifiabilité des paramètres a et γ^2 à δ connu** Comme précédemment, nous identifions les paramètres en utilisant les moments. Faisons l'hypothèse que $a > 2$ et $\delta > 2$, autrement dit $\theta_1 = \delta \in]2, +\infty[$ et $\theta_2 = (a, \gamma) \in]2, +\infty[\times]0, +\infty[$. Sous ces conditions, la variable aléatoire $F \sim \mathcal{F}(\delta, a)$ admet un moment d'ordre 1 et un mode \bar{m}_F . Notons \bar{m}_{t_1} le mode de Z_{t_1} . Ainsi nous pouvons écrire :

$$\begin{cases} E[Z_{t_1}] &= \frac{\gamma^2}{2} \frac{\delta}{a} E[F] &= \frac{\delta}{2} \frac{\gamma^2}{a} \frac{a}{a-2} &= \frac{\delta}{2} \frac{\gamma^2}{a-2} \\ \bar{m}_{t_1} &= \frac{\gamma^2}{2} \frac{\delta}{a} \bar{m}_F &= \frac{\gamma^2}{2} \frac{\delta}{a} \frac{\delta-2}{\delta} \frac{a}{a+2} &= \frac{\delta-2}{2} \frac{\gamma^2}{a+2} \end{cases}$$

Soient (a, γ) et (a', γ') deux paires de paramètres.

$$\begin{cases} \frac{\gamma^2}{a-2} = \frac{\gamma'^2}{a'-2} \\ \frac{\gamma^2}{a+2} = \frac{\gamma'^2}{a'+2} \end{cases} \Leftrightarrow \begin{cases} \frac{\gamma^2}{\gamma'^2} = \frac{a-2}{a'-2} \\ \frac{a-2}{a+2} = \frac{a'-2}{a'+2} \end{cases}$$

La fonction $a \mapsto \frac{a-2}{a+2}$ est bijective sur $]2, +\infty[$ donc $a = a'$ et $\gamma = \gamma'$. Nous avons ainsi obtenu l'identifiabilité de $\theta_2 = (a, \gamma)$ sur l'ensemble $\Theta_2 =]2, +\infty[\times]0, +\infty[$.

□

8.2.4 Processus brownien géométrique

Considérons cette fois-ci l'équation (8.2.8)

$$dZ_t = [2\Phi Z_t + \delta\gamma^2] dt + 2\gamma\sqrt{Z_t(1-Z_t)}dB_t, \quad \delta > 0, \Phi < 0 \quad (8.2.8)$$

Détermination de la loi stationnaire Posons $s(x) = \exp \left[- \int 2 \frac{2\Phi x + \delta \gamma^2}{4\gamma^2 x(1-x)} dx \right]$. Alors $s(x) = (1-x)^{\frac{\delta}{2} - \frac{|\Phi|}{\gamma^2}} x^{-\frac{\delta}{2}}$ et on obtient pour loi stationnaire : $\mu(x) \propto (1-x)^{\frac{|\Phi|}{\gamma^2} - \frac{\delta}{2} - 1} x^{\frac{\delta}{2} - 1}$ soit une loi beta :

$$Z_t | \Phi \sim \mathcal{B} \left(\frac{\delta}{2}, \frac{|\Phi|}{\gamma^2} - \frac{\delta}{2} \right)$$

Distribution marginale et identifiabilité La loi conditionnelle de $Z_t | \Phi$ étant une loi beta, il semble alors difficile de trouver une distribution sur l'un des paramètres (δ , β ou γ) permettant d'avoir une expression explicite de la loi marginale aboutissant ainsi à l'identifiabilité des paramètres.

Ainsi même avec une unique observation, nous tombons rapidement sur un processus dont nous ne pouvons exhiber la distribution marginale. Nous ne pouvons donc pas identifier les paramètres par cette méthode.

8.3 Cas d'observations multiples

Nous avons montré que dans le cas d'une unique observation, il est possible dans certain cas d'exhiber des couples EDS/distribution $\pi(\cdot; \beta)$ sur Φ pour lesquels la distribution marginale est explicite et le modèle identifiable. Qu'en est-il dans le cas d'observations multiples ? Peut-on obtenir des généralisations des résultats précédents ?

8.3.1 Cas du processus d'Ornstein-Uhlenbeck

Soit Z un processus défini de la façon suivante :

$$\begin{cases} dZ_t &= \Phi Z_t + \gamma dB_t \\ Z_0 &\sim \mathcal{N} \left(0, -\frac{\gamma^2}{2\Phi} \right) \\ -\Phi &\sim \Gamma \left(\frac{n}{2}, \frac{d}{2} \right) \end{cases}$$

Nous avons montré au paragraphe 8.2.1 que, dans le cas d'une unique observation, la distribution marginale de Z_t est une distribution de Fisher décentrée. Ce résultat repose sur le lemme 1. Considérons maintenant (t_1, t_2) , deux instants d'observation. Notons Z_1 et Z_2 les observations du processus Z précédemment décrit en ces instants. Alors, conditionnellement au paramètre Φ , la distribution conjointe des observations (Z_1, Z_2) est une gaussienne :

$$\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} | \Phi \sim \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} -\frac{\gamma^2}{2\Phi} & -e^{\Phi(t_2-t_1)} \frac{\gamma^2}{2\Phi} \\ -e^{\Phi(t_2-t_1)} \frac{\gamma^2}{2\Phi} & -\frac{\gamma^2}{2\Phi} \end{pmatrix} \right)$$

Le lemme 1 nous a permis de déterminer la distribution marginale de Z_t dans le cas d'une unique observation. Cherchons à appliquer la généralisation du lemme 1 qui s'exprime sous la forme du lemme suivant (voir [128], page 665) :

Lemme 2. Soit Φ une variable aléatoire de loi $\Gamma \left(\frac{n}{2}, \frac{d}{2} \right)$ et X un vecteur de \mathbb{R}^p tel que

$$X | \Phi \sim \mathcal{N}(\mathbf{m}, \mathbf{C}\Phi^{-1})$$

où \mathbf{m} est un vecteur de \mathbb{R}^p et \mathbf{C} est une matrice symétrique définie positive de taille p . Alors posons $\mathbf{R} = \frac{d}{n}\mathbf{C}$. X a pour densité marginale

$$p(X) \propto [n + (X - \mathbf{m})' \mathbf{R}^{-1} (X - \mathbf{m})]^{-(n+p)/2}.$$

Ainsi, X a pour distribution marginale une loi de Student à n degrés de liberté, de mode \mathbf{m} et de matrice d'échelle \mathbf{R} .

Cependant, dans notre cas, la matrice de variance-covariance vaut

$$\begin{pmatrix} \frac{-\gamma^2}{2\Phi} & -e^{\Phi(t_2-t_1)} \frac{\gamma^2}{2\Phi} \\ -e^{\Phi(t_2-t_1)} \frac{\gamma^2}{2\Phi} & \frac{-\gamma^2}{2\Phi} \end{pmatrix} = -\frac{1}{\Phi} \begin{pmatrix} \frac{\gamma^2}{2} & e^{\Phi(t_2-t_1)} \frac{\gamma^2}{2} \\ e^{\Phi(t_2-t_1)} \frac{\gamma^2}{2} & \frac{\gamma^2}{2} \end{pmatrix} = |\Phi|^{-1} \mathbf{C}(\Phi)$$

La dépendance de \mathbf{C} en le paramètre Φ ne nous permet pas d'appliquer ce résultat. Nous sommes dans l'incapacité d'exhiber la distribution stationnaire et donc d'étudier l'identifiabilité à partir de celle-ci.

Ainsi pour l'exemple le plus simple –le processus de Ornstein–Uhlenbeck– le problème de l'identifiabilité s'avère rapidement insoluble à partir de cette méthode.

8.3.2 Cas des modèles à effets mixtes

Au chapitre 7, nous avons appliqué les méthodes d'estimation proposées (SAEM-MCMC) sur un modèle à effets mixtes dans le cadre d'une application en pharmacocinétique. Les modèles à effets mixtes sont des modèles largement répandus en PK-PD pour étudier les données longitudinales puisqu'ils permettent de distinguer la variabilité intra-sujets de la variabilité inter-sujets. De nombreuses méthodes d'estimation des paramètres ont été proposées dans la littérature sans que la question de l'identifiabilité des modèles ait été complètement résolue. Nous rappelons ici les principaux résultats existants dans la littérature sur l'identifiabilité des modèles mixtes linéaires et non linéaires puis nous exhibons un modèle de diffusion mixte avec et sans bruit d'observation pour lesquels les paramètres sont identifiables.

Modèles linéaires à effets mixtes

Dans le cas des modèles linéaires mixtes, cette question se résume à un problème d'algèbre linéaire. En effet, pour $i = 1 \dots I$, soit $\mathbf{y}_i \in \mathbb{R}^{n_i}$ le vecteur des observations du sujet i aux instants $(t_{i1}, \dots, t_{in_i})$. Soit $\Phi_i \in \mathbb{R}^d$ le vecteur des paramètres individuels du sujet i : $\Phi_i \sim_{\text{i.i.d.}} \mathcal{N}(\mu, \Sigma)$ ($\mu \in \mathbb{R}^d$, $\Sigma \in \mathcal{M}_d(\mathbb{R})$). Soit $X_i \in \mathbb{R}^{n_i \times d}$ la matrice de protocole expérimental pour l'individu i (connue). Soit $\varepsilon_i \in \mathbb{R}^{n_i}$ un vecteur gaussien représentant le bruit de mesure. Alors le modèle linéaire à effets mixtes s'écrit :

$$\begin{cases} \mathbf{y}_i &= X_i \Phi_i + \varepsilon_i & \text{pour } i = 1 \dots I \\ \Phi_i &\sim \text{i.i.d.} \mathcal{N}(0, \Sigma) \\ \varepsilon_i &\sim \text{i.i.d.} \mathcal{N}(0, \sigma^2 \mathbf{I}_{n_i}) \end{cases} \quad (8.3.9)$$

Posons $n = \sum_{i=1}^I n_i$ et

$$\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_I \end{pmatrix} \in \mathbb{R}^n, \quad \mathbf{X}^{(1)} = \begin{pmatrix} X_1 \\ \vdots \\ X_I \end{pmatrix} \in \mathcal{M}_{n,d}(\mathbb{R}),$$

$$\mathbf{X}^{(2)} = \begin{pmatrix} X_1 & & \\ & \ddots & \\ & & X_I \end{pmatrix} \in \mathcal{M}_{n,d \times I}(\mathbb{R}), \quad \Sigma = \begin{pmatrix} \Sigma & & \\ & \ddots & \\ & & \Sigma \end{pmatrix} \in \mathcal{M}_{dI,dI}(\mathbb{R})$$

Nous pouvons écrire

$$\mathbf{y} \sim \mathcal{N} \left(\mathbf{X}^{(1)} \mu, \mathbf{X}^{(2)} \Sigma \mathbf{X}^{(2)'} + \sigma^2 \mathbf{I}_n \right),$$

Proposition 2. (Modèles linéaires mixtes, Nuñez et Concordet [33])

Le modèle linéaire mixte défini par (8.3.9) est identifiable dès lors que la matrice $\mathbf{X}^{(1)}$ est injective et qu'il existe un individu i ($i \in \{1, \dots, I\}$) pour lequel $n_i \geq d + 1$.

Démonstration. Nous avons l'identifiabilité si pour tous triplets (μ, Σ, σ^2) et $(\mu_0, \Sigma_0, \sigma_0^2)$

$$\begin{cases} \mathbf{X}^{(1)}\mu & = \mathbf{X}^{(1)}\mu_0 \\ \mathbf{X}^{(2)}\Sigma\mathbf{X}^{(2)'} + \sigma^2\mathbf{I}_n & = \mathbf{X}^{(2)}\Sigma_0\mathbf{X}^{(2)'} + \sigma_0^2\mathbf{I}_n \end{cases} \Rightarrow (\mu, \Sigma, \sigma^2) = (\mu_0, \Sigma_0, \sigma_0^2)$$

Une condition suffisante pour l'identifiabilité de μ est l'injectivité de la matrice $\mathbf{X}^{(1)}$. D'autre part, $X_i\Sigma X_i + \sigma^2\mathbf{I}_{n_i} = X_i\Sigma_0 X_i' + \sigma_0^2\mathbf{I}_{n_i}$ pour tout $i = 1, \dots, I$ si et seulement si $\forall i = 1, \dots, I$ $X_i(\Sigma - \Sigma_0)X_i' = (\sigma_0^2 - \sigma^2)\mathbf{I}_{n_i}$. Or la matrice $X_i(\Sigma - \Sigma_0)X_i'$ est de rang au plus d et \mathbf{I}_{n_i} est de rang n_i . Donc $\sigma^2 = \sigma_0^2$ et $\Sigma = \Sigma_0$ si il existe un individu i pour lequel $n_i \geq d + 1$. \square

Modèles non linéaires à effets mixtes

Nuñez et Concordet [33] ont étudié le problème de l'identifiabilité des modèles non linéaires à effets mixtes. Considérons le modèle non linéaire mixte suivant :

$$\begin{cases} y_{ij} & = m_{ij}(\Phi_i) + \varepsilon_{ij} \\ \varepsilon_{ij} & \sim \mathcal{N}(0, \sigma^2) \\ \Phi_i & \sim Q \end{cases}$$

où y_{ij} est l'observation du sujet i , ($i = 1, \dots, I$) à l'instant t_j , ($j = 1, \dots, n$). $\Phi_i \in \mathbb{R}^d$ est le vecteur des paramètres individuels du sujet i . m_{ij} est une fonction de protocole non-linéaire, connue. Q est une fonction de distribution appartenant à l'ensemble $\mathcal{P}(\mathbb{R}^d)$ des mesures de probabilités ayant pour support \mathbb{R}^d . Le but est de trouver des conditions sur le protocole $((m_{ij})_{j=1, \dots, n})$ assurant l'identifiabilité de (Q, σ^2) .

Soit $m_i = (m_{ij})_{j=1, \dots, n_i}$. Définissons les quatre conditions suivantes :

- (1) Il existe des composantes de m_i , $m_i^* = (m_{ij_1}, \dots, m_{ij_r})$ et un ouvert non vide $O_1 \subset \mathbb{R}^r$ ($O_1 \neq \mathbb{R}^r$) telles que $m_i^*(\phi) \in O_1 \forall \phi \in \mathbb{R}^d$.
- (2) Il existe O_2 un ouvert non vide de \mathbb{R}^d de mesure de Lebesgue non nulle et tel que $Q(O_2) > 0$
- (3) Le nombre n_i de composantes de m_i est supérieur ou égal à $d + 1$.
- (4) Il existe $j, k \in \{1, \dots, n_i\}$ tels que $m_{ij}(\phi) = m_{ik}(\phi)$ pour tout ϕ .

Remarque 26. Dans la condition (1), la distribution de $m_i(\Phi)$ ne pondère pas une région de l'espace alors que ε_i le fait. Le cas (2) traite du cas où il existe des ensembles de mesure de Lebesgue nulle pondérés par Q (ce que ne fait pas la distribution gaussienne de ε_i). La condition (4) peut être interprétée en terme de répétition de mesures pour le sujet i .

Proposition 3. (Modèles non-linéaires mixtes [33]) *Supposons qu'il existe $i \in \{1, \dots, I\}$, telle que m_i soit injective. Si (2) est vraie ou si il existe $j \in \{1, \dots, I\}$ telle que (1), (3) ou (4) soient vraies alors (Q, σ^2) est identifiable à partir de la distribution jointe de tous les individus.*

La preuve de ce résultat fait l'objet de l'article de Nuñez et Concordet [33].

Exemple Considérons le modèle suivant ([100]) :

$$y_{ij} = \frac{\Phi_{i1}t_j}{\Phi_{i2} + t_j} + \varepsilon_{ij}.$$

Nous considérons un cas où tous les individus ont le même protocole expérimental (les instants d'observations ne dépendent pas de l'individu). Les paramètres individuels $\Phi_i \in \mathbb{R}^2$ ont pour distribution Q que l'on suppose absolument continue par rapport à la mesure de Lebesgue sur \mathbb{R}^2 . Alors on peut montrer que Q et σ^2 sont identifiables dès lors que le protocole expérimental contient au moins trois instants d'observation distincts dont deux strictement positifs.

Remarque 27. *Les conditions énoncées précédemment sont suffisantes mais sans doute non nécessaires en particulier dans le cas où les sujets ont des protocoles différents. En effet, la structure de modèle mixte n'est pas utilisée au mieux : Q et σ^2 sont identifiés à partir de la distribution d'un des individus et non en considérant des critères sur l'ensemble des protocoles.*

Exemple de diffusion mixte sans bruit d'observation

Nous avons vu qu'en pharmacocinétique, la fonction de régression est souvent définie comme solution d'un système différentiel. Considérons par exemple l'élimination du premier ordre d'un médicament avec décroissance mono-exponentielle :

$$\frac{dC(t)}{dt} = -kC(t) ; C(0) = D/V$$

où $C(t)$ est la concentration du médicament dans le plasma à l'instant t après injection, k est la constante d'élimination, D est la dose injectée au temps $t = 0$ et V est le volume. Cette équation a pour solution $C(t) = C(0)e^{-kt}$.

Ditlevsen et De Gaetano [38] proposent de modéliser le comportement erratique du processus de métabolisation, en autorisant k à évoluer de façon aléatoire au cours du temps : $k + \xi(t)$ où $\xi(t)$ est un processus de bruit blanc. Alors en posant $Z_t = C(t)$ et $\Phi = -k$, nous obtenons l'équation différentielle stochastique suivante :

$$dZ_t = \Phi Z_t dt + \gamma Z_t dB_t ; Z_0 = D/V$$

où B_t est un mouvement brownien.

Remarque 28. *L'identifiabilité de ce modèle a été traitée dans le cas d'une unique observation au paragraphe 8.2.2.*

Soit $Y_t = \log Z_t$, nous obtenons l'EDS suivante :

$$dY_t = (\Phi - \frac{1}{2}\gamma^2)dt + \gamma dB_t ; Y_0 = \log(D/V)$$

Supposons que I sujets sont suivis au cours de l'étude pendant laquelle la concentration en médicament dans le plasma est mesurée à différents instants après l'injection. $\forall i = 1 \dots I$, soient $(t_{i1}, \dots, t_{in_i})$ les instants de mesure du sujet i et Φ_i , le paramètre individuel du sujet i . Supposons que $\Phi_i \sim_{\text{i.i.d}} \mathcal{N}(\mu, \omega^2)$, pour $i = 1 \dots I$. Soit $Z_{t_{ij}}^i$ l'observation de l'individu i à l'instant t_{ij} et $Y_{t_{ij}}^i = \log Z_{t_{ij}}^i$. Alors après intégration de l'EDS, le modèle statistique s'écrit :

$$\begin{cases} Y_{t_{ij}}^i &= Y_0 + (\Phi_i - \frac{1}{2}\gamma^2)t_{ij} + \gamma B_{t_{ij}}, \forall i \in \{1 \dots I\}, \forall j \in \{1 \dots n_i\} \\ B_{t_{ij}} &\sim \text{i.i.d} \mathcal{N}(0, t_{ij}) \\ \Phi_i &\sim \text{i.i.d} \mathcal{N}(\mu, \omega^2) \end{cases}$$

Les paramètres de ce modèle sont $\theta_2 = (\mu, \omega^2, \gamma^2) \in \mathbb{R} \times \mathbb{R}^{*+} \times \mathbb{R}^{*+}$.

Résultat 6. $\forall i \in \{1, \dots, I\}$, le vecteur des log-observations du sujet i a pour distribution marginale une gaussienne :

$$\begin{pmatrix} Y_{t_{i1}}^i \\ Y_{t_{i2}}^i \\ \vdots \\ Y_{t_{in_i}}^i \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} Y_0 + (\mu - \frac{1}{2}\gamma^2)t_{i1} \\ Y_0 + (\mu - \frac{1}{2}\gamma^2)t_{i2} \\ \vdots \\ Y_0 + (\mu - \frac{1}{2}\gamma^2)t_{in_i} \end{pmatrix}, W^{(i)} \right)$$

avec $W^{(i)} \in \mathcal{M}_{n_i}(\mathbb{R})$ telle que $W_{k,l}^{(i)} = \gamma^2 t_{i \min(l,k)} + \omega^2 t_k t_l, \forall (k, l) \in \{1, \dots, n_i\}$.

De plus, les paramètres $\theta_2 = (\mu, \omega^2, \gamma^2)$ sont identifiables dès lors qu'il existe un individu i possédant au moins deux instants d'observations distincts $t_{i'l}$ et $t_{i'l'}$ (différents de 0).

Démonstration. • **Détermination de la distribution marginale des observations**

$\forall i \in \{1 \dots I\}$, considérons les accroissements $\left(Y_{t_{ij}}^i - Y_{t_{i,j-1}}^i \right)_{j=2 \dots n_i}$,

$$\begin{cases} Y_{t_{i1}}^i & = Y_0 + (\Phi_i - \frac{1}{2}\gamma^2)t_{i1} + \gamma B_{t_{i1}}^i \\ Y_{t_{i2}}^i - Y_{t_{i1}}^i & = (\Phi_i - \frac{1}{2}\gamma^2)(t_{i2} - t_{i1}) + \gamma (B_{t_{i2}}^i - B_{t_{i1}}^i) \\ & \vdots \\ Y_{t_{in_i}}^i - Y_{t_{in_i-1}}^i & = (\Phi_i - \frac{1}{2}\gamma^2)(t_{in_i} - t_{in_i-1}) + \gamma (B_{t_{in_i}}^i - B_{t_{in_i-1}}^i) \end{cases}$$

Les accroissements $(B_{t_{ij}}^i - B_{t_{i,j-1}}^i)$ étant indépendants, nous avons conditionnellement à Φ_i :

$$\begin{pmatrix} Y_{t_{i1}}^i \\ Y_{t_{i2}}^i - Y_{t_{i1}}^i \\ \vdots \\ Y_{t_{in_i}}^i - Y_{t_{in_i-1}}^i \end{pmatrix} | \Phi_i \sim \mathcal{N} \left(\begin{pmatrix} Y_0 + (\Phi_i - \frac{1}{2}\gamma^2)t_{i1} \\ (\Phi_i - \frac{1}{2}\gamma^2)(t_{i2} - t_{i1}) \\ \vdots \\ (\Phi_i - \frac{1}{2}\gamma^2)(t_{in_i} - t_{in_i-1}) \end{pmatrix}, D^{(i)} \right)$$

où $D^{(i)} \in \mathcal{M}_{n_i}(\mathbb{R})$ est une matrice diagonale telle que, $\forall j \in \{1, \dots, n_i\}$, $D_{j,j}^{(i)} = \gamma^2(t_{ij} - t_{i,j-1})$ (avec $t_{i0} = 0$). Le paramètre Φ_i est gaussien de paramètres μ et ω^2 , nous obtenons donc la distribution marginale des accroissements :

$$\begin{pmatrix} Y_{t_{i1}}^i \\ Y_{t_{i2}}^i - Y_{t_{i1}}^i \\ \vdots \\ Y_{t_{in_i}}^i - Y_{t_{in_i-1}}^i \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} Y_0 + (\mu - \frac{1}{2}\gamma^2)t_{ij} \\ (\mu - \frac{1}{2}\gamma^2)(t_{i2} - t_{i1}) \\ \vdots \\ (\mu - \frac{1}{2}\gamma^2)(t_{in_i} - t_{in_i-1}) \end{pmatrix}, \Gamma^{(i)} + D^{(i)} \right)$$

où $\Gamma^{(i)} \in \mathcal{M}_{n_i}(\mathbb{R})$ est telle que, $\forall (j, k) \in \{1, \dots, n_i\}$, $\Gamma_{j,k}^{(i)} = \omega^2(t_{ij} - t_{i,j-1})(t_{ik} - t_{i,k-1})$.

Par conséquent, posons $L \in \mathcal{M}_{n_i}(\mathbb{R})$ telle que :

$$L = \begin{pmatrix} 1 & 0 & \dots & \dots & \dots & 0 \\ 1 & 1 & 0 & \dots & \dots & 0 \\ 1 & 1 & 1 & 0 & \dots & 0 \\ \vdots & & \ddots & & & \vdots \\ 1 & 1 & 1 & 1 & \dots & 1 \end{pmatrix}.$$

Nous avons

$$\begin{pmatrix} Y_{t_{i1}}^i \\ Y_{t_{i2}}^i \\ \vdots \\ Y_{t_{in_i}}^i \end{pmatrix} = L \begin{pmatrix} Y_{t_{i1}}^i \\ Y_{t_{i2}}^i - Y_{t_{i1}}^i \\ \vdots \\ Y_{t_{in_i}}^i - Y_{t_{in_i-1}}^i \end{pmatrix}$$

Donc

$$\begin{pmatrix} Y_{t_{i1}}^i \\ Y_{t_{i2}}^i \\ \vdots \\ Y_{t_{in_i}}^i \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} Y_0 + (\mu - \frac{1}{2}\gamma^2)t_{i1} \\ Y_0 + (\mu - \frac{1}{2}\gamma^2)t_{i2} \\ \vdots \\ Y_0 + (\mu - \frac{1}{2}\gamma^2)t_{in_i} \end{pmatrix}, W^{(i)} \right)$$

avec $W^{(i)} \in \mathcal{M}_{n_i}(\mathbb{R}) = L (\Gamma^{(i)} + D^{(i)}) L'$ telle que $W_{k,l}^{(i)} = \gamma^2 t_{i \min(l,k)} + \omega^2 t_k t_l$, $\forall (k, l) \in \{1, \dots, n_i\}$.

• **Identifiabilité du modèle**

Montrons que le modèle est identifiable. Soient $(\mu, \omega^2, \gamma^2)$ et $(\mu', \omega'^2, \gamma'^2)$ deux ensembles de paramètres tels que les distributions marginales correspondantes des observations soient identiques.

Alors, pour tout individu $i \in \{1, \dots, I\}$, $\forall j \in \{1, \dots, n_i\}$, $\forall (k, l) \in \{1, \dots, n_i\}$

$$\begin{cases} Y_0 + (\mu - \frac{1}{2}\gamma^2)t_{ij} & = & Y_0 + (\mu' - \frac{1}{2}\gamma'^2)t_{ij} \\ \gamma^2 t_{i \min(l,k)} + \omega^2 t_{ik} t_{il} & = & \gamma'^2 t_{i \min(l,k)} + \omega'^2 t_{ik} t_{il} \end{cases}$$

Donc $\forall i \in \{1, \dots, I\}$, $\forall (l) \in \{1, \dots, n_i\}$

$$\begin{cases} Y_0 + (\mu - \frac{1}{2}\gamma^2)t_{ij} & = & Y_0 + (\mu' - \frac{1}{2}\gamma'^2)t_{ij} \\ \gamma^2 - \gamma'^2 & = & (\omega'^2 - \omega^2)t_{il} \end{cases}$$

Cette dernière égalité est impossible dès lors qu'il existe un individu i possédant au moins deux instants d'observations distincts t_{il} et $t_{il'}$ (différents de 0). On obtient ainsi $(\mu, \omega^2, \gamma^2) = (\mu', \omega'^2, \gamma'^2)$. \square

Ainsi, dans ce cas précis, nous avons réussi à expliciter la distribution marginale des observations et à identifier les paramètres en profitant du fait que toutes les distributions étaient gaussiennes. Mais dans la plupart des modèles, nous sommes dans l'incapacité de déterminer la loi jointe des observations pour un individu donné, et donc l'identifiabilité n'est pas une question triviale.

Extension à un exemple de diffusion mixte avec bruit d'observation

Le modèle étudié précédemment peut être facilement étendu au modèle de diffusion mixte dans lequel les observations sont perturbées par un bruit d'observation. Plus précisément considérons le processus suivant :

$$\begin{cases} Z_t & = & \tilde{\varepsilon}_t \tilde{Z}_t \\ \log \tilde{\varepsilon}_t & \sim & \mathcal{N}(0, \sigma^2) \\ d\tilde{Z}_t & = & \Phi \tilde{Z}_t dt + \gamma \tilde{Z}_t dB_t, \tilde{Z}_0 = D/V \\ \Phi & \sim & \mathcal{N}(\mu, \omega^2) \end{cases}$$

Nous nous plaçons dans le cadre d'un modèle mixte et considérons I individus, observés aux temps $(t_{i0}, t_{i1}, \dots, t_{in_i})$. Alors, soit $Z_{t_{ij}}^i$ l'observation du sujet i au temps t_{ij} et $Y_{t_{ij}}^i = \log Z_{t_{ij}}^i$. Posons $\varepsilon_{t_{ij}} = \log \tilde{\varepsilon}_{t_{ij}} \sim_{\text{i.i.d}} \mathcal{N}(0, \sigma^2)$ le bruit d'observation au temps t_{ij} .

Après intégration de l'EDS, nous obtenons comme modèle statistique :

$$\begin{cases} Y_{t_{ij}}^i & = & Y_0 + (\Phi_i - \frac{1}{2}\gamma^2)t_{ij} + \gamma B_{t_{ij}} + \varepsilon_{t_{ij}}^i \\ B_{t_{ij}} & \sim & \text{i.i.d} \mathcal{N}(0, t_{ij}) \\ \varepsilon_{t_{ij}}^i & \sim & \text{i.i.d} \mathcal{N}(0, \sigma^2), \\ \Phi_i & \sim & \text{i.i.d} \mathcal{N}(\mu, \omega^2) \end{cases} \quad \forall i \in \{1 \dots I\}, \forall j \in \{0 \dots n_i\}$$

avec $Y_0 = \log(D/V)$.

Résultat 7. $\forall i \in \{1, \dots, I\}$, le vecteur $(Y_{t_{i0}}^i, Y_{t_{i1}}^i, \dots, Y_{t_{in_i}}^i)$ a pour distribution stationnaire une loi gaussienne :

$$\begin{pmatrix} Y_{t_{i0}}^i \\ Y_{t_{i1}}^i \\ Y_{t_{i2}}^i \\ \vdots \\ Y_{t_{in_i}}^i \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} Y_0 \\ Y_0 + (\mu - \frac{1}{2}\gamma^2)t_{i1} \\ Y_0 + (\mu - \frac{1}{2}\gamma^2)t_{i2} \\ \vdots \\ Y_0 + (\mu - \frac{1}{2}\gamma^2)t_{in_i} \end{pmatrix}, \widetilde{W}^{(i)} \right)$$

avec $\widetilde{W}^{(i)} \in \mathcal{M}_{n_i+1}(\mathbb{R})$ telle que

$$\widetilde{W}^{(i)} = \begin{pmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & W^{(i)} + \sigma^2 \mathbf{I}_{n_i} & \\ 0 & & & \end{pmatrix}$$

où $W^{(i)}$ a été définie précédemment. De plus, les paramètres $(\mu, \omega^2, \gamma^2, \sigma^2)$ sont identifiables si il existe un sujet i pour lequel un prélèvement est fait au temps $t_{i0} = 0$ et si il existe un individu i' possédant au moins deux instants d'observations distincts $t_{i'l}$ et $t_{i'l'}$ (différents de 0).

Démonstration. La distribution marginale découle directement du cas précédent (sans bruit d'observation). Les conditions sur l'identifiabilité de σ^2 sont alors immédiates : le paramètre σ^2 est identifiable dès lors qu'il existe un sujet i pour lequel un prélèvement est fait au temps $t_{i0} = 0$. Ensuite, nous nous retrouvons dans la même situation que précédemment : si il existe un individu i' possédant au moins deux instants d'observations distincts $t_{i'l}$ et $t_{i'l'}$ (différents de 0) les paramètres $(\mu, \omega^2, \gamma^2)$ sont identifiables. \square

8.4 Conclusion et perspectives

Nous avons exhibé dans ce chapitre des modèles de diffusion, solution d'une EDS à paramètres aléatoires pour lesquels nous sommes capable de déterminer la distribution marginale des observations et de prouver l'identifiabilité des paramètres. Cependant, ce ne sont que des cas très particuliers, et les raisonnements empruntés ne peuvent pas être généralisés. L'identifiabilité est ainsi une question non triviale pourtant nécessaire, pour la consistance des estimateurs par exemple. Afin d'approfondir ce travail, plusieurs pistes s'offrent à nous.

- Dans un premier temps, il paraît intéressant d'étudier le modèle qui semble le plus simple, i.e. le processus d'Ornstein-Uhlenbeck $dZ_t = \Phi Z_t + \gamma dB_t$ $Z_0 \sim \mathcal{N}(0, \frac{\sigma^2}{2\Phi})$, avec $\Phi \sim \Gamma(\frac{n}{2}, \frac{d}{2})$ dans le cas d'observations multiples.
- Par ailleurs, le problème étant rendu difficile par le manque de connaissances sur la solution de l'EDS, à paramètre Φ fixé, il peut être intéressant de considérer une approximation de ce processus par un schéma d'Euler-Maruyama par exemple (comme cela a été fait au chapitre 7). Le problème revient alors à étudier l'identifiabilité des paramètres pour des processus du type

$$Z_{t_n} = Z_{t_{n-1}} + (t_n - t_{n-1})F(Z_{t_{n-1}}, t_{n-1}, \Phi, \eta) + G(Z_{t_{n-1}}, t_{n-1}, \Phi, \gamma)(B_{t_n} - B_{t_{n-1}}).$$

Si $G(Z_{t_{n-1}}, t_{n-1}, \Phi, \gamma) = \gamma$, ces processus entrent dans la catégorie des processus auto-régressifs à paramètres aléatoires pour lesquels quelques résultats existent dans la littérature.

8.5 Résultats techniques

8.5.1 Preuve du lemme 1

Nous cherchons à démontrer le lemme suivant :

Lemme 1. *Soit Φ une variable aléatoire de loi Γ de paramètres $(\frac{n}{2}, \frac{d}{2})$ et X une variable aléatoire à valeurs dans \mathbb{R} telle que sa distribution conditionnellement à Φ soit :*

$$X|\Phi \sim \mathcal{N}(m, \frac{c}{\Phi})$$

avec $m \in \mathbb{R}$ et $c > 0$. Alors, posons $R = cd/n$. La distribution marginale de la variable aléatoire $\frac{X-m}{R^{1/2}}$ est une Student à n degrés de liberté :

$$\frac{X-m}{R^{1/2}} \sim T_n([0, 1])$$

Démonstration. 1. Puisque $\Phi \sim \Gamma(\frac{n}{2}, \frac{d}{2})$, $p(\phi)$ s'écrit :

$$p(\phi) = \frac{d^{n/2}}{2^{n/2}\Gamma(n/2)} \phi^{n/2-1} e^{-\phi d/2}.$$

Par conséquent, d'après le théorème de Bayes, la distribution conjointe de (X, ϕ) est :

$$\begin{aligned} p(X, \phi) &= p(X|\phi)p(\phi) \\ &= \left(\frac{\phi}{2\pi c}\right)^{1/2} \exp\left[-\frac{\phi}{2c}(X-m)^2\right] \frac{d^{n/2}}{2^{n/2}\Gamma(n/2)} \phi^{n/2-1} e^{-\phi d/2} \\ &\propto \phi^{\frac{n+1}{2}-1} \exp\left[-\frac{\phi}{2}\left\{\frac{(X-m)^2}{c} + d\right\}\right] \end{aligned}$$

2. Par conséquent, la distribution de ϕ conditionnellement à X vaut :

$$p(\phi|X) = \frac{p(X, \phi)}{p(X)} \propto \phi^{\frac{n+1}{2}-1} \exp\left[-\frac{\phi}{2}\left\{\frac{(X-m)^2}{c} + d\right\}\right]$$

On reconnaît une loi gamma de paramètres $\Gamma(\frac{n^*}{2}, \frac{d^*}{2})$ où

$$n^* = n + 1 \quad , \quad d^* = d + \frac{(X-m)^2}{c}$$

3. Au final,

$$\begin{aligned} p(X) &= p(X, \phi)/p(\phi|X) \\ &\propto \frac{\exp\left[-\frac{\phi}{2}\left\{\frac{(X-m)^2}{c} + d\right\}\right] \phi^{\frac{n+1}{2}-1}}{\left[d + \frac{(X-m)^2}{c}\right]^{\frac{n+1}{2}} \exp\left[-\frac{\phi}{2}\left\{\frac{(X-m)^2}{c} + d\right\}\right] \phi^{\frac{n+1}{2}-1}} \\ &\propto \left[d + \frac{(X-m)^2}{c}\right]^{-\frac{n+1}{2}} \\ &\propto \left[n + \frac{(X-m)^2}{cd/n}\right]^{-\frac{n+1}{2}} \end{aligned}$$

En posant $R = cd/n$ et $Y = \frac{X-m}{R^{1/2}}$, on reconnaît que Y suit une loi de Student à n degrés de liberté. □

8.5.2 Croissance 1

Lemme 3. la fonction $\Psi : x \mapsto \frac{\Gamma(\frac{x+1}{2})}{\Gamma(\frac{x}{2})}$ est strictement croissante sur \mathbb{R}^{*+}

Démonstration. La fonction Ψ est dérivable sur \mathbb{R}^{*+} de fonction dérivée :

$$\begin{aligned} \Psi'(x) &= \frac{\frac{1}{2}\Gamma'(\frac{x+1}{2})\Gamma(\frac{x}{2}) - \frac{1}{2}\Gamma(\frac{x+1}{2})\Gamma'(\frac{x}{2})}{\Gamma(\frac{x}{2})^2} \\ &= \frac{\frac{1}{2}\Gamma(\frac{x}{2})\Gamma(\frac{x+1}{2})}{\Gamma(\frac{x}{2})^2} \left[\frac{\Gamma'(\frac{x+1}{2})}{\Gamma(\frac{x+1}{2})} - \frac{\Gamma'(\frac{x}{2})}{\Gamma(\frac{x}{2})} \right] \\ &= \frac{\frac{1}{2}\Gamma(\frac{x+1}{2})}{\Gamma(\frac{x}{2})} \left[\psi_0\left(\frac{x+1}{2}\right) - \psi_0\left(\frac{x}{2}\right) \right] \end{aligned}$$

où ψ_0 est la fonction digamma définie comme suit : $\psi_0(x) = \frac{\Gamma'(x)}{\Gamma(x)}$ avec Γ définissant la fonction Gamma (voir [1]). Par définition de la fonction Γ , $\frac{1}{2}\Gamma(\frac{x+1}{2}) > 0$ pour tout $x > 0$. De plus, la fonction ψ_0 est strictement croissante sur \mathbb{R}^{*+} (voir [1]). Donc $\psi_0(\frac{x+1}{2}) - \psi_0(\frac{x}{2}) > 0$ sur \mathbb{R}^{*+} . Donc $\Psi'(x) > 0$, $\forall x > 0$. D'où la conclusion. \square

8.5.3 Distribution Inverse Gamma

Lemme 4. (Inverse Gamma) Si Z_t a pour densité

$$\mu(x) \propto \exp^{-\beta \frac{1}{x}} x^{-\alpha-2}$$

alors

$$\frac{1}{Z_t} \sim \Gamma(\alpha + 1, \beta)$$

Démonstration. Soit f une fonction mesurable bornée.

$$\begin{aligned} E \left[f \left(\frac{1}{Z_t} \right) \right] &= \int_0^{+\infty} f \left(\frac{1}{x} \right) C \exp^{-\beta \frac{1}{x}} x^{-\alpha-2} dx \\ &= \int_0^{+\infty} f(y) C \exp^{-\beta y} y^{\alpha+2} \frac{1}{y^2} dy \quad \text{en posant } y = \frac{1}{x} \\ &= \int_0^{+\infty} f(y) C \exp^{-\beta y} y^{(\alpha+1)-1} dy \end{aligned}$$

La densité de $\frac{1}{Z_t}$ est proportionnelle à $\exp^{-\beta y} y^{(\alpha+1)-1}$. On reconnaît la densité d'une $\Gamma((\alpha + 1), \beta)$ \square

8.5.4 Correspondance $\Gamma - \chi^2$

Lemme 5. Correspondance $\Gamma - \chi^2$

$$\forall (a, \sigma) \in \mathbb{R}^{+*2},$$

$$\Gamma \left(\frac{a}{2}, \frac{1}{2\sigma^2} \right) = \sigma^2 \chi^2(a). \quad (8.5.10)$$

Démonstration. Soit $Z \sim \sigma^2 \chi^2(a)$. Alors $Z = \sigma^2 X$ ou $X \sim \chi^2(a)$. Alors, soit f fonction mesurable

bornée.

$$\begin{aligned}
 E[f(Z)] &= E[f(\sigma^2 X)] \\
 &= \int_0^{+\infty} f(\sigma^2 x) p_X(x) dx \\
 &= \int_0^{+\infty} f(\sigma^2 x) \frac{(1/2)^{a/2}}{\Gamma(a/2)} x^{a/2-1} e^{-x/2} dx \\
 &= \int_0^{+\infty} f(y) \frac{(1/2)^{a/2}}{\Gamma(a/2)} \frac{1}{(\sigma^2)^{a/2-1}} y^{a/2-1} e^{-\frac{y}{2\sigma^2}} \frac{1}{\sigma^2} dy \quad \text{en posant } y = \sigma^2 x
 \end{aligned}$$

On reconnaît la densité d'une Gamma de paramètres $\Gamma\left(\frac{a}{2}, \frac{1}{2\sigma^2}\right)$.

□

Cinquième partie

Conclusion Générale et Bibliographie

Chapitre 9

Conclusion générale et perspectives

Ce travail de thèse est consacré au développement de méthodes statistiques permettant l'analyse de données IRMf, en tenant compte de variabilités ignorées auparavant dans les modèles de la littérature. Les méthodes statistiques utilisées reposent sur l'algorithme EM de Demspster *et al.* [36] et l'algorithme SAEM-MCMC développé par Kuhn et Lavielle [73]. Nous avons proposé deux extensions de ce dernier algorithme aux modèles respectivement définis par équations différentielles ordinaires et équations différentielles stochastiques. Ces extensions sont appuyées par des développements mathématiques et les propriétés de convergence de l'algorithme SAEM-MCMC ont été étendues à ces différents cas.

Sous l'hypothèse de linéarité du signal BOLD, nous avons proposé une variante du modèle de convolution standard permettant d'inclure une variabilité inter-occurrences. Nous avons mis en évidence la validité de cette hypothèse pour différents sujets et différents aires cérébrales. Cependant, de nombreuses questions restent ouvertes notamment sur l'origine de cette variabilité. Afin de pouvoir envisager une étude à plus grande échelle (davantage de sujets, davantage de régions d'intérêt, davantage de types de stimuli), nous devons faciliter l'utilisation pratique de nos programmes informatiques, en les mettant par exemple au format standard du logiciel SPM.

Dans le contexte de modèles définis par équations différentielles ordinaires à paramètres non-observés, nous avons proposé des méthodes d'estimation à la fois dans un cadre bayésien et du maximum de vraisemblance, intégrant une méthode d'intégration numérique du système différentiel. Nous avons montré la convergence des estimateurs proposés et avons borné l'erreur sur ces estimateurs induite par l'approximation numérique de la solution de l'EDO.

La méthode d'estimation par maximum de vraisemblance a été appliquée dans le cadre de données IRMf sur le "*Balloon model*". Dans ce cadre, nous avons introduit une variabilité supplémentaire, considérant ainsi, sur données réelles, des efficacités neuronales dépendant du type de stimuli. Sur ce jeu de données réelles, nous avons pu valider par un test du maximum de vraisemblance l'hypothèse de variabilité des efficacités neuronales. D'autre part, nous avons mis en évidence un phénomène de *suppression à la répétition*, au niveau neuronal.

La mise en évidence de ce dernier phénomène, jusque là avant tout observé au niveau hémodynamique, ouvre de nombreuses perspectives. Des études plus vastes doivent être envisagées (plus de sujets, autres types de stimuli, autres aires cérébrales) pour confirmer ces premières observations. L'introduction de nouvelles variabilités semble être la prochaine étape essentielle.

Cependant, le temps de calcul est un frein indéniable à l'utilisation des modèles physiologiques. En effet, en raison de la complexité du système dynamique résolu, une estimation requiert plusieurs jours de temps de calcul. Le facteur limitant essentiel semble être la puissance de l'ordinateur utilisé. Cependant, nous pouvons espérer des progrès au niveau algorithmique, par exemple en résolvant de façon plus judicieuse les équations différentielles.

D'autre part, alors que sur des paradigmes "blocs" (tels que celui présenté au chapitre 6) le "*Balloon model*" fournit de bons ajustements, les premiers essais réalisés sur données réelles dans le cadre d'un paradigme événementiel composé de stimuli auditifs, clics, etc, ont été peu concluants, nous laissant à penser que le "*Balloon model*" n'est pas adapté à de telles conditions, pour des raisons encore inexpliquées. Cette piste doit être explorée.

Nous avons considéré des modèles définis par équations différentielles stochastiques à paramètres aléatoires. Des méthodes d'estimation bayésienne et par maximum de vraisemblance ont été développées. Elles reposent sur une approximation de la diffusion par la méthode d'Euler-Maruyama, approximation obtenue en introduisant des temps intermédiaires entre les instants de mesure. La convergence des algorithmes a été démontrée sous des conditions générales. L'erreur induite par l'approximation d'Euler-Maruyama a été contrôlée par le pas de la discrétisation. La précision de la méthode d'estimation par maximum de vraisemblance a été illustrée par une étude sur données simulées à partir d'un modèle non-linéaire à effets mixtes issu de la pharmacocinétique. En outre, l'analyse du jeu de données réelles Theophyllin illustre la pertinence de l'approche par SDE par rapport à l'approche déterministe (par ODE).

Le modèle défini par équations différentielle stochastiques dépendant de paramètres aléatoires n'a pas été appliqué aux données IRMf. En effet, les résultats ont été démontrés dans le cadre de processus de diffusion scalaires. Il pourrait être intéressant d'un point de vue théorique de prolonger cette étude en considérant des processus de diffusions vectoriels, en les approchant par exemple par le schéma de linéarisation locale proposé par Biscay *et al.* [18]. A l'heure actuelle, pour des raisons de temps de calcul, une application dans le cadre des données IRMf semble inenvisageable. D'autres améliorations algorithmiques (en terme de temps de calcul) devraient être d'abord envisagées.

L'étude de l'identifiabilité des modèles définis par équation différentielle stochastique dépendant d'un paramètre aléatoire est un travail difficile et aucune référence n'existe dans la littérature sur le sujet. Nous avons proposé quelques modèles pour lesquels nous sommes capables de prouver des résultats d'identifiabilité. Cependant beaucoup de travail reste à faire. Plusieurs voies semblent intéressantes à explorer, notamment l'étude complète de cas simples, et la discrétisation du processus permettant d'utiliser la théorie de processus auto-régressifs à coefficients aléatoires.

Bibliographie

- [1] M. Abramowitz and I.A. Stegun. *Handbook of mathematical functions*. Applied Mathematics. National bureau of standards, 55 edition, 1964.
- [2] G. K. Aguirre, E. Zarahn, and M. D’Esposito. The variability of human bold hemodynamic responses. *Neuroimage*, 7 :574, 1998.
- [3] Yacine Aït-Sahalia. Maximum likelihood estimation of discretely sampled diffusions : a closed-form approximation approach. *Econometrica*, 70(1) :223–262, 2002.
- [4] J. L. R. Andersson, Chloe Hutton, John Ashburner, Robert Turner, and Karl J. Friston. Modelling geometric deformations in epi times series. *Neuroimage*, 13 :903–919, 2001.
- [5] C Andrieu and E Moulines. On the ergodicity properties of some adaptive mcmc algorithms. *to appear*, 2005.
- [6] A. Arieli, A. Sterkin, A. Grinvald, and A. Aertsen. Dynamics of ongoing activity : explanation of the large variability in evoked cortical responses. *Science*, 273(5283) :1868–71, Sep 1996.
- [7] A. Aubert and R. Costalat. A model of the coupling between brain electrical activity, metabolism, and hemodynamics : application to the interpretation of functional neuroimaging. *Neuroimage*, 17 :1162–1181, 2002.
- [8] V Bally and D. Talay. The law of the Euler Scheme for Stochastic Differential Equations : I. Convergence Rate of the Density. Technical Report 2675, INRIA, 1995.
- [9] V. Bally and D. Talay. The law of the Euler scheme for stochastic differential equations (II) : convergence rate of the density. *Monte Carlo Methods Appl.*, 2 :93–128, 1996.
- [10] SL Beal and LB Sheiner. Estimating population kinetics. *Crit Rev Biomed Eng*, 8(3) :195–222, 1982.
- [11] J. E. Bennet, A. Racine-Poon, and J. C. Wakefield. *MCMC for nonlinear hierarchical models*, pages 339–358. Chapman & Hall, London, 1996.
- [12] G.S. Berns, A.W. Song, and H. Mao. Continuous functional magnetic resonance imaging reveals dynamic nonlinearities of dose-response curve for finger opposition. *J. Neurosci.*, 19 :RC–17, 1999.
- [13] A. Beskos, O. Papaspiliopoulos, G. O. Roberts, and P. Fearnhead. Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes. *J. R. Stat. Soc. B*, 68 :333–382, 2006.
- [14] A. Beskos and G.O. Roberts. Exact simulation of diffusions. *Ann. Appl. Prob.*, 15 :2422–2444, 2005.

-
- [15] Bo Martin Bibby and Michael Sørensen. Martingale estimation functions for discretely observed diffusion processes. *Bernoulli*, 1(1-2) :17–39, 1995.
- [16] R. Birn, R.W. Cox, and P. A. Bandettini. Detection versus estimation in event-related fMRI : choosing the optimal stimulus timing. *Neuroimage*, 15(1) :252–264, January 2002. Rapid communication.
- [17] R. Birn, Z.S. Saad, and P. A. Bandettini. Spatial heterogeneity of nonlinear dynamics in the fMRI BOLD response. *Neuroimage*, 14 :817–826, 2001.
- [18] R. Biscay, J. C. Jimenez, J. J. Riera, and P. A. Valdes. Local linearization method for the numerical solution of stochastic differential equations. *Ann. Inst. Statist. Math.*, 48(4) :631–644, 1996.
- [19] J. G. Booth and J. P. Hobert. Maximizing generalized linear mixed model likelihoods with an automated monte carlo em algorithm. *Journal of the Royal Statistical Society Series B*, 61 :265–285, 1999.
- [20] G. M. Boynton, S. A. Engel, G. H. Glover, and D. J. Heeger. Linear systems analysis of functional magnetic resonance imaging in human V1. *The Journal of Neuroscience*, 16 :4207–4221, 1996.
- [21] Paul Broca. Remarques sur le siège de la faculté du langage articulé, suivies d’une observation d’aphémie. *Bul. Mém. Soc. Anat.*, 2 :330–357, 1861.
- [22] Edward T. Bullmore, Michael Brammer, Steve C.R. Williams, Sophia Rabe-Hesketh, Nicolas Janot, Anthony David, John Mellers, Robert Howard, and Pak Sham. Statistical methods of estimation and inference for functional MR image analysis. *Magnetic Resonance in Medicine*, 35 :261–277, 1996.
- [23] R. Buxton and L. Frank. A model for the coupling between cerebral blood flow and oxygen metabolism during neural stimulation. *Journal of Cerebral Blood Flow Metabolism*, 17(1) :64–72, 1997.
- [24] R. B Buxton, K. Miller, L.R. Frank, and E.C. Wong. Bold signal dynamics : the balloon model with viscoelastic effects. In *Sixth Meeting, International Society for Magnetic Resonance in Medicine, Sydney, Australia*, page 1441, 1998.
- [25] R. B. Buxton, E. C. Wong, and L. R. Franck. Dynamics of blood flow and oxygenation changes during brain activation : The balloon model. *MRM*, 39 :855–864, 1998.
- [26] R. B. Buxton, E. C. Wong, and Frank. L. Dynamics of blood flow and oxygenation changes during brain activation : the balloon model. *Magnetic Resonance in Medicine*, 39 :855–864, June 1998.
- [27] Richard B. Buxton, Kâmil Uludag, David J. Dubowitz, and Thomas T. Liu. Modeling the hemodynamic response to brain activation. *Neuroimage*, 23 :220–233, 2004.
- [28] B. P. Carlin and T. A. Louis. *Bayes and empirical Bayes methods for data analysis*, volume 69 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London, 2000.
- [29] G. Celeux and J. Diebolt. The SEM algorithm : a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem. *Computational. Statistics Quaterly*, 2 :73–82, 1985.
- [30] H. Chernoff. On the distribution of the likelihood ratio. *Annals of mathematical Statistics*, 25 :573–578, 1954.

-
- [31] P. Ciuciu, J.-B. Poline, G. Marrelec, J. Idier, C. Pallier, and H. Benali. Robust estimation of the hemodynamic response function in asynchronous multitasks multiblocks event-related fMRI paradigms. in revision to *IEEE Transactions on Medical Imaging*, SHFJ/CEA, Orsay, France, August 2002.
- [32] Mark S. Cohen. Parametric analysis of MRI data using linear systems methods. *Neuroimage*, 6 :93–103, 1997.
- [33] D. Concordet and O. Nu nez. When is a nonlinear mixed-effects model identifiable? *Preprint*, [http : //www.biostat.envt.fr/](http://www.biostat.envt.fr/), 2002.
- [34] D. Dacunha-Castelle and D. Florens-Zmirou. Estimation of the coefficients of a diffusion from discrete observations. *Stochastics*, 19(4) :263–284, 1986.
- [35] B. Delyon, M. Lavielle, and E. Moulines. Convergence of a stochastic approximation version of the EM algorithm. *Ann. Statist.*, 27 :94–128, 1999.
- [36] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B*, 39(1) :1–38, 1977. With discussion.
- [37] T. Deneux. *Hemodynamic Models :Investigation and Application to Brain Imagery Analysis*. PhD thesis, Ecole Doctorale Polytechnique, 2005.
- [38] S Ditlevsen and Andrea De Gaetano. Mixed effects in stochastic differential equation models. *REVSTAT- Statistical Journal*, 3(2) :137–153, 2005.
- [39] JR Duann, TP Jung, WJ Kuo, TC Yeh, S Makeig, Hsieh JC, and Sejnowski TJ. Single-trial variability in event-related bold signals. *Neuroimage*, 15(4) :823–35, April 2002.
- [40] Ola Elerian, Siddhartha Chib, and Neil Shephard. Likelihood inference for discretely observed nonlinear diffusions. *Econometrica*, 69(4) :959–993, 2001.
- [41] Bjorn Eraker. MCMC analysis of diffusion models with application to finance. *J. Bus. Econ. Statist.*, 19(2) :177–191, 2001.
- [42] A. Evans. 3D multimodality human brain mapping : past, present and future. In K. Uemura et al., editor, *Quantification of Brain Function. Tracer Kinetics and Image Analysis in Brain PET*, pages 373–389. Elsevier Science, 1993.
- [43] G. Flandin, W. Penny, X. Pennec, N. Ayache, and J.-B. Poline. A multisubject anatomo-functional parcellation of the brain. In *Proceedings 9th HBM*, New York,, June19–22 2003.
- [44] Guillaume Flandin. *Utilisation d’informations géométriques pour l’analyse statistique des données d’IRM fonctionnelle*. PhD thesis, Université Sophia-Antipolis, Ecole Doctorale STIC, 2004.
- [45] R. S. J. Frackowiak, K. J. Friston, C. D. Dolan, and J. C. Maziotta. *Human Brain Function*. Academic Press, San Diego,, 1997.
- [46] Karl J. Friston, Oliver Josephs, Eric Zarahn, Andrew P Holmes, Stéphanie Rouquette, and Jean-Baptiste Poline. To smooth or not to smooth? bias ans efficiency in fMRI time series analysis. *Neuroimage*, 12 :196–208, 2000.
- [47] Karl J. Friston, Steeve C.R Williams, Robert Howard, Richard S.J. Frackowiak, and Robert Turner. Movements related effects in fMRI times series. *Magnetic Resonance in Medicine*, 35 :346–355, 1996.

- [48] K.J. Friston. Statistical parametric mapping. In R.W. Thatcher, M. Hallet, T. Zeffiro, E.R. John, and M. Huerta, editors, *Functional Neuroimaging : Technical Foundations*, pages 79–93, 1994.
- [49] K.J. Friston. Bayesian estimation of dynamical systems : an application to fMRI. *Neuroimage*, 16 :513–530, 2002.
- [50] K.J. Friston, A. Mechelli, R. Turner, and C. J. Price. Nonlinear responses in fMRI : the balloon model, Volterra kernels, and other hemodynamics. *Neuroimage*, 12 :466–477, June 2000.
- [51] Alan E. Gelfand and Adrian F. M. Smith. Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.*, 85(410) :398–409, 1990.
- [52] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter. *Markov chain Monte Carlo in practice*. Interdisciplinary Statistics. Chapman & Hall, London, 1996.
- [53] Arnaud Gloter and Jean Jacod. Diffusions with measurement errors. I. Local asymptotic normality. *ESAIM Probab. Stat.*, 5 :225–242, 2001.
- [54] G. H. Glover. Deconvolution of impulse response in event-related BOLD fMRI. *Neuroimage*, 9 :416–429, 1999.
- [55] Eugene V. Golanov, Seiji Yamamoto, , and Donald J. Reis. Spontaneous waves of cerebral blood flow associated with pattern of electrocortical activity. *The American Physiology Society*, pages 204–214, 1994.
- [56] C. Gössl, L. Fahrmeir, and D. P. Auer. Bayesian modeling of the hemodynamic response function in BOLD fMRI. *Neuroimage*, 14 :140–148, 2001.
- [57] C. Gourieroux and Monfort. *Statistique et modèles économétriques volume 2 Tests, régions de confiance, choix de modèles, théorie asymptotique*. Econometrica, 1995.
- [58] C. Goutte, F. A. Nielsen, and L. K. Hansen. Modeling the haemodynamic response in fMRI using smooth filters. *IEEE Transactions on Medical Imaging*, 19(12) :1188–1201, December 2000.
- [59] E. Hairer, S.P. Nørsett, and G. Wanner. *Solving Ordinary Differential Equations I, Nonstiff Problems*. Springer, Berlin, 1987.
- [60] J.M Hammersley and D.C Handscomb. *Monte Carlo Methods*. Methuen, London, 1964.
- [61] Rick N.A. Henson, Christian Büchel, Olivier Josephs, and Karl J. Friston. The slice-timing problem in event-related fMRI. *Neuroimage Human Brain Mapping conference (HBM99)*, 1999.
- [62] D. L. G. Hill, P. G. Batchelor, M. Holden, and D.J. Hawkes. Medical image registration. *Phys. Med. Biol.*, 46 :R1–R45, 2001.
- [63] R. Hoge, B. Atkinson, B. Gill, and G. Crelier. Linear coupling between cerebral blood flow and oxygen consumption in activated human cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 96(16) :9403–9408, 1999.
- [64] Olivier Houdé, Mazoyer Bernard, and Nathalie Tzourio-Mazoyer. *Cerveau et Psychologie : Introduction à l’Imagerie Cérébrale Anatomique et Fonctionnelle*. Presses Universitaires de France, 2002.
- [65] Y. Huang, D. Liu, and H. Wu. Hierarchical bayesian methods for estimation of parameters in a longitudinal hiv dynamic system. *Technical Report 04/06*, 2004.

-
- [66] S.A. Huettel and G. McCarthy. Evidence for a refractoriness period in the hemodynamic response to visual stimuli as measured by MRI. *Neuroimage*, 11 :547–553, 2000.
- [67] J.L. Jensen and N.V. Petersen. Asymptotical normality of the maximum likelihood estimator in state space models. *Ann. Statist.*, 27 :514–535, 1999.
- [68] J. C. Jimenez. A simple algebraic expression to evaluate the local linearization schemes for stochastic differential equations. *Appl. Math. Lett.*, 15(6) :775–780, 2002.
- [69] J. C. Jimenez, R. Biscay, C. Mora, and L. M. Rodriguez. Dynamic properties of the local linearization method for initial-value problems. *Appl. Math. Comput.*, 126(1) :63–81, 2002.
- [70] P.E. Kloeden and E. Platen. *Numerical solution of stochastic differential equations*. Springer-Verlag, Berlin, 1992.
- [71] R. Krishna. *Applications of Pharmacokinetic principles in drug development*. Kluwer Academic/Plenum Publishers, New York, 2004.
- [72] E. Kuhn and M. Lavielle. Coupling a stochastic approximation version of EM with a MCMC procedure. *ESAIM Probab. Stat.*, 8 :115–131, 2004.
- [73] E Kuhn and M Lavielle. Maximum likelihood estimation in nonlinear mixed effects models. *Comput. Statist. Data Anal.*, 49 :1020–1038, 2005.
- [74] S. Kusuoka and D. Stroock. Applications of the malliavin calculus, part II. *J. Fac. Sci. Univ. Tokyo. Sect. IA, Math.*, 32 :1–76, 1985.
- [75] T. Kutoyants. *Parameter estimation for stochastic processes*. Helderman Verlag Berlin, 1984.
- [76] K. K. Kwong, J.W. Belliveau, D.A. Chesler, I.E. Goldberg, R.M. Weisskoff, B.P. Poncelet, D.N. Kennedy, B.E. Hoppel, M.S. Cohen, and R. Turner. Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. In *Proceedings of the National Academy of Sciences of the United States of America*, volume 89, pages 5675–5679, 1992.
- [77] N. Lange. Empirical and substantive models, the Bayesian paradigm, and meta-analysis in functional brain imaging. *Human Brain Mapping*, 5 :259–263, 1997.
- [78] Paul Lauterbur. Image formation by local induced local interactions : examples employing nuclear magnetic resonance -correlated fMRI of human alpha activity. *Nature*, 242 :190–191, 1973.
- [79] David A Leopold, Yusuke Murayama, and Nikos K Logothetis. Very slow activity fluctuations in monkey visual cortex : implications for functional brain imaging. *Cereb Cortex*, 13(4) :422–33, Apr 2003.
- [80] T. A. Louis. Finding the observed information matrix when using the EM algorithm. *J. Roy. Statist. Soc. Ser. B*, 44(2) :226–233, 1982.
- [81] Yingli Lu, Tianzi Jiang, and Yufeng Zang. Single-trial variable model for event-related fMRI data analysis. *IEEE Trans Med Imaging*, 24(2) :236–45, Feb 2005.
- [82] S. Makni, P. Ciuciu, J. Idier, and J.-B. Poline. Semi-blind deconvolution of neural impulse response in event-related fMRI using Gibbs sampler. In *Proceedings 2th International Symposium on Biomedical Imaging*, pages 860–863, Arlington, VA, April 2004.

- [83] J.B. Mandeville, J.J.A. Marota, C. Ayata, C. Zaharchuck, M.A. Moskowitz, B.R. Rosen, and R.M Weisskoff. Evidence of a cerebro-vascular post-arteriole windkessel with delayed compliance. *Journal of Cerebral Blood Flow Metabolism*, 19 :679–689, 1999.
- [84] D. W. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *SIAM J. Appl. Math.*, 11 :431–441, 1963.
- [85] G. Marrelec, H. Benali, P. Ciuciu, and J.-B. Poline. Bayesian estimation of the hemodynamic response function in functional MRI. In Robert Fry, editor, *Bayesian Inference and Maximum Entropy Methods*, Baltimore, MD, August 2001. MaxEnt Workshops.
- [86] G. Marrelec, M. Péligrini-Issac, and H. Benali. Conditional independence graphs for functional connectivity in functional MRI. In *Proceedings of the first International Symposium on Biomedical Imaging*, pages 903–906, Washington, DC, July 2002.
- [87] C Mathiesen, K Caesar, N Akgoren, and M Lauritzen. Modification of activity- dependent increases of cerebral blood flow by excitatory synaptic activity and spikes in rat cerebellar cortex. *J Physiol*, 512 :555–566, 1995.
- [88] Bernard Mazoyer. *L'imagerie Cérébrale Fonctionnelle*. Presses Universitaires de France, 2001.
- [89] Xiao-Li Meng and Donald B. Rubin. Maximum likelihood estimation via the ECM algorithm : a general framework. *Biometrika*, 80(2) :267–278, 1993.
- [90] S.P. Meyn and R.L. Tweedie. *Markov chains and stochastic stability*. Comm. Control. Engrg. Ser. Springer-Verlag London Ltd., London, 1993.
- [91] Karla L. Miller, Wen-Ming Luh, Thomas T. Liu, Antigona Martinez, Takayuki Obata, Eric C. Wong, Lawrence R. Frank, and Richard B. Buxton. Nonlinear temporal dynamics of the cerebral blood flow response. *Humain Brain Mapping*, 13 :1–12, 2001.
- [92] F. A. Nielsen, L. K. Hansen, P. Toft, C. Goutte, N. Lange, S. C. Stroher, N. Morch, C. Svarer, R. Savoy, B. Rosen, E. Rostrup, and P. Born. Comparison of two convolution models for fMRI time series. *Neuroimage*, 5 :S473, 1997.
- [93] T. Obata, T. T. Liu, K. L. Miller, W. M. Luh, E. C. Wong, and R. B. Buxton. Discrepancies between bold and flow dynamics in primary and supplementary motor areas : application of the balloon model to the interpretation of bold transients. *NeuroImage*, 21 :144–153, 2004.
- [94] S. Ogawa, D.W. Tank, R. Menon, J.M. Ellermann, S-G. Kim, H. Merkle, and K. Ugurbil. Intrinsic signal changes accompanying sensory stimulation : functional brain mapping with magnetic resonance imaging. *Proceedings of the National Academy of Sciences of the United States of America*, 89 :5951–5955, 1992.
- [95] R.V. Overgaard, N. Jonsson, C.W. Tornøe, and H. Madsen. Non-linear mixed-effects models with stochastic differential equations : Implementation of an estimation algorithm. *J Pharmacokinet. Pharmacodyn.*, 32(1) :85–107, 2005.
- [96] A.-L. Paradis. *Exploration par IRM fonctionnelle de la perception visuelle des formes tridimensionnelles ; dessin de paradigmes d'acquisition et outils d'analyse appliqués au domaine de la vision*. Phd thesis, Univ. de Paris-VI, Paris, France, March 2001.
- [97] Asger Roer Pedersen. A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scand. J. Statist.*, 22(1) :55–71, 1995.
- [98] W. Penfield and T. Rassmussen. *The cerebral cortex of man*. Macmillan, 1950.

- [99] J.C. Pinheiro and D.M. Bates. Approximations to the log-likelihood function in the non-linear mixed-effect models. *J. Comput. Graph. Statist.*, 4 :12–35, 1995.
- [100] J.C. Pinheiro and D.M. Bates. *Mixed-effect models in S and Splus*. Springer-Verlag, 2000.
- [101] R. Poulsen. Approximate maximum likelihood estimation of discretely observed diffusion process. *Center for Analytical Finance*, Working paper 29, 1999.
- [102] B.L.S. Prakasa Rao. *Statistical Inference for Diffusion Type Processes*. Arnold Publisher, 1999.
- [103] J. C. Rajapakse, F. Kruggel, J. M. Maisog, and D.Y. Von Cramon. Modeling hemodynamic response for analysis of functional MRI time-series. *Human Brain Mapping*, 6 :283–300, 1998.
- [104] J. I. Ramos and C. M. García-López. Piecewise-linearized methods for initial-value problems. *Appl. Math. Comput.*, 82(2-3) :273–302, 1997.
- [105] J.J. Riera, J. Watanabe, I. Kazuki, M. Naoki, E. Aubert, T. Ozaki, and R. Kawashima. A state-space model of the hemodynamic approach : nonlinear filtering of BOLD signal. *Neuroimage*, 21 :547–567, 2004.
- [106] C.P. Robert and G. Casella. *Monte Carlo statistical methods*. Springer-Verlag, 2002.
- [107] G. O. Roberts and J. S. Rosenthal. Coupling and ergodicity of adaptive mcmc. *to appear*, 2005.
- [108] G. O. Roberts and O. Stramer. On inference for partially observed nonlinear diffusion models using the Metropolis-Hastings algorithm. *Biometrika*, 88(3) :603–621, 2001.
- [109] A. Roche. *Recalage d’images médicales par inférence statistique*. PhD thesis, Univ. de Nice-Sophia Antipolis, France, February 2001.
- [110] A. Roche, G. Malandain, X. Pennec, and N. Ayache. The correlation ratio as a new similarity measure for multimodal image registration. In *Proceedings 1st International Conference on Medical Image Computing and Computer Assisted Intervention*, volume 1496 of *Lecture Notes in Computer Science*, pages 1115–1124, Cambridge, MA, October 1998. Springer Verlag.
- [111] C.W. Roy and C.S. Sherrington. On the regulation of the blood supply of the brain. *Journal of Physiology*, 11 :85–108, 1890.
- [112] F. Schwegge. Evaluation of likelihood function for gaussian signals. *IEEE Trans. Inf. Theory*, 11 :61–70, 1965.
- [113] H. Singer. Continuous-time dynamical systems with sampled data, error of measurement and unobserved components. *J. Time Series Anal.*, 14 :527–545, 1993.
- [114] H. Sørensen. Parametric inference for diffusion processes observed at discrete points in time : a survey. *Int. Stat. Rev.*, 72(3) :337–354, 2004.
- [115] Michael Sørensen. Prediction-based estimating functions. *Econom. J.*, 3(2) :123–147, 2000.
- [116] D.J. Spiegelhalter, A. Thomas, N.G. Best, and W.R. Gilks. Bugs : Bayesian inference using gibbs sampling, version 0.5. Technical report, 1996.
- [117] David D. Stark and William G. Bradley. *Magnetic Resonance Imaging*. Mosby-Year Book, 1992.
- [118] J. Talairach and P. Tournoux. *Co-Planar Stereotaxic Atlas of the Human Brain. 3-Dimensional Proportional System : An Approach to Cerebral Imaging*. Thieme Medical Publishers, Inc., Georg Thieme Verlag, Stuttgart, New York, 1988.

-
- [119] J. Talairach and P. Tournoux. *Referentially Oriented Cerebral MRI*. Georg Thieme Verlag, Stuttgart, New York, 1993.
- [120] J.K. Thompson, M.R. Peterson, and R.D. Freeman. Single-neuron activity and tissue oxygenation in the cerebral cortex. *Science*, 299 :1070–1072, 2003.
- [121] Luke Tierney. Markov chains for exploring posterior distributions. *Ann. Statist.*, 22(4) :1701–1762, 1994.
- [122] CW Tornøe, RV Overgaard, H Agersø, HA Nielsen, H Madsen, and EN Jonsson. Stochastic differential equations in NONMEM : implementation, application, and comparison with ordinary differential equations. *Pharm. Res.*, 22(8) :1247–58, 2005.
- [123] WG Tracewell, DL Trump, WP Vaughan, DC Smith, and PR Gwilt. Population pharmacokinetics of hydroxyurea in cancer patients. *Cancer Chemother. Pharmacol.*, 35(5) :417–22, 1995.
- [124] Pierre-François Van de Moortele. *IRM Fonctionnelle à 3 Tesla : développements méthodologiques*. PhD thesis, Université Paris XI, 1999.
- [125] E. F. Vonesh. A note on the use of Laplace’s approximation for nonlinear mixed-effects models. *Biometrika*, 83(2) :447–452, 1996.
- [126] J. Wakefield, A. Smith, A. Racine-Poon, and A. Gelfand. Bayesian analysis of linear and nonlinear population models using the gibbs sampler. *Appl. Statist*, 43 :201–221, 1994.
- [127] G. C. G. Wei and M. A. Tanner. Calculating the content and boundary of the highest posterior density region via data augmentation. *Biometrika*, 77(3) :649–652, 1990.
- [128] Mike West and Jeff Harrison. *Bayesian forecasting and dynamic models*. Springer Series in Statistics. Springer-Verlag, New York, second edition, 1997.
- [129] R. P. Woods, J. C. Mazziotta, and S. R. Cherry. Automated image registration. In K. Uemura et al., editor, *Quantification of Brain Function. Tracer Kinetics and Image Analysis in Brain PET*, pages 391–399. Elsevier Science, 1993.
- [130] M. Woolrich, T. Behrens, Ch. Beckmann, and S. Smith. Mixture models with adaptive spatial regularization for segmentation with an application to fMRI data. *IEEE Transactions on Medical Imaging*, 24(1) :1–11, January 2005.
- [131] M. Woolrich, M. Jenkinson, J. Brady, and S. Smith. Fully Bayesian spatio-temporal modelling of fMRI data. *IEEE Transactions on Medical Imaging*, 23(2) :213–231, February 2004.
- [132] K. Worsley, C.H. Liao, J. Aston, V. Petre, G.H. Duncan, F. Morales, and A.C. Evans. A general statistical analysis for fMRI data. *Neuroimage*, 15(1) :1–15, January 2002.
- [133] Jeff C.-F. Wu. On the convergence properties of the EM algorithm. *Ann. Statist*, 11 :95–103, 1983.
- [134] L. Wu. Exact and approximate inferences for nonlinear mixed-effects models with missing covariates. *J. Amer. Statist. Assoc.*, 99(467) :700–709, 2004.
- [135] E. Zarahn, G. K. Aguirre, and M. D’Esposito. Empirical analysis of BOLD fMRI statistics. I. Spatially unsmoothed data collected under null-hypothesis conditions. *Neuroimage*, 5 :179–197, 1997.
- [136] Ying Zheng, David Johnston, Jason Berwick, Danmei Chen, Steve Billings, and John Mayhew. A three-compartment model of the hemodynamic response and oxygen delivery to brain. *NeuroImage*, 28 :925–939, 2005.

Résumé

Cette thèse est consacrée à l'analyse de données d'Imagerie par Résonance Magnétique fonctionnelle (IRMf). Dans le cadre du modèle classique de convolution, nous testons l'hypothèse de variabilité inter-occurrences des amplitudes des réponses hémodynamiques. L'estimation des paramètres de ce nouveau modèle requiert le recours à l'algorithme *Expectation-Maximisation*. Nous comparons ce modèle au modèle sans variabilité inter-occurrences par un test du rapport des vraisemblances, sur un grand nombre de jeu de données réelles. Le modèle linéaire souffrant d'un manque de fondement biologique, nous considérons un modèle physiologique aboutissant à l'écriture du signal IRMf comme la somme d'un terme de régression, solution d'une équation différentielle ordinaire (EDO), sans solution analytique dépendant d'un paramètre aléatoire, et d'un bruit de mesure gaussien. Nous proposons une méthode générale d'estimation paramétrique des modèles définis par EDO à données non-observées, intégrant une méthode de résolution numérique du système dynamique et reposant sur une version stochastique de l'algorithme EM. Nous montrons la convergence des estimateurs des paramètres produits par cet algorithme, et contrôlons l'erreur induite par l'approximation de la solution du système différentiel sur l'estimation des paramètres. Nous appliquons cette méthode à la fois sur données d'IRMf simulées et réelles. Enfin, nous considérons des modèles définis par équations différentielles stochastiques (EDS) dépendant d'un paramètre aléatoire. En approchant la diffusion par un schéma numérique, nous proposons une méthode d'estimation des paramètres du modèle. La précision de cette méthode est illustrée sur une étude sur données simulées dans le cadre d'un modèle à effets mixtes, issus de la pharmacocinétique. Une étude sur données réelle démontre la pertinence de l'approche stochastique. Finalement, nous nous intéressons à l'identifiabilité des modèles définis par EDS dépendant de paramètres aléatoires.

Mots-Clefs : Donnée IRMf - Modèles à données non-observées- Algorithmes MCMC - Estimation par maximum de vraisemblance - Algorithme EM/SAEM - Tests d'hypothèses - Estimation Bayésienne - Equation différentielle ordinaire- Equation différentielle stochastique - Processus de diffusion - Diffusions mixtes

Classification AMS(2000) : 62J02, 62J12, 62F03, 62F10, 62F15, 62M99, 62P10, 68W01

Abstract

This thesis is devoted to the analysis of functional Magnetic Resonance Imaging data (fMRI). In the framework of standard convolution models, we test a model that allows for the variation of the magnitudes of the hemodynamic response. To estimate the parameters of this model, we have to resort to the *Expectation-Maximisation* algorithm. We test this model against the standard one –with constant magnitudes– on several real data, set by a likelihood ratio test. The linear model suffers from a lack of biological basis, hence we consider a physiological model. In this framework, we describe the data as the sum of a regression term, defined as the non-analytical solution of an ordinary differential equation (ODE) depending on random parameters, and a gaussian observation noise. We develop a general method to estimate the parameters of a statistical model defined by ODE with non-observed parameters. This method, integrating a numerical resolution of the ODE, relies on a stochastic version of the EM algorithm. The convergence of the algorithm is proved and the error induced by the numerical solving method is controlled. We apply this method on simulated and real data sets. Subsequently, we consider statistical models defined by stochastic differential equations (SDE) depending on random parameters. We approximate the diffusion process by a numerical scheme and propose a general estimation method. Results of a pharmacokinetic mixed model study (on simulated and real data set) illustrate the accuracy of the estimation and the relevance of the SDE approach. Finally, the identifiability of statistical models defined by SDE with random parameters is studied.