HAL
open science

Modèles d'encodage parcimonieux de l'activité cérébrale mesurée par IRM fonctionnelle

Christine Bakhous

To cite this version:

Christine Bakhous. Modèles d'encodage parcimonieux de l'activité cérébrale mesurée par IRM fonctionnelle. Mathématiques générales [math.GM]. Université de Grenoble, 2013. Français. NNT : 2013GRENM071 . tel-00933426v2

HAL Id: tel-00933426

https://theses.hal.science/tel-00933426v2

Submitted on 13 May 2014

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

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

UNIVERSITÉ DE GRENOBLE

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Mathématiques appliquées**

Arrêté ministériel : 7 août 2006

Présentée par

Christine BAKHOUS

Thèse dirigée par **Florence FORBES**
et codirigée par **Michel DOJAT**

préparée au sein **de INRIA Grenoble, Laboratoire Jean Kuntzmann**
et de **l'École Doctorale Mathématiques, Sciences et Technologies de l'Information, Informatique**

Modèles d'encodage parcimonieux de l'activité cérébrale mesurée par IRM fonctionnelle

Thèse soutenue publiquement le **10 décembre 2013**,
devant le jury composé de :

M. Jérome IDIER
Directeur de Recherche, IRCCyN, UMR CNRS 6597, Président

M. Habib BENALI
Directeur de Recherche, Laboratoire d'Imagerie Fonctionnelle - UMR-S 678, INSERM - Université Pierre et Marie Curie, Rapporteur

M. Jean-Yves TOURNERET
Professeur, INP - ENSEEIHT Toulouse, Rapporteur

M. Philippe CIUCIU
Chercheur, CEA-NeuroSpin, INRIA Saclay, Examinateur

M. Michel DOJAT
Ingénieur de Recherche, GIN, UMR-S 836, INSERM, Université de Grenoble, Co-Directeur de thèse

Mme. Florence FORBES
Directeur de Recherche, LJK, INRIA Rhône-Alpes, Directeur de thèse

Remerciements

"Nous avons besoin des difficultés, car sans elles,
le succès n'aurait pas la même saveur"

Un grand merci à ma mère Micheline et mon père Georges pour leurs encouragement et confiance. Leur fort soutien moral m'a permi en complétant le support scientifique apporté par mes directeurs Florence, Michel et Philippe d'aller jusqu'au bout de ce travail de thèse. Merci à tous les trois pour leurs attention, expérience et disponibilité.

Les agréables moments étant un élément important de la réussite, je tiens à remercier les *mistisiens* : Gildas, Angelika, Jonathan, Flor, Senan, Thomas, Lotfi, Farida, Alessandro, El-Hadji, Nourou, Marie-José, Jean-Baptiste, Stéphane, Imma, Françoise, Huu-Giao, Ludovic, Laurent, Kai, Laure et Darren ainsi que tous ceux que j'ai eu le plaisir de rencontrer à Grenoble : Meryam, Ahmad, Alexandre, Fédérico et mon équipe du GIN.

J'aimerai également adresser un merci exceptionnel à mon adorable frère, Christophe, et une très forte pensée pour mes amis syriens : Tarek, Kousai, Leona, Rania, Rawan, Mirna et tous ceux avec qui j'ai passé des moments inoubliables et que j'espère revoir très bientôt.

Enfin, je n'oublierai jamais mes professeurs Said, Bassel et Valérie ayant fortement marqué mon arrivée en France ainsi que mes premiers pas dans le domaine de la recherche.

Table des matières

Résumé **9**

Abstract **11**

Introduction **13**

I Imagerie par résonance magnétique fonctionnelle **18**

I.1 Système nerveux central 18
I.2 Imagerie médicale 20
I.3 Imagerie par résonance magnétique - IRM 22
I.4 Imagerie par résonance magnétique fonctionnelle - IRMf 26
I.5 Paradigme expérimental et acquisition des données 28
I.6 Pré-traitements des données IRMf 30
I.6.1 Correction du décalage temporel - Slice timing 30
I.6.2 Correction du mouvement - Recalage 30
I.6.3 Recalage anatomo-fonctionnel 32
I.6.4 Normalisation spatiale 33
I.6.5 Lissage spatial 34
I.7 Conclusion 35

II Analyse Statistique en IRMf **37**

II.1 Méthodes classiques de détection 37
II.1.1 Modèle Linéaire Général - MLG 37
II.1.1.1 Estimation univariée 40
II.1.1.2 Tests statistiques univariés 41
II.1.2 Extensions multivariées du MLG 44
II.1.3 Modélisation de la corrélation temporelle 45
II.2 Modélisation de la FRH 45

II.2.1 Modèles physiologiques 46
II.2.2 Modèles non-physiologiques 46
II.2.2.1 Modèles paramétriques 46
II.2.2.2 Modèles semi-paramétriques 47
II.2.2.3 Modèles non-paramétriques 48
II.3 Extensions bayésiennes du MLG 49
II.3.1 Estimation multivariée 51
II.3.2 RIF régularisée temporellement 52
II.3.3 Cartes de probabilités *a posteriori* (PPMs) 53
II.4 Détection-Estimation Conjointe (DEC) 54
II.4.1 Modèle bayésien hiérarchique 55
II.4.2 Choix de la parcellisation 60
II.5 Conclusion 61

III Approches parcimonieuses en IRMf **64**
III.1 MLG creux (*sparse*) 66
III.2 Approches DEC parcimonieuses 67
III.2.1 Introduction d'une variable de pertinence 67
III.2.2 Connexion aux variables de classification 70
III.2.2.1 Simulations de Monte Carlo par Chaînes de Markov - MCMC 70
III.2.2.2 Algorithme EM Variationnel - VEM 79
III.2.3 Connexion aux moyennes d'activation 89
III.2.4 Automatic Relevance Determination - ARD 92
III.2.5 Energie Libre en VEM 97
III.3 Conclusion : 98

IV Expériences et résultats **100**
IV.1 Données simulées 101
IV.1.1 Procédure de simulation 101
IV.1.2 Résultats d'analyse et comparaison de modèles 102
IV.1.2.1 Technique stochastique - MCMC 104
IV.1.2.2 Technique variationnelle - VEM 108
IV.1.3 Conclusion 124
IV.2 Données réelles 124
IV.2.1 Premier sujet 126

IV.2.1.1 Technique stochastique - MCMC 127
IV.2.1.2 Technique variationnelle - VEM 129
IV.2.1.3 Conclusion 134
IV.2.2 Deuxième sujet 135
IV.2.2.1 ROI motrice droite 137
IV.2.2.2 ROI motrice gauche 137
IV.2.2.3 Conclusion 144

V Analyses cerveau entier **147**
V.1 Résultats préliminaires 147
V.2 Conclusion 154

VI Discussion et Perspectives **158**

A Matrice de covariance de l'approche RIF régularisée **162**

B Matrice de dessin de l'approche DEC **165**

C Simulations de Monte Carlo par Chaînes de Markov **170**

D Algorithme EM Variationnel **184**
D.1 Connexion aux variables de classification 184
D.2 Connexion aux moyennes d'activations 203

E Energie libre **208**

F Rappels utiles **214**
F.1 Algorithme de Metropolis-Hasting 214
F.2 Distribution gaussienne tronquée 215
F.3 *a priori* de Jeffrey 217
F.4 Algorithme de dichotomie 217
F.5 Algorithme du gradient 218
F.6 Entropie 219
F.7 Échantillonneur de Gibbs 220

Liste des publications issues de cette thèse **223**

Bibliographie **225**

Résumé

L'imagerie par résonance magnétique fonctionnelle (IRMf) est une technique non invasive permettant l'étude de l'activité cérébrale au travers des changements hémodynamiques associés. Récemment, une technique de détection-estimation conjointe (DEC) a été développée permettant d'alterner (1) la détection de l'activité cérébrale induite par une stimulation ainsi que (2) l'estimation de la fonction de réponse hémodynamique caractérisant la dynamique vasculaire ; deux problèmes qui sont généralement traités indépendamment. Cette approche considère une parcellisation *a priori* du cerveau en zones fonctionnellement homogènes et alterne (1) et (2) sur chacune d'entre elles séparément.

De manière standard, l'analyse DEC suppose que le cerveau entier peut être activé par tous les types de stimuli (visuel, auditif, etc.). Cependant la spécialisation fonctionnelle des régions cérébrales montre que l'activité d'une région n'est due qu'à certains types de stimuli. La prise en compte de stimuli non pertinents dans l'analyse, peut dégrader les résultats. La sous-famille des types de stimuli pertinents n'étant pas la même à travers le cerveau une procédure de sélection de modèles serait très coûteuse en temps de calcul. De plus, une telle sélection *a priori* n'est pas toujours possible surtout dans les cas pathologiques. Ce travail de thèse propose une extension de l'approche DEC permettant la sélection automatique des conditions (types de stimuli) pertinentes selon l'activité cérébrale qu'elles suscitent, cela simultanément à l'analyse et adaptativement à travers les régions cérébrales.

Des exemples d'analyses sur des jeux de données simulés et réels, illustrent la capacité de l'approche DEC parcimonieuse proposée à sélectionner les conditions pertinentes ainsi que son intérêt par rapport à l'approche DEC standard.

Abstract

Functional magnetic resonance imaging (fMRI) is a noninvasive technique allowing the study of brain activity via the measurement of hemodynamic changes. Recently, a joint detection-estimation (JDE) framework was developed and relies on both (1) the brain activity detection and (2) the hemodynamic response function estimation, two steps that are generally addressed in a separate way. The JDE approach is a parcel-based model that alternates (1) and (2) on each parcel successively.

The JDE analysis assumes that all delivered stimuli (e.g. visual, auditory, etc.) possibly generate a response everywhere in the brain although activation is likely to be induced by only some of them in specific brain areas. Inclusion of irrelevant events may degrade the results. Since the relevant conditions or stimulus types can change between different brain areas, a model selection procedure will be computationally expensive. Furthermore, criteria are not always available to select the relevant conditions prior to activation detection, especially in pathological cases. The goal of this work is to develop a JDE extension allowing an automatic selection of the relevant conditions according to the brain activity they elicit. This condition selection is done simultaneously to the analysis and adaptively through the different brain areas.

Analysis on simulated and real datasets illustrate the ability of our model to select the relevant conditions and its interest compare to the standard JDE analysis.

Introduction

La compréhension du cerveau, cet organe extrêmement complexe, représente une problématique interdisciplinaire située à la croisée de nombreuses disciplines, entre autres : biologie et médecine, informatique et mathématiques. Apparue à la fin du XIXème siècle, l'imagerie médicale est de nos jours très utilisée dans le domaine des neurosciences. Plus précisément, l'Imagerie par Résonance Magnétique fonctionnelle (IRMf) est apparue à la fin du XXième siècle pour la compréhension du cerveau sain ou pathologique ainsi que pour l'évaluation du risque chirurgical potentiel. Cette technique repose sur la mesure non-invasive du signal BOLD (*Blood Oxygen Level Dependent*) reflétant les variations hémodynamiques induites par l'activité neuronale du cerveau et donc sur l'existence d'un couplage neuro-vasculaire.

L'encodage et le décodage de l'activité cérébrale constituent deux axes de recherche importants en neurosciences. Le premier, s'intéresse à la réaction neuronale suite à différents types de stimuli, alors que le second vise la prédiction des éventuelles causes de cette activité. L'encodage, auquel on s'intéresse dans ce travail, étudie (i) la détection et la localisation de l'activité cérébrale correspondant aux différents types de stimuli ainsi que (ii) l'estimation de la Fonction de Réponse Hémodynamique (FRH) caractérisant la dynamique vasculaire.

Des modèles physiologiques (modèle du ballon et modèle de Windkessel) permettant une modélisation du couplage neuro-vasculaire sont apparus à la fin des années 90. Ces modèles réalistes et explicites sont néanmoins complexes et souffrent de problèmes d'identifiabilité et de complexité calculatoire. Dans une autre gamme de modèles, des approches statistiques ont été développées, qui, elles, ne cherchent pas nécessairement à modéliser finement le couplage neuro-vasculaire mais à en rendre compte en proposant un modèle des signaux mesurés.

Aujourd'hui, l'approche la plus classique est basée sur l'utilisation d'un Modèle Linéaire Général (MLG) qui fait intervenir la FRH mais en présuppose une forme canonique constante à travers le cerveau entier. Ceci ne reflète pas toujours l'état physiologique surtout dans les cas pathologiques et peut donc dégrader les résultats de détection. Des approches semi-paramétriques ont permis la prise en compte de petites variabilités inter-voxel, mais restent insuffisantes pour caractériser l'hémodynamique de certaines régions cérébrales ou certains cas pathologiques. D'autres approches non-paramétriques (e.g. Réponse Impulsionnelle Finie ou RIF) estiment une FRH par voxel permettant une meilleure prise en compte de sa variabilité. Des extensions de RIF régularisée temporellement (RIF-R) ont également été développées et consistent en l'introduction d'une structure d'auto-corrélation temporelle permettant d'avoir des FRH lisses de formes physiologiques réalistes. Les approches RIF et RIF-R nécessitent cependant l'augmentation de la durée de l'expérience afin de fournir suffisamment de données par rapport au nombre de paramètres à estimer. De plus, elles traitent chaque voxel indépendamment et n'utilisent pas l'information spatiale inter-voxel. Récemment, un modèle de Détection-Estimation Conjointe (DEC) basé sur une parcellisation *a priori* du cerveau en zones fonctionnellement homogènes (parcelles), a été développé et alterne entre la détection de l'activité cérébrale et l'estimation de la réponse hémodynamique en prenant en compte la corrélation spatiale de l'activité cérébrale à l'intérieur de chaque parcelle séparément. Cette approche utilise également la régularité temporelle proposée dans RIF-R. De plus, la définition d'une FRH par parcelle (groupe de voxels) permet d'augmenter la robustesse de l'estimation.

De manière standard, ces analyses supposent que chaque zone du cerveau peut être activée par tous les types de stimuli (visuel, auditif, etc.). Cependant le principe de la spécialisation fonctionnelle des régions cérébrales suggère que l'activité d'une région n'est due qu'à certains types de stimuli. La prise en compte de stimuli non pertinents dans l'analyse peut dégrader les résultats, en particulier quand la FRH est conjointement estimée. De plus la sélection *a priori* des conditions pertinentes pour chaque région cérébrale n'est pas toujours possible, comme dans le cas d'une pathologie par exemple. *Donnet et al.* [Donnet 2004] ont déjà traité ce problème en 2004 dans un cadre de sélection de modèles qui nécessite la comparaison de plusieurs modèles afin de choisir le meilleur. Cette procédure

nécessite l'exécution de tous les modèles possibles (2^M si M est le nombre de stimuli étudiés), de plus la famille de stimuli pertinents change entre les régions cérébrales. Une procédure de type sélection de variables permettrait une sélection plus adaptative de cette famille en évitant l'éxecution de tous les modèles. Une telle procédure a déjà été utilisée en IRMf par *Smith et al.* [Smith 2007] mais uniquement pour le problème de détection (voxels activés vs non-activés) et non pour le problème de sélection de stimuli pertinents. Les travaux de *Luo et al.* [Luo 2005] et de *Oikonomou et al.* [Oikonomou 2012] ont proposé, dans le contexte du MLG, une modélisation sélective des dérives basses fréquences, une source de bruit en IRMf due aux phénomènes physiques et physiologiques. Cela à travers une procédure RVM (*Relevance Vector Machine*) généralement utilisée pour des problèmes de régression et classification parcimonieuse dans l'apprentissage automatique bayésien (*machine learning*). Leurs travaux n'ont cependant pas abordé la sélection des conditions pertinentes.

Le but de ce travail de thèse est de mettre en place une procédure adaptative permettant la sélection automatique des conditions expérimentales impliquées dans l'activité cérébrale tout en permettant l'estimation de la FRH. Pour cela, nous proposons une extension de l'approche DEC, dans laquelle nous ajoutons une étape de sélection de variables permettant de sélectionner automatiquement les types de stimuli pertinents dans chacune des régions cérébrales individuellement. Ceci est réalisé en introduisant une variable binaire supplémentaire exprimant la pertinence (en terme d'activité évoquée) de chaque type de stimuli. Le modèle proposé sera par la suite appelé modèle *parcimonieux* et sera comparé au modèle *complet* qui prend lui en compte tous les types de stimuli. La question étant *Quel intérêt pourrait avoir l'élimination des conditions non pertinentes sur les résultats d'analyse : une meilleure estimation de la FRH, une meilleure détection des activations pour les conditions pertinentes, ou autre ?* Il est néanmoins important de signaler que toute amélioration pouvant être apportée par le modèle *parcimonieux* dépend de l'impact de la présence des conditions non pertinentes sur la qualité du modèle. Plus l'impact sera fort, plus l'approche *parcimonieuse* sera pertinente. Des analyses sur des jeux de données simulés et réels permettront d'éclaircir ce point et de répondre à la question principale qui nous intéresse.

Ce manuscrit de thèse est organisé comme suit. Le premier chapitre est consacré aux principes de l'IRMf et aux prétraitements nécessaires avant l'analyse

statistique des données détaillée dans le deuxième chapitre. Ce dernier explique les approches classiques et bayésiennes utilisées en IRMf pour la détection de l'activité cérébrale ainsi que pour la modélisation de la réponse hémodynamique. La problématique de sélection de conditions pertinentes est quant à elle détaillée et discutée dans le troisième chapitre qui présente les approches d'analyse *parcimonieuses* et plus spécifiquement celles proposées dans le cadre de cette thèse. Nous détaillons les deux techniques d'estimation mises en place dans le cadre de l'approche DEC : les méthodes à base de simulations stochastiques ou Méthodes de Monte Carlo par Chaînes de Markov [Bakhous 2012b, Bakhous 2012a] et les méthodes basées sur des approximations déterministes des dépendances stochastiques, soient les méthodes dites variationnelles (algorithme Expectation-Maximisation Variationnel) [Bakhous 2013b, Bakhous 2013a]. Pour mettre en évidence l'intérêt de cette contribution, des expériences sur données simulées et réelles, ainsi que des comparaisons entre les différentes approches, figurent dans le quatrième chapitre. Une analyse cerveau entier et les résultats préliminaires obtenus sont présentés dans le cinquième chapitre. Enfin, une discussion autour des différents travaux et perspectives à venir est développée en conclusion de ce manuscrit.

Chapitre I

Imagerie par résonance magnétique fonctionnelle

Avant de s'intéresser à l'essence de ce travail de thèse à travers les algorithmes et traitements statistiques des données, nous allons détailler l'origine des données auxquelles nous nous intéressons ainsi qu'à leur mode d'acquisition. Comme introduit précédemment notre domaine de recherche concerne l'interaction entre les statistiques et le domaine médical, plus précisément l'imagerie cérébrale dans le but d'une meilleure compréhension du cerveau. Pour cela nous allons commencer par introduire le système nerveux central dont fait partie notre cerveau pour passer ensuite aux principes physiques à la base de l'imagerie par résonance magnétique (IRM) et son extension fonctionnelle (IRMf). Pour finir, nous expliquerons comment est effectuée l'acquisition des données IRMf et les pré-traitements nécessaires à la correction de certains artéfacts qui précèdent le traitement statistique de ces données.

I.1 Système nerveux central

Le système nerveux central (figure I.1) est composé de l'encéphale, de la moëlle épinière et des nerfs spinaux (rachidiens). L'encéphale désigne trois parties :

1- Le cerveau qui représente le moteur principal de notre organisme et auquel on va s'intéresser dans ce travail de thèse.

2- Le cervelet, situé en dessous du cerveau et derrière le tronc cérébral. Il intervient principalement dans le contrôle moteur.

3- Le tronc cérébral, situé en dessous du cerveau et à l'avant du cervelet. Il contrôle certaines fonctions automatiques (indépendantes de la volonté) comme la respiration, le rythme cardiaque, les activités motrices involontaires (réflexes), etc.

La moëlle épinière quant à elle représente la continuation du tronc cérébral. Elle est située dans le canal rachidien protégé par la colonne vertébrale. Sa mission principale est la transmission de l'influx nerveux entre le cerveau et le reste du corps à travers des nerfs spinaux (rachidiens). Elle est aussi responsable de certains réflexes involontaires.

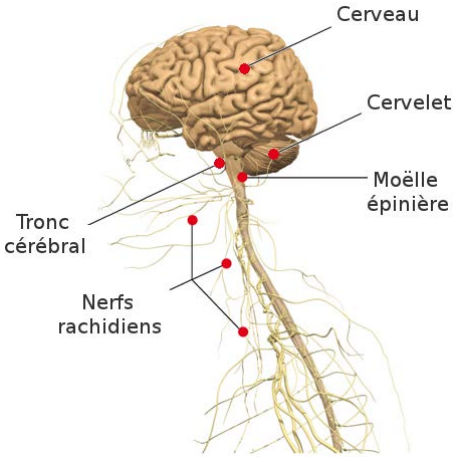


FIGURE I.1 – Système nerveux central humain composé de l'encéphale (cerveau, cervelet, tronc cérébral) et de la moëlle épinière ratachée aux nerfs spinaux (rachidiens) lui permettant la connexion avec le reste du corps (http://www.lecorpshumain.fr/).

Deux principaux types de cellules constituent le système nerveux :

1- Les neurones (figure I.2) représentent les cellules nerveuses excitables. Chacun est constitué principalement (i) du corps cellulaire, (ii) des dendrites, assurant sa connexion aux neurones voisins et permettant de conduire l'information portée par l'influx nerveux au corps cellulaire et (iii) un axone autorisant la circulation de l'information nerveuse du corps cellulaire aux synapses constituant des terminaisons permettant de passer l'information aux organes ou à d'autres neurones voisins.

2- Les cellules gliales qui jouent le rôle d'un environnement protecteur et nourrissier pour les neurones.

Comme le montre la figure I.3, le cerveau est constitué principalement de deux types de tissus (i) la substance grise ou le cortex contenant les corps cellulaires ainsi que les dendrites des neurones et (ii) la substance blanche contenant les axones des neurones et permettant ainsi de relier, entre elles, différentes régions de la substance grise.

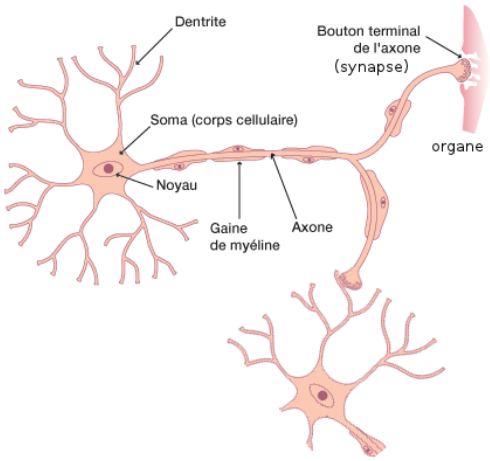


FIGURE I.2 – Cellule nerveuse (neurone) composée d'un corps cellulaire (soma), de dendrites assurant la connexion avec d'autres neurones et permettant la circulation de l'influx nerveux et d'un axone se terminant par des synapses permettant le transfert des signaux nerveux aux organes ou aux neurones voisins (http://opiaces-tpe.e-monsite.com/pages/morphine-et-heroine/mode-d-action.html).

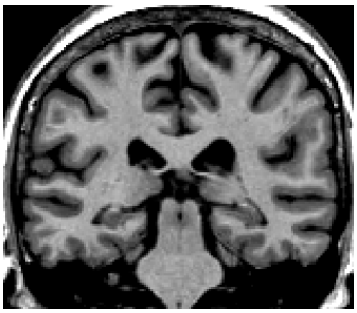


FIGURE I.3 – Coupe cérébrale coronale où deux tissus apparaîssent : la substance grise où se trouvent les corps cellulaires des neurones et leurs dendrites et la substance blanche contenant leurs axones.

L'organisme humain attire, par son extrême complexité, la curiosité des médecins et des chercheurs qui à travers des décennies ont développé des techniques, comme l'imagerie médicale par exemple, permettant de l'étudier dans le but de comprendre son fonctionnement.

I.2 Imagerie médicale

L'imagerie médicale, apparue en 1895 à travers la radiographie, représente de nos jours un domaine incontournable fournissant à la médecine une aide complémentaire au diagnostique, à la chirurgie ainsi qu'au suivi des patients. Elle constitue aussi un outil additionnel permettant aux chercheurs une meilleure compréhension de l'organisme humain et de son fonctionnement. Plus précisément, l'imagerie cérébrale a connu durant le XXième scièle, un développement rapide aboutissant à l'apparition de différentes techniques d'imagerie parmi lesquelles figurent des techniques anatomiques (scanner, TDM ou tomodensitométrie, IRM, ...) s'intéressant aux structures cérébrales ainsi qu'aux éléments pouvant les perturber (tumeurs, caillots ou autres), ou fonctionnelles (EEG, MEG, IRMf, TEP, ...) s'in-

téressant plutôt à l'activité de ces structures suites à des stimulations. Certaines d'entre elles sont dites invasives car elles nécessitent l'injection intraveineuse d'un produit de contraste ou d'un produit radioactif comme l'imagerie TEP ou même l'amincissement ou le retrait d'une partie du crâne comme c'est le cas pour l'imagerie optique, contrairement à d'autres dites non-invasives comme l'IRM, l'EEG, la MEG ou autres et qui sont donc préférables pour le sujet. De plus ces techniques diffèrent par rapport à leur résolution spatio-temporelle comme le montre la figure I.4. Parmi les techniques d'imagerie cérébrale fonctionnelle, il existe :

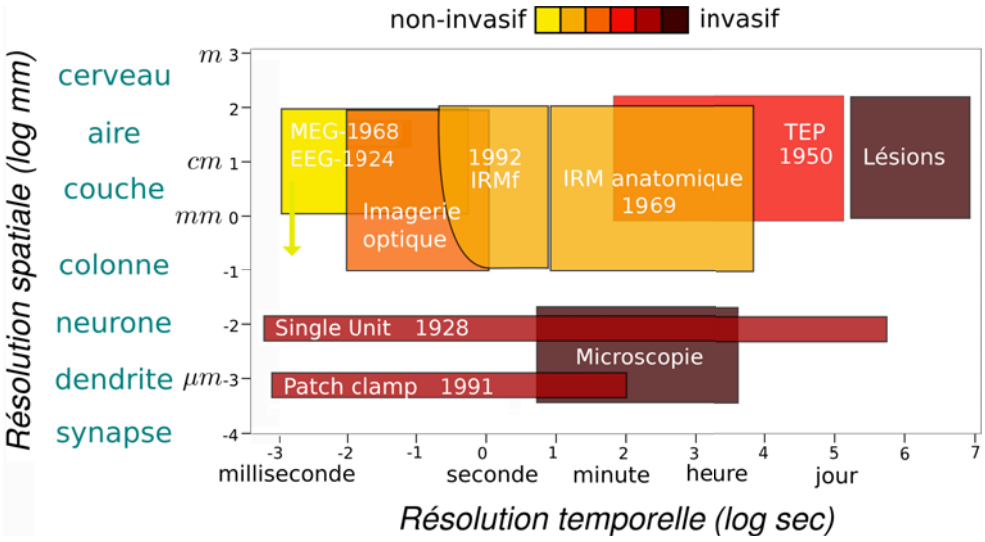


FIGURE I.4 – Comparaison de plusieurs techniques d'imagerie cérébrale par rapport à leur résolution spatio-temporelle.

- La tomographie par émission de positrons (TEP) est apparue en imagerie médicale dès 1950. Elle est basée sur la détection en coïncidence de deux photons obtenus du phénomène d'annihilation entre un électron du milieu et un positron émis par un traceur radioactif, dont l'injection intraveineuse rend cette technique invasive.
- L'électro-encéphalographie (EEG), découverte en 1924, permet de mesurer l'activité électrique du réseau neuronal à travers des électrodes (8-21 ou plus) placés sur le cuir chevelu.
- La magnéto-encéphalographie (MEG) a été mesurée pour la première fois par le physicien David Cohen en 1968 et est principalement utilisée dans la localisation pré-opératoire des foyers épileptiques. Cette technique permet la mesure des champs magnétiques induits par l'activité électrique des

neurones, ceci à travers une centaine de canaux couvrant le cerveau entier. L'EEG et la MEG bénéficient d'une résolution temporelle assez élevée (< 1 seconde) contrairement à leur faible résolution spatiale limitée par le nombre d'électrodes ou de canaux utilisés.

- L'imagerie par résonance magnétique (IRM), introduite en 1969 dans le domaine médical, se base sur les propriétés magnétiques des tissus vitaux et leurs interactions avec un champ magnétique externe permettant ainsi de les visualiser sous différents niveaux de gris. La branche fonctionnelle de l'IRM (IRMf) quant à elle est apparue en 1992 suite à des travaux concernant le magnétisme de l'hémoglobine qui diffère selon son niveau d'oxygénation. Contrairement à l'EEG et à la MEG, l'IRMf reflète indirectement l'activité cérébrale corrélée au changement du débit sanguin, fournissant ainsi un nouvel outil d'exploitation du fonctionnement cérébral.
- L'imagerie optique est spécifique aux études de la surface corticale, étant donné l'atténuation du signal avec la profondeur. C'est aussi une technique invasive nécessitant le retrait d'une partie du crâne. Ces techniques sont donc moins répandues que les précédentes et surtout utilisées dans les études pré-cliniques.

Dans ce travail de thèse nous allons nous concentrer sur l'IRMf qui représente une des méthodes les plus utilisées dans l'étude de l'activité cérébrale *in vivo*.

I.3 Imagerie par résonance magnétique - IRM

Cette technique d'imagerie est basée sur le principe d'interaction entre un champ magnétique externe et les moments magnétiques principalement des protons de noyaux d'hydrogène présents dans le corps humain (eau, graisse, muscle). Afin de bien comprendre le principe de l'IRM nous allons commencer par expliquer le *spin* d'une particule pour ensuite passer au phénomène de résonance à la base du signal IRM.

Le *spin*. En physique, le moment angulaire $\vec{L}$ est lié à la quantité de mouvement d'un objet tournant autour de son axe (moment angulaire propre) ou d'un axe extérieur (moment angulaire orbital). Dans le cas où l'objet contient une charge électrique son mouvement induit un champ magnétique auquel on associe un moment magnétique $\vec{M}$. L'objet est donc assimilé à un petit aimant qui, en présence

d'un champ magnétique externe, subit une force magnétique causant une déviation dans son mouvement. Les particules comme les électrons, les protons, les neutrons, etc, étant ponctuelles (élémentaires, de dimension zéro) elles ne peuvent pas tourner autour d'elles-mêmes mais possèdent néanmoins un moment angulaire quantique appelé le *spin* découvert par Stern et Gerlach en 1922. Le *spin* est donc une notion quantique n'ayant pas encore d'équivalent en mécanique classique.

Mouvement de précession. Dans le cas général (pas de champ magnétique externe), les *spins* des protons d'hydrogène sont orientés aléatoirement et leurs moments magnétiques se compensent (figure I.5-(a)). En présence d'un champ magnétique externe $\vec{B_0}$, les *spins* s'orientent dans une direction parallèle ou anti-parallèle à $\vec{B_0}$ comme le montre la figure I.5-(b) avec plus de *spins* alignés en position parallèle étant donné qu'elle nécessite moins d'énergie que l'autre. Cela crée une différence d'énergie induisant l'apparition d'un moment magnétique $\vec{M_0}$ parallèle au champ externe $\vec{B_0}$. Une fois orientés, les *spins* effectuent un mouve-

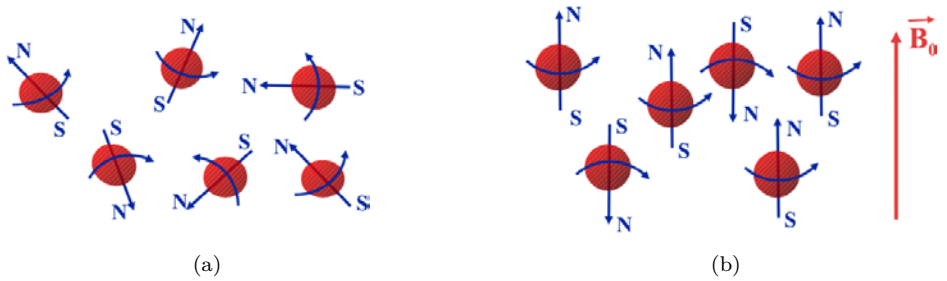


FIGURE I.5 – Illustration de l'effet d'un champ magnétique externe $\vec{B_0}$ sur l'orientation des *spins* des protons. (a) orientation aléatoire sans présence de champ externe. (b) orientation parallèle et anti-parallèle au champ externe avec plus de *spin* en parallèle.

ment de précession autour de l'axe de $\vec{B_0}$ (figure I.6-(a,b)) avec une fréquence de rotation appelée fréquence de Larmor $w_0 = 2\pi f_0 = \gamma\ B_0$ où B_0 représente l'intensité du champ magnétique externe et $\gamma = \frac{q\ g}{2\ m}$ le rapport gyromagnétique dépendant de la charge du proton q, sa masse m et le facteur de Landé g reliant le *spin* au moment magnétique. Dans le cas du proton d'hydrogène $\gamma = 42.5$ MHz/Tesla ce qui donne une fréquence de rotation de $63,75$ MHz pour une champ externe de $1,5$ Tesla. Notons que les *spins* ne tournent pas en phase autour de l'axe de $\vec{B_0}$ ce qui induit la compensation de leurs composantes transversales $\vec{M_T} = 0$ et $\vec{M_0} = \vec{M_L}$ avec $\vec{M_L}$ la composante longitudinale du moment magnétique $\vec{M_0}$

induit par les *spins* (figure I.6-(c)).

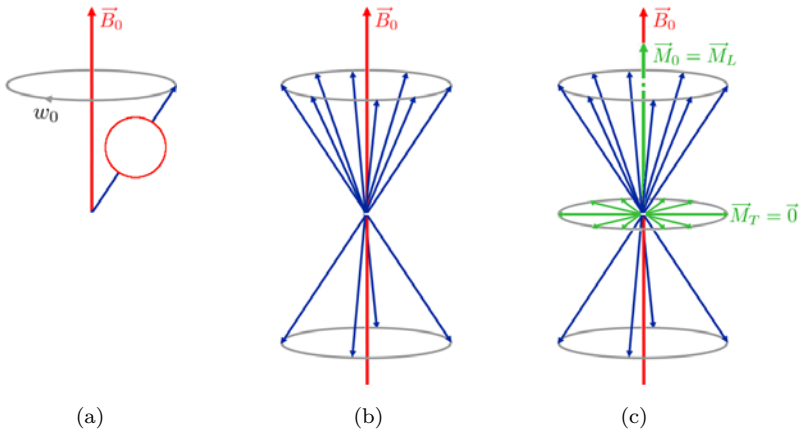


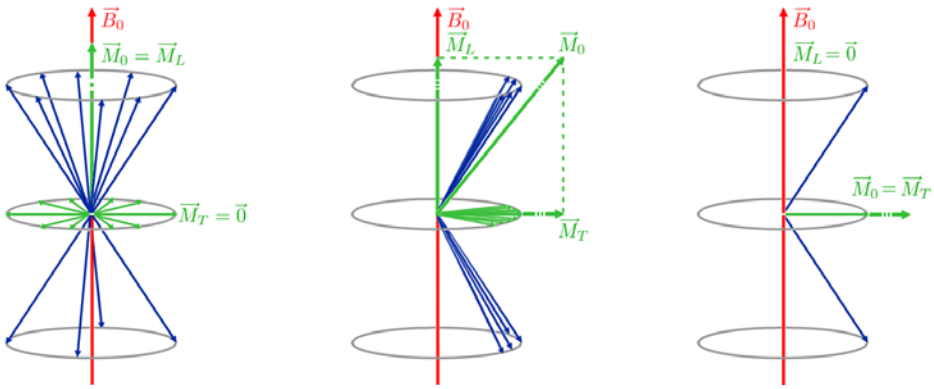
FIGURE I.6 – (a) Mouvement de précession d'un *spin* sous l'effet d'un champ magnétique externe $\vec{B_0}$, ce mouvement est caractérisé par une rotation de fréquence angulaire w_0 autour de l'axe de $\vec{B_0}$. (b) Mouvement de précession de plusieurs *spins* orientés parallèlement ou anti-parallèlement à $\vec{B_0}$. (c) Illustration des composantes longitudinale $\vec{M_L}$ et transversale $\vec{M_T}$ du moment magnétique induit par la différence entre le nombre de *spins* d'orientation parallèle et anti-parallèle. Notons que $\vec{M_T} = 0$ étant donné les rotations déphasées des *spins*.

Excitation et résonance des *spins*. Grâce à des émetteurs traversés par un courant électrique alternatif de fréquence d'oscillation équivalente à la fréquence de Larmor f_0, une onde électromagnétique et perpendiculaire au champ $\vec{B_0}$ est générée. Cela perturbe la rotation des *spins* (figure I.7) qui rentrent dans un état de résonance causant :

- Le passage d'un certain nombre de *spins* d'orientation parallèle à une orientation anti-parallèle à l'aide de l'énergie apportée par l'onde excitatrice, ce qui induit la diminution de la composante longitudinale $\vec{M_L}$ jusqu'à sa disparition (nombre égale de *spins* dans les deux orientations).
- La rotation en phase des *spins* induisant l'apparition d'une composante transversale $\vec{M_T}$.

$\vec{M_0}$ devient donc perpendiculaire à $\vec{B_0}$ ($\vec{M_0} = \vec{M_T}$).

Relaxation. Après un certain temps, l'onde excitatrice est coupée, permettant au système de revenir à son état initial dans un mouvement de relaxation émettant de l'énergie. La composante longitudinale M_L augmente pour revenir à sa valeur



(a) (b) (c)

FIGURE I.7 – (a) Etat initial (sans excitation) des *spins* dans un champ magnétique externe $\vec{B_0}$: mouvement de précession autour de $\vec{B_0}$ induisant un moment magnétique $\vec{M_0}$ de composante transversale $\vec{M_T} = \vec{0}$. (b) Etat intermédiaire de l'excitation des *spins* par une onde électromagnétique perpendiculaire à $\vec{B_0}$ et de fréquence égale à la fréquence de précession des *spins* : diminution de la composante longitudinale $\vec{M_L}$ et augmentation de la composante transversale $\vec{M_T}$ induites par la précession des *spins*. (c) Etat final de l'excitation des *spins* (résonance) correspondant à une composante longitudinale $\vec{M_L} = \vec{0}$ et à un champ total $\vec{M_0} = \vec{M_T}$ perpendiculaire à $\vec{B_0}$.

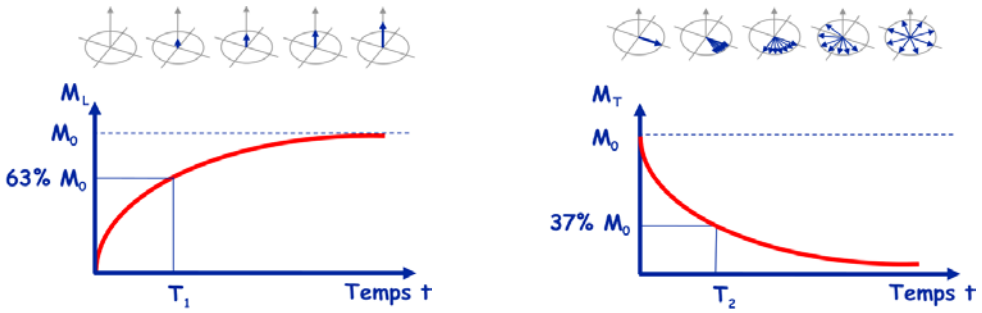
initiale suivant une loi exponentielle de temps caractéristique T_1 (figure I.8-(a))

$$M_L = M_0 \left[1 - \exp\left(-\frac{t}{T_1}\right)\right] \tag{I.1}$$

Pour $t = T_1$ on a $M_L = 0.63\ M_0$, ce qui veut dire que T_1 correspond au temps nécessaire pour que la composante longitudinale de $\vec{M_0}$ récupère 63% de sa valeur initiale. La composante transversale M_T quant à elle diminue suivant une loi exponentielle de temps caractéristique T_2 (figure I.8-(b)).

$$M_T = M_0 \exp\left(-\frac{t}{T_2}\right) \tag{I.2}$$

Pour $t = T_2$ on a $M_T = 0.37\ M_0$, ceci signifie qu'après un temps de relaxation T_2, la composante transversale du champ $\vec{M_0}$ ne possède plus que 37% de son énergie. Les temps de relaxation T_1 et T_2 diffèrent entre les tissus biologiques permettant ainsi, suite à un traitement informatique approprié, de determiner la nature de la structure analysée (substance blanche, substance grise et LCR) et d'obtenir une coupe cérébrale anatomique 2D dont chaque point est appelé pixel. La combinaison de plusieurs coupes nous permet ensuite d'obtenir le volume



(a) (b)

FIGURE I.8 – Illustration des phénomènes de relaxation des *spins* résonants après coupure de l'onde excitatrice : (a) relaxation longitudinale permettant à la composante $\vec{M}_L$ de revenir à son état initial $\vec{M}_L = \vec{M}_0$ et (b) relaxation transversale induisant l'annulation de la composante $\vec{M}_T$ pour le retour à l'état initial.

cérébral 3D composé d'un millier de voxels (pixels en 3D).

I.4 Imagerie par résonance magnétique fonctionnelle - IRMf

Développée au début des années 90, l'IRMf représente une technologie complémentaire à l'IRM précédemment abordée. Elle permet d'étudier le comportement cérébral suite à des tâches expérimentales précises (motrices, auditives, visuelles, cognitives, etc), ceci en se basant sur les propriétés magnétiques de l'hémoglobine, une molécule contenue dans les globules rouges du sang et dont la fonction principale est le transfert de l'oxygène. Quand une tâche quelconque est effectuée, un réseau de neurones, local ou distribué selon la complexité de la tâche, est suscité et nécessite de l'énergie pour accomplir son rôle. Ce qui entraîne une augmentation de la consommation d'oxygène et de glucose et donc du flux sanguin dans la région cérébrale en activité (couplage neurovasculaire). L'augmentation du flux sanguin étant plus importante que la consommation en oxygène, l'activité neuronale se traduit paradoxalement sous la forme d'une augmentation de la concentration en oxyhémoglobine dans le réseau veineux par rapport à la déoxyhémoglobine. Cette dernière, paramagnétique, cause une perturbation du champ initial. Ces inhomogénéités de champ magnétique sont corrélées à la concentration en déoxyhémoglobine et induisent une fréquence de précession de *spins* différente de f_0 et une relaxation transversale plus rapide T_2^* (figure I.9).

Ce sont ces faibles variations de signal appelé BOLD (*Blood Oxygen Level Dependent*) qui sont mesurées en IRMf. Ainsi, plus l'activité neuronale est forte, plus la concentration en déoxyhémoglobine est faible dans le réseau veineux, plus le signal BOLD mesuré est important en aval de la zone activée (figure I.10). [Ogawa 1990, Ogawa 1992, Bandettini 1993].

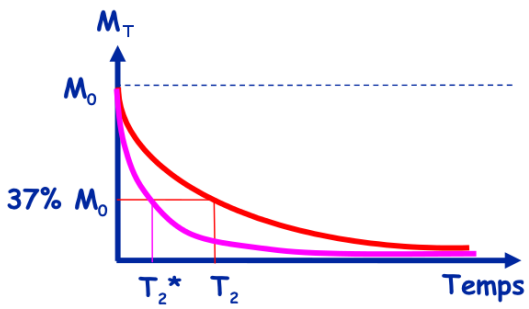


FIGURE I.9 – Illustration du phénomène de relaxation de la composante transversale $\vec{M}_T$ des *spins* résonants après coupure de l'onde excitatrice dans le cas d'une activité cérébrale causant une inhomogénéité du champ magnétique externe (en rose) et qui est plus rapide que la relaxation dans le cas d'un champ magnétique externe homogène (en rouge).

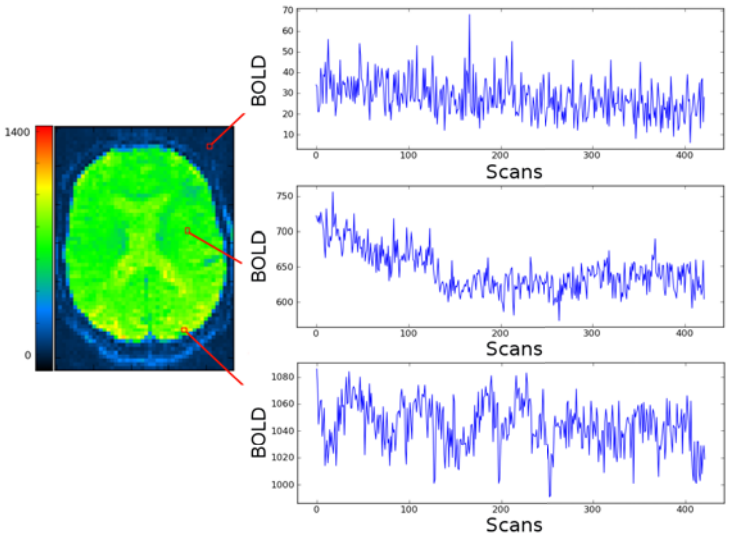


FIGURE I.10 – Exemple d'une coupe axiale de données IRMf à un instant précis et de signaux BOLD acquis à des voxels situés au fond de l'image (signal du haut) et à l'intérieur du cerveau ayant une activation moyenne (signal du milieu) et forte (signal du bas) reflétant une activité neuronale suite à une stimulation visuelle.

I.5 Paradigme expérimental et acquisition des données

L'étude de l'activité cérébrale d'un sujet nécessite, sauf si on s'intéresse à l'état de repos, l'exécution de certaines tâches appelées conditions expérimentales ou types de stimuli et pouvant être motrices, auditives, visuelles ou autre selon l'étude en question. Ces conditions sont présentées au sujet de façon répétitive afin d'obtenir des résultats statistiquement significatifs. Cela à travers ce que l'on appelle un paradigme expérimental dont il existe deux types principaux [Rosen 1998] :

Paradigme en bloc. Ce paradigme consiste en la répétition de stimulations de longue durée (plusieurs secondes) dans le but de maximiser le rapport signal sur bruit (RSB) permettant ainsi l'obtention de fortes activations. Généralement ces paradigmes ne contiennent que deux ou trois types de stimuli.

Paradigme événementiel. Contrairement au paradigme en bloc, le paradigme événementiel permet la répétition de stimuli brefs de durée inférieure au temps d'acquisition du signal. Cela permet de caractériser la forme de la réponse hémodynamique à l'inverse des paradigmes en bloc où une saturation est causée par la durée des stimulations. Dans le cas d'un paradigme événementiel lent, l'intervalle inter-stimuli (IIS) est de l'ordre d'une dizaine de secondes, permettant au système de revenir à son état de repos avant l'arrivée du prochain stimulus ce qui empêche d'avoir une interaction entre les réponses hémodynamiques des différents types de stimuli. Ces paradigmes limitent néanmoins le nombre de stimuli présentés au sujet durant l'expérience et donc la reproductibilité des résultats. Ils laissent, de plus, place à une distraction possible pouvant influencer l'état cognitif du sujet. A l'inverse, un paradigme évenementiel rapide (IIS de 2 à 4 secondes), permet de présenter plus de stimuli. Notons que les instants d'arrivées des stimuli peuvent être calés (synchrones) ou non (asynchrones) sur les temps d'acquisition du signal BOLD. De plus, il est important d'optimiser le paradigme expérimental par rapport à l'ordre et au nombre des stimuli ainsi qu'au choix de l'IIS comme expliqué dans [Joseph 1999] et qu'on ne discutera pas ici étant donné que ce n'est pas le sujet principal de cette thèse.

Une fois le paradigme mis en place, l'acquisition du volume cérébral 3D (scan)

est effectuée à travers l'acquisition consécutive de plusieurs coupes 2D constituées chacune de plusieurs pixels. La figure I.11 illustre ce principe avec un exemple de paradigme évenementiel rapide asynchrone contenant trois conditions expérimentales qu'on différencie par des couleurs (rouge, vert et bleu). Notons que la durée d'acquisition de chaque volume, ou temps de répétition TR, varie de 1 à 4 secondes dans les expériences IRMf. Après des traitements de reconstruction de données on obtient en chaque voxel du volume cérébral, un signal BOLD reflétant les activations cérébrales correspondant aux tâches expérimentales effectuées par le sujet. Plusieurs artéfacts sont néanmoins présents dans les signaux BOLD ac-

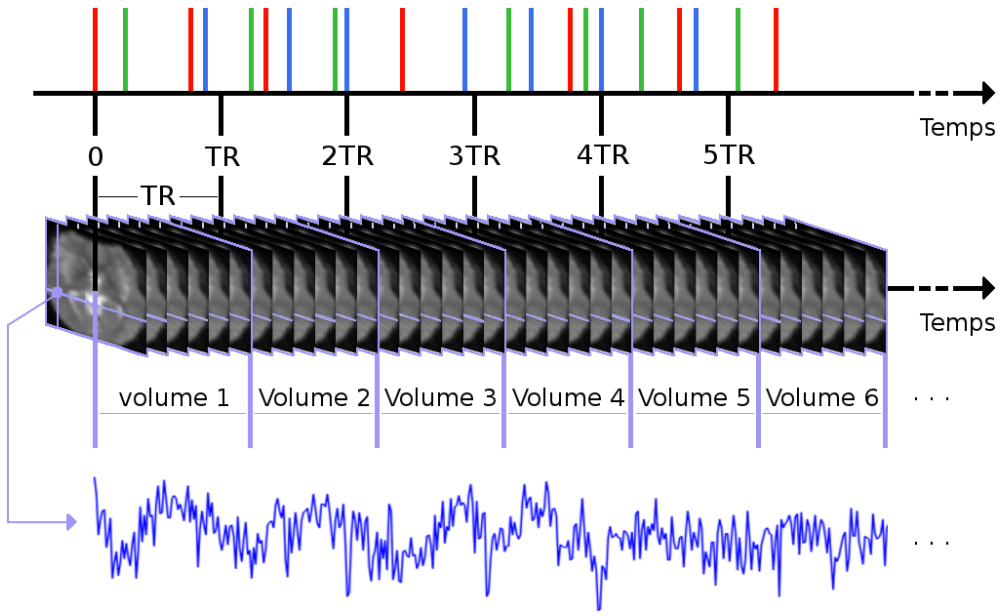


FIGURE I.11 – Illustration du principe d'acquisition des données IRMf avec un exemple de paradigme événementiel rapide asynchrone contenant 3 conditions expérimentales de couleurs différentes (rouge, vert et bleu) (en haut). Chaque volume cérébral est composé de plusieurs coupes 2D et est acquis sur une durée TR (temps de répétition) (au milieu) permettant ainsi d'obtenir, en chaque voxel du volume, un signal BOLD reflétant les changements hémodynamiques corrélés au paradigme en question (en bas).

quis et nécessitent certaines corrections à travers une séquence de pré-traitement effectuée avant le traitement statistique des données.

I.6 Pré-traitements des données IRMf

I.6.1 Correction du décalage temporel - Slice timing

Les coupes d'un volume cérébral ne sont pas acquises au même instant mais sur une certaine durée temporelle où l'acquisition peut être effectuée sous un mode séquentiel (de la première à la dernière) ou entrelacé (acquisition des coupes impaires ensuite paires). Dans les deux cas a lieu un décalage temporel d'acquisition entre les coupes du volume. La figure I.12 illustre ce décalage dans le cas d'une acquisition séquentielle. Chaque volume est représenté par une couleur différente. L'acquisition contient N volumes dont chacun possède C coupes. On s'aperçoit, que les C coupes du premier volume par exemple ne sont pas toutes acquises à l'instant t_0^0 mais plutôt aux instants $t_0^0, t_0^1, ..., t_0^C$, ce qui induit pour les voxels de la première coupe un signal temporel acquis au instants $t_0^0, t_1^0, ..., t_N^0$ et pour les voxels de la deuxième coupe un signal temporel acquis à des instants différents $t_0^1, t_1^1, ..., t_N^1$, on observe donc un décalage dans les signaux temporels acquis. Le slice timing est une technique de correction basée sur l'interpolation du signal mesuré au cours du temps, ce qui permet le recalage des signaux temporels afin de pouvoir, lors de la modélisation, considérer toutes les coupes d'un volume comme acquises au même instant.

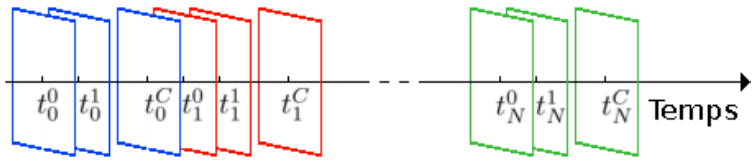


FIGURE I.12 – Illustration du décalage temporel inter-coupes durant l'acquisition séquentielle des volumes cérébraux en IRMf.

I.6.2 Correction du mouvement - Recalage

Le mouvement de la tête du sujet durant l'expérience est l'un des artéfacts non négligable et inévitable étant donné un temps d'acquisition de plusieurs minutes. La figure I.13 montre l'effet d'un brusque mouvement de la tête d'un sujet sur le signal BOLD acquis en un voxel j se situant à la frontière cérébrale. Cette baisse brusque dans l'intensité du signal (ou dans d'autre cas l'augmentation brusque) peut engendrer des faux négatifs (ou positifs dans le cas d'une augmentation) dans

les résultats de détection. Le recalage est une famille de méthodes permettant de

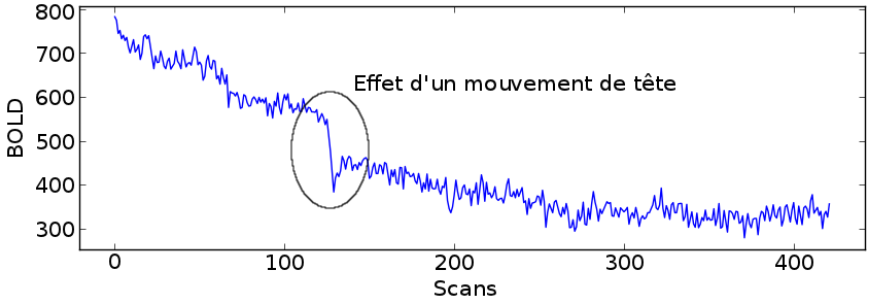


FIGURE I.13 – Effet d'un mouvement brusque de la tête sur le signal BOLD acquis en un voxel j situé à la frontière du volume cérébral (scan).

réaligner les objets entre eux. Dans le cas des séquences d'images fonctionnelles un recalage rigide permet la correction en partie des artéfacts de mouvement. Ceci à travers deux étapes consécutives :

- L'estimation des 6 paramètres de translation (figure I.14-(a)) et de rotation (figure I.14-(b)) basée sur la minimisation de la somme des différences quadratiques entre les scans fonctionnels successifs et un scan de référence (généralement le premier scan).
- L'application de la transformation estimée aux données fonctionnelles à travers une étape d'interpolation trilinéaire, sinusoïdale ou B-spline.

Des artéfacts liés aux mouvements restent cependant présents dans les données fonctionnelles. Ces artéfacts sont dus à plusieurs facteurs comme par exemple l'erreur d'interpolation, la précision de l'estimation des paramètres de mouvement, le repliement de spectre correspondant aux mouvements, etc. Ces artéfacts ne pouvant pas être corrigés par le recalage rigide peuvent engendrer de significatifs faux positifs dans les résultats de traitement, ceci à partir de 1 à 2 pourcents de changement induit dans le signal BOLD. Dans le cas des expériences où certaines conditions pourraient causer un mouvement de tête corrélé au paradigme expérimental (tâches motrices, langage), la correction de mouvement devient particulièrement importante et risque de causer la perte d'une partie du signal utile corrélé au mouvement en question. Plus de détails sur ce recalage sont disponibles dans [Ashburner 2003, Chap. 2].

Les deux pré-traitements précédents sont corrélés dans le sens où leur ordre d'application dépend du mode d'acquisition. Le choix de cet ordre est basé sur la considération qu'un grand changement d'intensité en un voxel j entre 2 coupes

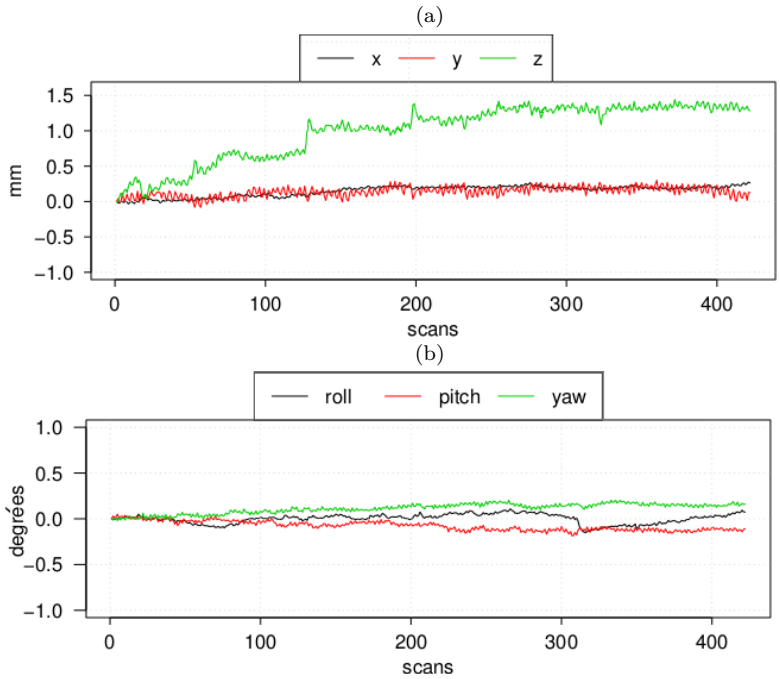


FIGURE I.14 – Paramètres de mouvements estimés par un recalage rigide, ceci concernant la translation (a) et la rotation (b) de la tête d'un sujet durant l'acquisition IRMf.

successives (acquisition séquentielle) est davantage du au mouvement qu'au décalage temporel d'acquisition, ce pourquoi il vaut mieux, dans ce cas, effectuer la correction du mouvement avant celle du slice-timing afin d'éviter que cette dernière concerne des zones cérébrales différentes. A l'inverse dans le cas d'une acquisition entrelacée, le décalage temporel inter-coupes est plus élevé et il est préférable d'effectuer la correction du slice-timing en premier pour éviter de déplacer des voxels dont les signaux sont acquis à des instants différents. Pour cela, des travaux se sont intéressés au développement d'un recalage spatio-temporel effectuant conjointement les deux types de corrections [Bannister 2004].

I.6.3 Recalage anatomo-fonctionnel

Ce recalage permet d'identifier les régions cérébrales anatomiques correspondant aux foyers d'activation détectés par le traitement, surtout qu'une visualisation conjointe des images anatomiques et fonctionnelles est impossible étant donnée leur différence de résolution $1\times1\times1\ mm^3$ pour l'anatomie et $3\times3\times3\ mm^3$ pour les fonctionnelles. Le recalage anatomo-fonctionnel est effectué entre le vo-

lume anatomique et la moyenne des volumes fonctionnels calculée à la sortie des étapes de correction spatio-temporelle précédentes. Comme les modalités sont différentes, la méthode de mesure de similarité utilisée se base généralement sur l'information mutuelle.

I.6.4 Normalisation spatiale

Dans le cadre des études de groupes et étant donné la différence morphologique cérébrale inter-sujets (figure I.15), une étape de recalage élastique (non-rigide) vers un volume de référence (référentiel ou *template*) est nécessaire afin de pouvoir comparer les résultats de traitement et de déterminer des comportements généraux à travers une population. Cette normalisation spatiale (recalage vers une norme) permet donc d'identifier des foyers d'activations à l'aide de leurs coordonnées euclidiennes dans un espace standard et commun à tous les sujets. Plusieurs référentiels ont été proposés dans la littérature dont les plus connus sont le référentiel de Talairach [Talairach 1988] et le référentiel développé par le *Montreal Neurological Institut* (MNI) [Fonov 2009]. Généralement, le recalage est

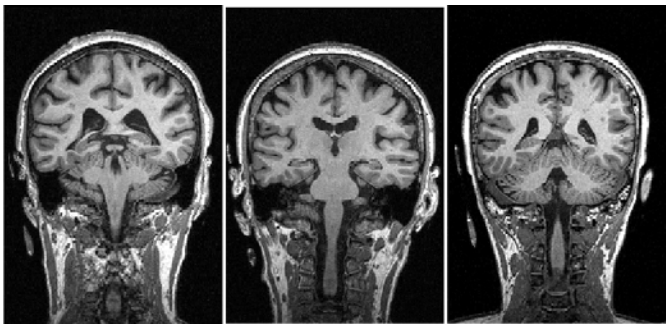


FIGURE I.15 – Illustration de la différence anatomique inter-sujets sur une coupe coronale ayant, pour les trois sujets, le même positionnement dans le volume.

effectué à travers les étapes suivantes :

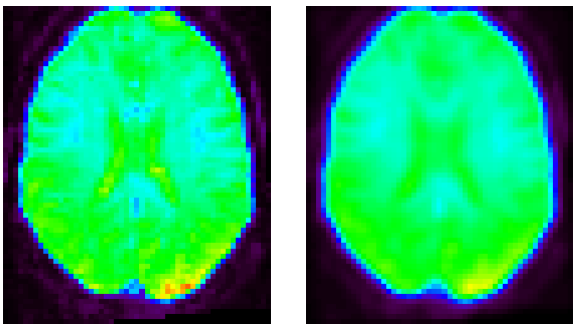
- Estimation des paramètres de recalage entre le référentiel et l'image anatomique étant donné sa meilleure résolution par rapport à l'image fonctionnelle.
- Application des paramètres estimés sur les données fonctionnelles recalées sur l'image anatomique précédente.

Une autre technique de normalisation appelée Dartel et montrant une bonne précision de recalage a été proposée dans [Ashburner 2007]. Cette méthode permet l'estimation conjointe d'un référentiel issu de la population d'étude et des

paramètres de recalage vers ce référentiel. Pour plus de détails sur les techniques de recalage, le lecteur peut se reporter à [Ashburner 2003, Chap. 3].

I.6.5 Lissage spatial

Cette étape de pré-traitement consiste en une convolution avec un noyau gaussien de largeur à mi-hauteur (*Full Width at Half Maximum* ou FWHM) généralement fixée à 1.5 ou 2 fois la taille des voxels. Ce lissage permet l'augmentation du RSB, causant néanmoins une diminution de la résolution spatiale ainsi que du contraste des données comme le montre la figure I.16. L'une des



(a) (b)

FIGURE I.16 – Illustration de l'effet du lissage spatial gaussien sur la résolution d'une coupe axiale d'un volume de données IRMf. L'exemple concerne un filtrage isotrope avec un FWHM de 5 mm par rapport à des voxels de taille $3 \times 3 \times 3$ mm^3. Une diminution du contraste et de la résolution spatiale est visible entre l'image avant (a) et après (b) lissage.

raisons principales nécessitant le lissage préalable des données est de forcer leur gaussianité pour la prise en compte de la corrélation spatiale inter-voxels lors de leur traitement statistique et plus précisément lors de la correction du problème de comparaisons multiples qu'on expliquera par la suite. Notons que ce lissage a aussi pour effet de gommer les différences inter-individus. Il existe bien sûr d'autres approches permettant la prise en compte de la corrélation spatiale des données IRMf à travers l'utilisation des champs de Markov, comme on le verra dans les chapitres suivants, et qui permet ainsi d'éviter cette étape de lissage préalable des données.

I.7 Conclusion

Ce chapitre nous a permi d'introduire les origines physiques des données IRMf auxquelles on s'intéresse dans ce travail de thèse ainsi que leurs pré-traitements nécessaires à la correction de quelques artéfacts d'acquisition limitant ainsi leurs influences sur les résultats de traitement statistique des données IRMf qui représente un grand domaine de recherche comme le montre le chapitre suivant.

Chapitre II

Analyse Statistique en IRMf

Une analyse statistique, successive aux pré-traitements des données, est nécessaire pour la détection des activités cérébrales correspondant aux différentes conditions (types de stimuli) présentes dans le paradigme expérimental. Plusieurs techniques ont été proposées parmi lesquelles figure le Modèle Linéaire Général (MLG), largement diffusé à travers le logiciel Statistical Parametric Mapping (SPM, http://www.fil.ion.ucl.ac.uk/spm/), et qui est, de nos jours, le plus utilisé pour la détection de l'activité cérébrale en IRMf. Dans ce chapitre nous allons rappeler les principes du MLG et les techniques de détection de l'activité cérébrale et de l'estimation de la FRH qui sont généralement traitées séparément, ceci en comparant les approches classiques et bayésiennes. Finalement nous introduirons l'approche de Détection-Estimation Conjointe (DEC) qui traite conjointement ces deux problèmes afin d'améliorer les résultats d'analyse. Par la suite les matrices et vecteurs sont notés en gras avec les matrices en majuscule et les vecteurs en minuscule. Un vecteur est par convention un vecteur colonne. La transposé est notée ^{t}.

II.1 Méthodes classiques de détection

II.1.1 Modèle Linéaire Général - MLG

Le MLG est un modèle convolutif reliant le signal BOLD aux instants d'occurences des conditions expérimentales à travers une fonction de réponse hémodynamique (FRH) qui reflète les variations du débit sanguin induite par une stimulation. Ce modèle est une régression linéaire multiple selon laquelle le signal BOLD

$\boldsymbol{y}_j$ acquis au voxel j est défini à travers trois types de signaux :

- Le signal utile qu'on suppose être la somme des effets induits par les différentes conditions (type de stimuli) du paradigme expérimental.
- Les dérives basses fréquences dues aux phénomènes physiques (bruit thermique, ...), physiologiques (e.g. repliement des fréquences respiratoires d'environ 1 Hz et cardiaques d'environ 0.25 Hz du à la faible fréquence d'échantillonnage $F = 1/TR \in [0.25, 1]$ Hz, et qui n'atteint pas forcément la fréquence limite de Nyquist égale à 2 Hz) et les effets induits par l'interaction entre les mouvements et le champ magnétique statique [Ashburner 2003, Chap. 10].
- Le bruit exprimant l'erreur de modélisation, que l'on suppose généralement gaussien et dépendant du voxel j.

$$\boldsymbol{y}_j = \boldsymbol{X}\boldsymbol{\beta}_j + \boldsymbol{b}_j = \sum_{r=1}^{R} \beta_j^r \boldsymbol{x}^r + \boldsymbol{b}_j \tag{II.1}$$

où $\boldsymbol{y}_j$ est un vecteur de dimension N (le nombre de scans) représentant le signal BOLD pré-traité au voxel j. Le signal utile ainsi que les dérives basses fréquences sont représentés par $\boldsymbol{X}\boldsymbol{\beta}_j$, où $\boldsymbol{X}$ est une matrice de dimension $N \times R$ appelée la matrice de dessin dont la construction est détaillée ci-dessous et $\boldsymbol{\beta}_j$ sont les coefficients de régression linéaire (vecteur de dimension R). Enfin, $\boldsymbol{b}_j$ représente le bruit.

Construction de la matrice de dessin du MLG. La matrice $\boldsymbol{X}$ contient, en plus d'un régresseur constant représentant la ligne de base, deux types de régresseurs : (i) les régresseurs stimulus-induits construits à partir du paradigme expérimental et dont le nombre est égal au nombre de conditions M et (ii) les régresseurs de non intérêt supplémentaires ($R - M - 1$ régresseurs) représentant des composantes qui covarient avec le signal BOLD et dont la prise en compte peut améliorer la modélisation, comme par exemple les paramètres de mouvement ou les régresseurs des dérives basses fréquences qui parfois sont écartés à travers un filtrage passe-haut dont la fréquence de coupure est généralement fixée à $F_0 = 1/T_0 = 1/128 = 0.007$Hz éliminant ainsi tous les signaux ayant une période supérieure à T_0. Dans ce cas, il est important de vérifier que cette valeur n'est pas trop élevée par rapport au signal stimulus-induit pour éviter de perdre des informations liées à l'activité cérébrale. Chaque régresseur stimulus-induit $\{\boldsymbol{x}^r\}_{r=1:M}$ (vecteur de dimension N) représente la convolution entre une Fonc-

tion de Réponse Hémodynamique (FRH) et un signal de stimulation. La FRH est généralement fixée à une forme canonique constante à travers le cerveau entier et obtenue comme la soustraction de deux fonctions Gamma (figure II.1) [Friston 1998]. Le signal de stimulation quant à lui est construit à partir du paradigme, représentant les temps d'occurence du $r$ième type de stimuli, à travers une discrétisation sur une grille d'échantillonnage de pas $\Delta t = \frac{\mathrm{TR}}{N_t} = \frac{TR}{16}$ avec N_t le nombre d'intervalles de temps (16 par défaut dans SPM). Son produit de convolution avec la FRH est ensuite sous-echantillonné d'un pas égal au TR afin d'obtenir le régresseur $\boldsymbol{x}^r$ en question [Ashburner 2003, Chap. 7]. Un exemple de la matrice de dessin est représenté par la figure II.2 pour $N = 150$ scans et $R = 5$ régresseurs dont 4 régresseurs stimulus-induits et 1 régresseur constant.

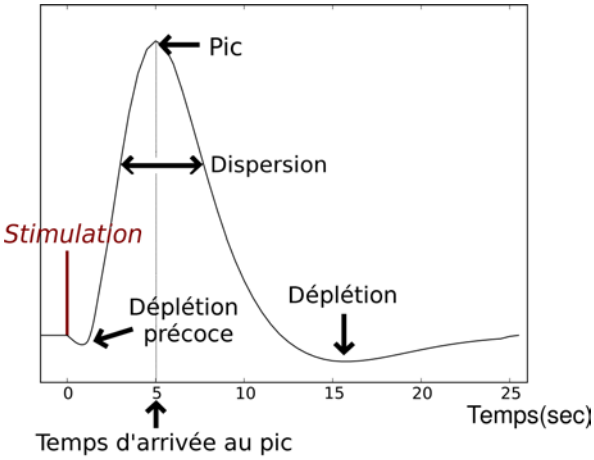


FIGURE II.1 – Forme canonique de la Fonction de Réponse Hémodynamique (FRH) définie comme la soustraction de deux fonctions Gamma et caractérisée principalement par : (1) un pic dont le temps d'arrivée est autour de 5 secondes après la stimulation, (2) une déplétion autour de 15 secondes causée par une diminution du débit sanguin plus rapide que celle du volume des capillaires (3) une déplétion précoce reflétant une autre diminution du débit sanguin due à la consommation immédiate de l'oxygène après l'excitation neuronale et avant l'apport du sang frais.

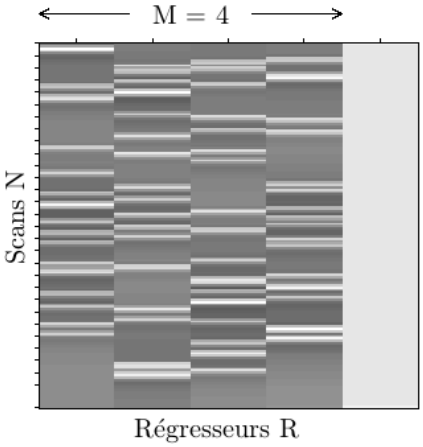


FIGURE II.2 – Matrice de dessin de dimensions $N \times R$ dans le cadre du Modèle Linéaire Général (MLG) avec $N = 150$ scans et $R = 5$ régresseurs dont $M = 4$ régresseurs stimulus-induits où chacun représente la convolution du paradigme correspondant avec la FRH canonique et 1 régresseur constant (à droite) représentant la ligne de base.

Une fois le modèle défini, il existe deux principales approches pour l'estimation des paramètres, l'approche classique s'appuyant sur le principe du Maximum de Vraisemblance (MV) et l'approche bayésienne permettant l'introduction d'informations *a priori*.

II.1.1.1 Estimation univariée

Les approches classiques consistent à estimer les paramètres du MLG selon les moindres carrés qui visent à minimiser l'erreur quadratique entre le signal observé $\boldsymbol{y}_j$ et celui estimé $\boldsymbol{X}\widehat{\boldsymbol{\beta}}_j$. Dans le cas d'un bruit blanc gaussien $\boldsymbol{b}_j \sim \mathcal{N}\left(\boldsymbol{0}, \sigma_j^2 \boldsymbol{I}_N\right)$, on a $\boldsymbol{y}_j \sim \mathcal{N}\left(\boldsymbol{X}\boldsymbol{\beta}_j, \sigma_j^2 \boldsymbol{I}_N\right)$ et l'estimation de $\boldsymbol{\beta}_j$ revient à minimiser le terme $(\boldsymbol{y}_j - \boldsymbol{X}\boldsymbol{\beta}_j)^{\mathrm{t}}(\boldsymbol{y}_j - \boldsymbol{X}\boldsymbol{\beta}_j)$ par rapport à $\boldsymbol{\beta}_j$ ce qui donne :

$$\widehat{\boldsymbol{\beta}}_j = \left(\boldsymbol{X}^{\mathrm{t}}\boldsymbol{X}\right)^{-1} \boldsymbol{X}^{\mathrm{t}}\boldsymbol{y}_j \tag{II.2}$$

$$\mathrm{Var}(\widehat{\boldsymbol{\beta}}_j) = (\boldsymbol{X}^{\mathrm{t}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{t}}\mathrm{Var}(\boldsymbol{y}_j)\boldsymbol{X}(\boldsymbol{X}^{\mathrm{t}}\boldsymbol{X})^{-1} = \sigma_j^2(\boldsymbol{X}^{\mathrm{t}}\boldsymbol{X})^{-1} \tag{II.3}$$

La contrainte étant que $\boldsymbol{X}^{\mathrm{t}}\boldsymbol{X}$ soit inversible, ce qui n'est pas toujours le cas. Si elle est de rang déficient on remplace le terme $(\boldsymbol{X}^{\mathrm{t}}\boldsymbol{X})^{-1}$ dans les équations (II.2,II.3) par la pseudo-inverse de Moore-Penrose $(\boldsymbol{X}^{\mathrm{t}}\boldsymbol{X})^{-}$ [Ashburner 2003, Chap. 7]. L'estimateur précédent de $\boldsymbol{\beta}_j$ est sans biais car :

$$\mathbb{E}[\widehat{\boldsymbol{\beta}}_j] = \left(\boldsymbol{X}^{\mathrm{t}}\boldsymbol{X}\right)^{-1} \boldsymbol{X}^{\mathrm{t}}\mathbb{E}[\boldsymbol{y}_j] = \left(\boldsymbol{X}^{\mathrm{t}}\boldsymbol{X}\right)^{-1} \boldsymbol{X}^{\mathrm{t}}\boldsymbol{X}\boldsymbol{\beta}_j = \boldsymbol{\beta}_j \tag{II.4}$$

La variance du bruit, quant à elle est estimée par :

$$\hat{\sigma}_j^2 = \frac{\left(\boldsymbol{y}_j - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_j\right)^{\mathrm{t}}\left(\boldsymbol{y}_j - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_j\right)}{N} \tag{II.5}$$

Cet estimateur est cependant biaisé car étant donné que $\mathbb{E}\left[\boldsymbol{y}_j - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_j\right] = \boldsymbol{0}$, $\mathrm{tr}\,[\boldsymbol{A}\boldsymbol{B}] = \mathrm{tr}\,[\boldsymbol{B}\boldsymbol{A}]$ et $\mathrm{tr}\,(\boldsymbol{A}+\boldsymbol{B}) = \mathrm{tr}\,(\boldsymbol{A}) + \mathrm{tr}\,(\boldsymbol{B})$ où $\mathrm{tr}\,(\boldsymbol{A})$ la trace de la matrice $\boldsymbol{A}$, on peut donc écrire :

$$\begin{aligned}
\mathbb{E}\left(\hat{\sigma}_j^2\right) &= \frac{\mathbb{E}[(\boldsymbol{y}_j - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_j)^{\mathrm{t}}(\boldsymbol{y}_j - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_j)]}{N} = \frac{\mathrm{tr}\,(\mathrm{Var}[(\boldsymbol{y}_j - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_j)])}{N} \\
&= \sigma_j^2 \frac{\mathrm{tr}\,(\boldsymbol{I}_N - \boldsymbol{X}(\boldsymbol{X}^{\mathrm{t}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{t}})}{N} = \sigma_j^2 \frac{\mathrm{tr}\,(\boldsymbol{I}_N) - \mathrm{tr}\,(\boldsymbol{X}^{\mathrm{t}}\boldsymbol{X}(\boldsymbol{X}^{\mathrm{t}}\boldsymbol{X})^{-1})}{N} \\
&= \sigma_j^2 \frac{\mathrm{tr}\,(\boldsymbol{I}_N) - \mathrm{tr}\,(\boldsymbol{I}_R)}{N} = \frac{N-R}{N}\sigma_j^2
\end{aligned} \tag{II.6}$$

L'estimateur non biaisé de la variance du bruit est donc :

$$\hat{\sigma}_j^2 = \frac{\left(\boldsymbol{y}_j - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_j\right)^{\mathrm{t}}\left(\boldsymbol{y}_j - \boldsymbol{X}\widehat{\boldsymbol{\beta}}_j\right)}{N-R} \tag{II.7}$$

II.1.1.2 Tests statistiques univariés

Une fois les paramètres du modèle estimés, des tests statistiques permettent la vérification de la significativité de cette estimation. Deux types de test sont généralement utilisés dans l'approche classique : le test de Student (test-T) et le test de Fisher (test-F).

Test-T. Ce test permet de vérifier si une combinaison linéaire univariée (un contraste scalaire) des paramètres estimés $\widehat{\boldsymbol{\beta}}_j$ est significativement différente d'une valeur d, ceci à travers les hypothèses nulle et alternative suivantes :

$$\begin{aligned} \mathcal{H}_0 &: \boldsymbol{c}^{\mathrm{t}}\boldsymbol{\beta}_j = d \\ \mathcal{H}_1 &: \boldsymbol{c}^{\mathrm{t}}\boldsymbol{\beta}_j > d \\ &\text{ou} \\ &\boldsymbol{c}^{\mathrm{t}}\boldsymbol{\beta}_j < d \end{aligned} \tag{II.8}$$

où $\boldsymbol{c}^{\mathrm{t}} = [c_r; r = 1 : R]$ avec $c_r \in \{0, 1, -1\}$ est un vecteur ligne déterminant les coefficients de la combinaison linéaire souhaitée. On peut par exemple vérifier si :

- La condition m entraîne une activité significative au voxel j, dans ce cas $c_r = 1$ pour $r = m$ et $c_r = 0$ sinon $\Rightarrow \boldsymbol{c}^{\mathrm{t}}\boldsymbol{\beta}_j = \beta_j^m$.
- La condition m_1 entraîne une activité significativement plus importante que la condition m_2, dans ce cas $c_{m_1} = 1$, $c_{m_2} = -1$ et $c_r = 0$ pour $r \neq m_1, m_2$ $\Rightarrow \boldsymbol{c}^{\mathrm{t}}\boldsymbol{\beta}_j = \beta_j^{m_1} - \beta_j^{m_2}$.
- La condition m_1 entraîne significativement une activité plus importante que la condition m_2 indépendamment de deux modalités de présentations A et B (e.g. conditions motrices présentées sous forme visuelle et auditive), dans ce cas on a deux régresseurs pour chaque condition, chacun représentant une modalité et $\boldsymbol{c}^{\mathrm{t}}\boldsymbol{\beta}_j = \beta_j^{m_1,A} + \beta_j^{m_1,B} - \beta_j^{m_2,A} - \beta_j^{m_2,B}$.

Déterminer la loi de la statistique de test nécessite la connaissance de la loi de $\widehat{\boldsymbol{\beta}}_j$. Cette dernière se déduit du fait que $\boldsymbol{y}_j \sim \mathcal{N}(\boldsymbol{X}\boldsymbol{\beta}_j, \sigma_j^2\boldsymbol{I}_N)$ et que $\widehat{\boldsymbol{\beta}}_j =$

$(\boldsymbol{X}^{\mathrm{t}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{t}}\boldsymbol{y}_j$, ce qui donne $\widehat{\boldsymbol{\beta}}_j \sim \mathcal{N}(\boldsymbol{\beta}_j, \sigma_j^2(\boldsymbol{X}^{\mathrm{t}}\boldsymbol{X})^{-1})$ et donc :

$$\boldsymbol{c}^{\mathrm{t}}\widehat{\boldsymbol{\beta}}_j \sim \mathcal{N}(\boldsymbol{c}^{\mathrm{t}}\boldsymbol{\beta}_j, \sigma_j^2\boldsymbol{c}^{\mathrm{t}}(\boldsymbol{X}^{\mathrm{t}}\boldsymbol{X})^{-1}\boldsymbol{c}) \tag{II.9}$$

Notons aussi que $\frac{\hat{\sigma}_j^2}{\sigma_j^2} \sim \frac{\chi^2_{N-R}}{N-R}$. Enfin, l'indépendance entre $\widehat{\boldsymbol{\beta}}_j$ et $\hat{\sigma}_j^2$ permet de déduire que la statistique de test suivante suit une loi de Student à $N-R$ degrés de liberté :

$$T_j = \frac{\boldsymbol{c}^{\mathrm{t}}\widehat{\boldsymbol{\beta}}_j - \boldsymbol{c}^{\mathrm{t}}\boldsymbol{\beta}_j}{\sqrt{\hat{\sigma}_j^2\boldsymbol{c}^{\mathrm{t}}(\boldsymbol{X}^{\mathrm{t}}\boldsymbol{X})^{-1}\boldsymbol{c}}} \sim \mathcal{T}_{N-R} \tag{II.10}$$

Sous l'hypothèse nulle définie précédemment (Eq.II.8), on a :

$$T_j = \frac{\boldsymbol{c}^{\mathrm{t}}\widehat{\boldsymbol{\beta}}_j - d}{\sqrt{\hat{\sigma}_j^2\boldsymbol{c}^{\mathrm{t}}(\boldsymbol{X}^{\mathrm{t}}\boldsymbol{X})^{-1}\boldsymbol{c}}} \tag{II.11}$$

Dans SPM tous les tests-T sont des tests unilatéraux avec $d = 0$. Une fois la statistique de test calculée, la significativité en chaque voxel est determinée à travers un seuillage de valeur u_α associé à un risque d'erreur de première espèce α (généralement 0.05) contrôlant le taux de faux positifs :

$$\alpha = p_{\mathcal{H}_0}(T > u_\alpha) \tag{II.12}$$

L'hypothèse nulle est donc rejetée au voxel j si la valeur de la statistique observée en ce voxel est telle que $T_j > u_\alpha$. On peut arriver à la même conclusion en calculant la p-valeur qui exprime la probabilité d'observer une valeur supérieure à T_j sous $\mathcal{H}_0$.

$$\text{p-valeur} = p_{\mathcal{H}_0}(T > |T_j|) \tag{II.13}$$

Si la p-valeur $< \alpha$ on rejette $\mathcal{H}_0$ et l'effet estimé au voxel j est considéré significatif (figure II.3).

Test-F. Ce test est basé sur la statistique de Fisher et permet d'effectuer un test d'hypothèse multiple où $\boldsymbol{c}^{\mathrm{t}}$ devient une matrice de dimension $P \times R$ avec P le nombre de combinaisons linéaires qu'on souhaite tester. Supposons par exemple qu'on veuille tester si au moins une des conditions, parmi les M conditions expérimentales présentes dans le paradigme, induit une activité significative au voxel

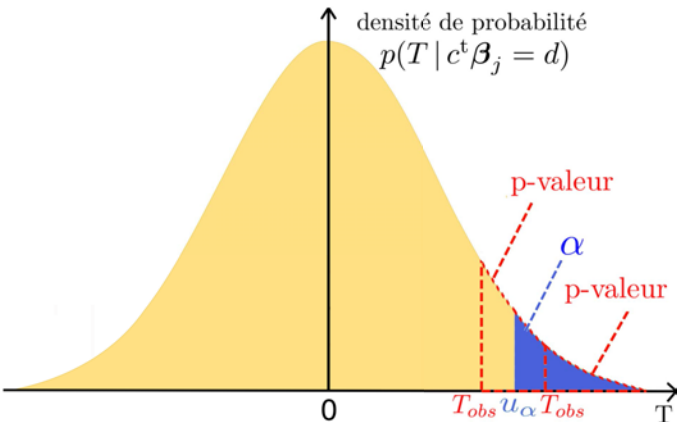


FIGURE II.3 – Seuillage de la distribution de Student dans un test statistique. α représente l'erreur de première espèce qui définit la région de rejet (en bleu) de l'hypothèse nulle $\mathcal{H}_0$. u_α est la valeur de la statistique associée à α et T_{obs} est la valeur de la statistique observée à laquelle on associe une p-valeur (surface entourée en rouge). Les assertions $T_{obs} > u_\alpha$, p-valeur $< \alpha$ et $\mathcal{H}_0$ est rejetée sont équivalentes.

j, on pourrait dans ce cas effectuer un test-F où $\boldsymbol{c}^{\mathrm{t}}$ est une matrice de dimension $M \times R$ avec $c_{m,m'} = \delta(m - m')$ où δ est la fonction Dirac qui vaut 1 en 0 (c'est à dire quand $m = m'$), ce qui éviterait d'effectuer M tests-T (un pour chaque condition). Le problème étant que dans le cas d'un test-F si $\mathcal{H}_0$ est rejetée on sait qu'au moins un des régresseurs exprime une activité significative mais on ne sait pas lequel ! Ce test peut également être utilisé pour la comparaison des modèles réduit (avec une sous-famille des M conditions) et complet (avec la totalité des M conditions) afin de vérifier lequel s'ajuste mieux aux données. Dans ce cas, la statistique de test est définie par :

$$\mathcal{F} = \frac{\frac{S_{\text{réduit}} - S_{\text{complet}}}{R - R'}}{\frac{S_{\text{complet}}}{N - R}} \sim \mathcal{F}_{R-R',N-R} \tag{II.14}$$

où S_{complet} (resp. $S_{\text{réduit}}$) représente la somme des erreurs quadratiques pour le modèle complet (resp. réduit) et R (resp. R') représente le rang de la matrice de dessin du modèle complet (resp. réduit). Pour plus de détails sur ce type de test le lecteur peut se reporter à [Ashburner 2003, Chap. 7,8].

Problème de comparaisons multiples. C'est le problème induit par l'accumulation des risques d'erreur voxelliques sur le cerveau entier et à travers lequel on peut s'attendre à observer un nombre élevé de faux positifs de l'ordre de $J\alpha$ où J est le nombre de voxels du cerveau ($J \approx 50000$). Il existe différents types de corrections parmi lesquelles la correction de Bonferroni qui permet de contrôler la FWER (*Family Wise Error rate*), c'est à dire le risque d'observer un

faux positif dans tout le volume de recherche, parmi les J voxels. Dans cette approche le risque d'erreur voxellique est fixé à α/J en supposant les tests indépendants ce qui n'est pas vraiment le cas étant donnée la corrélation spatiale des données en partie induite par le lissage spatial effectué en pré-traitement. Cette approche est très conservatrice. On lui préfère une approche plus sophistiquée qui utilise la théorie des champs aléatoires gaussiens GRF (Gaussian Random Field) [Ashburner 2003, Chap. 14] qui permet d'effectuer une correction du risque d'erreur global et donc du seuillage des cartes statistiques en se basant sur leur degré de régularité spatiale. D'autres approches ont aussi été développées comme la procédure contrôlant le *False Discovery Rate* (FDR) où la correction du seuillage des p-valeurs observées devient adaptative. Autrement dit, le seuillage commence au niveau de Bonferroni pour les p-valeurs les plus petites (les voxels les plus significatifs) et est ensuite multiplié par 2 au fur et à mesure que les p-valeurs augmentent [Benjamini 1995]. Une extension de l'approche FDR tenant compte de la dépendance entre les tests d'hypothèse voxelliques à été proposée dans [Benjamini 2001].

II.1.2 Extensions multivariées du MLG

Le caractère univarié de l'inférence classique suppose une indépendance spatiale entre les voxels lors de l'estimation (section II.1.1.1) ainsi que lors des tests statistiques (section II.1.1.2). Le signal BOLD résulte cependant du changement d'oxygénation sanguine dans les capillaires se situant à proximité du réseau neuronal activé et dont la géométrie spatiale induit une corrélation spatiale inter-voxels dans le signal BOLD acquis [Turner 2002]. Des approches d'inférence multivariées ont donc été développées, afin de prendre en compte cette corrélation spatiale et d'améliorer les résultats de traitement, selon deux axes différents :

- Au niveau des tests statistiques où l'information spatiale est prise en compte à travers l'application de la théorie GRF [Ashburner 2003, Chap. 14], comme expliqué précédemment. Une autre alternative est d'utiliser les champs de Markov lors des tests statistiques et d'éviter le lissage préalable des données [Woolrich 2005]. De plus ceci permettrait une approche plus raffinée par rapport au lissage préalable des données qui pourrait masquer les variations locales de la dépendance spatiale dont la taille est inférieure à celle du noyau gaussien utilisé.

– Au niveau de l'estimation des paramètres du modèle où la prise en compte de l'information spatiale est effectuée dans le cadre bayésien à travers l'introduction d'un *a priori* de champ de Markov (section II.3.1).

II.1.3 Modélisation de la corrélation temporelle

Jusque là le bruit modélisé est un bruit blanc gaussien $\boldsymbol{b}_j \sim \mathcal{N}\left(\mathbf{0}, \sigma_j^2 \boldsymbol{I}_N\right)$ (section II.1.1.1). Deux raisons principales renforcent cependant la présence d'une corrélation temporelle. La première concerne les mouvements de la tête lors de l'acquisition des données et les dérives basses fréquences et la deuxième la partie mal modélisée du signal stimulus-induit, due par exemple à la mauvaise modélisation de la FRH, ce signal étant corrélé temporellement à ceux de la physiologie sous-jacente [Ashburner 2003, Chap. 10]. Comme expliqué précédemment, les dérives peuvent être éliminées par un filtrage passe haut ou estimées comme on le verra par la suite. Il reste cependant la deuxième cause ainsi que les erreurs dues aux grands mouvements de tête ou aux mouvements abrupts et qui ne sont pas éliminés par la procédure de réalignement effectuée en pré-traitement. Pour cela, certains travaux proposent de prendre en compte la corrélation temporelle du bruit à travers une procédure de pré-blanchiement (*pre-whitening*) [Worsley 2002] où la structure d'auto-covariance du bruit $\boldsymbol{K}$ est estimée par une procédure de MV sous contrainte en supposant un modèle de bruit autorégressif. Cette structure est ensuite introduite au modèle linéaire $\boldsymbol{K}\boldsymbol{y}_j = \boldsymbol{K}\boldsymbol{X}\boldsymbol{\beta}_j + \boldsymbol{K}\boldsymbol{b}_j$ permettant de considérer $\boldsymbol{K}\boldsymbol{b}_j$ comme un bruit blanc. D'autre méthodes modélisent le bruit au voxel j et au scan n comme une structure autorégressive d'ordre P, $\sum_{p=1}^{P} a_{j,p} b_{j,n-p}$, de paramètres inconnus $\{a_{j,p}\}_{p=1:P}$, combinée à un bruit blanc $\epsilon_{j,n} \sim \mathcal{N}(0, \sigma_j^2)$ [Purdon 1998]. L'estimation peut être effectuée dans le cadre du MV sous contrainte [Friston 2002] ou de l'inférence bayésienne [Woolrich 2002, Penny 2003]. Finalement, une meilleure modélisation du signal stimulus-induit tenant compte de la variabilité de la FRH permet aussi de limiter les erreurs de modélisation et donc d'éliminer une partie de l'auto-corrélation du bruit.

II.2 Modélisation de la FRH

La modélisation de cette fonction est cruciale du fait de sa contribution à l'ajustement du signal stimulus-induit au signal BOLD. Comme indiqué précé-

demment, les premiers travaux concernant la détection de l'activité cérébrale se sont basés sur la forme canonique de cette réponse (figure II.1). Depuis, différents travaux ont montré que cette réponse varie entre les régions cérébrales et entre les sujets [Aguirre 1998, Buckner 1998, Glover 1999, Miezin 2000, Handwerker 2004] et ont donc incité le développement de plusieurs approches visant à améliorer la modélisation de cette FRH en s'orientant vers son estimation explicite. On trouve alors des modèles physiologiques (modèle du ballon, modèle de Windkessel) et d'autres non-physiologiques paramétriques, semi-paramétriques (dérives de la FRH, bases de fonctions) ou non-paramétriques estimant explicitement une FRH voxellique (Réponse Impulsionnelle Finie - RIF et RIF régularisée temporellement) ou parcellique (Détection-Estimation Conjointe - DEC).

II.2.1 Modèles physiologiques

A l'inverse des modèles linéaires, ces modèles permettent (i) de prendre en compte la non-linéarité du signal BOLD ainsi que (ii) de modéliser finement les mécanismes physiologiques impliqués dans la génération du signal BOLD et donc sa relation avec le réseau neuronal sous-jacent. Parmi ces modèles figurent le modèle du ballon [Buxton 1997, Buxton 1998] et le modèle de Windkessel [Mandeville 1999] qui s'intéressent, à travers des équations différentielles non-linéaires, à l'étude de différents facteurs comme le volume capillaire, le flux sanguin et la concentration en déoxyhémoglobine. Ces modèles permettent une modélisation plus réaliste mais souffrent cependant de certains problèmes d'identifiabilité ainsi que des problèmes liés au coût de calcul et à la compréhension incomplète des relations entre tous les paramètres physiologiques. Ils restent donc un axe de recherche très important étant donné l'information qu'ils apportent par rapport aux modèles non-physiologiques plus répandus pour l'analyse en IRMf.

II.2.2 Modèles non-physiologiques

II.2.2.1 Modèles paramétriques

Ces modèles sont à la base de l'estimation de la forme canonique de la FRH (figure II.1). Dans un modèle paramétrique, la FRH est modélisée à travers une fonction analytique dépendant d'un certain nombre de paramètres. Plusieurs fonctions ont été proposées dans la litérature parmi lesquelles la fonction Poisson [Friston 1994], la fonction gamma [Boynton 1996, Cohen 1997, Friston 1998] dont

l'estimation des paramètres rend le modèle non-linéaire [de Pasquale 2008] et la fonction gaussienne [Rajapakse 1998]. Afin de capturer plus de variabilité dans la forme de cette FRH, des modèles semi-paramétriques ont été développés.

II.2.2.2 Modèles semi-paramétriques

Ces modèles permettent, dans le cadre du MLG, la modélisation des variations de la FRH d'un voxel à l'autre, ceci en incluant des régresseurs supplémentaires dans la matrice de dessin.

Dérives de la FRH. Cette approche consiste à ajouter au modèle, la dérivée première (temporelle) de la FRH ($\boldsymbol{h}'_t = dh(t,s)/dt$) autorisant ainsi un décalage temporel de son pic et sa dérivée par rapport au paramètre de dispersion s ($\boldsymbol{h}'_s = dh(t,s)/ds$) permettant la prise en compte des variations de la largeur du pic (figure II.5-(a)). Ceci en ajoutant deux régresseurs par conditions à la matrice de dessin (figure II.6-(b)) permettant de représenter la convolution du paradigme avec chacune des dérivées respectivement. Dans ce cas, le signal BOLD au voxel j est exprimé par l'équation :

$$\boldsymbol{y}_j = \sum_{m=1}^{M} \left[(\boldsymbol{h} \star \boldsymbol{x}^m)\boldsymbol{\beta}^m_{1,j} + (\boldsymbol{h}'_t \star \boldsymbol{x}^m)\boldsymbol{\beta}^m_{2,j} + (\boldsymbol{h}'_s \star \boldsymbol{x}^m)\boldsymbol{\beta}^m_{3,j} \right] + \boldsymbol{b}_j \tag{II.15}$$

où $\boldsymbol{x}^m$ représente ici une séquence binaire reflétant les instants d'occurence de la $m$ième condition expérimentale. Le nombre de régresseurs stimulus-induits devient donc $3 \times M$ et le degré de liberté de la loi de Student de la statistique de test diminue de $2 \times M$. Cependant, cette approche n'autorise pas suffisamment de variations dans la forme de la FRH. Cela est dû au fait qu'elle s'interprète comme un développement de Taylor au voisinage d'un instant donné et donc n'est valide que pour prendre en compte les petites fluctuations. Un grand décalage du temps d'arrivée du pic par exemple, n'est pas pris en compte dans l'ensemble des formes possibles obtenues par les combinaisons de la FRH canonique $\boldsymbol{h}$ et sa dérivée temporelle $\boldsymbol{h}'$ ($\beta^m_{j,1}\boldsymbol{h} + \beta^m_{j,2}\boldsymbol{h}'$) pour lesquelles ce temps varie à peu près entre 3 et 6 secondes, ce qui est équivalent à un décalage de l'ordre de 1 seconde par rapport au pic canonique (figure II.4).

Bases de fonctions. L'approche précédente peut être généralisée en utilisant une base de fonction $\{\phi_c\}_{c=1:C}$ afin d'autoriser plus de variations que celles prises en compte par les dérivées de la FRH. Le signal BOLD au voxel j est dans ce cas

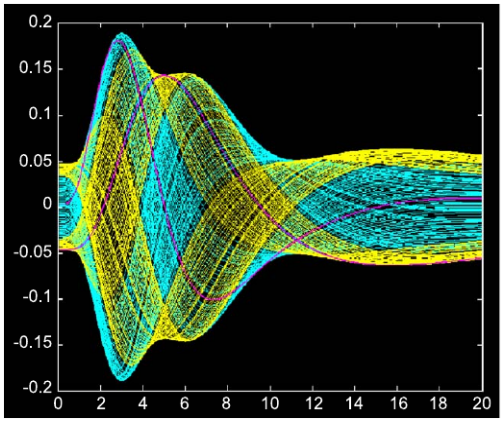


FIGURE II.4 – Ensemble des formes de FRH autorisées par la combinaison $\beta_{j,1}^m \boldsymbol{h} + \beta_{j,2}^m \boldsymbol{h}'$; Où $\boldsymbol{h}$ et $\boldsymbol{h}'$ sont respectivement la FRH canonique et sa dérivée temporelle (en rose). Les courbes en jaune représentent les cas où $\beta_{j,2}^m < \beta_{j,1}^m$ et celles en bleu les autres. A noter qu'une grande partie des courbes en bleu est bimodale ce qui complique la distinction entre *activation* et *déactivation* (activation négative) (Calhoun 2004).

exprimé par l'équation suivante :

$$\boldsymbol{y}_j = \sum_{m=1}^{M} \sum_{c=1}^{C} \left[(\boldsymbol{\phi}_c \star \boldsymbol{x}^m) \boldsymbol{\beta}_{c,j}^m \right] + \boldsymbol{b}_j \tag{II.16}$$

où $\boldsymbol{x}^m$ est une séquence binaire reflétant les instants d'occurence de la $m$ième condition expérimentale et $\boldsymbol{\beta}_{c,j}^m$ le paramètre de régression correspondant au voxel j, à la condition m et à la composante c de la base de fonction. Parmi ces bases on trouve la base de fonctions gamma (figure II.5-(b)), la base de Fourier (figure II.5-(c)) qui contient $N_b = 2N_s + 1$ fonctions : N_s fonction de sinus et N_s fonction de cosinus de periodes T_{FRH}, $T_{\text{FRH}}/2$, ..., T_{FRH}/N_s et une fonction représentant leur moyenne, où T_{FRH} représente la durée en seconde de la FRH. Parmi ces deux bases, celle utilisant des fonctions gamma est considérée comme la plus parcimonieuse étant donné qu'elle contraint davantage les formes générées limitant ainsi celles non plausibles physiologiquement. Il existe aussi des techniques semi-paramétriques non-linéaires utilisant des bases de fonction inverse logit [Lindquist 2009].

II.2.2.3 Modèles non-paramétriques

Appelés aussi modèles à Réponse Impulsionnelle Finie (RIF), à l'inverse des modèles paramétriques et semi-paramétriques, ceux-ci n'imposent aucune contrainte sur la forme de la FRH. La base RIF (figure II.5-(d)) est la base de fonctions la plus flexible contenant N_k fonctions créneau de largeur T_{FRH}/N_k secondes chacune. Un cas particulier serait l'utilisation de fonctions de Dirac (créneau de largeur nulle), ce qui reviendrait à modéliser la FRH par un vecteur de coefficients $\boldsymbol{h}_j = \{h_{j,d\Delta t}\}_{d=0:D}$ de dimension $D+1$. Le modèle devient donc pour un instant

n :

$$\begin{aligned} y_{j,n} &= \sum_{m=1}^{M} x_n^m * \boldsymbol{h}_j + \beta_{j,0} + b_{j,n} \\ &= \sum_{m=1}^{M} \sum_{d=0}^{D} x_{n-d}^m h_{j,d} + \beta_{j,0} + b_{j,n} \end{aligned} \tag{II.17}$$

où les effets $\{\beta_j^m\}_{m=1:M}$ stimulus-induits sont inclus dans le vecteur $\boldsymbol{h}_j$ et $\beta_{j,0}$ représente la ligne de base. L'inférence peut être faite au sens du MV ou dans le cadre d'une inférence bayésienne permettant d'introduire une contrainte de régularité temporelle *a priori* sur $\boldsymbol{h}$ (section II.3.2).

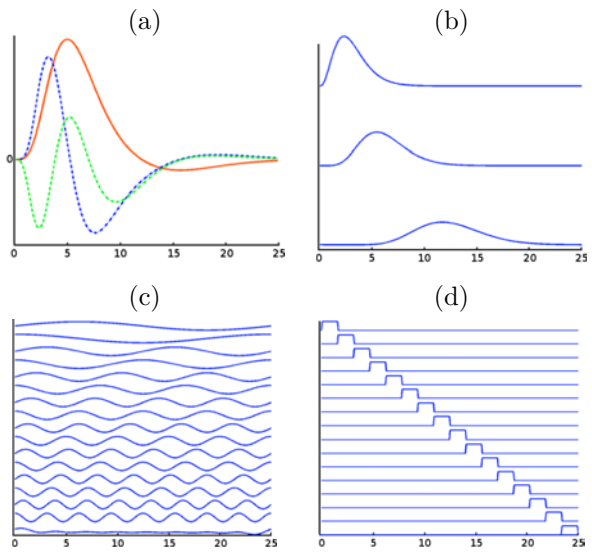


FIGURE II.5 – Familles des fonctions de base utilisées pour prendre en compte la variabilité de la FRH dans les approches semi-paramétriques : (a) FRH canonique (en rouge) et ses dérivées première (en bleu) et seconde (en vert). (b) base de fonctions gamma, (c) base de Fourier et les approches non-paramétriques (d) base de fonctions créneau (RIF).

II.3 Extensions bayésiennes du MLG

Par rapport à l'inférence classique, le cadre bayésien permet l'introduction d'informations *a priori* nécessaires à l'amélioration du traitement des données

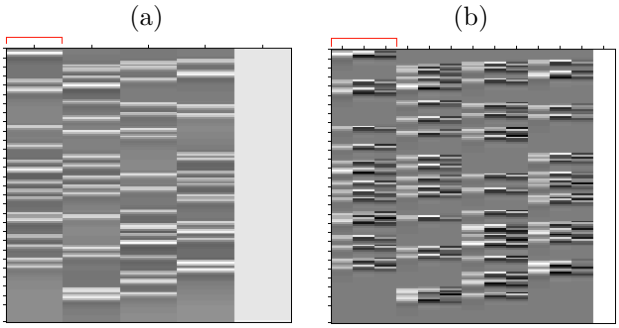


FIGURE II.6 – Matrices de dessin dans le cadre du Modèle Linéaire Général (MLG) avec $M = 4$ conditions et $N = 150$ scans. Chaque matrice contient un régresseur constant (à droite) modélisant la ligne de base. La partie indiquée en rouge représente les régresseurs modélisant la condition $m = 1$: (a) Cas standard où 1 régresseur stimulus-induit représente la convolution du paradigme correspondant avec la FRH canonique. (b) Ajout de deux régresseurs stimulus-induits afin d'inclure la dérivée première et seconde de la FRH. De même pour les 3 autres conditions.

IRMf, telles que la prise en compte de la corrélation spatiale inter-voxels (section II.3.1) ou la régularité temporelle de la FRH estimée (section II.3.2). Ces approches déduisent donc de la combinaison de ces informations *a priori* avec la vraisemblance, une distribution *a posteriori* de tous les paramètres et variables du modèle conditionnellement aux données, à partir de laquelle les estimateurs de ces derniers sont explicités. Elles nécessitent donc la définition de la vraisemblance ainsi que des lois *a priori* des variables et paramètres du modèles. L'inférence est effectuée au sens du maximum *a posteriori* (MAP) ou de la moyenne *a posteriori*, où plusieurs techniques permettent le calcul de ces estimateurs comme les algorithmes de simulation stochastique (MCMC : Méthodes de Monte Carlo par Chaînes de Markov) ou les techniques d'approximations deterministes (EM : Expectation-Maximisation ou ses variantes variationnelles) qu'on verra en détail dans le chapitre III.

De plus ces approches permettent d'éviter plusieurs problèmes induits par l'inférence statistique utilisée dans l'approche classique [Ashburner 2003, Chap. 10] et dans laquelle la p-valeur attribuée à un effet particulier ne reflète pas la probabilité que cet effet soit présent mais simplement la probabilité d'observer les données quand l'effet est absent (ou égale à une valeur précise d). Si cette p-valeur est suffisamment petite, l'hypothèse nulle est rejetée et l'effet est considéré comme significatif. Plusieurs défauts ont été reprochés à cette approche que l'inférence bayésienne évite en utilisant les cartes de probabilités *a posteriori* (PPMs, section II.3.3). Les deux premiers défauts viennent de la définition de l'hypothèse nulle

$\mathcal{H}_0 : \boldsymbol{c}^{\mathrm{t}}\boldsymbol{\beta}_j = d$ et du fait que la probabilité qu'un effet soit exactement égal à une valeur précise est elle-même nulle, ce qui induit :

- Qu'on ne peut jamais rejeter l'hypothèse alternative et constater qu'une activation n'a pas lieu car la probabilité qu'un effet soit exactement égal à une valeur $d = 0$ est nulle. Ceci est problématique, par exemple dans le cas où l'on veut établir une ségrégation fonctionnelle, on ne peut jamais dire « une région répond à la couleur mais pas au mouvement » ou l'inverse.
- Un léger écart à cette hypothèse peut être considéré comme significatif avec suffisamment de données (scans ou sujets), ce qui augmente le degré de liberté et rend la variabilité des activations estimées suffisamment petite.

Le troisième défaut quant à lui est spécifique à SPM et concerne la correction GRF du problème de comparaisons multiples. Cette correction étant basée sur le degré de lissage spatial préalable des données, elle fournit un seuil qui augmente avec la taille du noyau gaussien du lissage préalable, rendant ainsi l'inférence peu stable. Il est cependant peu réaliste que la probabilité d'activation d'un voxel dépende de cette procédure de lissage.

II.3.1 Estimation multivariée

Comme l'explique la section II.1.2, une façon de tenir compte de la corrélation spatiale des donnée consiste en l'introduction d'un *a priori* de champ de Markov. Une possibilité est d'utiliser un champ de Markov gaussien sur les paramètres $\boldsymbol{\beta}$ [Penny 2005, Oikonomou 2012] :

$$p(\boldsymbol{\beta}) = \prod_{r=1}^{R} \frac{1}{Z(\boldsymbol{\xi})} \exp(-U(\boldsymbol{\beta}^r)) = \prod_{r=1}^{R} \frac{1}{Z(\boldsymbol{\xi})} \exp\left(\frac{-\sum_j \xi_j \sum_{k \in N_j} ||\beta_j^r - \beta_k^r||^2}{2} \right) \tag{II.18}$$

où Z est la fonction de partition dépendant uniquement des paramètres de régularisation spatiale $\boldsymbol{\xi} = \{\xi_j; j = 1 : J\}$. Cette définition permet, selon la valeur de ξ_j, d'attribuer une grande probabilité *a priori* aux configurations où les paramètres sont proches entre les voxels voisins. Ce paramètre peut être supposé voxel dépendant [Oikonomou 2012] ou factorisé sur tous les voxels en étant régresseur dépendant [Penny 2005]. Une autre possibilité est de modéliser un champ de Markov discret sur des variables latentes [Smith 2007, Vincent 2010a, Vincent 2010b, Vincent 2007] comme nous le verrons dans la section II.4. Pour plus de détails sur d'autres techniques d'estimation multivariées qui ne seront pas détaillées ici

vu qu'elles ne représentent pas le sujet principal de cette thèse, le lecteur peut se reporter à [Vincent 2010a] .

II.3.2 RIF régularisée temporellement

Par rapport aux modèles non-paramétriques cités dans la section II.2.2.3, les coefficients de la FRH ne sont plus considérés comme indépendants mais caractérisés par une structure d'auto-corrélation temporelle permettant d'introduire un effet de lissage sur les formes obtenues. On est donc dans un cadre d'une RIF régularisée temporellement [Goutte 2000, Marrelec 2001, Marrelec 2003]. Ceci peut se faire en introduisant un *a priori* sur $\boldsymbol{h}_j = \{h_{j,d\Delta t}\}_{d=0:D}$ de l'équation (II.17) sous la forme :

$$p(\boldsymbol{h}_j \mid v_h) \propto \exp\left(-\frac{1}{2v_h}\boldsymbol{h}_j^{\mathrm{t}}\boldsymbol{R}^{-1}\boldsymbol{h}_j\right) \tag{II.19}$$

où v_h est la variance inconnue de la FRH et $\boldsymbol{R}$ sa matrice de covariance (voir détails en annexe A) exprimant une contrainte sur la dérivée seconde afin de pénaliser les fortes variations de pentes. En plus de cette régularité temporelle, [Ciuciu 2003] rajoute les contributions suivantes :

- Ajout d'une contrainte de bord fixant le premier et le dernier coefficient de la FRH à zéro ($h_0 = h_{D\Delta t} = 0$), modélisant ainsi le fait que cette dernière représente un processus causal de durée finie. La FRH débute donc après le stimulus et revient à la ligne de base après une certaine durée.
- Extension au cas des paradigmes multi-conditions en modélisant une FRH par conditions $\boldsymbol{h}^m$ ainsi qu'aux paradigmes asynchrones où les occurences des stimuli ne coïncident plus avec les temps d'acquisition.
- Modélisation des dérives basses fréquences $\boldsymbol{P}\boldsymbol{\ell}_j$ en chaque voxel.

Le modèle devient donc en un point temporel donné n :

$$y_{j,n} = \sum_{m=1}^{M}\sum_{d=1}^{D} x_{n-d}^m h_{j,d}^m + \boldsymbol{P}\boldsymbol{\ell}_j + b_{j,n} \tag{II.20}$$

où $\boldsymbol{P}$ est une matrice de dimension $N \times O$ dont les colonnes représentent une base de fonctions orthogonales basses fréquences et $\boldsymbol{\ell}_j$ les coefficients correspondants. Plusieurs modèles de bruit $\boldsymbol{b}_j = \{b_{j,n}\}_{n=1:N}$ ont été étudiés (blanc gaussien, AR(1), etc.) ainsi que leurs impacts sur la complexité des calculs. Pour plus de détails, le lecteur peut se reporter à l'étude comparative réalisée par [Casanova 2008,

Casanova 2009].

Notons qu'une bonne estimation de la FRH nécessite sa reproductibilité à travers les occurences des stimuli. La considération d'une FRH par voxel et par condition incite donc à l'augmentation de la durée de l'expérience afin d'augmenter le nombre de données disponibles par rapport au nombre de paramètres à estimer afin d'améliorer la robustesse de l'estimation. De plus, les travaux de [Ciuciu 2003, Ciuciu 2004] ont montré la proximité, à un facteur d'échelle près, entre les formes de FRH estimées dans des voxels voisins et pour différents types de stimuli.

II.3.3 Cartes de probabilités *a posteriori* (PPMs)

Les cartes de probabilités *a posteriori* (PPMs) représentent une alternative complémentaire aux cartes statistiques (cartes T) calculées dans l'inférence classique (section II.1.1.2). Elles expriment la probabilité sachant les données qu'un voxel soit activé ou que son activation (exprimée par les effets $\boldsymbol{\beta}_j$) soit supérieure à un seuil donné. Leur intérêt principal en neuroimagerie, par rapport aux tests statistiques (test-T) effectués selon l'approche classique, réside dans l'absence du problème de comparaisons multiples étant donné la disparition de la notion de test d'hypothèse. La probabilité qu'un voxel soit activé étant la même indépendamment du fait qu'on ait analysé uniquement ce voxel ou le cerveau entier [Ashburner 2003, Chap. 17], cette PPM est définie par :

$$PPM_j^{\gamma} = p(\boldsymbol{c}^{\mathrm{t}}\boldsymbol{\beta}_j > \gamma \,|\, \boldsymbol{y}_j) \tag{II.21}$$

Rappelons que le contraste $\boldsymbol{c}^{\mathrm{t}}\boldsymbol{\beta}_j$ représente une combinaison linéaire des R régresseurs du vecteur d'effets $\boldsymbol{\beta}_j$. Ces cartes de probabilité ($PPM^{\gamma} = \{PPM_j^{\gamma}; j = 1 : J\}$) peuvent ensuite être seuillées à $\alpha = 95\%$ par exemple gardant ainsi uniquement les voxels dont la probabilité d'activation *a posteriori* est supérieure à 95%.

La figure II.7 représente la différence entre les PPMs et le test-T.

Pour conclure, nous précisons qu'il n'est pas nécessaire de chercher à relier les PPMs et les tests-T car ils répondent à des questions différentes bien qu'il existe entre eux un lien dans des cas particuliers du MLG (bruit blanc gaussien et *a priori* bayésien non informatif).

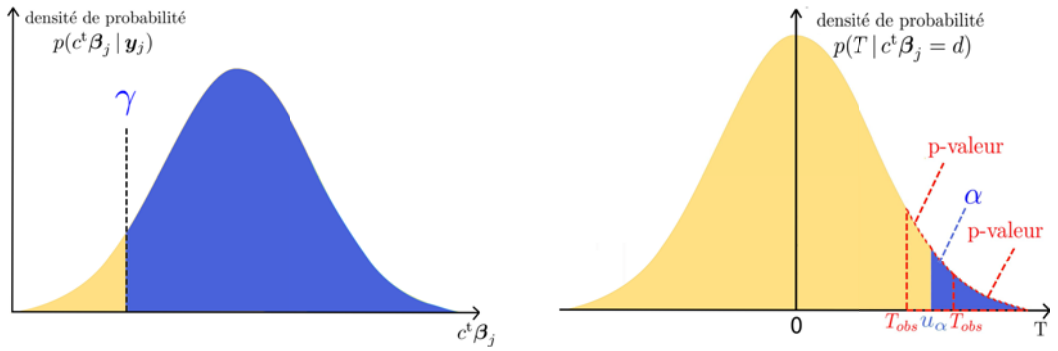


FIGURE II.7 – Différence entre le calcul des probabilités *a posteriori* (PPMs, à gauche) utilisées dans l'inférence bayésienne et le test-T (à droite) de l'approche classique.

II.4 Détection-Estimation Conjointe (DEC)

Apparue à travers les travaux de thèse de Salima Makni [Makni 2006] et [Makni 2005, Makni 2008], cette approche bayésienne regroupe d'une certaine manière plusieurs critères de modélisation cités dans les précédentes approches, elle combine donc :

- la détection et la localisation des activités cérébrales correspondantes aux conditions expérimentales du paradigme ;
- l'estimation de la FRH au niveau d'une région (parcelle) pour la robustesse, ce qui permet de prendre en compte sa variabilité intra et inter-sujets ;
- la modélisation des dérives basses fréquences ;
- la modélisation adaptative de la corrélation spatiale inter-voxels introduite à l'approche DEC par les travaux de thèse de Thomas Vincent [Vincent 2010a, Vincent 2010b, Vincent 2007] ;
- la prise en compte de la régularité temporelle de la FRH.

Son originalité concerne donc (i) la conjonction des étapes de détection et d'estimation qui sont habituellement effectuées séparément et séquentiellement (d'abord la détection puis l'estimation), et (ii) la considération d'une FRH parcellique ayant une forme constante et une amplitude variable sur un groupe de voxels voisins (parcelle) et pour tous les types de stimuli. Cela permet d'augmenter la robustesse de l'estimation et de diminuer le risque d'un sur-ajustement (*overfitting*) des données tout en exprimant la différence d'échelle entre les voxels voisins ainsi qu'entre les types de stimuli [Ciuciu 2003, Ciuciu 2004] à travers une nouvelle variable $\boldsymbol{A} = \{a_j^m; m = 1 : M, j = 1 : J\}$ appelés Niveaux de Réponse

Neuronale (NRNs) étant donné qu'ils expriment le niveau de réponse vasculaire induit par l'excitation neuronale.

L'approche DEC se base sur une parcellisation *a priori* définissant un partitionnement du cerveau en zones fonctionnellement homogènes (parcelles) de manière à ce que l'on puisse considérer la forme de la FRH constante sur l'ensemble des voxels de chaque parcelle.

Le modèle DEC représenté par la figure II.8 est donc défini, pour chaque voxel j d'une parcelle cérébrale donnée $\mathcal{P}\gamma$, par l'équation suivante :

$$\boldsymbol{y}_j = \boldsymbol{S}_j\boldsymbol{h} + \boldsymbol{P}\boldsymbol{\ell}_j + \boldsymbol{b}_j \quad \text{avec} \quad \boldsymbol{S}_j = \sum_{m=1}^{M} a_j^m \boldsymbol{X}^m \tag{II.22}$$

où :

- $\boldsymbol{y}_j$ est, comme défini précédemment, un vecteur de dimension N représentant le signal BOLD pré-traité au voxel j.
- a_j^m est le NRN correspondant à la condition m et au voxel j.
- $\boldsymbol{X}^m$ est une matrice binaire connue *a priori* qui représente les occurrences du $m$ième stimulus, sa dimension étant $N \times (D+1)$ et sa construction détaillée dans l'annexe B.
- $\boldsymbol{h}$ est un vecteur de dimension $D+1$ représentant la FRH commune à tous les voxels $j \in \mathcal{P}_\gamma$ ainsi qu'à tous les types de stimuli $m = 1 : M$.
- $\boldsymbol{P}\boldsymbol{\ell}_j$ modélise les dérives basses fréquences au voxel j où $\boldsymbol{P}$ est une matrice de dimensions $N \times O$ et dont les colonnes forment une base orthogonale de fonctions basses fréquences, $\boldsymbol{\ell}_j$ sont les coefficients correspondants.
- $\boldsymbol{b}_j$ est un vecteur de dimension N représentant le bruit.

II.4.1 Modèle bayésien hiérarchique

L'inférence bayésienne, comme précisé précédemment, nécessite la définition de la vraisemblance ainsi que des lois *a priori* des variables et paramètres du modèle.

Vraisemblance. En considérant un bruit gaussien $\boldsymbol{b}_j \sim \mathcal{N}(\boldsymbol{0}, \sigma_j^2\boldsymbol{\Lambda}_j^{-1})$, où $\boldsymbol{\Lambda}_j = \boldsymbol{I}_N$ dans le cas d'un bruit blanc gaussien indépendamment distribué, la distribution conditionnelle des données est également gaussienne $\boldsymbol{y}_j \,|\, \boldsymbol{a}_j, \boldsymbol{h}, \boldsymbol{\ell}_j, \sigma_j^2 \sim \mathcal{N}(\boldsymbol{S}_j\boldsymbol{h} + \boldsymbol{P}\boldsymbol{\ell}_j, \sigma_j^2\boldsymbol{\Lambda}_j^{-1})$, ce qui donne avec une hypothèse d'indépendance condi-

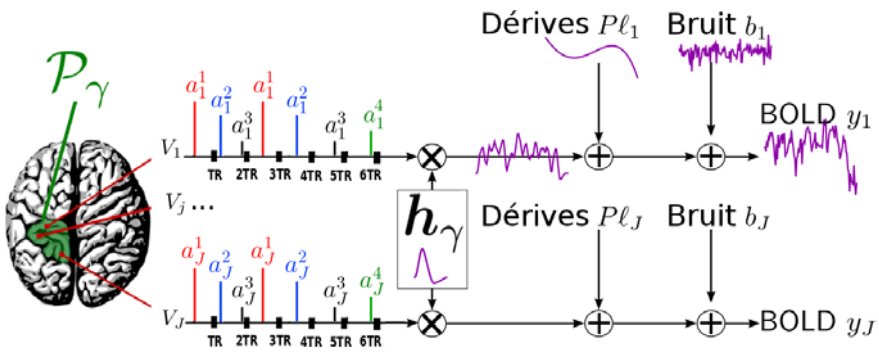


FIGURE II.8 – Illustration du modèle de Détection-Estimation Conjointe (DEC) pour une parcelle donnée $\mathcal{P}_\gamma$ et $M = 4$ types de stimuli, où a_j^m représente le Niveau de Réponse Neuronale (NRN) spécifique au voxel j et au type de stimuli m. La FRH $\boldsymbol{h}$ est quant à elle spécifique à la parcelle, tous deux contribuant à la formation du signal stimulus-induit auquel s'ajoutent les dérives basses fréquences $\boldsymbol{P\ell}_j$ et le bruit $\boldsymbol{b}_j$ afin d'obtenir le signal BOLD $\boldsymbol{y}_j$.

tionnelle inter-voxels :

$$p\left(\boldsymbol{Y} \mid \boldsymbol{A}, \boldsymbol{h}, \boldsymbol{L}, \boldsymbol{\sigma}^2\right) = \prod_{j=1}^{J} p\left(\boldsymbol{y}_j \mid \boldsymbol{a}_j, \boldsymbol{h}, \boldsymbol{\ell}_j, \sigma_j^2\right) \propto \prod_{j=1}^{J} \sigma_j^{-N} |\boldsymbol{\Lambda}_j|^{-\frac{1}{2}} \exp\left(\frac{-\widetilde{\boldsymbol{y}}_j^{\mathrm{t}} \boldsymbol{\Lambda}_j \widetilde{\boldsymbol{y}}_j}{2\sigma_j^2}\right) \quad \text{(II.23)}$$

avec $\boldsymbol{Y} = \{\boldsymbol{y}_j; j = 1 : J\}$, $\boldsymbol{L} = \{\boldsymbol{\ell}_j; j = 1 : J\}$ et $\widetilde{\boldsymbol{y}}_j = \boldsymbol{y}_j - \boldsymbol{S}_j \boldsymbol{h} - \boldsymbol{P\ell}_j$. Notons que la variable $\boldsymbol{A}$ peut être définie sous la forme $\boldsymbol{A} = \{\boldsymbol{a}_j; j = 1 : J\}$ avec $\boldsymbol{a}_j = \{a_j^m; m = 1 : M\}$ pour mettre en évidence les voxels ou sous la forme $\boldsymbol{A} = \{\boldsymbol{a}^m; m = 1 : M\}$ avec $\boldsymbol{a}^m = \{a_j^m; j = 1 : J\}$ pour mettre en évidence les conditions.

Distribution *a priori* de la FRH. L'approche DEC utilise exactement le même *a priori* gaussien défini dans l'approche RIF régularisée (section II.3.2) et qui permet l'introduction d'un lissage temporel induisant des formes plus cohérentes d'un point de vue physiologique. La seule différence concerne le regroupement des FRHs à travers les voxels voisins ainsi que les différentes conditions pour définir, à la place, une seule fonction par parcelle :

$$p\left(\boldsymbol{h} \mid v_h\right) \propto v_h^{-\frac{D-1}{2}} \exp\left(-\frac{1}{2v_h} \boldsymbol{h}^{\mathrm{t}} \boldsymbol{R}^{-1} \boldsymbol{h}\right) \quad \text{(II.24)}$$

où $\boldsymbol{R} = \left\{\frac{\boldsymbol{D}_2^{\mathrm{t}} \boldsymbol{D}_2}{(\Delta t)^4}\right\}^{-1}$ est la matrice de covariance de $\boldsymbol{h}$ avec $\boldsymbol{D}_2$ la matrice des différences secondes associées au vecteur $\boldsymbol{h}$. Comme dans le cas de la RIF régularisée temporellement (section II.3.2), la matrice $\boldsymbol{R}$ permet d'exprimer une contrainte

sur la dérivée seconde pénalisant ainsi les fortes variations de pentes. Cela dans le but d'obtenir des FRH lisses.

Distribution *a priori* des NRNs. Différentes distributions ont été testées dans la thèse de Salima Makni [Makni 2006] en commençant par une distribution gaussienne pour chaque condition expérimentale m. Cette distribution a ensuite été remplacée par un mélange de deux gaussiennes (figure II.9-(a)) étant donné qu'une parcelle en question peut parfois contenir des voxels activés et inactivés par une condition. Pour ce faire, il a été introduit des variables binaires latentes $\boldsymbol{Q}$ qu'on appellera labels par la suite. En tenant compte de l'indépendance inter-voxels des NRNs conditionnellement aux labels et en faisant l'hypothèse d'indépendance *a priori* des NRNs inter-conditions on peut écrire :

$$\begin{aligned} p\left(\boldsymbol{A} \,|\, \boldsymbol{\theta}_A\right) &= \prod_{m=1}^{M} p\left(\boldsymbol{a}^m \,|\, \boldsymbol{\theta}_m\right) = \prod_{m=1}^{M} \sum_{\boldsymbol{q}^m} p\left(\boldsymbol{a}^m \,|\, \boldsymbol{q}^m, \boldsymbol{\theta}_m\right) p(\boldsymbol{q}^m \,|\, \boldsymbol{\theta}_m) \\ &= \prod_{m=1}^{M} \sum_{\boldsymbol{q}^m} \left[\prod_{j=1}^{J} p(a_j^m \,|\, q_j^m, \boldsymbol{\theta}_m) \right] p(\boldsymbol{q}^m \,|\, \boldsymbol{\theta}_m) \end{aligned} \tag{II.25}$$

avec

$$p(a_j^m \,|\, q_j^m = i, \boldsymbol{\theta}_m) = \mathcal{N}(a_j^m; \mu_i^m, v_i^m) \quad \text{pour} \quad i \in \{0, 1\} \tag{II.26}$$

où $\boldsymbol{A} = \{\boldsymbol{a}^m; m = 1 : M\}$ avec $\boldsymbol{a}^m = \{a_j^m; j = 1 : J\}$, $\boldsymbol{\theta}_A = \{\boldsymbol{\theta}_m; m = 1 : M\}$ avec $\boldsymbol{\theta}_m = \{\mu_1^m, v_1^m, \mu_0^m, v_0^m\}$, une contrainte étant l'imposition de la moyenne de la classe inactive à 0 ($\mu_0^m = 0$). Notons que $\boldsymbol{q}^m = \{q_j^m; j = 1 : J\}$ où pour $q_j^m = 1$ (resp. 0), $a_j^m \sim \mathcal{N}(a_j^m; \mu_1^m, v_1^m)$ (resp. $\mathcal{N}(a_j^m; \mu_0^m, v_0^m)$) indiquant ainsi l'appartenance du voxel j à la classe active (resp. inactive) pour la condition m, cela nous permet d'écrire l'équation (II.26) sous la forme suivante :

$$p(a_j^m \,|\, q_j^m, \boldsymbol{\theta}_m) = (1 - q_j^m)\, \mathcal{N}(a_j^m; \mu_0^m, v_0^m) + q_j^m\, \mathcal{N}(a_j^m; \mu_1^m, v_1^m) \tag{II.27}$$

Dans le cas où l'une des conditions est non pertinente pour une parcelle donnée au sens où elle ne suscite pas d'activations significatives, les deux gaussiennes du mélange gaussien *a posteriori* des NRNs se recouvrent largement et les classes d'activation deviennent difficilement séparables (figure II.9-(b)) causant une classification aléatoire des voxels et donc un taux de faux positifs élevé. L'utilisation d'un modèle de mélange gamma-gaussien (figure II.10-(a)) permet d'éviter ce genre de problème étant donnée la contrainte de positivité introduite par la

distribution gamma sur les NRNs associés aux activations des voxels activés. Une autre approche propose un mélange à trois classes (2 gammas et 1 gaussienne) (figure II.10-(b)) permettant de modéliser les voxels activés et déactivés (avec une activation négative) par deux classes différentes, rendant ainsi le modèle plus robuste dans les cas où des voxels activés et déactivés co-existent dans la parcelle étudiée. Cette modélisation est néanmoins moins parcimonieuse que celle à deux classes dans certains cas où le passage de quelques voxels inactivés dans la classe de déactivation, avec de faibles valeurs des NRNs associées, indique l'absence d'une déactivation significative [Vincent 2010a]. Ceci peut également être abordé par une procédure de comparaison de modèles [Raftery 2007] qui, dans le cas de l'approche DEC, doit être effectuée parcelle par parcelle si on veut modéliser trois classes. Une autre possibilité utilisée dans [Smith 2007] est le mélange bernoulli-gaussien, où la classe active est modélisée par une gaussienne centrée autour d'une valeur non nulle et l'inactive est modélisée par une Dirac en 0, ce qui correspond à la version dégénérée du modèle à 2 classes gaussiennes lorsque la variance de la classe modélisant les voxels inactivés tend vers 0. Un mélange bernoulli-gaussien risque cependant de causer une augmentation de la variance estimée de la classe active afin d'englober les voxels inactivés se situant autour de 0.

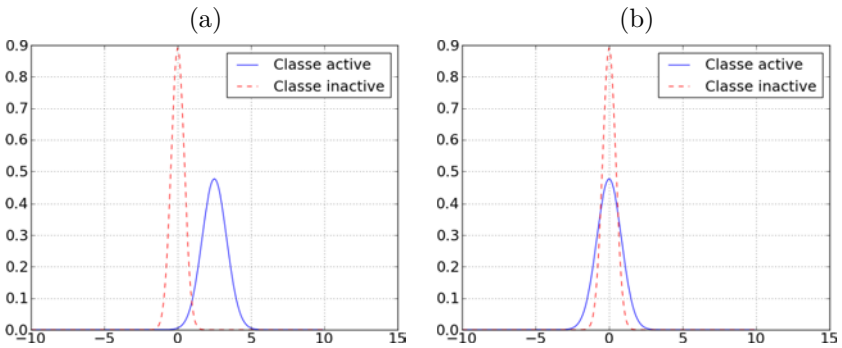


FIGURE II.9 – (a) Illustration de l'*a priori* de mélange gaussien pour les NRNs avec une gaussienne centrée en 0 représentant la classe inactive et une autre centrée autour d'une moyenne non nulle représentant la classe active. (b) Illustration du problème de la perte d'identifiabilité du mélange gaussien *a posteriori*, notons la confusion entre les deux classes d'activation dans le cas d'une condition non pertinente.

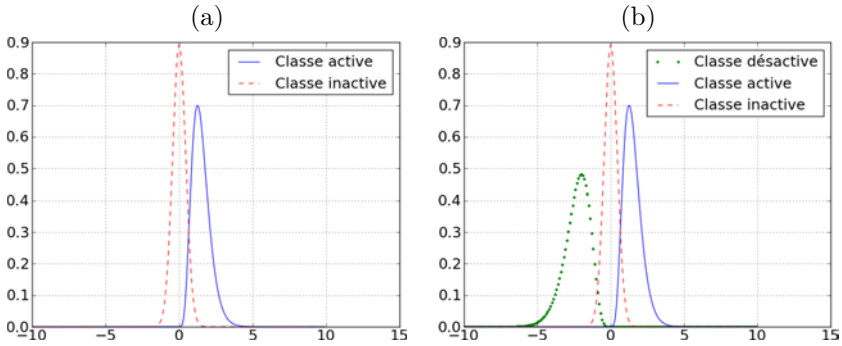


FIGURE II.10 – Illustration de l'*a priori* de mélange gamma-gaussien (a) et de mélange à trois classes (2 gammas et 1 gausienne) (b) pour les NRNs. Une gaussienne centrée en 0 représente la classe de voxels inactivés dans les deux cas alors que deux distributions gamma permettent de modéliser deux classes d'activation des voxels (activés et déactivés) dans (b) au lieu d'une seule classe modélisant uniquement les voxels activés dans (a).

Distribution *a priori* des labels $\boldsymbol{Q}$. Les premiers travaux sur l'approche DEC [Makni 2005], ont considéré un mélange gaussien indépendant sur les NRNs en supposant une indépendance inter-voxels sur les variables latentes $\boldsymbol{Q}$:

$$p\left(\boldsymbol{Q} \,|\, \boldsymbol{\xi}\right) = \prod_{m=1}^{M} p\left(\boldsymbol{q}^m \,|\, \xi_m\right) = \prod_{m=1}^{M} \prod_{j} p\left(q_j^m \,|\, \xi_m\right) \tag{II.28}$$

L'extension de l'approche DEC tenant compte de la modélisation de l'information spatiale a été mise en place à travers les travaux de thèse de Thomas Vincent [Vincent 2010a, Vincent 2010b, Vincent 2007] par la définition *a priori* d'un champ de Markov discret (champ de Ising pour le cas de 2 classes) sur les labels, ce qui a permi la modélisation d'un mélange spatial sur les NRNs. Cet *a priori* est exprimé par l'équation suivante, en gardant l'hypothèse d'indépendance entre les différentes conditions :

$$p\left(\boldsymbol{Q} \,|\, \boldsymbol{\xi}\right) = \prod_{m=1}^{M} p\left(\boldsymbol{q}^m \,|\, \xi_m\right) = \prod_{m=1}^{M} \frac{1}{Z(\xi_m)} \exp(\xi_m U(\boldsymbol{q}^m)) \tag{II.29}$$

avec

$$Z(\xi_m) = \sum_{\boldsymbol{q}^m} \exp(\xi_m U(\boldsymbol{q}^m)) \tag{II.30}$$

$$U(\boldsymbol{q}^m) = \frac{1}{2} \sum_{j} \sum_{k \in N_j} u_{j,k} \; I(q_j^m = q_k^m) \tag{II.31}$$

où $I(x) = 1$ si x est vrai et 0 sinon, N_j définit les voxels k voisins de j et $u_{j,k}$ sont les poids associés à chaque paire de voisinage (j, k). Dans (Vincent 2010) le voisinage choisi est de type six-connexité en $3D$ permettant de fixer $u_{j,k} = 1$ pour $k \in N_j$ et 0 sinon. On pourrait également prendre en compte l'anisotropie des images selon une des directions (généralement z), en affectant des valeurs de poids différents aux voisins n'ayant pas la même distance de voisinage avec le voxel en question. Enfin, la régularité spatiale est représentée par $\boldsymbol{\xi} = \{\xi_m; m = 1 : M\}$ où $\xi_m \geq 0$ contrôle, pour la $m$ième condition, la régularité spatiale entre les composantes de $\boldsymbol{q}^m$. Une valeur nulle de ξ_m par exemple, n'impose aucune corrélation inter-voxels. Plus la valeur de ξ_m augmente, plus on favorise la présence de clusters de voxels jusqu'à ce que la valeur de ξ_m devienne assez grande pour favoriser l'appartenance de tous les voxels à la même classe. Cette approche est d'un côté (i) non-supervisée dans le sens où le paramètre de régularité spatiale ξ_m est estimé et d'un autre (ii) adaptative dans le sens où ξ_m varie entre les différentes conditions $\xi_m \neq \xi_{m'}$ ainsi qu'entre les différentes parcelles cérébrales.

Densité *a posteriori*. Etant données les définitions ci-dessus de la vraisemblance et des lois *a priori*, et en considérant un bruit blanc gaussien, cette densité est définie par :

$$\begin{aligned} p(\boldsymbol{A}, \boldsymbol{Q}, \boldsymbol{h}, \boldsymbol{L}, \boldsymbol{\theta} \,|\, \boldsymbol{Y}) \propto p(\boldsymbol{Y} \,|\, \boldsymbol{A}, \boldsymbol{h}, \boldsymbol{L}, \boldsymbol{\sigma}^2)\; p(\boldsymbol{A} \,|\, \boldsymbol{Q}, \boldsymbol{\theta}_A)\; p(\boldsymbol{Q} \,|\, \boldsymbol{\xi}) \\ p(\boldsymbol{h} \,|\, v_h)\; p(\boldsymbol{L} \,|\, v_\ell)\; p(\boldsymbol{\theta}) \end{aligned} \tag{II.32}$$

où $\boldsymbol{\theta} = \{\boldsymbol{\sigma}^2, \boldsymbol{\theta}_A, \boldsymbol{\xi}, v_h, v_\ell\}$ et $p(\boldsymbol{\theta})$ permet l'introduction d'*a priori* sur les paramètres du modèle et dont la définition sera détaillée dans les sections d'inférence (chapitre III). Le terme $p(\boldsymbol{L} \,|\, v_\ell)$ représente la loi *a priori* sur les coefficients $\boldsymbol{L}$ des dérives basses fréquences, cette loi étant définie selon l'équation suivante et en supposant une indépendance spatiale inter-voxels :

$$p\left(\boldsymbol{L} \,|\, v_\ell\right) = \prod_j p\left(\boldsymbol{\ell}_j \,|\, v_\ell\right) = \prod_j \mathcal{N}(\boldsymbol{\ell}_j; \boldsymbol{0}, v_\ell \boldsymbol{I}_O) \tag{II.33}$$

avec O la dimension du vecteur $\boldsymbol{\ell}_j$ et $v_\ell \boldsymbol{I}_O$ sa matrice de variance-covariance.

II.4.2 Choix de la parcellisation

Deux critères principaux concernant la parcellisation ont été étudiés. Le premier s'intéresse au nombre de parcelles optimal qui, selon un Critère d'Informa-

tion Bayésien (BIC) et une procédure de validation croisée, est d'environ 200 parcelles par hémisphère [Thyreau 2006]. Le deuxième concerne l'influence de la position des parcelles qui serait majorée s'il y avait un changement brusque de la réponse vasculaire aux frontières entres parcelles. D'après une analyse réalisée dans [Vincent 2010a], la forme de la FRH semble avoir un caractère diffus pas très sensible au positionnement des parcelles à l'inverse de son amplitude captée par les NRNs et qui est liée à la parcellisation des données.

Plusieurs travaux de recherche se sont intéressés à la parcellisation cérébrale. Celle proposée dans [Flandin 2002a, Flandin 2002b], par exemple, est basée sur les *K-means* utilisant des distances géodésiques où l'introduction de contraintes anatomiques et fonctionnelles est effectuée à travers l'utilisation de poids sur ces distances. Certains travaux proposent une parcellisation anatomique basée sur les gyri [Cachia 2003] ou les sillons [Perrot 2009]. D'autres l'effectuent à partir des données fonctionnelles et des cartes statistiques issues du MLG classique [Thirion 2005, Thirion 2006, Tucholka 2008]. Ce dernier souffre cependant d'un problème de modélisation de la FRH, ce qui risque d'influencer la parcellisation obtenue. Il existe aussi, des atlas anatomo-fonctionnels : l'AAL, l'atlas Harvard–Oxford, l'altlas sulci, Ncuts, les réseaux *resting-state* [Varoquaux 2013]. De récents travaux s'intéressent au raffinement d'une parcellisation initiale au cours de l'analyse DEC [Chaari 2012]. Enfin, nous citons les travaux de thèse de Yongnan Ji [Ji 2010] qui s'intéressent à l'optimisation de la parcellisation commune à un groupe de sujets, ceci en se basant sur le principe de partitionnement d'un graphe multipartite. L'étude de l'optimalité de la parcellisation ne faisant pas partie de ce travail de thèse, nous avons utilisé une parcellisation obtenue à partir de l'algorithme de clustering hiérarchique de Ward basé sur la minimisation de la variance intra-classe en tenant compte d'une contrainte de connexité.

II.5 Conclusion

Ce chapitre constitue en première partie un état de l'art sur les méthodes d'analyses les plus répandues en IRMf parmi lesquelles certaines s'intéressent uniquement à la détection des activités cérébrales induites par les différentes conditions (types de stimuli) présentes dans le paradigme expérimental. D'autres se focalisent sur l'estimation de la FRH reflétant l'hémodynamique qui est à la base du signal BOLD et qui varie entre les différentes régions cérébrales et entre

les différents sujets.

Nous avons aussi vu l'intérêt des approches bayésiennes par rapport aux inférences classiques dans le sens où elles permettent la prise en compte d'informations *a priori* dans le but d'améliorer les résultats d'analyse.

Les aspects de détection et d'estimation étant inter-dépendants, une approche DEC les réunissant a été introduite et dont plus de détails seront discutés dans les chapitres suivants étant donné que cette approche constitue le socle commun à mes contributions.

Chapitre III

Approches parcimonieuses en IRMf

La non pertinence d'une condition expérimentale (type de stimuli) en IRMf est définie comme l'absence d'activité significative induite par celle-ci. Différents travaux se sont intéressés à cette notion ainsi qu'à son influence sur les résultats d'analyse. Dans [Donnet 2004], les auteurs proposent une procédure de sélection de modèles permettant de déterminer la sous-famille des conditions pertinentes. Elle nécessite cependant l'exécution de tous les modèles possibles (2^M si M est le nombre de conditions) et donc un temps de calcul élevé. De plus, elle n'autorise pas la famille de conditions pertinentes à changer entre les régions cérébrales. Une autre possibilité est d'effectuer cette sélection adaptativement et simultanément à l'analyse. Dans le contexte du MLG standard, les travaux de [Luo 2005, Oikonomou 2012] proposent d'introduire un *a priori* ARD (*Automatic Relevance Determination*) sur les effets $\boldsymbol{\beta}$. Cette procédure est utilisée dans la méthode RVM (*Relevance Vector Machine*) qui représente une technique d'apprentissage bayésien permettant de résoudre des problèmes de régression et de classification parcimonieuses [Tipping 2001]. Les auteurs s'intéressent principalement à la modélisation parcimonieuse des dérives basses fréquences et non du signal stimulus-induit. De plus, ils utilisent une FRH canonique fixe à travers le cerveau entier.

Dans ce travail de thèse, nous nous sommes intéressés à la sélection adaptative et automatique des conditions pertinentes dans l'approche DEC dans laquelle la FRH est estimée. Comme nous l'avons vu dans le chapitre précédent, cette approche suppose, pour chaque condition m, que les NRNs suivent une loi de

mélange de deux gaussiennes, une représentant la classe des voxels activés et une autre centrée autour de 0 représentant la classe des voxels inactivés, ceci indépendamment du fait que la condition m soit pertinente ou non. Or, si m est non pertinente, les NRNs seront tous proches de 0 mais le modèle cherchera néanmoins à estimer deux gaussiennes au lieu d'une. Afin de pallier ce problème de surajustement du modèle, nous proposons d'introduire un groupe de variables binaires supplémentaires $\boldsymbol{w} = \{w^m; m = 1 : M\}$ (section III.2.1) où w^m exprime la pertinence de la $m$ième condition dans une parcelle donnée $\mathcal{P}_\gamma$. Ceci nous ramène à une procédure de sélection de variables [OHara 2009]. Une difficulté est le choix de la loi de $\boldsymbol{w}$ en fonction des autres variables et paramètres du modèle pour traduire la notion de pertinence souhaitée. Ici, nous proposons une loi de Bernoulli dans laquelle la probabilité de succès (la condition est pertinente) est une fonction sigmoïde $S(x)$ fournissant un passage de 0 à 1 moins brusque et donc plus tolérant aux incertitudes qu'une fonction escalier. On appellera l'abscisse du point d'inflexion de la sigmoïde le seuil de pertinence étant donné qu'au-dessus de cet abscisse la probabilité qu'une condition soit considérée comme pertinente est élevée (> 0.5). La question étant le choix de x parmi les variables ou paramètres du modèle ainsi que le choix du seuil de pertinence. Deux propositions ont été testées dans le cadre de ce travail de thèse, la première consiste à lier $\boldsymbol{w}$ aux labels $\boldsymbol{Q}$ (section III.2.2) et la deuxième à lier $\boldsymbol{w}$ aux paramètres $\boldsymbol{\mu}_1 = \{\mu_1^m; m = 1 : M\}$ où μ_1^m représente la moyenne de la classe active pour la $m$ième condition (section III.2.3).

Pour commencer nous allons présenter le MLG creux (*sparse*) proposé par [Luo 2005, Oikonomou 2012]. Nous discuterons ensuite, l'approche DEC parcimonieuse qui représente la contribution principale de cette thèse en détaillant les deux propositions de positionnement de $\boldsymbol{w}$ par rapport aux variables et paramètres du modèle ainsi que les deux techniques d'inférences (MCMC et VEM) sur lesquelles va se baser notre analyse. Enfin, l'*a priori* ARD utilisé dans la littérature pour gérer le problème de pertinence dans le cadre de la modélisation des dérives basses fréquences, a également été adapté (section III.2.4) à l'approche DEC, où les $\boldsymbol{w}$ ne sont plus des variables binaires mais gaussiennes. Ceci permet de donner des scores de pertinence aux conditions expérimentales plutôt que des valeurs binaires. De plus l'*a priori* ARD représente un exemple de loi où $\boldsymbol{w}$ n'est plus reliée aux variables et paramètres du modèle comme c'est le cas dans le choix de la fonction sigmoïde.

Par la suite les matrices et vecteurs sont notés en gras avec les matrices en majuscule et les vecteurs en minuscule. Un vecteur est par convention un vecteur colonne. La transposé est notée ^{t}. Notons aussi que le modèle DEC standard qui considère toutes les conditions comme pertinentes sera désigné par la suite par le modèle *complet*.

III.1 MLG creux (*sparse*)

Comme expliqué dans le chapitre II, le MLG est exprimé par l'équation suivante :

$$\forall j \; ; \quad \boldsymbol{y}_j = \boldsymbol{X}\boldsymbol{\beta}_j + \boldsymbol{b}_j \tag{III.1}$$

où $\boldsymbol{y}_j = \{y_{n,j}; n = 1 : N\}^{\mathrm{t}}$ représente le signal BOLD au voxel j avec N le nombre de scans, $\boldsymbol{X}$ est la matrice de dessin de taille $N \times R$ et $\boldsymbol{\beta}_j = \{\beta_j^r; r = 1 : R\}^{\mathrm{t}}$ avec R le nombre de régresseurs. Dans ce modèle les dérives basses fréquences représentant les artéfacts physiologiques peuvent être écartées à travers un filtrage passe-haut dont la fréquence de coupure f_0 doit être choisie correctement ou bien en les modélisant dans la matrice de dessin $\boldsymbol{X}$ à travers une base de fonctions. La difficulté étant, dans ce dernier cas, le choix du nombre de fonctions de base évitant à la fois le sur et le sous apprentissage. Afin de faire face à cette difficulté [Luo 2005, Oikonomou 2012] proposent une extension creuse du MLG où une procédure d'apprentissage bayésien permet, à partir des données, la selection des régresseurs pertinents nécessaires à la modélisation du signal observé, rendant ainsi la matrice de dessin flexible en sélectionnant simultanément à l'analyse la sous famille R' des R régresseurs exprimant au mieux les données. Les auteurs s'intéressent principalement à la sélection des fonctions de base modélisant les dérives, ceci en définissant la matrice de dessin comme suit :

$$\boldsymbol{X} = [\boldsymbol{x}_1|\boldsymbol{x}_2|...|\boldsymbol{x}_n|...|\boldsymbol{x}_N|\boldsymbol{s}|\boldsymbol{1}] \tag{III.2}$$

avec $\boldsymbol{x}_n$ pour $n = 1, ..., N$ représentent les régresseurs des dérives basses fréquences, $\boldsymbol{s}$ le régresseur stimulus-induit représentant la convolution du paradigme avec la FRH canonique et le dernier régresseur étant un régresseur constant représentant la ligne de base. Le but est donc de réduire la matrice de dessin de dimension $N \times R$ avec $R = (N + 2)$ à la dimension $N \times R'$, ceci en introduisant un *a priori*

ARD sur les effets $\boldsymbol{\beta}_j$:

$$p(\boldsymbol{\beta}_j \,|\, \boldsymbol{\alpha}_j) = \prod_{r=1}^{R} \mathcal{N}\left(\beta_j^r; 0, \frac{1}{\alpha_{j,r}}\right) \tag{III.3}$$

où $\boldsymbol{\alpha}_j = \{\alpha_{j,r}; r = 1 : R\}$ et $p(\boldsymbol{\alpha}_j)$ est supposé uniforme dans [Luo 2005] et de loi gamma dans [Oikonomou 2012] $p(\boldsymbol{\alpha}_j) \propto \prod_{r=1}^{R} \alpha_{j,r}^{a-1} \exp(-b\ \alpha_{j,r})$ avec $a, b \to 0$ pour obtenir une loi non informative. L'analyse montre que la plupart des hyper-paramètres $\alpha_{j,r}$ tendent vers de grandes valeurs réduisant les effets correspondants β_j^r à 0 et éliminant ainsi les régresseurs en question. Des analyses sur données simulées et réelles [Luo 2005, Oikonomou 2012] montrent l'intérêt de cette sélection sur les résultats de traitement où une baisse du taux de faux positifs est remarquée. L'approche est efficace concernant la détermination du nombre de fonctions de base. Elle considère cependant une FRH canonique et ne discute pas la sélection de régresseurs stimulus-induits (un seul type de condition expérimentale est présent dans les données traitées). Ici nous allons nous intéresser à la pertinence de modélisation du signal stimulus-induit dans le cadre de l'approche DEC où la FRH est estimée. Etant donné que le nombre de conditions expérimentales dans les expériences IRMf n'est pas énorme, l'approche proposée sera désignée comme *parcimonieuse* au lieu de creuse.

III.2 Approches DEC parcimonieuses

III.2.1 Introduction d'une variable de pertinence

Comme expliqué précédemment, nous avons proposé de modéliser la pertinence des conditions expérimentales à travers un vecteur de variables supplémentaires $\boldsymbol{w} = \{w^m; m = 1 : M\}$ où $w^m = 0$ exprime la non pertinence de la $m$ième condition dans la parcelle $\mathcal{P}_\gamma$ et induit donc la suppression du terme correspondant dans la modélisation du signal stimulus-induit $\sum_{m=1}^{M} a_j^m w^m \boldsymbol{X}^m$. Le modèle DEC devient donc :

$$\boldsymbol{y}_j = \boldsymbol{S}_j \boldsymbol{h} + \boldsymbol{P}\boldsymbol{\ell}_j + \boldsymbol{b}_j \quad \text{avec} \quad \boldsymbol{S}_j = \sum_{m=1}^{M} a_j^m w^m \boldsymbol{X}^m \tag{III.4}$$

La seule différence concernant la vraisemblance par rapport au modèle DEC standard (équation (II.23)) est la définition du terme $\boldsymbol{S}_j$ qui inclut le groupe de

variables supplémentaires $\boldsymbol{w}$. Les distributions *a priori* sur $\boldsymbol{h}$ et $\boldsymbol{Q}$ ne changent pas par rapport à l'approche DEC standard (équations (II.24) et (II.29)). La distribution *a priori* des NRNs est cependant modifiée afin de prendre en compte la nouvelle variable de pertinence. Dans la section II.4.1, plusieurs *a priori* ont été testés parmi lesquels le mélange gaussien semblait être le plus parcimonieux. Cette définition souffre cependant d'un problème de perte d'identifiabilité du mélange gaussien *a posteriori* dans le cas d'une condition non pertinente. Une solution possible, proposée par Salima Makni (Makni 2006), est de remplacer ce mélange par un mélange gamma-gaussien afin de garantir la positivité de la classe active évitant ainsi sa confusion avec la gaussienne inactive et donc la classification aléatoire des voxels induisant un taux de faux positifs élevé quand m est non pertinente. Néanmoins, cette définition ne tient pas compte de la présence de voxels déactivés, ce qui pourait être faisable avec une modélisation à trois classes (section II.4.1). Ici, nous revenons au mélange gaussien en proposant de résoudre le problème de pertinence différemment à travers l'ajout de $\boldsymbol{w}$ et la modélisation *a priori* du principe que la plupart des voxels sont inactivés quand une condition m est non pertinente ($w^m = 0$) dans une parcelle donnée $\mathcal{P}_\gamma$, ceci indépendamment des étiquettes $\boldsymbol{q}^m = \{q_j^m; j \in \mathcal{P}_\gamma\}$. Pour cela, l'*a priori* des NRNs devient :

$$p\left(\boldsymbol{A} \,|\, \boldsymbol{\theta}_A\right) = \prod_{m=1}^{M} \sum_{\boldsymbol{q}^m, w^m} p\left(\boldsymbol{a}^m \,|\, \boldsymbol{q}^m, w^m, \boldsymbol{\theta}_m\right) p(\boldsymbol{q}^m, w^m \,|\, \boldsymbol{\theta}_m) \tag{III.5}$$

$$p\left(\boldsymbol{a}^m \,|\, \boldsymbol{q}^m, w^m, \boldsymbol{\theta}_m\right) = \prod_{j=1}^{J} \left[(1 - q_j^m w^m)\mathcal{N}(\mu_0^m, v_0^m) + q_j^m w^m \mathcal{N}(\mu_1^m, v_1^m)\right] \tag{III.6}$$

Pour $i \in \{0, 1\}$, $\left(a_j^m \,|\, w^m = 1, q_j^m = i\right) \sim \mathcal{N}(\mu_i^m, v_i^m)$ et $\left(a_j^m \,|\, w^m = 0, q_j^m = i\right) \sim \mathcal{N}(\mu_0^m, v_0^m)$, les NRNs ne dépendent donc plus des labels quand $w^m = 0$ mais sont tous distribués selon la distribution centrée en 0 de la classe inactive d'où le rôle de la variable de pertinence w^m.

Définition de la loi sur $\boldsymbol{w}$. Comme pour les autres variables du modèle, on suppose une indépendance entre les différents types de stimuli (la pertinence d'une condition est indépendante de la pertinence d'une autre dans une parcelle donnée).

- Cas d'un w^m binaire : on utilise une distribution de Bernoulli $\mathcal{B}(p_m)$ avec $p_m = S(x^m)$ une fonction sigmoïde dépendant de x^m. Cette dernière sera définie par la suite à l'aide des variables et paramètres du modèle. Notons $\boldsymbol{\tau} = \{\tau_1, \tau_2\}$ les

paramètres de la sigmoïde avec τ_1 le paramètre contrôlant la pente de la sigmoïde et τ_2 l'abscisse de son point d'inflexion considéré comme le seuil de pertinence au-dessus duquel la condition a une grande probabilité (> 0.5) d'être considérée comme pertinente. La loi sur w^m est donc exprimée par l'équation suivante :

$$p(\boldsymbol{w} \,|\, \boldsymbol{x}, \boldsymbol{\tau}) = \prod_{m=1}^{M} p(w^m \,|\, x^m, \boldsymbol{\tau}) = \prod_{m=1}^{M} [w^m p_m + (1 - w^m)(1 - p_m)] \tag{III.7}$$

avec

$$p_m = p(w^m = 1 \,|\, x^m, \boldsymbol{\tau}) = S(x^m) = \frac{1}{1 + \exp\left[-\tau_1 \left(x^m - \tau_2\right)\right]} \tag{III.8}$$

Avant de passer à l'étape d'inférence, deux points principaux concernant la définition de p_m doivent être discutés :

- Le choix de x^m dans l'équation (III.8) qui doit être cohérent avec la définition de la non pertinence selon laquelle une condition m non pertinente ne cause pas d'activité significative. Ceci peut se traduire par un faible nombre de voxels activés $\sum_{j=1}^{J} q_j^m$ où la présence de quelques voxels activés ($q_j^m = 1$) serait induit par le bruit. Une possibilité est donc de choisir $x^m = \sum_{j=1}^{J} q_j^m$, où J représente le nombre de voxels de la parcelle $\mathcal{P}_\gamma$ (section III.2.2). Une autre possibilité serait le choix $x^m = (\mu_1^m)^2$ (section III.2.3) où une faible valeur de μ_1 (moyenne de la classe active) reflèterait l'absence d'activité significative. Ces deux propositions ont été développées et testées sur des jeux de données simulées et réelles qu'on détaillera dans les chapitres IV et V.
- Le choix des paramètres de la sigmoïde τ_1, τ_2 qui sera détaillé dans le chapitre IV.

• Cas d'un *a priori* ARD : dans le précédent choix de la loi sur w^m nous avons proposé de lier cette variable à d'autres variables du modèle à travers une fonction sigmoïde. Une autre possibilité consiste à définir un *a priori* spécifique à w^m comme c'est le cas de l'*a priori* ARD déjà utilisé pour la détection de la pertinence d'un régresseur de la matrice de dessin du MLG où il a été défini sur les effets $\boldsymbol{\beta}$ (section III.1). Une approche comparable en DEC serait l'application de cet *a priori* sur les variables $\boldsymbol{A}$ représentant les NRNs. Etant donné l'*a priori* de mélange gaussien déjà défini sur ces variables, nous proposons d'appliquer l'*a priori* ARD sur les variables de pertinence $\boldsymbol{w}$. Une telle définition induirait des

$\boldsymbol{w}$ réelles et non plus binaires, ce qui revient dans le cas où w^m n'est pas estimée à 0 pour la condition m à factoriser une partie de $\boldsymbol{A}$ dans $\boldsymbol{w}$ causant ainsi un problème d'identifiabilité. Notons alors que quantativement la variable $\boldsymbol{A}$ dans le modèle *complet* serait équivalente au produit $\boldsymbol{Aw}$ dans le modèle *parcimonieux* avec un *a priori* ARD sur $\boldsymbol{w}$. En gardant la notion d'indépendance entre les conditions expérimentales l'*a priori* sur $\boldsymbol{w}$ devient :

$$p(\boldsymbol{w}|\boldsymbol{\alpha}) = \prod_{m=1}^{M} p(w^m|\alpha_m) = \prod_{m=1}^{M} \mathcal{N}\left(w^m; 0, \frac{1}{\alpha_m}\right) \tag{III.9}$$

où α_m représente la précision.

III.2.2 Connexion aux variables de classification

La non pertinence étant définie comme l'absence d'activations significatives pour une condition expérimentale m dans une parcelle donnée $\mathcal{P}_\gamma$, on peut donc considérer la condition non pertinente si le nombre de voxels activés par celle-ci est suffisamment petit. Ceci peut être effectué en faisant dépendre les variables de pertinence $\boldsymbol{w}$ des labels (variables de classification) $\boldsymbol{Q}$. Dans ce cas, on peut choisir $x^m = \sum_{j=1}^{J} q_j^m$ et la sigmoïde précédemment définie devient :

$$p_m = S\left(\sum_{j=1}^{J} q_j^m\right) = \frac{1}{1 + \exp\left[-\tau_1\left(q_{/j}^m + q_j^m - \tau_2\right)\right]} \tag{III.10}$$

avec $q_{/j}^m = \sum_{j' \neq j} q_{j'}^m$

Comme indiqué précédemment, l'inférence a été effectuée à travers deux techniques MCMC, en continuation des travaux de thèse de Salima Makni et Thomas Vincent [Makni 2006, Vincent 2010a], et VEM, en continuation des travaux de Lotfi Chaari [Chaari 2013].

III.2.2.1 Simulations de Monte Carlo par Chaînes de Markov - MCMC

Cette technique permet la simulation d'échantillons tirés selon une distribution complexe, et impossible à simuler par tirages indépendants. Son nom vient du fait que la simulation d'une nouvelle réalisation est basée sur la réalisation précédente générant ainsi une chaîne de Markov où la probabilité de transition

vers une nouvelle réalisation dépend uniquement de la réalisation la plus récente.

Densité *a posteriori*. En tenant compte de la définition de la vraisemblance ainsi que des lois *a priori* détaillées ci-dessus nous pouvons écrire la densité *a posteriori* à laquelle nous nous intéressons de la manière suivante :

$$\begin{aligned} p(\boldsymbol{A}, \boldsymbol{w}, \boldsymbol{Q}, \boldsymbol{h}, \boldsymbol{L}, \boldsymbol{\theta} \,|\, \boldsymbol{Y}) \propto\; & p(\boldsymbol{Y} \,|\, \boldsymbol{A}, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{L}, \boldsymbol{\sigma}^2)\; p(\boldsymbol{A} \,|\, \boldsymbol{w}, \boldsymbol{Q}, \boldsymbol{\theta}_A)\; p(\boldsymbol{Q} \,|\, \boldsymbol{\xi}) \\ & p(\boldsymbol{w} \,|\, \boldsymbol{Q}, \boldsymbol{\tau})\; p(\boldsymbol{h} \,|\, v_h)\; p(\boldsymbol{L} \,|\, v_l) p(\boldsymbol{\theta}) \end{aligned} \tag{III.11}$$

où $p(\boldsymbol{L} \,|\, v_l)$ représente l'*a priori* sur les coefficients $\boldsymbol{L} = \{\boldsymbol{\ell}_j; j = 1 : J\}$ des dérives basses fréquences (équation II.33). Rappelons que $\boldsymbol{\theta} = \{\boldsymbol{\sigma}^2, \boldsymbol{\theta}_A, \boldsymbol{\tau}, \boldsymbol{\xi}, v_h, v_l\}$. Ici, les paramètres de la sigmoïde $\boldsymbol{\tau} = \{\tau_1, \tau_2\}$ sont fixés et la distribution *a priori* de $\boldsymbol{\theta}$ est définie par :

$$\begin{aligned} p(\boldsymbol{\theta}) &= p(\boldsymbol{\sigma}^2)\; p(\boldsymbol{\theta}_A)\; p(\boldsymbol{\xi})\; p(v_h)\; p(v_l) \\ &= p(v_h)\; p(v_l) \prod_j p(\sigma_j^2) \prod_m [p(\boldsymbol{\theta}_m) p(\xi_m)] \end{aligned} \tag{III.12}$$

où $p(\sigma_j^2) \propto \frac{1}{\sigma_j^2}$, $p(v_h) \propto \frac{1}{v_h}$ et $p(v_l) \propto \frac{1}{v_l}$ sont des *a priori* non informatifs de Jeffrey (Annexe F.3). L'*a priori* sur le paramètre de régularité spatiale du champ de Ising est uniforme $p(\xi_m) = \mathcal{U}_{[0, \xi_m^{\max}]}$ où $\xi_m^{\max} = 2$ car le champ de Ising est totalement corrélé au delà de cette valeur [Vincent 2010a]. Enfin l'*a priori* sur les paramètres du mélange gaussien sur les NRNs est $p(\boldsymbol{\theta}_m) = \mathcal{N}(\mu_1^m; m_{\mu_1^m}, v_{\mu_1^m}) \prod_{i \in \{0,1\}} \mathcal{IG}(v_i^m; \alpha_{v_i^m}, \lambda_{v_i^m})$. La représentation graphique du modèle *parcimonieux* est illustrée par la figure III.1-(b) où un bloc exprimant la pertinence a été ajouté par rapport au graphe représentant le modèle DEC *complet* (figure III.1-(a)). Nous allons par la suite introduire, dans le cadre de l'approche DEC parcimonieuse, l'échantillonneur de Gibbs (Annexe F.7), qui est une méthode MCMC particulière qui a été utilisée avec succès dans de nombreuses analyses bayésiennes.

On rappelle que les variables et paramètres à estimer sont $(\boldsymbol{A}, \boldsymbol{Q}), \boldsymbol{\xi}, \boldsymbol{\theta}_A, \boldsymbol{w}, \boldsymbol{h}, v_h, \boldsymbol{L}, v_l, \boldsymbol{\sigma}^2$, ce qui nécessite la définition des distributions conditionnelles suivantes pour l'échantillonneur de Gibbs :

- Pour chaque voxel j et chaque condition m :
  $p\left(a_j^m, q_j^m \,|\, \boldsymbol{Y}, \boldsymbol{a}_{\setminus j}^m, \boldsymbol{A}^{\setminus m}, \boldsymbol{q}_{\setminus j}^m, \boldsymbol{Q}^{\setminus m}, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{L}, \boldsymbol{\theta}\right)$
- Pour chaque condition m :
  - $p\left(\xi_m \,|\, \boldsymbol{q}^m\right)$
  - $p\left(\boldsymbol{\theta}_m \,|\, \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{Q}, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{L}, \boldsymbol{\theta}_{\setminus \boldsymbol{\theta}_m}\right)$

- $p\left(w^m \,|\, \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{Q}, \boldsymbol{w}^{\setminus m}, \boldsymbol{h}, \boldsymbol{L}, \boldsymbol{\theta}\right)$
- $p\left(\boldsymbol{h} \,|\, \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{Q}, \boldsymbol{w}, \boldsymbol{L}, \boldsymbol{\theta}\right)$
- $p\left(v_h \,|\, \boldsymbol{h}\right)$
- $p\left(\boldsymbol{L} \,|\, \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{Q}, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{\theta}\right)$
- $p\left(v_l \,|\, \boldsymbol{L}\right)$
- $p\left(\boldsymbol{\sigma}^2 \,|\, \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{Q}, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{L}, \boldsymbol{\theta}_{\setminus \boldsymbol{\sigma}^2}\right)$

où $\boldsymbol{x}_{\setminus j} = [\boldsymbol{x}_{j'}; j' = 1 : J \text{ et } j' \neq j]$, $\boldsymbol{x}^{\setminus m} = [\boldsymbol{x}^{m'}; m' = 1 : M \text{ et } m' \neq m]$ et $\boldsymbol{\theta}_{\setminus x}$ contient tous les élements de $\boldsymbol{\theta}$ sauf x. Ci-dessous les formes obtenues pour

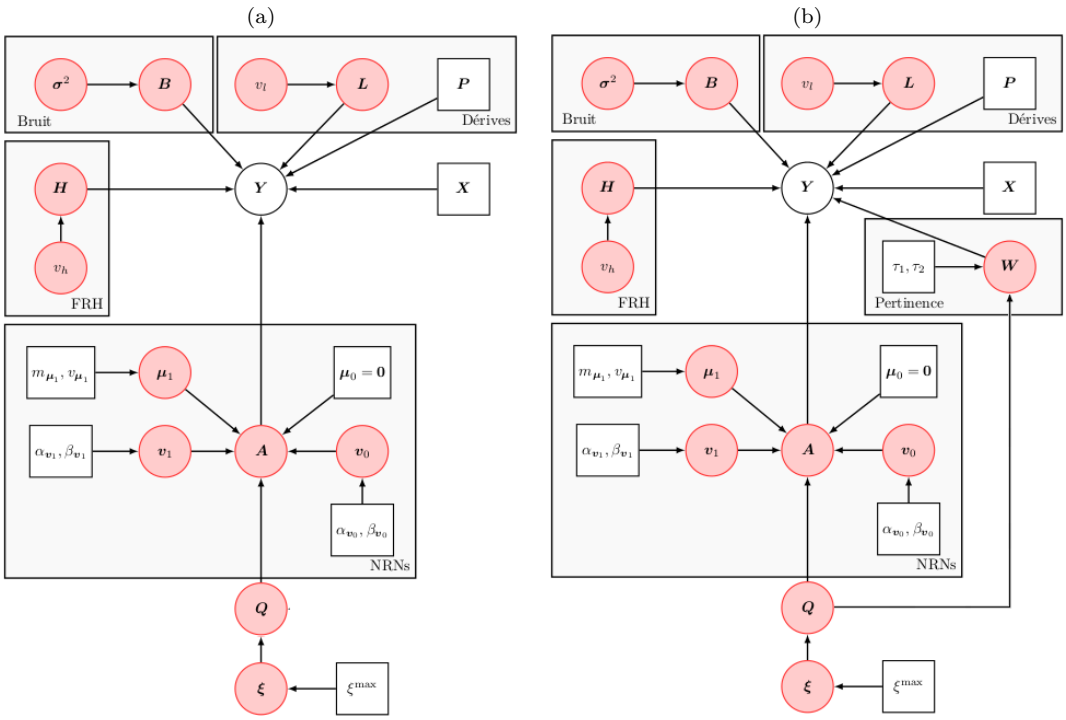


FIGURE III.1 – Modèle graphique illustrant, pour une parcelle $\mathcal{P}_\gamma$, les dépendances entre les variables aléatoires (cercles) et les paramètres (rectangles) du modèle DEC *complet* (a) et *parcimonieux* (b) dans le cadre de l'inférence MCMC : les variables observées Y ainsi que les paramètres fixés sont en blanc et les variables non observées sont en rouge.

chacune de ces distributions et dont le calcul détaillé se trouve dans l'annexe C. Notons que nous garderons, dans le conditionnement, uniquement les termes dépendant de la variable d'intérêt.

♣ Distribution conditionnelle de $\left(a_j^m, q_j^m\right)$

L'idée ici consiste à simuler la distribution conditionnelle du couple de variables $\left(a_j^m, q_j^m\right)$ comme expliqué dans la section F.7, ce qui permettra un meilleur taux

de mélange de la chaîne (Rao Blackwellisation) donc une convergence plus rapide [Dyk 2008, Park 2009]. La loi du couple s'écrit sous la forme :

$$\begin{aligned} &p\left(a_j^m, q_j^m \,|\, \boldsymbol{y}_j, \boldsymbol{a}_j^{\setminus m}, \boldsymbol{w}, \boldsymbol{q}_{\setminus j}^m, \boldsymbol{h}, \boldsymbol{\ell}_j, \sigma_j^2, \boldsymbol{\theta}_m, \xi_m\right) \\ &= p\left(a_j^m | q_j^m, \boldsymbol{y}_j, \boldsymbol{a}_j^{\setminus m}, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{\ell}_j, \sigma_j^2, \boldsymbol{\theta}_m\right) p\left(q_j^m | \boldsymbol{y}_j, \boldsymbol{a}_j^{\setminus m}, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{\ell}_j, \sigma_j^2, \boldsymbol{q}_{\setminus j}^m, \boldsymbol{\theta}_m, \xi_m\right) \end{aligned} \tag{III.13}$$

En faisant passer q_j^m de l'autre côté du conditionnement on obtient la distribution conditionnelle de A_j^m qui s'écrit sous la forme suivante :

$$\begin{aligned} &p\left(a_j^m \,|\, q_j^m, \boldsymbol{y}_j, \boldsymbol{a}_j^{\setminus m}, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{\ell}_j, \sigma_j^2, \boldsymbol{\theta}_m\right) \\ &\qquad = (1 - q_j^m w^m)\, \mathcal{N}(a_j^m; \mu_{0,j}^m, v_{0,j}^m) + q_j^m w^m\, \mathcal{N}(a_j^m; \mu_{1,j}^m, v_{1,j}^m) \end{aligned} \tag{III.14}$$

Les paramètres du mélange étant $v_{i,j}^m = \left(w^m \boldsymbol{h}^{\mathrm{t}} \boldsymbol{X}^{m\mathrm{t}} \boldsymbol{\Gamma}_j \boldsymbol{X}^m \boldsymbol{h} + \frac{1}{v_i^m}\right)^{-1}$ et $\mu_{i,j}^m = v_{i,j}^m \left(w^m \boldsymbol{e}_j^{m\mathrm{t}} \boldsymbol{\Gamma}_j \boldsymbol{X}^m \boldsymbol{h} + \frac{\mu_i^m}{v_i^m}\right)$, où $\boldsymbol{e}_j^m = \boldsymbol{y}_j - \sum\limits_{m' \neq m} a_j^{m'} w^{m'} \boldsymbol{X}^{m'} \boldsymbol{h} - \boldsymbol{P}\boldsymbol{\ell}_j$ et $\boldsymbol{\Gamma}_j = \frac{\Lambda_j}{\sigma_j^2}$ la matrice de variance-covariance du bruit avec $\Lambda_j = \boldsymbol{I}_N$ dans le cas d'un bruit blanc gaussien.

La distribution conditionnelle marginale de q_j^m est quant à elle obtenue en intégrant la loi du couple (a_j^m, q_j^m) (équation (III.13)) par rapport à a_j^m :

$$\begin{aligned} &p\left(q_j^m \,|\, \boldsymbol{y}_j, \boldsymbol{a}_j^{\setminus m}, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{\ell}_j, \sigma_j^2, \boldsymbol{q}_{\setminus j}^m, \boldsymbol{\theta}_m, \xi_m\right) \\ &\qquad = \int\limits_{a_j^m} p\left(a_j^m, q_j^m \,|\, \boldsymbol{y}_j, \boldsymbol{a}_j^{\setminus m}, \boldsymbol{w}, \boldsymbol{q}_{\setminus j}^m, \boldsymbol{h}, \boldsymbol{\ell}_j, \sigma_j^2, \boldsymbol{\theta}_m, \xi_m\right) da_j^m \end{aligned}$$

La probabilité qu'un voxel soit considéré comme activé conditionnellement aux données ainsi qu'aux variables et paramètres du modèle est donc exprimée par :

$$\begin{aligned} &p\left(q_j^m = 1 \,|\, \boldsymbol{y}_j, \boldsymbol{a}_j^{\setminus m}, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{\ell}_j, \sigma_j^2, \boldsymbol{q}_{\setminus j}^m, \boldsymbol{\theta}_m, \xi_m\right) \\ &= \left[1 + \frac{\frac{p(q_j^m=0 \,|\, \boldsymbol{q}_{\setminus j}^m, \xi_m)\, p(w^m \,|\, q_j^m=0, \boldsymbol{q}_{\setminus j}^m, \boldsymbol{\tau})}{p(q_j^m=1 \,|\, \boldsymbol{q}_{\setminus j}^m, \xi_m)\, p(w^m \,|\, q_j^m=1, \boldsymbol{q}_{\setminus j}^m, \boldsymbol{\tau})}}{1 - w^m \left\{1 - \sqrt{\frac{v_{1,j}^m v_0^m}{v_{0,j}^m v_1^m}} \exp\left(-\frac{1}{2}\left[\frac{(\mu_1^m)^2}{v_1^m} - \frac{(\mu_0^m)^2}{v_0^m} - \frac{(\mu_{1,j}^m)^2}{v_{1,j}^m} + \frac{(\mu_{0,j}^m)^2}{v_{0,j}^m}\right]\right)\right\}}\right]^{-1} \end{aligned} \tag{III.15}$$

♣ Distribution conditionnelle de w^m

Le calcul détaillé de cette distribution nous permet d'obtenir l'équation suivante exprimant la probabilité qu'une condition m soit considérée comme pertinente

conditionnellement aux données ainsi qu'aux variables et paramètres du modèle :

$$\begin{aligned} & p(w^m = 1 \,|\, \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{w}^{\setminus m}, \boldsymbol{q}^m, \boldsymbol{h}, \boldsymbol{L}, \boldsymbol{\sigma}^2, \boldsymbol{\theta}_m, \boldsymbol{\tau}) \\ & = \left(1 + \frac{\frac{p(w^m=0\,|\,\boldsymbol{q}^m,\boldsymbol{\tau})}{p(w^m=1\,|\,\boldsymbol{q}^m,\boldsymbol{\tau})} \left(\frac{v_0^m}{v_1^m}\right)^{-\frac{J_1^m}{2}} \exp\left(-\sum_{j\in J_1^m} \left[\frac{\left(a_j^m-\mu_0^m\right)^2}{2v_0^m} - \frac{\left(a_j^m-\mu_1^m\right)^2}{2v_1^m} \right] \right)}{\exp\left[-\sum_{j\in J} \frac{1}{2} \left(-2a_j^m \, \boldsymbol{e}_j^{m\mathrm{t}} \boldsymbol{\Gamma}_j \boldsymbol{X}^m \boldsymbol{h} + \left(a_j^m\right)^2 \boldsymbol{h}^{\mathrm{t}} \boldsymbol{X}^{m\mathrm{t}} \boldsymbol{\Gamma}_j \boldsymbol{X}^m \boldsymbol{h} \right) \right]} \right)^{-1} \end{aligned} \tag{III.16}$$

♣ Distribution conditionnelle de $\boldsymbol{h}_\gamma$

La vraisemblance ainsi que l'*a priori* sur $\boldsymbol{h}$ étant gaussiens, on peut montrer que la distribution *a posteriori* conditionnelle de $\boldsymbol{h}_\gamma$ est également gaussienne :

$$p\left(\boldsymbol{h} \,|\, \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{w}, \boldsymbol{L}, \boldsymbol{\sigma}^2, v_h\right) = \mathcal{N}\left(\boldsymbol{h}; \boldsymbol{\mu}_{\boldsymbol{h}}, \boldsymbol{\Sigma}_{\boldsymbol{h}}\right) \tag{III.17}$$

avec $\boldsymbol{\Sigma}_{\boldsymbol{h}} = \left(\frac{\boldsymbol{R}^{-1}}{v_h} + \sum_j \boldsymbol{S}_j^{\mathrm{t}} \boldsymbol{\Gamma}_j \boldsymbol{S}_j \right)^{-1}$ et $\boldsymbol{\mu}_{\boldsymbol{h}} = \boldsymbol{\Sigma}_{\boldsymbol{h}} \sum_j \boldsymbol{S}_j^{\mathrm{t}} \boldsymbol{\Gamma}_j (\boldsymbol{y}_j - \boldsymbol{P}\boldsymbol{\ell}_j)$. On rappelle que la matrice de covariance $\boldsymbol{R} = \left(\frac{\boldsymbol{D}_2^{\mathrm{t}} \boldsymbol{D}_2}{(\Delta t)^4} \right)^{-1}$ (Annexe A) exprime une contrainte sur la dérivée seconde afin de pénaliser les fortes variations de pente et obtenir des FRH lisses plus plausibles physiologiquement.

♣ Distribution conditionnelle de v_h

En revenant aux distributions *a priori* sur $\boldsymbol{h}_\gamma$ et v_h on peut facilement conclure que la loi conditionnelle de v_h est une loi inverse Gamma :

$$p\left(v_h \,|\, \boldsymbol{h}\right) = \mathcal{IG}(v_h; \alpha_{v_h}, \lambda_{v_h}) \tag{III.18}$$

avec $\alpha_{v_h} = \frac{D-1}{2}$ et $\lambda_{v_h} = \frac{1}{2} \, \boldsymbol{h}^{\mathrm{t}} \boldsymbol{R}^{-1} \boldsymbol{h}$.

♣ Distribution conditionnelle de $\boldsymbol{\sigma}^2$

Ce terme dépend uniquement de la vraisemblance ainsi que de l'*a priori* sur $\boldsymbol{\sigma}^2$ à partir desquelles la distribution conditionnelle obtenue pour chaque paramètre σ_j^2 est une loi inverse Gamma, où l'indépendance voxellique est due à celle présente

dans la vraisemblance (équation (II.23)) et l'*a priori* (équation (III.12)) :

$$p\left(\boldsymbol{\sigma}^2 \,|\, \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{L}\right) = \prod_j \mathcal{IG}(\sigma_j^2; \alpha_{\sigma_j^2}, \lambda_{\sigma_j^2}) \tag{III.19}$$

avec $\alpha_{\sigma_j^2} = \frac{N}{2}$, $\lambda_{\sigma_j^2} = \frac{\widetilde{\boldsymbol{y}}_j^{\mathrm{t}} \boldsymbol{\Lambda}_j \widetilde{\boldsymbol{y}}_j}{2}$ et $\boldsymbol{\Gamma}_j = \frac{\boldsymbol{\Lambda}_j}{\sigma_j^2}$.

♣ Distribution conditionnelle de $\boldsymbol{\theta}_m$

Les *a priori* précédemment définis sur $\boldsymbol{\theta}_m = \{\mu_1^m, v_1^m, v_0^m\}$ étant conjugués (gaussienne pour μ_1^m et inverse Gamma pour v_1^m et v_0^m), on en déduit des distributions conditionnelles de même nature avec des paramètres différents.

- Moyenne de la classe active μ_1^m

$$\begin{aligned} &p\left(\mu_1^m \,|\, \boldsymbol{a}^m, w^m, \boldsymbol{q}^m, v_1^m\right) \\ &\quad = (1 - w^m)\, \mathcal{N}\left(\mu_1^m; m_{\mu_1^m}, v_{\mu_1^m}\right) + w^m\, \mathcal{N}\left(\mu_1^m; M_{\mu_1^m}, V_{\mu_1^m}\right) \end{aligned} \tag{III.20}$$

avec $V_{\mu_1^m} = \left(\frac{1}{v_{\mu_1^m}} + \frac{J_1^m}{v_1^m}\right)^{-1}$ et $M_{\mu_1^m} = V_{\mu_1^m}\left(\frac{m_{\mu_1^m}}{v_{\mu_1^m}} + \frac{\sum_{j \in J_1^m} a_j^m}{v_1^m}\right)$. Rappelons que $m_{\mu_1^m}$ et $v_{\mu_1^m}$ représentent les moyennes et variances de la loi *a priori* sur μ_1^m.

- Variance de la classe inactive v_0^m :

$$\begin{aligned} &p\left(v_0^m \,|\, \boldsymbol{a}^m, w^m, \boldsymbol{q}^m, \mu_0^m\right) \\ &\quad = (1 - w^m)\, \mathcal{IG}(v_0^m; A_{0,v_0^m}, B_{0,v_0^m}) + w^m\, \mathcal{IG}(v_0^m; A_{1,v_0^m}, B_{1,v_0^m}) \end{aligned} \tag{III.21}$$

avec $A_{0,v_0^m} = \alpha_{v_0^m} + \frac{J}{2}$, $B_{0,v_0^m} = \beta_{v_0^m} + \sum_{j \in J} \frac{\left(a_j^m - \mu_0^m\right)^2}{2}$, $A_{1,v_0^m} = \alpha_{v_0^m} + \frac{J_0^m}{2}$ et $B_{1,v_0^m} = \beta_{v_0^m} + \sum_{j \in J_0^m} \frac{\left(a_j^m - \mu_0^m\right)^2}{2}$, où J_0^m désigne l'ensemble des voxels de la classe inactive ($i = 0$) correspondant à la $m$ième condition dans la parcelle $\mathcal{P}_\gamma$. Comme désigné précédemment, $\alpha_{v_0^m}$ et $\beta_{v_0^m}$, sont les hyperparamètres de la loi *a priori* sur v_0^m.

- Variance de la classe active v_1^m :

$$\begin{aligned} &p\left(v_1^m \,|\, \boldsymbol{a}^m, w^m, \boldsymbol{q}^m, \mu_1^m\right) \\ &\quad = (1 - w^m)\, \mathcal{IG}(v_1^m; \alpha_{v_1^m}, \beta_{v_1^m}) + w^m\, \mathcal{IG}(v_1^m; A_{1,v_1^m}, B_{1,v_1^m}) \end{aligned} \tag{III.22}$$

avec $A_{1,v_1^m} = \alpha_{v_1^m} + \frac{J_1^m}{2}$, $B_{1,v_1^m} = \beta_{v_1^m} + \sum_{j \in J_1^m} \frac{\left(a_j^m - \mu_1^m\right)^2}{2}$, où J_1^m désigne l'ensemble des

voxels de la classe active ($i = 1$) correspondant à la $m$ième condition dans la parcelle $\mathcal{P}_\gamma$. Rappelons que, $\alpha_{v_1^m}$ et $\beta_{v_1^m}$, sont les hyperparamètres de la loi *a priori* sur v_1^m.

Notons que les hyperparamètres des lois *a priori* sont choisis de manière à obtenir des *a priori* non informatifs.

♣ Distribution conditionnelle de $\boldsymbol{L}$

Etant donné l'indépendance spatiale de la vraisemblance ainsi que de l'*a priori* sur $\boldsymbol{L}$ et tenant compte de la nature gaussienne des lois voxelliques correspondantes, la distribution conditionnelle de chaque $\boldsymbol{\ell}_j$ est gaussienne :

$$p\left(\boldsymbol{L} \,|\, \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{\sigma}^2, v_\ell\right) = \prod_j \mathcal{N}(\boldsymbol{\ell}_j; \boldsymbol{\mu}_{\boldsymbol{\ell}_j}, \boldsymbol{\Sigma}_{\boldsymbol{\ell}_j}) \tag{III.23}$$

avec $\boldsymbol{\Sigma}_{\boldsymbol{\ell}_j} = \left(\frac{\boldsymbol{I}_O}{v_\ell} + \boldsymbol{P}^{\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{P}\right)^{-1}$ et $\boldsymbol{\mu}_{\boldsymbol{\ell}_j} = \boldsymbol{\Sigma}_{\boldsymbol{\ell}_j}\boldsymbol{P}^{\mathrm{t}}\boldsymbol{\Gamma}_j\left(\boldsymbol{y}_j - \boldsymbol{S}_j\boldsymbol{h}\right)$.

♣ Distribution conditionnelle de v_ℓ

Comme pour v_h la loi conditionnelle de v_ℓ ne dépend que de son *a priori* ainsi que de celui de $\boldsymbol{L}$, desquels on déduit facilement la loi inverse Gamma suivante :

$$p(v_\ell \,|\, \boldsymbol{L}) = \mathcal{IG}(v_\ell; \alpha_{v_\ell}, \lambda_{v_\ell}) \tag{III.24}$$

avec $\alpha_{v_\ell} = \frac{OJ}{2}$ et $\lambda_{v_\ell} = \sum_j \frac{||\boldsymbol{\ell}_j||^2}{2}$.

♣ Distribution conditionnelle de $\boldsymbol{\xi}$

L'indépendance inter-conditions des *a priori* définis sur les labels $\boldsymbol{Q}$ et les paramètres $\boldsymbol{\xi}$ est conservée dans la loi conditionnelle de ce dernier, ce qui donne :

$$p\left(\boldsymbol{\xi} \,|\, \boldsymbol{Q}\right) \propto \prod_{m=1}^{M} \frac{1}{Z\left(\xi_m\right)} \exp\left[\xi_m U\left(\boldsymbol{q}^m\right)\right] \mathcal{U}_{[0,\xi_m^{\max}]}(\xi_m) \tag{III.25}$$

L'estimation de ξ_m est réalisée à travers une étape de Metropolis-Hasting (Annexe C) [Pereyra 2013, Vincent 2010a].

Calcul de la fonction de partition $Z(\xi_m)$. Le calcul de la constante de normal-

isation du champ de Ising $Z(\xi_m) = \sum_{\boldsymbol{q}^m} \exp\left[\xi_m U(\boldsymbol{q}^m)\right]$ nécessite la prise en compte de toutes les configurations possibles de $\boldsymbol{q}^m$ qui s'élèvent à 2^J et qui atteignent déjà un millier de configurations avec seulement une dizaine de voxels. Son évaluation à chaque itération de l'algorithme demeure trop coûteuse en temps de calcul mais elle demeure indépendante des données. Elle peut donc être tabulée à l'avance à travers une procédure de *path sampling* [Vincent 2010a] sur une grille discrète dense de valeurs de ξ ($\xi^g = g\xi^{\max}/G; g = \{0, 1.., G\}$ avec $G = 10^3$). Une interpolation linéaire est ensuite effectuée pour les valeurs de ξ en dehors de la grille.

L'estimation, préalable à l'échantillonnage, de la fonction de partition $Z(\xi)$ du champ de Ising ainsi que l'échantillonneur de Gibbs sont décrits par les tables III.1 et III.2 respectivement.

a-	Estimation de la fonction de partition $Z(\xi)$ sur une grille de valeur de ξ.
b-	Interpolation linéaire de $Z(\xi)$ en dehors des points de la grille.

Table III.1 – Étapes de l'estimation préalable de la fonction de partition du champs de Ising

a- Initialisation $\boldsymbol{A}^{(0)}, \boldsymbol{Q}^{(0)}, \boldsymbol{w}^{(0)}, \boldsymbol{\theta}_A^{(0)}, \boldsymbol{\xi}^{(0)}, \boldsymbol{h}^{(0)}, v_h^{(0)}, \boldsymbol{L}^{(0)}, v_\ell^{(0)}, \boldsymbol{\sigma}^{2(0)}$.

b- Itération r :

1- NRNs : pour chaque condition m et chaque voxel j :
- $q_j^{m(r)} \sim p\left(q_j^m \,|\, \boldsymbol{y}_j, \boldsymbol{a}_j^{\setminus m(r-1)}, \boldsymbol{w}^{(r-1)}, \boldsymbol{h}^{(r-1)}, \boldsymbol{\ell}_j^{(r-1)}, \sigma_j^{2(r-1)}, \boldsymbol{q}_{\setminus j}^{m(r-1)}, \boldsymbol{\theta}_m^{(r-1)}, \xi_m^{(r-1)}\right)$
- $a_j^{m(r)} \sim p\left(a_j^m \,|\, q_j^{m(r)}, \boldsymbol{y}_j, \boldsymbol{a}_j^{\setminus m(r-1)}, \boldsymbol{w}^{(r-1)}, \boldsymbol{h}^{(r-1)}, \boldsymbol{\ell}_j^{(r-1)}, \sigma_j^{2(r-1)}, \boldsymbol{\theta}_m^{(r-1)}\right)$

2- Paramètres de régularité spatiale du champ de Ising : $\boldsymbol{\xi}^{(r)} \sim p\left(\boldsymbol{\xi} \,|\, \boldsymbol{Q}^{(r)}\right)$

3- Paramètres du mélange $\boldsymbol{\theta}_A$: pour chaque condition $m = 1$ à M :
- Moyenne des NRNs activés :
  $\mu_1^{m(r)} \sim p(\mu_1^m \,|\, \boldsymbol{a}^{m(r)}, w^{m(r-1)}, \boldsymbol{q}^{m(r)}, v_1^{m(r-1)})$
- Variance des NRNs activés :
  $v_1^{m(r)} \sim p(v_1^m \,|\, \boldsymbol{a}^{m(r)}, w^{m(r-1)}, \boldsymbol{q}^{m(r)}, \mu_1^{m(r)})$
- Variance des NRNs inactivés :
  $v_0^{m(r)} \sim p(v_0^m \,|\, \boldsymbol{a}^{m(r)}, w^{m(r-1)}, \boldsymbol{q}^{m(r)}, \mu_0^m = 0)$

4- variable de pertinence : pour chaque condition $m = 1$ à M,
$$w^{m(r)} \sim p\left(w^m \,|\, \boldsymbol{Y}, \boldsymbol{A}^{(r)}, \boldsymbol{w}_\gamma^{\setminus m(r-1)}, \boldsymbol{q}^{m(r)}, \boldsymbol{h}^{(r-1)}, \boldsymbol{\sigma}^{2(r-1)}, \boldsymbol{\theta}_m^{(r)}, \boldsymbol{\tau}\right)$$

5- FRH : $\boldsymbol{h}^{(r)} \sim p\left(\boldsymbol{h} \,|\, \boldsymbol{Y}, \boldsymbol{A}^{(r)}, \boldsymbol{w}^{(r)}, \boldsymbol{L}^{(r-1)}, \boldsymbol{\sigma}^{2(r-1)}, v_h^{(r-1)}\right)$

6- Variance de la FRH : $v_h^{(r)} \sim p\left(v_h \,|\, \boldsymbol{h}^{(r)}\right)$

7- Coefficients des dérives basses fréquences :
$\boldsymbol{L}^{(r)} \sim p\left(\boldsymbol{L} \,|\, \boldsymbol{Y}, \boldsymbol{A}^{(r)}, \boldsymbol{w}^{(r)}, \boldsymbol{h}^{(r)}, \boldsymbol{\sigma}^{2(r-1)}, v_\ell^{(r-1)}\right)$

8- Variance des coefficients $\boldsymbol{L}$: $v_\ell^{(r)} \sim p\left(v_\ell \,|\, \boldsymbol{L}^{(r)}\right)$

9- Variance du bruit : $\boldsymbol{\sigma}^{2(r)} \sim p\left(\boldsymbol{\sigma}^2 \,|\, \boldsymbol{Y}, \boldsymbol{A}^{(r)}, \boldsymbol{w}^{(r)}, \boldsymbol{h}^{(r)}, \boldsymbol{L}^{(r)}\right)$

c- Après R itérations à la suite d'une periode de chauffe de R_0 itérations :
- On estime les quantités d'intérêt par la moyenne *a posteriori* :
  $\forall x \in \{\boldsymbol{A}, \boldsymbol{h}, \boldsymbol{L}, \boldsymbol{\theta}\}$, $\hat{x} = \frac{1}{R-R_0} \sum_{r=R_0+1}^{R} x^{(r)}$
- Classification des voxels et conditions par maximum *a posteriori*
  $\hat{q}_j^m \approx \arg\max_i \frac{1}{R-R_0} \sum_{r=R_0+1}^{R} I(q_j^{m(r)} = i)$
  $\hat{w}_\gamma^m \approx \arg\max_i \frac{1}{R-R_0} \sum_{r=R_0+1}^{R} I(w^{m(r)} = i)$

TABLE III.2 – Étapes de l'échantillonnage de Gibbs pour le modèle *parcimonieux*

Les travaux de Lotfi Chaari [Chaari 2013] comparent les inférences MCMC et VEM dans le cadre du modèle DEC *complet* et mettent en évidence le gain obtenu en temps de calcul dans le cas de l'algorithme VEM comme le montre la figure III.2. L'étude montre que la différence en temps de calcul $\Delta t = t_{\text{MCMC}} - t_{\text{VEM}}$, en faveur de VEM, augmente avec la taille de la parcelle et est plus élevée pour les parcelles non activées (faible SNR) (en bleu) (figure III.2-(a)). De même la figure III.2-(b) illustre le gain en temps de calcul $G_f = t_{\text{MCMC}}/t_{\text{VEM}}$ supérieur à 1 (ligne horizontale) pour toutes les parcelles ($G_f \in [2.7, 80]$). Comme pour Δt, ce gain est plus élevé dans le cas des parcelles non activées. Nous proposons donc

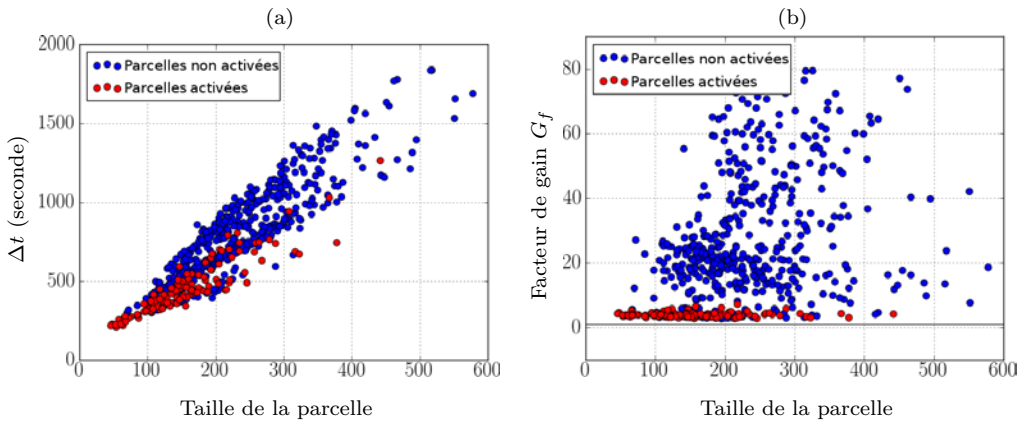


FIGURE III.2 – Comparaison des algorithmes MCMC et VEM par rapport au temps de calcul : (a) différence en temps de calcul $\Delta t = t_{\text{MCMC}} - t_{\text{VEM}}$, (b) gain en temps de calcul $G_f = t_{\text{MCMC}}/t_{\text{VEM}}$ supérieur à 1 (ligne horizontale). Les deux critères, confirment l'efficacité de VEM par rapport à MCMC. Ils augmentent avec la taille de la parcelle et sont plus élevés pour les parcelles non activées (en bleu) par rapport à celles activées (en rouge).

dans la prochaine partie d'estimer les paramètres de notre modèle *parcimonieux* à l'aide d'une inférence effectuée via un algorithme VEM.

III.2.2.2 Algorithme EM Variationnel - VEM

L'algorithme EM (*Expectation-Maximization*) est généralement utilisé dans les problèmes à données manquantes, qui sont dans le cadre de l'approche DEC $\boldsymbol{A} \in \mathcal{A}, \boldsymbol{Q} \in \mathcal{Q}, \boldsymbol{w} \in \mathcal{W}$ et $\boldsymbol{h} \in \mathcal{H}$. Notons $\mathcal{D}$ l'espace de toutes les distributions de probabilité sur $\mathcal{A} \times \mathcal{Q} \times \mathcal{W} \times \mathcal{H}$. L'algorithme EM peut être vu comme une procédure de maximisation alternée d'une fonction $\mathcal{F}$, appelée énergie libre, telle

que pour toute distribution $\widetilde{p} \in \mathcal{D}$:

$$F(\widetilde{p}, \boldsymbol{\theta}) = \mathbb{E}_{\widetilde{p}}[\log p(\boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{Q}, \boldsymbol{w}, \boldsymbol{h} \,|\, \boldsymbol{\Theta})] + I[\widetilde{p}] \tag{III.26}$$

où $\boldsymbol{Y}$ sont les données observées, $I[\widetilde{p}] = -\mathbb{E}_{\widetilde{p}}[\log \widetilde{p}(\boldsymbol{A}, \boldsymbol{Q}, \boldsymbol{w}, \boldsymbol{h})]$ est l'entropie de $\widetilde{p}$ et $\mathbb{E}_{\widetilde{p}}[.]$ désigne l'espérance par rapport à $\widetilde{p}$. Notons les valeurs courantes des paramètres du modèle $\boldsymbol{\theta}^{(r)}$, l'algorithme EM procède comme suit :

$$\textbf{E :} \quad \widetilde{p}^{(r)} = \arg\max_{\widetilde{p} \in \mathcal{D}} F(\widetilde{p}, \boldsymbol{\theta}^{(r)}) \tag{III.27}$$

$$\textbf{M :} \quad \boldsymbol{\theta}^{(r+1)} = \arg\max_{\boldsymbol{\theta} \in \Theta} F(\widetilde{p}^{(r)}, \boldsymbol{\theta}) \tag{III.28}$$

La complexité de la distribution $\widetilde{p}(\boldsymbol{A}, \boldsymbol{Q}, \boldsymbol{w}, \boldsymbol{h}) = p(\boldsymbol{A}, \boldsymbol{Q}, \boldsymbol{w}, \boldsymbol{h}; \boldsymbol{\theta}^{(r)} \,|\, \boldsymbol{Y})$ (solution de l'étape **E**) dans notre cas, rend l'étape **M** difficile. Pour ceci nous proposons d'utiliser une variante de cet algorithme dans laquelle l'étape E est effectuée sur une classe de distributions de probabilité plus restreinte $\widetilde{\mathcal{D}}$ regroupant les distributions $\widetilde{p}$ telles que $\widetilde{p}(\boldsymbol{A}, \boldsymbol{Q}, \boldsymbol{w}, \boldsymbol{h}) = \widetilde{p}_{\boldsymbol{A}}(\boldsymbol{A})\ \widetilde{p}_{\boldsymbol{Q}}(\boldsymbol{Q})\ \widetilde{p}_{\boldsymbol{w}}(\boldsymbol{w})\ \widetilde{p}_{\boldsymbol{h}}(\boldsymbol{h})$ où $\widetilde{p}_{\boldsymbol{A}} \in \mathcal{D}_{\mathcal{A}}$, $\widetilde{p}_{\boldsymbol{Q}} \in \mathcal{D}_{\mathcal{Q}}$, $\widetilde{p}_{\boldsymbol{w}} \in \mathcal{D}_{\mathcal{W}}$ et $\widetilde{p}_{\boldsymbol{h}} \in \mathcal{D}_{\mathcal{H}}$ représentent les distributions de probabilité sur $\boldsymbol{A}, \boldsymbol{Q}, \boldsymbol{w}$ et $\boldsymbol{h}$ respectivement. Notons que pour des raisons de complexité de calcul nous supposons aussi que $\widetilde{p}_{\boldsymbol{Q}}(\boldsymbol{Q}) = \prod_m \prod_j \widetilde{p}_{q_j^m}(q_j^m)$ et que $\widetilde{p}_{\boldsymbol{w}}(\boldsymbol{w}) = \prod_m \widetilde{p}_{w^m}(w^m)$. L'étape E précédente peut donc être décomposée en 4 étapes. A l'itération (r), avec les estimations courantes notées $\widetilde{p}_{\boldsymbol{A}}^{(r-1)}, \widetilde{p}_{\boldsymbol{Q}}^{(r-1)}, \widetilde{p}_{\boldsymbol{w}}^{(r-1)}$ et $\theta^{(r)}$ on a :

$$\textbf{E-H :} \quad \widetilde{p}_{\boldsymbol{h}}^{(r)} = \arg\max_{\widetilde{p}_{\boldsymbol{h}} \in \mathcal{D}_{\mathcal{H}}} F\left(\widetilde{p}_{\boldsymbol{A}}^{(r-1)} \widetilde{p}_{\boldsymbol{Q}}^{(r-1)} \widetilde{p}_{\boldsymbol{w}}^{(r-1)} \widetilde{p}_{\boldsymbol{h}}, \boldsymbol{\theta}^{(r)}\right) \tag{III.29}$$

$$\textbf{E-A :} \quad \widetilde{p}_{\boldsymbol{A}}^{(r)} = \arg\max_{\widetilde{p}_{\boldsymbol{A}} \in \mathcal{D}_{\mathcal{A}}} F\left(\widetilde{p}_{\boldsymbol{A}} \widetilde{p}_{\boldsymbol{Q}}^{(r-1)} \widetilde{p}_{\boldsymbol{w}}^{(r-1)} \widetilde{p}_{\boldsymbol{h}}^{(r)}, \boldsymbol{\theta}^{(r)}\right) \tag{III.30}$$

$$\textbf{E-W :} \quad \widetilde{p}_{\boldsymbol{w}}^{(r)} = \arg\max_{\widetilde{p}_{\boldsymbol{w}} \in \mathcal{D}_{\mathcal{W}}} F\left(\widetilde{p}_{\boldsymbol{A}}^{(r)} \widetilde{p}_{\boldsymbol{Q}}^{(r-1)} \widetilde{p}_{\boldsymbol{w}} \widetilde{p}_{\boldsymbol{h}}^{(r)}, \boldsymbol{\theta}^{(r)}\right) \tag{III.31}$$

$$\textbf{E-Q :} \quad \widetilde{p}_{\boldsymbol{Q}}^{(r)} = \arg\max_{\widetilde{p}_{\boldsymbol{Q}} \in \mathcal{D}_{\mathcal{Q}}} F\left(\widetilde{p}_{\boldsymbol{A}}^{(r)} \widetilde{p}_{\boldsymbol{Q}} \widetilde{p}_{\boldsymbol{w}}^{(r)} \widetilde{p}_{\boldsymbol{h}}^{(r)}, \boldsymbol{\theta}^{(r)}\right). \tag{III.32}$$

Les deux dernières étapes sont divisées en M et $M \times J$ sous-étapes respectivement :

pour $m = 1 : M$

$$\textbf{E-}W^m \textbf{ :} \quad \widetilde{p}_{w^m}^{(r)} = \arg\max_{\widetilde{p}_{w^m} \in \mathcal{D}_{\mathcal{W}^m}} F\left(\widetilde{p}_{\boldsymbol{A}}^{(r)} \widetilde{p}_{\boldsymbol{Q}}^{(r-1)} \widetilde{p}_{\boldsymbol{w}^{\setminus m}}^{(r)} \widetilde{p}_{w^m} \widetilde{p}_{\boldsymbol{h}}^{(r)}, \boldsymbol{\theta}^{(r)}\right) \tag{III.33}$$

pour $m = 1 : M, j = 1 : J$

$$\textbf{E-}Q_j^m \textbf{ :} \quad \widetilde{p}_{q_j^m}^{(r)} = \arg\max_{\widetilde{p}_{q_j^m} \in \mathcal{D}_{\mathcal{Q}_j^m}} F\left(\widetilde{p}_{\boldsymbol{A}}^{(r)} \widetilde{p}_{\boldsymbol{Q}_{\setminus j}}^{(r)} \widetilde{p}_{\boldsymbol{q}_j^{\setminus m}}^{(r)} \widetilde{p}_{q_j^m} \widetilde{p}_{\boldsymbol{w}}^{(r)} \widetilde{p}_{\boldsymbol{h}}^{(r)}, \boldsymbol{\theta}^{(r)}\right) \tag{III.34}$$

où pour alléger la notation nous avons noté :

$$\widetilde{p}^{(r)}_{\boldsymbol{w}^{\setminus m}} = \prod_{m'=1}^{m-1} \widetilde{p}^{(r)}_{\boldsymbol{w}^{m'}} \prod_{m'=m+1}^{M} \widetilde{p}^{(r-1)}_{\boldsymbol{w}^{m'}}$$
$$\widetilde{p}^{(r)}_{\boldsymbol{Q}^{\setminus j}} = \prod_{j'=1}^{j-1} \widetilde{p}^{(r)}_{\boldsymbol{Q}^{\setminus j}} \prod_{j'=j+1}^{J} \widetilde{p}^{(r-1)}_{\boldsymbol{Q}^{\setminus j}}$$
$$\widetilde{p}^{(r)}_{\boldsymbol{q}_j^{\setminus m}} = \prod_{m'=1}^{m-1} \widetilde{p}^{(r)}_{\boldsymbol{q}_j^{\setminus m}} \prod_{m'=m+1}^{M} \widetilde{p}^{(r-1)}_{\boldsymbol{q}_j^{\setminus m}}$$

La fonction F peut être aussi reliée à la distance de Kullback-Leibler D entre la loi recherchée $p(\boldsymbol{A}, \boldsymbol{Q}, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{\theta} \,|\, \boldsymbol{Y})$, désignée comme p, et son approximation $\widetilde{p}$. Cette distance est définie selon l'équation suivante :

$$F(\widetilde{p}, \boldsymbol{\theta}) = L(\boldsymbol{\theta}) - D(\widetilde{p}||p) \tag{III.35}$$

où $L(\boldsymbol{\theta}) = \log p(y \,|\, \boldsymbol{\theta})$ représente le log vraisemblance. Maximiser F revient donc à minimiser $D(\widetilde{p}||p)$ qui est définie, pour $\widetilde{p}(\boldsymbol{A}, \boldsymbol{Q}, \boldsymbol{w}, \boldsymbol{h}) = \widetilde{p}_{\boldsymbol{A}}(\boldsymbol{A})\, \widetilde{p}_{\boldsymbol{Q},\boldsymbol{w},\boldsymbol{h}}(\boldsymbol{Q}, \boldsymbol{w}, \boldsymbol{h})$, selon l'équation :

$$D(\widetilde{p}||p) = \int \widetilde{p}_{\boldsymbol{A}}(\boldsymbol{A}) \widetilde{p}_{\boldsymbol{Q},\boldsymbol{w},\boldsymbol{h}}(\boldsymbol{Q}, \boldsymbol{w}, \boldsymbol{h}) \log \frac{\widetilde{p}_{\boldsymbol{A}}(\boldsymbol{A}) \widetilde{p}_{\boldsymbol{Q},\boldsymbol{w},\boldsymbol{h}}(\boldsymbol{Q}, \boldsymbol{w}, \boldsymbol{h})}{p(\boldsymbol{A}, \boldsymbol{Q}, \boldsymbol{w}, \boldsymbol{h}; \boldsymbol{\theta} \,|\, \boldsymbol{Y})} d\boldsymbol{A}\, d\boldsymbol{Q}\, d\boldsymbol{w}\, d\boldsymbol{h}. \tag{III.36}$$

D'après les propriétés de la distance de Kullback-Leibler et comme détaillé dans [Chaari 2013, Annexe A], les étapes **E** précédentes s'écrivent de la manière sui-

vante :

$$\textbf{E-H :}\ \ \widetilde{p}_{\boldsymbol{h}}^{(r)}(\boldsymbol{h}) \propto \exp\left(\mathbb{E}_{\widetilde{p}_{\boldsymbol{A}}^{(r-1)}\widetilde{p}_{\boldsymbol{Q}}^{(r-1)}\widetilde{p}_{\boldsymbol{w}}^{(r-1)}}\left[\log p\left(\boldsymbol{h}\,|\,\boldsymbol{Y},\boldsymbol{A},\boldsymbol{Q},\boldsymbol{w};\boldsymbol{\theta}^{(r)}\right)\right]\right) \tag{III.37}$$

$$\textbf{E-A :}\ \ \widetilde{p}_{\boldsymbol{A}}^{(r)}(\boldsymbol{A}) \propto \exp\left(\mathbb{E}_{\widetilde{p}_{\boldsymbol{Q}}^{(r-1)}\widetilde{p}_{\boldsymbol{w}}^{(r-1)}\widetilde{p}_{\boldsymbol{h}}^{(r)}}\left[\log p\left(\boldsymbol{A}\,|\,\boldsymbol{Y},\boldsymbol{Q},\boldsymbol{w},\boldsymbol{h};\boldsymbol{\theta}^{(r)}\right)\right]\right) \tag{III.38}$$

pour $m = 1 : M$

$$\textbf{E-}W^m\textbf{ :}\ \ \widetilde{p}_{w^m}^{(r)}(w^m) \propto \exp\left(\mathbb{E}_{\widetilde{p}_{\boldsymbol{A}}^{(r)}\widetilde{p}_{\boldsymbol{Q}}^{(r-1)}\widetilde{p}_{\boldsymbol{w}^{\setminus m}}^{(r-1)}\widetilde{p}_{\boldsymbol{h}}^{(r)}}\left[\log p\left(w^m\,|\,\boldsymbol{Y},\boldsymbol{A},\boldsymbol{Q},\boldsymbol{w}^{\setminus m},\boldsymbol{h};\boldsymbol{\theta}^{(r)}\right)\right]\right) \tag{III.39}$$

pour $m = 1 : M, j = 1 : J$

$$\textbf{E-}Q_j^m\textbf{ :}\ \ \widetilde{p}_{q_j^m}^{(r)}(q_j^m) \propto \exp\left(\mathbb{E}_{\widetilde{p}_{\boldsymbol{A}}^{(r)}\widetilde{p}_{\boldsymbol{Q}_{\setminus j}}^{(r-1)}\widetilde{p}_{\boldsymbol{q}_{\setminus j}^{\setminus m}}^{(r-1)}\widetilde{p}_{\boldsymbol{w}}^{(r)}\widetilde{p}_{\boldsymbol{h}}^{(r)}}\left[\log p\left(q_j^m\,|\,\boldsymbol{Y},\boldsymbol{A},\boldsymbol{q}_{\setminus j}^m,\boldsymbol{Q}^{\setminus m},\boldsymbol{w},\boldsymbol{h};\boldsymbol{\theta}^{(r)}\right)\right]\right) \tag{III.40}$$

Avant de détailler les calculs, nous précisons que pour des raisons de simplification dues à la présence du logarithme dans les équations précédentes nous allons utiliser les équations (III.6) et (III.7) de la manière suivante :

$$p\left(\boldsymbol{a}^m\,|\,\boldsymbol{q}^m, w^m, \boldsymbol{\theta}_m\right) = \prod_{j=1}^{J} \mathcal{N}(\mu_0^m, v_0^m)^{(1-q_j^m w^m)}\ \mathcal{N}(\mu_1^m, v_1^m)^{q_j^m w^m} \tag{III.41}$$

$$p(w^m\,|\,x^m, \boldsymbol{\tau}) = p_m^{w^m}\ (1-p_m)^{(1-w^m)} \tag{III.42}$$

De plus et afin d'améliorer l'estimation des paramètres v_h, la variance *a priori* de la FRH, et ξ_m, le paramètre de régularité spatiale du champ de Ising, qui ont tendance à être sur-estimés en VEM causant, une FRH trop lisse et un champ de Ising trop homogène, des *a priori* sous forme de lois exponentielles de paramètres λ_{v_h} et λ_{ξ_m} ont été respectivement introduits

$$p(v_h\,|\,\lambda_{v_h}) = \lambda_{v_h}\exp(-\lambda_{v_h} v_h) \tag{III.43}$$

$$p(\boldsymbol{\xi}\,|\,\lambda_\xi) = \prod_{m=1}^{M} p(\xi_m\,|\,\lambda_{\xi_m}) = \prod_{m=1}^{M} \lambda_{\xi_m}\exp(-\lambda_{\xi_m}\xi_m). \tag{III.44}$$

Contrairement à l'algorithme MCMC, en VEM aucun *a priori* n'est défini sur les paramètres $\boldsymbol{\sigma}^2, \boldsymbol{\theta}_A$, ni sur les coefficients des dérives basses fréquences $\boldsymbol{L}$ qui sont considérés comme paramètres et non comme variables aléatoires. La densité

a posteriori devient donc :

$$p(\boldsymbol{A}, \boldsymbol{w}, \boldsymbol{Q}, \boldsymbol{h}, \boldsymbol{\theta} \,|\, \boldsymbol{Y}) \propto p(\boldsymbol{Y} \,|\, \boldsymbol{A}, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{\sigma}^2)\; p(\boldsymbol{A} \,|\, \boldsymbol{w}, \boldsymbol{Q}, \boldsymbol{\theta}_A)\; p(\boldsymbol{w} \,|\, \boldsymbol{Q}, \boldsymbol{\tau}) \\ \times\, p(\boldsymbol{Q} \,|\, \boldsymbol{\xi})\; p(\boldsymbol{h} \,|\, v_h)\; p(v_h \,|\, \lambda_{v_h})\; p(\boldsymbol{\xi} \,|\, \lambda_\xi) \qquad \text{(III.45)}$$

où $\boldsymbol{\theta} = \{\boldsymbol{\sigma}^2, \boldsymbol{\theta}_A, \boldsymbol{\tau}, \boldsymbol{\xi}, v_h, \boldsymbol{L}\}$. La représentation graphique du modèle *parcimonieux*, dans le cadre VEM, est illustrée par la figure III.3-(b) où un bloc exprimant la pertinence a été ajouté par rapport au graphe représentant le modèle DEC *complet* (figure III.3-(a)).

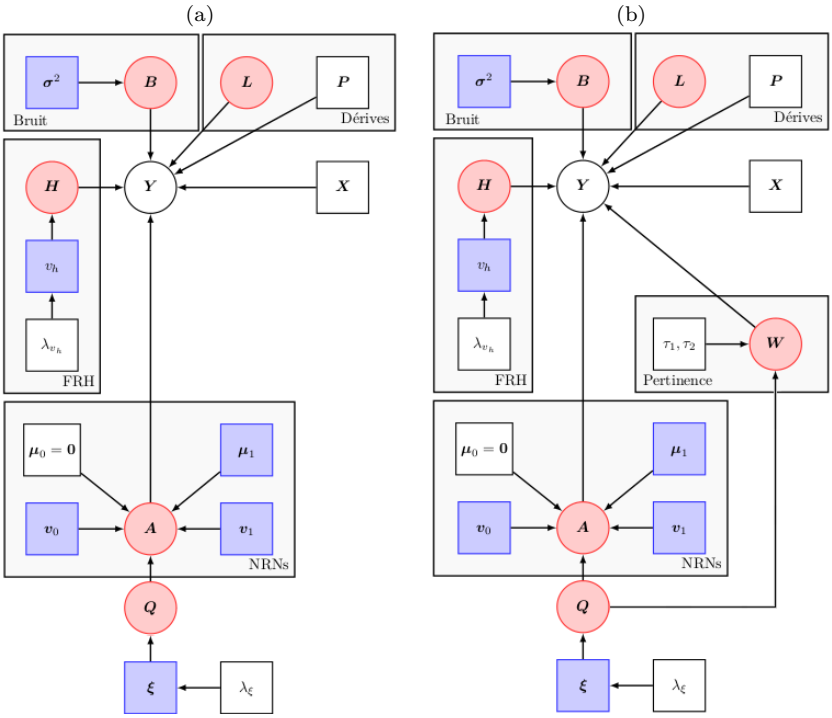


FIGURE III.3 – Modèle graphique illustrant, pour une parcelle $\mathcal{P}_\gamma$, les dépendances entre les variables aléatoires (cercles) et les paramètres (rectangles) du modèle DEC *complet* (a) et *parcimonieux* (b) dans le cadre de l'inférence VEM : les variables observées Y ainsi que les paramètres fixés sont en blanc, les variables non observées sont en rouge et les paramètres à estimer sont en bleu.

Dans les pages suivantes, on trouvera les distributions *a posteriori* des variables du modèle et les estimations de ces paramètres dont le calcul détaillé figure dans l'annexe (D.1). Par souci de compacité des notations, $\widetilde{p}_{w^m}(w^m = 1)$ et $\widetilde{p}_{q_j^m}(q_j^m = 1)$ seront respectivement notés $\widetilde{p}_{w^m}(1)$ et $\widetilde{p}_{q_j^m}(1)$.

♣ E-H

A partir de l'équation (III.37) et étant donné la nature gaussienne de la vraisemblance ainsi que de l'*a priori* sur $\boldsymbol{h}$, on peut déduire la loi gaussienne suivante :

$$\widetilde{p}_{\boldsymbol{h}}(\boldsymbol{h}) = \mathcal{N}(\boldsymbol{m}_{\boldsymbol{h}}, \boldsymbol{V}_{\boldsymbol{h}}) \tag{III.46}$$

où la moyenne et la variance ont des formes semblables à celle obtenues par l'inférence MCMC en moyennant les termes dépendant de a_j^m et w^m par rapport à $\widetilde{p}_{\boldsymbol{A}}$ et $\widetilde{p}_{\boldsymbol{w}}$ respectivement. Ce qui donne : $\boldsymbol{V}_{\boldsymbol{h}} = \left(\frac{\boldsymbol{R}^{-1}}{v_h} + E_{\widetilde{p}_{\boldsymbol{A}},\widetilde{p}_{\boldsymbol{w}}}\left[\sum_{j\in J} \boldsymbol{S}_j^{\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{S}_j\right]\right)^{-1}$ et $\boldsymbol{m}_{\boldsymbol{h}} = \boldsymbol{V}_{\boldsymbol{h}} E_{\widetilde{p}_{\boldsymbol{A}},\widetilde{p}_{\boldsymbol{w}}}\left[\sum_{j\in J} \boldsymbol{S}_j^{\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j - \boldsymbol{P}\boldsymbol{\ell}_j)\right]$,
avec

$$E_{\widetilde{p}_{\boldsymbol{A}},\widetilde{p}_{\boldsymbol{w}}}\left[\sum_{j\in J} \boldsymbol{S}_j^{\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{S}_j\right] = \sum_{j\in J}\sum_{m,m'}\left(m_{a_j^m}m_{a_j^{m'}} + V_{\boldsymbol{a}_j^{m,m'}}\right)\widetilde{p}_{w^m}(1)\widetilde{p}_{w^{m'}}(1)\boldsymbol{X}^{m\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{X}^{m'} \tag{III.47}$$

$$E_{\widetilde{p}_{\boldsymbol{A}},\widetilde{p}_{\boldsymbol{w}}}\left[\sum_{j\in J} \boldsymbol{S}_j^{\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j - \boldsymbol{P}\boldsymbol{\ell}_j)\right] = \sum_{j\in J}\sum_{m=1}^{M} m_{a_j^m}\widetilde{p}_{w^m}(1)\boldsymbol{X}^{m\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j - \boldsymbol{P}\boldsymbol{\ell}_j) \tag{III.48}$$

où $m_{a_j^m} = E_{\widetilde{p}_{a_j^m}}[a_j^m]$ et $V_{\boldsymbol{a}_j^{m,m'}}$ est la covariance entre a_j^m et $a_j^{m'}$ selon $\widetilde{p}_{\boldsymbol{a}_j^{m,m'}}$. Ces quantités seront définies dans l'étape **E-A**.

♣ E-A

Le calcul détaillé à partir de l'équation (III.38) nous permet d'obtenir la distribution *a posteriori* suivante pour les NRNs :

$$\widetilde{p}_{\boldsymbol{A}}(\boldsymbol{A}) = \prod_j \mathcal{N}\left(\boldsymbol{m}_{\boldsymbol{a}_j}, \boldsymbol{V}_{\boldsymbol{a}_j}\right) \tag{III.49}$$

et dont les paramètres sont définis par $\boldsymbol{V}_{\boldsymbol{a}_j} = \left(E_{\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{\boldsymbol{w}}}[\boldsymbol{G}^{\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{G}] + \sum_{i=0}^{1}\Delta_{i,j}\right)^{-1}$ et $\boldsymbol{m}_{\boldsymbol{a}_j} = \boldsymbol{V}_{\boldsymbol{a}_j}\left(E_{\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{\boldsymbol{w}}}[\boldsymbol{G}^{\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j - \boldsymbol{P}\boldsymbol{\ell}_j)] + \sum_{i=0}^{1}\Delta_{i,j}\boldsymbol{\mu}_i\right)$ où $\boldsymbol{G} = [\boldsymbol{g}^1|\boldsymbol{g}^2|...|\boldsymbol{g}^M]$ avec $\boldsymbol{g}^m = w^m\boldsymbol{X}^m\boldsymbol{h}$.
Notons $\boldsymbol{\mu}_i = [\mu_i^1, ..., \mu_i^M]^{\mathrm{t}}$ pour $i \in \{0,1\}$, $\Delta_{0,j} = \mathrm{diag}\left[\frac{1-\widetilde{p}_{q_j^m}(1)\widetilde{p}_{w^m}(1)}{v_0^m}\right]_{m=1:M}$ et $\Delta_{1,j} = \mathrm{diag}\left[\frac{\widetilde{p}_{q_j^m}(1)\widetilde{p}_{w^m}(1)}{v_1^m}\right]_{m=1:M}$. $E_{\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{\boldsymbol{w}}}[\boldsymbol{G}^{\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{G}]$ est une matrice de dimension

$M \times M$ dont chaque élément (m, m') s'écrit sous la forme suivante :

$$\begin{aligned} & E_{\widetilde{p}_{\boldsymbol{h}}, \widetilde{p}_{\boldsymbol{w}^{m,m'}}} \left[\boldsymbol{g}^{m\mathrm{t}} \boldsymbol{\Gamma}_j \boldsymbol{g}^{m'} \right] \\ & \quad = \widetilde{p}_{w^m}(1) \widetilde{p}_{w^{m'}}(1) \left[\boldsymbol{m}_{\boldsymbol{h}}^{\mathrm{t}} \boldsymbol{X}^{m\mathrm{t}} \boldsymbol{\Gamma}_j \boldsymbol{X}^{m'} \boldsymbol{m}_{\boldsymbol{h}} + \mathrm{tr} \left(\boldsymbol{V}_{\boldsymbol{h}} \boldsymbol{X}^{m\mathrm{t}} \boldsymbol{\Gamma}_j \boldsymbol{X}^{m'} \right) \right] \end{aligned} \tag{III.50}$$

et $E_{\widetilde{p}_{\boldsymbol{h}}, \widetilde{p}_{\boldsymbol{w}}} \left[\boldsymbol{G}^{\mathrm{t}} \boldsymbol{\Gamma}_j (\boldsymbol{y}_j - \boldsymbol{P} \boldsymbol{\ell}_j) \right]$ est un vecteur de dimension M dont chaque élément m est définit comme suit :

$$E_{\widetilde{p}_{\boldsymbol{h}}, \widetilde{p}_{w^m}} \left[\boldsymbol{g}^{m\mathrm{t}} \boldsymbol{\Gamma}_j (\boldsymbol{y}_j - \boldsymbol{P} \boldsymbol{\ell}_j) \right] = \widetilde{p}_{w^m}(1) \, \boldsymbol{m}_{\boldsymbol{h}}^{\mathrm{t}} \boldsymbol{X}^{m\mathrm{t}} \boldsymbol{\Gamma}_j (\boldsymbol{y}_j - \boldsymbol{P} \boldsymbol{\ell}_j) \tag{III.51}$$

Notons que $\widetilde{p}_{\boldsymbol{w}^{m,m'}}$ est le raccourcie de $\widetilde{p}_{w^m} \widetilde{p}_{w^{m'}}$.

♣ E-W

Rappelons que cette étape est décomposée en M sous-étapes (équation (III.39)) dont chacune permet l'estimation des probabilités *a posteriori* qu'une condition m soit considérée comme pertinente $\widetilde{p}_{w^m}(1)$ ou non pertinente $\widetilde{p}_{w^m}(0)$. Selon le calcul détaillé nous avons $\widetilde{p}_{w^m}(w^m) \propto f(w^m)$ avec

$$f(w^m = 0) = \exp \left(-\tau_1 \left[\sum_j \widetilde{p}_{q_j^m}(1) - \tau_2 \right] \right) \tag{III.52}$$

$$\begin{aligned} f(w^m = 1) = \exp \Bigg(\sum_j \Bigg\{ & \widetilde{p}_{q_j^m}(1) \left[\frac{V_{\boldsymbol{a}_j^{m,m}}}{2} \left(\frac{1}{v_0^m} - \frac{1}{v_1^m} \right) + \log \frac{\mathcal{N}(m_{a_j^m}; \mu_1^m, v_1^m)}{\mathcal{N}(m_{a_j^m}; \mu_0^m, v_0^m)} \right] \\ & - \frac{m_{a_j^m}^2 + V_{\boldsymbol{a}_j^{m,m}}}{2} \left[\boldsymbol{m}_{\boldsymbol{h}}^{\mathrm{t}} \boldsymbol{X}^{m\mathrm{t}} \boldsymbol{\Gamma}_j \boldsymbol{X}^m \boldsymbol{m}_{\boldsymbol{h}} + tr \left(\boldsymbol{V}_{\boldsymbol{h}} \boldsymbol{X}^{m\mathrm{t}} \boldsymbol{\Gamma}_j \boldsymbol{X}^m \right) \right] \\ & - \sum_{m' \neq m} \left(m_{a_j^m} m_{a_j^{m'}} + V_{\boldsymbol{a}_j^{m,m'}} \right) \widetilde{p}_{w^{m'}}(1) \Big[\boldsymbol{m}_{\boldsymbol{h}}^{\mathrm{t}} \boldsymbol{X}^{m\mathrm{t}} \boldsymbol{\Gamma}_j \boldsymbol{X}^{m'} \boldsymbol{m}_{\boldsymbol{h}} \\ & + tr \left(\boldsymbol{V}_{\boldsymbol{h}} \boldsymbol{X}^{m\mathrm{t}} \boldsymbol{\Gamma}_j \boldsymbol{X}^{m'} \right) \Big] + m_{a_j^m} \boldsymbol{m}_{\boldsymbol{h}}^{\mathrm{t}} \boldsymbol{X}^{m\mathrm{t}} \boldsymbol{\Gamma}_j (\boldsymbol{y}_j - \boldsymbol{P} \boldsymbol{\ell}_j) \Bigg\} \Bigg) \end{aligned} \tag{III.53}$$

Finalement, après normalisation on obtient :

$$\widetilde{p}_{w^m}(1) = \frac{f(w^m = 1)}{f(w^m = 1) + f(w^m = 0)} \tag{III.54}$$

♣ **E-Q**

Cette étape comprend $M \times J$ sous-étapes chacune permettant l'estimation des probabilités *a posteriori* qu'un voxel soit activé $\widetilde{p}_{q_j^m}(1)$ ou inactivé $\widetilde{p}_{q_j^m}(0)$ pour une condition m. Un calcul détaillé nous permet d'obtenir $\widetilde{p}_{q_j^m}(q_j^m) \propto f(q_j^m)$ avec :

$$f(q_j^m = 0) = \exp\left(-E_{\widetilde{p}_{\boldsymbol{q}_{\setminus j}^m}}\left(q_j^m = 0\right) + \xi_m \sum_{k \in N(j)} \widetilde{p}_{q_k^m}(0)\right) \tag{III.55}$$

$$f(q_j^m = 1) = \exp\left(\widetilde{p}_{w^m}(1)\left\{\frac{V_{\boldsymbol{a}_j^{m,m}}}{2}\left(\frac{1}{v_0^m} - \frac{1}{v_1^m}\right) + \log\frac{\mathcal{N}(m_{a_j^m}; \mu_1^m, v_1^m)}{\mathcal{N}(m_{a_j^m}; \mu_0^m, v_0^m)}\right\}\right.$$
$$\left. -\tau_1\, \widetilde{p}_{w^m}(0) - E_{\widetilde{p}_{\boldsymbol{q}_{\setminus j}^m}}\left(q_j^m = 1\right) + \xi_m \sum_{k \in N(j)} \widetilde{p}_{q_k^m}(1)\right) \tag{III.56}$$

où $E_{\widetilde{p}_{\boldsymbol{q}_{\setminus j}^m}}(q_j^m) = E_{\widetilde{p}_{\boldsymbol{q}_{\setminus j}^m}}[\log(1 + \exp[-\tau_1(\sum_j q_j^m - \tau_2)])]$ et dont le calcul nécessite une étape de Monte Carlo (MC, Annexe D.1). Comme pour w^m on obtient après normalisation :

$$\widetilde{p}_{q_j^m}(1) = \frac{f(q_j^m = 1)}{f(q_j^m = 1) + f(q_j^m = 0)} \tag{III.57}$$

Passons maintenant à l'étape M permettant d'estimer les paramètres du modèle $\{\boldsymbol{\theta}_A, v_h, \boldsymbol{\sigma}^2, \boldsymbol{L}, \boldsymbol{\xi}\}$ et qui est divisée en 5 sous-étapes :

♣ **M-$\boldsymbol{\theta}_A$**

$$\hat{\boldsymbol{\theta}}_A = \arg\max_{\boldsymbol{\theta}_A} E_{\widetilde{p}_{\boldsymbol{A}}, \widetilde{p}_{\boldsymbol{w}}, \widetilde{p}_{\boldsymbol{Q}}}\left[\log p(\boldsymbol{A} \,|\, \boldsymbol{w}, \boldsymbol{Q}, \boldsymbol{\theta}_A)\right] \tag{III.58}$$

Etant donné l'indépendance inter-conditions représentée dans $p(\boldsymbol{A} \,|\, \boldsymbol{w}, \boldsymbol{Q}, \boldsymbol{\theta}_A)$ (équation (III.5)), le calcul détaillé nous permet d'obtenir pour chaque condition m :

$$\hat{\mu}_1^m = \frac{\sum_j \widetilde{p}_{q_j^m}(1) m_{a_j^m}}{\sum_j \widetilde{p}_{q_j^m}(1)} \tag{III.59}$$

$$\hat{v}_1^m = \frac{\sum_j \widetilde{p}_{q_j^m}(1)\left[(m_{a_j^m} - \mu_1^m)^2 + V_{\boldsymbol{a}_j^{m,m}}\right]}{\sum_j \widetilde{p}_{q_j^m}(1)}, \tag{III.60}$$

et

$$\hat{v}_0^m = \frac{\sum_j \left(1 - \widetilde{p}_{q_j^m}(1)\widetilde{p}_{w^m}(1)\right)\left[(m_{a_j^m} - \mu_0^m)^2 + V_{\boldsymbol{a}_j^{m,m}}\right]}{\sum_j \left(1 - \widetilde{p}_{q_j^m}(1)\widetilde{p}_{w^m}(1)\right)} \tag{III.61}$$

♣ **M-v_h**

Cette variance est estimée à partir de l'équation suivante :

$$\hat{v}_h = \arg\max_{v_h} \left\{E_{\widetilde{p}_{\boldsymbol{h}}}\left[\log p(\boldsymbol{h} \mid v_h)\right] + \log p(v_h \mid \lambda_{v_h})\right\} \tag{III.62}$$

ce qui donne en tenant compte des équations (II.24) et (III.43) :

$$\hat{v}_h = \frac{(1-D) + \sqrt{(D-1)^2 + 8\lambda_{v_h}\mathrm{tr}\left[(\boldsymbol{m_h}\boldsymbol{m_h}^{\mathrm{t}} + \boldsymbol{V_h})\,\boldsymbol{R}^{-1}\right]}}{4\lambda_{v_h}} \tag{III.63}$$

♣ **M-$\boldsymbol{\sigma}^2$**

$$\hat{\boldsymbol{\sigma}}^2 = \arg\max_{\boldsymbol{\sigma}^2} E_{\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{\boldsymbol{A}},\widetilde{p}_{\boldsymbol{w}}}\left[\log p\left(\boldsymbol{Y} \mid \boldsymbol{A}, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{L}, \boldsymbol{\sigma}^2\right)\right] \tag{III.64}$$

L'indépendance conditionnelle des données $\boldsymbol{y}_j$ sachant $\boldsymbol{a}_j, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{\ell}_j, \sigma_j^2$ pour $j = 1 : J$ permet d'effectuer le calcul de la variance du bruit au niveau du voxel et d'obtenir :

$$\hat{\sigma_j}^2 = \frac{\boldsymbol{m}_{\boldsymbol{a}_j}^{\mathrm{t}} C_1 \boldsymbol{m}_{\boldsymbol{a}_j} + tr\left(\boldsymbol{V}_{\boldsymbol{a}_j} C_1\right) - 2\boldsymbol{m}_{\boldsymbol{a}_j}^{\mathrm{t}} C_2 + (\boldsymbol{y}_j - \boldsymbol{P}\boldsymbol{\ell}_j)^{\mathrm{t}}\boldsymbol{\Lambda}_j(\boldsymbol{y}_j - \boldsymbol{P}\boldsymbol{\ell}_j)}{N} \tag{III.65}$$

où $C_1 = E_{\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{\boldsymbol{w}}}\left[\boldsymbol{G}^{\mathrm{t}}\boldsymbol{\Lambda}_j\boldsymbol{G}\right]$ est une matrice de dimension $M \times M$ dont chaque élément (m, m') est donné par l'équation (III.50) et $C_2 = E_{\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{\boldsymbol{w}}}\left[\boldsymbol{G}^{\mathrm{t}}\boldsymbol{\Lambda}_j(\boldsymbol{y}_j - \boldsymbol{P}\boldsymbol{\ell}_j)\right]$ est un vecteur de dimension M dont chaque élément m est donné par l'équation (III.51). Cela en remplaçant, dans les deux cas, $\boldsymbol{\Gamma}_j$ par $\boldsymbol{\Lambda}_j$ où $\boldsymbol{\Gamma}_j = \frac{1}{\sigma_j^2}\boldsymbol{\Lambda}_j$.

♣ **M-$\boldsymbol{L}$**

$$\hat{\boldsymbol{L}} = \arg\max_{\boldsymbol{L}} E_{\widetilde{p}_{\boldsymbol{A}},\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{\boldsymbol{w}}}\left[\log p(\boldsymbol{Y} \mid \boldsymbol{A}, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{L}, \sigma_j^2)\right] \tag{III.66}$$

Pour la même raison d'indépendance conditionnelle des données $\boldsymbol{y}_j$ sachant $\boldsymbol{a}_j, \boldsymbol{w}$,

$\boldsymbol{h}, \boldsymbol{\ell}_j, \sigma_j^2$ pour $j = 1 : J$ on peut effectuer le calcul au niveau du voxel directement pour obtenir :

$$\hat{\boldsymbol{\ell}}_j = (\boldsymbol{P}^{\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{P})^{-1}\boldsymbol{P}^{\mathrm{t}}\left(\boldsymbol{y}_j - \sum_m m_{a_j^m}\widetilde{p}_{w^m}(1)\boldsymbol{X}^m\boldsymbol{m}_{h_\gamma}\right) \tag{III.67}$$

♣ **M-$\boldsymbol{\xi}$**

$$\hat{\boldsymbol{\xi}} = \arg\max_{\boldsymbol{\xi}} \left\{E_{\widetilde{p}_{\boldsymbol{Q}}}[\log p(\boldsymbol{Q} \mid \boldsymbol{\xi})] + \log p(\boldsymbol{\xi} \mid \lambda_\xi)\right\} \tag{III.68}$$

En tenant compte de l'indépendance *a priori* inter-conditions, le gradient correspondant est exprimé par l'équation :

$$\frac{d(.)}{d\xi_m} = -\frac{d\log Z(\xi_m)}{d\xi_m} + E_{\widetilde{p}_{Q^m}}[U(\boldsymbol{q}^m)] - \lambda_{\xi_m} \tag{III.69}$$

où étant donné la complexité de $Z(\xi_m)$ une approximation champ moyen est nécessaire [Celeux 2003] et l'estimation de ξ_m se fait à travers l'algorithme de descente du gradient (Annexe F.5).

Comme on le verra à travers les résultats de traitement sur données simulées et réelles (chapitre IV), le modèle parcimonieux précédent où les variables de pertinence $\boldsymbol{w}$ sont connectées aux labels $\boldsymbol{Q}$, souffre de deux problèmes principaux :

- Seuil de pertinence τ_2 élevé : la classification aléatoire des voxels activés et inactivés due à la perte d'identifiabilité du mélange gaussien *a posteriori* dans le cas d'une condition non pertinente, nécessite, sur certains jeux de données, le choix d'une valeur élevée du seuil τ_2 pour éliminer les conditions non pertinentes. Ceci pose deux problèmes : (i) l'incohérence avec la définition de la non pertinence (nombre faible ou nul de voxels activés) et (ii) le risque d'élimination des conditions pertinentes.
- L'étape MC nécessaire au calcul de la distribution *a posteriori* des labels dans le cadre de l'algorithme VEM désavantage l'algorithme par rapport au temps de calcul qui est la principale raison du passage de MCMC à VEM.

Pour ces deux raisons, nous avons proposé une autre approche qui profite du phénomène de superposition des gaussiennes d'activation pour la détection de la non pertinence. Cela en connectant les variables $\boldsymbol{w}$ à $\boldsymbol{\mu}_1 = \{\mu_1^m; m = 1 : M\}$ où μ_1^m est la moyenne de la classe active qui, comme on l'a vu, s'annule dans le cas

des conditions non pertinentes.

III.2.3 Connexion aux moyennes d'activation

Le recouvrement des deux gaussiennes du mélange gaussien *a posteriori* dans le cas d'une condition m non pertinente induit une moyenne μ_1^m proche de 0. Ce phénomène étant assez logique dans le sens où cela reflète l'absence d'activation induite par la condition en question et le fait qu'une seule gaussienne est censée représenter les NRNs dans ce cas, la gaussienne inactive. Nous proposons donc de bénéficier de ce phénomène afin de guider le modèle vers la sélection de la sous-famille des conditions pertinentes. Ceci peut se faire en connectant les variables de pertinence $\boldsymbol{w}$ aux paramètres $\boldsymbol{\mu}_1$ à travers le choix $x^m = (\mu_1^m)^2$ dans la sigmoïde :

$$p_m = p(w^m = 1 \,|\, \mu_1^m, \boldsymbol{\tau}) = \frac{1}{1 + \exp\{-\tau_1[(\mu_1^m)^2 - \tau_2]\}} \tag{III.70}$$

En plus de ce changement, le paramètre de la sigmoïde τ_2 est estimé contrairement à l'approche précédente où il était fixé. Pour ceci nous avons proposé d'introduire un *a priori* Gamma sur ce paramètre dans le but de diminuer la probabilité d'obtenir des petites valeurs insuffisantes à la détection de la non pertinence ou des grandes valeurs pouvant induire des faux négatifs (considérer la condition m comme non pertinente alors qu'elle l'est) :

$$p(\tau_2|\alpha_{\tau_2}, \lambda_{\tau_2}) = \frac{\lambda_{\tau_2}^{\alpha}}{\Gamma(\alpha_{\tau_2})} \tau_2^{\alpha_{\tau_2}-1} e^{-\lambda_{\tau_2}\tau_2} \tag{III.71}$$

La densité *a posteriori* est donc la même que celle définie par l'équation (III.45) en remplaçant le terme $p(\boldsymbol{w} \,|\, \boldsymbol{Q}, \boldsymbol{\tau})$ par $p(\boldsymbol{w} \,|\, \boldsymbol{\mu}_1, \boldsymbol{\tau})$ et en rajoutant l'*a priori* sur τ_2

$$\begin{aligned} p(\boldsymbol{A}, \boldsymbol{w}, \boldsymbol{Q}, \boldsymbol{h}, \boldsymbol{\theta} \,|\, \boldsymbol{Y}) \propto\; & p(\boldsymbol{Y} \,|\, \boldsymbol{A}, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{\sigma}^2)\; p(\boldsymbol{A} \,|\, \boldsymbol{w}, \boldsymbol{Q}, \boldsymbol{\theta}_A)\; p(\boldsymbol{w} \,|\, \boldsymbol{\mu}_1, \boldsymbol{\tau}) \\ & \times p(\boldsymbol{Q} \,|\, \boldsymbol{\xi})\; p(\boldsymbol{h} \,|\, v_h)\; p(v_h \,|\, \lambda_{v_h})\; p(\boldsymbol{\xi} \,|\, \lambda_\xi)\; p(\tau_2 \,|\, \alpha_{\tau_2}, \lambda_{\tau_2}) \end{aligned} \tag{III.72}$$

Le graphe représentant les dépendances entre les variables et paramètres du modèle est illustré par la figure III.4-(a) et comparé au graphe de l'approche *parcimonieuse* où $\boldsymbol{w}$ est connecté aux labels $\boldsymbol{Q}$ (figure III.4-(b)). Par rapport à cette dernière, les étapes E et M qui changent sont celles de $\boldsymbol{w}$, $\boldsymbol{Q}$ et $\boldsymbol{\mu}_1$ en plus d'une étape de maximisation supplémentaire permettant l'estimation du seuil de

pertinence τ_2 qui n'est plus à fixer par l'utilisateur.

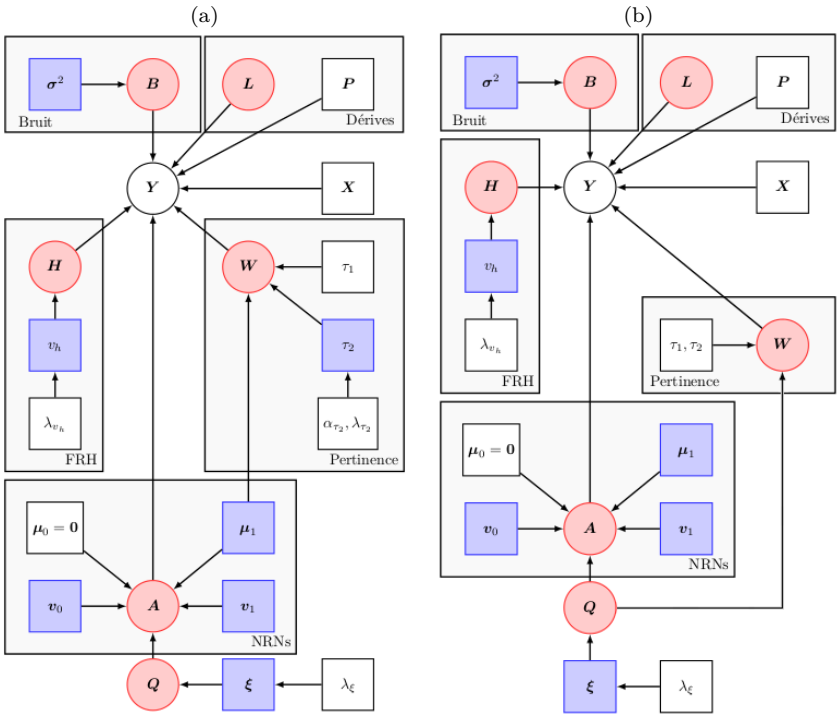


FIGURE III.4 – Comparaison des modèles graphiques illustrant, pour une parcelle $\mathcal{P}_\gamma$, les dépendances entre les variables aléatoires (cercles) et les paramètres (rectangles) des deux approches DES *parcimonieuses* dans le cadre de l'inférence VEM : (a) connexion entre $\boldsymbol{w}$ et $\boldsymbol{\mu}_1$, (b) connexion entre $\boldsymbol{w}$ et $\boldsymbol{Q}$. Les variables observées Y ainsi que les paramètres fixés sont en blanc, les variables non observées sont en rouge et les paramètres à estimer sont en bleu.

♣ E-W

Comme pour la sous-section III.2.2.2, cette étape est décomposée en M sous-étapes chacune permettant l'estimation des probabilités *a posteriori* de pertinence $\widetilde{p}_{w^m}(1)$ et de non pertinence $\widetilde{p}_{w^m}(0)$ d'une condition m. Dans ce cas, les termes $f(w^m = 0)$ et $f(w^m = 1)$ des équations (III.52) et (III.53) deviennent :

$$f(w^m = 0) = 1 - p_m \tag{III.73}$$

$$
\begin{aligned}
f(w^m = 1) = p_m \ \exp\Bigg(\sum_j \Bigg\{ & \widetilde{p}_{q_j^m}(1) \left[\frac{V_{\boldsymbol{a}_j^{m,m}}}{2} \left(\frac{1}{v_0^m} - \frac{1}{v_1^m} \right) + \log \frac{\mathcal{N}(m_{a_j^m}; \mu_1^m, v_1^m)}{\mathcal{N}(m_{a_j^m}; \mu_0^m, v_0^m)} \right] \\
& - \frac{m_{a_j^m}^2 + V_{\boldsymbol{a}_j^{m,m}}}{2} \left[\boldsymbol{m}_{\boldsymbol{h}}^{\mathrm{t}} \boldsymbol{X}^{m\mathrm{t}} \boldsymbol{\Gamma}_j \boldsymbol{X}^m \boldsymbol{m}_{\boldsymbol{h}} + tr\left(\boldsymbol{V}_{\boldsymbol{h}} \boldsymbol{X}^{m\mathrm{t}} \boldsymbol{\Gamma}_j \boldsymbol{X}^m \right) \right] \\
& - \sum_{m' \neq m} \left(m_{a_j^m} m_{a_j^{m'}} + V_{\boldsymbol{a}_j^{m,m'}} \right) \widetilde{p}_{w^{m'}}(1) \Big[\boldsymbol{m}_{\boldsymbol{h}}^{\mathrm{t}} \boldsymbol{X}^{m\mathrm{t}} \boldsymbol{\Gamma}_j \boldsymbol{X}^{m'} \boldsymbol{m}_{\boldsymbol{h}} \\
& + tr\left(\boldsymbol{V}_{\boldsymbol{h}} \boldsymbol{X}^{m\mathrm{t}} \boldsymbol{\Gamma}_j \boldsymbol{X}^{m'} \right) \Big] + \ m_{a_j^m} \boldsymbol{m}_{\boldsymbol{h}}^{\mathrm{t}} \boldsymbol{X}^{m\mathrm{t}} \boldsymbol{\Gamma}_j (\boldsymbol{y}_j - \boldsymbol{P}\boldsymbol{\ell}_j) \Bigg\} \Bigg)
\end{aligned}
\tag{III.74}
$$

Rappelons que p_m est défini par l'équation (III.70). En divisant les termes précédants par p_m on obtient les mêmes équations (III.52) et (III.53) en remplaçant le terme $\sum_j \widetilde{p}_{q_j^m}(1)$ par $(\mu_1^m)^2$ reflétant ainsi le changement d'*a priori* sur $\boldsymbol{w}$. Cette division n'étant pas problématique étant donné que $f(w^m = 0)$ et $f(w^m = 1)$ sont normalisés par la suite.

♣ E-Q

De même que la sous-section III.2.2.2, cette étape contient $M \times J$ sous-étapes où chacune estime les probabilités *a posteriori* qu'un voxel j de la condition m appartienne à la classe active $\widetilde{p}_{q_j^m}(1)$ ou inactive $\widetilde{p}_{q_j^m}(0)$. Dans ce nouveau modèle *parcimonieux*, les équations (III.55) et (III.56) deviennent :

$$
f(q_j^m = 0) = \exp\left(\xi_m \sum_{k \in N(j)} \widetilde{p}_{q_k^m}(0) \right) \tag{III.75}
$$

$$
\begin{aligned}
f(q_j^m = 1) = \exp\Bigg(& \widetilde{p}_{w^m}(1) \left\{ \frac{V_{\boldsymbol{a}_j^{m,m}}}{2} \left(\frac{1}{v_0^m} - \frac{1}{v_1^m} \right) + \log \frac{\mathcal{N}(m_{a_j^m}; \mu_1^m, v_1^m)}{\mathcal{N}(m_{a_j^m}; \mu_0^m, v_0^m)} \right\} \\
& + \xi_m \sum_{k \in N(j)} \widetilde{p}_{q_k^m}(1) \Bigg)
\end{aligned}
\tag{III.76}
$$

Ici, $\boldsymbol{w}$ ne dépend plus de $\boldsymbol{Q}$ *a priori*. On remarque donc, que par rapport aux équations (III.55) et (III.56), il manque les termes $-E_{\widetilde{p}_{\boldsymbol{q}_{\setminus j}^m}}\left(q_j^m = 0\right)$ et $-\tau_1 \ \widetilde{p}_{w^m}(0) - E_{\widetilde{p}_{\boldsymbol{q}_{\setminus j}^m}}\left(q_j^m = 1\right)$ venant de la connexion *a priori* entre $\boldsymbol{w}$ et $\boldsymbol{Q}$.

♣ M-μ_1

En tenant compte de l'indépendance inter-conditions et du terme supplémentaire

$p(w^m \,|\, \mu_1^m, \boldsymbol{\tau})$, le gradient obtenu est de la forme suivante :

$$\frac{d(.)}{d\mu_1^m} = B + B' \; \mu_1^m - 2 \; \tau_1 \; \mu_1^m \; S(\mu_1^m) = 0 \tag{III.77}$$

avec $B' = 2 \; \tau_1 \; \widetilde{p}_{w^m}(1) - \frac{\widetilde{p}_{w^m}(1)}{v_1^m} \sum_j \widetilde{p}_{q_j^m}(1)$ et $B = \frac{\widetilde{p}_{w^m}(1)}{v_1^m} \sum_j \widetilde{p}_{q_j^m}(1) m_{a_j^m}$. Contrairement à l'équation (III.59), ici la formule exprimant μ_1^m n'est plus explicite. Pour des raisons de problèmes numériques dans l'algorithme de descente du gradient, on a eu recourt à celle de dichotomie (Annexe F.4).

♣ **M-$\boldsymbol{\tau}_2$**

$$\hat{\tau}_2 = \arg\max_{\tau_2} \left\{ E_{\widetilde{p}_{\boldsymbol{w}}} \left[\log p(\boldsymbol{w} \,|\, \boldsymbol{\mu}_1, \boldsymbol{\tau}) \right] + \log p(\tau_2 | \alpha_{\tau_2}, \lambda_{\tau_2}) \right\} \tag{III.78}$$

De même que pour l'estimation de μ_1, une étape de dichotomie (Annexe F.4) permet l'estimation du paramètre τ_2 à partir du gradient suivant :

$$\frac{d(.)}{d\tau_2} = \tau_1 \sum_{m=1}^{M} \left[S(\mu_1^m, \tau_2) - \widetilde{p}_{w^m}(1) \right] - \lambda_{\tau_2} + \frac{\alpha_{\tau_2} - 1}{\tau_2} = 0 \tag{III.79}$$

III.2.4 Automatic Relevance Determination - ARD

Comme détaillé dans la sous-section III.2.1, dans le cas d'un *a priori* ARD, les variables supplémentaires $\boldsymbol{w} = \{w^m, m = 1 : M\}$ deviennent des variables gaussiennes et ne sont plus reliées aux autres variables ou paramètres du modèle. Leur loi *a priori* est définie par l'équation (III.9) et tient compte de l'indépendance inter-conditions étant donné que la pertinence d'une condition m ne dépend pas de la pertinence d'une autre dans une parcelle donnée. Nous choisissons également de changer l'*a priori* sur $\boldsymbol{A}$ qui devient :

$$p(\boldsymbol{A} \,|\, \boldsymbol{\theta}_A) = \prod_{m=1}^{M} p(\boldsymbol{a}^m \,|\, \boldsymbol{\theta}_m) = \prod_{m=1}^{M} \sum_{\boldsymbol{q}^m} p(\boldsymbol{a}^m \,|\, \boldsymbol{q}^m, \boldsymbol{\theta}_m) p(\boldsymbol{q}^m \,|\, \boldsymbol{\theta}_m) \tag{III.80}$$

avec :

$$p(\boldsymbol{a}^m \,|\, \boldsymbol{q}^m, \boldsymbol{\theta}_m) = \prod_{j=1}^{J} p(a_j^m | q_j^m, \theta_m) = \prod_{j=1}^{J} \mathcal{N}(a_j^m; \mu_0^m, v_0^m)^{1-q_j^m} \mathcal{N}(a_j^m; \mu_1^m, v_1^m)^{q_j^m} \tag{III.81}$$

ce qui est équivalent au cas du modèle DEC *complet*. Les autres *a priori* restent inchangés par rapport aux modèles DEC *parcimonieux* précédents et la densité *a posteriori* est définie comme suit :

$$p(\boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{w}, \boldsymbol{Q}, \boldsymbol{h}, \boldsymbol{\theta}) = p(\boldsymbol{Y} \mid \boldsymbol{A}, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{\sigma}^2)\, p(\boldsymbol{A} \mid \boldsymbol{Q}, \boldsymbol{\theta}_A)\, p(\boldsymbol{w} \mid \boldsymbol{\alpha})\, p(\boldsymbol{Q} \mid \boldsymbol{\xi}) \\ p(\boldsymbol{h} \mid v_h) p(v_h \mid \lambda_{v_h}) p(\boldsymbol{\xi} \mid \lambda_\xi) p(\boldsymbol{\alpha} \mid \nu_\alpha, \lambda_\alpha) \tag{III.82}$$

avec $\boldsymbol{\theta} = \{\boldsymbol{\sigma}^2, \boldsymbol{\theta}_A, \boldsymbol{\xi}, \boldsymbol{\alpha}, v_h, \boldsymbol{L}\}$, $p(v_h \mid \lambda_{v_h})$ et $p(\boldsymbol{\xi} \mid \lambda_\xi)$ ayant la même définition que les approches VEM parcimonieuses précédentes (équations (III.43) et (III.44) respectivement) et

$$p(\boldsymbol{\alpha} \mid \nu_\alpha, \lambda_\alpha) = \prod_{m=1}^{M} \mathcal{G}(\alpha_m; \nu_\alpha, \lambda_\alpha) = \prod_{m=1}^{M} \frac{\lambda_\alpha^{\nu_\alpha}}{\Gamma(\nu_\alpha)} (\alpha^m)^{\nu_\alpha - 1} \exp(-\lambda_\alpha \alpha_m) \tag{III.83}$$

L'estimation de la distribution *a posteriori* de la variable $\boldsymbol{w}$ ainsi que de sa précision *a priori* $\boldsymbol{\alpha}$ est effectuée selon les deux étapes suivantes, où **E-W** est, comme précédemment, décomposée en M sous-étapes.

♣ E-W

Pour chaque condition m, la probabilité *a posteriori* de la variable de pertinence correspondante est donnée par :

$$\widetilde{p}_{w^m}(w^m) \propto \exp\Big(\mathbb{E}_{\widetilde{p}_{\boldsymbol{h}}, \widetilde{p}_{\boldsymbol{A}}, \widetilde{p}_{w^{\backslash m}}} \Big[\log p\Big(\boldsymbol{Y} \mid \boldsymbol{A}, w^m, \boldsymbol{w}^{\backslash m}, \boldsymbol{h}, \boldsymbol{L}, \boldsymbol{\sigma}^2\Big)\Big]\Big) p(w^m \mid \alpha^m) \tag{III.84}$$

Le calcul du premier terme est donné dans l'annexe D.1 par l'équation (D.24) en tenant compte du fait que w^m n'est plus une variable binaire mais gaussienne et donc $\mathbb{E}_{\widetilde{p}_{w^m}}[w^m] = m_{w^m}$ ce qui donne :

$$\begin{aligned}
&\mathbb{E}_{\widetilde{p}_{\boldsymbol{h}}, \widetilde{p}_{\boldsymbol{A}}, \widetilde{p}_{\boldsymbol{w}^{\backslash m}}} \left[\log p\left(\boldsymbol{Y} \mid \boldsymbol{A}, w^m, \boldsymbol{w}^{\backslash m}, \boldsymbol{h}, \boldsymbol{L}, \boldsymbol{\sigma}^2\right)\right] \\
&= C'_{2,w^m} - \frac{1}{2} \sum_{j \in J} \Big\{ -2\, w^m m_{a_j^m} \boldsymbol{m}_{\boldsymbol{h}}^{\mathrm{t}} \boldsymbol{X}^{m\mathrm{t}} \boldsymbol{\Gamma}_j (\boldsymbol{y}_j - \boldsymbol{P}\boldsymbol{\ell}_j) \\
&+ (w^m)^2 \left(m_{a_j^m}^2 + V_{\boldsymbol{a}_j^{m,m}}\right) \left[\boldsymbol{m}_{\boldsymbol{h}}^{\mathrm{t}} \boldsymbol{X}^{m\mathrm{t}} \boldsymbol{\Gamma}_j \boldsymbol{X}^m \boldsymbol{m}_{\boldsymbol{h}} + tr\left(\boldsymbol{V}_{\boldsymbol{h}} \boldsymbol{X}^{m\mathrm{t}} \boldsymbol{\Gamma}_j \boldsymbol{X}^m\right)\right] \\
&+ 2\, w^m \sum_{m' \neq m} \left(m_{a_j^m} m_{a_j^{m'}} + V_{\boldsymbol{a}_j^{m,m'}}\right) m_{w^{m'}} \Big[\boldsymbol{m}_{\boldsymbol{h}}^{\mathrm{t}} \boldsymbol{X}^{m\mathrm{t}} \boldsymbol{\Gamma}_j \boldsymbol{X}^{m'} \boldsymbol{m}_{\boldsymbol{h}} \\
&+ \mathrm{tr}\left(\boldsymbol{V}_{\boldsymbol{h}} \boldsymbol{X}^{m\mathrm{t}} \boldsymbol{\Gamma}_j \boldsymbol{X}^{m'}\right)\Big]\Big\}.
\end{aligned} \tag{III.85}$$

Cela donne après la prise en compte de $p(w^m \,|\, \alpha^m)$:

$$\widetilde{p}_{w^m}(w^m) \propto \mathcal{N}(w^m; m_{w^m}, v_{w^m}) \tag{III.86}$$

avec :

$$v_{w^m} = \left\{ \sum_j \left(m^2_{a_j^m} + V^2_{\boldsymbol{a}_j^{m,m}} \right) \left[\boldsymbol{m}_{\boldsymbol{h}}^{\mathrm{t}} \boldsymbol{X}^{m\mathrm{t}} \boldsymbol{\Gamma}_j \boldsymbol{X}^m \boldsymbol{m}_{\boldsymbol{h}} + \mathrm{tr}\left(\boldsymbol{V}_{\boldsymbol{h}} \boldsymbol{X}^{m\mathrm{t}} \boldsymbol{\Gamma}_j \boldsymbol{X}^m \right) \right] + \alpha_m \right\}^{-1} \tag{III.87}$$

et

$$\begin{aligned} m_{w^m} = v_{w^m} \Bigg\{ \sum_j m_{a_j^m} \boldsymbol{m}_{\boldsymbol{h}}^{\mathrm{t}} \boldsymbol{X}^{m\mathrm{t}} \boldsymbol{\Gamma}_j (\boldsymbol{y}_j - \boldsymbol{P}\boldsymbol{\ell}_j) - \sum_{m' \neq m} \Bigg(\left[m_{a_j^m} m_{a_j^{m'}} + V_{\boldsymbol{a}_j^{m,m'}} \right] \\ \times m_{w^{m'}} \left[\boldsymbol{m}_{\boldsymbol{h}}^{\mathrm{t}} \boldsymbol{X}^{m\mathrm{t}} \boldsymbol{\Gamma}_j \boldsymbol{X}^{m'} \boldsymbol{m}_{\boldsymbol{h}} + \mathrm{tr}\left(\boldsymbol{V}_{\boldsymbol{h}} \boldsymbol{X}^{m\mathrm{t}} \boldsymbol{\Gamma}_j \boldsymbol{X}^{m'} \right) \right] \Bigg) \Bigg\} \end{aligned} \tag{III.88}$$

Rappelons qu'étant donné le produit $a_j^m w^m$ dans la définition du signal stimulus-induit, un w^m gaussien risque de récupérer une partie de a_j^m quand la condition m est pertinente et donc quand le w^m correspondant n'est pas estimé à 0. La question étant : *Cette bilinéarité serait-elle problématique ?* Ce détail sera discuté dans le chapitre V.

♣ M-$\boldsymbol{\alpha}$

$$\widehat{\boldsymbol{\alpha}} = \arg\max_{\boldsymbol{\alpha}} \left\{ \mathbb{E}_{\widetilde{p}_{\boldsymbol{w}}}[\log p(\boldsymbol{w} \,|\, \boldsymbol{\alpha})] + \log p(\boldsymbol{\alpha} \,|\, \nu_\alpha, \lambda_\alpha) \right\} \tag{III.89}$$

L'indépendance inter-conditions exprimée dans les équations (III.83) et (III.9) nous permet de calculer directement :

$$\begin{aligned} \widehat{\alpha}^m &= \arg\max_{\alpha^m} \left\{ \mathbb{E}_{\widetilde{p}_{w^m}}[\log p(w^m \,|\, \alpha^m)] + \log p(\alpha^m \,|\, \nu_\alpha, \lambda_\alpha) \right\} \\ &= \arg\max_{\alpha^m} \left\{ C_{\alpha_m} - \left(\frac{m^2_{w^m} + v_{w^m}}{2} + \lambda_\alpha \right) \alpha_m + \left(\nu_\alpha - \frac{1}{2} \right) \log \alpha^m \right\} \end{aligned} \tag{III.90}$$

où C_{α_m} est une constante en α_m. En dérivant cette expression par rapport à α_m et en annulant la dérivée on obtient :

$$-\left(\frac{m^2_{w^m} + v_{w^m}}{2} + \lambda_\alpha \right) + \left(\nu_\alpha - \frac{1}{2} \right) \frac{1}{\alpha^m} = 0 \tag{III.91}$$

Ce qui donne :

$$\widehat{\alpha}^m = \frac{\nu_\alpha - \frac{1}{2}}{\lambda_\alpha + \frac{1}{2}(m_{w^m}^2 + V_{w^m})} \tag{III.92}$$

Le changement de la loi *a posteriori* de $\boldsymbol{w}$ se reflète sur le reste des variables et paramètres du modèle en effectuant les changements suivants dans les étapes **E-H**, **M-$\boldsymbol{\sigma}^2$** et **M-$\boldsymbol{L}$** de la précédente approche DEC *parcimonieuse* (sous-section III.2.2.2) :

$$\mathbb{E}_{\widetilde{p}_{w^m}}[w^m] = \widetilde{p}_{w^m}(1) \quad \rightarrow \quad \mathbb{E}_{\widetilde{p}_{w^m}}[w^m] = m_{w^m}$$
$$\mathbb{E}_{\widetilde{p}_{\boldsymbol{w}^{m,m'}}}[w^m w^{m'}] = \widetilde{p}_{w^m}(1)\widetilde{p}_{w^{m'}}(1) \quad \rightarrow \quad \mathbb{E}_{\widetilde{p}_{\boldsymbol{w}^{m,m'}}}[w^m w^{m'}] = m_{w^m} m_{w^{m'}} + v_{w^{m,m'}}$$

où $v_{w^{m,m'}} = v_{w^m}$ si $m = m'$ et 0 sinon. En plus de ces changements, le nouvel *a priori* sur $\boldsymbol{A}$ (équation (III.81)) induit un changement au niveau des étapes **E-A**, **E-Q** et M-$\boldsymbol{\theta}_A$.

♣ **E-A**

$$\widetilde{p}_{\boldsymbol{A}}(\boldsymbol{A}) = \prod_j \mathcal{N}\left(\boldsymbol{m}_{\boldsymbol{a}_j}, \boldsymbol{V}_{\boldsymbol{a}_j}\right) \tag{III.93}$$

où les paramètres de cette loi gaussienne sont $\boldsymbol{V}_{\boldsymbol{a}_j} = \left(E_{\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{\boldsymbol{w}}}[\boldsymbol{G}^{\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{G}] + \sum_{i=0}^{1}\Delta_{i,j}\right)^{-1}$, $\boldsymbol{m}_{\boldsymbol{a}_j} = \boldsymbol{V}_{\boldsymbol{a}_j}\left(E_{\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{\boldsymbol{w}}}[\boldsymbol{G}^{\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j - \boldsymbol{P}\boldsymbol{\ell}_j)] + \sum_{i=0}^{1}\Delta_{i,j}\boldsymbol{\mu}_i\right)$. Notons pour $i \in \{0,1\}$ $\boldsymbol{\mu}_i = [\mu_i^1, ..., \mu_i^M]^{\mathrm{t}}$ et $\Delta_{i,j} = \mathrm{diag}\left[\frac{\widetilde{p}_{\boldsymbol{Q}_j^1}(i)}{v_i^1}, ..., \frac{\widetilde{p}_{\boldsymbol{Q}_j^M}(i)}{v_i^M}\right]$. $E_{\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{\boldsymbol{w}}}[\boldsymbol{G}^{\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{G}]$ est une matrice de dimension $M \times M$ dont chaque élément (m, m') s'écrit sous la forme suivante :

$$\begin{aligned} &E_{\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{\boldsymbol{w}^{m,m'}}}\left[\boldsymbol{g}^{m\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{g}^{m'}\right] \\ &= (m_{w^m}m_{w^{m'}} + V_{w^{m,m'}})\left[\boldsymbol{m}_{\boldsymbol{h}}^{\mathrm{t}}\boldsymbol{X}^{m\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{X}^{m'}\boldsymbol{m}_{\boldsymbol{h}} + \mathrm{tr}\left(\boldsymbol{V}_{\boldsymbol{h}}\boldsymbol{X}^{m\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{X}^{m'}\right)\right] \end{aligned} \tag{III.94}$$

et $E_{\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{\boldsymbol{w}}}[\boldsymbol{G}^{\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j - \boldsymbol{P}\boldsymbol{\ell}_j)]$ est un vecteur de dimension M dont chaque élément m est définit comme suit :

$$E_{\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{w^m}}\left[\boldsymbol{g}^{m\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j - \boldsymbol{P}\boldsymbol{\ell}_j)\right] = m_{w^m}\, \boldsymbol{m}_{\boldsymbol{h}}^{\mathrm{t}}\boldsymbol{X}^{m\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j - \boldsymbol{P}\boldsymbol{\ell}_j) \tag{III.95}$$

♣ E-Q

Ici, la probabilité $\widetilde{p}_{q_j^m}(q_j^m)$ au voxel j et à la condition m est exprimée par l'équation suivante :

$$\begin{aligned}\widetilde{p}_{q_j^m}(q_j^m) &\propto \exp\left(E_{\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{\boldsymbol{A}},\widetilde{p}_{\boldsymbol{q}_{\setminus j}^m},\widetilde{p}_{\boldsymbol{Q}^{\setminus m}}}\left[\log p(q_j^m \mid \boldsymbol{q}_{\setminus j}^m, \boldsymbol{Q}^{\setminus m}, \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{h}, \boldsymbol{\theta})\right]\right) \\ &\propto f(q_j^m) \end{aligned} \tag{III.96}$$

où l'on note :

$$f(q_j^m) = \exp\left(E_{\widetilde{p}_{a_j^m}}\left[\log p(a_j^m \mid q_j^m, \boldsymbol{\theta}_m)\right] + E_{\widetilde{p}_{\boldsymbol{q}_{\setminus j}^m}}\left[\log p(\boldsymbol{q}^m \mid \xi_m)\right]\right) \tag{III.97}$$

avec

$$f(q_j^m = 0) = \exp\left(\xi_m \sum_{k \in N(j)} \widetilde{p}_{q_k^m}(0)\right) \tag{III.98}$$

$$\begin{aligned} f(q_j^m = 1) = \exp\Bigg(&\frac{V_{\boldsymbol{a}_j^{m,m}}}{2}\left(\frac{1}{v_0^m} - \frac{1}{v_1^m}\right) + \log\frac{\mathcal{N}(m_{a_j^m}; \mu_1^m, v_1^m)}{\mathcal{N}(m_{a_j^m}; \mu_0^m, v_0^m)} \\ &+ \xi_m \sum_{k \in N(j)} \widetilde{p}_{q_k^m}(1)\Bigg) \end{aligned} \tag{III.99}$$

La connexion *a priori* entre $\boldsymbol{A}$ et $\boldsymbol{w}$ ayant disparue, les deux étapes précédentes (**E-A** et **E-Q**) sont équivalentes au cas du modèle DEC *complet*.

♣ M-$\boldsymbol{\theta}_A$

L'estimation de ces paramètres est effectuée à travers la maximisation suivante :

$$\hat{\boldsymbol{\theta}}_A = \arg\max_{\boldsymbol{\theta}_A} E_{\widetilde{p}_{\boldsymbol{A}},\widetilde{p}_{\boldsymbol{Q}}}\left[\log p(\boldsymbol{A} \mid \boldsymbol{Q}, \boldsymbol{\theta}_A)\right] \tag{III.100}$$

En effectuant le même calcul que celui détaillé dans l'annexe D.1 après avoir

supprimer la dépendance avec $\boldsymbol{w}$ ont obtient :

$$\hat{v}_i^m = \frac{\sum_j \widetilde{p}_{q_j^m}(i)\left[(m_{a_j^m} - \mu_i^m)^2 + V_{\boldsymbol{a}_j^{m,m}}\right]}{\sum_j \widetilde{p}_{q_j^m}(i)} \quad ; \quad i \in \{0,1\} \tag{III.101}$$

$$\hat{\mu}_1^m = \frac{\sum_j \widetilde{p}_{q_j^m}(1) m_{a_j^m}}{\sum_j \widetilde{p}_{q_j^m}(1)}. \tag{III.102}$$

De même que pour les deux étapes précédentes, **M-$\boldsymbol{\theta}_A$** dans le cas de l'*a priori* ARD est équivalente au cas du modèle DEC *complet*.

Une discussion plus détaillée concernant les trois modèles DEC *parcimonieux* proposés précédemment sera effectuée dans les chapitres IV et V.

III.2.5 Energie Libre en VEM

En VEM, l'énergie libre $\mathcal{F}(\widetilde{p}, \boldsymbol{\theta})$ représente le critère maximisé, son calcul pourrait donc être utile à la vérification de la convergence de l'algorithme. Rappelons que l'énergie libre est définie par l'équation :

$$F(\widetilde{p}, \boldsymbol{\theta}) = \mathbb{E}_{\widetilde{p}}[\log p(\boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{Q}, \boldsymbol{w}, \boldsymbol{h} \,|\, \boldsymbol{\theta})] - \mathbb{E}_{\widetilde{p}}[\log \widetilde{p}(\boldsymbol{A}, \boldsymbol{Q}, \boldsymbol{w}, \boldsymbol{h})] \tag{III.103}$$

où le premier terme est donné par :

- Cas des modèles III.2.2.2 et III.2.3 :

$$\begin{aligned} &E_{\widetilde{p}}[\log p(\boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{Q}, \boldsymbol{w}, \boldsymbol{h} \,|\, \boldsymbol{\theta})] \\ &= E_{\widetilde{p}_{\boldsymbol{A}}\widetilde{p}_{\boldsymbol{w}}\widetilde{p}_{\boldsymbol{h}}}\left[\log p(\boldsymbol{Y} \,|\, \boldsymbol{A}, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{L}, \boldsymbol{\sigma}^2)\right] + E_{\widetilde{p}_{\boldsymbol{A}}\widetilde{p}_{\boldsymbol{w}}\widetilde{p}_{\boldsymbol{Q}}}[\log p(\boldsymbol{A} \,|\, \boldsymbol{Q}, \boldsymbol{w}, \boldsymbol{\theta}_A)] \\ &\quad + E_{\widetilde{p}_{\boldsymbol{h}}}[\log p(\boldsymbol{h} \,|\, v_h)] + E_{\widetilde{p}_{\boldsymbol{w}}}[\log p(\boldsymbol{w} \,|\, \boldsymbol{x}, \boldsymbol{\tau})] + E_{\widetilde{p}_{\boldsymbol{Q}}}[\log p(\boldsymbol{Q} \,|\, \boldsymbol{\xi})] \end{aligned} \tag{III.104}$$

avec $\boldsymbol{x} = \boldsymbol{Q}$ ou $\boldsymbol{x} = \boldsymbol{\mu}_1$ selon le modèle.

- Cas du modèle III.2.4 :

$$\begin{aligned} &E_{\widetilde{p}}[\log p(\boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{Q}, \boldsymbol{w}, \boldsymbol{h} \,|\, \boldsymbol{\theta})] \\ &= E_{\widetilde{p}_{\boldsymbol{A}}\widetilde{p}_{\boldsymbol{w}}\widetilde{p}_{\boldsymbol{h}}}\left[\log p(\boldsymbol{Y} \,|\, \boldsymbol{A}, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{L}, \boldsymbol{\sigma}^2)\right] + E_{\widetilde{p}_{\boldsymbol{A}}\widetilde{p}_{\boldsymbol{Q}}}[\log p(\boldsymbol{A} \,|\, \boldsymbol{Q}, \boldsymbol{\theta}_A)] \\ &\quad + E_{\widetilde{p}_{\boldsymbol{h}}}[\log p(\boldsymbol{h} \,|\, v_h)] + E_{\widetilde{p}_{\boldsymbol{w}}}[\log p(\boldsymbol{w} \,|\, \boldsymbol{\alpha})] + E_{\widetilde{p}_{\boldsymbol{Q}}}[\log p(\boldsymbol{Q} \,|\, \boldsymbol{\xi})] \end{aligned} \tag{III.105}$$

et le deuxième terme représentant l'entropie est donné par :

$$\begin{aligned} & E_{\widetilde{p}}[\log \widetilde{p}\,(\boldsymbol{A}, \boldsymbol{Q}, \boldsymbol{w}, \boldsymbol{h})] \\ & = E_{\widetilde{p}_{\boldsymbol{A}}}[\log \widetilde{p}_{\boldsymbol{A}}(\boldsymbol{A})] + E_{\widetilde{p}_{\boldsymbol{Q}}}[\log \widetilde{p}_{\boldsymbol{Q}}(\boldsymbol{Q})] + E_{\widetilde{p}_{\boldsymbol{w}}}[\log \widetilde{p}_{\boldsymbol{w}}(\boldsymbol{w})] + E_{\widetilde{p}_{\boldsymbol{h}}}[\log \widetilde{p}_{\boldsymbol{h}}(\boldsymbol{h})] \end{aligned} \quad \text{(III.106)}$$

Le calcul détaillé des termes précédents se trouve dans l'annexe E.

III.3 Conclusion :

Ce chapitre discute la problématique de la pertinence des conditions expérimentales dans l'analyse des données IRMf et dans le cadre de l'approche DEC. Comme précisé au début du chapitre, la littérature s'est plutôt intéressée à la pertinence de modélisation des dérives basses fréquences dans le contexte du MLG classique à travers l'ajout d'un *a priori* ARD. Ici, deux définitions de la non pertinence ont été proposées, la première (def-1) est basée sur le nombre de voxels activés induit par la condition en question. La deuxième (def-2) quant à elle se base sur la moyenne de l'activité détectée pour cette condition et qui s'avère plus performante que def-1 en temps de calcul étant donnée la disparition de l'étape MC dans cette dernière. Nous avons également adapté l'*a priori* ARD, utilisé dans la littérature, à l'approche DEC. Des résultats sur données simulées et réelles ainsi qu'une comparaison détaillée avec le modèle *complet* seront exploités dans le chapitre IV. De plus des résultats préliminaires d'analyses cerveau entier figurent dans le chapitre V. Cela pour les trois modèles *parcimonieux* précédemment détaillés.

Chapitre IV

Expériences et résultats

Le but principal de ce chapitre est d'étudier l'intérêt du modèle *parcimonieux* capable de sélectioner automatiquement la sous-famille des conditions pertinentes par rapport au modèle *complet* qui considère toutes les conditions comme pertinentes. Rappelons que deux modèles *parcimonieux* ont été développés. Le premier, appelé par la suite modèle *parcimonieux 1*, permet la connexion entre la variable de pertinence $\boldsymbol{w}$ et les labels $\boldsymbol{Q}$ tandis que le second, appelé par la suite modèle *parcimonieux 2* (développé uniquement en VEM), permet la connexion entre $\boldsymbol{w}$ et la moyenne de la classe active $\boldsymbol{\mu}_1$. Nous allons donc à travers différents jeux de données simulées et réelles comparer nos trois modèles en nous basant sur plusieurs critères quantitatifs comme les intervalles de crédibilité [1], les erreurs quadratiques moyennes, les courbes ROC et les PPMs (*Posterior Probability Maps*).

Par la suite, les matrices et vecteurs sont notés en gras avec les matrices en majuscule et les vecteurs en minuscule. Un vecteur est par convention un vecteur colonne. La transposée est notée ^{t}.

1. L'intervalle de crédibilité (IC) reflète, dans le cadre de l'aproche bayésienne, la confiance que l'on peut avoir sur la valeur du paramètre estimé. Par définition, un IC au niveau $1-\alpha$ est un intervalle tel que la probabilité qu'a le paramètre de lui appartenir, selon la distribution a posteriori, est de α.

IV.1 Données simulées

IV.1.1 Procédure de simulation

Les simulations sont effectuées en 2D sur $J = 400$ pixels, à partir du modèle défini par l'équation :

$$\boldsymbol{y}_j = \sum_{m=1}^{M} a_j^m w^m \boldsymbol{X}^m \boldsymbol{h} + \boldsymbol{P}\boldsymbol{\ell}_j + \boldsymbol{b}_j \tag{IV.1}$$

L'étape initiale consiste à simuler des labels ou étiquettes (variables cachées $\{q_j^m \ ; \ m = 1 : M, j = 1 : J\}$) qui représentent les deux classes d'activation. La figure IV.1 montre des exemples de labels (première ligne) pour deux conditions pertinentes $m = 1, 2$ (figures IV.1-(a,b)) et une non pertinente $m = 3$ (figure IV.1-(c)) où la classe active ($q_j^m = 1$) est représentée en rouge et l'inactive ($q_j^m = 0$) en noir. Une fois les labels définis, les NRNs ($\{a_j^m \ ; \ m = 1 : M, j = 1 : J\}$) sont simulés par un mélange gaussien, où pour chaque pixel $j \in \mathcal{P}_\gamma$, $a_j^m \sim q_j^m \mathcal{N}(\mu_1^m, v_1^m) + (1 - q_j^m)\mathcal{N}(0, v_0^m)$. En parallèle, la séquence binaire représentant la projection du paradigme expérimental sur une grille d'échantillonnage de pas Δt (Annexe B) est convoluée avec la FRH choisie par l'utilisateur qui peut être de forme canonique ou autre. Le résultat est ensuite multiplié par $a_j^m w^m$ formant ainsi le signal stimulus-induit $a_j^m w^m \boldsymbol{X}^m \boldsymbol{h}$ au pixel j pour la condition m. Notons que $w^m = 1$ correspond à une condition pertinente et $w^m = 0$ à une condition non pertinente, ceci en correspondance avec la définition de la non pertinence (absence d'activation significative). Un exemple de ce signal est représenté dans la figure IV.2 pour des pixels inactivés (figures IV.2-(d,e)) et activés (figures IV.2-(f,g)) ceci pour chacune des deux conditions pertinentes $m = 1, 2$ (figures IV.2-(a,b) respectivement). Ce signal est cependant nul pour tous les pixels de la condition non pertinente $m = 3$ ($w^3 = 0$) comme le montre la figure IV.2-(c). Une fois les signaux stimulus-induits de chaque condition construits, leur somme nous permet d'obtenir le signal total $\sum_{m=1}^{3} a_j^m w^m \boldsymbol{X}^m \boldsymbol{h}$ représenté en noir dans la figure IV.2-(h). Le signal obtenu étant construit à partir d'une grille d'échantillonnage Δt, un sous échantillonnage de pas $TR/\Delta t$ est nécessaire à l'obtention des points temporels correspondants aux N scans (figure IV.2-(i)). Ensuite sont ajoutées les dérives basses fréquences $\boldsymbol{P}\boldsymbol{\ell}_j$ où $\boldsymbol{P}$ est une base orthogonale de fonctions polynomiales de degré $S = 4$ et $\boldsymbol{\ell}_j \sim \mathcal{N}(0, v_l \boldsymbol{I}_4)$ (figure IV.2-(j) où $v_l = 1$).

La figure IV.2-(k) montre l'effet de ces dérives en représentant le signal bruité en noir et le signal stimulus-induit total en vert. Finalement, on ajoute du bruit blanc $\boldsymbol{b}_j \sim \mathcal{N}(0, \sigma^2 \boldsymbol{I}_N)$ (figure IV.2-(l)) afin d'obtenir le signal BOLD au voxel j (figure IV.2-(m) où $\sigma^2 = 1$).

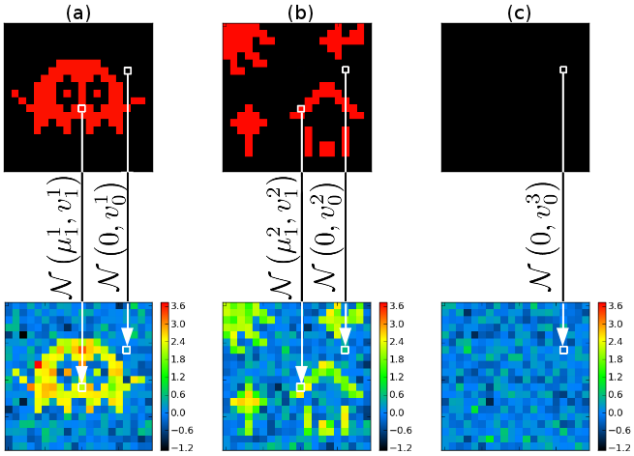


FIGURE IV.1 – Simulation des NRNs (en bas) à partir des labels (en haut), où les pixels appartenant à la classe active (labels rouges) sont simulés selon la loi gaussienne $\mathcal{N}(\mu_1^m, v_1^m)$ et ceux appartenant à la classe inactive (labels noirs) sont simulés selon la gaussienne $\mathcal{N}(0, v_0^m)$, ceci pour deux conditions pertinentes $m = 1, 2$ (a,b) et une condition non pertinente $m = 3$ (c) où tous les pixels sont inactifs selon la définition de la non pertinence.

IV.1.2 Résultats d'analyse et comparaison de modèles

Comme expliqué dans le chapitre II, dans le cas du modèle *complet*, quand une condition est non pertinente, nous observons une superposition des gaussiennes représentant les classes actives et inactives. Cela induit une classification aléatoire des pixels et donc un nombre de pixels actifs ($\sum_{j \in \mathcal{P}_\gamma} q_j^m$) tendant vers $\frac{J}{2} = 200$ pour chacune des deux conditions non pertinentes. Etant donné la connexion entre $\boldsymbol{w}$ et $\boldsymbol{Q}$ dans le modèle *parcimonieux 1*, ce phénomène pourrait, comme on le verra par la suite, être problématique pour le choix du paramètre τ_2 contrôlant le point d'inflexion de la sigmoïde et donc le seuil au-delà duquel la condition à une forte probabilité *a priori* d'être considérée comme pertinente.

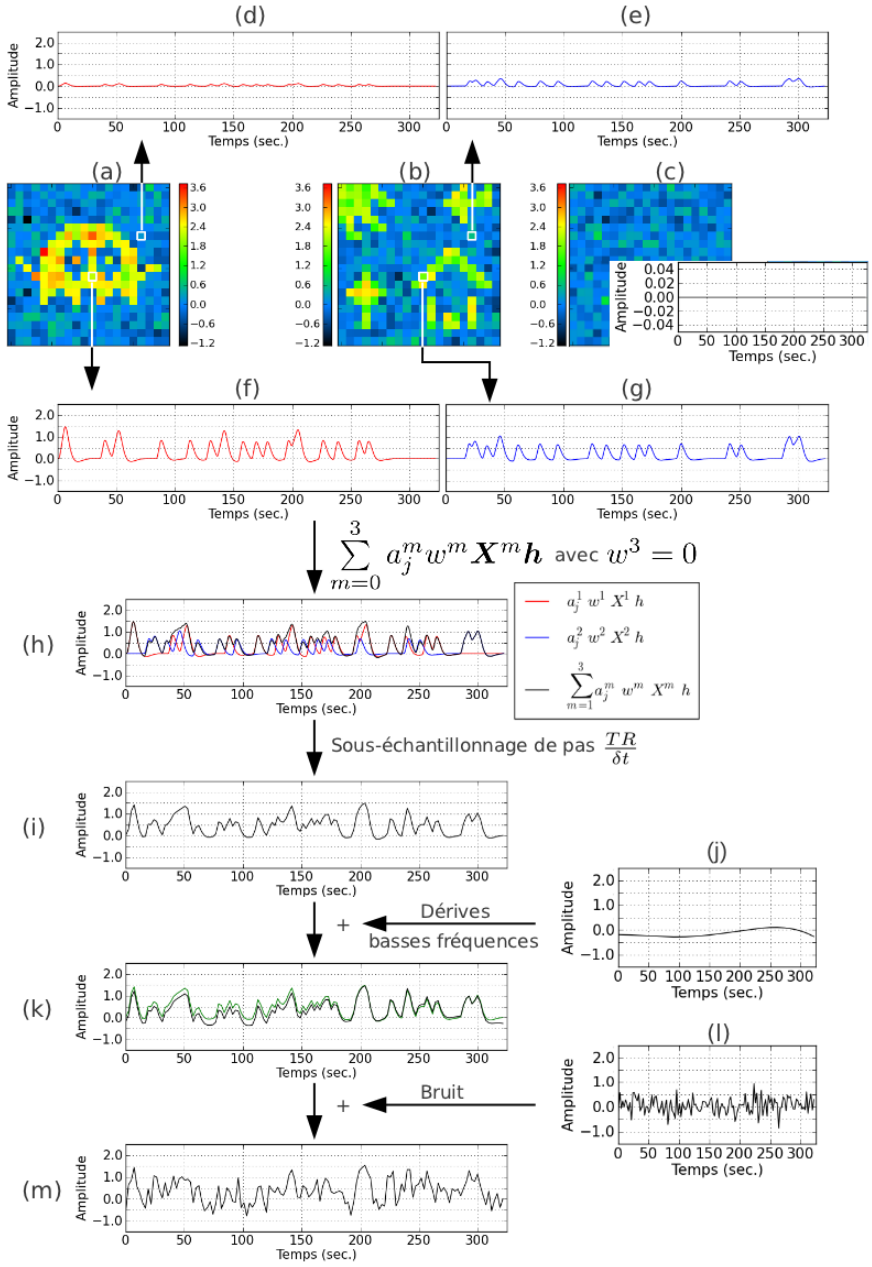


FIGURE IV.2 – Étapes de simulation du signal BOLD : signal stimulus-induit $a_j^m w^m \boldsymbol{X}^m \boldsymbol{h}$ en un pixel activé (f,g) et inactivé (d,e) des deux conditions pertinentes $m = 1, 2$ dont les NRNs sont représentés dans (a,b) respectivement. (c) NRNs de la condition non pertinente $m = 3$ et le signal stimulus-induit nul pour chacun de ses pixels car $w^3 = 0$. (h) Signal stimulus-induit total en noir représentant la somme des signaux rouge $a_j^1 w^1 \boldsymbol{X}^1 \boldsymbol{h}$ et bleu $a_j^2 w^2 \boldsymbol{X}^2 \boldsymbol{h}$. (i) Sous échantillonnage du signal stimulus-induit total d'un pas $TR/\Delta t = 4$ pour obtenir les points temporels correspondants aux scans. (k) Ajout des dérives basses fréquences (j) au signal stimulus-induit total représenté en vert afin de voir l'effet des dérives. (m) Ajout du bruit (l).

IV.1.2.1 Technique stochastique - MCMC

Cette simulation comporte $M = 3$ conditions expérimentales, 1 pertinente avec $\mu_1^1 = 2.5$ et 2 non pertinentes avec $\mu_1^2 = \mu_1^3 = 0$. Les moyennes des gaussiennes inactives sont fixées à 0 et les variances des gaussiennes actives v_1^m et inactives v_0^m ont été choisies à 0.1 pour toutes les conditions ($m = 1 : 3$). Les labels et NRNs simulés sont représentés par la figure IV.3 en première et deuxième ligne, respectivement. La FRH choisie est de forme non canonique avec un décalage et élargissement du pic (figure IV.4). Les dérives basses fréquences simulées sont $\boldsymbol{\ell}_j \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I_4})$ pour $j = 1 : 400$ et le bruit ϵ_j est un bruit blanc gaussien dont la variance est $\sigma_j^2 = \sigma^2 = 1$ pour tous les pixels j. Enfin la valeur du TR est de 2.4 secondes et celle du Δt est de 0.6 secondes.

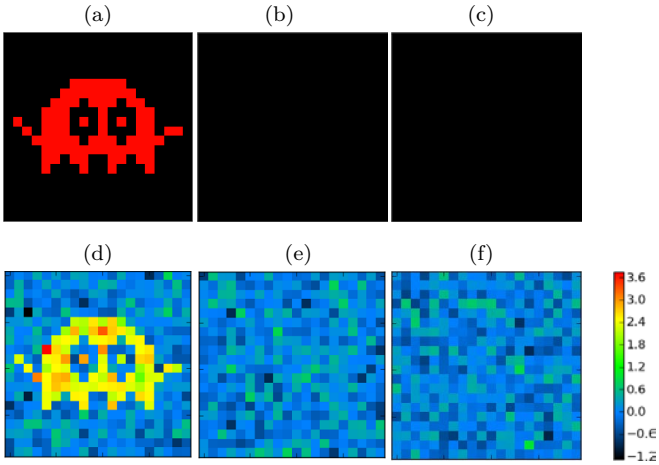


FIGURE IV.3 – Labels (en haut) et NRNs (en bas) simulés pour $M = 3$ conditions expérimentales dont une pertinente (colonne 1) et deux non pertinentes (colonnes 2 et 3).

Comme précisé dans le chapitre III, les paramètres de la sigmoïde sont fixés. Dans un premier temps nous allons fixer le paramètre τ_1 contrôlant la pente de la sigmoïde à 1 et nous préoccuper uniquement de τ_2 contrôlant l'abscisse de son point d'inflexion. Etant donné notre connaissance *a priori* des conditions non pertinentes nous allons suivre la procédure suivante pour déterminer la valeur de τ_2 suffisante à l'élimination des 2 conditions non pertinentes.

Procédure du choix de τ_2 :

1- Choisir $\tau_2 = 0$.

2- Traiter les données en question avec le modèle *parcimonieux 1* pour le τ_2 choisi.

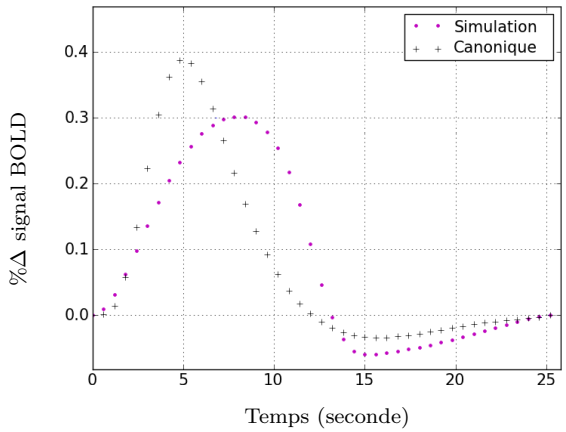


FIGURE IV.4 – FRH simulée (en violet) avec un décalage et un élargissement au niveau du pic par rapport à la forme canonique (en noir).

3- Si $w^m_{\text{estimé}} = 0$ pour les conditions non pertinentes (ici $m = 2, 3$), on conclut que la valeur précédente de τ_2 est suffisante. Sinon on augmente τ_2 et on revient à l'étape (2).

Cette procédure n'est pas idéale car elle nécessite de répéter l'analyse plusieurs fois, mais elle permet dans un premier temps d'étudier le comportement du modèle *parcimonieux 1*. Dans le cas de notre simulation, on obtient $\tau_2 = 45$, ce qui est équivalent à 11% de la taille de la ROI simulée (400 pixels).

Afin de vérifier l'intérêt d'éliminer les conditions non pertinentes par rapport au modèle *complet*, nous allons comparer la simulation et les estimations des deux modèles pour les labels, les PPM, et la FRH.

♣ Comparaison des labels

La figure IV.5 illustre les labels simulés (colonne 1) et l'estimation de leurs probabilités *a posteriori* d'appartenance à la classe active pour les modèles *complet* (colonne 2) et *parcimonieux 1* (colonne 3), ceci pour les trois conditions expérimentales $m = 1 : 3$. Les résultats d'estimation pour chaque condition sont illustrés sur deux lignes, celle du haut représente les probabilités *a posteriori* $p(q^m_j = 1 \,|\, \boldsymbol{y}_j)$ pour $j = 1$ à J, et celle du bas représente le seuillage de ces probabilités à 0.5. Notons que les deux modèles montrent un comportement équivalent par rapport à la condition pertinente $m = 1$ (figures IV.5-(b,c)). Le modèle *complet* souffre cependant d'un nombre élevé de faux positifs dans le cas des conditions non pertinentes $m = 2$ (figures IV.5-(e,f)) et $m = 3$ (figures IV.5-(h,i)).

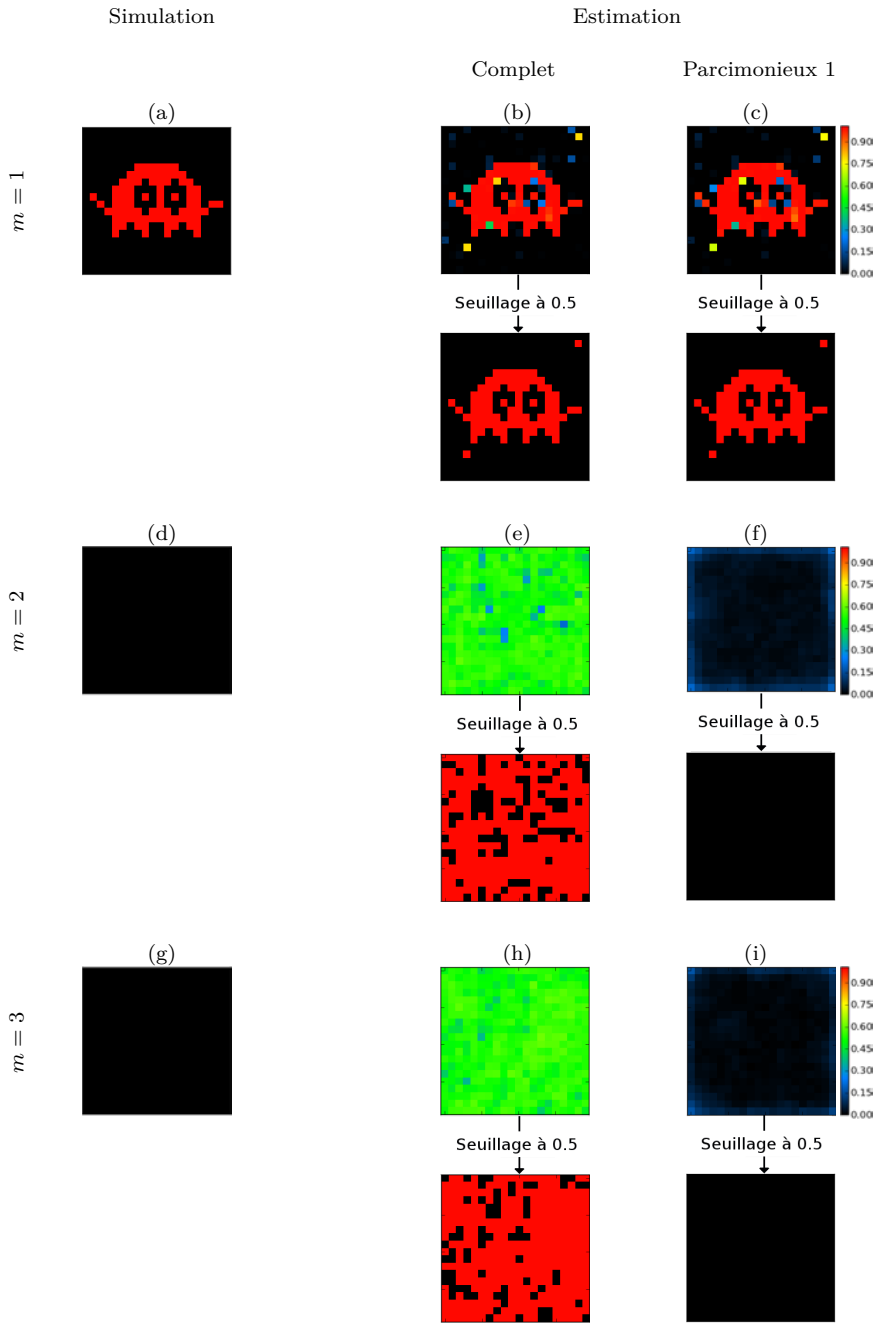


FIGURE IV.5 – Comparaison entre les labels simulés (colonne 1) et l'estimation de leurs probabilités *a posteriori* d'appartenance à la classe active pour les modèles *complet* (colonne 2) et *parcimonieux 1* (colonne 3). Chaque ligne représente une des trois conditions. Notons que les deux estimateurs ont des comportements équivalents dans le cas de la condition pertinente ($m = 1$) contrairement au cas des conditions non pertinentes ($m = 2, 3$) où le modèle *complet* souffre d'un nombre élevé de faux positifs.

♣ Comparaison des cartes de probabilité *a posteriori*- PPMs

Rappelons qu'en inférence bayésienne, les cartes de probabilité *a posteriori* permettent la comparaison entre différents modèles. Dans le cadre de l'approche DEC, ces probabilités correspondent à :

$$PPM_{j,m}^{\gamma} = p(\hat{a}_j^m \geq \gamma \,|\, \boldsymbol{y}_j) = p\left(\frac{\hat{a}_j^m - m_{a_j^m}}{\sqrt{V_{\boldsymbol{a}_j^{m,m}}}} \geq \frac{\gamma - m_{a_j^m}}{\sqrt{V_{\boldsymbol{a}_j^{m,m}}}} \,|\, \boldsymbol{y}_j\right) \quad ; \quad \forall j = 1 : J \tag{IV.2}$$

Ici, nous avons effectué une procédure qui sera désignée par la suite par la *procédure PPMs* afin de comparer nos deux modèles.

Procédure PPMs :

1- Calcul des PPM^{γ} en faisant varier γ de 0 jusqu'à $\gamma_{\max}$ où plus aucun voxel n'est détecté.

2- Seuillage des PPM^{γ} ($PPM_{j,m}^{\gamma} \geq \alpha$) pour un α choisi.

3- Calcul du nombre de voxels N^{γ} obtenus pour $\gamma = 0 : \gamma_{\max}$

4- Construction de la courbe (γ, N^{γ})

Ici, avec $\gamma_{\max} = 3$, la procédure PPMs nous permet d'obtenir la figure IV.6-(a) où le modèle *complet* est représenté en vert et le modèle *parcimonieux 1* en rouge. On remarque que pour des valeurs de $\gamma \in \{1.8, 2.4\}$ le modèle *parcimonieux 1* détecte plus de voxels que le *complet*.

Une fois les PPMs calculées et seuillées, il est nécessaire de regarder l'emplacement des voxels obtenus afin de vérifier que ce ne sont pas des faux positifs ou faux négatifs. Pour cela, nous avons comparé les résultats des PPMs seuillées à $\alpha = 0.95$ avec les labels et determiner les faux positifs (FP), faux négatifs (FN), vrais positifs (VP) et vrais négatifs (VN) que nous illustrons par la figure IV.6-(c,d) pour le cas $\gamma = 1.9$ où le modèle *complet* détecte 39 pixels par rapport à 49 détectés par le modèle *parcimonieux 1*. Finalement la figure IV.6-(b) représente l'histogramme des $PPM^{1.9}$ et confirme le résultat précédant en indiquant l'augmentation du nombre de pixels correspondant aux $PPM^{1.9} \geq 0.95$ pour le modèle *parcimonieux 1*.

♣ Comparaison des FRHs

D'un point de vue estimation de la FRH, les modèles *complet* et *parcimonieux*

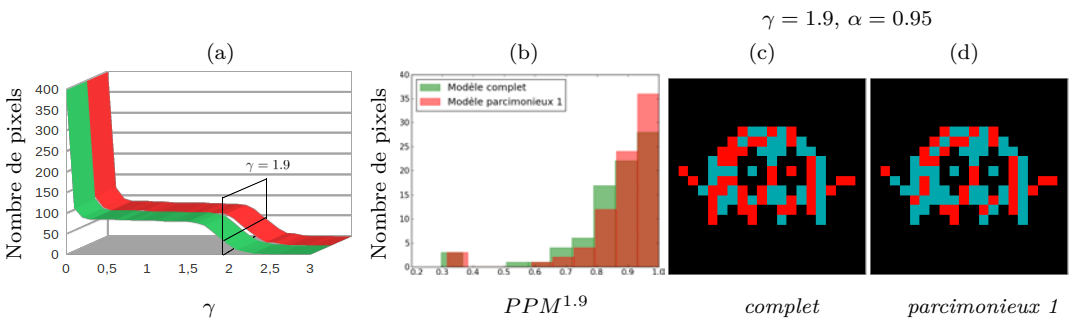


FIGURE IV.6 – Comparaison inter-modèles pour la condition pertinente $m = 1$: (a) Nombre de pixels obtenu après seuillage des PPM^{γ} ($p(\hat{a}_j^m \geq \gamma \,|\, \boldsymbol{y}_j) \geq \alpha$ pour $j = 1 : J$) avec $\gamma = 0 : 3$ et $\alpha = 0.95$. Notons que le modèle *parcimonieux 1* (en rouge) détecte plus de pixels que le modèle *complet* (en vert) pour des valeurs de γ entre 1.8 et 2.4. Pour $\gamma = 1.9$: (c-d) Illustrent les VPs (en bleu), VNs (en noir), FPs (en vert) et FNs (en rouge) dans les cartes $PPM^{1.9} \geq 0.95$ où le modèle *complet* (c) souffre d'un nombre plus élevé de FNs et moins élevé de VPs par rapport au *parcimonieux 1* (d). Cela est induit par des $PPM^{1.9}$ moins fortes pour le premier, ce qui est confirmé par l'histogramme (b) illustrant les $PPM^{1.9}$ des modèles *parcimonieux 1* (en rouge) et *complet* (en vert) pour les voxels activés ($q^m_{j,\text{simulé}} = 1$).

1 ne montrent aucune différence de comportement comme le confirme la figure IV.7 qui compare les estimations parfaitement superposées des deux modèles (*complet* en vert et *parcimonieux 1* en rouge) à la FRH simulée (en violet).

Par la suite nous allons nous intéresser à l'inférence VEM permettant un gain en temps de calcul comme indiqué dans le chapitre III.

IV.1.2.2 Technique variationnelle - VEM

Rappelons que dans le cadre de l'inférence VEM, nous avons développé deux modèles *parcimonieux*. Le premier, désigné par la suite par *parcimonieux 1*, est équivalent à celui développé en MCMC et le deuxième, désigné par la suite par *parcimonieux 2*, propose une autre définition de la pertinence où $\boldsymbol{w}$ est connecté à $\boldsymbol{\mu}_1$.

Pour la même simulation que précédemment, $\tau_2 = 0$ est suffisant pour la détection des deux conditions non pertinentes dans le cadre d'une analyse basée sur l'algorithme VEM et le modèle *parcimonieux 1*. Cette valeur de τ_2 est moins élevée que celle observée en MCMC. Cela pourrait être dû au fait qu'en utilisant une méthode MCMC, on manipule des simulations de variables aléatoires alors qu'en VEM on manipule leurs moments. En effet, en revenant aux équations permettant le calcul des probabilités *a posteriori* d'obtenir un w^m égal à 1 dans

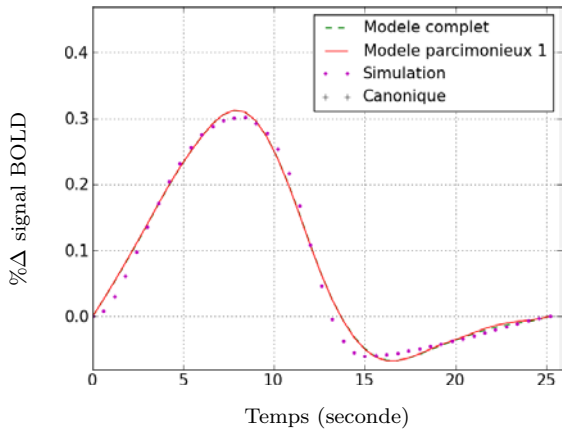


FIGURE IV.7 – Similarité des FRHs estimées par les modèles *parcimonieux 1* (en rouge) et *complet* (en vert, parfaitement superposée à la courbe rouge). La FRH simulée est représentée en violet.

les deux techniques d'inférence, on s'aperçoit que l'on peut les écrire de la manière suivante :

$$\text{MCMC} : p\left(w^m = 1 \,|\, \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{w}^{\backslash m}, \boldsymbol{q}^m, \boldsymbol{h}, \boldsymbol{\sigma}^2, \boldsymbol{\theta}_m, \boldsymbol{\tau}\right) = \left[1 + \frac{\exp(A)}{\exp(B)}\right]^{-1}$$

$$\text{VEM} : \widetilde{p}_{w^m}(w^m = 1) = \left[1 + \frac{\exp(A')}{\exp(B')}\right]^{-1}$$

où $A = -\tau_1(\sum_j q_j^m - \tau_2)$ et $A' = E_{\widetilde{p}}[A] = -\tau_1(\sum_j \widetilde{p}_{q_j^m}(1) - \tau_2)$ et où les termes B et B' dépendent des données ainsi que des autres variables et paramètres du modèle avec $B' = E_{\widetilde{p}}[B]$. Notons que $\widetilde{p}$ désigne le raccourci de la loi $\widetilde{p}(\boldsymbol{A}, \boldsymbol{Q}, \boldsymbol{w}, \boldsymbol{h})$ qui représente l'approximation de la loi *a posteriori* en VEM. Le terme B' étant la moyenne de B, il est possible d'avoir des cas où $B > B'$, ce qui nécessite d'augmenter A par rapport à A' afin d'avoir la même probabilité d'obtenir un $w^m = 1$ dans MCMC et VEM. On a donc :

$$\begin{aligned}
& A > A' \\
\iff & -\tau_1\left(\sum_j q_j^m - \tau_{2,MCMC}\right) > -\tau_1\left(\sum_j \widetilde{p}_{q_j^m}(1) - \tau_{2,VEM}\right) \\
\iff & \sum_j q_j^m - \tau_{2,MCMC} < \sum_j \widetilde{p}_{q_j^m}(1) - \tau_{2,VEM}
\end{aligned}$$

ce qui donne en supposant que les quantités $\sum_j q_j^m$ et $\sum_j \widetilde{p}_{q_j^m}(1)$ sont proches :

$$\tau_{2,MCMC} > \tau_{2,VEM}$$

La problématique est un peu plus complexe mais la précédente discussion pourrait être une cause éventuelle de la différence entre MCMC et VEM par rapport à la valeur de τ_2 nécessaire à la détection des conditions non pertinentes.

Nous allons par la suite compliquer la simulation en lui ajoutant une deuxième condition pertinente représentant une désactivation avec $\mu_1^2 = -3.4$ et $v_1^m = v_0^m = 0.1$, ce qui revient à simuler $M = 4$ conditions expérimentales. Cet ajout permet de voir le comportement du modèle *parcimonieux* dans le cas où plusieurs conditions pertinentes sont présentes ainsi que dans le cas d'une désactivation. Les labels et NRNs simulés sont représentés par la figure IV.8 en première et deuxième ligne respectivement. Comme pour la simulation précédente, la FRH choisie est de forme non canonique avec un décalage et élargissement du pic (figure IV.4), les dérives basses fréquences simulées sont $\boldsymbol{\ell}_j \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I_4})$ pour $j = 1 : 400$ et le bruit $\boldsymbol{b}_j$ est un bruit blanc gaussien. Ici, la variance du bruit a la même valeur σ^2 pour tous les pixels et elle varie de 0.5 à 2.5 avec un pas de 0.5 permettant ainsi de voir le comportement des modèles (*parcimonieux 1*, *parcimonieux 2* et *complet*) avec l'augmentation du bruit.

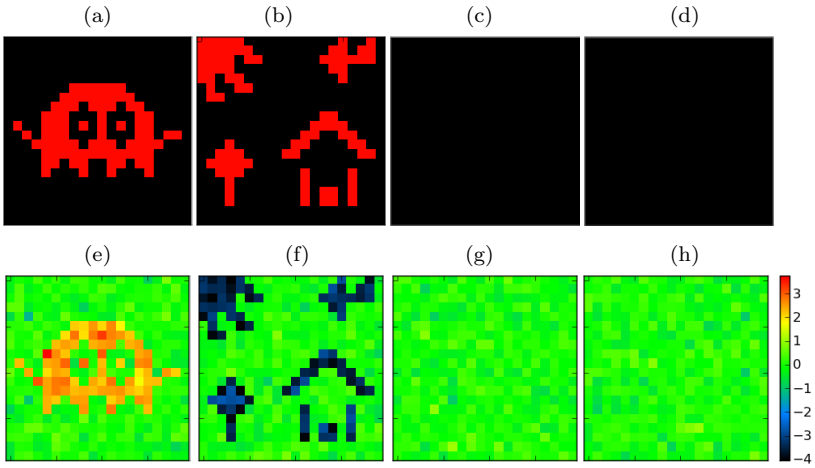


FIGURE IV.8 – Labels (en haut) et NRNs (en bas) simulés pour $M = 4$ conditions expérimentales dont deux pertinentes (colonnes 1 et 2) et deux non pertinentes (colonnes 3 et 4). L'une des conditions pertinentes (colonne 1) représente un activation (NRNs de moyenne positive) et l'autre (colonne 2) représente une déactivation (NRNs de moyenne négative).

Le nombre d'itérations minimum est de 100 avec deux critères d'arrêt : l'énergie libre et le produit $\boldsymbol{Ah}$ qui converge plus rapidement que chacune des deux variables $\boldsymbol{A}$ et $\boldsymbol{h}$ étant donné le problème de bilinéarité entre ces deux variables (la même solution peut être obtenue avec $(\boldsymbol{A}_1, \boldsymbol{h}_1)$ et $(s\boldsymbol{A}_1, \boldsymbol{h}_1/s)$ où s représente un facteur d'échelle). Ces critères consistent en la comparaison entre les quantités estimées sur deux itérations successives. L'algorithme s'arrête si la norme quadratique de l'erreur relative est inférieure à 10^{-5}. Pour le produit $\boldsymbol{Ah}$, par exemple le critère d'arrêt s'écrit sous la forme :

$$\frac{||\boldsymbol{Ah}^{(r)} - \boldsymbol{Ah}^{(r-1)}||^2}{||\boldsymbol{Ah}^{(r-1)}||^2} \leq 10^{-5} \tag{IV.3}$$

où $\boldsymbol{Ah}$ est de dimension $(M \times J) \times (D+1)$ avec $\boldsymbol{Ah} = \{a_j^m \boldsymbol{h}\ ; j = 1 : J\ , m = 1 : M\}$.

Comme pour l'inférence MCMC, le paramètre τ_1 de la sigmoïde est fixé à 1. Pour le modèle *parcimonieux 1*, le paramètre τ_2 est choisi avec la même procédure détaillée précédemment où une valeur de $\tau_2 = 0$ est suffisante pour éliminer les conditions non pertinentes. Ce dernier étant estimé dans le modèle *parcimonieux 2*, il nécessite le choix des hyperparamètres de son *a priori* gamma et donc de α_{τ_2} et λ_{τ_2} dans l'équation (III.71). Pour cela nous avons choisi $\alpha_{\tau_2} = 3$ et $\lambda_{\tau_2} = 4$, ce qui correspond à une loi gamma de mode relativement petit égal à 0.5 et de variance 0.2 limitant ainsi les grosses valeurs de τ_2 pouvant influencer les conditions pertinentes et les petites valeurs de τ_2 insuffisantes pour la détection de la non pertinence. Les deux modèles *parcimonieux* détectent les conditions non pertinentes avec des probabilités *a posteriori* $\widetilde{p}_{W^m}(w^m = 1)$ proches de 0 et les pertinentes avec des probabilités *a posteriori* $\widetilde{p}_{W^m}(w^m = 1)$ proches de 1, τ_2 étant estimé à 0.45 dans le modèle *parcimonieux 2*. Afin de vérifier l'intérêt de cette sélection par rapport au modèle *complet*, plusieurs aspects ont été étudiés permettant de comparer les modèles entre eux (PPMs, énergie libre) ainsi qu'avec la vérité terrain (Intervalle de crédibilité, erreur quadratique moyenne et courbe ROC).

♣ Courbes ROC

Les courbes ROC (*Receiver Operating Characteristics*) permettent dans notre cas de caractériser l'estimateur des labels en le comparant avec les labels simulés

afin de détecter, pour un seuil variable, les bonnes et mauvaises classifications. Ceci en se basant sur 4 quantités : vrais positifs (VP), vrais négatifs (VN), faux positifs (FP) et faux négatifs (FN) dont la définition est rappelée dans le tableau ci-dessous.

	$q^m_{j,\text{ simulé}} = 1$	$q^m_{j,\text{ simulé}} = 0$
$q^m_{j,\text{ estimé}} = 1$	VP	FP
$q^m_{j,\text{ estimé}} = 0$	FN	VN

Une courbe ROC exprime le taux de vrais positifs $\frac{VP}{VP+FN}$ (sensibilité) en fonction du taux de faux positifs $\frac{FP}{FP+VN}$ (1-spécificité). Pour un taux de faux positifs fixé, plus le taux de vrais positifs est élevé, meilleur est le détecteur. Nous avons comparé les courbes ROC obtenues pour les modèles *complet*, *parcimonieux 1* et *parcimonieux 2* pour les deux conditions pertinentes, $m = 1$ et $m = 2$. Cela pour des niveaux de bruit différents ($\sigma^2 = 0.5 : 2.5$ par pas de 0.5). Aucune différence significative n'est observée entre les modèles indiquant ainsi le même comportement de l'estimateur des labels pour les conditions pertinentes. La figure IV.9 illustre cette comparaison pour le cas $m = 1$ et $\sigma^2 = 0.5$ et confirme la similarité des modèles *parcimonieux 1* (en rouge), *parcimonieux 2* (en bleu) et *complet* (en vert). L'illustration n'a pas été faite pour la deuxième condition pertinente ou pour les autres niveaux de bruit étant donné la similarité observée entre les courbes ROC correspondant aux trois modèles.

Ceci est également confirmé par les deux premières lignes de la figure IV.10 illustrant, pour le cas $\sigma^2 = 0.5$ et pour les deux conditions pertinentes $m = 1, 2$, les labels simulés (colonne 1) et leurs probabilités *a posteriori* d'appartenance à la classe active pour le modèle *complet* (colonne 2), *parcimonieux 1* (colonne 3) et *parcimonieux 2* (colonne 4). A l'inverse, les modèles diffèrent entre eux dans le cas des conditions non pertinentes où un nombre élevé de FPs est observé pour le modèle *complet* (figures IV.10-(j,n)), ceci étant dû à la classification aléatoire induite par le phénomène de superposition des classes d'activation ($\widetilde{p}_{q^m_j}(q^m_j = 1) \approx 0.5$). La connexion entre $\boldsymbol{w}$ et $\boldsymbol{Q}$ permet au modèle *parcimonieux 1* de maitriser l'estimation des labels non pertinents (figures IV.10-(k,o)) et dont l'estimation est mauvaise ($\widetilde{p}_{q^m_j}(q^m_j = 1) = 0.5$) dans le modèle *parcimonieux 2* (figures IV.10-(l,p)). Cela n'étant pas un problème étant donné la présence de $\boldsymbol{w}$ qui indique la non pertinence de ces labels.

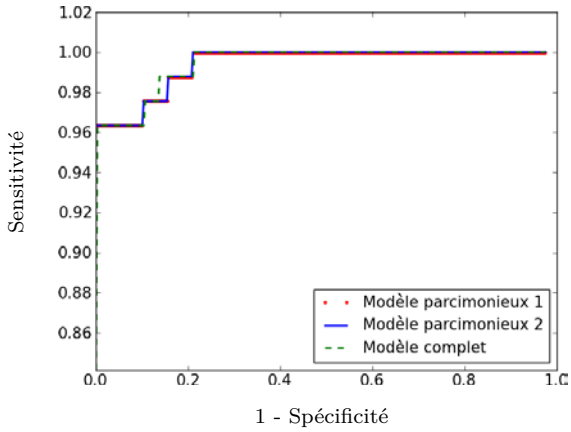


FIGURE IV.9 – Illustration de la similarité entre les courbes ROC des modèles *parcimonieux 1* (en rouge), *parcimonieux 2* (en bleu) et *complet* (en vert) pour la première condition pertinente $m = 1$ et un niveau de bruit correspondant à $\sigma^2 = 0.5$. Cela reflète la similarité de comportement des modèles par rapport à l'estimation des labels. Le comportement étant le même pour les autres niveaux de bruit et pour la deuxième condition pertinente $m = 2$, leurs courbes n'ont pas été illustrées.

♣ Intervalle de crédibilité

D'après la loi *a posteriori* de la variable $\boldsymbol{A}$ (équation (III.49)), on peut déduire qu'en chaque voxel j, $\hat{a}_j^m \sim \mathcal{N}(m_{a_j^m}, V_{\boldsymbol{a}_j^{m,m}})$. On peut donc calculer l'intervalle $IC_\alpha = \left[\, m_{a_j^m} \pm \delta \right]$ où :

$$\begin{aligned}
\widetilde{p}\left(\hat{a}_j^m \in [m_{a_j^m} \pm \delta]\right) &= \widetilde{p}\left(m_{a_j^m} - \delta \leq \hat{a}_j^m \leq m_{a_j^m} + \delta\right) \\
&= \widetilde{p}\left(\frac{-\delta}{\sqrt{V_{\boldsymbol{a}_j^{m,m}}}} \leq \frac{\hat{a}_j^m - m_{a_j^m}}{\sqrt{V_{\boldsymbol{a}_j^{m,m}}}} \leq \frac{\delta}{\sqrt{V_{\boldsymbol{a}_j^{m,m}}}}\right) \\
&= \widetilde{p}\left(X \leq \frac{\delta}{\sqrt{V_{\boldsymbol{a}_j^{m,m}}}}\right) - \widetilde{p}\left(X \leq \frac{-\delta}{\sqrt{V_{\boldsymbol{a}_j^{m,m}}}}\right) \\
&= \mathcal{F}\left(\frac{\delta}{\sqrt{V_{\boldsymbol{a}_j^{m,m}}}}\right) - \mathcal{F}\left(\frac{-\delta}{\sqrt{V_{\boldsymbol{a}_j^{m,m}}}}\right) = \alpha, \qquad \text{(IV.4)}
\end{aligned}$$

où $X = \frac{\hat{a}_j^m - m_{a_j^m}}{\sqrt{V_{\boldsymbol{a}_j^{m,m}}}}$ et $\mathcal{F}$ est la fonction de répartition de la loi $\mathcal{N}(0,1)$. Notons S_α^m le nombre de voxels tels que $a_{j,\text{simulé}}^m \in IC_\alpha$ pour $j \in \mathcal{P}_\gamma$. La table de $\mathcal{F}$ nous permet d'obtenir δ pour un α donné et donc de calculer l'intervalle de crédibilité qui permettra ensuite de calculer la quantité S_α^m. La figure IV.11 illustre les

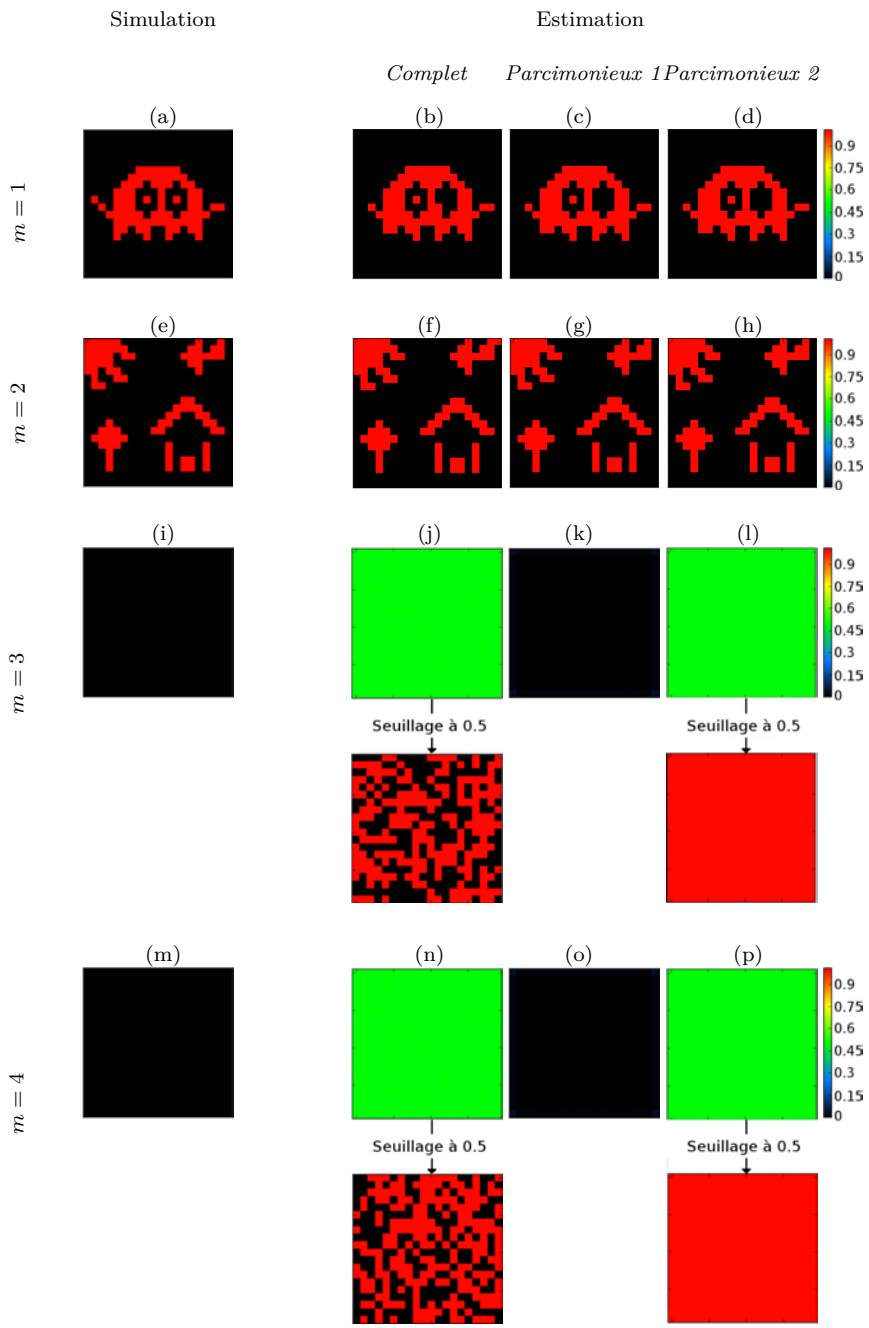


FIGURE IV.10 – Comparaison, pour un niveau de bruit $\sigma^2 = 0.5$, entre les labels simulés (colonne 1) et l'estimation de leurs probabilités *a posteriori* d'appartenance à la classe active pour les modèles *complet* (colonne 2), *parcimonieux 1* (colonne 3) et *parcimonieux 2* (colonne 4). Chaque ligne représente une des quatre conditions. Notons que les trois estimateurs ont des comportements équivalents dans le cas des conditions pertinentes ($m = 1, 2$) contrairement au cas des conditions non pertinentes ($m = 3, 4$) où le modèle *complet* souffre d'un nombre élevé de FPs. La mauvaise estimation des labels par le modèle *parcimonieux 2* ne représente pas un problème étant donné la présence de w^m estimée à 0 pour $m = 3, 4$ indiquant ainsi la non significativité de l'estimation des labels correspondants.

valeurs de S_α^m en fonction de $\alpha = 0 : 1$, ceci pour les deux conditions pertinentes, $m = 1$ à gauche et $m = 2$ à droite et pour les voxels activés uniquement (où $q_{j,\text{simulé}}^m = 1$). Les différentes lignes correspondent à des niveaux de bruit différents ($\sigma^2 = 0.5 : 2.5$ par pas de 0.5).

On remarque que le modèle *complet* (en vert) est toujours meilleur ou équivalent aux modèles *parcimonieux 1* (en rouge) et *2* (en bleu), dans le sens où il correspond à un nombre plus élevé de pixels réalisant la condition $a_{j,\text{simulé}}^m \in IC_\alpha$. Pour cela, nous allons étudier plus en détails les voxels problématiques des modèles *parcimonieux* pour des valeurs de RSB et α précises. Etant donné la similarité entre les deux modèles *parcimonieux*, concentrons nous sur le modèle *2* par exemple, pour $\sigma^2 = 0.5$ et $\alpha = 0.8$, un des cas où la différence avec le *complet* est la plus élevée (figure IV.11, première ligne). Cette différence concerne néanmoins un petit nombre de pixels (8 pour $m = 1$ et 15 pour $m = 2$ sur 400 pixels). La figure IV.12 illustre les moyennes estimées $m_{a_j^m}$ et les $IC_{0.8}$ correspondants. Il est clair que les intervalles de crédibilité sont plus larges pour le modèle *complet* indiquant des variances plus élevées par rapport au modèle *parcimonieux* ce qui justifie le fait que la valeur simulée du NRN en question (en rouge) soit incluse dans l'intervalle correspondant au modèle *complet* et non dans celui correspondant au modèle parcimonieux.

♣ Erreur quadratique moyenne - ESE

L'ESE (*Expected Squared Error*) permet de comparer une estimation à la réalité terrain. Elle consiste à calculer, en chaque pixel j, la quantité suivante :

$$ESE_{a_j^m} = E_{\widetilde{p}_{a_j^m}}[L(a_j^m)] = \int_{a_j^m=-\infty}^{\infty} L(a_j^m)\, \mathcal{N}(a_j^m; m_{a_j^m}, V_{\boldsymbol{a}_j^{m,m}})\, da_j^m \tag{IV.5}$$

Dans le cas d'un coût quadratique $L(a_j^m) = (a_j^m - a_{j,simu}^m)^2$ on a :

$$ESE_{a_j^m} = E_{\widetilde{p}_{a_j^m}}[(a_j^m - a_{j,simu}^m)^2] = V_{\boldsymbol{a}_j^{m,m}} + (m_{a_j^m} - a_{j,simu}^m)^2 \tag{IV.6}$$

Les histogrammes de la figures IV.13 illustrent les quantités $ESE_{a_{j,c}^m} - ESE_{a_{j,p_1}^m}$ en bleu et les quantités $ESE_{a_{j,c}^m} - ESE_{a_{j,p_2}^m}$ en vert, où l'indice c désigne le modèle *complet* et les indices p_1 et p_2 désignent les modèles *parcimonieux 1* et *2*, respectivement. Cela pour la première condition pertinente $m = 1$ (à gauche) et la deuxième condition pertinente $m = 2$ à droite. Chaque ligne de la figure IV.13 correspond à un niveau de bruit ($\sigma^2 = 0.5 : 2.5$ par pas de 0.5). Notons que

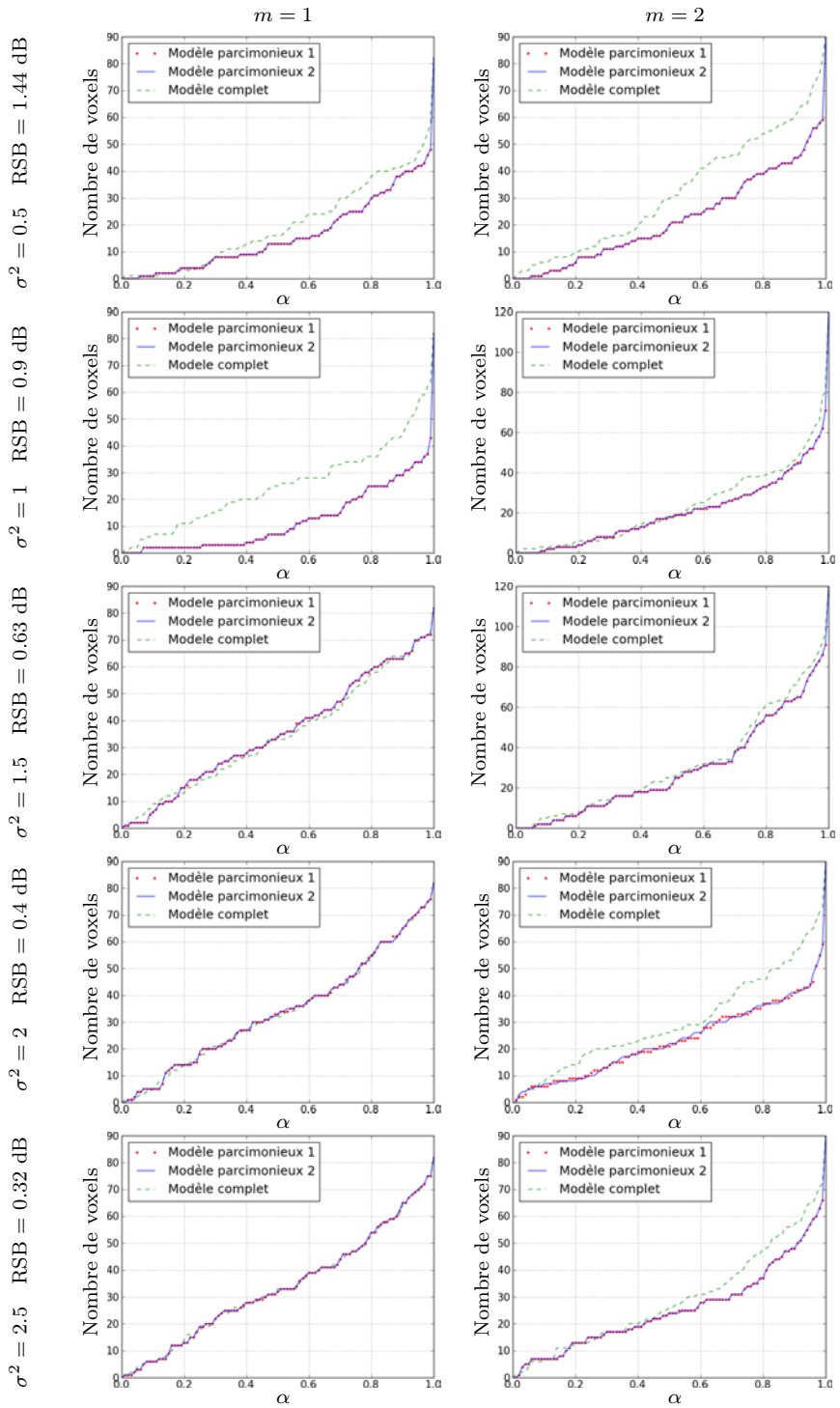


FIGURE IV.11 – Représentation de S_α^m, le nombre de pixels pour lesquels $a_{j,\text{simulé}}^m \in IC_\alpha$ pour $j \in \mathcal{P}_\gamma$, où IC_α est l'intervalle de crédibilité correspondant. Ceci pour chacunes des deux conditions pertinentes, $m = 1$ (à gauche) et $m = 2$ (à droite) et pour les trois modèles : *parcimonieux 1* (en rouge), *parcimonieux 2* (en bleu) et *complet* (en vert). Chaque ligne correspond à un niveau de bruit différent. Notons que dans certains cas $a_{j,\text{simulé}}^m \in IC_\alpha$ pour le modèle *complet* et non pour les modèles *parcimonieux 1,2*, ceci étant limité à une dizaine de pixels maximum.

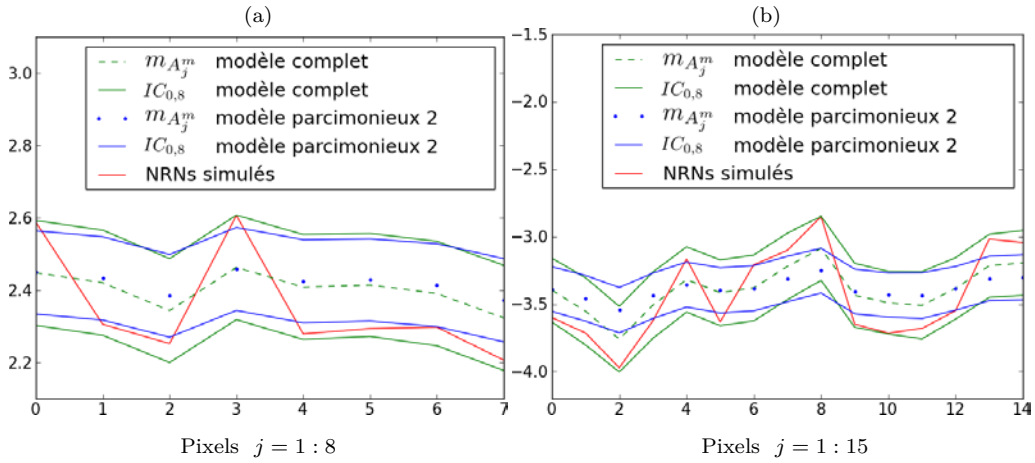


FIGURE IV.12 – Moyennes estimées des NRNs (en pointillé) et Intervalles de crédibilité pour $\alpha = 0.8$ ($IC_{0.8}$, en continu) pour les voxels dont les valeurs des NRNs simulés (en rouge) ne sont pas incluses dans $IC_{0.8}$ pour le modèle *parcimonieux 2* (en bleu) (le *1* ayant le même comportement) contrairement au modèle *complet* (en vert). Cela pour chacune des deux conditions pertinentes $m = 1$ (a) et $m = 2$ (b). Notons que ceci est dû à la variance des NRNs moins élevée pour le modèle *parcimonieux* induisant un $IC_{0.8}$ moins large par rapport au modèle *complet*.

les modèles *parcimonieux* ont des comportements assez similaires (superposition des histogrammes bleu et vert) et que les différences d'erreur entre les modèles *complet* et *parcimonieux* sont faibles car elles appartiennent globalement à l'intervalle $[-0.05, 0.05]$. La plupart de ces différences sont aux alentours de 0. Pour les différences non nulles on remarque, que dans certains cas il y a autant de différences négatives que positives et que dans d'autres les modèles *parcimonieux* induisent des ESE un peu moins élevées que le modèle *complet*, surtout pour les deux dernières lignes correspondantes à un faible RSB.

♣ Estimation de la FRH

La figure IV.14 compare les FRHs estimées entre le modèle *complet* (en vert) et chacun des modèles *parcimonieux 1* (en rouge) et *parcimonieux 2* (en bleu) ceci pour des niveaux de bruit différents correspondant à une variance de bruit σ^2 allant de 0.5 à 2.5 par pas de 0.5. Les estimations sont en plus, comparées à la FRH simulée (en violet) différente de la FRH canonique (en noir). Notons que les trois modèles fournissent une estimation assez proche de la FRH simulée et ne montrent aucune différence significative entre eux.

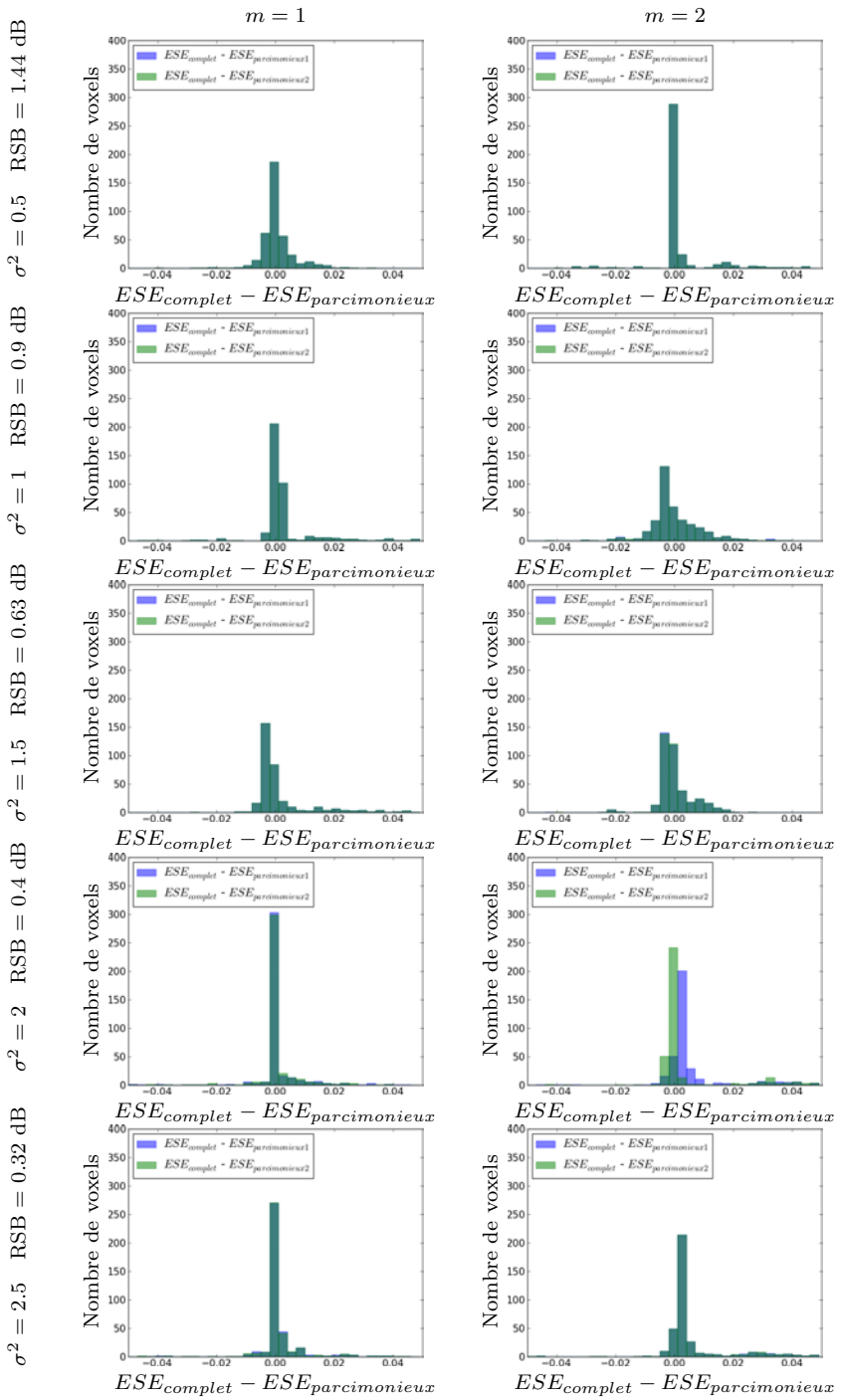


FIGURE IV.13 – Histogrammes illustrant les différences d'erreur quadratique moyenne (ESE) entre les modèles *complet* et *parcimonieux 1* (en bleu) et entre les modèles *complet* et *parcimonieux 2* (en vert). Cela pour les deux conditions pertinentes $m = 1$ (à gauche) et $m = 2$ (à droite). Chaque ligne correspond à un niveau de bruit. Notons que les deux modèles *parcimonieux* ont des comportements similaires (histogrammes bleu et vert superposés) et que les différences des ESE avec le modèle *complet* sont faibles (la plupart sont autour de 0) mais avec plus de valeurs positives que négatives confirmant ainsi des ESE moins élevées pour les modèles *parcimonieux* (cas où $\sigma^2 = 2$ et 2.5).

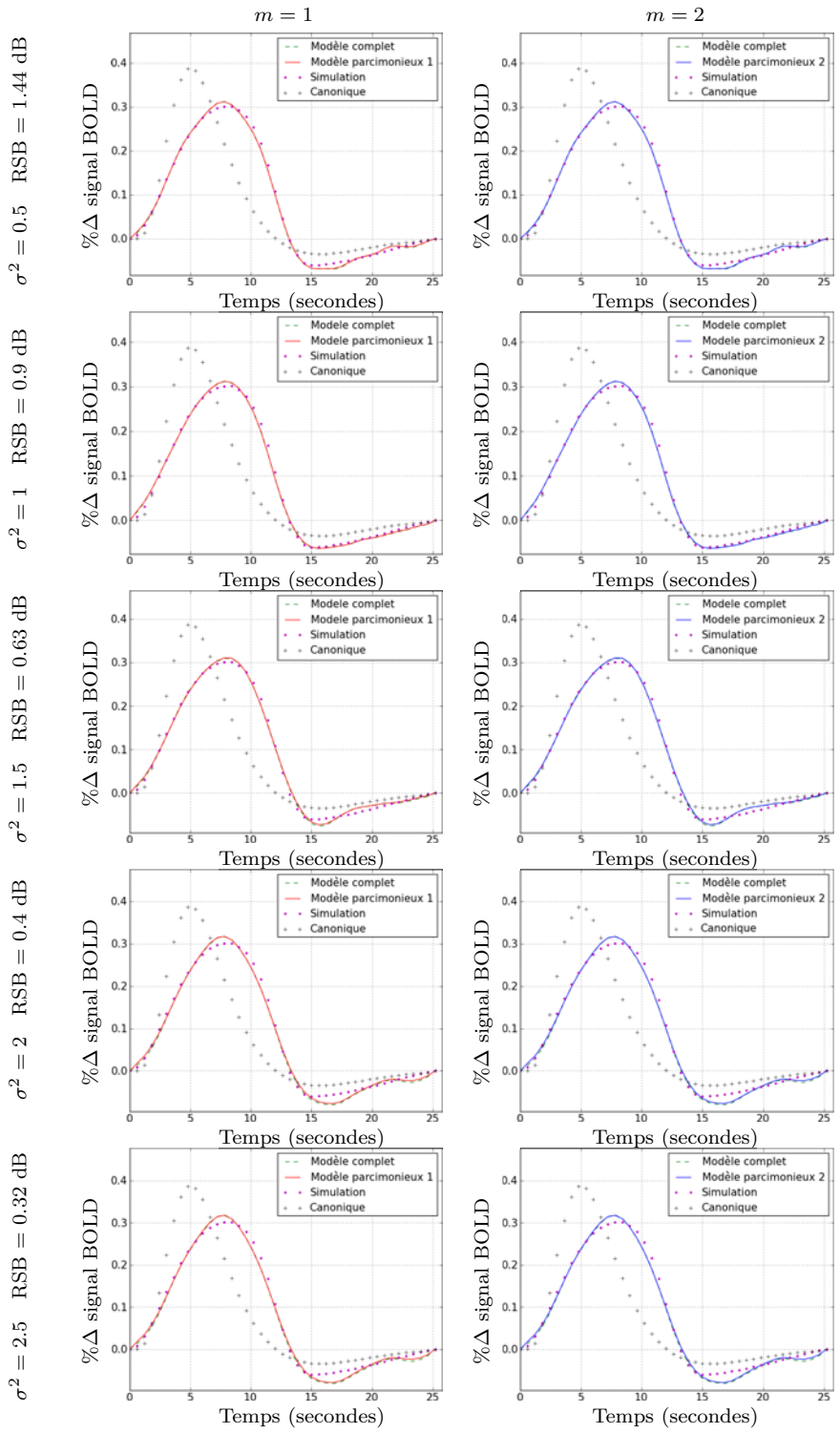


FIGURE IV.14 – Comparaison entre la FRH simulée (en violet) et les FRHs estimées par les modèles *complet* (en vert) et *parcimonieux* (modèle *1* en rouge à gauche et modèle *2* en bleu à droite). La FRH canonique est également illustrée en noir pour montrer la différence avec celle simulée. Chaque ligne correspond à un niveau de bruit différent. Aucune différence significative n'est observée indiquant une similarité entre les modèles.

♣ Cartes de probabilité *a posteriori*- PPMs

Ici, nous avons effectué la *procédure PPMs* avec $\gamma = 0 : 4$ (aucun voxel n'est détecté au delà de 4) et $\alpha \in \{0.9, 0.95, 0.99\}$ pour obtenir les courbes illustrées par la figure IV.15 pour $m = 1$ et par la figure IV.16 pour $m = 2$. Le modèle *complet* est représenté en vert, le modèle *parcimonieux 1* en rouge et le modèle *parcimonieux 2* en bleu.

On remarque que les PPMs des modèles *parcimonieux* sont toujours plus élevées ou similaires à celles du modèle *complet*. Cela induit dans le cas *parcimonieux* et pour des seuils γ et α donnés, un nombre d'activations supérieur ou équivalent à celui obtenu dans le cas *complet*. Notons que l'amélioration obtenue par la suppression des conditions non pertinentes est répartie entre les deux conditions pertinentes (elle peut être plus forte sur l'une ou sur l'autre).

En nous focalisant sur le cas $\sigma^2 = 0.5$, $\gamma = 2.2$ et $m = 1$ par exemple, on remarque qu'un seuillage à $\alpha \in \{0.9, 0.95, 0.99\}$ donne respectivement 69, 56, 9 pixels dans le cas *complet* par rapport à 79, 79, 57 dans le cas *parcimonieux*. Ces différences sont également mises en évidence par l'histogramme des $PPM^{2.2}$ pour $m = 1$ (figure IV.17-(g)). De même pour $\sigma = 0.5$, $\gamma = 3$ et $m = 2$, un seuillage à $\alpha = 0.9, 0.95, 0.99$ donne respectivement 72, 53, 28 pixels dans le cas *complet* par rapport à 90, 90, 78 dans le cas *parcimonieux*. L'histogramme des PPM^3 pour $m = 3$ est illustré par la figure IV.18-(g) et montre des PPMs plus fortes dans le cas *parcimonieux* (en rouge) par rapport au *complet* (en vert).

Comme dans le cadre de l'inférence MCMC, nous avons regardé l'emplacement des voxels obtenus afin de déterminer les VP, VN, FP et FN. Les résultats sont illustrés par la figure IV.17 pour le cas $\sigma^2 = 0.5, \gamma = 2.2$ et $m = 1$ par exemple. Le modèle *complet* est représenté par la première ligne et les modèles *parcimonieux 1,2* (comportements similaires) par la deuxième. Chaque colonne concerne un niveau de seuillage des PPMs ($\alpha = 0.9, 0.95$ et 0.99). Notons que le modèle *complet* souffre, dans les trois cas, de plus de FNs que les modèles *parcimonieux*, ceci étant cohérent avec le fait d'avoir des PPMs plus élevées dans le cas *parcimonieux*. De même, pour la deuxième condition pertinente $m = 2$ représentée par la figure IV.18.

♣ Energie libre

L'énergie libre étant le critère maximisé dans VEM (section III.2.5), son calcul nous permet de vérifier la convergence de l'algorithme. La figure IV.19, illustre

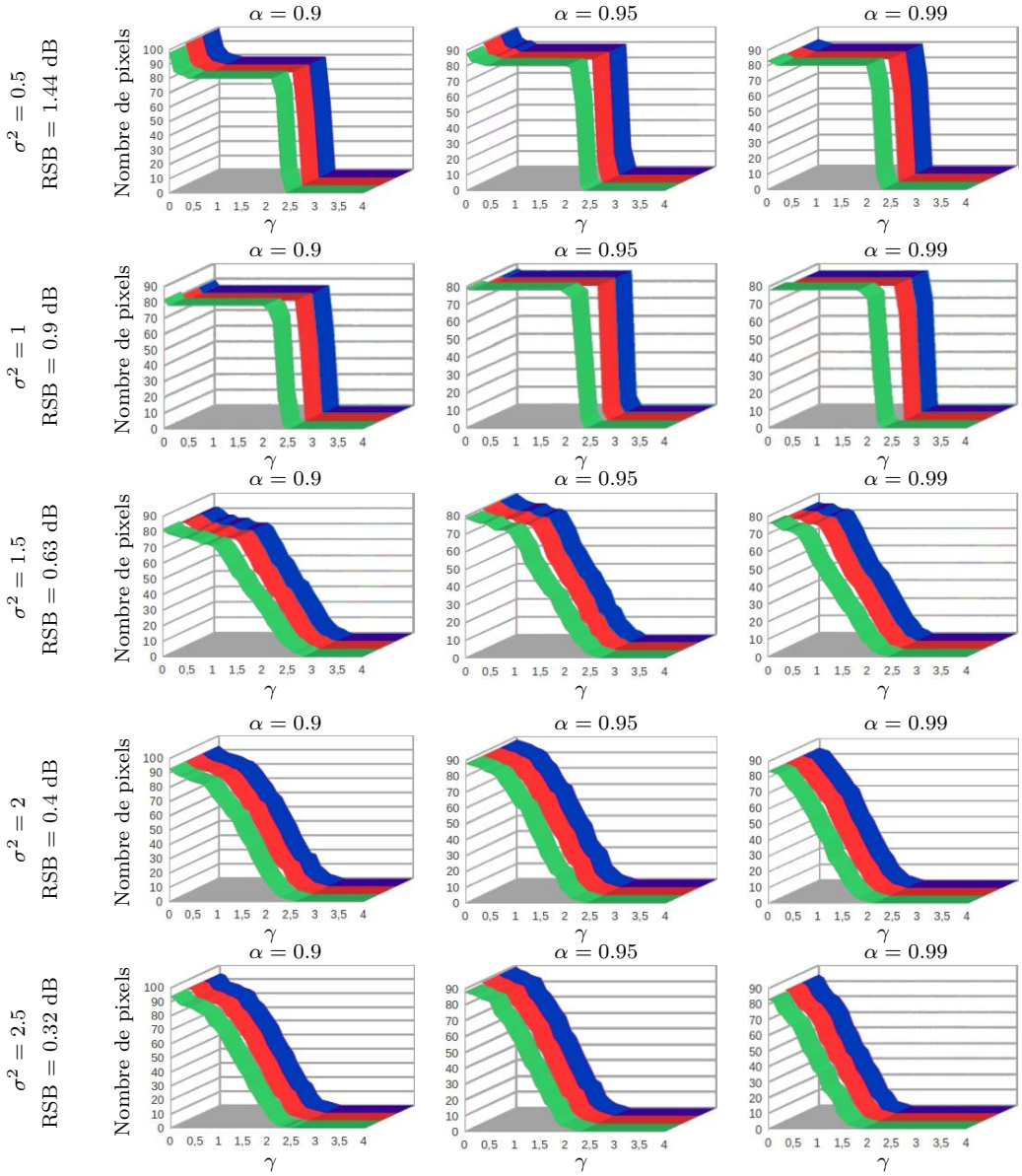


FIGURE IV.15 – Comparaison inter-modèles, pour la première condition pertinente $m = 1$, du nombre de pixels obtenu après seuillage des PPM^{γ} ($p(\hat{a}_j^m \geq \gamma \,|\, \boldsymbol{y}_j)$ pour $j = 1 : J$) avec $\gamma = 0 : 4$. Cela pour trois niveaux de seuillage $\alpha = 0.9$ (colonne de gauche), $\alpha = 0.95$ (colonne du milieu) et $\alpha = 0.99$ (colonne de droite). Notons que pour un seuil donné γ le nombre d'activations obtenues dans le cas *parcimonieux* (*1* en rouge et *2* en bleu) est équivalent ou plus élevé que celui obtenu dans le cas *complet* (en vert).

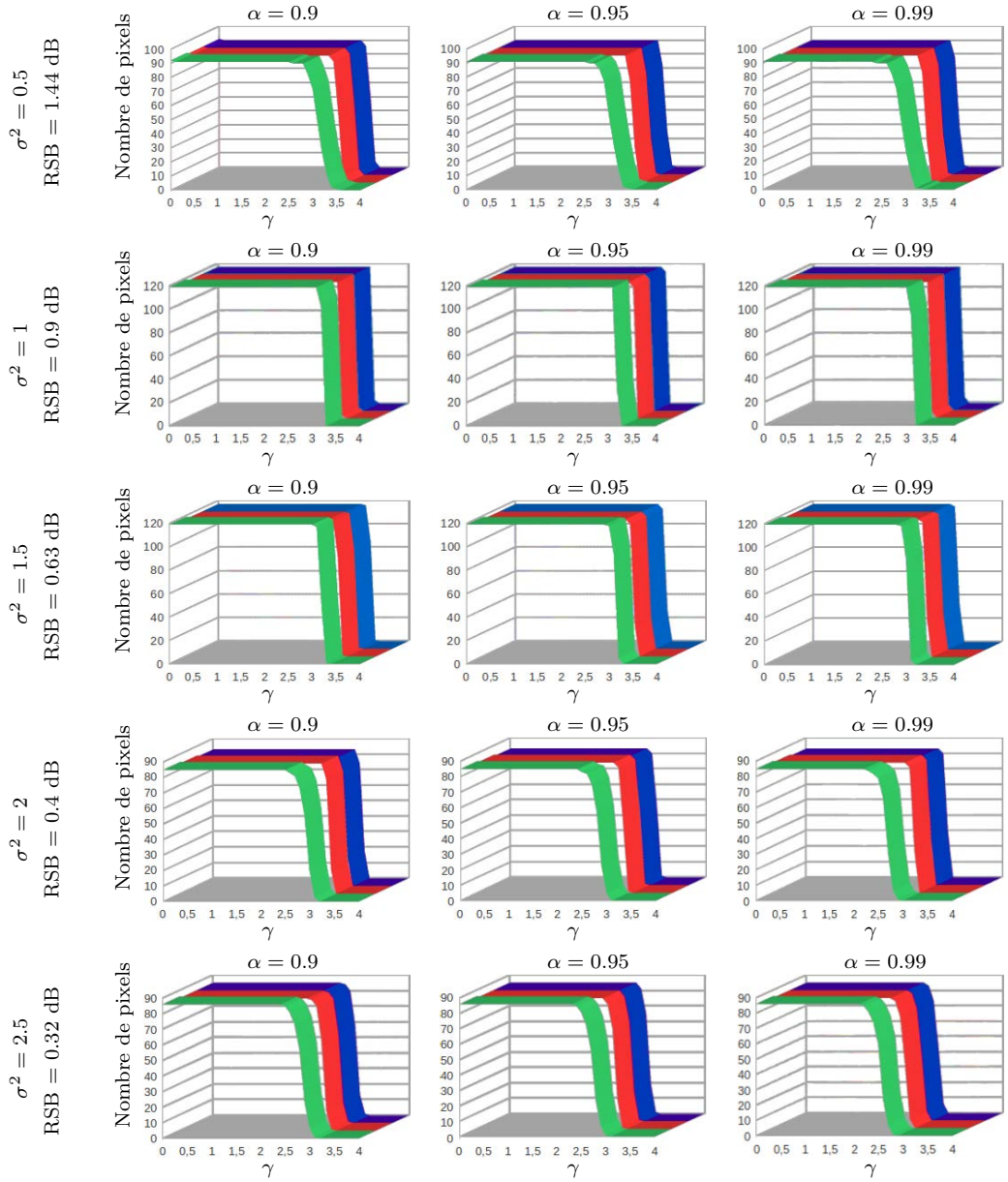


FIGURE IV.16 – Comparaison inter-modèles, pour la deuxième condition pertinente $m = 2$, du nombre de pixels obtenu après seuillage des PPM^{γ} ($p(\hat{a}_j^m \geq \gamma \,|\, \boldsymbol{y}_j)$ pour $j = 1 : J$) avec $\gamma = 0 : 4$. Cela pour trois niveaux de seuillage $\alpha = 0.9$ (colonne de gauche), $\alpha = 0.95$ (colonne du milieu) et $\alpha = 0.99$ (colonne de droite). Notons que pour un seuil donné γ le nombre d'activations obtenues dans le cas *parcimonieux* (*1* en rouge et *2* en bleu) est équivalent ou plus élevé que celui obtenu dans le cas *complet* (en vert).

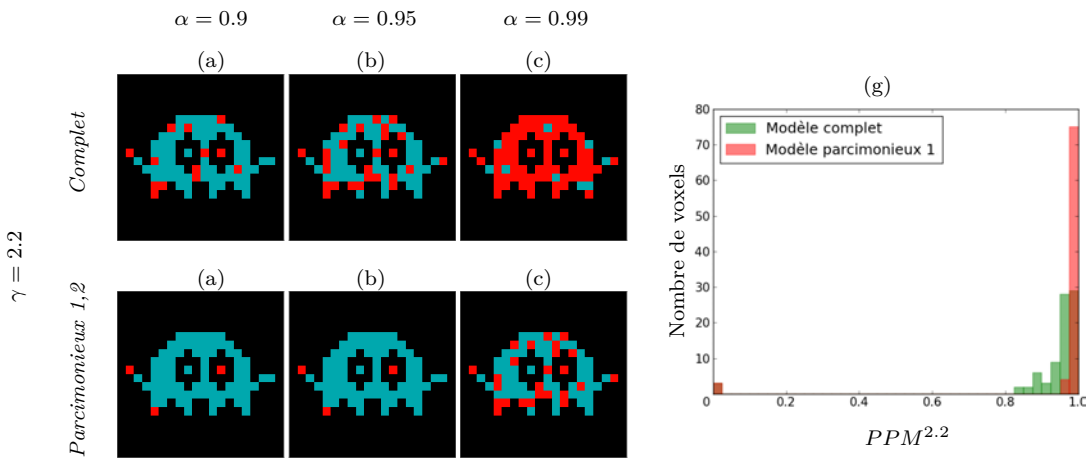


FIGURE IV.17 – Illustration, des VP (en bleu), VN (en noir), FP (en vert) et FN (en rouge) dans les cartes $PPM^{\gamma} \geq \alpha$ pour $\gamma = 2.2$ et $\alpha = 0.9$ (colonne 1), $\alpha = 0.95$ (colonne 2) et $\alpha = 0.99$ (colonne 3). Les modèles *parcimonieux 1*,2 ont des comportement équivalents et meilleurs par rapport au modèle *complet*, ce qui indique des PPMs moins élevées pour ce dernier. Cela est également confirmé par l'histogramme des $PPM^{2.2}$ (g). Ceci dans le cas de la première condition pertinente $m = 1$ et d'un niveau de bruit $\sigma^2 = 0.5$.

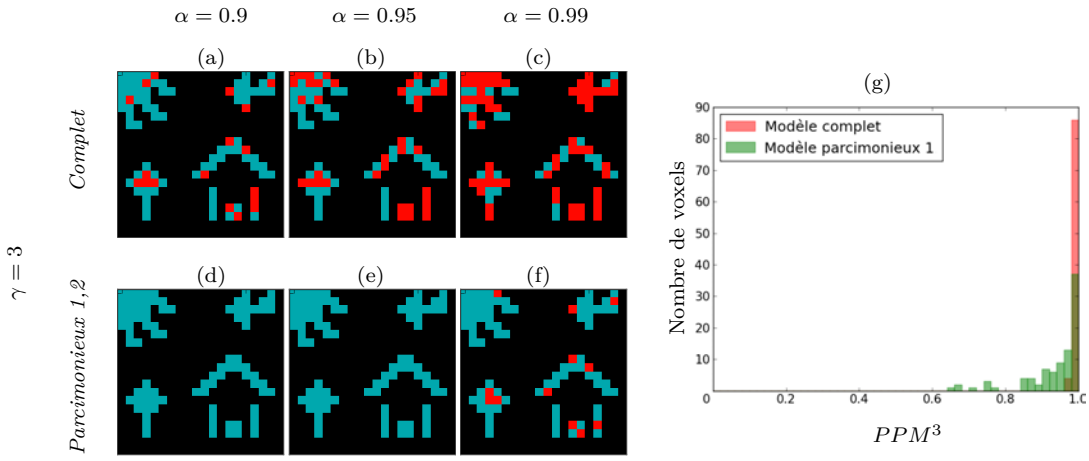


FIGURE IV.18 – Illustration, des VP (en bleu), VN (en noir), FP (en vert) et FN (en rouge) dans les cartes $PPM^{\gamma} \geq \alpha$ pour $\gamma = 3$ et $\alpha = 0.9$ (colonne 1), $\alpha = 0.95$ (colonne 2) et $\alpha = 0.99$ (colonne 3). Les modèles *parcimonieux 1*,2 ont des comportement équivalents et meilleurs par rapport au modèle *complet*, ce qui indique des PPMs moins élevées pour ce dernier. Cela également est confirmé par l'histogramme des PPM^{3} (g). Ceci dans le cas de la deuxième condition pertinente $m = 2$ et d'un niveau de bruit $\sigma^2 = 0.5$.

les énergies libres correspondantes aux 5 niveaux de bruit étudiés et montre la convergence similaire des trois modèles, *parcimonieux* (*1* en rouge et *2* en bleu) et *complet* (en vert).

D'autres types de simulations ont également été effectuées mais ne donnent pas de résultats différents dans le sens où les modèles *parcimonieux* ont toujours un comportement équivalent ou meilleur que le modèle *complet* avec des PPMs similaires ou plus fortes. Parmi ces simulations, on a testé le cas d'un bruit réel et deux cas avec un grand nombre de conditions non pertinentes (4 non pertinentes vs 1 pertinente et 7 non pertinentes vs 1 pertinente).

IV.1.3 Conclusion

Le problème principal observé dans le modèle *complet* en présence de conditions non pertinentes est l'estimation aléatoire de leurs labels et donc la détection d'un grand nombre de FP dans ces derniers. Ce problème est résolu par le modèle *parcimonieux* qui détecte la non pertinence et supprime donc ces FPs. Ce problème étant induit par la superposition des deux gaussiennes (l'annulation de la moyenne de la classe active), les NRNs non pertinents estimés sont faibles et ne gènent pas l'estimation de la FRH. C'est la raison pour laquelle aucune modification n'est observée sur cette dernière après élimination des conditions non pertinentes. Cependant et comme le montre l'étude effectuée ci-dessus, la suppression de ces conditions permet d'améliorer les PPMs, et donc les NRNs, correspondants aux conditions pertinentes et de diminuer le nombre de FN. Notons aussi que, pour le modèle *parcimonieux 1*, l'inférence variationnelle permet, par rapport à l'inférence par MCMC, la baisse du τ_2 nécessaire à la détection de la non pertinence.

IV.2 Données réelles

Nous considérons ci-dessous deux exemples de sujets appartenant à deux jeux de données réelles différents non lissés et issus d'un paradigme événementiel rapide. Le premier étant une jeune fille de 25 ans et le second une dame de 65 ans. Nous allons nous concentrer sur des régions d'intérêt (ROIs) motrices étant donné nos connaissances *a priori* sur les conditions pertinentes et non pertinentes et donc sur les résultats attendus dans ces ROIs. Les figures sont en convention neurologique, le côté gauche sur les figures correspond au côté gauche du cerveau.

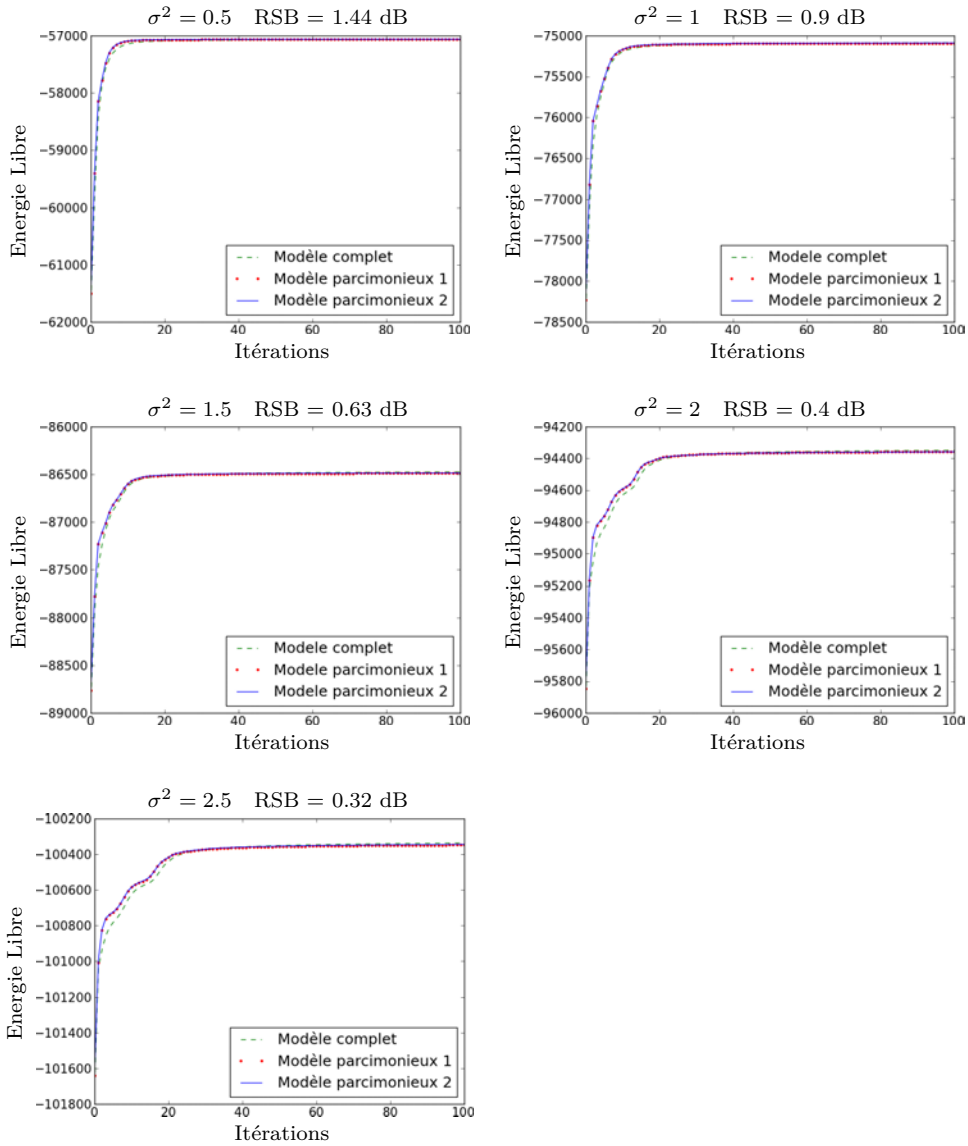


FIGURE IV.19 – Energie libre des modèles *parcimonieux 1* (en rouge), *parcimonieux 2* (en bleu) et *complet* (en vert) pour 5 niveaux de bruit différents ($\sigma^2 \in \{0.5, 1, 1.5, 2, 2.5\}$). Une petite différence en faveur des modèles *parcimonieux* est observée sur les premières itérations.

IV.2.1 Premier sujet

Ce sujet est issu de la base de données OPTIMED acquise à Neurospin à l'aide d'un scanner IRM 3 Tesla (Tim Trio, Siemens) sur 15 volontaires sains d'âge moyen de 23.2 ans (avec un écart type de 2 ans). Le volume anatomique, avec des voxels de taille $1\times1\times1$ mm^3, a été acquis avec un temps de répétition (TR) de 2400 ms et un temps d'écho (TE) de 2.98 ms. Les données fonctionnelles, quant à elles, ont été acquises avec un TR de 2400 ms et un TE de 30 ms. Elles comprennent une seule session de $N=128$ scans, chacun représentant un volume 3D de dimension $96\times96\times40$ voxels avec des voxels de taille $2\times2\times3$ mm^3. Ces données ont été collectées avec une antenne 32 canaux permettant d'effectuer une imagerie parallèle grâce à laquelle on peut réaliser un compromis entre un TR raisonable et une bonne résolution spatiale. Deux facteurs d'accélération ont été testés $R = 2$ et $R = 4$ où plus R est élevé plus le rapport signal sur bruit (RSB) est faible [Pruessmann 1999]. Ici, nous avons traité un cas avec $R = 2$.

Le paradigme expérimental est un paradigme événementiel rapide connu sous le nom de *Localizer* [Pinel 2007]. Ici, ce protocole est d'une durée de 5 minutes et comprend 100 stimuli avec un intervalle inter-stimuli (IIS) de 3.76 secondes de moyenne et de 1.99 secondes d'écart-type. Ce protocole comprend, en plus des stimuli visuels de damiers horizontaux et verticaux, des activités motrices ainsi que des tâches de calcul mental et de compréhension de phrase toutes données sous deux modalités, auditive et visuelle. On distingue donc 10 types de stimuli : (1,2) damiers horizontaux (10 occurences) et verticaux (10 occurences), (3,4) trois clics droits (resp. gauches) donnés par ordre visuel (5 occurences), (5,6) trois clics droits (resp. gauches) donnés par ordre auditif (5 occurences), (7,8) lire et écouter des phrases (10 occurences chacuns) et (9,10) calcul mental donné sous formes visuelles et auditives (10 occurences chacunes). Enfin, 20 périodes de repos (écran noir) sont insérées dans la séquence pour améliorer l'estimation de la réponse hémodynamique.

Afin d'obtenir un fort contraste, les stimuli ont été regroupés en quatres catégories : calcul, phrase, clic et damier. On s'intéresse à la région motrice droite (162 voxels) dans laquelle il est attendu que les conditions damier, phrase et calcul soient considérées comme non pertinentes. Cette région a été extraite à partir d'une analyse de groupe effectuée sur les 15 sujets.

Le bruit étant plus fort et de nature différente par rapport aux données simulées, la valeur de τ_2 nécessaire à l'élimination des conditions non pertinentes

est plus élevée comme on le verra par la suite.

IV.2.1.1 Technique stochastique - MCMC

En appliquant la même procédure que celle utilisée sur les données simulées, la valeur de τ_2 nécessaire pour le modèle *parcimonieux 1* est de 82, ce qui est équivalent à 50% de la taille de la ROI motrice en question (162 voxels). Pour cette valeur, le modèle *parcimonieux 1* détecte les trois conditions : calcul, phrase et damier comme non pertinentes avec des probabilités *a posteriori* $p(w^m = 1 \,|\, \boldsymbol{y}_j)$ égales à 0 et la condition clic comme pertinente avec une probabilité *a posteriori* $p(w^m = 1 \,|\, \boldsymbol{y}_j) = 1$.

La *procédure PPMs* a été effectuée avec $\gamma = 0 : 8$ induisant la figure IV.20-(a) où le modèle *parcimonieux 1* (en rouge) détecte considérablement plus de voxels dont le NRN est supérieur à γ que le modèle *complet* (en vert). Si on se focalise par exemple sur le cas $\gamma = 3$, on s'aperçoit que le modèle *parcimonieux 1* détecte 84 voxels par rapport à 24 détectés par le modèle *complet*. L'histogramme des PPM^3 est illustré par la figure IV.20-(b) indiquant effectivement des PPMs plus fortes dans le cas *parcimonieux*. Cette amélioration est également confirmée par les figures IV.20-(d,e) représentant les PPM^3 pour les modèles *complet* et *parcimonieux* respectivement. Pour plus de visibilité, un zoom a été effectué sur la partie entourée en rouge dans la figure IV.20-(c) et qui contient le sillon central impliqué dans les processus moteurs. Contrairement aux précédentes observations concernant l'estimation de la FRH, ici une différence significative est observée entre les deux modèles comme l'indique la figure IV.21 où l'estimation du modèle *parcimonieux 1* est en rouge et celle du *complet* en vert. Notons une différence d'environ 1 seconde dans le temps d'arrivée du pic. Il est clair que les deux diffèrent de la forme canonique (en noir). Rappelons que dans les précédents cas, les NRNs des conditions non pertinentes étaient faibles à cause de la superposition des gaussiennes représentant les classes d'activation. Ici, les NRNs non pertinents (calcul, phrase et damier) du modèle *complet* ont de fortes valeurs négatives comme on peut le voir à travers leurs histogrammes illustrés par les figures IV.22-(a,b,d) respectivement. Ces valeurs non nulles influencent clairement l'estimation de la FRH et diminuent également les NRNs de la condition pertinente (clic) comme apparaît dans l'histogramme IV.22-(c).

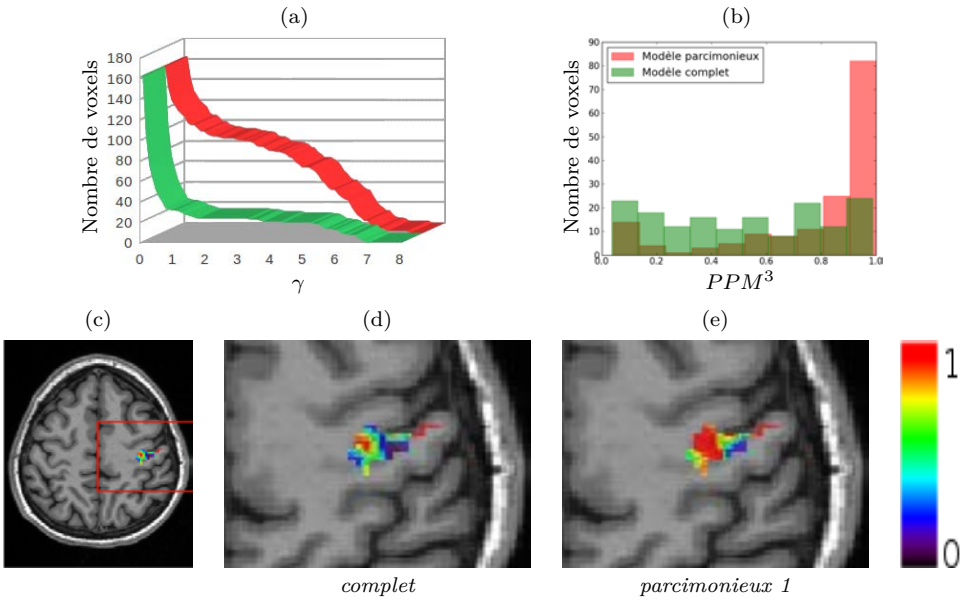


FIGURE IV.20 – (a) Nombre de voxels obtenu après seuillage des PPM^{γ} à 0.95 avec $\gamma = 0 : 8$ où le modèle *parcimonieux 1* (en rouge) détecte plus de voxels que le *complet* (en vert). (b) Histogramme des PPM^3 confirmant des PPMs *parcimonieuses* (en rouge) plus fortes que celle du *complet* (en vert). (c) Illustration de la région contenant le sillon central et dont le zoom permet d'obtenir les figures (d) et (e) représentant la superposition des PPM^3 au volume anatomique, ceci pour les modèles *complet* et *parcimonieux 1* respectivement.

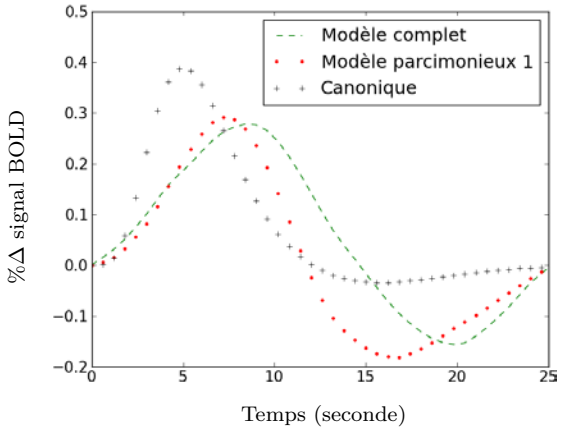


FIGURE IV.21 – Comparaison des FRHs estimées par chacun des modèles *parcimonieux 1* (en rouge) et *complet en vert*. La FRH canonique est illustrée en noir afin de monter la différence avec celles estimées.

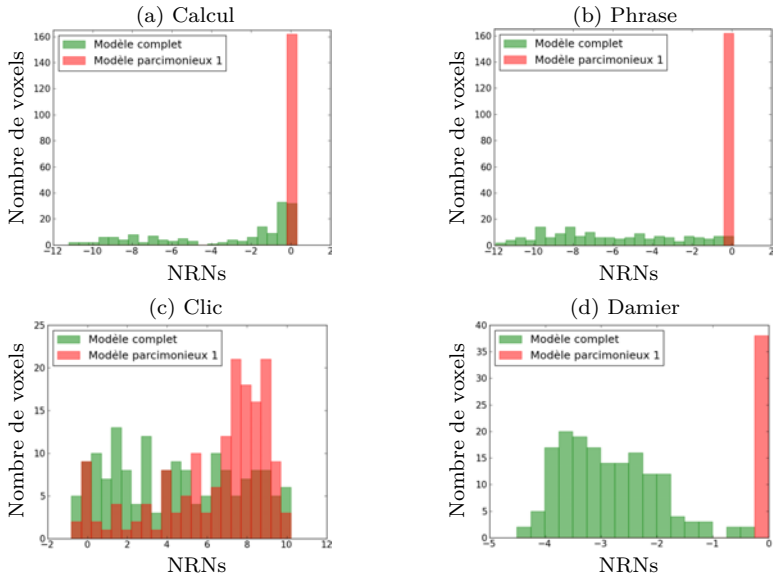


FIGURE IV.22 – Histogrammes des NRNs estimés par les modèles *complet* (en vert) et *parcimonieux 1* (en rouge) pour les trois conditions non pertinentes : calcul (a), phrase (b) et damier (d) et la condition pertinente clic (c). Les NRNs des conditions non pertinentes du modèle *complet* ont de fortes valeurs négatives non nulles causant la diminution des NRNs de la condition pertinente. L'élimination de ces trois conditions par le modèle *parcimonieux 1* améliore considérablement les NRNs de la condition pertinente.

IV.2.1.2 Technique variationnelle - VEM

Le nombre d'itérations minimum est de 100 avec les mêmes critères d'arrêt que ceux utilisés pour le traitement des données simulées. Dans le cas du modèle *parcimonieux 1*, un $\tau_2 = 0$ n'est plus suffisant pour l'élimination des conditions non pertinentes. La procédure du choix de τ_2 utilisée pour les autres cas, nous donne un $\tau_2 = 25$ (15% de la taille de notre ROI), ce qui est beaucoup moins élevé que la valeur obtenue en MCMC ($\tau_2 = 82$). Pour cette valeur de τ_2, le modèle *parcimonieux 1* détecte les conditions : damier, phrase et calcul comme non pertinentes avec des probabilités *a posteriori* $\widetilde{p}_{W^m}(w^m = 1) = 0$ et la condition clic comme pertinente avec une probabilité *a posteriori* de 1.

Concernant le modèle *parcimonieux 2*, 4 cas ont été testés et comparés. Pour chaque cas, nous illustrons le comportement de l'estimateur de $\widetilde{p}_{W^m}(w^m = 1)$ pour $m = 1 : 4$ (figures IV.24-(a,d,g,j)), l'estimateur de τ_2 (figures IV.24-(b,e,h,k)) ainsi que les résultats de la *procédure PPMs* à $\alpha = 0.95$ (figures IV.24-(c,f,i,l)). Rappelons que le modèle *complet* est représenté en vert, le *parcimonieux 1* en rouge et le *parcimonieux 2* en bleu.

1- Le premier cas consiste à utiliser les mêmes valeurs que le cas simulé : $\tau_1 = 1$, $\alpha_{\tau_2} = 3$ et $\lambda_{\tau_2} = 4$. Comme le montre la figure IV.24-(a) les probabilités *a posteriori* $\widetilde{p}_{W^m}(w^m = 1)$ des conditions non pertinentes (calcul, phrase et damier) sont estimées à 0.34, 0.28 et 0.34 respectivement et ne sont pas complètement nulles. Ceci donne des poids aux trois conditions et ne les élimine pas complètement du signal stimulus-induit induisant un comportement similaire des modèles *complet* et *parcimonieux 2* car les NRNs non pertinents du modèle *complet* sont déjà faibles et leur donner des poids ne change pas grand chose. Dans ce cas τ_2 est estimé à 0.52 ce qui donne pour $\hat{\mu}_1^m = 0$ une probabilité $p_0 = \frac{1}{1+\exp(\tau_1\tau_2)} = \frac{1}{1+\exp(0.52)} = 0.37$ (figure IV.23-(a), courbe en bleu). Cette probabilité représente la probabilité de considérer une condition comme pertinente alors que la moyenne d'activation μ_1^m est nulle, une valeur de 0.37 est donc élevée. Théoriquement et selon notre définition de la non pertinence, cette valeur doit être nulle, nous allons donc la fixer à 10^{-3} (figure IV.23-(a), courbe en rouge) comme on le verra dans les cas suivants. Ce problème n'étant pas présent dans le cas du modèle *parcimonieux 1* où $\tau_2 = 25$ (suffisamment loin de l'axe des ordonnées).

2- Le deuxième cas consiste en la diminution de la probabilité p_0 en augmentant la pente de la sigmoïde τ_1. Pour ceci nous proposons de fixer ce paramètre selon l'équation $\tau_1 = \frac{1}{\tau_2} \log \frac{1-p_0}{p_0}$. Etant donné que τ_2 est estimé, on peut utiliser sa valeur initiale où le mode de son *a priori* pour fixer la valeur de τ_1. Ici nous avons testé $p_0 = 10^{-3}$ et un *a priori* sur τ_2 d'hyperparamètres $\alpha_{\tau_2} = 3$ et $\lambda_{\tau_2} = 4$. La figure IV.24-(d) montre l'intérêt de la diminution de p_0 qui permet l'élimination complète des conditions non pertinentes ($\widetilde{p}_{W^m}(w^m = 1) = 0$) mais tardivement (après 70 itérations) empêchant le modèle *parcimonieux 2* de s'améliorer par rapport au modèle *complet* comme le montre la *procédure PPMs* illustrée par la figure IV.24-(f).

3- Le troisième cas vise à accélérer la détection de la non pertinence par le modèle *parcimonieux 2* afin de lui permettre de s'améliorer par rapport au modèle *complet*. Comme le montre la figure IV.24-(e), $\hat{\tau}_2$ tend vers 0.5 (le mode de son *a priori*), une proposition serait donc d'accélérer cette convergence en utilisant un *a priori* plus strict $\alpha_{\tau_2} = 9$ et $\lambda_{\tau_2} = 16$, où le mode est de 0.5 mais la variance de 0.03 (figure IV.23-(b), courbe en rouge), presque 10 fois plus petite que l'*a priori* précédent (figure IV.23-(b), courbe en bleu). Ceci accélère clairement la convergence (figure IV.24-(g,h)) permettant au modèle

parcimonieux 2 un comportement très proche de celui du modèle *parcimonieux 1* et meilleur que celui du modèle *complet* qui détecte moins de voxels que les modèles *parcimonieux* quand γ augmente.

4- Le quatrième et dernier cas permet de répondre à la question : *étant donné que* $\hat{\tau}_2$ *tend vers* 0.5*, pourquoi ne pas le fixer à cette valeur dès le début au lieu de l'estimer ?* Selon la figure IV.24-(j) ceci permetrait une élimination beaucoup plus rapide des conditions non pertinentes mais cause néanmoins une perte de quelques activations (figure IV.24-(l)). Cela peut être justifié par le fait que μ_1^m, auquel est connecté w^m, est de l'ordre de 0.5 ou inférieur à cause du problème de bilinéarité entre $\boldsymbol{A}$ et $\boldsymbol{h}$ qui se traduit par de grandes valeurs de $\boldsymbol{h}$ et de petites valeurs de $\boldsymbol{A}$ durant les premières itérations. Estimer τ_2 permet de s'adapter à ce problème car comme le montre les figures IV.24-(b,e,h) ce paramètre prend des petites valeurs durant les premières itérations.

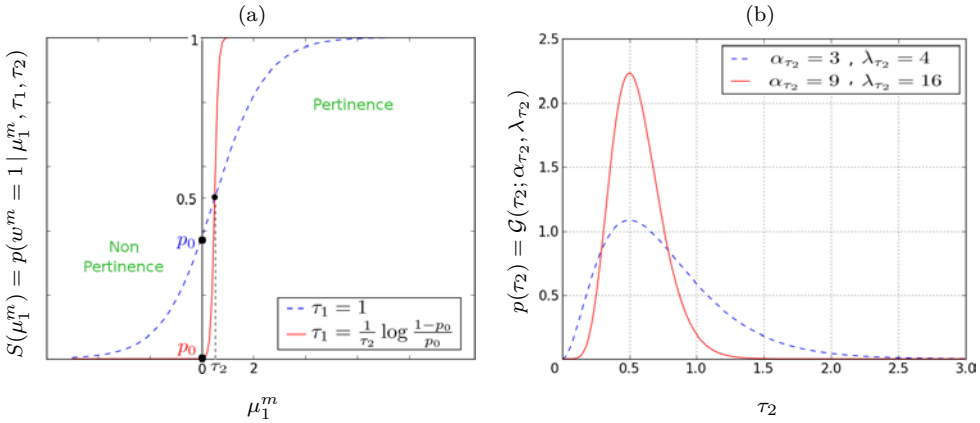


FIGURE IV.23 – (a) Illustration de l'influence de la pente de la sigmoïde, contrôlée par τ_1, sur la probabilité à l'origine $p_0 = p(w^m = 1 \,|\, \mu_1^m = 0, \tau_1, \tau_2)$. Pour un τ_2 donné, plus τ_1 est grand, plus la pente est forte et p_0 faible (courbe rouge). (b) Comparaison de deux fonctions gamma dont lo mode est équivalent 0.5, $\mathcal{G}(3, 4)$ (en bleu) a une variance plus élevée que $\mathcal{G}(9, 16)$ (en rouge), permettant ainsi plus de fluctuation dans l'estimation de τ_2 et donc un ralentissement de convergence vers sa valeur finale.

La précédente analyse nous permet de conclure, que dans le cas des données réelles il est préférable d'augmenter la pente de la sigmoïde pour le modèle *parcimonieux 2* et d'utiliser un *a priori* gamma plus strict afin d'accélerer la détection des conditions non pertinentes. La figure IV.25 illustre les PPM^5 pour la condition pertinente (clic) dans la ROI motrice droite. Pour une meilleure visibilité, un zoom a été effectué sur la partie entourée en rouge dans la figure IV.25-(a) où

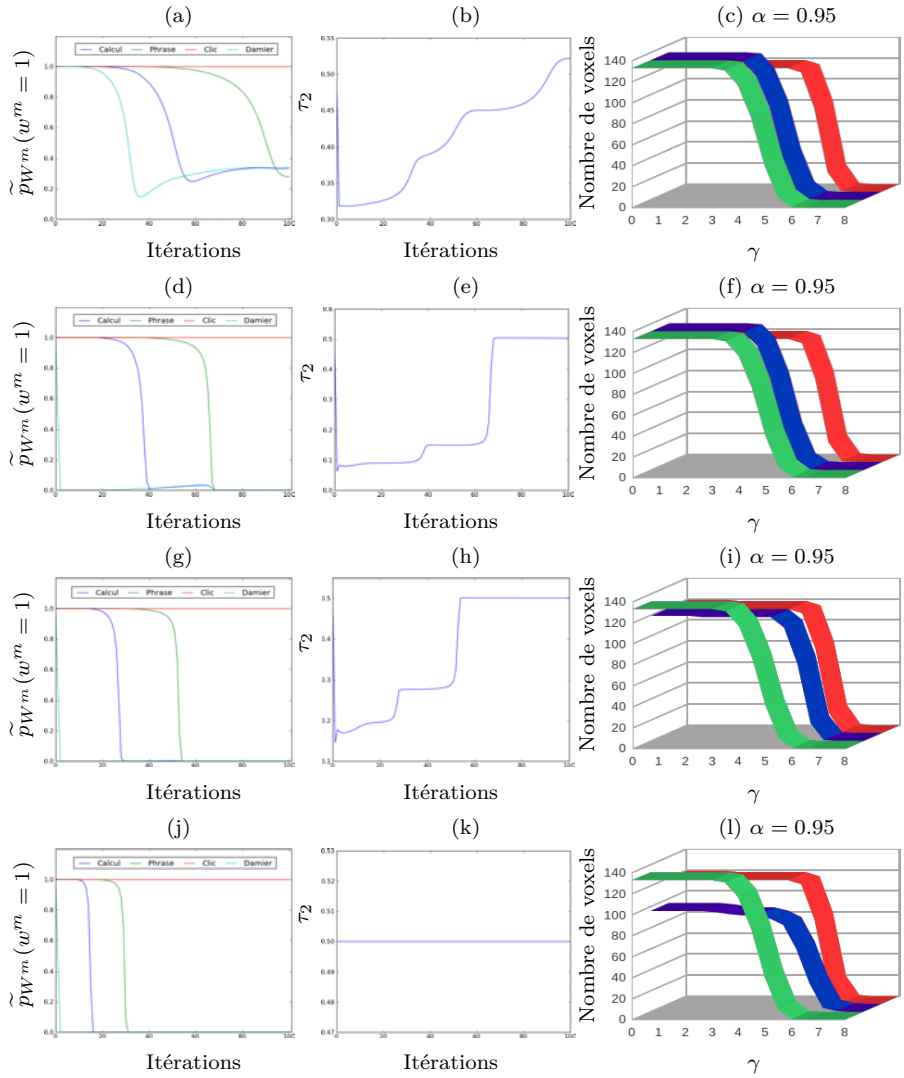


FIGURE IV.24 – Colonne 1 : comportement des estimateurs de $\widetilde{p}_{W^m}(w^m = 1)$ pour le modèle *parcimonieux 2* durant les itérations pour les quatre conditions : calcul en bleu, phrase en vert, clic en rouge et damier en bleu clair. Colonne 2 : comportement de l'estimateur de τ_2 pour le modèle *parcimonieux 2* durant les itérations. Colonne 3 : nombre de voxels obtenu après seuillage des PPM^γ à $\alpha = 0.95$ avec $\gamma = 0 : 8$, en vert le modèle *complet*, en rouge le modèle *parcimonieux 1* et en bleu le 2. Notons que le modèle *parcimonieux 1* permet d'obtenir des PPMs plus forts que le modèle *complet* (un nombre plus élevé d'activations est obtenu pour $\gamma > 4$). Les lignes correspondent aux cas suivants pour le modèle *parcimonieux 2* : (1) $p_0 = 0.37$ et $p(\tau_2) = \mathcal{G}(3, 4)$ où $\widetilde{p}_{W^m}(w^m = 1) \approx 0.3$, (2) $p_0 = 10^{-3}$ et $p(\tau_2) = \mathcal{G}(3, 4)$ avec une convergence tardive de $\widetilde{p}_{W^m}(w^m = 1)$ vers 0. Ces deux cas induisent une équivalence avec le modèle *complet*. (3) $p_0 = 10^{-3}$ et $p(\tau_2) = \mathcal{G}(9, 16)$ accélére la convergence des probabilités vers 0 par rapport à (2) induisant une amélioration du modèle *parcimonieux 2* qui devient proche du *parcimonieux 1*. (4) $p_0 = 10^{-3}$ et $\tau_2 = 0.5$, ce cas cause la baisse de performance du modèle *parcimonieux 2* pour $\gamma < 5.5$ à cause de la bilinéarité entre $\boldsymbol{A}$ et $\boldsymbol{h}$ induisant un μ_1^m de l'ordre de 0.5 sur les premières itérations.

se trouve le sillon central droit. Les figures IV.25-(d,e,f) représentent les PPM^5 sur une coupe axiale pour les modèles *complet*, *parcimonieux 1* et *parcimonieux 2* respectivement et montrent des PPMs plus fortes pour les modèles *parcimonieux*. Les histogrammes (figures IV.25-(b,c)) permettent d'avoir une idée sur les PPM^5 estimées dans la ROI en entier et non seulement sur une seule coupe. Ces histogrammes permettent de conclure, comme précédemment, que les PPM^5 des modèles *parcimonieux 1* (histogramme en rouge) et *2* (histogramme en bleu) sont supérieures à ceux du modèle *complet* (histogramme en vert). En seuillant ces PPMs à $\alpha = 0.95$, on obtient 41 voxels pour le modèle *complet* par rapport à 114 pour le *parcimonieux 1* et 108 pour le *2* (où τ_2 est estimé contrairement au modèle *1*).

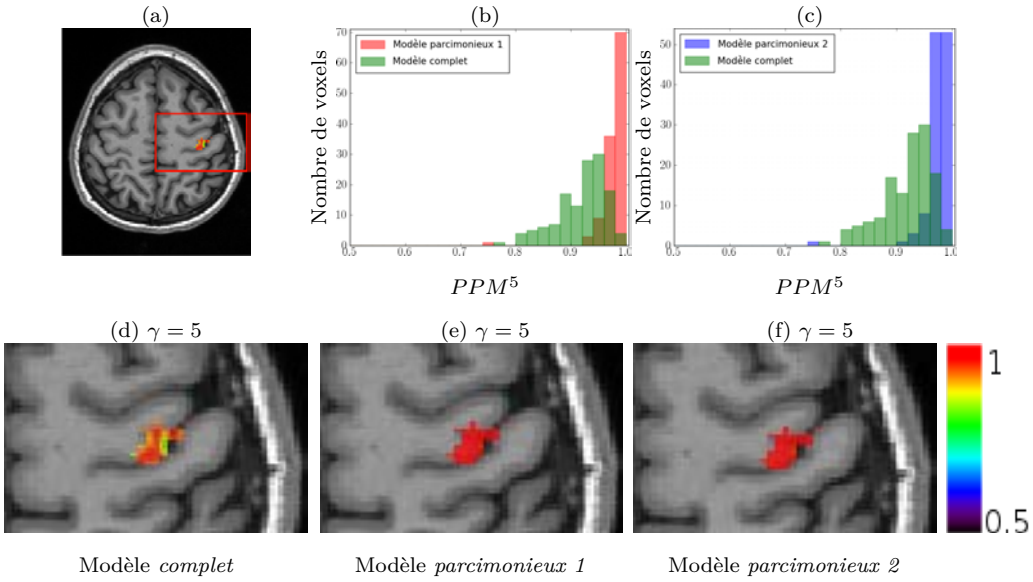


FIGURE IV.25 – (a) Illustration, en rouge, de la zone contenant le sillon central et sur laquelle un zoom a été effectué pour plus de visibilité. (b,c) Comparaison des histogrammes correspondant aux PPM^5 pour les modèles *complet* (en vert),*parcimonieux 1* (en rouge) et *2* (en bleu). Ces derniers fournissent des PPMs supérieures à celles du modèle *complet*, ce qui est également confirmé par les figures (d,e,f) où ces PPMs sont superposées sur le volume anatomique.

Finalement la figure IV.26 compare les estimations des FRHs entre le modèle *complet* (en vert) et chacun des deux modèles *parcimonieux 1* (en rouge) et *parcimonieux 2* (en bleu). Les trois estimations sont assez proches et différentes de la FRH canonique (en noir). Notons que l'estimation de la FRH est en cohérence

avec l'analyse de groupe effectuée dans [Badillo 2013b] où la FRH motrice correspond à un temps d'arrivée du pic supérieur à celui de la forme canonique.

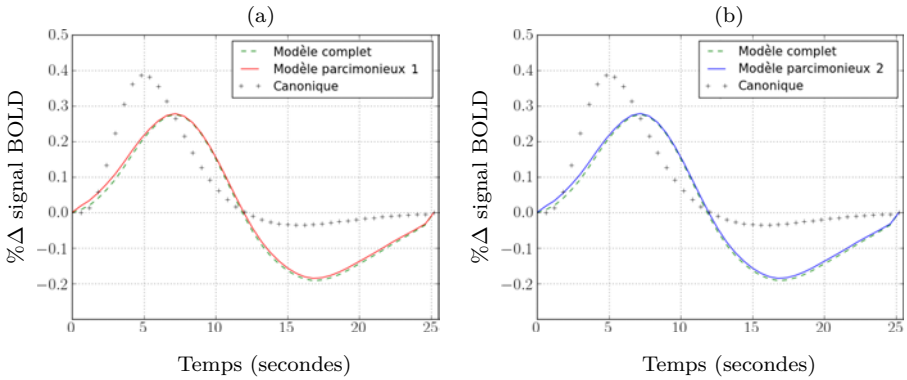


FIGURE IV.26 – Comparaison des FRHs estimées par les modèles *complet* (en vert), *parcimonieux 1* (en rouge) et *parcimonieux 2* (en bleu) dans la ROI motrice droite. La FRH canonique est affichée en noir pour montrer la différence avec les FRHs estimées. Les estimations des trois modèles sont assez proches.

IV.2.1.3 Conclusion

Les résultats de l'analyse précédente dans le cadre d'une inférence MCMC illustrent de nouvelles observations par rapport à celles des données simulées. Ici, l'amélioration obtenue au niveaux des PPMs en faveur du modèle *parcimonieux* est beaucoup plus importante que celle observée dans les simulations. Cette dernière étant due à des NRNs non pertinents non nuls et à de fortes valeurs négatives dans le cas du modèle *complet*, ce qui influence l'estimation des NRNs pertinents et effectivement celle de la FRH. En éliminant les conditions non pertinentes, le modèle *parcimonieux 1* estime des NRNs non pertinents faibles et améliore ainsi les NRNs pertinents. Son estimation de la FRH est également différente de celle du modèle *complet*. Notons que la FRH estimée par le modèle *parcimonieux* est plus cohérente avec les résultats de l'analyse de groupe effectuée dans [Badillo 2013b] où le temps d'arrivée du pic dans la ROI motrice est estimé aux alentours de 6.5 secondes. De plus la FRH estimée par le modèle *parcimonieux* en MCMC est plus proche de celle obtenue dans l'analyse variationnelle.

A l'inverse, dans le cadre de l'inférence variationnelle, le modèle *complet* estime

des NRNs non pertinents faibles dus à la superposition des gaussiennes d'activation. L'amélioration des PPMs de la condition pertinente est donc moins forte que celle observée en MCMC mais reste quand même présente illustrant l'intérêt du modèle *parcimonieux*. De plus, la valeur du paramètre τ_2 nécessaire à l'élimination des conditions non pertinentes par le modèle *parcimonieux 1* est moins élevée dans l'analyse VEM ($\tau_2 = 25$) par rapport à l'analyse MCMC ($\tau_2 = 82$) ce qui avantage, dans ce cas, l'analyse variationnelle par rapport à la méthode stochastique.

IV.2.2 Deuxième sujet

Ce sujet est issu d'une autre base de données acquises à NeuroSpin à l'aide d'un scanner IRM 3 Tesla (Tim Trio Siemens) et à travers une nouvelle technique d'acquisition de sequence EPI (*Multi-slice Echo-Planar Imaging*). Cette technique, développée par le CMRR (*Center for Magnetic Resonance Research*, université de Minnesota, Minneapolis) [Feinberg 2010], permet l'acquisition simultanée de plusieurs coupes induisant une baisse du TR à 1 seconde. Les données fonctionnelles sont acquises sur une seule session de $N = 422$ scans fonctionnels avec un TE de 30 ms. Chaque scan représente un volume 3D de dimension $64 \times 64 \times 48$ voxels avec des voxels de taille $3 \times 3 \times 3$ mm^3. L'ISI moyen est de 5.12 secondes avec un écart-type de 3.48 secondes.

Le paradigme expérimental est un paradigme événementiel rapide, d'une durée de 7 minutes et 9 secondes, comprenant 80 stimuli répartis sous 8 types différents où, par rapport au précédent paradigme, les damiers horizontaux et verticaux ont été supprimés : (1,2) trois clics droits (resp. gauches) donnés par ordre visuel (10 occurences), (3,4) trois clics droits (resp. gauches) donnés par ordre auditif (10 occurences), (5,6) lire et écouter des phrases (10 occurences chacuns) et (7,8) calcul mental donné sous formes visuelle et auditive (10 occurences chacunes). De plus 20 périodes de repos permettent d'améliorer l'estimation de la queue de la FRH.

Afin d'obtenir un fort contraste, les stimuli ont été regroupés en quatres catégories : calcul, phrase, clic droit et clic gauche. Ici, nous nous sommes intéressés à deux ROIs motrices droite (128 voxels) et gauche (130 voxels) extraites respectivement à partir des contrastes T (clic gauche - clic droit) et (clic droit - clic gauche) issus du MLG classique, ceci à travers la toolbox Marsbar de SPM.

♣ **Extraction de ROIs avec Marsbar :** Cette toolbox nous permet de créer une ROI basée sur le cluster d'activation individuelle détecté par le logiciel SPM à partir du contraste (clic gauche - clic droit) si on s'intéresse aux voxels activés majoritairement par le clic gauche et qui se trouvent dans l'hémisphère contralatéral droit. Une fois le contraste défini on peut choisir d'effectuer ou non une correction FWER (*Family Wise Error rate*) pour affiner notre ROI. Ici nos ROIs correspondent à une FWER seuillée à 0.05.

Nous remarquons un comportement différent par rapport au sujet précédent, ceci pour deux points principaux qu'on confirmera à travers des figures par la suite :

1- La classification aléatoire des activations correspondant aux conditions non pertinentes influence négativement le choix de τ_2 pour le modèle *parcimonieux 1* et qui atteint 60% de la taille de la ROI en question (valeur obtenue avec la même procédure que celle effectuée pour le sujet précédent). Le point positif étant que les données empêchent cette valeur élevée de τ_2 d'influencer la condition pertinente.

2- L'élimination des conditions non pertinentes n'induit aucune amélioration dans l'estimation des variables de la condition pertinente, à l'inverse de ce que l'on a observé dans les données simulées et le premier sujet. Les conditions éliminées par le modèle *parcimonieux* correspondent néanmoins à des activations dans le modèle *complet*. La question discutée par la suite étant : *Sommes nous dans un cadre de FN pour le modèle parcimonieux ou plutôt dans un cadre de FP pour le complet ?*

Pour le modèle *parcimonieux 2*, les mêmes valeurs que celles du permier sujet ont été utilisées pour les paramètres de la sigmoïde ($p_0 = 10^{-3}$ et $p(\tau_2) = \mathcal{G}(9, 16)$).

La *procédure PPMs* avec $\alpha = 0.95$ permet de comparer les trois modèles. Ici nous nous intéressons aussi aux PPMs obtenues pour les conditions non pertinentes et qui normalement doivent être nulles (comme c'est le cas pour le premier sujet, raison pour laquelle nous n'avons pas abordé ce point précédemment). Dans les résultats suivants, nous illustrons uniquement le modèle *complet* et le modèle *parcimonieux 1*. Le *parcimonieux 2* ayant exactement le même comportement que le *1*, il n'est pas détaillé.

IV.2.2.1 ROI motrice droite

Dans la ROI motrice droite, les conditions : calcul, phrase et clic droit sont considérées, par chacun des modèles *parcimonieux*, comme non pertinentes avec des probabilités *a posteriori* $\widetilde{p}_{W^m}(w^m = 1)$ à 0 et la condition clic gauche comme pertinente avec une probabilité *a posteriori* à 1.

La figure IV.27 illustre les histogrammes des PPM^0 pour les quatres conditions expérimentales et pour les modèles *complet* (en vert) et *parcimonieux 1* (en rouge). On remarque que les deux modèles sont équivalents par rapport à la condition pertinente (clic gauche, figure IV.27-(d)) et ont un comportement assez proche pour les deux conditions non pertinentes (calcul et phrase, figures IV.27-(a,b) respectivement). Cependant le modèle *complet* détecte des activations ipsilatérales, des voxels de $PPM^0 \geq 0.9$ pour le clic droit qui a été considéré comme non pertinent par le modèle parcimonieux. Les figures IV.28-(a,d) représentent les résultats de la *procédure PPMs* pour le clic droit et gauche respectivement et pour un niveau de seuillage $\alpha = 0.95$. Le modèle *complet* est en vert et les *parcimonieux 1* et *2* en rouge et bleu, respectivement. On remarque que les activations ipsilatérales détectées par le modèle *complet* concernant le clic droit dans la ROI motrice droite ne sont pas fortes, elles disparaissent au-delà de $\gamma = 4$ (figures IV.28-(a)) par rapport à un $\gamma = 41$ pour les activations contralatérales (figures IV.28-(d)). Les PPM^0 pour le clic droit ont été superposées sur le volume anatomique du sujet afin de voir leurs emplacements, comme le montrent les figures IV.28-(b,c) pour le modèle *complet* et *parcimonieux 1*, respectivement. La considération de ce dernier du clic droit comme non pertinent induit des PPMs autour de 0.5 ($\int_0^\infty \mathcal{N}(0,1) = 0.5$). Le comportement similaire des modèles dans le cas du clic gauche est confirmé également par les résultats de la *procédure PPMs* illustrée par la figure IV.28-(d). Prenons par exemple le cas PPM^{15} dont la superposition sur le volume anatomique nous permet d'obtenir les figures IV.28-(e,f) équivalentes pour les modèles *complet* et *parcimonieux 1*, respectivement.

IV.2.2.2 ROI motrice gauche

Dans la ROI motrice gauche, les conditions : calcul, phrase et clic gauche sont considérées, par les modèles *parcimonieux*, comme non pertinentes avec des probabilité *a posteriori* $\widetilde{p}_{W^m}(w^m = 1)$ à 0 et la condition clic droit comme pertinente avec une probabilité *a posteriori* à 1.

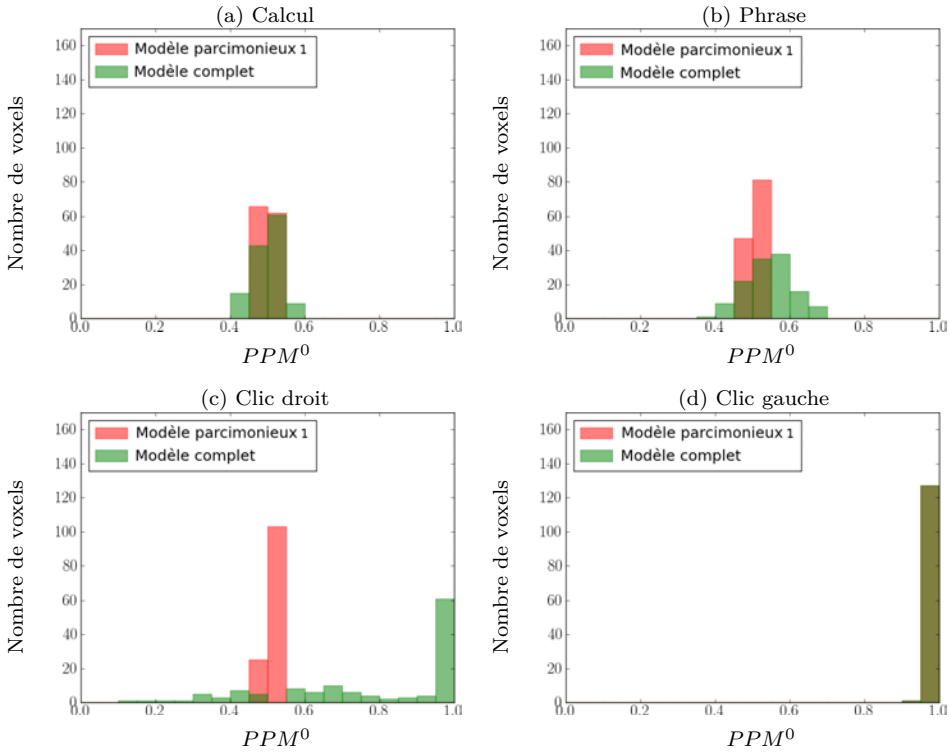


FIGURE IV.27 – Histogrammes des PPM^γ avec $\gamma = 0$ dans la ROI motrice droite pour les quatres conditions expérimentales (calcul, phrase, clic droit et clic gauche). Le modèle *complet* est représenté en vert et le modèle *parcimonieux 1* en rouge (le modèle *parcimonieux 2* ayant les mêmes résultats que le premier il n'a pas été représenté). Les modèles *parcimonieux* et *complet* ont des comportements équivalents dans le cas de la condition pertinente (clic gauche, (d)) et les deux conditions non pertinentes (calcul et phrase, (a) et (b) respectivement) à l'inverse de la condition clic droit (c) pour laquelle le modèle *complet* détecte des activations ($PPM^0 > 0.9$) alors que les modèles *parcimonieux* la considèrent comme non pertinente.

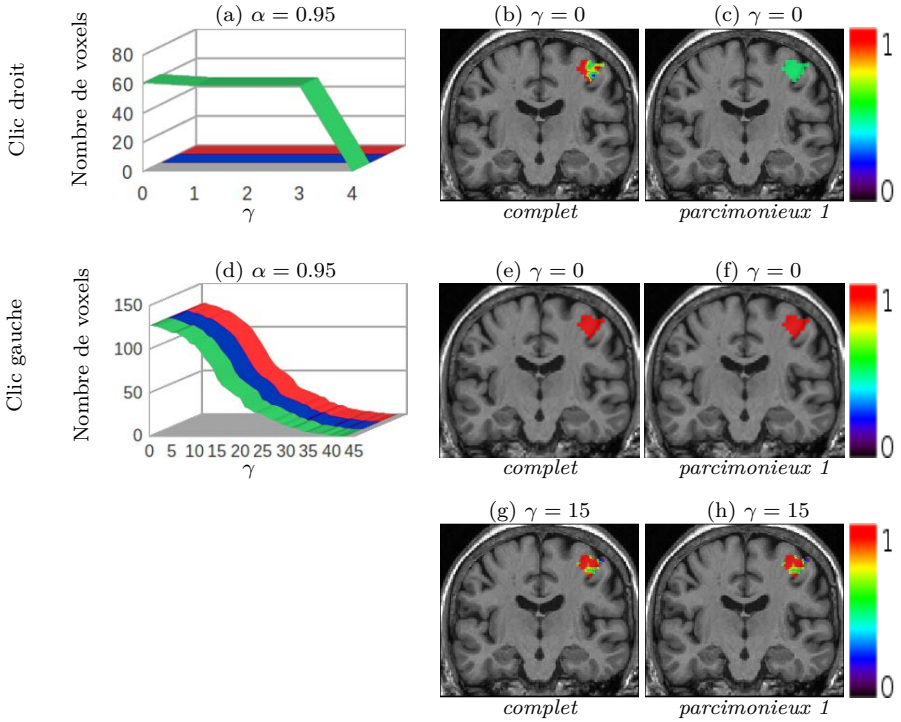


FIGURE IV.28 – (a,d) Nombre de voxels obtenu après seuillage des PPM^{γ} avec $\gamma = 0 : 4$ pour le clic droit (a) et $\gamma = 0 : 41$ pour le clic gauche (d), ceci dans la ROI motrice droite et pour un niveau de seuillage $\alpha = 0.95$. Le modèle *complet* est représenté en vert, le *parcimonieux 1* en rouge et le *parcimonieux 2* en bleu. Le modèle *complet* détecte de faibles activations ipsilatérales induites par le clic droit dans la ROI motrice droite à l'inverse des modèles *parcimonieux* qui considèrent cette condition comme non pertinente. Cette différence de comportement entre les modèles est également illustrée par (b,c) représentant la superposition avec le volume anatomique des PPM^0 pour le clic droit et pour les modèles *complet* et *parcimonieux*, respectivement. Superposition des PPM^0 (e,f) et des PPM^{15} (g,h) pour le clic gauche et pour les modèles *complet* (à gauche) et *parcimonieux 1* (à droite), ces derniers ayant des comportement similaires.

Comme pour la ROI motrice droite nous avons calculé, pour les quatres conditions expérimentales, les PPM^0 dont les histogrammes sont illustrés par la figure IV.29 avec le modèle *complet* en vert et le *parcimonieux 1* en rouge. Notons que les modèles sont équivalents par rapport à la condition pertinente (clic droit, IV.29-(c)) et la condition non pertinente (calcul, IV.29-(a)). La phrase et le clic gauche sont néanmoins considérés comme non pertinents par les modèles *parcimonieux* mais induisent quelques activations d'après le modèle *complet* ($PPM^0 \geq 0.9$).

Encore une fois la *procédure PPMs* illustrée par la figure IV.30 montre que les activations détectés par le modèle *complet* (en bleu) pour la condition phrase, sont faibles et disparaissent au-delà de $\gamma = 2.5$ (figures IV.30-(a)). A l'inverse, pour le clic gauche, ces activations sont plus fortes mais se limitent à une dizaine de voxels au-delà de $\gamma = 9$ (figures IV.30-(d)). Dans les deux cas, les modèles *parcimonieux* ne détectent aucune activation étant donné que ces conditions ont été considérées comme non pertinentes. Cela est également mis en évidence par les figures IV.30-(b,c) qui superposent les PPM^0 de la condition phrase au volume anatomique pour les modèles *complet* et *parcimonieux 1*, respectivement. Comme remarqué précédemment, ce dernier fourni des $PPM^0 \approx 0.5$ pour les conditions qu'il considère comme non pertinentes. De même, les PPM^0 superposées au volume anatomique pour le clic gauche sont illustrées par les figures IV.30-(e,f) correspondantes aux modèles *complet* et *parcimonieux 1*, respectivement. Encore une fois, ces deux modèles ont un comportement équivalent par rapport à la condition pertinente (clic droit) comme l'indique la *procédure PPMs* avec $\alpha = 0.95$ (figure IV.30-(g)) ainsi que la superposition des PPM^0 (figure IV.30-(h,i)) et des PPM^{15} sur le volume anatomique (figures IV.30-(j,k)) pour les modèles *complet* et *parcimonieux 1*, respectivement.

Enfin, les estimations des FRHs par les modèles *complet* (en vert), *parcimonieux 1* (en rouge) et *parcimonieux 2* (en bleu) sont illustrées par la figure IV.31 et comparées à la FRH canonique (en noir). Ceci est effectué pour la ROI motrice droite (en haut) et gauche (en bas). Comme pour les autres données, la différence entre les modèles n'est pas très significative. Leurs estimations de la FRH sont cependant différentes de la forme canonique.

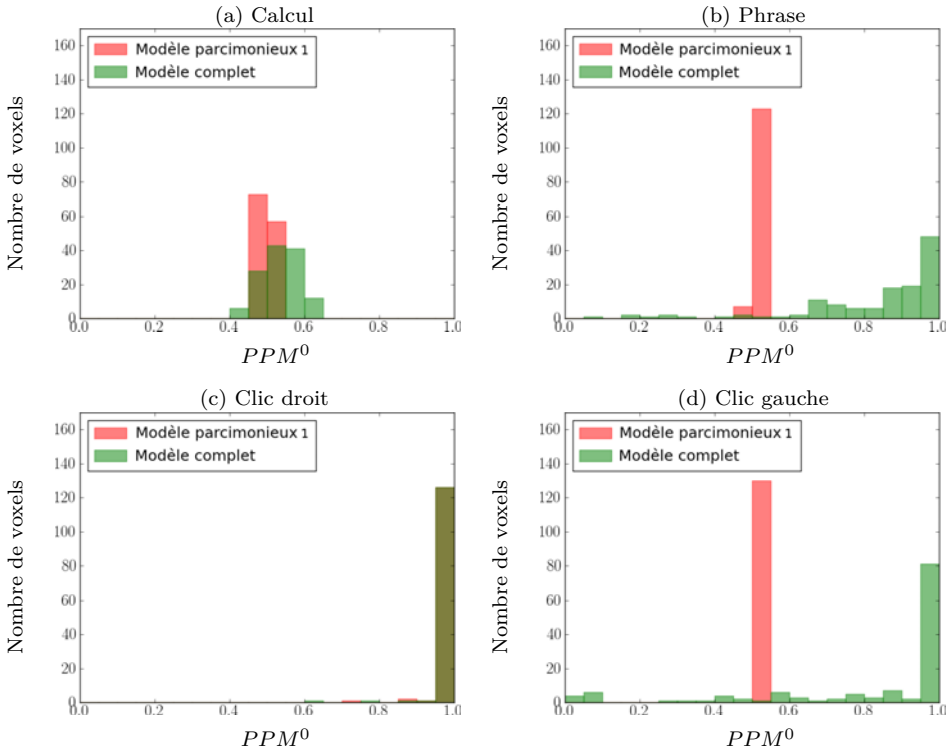


FIGURE IV.29 – Histogrammes des PPM^γ avec $\gamma = 0$ dans la ROI motrice gauche pour les quatres conditions expérimentales (calcul, phrase, clic droit et clic gauche). Le modèle *complet* est représenté en vert et le modèle *parcimonieux 1* en rouge (le modèle *parcimonieux 2* ayant les mêmes résultats que le *1* il n'a pas été représenté). Les modèles *parcimonieux* et *complet* ont des comportements équivalents dans le cas de la condition pertinente (clic droit, c) et de la condition non pertinente (calcul, a) à l'inverse des deux conditions non pertinentes (phrase, b et clic gauche, d) pour lesquelles le modèle *complet* détecte des activations ($PPM^\gamma > 0.9$) alors que les modèles *parcimonieux* les considèrent comme non pertinentes.

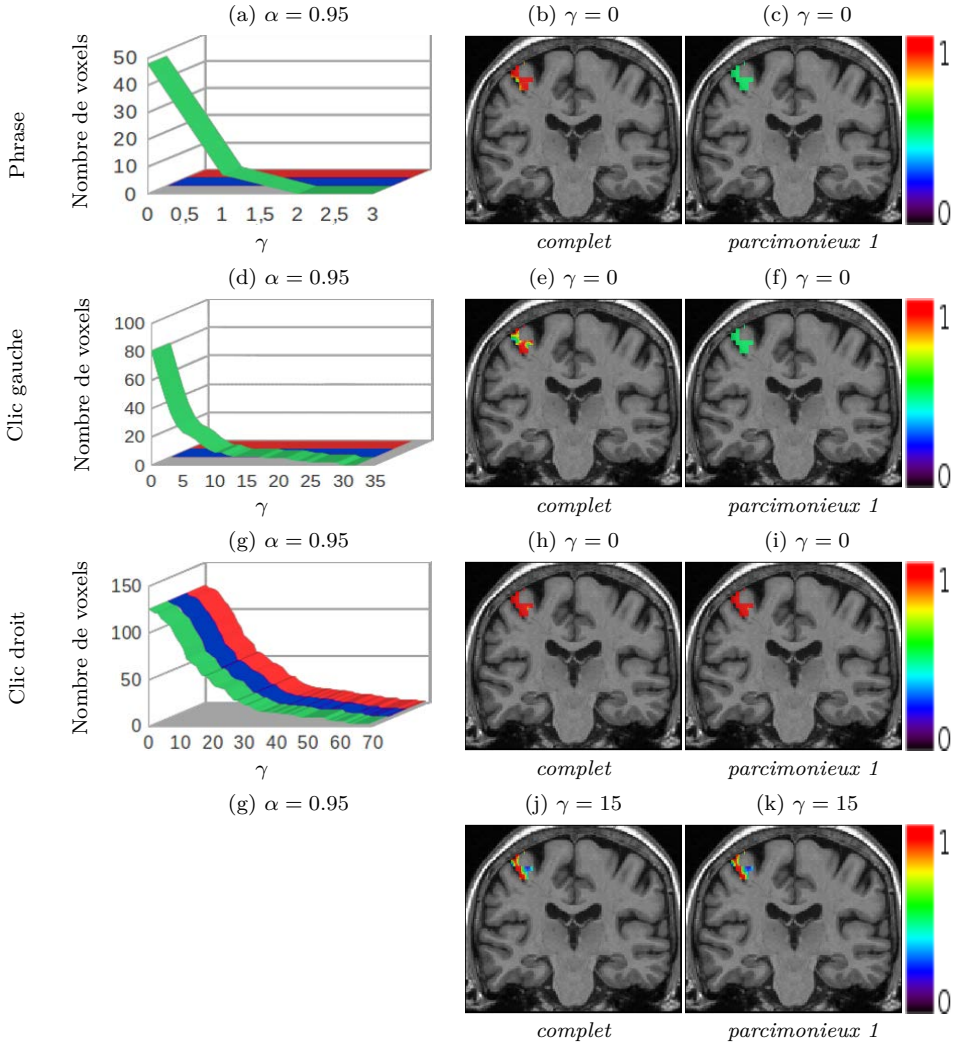


FIGURE IV.30 – (a,d,g) Nombre de voxels obtenu après seuillage des PPM^{γ} avec $\gamma = 0 : 3$ pour la phrase (a), $\gamma = 0 : 35$ pour le clic gauche (d) et $\gamma = 0 : 70$ pour le clic droit (g), ceci dans la ROI motrice gauche et pour un niveau de seuillage $\alpha = 0.95$. Le modèle *complet* est représenté en vert, le *parcimonieux 1* en rouge et le *parcimonieux 2* en bleu. Le modèle *complet* détecte des activations correspondantes aux conditions phrase et clic gauche considérées comme non pertinentes par les modèles *parcimonieux*. Cette différence de comportement entre les modèles est également illustrée par (b,c) représentant la superposition avec le volume anatomique des PPM^0 pour la phrase et pour les modèles *complet* et *parcimonieux* respectivement. De même pour les figures (e,f) correspondantes au clic gauche. Superposition des PPM^0 (h,i) et des PPM^{15} (j,k) pour le clic droit et pour les modèles *complet* (à gauche) et *parcimonieux 1* (à droite), ces derniers ayant des comportement similaires.

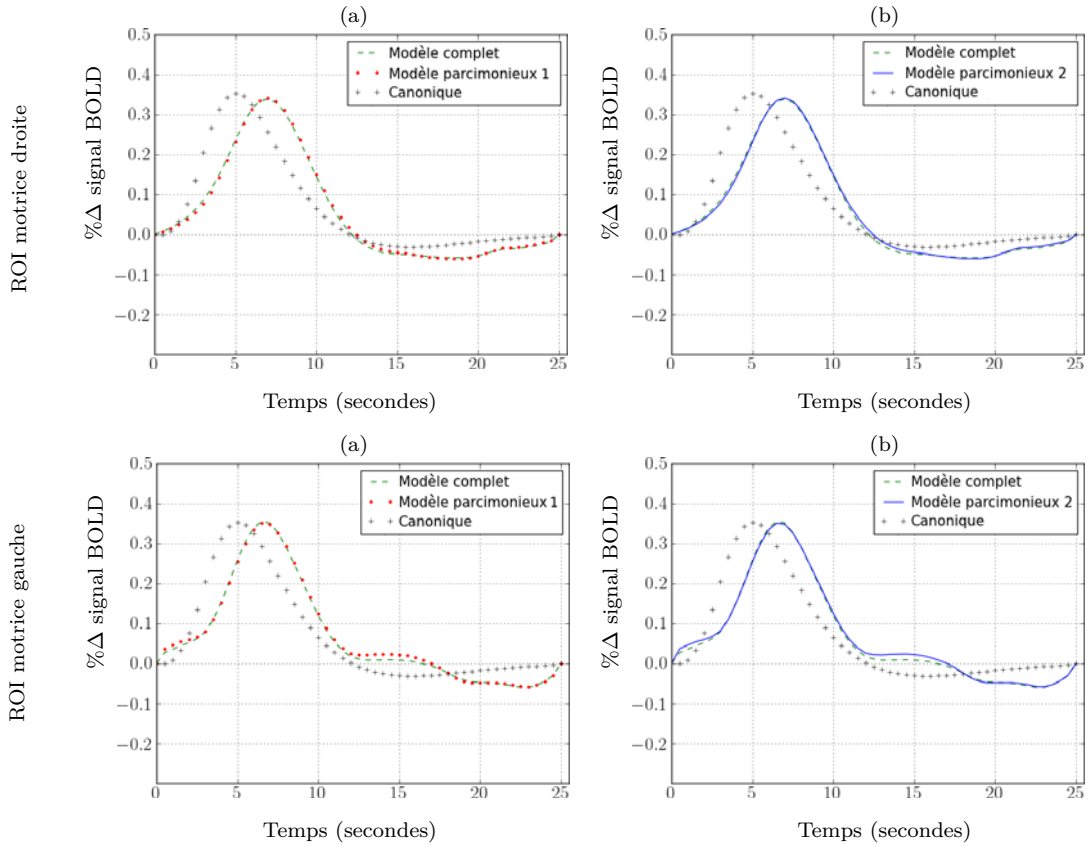


FIGURE IV.31 – Comparaison des FRHs estimées par les modèles *complet* (en vert), *parcimonieux 1* (en rouge) et *parcimonieux 2* (en bleu) dans la ROI motrice droite (en haut) et gauche (en bas). La FRH canonique est affichée en noir pour montrer la différence avec les FRHs estimées. Les estimations des trois modèles sont très proches.

IV.2.2.3 Conclusion

Pour le second sujet, les modèles *complet* et *parcimonieux* ont des comportements similaires par rapport aux conditions pertinentes. Leurs différences se focalisent sur les conditions considérées comme non pertinentes par le modèle *parcimonieux*. Ces conditions étant le clic droit dans le ROI motrice droite et la phrase et le clic gauche dans la ROI motrice gauche. Il est cependant connu que, pour des tâches motrices, de faibles activations ipsilatérales sont détectées [Biswal 1995]. On conclut donc deux possibilités :

1- Soit la ROI motrice droite (resp. gauche) extraite ne contient pas les voxels représentant les activations ipsilatérales dans le sens où elle est extraite à partir du contraste T *clic gauche - clic droit* (resp. *clic droit - clic gauche*). Dans ce cas, les voxels ipsilatéraux détectés par le modèle *complet* représentent des FP que le modèle *parcimonieux* élimine.

2- Soit ces voxels sont bien présents car un contraste *clic gauche - clic droit* (resp. *clic droit - clic gauche*) illustre les voxels majoritairement activés par le clic gauche (resp. clic droit) mais pas les voxels activés par le clic gauche (resp. clic droit) et non par le clic droit (resp. clic gauche). Dans ce cas, les activations ipsilatérales étant faibles, elles ont été éliminées par notre modèle *parcimonieux*, qui dans un certain sens effectue une sorte de seuillage. Cela induit un cas de FN dans le modèle *parcimonieux* mais sans grande conséquence.

Finalement, on note que le choix du paramètre τ_2 paraît plus compliqué pour le modèle *parcimonieux 1* que pour le *parcimonieux 2* dans le sens où sa valeur change d'un sujet à l'autre. Pour une analyse VEM, un τ_2 équivalent à 15% de la taille de la ROI était suffisant à l'élimination des conditions non pertinentes dans le cas du premier sujet alors qu'il a fallu un τ_2 équivalent à 60% de la taille de la ROI pour le deuxième sujet. Etant donné la procédure utilisée pour le choix de la valeur de τ_2, son application pour chaque sujet et chaque ROI est très coûteuse en temps. A l'inverse, le modèle *parcimonieux 2* paraît plus stable de ce point de vue dans le sens où aucun changement de paramètre n'a été nécessaire entre les deux sujets. Evidemment cette conclusion devrait être vérifiée sur un plus grand nombre de ROIs et de sujets avant d'être confirmée. De plus le modèle *parcimonieux 2* nous permet un gain en temps de calcul étant donné la présence d'une étape MC dans le *parcimonieux 1*. Notons aussi que l'élimination des activations ipsilatérales par le modèle *parcimonieux* ne vient pas du fait que

τ_2 est élevé car nous avons fait une analyse avec $\tau_2 = 0$ dans laquelle la condition clic droit (resp. clic gauche) dans la ROI motrice droite (resp. gauche) s'éteint directement dans le modèle *parcimonieux* et ce sont le calcul et la phrase qui nécessitent l'augmentation du τ_2 pour être éliminés. Sachant que ces conditions ne sont pas supposées induire d'activations sur le sillon central on peut confirmer que ce comportement vient de l'influence du bruit.

Chapitre V

Analyses cerveau entier

Les résultats illustrés dans le chapitre IV concernant les données réelles se focalisent sur une région cérébrale en particulier (la région motrice). Nous allons dans ce chapitre illustrer quelques résultats préliminaires concernant des analyses cerveau entier et dont la compréhension nécessite des études plus approfondies ainsi qu'une comparaison avec la littérature. Ces analyses ont été effectuées dans le cadre de l'inférence VEM et avec les trois modèles *parcimonieux* précédemment décrits : le modèle *parcimonieux 1* exprimant la pertinence en terme du nombre de voxels activés, le modèle *parcimonieux 2* l'exprimant en terme de moyenne d'activation, et enfin, le modèle *parcimonieux* utilisant un *a priori* ARD sur les variables de pertinence $\boldsymbol{w}$ au lieu de la loi de Bernoulli utilisée pour les deux premiers.

V.1 Résultats préliminaires

Les données auxquelles on s'intéresse par la suite correspondent à celle du deuxième sujet, une femme de 65 ans, et dont les régions d'intérêt motrices droite et gauche ont été traitées dans la sous-section IV.2.2 du chapitre IV. Rappelons que ces données sont issues d'un paradigme événementiel rapide composé de 8 conditions expérimentales qui ont été regroupées en quatres types pour obtenir plus de contraste. Ces quatres conditions étant : calcul, phrase, clic droit et clic gauche, chacune est représentée sous deux modalités, auditive et visuelle. La parcellisation cerveau entier utilisée est obtenue à partir d'un algorithme de clustering hiérarchique de Ward basé sur la minimisation de la variance intra-classe en tenant compte d'une contrainte de connexité. C'est une parcellisation fonction-

nelle basée sur les cartes de contraste de SPM. Le nombre de parcelles pouvant être choisi par l'utilisateur nous avons choisi d'effectuer la parcellisation sur 400 parcelles, en cohérence avec ce qui a été proposé dans [Thyreau 2006].

Modèle *parcimonieux 1.* Dans le cadre de cette approche, le seuil de pertinence exprimé par le paramètre τ_2 est choisi par l'utilisateur. Rappelons que ce paramètre représente le seuil au-dessus duquel une condition a une forte chance *a priori* d'être considérée comme pertinente. Notons que pour les ROIs motrices de ce sujet la valeur de τ_2 nécessaire à l'élimination des conditions non pertinentes était de 60% de la taille de la ROI. Nous avons donc effectué l'analyse cerveau entier avec cette même valeur. Cela nous a permis d'obtenir les figures V.1-V.4(a) qui illustrent la superposition des probabilités de pertinences *a posteriori* estimées $\widetilde{p}_{w^m}(1)$, sur le volume anatomique, et cela pour les quatres conditions expérimentales : calcul ($m = 1$), phrase ($m = 2$), clic droit ($m = 3$) et clic gauche ($m = 4$), respectivement. Ces cartes de probabilités indiquent les ROIs où les conditions ont été considérées comme pertinentes avec, en couleur, la valeur estimée de la probabilité *a posteriori* $\widetilde{p}_{w^m}(1)$.

Modèle *parcimonieux 2.* Ici, contrairement au modèle *parcimonieux 1* le paramètre τ_2 est estimé. Il est cependant nécessaire de choisir les hyperparamètres de sa loi *a priori* gamma. Dans le chapitre IV, nous avons conclu que pour la ROI motrice les meilleures valeurs étaient $\alpha_{\tau_2} = 9$, $\lambda_{\tau_2} = 16$ illustrant un *a priori* de mode égale à 0.5 et de variance égale à 0.03. Rappelons aussi qu'étant données les faibles valeurs de τ_2 nous avons augmenté la pente de la sigmoïde illustrée par τ_1 de manière à ce que la probabilité à l'origine soit égale à $p_0 = 10^{-3}$. L'analyse cerveau entier a été effectuée pour ces mêmes valeurs. Comme pour le modèle *parcimonieux 1*, les figures V.1-V.4(b) illustrent la superposition des probabilités de pertinences *a posteriori* estimées $\widetilde{p}_{w^m}(1)$, sur le volume anatomique, et cela pour les quatres conditions expérimentales : calcul ($m = 1$), phrase ($m = 2$), clic droit ($m = 3$) et clic gauche ($m = 4$), respectivement.

Notons que les figures illustrées sont en convention neurologique, le côté gauche sur les figures correspond au côté gauche du cerveau.

♣ Condition calcul

Dans les figures V.1-(a,b) illustrant, pour les modèles *parcimonieux 1* et *2* respectivement, la probabilité de pertinence *a posteriori* $\widetilde{p}_{w^m}(1)$ estimée pour la

condition calcul, on remarque la présence d'activations frontales et pariétales assez attendues dans le cas d'un calcul mental. Il faudrait néanmoins effectuer une étude plus appronfondie permettant de se comparer à la littérature, comme par exemple aux travaux effectués dans [Knops 2009] qui s'intéressent au calcul mental et qui, de plus, mettent en évidence l'implication d'aires cérébrales correspondantes à l'attention spatiale. Rappelons aussi que les modalités auditive et visuelle de présentation de cette condition induisent des activations dans les lobes occipital et temporal.

♣ Condition phrase

De même que pour le calcul, les figures V.2-(a,b) illustrent, pour les modèles *parcimonieux 1* et *2* respectivement, les parcelles où la condition phrase a été considérée comme pertinente. Les probabilités de pertinence *a posteriori* $\widetilde{p}_{w^m}(1)$ sont codées en couleur. Là aussi, on observe un réseau d'activation assez distribué qui nécessite d'être étudié finement pour le comparer à la littérature et vérifier si nos résultats coïncident en terme de coordonnées dans l'atlas MNI sur lequel sont normalisées nos données. Il serait bien, par exemple, de vérifier l'activation des aires de Broca et Wernicke impliquées dans le langage en plus du sillon temporal supérieur (STS). Il ne faut bien sûr pas omettre les activations occipitales et temporales dues aux modalités de présentation des conditions.

♣ Conditions clic droit et gauche

Ici, le modèle *parcimonieux 1* fournit des cartes de pertinence plus spécifiques que le modèle *parcimonieux 2*, dans le sens où il détecte plus de non pertinence que ce dernier surtout pour le clic droit. Ce qui est assez remarquable est le fait que dans les figures V.3-(b) pour le clic droit et V.4-(a,b) pour le clic gauche, la plupart du cerveau est considérée comme activée par ces conditions.

Si on se focalise sur le clic droit par exemple, on remarque d'après les NRNs illustrés dans les figures V.3-(c,d) pour les modèles *parcimonieux 1* et *2*, respectivement, qu'il y a bien plus d'activation détectée par le deuxième modèle, la question étant : *lequel des deux est le plus proche de la réalité ?* Une réponse nécessite une étude détaillée et comparative des deux modèles et surtout une comparaison avec la littérature.

On remarque également l'apparition de la SMA (*Supplementary Motor Area*) ainsi que de l'activation contralatérale. De plus, pour le clic droit dans le cas du modèle *parcimonieux 1* par exemple (figure V.3-(c)), on retrouve la ROI motrice droite traitée dans le chapitre IV (entourée en jaune) et dans laquelle le clic droit a été considéré comme non pertinent. L'activation ipsilatérale n'est cependant pas complètement perdue, car le modèle *parcimonieux* déclare le clic droit comme pertinent dans la partie du sillon central droit entourée en bleu dans les figures V.3-(a,c). Cette activation n'est pas évidente à voir car les NRNs correspondants sont de l'ordre de 2 ou 3 par rapport à 75. Signalons aussi que dans le paradigme en question les tâches motrices consistent à cliquer 3 fois sur un bouton, il est donc possible d'observer des activations dues au fait que certains sujets comptent en effectuant ces tâches.

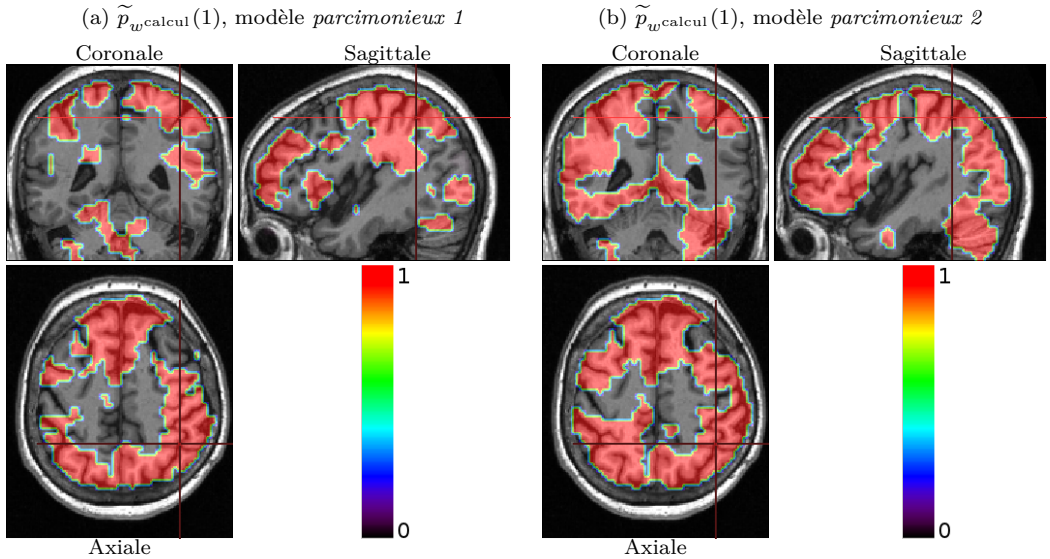


FIGURE V.1 – Illustration, pour la condition calcul ($m = 1$) présentée sous forme auditive et visuelle, des probabilités de pertinence *a posteriori* $\widetilde{p}_{w^m}(1)$ estimées par le modèle *parcimonieux 1* (a) et *2* (b). Notons que ce dernier à tendance à détecter plus de pertinence que le premier.

Modèle *parcimonieux* avec ARD. Rappelons que dans ce modèle l'*a priori* de Bernoulli utilisée dans les approches *parcimonieuses* précédentes a été remplacé par un *a priori* gaussien ($\boldsymbol{w} \sim \prod_m \mathcal{N}(0, \alpha_m^{-1})$) où les variables de pertinences $\boldsymbol{w}$ ne sont plus binaires et n'ont plus de connexion *a priori* avec les variables et paramètres du modèle. Ici, le modèle nécessite le choix des hyperparamètres $(k_\alpha, \lambda_\alpha)$ de l'*a priori* gamma sur la précision α_m qui ont été choisis de façon

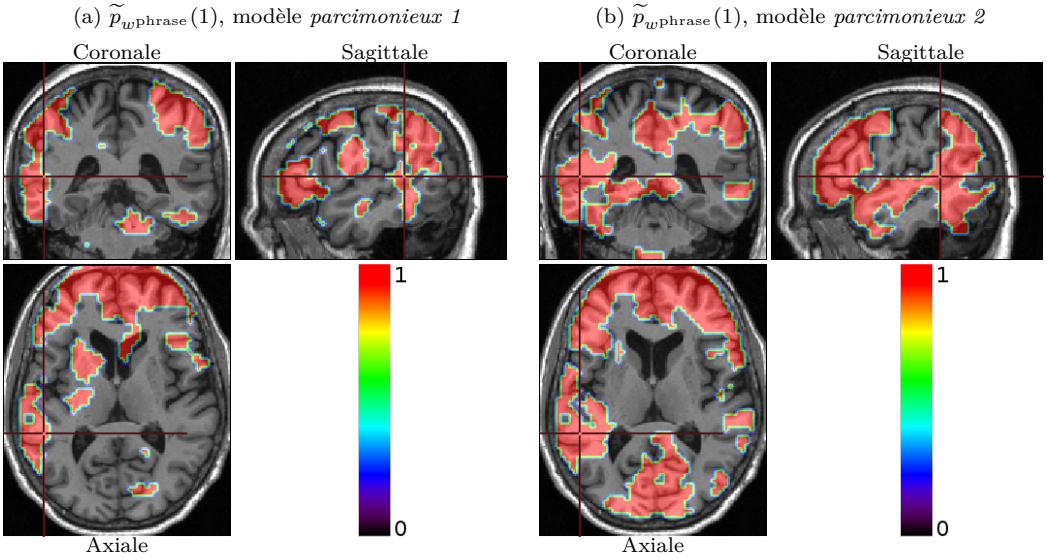


FIGURE V.2 – Illustration, pour la condition phrase ($m = 2$) présentée sous forme auditive et visuelle, des probabilités de pertinence *a posteriori* $\widetilde{p}_{w^m}(1)$ estimées par le modèle *parcimonieux 1* (a) et *2* (b). Notons que ce dernier à tendance à détecter plus de pertinence que le premier.

à obtenir un *a priori* non informatif ($k_\alpha, \lambda_\alpha \to 0$), ce qui n'était pas possible pour le paramètre τ_2 dans le cas du modèle *parcimonieux 2*. En procédant ainsi, on se libère du problème de choix des paramètres de la sigmoïde ainsi que des hyperparamètres de leur *a priori*. Ce détail pourrait représenter l'intérêt de ce nouveau modèle par rapport aux deux autres approches *parcimonieuses*. On remarque en revanche, un autre problème induit par le produit $\boldsymbol{Aw}$. Etant donné cette bilinéarité, une partie de $\boldsymbol{a}^m = \{a_j^m; j = 1 : J\}$ est factorisée dans w^m pour chaque condition m (ce problème n'étant pas présent avec des $\boldsymbol{w}$ binaires). Une comparaison avec le modèle *complet* nécessite donc la considération du produit $\boldsymbol{Aw}$ et non seulement de la variable $\boldsymbol{A}$. Cela complique un peu le calcul des PPMs dans le sens où, pour chaque voxel j et chaque condition m, le calcul de la quantité $PPM_j^\gamma = p(a_j^m w^m > \gamma)$ nécessite la connaissance de la loi du couple $a_j^m w^m$ qui est celle du produit de deux variables gaussiennes.

Calcul de PPM pour le produit de deux variables gaussiennes. Le produit de deux variables aléatoires gaussiennes peut être représenté comme la combinaison de deux variables aléatoires suivant des lois de χ^2. On peut, en effet,

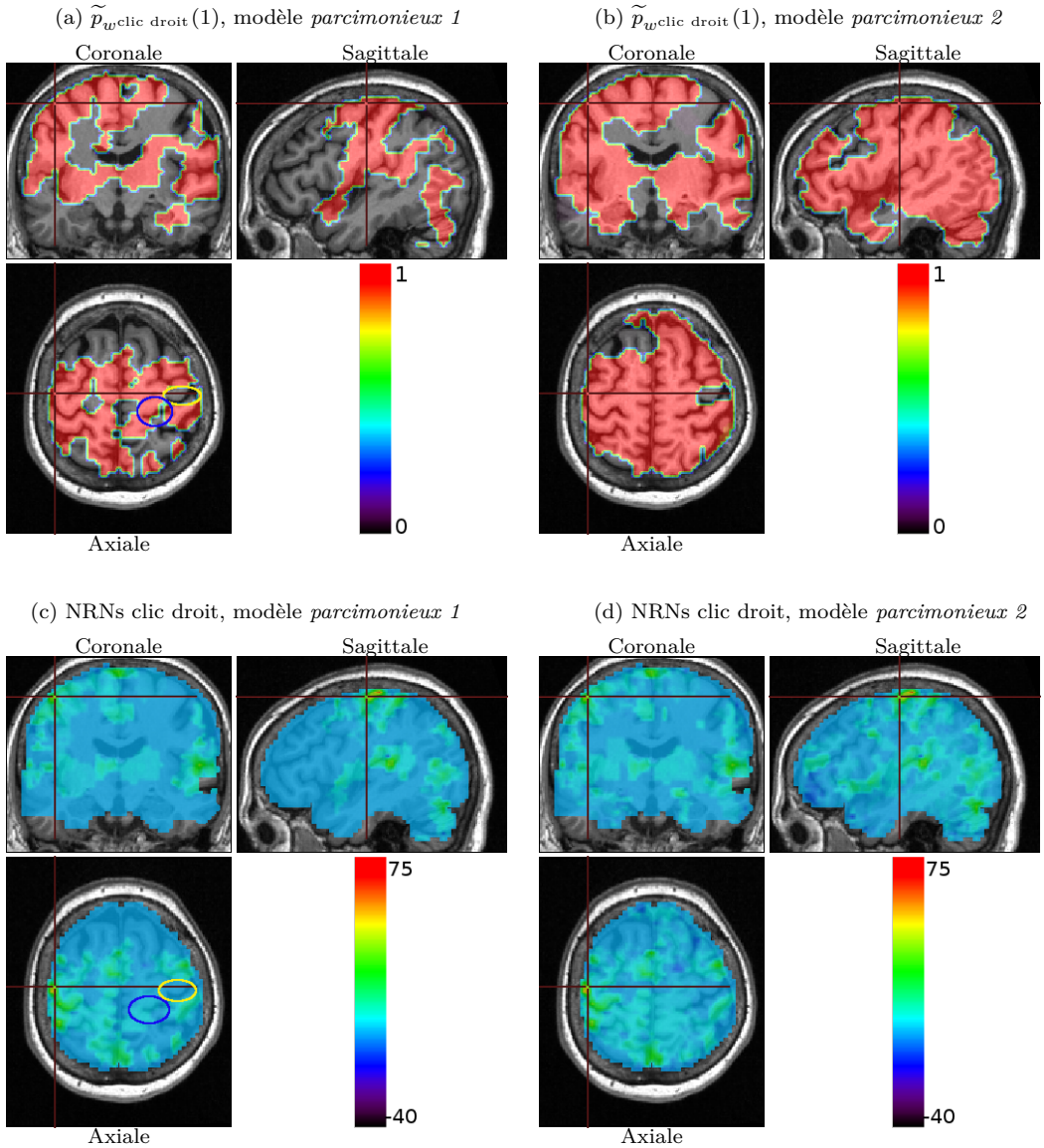


FIGURE V.3 – Illustration, pour la condition clic droit ($m = 3$) présentée sous forme auditive et visuelle, des probabilités de pertinence *a posteriori* $\widetilde{p}_{w^m}(1)$ (a,b) ainsi que des NRNs correspondants (c,d). Cela pour les modèles *parcimonieux 1* (a,c) et *2* (b,d). Notons que ce dernier à tendance à détecter plus de pertinence que le premier.

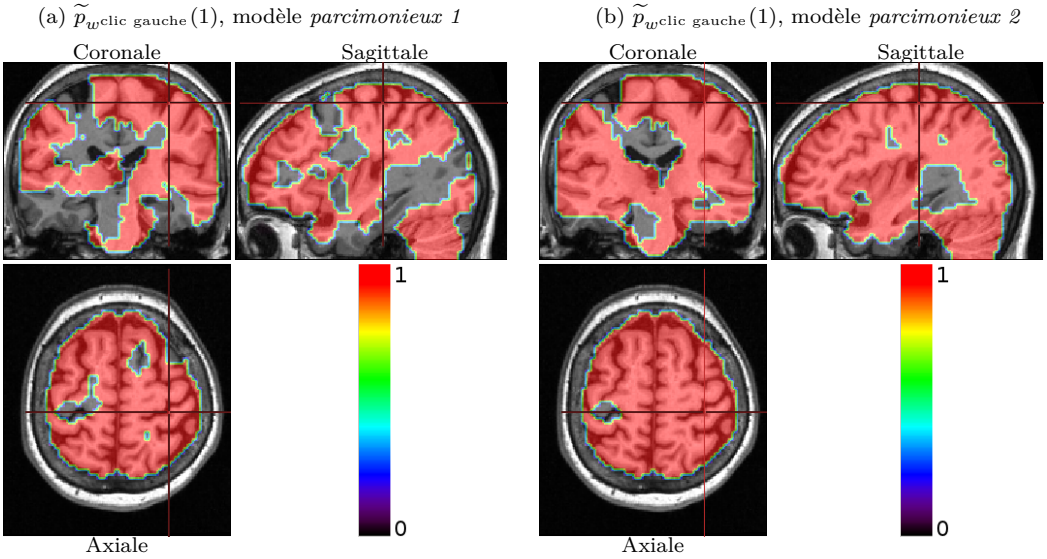


FIGURE V.4 – Illustration, pour la condition clic gauche ($m = 4$) présentée sous forme auditive et visuelle, des probabilités de pertinence *a posteriori* $\widetilde{p}_{w^m}(1)$ (a,b) pour les modèles *parcimonieux 1* et *2*, respectivement. Notons que ce dernier à tendance à détecter plus de pertinence que le premier.

écrire :

$$a_j^m w^m = \frac{1}{4}(a_j^m + w^m)^2 - \frac{1}{4}(a_j^m - w^m)^2$$

a_j^m et w^m étant de lois $\mathcal{N}(m_{a_j^m}, v_{\boldsymbol{a}_j^{m,m}})$ et $\mathcal{N}(m_{w^m}, v_{w^{m,m}})$ respectivement, les variables aléatoires $(a_j^m + w^m)^2$ et $(a_j^m - w^m)^2$ suivent des lois de χ^2. Ces dernières ne sont pas centrées car $m_{a_j^m} \neq 0$ et $m_{w^m} \neq 0$ en général. La distribution résultante peut être identifiée [Castaño-Martínez 2005] mais une manière simple d'estimer les PPMs est de le faire par simulation. On peut simuler des échantillons de taille S de chacune de nos deux variables afin d'obtenir ensuite l'échantillon $\{a_{j,1}^m w_1^m, ..., a_{j,S}^m w_S^m\}$ à l'aide duquel on peut calculer la fonction de répartition de la loi du couple $F_{a_j^m w^m}(\gamma)$. Cela permettra de calculer la PPM en question :

$$PPM_j^\gamma = p(a_j^m w^m > \gamma) = 1 - p(a_j^m w^m < \gamma) = 1 - F_{a_j^m w^m}(\gamma) \tag{V.1}$$

Cela peut être couteux en temps de calcul car la procédure doit être effectuée pour chaque voxel j et chaque condition m.

Les figures V.5 illustrent le produit $\boldsymbol{m}_{\boldsymbol{a}^m} w^m$ avec $\boldsymbol{m}_{\boldsymbol{a}^m} = \{m_{a_j^m}; j = 1 : J\}$ où $m_{a_j^m}$ représente la moyenne estimée de la distribution *a posteriori* du NRN

correspondant au voxel j et à la condition m. Cela est effectué pour les quatre conditions expérimentales : calcul $m = 1$ (a), phrase $m = 2$ (b), clic droit $m = 3$ (c) et clic gauche $m = 4$ (d). Leurs superpositions sur le volume anatomique indique la distribution spatiale des activations détectées. Comme pour les précédents résultats, une étude plus détaillée reste encore à faire afin de mieux comprendre ces résultats.

V.2 Conclusion

Une différence de comportement entre les modèles *parcimonieux* est observée sur les résultats préliminaires de l'analyse cerveau entier présentée précédemment. En effet, le modèle *parcimonieux 1* détecte plus de non pertinence que le modèle *parcimonieux 2*. Ce dernier considère par exemple les conditions clic gauche et clic droit comme pertinentes pour l'ensemble du cerveau. Notons que le seuillage effectué à travers la sigmoïde n'est pas comparable entre les deux modèles. Dans le premier modèle, le seuil de pertinence τ_2 est équivalent à 60% de la taille d'une ROI, ce qui représente un *a priori* plus contraint que celui du deuxième modèle. Nous avons donc effectué une deuxième analyse avec le modèle *parcimonieux 2* en lui imposant un *a priori* plus contraint. Pour cela nous avons choisi de fixer les hyperparamètres de la loi *a priori* sur τ_2 à $\alpha_{\tau_2} = 61$, $\lambda_{\tau_2} = 12$. Ici le mode de cet *a priori* est de 5 avec une variance de 0.4 par rapport à un mode de 0.5 et une variance de 0.03 dans le cas précédent. Les valeurs estimées de τ_2 sont alors dans l'intervalle $[3.43, 5]$. Malheureusement nous n'avons observé aucune différence par rapport à l'analyse précédente. Dans le cadre de ses observations deux importantes pistes sont à explorer :

1- **La modélisation d'un bruit autorégressif.** Nos modèles ont été développés avec une modélisation de bruit blanc $\boldsymbol{b}_j \sim \mathcal{N}\left(\mathbf{0}, \sigma_{\mathbf{j}}^{\mathbf{2}}\boldsymbol{\Lambda}_{\mathbf{j}}\right)$ où $\boldsymbol{\Lambda}_j = \boldsymbol{I}_N$. Cependant, et comme précisé dans la section II.1.3 du chapitre II, il est préférable de modéliser la corrélation temporelle du bruit. Une telle modélisation améliorerait l'adéquation de notre modèle aux données et cela pourrait apporter également des améliorations en terme de détection de pertinence. Quand un bruit de nature autorégressive est modélisé comme indépendant, il se peut que la partie mal modélisée de ce bruit passe dans le signal stimulus-induit compliquant ainsi l'élimination des conditions non pertinentes. Il serait donc nécessaire de vérifier l'influence de la modélisation du bruit sur nos modèles *parci-*

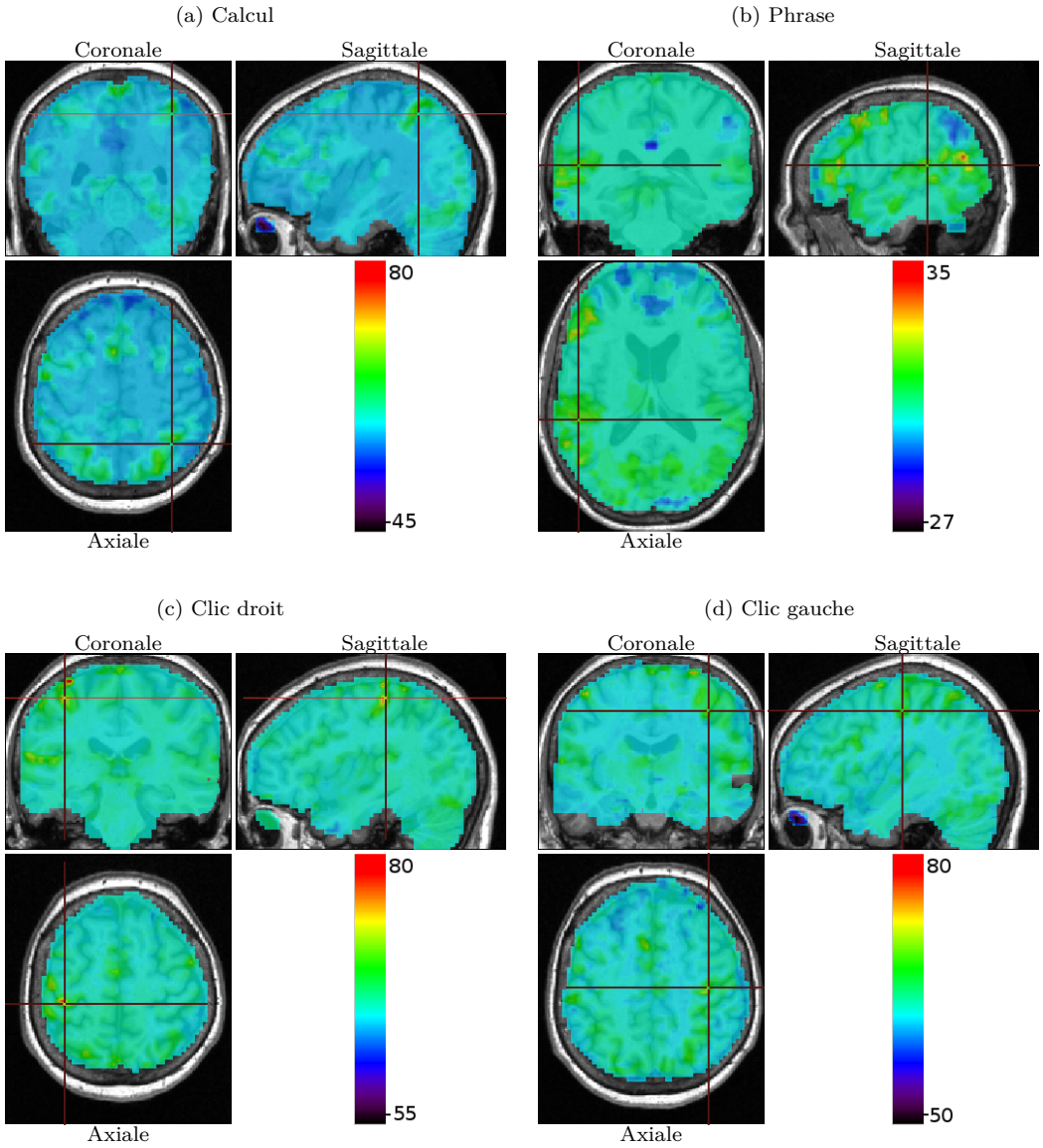


FIGURE V.5 – Illustration, du produit $\boldsymbol{m}_{\boldsymbol{a}^m} w^m$ superposé sur le volume anatomique avec $\boldsymbol{m}_{\boldsymbol{a}^m} = \{m_{a_j^m}; j = 1 : J\}$ où $m_{a_j^m}$ est la moyenne estimée de la distribution *a posteriori* du NRN correspondant au voxel j et à la condition m. Cela pour les quatre conditions expérimentales : calcul $m = 1$ (a), phrase $m = 2$ (b), clic droit $m = 3$ (c) et clic gauche $m = 4$ (d). Ces figures représentent des résultats préliminaires obtenus avec le modèle *parcimonieux* utilisant un *a priori* ARD sur la variable de pertinence $\boldsymbol{w}$.

monieux. Plusieurs types de modélisation sont possibles comme nous l'avons discuté dans la section II.1.3 du chapitre II. Une possibilité serait aussi, pour un bruit autorégressif d'ordre 1 (AR(1)), de définir la matrice $\mathbf{\Lambda}_j$ de la même manière que [Chaari 2013] où ce n'est plus une matrice identité mais une matrice tridiagonale symmétrique dépendant des paramètres du bruit AR(1) ρ_j tel que : $(\mathbf{\Lambda}_j)_{1,1} = (\mathbf{\Lambda}_j)_{N,N} = 1$, $(\mathbf{\Lambda}_j)_{n,n} = 1 + \rho_j^2$ pour $n = 2 : N-1$, $(\mathbf{\Lambda}_j)_{n+1,n} = (\mathbf{\Lambda}_j)_{n,n+1} = -\rho_j$ pour $n = 1 : N-1$ et $(\mathbf{\Lambda}_j)_{.,.} = 0$ autrement.

2- **L'optimisation du paradigme expérimental.** Le *Localizer* en question représente, par la diversité de ses conditions expérimentales, un exemple assez complexe. Rappelons que ce paradigme comprend 4 types de conditions : calcul, phrase, clic droit et clic gauche, chacune présentée sous deux modalités, auditive et visuelle. De plus, même dans la condition clic droit visuel par exemple, il y a, en quelque sorte, une activité de lecture et de compréhension, sans omettre le fait que certains sujets peuvent compter en cliquant lorsque plusieurs clics sont demandés à la fois. Tout cela pour mettre en évidence la complexité du paradigme par rapport au problème de pertinence qu'on cherche à valider. Il serait donc intéressant de mettre en place un paradigme plus simple comme par exemple de demander au sujet de cliquer une seule fois (pour éviter de compter) lorsqu'une croix apparaît sur l'écran. Il est évident qu'un paradigme de la sorte n'est pas ce que l'on fait en général dans les expériences IRMf qui s'intéressent à des tâches plus complexes, mais cela permettrait d'asseoir la validité du modèle *parcimonieux* sur des cas bien maîtrisés, pour ensuite étudier des paradigmes plus sophistiqués.

Chapitre VI

Discussion et Perspectives

Ce travail de thèse s'est intéressé principalement à la pertinence des conditions expérimentales dans le traitement des données IRM fonctionnelle. Etant donné que la sous-famille des conditions pertinentes change d'une région cérébrale à l'autre et d'un cas sain à un cas pathologique, nous avons proposé d'effectuer la sélection adaptativement et simultanément à l'analyse. Cette sélection est opérée dans un cadre bayésien de Détection-Estimation Conjointe (DEC) permettant, en plus de la détection de l'activité cérébrale, l'estimation de la Fonction de Réponse Hémodynamique (FRH), ce qui présente un avantage comparativement au MLG par exemple. Deux modèles *parcimonieux* ont principalement été développés et testés sur données simulées et réelles et comparés au modèle *complet* qui considère toutes les conditions comme pertinentes.

Les deux modèles *parcimonieux 1* et *2* se posent dans le cadre d'une sélection de variables en introduisant un groupe de variables binaires $\boldsymbol{w} = \{w^m; m = 1 : M\}$ où w^m exprime la pertinence de la $m$ième condition expérimentale du paradigme. Ces modèles proposent, pour chaque condition m, un *a priori* de Bernoulli avec une probabilité de succès donnée par une sigmoïde. La particularité de cette dernière étant le passage non abrupt de 0 à 1 autour de son point d'inflexion et dont l'abscisse peut être vue comme un seuil de pertinence au delà duquel la condition m à une forte probabilité *a priori* d'être considérée comme pertinente ($w^m = 1$). Le premier modèle exprime la pertinence en terme du nombre de voxels activés, autrement dit, si la condition m correspond à un nombre d'activations supérieur au seuil de pertinence, la probabilité *a priori* que w^m soit estimée à 1 est élevée. Le second modèle *parcimonieux* quant à lui exprime la pertinence en terme de moyenne d'activation.

Les analyses effectuées sur les données simulées et réelles (principalement des régions motrices) nous ont permis d'observer, dans le cadre du modèle *parcimonieux*, une grande amélioration quant à la classification des voxels, tous attribués à la classe inactive quand une condition est jugée non pertinente, contrairement au modèle *complet* où une classification aléatoire induit beaucoup de faux positifs.

Nous nous sommes également intéressés aux PPMs qui représentent les probabilités *a posteriori* que les niveaux d'activations estimés soient supérieurs à un seuil donné. Ici, nous avons observé que l'élimination des conditions non pertinentes par le modèle *parcimonieux* permettrait l'obtention de PPMs plus élevées pour les conditions pertinentes. Il est important de comprendre que toute amélioration pouvant être apportée par le modèle *parcimonieux* dépend de l'importance de la dégradation occasionnée par la présence des conditions non pertinentes. Plus ces dernières auront d'influence sur l'estimation des paramètres du modèle plus nos approches *parcimonieuses* seront intéréssantes.

Un troisième modèle *parcimonieux* a également été développé en remplaçant l'*a priori* de Bernoulli par un *a priori* ARD s'éloignant ainsi du principe de sélection de variable. L'idée étant de pouvoir diminuer la complexité du choix de certains paramètres due à la présence de la sigmoïde.

D'un autre côté, l'analyse des résultats préliminaires sur cerveau entier indique différents axes pour améliorer nos approches *parcimonieuses* et renforcer leur validation. Le premier concerne la modélisation d'un bruit autorégressif permettant de diminuer l'influence d'une mauvaise modélisation du bruit sur la détection de la pertinence. Le second, quant à lui, se focalise sur l'optimisation et la simplification du paradigme expérimental par rapport à la diversité des types de stimuli afin d'augmenter la confiance en les résultats attendus et de faciliter ainsi, dans un premier temps, la validation de l'approche *parcimonieuse*.

De plus, une perspective à long terme serait de se focaliser sur l'analyse de groupe. Cette dernière étant incontournable en IRMf car elle permet, en étudiant un grand nombre de sujets, de généraliser un certain comportement ou même de détecter un caractère de faible amplitude pouvant être masqué par l'analyse statistique individuelle. C'est sur ce dernier point que nous allons nous concentrer, étant donné que la présence de seuillage dans nos modèles *parcimonieux* pourrait être particulièrement problématique dans le cas d'une analyse de groupe car l'effet en question pourrait être éliminé au niveau de chaque sujet. Pour cela il serait

nécessaire de généraliser notre approche de manière à ce que la décision de la pertinence soit prise au niveau du groupe. Solveig Badillo propose par exemple un modèle DEC multi-sujets dans ses travaux de thèse [Badillo 2013a].

Finalement, on tient à préciser que nos modèles ont été implémentés en langage Python et C dans le logiciel PyHRF (http://pyhrf.org/).

Annexe A

Matrice de covariance de l'approche RIF régularisée

La matrice de covariance de la FRH dans l'approche RIF régularisée temporellement (section II.3.2) est définie par $\boldsymbol{R} = \left(\frac{\boldsymbol{D}_2^{\mathrm{t}}\boldsymbol{D}_2}{(\Delta t)^4}\right)^{-1}$. Cette matrice permet d'introduire une contrainte sur la dérivée seconde afin de pénaliser les fortes variations de pentes et d'obtenir donc des formes de FRH lisses. Δt est le pas de discrétisation de la FRH et $\boldsymbol{D}_2$ la matrice des différences secondes associées au vecteur $\boldsymbol{h}_j^m$ et calculée à partir de l'équation suivante :

$$\frac{\mathrm{d}^2\boldsymbol{h}_j^m}{\mathrm{d}t^2}(d\Delta t) \approx \left\{\frac{h_{j,(d+1)\Delta t}^m - 2h_{j,d\Delta t}^m + h_{j,(d-1)\Delta t}^m}{(\Delta t)^2}\ ;\ d{=}1{:}D-1\right\} \approx \frac{\boldsymbol{D}_2\boldsymbol{h}_j^m}{(\Delta t)^2} \quad \text{(A.1)}$$

On en déduit la matrice $\boldsymbol{D}_2$ de dimensions $(D-1)\times(D+1)$:

$$\boldsymbol{D}_2 = \begin{pmatrix} 1 & -2 & 1 & 0 & ... & ... & ... & 0 \\ 0 & 1 & -2 & 1 & 0 & ... & ... & 0 \\ & & . & . & & . & & \\ & & . & . & & . & & \\ 0 & ... & ... & 0 & 1 & -2 & 1 & 0 \\ 0 & ... & ... & ... & 0 & 1 & -2 & 1 \end{pmatrix} \quad \text{(A.2)}$$

En tenant compte de la contrainte de bords à 0, la matrice $\boldsymbol{D}_2$ passe à une dimension de $(D-1)\times(D-1)$.

$$\boldsymbol{D}_2 = \begin{pmatrix} -2 & 1 & 0 & ... & ... & ... \\ 1 & -2 & 1 & 0 & ... & ... \\ & & . & & . & \\ & & . & & . & \\ 0 & ... & 0 & 1 & -2 & 1 \\ 0 & ... & ... & 0 & 1 & -2 \end{pmatrix} \tag{A.3}$$

Finalement, on obtient la matrice de covariance *a priori* de la FRH.

$$\left\| \frac{\mathrm{d}^2 \boldsymbol{h}_j^m}{\mathrm{d}t^2} \right\|^2 = (\boldsymbol{h}_j^m)^{\mathrm{t}} \left(\frac{\boldsymbol{D}_2^{\mathrm{t}} \boldsymbol{D}_2}{(\Delta t)^4} \right) \boldsymbol{h}_j^m = (\boldsymbol{h}_j^m)^{\mathrm{t}} \boldsymbol{R}^{-1} \boldsymbol{h}_j^m \text{ avec } \boldsymbol{R} = \left(\frac{\boldsymbol{D}_2^{\mathrm{t}} \boldsymbol{D}_2}{(\Delta t)^4} \right)^{-1} \tag{A.4}$$

$$\text{Soit :} \qquad \boldsymbol{R}^{-1} = \frac{1}{(\Delta t)^4} \begin{pmatrix} 5 & -4 & 1 & 0 & ... & ... & ... & ... & 0 \\ -4 & 6 & -4 & 1 & 0 & ... & ... & ... & 0 \\ 1 & -4 & 6 & -4 & 1 & 0 & ... & ... & 0 \\ 0 & 1 & -4 & 6 & -4 & 1 & 0 & ... & 0 \\ & . & & . & & . & & . & \\ & . & & . & & . & & . & \\ 0 & ... & 0 & 1 & -4 & 6 & -4 & 1 & 0 \\ 0 & ... & ... & 0 & 1 & -4 & 6 & -4 & 1 \\ 0 & ... & ... & ... & 0 & 1 & -4 & 6 & -4 \\ 0 & ... & ... & ... & ... & 0 & 1 & -4 & 5 \end{pmatrix} \tag{A.5}$$

Annexe B

Matrice de dessin de l'approche DEC

La matrice de dessin $\boldsymbol{X}$ est de dimension $(M \times N) \times (D+1)$, où M représente le nombre de conditions expérimentales, N le nombre de scans et $(D+1)$ le nombre de coefficients de la FRH (la dimension du vecteur $\boldsymbol{h}$). Chaque matrice $\boldsymbol{X}^m$ de dimension $N \times (D+1)$ est une matrice binaire codant, pour la $m$ième condition, les temps d'arrivée echantillonnés sur une grille de résolution Δt. La figure B.1(a) montre un exemple de paradigme événementiel (stimuli de durées nulles) composé de trois conditions ayant chacune 4 occurences, ainsi que sa projection sur une grille d'échantillonnage de pas Δt (figure B.1(b)) et qui correspond à une interpolation d'ordre 0. Dans cet exemple, le temps de répétition $TR = 2,4\,\text{sec}$ et le pas d'échantillonnage $\Delta t = 0,6\,\text{sec}$, le premier est fixé par l'expérience et le deuxième est choisi arbitrairement avec la contrainte d'être un diviseur de TR pour minimiser l'approximation. Afin de bien comprendre les étapes de construction de la matrice de dessin, nous allons nous focaliser sur l'une des trois conditions (ici la première) (figure B.2). Après la projection sur la grille d'échantillonnage, une séquence binaire codant les temps d'arrivée de cette condition est construite pour former la première colonne de la matrice. Les $D+1$ colonnes restantes sont obtenues grâce à un décalage successif de la première colonne, ainsi une simple multiplication $\boldsymbol{X}^m\boldsymbol{h}$ représente le produit de convolution entre le paradigme et la FRH. Enfin, un sous échantillonnage en ligne d'un pas égal au TR permet d'obtenir la matrice finale. De la même façon, on obtient les matrices de dessin des deux autres stimuli (figure B.3). Dans l'exemple illustratif précédent, le paradigme n'est pas optimal, ce qui fait que la matrice $\boldsymbol{X}$ ne

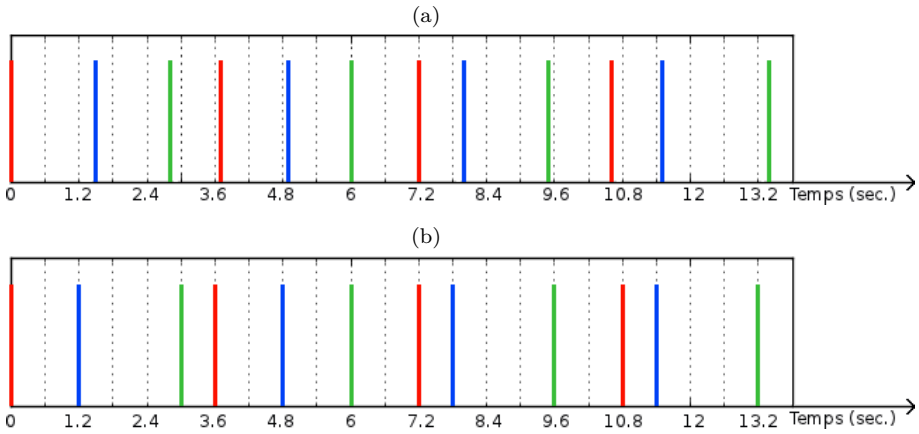


FIGURE B.1 – (a) Paradigme expérimental événementiel contenant 3 conditions ayant chacune 4 occurences. Chaque condition est codée sur une couleur (rouge, bleu, vert). (b) Projection du paradigme sur une grille d'échantillonnage de pas $\Delta t = 0,6$

contient pas de points permettant l'échantillonnage de la queue de la FRH par exemple. La figure B.4 montre un exemple de matrice que l'on obtient avec un paradigme plus approprié d'une durée de presque 300 sec (au lieu de $13,5$ sec) et dont chaque condition a 20 occurrences (au lieu de 4). L'ordre des dimensions ayant été changé pour une meilleure visualisation de la matrice. Dans l'exemple précédent du paradigme événementiel, une valeur binaire, "1", correspond à chaque stimulus. Pour les protocoles en blocs, la durée du bloc est codée à travers une suite de "1".

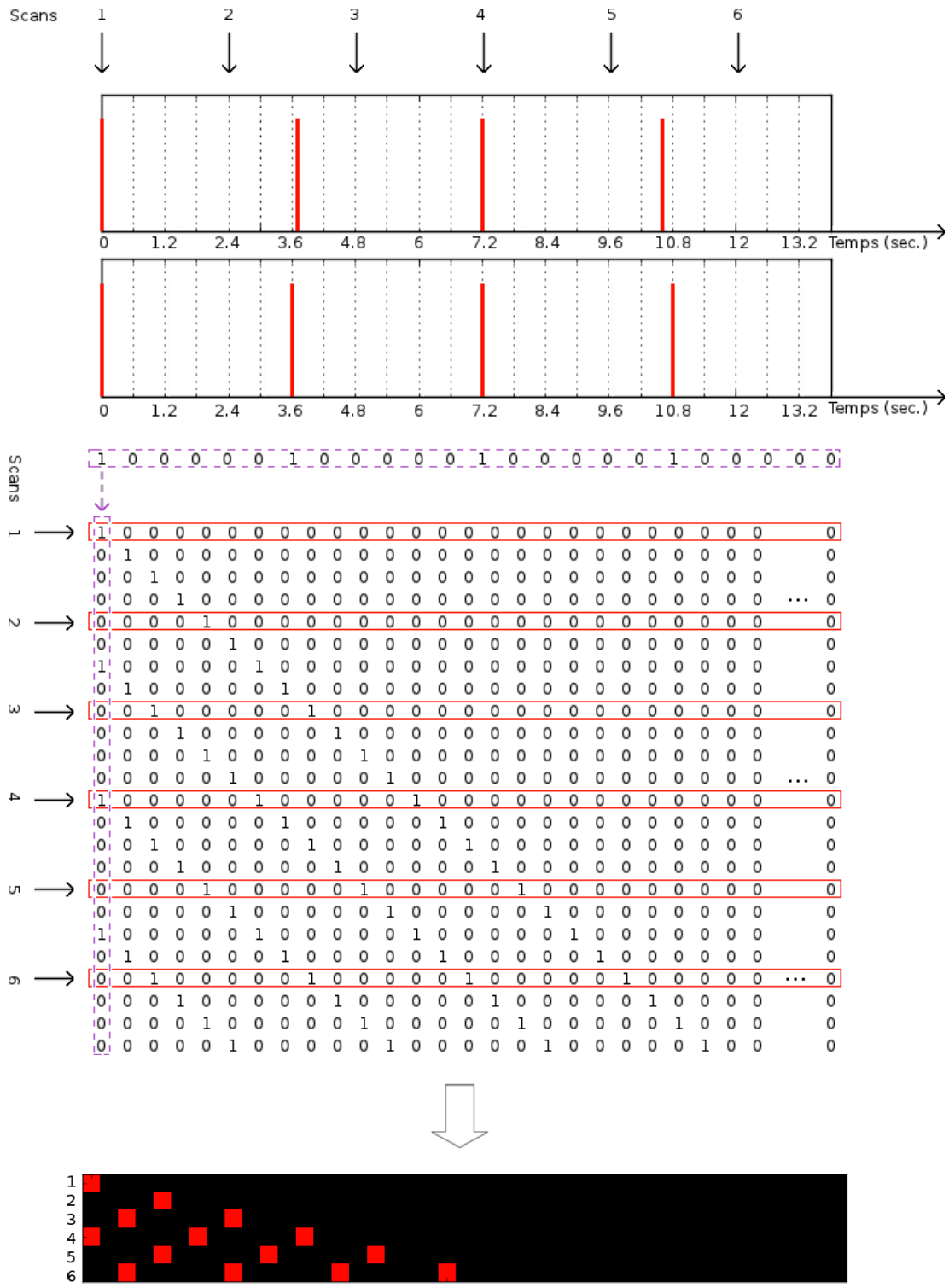


FIGURE B.2 – Etapes de construction de la matrice de dessin correspondante à la première condition expérimentale à partir du paradigme expérimental.

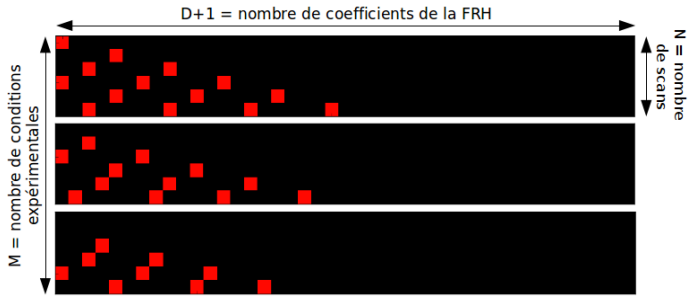


FIGURE B.3 – Matrice de dessin correspondante aux trois conditions expérimentales de 4 occurences chacunes et d'une durée totale de $13,5$ sec.

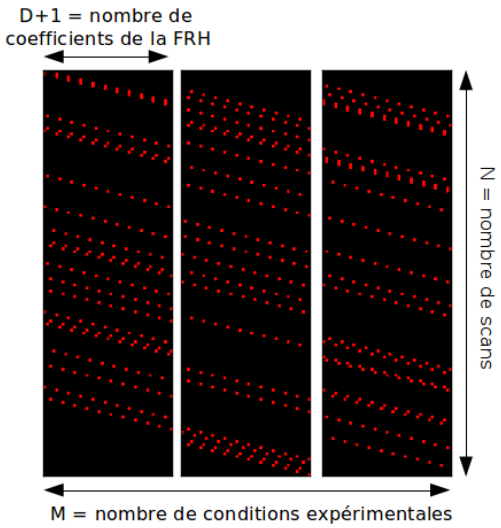


FIGURE B.4 – Matrice de dessin correspondante à trois conditions expérimentales de 20 occurences chacune et d'une durée totale de 300 sec. L'ordre des dimensions ayant changé par rapport à la figure B.3 pour des raisons de visualisation.

Annexe C

Simulations de Monte Carlo par Chaînes de Markov

Comme nous l'avons préalablement expliqué, l'inférence MCMC de l'approche DEC est basée sur un échantillonneur de Gibbs nécessitant le calcul des lois des variables du modèle conditionnellement aux données ainsi qu'aux autres variables. Les formules obtenues ont été citées dans la section III.2.2.1 du chapitre III. Leur calcul détaillé est présenté ci-dessous.
Dans ce qui suit les termes C_x représentent des termes constants par rapport à la variable x (parfois $C_{1,x}, C_{2,x}, etc.$ quand plusieurs constantes différentes interviennent dans une même formule). Ces termes ne seront détaillés que si on en a besoin pour le calcul. Par la suite les matrices et vecteurs sont notés en gras avec les matrices en majuscule et les vecteurs en minuscule. Un vecteur est par convention un vecteur colonne. La transposé est notée ^{t}.

♣ Distribution conditionnelle de $\left(a_j^m, q_j^m\right)$

Rappelons qu'il sagit ici de la simulation du couple de variables $\left(a_j^m, q_j^m\right)$:

$$\begin{aligned}
&p\left(a_j^m, q_j^m \mid \boldsymbol{y}_j, \boldsymbol{a}_j^{\setminus m}, \boldsymbol{w}, \boldsymbol{q}_{\setminus j}^m, \boldsymbol{h}, \boldsymbol{\ell}_j, \sigma_j^2, \boldsymbol{\theta}_m, \xi_m\right) \\
&= p\Big(a_j^m \mid q_j^m, \boldsymbol{y}_j, \boldsymbol{a}_j^{\setminus m}, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{\ell}_j, \sigma_j^2, \boldsymbol{\theta}_m\Big) p\Big(q_j^m \mid \boldsymbol{y}_j, \boldsymbol{a}_j^{\setminus m}, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{\ell}_j, \sigma_j^2, \boldsymbol{q}_{\setminus j}^m, \boldsymbol{\theta}_m, \xi_m\Big) \\
&\propto p\Big(\boldsymbol{y}_j \mid \boldsymbol{a}_j, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{\ell}_j, \sigma_j^2\Big)\, p\Big(a_j^m \mid w^m, q_j^m, \theta_m\Big)\, p\Big(q_j^m \mid \boldsymbol{q}_{\setminus j}^m, \xi_m\Big)\, p(w^m \mid \boldsymbol{q}^m, \boldsymbol{\tau}) \qquad \text{(C.1)}
\end{aligned}$$

- $p\left(a_j^m \,|\, q_j^m, \boldsymbol{y}_j, \boldsymbol{a}_j^{\setminus m}, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{\ell}_j, \sigma_j^2, \boldsymbol{\theta}_m\right)$

$$\begin{aligned}
&\propto p\left(\boldsymbol{y}_j \,|\, \boldsymbol{a}_j, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{\ell}_j, \sigma_j^2\right)\; p\left(a_j^m \,|\, w^m, q_j^m, \theta_m\right) \\
&\propto \exp\left(-\frac{\widetilde{\boldsymbol{y}_j}^{\mathrm{t}}\boldsymbol{\Gamma}_j\widetilde{\boldsymbol{y}_j}}{2}\right)\left\{\frac{(1-q_j^m w^m)\exp\left[\frac{-(a_j^m-\mu_0^m)^2}{2v_0^m}\right]}{\sqrt{v_0^m}}+\frac{q_j^m w^m \exp\left[\frac{-(a_j^m-\mu_1^m)^2}{2v_1^m}\right]}{\sqrt{v_1^m}}\right\}
\end{aligned} \tag{C.2}$$

avec $\boldsymbol{a}_j^{\setminus m} = \{a_j^{m'}; m' = 1 : M \text{ et } m' \neq m\}$ et $\widetilde{\boldsymbol{y}_j} = \boldsymbol{y}_j - \boldsymbol{S}_j\boldsymbol{h} - \boldsymbol{P}\boldsymbol{\ell}_j = \boldsymbol{e}_j^m - a_j^m w^m \boldsymbol{X}^m\boldsymbol{h}$ où $\boldsymbol{e}_j^m = \boldsymbol{y}_j - \sum\limits_{m'\neq m} a_j^{m'} w^{m'}\boldsymbol{X}^{m'}\boldsymbol{h} - \boldsymbol{P}\boldsymbol{\ell}_j$ ce qui donne :

$$\widetilde{\boldsymbol{y}_j}^{\mathrm{t}}\boldsymbol{\Gamma}_j\widetilde{\boldsymbol{y}_j} = \boldsymbol{e}_j^{m\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{e}_j^m - 2a_j^m w^m \boldsymbol{e}_j^{m\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{X}^m\boldsymbol{h} + (a_j^m)^2 w^m \boldsymbol{h}^{\mathrm{t}}\boldsymbol{X}^{m\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{X}^m\boldsymbol{h} \tag{C.3}$$

En remplaçant $\widetilde{\boldsymbol{y}_j}^{\mathrm{t}}\boldsymbol{\Gamma}_j\widetilde{\boldsymbol{y}_j}$ dans l'équation (C.2) on obtient :

$$\begin{aligned}
&p\left(a_j^m \,|\, q_j^m, \boldsymbol{y}_j, \boldsymbol{a}_j^{\setminus m}, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{\ell}_j, \sigma_j^2, \boldsymbol{\theta}_m\right) \\
&\quad\propto (1-q_j^m w^m)\sqrt{\frac{v_{0,j}^m}{v_0^m}}\exp\left(-\frac{1}{2}\left[\frac{(\mu_0^m)^2}{v_0^m}-\frac{(\mu_{0,j}^m)^2}{v_{0,j}^m}\right]\right)\mathcal{N}(a_j^m;\mu_{0,j}^m, v_{0,j}^m) \\
&\qquad+\; q_j^m w^m\sqrt{\frac{v_{1,j}^m}{v_1^m}}\exp\left(-\frac{1}{2}\left[\frac{(\mu_1^m)^2}{v_1^m}-\frac{(\mu_{1,j}^m)^2}{v_{1,j}^m}\right]\right)\mathcal{N}(a_j^m;\mu_{1,j}^m, v_{1,j}^m)
\end{aligned} \tag{C.4}$$

d'où on peut déduire que :

$$\begin{aligned}
&p\left(a_j^m \,|\, q_j^m, \boldsymbol{y}_j, \boldsymbol{a}_j^{\setminus m}, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{\ell}_j, \sigma_j^2, \boldsymbol{\theta}_m\right) \\
&\quad= (1-q_j^m w^m)\,\mathcal{N}(a_j^m;\mu_{0,j}^m, v_{0,j}^m) + q_j^m w^m\,\mathcal{N}(a_j^m;\mu_{1,j}^m, v_{1,j}^m)
\end{aligned} \tag{C.5}$$

avec

$$v_{i,j}^m = \left(w^m\boldsymbol{h}^{\mathrm{t}}\boldsymbol{X}^{m\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{X}^m\boldsymbol{h} + \frac{1}{v_i^m}\right)^{-1} \tag{C.6}$$

$$\mu_{i,j}^m = v_{i,j}^m\left(w^m\boldsymbol{e}_j^{m\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{X}^m\boldsymbol{h} + \frac{\mu_i^m}{v_i^m}\right) \tag{C.7}$$

- $p\left(q_j^m \,|\, \boldsymbol{y}_j, \boldsymbol{a}_j^{\setminus m}, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{\ell}_j, \sigma_j^2, \boldsymbol{q}_{\setminus j}^m, \boldsymbol{\theta}_m, \xi_m\right)$

$$= \int\limits_{a_j^m} p\left(a_j^m, q_j^m \,|\, \boldsymbol{y}_j, \boldsymbol{a}_j^{\setminus m}, \boldsymbol{w}, \boldsymbol{q}_{\setminus j}^m, \boldsymbol{h}, \boldsymbol{\ell}_j, \sigma_j^2, \boldsymbol{\theta}_m, \xi_m\right) da_j^m \propto f(q_j^m) \tag{C.8}$$

avec

$$f(q_j^m) = p(q_j^m \mid \boldsymbol{q}_{\setminus j}^m, \xi_m) p(w^m \mid \boldsymbol{q}^m, \boldsymbol{\tau}) \Bigg\{ (1 - q_j^m w^m) \sqrt{\frac{v_{0,j}^m}{v_0^m}} \exp\left(-\frac{1}{2} \left[\frac{(\mu_0^m)^2}{v_0^m} - \frac{(\mu_{0,j}^m)^2}{v_{0,j}^m} \right] \right) + q_j^m w^m \sqrt{\frac{v_{1,j}^m}{v_1^m}} \exp\left(-\frac{1}{2} \left[\frac{(\mu_1^m)^2}{v_1^m} - \frac{(\mu_{1,j}^m)^2}{v_{1,j}^m} \right] \right) \Bigg\} \tag{C.9}$$

On peut maintenant calculer les expressions suivantes :

$$f(q_j^m = 0) = p(q_j^m = 0 \mid \boldsymbol{q}_{\setminus j}^m, \xi_m) p(w^m \mid q_j^m = 0, \boldsymbol{q}_{\setminus j}^m, \boldsymbol{\tau}) \times \sqrt{\frac{v_{0,j}^m}{v_0^m}} \exp\left(-\frac{1}{2} \left[\frac{(\mu_0^m)^2}{v_0^m} - \frac{(\mu_{0,j}^m)^2}{v_{0,j}^m} \right] \right) \tag{C.10}$$

$$f(q_j^m = 1) = p(q_j^m = 1 \mid \boldsymbol{q}_{\setminus j}^m, \xi_m) p(w^m \mid q_j^m = 1, \boldsymbol{q}_{\setminus j}^m, \boldsymbol{\tau}) \times \Bigg\{ (1 - w^m) \sqrt{\frac{v_{0,j}^m}{v_0^m}} \exp\left(-\frac{1}{2} \left[\frac{(\mu_0^m)^2}{v_0^m} - \frac{(\mu_{0,j}^m)^2}{v_{0,j}^m} \right] \right) + w^m \sqrt{\frac{v_{1,j}^m}{v_1^m}} \exp\left(-\frac{1}{2} \left[\frac{(\mu_1^m)^2}{v_1^m} - \frac{(\mu_{1,j}^m)^2}{v_{1,j}^m} \right] \right) \Bigg\} \tag{C.11}$$

Une fois les calculs effectués, les probabilités recherchées sont obtenues en normalisant les deux quantités précédentes (équations (C.10) et (C.11)) afin de garantir des valeurs dans $]0, 1[$:

$$p\left(q_j^m = 0 \mid \boldsymbol{y}_j, \boldsymbol{a}_j^{\setminus m}, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{\ell}_j, \sigma_j^2, \boldsymbol{q}_{\setminus j}^m, \boldsymbol{\theta}_m, \xi_m\right) = \frac{f(q_j^m = 0)}{f(q_j^m = 0) + f(q_j^m = 1)} \tag{C.12}$$

$$p\left(q_j^m = 1 \mid \boldsymbol{y}_j, \boldsymbol{a}_j^{\setminus m}, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{\ell}_j, \sigma_j^2, \boldsymbol{q}_{\setminus j}^m, \boldsymbol{\theta}_m, \xi_m\right) = \frac{f(q_j^m = 1)}{f(q_j^m = 0) + f(q_j^m = 1)} \tag{C.13}$$

Ce qui donne :

$$p\left(q_j^m = 1 \mid \boldsymbol{y}_j, \boldsymbol{a}_j^{\setminus m}, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{\ell}_j, \sigma_j^2, \boldsymbol{q}_{\setminus j}^m, \boldsymbol{\theta}_m, \xi_m\right) = \left[1 + \frac{\frac{p(q_j^m = 0 \mid \boldsymbol{q}_{\setminus j}^m, \xi_m)\, p(w^m \mid q_j^m = 0, \boldsymbol{q}_{\setminus j}^m, \boldsymbol{\tau})}{p(q_j^m = 1 \mid \boldsymbol{q}_{\setminus j}^m, \xi_m)\, p(w^m \mid q_j^m = 1, \boldsymbol{q}_{\setminus j}^m, \boldsymbol{\tau})}}{1 - w^m \left\{ 1 - \sqrt{\frac{v_{1,j}^m v_0^m}{v_{0,j}^m v_1^m}} \exp\left(-\frac{1}{2} \left[\frac{(\mu_1^m)^2}{v_1^m} - \frac{(\mu_0^m)^2}{v_0^m} - \frac{(\mu_{1,j}^m)^2}{v_{1,j}^m} + \frac{(\mu_{0,j}^m)^2}{v_{0,j}^m} \right] \right) \right\}} \right]^{-1} \tag{C.14}$$

et

$$p\left(q_j^m = 0 \,|\, \boldsymbol{y}_j, \boldsymbol{a}_j^{\setminus m}, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{\ell}_j, \sigma_j^2, \boldsymbol{q}_{\setminus j}^m, \boldsymbol{\theta}_m, \xi_m\right)$$
$$= 1 - p\left(q_j^m = 1 \,|\, \boldsymbol{y}_j, \boldsymbol{a}_j^{\setminus m}, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{\ell}_j, \sigma_j^2, \boldsymbol{q}_{\setminus j}^m, \boldsymbol{\theta}_m, \xi_m\right) \tag{C.15}$$

avec $\boldsymbol{q}_{\setminus j}^m = \{q_{j'}^m; j' = 1 : J \text{ et } j' \neq j\}$.

- Calcul de $p(q_j^m = 0 \,|\, q_{V_j}^m, \xi_m)/p(q_j^m = 1 \,|\, q_{V_j}^m, \xi_m)$:

Comme précisé dans la section II.4.1 l'*a priori* sur les labels $\boldsymbol{Q}$ est un champ de Ising ce qui nous permet d'écrire :

$$p\left(q_j^m \,|\, \boldsymbol{q}_{\setminus j}^m, \xi_m\right) = p\left(q_j^m \,|\, \boldsymbol{q}_{N_j}^m, \xi_m\right) = \frac{\exp\left[\xi_m U(q_j^m)\right]}{Z(\xi_m)} \tag{C.16}$$

avec $U(q_j^m) = \sum_{k \in N_j} \mathbb{1}(q_j^m = q_k^m)$ et $Z(\xi_m) = \sum_{q_j^m \in \{0,1\}} \exp\left(\xi_m U(q_j^m)\right)$, on a donc :

$$\frac{p(q_j^m = 0 \,|\, \boldsymbol{q}_{N_j}^m, \xi_m)}{p(q_j^m = 1 \,|\, \boldsymbol{q}_{N_j}^m, \xi_m)} = \exp\left\{\xi_m \left(\sum_{k \in N_j} \mathbb{1}(q_k^m = 0) - \sum_{k \in N_j} \mathbb{1}(q_k^m = 1)\right)\right\}$$
$$= \exp\left[\xi_m \left(\Omega_0^j - \Omega_1^j\right)\right] = \exp\left[\xi_m \left(\Omega^j - 2\Omega_1^j\right)\right] \tag{C.17}$$

avec $\Omega^j = |N_j|$, $\Omega_i^j = \text{Card}(\{q_k^m = i; k \in N_j\})$ et $\boldsymbol{q}_{N_j}^m = \{q_k^m; k \in N_j\}$ où N_j définit les voxels k voisins de j.

♣ Distribution conditionnelle de w^m

Cette distribution est définie par l'équation suivante :

$$p\left(w^m \,|\, \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{w}^{\setminus m}, \boldsymbol{q}^m, \boldsymbol{h}, \boldsymbol{\sigma}^2, \boldsymbol{\theta}_m, \boldsymbol{\tau}\right)$$
$$\propto p\left(w^m \,|\, \boldsymbol{q}^m, \boldsymbol{\tau}\right) \prod_j \left[p(\boldsymbol{y}_j \,|\, \boldsymbol{a}_j, \boldsymbol{h}, \boldsymbol{w}, \sigma_j^2) p\left(a_j^m \,|\, w^m, q_j^m, \theta_m\right)\right]$$
$$\propto f(w^m) \tag{C.18}$$

en notant

$$f(w^m) = p\left(w^m \,|\, \boldsymbol{q}^m, \boldsymbol{\tau}\right) \prod_j \exp\left(-\frac{\widetilde{\boldsymbol{y}}_j^{\mathrm{t}} \boldsymbol{\Gamma}_j \widetilde{\boldsymbol{y}}_j}{2}\right)$$
$$\times \left\{ (1 - q_j^m w^m) \frac{\exp\left[\frac{-(a_j^m - \mu_0^m)^2}{2v_0^m}\right]}{(v_0^m)^{-\frac{1}{2}}} + q_j^m w^m \frac{\exp\left[\frac{-(a_j^m - \mu_1^m)^2}{2v_1^m}\right]}{(v_1^m)^{-\frac{1}{2}}} \right\} \tag{C.19}$$

où $\boldsymbol{w}^{\backslash m} = \{w^{m'}; m' = 1 : M \text{ et } m' \neq m\}$ et

$$\begin{aligned} \widetilde{\boldsymbol{y}}_j^{\mathrm{t}} \boldsymbol{\Gamma}_j \widetilde{\boldsymbol{y}}_j &= \boldsymbol{e}_j^{m\mathrm{t}} \boldsymbol{\Gamma}_j \boldsymbol{e}_j^m - 2a_j^m w^m \boldsymbol{e}_j^{m\mathrm{t}} \boldsymbol{\Gamma}_j \boldsymbol{X}^m \boldsymbol{h} + (a_j^m)^2 (w^m)^2 \boldsymbol{h}^{\mathrm{t}} \boldsymbol{X}^{m\mathrm{t}} \boldsymbol{\Gamma}_j \boldsymbol{X}^m \boldsymbol{h} \\ &= \boldsymbol{C}_{w^m} - w^m \left[2a_j^m \boldsymbol{e}_j^{m\mathrm{t}} \boldsymbol{\Gamma}_j \boldsymbol{X}^m \boldsymbol{h} - (a_j^m)^2 \boldsymbol{h}^{\mathrm{t}} \boldsymbol{X}^{m\mathrm{t}} \boldsymbol{\Gamma}_j \boldsymbol{X}^m \boldsymbol{h}\right] \end{aligned} \tag{C.20}$$

ce qui donne

$$f(w^m = 0) = p\left(w^m = 0 \,|\, \boldsymbol{q}^m, \boldsymbol{\tau}\right) \frac{\exp\left[-\sum_j \frac{\left(a_j^m - \mu_0^m\right)^2}{2v_0^m}\right]}{(v_0^m)^{J/2}} \tag{C.21}$$

$$\begin{aligned} f(w^m = 1) &= p(w^m = 1 \,|\, \boldsymbol{q}^m, \boldsymbol{\tau}) \left(v_1^m\right)^{-J_1^m/2} \left(v_0^m\right)^{-J_0^m/2} \\ &\times \exp\left[\frac{-1}{2} \sum_j \left(\frac{(a_j^m - \mu_{q_j^m}^m)^2}{v_{q_j^m}^m} + (a_j^m)^2 \boldsymbol{h}^{\mathrm{t}} \boldsymbol{X}^{m\mathrm{t}} \boldsymbol{\Gamma}_j \boldsymbol{X}^m \boldsymbol{h} - 2a_j^m \boldsymbol{e}_j^{m\mathrm{t}} \boldsymbol{\Gamma}_j \boldsymbol{X}^m \boldsymbol{h}\right)\right] \end{aligned} \tag{C.22}$$

Reste à normaliser les termes précédents comme dans le cas des labels. En écrivant les termes $\prod_{j \in J}(.)$ sous la forme $\prod_{j \in J_0^m}(.) \prod_{j \in J_1^m}(.)$ où $J = J_0^m + J_1^m$ pour $m = 1$ à M, on obtient :

$$\begin{aligned} &p(w^m = 1 \,|\, \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{w}^{\backslash m}, \boldsymbol{q}^m, \boldsymbol{h}, \boldsymbol{\sigma}^2, \boldsymbol{\theta}_m, \boldsymbol{\tau}) \\ &= \left(1 + \frac{p(w^m = 0 \,|\, \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{w}^{\backslash m}, \boldsymbol{q}^m, \boldsymbol{h}, \boldsymbol{\theta}, \boldsymbol{\tau})}{p(w^m = 1 \,|\, \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{w}^{\backslash m}, \boldsymbol{q}^m, \boldsymbol{h}, \boldsymbol{\theta}, \boldsymbol{\tau})}\right)^{-1} \\ &= \left(1 + \frac{\frac{p(w^m=0 \,|\, \boldsymbol{q}^m, \boldsymbol{\tau})}{p(w^m=1 \,|\, \boldsymbol{q}^m, \boldsymbol{\tau})} \left(\frac{v_0^m}{v_1^m}\right)^{-\frac{J_1^m}{2}} \exp\left(-\sum_{j \in J_1^m} \left[\frac{\left(a_j^m - \mu_0^m\right)^2}{2v_0^m} - \frac{\left(a_j^m - \mu_1^m\right)^2}{2v_1^m}\right]\right)}{\exp\left[-\sum_{j \in J} \frac{1}{2}\left(-2\, a_j^m\, \boldsymbol{e}_j^{m\mathrm{t}} \boldsymbol{\Gamma}_j \boldsymbol{X}^m \boldsymbol{h} + (a_j^m)^2 \boldsymbol{h}^{\mathrm{t}} \boldsymbol{X}^{m\mathrm{t}} \boldsymbol{\Gamma}_j \boldsymbol{X}^m \boldsymbol{h}\right)\right]}\right)^{-1} \end{aligned} \tag{C.23}$$

avec $\frac{p(w^m=0 \,|\, \boldsymbol{q}^m, \boldsymbol{\tau})}{p(w^m=1 \,|\, \boldsymbol{q}^m, \boldsymbol{\tau})} = \frac{1-p_m}{p_m} = \exp\left[-\tau_1\left(\sum_j q_j^m - \tau_2\right)\right]$, p_m étant défini par l'équation (III.10).

♣ Distribution conditionnelle de $\boldsymbol{h}$

Rappelons que l'*a priori* sur la FRH est une gaussienne $\boldsymbol{h} \sim \mathcal{N}(0, v_{\boldsymbol{h}}\boldsymbol{R})$ (équation (II.24)) et que la matrice de covariance $\boldsymbol{R} = \left(\frac{\boldsymbol{D}_2^{\mathrm{t}}\boldsymbol{D}_2}{(\Delta t)^4}\right)^{-1}$ (annexe A) exprime une contrainte sur la dérivée seconde afin de pénaliser les fortes variations de pentes et obtenir des FRH lisses plus plausibles physiologiquement.

$$\begin{aligned} p\left(\boldsymbol{h} \,|\, \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{w}, \boldsymbol{L}, \boldsymbol{\sigma}^2, v_h\right) &\propto p\left(\boldsymbol{h} \,|\, v_h\right) p\left(\boldsymbol{Y} \,|\, \boldsymbol{A}, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{L}, \boldsymbol{\sigma}^2\right) \\ &\propto v_h^{-\frac{D-1}{2}} \exp\left(-\frac{\boldsymbol{h}^{\mathrm{t}}\boldsymbol{R}^{-1}\boldsymbol{h}}{2v_h}\right)\prod_j \exp\left(-\frac{\widetilde{\boldsymbol{y}}_j^{\mathrm{t}}\boldsymbol{\Gamma}_j\widetilde{\boldsymbol{y}}_j}{2}\right) \\ &\propto v_h^{-\frac{D-1}{2}} \exp\left(-\frac{1}{2}\left[\boldsymbol{h}^{\mathrm{t}}\frac{\boldsymbol{R}^{-1}}{v_h}\boldsymbol{h} + \sum_j \widetilde{\boldsymbol{y}}_j^{\mathrm{t}}\boldsymbol{\Gamma}_j\widetilde{\boldsymbol{y}}_j\right]\right) \end{aligned} \tag{C.24}$$

où on écrit cette fois, pour faire ressortir les termes en $\boldsymbol{h}$:

$$\begin{aligned} \widetilde{\boldsymbol{y}}_j^{\mathrm{t}}\boldsymbol{\Gamma}_j\widetilde{\boldsymbol{y}}_j &= [(\boldsymbol{y}_j - \boldsymbol{P}\boldsymbol{\ell}_j) - \boldsymbol{S}_j\boldsymbol{h}]^{\mathrm{t}}\boldsymbol{\Gamma}_j[(\boldsymbol{y}_j - \boldsymbol{P}\boldsymbol{\ell}_j) - \boldsymbol{S}_j\boldsymbol{h}] \\ &= (\boldsymbol{y}_j - \boldsymbol{P}\boldsymbol{\ell}_j)^{\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j - \boldsymbol{P}\boldsymbol{\ell}_j) - 2\boldsymbol{h}^{\mathrm{t}}\boldsymbol{S}_j^{\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j - \boldsymbol{P}\boldsymbol{\ell}_j) + \boldsymbol{h}^{\mathrm{t}}\boldsymbol{S}_j^{\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{S}_j\boldsymbol{h} \\ &= \boldsymbol{h}^{\mathrm{t}}\boldsymbol{S}_j^{\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{S}_j\boldsymbol{h} - 2\boldsymbol{h}^{\mathrm{t}}\boldsymbol{S}_j^{\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j - \boldsymbol{P}\boldsymbol{\ell}_j) + \boldsymbol{C}_{\boldsymbol{h}} \end{aligned} \tag{C.25}$$

En remplaçant dans l'équation (C.24) on trouve :

$$\begin{aligned} p\left(\boldsymbol{h} \,|\, \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{w}, \boldsymbol{L}, \boldsymbol{\sigma}^2, v_h\right) \propto v_h^{-(D-1)/2} \exp\Bigg\{-\frac{1}{2}\Bigg[&\boldsymbol{h}^{\mathrm{t}}\left(\frac{\boldsymbol{R}^{-1}}{v_h} + \sum_j \boldsymbol{S}_j^{\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{S}_j\right)\boldsymbol{h} \\ &-2\boldsymbol{h}^{\mathrm{t}}\sum_j \boldsymbol{S}_j^{\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j - \boldsymbol{P}\boldsymbol{\ell}_j)\Bigg]\Bigg\}, \end{aligned} \tag{C.26}$$

d'où on déduit que

$$p\left(\boldsymbol{h} \,|\, \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{w}, \boldsymbol{L}, \boldsymbol{\sigma}^2, v_h\right) = \mathcal{N}\left(\boldsymbol{h}; \boldsymbol{\mu}_{\boldsymbol{h}}, \boldsymbol{\Sigma}_{\boldsymbol{h}}\right) \tag{C.27}$$

avec

$$\boldsymbol{\Sigma}_{\boldsymbol{h}} = \left(\frac{\boldsymbol{R}^{-1}}{v_h} + \sum_j \boldsymbol{S}_j^{\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{S}_j\right)^{-1} \tag{C.28}$$

$$\boldsymbol{\mu}_{\boldsymbol{h}} = \boldsymbol{\Sigma}_{\boldsymbol{h}}\sum_j \boldsymbol{S}_j^{\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j - \boldsymbol{P}\boldsymbol{\ell}_j) \tag{C.29}$$

♣ Distribution conditionnelle de v_h

En supposant un *a priori* non informatif de Jeffrey sur v_h tel que $p(v_h) = \frac{1}{v_h}$ (annexe F.3) on trouve :

$$\begin{aligned} p\left(v_h \,|\, \boldsymbol{h}\right) &\propto p\left(\boldsymbol{h} \,|\, v_h\right) p(v_h) \\ &\propto v_h^{-\frac{D+1}{2}} \exp\left(\frac{-\boldsymbol{h}^{\mathrm{t}}\boldsymbol{R}^{-1}\boldsymbol{h}}{2v_h}\right) \end{aligned} \tag{C.30}$$

c'est à dire :

$$p\left(v_h \,|\, \boldsymbol{h}\right) = \mathcal{IG}(v_h; \alpha_{v_h}, \lambda_{v_h}) \tag{C.31}$$

où $\alpha_{v_h} = \frac{D-1}{2}$ et $\lambda_{v_h} = \frac{1}{2}\, \boldsymbol{h}^{\mathrm{t}}\boldsymbol{R}^{-1}\boldsymbol{h}$

♣ Distribution conditionnelle de $\boldsymbol{\sigma}^2$

Avec un *a priori* non informatif de Jeffrey (annexe F.3) de la forme :

$$p(\boldsymbol{\sigma}^2) = \prod_j p(\sigma_j^2) = \prod_j \frac{1}{\sigma_j^2} \tag{C.32}$$

et tenant compte de la définition de la vraisemblance (équation (II.23)), on peut écrire :

$$\begin{aligned} p\left(\boldsymbol{\sigma}^2 \,|\, \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{L}\right) &\propto \prod_j p\left(\sigma_j^2 \,|\, \boldsymbol{y}_j, \boldsymbol{a}_j, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{\ell}_j\right) \\ &\propto \prod_j p\left(\boldsymbol{y}_j \,|\, \boldsymbol{a}_j, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{\ell}_j, \sigma_j^2\right) p(\sigma_j^2) \\ &\propto \prod_j \left(\sigma_j^2\right)^{-\frac{N+2}{2}} \exp\left(-\frac{1}{2\sigma_j^2}\, \widetilde{\boldsymbol{y}}_j^{\mathrm{t}}\boldsymbol{\Lambda}_j\widetilde{\boldsymbol{y}}_j\right) \end{aligned} \tag{C.33}$$

d'où

$$p\left(\boldsymbol{\sigma}^2 \,|\, \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{L}\right) = \prod_j \mathcal{IG}(\sigma_j^2; \alpha_{\sigma_j^2}, \lambda_{\sigma_j^2}) \tag{C.34}$$

où $\boldsymbol{\Gamma}_j = \frac{\boldsymbol{\Lambda}_j}{\sigma_j^2}$, $\alpha_{\sigma_j^2} = \frac{N}{2}$ et $\lambda_{\sigma_j^2} = \frac{\widetilde{\boldsymbol{y}}_j^{\mathrm{t}}\boldsymbol{\Lambda}_j\widetilde{\boldsymbol{y}}_j}{2}$

♣ Distribution conditionnelle de $\boldsymbol{\theta}_m$

En gardant l'indépendance inter-conditions (types de stimuli), on introduit un *a priori* gaussien sur la moyenne de la classe active μ_1^m et des *a priori* inverse-gamma sur les variances v_i^m des classes active ($i = 1$) et inactive ($i = 0$) :

$$p(\boldsymbol{\mu}_1) = \prod_m p(\mu_1^m) = \prod_m \mathcal{N}\left(\mu_1^m; m_{\mu_1^m}, v_{\mu_1^m}\right) \tag{C.35}$$

$$p(\boldsymbol{v}_i) = \prod_m p(v_i^m) = \prod_m \mathcal{IG}\left(v_i^m; \alpha_{v_i^m}, \lambda_{v_i^m}\right) \tag{C.36}$$

- Moyenne de la classe active μ_1^m :

$$p\left(\mu_1^m \mid \boldsymbol{a}^m, w^m, \boldsymbol{q}^m, v_1^m\right) \propto p\left(\mu_1^m\right) \prod_j p(a_j^m \mid w^m, q_j^m, \theta_m) \tag{C.37}$$

on a vu que

$$\prod_j p(a_j^m \mid w^m, q_j^m, \theta_m) = \prod_j \left[(1 - q_j^m w^m)\mathcal{N}(a_j^m; \mu_0^m, v_0^m) + q_j^m w^m \mathcal{N}(a_j^m; \mu_1^m, v_1^m)\right]$$

qu'on peut également écrire d'une autre manière $(1 - w^m) \prod_{j \in J} \mathcal{N}(a_j^m; \mu_0^m, v_0^m) + w^m \prod_{j \in J_0^m} \mathcal{N}(a_j^m; \mu_0^m, v_0^m) \prod_{j \in J_1^m} \mathcal{N}(a_j^m; \mu_1^m, v_1^m)$ et qui est plus pratique pour le calcul de l'*a posteriori* conditionnel de μ_1^m, où J_1^m (resp. J_0^m) désigne l'ensemble des voxels actifs (resp. inactifs) pour la condition m dans la parcelle γ. On obtient donc :

$$\begin{aligned} p\left(\mu_1^m \mid \boldsymbol{a}^m, w^m, \boldsymbol{q}^m, v_1^m\right) \propto\ & (1 - w^m)\mathcal{N}\left(\mu_1^m; m_{\mu_1^m}, v_{\mu_1^m}\right) C_{1,\mu_1^m} \\ & + w^m \mathcal{N}\left(\mu_1^m; m_{\mu_1^m}, v_{\mu_1^m}\right) C_{2,\mu_1^m} \prod_{j \in J_1} \mathcal{N}(a_j^m; \mu_1^m, v_1^m). \end{aligned}$$

Soit

$$\begin{aligned} & p\left(\mu_1^m \mid \boldsymbol{a}^m, w^m, \boldsymbol{q}^m, v_1^m\right) \\ & \quad = (1 - w^m)\mathcal{N}\left(\mu_1^m; m_{\mu_1^m}, v_{\mu_1^m}\right) + w^m \mathcal{N}\left(\mu_1^m; M_{\mu_1^m}, V_{\mu_1^m}\right) \end{aligned} \tag{C.38}$$

avec

$$V_{\mu_1^m} = \left(\frac{1}{v_{\mu_1^m}} + \frac{J_1^m}{v_1^m}\right)^{-1} \quad , \quad M_{\mu_1^m} = V_{\mu_1^m} \left(\frac{m_{\mu_1^m}}{v_{\mu_1^m}} + \frac{\sum_{j \in J_1^m} a_j^m}{v_1^m}\right) \tag{C.39}$$

- Variance de la classe inactive v_0^m :

$$
\begin{aligned}
&p\left(v_0^m \mid \boldsymbol{a}^m, w^m, \boldsymbol{q}^m, \mu_0^m\right) \\
&\propto p\left(v_0^m\right) \prod_j p(a_j^m \mid w^m, q_j^m, \theta_m) \\
&\propto p\left(v_0^m\right) \left\{ (1-w^m) \prod_{j\in J} \mathcal{N}(a_j^m; \mu_0^m, v_0^m) + w^m C_{v_0^m} \prod_{j\in J_0^m} \mathcal{N}(a_j^m; \mu_0^m, v_0^m) \right\} \\
&\propto \frac{\exp\left(\frac{-\beta_{v_0^m}}{v_0^m}\right)}{(v_0^m)^{\alpha_{v_0^m}+1}} \left\{ (1-w^m) \frac{\exp\left[\frac{-\sum_{j\in J}(a_j^m-\mu_0^m)^2}{2v_0^m}\right]}{(v_0^m)^{\frac{J}{2}}} + w^m C_{v_0^m} \frac{\exp\left[\frac{-\sum_{j\in J_0}(a_j^m-\mu_0^m)^2}{2v_0^m}\right]}{(v_0^m)^{\frac{J_0}{2}}} \right\}.
\end{aligned}
$$

Soit

$$
\begin{aligned}
&p\left(v_0^m \mid \boldsymbol{a}^m, w^m, \boldsymbol{q}^m, \mu_0^m\right) \\
&\qquad = (1-w^m)\, \mathcal{IG}(v_0^m; A_{0,v_0^m}, B_{0,v_0^m}) + w^m\, \mathcal{IG}(v_0^m; A_{1,v_0^m}, B_{1,v_0^m})
\end{aligned} \tag{C.40}
$$

avec

$$
A_{0,v_0^m} = \alpha_{v_0^m} + \frac{J}{2} \quad , \quad B_{0,v_0^m} = \beta_{v_0^m} + \sum_{j\in J} \frac{\left(a_j^m - \mu_0^m\right)^2}{2} \tag{C.41}
$$

$$
A_{1,v_0^m} = \alpha_{v_0^m} + \frac{J_0^m}{2} \quad , \quad B_{1,v_0^m} = \beta_{v_0^m} + \sum_{j\in J_0^m} \frac{\left(a_j^m - \mu_0^m\right)^2}{2} \tag{C.42}
$$

où J_0^m désigne l'ensemble des voxels de la classe inactive ($i = 0$) correspondant à la $m$ième condition dans la parcelle $\mathcal{P}_\gamma$.

- Variance de la classe active v_1^m :

$$
\begin{aligned}
&p\left(v_1^m \mid \boldsymbol{a}^m, w^m, \boldsymbol{q}^m, \mu_1^m\right) \\
&\propto p\left(v_1^m\right) \prod_j p(a_j^m \mid w^m, q_j^m, \theta_m) \\
&\propto p\left(v_1^m\right) \left\{ (1-w^m) C_{v_1^m} + w^m C_{v_1^m} \prod_{j\in J_1^m} \mathcal{N}(a_j^m; \mu_1^m, v_1^m) \right\} \\
&\propto \frac{\exp\left(\frac{-\beta_{v_1^m}}{v_1^m}\right)}{(v_1^m)^{\alpha_{v_1^m}+1}} \left\{ (1-w^m) C_{v_1^m} + w^m C_{v_1^m} \frac{\exp\left[\frac{-\sum_{j\in J_1}(a_j^m-\mu_1^m)^2}{2v_1^m}\right]}{(v_1^m)^{\frac{J_1}{2}}} \right\}.
\end{aligned}
$$

Soit

$$p\left(v_1^m \,|\, \boldsymbol{a}^m, w^m, \boldsymbol{q}^m, \mu_1^m\right) = (1-w^m)\,\mathcal{IG}(v_1^m; \alpha_{v_1^m}, \beta_{v_1^m}) + w^m\,\mathcal{IG}(v_1^m; A_{1,v_1^m}, B_{1,v_1^m}) \tag{C.43}$$

avec

$$A_{1,v_1^m} = \alpha_{v_1^m} + \frac{J_1^m}{2} \quad , \quad B_{1,v_1^m} = \beta_{v_1^m} + \sum_{j\in J_1^m} \frac{\left(a_j^m - \mu_1^m\right)^2}{2} \tag{C.44}$$

où J_1^m désigne l'ensemble des voxels de la classe active ($i = 1$) correspondant à la $m$ième condition dans la parcelle $\mathcal{P}_\gamma$.

♣ Distribution conditionnelle de $\boldsymbol{\ell}_j$

Rappelons l'*a priori* gaussien avec indépendance spatiale sur les dérives basses fréquences $\boldsymbol{L}$:

$$p\left(\boldsymbol{L} \,|\, v_\ell\right) = \prod_j p\left(\boldsymbol{\ell}_j \,|\, v_\ell\right) = \prod_j \mathcal{N}(\boldsymbol{\ell}_j; \boldsymbol{0}, v_\ell \boldsymbol{I}_O) \tag{C.45}$$

avec O la dimension du vecteur $\boldsymbol{\ell}_j$. Cela nous permet d'écrire en tenant compte de l'indépendance conditionnelle spatiale des données :

$$\begin{aligned} p\left(\boldsymbol{L} \,|\, \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{\sigma}^2, v_\ell\right) &\propto \prod_j p\left(\boldsymbol{\ell}_j \,|\, \boldsymbol{y}_j, \boldsymbol{a}_j, \boldsymbol{w}, \boldsymbol{h}, \sigma_j^2, v_\ell\right) \\ &\propto \prod_j p\left(\boldsymbol{y}_j \,|\, \boldsymbol{a}_j, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{\ell}_j, \sigma_j^2\right) p\left(\boldsymbol{\ell}_j \,|\, v_\ell\right) \\ &\propto \prod_j \exp\left[-\frac{1}{2}\,\left(\widetilde{\boldsymbol{y}}_j^{\mathrm{t}} \boldsymbol{\Gamma}_j \widetilde{\boldsymbol{y}}_j + \frac{1}{v_\ell}\,\boldsymbol{\ell}_j^{\mathrm{t}} \boldsymbol{\ell}_j\right)\right] \end{aligned} \tag{C.46}$$

où, pour faire ressortir les termes en $\boldsymbol{\ell}_j$, on écrit $\widetilde{\boldsymbol{y}}_j^{\mathrm{t}} \boldsymbol{\Gamma}_j \widetilde{\boldsymbol{y}}_j$ de la manière suivante :

$$\begin{aligned} \widetilde{\boldsymbol{y}}_j^{\mathrm{t}} \boldsymbol{\Gamma}_j \widetilde{\boldsymbol{y}}_j &= (\boldsymbol{y}_j - \boldsymbol{S}_j \boldsymbol{h} - \boldsymbol{P}\boldsymbol{\ell}_j)^{\mathrm{t}} \boldsymbol{\Gamma}_j (\boldsymbol{y}_j - \boldsymbol{S}_j \boldsymbol{h} - \boldsymbol{P}\boldsymbol{\ell}_j) \\ &= (\boldsymbol{y}_j - \boldsymbol{S}_j \boldsymbol{h})^{\mathrm{t}} \boldsymbol{\Gamma}_j (\boldsymbol{y}_j - \boldsymbol{S}_j \boldsymbol{h}) - 2\boldsymbol{\ell}_j^{\mathrm{t}} \boldsymbol{P}^{\mathrm{t}} \boldsymbol{\Gamma}_j (\boldsymbol{y}_j - \boldsymbol{S}_j \boldsymbol{h}) + \boldsymbol{\ell}_j^{\mathrm{t}} \boldsymbol{P}^{\mathrm{t}} \boldsymbol{\Gamma}_j \boldsymbol{P} \boldsymbol{\ell}_j \\ &= \boldsymbol{\ell}_j^{\mathrm{t}} (\boldsymbol{P}^{\mathrm{t}} \boldsymbol{\Gamma}_j \boldsymbol{P}) \boldsymbol{\ell}_j - 2\boldsymbol{\ell}_j^{\mathrm{t}} \boldsymbol{P}^{\mathrm{t}} \boldsymbol{\Gamma}_j\,(\boldsymbol{y}_j - \boldsymbol{S}_j \boldsymbol{h}) + \boldsymbol{C}_{\ell_j} \end{aligned}$$

En remplaçant le terme précédent dans l'équation (C.46) on obtient :

$$p\left(\boldsymbol{L} \,|\, \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{\sigma}^2, v_\ell\right) \propto \prod_j \exp\left\{\frac{-1}{2}\left[\boldsymbol{\ell}_j^{\mathrm{t}}\left(\frac{\boldsymbol{I}_O}{v_\ell} + \boldsymbol{P}^{\mathrm{t}} \boldsymbol{\Gamma}_j \boldsymbol{P}\right)\boldsymbol{\ell}_j - 2\boldsymbol{\ell}_j^{\mathrm{t}} \boldsymbol{P}^{\mathrm{t}} \boldsymbol{\Gamma}_j\,(\boldsymbol{y}_j - \boldsymbol{S}_j \boldsymbol{h})\right]\right\}.$$

Soit

$$p\left(\boldsymbol{L}\,|\,\boldsymbol{Y},\boldsymbol{A},\boldsymbol{w},\boldsymbol{h},\boldsymbol{\sigma}^2,v_\ell\right)=\prod_j \mathcal{N}(\boldsymbol{\ell}_j;\boldsymbol{\mu}_{\boldsymbol{\ell}_j},\boldsymbol{\Sigma}_{\boldsymbol{\ell}_j}) \tag{C.47}$$

avec

$$\boldsymbol{\Sigma}_{\boldsymbol{\ell}_j}=\left(\frac{\boldsymbol{I}_O}{v_\ell}+\boldsymbol{P}^{\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{P}\right)^{-1} \tag{C.48}$$

$$\boldsymbol{\mu}_{\boldsymbol{\ell}_j}=\Sigma_{\boldsymbol{\ell}_j}\boldsymbol{P}^{\mathrm{t}}\boldsymbol{\Gamma}_j\left(\boldsymbol{y}_j-\boldsymbol{S}_j\boldsymbol{h}\right) \tag{C.49}$$

♣ Distribution conditionnelle de v_l

En supposant un *a priori* non informatif de Jeffrey sur v_l tel que $p(v_\ell)=\frac{1}{v_\ell}$ on trouve :

$$\begin{aligned}p(v_\ell\,|\,\boldsymbol{L})&\propto p(v_\ell)\,p(\boldsymbol{L}\,|\,v_\ell)\\&\propto v_\ell^{-1}\prod_j v_\ell^{-O/2}\exp\left(-\frac{1}{2v_\ell}\,\boldsymbol{\ell}_j^{\mathrm{t}}\boldsymbol{\ell}_j\right)\\&\propto v_\ell^{-(OJ+2)/2}\exp\left(-\frac{1}{v_\ell}\,\sum_j\frac{||\boldsymbol{\ell}_j||^2}{2}\right).\end{aligned}$$

Soit

$$p(v_\ell\,|\,\boldsymbol{L})=\mathcal{IG}(v_\ell;\alpha_{v_\ell},\lambda_{v_\ell}) \tag{C.50}$$

avec $\alpha_{v_\ell}=\frac{OJ}{2}$ et $\lambda_{v_\ell}=\sum_j\frac{||\boldsymbol{\ell}_j||^2}{2}$.

♣ Distribution conditionnelle de ξ_m

On introduit un *a priori* uniforme sur $\boldsymbol{\xi}$ en supposant l'indépendance inter-conditions :

$$p\left(\boldsymbol{\xi}\right)=\prod_m p\left(\xi_m\right)=\prod_m \mathcal{U}_{[\xi_m^{\min},\xi_m^{\max}]} \tag{C.51}$$

avec $\xi_m^{\min}=0$ (indépendance spatiale) et $\xi_m^{\max}$ fixé à une valeur au délà de laquelle le champ est totalement corrélé, par exemple $\xi_m^{\max}=2$ dans le cas de deux classes

(active, inactive). La densité conditionnelle *a posteriori* de ξ s'écrit donc de la manière suivante :

$$p\left(\boldsymbol{\xi} \,|\, \boldsymbol{Q}\right) = \prod_{m=1}^{M} p\left(\xi_m \,|\, \boldsymbol{q}^m\right) \propto \prod_{m=1}^{M} p\left(\boldsymbol{q}^m \,|\, \xi_m\right)\; p\left(\xi_m\right) \tag{C.52}$$

avec

$$p\left(\xi_m \,|\, \boldsymbol{q}^m\right) \propto \frac{1}{Z\left(\xi_m\right)} \exp\left[\xi_m U\left(\boldsymbol{q}^m\right)\right]\; \mathcal{U}_{[\xi_m^{\min},\xi_m^{\max}]} \tag{C.53}$$

L'estimation de ξ_m est réalisée à travers une étape de Metropolis-Hasting (annexe F.1) :

1- Pour une valeur courante $\xi_m^{(t)}$ on génère un candidat

$$\xi_m^{(t+1/2)} \;\sim\; g\left(\xi_m^{(t+1/2)} \,|\, \xi_m^{(t)}\right) = \mathcal{N}_{[\xi^{\min},\xi^{\max}]}\left(\xi_m^{(t)}, \eta^2\right)$$

où $g\left(.\right)$ est la loi instrumentale définie comme une gaussienne tronquée (annexe F.2) de manière similaire à [Vincent 2010a].

2- On calcul la probabilité d'acceptation de la valeur candidate $\xi_m^{(t+1/2)}$

$$\alpha\left(\xi_m^{(t)} \to \xi_m^{(t+1/2)}\right) = \min\left(1, A_{t,t+1/2}\right) \tag{C.54}$$

$$\begin{aligned}
\text{où} \quad A_{t,t+1/2} &= \frac{p\left[\xi_m^{(t+1/2)} \mid \left(\boldsymbol{q}^m\right)^{(t)}\right]}{p\left[\xi_m^{(t)} \mid \left(\boldsymbol{q}^m\right)^{(t)}\right]} \frac{g\left(\xi_m^{(t)} \,|\, \xi_m^{(t+1/2)}\right)}{g\left(\xi_m^{(t+1/2)} \,|\, \xi_m^{(t)}\right)} \\
&= \frac{p\left[\left(\boldsymbol{q}_m\right)^{(t)} \,|\, \xi_m^{(t+1/2)}\right]\; p\left(\xi_m^{(t+1/2)}\right)}{p\left[\left(\boldsymbol{q}_m\right)^{(t)} \,|\, \xi_m^{(t)}\right]\; p\left(\xi_m^{(t)}\right)} \frac{g\left(\xi_m^{(t)} \,|\, \xi_m^{(t+1/2)}\right)}{g\left(\xi_m^{(t+1/2)} \,|\, \xi_m^{(t)}\right)} \\
&= \frac{\frac{\exp\left\{\xi_m^{(t+1/2)} U\left[\left(\boldsymbol{q}^m\right)^{(t)}\right]\right\}}{Z\left(\xi_m^{(t+1/2)}\right)}\; p\left(\xi_m^{(t+1/2)}\right)}{\frac{\exp\left\{\xi_m^{(t)} U\left[\left(\boldsymbol{q}^m\right)^{(t)}\right]\right\}}{Z\left(\xi_m^{(t)}\right)}\; p\left(\xi_m^{(t)}\right)} \frac{g\left(\xi_m^{(t)} \,|\, \xi_m^{(t+1/2)}\right)}{g\left(\xi_m^{(t+1/2)} \,|\, \xi_m^{(t)}\right)}
\end{aligned} \tag{C.55}$$

L'*a priori* sur ξ_m étant une loi uniforme, on a $p\left(\xi_m^{(t+1/2)}\right) = p\left(\xi_m^{(t)}\right)$. Soit :

$$A_{t,t+1/2} = \frac{Z\left(\xi_m^{(t)}\right)}{Z\left(\xi_m^{(t+1/2)}\right)} \exp\left\{\left(\xi_m^{(t+1/2)} - \xi_m^{(t)}\right) U\left[\left(\boldsymbol{q}^m\right)^{(t)}\right]\right\}\; B_{t,t+1/2} \tag{C.56}$$

avec $B_{t,t+1/2} = \frac{g\left(\xi_m^{(t)} \mid \xi_m^{(t+1/2)}\right)}{g\left(\xi_m^{(t+1/2)} \mid \xi_m^{(t)}\right)} = \frac{\mathcal{N}_{[\xi_{min},\xi_{max}]}\left(\xi_m^{(t+1/2)},\eta^2\right)}{\mathcal{N}_{[\xi_{min},\xi_{max}]}\left(\xi_m^{(t)},\eta^2\right)}$, où

$$\begin{aligned}\mathcal{N}_{[\xi_{min},\xi_{max}]}\left(\xi_m^{(t+1/2)},\eta^2\right) &= \frac{\mathcal{N}\left(\frac{\xi_m^{(t)}-\xi_m^{(t+1/2)}}{\eta}\right)}{\frac{\eta}{2}\left[\operatorname{erf}\left(\frac{\xi_{max}-\xi_m^{(t+1/2)}}{\sqrt{2}\,\eta}\right)-\operatorname{erf}\left(\frac{\xi_{min}-\xi_m^{(t+1/2)}}{\sqrt{2}\,\eta}\right)\right]} \\ &= \frac{\exp\left[-\frac{1}{2}\left(\frac{\xi_m^{(t)}-\xi_m^{(t+1/2)}}{\eta}\right)^2\right]}{\frac{\eta\sqrt{2\pi}}{2}\left[\operatorname{erf}\left(\frac{\xi_{max}-\xi_m^{(t+1/2)}}{\sqrt{2}\,\eta}\right)-\operatorname{erf}\left(\frac{\xi_{min}-\xi_m^{(t+1/2)}}{\sqrt{2}\,\eta}\right)\right]}\end{aligned} \tag{C.57}$$

et

$$\begin{aligned}\mathcal{N}_{[\xi_{min},\xi_{max}]}\left(\xi_m^{(t)},\eta^2\right) &= \frac{\mathcal{N}\left(\frac{\xi_m^{(t+1/2)}-\xi_m^{(t)}}{\eta}\right)}{\frac{\eta}{2}\left[\operatorname{erf}\left(\frac{\xi_{max}-\xi_m^{(t)}}{\sqrt{2}\,\eta}\right)-\operatorname{erf}\left(\frac{\xi_{min}-\xi_m^{(t)}}{\sqrt{2}\,\eta}\right)\right]} \\ &= \frac{\exp\left[-\frac{1}{2}\left(\frac{\xi_m^{(t+1/2)}-\xi_m^{(t)}}{\eta}\right)^2\right]}{\frac{\eta\sqrt{2\pi}}{2}\left[\operatorname{erf}\left(\frac{\xi_{max}-\xi_m^{(t)}}{\sqrt{2}\,\eta}\right)-\operatorname{erf}\left(\frac{\xi_{min}-\xi_m^{(t)}}{\sqrt{2}\,\eta}\right)\right]}\end{aligned} \tag{C.58}$$

Ce qui donne :

$$B_{t,t+1/2} = \frac{\operatorname{erf}\left(\frac{\xi_{max}-\xi_m^{(t)}}{\sqrt{2}\,\eta}\right)-\operatorname{erf}\left(\frac{\xi_{min}-\xi_m^{(t)}}{\sqrt{2}\,\eta}\right)}{\operatorname{erf}\left(\frac{\xi_{max}-\xi_m^{(t+1/2)}}{\sqrt{2}\,\eta}\right)-\operatorname{erf}\left(\frac{\xi_{min}-\xi_m^{(t+1/2)}}{\sqrt{2}\,\eta}\right)}. \tag{C.59}$$

Ainsi on met à jour :

$$\xi_m^{(t+1)} = \begin{cases} \xi_m^{(t+1/2)} & \text{avec la probabilité} \quad \alpha\left(\xi_m^{(t)} \to \xi_m^{(t+1/2)}\right), \\ \xi_m^{(t)} & \text{avec la probabilité} \quad 1-\alpha\left(\xi_m^{(t)} \to \xi_m^{(t+1/2)}\right). \end{cases} \tag{C.60}$$

Annexe D

Algorithme EM Variationnel

Ci-dessous le détail des calculs concernant les étapes **E** et **M** de l'inférence par VEM dans le cadre de l'approche DEC (section III.2.2.2 du chapitre III). Rappelons que les termes constants en x seront notés C_x ou parfois $C_{1,x}, C_{2,x}, etc$ quand plusieurs constantes différentes interviennent dans une même formule. Ils ne seront détaillés que s'ils sont nécessaires pour l'étape de maximisation. Rappelons aussi que les termes $\widetilde{p}_{w^m}(w^m = 1)$ et $\widetilde{p}_{q_j^m}(q_j^m = 1)$ seront respectivement notés $\widetilde{p}_{w^m}(1)$ et $\widetilde{p}_{q_j^m}(1)$ pour plus de simplicité. Par la suite les matrices et vecteurs sont notés en gras avec les matrices en majuscule et les vecteurs en minuscule. Un vecteur est par convention un vecteur colonne. La transposé est notée ^{t}.

D.1 Connexion aux variables de classification

Rappelons que dans cette approche DEC *parcimonieuse*, les variables de pertinence $\boldsymbol{w}$ sont connectées aux labels $\boldsymbol{Q}$.

♣ E-H

$$\begin{aligned}\widetilde{p}_{\boldsymbol{h}}(\boldsymbol{h}) &\propto \exp\left(E_{\widetilde{p}_{\boldsymbol{A}},\widetilde{p}_{\boldsymbol{Q}},\widetilde{p}_{\boldsymbol{w}}}[\log p(\boldsymbol{h}\,|\,\boldsymbol{Y},\boldsymbol{A},\boldsymbol{w},\boldsymbol{Q},\boldsymbol{L},\boldsymbol{\theta})]\right)\\ &\propto \exp\left(E_{\widetilde{p}_{\boldsymbol{A}},\widetilde{p}_{\boldsymbol{w}}}[\log p\left(\boldsymbol{Y}\,|\,\boldsymbol{A},\boldsymbol{w},\boldsymbol{h},\boldsymbol{L},\boldsymbol{\sigma}^2\right)] + \log p\left(\boldsymbol{h}\,|\,v_h\right)\right)\end{aligned} \tag{D.1}$$

1) Calcul de $\log p\left(\boldsymbol{Y}\,|\,\boldsymbol{A},\boldsymbol{w},\boldsymbol{h},\boldsymbol{L},\boldsymbol{\sigma}^2\right)$:
Selon la définition de la vraisemblance vue dans le chapitre II, section II.4, et en rappelant que $\widetilde{\boldsymbol{y}}_j = \boldsymbol{y}_j - \boldsymbol{S}_j\boldsymbol{h} - \boldsymbol{P}\boldsymbol{\ell}_j$ et qu'on ne s'intéresse qu'aux termes dépendant de $\boldsymbol{h}$, on peut écrire :

$$\begin{aligned}
&\log p\left(\boldsymbol{Y} \mid \boldsymbol{A}, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{L}, \boldsymbol{\sigma}^2\right) \\
&\qquad = C_{1,H} - \frac{1}{2} \sum_{j \in J} \widetilde{\boldsymbol{y}}_j^{\mathrm{t}} \boldsymbol{\Gamma}_j \widetilde{\boldsymbol{y}}_j \\
&\qquad = C_{2,H} - \frac{1}{2} \sum_{j \in J} \left[\boldsymbol{h}^{\mathrm{t}} \boldsymbol{S}_j^{\mathrm{t}} \boldsymbol{\Gamma}_j \boldsymbol{S}_j \boldsymbol{h} - 2 \boldsymbol{h}^{\mathrm{t}} \boldsymbol{S}_j^{\mathrm{t}} \boldsymbol{\Gamma}_j (\boldsymbol{y}_j - \boldsymbol{P} \boldsymbol{\ell}_j) \right]
\end{aligned} \tag{D.2}$$

Où $\boldsymbol{\Gamma}_j = \frac{1}{\sigma_j^2} \boldsymbol{\Lambda}_j$, $C_{1,\boldsymbol{h}} = -\frac{NJ}{2} \log 2\pi + \frac{J}{2} \log |\boldsymbol{\Lambda}_j| - N \sum_{j \in J} \log \sigma_j$ et $C_{2,\boldsymbol{h}} = C_{1,\boldsymbol{h}} - \frac{1}{2} \sum_{j \in J} (\boldsymbol{y}_j - \boldsymbol{P} \boldsymbol{\ell}_j)^{\mathrm{t}} \boldsymbol{\Gamma}_j (\boldsymbol{y}_j - \boldsymbol{P} \boldsymbol{\ell}_j)$.

2) Calcul de $\log p\left(\boldsymbol{h} \mid v_h\right)$:

Ce calcul est basé sur la définition de l'*a priori* sur la FRH et qu'on a vu dans la chapitre II, équation (II.24) :

$$\log p\left(\boldsymbol{h} \mid v_h\right) = C_{3,\boldsymbol{h}} - \frac{1}{2 v_h} \boldsymbol{h}^{\mathrm{t}} \boldsymbol{R}^{-1} \boldsymbol{h} \tag{D.3}$$

avec $C_{3,\boldsymbol{h}} = -\frac{D+1}{2} \log 2\pi - \frac{D-1}{2} \log v_h - \frac{1}{2} \log |\boldsymbol{R}|$.

En remplaçant les équations (D.2) et (D.3) dans l'équation (D.1) on obtient :

$$\begin{aligned}
\widetilde{p}_{\boldsymbol{h}}(\boldsymbol{h}) \propto \exp \Bigg(-\frac{1}{2} \Bigg[\boldsymbol{h}^{\mathrm{t}} \left(\frac{\boldsymbol{R}^{-1}}{v_h} + E_{\widetilde{p}_{\boldsymbol{A}}, \widetilde{p}_{\boldsymbol{w}}} \left[\sum_{j \in J} \boldsymbol{S}_j^{\mathrm{t}} \boldsymbol{\Gamma}_j \boldsymbol{S}_j \right] \right) \boldsymbol{h} \\
-2 \boldsymbol{h}^{\mathrm{t}} E_{\widetilde{p}_{\boldsymbol{A}}, \widetilde{p}_{\boldsymbol{w}}} \left[\sum_{j \in J} \boldsymbol{S}_j^{\mathrm{t}} \boldsymbol{\Gamma}_j (\boldsymbol{y}_j - \boldsymbol{P} \boldsymbol{\ell}_j) \right] \Bigg] \Bigg) .
\end{aligned} \tag{D.4}$$

D'où

$$\widetilde{p}_{\boldsymbol{h}}(\boldsymbol{h}) = \mathcal{N}(\boldsymbol{m}_H, \boldsymbol{\Sigma}_H) \tag{D.5}$$

avec :

$$\boldsymbol{\Sigma}_{\boldsymbol{h}} = \left(\frac{\boldsymbol{R}^{-1}}{v_h} + E_{\widetilde{p}_{\boldsymbol{A}}, \widetilde{p}_{\boldsymbol{w}}} \left[\sum_{j \in J} \boldsymbol{S}_j^{\mathrm{t}} \boldsymbol{\Gamma}_j \boldsymbol{S}_j \right] \right)^{-1} \tag{D.6}$$

$$\boldsymbol{m}_{\boldsymbol{h}} = \boldsymbol{\Sigma}_H E_{\widetilde{p}_{\boldsymbol{A}}, \widetilde{p}_{\boldsymbol{w}}} \left[\sum_{j \in J} \boldsymbol{S}_j^{\mathrm{t}} \boldsymbol{\Gamma}_j (\boldsymbol{y}_j - \boldsymbol{P} \boldsymbol{\ell}_j) \right] \tag{D.7}$$

- $E_{\widetilde{p}_{\boldsymbol{A}},\widetilde{p}_{\boldsymbol{w}}}\left[\sum_{j\in J}\boldsymbol{S}_j^{\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{S}_j\right]$

$$\begin{aligned}
&= E_{\widetilde{p}_{\boldsymbol{A}},\widetilde{p}_{\boldsymbol{w}}}\left[\sum_{j\in J}\sum_{m,m'} a_j^m w^m \boldsymbol{X}^{m\mathrm{t}}\boldsymbol{\Gamma}_j a_j^{m'} w^{m'}\boldsymbol{X}^{m'}\right]\\
&= \sum_{j\in J}\sum_{m,m'} E_{\widetilde{p}_{A_j^{m,m'}}}\left[a_j^m a_j^{m'}\right] E_{\widetilde{p}_{w^{m,m'}}}\left[w^m w^{m'}\right]\boldsymbol{X}^{m\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{X}^{m'}\\
&= \sum_{j\in J}\sum_{m,m'}\left(m_{a_j^m}m_{a_j^{m'}} + V_{\boldsymbol{a}_j^{m,m'}}\right)\widetilde{p}_{w^m}(1)\widetilde{p}_{w^{m'}}(1)\boldsymbol{X}^{m\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{X}^{m'} \qquad \text{(D.8)}
\end{aligned}$$

- $E_{\widetilde{p}_{\boldsymbol{A}},\widetilde{p}_{\boldsymbol{w}}}\left[\sum_{j\in J}\boldsymbol{S}_j^{\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j-\boldsymbol{P}\boldsymbol{\ell}_j)\right]$

$$\begin{aligned}
&= E_{\widetilde{p}_{\boldsymbol{A}},\widetilde{p}_{\boldsymbol{w}}}\left[\sum_{j\in J}\sum_{m=1}^{M} a_j^m w^m \boldsymbol{X}^{m\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j-\boldsymbol{P}\boldsymbol{\ell}_j)\right]\\
&= \sum_{j\in J}\sum_{m=1}^{M} E_{\widetilde{p}_{a_j^m}}\left[a_j^m\right] E_{\widetilde{p}_{w^m}}\left[w^m\right]\boldsymbol{X}^{m\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j-\boldsymbol{P}\boldsymbol{\ell}_j)\\
&= \sum_{j\in J}\sum_{m=1}^{M} m_{a_j^m}\widetilde{p}_{w^m}(1)\boldsymbol{X}^{m\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j-\boldsymbol{P}\boldsymbol{\ell}_j) \qquad \text{(D.9)}
\end{aligned}$$

♣ **E-A**

$$\begin{aligned}
\widetilde{p}_{\boldsymbol{A}}(\boldsymbol{A}) &\propto \exp\left(E_{\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{\boldsymbol{Q}},\widetilde{p}_{\boldsymbol{w}}}\left[\log p(\boldsymbol{A}\,|\,\boldsymbol{Y},\boldsymbol{w},\boldsymbol{h},\boldsymbol{Q},\boldsymbol{L},\boldsymbol{\theta})\right]\right)\\
&\propto \exp\left(E_{\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{\boldsymbol{w}}}\left[\log p\left(\boldsymbol{Y}\,|\,\boldsymbol{A},\boldsymbol{w},\boldsymbol{h},\boldsymbol{L},\boldsymbol{\sigma}^2\right)\right] + E_{\widetilde{p}_{\boldsymbol{Q}},\widetilde{p}_{\boldsymbol{w}}}\left[\log p(\boldsymbol{A}\,|\,\boldsymbol{w},\boldsymbol{Q},\boldsymbol{\theta}_A)\right]\right)
\end{aligned} \qquad \text{(D.10)}$$

1) Calcul de $E_{\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{\boldsymbol{w}}}\left[\log p\left(\boldsymbol{Y}\,|\,\boldsymbol{A},\boldsymbol{w},\boldsymbol{h},\boldsymbol{L},\boldsymbol{\sigma}^2\right)\right]$:
De la même façon que dans la section précédente la log vraisemblance s'écrit :

$$\begin{aligned}
\log p\left(\boldsymbol{Y}\,|\,\boldsymbol{A},\boldsymbol{w},\boldsymbol{h},\boldsymbol{L},\boldsymbol{\sigma}^2\right) &= C_{1,\boldsymbol{A}} - \frac{1}{2}\sum_{j\in J}\widetilde{\boldsymbol{y}}_j^{\mathrm{t}}\boldsymbol{\Gamma}_j\widetilde{\boldsymbol{y}}_j\\
&= C_{2,\boldsymbol{A}} - \frac{1}{2}\sum_{j\in J}\left(\boldsymbol{a}_j^{\mathrm{t}}\boldsymbol{G}^{\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{G}\boldsymbol{a}_j - 2\boldsymbol{a}_j^{\mathrm{t}}\boldsymbol{G}^{\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j-\boldsymbol{P}\boldsymbol{\ell}_j)\right)
\end{aligned} \qquad \text{(D.11)}$$

où $C_{1,\boldsymbol{A}} = C_{1,\boldsymbol{h}}$ est déjà définie dans la section précédente et $C_{2,\boldsymbol{A}} = C_{1,\boldsymbol{A}} - \frac{1}{2}\sum_{j\in J}(\boldsymbol{y}_j-\boldsymbol{P}\boldsymbol{\ell}_j)^{\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j-\boldsymbol{P}\boldsymbol{\ell}_j)$. $\boldsymbol{a}_j = [a_j^1,...,a_j^M]^{\mathrm{t}}$ et $\boldsymbol{G} = [\boldsymbol{g}^1|\boldsymbol{g}^2|...|\boldsymbol{g}^M]$ avec $\boldsymbol{g}^m =$

$w^m \boldsymbol{X}^m \boldsymbol{h}$.

On peut maintenant moyenner le terme précédent (équation (D.11)) part rapport aux autres variables du modèle :

$$\begin{aligned}
&E_{\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{\boldsymbol{w}}}\left[\log p\left(\boldsymbol{Y} \mid \boldsymbol{A}, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{L}, \boldsymbol{\sigma}^2\right)\right] \\
&= C_{2,\boldsymbol{A}} - \frac{1}{2}\sum_{j\in J}\Big(\boldsymbol{a}_j^{\mathrm{t}} E_{\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{\boldsymbol{w}}}\left[\boldsymbol{G}^{\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{G}\right]\boldsymbol{a}_j - 2\boldsymbol{a}_j^{\mathrm{t}} E_{\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{\boldsymbol{w}}}\left[\boldsymbol{G}^{\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j - \boldsymbol{P}\boldsymbol{\ell}_j)\right]\Big)
\end{aligned} \tag{D.12}$$

- $E_{\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{\boldsymbol{w}}}\left[\boldsymbol{G}^{\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{G}\right]$ est une matrice de dimension $M \times M$ dont chaque élément (m, m') s'écrit sous la forme suivante :

$$\begin{aligned}
&E_{\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{w^{m,m'}}}\left[\boldsymbol{g}^{m\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{g}^{m'}\right] \\
&= E_{\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{w^{m,m'}}}\left[w^m\boldsymbol{h}^{\mathrm{t}}\boldsymbol{X}^{m\mathrm{t}}\boldsymbol{\Gamma}_j w^{m'}\boldsymbol{X}^{m'}\boldsymbol{h}\right] \\
&= E_{\widetilde{p}_{w^{m,m'}}}\left[w^m w^{m'}\right]\left[E_{\widetilde{p}_{\boldsymbol{h}}}\left[\boldsymbol{h}^{\mathrm{t}}\right]\boldsymbol{X}^{m\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{X}^{m'}E_{\widetilde{p}_{\boldsymbol{h}}}\left[\boldsymbol{h}\right] + \mathrm{tr}\left(\boldsymbol{\Sigma}_H\boldsymbol{X}^{m\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{X}^{m'}\right)\right] \\
&= \widetilde{p}_{w^m}(1)\widetilde{p}_{w^{m'}}(1)\left(\boldsymbol{m}_H^{\mathrm{t}}\boldsymbol{X}^{m\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{X}^{m'}\boldsymbol{m}_H + \mathrm{tr}\left(\boldsymbol{\Sigma}_H\boldsymbol{X}^{m\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{X}^{m'}\right)\right)
\end{aligned} \tag{D.13}$$

- $E_{\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{\boldsymbol{w}}}\left[\boldsymbol{G}^{\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j - \boldsymbol{P}\boldsymbol{\ell}_j)\right]$ est un vecteur de dimension $M \times 1$ dont chaque élément $(m, 1)$ est définit comme suit :

$$\begin{aligned}
E_{\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{w^m}}\left[\boldsymbol{g}^{m\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j - \boldsymbol{P}\boldsymbol{\ell}_j)\right] &= E_{\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{w^m}}\left[w^m\boldsymbol{h}^{\mathrm{t}}\boldsymbol{X}^{m\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j - \boldsymbol{P}\boldsymbol{\ell}_j)\right] \\
&= E_{\widetilde{p}_{w^m}}\left[w^m\right]E_{\widetilde{p}_{\boldsymbol{h}}}\left[\boldsymbol{h}^{\mathrm{t}}\right]\boldsymbol{X}^{m\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j - \boldsymbol{P}\boldsymbol{\ell}_j) \\
&= \widetilde{p}_{w^m}(1)\,\boldsymbol{m}_H^{\mathrm{t}}\boldsymbol{X}^{m\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j - \boldsymbol{P}\boldsymbol{\ell}_j)
\end{aligned} \tag{D.14}$$

2) Calcul de $E_{\widetilde{p}_{\boldsymbol{Q}},\widetilde{p}_{\boldsymbol{w}}}\left[\log p(\boldsymbol{A} \mid \boldsymbol{w}, \boldsymbol{Q}, \boldsymbol{\theta}_A)\right]$:

En exploitant les équations (III.5) et (III.6), on peut écrire :

$$\begin{aligned}
&\log p(\boldsymbol{A} \mid \boldsymbol{w}, \boldsymbol{Q}, \boldsymbol{\theta}_A) \\
&= \sum_m\sum_j\left[(1 - q_j^m w^m)\log\left(\mathcal{N}(\mu_0^m, v_0^m)\right) + q_j^m w^m \log\left(\mathcal{N}(\mu_1^m, v_1^m)\right)\right] \\
&= C_{3,\boldsymbol{A}} - \frac{1}{2}\sum_m\sum_j\left[(1 - q_j^m w^m)\left(\frac{(a_j^m - \mu_0^m)^2}{v_0^m}\right) + q_j^m w^m\left(\frac{(a_j^m - \mu_1^m)^2}{v_1^m}\right)\right]
\end{aligned} \tag{D.15}$$

où $C_{3,\boldsymbol{A}} = \sum_m\sum_j\left[(1 - q_j^m w^m)\log\frac{1}{\sqrt{2\pi v_0^m}} + q_j^m w^m \log\frac{1}{\sqrt{2\pi v_1^m}}\right]$

$$
\begin{aligned}
&E_{\widetilde{p}_{\boldsymbol{Q}},\widetilde{p}_{\boldsymbol{w}}}\left[\log p(\boldsymbol{A}\,|\,\boldsymbol{w},\boldsymbol{Q},\boldsymbol{\theta}_A)\right] \\
&\quad = C'_{3,\boldsymbol{A}} - \frac{1}{2}\sum_m\sum_j \left[E_{\widetilde{p}_{q_j^m},\widetilde{p}_{w^m}}\left[(1-q_j^m w^m)\right]\left(\frac{(a_j^m-\mu_0^m)^2}{v_0^m}\right)\right. \\
&\qquad\qquad \left. + E_{\widetilde{p}_{q_j^m},\widetilde{p}_{w^m}}\left[q_j^m w^m\right]\left(\frac{(a_j^m-\mu_1^m)^2}{v_1^m}\right)\right] \\
&\quad = C'_{3,\boldsymbol{A}} - \frac{1}{2}\sum_m\sum_j \left[\left(1-\widetilde{p}_{q_j^m}(1)\widetilde{p}_{w^m}(1)\right)\left(\frac{(a_j^m-\mu_0^m)^2}{v_0^m}\right)\right. \\
&\qquad\qquad \left. + \widetilde{p}_{q_j^m}(1)\widetilde{p}_{w^m}(1)\left(\frac{(a_j^m-\mu_1^m)^2}{v_1^m}\right)\right]
\end{aligned}
\tag{D.16}
$$

où $C'_{3,\boldsymbol{A}} = E_{\widetilde{p}_{\boldsymbol{Q}},\widetilde{p}_{\boldsymbol{w}}}\left[C_{3,\boldsymbol{A}}\right]$.

Nous allons maintenant écrire notre équation sous forme matricielle afin de s'adapter à la première partie du calcul (équation (D.11)). En notant : $\boldsymbol{a}_j = [a_j^1, ..., a_j^M]^{\mathrm{t}}$, $\boldsymbol{q}_j = [q_j^1, ..., q_j^M]^{\mathrm{t}}$, $\boldsymbol{w} = [w^1, ..., w^M]^{\mathrm{t}}$, $\boldsymbol{\mu}_i = [\mu_i^1, ..., \mu_i^M]^{\mathrm{t}}$ pour $i \in \{0,1\}$ et $\Delta_{0,j} = \mathrm{diag}\left[\frac{1-\widetilde{p}_{Q_j^1}(1)\widetilde{p}_{W^1}(1)}{v_0^1}, ..., \frac{1-\widetilde{p}_{Q_j^M}(1)\widetilde{p}_{W^M}(1)}{v_0^M}\right]$, $\Delta_{1,j} = \mathrm{diag}\left[\frac{\widetilde{p}_{Q_j^1}(1)\widetilde{p}_{W^1}(1)}{v_1^1}, ..., \frac{\widetilde{p}_{Q_j^M}(1)\widetilde{p}_{W^M}(1)}{v_1^M}\right]$, l'équation (D.16) devient :

$$
\begin{aligned}
&E_{\widetilde{p}_{\boldsymbol{Q}},\widetilde{p}_{\boldsymbol{w}}}\left[\log p(\boldsymbol{A}\,|\,\boldsymbol{w},\boldsymbol{Q},\boldsymbol{\theta}_A)\right] \\
&\quad = C''_{3,\boldsymbol{A}} - \frac{1}{2}\sum_j\sum_{i=0}^{1}(\boldsymbol{a}_j-\boldsymbol{\mu}_i)^{\mathrm{t}}\Delta_{i,j}(\boldsymbol{a}_j-\boldsymbol{\mu}_i) \\
&\quad = C''_{3,\boldsymbol{A}} + C_{4,\boldsymbol{A}} - \frac{1}{2}\sum_j\left[\boldsymbol{a}_j^{\mathrm{t}}\left(\sum_{i=0}^{1}\Delta_{i,j}\right)\boldsymbol{a}_j - 2\boldsymbol{a}_j^{\mathrm{t}}\left(\sum_{i=0}^{1}\Delta_{i,j}\boldsymbol{\mu}_i\right)\right]
\end{aligned}
\tag{D.17}
$$

où $C''_{3,\boldsymbol{A}}$ est la forme matricielle de $C'_{3,\boldsymbol{A}}$, $C_{4,\boldsymbol{A}} = -\frac{1}{2}\sum_j\sum_{i=0}^{1}\boldsymbol{\mu}_i^{\mathrm{t}}\Delta_{i,j}\boldsymbol{\mu}_i$.

En remplaçant les équations (D.12) et (D.17) dans l'équation (D.10) on obtient :

$$
\widetilde{p}_{\boldsymbol{A}}(\boldsymbol{A}) \propto \exp\left(-\frac{1}{2}\sum_j(\boldsymbol{a}_j-\boldsymbol{m}_{\boldsymbol{a}_j})^{\mathrm{t}}\boldsymbol{V}_{\boldsymbol{a}_j}^{-1}(\boldsymbol{a}_j-\boldsymbol{m}_{\boldsymbol{a}_j})\right).
$$

Soit

$$
\widetilde{p}_{\boldsymbol{A}}(\boldsymbol{A}) = \prod_j \mathcal{N}\left(\boldsymbol{m}_{\boldsymbol{a}_j}, \boldsymbol{V}_{\boldsymbol{a}_j}\right) \tag{D.18}
$$

avec :

$$\boldsymbol{V}_{\boldsymbol{a}_j} = \left(E_{\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{\boldsymbol{w}}} \left[\boldsymbol{G}^{\mathrm{t}} \boldsymbol{\Gamma}_j \boldsymbol{G} \right] + \sum_{i=0}^{1} \Delta_{i,j} \right)^{-1} \tag{D.19}$$

$$\boldsymbol{m}_{\boldsymbol{a}_j} = \boldsymbol{V}_{\boldsymbol{a}_j} \left(E_{\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{\boldsymbol{w}}} \left[\boldsymbol{G}^{\mathrm{t}} \boldsymbol{\Gamma}_j (\boldsymbol{y}_j - \boldsymbol{P}\boldsymbol{\ell}_j) \right] + \sum_{i=0}^{1} \Delta_{i,j} \boldsymbol{\mu}_i \right) \tag{D.20}$$

♣ **E-W**

$$\begin{aligned} \widetilde{p}_{w^m}(w^m) &\propto \exp\left(E_{\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{\boldsymbol{A}},\widetilde{p}_{\boldsymbol{Q}},\widetilde{p}_{\boldsymbol{w}^{\backslash m}}} \left[\log p(w^m \mid \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{Q}, \boldsymbol{h}, \boldsymbol{w}^{\backslash m}, \boldsymbol{L}, \boldsymbol{\theta}) \right] \right) \\ &\propto f(w^m) \end{aligned} \tag{D.21}$$

où on note :

$$\begin{aligned} f(w^m) = \exp\Big(& E_{\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{\boldsymbol{A}},\widetilde{p}_{\boldsymbol{w}^{\backslash m}}} \left[\log p(\boldsymbol{Y} \mid \boldsymbol{A}, \boldsymbol{h}, w^m, \boldsymbol{w}^{\backslash m}, \boldsymbol{L}, \boldsymbol{\sigma}^2) \right] \\ & + E_{\widetilde{p}_{\boldsymbol{a}^m},\widetilde{p}_{\boldsymbol{q}^m}} \left[\log p(\boldsymbol{a}^m \mid w^m, \boldsymbol{q}^m, \boldsymbol{\theta}_m) \right] \\ & + E_{\widetilde{p}_{\boldsymbol{q}^m}} \left[\log p(w^m \mid \boldsymbol{q}^m, \boldsymbol{\tau}) \right] \Big) \end{aligned} \tag{D.22}$$

et où $\widetilde{p}_{\boldsymbol{w}^{\backslash m}}$ désigne la loi de $\left\{ \boldsymbol{w}^{m'} \right\}_{m' \neq m}$.

Nous allons par la suite détailler le calcul des trois termes de l'exponentielle précédente.

1) Calcul de $E_{\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{\boldsymbol{A}},\widetilde{p}_{\boldsymbol{w}^{\backslash m}}} \left[\log p(\boldsymbol{Y} \mid \boldsymbol{A}, \boldsymbol{h}, w^m, \boldsymbol{w}^{\backslash m}, \boldsymbol{L}, \boldsymbol{\sigma}^2) \right]$:

$$\begin{aligned} & \log p(\boldsymbol{Y} \mid \boldsymbol{A}, \boldsymbol{h}, w^m, \boldsymbol{w}^{\backslash m}, \boldsymbol{L}, \boldsymbol{\sigma}^2) \\ & = C_{1,w^m} - \frac{1}{2} \sum_{j \in J} \widetilde{\boldsymbol{y}}_j^{\mathrm{t}} \boldsymbol{\Gamma}_j \widetilde{\boldsymbol{y}}_j \\ & = C_{2,w^m} - \frac{1}{2} (w^m)^2 \sum_{j \in J} (a_j^m)^2 \boldsymbol{h}^{\mathrm{t}} \boldsymbol{X}^{m\mathrm{t}} \boldsymbol{\Gamma}_j \boldsymbol{X}^m \boldsymbol{h} \\ & \quad + w^m \sum_{j \in J} \left(a_j^m \boldsymbol{h}^{\mathrm{t}} \boldsymbol{X}^{m\mathrm{t}} \boldsymbol{\Gamma}_j (\boldsymbol{y}_j - \boldsymbol{P}\boldsymbol{\ell}_j) - \sum_{m' \neq m} a_j^m a_j^{m'} w^{m'} \boldsymbol{h}^{\mathrm{t}} \boldsymbol{X}^{m\mathrm{t}} \boldsymbol{\Gamma}_j \boldsymbol{X}^{m'} \boldsymbol{h} \right) \end{aligned} \tag{D.23}$$

$$
\begin{aligned}
&E_{\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{\boldsymbol{A}},\widetilde{p}_{\boldsymbol{w}^{\backslash m}}}\left[\log p(\boldsymbol{Y}\mid \boldsymbol{A},\boldsymbol{h},w^m,\boldsymbol{w}^{\backslash m},\boldsymbol{L},\boldsymbol{\sigma}^2)\right]\\
&\quad= C'_{2,w^m} - \frac{1}{2}(w^m)^2\sum_{j\in J} E_{\widetilde{p}_{\boldsymbol{a}_j^{m,m}}}\left[(a_j^m)^2\right] E_{\widetilde{p}_{\boldsymbol{h}}}\left[\boldsymbol{h}^{\mathrm{t}}\boldsymbol{X}^{m\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{X}^m\boldsymbol{h}\right]\\
&\qquad+ w^m\sum_{j\in J}\left(E_{\widetilde{p}_{a_j^m}}\left[a_j^m\right] E_{\widetilde{p}_{\boldsymbol{h}}}\left[\boldsymbol{h}^{\mathrm{t}}\right]\boldsymbol{X}^{m\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j-\boldsymbol{P}\boldsymbol{\ell}_j)\right)\\
&\qquad- w^m\sum_{j\in J}\sum_{m'\neq m}\left(E_{\widetilde{p}_{\boldsymbol{a}_j^{m,m'}}}\left[a_j^m a_j^{m'}\right] E_{\widetilde{p}_{w^{m'}}}\left[w^{m'}\right] E_{\widetilde{p}_{\boldsymbol{h}}}\left[\boldsymbol{h}^{\mathrm{t}}\boldsymbol{X}^{m\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{X}^{m'}\boldsymbol{h}\right]\right)\\
&\quad= C'_{2,w^m} - \frac{1}{2}w^m\sum_{j\in J}\Big\{-2\, m_{a_j^m}\boldsymbol{m}_H^{\mathrm{t}}\boldsymbol{X}^{m\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j-\boldsymbol{P}\boldsymbol{\ell}_j)\\
&\qquad+\left(m_{a_j^m}^2+V_{\boldsymbol{a}_j^{m,m}}\right)\left[\boldsymbol{m}_H^{\mathrm{t}}\boldsymbol{X}^{m\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{X}^m\boldsymbol{m}_H + tr\left(\boldsymbol{\Sigma}_H\boldsymbol{X}^{m\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{X}^m\right)\right]\\
&\qquad+ 2\sum_{m'\neq m}\left(m_{a_j^m}m_{a_j^{m'}}+V_{\boldsymbol{a}_j^{m,m'}}\right)\widetilde{p}_{w^{m'}}(1)\left[\boldsymbol{m}_H^{\mathrm{t}}\boldsymbol{X}^{m\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{X}^{m'}\boldsymbol{m}_H\right.\\
&\qquad\left.+\operatorname{tr}\left(\boldsymbol{\Sigma}_H\boldsymbol{X}^{m\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{X}^{m'}\right)\right]\Big\}
\end{aligned}
\tag{D.24}
$$

où $C'_{2,w^m} = E_{\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{\boldsymbol{A}},\widetilde{p}_{\boldsymbol{w}^{\backslash m}}}\left[C_{2,w^m}\right]$

2) Calcul de $E_{\widetilde{p}_{\boldsymbol{a}^m},\widetilde{p}_{\boldsymbol{q}^m}}\left[\log p(\boldsymbol{A}\mid w^m,\boldsymbol{q}^m,\boldsymbol{\theta}_m)\right]$:
$\log p(\boldsymbol{a}^m\mid w^m,\boldsymbol{q}^m,\boldsymbol{\theta}_m)$

$$
\begin{aligned}
&= \sum_j\left[\left(1-q_j^m w^m\right)\log\mathcal{N}(\mu_0^m,v_0^m) + q_j^m w^m\log\mathcal{N}(\mu_1^m,v_1^m)\right]\\
&= C_{4,w^m} + \frac{1}{2}\sum_j\left[q_j^m w^m\left(\log(2\pi v_0^m) + \frac{(a_j^m)^2-2a_j^m\mu_0^m+(\mu_0^m)^2}{v_0^m}\right.\right.\\
&\qquad\left.\left.-\log(2\pi v_1^m) - \frac{(a_j^m)^2-2a_j^m\mu_1^m+(\mu_1^m)^2}{v_1^m}\right)\right]
\end{aligned}
\tag{D.25}
$$

$E_{\widetilde{p}_{\boldsymbol{a}^m},\widetilde{p}_{\boldsymbol{q}^m}}\left[\log p(\boldsymbol{a}^m\mid w^m,\boldsymbol{q}^m,\boldsymbol{\theta}_m)\right]$

$$
\begin{aligned}
&= C'_{4,w^m} + w^m\sum_j\widetilde{p}_{q_j^m}(1)\left(\log\sqrt{2\pi v_0^m} + \frac{(m_{a_j^m}-\mu_0^m)^2+V_{\boldsymbol{a}_j^{m,m}}}{2v_0^m}\right.\\
&\qquad\left.-\log\sqrt{2\pi v_1^m} - \frac{(m_{a_j^m}-\mu_1^m)^2+V_{\boldsymbol{a}_j^{m,m}}}{2v_1^m}\right)\\
&= C'_{4,w^m} + w^m\sum_j\widetilde{p}_{q_j^m}(1)\left\{\frac{V_{\boldsymbol{a}_j^{m,m}}}{2}\left(\frac{1}{v_0^m}-\frac{1}{v_1^m}\right) + \log\frac{\mathcal{N}(m_{a_j^m};\mu_1^m,v_1^m)}{\mathcal{N}(m_{a_j^m};\mu_0^m,v_0^m)}\right\}
\end{aligned}
\tag{D.26}
$$

où $C'_{4,w^m} = E_{\widetilde{p}_{\boldsymbol{a}^m},\widetilde{p}_{\boldsymbol{q}^m}}\left[C_{4,w^m}\right]$.

3) Calcul de $E_{\widetilde{p}_{\boldsymbol{q}^m}}\left[\log p(w^m \mid \boldsymbol{q}^m, \boldsymbol{\tau})\right]$:

En revenant à la distribution *a priori* de $\boldsymbol{w}$ (équation (III.7)) on peut écrire :

$$\log p(w^m \mid \boldsymbol{q}^m, \boldsymbol{\tau}) = w^m \log p_m + (1 - w^m)\log(1 - p_m) \tag{D.27}$$

Selon l'équation (III.8), $p_m = \frac{1}{1+\exp[-\tau_1(\sum_j q_j^m - \tau_2)]}$, donc $1-p_m = \frac{\exp[-\tau_1(\sum_j q_j^m - \tau_2)]}{1+\exp[-\tau_1(\sum_j q_j^m - \tau_2)]}$, ce qui donne :

$$\log p(w^m \mid \boldsymbol{q}^m, \boldsymbol{\tau}) = C_{5,w^m} + (1 - w^m)\left[-\tau_1\left(\sum_j q_j^m - \tau_2\right)\right] \tag{D.28}$$

$$E_{\widetilde{p}_{\boldsymbol{q}^m}}[\log p(w^m \mid \boldsymbol{q}^m, \boldsymbol{\tau})] = C'_{5,w^m} + (1 - w^m)\left[-\tau_1\left(\sum_j \widetilde{p}_{q_j^m}(1) - \tau_2\right)\right] \tag{D.29}$$

où $C'_{5,w^m} = E_{\widetilde{p}_{\boldsymbol{q}^m}}[C_{5,w^m}]$.

Une fois les calculs effectués, les probabilités *a posteriori* $\widetilde{p}_{w^m}(0)$ et $\widetilde{p}_{w^m}(1)$ sont données par :

$$\widetilde{p}_{w^m}(0) = \frac{f(w^m = 0)}{f(w^m = 0) + f(w^m = 1)} \tag{D.30}$$

$$\widetilde{p}_{w^m}(1) = 1 - \widetilde{p}_{w^m}(0) = \frac{f(w^m = 1)}{f(w^m = 0) + f(w^m = 1)} \tag{D.31}$$

Afin de calculer les termes $f(w^m = 0)$ et $f(w^m = 1)$, on remplace $w^m = 0$ et 1 dans les équations (D.24), (D.26) et (D.29).

$$f(w^m = 0) = \exp\left(-\tau_1\left[\sum_j \widetilde{p}_{q_j^m}(1) - \tau_2\right]\right) \tag{D.32}$$

$$\begin{aligned} f(w^m = 1) = \exp\Bigg(\sum_j \Bigg\{ & \widetilde{p}_{q_j^m}(1)\left[\frac{V_{\boldsymbol{a}_j^{m,m}}}{2}\left(\frac{1}{v_0^m} - \frac{1}{v_1^m}\right) + \log\frac{\mathcal{N}(m_{a_j^m}; \mu_1^m, v_1^m)}{\mathcal{N}(m_{a_j^m}; \mu_0^m, v_0^m)}\right] \\ & - \frac{m_{a_j^m}^2 + V_{\boldsymbol{a}_j^{m,m}}}{2}\left[\boldsymbol{m}_H^{\mathrm{t}}\boldsymbol{X}^{m\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{X}^m\boldsymbol{m}_H + tr\left(\boldsymbol{\Sigma}_H\boldsymbol{X}^{m\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{X}^m\right)\right] \\ & - \sum_{m' \neq m}\left(m_{a_j^m}m_{a_j^{m'}} + V_{\boldsymbol{a}_j^{m,m'}}\right)\widetilde{p}_{w^{m'}}(1)\Big[\boldsymbol{m}_H^{\mathrm{t}}\boldsymbol{X}^{m\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{X}^{m'}\boldsymbol{m}_H \\ & + tr\left(\boldsymbol{\Sigma}_H\boldsymbol{X}^{m\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{X}^{m'}\right)\Big] + m_{a_j^m}\boldsymbol{m}_H^{\mathrm{t}}\boldsymbol{X}^{m\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j - \boldsymbol{P}\boldsymbol{\ell}_j)\Bigg\}\Bigg) \end{aligned} \tag{D.33}$$

♣ **E-Q**

$$\begin{aligned}\widetilde{p}_{q_j^m}(q_j^m) &\propto \exp\left(E_{\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{\boldsymbol{A}},\widetilde{p}_{\boldsymbol{w}},\widetilde{p}_{\boldsymbol{q}_{\setminus j}^m},\widetilde{p}_{\boldsymbol{Q}^{\setminus m}}}\left[\log p(q_j^m \mid \boldsymbol{q}_{\setminus j}^m, \boldsymbol{Q}^{\setminus m}, \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{h}, \boldsymbol{w}, \boldsymbol{\theta})\right]\right) \\ &\propto f(q_j^m) \end{aligned} \tag{D.34}$$

où on note

$$\begin{aligned} f(q_j^m) = \exp\Big(&E_{\widetilde{p}_{a_j^m},\widetilde{p}_{w^m}}\left[\log p(a_j^m \mid w^m, q_j^m, \boldsymbol{\theta}_m)\right] \\ &+ E_{\widetilde{p}_{w^m},\widetilde{p}_{\boldsymbol{q}_{\setminus j}^m}}\left[\log p(w^m \mid q_j^m, \boldsymbol{q}_{\setminus j}^m, \boldsymbol{\tau})\right] \\ &+ E_{\widetilde{p}_{\boldsymbol{q}_{\setminus j}^m}}\left[\log p(\boldsymbol{q}^m \mid \xi_m)\right]\Big) \end{aligned} \tag{D.35}$$

1) Calcul de $E_{\widetilde{p}_{a_j^m},\widetilde{p}_{w^m}}\left[\log p(a_j^m \mid w^m, q_j^m, \boldsymbol{\theta}_m)\right]$:

$$\begin{aligned}\log p(a_j^m \mid w^m, q_j^m, \boldsymbol{\theta}_m) &= C_{1,q_j^m} - q_j^m w^m \left[\log \mathcal{N}(\mu_0^m, v_0^m) - \log \mathcal{N}(\mu_1^m, v_1^m)\right] \\ &= C_{1,q_j^m} + \frac{1}{2} q_j^m w^m \Bigg(\log(2\pi v_0^m) + \frac{(a_j^m)^2 - 2a_j^m\mu_0^m + (\mu_0^m)^2}{v_0^m} \\ &\quad - \log(2\pi v_1^m) - \frac{(a_j^m)^2 - 2a_j^m\mu_1^m + (\mu_1^m)^2}{v_1^m}\Bigg) \end{aligned} \tag{D.36}$$

$$\begin{aligned} &E_{\widetilde{p}_{a_j^m},\widetilde{p}_{w^m}}\left[\log p(a_j^m \mid w^m, q_j^m, \boldsymbol{\theta}_m)\right] \\ &= C'_{1,q_j^m} + q_j^m \widetilde{p}_{w^m}(1)\left\{\frac{V_{\boldsymbol{a}_j^{m,m}}}{2}\left(\frac{1}{v_0^m} - \frac{1}{v_1^m}\right) + \log\frac{\mathcal{N}(m_{a_j^m}; \mu_1^m, v_1^m)}{\mathcal{N}(m_{a_j^m}; \mu_0^m, v_0^m)}\right\} \end{aligned} \tag{D.37}$$

2) Calcul de $E_{\widetilde{p}_{w^m},\widetilde{p}_{\boldsymbol{q}_{\setminus j}^m}}\left[\log p(w^m \mid q_j^m, \boldsymbol{q}_{\setminus j}^m, \boldsymbol{\tau})\right]$:

$$\begin{aligned} &\log p(w^m \mid q_j^m, \boldsymbol{q}_{\setminus j}^m, \boldsymbol{\tau}) \\ &\quad = w^m \log p_m + (1 - w^m)\log(1 - p_m) \\ &\quad = C_{2,q_j^m} - (1 - w^m)\tau_1 q_j^m - \log\left(1 + \exp\left[-\tau_1\left(\sum_j q_j^m - \tau_2\right)\right]\right) \end{aligned} \tag{D.38}$$

$$E_{\widetilde{p}_{w^m},\widetilde{p}_{\boldsymbol{q}^m_{\setminus j}}}\left[\log p(w^m \mid q^m_j, \boldsymbol{q}^m_{\setminus j}, \boldsymbol{\tau})\right]$$

$$= C'_{2,q^m_j} - \widetilde{p}_{w^m}(0)\tau_1 q^m_j - E_{\widetilde{p}_{\boldsymbol{q}^m_{\setminus j}}}\left[\log\left(1+\exp\left[-\tau_1\left(\sum_j q^m_j - \tau_2\right)\right]\right)\right] \quad \text{(D.39)}$$

Vu la difficulté du calcul $E_{\widetilde{p}_{\boldsymbol{q}^m_{\setminus j}}}(q^m_j) = E_{\widetilde{p}_{\boldsymbol{q}^m_{\setminus j}}}[\log(1+\exp[-\tau_1(\sum_j q^m_j - \tau_2)])]$ on propose de l'effectuer à travers une étape MC (Monte Carlo). Chaque simulation consiste en :

- La simulation des étiquettes $\boldsymbol{q}^m_{\setminus j}$ à partir de leurs probabilités $\widetilde{p}_{\boldsymbol{q}^m_{\setminus j}}$.
- Le calcul du terme $A_s(q^m_j) = \log(1+\exp[-\tau_1(\sum_{j'\neq j} q^m_{j'}(s) + q^m_j - \tau_2)])$
- Le calcul de la moyenne empirique $E_{\widetilde{p}_{\boldsymbol{q}^m_{\setminus j}}}(q^m_j) = \frac{1}{S}\sum_{s=1}^{S} A_s$.

Ce qui nous permet d'écrire :

$$E_{\widetilde{p}_{w^m},\widetilde{p}_{\boldsymbol{q}^m_{\setminus j}}}\left[\log p(w^m \mid q^m_j, \boldsymbol{q}^m_{\setminus j}, \boldsymbol{\tau})\right] = C'_{2,q^m_j} - \widetilde{p}_{w^m}(0)\tau_1 q^m_j - E_{\widetilde{p}_{\boldsymbol{q}^m_{\setminus j}}}(q^m_j) \quad \text{(D.40)}$$

3) Calcul de $E_{\widetilde{p}_{\boldsymbol{q}^m_{\setminus j}}}[\log p(\boldsymbol{q}^m \mid \xi_m)]$:

D'après la définition de la distribution *a priori* de la variable $\boldsymbol{Q}$ (équation (II.29))

$$\begin{aligned}
\log p(\boldsymbol{q}^m \mid \xi_m) &= \log p(q^m_j \mid \boldsymbol{q}^m_{\setminus j}, \xi_m)\, p(\boldsymbol{q}^m_{\setminus j} \mid \xi_m) \\
&= \log Z(\xi_m)^{-1} \exp\left(\xi_m U(q^m_j)\right) + \log p(\boldsymbol{q}^m_{\setminus j} \mid \xi_m) \\
&= \xi_m U(q^m_j) + C_{3,q^m_j} \\
&= \xi_m \sum_{k\in N(j)} I(q^m_j = q^m_k) + C_{3,q^m_j} \qquad \text{(D.41)}
\end{aligned}$$

où $N(j)$ désigne le voisinage de j et où $C_{3,q^m_j} = \log Z(\xi_m)^{-1} + \log p(\boldsymbol{q}^m_{\setminus j} \mid \xi_m)$.

$$\begin{aligned}
E_{\widetilde{p}_{\boldsymbol{q}^m_{\setminus j}}}[\log p(\boldsymbol{q}^m \mid \xi_m)] &= C'_{3,q^m_j} + \xi_m \sum_{k\in N(j)} E_{\widetilde{p}_{q^m_k}}\left[I(q^m_j = q^m_k)\right] \\
&= C'_{3,q^m_j} + \xi_m \sum_{k\in N(j)} \sum_{i=0}^{1} \left[I(q^m_j = i)\widetilde{p}_{q^m_k}(i)\right] \qquad \text{(D.42)}
\end{aligned}$$

avec $C'_{3,q^m_j} = E_{\widetilde{p}_{\boldsymbol{q}^m_{\setminus j}}}\left[C_{3,q^m_j}\right]$

Comme dans le cas de $\boldsymbol{w}$ les probabilités $\widetilde{p}_{q^m_j}(0)$ et $\widetilde{p}_{q^m_j}(1)$ *a posteriori* de la

variable binaire q_j^m sont :

$$\widetilde{p}_{q_j^m}(0) = \frac{f(q_j^m = 0)}{f(q_j^m = 0) + f(q_j^m = 1)} \tag{D.43}$$

$$\widetilde{p}_{q_j^m}(1) = 1 - \widetilde{p}_{q_j^m}(0) = \frac{f(q_j^m = 1)}{f(q_j^m = 0) + f(q_j^m = 1)} \tag{D.44}$$

où le calcul des quantités $f(q_j^m = 0)$ et $f(q_j^m = 1)$ se fait en remplaçant $q_j^m = 0$ et 1 respectivement dans les équations (D.37), (D.40) et (D.42) on trouve :

$$f(q_j^m = 0) = \exp\left(-E_{\widetilde{p}_{\boldsymbol{q}_{\setminus j}^m}}\left(q_j^m = 0\right) + \xi_m \sum_{k \in N(j)} \widetilde{p}_{q_k^m}(0)\right) \tag{D.45}$$

$$f(q_j^m = 1) = \exp\left(\widetilde{p}_{w^m}(1)\left\{\frac{V_{\boldsymbol{a}_j^{m,m}}}{2}\left(\frac{1}{v_0^m} - \frac{1}{v_1^m}\right) + \log\frac{\mathcal{N}(m_{a_j^m}; \mu_1^m, v_1^m)}{\mathcal{N}(m_{a_j^m}; \mu_0^m, v_0^m)}\right\}\right.$$
$$\left. -\tau_1\, \widetilde{p}_{w^m}(0) - E_{\widetilde{p}_{\boldsymbol{q}_{\setminus j}^m}}\left(q_j^m = 1\right) + \xi_m \sum_{k \in N(j)} \widetilde{p}_{q_k^m}(1)\right) \tag{D.46}$$

♣ **M-$\boldsymbol{\theta}_A$**

$$\hat{\boldsymbol{\theta}}_A = \arg\max_{\boldsymbol{\theta}_A} E_{\widetilde{p}_{\boldsymbol{A}}, \widetilde{p}_{\boldsymbol{w}}, \widetilde{p}_{\boldsymbol{Q}}}\left[\log p(\boldsymbol{A} \,|\, \boldsymbol{w}, \boldsymbol{Q}, \boldsymbol{\theta}_A)\right] \tag{D.47}$$

Selon l'équation (D.16) et en moyennant aussi la constante $C_{3,\boldsymbol{A}}$ (qui n'est plus constant en $\boldsymbol{\theta}_A$) par rapport à $\boldsymbol{Q}$ et $\boldsymbol{w}$ on obtient :

$$E_{\widetilde{p}_{\boldsymbol{Q}}, \widetilde{p}_{\boldsymbol{w}}}\left[\log p(\boldsymbol{A} \,|\, \boldsymbol{w}, \boldsymbol{Q}, \boldsymbol{\theta}_A)\right] = -\frac{1}{2}\sum_m \sum_j \left\{\left(1 - \widetilde{p}_{q_j^m}(1)\widetilde{p}_{w^m}(1)\right)\right.$$
$$\left.\times\left(\log(2\pi v_0^m) + \frac{(a_j^m - \mu_0^m)^2}{v_0^m}\right) + \widetilde{p}_{q_j^m}(1)\widetilde{p}_{w^m}(1)\left(\log(2\pi v_1^m) + \frac{(a_j^m - \mu_1^m)^2}{v_1^m}\right)\right\} \tag{D.48}$$

En moyennant le terme précédent par rapport à la variable $\boldsymbol{A}$ on obtient :

$$E_{\widetilde{p}_{\boldsymbol{A}}}\left[E_{\widetilde{p}_{\boldsymbol{Q}}, \widetilde{p}_{\boldsymbol{w}}}\left[\log p(\boldsymbol{A} \,|\, \boldsymbol{w}, \boldsymbol{Q}, \boldsymbol{\theta}_A)\right]\right] = -\frac{1}{2}\sum_m \left[f(\mu_0^m, v_0^m) + f(\mu_1^m, v_1^m)\right] \tag{D.49}$$

avec

$$f(\mu_0^m, v_0^m) = \sum_j \left(1 - \widetilde{p}_{q_j^m}(1)\widetilde{p}_{w^m}(1)\right)\left(\log(2\pi v_0^m) + \frac{(m_{a_j^m} - \mu_0^m)^2 + V_{\boldsymbol{a}_j^{m,m}}}{v_0^m}\right), \tag{D.50}$$

et

$$f(\mu_1^m,v_1^m)=\sum_j \widetilde{p}_{q_j^m}(1)\widetilde{p}_{w^m}(1)\left(\log(2\pi v_1^m)+\frac{(m_{a_j^m}-\mu_1^m)^2+V_{\boldsymbol{a}_j^{m,m}}}{v_1^m}\right) \tag{D.51}$$

on peut donc calculer chacun des paramètres $\hat{v}_0^m, \hat{v}_1^m, \hat{\mu}_1^m$ à partir des équations suivantes :

$$\hat{v}_0^m = \arg\max_{v_0^m} f(\mu_0^m, v_0^m) \tag{D.52}$$

$$\hat{v}_1^m = \arg\max_{v_1^m} f(\mu_1^m, v_1^m) \tag{D.53}$$

$$\hat{\mu}_1^m = \arg\max_{\mu_1^m} f(\mu_1^m, v_1^m) \tag{D.54}$$

1) Calcul de $\hat{v}_0^m$:

$$\frac{\partial f(\mu_0^m,\hat{v}_0^m)}{\partial \hat{v}_0^m}=\sum_j\left(1-\widetilde{p}_{q_j^m}(1)\widetilde{p}_{w^m}(1)\right)\left(1-\frac{(m_{a_j^m}-\mu_0^m)^2+V_{\boldsymbol{a}_j^{m,m}}}{\hat{v}_0^m}\right)=0$$

$$\Rightarrow \hat{v}_0^m=\frac{\sum_j\left(1-\widetilde{p}_{q_j^m}(1)\widetilde{p}_{w^m}(1)\right)\left[(m_{a_j^m}-\mu_0^m)^2+V_{\boldsymbol{a}_j^{m,m}}\right]}{\sum_j\left(1-\widetilde{p}_{q_j^m}(1)\widetilde{p}_{w^m}(1)\right)} \tag{D.55}$$

2) Calcul de $\hat{v}_1^m$:

$$\frac{\partial f(\mu_1^m,\hat{v}_1^m)}{\partial \hat{v}_1^m}=\sum_j\widetilde{p}_{q_j^m}(1)\widetilde{p}_{w^m}(1)\left(1-\frac{(m_{a_j^m}-\mu_1^m)^2+V_{\boldsymbol{a}_j^{m,m}}}{\hat{v}_1^m}\right)=0$$

$$\Rightarrow \hat{v}_1^m=\frac{\sum_j\widetilde{p}_{q_j^m}(1)\left[(m_{a_j^m}-\mu_1^m)^2+V_{\boldsymbol{a}_j^{m,m}}\right]}{\sum_j\widetilde{p}_{q_j^m}(1)} \tag{D.56}$$

3) Calcul de $\hat{\mu}_1^m$:

$$\frac{\partial f(\hat{\mu}_1^m,v_1^m)}{\partial \hat{\mu}_1^m}=\sum_j\widetilde{p}_{q_j^m}(1)\widetilde{p}_{w^m}(1)\left(\frac{-2m_{a_j^m}+2\hat{\mu}_1^m}{v_1^m}\right)=0 \tag{D.57}$$

$$\Rightarrow \hat{\mu}_1^m=\frac{\sum_j\widetilde{p}_{q_j^m}(1)m_{a_j^m}}{\sum_j\widetilde{p}_{q_j^m}(1)} \tag{D.58}$$

♣ **M-v_h**

$$\hat{v}_h = \arg\max_{v_h} f(v_h) = \arg\max_{v_h} \left\{ E_{\widetilde{p}_{\boldsymbol{h}}} \left[\log p(\boldsymbol{h} \,|\, v_h) \right] + \log p(v_h \,|\, \lambda_{v_h}) \right\} \quad \text{(D.59)}$$

Où l'*a priori* $p(v_h \,|\, \lambda_{v_h}) = \lambda_{v_h} \exp(-\lambda_{v_h} v_h)$ permet d'améliorer l'estimation de v_h [Chaari 2013].

$$E_{\widetilde{p}_{\boldsymbol{h}}} \left[\log p(\boldsymbol{h} \,|\, v_h) \right] = C_{v_h} - \frac{D-1}{2} \log v_h - \frac{\boldsymbol{m}_H^{\mathrm{t}} \boldsymbol{R}^{-1} \boldsymbol{m}_H + tr\left(\boldsymbol{\Sigma}_H \boldsymbol{R}^{-1}\right)}{2 v_h}$$

$$\frac{\partial f(v_h)}{\partial v_h} = -\frac{D-1}{2 v_h} + \frac{\boldsymbol{m}_H^{\mathrm{t}} \boldsymbol{R}^{-1} \boldsymbol{m}_H + tr\left(\boldsymbol{\Sigma}_H \boldsymbol{R}^{-1}\right)}{2 v_h^2} - \lambda_{v_h} = 0$$

$$\Rightarrow \quad v_h = \frac{(1-D) + \sqrt{(D-1)^2 + 8\lambda_{v_h} \mathrm{tr}\left[\left(\boldsymbol{m}_H \boldsymbol{m}_H^{\mathrm{t}} + \boldsymbol{\Sigma}_H\right) \boldsymbol{R}^{-1}\right]}}{4\lambda_{v_h}} \quad \text{(D.60)}$$

♣ **M-$\boldsymbol{\sigma}^2$**

$$\hat{\boldsymbol{\sigma}}^2 = \arg\max_{\boldsymbol{\sigma}^2} E_{\widetilde{p}_{\boldsymbol{h}}, \widetilde{p}_{\boldsymbol{A}}, \widetilde{p}_{\boldsymbol{w}}} \left[\log p\left(\boldsymbol{Y} \,|\, \boldsymbol{A}, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{L}, \boldsymbol{\sigma}^2 \right) \right] \quad \text{(D.61)}$$

Etant donnée l'indépendance entre $\{\boldsymbol{y}_j \,|\, \boldsymbol{a}_j, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{\ell}_j, \sigma_j^2\}_{j=1:J}$ on peut écrire :

$$\hat{\sigma_j}^2 = \arg\max_{\sigma_j^2} E_{\widetilde{p}_{\boldsymbol{h}}, \widetilde{p}_{\boldsymbol{a}_j}, \widetilde{p}_{\boldsymbol{w}}} \left[\log p\left(\boldsymbol{y}_j \,|\, \boldsymbol{a}_j, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{\ell}_j, \sigma_j^2 \right) \right] \quad \text{(D.62)}$$

A partir de la définition de la vraisemblance (équation (II.23)) on peut calculer :

$$E_{\widetilde{p}_{\boldsymbol{h}}, \widetilde{p}_{\boldsymbol{a}_j}, \widetilde{p}_{\boldsymbol{w}}} \left[\log p\left(\boldsymbol{y}_j \,|\, \boldsymbol{a}_j, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{\ell}_j, \sigma_j^2 \right) \right] = C_{\sigma_j^2} - \frac{N}{2} \log \sigma_j^2 - \frac{E_{\widetilde{p}_{\boldsymbol{h}}, \widetilde{p}_{\boldsymbol{a}_j}, \widetilde{p}_{\boldsymbol{w}}} \left[\widetilde{\boldsymbol{y}}_j^{\mathrm{t}} \boldsymbol{\Gamma}_j \widetilde{\boldsymbol{y}}_j \right]}{2}$$

- **Calcul de $E_{\widetilde{p}_{\boldsymbol{h}}, \widetilde{p}_{\boldsymbol{a}_j}, \widetilde{p}_{\boldsymbol{w}}} \left[\widetilde{\boldsymbol{y}}_j^{\mathrm{t}} \boldsymbol{\Gamma}_j \widetilde{\boldsymbol{y}}_j \right]$:**

Selon l'équation (D.11) et en tenant compte des termes de $C_{2,\boldsymbol{A}}$ qui dépendent de σ_j^2 on a :

$$\begin{aligned} E_{\widetilde{p}_{\boldsymbol{h}}, \widetilde{p}_{\boldsymbol{w}}} \left[\log p\left(\boldsymbol{Y} \,|\, \boldsymbol{A}, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{L}, \boldsymbol{\sigma}^2 \right) \right] &= C_{\sigma_j^2} - \frac{N}{2} \sum_j \log \sigma_j^2 \\ &- \frac{1}{2} \sum_j \Big\{ \boldsymbol{a}_j^{\mathrm{t}} E_{\widetilde{p}_{\boldsymbol{h}}, \widetilde{p}_{\boldsymbol{w}}} \left[\boldsymbol{G}^{\mathrm{t}} \boldsymbol{\Gamma}_j \boldsymbol{G} \right] \boldsymbol{a}_j - 2 \boldsymbol{a}_j^{\mathrm{t}} E_{\widetilde{p}_{\boldsymbol{h}}, \widetilde{p}_{\boldsymbol{w}}} \left[\boldsymbol{G}^{\mathrm{t}} \boldsymbol{\Gamma}_j (\boldsymbol{y}_j - \boldsymbol{P} \boldsymbol{\ell}_j) \right] \\ &+ (\boldsymbol{y}_j - \boldsymbol{P} \boldsymbol{\ell}_j)^{\mathrm{t}} \boldsymbol{\Gamma}_j (\boldsymbol{y}_j - \boldsymbol{P} \boldsymbol{\ell}_j) \Big\} \end{aligned} \quad \text{(D.63)}$$

$$
\begin{aligned}
E_{\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{\boldsymbol{w}},\widetilde{p}_{\boldsymbol{A}}}\left[\log p\left(\boldsymbol{Y}\mid\boldsymbol{A},\boldsymbol{w},\boldsymbol{h},\boldsymbol{L},\boldsymbol{\sigma}^2\right)\right] &= C_{\sigma_j^2} - \frac{N}{2}\sum_j \log\sigma_j^2 \\
&- \frac{1}{2}\sum_j \left[\boldsymbol{m}_{\boldsymbol{a}_j}^{\mathrm{t}} E_{\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{\boldsymbol{w}}}\left[\boldsymbol{G}^{\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{G}\right]\boldsymbol{m}_{\boldsymbol{a}_j} + tr\left(\boldsymbol{V}_{\boldsymbol{a}_j} E_{\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{\boldsymbol{w}}}\left[\boldsymbol{G}^{\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{G}\right]\right)\right. \\
&\left. -2\,\boldsymbol{m}_{\boldsymbol{a}_j}^{\mathrm{t}} E_{\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{\boldsymbol{w}}}\left[\boldsymbol{G}^{\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j-\boldsymbol{P}\boldsymbol{\ell}_j)\right] + (\boldsymbol{y}_j-\boldsymbol{P}\boldsymbol{\ell}_j)^{\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j-\boldsymbol{P}\boldsymbol{\ell}_j)\right]
\end{aligned}
\tag{D.64}
$$

On sait que $\boldsymbol{\Gamma}_j = \frac{1}{\sigma_j^2}\boldsymbol{\Lambda}_j$, on peut donc écrire :

$$
\begin{aligned}
&E_{\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{\boldsymbol{w}},\widetilde{p}_{\boldsymbol{A}}}\left[\log p\left(\boldsymbol{Y}\mid\boldsymbol{A},\boldsymbol{w},\boldsymbol{h},\boldsymbol{L},\boldsymbol{\sigma}^2\right)\right] \\
&= C_{\sigma_j^2} + \sum_j \left\{-\frac{N}{2}\log\sigma_j^2 - \frac{1}{2\sigma_j^2}\boxed{\left[\boldsymbol{m}_{\boldsymbol{a}_j}^{\mathrm{t}} E_{\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{\boldsymbol{w}}}\left[\boldsymbol{G}^{\mathrm{t}}\boldsymbol{\Lambda}_j\boldsymbol{G}\right]\boldsymbol{m}_{\boldsymbol{a}_j}\right.}\right. \\
&\boxed{+tr\left(\boldsymbol{V}_{\boldsymbol{a}_j} E_{\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{\boldsymbol{w}}}\left[\boldsymbol{G}^{\mathrm{t}}\boldsymbol{\Lambda}_j\boldsymbol{G}\right]\right) - 2\boldsymbol{m}_{\boldsymbol{a}_j}^{\mathrm{t}} E_{\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{\boldsymbol{w}}}\left[\boldsymbol{G}^{\mathrm{t}}\boldsymbol{\Lambda}_j(\boldsymbol{y}_j-\boldsymbol{P}\boldsymbol{\ell}_j)\right]} \\
&\left.\boxed{\left.+(\boldsymbol{y}_j-\boldsymbol{P}\boldsymbol{\ell}_j)^{\mathrm{t}}\boldsymbol{\Lambda}_j(\boldsymbol{y}_j-\boldsymbol{P}\boldsymbol{\ell}_j)\right]}\right\}
\end{aligned}
\tag{D.65}
$$

où $E_{\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{\boldsymbol{w}}}\left[\boldsymbol{G}^{\mathrm{t}}\boldsymbol{\Lambda}_j\boldsymbol{G}\right]$ et $E_{\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{\boldsymbol{w}}}\left[\boldsymbol{G}^{\mathrm{t}}\boldsymbol{\Lambda}_j(\boldsymbol{y}_j-\boldsymbol{P}\boldsymbol{\ell}_j)\right]$ donnés par les équations (D.13) et (D.14) en remplaçant $\boldsymbol{\Gamma}_j$ par $\boldsymbol{\Lambda}_j$. Notons V le terme entouré en bleu.

$$
\frac{\partial E_{\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{\boldsymbol{w}},\widetilde{p}_{\boldsymbol{a}_j}}\left[\log p\left(\boldsymbol{y}_j\mid\boldsymbol{a}_j,\boldsymbol{w},\boldsymbol{h},\boldsymbol{\ell}_j,\hat{\sigma}_j^2\right)\right]}{\partial\hat{\sigma}_j{}^2} = -N + \frac{1}{\hat{\sigma}_j^2}V = 0 \Rightarrow \hat{\sigma}_j^2 = \frac{V}{N}
\tag{D.66}
$$

♣ **M-$\boldsymbol{L}$**

$$
\hat{\boldsymbol{L}} = \arg\max_{\boldsymbol{L}} E_{\widetilde{p}_{\boldsymbol{A}},\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{\boldsymbol{w}}}\left[\log p(\boldsymbol{Y}\mid\boldsymbol{A},\boldsymbol{w},\boldsymbol{h},\boldsymbol{L},\sigma_j^2)\right]
\tag{D.67}
$$

Pour la même raison d'indépendance entre $\{\boldsymbol{y}_j\mid\boldsymbol{a}_j,\boldsymbol{w},\boldsymbol{h},\boldsymbol{\ell}_j,\sigma_j^2\}_{j=1:J}$ on peut calculer :

$$
\hat{\boldsymbol{\ell}}_j = \arg\max_{\boldsymbol{\ell}_j} E_{\widetilde{p}_{\boldsymbol{A}},\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{\boldsymbol{w}}}\left[\log p(\boldsymbol{y}_j\mid\boldsymbol{a}_j,\boldsymbol{w},\boldsymbol{h},\boldsymbol{\ell}_j,\sigma_j^2)\right]
\tag{D.68}
$$

$\log p(\boldsymbol{y}_j\mid\boldsymbol{a}_j,\boldsymbol{w},\boldsymbol{h},\boldsymbol{\ell}_j,\sigma_j^2)$

$$
\begin{aligned}
&= C_{\boldsymbol{\ell}_j} + \left(\boldsymbol{y}_j - \sum_m a_j^m w^m \boldsymbol{X}^m\boldsymbol{h} - \boldsymbol{P}\boldsymbol{\ell}_j\right)^{\mathrm{t}}\boldsymbol{\Gamma}_j\left(\boldsymbol{y}_j - \sum_m a_j^m w^m \boldsymbol{X}^m\boldsymbol{h} - \boldsymbol{P}\boldsymbol{\ell}_j\right) \\
&= C_{\boldsymbol{\ell}_j} + \bar{\boldsymbol{y}}_j{}^{\mathrm{t}}\boldsymbol{\Gamma}_j\bar{\boldsymbol{y}}_j - \bar{\boldsymbol{y}}_j{}^{\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{P}\boldsymbol{\ell}_j - \boldsymbol{\ell}_j^{\mathrm{t}}\boldsymbol{P}^{\mathrm{t}}\boldsymbol{\Gamma}_j\bar{\boldsymbol{y}}_j + \boldsymbol{\ell}_j^{\mathrm{t}}\boldsymbol{P}^{\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{P}\boldsymbol{\ell}_j
\end{aligned}
\tag{D.69}
$$

où $\bar{\boldsymbol{y}}_j = \boldsymbol{y}_j - \sum_m a_j^m w^m \boldsymbol{X}^m\boldsymbol{h}$.

$$E_{\widetilde{p}_{\boldsymbol{A}},\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{\boldsymbol{w}}}\left[\log p(\boldsymbol{y}_j \mid \boldsymbol{a}_j, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{\ell}_j, \sigma_j^2)\right]$$

$$= C_{\boldsymbol{\ell}_j} + E_{\widetilde{p}_{\boldsymbol{A}},\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{\boldsymbol{w}}}\left[\bar{\boldsymbol{y}}_j{}^{\mathrm{t}}\boldsymbol{\Gamma}_j\bar{\boldsymbol{y}}_j\right] - 2\,\boldsymbol{\ell}_j^{\mathrm{t}}\boldsymbol{P}^{\mathrm{t}}\boldsymbol{\Gamma}_j E_{\widetilde{p}_{\boldsymbol{A}},\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{\boldsymbol{w}}}[\bar{\boldsymbol{y}}_j] + \boldsymbol{\ell}_j^{\mathrm{t}}\boldsymbol{P}^{\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{P}\boldsymbol{\ell}_j \tag{D.70}$$

En dérivant l'équation précédente par rapport à $\boldsymbol{\ell}_j$ on obtient :

$$\frac{\partial(.)}{\partial\boldsymbol{\ell}_j} = -2\boldsymbol{P}^{\mathrm{t}}\boldsymbol{\Gamma}_j E_{\widetilde{p}_{\boldsymbol{A}},\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{\boldsymbol{w}}}[\bar{\boldsymbol{y}}_j] + 2\boldsymbol{P}^{\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{P}\boldsymbol{\ell}_j = \mathrm{Vect}\,0 \tag{D.71}$$

$$\begin{aligned}\Rightarrow \quad \hat{\boldsymbol{\ell}}_j &= (\boldsymbol{P}^{\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{P})^{-1}\boldsymbol{P}^{\mathrm{t}}E_{\widetilde{p}_{\boldsymbol{A}},\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{\boldsymbol{w}}}[\bar{\boldsymbol{y}}_j] \\ &= (\boldsymbol{P}^{\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{P})^{-1}\boldsymbol{P}^{\mathrm{t}}\left(\boldsymbol{y}_j - \sum_m m_{a_j^m}\widetilde{p}_{w^m}(1)\boldsymbol{X}^m\boldsymbol{m}_h\right)\end{aligned} \tag{D.72}$$

♣ **M-$\boldsymbol{\xi}$**

$$\begin{aligned}\hat{\xi}_m &= \arg\max_{\xi_m} E_{\widetilde{p}_{\boldsymbol{q}^m}}[\log p(\boldsymbol{q}^m \mid \xi_m)] + \log p(\xi_m \mid \lambda_{\xi_m}) \\ &= \arg\max_{\xi_m}\left\{-\log Z(\xi_m) + \xi_m\left[E_{\widetilde{p}_{\boldsymbol{q}^m}}[U(\boldsymbol{q}^m)] - \lambda_{\xi_m}\right] + C_{\xi_m}\right\} \\ &= \arg\max_{\xi_m} f(\xi_m)\end{aligned} \tag{D.73}$$

La maximisation de $f(\xi_m)$ nécessite le calcul de sa dérivée par rapport à ξ_m :

$$\frac{df(\xi_m)}{d\xi_m} = -\frac{d\log Z(\xi_m)}{d\xi_m} + E_{\widetilde{p}_{\boldsymbol{q}^m}}[U(\boldsymbol{q}^m)] - \lambda_{\xi_m} \tag{D.74}$$

où

$$Z(\xi_m) = \sum_{\boldsymbol{q}^m}\exp(\xi_m U(\boldsymbol{q}^m)) \tag{D.75}$$

$$U(\boldsymbol{q}^m) = \sum_{j\sim k}\left\langle \boldsymbol{q}_j^m, \boldsymbol{q}_k^m\right\rangle = \frac{1}{2}\sum_j\sum_{k\in N_j}\left\langle \boldsymbol{q}_j^m, \boldsymbol{q}_k^m\right\rangle = \frac{1}{2}\sum_j\left\langle \boldsymbol{q}_j^m, \sum_{k\in N_j}\boldsymbol{q}_k^m\right\rangle \tag{D.76}$$

Rappelons que N_j représente les voxels k voisins de j. Pour des raisons de simplicité, $\boldsymbol{q}_j^m$ n'est plus un scalaire mais un vecteur de dimension $I = 2$, le nombre de classes, avec $\boldsymbol{q}_j^m = [e_1 e_2 ... e_I]^{\mathrm{t}}$ où $e_i = 1$ si q_j^m appartient à la $i$ème classe et 0 sinon.

Il existe plusieurs possibilités pour calculer le gradient précédent (équation (D.74)) comme nous le montrons ci-dessous. La première consiste en l'appro-

ximation champ moyen [Celeux 2003] de $Z(\xi_m)$ avant de dériver contrairement à la deuxième qui effectue la dérivée avant l'approximation. La deuxième méthode est généralement préférable étant donné qu'elle diminue l'erreur d'approximation commise, cependant les deux mènent au même résultat dans notre cas, comme nous le verrons par la suite. La troisième et quatrième méthodes sont néanmoins d'une approximation plus forte et donc moins bonnes comparées aux deux premières. Une fois le gradient obtenu, par l'une des méthodes, l'estimation de ξ_m est effectuée à l'aide de l'algorithme de descente du gradient (annexe F.5).

Méthode 1. L'approximation champ moyen consiste à remplacer les voisins aléatoires $\boldsymbol{q}_k^m$ par des quantités fixées $\widetilde{\boldsymbol{q}}_k^m$ permettant ainsi l'approximation de $U(\boldsymbol{q}^m)$ par $U_{mf}(\boldsymbol{q}^m)$ [Forbes 2003] :

$$U_{mf}(\boldsymbol{q}^m) = \sum_j \sum_{k \in N_j} \left\langle \boldsymbol{q}_j^m, \widetilde{\boldsymbol{q}}_k^m \right\rangle = \sum_j \left\langle \boldsymbol{q}_j^m, \sum_{k \in N_j} \widetilde{\boldsymbol{q}}_k^m \right\rangle \tag{D.77}$$

d'où la simplification de la constante

$$\begin{aligned} Z_{mf}(\xi_m) &= \sum_{\boldsymbol{q}^m} \exp(\xi_m U_{mf}(\boldsymbol{q}^m)) = \sum_{\boldsymbol{q}^m} \exp\left(\xi_m \sum_j \left\langle \boldsymbol{q}_j^m, \sum_{k \in N_j} \widetilde{\boldsymbol{q}}_k^m \right\rangle\right) \\ &= \sum_{\boldsymbol{q}^m} \prod_j \exp\left(\xi_m \left\langle \boldsymbol{q}_j^m, \sum_{k \in N_j} \widetilde{\boldsymbol{q}}_k^m \right\rangle\right) = \prod_j \sum_{\boldsymbol{q}_j^m} \exp\left(\xi_m \left\langle \boldsymbol{q}_j^m, \sum_{k \in N_j} \widetilde{\boldsymbol{q}}_k^m \right\rangle\right) \\ &= \prod_j \sum_{i \in \{0,1\}} \exp\left(\xi_m \sum_{k \in N_j} \widetilde{\boldsymbol{q}}_k^m(i)\right) \end{aligned} \tag{D.78}$$

Ceci nous permet d'approximer $Z(\xi_m)$ comme suit :

$$\begin{aligned} Z(\xi_m) &= \sum_{\boldsymbol{q}^m} \exp(\xi_m[U(\boldsymbol{q}^m) - U_{mf}(\boldsymbol{q}^m)]) \exp(\xi_m U_{mf}(\boldsymbol{q}^m)) \\ &= Z_{mf}(\xi_m) \sum_{\boldsymbol{q}^m} \exp(\xi_m[U(\boldsymbol{q}^m) - U_{mf}(\boldsymbol{q}^m)]) \frac{\exp[\xi_m U_{mf}(\boldsymbol{q}^m)]}{Z_{mf}(\xi_m)} \\ &= Z_{mf}(\xi_m) E_{p_{mf}}\left[\exp(\xi_m[U(\boldsymbol{q}^m) - U_{mf}(\boldsymbol{q}^m)])\right] \end{aligned} \tag{D.79}$$

où $p_{mf}(\boldsymbol{q}^m \mid \xi_m) = \frac{\exp[\xi_m U_{mf}(\boldsymbol{q}^m))}{Z_{mf}(\xi_m]}$. Selon le développement de Tailor de l'exponentiel

$$\exp(x) \simeq 1 + x \;\Rightarrow\; \mathbb{E}[\exp(x)] \simeq 1 + \mathbb{E}[x] \simeq \exp(\mathbb{E}[x]) \tag{D.80}$$

on peut écrire :

$$Z(\xi_m) \simeq Z_{mf}(\xi_m) \exp\left\{\xi_m E_{p_{mf}}[U(\boldsymbol{q}^m) - U_{mf}(\boldsymbol{q}^m)]\right\} \tag{D.81}$$

Notons que x doit avoir une petite valeur pour minimiser l'erreur induite par l'approximation de Taylor, comme le montre la figure D.1, d'où l'intérêt de bien choisir les $\{\widetilde{\boldsymbol{q}}_k^m\}_{k\in N_j}$ de manière à ce que $U_{mf}(\boldsymbol{q}^m)$ soit le plus proche possible de $U(\boldsymbol{q}^m)$. Pour ceci nous proposons de choisir les valeurs courantes des probabilités *a posteriori* $\widetilde{p}_{q_k^m}$. En dérivant le logarithme de l'équation (D.81) on obtient :

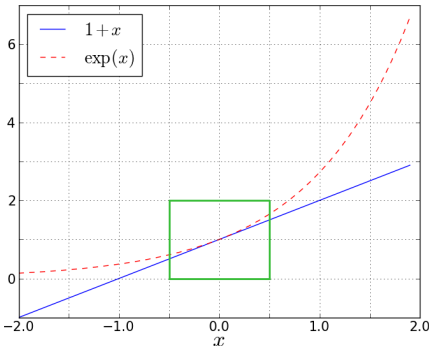


FIGURE D.1 – Représentation des fonctions $1 + x$ en bleu et $\exp(x)$ en rouge. La partie entourée en vert indique les valeurs de x pour lesquelles l'erreur induite par l'approximation de Taylor est petite. Plus x augmente, plus cette erreur est grande.

$$\log Z(\xi_m) \simeq \log Z_{mf}(\xi_m) + \xi_m \; E_{p_{mf}}[U(\boldsymbol{q}^m) - U_{mf}(\boldsymbol{q}^m)] \tag{D.82}$$

$$\begin{aligned}\frac{d\log Z(\xi_m)}{d\xi_m} &\simeq E_{p_{mf}}[U_{mf}(\boldsymbol{q}^m)] + E_{p_{mf}}[U(\boldsymbol{q}^m)] - E_{p_{mf}}[U_{mf}(\boldsymbol{q}^m)] \\ &= E_{p_{mf}}[U(\boldsymbol{q}^m)] \end{aligned} \tag{D.83}$$

où

$$\begin{aligned}\frac{d\log Z_{mf}(\xi_m)}{d\xi_m} &= \frac{\frac{dZ_{mf}(\xi_m)}{d\xi_m}}{Z_{mf}(\xi_m)} = \sum_{\boldsymbol{q}^m} U_{mf}(\boldsymbol{q}^m)\frac{\exp[\xi_m U_{mf}(\boldsymbol{q}^m)]}{Z_{mf}(\xi_m)} \\ &= \sum_{\boldsymbol{q}^m} U_{mf}(\boldsymbol{q}^m) p_{mf}(\boldsymbol{q}^m) = E_{p_{mf}}[U_{mf}(\boldsymbol{q}^m)] \end{aligned} \tag{D.84}$$

En remplaçant l'équation (D.83) dans l'équation (D.74) on obtient :

$$\frac{df(\xi_m)}{d\xi_m} = -E_{p_{mf}}[U(\boldsymbol{q}^m)] + E_{\widetilde{p}_{\boldsymbol{q}^m}}[U(\boldsymbol{q}^m)] - \lambda_{\xi_m} \tag{D.85}$$

Méthode 2. Ici la dérivée est calculée avant l'approximation :

$$\begin{aligned}\frac{d\log Z(\xi_m)}{d\xi_m} &= \frac{\frac{dZ(\xi_m)}{d\xi_m}}{Z(\xi_m)} = \frac{\sum\limits_{\boldsymbol{q}^m} U(\boldsymbol{q}^m)\exp(\xi_m U(\boldsymbol{q}^m))}{Z(\xi_m)} \\ &= \sum_{\boldsymbol{q}^m} U(\boldsymbol{q}^m) p(\boldsymbol{q}^m \mid \xi_m) = E_{p_{\boldsymbol{q}^m}}[U(\boldsymbol{q}^m)] \end{aligned} \tag{D.86}$$

- Calcul de $E_{p_{\boldsymbol{q}^m}}[U(\boldsymbol{q}^m)]$

$$E_{p_{\boldsymbol{q}^m}}[U(\boldsymbol{q}^m)] = \frac{1}{2}\sum_j \sum_{k\in N_j} E_{p_{\boldsymbol{q}^m}}[\left\langle \boldsymbol{q}_j^m, \boldsymbol{q}_k^m \right\rangle] = \frac{1}{2}\sum_j \sum_{k\in N_j} \sum_{\boldsymbol{q}_j^m, \boldsymbol{q}_k^m} \left\langle \boldsymbol{q}_j^m, \boldsymbol{q}_k^m \right\rangle p_{\boldsymbol{q}^m}(\boldsymbol{q}_j^m, \boldsymbol{q}_k^m)$$

Le produit scalaire $\left\langle \boldsymbol{q}_j^m, \boldsymbol{q}_k^m \right\rangle = 1$ uniquement si $\boldsymbol{q}_j^m = \boldsymbol{q}_k^m$ on peut écrire :

$$E_{p_{\boldsymbol{q}^m}}[U(\boldsymbol{q}^m)] = \frac{1}{2}\sum_j \sum_{k\in N_j} \sum_{\boldsymbol{q}_j^m = \boldsymbol{q}_k^m = \boldsymbol{c}} p_{\boldsymbol{q}^m}(\boldsymbol{q}_j^m = \boldsymbol{c}, \boldsymbol{q}_k^m = \boldsymbol{c}) \tag{D.87}$$

Etant donné la compléxité de calcul de $p_{\boldsymbol{q}^m}(\boldsymbol{q}_j^m = \boldsymbol{c}, \boldsymbol{q}_k^m = \boldsymbol{c})$ on propose d'effectuer une approximation champ moyen qui consiste à remplacer $p_{\boldsymbol{q}^m}(\boldsymbol{q}^m)$ par $p_{mf}(\boldsymbol{q}^m)$. L'équation (D.74) devient donc :

$$\frac{df(\xi_m)}{d\xi_m} = -E_{p_{mf}}[U(\boldsymbol{q}^m)] + E_{\widetilde{p}_{\boldsymbol{q}^m}}[U(\boldsymbol{q}^m)] - \lambda_{\xi_m} \tag{D.88}$$

et qui est équivalente à l'équation (D.85).

Reste à calculer le terme $E_{p_{mf}}[U(\boldsymbol{q}^m)]$ afin d'obtenir la formule du gradient permettant l'estimation de ξ_m.

$$\begin{aligned} p_{mf}(\boldsymbol{q}^m) &= \frac{\exp\left(\xi_m U_{mf}(\boldsymbol{q}^m)\right)}{Z_{mf}(\xi_m)} \\ &= \prod_j \frac{\exp\left(\xi_m \left\langle \boldsymbol{q}_j^m, \sum\limits_{k\in N_j} \widetilde{\boldsymbol{q}}_k^m \right\rangle\right)}{\sum\limits_{\boldsymbol{q}_j^m} \exp\left(\xi_m \left\langle \boldsymbol{q}_j^m, \sum\limits_{k\in N_j} \widetilde{\boldsymbol{q}}_k^m \right\rangle\right)} = \prod_j P_{mfj}(\boldsymbol{q}_j^m) \end{aligned} \tag{D.89}$$

avec $P_{mfj}(i) = \dfrac{\exp\left(\xi_m \sum\limits_{k\in N_j} \widetilde{\boldsymbol{q}}_k^m(i)\right)}{\sum\limits_{i\in\{0,1\}} \exp\left(\xi_m \sum\limits_{k\in N_j} \widetilde{\boldsymbol{q}}_k^m(i)\right)}$ et $\widetilde{\boldsymbol{q}}_k^m = \widetilde{p}_{q_k^m}$.

Ce qui donne en remplaçant $p_{\boldsymbol{q}^m}$ par p_{mf} dans l'équation (D.87) :

$$E_{p_{mf}}[U(\boldsymbol{q}^m)] = \frac{1}{2}\sum_j \sum_{k\in N_j} \sum_{i\in\{0,1\}} P_{mfj}(i)P_{mfk}(i) \tag{D.90}$$

- Calcul de $E_{\widetilde{p}_{\boldsymbol{q}^m}}[U(\boldsymbol{q}^m)]$

En remplaçant $p_{\boldsymbol{q}^m}$ par $\widetilde{p}_{\boldsymbol{q}^m}$ dans l'équation (D.87) et en tenant compte de $\widetilde{p}_{\boldsymbol{q}^m}(\boldsymbol{q}^m) = \prod_j \widetilde{p}_{q_j^m}(\boldsymbol{q}_j^m)$, on obtient :

$$E_{\widetilde{p}_{\boldsymbol{q}^m}}[U(\boldsymbol{q}^m)] = \frac{1}{2}\sum_j \sum_{i\in\{0,1\}} \sum_{k\in N_j} \widetilde{p}_{q_j^m}(i)\widetilde{p}_{q_k^m}(i) \tag{D.91}$$

En remplaçant les équations (D.90) et (D.91) dans l'équation du gradient (D.85) on trouve :

$$\frac{df(\xi_m)}{d\xi_m} = -\frac{1}{2}\sum_j \sum_{k\in N_j} \sum_{i\in\{0,1\}} \left\{p_{mfj}(i)p_{mfk}(i) - \widetilde{p}_{q_j^m}(i)\widetilde{p}_{q_k^m}(i)\right\} - \lambda_{\xi_m} \tag{D.92}$$

Méthode 3. Une autre possibilité d'approximation serait $Z(\xi_m) \simeq Z_{mf}(\xi_m)$ qui est d'ordre 0 ($\exp(x) \simeq 1$) et donc moins bonne que la précédente qui est d'ordre 1 ($\exp(x) \simeq 1 + x$) [Forbes 2003]. Dans ce cas :

$$\frac{d\log Z(\xi_m)}{d\xi_m} \simeq \frac{d\log Z_{mf}(\xi_m)}{d\xi_m} = E_{p_{mf}}[U_{mf}(\boldsymbol{q}^m)] \tag{D.93}$$

avec

$$\begin{aligned} E_{p_{mf}}[U_{mf}(\boldsymbol{q}^m)] &= \sum_j \sum_{k\in N_j} E_{p_{mf}}[\left\langle \boldsymbol{q}_j^m, \widetilde{\boldsymbol{q}}_k^m \right\rangle] = \sum_j \sum_{k\in N_j} < E_{p_{mf}}[\boldsymbol{q}_j^m], \widetilde{\boldsymbol{q}}_k^m > \\ &= \sum_j \sum_{k\in N_j} \sum_{i\in\{0,1\}} p_{mfj}(i)\widetilde{\boldsymbol{q}}_k^m(i) \end{aligned} \tag{D.94}$$

Ce qui donne en remplaçant dans l'équation (D.85)

$$\frac{df(\xi_m)}{d\xi_m} = -\sum_j \sum_{k\in N_j} \sum_{i\in\{0,1\}} \left\{p_{mfj}(i)\widetilde{\boldsymbol{q}}_k^m(i) - \frac{1}{2}\widetilde{p}_{q_j^m}(i)\widetilde{p}_{q_k^m}(i)\right\} - \lambda_{\xi_m} \tag{D.95}$$

En choisissant $\widetilde{\boldsymbol{q}}_k^m = \widetilde{p}_{q_k^m}$ on obtient :

$$\frac{df(\xi_m)}{d\xi_m} = -\sum_j \sum_{k \in N_j} \sum_{i \in \{0,1\}} \widetilde{p}_{q_k^m}(i) \left\{ p_{mfj}(i) - \frac{1}{2} \widetilde{p}_{q_j^m}(i) \right\} - \lambda_{\xi_m} \tag{D.96}$$

Méthode 4. La dernière possibilité qui s'avère bien sûr la moins bonne de toutes est l'approximation :

$$\frac{df(\xi_m)}{d\xi_m} = -E_{p_{mf}}[U_{mf}(\boldsymbol{q}^m)] + E_{\widetilde{p}_{\boldsymbol{q}^m}}[U_{mf}(\boldsymbol{q}^m)] - \lambda_{\xi_m} \tag{D.97}$$

où

$$\begin{aligned} E_{\widetilde{p}_{\boldsymbol{q}^m}}[U_{mf}(\boldsymbol{q}^m)] &= \sum_j \sum_{k \in N_j} E_{\widetilde{p}_{\boldsymbol{q}^m}}[\left\langle \boldsymbol{q}_j^m, \widetilde{\boldsymbol{q}}_k^m \right\rangle] = \sum_j \sum_{k \in N_j} \left\langle E_{\widetilde{p}_{\boldsymbol{q}^m}}[\boldsymbol{q}_j^m], \widetilde{\boldsymbol{q}}_k^m \right\rangle \\ &= \sum_j \sum_{k \in N_j} \sum_{i \in \{0,1\}} \widetilde{p}_{q_j^m}(i) \widetilde{\boldsymbol{q}}_k^m(i) \end{aligned} \tag{D.98}$$

Ce qui donne en remplaçant les équations (D.94) et (D.98) dans l'équation (D.97) et en choisissant $\widetilde{\boldsymbol{q}}_k^m = \widetilde{p}_{q_k^m}$:

$$\frac{df(\xi_m)}{d\xi_m} = -\sum_j \sum_{i \in \{0,1\}} \left(\sum_{k \in N_j} \widetilde{p}_{q_k^m}(i) \right) \left\{ p_{mfj}(i) - \widetilde{p}_{q_j^m}(i) \right\} - \lambda_{\xi_m} \tag{D.99}$$

D.2 Connexion aux moyennes d'activations

Ici, par rapport à l'approche DEC *parcimonieuse* précédente (section D.1), le changement concerne la connexion entre les variables de pertinence $\boldsymbol{w}$ et $\boldsymbol{Q}$. Cette dernière est remplacée par une autre connexion qui relie $\boldsymbol{w}$ à $\boldsymbol{\mu}_1 = \{\mu_1^m; m = 1 : M\}$ où μ_1^m exprime la moyenne de la gaussienne représentant la classe des voxels activés. Ce changement induit une modification dans les étapes **E-Q**, **E-W** et **M-$\boldsymbol{\mu}_1$** ainsi que l'ajout d'une étape **M-$\boldsymbol{\tau}_2$** nécessaire à l'estimation du paramètre τ_2 de la sigmoïde et qui représente également une différence par rapport à l'approche *parcimonieuse* précédente (section D.1).

♣ **E-Q**

$$\begin{aligned}\widetilde{p}_{q_j^m}(q_j^m) &\propto \exp\left(E_{\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{\boldsymbol{A}},\widetilde{p}_{\boldsymbol{w}},\widetilde{p}_{\boldsymbol{q}_{\setminus j}^m},\widetilde{p}_{\boldsymbol{Q}^{\setminus m}}}\left[\log p(q_j^m \mid \boldsymbol{q}_{\setminus j}^m, \boldsymbol{Q}^{\setminus m}, \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{h}, \boldsymbol{w}, \boldsymbol{\theta})\right]\right)\\ &\propto \exp\left(E_{\widetilde{p}_{a_j^m},\widetilde{p}_{w^m},\widetilde{p}_{\boldsymbol{q}_{\setminus j}^m}}\left[\log p(a_j^m \mid w^m, q_j^m, \boldsymbol{\theta}^m)\ p(\boldsymbol{q}^m \mid \xi_m)\right]\right)\\ &\propto f(q_j^m) \end{aligned} \tag{D.100}$$

où on note

$$f(q_j^m) = \exp\left(E_{\widetilde{p}_{a_j^m},\widetilde{p}_{w^m}}\Big[\log p(a_j^m \mid \boldsymbol{w}^m, q_j^m, \boldsymbol{\theta}_m)\Big] + E_{\widetilde{p}_{\boldsymbol{q}_{\setminus j}^m}}[\log p(\boldsymbol{q}^m \mid \xi_m)]\right) \tag{D.101}$$

Par rapport à l'équation (D.35), il manque le terme $p(w^m \mid \boldsymbol{q}^m, \boldsymbol{\tau})$ car $\boldsymbol{w}$ ne dépend plus de $\boldsymbol{Q}$ *a priori*. Ce qui donne :

$$f(q_j^m = 0) = \exp\left(\xi_m \sum_{k\in N(j)} \widetilde{p}_{q_k^m}(0)\right) \tag{D.102}$$

$$\begin{aligned} f(q_j^m = 1) = \exp\Bigg(&\widetilde{p}_{w^m}(1)\left\{\frac{V_{\boldsymbol{a}_j^{m,m}}}{2}\left(\frac{1}{v_0^m} - \frac{1}{v_1^m}\right) + \log\frac{\mathcal{N}(m_{a_j^m}; \mu_1^m, v_1^m)}{\mathcal{N}(m_{a_j^m}; \mu_0^m, v_0^m)}\right\}\\ &+\xi_m \sum_{k\in N(j)} \widetilde{p}_{q_k^m}(1)\Bigg) \end{aligned} \tag{D.103}$$

♣ **E-W**

$$\begin{aligned}\widetilde{p}_{w^m}(w^m) &\propto \exp\left(E_{\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{\boldsymbol{A}},\widetilde{p}_{\boldsymbol{Q}},\widetilde{p}_{\boldsymbol{w}^{\setminus m}}}\left[\log p(w^m \mid \boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{Q}, \boldsymbol{h}, \boldsymbol{w}^{\setminus m}, \boldsymbol{L}, \boldsymbol{\theta})\right]\right)\\ &\propto f(w^m) \end{aligned} \tag{D.104}$$

où on note

$$\begin{aligned} f(w^m) = p(w^m \mid \mu_1^m, \boldsymbol{\tau}) \exp\Big(&E_{\widetilde{p}_{\boldsymbol{h}},\widetilde{p}_{\boldsymbol{A}},\widetilde{p}_{\boldsymbol{w}^{\setminus m}}}\left[\log p(\boldsymbol{Y} \mid \boldsymbol{A}, \boldsymbol{h}, w^m, \boldsymbol{w}^{\setminus m}, \boldsymbol{L}, \boldsymbol{\sigma}^2)\right]\\ &+ E_{\widetilde{p}_{\boldsymbol{a}^m},\widetilde{p}_{\boldsymbol{q}^m}}\left[\log p(\boldsymbol{a}^m \mid w^m, \boldsymbol{q}^m, \boldsymbol{\theta}_m)\right]\Big) \end{aligned} \tag{D.105}$$

On rappelle que $\widetilde{p}_{\boldsymbol{w}^{\setminus m}}$ désigne la loi de $\left\{\boldsymbol{w}^{m'}\right\}_{m'\neq m}$ et que $p(w^m \mid \mu_1^m, \boldsymbol{\tau}) = p_m^{w^m}\ (1-p_m)^{(1-w^m)}$ où p_m est donné par l'équation (III.70). Les deux termes de l'exponentielle ont déjà été calculés dans l'annexe D.1 ce qui donne en incluant le troisième terme les expressions suivantes qui sont ensuite normalisées afin de garantir des

probabilités dans $]0,1[$

$$f(w^m=0) = 1-p_m \tag{D.106}$$

$$\begin{aligned} f(w^m=1) = p_m \ \exp\Bigg(\sum_j \Bigg\{ & \widetilde{p}_{q_j^m}(1)\left[\frac{V_{\boldsymbol{a}_j^{m,m}}}{2}\left(\frac{1}{v_0^m}-\frac{1}{v_1^m}\right)+\log\frac{\mathcal{N}(m_{a_j^m};\mu_1^m,v_1^m)}{\mathcal{N}(m_{a_j^m};\mu_0^m,v_0^m)}\right] \\ & -\frac{m_{a_j^m}^2+V_{\boldsymbol{a}_j^{m,m}}}{2}\left[\boldsymbol{m}_H^{\mathrm{t}}\boldsymbol{X}^{m\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{X}^m\boldsymbol{m}_H+tr\left(\boldsymbol{V}_H\boldsymbol{X}^{m\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{X}^m\right)\right] \\ & -\sum_{m'\neq m}\left(m_{a_j^m}m_{a_j^{m'}}+V_{\boldsymbol{a}_j^{m,m'}}\right)\widetilde{p}_{w^{m'}}(1)\left[\boldsymbol{m}_H^{\mathrm{t}}\boldsymbol{X}^{m\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{X}^{m'}\boldsymbol{m}_H\right. \\ & \left.+tr\left(\boldsymbol{V}_H\boldsymbol{X}^{m\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{X}^{m'}\right)\right]+\ m_{a_j^m}\boldsymbol{m}_H^{\mathrm{t}}\boldsymbol{X}^{m\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j-\boldsymbol{P}\boldsymbol{\ell}_j)\Bigg\}\Bigg) \end{aligned} \tag{D.107}$$

♣ **M-μ_1**

Dans la précédente approche DEC parcimonieuse, et en rappelant l'indépendance inter-conditions, le paramètre μ_1^m était estimé à travers la maximisation de la fonction $f(\mu_1^m, v_1^m)$ (équation (D.51)). Maintenant que la variable w^m est connectée à μ_1^m *a priori* l'estimation se fait comme suit :

$$\hat{\mu}_1^m = \arg\max_{\mu_1^m}\left(-\frac{1}{2}f(\mu_1^m,v_1^m)+E_{\widetilde{p}_{w^m}}\left[\log\left(p_m^{w^m}(1-p_m)^{1-w^m}\right)\right]\right) \tag{D.108}$$

- $E_{\widetilde{p}_{w^m}}\left[\log\left(p_m^{w^m}(1-p_m)^{1-w^m}\right)\right]$

$$\begin{aligned} &= \widetilde{p}_{w^m}(1)\log p_m+(1-\widetilde{p}_{w^m}(1))\log(1-p_m) \\ &= -(1-\widetilde{p}_{w^m}(1))[\tau_1((\mu_1^m)^2-\tau_2)]-\log(1+\exp[-\tau_1((\mu_1^m)^2-\tau_2)]) \end{aligned} \tag{D.109}$$

En dérivant le terme à maximiser dans l'équation (D.108) on obtient :

$$\begin{aligned} \frac{\partial(.)}{\partial\mu_1^m} &= -\frac{1}{2}\sum_j \widetilde{p}_{q_j^m}(1)\widetilde{p}_{w^m}(1)\left(\frac{-2(m_{a_j^m}-\mu_1^m)}{v_1^m}\right) \\ &\qquad -2(1-\widetilde{p}_{w^m}(1))\tau_1\mu_1^m-\frac{-2\tau_1\mu_1^m\exp[-\tau_1((\mu_1^m)^2-\tau_2)]}{1+\exp[-\tau_1((\mu_1^m)^2-\tau_2)]} \\ &= \frac{\widetilde{p}_{w^m}(1)}{v_1^m}\sum_j\widetilde{p}_{q_j^m}(1)m_{a_j^m}-\mu_1^m\frac{\widetilde{p}_{w^m}(1)}{v_1^m}\sum_j\widetilde{p}_{q_j^m}(1) \\ &\qquad -2\ \tau_1\ (1-\widetilde{p}_{w^m}(1))\mu_1^m+2\ \tau_1\ \mu_1^m-2\ \tau_1\ \mu_1^m S(\mu_1^m) \\ &= B+B'\ \mu_1^m-2\ \tau_1\ \mu_1^m\ S(\mu_1^m)=0 \end{aligned} \tag{D.110}$$

avec

$$B' = 2\,\tau_1\,\widetilde{p}_{w^m}(1) - \frac{\widetilde{p}_{w^m}(1)}{v_1^m}\sum_j \widetilde{p}_{q_j^m}(1) \tag{D.111}$$

$$B = \frac{\widetilde{p}_{w^m}(1)}{v_1^m}\sum_j \widetilde{p}_{q_j^m}(1) m_{a_j^m} \tag{D.112}$$

On remarque l'absence de formule explicite permettant l'estimation directe de μ_1^m, dans ce cas un simple algorithme de dichotomie suffit (Annexe F.4).

♣ **M-$\boldsymbol{\tau}_2$**

$$\begin{aligned}\hat{\tau}_2 &= \arg\max_{\tau_2}\left\{E_{\widetilde{p}_{\boldsymbol{w}}}\left[\log p(\boldsymbol{w}\,|\,\boldsymbol{\mu}_1,\boldsymbol{\tau})\right] + \log p(\tau_2|\alpha_{\tau_2},\lambda_{\tau_2})\right\}\\ &= \arg\max_{\tau_2}\left\{\sum_{m=1}^{M}\left\{\left([1-\widetilde{p}_{w^m}(1)]\tau_1 - \lambda_{\tau_2}\right)\tau_2 + (\alpha_{\tau_2}-1)\log\tau_2\right.\right.\\ &\qquad \left.\left. - \log\left(1+\exp\left[-\tau_1\left((\mu_1^m)^2-\tau_2\right)\right]\right) + C_{\tau_2}\right\}\right\}\end{aligned} \tag{D.113}$$

$$\begin{aligned}&\Rightarrow \sum_{m=1}^{M}\left[(1-\widetilde{p}_{w^m}(1))\tau_1 - \frac{\exp(-\tau_1((\mu_1^m)^2-\tau_2))\tau_1}{1+\exp(-\tau_1((\mu_1^m)^2-\tau_2))}\right] - \lambda_{\tau_2} + \frac{\alpha_{\tau_2}-1}{\tau_2} = 0\\ &\sum_{m=1}^{M}\left[(1-\widetilde{p}_{w^m}(1))\tau_1 - (1-S(\mu_1^m,\tau_2))\tau_1\right] - \lambda_{\tau_2} + \frac{\alpha_{\tau_2}-1}{\tau_2} = 0\\ &\tau_1\sum_{m=1}^{M}\left[S(\mu_1^m,\tau_2) - \widetilde{p}_{w^m}(1)\right] - \lambda_{\tau_2} + \frac{\alpha_{\tau_2}-1}{\tau_2} = 0\end{aligned} \tag{D.114}$$

Comme pour l'estimation de μ_1, une étape de dichotomie (annexe F.4) permet l'estimation du paramètre τ_2.

Annexe E

Energie libre

Comme nous l'avons précisé dans la section III.2.5 du chapitre III, l'énergie libre représente le terme maximisé en VEM, son calcul nous permet donc de vérifier la convergence de l'algorithme. Par la suite les matrices et vecteurs sont notés en gras avec les matrices en majuscule et les vecteurs en minuscule. Un vecteur est par convention un vecteur colonne. La transposé est notée ^{t}.

Rappelons que l'énergie libre est définie par l'équation suivante :

$$F(\widetilde{p}, \boldsymbol{\theta}) = \mathbb{E}_{\widetilde{p}}[\log p(\boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{Q}, \boldsymbol{w}, \boldsymbol{h} \,|\, \boldsymbol{\theta})] - \mathbb{E}_{\widetilde{p}}[\log \widetilde{p}(\boldsymbol{A}, \boldsymbol{Q}, \boldsymbol{w}, \boldsymbol{h})] \tag{E.1}$$

- **Calcul de l'entropie** $E_{\widetilde{p}}[\log \widetilde{p}\,(\boldsymbol{A}, \boldsymbol{Q}, \boldsymbol{w}, \boldsymbol{h})]$ **:**

$$\begin{aligned}
&E_{\widetilde{p}}[\log \widetilde{p}\,(\boldsymbol{A}, \boldsymbol{Q}, \boldsymbol{w}, \boldsymbol{h})] \\
&= E_{\widetilde{p}}[\log \widetilde{p}_{\boldsymbol{A}}(\boldsymbol{A})\ \widetilde{p}_{\boldsymbol{Q}}(\boldsymbol{Q})\ \widetilde{p}_{\boldsymbol{w}}(\boldsymbol{w})\ \widetilde{p}_{\boldsymbol{h}}(\boldsymbol{h})] \\
&= E_{\widetilde{p}_{\boldsymbol{A}}}[\log \widetilde{p}_{\boldsymbol{A}}(\boldsymbol{A})] + E_{\widetilde{p}_{\boldsymbol{Q}}}[\log \widetilde{p}_{\boldsymbol{Q}}(\boldsymbol{Q})] + E_{\widetilde{p}_{\boldsymbol{w}}}[\log \widetilde{p}_{\boldsymbol{w}}(\boldsymbol{w})] + E_{\widetilde{p}_{\boldsymbol{h}}}[\log \widetilde{p}_{\boldsymbol{h}}(\boldsymbol{h})]
\end{aligned} \tag{E.2}$$

Selon les équations (D.5) et (D.18), $\widetilde{p}_{\boldsymbol{h}}$ et $\widetilde{p}_{\boldsymbol{A}}$ sont des gaussiennes. Les variables $\boldsymbol{Q}$ et $\boldsymbol{w}$ sont quant à elles des groupes de variables binaires, ce qui donne (annexe F.6) :

$$\begin{aligned}
E_{\widetilde{p}_{\boldsymbol{A}}}\left[\log \widetilde{p}_{\boldsymbol{A}}(\boldsymbol{A})\right] &= \sum_j E_{\widetilde{p}_{\boldsymbol{a}_j}}\left[\log \widetilde{p}_{\boldsymbol{a}_j}(\boldsymbol{A})\right] = -\sum_j I\left[\widetilde{p}_{\boldsymbol{a}_j}\right] \\
&= -\sum_j \log\left(\sqrt{(2\pi e)^M |\boldsymbol{V}_{\boldsymbol{a}_j}|}\right)
\end{aligned}$$

$$
\begin{aligned}
E_{\widetilde{p}_{\boldsymbol{h}}}\left[\log \widetilde{p}_{\boldsymbol{h}}(\boldsymbol{h})\right] &= -I\left[\widetilde{p}_{\boldsymbol{h}}(\boldsymbol{h})\right] = -\log\left(\sqrt{(2\pi e)^D|\boldsymbol{V}_{\boldsymbol{h}}|}\right) \\
E_{\widetilde{p}_{\boldsymbol{Q}}}\left[\log \widetilde{p}_{\boldsymbol{Q}}(\boldsymbol{Q})\right] &= \sum_m \sum_j E_{\widetilde{p}_{q_j^m}}\left[\log \widetilde{p}_{q_j^m}(q_j^m)\right] = -\sum_m \sum_j I\left[\widetilde{p}_{q_j^m}\right] \\
&= \sum_m \sum_j \left[\widetilde{p}_{q_j^m}(1)\log \widetilde{p}_{q_j^m}(1) + \widetilde{p}_{q_j^m}(0)\log \widetilde{p}_{q_j^m}(0)\right] \\
E_{\widetilde{p}_{\boldsymbol{w}}}\left[\log \widetilde{p}_{\boldsymbol{w}}(\boldsymbol{w})\right] &= \sum_m E_{\widetilde{p}_{w^m}}\left[\log \widetilde{p}_{w^m}(w^m)\right] = -\sum_m I\left[\widetilde{p}_{w^m}\right] \\
&= \sum_m \left[\widetilde{p}_{w^m}(1)\log \widetilde{p}_{w^m}(1) + \widetilde{p}_{w^m}(0)\log \widetilde{p}_{w^m}(0)\right]
\end{aligned}
$$

- **Calcul de** $E_{\widetilde{p}}[\log p\,(\boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{Q}, \boldsymbol{w}, \boldsymbol{h} \,|\, \boldsymbol{\theta})]$ **:**

$$
\begin{aligned}
&E_{\widetilde{p}}[\log p\,(\boldsymbol{Y}, \boldsymbol{A}, \boldsymbol{Q}, \boldsymbol{w}, \boldsymbol{h} \,|\, \boldsymbol{\theta})] \\
&= E_{\widetilde{p}}\Big[\log p(\boldsymbol{Y} \,|\, \boldsymbol{A}, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{L}, \boldsymbol{\sigma}^2)\; p(\boldsymbol{A} \,|\, \boldsymbol{Q}, \boldsymbol{w}, \boldsymbol{\theta}_A)\; p(\boldsymbol{h} \,|\, v_h)\; p(\boldsymbol{w} \,|\, \boldsymbol{x}, \boldsymbol{\tau})\; p(\boldsymbol{Q} \,|\, \boldsymbol{\xi})\Big] \\
&= E_{\widetilde{p}_{\boldsymbol{A}}\widetilde{p}_{\boldsymbol{w}}\widetilde{p}_{\boldsymbol{h}}}\Big[\log p(\boldsymbol{Y} \,|\, \boldsymbol{A}, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{L}, \boldsymbol{\sigma}^2)\Big] + E_{\widetilde{p}_{\boldsymbol{A}}\widetilde{p}_{\boldsymbol{w}}\widetilde{p}_{\boldsymbol{Q}}}\left[\log p(\boldsymbol{A} \,|\, \boldsymbol{Q}, \boldsymbol{w}, \boldsymbol{\theta}_A)\right] \\
&\quad + E_{\widetilde{p}_{\boldsymbol{h}}}\left[\log p(\boldsymbol{h} \,|\, v_h)\right] + E_{\widetilde{p}}\left[\log p(\boldsymbol{w} \,|\, \boldsymbol{x}, \boldsymbol{\tau})\right] + E_{\widetilde{p}_{\boldsymbol{Q}}}\left[\log p(\boldsymbol{Q} \,|\, \boldsymbol{\xi})\right]
\end{aligned} \tag{E.3}
$$

avec $\boldsymbol{x} = \boldsymbol{Q}$ ou $\boldsymbol{x} = \boldsymbol{\mu}_1$ selon le modèle.

Nous allons par la suite calculer les 5 termes de l'équation (E.3).

1) $E_{\widetilde{p}_{\boldsymbol{A}}\widetilde{p}_{\boldsymbol{w}}\widetilde{p}_{\boldsymbol{h}}}\left[\log p(\boldsymbol{Y} \,|\, \boldsymbol{A}, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{L}, \boldsymbol{\sigma}^2)\right]$ **:** Selon l'équation (D.2) on a,

$$
\log p\left(\boldsymbol{Y} \,|\, \boldsymbol{A}, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{L}, \boldsymbol{\sigma}^2\right) = C_{2,\boldsymbol{h}} - \frac{1}{2}\sum_{j\in J}\Big[\boldsymbol{h}^{\mathrm{t}}\boldsymbol{S}_j^{\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{S}_j\boldsymbol{h} - 2\,\boldsymbol{h}^{\mathrm{t}}\boldsymbol{S}_j^{\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j - \boldsymbol{P}\boldsymbol{\ell}_j)\Big]
$$

Ce qui donne :

$$
\begin{aligned}
E_{\widetilde{p}_{\boldsymbol{A}}\widetilde{p}_{\boldsymbol{w}}\widetilde{p}_{\boldsymbol{h}}}\Big[\log p(\boldsymbol{Y} \,|\, \boldsymbol{A}, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{L}, \boldsymbol{\sigma}^2)\Big] = C_{2,\boldsymbol{h}} - \frac{1}{2}\Bigg\{&\boldsymbol{m}_H^{\mathrm{t}} E_{\widetilde{p}_{\boldsymbol{A}}\widetilde{p}_{\boldsymbol{w}}}\left[\sum_j \boldsymbol{S}_j^{\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{S}_j\right]\boldsymbol{m}_H \\
&+ \operatorname{tr}\left(\boldsymbol{V}_H E_{\widetilde{p}_{\boldsymbol{A}}\widetilde{p}_{\boldsymbol{w}}}\left[\sum_j \boldsymbol{S}_j^{\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{S}_j\right]\right) - 2\,\boldsymbol{m}_H^{\mathrm{t}} E_{\widetilde{p}_{\boldsymbol{A}}\widetilde{p}_{\boldsymbol{w}}}\left[\sum_j \boldsymbol{S}_j^{\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{y}_j\right]\Bigg\}
\end{aligned}
$$

En remplaçant les Eqs. (D.8) et (D.9) ainsi que la constante $C_{2,\boldsymbol{h}}$ dans l'équation

précédente on obtient :

$$
\begin{aligned}
&E_{\widetilde{p}_{\boldsymbol{A}}\widetilde{p}_{\boldsymbol{w}}\widetilde{p}_{\boldsymbol{h}}}\left[\log p(\boldsymbol{Y}\,|\,\boldsymbol{A},\boldsymbol{w},\boldsymbol{h},\boldsymbol{L},\boldsymbol{\sigma}^2)\right] = -\frac{NJ}{2}\log 2\pi + \frac{J}{2}\log|\boldsymbol{\Lambda}_j| - N\sum_{j\in J}\log\sigma_j \\
&+\boldsymbol{m}_H^{\mathrm{t}}\sum_{j\in J}\sum_{m=1}^{M} m_{a_j^m}\widetilde{p}_{w^m}(1)\boldsymbol{X}^{m\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j-\boldsymbol{P}\boldsymbol{\ell}_j) - \frac{1}{2}\sum_j(\boldsymbol{y}_j-\boldsymbol{P}\boldsymbol{\ell}_j)^{\mathrm{t}}\boldsymbol{\Gamma}_j(\boldsymbol{y}_j-\boldsymbol{P}\boldsymbol{\ell}_j) \\
&-\frac{1}{2}\sum_{j\in J}\sum_{m,m'}\left\{\left(m_{a_j^m}m_{a_j^{m'}}+V_{\boldsymbol{a}_j^{m,m'}}\right)\widetilde{p}_{w^m}(1)\widetilde{p}_{w^{m'}}(1)\right. \\
&\left.\times\left[\boldsymbol{m}_H^{\mathrm{t}}\boldsymbol{X}^{m\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{X}^{m'}\boldsymbol{m}_H + \mathrm{tr}\left(\boldsymbol{V}_H\boldsymbol{X}^{m\mathrm{t}}\boldsymbol{\Gamma}_j\boldsymbol{X}^{m'}\right)\right]\right\}
\end{aligned}
\tag{E.4}
$$

2) $E_{\widetilde{p}_{\boldsymbol{A}}\widetilde{p}_{\boldsymbol{Q}}\widetilde{p}_{\boldsymbol{w}}}\left[\log p(\boldsymbol{A}\,|\,\boldsymbol{Q},\boldsymbol{w},\boldsymbol{\theta}_A)\right]$ **:** Selon l'équation (D.16) on a,

$$
\begin{aligned}
&E_{\widetilde{p}_{\boldsymbol{Q}},\widetilde{p}_{\boldsymbol{w}}}\left[\log p(\boldsymbol{A}\,|\,\boldsymbol{Q},\boldsymbol{w},\boldsymbol{\theta}_A)\right] \\
&= C'_{3,\boldsymbol{A}} - \frac{1}{2}\sum_m\sum_j\left\{\left[1-\widetilde{p}_{q_j^m}(1)\widetilde{p}_{w^m}(1)\right]\left[\frac{(a_j^m-\mu_0^m)^2}{v_0^m}\right] + \widetilde{p}_{q_j^m}(1)\widetilde{p}_{w^m}(1)\left[\frac{(a_j^m-\mu_1^m)^2}{v_1^m}\right]\right\}
\end{aligned}
$$

avec

$$
\begin{aligned}
C'_{3,\boldsymbol{A}} &= E_{\widetilde{p}_{\boldsymbol{Q}},\widetilde{p}_{\boldsymbol{w}}}\left[C_{3,\boldsymbol{A}}\right] \\
&= \sum_{m,j}\left\{\left[1-\widetilde{p}_{q_j^m}(1)\widetilde{p}_{w^m}(1)\right]\log\frac{1}{\sqrt{2\pi v_0^m}} + \widetilde{p}_{q_j^m}(1)\widetilde{p}_{w^m}(1)\log\frac{1}{\sqrt{2\pi v_1^m}}\right\}
\end{aligned}
\tag{E.5}
$$

Ce qui donne :

$$
\begin{aligned}
&E_{\widetilde{p}_{\boldsymbol{A}}\widetilde{p}_{\boldsymbol{w}}\widetilde{p}_{\boldsymbol{Q}}}\left[\log p(\boldsymbol{A}\,|\,\boldsymbol{w},\boldsymbol{Q},\boldsymbol{\theta}_A)\right] \\
&= \sum_m\sum_j\left\{\left[1-\widetilde{p}_{q_j^m}(1)\widetilde{p}_{w^m}(1)\right]\left[\log\frac{1}{\sqrt{2\pi v_0^m}} - \frac{(m_{a_j^m}-\mu_0^m)^2+V_{\boldsymbol{a}_j^{m,m}}}{2v_0^m}\right]\right. \\
&\left.\quad + \widetilde{p}_{q_j^m}(1)\widetilde{p}_{w^m}(1)\left[\log\frac{1}{\sqrt{2\pi v_1^m}} - \frac{(m_{a_j^m}-\mu_1^m)^2+V_{\boldsymbol{a}_j^{m,m}}}{2v_1^m}\right]\right\}
\end{aligned}
\tag{E.6}
$$

3) $E_{\widetilde{p}_{\boldsymbol{h}}}\left[\log p(\boldsymbol{h}\,|\,v_h)\right]$ **:** Selon l'équation (D.3) et en remplaçant la constante $C_{3,\boldsymbol{h}}$ par sa valeur on a,

$$
\log p\left(\boldsymbol{h}\,|\,v_h\right) = -\frac{D+1}{2}\log 2\pi - \frac{D-1}{2}\log v_h - \frac{1}{2}\log|\boldsymbol{R}| - \frac{1}{2v_h}\boldsymbol{h}^{\mathrm{t}}\boldsymbol{R}^{-1}\boldsymbol{h}
$$

$$\Rightarrow E_{\widetilde{p}_{\boldsymbol{h}}}\left[\log p(\boldsymbol{h}\,|\,v_h)\right] = -\frac{D+1}{2}\log 2\pi - \frac{D-1}{2}\log v_h - \frac{1}{2}\log|\boldsymbol{R}| - \frac{\boldsymbol{m}_H^{\mathrm{t}}\boldsymbol{R}^{-1}\boldsymbol{m}_H + \mathrm{tr}\,(\boldsymbol{V}_H\boldsymbol{R}^{-1})}{2v_h} \tag{E.7}$$

4) $E_{\widetilde{p}}[\log p(\boldsymbol{w}\,|\,\boldsymbol{x},\boldsymbol{\tau})]$ **:**

- Cas de connexion avec les variables de classification $\boldsymbol{x} = \boldsymbol{Q}$:

$$E_{\widetilde{p}_{\boldsymbol{w}}\widetilde{p}_{\boldsymbol{Q}}}\left[\log p(\boldsymbol{w}\,|\,\boldsymbol{Q},\boldsymbol{\tau})\right] = \sum_m E_{\widetilde{p}_{w^m}\widetilde{p}_{\boldsymbol{q}^m}}\left[\log p(w^m\,|\,\boldsymbol{q}^m,\boldsymbol{\tau})\right]$$

D'après l'équation (D.29) on a,

$$E_{\widetilde{p}_{\boldsymbol{q}^m}}\left[\log p(w^m\,|\,\boldsymbol{q}^m,\boldsymbol{\tau})\right] = C'_{5,w^m} + (1-w^m)\left[-\tau_1\left(\sum_j \widetilde{p}_{q_j^m}(1) - \tau_2\right)\right]$$

avec $C'_{5,w^m} = E_{\widetilde{p}_{\boldsymbol{q}^m}}[C_{5,w^m}] = E_{\widetilde{p}_{\boldsymbol{q}^m}}[-\log\{1+\exp[-\tau_1(\sum_j q_j^m - \tau_2)]\}]$ qui est calculée à travers une étape MC comme dans l'annexe D.1. On obtient donc :

$$E_{\widetilde{p}_{\boldsymbol{w}}\widetilde{p}_{\boldsymbol{Q}}}\left[\log p(\boldsymbol{w}\,|\,\boldsymbol{Q},\boldsymbol{\tau})\right] = \sum_m\left[C'_{7,w^m} - \tau_1\,\widetilde{p}_{w^m}(0)\left(\sum_j \widetilde{p}_{q_j^m}(1) - \tau_2\right)\right] \tag{E.8}$$

- Cas de connexion avec les moyennes d'activation $\boldsymbol{x} = \boldsymbol{\mu}_1$:

$$\begin{aligned} &E_{\widetilde{p}_{\boldsymbol{w}}}\left[\log p(\boldsymbol{w}\,|\,\boldsymbol{\mu}_1,\boldsymbol{\tau})\right] \\ &= \sum_m E_{\widetilde{p}_{w^m}}\left[\log p(w^m\,|\,\mu_1^m,\boldsymbol{\tau})\right] \\ &= \sum_m\left\{-\log\left[1+\exp\left(-\tau_1\left[(\mu_1^m)^2-\tau_2\right]\right)\right] - \tau_1\left[(\mu_1^m)^2-\tau_2\right]\widetilde{p}_{w^m}(0)\right\} \end{aligned} \tag{E.9}$$

5) $E_{\widetilde{p}_{\boldsymbol{Q}}}[\log p(\boldsymbol{Q}\,|\,\boldsymbol{\xi})]$ **:**

$$E_{\widetilde{p}_{\boldsymbol{Q}}}\left[\log p(\boldsymbol{Q}\,|\,\boldsymbol{\xi})\right] = \sum_m\left\{-\log Z(\xi^m) + \xi^m E_{\widetilde{p}_{\boldsymbol{q}^m}}\left[U(\boldsymbol{q}^m)\right]\right\}$$

avec

$$E_{\widetilde{p}_{\boldsymbol{q}^m}}[U(\boldsymbol{q}^m)] = \frac{1}{2}\sum_j\sum_{k\in N(j)} E_{\widetilde{p}_{q_j^m}\widetilde{p}_{q_k^m}}\left[I(q_j^m = q_k^m)\right] = \frac{1}{2}\sum_j\sum_{k\in N(j)}\sum_{i=0}^{1}\widetilde{p}_{q_j^m}(i)\widetilde{p}_{q_k^m}(i)$$

• **Calcul de** $Z(\xi^m)$ **:** Etant donné la compléxité de ce terme, son calcul est effectué à travers une approximation champs moyen permettant d'obtenir l'équation

(D.81) dont le calcul a déjà été détaillé dans l'annexe D.1 :

$$Z(\xi^m) \simeq Z_{mf}(\xi^m) \exp\left(\xi^m E_{p_{mf}}\left[U(\boldsymbol{q}^m) - U_{mf}(\boldsymbol{q}^m)\right]\right) \tag{E.10}$$

où $U_{mf}(\boldsymbol{q}^m)$,$Z_{mf}(\xi^m)$ et p_{mf} sont donnés par les équations (D.77), (D.78) et (D.89) respectivement. Et les termes $E_{p_{mf}}[U(\boldsymbol{q}^m)]$ et $E_{p_{mf}}[U_{mf}(\boldsymbol{q}^m)]$ sont donnés par les équations (D.90) et (D.94). Ce qui donne en tenant compte du choix $\widetilde{q}_k^m = \widetilde{p}_{q_k^m}$.

$$\begin{aligned} Z(\xi^m) \;&\simeq\; \prod_j \sum_{i=0}^{1} \exp\left(\xi^m \sum_{k\in N(j)} \widetilde{p}_{q_k^m}(i)\right) \\ &\times \exp\left(\xi^m \sum_j \sum_{k\in N(j)} \sum_{i=0}^{1} \left[p_{mfj}(i)\left(\frac{p_{mfk}(i)}{2} - \widetilde{p}_{q_k^m}(i)\right)\right]\right) \end{aligned} \tag{E.11}$$

Remarque : dans le cas d'un *a priori* ARD, les calculs précédents restent les mêmes avec les changements suivants :

- Entropie de $\boldsymbol{w}$:

$$E_{\widetilde{p}_{\boldsymbol{w}}}\left[\log \widetilde{p}_{\boldsymbol{w}}(\boldsymbol{w})\right] = \sum_m E_{\widetilde{p}_{w^m}}\left[\log \widetilde{p}_{w^m}(w^m)\right] = -\sum_m \log\left(\sqrt{2\pi e v_{w^m}}\right)$$

- Dans l'équation (E.4), on remplace $\widetilde{p}_{w^m}(1)$ par m_{w^m} et $\widetilde{p}_{w^m}(1)\widetilde{p}_{w^{m'}}(1)$ par $m_{w^m}m_{w^{m'}} + v_{w^{m,m'}}$ où $v_{w^{m,m'}} = v_{w^m}$ si $m = m'$ et 0 sinon.
- Dans l'équation (E.6), on fixe $\widetilde{p}_{w^m}(1) = 1$ étant donné que $\boldsymbol{A}$ ne dépend plus de $\boldsymbol{w}$ (équation (III.81)).
- Le terme $E_{\widetilde{p}_{\boldsymbol{w}}}\left[\log p(\boldsymbol{w} \,|\, \boldsymbol{\alpha})\right]$:

$$\begin{aligned} E_{\widetilde{p}_{\boldsymbol{w}}}\left[\log p(\boldsymbol{w} \,|\, \boldsymbol{\alpha})\right] &= \sum_m E_{\widetilde{p}_{w^m}}\left[\log p(w^m \,|\, \alpha^m)\right] \\ &= \sum_m \left\{\frac{1}{2}\log\frac{\alpha^m}{2\pi} - \frac{1}{2}\alpha^m E_{\widetilde{p}_{w^m}}\left[(w^m)^2\right]\right\} \\ &= \sum_m \left\{\frac{1}{2}\log\frac{\alpha^m}{2\pi} - \frac{1}{2}\alpha^m\left(m_{w^m}^2 + v_{w^m}\right)\right\} \end{aligned} \tag{E.12}$$

Annexe F

Rappels utiles

F.1 Algorithme de Metropolis-Hasting

Cet algorithme permet d'obtenir un échantillon d'une distribution de probabilité complexe $p(x)$ non simulable directement. Nous allons commencé par décrire l'algorithme de Metropolis qui a ensuite été généralisé pour obtenir celui de Metropolis-Hasting. Supposons que notre but est d'obtenir des réalisations d'une distribution $p(\theta) = \frac{f(\theta)}{C}$ où la constante de normalisation C est généralement inconnue et complexe. L'algorithme de Metropolis [Metropolis 1949, Metropolis 1953] génère une séquence de réalisations de cette distribution selon les étapes suivantes :

1- On commence par une valeur initiale θ_0 où $f(\theta_0) > 0$

2- A l'itération r :
   - En utilisant la valeur courante de θ_{r-1} on calcul une valeur candidate θ^* à partir d'une distribution candidate simulable $g(\theta_1, \theta_2)$ représentant la probabilité de retourner la valeur θ_2 sachant la valeur précédente θ_1. La seule contrainte sur le choix de cette densité est la symmetrie $g(\theta_1, \theta_2) = g(\theta_2, \theta_1)$.
   - Etant donné la valeur candidate θ^* et courante θ_{r-1} on calcule la probabilité d'acceptation :

$$\alpha = \min\left(\frac{p(\theta^*)}{p(\theta_{r-1})}, 1\right) = \min\left(\frac{f(\theta^*)}{f(\theta_{r-1})}, 1\right) \tag{F.1}$$

   On accepte la valeur candidate avec la probabilité α en simulant u selon une loi uniforme sur $[0, 1]$.

On obtient donc une chaîne de Markov $(\theta_0, \theta_1, ..., \theta_r, ..., \theta_R)$ où la probabilité de transition de θ_r à θ_{r+1} dépend uniquement de θ_r. Après une periode de chauffe suffisante R_0, on peut considérer $(\theta_{R_0+1}, ..., \theta_R)$ comme un échantillon de $p(x)$. [Hastings 1970] a généralisé l'algorithme de Metropolis en utilisant une distribution candidate arbitraire $g(\theta_1, \theta_2) = \Pr(\theta_1 \to \theta_2)$. La probabilité d'acceptation d'une valeur candidate θ^* devient alors :

$$\alpha = \min\left(\frac{f(\theta^*)g(\theta^*, \theta_{t-1})}{f(\theta_{t-1})g(\theta_{t-1}, \theta^*)}, 1\right) \tag{F.2}$$

En supposant que $g(\theta_1, \theta_2)$ est symmetrique on revient à l'algorithme de Metropolis.

F.2 Distribution gaussienne tronquée

Dans l'approche DEC basée sur l'inférence MCMC [Vincent 2010a], la distribution gaussienne tronquée est la loi instrumentale de l'algorithme de Metropolis-Hasting utilisée pour la simulation de $\boldsymbol{\xi}$, le paramètre de régularité spatiale du champ de Ising. Ci-dessous la définition de la densité de probabilité de la gaussienne tronquée :

$$x \sim \mathcal{N}_{[a,b]}\left(\mu, \eta^2\right) \quad \text{avec} \quad -\infty \leq a < b \leq \infty$$
$$f\left(x; \mu, \eta, a, b\right) = \frac{1/\eta}{\Phi\left(\frac{b-\mu}{\eta}\right) - \Phi\left(\frac{a-\mu}{\eta}\right)} \varphi\left(\frac{x-\mu}{\eta}\right)$$

où
$\varphi\left(.\right)$ est la densité de probabilité de la loi normale et $\Phi\left(.\right)$ est sa fonction de répartition.
$\varphi\left(\frac{x-\mu}{\eta}\right)$ étant une loi normale centrée réduite, on peut écrire la fonction de répartition $\Phi\left(.\right)$ à l'aide de la fonction d'erreur erf $\left(\operatorname{erf}\left(x\right) = \frac{2}{\sqrt{\pi}} \int_0^x \exp\left(-y^2\right)\ dy\right)$:

$$\Phi\left(\frac{x-\mu}{\eta}\right) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x-\mu}{\sqrt{2}\,\eta}\right)\right]$$

<u>Démonstration :</u>

$$\Phi\left(z\right) = \int_{-\infty}^{z} \varphi\left(t\right)\ dt$$

avec $\varphi(t) \sim \mathcal{N}(0,1)$ on a :

$$\begin{aligned} &= \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) \, dt \\ &= \frac{1}{\sqrt{2\pi}} \left[\int_{-\infty}^{0} \exp\left(-\frac{t^2}{2}\right) \, dt + \int_{0}^{z} \exp\left(-\frac{t^2}{2}\right) \, dt \right] \end{aligned} \tag{F.3}$$

Selon l'intégrale de Gauss $\left(\int_{-\infty}^{\infty} \exp\left(-\alpha x^2\right) \, dx = \sqrt{\frac{\pi}{\alpha}}\right)$:

$$\begin{aligned} \int_{-\infty}^{0} \exp\left(-\frac{t^2}{2}\right) \, dt &= \frac{1}{2} \int_{-\infty}^{\infty} \exp\left(-\frac{t^2}{2}\right) \, dt \quad \text{(intégrale pair)} \\ &= \frac{1}{2} \sqrt{\frac{\pi}{1/2}} = \sqrt{\frac{\pi}{2}} \end{aligned} \tag{F.4}$$

Selon la définition de la fonction d'erreur $\left(\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp\left(-y^2\right) \, dy\right)$

$$\begin{aligned} \int_{0}^{z} \exp\left(-\frac{t^2}{2}\right) \, dt &= \int_{0}^{z} \exp\left(-\left(\frac{t}{\sqrt{2}}\right)^2\right) \, dt \\ &= \sqrt{2} \int_{0}^{z/\sqrt{2}} \exp\left(-y^2\right) \, dy \qquad \text{avec} \quad y = \frac{t}{\sqrt{2}} \\ &= \sqrt{2} \, \frac{\operatorname{erf}\left(\frac{z}{\sqrt{2}}\right)}{2/\sqrt{\pi}} \\ &= \sqrt{\frac{\pi}{2}} \operatorname{erf}\left(\frac{z}{\sqrt{2}}\right) \end{aligned} \tag{F.5}$$

En remplaçant les Eqs. (F.4) et (F.5) dans (F.3) on trouve :

$$\begin{aligned} \Phi(z) &= \frac{1}{\sqrt{2\pi}} \left[\sqrt{\frac{\pi}{2}} + \sqrt{\frac{\pi}{2}} \operatorname{erf}\left(\frac{z}{\sqrt{2}}\right) \right] \\ &= \frac{1}{2} \left[1 + \operatorname{erf}\left(\frac{z}{\sqrt{2}}\right) \right] \end{aligned} \tag{F.6}$$

On peut donc écrire la distribution gaussienne tronquée sous la forme suivante :

$$f(x; \mu, \eta, a, b) = \frac{1/\eta}{\frac{1}{2}\left[\operatorname{erf}\left(\frac{b-\mu}{\sqrt{2}\,\eta}\right) - \operatorname{erf}\left(\frac{a-\mu}{\sqrt{2}\,\eta}\right)\right]} \mathcal{N}\left(\frac{x-\mu}{\eta}\right)$$

Cette dernière est celle utilisée dans l'approche DEC développée dans les travaux de thèse de Thomas Vincent [Vincent 2010a].

F.3 *a priori* de Jeffrey

Ci-dessous la définition et la démonstration de l'*a priori* de Jeffrey choisi comme *a priori* sur la variance de bruit ($\boldsymbol{\sigma}^2$), celle de la FRH (v_h) et celle des coefficients des dérives basses fréquences (v_l), dans l'approche DEC basée sur l'inférence MCMC. L'*a priori* de Jeffrey sur un paramètre θ est un *a priori* non informatif défini à partir de l'informaiton de Fisher $I_f(\theta)$ comme suit :

$$p(\theta) \propto \sqrt{I_f(\theta)} = \sqrt{-\mathbb{E}_{p_{\boldsymbol{x}}}\left[\frac{d^2 \log p(X \mid \theta)}{d\theta^2}\right]} \tag{F.7}$$

Supposons $\boldsymbol{x}$ une variable aléatoire de dimension $C \times 1$ et de distribution gaussienne $\mathcal{N}(\boldsymbol{x}; \boldsymbol{m}, \sigma^2 \boldsymbol{V})$. Nous allons calculer l'*a priori* de Jeffrey sur la variance σ^2.

$$p_{\boldsymbol{x}}(\boldsymbol{x} \mid \sigma^2) = \frac{1}{(2\pi)^{C/2} |\sigma^2 \boldsymbol{V}|^{1/2}} \exp\left(-\frac{(\boldsymbol{x}-\boldsymbol{m})^{\mathrm{t}} \boldsymbol{V}^{-1} (\boldsymbol{x}-\boldsymbol{m})}{2\sigma^2}\right) \tag{F.8}$$

$$\frac{d \log p(\boldsymbol{x} \mid \sigma^2)}{d\sigma^2} = -\frac{C}{2\sigma^2} + \frac{(\boldsymbol{x}-\boldsymbol{m})^{\mathrm{t}} \boldsymbol{V}^{-1} (\boldsymbol{x}-\boldsymbol{m})}{2\sigma^4} \tag{F.9}$$

$$\frac{d^2 \log p(\boldsymbol{x} \mid \sigma^2)}{(d\sigma^2)^2} = \frac{C}{2\sigma^4} - \frac{(\boldsymbol{x}-\boldsymbol{m})^{\mathrm{t}} \boldsymbol{V}^{-1} (\boldsymbol{x}-\boldsymbol{m})}{\sigma^6} \tag{F.10}$$

$$\begin{aligned}
\mathbb{E}_{p_{\boldsymbol{x}}}\left[\frac{d^2 \log p(\boldsymbol{x} \mid \sigma^2)}{(d\sigma^2)^2}\right] &= \frac{C}{2\sigma^4} - \frac{\mathbb{E}_{p_{\boldsymbol{x}}}[(\boldsymbol{x}-\boldsymbol{m})^{\mathrm{t}} \boldsymbol{V}^{-1} (\boldsymbol{x}-\boldsymbol{m})]}{\sigma^6} \\
&= \frac{C}{2\sigma^4} - \frac{\mathbb{E}_{p_{\boldsymbol{x}}}[(\boldsymbol{x}-\boldsymbol{m})^{\mathrm{t}}] \boldsymbol{V}^{-1} \mathbb{E}_{p_{\boldsymbol{x}}}[\boldsymbol{x}-\boldsymbol{m}] + \mathrm{tr}\,(\sigma^2 \boldsymbol{V} \boldsymbol{V}^{-1})}{\sigma^6} \\
&= \frac{C - 2\,\mathrm{tr}\,(\boldsymbol{I}_C)}{2\sigma^4}
\end{aligned} \tag{F.11}$$

avec $\mathbb{E}_{p_{\boldsymbol{x}}}[(\boldsymbol{x}-\boldsymbol{m})^{\mathrm{t}}] = \mathbb{E}_{p_{\boldsymbol{x}}}[\boldsymbol{x}-\boldsymbol{m}] = 0$.

$$p(\sigma^2) = \sqrt{\frac{C}{2\sigma^4}} \propto \frac{1}{\sigma^2} \tag{F.12}$$

F.4 Algorithme de dichotomie

Dans l'approche DEC *parcimonieuse* utilisant l'inférence VEM, nous avons eu recourt à l'algorithme de dichotomie pour l'estimation du paramètre τ_2, l'abscisse

du point d'inflexion de la sigmoïde, ainsi que $\boldsymbol{\mu}_1$, la moyenne de la gaussienne représentant la classe des voxels activés. Cet algorithme permet d'effectuer une recherche de solution en divisant itérativement l'espace de recherche en deux parties (de la même taille pour un nombre d'itération optimal), d'où son nom (division en deux ; en grec). Il permet par exemple de chercher le zéro $\hat{x}$ d'une fonction $f(x)$ à une précision donnée ϵ, selon les étapes suivantes :

1- Initialisation : x_1, x_2 tel que $f(x_1)f(x_2) < 0$

2- Calcul de $x_3 = \frac{x_1+x_2}{2}$

3- Si $\quad |f(x_3)| \leq \epsilon \Rightarrow \hat{x} = x_3$

$$\text{Sinon} \begin{cases} \text{Si } f(x_1)f(x_3) < 0 \text{ la solution est dans l'intervalle } [x_1, x_3] \Rightarrow x_2 = x_3 \\ \text{Sinon, la solution est dans l'intervalle } [x_3, x_2] \Rightarrow x_1 = x_3 \end{cases}$$

4- Retour à l'étape 2.

La figure F.1 donne un apperçu de l'algorithme en montrant le positionnement des principaux termes cités ci-dessus $x_1, x_2, f(x_1)$ et $f(x_2)$.

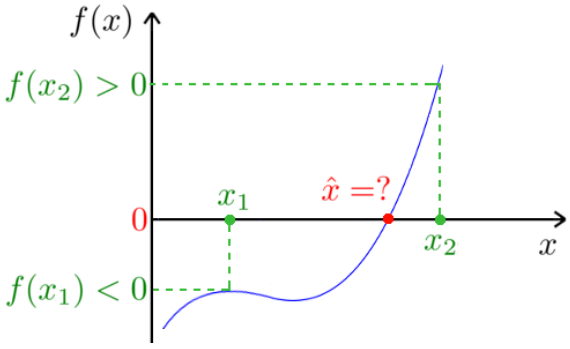


FIGURE F.1 – Illustration de l'algorithme itératif de dichotomie visant à trouver $\hat{x}$ tel que $f(\hat{x}) = 0$ à une précision ϵ.

F.5 Algorithme du gradient

L'algorithme du gradient a été utilisé pour l'estimation du paramètre de régularité spatial $\boldsymbol{\xi}$ dans l'approche DEC basée sur l'inférence VEM. C'est un algorithme d'optimisation permettant la minimisation d'une fonction différentiable $f(x)$. Le but étant l'estimation de la valeur de x correspondant au minimum $f_{\min}(x)$ et pour laquelle le gradient $\bigtriangledown f(x)$ (la pente de la tangente au point x) est nul. Ceci en suivant les étapes :

1- Initialisation : x_0, $\epsilon \geq 0$ (seuil de tolérance)

2- A l'itération t : Calcul de $\bigtriangledown f(x_t)$

3- Si $|| \bigtriangledown f(x_t)|| \leq \epsilon$ alors $\hat{x} = x_t$
Sinon calcul de $x_{t+1} = x_t - \alpha \bigtriangledown f(x_t)$ et retour à l'étape 2.

Où α est le pas de l'algorithme qui ne doit être ni trop petit pour éviter un retard de convergence, ni trop grand pour ne pas osciller autour du minimum. Dans l'approche DEC, ce pas a été fixé à 0.003.

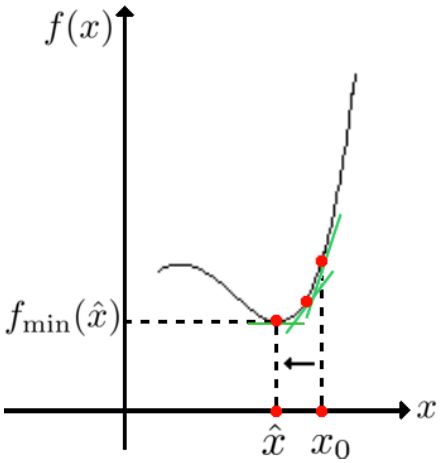


FIGURE F.2 – Illustration de l'algorithme de descente du gradient permettant de minimiser une fonction $f(x)$ en trouvant la valeur $\hat{x}$ correspondante. Ceci en partant d'une valeur initiale x_0 qui est ensuite modifiée itérativement avec un certain pas jusqu'à ce que le gradient (la pente de la tangente (en vert) au point x) devienne inférieure à une valeur ϵ prédéfinie.

F.6 Entropie

Ci-dessous les définitions des entropies des variables gaussiennes et binaires nécessaires au calcul de l'énergie libre dans l'approche DEC basée sur l'inférence VEM :

♣ Entropie d'une variable gaussienne :

Soit $\boldsymbol{x}$ un vecteur de R^N suivant une loi gaussienne $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. L'entropie est donnée par l'équation suivante :

$$I\left[p(\boldsymbol{x})\right] = -E_{p(\boldsymbol{x})}\left[\log p(\boldsymbol{x})\right] = \log\left(\sqrt{(2\,\pi e)^N\,|\boldsymbol{\Sigma}|}\right) \tag{F.13}$$

Dans le cas particulier où x est une variable aléatoire gaussienne de loi $\mathcal{N}(\mu, \sigma^2)$, l'entropie est :

$$I\left[p(x)\right] = -E_{p(x)}\left[\log p(x)\right] = \log\left(\sqrt{2\,\pi e\,\sigma^2}\right) \tag{F.14}$$

♣ **Entropie d'une variable binaire :**

Soit x une variable binaire de loi $p(x)$. L'entropie de cette variable est définie de la manière suivante :

$$I\left[p(x)\right] = -E_{p(x)}\left[\log p(x)\right] = -\sum_{x=0}^{1} p(x)\log p(x) \tag{F.15}$$

F.7 Échantillonneur de Gibbs

Utilisé dans le cadre de l'approche DEC, l'échantillonneur de Gibbs représente une technique particulière des méthodes MCMC. Introduit dans le contexte du traitement d'image par [Geman 1984] cet échantillonneur permet de simuler un échantillon de variables aléatoires x, y, z (3 dans notre exemple) de loi jointe $p(x, y, z)$ complexe à partir de leurs distributions conditionnelles univariées $p(x \mid y, z)$, $p(y \mid x, z)$ et $p(z \mid x, y)$ ayant généralement des formes analytiques moins complexes que $p(x, y, z)$. L'idée repose donc sur la simulation successive de n variables aléatoires (ici $n = 3$) à partir de n distributions conditionnelles univariées au lieu de générer un vecteur de taille n de variables aléatoires à partir de leur distribution jointe complexe. Ceci s'opère à travers les étapes suivantes :

1- Initialisation : $x^{(0)}$, $y^{(0)}$, $z^{(0)} \sim p(z \mid x^{(0)}, y^{(0)})$

2- Itération r :

$$\begin{aligned} x^{(r)} &\sim p\left(x \mid y^{(r-1)}, z^{(r-1)}\right) \\ y^{(r)} &\sim p\left(y \mid x^{(r)}, z^{(r-1)}\right) \\ z^{(r)} &\sim p\left(z \mid x^{(r)}, y^{(r)}\right) \end{aligned}$$

La suite obtenue $\{x^{(r)}, y^{(r)}, z^{(r)}\}_{r=R_0+1:R}$ forme une chaîne de Markov dont la loi stationnaire est la loi $p(x, y, z)$ que l'on cherche à simuler. R_0 est le nombre d'itérations représentant la période de chauffe et dont l'élimination permet de se rapprocher de l'indépendance par rapport aux valeurs initiales et de la loi stationnaire. Dans certains cas, la loi d'un groupe de variables ((x, y) par exem-

ple) peut être simulable facilement ce qui permet de grouper les deux étapes de l'échantillonneur simulant x et y en une et donc d'accélerer la convergence :

$$\begin{aligned}\left(x^{(r)}, y^{(r)}\right) &\sim p\left(x, y \mid z^{(r-1)}\right)\\ z^{(r)} &\sim p\left(z \mid x^{(r)}, y^{(r)}\right)\end{aligned}$$

Etant donné que la loi $p(x, y \mid z)$ est égale à $p(x \mid y, z)p(y \mid z)$ où $p(y \mid z) = \int p(x, y \mid z)dx$, la simulation peut également être effectuée selon les étapes suivantes :

$$\begin{aligned}y^{(r)} &\sim p\left(y \mid z^{(r-1)}\right)\\ x^{(r)} &\sim p\left(x \mid y^{(r)}, z^{(r-1)}\right)\\ z^{(r)} &\sim p\left(z \mid x^{(r)}, y^{(r)}\right)\end{aligned}$$

On peut donc ou bien simuler en une seule étape le couple (x, y) ou bien en deux étapes en simulant $y \mid z$ et ensuite $x \mid y, z$. C'est le deuxième cas qui est effectué dans l'approche DEC pour la simulation du couple (a_j^m, q_j^m).

Liste des publications issues de cette thèse

(1) C. Bakhous, F. Forbes, T. Vincent, L. Chaari, M. Dojat and P. Ciuciu. Sélection de variables dans un cadre bayésien de traitement de données d'IRM fonctionnelle. In Journées de Statistique de la Société Française de Statistique (SFdS), Brussels, Belgium, May 21-25 2012.

(2) C. Bakhous, F. Forbes, T. Vincent, L. Chaari, M. Dojat and P. Ciuciu. Adaptive experimental condition selection in event-related fMRI. In 9th Proc. IEEE ISBI, pages 1755–1758, Barcelona, Spain, May 2012.

(3) C. Bakhous, F. Forbes, F. Enikeeva, T. Vincent, M. Dojat and P. Ciuciu. Analyse parcimonieuse des données d'IRM fonctionnelle dans un cadre bayésien variationnel. In Journées de Statistique de la Société Française de Statistique (SFdS), Toulouse, France, May 27-31 2013.

(4) C. Bakhous, F. Forbes, T. Vincent, M. Dojat and P. Ciuciu. Variational variable selection to assess experimental condition relevance in event-related fMRI. In 10th Proc. IEEE ISBI, pages 1500–1503, San Francisco, CA, Apr. 2013.

Bibliographie

[Aguirre 1998] G. K. Aguirre, E. Zarahn et M. D'Esposito. *The variability of human BOLD hemodynamic responses.* Neuroimage, vol. 8, no. 4, pages 360–369, Nov. 1998. 46

[Ashburner 2003] J. Ashburner, K.J. Friston et W. Penny, editors. Human brain function, 2nd edition. Academic press, 2003. 31, 34, 38, 39, 40, 43, 44, 45, 50, 53

[Ashburner 2007] John Ashburner. *A fast diffeomorphic image registration algorithm.* Neuroimage, vol. 38, no. 1, pages 95–113, Oct. 2007. 33

[Badillo 2013a] S. Badillo. *Etude de la variabilité hémodynamique chez l'enfant et l'adulte sains en IRMf.* PhD thesis, Université de Paris–Sud, Orsay, France, 2013. 160

[Badillo 2013b] S. Badillo, T. Vincent et P. Ciuciu. *Group-level impacts of within- and between-subject hemodynamic variability in fMRI.* to appear in Neuroimage, (j.neuroimage.2013.05.100), June 2013. 134

[Bakhous 2012a] C. Bakhous, F. Forbes, T. Vincent, L. Chaari, M. Dojat et P. Ciuciu. *Sélection de variables dans un cadre bayésien de traitement de données d'IRM fonctionnelle.* In Journées de Statistique de la Société Française de Statistique (SFdS), Brussels, Belgium, May 21-25 2012. 16

[Bakhous 2012b] C. Bakhous, F. Forbes, T. Vincent, L. Chaari, M. Dojat et P. Ciuciu. *Adaptive experimental condition selection in event-related fMRI.* In 9th Proc. IEEE ISBI, pages 1755–1758, Barcelona, Spain, May 2012. 16

[Bakhous 2013a] C. Bakhous, F. Forbes, F. Enikeeva, T. Vincent, M. Dojat et P. Ciuciu. *Analyse parcimonieuse des données d'IRM fonctionnelle dans un cadre bayésien variationnel* . In Journées de Statistique de la Société Française de Statistique (SFdS), Toulouse, France, May 27-31 2013. 16

[Bakhous 2013b] C. Bakhous, F. Forbes, T. Vincent, M. Dojat et P. Ciuciu. *Variational variable selection to assess experimental condition relevance in event-related fMRI.* In 10th Proc. IEEE ISBI, pages 1500–1503, San Francisco, CA, Apr. 2013. 16

[Bandettini 1993] P. A. Bandettini, A. Jesmanowicz, E. C. Wong et J. S. Hyde. *Processing strategies for time-course data sets in functional MRI of the human brain.* Magn. Reson. Med., vol. 30, pages 161–173, 1993. 27

[Bannister 2004] P.R. Bannister, J.M. Brady et M. Jenkinson. *TIGER – A New Model for Spatio-temporal Realignment of FMRI Data.* In Computer Vision and Mathematical Methods in Medical and Biomedical Image Analysis, volume 3117 of *Lecture Notes in Computer Science*, pages 292–303. Springer Berlin / Heidelberg, 2004. 32

[Benjamini 1995] Y. Benjamini et Y. Hochberg. *Controlling the false discovery rate : a practical and powerful approach to multiple testing.* Journal of the Royal Statistical Society, vol. 57, no. 1, pages 289–300, 1995. 44

[Benjamini 2001] Y. Benjamini et D. Yekutieli. *The control of the false discovery rate in multiple testing under dependency.* The Annals of Statistics, vol. 29, no. 4, pages 1165–1188, 2001. 44

[Biswal 1995] B. Biswal, F. Z. Yetkin, V. M. Haughton et J. S. Hyde. *Functional connectivity in the motor cortex of resting human brain using echo-planar MRI.* Magn. Reson. Med., vol. 34, no. 4, pages 537–541, Oct. 1995. 144

[Boynton 1996] G. M. Boynton, S. A. Engel, G. H. Glover et D. J. Heeger. *Linear systems analysis of functional magnetic resonance imaging in human V1.* J. Neurosci., vol. 16, no. 13, pages 4207–4221, July 1996. 46

[Buckner 1998] R. L. Buckner, J. Goodman, M. Burock, M. Rotte, W. Koutstaal, D. L. Schachter, B. R. Rosen et A. M. Dale. *Functional-anatomic correlates of object priming in humans revealed by rapid presentation event-related f*MRI. Neuron, vol. 20, pages 285–296, Feb. 1998. 46

[Buxton 1997] R. Buxton et L. Frank. *A model for the coupling between cerebral blood flow and oxygen metabolism during neural stimulation.* J. Cereb. Blood Flow Metab., vol. 17, no. 1, pages 64–72, Jan. 1997. 46

[Buxton 1998] R. B. Buxton, E. C. Wong et Frank. L. R. *Dynamics of blood flow and oxygenation changes during Brain activation : the balloon model.* Magn. Reson. Med., vol. 39, pages 855–864, June 1998. 46

[Cachia 2003] A. Cachia, J.-F. Mangin, D. Rivière, D. Papadopoulos-Orfanos, F. Kherif, I. Bloch et J. Régis. *A generic framework for parcellation of the cortical surface into gyri using geodesic Voronoi diagrams.* Medical Image Analysis, vol. 7, no. 4, pages 403–416, 2003. 61

[Casanova 2008] R. Casanova, S. Ryali, J. Serences, L. Yang, R. Kraft, P.J. Laurienti et J.A. Maldjian. *The impact of temporal regularization on estimates of the BOLD hemodynamic response function : a comparative analysis.* Neuroimage, vol. 40, no. 4, pages 1606–1618, May 2008. 53

[Casanova 2009] R. Casanova, L. Yang, W. D. Hairston, P. J. Laurienti et J. A. Maldjian. *Evaluating the impact of spatio-temporal smoothness constraints on the BOLD hemodynamic response function estimation : an analysis based on Tikhonov regularization.* Physiol Meas, vol. 30, no. 5, pages N37–N51, May 2009. 53

[Castaño-Martínez 2005] A. Castaño-Martínez et F. López-Blázquez. *Distribution of a sum of weighted noncentral chi-square variables.* An Official Journal of the Spanish Society of Statistics and Operations Research, vol. 14, no. 2, pages 397–415, Dec. 2005. 153

[Celeux 2003] G. Celeux, F. Forbes et N. Peyrard. *EM procedures using mean field-like approximations for Markov model-based image segmentation.* Pat. Rec., vol. 36, no. 1, pages 131–144, 2003. 88, 199

[Chaari 2012] L. Chaari, F. Forbes, T. Vincent et P. Ciuciu. *Hemodynamic-informed parcellation of fMRI data in a variational joint detection estimation framework.* In Proc. 15th MICCAI, LNCS 7512, pages 180–188, Nice, France, Oct. 2012. Springer Verlag Berlin Heidelberg. 61

[Chaari 2013] L. Chaari, T. Vincent, F. Forbes, M. Dojat et P. Ciuciu. *Fast joint detection-estimation of evoked brain activity in event-related fMRI using a variational approach.* IEEE Trans. Med. Imag., vol. 32, no. 5, pages 821–837, May 2013. 70, 79, 81, 156, 196

[Ciuciu 2003] P. Ciuciu, J.-B. Poline, G. Marrelec, J. Idier, Ch. Pallier et H. Benali. *Unsupervised robust non-parametric estimation of the hemodynamic response function for any fMRI experiment.* IEEE Trans. Med. Imag., vol. 22, no. 10, pages 1235–1251, Oct. 2003. 52, 53, 54

[Ciuciu 2004] P. Ciuciu, J. Idier, A. Roche et Ch. Pallier. *Outlier detection for robust region-based estimation of the Hemodynamic Response Function in*

event-related fMRI. In 2th Proc. IEEE ISBI, pages 392–395, Arlington, VA, Apr. 2004. 53, 54

[Cohen 1997] Mark S. Cohen. *Parametric analysis of* MRI *data using linear systems methods.* Neuroimage, vol. 6, pages 93–103, 1997. 46

[de Pasquale 2008] F. de Pasquale, C. Del Gratta et G. L. Romani. *Empirical Markov Chain Monte Carlo Bayesian analysis of MRI data.* Neuroimage, vol. 42, no. 1, pages 99–111, Aug. 2008. 47

[Donnet 2004] S. Donnet. *Inversion de données IRMf : estimation et sélection de modèles.* thèse de doctorat, Université de Toulouse, 2004. 14, 64

[Dyk 2008] D.A. Van Dyk et T. Park. *Partially Collapsed Gibbs Samplers : Theory and Methods.* Journal of the American Statistical Association, vol. 103, pages 790–796, 2008. 73

[Feinberg 2010] DA. Feinberg, S. Moeller, SM. Smith, E. Auerbach, S. Ramanna, M. Gunther, MF. Glasser, KL. Miller, K. Ugurbil et E. Yacoub. *Multiplexed echo planar imaging for sub-second whole brain FMRI and fast diffusion imaging.* PLoS One, vol. 5, no. 12, page e15710, Dec. 2010. 135

[Flandin 2002a] G. Flandin, F. Kherif, X. Pennec, G. Malandain, N. Ayache et J.-B. Poline. *Improved detection sensitivity of functional MRI data using a brain parcellation technique.* In Proc. 5th MICCAI, LNCS 2488 (Part I), pages 467–474, Tokyo, Japan, Sep. 2002. Springer Verlag. 61

[Flandin 2002b] G. Flandin, F. Kherif, X. Pennec, D. Rivière, N. Ayache et J.-B. Poline. *Parcellation of brain images with anatomical and functional constraints for fMRI data analysis.* In Proc. 1st Proc. IEEE ISBI, pages 907–910, Washington, DC, July 2002. 61

[Fonov 2009] VS Fonov, AC Evans, RC McKinstry, CR Almli et DL Collins. *Unbiased nonlinear average age-appropriate brain templates from birth to adulthood.* NeuroImage, vol. 47, no. Supplement 1, pages S102 – S102, 2009. Organization for Human Brain Mapping 2009 Annual Meeting. 33

[Forbes 2003] F. Forbes et N. Peyrard. *Hidden Markov Random Field Model Selection Criteria Based on Mean Field-Like Approximations.* IEEE Trans. Pattern Anal. Mach. Intell., vol. 25, no. 9, pages 1089–1101, Sep. 2003. 199, 202

[Friston 1994] K.J. Friston, P. Jezzard et R. Turner. *Analysis of functional MRI time-series.* Hum. Brain Mapp., vol. 1, pages 153–171, 1994. 46

[Friston 1998] K. Friston, P. Fletcher, A. Joseph O. andHolmes, M. Rugg et R. Turner. *Event-related responses in fMRI : characterising differential responses.* Neuroimage, vol. 7, pages 30–40, 1998. 39, 46

[Friston 2002] K.J. Friston, D. E. Glaser, R. N. A. Henson, S. Kiebel, C. Phillips et J. Ashburner. *Classical and Bayesian Inference in Neuroimaging : Applications.* Neuroimage, vol. 16, no. 2, pages 484–512, 2002. 45

[Geman 1984] S. Geman et D. Geman. *Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images.* IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. PAMI-6, no. 6, pages 721–741, Nov. 1984. 220

[Glover 1999] G. H. Glover. *Deconvolution of impulse response in event-related BOLD fMRI.* Neuroimage, vol. 9, pages 416–429, 1999. 46

[Goutte 2000] C. Goutte, F. A. Nielsen et L. K. Hansen. *Modeling the haemodynamic response in fMRI using smooth filters.* IEEE Trans. Med. Imag., vol. 19, no. 12, pages 1188–1201, Dec. 2000. 52

[Handwerker 2004] D. A. Handwerker, J.M. Ollinger et M. D'Esposito. *Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses.* Neuroimage, vol. 21, pages 1639–1651, Apr. 2004. 46

[Hastings 1970] W. K. Hastings. *Monte Carlo sampling methods using Markov chains and their applications.* Biometrika, vol. 57, no. 1, pages 97–109, April 1970. 215

[Ji 2010] Y. Ji. *Data-driven fMRI data analysis based on parcellation.* PhD thesis, University of Nottingham, Oct. 2010. 61

[Joseph 1999] O. Joseph et R. N.A. Henson. *Event-related functional magnetic resonance imaging : modelling, inference and optimization.* Philos. Trans. Roy. Soc. London Ser. B : Biol., vol. 354, no. 1387, pages 1215–1228, 1999. 28

[Knops 2009] A. Knops, B. Thirion, E. M. Hubbard, V. Michel et S. Dehaene. *Recruitment of an area involved in eye movements during mental arithmetic.* Science, vol. 324, pages 1583–5, 2009. 149

[Lindquist 2009] M. Lindquist, J. Meng Loh, L.Y. Atlas et T.D. Wager. *Modeling the hemodynamic response function in fMRI : efficiency, bias and mis-*

modeling. Neuroimage, vol. 45, no. 1 Suppl, pages S187–98, Mar. 2009. 48

[Luo 2005] H. Luo et S. Puthusserypady. *A Sparse Bayesian Method for Determination of Flexible Design Matrix for fMRI Data Analysis*. IEEE Trans. on Circuits and systems–I : regular papers, vol. 52, no. 12, pages 2699–2706, 2005. 15, 64, 65, 66, 67

[Makni 2005] S. Makni, P. Ciuciu, J. Idier et J.-B. Poline. *Joint detection-estimation of brain activity in functional MRI : a multichannel deconvolution solution*. IEEE Trans. Signal Processing, vol. 53, no. 9, pages 3488–3502, Sep. 2005. 54, 59

[Makni 2006] S. Makni. *Détection-estimation conjointe de l'activité cérébrale en IRMf*. PhD thesis, Université de Paris–Sud, Orsay , France, Nov. 2006. 54, 57, 70

[Makni 2008] S. Makni, J. Idier, T. Vincent, B. Thirion, G. Dehaene-Lambertz et P. Ciuciu. *A fully Bayesian approach to the parcel-based detection-estimation of brain activity in fMRI*. Neuroimage, vol. 41, no. 3, pages 941–969, July 2008. 54

[Mandeville 1999] J.B. Mandeville, J.J.A. Marota, C. Ayata, C. Zaharchuck, M.A. Moskowitz, B.R. Rosen et R.M Weisskoff. *Evidence of a cerebrovascular post-arteriole Windkessel with delayed compliance*. J. Cereb. Blood Flow Metab., vol. 19, pages 679–689, 1999. 46

[Marrelec 2001] G. Marrelec, H. Benali, P. Ciuciu et 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, Aug. 2001. MaxEnt Workshops. 52

[Marrelec 2003] G. Marrelec, H. Benali, P. Ciuciu, M. Pélégrini-Issac et J.-B. Poline. *Robust Bayesian estimation of the hemodynamic response function in event-related BOLD MRI using basic physiological information*. Hum. Brain Mapp., vol. 19, no. 1, pages 1–17, May 2003. 52

[Metropolis 1949] N. Metropolis et S. Ulam. *The Monte Carlo Method*. J. Amer. Statist. Assoc. 44, pages 335–341, 1949. 214

[Metropolis 1953] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller et E. Teller. *Equation of State Calculations by Fast Computing*

Machines. Journal of Chemical Physics, vol. 21, no. 6, pages 1087–1092, June 1953. 214

[Miezin 2000] F. M. Miezin, L. Maccotta, J. M. Ollinger, S. E. Petersen et R. L. Buckner. *Characterizing the hemodynamic response : effects of presentation rate, sampling procedure, and the possibility of ordering brain activity based on relative timing.* Neuroimage, vol. 11, pages 735–759, 2000. 46

[Ogawa 1990] S. Ogawa, T. Lee, A. Kay et D. Tank. *Brain Magnetic Resonance Imaging with contrast dependent on blood oxygenation.* Proc. Natl. Acad. Sci. USA, vol. 87, no. 24, pages 9868–9872, 1990. 27

[Ogawa 1992] S. Ogawa, D.W. Tank, R. Menon, J.M. Ellermann, S-G. Kim, H. Merkle et K. Ugurbil. *Intrinsic signal changes accompanying sensory stimulation : functional brain mapping with magnetic resonance imaging.* Proc. Natl. Acad. Sci. USA, vol. 89, pages 5951–5955, 1992. 27

[OHara 2009] R.B. OHara et M.J. Sillanpaa. *A Review of Bayesian Variable Selection Methods : What, How and Which.* Bayesian Analysis, vol. 4, no. 1, pages 85–118, 2009. 65

[Oikonomou 2012] V.P. Oikonomou, K. Blekas et L. Astrakas. *A Sparse and Spatially Constrained Generative Regression Model for fMRI Data Analysis.* IEEE Trans. Biomed. Eng., vol. 59, no. 1, pages 58–67, 2012. 15, 51, 64, 65, 66, 67

[Park 2009] T. Park et D.A. Van Dyk. *Partially Collapsed Gibbs Samplers : Illustrations and Applications.* Journal of Computational and Graphical Statistics, vol. 18, pages 283–305, 2009. 73

[Penny 2003] W. D. Penny, S. Kiebel et K. J. Friston. *Variational Bayesian inference for fMRI time series.* Neuroimage, vol. 19, no. 3, pages 727–741, July 2003. 45

[Penny 2005] W. D. Penny, N. Trujillo-Barreto et K. J. Friston. *Bayesian fMRI time series analysis with spatial priors.* Neuroimage, vol. 23, no. 2, pages 350–362, 2005. 51

[Pereyra 2013] M. Pereyra, N. Dobigeon, H. Batatia et J.-Y. Tourneret. *Estimating the granularity coefficient of a Potts-Markov random field within an MCMC algorithm.* IEEE Trans. on Image Process., vol. 22, no. 6, pages 2385–2397, June 2013. 76

[Perrot 2009] M. Perrot. *Reconnaissance Automatique des Sillons Corticaux.* PhD thesis, ENS de Cachan, Cachan , France, Oct. 2009. 61

[Pinel 2007] P. Pinel, B. Thirion, S. Mériaux, A. Jobert, J. Serres, D. Le Bihan, J.-B. Poline et S. Dehaene. *Fast reproducible identification and large-scale databasing of individual functional cognitive networks.* BMC Neurosci., vol. 8, no. 1, page 91, Oct. 2007. 126

[Pruessmann 1999] K. P. Pruessmann, M. Weiger, M.W. Scheidegger et P. Boesiger. *SENSE : Sensitivity encoding for fast MRI.* Magn. Reson. Med., vol. 42, pages 952–962, Nov 1999. 126

[Purdon 1998] P. L. Purdon et R. M. Weisskoff. *Effect of temporal autocorrelation due to physiological noise and stimulus paradigm on voxel-level false-positive rates in fMRI.* Hum. Brain Mapp., vol. 6, no. 4, pages 239–249, 1998. 45

[Raftery 2007] Adrian E. Raftery, Michael A. Newton, Jaya M. Satagopan et Pavel N. Krivitsky. *Estimating the integrated likelihood via posterior simulation using the harmonic mean identity.* In J.M. Bernardo, M.J. Bayarri, O. Berger, A.P. David, D. Heckermann, A.F.M. Smith et M. West, editors, Bayesian statistics 8, pages 1–45. Oxford University Press, 2007. 58

[Rajapakse 1998] J. C. Rajapakse, F. Kruggel, J. M. Maisog et D.Y. Von Cramon. *Modeling Hemodynamic Response for Analysis of Functional* MRI *Time-Series.* Hum. Brain Mapp., vol. 6, pages 283–300, 1998. 47

[Rosen 1998] B. R. Rosen, R L. Buckner et A. M. Dale. *Event-related functional MRI : past, present and future.* Proc. Natl. Acad. Sci. USA, vol. 95, pages 773–780, 1998. 28

[Smith 2007] D. Smith et L. Fahrmeir. *Spatial Bayesian Variable Selection With Application to Functional Magnetic Resonance Imaging.* J. Amer. Statist. Assoc., vol. 102, no. 478, pages 417–431, 2007. 15, 51, 58

[Talairach 1988] J. Talairach et 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. 33

[Thirion 2005] B. Thirion, G. Flandin, P. Pinel et J.-B. Poline. *Spatial relaxation increases the sensitivity of random effects analyses : the benefit of inter-*

subject parcellation. In Proc. 11th HBMCD-RomNeuroimage vol 26 (1), Toronto, Canada, June12-16 2005. 61

[Thirion 2006] B. Thirion, S. Dodel et J.-B. Poline. *Detection of signal synchronizations in resting-state fMRI datasets.* Neuroimage, vol. 29, no. 1, pages 321–327, Jan. 2006. 61

[Thyreau 2006] B. Thyreau, B. Thirion, G. Flandin et J.-B. Poline. *Anatomo-functional description of the brain : a probabilistic approach.* In Proc. 31th Proc. IEEE ICASSP, volume V, pages 1109–1112, Toulouse, France, May 2006. 61, 148

[Tipping 2001] M.E. Tipping. *Sparse Bayesian learning and the relevance vector machine.* J. Mach. Learn. Res., vol. 1, pages 211–244, 2001. 64

[Tucholka 2008] A. Tucholka, B. Thirion, M. Perrot, P. Pinel, J.-F. Mangin et J.-B. Poline. *Probabilistic anatomo-functional parcellation of the cortex : how many regions ?* In 11thProc. MICCAI, LNCS Springer Verlag, New-York, USA, 2008. 61

[Turner 2002] Robert Turner. *How Much Cortex Can a Vein Drain ? Downstream Dilution of Activation-Related Cerebral Blood Oxygenation Changes.* Neuroimage, vol. 16, pages 1062–1067, 2002. 44

[Varoquaux 2013] G. Varoquaux et R. Cameron Craddock. *Learning and comparing functional connectomes across subjects.* NeuroImage, vol. 80, pages 405–415, oct. 2013. 61

[Vincent 2007] T. Vincent, P. Ciuciu et J. Idier. *Spatial mixture modelling for the joint detection-estimation of brain activity in fMRI.* In 32th Proc. IEEE ICASSP, volume I, pages 325–328, Honolulu, Hawaii, Apr. 2007. 51, 54, 59

[Vincent 2010a] T. Vincent. *Modèles hémodynamiques spatiaux adaptatifs pour l'imagerie cérébrale fonctionnelle.* PhD thesis, Université de Paris–Sud, Orsay , France, 2010. 51, 52, 54, 58, 59, 61, 70, 71, 76, 77, 181, 215, 216

[Vincent 2010b] T. Vincent, L. Risser et P. Ciuciu. *Spatially adaptive mixture modeling for analysis of within-subject fMRI time series.* IEEE Trans. Med. Imag., vol. 29, no. 4, pages 1059–1074, Apr. 2010. 51, 54, 59

[Woolrich 2002] M. Woolrich et S. Smith. *Spatio-temporal noise model selection in fMRI.* In Proc. 8th HBM, Sendai, Japan, June 2002. 45

[Woolrich 2005] M. Woolrich, T. Behrens, Ch. Beckmann et S. Smith. *Mixture models with adaptive spatial regularization for segmentation with an application to fMRI data.* IEEE Trans. Med. Imag., vol. 24, no. 1, pages 1–11, Jan. 2005. 44

[Worsley 2002] K.J. Worsley, C.H. Liao, J. Aston, V. Petre, G.H. Duncan, F. Morales et A.C. Evans. *A general statistical analysis for fMRI data.* Neuroimage, vol. 15, no. 1, pages 1–15, Jan. 2002. 45