Modèles bayésiens pour la détection de synchronisations au sein de signaux électro-corticaux

Maxime Rio

▶ **To cite this version:**

Maxime Rio. Modèles bayésiens pour la détection de synchronisations au sein de signaux électro-corticaux. Autre [cs.OH]. Université de Lorraine, 2013. Français. NNT : 2013LORR0090 . tel-01749865v1

HAL Id: tel-01749865

https://hal.univ-lorraine.fr/tel-01749865v1

Submitted on 29 Mar 2018 (v1), last revised 3 Oct 2013 (v3)

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 LORRAINE

École doctorale IAEM Lorraine

Modèles bayésiens pour la détection de synchronisations au sein de signaux électro-corticaux

THÈSE

présentée et soutenue publiquement le 16 juillet 2013

pour l'obtention du

Doctorat de l'université de Lorraine

(mention informatique)

par

Maxime Rio

Composition du jury

Président :	Fabrice Rossi, Professeur, Université Paris 1 Panthéon-Sorbonne
Rapporteurs :	Marc Gelgon, Professeur, Université de Nantes Maureen Clerc, Directrice de recherche, INRIA Sophia Antipolis
Examinateur :	Michael Wibral, Professeur, Université de Francfort
Directeurs de thèse :	Bernard Girau, Professeur, Université de Lorraine Axel Hutt, Chargé de recherche, INRIA Nancy-Grand Est

Laboratoire Lorrain de Recherche en Informatique et ses Applications — UMR 7503

Loria
Laboratoire lorrain de recherche en informatique et ses applications

Et... ne pas oublier de rêver...

Remerciements

Mes remerciements les plus chaleureux vont tout d'abord à mes encadrants, Axel Hutt et Bernard Girau, qui ont su m'apporter conseils et écoute au cours de ces années de thèse. Je tiens aussi à remercier les membres de mon jury, MM. Fabrice Rossi et Michael Wibral en tant qu'examinateurs et M. Marc Gelgon et Mme Maureen Clerc en tant que rapporteurs. Je souhaite témoigner une sincère reconnaissance envers ces derniers, pour leur patience et leurs encouragements.

Ma gratitude va aussi à ma famille et à mes amis proches, qui sont restés à mes côtés dans les bons et les moins bons moments. Pour avoir été les arcs-boutants de mon moral, pour tous ces bons petits plats que vous m'avez préparés, pour la relecture du présent document, et pour l'aide inestimable apportée à la confection du pot de thèse, je vous remercie Maman, Papa, Julien, Mathieu, Vincent, Jean-Charles, Jérémie, Lolita et Nicolas.

Au cours de mon séjour au Loria, j'ai croisé des personnes formidablement intéressantes, et cette thèse n'aurait pas la même saveur sans ces rencontres. Aussi, j'adresse mes remerciements à tous les collègues, passés et présents de l'équipe Cortex, et plus largement aux collègues du laboratoire et aux doctorants du campus de science, camarades du pique-nique du mercredi midi. Pour avoir pris le temps de la discussion, je tiens à exprimer une reconnaissance particulière à Bruno, Mathieu, Jean-Charles et Laure.

Enfin, je remercie Sophie, à qui cette thèse doit beaucoup et sans laquelle elle n'aurait pas pu se conclure.

Table des matières

Remerciements **iii**

1. Introduction générale **1**
1.1. Analyse de l'activité cérébrale aux échelles macroscopique et mésoscopique 1
1.1.1. Électrophysiologie du cerveau et enregistrements 1
1.1.2. Analyse fréquentielle de signaux cérébraux 3
1.1.3. Analyse temporelle de signaux cérébraux 5
1.1.4. Analyse temps-fréquence et synchronisations 6
1.1.5. Analyse de la phase et synchronisations 8
1.2. Présentation de la thèse 8
1.2.1. Motivation de la thèse 8
1.2.2. But de la thèse 9
1.2.3. Organisation de la thèse 9

2. Outils théoriques **11**
2.1. L'analyse temps-fréquence avec la transformée en ondelettes continue . . . 11
2.1.1. La transformée de Fourier 12
2.1.2. Transformée de Fourier à court terme 13
2.1.3. La transformée en ondelettes continue 14
2.1.4. L'ondelette de Morlet complexe 17
2.2. Modélisation bayésienne 19
2.2.1. Modèles bayésiens : distribution a priori des paramètres 19
2.2.2. Théorie de la décision, estimation et sélection de modèle 24
2.2.3. Calcul des probabilités a posteriori et marginale 27
2.3. Résumé 33

3. Statistiques dans le plan temps-fréquence et modèles de mélange **35**
3.1. Distribution des coefficients d'ondelettes de Morlet d'un signal aléatoire . 36
3.1.1. Distribution du module en un point temps-fréquence 36
3.1.2. Corrélation des coefficients adjacents 40
3.2. Modèles de mélange 41
3.2.1. Construction d'une distribution de mélange 41
3.2.2. Inférence bayésienne avec l'algorithme VMP pour un modèle de mélange 42
3.2.3. Utilisation pour partitionner des données 44
3.3. Modèle de mélange gaussien 46
3.3.1. Définition du modèle et inférence bayésienne par l'algorithme VMP 46

3.3.2. Application simple du modèle bayésien de mélange gaussien 47
3.4. Modèle de mélange ricien 47
3.4.1. Inférence bayésienne par l'algorithme VMP pour la distribution de Rice 49
3.4.2. Exemple de modèle ricien et application 51
3.4.3. Inférence bayésienne par l'algorithme VMP pour une distribution de mélange ricien 54
3.4.4. Exemple de modèle de mélange ricien et application 56
3.5. Banc d'essai des modèles bayésiens de mélange gaussien et ricien 58
3.5.1. Jeu de données artificielles et modèles évalués 59
3.5.2. Évaluation sur un petit jeu de données 60
3.5.3. Évaluation sur un grand jeu de données 62
3.5.4. Discussion sur l'utilisation des différents modèles de mélange . . . 63
3.6. Résumé 63

4. Méthodes d'analyse de synchronisations dans le plan temps-fréquence 65
4.1. Méthode d'analyse de synchronisations cooccurrentes par modèles de mélange gaussien 66
4.1.1. Détection des groupes d'électrodes à synchronisation cooccurrente 68
4.1.2. Validation des groupes d'électrodes détectés via une mesure de stabilité 69
4.2. Méthode d'analyse de synchronisations univariées par modèles de mélange ricien 72
4.2.1. Détection des états hauts/bas dans une bande de fréquence du signal d'une électrode 74
4.2.2. Validation fréquentiste par calcul de valeur-p 75
4.2.3. Validation bayésienne par modélisation hiérarchique 78
4.3. Résumé 86

5. Applications 87
5.1. Méthodes employées 87
5.1.1. Méthodes ERSP 87
5.1.2. Méthode d'analyse de synchronisations cooccurrentes 90
5.1.3. Méthode d'analyse de synchronisations univariées 91
5.2. Données artificielles 91
5.2.1. Données multivariées 91
5.2.2. Données univariées 95
5.3. Données expérimentales 101
5.3.1. Présentation du jeu de données 101
5.3.2. Résultats des méthodes d'analyse 103
5.3.3. Mise en correspondance des résultats et interprétation 106
5.4. Discussion générale des résultats 111
5.4.1. Qualité des résultats 111
5.4.2. Temps de calcul et complexités algorithmiques 116

5.5. Résumé 118

6. Conclusions et perspectives **121**
6.1. Conclusions 121
6.1.1. Rappel des motivations 121
6.1.2. Résumé des contributions 121
6.1.3. Conclusions sur les méthodes et les modèles 123
6.2. Perspectives 123
6.2.1. Amélioration des modèles 124
6.2.2. Amélioration des algorithmes d'inférence 124
6.2.3. Extension du domaine d'application 125

A. Variational Message Passing : messages et équations de ré-estimations **127**
A.1. Distribution gaussienne 127
A.2. Distribution gaussienne rectifiée 129
A.3. Distribution gamma 129
A.4. Distribution catégorique 130
A.5. Distribution de Dirichlet 131
A.6. Distribution de mélange 132
A.7. Nœud déterministe de sélection 133
A.8. Nœud déterministe d'addition 134

B. Mise en œuvre pratique de modèles bayésiens avec l'algorithme VMP **135**
B.1. Convergence des modèles 135
B.2. A priori faiblement informatifs et pré-traitement des données 135
B.3. Modèle de mélange gaussien 136
B.4. Modèle ricien 136
B.5. Modèle de mélange ricien 136
B.6. Modèle de mélange ricien hiérarchique 137

C. Résultats des expériences sur le jeu de données artificielles univariées **139**

Bibliographie **149**

Chapitre 1.

Introduction générale

> Mon âme est un orchestre caché ; je ne sais pas quels instruments vibrent et jouent en moi, cordes et harpes, timbales et tambours. Je ne peux me connaître qu'en tant que symphonie.
>
> *Fernando Pessoa – Le livre de l'intranquilité*

Le 6 juillet 1924, Hans Berger réalisa le premier enregistrement d'un électroencéphalogramme sur un jeune homme de 17 ans, au cours d'une opération de neurochirurgie. Ce psychiatre allemand découvrit alors sur le tracé enregistré des oscillations d'une fréquence de 10 Hz. Il fallut encore attendre cinq ans avant qu'il ne publie le premier article sur l'électroencéphalographie et fasse mention des activités oscillatoires particulières qu'il avait découvertes [Berger, 1929]. Depuis ces travaux fondateurs, le monde scientifique n'a eu de cesse d'étudier les signaux électriques cérébraux et plus particulièrement les motifs oscillatoires complexes qui les composent. En effet, ces derniers apparaissent comme un indicateur essentiel du déroulement de processus cognitifs et de la propagation des flux d'information au sein du cerveau.

Pour mieux capturer les mécanismes en œuvre au sein du cerveau, et notamment la manière dont les informations sont encodées et diffusées au sein de celui-ci, cette thèse a le projet de développer des méthodes aidant à analyser un aspect particulier des mécanismes neuronaux prenant place au sein des oscillations cérébrales : la synchronisation.

1.1. Analyse de l'activité cérébrale aux échelles macroscopique et mésoscopique

1.1.1. Électrophysiologie du cerveau et enregistrements

L'origine de l'activité électrique enregistrable du cerveau se situe dans les cellules nerveuses, les neurones, qui composent cet organe. Un neurone est composé de différents compartiments : les dendrites, le corps cellulaire, l'axone et les synapses (voir figure 1.1). Les différentes concentrations d'ions de part et d'autre de la membrane des cellules entrainent la création d'un potentiel de membrane. Dans le cas de cellules nerveuses, ce potentiel peut être localement modifié en différents endroits de la membrane, typiquement dans les synapses. Une synapse est un point de jonction entre l'axone d'une cellule nerveuse présynaptique et une dendrite d'une cellule nerveuse postsynaptique.

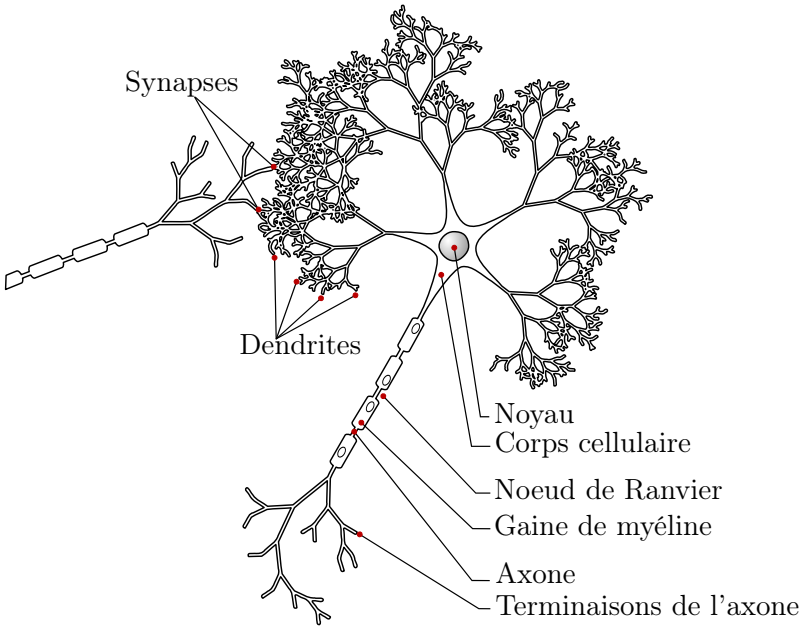

FIGURE 1.1. – Schéma d'un neurone biologique (adapté de Wikipédia – Nicolas Rougier)

Au niveau d'une synapse, la libération de messagers chimiques, les neurotransmetteurs, par la cellule présynaptique peut entrainer une augmentation ou une diminution du potentiel de membrane de la cellule postsynaptique, en fonction des flux d'ions provoqués entre l'intérieur et l'extérieur de cette cellule. Cette perturbation locale du potentiel de membrane se propage alors jusqu'au corps cellulaire. Au niveau du corps cellulaire, si le potentiel de membrane résultant dépasse un seuil, la cellule émet un pic de dépolarisation, appelé potentiel d'action, qui se propagera sur son axone. Les synapses en terminaison de l'axone libèreront à leur tour des neurotransmetteurs sous l'effet de l'arrivée du potentiel d'action. Le fonctionnement d'une cellule nerveuse implique donc des déplacements de charges électriques. Ces déplacements créent des courants électriques, qu'il est possible d'enregistrer.

Dans le cas d'un enregistrement d'électroencéphalographie (EEG), les électrodes placées sur le scalp du sujet enregistrent des différences de potentiel par rapport à une électrode de référence. Ces potentiels sont la conséquence des courants électriques décrits ci-dessus. Cependant, pour qu'un effet soit mesurable au travers du crâne et du scalp, l'activité électrique conjointe des cellules nerveuses doit être suffisamment grande. Ceci peut se produire quand les cellules sont alignées, ce qui permet à leurs activités respectives de se cumuler. Il s'agit typiquement des cellules nerveuses pyramidales des différentes couches du cortex, dont les arbres dendritiques sont parallèles. L'activité électrique enregistrée issue de ces cellules pyramidales est donc de nature postsynaptique [Nunez and Srinavasan, 2006].

L'activité conjointe d'une population localisée de cellules pyramidales se propage dans la boîte crânienne et est enregistrée par plusieurs électrodes du dispositif EEG. Comme

plusieurs populations sont actives en même temps, le signal EEG recueilli est un mélange des activités de chaque population. Il est intéressant de noter que suivant l'orientation de l'ensemble des neurones générateurs, l'activité correspondante enregistrée sur le scalp sera plus ou moins grande. Les ensembles de neurones orientés perpendiculairement au scalp sont les constituants principaux du signal EEG, même s'il est aussi possible de retrouver la trace de l'activité d'ensembles de neurones tangents au scalp [Nunez and Srinavasan, 2006]. En définitive, le signal EEG reste un signal très bruité, dont les amplitudes sont de l'ordre de grandeur du μV.

Un autre type d'enregistrement de l'activité électrique du cerveau est l'enregistrement de *potentiels de champs locaux* (local field potentials ou LFP). Il s'agit d'enregistrements intracrâniens, réalisés à l'aide d'électrodes directement insérées dans le milieu extracellulaire. Ces électrodes à faible impédance enregistrent l'activité électrique dans une zone donnée. Contrairement à l'EEG, la provenance des activités électriques enregistrées par ces électrodes est plus confuse. En effet, les électrodes intracrâniennes enregistrent à la fois les activités présynaptiques et postsynaptiques des neurones avoisinants, perçues différemment suivant la disposition et l'éloignement de ces neurones. Pour obtenir les potentiels de champs locaux, le signal issu des électrodes est filtré par un filtre passe-bas avec une fréquence de coupure entre 300 et 400 Hz, ceci afin de supprimer du signal les potentiels d'action des neurones avoisinants et éviter d'analyser par la suite les activités individuelles des neurones [Vialatte, 2005]. Les données exploitées dans cette thèse sont issues d'enregistrements de potentiels de champs locaux.

Dans le cas des enregistrements EEG comme dans celui des potentiels de champs locaux, l'intérêt des signaux réside dans leur capacité à refléter l'activité électrique conjointe de groupes de neurones localisés, les échelles et les contraintes de chaque technique étant différentes. Les signaux EEG ne donnent accès qu'à une vue projetée sur le scalp de l'activité du cerveau, mais cette vue est entière. Au contraire, les potentiels de champs locaux reflètent plus directement l'activité des populations de neurones proches des électrodes, dans un volume allant du mm^3 [Vialatte, 2005] au μm^3 [Denker et al., 2011, Le Van Quyen and Bragin, 2007a] suivant le type d'électrode employée. En contrepartie, le nombre de zones qui peuvent être étudiées est limité par le nombre d'électrodes implantées dans le sujet.

1.1.2. Analyse fréquentielle de signaux cérébraux

Les signaux EEG et les signaux de potentiels de champs locaux présentent couramment un aspect oscillatoire. Historiquement, les premières analyses de ces oscillations se sont faites en comptant le nombre d'intersections avec l'axe des abscisses du tracé [Nunez and Srinavasan, 2006]. Ce nombre d'intersections devait correspondre à la fréquence des oscillations. Malheureusement, dans le cas où le signal est un mélange de plusieurs oscillations, cette méthode est inefficace.

Ainsi, pour analyser les fréquences des oscillations des signaux électro-corticaux, la transformée de Fourier est l'outil idéal. Il s'agit d'une décomposition du signal comme une somme pondérée de sinusoïdes complexes. Les coefficients étant complexes, il est

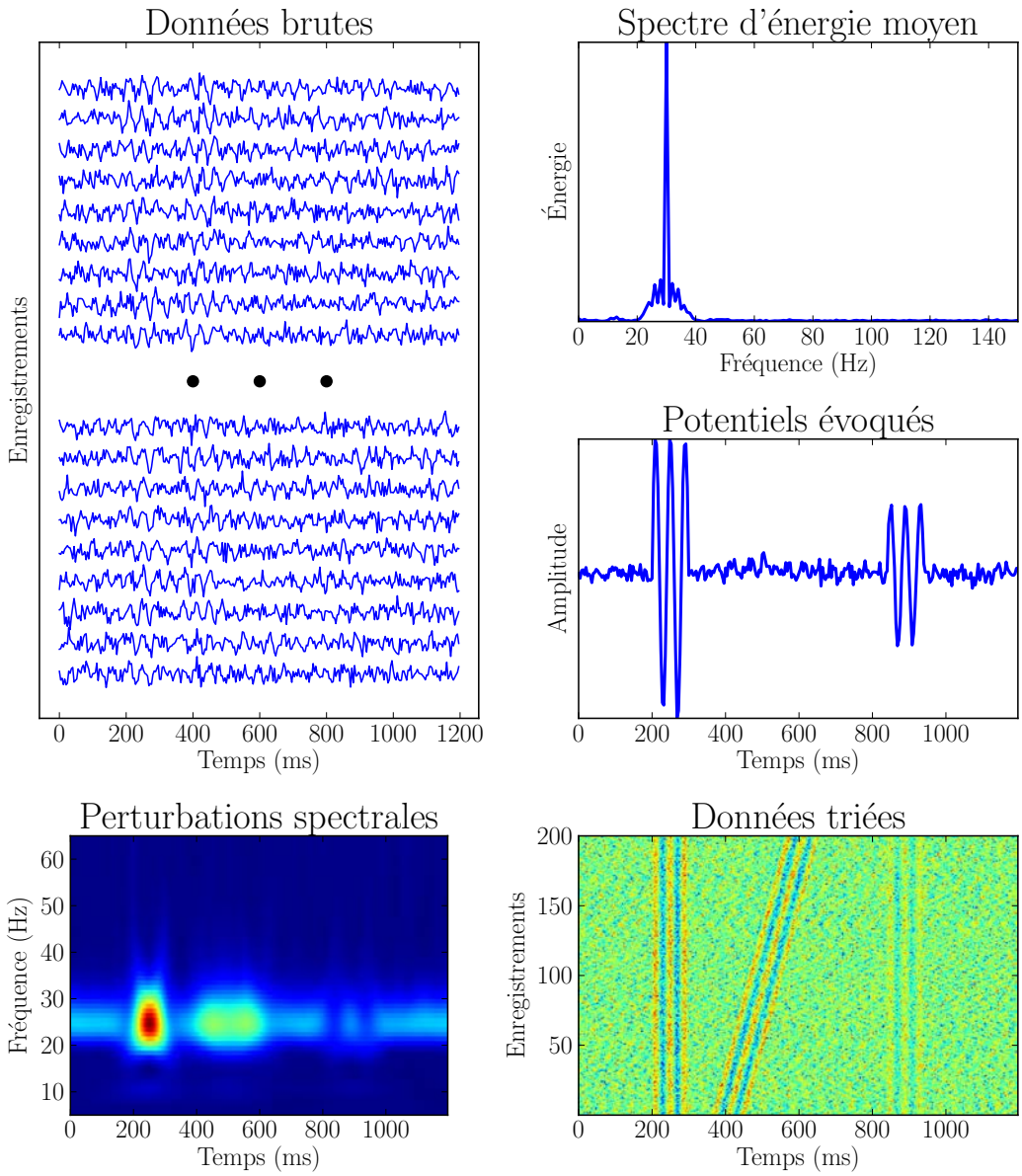

FIGURE 1.2. – Enregistrements simulés d'un signal recueilli par une électrode. Le jeu de données comporte 200 enregistrements simulés. Le signal est constitué d'une sinusoïde à 25 Hz additionnée d'un bruit. Entre 200 et 300 ms, a lieu un potentiel évoqué dû à un mécanisme de bruit additif. Entre 360 et 660 ms, apparaît un potentiel induit, dû à un décalage régulier de la phase entre chaque enregistrement. Entre 840 et 940 ms, un second potentiel évoqué apparaît, dû à un mécanisme de remise à zéro de la phase. Les signaux simulés sont traités avec les différentes méthodes d'analyse présentées dans le texte.

possible de calculer une amplitude et une phase pour chaque sinusoïde [1]. L'information d'amplitude peut alors être utilisée pour illustrer le contenu fréquentiel du signal dans un spectre d'énergie. Le signal de base étant bruité, le spectre d'énergie est en général obtenu en moyennant les spectres des signaux de plusieurs enregistrements d'une même expérience (exemple à la figure 1.2).

L'étude des spectres d'énergie de signaux corticaux a permis de mettre en relief l'existence de bandes de fréquence spécifiques liées à des observations comportementales. Par exemple, les oscillations à 10 Hz découvertes par Hans Berger font parties de la bande de fréquence des ondes alpha, présentes quand le sujet est relâché. Les principales bandes de fréquence identifiées chez l'homme sont les bandes des ondes delta (1 à 4 Hz), théta (4 à 8 Hz), alpha (8 à 12 Hz), bêta (12 à 30 Hz) et gamma (30 à 120 Hz) [Dalal et al., 2011].

Les oscillations du signal électrique ont pour origine des activités coordonnées de neurones au sein d'une population [Pfurtscheller and Lopes da Silva, 1999, Denker et al., 2011]. Plusieurs théories existent pour expliquer la génération de ces oscillations, en fonction des types de neurones, de leur agencement et du type d'oscillations. Mes travaux se focalisant sur l'étude du phénomène, je ne détaillerai pas ici ces différents mécanismes.

Une caractéristique étonnante du spectre d'énergie de signaux électro-corticaux est la distribution décroissante des énergies [Pfurtscheller and Lopes da Silva, 1999, Vialatte, 2005]. En effet, une composante fréquentielle voit son pic d'énergie être d'autant plus faible que sa fréquence est élevée. Une explication théorique avancée est que l'amplitude d'une composante fréquentielle reflète, au moins en partie, la quantité de neurones impliqués dont l'activité conjointe génère ces oscillations [Pfurtscheller and Lopes da Silva, 1999]. Dans le cas de basses fréquences, la quantité de neurones recrutés serait dès lors beaucoup plus grande que pour des oscillations de hautes fréquences, plus transitoires, qui ne mettraient en jeu que des petits groupes de neurones.

1.1.3. Analyse temporelle de signaux cérébraux

Une autre approche de l'analyse de signaux cérébraux consiste à étudier le décours temporel du signal électrique dans le cadre de paradigmes expérimentaux impliquant un stimulus et éventuellement une réponse du sujet. L'idée est de détecter dans le signal des événements liés à la présentation du stimulus ou à la réponse par le sujet.

En pratique, un grand nombre d'enregistrements, supérieur à une centaine, d'une même expérience est réalisé. Puis, le signal de tous les enregistrements est aligné sur le temps du stimulus ou de la réponse avant de calculer une moyenne en chaque point temporel sur l'ensemble des enregistrements. Le signal moyen obtenu est dit « verrouillé » sur le stimulus ou sur la réponse [Picton et al., 2000].

Si le nombre d'enregistrements utilisés pour la moyenne est suffisamment grand, alors il est possible de voir apparaître dans le signal moyen de larges déviations du potentiel enregistré. Les pics ainsi obtenus sont appelés *potentiels évoqués* [2], ou encore « ondes », et

1. La transformée de Fourier sera étudiée plus en détail au chapitre 2.

2. Sous ce terme, je désigne à la fois les potentiels évoqués précoces (evoked potentials) et les potentiels évoqués cognitifs, plus tardifs (event-related potentials ou ERP), lesquels sont parfois différenciés dans la littérature, parce qu'ils relèvent de structures corticales différentes.

sont désignés par la lettre P ou N suivi de la latence du phénomène en ms [Picton et al., 2000]. La lettre P fait référence à un potentiel évoqué dont le pic suit une augmentation du signal, et inversement pour la lettre N. Par exemple, l'onde P300 est un potentiel évoqué issu d'un accroissement du signal culminant aux environs de 300 ms après le stimulus.

Les potentiels évoqués sont ainsi dénommés car ils représentent des événements survenant à une latence fixe, très précise, après le stimulus ou avant la réponse. Les potentiels évoqués identifiés dans le signal moyen ne sont pas discernables dans les enregistrements individuels. Ce phénomène est expliqué dans la littérature par différents mécanismes dont les deux principaux sont :

– le modèle du signal avec un bruit additif,
– le modèle de la remise à zéro de la phase.

Dans le cas du bruit additif, l'hypothèse sous-jacente est que le signal correspondant aux potentiels évoqués est présent dans chaque enregistrement mais contaminé par un haut niveau de bruit [Tallon-Baudry and Bertrand, 1999, Turi et al., 2012, Makeig et al., 2004]. Par conséquent, l'utilisation d'une moyenne permet de réduire ce bruit et fait apparaître les potentiels évoqués. Dans le cas d'une remise à zéro de la phase, on considère que le signal de chaque enregistrement présente une activité oscillatoire de puissance constante et de phase aléatoire, sauf aux latences des potentiels évoqués, où cette phase est remise à zéro [Penny et al., 2002, Turi et al., 2012, Makeig et al., 2004]. Par conséquent, les signaux s'annulent lors du calcul de la moyenne là où la phase est aléatoire, et se cumulent là où la phase est identique, révélant ainsi des potentiels évoqués. Il a été mis en relief que ces deux mécanismes sont observables, l'un n'excluant pas l'autre (exemple à la figure 1.2).

Le calcul d'une moyenne sur les enregistrements n'est pas la seule approche pour l'analyse temporelle d'un signal électro-cortical. Une autre approche répandue consiste à ordonner les différents enregistrements suivant un critère, tel que le temps de réponse, puis à afficher l'amplitude du signal de tous les enregistrements dans un graphique à deux dimensions ayant pour axes le temps et les enregistrements triés. Cette approche porte de le nom d'*analyse simple essai* (single trial analysis) [Delorme and Makeig, 2004]. L'intérêt de celle-ci vient du fait que certains potentiels ne sont pas évoqués mais *induits* par le stimulus, c'est-à-dire qu'ils sont provoqués par le stimulus mais à une latence variable (exemple à la figure 1.2). Dès lors, l'utilisation de la moyenne est à proscrire, car elle entraîne une disparition de ces potentiels induits dans le signal moyen final.

1.1.4. Analyse temps-fréquence et synchronisations

L'analyse fréquentielle des signaux cérébraux présente l'inconvénient de ne pas montrer l'évolution du contenu fréquentiel du signal au cours du temps, puisque l'information temporelle disparaît totalement. L'analyse des potentiels évoqués, quant à elle, ne permet pas de détecter des phénomènes induits au sein des signaux. L'utilisation d'une transformation dite temps-fréquence permet de dépasser les limitations de ces deux techniques d'analyse, et de créer une méthode d'*analyse des perturbations spectrales événementielles* (event-related spectral perturbations ou ERSP [Makeig et al., 2004, Grandchamp and Delorme, 2011]). Le contexte expérimental est le même que celui de l'analyse des potentiels évoqués, avec la présentation d'un stimulus et l'attente éventuelle d'une réponse à chaque

enregistrement.

Une méthode d'analyse des ERSP se fonde sur une transformée temps-fréquence pour calculer dans une première phase la densité d'énergie temps-fréquence du signal de chaque enregistrement. Il s'agit d'obtenir un spectre d'énergie fréquentiel en chaque point temporel du signal. Une transformée temps-fréquence usuelle est la transformée de Fourier à court terme, consistant à calculer la transformée de Fourier d'une fenêtre temporelle centrée en chaque point temporel. Une autre transformée très utilisée est la transformée en ondelettes continue. Ces deux transformées seront présentées plus avant au chapitre 2.

La seconde étape d'une méthode d'analyse ERSP est le calcul d'une moyenne des spectres d'énergie temps-fréquence sur l'ensemble des enregistrements, en verrouillant sur le stimulus ou sur la réponse à la manière de l'analyse des potentiels évoqués. Dès lors, les phénomènes d'intérêt étudiés sont les augmentations et diminutions du spectre d'énergie moyen par rapport à l'activité pré-stimulus, considérée comme représentative d'une activité de base du cerveau [Pfurtscheller and Lopes da Silva, 1999]. Ainsi, dans chaque bande de fréquence, le spectre d'énergie moyen est normalisé à l'aide du spectre d'énergie moyen mesuré dans la période pré-stimulus, par exemple en centrant et réduisant avec la moyenne et l'écart-type de cette activité pré-stimulus [Grandchamp and Delorme, 2011]. Un intérêt de cette normalisation est de compenser la décroissance de l'énergie du signal dans les hautes fréquences, permettant ainsi de comparer de la même manière les fluctuations du spectre d'énergie moyen dans tout le plan temps-fréquence.

Les augmentations ou diminutions significatives du spectre d'énergie au cours du temps, dans les différentes bandes de fréquence, sont détectées à l'aide de tests statistiques comparant l'activité pré-stimulus et l'activité post-stimulus. Une augmentation significative correspond à un épisode de *synchronisation*, aussi appelé synchronisation événementielle (event-related synchronisation ou ERS [Pfurtscheller and Lopes da Silva, 1999]). De même, une diminution significative correspond à un épisode de *désynchronisation*, ou désynchronisation événementielle (event-related desynchronisation ou ERD). Ces dénominations sont issues de l'interprétation associée à ces fluctuations du spectre d'énergie moyen. En effet, une hypothèse avancée pour expliquer un épisode de synchronisation est que l'augmentation localisée de l'amplitude des oscillations correspond à une augmentation de la coordination de l'activité des neurones impliqués dans la génération de ces oscillations [Pfurtscheller and Lopes da Silva, 1999]. Inversement, un épisode de désynchronisation correspondrait à une baisse de la coordination des neurones générant les oscillations. À l'exception de la section suivante, le reste de ce manuscrit s'intéressera spécifiquement à ce type d'événements de synchronisation.

En comparaison de l'analyse temporelle des potentiels évoqués, l'analyse temps-fréquence à l'aide de méthode de type ERSP permet de détecter certains phénomènes évoqués et les phénomènes induits. Les phénomènes évoqués issus d'un mécanisme de bruit additif sont détectables étant donné que leur signature énergétique dans le spectre temps-fréquence de chaque enregistrement est débruitée par le calcul du spectre moyen et correspond au final à une augmentation dans le spectre d'énergie moyen. Au contraire, les phénomènes évoqués issus d'une remise à zéro de la phase n'ont pas de signature en énergie dans le spectre de chaque enregistrement et par conséquent n'apparaissent pas dans le spectre moyen. Concernant les phénomènes induits, ceux-ci présentent une signature en énergie à

chaque enregistrement dans des intervalles de temps se chevauchant. Le spectre d'énergie étant strictement positif, le décalage en temps d'un phénomène induit dans les enregistrements n'entraîne pas de disparition du phénomène lors du calcul du spectre moyen, mais seulement un étalement en temps et une diminution de l'énergie moyenne du phénomène induit, en comparaison d'un phénomène évoqué (exemple à la figure 1.2) [Tallon-Baudry and Bertrand, 1999].

1.1.5. Analyse de la phase et synchronisations

La transformée de Fourier et les transformées temps-fréquence génèrent à la fois un spectre d'énergie et un spectre de phase. L'analyse de la phase des signaux électro-corticaux donne lieu à l'étude des synchronisations de phase. Mes travaux de thèse ne concernent pas ce type de synchronisation mais il s'agit d'une des extensions possibles de ceux-ci.

À partir de la phase du signal dans chaque enregistrement, il est possible de calculer pour une même électrode un indice de verrouillage de la phase entre les enregistrements (inter-trial coherence ou ITC [Delorme and Makeig, 2004, Makeig et al., 2004]), à valeur dans $[0; 1]$ et représentant la similarité de la phase dans les différents enregistrements. Une valeur de 1 signifie que la phase du signal est identique dans tous les enregistrements. Si l'ITC est calculé sur un spectre de phase obtenu avec une transformée temps-fréquence, l'évolution de l'ITC au cours du temps met en valeur des phénomènes évoqués issus d'un mécanisme de remise à zéro de la phase.

La phase peut aussi être utilisée pour détecter des mécanismes de verrouillage de phase au sein du signal d'une paire d'électrodes. Ainsi, la valeur de verrouillage de la phase (phase locking value ou PLV [Lachaux et al., 1999]) est un indice couramment utilisé pour tester si la différence de phase entre deux électrodes reste constante au cours des différents enregistrements. La PLV est définie dans l'intervalle $[0; 1]$: une haute valeur dénote un phénomène de synchronisation de la phase entre les signaux enregistrés aux deux électrodes. Cette approche sert à détecter des mécanismes de coordination entre différentes aires cérébrales, par des mécanismes de synchronisation à longue distance [Rodriguez et al., 1999].

1.2. Présentation de la thèse

1.2.1. Motivation de la thèse

L'analyse de l'évolution du contenu fréquentiel des signaux cérébraux, telle que réalisée dans des techniques conventionnelles de type ERSP, présente des limitations de par les hypothèses fortes que ces techniques posent sur les signaux.

La première hypothèse contestable est celle de la stabilité du signal d'intérêt, portant l'information de synchronisation et de désynchronisation, entre plusieurs enregistrements d'une même expérience. Cette hypothèse de stabilité entraîne l'emploi d'une moyenne sur les enregistrements pour réduire l'influence des variations entre enregistrements considérées comme un bruit additif. Cette moyenne pose problème car :

– elle requiert un grand nombre d'enregistrements,

– elle est sensible aux points aberrants [Grandchamp and Delorme, 2011],
– elle ne permet pas de tenir compte de la variabilité intrinsèque des signaux cérébraux, qui se manifeste sous la forme de différences d'amplitude, de latence et de nature même des phénomènes de synchronisation et désynchronisation d'un enregistrement à l'autre [Onton et al., 2005].

La seconde hypothèse critiquable concerne le fait que les épisodes de synchronisations et de désynchronisations sont détectés par rapport à une activité de base antérieure à la présentation d'un stimulus. Cette hypothèse présente l'activité du cerveau avant le stimulus comme indépendante du stimulus et de l'activité après ce stimulus. Or des études récentes montrent que l'activité de base du cerveau a des répercussions sur l'activité post-stimulus et ne peut pas être traitée comme un simple bruit [Sadaghiani et al., 2010].

1.2.2. But de la thèse

Pour améliorer l'analyse des synchronisations au sein de signaux électro-corticaux, mes travaux de thèse se sont focalisés sur la suppression de l'étape de calcul d'une moyenne et de l'utilisation de l'information pré-stimulus pour mettre au point de nouvelles méthodes d'analyse.

Pour éviter l'emploi de la moyenne des enregistrements, l'approche que je propose repose sur une analyse de type simple essai de chaque enregistrement pour y détecter des épisodes potentiels de synchronisation. Ces épisodes potentiels forment alors une information symbolique qu'il est possible de combiner entre les enregistrements pour détecter les épisodes de synchronisation invariants vis-à-vis de l'expérience étudiée.

Au cœur de l'analyse simple essai, la détection des épisodes potentiels de synchronisation n'emploie pas l'information sur l'activité de la période pré-stimulus. En effet, l'analyse de chaque enregistrement se fonde sur une modélisation statistique pour extraire de manière non supervisée l'information de synchronisation. Pour réaliser cette modélisation statistique, je propose deux stratégies différentes :

– La première stratégie consiste à effectuer une comparaison entre les signaux enregistrés par différentes électrodes en un même point temps-fréquence, pour obtenir des groupes d'électrodes aux épisodes de synchronisations co-occurents. Ces groupes sont validés ou invalidés par une mesure utilisant l'information contenue dans une fenêtre minimale définie en temps et en fréquence autour de chaque point temps-fréquence analysé.
– La seconde stratégie consiste à modéliser entièrement le signal d'une électrode dans une bande de fréquence pour étiqueter chaque point temporel comme un potentiel épisode de synchronisation ou de désynchronisation.

Chacune de ces deux stratégies donne lieu au développement d'une technique spécifique pour combiner les informations recueillies au travers des enregistrements, formant ainsi deux méthodes distinctes d'analyse de synchronisations.

1.2.3. Organisation de la thèse

Le manuscrit de thèse est divisé en six chapitres. Le contenu de ces chapitres est structuré de la manière suivante.

Le chapitre 1 introduit les éléments de base de l'étude de signaux électriques du cerveau et en particulier, l'étude de la répartition de l'énergie de ces signaux dans un plan temps-fréquence pour la détection d'épisodes de synchronisations et désynchronisations dans des régions corticales. Les techniques d'analyse conventionnelles de ces phénomènes de synchronisation, telles que les méthodes ERSP, sont présentées avec leurs limitations.

Le chapitre 2 détaille deux fondements théoriques des méthodes développées dans ma thèse : la transformée en ondelettes continue avec l'ondelette de Morlet et les modèles bayésiens calculés à l'aide de l'algorithme Variational Message Passing.

Le chapitre 3 développe des résultats théoriques sur la distribution de Rice du module des coefficients d'ondelettes d'un signal bruité, puis introduit les modèles bayésiens de mélange gaussien et de mélange ricien, utilisés comme outil statistique pour détecter de manière non supervisée des groupes d'amplitudes différentes au sein d'un ensemble de coefficients d'ondelettes. Le modèle bayésien de mélange ricien et les techniques numériques de calcul associés sont un apport de ma thèse.

Le chapitre 4 expose deux méthodes d'analyse des variations de l'énergie de signaux cérébraux dans le plan temps-fréquence dont le but est la détection de synchronisations. Ces méthodes se fondent sur des modèles bayésiens de mélange gaussien et ricien, appliqués à la transformée en ondelette des signaux cérébraux. Une approche nouvelle reposant sur l'emploi de l'information contenue dans une fenêtre temps-fréquence minimale est introduite avec la première de ces deux méthodes. La seconde méthode présente un modèle hiérarchique bayésien permettant d'obtenir une méthode complète de détection des phénomènes d'intérêt sans usage des statistiques fréquentistes.

Le chapitre 5 présente l'application de mes méthodes d'analyse sur des signaux artificiels pour évaluer leur performance, et sur des signaux expérimentaux pour montrer leur capacité à extraire des informations de synchronisation physiologiquement plausibles. Les deux méthodes se révèlent performantes sur des signaux artificiels dont les hypothèses correspondent aux leurs. Sur les signaux expérimentaux, l'une des méthodes présente quelques limitations, signe de l'inadéquation de ses hypothèses, et l'autre méthode parvient à extraire une information fiable là où les méthodes standards échouent.

Le chapitre 6 conclut le manuscrit avec un résumé des apports nouveaux de ma thèse et une discussion sur la manière d'étendre mes méthodes d'analyse, dans leurs champs d'application et dans les phénomènes détectés.

Chapitre 2.

Outils théoriques

> Les amants inventent leur propre vocabulaire,
> mais il n'a de signification que pour eux.
>
> *René Barjavel – L'enchanteur*

Ce chapitre se veut une présentation de deux blocs théoriques. Ceux-ci forment ensemble la base de l'analyse statistique de signaux cérébraux que j'ai mise au point et qui sera le sujet des chapitres suivants.

Le premier bloc théorique concerne l'analyse temps-fréquence et se concentrera plus particulièrement sur la transformée en ondelettes continue. Cette dernière a déjà été évoquée dans le chapitre précédent, comme elle constitue la première étape de nombreuses chaînes de traitement de signaux cérébraux.

Les modèles statistiques bayésiens forment le second bloc théorique dont il sera question dans ce chapitre. Ce type de modèles me sert à analyser les signaux traités par la transformée en ondelettes continue. Je décrirai dans cette partie la manière de construire des modèles bayésiens et de mener des calculs numériques sur ceux-ci pour extraire des informations à partir des données.

2.1. L'analyse temps-fréquence avec la transformée en ondelettes continue

Dans le chapitre précédent, une caractéristique essentielle des signaux cérébraux a été mise en avant : la présence d'ondes de différentes fréquences. L'étude de la fréquence, de l'amplitude et de la phase de ces ondes offre une porte d'accès pour la détection de certains types de synchronisations cérébrales, comme l'atteste les techniques présentées au chapitre 1 se basant pour une majorité sur l'information fréquentielle.

Je m'attacherai à décrire ici trois techniques usuelles de traitement du signal servant à extraire ces caractéristiques fréquentielles : la transformée de Fourier, la transformée de Fourier à court terme et la transformée en ondelettes continue. Par la suite, dans mes travaux je n'emploierai que la transformée en ondelettes continue associée à l'ondelette de Morlet complexe. Il est néanmoins essentiel de présenter les deux autres transformées pour mieux saisir les avantages de la transformée en ondelettes continue.

2.1.1. La transformée de Fourier

Le principe de l'analyse de Fourier est de considérer toute fonction s à énergie finie comme une somme pondérée infinie de sinusoïdes complexes $e^{i2\pi ft}$:

$$s(t) = \int_{-\infty}^{+\infty} \hat{s}(f)\, e^{i2\pi ft}\, df, \tag{2.1}$$

où $\hat{s}(f)$ est le coefficient de pondération complexe associé à la sinusoïde de fréquence f. Ces coefficients reflètent la projection de la fonction s sur chaque sinusoïde, et sont calculés via la transformée de Fourier $\hat{s}$ de la fonction s :

$$\hat{s}(f) = \int_{-\infty}^{+\infty} s(t)\, e^{-i2\pi ft}\, dt. \tag{2.2}$$

Dans le cadre de l'étude de l'activité électrique cérébrale, la fonction s correspond à un signal d'électrophysiologie enregistré. Ce signal étant borné en temps et en amplitude, son énergie est finie. Au niveau des grandeurs physiques, t est exprimé en secondes, f en Hertz et $s(t)$ en volts. Pour la suite du manuscrit, ces unités ne seront pas mentionnées.

Les coefficients $\hat{s}(f)$ de la transformée de Fourier d'un signal s ne dépendent que d'un unique paramètre : la fréquence f. Aussi forment-ils une représentation du signal dans un espace indexé par des fréquences, c'est-à-dire un espace fréquentiel.

Ces coefficients étant complexes, ils sont décomposables en un module $|\hat{s}(f)|$ et une phase $\phi(\hat{s}(f))$ tels que

$$\hat{s}(f) = |\hat{s}(f)|\, e^{i\phi(\hat{s}(f))}. \tag{2.3}$$

Pour chaque fréquence f, le module $|\hat{s}(f)|$ correspond à l'amplitude de la sinusoïde complexe de fréquence f composant le signal, et la phase $\phi(\hat{s}(f))$ correspond au décalage temporel initial de cette sinusoïde.

L'égalité de *Parseval* énonce que l'énergie du signal dans l'espace fréquentiel et l'énergie du signal dans l'espace temporel sont équivalentes [Mallat, 2008] :

$$\int_{-\infty}^{+\infty} |s(t)|^2\, dt = \int_{-\infty}^{+\infty} |\hat{s}(f)|^2\, df. \tag{2.4}$$

Ainsi le module au carré $|\hat{s}(f)|^2$, aussi appelé *densité spectrale d'énergie*, forme une représentation de la distribution de l'énergie du signal dans l'espace fréquentiel. Il s'agit d'une représentation privilégiée pour étudier la composition fréquentielle d'un signal.

Un inconvénient majeur de la transformée de Fourier est qu'elle ne permet pas de localiser en temps les composantes fréquentielles d'un signal, l'information temporelle étant absente de la représentation fréquentielle. Cette limitation est problématique pour l'étude de signaux dont le contenu fréquentiel change au cours du temps, tel que les signaux d'enregistrements électriques du cerveau.

Pour remédier à cette limitation, des représentations dites *temps-fréquence* ont été mises au point. Les sections suivantes présentent deux approches différentes pour construire ces représentations.

2.1.2. Transformée de Fourier à court terme

Pour construire une représentation temps-fréquence d'un signal, une première approche consiste à localiser en temps les sinusoïdes complexes de la transformée de Fourier et à projeter le signal sur cette nouvelle base de fonctions localisées en temps et en fréquence.

À l'aide d'une fonction réelle paire g d'énergie finie, qui fait office de « fenêtre » temporelle, on construit une base de sinusoïdes complexes délimitées dans le temps $g_{\tau,f}$:

$$g_{\tau,f}(t) = g(t-\tau)\, e^{i2\pi ft}. \tag{2.5}$$

Le paramètre f sert à fixer la fréquence de la sinusoïde et le paramètre τ à translater la fenêtre g sur l'axe du temps. La fonction g est choisie normalisée en énergie :

$$\|g\| = \int_{-\infty}^{+\infty} |g(t)|^2\, dt = 1, \tag{2.6}$$

de telle sorte que toutes les fonctions $g_{\tau,f}$ le sont aussi. Une fonction de fenêtrage typique est la fonction gaussienne normalisée en énergie :

$$g(t) = \pi^{-\frac{1}{4}}\, e^{-\frac{t^2}{2}}. \tag{2.7}$$

Tout signal s d'énergie finie peut alors être décomposé en une somme pondérée infinie de fonctions $g_{\tau,f}$:

$$s(t) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \mathrm{ST}s(\tau,f)\, g_{\tau,f}(t)\, df d\tau. \tag{2.8}$$

Les coefficients $\mathrm{ST}s(\tau,f)$ sont calculés via la transformée de Fourier fenêtrée, aussi appelée *transformée de Fourier à court terme*, et correspondent à la projection de s sur la base des fonctions $g_{\tau,f}$:

$$\mathrm{ST}s(\tau,f) = \int_{-\infty}^{+\infty} s(t)\, \overline{g}_{\tau,f}(t)\, dt = \int_{-\infty}^{+\infty} s(t)\, g(t-\tau)\, e^{-i2\pi ft}\, dt, \tag{2.9}$$

où $\overline{g}_{\tau,f}(t)$ est le conjugué de $g_{\tau,f}(t)$. Ces coefficients sont indexés en temps par τ et en fréquence par f : ils représentent donc le signal dans un espace temps-fréquence.

Les coefficients $\mathrm{ST}s(\tau,f)$ sont des nombres complexes. La transformée de Fourier à court terme d'un signal a donc un module et une phase.

Il est possible de démontrer la conservation de l'énergie par la transformée de Fourier à court terme [Vetterli and Kovačevic, 1995] :

$$\int_{-\infty}^{+\infty} |s(t)|^2\, dt = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |\mathrm{ST}s(\tau,f)|^2\, df d\tau, \tag{2.10}$$

ce qui implique que le module au carré $|\mathrm{ST}s(\tau,f)|^2$ est une densité d'énergie du signal, représentant la distribution de l'énergie du signal dans un espace temps-fréquence. Cette densité d'énergie est nommée *spectrogramme* du signal.

Boîte d'Heisenberg

La résolution en temps et en fréquence de la transformée de Fourier à court terme n'est pas infinie. À cause de la dispersion en temps et en fréquence de la densité d'énergie des fonctions de base, chaque fonction de base $g_{\tau,f}$ associe une imprécision au coefficient $\mathrm{ST}s(\tau, f)$ sous la forme d'une région du plan temps-fréquence centrée en (τ, f). Une méthode standard pour quantifier la taille de cette région est de calculer la variance en temps et en fréquence de l'énergie de la fonction de base $g_{\tau,f}$ [1] :

$$\begin{aligned}\sigma_t^2 &= \int_{-\infty}^{+\infty} (t-\tau)^2 \, |g_{\tau,f}(t)|^2 \, dt = \int_{-\infty}^{+\infty} t^2 \, |g(t)|^2 \, dt \\ \sigma_f^2 &= \int_{-\infty}^{+\infty} (f'-f)^2 \, |\hat{g}_{\tau,f}(f')|^2 \, df' = \int_{-\infty}^{+\infty} f'^2 \, |\hat{g}(f')|^2 \, df'.\end{aligned} \tag{2.11}$$

Dans le plan temps-fréquence, le rectangle centré en (τ, f) et de dimension $\sigma_t \times \sigma_f$ est appelé *boîte d'Heisenberg* de la transformée de Fourier à court terme, et représente la résolution temps-fréquence de cette transformée. Le principe d'incertitude d'Heisenberg énonce que l'aire de ce rectangle est bornée par une valeur inférieure [Mallat, 2008] :

$$\sigma_t \sigma_f \geq \frac{1}{4\pi}. \tag{2.12}$$

Cette borne inférieure est atteinte quand la fonction de fenêtrage g est une gaussienne.

Il est intéressant de noter que les dimensions de la boîte d'Heisenberg de la transformée de Fourier à court terme sont identiques pour tous les coefficients $\mathrm{ST}s(\tau, f)$, quelle que soit leur localisation dans le plan temps-fréquence. Cette propriété peut se révéler désavantageuse si l'on souhaite faire varier la résolution temps-fréquence pour mieux capturer les caractéristiques d'un signal. Il faut alors se tourner vers d'autres représentations temps-fréquence, par exemple la transformée en ondelettes continue.

2.1.3. La transformée en ondelettes continue

Comme la transformée de Fourier à court terme, la transformée en ondelettes continue d'un signal est une projection de ce signal sur une base de fonctions oscillantes localisées en temps et en fréquence. Ces fonctions de base se nomment *ondelettes*.

Pour construire la base de fonctions sur laquelle projeter le signal, la première étape consiste à définir une ondelette mère ψ. Celle-ci doit répondre à une *condition d'admissibilité* [Kaiser, 1994]

$$\mathcal{C}_\psi = \int_{-\infty}^{+\infty} \frac{|\hat{\psi}(f)|^2}{|f|} \, df < +\infty, \tag{2.13}$$

1. Sachant que $\|g_{\tau,f}\| = \|\hat{g}_{\tau,f}\| = 1$, une analogie est faite entre les densités temporelle et fréquentielle d'énergie de $g_{\tau,f}$ et des densités de probabilité. Les dispersions temporelle et fréquentielle de la fonction de base $g_{\tau,f}$ sont alors calculées à la manière de la variance d'une variable aléatoire.

qui consiste, pour une fonction ψ d'énergie finie continument dérivable, à avoir une moyenne nulle [Kaiser, 1994] :

$$\int_{-\infty}^{+\infty} \psi(t)\, dt = \hat{\psi}(0) = 0. \tag{2.14}$$

Ceci assure que l'ondelette mère est une fonction non nulle sur une portion de l'axe temporel et dont l'amplitude oscille autour de 0. Il est d'usage de choisir une ondelette mère centrée temporellement en $t_0 = 0$ et d'énergie totale unitaire, soit $\|\psi\| = 1$.

Des ondelettes filles $\psi_{\tau,a}$ sont créées à partir de l'ondelette mère, par translation et compression de cette dernière :

$$\psi_{\tau,a}(t) = \frac{1}{\sqrt{a}}\, \psi\left(\frac{t-\tau}{a}\right), \tag{2.15}$$

ce qui se traduit au niveau de la transformée de Fourier des ondelettes filles par une translation et une dilatation de la transformée de Fourier de l'ondelette mère :

$$\hat{\psi}_{\tau,a}(f) = \sqrt{a}\, \hat{\psi}(af)\, e^{i2\pi f\tau}. \tag{2.16}$$

Ainsi, pour une ondelette mère centrée dans le plan temps-fréquence en (t_0, f_0), l'ondelette fille $\psi_{\tau,a}$ est centrée en $(t_0 + \tau, f_0/a)$. Le paramètre a est appelé *échelle* de l'ondelette. Par construction, les ondelettes filles sont aussi normalisées en énergie, soit $\|\psi_{\tau,a}\| = 1$.

La transformée en ondelettes continue d'un signal s correspond alors à la projection du signal s sur la base de fonctions formée par les ondelettes filles $\psi_{\tau,a}$. Les *coefficients d'ondelettes* $\mathrm{W}s(\tau, a)$ de la transformée en ondelettes continue s'écrivent :

$$\mathrm{W}s(\tau, a) = \int_{-\infty}^{+\infty} s(t)\, \overline{\psi}_{\tau,a}(t)\, dt. \tag{2.17}$$

Comme ces coefficients sont indexés par un paramètre de temps τ et d'échelle a, ils forment une représentation *temps-échelle* du signal. En reparamétrant les ondelettes filles par leur fréquence centrale $f = f_0/a$, on obtient des coefficients d'ondelettes $\mathrm{W}s(\tau, f_0/f)$ qui forment une représentation temps-fréquence du signal.

Dans le cas où l'ondelette mère est complexe, les coefficients d'ondelettes $\mathrm{W}s(\tau, a)$ sont eux aussi complexes et par conséquent séparables en un module et une phase. Pour la suite, je me concentrerai uniquement sur ce cas : celui de la transformée en ondelettes continue complexe.

La reconstruction du signal à partir des coefficients d'ondelettes est un peu plus compliquée que dans le cas de la transformée de Fourier à court terme :

$$s(t) = \frac{1}{\mathcal{C}_\psi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \mathrm{W}s(\tau, a)\, \psi_{\tau,a}(t)\, \frac{da\, d\tau}{a^2}, \tag{2.18}$$

mais un changement de variable $f = f_0/a$ fait apparaître une somme pondérée infinie de coefficients d'ondelettes :

$$s(t) = \frac{1}{\mathcal{C}_\psi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \mathrm{W}s\,(\tau, f_0/f)\ \psi_{\tau,f_0/f}(t)\, df\, d\tau. \tag{2.19}$$

Il existe aussi une relation de conservation de l'énergie du signal pour la transformée en ondelettes continue [Kaiser, 1994] :

$$\begin{aligned}\int_{-\infty}^{+\infty} |s(t)|^2 \, dt &= \frac{1}{\mathcal{C}_\psi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |\mathrm{W}s(\tau, a)|^2 \, \frac{da \, d\tau}{a^2} \\ &= \frac{1}{\mathcal{C}_\psi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |\mathrm{W}s(\tau, f_0/f)|^2 \, df \, d\tau. \end{aligned} \tag{2.20}$$

Il en découle que $|\mathrm{W}s(\tau, f_0/f)|^2$ est une densité d'énergie représentant la répartition de l'énergie du signal dans le plan temps-fréquence. Cette densité d'énergie est appelée *scalogramme* du signal.

Résolution temps-fréquence variable

La particularité de la transformée en ondelettes continue vient de la manière de construire ses fonctions de base, les ondelettes filles. L'emploi d'un paramètre d'échelle a permet, en dilatant ou compressant l'ondelette mère, de faire varier la dispersion temporelle et fréquentielle de l'énergie des ondelettes filles. Cela se traduit pour les coefficients d'ondelettes calculés avec ces ondelettes filles par une variation de leur résolution en temps et en fréquence en fonction de leur localisation en fréquence.

La résolution en temps et en fréquence associée à une ondelette fille $\psi_{\tau,a}$, centrée en $(\tau, f_0/a)$, correspond à la dispersion de son énergie dans le plan temps-fréquence [Mallat, 2008] :

$$\begin{aligned}\sigma^2_{\psi_{\tau,a},t} &= \int_{-\infty}^{+\infty} (t-\tau)^2 \, |\psi_{\tau,a}(t)|^2 \, dt = a\sigma_{\psi,t} \\ \sigma^2_{\psi_{\tau,a},f} &= \int_{-\infty}^{+\infty} \left(f - \frac{f_0}{a}\right)^2 \, |\hat{\psi}_{\tau,a}(f)|^2 \, df = \frac{\sigma_{\psi,t}}{a}, \end{aligned} \tag{2.21}$$

où $\sigma_{\psi,t}$ et $\sigma_{\psi,f}$ correspondent à la dispersion de l'énergie de l'ondelette mère ψ dans le plan temps-fréquence. La boîte d'Heisenberg correspondante pour l'ondelette fille $\psi_{\tau,a}$ est donc un rectangle centré en $(\tau, f_0/a)$ de dimension $a\,\sigma_{\psi,t} \times \sigma_{\psi,f}/a$. Cette boîte est soumise au principe d'incertitude d'Heisenberg et son aire est bornée par la formule (2.12).

Ainsi, la transformée en ondelettes continue présente une résolution fréquentielle élevée et une résolution temporelle faible à basses fréquences, et inversement à hautes fréquences. De ce fait, la transformée en ondelettes continue capture avec précision la temporalité des événements transitoires, de hautes fréquences, et privilégie une bonne résolution fréquentielle pour des événements à caractère stationnaire, en basses fréquences.

Cette propriété de *multirésolution* fait de la transformée en ondelettes continue un outil de choix pour l'analyse de signaux non stationnaires au contenu fréquentiel riche. En comparaison, la transformée de Fourier à court terme dispose d'une résolution temps-fréquence fixe et nécessite donc une meilleure connaissance a priori du signal pour régler correctement cette résolution.

Dans le cas des signaux d'enregistrement de l'activité cérébrale, le fait qu'ils soient non stationnaires et qu'ils se répartissent en bandes de fréquences d'intérêt appelle à l'utilisation

d'une transformée multirésolution. De plus, comme détaillé à la section 1.1.2, les bandes de fréquence d'intérêt s'élargissent pour des fréquences plus élevées. Ainsi la transformée en ondelettes continue est particulièrement bien adaptée pour l'étude des signaux cérébraux, comme l'atteste une abondante littérature [Lachaux et al., 2005, Le Van Quyen and Bragin, 2007b, Tallon-Baudry and Bertrand, 1999, Vialatte, 2005, Herrmann et al., 2005]. Parmi les ondelettes utilisées pour l'analyse de signaux cérébraux, l'usage de l'une d'entre elles est prépondérant : il s'agit de l'ondelette de Morlet complexe.

Il est intéressant de noter qu'il existe aussi une version discrète de la transformée en ondelettes. Il s'agit alors d'une projection du signal sur une base de fonctions formée d'un nombre fini d'ondelettes filles. Cette transformée discrète est numériquement exacte et plus parcimonieuse que la transformée contenue. Elle est habituellement utilisée pour compresser des signaux, et dans le cas de signaux électro-corticaux, pour débruiter ces signaux [Quiroga and Garcia, 2003]. Cette transformée n'est pas adaptée à l'analyse des signaux, comme elle n'est pas invariante par translation du signal [Mallat, 2008].

2.1.4. L'ondelette de Morlet complexe

L'ondelette de Morlet complexe est définie comme une sinusoïde complexe fenêtrée par une fonction gaussienne [Goupillaud et al., 1984, Jordan et al., 1997]. J'emploie dans mes travaux la version normalisée de cette ondelette [Simonovski and Boltežar, 2003], qui définit l'ondelette mère ψ comme

$$\psi(t) = \pi^{-\frac{1}{4}}\, e^{-\frac{t^2}{2}}\, e^{i\omega_0 t}, \tag{2.22}$$

et dont la transformée de Fourier vaut

$$\hat{\psi}(f) = \pi^{\frac{1}{4}}\, \sqrt{2}\, e^{-2\pi^2 (f-\frac{\omega_0}{2\pi})^2}. \tag{2.23}$$

Le paramètre ω_0 correspond au nombre de périodes de la sinusoïde dont l'amplitude n'est pas significativement atténuée par la fenêtre gaussienne. De plus, pour $\omega_0 \geq 5$, la moyenne de l'ondelette de Morlet complexe est négligeable, permettant à l'ondelette de Morlet complexe de respecter approximativement la condition d'admissibilité [Simonovski and Boltežar, 2003]. Enfin, la fréquence centrale f_0 de l'ondelette mère est proportionnelle à ω_0, $f_0 = \omega_0/2\pi$.

L'ondelette de Morlet complexe est très similaire aux fonctions de base de la transformée de Fourier à court terme employant une fenêtre gaussienne. Dans le cas de l'ondelette de Morlet complexe, le nombre d'oscillations significatives est fixe et la taille de la fenêtre gaussienne varie avec la fréquence centrale de l'ondelette. Dans le cas de la transformée de Fourier à court terme, la taille de la fenêtre gaussienne est fixe et le nombre d'oscillations significatives varie avec la fréquence centrale de la fonction de base.

La résolution temps-fréquence de l'ondelette de Morlet complexe est excellente. Les résolutions en temps et en fréquence de l'ondelette mère valent :

$$\sigma_{\psi,t} = \frac{1}{\sqrt{2}} \quad \text{et} \quad \sigma_{\psi,f} = \frac{1}{2\sqrt{2}\pi}. \tag{2.24}$$

Dès lors, l'aire de la boîte d'Heisenberg associée à cette ondelette est minimale au regard du principe d'incertitude d'Heisenberg : $\sigma_{\psi,t}\, \sigma_{\psi,f} = \frac{1}{4\pi}$.

Quelques détails d'implémentation

Pour mes travaux, j'utilise une implémentation décrite par Jordan, Miksad et Powers [Jordan et al., 1997], utilisant la transformée de Fourier rapide pour le calcul numérique de la transformée en ondelettes continue avec l'ondelette de Morlet complexe. Deux points spécifiques de cette implémentation nécessitent d'être abordés, parce qu'ils auront un impact majeur sur la représentation temps-fréquence obtenue : le choix des échelles et la gestion des bords du signal.

Dans le cadre d'un calcul numérique, le signal traité est projeté sur un ensemble fini d'échelles a_i. La manière la plus simple de choisir les a_i est d'opter pour un échantillonnage linéaire, c'est-à-dire un espacement constant $\Delta a = a_{i+1} - a_i$ entre chaque échelle. À cause de la résolution temps-fréquence variable de la transformée en ondelette, cet échantillonnage n'est pas efficace :

- un échantillonnage fin entraîne une très grande redondance dans les hautes fréquences, où la résolution en fréquence des ondelettes filles est faible,
- et un échantillonnage grossier provoque une couverture insuffisante dans les basses fréquences, où la résolution en fréquence des ondelettes filles est élevée.

Pour éviter ces écueils, Jordan, Miksad et Powers [Jordan et al., 1997] ont proposé un échantillonnage géométrique des échelles, où le rapport $\Delta a = a_{i+1}/a_i$ entre deux échelles successives est constant.

Une autre subtilité de l'implémentation de la transformée en ondelettes concerne la gestion des effets de bord. Un signal enregistré est défini sur une période de temps finie $[t_{min}, t_{max}]$. Cela pose problème pour calculer des coefficients d'ondelettes correspondant à des ondelettes dont le support déborde de cet intervalle temporel.

Théoriquement, le support des ondelettes de Morlet est infini, étant défini par une fonction exponentielle décroissante. Cependant l'amplitude de ces ondelettes est négligeable en dehors d'un certain *support effectif*. Le problème des effets de bord persiste donc pour les ondelettes de Morlet dont le support effectif dépasse $[t_{min}, t_{max}]$.

Soit une ondelette fille $\psi_{\tau,a}$ de Morlet, son support effectif est défini comme l'intervalle $[\tau - \Delta t_{\tau,a}, \tau + \Delta t_{\tau,a}]$, où $\Delta t_{\tau,a}$ est défini de telle sorte que l'amplitude de $\psi_{\tau,a}$ est négligeable en $\tau \pm \Delta t_{\tau,a}$. Cette dernière propriété est respectée si, pour un α petit fixé,

$$\alpha \geq \frac{|\psi_{\tau,a}(\tau \pm \Delta t_{\tau,a})|}{|\psi_{\tau,a}(\tau)|} \Leftrightarrow \Delta t_{\tau,a} \leq a\sqrt{-2\ln\alpha}. \tag{2.25}$$

Une approche courante pour gérer le problème des bords d'un signal consiste à considérer que $s(t) = 0$ pour $t \notin [t_{min}, t_{max}]$. L'inconvénient de cette approche, dite de *bourrage de zéros*, est qu'elle ajoute de l'information au signal. Les coefficients d'ondelettes obtenus aux bords ne sont donc pas entièrement déterminés par le contenu du signal. Par conséquent, je préfère employer une autre approche qui consiste à ne pas prendre en compte les coefficients d'ondelettes calculés au bord. Dans le cas des ondelettes de Morlet, cela signifie rejeter les coefficients issus d'ondelettes filles $\psi_{\tau,a}$ si

$$[\tau - \Delta t_{\tau,a}, \tau + \Delta t_{\tau,a}] \not\subset [t_{min}, t_{max}]. \tag{2.26}$$

où $\Delta t_{\tau,a}$ est calculé à l'aide de l'équation (2.25), avec typiquement $\alpha = 0{,}001$.

2.2. Modélisation bayésienne

En science expérimentale, le recueil des données s'accompagne souvent d'une modélisation de celles-ci. Le modèle est une simplification des données, une réduction à un ensemble de variables, dont les données sont les observations, et de paramètres, le tout étant lié par des équations mathématiques. Dans le cas d'un modèle statistique, le modèle comprend des variables aléatoires pour rendre compte des données observées. Pour ces variables aléatoires, on suppose qu'elles peuvent prendre différentes valeurs de manière aléatoire, même si une seule réalisation est observée dans les données.

La fluctuation aléatoire qui est capturée par ce type de modèle peut être de nature très diverse : bruit de mesure d'un appareil, aléa quantique mais aussi phénomènes chaotiques peu prévisibles. Il peut aussi et surtout s'agir de tout ce que le modèle ne peut pas modéliser finement et est donc considéré comme « bruit ».

Exemple : Une électrode mesure un potentiel dans une région du cerveau d'un macaque rhésus. Un modèle statistique, fantaisiste, peut considérer que les valeurs de potentiel échantillonnées sont issues d'un simple modèle gaussien. Soit l'ensemble $\mathbf{x} = \{\mathrm{x}_n\}_{n=1}^N$ des valeurs de potentiel observées, alors le modèle est un ensemble de variables aléatoires $\boldsymbol{X} = \{X_n\}_{n=1}^N$ dont les $\mathbf{x}$ sont les observations. Chaque variable aléatoire X_n est associée à une distribution gaussienne de paramètre de moyenne μ et de paramètre d'écart-type σ. La densité de probabilité des X_n s'écrit alors :

$$p(x_n) = \mathcal{N}(x_n|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}\ \mathrm{e}^{-\frac{1}{2}\left(\frac{x_n-\mu}{\sigma}\right)^2} \tag{2.27}$$

pour tout réel x_n. Deux questions se posent alors immédiatement vis-à-vis de ce modèle : est-il approprié pour décrire les données ? Si oui, quelles sont les valeurs de μ et de σ^2 ?

L'*inférence statistique* regroupe les techniques servant à répondre à ce type de questions, telles que l'estimation de paramètres d'un modèle statistique ou la sélection d'un modèle adapté aux données. L'inférence bayésienne constitue une branche particulière qui utilise la théorie bayésienne pour mener à bien cette tâche.

Ainsi, dans cette seconde partie, l'élaboration de modèles statistiques dans le contexte bayésien sera décrite dans un premier temps. Dans un deuxième temps la *théorie de la décision*, qui sert à répondre aux questions d'estimation de paramètres et de sélection de modèle à l'aide des modèles bayésiens, sera introduite. Dans un dernier temps, l'algorithme VMP qui me sert à calculer numériquement les valeurs nécessaires à l'inférence bayésienne sera présenté.

2.2.1. Modèles bayésiens : distribution a priori des paramètres

Ce qui distingue un modèle dit « bayésien » d'un modèle statistique fréquentiste est le fait que les paramètres sont traités comme des variables aléatoires. En effet, une distribution est associée à chaque paramètre du modèle aléatoire. Ces distributions sur les paramètres sont dites *a priori* car elles ne dépendent pas des variables aléatoires modélisant les

données observées et reflètent ainsi les connaissances sur les paramètres avant observation des données. Les paramètres non aléatoires des distributions des paramètres aléatoires sont appelés *hyperparamètres*[2].

Un modèle bayésien est caractérisé par sa probabilité jointe[3]. Pour un modèle définissant un ensemble de variables aléatoires $\boldsymbol{Y}$ et de paramètres $\boldsymbol{\theta}$, cette probabilité jointe se décompose en

$$p(\boldsymbol{y}, \boldsymbol{\theta}) = p(\boldsymbol{y}|\boldsymbol{\theta})\, p(\boldsymbol{\theta}) \tag{2.28}$$

où $p(\boldsymbol{y}|\boldsymbol{\theta})$ est la probabilité conditionnelle de $\boldsymbol{Y}$ sachant $\boldsymbol{\theta}$ et $p(\boldsymbol{\theta})$ est la probabilité a priori de $\boldsymbol{\theta}$. La probabilité a priori de $\boldsymbol{\theta}$ reflète bien le fait que la distribution a priori des paramètres $\boldsymbol{\theta}$ ne dépend pas des variables aléatoires $\boldsymbol{Y}$.

Exemple (suite) : Pour transformer le modèle gaussien simple en modèle bayésien, il suffit d'associer une distribution a priori pour μ et pour σ, de densité de probabilité $p(\mu, \sigma)$. La densité de probabilité jointe du modèle s'écrit alors :

$$p(\boldsymbol{x}, \mu, \sigma) = p(\boldsymbol{x}|\mu, \sigma)\, p(\mu, \sigma) = \Pi_{n=1}^{N}\, p(x_n|\mu, \sigma)\, p(\mu, \sigma), \tag{2.29}$$

en supposant l'indépendance entre les variables X_n.

Il est possible de conditionner la distribution d'un paramètre sur celle d'un autre pour décomposer davantage la distribution jointe. En conditionnant μ sur σ, la densité de probabilité jointe des paramètres devient $p(\mu, \sigma) = p(\mu|\sigma)\, p(\sigma)$. Une autre possibilité consiste à ajouter une hypothèse d'indépendance entre les paramètres, ce qui se traduit par la densité de probabilité jointe $p(\mu, \sigma) = p(\mu)\, p(\sigma)$.

La distribution a priori des paramètres est le lieu où les hypothèses sur le modèle et les connaissances sur les données sont insérées. Cela se traduit dans le choix de la forme paramétrique des distributions a priori, mais aussi dans la manière de subdiviser la distribution jointe des paramètres, par conditionnement de certains paramètres sur d'autres ou l'ajout d'hypothèses d'indépendances entre des paramètres.

Un modèle bayésien ne se limite pas à l'ajout de distributions a priori sur les paramètres des distributions des variables aléatoires liées aux observations. Il est aussi possible de définir des distributions a priori sur les paramètres des distributions a priori des paramètres, et ainsi de suite. Le modèle est alors dit *hiérarchique*, empilant une hiérarchie de variables aléatoires, les unes étant les paramètres des distributions des autres.

Au niveau de la terminologie, les paramètres d'un modèle bayésien étant des variables aléatoires, le terme de *variables cachées* ou *variables latentes* pourra être employé pour

2. Dans la suite du manuscrit, je ferai référence à l'ensemble des hyperparamètres des distributions des paramètres d'un modèle sous le terme d'hyperparamètres du modèle.

3. Le terme « probabilité jointe » est ici utilisé pour désigner une densité de probabilité ou une fonction de masse, suivant qu'il s'agisse de variables aléatoires continues ou discrètes. Il en est de même pour les probabilités conditionnelles, a priori, a posteriori et marginales employées plus tard. Cet usage, pour abusif qu'il soit, permet d'alléger la terminologie dans les cas où aucune distinction sur la nature discrète ou continue des variables aléatoire n'est nécessaire.

les désigner. Les variables aléatoires du modèle associées à des données observées sont quant à elles appelées *variables observées*. Ces termes reflètent le fait que variables cachées et observées sont de même nature mathématique, l'unique distinction venant de la disponibilité de données observées pour certaines variables et non pour d'autres.

Probabilités a posteriori et probabilité marginale

Le *théorème de Bayes* [Bayes, 1763] énonce que, pour deux variables aléatoires X et Y,

$$p(y|x) = \frac{p(x|y)\,p(y)}{p(x)} \tag{2.30}$$

où $p(y|x)$ est la probabilité *a posteriori* de Y. En utilisant la formule des probabilités totales, le dénominateur $p(x)$ peut être réécrit, dans le cas de variables aléatoires continues, sous la forme suivante :

$$p(x) = \int_{\mathcal{D}} p(x|y)\,p(y)\,dy \tag{2.31}$$

où $\mathcal{D}$ est le domaine de définition de la distribution a priori de Y. Dans le cas de variables aléatoires discrètes, l'intégrale est remplacée par une somme sur toutes les valeurs possibles de Y. La probabilité $p(x)$ ainsi obtenue par intégration ou somme est appelée *probabilité marginale* de X.

Il est important de rappeler la signification de la notion de probabilité pour les modèles bayésiens. Une probabilité correspond à un niveau de *croyance*. Dès lors, la probabilité a priori capture la croyance préliminaire sur les variables cachées d'un modèle bayésien, c'est-à-dire l'expertise que l'on a sur les valeurs possibles de ces variables. La probabilité a posteriori correspond alors à cette croyance sur les variables cachées mise à jour au vu des données observées. Le mécanisme de passage d'une croyance a priori à une croyance a posteriori est appelé *révision des croyances*.

Les probabilités a posteriori et marginales sont des densités de probabilité ou des fonctions de masse suivant la nature des variables aléatoires qu'elles qualifient. Aussi, la probabilité a posteriori d'une variable cachée d'un modèle définit une distribution a posteriori pour cette variable. De la même manière, la probabilité marginale d'une variable observée définit une distribution marginale pour cette variable.

La théorie de la décision présentée à la section 2.2.2 exploitera ces distributions a posteriori et marginale pour répondre aux questions d'estimation de paramètres et de sélection de modèle.

Distributions a priori et modèle exponentiel-conjugué

À première vue, le choix des distributions a priori des variables cachées d'un modèle bayésien peut sembler arbitraire. Cependant, il est gouverné par l'information disponible a priori sur les variables cachées, avant toute observation de données. Si beaucoup d'information est disponible, il suffit de la traduire sous la forme d'une distribution adéquate. En cas de manque d'information, la distribution a priori est choisie pour son influence minimale sur la distribution a posteriori [Robert, 2006].

Une autre contrainte, plus artificielle, qui pèse sur le choix des distributions a priori concerne la facilité de calcul des distributions a posteriori. En effet, une difficulté inhérente à l'approche bayésienne provient du calcul de la probabilité marginale servant à normaliser la probabilité a posteriori dans la formule de Bayes. Cette probabilité marginale nécessite le calcul d'une intégrale, définie à l'équation (2.31), qui n'est pas toujours calculable analytiquement. L'algorithme VMP décrit en section 2.2.3 permet de s'affranchir de ces difficultés via une approximation déterministe. Cependant, cet algorithme ne fonctionne qu'avec un *modèle exponentiel-conjugué* : il s'agit d'un modèle bayésien dont toutes les variables ont une distribution issue d'une famille *exponentielle* et dont les distributions a priori des variables cachées sont conjuguées.

Une famille exponentielle est une famille de distributions paramétriques dont la densité de probabilité dans le cas continu, ou la fonction de masse dans le cas discret, peut s'exprimer sous la forme

$$p(x|\boldsymbol{\theta}) = \exp(\eta(\boldsymbol{\theta})^\top \cdot \mathrm{u}(x) + f(x) + g(\boldsymbol{\theta})), \tag{2.32}$$

où $\boldsymbol{\theta}$ représente l'ensemble des paramètres des distributions de cette famille. Les vecteurs $\eta(\boldsymbol{\theta})$ et $\mathrm{u}(x)$ sont respectivement les *paramètres naturels* et les *statistiques naturelles* des distributions de la famille, et $g(\boldsymbol{\theta})$ est la constante de normalisation [Winn, 2003]. La fonction $f(x)$ est spécifique à la famille exponentielle, et contient les termes du logarithme de la densité de probabilité qui dépendent de x et n'impliquent pas $\boldsymbol{\theta}$. Il est courant de désigner les distributions d'une famille exponentielle par le nom de la famille. Par exemple, une distribution gaussienne de moyenne $\mu = 2$ et d'écart-type $\sigma = 5$ est une distribution de la famille des distributions gaussiennes.

La distribution a priori d'une variable cachée est conjuguée si elle a la même forme que la distribution a posteriori de cette variable [Beal, 2003]. Ainsi, si X est une variable observée ayant comme paramètres de sa distribution la variable cachée Y et d'autres variables $\boldsymbol{\beta}$, la distribution conjuguée de Y est définie de telle sorte que

$$p(y|x, \boldsymbol{\beta}) = f(y|\boldsymbol{\theta'}) \;\propto\; p(x|y, \boldsymbol{\beta})\, f(y|\boldsymbol{\theta}), \tag{2.33}$$

où f est une densité de probabilité ou une fonction de masse. Dans le cas particulier où X et Y ont des distributions issues de familles exponentielles, cela signifie que les distributions a priori et a posteriori de Y sont issues de la même famille exponentielle. De plus, la distribution de X peut s'écrire en fonction du vecteur de statistiques naturelles $\mathrm{u}_y(y)$ de Y [Winn, 2003] :

$$p(x|y, \boldsymbol{\beta}) = \exp(\ell_{x,y}(x, \boldsymbol{\beta})^\top \cdot \mathrm{u}_y(y) + h_{x,y}(x, \boldsymbol{\beta})), \tag{2.34}$$

où $\ell_{x,y}$ et $h_{x,y}$ sont spécifiques à la famille exponentielle de X.

De par ses contraintes sur sa structure, un modèle exponentiel-conjugué apporte de l'information aux distributions a priori du modèle, qui sont alors nécessairement informatives [Robert, 2006]. Il est cependant possible de rendre ces distributions a priori faiblement informatives, ou *vagues*, en choisissant de manière appropriée la valeur de leurs paramètres non aléatoires pour réduire leur impact sur la distribution a posteriori.

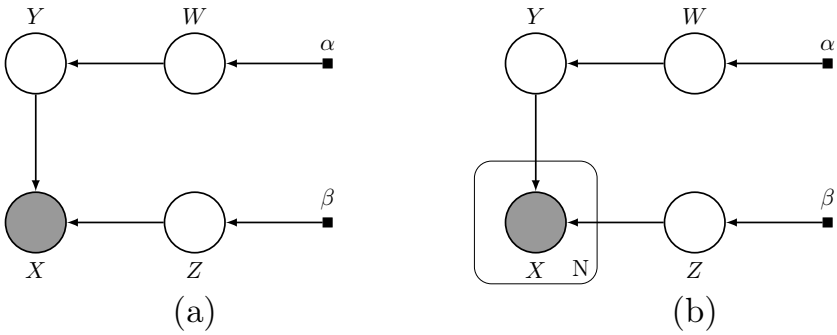

FIGURE 2.1. – Modèles graphiques probabilistes représentant la distribution jointe des modèles décris par (a) l'équation (2.35) et (b) l'équation (2.36).

Modèles graphiques

Tout modèle bayésien peut être représenté sous la forme d'un graphe acyclique dirigé. Plus précisément, la représentation graphique correspond à la décomposition de la probabilité jointe du modèle, avec la factorisation induite par les différentes hypothèses d'indépendance conditionnelle faites sur les variables cachées et observées du modèle [Wainwright and Jordan, 2008].

Au sein du graphe, chaque nœud correspond à une variable cachée ou observée du modèle associé au graphe. Une couleur différente distingue les variables observées des variables cachées. Il est aussi possible de faire figurer les hyperparamètres du modèle sous la forme de nœuds ponctuels. Les flèches entre les nœuds indiquent les relations de conditionnement des variables entre elles : le nœud à la base de la flèche, ou *nœud parent*, symbolise une variable paramètre de la distribution de la variable du *nœud enfant* où pointe la flèche.

Une notation supplémentaire en « plaque » peut être introduite si certaines variables du modèle sont similaires et conditionnellement indépendantes les unes des autres. Ces variables sont alors représentées par un unique nœud au sein d'une plaque étiquetée par le nombre de nœuds similaires.

Exemple : Soit un modèle bayésien comprenant quatre variables aléatoires W, X, Y et Z, dont la probabilité jointe est définie par

$$p(x, y, z) = p(x|y, z)\, p(y|w)\, p(w|\alpha)\, p(z|\beta), \tag{2.35}$$

où α et β sont les hyperparamètres du modèle, paramètres respectivement des distributions de W et Z. La figure 2.1 (a) illustre le graphe correspondant à ce modèle.

En considérant une variante de ce modèle où X est remplacé par un ensemble de N variables aléatoires $\boldsymbol{X} = \{X_n\}_{n=1}^{N}$, la probabilité jointe du modèle s'écrit alors

$$p(\boldsymbol{x}, y, z) = \Pi_{n=1}^{N}\, p(x_n|y, z)\, p(y|w)\, p(w|\alpha)\, p(z|\beta). \tag{2.36}$$

Le graphe correspondant à ce modèle est illustré à la figure 2.1 (b).

La représentation en graphe d'un modèle bayésien permet d'accéder rapidement à l'information de structure du modèle. Ceci est d'autant plus utile pour des modèles complexes tels que les modèles hiérarchiques. Au-delà de cette facilitation de la lecture du modèle, le graphe peut aussi servir de support pour l'inférence du modèle, c'est-à-dire le calcul des probabilités a posteriori et marginales. L'algorithme VMP présenté plus bas utilise la structure du graphe pour découper ce calcul en calculs locaux aux nœuds du graphe.

2.2.2. Théorie de la décision, estimation et sélection de modèle

Comme présenté en introduction de cette partie sur les modèles bayésiens, l'utilisation d'un modèle pour caractériser des données s'accompagne de deux principales questions :

- Quelles sont les valeurs des paramètres du modèle ?
- Le modèle est-il approprié pour décrire les données recueillies dans l'absolu ? ou par rapport à un autre modèle ?

Suivant la manière dont le modèle est construit, celui-ci incorpore plus ou moins d'hypothèses sur les données. Répondre à ces questions permet alors de valider ou d'invalider ces hypothèses, en acceptant ou non le modèle et en caractérisant ses paramètres. Ce processus engendre une création de connaissance sur des données.

La théorie de la décision introduit un cadre formel qui permet d'exploiter les informations apportées par le modèle bayésien afin de répondre à ces questions. Je n'introduirai ici que quelques éléments relatifs à l'approche bayésienne, même si cette théorie s'étend aussi aux statistiques classiques.

Fonction de coût et coût moyen a posteriori

La première quantité que l'on définit avant de prendre une décision est la *fonction de coût*, aussi appelée *fonction de perte*. Cette fonction est notée $L(\theta, \delta)$, où la variable θ est l'objet de la décision et δ la règle de décision utilisée. Cette fonction quantifie les répercussions d'une mauvaise décision sur θ, par exemple une perte financière, de temps, etc.

Le *coût moyen a posteriori* correspond alors à l'espérance de la fonction de coût par rapport à la probabilité a posteriori de θ :

$$\mathbb{E}_{p(\theta|x)}[L(\theta, \delta)] = \int_{\mathcal{D}} L(\theta, \delta)\, p(\theta|x)\, dx, \tag{2.37}$$

avec $\mathcal{D}$ l'espace de définition des variables observées X du modèle bayésien traité.

À partir du coût moyen a posteriori, une règle de décision optimale δ^* peut être dérivée en tant que règle minimisant ce coût moyen [Robert, 2006] :

$$\delta^*(x) = \underset{\delta}{\operatorname{argmin}}\ \mathbb{E}_{p(\theta|x)}[L(\theta, \delta)]. \tag{2.38}$$

Estimation ponctuelle

Dans un modèle bayésien, les paramètres du modèle sont traités comme des variables aléatoires, ce sont les variables cachées du modèle. Dès lors, l'inférence bayésienne fournit déjà une information riche sur les paramètres du modèle à travers la probabilité a posteriori des variables cachées. Il arrive cependant que l'on souhaite résumer chaque paramètre par une valeur unique au lieu d'une distribution.

Pour l'estimation ponctuelle d'un paramètre θ, les fonctions de coût usuelles sont définies par l'écart entre la valeur d'un estimateur δ et la valeur de θ [Robert, 2006].

Le coût quadratique s'écrit $L(\theta, \delta) = (\theta - \delta(x))^2$. Ce coût est minimisé par l'espérance de la distribution a posteriori de θ.

Le coût absolu s'écrit $L(\theta, \delta) = |\theta - \delta(x)|$ et est minimisé par la médiane de la distribution a posteriori de θ.

Il est intéressant de noter que les estimateurs bayésiens définis dans le cadre de la théorie de la décision correspondent à des estimateurs connus en statistique classique. La théorie de la décision apporte ainsi une justification différente pour l'emploi de ces mêmes estimateurs, au travers des critères d'optimalité qu'elle définit.

Les estimateurs bayésiens décrits ci-dessus, la moyenne et la médiane a posteriori, sont convergents. Cela signifie que pour un nombre d'observations tendant vers l'infini ils tendent vers la valeur du paramètre estimé. Asymptotiquement, la contribution de la distribution a priori des paramètres dans leur estimation disparaît, et les estimateurs bayésiens deviennent sensiblement équivalents à des estimateurs du maximum de vraisemblance [Robert, 2006].

Validation de modèles bayésiens

Le problème de la validation d'un modèle vis-à-vis de données peut se décomposer en deux sous-problèmes :

- un problème de comparaison de modèles, pour connaître le modèle rendant le mieux compte des données parmi un ensemble de modèles,
- un problème d'adéquation du modèle, pour savoir si dans l'absolu un modèle est raisonnable pour caractériser des données.

Une grande force de l'approche bayésienne est de pouvoir répondre facilement au premier sous-problème. Une faiblesse de cette même approche est sa difficulté à répondre au second sous-problème [Little, 2006].

La comparaison de modèles peut se traduire sous la forme d'un problème d'estimation d'une variable cachée discrète. Dans le cas de la comparaison de deux modèles $\mathcal{M}_1$ et $\mathcal{M}_2$ sur un même jeu de données, un méta-modèle les regroupant tous deux est construit, dont la probabilité jointe s'écrit :

$$p(\boldsymbol{x}, m) = p(\boldsymbol{x}|m)\, p(m), \tag{2.39}$$

avec $\boldsymbol{X}$ un ensemble de variables observées liées aux données et M une variable cachée discrète à valeur dans $\{\mathcal{M}_1, \mathcal{M}_2\}$. Dès lors, choisir un des deux modèles correspond à

minimiser un coût moyen a posteriori $\mathbb{E}_{p(m|x)}[L(m,\delta)]$ pour l'estimation de la valeur de M, avec une fonction de coût dite « 0-1 » :

$$L(m,\delta) = \begin{cases} 1 & \text{si } m \neq \delta(\boldsymbol{x}), \\ 0 & \text{sinon.} \end{cases} \tag{2.40}$$

L'estimateur δ minimisant ce coût moyen correspond au maximum a posteriori, aussi appelé *estimateur MAP* [Robert, 2006] :

$$\delta^*(\boldsymbol{x}) = \underset{m}{\operatorname{argmax}}\, p(m|\boldsymbol{x}). \tag{2.41}$$

Sous l'hypothèse d'équiprobabilité a priori des deux modèles, l'estimateur δ devient le maximum a priori :

$$\delta^*(\boldsymbol{x}) = \underset{m}{\operatorname{argmax}}\, p(\boldsymbol{x}|m) \quad \text{si} \quad p(m) = \frac{1}{2} \quad \forall m \in \{\mathcal{M}_1, \mathcal{M}_2\}. \tag{2.42}$$

Ceci est équivalent à une comparaison des probabilités marginales des variables observées $\boldsymbol{X}$ dans chaque modèle, $p(\boldsymbol{x}|\mathcal{M}_1)$ et $p(\boldsymbol{x}|\mathcal{M}_2)$, de telle sorte que le modèle choisi est celui qui a la plus grande probabilité marginale. Ce résultat se généralise pour une comparaison de multiples modèles vis-à-vis d'un même jeu de données.

La comparaison de modèles via leur probabilité marginale entraîne, lors d'un choix entre des modèles de complexité différente, un compromis entre simplicité du modèle et surajustement aux données. Cette propriété vient de l'intégration sur les variables cachées du modèle lors du calcul de la probabilité marginale, tel que défini à l'équation (2.31). Dans cette intégration, la probabilité a priori des variables cachées joue un rôle de terme pénalisant vis-à-vis de la complexité du modèle. Le compromis obtenu par la comparaison des probabilités marginales est souvent comparé au *Rasoir d'Ockham* [Jefferys and Berger, 1991] comme les modèles privilégiés font l'économie d'une complexité inutile. Une mise en pratique de ce principe sera présentée avec les modèles de mélange, au chapitre 3.

Concernant l'évaluation de l'adéquation d'un modèle bayésien à un jeu de données, l'approche bayésienne n'offre pas de solution simple. Certains auteurs comme Little préconisent l'emploi des statistiques classiques pour l'évaluation des modèles [Little, 2006], dans une approche dite *bayésienne ajustée* (Calibrated Bayes). Dans cette optique, des outils fondés sur la distribution prédictive a posteriori ont notamment été développés [Gelman et al., 2004]. Une autre démarche, totalement bayésienne, a été proposée par Starkman *et al.*, dans laquelle l'évaluation d'un modèle est envisagée comme une comparaison de ce modèle avec un modèle totalement inconnu potentiellement meilleur [Starkman et al., 2008, March et al., 2011]. Cette dernière approche semble prometteuse même si la définition du modèle inconnu fait appel à de nombreuses approximations.

De manière plus pragmatique, il est aussi possible d'évaluer un modèle vis-à-vis d'une tâche précise, l'adéquation du modèle aux données correspondant à sa performance sur cette tâche. Il faut cependant garder à l'esprit que la validation par la performance est effectuée sur des données artificielles dont le processus de génération est totalement connu, mais qui n'est pas toujours le reflet du processus de génération des données expérimentales

qui seront traitées par la suite. Concernant les modèles bayésiens que j'ai développés, ceux-ci ont pour finalité la détection de certains types de synchronisations cérébrales. Je m'appuierai donc sur l'efficacité de cette détection pour juger de la pertinence des différents modèles.

2.2.3. Calcul des probabilités a posteriori et marginale

L'inférence statistique est conceptuellement simple avec les modèles bayésiens. En effet, celle-ci se résume au calcul des probabilités a posteriori et marginale des modèles, à l'aide des équations (2.30) et (2.31), puis à l'application de la théorie de la décision. Hélas, l'intégrale (2.31) de la probabilité marginale n'est pas toujours calculable de manière analytique [Bishop, 2006]. Pour contourner ce problème, différentes approximations des probabilités a posteriori et marginale ont été mises au point. Deux grands courants se distinguent dans les méthodes d'approximations : les méthodes d'approximations stochastiques et les méthodes d'approximations déterministes.

Les méthodes d'approximations stochastiques reposent sur des méthodes de Monte Carlo, la plus connue étant la méthode de Monte Carlo à base de chaînes de Markov (Markov Chain Monte Carlo ou MCMC). L'intégrale à calculer est alors approchée par une somme d'échantillons générés aléatoirement par une chaîne de Markov dont la distribution stationnaire correspond à la distribution à approximer [Neal, 1993, MacKay, 1998]. Les méthodes MCMC ont favorisé l'essor des modèles bayésiens en permettant le calcul de modèles hiérarchiques complexes. Elles présentent néanmoins des inconvénients majeurs, tels qu'un temps de calcul et une consommation mémoire élevée, ainsi que des problèmes d'évaluation de la convergence de la chaîne de Markov [Gill, 2008, Grimmer, 2011].

Les méthodes d'approximations déterministes consistent à approcher les probabilités a posteriori et marginales de manière analytique. Une large part de ces méthodes sont formulées comme la minimisation d'une distance, ou plus généralement d'une divergence, entre la probabilité a posteriori du modèle et une fonction d'approximation [Minka, 2005]. Ces méthodes sont dites *variationnelles*, parce qu'elles se résument à un problème d'optimisation sur un ensemble de fonctions. Le caractère approximatif de ces méthodes vient alors des restrictions posées sur la fonction d'approximation : il peut s'agir d'une forme paramétrique particulière, par exemple la densité de probabilité d'une distribution gaussienne multivariée [Opper and Archambeau, 2009], ou reposer sur des contraintes structurelles comme celle d'avoir une probabilité a posteriori factorisable [Attias, 2000]. Cette dernière approximation s'appelle *l'approximation naïve par champ moyen*. Cette méthode s'étant révélée efficace numériquement pour un large spectre de modèles, tels que l'analyse en composantes principales [Bishop, 1999], l'analyse en composantes indépendantes [Miskin, 2000], les modèles de Markov [MacKay, 1997], et de nombreux autres [Winn, 2003, Beal, 2003], j'ai décidé de l'employer pour l'inférence de mes modèles bayésiens.

L'approximation naïve par champ moyen

La méthode d'approximation naïve par champ moyen se fonde sur la minimisation de la *divergence de Kullback-Leibler*, aussi appelée divergence KL, entre la probabilité a posteriori d'un modèle et son approximation. Pour un modèle bayésien définissant un ensemble de variables observées $\boldsymbol{X}$ et cachées $\boldsymbol{Y}$, l'approximation q de la probabilité a posteriori minimise la divergence KL

$$\mathrm{KL}(q||p) = \int q(\boldsymbol{y}) \ln \frac{q(\boldsymbol{y})}{p(\boldsymbol{y}|\boldsymbol{x})}\, d\boldsymbol{y}. \tag{2.43}$$

Cette divergence KL est nulle si et seulement si $q(\boldsymbol{y}) = p(\boldsymbol{y}|\boldsymbol{x})\ \forall x$, et strictement positive dans tout autre cas.

En substituant $p(\boldsymbol{y}|\boldsymbol{x}) = p(\boldsymbol{y}, \boldsymbol{x})/p(\boldsymbol{x})$ dans l'équation (2.43), la probabilité marginale de $\boldsymbol{X}$ apparaît dans la formule de la divergence [Winn, 2003] :

$$\mathrm{KL}(q||p) = \ln p(\boldsymbol{x}) - \underbrace{\int q(\boldsymbol{y}) \ln \frac{p(\boldsymbol{y}, \boldsymbol{x})}{q(\boldsymbol{y})}\, d\boldsymbol{y}}_{\mathcal{L}(q)}. \tag{2.44}$$

Sachant que la divergence KL est positive ou nulle, en réarrangeant l'équation (2.44) sous la forme

$$\ln p(\boldsymbol{x}) = \mathrm{KL}(q||p) + \mathcal{L}(q), \tag{2.45}$$

il apparaît plus clairement que $\mathcal{L}(q)$ est une *borne inférieure* du logarithme de la probabilité marginale, $\ln p(\boldsymbol{x}) \geq \mathcal{L}(q)$. La minimisation de $\mathrm{KL}(q||p)$ est équivalente à la maximisation de la borne inférieure $\mathcal{L}(q)$.

L'hypothèse de champ moyen impose que l'approximation q soit une distribution factorisable sur l'ensemble des variables $\boldsymbol{Y}$:

$$p(\boldsymbol{y}|\boldsymbol{x}) \approx q(\boldsymbol{y}) \quad \text{où} \quad q(\boldsymbol{y}) = \prod_{y_i \in \boldsymbol{y}} q_i(y_i). \tag{2.46}$$

Les q_i sont donc les probabilités a posteriori approchées de chaque variable cachée Y_i, considérées comme indépendantes a posteriori.

La substitution de q par sa forme factorisée dans l'équation de $\mathcal{L}(q)$ donne

$$\mathcal{L}(q) = \int \prod_{y_i \in \boldsymbol{y}} q_i(y_i) \ln \frac{p(\boldsymbol{y}, \boldsymbol{x})}{\prod_{y_i \in \boldsymbol{y}} q_i(y_i)}\, d\boldsymbol{y}. \tag{2.47}$$

Cette décomposition peut être reformulée pour se focaliser sur une seule des variables

cachées, notée Y_j, et sa probabilité a posteriori approchée q_j :

$$\begin{aligned}\mathcal{L}(q) &= \int q_j(y_j) \Bigg[\underbrace{\left(\int \prod_{y_i \in \boldsymbol{y}\backslash\{y_j\}} q_i(y_i) \ln p(\boldsymbol{y},\boldsymbol{x})\, d(\boldsymbol{y}\backslash\{y_j\}) \right)}_{\mathbb{E}_{q(\boldsymbol{y}\backslash\{y_j\})}[\ln p(\boldsymbol{x},\boldsymbol{y})]} - \ln q_j(y_j) \Bigg] dy_j \\ &\quad - \sum_{y_i \in \boldsymbol{y}\backslash\{y_j\}} \int q_i(y_i) \ln q_i(y_i)\, dy_i \\ &= \underbrace{- \int q_j(y_j) \ln \frac{q_j(y_j)}{\exp\left(\mathbb{E}_{q(\boldsymbol{y}\backslash\{y_j\})}[\ln p(\boldsymbol{x},\boldsymbol{y})]\right)} dy_j}_{\text{divergence KL}} \underbrace{- \sum_{y_i \in \boldsymbol{y}\backslash\{y_j\}} \int q_i(y_i) \ln q_i(y_i)\, dy_i}_{\text{constant par rapport à } q_j}\end{aligned} \tag{2.48}$$

où $\boldsymbol{y}\backslash\{y_j\}$ est l'ensemble des variables $\boldsymbol{y}$ privé de y_j.

La maximisation de $\mathcal{L}(q)$ par rapport à q_j revient à minimiser une divergence KL qui s'annule en q_j^*, définie comme suit :

$$\begin{aligned}\ln q_j^*(y_j) &= \mathbb{E}_{q(\boldsymbol{y}\backslash\{y_j\})}[\ln p(\boldsymbol{x},\boldsymbol{y})] + \text{const.} \\ \Leftrightarrow q_j^*(y_j) &= \frac{\exp\left(\mathbb{E}_{q(\boldsymbol{y}\backslash\{y_j\})}[\ln p(\boldsymbol{x},\boldsymbol{y})]\right)}{\int \exp\left(\mathbb{E}_{q(\boldsymbol{y}\backslash\{y_j\})}[\ln p(\boldsymbol{x},\boldsymbol{y})]\right) dy_i},\end{aligned} \tag{2.49}$$

comme q_j^* est une densité de probabilité dont l'intégrale vaut 1. La forme analytique de q_j^* est dépendante d'espérances calculées avec les autres q_i pour $i \neq j$.

Dès lors, $\mathcal{L}(q)$ peut être maximisée en égalisant successivement chaque q_i au q_i^* correspondant, calculé via l'équation (2.49). La valeur de $\mathcal{L}(q)$ est ainsi optimisée tour à tour pour chaque q_i et croît de manière monotone. Le suivi de la valeur de $\mathcal{L}(q)$ permet alors d'évaluer la convergence de la borne inférieure en un maximum.

L'algorithme Variational Message Passing

Dans le cas d'un modèle bayésien exponentiel-conjugué, Winn et Bishop ont développé l'algorithme *Variational Message Passing* (ou algorithme VMP) [Winn, 2003, Winn and Bishop, 2005] pour dériver automatiquement les formes analytiques des q_i^* et les optimiser, c'est-à-dire maximiser la borne $\mathcal{L}(q)$. Cet algorithme fonctionne sur des modèles graphiques en utilisant les nœuds comme support pour le calcul, par passage de messages entre les nœuds.

L'ensemble des variables parents de la variable cachée Y_j, c'est-à-dire ses paramètres, sera noté $\boldsymbol{Pa}_j$. De la même manière, l'ensemble des variables enfants de Y_j, c'est-à-dire les variables paramétrées par Y_j, sera noté $\boldsymbol{Ch}_j$. Soit une variable enfant Ch de Y_j, l'ensemble des co-parents de Y_j par rapport à Ch sera noté $\boldsymbol{Cp}_{ch,j}$ [4].

4. Les notations en minuscules $\boldsymbol{pa}_j$, $\boldsymbol{ch}_j$ et $\boldsymbol{cp}_{ch,j}$ seront utilisées pour désigner respectivement les valeurs des variables de $\boldsymbol{Pa}_j$, $\boldsymbol{Ch}_j$ et $\boldsymbol{Cp}_{ch,j}$.

Au sein de l'équation (2.49), l'espérance $\mathbb{E}_{q(\boldsymbol{y}\setminus\{y_j\})}[\ln p(\boldsymbol{x},\boldsymbol{y})]$ peut alors être décomposée pour ne garder que les termes dépendants de Y_j, via sa probabilité a priori ou la probabilité a priori de ses variables enfants, tout le reste étant mis de côté dans le terme constant :

$$\ln q_j^*(y_j) = \mathbb{E}_{q(\boldsymbol{pa}_j)}[\ln p(y_j|\boldsymbol{pa}_j)] + \sum_{ch\,\in\,\boldsymbol{ch}_j} \mathbb{E}_{q(ch,\boldsymbol{cp}_{ch,j})}[\ln p(ch|y_j,\boldsymbol{cp}_{ch,j})] + \text{const.} \quad (2.50)$$

Si le modèle est exponentiel-conjugué, alors la probabilité a priori de Y_j peut s'écrire sous la forme :

$$\ln p(y_j|\boldsymbol{pa}_j) = \eta_j(\boldsymbol{pa}_j)^\top \cdot \mathrm{u}_j(y_j) + f_j(y_j) + g_j(\boldsymbol{pa}_j). \quad (2.51)$$

La variable Y_j étant conjuguée avec ses variables enfants $\boldsymbol{Ch}_j$, il est possible d'utiliser l'équation (2.34) pour reformuler la probabilité a priori de chaque variable enfant Ch :

$$\ln p(ch|y_j,\boldsymbol{cp}_{ch,j}) = \ell_{ch,j}(ch,\boldsymbol{cp}_{ch,j})^\top \cdot \mathrm{u}_j(y_j) + h_{ch,j}(ch,\boldsymbol{cp}_{ch,j}). \quad (2.52)$$

L'utilisation des équations (2.51) et (2.52) dans l'équation (2.50) donne la forme analytique de q_j^* suivante :

$$\begin{aligned}\ln q_j^*(y_j) = &\left(\mathbb{E}_{q(\boldsymbol{pa}_j)}[\eta_j(\boldsymbol{pa}_j)] + \sum_{ch\,\in\,\boldsymbol{ch}_j} \mathbb{E}_{q(ch,\boldsymbol{cp}_{ch,j})}[\ell_{ch,j}(ch,\boldsymbol{cp}_{ch,j})]\right)^\top \cdot \mathrm{u}_j(y_j) \\ &+ f_j(y_j) + \text{const.}\end{aligned} \quad (2.53)$$

Soient $\tilde{\eta}_j$ et $\tilde{\ell}_{ch,j}$ des re-paramétrisations respectivement de η_j et $\ell_{ch,j}$ en fonction des statistiques naturelles de leurs paramètres :

$$\eta_j(\boldsymbol{pa}_j) = \tilde{\eta}_j\Big(\Big\{\mathrm{u}_i(y_i)\Big\}_{y_i\,\in\,\boldsymbol{pa}_j}\Big) \quad \text{et} \quad \ell_{ch,j}(ch,\boldsymbol{cp}_{ch,j}) = \tilde{\ell}_{ch,j}\Big(\mathrm{u}_{ch}(ch), \Big\{\mathrm{u}_k(y_k)\Big\}_{y_k\,\in\,\boldsymbol{cp}_{ch,j}}\Big). \quad (2.54)$$

Par construction, $\tilde{\eta}_j$ et $\tilde{\ell}_{ch,j}$ sont multi-linéaires par rapport à ces statistiques naturelles. Par conséquent, en prenant en compte l'hypothèse de champ moyen qui force l'indépendance a posteriori des variables cachées $\boldsymbol{Y}$, la forme analytique de q_j^* s'écrit

$$\begin{aligned}&\ln q_j^*(y_j) \\ &= \left(\tilde{\eta}_j\Big(\Big\{\mathbb{E}_{q(y_i)}[\mathrm{u}_i(y_i)]\Big\}_{y_i\,\in\,\boldsymbol{pa}_j}\Big) + \sum_{ch\,\in\,\boldsymbol{ch}_j} \tilde{\ell}_{ch,j}\Big(\mathbb{E}_{q(ch)}[\mathrm{u}_{ch}(ch)], \Big\{\mathbb{E}_{q(y_k)}[\mathrm{u}_k(y_k)]\Big\}_{y_k\,\in\,\boldsymbol{cp}_{ch,j}}\Big)\right)^\top \cdot \mathrm{u}_j(y_j) \\ &\quad + f_j(y_j) + \text{const.}\end{aligned} \quad (2.55)$$

L'algorithme VMP spécifie alors des messages entre les nœuds du modèle graphique, de telle sorte que q_j^* peut être défini en fonction de ceux-ci :

$$\ln q_j^*(y_j) = \left(\tilde{\eta}_j\Big(\Big\{m_{Y_i\to Y_j}\Big\}_{Y_i\,\in\,\boldsymbol{Pa}_j}\Big) + \sum_{Ch\,\in\,\boldsymbol{Ch}_j} m_{Ch\to Y_j}\right)^\top \cdot \mathrm{u}_j(y_j) + f_j(y_j) + \text{const.} \quad (2.56)$$

où $m_{Y_i \to Y_j}$ est un message d'un nœud parent Y_i vers son nœud enfant Y_j et $m_{Ch \to Y_j}$ est un message d'un nœud enfant Ch vers son nœud parent Y_j :

$$\begin{aligned} m_{Y_i \to Y_j} &= \mathbb{E}_{q(y_i)}[\mathrm{u}_i(y_i)] \\ m_{Ch \to Y_j} &= \tilde{\ell}_{ch,j}\Big(\mathbb{E}_{q(ch)}[\mathrm{u}_{ch}(ch)], \Big\{ m_{Y_k \to Ch} \Big\}_{Y_k \in \boldsymbol{Cp}_{ch,j}}\Big). \end{aligned} \tag{2.57}$$

D'après l'équation (2.56), la probabilité a posteriori approximée de Y_j est de la même famille exponentielle que sa probabilité a priori, comme elles partagent toutes deux le même type de statistiques naturelles u_j et la même fonction f_j [5]. L'approximation optimale q_j^* est alors une densité de probabilité, ou une fonction de masse, ayant pour paramètres

$$\boldsymbol{pa}_j^* = \eta_j^{-1}\left(\tilde{\eta}_j\Big(\Big\{ m_{Y_i \to Y_j} \Big\}_{Y_i \in \boldsymbol{Pa}_j}\Big) + \sum_{Ch \in \boldsymbol{Ch}_j} m_{Ch \to Y_j} \right), \tag{2.58}$$

qui se calculent en récupérant les messages auprès des nœuds parents et enfants de Y_j. L'espérance a posteriori de Y_j, nécessaire pour les messages de Y_j vers ses nœuds enfants, se calcule avec une fonction déterministe $\mathfrak{u}_j$ des paramètres $\boldsymbol{pa}_j^*$, définie pour chaque famille exponentielle :

$$\mathbb{E}_{q_j^*(y_j)}[\mathrm{u}_j(y_j)] = \mathfrak{u}_j(\boldsymbol{pa}_j^*). \tag{2.59}$$

Ainsi, l'utilisation de l'algorithme VMP pour calculer l'approximation de la probabilité a posteriori d'un modèle exponentiel-conjugué se résume à :

1. spécifier le modèle bayésien, ce qui permet de déduire automatiquement la forme analytique de chaque q_i^*,
2. mettre à jour, tour à tour, chaque q_i en l'égalisant à q_i^* c'est-à-dire
   a) récupérer les messages des nœuds parents et enfants avec les équations (2.57),
   b) calculer les nouveaux paramètres $\boldsymbol{pa}_i^*$ associés à q_i^* avec l'équation (2.58).

Concernant les messages des nœuds enfant, si un nœud enfant correspond à une variable observée X_k, associée à une valeur observée x_k, alors l'espérance $\mathbb{E}_{q(x_k)}[\mathrm{u}_k(x_k)]$ est remplacée par $\mathrm{u}_k(\mathrm{x}_k)$.

Par ailleurs, il est possible de calculer la borne inférieure $\mathcal{L}(q)$ pour évaluer l'état de convergence de l'algorithme. La borne inférieure $\mathcal{L}(q)$ peut être décomposée en une somme de la contribution locale de chaque nœud du modèle graphique [Winn, 2003] :

$$\mathcal{L}(q) = \sum_{y_j \in \boldsymbol{y}} \mathcal{L}_{y_j} + \sum_{x_k \in \boldsymbol{x}} \mathcal{L}_{x_k}, \tag{2.60}$$

où $\mathcal{L}_{y_j}$ est la contribution de la variable cachée Y_j :

$$\mathcal{L}_{y_j} = \left(\tilde{\eta}_j\Big(\Big\{ m_{Y_i \to Y_j} \Big\}_{Y_i \in \boldsymbol{Pa}_j}\Big) - \eta_j(\boldsymbol{pa}_j^*) \right)^\top \cdot \mathfrak{u}_j(\boldsymbol{pa}_j^*) + \tilde{g}_j\Big(\Big\{ m_{Y_i \to Y_j} \Big\}_{Y_i \in \boldsymbol{Pa}_j}\Big) - g_j(\boldsymbol{pa}_j^*) \tag{2.61}$$

5. La constante de normalisation g_j de la famille exponentielle se déduit de u_j et f_j.

et $\mathcal{L}_{x_k}$ est la contribution de la variable observée X_k, associée à la donnée observée x_k :

$$\mathcal{L}_{x_k} = \tilde{\eta}_k\Big(\Big\{m_{Y_i \to X_k}\Big\}_{Y_i \in \boldsymbol{Pa}_k}\Big)^\top \cdot \mathrm{u}_k(\mathrm{x}_k) + f_k(\mathrm{x}_k) + \tilde{g}_k\Big(\Big\{m_{Y_i \to X_k}\Big\}_{Y_i \in \boldsymbol{Pa}_k}\Big). \tag{2.62}$$

Les fonctions $\tilde{g}_j$ et $\tilde{g}_k$ sont des re-paramétrisations de g_j et g_k pour prendre en paramètres les statistiques naturelles des variables parents.

Par son approche fondée sur les modèles graphiques et des calculs locaux, l'algorithme VMP permet l'élaboration de boîtes à outils pour la construction et le calcul de modèles bayésiens. En effet, il suffit d'implémenter les messages des nœuds de différentes familles exponentielles, ainsi qu'un mécanisme de gestion d'un graphe de nœuds [6], pour pouvoir rapidement expérimenter avec des modèles variés. La forme analytique de chaque distribution a posteriori étant calculée automatiquement, le risque d'erreur d'analyse est réduit, accélérant ainsi la mise au point des modèles.

À ma connaissance, il existe deux principales implémentations de l'algorithme VMP : Vibes [7] en Java, et Infer.NET [8] sur la plateforme .NET. Le premier n'est plus maintenu et le second ne permet pas d'examiner ses codes sources. Pour l'évaluation de mes modèles bayésiens, j'ai développé ma propre bibliothèque dans le langage Python, reposant sur les bibliothèques de calcul Numpy [9] et Scipy [10].

Propriétés de l'approximation par champ moyen

L'utilisation de la divergence de KL comme objectif d'optimisation n'est pas sans répercussions sur la forme prise par l'approximation de la distribution a posteriori. La divergence KL(q||p) présentée à l'équation (2.43) a pour conséquence que l'approximation q est un « chercheur de mode » de la probabilité a posteriori. Cela signifie que q est centré sur un mode de $p(\boldsymbol{y}|\boldsymbol{x})$, comme la divergence KL force $q(\boldsymbol{y}) = 0$ là où $p(\boldsymbol{y}|\boldsymbol{x}) = 0$ [Minka, 2005, Bishop, 2006]. Par conséquent, il faudra faire attention à l'estimation de la dispersion de la distribution a posteriori qui sera sous-estimée par cette approximation.

Au niveau de l'optimisation, l'hypothèse de champ moyen implique que la borne inférieure optimisée est non convexe en général [Wainwright and Jordan, 2008]. La convergence n'est donc assurée que pour un maximum local. L'approche conventionnelle consiste alors à démarrer plusieurs fois la boucle de ré-estimation des q_i^* à partir d'initialisations aléatoires, puis à conserver l'approximation q qui amène à la plus haute borne inférieure $\mathcal{L}(q)$ après convergence. Il est aussi possible de guider l'optimisation vers de meilleurs résultats en choisissant un point de départ de bonne qualité, c'est-à-dire proche du maximum. De tels points de départ peuvent être obtenus pour des modèles complexes à partir de résultats de modèles plus simples.

6. Il s'agit là de vérifier à leur création que les nœuds respectent la contrainte de conjugaison, puis de gérer le passage des messages pour la mise à jour successive de chaque nœud.
7. http://vibes.sourceforge.net/
8. https://research.microsoft.com/en-us/um/cambridge/projects/infernet/
9. http://www.numpy.org/
10. http://www.scipy.org/

Dans le cadre de la comparaison de modèles, la borne inférieure $\mathcal{L}(q)$ peut être utilisée comme approximation du logarithme de la probabilité marginale. Beal et Ghahramani [Beal, 2003, Beal and Ghahramani, 2006] ont montré que cette borne inférieure se révèle plus performante pour la sélection de modèles que d'autres critères usuels tels que le *critère d'information bayésien* (Bayesian information criterion ou BIC [Schwarz, 1978]), bien qu'elle soit biaisée et favorise des modèles plus simples. La borne inférieure $\mathcal{L}(q)$ tend asymptotiquement vers le BIC, quand le nombre de variables observées tend vers l'infini.

Enfin, concernant la rapidité de l'algorithme, celle-ci est bien supérieure aux méthodes d'approximation stochastique [Beal, 2003]. Cependant la vitesse de convergence vers le point fixe peut être sévèrement impactée dans le cas de fortes corrélations entre les variables [Qi and Jaakkola, 2006].

2.3. Résumé

Ce chapitre a présenté en détail les fondements théoriques des deux outils principaux, la transformée en ondelettes continue et les modèles bayésiens, sur lesquels s'appuient mes méthodes de détection de synchronisations au sein de signaux cérébraux.

La transformée en ondelettes sert à projeter le signal dans un espace temps-fréquence pour révéler l'évolution de son contenu fréquentiel au cours du temps et ainsi faciliter l'étude de ce contenu fréquentiel. La force de cette transformée vient de sa capacité de multirésolution, c'est-à-dire sa capacité d'adaptation de la résolution dans le plan temps-fréquence. Ainsi les coefficients d'ondelettes en basses fréquences ont une résolution en temps faible et une résolution en fréquence élevée pour capturer les caractéristiques de signaux stationnaires. Inversement, les coefficients d'ondelettes en hautes fréquences disposent d'une résolution en fréquence faible mais d'une résolution en temps élevée, pour permettre de détecter des phénomènes transitoires. Pour l'étude de signaux cérébraux, la transformée en ondelettes a donc l'avantage de pouvoir détecter une grande diversité de phénomènes, tout en restant faiblement paramétrique.

L'ondelette qui sera utilisée par la suite est l'ondelette de Morlet complexe. L'avantage de cette ondelette est qu'elle présente une résolution temps-fréquence optimale, c'est-à-dire que l'aire de la boîte d'Heisenberg associée aux ondelettes filles est minimale.

Concernant les modèles bayésiens, ils permettent de rendre compte de données observées sous la forme d'équations paramétriques. L'intérêt spécifique de l'approche bayésienne provient du fait qu'elle modélise aussi la croyance associée aux paramètres du modèle et au modèle lui-même, sous la forme de probabilités a posteriori des paramètres et de probabilité marginale du modèle. Associées à la théorie de la décision, ces probabilités a posteriori et marginale servent à prendre des décisions optimales respectivement pour l'estimation des paramètres d'un modèle et le choix parmi plusieurs modèles de celui expliquant au mieux un jeu de données.

Lors de l'utilisation de modèles bayésiens, le calcul des probabilités a posteriori et marginale requiert le calcul d'intégrales n'ayant parfois pas de forme close. Dans ce contexte, l'algorithme Variational Message Passing (ou algorithme VMP) est une solution pour approcher les probabilités a posteriori et marginale. Cet algorithme offre l'avantage

d'être totalement automatique pour des modèles graphiques exponentiel-conjugués.

Chapitre 3.

Statistiques dans le plan temps-fréquence et modèles de mélange

> Tous les modèles sont faux, mais certains sont utiles.
>
> *Georges Box*

Le chapitre 1 a montré que l'étude de synchronisations locales au sein de signaux d'électrophysiologie cérébrale passe par l'étude de la répartition de l'énergie dans le plan temps-fréquence de ces signaux. Cette répartition d'énergie peut être identifiée à un scalogramme, obtenu à partir de coefficients d'ondelettes continues, tel que détaillé au chapitre 2.

Si l'on considère qu'une partie de la grande variabilité des signaux cérébraux est d'origine aléatoire, alors le scalogramme associé à ces signaux est lui aussi aléatoire et les coefficients d'ondelettes du signal correspondent à des variables aléatoires. Pour pouvoir étudier finement le scalogramme, une modélisation statistique des coefficients d'ondelettes est alors essentielle. Ainsi dans une première partie de ce chapitre, une étude de la distribution aléatoire des coefficients d'ondelettes sera proposée.

L'étude de la répartition temps-fréquence de l'énergie des signaux cérébraux porte notamment sur la comparaison de valeurs d'énergie entre des signaux, des enregistrements et des points temps-fréquence différents pour trouver des similarités et des différences. Ces comparaisons de valeurs d'énergie peuvent être formulées comme un problème de partitionnement, c'est-à-dire de séparation des valeurs en partitions d'énergie similaire. Dans le cadre d'une modélisation statistique des signaux, la recherche de ces partitions est possible en associant les valeurs d'énergie comparées à un modèle de mélange. La deuxième partie de ce chapitre sera consacrée à la présentation des modèles de mélange dans le cadre de la théorie bayésienne. Les parties suivantes présenteront deux modèles de mélanges adaptés à la distribution des coefficients d'ondelettes : le modèle de mélange gaussien et le modèle de mélange ricien. Le modèle de mélange ricien est un élément nouveau issu de mes travaux de thèse. Une dernière section exposera une évaluation de ces deux modèles sur des données artificielles, pour connaître leur capacité de détection de partitions dans un contexte appliqué.

3.1. Distribution des coefficients d'ondelettes de Morlet d'un signal aléatoire

Pour étudier le scalogramme d'un signal bruité, une première étape consiste à formuler la distribution du module au carré des coefficients d'ondelettes de ce signal. Pour trouver cette distribution, il suffit dans un premier temps de calculer la distribution du module des coefficients d'ondelettes.

3.1.1. Distribution du module en un point temps-fréquence

La première étape de la modélisation statistique d'un signal aléatoire consiste à définir la partie déterministe du signal et les hypothèses faites sur la partie aléatoire, aussi appelée bruit.

Soit $S_e(t)$ une famille de variables aléatoires représentant un signal s contaminé par un bruit blanc gaussien additif, le signal contaminé s'écrit :

$$S_e(t) = s(t) + B(t), \tag{3.1}$$

où $B(t)$ est une famille de variables aléatoires gaussiennes de moyenne nulle et non corrélées en temps :

$$p(b(t)) = \mathcal{N}(b(t)|0, \sigma_B) \quad \text{et} \quad \mathrm{Cov}(B(t), B(t')) = 0 \quad \forall t \neq t'. \tag{3.2}$$

Dans ce cas précis, Soltani Bozchalooi et Liang ont montré que le module des coefficients d'ondelettes $|\mathrm{W}s_e(\tau, a)|$ calculé avec l'ondelette de Morlet complexe est une famille de variables aléatoires suivant une distribution de Rice [Soltani Bozchalooi and Liang, 2007] :

$$p(|\mathrm{W}s_e(\tau, a)|) = \mathcal{R}(x|\nu_{\mathrm{W}}, \sigma_{\mathrm{W}}), \tag{3.3}$$

avec $\nu_{\mathrm{W}} = |\mathrm{W}s(\tau, a)|$ et $\sigma_{\mathrm{W}} = \sigma_B/\sqrt{2}$.

La distribution de Rice [Rice, 1945] est définie sur $[0, +\infty)$ et a pour densité de probabilité

$$\begin{aligned} p(x|\nu, \sigma) &= \mathcal{R}(x|\nu, \sigma) \\ &= I_0\left(\frac{x\nu}{\sigma^2}\right)\frac{x}{\sigma^2}\exp\left(\frac{-(x^2+\nu^2)}{2\sigma^2}\right) \\ &= \exp\left(\ln I_0\left(\frac{x\nu}{\sigma^2}\right) + \ln x - 2\ln\sigma - \frac{x^2+\nu^2}{2\sigma^2}\right) \quad \forall\, x \geq 0 \end{aligned} \tag{3.4}$$

où I_0 est la fonction modifiée de Bessel de première espèce d'ordre zéro. Le support des paramètres ν et σ est $(0, +\infty)$.

À l'origine, la distribution de Rice correspond à la distribution de l'enveloppe d'un signal périodique d'amplitude ν contaminé par un bruit gaussien additif d'écart-type σ. Avec cette définition, le rapport ν/σ correspond à un rapport signal sur bruit. Pour un haut rapport signal sur bruit, typiquement $\nu/\sigma > 3$, la distribution de Rice se rapproche

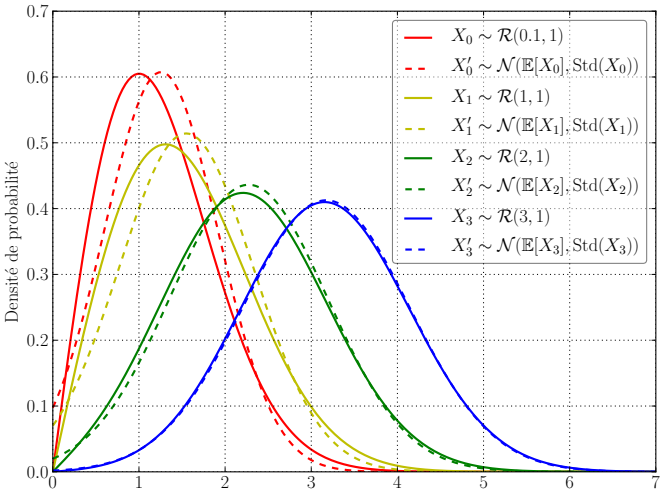

FIGURE 3.1. – Densité de probabilité de la distribution de Rice, pour différents rapports signal sur bruit. La densité de probabilité de la distribution gaussienne de même espérance et écart-type est aussi représentée pour comparaison.

d'une distribution gaussienne. À l'inverse, pour un faible rapport signal sur bruit, elle tend vers une distribution de Rayleigh, qu'elle atteint pour $\nu/\sigma = 0$. La figure 3.1 présente les densités de probabilité de cas plus ou moins bruités, côte à côte avec la distribution gaussienne de même espérance et écart-type. Il est intéressant de noter l'asymétrie de la distribution quand le rapport signal sur bruit est faible, la densité de probabilité étant décalée vers la gauche avec une longue queue à droite.

Si une variable aléatoire suit une distribution de Rice, alors le carré de cette variable aléatoire suit une distribution du Chi-2 non centrée, à deux degrés de liberté et de paramètre de décentrage égale à ν [Kay, 1993]. Ainsi, sous l'hypothèse d'un bruit blanc additif gaussien, le spectrogramme d'un signal bruité S_e calculé avec le carré des coefficients d'ondelettes suit une distribution du Chi-2 non centrée.

La figure 3.2 présente la dégradation du module des coefficients d'ondelettes de Morlet d'un signal sinusoïdal pur, pour différents niveaux de bruit gaussien. Cette expérience a été reproduite 10 000 fois pour obtenir la distribution empirique du module des coefficients d'ondelettes à différents points temps-fréquence $f0$, $f1$ et $f2$. Cette distribution empirique est illustrée par les histogrammes de la figure 3.3 et reflète l'adéquation du module des coefficients d'ondelettes à la distribution de Rice. Pour les zones de haut rapport signal sur bruit, en $f0$ et $f1$, la distribution du module des coefficients semble gaussienne. Au contraire, au point temps-fréquence $f2$, le rapport signal sur bruit est beaucoup plus faible, comme ce point est éloigné de la fréquence centrale du signal. Ceci se traduit dans l'histogramme associé par un décalage à gauche et une asymétrie caractéristique de la distribution de Rice.

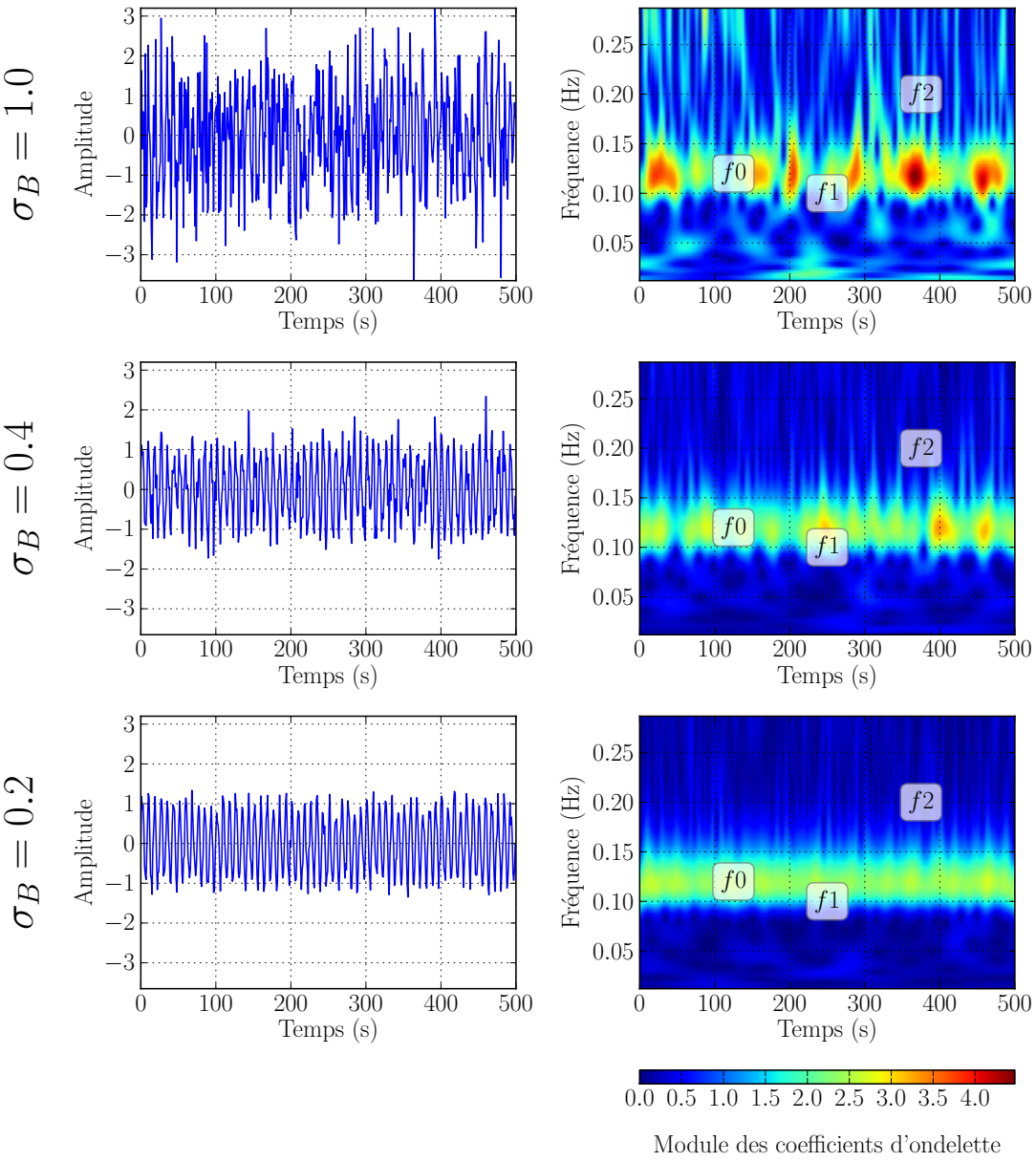

FIGURE 3.2. – Exemples d'observations de signaux aléatoires. Le signal non bruité est une sinusoïde pure à 0,12 Hz, auquel est additionné un bruit blanc gaussien d'écart-type σ_B. La représentation temps-fréquence correspond au module des coefficients d'ondelettes calculés avec des ondelettes de Morlet.

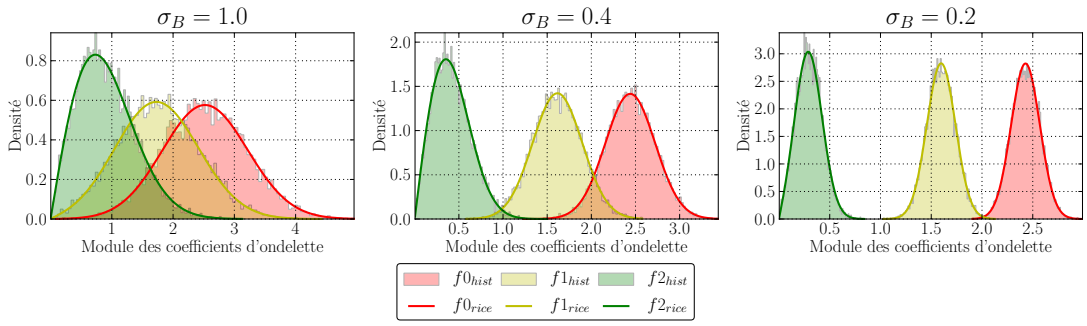

FIGURE 3.3. – Histogrammes du module des coefficients d'ondelettes d'un signal sinusoïdal contaminé par un bruit blanc gaussien additif, pour plusieurs niveaux de bruit σ_B. Plusieurs points du plan temps-fréquence ont été échantillonnés, voir la figure 3.2. Pour chaque histogramme, la densité de probabilité de la distribution de Rice théorique associée est aussi représentée.

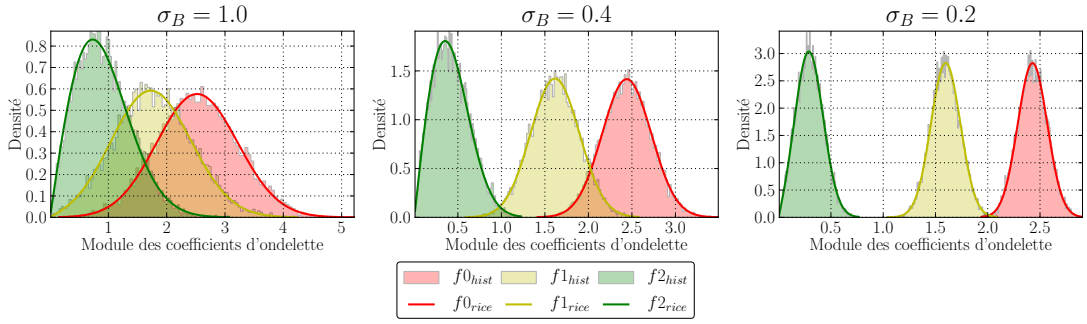

FIGURE 3.4. – Histogrammes du module des coefficients d'ondelettes d'un signal sinusoïdal contaminé par un bruit blanc uniforme additif, pour plusieurs niveaux de bruit σ_B. Plusieurs points du plan temps-fréquence ont été échantillonnés, voir la figure 3.2. Pour chaque histogramme, la densité de probabilité de la distribution de Rice correspondant à un bruit blanc gaussien additif de même écart-type est aussi représentée.

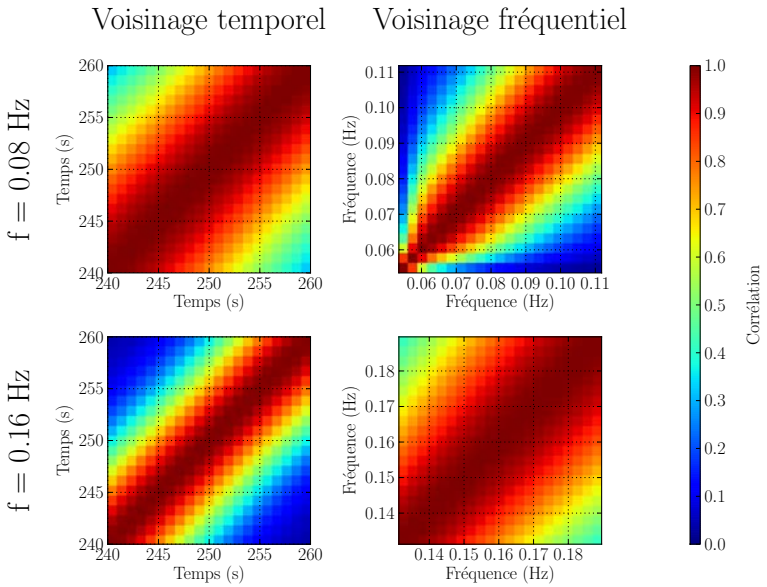

FIGURE 3.5. – Corrélations estimées entre le module des coefficients d'ondelettes de Morlet à des points temps-fréquence proches. Les corrélations sont estimées sur la transformée en ondelettes de Morlet d'observations de signaux aléatoires. Ces signaux sont composés d'une sinusoïde pure à 0,16 Hz ou à 0,08 Hz additionnée d'un bruit blanc gaussien d'écart-type $\sigma_B = 0{,}5$. La transformée en ondelettes est calculée avec un échantillonnage linéaire des échelles des ondelettes filles.

La même expérience a été reproduite en changeant la distribution du bruit par une distribution uniforme, dont l'intervalle est fixé dans le but d'obtenir un bruit de même écart-type que dans le cas gaussien précédent. De manière étonnante, les histogrammes du module des coefficients d'ondelettes présentent les mêmes caractéristiques que dans le cas gaussien : une forme gaussienne pour les points temps-fréquence correspondant à un rapport signal sur bruit élevé et une asymétrie à gauche marquée pour les faibles rapports signal sur bruit. L'adéquation à une distribution de Rice semble possible.

3.1.2. Corrélation des coefficients adjacents

Les ondelettes filles utilisées pour calculer la transformée en ondelette continue présente une grande corrélation entre elles lorsqu'elles sont centrées en des points temps-fréquence proches [Mallat, 2008]. De ce fait, les coefficients d'ondelettes calculés en des points temps-fréquences proches sont eux aussi fortement corrélés.

Dans le cas d'un signal aléatoire, cette corrélation déterministe se traduit par une corrélation statistique du module des coefficients d'ondelettes proches en temps et en fréquence. La figure 3.5 illustre ces corrélations dans le cas d'un signal sinusoïdal ayant un bruit blanc additif gaussien. Il est intéressant de noter que les corrélations entre coefficients

d'ondelettes distants en temps diminuent plus vite à haute fréquence qu'à basse fréquence. Au contraire, la corrélation entre coefficients d'ondelettes distants en fréquence diminue plus vite à basse fréquence qu'à haute fréquence. Les mêmes phénomènes sont observés avec un bruit plus faible ou plus fort, et même avec un bruit blanc additif uniforme au lieu du bruit gaussien.

Ces phénomènes s'expliquent par la multirésolution de la transformée en ondelette. En effet, la haute résolution fréquentielle des ondelettes filles à basses fréquences diminue la corrélation fréquentielle des coefficients d'ondelettes associés, et la basse résolution temporelle de ces ondelettes filles augmente la corrélation temporelle. Le phénomène inverse se produit avec les ondelettes filles à hautes fréquences, ayant une résolution en temps élevée et une résolution en fréquence faible.

3.2. Modèles de mélange

Un modèle de mélange est un modèle statistique, où les données associées au modèle sont considérées comme issues de tirages aléatoires d'une distribution de mélange [Bishop, 2006, Verbeek, 2004].

Les modèles de mélange sont utilisés dans un grand nombre de domaines de l'apprentissage automatique, tels que le partitionnement avec des modèles de mélange gaussien [Verbeek, 2004, Corduneanu and Bishop, 2001], la réduction de dimension avec des modèles de mélange d'analyseurs en composantes principales [Tipping and Bishop, 1998] ou d'analyseurs factoriels [Beal, 2003, Ghahramani and Beal, 2000], ou encore la régression avec des modèles de mélange d'experts [Ueda and Ghahramani, 2002].

Les modèles de mélange employés ici servent exclusivement pour des tâches de partitionnement. Les données à partitionner sont des valeurs d'énergie de signaux cérébraux dans le plan temps-fréquence, entre différentes électrodes et différents points temps-fréquence.

3.2.1. Construction d'une distribution de mélange

Une distribution de mélange est définie comme la somme pondérée de plusieurs distributions, nommées *composantes* du modèle de mélange.

Soit une distribution de mélange à K composantes, θ_k désigne le vecteur de paramètres d'une composante et π_k le coefficient de pondération associé. Les coefficients π_k, aussi appelés *proportions de mélange*, sont définis tel que :

$$\pi_k \geq 0 \ \forall k \in [1, K] \quad \text{et} \quad \sum_{k=1}^{K} \pi_k = 1, \tag{3.5}$$

et sont regroupés dans un vecteur $\pi = (\pi_1, \ldots, \pi_K)$.

Soit X une variable aléatoire suivant une distribution de mélange à K composantes, sa densité de probabilité s'écrit de la manière suivante :

$$p(x|\pi, \boldsymbol{\theta}) = \sum_{k=1}^{K} \pi_k \, p(x|\theta_k), \tag{3.6}$$

où $\boldsymbol{\theta} = \{\theta_k\}_{k=1}^K$ est l'ensemble des paramètres des composantes.

Il est possible de reformuler cette distribution en introduisant une variable cachée Z représentant la composante de la distribution de mélange utilisée pour générer un tirage de la distribution de mélange [Bishop, 2006]. Z est un vecteur aléatoire à K valeurs, $Z = (Z_1, \ldots, Z_K)$. Pour un tirage aléatoire z issu de Z, un seul z_k vaut un, tous les autres valant zéro. Z est décrit par une distribution catégorique :

$$p(z|\pi) = p(z_k = 1|\pi) = \Pi_{k=1}^K \pi_k^{z_k}. \tag{3.7}$$

La densité conditionnelle de X sachant Z s'écrit :

$$p(x|z_k = 1, \boldsymbol{\theta}) = p(x|\theta_k) \;\Leftrightarrow\; p(x|z, \boldsymbol{\theta}) = \Pi_{k=1}^K p(x|\theta_k)^{z_k}, \tag{3.8}$$

et la densité jointe de X et Z

$$p(x, z|\pi, \boldsymbol{\theta}) = p(x|z, \boldsymbol{\theta})\, p(z|\pi) = \Pi_{k=1}^K p(x|\theta_k)^{z_k}\, \pi_k^{z_k}. \tag{3.9}$$

La densité marginale de X, présentée dans l'équation (3.6), correspond alors à la somme sur tous les états possibles de Z de la densité jointe de X et Z :

$$p(x|\pi, \boldsymbol{\theta}) = \sum_z \left(\Pi_{k=1}^K p(x|\theta_k)^{z_k}\, \pi_k^{z_k}\right) = \sum_{k=1}^K \pi_k\, p(x|\theta_k). \tag{3.10}$$

Ainsi, la génération d'un tirage aléatoire à partir d'une distribution de mélange peut s'interpréter comme suit :

1. le tirage aléatoire d'une valeur de Z, représentant le choix aléatoire d'une des K composantes, avec une probabilité π_k de choisir chaque composante,
2. puis la génération aléatoire d'un échantillon à partir de la composante sélectionnée.

3.2.2. Inférence bayésienne avec l'algorithme VMP pour un modèle de mélange

Pour modéliser un échantillon de N observations avec une distribution de mélange, il faut considérer un modèle comprenant un ensemble $\boldsymbol{X} = \{X_n\}_{n=1}^N$ de N variables aléatoires indépendantes et identiquement distribuées, associées à chaque observation. Chaque X_n suit la même distribution de mélange de paramètres π et $\boldsymbol{\theta}$. À chaque variable X_n est associée une variable cachée Z_n :

$$p(x_n|z_n, \boldsymbol{\theta}) = \Pi_{k=1}^K p(x_n|\theta_k)^{z_{n,k}} \quad \text{et} \quad p(z_n|\pi) = \Pi_{k=1}^K \pi_k^{z_{n,k}}. \tag{3.11}$$

Pour pouvoir pratiquer l'inférence bayésienne sur ce modèle statistique, il faut préciser les distributions a priori des paramètres du modèle, c'est-à-dire π et $\boldsymbol{\theta}$. Soient α et ϵ les paramètres respectivement des distributions a priori de π et $\boldsymbol{\theta}$.

La probabilité jointe d'un modèle de mélange associé à un échantillon de N observations s'écrit alors :

$$\begin{aligned} p(\boldsymbol{x}, \boldsymbol{z}, \pi, \boldsymbol{\theta}|\alpha, \epsilon) &= p(\boldsymbol{x}|\boldsymbol{z}, \pi, \boldsymbol{\theta})\, p(\boldsymbol{z}|\pi)\, p(\pi|\alpha)\, p(\boldsymbol{\theta}|\epsilon) \\ &= \Pi_{n=1}^N p(x_n|z_n, \boldsymbol{\theta})\, p(z_n|\pi)\, p(\pi|\alpha)\, \Pi_{k=1}^K p(\theta_k|\epsilon), \end{aligned} \tag{3.12}$$

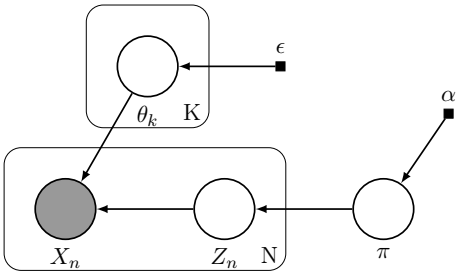

FIGURE 3.6. – Modèle graphique probabiliste représentant la distribution jointe d'un modèle de mélange décrite par l'équation (3.12)

où $\boldsymbol{x} = \{x_n\}_{n=1}^N$ et $\boldsymbol{z} = \{z_n\}_{n=1}^N$. Les vecteurs de paramètres θ_k des composantes sont ici considérés comme indépendants les autres des autres, d'où la factorisation des $p(\theta_k|\epsilon)$. La figure 3.6 présente le modèle graphique associé à cette distribution jointe.

La probabilité jointe a posteriori de π, $\boldsymbol{\theta}$ et des Z_n peut être approximée de nombreuses manières. Seuls les éléments relatifs à l'algorithme VMP, introduit à la section 2.2.3, seront ici présentés.

L'algorithme VMP nécessite que le modèle soit de type exponentiel-conjugué. Dans le cas d'un modèle de mélange, cette contrainte déterminera les familles exponentielles des distributions a priori de π et $\boldsymbol{\theta}$.

Les variables Z_n suivent une distribution catégorique, laquelle est une famille exponentielle. Elles contraignent par conjugaison la distribution a priori de π : il s'agit d'une distribution de Dirichlet [Bishop, 2006]. La distribution de Dirichlet est aussi une famille exponentielle :

$$p(\pi|\alpha) = \mathcal{D}ir(\pi|\alpha) = C(\alpha)\,\Pi_{k=1}^K \pi_k^{\alpha_k - 1} \quad \text{où} \quad C(\alpha) = \frac{\Gamma(\sum_{k=1}^K \alpha_k)}{\Pi_{k=1}^K \Gamma(\alpha_k)}, \tag{3.13}$$

avec le vecteur de paramètres $\alpha = (\alpha_1, \ldots, \alpha_K)$.

Concernant les variables X_n, elles suivent une distribution de mélange, laquelle n'est pas en soi une famille exponentielle. Mais la distribution conditionnelle de X_n sachant Z_n forme une famille exponentielle si les composantes font aussi partie de familles exponentielles [Winn, 2003]. Les contraintes de conjugaison sur les distributions des paramètres θ_k sont imposées par le type de distribution des composantes.

La probabilité jointe approximée par l'algorithme VMP a la forme factorisée suivante :

$$p(\boldsymbol{z}, \pi, \boldsymbol{\theta}|\boldsymbol{x}, \alpha, \epsilon) \approx \Pi_{n=1}^N q(z_n)\, q(\pi)\, \Pi_{k=1}^K q(\theta_k). \tag{3.14}$$

Par construction, les distributions a posteriori approximées optimales sont du même type que les distributions a priori, d'où les densités de probabilité a posteriori optimales :

$$q^*(z_n) = \Pi_{k=1}^K (\pi_{n,k}^*)^{z_{n,k}} \qquad q^*(\pi) = \mathcal{D}ir(\pi|\alpha^*) \qquad q^*(\theta_k) = p(\theta_k|\epsilon_k^*), \tag{3.15}$$

avec $\alpha^* = (\alpha_1^*, \ldots, \alpha_K^*)$.

Les paramètres $\pi^*_{n,k}$, α^* et ϵ^*_k des distributions a posteriori sont calculés par passage de messages entre les nœuds du graphe probabiliste du modèle, illustré à la figure 3.6. Les messages utilisés par l'algorithme VMP pour les nœuds représentant les variables Z_n et π sont détaillés en annexes A.4 et A.5. Si toutes les composantes de la distribution de mélange sont du même type, les messages utilisés pour les nœuds représentant les variables X_n sont dérivables automatiquement, voir l'annexe A.6 pour plus de détails.

Par ailleurs, l'algorithme VMP permet aussi de calculer une borne inférieure $\mathcal{L}(q)$ sur le logarithme de la probabilité marginale $\ln p(\boldsymbol{x}|\alpha, \epsilon)$ du modèle.

3.2.3. Utilisation pour partitionner des données

Pour utiliser un modèle de mélange comme technique de partitionnement de données il suffit d'identifier chaque composante de la distribution de mélange comme une partition de l'échantillon observé.

Les caractéristiques des partitions sont données par les distributions a posteriori des vecteurs de paramètres θ_k. Il est possible d'obtenir une estimation ponctuelle de ces paramètres en utilisant les estimateurs définis dans le cadre de la théorie bayésienne de la décision, comme décrit à la section 2.2.2. Pour une fonction de coût quadratique, les estimateurs ponctuels optimaux θ^*_k correspondent à l'espérance des distributions a posteriori des paramètres θ_k.

Le degré d'appartenance de chaque observation de l'échantillon à chaque partition est reflété par la distribution a posteriori des variables cachées Z_n, via la probabilité a posteriori $p(z_{n,k} = 1|\boldsymbol{x})$. Il s'agit d'un partitionnement doux. Un partitionnement dur est obtenu en utilisant un estimateur du maximum a posteriori, ou estimateur MAP, sur cette distribution [1]. L'estimation z^*_n de la variable cachée Z_n associée à la n-ième observation est donc calculée comme suit :

$$z^*_n = \underset{z_n}{\operatorname{argmax}}\ p(z_n|\boldsymbol{x}). \tag{3.16}$$

Ainsi chaque observation est associée à une et une seule partition.

Dans le cas d'une utilisation de l'algorithme VMP pour approximer les distributions a posteriori, les estimateurs θ^*_k et z^*_n sont déterminés avec les distributions a posteriori approximées des variables θ_k et Z_n :

$$\theta^*_k \approx \mathbb{E}_{q(\theta_k)}[\theta_k] \quad \text{et} \quad z^*_n \approx \underset{z_n}{\operatorname{argmax}}\ q(z_n). \tag{3.17}$$

Sélection de modèle et choix du nombre de partitions

La question du choix du nombre de partitions est un des problèmes fondamentaux des techniques de partitionnement de données. L'utilisation de modèles bayésiens apporte une réponse efficace à ce problème.

1. Pour rappel, l'estimateur MAP minimise le coût moyen a posteriori de l'erreur d'estimation, pour une fonction de coût 0-1.

Pour déterminer le nombre de partitions adéquat, il suffit de comparer les probabilités marginales $p(\boldsymbol{x}|\alpha, \epsilon)$ de différents modèles de mélange, avec différents nombres de composantes. Le modèle le plus apte à représenter les observations de l'échantillon étudié présentera la plus grande probabilité marginale. L'équilibre entre le nombre de composantes et l'ajustement aux données est réalisé grâce aux probabilités a priori du modèle de mélange, qui pénalisent l'utilisation d'un trop grand nombre de composantes et le sur-ajustement du modèle aux données. L'utilisation de l'algorithme VMP permet d'approcher la valeur du logarithme de la probabilité marginale $\ln p(\boldsymbol{x}|\alpha, \epsilon)$ par une borne inférieure $\mathcal{L}(q)$ et d'utiliser cette dernière pour sélectionner le meilleur modèle.

En comparaison, les modèles de mélange ajustés par des techniques de maximum de vraisemblance ne peuvent pas être sélectionnés sur la seule base de leur vraisemblance, cette valeur croissant avec le nombre de composantes du modèle. Ces modèles emploient alors des critères de pénalisation de leur log-vraisemblance pour déterminer un nombre raisonnable de composantes [Robert, 2006, Saporta, 2011]. Les plus connus sont le *critère d'information d'Akaike* (Akaike information criterion ou AIC [Akaike, 1973]) et le BIC, précédemment évoqué à la section 2.2.3. Néanmoins, l'utilisation de ces critères est moins performante que l'approximation de la probabilité marginale par la borne inférieure [Beal, 2003].

Suppression automatique de composantes

Une autre approche pour choisir automatiquement le nombre de composantes consiste en la suppression automatique des composantes inutiles d'un modèle. Cette suppression automatique est mise en place via le choix de la probabilité a priori de π.

Dans le cas d'un modèle exponentiel-conjugué tel que décrit précédemment, la distribution de Dirichlet utilisée comme a priori pour π permet, pour certaines valeurs de son paramètre α, de « désactiver » les composantes inutiles. Pour $\alpha_k \to 0 \ \forall k$, la distribution a priori de π favorise des proportions de mélange π_k proches de zéro, rendant ainsi le modèle de mélange parcimonieux [Bishop, 2006]. En pratique, j'utilise la valeur arbitraire $\alpha_k = 10^{-2} \ \forall k$ pour mes modèles de mélange.

Il suffit alors d'utiliser des modèles de mélange avec un grand nombre de composantes, d'approximer la distribution a posteriori de ces modèles avec l'algorithme VMP et d'inspecter cette distribution a posteriori pour détecter le nombre effectif de composantes utilisées.

Ainsi, pour chaque composante, on vérifie si la partition correspondante, déterminée par partitionnement dur, est vide. Soient $q(z_n)$ la densité a posteriori approximée de Z_n et z_n^* l'estimation de Z_n associée à la n-ième observation x_n d'un échantillon de N éléments, alors le sous-ensemble des données correspondant à la k-ième partition s'écrit :

$$\mathbf{x}_k = \{\mathrm{x}_n\}_{z_{n,k}^*=1}, \tag{3.18}$$

où $n \in [1, N]$. Si ce sous-ensemble $\mathbf{x}_k$ est vide, la composante est considérée comme inutilisée.

3.3. Modèle de mélange gaussien

Le modèle de mélange gaussien est un modèle de mélange dont les composantes sont toutes des distributions gaussiennes.

En faisant l'hypothèse d'un bruit blanc gaussien additif, le modèle de mélange gaussien peut être adapté au partitionnement du module des coefficients d'ondelettes de signaux bruités si leur distribution de Rice présente un haut rapport signal sur bruit. Dans un cas plus général, si la distribution du bruit des signaux est inconnu, le modèle de mélange gaussien peut tout de même être utilisé en première approximation.

3.3.1. Définition du modèle et inférence bayésienne par l'algorithme VMP

Soit un modèle de mélange gaussien associé à un échantillon de N observations, la densité de probabilité des variables X_n associées à chaque observation vaut :

$$p(x_n|z_n, \boldsymbol{\mu}, \boldsymbol{\lambda}) = \Pi_{k=1}^{K} \mathcal{N}(x_n|\mu_k, \lambda_k)^{z_{n,k}}, \tag{3.19}$$

avec $\boldsymbol{\mu} = \{\mu_k\}_{k=1}^{K}$ et $\boldsymbol{\lambda} = \{\lambda_k\}_{k=1}^{K}$. Le paramètre μ_k est la moyenne de la distribution gaussienne et le paramètre λ_k est sa *précision*, c'est-à-dire l'inverse de sa variance.

Pour obtenir un modèle exponentiel-conjugué, il faut choisir une distribution gaussienne comme distribution a priori des μ_k et une distribution gamma comme distribution a priori des λ_k des composantes [Bishop, 2006] :

$$p(\mu_k|m_\mu, l_\mu) = \mathcal{N}(\mu_k|m_\mu, l_\mu) \quad \text{et} \quad p(\lambda_k|a_\lambda, b_\lambda) = \mathcal{G}(\lambda_k|a_\lambda, b_\lambda). \tag{3.20}$$

Comme expliqué précédemment à la section 3.2.2, la distribution a priori des variables Z_n est une distribution catégorique de paramètre π, ce dernier paramètre ayant lui-même pour distribution a priori une distribution de Dirichlet :

$$p(z_n|\pi) = \Pi_{k=1}^{K} \pi_k^{z_{n,k}} \quad \text{et} \quad p(\pi|\alpha) = \mathcal{D}ir(\pi|\alpha). \tag{3.21}$$

La probabilité jointe du modèle de mélange gaussien s'écrit alors :

$$\begin{aligned} &p(\boldsymbol{x}, \boldsymbol{z}, \pi, \boldsymbol{\mu}, \boldsymbol{\lambda}|\alpha, m_\mu, l_\mu, a_\lambda, b_\lambda) \\ &= \Pi_{n=1}^{N} p(x_n|z_n, \mu, \lambda)\, p(z_n|\pi)\, p(\pi|\alpha)\, \Pi_{k=1}^{K} p(\mu_k|m_\mu, l_\mu)\, p(\lambda_k|a_\lambda, b_\lambda). \end{aligned} \tag{3.22}$$

où $\boldsymbol{x} = \{x_n\}_{n=1}^{N}$ et $\boldsymbol{z} = \{z_n\}_{n=1}^{N}$. La figure 3.7 représente la distribution jointe du modèle sous la forme d'un modèle graphique probabiliste.

L'algorithme VMP permet alors d'approximer la probabilité jointe a posteriori comme suit :

$$p(\boldsymbol{z}, \pi, \boldsymbol{\mu}, \boldsymbol{\lambda}|\boldsymbol{x}, \alpha, m_\mu, l_\mu, a_\lambda, b_\lambda) \approx \Pi_{n=1}^{N} q(z_n)\, q(\pi)\, \Pi_{k=1}^{K} q(\mu_k)\, q(\lambda_k). \tag{3.23}$$

Les densités des distributions a posteriori optimales valent alors :

$$\begin{aligned} q^*(z_n) &= \Pi_{k=1}^{K} (\pi_{n,k}^*)^{z_{n,k}} & q^*(\mu_k) &= \mathcal{N}(\mu_k|m_{\mu_k}^*, l_{\mu_k}^*) \\ q^*(\pi) &= \mathcal{D}ir(\pi|\alpha^*) & q^*(\lambda_k) &= \mathcal{G}(\lambda|a_{\lambda_k}^*, b_{\lambda_k}^*), \end{aligned} \tag{3.24}$$

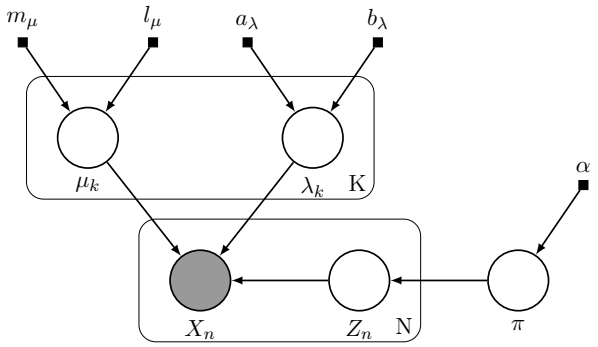

FIGURE 3.7. – Modèle graphique probabiliste représentant la distribution jointe d'un modèle de mélange gaussien décrite par l'équation (3.22)

avec $\alpha^* = (\alpha_1^*, \ldots, \alpha_K^*)$.

Le graphe probabiliste, illustré par la figure 3.7, est utilisé pour calculer par passage de messages entre les nœuds, avec l'algorithme VMP, la valeur des paramètres $\pi_{n,k}^*$, α^*, $m_{\mu_k}^*$, $l_{\mu_k}^*$, $a_{\lambda_k}^*$ et $b_{\lambda_k}^*$. Les messages sont décrits en annexe A.

3.3.2. Application simple du modèle bayésien de mélange gaussien

Dans l'exemple suivant, j'ai généré deux cents points à partir d'une distribution de mélange gaussien à deux composantes, ayant pour moyennes $\mu_1 = 0$ et $\mu_2 = 10$, pour précisions $\lambda_1 = 1$ et $\lambda_2 = 0{,}25$ et pour proportions $\pi_1 = 0{,}4$ et $\pi_2 = 0{,}6$. J'ai ensuite approximé avec l'algorithme VMP la distribution a posteriori des paramètres d'un modèle bayésien de mélange gaussien appliqué à ces données. Pour ces calculs, le nombre de composantes du modèle de mélange a été fixé à $K = 5$. Les autres détails pratiques de mise en œuvre de l'algorithme VMP (initialisation, convergence, etc.) pour ce modèle sont donnés en annexe B.3.

En utilisant le mécanisme de suppression automatique des composantes inutiles, décrit en section 3.2.3, la distribution a posteriori obtenue ne comporte que deux composantes. Ainsi, les trois composantes gaussiennes surnuméraires ont été éliminées.

L'approximation des distributions a posteriori des paramètres et une estimation de la densité la distribution de mélange gaussien utilisée pour générer les observations sont illustrés dans la figure 3.8. Les *intervalles de crédibilité* à 95 % représentés correspondent aux intervalles interquantiles des probabilités a posteriori des paramètres dont la probabilité vaut 95 %. Ils illustrent les régions où la croyance a posteriori sur les valeurs des paramètres du modèle est la plus grande.

3.4. Modèle de mélange ricien

Le modèle de mélange ricien est un modèle de mélange dont les composantes sont des distributions de Rice. Pour le partitionnement du module des coefficients d'ondelettes de

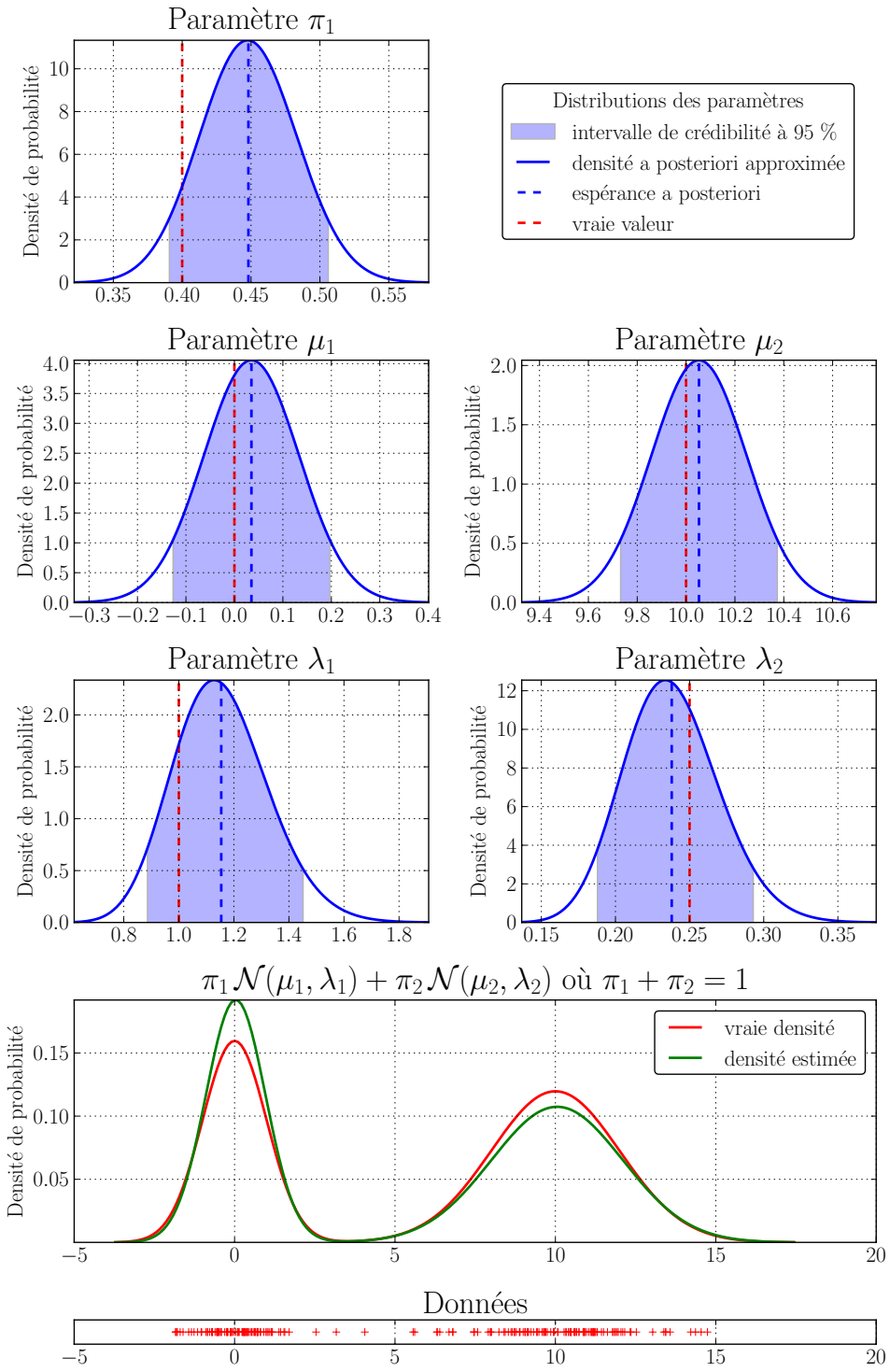

FIGURE 3.8. – Distributions a posteriori approximées des paramètres d'un modèle de mélange gaussien appliqué à un échantillon de 200 observations, issues d'une distribution de mélange gaussien. La densité de cette distribution de mélange est estimée à partir des espérances a posteriori des paramètres.

signaux bruités, ce modèle est théoriquement le plus adapté si le bruit est blanc, gaussien et additif.

La distribution de Rice n'étant pas une famille exponentielle, le choix des distributions a priori de ses paramètres et le calcul de leur distribution a posteriori n'est pas trivial.

Dans un premier temps, je vais montrer qu'il est possible par une *approximation variationnelle locale* d'utiliser la distribution de Rice dans un modèle exponentiel-conjugué. Dans un second temps, je détaillerai les mêmes mécanismes mais pour la distribution de mélange ricien. Les éléments présentés dans ces deux sections sont des apports originaux de ma thèse.

3.4.1. Inférence bayésienne par l'algorithme VMP pour la distribution de Rice

La technique d'approximation variationnelle locale dont il est question ici a été mise en œuvre par Bishop et Svensen, pour des modèles bayésiens de mélanges hiérarchiques d'experts [Bishop and Svensen, 2003]. Cette technique d'approximation variationnelle locale a été adaptée pour l'algorithme VMP par Winn [Winn and Bishop, 2005]. Dans les deux cas, il s'agissait d'approximer des distributions conditionnelles mettant en jeu une fonction sigmoïde.

La distribution de Rice a été présentée en section 3.1.1. Soit X une variable aléatoire suivant une distribution de Rice, sa densité de probabilité vaut :

$$p(x|\nu,\sigma) = \mathcal{R}(x|\nu,\sigma) = \exp\left(\ln I_0\left(\frac{x\nu}{\sigma^2}\right) + \ln x - 2\ln\sigma - \frac{x^2+\nu^2}{2\sigma^2}\right). \tag{3.25}$$

Pour faciliter par la suite la recherche de distributions a priori, la distribution de Rice est re-paramétrisée avec un paramètre de précision λ à la place du paramètre de dispersion σ, tel que $\lambda = 1/\sigma^2$:

$$\begin{aligned} p(x|\nu,\lambda) &= \mathcal{R}(x|\nu,\lambda) \\ &= \exp\left(\ln I_0(x\nu\lambda) + \ln x + \ln\lambda - \frac{1}{2}\lambda(x^2+\nu^2)\right). \end{aligned} \tag{3.26}$$

Afin d'utiliser cette distribution dans un modèle exponentiel-conjugué, sa densité de probabilité est approximée par une borne inférieure variationnelle $\mathcal{F}$:

$$p(x|\nu,\lambda) \geq \mathcal{F}(x,\nu,\lambda,\xi) \tag{3.27}$$

où ξ est le paramètre variationnel de la borne. La borne recherchée doit être définie de telle sorte qu'elle puisse s'exprimer à la manière d'une famille exponentielle :

$$\ln\mathcal{F}(x,\nu,\lambda,\xi) = \eta_x(\nu,\lambda,\xi)^\top \cdot \mathrm{u}_x(x) + f_x(x) + g_x(\nu,\lambda,\xi). \tag{3.28}$$

Il faut noter que dans ce cas $g_x(\nu,\lambda,\xi)$ n'est pas une constante de normalisation, la borne elle-même n'étant pas une densité de probabilité.

La borne $\mathcal{F}$ peut s'obtenir à partir de l'équation (3.25) en bornant $\ln I_0$. La fonction $\ln I_0$ étant convexe [Neuman, 1992], une borne inférieure simple de $\ln I_0$ est donnée par

son développement de Taylor du premier ordre : il s'agit de la tangente à la courbe de $\ln I_0$ en chaque point [Bishop, 2006] :

$$\ln I_0(x\nu\lambda) \geq \ln I_0(\xi) + \frac{d \ln I_0(\xi)}{d\,\xi}(x\nu\lambda - \xi). \tag{3.29}$$

La dérivée de I_0 étant I_1 [Abramowitz and Stegun, 1964], la borne sur $\ln I_0$ est :

$$\ln I_0(x\nu\lambda) \geq \ln I_0(\xi) + \frac{I_1(\xi)}{I_0(\xi)}(x\nu\lambda - \xi). \tag{3.30}$$

La borne $\mathcal{F}$ s'obtient alors en introduisant cette borne de $\ln I_0$ dans l'équation (3.25) :

$$\begin{aligned}\ln \mathcal{F}(x,\nu,\lambda,\xi) &= \ln I_0(\xi) + \frac{I_1(\xi)}{I_0(\xi)}(x\nu\lambda - \xi) + \ln x - \ln\lambda - \frac{1}{2}\lambda(x^2+\nu^2) \\ &= \underbrace{\begin{pmatrix} \frac{I_1(\xi)}{I_0(\xi)}\nu\lambda \\ -\frac{1}{2}\lambda \end{pmatrix}^\top}_{\eta_x(\nu,\lambda,\xi)} \cdot \underbrace{\begin{pmatrix} x \\ x^2 \end{pmatrix}}_{\mathrm{u}_x(x)} + \underbrace{\ln x}_{f_x(x)} \underbrace{- \frac{1}{2}\lambda\nu^2 + \ln\lambda - \frac{I_1(\xi)}{I_0(\xi)}\xi + \ln I_0(\xi)}_{g_x(\nu,\lambda,\xi)},\end{aligned} \tag{3.31}$$

où les différentes parties de l'équation (3.28) sont identifiables. Par construction, comme la fonction exponentielle est strictement croissante, l'inégalité (3.27) est vérifiée.

Pour utiliser la distribution de Rice dans un modèle, il suffit alors d'employer la borne $\mathcal{F}$ au lieu de sa densité de probabilité lors de la définition de la probabilité jointe du modèle. Les contraintes sur les distributions a priori conjuguées pour les paramètres ν et λ se déduisent de la forme de $\mathcal{F}$. De même, les messages utilisés par l'algorithme VMP entre le nœud du graphe probabiliste représentant X et les nœuds du graphe représentant les paramètres ν et λ sont obtenus à partir de la forme de $\mathcal{F}$.

Ainsi, pour le paramètre ν, une réécriture adéquate de $\ln \mathcal{F}$ met en valeur le type de distribution a priori et le message pour l'algorithme VMP :

$$\ln \mathcal{F}(x,\nu,\lambda,\xi) = \underbrace{\begin{pmatrix} \frac{I_1(\xi)}{I_0(\xi)}\lambda x \\ -\frac{1}{2}\lambda \end{pmatrix}^\top}_{\ell_{x,\nu}(x,\lambda,\xi)} \cdot \underbrace{\begin{pmatrix} \nu \\ \nu^2 \end{pmatrix}}_{\mathrm{u}_\nu(\nu)} + h_{x,\nu}(x,\lambda,\xi), \tag{3.32}$$

où $h_{x,\nu}$ rassemble tous les termes ne dépendants pas de ν.

La distribution a priori de ν, faisant partie d'une famille exponentielle, doit donc avoir comme vecteur de statistiques naturelles $\mathrm{u}_\nu(\nu) = (\nu, \nu^2)$. Un exemple typique, respectant aussi la contrainte de support $\nu \in (0, +\infty)$, est la distribution gaussienne rectifiée. Il s'agit d'une distribution gaussienne tronquée dont la densité est nulle sur l'intervalle $(-\infty, 0)$ [Miskin, 2000].

Le message vers le nœud du graphe probabiliste représentant ν est donné par l'espérance a posteriori de $\ell_{x,\nu}(x,\lambda,\xi)$. Dans le cas où X est une variable observée, ce message s'écrit :

$$m_{X\to\nu} = \mathbb{E}_{q(\lambda)}[\ell_{x,\nu}(x,\lambda,\xi)] = \begin{pmatrix} \frac{I_1(\xi)}{I_0(\xi)}\mathbb{E}_{q(\lambda)}[\lambda]x \\ -\frac{1}{2}\mathbb{E}_{q(\lambda)}[\lambda] \end{pmatrix}. \tag{3.33}$$

Le même procédé est employé pour le paramètre λ :

$$\ln \mathcal{F}(x,\nu,\lambda,\xi) = \underbrace{\begin{pmatrix} \frac{I_1(\xi)}{I_0(\xi)}\nu x - \frac{1}{2}(x^2+\nu^2) \\ 1 \end{pmatrix}^\top}_{\ell_{x,\lambda}(x,\nu,\xi)} \cdot \underbrace{\begin{pmatrix} \lambda \\ \ln\lambda \end{pmatrix}}_{\mathrm{u}_\lambda(\lambda)} + h_{x,\lambda}(x,\nu,\xi), \tag{3.34}$$

où $h_{x,\lambda}$ rassemble tous les termes ne dépendants pas de λ. Le vecteur des statistiques naturelles de la distribution a priori de λ doit être $\mathrm{u}_\lambda(\lambda) = (\lambda, \ln\lambda)$, ce qui correspond par exemple à une distribution gamma. Si X est une variable observée, le message vers le nœud représentant λ est :

$$m_{X\to\lambda} = \mathbb{E}_{q(\nu)}[\tilde{\phi}_{x,\lambda}(x,\nu,\xi)] = \begin{pmatrix} \frac{I_1(\xi)}{I_0(\xi)}\mathbb{E}_{q(\nu)}[\nu]x - \frac{1}{2}(x^2+\mathbb{E}_{q(\nu)}[\nu^2]) \\ 1 \end{pmatrix}. \tag{3.35}$$

Pour pleinement utiliser la distribution de Rice avec l'algorithme VMP, il reste à déterminer la contribution de la variable aléatoire X à la borne inférieure $\mathcal{L}(q)$ du modèle et à estimer la valeur de ξ. Si X est une variable observée, la contribution de $\mathcal{F}$ à $\mathcal{L}(q)$ vaut :

$$\mathcal{L}_\mathcal{F} = \mathbb{E}_{q(\nu),q(\lambda)}[\eta_x(\nu,\lambda,\xi)]^\top \cdot \mathrm{u}_x(x) + f_x(x) + \mathbb{E}_{q(\nu),q(\lambda)}[g_x(\nu,\lambda,\xi)]. \tag{3.36}$$

La valeur de ξ est estimée de telle sorte qu'elle maximise la borne inférieure $\mathcal{L}(q)$. Le seul élément de $\mathcal{L}(q)$ dépendant de ξ étant $\mathcal{L}_\mathcal{F}$, la valeur de ξ doit maximiser $\mathcal{L}_\mathcal{F}$. Les dérivées première et seconde de $\mathcal{L}_\mathcal{F}$ par rapport à ξ valent :

$$\frac{d\,\mathcal{L}_\mathcal{F}}{d\,\xi} = \frac{d^2\,\ln I_0(\xi)}{d\,\xi^2}\left(\mathbb{E}_{q(\nu)}[\nu]\mathbb{E}_{q(\lambda)}[\lambda]\,x - \xi\right) \quad \text{et} \quad \frac{d^2\,\mathcal{L}_\mathcal{F}}{d\,\xi^2} = -\frac{d^2\,\ln I_0(\xi)}{d\,\xi^2}. \tag{3.37}$$

La dérivée première s'annule en $\xi = \xi^*$ tel que

$$\xi^* = \mathbb{E}_{q(\nu)}[\nu]\mathbb{E}_{q(\lambda)}[\lambda]\,x, \tag{3.38}$$

et il s'agit d'un maximum. En effet, la fonction $\ln I_0$ étant convexe, alors $\mathcal{L}_\mathcal{F}$ est concave. Dès lors l'équation (3.38) est utilisée dans l'algorithme VMP pour mettre à jour la valeur de ξ juste avant d'envoyer les messages $m_{X\to\nu}$ et $m_{X\to\lambda}$ aux nœuds correspondants.

Enfin, si X est une variable aléatoire cachée, il est aussi possible de déduire de la forme de $\ln\mathcal{F}$ la densité de probabilité de la distribution a posteriori de X. Mes modèles n'utilisant la distribution de Rice que pour des variables observées, ce point ne sera pas développé ici.

Dans la partie qui suit, je mettrai en application l'approximation variationnelle locale de la densité d'une distribution de Rice au sein d'un modèle bayésien ricien minimaliste.

3.4.2. Exemple de modèle ricien et application

Soit un modèle bayésien ricien associé à un échantillon de N observations, la densité de probabilité des variables X_n associées à chaque observation s'écrit :

$$\begin{aligned} p(x_n|\nu,\lambda) &= \mathcal{R}ice(x_n|\nu,\lambda) \\ &\geq \mathcal{F}(x_n,\nu,\lambda,\xi_n). \end{aligned} \tag{3.39}$$

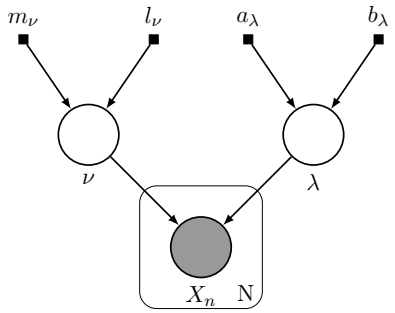

FIGURE 3.9. – Modèle graphique probabiliste représentant la distribution jointe d'un modèle ricien décrite par l'équation (3.41)

Pour former un modèle exponentiel-conjugué, les distributions a priori associées aux paramètres ν et λ sont respectivement une distribution gaussienne rectifiée et une distribution gamma :

$$p(\nu|m_\nu, l_\nu) = \mathcal{N}^R(\nu|m_\nu, l_\nu) \quad \text{et} \quad p(\lambda_k|a_\lambda, b_\lambda) = \mathcal{G}(\lambda_k|a_\lambda, b_\lambda). \tag{3.40}$$

La probabilité jointe du modèle ricien s'écrit :

$$\begin{aligned} p(\boldsymbol{x}, \nu, \lambda|m_\nu, l_\nu, a_\lambda, b_\lambda) &= \Pi_{n=1}^N \, p(x_n|\nu, \lambda) \, p(\nu|m_\nu, l_\nu) \, p(\lambda|a_\lambda, b_\lambda) \\ &\geq \Pi_{n=1}^N \, \mathcal{F}(x_n, \nu, \lambda, \xi_n) \, p(\nu|m_\nu, l_\nu) \, p(\lambda|a_\lambda, b_\lambda). \end{aligned} \tag{3.41}$$

avec $\boldsymbol{x} = \{x_n\}_{n=1}^N$. La figure 3.9 illustre sous la forme d'un modèle graphique probabiliste cette probabilité jointe.

La probabilité a posteriori approximée par l'algorithme VMP est formulée comme suit :

$$p(\nu, \lambda|\boldsymbol{x}, m_\nu, l_\nu, a_\lambda, b_\lambda) \approx q(\nu) \, q(\lambda), \tag{3.42}$$

et les densités a posteriori optimales s'écrivent

$$q^*(\nu) = \mathcal{N}^R(\nu|m_\nu^*, l_\nu^*) \quad \text{et} \quad q^*(\lambda_k) = \mathcal{G}(\lambda|a_\lambda^*, b_\lambda^*). \tag{3.43}$$

Le graphe probabiliste associé au modèle, représenté par la figure 3.9, est utilisé avec l'algorithme VMP pour calculer les valeurs optimales des paramètres m_ν^*, l_ν^*, a_λ^* et b_λ^* par passage de messages entre les nœuds du graphe. Les messages du nœud associé à la distribution de Rice ont été présentés aux équations (3.32) et (3.34). L'intégralité des messages mis en jeu par le modèle, pour chaque type de distribution, sont données en annexe A.

La figure 3.10 illustre l'application du modèle bayésien ricien à un jeu de données constitué de deux cents points, générés à partir d'une distribution de Rice de paramètres $\nu = 3$ et $\lambda = 0{,}25$. La distribution a posteriori des paramètres d'un modèle bayésien ricien a été approximée avec l'algorithme VMP. Les conditions de calcul sont données en annexe B.4. Sur cet exemple simple, les paramètres de la distribution de Rice sont retrouvés avec précision.

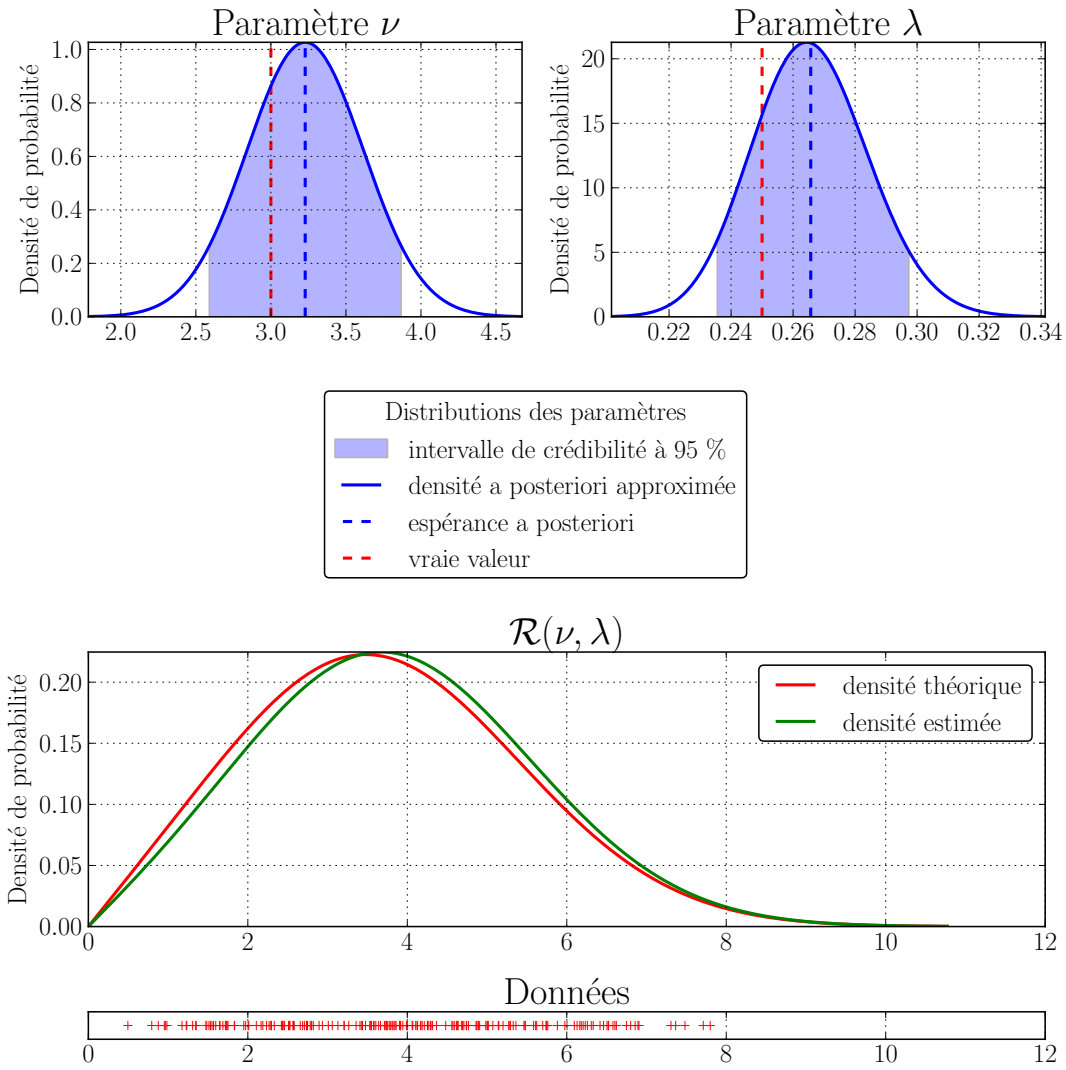

FIGURE 3.10. – Distributions a posteriori approximées des paramètres d'un modèle ricien appliqué à un échantillon de 200 observations, issues d'une distribution de Rice. La densité de cette distribution de Rice est estimée à partir des espérances a posteriori des paramètres.

3.4.3. Inférence bayésienne par l'algorithme VMP pour une distribution de mélange ricien

La distribution de mélange ricien est une distribution de mélange qui utilise pour ses composantes des distributions de Rice. Soit X une variable aléatoire suivant une distribution de mélange ricien, sa densité de probabilité s'écrit :

$$\begin{aligned} p(x|z,\boldsymbol{\nu},\boldsymbol{\lambda}) &= \Pi_{k=1}^{K}\, \mathcal{R}ice(x|\nu_k,\lambda_k)^{z_k} \\ &= \exp\left(\sum_{k=1}^{K} z_k \left(\ln I_0(x\nu_k\lambda_k) + \ln x + \ln\lambda_k - \frac{1}{2}\lambda_k(x^2+\nu_k^2)\right)\right), \end{aligned} \tag{3.44}$$

avec $\boldsymbol{\nu} = \{\nu_k\}_{k=1}^{K}$ et $\boldsymbol{\lambda} = \{\lambda_k\}_{k=1}^{K}$.

Comme la distribution de Rice, la distribution de mélange ricien ne forme pas une famille exponentielle. Dans ce cas également, l'approximation variationnelle locale permet d'utiliser cette distribution dans un modèle exponentiel-conjugué.

Pour construire une borne inférieure variationnelle $\mathcal{F}_m$, il suffit d'utiliser dans l'équation (3.44) la borne inférieure de $\ln I_0$ définie à l'équation (3.30). Ainsi le logarithme de la borne $\mathcal{F}_m$ est formulé comme suit :

$$\begin{aligned} &\ln \mathcal{F}_m(x,z,\boldsymbol{\nu},\boldsymbol{\lambda},\xi) \\ &= \sum_{k=1}^{K} z_k \left(\ln I_0(\xi_k) + \frac{I_1(\xi_k)}{I_0(\xi_k)}(x\nu_k\lambda_k - \xi_k) + \ln x - \ln\lambda_k - \frac{1}{2}\lambda_k(x^2+\nu_k^2)\right) \\ &= \underbrace{\begin{pmatrix} \sum_{k=1}^{K} z_k \frac{I_1(\xi_k)}{I_0(\xi_k)}\nu_k\lambda_k \\ -\frac{1}{2}\sum_{k=1}^{K} z_k\lambda_k \end{pmatrix}^{\top}}_{\eta_x(z,\boldsymbol{\nu},\boldsymbol{\lambda},\xi)} \cdot \underbrace{\begin{pmatrix} x \\ x^2 \end{pmatrix}}_{\mathrm{u}_x(x)} + \underbrace{\ln x}_{f_x(x)} \\ &\qquad + \underbrace{\sum_{k=1}^{K} z_k \left(-\frac{1}{2}\lambda_k\nu_k^2 + \ln\lambda_k - \frac{I_1(\xi_k)}{I_0(\xi_k)}\xi_k + \ln I_0(\xi_k)\right)}_{g_x(z,\boldsymbol{\nu},\boldsymbol{\lambda},\xi)}. \end{aligned} \tag{3.45}$$

La borne $\mathcal{F}_m$ ne dépend pas d'un seul paramètre variationnel mais d'un vecteur $\xi = (\xi_1,\ldots,\xi_K)$ de K paramètres variationnels. Par construction, la borne $\mathcal{F}_m$ respecte l'inégalité $p(x|z,\boldsymbol{\nu},\boldsymbol{\lambda}) \geq \mathcal{F}_m(x,z,\boldsymbol{\nu},\boldsymbol{\lambda},\xi)$. De plus, la borne $\mathcal{F}_m$ présente la même forme que la densité d'une distribution de la famille exponentielle et peut donc être utilisée avec l'algorithme VMP.

La réécriture de $\ln \mathcal{F}_m$ vis-à-vis du paramètre ν_k donne :

$$\ln \mathcal{F}_m(x, z, \boldsymbol{\nu}, \boldsymbol{\lambda}, \xi) = \underbrace{\begin{pmatrix} z_k \frac{I_1(\xi_k)}{I_0(\xi_k)} \lambda_k \, x \\ -\frac{1}{2} z_k \lambda_k \end{pmatrix}^\top}_{\ell_{x,\nu_k}(x, z, \boldsymbol{\nu} \backslash \{\nu_k\}, \boldsymbol{\lambda}, \xi)} \cdot \underbrace{\begin{pmatrix} \nu_k \\ \nu_k^2 \end{pmatrix}}_{\mathrm{u}_{\nu_k}(\nu_k)} + h_{x,\nu_k}(x, z, \boldsymbol{\nu} \backslash \{\nu_k\}, \boldsymbol{\lambda}, \xi). \tag{3.46}$$

Le vecteur de statistiques naturelles de la distribution a priori de ν_k est fixé à $\mathrm{u}_{\nu_k}(\nu_k) = (\nu_k, \nu_k^2)$. Sachant que ν_k est positif, il peut donc s'agir par exemple d'une distribution gaussienne rectifiée. Si X est une variable observée, le message employé par l'algorithme VMP vers le nœud du graphe probabiliste représentant ν_k vaut :

$$m_{X \to \nu_k} = \mathbb{E}_{q(z), q(\lambda_k)}[\ell_{x,\nu_k}(x, z, \boldsymbol{\nu} \backslash \{\nu_k\}, \boldsymbol{\lambda}, \xi)]. \tag{3.47}$$

En procédant de la même manière pour les paramètres λ_k, on obtient :

$$\ln \mathcal{F}_m(x, z, \boldsymbol{\nu}, \boldsymbol{\lambda}, \xi) \tag{3.48}$$

$$= \underbrace{\begin{pmatrix} z_k \left(\frac{I_1(\xi_k)}{I_0(\xi_k)} \nu_k x - \frac{1}{2}(x^2 + \nu_k^2) \right) \\ z_k \end{pmatrix}^\top}_{\ell_{x,\lambda_k}(x, z, \boldsymbol{\nu}, \boldsymbol{\lambda} \backslash \{\lambda_k\}, \xi)} \cdot \underbrace{\begin{pmatrix} \lambda_k \\ \ln \lambda_k \end{pmatrix}}_{\mathrm{u}_{\lambda_k}(\lambda_k)} + h_{x,\lambda_k}(x, z, \boldsymbol{\nu}, \boldsymbol{\lambda} \backslash \{\lambda_k\}, \xi). \tag{3.49}$$

La distribution a priori de λ_k a pour vecteur de statistiques naturelles $\mathrm{u}_{\lambda_k}(\lambda_k) = (\lambda_k, \ln \lambda_k)$, correspondant typiquement à une distribution gamma. En considérant le cas où X est une variable observée, le message vers le nœud du graphe probabiliste représentant λ_k est :

$$m_{X \to \lambda_k} = \mathbb{E}_{q(z), q(\nu_k)}[\ell_{x,\lambda_k}(x, z, \boldsymbol{\nu}, \boldsymbol{\lambda} \backslash \{\lambda_k\}, \xi)]. \tag{3.50}$$

S'agissant du paramètre z, la réécriture de $\ln \mathcal{F}_m$ est la suivante :

$$\ln \mathcal{F}_m(x, z, \boldsymbol{\nu}, \boldsymbol{\lambda}, \xi)$$

$$= \underbrace{\begin{pmatrix} \ln I_0(\xi_1) + \frac{I_1(\xi_1)}{I_0(\xi_1)}(x \nu_1 \lambda_1 - \xi_1) - \ln \lambda_1 - \frac{1}{2}\lambda_1(x^2 + \nu_1^2) \\ \vdots \\ \ln I_0(\xi_K) + \frac{I_1(\xi_K)}{I_0(\xi_K)}(x \nu_K \lambda_K - \xi_K) - \ln \lambda_K - \frac{1}{2}\lambda_K(x^2 + \nu_K^2) \end{pmatrix}^\top}_{\ell_{x,z}(x, \boldsymbol{\nu}, \boldsymbol{\lambda}, \xi)} \cdot \underbrace{\begin{pmatrix} z_1 \\ \vdots \\ z_K \end{pmatrix}}_{\mathrm{u}_z(z)} + \ln x. \tag{3.51}$$

Le vecteur des statistiques naturelles $\mathrm{u}_z(z)$ contraint la distribution a priori de Z. Une distribution habituelle respectant cette contrainte est la distribution catégorique (categorical distribution). Si X est une variable observée, le message vers le nœud représentant Z est :

$$m_{X \to Z} = \mathbb{E}_{q(\boldsymbol{\nu}), q(\boldsymbol{\lambda})}[\ell_{x,z}(x, \boldsymbol{\nu}, \boldsymbol{\lambda}, \xi)]. \tag{3.52}$$

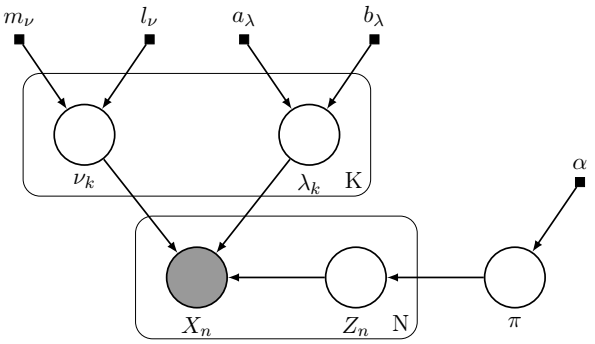

FIGURE 3.11. – Modèle graphique probabiliste associé à la distribution jointe d'un modèle de mélange ricien tel que décrit par l'équation (3.55)

Toujours dans le cas où X est une variable observée, la contribution de $\mathcal{F}_m$ à la valeur de la borne inférieure $\mathcal{L}(q)$ maximisée par l'algorithme VMP a pour valeur :

$$\mathcal{L}_{\mathcal{F}_m} = \mathbb{E}_{q(z),q(\boldsymbol{\nu}),q(\boldsymbol{\lambda})}[\eta_x(z,\boldsymbol{\nu},\boldsymbol{\lambda},\xi)]^\top \cdot \mathrm{u}_x(x) + f_x(x) + \mathbb{E}_{q(z),q(\boldsymbol{\nu}),q(\boldsymbol{\lambda})}[g_x(z,\boldsymbol{\nu},\boldsymbol{\lambda},\xi)]. \tag{3.53}$$

Le vecteur des paramètres variationnels ξ qui maximise la borne inférieure $\mathcal{L}(q)$ est le même que celui qui maximise $\mathcal{L}_{\mathcal{F}_m}$. En examinant la forme de $\ln \mathcal{F}_m$ donnée à l'équation (3.45), il est évident que chaque ξ_k peut être calculé indépendamment, en maximisant les différents termes de la somme. Le calcul est similaire à celui conduit dans le cas de la distribution de Rice. Il en résulte l'estimateur $\xi_k^* = \mathbb{E}_{q(\nu_k)}[\nu_k]\,\mathbb{E}_{q(\lambda_k)}[\lambda_k]x$, mis en jeu dans l'algorithme VMP pour ré-estimer les ξ_k avant d'envoyer les messages $m_{X\to\nu_k}$ et $m_{X\to\lambda_k}$ aux nœuds respectifs.

Les modèles que j'utilise ne mettant pas en jeu de variables aléatoires cachées suivant des distributions de mélange ricien, les calculs associés à la distribution a posteriori d'une distribution de mélange ricien ne seront pas détaillés.

3.4.4. Exemple de modèle de mélange ricien et application

Soient N observations associées à un ensemble de variables aléatoires $\boldsymbol{X} = \{X_n\}_{n=1}^N$. Chaque variable aléatoire X_n suit une distribution de mélange ricien à K composantes. Leur densité de probabilité est alors formulée comme suit :

$$\begin{aligned} p(x_n|z_n,\boldsymbol{\nu},\boldsymbol{\lambda}) &= \Pi_{k=1}^K \mathcal{R}ice(x_n|\nu_k,\lambda_k)^{z_{n,k}} \\ &\geq \mathcal{F}_m(x_n,z_n,\boldsymbol{\nu},\boldsymbol{\lambda},\xi_n) \end{aligned} \tag{3.54}$$

avec $\boldsymbol{\nu} = \{\nu_k\}_{k=1}^K$ et $\boldsymbol{\lambda} = \{\lambda_k\}_{k=1}^K$.

Un modèle bayésien de mélange ricien est construit en définissant les probabilités a priori des variables Z_n, ν_k et λ_k. Pour obtenir un modèle exponentiel-conjugué, il suffit d'associer comme distributions a priori :

– une distribution catégorique de paramètre π pour chaque Z_n,

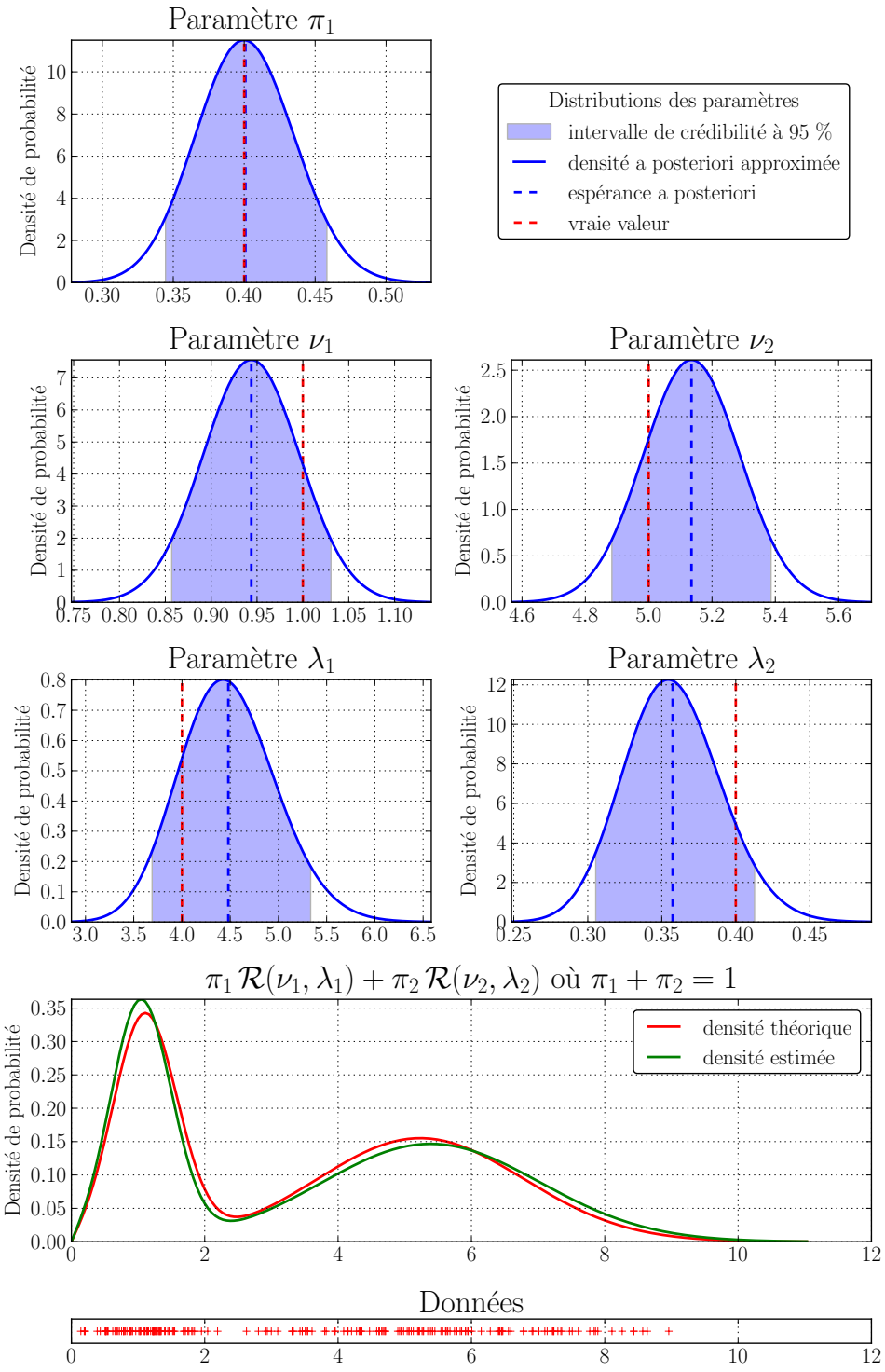

FIGURE 3.12. – Distributions a posteriori approximées des paramètres d'un modèle de mélange ricien appliqué à un échantillon de 200 observations, issues d'une distribution de mélange ricien. La densité de cette distribution de mélange est estimée à partir des espérances a posteriori des paramètres.

– une distribution de Dirichlet pour ce paramètre π,
– une distribution gaussienne rectifiée pour ν_k,
– une distribution gamma pour λ_k.

La densité jointe du modèle a pour formule :

$$\begin{aligned} & p(\boldsymbol{x}, \boldsymbol{z}, \pi, \boldsymbol{\nu}, \boldsymbol{\lambda} | \alpha, m_\nu, l_\nu, a_\lambda, b_\lambda) \\ & = \Pi_{n=1}^{N} p(x_n | z_n, \boldsymbol{\nu}, \boldsymbol{\lambda}) \, p(z_n | \pi) \, p(\pi | \alpha) \, \Pi_{k=1}^{K} p(\nu_k | m_\nu, l_\nu) \, p(\lambda_k | a_\lambda, b_\lambda) \\ & \geq \Pi_{n=1}^{N} \mathcal{F}_m(x_n, z_n, \nu, \lambda, \xi) \, p(z_n | \pi) \, p(\pi | \alpha) \, \Pi_{k=1}^{K} p(\nu_k | m_\nu, l_\nu) \, p(\lambda_k | a_\lambda, b_\lambda) \end{aligned} \tag{3.55}$$

avec $\boldsymbol{x} = \{x_n\}_{n=1}^{N}$, $\boldsymbol{z} = \{z_n\}_{n=1}^{N}$ et les densités a priori

$$\begin{aligned} p(z_n | \pi) &= \Pi_{k=1}^{K} \pi_k^{z_{n,k}} & p(\nu_k | m_\nu, l_\nu) &= \mathcal{N}^R(\nu_k | m_\nu, l_\nu) \\ p(\pi | \alpha) &= \mathcal{D}ir(\pi | \alpha) & p(\lambda_k | a_\lambda, b_\lambda) &= \mathcal{G}(\lambda_k | a_\lambda, b_\lambda) \end{aligned} \tag{3.56}$$

Il est alors possible d'approximer la probabilité a posteriori du modèle avec l'algorithme VMP :

$$p(\boldsymbol{z}, \pi, \boldsymbol{\nu}, \boldsymbol{\lambda} | \boldsymbol{x}, \alpha, m_\nu, l_\nu, a_\lambda, b_\lambda) \approx \Pi_{n=1}^{N} q(z_n) \, q(\pi) \, \Pi_{k=1}^{K} q(\nu_k) \, q(\lambda_k) \tag{3.57}$$

avec les densités a posteriori approximées optimales

$$\begin{aligned} q^*(z_n) &= \Pi_{k=1}^{K} (\pi_{n,k}^*)^{z_{n,k}} & q^*(\nu_k) &= \mathcal{N}^R(\nu_k | m_{\nu_k}^*, l_{\nu_k}^*) \\ q^*(\pi) &= \mathcal{D}ir(\pi | \alpha^*) & q^*(\lambda_k) &= \mathcal{G}(\lambda_k | a_{\lambda_k}^*, b_{\lambda_k}^*) \end{aligned} \tag{3.58}$$

La figure 3.11 présente le modèle graphique probabiliste correspondant au modèle bayésien de mélange ricien tel que présenté ici. Ce graphe est mis en jeu par l'algorithme VMP pour calculer les paramètres des distributions a posteriori approximées.

Un exemple simple d'utilisation du modèle de mélange ricien avec l'algorithme VMP est présenté à la figure 3.12. Le jeu de données utilisé comporte deux cents points, générés à partir d'une distribution de mélange ricien à deux composantes, de paramètres $\nu_1 = 1$, $\nu_2 = 5$, $\lambda_1 = 4$, $\lambda_2 = 0{,}4$, $\pi_1 = 0{,}4$ et $\pi_2 = 0{,}6$. Pour le calcul de la distribution a posteriori approximée avec l'algorithme VMP, le nombre de composantes est fixé à $K = 5$. Les autres détails de calcul, tels que l'initialisation des hyperparamètres du modèle et la détection de la convergence, sont donnés en annexe B.5. Grâce au mécanisme de suppression des composantes inutilisées (cf. section 3.2.3), la distribution a posteriori approximée ne comporte que deux composantes. Dans cet exemple, les paramètres des composantes sont retrouvés avec précision.

3.5. Banc d'essai des modèles bayésiens de mélange gaussien et ricien

La section 3.1.1 a mis en évidence que le module des coefficients d'ondelettes d'un signal contaminé par un bruit blanc gaussien additif suit une distribution de Rice. Dès lors, un échantillon contenant le module des coefficients d'ondelettes du même point

temps-fréquence de différentes séries temporelles suit une distribution de Rice si toutes ces séries contiennent le même signal et sont contaminés par un même bruit blanc gaussien additif. Si les séries temporelles ne contiennent pas le même signal, mais peuvent être groupées en classe de signaux similaires, alors l'échantillon peut être partitionné pour retrouver ces classes.

Le but de la présente section est d'évaluer sur un exemple simple la capacité des différents modèles de mélange à partitionner un tel échantillon de module de coefficients d'ondelettes, pour retrouver la classe d'appartenance des séries temporelles.

3.5.1. Jeu de données artificielles et modèles évalués

Un jeu de données est constitué à partir de deux classes de séries temporelles : une classe sans signal et une classe avec un signal de niveau arbitraire. Le niveau de bruit est identique dans les deux classes. Ainsi, le module des coefficients d'ondelettes issus de la première classe est obtenu par tirages aléatoires d'une distribution de Rice de paramètres $\nu = 0$ et $\sigma = 1$. Le module des coefficients issus de la seconde classe est issu de tirages aléatoires d'une distribution de Rice de paramètres $\nu = \nu_0$ et $\sigma = 1$, où ν_0 est une valeur arbitraire.

Deux variantes de jeu de données sont utilisées. La première variante restreint le jeu de données à 20 échantillons, tandis que la seconde variante utilise 500 échantillons. Pour chaque variante, 5 000 jeux de données sont générés. Chacun de ces jeux de données comporte un nombre aléatoire d'échantillons issus de chaque classe, une classe ne pouvant pas être vide et le nombre total d'échantillons étant fixé par la variante du jeu de données. De plus, la valeur ν_0 de chaque jeu de données est générée à partir d'une distribution uniforme sur l'intervalle $[0{,}1; 6]$, pour obtenir des cas avec une plus ou moins bonne séparation des classes [2].

Pour chaque variante de jeu de données, 5 000 jeux de données supplémentaires sont générés, ne comportant que des échantillons issus de la seconde classe. Ces jeux de données servent à vérifier que les modèles de mélange sont capables d'identifier la présence d'une seule classe.

Les modèles de mélange testés sont les modèles bayésiens de mélange gaussien et ricien à deux composantes, respectivement désignés par les abréviations G2M et R2M. Les détails de mise en œuvre de ces modèles, tels que l'initialisation et l'analyse de la convergence, sont donnés en annexes B.3 et B.5. Pour le modèle de mélange gaussien, des pré-traitements sur les données sont aussi testés :

- l'utilisation du carré du module des coefficients d'ondelettes ($\mathrm{G2M_{pow}}$) au lieu du simple module, pour vérifier si l'erreur commise empire en partitionnant directement la distribution d'énergie des signaux,
- l'utilisation du logarithme du module des coefficients d'ondelettes ($\mathrm{G2M_{log}}$), comme cette transformée est parfois utilisée en analyse de signaux corticaux afin de rendre plus gaussiennes les données [Grandchamp and Delorme, 2011].

Pour le modèle de mélange ricien, une variante testée consiste à ne faire l'inférence qu'à

2. La séparation des classes est faible pour $\nu_0 = 0{,}1$ et grande pour $\nu_0 = 6$.

	G2M	G2M_{pow}	G2M_{log}	R2M	R2M_{med}
bonne détection de 1 classe	85,24 %	85,10 %	85,14 %	97,70 %	88,42 %
bonne détection de 2 classes	28,58 %	29,32 %	29,26 %	22,98 %	53,42 %
précision $\geq 90\,\%$	20,80 %	21,06 %	20,88 %	20,42 %	39,42 %

Tableau 3.1. – Performances des différents modèles de mélange pour des jeux de données à 20 échantillons, pour la bonne détection d'une ou deux classes. La précision ne concerne que les jeux de données à deux classes. Voir la section 3.5.1 pour les abréviations.

partir d'une seule initialisation, fondée sur les classes des données à partitionner avec leur position, supérieure ou inférieure, à la médiane de ces données (R2M_{med}). L'intérêt de cette heuristique est de réduire le nombre de calculs.

L'évaluation des modèles porte tout d'abord sur leur capacité à détecter si le jeu de données comporte une ou deux classes. Pour les jeux de données à deux classes, le taux de bonne classification, ou *précision*, des modèles est aussi calculé. Quand un modèle ne détecte qu'une composante pour un jeu de données à deux classes, la précision est considérée à 50 %, le modèle n'apportant aucune information par rapport à un partitionnement aléatoire.

3.5.2. Évaluation sur un petit jeu de données

Les résultats des modèles de mélange sur des jeux de données à 20 échantillons sont présentés dans le tableau 3.1 et la figure 3.13. Une première observation est que l'emploi d'un pré-traitement des données n'influence pas sensiblement le résultat des modèles de mélange gaussien.

Concernant la détection de la présence d'une seule classe pour les jeux de données à une classe, les modèles de mélange ricien obtiennent les meilleures performances. Tous les modèles restent supérieurs à 85 % de bonne détection.

Au contraire, la détection de la présence de deux classes est très mauvaises pour l'ensemble des modèles testés. Le meilleur cas, R2M_{med}, n'obtient qu'un peu plus de 50 % de bonne détection de la situation. Les autres modèles sont en dessous de 30 % de bonne détection. Cette déficience des modèles peut s'expliquer par la très faible quantité d'information contenue dans seulement 20 échantillons, ce qui favorise des modèles plus parcimonieux, avec une seule composante. Cette tendance est accentuée par l'approximation par champ moyen, comme discuté à la section 2.2.3, page 32.

Un examen de la figure 3.13 met en relief une meilleure précision des modèles de mélange quand les deux classes sont bien séparées, à partir de $\nu_0 \approx 3$. À l'exception de R2M_{med}, tous les autres modèles présentent néanmoins des difficultés à détecter les deux classes avec précision quand le nombre d'échantillons avec et sans signal est similaire.

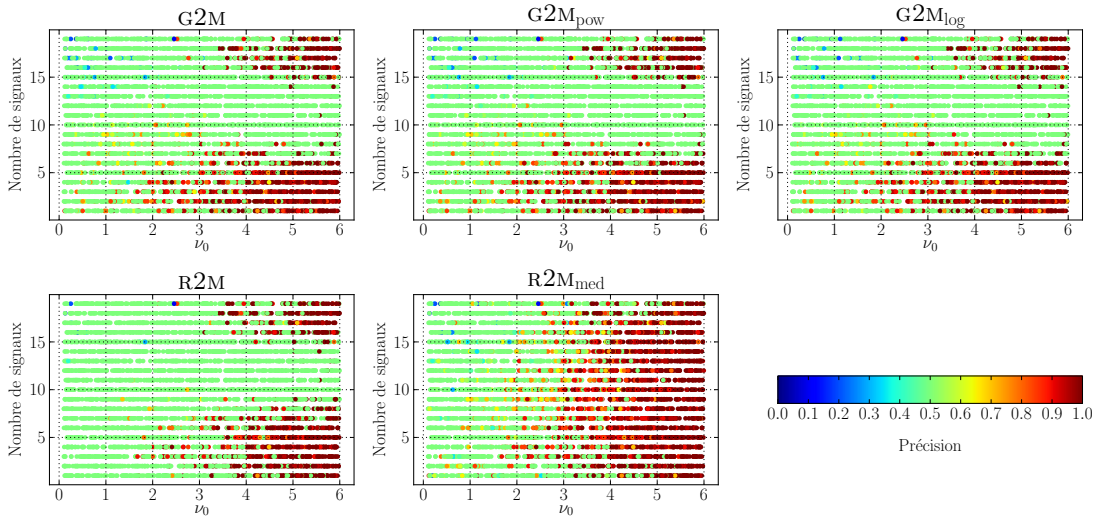

FIGURE 3.13. – Comparaison des performances des différents modèles de mélange pour des jeux de données à 20 échantillons, en fonction de l'amplitude ν_0 du signal et du nombre d'échantillons avec signal. Voir la section 3.5.1 pour les abréviations.

	G2M	G2M$_{pow}$	G2M$_{log}$	R2M	R2M$_{med}$
bonne détection de 1 classe	0,00 %	0,00 %	0,00 %	97,88 %	90,92 %
bonne détection de 2 classes	95,96 %	96,10 %	96,04 %	60,18 %	64,46 %
précision $\geq$ 90 %	43,40 %	43,42 %	43,40 %	48,30 %	48,64 %

Tableau 3.2. – Performances des différents modèles de mélange pour des jeux de données à 500 échantillons, pour la détection d'une ou deux classes. La précision ne concerne que les jeux de données à deux classes. Voir la section 3.5.1 pour les abréviations.

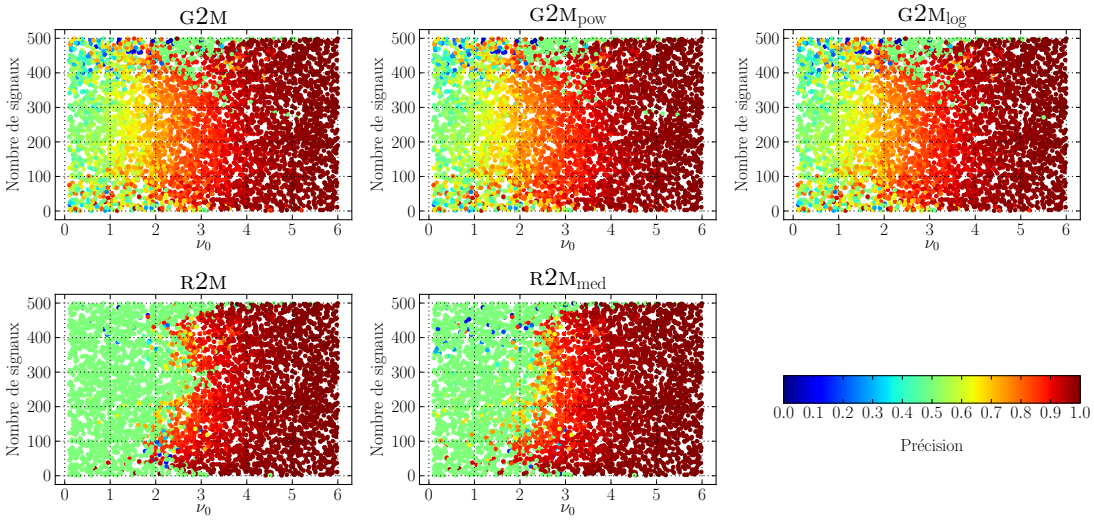

FIGURE 3.14. – Comparaison des performances des différents modèles de mélange pour des jeux de données à 500 échantillons, en fonction de l'amplitude ν du signal et du nombre d'échantillons avec signal. Voir la section 3.5.1 pour les abréviations.

3.5.3. Évaluation sur un grand jeu de données

Les résultats pour les jeux de données à 500 échantillons sont compilés dans le tableau 3.2 et la figure 3.14.

Dans le cas de jeux de données avec de nombreux échantillons, les modèles de mélange gaussien ne sont plus à même de détecter les jeux de données avec une seule classe. À l'inverse, les modèles de mélange ricien ont des résultats supérieurs à 90 % de bonne détection. L'échec des modèles de mélange gaussien vient de l'asymétrie des distributions riciennes qui ne peuvent être approchées avec une distribution symétrique telle que la distribution gaussienne. Le grand nombre d'échantillons fait apparaître plus fortement cette asymétrie et force ainsi les modèles de mélange gaussien à utiliser deux composantes pour modéliser les données. La transformation par le logarithmique n'améliore pas cette situation.

Le biais des modèles de mélange gaussien leur est favorable pour la détection de jeux de données à deux classes. Les modèles de mélange ricien subissent quant à eux le biais dû à l'approximation par champ moyen, et favorisent plus les modèles à une composante, d'où un taux de bonne détection de l'ordre de 60 %. Cependant, sur la figure 3.14, il est visible que les modèles de mélange ricien affectés sont ceux où la séparation des deux classes est mauvaise, pour $\nu_0 < 2$. Dans ce même contexte, les modèles de mélange gaussien détectent bien deux classes, mais leur précision reste faible, de l'ordre de 60 %.

3.5.4. Discussion sur l'utilisation des différents modèles de mélange

D'après les résultats précédents, dans le contexte d'un petit échantillon, les modèles de mélange ricien avec une heuristique d'initialisation à l'aide de la médiane sont les modèles les plus performants. Dans le cas où l'on ne souhaite pas se reposer sur cette heuristique, les modèles de mélange gaussien se révèlent alors aussi attractifs que les modèles de mélange ricien, avec une implémentation plus simple que ces derniers.

De même, dans le contexte d'un grand échantillon, l'heuristique d'initialisation des modèles de mélange ricien reste l'option la plus pertinente. L'usage des modèles de mélange gaussien peut être justifié s'il est assuré que les données comportent deux classes.

3.6. Résumé

Ce chapitre a introduit la distribution de Rice en tant que distribution théorique du module des coefficients d'ondelettes continues d'un signal bruité. Il a par ailleur présenté des modèles bayésiens de mélange utilisant cette connaissance pour partitionner ces coefficients d'ondelettes. Ces apports sont fondamentaux pour modéliser finement les signaux corticaux et sont ainsi au cœur de mes méthodes d'analyse de synchronisations.

Sous l'hypothèse qu'un enregistrement électro-cortical peut être considéré comme un signal contaminé par un bruit blanc gaussien additif, alors le module de chaque coefficient de sa transformée en ondelettes continue suit une distribution de Rice. Cette propriété semble conservée pour d'autres modèles de bruit blanc additif, et peut donc être utilisée en première approximation pour modéliser le module des coefficients d'ondelettes.

Dès lors, pour différencier parmi un ensemble de coefficients d'ondelettes celles représentant un même signal bruité, il suffit de modéliser la répartition du module de ces coefficients avec des modèles bayésiens de mélange. L'intérêt d'un partitionnement par un modèle bayésien provient principalement de la faculté de ce type de modèle de sélectionner automatiquement le nombre de composantes du modèle. Deux modèles de mélange ont été discutés : le modèle de mélange gaussien et le modèle de mélange ricien.

Le modèle bayésien de mélange gaussien est un modèle connu dans la littérature. Bien que théoriquement inadéquat pour partitionner des coefficients d'ondelettes, il peut facilement être mis en œuvre avec l'algorithme VMP et ses résultats peuvent être utilisés sous certaines conditions.

Le modèle bayésien de mélange ricien est un nouveau modèle spécialement adapté au partitionnement des coefficients d'ondelettes. Pour utiliser l'algorithme VMP avec ce modèle, une approximation variationnelle locale a été employée pour dériver les équations nécessaires à l'inférence des paramètres de la distribution de mélange ricien.

Chapitre 4.

Méthodes d'analyse de synchronisations dans le plan temps-fréquence

> - Homer : Les enfants, y a trois façons de faire les choses. La bonne, la mauvaise et celle de Max Puissant.
> - Bart : C'est pas la mauvaise ?
> - Homer : Si... mais en plus rapide.
>
> *Les Simpson – Épisode 13, saison 10*

Au chapitre 1, les outils conventionnels d'étude des synchronisations au sein de signaux électro-corticaux ont été présentés, et leurs limitations exposées. Ces limitations reposent principalement sur le moyennage des scalogrammes entre les enregistrements d'une même expérience et la détection de phénomènes d'intérêt par une comparaison avec l'activité pré-stimulus.

Les méthodes que j'ai mises au point pour étudier la synchronisation sans avoir recours à ces techniques reposent sur une approche en deux étapes principales :

1. l'analyse *simple essai* des scalogrammes de chaque enregistrement, pour en extraire une information symbolique portant sur la possible présence d'épisodes de synchronisation,
2. la validation statistique des résultats symboliques obtenus dans chaque enregistrement, pour détecter les événements liés à l'expérience.

Les analyses à l'échelle de chaque enregistrement permettent de prendre en compte finement la variabilité de ceux-ci. Les outils mis en jeu à cette étape sont les modèles bayésiens de mélange gaussien et ricien, évoqués au chapitre 3. Mes deux méthodes se servent de ces modèles de mélange mais y associent des hypothèses différentes sur les données. L'étape de validation de chaque méthode est propre au type d'information extraite à l'étape précédente.

La première partie de ce chapitre détaillera une méthode d'analyse multivariée, fondée sur l'hypothèse qu'il est possible de détecter un épisode de synchronisation s'il apparaît simultanément dans l'enregistrement d'un sous-ensemble des électrodes.

La seconde partie sera consacrée à une méthode d'analyse univariée, fondée sur l'hypothèse que l'ensemble du signal d'une électrode, du stimulus à la réponse, peut être modélisé comme un système à deux états, haut et bas.

4.1. Méthode d'analyse de synchronisations cooccurrentes par modèles de mélange gaussien

La première méthode d'analyse que j'ai mise au point utilise les caractéristiques jointes des électrodes pour identifier des épisodes de synchronisations ou désynchronisations au sein de chaque électrode. L'hypothèse sous-jacente est qu'une fluctuation de la densité d'énergie temps-fréquence du signal peut être coïncidente entre plusieurs électrodes. Ainsi, le terme de « synchronisation cooccurrente » désignera un épisode de synchronisation, ou de désynchronisation, détecté en un point temps-fréquence au sein du signal de plusieurs électrodes. Avec cette hypothèse, l'analyse des synchronisations au sein des électrodes n'a plus besoin d'une comparaison avec le signal pré-stimulus, remplacée par une comparaison entre les électrodes.

Par conséquent, l'information extraite au sein de chaque enregistrement porte sur les groupes d'électrodes présentant une synchronisation cooccurrente, en chaque point temps-fréquence du signal. La combinaison de l'information récoltée dans chaque enregistrement porte alors sur l'invariance en temps et en fréquence des groupes d'électrodes détectés. Cette combinaison ne porte que sur l'information structurelle extraite des enregistrements, et permet donc une plus grande robustesse vis-à-vis de la variabilité du signal de ces enregistrements.

Ma méthode d'analyse s'articule donc autour des deux étapes suivantes :

1. une première étape de *détection* des groupes d'électrodes à synchronisation cooccurrente dans chaque enregistrement, à chaque point temps-fréquence, elle-même décomposée en deux phases
   a) l'estimation de la densité d'énergie temps-fréquence des signaux des électrodes dans les différents enregistrements par la transformée en ondelettes continue, avec l'ondelette de Morlet normalisée,
   b) le partitionnement de la densité d'énergie des signaux des électrodes à chaque point temps-fréquence de chaque enregistrement, à l'aide de modèles de mélange gaussien
2. une seconde étape de *validation* des groupes d'électrodes détectés à chaque point temps-fréquence de chaque enregistrement
   a) par une *mesure de stabilité locale* permettant de rejeter au sein de chaque enregistrement les groupes d'électrodes détectés fallacieux,
   b) suivie d'une *mesure de stabilité globale* des groupes d'électrodes retenus, pour extraire les motifs de synchronisations cooccurrentes pertinents vis-à-vis de l'expérience analysée.

L'enchaînement de ces différentes étapes est résumé par la figure 4.1. Cette méthode d'analyse a fait l'objet d'une présentation à la conférence *Neurocomp 2010* [Rio et al., 2010] et d'une publication dans le *Journal of Physiology – Paris* [Rio et al., 2011]. Les sections suivantes détaillent les étapes de la méthode.

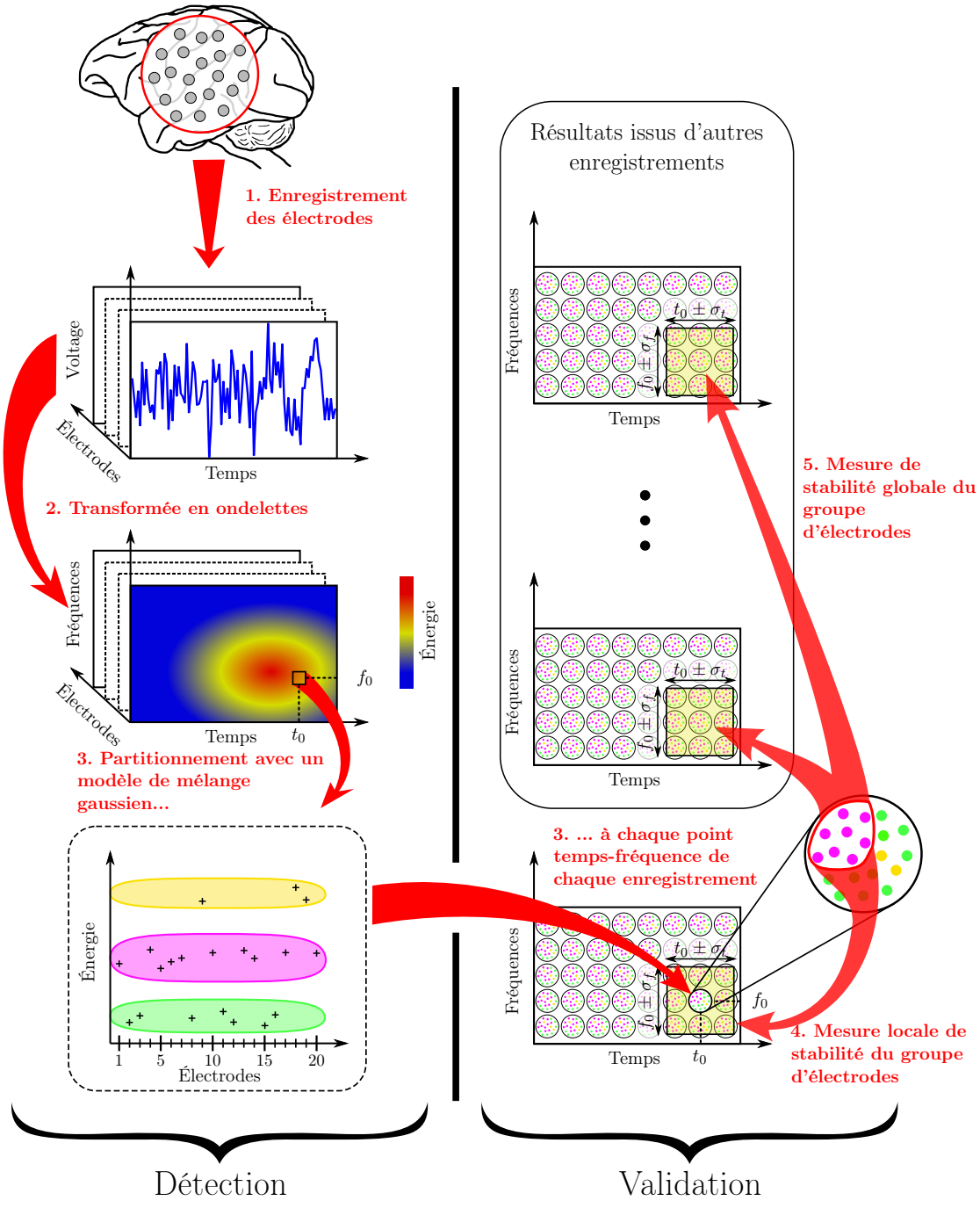

FIGURE 4.1. – Résumé graphique des étapes de la méthode d'analyse de synchronisations cooccurrentes.

4.1.1. Détection des groupes d'électrodes à synchronisation cooccurrente

L'étape de détection est une étape réalisée de manière indépendante pour chaque enregistrement d'une expérience étudiée.

Dans un premier temps, le signal est projeté dans le plan temps-fréquence à l'aide de la transformée en ondelettes continue, en utilisant une ondelette de Morlet complexe normalisée. Cette transformation, présentée en détail à la section 2.1.4, permet de représenter la répartition de l'énergie du signal en capturant à la fois les aspects transitoires du signal en hautes fréquences et les aspects stationnaires en basses fréquences, par un mécanisme de multirésolution. Dans ma méthode d'analyse, l'échantillonnage des échelles des ondelettes filles est linéaire. Cet échantillonnage sera utile pour les calculs de l'étape de validation.

Au point temps-fréquence (t, f) du r-ième enregistrement, un ensemble de données $\mathbf{x}_{r,t,f}$ contenant le module au carré des coefficients d'ondelettes du signal de toutes les électrodes est défini tel que

$$\mathbf{x}_{r,t,f} = \left\{ \mathrm{x}_{r,t,f,d} \right\}_{d \in [1,D]} \quad \text{et} \quad \mathrm{x}_{r,t,f,d} = |\mathrm{W}s_{r,d}(t, f_0/f)|^2, \tag{4.1}$$

où D est le nombre d'électrodes, et $s_{r,d}$ le signal enregistré à l'électrode d dans l'enregistrement r.

Le partitionnement de la densité d'énergie des signaux des électrodes correspond alors à un partitionnement de chaque ensemble de données $\mathbf{x}_{r,t,f}$, indépendamment les uns des autres. Pour ce partitionnement, des modèles de mélange gaussien ont été choisis en première approximation, en raison de leur facilité d'implémentation et de leur rapidité d'exécution. Les fausses partitions détectées dues à cette approximation seront filtrées à l'étape suivante de validation.

Ainsi, un modèle de mélange gaussien $\mathcal{M}_{r,t,f}$ est associé à chaque ensemble de données $\mathbf{x}_{r,t,f}$, et la probabilité a posteriori de ce modèle est calculée avec l'algorithme VMP, comme décrit en section 3.3. Chaque modèle de mélange gaussien est initialisé avec un nombre de composantes gaussiennes élevé par rapport au nombre de données, c'est-à-dire au nombre d'électrodes, pour permettre un mécanisme de suppression automatique des composantes inutiles. Comme l'algorithme VMP converge en un maximum local, plusieurs initialisations aléatoires sont nécessaires pour chaque modèle, pour obtenir la meilleure convergence possible. De plus, comme les signaux sont fortement corrélés en des points temps-fréquence proches, chaque modèle de mélange gaussien est aussi initialisé avec le résultat de partitionnement des autres modèles de mélange gaussien proches en temps et en fréquence, pour essayer d'améliorer un peu plus le résultat de l'algorithme VMP. Les autres détails pratiques de la mise en œuvre de l'algorithme VMP sont en annexe B.3.

Pour chaque modèle de mélange gaussien $\mathcal{M}_{r,t,f}$, à K composantes gaussiennes, la quantité d'intérêt retenue est la probabilité a posteriori approchée des variables cachées Z_d associées aux électrodes. Cette probabilité a posteriori, notée $q_{r,t,f}(z_{d,k})$ pour l'électrode d, porte l'information d'appartenance de cette électrode à la composante gaussienne k du modèle $\mathcal{M}_{r,t,f}$. Un groupe d'électrodes à synchronisation cooccurrente correspond aux électrodes appartenant à une même composante gaussienne d'un même modèle.

4.1.2. Validation des groupes d'électrodes détectés via une mesure de stabilité

À l'issue de l'étape de détection, en chaque point temps-fréquence de chaque enregistrement, une information a été extraite sous la forme de groupes d'électrodes. L'étape suivante consiste alors à chercher quels groupes d'électrodes sont pertinents vis-à-vis de l'expérience étudiée. Il s'agit donc, parmi tous les groupes d'électrodes isolés au sein de chaque enregistrement, de trouver ceux qui sont invariants en temps et en fréquence à travers les différents enregistrements.

L'étape de validation des groupes d'électrodes se divise en deux phases. La première phase consiste à filtrer au sein de chaque enregistrement les groupes d'électrodes pour éliminer d'éventuels groupes fallacieux. C'est une phase de validation locale. La seconde phase, qui est une validation globale, est une recherche de groupes d'électrodes similaires au travers des différents enregistrements.

Validation locale : filtrage des groupes d'électrodes dans chaque enregistrement

L'emploi d'une technique de partitionnement présuppose l'existence d'une structure naturelle dans les données partitionnées. Par conséquent, de nombreuses techniques de partitionnement fournissent un partitionnement des données même en l'absence de réelle structure dans ces données [Handl et al., 2005].

Le modèle bayésien de mélange gaussien utilisé pour la détection des sous-ensembles d'électrodes n'est pas exempt de ce problème et il faut donc vérifier la pertinence de ses résultats. Ce constat peut néanmoins être pondéré par deux observations propres à ce modèle.

La première observation est qu'il s'agit d'un modèle bayésien de mélange élaboré de manière à sélectionner automatiquement un nombre de composantes en adéquation avec les données, comme détaillé en section 3.2.3. De plus, l'utilisation de l'algorithme VMP pour le calcul des probabilités a posteriori et marginale du modèle de mélange gaussien biaise la sélection du nombre de composantes vers des modèles plus simples. Ainsi, les partitions trouvées par le modèle de mélange seront d'autant plus probables que ce modèle a une tendance à se simplifier en modèle à une seule partition.

Cependant, une seconde observation contrebalance l'aspect rassurant de la première observation : le modèle de mélange gaussien n'est pas forcément adapté aux données traitées, ce qui peut entraîner la détection de fausses partitions. Comme expliqué précédemment, ce modèle a été choisi pour sa facilité d'emploi et sert de première approximation avant une étude plus approfondie de la distribution du module des coefficients d'ondelettes d'un signal bruité. Dans le cas d'un bruit blanc gaussien additif, le banc d'essai présenté en section 3.5 montre que le modèle bayésien de mélange gaussien pour partitionner le module de coefficients d'ondelettes détecte des partitions même si les données n'en contiennent pas, ce phénomène étant amplifié lorsque la quantité de données partitionnées augmente. Ce dernier argument justifie l'emploi d'une technique de validation locale pour les résultats de chaque modèle de mélange.

Les techniques conventionnelles de validation de partitions utilisent des critères de qualité obtenus à partir des caractéristiques des partitions et à partir des données ayant servi à calculer ces partitions[Handl et al., 2005, Kerr and Churchill, 2001, Schmitzer-Torbert et al., 2005]. Ces approches emploient donc deux fois les mêmes données, pour calculer les partitions et un critère de qualité sur celles-ci, alors que le critère de qualité pourrait servir directement de critère d'optimisation pour l'algorithme de partitionnement.

Dans le cas particulier de la validation de groupes d'électrodes détectés, une information supplémentaire est disponible et permet d'envisager une autre approche pour valider ces groupes d'électrodes. Il s'agit de l'information de partitionnement en des points temps-fréquence voisins. À cause de la corrélation des coefficients d'ondelettes proches en temps et en fréquence, les partitions intrinsèques aux données existent aussi en des points temps-fréquence proches. Dès lors, un critère de validation des groupes d'électrodes correspond à une vérification de la présence de ces groupes au-delà du point temps-fréquence où ils ont été détectés, dans un voisinage temps-fréquence. Le critère de validation que j'ai mis au point est donc une mesure de *stabilité* temps-fréquence des groupes d'électrodes détectés.

Le voisinage $\boldsymbol{\mathcal{V}}_{t,f}$ d'un point temps-fréquence (t, f) du scalogramme est défini comme l'ensemble des autres points temps-fréquence contenus dans la boîte d'Heisenberg de l'ondelette fille $\psi_{\tau,a}$ centrée en (t, f) [1] :

$$\boldsymbol{\mathcal{V}}_{t,f} = \{(t', f')\}_{\substack{t' \in [t - a\sigma_{\psi,t},\, t + a\sigma_{\psi,t}] \\ f' \in \left[f - \frac{\sigma_{\psi,f}}{a},\, f + \frac{\sigma_{\psi,f}}{a}\right]}}, \tag{4.2}$$

où $\sigma_{\psi,t}$ et $\sigma_{\psi,f}$ sont les résolutions en temps et en fréquence de l'ondelette mère ψ. Le fait d'utiliser un échantillonnage linéaire des échelles des ondelettes filles dans l'étape de détection permet de s'assurer que le nombre d'éléments présents dans les voisinages $\mathcal{V}_{t,f}$ est approximativement équivalent.

La vérification de la présence d'un même groupe d'électrodes en deux points temps-fréquence voisins passe par une *mesure de similarité* entre les groupes d'électrodes détectés en ces deux points temps-fréquence. Cette similarité entre groupes d'électrodes correspond à une similarité entre les composantes des modèles de mélange gaussien calculés en ces points temps-fréquence. Ainsi, la mesure de similarité entre la composante k du modèle $\mathcal{M}_{r,t,f}$ et la composante k' du modèle voisin $\mathcal{M}_{r,t',f'}$ s'écrit

$$\mathrm{Sim}(k, k') = 1 - \frac{\mathrm{Dist}_1(k, k')}{D}, \tag{4.3}$$

où Dist_1 est la distance de *Manhattan* entre les vecteurs de probabilité a posteriori des D électrodes pour chaque composante k et k',

$$\mathrm{Dist}_1(k, k') = \sum_{d=1}^{D} |q_{r,t,f}(z_{d,k}) - q_{r,t',f'}(z_{d,k'})|. \tag{4.4}$$

Cette distance équivaut au nombre d'électrodes différentes entre les deux groupes. La mesure de similarité $\mathrm{Sim}(k, k')$ est définie dans l'intervalle $[0, 1]$. Elle est nulle si les

1. Pour rappel, l'ondelette fille $\psi_{\tau,a}$ est centrée en (t, f) si $t = \tau$ et $f = f_0/a$, avec f_0 la fréquence centrale de l'ondelette mère ψ.

composantes ne partagent aucune électrode en commun, et vaut 1 si les deux composantes définissent le même groupe d'électrodes.

La stabilité Stab(k) d'un groupe d'électrodes défini par une composante gaussienne k au point temps-fréquence (t, f) d'un enregistrement r est alors calculée comme la moyenne des meilleures similarités entre cette composante k et les composantes k' de chaque modèle $\mathcal{M}_{r,t',f'}$ du voisinage $\boldsymbol{\mathcal{V}}_{t,f}$:

$$\text{Stab}(k) = \frac{1}{\text{card}(\boldsymbol{\mathcal{V}}_{t,f})} \sum_{(t',f') \in \boldsymbol{\mathcal{V}}_{t,f}} \max_{k' \text{ comp. de } \mathcal{M}_{r,t',f'}} \left(\left\{ \text{Sim}(k, k') \right\} \right), \tag{4.5}$$

où card($\boldsymbol{\mathcal{V}}_{t,f}$) dénote le nombre d'éléments du voisinage $\boldsymbol{\mathcal{V}}_{t,f}$.

La mesure de stabilité Stab(k) est définie dans l'intervalle $[0, 1]$. Elle est proche de 1 si le groupe d'électrodes associé à la composante k se retrouve dans les modèles de mélange gaussien voisins en temps et en fréquence. Les groupes d'électrodes dont la mesure de stabilité est faible couvrent donc une faible région temps-fréquence et seront par conséquent considérés comme du bruit, dû soit au partitionnement par modèle de mélange gaussien soit à du bruit dans les données.

La validation locale des groupes d'électrodes détectés s'effectue par filtrage de leur mesure de stabilité, avec un seuil ϵ_l :

$$\text{Stab(k)} > 1 - \frac{\epsilon_l}{D}. \tag{4.6}$$

Ce seuil équivaut à une distance de Manhattan moyenne de ϵ_l entre la composante k et les composantes les plus similaires au sein des modèles de mélange voisins. Dès lors, pour $\epsilon_l = 1$, moins d'une électrode est différente entre la composante k et les composantes les plus similaires des autres modèles. Ce réglage strict est celui adopté par la suite pour limiter au maximum les composantes fallacieuses.

Validation globale : reconnaissance des groupes d'électrodes entre les enregistrements

La dernière phase de la méthode d'analyse consiste à agréger les résultats obtenus dans chaque enregistrement, à chaque point temps-fréquence, pour trouver des groupes d'électrodes invariants. Cette invariance est évaluée à l'aide d'une mesure de stabilité globale qui associe un score à chaque groupe d'électrodes, indiquant le nombre d'enregistrements dans lesquels le même groupe est retrouvé au même point temps-fréquence. Pour déterminer si un groupe d'électrodes détecté et validé localement est présent dans un autre enregistrement, la mesure de similarité introduite à l'équation (4.3) est utilisée. Avant le calcul de la stabilité globale des groupes d'électrodes, l'ensemble des enregistrements est verrouillé en temps sur le stimulus ou la réponse.

Soit un groupe d'électrodes validé localement, détecté au point temps-fréquence (t, f) de l'enregistrement r et associé à une composante k. Ce groupe d'électrodes est considéré comme présent dans un autre enregistrement r' si la meilleure similarité entre ce groupe et

les groupes détectés dans un voisinage temps-fréquence $\boldsymbol{\mathcal{V}}_{t,f}$ dans l'autre enregistrement est supérieure à un seuil $1 - \frac{\epsilon_g}{D}$:

$$\text{Score}(k, r') = \begin{cases} 1 & \text{si } \max\left(\left\{ \max\limits_{\substack{k' \text{ composante} \\ \text{validée de } \mathcal{M}_{r',t',f'}}} \left(\left\{ \text{Sim}(k, k') \right\} \right) \right\}_{(t',f') \in \boldsymbol{\mathcal{V}}_{t,f}} \right) > 1 - \frac{\epsilon_g}{D}, \\ 0 & \text{sinon.} \end{cases} \tag{4.7}$$

L'utilisation d'un voisinage temps-fréquence se justifie ici par les décalages en temps et en fréquence possibles entre deux enregistrements. Les composantes k' des modèles $\mathcal{M}_{r',t',f'}$ considérés doivent avoir été validées localement pour être prises en compte. Dans le cas où la phase précédente de validation locale a filtré toutes les composantes de tous les modèles dans le voisinage $\boldsymbol{\mathcal{V}}_{t,f}$ de l'enregistrement r', la valeur de $\text{Score}(k, r')$ est nulle. Comme pour le seuil local ϵ_l utilisé à l'équation (4.6), le seuil global ϵ_g sert à quantifier le degré de différence acceptable entre deux groupes d'électrodes différents pour lequel ces deux groupes sont considérés comme identiques. Typiquement, pour $\epsilon_g = 1$, les groupes identiques sont ceux ayant moins d'une électrode de différence.

La mesure de stabilité globale $\text{Stab}_g(k)$ d'un groupe d'électrodes associé à une composante k est alors défini comme la somme des scores obtenus par la composante k avec tous les autres enregistrements r' :

$$\text{Stab}_g(k) = \sum_{r' \neq r} \text{Score}(k, r'). \tag{4.8}$$

L'étude de la distribution des scores dans l'ensemble des enregistrements permet alors de mettre en relief les groupes d'électrodes aux scores statistiquement plus élevés que les autres. Les groupes d'électrodes dont la mesure de stabilité globale est supérieure ou égale au 99$^{\text{e}}$ centile de la distribution empirique des mesures de stabilité sont alors considérés comme représentant des groupes d'électrodes à synchronisation cooccurrente. Les groupes d'électrodes vides ou correspondant à l'ensemble des électrodes ne sont pas pris en compte, puisqu'ils n'apportent aucune information d'intérêt pour l'analyse des synchronisations cooccurrentes.

4.2. Méthode d'analyse de synchronisations univariées par modèles de mélange ricien

La seconde méthode d'analyse que j'ai mise au point dans le cadre de mes travaux de thèse explore une autre voie pour la détection de synchronisation au sein de signaux cérébraux. L'analyse porte ici sur le signal de chaque électrode indépendamment des autres, d'où la dénomination de « synchronisation univariée ». L'hypothèse à la base de cette méthode est que l'intégralité du signal d'une électrode dans une bande de fréquence peut être classée en deux états, haut et bas. Dès lors, un phénomène de synchronisation, qui se traduit usuellement par une augmentation en moyenne du signal en un point temps-fréquence, équivaut à une forte probabilité d'apparition de l'état haut en ce point

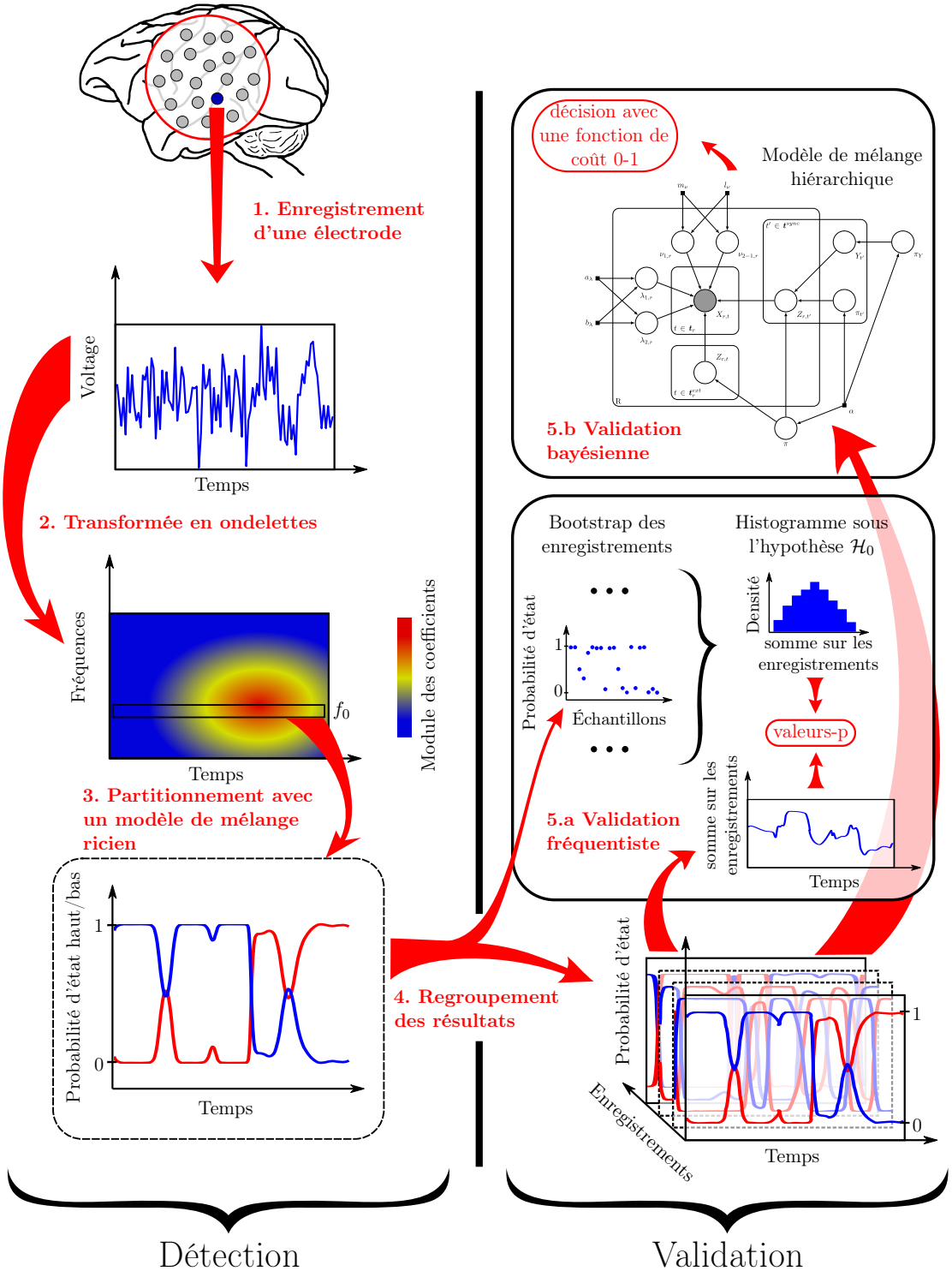

FIGURE 4.2. – Résumé graphique des étapes de la méthode d'analyse de synchronisations univariées.

temps-fréquence. Inversement, une forte probabilité d'apparition de l'état bas en un point temps-fréquence sera associée à un phénomène de désynchronisation.

Par rapport à une analyse classique de type ERSP, l'information du signal sur la période pré-stimulus est remplacée par l'information du signal sur la totalité de la période de temps entre le stimulus et la réponse. La moyenne entre les enregistrements est quant à elle remplacée par une analyse de la probabilité de niveaux hauts et bas en chaque point temporel.

L'approche générale reste la même que dans la première méthode, c'est-à-dire une approche en deux étapes, avec en premier lieu la détection de phénomènes d'intérêt dans chaque enregistrement puis la mise en commun de cette information entre tous les enregistrements pour faire émerger les phénomènes invariants en temps et en fréquence propres à l'expérience étudiée.

Dans l'étape de détection, le signal de l'électrode est projeté dans un espace temps-fréquence à l'aide d'une transformée en ondelettes continue, avec l'ondelette de Morlet complexe. Puis, pour chaque bande de fréquence, le module des coefficients d'ondelettes du signal est partitionné en deux états, haut et bas, par un modèle de mélange ricien.

L'étape de validation consiste alors à évaluer dans chaque bande de fréquence, pour chaque point temporel, si le nombre d'états hauts ou bas détectés à travers les enregistrements peut être expliqué statistiquement avec ou sans indépendance par rapport au temps. Deux possibilités ont été explorées pour effectuer ce test statistique :

- en approchant par *bootstrap* la distribution du nombre d'états hauts ou bas observés sous l'hypothèse nulle qu'il existe une indépendance vis-à-vis du temps,
- en utilisant un modèle bayésien hiérarchique qui embarque les deux hypothèses sous la forme d'un sous-modèle de mélange.

La figure 4.2 résume l'enchaînement des différentes étapes de la méthode d'analyse proposée. Ces étapes sont détaillées dans les sections suivantes.

4.2.1. Détection des états hauts/bas dans une bande de fréquence du signal d'une électrode

Tout d'abord, pour chaque enregistrement de l'électrode étudiée, les coefficients d'ondelettes du signal sont calculés à l'aide d'une transformée en ondelette continue, avec l'ondelette de Morlet complexe. Un échantillonnage géométrique est utilisé pour choisir les échelles des ondelettes filles.

Les ondelettes filles $\psi_{\tau,a}$ définies à l'échelle a analysent le signal dans une bande de fréquence centrée en $f = f_0/a$. Soit l'enregistrement r de l'électrode étudiée, l'ensemble $\mathbf{x}_{r,f}$ contenant le module des coefficients d'ondelettes de la bande de fréquence centrée en f est défini comme suit :

$$\mathbf{x}_{r,f} = \left\{ \mathrm{x}_{r,f,t} \right\}_{t \in \boldsymbol{t}_r} \quad \text{et} \quad \mathrm{x}_{r,f,t} = |\mathrm{W}s_r(t, f_0/f)|, \tag{4.9}$$

où s_r est le signal de l'électrode et l'ensemble $\boldsymbol{t}_r$ représente les points temporels du signal échantillonné définis entre le temps du stimulus $t_{stim,r}$ et le temps de la réponse $t_{resp,r}$.

Si, dans une bande de fréquence centrée en f, le signal est constitué d'une alternance aléatoire de deux sinusoïdes d'amplitudes différentes additionnées d'un bruit blanc gaussien,

alors la distribution du module de chaque coefficient d'ondelette $|\mathrm{W}s_r(t, f_0/f)|$ est un mélange ricien à deux composantes.

Dès lors, pour partitionner en deux classes $\mathbf{x}_{r,f}$, un modèle bayésien de mélange ricien à deux composantes, tel que détaillé en section 3.4, est utilisé. Il faut noter que ce modèle, même dans le cas où le signal suit le modèle à deux sinusoïdes évoqué ci-dessus, est une approximation de la distribution de $\mathbf{x}_{r,f}$: il ne prend pas en compte les dépendances entre coefficients d'ondelettes proches en temps. Le modèle associé à $\mathbf{x}_{r,f}$ est noté $\mathcal{M}_{r,f}$.

Pour calculer la probabilité a posteriori et marginale du modèle de mélange ricien $\mathcal{M}_{r,f}$, l'algorithme VMP est employé. L'initialisation du modèle se fait à l'aide de l'heuristique reposant sur la médiane des données, présentée en section 3.5.1. Il s'agit d'initialiser les deux composantes du modèle de mélange en considérant les éléments de $\mathbf{x}_{r,f}$ en dessous de la médiane comme appartenant à l'une des composantes riciennes, et les éléments supérieurs à la médiane comme appartenant à l'autre composante ricienne. De surcroît, pour tenter d'obtenir une meilleure convergence de l'algorithme VMP, le modèle de mélange est aussi initialisé avec les résultats de partitionnement obtenus dans les bandes de fréquence adjacentes. Les autres détails d'utilisation de l'algorithme VMP avec les modèles de mélange ricien sont données en annexe B.5.

Pour chaque modèle de mélange ricien $\mathcal{M}_{r,f}$ à deux composantes, la quantité d'intérêt retenue est la probabilité a posteriori approchée des variables cachées Z_t associées aux éléments de $\mathbf{x}_{r,f}$, notée $q_{r,f}(z_t)$. Les composantes riciennes de chaque modèle de mélange sont réordonnées de telle sorte que la première composante est de moyenne inférieure à la seconde. Ainsi la première composante de chaque modèle $\mathcal{M}_{r,f}$ correspond à l'état bas et la seconde composante à l'état haut. De ce fait, $q_{r,f}(z_{t,1})$ est la probabilité a posteriori que l'élément de $\mathbf{x}_{r,f}$ défini au point temporel t appartienne à l'état bas, et $q_{r,f}(z_{t,2})$ est la probabilité a posteriori que ce même élément appartienne à l'état haut.

4.2.2. Validation fréquentiste par calcul de valeur-p

À l'issue de l'étape de détection, l'information disponible sur le signal de l'électrode étudiée est donc un ensemble de probabilités d'états hauts et bas détectées en chaque point temps-fréquence de chaque enregistrement. L'étape de validation consiste à évaluer si, en chaque point temps-fréquence, le nombre moyen d'états hauts et bas détectés dans les enregistrements est significativement différent du nombre moyen d'états hauts et bas qui seraient détectés sous une hypothèse de stationnarité de ces états. Un nombre moyen significativement élevé d'états hauts et faible d'états bas correspond à un épisode de synchronisation. Inversement, un nombre moyen significativement faible d'états hauts et élevé d'états bas correspond à un épisode de désynchronisation.

Pour révéler les épisodes de synchronisation ou de désynchronisation, un test statistique est utilisé en chaque point temps-fréquence (t, f). Seule l'information des états hauts est utilisée, celle des états bas étant redondante. Ce test statistique se fonde sur les hypothèses suivantes :

$\mathcal{H}_0$: « les probabilités d'états hauts dans les enregistrements sont issues de distributions stationnaires dans chaque enregistrement »

$\mathcal{H}_1$: « les probabilités d'états hauts dans les enregistrements sont issues de distributions non stationnaires dans chaque enregistrement »

La non-stationnarité révélée en cas de rejet de l'hypothèse nulle $\mathcal{H}_0$ est considérée comme un indice de la présence de synchronisation ou désynchronisation.

La statistique associée au test est le nombre moyen d'états hauts détectés $N_{t,f}$ dans tous les enregistrements en un point temps-fréquence (t, f). La valeur observée $n_{t,f}$ de cette statistique est calculée en verrouillant les résultats de l'étape de détection sur le stimulus ou la réponse dans tous les enregistrements, puis en additionnant les probabilités a posteriori $q_{r,f}(z_{t,2})$:

$$n_{t,f} = \sum_{r=1}^{R} q_{r,f}(z_{t,2}). \tag{4.10}$$

La valeur-p bilatérale symétrique associée au test bilatéral s'écrit alors :

$$\text{valeur-p} = 2 \times \min\left(P(n_{t,f} < N_{t,f}|\mathcal{H}_0), P(n_{t,f} > N_{t,f}|\mathcal{H}_0)\right). \tag{4.11}$$

Même si la distribution de $N_{t,f}$ n'est pas connue analytiquement sous l'hypothèse $\mathcal{H}_0$, un test statistique de Monte-Carlo est réalisable s'il est possible de générer des échantillons de cette distribution [Davison and Hinkley, 1997]. Soit $n_{t,f,(i)}$ un échantillon issu de la distribution de $N_{t,f}$ sous l'hypothèse $\mathcal{H}_0$, les probabilités utilisées dans l'équation (4.11) sont approchées comme suit :

$$\begin{aligned} P(n_{t,f} < N_{t,f}|\mathcal{H}_0) &\approx \frac{1}{M}\sum_{i=1}^{M} \mathbb{1}(n_{t,f,(i)} < n_{t,f}) \\ P(n_{t,f} > N_{t,f}|\mathcal{H}_0) &\approx \frac{1}{M}\sum_{i=1}^{M} \mathbb{1}(n_{t,f,(i)} > n_{t,f}), \end{aligned} \tag{4.12}$$

où M est le nombre d'échantillons utilisés pour l'approximation et $\mathbb{1}$ la fonction indicatrice. Chaque échantillon $n_{t,f,(i)}$ est généré comme une somme d'échantillons $q_{r,f,(i)}$:

$$n_{t,f,(i)} = \sum_{r=1}^{R} q_{r,f,(i)}. \tag{4.13}$$

où les échantillons $q_{r,f,(i)}$ proviennent de la distribution de la probabilité d'états hauts dans l'enregistrement r et dans la bande de fréquence centrée en f, sous l'hypothèse de stationnarité $\mathcal{H}_0$. Sous cette hypothèse $\mathcal{H}_0$, les probabilités a posteriori $q_{r,f}(z_{t,2})$ calculées à l'étape de détection sont des observations de cette distribution de probabilité d'états hauts. Aussi la distribution empirique des $q_{r,f}(z_{t,2})$ dans l'enregistrement r et la bande de fréquence centrée en f peut être considérée comme suffisamment représentative de la vraie distribution de la probabilité d'états hauts pour que les échantillons $q_{r,t,(i)}$ soient échantillonnés par *bootstrap* à partir de l'ensemble $\boldsymbol{q}_{r,f} = \{q_{r,f}(z_{t,2})\}_{t \in \boldsymbol{t}_r}$ [Efron and Tibshirani, 1993, Davison and Hinkley, 1997]. Cet échantillonnage par bootstrap correspond à un tirage avec remise sur l'ensemble $\boldsymbol{q}_{r,f}$.

L'hypothèse nulle $\mathcal{H}_0$ est rejetée en un point temps-fréquence (t, f) si la valeur-p du test statistique calculée en ce point est inférieure à un niveau de signification α. La valeur de α reflète la probabilité de faux positifs maximale acceptée dans le cas où l'hypothèse nulle $\mathcal{H}_0$ est vraie, et est habituellement fixé à 0,05. Si l'hypothèse nulle est rejetée et la valeur-p est issue de la queue supérieure de la distribution de $N_{t,f}$ sous $\mathcal{H}_0$, alors cela signifie que le nombre moyen d'états hauts détectés est anormalement élevé, ce qui correspond à un épisode de synchronisation. Inversement, le rejet de l'hypothèse nulle avec une valeur-p issue de la queue inférieure de la distribution de $N_{t,f}$ sous $\mathcal{H}_0$ met en valeur à un nombre moyen d'états hauts détectés anormalement faible, correspondant à une désynchronisation.

Correction de tests d'hypothèses multiples

L'utilisation de tests statistiques multiples pour l'identification des épisodes de synchronisation et désynchronisation entraîne des problèmes majeurs de fausses découvertes. En effet, le seuil α utilisé pour limiter la probabilité de faux positifs implique que, dans le cas de nombreux tests statistiques et d'une hypothèse nulle vraie, un pourcentage égal à α des tests statistiques rejetteront l'hypothèse nulle à tort et seront interprétés comme des découvertes alors qu'il n'y a rien [2].

Pour éviter ce phénomène de fausse découverte dans un contexte de tests statistiques multiples, une approche conservatrice consiste à fixer la probabilité d'un faux positif sur tous les tests statistiques au lieu de fixer la probabilité α d'un faux positif dans chaque test statistique. Cette approche donne lieu à la *correction de Bonferroni*, qui remplace le seuil α par α/K dans chaque test, où K est le nombre de tests multiples corrigés [Verhoeven et al., 2005]. Cette correction est très conservatrice et se révèle inutilisable quand le nombre de tests corrigés est très grand.

Une autre approche moins conservatrice a été proposée par Benjamini et Hochberg [Benjamini and Hochberg, 1995], et contrôle le *taux de fausses découvertes* (False discovery rate ou FDR) au lieu de la probabilité de faux positifs. Le taux de fausses découvertes correspond à l'espérance de la proportion de fausses découvertes. Cette proportion est le pourcentage de faux positifs parmi toutes les découvertes, c'est-à-dire parmi tous les tests rejetant l'hypothèse nulle.

La procédure de Benjamini et Hochberg consiste à classer les valeur-p des tests statistiques par ordre croissant et à chercher le plus grand k tel que :

$$k\text{-ième valeur-p} \leq \frac{k}{K}\,\gamma, \tag{4.14}$$

avec K le nombre de tests statistiques et γ un seuil fixé. Tous les tests associés aux k plus petites valeurs-p rejettent alors l'hypothèse nulle et sont considérés comme ayant fait une découverte. Cette procédure assure que le taux de fausses découvertes est inférieur ou égal à γ. Dans la littérature, γ est conventionnellement fixé à 0,10.

2. Ainsi, en oubliant la correction de tests multiples, il est possible de trouver de fausses activations corticales dans une expérience d'IRM fonctionnel sur un cerveau de saumon mort [Bennett et al., 2009].

La procédure de Benjamini et Hochberg est valide dans le cas où les tests statistiques corrigés sont indépendants ou dépendants positivement [Benjamini and Yekutieli, 2001]. Une procédure semblable existe dans le cas d'une dépendance quelconque entre les tests statistiques [Benjamini and Yekutieli, 2001]. Néanmoins, cette dernière procédure peut se révéler plus conservatrice [Farcomeni, 2008].

Dans le cas de la validation fréquentiste des états hauts et bas détectés, la correction des tests d'hypothèses privilégiée est la procédure de Benjamini et Hochberg, de par son efficacité et sa simplicité de mise en œuvre. L'hypothèse de dépendance positive entre les tests statistiques semble raisonnable. En effet, la dépendance de ces tests statistiques provient des données sous-jacentes, c'est-à-dire en premier lieu des coefficients d'ondelettes, or ceux-ci sont corrélés positivement, comme décrit en section 3.1.2.

4.2.3. Validation bayésienne par modélisation hiérarchique

La validation fréquentiste présentée dans la section précédente utilise des tests d'hypothèses fréquentistes. Ce type de tests employant des valeurs-p, bien qu'au cœur de la méthode scientifique dans les domaines expérimentaux, est la cible de vives critiques [Berger and Berry, 1988, Johnson, 1999], les principales pouvant être formulées comme suit :

- Le test d'hypothèse est construit de sorte à rejeter l'hypothèse nulle. Il n'est pas possible de prouver l'hypothèse nulle.
- La décision est faite en considérant la probabilité d'obtenir une valeur d'une statistique plus extrême que celle observée. Ainsi, cette décision se fonde sur des données non observées.
- Le test d'hypothèse fréquentiste ne respecte pas le *principe de vraisemblance*, qui stipule que toute l'information d'un échantillon est contenue dans sa fonction de vraisemblance [Birnbaum, 1962, Robert, 2006]. Or la valeur-p d'un test dépend aussi de l'intention de l'expérimentateur, suivant la manière dont le recueil des données a eu lieu.
- La valeur-p est très souvent interprétée à tort comme la probabilité de l'hypothèse nulle conditionnée sur les données expérimentales.

De plus, le fait de fixer un niveau de signification α pour limiter la probabilité de faux positifs entraîne automatiquement une augmentation de la proportion de fausses découvertes avec le nombre de tests calculés [Farcomeni, 2008], d'où la nécessité de procédures de correction de tests d'hypothèses multiples.

Cette section décrit une autre approche pour la validation des résultats de la méthode d'analyse de synchronisations univariées, une alternative à la validation fréquentiste. Cette approche est fondée sur un modèle bayésien hiérarchique couplé à la théorie de la décision. Elle permet d'éviter les défauts de l'approche fréquentiste et a l'avantage d'utiliser un unique cadre théorique pour l'ensemble de la partie statistique de la méthode d'analyse de synchronisations univariées.

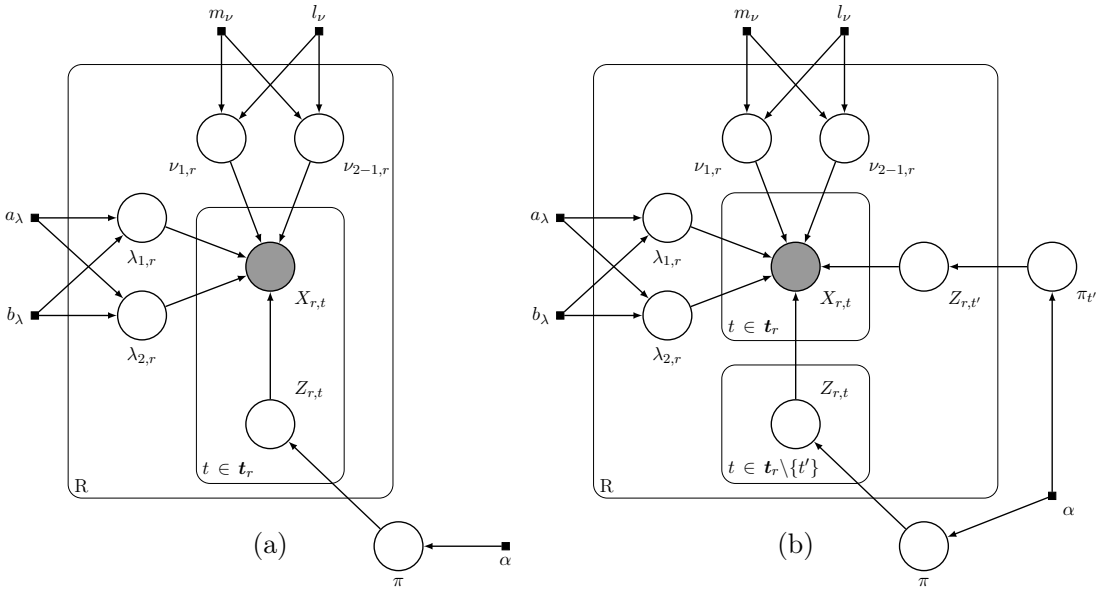

FIGURE 4.3. – Modèle graphique probabiliste représentant la distribution jointe des modèles de mélange ricien (a) $\mathcal{M}_{0,f}$ avec l'hypothèse de stationnarité des états hauts et bas, décrit par l'équation (4.21), (b) $\mathcal{M}_{1,f,t'}$ sans l'hypothèse de stationnarité au point temporel t', décrit par l'équation (4.22)

Test d'hypothèses bayésien avec une comparaison de modèles

La définition du modèle bayésien hiérarchique au cœur de l'étape de validation bayésienne passe dans un premier temps par la reformulation du test d'hypothèses fréquentiste introduit à la section 4.2.2 en un test bayésien. Il est possible de créer ce test bayésien si les hypothèses $\mathcal{H}_0$ et $\mathcal{H}_1$ du test fréquentiste peuvent chacune être formalisées sous la forme de modèles statistiques bayésiens. Dès lors, le test bayésien correspond à une simple comparaison de modèle [Robert, 2006], telle que décrite à la section 2.2.2.

L'hypothèse de stationnarité testée portant sur l'axe temporel du signal, l'étude des différentes bandes de fréquence peut être effectuée indépendamment pour chacune. Ainsi, les modèles mis au point pour correspondre aux hypothèses $\mathcal{H}_0$ et $\mathcal{H}_1$ sont associés aux données de l'électrode étudiée pour l'ensemble des enregistrements et des points temporels mais dans une seule bande de fréquence à la fois. L'ensemble $\mathbf{x}_f$ des données dans la bande de fréquence centrée en f est défini comme suit :

$$\mathbf{x}_f = \left\{\mathrm{x}_{r,f,t}\right\}_{\substack{r \in [1,R] \\ t \in \boldsymbol{t}_r}}. \tag{4.15}$$

Par souci de concision, comme les modèles bayésiens présentés dans la suite sont tous indicés par f, cet indice sera omis dans la description de leurs variables observées et cachées.

Dans le cas de l'hypothèse $\mathcal{H}_0$ de stationnarité des états hauts et bas, le modèle $\mathcal{M}_{0,f}$ associé à $\mathbf{x}_f$ est très similaire au modèle bayésien de mélange ricien à deux composantes

utilisé à l'étape de détection. Plus précisément, le module des coefficients d'ondelettes est modélisé par une distribution de mélange ricien à deux composantes dans chaque enregistrement. La densité de probabilité a priori associée à chaque donnée $\mathrm{x}_{r,t}$ de la bande de fréquence centrée en f vaut alors :

$$p(x_{r,t}|\nu_{1,r},\nu_{2-1,r},\lambda_{1,r},\lambda_{2,r},z_{r,t}) = \mathcal{R}ice(x_{r,t}|\nu_{1,r},\lambda_{1,r})^{z_{r,t,1}} \times \mathcal{R}ice(x_{r,t}|\nu_{2,r},\lambda_{2,r})^{z_{r,t,2}}. \tag{4.16}$$

Les paramètres $\nu_{1,r}$, $\nu_{2,r}$, $\lambda_{1,r}$ et $\lambda_{2,r}$ des composantes riciennes sont différents dans chacun des R enregistrements. Les probabilités a priori des variables cachées du modèle $\mathcal{M}_{0,f}$ sont donc les mêmes que celles définies pour un modèle de mélange ricien basique :

$$\begin{aligned} p(\nu_{1,r}|m_\nu,l_\nu) &= \mathcal{N}^R(\nu_{1,r}|m_\nu,l_\nu) & p(\nu_{2,r}|m_\nu,l_\nu) &= \mathcal{N}^R(\nu_{2,r}|m_\nu,l_\nu) \\ p(\lambda_{1,r}|a_\lambda,b_\lambda) &= \mathcal{G}(\lambda_{1,r}|a_\lambda,b_\lambda) & p(\lambda_{2,r}|a_\lambda,b_\lambda) &= \mathcal{G}(\lambda_{2,r}|a_\lambda,b_\lambda) \\ p(z_{r,t}|\pi) &= \pi_1^{z_{r,t,1}} \times \pi_2^{z_{r,t,2}} & p(\pi|\alpha) &= \mathcal{D}ir(\pi|\alpha). \end{aligned} \tag{4.17}$$

Le paramètre π de proportions de mélange pour les états hauts et bas est mis en commun pour tous les enregistrements, pour faciliter par la suite la création et l'interprétation du modèle associé à l'hypothèse alternative $\mathcal{H}_1$.

Le modèle $\mathcal{M}_{0,f}$ est alors un modèle exponentiel-conjugué dont les probabilités marginale et a posteriori sont calculables avec l'algorithme VMP, quand la distribution de mélange ricien est approchée par une borne inférieure variationelle, comme décrit précédemment en section 3.4.3.

Néanmoins, une adaptation de ce modèle est nécessaire pour s'assurer que la première composante ricienne dans chaque enregistrement corresponde à l'état bas et la seconde composante à l'état haut. En effet, dans le modèle $\mathcal{M}_{0,f}$ tel que défini à présent, rien ne distingue a priori l'ordre des deux composantes, qui sont donc interchangeables par simple ré-étiquetage des composantes. Il s'agit là d'une source de non-*identifiabilité* commune à tous les modèles de mélange [Bishop, 2006, Beal, 2003]. L'ordre des composantes peut être contraint en modélisant le paramètre $\nu_{2,r}$ de la seconde composante comme la somme du paramètre $\nu_{1,r}$ et d'un paramètre $\nu_{2-1,r}$ strictement positif :

$$\nu_{2,r} = \nu_{1,r} + \nu_{2-1,r} \quad \text{et} \quad \nu_{2-1,r} \geq 0. \tag{4.18}$$

Dans le cadre de l'algorithme VMP, le paramètre $\nu_{2,r}$ est représenté par un nœud dit *déterministe*, comme $\nu_{2,r}$ est une fonction déterministe des paramètres $\nu_{1,r}$ et $\nu_{2-1,r}$. Pour un nœud déterministe, un message vers un nœud enfant est une fonction des messages depuis les nœuds parents. De même, un message vers un nœud parent est une fonction des messages depuis les autres nœuds parents et enfants. Ainsi, le nœud déterministe ne fait que transformer les messages afférents en messages efférents, sans maintenir d'état interne [Winn, 2003]. Dans le cas présent d'une addition, le nœud enfant $X_{r,t}$ contraint la

forme du message de $\nu_{2,r}$ vers $X_{r,t}$ à :

$$\begin{aligned} m_{\nu_{2,r}\to X_{r,t}} &= \begin{pmatrix} \mathbb{E}_{q(\nu_{2,r})}[\nu_{2,r}] \\ \mathbb{E}_{q(\nu_{2,r})}[\nu_{2,r}^2] \end{pmatrix} = \begin{pmatrix} \mathbb{E}_{q(\nu_{2,r})}[\nu_{1,r}+\nu_{2-1,r}] \\ \mathbb{E}_{q(\nu_{2,r})}[(\nu_{1,r}+\nu_{2-1,r})^2] \end{pmatrix} \\ &= \begin{pmatrix} \mathbb{E}_{q(\nu_{1,r})}[\nu_{1,r}] + \mathbb{E}_{q(\nu_{2-1,r})}[\nu_{2-1,r}] \\ \mathbb{E}_{q(\nu_{1,r})}[\nu_{1,r}^2] + \mathbb{E}_{q(\nu_{2-1,r})}[\nu_{2-1,r}^2] + 2\,\mathbb{E}_{q(\nu_{1,r})}[\nu_{1,r}]\,\mathbb{E}_{q(\nu_{2-1,r})}[\nu_{2-1,r}] \end{pmatrix}. \end{aligned} \tag{4.19}$$

Il est possible de déduire de ce message la forme des vecteurs de statistiques naturelles des distributions a priori de $\nu_{1,r}$ et $\nu_{2-1,r}$, respectivement $\mathrm{u}_{\nu_{1,r}}(\nu_{1,r}) = (\nu_{1,r}, \nu_{1,r}^2)$ et $\mathrm{u}_{\nu_{2-1,r}}(\nu_{2-1,r}) = (\nu_{2-1,r}, \nu_{2-1,r}^2)$. Les paramètres $\nu_{1,r}$ et $\nu_{2-1,r}$ ayant une contrainte de positivité, leur distribution a priori ne peut être qu'une distribution gaussienne rectifiée, ce qui ne change pas la distribution a priori choisie précédemment pour $\nu_{1,r}$:

$$p(\nu_{1,r}|m_\nu, l_\nu) = \mathcal{N}^R(\nu_{1,r}|m_\nu, l_\nu) \quad \text{et} \quad p(\nu_{2-1,r}|m_\nu, l_\nu) \quad = \mathcal{N}^R(\nu_{2-1,r}|m_\nu, l_\nu). \tag{4.20}$$

Les autres messages employés par l'algorithme VMP pour le nœud déterministe d'addition associé à $\nu_{2,r}$ sont décrits en annexe A.8.

La probabilité jointe du modèle $\mathcal{M}_{0,f}$ rectifié se décompose dès lors de la manière suivante :

$$\begin{aligned} &p(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{\nu}_1, \boldsymbol{\nu}_{2-1}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \pi | m_\nu, l_\nu, a_\lambda, b_\lambda, \alpha) \\ &= \Pi_{r=1}^R \prod_{t \in \boldsymbol{t}_r} p(x_{r,t}|\nu_{1,r}, \nu_{2-1,r}, \lambda_{1,r}, \lambda_{2,r}, z_{r,t}) \\ &\quad p(\nu_{1,r}|m_\nu, l_\nu)\, p(\nu_{2-1,r}|m_\nu, l_\nu)\, p(\lambda_{1,r}|a_\lambda, b_\lambda)\, p(\lambda_{2,r}|a_\lambda, b_\lambda) \\ &\quad p(z_{r,t}|\pi)\, p(\pi|\alpha), \end{aligned} \tag{4.21}$$

où

$$\begin{aligned} \boldsymbol{x} &= \{x_{r,t}\}_{\substack{t \in \boldsymbol{t}_r \\ r \in [1,R]}} & \boldsymbol{\nu}_1 &= \{\nu_{1,r}\}_{r \in [1,R]} & \boldsymbol{\nu}_{2-1} &= \{\nu_{2-1,r}\}_{r \in [1,R]} \\ \boldsymbol{z} &= \{z_{r,t}\}_{\substack{t \in \boldsymbol{t}_r \\ r \in [1,R]}} & \boldsymbol{\lambda}_1 &= \{\lambda_{1,r}\}_{r \in [1,R]} & \boldsymbol{\lambda}_2 &= \{\lambda_{2,r}\}_{r \in [1,R]}. \end{aligned}$$

La figure 4.3 (a) représente sous la forme d'un modèle graphique probabiliste cette probabilité jointe.

Dans le cas de l'hypothèse $\mathcal{H}_1$, le modèle $\mathcal{M}_{1,f,t'}$ associé à celle-ci est une modification du modèle $\mathcal{M}_{0,f}$ où l'on considère qu'en un point temporel t' la probabilité d'états hauts et bas n'est pas stationnaire mais liée à ce point temporel. Le paramètre des proportions de mélange π reste commun à tous les points temporels sauf en t' où un autre paramètre $\pi_{t'}$ entre en jeu. La probabilité jointe du modèle $\mathcal{M}_{1,f,t'}$, illustrée à la figure 4.3 (b), s'écrit

alors :

$$\begin{aligned}
&p(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{\nu}_1, \boldsymbol{\nu}_{2-1}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \pi, \pi_{t'} | m_\nu, l_\nu, a_\lambda, b_\lambda, \alpha) \\
&= \Pi_{r=1}^{R} \left(\prod_{t \in \boldsymbol{t}_r} p(x_{r,t} | \nu_{1,r}, \nu_{2-1,r}, \lambda_{1,r}, \lambda_{2,r}, z_{r,t}) \right) \\
&\quad p(\nu_{1,r} | m_\nu, l_\nu)\, p(\nu_{2-1,r} | m_\nu, l_\nu)\, p(\lambda_{1,r} | a_\lambda, b_\lambda)\, p(\lambda_{2,r} | a_\lambda, b_\lambda) \\
&\quad \left(\prod_{t \in \boldsymbol{t}_r \setminus \{t'\}} p(z_{r,t} | \pi) \right) p(z_{r,t'} | \pi_{t'})\, p(\pi | \alpha)\, p(\pi_{t'} | \alpha),
\end{aligned} \tag{4.22}$$

avec les densités a priori de $z_{r,t'}$ et $\pi_{t'}$

$$p(z_{r,t'} | \pi_{t'}) = \pi_{t',1}^{z_{r,t',1}} \times \pi_{t',2}^{z_{r,t',2}} \quad \text{et} \quad p(\pi_{t'} | \alpha) = \mathcal{D}ir(\pi_{t'} | \alpha). \tag{4.23}$$

Le point temporel t' appartient à la période commune des signaux, verrouillés sur le stimulus ou la réponse. Plus formellement, si les signaux sont verrouillés sur le stimulus, alors ils partagent le même point temporel de départ : $t_{min} = t_{min,r}\ \forall r \in [1, R]$. Le point t' appartient alors à l'ensemble des points temporels communs $\boldsymbol{t}^{sync}$ si $t' \in [t_{min}, \min(\{t_{max,r}\}_{r \in [1,R]})]$.

L'étape de validation bayésienne par test d'hypothèses bayésien consiste alors, pour chaque bande de fréquence centrée en f et chaque point temporel t' de l'ensemble commun $\boldsymbol{t}^{sync}$, à :

1. calculer la probabilité marginale du modèle $\mathcal{M}_{0,f}$,
2. calculer la probabilité marginale du modèle $\mathcal{M}_{1,f,t'}$,
3. déterminer le modèle le plus probable en comparant les probabilités marginales des deux modèles,
4. – conclure à l'absence d'épisode de synchronisation ou désynchronisation au point temps-fréquence (t', f) si $\mathcal{M}_{0,f}$ est le plus probable,
   – ou conclure à la présence d'un épisode de synchronisation ou désynchronisation au point temps-fréquence (t', f) si $\mathcal{M}_{1,f,t'}$ est le plus probable.

Dans ce dernier cas, la comparaison des probabilités a posteriori des paramètres π et $\pi_{t'}$ pour le modèle $\mathcal{M}_{1,f,t'}$ permet de distinguer un épisode de synchronisation d'un épisode de désynchronisation. Si la proportion de mélange stationnaire des états hauts π_2 est estimée a posteriori inférieure à la proportion de mélange au point t' des états hauts $\pi_{t',2}$, soit

$$\mathbb{E}_{q(\pi)}[\pi_2] < \mathbb{E}_{q(\pi_{t'})}[\pi_{t',2}], \tag{4.24}$$

alors il est possible de conclure à un épisode de synchronisation, et inversement pour un épisode de désynchronisation.

En pratique, cette approche d'une validation avec des tests d'hypothèses bayésiens n'est pas envisageable. En effet, le calcul des probabilités marginales des modèles avec l'algorithme VMP ne permet d'obtenir qu'une borne inférieure, laquelle est biaisée vers les modèles plus simples [Beal, 2003], et donc risque de favoriser systématiquement $\mathcal{M}_{0,f}$. Pour remédier à ce problème, il est possible de combiner les deux modèles en un unique modèle bayésien hiérarchique, comportant un paramètre pour décider du modèle effectif.

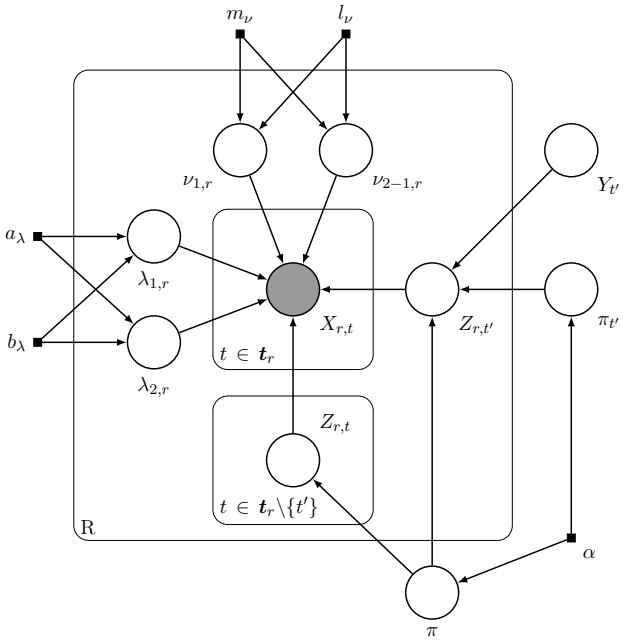

FIGURE 4.4. – Modèle graphique probabiliste représentant la distribution jointe du modèle hiérarchique $\mathcal{M}_{f,t'}$ qui combine les modèles $\mathcal{M}_{0,f}$ et $\mathcal{M}_{1,f,t'}$, décrit par l'équation (4.29)

Transformation en modèle de mélange hiérarchique

Les deux modèles bayésiens $\mathcal{M}_{0,f}$ et $\mathcal{M}_{1,f,t'}$ sont définis par des probabilités jointes très proches, dont l'unique élément qui diffère est la probabilité a priori de la variable cachée $Z_{r,t'}$. Pour combiner les deux modèles en un modèle $\mathcal{M}_{f,t'}$, il suffit alors de remplacer la distribution a priori de $Z_{r,t'}$ par une distribution de mélange catégorique à deux composantes et d'introduire une variable $Y_{t'}$ servant à discriminer les deux cas :

$$p(z_{r,t'}|\pi, \pi_{t'}, y_{t'}) = (\pi_{t',1}^{z_{r,t',1}} \times \pi_{t',2}^{z_{r,t',2}})^{y_{t',1}} \ (\pi_1^{z_{r,t',1}} \times \pi_2^{z_{r,t',2}})^{y_{t',2}}. \tag{4.25}$$

En pratique, la distribution a priori de $Z_{r,t'}$ est une distribution catégorique normale mais ayant un paramètre de proportions de mélange $\kappa_{t'}$ qui est une fonction déterministe de sélection dépendant de π, $\pi_{t'}$ et $Y_{t'}$:

$$p(z_{r,t'}|\kappa_{t',k}) = \kappa_{t',1}^{z_{r,t',1}} \times \kappa_{t',2}^{z_{r,t',2}} \quad \text{et} \quad \kappa_{t',k} = (\pi_{t',k} \times y_{t',1}) + (\pi_k \times y_{t',2}) \ \forall k \in \{1,2\}. \tag{4.26}$$

Cette modification mineure simplifie l'implémentation du modèle mais ne change pas ses propriétés. Les messages utilisés par l'algorithme VMP pour un nœud associé à une fonction déterministe de sélection sont détaillés à l'annexe A.7.

La variable $Y_{t'}$ est un vecteur aléatoire à deux éléments, $Y_{t'} = (Y_{t',1}, Y_{t',2})$. Pour un tirage aléatoire de $Y_{t'}$, un seul des éléments du vecteur est égal à un et l'autre vaut zéro.

La probabilité a priori de $Y_{t'}$ est choisie pour assurer une équiprobabilité entre les deux hypothèses :

$$p(y_{t'}) = \left(\frac{1}{2}\right)^{y_{t',1}} \times \left(\frac{1}{2}\right)^{y_{t',2}}. \tag{4.27}$$

À partir de la probabilité a posteriori de la variable $Y_{t'}$, il est possible d'obtenir une estimation ponctuelle de la valeur de cette variable. Avec une fonction de coût 0-1, l'estimateur de la valeur de $Y_{t'}$ minimisant le coût moyen a posteriori est l'estimateur MAP :

$$y^*_{t'} = \underset{y_{t'}}{\operatorname{argmax}}\ p(y_{t'}|\boldsymbol{x}) \approx \underset{y_{t'}}{\operatorname{argmax}}\ q(y_{t'}). \tag{4.28}$$

Si $y^*_{t'}$ vaut $(1,0)$, alors le modèle $\mathcal{M}_{f,t'}$ privilégie l'hypothèse de non-stationnarité au point temporel t', ce qui correspond à un épisode de synchronisation ou désynchronisation. Inversement, l'hypothèse de stationnarité est privilégiée si $y^*_{t'}$ vaut $(0,1)$. Comme précédemment, la comparaison des estimations ponctuelles de $\pi_{t'}$ et π permet de discriminer un épisode de synchronisation d'un épisode de désynchronisation.

La probabilité jointe du modèle $\mathcal{M}_{f,t'}$ est définie de la manière suivante :

$$\begin{aligned}
&p(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{\nu}_1, \boldsymbol{\nu}_{2-1}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \pi, \pi_{t'}, y_{t'}|m_\nu, l_\nu, a_\lambda, b_\lambda, \alpha) \\
&= \Pi_{r=1}^{R} \left(\prod_{t \in \boldsymbol{t}_r} p(x_{r,t}|\nu_{1,r}, \nu_{2-1,r}, \lambda_{1,r}, \lambda_{2,r}, z_{r,t}) \right) \\
&\quad p(\nu_{1,r}|m_\nu, l_\nu)\, p(\nu_{2-1,r}|m_\nu, l_\nu)\, p(\lambda_{1,r}|a_\lambda, b_\lambda)\, p(\lambda_{2,r}|a_\lambda, b_\lambda) \\
&\quad \left(\prod_{t \in \boldsymbol{t_r} \setminus \{t'\}} p(z_{r,t}|\pi) \right) p(z_{r,t'}|\pi, \pi_{t'}, y_{t'})\, p(\pi|\alpha)\, p(\pi_{t'}|\alpha)\, p(y_{t'}),
\end{aligned} \tag{4.29}$$

et est représentée graphiquement à la figure 4.4.

Une dernière amélioration du modèle de mélange hiérarchique consiste à tester tous les points temporels de $\boldsymbol{t}^{sync}$ dans un même modèle. Par conséquent, il n'y a plus à évaluer qu'un seul modèle $\mathcal{M}_f$ par bande de fréquence, pour tester simultanément la présence d'épisodes de synchronisation et désynchronisation sur tous ces points temporels. La probabilité jointe du modèle $\mathcal{M}_f$ s'écrit comme suit :

$$\begin{aligned}
&p(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{\nu}_1, \boldsymbol{\nu}_{2-1}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \pi, \pi_{t'}, \boldsymbol{y}, \pi_Y|m_\nu, l_\nu, a_\lambda, b_\lambda, \alpha) \\
&= \Pi_{r=1}^{R} \left(\prod_{t \in \boldsymbol{t}_r} p(x_{r,t}|\nu_{1,r}, \nu_{2-1,r}, \lambda_{1,r}, \lambda_{2,r}, z_{r,t}) \right) \\
&\quad p(\nu_{1,r}|m_\nu, l_\nu)\, p(\nu_{2-1,r}|m_\nu, l_\nu)\, p(\lambda_{1,r}|a_\lambda, b_\lambda)\, p(\lambda_{2,r}|a_\lambda, b_\lambda) \\
&\quad \left(\prod_{t' \in \boldsymbol{t}^{sync}} p(z_{r,t'}|y_{t'}, \pi, \pi_{t'})\, p(y_{t'}|\pi_Y)\, p(\pi_{t'}|\alpha) \right) p(\pi_Y|\alpha) \\
&\quad \left(\prod_{t \in \boldsymbol{t}_r^{ext}} p(z_{r,t}|\pi) \right) p(\pi|\alpha),
\end{aligned} \tag{4.30}$$

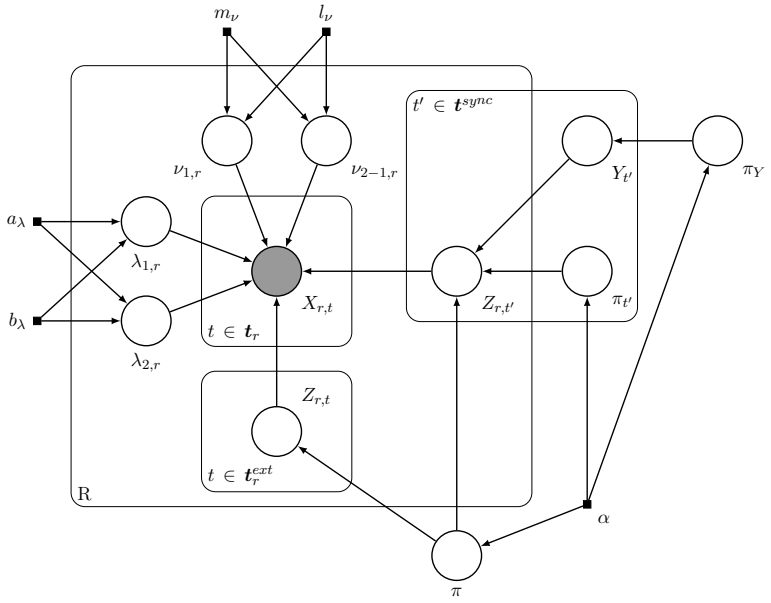

FIGURE 4.5. – Modèle graphique probabiliste représentant la distribution jointe du modèle hiérarchique $\mathcal{M}_f$ décrit par l'équation (4.30)

où $\boldsymbol{y} = \{y_{t'}\}_{t' \in \boldsymbol{t}^{sync}}$, et $\boldsymbol{t}_r^{ext} = \boldsymbol{t}_r \backslash \boldsymbol{t}^{sync}$. Cette probabilité jointe est représentée à la figure 4.5. Le fait que plusieurs variables $Y_{t'}$ soient présentes en même temps dans le modèle rend intéressant l'ajout d'une variable π_Y comme paramètre de la distribution a priori de $Y_{t'}$:

$$p(y_{t'}) = \pi_{Y,1}^{y_{t'},1} \times \pi_{Y,2}^{y_{t'},2} \quad \text{et} \quad p(\pi_Y|\alpha) = \mathcal{D}ir(\pi_Y|\alpha). \tag{4.31}$$

La variable π_Y représente les proportions de points temporels avec et sans épisodes de synchronisation ou désynchronisation. En introduisant ce paramètre dans le modèle, les probabilités a posteriori des variables $Y_{t'}$ sont automatiquement pénalisées suivant la quantité inférée d'épisodes temporels avec et sans épisodes de synchronisation ou désynchronisaton. Cette pénalisation embarquée dans le modèle est une technique pour compenser une situation de tests multiples dans un cadre bayésien [Scott and Berger, 2006].

Le modèle final $\mathcal{M}_f$ étant complexe, le calcul de sa probabilité a posteriori avec l'algorithme VMP avec des initialisations aléatoires n'est pas envisageable. Pour l'initialisation, les résultats obtenus avec les modèles de mélange ricien à deux composantes à l'étape de détection de la méthode sont utilisés. Les valeurs des hyperparamètres du modèle ainsi que l'analyse de la convergence du modèle sont détaillées en annexe B.6.

4.3. Résumé

Ce chapitre a présenté deux méthodes d'analyse de synchronisations corticales, qui reposent sur des hypothèses alternatives à celles employées dans les méthodes standards. Plus précisément, les deux méthodes reposent sur une étape d'analyse individuelle du scalogramme de chaque enregistrement pour détecter des phénomènes d'intérêt, suivie d'une étape de validation statistique en comparant les résultats entre les enregistrements. Les deux méthodes mettent en œuvre des modèles bayésiens de mélange à l'étape de détection, pour partitionner les données.

La première méthode d'analyse développée recherche des sous-ensembles d'électrodes présents aux mêmes points temps-fréquence dans les différents enregistrements. Ainsi, l'étape de détection de cette méthode consiste à partitionner les valeurs du scalogramme des électrodes en chaque point temps-fréquence avec un modèle bayésien de mélange gaussien, pour séparer les électrodes en sous-groupes. L'étape de validation met alors en jeu une mesure de stabilité pour vérifier que les sous-groupes identifiés sont récurrents à travers les enregistrements. Un sous-groupe d'électrodes validé de cette manière correspond à un épisode de synchronisations cooccurrentes dans ces électrodes, c'est-à-dire qu'un phénomène de synchronisation ou de désynchronisation a lieu au même point temps-fréquence dans les électrodes du sous-groupe. Dans cette méthode, c'est l'information de l'activité simultanée de toutes les électrodes qui permet d'isoler des phénomènes d'intérêt.

Cette méthode d'analyse a fait l'objet d'une communication à la conférence Neurocomp 2010 [Rio et al., 2010] et d'un article dans le Journal of Physiology – Paris [Rio et al., 2011].

La seconde méthode d'analyse développée étudie individuellement le scalogramme de chaque électrode. L'étape de détection de cette méthode est un partitionnement du module des coefficients d'ondelette dans chaque bande de fréquence de chaque enregistrement, avec un modèle de mélange ricien à deux composantes. Le module de chaque coefficient d'ondelette est étiqueté avec un état haut ou bas, suivant la composante ricienne qui lui est associée dans le modèle de mélange. L'étape de validation qui suit sert à vérifier si en chaque point temps-fréquence le nombre d'états hauts ou bas détectés dans tous les enregistrements peut être expliqué statistiquement avec une hypothèse de stationnarité. Si l'hypothèse de stationnarité est rejetée en un point temps-fréquence, alors on considère qu'un épisode de synchronisation ou de désynchronisation y est présent. Deux approches différentes ont été proposées pour exécuter la validation, soit avec des tests d'hypothèses fréquentistes, soit avec un modèle bayésien hiérarchique. Dans ce dernier cas, l'intégralité de la méthode d'analyse est développée dans le cadre de l'inférence bayésienne.

Chapitre 5.

Applications

> Ce n'est qu'en essayant continuellement que l'on finit par réussir. Autrement dit : plus ça rate, plus on a de chances que ça marche.
>
> *devise Shadok*

Le chapitre précédent a introduit deux méthodes d'analyse de synchronisation fondées sur des modèles de mélanges. Le présent chapitre regroupe les applications de ces méthodes sur des données artificielles et expérimentales.

Une première partie sera dédiée à une présentation synthétique des méthodes utilisées dans ce chapitre. Dans une seconde partie, les performances de détection de ces méthodes seront évaluées à l'aide de données artificielles dont les caractéristiques sont connues. Ces données ont pour but de mettre en relief les conditions dans lesquelles les méthodes d'analyses mises au point détectent des épisodes de synchronisation. La troisième partie sera consacrée à l'application des méthodes d'analyse à des données expérimentales. Ces données sont issues d'enregistrements de potentiels de champs locaux mesurés pendant une tâche visuo-motrice sur des singes. Enfin, la dernière partie de ce chapitre sera consacrée à une discussion générale des résultats des méthodes.

5.1. Méthodes employées

Les méthodes utilisées pour analyser les données artificielles et expérimentales regroupent des méthodes conventionnelles de type ERSP, discutées au chapitre 1, ainsi que mes méthodes, présentées au chapitre 4.

5.1.1. Méthodes ERSP

Il existe de nombreuses variantes des méthodes de type ERSP pour étudier les synchronisations au sein de signaux cérébraux. Grandchamp et Delorme ont publié une étude détaillée de ses variantes avec une évaluation de la robustesse de ces méthodes face à des signaux très bruités [Grandchamp and Delorme, 2011]. De cette étude, j'ai retenu pour mes évaluations une méthode historique et la méthode la plus robuste face au bruit.

Comme présenté à la section 1.1.4, toutes les méthodes ERSP se fondent sur une moyenne de la densité d'énergie temps-fréquence des enregistrements d'une électrode. Elles utilisent originellement le spectrogramme comme estimation de la densité d'énergie,

mais peuvent être utilisées sans modification sur le scalogramme d'un signal [Pfurtscheller and Lopes da Silva, 1999]. J'ai choisi cette dernière option pour permettre une comparaison plus aisée avec mes méthodes d'analyse. La moyenne calculée en chaque point temps-fréquence, en sommant les scalogrammes des différents enregistrements, est réalisée après verrouillage de tous les enregistrements par rapport au stimulus ou à la réponse. Ainsi, le scalogramme moyen $\overline{|\mathrm{W}s(t,f_0/f)|^2}$ du signal s d'une électrode est défini au point temps-fréquence (t,f) comme suit :

$$\overline{|\mathrm{W}s(t,f_0/f)|^2} = \frac{1}{R}\sum_{r=1}^{R}|\mathrm{W}s_r(t,f_0/f)|^2, \tag{5.1}$$

où s_r est le signal de l'électrode à l'enregistrement r et R le nombre total d'enregistrements. Les méthodes ERSP retenues se fondent alors sur des modèles différents pour les épisodes de synchronisation et désynchronisation, ce qui amène à différentes méthodes de calcul des ERSP.

Modèle multiplicatif : Ce modèle est un modèle à gain, où les épisodes de synchronisation et désynchronisation multiplient l'activité de base. Dès lors, les ERSP se calculent en divisant le scalogramme par l'activité pré-stimulus moyenne :

$$\mathrm{ERSP}_{\%}(t,f) = \frac{\overline{|\mathrm{W}s(t,f_0/f)|^2}}{\mu_B(f)} \tag{5.2}$$

avec

$$\mu_B(f) = \frac{1}{R \times \mathrm{card}(\boldsymbol{t}_B)}\sum_{r=1}^{R}\sum_{t\in\boldsymbol{t}_B}|\mathrm{W}s_r(t,f_0/f)|^2, \tag{5.3}$$

où $\boldsymbol{t}_B$ est l'ensemble des points temporels de la période pré-stimulus.

Modèle multiplicatif avec normalisation simple essai : Ce modèle est très similaire au modèle multiplicatif simple, mais considère en plus un gain possible entre chaque enregistrement. Par conséquent chaque bande de fréquence du scalogramme de chaque enregistrement est normalisée en divisant par la moyenne sur toute la période du signal, avant le calcul du scalogramme moyen. L'équation (5.1) est remplacée par

$$\overline{|\mathrm{W}s(t,f_0/f)|^2}_{\mathrm{FT}} = \frac{1}{R}\sum_{r=1}^{R}\frac{|\mathrm{W}s_r(t,f_0/f)|^2}{\mu_r(f)}, \tag{5.4}$$

avec

$$\mu_r(f) = \frac{1}{\mathrm{card}(\boldsymbol{t}_r)}\sum_{t\in\boldsymbol{t}_r}|\mathrm{W}s_r(t,f_0/f)|^2, \tag{5.5}$$

où $\boldsymbol{t}_r$ est l'ensemble des points temporels de l'enregistrement r. Les ERSP sont alors calculés de la même manière que dans le cas du modèle multiplicatif simple, en divisant le scalogramme moyen par l'activité pré-stimulus moyenne :

$$\mathrm{ERSP}_{\mathrm{FT}}(t,f) = \frac{\overline{|\mathrm{W}s(t,f_0/f)|^2}_{\mathrm{FT}}}{\mu_B(f)}. \tag{5.6}$$

Cette dernière méthode est particulièrement robuste dans le cas où certains enregistrements sont contaminés avec un haut niveau de bruit [Grandchamp and Delorme, 2011].

Les épisodes de synchronisation et désynchronisation sont détectés avec un test statistique, comme des points temps-fréquence significatifs. L'hypothèse nulle de ce test est que les ERSP calculés entre le stimulus et la réponse sont issus de la même distribution que des ERSP dus à l'activité de base. La distribution des ERSP de l'activité de base est approchée par une distribution empirique obtenue par rééchantillonnage. Les échantillons sont générés en calculant des ERSP sur la période pré-stimulus après permutation aléatoire en temps des valeurs des scalogrammes des enregistrements dans cette période pré-stimulus.

Chaque point temps-fréquence du signal post-stimulus est alors considéré comme significatif si la valeur-p d'égales queues calculée avec la distribution empirique pour l'ERSP du point temps-fréquence est inférieure à 5 %. Il s'agit d'un épisode de synchronisation ou désynchronisation en fonction de la queue utilisée pour le calcul de la valeur-p.

Correction de tests d'hypothèses multiples

Comme il y a un test statistique par point temps-fréquence étudié dans le signal entre le stimulus et la réponse, une correction de tests multiples peut être utilisée. Ainsi la procédure de Benjamini et Hochberg, présentée en section 4.2.2, est utilisée pour limiter le taux de fausses découvertes. La procédure de Benjamini et Yekutieli, plus conservatrice, peut aussi être utilisée, pour éviter de faire l'hypothèse que les tests statistiques sont indépendants ou seulement dépendants positivement. Ces deux procédures sont respectivement notées « B-H » et « B-Y » dans les résultats. Dans les deux cas, le seuil γ est fixé à 0,10.

Restriction de la période pré-stimulus

Dans certaines expériences, le nombre de points temporels disponibles dans la période pré-stimulus a été volontairement réduit pour étudier la sensibilité des méthodes ERSP à ce paramètre. Dans ce cas, pour la présentation des résultats, le nombre de points utilisés est indiqué entre parenthèses, par exemple « $\mathrm{ERSP}_{\mathrm{FT}}$ (100) ».

Synchronisations cooccurrentes

Les méthodes de type ERSP détectent des épisodes de synchronisation ou de désynchronisation au sein du signal de chaque électrode, indépendamment les unes des autres.

Aussi, pour détecter des épisodes de synchronisation cooccurrente, à la manière de ma méthode présentée en partie 4.1, il faut effectuer un traitement sur les résultats des méthodes de type ERSP pour en extraire des sous-ensembles d'électrodes à synchronisation cooccurrente.

Ce traitement consiste, en chaque point temps-fréquence, à regrouper dans un même sous-ensemble les électrodes présentant un épisode de synchronisation, dans un autre sous-ensemble les électrodes présentant un épisode de désynchronisation, et dans un dernier sous-ensemble les électrodes pour lesquelles aucun événement significatif n'a été détecté à ce point temps-fréquence. Des épisodes de synchronisation cooccurrente ne sont détectés que si plusieurs sous-ensembles d'électrodes coexistent en un même point temps-fréquence. Dans le cas où toutes les électrodes appartiennent au même ensemble, pour un point temps-fréquence, aucun épisode de synchronisation cooccurrente n'est détecté.

5.1.2. Méthode d'analyse de synchronisations cooccurrentes

La première méthode développée qui a été évaluée est celle servant à détecter des épisodes de synchronisation cooccurrente, introduite en section 4.1. Dans la suite de ce document, cette méthode sera désignée par l'abréviation « GMM ». Pour rappel, cette méthode se décompose en deux phases : une phase de détection et une phase de validation.

La phase de détection utilise la transformée en ondelettes de Morlet continue, avec un échantillonnage linéaire des fréquences. Des sous-ensembles d'électrodes à synchronisation cooccurrente sont détectés en chaque point temps-fréquence de chaque enregistrement à l'aide de modèles de mélange gaussien. En chaque point temps-fréquence de chaque enregistrement, un modèle est appliqué aux valeurs de la densité d'énergie du signal des électrodes. Les paramètres standards ainsi que les conditions de calcul avec l'algorithme VMP des modèles de mélange gaussien sont détaillés en annexe B.3. Les données étudiées comportant un faible nombre d'électrodes, le nombre de composantes maximum de chaque modèle est fixé à $K = 3$.

Une seconde passe peut être utilisée pour améliorer la convergence des modèles de mélange, en utilisant les résultats des points temps-fréquence adjacents. Le nombre maximum d'itérations est fixé à 30. À chaque itération, en chaque point temps-fréquence de chaque enregistrement, le modèle de mélange gaussien initialisé avec les modèles adjacents est calculé avec un maximum de 10 000 pas de calculs, et la convergence est considérée comme atteinte si l'accroissement de la borne inférieure du modèle est inférieure à 10^{-10}. L'utilisation de cette seconde passe d'optimisation est notée « $\mathrm{GMM}_{\mathrm{2nd}}$ ».

La seconde phase, de validation, comporte une étape de validation locale et globale, à l'aide de mesures de stabilité des sous-ensembles d'électrodes détectés. Les seuils utilisés pour le calcul des mesures de stabilité des sous-ensembles d'électrodes dans ces deux étapes sont fixés à $\epsilon_l = 1$ et $\epsilon_g = 1$. Les groupes d'électrodes retenus à l'issue de l'étape de validation globale sont ceux dont la mesure de stabilité globale est supérieure ou égale au 99$^{\mathrm{e}}$ centile de la distribution empirique des mesures de stabilité globale.

5.1.3. Méthode d'analyse de synchronisations univariées

La seconde méthode développée sert à détecter des épisodes de synchronisation et désynchronisation univariées, c'est-à-dire indépendamment au sein du signal de chaque électrode. Cette méthode est détaillée en section 4.2. Elle sera désignée dans la suite du document par l'abréviation « RMM ». Comme pour la précédente méthode, cette méthode se décompose en une phase de détection et une phase de validation.

Dans la phase de détection, la densité d'énergie temps-fréquence du signal de chaque électrode dans chaque enregistrement est calculée avec la transformée en ondelettes de Morlet continue, avec un échantillonnage géométrique des fréquences. Pour chaque électrode et chaque enregistrement, un modèle de mélange ricien à deux composantes est utilisé sur le module des coefficients d'ondelettes. Les paramètres standards et les conditions de calcul avec l'algorithme VMP des modèles de mélange ricien sont donnés en annexe B.5.

L'utilisation d'une seconde passe pour améliorer la convergence des modèles de mélange est signifiée par la notation « RMM_{2nd} ». Le nombre maximum d'itérations est fixé à 30. À chaque itération, pour chaque modèle de mélange, les résultats des modèles des fréquences adjacentes du même enregistrement et de la même électrode sont utilisés pour initialiser le modèle. Le nombre de pas de calcul maximum est fixé à 10 000, et l'accroissement maximum de la borne inférieure du modèle correspondant à la convergence est fixé à 10^{-10}.

Pour la phase de validation, deux approches sont possibles : une validation fréquentiste et une validation bayésienne. La validation fréquentiste nécessite une distribution stationnaire calculée par bootstrap, avec 1 000 000 d'échantillons. Pour les tests d'hypothèse, le niveau de signification α est fixé à 0,05. Une correction de tests multiples peut aussi être employée, avec la procédure de Benjamini et Hochberg ou la procédure de Benjamini et Yekutieli, respectivement notées « B-H » et « B-Y ». Le seuil γ est alors fixé à 0,10. Concernant la validation bayésienne, celle-ci met en jeu le modèle hiérarchique décrit en section 4.2.3. Les paramètres du modèle ainsi que les conditions de calcul avec l'algorithme VMP sont décrits en annexe B.6. L'emploi de la validation bayésienne est notée « Bayes ».

5.2. Données artificielles

Afin d'évaluer les performances des méthodes d'analyse que j'ai développées, celles-ci ont été testées sur des jeux de données artificielles dont le contenu en terme d'épisodes de synchronisation est connu à l'avance. Le type d'épisodes de synchronisation détectés par chacune des deux méthodes d'analyse n'est pas exactement le même, aussi deux jeux de données spécifiques ont été créés pour refléter cette différence.

5.2.1. Données multivariées

Pour tester la détection d'épisodes de synchronisation cooccurrente, un premier jeu de données dites « multivariées » a été mis au point. Celui-ci simule de manière simple

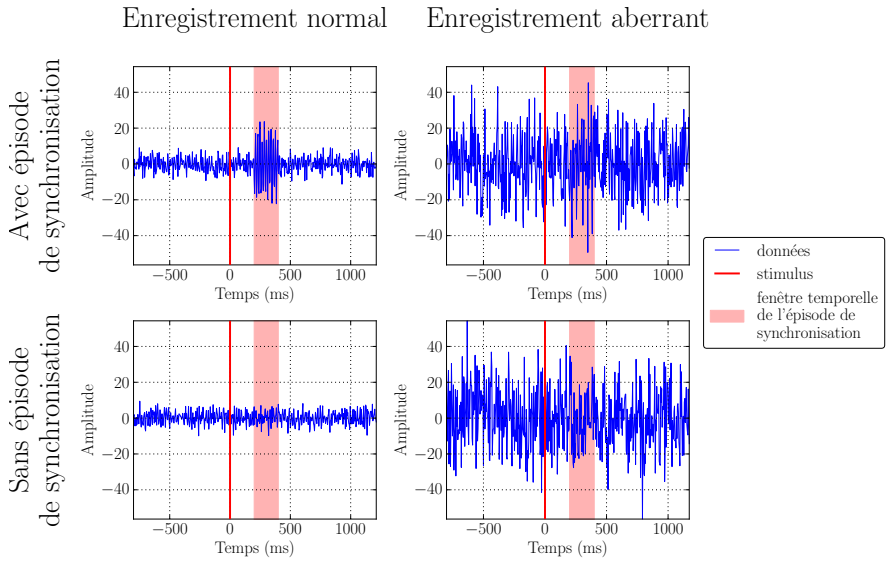

FIGURE 5.1. – Exemple de données artificielles utilisées pour l'évaluation de la détection de synchronisations cooccurrentes. Voir le texte pour les détails de génération des données.

des enregistrements électro-corticaux simultanés mettant en jeu plusieurs électrodes, et présentant un épisode de synchronisation cooccurrente.

Présentation du jeu de données

Ce jeu de données artificielles comporte 20 enregistrements, d'une durée aléatoire comprise entre 1 920 ms et 2 080 ms, soit de 480 à 520 échantillons pour un échantillonnage à 250 Hz. Dans chaque enregistrement, le temps d'apparition du stimulus est fixé à $t_s = 800$ ms. L'utilisation d'une durée aléatoire reflète les durées expérimentales variables entre un stimulus et la réponse du sujet pour une même tâche donnée.

Chaque enregistrement se compose des signaux générés pour 25 électrodes. Le signal de chaque électrode est constitué d'une sinusoïde à 55 Hz, d'amplitude 1. Les 7 premières électrodes présentent un épisode de synchronisation cooccurrente 200 ms après le stimulus, d'une durée de 200 ms. Cet épisode se traduit par un facteur multiplicatif de 6 appliqué à la sinusoïde de base. Un bruit blanc gaussien additif d'écart-type 1 est ajouté au signal de chaque électrode.

Afin de représenter la variabilité d'amplitude entre les enregistrements, un coefficient multiplicateur supplémentaire est ajouté à l'ensemble des signaux de chaque enregistrement. Ce coefficient est tiré aléatoirement dans l'intervalle $[1; 3]$, pour chaque enregistrement.

Pour évaluer la robustesse des méthodes d'analyse, certains enregistrements sont contaminés avec un bruit additionnel élevé. Un bruit blanc gaussien additif dont l'écart-type vaut 5 fois l'écart-type de l'ensemble des données simulées est ajouté aux signaux

d'un nombre fixé d'enregistrements. Ceci sert à refléter des enregistrements dont la mesure est aberrante, en comparaison des autres enregistrements.

La figure 5.1 illustre les différents types de signaux constituant le jeu de données artificielles.

Critères d'évaluation

À partir du jeu de données décrit ci-dessus, la qualité de détection des synchronisations cooccurrentes d'une methode est appréciée comme suit :

- La méthode doit détecter deux sous-ensembles d'électrodes aux points temporels correspondant à l'épisode de synchronisation. Le premier sous-groupe d'électrodes se compose des 7 premières électrodes, tandis que le second se compose des 18 électrodes restantes.
- La méthode ne doit détecter aucun phénomène en dehors de la période temporelle de l'épisode de synchronisation.

La qualité de détection n'est estimée que dans la bande de fréquence de la transformée en ondelettes de Morlet centrée en $f_0 = 55\,\mathrm{Hz}$. L'intervalle temporel étudié est délimité par le stimulus et le temps de réponse le plus court parmi tous les enregistrements.

Dans ce contexte, la détection des synchronisations cooccurrentes correspond à un problème de classification à plusieurs classes. En effet, chaque méthode d'analyse attribue aux différents points temporels étudiés l'une des trois classes suivantes :

- C_{rien}, qui traduit l'absence d'épisode de synchronisation cooccurrente,
- C_{sync}, correspondant à la présence de deux groupes d'électrodes, comprenant respectivement les 7 premières électrodes et les 18 autres,
- C_{autre}, pour tout autre ensemble de groupes d'électrodes.

Ce problème de classification à trois classes est reformulé comme trois problèmes de classification binaires, un pour chaque classe. Cette approche permet d'utiliser les outils standards d'évaluation de classifieurs binaires [Fawcett, 2006]. Pour chaque problème de classification binaire, les nombres de vrais positifs (VP_{classe}), faux positifs (FP_{classe}), vrais négatifs (VN_{classe}) et faux négatifs (FN_{classe}) obtenus par une méthode sur le jeu de données sont calculés par rapport à la classe spécifique du problème binaire. La *sensibilité* est définie alors comme la proportion de vrais positifs parmi l'ensemble des positifs [Fawcett, 2006, Grandchamp and Delorme, 2011] :

$$S_{classe} = \frac{\mathrm{VP}_{classe}}{\mathrm{VP}_{classe} + \mathrm{FN}_{classe}}. \tag{5.7}$$

Cette dernière mesure correspond à la capacité d'une méthode à détecter une classe, indépendamment de la fréquence de cette classe. La classe C_{autre} étant absente du jeu de données, seules les sensibilités S_{rien} et S_{sync} des classes C_{rien} et C_{sync} sont calculées [1].

La performance d'une méthode est aussi évaluée vis-à-vis de sa proportion de fausses découvertes. Dans le cas d'un problème binaire, cette proportion est le rapport du nombre

1. Il est aussi possible de calculer la *spécificité*, ou proportion de vrais négatifs, pour chaque classe. Cette mesure n'est pas utilisée pour ce jeu de données car elle n'apporte que peu d'information par rapport aux sensibilités calculées pour chaque classe.

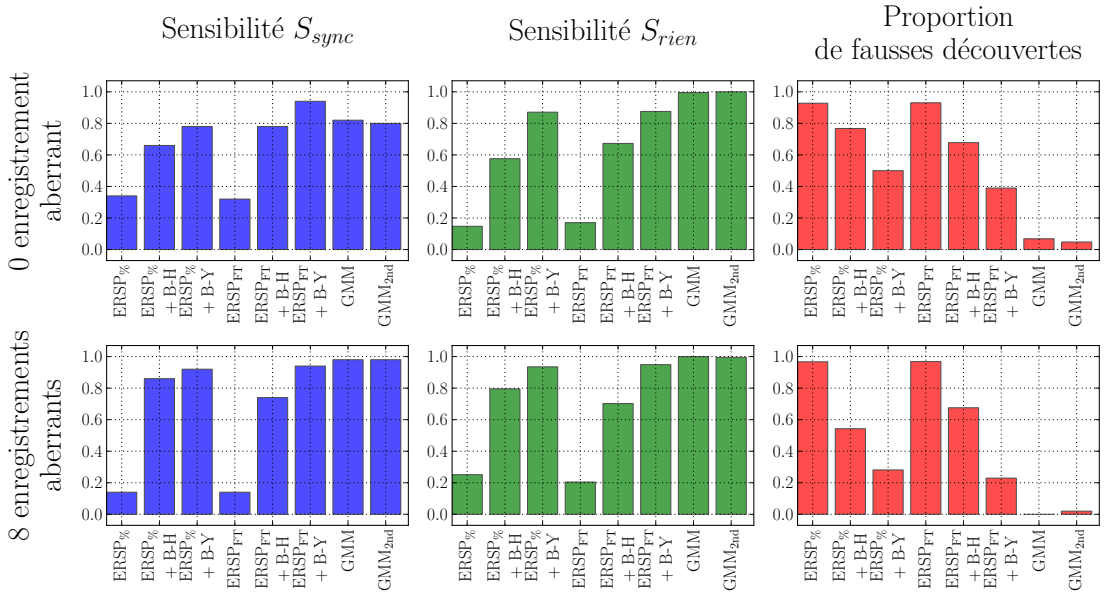

FIGURE 5.2. – Comparaison des performances pour la détection de synchronisations cooccurrentes des méthodes conventionnelles ERSP et de mes méthodes d'analyse de synchronisations cooccurrentes, sur le jeu de données artificielles multivariées. Les abréviations sont décrites à la section 5.1.

de faux positifs sur le nombre total de découvertes. Dans le cas présent, à trois classes, cette mesure est adaptée suivant la signification des classes. La détection des classes C_{sync} et C_{autre} correspond à une découverte, mais pour le jeu de données présent, seule la détection de la classe C_{sync} dans le bon intervalle de temps est une vraie découverte. Ainsi la proportion de fausses découvertes s'écrit :

$$\mathrm{PFD} = \begin{cases} 0 & \text{si} \quad \mathrm{VP}_{sync} + \mathrm{FD} = 0 \\ \frac{\mathrm{FD}}{\mathrm{VP}_{sync} + \mathrm{FD}} & \text{sinon.} \end{cases}, \tag{5.8}$$

où le nombre de fausses découvertes FD vaut

$$\mathrm{FD} = \mathrm{FP}_{sync} + \mathrm{VP}_{autre} + \mathrm{FP}_{autre}. \tag{5.9}$$

Résultats

Deux variantes du jeu de données multivariées ont été utilisées pour tester les méthodes d'analyses : une variante sans aucun enregistrement aberrant et une variante avec 8 enregistrements aberrants.

Les méthodes d'analyse mises en jeu sont ma méthode d'analyse de synchronisations cooccurrentes, avec et sans seconde passe d'optimisation, et les méthodes ERSP à modèle multiplicatif, avec et sans normalisation simple essai, avec et sans correction de tests multiples.

La figure 5.2 présente les résultats obtenus sur les deux variantes du jeu de données multivariées. Les résultats de la méthode ERSP à modèle multiplicatif simple ($\text{ERSP}_{\%}$) étant très similaires aux résultats de la méthode ERSP avec normalisation simple essai (ERSP_{FT}), les résultats des deux méthodes ERSP ne sont pas différenciés dans la description suivante.

Les méthodes ERSP non corrigées pour les tests multiples ont une faible sensibilité à l'épisode de synchronisation cooccurrente et à son absence. De plus, leur proportion de fausses découvertes est supérieure à 90 %. Ces résultats sont dûs au fait que ces méthodes détectent de petits épisodes fallacieux de synchronisation, isolément dans les différentes électrodes. Ceci entraîne la détection de nombreux faux épisodes de synchronisation cooccurrente et leurs groupes d'électrodes associés.

Avec la correction de tests multiples, notamment par la procédure de Benjamini et Yekutieli, les résultats des méthodes ERSP sont sensiblement améliorés. Leur sensibilité à l'épisode de synchronisation cooccurrente et à l'absence de synchronisation augmente, dépassant 90 % pour la meilleure configuration. Inversement, leur proportion de fausses découvertes diminue, bien que restant supérieure à 20 % dans tous les cas.

En comparaison, la méthode d'analyse de synchronisations cooccurrentes a aussi une sensibilité élevée à l'épisode de synchronisation. Pour le jeu de données sans enregistrement aberrant, cette sensibilité est inférieure à celle de la meilleure configuration des méthodes ERSP. Dans le cas du jeu de données avec 8 enregistrements aberrants, cette sensibilité atteint 98 % et dépasse celle des méthodes ERSP.

La sensibilité de ma méthode à l'absence de synchronisation est supérieure à 99 % dans les deux variantes du jeu de données. Cette très grande sensibilité surpasse celle des méthodes ERSP et témoigne de l'aspect conservateur de ma méthode. Une conséquence logique de cet aspect conservateur est la très faible proportion de fausses découvertes, inférieure à 7 %. Cette dernière caractéristique assure que les découvertes mises en valeur par ma méthode reflètent de manière très probable un véritable épisode de synchronisation cooccurrente au sein des données.

L'utilisation de la seconde passe d'optimisation n'améliore pas sensiblement les résultats de ma méthode.

5.2.2. Données univariées

Le second jeu de données artificielles mis au point sert à tester la détection d'épisodes de synchronisation et de désynchronisation quand l'étude des signaux ne porte que sur une seule électrode à la fois. L'étude de cette synchronisation univariée ne nécessite alors la simulation que d'une seule électrode dans le jeu de données.

Présentation du jeu de données

Le jeu de données artificielles se compose de 20 enregistrements d'une seule électrode. Chaque enregistrement a une durée aléatoire comprise entre 1 700 ms et 2 000 ms, et est échantillonné à 1 000 Hz. Cette durée aléatoire sert à simuler sommairement des temps de réponse variables. Le temps d'apparition du stimulus est fixé à $t_s = 800\,\text{ms}$.

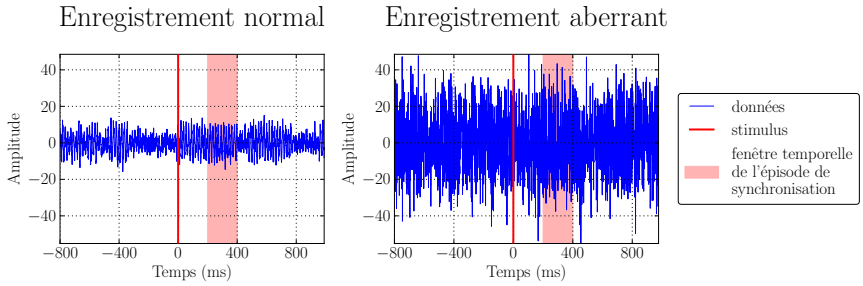

FIGURE 5.3. – Exemple de données artificielles utilisées pour l'évaluation de la détection de synchronisations univariées. Voir le texte pour les détails de génération des données.

Le signal de l'électrode est formé d'une sinusoïde à 55 Hz dont l'amplitude est modulée. Cette modulation est structurée sous la forme d'une alternance aléatoire de deux états :

- un état bas correspondant à un coefficient multiplicateur de 1,
- et un état haut correspondant à un coefficient multiplicateur tiré aléatoirement dans l'intervalle $[1;3]$ pour chaque enregistrement.

L'aléa sur le coefficient de l'état haut permet d'obtenir un rapport signal sur bruit variable pour chaque enregistrement. La durée de chaque état est échantillonnée à partir d'une distribution binomiale négative [2], avec une probabilité de succès $p = 0{,}985$ et un nombre d'échecs $n = 2$. Avec cette distribution, la durée moyenne d'un état vaut 131,34 ms, avec un écart-type de 93,57 ms. Un épisode de synchronisation est inséré en fixant à l'état haut le signal 200 ms après le stimulus, sur une durée de 200 ms.

Le signal de chaque enregistrement est contaminé par un bruit blanc gaussien additif d'écart-type 1. Ce signal est ensuite multiplié par un coefficient tiré aléatoirement dans l'intervalle $[1;3]$ pour chaque enregistrement, afin d'introduire une variabilité supplémentaire d'amplitude entre les enregistrements.

Comme pour le jeu de données multivariées, la robustesse des méthodes d'analyse est étudiée en injectant un bruit élevé supplémentaire sur un nombre fixé d'enregistrements. Il s'agit d'un bruit blanc gaussien additif d'écart-type égal à 5 fois l'écart-type de l'ensemble des données simulées. La figure 5.3 représente des signaux du jeu de données, avec et sans ce bruit supplémentaire.

Critères d'évaluation

Les méthodes d'analyses sont évaluées sur leur capacité à détecter l'épisode de synchronisation inséré dans les données, et à ne rien détecter d'autre. L'analyse ne porte que sur

2. En réalité, la suite d'états hauts et bas est obtenue via un modèle de Markov caché à états étendus (expanded state hidden Markov model ou ESHMM [Johnson, 2005]), c'est-à-dire où chaque état de la chaîne de Markov est lui-même modélisé par une chaîne de Markov. Dans le cas présent, ces sous-chaînes sont de type gauche-droite et ont deux états ayant chacun la même probabilité d'auto-transition. La durée de telles sous-chaînes suit une distribution binomiale négative [Bilmes, 2004].

la bande de fréquence de la transformée en ondelettes de Morlet centrée en $f_0 = 55\,\mathrm{Hz}$. La période du signal analysée est délimitée par la période la plus courte entre le stimulus et le temps de réponse dans tous les enregistrements.

Pareillement au jeu de données multivariées, le problème de détection présenté par le jeu de données univariées est traité comme un problème de classification à trois classes par les méthodes d'analyse :

- C_{rien}, pour l'absence d'épisode de synchronisation et de désynchronisation,
- C_{sync}, pour la détection d'un épisode de synchronisation,
- C_{dsync}, pour la détection d'un épisode de désynchronisation.

Comme le jeu de données ne contient pas d'épisode de désynchronisation, seules les sensibilités aux classes C_{rien} et C_{sync}, respectivement notées S_{rien} et S_{sync}, sont calculées.

La proportion de fausses découvertes est aussi relevée pour chaque méthode d'analyse. La proportion de fausses découvertes est calculée avec l'équation (5.8). La classification d'un point temporel dans la classe C_{dsync} étant une fausse découverte, le nombre de fausses découvertes FD s'écrit alors :

$$\mathrm{FD} = \mathrm{FP}_{sync} + \mathrm{VP}_{dsync} + \mathrm{FP}_{dsync}. \tag{5.10}$$

Résultats

Trois variantes du jeu de données univariées ont été mises en place pour évaluer la qualité de détection des méthodes d'analyse. Ces trois variantes se distinguent par le nombre d'enregistrements aberrants insérés, avec respectivement aucun, 4 et 8 enregistrements aberrants. Pour chaque variante, 50 jeux de données ont été générés, pour obtenir une estimation statistique des performances des méthodes. Les méthodes étant appliquées aux mêmes jeux de données, il est possible de comparer deux à deux les méthodes, en groupant par paires leurs résultats.

Les méthodes d'analyse étudiées sont les méthodes ERSP à modèle multiplicatif simple ($\mathrm{ERSP}_{\%}$) et avec normalisation simple essai ($\mathrm{ERSP}_{\mathrm{FT}}$), ainsi que ma méthode d'analyse de synchronisations univariées, avec validation fréquentiste (RMM) ou bayésienne (« RMM et Bayes »). Les méthodes ERSP ainsi que ma méthode avec la validation fréquentiste sont testées avec un seuillage simple des valeurs-p, mais aussi avec différentes corrections de tests multiples (B-H et B-Y). Par ailleurs, les méthodes ERSP sont également utilisées en ne prenant en compte qu'une période pré-stimulus réduite à 100, 200 ou 300 points temporels. Enfin, ma méthode d'analyse peut employer une seconde passe d'optimisation pour la recherche des meilleurs modèles de mélange ricien ($\mathrm{RMM}_{\mathrm{2nd}}$).

La figure 5.4 illustre les principaux résultats de ces expériences. Elle regroupe les résultats d'une partie de toutes les méthodes testées afin de mettre en relief l'influence des variations entre les méthodes. Un ensemble plus complet de figures représentant tous les résultats individuels et entre versions différentes des méthodes est fourni en annexe C.

Cette figure emploie un *diagramme haricot* (ou beanplot [Kampstra, 2008]) pour représenter la distribution des séries de données. Au sein de chaque « haricot » d'un diagramme haricot, chaque trait fin noir correspond une donnée de la série illustrée, un trait plus large rouge correspond à la moyenne de la série. L'enveloppe du « haricot »

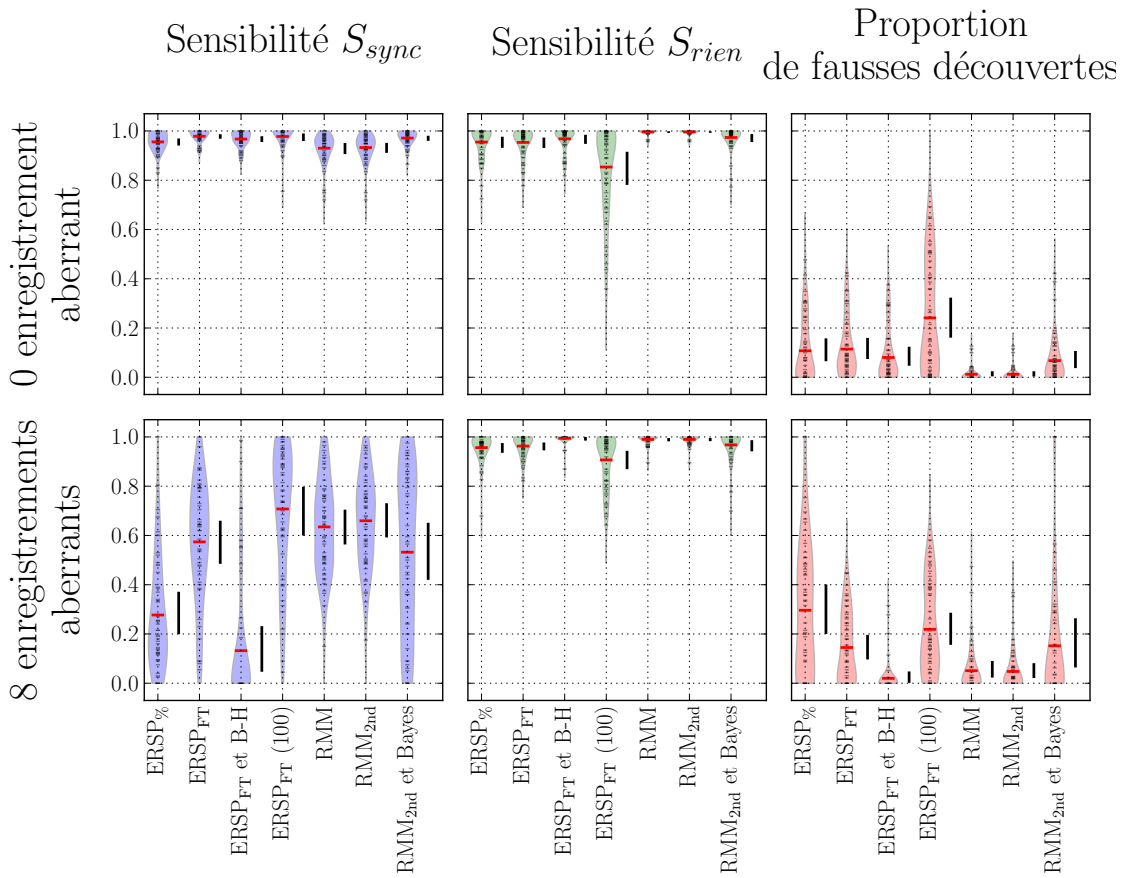

FIGURE 5.4. – Comparaison des performances pour la détection de synchronisations univariées des méthodes conventionnelles ERSP et de mes méthodes d'analyse de synchronisations univariées, sur le jeu de données artificielles univariées. Ce graphique ne présente qu'une partie de toutes les méthodes comparées. Les abréviations sont décrites à la section 5.1. La lecture du diagramme haricot est expliquée en section 5.2.2, page 97.

est formé par une estimation par noyau gaussien de la densité de la série et son reflet par symétrie axiale. Ce type de graphique permet de faire apparaître visuellement les particularités des séries de données telles qu'une distribution bimodale, ce que ne permet pas la boîte à moustaches par exemple. À côté de chaque série représentée par un haricot, un trait noir épais figure l'intervalle de confiance à 99 % sur la moyenne [3].

Pour la variante du jeu de données sans enregistrement aberrant, la majorité des méthodes d'analyse obtiennent une sensibilité moyenne à l'épisode de synchronisation et à l'absence de synchronisation supérieures à 80 %, et la proportion moyenne de fausses découvertes est inférieure à 20 %. Quelques méthodes employant une correction de tests d'hypothèses multiples obtiennent une sensibilité moyenne à l'épisode de synchronisation inférieure à 80 %, pouvant même décroître jusqu'à 20% (RMM_{2nd} et B-Y). D'autre part, quelques méthodes de type ERSP avec une restriction de période pré-stimulus à 100 points présentent une proportion de fausses découvertes supérieure à 20 %.

Les résultats des méthodes, globalement satisfaisants, se dégradent pour les variantes du jeu de données avec des enregistrements aberrants. Ainsi la variante du jeu de données avec 8 enregistrements aberrants fait apparaître distinctement les méthodes les moins robustes. Avec cette variante, la sensibilité moyenne à l'épisode de synchronisation décroît pour toutes les méthodes. Parmi les méthodes de type ERSP, la méthode ERSP avec normalisation simple essai sans autre modification ($ERSP_{FT}$) présente les meilleurs résultats, avec une sensibilité moyenne à l'épisode de synchronisation de 57,5 % et une proportion moyenne de fausses découvertes de 14,5 %. D'autres méthodes de type ERSP restent plus sensibles à l'épisode de synchronisation mais au prix d'une proportion moyenne de fausses découvertes élevée, supérieure à 20 %.

Pour cette même variante du jeu de données avec 8 enregistrements aberrants, ma méthode d'analyse avec validation fréquentiste et une seconde passe d'optimisation (RMM_{2nd}) obtient des résultats en moyenne meilleurs que la meilleure méthode de type ERSP. Sa sensibilité moyenne à l'épisode de synchronisation vaut 66 % et sa proportion moyenne de fausses découvertes vaut 0,05 %. Un examen attentif des intervalles de confiances des moyennes révèle néanmoins que cette méthode n'est pas significativement meilleure en moyenne que la meilleure méthode de type ERSP. De même, ma méthode avec une validation bayésienne (« RMM_{2nd} et Bayes ») obtient des résultats en moyenne inférieurs à la meilleure méthode de type ERSP, mais les différences de moyenne ne sont pas significatives.

La figure 5.5 présente les différences appariées entre la meilleure méthode ERSP, c'est-à-dire la méthode ERSP à modèle multiplicatif avec normalisation simple essai sans restriction de période pré-stimulus ni correction de tests multiples, et mes meilleures méthodes avec validation fréquentiste et bayésienne. Il est intéressant de noter pour ces trois méthodes, dans le cas avec 8 enregistrements aberrants, que les différences appariées des sensibilités à l'épisode de synchronisation sont très dispersées bien que les distributions respectives de cette sensibilité, visibles sur la figure 5.4, soient très semblables. Ceci reflète le fait que mes méthodes présentent de bons résultats sur des enregistrements où la

3. L'intervalle de confiance est un intervalle de confiance bootstrap bilatéral reposant sur les percentiles, calculé avec 2 000 échantillons [Davison and Hinkley, 1997].

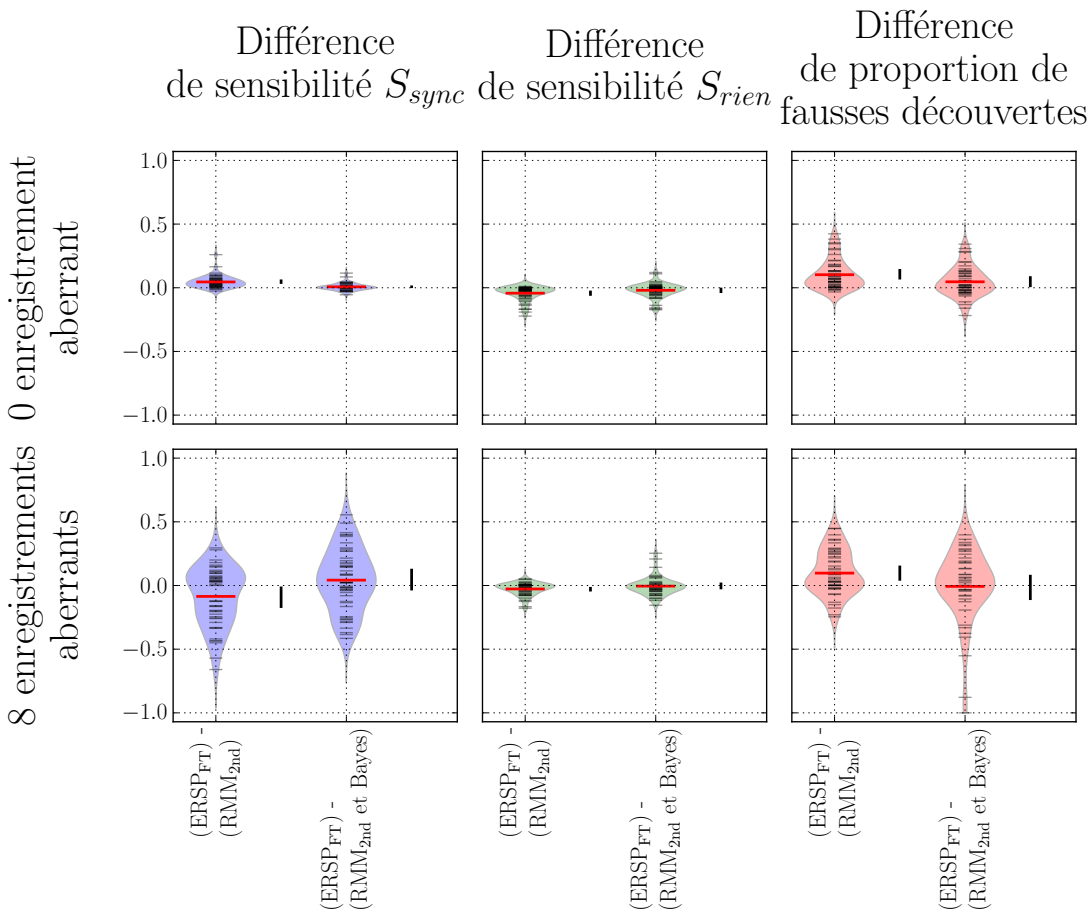

FIGURE 5.5. – Comparaison deux à deux de la meilleure variante des méthodes ERSP et des variantes de mes méthodes d'analyse de synchronisations univariées, avec validation fréquentiste et bayésienne. Le diagramme haricot représente les séries de différences appariées des résultats obtenus sur chaque jeu de données. Voir la section 5.1 pour une description des abréviations, et la section 5.2.2, page 97, pour une explication sur la lecture du diagramme haricot.

méthode ERSP n'est pas efficace et inversement.

Par ailleurs, la comparaison par paire des résultats des différentes modifications des méthodes d'analyse permet les observations suivantes :

- La méthode ERSP à modèle multiplicatif avec normalisation simple essai est beaucoup plus robuste au bruit que la version sans normalisation simple essai, exhibant une sensibilité à l'épisode de synchronisation supérieure et une proportion de fausses découvertes moindre. La sensibilité à l'absence de synchronisation est équivalente.
- La réduction de la période pré-stimulus dans les méthodes ERSP entraîne une augmentation de la sensibilité à l'épisode de synchronisation. Cette amélioration est contrebalancée par une diminution de la sensibilité à l'absence de synchronisation et une augmentation de la proportion de fausses découvertes. Les écarts sont d'autant plus grands que la période pré-stimulus est réduite. Ainsi, les méthodes ERSP deviennent trop sensibles à la synchronisation et produisent de nombreux faux positifs quand la période pré-stimulus est trop courte.
- L'ajout d'une seconde passe d'optimisation pour les modèles de mélange ricien dans mes méthodes d'analyse n'apporte aucune amélioration significative.
- L'emploi de procédures de correction de tests multiples, avec les méthodes ERSP et ma méthode avec validation fréquentiste provoque une diminution de la proportion de fausses découvertes, mais aussi de la sensibilité à l'épisode de synchronisation. Les méthodes corrigées deviennent plus conservatrices et privilégient l'absence de synchronisation. La procédure de Benjamini et Yekutieli rend les méthodes plus conservatrices que la procédure de Benjamini et Hochberg. Dans les deux cas, le biais induit par ces procédures est tel que pour un jeu de données avec 8 enregistrements aberrants les méthodes d'analyse ne détectent plus que très faiblement l'épisode de synchronisation.

Ces phénomènes sont observables sur l'échantillon présenté dans la figure 5.4 et à l'annexe C.

5.3. Données expérimentales

L'évaluation des méthodes d'analyse sur les jeux de données artificielles a permis de montrer que mes méthodes d'analyse sont capables de détecter certains types de phénomènes de synchronisation, avec une robustesse égalant les meilleures méthodes de type ERSP.

Cependant, ces jeux de données artificielles sont une représentation partielle et simplifiée de vraie données d'enregistrements électro-corticaux. Il est donc indispensable de tester mes méthodes sur de vraies données biologiques et de comparer leurs résultats avec ceux obtenus par des méthodes de type ERSP.

5.3.1. Présentation du jeu de données

Les données expérimentales sont des enregistrements intra-corticaux recueillis sur des macaques crabiers (macaca fascicularis) au cours d'une tâche visuo-motrice. Ces

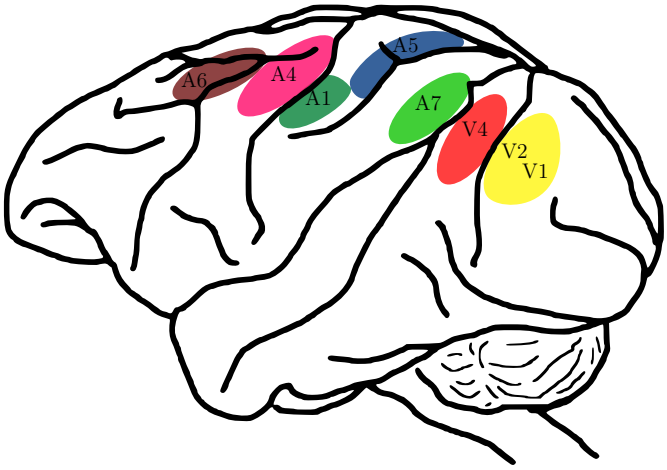

FIGURE 5.6. – Aires cérébrales d'implantation des électrodes dans le cortex d'un macaque.

enregistrements ont été fournis par le Dr. Matthias Munk et réalisés au Max Planck Institute for Brain Research à Frankfurt [Hutt and Munk, 2009].

Les signaux électriques enregistrés sont issus d'électrodes bipolaires implantées de manière chronique dans le cortex du macaque. Les 17 électrodes sont réparties dans les aires visuelles V1, V2, V4, l'aire pariétale A7, l'aire somato-sensorielle secondaire A5, l'aire somato-sensorielle post-centrale A1, le cortex moteur primaire A4 et le cortex pré-moteur A6. La figure 5.6 illustre les aires d'implantation des électrodes. Ces signaux sont échantillonnés à 1 kHz et traités avec un filtre passe-bande entre 5 Hz et 150 Hz (3 dB/octave). Un filtre réjecteur de bande de Chebyshev de type 2 est appliqué pour supprimer le bruit à 50 Hz dû à l'alimentation.

Le singe est installé face à un écran, à portée d'un levier à plusieurs positions. La tâche visuo-motrice se décompose alors en une séquence de stimuli visuels présentés au singe via l'écran, auxquels il doit répondre de manière appropriée par une commande motrice sur le levier. Chaque enregistrement commence par l'affichage d'un point de fixation sur l'écran. Un motif rayé en mouvement est alors affiché en position parafovéale, 200 ms après la fixation par le singe. Suivant la direction du mouvement du stimulus, le singe doit déplacer le levier à une position donnée, en avant ou en arrière. Si le singe positionne correctement le levier à la position désirée pendant 200 ms, un nouveau stimulus est présenté à l'écran sous la forme d'un changement de vitesse et de direction du motif rayé. Cette séquence stimulus/réponse est répétée quatre fois au total. Si le singe réussit les quatre séquences de l'enregistrement, il reçoit une petite quantité d'eau en récompense.

Le jeu de données étudié comporte 40 enregistrements issus de d'un sujet. Dans ce jeu de données, seule la période entre le premier stimulus et la première réponse est considérée. La durée de cette période varie entre 644 ms et 1 292 ms. La période pré-stimulus fournie est de 300 ms. Les résultats détaillés par la suite sont tous obtenus après verrouillage des enregistrements sur le stimulus.

5.3.2. Résultats des méthodes d'analyse

Les données expérimentales ont été analysées avec les méthodes ERSP standards et les deux méthodes développées dans cette thèse. Cette section aborde les conditions de mise en œuvre des méthodes ainsi que leurs résultats bruts. L'intérêt des résultats résidant dans les corrélations entre les méthodes, la section suivante présentera sous cet angle un examen approfondi des résultats.

Résultats des méthodes ERSP

Dans un premier temps, les données expérimentales ont été étudiées à l'aide des méthodes ERSP standard décrites précédemment. Les données ont été analysées dans les bandes des ondes bêta et gamma, c'est-à-dire entre 15 et 90 Hz. Seuls les résultats de la méthode ERSP utilisant un modèle multiplicatif et une normalisation simple essai seront détaillés, cette méthode étant la plus robuste, d'après la littérature [Grandchamp and Delorme, 2011] et les résultats obtenus sur les jeux de données artificielles.

La figure 5.7 présente les épisodes de synchronisation et désynchronisation détectés par la méthode ERSP au sein de chaque électrode. Différentes procédures de correction de tests multiples ont été appliquées, plus ou moins conservatrices. Celles-ci permettent d'assurer un faible taux de fausses découvertes.

L'emploi de ces procédures est ici capital, car la quantité de données disponible dans la période pré-stimulus est faible. En effet, au maximum, la méthode ERSP dispose de 300 points dans la période pré-stimulus. La gestion des effets de bord lors du calcul de la transformée en ondelettes continue retire des points supplémentaires dans la période pré-stimulus (cf. section 2.1.4). Ainsi, celle-ci se réduit à une centaine de points pour les plus basses fréquences traitées. Or les expériences menées sur le jeu de données artificielles univariées ont montré précédemment que dans ce contexte, la méthode ERSP engendre de nombreux faux positifs.

Résultats de la méthode d'analyse de synchronisations cooccurrentes

La méthode d'analyse de synchronisations cooccurrentes a été appliquée au jeu de données après un sous-échantillonnage de celui-ci à 250 Hz. Ce sous-échantillonnage permet de réduire le nombre de points temporels traités et diminue ainsi le temps de calcul. Il limite par contre la fréquence maximum analysable avec la transformée en ondelettes continue. Ainsi, le jeu de données est analysé entre 15 et 70 Hz. La seconde passe d'optimisation des modèles de mélange gaussien n'a pas été utilisée, celle-ci semblant inutile d'après les résultats sur les données artificielles.

La méthode d'analyse a détecté de nombreux sous-groupes d'électrodes présentant des épisodes de synchronisation cooccurrente. La majorité de ces sous-groupes est composée de groupes d'une seule électrode ou au contraire de groupes rassemblant toutes les électrodes sauf une. Les autres sous-groupes d'électrodes détectés sont constitués de deux électrodes ou de l'ensemble complémentaire des électrodes. Pour l'examen des résultats, j'ai choisi de me focaliser sur les groupes d'électrodes à une ou deux électrodes, et de laisser de côté

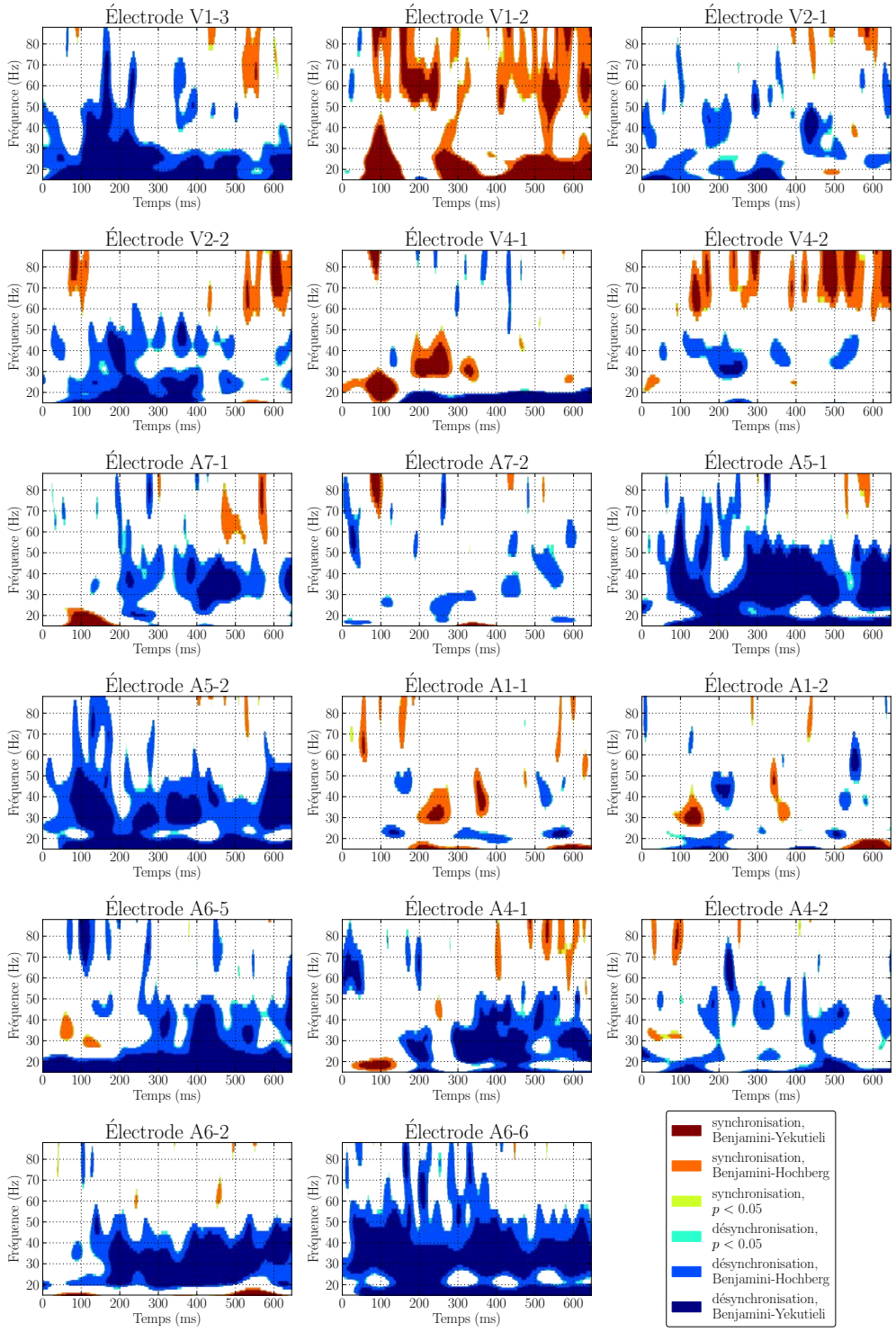

FIGURE 5.7. – Épisodes de synchronisations et désynchronisations verrouillés sur le stimulus visuel, extraits par méthode ERSP à modèle multiplicatif avec normalisation simple essai. Différents seuillages des valeurs-p sont présentés, suivant la méthode de correction de tests multiples appliquée.

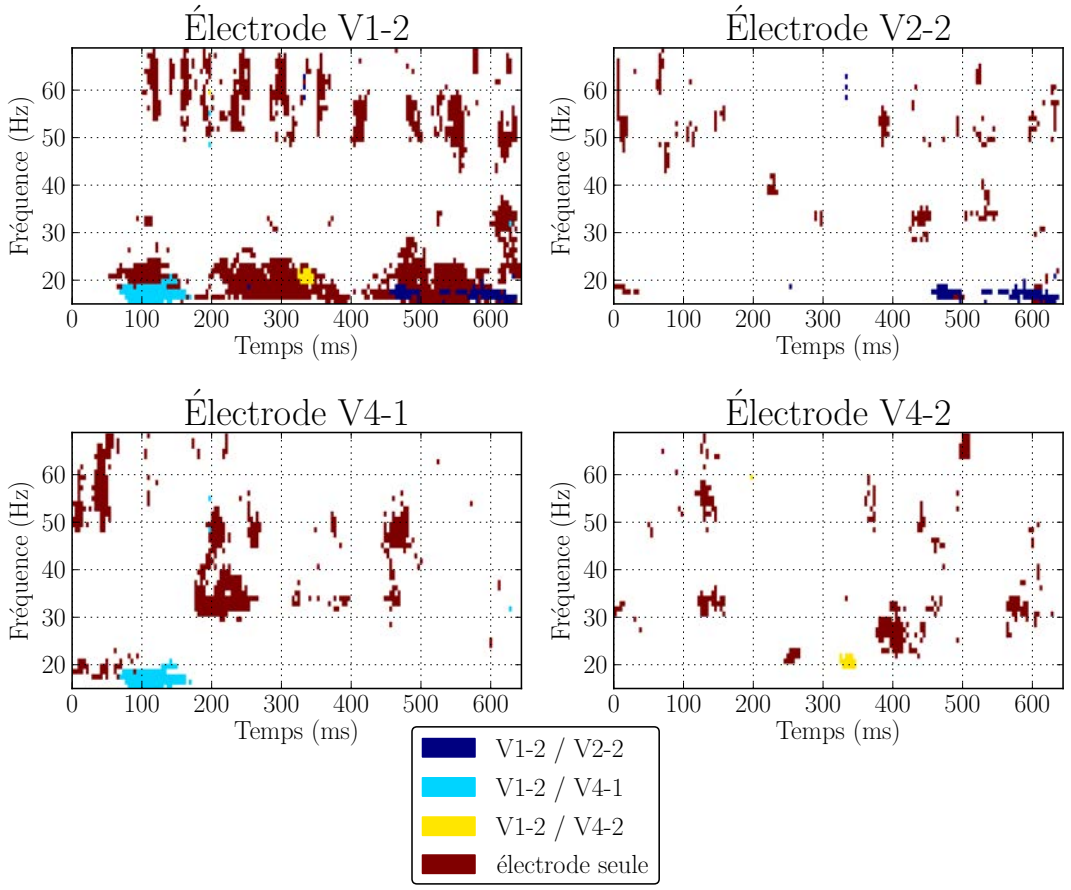

FIGURE 5.8. – Électrodes participant à des petits sous-groupes d'électrodes détectés par la méthode d'analyse de synchronisations cooccurrentes. Les régions du plan temps-fréquence où une électrode forme un sous-groupe à elle seule sont en rouge. Des couleurs spécifiques représentent les points temps-fréquence où les électrodes forment une paire d'électrodes à synchronisation cooccurrente.

leur complémentaire, vu qu'il apparaît nécessairement puisque les modèles de mélange gaussien mis en jeu ne détectent au mieux que deux composantes.

Les électrodes présentes dans les petits sous-groupes d'électrodes sont au nombre de quatre : V1-2, V2-2, V4-1 et V4-2. La figure 5.8 illustre les points temps-fréquence où ces électrodes apparaissent dans des sous-groupes d'électrodes, seules ou en paire avec une autre électrode.

Résultats de la méthode d'analyse de synchronisations univariées

La méthode d'analyse de synchronisations univariées a été mise en jeu sur les données expérimentales. Comme pour les méthodes ERSP, les bandes de fréquence étudiées englobent les ondes bêta et les ondes gamma, soit de 15 à 90 Hz.

Les deux techniques de validation de cette méthode d'analyse ont été utilisée : l'approche fréquentiste utilisant le bootstrap et l'approche bayésienne utilisant un modèle hiérarchique. L'approche fréquentiste générant des valeurs-p, différentes procédures de correction de tests multiples ont été appliquées, même si les résultats des expériences sur les données artificielles permettent de penser qu'elles sont inutiles avec cette méthode.

Les épisodes de synchronisation et de désynchronisation détectés sont illustrés aux figures 5.9 et 5.10.

5.3.3. Mise en correspondance des résultats et interprétation

À partir des résultats des différentes méthodes, il est intéressant de chercher les épisodes de synchronisation et de désynchronisation communs et ceux divergents. Les premiers dénotent des événements très probables, car révélés sous différentes hypothèses. Les seconds mettent en lumière les particularités de chaque méthode. Pour tous les épisodes détectés, il est important de vérifier leur plausibilité au regard de la littérature.

Épisodes de synchronisation et désynchronisation communs

Un premier épisode de synchronisation, détecté par toutes les méthodes, se situe à 20 Hz et 100 ms après le stimulus, et est présent simultanément aux électrodes V1-2 et V4-1. Cet épisode est détecté même après l'application de procédures de correction de tests multiples. Il est aussi détecté par ma méthode d'analyse de synchronisations cooccurrentes, comme il se produit simultanément aux deux électrodes V1-2 et V4-1. La figure 5.11 symbolise ce sous-ensemble d'électrodes.

De plus, toutes les méthodes d'analyse, à l'exception de ma méthode d'analyse de synchronisations cooccurrentes, détectent les épisodes de synchronisation et de désynchronisation suivants :

- un épisode de synchronisation aux électrodes A4-1 et A7-1, également à 20 Hz et 100 ms après le stimulus,
- un épisode de synchronisation aux électrodes A6-5, entre 25 et 40 Hz, de 50 et à 150 ms après le stimulus,
- puis un épisode de désynchronisation aux électrodes A6-5, A6-2 et A4-1, aussi entre 25 et 40 Hz, mais de 300 à 500 ms après le stimulus,

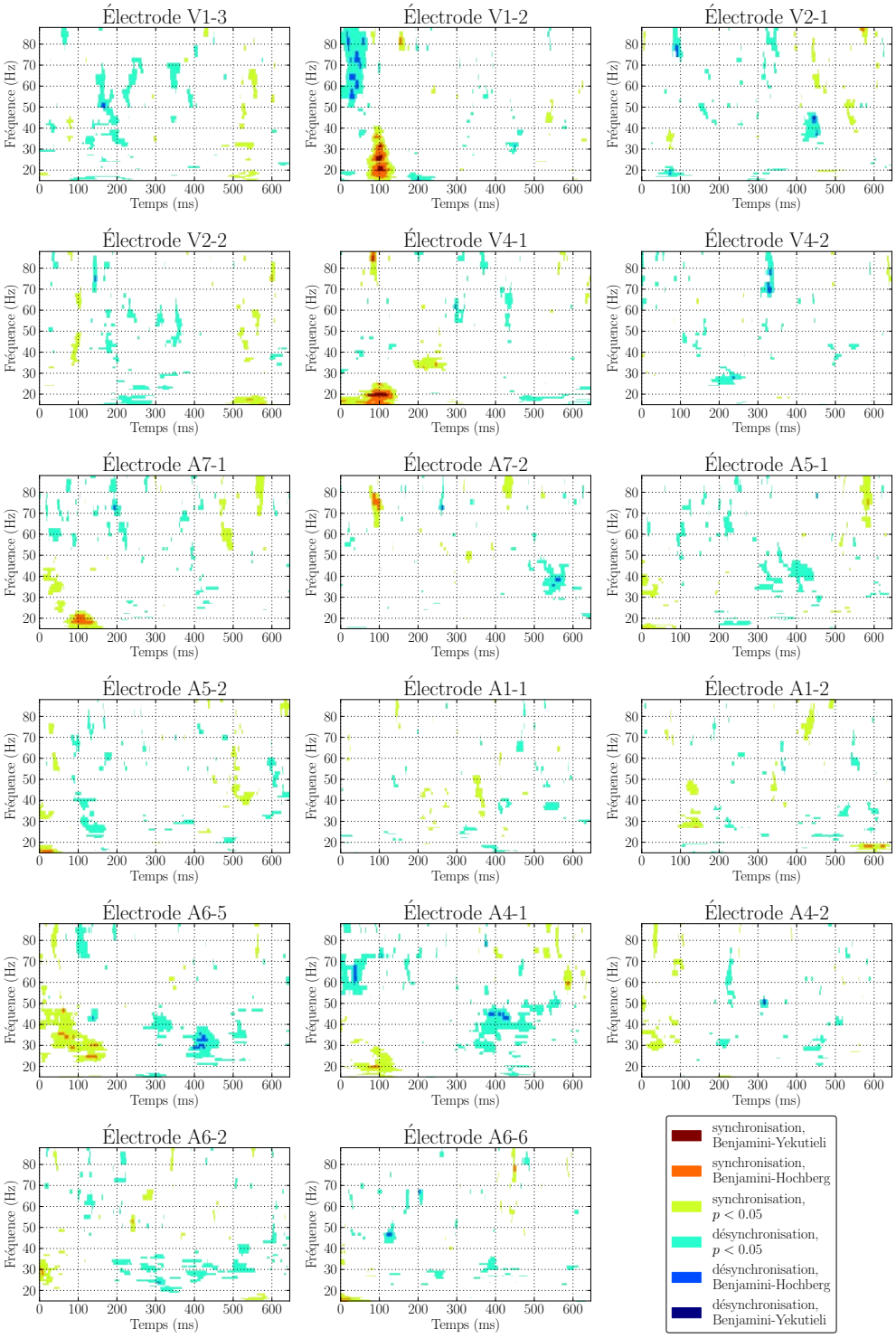

FIGURE 5.9. – Épisodes de synchronisations et désynchronisations verrouillées sur le stimulus visuel, extraits par le modèle d'analyse de synchronisations univariées, avec une validation fréquentiste par bootstrap. Différents seuillages des valeurs-p sont représentés, suivant la méthode de correction de tests multiples employées.

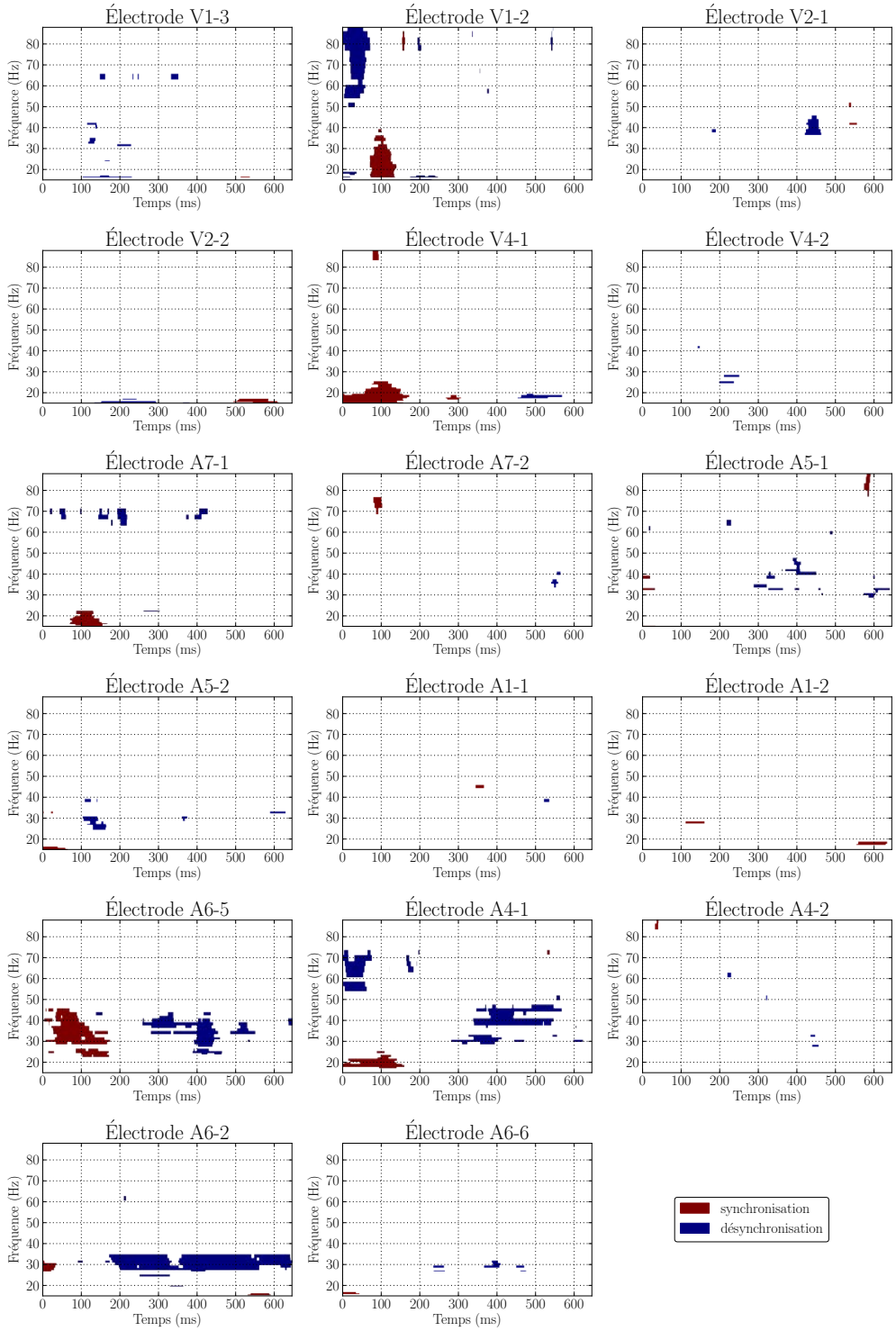

FIGURE 5.10. – Épisodes de synchronisations et désynchronisations verrouillées sur le stimulus visuel, extraits par le modèle d'analyse de synchronisations univariées, avec une validation bayésienne par modèlisation hiérarchique.

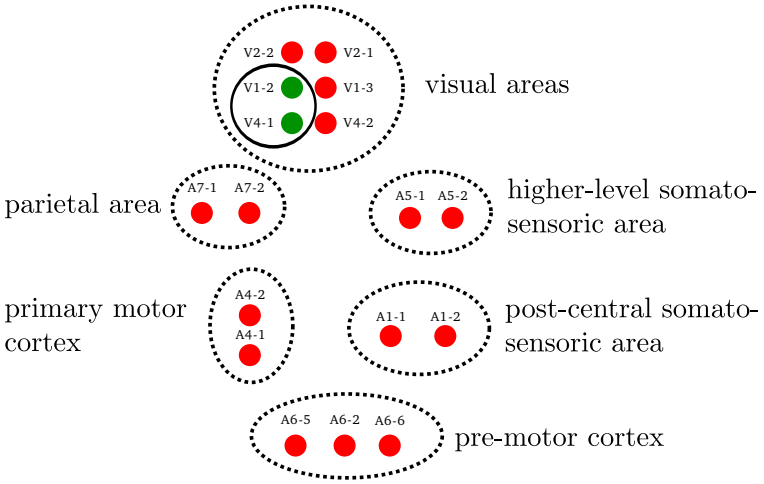

FIGURE 5.11. – Représentation schématique du groupe d'électrodes ayant un épisode de synchronisation au même point temps-fréquence, à 100 ms après le stimulus et à 20 Hz, détecté par toutes les méthodes testées. Chaque cercle coloré représente une électrode. Les cercles en pointillés symbolisent les aires corticales où sont situées les électrodes. Le cercle en trait plein symbolise le groupe d'électrode synchronisées en V1-2 et V4-1

– un épisode de désynchronisation en V2-1, de 35 à 45 Hz et de 420 à 460 ms après le stimulus,
– un épisode de désynchronisation en V1-2, entre 60 et 80 Hz, à 50 ms après le stimulus.

L'étendue temps-fréquence de ces événements est délimitée par le plus petit espace du plan temps-fréquence détecté conjointement par les différentes méthodes.

Par ailleurs, ma méthode d'analyse de synchronisations cooccurrentes détecte certains épisodes corroborés par la méthode ERSP. Ces groupes d'électrodes ne sont en fait constitués que d'une seule électrode :

– l'électrode V1-2 présente des épisodes de synchronisation intermittents, à 50 Hz et plus, à partir de 100 ms après le stimulus jusqu'à la fin,
– cette même électrode présente aussi un épisode de synchronisation de 15 à 25 Hz, de 250 ms après le stimulus jusqu'à la fin,
– l'électrode V4-1 exhibe un épisode de synchronisation entre 30 et 40 Hz, de 190 à 270 ms après le stimulus.

Ce dernier épisode est également détecté par ma méthode d'analyse de synchronisations univariées, mais uniquement avec la validation fréquentiste.

Épisodes de synchronisation et désynchronisation différents

Le cas le plus remarquable d'événements détectés par une unique méthode concerne la méthode ERSP. Plus précisément, celle-ci détecte des épisodes de désynchronisations dans le bas de la bande de fréquence des ondes bêta, entre 15 et 20 Hz, sur quasiment l'intégralité

de la durée du signal aux électrodes V1-3, V4-1, A5-1, A5-2, A6-5, A4-2 et A6-6. De même, cette méthode est la seule à détecter de larges épisodes de désynchronisation entre 25 et 50 Hz aux électrodes V1-3, V4-2, V2-2, A7-1, A5-1, A5-2 et A6-6. Comme décrit précédemment, il est probable que ces résultats soient en grande partie des faux positifs. La discussion à la section 5.4.1 apportera des arguments supplémentaires favorables à cette hypothèse.

Plausibilité des résultats

Les résultats obtenus sont dépendants du type de sujet, de la localisation des électrodes, du type de signal électrique recueilli, du stimulus visuel et de la tâche motrice. Aussi, pour illustrer la plausibilité des résultats, j'ai sélectionné trois études proches de par leur dispositif expérimental.

Ainsi les travaux de Juergens, Guettler et Eckhorn [Juergens et al., 1999] mettent en jeu une expérience relativement similaire, dans laquelle un macaque rhésus (macaca mulatta) doit fixer un motif rayé d'abord immobile puis en mouvement. L'analyse des potentiels de champs locaux porte sur des enregistrements intracrâniens réalisés à l'aide de sept électrodes implantées dans l'aire V1. Il est intéressant de noter parmi leurs résultats la détection d'une activité évoquée correspondant à un accroissement de la puissance du signal de 80 à 170 ms après l'apparition du stimulus immobile, dans les bandes des ondes bêta et gamma (jusqu'à 100 Hz). En particulier, les auteurs ont noté un pic de la puissance à 20 Hz et à 100 ms après le stimulus. De plus, les auteurs ont aussi mentionné la présence d'une importante modulation de la puissance du signal entre 40 et 100 Hz, correspondant d'abord à une activité évoquée à 80 ms après l'apparition du stimulus puis devenant progressivement une activité induite qui se poursuit au-delà de 360 ms après l'apparition du stimulus.

Vis-à-vis de mes résultats, le pic observé à 20 Hz et 100 ms après le stimulus correspond très précisément à un résultat commun à toutes les méthodes d'analyse de synchronisations, pour l'électrode V1-2. Quant à l'accroissement soutenu de la puissance du signal dans la bande des ondes gamma, cette observation pourrait correspondre au résultat que j'ai obtenu conjointement par la méthode ERSP et ma méthode d'analyse de synchronisations cooccurrentes, aussi pour l'électrode V1-2.

Dans les travaux de Turi, Gotthardt, Singer, Vuong, Munk et Wibral [Turi et al., 2012], les auteurs emploient un dispositif expérimental quasiment identique à celui que j'ai étudié. En effet, une de leurs expériences consiste en une tâche visuo-motrice où un macaque doit fixer un point sur un écran puis, à l'apparition d'un motif rayé en mouvement, bouger un levier à des positions déterminées, suivant la vitesse du motif. Des potentiels de champs locaux sont enregistrés à l'aide d'électrodes implantées dans les aires V4 et A4. L'analyse faite par les auteurs des potentiels évoqués révèle un événement à 100 ms après le stimulus dans les deux aires enregistrées. L'étude ne portant que sur la moyenne temporelle du signal, la bande de fréquence des potentiels évoqués détectés n'est pas donnée.

Cette observation corrobore l'épisode de synchronisation détecté aux électrodes V4-1 et A4-1 dans le jeu de données que j'ai analysé. Il faut cependant noter que les auteurs ont mis en relief que le potentiel évoqué détecté dans l'aire A4 est dû à un phénomène de

remise à zéro de la phase. Aussi, s'il s'agit du même phénomène, ce phénomène n'aurait pas dû être détecté par les méthodes d'analyse que j'ai employées, celles-ci étant insensibles à la phase du signal.

La dernière étude que je souhaite présenter a été réalisée par Sanes et Donoghue [Sanes and Donoghue, 1993]. Dans celle-ci, le macaque crabier (macaca fascicularis) est installé face à un écran où apparaît d'abord une instruction sur un mouvement à réaliser pendant 2 à 3 s puis un signal de départ pour effectuer le mouvement. Ce mouvement consiste en une flexion du poignet ou d'un doigt. Les potentiels de champs locaux sont enregistrés à l'aide de 12 électrodes insérées dans le cortex moteur primaire A4 et le cortex pré-moteur A6. L'analyse temps-fréquence de ces potentiels révèle un accroissement des oscillations dans la bande de fréquence entre 15 et 50 Hz, mais sans lien direct avec l'apparition des différentes instructions à l'écran. Néanmoins, la cessation de ces oscillations est fortement corrélée avec le début du mouvement. Les fréquences dominantes dans ces enregistrements varient entre 20 et 35 Hz suivant le singe et le type de tâche motrice.

Dans le jeu de données que j'ai étudié, le temps du début du mouvement du singe n'est pas reporté. Il est cependant possible de faire l'hypothèse que la fin de la période commune à tous les enregistrements contient la période du mouvement du singe sur le levier. Aussi, les épisodes de désynchronisation détectés dans le signal des électrodes A6-5, A6-2 et A4-1 de 25 à 40 Hz entre 300 et 500 ms pourraient être de même nature que le phénomène observé par Sanes et Donoghue, et dus au mouvement du bras du singe.

5.4. Discussion générale des résultats

Les sections précédentes ont montré que les méthodes d'analyse que j'ai développées sont capables d'extraire de jeux de données artificiels et expérimentaux des informations comparables aux méthodes standards de type ERSP, sans avoir recours aux hypothèses sous-jacentes à ces dernières.

La présente section a pour but de mettre en relief de manière plus qualitative les avantages et inconvénients de chaque méthode, et d'évaluer la confiance qu'il est possible d'accorder aux résultats obtenus grâce à celles-ci.

5.4.1. Qualité des résultats

Les expériences sur les jeux de données artificielles témoignent des bons résultats obtenus par les nouvelles méthodes introduites dans la présente thèse. Il faut admettre que cette réussite était attendue puisque les données ont été générées d'une part pour correspondre aux hypothèses faites par ces méthodes et d'autre part pour présenter une variabilité entre les enregistrements. De manière plus surprenante, les méthodes standards se sont révélées assez robuste à cette variabilité, mais défaillantes dans un cas qui n'était pas prévu à l'origine : la faible quantité de données dans la période pré-stimulus. En effet, l'étude des données expérimentales a mis en valeur ce problème qui a été ajouté par la suite aux cas de tests sur les jeux de données artificielles pour le quantifier.

Ainsi, si les résultats de l'analyse des données expérimentales par les différentes méthodes présentent de nombreuses similarités, la question se pose de l'origine des différences des

résultats et du crédit à porter à chaque méthode vis-à-vis de ces différences.

Parcimonie de la méthode d'analyse de synchronisations cooccurrentes

La méthode d'analyse de synchronisations cooccurrentes utilise pour détecter les phénomènes de synchronisation l'hypothèse que certaines électrodes ont des épisodes de synchronisation ou désynchronisation aux mêmes instants, formant ainsi des groupes d'électrodes identiques à travers les différents enregistrements. Cette première approche m'a permis de développer une méthode qui n'emploie ni l'information de la période pré-stimulus, ni une moyenne des scalogrammes des enregistrements.

Cependant, deux limitations inhérentes à cette hypothèse de cooccurrence des synchronisations apparaissent à l'usage. La première limitation est qu'il n'est pas possible de savoir si les groupes d'électrodes détectés par la méthode sont issus d'un épisode commun de synchronisation ou de désynchronisation, comme aucune notion d'ordre n'entre en jeu entre les groupes d'électrodes détectés. Une seconde limitation apparaît quand un faible nombre d'électrodes est en jeu, comme c'était le cas avec les données expérimentales. Dans ce cas, si un groupe d'électrodes est détecté, son complémentaire rassemblant toutes les autres électrodes sera aussi systématiquement détecté. Ainsi, parmi les résultats de la méthode, apparaîtront simultanément le groupe d'électrodes à synchronisation cooccurrente et son complémentaire illégitime. Le parti que j'ai pris pour l'analyse des données expérimentales a été de ne rapporter que le plus petit des deux groupes d'électrodes. Ce choix, contestable, était supporté par le fait que ces résultats étaient corrélés avec ceux des autres méthodes, ce qui les rendait plus plausibles.

En examinant les résultats obtenus sur les données expérimentales, il apparaît que la méthode d'analyse de synchronisations cooccurrentes détecte moins de phénomènes que toutes les autres méthodes. Une première explication vient du fait qu'une part des phénomènes de synchronisation et désynchronisation ont lieu dans une seule électrode en un point temps-fréquence. La méthode est capable de détecter ces groupes d'électrodes ne contenant qu'une seule électrode, comme le prouve les résultats, mais ceci est plus difficile. En effet, les modèles bayésiens de mélange gaussien au cœur de la méthode privilégient facilement des modèles à une seule composante plutôt que des modèles à deux composantes dont l'une est quasiment vide. Une seconde explication vient de l'utilisation de mesures de stabilité pour valider les groupes d'électrodes détectés. Si ces mesures sont adéquates dans le cadre de l'analyse des données artificielles, il est fort probable qu'elles soient trop sévères pour l'analyse des données expérimentales. Ainsi, les résultats de la méthode sont peu nombreux mais le degré de confiance que l'on peut leur accorder est très élevé.

Éparpillement de la méthode d'analyse de synchronisations univariées

La seconde méthode que j'ai développée se fonde sur l'hypothèse que la puissance d'un signal dans une bande de fréquence peut être binarisée en deux états et qu'un épisode de synchronisation ou désynchronisation correspond alors à une proportion anormalement élevée ou faible d'états hauts ou bas en un point temps-fréquence. Le fait que l'analyse

porte sur chaque bande de fréquence indépendamment et que les résultats soient des détections d'épisodes de synchronisation ou de désynchronisation rapproche cette méthode des méthodes standards et permet une comparaison des résultats plus aisée.

Numériquement, sur les données artificielles, les résultats de cette méthode sont comparables avec la meilleure méthode ERSP et dépassent ceux des méthodes ERSP courantes. Ceci n'est pas valable lorsque des techniques de correction de tests multiples sont employées conjointement avec la variante de la méthode avec une validation fréquentiste. En effet, la correction de tests multiples est une procédure très conservatrice, qui entraîne la disparition de nombreux résultats d'intérêt, voire de la totalité des résultats dans le cas d'un signal très bruité. L'emploi de ces techniques de correction semble donc superflu, d'autant plus que la proportion de fausses découvertes de la méthode non corrigée n'est pas assez élevé pour justifier leur emploi.

Sur les données expérimentales, les épisodes de synchronisation ou désynchronisation détectés par la méthode d'analyse correspondent pour la grande majorité à des épisodes aussi détectés avec la méthode ERSP. Cependant, la méthode d'analyse de synchronisations univariées est plus parcimonieuse que la méthode ERSP, sans pour autant être aussi restrictive que ma première méthode. Les bonnes caractéristiques de la méthode en termes de proportion de fausses découvertes sur les données artificielles donne confiance dans ses résultats sur les données expérimentales.

Par ailleurs, une différence qualitative entre les variantes de la méthode employant une validation fréquentiste ou bayésienne apparaît avec les résultats sur les données expérimentales. L'éparpillement des épisodes détectés varie en fonction de la validation utilisée. Ainsi, la validation fréquentiste conserve plus d'épisodes de synchronisation et désynchronisation mais ceux-ci sont plus dispersés dans le plan temps-fréquence. Au contraire, les épisodes retenus après la validation bayésienne sont plus étendus et connexes mais moins nombreux. Dès lors, la validation bayésienne permet d'obtenir des résultats reproduisant mieux la forte corrélation des signaux en des points temps-fréquence voisins, mais est plus conservatrice que la validation fréquentiste.

Robustesse et fausses découvertes des méthodes ERSP

La méthode ERSP à base d'un modèle multiplicatif est une méthode classique pour détecter les épisodes de synchronisation et de désynchronisation dans des signaux électrocorticaux. Les résultats sur les jeux de données artificielles témoignent du manque de robustesse de cette méthode en présence d'enregistrements aberrants. Une étape de détection et de rejet des enregistrements contenant trop d'artefacts est donc obligatoire pour l'usage de cette méthode. Cette étape est intégrée dans les chaînes de traitement de signaux corticaux [Delorme and Makeig, 2004], mais reste fastidieuse car réalisée manuellement.

Concernant la seconde méthode ERSP testée, utilisant une normalisation simple essai, sa robustesse décrite dans la littérature s'est vue confirmée dans l'analyse des données artificielles. Les performances de cette méthode sont comparables à celles de mes méthodes d'analyse. Cependant, les résultats sur les données expérimentales peuvent paraître intrigants de par la grande quantité d'épisodes de synchronisation et de désynchronisation

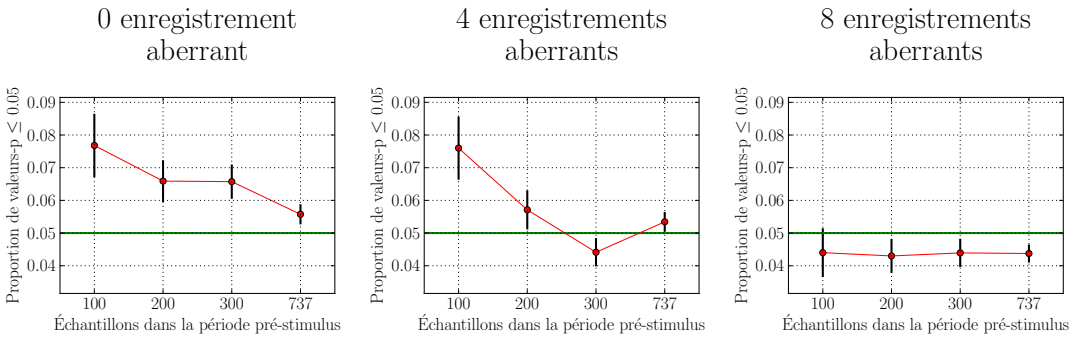

FIGURE 5.12. – Proportion des valeurs-p inférieures à 0,05 parmi celles calculées dans la période pré-stimulus par la méthode ERSP avec normalisation simple essai avec les jeux de données artificielles univariées, en fonction du nombre de points temporels disponibles dans la période pré-stimulus de chaque enregistrement. Chaque proportion estimée est accompagnée d'un intervalle de confiance à 99 %. La ligne verte est un guide visuel pour la valeur de seuil à $\alpha = 0{,}05$.

détectés, en comparaison des résultats de mes méthodes, notamment entre 25 et 50 Hz.

Afin de tester le lien entre ces résultats expérimentaux et la faible quantité de données de la période pré-stimulus des données expérimentales, des cas supplémentaires ont été ajoutés aux tests sur le jeu de données artificielles univariées, en restreignant la quantité de données accessibles dans la période pré-stimulus pour les méthodes ERSP. Le résultat inattendu de ces tests est que les méthodes ERSP voient leur taux de faux positifs s'accroître d'autant plus que la période pré-stimulus est courte, entraînant une hausse de la proportion de fausses découvertes.

Une possible explication serait une mauvaise distribution des valeurs-p quand la période pré-stimulus est trop courte. En effet, la distribution théorique des valeurs-p sous l'hypothèse nulle est une distribution uniforme sur l'intervalle $[0; 1]$, mais dans le cas d'une approximation par bootstrap de la distribution de la statistique de test sous l'hypothèse nulle, la distribution des valeurs-p peut s'éloigner de la distribution théorique [Davison and Hinkley, 1997]. Cette hypothèse est vérifiable en observant la distribution des valeurs-p calculées par la méthode ERSP dans la période pré-stimulus. En effet, le test statistique utilisé avec la méthode ERSP est construit de sorte que la période pré-stimulus corresponde à l'hypothèse nulle. Alors, pour n'importe quelle valeur de α, la probabilité d'obtenir une valeur-p inférieure ou égale à α dans la période pré-stimulus doit valoir α :

$$P(\text{valeur-p} \leq \alpha | \mathcal{H}_0) = \alpha. \tag{5.11}$$

Les figures 5.12 et 5.13 représentent des estimations de cette probabilité pour le cas $\alpha = 0{,}05$ dans les jeux de données artificielles univariées et le jeu de données expérimentales. Dans les deux cas, la proportion de valeurs-p inférieures à 0,05 est mal calibrée et ne correspond pas à la fréquence attendue de 0,05, même approximativement. Pour les jeux de

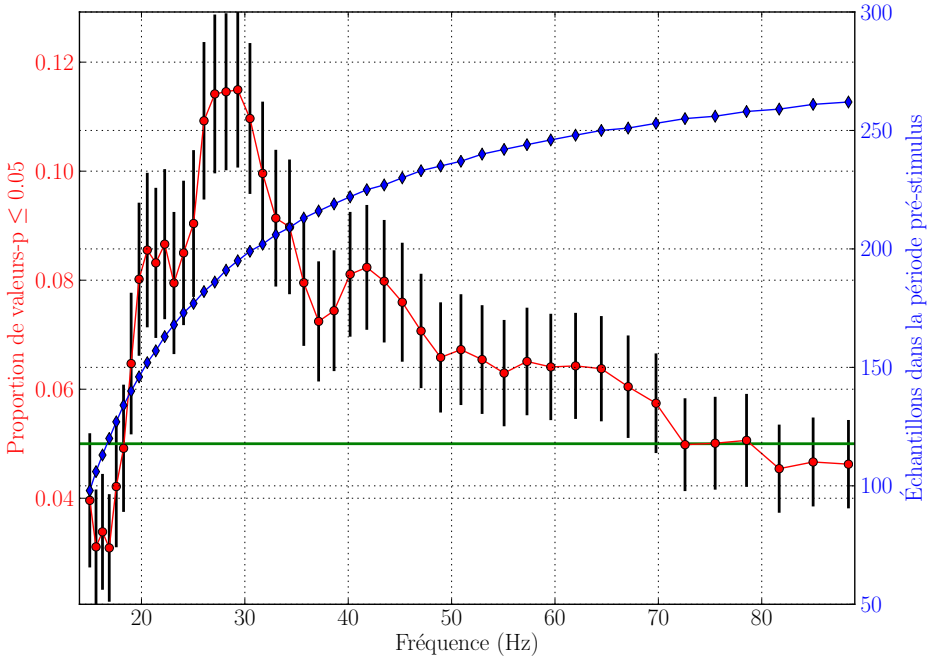

FIGURE 5.13. – Proportion des valeurs-p inférieures à 0,05 parmi celles calculées dans la période pré-stimulus du jeu de données expérimentales, en fonction de la bande de fréquence. Un intervalle de confiance à 99 % accompagne chaque proportion estimée. Le nombre de points temporels disponibles dans la période pré-stimulus est aussi reporté en fonction de la bande de fréquence. La ligne verte est un repère visuel pour la valeur de seuil à $\alpha = 0{,}05$.

méthode	étape	temps
$\text{ERSP}_{\%}$	ondelettes	1 min 58 s
	valeurs-p	26 min 7 s
ERSP_{FT}	ondelettes	1 min 58 s
	valeurs-p	26 min 40 s
méthode 1	ondelettes	16 s
	détection	5 j 12 h 32 min 51 s
	validation	58 min 16 s
méthode 2	ondelettes	2 min 33 s
	détection	7 h 25 min 58 s
	validation fréquentiste	42 min 44 s
	validation bayésienne	3 j 12 h 44 min 33 s

Tableau 5.1. – Temps de calcul des méthodes ERSP ($\text{ERSP}_{\%}$ et ERSP_{FT}), de la méthode d'analyse de synchronisations cooccurrentes (méthode 1) et de la méthode d'analyse de synchronisations univariées (méthode 2) appliquées aux données expérimentales. Voir le texte pour la configuration matérielle utilisée.

données artificielles avec zéro ou quatre enregistrements aberrants, une faible quantité de points temporels dans la période pré-stimulus entraîne une hausse des valeurs-p inférieures à 0,05, d'où une augmentation du nombre de faux positifs lorsque le seuil de signification du test statistique est fixé à $\alpha = 0{,}05$. Le même phénomène est visible pour les valeurs-p calculées dans la période pré-stimulus des données expérimentales entre 20 et 70 Hz, avec un biais maximal autour de 30 Hz [4]. Ceci confirme les réserves émises vis-à-vis des nombreux épisodes de désynchronisation détectés autour de 30 Hz qui ont une plus grande probabilité d'être de faux positifs.

Une solution envisageable pour atténuer ce problème est d'effectuer un *double bootstrap* sur la statistique de test pour rétablir la distribution uniforme des valeurs-p [Davison and Hinkley, 1997]. L'inconvénient majeur de cette approche est l'accroissement considérable de la quantité d'échantillons à générer aléatoirement pour calculer les valeurs-p.

5.4.2. Temps de calcul et complexités algorithmiques

Un dernier point à examiner concernant mes méthodes d'analyse est leur temps de calcul. Le tableau 5.1 expose les temps de calcul des différentes étapes de mes méthodes, appliquées aux données expérimentales. Les calculs ont été effectués sur un ordinateur équipé de 8 Go de mémoire vive et d'un processeur Xeon E5620 cadencé à 2,4 GHz et disposant de quatre cœurs de calcul pouvant exécuter deux threads en parallèle. Mes

4. À cause de la gestion des effets de bord de la transformation en ondelettes continue, la période pré-stimulus disponible est d'autant plus réduite que la fréquence étudiée est basse.

méthodes d'analyse comprenant un grand nombre de calculs totalement indépendants les uns des autres, par exemple sur les différentes bandes de fréquence du signal étudié, elles tirent aisément parti de cette architecture parallèle.

La lenteur avérée de mes méthodes peut s'expliquer par deux causes principales :

- les choix technologiques concernant l'implémentation des méthodes d'analyse,
- la complexité algorithmique des méthodes d'analyse.

Concernant les choix technologiques, l'ensemble des programmes et bibliothèques mises en jeu par mes méthodes utilisent le langage Python et les bibliothèques de calcul scientifique Numpy et Scipy. L'utilisation de Python avec ces bibliothèques riches en fonctionnalités permet un prototypage rapide, avec un coût en performance minime tant que la majorité du code repose sur des appels aux fonctions optimisées de ces bibliothèques. Or mon implémentation de l'algorithme VMP consiste en une bibliothèque Python introduisant une couche d'abstraction pour rendre aisés les calculs de modèles bayésiens complexes. Ainsi, si ma bibliothèque de calcul est une source de lenteur dans mes méthodes d'analyse, c'est elle qui m'a permis de mettre rapidement au point des modèles aussi élaborés que le modèle de mélange hiérarchique servant à la validation bayésienne de ma seconde méthode d'analyse. Une ré-implémentation de cette bibliothèque dans un langage compilé tel que le langage C permettrait de mesurer l'impact réel de l'implémentation sur les performances des méthodes d'analyse développées.

Pour l'analyse de la complexité algorithmique de mes méthodes, seules les goulots d'étranglement des méthodes seront détaillés. Il s'agit de l'étape de détection pour les deux méthodes ainsi que de l'étape de validation bayésienne pour la méthode d'analyse de synchronisations univariées. Toutes ces étapes sont des étapes d'inférence de modèles bayésiens.

Dans la méthode d'analyse de synchronisations cooccurrentes, un modèle de mélange gaussien est calculé pour chaque point temps-fréquence de chaque enregistrement. La complexité « grand-O » de l'étape de détection est $O(T \times F \times R \times S \times O(\text{GMM}))$ où :

- T est le nombre de points temporels maximum du signal entre le stimulus et la réponse,
- F est le nombre de bandes de fréquence étudiées,
- R est le nombre d'enregistrements,
- S est le nombre d'initialisations aléatoires pour calculer le meilleur modèle de mélange gaussien,
- $O(\text{GMM})$ est la complexité du calcul de la probabilité a posteriori du modèle de mélange gaussien avec l'algorithme VMP.

La complexité d'une itération de l'algorithme VMP pour l'inférence du modèle de mélange est linéaire vis-à-vis de la quantité de données observées et du nombre de composantes gaussiennes K. Le nombre maximal d'itérations I étant fixé, la complexité de l'inférence du modèle bayésien de mélange vaut $O(K \times D \times I)$, avec D le nombre d'électrodes. La complexité de l'étape de détection est donc $O(T \times F \times R \times S \times K \times D \times I)$.

De manière similaire, dans le cas de la méthode d'analyse de synchronisations univariées, l'étape de détection emploie un modèle de mélange ricien à deux composantes pour chaque bande de fréquence de chaque enregistrement de chaque électrode. Dès lors, la complexité de l'inférence du modèle de mélange ricien étant $O(T \times I)$, la complexité de cette étape

			méthode 1				méthode 2	
R	D	I	T	F	K	S	T	F
40	17	10 000	324	50	3	50	1296	46

Tableau 5.2. – Tailles des données et des modèles utilisés pour l'utilisation de la méthode d'analyse de synchronisations cooccurrentes (méthode 1) et de la méthode d'analyse de synchronisations univariées (méthode 2), sur le jeu de données expérimentales. Voir le texte pour la signification de chaque colonne.

vaut $O(F \times R \times D \times O(\text{RMM})) = O(F \times R \times D \times T \times I)$. Il est intéressant de noter que cette complexité ne dépend pas du nombre de composantes K du modèle de mélange, comme il est fixé à deux pour tout jeu de données, ni d'un nombre d'initialisations S, comme il n'y a qu'une seule initialisation pour chaque modèle de mélange, avec l'heuristique de la médiane.

Pour l'étape de validation bayésienne, un modèle de mélange hiérarchique est inféré pour chaque bande de fréquence de chaque électrode. Aussi la complexité de l'étape de validation égale à $O(F \times D \times O(\mathcal{M}))$, où $O(\mathcal{M})$ est la complexité de l'inférence du modèle hiérarchique. La complexité de chaque itération de l'algorithme VMP vaut $O(R \times T)$ d'où $O(\mathcal{M}) = O(R \times T \times I)$. La complexité grand-$O$ de l'étape de validation bayésienne est donc $O(F \times D \times R \times T \times I)$, c'est-à-dire la même complexité que l'étape de détection.

Le tableau 5.2 résume les tailles des données et des modèles pour les deux méthodes appliquées au jeu de données expérimentales. Il est notable que le nombre de points temporels utilisés pour la méthode d'analyse de synchronisations cooccurrentes a été réduit via un sous-échantillonnage de 1 000 Hz à 250 Hz, ceci afin de garder des temps de calcul raisonnables.

5.5. Résumé

Ce chapitre a présenté des applications des méthodes d'analyse développées à des jeux de données artificielles et expérimentales. Ces résultats ont été comparés avec ceux obtenus via une méthode ERSP standard à modèle multiplicatif et via une méthode ERSP plus robuste avec une normalisation simple essai.

L'application à un jeu de données artificielles a permis de mettre en relief les bonnes performances de mes méthodes vis-à-vis de données qui répondent aux hypothèses sur lesquelles mes méthodes sont bâties. De plus, l'introduction d'enregistrements aberrants au sein des données artificielles a montré la robustesse de mes méthodes face à ce type de contamination des jeux de données. En comparaison, les performances de la méthode ERSP standard à modèle multiplicatif se dégradent rapidement avec ce type de contamination. La méthode ERSP robuste, quant à elle, conserve des performances similaires à mes méthodes même sur les données très bruitées, tant que la quantité de données pré-stimulus est suffisante. En effet, un manque de données dans la période pré-stimulus entraîne une augmentation de la proportion de fausses découvertes de cette méthode.

Le jeu de données expérimentales consiste en des enregistrements intracrâniens de potentiels de champs locaux réalisés sur des macaques effectuant une tâche visuo-motrice. Les résultats obtenus avec mes méthodes correspondent à ceux obtenus avec la méthode ERSP robuste, mais sont plus parcimonieux. La faible quantité de données disponible dans la période pré-stimulus laisse à penser que la quantité de faux positifs parmi les résultats la méthode ERSP est anormalement élevée. Une partie des résultats communs à toutes les méthodes est supportée par la littérature.

Chapitre 6.

Conclusions et perspectives

> La vérité vaut bien qu'on passe quelques années sans la trouver.
>
> *Jules Renard – Journal du 7 février 1901*

Le but de cette thèse est de créer des méthodes d'analyse capables de détecter des phénomènes de synchronisation et de désynchronisation au sein de signaux électro-corticaux, ces méthodes reposant sur des hypothèses différentes et théoriquement moins contraignantes que celles utilisées dans les méthodes classiques. Dans ce dernier chapitre, je conclurai sur les contributions directes de ma thèse vis-à-vis de cet objectif ainsi que sur les développements annexes ayant servi ce dernier mais ayant une portée plus générale. Enfin, je suggérerai plusieurs pistes d'améliorations de mon travail, pour mieux remplir l'objectif initial et même dépasser celui-ci.

6.1. Conclusions

6.1.1. Rappel des motivations

Le postulat qui a présidé à ces travaux de thèse est que les techniques standards d'analyse de synchronisations et désynchronisations au sein de signaux corticaux appellent des hypothèses fortes qui ne sont pas nécessaires pour caractériser les mêmes phénomènes. En effet, les méthodes de type ERSP se fondent sur l'hypothèse que les signaux issus de divers enregistrements contiennent un signal d'intérêt invariant en amplitude entre différents enregistrements et dont seules les fluctuations statistiquement différentes par rapport à une période pré-stimulus constituent des indices de phénomènes de synchronisation ou désynchronisation. Cette approche ne laisse ainsi pas de place pour une variabilité du signal au sein de chaque enregistrement. De même, la période pré-stimulus n'est pas considérée comme une période où des phénomènes d'intérêt peuvent se dérouler mais est reléguée à un état de bruit. Ainsi j'ai cherché à répondre à ces limitations théoriques en mettant au point des méthodes qui ne font pas appel aux hypothèses des méthodes standards.

6.1.2. Résumé des contributions

Mes contributions ont été exposées aux chapitres 3, 4 et 5. La teneur de celles-ci, résumée chapitre par chapitre, est la suivante.

Chapitre 3 : La distribution du module des coefficients d'ondelette d'un signal contaminé par un bruit blanc gaussien additif est une distribution de Rice. Pour modéliser ce type d'observation, j'ai mis au point un modèle bayésien ricien. Le calcul numérique des probabilités marginales et a posteriori de ce modèle est possible grâce à une adaptation de l'algorithme VMP pour la distribution de Rice, employant une approximation variationnelle locale. Pour discriminer plusieurs signaux bruités à partir du module de leurs coefficients d'ondelettes, j'ai mis au point un modèle bayésien de mélange ricien dont l'inférence est aussi menée grâce à une adaptation spécifique de l'algorithme VMP pour la distribution de mélange ricien.

Chapitre 4 : À partir des modèles bayésiens de mélange gaussien et ricien, j'ai mis au point deux méthodes d'analyse de synchronisations corticales. Ces deux méthodes se décomposent en une étape de détection des phénomènes de synchronisations dans chaque enregistrement, suivie d'une étape de validation de ces phénomènes par une mesure prenant en compte les résultats de tous les enregistrements. Ma première méthode d'analyse détecte des sous-groupes d'électrodes en chaque point temps-fréquence à l'aide de modèles de mélange gaussien, puis les valide si ces sous-groupes se retrouvent dans un même voisinage temps-fréquence dans différents enregistrements. Les électrodes faisant partie d'un même sous-groupe validé présentent donc un épisode de synchronisation ou désynchronisation qui est dit « cooccurrent », car il a lieu au même point temps-fréquence pour toutes ces électrodes. Ma seconde méthode d'analyse modélise chaque bande de fréquence du scalogramme de chaque électrode dans chaque enregistrement par un modèle de mélange ricien à deux composantes. L'appartenance à l'une des deux composantes en un point temps-fréquence symbolise un potentiel épisode de synchronisation ou désynchronisation, selon qu'il s'agit de la composante haute ou basse. La validation repose alors sur la probabilité d'observer le même type d'épisode potentiel de synchronisation ou de désynchronisation en un même point temps-fréquence de l'ensemble des enregistrements. Deux variantes de validation ont été testées, l'une employant des tests d'hypothèses fréquentistes et l'autre un modèle bayésien hiérarchique.

Chapitre 5 : À l'aide de jeux de données artificielles, j'ai pu mettre en relief que mes méthodes sont capables de détecter de manière robuste des épisodes de synchronisation et de désynchronisation au sein de données, dès lors que ces données se conforment aux hypothèses de mes modèles. Dans ce cadre, mes méthodes obtiennent des résultats aussi bons que les meilleures méthodes de type ERSP. Appliquées à des données expérimentales, mes méthodes d'analyse réussissent à extraire des informations pertinentes, de même nature que les méthodes standards. Cependant, les hypothèses de ma première méthode d'analyse se révèlent alors très contraignantes et limitent fortement la quantité de résultats que peut extraire cette méthode. Ma seconde méthode ne présente pas cet inconvénient et évite même un écueil des méthodes standards qui semblent produire sur ce jeu de données de nombreux faux positifs.

6.1.3. Conclusions sur les méthodes et les modèles

Mes méthodes d'analyse paraissent remplir l'objectif central de mes travaux de thèse. Cependant, il est important de faire la critique des hypothèses que j'ai placées à la base de mes méthodes. Ces hypothèses sont-elles réellement moins contraignantes que celles des méthodes standards ? Concernant ma première méthode d'analyse, l'hypothèse de l'existence de groupes d'électrodes présentant des synchronisations cooccurrentes, bien qu'exploitable sur des données expérimentales, se révèle trop restrictive et ne permet pas d'embrasser une quantité de phénomènes aussi large que les méthodes standards. Néanmoins elle a le mérite de présenter une approche nouvelle pour traiter les signaux cérébraux, en s'intéressant spécifiquement à l'aspect cooccurrent des phénomènes étudiés. Dans le cas de ma seconde méthode, l'hypothèse de la décomposabilité du signal en deux états, haut et bas, dans une bande de fréquence s'avère au moins aussi puissante que les hypothèses classiques. Cette hypothèse apparaît comme moins contraignante car elle laisse une plus grande variabilité au signal entre les enregistrements.

Au-delà du choix des hypothèses minimales pour mes méthodes, ma démarche pour la mise au point de ces méthodes a aussi été de minimiser la quantité de paramètres ad hoc. Il s'agit bien sûr de méthodes paramétriques, mais les valeurs des paramètres ne doivent influer que marginalement sur les résultats. Le choix de modèles bayésiens dont les hyperparamètres fixent des probabilités a priori vagues fait partie de cette démarche. Pourtant ma première méthode d'analyse est dépendante de valeurs de seuils, qui, bien que choisies de manière argumentée, restent des valeurs arbitraires. Dans ma seconde méthode d'analyse, le même type de reproche peut être fait pour le choix des seuils de signification des tests d'hypothèses fréquentistes. En revanche, cette seconde méthode d'analyse avec la validation bayésienne dépasse ces limitations, étant donné qu'elle n'emploie que des modèles bayésiens aux paramètres peu sensibles ainsi qu'une fonction de coût 0-1 pour estimer les épisodes de synchronisation. Le prix à payer pour cette robustesse est la mise au point d'un modèle bayésien hiérarchique complexe et difficile à calculer. L'étape de détection à l'aide de modèles de mélange ricien ne constitue alors qu'une aide à l'initialisation de ce modèle hiérarchique pour calculer sa probabilité a posteriori. Un avantage théorique de cette approche est que, mise à part la transformée en ondelettes continue, l'intégralité de la méthode réside dans le même cadre théorique, apportant ainsi une plus grande cohérence à la méthode.

6.2. Perspectives

À l'issue de mes travaux de thèse, le modèle bayésien hiérarchique que j'ai mis au point constitue une formidable base pour développer des améliorations en vue d'étendre le spectre des phénomènes détectables par ce modèle. Celui-ci et ses dérivés étant complexes, des améliorations de l'algorithme VMP sont aussi à envisager pour rendre pratique leur emploi.

6.2.1. Amélioration des modèles

Une caractéristique des données qui n'est pas modélisée dans mon modèle hiérarchique est la corrélation du module des coefficients d'ondelette dans un même voisinage temps-fréquence. Ainsi une modification possible consisterait à modéliser en chaque point temps-fréquence le module de tous les coefficients du voisinage avec une distribution ricienne multivariée [Beaulieu and Hemachandra, 2010].

Un aspect des données plus crucial à modéliser concerne la différence dans le temps de réponse à chaque enregistrement. Il s'agit d'un aspect essentiel de la variabilité du signal qui n'est pas pris en compte ni dans les méthodes standards ni dans mes méthodes. Pour remédier à cette lacune, une idée serait d'introduire dans le modèle hiérarchique une chaîne de Markov gauche-droite, pour former un modèle de Markov caché. Plus précisément, pour s'assurer que chaque état de la chaîne de Markov ait une durée minimale, il faudrait considérer un modèle de Markov caché à états étendus [Russell and Cook, 1987, Johnson, 2005].

En suivant la même logique de construction de modèle que pour le modèle hiérarchique, il serait aussi possible de construire un modèle pour détecter les épisodes de synchronisation de phase entre différents enregistrements. Au lieu de la distribution de Rice, la distribution de von Mises serait utilisée pour modéliser les phases [Guttorp and Lockhart, 1988]. Le même type d'approximation variationnelle locale est utilisable avec cette distribution circulaire.

6.2.2. Amélioration des algorithmes d'inférence

Même si les approches bayésiennes variationnelles, dont fait partie l'algorithme VMP, sont réputées plus rapides que d'autres techniques d'inférence telles que les méthodes de Monte Carlo par chaînes de Markov, leur convergence peut s'avérer lente, d'autant plus que les variables cachées du modèle en jeu présentent une dépendance forte. Pour remédier à ce problème, il pourrait être intéressant d'exploiter des techniques d'accélération d'algorithmes de point fixe ayant fait leurs preuves sur un algorithme proche tel que l'algorithme d'espérance-maximisation [Salakhutdinov and Roweis, 2003, Huang et al., 2005].

Une autre difficulté de l'algorithme VMP consiste à bien l'initialiser pour qu'il converge vers la meilleure solution possible. Relancer l'algorithme avec plusieurs initialisations est possible pour des petits modèles mais n'est pas faisable en pratique pour de grands modèles. Une autre solution déjà exploitée dans les approches bayésiennes variationnelles est d'ajouter un mécanisme de recuit déterministe [Beal, 2003, Katahira et al., 2008]. Cette solution pourrait être incorporée directement dans l'algorithme VMP.

Enfin, l'algorithme VMP permet de calculer une probabilité a posteriori approximée par une distribution factorisée. Pour rendre plus précis les résultats issus de ces calculs, la modification opérée par Jaakkola et Jordan sur l'approximation de champ moyen pourrait être transposée à l'algorithme VMP [Jaakkola and Jordan, 1998]. Cette modification consiste à remplacer l'approximation via une distribution factorisée par une approximation via un mélange de distributions factorisées.

6.2.3. Extension du domaine d'application

Si le domaine d'application de mes méthodes d'analyse est restreint aux signaux électrocorticaux, il n'en est pas de même des modèles statistiques développés pour celles-ci. Ainsi les modèles ricien et de mélange ricien présentent un intérêt pour des domaines plus vastes, pour estimer les paramètres de l'enveloppe de n'importe quel signal ou discriminer différents signaux sur les caractéristiques de leurs enveloppes. De même, les équations mises au point pour utiliser l'algorithme VMP avec les distributions de Rice et de mélange ricien trouveront un intérêt dans n'importe quel modèle statistique ayant besoin de rendre compte d'un bruit ricien. Par exemple, mes travaux pourraient s'appliquer à des modèles statistiques servant à réduire le bruit ricien contaminant des images IRM [Gudbjartsson and Patz, 1995].

Annexe A.

Variational Message Passing : messages et équations de ré-estimations

Cette annexe liste les distributions et les nœuds déterministes mis en jeu dans mes travaux de thèse avec l'algorithme VMP. Pour chaque distribution, la densité de probabilité est présentée sous sa forme usuelle puis sous la forme propre à une famille exponentielle, telle que présentée à la section 2.2.1. Pour chaque distribution et nœud déterministe, les messages vers les nœuds enfants et parents nécessaires pour l'inférence de mes modèles graphiques probabilistes avec l'algorithme VMP sont détaillés, ainsi que les contraintes induites par cet algorithme sur la distribution des nœuds parents. Sauf mention contraire, les informations générales sur les distributions sont tirées des écrits de Beal, Bishop, Miskin et Winn [Beal, 2003, Bishop, 2006, Miskin, 2000, Winn, 2003].

A.1. Distribution gaussienne

La distribution gaussienne, aussi appelée distribution *normale*, est définie par la densité de probabilité suivante :

$$p(x|\mu,\sigma) = \mathcal{N}(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \tag{A.1}$$

où μ est le paramètre de moyenne et σ^2 le paramètre de variance. Il est possible de formuler cette distribution avec un paramètre de *précision* λ au lieu de la variance, tel que $\lambda = 1/\sigma^2$:

$$p(x|\mu,\lambda) = \mathcal{N}(x|\mu,\lambda) = \sqrt{\frac{\lambda}{2\pi}} \exp\left(-\frac{\lambda}{2}(x-\mu)^2\right). \tag{A.2}$$

La réécriture de la densité de probabilité sous la forme propre a une famille exponentielle donne :

$$\ln p(x|\mu,\lambda) = \underbrace{\begin{pmatrix}\lambda\mu \\ -\frac{\lambda}{2}\end{pmatrix}^\top}_{\eta_x(\mu,\lambda)} \cdot \underbrace{\begin{pmatrix}x \\ x^2\end{pmatrix}}_{\mathrm{u}_x(x)} \underbrace{-\frac{1}{2}\ln 2\pi}_{f_x(x)} \underbrace{+\frac{1}{2}(\ln\lambda - \lambda\mu^2)}_{g_x(\mu,\lambda)}. \tag{A.3}$$

Contraintes et message pour le paramètre μ : Une réécriture de $\ln p(x)$ met en valeur le type de distribution a priori et le message pour l'algorithme VMP :

$$\ln p(x|\mu,\lambda) = \underbrace{\begin{pmatrix} \lambda x \\ -\frac{\lambda}{2} \end{pmatrix}^{\top}}_{\ell_{x,\mu}(x,\lambda)} \cdot \underbrace{\begin{pmatrix} \mu \\ \mu^2 \end{pmatrix}}_{\mathrm{u}_\mu(\mu)} + h_{x,\mu}(x,\lambda), \tag{A.4}$$

d'où le message vers μ

$$m_{X\to\mu} = \mathbb{E}_{q(\lambda,x)}[\ell_{x,\mu}(x,\lambda)] = \begin{pmatrix} \mathbb{E}_{q(\lambda)}[\lambda]\mathbb{E}_{q(x)}[x] \\ -\frac{1}{2}\mathbb{E}_{q(\lambda)}[\lambda] \end{pmatrix}. \tag{A.5}$$

La distribution du paramètre μ doit faire partie d'une famille exponentielle dont le vecteur de statistiques naturelles est de la forme $\mathrm{u}_\mu(\mu) = (\mu, \mu^2)$. Il s'agit typiquement d'une distribution gaussienne ou de mélange gaussien.

Contraintes et message pour le paramètre λ : De la même manière, $\ln p(x)$ peut être réécrit pour le paramètre λ :

$$\ln p(x|\mu,\lambda) = \underbrace{\begin{pmatrix} -\frac{1}{2}(x^2 - 2x\mu + \mu^2) \\ \frac{1}{2} \end{pmatrix}^{\top}}_{\ell_{x,\lambda}(x,\mu)} \cdot \underbrace{\begin{pmatrix} \lambda \\ \ln\lambda \end{pmatrix}}_{\mathrm{u}_\lambda(\lambda)} + h_{x,\lambda}(x,\mu), \tag{A.6}$$

d'où le message vers λ

$$m_{X\to\lambda} = \mathbb{E}_{q(\mu,x)}[\ell_{x,\lambda}(x,\mu)] = \begin{pmatrix} -\frac{1}{2}(\mathbb{E}_{q(x)}[x^2] - 2\mathbb{E}_{q(x)}[x]\mathbb{E}_{q(\mu)}[\mu] + \mathbb{E}_{q(\mu)}[\mu^2]) \\ \frac{1}{2} \end{pmatrix}. \tag{A.7}$$

Le paramètre λ doit avoir une distribution faisant partie d'une famille exponentielle dont le vecteur de statistiques naturelles est de la forme $\mathrm{u}_\lambda(\lambda) = (\lambda, \ln\lambda)$. Il s'agit typiquement d'une distribution gamma.

Message pour les nœuds enfants : Sous l'hypothèse de champ moyen, la distribution a posteriori calculée par l'algorithme VMP est une distribution gaussienne de paramètres μ^* et λ^*. Le message vers un nœud enfant Ch s'écrit alors :

$$m_{X\to Ch} = \begin{pmatrix} \mathbb{E}_{q(x)}[x] \\ \mathbb{E}_{q(x)}[x^2] \end{pmatrix} = \begin{pmatrix} \mu^* \\ (\mu^*)^2 + \frac{1}{\lambda^*} \end{pmatrix}. \tag{A.8}$$

A.2. Distribution gaussienne rectifiée

La distribution gaussienne rectifiée est similaire à la distribution gaussienne, exception faite qu'elle n'est définie que pour les réels positifs :

$$p(x|\mu,\lambda) = \mathcal{N}^R(x|\mu,\lambda) = \begin{cases} \dfrac{\sqrt{2\lambda}}{\sqrt{\pi}\,\mathrm{erfc}(-\mu\sqrt{\lambda/2})} \exp\left(-\frac{\lambda}{2}(x-\mu)^2\right) & \text{si } x \geq 0, \\ 0 & \text{sinon,} \end{cases} \tag{A.9}$$

où erfc est la fonction d'erreur complémentaire,

$$\mathrm{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty \exp(-t^2)dt. \tag{A.10}$$

La distribution gaussienne rectifiée dispose des mêmes vecteurs de paramètres et statistiques naturelles $\eta_x(\mu,\lambda)$ et $\mathrm{u}_x(x)$ que la distribution gaussienne. Elle diffère cependant au niveau de $f_x(x)$ et de la constante de normalisation $g_x(\mu,\lambda)$:

$$f_x(x) = \begin{cases} -\frac{1}{2}\ln\frac{\pi}{2} & \text{si } x \geq 0 \\ -\infty & \text{sinon} \end{cases} \quad \text{et} \quad g_x(\mu,\lambda) = \frac{1}{2}(\ln\lambda - \lambda\mu^2) - \ln\mathrm{erfc}(-\mu\sqrt{\lambda/2}). \tag{A.11}$$

Message pour les nœuds enfants : La distribution a posteriori obtenue avec l'approximation par champ moyen est aussi une distribution gaussienne rectifiée, de paramètres μ^* et λ^*. Le message vers un nœud enfant Ch est alors défini comme suit :

$$m_{X\to Ch} = \begin{pmatrix} \mathbb{E}_{q(x)}[x] \\ \mathbb{E}_{q(x)}[x^2] \end{pmatrix} = \begin{pmatrix} \mu^* + \sqrt{\dfrac{2}{\pi\lambda^*}} \dfrac{1}{\mathrm{erfcx}(-\mu^*\sqrt{\lambda^*/2})} \\ (\mu^*)^2 + \dfrac{1}{\lambda^*} + \sqrt{\dfrac{1}{\pi\lambda^*}} \dfrac{\mu^*}{\mathrm{erfcx}(-\mu^*\sqrt{\lambda^*/2})} \end{pmatrix}, \tag{A.12}$$

où erfcx est la fonction d'erreur complémentaire normalisée, $\mathrm{erfcx}(x) = \exp(x^2) \times \mathrm{erfc}(x)$.

Il peut se produire des erreurs numériques dans les cas où la quantité $\mathrm{erfcx}(-\mu^*\sqrt{\lambda^*/2})$ tend vers 0 ou vers $+\infty$ [Miskin, 2000]. Dans le premier cas, la distribution gaussienne rectifiée tend vers une distribution exponentielle, et les espérances tendent vers

$$\mathbb{E}_{q(x)}[x] \to -\frac{1}{\mu^*\lambda^*} \quad \text{et} \quad \mathbb{E}_{q(x)}[x^2] \to \frac{2}{(\mu^*\lambda^*)^2}. \tag{A.13}$$

Dans le second cas, on retrouve les mêmes espérances que celles d'une distribution gaussienne usuelle.

A.3. Distribution gamma

La distribution gamma est une distribution définie sur les réels positifs ayant pour densité :

$$p(x|a,b) = \mathcal{G}(x|a,b) = \frac{1}{\Gamma(a)} b^a x^{(a-1)} \exp(-bx), \tag{A.14}$$

où a et b sont strictement positifs et représentent respectivement le paramètre de *forme* et l'inverse du paramètre d'*échelle*.

La densité de probabilité peut s'écrire sous la forme propre aux familles exponentielles :

$$\ln p(x|a,b) = \underbrace{\begin{pmatrix} -b \\ a-1 \end{pmatrix}^{\top}}_{\eta_x(a,b)} \cdot \underbrace{\begin{pmatrix} x \\ \ln x \end{pmatrix}}_{\mathrm{u}_x(x)} + \underbrace{a \ln b - \ln \Gamma(a)}_{g_x(a,b)}. \tag{A.15}$$

Message pour les nœuds enfants : La distribution a posteriori pour un nœud ayant une distribution gamma est aussi une distribution gamma, de paramètres a^* et b^*, avec l'hypothèse de champ moyen. Le message vers un nœud enfant Ch a pour formule :

$$m_{X \to Ch} = \begin{pmatrix} \mathbb{E}_{q(x)}[x] \\ \mathbb{E}_{q(x)}[\ln x] \end{pmatrix} = \begin{pmatrix} \frac{a^*}{b^*} \\ \psi(a^*) - \ln b^* \end{pmatrix}, \tag{A.16}$$

où ψ est la fonction digamma, définie comme la dérivée de la fonction gamma,

$$\psi(x) = \frac{d \ln \Gamma(x)}{d\, x}. \tag{A.17}$$

A.4. Distribution catégorique

Soit Z un vecteur aléatoire ayant une distribution catégorique à K dimensions, sa probabilité est caractérisée comme suit :

$$p(z|\pi) = \Pi_{k=1}^{K} \pi_k^{z_k}, \tag{A.18}$$

où le paramètre $\pi = (\pi_1, \ldots, \pi_K)$ est un vecteur défini tel que

$$\pi_k \geq 0 \ \forall k \in [1, K] \quad \text{et} \quad \sum_{k=1}^{K} \pi_k = 1. \tag{A.19}$$

Le vecteur $z = (z_1, \ldots z_K)$ contient $K - 1$ éléments égaux à zéro et une unique valeur valant un.

L'écriture de cette distribution sous la forme propre à une famille exponentielle est :

$$\ln p(z|\pi) = \underbrace{\begin{pmatrix} \ln \pi_1 \\ \vdots \\ \ln \pi_K \end{pmatrix}^{\top}}_{\eta_x(\pi)} \cdot \underbrace{\begin{pmatrix} z_1 \\ \vdots \\ z_K \end{pmatrix}}_{\mathrm{u}_x(x)}. \tag{A.20}$$

Contraintes et messages pour le paramètre π : Il est possible de déduire directement de l'équation (A.20) que $\ell_x(x) = \mathrm{u}_x(x)$ et $\mathrm{u}_\pi(\pi) = \eta_x(\pi)$. Le message pour le paramètre π s'écrit donc :

$$m_{Z\to\pi} = \mathbb{E}_{q(x)}[\ell_x(x)] = \begin{pmatrix} \mathbb{E}_{q(z)}[z_1] \\ \vdots \\ \mathbb{E}_{q(z)}[z_K] \end{pmatrix}. \tag{A.21}$$

Ainsi la distribution du paramètre π a comme vecteur de statistiques naturelles $\mathrm{u}_\pi(\pi) = (\ln \pi_1, \ldots, \pi_K)$ et doit respecter les contraintes définies à l'équation (A.19). La distribution de Dirichlet répond à ces critères.

Message pour les nœuds enfants : Dans le cas où Z est une variable cachée, sa distribution a posteriori sous l'hypothèse de champ moyen est aussi une distribution catégorique, de paramètre π^*. Le message vers un nœud enfant Ch vaut alors :

$$m_{Z\to Ch} = \begin{pmatrix} \mathbb{E}_{q(z)}[z_1] \\ \vdots \\ \mathbb{E}_{q(z)}[z_K] \end{pmatrix} = \begin{pmatrix} \pi_1^* \\ \vdots \\ \pi_K^* \end{pmatrix}. \tag{A.22}$$

A.5. Distribution de Dirichlet

La distribution de Dirichlet est une distribution multivariée à K dimensions définie par la densité de probabilité qui suit :

$$p(x|\alpha) = \mathcal{D}ir(x|\alpha) = \frac{1}{C(\alpha)} \Pi_{k=1}^K x_k^{\alpha_k - 1}, \tag{A.23}$$

avec la constante de normalisation

$$C(\alpha) = \frac{\Pi_{k=1}^K \Gamma(a_k)}{\Gamma(\sum_{k=1}^K a_k)}. \tag{A.24}$$

Le paramètre $\alpha = (\alpha_1, \ldots, \alpha_K)$ est un vecteur dont tous les éléments α_k sont strictement positifs. Le vecteur $x = (x_1, \ldots, x_K)$ est sujet aux contraintes

$$x_k \geq 0 \ \forall k \in [1, K] \quad \text{et} \quad \sum_{k=1}^K x_k = 1. \tag{A.25}$$

La réécriture de la densité de probabilité sous la forme propre à une famille exponentielle donne la formule :

$$\ln p(x|\alpha) = \underbrace{\begin{pmatrix} \alpha_1 - 1 \\ \vdots \\ \alpha_K - 1 \end{pmatrix}^\top}_{\eta_x(\alpha)} \cdot \underbrace{\begin{pmatrix} \ln x_1 \\ \vdots \\ \ln x_K \end{pmatrix}}_{\mathrm{u}_x(x)} + \underbrace{\ln \Gamma\left(\sum_{k=1}^K a_k\right) - \sum_{k=1}^K \ln \Gamma(a_k)}_{g_x(\alpha)}. \tag{A.26}$$

Message pour les nœuds enfants : Avec l'hypothèse de champ moyen, la distribution a posteriori est aussi une distribution de Dirichlet, de paramètre α^*. Le message pour un nœud enfant Ch s'écrit :

$$m_{X\to Ch} = \begin{pmatrix} \mathbb{E}_{q(x)}[\ln x_1] \\ \vdots \\ \mathbb{E}_{q(x)}[\ln x_K] \end{pmatrix} = \begin{pmatrix} \psi(\alpha_1^*) - \psi\left(\sum_{k=1}^K \alpha_k^*\right) \\ \vdots \\ \psi(\alpha_K^*) - \psi\left(\sum_{k=1}^K \alpha_k^*\right) \end{pmatrix}, \tag{A.27}$$

où ψ est la fonction digamma définie précédemment.

A.6. Distribution de mélange

Une distribution de mélanges à K composantes a pour densité de probabilité :

$$p(x|z,\boldsymbol{\theta}) = \Pi_{k=1}^K p(x|\theta_k)^{z_k}, \tag{A.28}$$

où $p(x|\theta_k)$ est la densité de probabilité d'une composante, ayant pour vecteur de paramètres θ_k. Le vecteur $z = (z_1, \ldots, z_K)$ sert à sélectionner une composante, toutes ses valeurs e_k sont nulles sauf une qui vaut un. L'ensemble $\boldsymbol{\theta} = \{\theta_k\}_{k=1}^K$ regroupe les paramètres de toutes les composantes.

Si toutes les composantes sont issues de distributions appartenant à une même famille exponentielle, alors la densité de probabilité de la distribution de mélange peut s'écrire sous la forme :

$$\ln p(x|z,\boldsymbol{\theta}) = \underbrace{\left(\sum_{k=1}^K z_k\, \eta_x'(\theta_k)\right)^\top}_{\eta_x(z,\boldsymbol{\theta})} \cdot \mathrm{u}_x(x) + f_x(x) + \underbrace{\sum_{k=1}^K z_k\, g_x'(\theta_k)}_{g_x(z,\boldsymbol{\theta})}, \tag{A.29}$$

où $\eta_x'(\theta_k)$, $\mathrm{u}_x(x)$, $f_x(x)$ et $g_x'(\theta_k)$ sont propres à la famille exponentielle des composantes. La densité de probabilité de la distribution de mélange fait alors aussi partie d'une famille exponentielle.

Contraintes et messages pour le paramètre z : La densité de probabilité de la distribution de mélange peut être reformulée pour le paramètre z de la manière suivante :

$$\ln p(x|z,\boldsymbol{\theta}) = \underbrace{\begin{pmatrix} \eta_x'(\theta_1)^\top \cdot \mathrm{u}_x(x) + g_x'(\theta_1) \\ \vdots \\ \eta_x'(\theta_K)^\top \cdot \mathrm{u}_x(x) + g_x'(\theta_K) \end{pmatrix}^\top}_{\ell_{x,z}(x,\boldsymbol{\theta})} \cdot \underbrace{\begin{pmatrix} z_1 \\ \vdots \\ z_K \end{pmatrix}}_{\mathrm{u}_z(z)} + \underbrace{f_x(x)}_{h_{x,z}(x,\boldsymbol{\theta})}, \tag{A.30}$$

d'où le message pour le paramètre z

$$m_{X\to z} = \mathbb{E}_{q(x,\boldsymbol{\theta})}[\ell_{x,z}(x,\boldsymbol{\theta})] = \begin{pmatrix} \mathbb{E}_{q(\theta_1)}[\eta'_x(\theta_1)]^\top \cdot \mathbb{E}_{q(x)}[\mathrm{u}_x(x)] + \mathbb{E}_{q(\theta_1)}[g'_x(\theta_1)] \\ \vdots \\ \mathbb{E}_{q(\theta_K)}[\eta'_x(\theta_K)]^\top \cdot \mathbb{E}_{q(x)}[\mathrm{u}_x(x)] + \mathbb{E}_{q(\theta_K)}[g'_x(\theta_K)] \end{pmatrix}. \tag{A.31}$$

Le paramètre z est défini par une distribution appartenant à une famille exponentielle dont le vecteur de statistiques naturelles est $\mathrm{u}_z(z) = (z_1, \ldots, z_K)$, avec la contrainte qu'une seule valeur z_k est égale à un, les autres valant zéro. Il s'agit usuellement d'une distribution catégorique.

Contraintes et messages pour un paramètre d'une composante : Soit le paramètre $\theta_{k,i}$ un des éléments du vecteur de paramètres θ_k de la k-ième composante de la distribution de mélange. La densité de probabilité de la distribution de mélange formulée pour ce paramètre prend la forme qui suit :

$$\ln p(x|z,\boldsymbol{\theta}) = \underbrace{\big(z_k \;\ell'_{x,i}(x, \{\theta_{k,j}\}_{j\neq i})\big)^\top}_{\ell_{x,\theta_{k,i}}(x,\boldsymbol{\theta}\backslash\{\theta_{k,i}\},z)} \cdot \mathrm{u}_i(\theta_{k,i}) + \underbrace{z_k \;h'_{x,i}(x, \{\theta_{k,j}\}_{j\neq i})}_{h_{x,\theta_{k,i}}(x,\boldsymbol{\theta}\backslash\{\theta_{k,i}\},z)}, \tag{A.32}$$

où les fonctions $\ell'_{x,i}$ et $h'_{x,i}$ sont spécifiques au i-ième paramètre de la famille exponentielle des composantes de la distribution de mélange. Le message vers le paramètre $\theta_{k,i}$ s'écrit donc :

$$\begin{aligned} m_{X\to\theta_{k,i}} &= \mathbb{E}_{q(x,\boldsymbol{\theta}\backslash\{\theta_{k,i}\},z)}[\ell_{x,\theta_{k,i}}(x,\boldsymbol{\theta}\backslash\{\theta_{k,i}\},z)] \\ &= \mathbb{E}_{q(z)}[z_k]\,\mathbb{E}_{q(x,\{\theta_{k,j}\}_{j\neq i})}[\ell'_{x,i}(x,\{\theta_{k,j}\}_{j\neq i})]. \end{aligned} \tag{A.33}$$

La contrainte sur le vecteur des statistiques naturelles $\mathrm{u}_i(\theta_{k,i})$ de la distribution du paramètre $\theta_{k,i}$ dépend de la famille exponentielle des composantes.

Message pour les nœuds enfants : La distribution a posteriori calculée avec l'algorithme VMP possède le même vecteur de statistiques naturelles $\mathrm{u}_x(x)$ que la famille exponentielle des composantes. Il s'agit donc d'une distribution de la même famille exponentielle, avec un vecteur de paramètres θ^*. Dès lors, le message vers un nœud enfant pour la distribution de mélange a la même formule que le message vers un nœud enfant calculé pour une distribution de cette famille exponentielle, en utilisant θ^* comme paramètres.

A.7. Nœud déterministe de sélection

Soient $Z = (Z_1, \ldots, Z_K)$ un vecteur aléatoire issu d'une distribution catégorique et $\boldsymbol{X}$ un ensemble de K variables aléatoires issues de distributions appartenant à la même

famille exponentielle. Une variable Y associée à une fonction déterministe f de sélection sur ces variables correspond alors à :

$$Y = f(Z, \boldsymbol{X}) = \sum_{k=1}^{K} Z_k \times X_k = \begin{pmatrix} Z_1 \\ \vdots \\ Z_K \end{pmatrix}^{\top} \cdot \begin{pmatrix} X_1 \\ \vdots \\ X_K \end{pmatrix} \tag{A.34}$$

Étant donnée la nature du vecteur aléatoire Z, cette fonction consiste à ne choisir qu'un seul des éléments de l'ensemble $\boldsymbol{X}$, suivant le tirage issu de Z. Par construction, le vecteur de statistiques naturels de Y est du même type que celui des X_k.

En suivant la méthodologie décrite par Winn [Winn, 2003], il est possible de dériver les messages pour les nœuds enfants Ch et les nœuds parents suivants :

$$\begin{aligned} m_{Y \to Ch} &= \begin{pmatrix} \mathbb{E}_{q(x_1)}[\mathrm{u}_x(x_1)]^{\top} \\ \vdots \\ \mathbb{E}_{q(x_K)}[\mathrm{u}_x(x_K)]^{\top} \end{pmatrix}^{\top} \cdot \begin{pmatrix} \mathbb{E}_{q(z)}[z_1] \\ \vdots \\ \mathbb{E}_{q(z)}[z_K] \end{pmatrix} \\ m_{Y \to Z} &= \begin{pmatrix} \mathbb{E}_{q(x_1)}[\mathrm{u}_x(x_1)]^{\top} \\ \vdots \\ \mathbb{E}_{q(x_K)}[\mathrm{u}_x(x_K)]^{\top} \end{pmatrix} \cdot \left(\sum_{Ch} m_{Ch \to Y} \right) \\ m_{Y \to X_k} &= \left(\sum_{Ch} m_{Ch \to Y} \right) \times \mathbb{E}_{q(z)}[z_k]. \end{aligned} \tag{A.35}$$

A.8. Nœud déterministe d'addition

Soient X_1 et X_2 deux variables aléatoires associés à des distributions dont le vecteur de statistiques naturelles est de la forme $\mathrm{u}_x(x) = (x, x^2)$. La variable Y, définie comme la somme de ces variables, $Y = X_1 + X_2$, suit alors une distribution dont le vecteur de statistiques naturelles est aussi de la forme $\mathrm{u}_y(y) = (y, y^2)$. Les messages de l'algorithme VMP mis en jeu pour le nœud associé à Y s'écrivent comme suit :

$$\begin{aligned} m_{Y \to Ch} &= \begin{pmatrix} \mathbb{E}_{q(y)}[y] \\ \mathbb{E}_{q(y)}[y^2] \end{pmatrix} = \begin{pmatrix} \mathbb{E}_{q(x_1)}[x_1] + \mathbb{E}_{q(x_2)}[x_2] \\ \mathbb{E}_{q(x_1)}[x_1^2] + \mathbb{E}_{q(x_2)}[x_2^2] + 2\,\mathbb{E}_{q(x_1)}[x_1]\,\mathbb{E}_{q(x_2)}[x_2] \end{pmatrix} \\ m_{Y \to X_i} &= \begin{pmatrix} 1 & 2\,\mathbb{E}_{q(x_j)}[x_j] \\ 0 & 1 \end{pmatrix} \cdot \left(\sum_{Ch} m_{Ch \to Y} \right) \quad \forall\, i, j \in \{1, 2\},\, i \neq j, \end{aligned} \tag{A.36}$$

où Ch est un nœud enfant de Y.

Annexe B.

Mise en œuvre pratique de modèles bayésiens avec l'algorithme VMP

Cette annexe rassemble les détails concernant le choix des valeurs pour les hyperparamètres des modèles bayésiens présentés dans ce manuscrit, ainsi que la mise en œuvre pratique de l'algorithme VMP pour ces modèles.

B.1. Convergence des modèles

Pour tous les modèles bayésiens, la convergence du calcul de la probabilité a posteriori du modèle par l'algorithme VMP est vérifiée en examinant l'évolution de la différence $\Delta\mathcal{L}$ de la valeur de la borne inférieure $\mathcal{L}$ entre deux itérations de l'algorithme VMP. La convergence est considérée comme atteinte si $\Delta\mathcal{L} < 10^{-1}$. Pour borner le temps de calcul, le nombre d'itérations maximum est fixé à 10 000.

Dans le cas où la probabilité a posteriori d'un modèle appliqué à un jeu de données est calculée à partir de plusieurs initialisations aléatoires, le nombre d'itérations maximum pour chaque instance est limité à 1 000 itérations. Le nombre d'initialisations aléatoires est fixé à 50. Pour l'instance du modèle ayant atteint la plus haute valeur de borne inférieure $\mathcal{L}$, les calculs sont étendus sur 10 000 itérations supplémentaires si la convergence de $\mathcal{L}$ n'a pas déjà été atteinte.

B.2. A priori faiblement informatifs et pré-traitement des données

Les paramètres des distributions a priori sont fixés de telle sorte que celles-ci aient un vaste support effectif et apporte une faible quantité d'information a priori. On parle alors d'a priori vague. Les valeurs sont inspirées de ce qu'utilise Bishop pour ses modèles bayésiens [Bishop, 2006].

Cependant, il faut préciser que les distributions a priori ne sont vagues que relativement à la gamme des valeurs observées. Si la gamme des valeurs observées est trop grande, certaines valeurs risquent d'être en dehors de l'ensemble des valeurs supportées par la probabilité a priori vague. Ces valeurs étant plus faiblement probables au regard de l'a priori, le calcul de la probabilité a posteriori est biaisé et la probabilité a priori perd sa qualité de faible apport d'information.

Pour éviter ce phénomène, les valeurs observées associées à un modèle bayésien sont normalisées avant d'être utilisées. Concrètement, comme les données manipulées sont le module de coefficients d'ondelette, ou leur carré, l'ensemble des données est remis à l'échelle en divisant par la moyenne des données.

B.3. Modèle de mélange gaussien

Le modèle bayésien de mélange gaussien a été présenté à la section 3.3.2. Les paramètres des distributions a priori sont fixés à :

- $m_\mu = 0$ et $l_\mu = 10^{-3}$ pour la distribution gaussienne des variables μ_k,
- $a_\lambda = 10^{-3}$ et $b_\lambda = 10^{-3}$ pour la distribution gamma des variables λ_k,
- $\alpha_k = 10^{-2}\ \forall\, k$ pour la distribution de Dirichlet de π.

La probabilité a posteriori de ce modèle est calculée à l'aide de plusieurs instances initialisées aléatoirement, suivant les modalités détaillées en section B.1. L'aléa porte sur les valeurs initiales des paramètres $\pi^*_{n,k}$ des distributions a posteriori des variables cachées Z_n du modèle.

B.4. Modèle ricien

Le modèle ricien et les équations de l'algorithme VMP liées à ce modèle ont été développés à la section 3.4.2. Les paramètres des distributions a priori sont fixés à :

- $m_\nu = 0$ et $l_\nu = 10^{-3}$ pour la distribution gaussienne rectifiée de la variable ν,
- $a_\lambda = 10^{-3}$ et $b_\lambda = 10^{-3}$ pour la distribution gamma de la variable λ.

La probabilité a posteriori de ce modèle est calculée avec une seule instance, sans initialisation aléatoire, suivant les modalités détaillées en section B.1.

B.5. Modèle de mélange ricien

Le modèle bayésien de mélange ricien a été introduit à la section 3.4.4. Les paramètres des distributions a priori sont fixés à :

- $m_\nu = 0$ et $l_\nu = 10^{-3}$ pour la distribution gaussienne rectifiée des variables ν_k,
- $a_\lambda = 10^{-3}$ et $b_\lambda = 10^{-3}$ pour la distribution gamma des variables λ_k,
- $\alpha_k = 10^{-2}\ \forall\, k$ pour la distribution de Dirichlet de π.

La probabilité a posteriori de ce modèle peut être calculée à l'aide de plusieurs instances initialisées aléatoirement, suivant les modalités détaillées en section B.1 : les valeurs initiales des paramètres $\pi^*_{n,k}$ des distributions a posteriori des variables cachées Z_n du modèle sont tirées aléatoirement.

Dans le cas d'un modèle à deux composantes riciennes, une heuristique consiste à calculer la probabilité a posteriori du modèle à partir d'une seule initialisation, où la valeur initiale du paramètre $\pi^*_{n,1}$ correspond à 1 si l'observation x_n est inférieure à la médiane et à 0 sinon. La valeur du paramètre $\pi^*_{n,2}$ vaut alors $1 - \pi^*_{n,1}$.

B.6. Modèle de mélange ricien hiérarchique

Le modèle de mélange ricien hiérarchique a été formalisé à la section 4.2.3, page 83. Les paramètres des distributions a priori sont fixés à :

- $m_\nu = 0$ et $l_\nu = 10^{-3}$ pour la distribution gaussienne rectifiée des variables $\nu_{1,r}$ et des variables $\nu_{2-1,r}$,
- $a_\lambda = 10^{-3}$ et $b_\lambda = 10^{-3}$ pour la distribution gamma des variables $\lambda_{1,r}$ et des variables $\lambda_{2,r}$,
- $\alpha_1 = \alpha_2 = 10^{-2}$ pour la distribution de Dirichlet des variables π et π_Y ainsi que des variables $\pi_{t'}$.

Ce modèle étant complexe, son initialisation est faite à partir des résultats obtenus par des modèles de mélange ricien à deux composantes. Ainsi, la probabilité a posteriori de ce modèle est calculée à partir d'une seule initialisation.

Annexe C.

Résultats des expériences sur le jeu de données artificielles univariées

Cette annexe rassemble les résultats des expériences décrites à la section 5.2.2. Ces résultats sont présentés sous la forme de diagrammes haricot, complétés par un intervalle de confiance à 99 % de la moyenne dans le cas où les résultats sont des différences appariées. Cette représentation est expliquée à la page 97. Les abréviations utilisées sont décrites avec les méthodes d'analyses testées, à la section 5.1.

La figure C.1 contient l'intégralité des résultats. Les autres figures illustrent des différences appariées entre les résultats des méthodes. Les méthodes ainsi comparées ne diffèrent que par une caractéristique. Par exemple, dans la figure C.6, l'utilisation et l'absence d'utilisation de la procédure de Benjamini et Yekutieli sont comparées pour chaque méthode en faisant éventuellement usage.

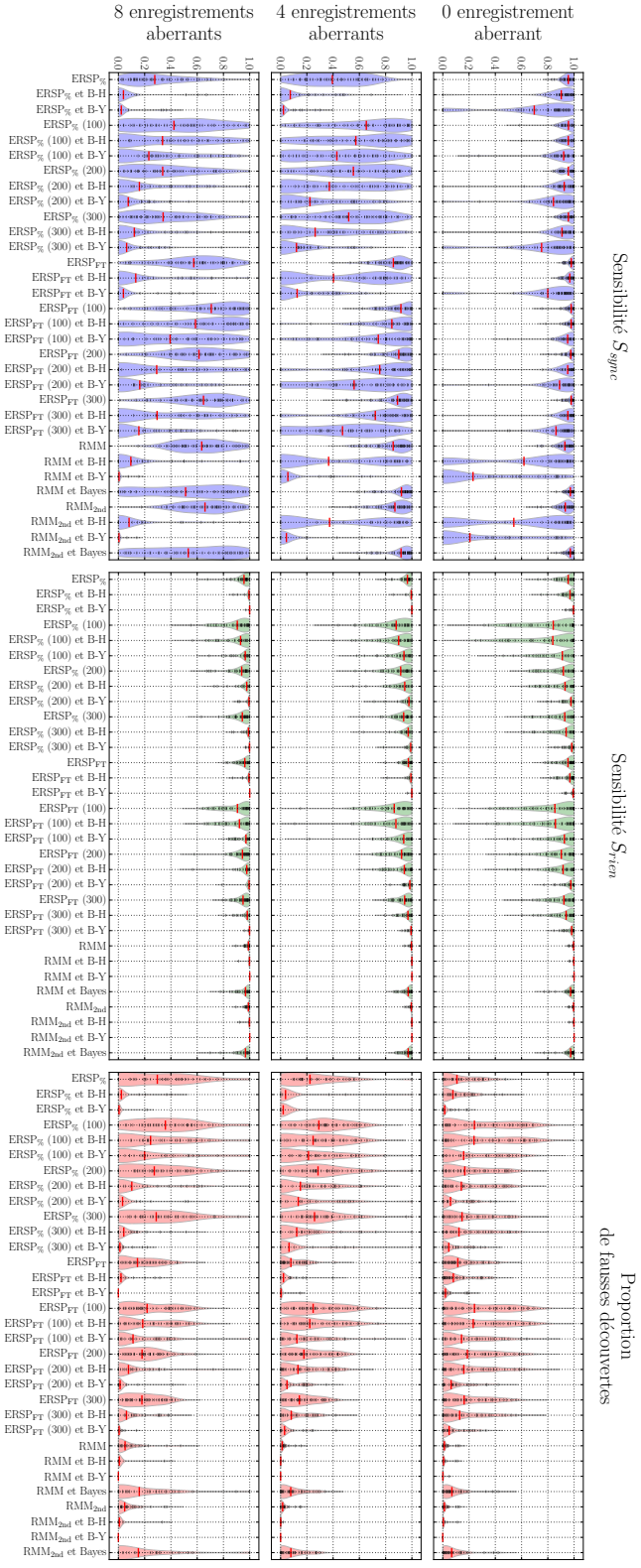

FIGURE C.1. – Comparaison des performances pour la détection de synchronisations univariées des méthodes conventionnelles ERSP et de mes méthodes d'analyse de synhcronisations univariées, sur le jeu de données artificielles univariées

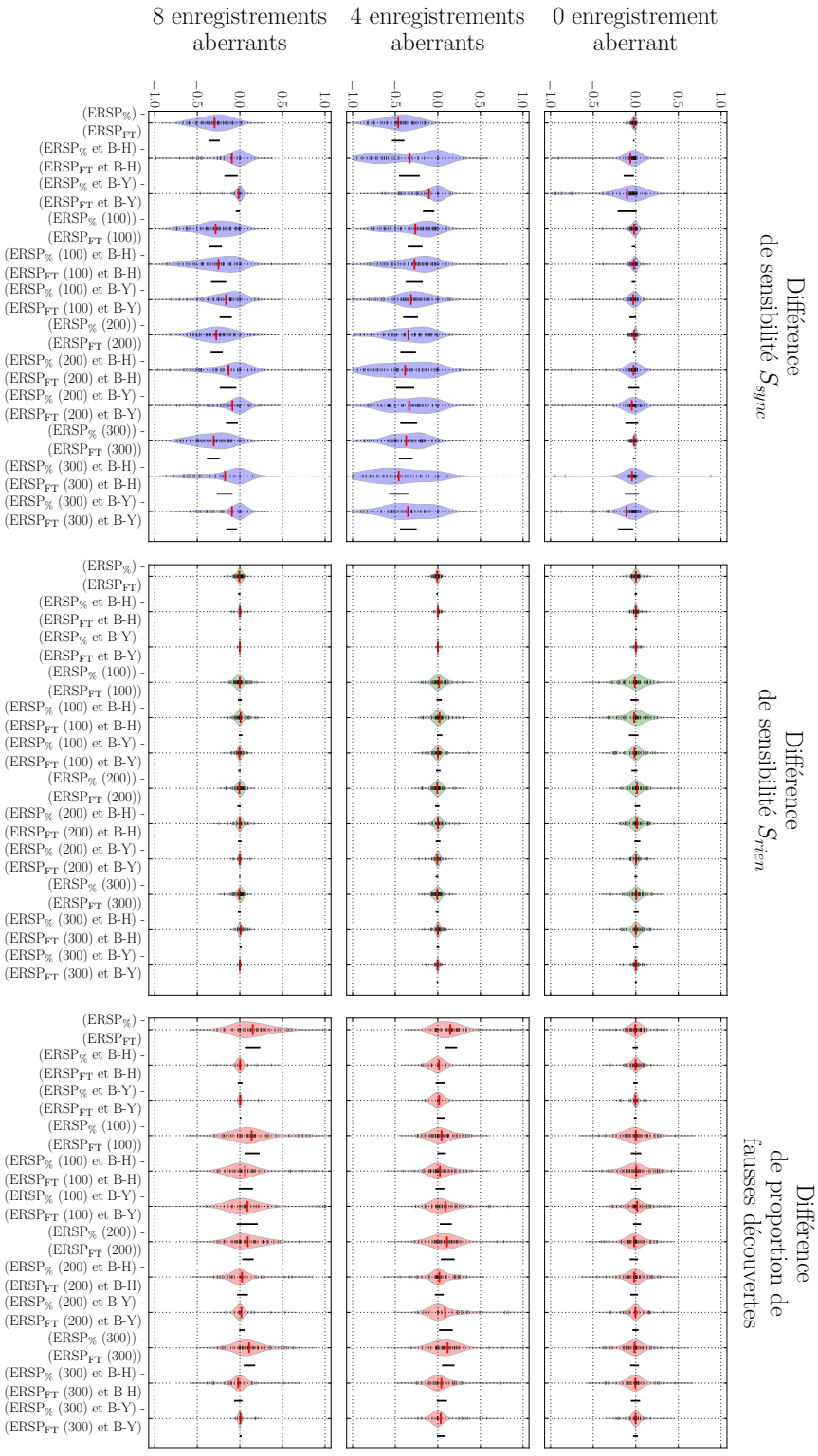

FIGURE C.2. – Comparaison deux à deux des résultats de la méthode ERSP à modèle multiplicatif simple ($\text{ERSP}_{\%}$) et de la méthode ERSP avec normalisation simple essai (ERSP_{FT}).

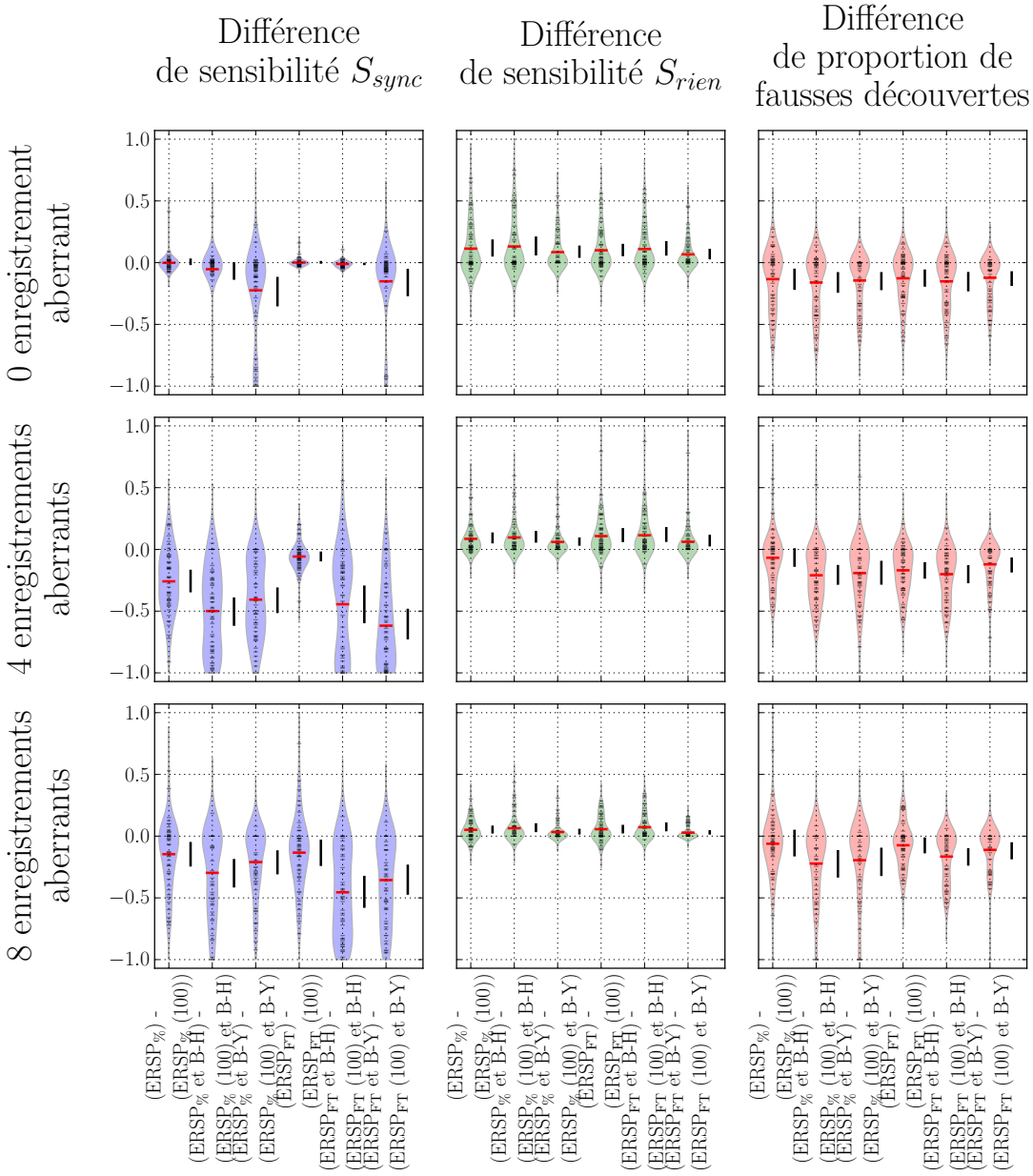

FIGURE C.3. – Comparaison deux à deux des résultats des méthodes ERSP sans et avec restriction de la période pré-stimulus à 100 points.

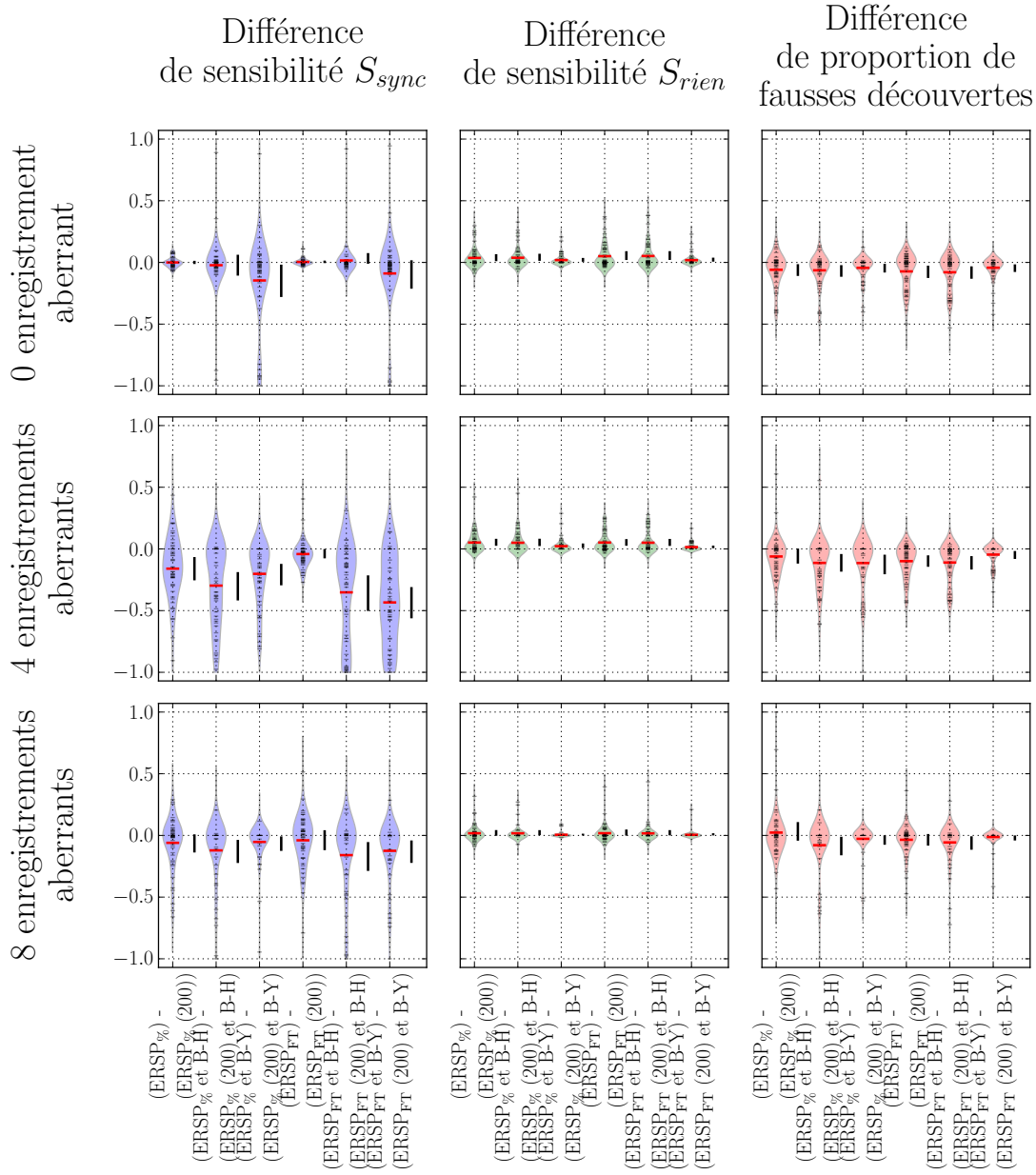

FIGURE C.4. – Comparaison deux à deux des résultats des méthodes ERSP sans et avec restriction de la période pré-stimulus à 200 points.

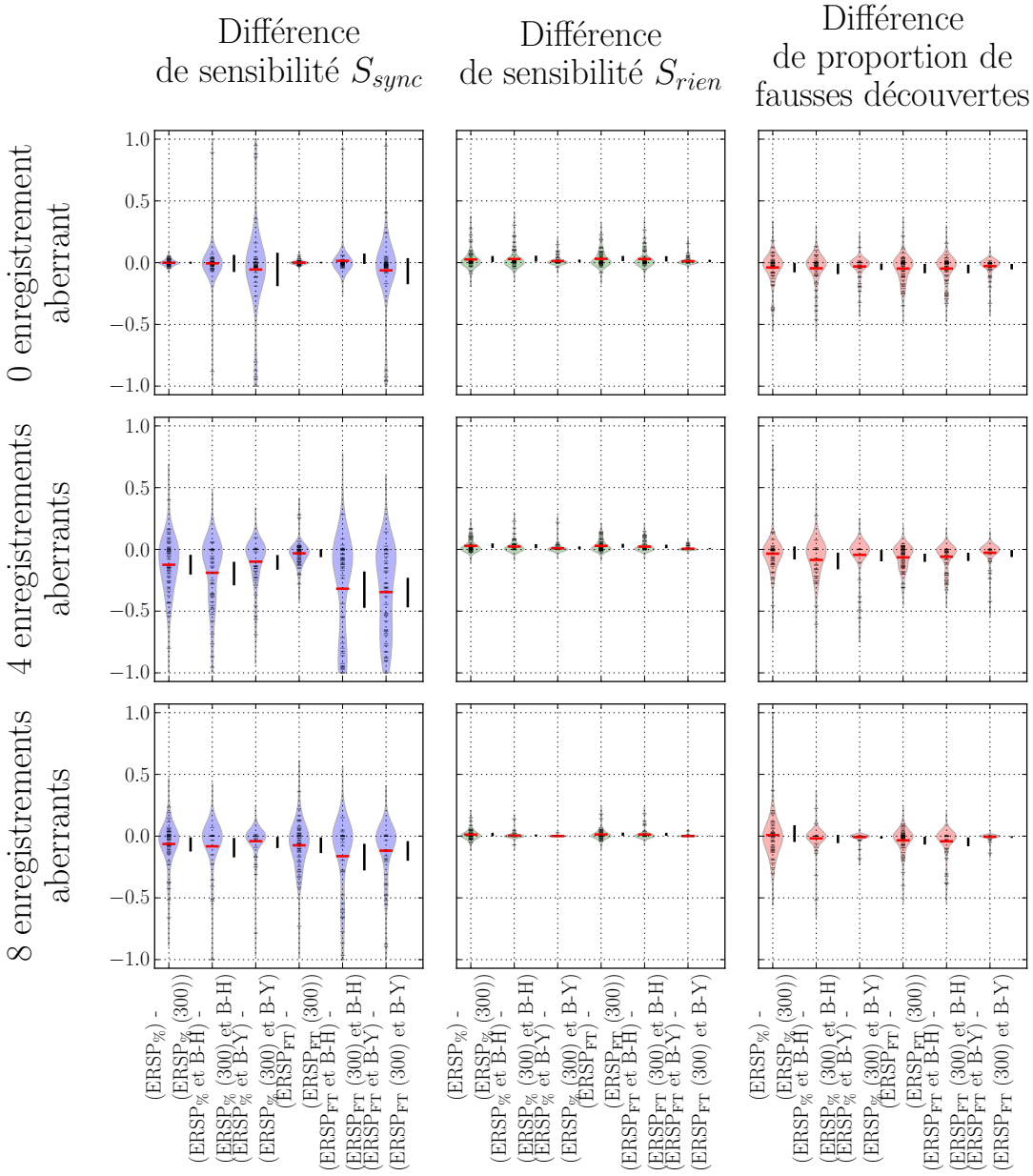

FIGURE C.5. – Comparaison deux à deux des résultats des méthodes ERSP sans et avec restriction de la période pré-stimulus à 300 points.

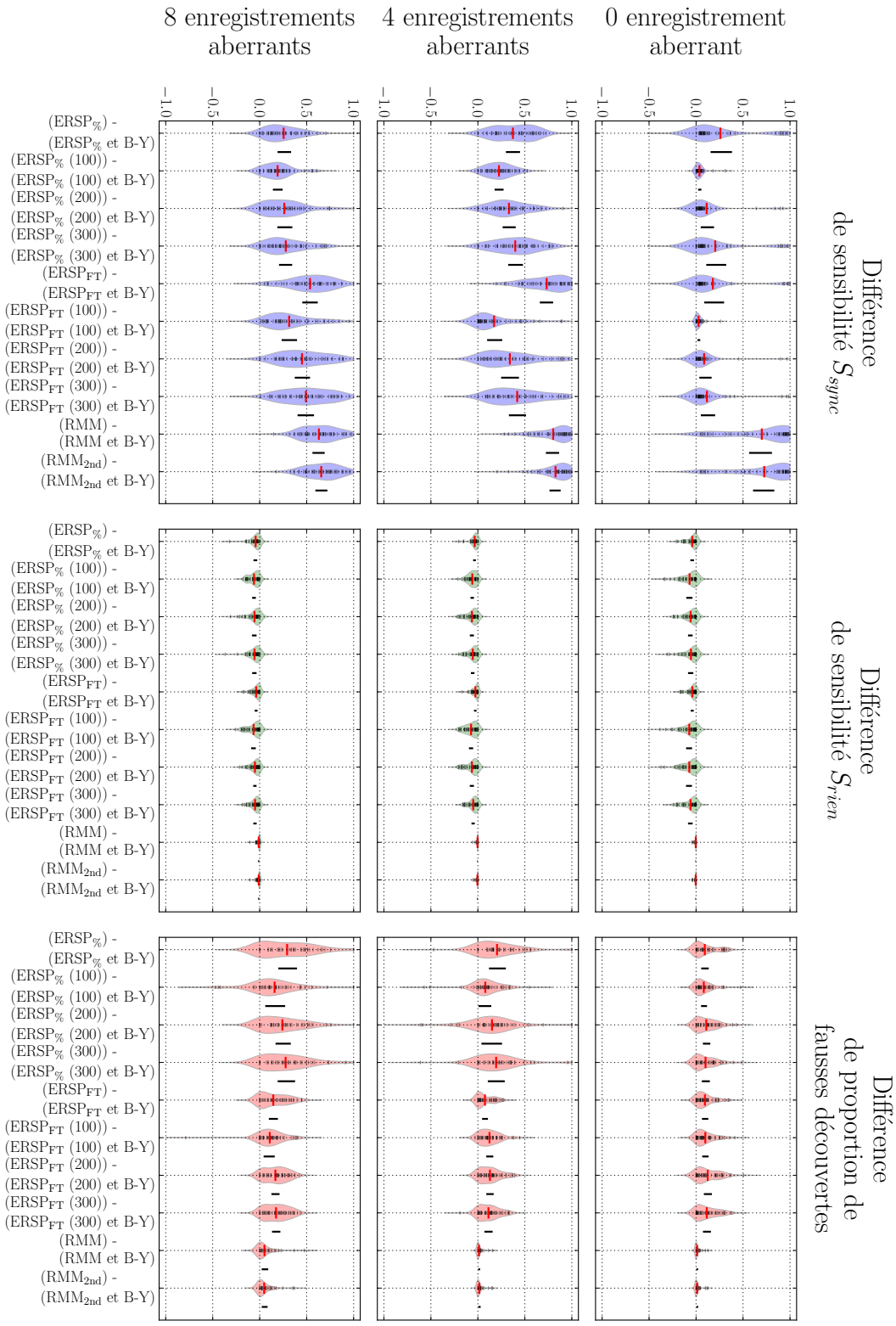

FIGURE C.6. – Comparaison deux à deux des résultats des méthodes d'analyse sans et avec correction des tests d'hypothèses multiples par la procédure de Benjamini et Yekutieli (B-Y).

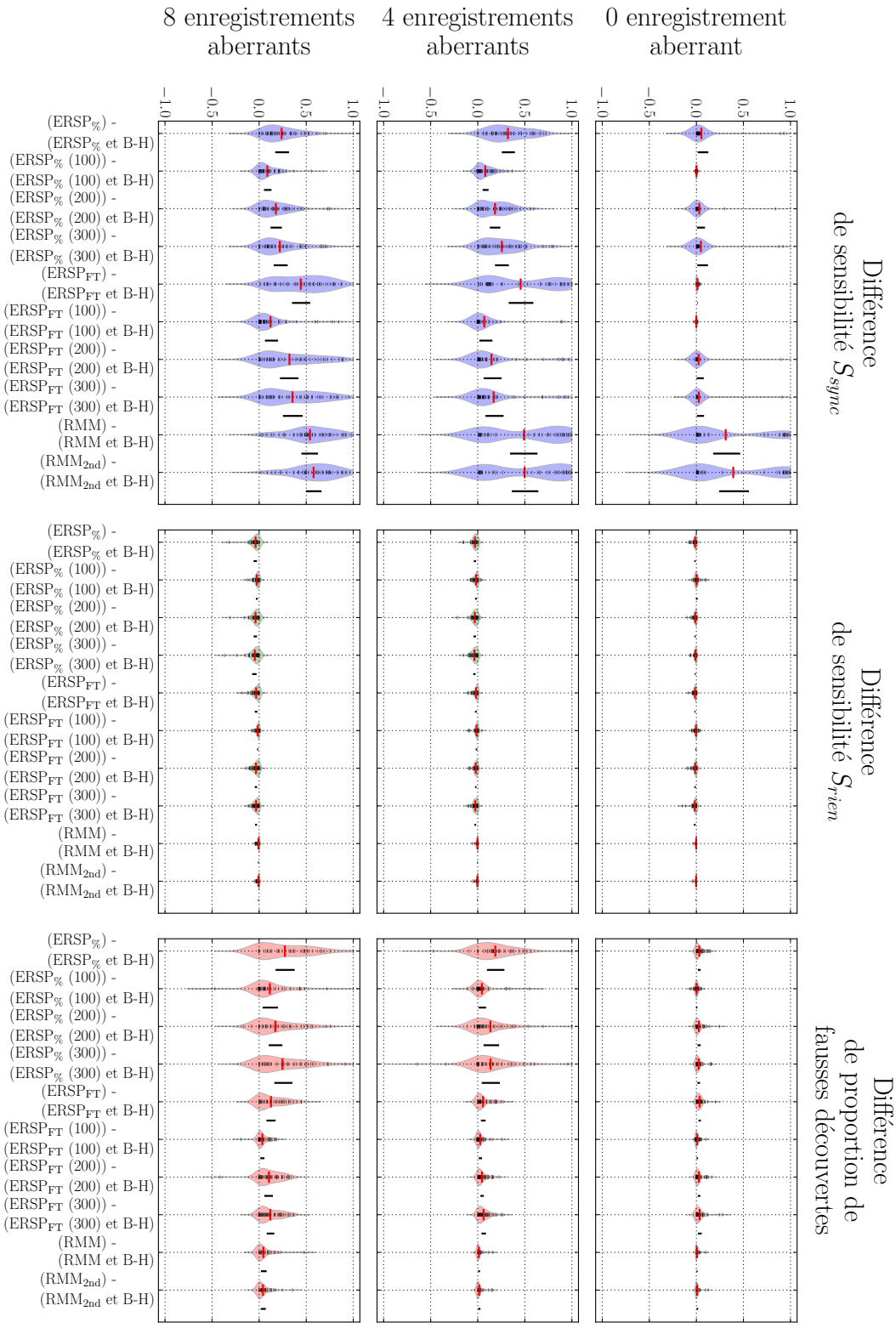

FIGURE C.7. – Comparaison deux à deux des résultats des méthodes d'analyse sans et avec correction des tests d'hypothèses multiples par la procédure de Benjamini et Hochberg (B-H).

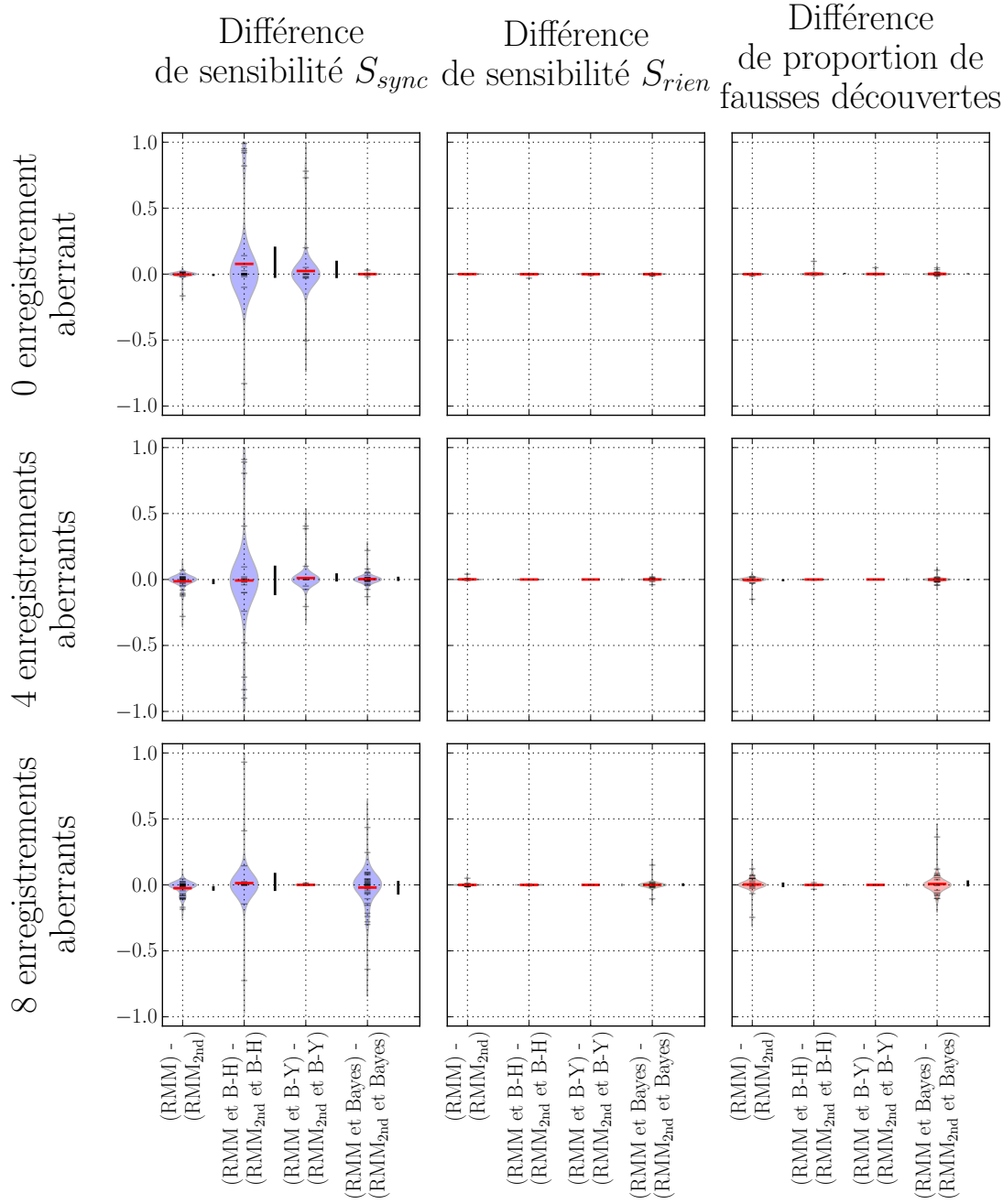

FIGURE C.8. – Comparaison deux à deux des résultats de mes méthodes d'analyse, sans et avec une seconde passe d'optimisation pour améliorer la convergence des modèles de mélange.

Bibliographie

[Abramowitz and Stegun, 1964] Abramowitz, M. and Stegun, I. (1964). Handbook of mathematical functions with formulas, graphs, and mathematical tables, vol. 55, of National Bureau of Standards Applied Mathematics Series. U.S. Government Printing Office, Washington, D.C.

[Akaike, 1973] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In 2nd International Symposium on Information Theory, (Petrov, B. N. and Csaki, F., eds), pp. 267–281,.

[Attias, 2000] Attias, H. (2000). A variational Bayesian framework for graphical models. In Advances in neural information processing systems vol. 12, pp. 209–215,.

[Bayes, 1763] Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. Philosophical Transactions of the Royal Society of London *53*, 370–418.

[Beal, 2003] Beal, M. (2003). Variational algorithm for approximate Bayesian inference. PhD thesis, University College London.

[Beal and Ghahramani, 2006] Beal, M. and Ghahramani, Z. (2006). Variational Bayesian Learning of Directed Graphical Models with Hidden Variables. Bayesian Analysis *1*, 793–832.

[Beaulieu and Hemachandra, 2010] Beaulieu, N. and Hemachandra, K. (2010). Novel Simple Forms for Multivariate Rayleigh and Rician Distributions with Generalized Correlation. In Global Telecommunications Conference (GLOBECOM 2010), 2010 IEEE pp. 1–6,.

[Benjamini and Hochberg, 1995] Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate : a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B (Methodological) *57*, 289–300.

[Benjamini and Yekutieli, 2001] Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. Annals of Statistics *29*, 1165–1188.

[Bennett et al., 2009] Bennett, C., Baird, A., Miller, M. and Wolford, G. (2009). Neural correlates of interspecies perspective taking in the post-mortem Atlantic Salmon : an argument for multiple comparisons correction (poster). Presented at Organization for Human Brain Mapping 2009 Annual Meeting.

[Berger, 1929] Berger, H. (1929). Über das Elektrenkephalogramm des Menschen (De l'électroencéphalogramme humain). Archiv für Psychiatrie und Nervenkrankheiten *87*, 527–570.

[Berger and Berry, 1988] Berger, J. and Berry, D. (1988). Statistical analysis and the illusion of objectivity. American Scientist *76*, 159–165.

[Bilmes, 2004] Bilmes, J. (2004). What HMMs Can't Do : A Graphical Model Perspective. In Beyond HMM : Workshop on Statiatical Modeling Approach for Speech Recognition, Kyoto, Japan. ATR Invited Paper and Lecture.

[Birnbaum, 1962] Birnbaum, A. (1962). On the Foundations of Statistical Inference. Journal of the American Statistical Association *57*, 269–306.

[Bishop, 1999] Bishop, C. (1999). Variational principal components. In Artificial Neural Networks, 1999. ICANN 99. Ninth International Conference on (Conf. Publ. No. 470) vol. 1, pp. 509–514, IET.

[Bishop, 2006] Bishop, C. (2006). Pattern Recognition and Machine Learning (Information Science and Statistics). Springer.

[Bishop and Svensen, 2003] Bishop, C. and Svensen, M. (2003). Bayesian hierarchical mixtures of experts. In Proceedings of the 19th Annual Conference on Uncertainty in Artificial Intelligence (UAI-03) pp. 57–64,.

[Corduneanu and Bishop, 2001] Corduneanu, A. and Bishop, C. (2001). Variational Bayesian model selection for mixture distributions. In Artificial Intelligence and Statistics pp. 27–34,.

[Dalal et al., 2011] Dalal, S., Vidal, J., Hamamé, C., Ossandón, T., Bertrand, O., Lachaux, J. and Jerbi, K. (2011). Spanning the rich spectrum of the human brain : slow waves to gamma and beyond. Brain Structure and Function *216*, 77–84.

[Davison and Hinkley, 1997] Davison, A. and Hinkley, D. (1997). Bootstrap Methods and Their Applications. Cambridge University Press.

[Delorme and Makeig, 2004] Delorme, A. and Makeig, S. (2004). EEGLAB : an open source toolbox for analysis of single-trial {EEG} dynamics including independent component analysis. Journal of Neuroscience Methods *134*, 9–21.

[Denker et al., 2011] Denker, M., Roux, S., Lindén, H., Diesmann, M., Riehle, A. and Grün, S. (2011). The Local Field Potential Reflects Surplus Spike Synchrony. Cerebral Cortex *21*, 2681–2695.

[Efron and Tibshirani, 1993] Efron, B. and Tibshirani, R. (1993). An introduction to the bootstrap. Monographs on statistics and applied probabilities, Chapman & Hall/CRC.

[Farcomeni, 2008] Farcomeni, A. (2008). A review of modern multiple hypothesis testing, with particular attention to the false discovery proportion. Statistical Methods in Medical Research *17*, 347–388.

[Fawcett, 2006] Fawcett, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters *27*, 861–874.

[Gelman et al., 2004] Gelman, A., Carlin, J., Stern, H. and Rubin, D. (2004). Bayesian data analysis. Chapman & Hall/CRC.

[Ghahramani and Beal, 2000] Ghahramani, Z. and Beal, M. (2000). Variational inference for Bayesian mixtures of factor analysers. Advances in neural information processing systems *12*, 449–455.

[Gill, 2008] Gill, J. (2008). Is Partial-Dimension Convergence a Problem for Inferences from MCMC Algorithms ? Political Analysis *16*, 153–178.

[Goupillaud et al., 1984] Goupillaud, P., Grossmann, A. and Morlet, J. (1984). Cycle-octave and related transforms in seismic signal analysis. Geoexploration *23*, 85–102.

[Grandchamp and Delorme, 2011] Grandchamp, R. and Delorme, A. (2011). Single-trial normalization for event-related spectral decomposition reduces sensitivity to noisy trials. Frontiers in Psychology *2*.

[Grimmer, 2011] Grimmer, J. (2011). An introduction to Bayesian inference via variational approximations. Political Analysis *19*, 32–47.

[Gudbjartsson and Patz, 1995] Gudbjartsson, H. and Patz, S. (1995). The rician distribution of noisy mri data. Magnetic Resonance in Medicine *34*, 910–914.

[Guttorp and Lockhart, 1988] Guttorp, P. and Lockhart, R. (1988). Finding the Location of a Signal : A Bayesian Analysis. Journal of the American Statistical Association *83*, 322–330.

[Handl et al., 2005] Handl, J., Knowles, J. and Kell, D. (2005). Computational cluster validation in post-genomic data analysis. Bioinformatics *21*, 3201–3212.

[Herrmann et al., 2005] Herrmann, C., Grigutsch, M. and Busch, N. (2005). EEG oscillations and wavelet analysis. In Event-related potentials : A methods handbook. MIT Press.

[Huang et al., 2005] Huang, H.-S., Yang, B.-H. and Hsu, C.-N. (2005). Triple jump acceleration for the EM algorithm. In Data Mining, Fifth IEEE International Conference on.

[Hutt and Munk, 2009] Hutt, A. and Munk, M. (2009). Detection of Phase Synchronization in Multivariate Single Brain Signals by a Clustering Approach. In Coordinated Activity in the Brain, (Velazquez, J. L. P. and Wennberg, R., eds), vol. 2, of Springer Series in Computational Neuroscience pp. 149–164. Springer New York.

[Jaakkola and Jordan, 1998] Jaakkola, T. and Jordan, M. (1998). Improving the Mean Field Approximation Via the Use of Mixture Distributions. In Learning in Graphical Models, (Jordan, M., ed.), vol. 89, of NATO ASI Series pp. 163–173. Springer Netherlands.

[Jefferys and Berger, 1991] Jefferys, W. and Berger, J. (1991). Sharpening Ockham's razor on a bayesian strop. Technical Report 91-44C Purdue University.

[Johnson, 1999] Johnson, D. H. (1999). The insignificance of statistical significance testing. The journal of wildlife management *63*, 763–772.

[Johnson, 2005] Johnson, M. (2005). Capacity and complexity of HMM duration modeling techniques. Signal Processing Letters, IEEE *12*, 407–410.

[Jordan et al., 1997] Jordan, D., Miksad, R. and Powers, E. (1997). Implementation of the continuous wavelet transform for digital time series analysis. Review of scientific instruments *68*, 1484–1494.

[Juergens et al., 1999] Juergens, E., Guettler, A. and Eckhorn, R. (1999). Visual stimulation elicits locked and induced gamma oscillations in monkey intracortical- and EEG-potentials, but not in human EEG. Experimental Brain Research *129*, 247–259.

[Kaiser, 1994] Kaiser, G. (1994). A friendly guide to wavelets. Birkhäuser.

[Kampstra, 2008] Kampstra, P. (2008). Beanplot : A Boxplot Alternative for Visual Comparison of Distributions. Journal of Statistical Software, Code Snippets *28*, 1–9.

[Katahira et al., 2008] Katahira, K., Watanabe, K. and Okada, M. (2008). Deterministic annealing variant of variational Bayes method. Journal of Physics : Conference Series *95*, 012015.

[Kay, 1993] Kay, S. (1993). Fundamental of Statistical Signal Processing, Volume II : Detection Theory. Prentice-Hall, Inc.

[Kerr and Churchill, 2001] Kerr, M. and Churchill, G. (2001). Bootstrapping cluster analysis : Assessing the reliability of conclusions from microarray experiments. Proceedings of the National Academy of Sciences *98*, 8961–8965.

[Lachaux et al., 1999] Lachaux, J., Rodriguez, E., Martinerie, J. and Varela, F. (1999). Measuring phase synchrony in brain signals. Human Brain Mapping *8*, 194–208.

[Lachaux et al., 2005] Lachaux, J.-P., George, N., Tallon-Baudry, C., Martinerie, J., Hugueville, L., Minotti, L., Kahane, P. and Renault, B. (2005). The many faces of the gamma band response to complex visual stimuli. NeuroImage *25*, 491–501.

[Le Van Quyen and Bragin, 2007a] Le Van Quyen, M. and Bragin, A. (2007a). Analysis of dynamic brain oscillations : methodological advances. Trends in neurosciences *30*, 365–373.

[Le Van Quyen and Bragin, 2007b] Le Van Quyen, M. and Bragin, A. (2007b). Analysis of dynamic brain oscillations : methodological advances. Trends in Neurosciences *30*, 365–373.

[Little, 2006] Little, R. (2006). Calibrated Bayes. The American Statistician *60*, 213–223.

[MacKay, 1997] MacKay, D. (1997). Ensemble learning for hidden Markov models. http://www.inference.phy.cam.ac.uk/mackay/abstracts/ensemblePaper.html.

[MacKay, 1998] MacKay, D. (1998). Introduction to Monte Carlo Methods. In Learning in Graphical Models, (Jordan, M. I., ed.), NATO Science Series pp. 175–204. Kluwer Academic Press.

[Makeig et al., 2004] Makeig, S., Debener, S., Onton, J. and Delorme, A. (2004). Mining event-related brain dynamics. Trends in cognitive sciences *8*, 204–210.

[Mallat, 2008] Mallat, S. (2008). A wavelet tour of signal processing. Academic Press.

[March et al., 2011] March, M., Starkman, G., Trotta, R. and Vaudrevange, P. (2011). Should we doubt the cosmological constant ? Monthly Notices of the Royal Astronomical Society *410*, 2488–2496.

[Minka, 2005] Minka, T. (2005). Divergence measures and message passing. Technical Report MSR-TR-2005-173 Microsoft Research Cambridge.

[Miskin, 2000] Miskin, J. (2000). Ensemble Learning for Independent Component Analysis. PhD thesis, University of Cambridge.

[Neal, 1993] Neal, R. (1993). Probabilistic Inference Using Markov Chain Monte Carlo Methods. Technical Report CRG-TR-93-1 University of Toronto.

[Neuman, 1992] Neuman, E. (1992). Inequalities involving modified Bessel functions of the first kind. Journal of mathematical analysis and applications *171*, 532–536.

[Nunez and Srinavasan, 2006] Nunez, P. and Srinavasan, R. (2006). Electric fields of the brain : the neurophysics of EEG. Second edition, Oxford University Press, Incorporated.

[Onton et al., 2005] Onton, J., Delorme, A. and Makeig, S. (2005). Frontal midline EEG dynamics during working memory. NeuroImage *27*, 341–356.

[Opper and Archambeau, 2009] Opper, M. and Archambeau, C. (2009). The variational Gaussian approximation revisited. Neural computation *21*, 786–792.

[Penny et al., 2002] Penny, W., Kiebel, S., Kilner, J. and Rugg, M. (2002). Event-related brain dynamics. Trends in Neurosciences *25*, 387–389.

[Pfurtscheller and Lopes da Silva, 1999] Pfurtscheller, G. and Lopes da Silva, F. (1999). Event-related EEG/MEG synchronization and desynchronization : basic principles. Clinical neurophysiology *110*, 1842–1857.

[Picton et al., 2000] Picton, T., Bentin, S., Berg, P., Donchin, E., Hillyard, S., Johnson, R., Miller, G., Ritter, W., Ruchkin, D., Rugg, M. and Taylor, M. (2000). Guidelines for using human event-related potentials to study cognition : Recording standards and publication criteria. Psychophysiology *37*, 127–152.

[Qi and Jaakkola, 2006] Qi, Y. and Jaakkola, T. (2006). Parameter Expanded Variational Bayesian Methods. In NIPS pp. 1097–1104,.

[Quiroga and Garcia, 2003] Quiroga, R. and Garcia, H. (2003). Single-trial event-related potentials with wavelet denoising. Clinical Neurophysiology *114*, 376–390.

[Rice, 1945] Rice, O. (1945). Mathematical Analysis of Random Noise. Bell System Technical Journal *24*, 46–156.

[Rio et al., 2010] Rio, M., Hutt, A. and Girau, B. (2010). Partial amplitude synchronization detection in brain signals using Bayesian Gaussian mixture models. In Proceedings of the fifth French conference on Computational Neuroscience, Lyon, (Neurocomp, ed.), pp. 109–113, Neurocomp, Lyon, France.

[Rio et al., 2011] Rio, M., Hutt, A., Munk, M. and Girau, B. (2011). Partial amplitude synchronization detection in brain signals using Bayesian Gaussian mixture models. Journal of Physiology - Paris *105*, 98–105.

[Robert, 2006] Robert, C. (2006). Le choix bayésien : Principes et pratique. Springer.

[Rodriguez et al., 1999] Rodriguez, E., George, N., Lachaux, J., Martinerie, J., Renault, B. and Varela, F. (1999). Perception's shadow : long-distance synchronization of human brain activity. Nature *397*, 430–433.

[Russell and Cook, 1987] Russell, M. and Cook, A. (1987). Experimental evaluation of duration modelling techniques for automatic speech recognition. In Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '87. vol. 12, pp. 2376–2379,.

[Sadaghiani et al., 2010] Sadaghiani, S., Hesselmann, G., Friston, K. and Kleinschmidt, A. (2010). The relation of ongoing brain activity, evoked neural responses, and cognition. Frontiers in Systems Neuroscience *4*.

[Salakhutdinov and Roweis, 2003] Salakhutdinov, R. and Roweis, S. (2003). Adaptive Overrelaxed Bound Optimization Methods. In Proceedings of the International Conference on Machine Learning vol. 20, pp. 664–671,.

[Sanes and Donoghue, 1993] Sanes, J. and Donoghue, J. (1993). Oscillations in local field potentials of the primate motor cortex during voluntary movement. Proceedings of the National Academy of Sciences *90*, 4470–4474.

[Saporta, 2011] Saporta, G. (2011). Probabilités, analyse des données et statistique. Editions Technip.

[Schmitzer-Torbert et al., 2005] Schmitzer-Torbert, N., Jackson, J., Henze, D., Harris, K. and Redish, A. (2005). Quantitative measures of cluster quality for use in extracellular recordings. Neuroscience *131*, 1–11.

[Schwarz, 1978] Schwarz, G. (1978). Estimating the dimension of a model. The annals of statistics *6*, 461–464.

[Scott and Berger, 2006] Scott, J. and Berger, J. (2006). An exploration of aspects of Bayesian multiple testing. Journal of Statistical Planning and Inference *136*, 2144–2162.

[Simonovski and Boltežar, 2003] Simonovski, I. and Boltežar, M. (2003). The norms and variances of the Gabor, Morlet and general harmonic wavelet functions. Journal of Sound and Vibration *264*, 545–557.

[Soltani Bozchalooi and Liang, 2007] Soltani Bozchalooi, I. and Liang, M. (2007). A smoothness index-guided approach to wavelet parameter selection in signal de-noising and fault detection. Journal of Sound and Vibration *308*, 246–267.

[Starkman et al., 2008] Starkman, G., Trotta, R. and Vaudrevange, P. (2008). Introducing doubt in bayesian model comparison. arXiv preprint arXiv :0811.2415.

[Tallon-Baudry and Bertrand, 1999] Tallon-Baudry, C. and Bertrand, O. (1999). Oscillatory gamma activity in humans and its role in object representation. Trends in Cognitive Sciences *3*, 151–162.

[Tipping and Bishop, 1998] Tipping, M. and Bishop, C. (1998). Mixtures of probabilistic principal component analyzers. Technical Report NCRG/97/003 Neural Computing Research Group Aston University.

[Turi et al., 2012] Turi, G., Gotthardt, S., Singer, W., Vuong, T., Munk, M. and Wibral, M. (2012). Quantifying additive evoked contributions to the event-related potential. NeuroImage *59*, 2607–2624.

[Ueda and Ghahramani, 2002] Ueda, N. and Ghahramani, Z. (2002). Bayesian model search for mixture models based on optimizing variational bounds. Neural Networks *15*, 1223–1241.

[Verbeek, 2004] Verbeek, J. (2004). Mixture Models for Clustering and Dimension Reduction. PhD thesis, University of Amsterdam.

[Verhoeven et al., 2005] Verhoeven, K., Simonsen, K. and McIntyre, L. (2005). Implementing false discovery rate control : increasing your power. Oikos *108*, 643–647.

[Vetterli and Kovačevic, 1995] Vetterli, M. and Kovačevic, J. (1995). Wavelets and subband coding. Prentice-Hall, Inc.

[Vialatte, 2005] Vialatte, F.-B. (2005). Modélisation en bosses pour l'analyse des motifs oscillatoires reproductibles dans l'activité de populations neuronales : applications à l'apprentissage olfactif chez l'animal et à la détection précoce de la maladie d'Alzheimer. PhD thesis, Université Pierre et Marie Curie - Paris 6.

[Wainwright and Jordan, 2008] Wainwright, M. and Jordan, M. (2008). Graphical models, exponential families, and variational inference. Foundations and Trends® in Machine Learning *1*, 1–305.

[Winn, 2003] Winn, J. (2003). Variational Message Passing and its applications. PhD thesis, University of Cambridge.

[Winn and Bishop, 2005] Winn, J. and Bishop, C. (2005). Variational Message Passing. Journal of Machine Learning Research *6*, 661—694.

Résumé

Cette thèse propose de nouvelles méthodes d'analyse d'enregistrements cérébraux intra-crâniens (potentiels de champs locaux), qui pallie les lacunes de la méthode temps-fréquence standard d'*analyse des perturbations spectrales événementielles* : le calcul d'une moyenne sur les enregistrements et l'emploi de l'activité dans la période pré-stimulus.

La première méthode proposée repose sur la détection de sous-ensembles d'électrodes dont l'activité présente des *synchronisations cooccurrentes* en un même point du plan temps-fréquence, à l'aide de modèles bayésiens de mélange gaussiens. Les sous-ensembles d'électrodes pertinents sont validés par une *mesure de stabilité* calculée entre les résultats obtenus sur les différents enregistrements.

Pour la seconde méthode proposée, le constat qu'un bruit blanc dans le domaine temporel se transforme en bruit ricien dans le domaine de l'amplitude d'une transformée temps-fréquence a permis de mettre au point une segmentation du signal de chaque enregistrement dans chaque bande de fréquence en deux niveaux possibles, haut ou bas, à l'aide de modèles bayésiens de mélange ricien à deux composantes. À partir de ces deux niveaux, une analyse statistique permet de détecter des régions temps-fréquence plus ou moins actives. Pour développer le modèle bayésien de mélange ricien, de nouveaux algorithmes d'inférence bayésienne variationnelle ont été créés pour les distributions de Rice et de mélange ricien.

Les performances des nouvelles méthodes ont été évaluées sur des données artificielles et sur des données expérimentales enregistrées sur des singes. Il ressort que les nouvelles méthodes génèrent moins de faux-positifs et sont plus robustes à l'absence de données dans la période pré-stimulus.

Mots-clefs : modèles bayésiens, synchronisations corticales, représentations temps-fréquence, analyse simple essai, distribution de Rice, inférence bayésienne variationnelle

Abstract

This thesis promotes new methods to analyze intracranial cerebral signals (local field potentials), which overcome limitations of the standard time-frequency method of *event-related spectral perturbations analysis*: averaging over the trials and relying on the activity in the pre-stimulus period.

The first proposed method is based on the detection of sub-networks of electrodes whose activity presents *cooccurring synchronisations* at a same point of the time-frequency plan, using bayesian gaussian mixture models. The relevant sub-networks are validated with a *stability measure* computed over the results obtained from different trials.

For the second proposed method, the fact that a white noise in the temporal domain is transformed into a rician noise in the amplitude domain of a time-frequency transform made possible the development of a segmentation of the signal in each frequency band of each trial into two possible levels, a high one and a low one, using bayesian rician mixture models with two components. From these two levels, a statistical analysis can detect time-frequency regions more or less active. To develop the bayesian rician mixture model, new algorithms of variational bayesian inference have been created for the Rice distribution and the rician mixture distribution.

Performances of the new methods have been evaluated on artificial data and experimental data recorded on monkeys. It appears that the new methods generate less false positive results and are more robust to a lack of data in the pre-stimulus period.

Keywords: bayesien models, cortical synchronisation, time-frequency representation, single trial analysis, Rice distribution, variational bayesian inference