

THÈSE

présentée

pour obtenir le titre de

DOCTEUR DE L'INSTITUT NATIONAL POLYTECHNIQUE DE TOULOUSE

Ecole Doctorale d'Informatique et de Télécommunications

Spécialité : Signal, Image, Acoustique et Optimisation

par

Nicolas DOBIGEON

MODÈLES BAYÉSIENS HIÉRARCHIQUES POUR LE TRAITEMENT MULTI-CAPTEUR

Soutenue le 19 octobre 2007 devant le jury composé de :

M. Patrick FLANDRIN	Directeur de Recherche C.N.R.S.	Président
M. Jérôme IDIER	Directeur de Recherche C.N.R.S.	Rapporteur
M. Eric MOULINES	Professeur E.N.S.T. Paris	Rapporteur
M. Jean-François CARDOSO	Directeur de Recherche C.N.R.S.	Examineur
M. Manuel DAVY	Chargé de Recherche C.N.R.S.	Examineur
M. Jean-Yves TOURNERET*	Professeur E.N.S.E.E.I.H.T. Toulouse	Examineur
M. Emmanuel CHRISTOPHE	Ingénieur de Recherche C.N.E.S.	Membre invité

* Directeur de thèse

préparée à l'Institut de Recherche en Informatique de Toulouse

IRIT/INP-ENSEEIHHT – 2, rue Charles Camichel – BP7122 – 31071 Toulouse Cedex 7

Remerciements

Je ne peux débiter ces quelques lignes sans adresser ma plus profonde gratitude à Jean-Yves Tourneret qui m'a proposé cette thèse un jour de novembre 2003. Outre un adversaire de couinche redoutable, il a été tout au long de ces trois années un directeur de thèse exemplaire, me laissant une grande liberté dans mes recherches, tout en étant prolix de conseils avisés. Il a su guider mon travail avec intelligence et faire preuve de disponibilité en toute circonstance. Favorisant des échanges nombreux avec la communauté scientifique française et internationale, je lui suis également redevable des nombreuses heures de vols qui m'ont conduit à Lille, Paris et Nantes, mais aussi à Louvain, Florence, Honolulu et Madison. Il restera sans aucun doute celui qui m'a donné la chance d'apprendre le métier de chercheur dans des conditions qui feraient de nombreux envieux.

Patrick Flandrin m'a fait l'honneur de présider le jury, qu'il soit remercié pour son intérêt pour mon travail. Pour avoir fait une lecture minutieuse du manuscrit, j'aimerais adresser mes remerciements à Jérôme Idier et Éric Moulines qui ont accepté de rapporter cette thèse, ainsi qu'à Manuel Davy et Jean-François Cardoso pour l'avoir examinée. Merci enfin à Emmanuel Christophe pour avoir accepté, « au pied levé, » de participer à ce jury.

Comme évoqué plus haut, j'ai eu la chance de bénéficier de l'aide de nombreux chercheurs, soit lors de visites dans leurs laboratoires, soit lors de discussions plus informelles. Parmi les collaborations entreprises, je remercie Jeffrey Scargle, chercheur à la NASA, pour les échanges fructueux que nous avons entretenus. Sa disponibilité, sa réactivité et son expertise ont permis de concrétiser une partie de ces travaux de thèse. Je remercie une nouvelle fois Manuel Davy, mais cette fois pour son accueil au LAGIS au mois de Juillet 2005. A un moment de doute, il a donné le petit coup de pouce alors

indispensable au déblocage de la situation. J'en profite pour adresser une pensée amicale à François Caron dont l'aide a par la suite été précieuse dans la rédaction de mes demandes de bourses post-doctorales. Cette visite à Lille a été financée par le GdR-ISIS au travers du « réseau Doctorants. » Je remercie donc Christian Doncarli, alors responsable du réseau, et Cédric Richard, son successeur aujourd'hui. Le GdR-ISIS a également accepté de soutenir le projet « Jeunes Chercheurs » sur les méthodes MCMC pour l'analyse d'images hyperspectrales. Ce projet m'a donné l'occasion de rendre une visite à l'équipe ADTSI de l'IRCCyN en décembre 2006. Je salue alors les membres nantais du projet, Eric Le Carpentier, une nouvelle fois Jérôme Idier, et particulièrement Saïd Moussaoui avec qui j'ai démarré une collaboration que j'espère encore fructueuse. Je le remercie chaleureusement pour ses conseils toujours pertinents et pour sa bonne humeur. Lors d'un séjour à l'ENST au début de l'année 2007, j'ai pu profiter des connaissances de Jean-François Cardoso en matière de séparation de sources. Je me permets donc de le remercier une nouvelle fois pour sa disponibilité. Parmi les contributeurs à ce travail de thèse, je tiens à remercier Gérard Letac, professeur émérite à l'Université Paul Sabatier, pour le temps qu'il a su m'accorder pour m'aider dans mes recherches. Alfred Hero, professeur à l'Université du Michigan, a accepté de me recevoir à Ann Arbor dans le cadre d'un post-doctorat. Je ne doute pas que notre collaboration sera constructive. Enfin, je tiens à saluer les jeunes chercheurs croisés ici ou là, en congrès ou ailleurs, notamment Sébastien Bourguignon pour nos discussions passionnés d'apprentis bayésiens.

Chaque jour (ou presque) de ces trois années, ce fut un réel plaisir de traîner mes souliers au laboratoire IRIT. Je me dois de profiter de l'occasion qu'il m'est offerte pour témoigner ma reconnaissance à ceux que j'y ai croisés, et qui font (ou ont fait) la vie de l'équipe SC. Merci à Corinne, Nathalie et Marie pour leur constante gentillesse. Merci à Martial, gastronome exigeant, et collaborateur sur le projet « Jeunes Chercheurs. » Merci aux doctorants et ATER, anciens et nouveaux, avec qui j'ai partagé ces moments privilégiés au Petit London, au Griffon ou ailleurs : Audrey, David, Jean-Pierre, Mathieu, Will, Vincent, Jean-Rémi, Jérôme, Olivier. Une pensée particulière pour Florent, compagnon de galère, si prompt à répondre à mes interrogations sur \LaTeX ou sur les lois multivariées. Évidemment, un grand merci à Sylvie, maître ès secrétariat, pour la vie plus facile qu'elle rend à tous ceux qui se perdent dans les méandres de l'administration.

J'ai également pu profiter de l'accueil, certes temporaire, du laboratoire TéSA. Il a, par la suite, financé quelques uns de mes déplacements. C'est donc avec plaisir que je salue les doctorants de l'équipe domiciliés au 14-16 Port Saint-Etienne : Anchalee, Mariana et Xavier. Merci également à Sarah et Marie-Jo pour leur professionnalisme sans faille. Merci encore à Emmanuel, expert en imagerie hyperspectrale et en Asie du Sud-Est pour ses inestimables conseils, dans un domaine ou dans l'autre.

Dans la vie, il n'y a pas que le boulot. Quel aurait été mon quotidien dans la ville rose sans ceux que je ne résiste pas à nommer maintenant ? Un grand merci aux ex-poitevins, Simon et Anaïs, Laurence et Al, Erman, Sébastien, Chi-Kien et Benoît pour toutes les soirées passées en leur compagnie. Merci à Anaïs et Clément, bientôt ex-proches voisins, pour les parties de cartes endiablées, pour les discussions passionnées, et pour leur hospitalité dans la maison du bonheur. Quelque part sur les hauteurs de Fitou, dans le désordre, j'ai pu y croiser Nathou, Mathieu, Sam, Suzanna, Karine et Kossy. Merci aussi à Julie et Nico pour leurs invitations répétées au soleil.

Merci à Hugues pour sa gentillesse hors norme et tout ce que nous avons partagé.

Partout ailleurs, quelques autres ont réussi à me faire oublier, un peu, mes problèmes d'inférence bayésienne. Merci notamment à Abde et à Sylvain pour leur accueil, respectivement, au Caire et à Singapour. Merci à David, Jérôme et Max pour ces quelques jours sur la route en Allemagne ou en ex-Yougoslavie.

Au cours de ces trois années, dans quelques uns de mes engagements, j'ai eu la chance de côtoyer des personnes remarquables. Je ne citerai que Diane et Gabriel, Claire et Geneviève pour lesquels j'ai énormément d'estime. Aussi, derrière ces hauts murs, une foule d'« anonymes » m'ont fait comprendre à quel point j'avais la chance de faire ce qui me plaisait. Qu'ils en soient remerciés.

Merci à Pierre et Annick, mes parents, pour leur soutien discret, attentif et précieux.

Enfin, merci à celle qui a accompagné ce travail à mes côtés. C'est bien peu de chose que de dire qu'il n'aurait jamais abouti sans elle. Je lui dois aujourd'hui beaucoup.

Nicolas.

Résumé

Afin de traiter la masse d'informations récoltée dans de nombreuses applications, il est nécessaire de proposer de nouvelles méthodes de traitement permettant d'exploiter le caractère « multi-capteur » des données observées. Le sujet de cette thèse consiste à étudier des algorithmes d'estimation dans un contexte multi-capteur où plusieurs signaux ou images issus d'une même application sont disponibles. Ce problème présente un grand intérêt puisqu'il permet d'améliorer les performances d'estimation par rapport à une analyse qui serait menée sur chaque signal indépendamment des autres. Nous avons développé dans ce contexte des méthodes d'inférence bayésienne hiérarchique afin de résoudre efficacement des problèmes de segmentation de signaux multiples et d'analyse d'images hyperspectrales. L'utilisation de méthodes de Monte Carlo par chaînes de Markov permet alors de surmonter les difficultés liées à la complexité calculatoire de ces méthodes d'inférence.

Abstract

In order to process the mass of information collected in many applications, it is necessary to propose new processing methods to exploit the “multi-sensor” feature of the observed data. The subject of this thesis consists in studying algorithms of estimation in a multi-sensor context where several signals or images resulting from the same application are available. This problem is of great interest since it makes it possible to improve the estimation performances compared to an analysis that would be carried out on each signal independently of the others. In this context, we have developed methods of hierarchical bayesian inference to perform segmentation of multiple signals and to analyze hyperspectral images. The use of Markov Chain Monte Carlo methods allows one to overcome the difficulties related to the computational complexity of these inference methods.

Abréviations et notations

Abréviations

AR	autorégressif
DAG	graphe acyclique orienté (<i>directed acyclic graph</i>)
EM	<i>expectation-maximisation</i>
EQM	erreur quadratique moyenne
i.i.d.	indépendants et identiquement distribués
MCMC	méthodes de Monte Carlo par chaînes de Markov
MAP	maximum <i>a posteriori</i>
MMSE	estimateur qui minimise l'erreur quadratique moyenne (<i>minimum mean square error</i>)
pdf	fonction densité de probabilité (<i>probability density function</i>)
RJMCMC	méthodes de Monte Carlo par chaînes de Markov à sauts réversibles (<i>reversible jump MCMC</i>)
RSB	rapport signal sur bruit
SAEM	<i>stochastic approximation</i> EM
SEM	<i>stochastic</i> EM

Notations standard

\propto	proportionnel à
\ll	très inférieur à
\gg	très supérieur à
$\mathbf{1}_{\mathbb{E}}(\mathbf{x})$	fonction indicatrice définie sur l'espace \mathbb{E} :
	$\begin{cases} \mathbf{1}_{\mathbb{E}}(\mathbf{x}) = 1, & \text{si } \mathbf{x} \in \mathbb{E}; \\ \mathbf{1}_{\mathbb{E}}(\mathbf{x}) = 0, & \text{sinon.} \end{cases}$
$\Gamma(\cdot)$	fonction Gamma,
\mathbb{R}	ensemble des réels
\mathbb{R}^+	ensemble des réels positifs

Notations matricielles

x	scalaire
\mathbf{x}	vecteur
\cdot^{\top}	transposé
$\mathbf{x}_{i:j}$	sous vecteur $[x_i, x_{i+1}, \dots, x_{j-1}, x_j]$ de \mathbf{x}
\mathbf{X}	matrice
$ \mathbf{X} $	déterminant de la matrice \mathbf{X}
$\text{tr}\{\mathbf{X}\}$	trace de la matrice \mathbf{X}
$\text{etr}\{\mathbf{X}\}$	valeur exponentielle de la trace de la matrice \mathbf{X}
\mathbf{X}_i	i -ième colonne de la matrice \mathbf{X}
\mathbf{X}_{-i}	matrice \mathbf{X} privée de sa i -ième colonne
\mathbf{I}_p	matrice identité de taille $p \times p$
$\mathbf{0}_p$	vecteur nul de \mathbb{R}^p
\mathbf{O}_p	matrice nulle de taille $p \times p$
$\ \mathbf{x}\ ^2$	norme L^2 standard $\ \mathbf{x}\ ^2 = \mathbf{x}^{\top}\mathbf{x}$
$\ \mathbf{x}\ _{\mathbf{P}}^2$	norme euclidienne $\ \mathbf{x}\ _{\mathbf{P}}^2 = \mathbf{x}^{\top}\mathbf{P}\mathbf{x}$

Notations relatives aux problèmes de segmentation

n	longueur du signal
j	indice du signal
J	nombre de signaux
$y_{j,i}$	i -ième réalisation du j -ième signal
k_j	indice du k -ième segment dans le j -ième signal
K_j	nombre de segments dans le j -ième signal
$l_{j,k}$	instant après lequel a lieu la k -ième rupture dans le j -ième signal
$n_{j,k}$	longueur du k -ième segment dans le j -ième signal

Notations relatives au problème de démélange spectral

r	indice du composant pur
R	nombre de composants purs
l	indice de la bande spectrale
L	nombre de bandes spectrales

Algorithmes d'estimation statistique

\hat{x}	estimation de x
$\mathbb{E}[x y]$	espérance de x conditionnellement à y
$\text{var}[x y]$	variance de x conditionnellement à y
$\text{P}[A]$	probabilité de l'événement A
\sim	distribué suivant
N_{MC}	longueur totale de la chaîne de Markov
N_{bi}	longueur de la période de chauffage de la chaîne de Markov
N_r	nombre d'itérations de la chaîne de Markov construite pour réaliser les estimations
$x^{(k)}$	k -ième échantillon du processus $(x^{(t)})_{t=1,\dots,N_{\text{MC}}}$

Distributions de probabilité usuelles

$\mathcal{U}_{\mathbb{E}}$	loi uniforme sur l'ensemble \mathbb{E}
$\mathcal{B}e(a, b)$	loi Beta de paramètres a et b
$\mathcal{N}(\mathbf{x}, \Sigma)$	loi normale de vecteur moyenne \mathbf{x} et de matrice de covariance Σ
$\phi(\cdot \mathbf{x}, \Sigma)$	fonction densité de probabilité de la loi normale $\mathcal{N}(\mathbf{x}, \Sigma)$
$\mathcal{N}_{\mathbb{E}}(\mathbf{x}, \Sigma)$	loi normale tronquée à l'ensemble \mathbb{E} de vecteur moyenne \mathbf{x} et de matrice de covariance Σ (voir la définition dans l'annexe C)
$\phi_{\mathbb{E}}(\cdot \mathbf{x}, \Sigma)$	fonction densité de probabilité de la loi normale tronquée $\mathcal{N}_{\mathbb{E}}(\mathbf{x}, \Sigma)$
$\mathcal{IG}(a, b)$	loi inverse Gamma de paramètres a et b (voir [RC99] pour la définition des paramètres a et b)
$\mathcal{G}(a, b)$	loi Gamma de paramètres a et b (voir [RC99] pour la définition des paramètres a et b)
$\mathcal{D}_p(\boldsymbol{\alpha})$	loi de Dirichlet de vecteur paramètre $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_p]$

Table des matières

Remerciements	iii
Résumé	vii
Abstract	ix
Abréviations et notations	xi
Introduction	1
1 Segmentation conjointe de signaux astronomiques	7
1.1 Brève taxinomie des méthodes de segmentation	7
1.2 Position du problème	9
1.3 Mises en forme des données	11
1.4 Description du modèle bayésien hiérarchique	13
1.5 Échantillonneur de Gibbs pour la détection de ruptures	20
1.6 Diagnostic de convergence	22
1.7 Segmentation de données synthétiques	24
1.8 Analyse de données astronomiques réelles	30
1.9 Conclusions	34
2 Segmentation conjointe de processus autorégressifs	37
2.1 Introduction	37

2.2	Modèle bayésien hiérarchique	39
2.3	Échantillonneur de Gibbs pour la segmentation conjointe de processus AR	45
2.4	Segmentation de données synthétiques	48
2.5	Modèles alternatifs	57
2.6	Ordres des modèles inconnus	60
2.7	Applications	65
2.8	Conclusions	71
3	Démélange linéaire d'images hyperspectrales	73
3.1	Introduction	73
3.2	Modèle de mélange linéaire	78
3.3	Modèle bayésien hiérarchique	80
3.4	Échantillonneur de Gibbs pour l'estimation des abondances	82
3.5	Résultats de simulations sur des données synthétiques	84
3.6	Démélange spectral d'une image AVIRIS	89
3.7	Cas d'un bruit gaussien corrélé	95
3.8	Estimation du nombre de pôles de mélange à l'aide d'un algorithme à sauts réversibles	105
3.9	Conclusions	112
	Conclusions et perspectives	115
A	Calcul et simulation rapides d'une loi normale multivariée	121
B	Calcul de la loi <i>a posteriori</i> des coefficients d'abondance	123
C	Loi multivariée normale tronquée	125
D	Simulation efficace d'une loi normale multivariée tronquée à un simplexe	127
D.1	Position du problème	127
D.2	Echantillonneur de Gibbs pour loi multivariée normale tronquée	127
D.3	Simulations	129

E	Loi de Wishart : définition et principales propriétés	131
E.1	Loi de probabilité	131
E.2	Principales propriétés	132
F	Calcul de la loi <i>a posteriori</i> de la matrice de covariance	133
G	Calcul des probabilités d'acceptation des mouvements réversibles	135
	Liste des publications	137
	Bibliographie	154

Table des figures

1.1	À gauche : satellite CGRO. À droite : capteur BATSE.	9
1.2	DAG pour les lois <i>a priori</i> (les hyperparamètres qui sont fixés apparaissent dans les carrés en pointillés).	19
1.3	Probabilités <i>a posteriori</i> des instants de ruptures pour une segmentation 1D (à gauche) et conjointe (à droite).	25
1.4	Lois <i>a posteriori</i> de K_1 et K_2 pour des segmentation unidimensionnelle (à gauche) et conjointe (à droite).	26
1.5	Lois <i>a posteriori</i> des paramètres de Poisson $\lambda_{1,k}$ (pour $k = 1, \dots, 4$) sachant $K_1 = 4$	27
1.6	Lois <i>a posteriori</i> des paramètres de Poisson $\lambda_{2,k}$ (pour $k = 1, 2$) sachant $K_2 = 2$	27
1.7	Lois <i>a posteriori</i> des hyperparamètres P_{00}, P_{01}, P_{10} et P_{11}	27
1.8	Suivi de la convergence de cinq chaînes de Markov.	28
1.9	Erreur quadratique moyenne entre l'estimateur de référence et l'estimateur obtenu après p itérations (trait plein). EQMs moyennées à partir de 64 chaînes (tirets).	29
1.10	Probabilités <i>a posteriori</i> des instants de ruptures pour deux valeurs de l'hyperparamètre γ : $\gamma = 0.05$ (à gauche) et $\gamma = 5$ (à droite).	30
1.11	En haut : données mises en forme et segmentées. En bas : loi <i>a posteriori</i> des instants de ruptures (données astronomiques unidimensionnelles).	31
1.12	Loi <i>a posteriori</i> du nombre de ruptures (données astronomiques unidimensionnelles).	32
1.13	Lois <i>a posteriori</i> des nombres de ruptures (données astronomiques 4D).	34
1.14	Probabilités <i>a posteriori</i> des instants de ruptures (données astronomiques 4D).	35

1.15	Représentation par blocs (données astronomiques 4D).	35
2.1	DAG pour les lois <i>a priori</i> (les hyperparamètres qui sont fixés apparaissent dans les carrés en pointillés).	44
2.2	Probabilités <i>a posteriori</i> des instants de ruptures obtenus pour des segmentations 1D (à gauche) et conjointe (à droite) obtenues après une période de chauffage de $N_{\text{bi}} = 200$ itérations et $N_r = 500$ itérations d'intérêt.	49
2.3	Lois <i>a posteriori</i> des nombres de rupture calculées après $N_r = 500$ itérations d'intérêt (moyenne en gris, moyenne \pm écart type en blanc et noir).	50
2.4	Lois <i>a posteriori</i> des variances de bruit $\sigma_{1,k}^2$ (pour $k = 1, \dots, 3$) sachant $K_1 = 3$ calculés à partir $N_r = 500$ itérations d'intérêt (trait plein). Lois <i>a posteriori</i> moyennées à partir des résultats fournis par 64 Markov chaînes (pointillés).	51
2.5	Lois <i>a posteriori</i> des variances $\sigma_{2,k}^2$ (pour $k = 1, 2$) sachant $K_2 = 2$ calculés à partir de $N_r = 500$ itérations d'intérêt (trait plein). Lois <i>a posteriori</i> moyennées sur 64 chaînes de Markov (pointillés),	51
2.6	Lois <i>a posteriori</i> des hyperparamètres P_ϵ (calculées à partir de $N_r = 500$ itérations d'intérêts de 64 chaînes de Markov).	52
2.7	Probabilités <i>a posteriori</i> des instants de ruptures pour des vecteurs de bruit corrélés avec $\rho = 0.1$ (en haut) et $\rho = 0.5$ (en bas) ($N_{\text{bi}} = 200$ et $N_r = 500$).	53
2.8	Diagnostic de convergence : exemples de $M = 5$ chaînes parallèles relatives au paramètre P_{00}	55
2.9	Erreur quadratique moyenne entre l'estimateur de référence et l'estimateur des probabilités <i>a posteriori</i> de ruptures en fonction de p (trait plein). EQMs moyennées à partir de 64 chaînes (tirets).	56
2.10	Probabilités <i>a posteriori</i> des instants de ruptures pour deux jeux d'hyperparamètres : $\delta_0^2 = 0.5$ (à gauche) et $\gamma = 5$ (à droite).	57
2.11	Probabilités <i>a posteriori</i> des instants de ruptures pour des algorithmes avec un hyperparamètre γ (gauche) et plusieurs hyperparamètres $\gamma_j, j = 1, \dots, J$ (right).	59

2.12	Probabilités <i>a posteriori</i> des instants de ruptures pour des <i>a priori</i> indépendants (à droite) et markovien (à droite) pour \mathbf{R}	61
2.13	Lois <i>a posteriori</i> des ordres des processus p_{1i} (pour $i = 1, \dots, 3$) sachant $K_1 = 3$	66
2.14	Lois <i>a posteriori</i> des ordres des processus p_{2i} (pour $i = 1, 2$) sachant $K_2 = 2$	66
2.15	Probabilités <i>a posteriori</i> des instants de ruptures estimées par l’algorithme à sauts réversibles.	67
2.16	Segmentations 1D (en haut) et 2D (au milieu et en bas) d’un signal de parole réel ($N_{\text{bi}} = 200$ et $N_{\text{MC}} = 600$).	68
2.17	Probabilités <i>a posteriori</i> des instants de ruptures pour la segmentation conjointe de signaux aéronautiques 3D ($N_{\text{bi}} = 100$ et $N_{\text{MC}} = 450$).	71
2.18	Probabilités <i>a posteriori</i> des instants de ruptures et résultat de segmentation des données aéronautiques triphasées ($N_{\text{bi}} = 100$ et $N_{\text{MC}} = 450$).	72
3.1	Un exemple d’image hyperspectrale.	74
3.2	Illustration d’un cube hyperspectral dont 3 pixels ont été extraits.	75
3.3	Spectres en réflectance du séquoia sec et vert.	76
3.4	En haut - spectres des composants purs : béton de construction (trait plein), herbe verte (tirets), terre grasse micacée (pointillés). En bas - spectre résultant du pixel mélangé.	85
3.5	Lois <i>a posteriori</i> des abondances estimées $[\alpha_1, \alpha_2, \alpha_3]^T$ (en trait plein) et histogramme des estimateurs FCLS (en pointillés).	86
3.6	Estimateurs MAP (croix) et écarts-types (barres verticales) des paramètres α_r ($r = 1, \dots, 3$) en fonction du RSB.	87
3.7	Taux d’acceptation théorique (trait plein) et expérimental (pointillé) du test d’acceptation/rejet en fonction du RSB.	88
3.8	Vérification de la convergence avec cinq réalisations de chaînes de Markov.	89
3.9	Erreur quadratique entre l’estimateur de référence et l’estimateur après p itérations en fonction de p (trait plein). Erreur quadratique moyennée sur 10 chaînes (tirets) ($N_{\text{bi}} = 100$).	90

3.10	Données hyperspectrales réelles : l'image de Moffett Field acquise par AVIRIS en 1997 (à gauche) et la région d'intérêt à la longueur d'onde $\lambda = 0.66\mu m$ représentée en échelle de gris (à droite).	90
3.11	Les $R = 3$ spectres des composants purs obtenus par l'algorithme N-FINDR.	92
3.12	En haut : lois <i>a posteriori</i> des abondance α_r ($r = 1, \dots, 3$) pour le pixel $\#(43, 35)$. En bas : 150 premières échantillons simulés par l'algorithme de Gibbs.	93
3.13	En haut : cartes d'abondances estimées par l'algorithme proposé. En bas : cartes d'abondances estimées par le logiciel ENVI.	94
3.14	Diagnostic de convergence : exemples de $M = 5$ chaînes parallèles relatives au paramètre σ^2	94
3.15	Facteurs d'échelle calculés pour chaque pixel de l'image de Moffett Field.	95
3.16	Image de Moffett Field segmentée à l'aide de l'algorithme K-means.	95
3.17	En haut : les $R = 3$ pôles de mélange obtenue par la procédure K-means/PPI (trait plein) et par l'algorithme du N-FINDR (pointillés). En bas : les cartes d'abondance des composants purs estimés par la procédure K-means/PPI.	96
3.18	Les 50×50 premiers coefficients de la matrice de covariance Σ pour différentes valeurs du paramètre η	100
3.19	En haut - spectres des composants purs : herbe verte (trait plein), brique rouge (tirets), métal galvanisé (pointillés). En bas - spectre résultant du pixel mélangé.	101
3.20	Lois <i>a posteriori</i> des abondances $[\alpha_1, \alpha_2, \alpha_3]^T$ estimées par l'algorithme proposé (trait plein) et par l'algorithme initial (pointillés).	102
3.21	Spectre d'un pixel résultant d'un mélange avec des variances de bruits différentes d'une région spectrale à l'autre.	104
3.22	Lois <i>a posteriori</i> des abondances $[\alpha_1, \alpha_2, \alpha_3]^T$ estimées par l'algorithme proposé (trait plein) et par l'algorithme ALGO. 3.3 (pointillés) avec des RSBs hétérogènes.	105
3.23	Spectres des pôles de mélange de la librairie.	112
3.24	Loi <i>a posteriori</i> de l'ordre du modèle estimé R	113
3.25	Lois <i>a posteriori</i> des abondances estimées $\alpha^+ = [\alpha_1, \alpha_2, \alpha_3]^T$ sachant $\mathbf{M}^+ = [\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3]$	114

D.1 En haut : histogramme de $N = 500000$ variables simulées avec l'échantillonneur de Gibbs proposé pour une loi normale multivariée tronquée au simplexe \mathbb{S} ($R = 3$). En bas : densité de probabilité théorique correspondante. 130

Liste des tableaux

1.1	Facteurs d'échelle relatifs aux paramètres P_ϵ (calculés à partir de $M = 64$ chaînes de Markov)	28
1.2	Facteurs d'échelle relatifs aux paramètres P_ϵ (calculés à partir de $M = 64$ chaînes de Markov)	36
2.1	Paramètres des modèles AR et variances des bruits pour chaque segment de chaque séquence.	49
2.2	Facteurs d'échelle relatifs aux paramètres P_ϵ , $\sigma_{1,k}^2$ et $\sigma_{2,k}^2$ (calculés à partir de $M = 64$ chaînes de Markov).	55
2.3	Paramètres des processus AR et variances des bruits pour chaque segment de chaque séquence.	66
2.4	Positions des ruptures estimées par différentes méthodes.	69
3.1	Variances des abondances estimées ($\times 10^{-5}$).	102
3.2	Valeurs réelles et estimées des hyperparamètres $\gamma_1, \gamma_2, \gamma_3$	104

Liste des algorithmes

1.1	Échantillonneur de Gibbs pour la segmentation conjointe de données poissonniennes	20
2.1	Échantillonneur de Gibbs pour la segmentation conjointe de processus AR	46
2.2	Algorithme MCMC à sauts réversibles pour la détection de ruptures	63
3.1	Échantillonneur de Gibbs pour le démélange spectral en imagerie hyperspectrale	82
3.2	Procédure d'acceptation/rejet pour la génération d'échantillons distribués suivant une loi multivariée normale tronquée	83
3.3	Echantillonneur de Gibbs pour le démélange spectral avec bruit corrélé	98
3.4	Échantillonneur de Gibbs hybride pour le démélange semi-supervisé d'images hyperspectrales	107
3.5	Algorithme MCMC à sauts réversibles pour l'imagerie hyperspectrale : mouvement de NAISSANCE	109
3.6	Algorithme MCMC à sauts réversibles pour l'imagerie hyperspectrale : mouvement de MORT	109
3.7	Algorithme MCMC à sauts réversibles pour l'imagerie hyperspectrale : mouvement d'ÉCHANGE	110
A.1	Calculs rapides de $T_{j,k}^2$	121
A.2	Echantillonnage rapide des paramètres autorégressifs $\mathbf{a}_{j,k}$	122

Introduction

Contexte et problématique de la thèse

Le travail de thèse rapporté dans ce document a été mené à l'Institut de Recherche en Informatique de Toulouse, au sein de l'équipe Signal et Communications. Il présente des méthodes de traitement statistique du signal mises en œuvre pour la résolution de problèmes en contexte « multi-capteur. » Depuis quelques années déjà, les réseaux de capteurs connaissent en effet un succès certain, dû en grande partie à la baisse des coûts de fabrication et d'utilisation des composants électroniques. Afin de traiter la masse d'informations récoltée par ces capteurs, il est nécessaire de proposer de nouvelles méthodes d'analyse permettant d'exploiter de manière efficace le caractère « multi-capteur » des données observées. Le sujet de cette thèse consiste à étudier des algorithmes d'estimation adaptés au contexte multi-capteur. Ce problème présente un grand intérêt puisqu'il permet d'améliorer les performances d'estimation par rapport à une analyse qui se ferait sur chaque signal indépendamment des autres.

De nombreux travaux abordent le traitement des signaux multiples comme un problème de fusion de données (voir [VV97] et les références citées). Nous avons développé dans cette thèse des méthodes d'inférence bayésienne permettant de résoudre différemment ce problème. Tout d'abord, une analyse statistique du phénomène physique à l'origine des données observées permet d'introduire une fonction de vraisemblance. Ensuite, les paramètres inconnus sont munis de lois *a priori* qui peuvent dépendre d'hyperparamètres. Les paramètres inconnus sont estimés à partir de leurs lois *a posteriori* via les estimateurs bayésiens standards, comme l'estimateur du maximum *a posteriori* (MAP) ou l'estimateur

qui minimise l'erreur quadratique moyenne appelé estimateur MMSE (pour *minimum mean square error*).

Très souvent, la complexité des lois *a posteriori* des paramètres ne permet pas d'obtenir des expressions simples des estimateurs bayésiens (MMSE, MAP). Il est alors nécessaire d'utiliser des méthodes de simulations appropriées, comme les méthodes de Monte Carlo par chaînes de Markov (méthodes MCMC) qui consistent à générer des échantillons distribués asymptotiquement suivant la loi d'intérêt [RC99]. Ces échantillons sont alors utilisés pour calculer des estimateurs des paramètres inconnus. Notons qu'une alternative a été récemment proposée par Fearnhead, baptisée méthode d'« échantillonnage parfait » qui permet de générer de manière exacte des échantillons suivant la loi d'intérêt [Fea05; Fea06]. Toutefois, cette méthode n'autorise pas la construction des modèles hiérarchiques étudiés dans cette thèse.

Par ailleurs, la détermination des hyperparamètres est un problème majeur qui conditionne la performance de l'estimation. Lorsqu'une information concernant les paramètres inconnus est disponible *a priori* (nombre de ruptures à estimer, valeurs moyennes des signaux observés...), ces hyperparamètres peuvent être fixés préalablement. Dans le cas où une telle information n'est pas accessible, les hyperparamètres doivent être estimés. Une première approche permettant cette estimation est de coupler les méthodes MCMC avec des algorithmes de type EM ou SAEM [KL04; LL01]. Malheureusement, ces algorithmes peuvent converger vers des maxima locaux [DI96, p. 259]. Nous préférons dans notre étude nous concentrer sur une autre approche, entièrement bayésienne, qui réside en la définition de lois *a priori* non informatives sur les hyperparamètres, ce qui introduit naturellement un second niveau de hiérarchie dans l'inférence bayésienne. On obtient alors des modèles bayésiens hiérarchiques [AD99; RG97; TG01; DDT02; SI06].

Organisation du manuscrit

Le premier chapitre du présent manuscrit est consacré à la segmentation conjointe de signaux astronomiques. Une modélisation bayésienne hiérarchique permet de concevoir un algorithme qui réalise

la détection des changements de brillance dans des données issues de comptage de photons. Nous proposons ainsi une alternative totalement non-supervisée à la stratégie adoptée par Scargle dans [Sca98]. L'algorithme développé ne nécessite pas en effet de règle d'arrêt. La procédure proposée permet également de traiter conjointement les données enregistrées par l'instrument BATSE dans quatre bandes d'énergie. Grâce à des simulations effectuées sur des signaux synthétiques et réels, nous montrons que ce traitement multi-capteur permet de proposer des modèles de segmentation plus simples que ceux initialement estimés dans [Sca98]. Notons que l'introduction de vecteurs d'indicatrices pour repérer les changements dans les séries temporelles observées permet d'éviter le recours à des méthodes MCMC à sauts réversibles, comme dans [PADF02]. Enfin, pour chacune des simulations réalisées, nous proposons des diagnostics de convergence afin de s'assurer des bonnes propriétés de mélange de l'échantillonneur de Gibbs et de la précision des estimations effectuées.

Le second chapitre est consacré à la segmentation conjointe de processus autorégressifs (AR). Nous proposons une extension des travaux de Punskeya *et al.* présentée dans [PADF02] au cas où les données traitées sont issues de capteurs multiples. Comme au premier chapitre, nous préférons introduire des vecteurs d'indicatrices plutôt que d'avoir recours à un algorithme à sauts réversibles pour repérer les ruptures dans les différentes séquences. Dans un premier temps, les ordres des modèles AR sont supposés connus mais cette hypothèse est relâchée dans la dernière partie de l'analyse. Dans un contexte multi-capteur, il est possible que les signaux mesurés soient caractérisés par des dynamiques très différentes. Pour tenir compte de ces différences, nous proposons un modèle alternatif basé sur un jeu d'hyperparamètres différents. Aussi, il arrive que les modèles de segmentation estimés comportent des changements trop proches. C'est le cas par exemple lorsqu'une donnée aberrante est présente dans le signal. Nous montrons comment un modèle markovien sur les indicatrices permet de résoudre ce problème grâce à l'introduction d'une contrainte sur la longueur minimale des segments. Deux applications sont envisagées pour illustrer les performances de l'algorithme. Tout d'abord, nous envisageons une application qui a fait l'objet de nombreux travaux dans la littérature du traitement du signal : la segmentation de la parole. Nous nous attachons dans ce manuscrit au cas moins classique où le signal de parole est multi-canal, en l'occurrence stéréo. Ensuite, une deuxième série de

simulations est consacrée à la segmentation de signaux électriques dans un contexte aéronautique. L'application visée est la détection du phénomène d' « arc tracking, » responsable d'un grand nombre d'accidents ces dernières années [FH01].

Le troisième et dernier chapitre de ce document est consacré à l'analyse d'images multi-bandes. Le pixel d'une image hyperspectrale est représenté non plus, comme dans le cas des images de télédétection classique, par une valeur scalaire mais par un vecteur de mesures appelé spectre. Un problème classique de l'imagerie hyperspectrale consiste en l'identification des composants purs présents dans l'image et leurs proportions respectives dans chacun des pixels. Par la nature même du mélange, des contraintes de positivité et d'additivité sont imposés aux coefficients du mélange. Nous posons ce problème de démixage spectral comme un problème de régression linéaire sous contraintes. Une loi *a priori* adéquate permet d'inclure ces contraintes dans le modèle. C'est, à notre connaissance, la première fois qu'un problème de régression linéaire sous ces contraintes est résolu dans un formalisme bayésien. La première partie de l'analyse est conduite sous l'hypothèse que les spectres des composants purs ont été préalablement identifiés. L'estimation des coefficients d'abondance est réalisée à partir de leur loi *a posteriori* et permet donc de calculer des intervalles de confiance pour ces estimations. Le coût calculatoire de l'algorithme est principalement dû à la simulation de variables normales tronquées à un simplexe. Afin d'évaluer ce coût, nous calculons le taux d'acceptation de l'algorithme d'acceptation/rejet mis en place. Dans le cas où il se montre inefficace, nous adoptons une stratégie basée sur les travaux de Robert [Rob95]. Un algorithme de Gibbs reporté en annexe permet alors de simuler successivement chaque composante du vecteur Gaussien. Nous envisageons également le cas où le bruit de mesure est corrélé et structuré par des rapports signal-sur-bruit différents dans plusieurs plages de longueur d'onde, comme c'est le cas avec certains spectro-imageurs réels. Le recours à des lois multivariées décrivant les propriétés statistiques de la matrice de covariance du bruit est alors indispensable. Enfin, dans un deuxième temps de l'analyse, les spectres des composants purs présents dans l'image sont supposés inconnus mais appartenant à une librairie spectrale qui joue le rôle de dictionnaire pour les régresseurs. Il convient alors d'estimer conjointement le nombre de matériaux présents dans le pixel, leurs natures et leurs proportions respectives. Il en

résulte un algorithme de démixage spectral semi-supervisé basé sur un algorithme à sauts réversibles.

Principales contributions

Les principales contributions de cette thèse sont rappelées ci-dessous.

Chapitre 1. L'originalité de nos travaux réside en le développement d'un algorithme n'ayant pas besoin de règle d'arrêt pour réaliser la segmentation de signaux astronomiques. Il est capable de traiter conjointement les données multi-capteur fournies par BATSE.

Chapitre 2. Nous introduisons un algorithme MCMC qui permet de segmenter conjointement des processus auto-régressifs enregistrés par plusieurs capteurs. Un modèle markovien sur les vecteurs d'indicateurs résout le problème de sur-segmentation dans le cas de données aberrantes. Nous traitons le cas où les signaux observés ont des dynamiques différentes. Les applications envisagées sont la segmentation de parole stéréo et de signaux aéronautiques.

Chapitre 3. Le problème de démixage spectral est abordé comme un problème de régression linéaire sous contraintes d'additivité et de positivité. Il se ramène à la simulation de lois normales tronquées à un simplexe qui est effectuée à l'aide d'un échantillonneur de Gibbs approprié. Le cas plus complexe d'un bruit d'observation structuré est résolu grâce à l'introduction de lois *a priori* judicieuses. Dans un contexte semi-supervisé, un algorithme MCMC à sauts réversibles est mis en place pour estimer les spectres des composants purs prenant part au mélange.

CHAPITRE 1

Segmentation conjointe de signaux astronomiques

Sommaire

1.1	Brève taxinomie des méthodes de segmentation	7
1.2	Position du problème	9
1.3	Mises en forme des données	11
1.4	Description du modèle bayésien hiérarchique	13
1.5	Échantillonneur de Gibbs pour la détection de ruptures	20
1.6	Diagnostic de convergence	22
1.7	Segmentation de données synthétiques	24
1.8	Analyse de données astronomiques réelles	30
1.9	Conclusions	34

1.1 Brève taxinomie des méthodes de segmentation

Les problèmes de détection de changements ont reçu un intérêt considérable dans la communauté du traitement du signal, et ce depuis plus de 25 ans. Une bibliographie complète des références publiés avant 2000 peut être trouvée dans les livres de Basseville et Nikiforov [BN93], Brodsky et Darkhovsky [BD93] ou Gustafsson [Gus00]. Cependant, depuis 2000, la mise en œuvre de nouveaux algorithmes performants s'avère possible, grâce notamment aux importantes charges calculatoires désormais supportées par les outils informatiques. D'une manière générale, on distingue deux types d'algorithmes de segmentation. La première classe de méthodes consiste en un traitement séquentiel des données observées, à mesure de leur disponibilité. On parle alors de segmentation « en ligne » : pour un taux de fausse alarme donné, un critère de décision est calculé successivement sur les portions de signal

$[y_1, \dots, y_i]$, puis $[y_1, \dots, y_{i+1}]$, puis $[y_1, \dots, y_{i+2}]$ etc..., jusqu'à ce qu'un changement soit détecté. Un grand nombre de ces algorithmes sont basés sur des tests d'hypothèses statistiques mettant en jeu des rapports de vraisemblance. Nous citerons par exemple les algorithmes CUSUM pour *Cumulative Sum* [BN93, p. 35] et GLR pour *Generalized Likelihood Ratio* [Nik01; WW04]. Notons que d'autres algorithmes de segmentation en ligne reposent par exemple sur des méthodes à noyau [DDD05] ou sur des techniques de filtrage [DF98; FL07].

Le travail de thèse rapporté dans ce manuscrit se concentre sur un deuxième type d'algorithme : les procédures de segmentation hors-ligne. Dans ce cas, nous supposons que l'ensemble du signal observé $[y_1, \dots, y_N]$ est connu. Il convient alors de déceler un ou plusieurs changements dans l'ensemble de cette séquence, de manière « rétrospective » (terme employé notamment par Inclan dans [IT94] et Rotondi dans [Rot02]). Là encore, une dichotomie des algorithmes de segmentation hors-ligne peut être opérée. Un grand nombre de stratégies développées se base sur une approche par sélection de modèle. Les algorithmes les plus connus sont basés sur les critères C_p introduit par Mallows [Mal73], AIC (*Akaike Information Criterion* [Aka74]) et BIC (*Bayesian Information Criterion* [Sch78]). Une description paramétrique est proposée, incluant les positions des ruptures entre un nombre inconnu de segments. Les changements sont alors estimés par optimisation d'un critère pénalisé approprié [Lav99; LM00]). Cette pénalisation, portant par exemple sur le nombre total de segments, permet de s'affranchir des problèmes de sur-segmentation. Plusieurs études ont été menées afin de choisir au mieux le terme de pénalisation [Lav98; LL00; BM01; Lav05; Leb05]. Une alternative à cette démarche de pénalisation réside en une formulation bayésienne du problème. Il est alors nécessaire de définir des lois *a priori* pour les paramètres inconnus. Ces lois induisent implicitement une pénalisation du terme d'attache aux données représenté par la vraisemblance. Comme nous l'avons précisé plus haut, l'utilisation de méthodes MCMC s'avère alors souvent nécessaire [MT93; Dju94; Chi98; LM99; WW03]. L'estimation des hyperparamètres est alors possible en introduisant un deuxième niveau de hiérarchie dans l'inférence [CGS92; Cap02]. C'est précisément cette approche totalement bayésienne qui a été retenue dans le présent chapitre.

1.2 Position du problème

Très peu de travaux issus de la littérature du traitement du signal sont consacrés à la segmentation de séries astronomiques. Pourtant, détecter les changements de brillance de différents objets est un problème fondamental en astronomie pour l'analyse temporelle des activités galactique et extra-galactique. Au début des années 1990, la NASA a lancé le satellite CGRO (*Compton Gamma Ray Observatory*, figure 1.1, à gauche) afin d'étudier plus précisément un phénomène méconnu jusqu'alors : les sursauts gamma. Embarqué à son bord, un instrument, appelé BATSE (*Burst and Transient Source Experiment*, figure 1.1, à droite), a pour mission d'enregistrer les instants d'arrivée des photons sur un capteur photo-sensible dans quatre bandes d'énergies différentes.

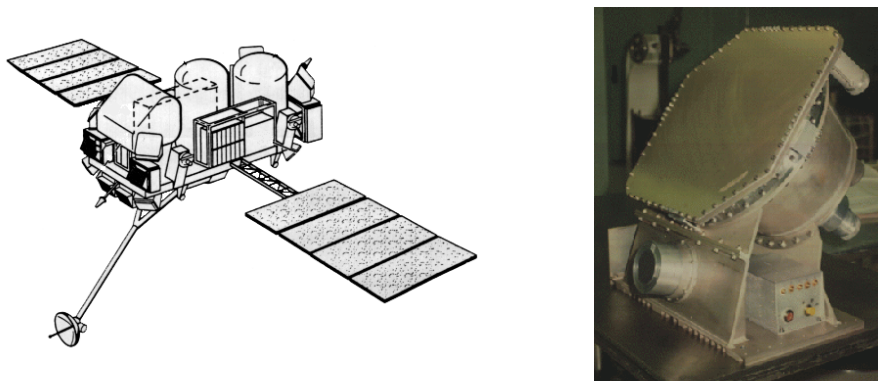


FIG. 1.1 – À gauche : satellite CGRO. À droite : capteur BATSE.

Un algorithme bayésien itératif basé sur un modèle de Poisson constant par morceaux a été récemment étudié pour résoudre ce problème [Sca98]. L'idée principale de cet algorithme est de décomposer le signal observé en deux sous-intervalles (en optimisant un critère approprié basé sur un rapport de vraisemblance), d'appliquer la même procédure sur les deux segments obtenus et de continuer cette opération autant de fois que nécessaire. L'avantage majeur de cette procédure est de choisir une seule rupture à chaque étape. Cependant, la performance de l'algorithme est limitée par sa complexité calculatoire et par le fait qu'une règle d'arrêt judicieuse a besoin d'être définie. L'algorithme à ruptures multiples présenté dans [JSB⁺05] permet de dépasser ces limitations, mais nécessite la connaissance d'une information *a priori* sur le nombre de ruptures. Il a également l'inconvénient de

ne pas fournir automatiquement d'information sur la précision des paramètres déterminés de façon optimale. Un autre inconvénient des algorithmes présentés dans [Sca98] et [JSB⁺05] est qu'ils ne permettent pas de traiter simultanément les quatre séquences de données fournies par le capteur BATSE.

Ce chapitre étudie un nouvel algorithme de segmentation bayésienne qui ne nécessite pas de règle d'arrêt et qui permet de segmenter conjointement plusieurs signaux issus de capteurs multiples. La stratégie proposée repose sur un modèle hiérarchique qui suppose que des lois *a priori* pour les paramètres inconnus sont disponibles. Des lois *a priori* vagues sont choisies pour les hyperparamètres qui sont soit intégrés dans la loi *a posteriori* dès que cela est possible, soit estimés grâce aux données observées. Des méthodes MCMC sont utilisées pour générer des échantillons distribués suivant la loi *a posteriori*. Les estimateurs bayésiens sont finalement calculés à partir de ces échantillons simulés. La méthodologie proposée est similaire au modèle bayésien hiérarchique développé dans [PADF02]. Pourtant, il convient de noter les deux points suivants. La méthode présentée dans [PADF02] traite de modèles de régression linéaire entachés d'un bruit additif gaussien et ne peut donc pas être directement appliquée à des données poissonniennes. Notre méthode peut donc être considérée comme une adaptation de cette approche à des données distribuées suivant une loi de Poisson. D'autre part, la procédure de segmentation que nous détaillons ici permet la segmentation conjointe de séquences enregistrées par plusieurs capteurs, ce qui n'était pas possible avec l'algorithme proposé dans [PADF02]. C'est, à notre connaissance, la première fois qu'un algorithme totalement bayésien est développé pour la segmentation conjointe de données poissonniennes.

Ce chapitre est organisé comme suit. Nous présentons au paragraphe 1.3 les différentes mises en forme possibles des données recueillies par l'appareil BATSE. Une inférence bayésienne hiérarchique est proposée au paragraphe 1.4 pour résoudre le problème de segmentation. La complexité de la loi *a posteriori* qui résulte du modèle hiérarchique est surmontée grâce à un échantillonneur de Gibbs détaillé au paragraphe 1.5. Les problèmes de diagnostics de convergence des méthodes MCMC sont traités au paragraphe 1.6. Enfin, des résultats de simulations menées sur des données synthétiques et réelles sont rapportées respectivement aux paragraphes 1.7 et 1.8.

1.3 Mises en forme des données

Comme cela est détaillé dans [Sca98], les données brutes fournies par des outils comme BATSE sont les temps d'arrivées de P photons $\{t_p, p = 1, \dots, P\}$ sur la surface sensible de l'appareil. Ces données peuvent être mises en forme selon différents modèles qui sont détaillés ci-après. Le mode TTE (*Time-Tagged Event*) est couramment utilisé en astronomie. Il suppose la connaissance d'une période d'échantillonnage δt qui correspond à la résolution temporelle de l'appareil. Cette constante est le délai minimal qui sépare l'arrivée de deux photons détectés. En d'autres termes, c'est le temps nécessaire au dispositif pour qu'il redevienne opérationnel après avoir détecté une particule. Ce temps est généralement très court comparé au phénomène astronomique qui nous intéresse (pour BATSE, $\delta t = 2\mu s$). Ainsi, les données observées pendant la durée $T = M\delta t$ sont représentées par leur indice d'arrivée $\mathcal{D}_{\text{TTE}} = \{m_p, p = 1, \dots, P\}$, où $t_p = m_p\delta t$ est le temps d'arrivée du p -ième photon. Une représentation élégante de ce mode consiste à introduire un processus binaire X_m ($m = 1, \dots, M$) tel que :

$$\begin{cases} X_m = 1 \text{ si un photon est détecté à l'instant } m\delta t, \\ X_m = 0 \text{ sinon.} \end{cases}$$

Ainsi X_m peut-il être considéré comme un processus de Bernoulli avec $P_0 = P[X_m = 0]$ et $P_1 = 1 - P_0$. Bien sûr, le mode TTE fournit les données les plus résolues du point de vue temporel. Cependant, les données binaires TTE semblent trop volumineuses pour donner lieu à une analyse facilement exploitable.

La deuxième manière de représenter les données est appelée *Time-To-Spill* (TTS). Pour réduire le volume occupé par les données brutes, seuls sont conservés les temps d'arrivée tous les S photons, où S est un entier (pour BATSE, $S = 64$). Les données disponibles sont alors les temps qui séparent l'arrivée de S photons consécutifs, $\mathcal{D}_{\text{TTS}} = \{\tau_l, l = 0, \dots, L - 1\}$, où $\tau_l\delta t$ est le temps écoulé entre l'enregistrement des $(lS + 1)$ -ième et $(lS + S)$ -ième photons. Classiquement, la loi statistique qui régit de telles données est une loi Gamma de paramètre Λ :

$$f(\tau_l|\Lambda) = \frac{\Lambda^S}{\Gamma(S)} \tau_l^{S-1} \exp(-\Lambda\tau_l) \mathbf{1}_{\mathbb{R}^+}(\tau_l), \quad (1.1)$$

où $\Gamma(\cdot)$ est la fonction Gamma (définie par $\Gamma(t) = \int_0^{+\infty} u^{t-1} e^{-u} du$, $t > 0$), $\mathbf{1}_{\mathbb{R}^+}(x)$ est la fonction

indicatrice définie sur \mathbb{R}^+ et $\Lambda = \frac{SL}{M}$ avec $M = \sum_{l=0}^{L-1} \tau_l$. Mais, dans une optique de traitement conjoint des séquences issues des quatre capteurs de BATSE, il est nécessaire que les phénomènes physiques aient des conséquences simultanées dans les quatre signaux après mise en forme. Lors d'un traitement des données TTS, il apparaît très vite que ce n'est plus le cas. À titre d'illustration, l'on peut citer simplement le fait qu'après mise en forme TTS, les quatre signaux numériques à traiter, bien qu'ils correspondent à une seule mesure, ne sont plus des séquences de même longueur.

Enfin, le dernier mode de représentation des données consiste à compter le nombre de photons reçus durant un intervalle de temps Λ . Les données peuvent alors être modélisées par un processus discret de comptage de Poisson. Le nombre de photons reçus dans n intervalles successifs régulièrement espacés par le j -ième capteur est noté $y_{j,i}$, où $y_{j,i}$ est le nombre de particules reçues dans l'intervalle i du signal j ($i = 1, \dots, n$ et $j = 1, \dots, J$). Les intervalles sont regroupés en segments, contenant un ou plusieurs intervalles, appelés blocs sur lesquels on suppose que le signal suit une loi de Poisson dont le paramètre peut varier d'un segment à l'autre, reflétant le changement éventuel de brillance de l'objet observé. Les propriétés statistiques de telles données peuvent donc être définies de la manière suivante :

$$y_{j,i} \sim \mathcal{P}(\lambda_{j,k}), \quad (1.2)$$

où $j = 1, \dots, J$, $k = 1, \dots, K_j$, $i \in I_{j,k} = \{l_{j,k-1} + 1, \dots, l_{j,k}\}$, et avec les notations suivantes :

- $\mathcal{P}(\lambda)$ désigne une loi de Poisson de paramètre λ ,
- J est le nombre de signaux à segmenter,
- K_j est le nombre de segments dans le j -ième signal observé,
- $l_{j,k}$ est l'instant après lequel le k -ième changement intervient dans le j -ième signal (par convention $l_{j,0} = 0$ et $l_{j,K_j} = n$ où n est le nombre de données observés dans chaque séquence). En d'autres termes, les instants de ruptures sont $t_{j,k} = l_{j,k}T + \tau$ avec $0 \leq \tau < T$, où T est la période d'échantillonnage.

De plus les séquences $\mathbf{y}_l = [y_{l,1}, \dots, y_{l,n}]$ et $\mathbf{y}_m = [y_{m,1}, \dots, y_{m,n}]$ sont supposées *a priori* indépendantes pour $l \neq m$. Segmenter conjointement les séries temporelles astronomiques $\mathbf{y}_j = [y_{j,1}, \dots, y_{j,n}]$ consiste à estimer le nombre de ruptures K_j et leurs positions $l_{j,k}$ (pour $j = 1, \dots, J$ et $k = 1, \dots, K_j$) à partir des observations $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_J]^T$.

1.4 Description du modèle bayésien hiérarchique

Les paramètres inconnus pour le problème de segmentation introduit précédemment sont les nombres de segments K_j , les positions des ruptures $l_{j,k}$ et les paramètres de Poisson $\lambda_{j,k}$ (avec $\boldsymbol{\lambda}_j = [\lambda_{j,1}, \dots, \lambda_{j,K_j}]^\top$ et $\boldsymbol{\Lambda} = \{\boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_J\}$). Un reparamétrage classique [TDL03; CTC03; RB06] consiste à introduire une variable indicatrice $r_{j,i}$, $j \in \{1, \dots, J\}$, $i \in \{1, \dots, n\}$ telle que :

$$\begin{cases} r_{j,i} = 1 \text{ s'il y a un changement à l'instant } i \text{ dans la séquence } j, \\ r_{j,i} = 0 \text{ sinon,} \end{cases}$$

avec $r_{j,n} = 1$ (cette condition assure que le nombre de ruptures et le nombre de segments du j -ième signal sont tous les deux égaux à $K_j = \sum_{i=1}^n r_{j,i}$). Le vecteur des paramètres inconnus issus de ce reparamétrage est $\boldsymbol{\Theta} = \{\boldsymbol{\Theta}_1, \dots, \boldsymbol{\Theta}_J\}$ où $\boldsymbol{\Theta}_j = \{\mathbf{r}_j, \boldsymbol{\lambda}_j\}$ et $\mathbf{r}_j = [r_{j,1}, \dots, r_{j,n}]$. Il est important de remarquer que le vecteur des paramètres inconnus $\boldsymbol{\Theta}$ appartient à un espace $\Theta = \{0, 1\}^{J \times n} \times \prod_{j=1}^J \mathbb{R}_+^{K_j}$ dont la dimension dépend des paramètres K_j , $j = 1, \dots, J$. Nous proposons ici d'estimer ce vecteur de paramètres $\boldsymbol{\Theta}$ dans le cadre de la théorie de l'estimation bayésienne. L'inférence bayésienne sur $\boldsymbol{\Theta}$ est basée sur la loi *a posteriori* $f(\boldsymbol{\Theta}|\mathbf{Y})$. Cette loi est liée à la fonction de vraisemblance des observations (terme d'attache aux données) et aux lois *a priori* des paramètres par la relation de Bayes $f(\boldsymbol{\Theta}|\mathbf{Y}) \propto f(\mathbf{Y}|\boldsymbol{\Theta})f(\boldsymbol{\Theta})$. Cette fonction de vraisemblance et ces lois *a priori* pour le problème de segmentation envisagé sont détaillées dans les paragraphes qui suivent.

1.4.1 Fonction de vraisemblance

Compte tenu de l'indépendance *a priori* des séquences observées d'une part, et des segments au sein de chaque signal d'autre part, la fonction de vraisemblance de la matrice des observations \mathbf{Y}

s'écrit :

$$\begin{aligned}
f(\mathbf{Y}|\Theta) &= \prod_{j=1}^J \prod_{k=1}^{K_j} \prod_{i \in I_{j,k}} \frac{\lambda_{j,k}^{y_{j,i}} \exp(-\lambda_{j,k})}{y_{j,i}!} \\
&= \frac{1}{\prod_{j=1}^J \prod_{i=1}^n y_{j,i}!} \prod_{j=1}^J \prod_{k=1}^{K_j} \lambda_{j,k}^{s_{j,k}(\mathbf{r}_j)} \exp(-\lambda_{j,k} n_{j,k}(\mathbf{r}_j)) \\
&\propto \prod_{j=1}^J \prod_{k=1}^{K_j} \lambda_{j,k}^{s_{j,k}(\mathbf{r}_j)} \exp(-\lambda_{j,k} n_{j,k}(\mathbf{r}_j)),
\end{aligned} \tag{1.3}$$

où \propto signifie « proportionnel à, » $s_{j,k}(\mathbf{r}_j) = \sum_{i \in I_{j,k}} y_{j,i}$ et $n_{j,k}(\mathbf{r}_j) = l_{j,k} - l_{j,k-1}$ (nombre de points dans le k -ième segment $I_{j,k}$ du j -ième signal).

1.4.2 Lois *a priori* des paramètres

Dans l'approche proposée, les ruptures sont repérées grâce aux fonctions indicatrices \mathbf{r}_j , $j = 1, \dots, J$. Nous rappelons au lecteur qu'une fonction indicatrice \mathbf{r}_j est définie pour chaque signal observé \mathbf{y}_j . Cette partie détaille les lois *a priori* choisies pour ces indicatrices, ainsi que pour les paramètres de Poisson $\lambda_1, \dots, \lambda_J$.

Vecteur des indicatrices

Les corrélations éventuelles entre les positions des ruptures dans les différents signaux observés sont modélisées par une loi *a priori* $f(\mathbf{R}|\mathbf{P})$ adaptée, où $\mathbf{R} = [\mathbf{r}_1, \dots, \mathbf{r}_J]^\top$ et \mathbf{P} est explicitée ci-dessous. Une réalisation de la matrice \mathbf{R} , de taille $J \times n$ et composée de 0 et de 1, est une configuration globale de changements qui peut être solution du problème de segmentation conjointe. On définit alors une configuration locale de ruptures, notée ϵ (où $\epsilon \in \mathcal{E} = \{0, 1\}^J$), comme une réalisation particulière d'une colonne de \mathbf{R} , qui traduit la présence ou l'absence de ruptures à un instant donné dans les J signaux. On note alors P_ϵ la probabilité d'avoir la configuration ϵ à l'instant i , c'est-à-dire d'avoir $\mathbf{R}_i = [r_{1,i}, \dots, r_{J,i}]^\top = \epsilon$. Nous supposons que la probabilité P_ϵ ne dépend pas de i . Par conséquent, en supposant aussi que $\mathbf{R}_i = [r_{1,i}, \dots, r_{J,i}]^\top$ est indépendant de $\mathbf{R}_{i'} = [r_{1,i'}, \dots, r_{J,i'}]^\top$ pour tout

$i \neq i'$, la loi *a priori* de la fonction indicatrice s'écrit :

$$f(\mathbf{R}|\mathbf{P}) = \prod_{\epsilon \in \mathcal{E}} P_{\epsilon}^{S_{\epsilon}(\mathbf{R})}, \quad (1.4)$$

où $\mathbf{P} = \{P_{\epsilon}\}_{\epsilon \in \mathcal{E}}$ et $S_{\epsilon}(\mathbf{R})$ est le nombre d'instant i tels que $[r_{1,i}, \dots, r_{J,i}]^{\top} = \epsilon$. Par exemple, dans le cas de deux signaux \mathbf{y}_1 et \mathbf{y}_2 (i.e., $J = 2$), la loi *a priori* de \mathbf{R} se réduit à :

$$f(\mathbf{R}|\mathbf{P}) = P_{00}^{S_{00}} P_{10}^{S_{10}} P_{01}^{S_{01}} P_{11}^{S_{11}}, \quad (1.5)$$

où $S_{00} = \sum_{i=1}^{n-1} (1 - r_{1,i})(1 - r_{2,i})$, $S_{11} = \sum_{i=1}^{n-1} r_{1,i}r_{2,i}$, $S_{10} = \sum_{i=1}^{n-1} r_{1,i}(1 - r_{2,i})$ et $S_{01} = \sum_{i=1}^{n-1} (1 - r_{1,i})r_{2,i}$. Avec cette loi *a priori*, une valeur élevée de P_{ϵ} caractérise une configuration locale de rupture $[r_{1,i}, \dots, r_{J,i}]^{\top} = \epsilon$ très probable pour tout $i = 1, \dots, n-1$. Ainsi, en choisissant de fortes valeurs pour $P_{0\dots 0}$ (respectivement $P_{1\dots 1}$), nous favoriserons l'absence (respectivement la présence) de changements simultanés dans tous les signaux. Ce choix introduit naturellement une corrélation entre les positions de ruptures parmi tous les signaux.

Remarque

Nous avons supposé dans ce chapitre l'indépendance des vecteurs \mathbf{R}_i et $\mathbf{R}_{i'}$ pour $i \neq i'$. Nous verrons au chapitre suivant qu'un modèle markovien des indicatrices, introduisant une corrélation temporelle des configurations de ruptures, peut être envisagé (voir paragraphe 2.5.2).

Paramètres de Poisson

En supposant l'indépendance des paramètres $\lambda_{j,k}$, la loi *a priori* de $\mathbf{\Lambda} = \{\boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_J\}$ est :

$$f(\mathbf{\Lambda}|\gamma) = \prod_{j=1}^J \prod_{k=1}^{K_j} f(\lambda_{j,k}|\nu, \gamma).$$

Des lois Gamma sont proposées comme lois *a priori* pour ces paramètres de Poisson :

$$\lambda_{j,k}|\nu, \gamma \sim \mathcal{G}(\nu, \gamma), \quad (1.6)$$

où $\nu = 1$ (comme dans [PADF02]), γ est un hyperparamètre que l'on peut régler, et $\mathcal{G}(a, b)$ désigne la loi Gamma de paramètres a et b . Le choix d'une telle loi peut être motivée par les deux points suivants.

Premièrement, la grande diversité des lois Gamma lorsque γ varie montre qu'il est possible de prendre en compte plus ou moins d'information *a priori* pour choisir les paramètres $\lambda_{j,k}$. Deuxièmement, la loi Gamma est la loi conjuguée des paramètres $\lambda_{j,k}$, ce qui permet d'intégrer ces paramètres de nuisance lors du calcul de la loi *a posteriori* (cf. paragraphe 1.10). Avec les hypothèses précédentes, la loi *a priori* du vecteur des paramètres des lois de Poisson $\mathbf{\Lambda}$ s'écrit¹ :

$$\begin{aligned} f(\mathbf{\Lambda}|\gamma) &= \prod_{j=1}^J \prod_{k=1}^{K_j} f(\lambda_{j,k}|\nu, \gamma) \\ &= \prod_{j=1}^J \prod_{k=1}^{K_j} \frac{\gamma^\nu}{\Gamma(\nu)} \lambda_{j,k}^{\nu-1} e^{-\gamma \lambda_{j,k}} \\ &= \prod_{j=1}^J \left[\frac{\gamma^{\nu K_j} e^{-\gamma \sum_{k=1}^{K_j} \lambda_{j,k}}}{\Gamma(\nu)^{K_j}} \prod_{k=1}^{K_j} \left(\lambda_{j,k}^{\nu-1} \mathbf{1}_{\mathbb{R}^+}(\lambda_{j,k}) \right) \right], \end{aligned}$$

où $\mathbf{1}_{\mathbb{R}^+}(\cdot)$ est la fonction indicatrice sur \mathbb{R}^+ (i.e. $\mathbf{1}_{\mathbb{R}^+}(x) = 1$ si $x \geq 0$ et $\mathbf{1}_{\mathbb{R}^+}(x) = 0$ sinon).

1.4.3 Lois *a priori* des hyperparamètres

Pourquoi un deuxième niveau de hiérarchie ?

Le vecteur des hyperparamètres associés aux lois *a priori* définies précédemment s'écrit $\mathbf{\Phi} = (\mathbf{P}, \gamma)$. Bien sûr, la qualité de la segmentation bayésienne dépend des valeurs choisies pour ces hyperparamètres. Dans des contextes bien particuliers – en l'occurrence lorsque l'on connaît la probabilité d'apparition des 2^J configurations ϵ et la « brillance » moyenne du flux astronomique observé – ces hyperparamètres peuvent être fixés par l'opérateur, comme dans [MT93]. Cependant, d'une manière beaucoup plus générale, lorsqu'aucune information *a priori* n'est disponible, ces hyperparamètres peuvent à leur tour être considérés comme des valeurs inconnues qu'il convient d'estimer. La procédure de segmentation conjointe devient alors entièrement non supervisée. Plusieurs approches peuvent être envisagées pour à la fois fournir une solution au problème de segmentation conjointe et estimer les hyperparamètres inconnus. La méthode retenue dans ce chapitre propose une description probabiliste

¹On peut noter qu'il aurait été possible de définir un jeu d'hyperparamètres $\gamma_j, j = 1, \dots, J$ pour des signaux dont les amplitudes diffèrent de manière significative d'une séquence à l'autre. Le choix d'un jeu d'hyperparamètres d'échelle sera discuté au paragraphe 2.5.1.

des hyperparamètres Φ par des lois *a priori* non informatives reflétant l'absence d'information concernant ces hyperparamètres. Cette stratégie, comportant plusieurs niveaux d'inférence bayésienne, fournit ce que l'on appelle des *modèles bayésiens hiérarchiques*. Comme suggéré plus haut, ils nécessitent la définition de lois *a priori* pour ces hyperparamètres : celles-ci sont détaillées dans le paragraphe suivant.

Hyperparamètre γ

Comme nous l'avons vu précédemment, il convient de définir pour γ une loi *a priori* qui soit non-informative. Une idée naturelle serait d'envisager une loi uniforme. Bien sûr, cette distribution est impropre lorsque le paramètre prend ses valeurs dans un domaine non borné. L'utilisation d'une telle loi peut tout de même être justifiée lorsque la loi jointe $f(\Theta, \Phi | \mathbf{Y})$ reste proprement définie [Rob92, p.31]. Cependant, il est légitime de s'interroger sur le réel caractère non informatif d'une loi uniforme qui n'est pas invariante par reparamétrisation. Par exemple, le choix d'une loi uniforme sur $[a, +\infty[$ pour un paramètre quelconque η (loi impropre) implique que η^2 est distribué suivant $f(\eta^2) \propto \frac{1}{\sqrt{\eta}}$ (loi également impropre) qui favorise les petites valeurs de η . Il convient alors de considérer un tout autre critère rendant compte de la nature non-informative d'une distribution *a priori*.

Une solution qui permet de s'affranchir de ce problème de paramétrisation consiste à utiliser comme loi *a priori* pour γ une loi non-informative dite de *Jeffrey*. Elle se résume à la racine carrée de l'information de Fisher du modèle [Rob92, p.103] :

$$f(\eta) = \sqrt{-\mathbb{E}_\eta \left[\frac{\partial^2 \log f(\mathbf{Y} | \eta)}{\partial \eta^2} \right]}$$

Couramment utilisée dans la littérature [PADF02; GR98; AD99], cette loi non-informative devient dans le cas du paramètre γ :

$$f(\gamma) = \frac{1}{\gamma} \mathbf{1}_{\mathbb{R}^+}(\gamma). \quad (1.7)$$

Hyperparamètre \mathbf{P}

La loi de Dirichlet est la distribution usuelle pour les paramètres positifs dont la somme est égale à 1. Elle a l'avantage de fournir une loi *a priori* vague ou informative selon la valeur de ses paramètres.

Dans le cas du modèle bayésien introduit, elle offre la possibilité d'intégrer les hyperparamètres P_ϵ dans la loi *a posteriori* jointe. Nous supposons donc ici que la loi *a priori* pour \mathbf{P} est une loi de Dirichlet de vecteur paramètre $\boldsymbol{\alpha} = [\alpha_{0\dots 0}, \dots, \alpha_{1\dots 1}]^\top$:

$$\mathbf{P}|\boldsymbol{\alpha} \sim \mathcal{D}_{2^J}(\boldsymbol{\alpha}). \quad (1.8)$$

Cette loi est définie sur le simplexe $\mathcal{P} = \{\mathbf{P} \mid \sum_{\epsilon \in \mathcal{E}} P_\epsilon = 1, P_\epsilon > 0\}$. Nous supposons ici que les valeurs α_ϵ sont égales. Dans ce cas, la loi de Dirichlet se ramène à une loi uniforme sur \mathcal{P} .

En supposant l'indépendance *a priori* des différents hyperparamètres, la loi *a priori* du vecteur Φ s'écrit :

$$f(\Phi|\boldsymbol{\alpha}) \propto \left(\prod_{\epsilon \in \mathcal{E}} P_\epsilon^{\alpha_\epsilon - 1} \right) \frac{1}{\gamma} \mathbf{1}_{\mathbb{R}^+}(\gamma) \mathbf{1}_{\mathcal{P}}(\mathbf{P}), \quad (1.9)$$

où $\alpha_\epsilon \in \{\alpha_{0\dots 0}, \dots, \alpha_{1\dots 1}\}$.

1.4.4 Loi *a posteriori* de Θ

La loi *a posteriori* du vecteur des paramètres inconnus $\Theta = \{\mathbf{\Lambda}, \mathbf{R}\}$ peut être obtenue grâce à la structure hiérarchique suivante :

$$f(\Theta|\mathbf{Y}) = \int f(\Theta, \Phi|\mathbf{Y}) d\Phi \propto \int f(\mathbf{Y}|\Theta) f(\Theta|\Phi) f(\Phi) d\Phi,$$

où $f(\mathbf{Y}|\Theta)$ et $f(\Phi)$ ont été définies dans (1.10) et (1.9). Cette structure hiérarchique peut être visualisée sur le graphe orienté (*directed acyclic graph* ou DAG en anglais) de la Figure 1.2. Elle permet d'intégrer les paramètres de nuisance $\mathbf{\Lambda}$ et \mathbf{P} dans la loi jointe $f(\Theta, \Phi|\mathbf{Y})$, ce qui permet d'écrire :

$$f(\mathbf{R}, \gamma|\mathbf{Y}) \propto \frac{1}{\gamma} C(\mathbf{R}|\mathbf{Y}) \prod_{j=1}^J \left[\left(\frac{\gamma^\nu}{\Gamma(\nu)} \right)^{K_j} \prod_{k=1}^{K_j} \frac{\Gamma(s_{j,k} + \nu)}{(n_{j,k} + \gamma)^{s_{j,k} + \nu}} \right] \mathbf{1}_{\mathbb{R}^+}(\gamma), \quad (1.10)$$

avec

$$C(\mathbf{R}|\mathbf{Y}) = \frac{\prod_{\epsilon \in \mathcal{E}} \Gamma(S_\epsilon(\mathbf{R}) + \alpha_\epsilon)}{\Gamma(\sum_{\epsilon \in \mathcal{E}} (S_\epsilon(\mathbf{R}) + \alpha_\epsilon))}.$$

Il est important de remarquer que l'intégration des paramètres de nuisance $\boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_J$ conduit à une loi *a posteriori* d'un vecteur paramètre $\{\mathbf{R}, \gamma\}$ appartenant à un espace de dimension fixe $\{0, 1\}^{J \times n} \times \mathbb{R}^+$. La loi *a posteriori* (1.10) a toutefois une forme trop complexe pour obtenir une

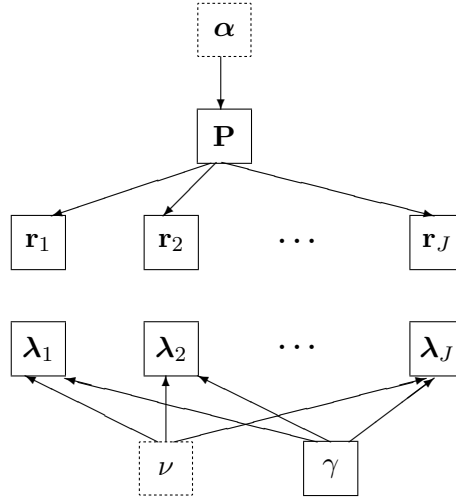


FIG. 1.2 – DAG pour les lois *a priori* (les hyperparamètres qui sont fixés apparaissent dans les carrés en pointillés).

expression simple des estimateurs bayésiens des paramètres inconnus, tels que l'estimateur qui minimise l'erreur quadratique moyenne (estimateur MMSE ou estimateur de la moyenne *a posteriori*) ou l'estimateur du Maximum *A Posteriori* (MAP) dont les expressions sont les suivantes :

$$\begin{aligned}\hat{\Theta}_{\text{MMSE}} &= \mathbb{E}[\Theta | \mathbf{Y}], \\ \hat{\Theta}_{\text{MAP}} &= \arg \max_{\Theta} f(\Theta | \mathbf{Y}).\end{aligned}\tag{1.11}$$

Dans ce cas, le recours à des techniques d'échantillonnage stochastique de type Monte Carlo par chaînes de Markov permet de dépasser les limites imposées par la complexité du modèle. Cette procédure génère des échantillons $\{\Theta^{(t)}\}_{t=1, \dots, N_{\text{MC}}}$ qui, après une période de chauffage, sont distribués suivant la loi d'intérêt. Ces échantillons peuvent alors être utilisés pour estimer les paramètres inconnus. Nous adoptons ici un échantillonneur de Gibbs similaire à celui décrit dans [PADF02]. Cependant, il est important de noter deux différences majeures. Tout d'abord, la procédure proposée ne nécessite pas le recours à un algorithme à sauts réversibles, généralement plus coûteux en temps de calcul et plus long à converger. Grâce au reparamétrage introduit au début de la partie 1.4, les paramètres inconnus intervenant dans (1.10) appartiennent en effet à un espace de dimension fixée. Ensuite, l'algorithme décrit dans ce chapitre permet de segmenter conjointement des séries temporelles issues de

plusieurs capteurs, ce qui n'était pas le cas dans [PADF02].

1.5 Échantillonneur de Gibbs pour la détection de ruptures

L'échantillonneur de Gibbs est une stratégie d'échantillonnage itérative qui consiste à générer des échantillons (notés $\cdot^{(t)}$ où t est l'indice de l'itération) distribués suivant la loi conditionnelle de chaque paramètre. Nous proposons d'échantillonner selon la loi $f(\mathbf{R}, \gamma | \mathbf{Y})$ définie par (1.10). Les principales étapes de l'algorithme 1.1 sont détaillées ci-dessous.

-
- **Initialisation :**
 - Échantillonner l'hyperparamètre $\gamma^{(0)}$ selon la loi définie dans (1.7),
 - Échantillonner l'hyperparamètre $\mathbf{P}^{(0)}$ selon la loi définie dans (1.8),
 - Pour $i = 1, \dots, n - 1$, échantillonner $\left[r_{1,i}^{(0)}, \dots, r_{J,i}^{(0)} \right]^T$ selon la loi définie dans (1.4),
 - Pour $j = 1, \dots, J$, $k = 1, \dots, K_j$, échantillonner $\lambda_{j,k}^{(0)}$ d'après les lois définies dans (1.6),
 - Poser $t \leftarrow 1$,
 - **Itérations :** Pour $t = 1, 2, 3, \dots$, faire
 - Pour $i = 1, \dots, n - 1$, tirer $\left[r_{1,i}^{(t)}, \dots, r_{J,i}^{(t)} \right]^T$ d'après les probabilités (1.12),
 - Pour $j = 1, \dots, J$, $k = 1, \dots, K_j$, échantillonner les paramètres $\lambda_{j,k}^{(t)}$ selon les lois définies dans (1.13),
 - Échantillonner le paramètre $\gamma^{(t)}$ selon les lois définies dans (1.14),
 - *Étape optionnelle :* Échantillonner l'hyperparamètre $\mathbf{P}^{(t)}$ selon la loi définie dans (1.15),
 - Poser $t \leftarrow t + 1$.
-

ALGO. 1.1: Échantillonneur de Gibbs pour la segmentation conjointe de données poissonniennes

Génération d'échantillons distribués suivant $f(\mathbf{R} | \gamma, \mathbf{Y})$

Cette génération peut elle même s'effectuer grâce à un échantillonnage de Gibbs qui permet de mettre à jour les colonnes de \mathbf{R} en $(n - 1)$ étapes selon $f(r_{1,i}, \dots, r_{J,i} | \gamma, \mathbf{Y})$:

- $\mathbf{R}_1^{(t+1)} \sim f\left(\mathbf{R}_1 | \mathbf{R}_2^{(t)}, \dots, \mathbf{R}_{n-1}^{(t)}, \gamma, \mathbf{Y}\right)$
- $\mathbf{R}_2^{(t+1)} \sim f\left(\mathbf{R}_2 | \mathbf{R}_1^{(t+1)}, \mathbf{R}_3^{(t)}, \dots, \mathbf{R}_{n-1}^{(t)}, \gamma, \mathbf{Y}\right)$
- $\mathbf{R}_3^{(t+1)} \sim f\left(\mathbf{R}_3 | \mathbf{R}_1^{(t+1)}, \mathbf{R}_2^{(t+1)}, \mathbf{R}_4^{(t)}, \dots, \mathbf{R}_{n-1}^{(t)}, \gamma, \mathbf{Y}\right)$
- ...

$$- \mathbf{R}_{n-1}^{(t+1)} \sim f\left(\mathbf{R}_{n-1}|\mathbf{R}_1^{(t+1)}, \dots, \mathbf{R}_{n-2}^{(t+1)}, \gamma, \mathbf{Y}\right)$$

La variable aléatoire $\mathbf{R}_i = [r_{1,i}, \dots, r_{J,i}]^\top$, $i = 1, \dots, n-1$, est discrète et prend ses valeurs dans $\mathcal{E} = \{0, 1\}^J$. Par conséquent, sa loi statistique est entièrement déterminée par les 2^J probabilités $P\left([r_{1,i}, \dots, r_{J,i}]^\top = \boldsymbol{\epsilon} | \gamma, \mathbf{Y}\right)$, $\boldsymbol{\epsilon} \in \mathcal{E}$. En notant \mathbf{R}_{-i} la matrice \mathbf{R} dont la i -ième colonne est supprimée, le résultat suivant peut être obtenu :

$$P\left([r_{1,i}, \dots, r_{J,i}]^\top = \boldsymbol{\epsilon} | \mathbf{R}_{-i}, \gamma, \mathbf{Y}\right) \propto f\left(\mathbf{R}_i(\boldsymbol{\epsilon}), \gamma | \mathbf{Y}\right), \quad (1.12)$$

où $\mathbf{R}_i(\boldsymbol{\epsilon})$ est la matrice \mathbf{R} dont i -ième colonne a été remplacée par le vecteur $\boldsymbol{\epsilon}$. On obtient alors, après normalisation, une expression analytique des probabilités $P\left([r_{1,i}, \dots, r_{J,i}]^\top = \boldsymbol{\epsilon} | \mathbf{R}_{-i}, \gamma, \mathbf{Y}\right)$.

Génération d'échantillons distribués suivant $f(\gamma | \mathbf{R}, \mathbf{Y})$

Échantillonner suivant $f(\gamma | \mathbf{R}, \mathbf{Y})$ ne peut se faire de manière directe puisque cette densité n'appartient pas à une famille de lois connue. En revanche, la loi jointe $f(\gamma, \boldsymbol{\Lambda} | \mathbf{R}, \mathbf{Y})$ permet d'obtenir deux lois conditionnelles $f(\gamma | \mathbf{R}, \mathbf{Y})$ et $f(\boldsymbol{\Lambda} | \mathbf{R}, \mathbf{Y})$ aisément simulables. Pour générer des échantillons distribués suivant $f(\gamma | \mathbf{R}, \mathbf{Y})$, il est alors possible d'utiliser un échantillonnage de Gibbs comme suit :

– Génération d'échantillons distribués suivant $f(\boldsymbol{\Lambda} | \mathbf{R}, \gamma, \mathbf{Y})$

En examinant la loi jointe $f(\boldsymbol{\Theta}, \boldsymbol{\Phi} | \mathbf{Y})$, on peut montrer que :

$$\lambda_{j,k} | \mathbf{R}, \gamma, \mathbf{Y} \sim \mathcal{G}(s_{j,k}(\mathbf{r}_j) + \nu, n_{j,k}(\mathbf{r}_j) + \gamma). \quad (1.13)$$

– Génération d'échantillons distribués suivant $f(\gamma | \mathbf{R}, \boldsymbol{\Lambda}, \mathbf{Y})$

Cette génération s'effectue de la manière suivante :

$$\gamma | \mathbf{R}, \boldsymbol{\Lambda}, \mathbf{Y} \sim \mathcal{G}\left(\nu \sum_{j=1}^J K_j, \sum_{j=1}^J \sum_{k=1}^{K_j} \lambda_{j,k}\right). \quad (1.14)$$

Loi *a posteriori* de l'hyperparamètre \mathbf{P}

L'hyperparamètre P_ϵ permet de connaître la probabilité d'avoir une configuration de ruptures $\boldsymbol{\epsilon}$. Par conséquent, son estimation peut être intéressante dans certaines applications pratiques. La loi *a posteriori* de cet hyperparamètre, connaissant la matrice des indicatrices \mathbf{R} , les observations \mathbf{Y} et les

paramètres fixés α , peut être facilement calculée. Il s'agit d'une loi de Dirichlet à 2^J dimensions de paramètres $(\alpha_\epsilon + S_\epsilon(\mathbf{R}))_{\epsilon \in \mathcal{E}}$:

$$\mathbf{P} | \mathbf{R}, \mathbf{Y}, \alpha \sim \mathcal{D}_{2^J}(\alpha_\epsilon + S_\epsilon(\mathbf{R})). \quad (1.15)$$

1.6 Diagnostic de convergence

Un problème capital inhérent à l'utilisation de méthodes MCMC est l'évaluation de la convergence des chaînes construites. L'échantillonneur de Gibbs permet de générer des échantillons $\{(\mathbf{R}^{(t)}, \gamma^{(t)})\}$ asymptotiquement distribués suivant $f(\mathbf{R}, \gamma | \mathbf{Y})$. Les probabilités *a posteriori* des instants de ruptures peuvent être estimées par la moyenne empirique :

$$\hat{\mathbf{R}}_{\text{MMSE}} = \frac{1}{N_r} \sum_{t=1}^{N_r} \mathbf{R}^{(N_{\text{bi}}+t)}, \quad (1.16)$$

où N_{bi} est le nombre d'itérations de chauffage et N_r est le nombre d'itérations d'intérêt. Cependant, il nous faut répondre à deux questions essentielles. Premièrement, à partir de quel moment pouvons-nous affirmer que les échantillons $\{\mathbf{R}^{(t)}\}$ sont distribués suivant la loi cible? Par ailleurs, combien d'échantillons sont nécessaires pour obtenir une estimation précise de \mathbf{R} lorsqu'on utilise (1.16)? Ce paragraphe présente quelques outils qui permettent de choisir des valeurs convenables des paramètres N_r et N_{bi} .

1.6.1 Choix de la période de chauffage N_{bi}

Construire en parallèle plusieurs chaînes de Markov initialisées de manières aléatoires permet de définir plusieurs mesures de convergence des méthodes MCMC [RR98]. Nous proposons d'utiliser ici le critère de variance inter/intra-chaîne pour assurer la convergence des chaînes construites. Cette méthode a initialement été présentée par Gelman et Rubin dans [GR92] puis a souvent été employée pour surveiller la convergence des chaînes de Markov (cf. par exemple [GR98], [DC02] ou [RR98, p. 33]). Ce critère nécessite de construire M chaînes en parallèle de longueur N_r avec différentes valeurs initiales. Les variances inter-chaîne B et intra-chaîne W pour ces M chaînes de Markov sont définies

respectivement par :

$$B = \frac{N_r}{M-1} \sum_{m=1}^M (\bar{\kappa}_m - \bar{\kappa})^2, \quad (1.17)$$

et

$$W = \frac{1}{M} \sum_{m=1}^M \frac{1}{N_r-1} \sum_{t=1}^{N_r} \left(\kappa_m^{(t)} - \bar{\kappa}_m \right)^2, \quad (1.18)$$

avec

$$\begin{cases} \bar{\kappa}_m = \frac{1}{N_r} \sum_{t=1}^{N_r} \kappa_m^{(t)}, \\ \bar{\kappa} = \frac{1}{M} \sum_{m=1}^M \bar{\kappa}_m, \end{cases} \quad (1.19)$$

où κ est le paramètre d'intérêt et $\kappa_m^{(t)}$ est le t -ième échantillon de la m -ième chaîne. La convergence des chaînes est alors mesurée par le critère $\hat{\rho}$, que nous appellerons facteur d'échelle et qui est défini dans [GCSR95, p. 332] :

$$\sqrt{\hat{\rho}} = \sqrt{\frac{1}{W} \left(\frac{N_r-1}{N_r} W + \frac{1}{N_r} B \right)}. \quad (1.20)$$

Une valeur de $\sqrt{\hat{\rho}}$ proche de 1 indique une bonne convergence de l'échantillonneur, c'est-à-dire qu'un nombre d'itérations de chauffage égal à N_{bi} est suffisant pour obtenir des échantillons $\{(\mathbf{R}^{(N_{\text{bi}}+t)}, \gamma^{(N_{\text{bi}}+t)})\}$, $t = 1, \dots, N_r$, distribués suivant la loi cible.

1.6.2 Choix du nombre d'itérations d'intérêt N_r

Une fois que le nombre d'itérations de chauffage a été choisi, il est nécessaire de déterminer le nombre d'itérations N_r requis pour obtenir un estimateur correct de \mathbf{R} en utilisant l'équation (1.16). Une approche *ad hoc* consiste à vérifier la convergence de la chaîne construite au moyen d'évaluations graphiques [RR98, p. 28]. Tout d'abord, un estimateur de référence noté $\tilde{\mathbf{R}}$ est calculé pour un grand nombre d'itérations afin d'assurer à la fois la convergence de la chaîne et la précision de l'estimation (1.16). L'erreur quadratique moyenne (EQM) entre cet estimateur de référence $\tilde{\mathbf{R}}$ et l'estimateur obtenu après $N_r = p$ itérations est ensuite calculée grâce à :

$$e_r^2(p) = \left\| \tilde{\mathbf{R}} - \frac{1}{p} \sum_{t=1}^p \mathbf{R}^{(\tilde{N}_{\text{bi}}+t)} \right\|^2. \quad (1.21)$$

Le nombre d'itérations N_r est alors fixé à une valeur p telle que l'erreur $e_r^2(p)$ soit inférieure à un seuil donné.

1.7 Segmentation de données synthétiques

Afin de valider l'algorithme de segmentation, nous avons procédé à des simulations conduites sur des données synthétiques. Celles présentées dans cette partie ont été obtenues pour $J = 2$ signaux de longueur $n = 120$ points. Les instants de ruptures pour ces deux séquences sont $\mathbf{l}_1 = (20, 50, 100, 120)$ et $\mathbf{l}_2 = (50, 120)$. Les paramètres des lois de Poisson pour les différents segments sont $\boldsymbol{\lambda}_1 = [19, 9, 16, 6]^\top$ et $\boldsymbol{\lambda}_2 = [8, 11]^\top$. Les hyperparamètres ont été fixés à $\nu = 2$ et $\alpha_\epsilon = \alpha = 1$ ($\forall \epsilon \in \mathcal{E}$).

Les hyperparamètres α_ϵ sont choisis tous égaux afin que la loi de Dirichlet se réduise à une loi uniforme sur le simplexe \mathcal{E} . De plus, la valeur choisie pour les hyperparamètres α_ϵ est $\alpha = 1 \ll n$ dans le but de réduire l'influence de cette constante dans la loi *a posteriori* de \mathbf{P} donnée par (1.15).

Toutes les figures ont été obtenues après avoir moyenné les résultats de 64 chaînes de Markov construites en parallèles. Le nombre total d'itérations pour chaque chaîne de Markov est $N_{MC} = 1000$, dont $N_{bi} = 200$ itérations de chauffage. Ainsi, seuls les 800 derniers échantillons issus de la chaîne sont utilisés pour réaliser les estimations. Les résultats fournis par la segmentation conjointe sont comparés à ceux de deux segmentations unidimensionnelles, obtenues en appliquant l'algorithme proposé séparément sur chacune des deux séquences. Notons que l'exécution de 100 itérations de la procédure conjointe prend environ 30 secondes avec une implantation MATLAB sur un PC Pentium IV cadencé à 2.8Ghz.

1.7.1 Lois *a posteriori* des positions de ruptures

Les estimateurs MMSE des instants de ruptures sont représentés sur la figure 1.3 pour des approches unidimensionnelle et conjointe. Ces estimateurs ont été calculés grâce à la moyenne empirique

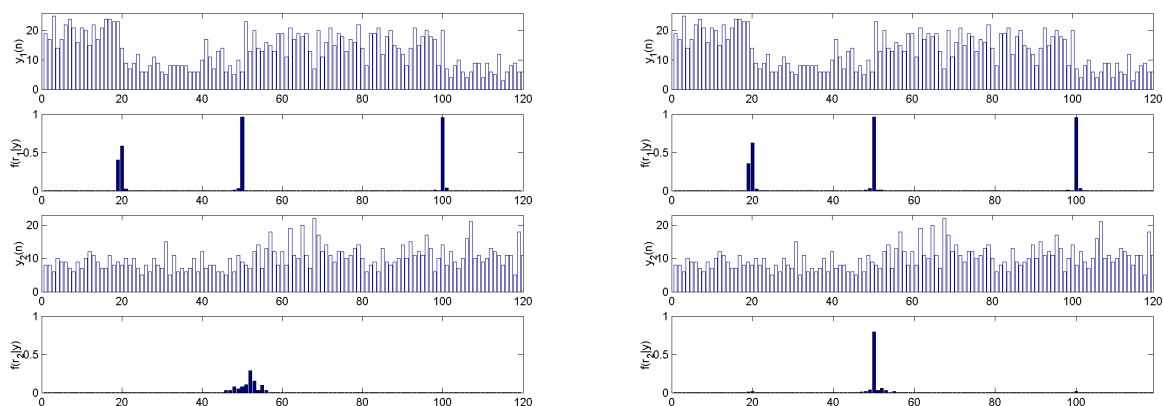


FIG. 1.3 – Probabilités *a posteriori* des instants de ruptures pour une segmentation 1D (à gauche) et conjointe (à droite).

(1.16). Il est important de remarquer que la moyenne de la matrice des indicatrices \mathbf{R} distribuées suivant $f(\mathbf{R}|\mathbf{Y})$ fournit les probabilités *a posteriori* d'avoir les ruptures aux différents instants². En effet, $r_{j,i}$ est une variable aléatoire binaire. Par exemple, il y a une très forte probabilité *a posteriori* pour que se produise un changement aux instants $i = 50$ et $i = 100$ dans la première séquence (selon les deux types de segmentation, unidimensionnelle et conjointe). Cependant, les deux méthodes semblent hésiter quant à la localisation de la première rupture (à l'instant $i = 20$) dans le premier signal. L'avantage majeur de la segmentation conjointe apparaît sur la figure 1.3 : le changement à l'instant $i = 50$ dans la deuxième séquence est beaucoup mieux estimé par la segmentation conjointe (tracé de droite) que par une segmentation unidimensionnelle (tracé de gauche). La présence d'une rupture au même instant dans la première séquence a en effet favorisé la détection d'une rupture dans la seconde. Ainsi, la segmentation conjointe donne-t-elle de bien meilleurs résultats que deux procédures indépendantes de segmentation unidimensionnelle sur cet exemple.

²Dans toute la suite du document, tout en étant conscient de la maladresse de langage, nous parlerons de lois *a posteriori* des instants (ou positions) de ruptures. Il s'agit pourtant bien des probabilités *a posteriori* des configurations globales de ruptures.

1.7.2 Lois *a posteriori* des nombres de ruptures (K_1, K_2)

Notre algorithme génère des échantillons $(\mathbf{R}^{(t)}, \gamma^{(t)})$ distribués selon la loi *a posteriori* $f(\mathbf{R}, \gamma | \mathbf{Y})$, ce qui permet de réaliser une sélection de modèle. Le nombre de ruptures dans chaque séquence peut en effet être estimé par $K_j^{(t)} = \sum_{i=1}^n r_{j,i}^{(t)}$. La figure 1.4 présente les lois *a posteriori* estimées de K_j dans chaque séquence (calculées à partir des 800 derniers états de la chaînes de Markov) pour des segmentations unidimensionnelle (figure de gauche) et conjointe (figure de droite). Les valeurs maximales de ces lois *a posteriori* permettent d'obtenir les estimateurs MAP des nombres de ruptures $\hat{K}_1 = 4$ et $\hat{K}_2 = 2$, qui correspondent aux nombres réels de changements dans les deux séquences (nous rappelons ici qu'il y a, par convention, une rupture à l'instant $i = n = 120$ dans toutes les séquences).

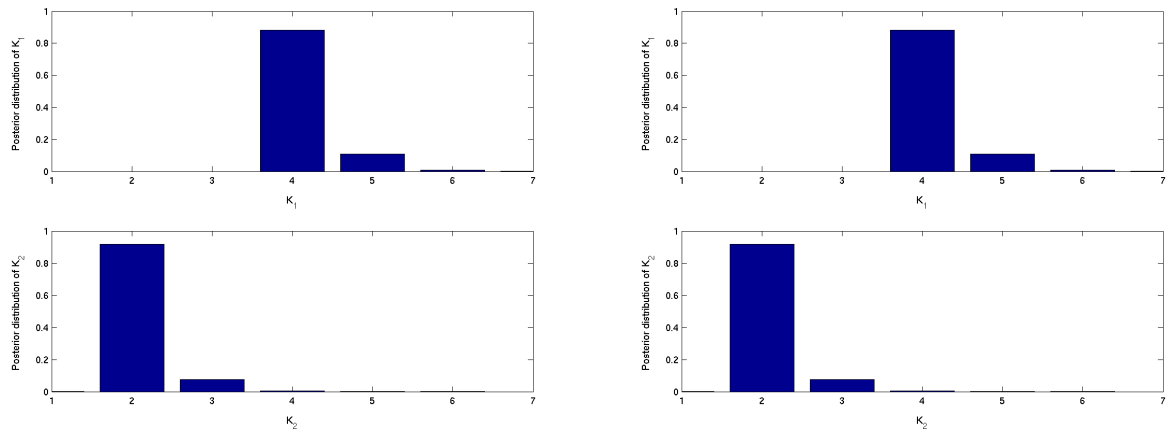
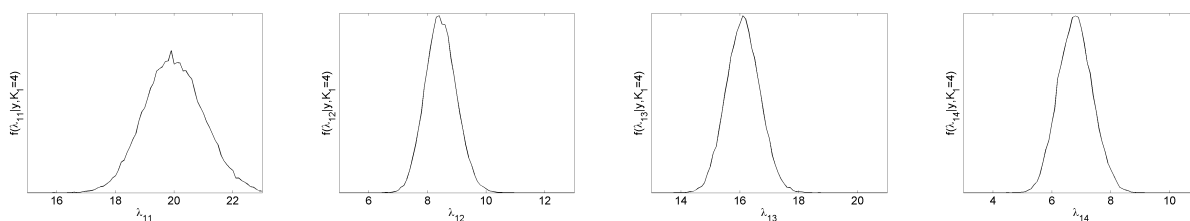
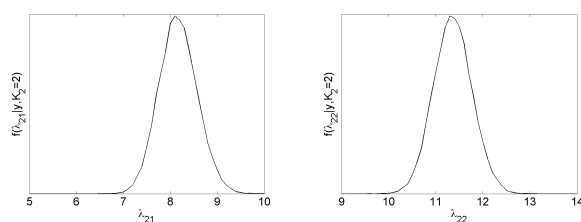


FIG. 1.4 – Lois *a posteriori* de K_1 et K_2 pour des segmentation unidimensionnelle (à gauche) et conjointe (à droite).

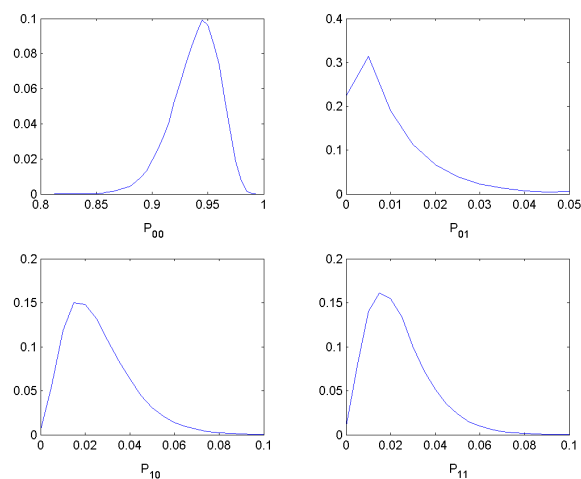
1.7.3 Estimation des paramètres des lois de Poisson

L'estimation des paramètres des lois de Poisson est intéressante notamment puisqu'elle rend possible la reconstruction des signaux. Les lois *a posteriori* des paramètres $\lambda_{1,k}$ et $\lambda_{2,k}$ sachant respectivement $K_1 = 4$ et $K_2 = 2$ sont représentées sur les figures 1.5 et 1.6. Elles sont clairement en accord avec les valeurs réelles des paramètres $\boldsymbol{\lambda}_1 = [19, 9, 16, 6]^\top$ et $\boldsymbol{\lambda}_2 = [8, 11]^\top$.

FIG. 1.5 – Lois *a posteriori* des paramètres de Poisson $\lambda_{1,k}$ (pour $k = 1, \dots, 4$) sachant $K_1 = 4$.FIG. 1.6 – Lois *a posteriori* des paramètres de Poisson $\lambda_{2,k}$ (pour $k = 1, 2$) sachant $K_2 = 2$.

1.7.4 Estimation de l'hyperparamètre \mathbf{P}

Les derniers résultats permettent d'illustrer la performance de la méthode pour l'estimation du vecteur d'hyperparamètre \mathbf{P} . Les lois *a posteriori* de P_ϵ ($\epsilon \in \mathcal{E}$) sont tracées sur la figure 1.7. Elles sont clairement en accord avec les lois *a posteriori* données par les lois de Dirichlet définies par (1.15).

FIG. 1.7 – Lois *a posteriori* des hyperparamètres P_{00}, P_{01}, P_{10} et P_{11} .

1.7.5 Convergence de l'échantillonneur

Comme il l'a été indiqué au paragraphe 1.6, nous nous assurons que l'algorithme de Gibbs génère des échantillons distribués suivant la loi cible en mesurant le facteur d'échelle défini par (1.20). Plusieurs choix pour le paramètre κ peuvent être considérés pour évaluer la convergence de la procédure de segmentation proposée. Nous proposons ici d'évaluer la convergence de l'échantillonneur de Gibbs grâce aux paramètres P_ϵ , $\epsilon \in \mathcal{E}$. Par exemple, les états de $M = 5$ chaînes de Markov pour le paramètre P_{00} sont représentés sur la figure 1.8. Les chaînes convergent clairement vers des valeurs identiques proches de 1. Elles confirment une probabilité élevée d'avoir $[r_{1,i}, r_{2,i}] = [0, 0]$ pour les signaux analysés dans ce paragraphe. Les critères de Gelman & Rubin $\sqrt{\hat{\rho}}$ pour tous les paramètres P_ϵ sont donnés dans le tableau 1.1. Ces valeurs de $\sqrt{\hat{\rho}}$ confirment la bonne convergence de l'échantillonneur puisqu'une valeur inférieure à 1.2 est conseillée pour assurer la convergence [GCSR95, p. 332].

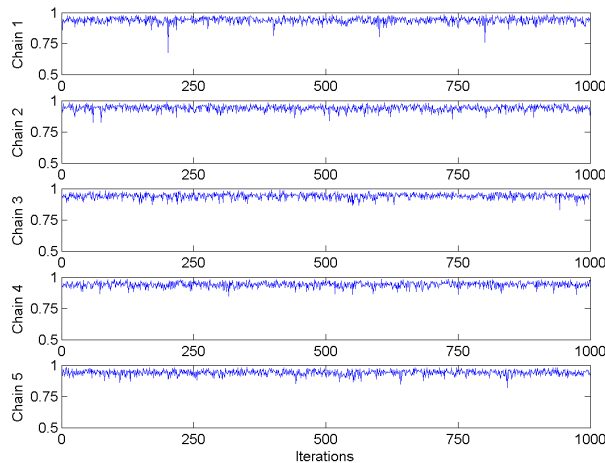


FIG. 1.8 – Suivi de la convergence de cinq chaînes de Markov.

TAB. 1.1 – Facteurs d'échelle relatifs aux paramètres P_ϵ (calculés à partir de $M = 64$ chaînes de Markov)

P_ϵ	P_{00}	P_{01}	P_{10}	P_{11}
$\sqrt{\hat{\rho}}$	0.9995	1.0000	0.9998	1.0001

Le nombre d'itérations N_r nécessaires pour calculer un estimateur efficace de \mathbf{R} selon (1.16) est fixé en évaluant l'erreur quadratique entre une estimation de référence $\tilde{\mathbf{R}}$ (obtenu avec $N_r = 10000$) et une estimation après $N_r = p$ itérations selon l'équation 1.21. La figure 1.9 montre alors qu'un nombre d'itérations d'intérêt égal à $N_r = 800$ est suffisant pour assurer une bonne estimation de la moyenne empirique (1.16).

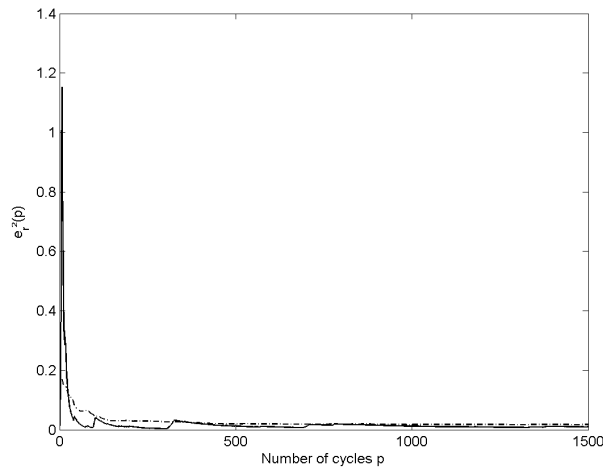


FIG. 1.9 – Erreur quadratique moyenne entre l'estimateur de référence et l'estimateur obtenu après p itérations (trait plein). EQMs moyennées à partir de 64 chaînes (tirets).

1.7.6 Comparaison avec un modèle bayésien non hiérarchique

Pour justifier le deuxième niveau de hiérarchie introduit dans l'inférence bayésienne, nous présentons les résultats de segmentation obtenus sur les signaux synthétiques précédents lorsque les hyperparamètres prennent des valeurs fixées. En ce qui concerne l'hyperparamètre \mathbf{P} , afin de traduire l'absence de connaissance *a priori* de la probabilité d'occurrence des différentes combinaisons $[0, 0]$, $[0, 1]$, $[1, 0]$ et $[1, 1]$, ces configurations locales de ruptures sont supposées équiprobables : $P_{00} = P_{01} = P_{10} = P_{11} = \frac{1}{4}$. Ensuite nous présentons sur la figure 1.10 les résultats de simulations pour deux valeurs différentes de l'hyperparamètre γ . Ces figures montrent bien que le résultat de segmentation est fortement lié aux valeurs de cet hyperparamètre. Notons que pour les deux jeux de

paramètres présentés, les résultats ne sont pas du tout satisfaisants.

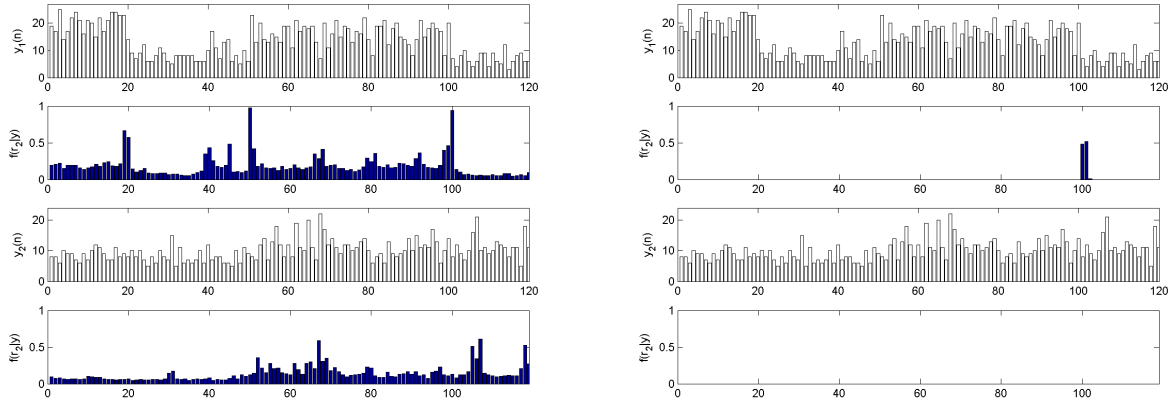


FIG. 1.10 – Probabilités *a posteriori* des instants de ruptures pour deux valeurs de l'hyperparamètre γ : $\gamma = 0.05$ (à gauche) et $\gamma = 5$ (à droite).

1.8 Analyse de données astronomiques réelles

1.8.1 Données unidimensionnelles

Description des enregistrements

Afin d'évaluer la capacité de notre algorithme à traiter des données réelles, cette partie présente l'analyse d'un jeu de données recueillies par l'instrument BATSE de la NASA [PMP+99]. Par la nature même des mesures effectuées (un comptage de photons), les séries temporelles peuvent être modélisées précisément par des processus de Poisson. Le paramètre de ces processus de Poisson varie à mesure que change l'intensité des sursauts gamma (*gamma-ray burst* ou GRB en anglais) émis par la source observée.

L'intensité du flux GRB en fonction du temps se comporte comme une série de brèves structures élémentaires, appelées impulsions. Le but de l'analyse est alors de déterminer les temps de début et de fin de ces impulsions, ainsi que les quantités caractéristiques de la représentation choisie, à savoir les paramètres des lois de Poisson supposés constants par morceaux.

Résultats

L'algorithme présenté dans les sections précédentes a tout d'abord été appliqué aux données astronomiques unidimensionnelles étudiées dans [Sca98]. Les données brutes (les temps d'arrivée d'environ 29000 photons) ont été converties en comptant le nombre de particules reçues par intervalle de temps de $3.68ms$ (soit 256 intervalles). Dans cet exemple, $J = 1$. Les résultats obtenus ont été moyennés sur 64 chaînes de Markov de longueur totale $N_{MC} = 1550$ incluant $N_{bi} = 50$ itérations de chauffage. Les valeurs de N_{MC} et de N_{bi} ont été choisies dans le but d'obtenir des valeurs correctes du critère de convergence de Gelman & Rubin pour les paramètres P_0 et P_1 (cf. la fin de ce paragraphe).

La première étape de l'analyse consiste à estimer la loi *a posteriori* des indicatrices \mathbf{R} pour le signal traité qui peut être visualisé sur la figure 1.11 (en haut). Cette loi estimée est représentée sur la figure 1.11 (en bas).

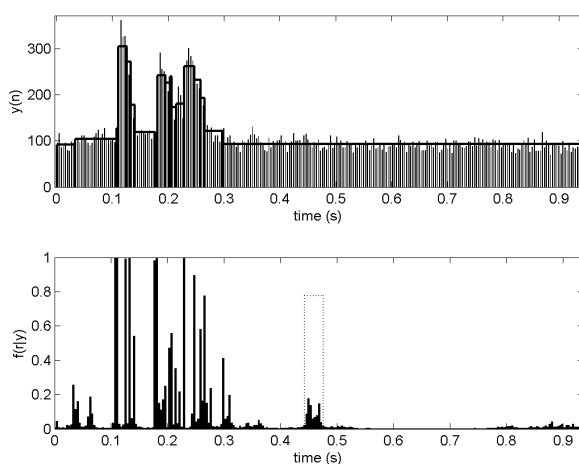


FIG. 1.11 – En haut : données mises en forme et segmentées. En bas : loi *a posteriori* des instants de ruptures (données astronomiques unidimensionnelles).

Dans un deuxième temps, le nombre de ruptures présentes dans le signal analysé est estimé. La loi *a posteriori* du nombre de changements, calculé à partir des 1500 derniers échantillons de la chaîne de Markov comme au paragraphe 1.7.2, est représentée sur la figure 1.12. L'estimateur MAP correspondant est alors $\hat{K} = 18$.

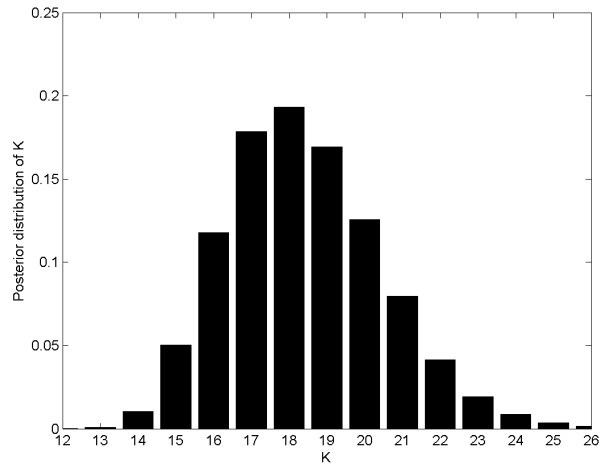


FIG. 1.12 – Loi *a posteriori* du nombre de ruptures (données astronomiques unidimensionnelles).

Enfin, les différents paramètres des lois de Poisson caractéristiques de la brillance de la source sont estimés sur chacun des segments repérés par les instants de rupture estimés. Plus précisément, les segments sont obtenus en conservant les $\hat{K} = 18$ plus grandes valeurs de la loi *a posteriori* représentée sur la figure 1.11 (en bas). Les paramètres des lois de Poisson sont alors approchés en moyennant le signal sur chaque segment. Cette procédure permet d’obtenir la représentation par blocs introduite dans [Sca98]. La figure 1.11 (en haut) représente les blocs obtenus après avoir gardé $\hat{K} = 18$ segments. Il est important de remarquer que la méthode proposée rend possible le calcul de la probabilité d’avoir une rupture dans un intervalle de temps donnée. Par exemple, la probabilité *a posteriori* d’avoir au moins un changement dans l’intervalle $[0.44; 0.47]$ apparaît en pointillés sur la figure 1.11 (en bas). La valeur élevée de cette probabilité permet d’introduire un nouveau modèle de segmentation qui admettrait une nouvelle rupture dans l’intervalle $[0.44; 0.47]$. Ces résultats sont en accord avec ceux présentés dans [Sca98].

La convergence de l’échantillonneur de Gibbs utilisé pour la segmentation des données de la figure 1.11 (en haut) a été examinée grâce au critère de Gelman & Rubin pour les paramètres P_0 et P_1 , obtenu à partir de 64 chaînes construites en parallèles, chacune de longueur $N_r = 1500$. Ils sont tous les deux égaux à $\sqrt{\hat{\rho}} = 0.9996$. La règle de convergence imposant $\sqrt{\hat{\rho}} < 1.2$ est donc satisfaite pour

cet exemple. Nous rappelons ici que notre algorithme, contrairement à celui présenté dans [Sca98], ne nécessite pas de règle d'arrêt. Il nous suffit juste de vérifier la bonne convergence des chaînes de Markov construites.

1.8.2 Données multidimensionnelles

Motivations

Un point majeur dans l'étude des *Gamma Ray Burst* est de savoir comment la variabilité du flux astronomique (variabilité observée et analysée notamment au paragraphe précédent) dépend de l'énergie du rayonnement. Dans le mode d'enregistrement qui a permis de recueillir les données précédemment étudiées, les photons sont captés par BATSE dans quatre bandes énergétiques : $25 - 60\text{keV}$, $60 - 110\text{keV}$, $110 - 325\text{keV}$, et $> 325\text{keV}$. L'évolution du rayonnement est, sensiblement, la même pour des flux respectivement de basses et de plus hautes énergies. Nous nous proposons, à l'aide de la procédure de segmentation conjointe introduite, de mettre en évidence ces similitudes de comportement dans les quatre bandes d'énergie considérées. Le traitement multi-varié que nous décrivons est une étape capitale dans l'analyse des flux gamma, analyse qui est encore, par ailleurs, dans sa phase exploratoire.

Résultats

Les données observées dans les quatre canaux ont donc été traitées par l'algorithme de segmentation conjointe proposé. Le nombre de ruptures et leurs positions sont obtenus après 3500 itérations incluant une période de chauffage de 200 itérations. La figure 1.13 montre les histogrammes des paramètres K_j . Leurs valeurs maximales, correspondant aux estimateurs MAP des nombres de ruptures sont $\widehat{K}_1 = 5$, $\widehat{K}_2 = 7$, $\widehat{K}_3 = 11$ et $\widehat{K}_4 = 8$. La loi *a posteriori* de \mathbf{R} représentée sur la figure 1.14 permet ensuite d'estimer la position des ruptures dans chaque canal (comme expliqué dans la section précédente). La représentation par blocs est alors illustrée sur la figure 1.15. Notez que les échelles de temps ne sont pas les mêmes pour les figures 1.14 et 1.15 par soucis de clarté (i.e. $[0, 0.33\text{s}]$ pour la figure 1.14 et $[0, 0.94\text{s}]$ pour la figure 1.15). Ces résultats sont en accord avec ceux présentés dans [SNJ06]. Cependant, l'approche conjointe permet de mettre en évidence des ruptures qui n'étaient pas

initialement détectées par la méthode itérative. Par exemple, les deuxième et troisième changements $l_{1,2}$ et $l_{1,3}$ dans le premier canal (respectivement à $0.1294s$ et $0.2316s$) sont détectés par l'approche conjointe mais ne le sont pas par l'approche 1D. La présence de ruptures au même instant dans d'autres canaux peut expliquer cette détection.

La convergence de l'échantillonneur de Gibbs pour la segmentation conjointe des données astronomiques réelles multidimensionnelles de la figure 1.14 a été étudiée. Les critères de Gelman & Rubin pour les paramètres P_{ϵ} , $\epsilon \in \{0,1\}^4$ sont rassemblés dans le tableau 1.2. La règle de convergence $\sqrt{\hat{\rho}} < 1.2$ est à nouveau satisfaite pour cet exemple.

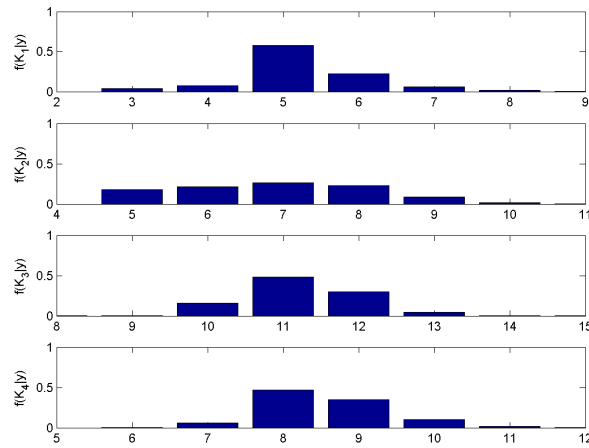


FIG. 1.13 – Lois *a posteriori* des nombres de ruptures (données astronomiques 4D).

1.9 Conclusions

Ce chapitre propose un modèle bayésien hiérarchique pour la segmentation de séries temporelles distribuées suivant des lois de Poisson dont les paramètres sont constants par morceaux. Nous avons vu qu'introduire un deuxième niveau de hiérarchie dans l'inférence bayésienne permet de s'affranchir du problème d'estimation des hyperparamètres. Des échantillons distribués suivant les lois *a posteriori* des paramètres inconnus sont générés grâce à un échantillonneur de Gibbs. Ils sont ensuite utilisés pour réaliser les estimations des paramètres d'intérêt.

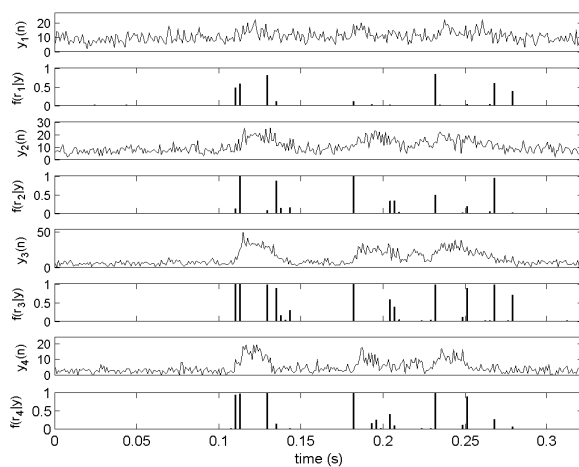


FIG. 1.14 – Probabilités *a posteriori* des instants de ruptures (données astronomiques 4D).

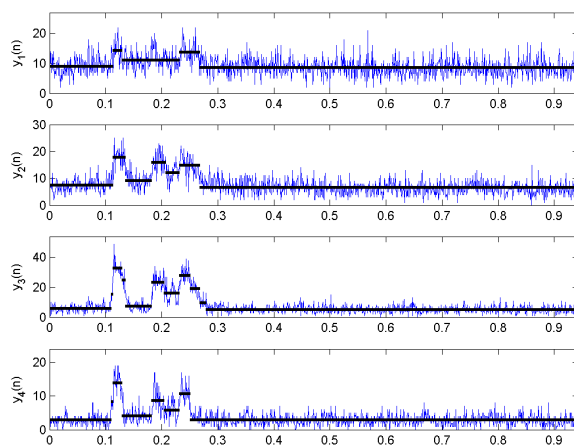


FIG. 1.15 – Représentation par blocs (données astronomiques 4D).

Un des aspects importants de ce travail est la possibilité de traiter conjointement plusieurs signaux et d'exploiter les possibles relations entre ces séries. Le choix d'une loi *a priori* adéquate sur les instants de ruptures a permis en effet de dépasser largement les performances d'un algorithme de segmentation qui serait appliqué sur chaque signal indépendamment des autres.

TAB. 1.2 – Facteurs d’échelle relatifs aux paramètres P_ϵ (calculés à partir de $M = 64$ chaînes de Markov)

P_ϵ	$\sqrt{\hat{\rho}}$	P_ϵ	$\sqrt{\hat{\rho}}$
P_{0000}	0.9999	P_{0100}	1.0008
P_{0001}	1.0004	P_{0101}	1.0015
P_{0010}	0.9999	P_{0110}	1.0002
P_{0011}	1.0000	P_{0111}	1.0013
P_{1000}	1.0013	P_{1100}	1.0021
P_{1001}	1.0000	P_{1101}	1.0000
P_{1010}	0.9999	P_{1110}	0.9999
P_{1011}	0.9998	P_{1111}	0.9999

Un autre point important du travail est que la stratégie d’échantillonnage stochastique développée dans ce chapitre permet d’explorer l’espace des paramètres d’intérêt. Les algorithmes de segmentation de données poissonniennes développés dans la littérature (comme dans [Sca98] ou [JSB⁺05]) ne permettent pas une telle exploration et sont donc dans l’incapacité de fournir des mesures de confiance des estimations réalisées.

CHAPITRE 2

Segmentation conjointe de processus autorégressifs

Sommaire

2.1	Introduction	37
2.2	Modèle bayésien hiérarchique	39
2.3	Échantillonneur de Gibbs pour la segmentation conjointe de processus AR	45
2.4	Segmentation de données synthétiques	48
2.5	Modèles alternatifs	57
2.6	Ordres des modèles inconnus	60
2.7	Applications	65
2.8	Conclusions	71

2.1 Introduction

Ce chapitre est consacré à la segmentation de processus autorégressifs multiples. Comme dans le chapitre précédent, nous considérons que J capteurs fournissent J signaux (appelées aussi *observations*), de longueur n . Ces signaux sont notés vectoriellement par $\mathbf{y}_j = [y_{j,1}, \dots, y_{j,n}]$ pour $j = 1, \dots, J$, où $y_{j,i}$ est la valeur du signal j à l'instant i . Chacun de ces signaux est modélisé comme un processus AR dont les paramètres sont constants par morceaux :

$$y_{j,i} = \sum_{l=1}^p a_{j,k,l} y_{j,i-l} + e_{j,i}, \quad (2.1)$$

où $k = 1, \dots, K_j$ est l'indice de la portion de signal (appelé *segment*) \mathbf{y}_j où le processus AR est stationnaire. Le k -ième segment dans le j -ième signal possède des frontières notées $[l_{j,k-1} + 1, l_{j,k}]$ où $l_{j,k}$ est l'indice des temps après lequel intervient une rupture, avec par convention, $l_{j,0} = 0$ et

$l_{j,K_j} = n$. Dans chacun de ces K_j segments, pour le j -ième signal, l'ensemble des paramètres AR est noté vectoriellement $\mathbf{a}_{j,k} = [a_{j,k,1}, \dots, a_{j,k,p}]^\top$. Les pôles de ces processus AR sont supposés à l'intérieur du cercle unité afin d'assurer leur stationnarité et leur causalité sur chaque segment. Enfin, $\mathbf{e}_j = [e_{j,1}, \dots, e_{j,n}]$ est une séquence de n réalisations indépendantes et identiquement distribuées d'un bruit gaussien centré de variance $\sigma_{j,k}^2$ sur le k -ième segment. Les vecteurs bruits $\mathbf{e}_1, \dots, \mathbf{e}_J$ sont supposés *a priori* indépendants.

La modélisation des observations à l'aide des processus AR est motivée par le fait que pour tout signal de densité spectrale $S(f)$, on peut trouver un processus AR dont la densité spectrale est aussi proche qu'on le veut de $S(f)$ [BD90, p. 130]. Un grand nombre d'auteurs ont suivi cette approche pour mettre au point des algorithmes de détection de ruptures, comme par exemple dans [And88] ou [Lav98]. Afin de simplifier l'énoncé du problème, nous supposerons, dans un premier temps de l'analyse, que les ordres des modèles AR dans (2.1) sont connus et tous égaux à p . Un modèle plus général, pour lequel les ordres de processus sont inconnus et différents d'un segment à l'autre, sera envisagé ultérieurement.

En utilisant la notation standard $\mathbf{x}_{j,i:i'} = [x_{j,i}, \dots, x_{j,i'}]$, l'ensemble des équations (2.1) peut être ré-écrit de manière matricielle :

$$\mathbf{y}_{j,l_{j,k-1}+1:l_{j,k}}^\top = \mathbf{Y}_{j,k} \mathbf{a}_{j,k} + \mathbf{e}_{j,l_{j,k-1}+1:l_{j,k}}^\top, \quad (2.2)$$

où $\mathbf{Y}_{j,k}$ est une matrice de taille $(l_{j,k} - l_{j,k-1}) \times p$:

$$\mathbf{Y}_{j,k} = \begin{bmatrix} y_{j,l_{j,k-1}} & y_{j,l_{j,k-1}-1} & \cdots & y_{j,l_{j,k-1}-p+1} \\ y_{j,l_{j,k-1}+1} & y_{j,l_{j,k-1}} & \cdots & y_{j,l_{j,k-1}-p+2} \\ \vdots & \vdots & \vdots & \vdots \\ y_{j,l_{j,k}-1} & y_{j,l_{j,k}-2} & \cdots & y_{j,l_{j,k}-p} \end{bmatrix}. \quad (2.3)$$

Comme dans le chapitre précédent, nous proposons une formulation bayésienne hiérarchique ainsi qu'un algorithme efficace pour estimer les positions des ruptures $l_{j,k}$ à partir des J séquences observées \mathbf{y}_j , $j = 1, \dots, J$.

Ce chapitre est organisé comme suit. Le paragraphe 2.2 décrit le modèle bayésien utilisé pour

réaliser la détection de rupture dans les signaux auto-régressifs. Classiquement, la méthodologie proposée consiste en un choix approprié de lois *a priori* pour les paramètres et hyperparamètres inconnus. Un échantillonneur de Gibbs, décrit dans le paragraphe 2.3, permet de générer des échantillons distribués suivant la loi *a posteriori* d'intérêt. Des simulations conduites sur des signaux synthétiques, rapportées dans le paragraphe 2.4, permettent d'illustrer la performance de l'algorithme proposé. Des modèles alternatifs sont envisagés dans le paragraphe 2.5. Dans le paragraphe 2.6, l'algorithme initial est généralisé au cas plus complexe où les ordres des modèles AR sont inconnus et peuvent varier d'un segment à l'autre. Deux applications de la segmentation conjointe de signaux AR sont étudiées au paragraphe 2.7 : la segmentation de signaux de paroles stéréo et la détection d'arc-tracking dans des signaux aéronautiques.

2.2 Modèle bayésien hiérarchique

Le problème de détection conjointe de ruptures introduit précédemment repose sur l'estimation des paramètres inconnus K_j (nombres de segments), $l_{j,k}$ (positions des ruptures), $\sigma_{j,k}^2$ (variances des bruits, avec $\boldsymbol{\sigma}_j^2 = [\sigma_{j,1}^2, \dots, \sigma_{j,K_j}^2]^\top$) et $\mathbf{a}_{j,k}$ (vecteurs des paramètres AR notés $\mathbf{A}_j = \{\mathbf{a}_{j,1}, \dots, \mathbf{a}_{j,K_j}\}$ pour le j -ième signal). Le reparamétrage présenté au paragraphe 1.4 permet d'introduire des variables indicatrices $r_{j,i}$ ($j \in \{1, \dots, J\}$, $i \in \{1, \dots, n\}$) telles que :

$$\begin{cases} r_{j,i} = 1 \text{ s'il y a un changement à l'instant } i \text{ dans la séquence } j, \\ r_{j,i} = 0 \text{ sinon,} \end{cases}$$

avec $r_{j,n} = 1$ (nous rappelons que cette condition assure que le nombre de rupture est égal au nombre de segments dans chaque signal, c'est-à-dire $K_j = \sum_{i=1}^n r_{j,i}$). En utilisant ces variables indicatrices, le vecteur des paramètres inconnus est $\Theta = \{\Theta_1, \dots, \Theta_J\}$, où $\Theta_j = (\mathbf{r}_j, \boldsymbol{\sigma}_j^2, \mathbf{A}_j)$ et $\mathbf{r}_j = [r_{j,1}, \dots, r_{j,n}]$. Comme dans le cas du modèle hiérarchique décrit au chapitre 1, il est important de remarquer que ce vecteur Θ appartient à un espace dont la dimension dépend de K_j , i.e., $\Theta \in \Theta = \{0, 1\}^{nJ} \times \prod_{j=1}^J (\mathbb{R}^+ \times \mathbb{R}^p)^{K_j}$. Nous proposons dans ce qui suit une approche bayésienne hiérarchique afin d'estimer ce vecteur inconnu Θ . L'inférence bayésienne sur Θ repose sur la loi *a*

posteriori $f(\Theta|\mathbf{Y})$, avec $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_J]^\top$, qui est reliée à la fonction de vraisemblance des observations et aux lois *a priori* des paramètres grâce à la règle de Bayes $f(\Theta|\mathbf{Y}) \propto f(\mathbf{Y}|\Theta)f(\Theta)$. La fonction de vraisemblance et les lois *a priori* utilisées dans ce problème de détection de ruptures sont donc présentées ci-dessous.

2.2.1 Fonction de vraisemblance approchée

Il est très délicat d'exprimer la vraisemblance associée à un modèle AR stationnaire par morceaux. Chaque segment stationnaire doit être en effet initialisé par les réalisations du segment précédent. Dans beaucoup de travaux, la dépendance de la vraisemblance exacte $f(\mathbf{y}_j|\Theta_j)$ aux p premières réalisations $\mathbf{y}_{j,1:p}$ est négligée (voir [Kay88, p. 186]). Nous avons décidé dans l'approche proposée d'utiliser cette approximation. En d'autres termes, en exploitant l'hypothèse d'indépendance entre les vecteurs bruits \mathbf{e}_j , $j \in \{1, \dots, J\}$, la vraisemblance exacte des observations \mathbf{Y} est approchée par l'expression :

$$\begin{aligned} f(\mathbf{Y}|\Theta) &\approx \prod_{j=1}^J f(\mathbf{y}_{j,p+1:n}|\mathbf{y}_{j,1:p}, \Theta_j) \\ &\approx \prod_{j=1}^J \prod_{k=1}^{K_j} \frac{1}{(2\pi\sigma_{j,k}^2)^{n_{j,k}(\mathbf{r}_j)/2}} \exp\left(-\frac{E_{j,k}(\mathbf{r}_j)}{2\sigma_{j,k}^2}\right), \end{aligned} \quad (2.4)$$

où $n_{j,k}(\mathbf{r}_j) = l_{j,k} - l_{j,k-1}$ est la longueur du k -ième segment dans le j -ième signal et

$$E_{j,k}(\mathbf{r}_j) \triangleq \left\| \mathbf{y}_{j,l_{j,k-1}+1:l_{j,k}}^\top - \mathbf{Y}_{j,k} \mathbf{a}_{j,k} \right\|^2, \quad (2.5)$$

avec $\|\mathbf{x}\|^2 = \mathbf{x}^\top \mathbf{x}$.

2.2.2 Lois *a priori* des paramètres

Les ruptures dans les J signaux sont détectées grâce à des variables indicatrices \mathbf{r}_j , $j = 1, \dots, J$ (une variable $r_{j,i}$ est en effet affectée à chaque signal j et à chaque instant $i = 1, \dots, n$). Ce paragraphe introduit les différentes lois *a priori* choisies pour ces variables indicatrices, ainsi que pour les variances des bruits et les paramètres AR.

Vecteurs des indicatrices

Les corrélations possibles entre les instants de ruptures dans les J signaux observés sont modélisées par une loi *a priori* appropriée $f(\mathbf{R}|\mathbf{P})$, où $\mathbf{R} = [\mathbf{r}_1, \dots, \mathbf{r}_J]^\top$ et \mathbf{P} est définie ci-dessous. Comme dans le chapitre précédent, nous définissons une *configuration globale de ruptures* comme suit : la matrice \mathbf{R} est constituée de 0 et de 1, et une configuration globale est une réalisation particulière de cette matrice $J \times n$. Avec notre formulation, celle-ci correspond à une solution particulière au problème de segmentation conjointe envisagé. Une *configuration locale de rupture*, notée ϵ (où $\epsilon \in \mathcal{E} = \{0, 1\}^J$), est une valeur particulière d'une colonne de la matrice \mathbf{R} : elle correspond à la présence (ou à l'absence) des ruptures à un instant donné, dans les J signaux.

Notons P_ϵ la probabilité d'avoir une configuration de rupture ϵ à l'instant i , c'est-à-dire la probabilité $P\left([r_{1,i}, \dots, r_{J,i}]^\top = \epsilon\right)$. Les probabilités de configurations de ruptures P_ϵ sont supposées indépendantes de l'instant i . Par ailleurs, nous supposons que $\mathbf{R}_i = [r_{1,i}, \dots, r_{J,i}]$ est indépendant de $\mathbf{R}_{i'} = [r_{1,i'}, \dots, r_{J,i'}]$ pour tout $i \neq i'$. Notons qu'une hypothèse markovienne plus générale sera envisagée sur \mathbf{R}_i au paragraphe 2.5.2. La loi *a priori* des indicatrices s'écrit :

$$f(\mathbf{R}|\mathbf{P}) = \prod_{\epsilon \in \mathcal{E}} P_\epsilon^{S_\epsilon(\mathbf{R})}, \quad (2.6)$$

où $\mathbf{P} = \{P_\epsilon\}_{\epsilon \in \mathcal{E}}$ et $S_\epsilon(\mathbf{R})$ est le nombre d'instant i tels que $[r_{1,i}, \dots, r_{J,i}]^\top = \epsilon$. Par exemple, dans le cas où deux signaux sont observés \mathbf{y}_1 et \mathbf{y}_2 (i.e., $J = 2$), la loi *a priori* de \mathbf{R} est :

$$f(\mathbf{R}|\mathbf{P}) = P_{00}^{S_{00}} P_{10}^{S_{10}} P_{01}^{S_{01}} P_{11}^{S_{11}}, \quad (2.7)$$

où $S_{00} = \sum_{i=1}^{n-1} (1 - r_{1,i})(1 - r_{2,i})$, $S_{11} = \sum_{i=1}^{n-1} r_{1,i}r_{2,i}$, $S_{10} = \sum_{i=1}^{n-1} r_{1,i}(1 - r_{2,i})$ et $S_{01} = \sum_{i=1}^{n-1} (1 - r_{1,i})r_{2,i}$. Avec cette loi *a priori*, une grande valeur de P_ϵ indique une configuration $[r_{1,i}, \dots, r_{J,i}]^\top = \epsilon$ très probable pour tout $i = 1, \dots, n$. Par exemple, en choisissant une grande valeur de $P_{0\dots 0}$ (respectivement $P_{1\dots 1}$), nous favorisons une absence (respectivement présence) simultanée de changements dans toutes les séquences observées.

Variances et paramètres AR

Des lois inverse Gamma sont choisies comme lois *a priori* des variances de bruit :

$$\sigma_{j,k}^2 \mid \nu, \gamma \sim \mathcal{IG} \left(\frac{\nu}{2}, \frac{\gamma}{2} \right), \quad (2.8)$$

où $\mathcal{IG}(a, b)$ désigne une loi inverse Gamma de paramètres a et b , $\nu = 2$ (comme dans [PADF02]) et γ est un hyperparamètre ajustable. Cette loi *a priori*, dite *conjuguée* pour le paramètre $\sigma_{j,k}^2$, a été utilisée avec succès dans [LP01], [ABD01], [PADF02], [GR98], [ADD01] ou plus récemment [DGI06]. Nous supposons que l'hyperparamètre γ est le même pour tous les signaux observés. Notons, cependant, qu'une analyse similaire peut être conduite avec un ensemble d'hyperparamètres différents $\gamma_j, j = 1, \dots, J$. Une telle analyse se montre intéressante lorsque les amplitudes des signaux sont significativement différentes d'un signal à l'autre et sera développée au paragraphe 2.5.1.

Les lois normales centrées ont été choisies comme distributions *a priori* pour les paramètres AR :

$$\mathbf{a}_{j,k} \mid \sigma_{j,k}^2, \delta_0^2 \sim \mathcal{N}(\mathbf{0}_p, \sigma_{j,k}^2 \delta_0^2 \mathbf{I}_p), \quad (2.9)$$

où \mathbf{I}_p est la matrice identité de taille $p \times p$, $\mathbf{0}_p$ est le vecteur de \mathbb{R}^p constitué de p zéros et δ_0^2 est un hyperparamètre à estimer. Une motivation principale pour le choix de ces lois *a priori* conjuguées est qu'elles permettent l'intégration des variances des bruits et des paramètres AR dans la loi *a posteriori* $f(\Theta \mid \mathbf{Y})$.

2.2.3 Lois *a priori* des hyperparamètres

Le vecteur d'hyperparamètres associé aux lois *a priori* définies précédemment s'écrit $\Phi = (\mathbf{P}, \delta_0^2, \gamma)$. Bien sûr, l'aptitude du modèle bayésien à détecter précisément les ruptures dans les J signaux dépend fortement de la valeur de ces hyperparamètres. Comme cela a été souligné dans la partie 2.1, ces hyperparamètres peuvent être considérés inconnus, puis estimés à partir des données, rendant l'estimation plus robuste, comme par exemple dans [PADF02], [ABD01] ou [AD99]. Le modèle hiérarchique qui en résulte nécessite de définir des lois *a priori*, détaillées dans ce qui suit.

Hyperparamètres δ_0^2 et γ

Une loi non informative de Jeffrey et une loi inverse Gamma vague (c'est-à-dire avec une grande variance) sont choisies comme lois *a priori* des hyperparamètres δ_0^2 et γ , ce qui traduit l'absence de connaissance *a priori* concernant ces hyperparamètres :

$$\begin{aligned} \delta_0^2 | \xi, \beta &\sim \mathcal{IG}(\xi, \beta), \\ f(\gamma) &= \frac{1}{\gamma} \mathbf{1}_{\mathbb{R}^+}(\gamma), \end{aligned} \quad (2.10)$$

où $\mathbf{1}_{\mathbb{R}^+}(x)$ est la fonction indicatrice définie sur \mathbb{R}^+ .

Hyperparamètre \mathbf{P}

La loi *a priori* pour le paramètre \mathbf{P} est, comme au chapitre 1, une loi de Dirichlet avec un vecteur de paramètres $\boldsymbol{\alpha} = [\alpha_{0\dots 0}, \dots, \alpha_{1\dots 1}]$ défini sur le simplexe $\mathcal{P} = \{\mathbf{P} \mid \sum_{\epsilon \in \mathcal{E}} P_\epsilon = 1, P_\epsilon \geq 0\}$. Nous rappelons que cette loi, notée $\mathbf{P} \sim D_{2^J}(\boldsymbol{\alpha})$, permet la marginalisation du paramètre \mathbf{P} dans la loi *a posteriori* $f(\boldsymbol{\Theta} | \mathbf{Y})$. De plus, en choisissant $\alpha_\epsilon = 1, \forall \epsilon \in \mathcal{E}$, la loi de Dirichlet se ramène à une loi uniforme sur le simplexe \mathcal{P} .

En supposant les hyperparamètres indépendants, la loi *a priori* des hyperparamètres Φ peut s'écrire (à une constante de normalisation près) :

$$f(\Phi | \boldsymbol{\alpha}, \xi, \beta) \propto \left(\prod_{\epsilon \in \mathcal{E}} P_\epsilon^{\alpha_\epsilon - 1} \right) \frac{1}{\gamma} \frac{\beta^\xi}{\Gamma(\xi) (\delta_0^2)^{\xi+1}} \exp\left(-\frac{\beta}{\delta_0^2}\right) \mathbf{1}_{\mathbb{R}^+}(\gamma) \mathbf{1}_{\mathbb{R}^+}(\delta_0^2) \mathbf{1}_{\mathcal{P}}(\mathbf{P}), \quad (2.11)$$

où \propto , pour mémoire, signifie « proportionnel à » et $\Gamma(\cdot)$ est la fonction Gamma.

2.2.4 Lois *a posteriori* des paramètres inconnus

La loi *a posteriori* du vecteur de paramètres inconnus $\boldsymbol{\Theta}$ défini au début du paragraphe 2.2 peut être calculée à partir de la structure hiérarchique visualisée sur la figure 2.1 et rappelée ci-dessous :

$$f(\boldsymbol{\Theta} | \mathbf{Y}) = \int f(\boldsymbol{\Theta}, \Phi | \mathbf{Y}) d\Phi \propto \int f(\mathbf{Y} | \boldsymbol{\Theta}) f(\boldsymbol{\Theta} | \Phi) f(\Phi) d\Phi, \quad (2.12)$$

où

$$f(\boldsymbol{\Theta} | \Phi) = f(\mathbf{R} | \mathbf{P}) \prod_{j=1}^J \prod_{k=1}^{K_j} f(\mathbf{a}_{j,k} | \sigma_{j,k}^2, \delta_0^2) f(\sigma_{j,k}^2 | \nu, \gamma), \quad (2.13)$$

et $f(\mathbf{Y}|\Theta)$ et $f(\Phi)$ sont définies dans (2.4) et (2.11).

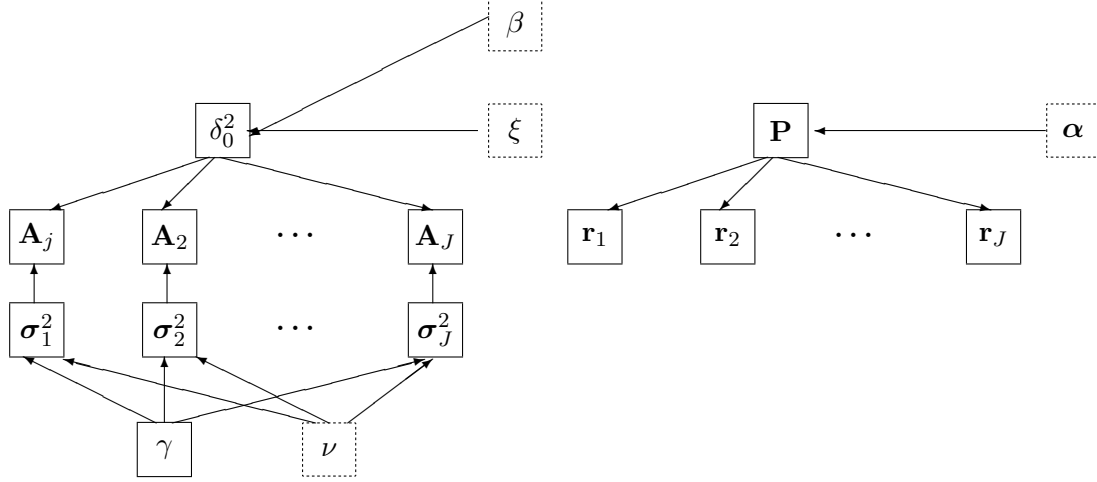


FIG. 2.1 – DAG pour les lois *a priori* (les hyperparamètres qui sont fixés apparaissent dans les carrés en pointillés).

Cette structure hiérarchique permet d'intégrer les paramètres de nuisance $\sigma^2 = \{\sigma_1^2, \dots, \sigma_J^2\}$, $\mathbf{A} = \{\mathbf{A}_1, \dots, \mathbf{A}_J\}$ et \mathbf{P} dans la loi jointe $f(\Theta, \Phi|\mathbf{Y})$, donnant :

$$f(\mathbf{R}, \gamma, \delta_0^2|\mathbf{Y}) \propto \frac{C(\mathbf{R}|\mathbf{Y}, \alpha)}{\gamma} \mathbf{1}_{\mathbb{R}^+}(\gamma) (\delta_0^2)^{-\frac{p}{2} \sum_{j=1}^J K_j(\mathbf{r}_j)} \times \prod_{j=1}^J \prod_{k=1}^{K_j(\mathbf{r}_j)} \left(\frac{\gamma^{\frac{\nu}{2}} |\mathbf{M}_{j,k}|^{\frac{1}{2}} \Gamma(\frac{\nu}{2} + \frac{1}{2} n_{j,k}(\mathbf{r}_j))}{(\gamma + T_{j,k}^2)^{\frac{\nu}{2} + \frac{1}{2} n_{j,k}(\mathbf{r}_j)}} \right) f(\delta_0^2|\xi, \beta), \quad (2.14)$$

avec

$$\begin{cases} T_{j,k}^2 = \mathbf{y}_{j,l_{j,k-1}+1:l_{j,k}}^\top \mathbf{Q}_{j,k} \mathbf{y}_{j,l_{j,k-1}+1:l_{j,k}}, \\ \mathbf{Q}_{j,k} = \mathbf{I}_p - \mathbf{Y}_{j,k} \mathbf{M}_{j,k} \mathbf{Y}_{j,k}^\top, \\ \mathbf{M}_{j,k} = \left(\mathbf{Y}_{j,k}^\top \mathbf{Y}_{j,k} + \frac{\mathbf{I}_p}{\delta_0^2} \right)^{-1}, \end{cases} \quad (2.15)$$

et

$$C(\mathbf{R}|\mathbf{Y}) = \frac{\prod_{\epsilon \in \{0,1\}^J} \Gamma(S_\epsilon(\mathbf{R}) + \alpha_\epsilon)}{\Gamma\left(\sum_{\epsilon \in \{0,1\}^J} (S_\epsilon(\mathbf{R}) + \alpha_\epsilon)\right)}. \quad (2.16)$$

Nous attirons l'attention du lecteur sur le fait que l'intégration des paramètres de nuisance conduit à une loi *a posteriori* d'un vecteur paramètre $\{\mathbf{R}, \gamma, \delta_0^2\}$ appartenant à un espace de dimension fixe.

Pourtant, cette loi reste trop complexe pour que les estimateurs bayésiens classiques (par exemple estimateurs MMSE et MAP) des paramètres inconnus puissent être exprimés de façon simple. De même qu’au chapitre 1, nous proposons un échantillonneur de Gibbs, semblable à celui décrit dans [PADF02], avec toutefois deux différences qu’il semble nécessaire de souligner. Tout d’abord, notre approche permet la segmentation conjointe de plusieurs signaux. Ensuite, l’utilisation des variables indicatrices permet à notre modèle d’évoluer dans un espace de dimension fixe, évitant donc le recours à un algorithme MCMC à sauts réversibles, plus coûteux en temps de calcul. Le paragraphe 2.3 présente l’algorithme MCMC conçu pour réaliser la détection conjointe de ruptures lorsque l’hyperparamètre γ et les ordres des processus AR sont les mêmes dans tous les signaux. Ces hypothèses seront relâchées aux paragraphes 2.5.1 et 2.6.

2.3 Échantillonneur de Gibbs pour la segmentation conjointe de processus AR

L’échantillonneur de Gibbs proposé génère des échantillons (notés $\cdot^{(t)}$ où t est l’indice de l’itération) distribués selon les lois conditionnelles de chaque paramètre. Nous proposons d’échantillonner la loi *a posteriori* $f(\mathbf{R}, \gamma, \delta_0^2 | \mathbf{Y})$ définie dans (2.14) grâce à la procédure décrite ci-dessous. Les étapes de l’algorithme 2.1, ainsi que les équations principales, sont détaillés dans les paragraphes 2.3.1 à 2.3.3.

2.3.1 Génération d’échantillons distribués suivant $f(\mathbf{R} | \gamma, \delta_0^2, \mathbf{Y})$

Cette étape est réalisée en utilisant un échantillonneur de Gibbs pour générer des échantillons distribués suivant $f([r_{1,i}, \dots, r_{J,i}]^T | \gamma, \delta_0^2, \mathbf{Y})$. Le vecteur $[r_{1,i}, \dots, r_{J,i}]^T$ est un vecteur aléatoire de booléens dans l’espace $\mathcal{E} = \{0, 1\}^J$. Par conséquent, sa distribution est entièrement caractérisée par les 2^J probabilités $P([r_{1,i}, \dots, r_{J,i}]^T = \boldsymbol{\epsilon} | \gamma, \delta_0^2, \mathbf{Y})$, $\boldsymbol{\epsilon} \in \mathcal{E}$. En utilisant la notation \mathbf{R}_{-i} pour représenter la matrice \mathbf{R} dans laquelle la colonne à l’instant i a été supprimée, le résultat suivant peut être obtenu :

$$P([r_{1,i}, \dots, r_{J,i}]^T = \boldsymbol{\epsilon} | \mathbf{R}_{-i}, \gamma, \delta_0^2, \mathbf{Y}) \propto f(\mathbf{R}_i(\boldsymbol{\epsilon}), \gamma, \delta_0^2 | \mathbf{Y}), \quad (2.17)$$

-
- Initialisation :
 - Échantillonner le vecteur d'hyperparamètres $\Phi^{(0)} = (\delta_0^{2(0)}, \gamma^{(0)}, \mathbf{P}^{(0)})$ selon la loi définie dans (2.11),
 - Pour $i = 1, \dots, n - 1$, échantillonner $[r_{1,i}^{(0)}, \dots, r_{J,i}^{(0)}]$ selon la loi dans (2.6),
 - Pour $j = 1, \dots, J$, $k = 1, \dots, K_j$, échantillonner $\sigma_{j,k}^{2(0)}$ et $\mathbf{a}_{j,k}^{(0)}$ selon les lois dans (2.8) et (2.9),
 - Poser $t \leftarrow 1$,
 - Itérations : Pour $t = 1, 2, 3, \dots$, faire
 - Pour chaque instant $i = 1, \dots, n - 1$, échantillonner la configuration locale de rupture $[r_{1,i}^{(t)}, \dots, r_{J,i}^{(t)}]$ d'après les probabilités (2.17),
 - Pour chaque séquence $j = 1, \dots, J$, et les segments $k = 1, \dots, K_j$, échantillonner les variances de bruit $\sigma_{j,k}^{2(t)}$ selon les lois conditionnelles données dans (2.18),
 - Échantillonner l'hyperparamètre $\gamma^{(t)}$ selon la loi (2.19),
 - Pour chaque séquence $j = 1, \dots, J$ et les segments $k = 1, \dots, K_j$, échantillonner les paramètres AR $\mathbf{a}_{j,k}^{(t)}$ selon les lois conditionnelles (2.20),
 - Échantillonner l'hyperparamètre $\delta_0^{2(t)}$ selon la loi (2.21),
 - *Étape optionnelle* : échantillonner l'hyperparamètre $\mathbf{P}^{(t)}$ selon la loi (2.22),
 - Poser $t \leftarrow t + 1$.
-

ALGO. 2.1: Échantillonneur de Gibbs pour la segmentation conjointe de processus AR

où $\mathbf{R}_i(\boldsymbol{\epsilon})$ est la matrice \mathbf{R} dont la colonne à l'instant i est remplacée par le vecteur $\boldsymbol{\epsilon}$. Ceci conduit à une expression littérale exacte des probabilités $P\left([r_{1,i}, \dots, r_{J,i}]^\top = \boldsymbol{\epsilon} | \mathbf{R}_{-i}, \gamma, \delta_0^2, \mathbf{Y}\right)$ après une normalisation adéquate.

2.3.2 Génération d'échantillons distribués suivant $f(\gamma, \delta_0^2 | \mathbf{R}, \mathbf{Y})$

Pour obtenir des échantillons distribués suivant $f(\gamma, \delta_0^2 | \mathbf{R}, \mathbf{Y})$, il est possible de générer des vecteurs distribués suivant $f(\gamma, \delta_0^2, \boldsymbol{\sigma}^2, \mathbf{A} | \mathbf{R}, \mathbf{Y})$ en utilisant un échantillonneur de Gibbs approprié. En étudiant la loi jointe $f(\boldsymbol{\Theta}, \boldsymbol{\Phi} | \mathbf{Y})$, cette étape peut se décomposer comme suit :

– **Génération d'échantillons distribués suivant $f(\gamma, \boldsymbol{\sigma}^2 | \mathbf{R}, \delta_0^2, \mathbf{Y})$**

Après avoir intégré les paramètres AR dans la loi jointe $f(\boldsymbol{\Theta}, \boldsymbol{\Phi} | \mathbf{Y})$, les résultats suivants peuvent être obtenus :

$$\sigma_{j,k}^2 | \mathbf{R}, \gamma, \delta_0^2, \mathbf{Y} \sim \mathcal{IG} \left(\frac{\nu + n_{j,k}(\mathbf{r}_j)}{2}, \frac{\gamma + T_{j,k}^2}{2} \right), \quad (2.18)$$

$$\gamma | \mathbf{R}, \boldsymbol{\sigma}^2 \sim \mathcal{G} \left(\frac{\nu}{2} \sum_{j=1}^J K_j(\mathbf{r}_j), \frac{1}{2} \sum_{j=1}^J \sum_{k=1}^{K_j(\mathbf{r}_j)} \frac{1}{\sigma_{j,k}^2} \right), \quad (2.19)$$

où $\mathcal{G}(a, b)$ est la loi Gamma de paramètres a et b .

– **Génération des échantillons distribués suivant $f(\delta_0^2, \mathbf{A} | \mathbf{R}, \boldsymbol{\sigma}^2, \mathbf{Y})$**

Cette génération est réalisée comme suit :

$$\mathbf{a}_{j,k} | \mathbf{R}, \boldsymbol{\sigma}^2, \delta_0^2, \mathbf{Y} \sim \mathcal{N}(\boldsymbol{\mu}_{j,k}, \sigma_{j,k}^2 \mathbf{M}_{j,k}), \quad (2.20)$$

$$\delta_0^2 | \mathbf{R}, \mathbf{A}, \boldsymbol{\sigma}^2 \sim \mathcal{IG} \left(\xi + \frac{p}{2} \sum_{j=1}^J K_j(\mathbf{r}_j), \beta + \sum_{j=1}^J \sum_{k=1}^{K_j(\mathbf{r}_j)} \frac{\|\mathbf{a}_{j,k}\|^2}{2\sigma_{j,k}^2} \right), \quad (2.21)$$

avec $\boldsymbol{\mu}_{j,k} = \mathbf{M}_{j,k} \mathbf{Y}_{j,k}^\top \mathbf{y}_{j,l_{j,k-1}+1:l_{j,k}}$.

2.3.3 Loi *a posteriori* de P_ϵ

Les hyperparamètres P_ϵ , $\epsilon \in \mathcal{E}$, contiennent une information concernant les corrélations entre les instants de ruptures dans les différentes séries temporelles. Par conséquent, il est intéressant en

pratique de les estimer à partir de leur loi *a posteriori* qui, en l'occurrence, est une loi de Dirichlet :

$$\mathbf{P}|\mathbf{R}, \mathbf{Y} \sim \mathcal{D}_{2J}(S_\epsilon(\mathbf{R}) + \alpha_\epsilon). \quad (2.22)$$

2.4 Segmentation de données synthétiques

Les simulations présentées dans cette partie ont été obtenues pour $J = 2$ avec un signal de taille $n = 300$. Les positions des ruptures sont $\mathbf{l}_1 = (60, 150, 300)$ pour le premier signal et $\mathbf{l}_2 = (60, 300)$ pour le second signal. Les paramètres des processus AR sont résumés dans le tableau 2.1. Les paramètres et hyperparamètres fixes sont choisis comme suit : $\nu = 2$ (comme dans [PADF02]), $\xi = 1$ et $\beta = 100$ (afin d'obtenir une loi *a priori* vague), $\alpha_\epsilon = \alpha = 1, \forall \epsilon \in \mathcal{E}$. Les hyperparamètres α_ϵ sont choisis tous égaux afin de s'assurer que la loi de Dirichlet se ramène à une loi uniforme sur le simplexe. De plus, la valeur commune des hyperparamètres α_ϵ est fixée à $\alpha = 1 \ll n$ pour réduire l'influence de ce paramètre dans la loi *a posteriori* (2.22). Afin d'accélérer les calculs, les quantités $T_{j,k}^2$, $\mathbf{Q}_{j,k}$ et $\mathbf{M}_{j,k}$ définies par (2.15) ont été calculées en suivant l'implémentation décrite dans [DI04] et rappelée dans l'annexe A. Toutes les figures ont été obtenues après avoir moyenné les résultats de 64 chaînes de Markov. La longueur de chaque chaîne est fixée à $N_{MC} = 700$, comprenant $N_{bi} = 200$ itérations de chauffage. Ainsi, seuls les 500 dernières itérations des chaînes de Markov sont utilisées pour réaliser les estimations (le choix des paramètres N_{MC} et N_{bi} sera discuté plus loin). Notons que l'exécution de 100 itérations de l'algorithme proposé pour la segmentation conjointe de deux signaux de taille $n = 300$ prend environ 2 minutes et 30 secondes avec une implémentation MATLAB sur PC Pentium IV 2.8 Ghz. Bien sûr, le coût calculatoire peut augmenter avec des séquences de signaux plus longues.

2.4.1 Lois *a posteriori* des positions de ruptures

Les premiers résultats de simulation illustrent l'intérêt de la procédure de segmentation conjointe proposée lorsqu'elle est comparée à une segmentation qui serait réalisée signal par signal¹. La figure 2.2 montre les lois *a posteriori* des instants de ruptures pour deux séquences. Comme nous

¹Pour réaliser une segmentation univariée, il suffit d'appliquer l'algorithme proposé avec $J = 1$, ce qui correspond à une stratégie similaire à celle présentée par Punskeya *et al.* dans [PADF02].

TAB. 2.1 – Paramètres des modèles AR et variances des bruits pour chaque segment de chaque séquence.

Séquence	Segment	$\sigma_{j,k}^2$	$a_{j,k,l}$					
$j = 1$	$k = 1$	0.50	0.0746	0.1664	-0.0693	-0.1571	-0.3543	-0.4277
	$k = 2$	0.52	0.0135	0.1525	0.8170	2.3037	3.5316	2.8567
	$k = 3$	3.80	0.0189	-0.0571	0.1502	-0.3173	0.4824	0.1607
$j = 2$	$k = 1$	0.81	0.0011	-0.0104	0.0538	-0.0646	0.3713	-0.0717
	$k = 2$	4.63	0.0074	0.0138	0.1244	0.2660	0.7677	0.8705

pouvons le voir, la rupture dans la deuxième séquence peut être détectée lorsqu'on utilise la segmentation conjointe (figure de droite) alors qu'elle ne l'est pas lorsqu'on applique deux procédures de segmentation indépendantes (figure de gauche). Lorsqu'une segmentation conjointe est réalisée, la rupture à l'instant $i = 60$ dans le second signal favorise en effet la détection d'une rupture au même instant dans l'autre signal.

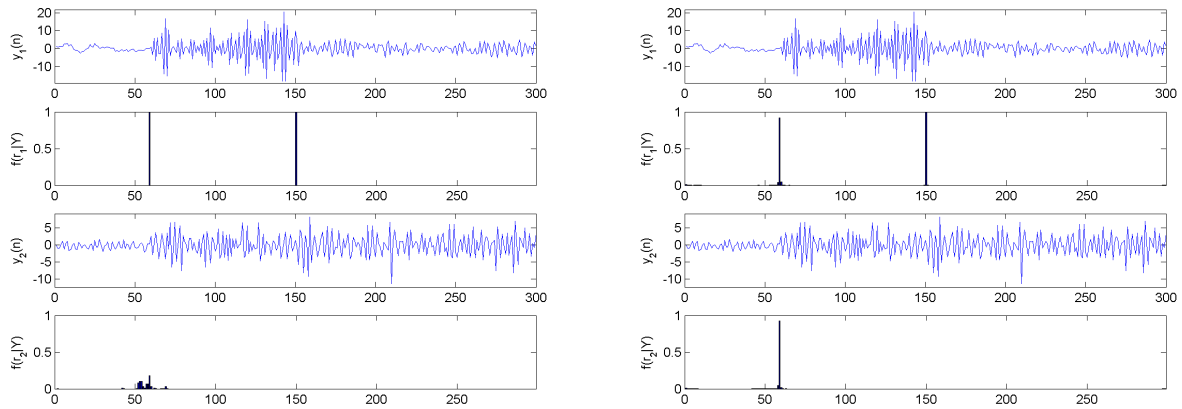


FIG. 2.2 – Probabilités *a posteriori* des instants de ruptures obtenus pour des segmentations 1D (à gauche) et conjointe (à droite) obtenues après une période de chauffage de $N_{bi} = 200$ itérations et $N_r = 500$ itérations d'intérêt.

2.4.2 Lois *a posteriori* des nombres de ruptures

L'estimation du nombre de ruptures présentes dans chacun des deux signaux est un problème important. L'algorithme proposé génère des échantillons $(\mathbf{R}^{(t)}, \gamma^{(t)}, \delta_0^{2(t)})$ distribués suivant la loi *a posteriori* $f(\mathbf{R}, \gamma, \delta_0^2 | \mathbf{Y})$, ce qui permet d'estimer K_j ($j = 1, \dots, J$). En effet, pour chaque matrice d'indicatrices générés $\mathbf{R}^{(t)}$, les nombres de ruptures sont $K_1^{(t)} = \sum_{i=1}^N r_{1,i}^{(t)}$ et $K_2^{(t)} = \sum_{i=1}^N r_{2,i}^{(t)}$. La figure 2.3 représente les moyennes des paramètres K_1 et K_2 ainsi que écarts types associés calculés à partir des 500 derniers échantillons de la chaîne de Markov. Ces histogrammes ont des valeurs maximales en $K_1 = 3$ et $K_2 = 2$ qui sont bien les nombres réels de changements dans les séquences observées.

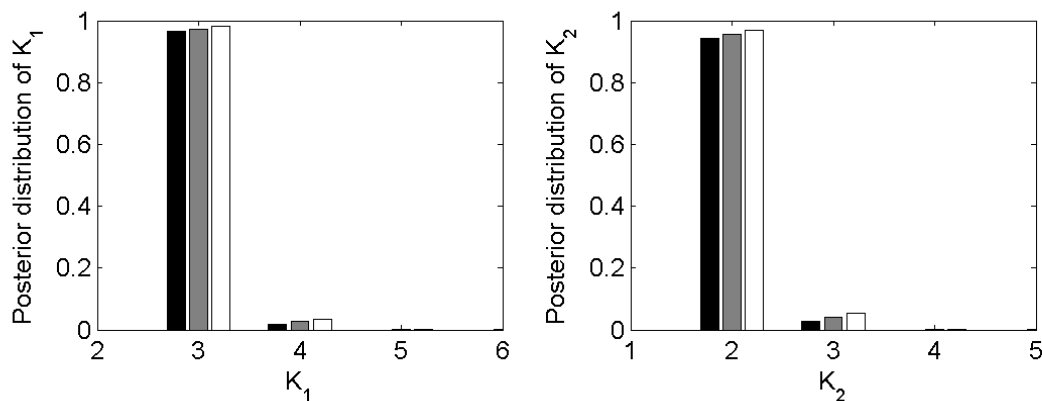


FIG. 2.3 – Lois *a posteriori* des nombres de rupture calculées après $N_r = 500$ itérations d'intérêt (moyenne en gris, moyenne \pm écart type en blanc et noir).

2.4.3 Variances de bruit et paramètres AR

L'estimation des variances des bruits et des paramètres AR peut s'avérer intéressante dans certaines applications pratiques. Les figures 2.4 et 2.5 représentent les lois *a posteriori* des paramètres $\{\sigma_{1,k}^2\}_{k=1,\dots,3}$ et $\{\sigma_{2,k}^2\}_{k=1,2}$ associés aux deux séries temporelles \mathbf{y}_1 et \mathbf{y}_2 . Ces lois sont en accord avec les valeurs réelles des paramètres $\sigma_{1,1}^2 = 0.50$, $\sigma_{1,2}^2 = 0.52$, $\sigma_{1,3}^2 = 3.80$ et $\sigma_{2,1}^2 = 0.81$, $\sigma_{2,2}^2 = 4.63$. Bien

sûr, des résultats similaires peuvent être obtenus pour les paramètres AR.

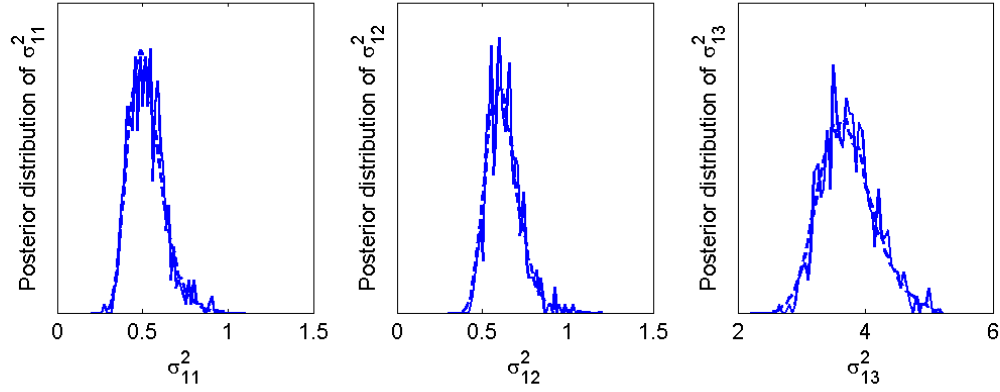


FIG. 2.4 – Lois *a posteriori* des variances de bruit $\sigma^2_{1,k}$ (pour $k = 1, \dots, 3$) sachant $K_1 = 3$ calculés à partir $N_r = 500$ itérations d'intérêt (trait plein). Lois *a posteriori* moyennées à partir des résultats fournis par 64 Markov chaînes (pointillés).

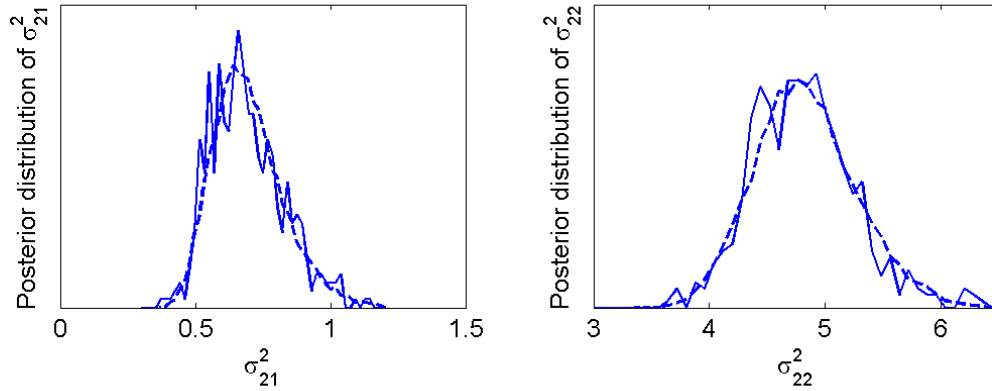


FIG. 2.5 – Lois *a posteriori* des variances $\sigma^2_{2,k}$ (pour $k = 1, 2$) sachant $K_2 = 2$ calculés à partir de $N_r = 500$ itérations d'intérêt (trait plein). Lois *a posteriori* moyennées sur 64 chaînes de Markov (pointillés),

2.4.4 Estimation de l'hyperparamètre \mathbf{P}

Il est possible d'étudier la performance de l'estimation du vecteur d'hyperparamètres \mathbf{P} . Les lois *a posteriori* estimées des probabilités P_{00} , P_{01} , P_{10} et P_{11} sont représentées sur la figure 2.6. Elles montrent que l'échantillonneur de Gibbs permet bien la génération d'échantillons distribués suivant la loi *a posteriori* théorique donnée par (2.22).

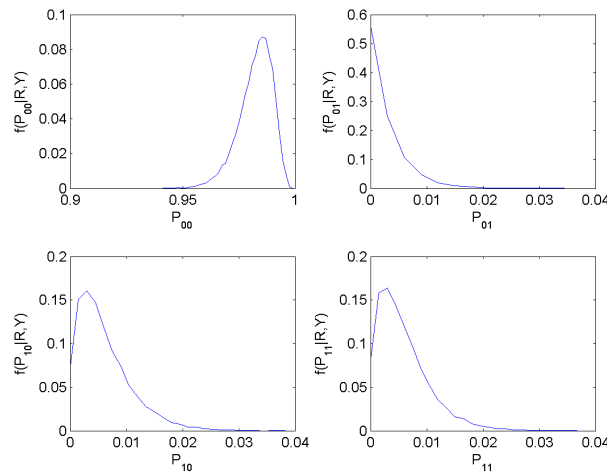


FIG. 2.6 – Lois *a posteriori* des hyperparamètres P_e (calculées à partir de $N_r = 500$ itérations d'intérêts de 64 chaînes de Markov).

2.4.5 Robustesse à des vecteurs de bruits corrélés

Cette partie se propose de démontrer la robustesse de la procédure de segmentation lorsqu'il existe une corrélation entre les vecteurs de bruit. Pour cela, nous supposons que deux vecteurs \mathbf{e}_1 et \mathbf{e}_2 sont

corrélés sur chacun des segments avec les matrices de covariances suivantes :

$$\Sigma_i = \begin{cases} \begin{pmatrix} \sigma_{1,1}^2 & \rho \\ \rho & \sigma_{2,1}^2 \end{pmatrix} & \text{pour } i = 1, \dots, 60, \\ \begin{pmatrix} \sigma_{1,2}^2 & \rho \\ \rho & \sigma_{2,2}^2 \end{pmatrix} & \text{pour } i = 61, \dots, 150, \\ \begin{pmatrix} \sigma_{1,3}^2 & \rho \\ \rho & \sigma_{2,2}^2 \end{pmatrix} & \text{pour } i = 151, \dots, 300. \end{cases}$$

Notons que pour $\rho = 0$, cet exemple se ramène aux données synthétiques précédentes. Les résultats de la procédure de segmentation conjointe sont représentés sur la Figure 2.7 pour deux valeurs du paramètre ρ , c'est-à-dire pour deux corrélations différentes. Les figures 2.2 et 2.7 montrent que des résultats similaires sont obtenus pour des vecteurs de bruit corrélés ou non-corrélés. Par conséquent, le modèle proposé semble robuste aux corrélations éventuelles entre les bruits des différents capteurs.

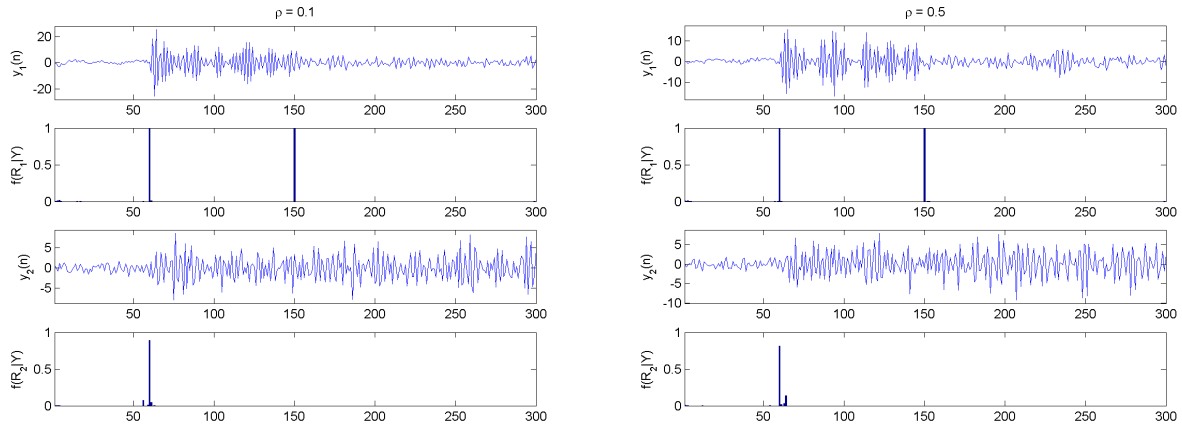


FIG. 2.7 – Probabilités *a posteriori* des instants de ruptures pour des vecteurs de bruit corrélés avec $\rho = 0.1$ (en haut) et $\rho = 0.5$ (en bas) ($N_{\text{bi}} = 200$ et $N_r = 500$).

2.4.6 Convergence de l'échantillonneur

L'échantillonneur de Gibbs permet de générer des échantillons $\left\{ \left(\mathbf{R}^{(t)}, \gamma^{(t)}, \delta_0^{2(t)} \right) \right\}$ asymptotiquement distribués suivant $f(\mathbf{R}, \gamma, \delta_0^2 | \mathbf{Y})$. Les probabilités *a posteriori* de ruptures peuvent donc être

estimées par la moyenne empirique :

$$\hat{\mathbf{R}}_{\text{MMSE}} = \frac{1}{N_r} \sum_{t=1}^{N_r} \mathbf{R}^{(N_{\text{bi}}+t)}, \quad (2.23)$$

où N_{bi} est le nombre d'itérations de chauffage. Cependant, nous avons vu au chapitre précédent qu'il est nécessaire de répondre à deux questions importantes que nous rappelons ici. Tout d'abord, quand peut-on décider que les échantillons $\{\mathbf{R}^{(t)}\}$ sont réellement distribués suivant la loi cible ? Par ailleurs, combien d'échantillons sont nécessaires pour obtenir une estimation précise \mathbf{R} lorsqu'est utilisée l'équation (2.23) ? Comme au chapitre précédent, nous utilisons le critère de variance intra/inter chaîne pour assurer la convergence de l'algorithme. Ce critère nécessite la génération de M chaînes en parallèle de longueur N_r avec différentes valeurs initiales. La convergence des chaînes est mesurée par le facteur d'échelle $\hat{\rho}$ défini par [GCSR95, p. 332] :

$$\sqrt{\hat{\rho}} = \sqrt{\frac{1}{W} \left(\frac{N_r - 1}{N_r} W + \frac{1}{N_r} B \right)}, \quad (2.24)$$

où B et W sont respectivement les variances inter-séquence et intra-séquence pour les M chaînes de Markov construites définies dans (1.17) et (1.18). Une valeur de $\sqrt{\hat{\rho}}$ proche de 1 indique une bonne convergence de l'échantillonneur. Comme au chapitre précédent, nous avons fait le choix de surveiller la convergence de l'échantillonneur de Gibbs avec les paramètres P_{ϵ} , $\epsilon \in \mathcal{E}$. Par exemple, les points générés P_{00} de $M = 5$ chaînes sont représentés sur la Figure 2.8. Ces chaînes convergent clairement vers des valeurs identiques. Le facteur d'échelle pour tous les paramètres P_{ϵ} sont donnés dans le tableau 2.2. Ces valeurs de $\sqrt{\hat{\rho}}$ confirment la bonne convergence de l'échantillonneur (un conseil pour assurer la convergence est une valeur de $\sqrt{\hat{\rho}}$ inférieure à 1.2 [GCSR95, p. 332]). Par ailleurs, il semble opportun de faire les deux remarques suivantes :

- Dans le cadre d'un problème de segmentation, l'information majeure est contenue dans la position des ruptures, ce qui a motivé le choix du paramètres P_{ϵ} pour la surveillance de la convergence. Cependant, pour des applications particulières nécessitant la reconstruction du signal (pour des problèmes inverses par exemple), les paramètres $a_{j,k,l}$ et les variances des bruits $\sigma_{j,k}^2$ sont des paramètres importants. Ainsi, les facteurs d'échelle $\sqrt{\hat{\rho}}$ calculés pour les variances estimées sont aussi indiqués dans le tableau 2.2. Les valeurs obtenues confirment le fait qu'une

TAB. 2.2 – Facteurs d'échelle relatifs aux paramètres P_ϵ , $\sigma_{1,k}^2$ et $\sigma_{2,k}^2$ (calculés à partir de $M = 64$ chaînes de Markov).

	P_ϵ				$\sigma_{1,k}^2$			$\sigma_{2,k}^2$	
	P_{00}	P_{01}	P_{10}	P_{11}	$\sigma_{1,1}^2$	$\sigma_{1,2}^2$	$\sigma_{1,3}^2$	$\sigma_{2,1}^2$	$\sigma_{2,2}^2$
$\sqrt{\hat{\rho}}$	1.0005	1.0002	1.0006	0.9997	1.0005	1.0002	1.0006	1.0005	1.0002

période de chauffage de 200 itérations est suffisante pour l'exemple étudié.

- D'autres exemples de simulations avec des changements moins « nets » peuvent conduire à des problèmes de convergence des méthodes MCMC. Dans de tels cas, une alternative basée sur des techniques de « perfect simulation » peut être envisagée [Fea05]. Elles n'ont pas été rapportées dans le présent document. De telles méthodes empêchent en effet toute utilisation de modèles bayésiens hiérarchiques, puisque ceux-ci ne respectent pas les hypothèses nécessaires à la mise en œuvre de telles techniques. Les hyperparamètres du modèle doivent, dans ce cas, soit être fixés préalablement (notamment à l'aide d'informations *a priori* disponibles sur les paramètres inconnus), soit être estimés par des méthodes MCMC classiques.

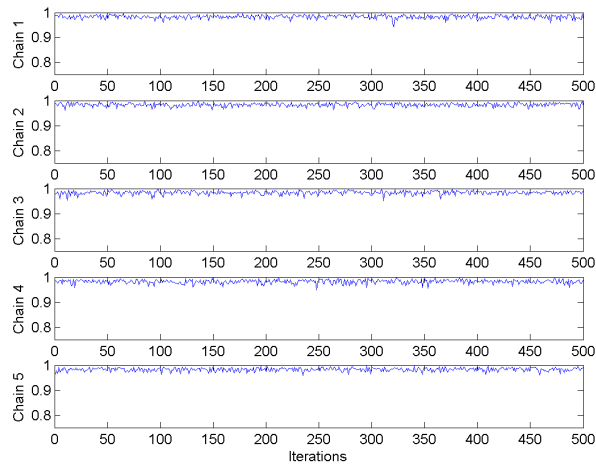


FIG. 2.8 – Diagnostic de convergence : exemples de $M = 5$ chaînes parallèles relatives au paramètre P_{00} .

Une fois que la période de chauffage a été identifiée, et afin de déterminer le nombre d'itérations nécessaires à l'obtention d'une estimation correcte de \mathbf{R} lors de l'utilisation de la moyenne empirique définie par (2.23), un estimateur de référence noté $\tilde{\mathbf{R}}$ est calculé pour un grand nombre d'itérations $N_r = 10000$ (avec $N_{bi} = 200$). La figure 2.9 montre l'erreur quadratique moyenne (MSE) entre cet estimateur de référence $\tilde{\mathbf{R}}$ et l'estimateur obtenu après $N_r = p$ itérations (avec $N_{bi} = 200$) :

$$e_r^2(p) = \left\| \tilde{\mathbf{R}} - \frac{1}{p} \sum_{t=1}^p \mathbf{R}^{(N_{bi}+t)} \right\|^2.$$

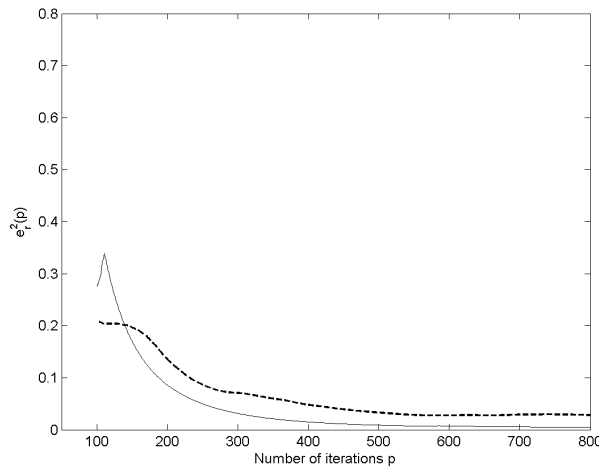


FIG. 2.9 – Erreur quadratique moyenne entre l'estimateur de référence et l'estimateur des probabilités *a posteriori* de ruptures en fonction de p (trait plein). EQMs moyennées à partir de 64 chaînes (tirets).

Cette figure indique qu'un nombre d'itérations égal à $N_r = 500$ est suffisant pour assurer une estimation précise de la moyenne empirique (2.23) pour cet exemple. Bien sûr, pour des problèmes plus délicats, un grand nombre d'itérations peut s'avérer nécessaire pour une estimation acceptable de la loi *a posteriori* des positions de ruptures.

2.4.7 Comparaison avec un modèle bayésien non hiérarchique

Pour justifier une nouvelle fois le deuxième niveau de hiérarchie du modèle bayésien, nous présentons des résultats de segmentation obtenus sur les signaux synthétiques précédents lorsque les

hyperparamètres prennent des valeurs fixées. Comme au chapitre 1, les configurations locales de ruptures sont supposées équiprobables : $P_{00} = P_{01} = P_{10} = P_{11} = \frac{1}{4}$. Ensuite nous présentons sur la figure 2.10 les résultats de simulations pour deux valeurs différentes de l'hyperparamètre δ_0^2 lorsque l'hyperparamètre γ est fixé à $\gamma = 2$. Ces figures montrent encore que le résultat de segmentation est fortement lié aux valeurs de ces hyperparamètres. Très clairement, il apparaît également que de telles valeurs des hyperparamètres conduisent à des modèles de segmentation très peu pertinents.

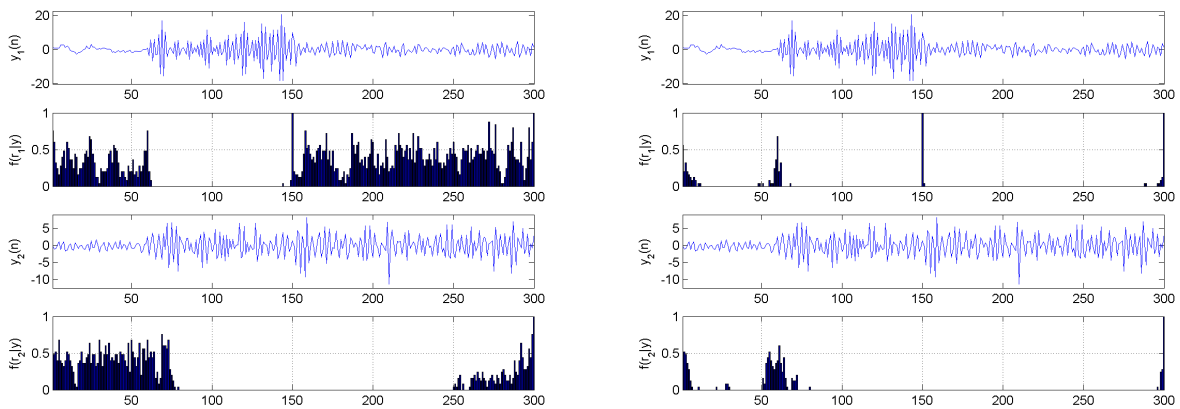


FIG. 2.10 – Probabilités *a posteriori* des instants de ruptures pour deux jeux d'hyperparamètres : $\delta_0^2 = 0.5$ (à gauche) et $\gamma = 5$ (à droite).

2.5 Modèles alternatifs

2.5.1 Signaux de dynamiques différentes

Nous avons supposé dans le modèle précédent que l'hyperparamètre γ était le même pour tous les signaux observés. Ce paragraphe s'attache à montrer qu'une étude similaire peut être menée avec un jeu d'hyperparamètres différents $\gamma_j, j = 1, \dots, J$. Une telle analyse se montre particulièrement intéressante lorsque les amplitudes des signaux auto-régressifs sont significativement différentes d'un signal à l'autre. Dans ce cas, des lois inverse Gamma sont choisies comme lois *a priori* des variances de bruit :

$$\sigma_{j,k}^2 \mid \nu, \gamma \sim \text{IG} \left(\frac{\nu}{2}, \frac{\gamma_j}{2} \right). \quad (2.25)$$

Les lois *a priori* pour les hyperparamètres $\gamma_j, j = 1, \dots, J$ demeurent comme dans l'étude précédente des lois de Jeffrey afin de traduire l'absence de connaissance concernant ces hyperparamètres :

$$f(\gamma_j) = \frac{1}{\gamma_j} \mathbf{1}_{\mathbb{R}^+}(\gamma_j). \quad (2.26)$$

En supposant l'indépendance entre γ_j et $\gamma_{j'}$ pour tout $j \neq j'$, la loi jointe *a priori* de $\boldsymbol{\gamma} = [\gamma_1, \dots, \gamma_J]^\top$ est :

$$f(\boldsymbol{\gamma}) = \prod_{j=1}^J \frac{1}{\gamma_j} \mathbf{1}_{\mathbb{R}^+}(\gamma_j). \quad (2.27)$$

Par conséquent, la loi *a posteriori* qui nous intéresse est donnée par :

$$\begin{aligned} f(\mathbf{R}, \boldsymbol{\gamma}, \delta_0^2 | \mathbf{Y}) &\propto (\delta_0^2)^{-\frac{p}{2} \sum_{j=1}^J K_j(\mathbf{r}_j)} f(\delta_0^2 | \xi, \beta) C(\mathbf{R} | \mathbf{Y}, \boldsymbol{\alpha}) \\ &\times \prod_{j=1}^J \left[\frac{1}{\gamma_j} \mathbf{1}_{\mathbb{R}^+}(\gamma_j) \prod_{k=1}^{K_j} \left(\frac{\gamma_j^{\frac{\nu}{2}} |\mathbf{M}_{j,k}|^{\frac{1}{2}} \Gamma\left(\frac{\nu}{2} + \frac{1}{2} n_{j,k}(\mathbf{r}_j)\right)}{(\gamma_j + T_{j,k}^2)^{\frac{\nu}{2} + \frac{1}{2} n_{j,k}(\mathbf{r}_j)}} \right) \right], \end{aligned} \quad (2.28)$$

où $T_{j,k}^2$, $\mathbf{Q}_{j,k}$, $\mathbf{M}_{j,k}$ et $C(\mathbf{R} | \mathbf{Y})$ ont été définies par (2.15) et (2.16) respectivement.

La génération d'échantillons distribués suivant $f(\boldsymbol{\gamma}, \boldsymbol{\sigma}^2 | \mathbf{R}, \delta_0^2, \mathbf{Y})$ selon l'algorithme 2.1 peut s'effectuer en remplaçant les étapes correspondantes du paragraphe 2.3.2 par :

$$\begin{aligned} \sigma_{j,k}^2 | \boldsymbol{\gamma}, \delta_0^2, \mathbf{R}, \mathbf{Y} &\sim \mathcal{IG} \left(\frac{\nu + n_{j,k}(\mathbf{r}_j)}{2}, \frac{\gamma_j + T_{j,k}^2}{2} \right), \\ \boldsymbol{\gamma}_j | \boldsymbol{\sigma}^2, \mathbf{R}, \mathbf{Y} &\sim \mathcal{G} \left(\frac{\nu}{2} K_j(\mathbf{r}_j), \frac{1}{2} \sum_{k=1}^{K_j(\mathbf{r}_j)} \frac{1}{\sigma_{j,k}^2} \right). \end{aligned} \quad (2.29)$$

Nous démontrons l'intérêt de la méthode en analysant les données synthétiques présentées au paragraphe 2.4. Cependant, l'amplitude du second signal a été multiplié par 0.005 pour produire une différence d'amplitude significative entre les deux signaux analysés. La figure 2.11 montre les lois *a posteriori* des instants de ruptures obtenues pour les deux séquences avec les deux approches. Il apparaît clairement que l'algorithme initial avec un seul hyperparamètre γ (figure de gauche) est incapable de détecter le changement qui se produit à l'instant $i = 60$ dans la première séquence. L'algorithme avec les différents hyperparamètres $\gamma_j, j = 1, \dots, J$ (figure de droite) permet de détecter tous les changements de manière précise.

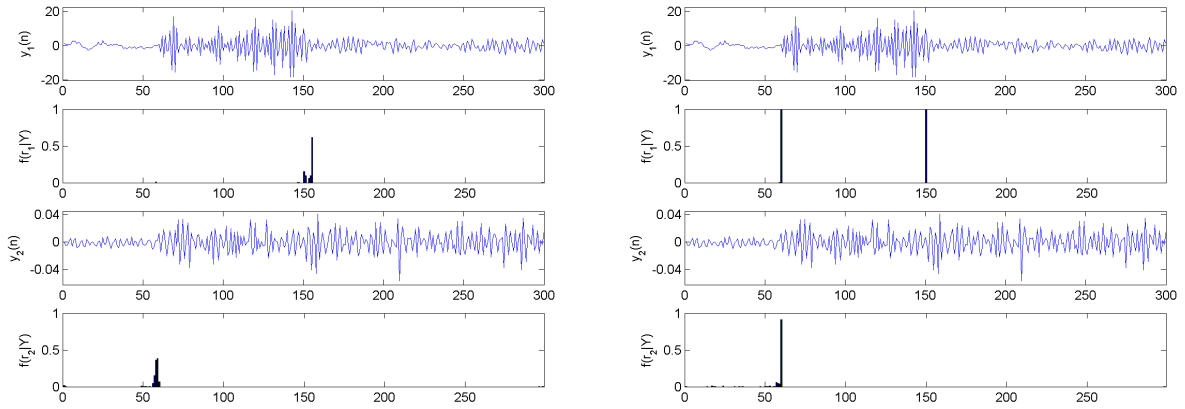


FIG. 2.11 – Probabilités *a posteriori* des instants de ruptures pour des algorithmes avec un hyperparamètre γ (gauche) et plusieurs hyperparamètres γ_j , $j = 1, \dots, J$ (right).

2.5.2 Modèle markovien

Les vecteurs $\mathbf{R}_i = [r_{1,i}, \dots, r_{J,i}]^T$ ($i = 1, \dots, n$) ont été précédemment supposés indépendants. Par conséquent, l'algorithme peut fournir des modèles de segmentation contenant des instants de ruptures très proches. Introduire des contraintes sur la longueur des segments peut être un moyen efficace d'interdire de telles solutions. Ce paragraphe présente un modèle de Markov à états finis sur \mathbf{R}_i qui rejette tous les modèles de segmentation comprenant des segments plus courts qu'une certaine longueur L . Dans un but de clarté et de concision, nous considérons par la suite le cas simple $L = 1$. Cependant, l'analyse peut facilement être généralisée à toute valeur de L .

Nous proposons donc un modèle markovien à 2^J états sur \mathbf{R}_i avec la matrice de transition suivante :

$$\mathbf{P} = \begin{pmatrix} P_{0\dots 0} & P_{0\dots 1} & \cdots & P_{1\dots 1} \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \cdots & 0 \end{pmatrix}. \quad (2.30)$$

En notant $\Pi(\mathbf{R}_i) = \prod_{\epsilon \in \mathcal{E}} P_\epsilon^{\delta(\mathbf{R}_i - \epsilon)}$, on peut montrer que :

$$f(\mathbf{R}_i | \mathbf{R}_{i-1}, \mathbf{P}) = \begin{cases} \Pi(\mathbf{R}_i) & \text{si } \delta(\mathbf{R}_{i-1} - \mathbf{0}) = 1, \\ \delta(\mathbf{R}_i - \mathbf{0}) & \text{sinon,} \end{cases} \quad (2.31)$$

avec $\delta(\mathbf{R}_i - \epsilon) = 1$ si $\mathbf{R}_i = \epsilon$, $\delta(\mathbf{R}_i - \epsilon) = 0$ sinon. Le facteur $C(\mathbf{R} | \mathbf{Y})$ dans la loi *a posteriori* (2.14) est maintenant :

$$C(\mathbf{R} | \mathbf{Y}) = \frac{\prod_{\epsilon \in \mathcal{E}} \Gamma(\tilde{S}_\epsilon(\mathbf{R}) + \alpha_\epsilon)}{\Gamma(\sum_{\epsilon \in \mathcal{E}} (\tilde{S}_\epsilon(\mathbf{R}) + \alpha_\epsilon))}, \quad (2.32)$$

où $\tilde{S}_\epsilon(\mathbf{R})$ est le nombre d'instants tels que $\mathbf{R}_i = \epsilon$ et $\mathbf{R}_{i-1} = \mathbf{0}$. La loi *a posteriori* de P_ϵ reste une loi de Dirichlet :

$$\mathbf{P} | \mathbf{R}, \mathbf{Y} \sim \mathcal{D}_{2J}(\tilde{S}_\epsilon(\mathbf{R}) + \alpha_\epsilon). \quad (2.33)$$

Nous illustrons la performance de la méthode en analysant les données synthétiques du paragraphe 2.4. Cependant, nous avons modifié la deuxième séquence en insérant un changement à l'instant $i = 69$. Nous appliquons l'algorithme proposé avec $L = 2p$, c'est-à-dire que les modèles de segmentation comprenant au moins un segment plus court que deux fois l'ordre p sont interdits. La figure 2.12 montre les lois *a posteriori* des instants de ruptures pour les deux séquences obtenues avec les deux approches. Il apparaît clairement que l'algorithme initial détecte bien des changements de modèles successifs entre les instants $i = 60$ et $i = 69$ (figure de gauche), contrairement à l'algorithme modifié qui inclut un *a priori* markovien pour \mathbf{R}_i (figure de droite). Notons qu'une stratégie similaire peut être conduite dans le cas de la segmentation de données poissonniennes rencontrée au chapitre 1. Une telle étude est décrite dans [DTS06].

2.6 Ordres des modèles inconnus

Cette partie généralise le modèle bayésien hiérarchique précédent à des processus auto-régressifs dont les ordres sont inconnus et peuvent changer d'un segment à l'autre.

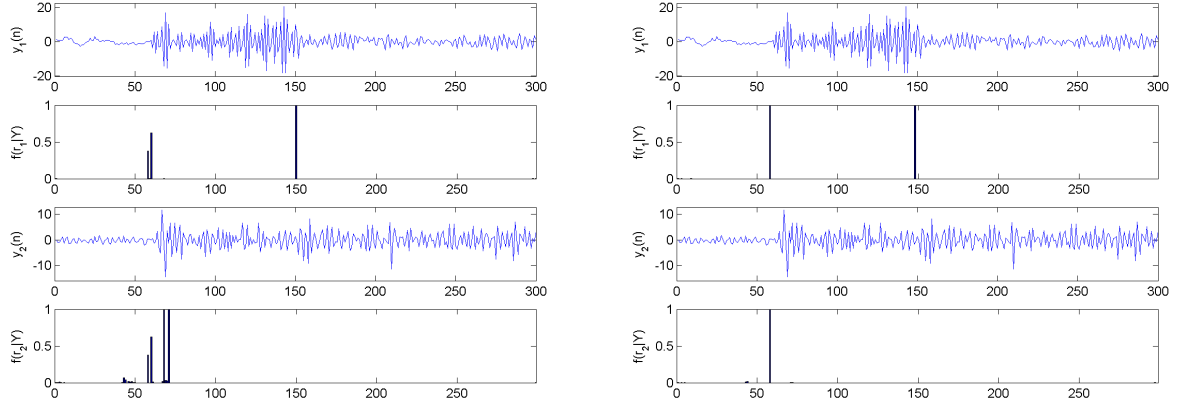


FIG. 2.12 – Probabilités *a posteriori* des instants de ruptures pour des *a priori* indépendants (à droite) et markovien (à gauche) pour \mathbf{R} .

2.6.1 Modèle bayésien étendu

Nous définissons des lois *a priori* pour les nouveaux paramètres à estimer. Une loi de Poisson tronquée est choisie comme loi *a priori* des ordres des modèles AR, notés $p_{j,k}$ sur le k -ième segment du j -ième signal :

$$f(p_{j,k}|\psi) = \frac{1}{\Psi_{p_{\max}}(\psi)} \frac{\psi^{p_{j,k}}}{p_{j,k}!} \mathbf{1}_{\{0, \dots, p_{\max}\}}(p_{j,k}),$$

$$\Psi_{p_{\max}}(\psi) = \sum_{p=0}^{p_{\max}} \frac{\psi^p}{p!}.$$
(2.34)

De manière classique, une loi Gamma conjuguée est choisie pour l'hyperparamètre ψ avec des paramètres μ et ρ fixés tels que la loi résultante soit non informative :

$$\psi|\mu, \rho \sim \mathcal{G}(\rho, \mu).$$
(2.35)

Par conséquent, en supposant l'indépendance entre p_{j_1, k_1} et p_{j_2, k_2} pour tout $j_1 \neq j_2$ et $k_1 \neq k_2$, et en notant $\mathbf{p} = \{\mathbf{p}_1, \dots, \mathbf{p}_J\}$ avec $\mathbf{p}_j = [p_{j,1}, \dots, p_{j, K_j}]^\top$, la loi *a posteriori* des paramètres inconnus

peut s'écrire :

$$f(\mathbf{R}, \mathbf{p}, \gamma, \delta^2, \psi | \mathbf{Y}) \propto \frac{C(\mathbf{R} | \mathbf{Y}, \boldsymbol{\alpha})}{\gamma} f(\delta_0^2 | \xi, \beta) \Psi_{p_{\max}}(\psi)^{-\sum_j K_j} \times \prod_{j=1}^J \prod_{k=1}^{K_j} \left(\frac{\gamma^{\frac{\nu}{2}} |\mathbf{M}_{j,k}|^{\frac{1}{2}} \Gamma\left(\frac{\nu}{2} + \frac{1}{2} n_{j,k}\right) \psi^{p_{j,k}}}{\left(\gamma + T_{j,k}^2\right)^{\frac{\nu}{2} + \frac{1}{2} n_{j,k}} p_{j,k}!} (\delta_0^2)^{-\frac{p_{j,k}}{2}} \right) \mathbf{1}_{\mathbb{R}^+}(\gamma), \quad (2.36)$$

où $C(\mathbf{R} | \mathbf{Y})$ a été définie précédemment dans (2.16). Nous remarquons que les dimensions des matrices $\mathbf{M}_{j,k}$ et donc les quantités $T_{j,k}^2$ définie dans (2.15) dépendent maintenant de l'ordre $p_{j,k}$.

2.6.2 Algorithme MCMC à sauts réversibles

La loi cible (2.36) nous impose de développer une stratégie efficace qui permette d'échantillonner suivant $f(\mathbf{R}, \mathbf{p}, \gamma, \delta_0^2, \psi | \mathbf{Y})$. Dans ce cas, les vecteurs à échantillonner $(\mathbf{R}, \mathbf{p}, \gamma, \delta_0^2, \psi)$ appartiennent à l'espace $\{0, 1\}^{nJ} \times \prod_{j=1}^J \{0, \dots, p_{\max}\}^{K_j} \times \mathbb{R}^+ \times \mathbb{R}$ dont la dimension dépend clairement des nombres de ruptures K_j . Afin de générer directement des vecteurs dans cet espace, nous proposons un algorithme de Gibbs hybride noté « algorithme 2.2 » dont les étapes principales sont détaillées ci-dessous :

a- Génération d'échantillons suivant $f(\mathbf{R} | \mathbf{p}, \gamma, \delta_0^2, \psi, \mathbf{Y})$

Comme dans le modèle initial, cette génération est accomplie en utilisant $n - 1$ étapes de Gibbs pour générer des échantillons distribués suivant $f(r_{1,i}, \dots, r_{J,i} | \mathbf{p}, \gamma, \delta_0^2, \psi, \mathbf{Y})$. Les 2^J probabilités $P\left([r_{1,i}, \dots, r_{J,i}]^\top = \boldsymbol{\epsilon} | \mathbf{R}_{-i}, \mathbf{p}, \gamma, \delta_0^2, \psi, \mathbf{Y}\right)$ peuvent être évaluées de manière exacte avec les deux règles de mise à jour suivantes pour le vecteur \mathbf{p} :

- si, dans la séquence j , les deux segments consécutifs k_1 et k_2 doivent être fusionnés, l'ordre $p_{j,k}^*$ du modèle AR défini sur le segment résultant est $p_{j,k}^* = p_{j,k_1} + p_{j,k_2}$ où p_{j,k_1} et p_{j,k_2} sont les ordres des modèles AR sur les segments k_1 et k_2 .
- si, dans la séquence j , le segment k sur lequel est défini un modèle AR d'ordre $p_{j,k}$ doit être divisé, les ordres p_{j,k_1}^* et p_{j,k_2}^* des processus AR sur les deux segments résultants sont choisis comme suit : $p_{j,k_1}^* \sim \mathcal{U}_{\{1, \dots, p_{j,k}\}}$ et $p_{j,k_2}^* = p_{j,k} - p_{j,k_1}^*$.

Ces choix assurent la réversibilité des mouvements proposés d'un espace à un espace de dimension différente.

-
- Initialisation :
 - Échantillonner le vecteur d'hyperparamètre $\Phi^{(0)} = (\delta_0^{2(0)}, \gamma^{(0)}, \mathbf{P}^{(0)})$ selon la loi définie par (2.11),
 - Échantillonner l'hyperparamètre $\psi^{(0)}$ selon la loi définie par (2.35),
 - Pour $i = 1, \dots, n-1$, échantillonner $[r_{1,i}^{(0)}, \dots, r_{J,i}^{(0)}]$ selon la loi (2.6),
 - Pour $j = 1, \dots, J$, $k = 1, \dots, K_j$, échantillonner $\sigma_{j,k}^{2(0)}$, $\mathbf{a}_{j,k}^{(0)}$ et $p_{j,k}^{(0)}$ selon les lois (2.8), (2.9) et (2.34),
 - Poser $t \leftarrow 1$,
 - Itérations : pour $t = 1, 2, 3, \dots$, faire
 - Pour chaque instant $i = 1, \dots, n-1$, échantillonner la configuration locale de rupture $[r_{1,i}^{(t)}, \dots, r_{J,i}^{(t)}]$ d'après les 2^J probabilités définie à l'étape a-
 - Pour chaque séquence $j = 1, \dots, J$, $k = 1, \dots, K_j$, mettre à jour l'ordre $p_{j,k}^{(t)}$ du modèle (voir étape b-) :
 - si $(u \sim \mathcal{U}_{[0,1]}) \leq b_{p_{j,k}^{(t-1)}}$, alors proposer $p_{j,k}^* = p_{j,k}^{(t-1)} + 1$,
 sinon si $(u \sim \mathcal{U}_{[0,1]}) \leq b_{p_{j,k}^{(t-1)}} + d_{p_{j,k}^{(t-1)}}$, alors proposer $p_{j,k}^* = p_{j,k}^{(t-1)} - 1$,
 - si $(v_p \sim \mathcal{U}_{[0,1]}) \leq \lambda_{p_{j,k} \rightarrow p_{j,k}^*}$ (voir (2.37)), $p_{j,k}^{(t)} = p_{j,k}^*$,
 sinon $p_{j,k}^{(t)} = p_{j,k}^{(t-1)}$,
 - Mettre à jour $\psi^{(t)}$ (voir étape c-) :
 - Proposer ψ^* selon la loi de proposition Gamma définie à l'étape d-
 - si $(v_\psi \sim \mathcal{U}_{[0,1]}) \leq \lambda_{\psi \rightarrow \psi^*}$ (voir (2.38)), alors $\psi^{(t)} = \psi^*$,
 sinon $\psi^{(t)} = \psi^{(t-1)}$,
 - Pour chaque séquence $j = 1, \dots, J$, $k = 1, \dots, K_j$, échantillonner $\sigma_{j,k}^{2(t)}$ d'après la loi (2.39),
 - Échantillonner $\gamma^{(t)}$ d'après la loi (2.40),
 - Pour $j = 1, \dots, J$, $k = 1, \dots, K_j$, échantillonner $\mathbf{a}_{j,k}^{(t)}$ d'après la loi (2.41),
 - Échantillonner $\delta_0^{2(t)}$ d'après la loi (2.42),
 - *Étape optionnel* : Échantillonner l'hyperparamètre $\mathbf{P}^{(t)}$ d'après la loi (2.43),
 - Poser $t \leftarrow t + 1$.
-

ALGO. 2.2: Algorithme MCMC à sauts réversibles pour la détection de ruptures

b- Génération d'échantillons distribués suivant $f(\mathbf{p} | \mathbf{R}, \gamma, \delta_0^2, \psi, \mathbf{Y})$

Afin de générer des vecteurs $\mathbf{p}^{(t)}$ évoluant dans des espaces de dimension différente, il est nécessaire de faire appel à des méthodes MCMC à sauts réversibles. Cette stratégie d'échantillonnage a été initialement introduite par Peter Green dans [Gre95] et a été largement utilisée dans la littérature du traitement du signal (voir [AD99], [ADD01], [LR02], [VADG04] ou plus récemment [DGI06]). Comme dans [PADF02], la mise à jour des ordres est effectuée grâce aux deux

mouvements suivants :

- un mouvement de naissance $p_{j,k}^* = p_{j,k} + 1$ est proposé avec une probabilité $b_{p_{j,k}}$,
- un mouvement de mort $p_{j,k}^* = p_{j,k} - 1$ est proposé avec une probabilité $d_{p_{j,k}}$.

La probabilité d'acceptation du nouvel état dans la chaîne de Monte Carlo est alors :

$$\begin{aligned} \lambda_{p_{j,k} \rightarrow p_{j,k}^*} &= \frac{\psi^{p_{j,k}^*} p_{j,k}^*!}{\psi^{p_{j,k}} p_{j,k}!} \left(\frac{1}{\delta_0^2} \right)^{\pm \frac{1}{2}} \frac{|\mathbf{M}_{j,k}(p_{j,k}^*)|^{\frac{1}{2}}}{|\mathbf{M}_{j,k}(p_{j,k})|^{\frac{1}{2}}} \\ &\times \frac{[T_{j,k}^2(p_{j,k}) + \gamma]^{\frac{\nu}{2} + \frac{1}{2}n_{j,k}}}{[T_{j,k}^2(p_{j,k}^*) + \gamma]^{\frac{\nu}{2} + \frac{1}{2}n_{j,k}}} \left(\frac{d_{p_{j,k}}}{b_{p_{j,k}}} \right)^{\pm 1}, \end{aligned} \quad (2.37)$$

avec $p_{j,k}^* = p_{j,k} \pm 1$.

c- **Génération d'échantillons distribués suivant $f(\psi | \mathbf{R}, \mathbf{p}, \gamma, \delta_0^2, \mathbf{Y})$**

En examinant avec intérêt la loi *a posteriori*, nous remarquons qu'il est possible de générer des échantillons ψ par une simple étape de Metropolis-Hastings avec une loi de proposition Gamma $\psi^* \sim \mathcal{G}(\mu + \sum_{j,k} p_{j,k}, \rho + \sum_j K_j)$. La probabilité d'acceptation est alors calculée :

$$\lambda_{\psi \rightarrow \psi^*} = \left[\frac{\Psi_{p_{\max}}(\psi)}{\Psi_{p_{\max}}(\psi^*)} \exp(\psi^* - \psi) \right]^{\sum_{j=1}^J K_j}. \quad (2.38)$$

d- **Génération d'échantillons distribués suivant $f(\gamma, \sigma^2 | \mathbf{R}, \mathbf{p}, \delta_0^2, \mathbf{Y})$**

Comme dans le modèle initial, après intégration adéquate, les lois *a posteriori* suivantes peuvent être obtenues :

$$\sigma_{j,k}^2 | \mathbf{R}, \mathbf{p}, \gamma, \delta_0^2, \mathbf{Y} \sim \mathcal{IG} \left(\frac{\nu + n_{j,k}}{2}, \frac{\gamma + T_{j,k}^2}{2} \right), \quad (2.39)$$

$$\gamma | \mathbf{R}, \sigma^2 \sim \mathcal{G} \left(\frac{\nu}{2} \sum_j K_j, \frac{1}{2} \sum_{j,k} \frac{1}{\sigma_{j,k}^2} \right). \quad (2.40)$$

e- **Génération d'échantillons distribués suivant $f(\delta_0^2, \mathbf{A} | \mathbf{R}, \mathbf{p}, \sigma^2, \mathbf{Y})$**

Cette génération est accomplie de la manière suivante :

$$\mathbf{a}_{j,k} | \mathbf{R}, \mathbf{p}, \sigma^2, \delta_0^2, \mathbf{Y} \sim \mathcal{N}(\boldsymbol{\mu}_{j,k}, \sigma_{j,k}^2 \mathbf{M}_{j,k}), \quad (2.41)$$

$$\delta_0^2 | \mathbf{R}, \mathbf{p}, \mathbf{A}, \sigma^2 \sim \mathcal{IG} \left(\xi + \sum_{j,k} \frac{p_{j,k}}{2}, \beta + \sum_{j,k} \frac{\|\mathbf{a}_{j,k}\|^2}{2\sigma_{j,k}^2} \right). \quad (2.42)$$

f- Génération d'échantillons distribués suivant $f(\mathbf{P} | \mathbf{R}, \mathbf{Y})$

Comme précédemment, la loi *a posteriori* suivante est obtenue :

$$\mathbf{P} | \mathbf{R}, \mathbf{Y} \sim \mathcal{D}_{2J}(S_\epsilon(\mathbf{R}) + \alpha_\epsilon). \quad (2.43)$$

Il est important de noter que la stratégie proposée ne nécessite qu'une étape de sélection d'ordre de modèle (c'est-à-dire une seule procédure à sauts réversibles), contrairement à l'approche présentée dans [PADF02] qui nécessite deux étapes à sauts réversibles.

2.6.3 Simulations

Afin de vérifier la performance de la méthode proposée, nous considérons $J = 2$ signaux synthétiques de longueur $n = 300$. Les positions de ruptures sont $\mathbf{l}_1 = (60, 150, 300)$ pour le premier signal et $\mathbf{l}_2 = (60, 300)$ pour le second signal. Les paramètres des processus AR (semblables à ceux présentés dans [PADF02]) sont résumés dans le tableau 2.3. Les paramètres et hyperparamètres fixes ont été choisis comme suit : $\nu = 2$, $\xi = 1$, $\mu = 1$, $\beta = 10^2$, $\rho = 10^{-2}$ (loi *a priori* des hyperparamètres vagues) et $\alpha_\epsilon = 1, \forall \epsilon \in \mathcal{E}$ afin d'obtenir une loi *a priori* uniforme sur le simplexe \mathcal{P} pour \mathbf{P} . Les histogrammes des ordres estimés des modèles AR associés à chacune des deux séquences sont représentés sur les figures 2.13 et 2.14. Les probabilités *a posteriori* de changements correspondantes sont illustrées sur la figure 2.15. L'algorithme proposé réussit parfaitement à estimer les changements dans les deux séquences. Par ailleurs, l'ordre du processus AR dans chaque segment est estimé avec précision.

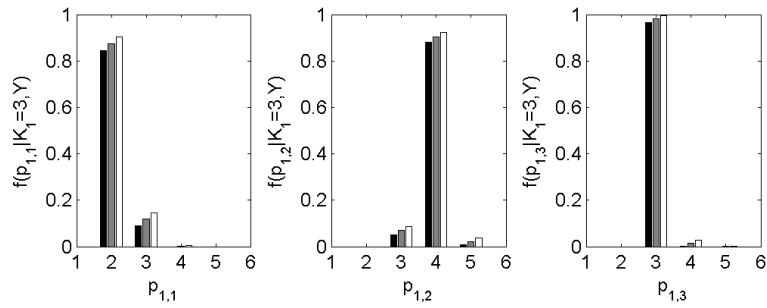
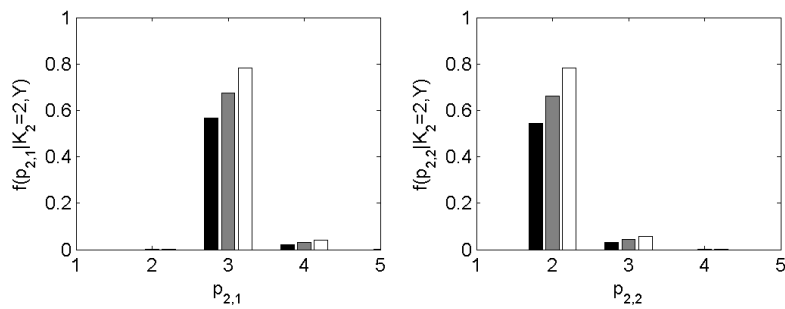
2.7 Applications

2.7.1 Segmentation de signaux de parole stéréo

Cette partie illustre la performance de l'algorithme proposé en l'appliquant à un signal de parole réel qui a reçu beaucoup d'intérêt dans la littérature (voir [BN93], [Gus00], [PADF02], [And88] et plus récemment [Fea05]). Comme il l'est expliqué dans [BN93, p. 401], ce signal appartient à une base de

TAB. 2.3 – Paramètres des processus AR et variances des bruits pour chaque segment de chaque séquence.

Séquence	Segment	$\sigma_{j,k}^2$	$p_{j,k}$	$a_{j,k,l}$			
$j = 1$	$k = 1$	1.7	2	-0.8000	0.5200	—	—
	$k = 2$	1.6	4	2.3000	2.6675	1.8437	0.5936
	$k = 3$	1.8	3	0.5000	-0.6100	-0.5850	—
$j = 2$	$k = 1$	0.5	3	-2.0000	1.6350	-0.5075	—
	$k = 2$	0.6	2	1.7000	0.7450	—	—

FIG. 2.13 – Lois *a posteriori* des ordres des processus p_{1i} (pour $i = 1, \dots, 3$) sachant $K_1 = 3$.FIG. 2.14 – Lois *a posteriori* des ordres des processus p_{2i} (pour $i = 1, 2$) sachant $K_2 = 2$.

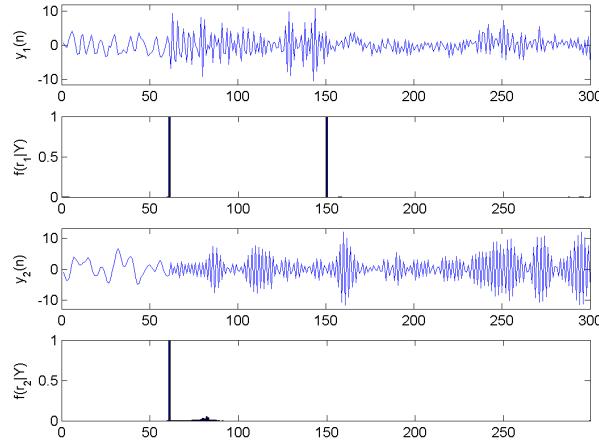


FIG. 2.15 – Probabilités *a posteriori* des instants de ruptures estimées par l'algorithme à sauts réversibles.

données conçue par le Centre National d'Études des Télécommunications (CNET). Il consiste en un signal de parole bruité enregistré dans une voiture à une fréquence d'échantillonnage de 8kHz et avec un niveau de quantification de 16 bits. Il a été filtré par un filtre passe-haut de fréquence de coupure égale à 150Hz. Les données unidimensionnelles $\mathbf{y} = [y_1, \dots, y_n]$ sont d'abord traitées par l'algorithme proposé avec $J = 1$. Le nombre estimé de ruptures et leurs positions sont obtenus après $N_{MC} = 600$ itérations dont $N_{bi} = 200$ itérations de chauffage. Les ruptures estimées sont représentées sur la figure 2.16 (figure du haut). Le tableau 2.4 permet de comparer ces estimations avec celles obtenues par différentes méthodes précédemment utilisées dans la littérature. Il apparaît clairement que la méthode proposée fournit des segmentations très similaires. Cependant, elle a également l'avantage de pouvoir traiter des signaux issus de plusieurs capteurs. Pour illustrer ce point, les données ont été converties en des données stéréo $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2]^T$ avec $\mathbf{y}_1 = [y_{1,1}, \dots, y_{1,n}]$ et $\mathbf{y}_2 = [y_{2,1}, \dots, y_{2,n}]$ en utilisant un convertisseur mono-stéréo standard. Les lois *a posteriori* des positions de ruptures pour les deux signaux \mathbf{y}_1 et \mathbf{y}_2 ont été fournies par l'algorithme avec $J = 2$. Les segments pour les deux séries temporelles peuvent comme précédemment être obtenus en conservant les plus grandes valeurs des probabilités de ruptures *a posteriori* (correspondant aux nombres \hat{K}_j de ruptures estimés dans

chaque séquence, $j = 1, 2$). Les résultats obtenus sont présentés dans le tableau 2.4 et sur la figure 2.16 (tracés du milieu et d'en bas). Ils sont en accord avec ceux d'une segmentation unidimensionnelle. Notons toutefois que la segmentation des données stéréo ne conduit pas à l'estimation du premier changement à $l_{1,1} = l_{2,1} = 448$ puisqu'il ne semble pas significatif dans les deux séquences.

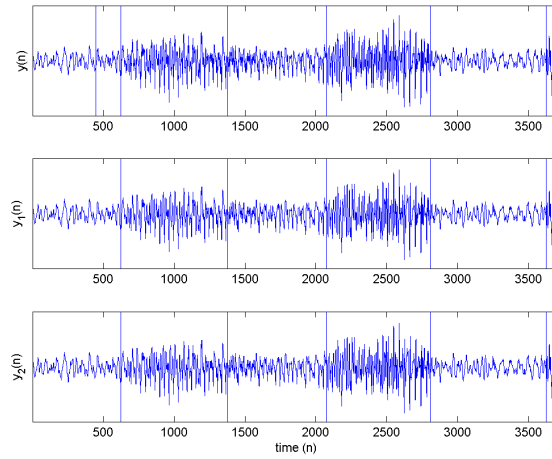


FIG. 2.16 – Segmentations 1D (en haut) et 2D (au milieu et en bas) d'un signal de parole réel ($N_{\text{bi}} = 200$ et $N_{\text{MC}} = 600$).

TAB. 2.4 – Positions des ruptures estimées par différentes méthodes.

Méthode	ordre AR	Positions des ruptures									
		445	—	645	1550	1800	2151	2797	—	3626	—
Divergence [BB83]	16	445	—	645	1550	1800	2151	2797	—	3626	—
GLR [AB83]	16	445	—	645	1550	1800	2151	2797	—	3626	—
GLR [AB83]	2	445	—	645	1550	1750	2151	2797	3400	3626	—
Approx. ML [Gus00]	16	445	—	626	1609	—	2151	2797	—	3627	—
1D MCMC [PADF02]	estimé	448	—	624	1377	—	2075	2807	—	3626	—
Conditional MAP [Fea05]	estimé	449	585	620	1365	1795	2144	2811	—	3624	3657
Approche proposée 1D	estimé	448	—	624	1377	—	2075	2807	—	3626	—
Approche proposée 2D $\begin{matrix} \swarrow \\ \searrow \end{matrix}$	Canal 1	—	—	624	1377	—	2075	2809	—	3626	—
	Canal 2	—	—	624	1377	—	2075	2809	—	3626	—

2.7.2 Détection d' « arc-tracking »

Dans ce paragraphe, nous nous proposons d'analyser des données aéronautiques réelles, pour lesquelles l'enjeu est de prévenir le phénomène appelé « arc-tracking. » Ce phénomène a été responsable d'un grand nombre d'accidents d'avions ces dernières années. Parmi ces accidents, on peut citer l'explosion en plein vol en 1996, au large des côtes de Long Island, du Boeing 747-131 du vol TWA 800 reliant New York – J.F.K. à Paris – Charles de Gaulle (voir [BL04], [TS01] et surtout le rapport d'enquête [Nat00, p. 143]). Les quelques centaines de kilomètres de câbles embarqués dans les avions civils et militaires sont sujets à de fortes contraintes chimiques, mécaniques et thermiques, provoquant des altérations des gaines isolantes. Ces défauts exposent les câbles à des courts-circuits qui peuvent mettre le feu aux câbles voisins [FH01]. Plusieurs méthodes pour détecter ces défauts d'isolation ont été étudiées dans la littérature : elles sont principalement basées sur les propriétés diélectriques [HK97] ou plus récemment électromagnétiques [LF05] des câbles. Nous proposons dans ce travail une procédure de détection d'arc-tracking basée sur la recherche de transitoires dans les câbles endommagés. Ces transitoires constituent en effet des phénomènes précurseurs caractéristiques de courts-circuits.

Les données réelles analysées sont issues de l'enregistrement de tensions triphasées (phases notées A , B et C) dans un réseau électrique dont la fréquence est f_0 . Le phénomène que nous cherchons à détecter affecte les signaux à des fréquences supérieures à f_c . Par conséquent, les $J = 3$ séquences de taille $n = 551$ sont filtrées par un filtre passe-haut pour mettre en évidence les transitoires, beaucoup moins énergétiques². La présence de ces transitoires dans les séries temporelles alors observées conduit à des changements dans les paramètres auto-régressifs.

Nous proposons donc de détecter les transitoires qui apparaissent dans les phases A , B et C entre les instants $t_1 = 0.04$ s et $t_2 = 0.17$ s. Les signaux correspondant aux trois phases mesurées sont traités par l'algorithme de segmentation conjointe proposé. Le nombre de ruptures et leurs positions dans chacun des 3 signaux sont obtenus après $N_{MC} = 450$ itérations comprenant une période de chauffage de $N_{bi} = 100$ itérations. Les paramètres N_{bi} et N_{MC} ont été choisis tels qu'une valeur acceptable soient obtenue pour le facteur d'échelle $\sqrt{\hat{\rho}}$ pour les hyperparamètres P_ϵ ($\epsilon \in \mathcal{E}$).

²Pour des raisons de confidentialité, les valeurs réelles de f_0 et f_c associées à ces données aéronautiques ne peuvent être fournies.

Dans un premier temps de l'analyse, nous estimons le nombre de ruptures dans chaque séquence comme au paragraphe 2.4.2. Les lois *a posteriori* des nombres de changements sont représentées sur la figure 2.17. Les estimateurs MAP correspondants sont $\hat{K}_1 = 4$, $\hat{K}_2 = 8$ et $\hat{K}_3 = 8$.

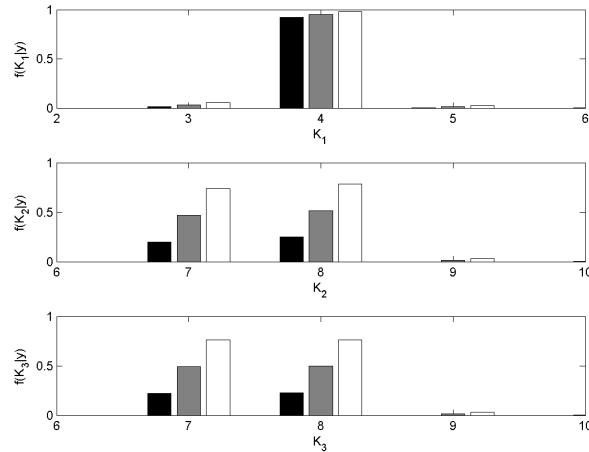


FIG. 2.17 – Probabilités *a posteriori* des instants de ruptures pour la segmentation conjointe de signaux aéronautiques 3D ($N_{bi} = 100$ et $N_{MC} = 450$).

La loi *a posteriori* des configurations globales de ruptures \mathbf{R} représentée sur la figure 2.18 peut ensuite être utilisée pour estimer le début et la fin des transitoires dans chaque phase. En effet, en conservant les \hat{K}_j plus grands pics dans la loi *a posteriori* $f(\mathbf{R}|\mathbf{Y})$, les segments correspondants aux différents transitoires (illustrés par les lignes verticales sur la figure 2.18) peuvent clairement être repérés.

2.8 Conclusions

Nous avons étudié dans ce chapitre une procédure bayésienne de détection de ruptures qui permet de segmenter des signaux enregistrés par différents capteurs. L'approche proposée suppose que les signaux peuvent être modélisés par des processus AR dont les paramètres sont constants par morceaux. Un modèle bayésien hiérarchique, grâce à des lois *a priori* choisies de manière adéquate, permet d'estimer conjointement les positions des ruptures dans les séquences observées. Nous avons montré

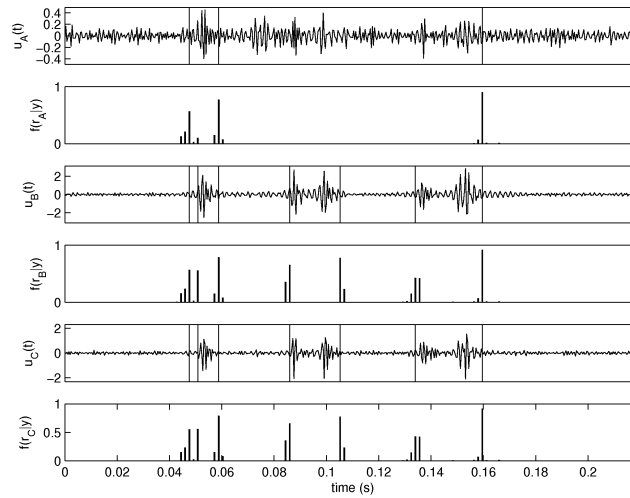


FIG. 2.18 – Probabilités *a posteriori* des instants de ruptures et résultat de segmentation des données aéronautiques triphasées ($N_{bi} = 100$ et $N_{MC} = 450$).

qu'un traitement conjoint peut améliorer les performances d'estimation.

Nous avons également étudié dans ce chapitre deux modèles permettant de segmenter des signaux de dynamiques très différentes ou d'imposer une durée minimale entre deux instants de ruptures consécutifs.

Une dernière partie du chapitre a été consacrée à l'étude des processus AR d'ordres inconnus pouvant varier d'un segment à l'autre. Nous avons vu que de tels signaux peuvent être segmentés à l'aide d'algorithmes MCMC à sauts réversibles.

Deux applications à des données réelles ont également été proposées. La première, fréquemment rencontrée dans la littérature, consiste en la segmentation de parole. Dans ce chapitre, nous illustrons l'intérêt d'un algorithme bayésien hiérarchique pour la segmentation moins classique des enregistrements stéréo. La seconde application est une méthode originale de détection d'effluves électriques dans des données aéronautiques. Un traitement multi-capteur des signaux tri-phasés permet de repérer les transitoires qui apparaissent dans les enregistrement de tension afin de prévenir l'apparition d'arcs électriques se propageant le long des câbles.

CHAPITRE 3

Démélange linéaire d'images hyperspectrales

Sommaire

3.1	Introduction	73
3.2	Modèle de mélange linéaire	78
3.3	Modèle bayésien hiérarchique	80
3.4	Échantillonneur de Gibbs pour l'estimation des abondances	82
3.5	Résultats de simulations sur des données synthétiques	84
3.6	Démélange spectral d'une image AVIRIS	89
3.7	Cas d'un bruit gaussien corrélé	95
3.8	Estimation du nombre de pôles de mélange à l'aide d'un algorithme à sauts réversibles	105
3.9	Conclusions	112

3.1 Introduction

3.1.1 Imagerie hyperspectrale

Description

L'observation de la Terre à partir de capteurs spatiaux ou aéroportés a d'abord été effectuée en acquisition monochrome, principalement dans le domaine militaire. Il est ensuite apparu que l'observation de la même scène dans plusieurs bandes spectrales, i.e. à différentes longueurs d'onde, permettait une meilleure exploitation des données. Un très grand nombre de capteurs multispectraux se sont ainsi développés, le premier étant Landsat au début des années 70. Les progrès réalisés en matière de capteurs d'images ont conduit naturellement à la naissance de l'imagerie hyperspectrale : la scène

n'est plus observée à trois ou quatre longueurs d'onde mais dans plusieurs centaines de bandes spectrales étroites et contiguës. À chaque pixel d'une image hyperspectrale est alors associée un vecteur de mesures formant un spectre. On parle alors de « pavé » hyperspectral ou « hypercube » pour décrire l'ensemble des données échantillonnées spatialement et spectralement (voir par exemple la figure 3.1).

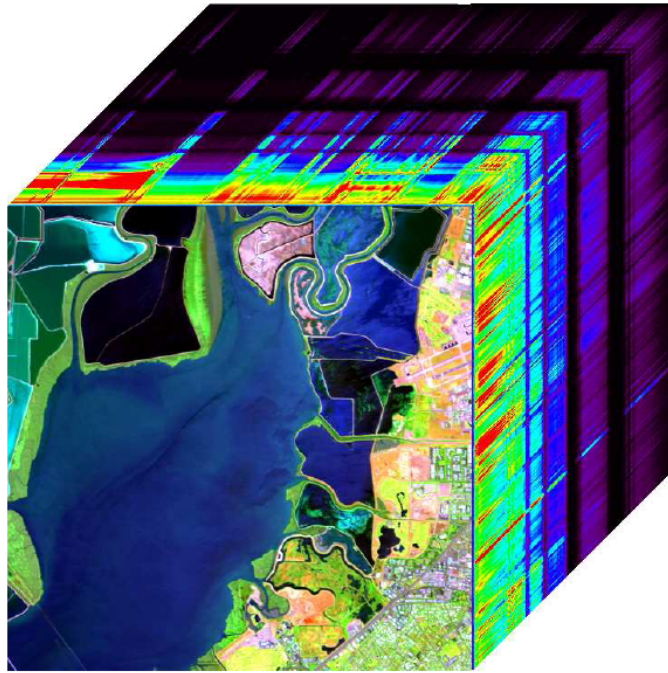


FIG. 3.1 – Un exemple d'image hyperspectrale.

Applications

L'imagerie hyperspectrale est utilisée dans des domaines divers, pour les applications civiles et militaires. Depuis plusieurs années, elle fait l'objet d'études avancées menées aux niveaux national et international par différents organismes de recherches tels l'ESA ou la NASA ou par de grands groupes industriels. On peut citer, parmi les imageurs hyperspectraux les plus performants, le capteur en 224 bandes AVIRIS (Airborne Visible/Infrared Imaging Spectrometer) du laboratoire Jet Propulsion de la NASA [GES⁺98], le capteur en 210 bandes HYDICE (HYperspectral Digital Imagery Collection

Experiment) du laboratoire Naval Research [BCA95] et le capteur HYMAP (HYperspectral MAPping) mis au point par l'entreprise australienne Integrated Spectronics [CJS+98].

L'exploitation des données fournies par le capteur permet d'effectuer une analyse d'une finesse jusqu'alors inégalée en télédétection. Chaque pixel, représenté par un vecteur de mesure, est en effet caractéristique des éléments macroscopiques qui le composent (figure 3.2).

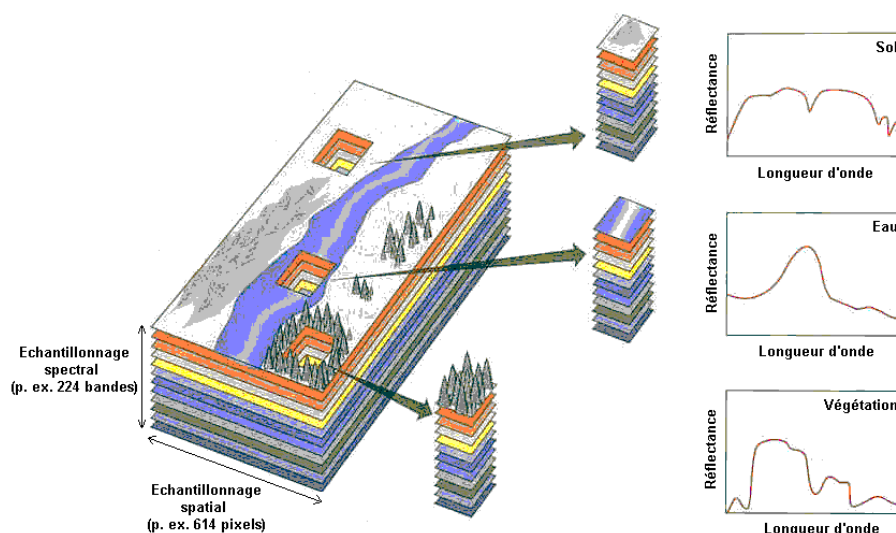


FIG. 3.2 – Illustration d'un cube hyperspectral dont 3 pixels ont été extraits.

Dans le cas d'analyses minéralogiques, il est possible par exemple de déterminer avec une grande précision les zones riches en un minéral particulier. Les raies d'absorptions de certains minéraux étant très fines, la résolution de l'outil hyperspectral est indispensable pour les discriminer. À partir des images hyperspectrales, on peut effectuer de grandes classifications de manière beaucoup plus rapide que par le biais des campagnes de prélèvements au sol (voir par exemple [JL02] ou [RDFZ04]). En matière d'agriculture ou de surveillance de l'environnement, on peut mesurer des paramètres macroscopiques qui permettent un contrôle efficace des plantations [LJdG01] ou des milieux naturels [DMN+03]. Notamment, les bandes d'absorptions de la chlorophylle et de l'eau liquide fournissent des informations sur la qualité de la photosynthèse et l'état hydrique des végétaux (figure 3.3).

Dans le domaine militaire, la détection de cible sub-pixel, c'est-à-dire dont la taille est inférieure

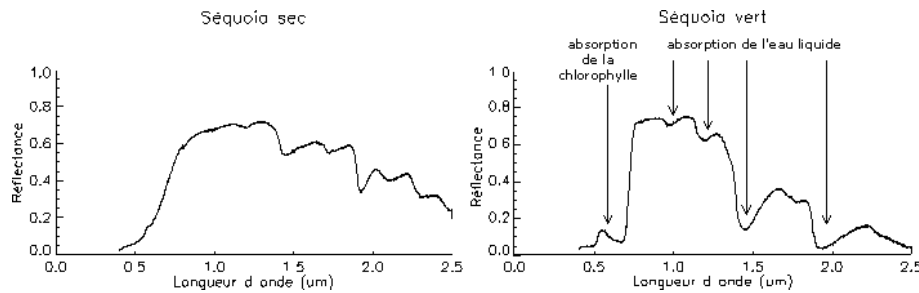


FIG. 3.3 – Spectres en réflectance du séquoia sec et vert.

à la résolution spatiale de l'imageur, a fait l'objet d'un grand nombre de travaux. Le lecteur intéressé est invité à se reporter à [C-03, part 2] et aux références qui y sont citées.

3.1.2 Analyse du problème

Le démélange spectral a largement été utilisé pour l'analyse d'images hyperspectrales [KM02]. Son hypothèse fondatrice est basé sur le fait que les vecteurs d'observation sont issus du mélange d'un certain nombre de constituants macroscopiques purs, appelés pôles de mélange, présents dans la scène observée. Suivant cette hypothèse, deux modèles de mélange différents ont été principalement étudiés dans le passé. Le premier est un mélange macrospectral qui décrit un pixel de l'image comme le mélange linéaire des pôles de mélange [SM79]. Le second modèle, introduit par Hapke [Hap81], fait référence à un mélange intrinsèque qui suppose une combinaison non linéaire de ces pôles de mélange. Cependant, comme il l'a été montré dans [JSTA83], ce mélange intrinsèque peut être linéarisé pour simplifier l'analyse. C'est pourquoi, nous nous sommes attaché dans le présent document à l'étude du démélange spectral linéaire des images hyperspectrales. Pour rendre cette analyse performante, trois points clés doivent être envisagés. Le premier est l'identification du nombre de composants purs supposés présents dans l'image. Le suivant consiste à estimer ces composants purs une fois leur nombre connu. Enfin, la troisième étape de l'analyse est la mise en œuvre d'algorithmes capables de réaliser le démélange linéaire (on parle alors d'étape d'inversion).

Alors qu'un grand nombre de travaux dans la littérature sont consacrés à l'étude des algorithmes d'inversion, les premières étapes de l'analyse ont largement été ignorées, notamment en supposant

qu'une information *a priori* concernant l'image est disponible. La plupart des techniques de démixage spectral développées actuellement dans la littérature le sont donc dans un cadre supervisé, c'est-à-dire qu'une information *a priori* est accessible concernant le nombre de pôles de mélange et leur nature respective. Le travail rapporté ici présente une approche de démixage spectral dans un contexte semi-supervisé. Elle permet de déterminer quels constituants purs ont participé au mélange spectral lorsqu'une bibliothèque spectrale est disponible. Pour certaines applications réelles, les pôles de mélange peuvent être obtenus directement grâce aux données elles-mêmes sans connaissance *a priori*. Dans ce cas, l'algorithme que nous proposons peut être associé à un algorithme d'extraction de pôles de mélange, comme l'algorithme N-finder (ou N-FINDR) développé par Winter [Win99].

Comme expliqué précédemment, l'étape d'inversion a reçu une attention considérable de la part de la communauté de la télédétection (voir [KM02] et les références citées). Le modèle de mélange linéaire est classiquement utilisé pour modéliser le spectre d'un pixel d'une scène observée. Ce modèle suppose que le spectre de ce pixel résulte du mélange linéaire des spectres des composants purs. Les coefficients de ce mélange linéaire sont appelés coefficients d'abondance ou bien abondances. Le problème d'inversion se résume alors à l'estimation de ces abondances à partir du pixel observé. Pour des raisons évidentes, les abondances vérifient des contraintes de positivité et d'additivité. Par conséquent, leur estimation nécessite des algorithmes de résolution sous contraintes d'égalité et d'inégalité. Plusieurs estimateurs, parmi lesquels les estimateurs des moindres carrés sous contraintes ou de variance minimum ont été proposés [HC01; TCC98].

Ce chapitre étudie un estimateur bayésien hiérarchique qui nous permet d'estimer les abondances dans le cadre d'un modèle de mélange linéaire. L'approche adoptée définit des lois *a priori* pour les paramètres du signal inconnus (les abondances et la variance du bruit) puis l'estimation de ces paramètres est réalisée à partir de leurs lois *a posteriori*. Comme dans les chapitres précédents, la complexité des lois *a posteriori* des paramètres inconnus impose le recours à des méthodes de simulations appropriées comme les méthodes de Monte Carlo par chaînes de Markov [GRS96].

Le problème de démixage spectral considéré ici peut être formulé comme un problème de régression linéaire sous contraintes. Or, les modèles bayésiens sont particulièrement bien appropriés pour ce genre de problèmes puisque les contraintes imposées peuvent guider naturellement le choix des lois

a priori. L'espace support de la loi *a posteriori* est alors un espace qui tient compte des contraintes. De nombreux exemples de problèmes sous contraintes ont récemment été étudiés dans la littérature. On pourra citer notamment les problèmes sous contraintes de monotonie, résolus par exemple par des lois *a priori* normales tronquées [CD96] ou les problèmes sous contraintes de non-négativité, faisant alors appel à des lois *a priori* Gamma [MBMC06] ou normales tronquées [RYDS02]. Il semble intéressant de remarquer également que des idées similaires ont récemment été exploitées afin d'étudier les modèles linéaires d'approximation parcimonieuse. Ainsi, la parcimonie peut être assurée en introduisant des lois *a priori* de Student [FG06] ou des lois de mélanges appropriées [BD07]. Nous nous intéressons dans le problème considéré à un modèle bayésien avec des lois *a priori* qui satisfont les contraintes de positivité et d'additivité inhérentes à l'imagerie hyperspectrale. À notre connaissance, c'est la première fois qu'un tel modèle est proposé.

Ce chapitre débute par la description du modèle de mélange linéaire qui caractérise les pixels observés en imagerie hyperspectrale (paragraphe 3.2). Le paragraphe 3.3 introduit le modèle bayésien hiérarchique utilisé pour résoudre le problème de démixage spectral. Un échantillonneur de Gibbs permettant de générer des échantillons suivant la loi d'intérêt est étudié au paragraphe 3.4. Des résultats de simulations sur des données synthétiques et réelles sont rapportés respectivement aux paragraphes 3.5 et 3.6. Nous envisageons le cas où les données sont corrompues par un bruit corrélé dans les bandes au paragraphe 3.7. Enfin, le paragraphe 3.8 traite le cas d'un démixage semi-supervisé pour lequel le nombre de composants dans le mélange ainsi que leur nature sont inconnus.

3.2 Modèle de mélange linéaire

Cette partie présente le modèle analytique utilisé pour réaliser le démixage spectral. Nous rappelons au lecteur que nous nous intéressons plus particulièrement au modèle linéaire, qui constitue une bonne approximation dans le domaine réflectif allant de $0.4\mu m$ à $2.5\mu m$ (voir [KM02], [MSS01] ou plus récemment [ND05]). Cependant, l'analyse proposée pourrait être étendue à des modèles de mélange non linéaires, par exemple en ayant recours à des approches basées sur des représentations par fonctions de base [DHMS02, p. 134]. Le modèle de mélange linéaire suppose que le L -spectre

$\mathbf{y} = [y_1, \dots, y_L]^\top$ d'un pixel mélangé est la combinaison linéaire de R spectres \mathbf{m}_r entaché d'un bruit blanc additif :

$$\mathbf{y} = \sum_{r=1}^R \mathbf{m}_r \alpha_r + \mathbf{n}, \quad (3.1)$$

où

- $\mathbf{m}_r = [m_{r,1}, \dots, m_{r,L}]^\top$ représente le spectre du r -ième matériau (appelé aussi pôle de mélange ou *endmember* en anglais),
- α_r est la proportion du r -ième matériau dans le pixel (ou r -ième abondance),
- R est le nombre de matériaux purs présents dans toute la scène observée,
- L est le nombre de bandes spectrales disponibles pour l'image,
- $\mathbf{n} = [n_1, \dots, n_L]^\top$ est une séquence de bruit blanc additif notée $\mathbf{n} \sim \mathcal{N}(\mathbf{0}_L, \sigma^2 \mathbf{I}_L)$, dont les réalisations n_l sont classiquement supposées indépendantes et identiquement distribuées suivant une loi normale centrée de variance σ^2 (\mathbf{I}_L est la matrice identité de dimension $L \times L$). Un modèle plus général faisant intervenir un bruit gaussien corrélé sera étudié au paragraphe 3.7.

Comme les coefficients α_r ($r = 1, \dots, R$) représentent des proportions (ou concentrations), le vecteur $\boldsymbol{\alpha}^+ = [\alpha_1, \dots, \alpha_R]^\top$ est soumis aux contraintes de positivité et d'additivité suivantes :

$$\begin{cases} \alpha_r \geq 0, \quad \forall r = 1, \dots, R, \\ \sum_{r=1}^R \alpha_r = 1. \end{cases} \quad (3.2)$$

Ainsi, le vecteur d'abondance peut-il être réécrit comme $\boldsymbol{\alpha}^+ = [\boldsymbol{\alpha}^\top, \alpha_R]^\top$ avec $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_{R-1}]^\top$ et $\alpha_R = 1 - \sum_{r=1}^{R-1} \alpha_r$. Par ailleurs, les R spectres des composants \mathbf{m}_r sont supposés connus dans la première partie de l'analyse. En conséquence, la méthodologie proposée doit être couplée avec une des nombreuses techniques d'estimation de ces spectres. Ces techniques sont soit basées sur des approches géométriques [Cra94; Win99] ou statistiques [SS97; BKL+04]. La seconde partie de ce chapitre proposera une extension de l'algorithme à des mélanges contenant un nombre inconnu de spectres appartenant à une bibliothèque connue.

3.3 Modèle bayésien hiérarchique

Cette partie fournit une description du modèle bayésien hiérarchique utilisé pour l'estimation du vecteur des paramètres inconnus sous les contraintes précisées par l'équation (3.2). Ces paramètres inconnus sont les abondances $\alpha_1, \dots, \alpha_R$ et la variance du bruit σ^2 . L'inférence bayésienne sur le vecteur des paramètres inconnus $\Theta = \{\boldsymbol{\alpha}, \sigma^2\}$ est basée sur la loi *a posteriori* $f(\Theta|\mathbf{y})$. Cette loi est liée à la vraisemblance des observations et aux lois *a priori* des paramètres par la relation de Bayes $f(\Theta|\mathbf{y}) \propto f(\mathbf{y}|\Theta) f(\Theta)$. Cette vraisemblance et ces lois *a priori* pour le problème de démixage envisagé sont détaillées dans les paragraphes qui suivent.

3.3.1 Vraisemblance

L'équation (3.1) montre que $\mathbf{y} \sim \mathcal{N}(\mathbf{M}^+ \boldsymbol{\alpha}^+, \sigma^2 \mathbf{I}_L)$, où $\mathbf{M}^+ = [\mathbf{m}_1, \dots, \mathbf{m}_R]$. Par conséquent, la vraisemblance du vecteur d'observations \mathbf{y} peut s'exprimer comme suit :

$$f(\mathbf{y}|\boldsymbol{\alpha}^+, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{L}{2}} \exp\left[-\frac{\|\mathbf{y} - \mathbf{M}^+ \boldsymbol{\alpha}^+\|^2}{2\sigma^2}\right], \quad (3.3)$$

où $\|\mathbf{x}\|^2 = \mathbf{x}^\top \mathbf{x}$ est la norme L^2 standard.

3.3.2 Lois *a priori* des paramètres

Les contraintes inhérentes au modèle de mélange linéaire (3.2) imposent que le vecteur $\boldsymbol{\alpha}$ appartienne au simplexe \mathbb{S} défini par :

$$\mathbb{S} = \left\{ \boldsymbol{\alpha} \left| \alpha_r \geq 0, \forall r = 1, \dots, R-1, \sum_{r=1}^{R-1} \alpha_r \leq 1 \right. \right\}. \quad (3.4)$$

Une loi uniforme sur \mathbb{S} est choisie comme loi *a priori* pour $\boldsymbol{\alpha}$ afin de traduire l'absence de connaissance *a priori* concernant ce vecteur inconnu. Notons que choisir une loi uniforme sur le simplexe pour $\boldsymbol{\alpha}$ est totalement équivalent à choisir une loi de Dirichlet $\mathcal{D}_R(1, \dots, 1)$ pour $\boldsymbol{\alpha}^+$ (voir [DHMS02, p. 237] pour la définition de la loi de Dirichlet $\mathcal{D}_R(1, \dots, 1)$).

Une loi non informative est choisie comme loi *a priori* pour σ^2 , c'est-à-dire une loi inverse Gamma avec les paramètres $\frac{\nu}{2}$ et $\frac{\gamma}{2}$:

$$\sigma^2 \left| \nu, \gamma \sim \mathcal{IG}\left(\frac{\nu}{2}, \frac{\gamma}{2}\right). \quad (3.5)$$

L'hyperparamètre ν sera fixé à $\nu = 2$ (comme dans [PADF02]) tandis que γ est un hyperparamètre ajustable.

3.3.3 Loi *a priori* de l'hyperparamètre

L'hyperparamètre associé à la loi *a priori* définie précédemment est γ . Comme nous l'avons précisé au paragraphe 1.4.3 puis illustré par des simulations aux paragraphes 1.7.6 et 2.4.7, la qualité de l'estimation dépend de la valeur de cet hyperparamètre. L'approche hiérarchique développée dans ce chapitre propose une loi non-informative de Jeffrey comme loi *a priori* de l'hyperparamètre γ :

$$f(\gamma) = \frac{1}{\gamma} \mathbf{1}_{\mathbb{R}^+}(\gamma), \quad (3.6)$$

où $\mathbf{1}_{\mathbb{R}^+}(\cdot)$ est la fonction indicatrice définie sur \mathbb{R}^+ .

3.3.4 Loi *a posteriori* de Θ

La loi *a posteriori* du vecteur de paramètres inconnus $\Theta = \{\alpha, \sigma^2\}$ peut être calculée suivant la structure hiérarchique suivante :

$$f(\Theta|\mathbf{y}) \propto \int f(\mathbf{y}|\Theta) f(\Theta|\gamma) f(\gamma) d\gamma, \quad (3.7)$$

où $f(\mathbf{y}|\Theta)$ et $f(\gamma)$ ont été définie par (3.3) et (3.6) respectivement. En supposant l'indépendance entre σ^2 et α , c'est-à-dire $f(\Theta|\gamma) = f(\alpha)f(\sigma^2|\nu, \gamma)$, l'hyperparamètre γ peut être intégré dans la loi jointe $f(\Theta, \gamma|\mathbf{y})$, permettant d'écrire :

$$f(\alpha, \sigma^2|\mathbf{y}) \propto \frac{1}{\sigma^{L+2}} \exp\left[-\frac{\|\mathbf{y} - \mathbf{M}^+ \alpha^+\|^2}{2\sigma^2}\right] \mathbf{1}_{\mathbb{S}}(\alpha), \quad (3.8)$$

où $\mathbf{1}_{\mathbb{S}}(\cdot)$ est la fonction indicatrice définie sur le simplexe \mathbb{S} . La troncature de la loi *a posteriori* à ce simplexe nous empêche d'obtenir une expression simple des estimateurs bayésiens. Le paragraphe suivant montre qu'une stratégie basée sur un échantillonneur de Gibbs permet de générer des échantillons distribués suivant la loi jointe $f(\alpha, \sigma^2|\mathbf{y})$. Ces échantillons pourront par la suite être utilisés pour réaliser les estimations souhaitées.

3.4 Échantillonneur de Gibbs pour l'estimation des abondances

Il est possible de générer des échantillons suivant $f(\boldsymbol{\alpha}, \sigma^2 | \mathbf{y})$ grâce à un échantillonneur de Gibbs dont les étapes principales sont détaillées dans les sections 3.4.1 et 3.4.2 (voir aussi l'algorithme 3.1).

– Initialisation :

- Échantillonner les paramètres $\sigma^{2(0)}$ et $\boldsymbol{\alpha}^{(0)}$ à partir des lois *a priori* définies à la section 3.3.2,
- Poser $t \leftarrow 1$,

– Itérations : pour $t = 1, 2, \dots$, faire

- Échantillonner $\boldsymbol{\alpha}^{(t)}$ selon la loi (3.11),
 - Échantillonner $\sigma^{2(t)}$ selon la loi (3.12),
 - Set $t \leftarrow t + 1$.
-

ALGO. 3.1: Échantillonneur de Gibbs pour le démélange spectral en imagerie hyperspectrale

3.4.1 Génération d'échantillons distribués suivant $f(\boldsymbol{\alpha} | \sigma^2, \mathbf{y})$

En notant $\mathbf{M} = [\mathbf{m}_1, \dots, \mathbf{m}_{R-1}]$, des calculs élémentaires détaillés dans l'annexe B conduisent au résultat suivant :

$$f(\boldsymbol{\alpha} | \sigma^2, \mathbf{y}) \propto \exp \left[-\frac{(\boldsymbol{\alpha} - \boldsymbol{\mu})^\top \boldsymbol{\Lambda}^{-1} (\boldsymbol{\alpha} - \boldsymbol{\mu})}{2} \right] \mathbf{1}_{\mathbb{S}}(\boldsymbol{\alpha}), \quad (3.9)$$

où

$$\begin{cases} \boldsymbol{\Lambda} &= \left[\frac{1}{\sigma^2} (\mathbf{M} - \mathbf{m}_R \mathbf{u}^\top)^\top (\mathbf{M} - \mathbf{m}_R \mathbf{u}^\top) \right]^{-1}, \\ \boldsymbol{\mu} &= \boldsymbol{\Lambda} \left[\frac{1}{\sigma^2} (\mathbf{M} - \mathbf{m}_R \mathbf{u}^\top)^\top (\mathbf{y} - \mathbf{m}_R) \right], \end{cases} \quad (3.10)$$

et $\mathbf{u} = [1, \dots, 1]^\top \in \mathbb{R}^{R-1}$. Par conséquent, $\boldsymbol{\alpha} | \sigma^2, \mathbf{y}$ est distribué suivant une loi multivariée normale tronquée au simplexe \mathbb{S} (définie dans l'annexe C) :

$$\boldsymbol{\alpha} | \sigma^2, \mathbf{y} \sim \mathcal{N}_{\mathbb{S}}(\boldsymbol{\mu}, \boldsymbol{\Lambda}). \quad (3.11)$$

La génération d'échantillons distribués suivant une loi multivariée normale tronquée peut s'effectuer simplement en utilisant une procédure d'acceptation/rejet décrite par l'algorithme 3.2, quand le nombre de composants purs dans l'image est relativement faible (comme dans l'exemple considéré).

Cependant, il est important de mentionner que des techniques de simulation plus efficaces doivent être envisagées pour des problèmes plus complexes (voir l'annexe D et [Rob95] pour plus de détails).

-
1. Générer $\alpha^* \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Lambda})$
 2. Test d'acceptation/rejet :
 - 2.1. si $\alpha^* \in \mathbb{S}$, poser $\alpha = \alpha^*$ (*accepter*),
 - 2.2. si $\alpha^* \notin \mathbb{S}$, aller à l'étape 1. (*rejeter*),
 3. poser $\alpha^+ = \left[\alpha^\top, 1 - \sum_{r=1}^{R-1} \alpha_r \right]^\top$.
-

ALGO. 3.2: Procédure d'acceptation/rejet pour la génération d'échantillons distribués suivant une loi multivariée normale tronquée

3.4.2 Génération d'échantillons distribués suivant $f(\sigma^2 | \boldsymbol{\alpha}, \mathbf{y})$

En regardant attentivement la loi jointe $f(\sigma^2, \boldsymbol{\alpha} | \mathbf{y})$, la loi conditionnelle de $\sigma^2 | \boldsymbol{\alpha}, \mathbf{y}$ apparaît clairement comme la loi inverse Gamma suivante :

$$\sigma^2 | \boldsymbol{\alpha}, \mathbf{y} \sim \text{IG} \left(\frac{L}{2}, \frac{\|\mathbf{y} - \mathbf{M}^+ \boldsymbol{\alpha}^+\|^2}{2} \right). \quad (3.12)$$

Remarque

Il est intéressant de remarquer que l'hyperparamètre γ n'apparaît pas dans le schéma de simulation présenté ci-dessus. L'intégration de ce paramètre dans la loi *a posteriori* conduit à un modèle similaire à celui obtenu lorsqu'une loi *a priori* de Jeffrey est choisie pour le paramètre σ^2 . En effet,

$$\begin{aligned} f(\sigma^2) &= \int f(\sigma^2 | \gamma) f(\gamma) d\gamma \\ &\propto \left(\frac{1}{\sigma^2} \right)^2 \int \exp \left(-\frac{\gamma}{2\sigma^2} \right) d\gamma \\ &\propto \frac{1}{\sigma^2}. \end{aligned} \quad (3.13)$$

Sans intégrer ce paramètre dans la loi jointe $f(\boldsymbol{\alpha}, \sigma^2, \gamma | \mathbf{y})$, la génération des échantillons σ^2 aurait été possible conditionnellement à γ . Il aurait été nécessaire d'échantillonner l'hyperparamètre γ selon sa loi *a posteriori*. La simulation selon $f(\sigma^2 | \boldsymbol{\alpha}, \mathbf{y})$ aurait alors été effectuée selon les étapes :

- génération d'échantillons suivant $\gamma | \sigma^2 \sim \mathcal{G}\left(\frac{\nu}{2}, \frac{1}{2\sigma^2}\right)$,
- génération d'échantillons suivant $\sigma^2 | \gamma, \boldsymbol{\alpha}, \mathbf{y} \sim \mathcal{IG}\left(\frac{\nu+L}{2}, \frac{\gamma + \|\mathbf{y} - \mathbf{M}^+ \boldsymbol{\alpha}^+\|^2}{2}\right)$.

Afin de limiter le temps d'exécution de la procédure de démélange (nous rappelons que cette procédure doit être appliquée sur chaque pixel de l'image), nous avons préféré ne pas échantillonner l'hyperparamètre γ . Des simulations conduites sur des mélanges synthétiques et réels montrent toutefois que les résultats sont sensiblement identiques pour les deux méthodes envisagées.

3.5 Résultats de simulations sur des données synthétiques

3.5.1 Estimation des abondances

La précision de la procédure d'estimation des abondances proposée précédemment est d'abord illustrée par le démélange d'un pixel synthétique issu de la combinaison de trois composants purs. Ces composants ont été extraits d'une bibliothèque spectrale fournie avec le logiciel ENVI [RSI03, p. 1035] et sont représentatifs d'un environnement urbain ou péri-urbain : béton de construction, herbe verte et terre grasse micacée. Les proportions de ces composants sont fixées comme suit : $\alpha_1 = 0.3$, $\alpha_2 = 0.6$ et $\alpha_3 = 0.1$. Les observations sont corrompues par un bruit gaussien additif de variance $\sigma^2 = 0.025$ (ce qui correspond à un rapport signal sur bruit d'environ $\text{RSB} = 15\text{dB}$). Les spectres des composants purs et le spectre bruité du pixel mélangé sont représentés sur la figure 3.4.

La figure 3.5 montre les lois *a posteriori* des coefficients d'abondance α_r ($r = 1, 2, 3$) obtenues pour $N_{\text{MC}} = 20000$ itérations (dont $N_{\text{bi}} = 100$ itérations de chauffage). Ces distributions sont clairement en accord avec les valeurs réelles des coefficients, i.e. $\boldsymbol{\alpha}^+ = [0.3, 0.6, 0.1]^T$. Pour comparaison, l'algorithme *fully constrained least-squares* (FCLS) détaillé dans [CJ01; HC01] a été exécuté N_{MC} fois sur des signaux similaires à celui présenté sur la figure 3.4 (en bas), avec bien sûr des réalisations de bruit différentes. Les histogrammes des N_{MC} estimations d'abondance par l'algorithme FCLS sont représentés sur la figure 3.5 (en pointillés). Ces histogrammes sont également en accord avec les lois *a posteriori* correspondantes estimées par l'algorithme bayésien hiérarchique proposé. Cependant, il est important de remarquer que les lois *a posteriori* tracées sur la figure 3.5 (trait continu) ont été obtenues à partir du spectre d'un pixel donné, tandis que l'algorithme FCLS doit être exécuté N_{MC}

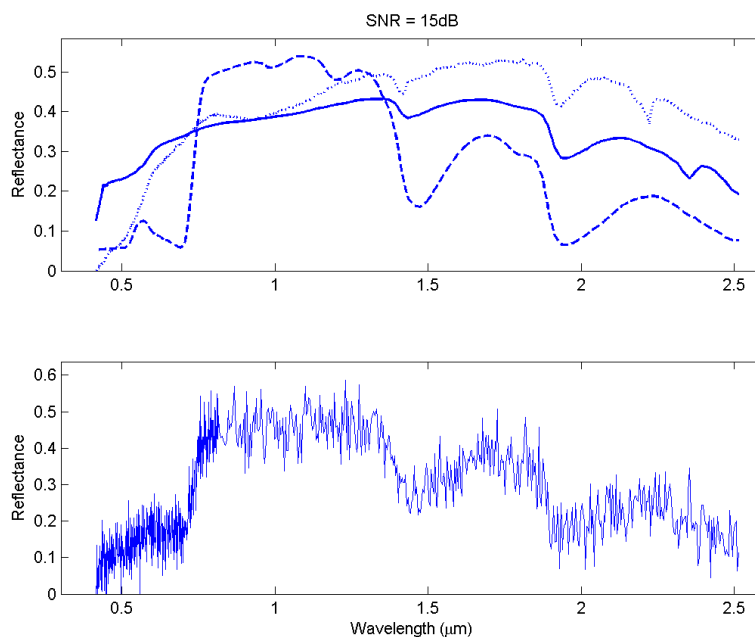


FIG. 3.4 – En haut - spectres des composants purs : béton de construction (trait plein), herbe verte (tirets), terre grasse micacée (pointillés). En bas - spectre résultant du pixel mélangé.

fois pour obtenir les histogrammes des coefficients d'abondances.

La figure 3.6 représente les estimateurs MAP des paramètres α_r et les écarts-types correspondant (calculés grâce à l'algorithme proposé) en fonction du rapport signal sur bruit. Cette figure permet d'évaluer les performances d'estimation pour un rapport signal-sur-bruit (RSB) donné. Il faut noter que les RSBs des spectro-imageurs réels comme AVIRIS ne sont pas inférieurs à 30dB lorsque les bandes d'absorption de la vapeur d'eau ont été enlevées [GES⁺98]. Par conséquent, les résultats de la figure 3.6 indiquent que l'algorithme proposé permet une estimation satisfaisante du vecteur des abondances.

3.5.2 Taux d'acceptation de l'échantillonneur

L'efficacité algorithmique de l'échantillonneur de Gibbs est gouvernée par le taux d'acceptation de la procédure d'acceptation/rejet dédiée à la génération d'échantillons distribués suivant la loi

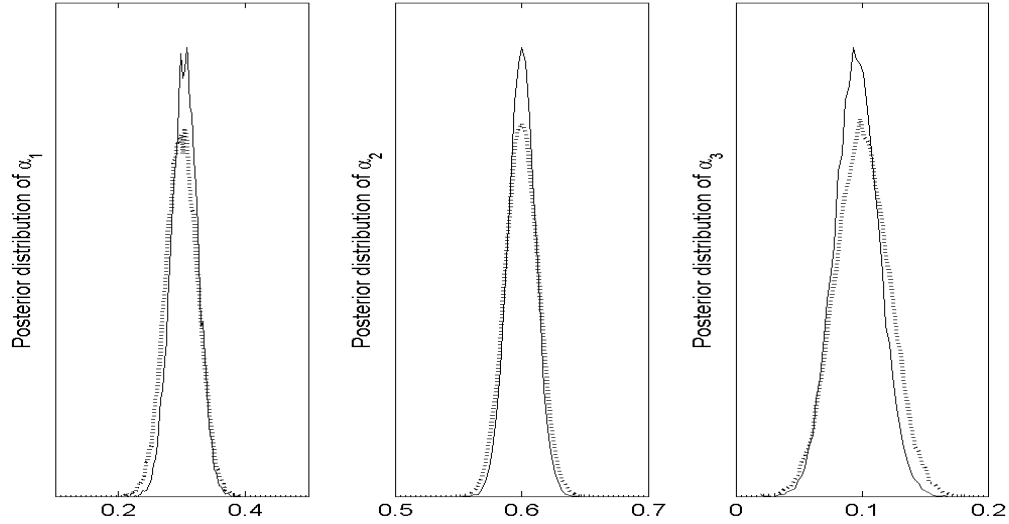


FIG. 3.5 – Lois *a posteriori* des abondances estimées $[\alpha_1, \alpha_2, \alpha_3]^\top$ (en trait plein) et histogramme des estimateurs FCLS (en pointillés).

multivariée normale tronquée. La probabilité d'accepter un échantillon généré est notée $P[\boldsymbol{\alpha} \in \mathbb{S}]$, où $\boldsymbol{\alpha} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Lambda})$ et $\boldsymbol{\mu}$ et $\boldsymbol{\Lambda}$ sont définies par (3.10). Des calculs élémentaires permettent d'obtenir le résultat suivant :

$$\begin{aligned}
 P[\boldsymbol{\alpha} \in \mathbb{S} | \boldsymbol{\alpha} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Lambda})] &= \int_{\mathbb{S}} \phi(\boldsymbol{\alpha} | \boldsymbol{\mu}, \boldsymbol{\Lambda}) d\boldsymbol{\alpha} \\
 &= \int_0^1 \int_0^{1-\alpha_1} \int_0^{1-\alpha_1-\alpha_2} \dots \int_0^{1-\sum_{r=1}^{R-2} \alpha_r} \phi(\boldsymbol{\alpha} | \boldsymbol{\mu}, \boldsymbol{\Lambda}) d\alpha_{R-1} d\alpha_{R-2} \dots d\alpha_1,
 \end{aligned} \tag{3.14}$$

où ϕ est la densité de probabilité de la loi normale définie dans l'annexe C. La figure 3.7 compare le taux d'acceptation théorique $P[\boldsymbol{\alpha} \in \mathbb{S}]$ à celui obtenu par simulation de 5000 variables gaussiennes. Ces résultats sont représentés en fonction du RSB pour une valeur donnée de $\boldsymbol{\alpha} = [0.3, 0.6, 0.1]^\top$. Cette figure montre que le taux d'acceptation $P[\boldsymbol{\alpha} \in \mathbb{S} | \boldsymbol{\alpha} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Lambda})]$ est une fonction croissante du RSB. Elle montre aussi que ce taux est très satisfaisant pour des valeurs de RSB fréquemment rencontrées en imagerie hyperspectrale (RSB > 30dB). Il faut noter que nous n'avons pas rencontré de problème de temps de calcul lors des simulations réalisées, puisque le nombre de pôles de mélange présents dans l'image est relativement faible. Cependant, pour des problèmes mettant en jeu un grand

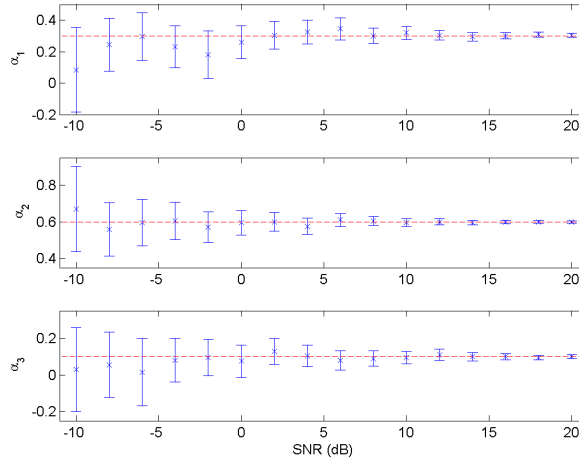


FIG. 3.6 – Estimateurs MAP (croix) et écarts-types (barres verticales) des paramètres α_r ($r = 1, \dots, 3$) en fonction du RSB.

nombre de composants purs, ou lorsque les valeurs de RSB sont beaucoup plus faibles, l'algorithme d'acceptation/rejet peut se montrer très vite inefficace. Dans ce cas, un échantillonneur de Gibbs adéquat peut s'avérer nécessaire pour générer efficacement des échantillons distribués suivant une loi multivariée normale tronquée (voir l'annexe D et [Rob95] pour les détails concernant la mise en œuvre d'une telle stratégie).

3.5.3 Diagnostic de convergence

L'échantillonneur de Gibbs proposé au paragraphe 3.4 permet de générer des échantillons $(\boldsymbol{\alpha}^{(t)}, \sigma^{2(t)})$ asymptotiquement distribués suivant $f(\boldsymbol{\alpha}, \sigma^2 | \mathbf{y})$. Le vecteur d'abondance peut alors être estimé grâce à la moyenne empirique suivante :

$$\hat{\boldsymbol{\alpha}}_{\text{MMSE}} = \frac{1}{N_r} \sum_{t=1}^{N_r} \boldsymbol{\alpha}^{(N_{\text{bi}}+t)}, \quad (3.15)$$

où N_{bi} et N_r sont les nombres d'itérations de chauffage et d'intérêt respectivement. Comme dans les chapitres précédents, il convient de répondre à deux questions pour assurer la qualité des estimations effectuées. Tout d'abord, à partir de quelle valeur de N_{bi} peut-on décider que les échantillons

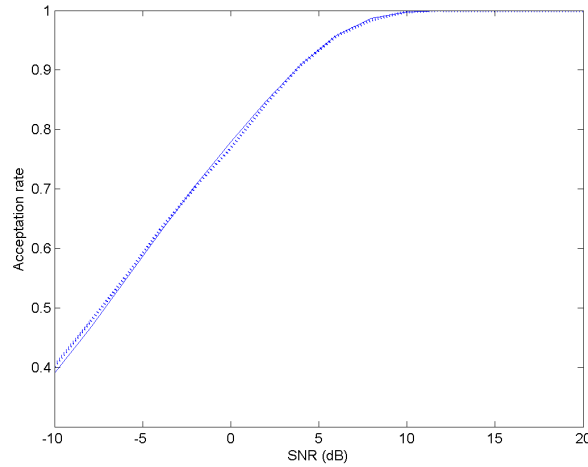


FIG. 3.7 – Taux d’acceptation théorique (trait plein) et expérimental (pointillé) du test d’acceptation/rejet en fonction du RSB.

$(\boldsymbol{\alpha}^{(t)}, \sigma^{2(t)})$ sont réellement distribués suivant la loi cible $f(\boldsymbol{\alpha}, \sigma^2 | \mathbf{y})$? Puis, combien d’itérations N_r sont nécessaires pour calculer la moyenne empirique (3.15) ?

Comme dans les chapitres précédents, la convergence de l’échantillonneur est assurée par le calcul du potentiel d’échelle :

$$\sqrt{\hat{\rho}(\kappa)} = \sqrt{\frac{1}{W(\kappa)} \left(\frac{N_r - 1}{N_r} W(\kappa) + \frac{1}{N_r} B(\kappa) \right)}. \quad (3.16)$$

où B et W sont respectivement les variances inter-séquence et intra-séquence définies dans (1.17) et (1.18). Différents choix pour le paramètre κ peuvent être envisagés pour la procédure de démélange spectral. Nous proposons de mesurer la convergence de l’algorithme de Gibbs à l’aide de la variance de bruit σ^2 (voir [GR98] pour un choix identique). Pour illustration, les échantillons de σ^2 générés par $M = 5$ chaînes de Markov sont représentés sur la figure 3.8. Ces chaînes convergent clairement vers une même valeur qui est en accord avec la valeur réelle de la variance $\sigma^2 = 0.025$. Le paramètre d’échelle pour ce paramètre σ^2 calculé à partir de $M = 10$ chaînes de Markov est égal à 0.9996. Cette valeur de $\sqrt{\hat{\rho}}$ confirme la bonne convergence de l’échantillonneur (nous rappelons qu’une valeur recommandée pour assurer la convergence est $\sqrt{\hat{\rho}} \leq 1.2$ [GCSR95, p. 332]).

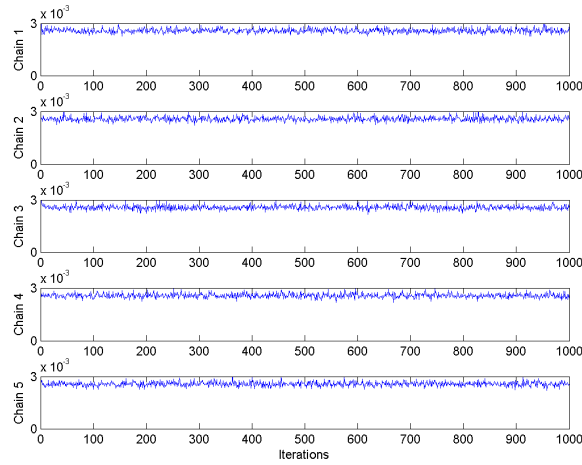


FIG. 3.8 – Vérification de la convergence avec cinq réalisations de chaînes de Markov.

Le nombre d'itérations N_r nécessaire pour calculer un estimateur efficace de α selon (3.15) est fixé en évaluant l'erreur quadratique entre une estimation de référence $\tilde{\alpha}$ (obtenu avec $N_r = 10000$) et une estimation après $N_r = p$ itérations. La figure 3.9 représente cette erreur quadratique en fonction du nombre d'itérations p (le nombre d'itérations de chauffage est $N_{bi} = 100$). Cette figure indique qu'un nombre d'itérations égal à $N_r = 500$ est suffisant pour assurer une estimation correcte de la moyenne empirique (3.15) pour cet exemple. Il faut noter que pour de telles valeurs de N_r et N_{bi} , réaliser le démélange du pixel considéré dans ce paragraphe prend environ 0.3 secondes pour une implémentation MATLAB sur un Pentium IV 2.8GHz.

3.6 Démélange spectral d'une image AVIRIS

Afin d'évaluer la performance de l'algorithme proposé sur des données réelles, cette partie présente l'analyse d'une image hyperspectrale qui a très souvent été utilisée dans les littératures de la télédétection [RDFZ04; CLM05; Che05] et du traitement d'image [TP04; AAM05]. Cette image, représentée sur la figure 3.10, est composée de 224 bandes spectrales de largeur de bande $10nm$. Elle a été acquise en 1997 par le spectromètre AVIRIS (Airborne Visible Infrared Imaging Spectrometer) au dessus de Moffett Field, à la pointe sud de la baie de San Francisco, en Californie (voir [AVI06] pour plus de

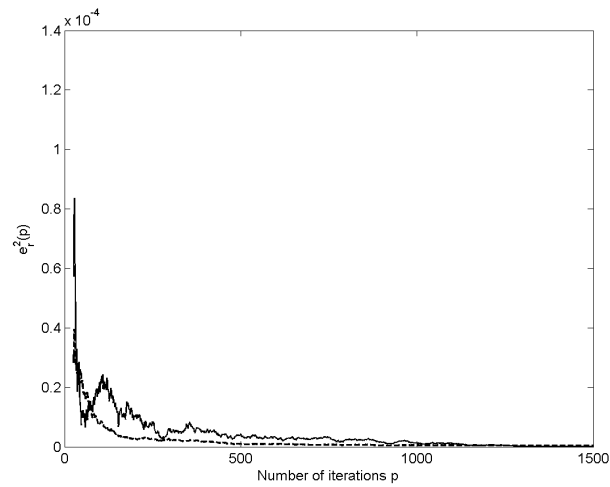


FIG. 3.9 – Erreur quadratique entre l'estimateur de référence et l'estimateur après p itérations en fonction de p (trait plein). Erreur quadratique moyennée sur 10 chaînes (tirets) ($N_{\text{bi}} = 100$).

détails). Elle comporte un large point d'eau (une partie d'un lac qui apparaît en foncé en haut de l'image) et une zone côtière composée de végétation et de sol nu.

Ce jeu de donnée a été réduit à $L = 189$ bandes spectrales en supprimant les bandes d'absorption de la vapeur d'eau. Une partie de cette image de taille 50×50 , représentée sur la figure 3.10 en échelle de gris à la longueur d'onde $\lambda = 0.66\mu\text{m}$ (bande 30), a été traitée par l'algorithme de démixage spectral proposé.

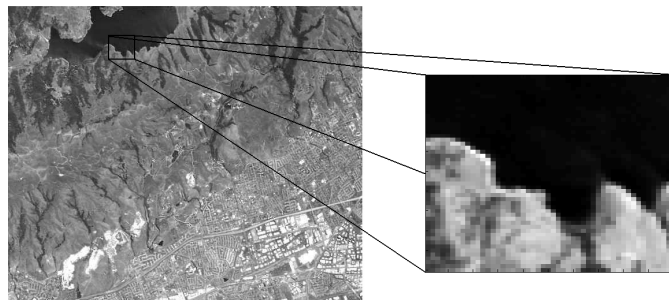


FIG. 3.10 – Données hyperspectrales réelles : l'image de Moffett Field acquise par AVIRIS en 1997 (à gauche) et la région d'intérêt à la longueur d'onde $\lambda = 0.66\mu\text{m}$ représentée en échelle de gris (à droite).

3.6.1 Détermination des pôles de mélange

La première étape de l'analyse permet d'identifier les matériaux purs présents dans la scène. On peut noter qu'une connaissance préliminaire de la géologie du terrain aurait permis d'utiliser une méthode d'extraction de pôles de mélange supervisée (par exemple en moyennant les pixels choisis sur des zones d'intérêt particulières). De telles informations n'étant pas disponibles, un algorithme totalement automatique a été utilisé. Cette procédure comprend une analyse en composantes principales (ACP) qui permet de réduire l'espace de travail et de connaître le nombre de pôles de mélange présents dans la scène comme expliqué dans [KM02]. Après avoir calculé la somme cumulée des valeurs propres, les données sont projetées sur les deux principaux axes associés aux deux valeurs propres les plus grandes. Les sommets du simplexe ainsi obtenus sont déterminés grâce à l'algorithme N-FINDR [Win99]. Les $R = 3$ pôles de mélanges estimés qui correspondent à la végétation, l'eau, et le sol nu, sont représentés sur la figure 3.11. Notons que d'autres méthodes d'extraction ont récemment été étudiées dans la littérature (voir par exemple [ND05] ou [CWL+06]).

3.6.2 Estimation des abondances

L'algorithme bayésien de démélange présenté aux paragraphes 3.3 et 3.4 a été appliqué à chaque pixel de l'image hyperspectrale en utilisant les pôles de mélange estimés au paragraphe 3.6.1. Plusieurs diagnostics de convergence ont montré qu'une période de chauffage très courte suffit pour cet exemple. Ceci est confirmé sur la figure 3.12 (en bas) qui illustre les chaînes de Markov des coefficients d'abondance pour un pixel donné. Par conséquent, la période de chauffage a été fixée à $N_{bi} = 10$ pour tous les résultats présentés dans cette partie. Les abondances sont estimées en moyennant les $N_r = 900$ derniers échantillons simulés pour chaque pixel, selon l'estimateur (3.15). Les lois *a posteriori* des abondance α_r ($r = 1, 2, 3$) sont représentées sur la figure 3.12 (en haut) pour le pixel $\#(43, 35)$. Ces lois *a posteriori* indiquent que ce pixel est composé de sol nu essentiellement, ce qui reflète bien que ce pixel est situé sur une zone côtière contenant très peu de végétation.

Les cartes d'abondance estimées par l'algorithme proposé pour les $R = 3$ matériaux purs sont représentés sur la figure 3.13 (en haut). Notons qu'un pixel blanc (respectivement noir) sur la carte

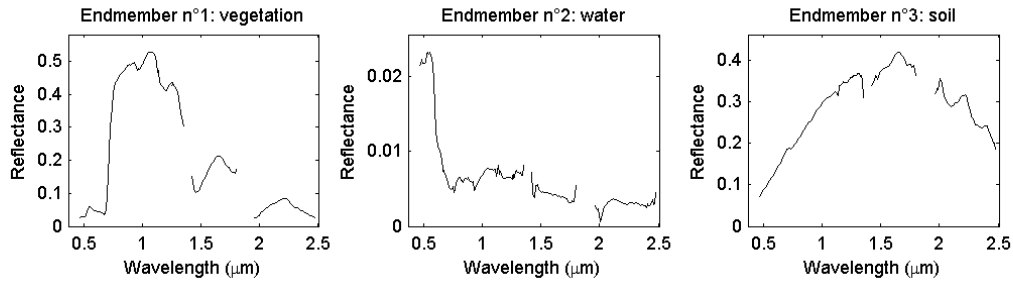


FIG. 3.11 – Les $R = 3$ spectres des composants purs obtenus par l’algorithme N-FINDR.

indique une forte (respectivement faible) valeur des coefficients d’abondance. La zone de lac (représentée par les pixels blancs sur la carte d’abondance de l’eau et en noir sur les autres cartes) est clairement retrouvée. Pour comparaison, les résultats obtenus par l’algorithme déterministe disponible avec le logiciel ENVI [RSI03, p. 739] sont représentés sur la figure 3.13 (en bas). Ces figures obtenues avec un algorithme de moindres carrés sous contraintes d’additivité et de positivité sont clairement en accord avec celles obtenues par l’approche proposée de la figure 3.13 (en haut). Cependant, il est important d’ajouter que l’algorithme bayésien hiérarchique permet d’estimer la loi *a posteriori* des coefficients d’abondances et de la variance de bruit. Cette loi peut ensuite être utilisée pour calculer des intervalles de confiance ou d’autres mesures de confiance associées aux estimations des abondances.

3.6.3 Convergence de l’échantillonneur

Comme expliqué dans la partie 1.6, la convergence de l’échantillonneur peut être vérifiée en surveillant des paramètres clés comme par exemple σ^2 . Les échantillons de 5 différentes chaînes de Markov pour le paramètre σ^2 sont représentés sur la figure 3.14 pour le pixel $\#(43, 35)$. Toutes les chaînes convergent clairement vers une même valeur. Le facteur d’échelle associé à la variance σ^2 est calculé à partir de $M = 10$ chaînes de Markov pour chaque pixel. Ces valeurs de $\sqrt{\hat{\rho}}$ sont représentés sur la figure 3.15. Elles sont toutes inférieures à 1.0028 (valeur obtenue pour le pixel $\#(10, 26)$), ce qui indique une bonne convergence de l’échantillonneur pour chaque pixel.

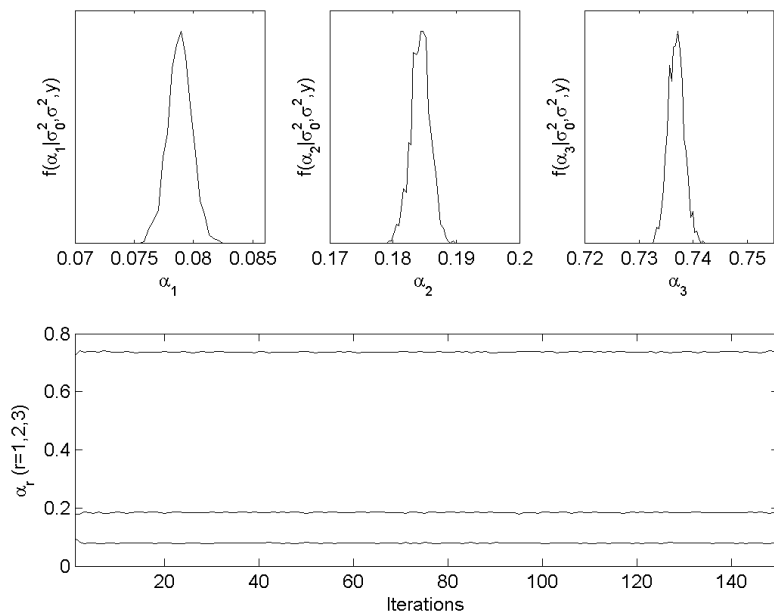


FIG. 3.12 – En haut : lois *a posteriori* des abondance α_r ($r = 1, \dots, 3$) pour le pixel $\#(43, 35)$. En bas : 150 premières échantillons simulés par l'algorithme de Gibbs.

3.6.4 Sensibilité à l'étape d'extraction de pôles de mélange

Afin d'étudier la sensibilité de l'algorithme à l'étape d'extraction de pôles de mélange, nous estimons les pôles de mélanges de l'image réelle précédente comme suit. Tout d'abord, l'image de Moffett Field a été segmentée par un algorithme K-means non supervisé initialisé avec 3 classes. Les résultats de cette segmentation sont représentés sur la figure 3.16 et montrent que les classes « végétation, » « eau » et « sol » sont clairement retrouvées.

Les pixels les plus « purs » appartenant à chaque classe sont alors identifiés par leur index de pureté respectif [BKG95] calculé après 15000 itérations. Leurs spectres sont alors moyennés pour définir les trois spectres des composants purs. Ceux-ci sont représentés sur la figure 3.17 (en haut, en trait plein). Pour rappel, ces mêmes spectres estimés par l'algorithme N-FINDR sont représentés sur cette même figure (en pointillés). On peut remarquer que les cartes d'abondance, tracées sur la figure 3.17 (en bas), sont semblables à celles obtenues avec la procédure N-FINDR.

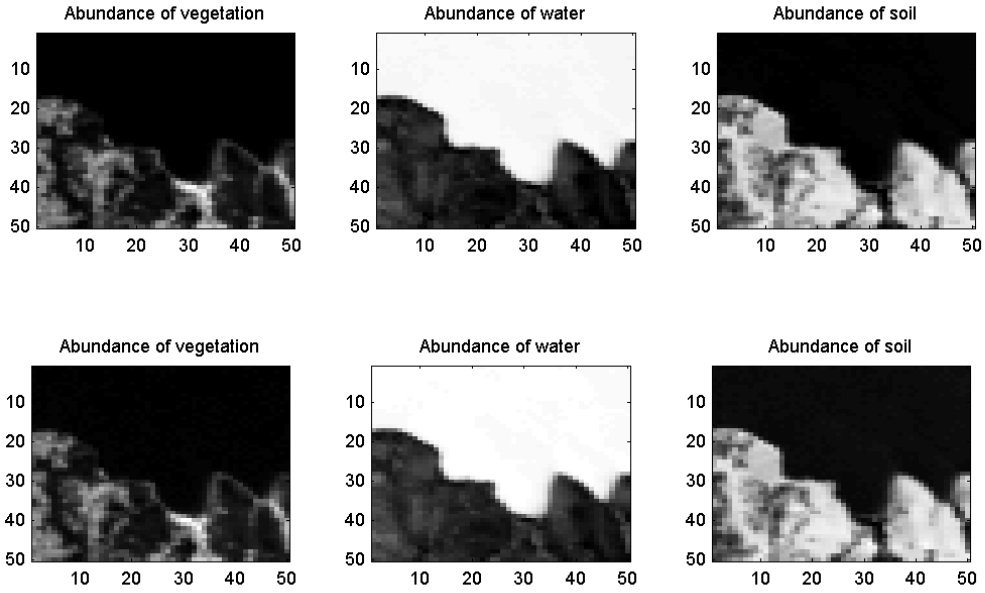


FIG. 3.13 – En haut : cartes d'abondances estimées par l'algorithme proposé. En bas : cartes d'abondances estimées par le logiciel ENVI.

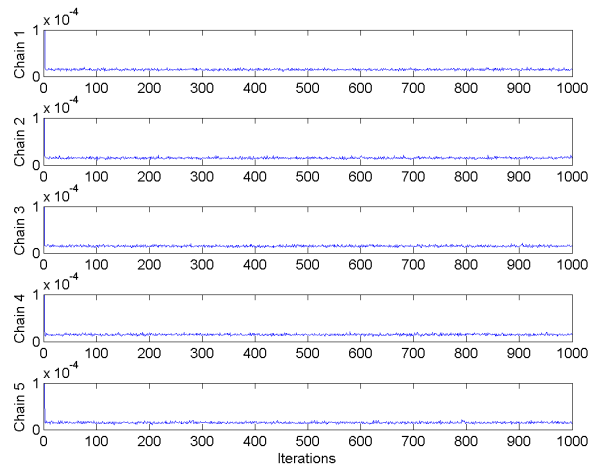


FIG. 3.14 – Diagnostic de convergence : exemples de $M = 5$ chaînes parallèles relatives au paramètre σ^2 .

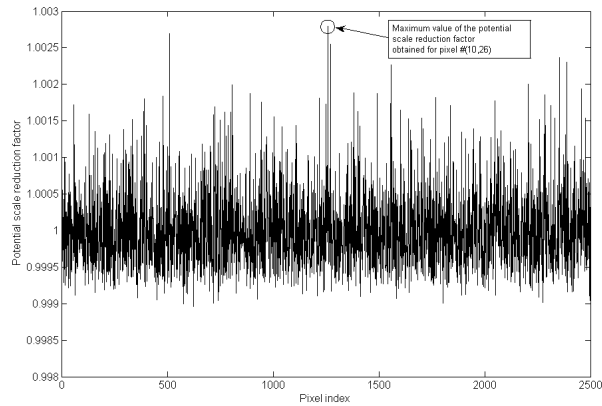


FIG. 3.15 – Facteurs d'échelle calculés pour chaque pixel de l'image de Moffett Field.

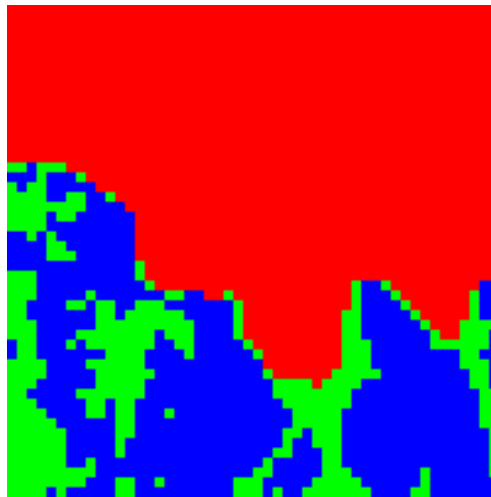


FIG. 3.16 – Image de Moffett Field segmentée à l'aide de l'algorithme K-means.

3.7 Cas d'un bruit gaussien corrélé

Nous avons présenté au paragraphe précédent un modèle de mélange linéaire entaché d'un bruit blanc gaussien centré. Nous présentons dans ce qui suit une adaptation du modèle précédent au cas où la séquence de bruit possède une corrélation inconnue.

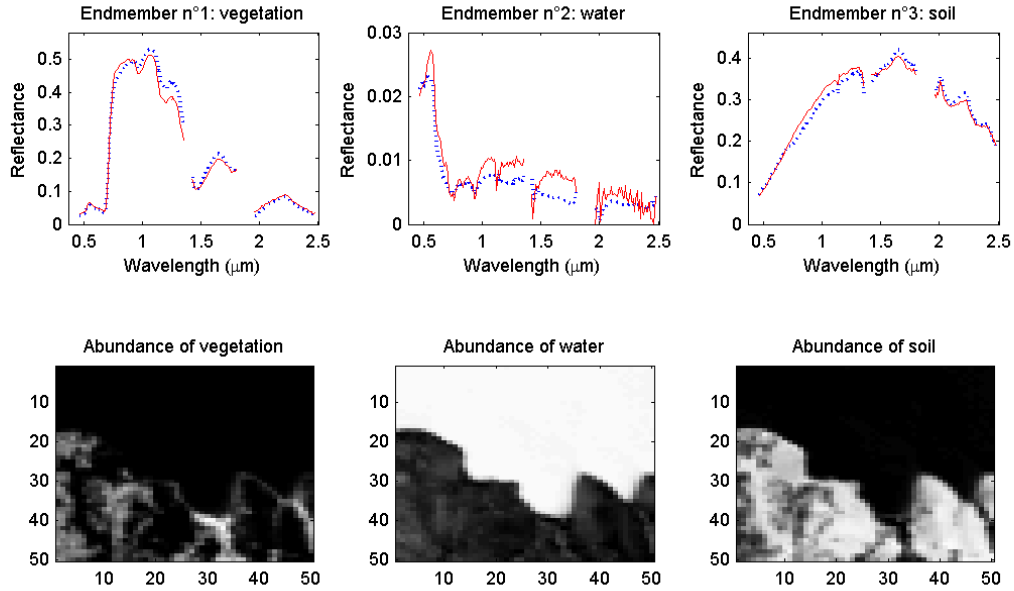


FIG. 3.17 – En haut : les $R = 3$ pôles de mélange obtenue par la procédure K-means/PPI (trait plein) et par l’algorithme du N-FINDR (pointillés). En bas : les cartes d’abondance des composants purs estimés par la procédure K-means/PPI.

3.7.1 Modèle bayésien hiérarchique

Nous reprenons le modèle de mélange linéaire décrit au paragraphe 3.2 :

$$\mathbf{y} = \sum_{r=1}^R \mathbf{m}_r \alpha_r + \mathbf{n}. \quad (3.17)$$

Le vecteur $\mathbf{n} = [n_1, \dots, n_L]^T$ est maintenant une séquence de bruit supposé gaussien centré et de matrice de covariance Σ de dimension $L \times L$. La fonction de vraisemblance du pixel observé dans L bandes devient :

$$f(\mathbf{y} | \boldsymbol{\alpha}^+, \Sigma^{-1}) = (2\pi)^{-\frac{L}{2}} |\Sigma^{-1}|^{\frac{1}{2}} \exp \left[-\frac{1}{2} \|\mathbf{y} - \mathbf{M}^+ \boldsymbol{\alpha}^+\|_{\Sigma^{-1}}^2 \right], \quad (3.18)$$

où

– $|\mathbf{A}|$ dénote le déterminant de la matrice \mathbf{A} ,

– $\|\cdot\|_{\mathbf{P}}$ est la norme euclidienne associée au \mathbf{P} -produit scalaire :

$$\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{P}} = \mathbf{x}^{\top} \mathbf{P} \mathbf{y}, \quad (3.19)$$

où \mathbf{P} est une matrice hermitienne $L \times L$ définie positive.

Il convient de définir la loi *a priori* de la matrice Σ . Le choix de cette loi *a priori* est guidée par la forme de la fonction de vraisemblance : il est judicieux, dans un but de simplification des calculs d'opter pour une loi *a priori* conjuguée inverse Wishart. Pour des raisons d'écriture, de manière totalement équivalente, nous choisirons une loi de Wishart comme loi *a priori* de l'inverse de la matrice de covariance Σ^{-1} :

$$\Sigma^{-1} | \nu, \mathbf{D} \sim \mathcal{W} \left(\nu, \frac{1}{\nu} \mathbf{D} \right), \quad (3.20)$$

où ν est un hyperparamètre fixé. La matrice \mathbf{D} sera supposée diagonale, proportionnelle à l'identité : $\mathbf{D} = \frac{1}{\gamma} \mathbf{I}_L$, où γ est, comme au paragraphe 3.3.2, un hyperparamètre ajustable. Nous justifions la forme particulière de la matrice \mathbf{D} par le fait qu'il n'est pas restrictif, dans le cas de l'imagerie hyperspectrale, de supposer les éléments non diagonaux de \mathbf{D} très faibles par rapport aux éléments diagonaux de cette même matrice (voir les propriétés de la loi de Wishart, rappelées dans l'annexe E).

La loi *a priori* de l'hyperparamètre γ est, comme précédemment, une loi de Jeffrey :

$$f(\gamma) = \frac{1}{\gamma} \mathbf{1}_{\mathbb{R}^+}(\gamma). \quad (3.21)$$

Si une loi uniforme sur le simplexe \mathbb{S} est de nouveau choisie comme loi *a priori* du vecteur d'abondances α , et en supposant l'indépendance entre α et Σ , la loi *a posteriori* de $f(\alpha, \Sigma, \gamma | \mathbf{y})$ s'écrit :

$$\begin{aligned} f(\Theta, \gamma | \mathbf{y}) &= |\Sigma^{-1}|^{\frac{\nu-L}{2}} \exp \left[-\frac{1}{2} \|\mathbf{y} - \mathbf{M}^+ \alpha^+\|_{\Sigma^{-1}}^2 \right] \\ &\times \text{etr} \left\{ -\frac{\gamma \nu}{2} \Sigma^{-1} \right\} \gamma^{\frac{L\nu}{2}-1} \mathbf{1}_{\mathbb{S}}(\alpha) \mathbf{1}_{\mathbb{R}^+}(\gamma) \end{aligned} \quad (3.22)$$

où $\text{etr}\{\mathbf{A}\}$ est l'exponentielle de la trace de la matrice \mathbf{A} . Nous proposons dans le paragraphe qui suit un échantillonneur de Gibbs qui permet de générer des échantillons distribués suivant cette loi *a posteriori* grâce aux lois conditionnelles des paramètres inconnus.

3.7.2 Algorithme de Gibbs

L'échantillonneur de Gibbs qui permet de générer des échantillons suivant la loi *a posteriori* d'intérêt (3.22) comporte trois étapes décrites ci-dessous. Le schéma de simulation est détaillé dans l'algorithme 3.3.

– Initialisation :

- Echantillonner l'hyperparamètre $\gamma^{(0)}$ d'après la loi (3.21),
- Echantillonner la matrice de covariance $\Sigma^{(0)}$ à partir de la loi *a priori* (3.20),
- Echantillonner le paramètre $\alpha^{(0)}$ à partir de la loi *a priori* définie à la section (3.3.2),
- Poser $t \leftarrow 1$,

– Itérations : pour $t = 1, 2, 3, \dots$, faire

- Echantillonner le paramètre $\alpha^{(t)}$ selon la loi (3.24),
 - Echantillonner l'hyperparamètre $\gamma^{(t)}$ selon la loi (3.26),
 - Echantillonner le paramètre $\Sigma^{(t)}$ selon la loi (3.27),
-

ALGO. 3.3: Echantillonneur de Gibbs pour le démélange spectral avec bruit corrélé

Génération d'échantillons distribués suivant $f(\alpha | \Sigma^{-1}, \mathbf{y})$

En notant $\mathbf{M} = [\mathbf{m}_1, \dots, \mathbf{m}_{R-1}]$, des calculs similaires à ceux présentés dans l'annexe B permettent d'écrire :

$$f(\alpha | \Sigma^{-1}, \mathbf{y}) \propto \exp \left[-\frac{(\alpha - \mu)^\top \Lambda^{-1} (\alpha - \mu)}{2} \right] \mathbf{1}_{\mathbb{S}}(\alpha), \quad (3.23)$$

où cette fois,

$$\begin{cases} \Lambda = \left[(\mathbf{M} - \mathbf{m}_R \mathbf{u}^\top)^\top \Sigma^{-1} (\mathbf{M} - \mathbf{m}_R \mathbf{u}^\top) \right]^{-1}, \\ \mu = \Lambda \left[(\mathbf{M} - \mathbf{m}_R \mathbf{u}^\top)^\top \Sigma^{-1} (\mathbf{y} - \mathbf{m}_R) \right], \end{cases} \quad (3.24)$$

avec, pour rappel, $\mathbf{u} = [1, \dots, 1]^\top \in \mathbb{R}^{R-1}$. Par conséquent, $\alpha | \Sigma^{-1}, \mathbf{y}$ est distribué suivant la loi normale tronquée suivante :

$$\alpha | \Sigma^{-1}, \mathbf{y} \sim \mathcal{N}_{\mathbb{S}}(\mu, \Lambda). \quad (3.25)$$

Génération d'échantillons distribués suivant $f(\gamma|\Sigma^{-1}, \alpha, \mathbf{y})$

La loi *a posteriori* de l'hyperparamètre γ est :

$$\gamma|\Sigma^{-1}, \mathbf{y} \sim \mathcal{G}\left(\frac{\nu L}{2}, \frac{\nu}{2} \text{tr}\{\Sigma^{-1}\}\right), \quad (3.26)$$

où $\mathcal{G}(a, b)$ est la loi Gamma de paramètres a et b et $\text{tr}\{\cdot\}$ est l'opérateur trace.

Génération d'échantillons distribués suivant $f(\Sigma^{-1}|\gamma, \alpha, \mathbf{y})$

Des calculs détaillés dans l'annexe F conduisent à la loi *a posteriori* de Wishart suivante :

$$\Sigma^{-1}|\gamma, \mathbf{y} \sim \mathcal{W}\left(\nu + 1, \left[\gamma\nu\mathbf{I}_L + \mathbf{z}\mathbf{z}^\top\right]^{-1}\right), \quad (3.27)$$

où $\mathbf{z} = \mathbf{y} - \mathbf{M}^+\alpha^+$.

3.7.3 Simulations

Afin d'illustrer l'intérêt de notre approche, un pixel résultant de la combinaison de trois composants purs est démêlé par le nouvel algorithme proposé. Ces composants sont, comme au paragraphe 3.5, extraits de la bibliothèque spectrale du logiciel ENVI [RSI03, p.1035] : herbe verte, brique rouge et métal galvanisé. Les proportions du mélange de ces composants sont respectivement $\alpha_1 = 0.05$, $\alpha_2 = 0.6$ et $\alpha_3 = 0.35$. Les observations ont été altérées par un bruit gaussien centré de matrice de covariance Σ . Cette matrice $L \times L$ définie positive a été générée aléatoirement selon la loi de Wishart $\mathcal{W}(\nu, \frac{1}{\nu}\mathbf{D})$ avec $\nu = \eta L$ et $\mathbf{D} = \frac{1}{\gamma}\mathbf{I}_L$. Pour comprendre la signification physique des paramètres η et γ , considérons le cas limite $\eta \gg 1$. Pour des grandes valeurs du paramètre η , les propriétés de la loi de Wishart donnée dans l'annexe E nous permettent d'écrire $\mathbb{E}[\Sigma^{-1}] = \mathbf{D}$ et $\text{var}[\Sigma^{-1}] \approx \mathbf{O}_L$, où \mathbf{O}_L est la matrice nulle de taille $L \times L$. Dans ce cas, le vecteur bruit \mathbf{n} se ramène à une séquence d'échantillons indépendants identiquement distribués suivant une loi normale centrée de variance γ . Les 50×50 premiers éléments de la matrice de covariance Σ sont représentés en échelle de gris sur la figure 3.18 pour différentes valeurs du paramètres η . Bien sûr, les coefficients les plus

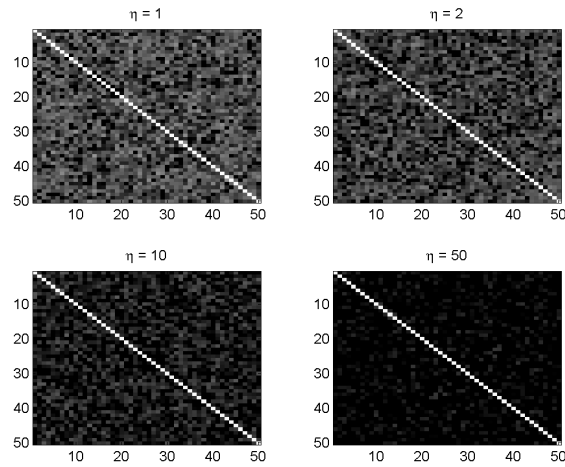


FIG. 3.18 – Les 50×50 premiers coefficients de la matrice de covariance Σ pour différentes valeurs du paramètre η .

grands, qui apparaissent en blanc, sont situés sur la diagonale et la matrice tend à être diagonale lorsque η augmente.

Pour les résultats de simulation que nous présentons, la valeur de γ a été ajusté à $\gamma = 4.8 \times 10^{-4}$, ce qui correspond au cas limite $\text{RSB} = 25\text{dB}$ lorsque $\eta \gg 1$. Cependant, afin d'introduire de la corrélation dans la séquence de bruit et de s'éloigner de ce cas limite, le paramètre η a été fixé à $\eta = 1.1$. Les spectres des composants purs et le spectre résultant du pixel mélangé sont représentés sur la figure 3.19.

La figure 3.20 illustre les lois *a posteriori* des abondances α_r ($r = 1, 2, 3$) obtenues pour $N_{\text{MC}} = 1000$ itérations (dont $N_{\text{bi}} = 100$ itérations de chauffage). Ces lois sont en accord avec les valeurs réelles des paramètres $\alpha^+ = [0.05, 0.6, 0.35]^T$. Pour comparaison, les lois *a posteriori* de ces mêmes coefficients estimées par la procédure de démélange spectral détaillé aux paragraphes 3.3 et 3.4 sont représentées sur cette même figure en pointillés. Ces lois montrent que la procédure permet d'accroître significativement les performances d'estimation. Les variances des estimations avec la méthode adaptée sont en effet plus faibles que celles des estimations obtenues avec l'algorithme initial. Les valeurs de ces variances sont reportées dans le tableau 3.1.

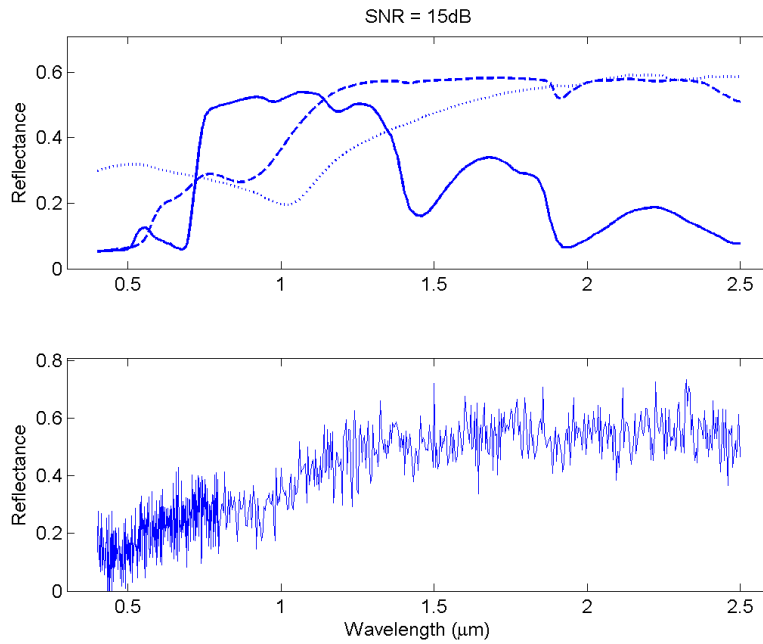


FIG. 3.19 – En haut - spectres des composants purs : herbe verte (trait plein), brique rouge (tirets), métal galvanisé (pointillés). En bas - spectre résultant du pixel mélangé.

3.7.4 Vers un modèle plus général pour des RSBs hétérogènes

Modèle bayésien hiérarchique et échantillonneur de Gibbs

À cause de leur fonctionnement intrinsèque, certains spectro-imageurs fournissent des images hyperspectrales dont les RSBs peuvent varier d'un ensemble de bandes spectrales à l'autre. Par exemple, dans le cas du capteur Hyperion de la NASA embarqué à bord du satellite EO-1, trois différentes plages de bandes spectrales (appelées régions spectrales par la suite) doivent être considérées (550 – 700nm, 700 – 1225nm et 1225 – 2125nm) avec différents RSBs [PBS+03]. Cette hétérogénéité des variances de bruit peut motiver l'introduction d'une loi *a priori* pour Σ^{-1} plus générale que celle présentée dans (3.20). Plus précisément, il est possible de redéfinir la matrice \mathbf{D} introduite précédemment comme

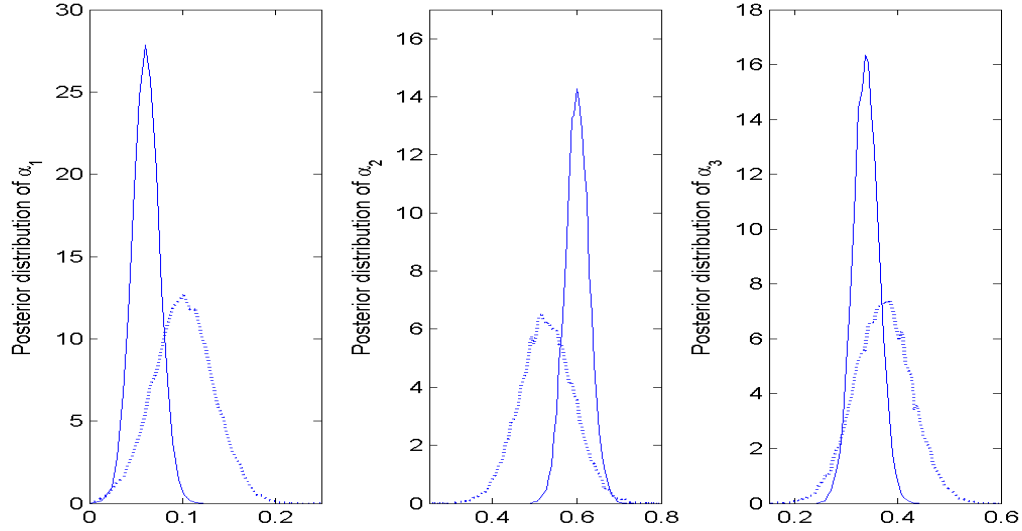


FIG. 3.20 – Lois *a posteriori* des abondances $[\alpha_1, \alpha_2, \alpha_3]^\top$ estimées par l'algorithme proposé (trait plein) et par l'algorithme initial (pointillés).

TAB. 3.1 – Variances des abondances estimées ($\times 10^{-5}$).

	var $[\hat{\alpha}_1]$	var $[\hat{\alpha}_2]$	var $[\hat{\alpha}_3]$
Approche initiale	10.28	40.43	30.44
Approche proposée	2.09	8.38	6.32

suit :

$$\mathbf{D} = \begin{pmatrix} \frac{1}{\gamma_1} \mathbf{I}_{L_1} & & & \\ & \frac{1}{\gamma_2} \mathbf{I}_{L_2} & & \\ & & \ddots & \\ & & & \frac{1}{\gamma_P} \mathbf{I}_{L_P} \end{pmatrix}, \quad (3.28)$$

où P est le nombre de régions spectrales, γ_p est lié à la variance du bruit dans la p -ième région spectrale, L_p est le nombre de bandes spectrales dans la p -ième région. Nous avons bien sûr $L = \sum_{p=1}^P L_p$

(nombre total de bandes de l'image). Dans ce cas, la loi *a priori* de Σ^{-1} est :

$$f(\Sigma^{-1}|\gamma, \nu) \propto |\Sigma^{-1}|^{\frac{1}{2}} \text{etr} \left\{ -\frac{\nu}{2} \mathbf{D}^{-1} \Sigma^{-1} \right\}, \quad (3.29)$$

avec $\gamma = [\gamma_1, \dots, \gamma_P]^T$. Comme dans le cas de la segmentation de signaux de dynamiques différentes traitée au paragraphe 2.5.1, il est nécessaire de définir des lois *a priori* différentes pour les hyperparamètres γ_p supposés indépendants ($p = 1, \dots, P$), en l'occurrence des lois de Jeffrey :

$$f(\gamma) = \prod_{p=1}^P \frac{1}{\gamma_p} \mathbf{1}_{\mathbb{R}^+}(\gamma_p). \quad (3.30)$$

La loi *a posteriori* s'écrit alors :

$$\begin{aligned} f(\Theta, \gamma|\mathbf{y}) &\propto |\Sigma^{-1}|^{\frac{\nu-L}{2}} \exp \left[-\frac{\|\mathbf{y} - \mathbf{M}^+ \boldsymbol{\alpha}^+\|_{\Sigma^{-1}}^2}{2} \right] \mathbf{1}_{\mathbb{S}}(\boldsymbol{\alpha}) \\ &\times |\mathbf{D}^{-1}|^{\frac{\nu}{2}} \text{etr} \left\{ -\frac{\nu}{2} \mathbf{D}^{-1} \Sigma^{-1} \right\} \prod_{p=1}^P \frac{1}{\gamma_p} \mathbf{1}_{\mathbb{R}^+}(\gamma_p), \end{aligned} \quad (3.31)$$

où \mathbf{D} est donnée par (3.28).

Dans l'algorithme de Gibbs précédent (ALGO. 3.3), il suffit donc de remplacer l'étape de simulation de l'hyperparamètre γ par les simulations de chacun des hyperparamètres γ_p . Pour identifier les lois *a posteriori* des hyperparamètres γ_p , il suffit d'écrire :

$$\text{tr} \{ \mathbf{D}^{-1} \Sigma^{-1} \} = \sum_{p=1}^P \text{tr}_{[p]} \{ \gamma_p \Sigma^{-1} \}, \quad (3.32)$$

où $\text{tr}_{[p]} \{ \mathbf{X} \} = \sum_{i \in \mathcal{R}_p} x_{i,i}$ est la trace de la matrice $\mathbf{X} = (x_{i,j})_{1 \leq i,j \leq L}$ évaluée dans la p -ième région spectrale. Il vient alors simplement le résultat suivant :

$$\gamma_p | \boldsymbol{\alpha}, \Sigma^{-1}, \mathbf{y} \sim \mathcal{G} \left(\frac{\nu L_p}{2}, \frac{\nu}{2} \text{tr}_{[p]} \{ \Sigma^{-1} \} \right). \quad (3.33)$$

Illustration à l'aide d'un pixel synthétique

Comme au paragraphe 3.7.3, la méthode proposée est utilisée pour réaliser l'analyse d'un pixel résultant du mélange de trois composants purs avec $\alpha_1 = 0.05$, $\alpha_2 = 0.6$ et $\alpha_3 = 0.35$. La matrice de covariance est générée aléatoirement selon une loi de Wishart avec \mathbf{D} comme matrice moyenne. Cette

matrice a été choisie de la forme spécifiée par (3.28) à partir de trois régions spectrales : $0.4 - 0.7\mu\text{m}$, $0.7 - 1.225\mu\text{m}$ et $1.225 - 2.5\mu\text{m}$. Dans chacune de ces régions, les hyperparamètres γ_p sont choisis comme suit : $\gamma_1 = 1.0e^{-3}$, $\gamma_2 = 1.5e^{-3}$ et $\gamma_3 = 3.8e^{-3}$. Les RSBs correspondants dans les cas limites sont 150dB, 100dB et 40dB. Le spectre de ce pixel est représenté sur la figure 3.21.

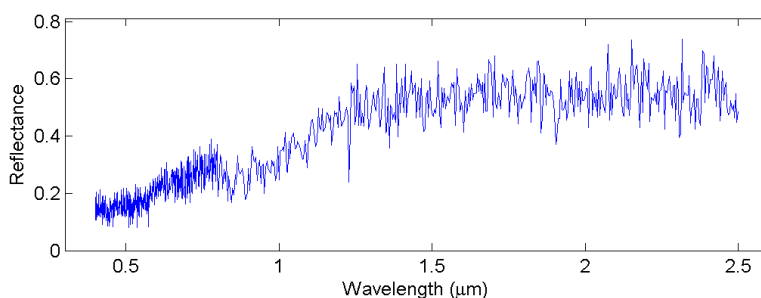


FIG. 3.21 – Spectre d'un pixel résultant d'un mélange avec des variances de bruits différentes d'une région spectrale à l'autre.

Les lois *a posteriori* des abondances estimées α_r ($r = 1, 2, 3$) après $N_{MC} = 20100$ itérations (dont $N_{bi} = 100$ itérations de chauffage) sont représentées sur la figure 3.22. Ces lois sont comparées à celles obtenus par l'algorithme précédent. La prise en compte de RSBs hétérogènes permet d'améliorer notablement les performances de l'estimation. Par ailleurs, l'algorithme proposé permet d'estimer des hyperparamètres γ_p dans les différentes régions spectrales. Les estimateurs MMSE de ces hyperparamètres sont comparés à leur valeurs réelles dans le tableau 3.2.

TAB. 3.2 – Valeurs réelles et estimées des hyperparamètres $\gamma_1, \gamma_2, \gamma_3$.

	γ_1	γ_2	γ_3
Valeurs réelles	$1.0e^{-3}$	$1.5e^{-3}$	$3.8e^{-3}$
Valeurs estimées	$9.0e^{-4}$	$1.7e^{-3}$	$3.8e^{-3}$

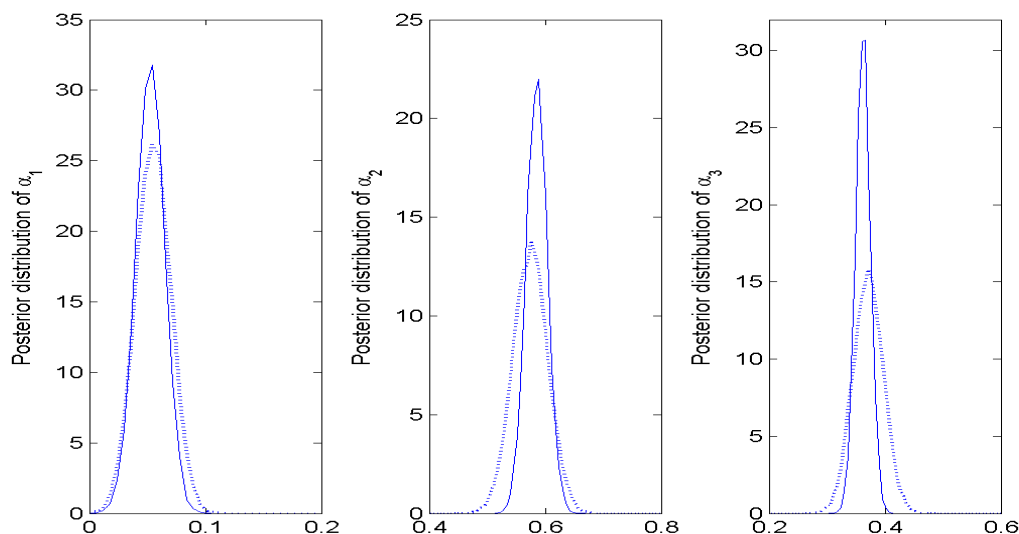


FIG. 3.22 – Lois *a posteriori* des abondances $[\alpha_1, \alpha_2, \alpha_3]^\top$ estimées par l'algorithme proposé (trait plein) et par l'algorithme ALGO. 3.3 (pointillés) avec des RSBs hétérogènes.

3.8 Estimation du nombre de pôles de mélange à l'aide d'un algorithme à sauts réversibles

Cette partie généralise l'échantillonneur présenté dans les paragraphes 3.3 et 3.4 aux cas où le nombre R de composants purs est inconnu. Nous supposons à présent que les R pôles de mélange appartiennent à une bibliothèque spectrale connue $\mathcal{S} = \{\mathbf{s}_1, \dots, \mathbf{s}_{R_{\max}}\}$ (où \mathbf{s}_r représente le L -spectre $[s_{r,1}, \dots, s_{r,L}]^\top$ du r -ième pôle de mélange). Cependant, le nombre de composants R ainsi que les spectres de la bibliothèque ayant pris part au mélange ne sont pas connus.

3.8.1 Modèle bayésien hiérarchique étendu

La loi *a posteriori* du vecteur des paramètres inconnus $\{\boldsymbol{\alpha}, \mathbf{M}^+, R, \sigma^2\}$ peut s'écrire :

$$f(\boldsymbol{\alpha}, \mathbf{M}^+, R, \sigma^2 | \mathbf{y}) \propto f(\mathbf{y} | \boldsymbol{\alpha}, \mathbf{M}^+, \sigma^2, R) \times f(\boldsymbol{\alpha} | R) f(\mathbf{M}^+ | R) f(\sigma^2) f(R), \quad (3.34)$$

où

$$f(\mathbf{y}|\boldsymbol{\alpha}^+, \sigma^2) = \left(\frac{1}{2\pi\sigma^2} \right)^{\frac{L}{2}} \exp \left[-\frac{\|\mathbf{y} - \mathbf{M}(R)^+ \boldsymbol{\alpha}(R)^+\|^2}{2\sigma^2} \right]. \quad (3.35)$$

et les dimensions de $\mathbf{M}(R)^+$ et $\boldsymbol{\alpha}(R)$ dépendent du paramètre inconnu R . Les lois *a priori* $f(\boldsymbol{\alpha}|R)$ et $f(\sigma^2)$ ont été précédemment définies au paragraphe 3.3.2. Une loi uniforme discrète sur $[2, \dots, R_{\max}]$ est choisie comme loi *a priori* du nombre de composants de mélange R :

$$f(R) = \frac{1}{R_{\max} - 1}, \quad R = 2, \dots, R_{\max}. \quad (3.36)$$

Par ailleurs, conditionnellement à R , toutes les combinaisons de R spectres appartenant à la bibliothèque \mathcal{S} sont supposées équiprobable :

$$\begin{aligned} f(\mathbf{M}^+ | R) &= \frac{1}{\binom{R_{\max}}{R}}, \\ &= \frac{\Gamma(R+1) \Gamma(R_{\max} - R + 1)}{\Gamma(R_{\max} + 1)}. \end{aligned} \quad (3.37)$$

3.8.2 Algorithme de Gibbs hybride

Cette partie décrit un algorithme de Gibbs capable de générer des échantillons distribués suivant $f(\boldsymbol{\alpha}, \mathbf{M}^+, \sigma^2, R|\mathbf{y})$. Les vecteurs à générer appartiennent à un espace dont la dimension dépend de R , nécessitant l'utilisation d'une stratégie permettant de passer d'un espace donné à un espace de dimension différente comme dans [RG97]. Plus précisément, l'algorithme noté algorithme 3.4 comporte trois mouvements principaux :

- 1) mise à jour des spectres des pôles de mélange \mathbf{M}^+ ,
- 2) mise à jour des vecteurs d'abondance $\boldsymbol{\alpha}$,
- 3) mise à jour de la variance du bruit σ^2 .

Ces trois mouvements possibles sont réalisés systématiquement comme dans [RG97] et sont détaillés ci-dessous.

Mise à jour des spectres des composants purs \mathbf{M}^+

Les spectres des pôles de mélange prenant part au mélange sont mis à jour à l'aide de trois sortes de mouvement, appelés mouvements de « NAISSANCE, » « MORT » et « ÉCHANGE » comme dans

-
- Initialisation :
 - Échantillonner le paramètre $R^{(0)}$,
 - Choisir $R^{(0)}$ spectres dans la bibliothèque \mathcal{S} pour construire $\mathbf{M}^{+(0)}$,
 - Échantillonner les paramètres $\sigma^{2(0)}$ et $\boldsymbol{\alpha}^{(0)}$,
 - Poser $t \leftarrow 1$,
 - Itérations : pour $t = 1, 2, \dots$, faire
 - mettre à jour la matrice des spectres $\mathbf{M}^{+(t)}$:
 - tirer $u_1 \sim \mathcal{U}_{[0,1]}$,
 - SI $w \leq b_{R^{(t-1)}}$, ALORS
 - proposer un mouvement de NAISSANCE (voir l'algorithme 3.5),
 - SINON SI $w \leq b_{\tilde{R}^{(t-1)}} + d_{\tilde{R}^{(t-1)}}$, ALORS
 - proposer un mouvement de MORT (voir l'algorithme 3.6),
 - SINON
 - proposer un mouvement d'ÉCHANGE (voir l'algorithme 3.7),
 - FIN SI,
 - tirer $u_2 \sim \mathcal{U}_{[0,1]}$,
 - SI $u_2 < \rho$ (voir (G.6) ou (3.41)) ALORS
 - poser $(\boldsymbol{\alpha}^{(t)}, \mathbf{M}^{+(t)}, R^{(t)}) = (\boldsymbol{\alpha}^*, \mathbf{M}^{+*}, R^*)$,
 - SINON
 - poser $(\boldsymbol{\alpha}^{(t)}, \mathbf{M}^{+(t)}, R^{(t)}) = (\boldsymbol{\alpha}^{(t-1)}, \mathbf{M}^{+(t-1)}, R^{(t-1)})$,
 - FIN SI,
 - Échantillonner $\boldsymbol{\alpha}^{(t)}$ d'après la loi (3.42),
 - Échantillonner $\sigma^{2(t)}$ d'après la loi (3.43),
 - Poser $t \leftarrow t + 1$.

ALGO. 3.4: Échantillonneur de Gibbs hybride pour le démelange semi-supervisé d'images hyperspectrales

[DHMS02, p. 53]. Les deux premiers mouvements consistent respectivement à augmenter ou à diminuer le nombre de composants purs R par 1. C'est pourquoi, ils nécessitent l'utilisation d'une méthode MCMC à sauts réversibles introduite par Green dans [Gre95] et déjà rencontré au chapitre 2. En revanche, le nombre R n'est pas altéré par le troisième mouvement proposé, ce qui conduit naturellement à l'utilisation d'une procédure de Metropolis-Hastings standard. Supposons qu'à l'itération t , le modèle courant est défini par $(\boldsymbol{\alpha}^{(t)}, \mathbf{M}^{+(t)}, \sigma^{2(t)}, R^{(t)})$. Les mouvements de « NAISSANCE », « MORT » et « ÉCHANGE » sont définis comme suit :

- NAISSANCE : un mouvement de *naissance* $R^* = R^{(t)} + 1$ est proposé avec la probabilité $b_{R^{(t)}}$ comme expliqué dans l'algorithme 3.5. Un nouveau spectre \mathbf{s}^* est aléatoirement choisi parmi les spectres des composants purs disponibles dans la bibliothèque \mathcal{S} pour construire $\mathbf{M}^{+*} = [\mathbf{M}^{+(t)}, \mathbf{s}^*]$. Un nouveau vecteur de coefficients d'abondance est proposé selon une règle inspirée de [RG97] :
 - tirer un nouveau coefficient d'abondance w^* d'après la loi Beta $\mathcal{Be}(1, R^{(t)})$,
 - re-normaliser les poids existants tels que tous les poids somment à 1,
 - construire $\boldsymbol{\alpha}^{+*} = [\alpha_1^*, \dots, \alpha_{R^{(t)}}^*, w^*]^\top$,
 Le choix d'une loi Beta $\mathcal{Be}(1, R^{(t)})$ s'explique simplement par le fait qu'en moyenne, les coefficients w^* proposés sont d'autant plus faibles que le nombre de composants $R^{(t)}$ est grand.
- MORT : un mouvement de *mort* $R^* = R^{(t)} - 1$ est proposé avec la probabilité $d_{R^{(t)}}$ comme décrit dans l'algorithme 3.6. Un des spectres de $\mathbf{M}^{+(t)}$ est enlevé, ainsi que l'abondance correspondante. Les coefficients restants sont re-normalisés pour sommer à 1,
- Mouvement d'ÉCHANGE : un mouvement d'*échange* est proposé avec la probabilité $u_{R^{(t)}}$ (voir Algorithme 3.7). Un spectre aléatoirement choisi dans $\mathbf{M}^{+(t)}$ est remplacé par un spectre aléatoirement choisi dans la bibliothèque \mathcal{S} .

À chaque itération, un des mouvements de « NAISSANCE », « MORT » ou d' « ÉCHANGE » est aléatoirement choisi avec les probabilités $b_{R^{(t)}}$, $d_{R^{(t)}}$ et $u_{R^{(t)}}$ respectivement (avec $b_{R^{(t)}} + d_{R^{(t)}} + u_{R^{(t)}} = 1$). Bien sûr, un mouvement de mort n'est pas autorisé pour $R = 2$ et le mouvement de naissance est

-
- poser $R^* = R^{(t)} + 1$,
 - choisir \mathbf{s}^* dans \mathcal{S} tel que $\mathbf{s}^* \neq \mathbf{m}_r^{(t)}$, $r = 1, \dots, R^{(t)}$,
 - ajouter \mathbf{s}^* à $\mathbf{M}^{+(t)}$, i.e. poser

$$\mathbf{M}^{+*} = \left[\mathbf{m}_1^{(t)}, \dots, \mathbf{m}_{R^{(t)}}^{(t)}, \mathbf{s}^* \right], \quad (3.38)$$

- tirer $w^* \sim \mathcal{Be}(1, R^{(t)})$,
- ajouter w^* à $\boldsymbol{\alpha}^{+(t)}$ et re-normaliser les autres coefficients poids, i.e. poser

$$\boldsymbol{\alpha}^{+*} = \left[\frac{\alpha_1^{(t)}}{C}, \dots, \frac{\alpha_{R^{(t)}}^{(t)}}{C}, w^* \right]^\top, \quad (3.39)$$

avec $C = \frac{1}{(1-w^*)}$.

ALGO. 3.5: Algorithme MCMC à sauts réversibles pour l'imagerie hyperspectrale : mouvement de NAISSANCE

- poser $R^* = R^{(t)} - 1$,
- tirer $j \sim \mathcal{U}_{\{1, \dots, R^{(t)}\}}$,
- enlever $\mathbf{m}_j^{(t)}$ de $\mathbf{M}^{+(t)}$, i.e. poser

$$\mathbf{M}^{+*} = \left[\mathbf{m}_1^{(t)}, \dots, \mathbf{m}_{j-1}^{(t)}, \mathbf{m}_{j+1}^{(t)}, \dots, \mathbf{m}_{R^{(t)}}^{(t)} \right],$$

- enlever $\alpha_j^{(t)}$ de $\boldsymbol{\alpha}^{+(t)}$ et re-normaliser les autres coefficients d'abondance, i.e. poser

$$\boldsymbol{\alpha}^{+*} = \left[\frac{\alpha_1^{(t)}}{C}, \dots, \frac{\alpha_{j-1}^{(t)}}{C}, \frac{\alpha_{j+1}^{(t)}}{C}, \dots, \frac{\alpha_{R^{(t)}}^{(t)}}{C} \right]^\top,$$

avec $C = \sum_{r \neq j} \alpha_r^{(t)}$.

ALGO. 3.6: Algorithme MCMC à sauts réversibles pour l'imagerie hyperspectrale : mouvement de MORT

impossible si $R = R_{\max}$ (i.e. $d_2 = b_{R_{\max}} = 0$). Par conséquent, les probabilités de mouvements suivantes ont été choisies : $b_R = d_R = u_R = \frac{1}{3}$ pour $R \in \{3, R_{\max} - 1\}$ et $b_2 = d_{R_{\max}} = u_2 = u_{R_{\max}} = \frac{1}{2}$. Quand le vecteur d'abondances $\boldsymbol{\alpha}^+$ est distribué *a priori* suivant une loi de Dirichlet $\mathcal{D}_R(\delta, \dots, \delta)$,

-
- tirer $j \sim \mathcal{U}_{\{1, \dots, R^{(t)}\}}$,
 - choisir \mathbf{s}^* dans \mathcal{S} tel que $\mathbf{s}^* \neq \mathbf{m}_r^{(t)}$, $r = 1, \dots, R^{(t)}$,
 - remplacer $\mathbf{m}_j^{(t)}$ dans $\mathbf{M}^{+(t)}$ par \mathbf{s}^* , i.e. poser

$$\mathbf{M}^{+*} = \left[\mathbf{m}_1^{(t)}, \dots, \mathbf{m}_{j-1}^{(t)}, \mathbf{s}^*, \mathbf{m}_{j+1}^{(t)}, \dots, \mathbf{m}_{R^{(t)}}^{(t)} \right],$$

- poser $\boldsymbol{\alpha}^{+*} = \boldsymbol{\alpha}^{+(t)}$ et $R^* = R^{(t)}$.
-

ALGO. 3.7: Algorithme MCMC à sauts réversibles pour l'imagerie hyperspectrale : mouvement d'ÉCHANGE

les probabilités d'acceptation pour les mouvements de NAISSANCE et MORT sont $\rho = \min\{1, A_b\}$ et $\rho = \min\{1, A_b^{-1}\}$ avec :

$$\begin{aligned} A_b &= \exp \left[-\frac{\|\mathbf{y} - \mathbf{M}^{+*} \boldsymbol{\alpha}^{+*}\|^2 - \|\mathbf{y} - \mathbf{M}^{+(t)} \boldsymbol{\alpha}^{+(t)}\|^2}{2} \right] \\ &\times \frac{d_{R^{(t)+1}}}{b_{R^{(t)}}} \frac{1}{g_{1, R^{(t)}}(w^*)} (1 - w^*)^{R^{(t)}-1} \\ &\times \frac{\Gamma(\delta R^{(t)} + \delta)}{\Gamma(\delta R^{(t)}) \Gamma(\delta)} w^{*\delta-1} (1 - w^*)^{(\delta-1)R^{(t)}}, \end{aligned} \quad (3.40)$$

où $g_{a,b}(\cdot)$ représente la densité de probabilité d'une loi Beta $\mathcal{B}e(a, b)$. Le détail des calculs conduisant à cette expression peut être trouvée en Annexe G.

La probabilité d'acceptation pour le mouvement d'ÉCHANGE est le rapport classique de Metropolis-Hastings $\rho = \min\{1, A_s\}$ avec :

$$A_s = \exp \left[-\frac{\|\mathbf{y} - \mathbf{M}^{+*} \boldsymbol{\alpha}^{+*}\|^2 - \|\mathbf{y} - \mathbf{M}^{+(t)} \boldsymbol{\alpha}^{+(t)}\|^2}{2} \right], \quad (3.41)$$

puisque le rapport des lois de propositions est égale à 1. Dans chaque direction, la probabilité de choisir un spectre dans la librairie est en effet $1/(R_{\max} - R^{(t)})$.

Générer des échantillons distribués suivant $f(\boldsymbol{\alpha} | \mathbf{M}^+, R, \sigma^2, \mathbf{y})$

Comme dans le modèle initial, la loi *a posteriori* suivante est obtenue :

$$\boldsymbol{\alpha} | \mathbf{M}^+, \sigma^2, R, \mathbf{y} \sim \mathcal{N}_{\mathbb{S}}(\boldsymbol{\mu}, \boldsymbol{\Lambda}). \quad (3.42)$$

Générer des échantillons distribués suivant $f(\sigma^2 | \boldsymbol{\alpha}, \mathbf{M}^+, \sigma^2, R, \mathbf{y})$

Ceci est effectué comme suit :

$$\sigma^2 | \boldsymbol{\alpha}, \mathbf{M}^+, R, \mathbf{y} \sim \mathcal{IG} \left(\frac{L}{2}, \frac{\|\mathbf{y} - \mathbf{M}^+ \boldsymbol{\alpha}^+\|^2}{2} \right). \quad (3.43)$$

3.8.3 Simulations

La précision de l'échantillonneur de Gibbs hybride est mesurée en étudiant le spectre du pixel synthétique utilisé au paragraphe 3.5. Nous rappelons que ce pixel est le résultat du mélange de trois pôles de mélange (béton de construction, herbe verte et terre grasse) avec le vecteur d'abondances $\boldsymbol{\alpha}^+ = [0.3, 0.6, 0.1]^\top$. Les observations sont entachées d'un bruit additif gaussien de rapport signal à bruit RSB = 15dB. Les résultats sont obtenus pour $N_{MC} = 20000$ itérations, dont $N_{bi} = 200$ itérations de chauffage. Cette simulation utilise une bibliothèque extraite du logiciel ENVI contenant six spectres : béton de construction, herbe verte, terre grasse, peinture verte, brique, acier galvanisé. Les spectres de ces composants purs sont représentés sur la figure 3.23.

La première étape de l'analyse consiste en l'estimation de l'ordre du modèle, c'est-à-dire du nombre de composants purs utilisés pour le mélange R . Cette estimation est effectuée par maximum *a posteriori*. La loi *a posteriori* de R représentée sur la figure 3.24 est clairement en accord avec la valeur réelle de R puisque le maximum est obtenu pour $R = 3$. La seconde étape de l'analyse estime les probabilités *a posteriori* des combinaisons de spectres sachant $R = 3$. Pour cette simulation, seulement deux vecteurs ont été générés après la période de chauffage : $[\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3]$ et $[\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_5]$ avec les probabilités respectives $P_{1,2,3} = 0.87$ et $P_{1,2,5} = 0.13$. La probabilité maximale correspond à la combinaison de spectres réellement employés pour réaliser le mélange. Les lois *a posteriori* des coefficients d'abondance correspondants sont finalement estimées conditionnellement à $\mathbf{M}^+ = [\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3]$ et représentées sur la figure 3.25. Ces distributions sont en accord avec les valeurs réelles des abondances $\boldsymbol{\alpha} = [0.3, 0.6, 0.1]^\top$.

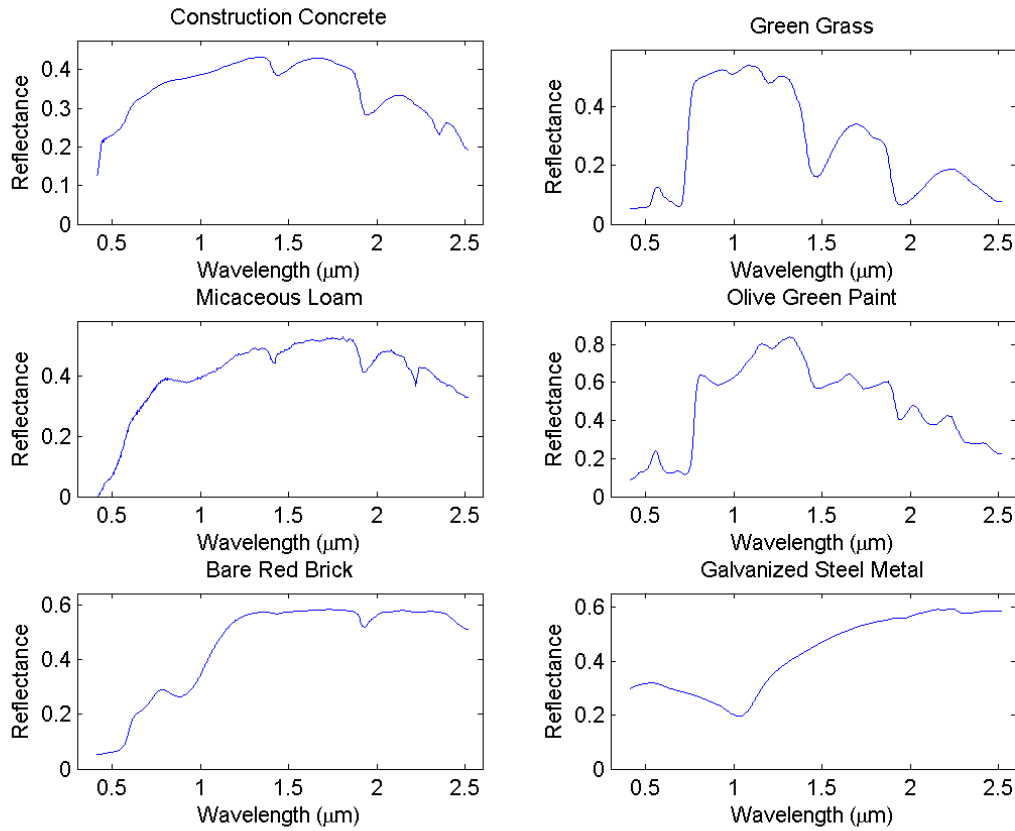
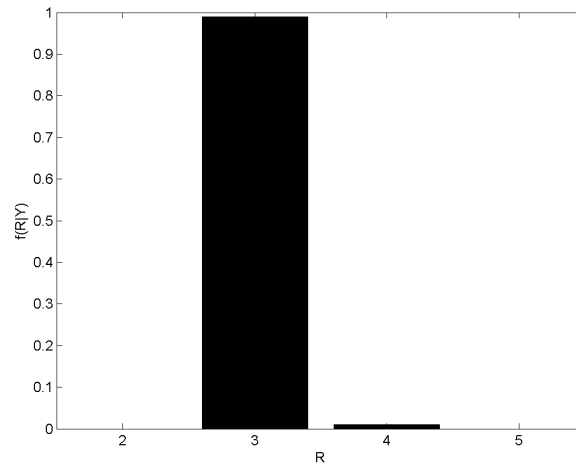


FIG. 3.23 – Spectres des pôles de mélange de la librairie.

3.9 Conclusions

Ce chapitre présente une approche bayésienne hiérarchique pour le démixage d'images multibandes. Des lois *a priori* appropriées permettent de satisfaire les contraintes imposées au modèle considéré. Les lois *a posteriori* des paramètres inconnus sont classiquement échantillonnées grâce à un algorithme de Gibbs. Ces lois fournissent à la fois des estimations des paramètres mais également une information sur la précision de ces estimations.

La stratégie proposée suppose tout d'abord que les spectres des composants purs prenant part au mélange sont préalablement identifiés, grâce par exemple à une connaissance de la vérité terrain. Dans

FIG. 3.24 – Loi *a posteriori* de l'ordre du modèle estimé R .

un deuxième temps, nous développons un algorithme qui permet de résoudre le cas où ces spectres ne sont pas connus. La connaissance *a priori* d'un dictionnaire suffit alors à retrouver le nombre de matériaux impliqués dans le mélange et leur nature respective.

Enfin, nous envisageons le cas plus complexe où les observations sont entachées d'un bruit corrélé. Le recours à des lois *a priori* de Wishart permet alors d'estimer correctement la matrice de covariance.

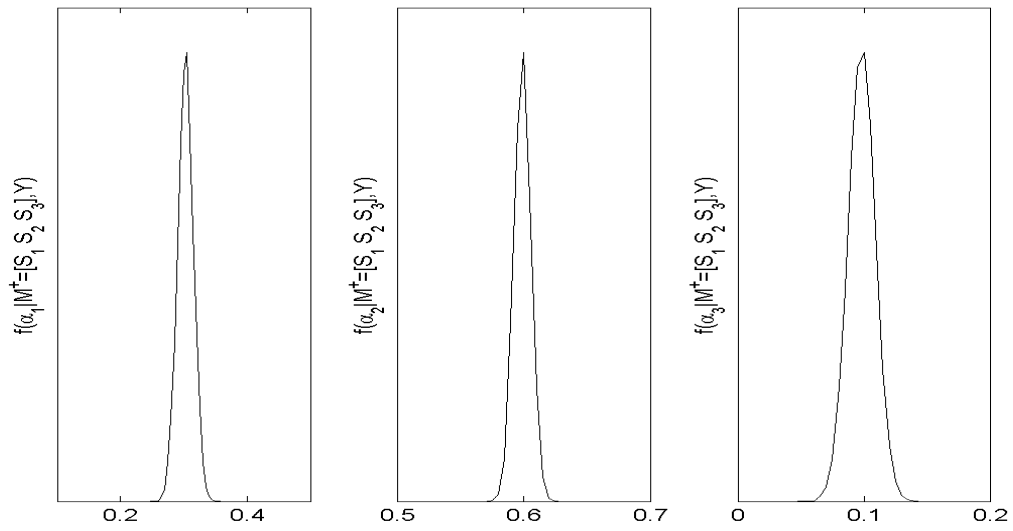


FIG. 3.25 – Lois *a posteriori* des abondances estimées $\boldsymbol{\alpha}^+ = [\alpha_1, \alpha_2, \alpha_3]^\top$ sachant $\mathbf{M}^+ = [\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3]$.

Conclusions et perspectives

Cette thèse a pour objet l'étude de modèles bayésiens hiérarchiques pour le traitement de signaux multi-capteur. La masse d'information récoltée par des capteurs de plus en plus nombreux nécessitent en effet de mettre en place des traitements efficaces, capables d'exploiter au mieux ce caractère multi-capteur. Plutôt que d'aborder un tel problème comme un problème de fusion de données, nous adoptons une stratégie de modélisation statistique de l'ensemble des signaux délivrés. Des lois *a priori* appropriées sont choisies pour les paramètres et hyperparamètres inconnus. La complexité des lois *a posteriori* obtenues est maîtrisée grâce à des méthodes de Monte Carlo par Chaînes de Markov. Elles consistent en la génération d'échantillons distribués asymptotiquement suivant la loi d'intérêt. Ces échantillons sont ensuite utilisés pour approcher les estimateurs bayésiens classiques (MAP ou MMSE).

Les modèles développés ont tout d'abord été utilisés pour des applications de segmentation de signaux multiples. Nous nous intéressons dans un premier chapitre à la détection de changements de brillance dans des séries astronomiques. Les séquences fournies par le télescope BATSE dans quatre bandes d'énergie sont supposées stationnaires par morceaux. La segmentation de ces séquences permet d'estimer le début et la fin des sursauts gamma, phénomènes sur lesquels se penchent de nombreux astrophysiciens. Grâce à la nature multi-capteur des données observées, les modèles de segmentation obtenus par notre approche semblent bien plus pertinents que ceux obtenus lors d'un traitement uni-dimensionnel.

Dans un deuxième chapitre, nous nous intéressons à la segmentation de signaux autorégressifs. Nous proposons une extension des travaux de Punskeya au cas multi-capteur, tout en simplifiant l'algorithme par l'introduction d'indicatrices pour repérer les ruptures. Nous proposons deux extensions

au modèle initial. Tout d'abord, nous montrons comment un modèle markovien sur ces indicatrices permet d'imposer des contraintes de longueur minimale pour les segments obtenus. Par ailleurs, nous traitons le cas où les signaux sont de dynamiques différentes. Les signaux autorégressifs ont fait l'objet de nombreux travaux dans le domaine du traitement du signal. Leur atout majeur est sans aucun doute leur aptitude à modéliser un grand nombre de processus réels complexes. Notamment, les algorithmes développés sont appliqués à des signaux de parole stéréo afin de les segmenter. Enfin, une application à la détection du phénomène d' « arc-tracking » est également envisagée.

Le troisième chapitre se concentre sur l'analyse d'images hyperspectrales, c'est-à-dire acquises dans un grand nombre de bandes spectrales. Lorsque les spectres des composants purs ont été préalablement identifiés, nous abordons le problème fondamental de démixage spectral comme un problème de régression linéaire sous contraintes d'additivité et de positivité. En revanche, lorsque ces matériaux sont *a priori* inconnus, nous basons notre étude sur la connaissance d'une bibliothèque spectrale. L'enjeu devient alors double puisqu'il s'agit de réaliser conjointement une sélection de modèle et une estimation des coefficients de régression. Notons que l'algorithme initialement introduit pour un bruit gaussien i.i.d. est étendu à des modèles de bruits plus généraux, parfois rencontrés lors de l'utilisation de spectro-imageurs réels.

Différentes perspectives des travaux menés peuvent être dégagées. Dans le premier chapitre, il conviendrait d'envisager le cas où un délai inconnu mais faible sépare des positions de ruptures dans des signaux différents. La résolution temporelle des appareils utilisés induit en effet un retard naturel dans la détection des photons qu'il semble important de prendre en compte.

Le second chapitre se concentre sur la segmentation de processus autorégressifs. La généralisation des travaux menés à des signaux plus complexes de type GARCH [Gue94] ou données à longue dépendance [RT02; LT06] pourrait être envisagée. Par ailleurs, dans un objectif de détection d' « arc-tracking, » il serait opportun d'adopter une stratégie de segmentation en-ligne, par méthodes de filtrage par exemple (voir par exemple [FL07]).

Enfin, dans le troisième chapitre, nous supposons la connaissance d'une bibliothèque spectrale dans laquelle sont regroupés les spectres des composants purs susceptibles d'apparaître dans l'image

analysée. Il en résulte un algorithme semi-supervisé de démixage spectral. Afin de rendre totalement non-supervisée la méthode de résolution, il devient impératif d'estimer également les spectres des matériaux. Des techniques de séparation aveugle de sources sous contraintes telles que celles proposées par Moussaoui dans [MBMC06; MCBMD06] peuvent alors être envisagées. Quelques résultats intéressants concernant ces techniques peuvent être trouvés dans [DMT07] et [DTM07].

Enfin, d'une manière plus générale, en ce qui concerne les méthodes de simulation utilisées tout au long de ce document, il serait intéressant d'approcher les lois *a posteriori* complexes rencontrées par des produits de lois séparables. L'utilisation de méthodes bayésiennes variationnelles [JJ00] permettraient alors de réduire les coûts calculatoires des algorithmes de Gibbs proposés dans ce manuscrit.

Annexes

ANNEXE A

Calcul et simulation rapides d'une loi normale multivariée

Nous détaillons dans cette annexe une stratégie efficace pour calculer les matrices $\mathbf{Q}_{j,k}$ et $\mathbf{M}_{j,k}$, ainsi que la variable scalaire $T_{j,k}^2$ définies par l'équation (2.15) et rappelées ci-dessous en notant $\mathbf{y}_{j,[k]} = \mathbf{y}_{j,l_{j,k-1}+1:l_{j,k}}^\top$:

$$\begin{cases} T_{j,k}^2 = \mathbf{y}_{j,[k]}^\top \mathbf{Q}_{j,k} \mathbf{y}_{j,[k]}, \\ \mathbf{Q}_{j,k} = \mathbf{I}_p - \mathbf{Y}_{j,k} \mathbf{M}_{j,k} \mathbf{Y}_{j,k}^\top, \\ \mathbf{M}_{j,k} = \left(\mathbf{Y}_{j,k}^\top \mathbf{Y}_{j,k} + \frac{\mathbf{I}_p}{\delta_0^2} \right)^{-1}, \end{cases} \quad (\text{A.1})$$

Les quantités ci-dessus peuvent en effet être calculées en suivant l'implémentation décrit par Davy *et al.* dans [DI04] et détaillée dans l'algorithme A.1.

-
- Calculer $\mathbf{M}_{j,k}^{-1} = \mathbf{Y}_{j,k}^\top \mathbf{Y}_{j,k} + \frac{\mathbf{I}_p}{\delta_0^2}$,
 - Calculer le facteur de Cholesky $\mathbf{C}_{j,k}$ tel que $\mathbf{C}_{j,k} \mathbf{C}_{j,k}^\top = \mathbf{M}_{j,k}$,
 - Calculer $\mathbf{u}_{j,k} = \mathbf{Y}_{j,k}^\top \mathbf{y}_{j,[k]}$,
 - Résoudre le système $\mathbf{C}_{j,k} \mathbf{v}_{j,k} = \mathbf{u}_{j,k}$ en $\mathbf{v}_{j,k}$,
 - Calculer $T_{j,k}^2 = \mathbf{y}_{j,[k]}^\top \mathbf{y}_{j,[k]} - \mathbf{v}_{j,k}^\top \mathbf{v}_{j,k}$.
-

ALGO. A.1: Calculs rapides de $T_{j,k}^2$

Une telle implémentation nous permet par ailleurs de développer une stratégie pour générer des échantillons $\mathbf{a}_{j,k} | \delta_0^2, \sigma_{j,k}^2, \mathbf{R}, \mathbf{Y}$ distribués suivant la loi normale $\mathcal{N}(\boldsymbol{\mu}_{j,k}, \sigma_{j,k}^2 \mathbf{M}_{j,k})$ (voir paragraphe 2.3.2 et Eq. (2.20)) selon l'algorithme efficace A.2.

Un autre avantage de ce schéma de simulation est qu'il n'est pas nécessaire de calculer directement les déterminants des matrices $\mathbf{M}_{j,k}$ qui apparaissent dans (2.14). En effet, $|\mathbf{M}_{j,k}| = |\mathbf{C}_{j,k}|^{-2}$ où $\mathbf{C}_{j,k}$ sont des matrices triangulaires supérieures dont les déterminants sont très faciles à calculer.

-
- Tirer un vecteur i.i.d. $\mathbf{w}_{j,k}$ selon $\mathcal{N}(\mathbf{0}, \sigma_{j,k}^2 \mathbf{I}_p)$
 - Résoudre le système $\mathbf{C}_{j,k} \tilde{\boldsymbol{\mu}}_{j,k} = \mathbf{w}_{j,k}$ en $\tilde{\boldsymbol{\mu}}_{j,k}$,
 - Résoudre le système $\mathbf{C}_{j,k}^\top \boldsymbol{\mu}'_{j,k} = \mathbf{Y}_{j,k}^\top \mathbf{y}_{j,[k]}$ en $\boldsymbol{\mu}'_{j,k}$,
 - Résoudre le système $\mathbf{C}_{j,k}^\top \boldsymbol{\mu}_{j,k} = \boldsymbol{\mu}'_{j,k}$ en $\boldsymbol{\mu}_{j,k}$,
 - Calculer $\mathbf{a}_{j,k} = \tilde{\boldsymbol{\mu}}_{j,k} + \boldsymbol{\mu}_{j,k}$.
-

ALGO. A.2: Echantillonnage rapide des paramètres autorégressifs $\mathbf{a}_{j,k}$

ANNEXE B

Calcul de la loi *a posteriori* des coefficients d'abondance

Nous détaillons dans cette annexe les calculs permettant d'exprimer la loi *a posteriori* du vecteur d'abondance donnée dans (3.9). En utilisant la règle de Bayes, la loi *a posteriori* $f(\boldsymbol{\alpha}|\sigma^2, \mathbf{y})$ peut s'écrire :

$$\begin{aligned}
 f(\boldsymbol{\alpha}|\sigma^2, \mathbf{y}) &\propto f(\mathbf{y}|\boldsymbol{\alpha}, \sigma^2) f(\boldsymbol{\alpha}), \\
 &\propto \exp\left[-\frac{\|\mathbf{y} - \mathbf{M}^+ \boldsymbol{\alpha}^+\|^2}{2\sigma^2}\right] \mathbf{1}_{\mathbb{T}}(\boldsymbol{\alpha}), \\
 &\propto \exp\left[-\frac{C(\boldsymbol{\alpha}|\mathbf{y}, \sigma^2)}{2\sigma^2}\right] \mathbf{1}_{\mathbb{T}}(\boldsymbol{\alpha}),
 \end{aligned} \tag{B.1}$$

avec

$$C(\boldsymbol{\alpha}|\mathbf{y}, \sigma^2) = \|\mathbf{y} - \mathbf{M}^+ \boldsymbol{\alpha}^+\|^2.$$

De simples calculs conduisent à :

$$\begin{aligned}
 \|\mathbf{y} - \mathbf{M}^+ \boldsymbol{\alpha}^+\|^2 &= \sum_{l=1}^L \left[y_l - \sum_{r=1}^{R-1} m_{r,l} \alpha_r - m_{R,l} \alpha_R \right]^2, \\
 &= \sum_{l=1}^L \left[\left(y_l - \sum_{r=1}^{R-1} m_{r,l} \alpha_r \right)^2 - 2 \left(y_l - \sum_{r=1}^{R-1} m_{r,l} \alpha_r \right) m_{R,l} \alpha_R + (m_{R,l} \alpha_R)^2 \right] \\
 &= \|\mathbf{y} - \mathbf{M} \boldsymbol{\alpha}\|^2 - 2(\mathbf{y} - \mathbf{M} \boldsymbol{\alpha})^\top \mathbf{m}_R (1 - \mathbf{u}^\top \boldsymbol{\alpha}) + \|\mathbf{m}_R (1 - \mathbf{u}^\top \boldsymbol{\alpha})\|^2,
 \end{aligned}$$

avec $\mathbf{u} = [1, \dots, 1]^\top \in \mathbb{R}^{R-1}$, et donc

$$\begin{aligned} C(\alpha|\mathbf{y}, \sigma^2) &= \|\mathbf{y} - \mathbf{M}^+ \alpha^+\|^2 \\ &= \left[\|\mathbf{y} - \mathbf{M}\alpha\|^2 - 2(\mathbf{y} - \mathbf{M}\alpha)^\top \mathbf{m}_R(1 - \mathbf{u}^\top \alpha) + \|\mathbf{m}_R(1 - \mathbf{u}^\top \alpha)\|^2 \right] \\ &= \left[(\mathbf{y}^\top \mathbf{y} - \alpha^\top \mathbf{M}^\top \mathbf{y} - \mathbf{y}^\top \mathbf{M}\alpha + \alpha^\top \mathbf{M}^\top \mathbf{M}\alpha) + 2(\mathbf{M}\alpha - \mathbf{y})^\top (\mathbf{m}_R - \mathbf{m}_R \mathbf{u}^\top \alpha) \right] \\ &\quad + \left[\|\mathbf{m}_R\|^2 (1 - 2\mathbf{u}^\top \alpha + \alpha^\top \mathbf{u} \mathbf{u}^\top \alpha) \right] \end{aligned}$$

En réorganisant les différents termes :

$$\begin{aligned} C(\alpha|\mathbf{y}, \sigma^2) &\propto \alpha^\top \left[(\mathbf{M}^\top \mathbf{M} - \mathbf{M}^\top \mathbf{m}_R \mathbf{u}^\top - \mathbf{u} \mathbf{m}_R^\top \mathbf{M} + \|\mathbf{m}_R\|^2 \mathbf{u} \mathbf{u}^\top) \right] \alpha \\ &\quad + \alpha^\top \left[(-\mathbf{M}^\top \mathbf{y} + \mathbf{M}^\top \mathbf{m}_R + \mathbf{u} \mathbf{m}_R^\top \mathbf{y} - \|\mathbf{m}_R\|^2 \mathbf{u}) \right] \\ &\quad + \left[(-\mathbf{M}^\top \mathbf{y} + \mathbf{M}^\top \mathbf{m}_R + \mathbf{u} \mathbf{m}_R^\top \mathbf{y} - \|\mathbf{m}_R\|^2 \mathbf{u}) \right]^\top \alpha, \end{aligned}$$

ou de manière équivalente :

$$\begin{aligned} C(\alpha|\mathbf{y}, \sigma^2) &\propto \alpha^\top \left[(\mathbf{M} - \mathbf{m}_R \mathbf{u}^\top)^\top (\mathbf{M} - \mathbf{m}_R \mathbf{u}^\top) \right] \alpha \\ &\quad - \alpha^\top \left[(\mathbf{M} - \mathbf{m}_R \mathbf{u}^\top)^\top (\mathbf{y} - \mathbf{m}_R) \right] \\ &\quad - \left[(\mathbf{M} - \mathbf{m}_R \mathbf{u}^\top)^\top (\mathbf{y} - \mathbf{m}_R) \right]^\top \alpha. \end{aligned}$$

En notant

$$\begin{aligned} \Lambda &= \left[\frac{1}{\sigma^2} (\mathbf{M} - \mathbf{m}_R \mathbf{u}^\top)^\top (\mathbf{M} - \mathbf{m}_R \mathbf{u}^\top) \right]^{-1}, \\ \boldsymbol{\mu} &= \Lambda \left[\frac{1}{\sigma^2} (\mathbf{M} - \mathbf{m}_R \mathbf{u}^\top)^\top (\mathbf{y} - \mathbf{m}_R) \right], \end{aligned}$$

la loi *a posteriori* $f(\alpha|\sigma^2, \mathbf{y})$ vérifie la relation :

$$f(\alpha|\sigma^2, \mathbf{y}) \propto \exp \left[-\frac{(\alpha - \boldsymbol{\mu})^\top \Lambda^{-1} (\alpha - \boldsymbol{\mu})}{2} \right] \mathbf{1}_{\mathbb{T}}(\alpha).$$

ANNEXE C

Loi multivariée normale tronquée

Nous rappelons dans cette annexe quelques propriétés concernant les lois normales multivariées et les lois normales multivariées tronquées. Soient $\mathbf{m} \in \mathbb{R}^n$ et Σ un opérateur défini positif non singulier sur \mathbb{R}^n , la loi normale multivariée sur \mathbb{R}^n est définie comme suit :

$$\phi(\mathbf{x}|\mathbf{m}, \Sigma) = (2\pi)^{-\frac{n}{2}} (\det \Sigma)^{-\frac{1}{2}} \exp \left[-\frac{1}{2} (\mathbf{x} - \mathbf{m})^\top \Sigma^{-1} (\mathbf{x} - \mathbf{m}) \right], \quad (\text{C.1})$$

La loi normale *standard* est $\phi_{\mathbb{R}^n}(\cdot) = \phi(\cdot|0, \mathbf{I}_n)$. Soit \mathbb{U} un ensemble de \mathbb{R}^n de mesure de Lebesgue non nulle. Considérons la loi sur \mathbb{U} :

$$\phi_{\mathbb{U}}(\mathbf{x}|\mathbf{m}, \Sigma) = \frac{\mathbf{1}_{\mathbb{U}}(\mathbf{x})}{\phi(\mathbb{U}|\mathbf{m}, \Sigma)} \phi(\mathbf{x}|\mathbf{m}, \Sigma), \quad (\text{C.2})$$

où

$$\phi(\mathbb{U}|\mathbf{m}, \Sigma) = \int_{\mathbf{x} \in \mathbb{U}} \phi(\mathbf{x}|\mathbf{m}, \Sigma) d\mathbf{x} \quad (\text{C.3})$$

Nous dirons que $\phi_{\mathbb{U}}(\cdot|\mathbf{m}, \Sigma)$ est la loi normale multivariée tronquée sur \mathbb{U} avec la *moyenne cachée* \mathbf{m} et la *covariance cachée* Σ . La raison de ces termes est que \mathbf{m} et Σ ne sont pas en général la moyenne et la covariance de la loi $\phi_{\mathbb{U}}(\cdot|\mathbf{m}, \Sigma)$. Par exemple, si \mathbb{U} est un convexe, un cas qui arrive très souvent en pratique, alors la moyenne de $\phi_{\mathbb{U}}(\cdot|\mathbf{m}, \Sigma)$ est nécessairement dans \mathbb{U} alors que \mathbf{m} peut être en dehors de \mathbb{U} . En général, si \mathbb{U} est connu et fixé, l'estimation des paramètres cachés \mathbf{m} et Σ à partir d'un échantillon n'est pas un problème facile. Si $\mathbb{R}^n = \mathbb{R}$, cette estimation est étudiée pour $\mathbb{U} = (0, 1)$ dans [Bar78, exemple 9.16] et pour $\mathbb{U} = (0, \infty)$ dans [del94] et dans [Let92, Chapter 2, Theorem 1.1].

ANNEXE D

Simulation efficace d'une loi normale multivariée tronquée à un simplexe

Note : cette annexe reprend principalement les résultats présentés dans le rapport technique [DT07].

D.1 Position du problème

Soit \mathbb{S} le simplexe suivant défini sur \mathbb{R}^{R-1} :

$$\mathbb{S} = \left\{ \boldsymbol{\alpha} \mid \alpha_r \geq 0, \forall r = 1, \dots, R-1, \sum_{r=1}^{R-1} \alpha_r \leq 1 \right\}, \quad (\text{D.1})$$

Notons $\mathcal{N}_{\mathbb{S}}(\mathbf{A}, \mathbf{B})$ la loi normale multivariée définie sur le simplexe \mathbb{S} avec le vecteur moyenne \mathbf{A} et la matrice de covariance \mathbf{B} . La fonction densité de probabilité de la loi normale multivariée $\mathcal{N}_{\mathbb{S}}(\mathbf{A}, \mathbf{B})$ notée $\phi_{\mathbb{S}}(\cdot \mid \mathbf{A}, \mathbf{B})$ satisfait la relation suivante :

$$\phi_{\mathbb{S}}(\boldsymbol{\alpha} \mid \mathbf{A}, \mathbf{B}) \propto \phi(\boldsymbol{\alpha} \mid \mathbf{A}, \mathbf{B}) \mathbf{1}_{\mathbb{S}}(\boldsymbol{\alpha}), \quad (\text{D.2})$$

où $\phi(\cdot \mid \mathbf{A}, \mathbf{B})$ est la fonction densité de probabilité de la loi normale de moyenne \mathbf{A} et de matrice de covariance $\mathbf{\Sigma}$.

Nous proposons dans cette annexe un moyen efficace de générer des échantillons $\boldsymbol{\alpha}^{(t)}$ distribués suivant $\mathcal{N}_{\mathbb{S}}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, où $\boldsymbol{\mu}$ est un vecteur moyenne $(R-1) \times 1$ et $\boldsymbol{\Sigma}$ est une matrice de covariance de taille $(R-1) \times (R-1)$.

D.2 Echantillonneur de Gibbs pour loi multivariée normale tronquée

Un moyen très simple de générer des échantillons distribués suivant $\phi_{\mathbb{S}}(\boldsymbol{\alpha} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma})$ est d'utiliser un méthode simple d'acceptation-rejet [RC99, p. 49] qui consiste à simuler suivant $\boldsymbol{\alpha} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ jusqu'à

ce que l'échantillon $\boldsymbol{\alpha}$ appartient à \mathbb{S} . Cependant, pour des espaces de grandes dimensions ou si $\boldsymbol{\mu} \notin \mathbb{S}$, cette stratégie peut s'avérer très vite inefficace. Le rapport d'acceptation est en effet lié à la probabilité $\mathbb{P}[\boldsymbol{\alpha} \in \mathbb{S} | \boldsymbol{\alpha} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})]$, ce qui peut conduire à rejeter les échantillons $\boldsymbol{\alpha}$ bien trop souvent. Pour dépasser ce problème, nous proposons un algorithme alternatif inspirée de [Rob95] et basée sur l'échantillonneur de Gibbs. La génération d'un vecteur $\boldsymbol{\alpha}^{(t)}$ distribué suivant la loi normale multivariée $\mathcal{N}_{\mathbb{S}}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ est alors réalisée en générant successivement les composants $\alpha_r^{(t)}$ ($r = 1, \dots, R - 1$) selon leurs lois conditionnelles, c'est-à-dire à l'étape $t + 1$:

1. $\alpha_1^{(t+1)} \sim f\left(\alpha_1^{(t+1)} | \alpha_2^{(t)}, \dots, \alpha_{R-1}^{(t)}\right),$
2. $\alpha_2^{(t+1)} \sim f\left(\alpha_2^{(t+1)} | \alpha_1^{(t+1)}, \alpha_3^{(t)}, \dots, \alpha_{R-1}^{(t)}\right),$
3. $\alpha_3^{(t+1)} \sim f\left(\alpha_3^{(t+1)} | \alpha_1^{(t+1)}, \alpha_2^{(t+1)}, \alpha_4^{(t)}, \dots, \alpha_{R-1}^{(t)}\right),$
- \vdots
- $R - 1.$ $\alpha_{R-1}^{(t+1)} \sim f\left(\alpha_{R-1}^{(t+1)} | \alpha_1^{(t+1)}, \dots, \alpha_{R-2}^{(t+1)}\right),$

En notant $\boldsymbol{\alpha}_{-r}$ le vecteur $\boldsymbol{\alpha}$ dans lequel le r -ième élément a été supprimé, les lois conditionnelles $f(\alpha_r | \boldsymbol{\alpha}_{-r})$ sont :

$$\alpha_r | \boldsymbol{\alpha}_{-r} \sim \mathcal{N}_-^+(\mu_r^*, \sigma_r^{2*}, \alpha_r^-, \alpha_r^+), \quad (\text{D.3})$$

où $\mathcal{N}_-^+(\mu_r^*, \sigma_r^{2*}, \alpha_r^-, \alpha_r^+)$ représente la loi normale univariée tronquée à gauche et à droite donnée par :

$$\phi_-^+(x | \mu_r^*, \sigma_r^{2*}, \alpha_r^-, \alpha_r^+) \propto \phi(x | \mu_r^*, \sigma_r^{2*}) \mathbf{1}_{\alpha_r^- \leq x \leq \alpha_r^+}(x) \quad (\text{D.4})$$

La moyenne μ_r^* et la variance σ_r^{2*} cachées des lois normales univariées doublement tronquées sont données en introduisant les matrices et moyennes partitionnées [Kay93, p. 324] :

$$\begin{cases} \mu_r^* &= \mu_r + \mathbf{s}_{-r}^\top \boldsymbol{\Sigma}_{-r}^{-1} (\boldsymbol{\alpha}_{-r} - \boldsymbol{\mu}_{-r}) \\ \sigma_r^{2*} &= \sigma_{r,r}^2 - \mathbf{s}_{-r}^\top \boldsymbol{\Sigma}_{-r}^{-1} \mathbf{s}_{-r}, \end{cases} \quad (\text{D.5})$$

où

- μ_r représente le r -ième élément de $\boldsymbol{\mu}$,
- \mathbf{s}_{-r} est le vecteur de taille $(R - 2) \times 1$ obtenu à partir de la r -ième colonne de $\boldsymbol{\Sigma}$ en supprimant le terme de la r -ième ligne,

- Σ_{-r} est la matrice $(R-2) \times (R-2)$ obtenue à partir de Σ en supprimant la r -ième ligne et la r^{th} colonne,
- $\boldsymbol{\mu}_{-r}$ est le vecteur $(R-2) \times 1$ obtenu à partir de $\boldsymbol{\mu}$ en supprimant le r -ième élément.

Pour la loi multivariée normale tronquée au simplexe donné dans l'équation (D.1), les frontières de la troncature α_r^- et α_r^+ sont simplement :

$$\begin{cases} \alpha_r^- = 0, \\ \alpha_r^+ = 1 - \sum_{i \neq r} \alpha_i. \end{cases} \quad (\text{D.6})$$

Notons que générer des échantillons distribués suivant une loi univariée normale doublement tronquée peut être effectuée simplement en utilisant l'algorithme décrit dans [Rob95].

D.3 Simulations

En exemple, nous considérons la génération de variables $\boldsymbol{\alpha} = [\alpha_1, \alpha_2]^T$ distribuées suivant la loi normale multivariée $\mathcal{N}_{\mathbb{S}}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ tronquée au simplexe $\mathbb{S} \in \mathbb{R}^2$ avec $\boldsymbol{\mu} = [0.2, 0.2]^T$ et $\boldsymbol{\Sigma} = \begin{pmatrix} 0.13 & 0.08 \\ 0.08 & 0.13 \end{pmatrix}$. L'histogramme des variables générées par l'algorithme proposé plus haut est représenté sur la figure D.1. Notons qu'une procédure d'acceptation-rejet standard pour la génération d'échantillons distribués suivant la loi utilisée serait d'environ $\rho = 0.59$.

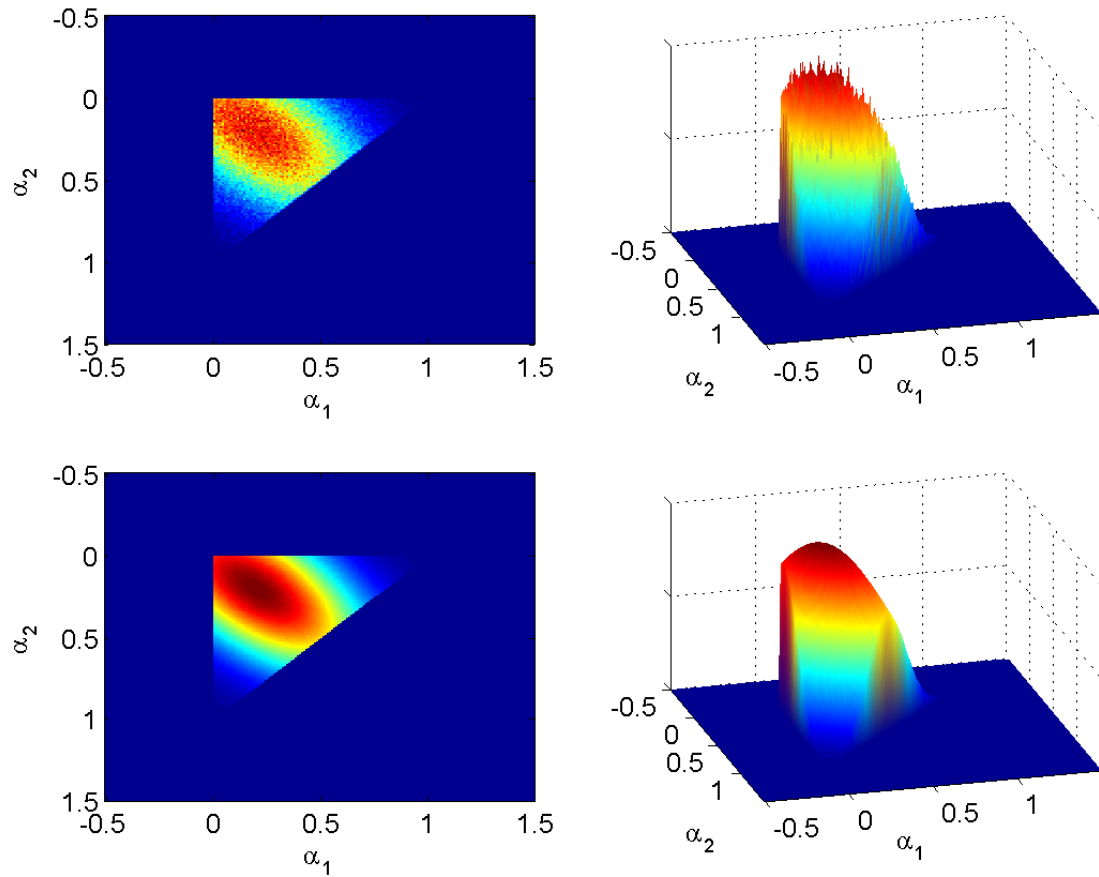


FIG. D.1 – En haut : histogramme de $N = 500000$ variables simulées avec l'échantillonneur de Gibbs proposé pour une loi normale multivariée tronquée au simplexe \mathbb{S} ($R = 3$). En bas : densité de probabilité théorique correspondante.

ANNEXE E

Loi de Wishart : définition et principales propriétés

Dans cette annexe, nous rappelons la définition de la loi de Wishart ainsi que ses principales propriétés.

E.1 Loi de probabilité

Soient $\mathbf{x}_i = [x_{1,i}, \dots, x_{n,i}]^\top$ ($i = 1, \dots, p$) p vecteurs gaussiens de \mathbb{R}^n distribués suivant la loi normale multivariée $\mathcal{N}(\mathbf{0}_n, \Psi)$ centrée et de matrice de covariance Ψ . La matrice aléatoire \mathbf{A} de taille $n \times n$ définie par $\mathbf{A} = \sum_{i=1}^p \mathbf{x}_i \mathbf{x}_i^\top$ suit une loi de Wishart notée $\mathcal{W}(p, \Psi)$ avec p degrés de liberté et de matrice paramètre Ψ . Sa fonction densité de probabilité est donnée par [Mui05, p.85] :

$$f(\mathbf{A}|p, \Psi) = \frac{1}{c(n, p)} |\Psi|^{-\frac{p}{2}} |\mathbf{A}|^{\frac{p-n-1}{2}} \text{etr} \left\{ -\frac{1}{2} \Psi^{-1} \mathbf{A} \right\} \mathbf{1}_{\{\mathbf{A} > 0\}}(\mathbf{A}), \quad (\text{E.1})$$

où $\{\mathbf{A} > 0\}$ désigne l'ensemble des matrices définies positives, $c(n, p)$ est la constante de normalisation :

$$c(n, p) = 2^{\frac{pn}{2}} \pi^{\frac{n(n-1)}{4}} \prod_{i=1}^n \Gamma \left(\frac{p-i+1}{2} \right), \quad (\text{E.2})$$

et $\Gamma(\cdot)$ est la fonction Gamma.

Notons que la matrice $n \times n$ aléatoire \mathbf{B} suit une loi inverse Wishart $\mathcal{IW}(\beta, \Psi^{-1})$ si la matrice inverse \mathbf{B}^{-1} est distribuée suivant la loi de Wishart $\mathcal{W}(p, \Psi)$ avec $p = \beta - n - 1$.

E.2 Principales propriétés

Si $\mathbf{A} \sim \mathcal{W}(p, \Psi)$, alors la moyenne et la variance de la matrice $n \times n$ aléatoire \mathbf{A} peuvent facilement être exprimées grâce aux moments donnés dans [LM04] :

- $E[\mathbf{A}] = p\Psi$,
- $\text{var}[\mathbf{A}] = p[\text{tr}\{\Psi\} \mathbf{I}_L + \Psi] \Psi$.

ANNEXE F

Calcul de la loi *a posteriori* de la matrice de covariance

Dans cette annexe, nous rapportons les calculs permettant d'exprimer la loi *a posteriori* de la matrice de covariance Σ donnée dans (3.27). En utilisant le théorème de Bayes, la loi *a posteriori* $f(\Sigma^{-1}|\gamma, \alpha, \mathbf{y})$ s'écrit :

$$\begin{aligned} f(\Sigma^{-1}|\gamma, \alpha, \mathbf{y}) &\propto f(\mathbf{y}|\alpha, \Sigma^{-1}) f(\Sigma^{-1}|\gamma) \\ &\propto |\Sigma^{-1}|^{\frac{1}{2}} \exp\left[-\frac{1}{2} \|\mathbf{y} - \mathbf{M}^+ \alpha^+\|_{\Sigma^{-1}}^2\right] \\ &\quad \times |\Sigma^{-1}|^{\frac{\nu-L-1}{2}} \text{etr}\left\{-\frac{\nu}{2} \mathbf{D}^{-1} \Sigma^{-1}\right\}, \end{aligned} \quad (\text{F.1})$$

avec $\mathbf{D} = \frac{1}{\gamma} \mathbf{I}_L$. En notant $\mathbf{z} = \mathbf{y} - \mathbf{M}^+ \alpha^+$ et réécrivant $\mathbf{z}^T \Sigma^{-1} \mathbf{z} = \text{tr}\{\mathbf{z} \mathbf{z}^T \Sigma^{-1}\}$, il vient simplement :

$$f(\Sigma^{-1}|\gamma, \alpha, \mathbf{y}) \propto |\Sigma^{-1}|^{\frac{\nu-L}{2}} \text{etr}\left\{-\frac{1}{2} [\gamma \nu \mathbf{I}_L + \mathbf{z} \mathbf{z}^T] \Sigma^{-1}\right\}. \quad (\text{F.2})$$

Par conséquent,

$$\Sigma^{-1}|\gamma, \alpha, \mathbf{y} \sim \mathcal{W}\left(\nu + 1, [\gamma \nu \mathbf{I}_L + \mathbf{z} \mathbf{z}^T]^{-1}\right). \quad (\text{F.3})$$

ANNEXE G

Calcul des probabilités d'acceptation des mouvements de naissance et de mort pour l'algorithme hybride de démélange spectral

Nous présentons dans cette partie annexe le détail des calculs permettant d'exprimer les probabilités d'acceptation des mouvements de « naissance » et de « mort » introduits au paragraphe 3.8. A l'itération t , considérons le mouvement de naissance de l'état $\{\boldsymbol{\alpha}^{(t)}, \mathbf{M}^{+(t)}, R^{(t)}\}$ au nouvel état $\{\boldsymbol{\alpha}^*, \mathbf{M}^{+*}, R^*\}$ avec $\boldsymbol{\alpha}^* = [(1 - w^*)\alpha_1, \dots, (1 - w^*)\alpha_{R^{(t)}}]^\top$, $\mathbf{M}^{+*} = [\mathbf{M}^{+(t)}, s^*]$ et $R^* = R^{(t)} + 1$. Le rapport d'acceptation pour ce mouvement de naissance est :

$$A_b = \frac{f(\boldsymbol{\alpha}^*, \mathbf{M}^{+*}, R^* | \mathbf{y}) p_{R^* \rightarrow R^{(t)}}}{f(\boldsymbol{\alpha}^{(t)}, \mathbf{M}^{+(t)}, R^{(t)} | \mathbf{y}) p_{R^{(t)} \rightarrow R^*}} \times \frac{q(\mathbf{M}^{+(t)}, \boldsymbol{\alpha}^{(t)} | \mathbf{M}^{+*}, \boldsymbol{\alpha}^*)}{q(\mathbf{M}^{+*}, \boldsymbol{\alpha}^* | \mathbf{M}^{+(t)}, \boldsymbol{\alpha}^{(t)})} |J(w^*)|, \quad (\text{G.1})$$

où $q(\cdot | \cdot)$ fait référence aux lois de proposition, $|J(w^*)|$ est le Jacobien de la transformation et p_{\dots} est la probabilité de transition, i.e. $p_{R^* \rightarrow R^{(t)}} = d_{R^*}$ et $p_{R^{(t)} \rightarrow R^*} = b_{R^{(t)}}$. D'après les mouvements précisés au paragraphe 3.8, le rapport des lois de propositions est :

$$\frac{q(\mathbf{M}^{+(t)}, \boldsymbol{\alpha}^{+(t)} | \mathbf{M}^{+*}, \boldsymbol{\alpha}^{+*})}{q(\mathbf{M}^{+*}, \boldsymbol{\alpha}^{+*} | \mathbf{M}^{+(t)}, \boldsymbol{\alpha}^{+(t)})} = \frac{1}{g_{1, R^{(t)}}(w^*)} \frac{R_{\max} - R^{(t)}}{R^{(t)} + 1}, \quad (\text{G.2})$$

où $g_{a,b}(\cdot)$ est la fonction densité de probabilité d'une loi Beta $\mathcal{B}e(a, b)$. La probabilité de choisir un nouvel élément dans la bibliothèque (mouvement de naissance) est en effet $1 / (R_{\max} - R^{(t)})$ et la

probabilité de supprimer un élément (mouvement de mort) est $1 / (R^{(t)} + 1)$.

Le rapport des lois *a posteriori* qui apparaît dans (G.1) peut être réécrit :

$$\frac{f(\boldsymbol{\alpha}^*, \mathbf{M}^{+*}, R^* | \mathbf{y})}{f(\boldsymbol{\alpha}^{(t)}, \mathbf{M}^{+(t)}, R^{(t)} | \mathbf{y})} = \frac{f(\mathbf{y} | \boldsymbol{\alpha}^*, \mathbf{M}^{+*}, R^*)}{f(\mathbf{y} | \boldsymbol{\alpha}^{(t)}, \mathbf{M}^{+(t)}, R^{(t)})} \times \frac{f(\boldsymbol{\alpha}^* | R^*)}{f(\boldsymbol{\alpha}^{(t)} | R^{(t)})} \frac{f(\mathbf{M}^{+*} | R^*)}{f(\mathbf{M}^{+(t)} | R^{(t)})} \frac{f(R^*)}{f(R^{(t)})}. \quad (\text{G.3})$$

Puisque la loi *a priori* du vecteur des coefficients d'abondance $\boldsymbol{\alpha}^+$ est une loi de Dirichlet $\mathcal{D}_R(\delta, \dots, \delta)$, le rapport des lois *a priori* est :

$$\frac{f(\boldsymbol{\alpha}^* | R^*)}{f(\boldsymbol{\alpha}^{(t)} | R^{(t)})} = \frac{\Gamma(\delta R^{(t)} + \delta)}{\Gamma(\delta R^{(t)}) \Gamma(\delta)} \times w^{*\delta-1} (1 - w^*)^{(\delta-1)R^{(t)}}. \quad (\text{G.4})$$

En choisissant des configurations équiprobables *a priori* pour \mathbf{M}^+ conditionnellement à R , le rapport des lois *a priori* de la matrice des spectres est :

$$\frac{f(\mathbf{M}^{+*} | R^*)}{f(\mathbf{M}^{+(t)} | R^{(t)})} = \frac{\binom{R_{\max}}{R^{(t)}}}{\binom{R_{\max}}{R^*}} = \frac{R^{(t)} + 1}{R_{\max} - R}. \quad (\text{G.5})$$

Le rapport des lois *a priori* uniformes liées au nombre de composants dans le mélange R associated se réduit à 1.

Finalement, le rapport d'acceptation pour le mouvement de NAISSANCE peut s'écrire :

$$A_b = \exp \left[- \frac{\|\mathbf{y} - \mathbf{M}^{+*} \boldsymbol{\alpha}^{+*}\|^2 - \|\mathbf{y} - \mathbf{M}^{+(t)} \boldsymbol{\alpha}^{+(t)}\|^2}{2} \right] \times \frac{d_{R^{(t)}+1}}{b_{R^{(t)}}} \frac{1}{g_{1,R^{(t)}}(w^*)} (1 - w^*)^{R^{(t)}-1} \times \frac{\Gamma(\delta R^{(t)} + \delta)}{\Gamma(\delta R^{(t)}) \Gamma(\delta)} w^{*\delta-1} (1 - w^*)^{(\delta-1)R^{(t)}}, \quad (\text{G.6})$$

Nous remarquons alors que l'équation (G.6) est bien sûr similaire à l'équation donnée dans [RG98] et que $\delta = 1$ quand $\boldsymbol{\alpha}$ a une loi *a priori* uniforme sur le simplexe \mathbb{S} .

Liste des publications

Articles de journaux

1. N. Dobigeon, J.-Y. Tournet and C.-I Chang, “Semi-supervised linear spectral using a hierarchical Bayesian model for hyperspectral imagery,” *IEEE Trans. Signal Processing*, under 2nd review, 2007
2. N. Dobigeon and J.-Y. Tournet, “Joint segmentation of wind speed and direction using a hierarchical model,” *Computational Statistics and Data Analysis*, vol. 51, no. 12, pp. 5603–5621, Aug. 2007. [Online]. Available : <http://dx.doi.org/10.1016/j.csda.2007.04.016>
3. N. Dobigeon, J.-Y. Tournet and M. Davy, “Joint segmentation of piecewise constant autoregressive processes by using a hierarchical model and a Bayesian sampling approach,” *IEEE Trans. Signal Processing*, vol. 55, no. 4, pp. 1251–1263, April 2007.
4. N. Dobigeon, J.-Y. Tournet and J. Scargle, “Joint segmentation of multivariate astronomical time series : Bayesian sampling with a hierarchical model,” *IEEE Trans. Signal Processing*, vol. 55, no. 2, pp. 414–423, Feb. 2007.

Articles de conférence

1. N. Dobigeon, J.-Y. Tournet and S. Moussaoui, “Blind unmixing of linear mixtures using a hierarchical Bayesian model. Application to spectroscopic signal analysis,” in *Proc. IEEE-SP Workshop Stat. and Signal Processing*, Madison, USA, Aug. 2007.
2. N. Dobigeon, S. Moussaoui and J.-Y. Tournet, “Séparation bayésienne de sources spectrales sous contraintes de positivité et d’additivité,” in *Actes du XXIème Colloque GRETSI*, Troyes,

- France, Sept. 2007, in french.
3. N. Dobigeon and J.-Y. Tourneret, “Spectral unmixing of hyperspectral images using a hierarchical Bayesian model,” in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Processing (ICASSP)*, vol. 3, Honolulu, Hawaii, USA, April 2007, pp. III1209–III1212.
 4. N. Dobigeon, J.-Y. Tourneret and J. D. Scargle, “Joint segmentation of multivariate poissonian time series. applications to burst and transient source experiments,” in *Proc. European Signal Processing Conf. (EUSIPCO)*, Florence, Italy, Sept. 2006.
 5. N. Dobigeon, J.-Y. Tourneret and M. Davy, “Joint segmentation of piecewise constant autoregressive processes by using a hierarchical model and a Bayesian sampling approach,” in *Proc. IEEE Int. Conf. Acoust., Speech and Signal Processing (ICASSP)*, vol. 3, Toulouse, France, May 2006, pp. III1–III4.
 6. N. Dobigeon and V. Achard, “Performance comparison of geometric and statistical methods for endmembers extraction in hyperspectral imagery,” in *Image and Signal Processing for Remote Sensing XI*, L. Bruzzone, Ed., vol. 5982, no. 1. SPIE, Oct. 2005, pp. 335–344. [Online]. Available : <http://link.aip.org/link/?PSI/5982/598213/1>
 7. N. Dobigeon and J.-Y. Tourneret, “Segmentation conjointe de données poissonniennes,” in *Actes du XXIème Colloque GRETSI*, Louvain-la-Neuve, Belgique, Sept. 2005, in french.
 8. N. Dobigeon, J.-Y. Tourneret and J. D. Scargle, “Change-point detection in astronomical data by using a hierarchical model and a bayesian sampling approach,” in *Proc. IEEE-SP Workshop Stat. and Signal Processing*, Bordeaux, France, July 2005, pp. 369–374.

Rapports techniques

1. N. Dobigeon and J.-Y. Tourneret, “Efficient sampling according to a multivariate Gaussian distribution truncated on a simplex,” IRIT/ENSEEIH/TeSA, Tech. Rep., March 2007.
2. N. Dobigeon, J.-Y. Tourneret and C.-I Chang, “Semi-supervised linear spectral using a hierarchical Bayesian model for hyperspectral imagery,” IRIT/ENSEEIH/TeSA, Tech. Rep., March 2007.

3. N. Dobigeon and J.-Y. Tourneret, "Truncated multivariate Gaussian distribution on a simplex," IRIT/ENSEEIH/TeSA, Tech. Rep., Jan. 2007.
4. N. Dobigeon, J.-Y. Tourneret and M. Davy, "Joint segmentation of piecewise constant autoregressive processes by using a hierarchical model and a Bayesian sampling approach," IRIT/ENSEEIH/TeSA and LAGIS, Tech. Rep., March 2006.

Bibliographie

- [AAM05] T. Akgun, Y. Altunbasak, and R. M. Mersereau. Super-resolution reconstruction of hyperspectral images. *IEEE Trans. Image Processing*, 14(11) :1860–1875, Nov. 2005.
- [AB83] U. Appel and A. V. Brandt. Adaptive sequential segmentation of piecewise stationary time series. *Inform. Sci.*, 29(1) :27–56, 1983.
- [ABD01] C. Andrieu, E. Barat, and A. Doucet. Bayesian deconvolution of noisy filtered point processes. *IEEE Trans. Signal Processing*, 49(1) :134–146, Jan. 2001.
- [AD99] C. Andrieu and A. Doucet. Joint Bayesian model selection and estimation of noisy sinusoids via reversible jump MCMC. *IEEE Trans. Signal Processing*, 47(10) :19–37, Oct. 1999.
- [ADD01] C. Andrieu, P. M. Djuric, and A. Doucet. Model selection by MCMC computation. *Signal Processing*, 81(1) :19–37, Jan. 2001.
- [Aka74] H. Akaike. A new look at the statistical model identification. *IEEE Trans. Autom. Contr.*, 19(6) :716–723, Dec. 1974.
- [And88] R. André-Obrecht. A new statistical approach for automatic segmentation of continuous speech. *IEEE Trans. Acoust., Speech, Signal Processing*, 36 :29–40, Jan. 1988.
- [AVI06] AVIRIS Free Data, 2006, Jet Propulsion Lab. (JPL), California Inst. Technol., Pasadena, CA. Available online at <http://aviris.jpl.nasa.gov/html/aviris.freedata.html>.

- [Bar78] O. E. Barndorff-Nielsen. *Information and Exponential Families in Statistical Theory*. Wiley, New-York, 1978.
- [BB83] M. Basseville and A. Benveniste. Design and comparative study of some sequential jump detection algorithms for digital signals. *IEEE Trans. Acoust., Speech, Signal Processing*, 31 :521–535, 1983.
- [BCA95] R. W. Basedow, D. C. Carmer, and M. E. Anderson. HYDICE system : implementation and performance. In M. R. Descour, J. M. Mooney, D. L. Perry, and L. R. Illing, editors, *Imaging Spectrometry*, volume 2480, pages 258–267. SPIE, 1995.
- [BD90] P. J. Brockwell and R. A. Davis. *Time Series : Theory and Methods*. Springer Verlag, New York, 1990.
- [BD93] B. Brodsky and B. Darkhovsky. *Nonparametric methods in change-point problems*. Kluwer Academic Publishers, Boston (MA), 1993.
- [BD07] T. Blumensath and M. E. Davies. Monte-Carlo methods for adaptive sparse approximations of time-series. *IEEE Trans. Signal Processing*, 2007. to be published.
- [BKG95] J. W. Boardman, F. A. Kruse, and R. O. Green. Mapping target signatures via partial unmixing of AVIRIS data. In *Sum. of the Fifth JPL Airborne Earth Science Workshop*, *JPL Publication 95-1*, pages 23–26, 1995.
- [BKL⁺04] M. Berman, H. Kiiveri, R. Lagerstrom, A. Ernst, R. Dunne, and J. F. Huntington. ICE : A statistical approach to identifying endmembers in hyperspectral images. *IEEE Trans. Geosci. and Remote Sensing*, 42(10) :2085–2095, Oct. 2004.
- [BL04] A. M. Bruning and W. G Linzey. High voltage gradients in aircraft low voltage systems. In *Proc. Int. Power Modulator Conf.*, pages 15–16, San Francisco, CA, USA, May 2004.
- [BM01] L. Birgé and P. Massart. Gaussian model selection. *Jour. Eur. Math. Soc.*, 3 :203–268, 2001.

- [BN93] M. Basseville and I. V. Nikiforov. *Detection of Abrupt Changes : Theory and Application*. Prentice-Hall, Englewood Cliffs NJ, 1993.
- [C.-03] C.-I Chang. *Hyperspectral imaging : techniques for spectral detection and classification*. Kluwer Academic, New York, 2003.
- [Cap02] O. Cappé. A Bayesian approach for simultaneous segmentation and classification of count data. *IEEE Trans. Signal Processing*, 50(2) :400–410, Feb. 2002.
- [CD96] M.-H. Chen and J. J. Deely. Bayesian analysis for a constrained linear multiple regression problem for predicting the new crop of apples. *J. of Agricultural, Biological, and Environmental Stat.*, 1 :467–489, Dec. 1996.
- [CGS92] B. P. Carlin, A. E. Gelfand, and A. F. M. Smith. Hierarchical Bayesian analysis of changepoint problems. *Appl. Statist.*, 41(2) :389–405, 1992.
- [Che05] F. W. Chen. Archiving and distribution of 2-D geophysical data using image formats with lossless compression. *IEEE Geosci. and Remote Letters*, 2(1) :64–68, Jan. 2005.
- [Chi98] S. Chib. Estimation and comparison of multiple change-point models. *J. of Econometrics*, 86(2) :221–241, Feb. 1998.
- [CJ01] C.-I Chang and B. Ji. Weighted abundance-constrained linear spectral mixture analysis. *IEEE Trans. Geosci. and Remote Sensing*, 44(2) :378–388, Feb. 2001.
- [CJS+98] T. Cocks, R. Jenssen, A. Stewart, I. Wilson, and T. Shields. The HYMAP airborne hyperspectral sensor : the system, calibration and performance. In *Proc. First EARSeL Workshop on Imaging Spectroscopy*, pages 37–42, Oct. 1998.
- [CLM05] E. Christophe, D. Léger, and C. Mailhes. Quality criteria benchmark for hyperspectral imagery. *IEEE Trans. Geosci. and Remote Sensing*, 43(9) :2103–2114, Sept. 2005.
- [Cra94] M. D. Craig. Minimum volume transforms for remotely sensed data. *IEEE Trans. Geosci. and Remote Sensing*, 32(3) :542–552, 1994.

- [CTC03] M. Chabert, J.-Y. Tourneret, and M. Coulon. Joint detection of variance changes using hierarchical Bayesian analysis. In *Proc. IEEE-SP Workshop Stat. and Signal Processing*, pages 593–596, Saint-Louis, Missouri, Sept. 2003.
- [CWL⁺06] F. Chaudhry, C.-C. Wu, W. Liu, C.-I Chang, and A. Plaza. Pixel purity index-based algorithms for endmember extraction from hyperspectral imagery. In C.-I Chang, editor, *Recent Advances in Hyperspectral Signal and Image Processing*, chapter 2. Research Signpost, Trivandrum, Kerala, India, 2006.
- [DC02] P. M. Djurić and J.-H. Chun. An MCMC sampling approach to estimation of non-stationary hidden Markov models. *IEEE Trans. Signal Processing*, 50(5) :1113–1123, 2002.
- [DDD05] F. Desobry, M. Davy, and C. Doncarli. An online kernel change detection algorithm. *IEEE Trans. Signal Processing*, 53(8) :2961–2974, Aug. 2005.
- [DDT02] M. Davy, C. Doncarli, and J.-Y. Tourneret. Classification of chirp signals using hierarchical Bayesian learning and MCMC methods. *IEEE Trans. Signal Processing*, 50(2) :377–388, Feb. 2002.
- [del94] J. del Castillo. The singly truncated normal distribution : a non-steep exponential family. *Annals of the Institute of Statistical Mathematics*, 46(1) :57–66, March 1994.
- [DF98] M. Daumer and M. Falk. On-line change-point detection (for state space models) using multi-process Kalman filters. *Linear Algebra and its Applications*, 284(1–3) :125–135, 1998.
- [DGI06] M. Davy, S. Godsill, and J. Idier. Bayesian analysis of polyphonic western tonal music. *J. Acoust. Soc. Amer.*, 119(4) :2498–2517, April 2006.
- [DHMS02] D. G. T. Denison, C. C. Holmes, B. K. Mallick, and A. F. M. Smith. *Bayesian methods for nonlinear classification and regression*. Wiley, Chichester, England, 2002.

- [DI96] J. Diebolt and E. H. S. Ip. Stochastic EM : method and application. In W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, editors, *Markov Chain Monte Carlo in Practice*, pages 259–273. Chapman & Hall, London, 1996.
- [DI04] M. Davy and J. Idier. Fast MCMC computations for the estimation of sparse processes from noisy observations. In *Int. Conf. Acoust., Speech, and Signal Processing (ICASSP)*, volume 2, pages 1041–1044, Montréal, Quebec, Canada, May 2004.
- [Dju94] P. M. Djurić. A MAP solution to off-line segmentation of signals. In *Int. Conf. Acoust., Speech, and Signal Processing (ICASSP)*, volume 4, pages 505–508, April 1994.
- [DMN⁺03] B. Datt, T. R. McVicar, T. G. Van Niel, D. L. B. Jupp, and J. S. Pearlman. Preprocessing EO-1 Hyperion hyperspectral data to support the application of agricultural indexes. *IEEE Trans. Geosci. and Remote Sensing*, 41(6) :1246–1259, June 2003.
- [DMT07] N. Dobigeon, S. Moussaoui, and J.-Y. Tournéret. Séparation bayésienne de sources spectrales sous contraintes de positivité et d’additivité. In *Actes du XXIème Colloque GRETSI*, Troyes, France, Sept. 2007. in French.
- [DT07] N. Dobigeon and J.-Y. Tournéret. Efficient sampling according to a multivariate Gaussian distribution truncated on a simplex. Technical report, IRIT/ENSEEIH/TESA, March 2007.
- [DTM07] N. Dobigeon, J.-Y. Tournéret, and S. Moussaoui. Blind unmixing of linear mixtures using a hierarchical Bayesian model. Application to spectroscopic signal analysis. In *Proc. IEEE-SP Workshop Stat. and Signal Processing*, Madison, USA, Aug. 2007.
- [DTS06] N. Dobigeon, J.-Y. Tournéret, and J. D. Scargle. Joint segmentation of multivariate Poissonian time series applications to burst and transient source experiments. In *Proc. European Signal Processing Conf. (EUSIPCO)*, Florence, Italy, Sept. 2006.
- [Fea05] P. Fearnhead. Exact Bayesian curve fitting and signal segmentation. *IEEE Trans. Signal Processing*, 53(6) :2160–2166, June 2005.

- [Fea06] P. Fearnhead. Exact and efficient Bayesian inference for multiple changepoint problems. *Statistics and Computing*, 16(2) :203–213, June 2006.
- [FG06] C. Févotte and S. J. Godsill. A Bayesian approach for blind separation of sparse sources. *IEEE Trans. Audio, Speech, Language Processing*, 14(6) :2174–2188, Nov. 2006.
- [FH01] C. Furse and R. Haupt. Down to the wire. *IEEE Spectrum*, 38(2) :35–39, Feb. 2001.
- [FL07] P. Fearnhead and Z. Liu. On-line inference for multiple change points problems. to appear in *J. Royal Statist. Soc., Series B*, 2007.
- [GCSR95] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian Data Analysis*. Chapman & Hall, London, 1995.
- [GES⁺98] R. O. Green, M. L. Eastwood, C. M. Sarture, T. G. Chrien, M. Aronsson, B. J. Chippendale, J. A. Faust, B. E. Pavri, C. J. Chovit, M. Solis, M. R. Olah, and O. Williams. Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (AVIRIS). *Remote Sens. Environ.*, 65(3) :227–248, Sept. 1998.
- [GR92] A. Gelman and D.B. Rubin. Inference from iterative simulation using multiple sequences. *Statist. Science*, 7(4) :457–511, 1992.
- [GR98] S. J. Godsill and P. J. W. Rayner. Statistical reconstruction and analysis of autoregressive signals in impulsive noise using the Gibbs sampler. *IEEE Trans. Speech, Audio Processing*, 6(4) :352–372, 1998.
- [Gre95] P. J. Green. Reversible jump MCMC computation and Bayesian model determination. *Biometrika*, 82(4) :711–732, Dec. 1995.
- [GRS96] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter. Introducing Markov Chain Monte Carlo. In W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, editors, *Markov Chain Monte Carlo in Practice*, pages 1–19. Chapman & Hall, London, 1996.
- [Gue94] D. Guegan. *Séries chronologiques non linéaires à temps discret*. Economica, Paris, 1994.

- [Gus00] F. Gustafsson. *Adaptive filtering and change detection*. Wiley, New-York, 2000.
- [Hap81] B. W. Hapke. Bidirectional reflectance spectroscopy. I. Theory. *J. Geophys. Res.*, 86 :3039–3054, 1981.
- [HC01] D. C. Heinz and C.-I Chang. Fully constrained least-squares linear spectral mixture analysis method for material quantification in hyperspectral imagery. *IEEE Trans. Geosci. and Remote Sensing*, 29(3) :529–545, March 2001.
- [HK97] J. Hanson and D. Koenig. Fault arc effects in cable bundles for space applications in vacuum. *IEEE Trans. Dielectrics and Electrical Insulation*, 4(4) :394–399, Aug. 1997.
- [IT94] C. Inclan and G. C. Tiao. Use of cumulative sums of squares for retrospective detection of changes of variance. *J. Am. Stat. Assoc.*, 89 :913–923, 1994.
- [JJ00] T. S. Jaakola and M. Jordan. Bayesian parameter estimation via variational methods. *Statistics and Computing*, 10(1) :25–37, Jan. 2000.
- [JL02] Q. Jackson and D. A. Landgrebe. An adaptive method for combined covariance estimation and classification. *IEEE Trans. Geosci. and Remote Sensing*, 40(5) :1082–1087, May 2002.
- [JSB⁺05] B. Jackson, J. D. Scargle, D. Barnes, S. Arabhi, A. Alt, P. Gioumousis, E. Gwin, P. Sangtrakulcharoen, L. Tan, and T. T. Tsai. An algorithm for optimal partitioning of data on an interval. *IEEE Signal Processing Lett.*, 12(2) :105–108, Feb. 2005.
- [JSTA83] P. E. Johnson, M. O. Smith, S. Taylor-George, and John B. Adams. A semiempirical method for analysis of the reflectance spectra of binary mineral mixtures. *J. Geophys. Res.*, 88 :3557–3561, 1983.
- [Kay88] S. M. Kay. *Modern spectral estimation*. Prentice Hall, Englewood Cliffs NJ, 1988.
- [Kay93] S. M. Kay. *Fundamentals of Statistical Signal Processing : Estimation theory*. Prentice Hall, Englewood Cliffs NJ, 1993.

- [KL04] E. Kuhn and M. Lavielle. Coupling a stochastic approximation version of EM with an MCMC procedure. *ESAIM Probab. Statist.*, 8 :115–131, 2004.
- [KM02] N. Keshava and J. F. Mustard. Spectral unmixing. *IEEE Signal Processing Magazine*, 19(1) :44–57, Jan. 2002.
- [Lav98] M. Lavielle. Optimal segmentation of random processes. *IEEE Trans. Signal Processing*, 46(5) :1365–1373, May 1998.
- [Lav99] M. Lavielle. Detection of multiple changes in a sequence of dependent variables. *Stochastic Processes and their Applications*, 83(1) :79–102, Sept. 1999.
- [Lav05] M. Lavielle. Using penalized contrasts for the change-point problem. *Signal Processing*, 85(8) :1501–1510, Aug. 2005.
- [Leb05] E. Lebarbier. Detecting multiple change-points in the mean of Gaussian process by model selection. *Signal Processing*, 85(4) :717–736, April 2005.
- [Let92] G. Letac. *Lectures on natural exponential families and their variance functions*, volume 50. Instituto de Matemática Pura e Aplicada : Monografias de Matemática, Rio de Janeiro, Brazil, 1992.
- [LF05] C. Lo and C. Furse. Noise-domain reflectometry for locating wiring faults. *IEEE Trans. Electromagn. Compat.*, 47(1) :97–104, Feb. 2005.
- [LJdG01] M. Lewis, V. Jooste, and A. A. de Gasparis. Discrimination of arid vegetation with airborne multispectral scanner hyperspectral imagery. *IEEE Trans. Geosci. and Remote Sensing*, 39(7) :1471–1479, July 2001.
- [LL00] M. Lavielle and C. Ludena. The multiple change-points problem for the spectral distribution. *Bernoulli*, 6(5) :845–869, Oct. 2000.
- [LL01] M. Lavielle and E. Lebarbier. An application of MCMC methods for the multiple change-points problem. *Signal Processing*, 81(1) :39–53, Jan. 2001.

- [LM99] M. Lavielle and D. Marquez. Gibbs sampling for parameters and change-points detection in inverse problem. Application to electromagnetic imaging. *Signal Processing*, 78(3) :349–362, Nov. 1999.
- [LM00] M. Lavielle and E. Moulines. Least squares estimation of an unknown number of shifts in a time series. *Jour. of Time Series Anal.*, 21(1) :33–59, Jan. 2000.
- [LM04] G. Letac and H. Massam. All invariant moments of the Wishart distribution. *Scandinavian Journal of Statistics*, 31(2) :295–318, June 2004.
- [LP01] D. Leporini and J.-C. Pesquet. Bayesian wavelet denoising : Besov priors and non-Gaussian noises. *Signal Processing*, 81(1) :55–67, Jan. 2001.
- [LR02] J.-R. Larocque and J. P. Reilly. Reversible jump MCMC for joint detection and estimation of sources in colored noise. *IEEE Trans. Signal Processing*, 50(2) :231–240, Feb. 2002.
- [LT06] M. Lavielle and G. Teyssière. Detection of multiple change-points in multivariate time series. *Lithuanian Mathematical Journal*, 46(3) :287–306, July 2006.
- [Mal73] C. L. Mallows. Some comments on Cp. *Technometrics*, 15(4) :661–675, Nov. 1973.
- [MBMC06] S. Moussaoui, D. Brie, A. Mohammad-Djafari, and C. Carteret. Separation of non-negative mixture of non-negative sources using a Bayesian approach and MCMC sampling. *IEEE Trans. Signal Processing*, 54(11) :4133–4145, Nov. 2006.
- [MCBMD06] S. Moussaoui, C. Carteret, D. Brie, and A. Mohammad-Djafari. Bayesian analysis of spectral mixture data using Markov chain Monte Carlo methods. *Chemometrics and Intelligent Laboratory Systems*, 81(2) :137–148, April 2006.
- [MSS01] D. Manolakis, C. Siracusa, and G. Shaw. Hyperspectral subpixel target detection using the linear mixing model. *IEEE Trans. Geosci. and Remote Sensing*, 39(7) :1392–1409, July 2001.

- [MT93] R. E. McCulloch and R. S. Tsay. Bayesian inference and prediction for mean and variance shifts in autoregressive time series. *J. Am. Stat. Assoc.*, 88(423) :968–978, 1993.
- [Mui05] R. J. Muirhead. *Aspects of Multivariate Statistical Theory*. Wiley, New-Jersey, 2005.
- [Nat00] National Transportation Safety Board. In-flight breakup over the Atlantic Ocean, Trans World Airlines Flight 800, Boeing 747-131, N93119, Near East Moriches, New York, July 17, 1996. Aircraft Accident Report NTSB/AAR-00/03, National Transportation Safety Board, Aug. 2000. Available online at <http://www.nts.gov/Publicctn/2000/AAR0003.pdf>.
- [ND05] J. M. P. Nascimento and J. M. Bioucas Dias. Vertex component analysis : A fast algorithm to unmix hyperspectral data. *IEEE Trans. Geosci. and Remote Sensing*, 43(4) :898–910, April 2005.
- [Nik01] I. V. Nikiforov. A simple change detection scheme. *Signal Processing*, 81(1) :149–172, Jan. 2001.
- [PADF02] E. Punskeya, C. Andrieu, A. Doucet, and W. Fitzgerald. Bayesian curve fitting using MCMC with applications to signal segmentation. *IEEE Trans. Signal Processing*, 50(3) :747–758, March 2002.
- [PBS⁺03] J. S. Pearlman, P. S. Barry, C. C. Segal, J. Shepanski, D. Beiso, and S. L. Carman. Hyperion, a space-based imaging spectrometer. *IEEE Trans. Geosci. and Remote Sensing*, 41(6) :1160–1173, June 2003.
- [PMP⁺99] W. S. Paciesas, C. A. Meegan, G. N. Pendleton, M. S. Briggs, C. Kouveliotou, T. M. Koshut, J. P. Lestrade, M. L. McCollough, J. J. Brainerd, J. Hakkila, W. Henze, R. D. Preece, V. Connaughton, R. M. Kippen, R. S. Malozzi, G. J. Fishman, G. A.

- Richardson, and M. Sahi. The fourth BATSE burst revised catalog. *Astrophysical Journal Supplement Series*, 122 :465–497, June 1999. Data available online at <http://coss.gsfc.nasa.gov/batse/>.
- [RB06] S. Reboul and M. Benjelloun. Joint segmentation of the wind speed and direction. *Signal Processing*, 86(4) :744–759, April 2006.
- [RC99] C. P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer-Verlag, New-York, 1999.
- [RDFZ04] G. Rellier, X. Descombes, F. Falzon, and J. Zerubia. Texture feature analysis using a Gauss Markov model in hyperspectral image classification. *IEEE Trans. Geosci. and Remote Sensing*, 42(7) :1543–1551, July 2004.
- [RG97] S. Richardson and P. J. Green. On Bayesian analysis of mixtures with unknown number of components. *J. Roy. Stat. Soc. B*, 59(4) :731–792, 1997.
- [RG98] S. Richardson and P. J. Green. Corrigendum : On Bayesian analysis of mixtures with unknown number of components. *J. Roy. Stat. Soc. B*, 60(3) :661, 1998.
- [Rob92] C. P. Robert. *L'analyse statistique bayésienne*. Economica, Paris, 1992.
- [Rob95] C. P. Robert. Simulation of truncated normal variables. *Statistics and Computing*, 5(2) :121–125, June 1995.
- [Rot02] R. Rotondi. On the influence of the proposal distributions on a reversible jump MCMC algorithm applied to the detection of multiple change-points. *Computational Statistics & Data Analysis*, 40(3) :633–653, Sept. 2002.
- [RR98] C. P. Robert and S. Richardson. Markov Chain Monte Carlo methods. In C. P. Robert, editor, *Discretization and MCMC Convergence Assessment*, pages 1–25. Springer Verlag, New York, 1998.

- [RSI03] RSI (Research Systems Inc.). *ENVI User's guide Version 4.0*. Boulder, CO 80301 USA, Sept. 2003.
- [RT02] B. K. Ray and R. S. Tsay. Bayesian methods for change-point detection in long-range dependent processes. *Jour. of Time Series Anal.*, 23(6) :687–705, Nov. 2002.
- [RYDS02] G. Rodriguez-Yam, R. A. Davis, and L. Scharf. A Bayesian model and Gibbs sampler for hyperspectral imaging. In *Proc. IEEE Sensor Array and Multichannel Signal Processing Workshop*, pages 105–109, Washington, D.C., Aug. 2002.
- [Sca98] J. D. Scargle. Studies in astronomical time series analysis : v. Bayesian blocks, a new method to analyze structure in photon counting data. *The Astrophysical Journal*, 504 :405–418, Sept. 1998.
- [Sch78] G. Schwarz. Estimating the dimension of a model. *Ann. Statist.*, 6(2) :461–464, March 1978.
- [SI06] H. Snoussi and J. Idier. Bayesian blind separation of generalized hyperbolic processes in noisy and underdeterminate mixtures. *IEEE Trans. Signal Processing*, 54(9) :3257–3269, Sept. 2006.
- [SM79] R. B. Singer and T. B. McCord. Mars : Large scale mixing of bright and dark surface materials and implications for analysis of spectral reflectance. In *Proc. Lunar Planet. Sci. Conf.*, volume 2, pages 1835–1848, 1979.
- [SNJ06] J. D. Scargle, J. Norris, and B. Jackson. Studies in astronomical time series analysis : vi. optimal segmentation : blocks, histograms and triggers. *in preparation*, 2006.
- [SS97] A. D. Strocker and P. Schaum. Application of stochastic mixing models to hyperspectral detection problems. In *Proc. SPIE, Algorithms for Multispectral and Hyperspectral Imagery III*, volume 3071, pages 47–60, Orlando, FL, 1997.

- [TCC98] T. M. Tu, C. H. Chen, and C.-I Chang. A noise subspace projection approach to target signature detection and extraction in an unknown background for hyperspectral images. *IEEE Trans. Geosci. and Remote Sensing*, 36(1) :171–181, Jan. 1998.
- [TDL03] J.-Y. Tourneret, M. Doisy, and M. Lavielle. Bayesian retrospective detection of multiple changepoints corrupted by multiplicative noise. application to SAR image edge detection. *Signal Processing*, 83(9) :1871–1887, Sept. 2003.
- [TG01] P. T. Troughton and S. J. Godsill. MCMC methods for restoration of nonlinearly distorted autoregressive signals. *Signal Processing*, 81(1) :83–97, Jan. 2001.
- [TP04] X. Tang and W. A. Pearlman. Lossy-to-lossless block-based compression of hyperspectral volumetric data. In *Int. Conf. Image Processing (ICIP)*, volume 5, pages 3283–3286, Oct. 2004.
- [TS01] C. Teal and D. Sorensen. Health management for avionics performance. In *Proc. Digital Avionics Systems Conf. (DASC)*, volume 1, pages 3B3/1 – 3B3/5, Daytona Beach, FL, USA, Oct. 2001.
- [VADG04] J. Vermaak, C. Andrieu, A. Doucet, and S. J. Godsill. Reversible jump Markov chain Monte Carlo strategies for Bayesian model selection in autoregressive processes. *Jour. of Time Series Anal.*, 25(6) :785–809, Nov. 2004.
- [VV97] R. Viswanathan and P. K. Varshney. Distributed detection with multiple sensors : Part I – Fundamentals. *Proc. of the IEEE*, 85(1) :54–63, Jan. 1997.
- [Win99] M. E. Winter. Fast autonomous spectral end-member determination in hyperspectral data. In *Proc. 13th Int. Conf. on Applied Geologic Remote Sensing*, volume 2, pages 337–344, Vancouver, April 1999.
- [WW03] Z. Wang and P. Willett. Two algorithms to segment white Gaussian data with piecewise constant variances. *IEEE Trans. Signal Processing*, 51(2) :373–385, Feb. 2003.

- [WW04] Z. J. Wang and P. Willett. Joint segmentation and classification of time series using class-specific features. *IEEE Trans. Systems, Man, Cybernet.*, 34(2) :1056–1067, Feb. 2004.