

Dans le cadre de l'Ecole Doctorale SSTO  
Sciences des Systèmes Technologiques et Organisationnels  
En partenariat avec l'Université de Technologie de Troyes

Spécialité :  
Mathématiques Appliquées

Par  
Pierre LANCHANTIN

Thèse présentée pour l'obtention du grade de Docteur  
de l'Institut National des Télécommunications

## **Chaînes de Markov Triplets et Segmentation Non Supervisée de Signaux**

Soutenue le 5 décembre 2006 devant le jury composé de :

Président	Igor NIKIFOROV	UTT, Troyes
Rapporteurs	Christophe COLLET	ULP/ENSPS, Strasbourg
	Patrick PEREZ	IRISA, Rennes
Examineur	Henri MAITRE	ENST, Paris
Directeur de thèse	Wojciech PIECZYNSKI	INT, Evry
Co-Encadrant	Fabien SALZENSTEIN	ULP, Strasbourg







# Remerciements

*Je tiens tout d'abord à exprimer ma reconnaissance à Wojciech Pieczynski, mon directeur de thèse, pour son encadrement, ses nombreux conseils et son soutien tout au long de ma thèse.*

*Je remercie également Fabien Salzenstein pour son encadrement sur la partie traitant des arbres de Markov flous multispectraux.*

*J'adresse ensuite mes remerciements aux membres du jury pour l'intérêt qu'ils ont porté à ce travail :*

- Igor Nikiforov qui m'a fait l'honneur de présider mon jury*
- Christophe Collet et Patrick Pérez qui ont accepté la tâche d'être rapporteurs, et qui ont porté des jugements très pertinents sur mon manuscrit, tant sur le fond que sur la forme et qui ont contribué à améliorer la qualité de ce mémoire.*
- Henri Maître qui a accepté d'examiner ce travail.*

*Je remercie les doctorants, chercheurs, enseignants, membres du personnel et élèves que j'ai pu côtoyer à l'INT ou à l'Université d'Orsay, pour leur sympathie et leur aide. Je tiens à témoigner tout particulièrement ma sympathie et ma reconnaissance à Nicolas avec qui j'ai partagé le même bureau pendant ces années de thèse.*

*Plus personnellement, un grand "Tack så mycket<sup>1</sup>" à Lisa, pour m'avoir supporté, aidé, motivé, encouragé pendant ces années.*

*Je termine par un grand remerciement à ma famille et surtout à mes parents pour leur confiance et leur soutien tout au long de mes études.*

---

<sup>1</sup>"merci beaucoup" en suédois.



# Table des matières

<b>Notations</b>	<b>v</b>
<b>Introduction</b>	<b>vii</b>
<b>1 Généralités</b>	<b>1</b>
1.1 Approche bayésienne du problème de la segmentation . . . . .	2
1.1.1 Modélisation probabiliste . . . . .	2
1.1.2 Stratégie bayésienne . . . . .	3
1.1.3 Estimateurs bayésiens fondés sur la fonction de perte $L_{0/1}$ . . . . .	4
1.2 Estimation des paramètres . . . . .	6
1.2.1 L'algorithme Espérance-Maximisation (EM) . . . . .	6
1.2.2 Estimation Conditionnelle Itérative (ECI) . . . . .	9
1.3 Conclusion . . . . .	11
<b>2 Modélisations de la loi jointe dans le cas d'un signal monodimensionnel</b>	<b>13</b>
2.1 Algorithmes d'inférence dans le cas général . . . . .	14
2.1.1 Calcul du MPM . . . . .	15
2.1.2 Calcul du MAP . . . . .	17
2.2 Modélisations de la loi de $Z$ . . . . .	18
2.2.1 Chaînes de Markov cachées à bruit indépendant . . . . .	20
2.2.2 Chaînes de Markov couples . . . . .	21
2.2.3 Chaînes couples partiellement de Markov . . . . .	28
2.3 Estimation des paramètres des CM Couples . . . . .	29
2.3.1 Estimation EM . . . . .	29
2.3.2 Estimation MCEM . . . . .	32
2.3.3 Estimation ECI . . . . .	32
2.4 Application à la segmentation de processus gaussiens centrés avec bruit à corrélacion longue . . . . .	33
2.4.1 Calcul des probabilités de transition dans les CCPM . . . . .	33
2.4.2 Segmentation supervisée d'un processus à corrélacion longue . . . . .	35
2.4.3 Segmentation non supervisée d'un processus à corrélacion longue . . . . .	38

2.5	Conclusion . . . . .	39
<b>3</b>	<b>Chaînes de Markov <math>M</math>-non stationnaires cachées à bruit indépendant</b>	<b>43</b>
3.1	Présentation générale des modèles triplets . . . . .	44
3.2	Chaînes de Markov $M$ -non stationnaires cachées à bruit indépendant . . . . .	45
3.3	Segmentation non supervisée de signaux et d'images non stationnaires. . . . .	47
3.3.1	Comparaison entre CMT et CMC-BI . . . . .	47
3.3.2	Segmentation d'images réelles . . . . .	53
3.4	Conclusion . . . . .	53
<b>4</b>	<b>Chaînes de Markov évidentielles cachées à bruit indépendant</b>	<b>59</b>
4.1	Généralités sur la théorie des croyances . . . . .	59
4.1.1	Affaiblissement . . . . .	62
4.2	Introduction de l'information spatiale . . . . .	64
4.2.1	Chaînes de Markov évidentielles cachées . . . . .	65
4.3	Fusion DS dans CMCouple . . . . .	69
4.4	Segmentation non supervisée des images non-stationnaires . . . . .	70
4.4.1	Chaînes de Markov cachées non stationnaires . . . . .	70
4.4.2	Image non stationnaire bruitée . . . . .	71
4.4.3	Image réelle . . . . .	73
4.5	Conclusion . . . . .	73
<b>5</b>	<b>Arbres de Markov cachés flous multispectraux</b>	<b>77</b>
5.1	Extensions des modèles monodimensionnels au cas multirésolution . . . . .	78
5.1.1	Calcul du MPM . . . . .	78
5.1.2	Modélisations de $Z$ dans le cas d'un signal multirésolution . . . . .	80
5.1.3	Estimation EM dans le cas d'un AMC-BI multispectral . . . . .	81
5.2	Arbres de Markov cachés flous multispectraux . . . . .	82
5.3	Expérimentations . . . . .	84
5.4	Conclusion . . . . .	85
	<b>Conclusion générale et perspectives</b>	<b>89</b>
	<b>Bibliographie</b>	<b>91</b>



# Notations

## Symboles

$\mathcal{N}$	Ensemble d'indices
$N$	Nombre d'indices
$\mathcal{X}$	Espace des états cachés
$\mathcal{Y}$	Espace des états observés
$\mathcal{U}$	Espace des états auxiliaires
$X_n, U_n, Y_n$	Variables aléatoires indicées par $n$
$X$	Processus caché
$Y$	Processus observé
$U$	Processus auxiliaire
$Z$	Processus joint de $X$ et $Y$
$V$	Processus joint de $X$ et $U$
$T$	Processus joint de $X, U$ et $Y$
$x$	Réalisation de $X$
$A \perp\!\!\!\perp B$	$A$ est indépendante de $B$
$A \perp\!\!\!\perp B   C$	$A$ est indépendante de $B$ conditionnellement à $C$
$p(\cdot)$	Densité d'une variable aléatoire relativement à une mesure de référence
$p(\cdot   \cdot)$	Densité conditionnelle d'une variable aléatoire
$\beta_n$	Quantité rétrograde
$\mathcal{N}(m, \sigma^2)$	Loi normale réelle de moyenne $m$ et de variance $\sigma^2$
$\Theta$	Espace des paramètres
$\mathbb{E}$	Espérance mathématique
$\mathcal{P}(\Omega)$	Ensemble des parties de $\Omega$
$m$	Fonction de masse
$\oplus$	Fusion de Dempster -Shafer
$s^-$	Père d'un site $s$
$s^+$	Enfants d'un site $s$
$s^{++}$	Ensemble des descendants d'un site $s$
$\underline{s}^{++}$	$\{s\} \cup s^{++}$

## Acronymes

EM	Espérance Maximisation
MCEM	Monte Carlo Espérance Maximisation
SEM	Stochastique Espérance Maximisation
ECI	Estimation conditionnelle itérative
MPM	Mode <i>a posteriori</i> des Marginales
MAP	Maximum <i>a posteriori</i>
FusionDS	Fusion de Dempster-Shafer

## Modèles

CM	Chaîne de Markov d'ordre 1
CMC-BI	Chaîne de Markov cachée à Bruit Indépendant
CMCouple	Chaîne de Markov Couple
CMT	Chaîne de Markov Triplet
CCPM	Chaîne Couple Partiellement de Markov
CMC-BCL	Chaîne de Markov Cachée à Bruit à Corrélation Longue
CMM-NS-BI	Chaîne de Markov $M$ Non Stationnaire à Bruit-Indépendant
AM	Arbre de Markov d'ordre 1
AMC-BI	Arbre de Markov caché à Bruit Indépendant
AMCouple	Arbre de Markov Couple
AMT	Arbre de Markov Triplet

# Introduction

La segmentation est un problème clé en traitement du signal et de l'image, l'objectif étant de déterminer certaines des caractéristiques cachées d'un signal à partir de ses caractéristiques observées et de créer ainsi une partition de ce signal en sous-ensembles possédant certaines propriétés d'homogénéité. Parmi les techniques de segmentation existantes, les méthodes statistiques bayésiennes sont bien adaptées à un grand nombre de situations car elles proposent des outils puissants et généraux, intéressants aussi bien pour les modélisations que les traitements dans des espaces de grandes dimensions. Celles-ci sont fondées sur une modélisation probabiliste du phénomène partiellement observé donnée par la définition de la loi jointe des processus observé et caché. On se fixe alors un critère de qualité des stratégies de segmentation, formalisé par la définition d'une fonction de perte, à partir duquel il est possible d'obtenir l'estimateur optimal du signal caché, à partir du signal observé, en recherchant la stratégie minimisant la perte moyenne. Cela confère aux méthodes statistiques bayésiennes une grande souplesse, étant donné que l'optimalité des solutions peut être adaptée à des préoccupations particulières en choisissant la fonction de perte appropriée. D'autre part, lorsque les paramètres du modèle sont inconnus, l'utilisation de méthodes d'estimation permet l'automatisation des traitements qui sont alors d'un grand intérêt pratique. Ces considérations d'ordre général et qui définissent le cadre de notre étude seront présentées dans le **chapitre 1**.

La loi jointe des processus caché et observé doit cependant être définie avec précaution. En effet, l'utilisation des estimateurs bayésiens, qui permettent d'estimer une réalisation du processus caché à partir des observations, nécessitent le calcul de probabilités *a posteriori*. Étant donné la taille des espaces de configurations considérés, il est en général impossible de calculer directement ces probabilités sans émettre d'hypothèses simplificatrices concernant la loi jointe. Il est donc nécessaire de définir des lois jointes telles que le calcul des probabilités *a posteriori* soit possible, tout en conservant des modèles assez riches pour pouvoir modéliser un grand nombre de situations et de comportements. Les modèles de Markov cachés introduits par Baum et Petrie dans le cas monodimensionnel [3] et par les frères Geman dans le cas bidimensionnel [53] répondent, la plupart du temps, à ces demandes de par leur capacité à modéliser l'information contextuelle avec parcimonie. Toutefois, ces derniers peuvent se révéler inadaptés à certaines applications. En particulier, leurs représentants les plus simples

et les plus couramment utilisés ne permettent pas de prendre en compte les corrélations, conditionnellement aux états cachés, entre observations. Ils ont donc été généralisés aux modèles de Markov couples [102, 93, 94] et couples partiellement de Markov [98, 99]. Ce dernier modèle permet, à titre d'exemple, la prise en compte de bruits à corrélation longue [101]. Notons que c'est également la motivation pour l'introduction des champs de Markov discriminants et conditionnels [67, 64] qui font actuellement l'objet de nombreux travaux. La présentation de ces différentes modélisations de la loi jointe, de généralités croissantes, sera présentée dans le **chapitre 2**. En particulier, nous y proposons une extension originale d'une caractérisation locale des chaînes de Markov cachées par rapport aux chaînes de Markov couples, utile dans l'appréhension intuitive de l'apport des modèles couples. Nous montrons également, via simulations informatiques, que l'information portée par les seules corrélations peut être suffisante, dans les modèles partiellement markoviens, pour retrouver le signal caché.

Par ailleurs, les modèles de Markov couples peuvent être étendus aux modèles de Markov triplets [95]. Le principe est d'ajouter un, voire plusieurs, processus auxiliaire(s) tel que la loi jointe du triplet « processus caché, processus auxiliaire, processus observé » soit celle d'une chaîne de Markov. Ces modèles très généraux permettent, entre autre, de palier une autre limitation des modèles classiques qui consiste à supposer la loi jointe stationnaire. En effet, en introduisant un processus auxiliaire contrôlant les changements de matrices de transition du processus, nous montrons par des simulations l'efficacité d'un tel modèle dans les situations où la loi jointe du processus caché et des observations n'est pas stationnaire [70]. L'étude de ce cas particulier de modèle triplet fait l'objet du **chapitre 3**.

Nous étudions également les possibilités d'extensions des modèles probabilistes classiques à un modèle « évidentiel », avec la loi *a posteriori* du processus caché donnée par la fusion de Dempster-Shafer généralisant la loi de Bayes. Il est possible d'avoir recours à ce formalisme par le biais des modèles triplets [69, 71, 100]. La présentation de la théorie des croyances dans le cas markovien et son application à la segmentation de processus non stationnaires sont présentées dans le **chapitre 4**, dont le contenu est original.

Finalement, les modèles que nous avons présentés dans le cas monodimensionnel peuvent également être utilisés dans le cas de signaux bidimensionnels (image) après transformation de l'ensemble des pixels à l'aide d'un parcours de Hilbert-Peano. Une autre possibilité est d'étendre ces modèles aux modèles d'arbres permettant la segmentation de données multi-résolutions. D'autre part la segmentation floue proposée dans le cas des champs de Markov [108, 107] et des chaînes de Markov [24] a été proposée afin de tenir compte de l'imprécision portant sur l'appartenance d'un site à une région thématique. Dans un signal « flou » cohabitent ainsi des zones homogènes (classes « dures ») avec des zones floues représentant des sites intermédiaires pouvant appartenir à plusieurs classes dures. L'originalité de ces modèles est caractérisée par le fait que leur distribution comprend une composante discrète et une composante continue, la composante discrète étant constituée par les masses de Dirac

et représentant le poids affecté à chaque classe dure et la composante continue correspondant aux sites flous (mesure de Lebesgue). Dans le **chapitre 5**, nous proposons un modèle d'arbres de Markov cachés flous multispectraux [72] fondé sur le même principe.

**Organisation du manuscrit** Dans le chapitre 1, nous présentons la modélisation probabiliste utilisée dans le manuscrit. Puis nous explicitons les principes généraux de la segmentation bayésienne ainsi que les méthodes d'estimation des paramètres permettant d'obtenir des méthodes de segmentation non supervisées.

Dans le chapitre 2, nous écrivons la loi jointe dans le cas général, et nous décrivons les algorithmes d'inférence dans ce cas. Puis nous introduisons des hypothèses d'indépendance conditionnelle de manière à ce que la loi *a posteriori* soit de Markov et donc calculable. Nous présentons successivement le modèle de chaîne de Markov à bruit indépendant, le modèle de chaîne de Markov couple, et le modèle de chaîne couple partiellement de Markov. Les algorithmes d'estimation de paramètres sont explicités pour le cas couple. Finalement, nous présentons une application des modèle couples partiellement de Markov en considérant un modèle particulier permettant la segmentation de processus gaussiens centrés à corrélation longue.

Dans le chapitre 3, nous introduisons le principe général des modèles triplet. Puis nous présentons les méthodes d'estimation du processus caché et des paramètres. Finalement nous proposons une série d'expérimentations mettant en évidence l'amélioration apportée par le nouveau modèle modélisant la non stationnarité.

Dans le chapitre 4, nous présentons dans un premier temps les principes de la théorie des croyances. Puis, nous décrivons le modèle utilisé pour la segmentation non supervisée avec *a priori* évidentiel. Nous définissons pour cela le modèle de chaîne de Markov évidentielle cachée. Nous explicitons alors les algorithmes d'estimation des paramètres utilisés. Finalement, nous proposons différentes simulations sur des données synthétiques et réelles mettant en évidence les améliorations apportées par le nouveau modèle.

Dans le chapitre 5, nous étendons dans un premier temps les modèles vus dans le cas monodimensionnel à des modèles multirésolutions. Nous développons les algorithmes d'inférence et d'estimation des paramètres. Puis, nous proposons un modèle d'arbres de Markov cachés flous multispectraux permettant d'obtenir une carte de segmentation floue à partir d'un ensemble d'images prises par différents capteurs.

Finalement, nous terminons par une conclusion générale mettant en évidence les apports méthodologiques et applicatifs et proposons quelques perspectives de poursuite de ce travail.



# Chapitre 1

## Généralités

Le traitement numérique des signaux nécessite une représentation discrète de ces derniers. Le signal que l'on désire traiter est donc échantillonné de  $\mathbb{R}^d$  dans  $\mathbb{Z}^d$  où  $d$  est la dimension du signal. Le signal discret obtenu, que nous noterons  $y$ , est indexé par un ensemble discret fini  $\mathcal{N}$  de  $N$  sites. A chaque site  $n$  est associé un descripteur  $y_n$  représentant son état et prenant ses valeurs dans un ensemble  $\mathcal{Y}$ . Le descripteur est en général une grandeur physique liée à la nature du capteur comme par exemple la tension dans le cas d'un capteur électrocardiogramme (ECG) ou la quantité de lumière émise par l'objet dans le cas d'un capteur à transfert de charge (charged-coupled device CCD). Il peut également s'agir d'une information plus complexe comme un vecteur dans le cas de l'utilisation de plusieurs capteurs, une matrice dans le cas de données polarimétriques ou plus abstraite, comme la forme des nucléotides dans une séquence d'ADN, ou encore la valeur d'un cours boursier. A partir du signal observé discrétisé  $y = (y_n)_{n \in \mathcal{N}}$ , il convient d'extraire des informations pertinentes en regard de l'application concernée, les traiter puis les interpréter. L'ensemble de ces opérations est appelé *analyse de signaux* [37]. En analyse de signaux, et plus particulièrement en analyse d'image [77], on distingue les traitements de haut-niveau et les traitements de bas-niveau, ces derniers opérant plutôt directement sur les données numériques sans tenir compte nécessairement, à l'inverse des premiers, de la réalité qu'elles représentent. Parmi les traitements de bas-niveau, la *segmentation* peut être cruciale car elle est à la base de la plupart des schémas d'interprétation et de décision. Ainsi, elle n'est souvent pas une fin en soi et sa qualité influe directement sur les résultats obtenus par les traitements situés en aval. Le principe général de la segmentation est de créer une partition du signal  $y$  observé en sous-ensembles possédant certaines caractéristiques d'homogénéité. Elle fait référence aux notions de différence et de similarité, c'est pourquoi on distingue généralement l'approche *frontière* et l'approche *région* [86, 37]. La première consiste à rechercher les frontières ou les contours (dans le cas de signaux bidimensionnels) susceptibles de délimiter les régions d'intérêt. Dans le cadre de la segmentation d'image, cette approche regroupe ainsi toutes les méthodes de détection de contours : méthodes dérivatives [41, 74, 33], surfaciques [59, 56],

morphologiques, markoviennes [52] et variationnelles [61, 112]. L’approche région consiste au contraire à trouver des groupements de points, des zones du signal, ayant des propriétés communes. A chaque site  $n$  du réseau  $\mathcal{N}$  est alors associé un autre descripteur appelé *label*, noté  $x_n$ , correspondant à la région d’appartenance du site, prenant ses valeurs dans l’ensemble  $\mathcal{X}$ . Nous obtenons de cette manière une représentation du signal segmenté  $x = (x_n)_{n \in \mathcal{N}}$ . L’approche région regroupe les méthodes de type croissance de région, division-fusion et statistiques. Outre les méthodes de reconnaissance de forme [47], de nombreuses méthodes statistiques reposent sur une approche bayésienne [53, 12, 96, 29]. La segmentation se ramène alors à l’estimation d’un processus aléatoire à valeurs discrètes à partir d’un processus observé. C’est ce type d’approche que nous adoptons et que nous décrivons de manière générale dans ce chapitre. Nous introduisons les principes de la segmentation bayésienne en décrivant dans un premier temps la modélisation probabiliste retenue. Puis nous présentons quelques méthodes d’estimation des paramètres, dans le cas où ceux du modèle choisi ne sont pas connus, de manière à proposer des méthodes de segmentation non supervisées.

## 1.1 Approche bayésienne du problème de la segmentation

La segmentation peut être vue comme un problème dans lequel nous recherchons des caractéristiques cachées à partir d’observations. La résolution de ce type de problème nécessite d’explicitier les liens entre les phénomènes caché et observé. Il n’existe manifestement pas de liens déterministes entre ces deux phénomènes néanmoins nous pouvons traiter le problème en adoptant une approche statistique. Nous devons pour cela définir une modélisation probabiliste du signal observé et du signal des caractéristiques cachées. Le problème de la segmentation se ramène alors à un problème statistique pouvant être abordé par une approche bayésienne.

### 1.1.1 Modélisation probabiliste

En chaque site  $n$  de  $\mathcal{N}$ , nous considérons  $x_n$  (resp.  $y_n$ ) comme la réalisation d’une variable aléatoire (v.a)  $X_n$  (resp.  $Y_n$ ) prenant ses valeurs dans l’ensemble  $\mathcal{X}$  (resp.  $\mathcal{Y}$ ). Nous définissons alors les deux processus aléatoires (p.a) suivants :

$$\begin{aligned} X &= \{X_n\}_{n \in \mathcal{N}} && : \text{le p.a caché prenant ses valeurs dans } \mathcal{X}^N \\ Y &= \{Y_n\}_{n \in \mathcal{N}} && : \text{le p.a observable prenant ses valeurs dans } \mathcal{Y}^N \end{aligned}$$

Dans le cadre probabiliste, le signal caché (resp. observé) correspond donc à une réalisation  $x$  (resp.  $y$ ) du p.a  $X$  (resp.  $Y$ ). Nous notons  $\kappa$  la mesure de comptage sur  $\mathcal{X}$  et  $\mu$  une mesure de référence sur  $\mathcal{Y}$  et nous supposons que la loi du processus couple  $Z = (X, Y)$  admet une densité  $p(z)$  relativement à la mesure produit  $(\kappa \otimes \mu)^N$ . Finalement, la loi de  $Z$  permet de modéliser l’ensemble des liens statistiques entre les deux p.a.  $X$  et  $Y$ . Nous pouvons dès



lors exprimer notre connaissance sur  $X$  lorsque  $y$  est observé à partir de la loi *a posteriori* calculée à partir de

$$p(x|y) = \frac{p(z)}{\sum_{x \in \mathcal{X}^N} p(z)} \quad (1.1.1)$$

Il est alors possible, via la définition des lois de probabilités conditionnelles à  $y$  sur  $\mathcal{X}^N$ , de modéliser des situations très diverses, allant du lien déterministe à l'indépendance totale entre  $X$  et  $Y$ , assurant ainsi une certaine "généralité" aux méthodes statistiques [96].

**Remarque** Dans le cadre bayésien il est classique de considérer la densité  $p(x)$  relative à la mesure  $\kappa^N$ , appelée loi *a priori*, et la densité  $p(y|x)$  relative à la mesure  $\mu^N$ , qui est la loi des observations conditionnellement à  $X = x$ . La loi de  $Z$  est alors donnée par  $p(z) = p(x)p(y|x)$ . Classiquement, lors de la modélisation, la loi *a priori*  $p(x)$  est choisie en fonction des caractéristiques quantitatives ou spatiales des classes souhaitées et permet par exemple de prendre en compte la dépendance spatiale des observations dans le cas de la segmentation d'une image. Ensuite, le processus de formation du phénomène observé est modélisé par la densité  $p(y|x)$ . Bien que cette manière de modéliser le problème soit intuitive, nous verrons par la suite que le fait de considérer directement la loi de  $Z$  permet de proposer des modèles plus généraux donnant, dans certaines situations, de meilleurs résultats de segmentation.

### 1.1.2 Stratégie bayésienne

Une fois définie la loi de  $Z = (X, Y)$  à valeurs dans  $(\mathcal{X} \times \mathcal{Y})^N$ , le problème de la segmentation se ramène au problème probabiliste suivant : comment estimer une réalisation invisible  $x$  du p.a  $X$  à partir de l'observation  $y$  du p.a  $Y$ . On introduit pour cela une *stratégie de classification*  $\hat{s} : \mathcal{Y}^N \rightarrow \mathcal{X}^N$ . Nous définissons une application appelée *fonction de perte*  $L : \mathcal{X}^N \times \mathcal{X}^N \rightarrow \mathbb{R}^+$  donnant la perte  $\lambda_{i,j}$  occasionnée lorsque l'on choisit  $i$  au lieu de la vraie valeur  $j$

$$\forall (i, j) \in \mathcal{X}^N \times \mathcal{X}^N, \begin{cases} L(i, j) = 0 & \text{si } i = j \\ L(i, j) = \lambda_{i,j} & \text{si } i \neq j \end{cases} \quad (1.1.2)$$

Le risque bayésien associé à  $\hat{s}$  et à  $L$  est donné par la perte moyenne :

$$\mathcal{R}(\hat{s}, L) = \mathbb{E} \{L(\hat{s}(Y), X)\} \quad (1.1.3)$$

La décision bayésienne  $\hat{s}_B$  est celle qui minimise ce risque

$$\mathbb{E} \{L(\hat{s}_B(Y), X)\} = \min_{\hat{s}} \mathbb{E} \{L(\hat{s}(Y), X)\} \quad (1.1.4)$$

On peut trouver  $\hat{s}_B$  en minimisant, pour chacune des observations  $Y = y$ , l'espérance conditionnelle  $\mathbb{E} \{L(\hat{s}(Y), X)|Y = y\}$ . Finalement, la stratégie bayésienne, relativement à la fonc-

tion de perte  $L$  choisie, est celle qui minimise  $\mathbb{E} \{L(\hat{s}(Y), X)|Y = y\}$  pour chaque observation  $Y = y$  :

$$\mathbb{E} \{L(\hat{s}_B(Y), X)|Y = y\} = \min_{\hat{s}} \mathbb{E} \{L(\hat{s}(Y), X)|Y = y\} \quad (1.1.5)$$

L'estimateur bayésien  $\hat{x}_B$  optimal par rapport à une fonction de perte de la forme (1.1.2), est donc donné par

$$\hat{x}_B(y) = \arg \min_{x' \in \mathcal{X}^N} \sum_{x \in \mathcal{X}^N} \lambda_{x', x} p(x|y) \quad (1.1.6)$$

**Remarque** Selon nos préoccupations, il peut être parfois préférable de définir la fonction de perte à partir d'une fonction de perte locale sous la forme additive suivante

$$\forall (i, j) = (i_n, j_n)_{n \in \mathcal{N}} \in \mathcal{X}^N \times \mathcal{X}^N, L(i, j) = \sum_{n \in \mathcal{N}} L_n(i_n, j_n) \quad (1.1.7)$$

Dans ce cas, on peut montrer que la stratégie bayésienne se ramène à la minimisation suivante

$$\sum_{n \in \mathcal{N}} \mathbb{E} \{L_n(\hat{s}_B(Y)_n, X_n)|Y = y\} = \min_{\hat{s}} \sum_{n \in \mathcal{N}} \mathbb{E} \{L_n(\hat{s}(Y)_n, X_n)|Y = y\} \quad (1.1.8)$$

Etant donné que tous les termes de la somme sont positifs, la stratégie bayésienne consiste à minimiser en chaque site l'espérance de la fonction de perte locale conditionnellement aux observations. Nous obtenons des estimateurs locaux optimaux  $(\hat{x}_n)_B$  qui doivent être calculés en chaque site  $n \in \mathcal{N}$

$$(\hat{x}_n)_B(y) = \arg \min_{x'_n \in \mathcal{X}} \sum_{x_n \in \mathcal{X}} \lambda_{x'_n, x_n} p(x_n|y) \quad (1.1.9)$$

Les méthodes statistiques de traitement présentent donc une certaine "souplesse" dans le sens où, devant une réalité objective donnée, modélisée par la loi de  $Z$ , l'optimalité des solutions est adaptable à des préoccupations particulières en choisissant la fonction de perte appropriée [96]. On peut ainsi, dans le cas d'une application particulière, définir une fonction de perte donnant des pertes différentes en fonction des classes. On privilégiera dans ce cas certaines classes par rapport à d'autres. Néanmoins, par la suite nous n'utiliserons que des estimateurs optimaux par rapport à la fonction de perte  $L_{0/1}$  définie dans la sous-section suivante.

### 1.1.3 Estimateurs bayésiens fondés sur la fonction de perte $L_{0/1}$

Considérons la fonction de perte  $L_{0/1}$  définie de la manière suivante

$$\forall i, j \in \mathcal{I} \times \mathcal{I}, L_{0/1}(i, j) = \mathbb{1}_{[i \neq j]} \quad (1.1.10)$$

A partir de cette dernière, nous pouvons définir deux estimateurs couramment utilisés en traitement du signal et des images : l'estimateur du *Maximum A Posteriori* (MAP) et l'estimateur du *Mode des Marginales a Posteriori* (MPM).

**Estimateur MAP** La fonction de perte du MAP est la fonction  $L_{0/1}$  définie dans (1.1.10) où  $\mathcal{I}$  est l'ensemble  $\mathcal{X}^N$  des configurations

$$\forall i, j \in \mathcal{X}^N \times \mathcal{X}^N, L_{MAP}(i, j) = L_{0/1}(i, j) \quad (1.1.11)$$

La fonction de perte du MAP traduit le fait que l'on pénalise de la même manière toutes les configurations estimées différentes de la vraie configuration. Ainsi une configuration estimée ne présentant qu'un seul site mal classé, sera pénalisée de la même manière qu'une configuration possédant un nombre important de sites mal classés. A partir de (1.1.6), nous obtenons alors l'estimateur du MAP

$$\hat{x}_{MAP}(y) = \arg \max_{x' \in \mathcal{X}^N} p(x'|y) \quad (1.1.12)$$

L'estimateur du MAP est donc celui qui maximise la probabilité *a posteriori*.

**Estimateur MPM** La fonction de perte du MPM est de la forme additive (1.1.7) avec  $L_n = L_{0/1}$  où  $\mathcal{I}$  est l'ensemble  $\mathcal{X}$

$$\forall (i, j) = (i_n, j_n)_{n \in \mathcal{N}} \in \mathcal{X}^N \times \mathcal{X}^N, L_{MPM}(i, j) = \sum_{n \in \mathcal{N}} L_{0/1}(i_n, j_n) \quad (1.1.13)$$

Comparativement à la fonction de perte du MAP (1.1.11), celle du MPM pénalise une configuration estimée de manière proportionnelle au nombre de sites mal classés. A partir de (1.1.9), nous obtenons les estimateurs locaux du MPM qui doivent être calculés en chaque site  $n$

$$(\hat{x}_n)_{MPM}(y) = \arg \max_{x'_n \in \mathcal{X}} p(x'_n|y) \quad (1.1.14)$$

L'estimateur du MPM est donc celui qui maximise la probabilité marginale *a posteriori* en chaque site. L'estimateur MPM sera bien adapté au problème de la segmentation si notre préoccupation est de minimiser le taux de sites mal classés.

Les estimateurs du MAP et du MPM, ou plus généralement de la forme (1.1.6) et (1.1.9), nécessitent le calcul de  $(p(x|y))_{x \in \mathcal{X}^N}$  dans le cas de (1.1.6) ou des  $(p(x_n|y))_{n \in \mathcal{N}, x_n \in \mathcal{X}}$  dans le cas de (1.1.9). Etant donné la grande taille des espaces de configurations considérés, il est impossible de calculer directement ces probabilités sans faire au préalable des hypothèses simplificatrices. Nous verrons qu'il est néanmoins possible de définir des lois de  $Z$  permettant le calcul de ces probabilités tout en étant suffisamment riches pour pouvoir modéliser un grand nombre de situations.

## 1.2 Estimation des paramètres

Nous avons supposé jusqu'à présent la loi de  $Z$  parfaitement connue. De manière générale, ce n'est pas le cas et on suppose que la densité  $p(z)$  est paramétrée. Il est alors nécessaire d'estimer ses paramètres avant de procéder à la segmentation du processus caché. Nous noterons  $\Phi$  l'espace des paramètres du modèle  $p(z|\phi)$  considéré avec  $\phi \in \Phi$ . En supposant que le processus  $Z$  est stationnaire il est alors possible d'estimer ces paramètres à partir des seules observations  $y$ . Différentes approches peuvent être utilisées notamment celles basées sur la maximisation de la vraisemblance comme l'algorithme Espérance Maximisation (EM) et ses consorts stochastiques ou d'autres comme l'algorithme Espérance Conditionnelle Itérative (ECI). Ils permettent la mise en place de méthodes de segmentation non supervisée d'un grand intérêt pratique notamment en imagerie satellitaire ou aérienne [48, 115], en imagerie sonar [79, 82, 80, 81], ou encore en imagerie médicale [88, 58, 46, 17]. L'objectif de cette section est de présenter de manière générale quelques algorithmes d'estimation couramment utilisés en imagerie. Dans les chapitres suivants, nous utiliserons certains de ces algorithmes pour l'estimation des paramètres de nos modèles afin d'obtenir des algorithmes de segmentation non supervisée.

### 1.2.1 L'algorithme Espérance-Maximisation (EM)

L'estimateur du maximum de vraisemblance (EMV) donne les paramètres du modèle maximisant la vraisemblance  $p(y|\phi)$  que nous supposons dérivable par rapport à  $\phi$ . Afin de calculer son maximum nous devons déterminer les valeurs de  $\phi$  pour lesquelles sa dérivée s'annule. La vraisemblance s'écrivant souvent sous la forme d'un produit de densités, il est parfois préférable de prendre son logarithme. La fonction logarithme étant croissante, il est alors équivalent de maximiser  $\log p(y|\phi)$  ou  $p(y|\phi)$ . Nous noterons  $\mathcal{L}(y; \phi) = \log p(y|\phi)$  la log-vraisemblance. L'estimateur du maximum de vraisemblance est alors donné par

$$\hat{\phi}(y) = \arg \max_{\phi} \mathcal{L}(y; \phi) \quad (1.2.1)$$

Pour trouver la valeur des paramètres maximisant la log-vraisemblance, on peut en principe utiliser n'importe quelle méthode d'optimisation. Néanmoins, les modèles que nous étudions possèdent des *variables latentes* ou *cachées*. La fonction à optimiser pour ce type de modèle comprenant souvent le logarithme d'une somme, il est en général difficile d'obtenir une solution analytique simple. On a donc recours à des méthodes numériques itératives. L'algorithme Espérance-Maximisation (EM [40, 78]) est une de ces méthodes. Il s'agit d'un algorithme général d'estimation par maximum de vraisemblance dans les cas où les données sont *incomplètes* ou lorsque la maximisation de la vraisemblance est rendue plus simple par

l'introduction de *variables latentes*. La log-vraisemblance complétée est donnée par

$$\mathcal{L}_c(z; \phi) = \log p(z|\phi) = \mathcal{L}(y; \phi) + \log p(x|y; \phi) \quad (1.2.2)$$

Supposons que la valeur courante, à l'itération  $q$ , des paramètres soit  $\phi^{[q]}$ . Notre but est de trouver  $\phi^{[q+1]}$  qui puisse augmenter la log-vraisemblance  $\mathcal{L}(y; \phi)$ . L'accroissement de la log-vraisemblance peut être écrit sous la forme suivante :

$$\mathcal{L}(y; \phi) - \mathcal{L}(y; \phi^{[q]}) = \mathcal{L}_c(z; \phi) - \mathcal{L}_c(z; \phi^{[q]}) + \log \frac{p(x|y; \phi^{[q]})}{p(x|y; \phi)} \quad (1.2.3)$$

Prenons l'espérance conditionnelle par rapport à  $y$ , à la valeur courante des paramètres  $\phi^{[q]}$ , à droite et à gauche de cette équation. Etant donné que le membre de gauche ne contient pas  $x$ , celui-ci n'est pas modifié et nous obtenons :

$$\mathcal{L}(y; \phi) - \mathcal{L}(y; \phi^{[q]}) = \mathbb{E}_{\phi^{[q]}} \left\{ \mathcal{L}_c(Z; \phi) - \mathcal{L}_c(Z; \phi^{[q]}) | Y = y \right\} + \underbrace{\mathbb{E}_{\phi^{[q]}} \left\{ \log \frac{p(X|y; \phi^{[q]})}{p(X|y; \phi)} | Y = y \right\}}_{\geq 0} \quad (1.2.4)$$

Le dernier terme correspond à la divergence de Kullback-Leibler de  $p(x|y; \phi^{[q]})$  et de  $p(x|y; \phi)$ , qui est toujours positive ou nulle. En posant ,  $\mathcal{Q}(\phi|\phi^{[q]}) = \mathbb{E}_{\phi^{[q]}} \{ \mathcal{L}_c(Z; \phi) | Y = y \}$ , nous en déduisons la majoration suivante :

$$\mathcal{L}(y; \phi) - \mathcal{L}(y; \phi^{[q]}) \geq \mathcal{Q}(\phi|\phi^{[q]}) - \mathcal{Q}(\phi^{[q]}|\phi^{[q]}) \quad (1.2.5)$$

Que nous pouvons réécrire sous la forme

$$\mathcal{L}(y; \phi) \geq l(\phi|\phi^{[q]}) \text{ avec } l(\phi|\phi^{[q]}) = \mathcal{L}(y; \phi^{[q]}) + \mathcal{Q}(\phi|\phi^{[q]}) - \mathcal{Q}(\phi^{[q]}|\phi^{[q]}) \quad (1.2.6)$$

La fonction  $l(\phi|\phi^{[q]})$  est donc bornée supérieurement par la log-vraisemblance  $\mathcal{L}(y; \phi)$ . D'autre part, on peut vérifier que les fonctions  $l(\phi|\phi^{[q]})$  et  $\mathcal{L}(y; \phi)$  sont égales pour  $\phi = \phi^{[q]}$ . Ainsi, toute valeur de  $\phi$  faisant croître la valeur de  $l(\phi|\phi^{[q]})$  fera croître la log-vraisemblance  $\mathcal{L}(y; \phi)$ . L'objectif étant de maximiser la log-vraisemblance  $\mathcal{L}(y; \phi)$  par rapport à  $\phi$ , l'algorithme EM sélectionne la valeur de  $\phi$  qui maximise  $l(\phi|\phi^{[q]})$  afin d'obtenir le plus grand accroissement de la valeur de la log-vraisemblance à chaque itération :

$$\begin{aligned} \phi^{[q+1]} &= \arg \max_{\phi} l(\phi|\phi^{[q]}) \\ &= \arg \max_{\phi} \left\{ \mathcal{Q}(\phi|\phi^{[q]}) + \underbrace{\mathcal{L}(y; \phi^{[q]}) - \mathcal{Q}(\phi^{[q]}|\phi^{[q]})}_{Cste} \right\} \end{aligned}$$

$\mathcal{L}(y; \phi^{[q]})$  et  $\mathcal{Q}(\phi^{[q]}|\phi^{[q]})$  ne contenant pas  $\phi$ , ils peuvent être traités comme des constantes. Finalement, nous obtenons

$$\phi^{[q+1]} = \arg \max_{\phi} \mathcal{Q}(\phi|\phi^{[q]}) \quad (1.2.7)$$

Le déroulement de l'algorithme EM est donc le suivant :  $\phi$  est d'abord initialisé à une valeur  $\phi^{[0]}$  des paramètres. Puis, on itère alternativement entre deux étapes notées  $E$  (Espérance) et  $M$  (Maximisation) :

- dans l'étape  $E$ , nous calculons la fonction  $\mathcal{Q}(\phi|\phi^{[q]})$ ,
- dans l'étape  $M$ , les paramètres sont estimés en maximisant la fonction  $\mathcal{Q}(\phi|\phi^{[q]})$  par rapport à  $\phi$  selon (1.2.7).

Ayant une nouvelle génération des valeurs des paramètres  $\phi^{[q+1]}$ , on peut dès lors répéter les étapes  $E$  et  $M$ . On construit ainsi, de manière itérative une suite de paramètres  $(\phi^{[q]})_{q>0}$  telle que  $(\mathcal{L}(y; \phi^{[q]}))_{q \geq 0}$  converge vers un maximum local de la log-vraisemblance  $\mathcal{L}(y; \phi)$  [78]. Dans la pratique, on arrête l'algorithme lorsque  $\mathcal{L}(y; \phi^{[q+1]}) - \mathcal{L}(y; \phi^{[q]}) \leq \varepsilon$  où  $\varepsilon$  est fixé.

En ce qui concerne les propriétés de convergence de l'algorithme EM, suite à la formulation générale de l'algorithme EM par Dempster et al. en 1977 [40], Wu [120] en a approfondi l'étude dans le cas général. Il a montré que sous certaines conditions, toute séquence  $(\mathcal{L}(y; \phi^{[q]}))_{q \geq 0}$  construite à partir de EM converge vers un point stationnaire de la fonction de vraisemblance. D'autre part, si cette dernière possède plusieurs points stationnaires, la convergence de la séquence EM dépend alors de la valeur initiale  $\phi^{[0]}$  des paramètres et la convergence vers le maximum global n'est pas garantie, l'algorithme pouvant rester piégé par un maximum local ou un point selle. Pour plus d'informations concernant la convergence de l'algorithme EM, on pourra consulter [78] ainsi que [120].

L'algorithme EM possède de nombreux avantages parmi lesquels sa stabilité numérique (la log-vraisemblance croissant à chaque itération), le fait qu'il soit, en général, analytiquement et numériquement simple à implémenter ou encore son coût calculatoire faible à chaque itération. Ces propriétés en ont fait une méthode de référence pour l'estimation des modèles à données manquantes. Cependant, EM possède des limitations comme sa sensibilité à la valeur initiale  $\phi^{[0]}$ , sa faible vitesse de convergence ou encore des difficultés de calcul possibles au niveau des étapes  $E$  (calcul analytique de  $\mathcal{Q}(\phi|\phi^{[q]})$  impossible) et  $M$ . Ces limitations ont motivé le développement de nouveaux algorithmes comme ceux que nous allons décrire à présent.

**Remarque** Dans la formulation de l'algorithme EM décrit précédemment,  $\phi^{[q+1]}$  a été choisi comme la valeur de  $\phi$  pour laquelle  $l(\phi|\phi^{[q]})$  est maximisée. Alors que cela assure le plus grand accroissement de  $\mathcal{L}(y; \phi)$ , il est cependant possible de relâcher cette contrainte de maximisation en recherchant simplement une valeur de  $\phi^{[q+1]}$  telle que  $\mathcal{Q}(\phi^{[q+1]}|\phi^{[q]}) \geq \mathcal{Q}(\phi^{[q]}|\phi^{[q]})$ . De cette manière, on fait croître la log-vraisemblance sans chercher à maximiser cet accroissement. Cette approche est connue sous le nom d'algorithme EM généralisé (GEM

[84]) et est souvent utilisée dans le cas où l'étape  $M$  de maximisation est difficile.

**Versions stochastiques de EM** Dans le cas où la fonction  $\mathcal{Q}(\phi|\phi^{[q]})$  n'est pas calculable analytiquement, une solution est d'avoir recours à des approximations de Monte Carlo menant à l'algorithme Monte Carlo EM (MCEM [117]). Le calcul de  $\mathcal{Q}(\phi|\phi^{[q]})$  est alors remplacé par sa moyenne empirique  $\mathcal{Q}^{[q+1]}(\phi|\phi^{[q]})$  fondée sur  $\tau$  tirages de  $x$  *a posteriori* pour une valeur  $\phi^{[q]}$  des paramètres. Lorsque  $\tau = 1$ , MCEM se réduit à l'algorithme EM stochastique (SEM [26]). Initialement, l'algorithme SEM fut proposé, non pas pour approcher le calcul de  $\mathcal{Q}(\phi|\phi^{[q]})$ , mais pour tenter de répondre aux limitations de EM évoquées précédemment. L'étape stochastique introduite est motivée par le principe d'attribution aléatoire suivant (dit *Stochastic Imputation Principle* (SIP [27])) : "générer des échantillons pseudo-complets en tirant des échantillons potentiellement inobservables à partir de leur densité conditionnelle aux observations". A chaque itération, il existe ainsi, contrairement à l'algorithme EM, une probabilité non nulle d'accepter un estimé  $\phi^{[q+1]}$  pour lequel la vraisemblance est inférieure à celle obtenue à partir de  $\phi^{[q]}$ . Cette propriété permet en principe d'éviter que la convergence de l'algorithme soit ralentie par un point selle de la vraisemblance ou bloquée dans un maximum local non significatif. L'étude des propriétés de convergence de l'algorithme SEM a été menée dans le cadre de l'identification de mélanges finis de densités [28] dans lequel les auteurs partent du fait que la suite aléatoire créée est une chaîne de Markov homogène, qui peut s'interpréter comme une perturbation aléatoire de l'algorithme EM. Le principal avantage de SEM sur EM est d'être moins sensible à l'initialisation que EM, bien que sa vitesse de convergence soit plus faible. D'autre part, SEM peut se comporter dans certains cas mieux que EM [25]. Le principe des algorithmes de détermination de l'EMV est résumé dans l'Algorithme 1.

### 1.2.2 Estimation Conditionnelle Itérative (ECI)

Le principe de ECI [91, 92, 9, 89] est différent de celui de EM et de ses versions stochastiques. Il consiste à prendre un estimateur  $\hat{\phi} = \hat{\phi}(z)$  défini à partir des données complètes  $z$  et à l'approcher par une fonction de  $y$  observé. La meilleure approximation au sens de l'erreur quadratique moyenne est l'espérance conditionnelle par rapport à  $y$ . L'espérance conditionnelle de  $\hat{\phi} = \hat{\phi}(z)$  par rapport à  $y$ , notée  $\mathbb{E}_\phi \left\{ \hat{\phi}(Z) | Y = y \right\}$ , dépend du paramètre  $\phi$ , elle n'est donc pas calculable explicitement. Pour la calculer, on utilise une procédure itérative qui consiste à calculer la valeur suivante des paramètres en fonction de leur valeur "courante" :

$$\phi^{[q+1]} = \mathbb{E}_{\phi^{[q]}} \left\{ \hat{\phi}(Z) | Y = y \right\} \quad (1.2.11)$$

Lorsque  $\mathbb{E}_{\phi^{[q]}} \left\{ \hat{\phi}(Z) | Y = y \right\}$  n'est pas explicitement calculable, on peut l'approcher (de la même manière que pour MCEM) par sa moyenne empirique : on génère un échantillon i.i.d

---

**Algorithme 1** Estimation des paramètres : détermination de l'EMV.

---

- Initialiser  $\phi^{[0]}$  ;
- Améliorer itérativement l'estimation de  $\phi$  en alternant entre les deux étapes suivantes :
  - $\triangleright$  **Etape E** : calculer  $\mathcal{Q}(\phi|\phi^{[q]}) = \mathbb{E}_{\phi^{[q]}} \{ \mathcal{L}_c(z; \phi) | y \}$  ;
    - Dans le cas où  $\mathcal{Q}(\phi|\phi^{[q]})$  n'est pas calculable analytiquement (MCEM), ou dans le but d'introduire une perturbation stochastique (SEM), générer un échantillon i.i.d  $x^{[q]}(1), \dots, x^{[q]}(\tau)$  à partir de  $p(x|y; \phi^{[q]})$  et en déduire  $\mathcal{Q}(\phi|\phi^{[q]})$  de la façon suivante

$$\mathcal{Q}(\phi|\phi^{[q]}) = \frac{1}{\tau} \sum_{t=1}^{\tau} \log p(x^{[q]}(t), y | \phi) \quad (1.2.8)$$

- $\triangleright$  **Etape M** : ré-estimer  $\phi$  en maximisant la fonction  $\mathcal{Q}$  ;

$$\phi^{[q+1]} = \arg \max_{\phi} \mathcal{Q}(\phi|\phi^{[q]}) \quad (1.2.9)$$

- Dans le cas où la maximisation est difficile (GEM), choisir  $\phi^{[q+1]}$  tel que

$$\mathcal{Q}(\phi^{[q+1]}|\phi^{[q]}) \geq \mathcal{Q}(\phi^{[q]}|\phi^{[q]}) \quad (1.2.10)$$

- Arrêter lorsque  $\mathcal{L}(\phi^{[q+1]}) - \mathcal{L}(\phi^{[q]}) \leq \varepsilon$ .
- 

---

**Algorithme 2** Estimation conditionnelle itérative

---

- Initialiser  $\phi^{(0)}$  ;
- Calculer  $\phi^{[q+1]}$  à partir de  $\phi^{[q]}$  et de  $y$  par

$$\phi^{[q+1]} = \mathbb{E}_{\phi^{[q]}} \left\{ \hat{\phi}(z) | Y = y \right\} \quad (1.2.12)$$

- Répéter jusqu'à convergence.
-



$x^{[q]}(1), \dots, x^{[q]}(\tau)$  à partir de  $p(x|y, \phi^{[q]})$  et on pose :

$$\phi^{[q+1]} = \frac{1}{\tau} \sum_{t=1}^{\tau} \hat{\phi}(x^{[q]}(t), y) \quad (1.2.13)$$

Il peut arriver que l'espérance soit calculable pour certaines composantes de  $\phi$  et ne le soit pas pour les autres. Dans ce cas, il suffit d'appliquer (1.2.13) pour les composantes dont l'espérance n'est pas calculable. Bien que la convergence de la procédure ECI n'ait pas été prouvée, différentes expériences et applications témoignent de son efficacité [9, 42]. Il s'agit d'une méthode très flexible car elle permet de choisir les estimateurs sur données complètes les plus adaptés aux modèles étudiés en évitant le calcul et la maximisation souvent délicats des vraisemblances. Enfin, lorsque le maximum de vraisemblance est utilisé pour  $\hat{\phi}(z)$ , l'ECI ne diffère de l'EM que par l'ordre des opérations de maximisation et de calcul de l'espérance conditionnelle. Ainsi, lorsque cet ordre peut être inversé, on aura l'équivalence entre les deux algorithmes [39].

### 1.3 Conclusion

La segmentation statistique est donc fondée sur une modélisation probabiliste des phénomènes caché et observé. L'ensemble des liens statistiques entre ces deux phénomènes est modélisé par la loi de  $Z = (X, Y)$  donnée par la densité paramétrée  $p(z|\theta)$ . On se fixe alors un critère de qualité des stratégies de segmentation formalisé par la définition d'une fonction de perte. On peut dès lors obtenir l'estimateur optimal du signal segmenté  $x$  à partir du signal observé  $y$  en recherchant la stratégie minimisant la perte moyenne conditionnellement à  $y$ . Les estimateurs du MAP et du MPM, couramment utilisés en segmentation de signaux et que nous utiliserons par la suite, sont fondés sur la fonction de perte  $L_{0/1}$ . Néanmoins, il est tout à fait envisageable d'utiliser d'autres fonctions de perte ce qui confère aux méthodes statistiques une certaine "souplesse" dans le sens où l'on peut adapter l'optimalité à des préoccupations particulières. Enfin, lorsque que les paramètres du modèle sont inconnus, différentes méthodes d'estimation peuvent être utilisées, permettant d'obtenir des algorithmes de segmentation non supervisée d'un grand intérêt pratique. D'autre part, nous avons vu que les méthodes de segmentation bayésienne nécessitent le calculs de probabilités *a posteriori* qu'il est impossible de calculer directement étant donné la taille des espaces de configurations considérés. Il est alors nécessaire d'introduire des hypothèses simplificatrices. Nous allons voir dans le chapitre suivant qu'il est possible de définir des lois de  $Z$  permettant le calcul rapide de ces probabilités *a posteriori*, dans certains cas de manière exacte, tout en étant suffisamment riches pour modéliser une gamme variée de comportements [104].



## Chapitre 2

# Modélisations de la loi jointe dans le cas d'un signal monodimensionnel

La segmentation statistique bayésienne repose sur une modélisation probabiliste des phénomènes caché et observé. On modélise ainsi l'ensemble des liens entre les deux phénomènes en définissant la loi jointe  $Z = (X, Y)$  du processus caché  $X$  et du processus observé  $Y$ . Néanmoins, celle-ci doit être définie avec précaution. En effet, l'utilisation des estimateurs bayésiens, qui permettent d'estimer une réalisation  $x$  du processus caché  $X$  à partir de l'observation  $y$  du processus  $Y$ , nécessitent le calcul de probabilités *a posteriori*. Etant donné la taille des espaces de configurations considérés, il est en général difficile, voire impossible, de calculer directement ces probabilités, d'une manière numérique, sans émettre d'hypothèses simplificatrices concernant la loi de  $Z$ . Il existe par exemple  $2^{256 \times 256} \simeq 10^{20000}$  configurations possibles pour une image de taille  $256 \times 256$  à deux classes. Un moyen simple de contourner le problème est de supposer l'indépendance des v.a  $Z_n$  pour tout  $n \in \mathcal{N}$ . On peut alors appliquer les méthodes d'estimation bayésienne site par site. Néanmoins, on perd dans ce cas une information non négligeable qui est l'information spatiale. Le problème est donc de définir des lois jointes telles que les probabilités *a posteriori* soit aisément calculables, tout en conservant des modèles assez riches pour pouvoir modéliser un grand nombre de situations et de comportements.

Dans ce chapitre, nous allons présenter des modèles, de généralités croissantes, pour lesquels le calcul des probabilités *a posteriori* peut être mené d'une manière récursive. Nous nous limiterons dans un premier temps à des modèles monodimensionnels. Bien que ces modèles soient principalement destinés à des problématiques de segmentation de signaux monodimensionnels, ils peuvent également fournir des méthodes rapides de segmentation d'images 2D. En effet, contrairement aux méthodes fondées sur les modèles bidimensionnels<sup>1</sup> faisant appel à des simulations itératives, souvent coûteuses en temps de calcul, pour le calcul des quantités d'intérêt, les modèles monodimensionnels permettent le calcul direct

---

<sup>1</sup>par exemple fondées sur des modèles de champs de Markov cachés.

de ces quantités. Pour cela, les images traitées seront dans un premier temps converties en processus monodimensionnels au moyen d'un parcours fractal de Hilbert-Peano [9]. Ce dernier a ainsi été utilisé dans le cadre de la segmentation des images fixes et animées [9], multisenseurs [54], ou encore multirésolutions [51]. On obtient de cette manière des méthodes concurrentes, plus rapides mais s'appuyant sur un modèle moins satisfaisant au plan intuitif, de celles utilisant des modèles de champs 2D. On trouvera dans [109] des comparaisons des résultats de segmentation d'images obtenus en considérant les modèles monodimensionnels et les modèles bidimensionnels ; il apparaît que les méthodes fondées sur les modèles monodimensionnels peuvent être compétitives. D'autre part, les résultats obtenus en considérant les modèles monodimensionnels peuvent servir d'initialisations, aussi bien au niveau de l'estimation des paramètres qu'au niveau de la segmentation. Enfin, dans le cadre de notre étude, la représentation des signaux segmentés sous la forme d'images facilitera grandement la comparaison qualitative des résultats de segmentation obtenus en considérant différents modèles.

Nous écrirons tout d'abord la loi de  $Z$  dans le cas général, sans faire d'hypothèses d'indépendance conditionnelle. Puis nous expliciterons les algorithmes de calculs des probabilités *a posteriori* dans le cas général. L'application numérique directe de ceux-ci n'étant pas possible, on aura recours à des hypothèses d'indépendance conditionnelle permettant de simplifier le modèle, l'objectif étant de réduire la quantité de calculs tout en conservant un modèle suffisamment riche pour prendre en compte les interactions entre les v.a des sites. Nous présenterons ainsi successivement le modèle de chaînes de Markov cachées à bruit indépendant (CMC-BI), le modèle de chaînes de Markov couples (CMCouple) et enfin le modèle de chaînes couples partiellement de Markov (CCPM). Puis, nous expliciterons les algorithmes d'estimation des paramètres dans le cas couple. Enfin nous terminerons ce chapitre par une application des CCPM à la segmentation de processus gaussiens centrés à corrélation longue.

## 2.1 Algorithmes d'inférence dans le cas général

Considérons le cas d'un signal monodimensionnel. Celui-ci est discrétisé sur une grille monodimensionnelle  $\mathcal{N}$  de longueur  $N$ . Pour simplifier les notations, nous noterons  $Y_{m:n}$  (resp.  $y_{m:n}$ ) l'ensemble des v.a  $\{Y_m, Y_{m+1}, \dots, Y_{n-1}, Y_n\}$  (resp. réalisations  $\{y_m, y_{m+1}, \dots, y_{n-1}, y_n\}$ ) comprises entre les indices  $m$  et  $n$  inclus. Le signal discrétisé observé est alors donné par  $y = y_{1:N}$  et la loi de  $Z = (X, Y)$  peut être écrite sous la forme factorisée suivante :

$$p(z) = p(z_1) \prod_{n=2}^N p(z_n | z_{1:n-1}) \quad (2.1.1)$$

La loi de  $Z$  est donc donnée par le produit de la loi  $p(z_1)$  et des lois de *transition*  $p(z_n | z_{1:n-1})$  pour  $n \in \mathcal{N} \setminus \{1\}$ . Nous allons à présent montrer comment estimer une réalisation du

processus caché à partir des observations au sens du MPM (1.1.14), puis du MAP (1.1.12), lorsque la loi de  $Z$  est de la forme générale (2.1.1).

### 2.1.1 Calcul du MPM

L'estimateur du MPM, décrit à la sous section 1.1.3, est celui qui maximise la probabilité marginale *a posteriori* en chaque site. Il est donné, rappelons-le, par

$$\forall n \in \mathcal{N}, (\hat{x}_n)_{MPM}(y) = \arg \max_{x'_n \in \mathcal{X}} p(x'_n|y) \quad (2.1.2)$$

L'estimation d'une réalisation du processus caché au sens du MPM nécessite donc le calcul des marginales *a posteriori*  $p(x_n|y)$  en chaque site  $n \in \mathcal{N}$ . De la même manière que pour la loi de  $Z$  dans (2.1.1), la loi du processus  $X$  *a posteriori* peut être écrite sous une forme factorisée :

$$p(x|y) = p(x_1|y) \prod_{n=2}^N p(x_n|x_{1:n-1}, y) \quad (2.1.3)$$

Cette dernière est donnée par le produit de la loi *a posteriori*  $p(x_1|y)$  et des lois de transition *a posteriori*  $\{p(x_n|x_{1:n-1}, y)\}_{n=2}^N$ . Dans le cas où ces lois sont connues, les probabilités marginales *a posteriori* peuvent être calculées en effectuant un parcours progressif de la chaîne appelé *passé avant*. En effet, connaissant  $p(x_1|y)$ , nous pouvons calculer  $p(x_n|y) = \sum_{x_{1:n-1} \in \mathcal{X}^{n-1}} p(x_{1:n}|y)$  avec  $p(x_{1:n}|y) = p(x_{1:n-1}|y)p(x_n|x_{1:n-1}, y)$  pour  $n = 2 \dots N$ . Le problème se ramène donc au calcul de ces probabilités. En ce qui concerne le calcul des probabilités de transition *a posteriori*, le calcul de celles-ci peut être écrit sous la forme suivante pour tout  $n \in \mathcal{N} \setminus \{1\}$  :

$$p(x_n|x_{1:n-1}, y) = p(z_n|z_{1:n-1}) \frac{\beta_n(z_{1:n})}{\beta_{n-1}(z_{1:n-1})} \quad (2.1.4)$$

avec  $\beta_{n-1}(z_{1:n-1}) = p(y_{n:N}|z_{1:n-1})$ .

En effet,

$$\begin{aligned} p(x_n|x_{1:n-1}, y) &= \frac{p(x_{1:n}, y)}{p(x_{1:n-1}, y)} \\ &= \frac{\sum_{x_{n+1:N}} p(z_{1:N})}{\sum_{x_{n:N}} p(z_{1:N})} \\ &= \frac{p(z_{1:n}) \sum_{x_{n+1:N}} p(z_{n+1:N}|z_{1:n})}{p(z_{1:n-1}) \sum_{x_{n:N}} p(z_{n:N}|z_{1:n-1})} \\ &= p(z_n|z_{1:n-1}) \frac{\sum_{x_{n+1:N}} p(z_{n+1:N}|z_{1:n})}{\sum_{x_{n:N}} p(z_{n:N}|z_{1:n-1})} \\ &= p(z_n|z_{1:n-1}) \frac{\beta_n(z_{1:n})}{\beta_{n-1}(z_{1:n-1})} \end{aligned} \quad (2.1.5)$$

La quantité  $\beta_{n-1}(z_{1:n-1})$ , dite *rétrograde* peut elle-même être exprimée en fonction de  $p(z_n|z_{1:n-1})$  et de  $\beta_n(z_{1:n})$ , en effet,

$$\begin{aligned}\beta_{n-1}(z_{1:n-1}) &= \sum_{x_{n:N}} p(z_{n:N}|z_{1:n-1}) \\ &= \sum_{x_n} p(z_n|z_{1:n-1}) \underbrace{\sum_{x_{n+1:N}} p(z_{n+1:N}|z_{1:n})}_{\beta_n(z_{1:n})}\end{aligned}$$

Il s'ensuit que les quantités  $\beta_n(z_{1:n})$  peuvent alors être calculées récursivement pour  $n \in \mathcal{N} \setminus [N]$  en effectuant un parcours rétrograde, ou *passé arrière*, sur la grille  $\mathcal{N}$  de la manière suivante :

$$\beta_{n-1}(z_{1:n-1}) \propto \begin{cases} \sum_{x_n \in \mathcal{X}} p(z_n|z_{1:n-1}) & \text{si } n = N \\ \sum_{x_n \in \mathcal{X}} p(z_n|z_{1:n-1}) \beta_n(z_{1:n}) & \text{sinon} \end{cases} \quad (2.1.6)$$

Nous pouvons en déduire l'ensemble des probabilités de transitions *a posteriori*  $p(x_n|x_{1:n-1}, y)$ , pour  $n \in \mathcal{N} \setminus \{1\}$ , à partir de (2.1.4). Il est néanmoins nécessaire que les probabilités des transitions  $p(z_n|z_{1:n-1})$ , pour  $n \in \mathcal{N} \setminus \{1\}$ , du processus  $Z$  soient connues ou calculables. D'autre part, lors du calcul des  $\beta_n(z_{1:n})$ , pour  $n \in \mathcal{N} \setminus \{N\}$ , il est très courant de rencontrer des problèmes de dépassement de possibilités de représentation numérique. En effet, d'après (2.1.6), les variables  $\beta_n(z_{1:n})$  ont une dynamique qui tend à croître de manière exponentielle avec  $N - n$ . Une solution classique (*scaling* [105]) consiste à les normaliser à chaque étape par  $c_n = \sum_{z_{1:n} \in (\mathcal{X} \times \mathcal{Y})^n} \beta_n(z_{1:n})$ . Finalement, la probabilité marginale *a posteriori* en  $n = 1$  peut être calculée à partir de  $\beta_1(z_1)$  de la manière suivante :

$$p(x_1|y) = p(z_1) \frac{\beta_1(z_1)}{\sum_{x_1 \in \mathcal{X}} p(z_1) \beta_1(z_1)} \quad (2.1.7)$$

L'algorithme de calcul des marginales *a posteriori* est résumé dans l'algorithme 3. Il consiste à calculer dans un premier temps les probabilités de transition *a posteriori* lors de la passe arrière de  $n = N$  à  $n = 2$ . La probabilité marginale *a posteriori* est ensuite calculée en  $n = 1$ . On peut alors en déduire les probabilités marginales *a posteriori* lors de la passe avant. Les probabilités marginales *a posteriori*  $\{p(x_n|y)\}_{n=1}^N$  étant connues, on peut estimer une réalisation des variables cachées au sens du MPM dont l'estimateur est donné par (2.1.2). Notons que la connaissance de ces probabilités marginales ainsi que des probabilités marginales  $p(x_{1:n}|y)$  sera également très utile pour l'estimation des paramètres du modèle. Enfin, cet algorithme de calcul des probabilités marginales *a posteriori* peut également être utilisé pour l'estimation d'une réalisation du processus caché au sens d'autres estimateurs locaux dont la forme générale est donnée par (1.1.9).

---

**Algorithme 3** Calcul des marginales *a posteriori* dans le cas d'un processus monodimensionnel.

---

◁ Passé arrière : pour  $n = N \dots 2$

$$\beta_{n-1}(z_{1:n-1}) \propto \begin{cases} \sum_{x_n \in \mathcal{X}} p(z_n | z_{1:n-1}) & \text{si } n = N \\ \sum_{x_n \in \mathcal{X}} p(z_n | z_{1:n-1}) \beta_n(z_{1:n}) & \text{sinon} \end{cases}$$

$$p(x_n | x_{1:n-1}, y) \propto \begin{cases} p(z_n | z_{1:n-1}) & \text{si } n = N \\ p(z_n | z_{1:n-1}) \frac{\beta_n(z_{1:n})}{\beta_{n-1}(z_{1:n-1})} & \text{sinon} \end{cases}$$

$$\underline{n=0} : p(x_1 | y) \propto p(z_1) \beta_1(z_1)$$

▷ Passé avant : pour  $n = 2 \dots N$

$$p(x_{1:n} | y) = p(x_{1:n-1} | y) p(x_n | x_{1:n-1}, y)$$

$$p(x_n | y) = \sum_{x_{1:n-1} \in \mathcal{X}^{n-1}} p(x_{1:n} | y)$$


---

### 2.1.2 Calcul du MAP

L'estimateur du MAP est celui qui maximise la probabilité *a posteriori*. Celui-ci est donné, rappelons-le, par

$$\hat{x}_{MAP}(y) = \arg \max_{x' \in \mathcal{X}^N} p(x' | y) \quad (2.1.8)$$

Dans le cas où l'on souhaite estimer une réalisation du processus caché au sens du MAP nous considérons la quantité suivante :

$$\delta_n(z_{1:n}) = \max_{x_{n+1:N} \in \mathcal{X}^{N-n}} p(z_{n+1:N} | z_{1:n}) \quad (2.1.9)$$

Celle-ci correspond à la probabilité du chemin le plus probable allant de  $n+1$  jusqu'à  $N$ . On recherche donc le MAP sur la sous-chaîne allant de l'instant  $n+1$  à  $N$  avec  $X_{n+1}$  fixé. On obtient la récurrence suivante que l'on applique lors d'une passe arrière pour  $n = N \dots 2$  :

$$\delta_{n-1}(z_{1:n-1}) = \max_{x_n \in \mathcal{X}} p(z_n | z_{1:n-1}) \delta_n(z_{1:n}) \quad (2.1.10)$$

en effet,

$$\begin{aligned} \delta_{n-1}(z_{1:n-1}) &= \max_{x_{n:N} \in \mathcal{X}^{N-n+1}} p(z_{n:N} | z_{1:n-1}) \\ &= \max_{x_{n:N} \in \mathcal{X}^{N-n+1}} p(z_{n+1:N} | z_{1:n}) p(z_n | z_{1:n-1}) \\ &= \max_{x_n \in \mathcal{X}} p(z_n | z_{1:n-1}) \max_{x_{n+1:N} \in \mathcal{X}^{N-n}} p(z_{n+1:N} | z_{1:n}) \\ &= \max_{x_n \in \mathcal{X}} p(z_n | z_{1:n-1}) \delta_n(z_{1:n}) \end{aligned}$$

Lors de cette passe arrière, il est nécessaire de conserver les valeurs d'état réalisant les maxima

$$\psi_n(x_n) = \arg \max_{x_n \in \mathcal{X}} p(z_n | z_{1:n-1}) \delta_{n+1}(z_{1:n}) \quad (2.1.11)$$

---

**Algorithme 4** Calcul du MAP dans le cas d'un processus monodimensionnel.

---

◁ Passé arrière : pour  $n = N \dots 2$

$$\delta_{n-1}(z_{1:n-1}) \propto \begin{cases} \max_{x_n \in \mathcal{X}} p(z_n | z_{1:n-1}) & \text{si } n = N \\ \max_{x_n \in \mathcal{X}} p(z_n | z_{1:n-1}) \delta_n(z_{1:n}) & \text{sinon} \end{cases}$$

$$\psi_n(x_n) \propto \begin{cases} \arg \max_{x_n \in \mathcal{X}} p(z_n | z_{1:n-1}) & \text{si } n = N \\ \arg \max_{x_n \in \mathcal{X}} p(z_n | z_{1:n-1}) \delta_n(z_{1:n}) & \text{sinon} \end{cases}$$

$$\underline{n=0} : \hat{x}_1 = \arg \max_{x_1 \in \mathcal{X}} p(z_1) \delta_1(z_1)$$

▷ Passé avant : pour  $n = 2 \dots N$

$$\hat{x}_{n+1} = \psi_n(\hat{x}_n)$$


---

Une fois connu l'état optimal à  $n = 1$  en calculant  $\hat{x}_1 = \arg \max_{x_1 \in \mathcal{X}} p(z_1) \delta_1(z_1)$ , la passe avant consiste simplement à récupérer de proche en proche toutes les autres composantes du MAP. L'algorithme est résumé dans l'algorithme 4.

Les deux algorithmes présentés permettent l'estimation d'une réalisation du processus caché au sens du MPM et du MAP. Ces derniers ne diffèrent qu'au niveau des sommations et des maximisations. Notons que ces algorithmes peuvent être également présentés en utilisant un formalisme à base de passage de "messages" [104]. Ces algorithmes sont applicables à condition que les probabilités  $p(z_1)$  et  $\{p(z_n | z_{1:n-1})\}_{n=2}^N$  soient calculables. Or, cela devient impossible pour de grandes valeurs de  $N$  étant donné le nombre considérable de probabilités de transition à calculer dû à la dépendance en  $z_{1:n-1}$ . Il est donc nécessaire d'introduire des hypothèses permettant de simplifier le modèle afin de rendre ces probabilités de transition calculables.

## 2.2 Modélisations de la loi de $Z$

Une première simplification de la forme générale de  $p(z)$  consiste à supposer l'indépendance des v.a  $Z_{1:N}$ , l'indépendance de deux variables aléatoires étant définie de la manière suivante :

**Définition 2.2.1.** *Deux v.a  $A$  et  $B$  à valeurs respectivement dans  $\mathcal{A}$  et  $\mathcal{B}$  sont dites statistiquement indépendantes si et seulement si*

$$\forall (a, b) \in \mathcal{A} \times \mathcal{B}, \quad p(a, b) = p(a)p(b) \quad (2.2.1)$$

Nous utiliserons dans ce cas la notation suivante  $A \perp\!\!\!\perp B$  pour indiquer que  $A$  est indépendante de  $B$ . Ainsi, nous faisons l'hypothèse suivante

$$\forall n \in \mathcal{N}, \quad Z_n \perp\!\!\!\perp Z_{\mathcal{N} \setminus n} \quad (2.2.2)$$

où  $Z_{\mathcal{N} \setminus n} = Z_{1:n-1} \cup Z_{n+1:N}$ . Les probabilités de transition de (2.1.1) se simplifient alors pour tout  $n \in \mathcal{N} \setminus 1$  en  $p(z_n | z_{1:n-1}) = p(z_n)$ . Le processus  $Z$  est alors dit *indépendant et*



*identiquement distribué* (i.i.d) et la loi de  $Z$  donnée par (2.1.1) est simplifiée de la manière suivante

$$p(z) = \prod_{n=1}^N p(z_n) \quad (2.2.3)$$

Il devient alors très simple de calculer les probabilités marginales *a posteriori* et d'estimer une réalisation de  $x_n$  en chaque site au sens du MPM (le MAP étant dans ce cas équivalent). Bien que cette modélisation soit séduisante d'un point de vue calculatoire, elle n'en est pas moins incomplète car nous ne tenons pas compte de la dépendance entre les v.a. Or la dépendance spatiale peut être cruciale, notamment en segmentation d'image pour laquelle un des objectifs est d'obtenir des classes spatialement homogènes. Une manière de s'affranchir de ce problème et d'utiliser la notion d'indépendance conditionnelle définie de la manière suivante

**Définition 2.2.2.** Une v.a  $A$  est dite *conditionnellement indépendante* d'une deuxième v.a  $B$  étant donné une troisième v.a  $C$  si la relation suivante est vérifiée :

$$\forall(a, b, c) \in \mathcal{A} \times \mathcal{B} \times \mathcal{C}, \quad p(a|b, c) = p(a|c) \quad (2.2.4)$$

Nous utiliserons dans ce cas la notation suivante  $A \perp\!\!\!\perp B|C$  pour indiquer que  $A$  est indépendante de  $B$  conditionnellement à  $C$ . L'indépendance conditionnelle de  $A$  et  $B$  étant donné  $C$  peut être interprétée de la manière intuitive suivante : si nous observons une réalisation de  $C$ , alors la connaissance d'une réalisation de  $B$  ne changera pas la connaissance sur  $A$ . Néanmoins, cela ne signifie pas que  $A$  soit indépendante de  $B$ . En effet, et c'est cette propriété qui sera fondamentale par la suite, si l'on supprime le conditionnement, l'indépendance n'a pas forcément lieu et  $A$  peut être dépendante de  $B$ . La notion d'indépendance conditionnelle est un concept puissant et elle permet de simplifier énormément les modèles statistiques comme nous allons le voir. D'autre part, une manière simple de représenter la structure de dépendance d'un modèle caché est le graphe d'indépendance conditionnelle [119]. Dans ce type de graphe, l'absence d'une arête entre deux nœuds indique que les deux variables concernées sont conditionnellement indépendantes par rapport aux autres variables.

**Remarque** Nous pouvons réécrire (2.2.2) sous la forme des deux hypothèses d'indépendance conditionnelle suivantes concernant la loi de  $X$  et la loi de  $Y$  :

$$\forall n \in \mathcal{N}, \quad \begin{cases} X_n \perp\!\!\!\perp X_{\mathcal{N} \setminus n} \\ Y_n \perp\!\!\!\perp Z_{\mathcal{N} \setminus n} | X_n \end{cases} \quad (2.2.5)$$

La loi de  $Z_n$  est donc donnée par  $p(z_n) = p(x_n)p(y_n|x_n)$ . Pour une classe  $k$  donnée,  $p(y_n|x_n = k)$  sera une densité paramétrée choisie afin de modéliser les caractéristiques observées de la classe  $k$ . Nous pouvons citer par exemple les lois normales, gamma, Rayleigh, Weibull pour les données sonar, K pour les données radar, ou les lois du système de Pearson [38, 54]. Par

la suite nous ne considérerons que le cas gaussien.

### 2.2.1 Chaînes de Markov cachées à bruit indépendant

Nous pouvons relâcher l'hypothèse sur  $X$  dans (2.2.5) en supposant que  $X_{n+1}$  est indépendant de  $X_{1:n-1}$  conditionnellement à  $X_n$  pour tout  $n \in \mathcal{N} \setminus N$  :

$$\forall n \in \mathcal{N}, \quad \begin{cases} X_{n+1:N} \perp\!\!\!\perp X_{1:n-1} | X_n \\ Y_n \perp\!\!\!\perp Z_{\mathcal{N} \setminus n} | X_n \end{cases} \quad (2.2.6)$$

$X$  est alors une *chaîne de Markov* d'ordre 1 (CM) étant donné que  $p(x_n | x_{1:n-1}) = p(x_n | x_{n-1})$  d'après (2.2.4). D'autre part, on peut remarquer que  $Z$  est également une CM étant donné que d'après (2.2.7) :

$$p(z_n | z_{1:n-1}) = p(x_n | x_{n-1}) p(y_n | x_n) = p(z_n | z_{n-1}) \quad (2.2.7)$$

On peut en déduire que la loi *a posteriori* est celle d'une CM. Pour cela, on part de la forme générale des transitions *a posteriori* donnée par (2.1.5). Étant donné que  $p(z_n | z_{1:n-1}) = p(z_n | z_{n-1})$ , le résultat est immédiat :

$$\begin{aligned} p(x_n | x_{1:n-1}, y) &= p(z_n | z_{1:n-1}) \frac{\sum_{x_{n+1:N}} p(z_{n+1:N} | z_{1:n})}{\sum_{x_{n:N}} p(z_{n:N} | z_{1:n-1})} \\ &= p(z_n | z_{n-1}) \frac{\sum_{x_{n+1:N}} p(z_{n+1:N} | z_n)}{\sum_{x_{n:N}} p(z_{n:N} | z_{n-1})} \text{ d'après (2.2.7)} \\ &= p(x_n | x_{n-1}, y) \end{aligned}$$

Les probabilités de transition *a posteriori* sont donc données en chaque  $n \in \mathcal{N} \setminus \{1\}$  par :

$$p(x_n | x_{n-1}, y) = p(x_n | x_{n-1}) p(y_n | x_n) \frac{\beta_n(x_n)}{\beta_{n-1}(x_{n-1})} \quad (2.2.8)$$

avec  $\beta_{n-1}(x_{n-1}) = p(y_{n:N} | x_{n-1})$  calculable récursivement pour tout  $n \in \mathcal{N} \setminus \{N\}$  de la manière suivante :

$$\beta_{n-1}(x_{n-1}) \propto \begin{cases} \sum_{x_n \in \mathcal{X}} p(x_n | x_{n-1}) p(y_n | x_n) & \text{si } n = N \\ \sum_{x_n \in \mathcal{X}} p(x_n | x_{n-1}) p(y_n | x_n) \beta_n(x_n) & \text{sinon} \end{cases} \quad (2.2.9)$$

Cette propriété de markovianité de la loi *a posteriori*, qui sera fondamentale pour la définition de modèles plus généraux, réduit considérablement la complexité des algorithmes d'inférence présentés précédemment dans le cas général et rend possible leur utilisation pratique. En effet, les probabilités de transition données par (2.2.7) deviennent calculables même pour de grandes valeurs de  $N$ . Il suffit pour cela de connaître les densités  $p(x_n | x_{n-1})$  et  $p(y_n | x_n)$

pour tout  $n \in \mathcal{N}$ . De plus, si on considère que  $Z$  est un processus stationnaire comme nous le ferons par la suite, alors on aura  $p(z_{n-1}, z_n) = p(z_1, z_2)$  pour tout  $n \in \mathcal{N} \setminus \{1\}$  et seule la connaissance de la densité  $p(z_1, z_2)$  sera nécessaire pour effectuer les calculs. On peut alors estimer une réalisation du processus caché en utilisant les algorithmes 3 et 4 avec (2.2.7). Nous obtenons ainsi un modèle simple permettant de tenir compte de la dépendance entre v.a de sites voisins.

Ce modèle, que nous appellerons *chaînes de Markov cachées à bruit indépendant* (CMC-BI) étant donné les hypothèses considérées (2.2.6), a été, et est encore, très utilisé en analyse et en traitement de signaux monodimensionnels. Les CMC-BI ont été introduites pour la première fois en 1966 par Baum et Petrie [2]. En 1970, ces derniers, en compagnie de Soules et Weiss [3, 1], ont proposé des récursions dites *progressives-rétrogrades* pour le calcul des probabilités marginales *a posteriori* qui sont en tout point équivalentes aux passes avant/arrière que nous avons présentées dans le cas général dans l’algorithme 3. Ils ont également proposé une procédure de calcul itérative pour l’estimation par maximum de vraisemblance des paramètres d’une CMC-BI utilisant les récursions progressives-rétrogrades. Cette procédure, souvent référencée sous le nom d’algorithme de Baum-Welch, correspond à l’algorithme Expectation-Maximisation (EM) de Dempster, Laird et Rubin [40] appliqué aux CMC-BI. Nous proposerons, dans le suite de ce chapitre, une extension de cet algorithme à l’estimation des paramètres de modèles plus généraux. Enfin, Viterbi [116] puis Fornay [50] ont proposé un algorithme de calcul du MAP semblable à l’algorithme 4. Une des premières applications des CMC-BI fut la reconnaissance automatique des caractères. Puis, vers le milieu des années 1970, une autre application majeure est apparue : la reconnaissance automatique de la parole. Les applications des CMC-BI à la reconnaissance automatique de la parole furent étudiées principalement par Ferguson et ses collègues à l’Institute for Defence Analysis et par Rabiner et son équipe aux laboratoires AT&T Bell [105]. Ces études ont popularisé la théorie des CMC-BI qui se sont alors étendues vers d’autres nombreuses applications dans les domaines tels que les biosciences [63, 85], la climatologie [114], les communications [23], l’écologie [10], l’économie et les finances [113] ou encore le traitement du signal et des images [20, 23, 106].

### 2.2.2 Chaînes de Markov couples

Nous venons de voir qu’une condition pour que les algorithmes d’inférence bayésienne soient applicables numériquement est que le processus *a posteriori* soit une CM. Dans le cas des CMC-BI, ceci est dû au fait que  $Z$  est une CM, la markovianité de  $Z$  étant due aux hypothèses (2.2.6). Cependant, ces hypothèses peuvent, dans certaines applications, être trop fortes. Tout d’abord, concernant la loi de  $X$ , le processus caché peut ne pas être markovien. D’autre part, concernant les observations, l’hypothèse d’indépendance conditionnellement aux états est discutable. Par exemple, en reconnaissance de parole (indépendante du locuteur), les mêmes états sont utilisés pour modéliser la parole de locuteurs possédant différentes

caractéristiques de voix. Or, ces caractéristiques spécifiques aux locuteurs se traduisent, en grande partie, par la corrélation entre les observations voisines. L'hypothèse d'indépendance des observations conditionnellement aux états ne permet pas de prendre en compte ces corrélations. Ces constatations ont conduit à la proposition d'autres modèles pour lesquels cette hypothèse d'indépendance n'est plus retenue (voir par exemple [118, 87]). L'idée des *chaînes de Markov couples* [97, 42] est de supposer directement la markovianité de  $Z$

$$\forall n \in \mathcal{N} \setminus N \quad Z_{n+1:N} \perp\!\!\!\perp Z_{1:n-1} | Z_n \quad (2.2.10)$$

De la même manière que pour les CMC-BI, nous pouvons en déduire que le processus  $a$  *posteriori* est une CM<sup>2</sup>, son noyau de transition en chaque  $n \in \mathcal{N} \setminus \{1\}$  étant donné par :

$$p(x_n | x_{n-1}, y) = p(z_n | z_{n-1}) \frac{\beta_n(z_n)}{\beta_{n-1}(z_{n-1})} \quad (2.2.11)$$

avec  $\beta_{n-1}(z_{n-1}) = p(y_{n:N} | z_{n-1})$  calculable récursivement pour tout  $n \in \mathcal{N} \setminus \{N\}$  de la manière suivante :

$$\beta_{n-1}(z_{n-1}) \propto \begin{cases} \sum_{x_n \in \mathcal{X}} p(z_n | z_{n-1}) & \text{si } n = N \\ \sum_{x_n \in \mathcal{X}} p(z_n | z_{n-1}) \beta_n(z_n) & \text{sinon} \end{cases} \quad (2.2.12)$$

En considérant l'hypothèse d'indépendance (2.2.10), les probabilités de transition de (2.1.1) se simplifient, pour tout  $n \in \mathcal{N} \setminus \{1\}$  :

$$p(z_n | z_{1:n-1}) = p(z_n | z_{n-1}) = p(x_n | z_{n-1}) p(y_n | x_n, z_{n-1}) \quad (2.2.13)$$

Elles sont alors calculables pour tout  $n \in \mathcal{N} \setminus \{1\}$ . En effet, elles peuvent être décomposées sous la forme suivante :

$$p(z_n | z_{n-1}) = p(x_n | x_{n-1}) \frac{p(y_{n-1:n} | x_{n-1:n})}{p(y_{n-1} | x_{n-1})} = \frac{p(x_{n-1:n}) p(y_{n-1:n} | x_{n-1:n})}{\sum_{x_n \in \mathcal{X}} p(x_{n-1:n}) p(y_{n-1} | x_{n-1:n})} \quad (2.2.14)$$

Il suffit alors de connaître les densités  $p(x_{n-1:n})$  et  $p(y_{n-1:n} | x_{1:n})$ , pour pouvoir calculer les probabilités de transition. Enfin, dans le cas où  $Z$  est stationnaire, il suffit de connaître  $p(z_{1:2}) = p(x_{1:2}) p(y_{1:2} | x_{1:2})$ .

**Remarque** Dans le cas gaussien où la densité  $p(y_{n-1:n} | x_{n-1:n})$  est gaussienne, de moyennes  $\mu_{n-1}^{x_{n-1}, x_n}$ ,  $\mu_n^{x_{n-1}, x_n}$  et d'écart-types  $\sigma_{n-1}^{x_{n-1}, x_n}$ ,  $\sigma_n^{x_{n-1}, x_n}$ , il peut être utile de décomposer la densité  $p(y_{n-1:n} | x_{n-1:n})$  sous la forme suivante :

$$p(y_{n-1:n} | x_{n-1:n}) = p(y_{n-1} | x_{n-1:n}) p(y_n | x_n, z_{n-1}) \quad (2.2.15)$$

---

<sup>2</sup>On peut également montrer que le processus  $Y$  conditionnel à  $X$  est une CM.

Dans ce cas,  $p(y_n|x_n, z_{n-1})$  est une densité gaussienne ayant pour moyenne et pour écart-type :

$$\begin{cases} \mu_n^{x_{n-1}, x_n} + \rho^{x_{n-1}, x_n} (\sigma_n^{x_{n-1}, x_n} / \sigma_{n-1}^{x_{n-1}, x_n}) (y_n - \mu_{n-1}^{x_{n-1}, x_n}) \\ \sigma_n^{x_{n-1}, x_n} \sqrt{1 - (\rho^{x_{n-1}, x_n})^2} \end{cases} \quad (2.2.16)$$

Il est alors possible d'appliquer les algorithmes d'inférence bayésienne vus précédemment avec (2.2.13). L'avantage par rapport au modèle CMC-BI est de pouvoir considérer des lois de  $Y$  conditionnellement à  $X$  plus complexes pouvant mieux correspondre aux données traitées et d'autre part que le processus caché  $X$  n'est plus nécessairement une CM d'où une plus grande généralité. Le lemme suivant, spécifie cette généralité.

Soit  $W = (G, H) = (G_n, H_n)_{n \in \mathcal{N}}$  un p.a pour lequel chaque  $W_n = (G_n, H_n)$  prend ses valeurs dans  $\mathcal{G} \times \mathcal{H}$ . Soit  $\gamma$  une mesure  $\sigma$ -additive sur  $\mathcal{G}$  et soit  $\eta$  une mesure  $\sigma$ -additive sur  $\mathcal{H}$ . Supposons que  $W$  soit une CM.

**Lemme 2.2.1.** *Pour une chaîne de Markov  $W = (G, H)$  vérifiant*

- $p(w_n, w_{n+1})$  ne dépend pas de  $n \in \mathcal{N}$
- $p(w_n = a, w_{n+1} = b) = p(w_n = b, w_{n+1} = a) \forall n \in \mathcal{N}, \forall (a, b) \in (\mathcal{G} \times \mathcal{H})^2$

Les trois propositions suivantes sont équivalentes :

1.  $G$  est une CM;
2.  $\forall n \in \mathcal{N} \setminus \{1\}, p(h_n | g_{n:n-1}) = p(h_n | g_n)$ ;
3.  $\forall n \in \mathcal{N}, p(h_n | g) = p(h_n | g_n)$ .

**Preuve :**

- 1 implique 2 :  $W$  étant une CM, nous pouvons écrire

$$\begin{aligned} p(w) &= p(w_1) \prod_{n=2}^N p(w_n | w_{n-1}) = p(w_1) \prod_{n=2}^N \frac{p(w_{n-1:n})}{p(w_{n-1})} \\ &= \frac{\prod_{n=2}^N p(w_{n-1:n})}{\prod_{n=3}^N p(w_{n-1})} = \frac{\prod_{n=2}^N p(h_{n-1:n}, g_{n-1:n})}{\prod_{n=3}^N p(h_{n-1}, g_{n-1})} \\ &= \underbrace{\frac{\prod_{n=2}^N p(g_{n-1:n})}{\prod_{n=3}^N p(g_{n-1})}}_{a(g)} \times \underbrace{\frac{\prod_{n=2}^N p(h_{n-1:n} | g_{n-1:n})}{\prod_{n=3}^N p(h_{n-1} | g_{n-1})}}_{b(h)} \end{aligned} \quad (2.2.17)$$

En particulier,  $W_{1:3}$  est également une CM et ainsi (2.2.17) est valide pour  $N = 3$ . Considérons donc (2.2.17) avec  $N = 3$ . Etant donné que  $G$  est une CM selon (1),  $a(g_{1:3})$  est la loi de  $G$ , ce qui signifie que l'intégrale de  $b(h_{1:3})$  vaut 1. Après avoir intégré  $b(h_{1:3})$  par rapport à  $h_1$  et  $h_3$  nous obtenons

$$\int_{\mathcal{H}} \frac{p(h_2 | g_{1:2}) p(h_2 | g_{2:3})}{p(h_2 | g_2)} d\nu(h_2) = 1 \quad (2.2.18)$$

Considérons à présent le produit scalaire  $\langle f^1, f^2 \rangle = \int \frac{f^1(h_2)f^2(h_2)}{p(h_2|g_2)} d\nu(h_2)$  et la norme associée  $\|f\| = \sqrt{\langle f, f \rangle}$ . Par rapport à (2.2.18), nous avons :

$$\begin{aligned} \|p(h_2|g_{1:2}) - p(h_2|g_{2:3})\|^2 &= \|p(h_2|g_{1:2})\|^2 + \|p(h_2|g_{2:3})\|^2 \\ &\quad - 2 \int_{\mathcal{H}} \frac{p(h_2|g_{1:2})p(h_2|g_{2:3})}{p(h_2|g_2)} d\nu(h_2) = 0 \end{aligned}$$

et ainsi  $p(h_2|g_{1:2}) = p(h_2|g_{2:3})$  pour tout  $g_1, g_2$ . Cela signifie que  $p(h_2|g_{1:2})$  est indépendant de  $g_1$ , ce qui donne 2.

- 2 implique 1 et 3 : étant donné que  $p(h_n|g_{n-1:n}) = p(h_n|g_{n:n+1}) = p(h_n|g_n)$ , l'intégration de  $b(h)$  par rapport à  $h_{1:N}$  donne 1, ce qui signifie que  $p(g) = a(g)$  et que  $G$  est une CM. De plus, l'intégration de  $b(h)$  par rapport à  $h_{\mathcal{N} \setminus n}$  donne  $p(g, h_n) = a(g)p(h_n|g_n) = p(g)p(h_n|g_n)$ . Ainsi,  $p(h_n|g) = \frac{p(g, h_n)}{p(g)} = \frac{p(g)p(h_n|g_n)}{p(g)} = p(h_n|g_n)$
- 3 implique 2 : nous avons :

$$\begin{aligned} p(h_n|g_{n-1}, g_n) &= \int_{H^{N-2}} p(g_{\mathcal{N} \setminus \{n-1, n\}}) p(h_n|g_{1:N}) d\gamma^{N-2}(g_{\mathcal{N} \setminus \{n-1, n\}}) \\ &= \int_{H^{N-2}} p(g_{\mathcal{N} \setminus \{n-1, n\}}) p(h_n|g_n) d\gamma^{N-2}(g_{\mathcal{N} \setminus \{n-1, n\}}) \\ &= p(h_n|g_n) \int_{H^{N-2}} p(g_{\mathcal{N} \setminus \{n-1, n\}}) d\gamma^{N-2}(g_{\mathcal{N} \setminus \{n-1, n\}}) \\ &= p(h_n|g_n) \end{aligned}$$

ce qui termine la preuve.

Ce lemme fut proposé tout d'abord dans [97] dans le cadre des chaînes de Markov couples puis étendu aux modèles de Markov triplets (que nous étudierons au chapitre suivant), dans [100] et [68]. Des travaux similaires ont par ailleurs été effectués dans le cas où  $\mathcal{X}$  est un ensemble continu [45, 44, 43]. Ce lemme est pratique pour évaluer le degré de généralité d'un modèle donné. Par exemple, nous nous intéressons ici à une CM Couple  $Z = (X, Y)$  telle que  $X_n$  prend ses valeurs dans un ensemble discret  $\mathcal{X}$ , chaque  $Y_n$  prenant ses valeurs dans l'ensemble des nombres réels  $\mathbb{R}$ . Quand  $Z$  est stationnaire, sa loi est donnée par  $p(z_{1:2})$ . Nous cherchons à savoir si  $X$  est une CM ou non ; pour cela, nous pouvons appliquer le lemme à  $G = X$ , avec  $\gamma$  la mesure de comptage sur  $\mathcal{G} = \mathcal{X}$  et  $H = Y$ , avec  $\eta$  la mesure de Lebesgue sur  $\mathbb{R}$ . En supposant que les hypothèses du lemme sont vérifiées et en écrivant  $p(z_{1:2}) = p(x_{1:2})p(y_{1:2}|x_{1:2})$ , nous pouvons dire que  $X$  est une CM si et seulement si  $p(y_1|x_{1:2}) = p(y_1|x_1)$ .

Connaissant les conditions pour lesquelles dans une CM Couple,  $X$  est une CM, nous pouvons à présent proposer différents modèles, de généralités croissantes, fondés sur le modèle CM Couple. Nous donnons pour chaque modèle la forme des probabilités de transition. Les graphes d'indépendance conditionnelle de chaque modèle, sont représentés sur la figure 2.2.1.

**Chaînes de Markov cachées à bruit indépendant (CMC-BI) :** Dans ce premier modèle,  $p(y_n|x_{n-1}, x_n)$  ne dépend que de  $x_n$  et les observations ne sont pas corrélées. D'après le lemme,  $X$  est une CM et

$$p(y_{n-1:n}|x_{n-1:n}) = p(y_n|x_n)p(y_{n-1}|x_{n-1}) \quad (2.2.19)$$

Nous en déduisons la forme des probabilités de transition pour le modèle CMC-BI :

$$p(z_n|z_{n-1}) = p(x_n|x_{n-1})p(y_n|x_n) \quad (2.2.20)$$

En effet, nous avons

$$\begin{aligned} p(z_n|z_{n-1}) &= p(x_n|x_{n-1}) \frac{p(y_{n-1:n}|x_{n-1:n})}{p(y_{n-1}|x_{n-1})} \\ &= p(x_n|x_{n-1}) \frac{p(y_n|x_n)p(y_{n-1}|x_{n-1})}{p(y_{n-1}|x_{n-1})} \text{ d'après (2.2.19)} \\ &= p(x_n|x_{n-1})p(y_n|x_n) \end{aligned}$$

Nous retrouvons le modèle CMC-BI décrit précédemment qui est un cas particulier de CM-Couple.

**Chaînes de Markov cachées (CMC)** Nous prenons à présent en compte les corrélations entre les observations tout en conservant l'hypothèse que  $p(y_n|x_n, x_{n-1})$  ne dépend que de  $x_n$ .  $X$  est donc toujours une CM et

$$p(y_{n-1:n}|x_{n-1:n}) = p(y_n|y_{n-1}, x_n)p(y_{n-1}|x_{n-1}) \quad (2.2.21)$$

Nous en déduisons la forme de probabilités de transitions pour le modèle CMC :

$$p(z_n|z_{n-1}) = p(x_n|x_{n-1})p(y_n|y_{n-1}, x_n) \quad (2.2.22)$$

En effet,

$$\begin{aligned} p(z_n|z_{n-1}) &= p(x_n|x_{n-1}) \frac{p(y_{n-1:n}|x_{n-1:n})}{p(y_{n-1}|x_{n-1})} \\ &= p(x_n|x_{n-1}) \frac{p(y_n|y_{n-1}, x_n)p(y_{n-1}|x_{n-1})}{p(y_{n-1}|x_{n-1})} \text{ d'après (2.2.21)} \\ &= p(x_n|x_{n-1})p(y_n|y_{n-1}, x_n) \end{aligned}$$

L'avantage du modèle CMC est qu'il permet de prendre en compte la corrélation entre observations tout en ayant un nombre de paramètres réduit.

**Chaînes de Markov couples à bruit indépendant (CMCouple-BI)** Dans ce modèle,  $p(y_n|x_{n-1}, x_n)$  dépend à la fois de  $x_n$  et de  $x_{n-1}$ . D'après le lemme,  $X$  n'est plus une CM d'ordre 1. Néanmoins, on peut montrer qu'il s'agit d'une CM d'ordre 2. Nous supposons également que les observations sont indépendantes conditionnellement à  $X$ . Nous avons donc

$$p(y_{n-1:n}|x_{n-1:n}) = p(y_n|x_{n-1:n})p(y_{n-1}|x_{n-1:n}) \quad (2.2.23)$$

Nous en déduisons la forme des probabilités de transitions dans le cas CMCouple-BI :

$$p(z_n|z_{n-1}) = p(x_n|z_{n-1})p(y_n|x_{n-1:n}) \quad (2.2.24)$$

En effet

$$\begin{aligned} p(z_n|z_{n-1}) &= p(x_n|x_{n-1}) \frac{p(y_{n-1:n}|x_{n-1:n})}{p(y_{n-1}|x_{n-1})} \\ &= p(x_n|x_{n-1}) \frac{p(y_n|x_{n-1:n})p(y_{n-1}|x_{n-1:n})}{p(y_{n-1}|x_{n-1})} \text{ d'après (2.2.23)} \\ &= \underbrace{p(x_n|x_{n-1}) \frac{p(y_{n-1}|x_{n-1:n})}{p(y_{n-1}|x_{n-1})}}_{p(x_n|z_{n-1})} p(y_n|x_{n-1:n}) \end{aligned}$$

**Chaînes de Markov couples** Enfin, dans le modèle le plus général que nous appelons chaînes de Markov couples, bien que les autres modèles présentés soient également des CM-Couples,  $p(y_n|x_{n-1}, x_n)$  dépend de  $x_n$  et de  $x_{n-1}$  et les observations sont corrélées. Dans ce cas,  $X$  n'est plus une CM d'ordre 1 et

$$p(y_{n-1:n}|x_{n-1:n}) = p(y_n|x_n, z_{n-1})p(y_{n-1}|x_{n-1:n}) \quad (2.2.25)$$

On en déduit que

$$p(z_n|z_{n-1}) = p(x_n|z_{n-1})p(y_n|x_n, z_{n-1}) \quad (2.2.26)$$

On peut donc prendre le modèle le plus général, en le réduisant, s'il y a lieu, à un modèle plus simple. Par exemple, considérer que  $p(y_n|x_{n-1}, x_n)$  dépend de  $x_n$  et également de  $x_{n-1}$  peut être utile dans une application de segmentation d'image pour modéliser le fait que la loi d'observation d'une classe donnée peut être différente près d'une frontière entre classes différentes et à l'intérieur d'un grand ensemble de sites d'une même classe, ce qui n'est pas possible en considérant le modèle CMC-BI. D'autre part la prise en compte de la corrélation entre observations permet dans certaines applications d'accroître considérablement la qualité de la classification. C'est notamment le cas en reconnaissance de parole [118, 87] ou encore en traitement d'image dans le cas de classes texturées [19, 42, 18, 20].



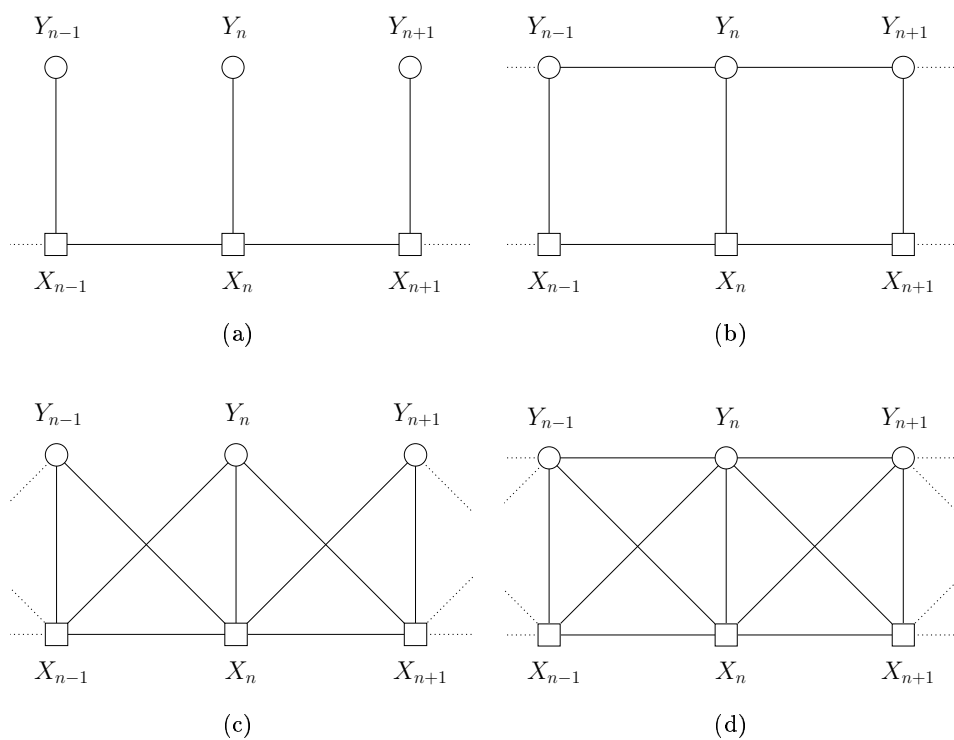


FIG. 2.2.1 – Graphes d'indépendance conditionnelle des différents modèles : (a) CMC-BI, (b) CMC, (c) CMCouple-BI, (d) CMCouple. Dans ce type de graphe, l'absence d'une arête entre deux nœuds indique que les deux variables concernées sont conditionnellement indépendantes par rapport aux autres variables.

### 2.2.3 Chaînes couples partiellement de Markov

Un autre extension envisageable du modèle classique, appelée *chaînes couples partiellement de Markov* (CCPM) et proposée dans [98], consiste à faire l'hypothèse d'indépendance conditionnelle suivante :

$$Z_{n+1:N} \perp\!\!\!\perp X_{1:n-1} | (Z_n, Y_{1:n-1}) \quad (2.2.27)$$

Ainsi, ni  $X$ , ni  $Z$  n'est supposé être une CM. Néanmoins, la loi *a posteriori* reste celle d'une CM<sup>3</sup>. En effet, d'après (2.2.27), les probabilités de transitions de (2.1.1) se simplifient et sont données par

$$p(z_n | z_{1:n-1}) = p(z_n | z_{n-1}, y_{1:n-2}) = p(x_n | z_{n-1}, y_{1:n-2}) p(y_n | x_n, z_{n-1}, y_{1:n-2}) \quad (2.2.28)$$

et d'après (2.1.5), nous avons

$$\begin{aligned} p(x_n | x_{1:n-1}, y) &= p(z_n | z_{1:n-1}) \frac{\sum_{x_{n+1:N}} p(z_{n+1:N} | z_{1:n})}{\sum_{x_{n:N}} p(z_{n:N} | z_{1:n-1})} \\ &= p(z_n | z_{n-1}, y_{1:n-2}) \frac{\sum_{x_{n+1:N}} p(z_{n+1:N} | z_n, y_{1:n-1})}{\sum_{x_{n:N}} p(z_{n:N} | z_{n-1}, y_{1:n-2})} \text{ d'après (2.2.28)} \\ &= p(x_n | x_{n-1}, y) \end{aligned}$$

Les probabilités de transition *a posteriori* sont donc données en chaque  $n \in \mathcal{N} \setminus \{1\}$  par :

$$p(x_n | x_{n-1}, y) = p(z_n | z_{n-1}, y_{1:n-2}) \frac{\beta_n(z_n, y_{1:n-1})}{\beta_{n-1}(z_{n-1}, y_{1:n-2})} \quad (2.2.29)$$

avec  $\beta_{n-1}(z_{n-1}, y_{1:n-2}) = p(y_{n:N} | z_{n-1}, y_{1:n-2})$  calculable récursivement pour tout  $n \in \mathcal{N} \setminus \{N\}$  de la manière suivante :

$$\beta_{n-1}(z_{n-1}, y_{1:n-2}) \propto \begin{cases} \sum_{x_n \in \mathcal{X}} p(z_n | z_{n-1}, y_{1:n-2}) & \text{si } n = N \\ \sum_{x_n \in \mathcal{X}} p(z_n | z_{n-1}, y_{1:n-2}) \beta_n(z_n, y_{1:n-1}) & \text{sinon} \end{cases} \quad (2.2.30)$$

Ainsi, les probabilités de transitions *a posteriori* sont calculables pour tout  $n \in \mathcal{N}$  si les transitions  $p(z_n | z_{n-1}, y_{n-2})$  le sont. Ces dernières peuvent être décomposées de la manière suivante :

$$\begin{aligned} p(z_n | z_{n-1}, y_{1:n-2}) &= p(x_n | x_{n-1}) \frac{p(y_{1:n} | x_{n-1:n})}{p(y_{1:n-1} | x_{n-1})} \\ &= \frac{p(x_{n-1}, x_n) p(y_{1:n} | x_{n-1:n})}{\sum_{x_n \in \mathcal{X}} p(x_{n-1}, x_n) p(y_{1:n-1} | x_{n-1:n})} \end{aligned} \quad (2.2.31)$$

Il est donc nécessaire de connaître les densités  $p(x_{n-1:n})$  et  $p(y_{n-1:n} | x_{1:n})$  afin de calculer les probabilités de transition. Dans certains cas particuliers, les  $p(y_{1:n} | x_{n-1:n})$  peuvent

---

<sup>3</sup>Mais contrairement aux CMCouples, le processus  $Y$  conditionnel à  $X$  n'est pas une CM.

être calculées. C'est par exemple le cas lorsque  $y$  est un processus gaussien. Le calcul de  $p(y_{1:n}|x_{n-1:n})$  peut alors être effectué lors d'une passe avant en utilisant à la décomposition suivante.

$$p(y_{1:n}|x_{n-1:n}) = p(y_1|x_{n-1:n}) \prod_{i=2}^n p(y_i|y_{1:i-1}, x_{n-1:n}) \quad (2.2.32)$$

Ce modèle, sur lequel nous reviendrons à la fin de ce chapitre, permet ainsi la prise en compte de bruits centrés à corrélation longue. Par ailleurs, on peut généraliser les marginales gaussiennes aux marginales quelconques en utilisant les copules comme cela est proposé dans [99].

Dans cette section, nous avons présenté différentes modélisations de la loi de  $Z$  telles que les algorithmes d'inférence proposés dans la sous-section 2.1 soient applicables. Il suffit pour cela que la loi du processus caché *a posteriori* soit une CM. Différents jeux d'hypothèses d'indépendance conditionnelle concernant la loi de  $Z$  permettent d'obtenir la markovianité du processus *a posteriori*. Ainsi, en relâchant certaines hypothèses du modèle classique CMC-BI nous avons présenté des extensions de ce modèle dont le plus général est le modèle CCPM. Les CCPM permettent ainsi la prise en compte de lois de  $Y$  conditionnellement à  $X$  plus complexes pouvant mieux correspondre aux données traitées comme par exemple des bruits à corrélation longue. D'autre part le processus caché  $X$  n'est plus nécessairement une CM d'où une plus grande généralité. Dans la section suivante, nous traitons l'estimation de paramètres pour le modèle de chaînes de Markov couples.

## 2.3 Estimation des paramètres des CMCouples

Nous avons vu au chapitre précédent différentes méthodes d'estimation des paramètres bien adaptées aux modèles à variables cachées. Nous nous proposons ici d'explicitier ces méthodes dans le cas des CMCouples, l'estimation des CCPM n'ayant été étudiée que dans un cas particulier que nous détaillerons par la suite. Nous supposons que le processus  $Z$  est stationnaire. Sa loi, donnée par  $p(z_1, z_2)$ , est décomposable comme le produit de deux densités :  $p(z_{1:2}) = p(y_{1:2}|x_{1:2})p(x_{1:2})$ . Il est alors nécessaire d'estimer d'une part la densité  $p(x_{1:2})$  et d'autre part d'estimer la densité  $p(y_{1:2}|x_{1:2})$ . Dans le cas considéré par la suite,  $\mathcal{X}$  sera un ensemble discret de classes et  $\mathcal{Y}$  sera l'ensemble des réels  $\mathbb{R}$ . Nous commencerons par présenter l'algorithme EM dans le cas couple, qui est un de nos apports, puis l'algorithme MCEM et nous finirons par la description de l'algorithme ECI.

### 2.3.1 Estimation EM

L'algorithme EM a été décrit dans la sous-section 1.2.1 du chapitre 1. Il consiste, rappelons le, à construire, de manière itérative, une suite de paramètres  $(\phi^{[q]})_{q>0}$  telle que  $(\mathcal{L}(y; \phi^{[q]}))_{q \geq 0}$  converge vers un maximum local de la log-vraisemblance  $\mathcal{L}(y; \phi)$  en alternant entre une étape  $E$  de calcul de la fonction  $\mathcal{Q}(\phi|\phi^{[q]}) = \mathbb{E}_{\phi^{[q]}} \{\mathcal{L}_c(Z; \phi)|Y = y\}$  et une étape

$M$  de maximisation de  $\mathcal{Q}(\phi|\phi^{[q]})$  par rapport à  $\phi$ . Pour cela, nous décomposons la loi de  $Z$ , dans le cas couple, de la manière suivante

$$p(z|\phi) = \prod_{n=2}^N p(x_{n-1:n}) \times \frac{1}{\prod_{n=2}^{N-1} p(z_n)} \times \prod_{n=2}^N p(y_{n-1:n}|x_{n-1:n}) \quad (2.3.1)$$

La vraisemblance complétée est alors donnée par

$$\mathcal{L}_c(\phi) = \sum_{n=2}^N \log p(x_{n-1:n}) - \sum_{n=2}^{N-1} \log p(z_n) + \sum_{n=2}^N \log p(y_{n-1:n}|x_{n-1:n}) \quad (2.3.2)$$

**Etape E** En ce qui concerne l'étape E de l'algorithme EM, nous devons donc calculer la fonction  $\mathcal{Q}(\phi|\phi^{[q]})$ . Pour simplifier les notations nous posons

$$\left\{ \begin{array}{l} b_j(y_n) = p(x_n = j, y_n) \\ c_{ij} = p(x_{n-1} = i, x_n = j) \\ f_{i,j}(y_{n-1}, y_n) = p(y_{n-1:n}|x_{n-1} = i, x_n = j) \\ \gamma_n^{[q]}(j) = p(x_n = j|y; \phi^{[q]}) \\ \psi_{n-1}^{[q]}(i, j) = p(x_{n-1} = i, x_n = j|y; \phi^{[q]}) \end{array} \right. \quad (2.3.3)$$

nous obtenons

$$\mathcal{Q}(\phi|\phi^{[q]}) = \sum_{n=2}^N \sum_{(i,j) \in \mathcal{X}^2} \psi_{n-1}^{[q]}(i, j) \log c_{ij} - \sum_{n=2}^{N-1} \sum_{j \in \mathcal{X}} \gamma_n^{[q]}(j) \log b_j(y_n) + \sum_{n=2}^N \sum_{(i,j) \in \mathcal{X}^2} \psi_{n-1}^{[q]}(i, j) \log f_{i,j}(y_{n-1}, y_n) \quad (2.3.4)$$

Les probabilités *a posteriori*  $\gamma_n^{[q]}(j)$  et  $\psi_{n-1}^{[q]}(i, j)$  sont calculées, à chaque itération, en utilisant l'algorithme 3 de calcul des marginales *a posteriori*.

**Etape M** L'étape M de l'algorithme EM nécessite de trouver les paramètres maximisant la fonction  $\mathcal{Q}(\phi|\phi^{[q]})$ . Pour maximiser la fonction  $\mathcal{Q}(\phi|\phi^{[q]})$ , nous utilisons la méthode des multiplicateurs de Lagrange sous les contraintes suivantes

$$\sum_{(i,j) \in \mathcal{X}^2} c_{ij} = 1, \quad \sum_{(i,l) \in \mathcal{X} \times \mathcal{Y}} c_{ij} f_{i,j}(l, m) = b_j(m) \quad \forall (j, m) \in \mathcal{X} \times \mathcal{Y}, \quad \sum_{(l,m) \in \mathcal{Y}^2} f_{i,j}(l, m) = 1 \quad (2.3.5)$$

Nous devons alors maximiser le lagrangien suivant :

$$\begin{aligned} L = & \mathcal{Q}(\phi|\phi^{[q]}) - \lambda_1 \left( \sum_{(i,j) \in \mathcal{X}^2} c_{ij} - 1 \right) \\ & - \sum_{(j,m) \in \mathcal{X} \times \mathcal{Y}} \lambda_{j,m} \left( \sum_{i,l \in \mathcal{X} \times \mathcal{Y}} c_{ij} f_{i,j}(l, m) - b_j(m) \right) - \lambda_3 \left( \int_{(l,m) \in \mathcal{Y}^2} f_{i,j}(l, m) - 1 \right) \end{aligned} \quad (2.3.6)$$

La dérivation de  $L$  dans (2.3.6) par rapport à  $c_{ij}$  et l'annulation du résultat conduit à :

$$c_{ij} = \frac{\sum_{n=2}^N \psi_{n-1}^{[q]}(i, j)}{\lambda_1 + \sum_{l,m \in \mathcal{Y}^2} \lambda_{j,m} f_{i,j}(l, m)} \quad (2.3.7)$$

Par ailleurs, d'après la première contrainte,

$$\sum_{(i,j) \in \mathcal{X}^2} c_{ij} = 1 = \sum_{(i,j) \in \mathcal{X}^2} \frac{\sum_{n=2}^N \psi_{n-1}^{[q]}(i, j)}{\lambda_1 + \sum_{l,m \in \mathcal{Y}^2} \lambda_{j,m} f_{i,j}(l, m)} = \frac{N-1}{\lambda_1 + \sum_{l,m \in \mathcal{Y}^2} \lambda_{j,m} f_{i,j}(l, m)}$$

Nous en déduisons  $c_{ij}^{[q+1]}$ . D'autre part, la dérivation de  $L$  par rapport à  $f_{i,j}(l, m)$  et l'annulation du résultat conduit à :

$$f_{i,j}(l, m) = \frac{\sum_{n=2: (y_{n-1}, y_n) = (l, m)}^N \psi_{n-1}^{[q]}(i, j)}{\lambda_3} \quad (2.3.8)$$

Par ailleurs,

$$\sum_{(l,m) \in \mathcal{Y}^2} f_{i,j}(l, m) = 1 = \sum_{(l,m) \in \mathcal{Y}^2} \frac{\sum_{n=2: (y_{n-1}, y_n) = (l, m)}^N \psi_{n-1}^{[q]}(i, j)}{\lambda_3} = \frac{\sum_{n=2}^N \psi_{n-1}^{[q]}(i, j)}{\lambda_3}$$

Nous en déduisons la formule de réestimation de  $f_{i,j}^{[q+1]}(l, m)$ . Finalement, nous obtenons les formules de réestimation suivantes

$$c_{ij}^{[q+1]} = \frac{1}{N-1} \sum_{n=2}^N \psi_{n-1}^{[q]}(i, j), \quad f_{i,j}^{[q+1]}(l, m) = \frac{\sum_{n=2: (y_{n-1}, y_n) = (l, m)}^N \psi_{n-1}^{[q]}(i, j)}{\sum_{n=2}^N \psi_{n-1}^{[q]}(i, j)} \quad (2.3.9)$$

Dans le cas d'observation réelles avec densité conditionnelles gaussiennes de la forme suivante

$$\forall (i, j) \in \mathcal{X}^2, \quad p(y_{n-1:n} | x_{n-1} = i, x_{n-1} = j) = \mathcal{N}(y_{n-1:n}; m_{ij}, \Gamma_{ij}) \quad (2.3.10)$$

nous obtenons les formules de réestimation suivantes :

$$\begin{cases} c_{ij}^{[q+1]} = \frac{1}{N-1} \sum_{n=2}^N \psi_{n-1}^{[q]}(i, j) \\ m_{ij}^{[q+1]} = \frac{\sum_{n=2}^N y_{n-1:n} \psi_{n-1}^{[q]}(i, j)}{\sum_{n=2}^N \psi_{n-1}^{[q]}(i, j)} \\ \Gamma_{ij}^{[q+1]} = \frac{\sum_{n=2}^N (y_{n-1:n} - m_{ij}^{[q+1]}) (y_{n-1:n} - m_{ij}^{[q+1]})^t \psi_{n-1}^{[q]}(i, j)}{\sum_{n=2}^N \psi_{n-1}^{[q]}(i, j)} \end{cases} \quad (2.3.11)$$

On peut remarquer que l'on retrouve les formules classiques dans le cas des CMC-BI (voir [9]).

### 2.3.2 Estimation MCEM

L'algorithme MCEM, décrit également dans la sous-section 1.2.1, consiste à remplacer le calcul de  $\mathcal{Q}(\phi|\phi^{[q]})$  par sa moyenne empirique  $\mathcal{Q}^{[q+1]}(\phi|\phi^{[q]})$  fondée sur  $\tau$  tirages de  $x$  *a posteriori* pour une valeur des paramètres  $\phi^{[q]}$ .

**Etape E** Nous approchons le calcul de  $\mathcal{Q}$  en simulant des réalisations de  $X$  *a posteriori*

$$\mathcal{Q}(\phi|\phi^{[q]}) = \mathbb{E} \left\{ \log p(X, Y|\phi) | y, \phi^{[q]} \right\} \approx \frac{1}{\tau} \sum_{t=1}^{\tau} \log p(x^{[q]}(t), y|\phi) = \mathcal{Q}^{[q+1]}(\phi|\phi^{[q]}) \quad (2.3.12)$$

et nous obtenons

$$\mathcal{Q}^{[q+1]}(\phi|\phi^{[q]}) = \frac{1}{\tau} \sum_{t=1}^{\tau} \left[ \sum_{n=2}^N \log p(x_{n-1:n}^{[q]}|\phi) - \sum_{n=2}^{N-1} \log p(x_n^{[q]}, y_n|\phi) + \sum_{n=2}^N \log p(y_{n-1:n}|x_{n-1:n}^{[q]}, \phi) \right] \quad (2.3.13)$$

**Etape M** Pour maximiser l'espérance, nous procédons de la même manière que pour EM en utilisant la méthode des multiplicateurs de Lagrange sous les mêmes contraintes. Nous obtenons les formules de réestimation dans le cas gaussien en remplaçant  $\psi_{n-1}^{[q]}(i, j)$  par  $\mathbb{1}_{[x_{n-1:n}^{[q]}(t)=(i,j)]}$  dans (2.3.11).

### 2.3.3 Estimation ECI

L'algorithme ECI, décrit à la sous-section 1.2.2, consiste à prendre un estimateur  $\hat{\phi} = \hat{\phi}(z)$  défini à partir des données complètes  $z$  et de l'approcher par l'espérance conditionnelle de  $\hat{\phi} = \hat{\phi}(z)$  par rapport à  $y$ . Celle-ci est calculée à partir de la procédure itérative suivante :  $\phi^{[q+1]} = \mathbb{E}_{\phi^{[q]}} \left\{ \hat{\phi}(Z) | Y = y \right\}$ . Nous utilisons les estimateurs empiriques suivants :

$$\hat{c}_{ij} = \frac{1}{N-1} \sum_{n=2}^N \mathbb{1}_{[(x_{n-1:n})=(i,j)]} \quad , \quad \hat{f}_{i,j}(l, m) = \frac{\sum_{n=2}^N \mathbb{1}_{[y_{n-1:n}=(l,m)]} \mathbb{1}_{[x_{n-1:n}=(i,j)]}}{\sum_{n=2}^N \mathbb{1}_{[x_{n-1:n}=(i,j)]}} \quad (2.3.14)$$

Nous utilisons dans un premier temps l'estimateur empirique  $\hat{c}_{ij}$  dont on calcule l'espérance par rapport à  $Y = y$  :

$$c_{ij}^{[q+1]} = \mathbb{E} \{ \hat{c}_{ij} | y \} = \frac{1}{N-1} \sum_{n=2}^N \psi_{n-1}^{[q]}(i, j) \quad (2.3.15)$$

Nous retrouvons la même formule de réestimation que celle obtenue par la méthode EM. Néanmoins, contrairement à EM, la méthode ECI n'utilise par la notion de vraisemblance et le calcul reste valable quelle que soit la forme du bruit. Le calcul de l'espérance de  $\hat{f}_{i,j}(l, m)$  ne peut par contre pas être fait directement à cause du numérateur, nous utilisons alors

l'approximation stochastique.

$$f_{i,j}^{[q+1]}(l, m) = \mathbb{E} \left\{ \hat{f}_{i,j}(l, m) | y \right\} \approx \frac{\sum_{t=1}^{\tau} \sum_{n=2}^N y_{n-1:n} = (l, m) \mathbb{1}_{[x_{n-1:n}^{[q]}(t) = (i, j)]}}{\sum_{t=1}^{\tau} \sum_{n=2}^N \mathbb{1}_{[x_{n-1:n}^{[q]}(t) = (i, j)]}} \quad (2.3.16)$$

Finalement, dans le cas d'observations réelles avec densités conditionnelles gaussiennes de la forme  $p(y_{n-1:n} | x_{n-1} = i, x_n = j) = \mathcal{N}(y_{n-1:n}; m_{ij}, \Gamma_{ij})$ ,  $(i, j) \in \mathcal{X}^2$ , nous obtenons les formules de réestimation suivantes (voir [42]) :

$$\begin{cases} c_{ij}^{[q+1]} = \frac{1}{N-1} \sum_{n=2}^N \psi_{n-1}^{[q]}(i, j) \\ m_{ij}^{[q+1]} = \frac{\sum_{t=1}^{\tau} \sum_{n=2}^N y_{n-1:n} \mathbb{1}_{[x_{n-1:n}^{[q]}(t) = (i, j)]}}{\sum_{t=1}^{\tau} \sum_{n=2}^N \mathbb{1}_{[x_{n-1:n}^{[q]}(t) = (i, j)]}} \\ \Gamma_{ij}^{[q+1]} = \frac{\sum_{t=1}^{\tau} \sum_{n=2}^N (y_{n-1:n} - m_{ij}^{[q+1]}) (y_{n-1:n} - m_{ij}^{[q+1]})^t \mathbb{1}_{[x_{n-1:n}^{[q]}(t) = (i, j)]}}{\sum_{t=1}^{\tau} \sum_{n=2}^N \mathbb{1}_{[x_{n-1:n}^{[q]}(t) = (i, j)]}} \end{cases} \quad (2.3.17)$$

Ainsi, dans le cas où  $\psi_{n-1}^{[q]}(i, j)$  est approché par Monte-Carlo, on obtient des formules de réestimations identiques à celles de MCEM. D'autre part, on retrouve également les formules classiques dans le cas des CMC-BI (voir [9]).

## 2.4 Application à la segmentation de processus gaussiens centrés avec bruit à corrélation longue

Une des applications possibles des CCPM proposée dans [99], et que nous avons étudiée dans [101], est de considérer des bruits à "corrélation longue" pour lesquels les corrélations décroissent "lentement", i.e tel que :  $\lim_{\tau \rightarrow \infty} \tau^\alpha \rho(\tau) = c_\rho$ , où  $\alpha \in ]0, 1[$ ,  $c_\rho > 0$  est une constante, et  $\rho(\tau)$  est une fonction d'autocorrélation. Ce type de bruit intervient dans de nombreuses situations et ne peut pas être pris en compte par les modèles de Markov cachés classiques [11, 22, 30].

### 2.4.1 Calcul des probabilités de transition dans les CCPM

Dans le cas des chaînes couples partiellement de Markov (CCPM), les probabilités de transition sont données, rappelons-le, par (2.2.31) :

$$p(z_n | z_{n-1}, y_{1:n-2}) = \frac{p(x_{n-1}, x_n) p(y_{1:n} | x_{n-1:n})}{\sum_{x_n \in \mathcal{X}} p(x_{n-1}, x_n) p(y_{1:n-1} | x_{n-1:n})} \quad (2.4.1)$$

avec

$$p(y_{1:n} | x_{n-1:n}) = p(y_1 | x_{n-1:n}) \prod_{i=2}^n p(y_i | y_{1:i-1}, x_{n-1:n}) \quad (2.4.2)$$

Le problème principal, pour pouvoir calculer les probabilités *a posteriori* nécessaires à l'estimation d'une réalisation du processus caché, est de pouvoir calculer ces probabilités de transitions. Les probabilités  $p(x_{n-1}, x_n)$  ne posent pas de problème particulier, par contre, les probabilités  $p(y_{1:n}|x_{n-1:n})$  semblent difficilement calculables au premier abord. Néanmoins, elles le sont pour certaines CCPM dites CCPM gaussiennes [99], dans lesquelles  $p(y|x)$  est gaussien. Pour cela, nous allons utiliser la propriété classique suivante

*Propriété 2.4.1.* Soit  $B = (B_1, \dots, B_N)$  un processus réel gaussien centré avec pour tout  $1 \leq n \leq N$ ,  $\Sigma_n = (\gamma_{kl})_{1 \leq k, l \leq n}$  matrice de covariance de  $B_{1:n} = (B_1, \dots, B_n)$  que nous pouvons décomposer de la façon suivante :

$$\Sigma_n = \begin{bmatrix} \Sigma_{n-1} & \Sigma_{1:n-1}^T \\ \Sigma_{1:n-1} & \gamma_{nn} \end{bmatrix} \quad (2.4.3)$$

Alors il est possible de calculer, pour tout  $n$ , la densité gaussienne  $p(y_{1:n})$  de matrice de covariance  $\Sigma_n$ . Classiquement, on procède de proche en proche en utilisant l'identité  $p(y_n) = p(y_{1:n-1})p(y_n|y_{1:n-1})$  et le fait que  $p(y_n|y_{1:n-1})$  est gaussienne de moyenne  $E(y_n|y_{1:n-1})$  et de variance  $V(y_n|y_{1:n-1})$  données par

$$\begin{cases} E(y_n|y_{1:n-1}) = \Sigma_{1:n-1} \Sigma_{n-1}^{-1} y_{1:n-1} \\ V(y_n|y_{1:n-1}) = \gamma_{nn} - \Sigma_{1:n-1} \Sigma_{n-1}^{-1} \Sigma_{1:n-1}^T \end{cases} \quad (2.4.4)$$

avec  $\Sigma_{1:n-1}^T = (\gamma_{1,n}, \dots, \gamma_{n-1,n})^T$  (la matrice  $\Sigma_{n-1}$  figurant dans  $p(y_n|y_{1:n-1})$  est donnée par la densité gaussienne  $p(y_{1:n-1})$ ).

Ainsi, les  $p(y_n|y_{1:n-1}, x_{n-1:n})$  sont calculables pour tout  $n \in \mathcal{N}$  en utilisant la propriété (2.4.1)  $|\mathcal{X} \times \mathcal{X}|$  fois. Toutefois, les calculs de  $E(y_n|y_{1:n-1})$  et de  $V(y_n|y_{1:n-1})$  nécessitent le stockage en mémoire et l'inversion de la matrice  $\Sigma_{n-1}$  ce qui devient rapidement ingérable lorsque  $n$  croît. Pour contourner le problème, nous pouvons utiliser le fait que la matrice  $\Sigma$  est Toeplitz symétrique (voir [75]). Pour le montrer, nous ne considérons à présent que le cas centré avec variance égales à 1.  $\Sigma$  dénotera donc à présent la matrice de corrélation d'une série temporelle gaussienne avec les autocorrélations  $r = (r_1, r_2, \dots)^T$  et  $r_0 = 1$ .  $\Sigma$  étant Toeplitz symétrique, l'espérance et la variance peuvent être exprimées en fonction des solutions des équations de Yule-Walker données par  $\Sigma_n \lambda = -r$ . Pour cela, nous définissons la matrice  $W$  comme la matrice carrée contenant des uns sur la seconde diagonale et des zéros partout ailleurs. Ainsi,  $W = W^T = W^{-1}$  et  $W \Sigma^{-1} = \Sigma^{-1} W$  pour toute matrice Toeplitz symétrique  $\Sigma$  étant donné que  $\Sigma^{-1}$  est symétrique et persymétrique. Les deux premiers moments de  $y_n$  peuvent alors être exprimés via la solution  $\lambda$  des équations de Yule-Walker  $\lambda = -\Sigma_n^{-1} r$ . Ainsi, nous avons le résultat classique suivant<sup>4</sup> qui permet le calcul récursif des moyennes et des variances conditionnelles :

---

<sup>4</sup>En effet,  $\begin{cases} E(y_{n+1}|y_{1:n}) = (Wr)^T \Sigma_n^{-1} y_{1:n} = r^T \Sigma_n^{-1} W y_{1:n} = -(W\lambda)^T y_{1:n} \\ V(y_{n+1}|y_{1:n}) = 1 - (Wr)^T \Sigma_n^{-1} W r = 1 - r^T \Sigma_n^{-1} r = 1 + r^T \lambda \end{cases}$



$$\begin{cases} E(y_n|y_{1:n-1}) = -(W\lambda)^T y_{1:n} \\ V(y_n|y_{1:n-1}) = 1 + r^T \lambda \end{cases} \quad (2.4.5)$$

Les solutions  $\lambda$  des équations de Yule-Walker peuvent alors être trouvées en utilisant l'algorithme de Durbin-Levinson [55]. Nous utiliserons cette méthode pour le calcul des probabilités  $p(y_n|y_{n-1}, x_{n-1:n})$  et pour la simulation de CCPM.

Nous présentons dans les sous-sections suivantes deux expérimentations basées sur une CCPM gaussienne particulière appelée *chaîne de Markov cachée à bruit à corrélation longue* CMC-BCL pour laquelle

$$p(z_n|z_{1:n-1}) = p(x_n|x_{n-1})p(y_n|y_{1:n-1}, x_n) \quad (2.4.6)$$

La première concerne la segmentation supervisée d'un processus à corrélation longue. Ce sera l'occasion de présenter les algorithmes de simulation de CCPM et de calcul des densités des bruits fondés sur l'algorithme de Durbin-Levinson. Enfin, nous ferons une étude comparative du modèle CMC-BI avec le modèle CMC-BCL.

## 2.4.2 Segmentation supervisée d'un processus à corrélation longue

Dans la première expérimentation, nous considérons le cas de la segmentation supervisée d'une image synthétique bruitée par un bruit à corrélation longue. L'image cachée possède 4 classes. Celle-ci est convertie en un processus monodimensionnel par l'utilisation d'un parcours de Hilbert-Peano représenté sur la figure 2.4.1 ; la chaîne obtenue est alors considérée comme une réalisation d'une CM  $X$ .

Les quatre lois gaussiennes de  $B = \{B_n\}_{n=1}^N$  seront supposées stationnaires, avec toutes les moyennes nulles et toutes les variances égales à 1. Ainsi les corrélations sont les seuls paramètres discriminants. Toutes les 4 autocorrélations possèdent la forme  $r(\tau) = |\tau + 1|^{-\alpha}$  où  $\tau = |j - i|$ ,  $\alpha_{\omega_1} = 0.99$ ,  $\alpha_{\omega_2} = 0.3$ ,  $\alpha_{\omega_3} = 0.05$ ,  $\alpha_{\omega_4} = 0.01$ . Les paramètres du modèle sont ainsi la loi  $p(x_1, x_2)$  sur  $\mathcal{X}^2 = \{\omega_1, \omega_2, \omega_3, \omega_4\}^2$ , et les paramètres  $\alpha_i > 0$ . La méthode de simulation de  $y$  pour une réalisation  $x$  donnée est résumée dans l'algorithme 5, le point important étant que pour chaque  $n \in \mathcal{N} \setminus \{N\}$ , on calcule l'ensemble des solutions  $\lambda(k)$  pour tout  $k$  qui seront utiles pour la simulation à l'itération suivante. Ainsi, la versions bruitée de 2.4.2(a) par le bruit à corrélation longue est présentée en 2.4.2(b).

Une réalisation du processus caché est alors estimé au sens du MPM en utilisant l'algorithme 3 du chapitre 1 avec (2.4.6). Les probabilités  $p(y_n|y_{1:n-1}, x_n)$  sont calculées en utilisant l'algorithme 6. Le résultat de la segmentation MPM supervisée est présentée en 2.4.2(c). Nous pouvons remarquer que bien que le bruitage soit relativement important, les corrélations étant les seuls paramètres discriminants, l'algorithme d'estimation supervisée donne un bon résultat avec un taux de sites mal classés égal à 6,9%.

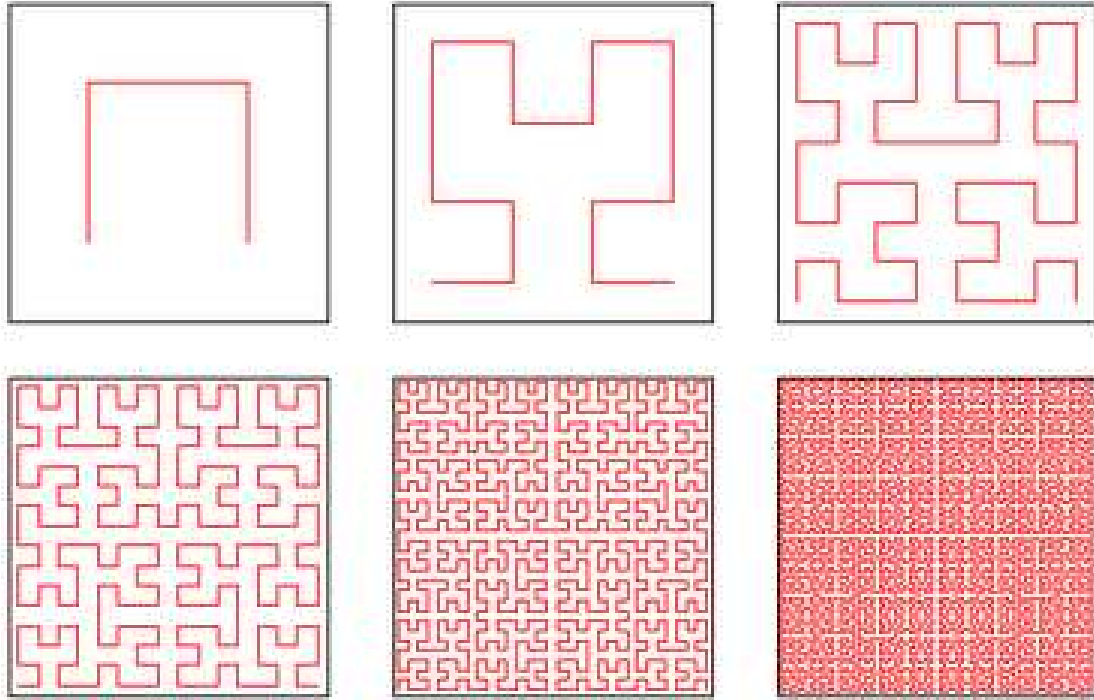


FIG. 2.4.1 – Parcours de Hilbert-Peano

---

**Algorithme 5** Simulation de  $y$  dans le cas d'une CMC-BCL gaussienne à partir d'une réalisation  $x$  donnée.  $\varepsilon_{1:N}$  est un vecteur i.i.d de v.a gaussiennes centrées de variance 1. Nous rappelons que  $1 : n + 1$  signifie  $1, \dots, n + 1$  et  $n + 1 : -1 : 1$  signifie  $n + 1, n, \dots, 1$ .

---

**pour**  $k \in \mathcal{X}$

*Initialisation* :  $\lambda_1(k) = -r_1(k); \beta(k) = 1; \alpha(k) = -r_1(k);$

**finpour**  $k$

$y_1 = \varepsilon_1; y_2 = -y_1 * \lambda_1 + \varepsilon_2 * \sqrt{\gamma_{11}(1 + r_1(x_2) * \lambda_1(x_2))};$

**pour**  $n = 1 : N - 2$

**pour**  $k \in \mathcal{X}$

*Durbin-Levinson* : calcul des solutions  $\lambda(k)$  des équations de Yule Walker

$\beta(k) = (1 - \alpha(k) * \alpha(k)) * \beta(k);$

$\alpha(k) = -(r_{n+1}(k) + r'_{n:-1:1}(k)\lambda_{1:n})/\beta(k);$

**pour**  $i = 1 : (n + 1)/2$

$h_1 = \lambda_i(k); h_2 = \lambda_{n+1-i}(k);$

$\lambda_i(k) = h_1 + \alpha(k) * h_2;$

$\lambda_{n+1-i}(k) = h_2 + \alpha(k) * h_1;$

**finpour**  $i$

$\lambda_{n+1}(k) = \alpha(k);$

**finpour**  $k$

$y_{n+2} = -y'_{1:n+1}\lambda_{n+1:-1:1}(x_{n+1}) + \varepsilon_{n+2}\sqrt{\gamma_{n+2,n+2}(1 + r'_{1:n+1}(x_{n+1})\lambda_{1:n+1}(x_{n+1}))};$

**finpour**  $n$

---

---

**Algorithme 6** Calcul des probabilités  $p(y_n|y_{1:n-1}, x_n)$  pour une CMC-BCL gaussienne.

---

**pour**  $k \in \mathcal{X}$

*Initialisation* :  $\lambda_1(k) = -r_1(k)$ ;  $\beta(k) = 1$ ;  $\alpha(k) = -r_1(k)$ ;  $\begin{cases} \mu_1(k) = 0 \\ \sigma_1(k) = 1 \end{cases}$  ;

$p(y_1|x_1 = k) = \frac{1}{\sigma_1(k)\sqrt{2\pi}} \exp\left(-\frac{(x-\mu_1(k))^2}{2\sigma_1^2(k)}\right)$  avec  $\begin{cases} \mu_1(k) = 0 \\ \sigma_1(k) = \sqrt{\gamma_{11}} \end{cases}$  ;

$p(y_2|x_2 = k, y_1) = \frac{1}{\sigma_2(k)\sqrt{2\pi}} \exp\left(-\frac{(x-\mu_2(k))^2}{2\sigma_2^2(k)}\right)$  avec  $\begin{cases} \mu_2(k) = -y_1 * \lambda_1(k) \\ \sigma_2(k) = \sqrt{\gamma_{22}(1 + r_1(k) * \lambda_1(k))} \end{cases}$  ;

**finpour**  $k$

**pour**  $n = 1 : N - 2$

**pour**  $k \in \mathcal{X}$

*Durbin-Levinson* : calcul des solution  $\lambda(k)$  des équations de Yule Walker

$\beta(k) = (1 - \alpha(k) * \alpha(k)) * \beta(k)$  ;

$\alpha(k) = -(r_{n+1}(k) + r'_{n:-1:1}(k)\lambda_{1:n}(k))/\beta(k)$  ;

**pour**  $i = 1 : (n + 1)/2$

$h_1 = \lambda_i(k)$ ;  $h_2 = \lambda_{n+1-i}(k)$  ;

$\lambda_i(k) = h_1 + \alpha(k) * h_2$  ;

$\lambda_{n+1-i}(k) = h_2 + \alpha(k) * h_1$  ;

**finpour**  $i$

$\lambda_{n+1}(k) = \alpha(k)$  ;

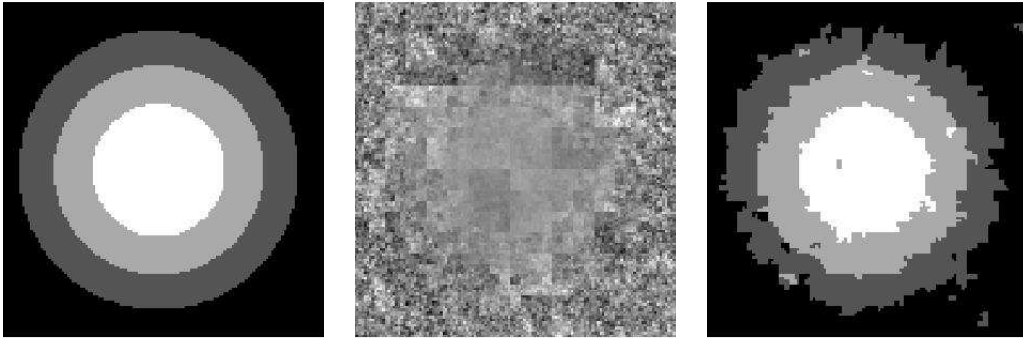
$p(y_{n+2}|x_{n+2} = k, y_{1:n+1}) = \frac{1}{\sigma_{n+2}(k)\sqrt{2\pi}} \exp\left(-\frac{(x-\mu_{n+2}(k))^2}{2\sigma_{n+2}^2(k)}\right)$

avec  $\begin{cases} \mu_{n+2}(k) = -y'_{1:n+1}\lambda_{n+1:-1:1}(k) \\ \sigma_{n+2}(k) = \sqrt{\gamma_{n+2,n+2}(1 + r'_{1:n+1}(k)\lambda_{1:n+1}(k))} \end{cases}$  ;

**finpour**  $k$

**finpour**  $n$

---



(a)

(b)

(c)

FIG. 2.4.2 – (a) l'image des classes, (b) sa version bruitée (moyennes et variances égales pour chacune des 4 classes), et (c) le résultat de la segmentation MPM obtenue à partir du modèle CMC-BCL. (a) et (b) sont convertis en processus monodimensionnel à partir d'un parcours de Hilbert-Peano.

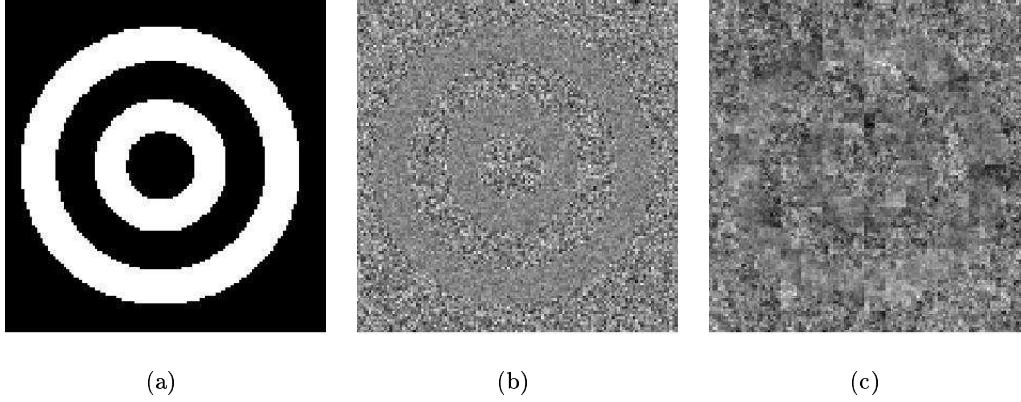


FIG. 2.4.3 – (a) image à deux classes, (b) sa version bruitée par un bruit indépendant et (c) sa version bruitée par un bruit à corrélation longue.

### 2.4.3 Segmentation non supervisée d'un processus à corrélation longue

Dans cette deuxième expérimentation, nous restaurons, de manière non supervisée, une image à deux classes avec 2 bruitages différents. De la même manière que dans la première expérimentation, nous supposons que  $X$  est une CM stationnaire obtenue après conversion de l'image (a) à partir d'un parcours de Hilbert-Peano. Dans les deux cas, les lois gaussiennes de  $B = (B_n)_{n=1}^N$  sont stationnaires, avec toutes les moyennes nulles.

La première image bruitée présentée sur la figure 2.4.3(b), est obtenue en utilisant un bruit indépendant avec des variances respectivement égales à 1 et 4 pour les deux classes. La deuxième image bruitée est obtenue en utilisant un bruit à corrélation longue, avec chaque variance égale à 1 et  $\alpha_{\omega_1} = 0.99$ ,  $\alpha_{\omega_2} = 0.2$ . Les paramètres sont estimés par un l'algorithme Espérance-Maximisation (EM) en considérant le modèle CMC-BI classique et le modèle CMC-BCL avec les moyennes fixées à 0 dans les deux cas. Les formules de réestimations concernant le modèle CMC-BI sont données par :

$$\begin{cases} c_{ij}^{[q+1]} = \frac{1}{N-1} \sum_{n=2}^N \psi_{n-1}^{[q]}(i, j) \\ (\sigma_i^2)^{[q+1]} = \frac{\sum_{n=1}^N y_n^2 \gamma_n^{[q]}(i)}{\sum_{n=1}^N \gamma_n^{[q]}(i)} \end{cases} \quad (2.4.7)$$

Les formules de réestimation de CMC-BCL sont identiques à celles de CMC-BI auxquelles on rajoute la formule de réestimation de  $r_1^{[q+1]}$  donnée par

$$r_1^{[q+1]}(i) = \frac{\sum_{n=2}^N y_{n-1} y_n \gamma_n^{[q]}(i)}{(\sigma_i^2)^{[q+1]} \sum_{n=2}^N \gamma_n^{[q]}(i)} \quad (2.4.8)$$

		CMC-BI		CMC-BCL	
		$\omega_1$	$\omega_2$	$\omega_1$	$\omega_2$
BI	$\sigma^2$	3.93	0.97	3.93	0.97
	$\alpha$	-	-	6.31	5.09
BCL	$\sigma^2$	1.25	0.33	1.02	0.70
	$\alpha$	-	-	0.97	0.32

TAB. 2.1 – Estimées EM de  $\sigma^2$  et de  $\alpha$  en considérant les deux modèles (CMC-BI et CMC-BCL) pour les deux bruits (BI et BCL).

$p(x_1, x_2)$	CMC-BI	CMC-BCL
BI	$\begin{pmatrix} 0.69 & 0.00 \\ 0.00 & 0.30 \end{pmatrix}$	$\begin{pmatrix} 0.69 & 0.00 \\ 0.00 & 0.30 \end{pmatrix}$
BCL	$\begin{pmatrix} 0.61 & 0.02 \\ 0.02 & 0.34 \end{pmatrix}$	$\begin{pmatrix} 0.67 & 0.00 \\ 0.02 & 0.32 \end{pmatrix}$

TAB. 2.2 – Estimées EM de  $p(x_1, x_2)$  des deux modèles (CMC-BI et CMC BCL) pour chacun des bruits (BI et BCL). La vraie valeur estimée à partir de l'image des classes initiale est  $\begin{pmatrix} 0.66 & 0.02 \\ 0.02 & 0.30 \end{pmatrix}$ .

à partir de laquelle on peut déduire les  $\alpha_i$  :

$$\alpha_i^{[q+1]} = -\frac{\log r_1^{[q+1]}(i)}{\log 2} \quad (2.4.9)$$

Les estimées sont données dans tab 2.1 et 2.2 et les processus cachés restaurés par MPM sont donnés dans la figure 2.4.4. Les résultats de la segmentation MPM concernant le bruit indépendant sont similaires lorsque l'on considère le modèle CMC-BI (a) et le modèle CMC-BCL (b), avec un taux de pixels mal classés égal à 5,3%. Nous notons également les valeurs élevées des estimées de  $\alpha_{\omega_1} = 6.31$  et  $\alpha_{\omega_2} = 5.09$ . En revanche, pour le bruit à mémoire longue, nous pouvons voir que le modèle CMC-BI (d) est incapable de prendre en compte la corrélation longue. Les mauvaises estimées se traduisent par une segmentation MPM pauvre avec un taux de sites mal classés égal à 27,6%. En revanche, le modèle CMC-BCL donne de meilleures estimées ce qui se traduit pas un bonne segmentation MPM (c) du processus caché, avec un taux de sites mal classés égal à 6,5%.

## 2.5 Conclusion

Nous avons présenté différentes modélisations possibles de la loi de  $Z$  et nous avons donné les algorithmes d'estimation des paramètres associés au modèle couple. On possède au final un modèle très général fondé sur l'unique hypothèse de markovianité du processus *a posteriori* permettant la prise en compte de lois plus complexes. Cela se traduit néanmoins par un

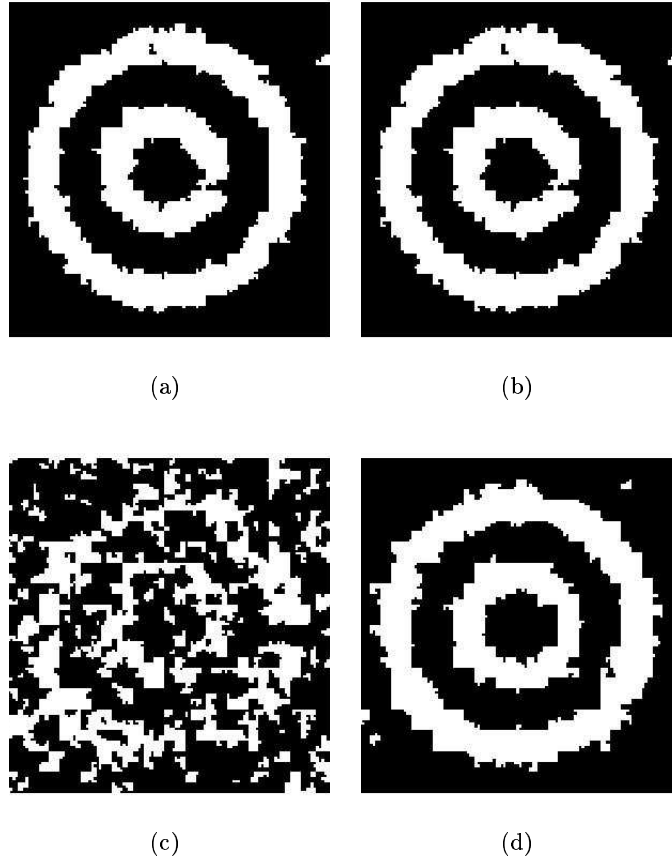


FIG. 2.4.4 – (a-b) Résultats de la segmentation MPM pour le bruit indépendant en considérant (a) le modèle CMC-BI (taux d’erreurs=5.3%) et (b) le modèle CMC-BCL (taux d’erreurs=5.3%). (c-d) Résultats de la segmentation MPM pour le bruit à corrélation longue en considérant (c) le modèle CMC-BI (taux d’erreurs=27.6%) et (d) le modèle CMC-BCL (taux d’erreurs=6.5%).

nombre de paramètres plus important et une plus grande complexité. Il faudra donc choisir le modèle en fonction de l'application visée. Enfin, nous avons présenté une application à la segmentation de processus à corrélation longue. Nous avons supposé que les processus était centré. Une extension intéressante serait de pouvoir adapter l'algorithme à la segmentation de bruits à corrélation longue non centrés pour ainsi l'appliquer à la segmentation d'image. Enfin, notons que l'estimation des variances par les formules de réestimation EM n'a pas été justifiée rigoureusement et que l'étude rigoureuse de l'application de EM aux CMC-BCL fait partie des perspectives pour la poursuite de l'étude.





## Chapitre 3

# Chaînes de Markov $M$ -non stationnaires cachées à bruit indépendant

Le calcul récursif des probabilités *a posteriori* d'un modèle de signal monodimensionnel  $Z = (X, Y)$ , dans lequel seul le processus  $Y$  est observé, est numériquement possible dès lors que le processus des états cachés *a posteriori* est une chaîne de Markov (CM). En partant de ce constat, le modèle classique de chaînes de Markov cachées à bruit indépendant (CMC-BI) a été étendu aux chaînes de Markov couples (CMCouples) puis aux chaînes couples partiellement de Markov (CCPM), de généralités croissantes. Par ailleurs, les modèles de Markov couples peuvent être étendus aux modèles de Markov triplets (CMT [95]). Pour cela, un processus auxiliaire est introduit tel que la loi jointe du triplet "processus caché, processus auxiliaire, processus observé" soit une CM. D'autres auteurs, dans le cadre de l'étude de modèles markoviens cachés monodimensionnels, ont eu recours à un processus auxiliaire pour modéliser par exemple l'influence d'un processus extérieur au modèle. C'est le cas de Hugues [60] qui utilise un processus auxiliaire pour modéliser l'influence des précipitations dans le cas d'un modèle non homogène pour des applications de pluviométrie. De même, le modèle de chaîne de Markov cachée à entrée/sortie (IOHMM [8]) inclut également un processus auxiliaire pour modéliser les données d'entrées d'une CMC-BI. Néanmoins, ces derniers ne considèrent pas directement la markovianité de la loi jointe des états cachés, des observations et du(des) processus auxiliaire(s). Une application possible des modèles triplets est la segmentation de processus cachés non stationnaires. Dans les modèles classiques, lors de l'estimation des paramètres, on suppose généralement que le processus caché est une CM stationnaire. Ainsi, lorsque le processus caché peut effectivement être modélisé par une CM stationnaire, les paramètres sont bien estimés et le processus caché l'est également. En revanche, lorsque le processus caché n'est pas stationnaire, les paramètres sont mal estimés, dû à une mauvaise correspondance entre le modèle estimé et le vrai modèle, ce qui conduit

généralement à une estimation du processus caché de mauvaise qualité. En introduisant un processus auxiliaire gouvernant les changements de régime du processus caché, nous montrons que les résultats de segmentation fondés sur le modèle triplet obtenu peuvent être améliorés. Nous définissons dans un premier temps le principe des modèles triplets. Nous explicitons alors les méthodes d'estimation bayésienne du processus caché dans le cas triplet. Enfin, nous proposons un modèle triplet particulier appelé *chaîne de Markov M-non stationnaire cachée à bruit indépendant* (CMM-NSC-BI) présenté dans [70, 68] permettant la segmentation de processus cachés non stationnaires.

### 3.1 Présentation générale des modèles triplets

Une extension possible des modèles présentés dans les chapitres précédents consiste à introduire un processus auxiliaire  $U = \{U_n\}_{n \in \mathcal{N}}$ , chaque  $U_n$  prenant ses valeurs dans  $\mathcal{U}$  qui sera dans notre étude fini. Dans le cas général, i.e. sans faire d'hypothèses d'indépendance conditionnelle, la loi de  $Z$  est alors donnée par  $p(z) = \sum_{u \in \mathcal{U}^{\mathcal{N}}} p(x, u, y) = \sum_{u \in \mathcal{U}^{\mathcal{N}}} p(t)$  que l'on peut réécrire sous la forme factorisée suivante

$$p(z) = \sum_{u \in \mathcal{U}^{\mathcal{N}}} p(t_1) \prod_{n=2}^{\mathcal{N}} p(t_n | t_{1:n-1}) \quad (3.1.1)$$

En considérant le processus  $T = (V, Y)$  avec  $V = (X, U)$ , l'algorithme 3 de calcul des marginales *a posteriori* peut être utilisé pour calculer les marginales *a posteriori*  $(p(v_n | y))_{n \in \mathcal{N}}$ . Il suffit pour cela de remplacer les  $x_n$  par les  $v_n$  pour tout  $n \in \mathcal{N}$ . Les marginales  $(p(x_n | y))_{n \in \mathcal{N}}$  nécessaire à l'estimation par MPM s'en déduisent alors aisément en intégrant les marginales  $p(v_n | y)$  par rapport aux variables  $u_n$  pour tout  $n \in \mathcal{N}$ . D'autre part, comme nous l'avons vu précédemment, il est nécessaire que le processus  $(X, U)$  *a posteriori* soit une CM pour que l'algorithme soit applicable numériquement. Chacun des modèles proposés précédemment peut ainsi être généralisé par l'introduction d'un processus auxiliaire  $U$  tel que :

- $V = (X, U)$  soit une CM. On parlera dans ce cas de *chaîne de Markov couple cachée* (CMCoC). Sa loi est de la forme 3.1.1 avec  $p(t_n | t_{1:n-1}) = p(v_n | v_{n-1})p(y_n | v_n)$
- $T = (Z, U)$  soit une CM. Ce modèle est appelé *chaîne de Markov triplet* (CMT [95]).
- $T_{n+1} \perp\!\!\!\perp V_{1:n-1} | (T_n, Y_{1:n-1})$ . Ce modèle est appelé *chaîne triplet partiellement de Markov* (CTPM [98]).

Considérons à présent une CMT  $T = (X, U, Y)$  telle que  $X_n$  prenne ses valeurs dans  $\mathcal{X}$ , chaque  $Y_n$  prenant ses valeurs dans l'ensemble des nombres réels  $\mathbb{R}$ , et chaque  $U_n$  prenant ses valeurs dans l'ensemble  $\mathcal{U}$ . Lorsque  $T$  est stationnaire, sa loi est donnée par  $p(t_1, t_2)$ . Nous cherchons à savoir si  $(X, U)$  est une CM ou non ; pour cela, nous appliquons le lemme 2.2.1 vu au chapitre 1 à  $G = (X, U)$ , avec  $\gamma$  la mesure de comptage sur  $\mathcal{G} = \mathcal{X} \times \mathcal{U}$  et  $H = Y$ , avec  $\eta$  la mesure de Lebesgue sur  $\mathbb{R}$ . En supposant que les hypothèses du lemme sont vérifiées et en écrivant  $p(t_1, t_2)$  comme  $p(t_1, t_2) = p(x_1, u_1, x_2, u_2)p(y_1, y_2 | x_1, u_1, x_2, u_2)$ , nous pouvons

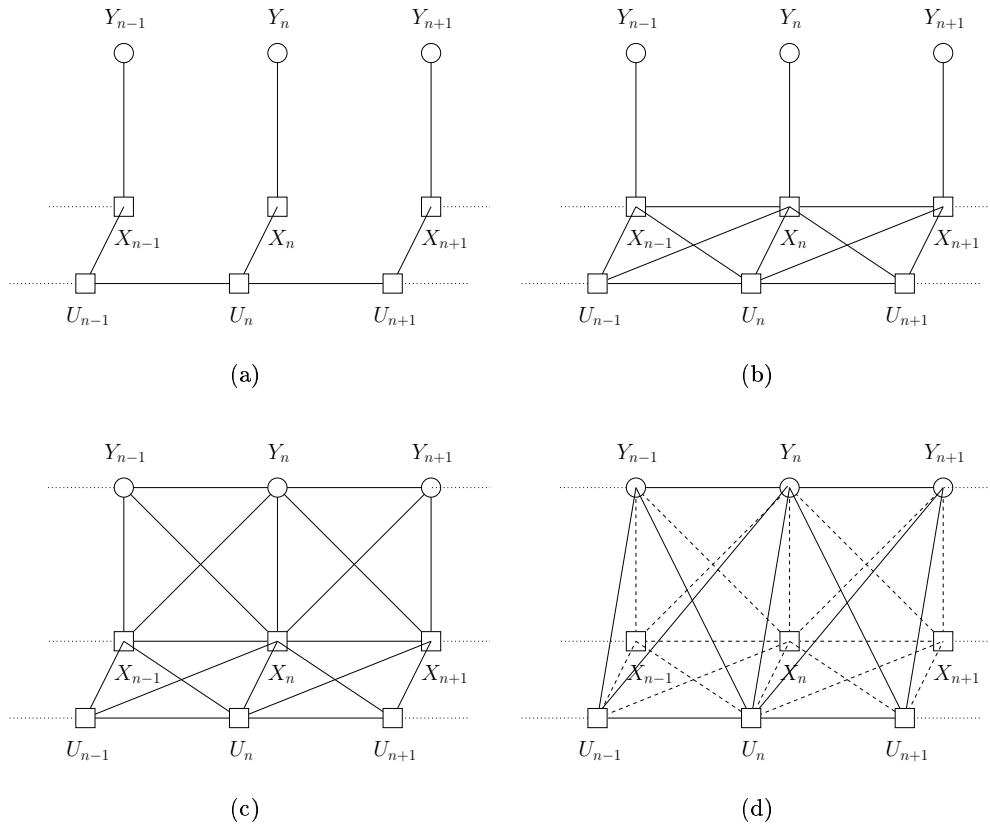


FIG. 3.1.1 – Graphes d'indépendance de cas particuliers de modèles triplet parmi lesquels : b) chaîne de Markov cachée non stationnaire, d) chaîne de Markov triplet

dire que  $(X, U)$  est une CM ssi  $p(y_1|x_1, u_1, x_2, u_2) = p(y_1|x_1, u_1)$ . D'autre part, nous pouvons également nous demander si  $X$  est une CM ; pour cela, nous pouvons appliquer le lemme avec  $G = X$  avec  $\gamma$  la mesure de comptage sur  $\mathcal{G} = \mathcal{X}$  et  $H = (U, Y)$ , avec  $\eta$  le produit des mesure de comptage sur  $\mathcal{U}$  et par la mesure de Lebesgue sur  $\mathbb{R}$ . En écrivant  $p(t_1, t_2)$  comme  $p(t_1, t_2) = p(x_1, x_2)p(u_1, y_1, u_2, y_2|x_1, x_2)$ , nous pouvons dire que  $X$  est une CM ssi  $p(u_1, y_1|x_1, x_2) = p(u_1, y_1|x_1)$ . Il est donc possible de définir des modèles pour lesquels ni  $X$ , ni  $(X, U)$ , ni  $Z$  ne sont des CM. La figure (3.1.1) donne les graphes d'indépendance conditionnelle de quelques exemples de modèles particuliers de CMT. Dans la section suivante, nous présentons un modèle original (dont le graphe d'indépendance conditionnelle correspond à 3.1.1(a)) permettant la segmentation de processus non stationnaires.

### 3.2 Chaînes de Markov $M$ -non stationnaires cachées à bruit indépendant

Lors de l'estimation du processus caché, nous faisons en général l'hypothèse que ce dernier est stationnaire. Ainsi, si le processus caché n'est pas stationnaire, le fait de le supposer

stationnaire conduit à une mauvaise estimation des paramètres, ce qui se traduit par une mauvaise segmentation. L'idée du modèle que nous proposons est de modéliser la non stationnarité du processus caché en utilisant un processus auxiliaire  $U$ . Ce processus auxiliaire gouverne ainsi les changements de matrices de transition et modélise de cette manière la non stationnarité de la chaîne  $Z = (X, Y)$  dans le sens suivant :

**Définition 3.2.1.** *Soit  $Z = (X, Y)$  une chaîne couple, où  $Y$  est observé et  $X$  est caché.  $Z$  sera dit "chaîne de Markov Couple  $M$ - non stationnaire" (CMCouple $M$ -NS) s'il existe un troisième p.a  $U = \{U_n\}_{n \in \mathcal{N}}$  chaque  $U_n$  prenant ses valeurs dans l'ensemble fini  $\mathcal{U}$  avec  $|\mathcal{U}| = M$  tel que  $T = (X, U, Y) = (X_n, U_n, Y_n)_{n \in \mathcal{N}}$  est une chaîne de Markov stationnaire.*

Nous devons préciser quelle est la différence entre une CMCouple non stationnaire classique et une CMCouple  $M$ -non stationnaire. Dans le cas de CMC-BI ou de CMCouple non stationnaires classiques, la chaîne  $Z$  est une CM et est non stationnaire, ce qui signifie que  $p(z_{n-1:n})$  dépend de  $n \in \mathcal{N} \setminus \{1\}$ . Dans une CMCouple $M$ -NS, la loi de la chaîne  $Z$  est la loi marginale d'un chaîne stationnaire  $T = (X, U, Y)$ ; Ainsi,  $Z$  n'est pas une CM dans le cas général d'une part et  $p(z_{n-1:n})$  ne dépend pas de  $n \in \mathcal{N} \setminus \{1\}$  d'autre part. En fait,  $p(z_{n-1:n})$  est la loi marginale de  $p(t_{n-1:n})$  qui ne dépend pas de  $n \in \mathcal{N} \setminus \{1\}$ . Cependant, en considérant une CMT  $T$  associée à une CMCouple $M$ -NS  $Z$ , nous pouvons dire que pour tout  $U = u$ , la loi  $p(z|u)$  de  $Z$  conditionnellement à  $U = u$  est une CM qui est non stationnaire dans le sens classique présenté précédemment.

Dans la suite, nous considérons le cas particulier où le processus  $V = (X, U)$  est une CM stationnaire et la loi  $p(y|v)$  de  $Y$  conditionnelle à  $V$  vérifie :  $p(y|v) = p(y|x) = \prod_{n=1}^N p(y_n|x_n)$ . Ce modèle appelé Chaîne de Markov  $M$ -non stationnaire cachée à bruit indépendant CMM-NSC-BI sera dénommé par la suite CMT afin de simplifier les notations. Notons qu'en dépit de la simplicité de  $p(y|v)$ , le processus couple  $Z$  n'est pas nécessairement une CMC-BI, car  $X$  n'est pas nécessairement une CM. D'autre part, supposons que  $p(y_n|x_n)$  soit gaussienne et indépendante de  $n \in \mathcal{N}$ . Ainsi, pour  $K$  classes, nous devons estimer  $K$  moyennes  $\mu_{1:K}$ , et  $K$  variances  $\sigma_{1:K}^2$  des  $K$  densités gaussiennes  $p(y_n|x_n = \omega_1), \dots, p(y_n|x_n = \omega_K)$ . De plus, chaque  $U_n$  prenant ses valeurs dans  $\mathcal{U}$ , la loi de la CM stationnaire  $V$  est donnée par  $(K \times M)^2 - 1$  paramètres  $p_{ij} = p(v_1 = i, v_2 = j)$ , qui est une probabilité sur  $(\mathcal{X} \times \mathcal{U})^2$ . Nous utilisons la méthode d'estimation SEM, présentée précédemment, qui pour ce type de modèle converge plus rapidement vers une solution satisfaisante, l'algorithme EM ayant une forte tendance à converger vers des maximums locaux. La méthode d'estimation se déroule de la manière suivante :

- (i) On considère une valeur initiale  $\phi^{[0]} = (p_{ij}^{[0]}, \mu_k^{[0]}, (\sigma_k^{[0]})^2)$ , pour  $1 \leq i \leq K \times M$ ,  $1 \leq j \leq K \times M$ , et  $1 \leq k \leq K$  ;
- (ii) Pour chaque  $q \in N^*$  :
  - simuler  $V = v^{[q]}$  par rapport à la CM  $p(v|y)$  basée sur  $\phi^{[q]}$  ;

- calculer  $\phi^{[q+1]} = (p_{ij}^{[q+1]}, \mu_k^{[q+1]}, (\sigma_k^{[q+1]})^2)$  avec

$$p_{ij}^{[q+1]} = \frac{1}{N-1} \sum_{n=2}^N 1_{[v_{n-1}=i, v_n=j]}$$

$$\mu_k^{[q+1]} = \frac{\sum_{n=1}^N y_n 1_{[x_n=\omega_k]}}{\sum_{n=1}^N 1_{[x_n=\omega_k]}}$$

$$(\sigma_k^{[q+1]})^2 = \frac{\sum_{n=1}^N (y_n - \mu_k^{[q+1]})^2 1_{[x_n=\omega_k]}}{\sum_{n=1}^N 1_{[x_n=\omega_k]}}$$

(iii) arrêter la procédure lorsque la séquence  $(\phi^{[q]})$  devient stable par rapport à un critère donné.

### 3.3 Segmentation non supervisée de signaux et d'images non stationnaires.

Nous présentons 4 séries d'expérimentations. Dans la première série, les données à segmenter sont entièrement simulées. Dans la deuxième série, le processus caché est une image binaire. Dans la troisième série nous traitons de la segmentation de textures de Brodatz. Finalement, dans la quatrième série, nous traitons le cas d'une image réelle.

#### 3.3.1 Comparaison entre CMT et CMC-BI

##### Processus synthétiques

Considérons  $\mathcal{X} = \{\omega_1, \omega_2\}$  et  $\mathcal{U} = \{\lambda_1, \lambda_2, \lambda_3\}$ . Il y a donc 2 classes et trois différentes stationnarités pour la CM  $X$ . La grille de sites considérée est constituée de  $N = 128 \times 128$  sites. Nous présentons 2 cas. Dans le premier cas, le processus  $V = (X, U)$  est une CM et suit donc le modèle proposé. Dans le deuxième cas,  $V$  ne suit pas le modèle proposé  $U$  étant simulé de manière arbitraire. Dans les deux cas, une réalisation  $X = x$  est simulée conditionnellement à la réalisation  $U = u$  à partir de (3.3.1) :

$$\begin{cases} \forall (x_1, u_1) \in \mathcal{X} \times \mathcal{U}, p(x_1|u_1) = 1/2 \\ \forall n \in \mathcal{N} \setminus \{1\}, \{p(x_n|x_{n-1}, u_n = \lambda_i)\}_{x_{n-1:n} \in \mathcal{X}^2} = M_{\lambda_i} \end{cases} \quad (3.3.1)$$

avec

$$M_{\lambda_1} = \begin{bmatrix} 0.98 & 0.02 \\ 0.02 & 0.98 \end{bmatrix}, M_{\lambda_2} = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}, M_{\lambda_3} = \begin{bmatrix} 0.02 & 0.98 \\ 0.98 & 0.02 \end{bmatrix} \quad (3.3.2)$$

Les observations  $Y = y$  sont alors échantillonnées par rapport à  $(p(y_n|x_n))_{n \in \mathcal{N}}$ , où  $p(y_n|x_n = \omega_1)$  est une gaussienne de moyenne 1 et de variance 1, et  $p(y_n|x_n = \omega_2)$  est une gaussienne

de moyenne 3 et de variance 1 :

$$\forall n \in \mathcal{N}, \begin{cases} p(y_n|x_n = \omega_1) = \mathcal{N}(1, 1) \\ p(y_n|x_n = \omega_2) = \mathcal{N}(3, 1) \end{cases} \quad (3.3.3)$$

Une réalisation  $X = x$  est alors estimée au sens du MPM à partir de  $Y = y$  de deux manières différentes :

- la première segmentation  $\hat{X} = \hat{x}_{CMC-BI}$  est obtenue en utilisant les paramètres estimés par l'algorithme SEM initialisé à partir d'un algorithmes des k-moyennes, en considérant que  $Z$  est une CMC-BI classique, le processus caché  $X$  étant supposé stationnaire,
- la deuxième segmentation  $\hat{X} = \hat{x}_{TMC}$  est obtenue en utilisant les paramètres estimés à partir de l'algorithme SEM en supposant que  $T$  est la CMT présentée précédemment avec 2 vraies classes et trois classes auxiliaires.

**1er cas :  $V = (X, U)$  est une CM** Dans ce premier cas, l'image simulée correspond au modèle non stationnaire étant donné que  $V = (X, U)$  est une CM dont la loi est donnée par  $p(v) = p(v_1) \prod_{n=2}^N p(v_n|v_{n-1})$  telle que

$$p(v_n|v_{n-1}) = p(x_n|x_{n-1}, u_n)p(u_n|u_{n-1}) \quad (3.3.4)$$

On simule dans une premier temps une réalisation de  $U = u$  à partir de (3.3.5) :

$$\begin{cases} \forall u_1 \in \mathcal{U}, p(u_1) = 1/3 \\ \forall n \in \mathcal{N} \setminus \{1\}, \{p(u_n|u_{n-1})\}_{u_{n-1:n} \in \mathcal{U}^2} = M_U \end{cases} \quad (3.3.5)$$

avec

$$M_U = \begin{bmatrix} 0.99 & 0.01 & 0 \\ 0 & 0.97 & 0.3 \\ 0.3 & 0 & 0.97 \end{bmatrix} \quad (3.3.6)$$

puis  $X = x$  et  $Y = y$  sont échantillonnées à partir de (3.3.1) et (3.3.3). Les réalisations  $U = u$ ,  $X = x$ , et  $Y = y$  sont représentées sur la figure 3.3.1 après transformation du processus monodimensionnel en image à partir du parcours de Peano. En considérant que le modèle est une CMC-BI, les paramètres sont mal estimés comme on peut le voir sur le tableau 3.1. Cela se traduit par une mauvaise segmentation  $\hat{x}_{CMC-BI}$  du processus caché avec un taux de pixels mal classifiés  $\tau = 19\%$ . Si l'on considère à présent que le modèle est une CMT, les paramètres sont mieux estimés et le taux de pixels mal classifiés chute à  $\tau = 4,5\%$ . D'autre part, on peut également estimer une réalisation du processus  $U = u$ , ce qui peut être d'un grand intérêt pratique, notamment en segmentation de texture. Les estimations  $\hat{x}_{CMC-BI}$ ,  $\hat{x}_{CMT}$  et  $\hat{u}$  sont également représentés sur la figure 3.3.1.

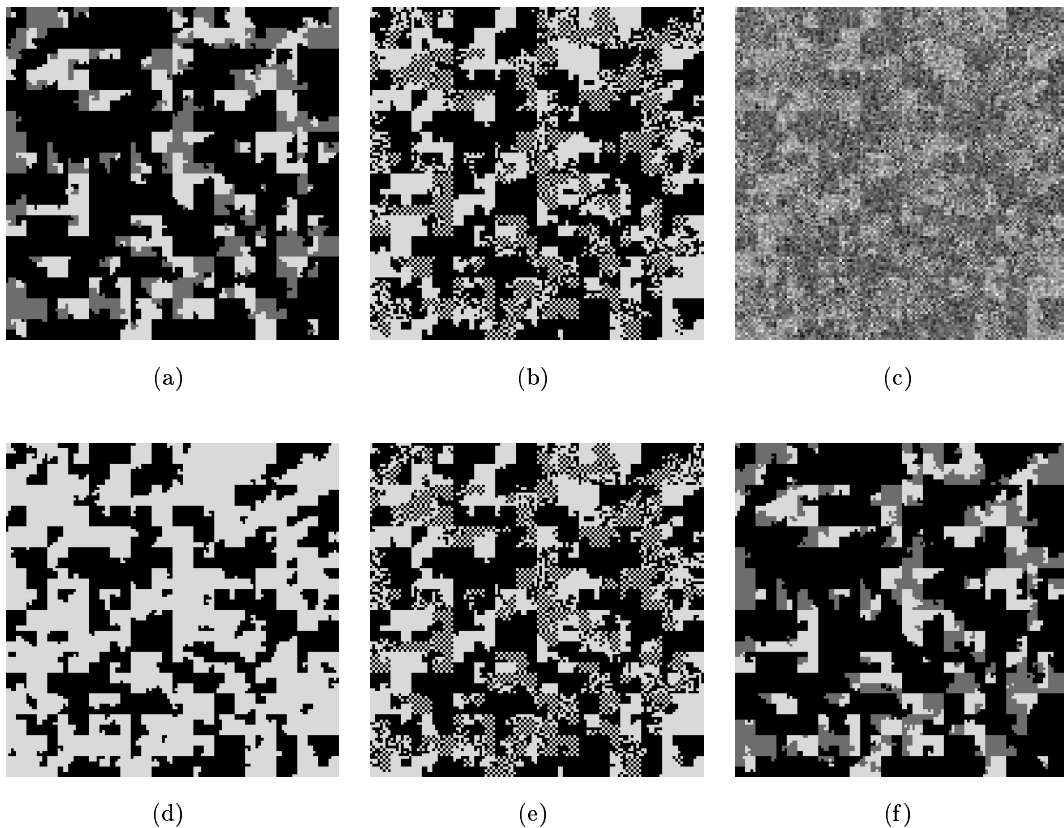


FIG. 3.3.1 – Segmentation non supervisée d’un processus synthétique pour lequel  $V$  est une CM : (a) simulation de  $U = u$ , (b) simulation de  $X = x$  conditionnellement à  $U = u$ , (c) simulation de  $Y = y$  à partir de  $X = x$ , (d) Résultat de l’estimation  $\hat{X} = \hat{x}_{CMC-BI}$  au sens du MPM obtenue en utilisant les paramètres estimés par l’algorithme SEM initialisé à partir d’un algorithmes des k-moyennes, en considérant que  $Z$  est une CMC-BI classique,  $\tau = 19\%$ , (e) Résultat de l’estimation de  $\hat{X} = \hat{x}_{TMC}$  au sens du MPM obtenue en utilisant les paramètres estimés par l’algorithme SEM initialisé à partir d’un algorithmes des k-moyennes, en considérant que  $Z$  est une CMT,  $\tau = 4,5\%$  et (f) estimation de  $U = u$ .

	$\mu_1$	$\sigma_1^2$	$\mu_2$	$\sigma_2^2$		$v_1$	$v_2$	$v_3$	$v_4$	$v_5$	$v_6$
Vrai	1.00	1.00	3.00	1.00	$p(v_n v_{n-1})$	<b>0.97</b>	<b>0.02</b>	0.00	0.00	0.00	0.00
CMC-BI	2.4	1.82	1.03	1.04	$v_1$	<b>0.00</b>	<b>0.98</b>	0.00	0.00	0.01	0.00
CMT	0.99	1.02	3.01	0.98	$v_2$	0.00	0.03	<b>0.00</b>	<b>0.94</b>	0.02	0.00
					$v_3$	0.03	0.00	<b>0.97</b>	<b>0.00</b>	0.00	0.00
					$v_4$	0.00	0.00	0.00	0.00	<b>0.46</b>	<b>0.54</b>
					$v_5$	0.00	0.00	0.80	0.00	<b>0.45</b>	<b>0.46</b>
					$v_6$	0.23	0.40	0.09	0.09	0.09	0.09
					$p(v_1)$						

TAB. 3.1 – Segmentation non supervisée d’un processus synthétique pour lequel  $V$  est une CM. A gauche : estimées des paramètres des lois d’observations, à droite : estimées de la matrice de transition  $p(v_n|v_{n-1})$  et de  $p(v_1)$ .

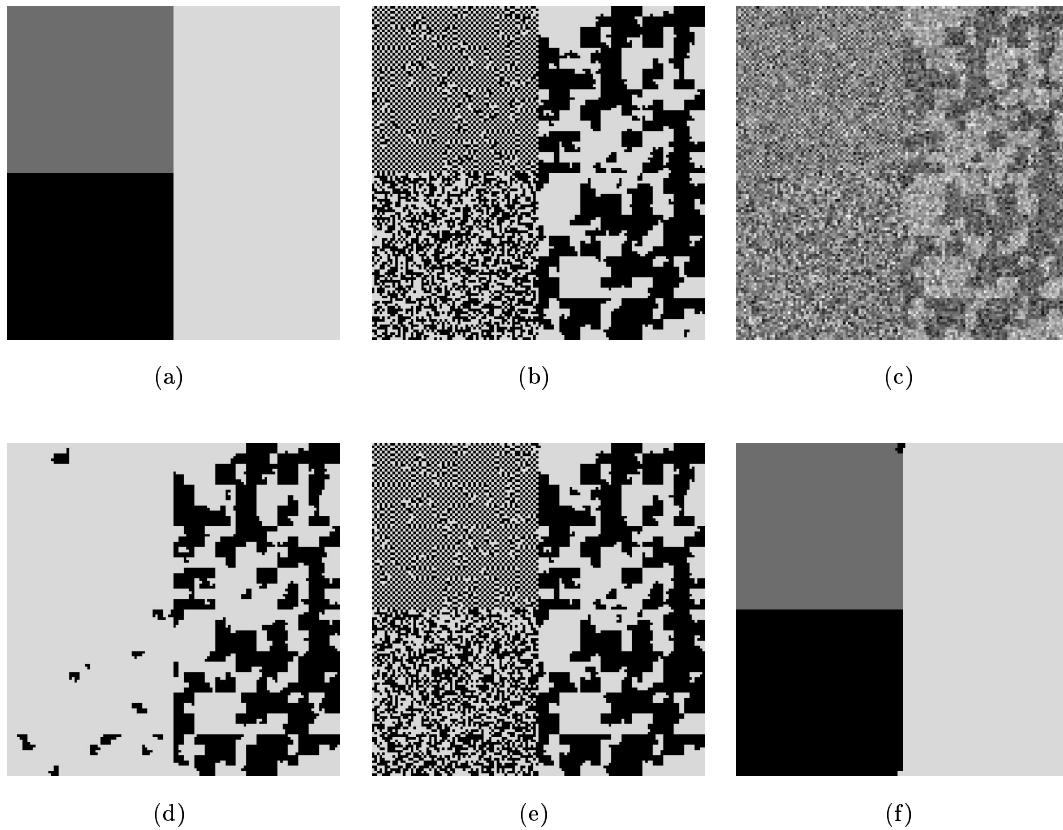


FIG. 3.3.2 – Segmentation non supervisée d'un processus synthétique pour lequel  $V$  n'est pas une CM : (a) simulation de  $U = u$ , (b) simulation de  $X = x$  conditionnellement à  $U = u$ , (c) simulation de  $Y = y$  à partir de  $X = x$ , (d) Résultat de l'estimation de  $\hat{X} = \hat{x}_{CMC-BI}$  au sens du MPM obtenue en utilisant les paramètres estimés par l'algorithme SEM initialisé à partir d'un algorithmes des k-moyennes, en considérant que  $Z$  est une CMC-BI classique,  $\tau = 26\%$ , (e) Résultat de l'estimation  $\hat{X} = \hat{x}_{TMC}$  au sens du MPM obtenue en utilisant les paramètres estimés par l'algorithme SEM initialisé à partir d'un algorithmes des k-moyennes, en considérant que  $Z$  est une CMT,  $\tau = 5\%$  et (f) estimation de  $U = u$ .

**2ème cas :  $V = (X, U)$  n'est pas une CM** Nous souhaitons à présent vérifier le comportement du modèle pour une image qui ne suit pas exactement le modèle. La réalisation  $U = u$  est simulée arbitrairement de la manière suivante : les sites appartenant au premier quart de  $U = u$  sont dans l'état  $\lambda_1$ , ceux appartenant au second quart dans l'état  $\lambda_2$  et enfin ceux de la dernière moitié sont dans l'état  $\lambda_3$ . Puis  $X = x$  et  $Y = y$  sont échantillonnés à partir de (3.3.1) et (3.3.3). Les réalisations  $U = u$ ,  $X = x$ , et  $Y = y$  sont représentées sur la figure 3.3.2. De manière similaire au premier cas, les paramètres sont mal estimés en considérant le modèle CMC-BI (voir tableau 3.2) ce qui se traduit par une mauvaise segmentation  $\hat{x}_{CMC-BI}$  du processus caché avec un taux de pixels mal classifiés  $\tau = 26\%$ . Ce taux diminue fortement en considérant le modèle CMT ( $\tau = 5\%$ ). Les estimations  $\hat{x}_{CMC-BI}$ ,  $\hat{x}_{CMT}$  et  $\hat{u}$  sont également représentés sur la figure 3.3.2.



					$p(v_n v_{n-1})$							
					$v_1$	$v_2$	$v_3$	$v_4$	$v_5$	$v_6$		
	$\mu_1$	$\sigma_1^2$	$\mu_2$	$\sigma_2^2$	$v_1$	<b>0.98</b>	<b>0.02</b>	0.00	0.00	0.00	0.00	
					$v_2$	<b>0.02</b>	<b>0.98</b>	0.00	0.00	0.01	0.00	
Vrai	1.00	1.00	3.00	1.00	$v_3$	0.00	0.00	<b>0.50</b>	<b>0.49</b>	0.00	0.00	
CMC-BI	0.98	1.03	2.33	1.91	$v_4$	0.00	0.00	<b>0.52</b>	<b>0.47</b>	0.00	0.00	
CMT	1.02	1.01	2.98	0.94	$v_5$	0.00	0.00	0.00	0.00	<b>0.00</b>	<b>0.99</b>	
					$v_6$	0.00	0.00	0.00	0.00	<b>0.95</b>	<b>0.04</b>	
					$p(v_1)$	0.26	0.24	0.13	0.12	0.12	0.13	

TAB. 3.2 – Segmentation non supervisée d’un processus synthétique pour lequel  $V$  n’est pas une CM. A gauche : estimées des paramètres des lois d’observations, à droite : estimées de la matrice de transition  $p(v_n|v_{n-1})$  et de  $p(v_1)$ .

Finalement, l’utilisation des CMT permet d’améliorer les résultats obtenus à partir du modèle CMC-BI classique en considérant 3 non stationnarités différentes même si le signal traité ne correspond pas exactement au modèle CMT. Bien sûr, dans les expérimentations présentées, les données ne sont pas fortement bruitées ce qui se traduit par de très bons résultats de segmentation. Cependant, ces exemples montrent que la non stationnarité de  $X$  peut avoir des conséquences très négatives sur l’efficacité des méthodes non supervisées fondées sur le modèle CMC-BI et ces résultats montrent également que ces conséquences négatives peuvent être évitées en utilisant une CMT.

### Processus caché réel

Dans la deuxième expérimentation, nous considérons l’image binaire d’un zèbre présenté sur la figure 3.3.3a avec  $N = 256 \times 256$ . La peau du zèbre possède des bandes de tailles et d’espacement variés. Par exemple, les bandes situées sur le cou et les jambes du zèbre sont plus fines que celles situées sur le corps. L’observation (figure 3.3.3b) est alors échantillonnée de la même manière que dans la première expérimentation et une réalisation  $X = x$  est estimée par MPM à partir de  $Y = y$  en utilisant les paramètres estimés par SEM en considérant le modèle CMC-BI puis le modèle CMT. En supposant la stationnarité de la chaîne cachée, nous pouvons voir sur la figure 3.3.3c que les bandes ne sont pas bien restaurées, ce qui donne  $\hat{X} = \hat{x}_{CMC-BI}$ ,  $\tau = 7,2\%$ . Si nous supposons à présent que  $T$  est une CMT avec 3 classes auxiliaires, nous pouvons prendre en compte 3 stationnarités différentes, ce qui est montré sur la figure 3.3.3e où sont représentées les estimées de  $\hat{U} = \hat{u}$  du processus auxiliaire : les labels noirs sont assignés au fond, les labels gris au corps qui présente des bandes larges et les labels blancs sont assignés au cou et aux jambes du zèbre qui présentent des bandes fines. En considérant différentes stationnarités, nous améliorons ainsi la qualité de la segmentation  $\hat{X} = \hat{x}_{CMT}$  avec  $\tau = 3,5\%$  (voir figure 3.3.3f).

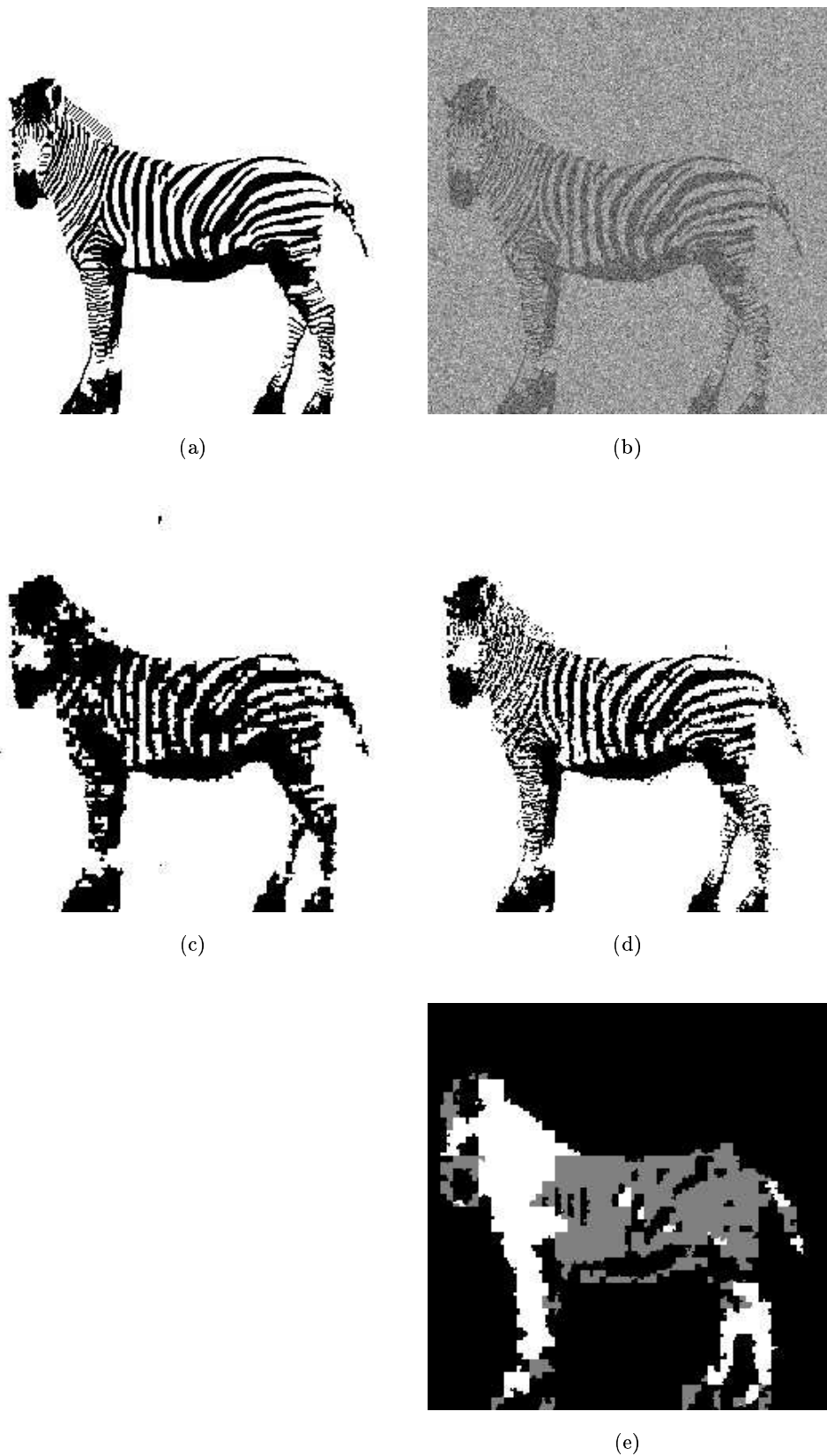


FIG. 3.3.3 – a) image initiale cachée, b) image bruitée, c) segmentation par MPM en considérant le modèle CMC-BI,  $\hat{X} = \hat{x}_{CMC-BI}$ ,  $\tau = 7.2\%$ , d) segmentation par MPM en considérant le modèle CMT,  $\hat{X} = \hat{x}_{CMT}$ ,  $\tau = 3.5\%$ , e) processus auxiliaire estimé  $\hat{U} = \hat{u}$ .

### 3.3.2 Segmentation d'images réelles

#### Textures de Brodatz

La troisième série d'expérimentations concerne la segmentation non supervisée de patterns de Brodatz qui sont souvent utilisés en segmentation de texture. Nous construisons une image  $Y = y$  (figure 3.3.4b) par rapport au processus  $U = u$  présenté sur la figure 3.3.4a en considérant 3 textures différentes de Brodatz. Nous supposons que  $T = (X, U, Y)$  est une CMT avec 3 classes auxiliaires et 4 classes réelles. Les résultats de la segmentation sont présentés sur la figure 3.3.4. Le point intéressant est le fait que  $\hat{U} = \hat{u}$  soit bien estimé avec  $\tau = 2.5\%$ , ce qui montre que les modèles CMT peuvent être utiles dans le contexte de la segmentation non supervisée de texture.

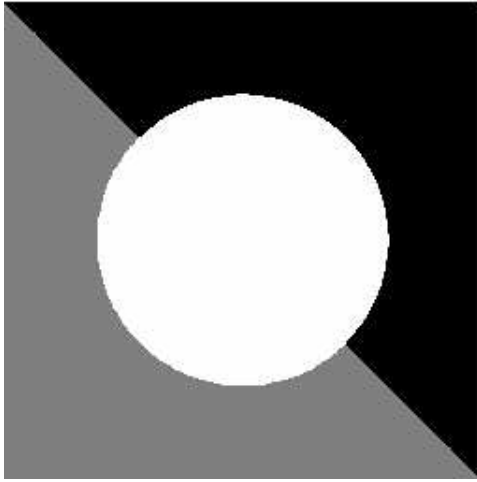
#### Image Tokyo

Notre quatrième expérimentation est un exemple de segmentation non supervisée d'une image satellite de Tokyo de taille  $256 \times 256$  présentée sur la figure 3.3.5. Nous segmentons l'image en considérant différents nombres de classes réelles et auxiliaires. Les comparaisons entre les modèles sont faites en utilisant le critère d'information de Bayes (BIC). Ce critère est défini par  $BIC = -2LL + q \log(N)$  où  $LL$  est la log-vraisemblance du modèle,  $q$  est le nombre de paramètres indépendants et  $N$  et le nombre de données. Nous ne prenons pas en compte les paramètres estimés égaux à 0 relativement à la convention établie dans [13]. Nous avons testé différents cas avec deux, trois et quatre classes réelles, le nombre d'états auxiliaires variant de un (qui donne le modèle CMC-BI) à trois. Les résultats sont présentés dans le tableau 3.3.

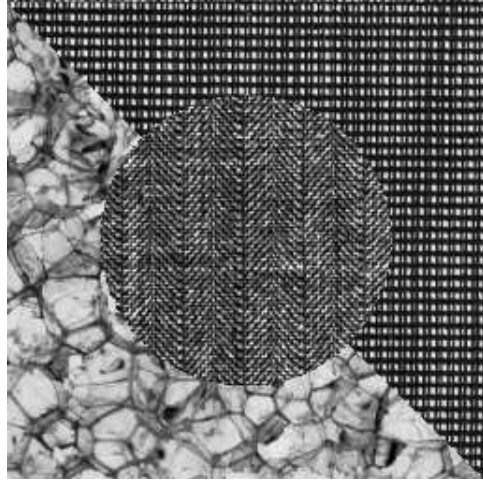
Etant donné que le meilleur modèle est celui donnant le BIC le plus faible, nous voyons que le cas avec 4 classes réelles et 2 états auxiliaires est le plus adapté parmi les modèles que nous avons testés. Les résultats de segmentation non supervisée correspondant à 4 classes pour 1, 2 et 3 états auxiliaires sont présentés sur la figure 3.3.6. D'autre part, en se référant aux valeurs calculées du BIC, nous pouvons remarquer qu'il est préférable pour cette image d'utiliser un modèle à 3 classes réelles et 2 classes auxiliaires plutôt qu'un modèle à 4 classes réelles et 1 classe auxiliaire.

## 3.4 Conclusion

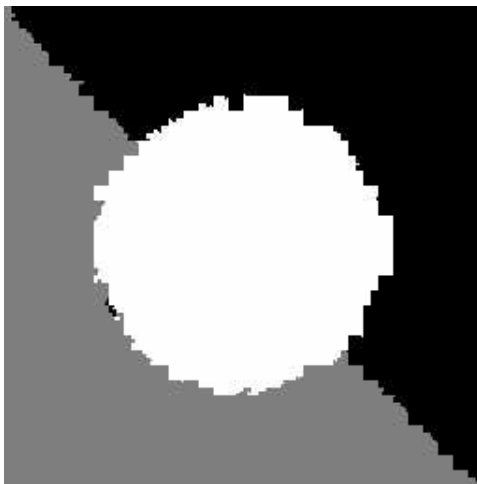
Nous avons montré dans ce chapitre comment les modèles proposés dans les chapitres précédents (à l'exception du modèle le plus général de CCPM) peuvent être étendus à des modèles triplets en ajoutant un ou plusieurs processus auxiliaires. Nous avons proposé un modèle de chaînes de Markov  $M$ - non stationnaires cachées à bruit indépendant permettant de segmenter des processus cachés non stationnaires. Nous avons finalement présenté différents résultats de simulation montrant l'amélioration obtenue en utilisant le nouveau modèle,



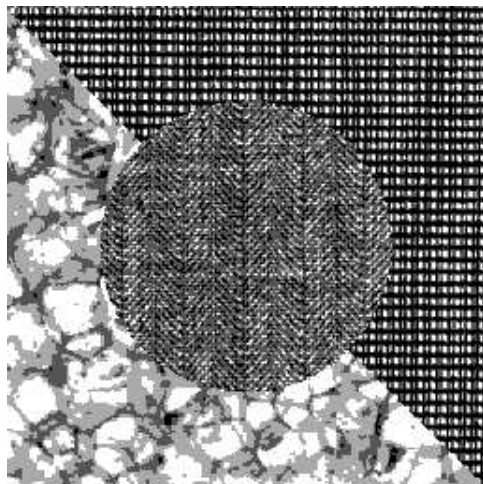
(a)



(b)



(c)



(d)

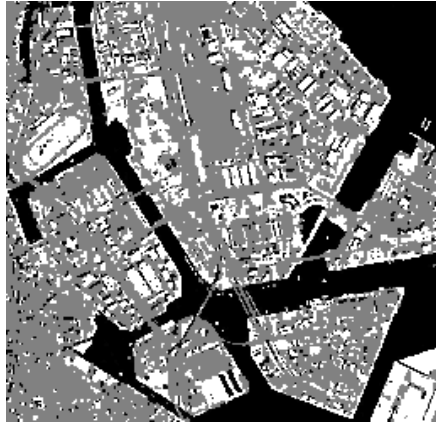
FIG. 3.3.4 – a) processus auxiliaire initial, b) image observée, c) processus auxiliaire estimé,  $\hat{U} = \hat{u}$ ,  $\tau = 2.5\%$ , d) processus caché estimé  $\hat{X}$ .

$K$	$M$	$LL$	$q$	BIC
2	1	-338750	7	677580
2	2	-337460	13	675060
2	3	-337360	23	<b>674980</b>
3	1	-334310	14	668780
3	2	-332070	27	664440
3	3	-331760	55	<b>664130</b>
4	1	-333020	21	666280
4	2	-331210	43	<b>662610</b>
4	3	-330830	121	663002

TAB. 3.3 – Comparaisons entre les modèles avec différentes valeurs de  $K$  (nombre de classes réelles) et différentes valeurs de  $M$  (nombre de classes auxiliaires)



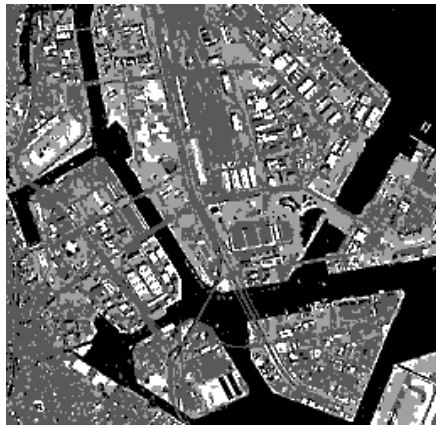
FIG. 3.3.5 – Image de tokyo.



(a)



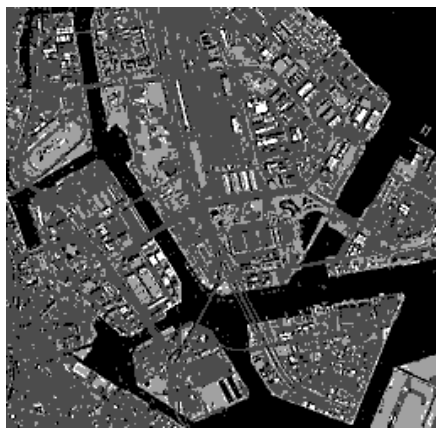
(b)



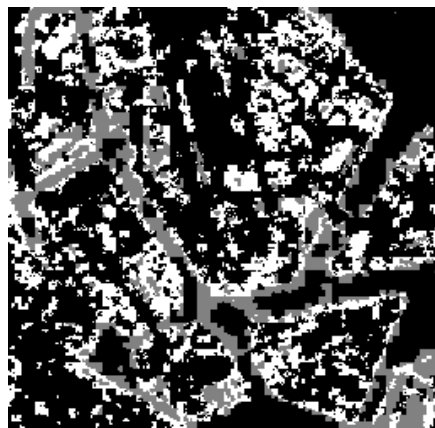
(c)



(d)



(e)



(f)

FIG. 3.3.6 – Estimation de réalisations  $\hat{x}$  en considérant 4 états réels et différents nombres d'états auxiliaires. Colonne de gauche : réalisations  $\hat{x}$  estimées, colonne de droite : réalisations de  $\hat{u}$  correspondantes. 1ère ligne : 1 état auxiliaire (CMC-BI), 2ème ligne : 2 états auxiliaires, 3ème ligne : 3 états auxiliaires.

au niveau de l'estimation des processus cachés, par rapport aux résultats obtenus en considérant le modèle classique. Finalement, nous avons montré que l'estimation du processus auxiliaire peut être utile dans le cadre de la segmentation de texture. Précisons qu'un autre cas particulier de modèle triplet permettant la segmentation de processus non stationnaires, les chaînes semi de Markov, a été récemment étudié dans [73]. Dans le chapitre suivant, nous proposons un autre modèle permettant la segmentation de processus non stationnaires fondé sur la théorie des croyances.





## Chapitre 4

# Chaînes de Markov évidentielles cachées à bruit indépendant

Nous proposons dans ce chapitre un autre modèle triplet [69, 70, 71], permettant la segmentation de processus non stationnaires, fondé sur la théorie des croyances. Par rapport au modèle CMM-NSC-BI dans lequel on considérait différentes zones de stationnarités correspondant chacune à un état auxiliaire, nous allons utiliser une méthode d'affaiblissement de la loi *a priori* propre à la théorie des croyances. Pour cela, nous introduisons dans un premier temps les principes de la théorie des croyances et en particulier la fusion de Dempster Shafer (fusionDS) qui généralise la règle de Bayes. Nous présentons alors la méthode d'affaiblissement que l'on utilisera par la suite du chapitre. Puis, nous montrons comment appliquer la fusionDS lorsque la loi *a priori* est évidentielle. Cela nous permettra de définir un modèle de chaîne de Markov évidentielle cachée permettant la segmentation de processus non stationnaire. Nous présenterons finalement des résultats d'expérimentations sur images simulées et image réelles.

### 4.1 Généralités sur la théorie des croyances

La théorie des croyances fut introduite dans les années 60 par Dempster puis reformulée dans les années 1970 par Shafer [110]. Ce dernier présenta ses travaux comme une généralisation de la théorie de Bayes au traitement de l'incertain. Nous noterons  $\Omega = \{\omega_1, \dots, \omega_K\}$  le cadre de discernement représentant l'ensemble des hypothèses en faveur desquelles une décision peut-être prise. Nous supposerons que celles-ci sont exhaustives et exclusives (hypothèse du monde fermé défini par Smets [111]). Dans la modélisation probabiliste, les probabilités sont définies sur cet ensemble en fonction de la connaissance que l'on possède de la source d'information. On modélise de cette façon l'incertitude sur les événements, l'imprécision étant considérée comme nulle. Ceci peut-être un inconvénient en fusion de données, plus particulièrement en présence de données imparfaites. Dans le cadre de la théorie des

croyances, les fonctions de masses ne sont pas définies sur l'ensemble  $\Omega$  mais sur l'ensemble des sous-ensembles  $\mathcal{P}(\Omega) = \{A_1, \dots, A_Q\}$ , avec  $Q = 2^K$ .

**Fonction de masses** Une fonction de masse élémentaire  $m$  est une fonction de  $\mathcal{P}(\Omega)$  dans  $\mathbb{R}^+$  vérifiant

$$\begin{cases} m(\emptyset) = 0 \\ \sum_{A \in \mathcal{P}(\Omega)} m(A) = 1 \end{cases} \quad (4.1.1)$$

Celle-ci représente la partie du degré de croyance placée sur une proposition  $A \subset \Omega$ . Nous pouvons alors affecter une masse aux hypothèses singletons, mais également aux hypothèses composites, ce qui a pour avantage de ne pas introduire d'information arbitraire pour forcer une répartition de masses sur des hypothèses qu'un capteur serait incapable de différencier. Nous obtenons alors une modélisation plus flexible permettant de représenter de multiples situations comme l'incomplétude, l'ambiguïté ou encore l'ignorance [57]. Nous donnons quelques exemples de jeux de masses :

- *Ignorance totale* : elle est représentée par le jeu de masses suivant :  $m(\Omega) = 1$  et  $m(A) = 0, \forall A \subset \Omega, A \neq \Omega$ . Ainsi, toute la masse est concentrée sur  $\Omega$ . Le discernement est alors impossible entre toutes les hypothèses de  $\Omega$ .
- *Certitude* : le jeu de masses correspondant est le suivant  $\exists H_i \in \Omega / m(H_i) = 1$  et  $\forall A \subset \Omega, A \neq H_i, m(A) = 0$ . Toute la masse est concentrée sur l'hypothèse singleton  $H_i$  ce qui se traduit par la certitude de la réalisation de l'hypothèse  $H_i$ .
- *Jeu de masses consonant* : les éléments focaux  $y$  sont emboîtés les uns dans les autres.

*Exemple 4.1.1.* Considérons le problème de la segmentation d'une image optique satellite ou aérienne en 2 classes  $\Omega = \{E, F\}$  "eau" et "forêt". La connaissance a priori est modélisée par  $m_1$ , qui est simplement une probabilité sur  $\Omega$ . Pour une observation  $y_n$ , qui est un nombre réel, sur un pixel  $n$ , nous avons trois possibilités "forêt", "eau", ou "nuage". La présence possible de nuages peut alors être modélisée par une mesure de probabilité  $m_2$  sur  $\mathcal{P}(\Omega) = \{\emptyset, E, F, \Omega\}$ , qui est une fonction de masse, définie par  $m_2(E) \propto f_E(y)$ ,  $m_2(F) \propto f_F(y)$  et  $m_2(\Omega) \propto f_\Omega(y)$  où  $f_\omega(y) = p(y|x = \omega), \forall \omega \in \mathcal{P}(\Omega)$ . Ainsi,  $m_2(\Omega)$  modélise l'ignorance attachée au fait qu'on ne peut pas voir au travers des nuages.

Nous pouvons voir que lorsque  $m$  est nulle en dehors des singletons, elle peut être assimilée à une probabilité et ainsi, elle apparaît comme une "généralisation" de ces dernières. Les fonctions de masses sont les seules que nous utiliserons par la suite. Toutefois, nous mentionnons par la suite deux autres représentations équivalentes de l'incertitude liée à une fonction de masse appelées fonctions de *crédibilité* et de *plausibilité*.

**Crédibilité, plausibilité** A partir d'une fonction de masse élémentaire  $m$ , on peut définir les fonctions de crédibilité  $Bel(A) = \sum_{B \subset A} m(B)$  et de plausibilité  $Pls(A) = \sum_{A \cap B \neq \emptyset} m(B)$ . La crédibilité représente la croyance totale placée sur une proposition  $A$ , la plausibilité

représentant la masse totale de croyance que l'on pourrait potentiellement transférer sur cette proposition. Ces deux fonctions représentent respectivement la valeur minimum et maximum de l'incertitude sur une proposition  $A$ . La longueur de l'intervalle  $[Bel(A), Pls(A)]$ , appelé intervalle de croyance, donne alors directement une mesure de l'imprécision sur l'incertitude sur une proposition  $A$ . Notons que la probabilité apparaît bien comme un cas particulier ; en effet, lorsqu'une masse est nulle partout sauf, éventuellement, sur les singletons, la crédibilité est alors égale à la plausibilité et c'est une probabilité. Une telle masse sera appelée dans la suite "bayésienne".

**Fusion de Dempster Shafer** Nous pouvons imaginer qu'une masse modélise l'information fournie par une "source", qui peut être, à titre d'exemple, un capteur en imagerie. Lorsque nous disposons de plusieurs masses  $m_1, \dots, m_l$  fournies par plusieurs sources, il se pose alors le problème de fusion les informations qu'elles modélisent. Une des possibilités que nous retenons pour la suite est la règle de combinaison de Dempster que nous appellerons fusion de Dempster Shafer par la suite (fusionDS), dont la règle est la suivante :

$$\begin{aligned} m(A) &= (m_1 \oplus m_2 \oplus \dots \oplus m_l)(A) \\ &\propto \sum_{B_1 \cap \dots \cap B_l = A} m_1(B_1) m_2(B_2) \dots m_l(B_l) \end{aligned} \quad (4.1.2)$$

Une des propriétés importantes de la fusionDS est que lorsque au moins l'une des masses impliquées dans la combinaison est bayésienne, le résultat de la combinaison est une masse bayésienne. Cette propriété permet de généraliser le calcul de la loi a posteriori dans le paragraphe suivant.

**Généralisation de la loi a posteriori** Soit  $p(x)$ ,  $x \in \Omega$ , la masse bayésienne modélisant l'information *a priori* et  $q(x) \propto p(y|x)$ ,  $x \in \Omega$ , ( $y \in \mathbb{R}$  est fixé), la masse bayésienne modélisant l'information issue de l'observation  $y \in \mathbb{R}$ . Le calcul de la loi *a posteriori* est un cas particulier de fusionDS. En effet, nous avons

$$(m \oplus q)(x) = \frac{p(x)p(y|x)}{\sum_{x' \in \Omega} p(x')p(y|x')} = p(x|y) \quad (4.1.3)$$

Nous pouvons alors généraliser le calcul de la loi *a posteriori* en remplaçant la masse bayésienne *a priori* par une masse évidentielle  $m(u)$ ,  $u \in \mathcal{P}(\Omega)$ . En effet, l'une des masses étant bayésienne (masse relative aux observations), le résultat de la combinaison de Dempster est une probabilité. Plus précisément, la combinaison de Dempster des masses  $m(u)$  et  $q(x)$

s'écrit (la somme  $\sum_{x \in u}$  est prise sur les  $u$  tels que  $x \in u$ )

$$(m \oplus q)(x) = \frac{\sum_{x \in u} m(u)q(x)}{\sum_{x' \in \Omega} \sum_{x' \in u'} m(u')q(x')} \quad (4.1.4)$$

*Exemple 4.1.2.* Reprenons l'exemple 4.1.1 : la fusion DS de  $m_1$  avec  $m_2$  donne alors une probabilité sur  $\Omega$ , qui généralise la probabilité *a posteriori* (que l'on retrouve lorsqu'il n'y a plus de nuages), et peut être utilisée pour effectuer une classification bayésienne. Un exemple d'une telle situation avec une représentation visuelle est traitée dans [7].

### 4.1.1 Affaiblissement

Lorsqu'une probabilité sur  $\Omega$  est connue avec imprécision, cette dernière peut être modélisée par une opération dite d'affaiblissement : on remplace la probabilité non fiable  $p(x)$  par la masse  $m(u)$  définie par :

$$m(u = \{\omega_i\}) = \alpha p(x = \{\omega_i\}), \quad m(\Omega) = 1 - \alpha \quad (4.1.5)$$

$\alpha$  étant le coefficient d'affaiblissement (ou de fiabilité) prenant ses valeurs dans l'intervalle  $[0, 1]$ . A titre d'exemple, la technique d'affaiblissement permet de prendre en compte la fiabilité relative des sources d'informations (plus le coefficient correspondant à une source est proche de 1, plus elle est considérée comme fiable). Inversement, une source non fiable aura une valeur  $\alpha$  proche de 0. L'opération d'affaiblissement a pour effet de pondérer l'importance d'une source d'information lors de sa combinaison avec d'autres sources.

*Exemple 4.1.3.* Revenons à l'exemple 4.1.1 : la connaissance a priori sur la région où a été prise l'image peut être modélisée par la masse bayésienne a priori  $(p(E), p(F))$ . La masse modélisant l'information issue de l'observation  $y \in \mathbb{R}$  est la masse bayésienne définie par

$$q(E) = \frac{f_E(y)}{f_E(y) + f_F(y)} \text{ et } q(F) = \frac{f_F(y)}{f_E(y) + f_F(y)} \quad (4.1.6)$$

En combinant ces deux masses par la règle de combinaison de Dempster, on obtient la masse bayésienne  $m' = m \oplus q$ , qui correspond dans ce cas à la probabilité *a posteriori* :

$$\begin{aligned} m'(E) &= \frac{p(E)f_E(y)}{\sum_{x' \in \{E, F\}} p(x')f_{x'}(y)} \\ m'(F) &= \frac{p(F)f_F(y)}{\sum_{x' \in \{E, F\}} p(x')f_{x'}(y)} \end{aligned} \quad (4.1.7)$$

Les deux fonctions de masses étant bayésiennes, on retrouve la formulation de la règle de Bayes.

Nous souhaitons à présent modéliser la fiabilité de la source d'information a priori. Pour cela, nous introduisons le coefficient  $\alpha$  et nous remplaçons la masse bayésienne a priori par

le nouveau jeu de masses  $m(E) = \alpha p(E)$ ,  $m(F) = \alpha p(F)$  et  $m(\Omega) = 1 - \alpha$ . Après calcul, la masse bayésienne  $m' = m \oplus q$  généralisant la loi a posteriori s'écrit :

$$\begin{aligned} m'(E) &= \frac{[\alpha p(E) + (1-\alpha)]f_E(y)}{\sum_{x' \in \{E, F\}} [\alpha p(x') + (1-\alpha)]f_{x'}(y)} \\ m'(F) &= \frac{[\alpha p(F) + (1-\alpha)]f_F(y)}{\sum_{x' \in \{E, F\}} [\alpha p(x') + (1-\alpha)]f_{x'}(y)} \end{aligned} \quad (4.1.8)$$

Cette masse peut alors être utilisée pour effectuer une classification bayésienne : en adoptant la fonction de coût  $L_{0/1}$ , on obtient la stratégie bayésienne suivante

$$\hat{s}_B(y) = \begin{cases} F & \text{si } f_F(y)/f_E(y) \geq \gamma \\ E & \text{si } f_E(y)/f_F(y) \leq \gamma \end{cases} \quad (4.1.9)$$

$\gamma$  étant la valeur du seuil de décision donnée par

$$\gamma = \frac{\alpha p(E) + (1 - \alpha)}{\alpha p(F) + (1 - \alpha)} \quad (4.1.10)$$

Sur la figure 4.1.1, nous représentons la valeur du seuil de décision  $\gamma$  en fonction du coefficient  $\alpha$  pour différentes valeurs du seuil de décision avant affaiblissement  $\gamma_b = p(E)/p(F)$ .

- Lorsque  $\alpha = 1$ , la source est fiable et la valeur du seuil de décision n'est pas modifiée.
- Lorsque  $\alpha$  diminue, la source devient de moins en moins fiable et la valeur du seuil de décision tend vers 1. L'importance de la source d'information *a priori* dans la fusion diminue.
- Lorsque  $\alpha = 0$ , la source n'est plus fiable et la valeur du seuil de décision est égale à 1. La source d'information *a priori* n'a donc plus d'influence sur le résultat de la classification.

L'opération d'affaiblissement a donc pour effet de pondérer l'importance d'une source d'information lorsque cette dernière n'est pas entièrement fiable. L'utilisation de ce fait dans un cadre markovien est le point crucial de ce chapitre. On souhaite en effet améliorer le résultat de la segmentation lorsque la loi *a priori* est mal connue en remplaçant cette dernière par une masse. On peut montrer que c'est le cas pour une séquence  $Z_{1:N} = (X_{1:N}, Y_{1:N})$  de v.a indépendantes, chaque  $X_n$  prenant ses valeurs dans  $\mathcal{X} = \Omega = \{E, F\}$ , et chaque  $Y_n$  prenant ses valeurs dans l'intervalle  $[0, 1]$ . Nous supposons que  $p_n = p(x_n = E)$  dépend de  $n$ , mais que les deux densités  $f_E(y) = 2(1 - y)$ ,  $f_F(y) = 2y$ , ne dépendent pas de  $n$ . Lorsque l'on utilise le vrai paramètre  $p_n$ , la stratégie bayésienne  $\hat{s}_B(y)$  correspondant à la fonction de perte  $L_{0/1}$  est donnée par

$$\hat{s}_B(y) = \begin{cases} F & \text{si } y \geq p_n \\ E & \text{si } y \leq p_n \end{cases} \quad (4.1.11)$$

ce qui nous donne la probabilité d'erreur<sup>1</sup>  $ERR(p_n) = p_n(1 - p_n)$ . Lorsque l'on utilise une

---

<sup>1</sup>En effet,  $ERR(p_n) = (1 - p_n) \int_0^{p_n} f_2(y) dy + p_n \int_{p_n}^1 f_1(y) dy = p_n(1 - p_n)$

valeur erronée  $r$  à la place de  $p_n$ , la probabilité d'erreur devient<sup>2</sup>  $ERR(p_n, r) = p_n(1 - p_n) + (p_n - r)^2$ .

Finalement, en tenant compte de ce qui va être fait dans le contexte markovien par la suite, nous remplaçons la valeur erronée  $r = p(x_n = E)$ ,  $1 - r = p(x_n = F)$  par une fonction de masse "affaiblie"  $m(E) = r - t$ ,  $m(F) = 1 - r - t$ ,  $m(\Omega) = 2t$ . La stratégie bayésienne  $\hat{s}'_B(y)$  est alors donnée par<sup>3</sup> :

$$\hat{s}'_B(y) = \begin{cases} F & \text{si } y \leq \frac{r+t}{1+2t} \\ E & \text{si } y \geq \frac{r+t}{1+2t} \end{cases} \quad (4.1.12)$$

ce qui nous mène à la probabilité d'erreur<sup>4</sup>  $ERR(p_n, r, t) = p_n(1 - p_n) + (p_n - \frac{r+t}{1+2t})^2$ .

Le problème est de savoir si un  $t > 0$  existe de telle manière que  $ERR(p_n, r, t) < ERR(p_n, r)$  : en d'autres termes, est-il possible de décroître la probabilité d'erreur en introduisant une fonction de masse  $m$  lorsque l'on a une valeur erronée  $r$  ? La réponse est positive dans le contexte suivant. Etant donné que  $p_{1:N}$  ne sont pas connues et peuvent varier avec  $n$ , nous supposons qu'elles varient avec  $n$  et sont la réalisation d'une v.a  $W$ , avec  $E[W] = \frac{1}{2}$ . Le problème est de voir si l'espérance de  $ERR[W, r, t]$  décroît lorsque l'on utilise une fonction de masse  $m$  à la place de  $r$  ou, en d'autres termes, lorsque  $t$  commence à partir de 0. Un calcul classique<sup>5</sup> donne  $\frac{d}{dt}[ERR(W, r, t)](0) = -(1 - 2r)^2$  ce qui montre que l'erreur moyenne décroît et donc, que pour une grande valeur de  $N$ , l'erreur décroît également lorsque l'on utilise  $m$  au lieu de  $r$ .

## 4.2 Introduction de l'information spatiale

La méthode que nous allons présenter consiste à remplacer la loi *a priori* par une masse évidentielle markovienne, ce qui permettra, d'après ce que nous avons vu précédemment, de gérer l'insuffisance de la connaissance de la loi a priori non stationnaire en introduisant un affaiblissement. Dans la suite, l'ensemble  $\mathcal{X}$  correspond au cadre de discernement  $\Omega$ , l'ensemble  $\mathcal{U}$  correspond à  $\mathcal{P}(\Omega)$  et  $\mathcal{Y}$  correspond à l'ensemble des réels  $\mathbb{R}$ .

<sup>2</sup>En effet,  $ERR(p_n, r) = (1 - p_n) \int_0^r f_2(y)dy + p_n \int_r^1 f_1(y)dy = p_n(1 - p_n) + (p_n - r)^2$ .

<sup>3</sup>En effet, étant donnée que  $(m \oplus q)(E) \propto 2(r + t)(1 - y)$  et  $(m \oplus q)(F) \propto 2(1 - r + t)y$ ,  $(m \oplus q)(E) \geq (m \oplus q)(F)$  signifie que  $(r + t)(1 - y) \geq (1 - r + t)y$ , ou encore  $y \leq \frac{r+t}{1+2t}$

<sup>4</sup>Le calcul est analogue à celui de  $ERR[p_n, r]$ , avec  $\frac{r+t}{1+2t}$  au lieu de  $r$ .

<sup>5</sup>En supposant que  $\frac{d}{dt}(E[ERR(p_n, r, t)]) = E[\frac{d}{dt}(ERR(p_n, r, t))]$ , nous avons  $\frac{d}{dt}[ERR(p_n, r, t)] = -2[p_n - \frac{(r+t)}{(1+2t)}][\frac{(1-2r)}{(1+2t)^2}]$ , et ainsi  $\frac{d}{dt}[ERR(p_n, r, t)](0) = -2(p_n - r)(1 - 2r)$ . En prenant l'espérance, nous obtenons  $\frac{d}{dt}[ERR(W, r, t)](0) = -2(E[W] - r)(1 - 2r) = -2(\frac{1}{2} - r)(1 - 2r) = -(1 - 2r)^2$

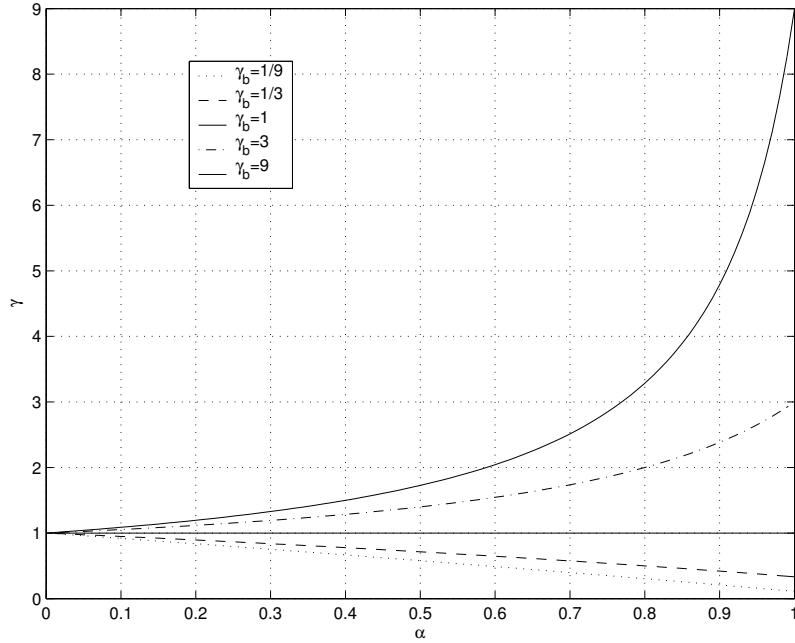


FIG. 4.1.1 – Effet du coefficient d'affaiblissement  $\alpha$  sur la valeur du seuil de décision.

#### 4.2.1 Chaînes de Markov évidentielles cachées

Une fonction de masse  $m$  définie sur  $\mathcal{P}(\Omega^N)$  est une chaîne de Markov évidentielle (CME) si elle est nulle en dehors de  $[\mathcal{P}(\Omega)]^N$  et si :

$$m(u) = m(u_1) \prod_{n=2}^N m(u_n|u_{n-1}) \quad (4.2.1)$$

où l'écriture  $m(u_n|u_{n-1})$  signifie que l'on associe à chaque  $u_n \subset \mathcal{P}(\Omega)$  une masse  $m(\cdot|u_{n-1})$ . Une CM classique est ainsi un cas particulier de CME, dans laquelle  $m(u_1)$  et toutes les  $m(u_n|u_{n-1})$  sont nulles en dehors des singletons. La fusion DS de  $m$  avec une loi de probabilité donnée par (4.2.2) est alors une probabilité qui généralise la loi *a posteriori*.

$$q(x) = \prod_{n=1}^N q(x_n) \text{ avec } q(x_n) = \frac{p(y_n|x_n)}{\sum_{x'_n \in \Omega} p(y_n|x'_n)} \quad (4.2.2)$$

Cette fusion s'écrit

$$(m \oplus q)(x) = \frac{\sum_{u \in [\mathcal{P}(\Omega)]^N} \mathbb{1}_{[x \in u]} m(u) q(x)}{\sum_{x' \in \Omega^N} \sum_{u' \in [\mathcal{P}(\Omega)]^N} \mathbb{1}_{[x' \in u']} m(u') q(x')} \quad (4.2.3)$$

Afin de montrer que  $m \oplus q$  est une marginale d'une CM, et que  $(m \oplus q)(x) = \sum_{u \in \mathcal{U}} p(x, u|y)$  énonçons le résultat classique suivant :

**Lemme 4.2.1.** *Soit  $V = \{V_n\}_{n \in \mathcal{N}}$  un processus stochastique, chaque  $V_n$  prenant ses valeurs dans un même ensemble fini. Alors  $V$  est une CM si et seulement si il existe  $N - 1$  fonctions positives sommables  $r_1, \dots, r_{N-1}$ , telles que la loi de  $V$  est proportionnelle à  $r_1(v_1, v_2) \dots r_{N-1}(v_{N-1}, v_N)$ . De plus, la loi de  $V$  est alors donnée par  $p(v_1) = \beta_1(v_1) / \sum_{v'} \beta_1(v')$  et les transitions  $p(v_n|v_{n-1}) = r_{n-1}(v_{n-1}, v_n) \beta_n(v_n) / \beta_{n-1}(v_{n-1})$ , où les fonctions  $\beta_1, \dots, \beta_N$ , sont définies par  $\beta_N(v_N) = 1$ , et  $\beta_{n-1}(v_{n-1}) = \sum_{v_n} r_{n-1}(v_{n-1}, v_n) \beta_n(v_n)$ .*

En considérant le processus  $V = (X, U)$ , avec  $V_n = (X_n, U_n)$ , posons

$$\begin{cases} r_1(v_1, v_2) = \mathbb{1}_{[x_1 \in u_1]} \mathbb{1}_{[x_2 \in u_2]} m(u_1, u_2) q(x_1) q(x_2) \\ r_{n-1}(v_{n-1}, v_n) = \mathbb{1}_{[x_{n-1} \in u_{n-1}]} \mathbb{1}_{[x_n \in u_n]} m(u_{n-1}, u_n) q(x_n) q(x_{n-1}) \text{ pour } 3 \leq n \leq N \end{cases} \quad (4.2.4)$$

La quantité  $\mathbb{1}_{[x \in u]} m(u) q(x)$  est de la forme  $r_1(v_1, v_2) \dots r_{N-1}(v_{N-1}, v_N)$  et donc (4.2.5) est, en vertu du Lemme 4.2.1, une CM avec les transitions calculables.

$$p(v) = \frac{\mathbb{1}_{[x \in u]} m(u) q(x)}{\sum_{x' \in \Omega^N} \sum_{u' \in [P(\Omega)]^N} \mathbb{1}_{[x' \in u']} m(u') q(x')} \quad (4.2.5)$$

Par ailleurs,  $(m \oplus q)(x) = \sum_{u \in [P(\Omega)]^N} p(x, u)$  d'après (4.2.3), ce qui montre que  $m \oplus q$  est une marginale de cette même CM. Sachant que  $(m \oplus q)(x)$  généralise la loi *a posteriori* de  $X$  classique ( $y$  figure dans  $(m \oplus q)(x)$ ), on a  $p(x_n|y) = (m \oplus q)(x_n) = \sum_{u_n \in P(\Omega)} p(x_n, u_n)$  qui est ainsi calculable, ce qui permet d'estimer une réalisation du processus caché au sens du MPM. Nous appellerons Chaîne de Markov évidentielle caché à bruit indépendant (CMEC-BI) la CMT correspondant à ce modèle particulier.

Concrètement, les quantités  $p(x_n|y)$  et  $p(v_n, v_{n+1}|y)$  se calculent à partir des quantités (4.2.4) (dans lesquels  $y$  figure implicitement, sachant que  $q(x_n) \propto p(y_n|x_n)$ ) de la manière suivante :

- On calcule la suite  $\beta_1, \dots, \beta_N$ , et les transitions  $p(v_n|v_{n-1})$  par les formules du Lemme 4.2.1.
- Puis, les marginales  $p(v_n)$  sont calculées récursivement par

$$\begin{cases} p(v_1) = \beta_1(v_1) \\ p(v_n) = \sum_{v_{n-1}} p(v_n|v_{n-1}) p(v_{n-1}) \end{cases} \quad (4.2.6)$$

ce qui donne  $p(x_n|y) = \sum_{u_n \in P(\Omega)} p(x_n, u_n)$  d'une part, et  $p(v_n, v_{n+1}|y) = p(v_n|v_{n-1}) p(v_{n-1})$ , qui seront utilisées dans la suite pour l'estimation des paramètres, d'autre part.



**Estimation des paramètres** Par ailleurs, considérer  $V = (X, U)$  comme étant le processus caché dans  $T = (X, U, Y)$  (qui devient donc une CMC-BI classique) permet d'appliquer l'une des méthodes classiques d'estimation des paramètres, comme EM, SEM, ou ICE.

Nous proposons la méthode d'estimation des paramètres suivante. La chaîne  $T = (X, U, Y)$  étant stationnaire, sa loi est donnée par  $p(t_n, t_{n+1})$ , qui ne dépend pas de  $n$ . Notons que c'est une loi inhabituelle dans les chaînes de Markov cachées dans le sens où  $p(y_n|x_n, u_n) = p(y_n|x_n)$ , ce qui signifie qu'il y a moins de densités de "bruit" que des états du processus caché  $V = (X, U)$ . Pour  $K$  classes  $\Omega = \{\omega_1, \dots, \omega_K\}$ , nous devons estimer  $K$  moyennes  $\mu_1, \dots, \mu_K$  et  $K$  variances  $\sigma_1^2, \dots, \sigma_K^2$  de  $K$  densités gaussiennes  $p(y_n|x_n = \omega_1), \dots, p(y_n|x_n = \omega_K)$ , ainsi que les  $(K + 1) \times (K + 1) - 1$  paramètres  $m_{ij} = p(u_1 = A_i, u_2 = A_j)$ , qui définissent une CME sur  $\mathcal{U} = \{\omega_1, \dots, \omega_K, \Omega\}$  (étant donné qu'on utilise ici une méthode d'affaiblissement).

En posant  $\gamma(x_n, u_n) = p(v_n|y)$  et  $\xi = (x_n, u_n, x_{n+1}, u_{n+1}) = p(v_n, v_{n+1}|y)$  (nous omettons  $y$  afin de simplifier les notations), le déroulement de l'algorithme EM utilisé est le suivant

- (i) initialisation :  $\theta^0 = (m_{ij}^0, \mu_k^0, (\sigma_k^0)^2)_{0 \leq i, j \leq K+1, 1 \leq k \leq K}$
- (ii) pour tout  $q \in \mathbb{N}$ ,  $\theta^{[q+1]}$  est calculé à partir de  $(y_1, \dots, y_N)$  et  $\theta^q$  par

$$m_{ij}^{[q+1]} = \frac{1}{N-1} \sum_{n=1}^{N-1} \sum_{(x_n, x_{n+1}) \in \Omega^2} \xi^{[q]}(x_n, u_n = A_i, x_{n+1}, u_{n+1} = A_j) \quad (4.2.7)$$

$$\mu_k^{[q+1]} = \frac{1}{N} \sum_{n=1}^N \sum_{u_n \in \mathcal{U}} \gamma^{[q]}(x_n = k, u_n) y_n \quad (4.2.8)$$

$$(\sigma_k^{[q+1]})^2 = \frac{1}{N} \sum_{n=1}^N \sum_{u_n \in \mathcal{U}} \gamma^{[q]}(x_n = k, u_n) \times (y_n - \mu_k^{[q+1]})^2 \quad (4.2.9)$$

Notons que la méthode utilisée donne l'algorithme EM classique lorsque la CME  $U$  ne charge que les singletons (on a alors  $U = X$  et la CMEC-BI devient une CMC-BI classique).

Revenons au problème de la segmentation non supervisée des chaînes de Markov cachées fortement non stationnaires. Nous avons deux manières de traiter le problème et l'objectif des résultats présentés à la Figure 4.2.1 est de montrer l'intérêt de la seconde :

1. on applique la méthode classique fondée sur le modèle de CMC-BI. Les paramètres de la CMC-BI  $Z = (X, Y)$ , supposée stationnaire, sont estimés par EM et la segmentation est effectuée, sur la base des paramètres estimés, par MPM ;
2. on considère la CMEC-BI  $T = (X, U, Y)$ , où la chaîne de Markov cachée stationnaire est  $V = (X, U)$ , comme décrit ci-dessus. Les paramètres sont estimés par EM, les marginales  $p(v_n|y) = p(x_n, u_n|y)$  sont calculées classiquement à partir des paramètres estimés en utilisant l'algorithme de calcul des probabilités marginales *a posteriori*. Ensuite, les marginales  $p(x_n|y)$  sont calculées à partir des  $p(x_n, u_n|y)$  en sommant sur

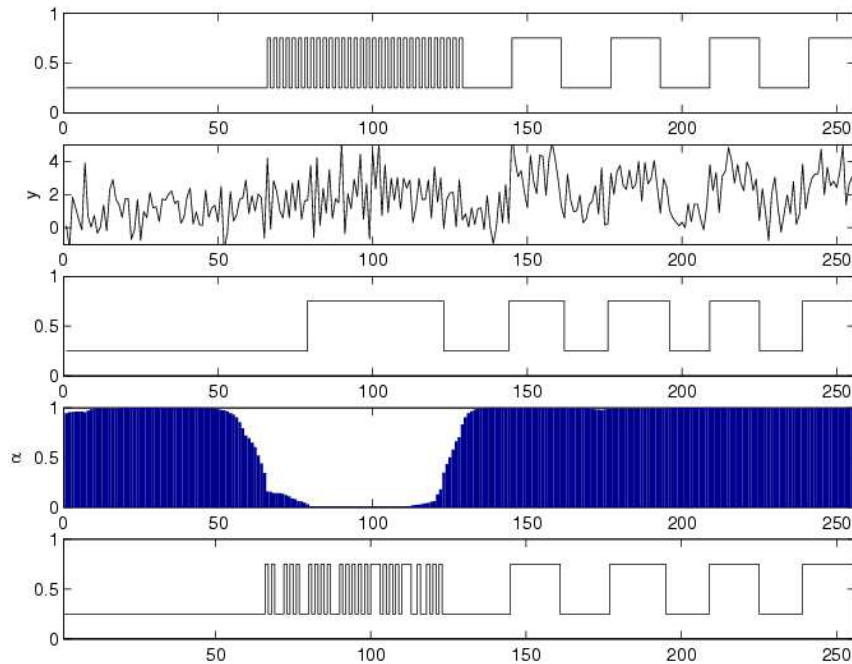


FIG. 4.2.1 – De haut en bas : (a) processus caché, (b) processus observé, (c) résultat de la segmentation non supervisée de la CMC, (d) variation du coefficient d'affaiblissement  $\alpha$  conditionnellement aux observations, (e) résultat de la segmentation non-supervisée de la CMEC-BI

$u_n$ , ce qui permet l'application de la méthode MPM.

La réalisation  $X = x$  du processus caché à deux classes, représentée sur le graphique (a) de la Figure 4.2.1, est fortement non stationnaire. Le bruitage classique consiste à considérer deux densités gaussiennes de variance égale à 1, et de moyennes 1 et 3 respectivement, ce qui donne une réalisation  $Y = y$  représentée sur le graphique (b). La réalisation du processus caché estimée par la première méthode est présentée sur le graphique (c) ; on constate que cette méthode est incapable d'estimer correctement l'endroit où les probabilités de transition d'un état à un état différent sont très fortes (les  $x_n$  pour  $n$  variant de 70 à 130). Par contre, la deuxième méthode, dont la trajectoire estimée est présentée sur le graphique (e), présente une estimation convenable de cette partie du processus caché. Afin de mieux cerner ce phénomène nous avons présenté sur le graphique (d) les coefficients d'affaiblissement appliqués, pour chaque point  $n$ , à la loi marginale a posteriori correspondante. Il s'avère que ces marginales ne sont pratiquement pas modifiées aux points où la première méthode donne des bons résultats (les points en dehors des  $n$  variant de 70 à 130) ; par contre, l'affaiblissement est maximal pour une bonne partie des points où la première méthode faillit (les  $n$  variant de 70 à 130). Ce phénomène, dont l'explication mathématiquement rigoureuse ne nous apparaît pas avec évidence, est intéressant pour les applications pratiques. En particulier, il réapparaît dans les segmentations non supervisées d'images présentées à la fin de ce chapitre.

**Remarque** Le modèle de la CMEC-BI, où la CME  $m(u)$  est fusionnée avec la loi de probabilité  $q(x)$  peut être généralisé au cas de plusieurs capteurs indépendants  $c_{1:r}$  de la façon suivante. Si tous les capteurs produisent des lois de probabilités, la généralisation est immédiate en remplaçant  $q(x_n) \propto p(y_n|x_n)$  par  $q(x_n) \propto p(y_n^1|x_n) \dots p(y_n^r|x_n)$ , qui est une fusion DS. Si certains des capteurs  $c_1, \dots, c_s$  (par exemple les  $s$  premiers) produisent des masses (de telles situations où, par exemple, un capteur optique est confronté à la présence de nuages, sont étudiées dans [57, 7]), ces dernières sont fusionnées avec les  $r - s$  probabilités données par  $p(y_n^{s+1}|x_n), \dots, p(y_n^r|x_n)$  avec, pour résultat, une probabilité  $q(x_n)$ . Cette dernière est alors fusionnée avec  $m(u)$  et la suite de la démarche est identique à celle correspondant au cas d'un seul capteur probabiliste traité dans ce chapitre.

### 4.3 Fusion DS dans CMCouple

Nous avons vu dans ce qui précède que lorsque dans une CMC-BI  $(X, Y)$  on remplace la loi *a priori* de  $X$  par une CME, la fusion de cette dernière avec une probabilité définie à partir des observations  $Y = y$  est formellement la loi de  $X$  conditionnelle à  $Y = y$  déduite d'une CMT  $T = (X, U, Y)$ , ce qui permet le calcul des probabilités marginales *a posteriori*  $p(x_n|y)$ . Sachant que les CMC-BI peuvent être étendues aux CMC, CMCouple, et CCPM, la question qui se pose est celle de savoir s'il est encore possible, dans ces extensions, de remplacer la loi *a priori* de  $X$  par une CME de manière à ce que le résultat de la combinaison de Dempster avec une probabilité définie par les observations  $Y = y$  étende, d'une part, la loi *a posteriori* de  $X$  et aboutisse, d'autre part, à un résultat exploitable comme dans le cas des CMC-BI. L'objet de cette section est de montrer que la réponse est positive. Pour cela, considérons une CMCouple  $Z = (X, Y)$ , avec  $p(z) = p(z_1)p(z_2|z_1) \dots p(z_N|z_{N-1}) = \frac{p(z_1, z_2) \dots p(z_{N-1}, z_N)}{p(z_2) \dots p(z_{N-1})}$ . Sachant que  $p(z_n, z_{n+1}) = p(x_n, x_{n+1})p(y_n, y_{n+1}|x_n, x_{n+1})$  et  $p(z_n) = p(x_n)p(y_n|x_n)$ , la loi de  $Z$  s'écrit également

$$p(z) = \underbrace{\frac{p(x_1, x_2) \dots p(x_{N-1}, x_N)}{p(x_1) \dots p(x_{N-1})}}_{a(x)} \times \underbrace{\frac{p(y_1, y_2|x_1, x_2) \dots p(y_{N-1}, y_N|x_{N-1}, x_N)}{p(y_2|x_2) \dots p(y_{N-1}|x_{N-1})}}_{b(x, y)} \quad (4.3.1)$$

L'écriture (4.3.1) est ainsi valable dans tout CMCouple, avec trois cas particuliers :

1. les  $p(y_{n:n+1}|x_{n:n+1})$  sont de la forme  $p(y_{n:n+1}|x_{n:n+1}) = p(y_n|x_n)p(y_{n+1}|x_{n+1})$  dans  $b(x, y)$ ; on retrouve le cas de CMC-BI traité dans la section précédente;
2. l'égalité  $p(y_n, y_{n+1}|x_n, x_{n+1}) = p(y_n|x_n)p(y_{n+1}|x_{n+1})$  n'a pas nécessairement lieu mais l'intégrale de  $b(x, y)$  par rapport à  $y$  vaut 1 : alors le modèle est une CMC et  $a(x)$  est la loi de  $X$ ;
3. l'intégrale de  $b(x, y)$  par rapport à  $y$  ne vaut pas 1 (ce qui revient à dire que  $a(x)$  n'est pas la loi de  $X$ , ou encore que  $X$  n'est pas de Markov et donc  $(X, Y)$  est une

CMCouple sans être une CMC). Notons que dans ce dernier cas  $p(x_n, x_{n+1})$  figurant dans (4.3.1) est bien la loi de  $(X_n, X_{n+1})$  pour tout  $1 \leq n \leq N - 1$ , mais  $a(x)$  n'est pas la loi de  $X$ .

Considérons le cas le plus général où la loi de  $Z = (X, Y)$  est donnée par (4.3.1). Ainsi  $a$  n'est plus nécessairement la loi de  $X$  et, de manière équivalente,  $b(x, y)$  n'est plus nécessairement la loi de  $Y$  conditionnelle à  $X = x$ . Cependant, en notant  $q$  la loi de probabilité obtenue à partir de  $b(x, y)$  par normalisation par rapport à  $x$  et à  $y$  fixés, on peut affirmer que la loi a posteriori  $p(x|y)$  est le résultat de combinaison de Dempster de  $a$  par  $q$ . En suivant la même démarche que celle appliquée au cas (1) dans la section précédente, on peut donc envisager de remplacer  $a$  par une chaîne évidentielle.

**Proposition 4.3.1.** *Soit  $Z$  une CMCouple et soient  $q$  la loi de probabilité définie par  $b(x, y)$  (après normalisation par rapport à  $x$  et  $y$  fixés) et  $m$  une CME définie par  $m(u) = \frac{m(u_1, u_2) \dots m(u_{N-1}, u_N)}{m(u_2) \dots m(u_{N-1})}$  ( $m$  étend la probabilité  $a$  de (4.3.1)).*

*Alors la fusion DS  $m \oplus q$ , qui généralise la loi a posteriori  $p(x|y)$  du cas classique où  $m = a$ , est la loi marginale de la CM  $p(x, u|y)$  définie par les fonction  $r_{n-1}(v_{n-1}, v_n) = r_{n-1}(x_{n-1}, u_{n-1}, x_n, u_n) = \mathbb{1}_{[x_{n-1} \in u_{n-1}]} \mathbb{1}_{[x_n \in u_n]} m(u_{n-1}, u_n) q_n(x_{n-1}, x_n)$  avec  $q_2(x_1, x_2) \propto p(y_1, y_2 | x_1, x_2)$  et  $q_n(x_{n-1}, x_n) \propto p(y_{n-1}, y_n | x_{n-1}, x_n) / p(y_{n-1} | x_{n-1})$  pour  $n \in \mathcal{N} \setminus \{2\}$  ( $y$  est fixé)*

Il en résulte que les transitions  $p(v_{n+1} | v_n, y)$  et les marginales  $p(v_n, v_{n+1} | y)$ ,  $p(v_n | y)$  sont calculables.

## 4.4 Segmentation non supervisée des images non-stationnaires

Dans cette section, nous comparons les résultats obtenus en considérant le modèle CMC-BI avec les résultats obtenus en considérant le modèle CMEC-BI. Nous montrons l'existence de situations pour laquelle la nouvelle méthode améliore de manière significative les résultats obtenus dans le cas classique. Nous présentons deux séries d'expérimentations. Dans la première, une CMC-BI non stationnaire est simulée. Dans la deuxième, nous considérons le problème de la segmentation d'image. Une image non stationnaire est segmentée par les méthodes classique et évidentielle, l'ensemble bidimensionnel des pixels étant transformé en ensemble mono-dimensionnel via un parcours de Hilbert Peano. Finalement, nous présentons les résultats de segmentation obtenus pour une image réelle.

### 4.4.1 Chaînes de Markov cachées non stationnaires

Nous considérons une CMC-BI  $Z$ , avec  $\Omega = \{1, 2\}$  et  $N = 1024$  et les trois matrices de transition suivantes

$$M_1 = \begin{bmatrix} 0.98 & 0.02 \\ 0.02 & 0.98 \end{bmatrix}, M_2 = \begin{bmatrix} 0.60 & 0.40 \\ 0.40 & 0.60 \end{bmatrix}, M_3 = \begin{bmatrix} 0.51 & 0.49 \\ 0.49 & 0.51 \end{bmatrix} \quad (4.4.1)$$

La CM  $X$  est non stationnaire de la manière suivante. Etant donné deux matrices de transitions  $M_1$  et  $M_2$ , et, pour  $s = 1, 2, \dots$ ,  $X^i = (X_{(i-1)s+1}, X_{(i-1)s+2}, \dots, X_{is})$ , la chaîne non stationnaire  $X = (X_1, \dots, X_N)$  vérifie alors :

1. La loi de  $X_1$  est  $(0.5, 0.5)$
2.  $M_1$  est la matrice de transition pour  $X^2, X^4, \dots$  et
3.  $M_2$  est la matrice de transition dans  $X^3, X^5, \dots$

Une réalisation  $X = x$  est simulée et  $Y = y$  est échantillonnée pour tout  $n \in \mathcal{N}$  par rapport à  $p(y_n|x_n)$ , où  $p(y_n|x_n = 1)$  est gaussienne avec une moyenne 0 et une variance 1, et  $p(y_n|x_n = 2)$  est gaussienne avec une moyenne 2 et une variance 1. la réalisation  $X = x$  est alors estimée par MPM à partir de  $Y = y$  de trois manières différentes :

1. La première segmentation est obtenue en utilisant les paramètres réels de la chaîne non stationnaire. Il s'agit d'un résultat de référence qui donne le taux d'erreur minimal  $\tau_{min}^{MPM}$ .
2. La deuxième segmentation est obtenue en utilisant les paramètres estimés par l'algorithme EM en considérant que  $(X, Y)$  est une CMC-BI (la chaîne cachée est supposée stationnaire). Le taux d'erreur est noté  $\tau_{HMC-EM}^{MPM}$ .
3. La troisième segmentation est obtenue en utilisant les paramètres estimés à partir de l'algorithme EM en considérant que  $(X, Y)$  est une CMEC-BI. Le taux d'erreur est noté  $\tau_{HEMC-DS}^{MPM}$  (Le CMEC-BI-EM est initialisé par les résultats de CMC-EM).

Nous présentons deux séries de résultats : La première, correspondant à  $M_2$  est donnée dans le tableau 4.1(gauche) et la deuxième correspondant à la même  $M_1$  et  $M_3$  à la place de  $M_2$ , est donnée dans le tableau 4.1(droite). D'après ces résultats, nous voyons que  $\tau_{HEMC-EM}^{MPM}$  est toujours inférieur à  $\tau_{HMC-EM}^{MPM}$  et, pour que pour certains  $s$ , le gain peut être important.

#### 4.4.2 Image non stationnaire bruitée

L'image  $128 \times 128$  de la figure 4.4.1 est bruitée par un bruit gaussien indépendant de variance 1 et de moyennes 0 et 2. L'ensemble bi-dimensionnel de pixels est converti, via un parcours de Hilbert-Peano, en un ensemble mono-dimensionnel  $1, 2, \dots, N = 128 \times 128$ . Les résultats présentés sur la figure 4.4.2 montrent que la segmentation MPM fondée sur les paramètres estimés par CMC-EM (c) ne permet pas de trouver les détails des ailes et de la queue de l'oiseau, dû à un trop fort taux de régularisation spatiale. Sur l'image (b), nous avons représenté la valeur de  $\{1 - p(u_n = \Omega|y)\}_{n \in \mathcal{N}}$  que nous appellerons coefficient

$s$	$\tau_{min}^{MPM}$	$\tau_{CMC-EM}^{MPM}$	$\tau_{CMEC-EM}^{MPM}$	$s$	$\tau_{min}^{MPM}$	$\tau_{CMC-EM}^{MPM}$	$\tau_{CMEC-EM}^{MPM}$
2	7.8	11.0	11.0	2	8.1	13.3	13.3
4	7.8	11.3	11.1	4	8.0	13.3	12.7
8	8.0	12.2	11.3	8	8.0	14.3	11.8
16	8.1	12.7	10.8	16	8.8	16.9	11.3
32	8.3	14.7	10.4	32	8.5	19.1	11.1
64	8.7	14.7	9.9	64	8.6	20.0	9.7
128	8.4	15.4	9.4	128	8.7	22.0	9.2
256	8.2	16.5	8.7	256	8.7	22.9	9.0
512	8.2	17.5	8.7	512	8.6	22.8	8.8

TAB. 4.1 – A gauche : Taux d’erreurs correspondant à  $M_1$ ,  $M_2$  et différentes valeurs de  $s$ . A droite : Taux d’erreur correspondant à  $M_1$ ,  $M_3$  et différentes valeurs de  $s$ .

CMC-BI			$N(0, 1)$	$N(2, 1)$
$p(\omega_1)$	0.28	$\mu$	-0.00	1.06
$p(\omega_2)$	0.72	$\sigma$	1.01	2.07
$p(x_1, x_2)$		$\omega_1$	$\omega_2$	
$\omega_1$		0.27	0.00	
$\omega_2$		0.00	0.71	
Paramètres de la CMC-BI estimés par EM				
CMEC-BI			$N(0, 1)$	$N(2, 1)$
$m(\{\omega_1\})$	0.02	$\mu$	-0.00	2.08
$m(\{\omega_2\})$	0.71	$\sigma$	0.92	1.01
$m(\{\omega_1, \omega_2\})$	0.26			
$m(u_1, u_2)$		$\{\omega_1\}$	$\{\omega_2\}$	$\{\omega_1, \omega_2\}$
$\{\omega_1\}$		0.02	0.00	0.00
$\{\omega_2\}$		0.00	0.70	0.00
$\{\omega_1\omega_2\}$		0.00	0.00	0.26
Paramètres de la CMEC-BI estimés par EM				

TAB. 4.2 – Paramètres estimés obtenus par EM dans le cas des CMC-BI et des CMEC-BI. Les densités gaussiennes réelles sont  $\mathcal{N}(0, 1)$  et  $\mathcal{N}(2, 1)$ .

d’affaiblissement *a posteriori* et qui est une manière approchée de visualiser le coefficient d’affaiblissement en fonction des zones de l’images. On constate que pour les zones comprenant des détails, celui-ci tend vers 0. Le résultat de la segmentation MPM basée sur le modèle CMEC-BI représenté sur l’image (d) montre bien l’amélioration obtenue. Les estimés des différents paramètres en considérant le modèle CMC puis le modèle CMEC sont présentés sur le tableau 4.2. Nous pouvons voir que CMEC-IN-EM est plus efficace dans l’estimation des paramètres du bruit. Un autre point est le fait que CMEC-BI-EM attribue une probabilité de 0.26 à  $\{\omega_1, \omega_2\} \times \{\omega_1, \omega_2\}$ , ce qui peut être vu, comme une mesure de non adaptabilité de la CMC stationnaire aux données. En fait, lorsque les données cachées conviennent à une CMC-BI, CMEC-BI-EM donne ce dernier modèle et ainsi, les probabilités attribuées à  $\{\omega_1\} \times \{\omega_1, \omega_2\}$ ,  $\{\omega_2\} \times \{\omega_1, \omega_2\}$  et  $\{\omega_1, \omega_2\} \times \{\omega_1, \omega_2\}$  tendent vers zero.

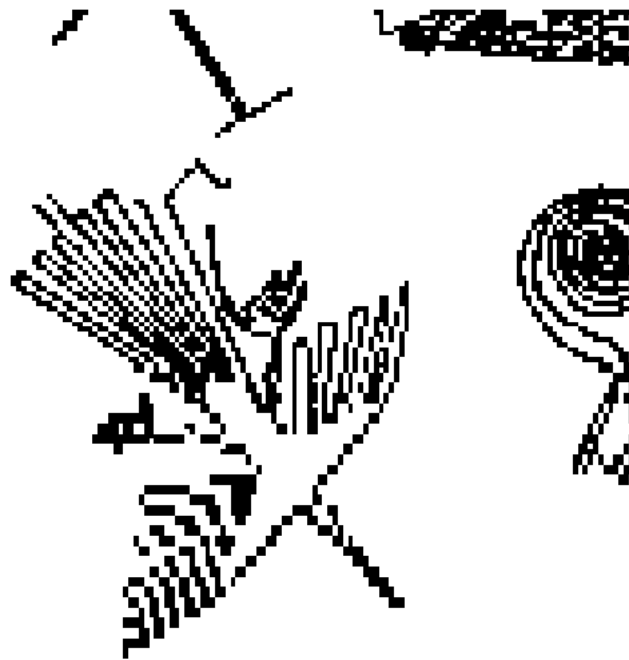


FIG. 4.4.1 – Image binaire Nazca

#### 4.4.3 Image réelle

Nous réeffectuons la même expérimentation à partir d'une image réelle. Les résultats sont présentés sur la figure 4.4.3. Il est difficile, en absence de la vérité terrain, de comparer quantitativement les résultats obtenus. Cependant, certains détails semblent mieux conservés par la nouvelle méthode. Notons qu'une façon plus rigoureuse de comparer les résultats serait de calculer le BIC comme au chapitre précédent. Par ailleurs, la carte des valeurs du coefficient d'affaiblissement *a posteriori* semble compatible avec les zones difficiles à segmenter.

### 4.5 Conclusion

Une des applications de la théorie des croyances consiste à introduire des fonctions de masse dans des modèles probabilistes lorsque ces derniers ne sont pas connus avec suffisamment de précision. Une telle situation se produit par exemple lorsque la loi *a priori* d'une chaîne de Markov cachée n'est pas stationnaire ; en effet, une telle loi peut être difficile à estimer avec précision. En remplaçant la loi *a priori* par une masse évidentielle markovienne (CME), nous avons montré qu'il est possible de gérer, de manière entièrement automatique, de fortes non stationnarités du processus caché. La nouvelle méthode proposée dans ce chapitre permet d'obtenir une segmentation plus fine en adaptant automatiquement, sur chaque site, l'affaiblissement. En particulier, son application à la segmentation non supervisée d'images permet de retrouver des détails fins, perdus lors des traitements classiques. Par

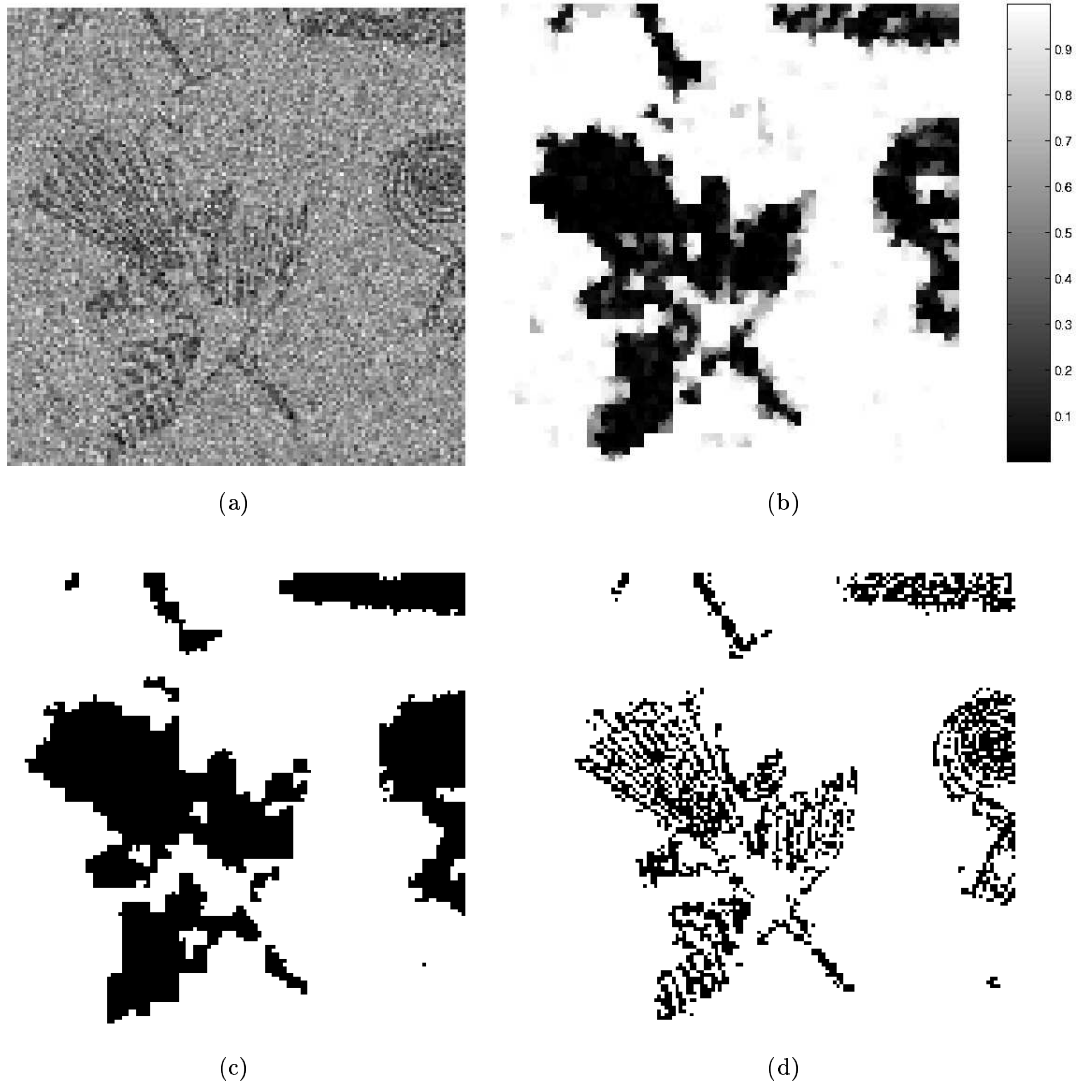


FIG. 4.4.2 – Segmentation non supervisées obtenues en utilisant les chaînes classiques et les chaînes évidentielles. (a) Image bruitée par deux gaussiennes  $\mathcal{N}(0, 1)$  et  $\mathcal{N}(2, 1)$ . (b) valeur du coefficient d'affaiblissement *a posteriori* . (c) segmentation non supervisée CMC-EM-MPM, taux d'erreurs :14,64%. (d) segmentation non supervisée CMEC-EM-MPM, taux d'erreurs : 5,05%



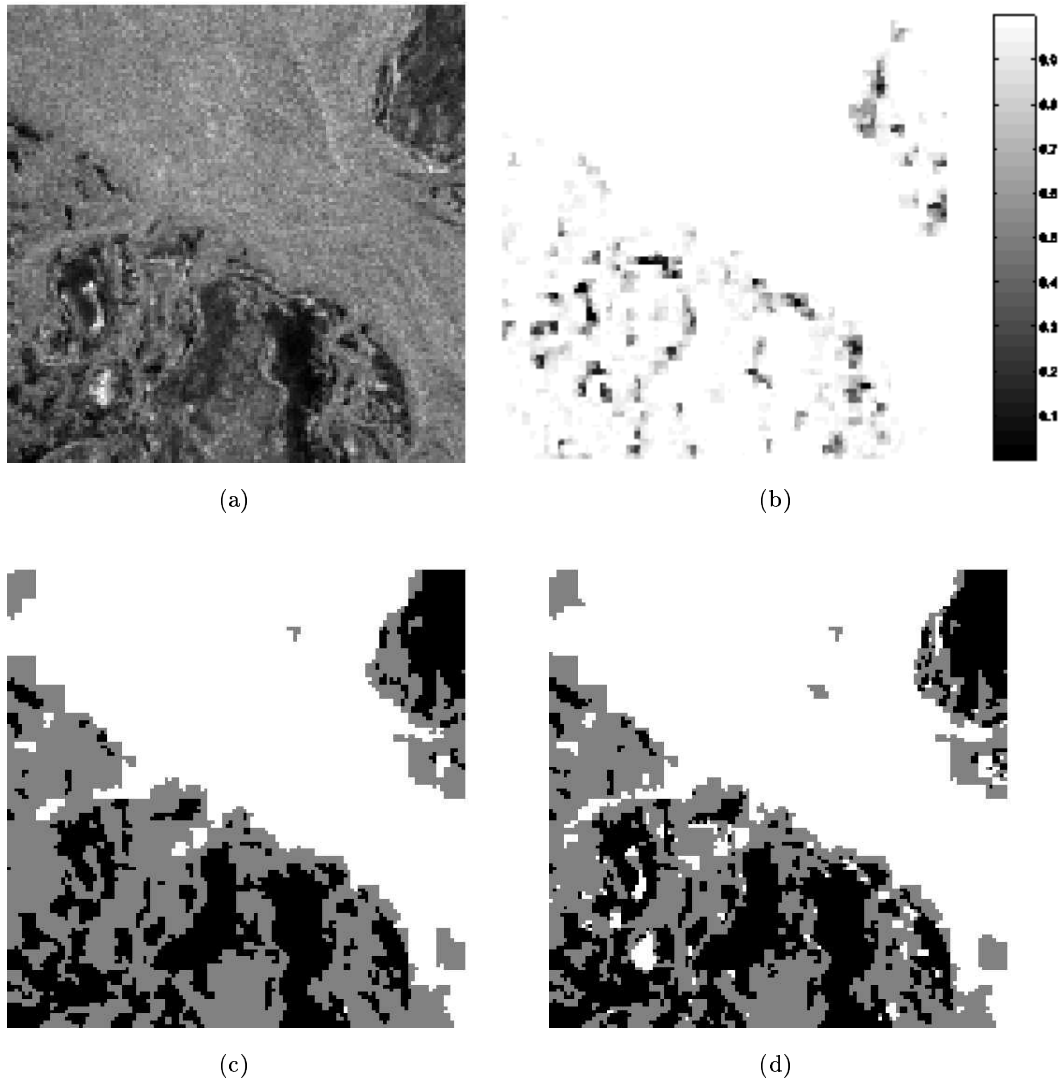


FIG. 4.4.3 – Image réelle et résultats de la segmentation à trois classes en considérant le modèle CMC-BI, pour le modèle CMEC-BI. (a) Image réelle, (b) Valeur du coefficient d'affaiblissement conditionnellement aux observations, (c) Segmentation non supervisée CMC-EM-MPM, (d) Segmentation non supervisée CMEC-EM-MPM.

rapport au modèle CMM-NSC-BI présenté au chapitre précédent, l'approche est très différente. Dans le premier cas, on considère différentes zones de stationnarités correspondant chacune à un état auxiliaire. Dans le modèle CMEC-BI, l'influence de la loi *a priori* et donc de l'information contextuelle est affaiblie par rapport à l'information issue des observations dans les zones présentant des détails fins. Ainsi, dans ces zones, les résultats de segmentation sont semblables à ceux obtenus en utilisant une méthode de segmentation aveugle<sup>6</sup>. Dans [109], l'auteur avait par ailleurs montré que pour des images présentant des détails fins, les méthodes de segmentation aveugles donnaient de meilleurs résultats que les méthodes markoviennes cachées. Le modèle CMEC-BI permet donc d'adapter automatiquement l'influence de l'information contextuelle en fonction des caractéristiques de l'image.

---

<sup>6</sup>i.e. aveugle du contexte, en considérant un modèle pour lequel les  $(Z_n)_{n \in \mathcal{N}}$  sont indépendantes.

## Chapitre 5

# Arbres de Markov cachés flous multispectraux

Nous avons considéré jusqu'à présent uniquement des modélisations monodimensionnelles de la loi de  $Z$ . Nous proposons dans ce chapitre d'étendre ces modèles au cas de signaux multirésolutions. D'autre part, nous nous intéressons à la segmentation statistique floue, bien adaptée aux caractéristiques de l'imagerie astronomique. En effet, les méthodes de segmentation statistique floue représentent une alternative aux techniques classiques dans le cas d'images contenant des structures diffuses. Les pixels  $y$  sont indexés au moyen d'ensembles discrets (classes thématiques homogènes) et/ou continus (régions intermédiaires). Ce type de classification, fondée sur la théorie des ensembles flous [121], fait correspondre à chaque pixel son degré d'appartenance aux régions thématiques. Cette approche permet de modéliser l'imprécision des données cachées, et l'incertitude sur la classification des pixels [14]. La segmentation floue a fait appel à des techniques fondées sur les champs [62, 108], puis récemment sur les chaînes de Markov [34, 24]. Les images astronomiques étant généralement de nature multirésolution et multispectrale, nous proposons dans ce chapitre un modèle d'arbres de Markov cachés flous multispectraux. Ainsi, nous traitons des données sous l'hypothèse de corrélation spectrale entre les bandes selon le modèle défini dans [107]. Les paramètres (lois d'attaches aux données supposées gaussiennes et paramètres de la distribution *a priori*) sont estimés au moyen de l'algorithme EM. L'étape de décision finale met en jeu le critère MPM [76] adapté au contexte du flou.

Nous étendons dans un premier temps les modèles monodimensionnels de la loi de  $Z$  au cas de signaux multirésolutions. Nous présentons l'algorithme de calcul des marginales *a posteriori* sans faire d'hypothèses d'indépendance conditionnelle. Puis, nous donnons la loi de  $Z$  pour les modèles d'arbres de Markov caché à bruit indépendant (AMC-BI), d'arbres de Markov couples (AMC Couple), d'arbres couples partiellement de Markov (ACPM) et nous présentons l'algorithme EM dans le cas AMC-BI. Enfin, nous proposons le modèle d'arbres de Markov cachés flous multispectraux et nous l'appliquons à la segmentation d'une image

de synthèse et d'une image astronomique multibande.

## 5.1 Extensions des modèles monodimensionnels au cas multi-résolution

Nous étendons les modèles présentés dans le chapitre 2 et 3 à des modèles multirésolutions. Nous suivons pour cela une démarche identique à celle suivie dans le chapitre 2. Nous donnons dans un premier temps la forme factorisée d'un signal multirésolution. Pour cela, nous devons introduire au préalable quelques notations relatives aux structures arborescentes. Soit  $\mathcal{S}$  un ensemble fini d'indices et  $\mathcal{S}^1, \dots, \mathcal{S}^N$  une partition de  $\mathcal{S}$ . Nous posons  $S = |\mathcal{S}|$  et pour  $1 \leq n \leq N$ , nous supposons que  $S^1 = 1 \leq S^2 \leq \dots \leq S^N$ . Pour  $1 \leq n \leq N - 1$ , nous associons à chaque  $s \in \mathcal{S}^n$  un sous ensemble  $s^+$  de  $\mathcal{S}^{n+1}$ , appelé "enfants de  $s$ " de manière à ce que  $(s^+)_{s \in \mathcal{S}^n}$  soit une partition de  $\mathcal{S}^{n+1}$ . L'unique élément de  $\mathcal{S}^1$  est appelé "racine". Par ailleurs,  $s^{++}$  désignera l'ensemble des descendants de  $s$  et pour tout  $s \in \mathcal{S} \setminus \mathcal{S}^1$ , son unique père sera noté  $s^-$ . Enfin, nous noterons  $\underline{s}^{++} = \{s\} \cup s^{++}$ . La loi d'un signal multirésolution  $Z$  peut donc, en tenant compte de ces notations, être écrite de la manière suivante

$$p(z) = p(z_1) \prod_{n=2}^N \prod_{s \in \mathcal{S}^n} p(z_s | z_{\mathcal{S} \setminus \underline{s}^{++}}) \quad (5.1.1)$$

Nous allons à présent décrire l'algorithme de calcul des probabilités marginales *a posteriori* dans le cas de signaux multirésolutions dont la loi est donnée par (5.1.1), permettant l'estimation d'une réalisation du processus caché au sens du MPM. Pour cela, nous procédons d'une manière similaire à celle suivie dans le cas des modèles monodimensionnels.

### 5.1.1 Calcul du MPM

L'estimateur du MPM, est, rappelons-le, celui qui, en chaque site  $s \in \mathcal{S}$ , maximise la probabilité marginale *a posteriori*. Il est donné, par

$$\forall s \in \mathcal{S}, (\hat{x}_s)_{MPM}(y) = \arg \max_{x'_s \in \mathcal{X}} p(x'_s | y) \quad (5.1.2)$$

Afin d'estimer une réalisation du processus caché au sens du MPM, il est donc nécessaire de calculer les probabilités marginales *a posteriori*  $p(x_s | y)$  en chaque site  $s \in \mathcal{S}$ . La loi du processus  $X$  *a posteriori* peut être écrite sous une forme factorisée suivante :

$$p(x | y) = p(x_1 | y) \prod_{n=2}^N \prod_{s \in \mathcal{S}^n} p(x_s | x_{\mathcal{S} \setminus \underline{s}^{++}}, y) \quad (5.1.3)$$

Cette dernière est donnée par le produit de la loi initiale *a posteriori*  $p(x_1 | y)$  et des lois de transition *a posteriori*  $\{p(x_s | x_{\mathcal{S} \setminus \underline{s}^{++}}, y)\}_{s \in \mathcal{S} \setminus \{1\}}$ . Dans le cas où celles-ci sont connues, on

peut calculer les probabilités marginales *a posteriori* en effectuant un parcours descendant de l'arbre appelé *passé descendante* [104]. En effet, connaissant  $p(x_1|y)$ , nous pouvons calculer  $p(x_s|y) = \sum_{x_{\mathcal{S}\setminus\underline{s}^{++}}} p(x_{\mathcal{S}\setminus\underline{s}^{++}}|y)$  avec  $p(x_{\mathcal{S}\setminus\underline{s}^{++}}|y) = p(x_{\mathcal{S}\setminus\underline{s}^{++}}|y)p(x_s|x_{\mathcal{S}\setminus\underline{s}^{++}}, y)$  pour  $n = 2 \dots N$  et  $s \in \mathcal{S}^n$ . Le problème se ramène donc au calcul de ces probabilités. En ce qui concerne le calcul des probabilités de transition *a posteriori*, celles-ci peuvent être calculées de la manière suivante pour tout  $n = 2 \dots N$ ,  $s \in \mathcal{S}^n$  :

$$p(x_s|x_{\mathcal{S}\setminus\underline{s}^{++}}, y) = p(z_s|z_{\mathcal{S}\setminus\underline{s}^{++}}) \frac{\prod_{t \in s^+} \beta_{t,s}(z_{\mathcal{S}\setminus\underline{s}^{++}})}{\beta_{s,s^-}(z_{\mathcal{S}\setminus\underline{s}^{++}})} \quad (5.1.4)$$

avec  $\beta_{s,s^-}(z_{\mathcal{S}\setminus\underline{s}^{++}}) = p(y_{\underline{s}^{++}}|z_{\mathcal{S}\setminus\underline{s}^{++}})$  la quantité dite *rétrograde*. En effet,

$$\begin{aligned} p(x_s|x_{\mathcal{S}\setminus\underline{s}^{++}}, y) &= \frac{p(x_{\mathcal{S}\setminus\underline{s}^{++}}, y_{\mathcal{S}\setminus\underline{s}^{++}}, y_{s^{++}})}{p(x_{\mathcal{S}\setminus\underline{s}^{++}}, y_{\mathcal{S}\setminus\underline{s}^{++}}, y_{\underline{s}^{++}})} \\ &= \frac{p(z_{\mathcal{S}\setminus\underline{s}^{++}}) p(y_{s^{++}}|z_{\mathcal{S}\setminus\underline{s}^{++}})}{p(z_{\mathcal{S}\setminus\underline{s}^{++}}) p(y_{\underline{s}^{++}}|z_{\mathcal{S}\setminus\underline{s}^{++}})} \\ &= p(z_s|z_{\mathcal{S}\setminus\underline{s}^{++}}) \frac{p(y_{s^{++}}|z_{\mathcal{S}\setminus\underline{s}^{++}})}{p(y_{\underline{s}^{++}}|z_{\mathcal{S}\setminus\underline{s}^{++}})} \\ &= p(z_s|z_{\mathcal{S}\setminus\underline{s}^{++}}) \frac{\prod_{t \in s^+} p(y_{t^{++}}|z_{\mathcal{S}\setminus\underline{s}^{++}})}{p(y_{\underline{s}^{++}}|z_{\mathcal{S}\setminus\underline{s}^{++}})} \\ &= p(z_s|z_{\mathcal{S}\setminus\underline{s}^{++}}) \frac{\prod_{t \in s^+} \beta_{t,s}(z_{\mathcal{S}\setminus\underline{s}^{++}})}{\beta_{s,s^-}(z_{\mathcal{S}\setminus\underline{s}^{++}})} \end{aligned}$$

$\beta_{s,s^-}(z_{\mathcal{S}\setminus\underline{s}^{++}})$  peut elle-même être exprimée en fonction de  $p(z_s|z_{\mathcal{S}\setminus\underline{s}^{++}})$  et de  $\prod_{t \in s^+} \beta_{t,s}(z_{\mathcal{S}\setminus\underline{s}^{++}})$ , en effet,

$$\begin{aligned} \beta_{s,s^-}(z_{\mathcal{S}\setminus\underline{s}^{++}}) &= p(y_{\underline{s}^{++}}|z_{\mathcal{S}\setminus\underline{s}^{++}}) \\ &= \sum_{x_s \in \mathcal{X}} p(z_s|z_{\mathcal{S}\setminus\underline{s}^{++}}) \prod_{t \in s^+} \underbrace{p(y_{t^{++}}|z_{\mathcal{S}\setminus\underline{s}^{++}})}_{\beta_{t,s}(z_{\mathcal{S}\setminus\underline{s}^{++}})} \end{aligned} \quad (5.1.5)$$

Ces quantités peuvent alors être calculées récursivement en effectuant un parcours montant sur l'arbre, ou *passé montante* de la manière suivante :

$$\beta_{s,s^-}(z_{\mathcal{S}\setminus\underline{s}^{++}}) \propto \begin{cases} \sum_{x_s \in \mathcal{X}} p(z_s|z_{\mathcal{S}\setminus\underline{s}^{++}}) & \text{si } n = N \\ \sum_{x_s \in \mathcal{X}} p(z_s|z_{\mathcal{S}\setminus\underline{s}^{++}}) \prod_{t \in s^+} \beta_{t,s}(z_{\mathcal{S}\setminus\underline{s}^{++}}) & \text{sinon} \end{cases} \quad (5.1.6)$$

On peut alors en déduire l'ensemble des probabilités de transitions *a posteriori*  $p(x_s|x_{\mathcal{S}\setminus\underline{s}^{++}}, y)$ , pour  $s \in \mathcal{S}^n$ ,  $n = 2 : N$  à partir de (5.1.4). Il est néanmoins nécessaire, de manière similaire aux chaînes, que les probabilité des transitions  $\{p(z_s|z_{\mathcal{S}\setminus\underline{s}^{++}})\}_{s \in \mathcal{S}^n, n=2:N}$  du processus  $Z$  soient connues ou calculables. Finalement, la probabilité marginale *a posteriori* à la racine

---

**Algorithme 7** Calcul des marginales a posteriori dans un arbre.

---

Passé montante : pour  $n = N \dots 2, \forall s \in \mathcal{S}^n$

$$\beta_{s,s^-}(z_{\mathcal{S} \setminus \underline{s}^{++}}) \propto \begin{cases} \sum_{x_s \in \mathcal{X}} p(z_s | z_{\mathcal{S} \setminus \underline{s}^{++}}) & \text{si } n = N \\ \sum_{x_s \in \mathcal{X}} p(z_s | z_{\mathcal{S} \setminus \underline{s}^{++}}) \prod_{t \in s^+} \beta_{t,s}(z_{\mathcal{S} \setminus s^{++}}) & \text{sinon} \end{cases}$$

$$p(x_s | x_{\mathcal{S} \setminus \underline{s}^{++}}, y) \propto \begin{cases} p(z_s | z_{\mathcal{S} \setminus \underline{s}^{++}}) & \text{si } n = N \\ p(z_s | z_{\mathcal{S} \setminus \underline{s}^{++}}) \prod_{t \in s^+} \beta_{t,s}(z_{\mathcal{S} \setminus s^{++}}) & \text{sinon} \end{cases}$$

Racine :  $p(x_1 | y) \propto p(z_1) \prod_{t \in 1^+} \beta_{t,1}(x_1)$

Passé descendante : pour  $n = 2 \dots N, \forall s \in \mathcal{S}^n$

$$p(x_{\mathcal{S} \setminus s^{++}} | y) = p(x_{\mathcal{S} \setminus \underline{s}^{++}} | y) p(x_s | x_{\mathcal{S} \setminus \underline{s}^{++}}, y)$$

$$p(x_s | y) = \sum_{x_{\mathcal{S} \setminus \underline{s}^{++}}} p(x_{\mathcal{S} \setminus s^{++}} | y)$$


---

peut être calculée à partir de  $p(z_1)$  et des  $\{\beta_{t,1}(z_1)\}_{t \in 1^+}$  de la manière suivante :

$$p(x_1 | y) \propto p(z_1) \prod_{t \in 1^+} \beta_{t,1}(z_1) \quad (5.1.7)$$

L'algorithme de calcul des marginales *a posteriori* est résumé dans l'algorithme 7. Il consiste à calculer dans un premier temps les probabilités de transition *a posteriori* lors de la passe montante  $n = N$  à  $n = 2, \forall s \in \mathcal{S}^n$ . La probabilité *a posteriori* à la racine est ensuite calculée. On peut alors en déduire les probabilités marginales *a posteriori* lors de la passe avant. Les probabilités marginales *a posteriori*  $\{p(x_s | y)\}_{s \in \mathcal{S}}$  étant connues, on peut estimer une réalisation des variables cachées au sens du MPM dont l'estimateur est donné par (5.1.2). Notons que la connaissance de ces probabilités marginales ainsi que des probabilités marginales  $p(x_{\mathcal{S} \setminus s^{++}} | y)$  sera également très utile pour l'estimation des paramètres du modèles.

### 5.1.2 Modélisations de $Z$ dans le cas d'un signal multirésolution

En procédant de la même manière que dans le second chapitre, nous pouvons citer les modèles suivants qui sont des extensions directes des modèles monodimensionnels.

\* Le modèle d'*Arbre de Markov caché à bruit indépendant* (AMC-BI) qui consiste à faire les hypothèses d'indépendance conditionnelle suivante : [35, 36, 15, 66, 32, 31, 65, 90, 104, 103]

$$\begin{aligned} X_s &\perp\!\!\!\perp X_{\mathcal{S} \setminus (\{s^-\} \cup \underline{s}^{++})} | X_{s^-} \\ Y_s &\perp\!\!\!\perp Z_{\mathcal{S} \setminus \{s\}} | X_s \end{aligned} \quad (5.1.8)$$

La loi de  $Z$  est alors donnée par (5.1.1) avec

$$p(z_s | z_{\mathcal{S} \setminus \underline{s}^{++}}) = p(x_s | x_{s^-}) p(y_s | x_s) \quad (5.1.9)$$

\* Le modèle d'*Arbre de Markov Couple* (AMC Couple) qui consiste à supposer directement la markovianité en échelle du processus  $Z$  : [94, 83, 49]

$$Z_s \perp\!\!\!\perp Z_{\mathcal{S} \setminus (s^- \cup \underline{s}^{++})} | Z_{s^-} \quad (5.1.10)$$

La loi de  $Z$  est alors donnée par (5.1.1) avec

$$p(z_s | z_{\mathcal{S} \setminus \underline{s}^{++}}) = p(z_s | z_{s^-}) \quad (5.1.11)$$

\*Le modèle d'*Arbre Couple Partiellement de Markov* (ACPM) qui consiste à faire l'hypothèse d'indépendance conditionnelle suivante : [98]

$$Z_s \perp\!\!\!\perp X_{\mathcal{S} \setminus (s^- \cup \underline{s}^{++})} | Z_{s^-}, Y_{\mathcal{S} \setminus (s^- \cup \underline{s}^{++})} \quad (5.1.12)$$

La loi de  $Z$  est alors donnée par (5.1.1) avec

$$p(z_s | z_{\mathcal{S} \setminus \underline{s}^{++}}) = p(z_s | x_{s^-}, y_{\mathcal{S} \setminus \underline{s}^{++}}) \quad (5.1.13)$$

Une réalisation du processus caché peut alors être estimée au sens du MPM en remplaçant  $p(z_s | z_{\mathcal{S} \setminus \underline{s}^{++}})$  par (5.1.9), (5.1.11) ou (5.1.13) dans l'algorithme 7. D'autre part, de la même manière que dans le chapitre 3, les AMCouples peuvent être étendus aux arbres de Markov triplet et les ACPM peuvent être étendus aux arbres triplets partiellement de Markov [98]. Nous allons à présent décrire l'algorithme EM dans le cas d'un AMC-BI multispectral. Cet algorithme nous sera utile dans la suite du chapitre lorsque nous aborderons l'estimation des AMC-BI flous multispectraux.

### 5.1.3 Estimation EM dans le cas d'un AMC-BI multispectral

Nous rappelons dans cette sous-section le développement permettant de trouver les formules de réestimation de l'algorithme EM dans le cas des AMC-BI multispectraux [65]. La loi jointe du processus est donnée par  $p(z|\phi) = \frac{\prod_{s \in \mathcal{S} \setminus \mathcal{S}^1} p(x_{s^-}, x_s)}{\prod_{s \in \mathcal{S} \setminus (\mathcal{S}^1 \cup \mathcal{S}^N)} p(x_s)} \times \prod_{s \in \mathcal{S}} p(y_s | x_s)$ . La vraisemblance complétée est donnée par

$$\mathcal{L}_c(\phi) = \sum_{s \in \mathcal{S} \setminus \mathcal{S}^1} \log p(x_{s^-}, x_s) - \sum_{s \in \mathcal{S} \setminus (\mathcal{S}^1 \cup \mathcal{S}^N)} \log p(x_s) + \sum_{s \in \mathcal{S}} \log p(y_s | x_s)$$

**Etape E** Afin de simplifier les notations, nous posons

$$\left\{ \begin{array}{l} \pi_i = p(x_s = i) \\ c_{ij} = p(x_{s^-} = i, x_s = j) \\ f_i(y_s) = p(y_s | x_s = i) \\ \gamma_s^{[q]}(i) = p(x_s = i | y; \phi^{[q]}) \\ \psi_{s^-, s}^{[q]}(i, j) = p(x_{s^-} = i, x_s = j | y; \phi^{[q]}) \end{array} \right.$$

où  $f_i(\cdot)$  est une loi paramétrée par rapport à une mesure de référence  $\nu$ , nous obtenons ainsi :

$$\mathcal{Q}(\phi | \phi^{[q]}) = \sum_{s \in \mathcal{S} \setminus \mathcal{S}^1} \sum_{(i, j) \in \mathcal{X}^2} \psi_{s^-, s}^{[q]}(i, j) \log c_{ij} - \sum_{s \in \mathcal{S} \setminus (\mathcal{S}^1 \cup \mathcal{S}^N)} \sum_{i \in \mathcal{X}} \gamma_s^{[q]}(i) \log \pi_i + \sum_{s \in \mathcal{S}} \sum_{i \in \mathcal{X}} \gamma_s^{[q]}(i) \log f_i(y_s) \quad (5.1.14)$$

**Etape M** Nous devons à présent maximiser la fonction  $\mathcal{Q}(\phi | \phi^{[q]})$ . Pour cela, nous utilisons la méthode des multiplicateurs de Lagrange sous les contraintes

$$\sum_{(i, j) \in \mathcal{X}^2} c_{ij} = 1, \quad \sum_{j \in \mathcal{X}} c_{ij} = \pi_i \quad \text{et} \quad \int_{l \in \mathcal{Y}} f_i(l) = 1$$

Finalement, nous obtenons les formules de réestimation suivantes, les  $\gamma_s(i)$  et les  $\psi_{s^-, s}^{[q]}(i, j)$  étant obtenus par l'algorithme 7 :

$$c_{ij}^{[q+1]} = \frac{1}{S-1} \sum_{s \in \mathcal{S} \setminus \mathcal{S}^1} \psi_{s^-, s}^{[q]}(i, j), \quad f_i^{[q+1]}(l) = \frac{\sum_{s \in \mathcal{S} : y_s = l} \gamma_s^{[q]}(i)}{\sum_{s \in \mathcal{S}} \gamma_s^{[q]}(i)} \quad (5.1.15)$$

Dans le cas d'une attache aux données suivant une loi gaussienne multispectrale, les formules de réestimation des vecteurs moyennes et des matrices de covariances sont données par :

$$\left\{ \begin{array}{l} \mu_i^{[q+1]} = \frac{\sum_{s \in \mathcal{S}} y_s \gamma_s^{[q]}(i)}{\sum_{s \in \mathcal{S}} \gamma_s^{[q]}(i)} \\ \Gamma_i^{[q+1]} = \frac{\sum_{s \in \mathcal{S}} \gamma_s^{[q]}(i) (y_s - \mu_i^{[q+1]})(y_s - \mu_i^{[q+1]})^T}{\sum_{s \in \mathcal{S}} \gamma_s^{[q]}(i)} \end{array} \right. \quad (5.1.16)$$

## 5.2 Arbres de Markov cachés flous multispectraux

Nous avons implicitement admis jusqu'à présent qu'un site donné ne pouvait appartenir qu'à une seule classe dans l'ensemble  $\Omega = \{\omega_1, \dots, \omega_K\}$ . Dans le cas de deux classes tel que  $\Omega = \{0, 1\}$ , on peut imaginer l'existence de site contenant un mélange des deux classes sans qu'aucune des deux classes ne s'impose. On peut alors poser  $x_s = 0$  si  $s$  appartient à la classe "0",  $x_s = 1$  si  $s$  appartient à la classe "1" et  $x_s = \varepsilon_s \in ]0, 1[$  si la proportion de la classe "0" dans le site est  $\varepsilon_s \in ]0, 1[$ . Nous obtenons de cette manière une appartenance "floue" du site. Nous dirons que le site est "purement flou" si  $x_s \in ]0, 1[$  est qu'il est "dur" si  $x_s \in \{0, 1\}$ .



On peut alors considérer cette appartenance comme aléatoire en définissant la loi de  $X_s$  sur l'intervalle  $[0, 1]$ . Une possibilité est de supposer que celle-ci est donnée par une densité  $p(x_s)$  par rapport à une mesure  $\nu$  incluant des composantes discrètes et continues : les fonction de Dirac  $\delta_0, \delta_1$  sur  $\{0, 1\}$  et la mesure de Lebesgue  $\mu$  sur  $]0, 1[$  [21] :

$$\nu = \delta_0 + \delta_1 + \mu \quad (5.2.1)$$

Précisons que d'autres travaux utilisent des modélisations comprenant un mélange de composantes continues et discrètes, notamment dans [16]. Nous avons alors

$$p(0) + p(1) + \int_0^1 p(\varepsilon)d\varepsilon = 1 \quad (5.2.2)$$

Nous considérons à présent un arbre de Markov  $X = (X)_{s \in \mathcal{S}}$  avec  $X_s \in [0, 1]$ . Dans ce contexte, la loi d'un *arbre de Markov flou*  $X$  par rapport à la mesure  $\nu^{\mathcal{S}}$  est définie par une loi initiale  $p(x_1)$  et les densités de transition  $p(x_s|x_{s-})$

$$p(x) = p(x_1) \prod_{n=2}^N \prod_{s \in \mathcal{S}^n} p(x_s|x_{s-}) \quad (5.2.3)$$

Nous souhaitons traiter par la suite des données multispectrales. Soit une réalisation  $y = (y_s^{(1)}, y_s^{(2)}, \dots, y_s^{(D)})_{s \in \mathcal{S}}$  du vecteur aléatoire  $Y = (Y_s^{(1)}, Y_s^{(2)}, \dots, Y_s^{(D)})_{s \in \mathcal{S}}$ ,  $Y = y$  représentant l'observation d'une scène unique par  $D$  capteurs. Conditionnellement aux états  $x_s$ , les observations sont supposées indépendantes mais spectralement corrélées. La distribution de  $Y_s = (Y_s^{(1)}, Y_s^{(2)}, \dots, Y_s^{(D)})$  conditionnellement à  $x_s$  est modélisée par une densité gaussienne multivariée

$$p(y_s|x_s) = \frac{1}{(2\pi)^{D/2}(\det \mathbf{\Gamma}_{x_s})^{1/2}} \cdot e^{(-\frac{1}{2}(y_s - \mu_{x_s})^T \mathbf{\Gamma}_{x_s}^{-1} (y_s - \mu_{x_s}))} \quad (5.2.4)$$

où  $\mu_{x_s} = [\mu_{x_s}^{(1)}, \mu_{x_s}^{(2)}, \dots, \mu_{x_s}^{(D)}]^T$  et  $\mathbf{\Gamma}_{x_s} \in \mathbb{R}^D \times \mathbb{R}^D$  définissent respectivement le vecteur moyenne et la matrice de variance-covariance pour chaque état  $x_s$ . Soient  $(\mu_0, \mu_1)$  et  $(\mathbf{\Gamma}_0, \mathbf{\Gamma}_1)$  les vecteurs moyennes et les matrices de covariance relatifs aux classes dures "0" et "1". Nous utilisons le modèle proposé dans [107], pour exprimer la matrice de variance covariance relative aux classes floues au moyen de celle relative aux classes dures. Le vecteur moyenne  $\mu_{\varepsilon_s}$  et la matrice  $\mathbf{\Gamma}_{\varepsilon_s}$  associées au site flou  $X_s = \varepsilon_s$  dépendent linéairement des paramètres durs  $(\mu_0, \mu_1), (\mathbf{\Gamma}_0, \mathbf{\Gamma}_1)$  :

$$\begin{cases} \mu_{\varepsilon_s} = (1 - \varepsilon_s)\mu_0 + \varepsilon_s\mu_1 \\ \mathbf{\Gamma}_{\varepsilon_s} = (1 - \varepsilon_s)^2\mathbf{\Gamma}_0 + \varepsilon_s^2\mathbf{\Gamma}_1 \end{cases} \quad (5.2.5)$$

---

**Algorithme 8** Calcul des marginales *a posteriori* dans un arbre de Markov caché flou

---

Passé montante : pour  $n = N \dots 2, \forall s \in \mathcal{S}^n$

$$\beta_{s,s^-}(z_{s^-}) \propto \begin{cases} \int_{\mathcal{X}} p(x_s|x_{s^-})p(y_s|x_s)d\nu(x_s) & \text{si } n = N \\ \int_{\mathcal{X}} p(x_s|x_{s^-})p(y_s|x_s) \prod_{t \in s^+} \beta_{t,s}(z_t)d\nu(x_s) & \text{sinon} \end{cases}$$
$$p(x_s|x_{s^-}, y) \propto \begin{cases} p(x_s|x_{s^-})p(y_s|x_s) & \text{si } n = N \\ p(x_s|x_{s^-})p(y_s|x_s) \prod_{t \in s^+} \beta_{t,s}(z_t) & \text{sinon} \end{cases}$$

Racine :  $p(x_1|y) \propto p(z_1) \prod_{t \in 1^+} \beta_{t,1}(x_t)$

Passé descendante : pour  $n = 2 \dots N, \forall s \in \mathcal{S}^n$

$$p(x_s, x_{s^-}|y) = p(x_{s^-}|y)p(x_s|x_{s^-}, y)$$

$$p(x_s|y) = \int_{\mathcal{X}} p(x_s, x_{s^-}|y)d\nu(x_{s^-})$$

---

Nous pouvons écrire à présent la loi de  $Z$  par rapport à la mesure  $\nu^S \otimes (\mu^S)^{(1)} \otimes (\mu^S)^{(2)} \otimes \dots \otimes (\mu^S)^{(D)}$ , où les  $(\mu^S)^{(i)}$  correspondent à la mesure de Lebesgue sur  $\mathbb{R}^S$  :

$$p(z) = p(x_1)p(y_1|x_1) \prod_{n=2}^N \prod_{s \in \mathcal{S}^n} p(x_s|x_{s^-})p(y_s|x_s) \quad (5.2.6)$$

On peut dès lors estimer une réalisation  $x \in [0, 1]^S$  au sens d'un estimateur défini à partir de la fonction de perte "distance en valeur absolue" donnée par  $L(x_s, \hat{x}_s) = |x_s - \hat{x}_s|$ . L'algorithme de calcul des marginales *a posteriori* est présenté dans l'algorithme 8.

Les intégrales ne pouvant pas être calculées de façon analytique, il convient de proposer une approximation numérique qui permette leur implémentation numérique. Pour cela, nous divisons l'intervalle  $[0, 1]$  en  $F$  sous-intervalles flous  $[a_0 = 0, a_1 = \frac{1}{F}]$ ,  $[a_1 = \frac{1}{F}, a_2 = \frac{2}{F}]$ ,  $\dots$ ,  $[a_{F-1} = \frac{F-1}{F}, a_F = 1]$ . A chaque sous-intervalle flou, nous associons la valeur médiane correspondante. D'un point de vue numérique et algorithmique, les traitements associés à un arbre de Markov caché flou à deux classes dures et  $F$  sous-intervalles flous sont donc très proches de ceux mis en oeuvre pour un AMC classique à  $F + 2$  états [65, 66], la différence étant que les paramètres des lois d'attache aux données floues sont contraints par ceux des classes dures.

L'arbre de Markov flou  $X$  étant stationnaire et homogène, sa loi est donnée par la densité  $p(x_s, x_{s^-})$  qui ne dépend pas de  $s$ . Nous avons utilisé l'algorithme EM présenté dans l'algorithme 9 et initialisé par un algorithme de k-moyennes, pour calculer les paramètres  $\phi$  de la manière suivante :

### 5.3 Expérimentations

Nous avons tout d'abord testé le modèle d'arbre de Markov flou sur une image de synthèse trispectrale (figure 5.4.2(a)-(c)) dont la composante "vérité-terrain" est un champ de Markov flou à 10 niveaux présenté sur la figure 5.4.1. L'algorithme est initialisé par un algorithme de k-moyennes. Les résultats de la segmentation sont donnés sur la figure 5.4.2(d). L'algorithme

---

**Algorithme 9** Algorithme d'estimation EM des paramètres d'un arbre de Markov flou multispectral.

---

-Initialisation :  $\phi^{[0]} = (c_{ij}^{[0]}, \mu_k^{[0]}, \Gamma_k^{[0]})_{0 \leq i, j \leq 1, 0 \leq k \leq 1}$

-Pour tout  $q \in \mathbb{N}^+$ ,  $\phi^{[q+1]}$  est calculé à partir de  $y$  et  $\phi^{[q]}$  par

$$c_{ij}^{[q+1]} = \frac{1}{\sum_{i=2}^n N^i} \sum_{i=2}^n \sum_{s \in S^i} p(x_s, x_{s-} | y)$$

-pour  $k \in \{0, 1\}$  (classes dures) : (d'après 5.1.16)

$$\mu_k^{[q+1]} = \frac{\sum_{s \in S} y_s \gamma_s^{[q]}(k)}{\sum_{s \in S} \gamma_s^{[q]}(k)}$$

$$\Gamma_k^{[q+1]} = \frac{\sum_{s \in S} \gamma_s^{[q]}(k) (y_s - \mu_k^{[q+1]}) (y_s - \mu_k^{[q+1]})^T}{\sum_{s \in S} \gamma_s^{[q]}(k)}$$

-pour  $k \in ]0, 1[$  (classes floues) :

$$\mu_k^{[q+1]} = (1 - k) \mu_0^{[q+1]} + k \mu_1^{[q+1]}$$

$$\Gamma_k^{[q+1]} = (1 - k)^2 \Gamma_0^{[q+1]} + k^2 \Gamma_1^{[q+1]}$$


---

restitue correctement l'image d'origine avec un taux d'erreur de 4.58%. Les paramètres des lois d'attache aux données estimés selon l'algorithme EM sont indiqués dans le tableau 5.1. L'information floue est correctement restituée (44.96% vs 53.96%) et de ce point de vue, cette méthode donne, dans le cas considéré, de meilleurs résultats que l'approche multibandes par champ de Markov [107].

D'autre part, la figure 5.4.3 (a)-(c) présente trois images qui se prêtent bien "visuellement" au modèle flou. Ce sont des images astronomiques dans le domaine infrarouge à 60, 100 et 170 microns, observées à partir d'un niveau de résolution unique. Il s'agit de champs de haute altitude de notre galaxie pour lesquels on cherche à mesurer les fluctuations induites par la présence de galaxies extérieures à la voie lactée, non résolues dans le cadre d'une étude cosmologique. Des fluctuations, provoquées par la poussière interstellaire de notre galaxie, gênent l'observation. Cette poussière interstellaire (en bas à droite des images multispectrales) possède une structure filamentaire que nous souhaitons mettre en évidence par une analyse floue multibande. Le résultat d'une segmentation MPM sur 10 niveaux donne la structure en arbre multirésolution observée figure 5.4.3 (d). Les paramètres des moyennes et covariances obtenus pour chaque bande sont donnés dans le tableau 5.2. Bien que le résultat de la segmentation ne permette pas de mettre en évidence la structure filamentaire de la poussière interstellaire, les résultats montrent qu'il est possible de restituer une gradation floue entre les régions homogènes.

## 5.4 Conclusion

Après avoir étendu les modèles monodimensionnels au cas multirésolution, nous avons proposé une nouvelle méthode de segmentation statistique non supervisée fondée sur une

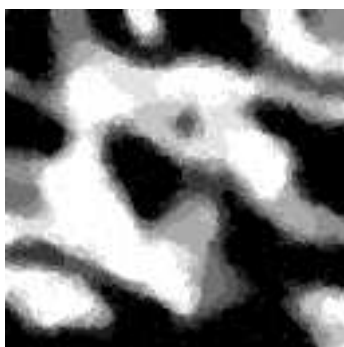


FIG. 5.4.1 – Image floue originale simulée à partir d’une champ de Markov flou avec 2 classes dures et 8 niveaux de flou. Le taux de flou est de 53.96%

$\mu_0$	120	120	120	$\mu_0$	119.9	119.8	119.9
$\sigma_0^2$	16	64	256	$\Gamma_0$	14.0	63.0	249.4
$\mu_1$	146	146	146	$\mu_1$	146.0	145.9	145.8
$\sigma_1^2$	16	64	256	$\Gamma_1$	13.8	64.0	244.0

TAB. 5.1 – Paramètres d’attache aux données vrais (tableau de gauche) et estimés (tableau de droite) correspondant à la figure 5.4.2

structure d’arbre de Markov caché flou. Ce modèle s’appuie localement sur une densité de probabilité par rapport à la mesure constituée d’une composante continue et de composantes discrètes. Nous avons supposé les capteurs corrélés, en exprimant les lois du bruit par des gaussiennes multivariées dont les matrices des variances-covariances des classes floues sont déduites linéairement des matrices de variances-covariances des classes dures. Les méthodes ont été testées sur des images de synthèse trispectrales puis appliquées à des données astronomiques. Ces résultats montrent l’intérêt des méthodes non supervisées floues en astronomie, en restituant ainsi une gradation floue entre les zones homogènes du ciel. On pourra envisager d’enrichir le modèle en introduisant d’autres lois paramétriques [38] afin de caractériser les observations.

$\mu_0$	51.6	34.0	14.0
$\sigma_0^2$	301.5	106.9	97.2
	106.9	175.0	44.3
	97.2	44.3	52.9
$\mu_1$	132.8	198.0	225.4
$\sigma_1^2$	1019.7	1094.5	462.2
	1094.5	1420.4	404.0
	462.2	404.0	297.6

TAB. 5.2 – Paramètres estimés dans le cas de la figure 5.4.3

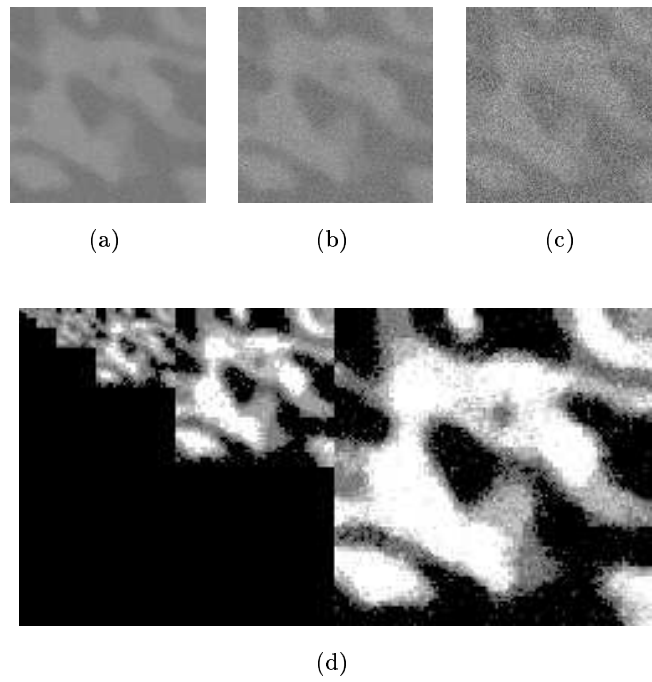


FIG. 5.4.2 – Segmentation de données synthétiques. Images du haut : les trois bandes d'observations. Image du bas : Résultats sur 8 niveaux de flou et 2 classes dures.

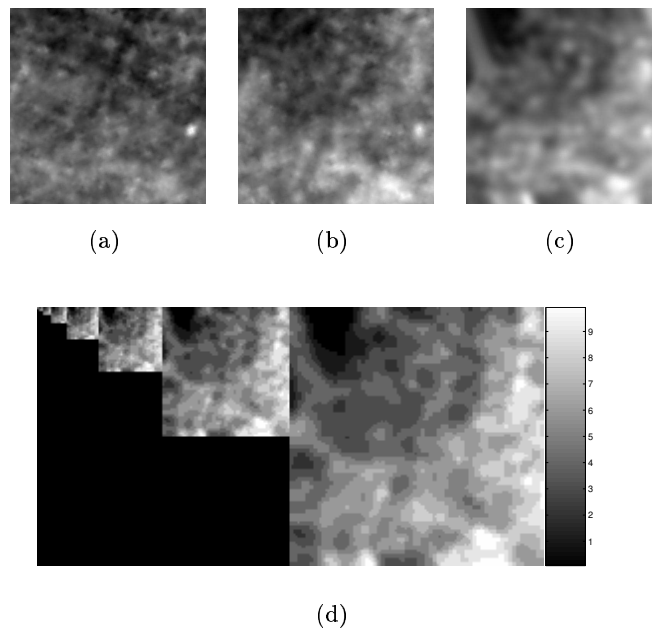


FIG. 5.4.3 – Segmentation d'une image astronomique multispectrale par arbre de Markov caché flou. (a)-(c) : les données multibandes ; (d) : résultats de la segmentation à 2 classes dures et 8 niveaux de flou.



# Conclusion générale et perspectives

Le contexte général de cette étude était la segmentation statistique de signaux et d'images. L'objectif était de proposer des méthodes originales de segmentation non supervisées de signaux et d'images fondées sur les modèles de Markov triplets et les modèles triplets partiellement de Markov. Nous avons apporté dans ce cadre un certain nombre de contributions.

Nous avons tout d'abord présenté différentes modélisations, de généralités croissantes, de la loi jointe du processus caché et des observations. Une première contribution a été de développer l'algorithme d'estimation EM dans le cas des chaînes de Markov Couples. Une deuxième contribution a été d'étudier un cas particulier de chaîne couple partiellement de Markov permettant la segmentation de processus gaussiens à corrélation longue. Les expérimentations sur données synthétiques sont très encourageantes et laissent percevoir de nettes améliorations par rapport aux modèles classiques lorsque les bruits sont effectivement à corrélation longue.

Néanmoins, la méthode de l'estimation des paramètres proposée n'est valable que pour le cas centré, ce qui nous a empêché de tester le modèle sur des images réelles. Des études devront donc être menées afin d'affiner notre méthode et de rendre possible la segmentation non supervisée de processus gaussiens dont les moyennes ne sont pas nécessairement nulles.

Nous avons ensuite présenté le principe général des modèles triplets. Nous avons, dans ce contexte, amené une troisième contribution qui fut de proposer un cas particulier de chaîne triplet permettant la segmentation non supervisée de processus non stationnaires [70, 68] et de montrer par la même occasion comment estimer les paramètres d'une chaîne de Markov triplet. Dans ce type de modèle le changement de matrice de transition du processus caché est gouverné par un processus auxiliaire. Ce modèle a été appliqué à la segmentation d'images synthétiques et réelles. Une première constatation est que ce modèle permet effectivement la prise en compte de régimes différents, ce qui se traduit par une amélioration de la qualité de segmentation dans le cas d'image possédant à la fois des zones homogènes étendues et des zones possédant des détails fins. Une deuxième constatation est qu'il est également possible d'obtenir une réalisation du processus auxiliaire par MPM. Ce type de représentation peut être très utile, notamment en segmentation de textures qui peuvent précisément être modélisées par le processus auxiliaire.

Une quatrième contribution, dans le cadre de l'étude des modèles triplets, fut d'étudier

les possibilités d'extension des modèles probabilistes classiques à un modèle « évidentiel », avec la loi *a posteriori* du processus caché donnée par la fusion de Dempster-Shafer [69, 71]. Nous avons alors également appliqué ce modèle évidentiel à la segmentation de processus non stationnaires. L'intérêt principal de notre approche est de montrer que, bien que la fusion de Dempster-Shafer détruise la markovianité dans le contexte de la chaîne cachée évidentielle, la segmentation bayésienne reste possible via les chaînes de Markov triplets. Notons que ces deux modèles triplets permettant la segmentation non supervisée de processus non stationnaires ont par la suite été repris et adaptés avec succès aux champs de Markov [5, 4, 6]. Une perspective intéressante que nous avons commencé à étudier est d'étendre le modèle de chaînes de Markov évidentielles au modèle d'arbres de Markov évidentiels.

Enfin, une dernière de nos contributions fut d'étendre les modèles de chaînes floues et de champs flous proposés auparavant aux modèles d'arbres flous. Nous avons ainsi proposé un modèle d'arbre de Markov flou caché multicapteur que nous avons appliqué à la segmentation d'images astronomiques [72]. Une perspective possible serait d'étendre ce modèle d'arbres de Markov cachés flous multispectraux à un modèle d'arbres de Markov couples flous multispectraux en s'appuyant par exemple sur les résultats obtenus dans [49].



# Bibliographie

- [1] L. E. Baum. An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes. In *Inequalities, III (Proc. 3rd Symp., Univ. Calif., Los Angeles, Calif., 1969; dedicated to the memory of Theodore S. Motzkin)*, pages 1–8, New York, 1972. Academic.
- [2] L. E. Baum and T. Petrie. Statistical inference for probabilistic functions of finite state Markov chains. *Ann. Math. Statist.*, 37 :1554–1563, 1966.
- [3] L.E. Baum, T. Petrie, G. Soules, and N. Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *Ann, Math. Statistic.*, 41 :164–171, 1970.
- [4] D. Benboudjema and W. Pieczynski. Segmenting non stationary images with triplet Markov fields. In *International Conference on Image Processing (ICIP 2005)*, Genova, Italy, September 11-14 2005.
- [5] D. Benboudjema and W. Pieczynski. Unsupervised image segmentation using triplet Markov fields. *Computer Vision and Image Understanding*, 3 :476–498, 2005.
- [6] D. Benboudjema and W. Pieczynski. Unsupervised segmentation of non stationary images with non Gaussian correlated noise using triplet Markov fields and the Pearson system. In *Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP'06)*, Toulouse, May 15-19 2006.
- [7] A. Bendjebbour, Y. Delignon, L. Fouque, V. Samson, and W. Pieczynski. Multisensor images segmentation using Dempster-Shafer fusion in Markov fields context. *IEEE Trans. on Geoscience and Remote Sensing*, 39(8) :1789–1798, 2001.
- [8] Y. Bengio and P. Frasconi. Input-output HMM's for sequence processing. *IEEE Transactions on Neural Networks*, 7(5) :1231–1249, September 1996.
- [9] B. Benmiloud and W. Pieczynski. Estimation des paramètres dans les chaînes de Markov cachées et segmentation d'images. *Traitement du Signal*, 12(5) :433–454, 1995.
- [10] F. Le Ber, M. Benoît, C. Schott, J.-F. Mari, and C. Mignolet. Studying crop sequences with carrotAge, a HMM-based data mining software. *Ecological Modelling*, 191(1) :170–185, 2006.

- [11] J. Beran and M. S. Taqqu. Statistics for long-memory processes. In *Monographs on Statistics and Applied Probability*. Chapman and Hall, New York, 1994.
- [12] J. Besag. On the statistical analysis of dirty pictures. *J. Royal Statist. Soc.*, B48(3) :259–302, 1986.
- [13] Y. M. Bishop, S.E. Fienberg, and P.W. Holland. *Discrete Multivariate Analysis*. MIT Press, 1975.
- [14] I. Bloch and H. Maître. Fusion de données en traitement d’images : modèles d’informations et décisions. *Traitement du Signal*, 11(6) :435–446, 1994.
- [15] C. Bouman and M. Shapiro. A multiscale image model for Bayesian image segmentation. *IEEE Trans. on Image Processing*, 3(2) :162–177, 1994.
- [16] P. Bouthemy, C. Hardouin, G. Piriou, and J.-F. Yao. Mixed-state auto-models and motion texture modeling. *Journal of Mathematical Imaging and Vision*, 25(3) :387–402, 2006.
- [17] S. Bricq, Ch Collet, and J.-P. Armspach. Triplet markov chain for 3D MRI brain segmentation using a probabilistic atlas. In *IEEE 2006 International Symposium on Biomedical Imaging*, April 6-9 2006.
- [18] N. Brunel and F. Barbaresco. Doppler and polarimetric statistical segmentation of radar clutter environment based on pairwise markov chains. In *Proceedings of International conference on Radar Systems (Radar 2004)*, 18-22 Octobre 2004.
- [19] N. Brunel and W. Pieczynski. Unsupervised signal restoration using copulas and pairwise Markov chains. In *Proceedings of IEEE Workshop on Statistical Signal Processing (SSP 2003)*, Saint Louis, Missouri, September28-October 1 2003.
- [20] N. Brunel and W. Pieczynski. Unsupervised signal restoration using hidden markov chains with copulas. *Signal Processing*, 85(12) :2304–2315, 2005.
- [21] H. Caillol, A. Hillion, and W. Pieczynski. Fuzzy random Markov fields and unsupervised image segmentation. *IEEE Trans. Geosci. Remote Sensing*, 4 :801–810, 1993.
- [22] O. Cappé, E. Moulines, J. C. Pesquet, A. Petropulu, and X. Yang. Long-range dependence and heavy-tail modeling for teletraffic data. *IEEE Signal Processing Magazine*, 19(3) :14–27, 2002.
- [23] O. Cappé, E. Moulines, and T. Ryden. *Inference in Hidden Markov Models*. Springer-Verlag, 2005.
- [24] C. Carincotte, S. Derrode, G. Sicot, and J.-M. Boucher. Unsupervised image segmentation based on a new fuzzy HMC model. In *ICASSP’04*, Montreal, Canada, May 17-21 2004.
- [25] G. Celeux, D. Chauveau, and J. Diebolt. Stochastic versions of the EM algorithm : an experimental study. *Journal of Statistical Computation and Simulation*, 55 :287–314, 1996.

- [26] G. Celeux and J. Diebolt. The sem algorithm : A probabilistic teacher algorithm derived from the em algorithm for the mixture problem. *Computational Statistic Quarterly*, 2 :73–82, 1985.
- [27] G. Celeux and J. Diebolt. A random imputation principle : the stochastic EM algorithm. Technical Report 901, INRIA Rocquencourt, september 1988.
- [28] G. Celeux and J. Diebolt. A stochastic aproximation type EM algorithm for the mixture problem. *Stochastics and Stochastics Reports*, 41 :119–134, 1992.
- [29] B. Chalmond. *Eléments de modélisation pour l'analyse d'images*, volume 33 of *Mathématiques et applications*. Springer-Verlag, 2000.
- [30] F. Chapeau-Blondeau and M. Guglielmi. Modèles de signaux à longue dépendance statistique. In *Signaux Aléatoires : Modélisation, Estimation, Détection*, chapter 4. Hermes, Paris, traité ic2 edition, 2004.
- [31] A. Chardin and P. Pérez. Mode of posteriori marginals with hierarchical models. In *Proc.Int. Conf. Image Processing*, volume I, pages 324–328, 1999.
- [32] A. Chardin and P. Pérez. Unsupervised image classification with a hierarchical em algorithm. In *Proc. Int. Conf. Computer Vision*, volume II, pages 969–974, 1999.
- [33] A. Chehikian. Recursive filters for gradient estimation and contour extraction using B-spline interpolation. *Traitement du signal*, 14(1) :29, 1997.
- [34] S. Chevalier, M. Kaynak, A. Cheok, K. Sengupta, and K. Chung. Use of a novel generalized fuzzy hidden markov model for speech recognition. In *Proceedings of the IEEE Fuzzy Systems Conference (FUZZ-IEEE 2001)*, volume 3, pages 1207–1220, dec. 2001.
- [35] K. Chou, S. Golden, and A. Willsky. Multiresolution stochastic models, data fusion and wavelet transforms. *IEEE Trans. on Signal Processing*, 34(3) :257–282, 1993.
- [36] K. Chou, A. Willsky, and A. Benveniste. Multiscale recursive estimation, data fusion, and regularization. *IEEE Trans. Automatic Control*, 39(3) :464–477, 1994.
- [37] J-P. Cocquerez and S. Philip. *Analyse d'images : filtrage et segmentation*. Masson, 1995.
- [38] Y. Delignon, A. Marzouki, and W. Pieczynski. Estimation of generalized mixtures and its application in image segmentation. *IEEE Transactions on Image Processing*, 6(10) :1364–1375, 1997.
- [39] J.-P. Delmas. An equivalence of the EM and ICE algorithm for exponential family. *IEEE Trans. on Signal Processing*, 45(10) :2613–2615, 1997.
- [40] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of Royal Statistical Society B*, 39(1) :1–38, 1977.
- [41] R. Deriche. Using Canny's criteria to derive a recursively implemented optimal edge detector. *Int. Journal of Computer Vision*, 1(2) :167–187, 1987.

- [42] S. Derrode and W. Pieczynski. Signal and image segmentation using pairwise Markov chains. *IEEE Trans. on Signal Processing*, 52(9) :2477–2489, 2004.
- [43] F. Desbouvries, J. Lecomte, and W. Pieczynski. Kalman filtering in pairwise Markov trees. *Signal Processing*, 86(5) :1049–1054, 2006.
- [44] F. Desbouvries and W. Pieczynski. Modèles de Markov triplet et filtrage de Kalman. *CRAS*, 336(8) :667–670, 2003.
- [45] F. Desbouvries and W. Pieczynski. Particle filtering with pairwise Markov processes. In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing (ICASSP'03)*, Hong-Kong, 2003.
- [46] X. Descombes, F. Kruggel, and D. Y. Von Cramon. Spatio-temporal fMRI analysis using markov random fields. *IEEE Trans. on Medical Imaging*, 17 :1028–1039, 1998.
- [47] R. Duda, P. Hart, and D. Stork. *Pattern Classification (second edition)*. Wiley, 2001.
- [48] R. Fjortoft, J. M. Boucher, Y. Delignon, R. Garello, J. M. LeCaillec, H. Maitre, J. M. Nicolas, W. Pieczynski, M. Sigelle, and F. Tupin. Unsupervised classification of radar images based on hidden Markov models and generalised mixture estimation. In *ProcSymposium on Remote Sensing Conference on SAR image analysis*, 2000.
- [49] F. Flitti. *Techniques de réduction de données et analyse d'images multispectrales astronomiques par arbres de Markov*. PhD thesis, Université Louis Pasteur - Strasbourg I, 2005.
- [50] G. D. Fornay. The Viterbi algorithm. *Proceedings of the IEEE*, 61(3) :268–277, 1973.
- [51] L. Fouque, A. Appriou, and W. Pieczynski. Multiresolution hidden Markov chain model and unsupervised image segmentation. In *Proceedings of IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI'2000)*, pages 121–125, Austin, Texas, Unisted States, 2000.
- [52] D. Geman. Random field and inverse problems in imaging. In P.L. Hennequin, editor, *Lecture notes in mathematics*, pages 117–193, Berlin, 1990. Ecole d'été de probabilités de Saint-Flour XVIII-1988, Springer-Verlag.
- [53] S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 6(6) :721–741, 1984.
- [54] N. Giordana and W. Pieczynski. Estimation of generalized multisensor hidden Markov chains and unsupervised image segmentation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(5) :465–475, 1997.
- [55] G.H. Golub and C.F. Van Loan. *Matrix Computations*. The John Hopkins University Press, 1989.
- [56] R. M. Haraclick. Digital step edges from zero-crossings of second directional derivative. *IEEE Trans. on Pattern Anal. Machine Intell.*, 6(1) :58–68, 1984.

- [57] S. L. Hégarat-Masclé, I. Bloch, and D. Vidal-Madjar. Introduction of neighborhood information on evidence theory and application to data fusion of radar and optical images with partial cloud cover. *Pattern Recognition*, 31(11) :1811–1823, 1998.
- [58] D. M. Higdon, J. E. Bowsher, V. E. Johnson, T. G. Turkington, T. G. Gilland, and R. J. Jazszczak. Fully Bayesian estimation of gibbs hyperparameters for emission computed tomography data. *IEEE Trans. on Medical Imaging*, 16(5) :516–526, 1997.
- [59] M. F. Hueckel. An operator which locates edges in digitized pictures. *J. Ass. Comput. Mach.*, 18(1) :113–125, 1971.
- [60] J. P. Hughes, P Guttorp, and S. P. Charles. A non-homogeneous hidden Markov model for precipitation occurrence. *Applied Statistics*, 48(1) :15–30, 1999.
- [61] M. Kass, A. Witkin, and D. Terzopoulos. Snakes : active contour models. *Int. Journal of Computer Vision*, 1(4) :321–331, 1988.
- [62] J.T. Kent and K.V. Mardia. Spatial classification using fuzzy membership. *IEEE, Trans. On Pattern Analysis and Machine Intelligence*, 10(5) :659–671, 1988.
- [63] T. Koski. *Hidden Markov models for Bioinformatics*. Kluwer Academic Publishers, 2001.
- [64] S. Kumar and M. Hebert. Discriminative random fields : A discriminative framework for contextual interaction in classification. *IEEE International Conference on Computer Vision (ICCV)*, 2003.
- [65] J.-M. Laferté, P. Pérez, and F. Heitz. Discrete Markov image modeling and inference on the quadtree. *IEEE Trans. on Image Processing*, 9(3) :390 – 404, 2000.
- [66] J.-M. Laferté, F. Heitz, P. Pérez, and E. Fabre. Hierarchical statistical models for the fusion of multiresolution image data. In *Proc. Int. Conf. Computer Vision*, Cambridge, 1995.
- [67] John Lafferty, Andrew McCallum, and Fernando Pereira. Conditional random fields : Probabilistic models for segmenting and labeling sequence data. In *Proc. 18th International Conf. on Machine Learning*, pages 282 – 289. Morgan Kaufmann, San Francisco, CA, 2001.
- [68] P. Lanchantin and W. Pieczynski. Unsupervised segmentation of non stationary Markov chains. *soumis à IEEE Transactions of Information Theory*.
- [69] P. Lanchantin and W. Pieczynski. Arbres de Markov triplet et théorie de l'évidence. In *Actes du colloque GRETSI'03*, Paris, France, 8-11 Septembre 2003.
- [70] P. Lanchantin and W. Pieczynski. Unsupervised non stationary image segmentation using triplet markov chain. In *Proceedings of Advanced Concepts for Intelligent Vision Systems (ACIVS 2004)*, Brussels, Belgium, August 31-September 3 2004.

- [71] P. Lanchantin and W. Pieczynski. Unsupervised restoration of non stationary markov chain using evidential priors. *IEEE Transactions on Signal Processing*, 53(8) :3091–3098, 2005.
- [72] P. Lanchantin and F. Salzenstein. Segmentation d’images multispectrales par arbre de Markov caché flou. In *Actes du colloque GRETSI’05*, Louvain-la-neuve, Belgique, 6-9 Septembre 2005.
- [73] J. Lapuyade-Lahorgue and W. Pieczynski. Unsupervised segmentation of hidden semi-markov non stationary chains. In *Twenty six International Workshop on Bayesian Inference and Maximum Entropy in Science and Engineering (MaxEnt06)*, France, July 8-13 2006.
- [74] W. Y. Liu, I. E. Magnin, and G. Gimenez. Un nouvel opérateur pour la détection de ruptures dans les signaux bruités. *Traitement du Signal*, 12(3) :225–236, 1995.
- [75] A.S. Marchenko and V.A. Ogorodnikov. The modelling of very long stationary gaussian sequences with an arbitrary correlation function. *USSR Comput. Maths. Math. Phys.*, 24(5) :141–144, 1984.
- [76] J. Marroquin, S. Mitter, and T. Poggio. Probabilistic solution of ill-posed problems in computational vision. *Journal of The American Statistical Association*, (82) :76–89, 1987.
- [77] H. Maître. Traitement des images. In H. Maître, editor, *Traitement du signal et de l’image*. Hermès, 2003.
- [78] G. J. McLachlan and T. Krishnan. Wiley series in probability and statistics. In *The EM algorithm and extensions*. Wiley Interscience, 1996.
- [79] M. Mignotte, C. Collet, P. Perez, and P. Bouthemy. Unsupervised markovian segmentation of sonar images. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP’97)*, 1997.
- [80] M. Mignotte, C. Collet, P. Pérez, and P. Bouthemy. Hybrid genetic optimization and statistical model-based approach for the classification of shadow shapes in sonar imagery. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(2) :129–141, February 2000.
- [81] M. Mignotte, C. Collet, P. Pérez, and P. Bouthemy. Markov random field and fuzzy logic Modeling in Sonar Imagery Application to the Classification of Underwater-Floor. *Journal of Computer Vision and Image Understanding*, 79 :4–24, july 2000.
- [82] M. Mignotte, Ch. Collet, P. Pérez, and P. Bouthemy. Three-class Markovian segmentation of high resolution sonar images. *Journal of Computer Vision and Image Understanding*, 76(3) :191–204, Decembre 1999.
- [83] E. Monfrini, J. Lecomte, F. Desbouvries, and W. Pieczynski. Image and signal restoration using pairwise Markov trees. In *IEEE Workshop on Statistical Signal Processing (SSP 2003)*, Saint Louis, Missouri, September, 28-October 1 2003.

- [84] R. Neal and G. Hinton. A view of the EM algorithm that justifies incremental, sparse, and other variants. In M.Jordan, editor, *Learning in Graphical Models*, pages 355–368, Dordrecht, 1998. Luwer Academic Publisher.
- [85] P. Nicolas, L. Bize, F. Muri-Majoube, M. Hoebeke, F. Rodolphe, S. Dusko Ehrlich, P. Brum, and P.Bessières. Mining bacillus subtilis chromosome heterogeneities using hidden Markov models. *Nucleic Acid Research*, 30 :1418–1426, 2002.
- [86] N.R. Pal and S.K. Pal. A review on image segmentation techniques. *Pattern Recognition*, 26(9) :1277–1294, 1993.
- [87] K.K. Paliwal. Use of temporal correlation between successive frames in hidden Markov model based speech recognizer. In *Proc. ICASSP*, pages 215–218, 1993.
- [88] C. Pellot, A. Herment, M. Sigelle, P. Horain, H. Maître, and P. Peronneau. A 3D reconstruction of vascular structures from two X-ray angiograms using an adapted annealing algorithm. *IEEE Trans. on Medical Imaging*, 13 :48–60, 1994.
- [89] A. Peng and W. Pieczynski. Adaptative mixture estimation and unsupervised local Bayesian image segmentation. *Graphical Models and Image Processing*, 57(5) :389–399, 1995.
- [90] P. Pérez, A. Chardin, and J.-M. Laferté. Noniterative manipulation of discrete energy-based models for image analysis. *Pattern Recognition*, 33(4) :573–586, 2000.
- [91] W. Pieczynski. Statistical image segmentation. *Machine Graphics and Vision*, 1(1/2) :261–268, 1992.
- [92] W. Pieczynski. Champs de Markov cachés et estimation conditionnelle itérative. *Traitement du Signal*, 11(2) :141–153, 1994.
- [93] W. Pieczynski. Arbres de Markov couple. *CRAS*, 335 :79–82, 2002.
- [94] W. Pieczynski. Arbres de markov couple, pairwise markov trees. *Comptes Rendus de l'Académie des Sciences - Mathématiques*, 335(1) :79–82, 2002.
- [95] W. Pieczynski. Chaînes de Markov triplet - Triplet Markov chains. *Comptes Rendus de l'Académie des Sciences - Mathématiques, Paris*, (I 335) :275–278, 2002.
- [96] W. Pieczynski. Modèles de Markov en traitement d'images. *Traitement du Signal*, 20(3) :255–278, 2003.
- [97] W. Pieczynski. Pairwise Markov chains. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25(5) :634–639, 2003.
- [98] W. Pieczynski. Triplet Partially Markov Chains and Tree. In *2nd International Symposium on Image/Video Communications over fixed and mobile networks (ISIVC'04)*, Brest, France, 7-9 July 2004.
- [99] W. Pieczynski. Copules gaussiennes dans les chaînes triplet partiellement de markov - gaussian copulas in triplet partially markov chains. *Comptes Rendus de l'Académie des Sciences - Mathématique*, 341(3) :189–194, 2005.

- [100] W. Pieczynski. Multisensor triplet Markov chains and theory of evidence. *International Journal of Approximate Reasoning*, 2006, à paraître.
- [101] W. Pieczynski and P. Lanchantin. Restoring hidden non stationary process using triplet partially Markov chain with long memory noise. In *Proceedings of Statistical Signal Processing (SSP2005)*, Bordeaux, France, July 17-20 2005.
- [102] W. Pieczynski and A.-N. Tebbache. Pairwise Markov random fields and segmentation of textured images. *Machine Graphics and Vision*, 9(3) :705–718, 2000.
- [103] J.-N. Provost, C. Collet, P. Rostaing, P. Pérez, and P. Bouthemy. Hierarchical Markovian segmentation of multispectral images for reconstruction of water depth maps. *Computer Vision and Image Understanding*, 93(2) :155–174, 2004.
- [104] P. Pérez. Modèles et algorithmes pour l’analyse probabiliste des images. *Habilitation à diriger des recherches de l’Université de Rennes I*, Décembre 2003.
- [105] L.R. Rabiner. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of IEEE*, (2) :257–286, 1989.
- [106] C. Raphael. Automatic segmentation of acoustic musical signals using hidden Markov models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(4) :360–370, 1999.
- [107] F. Salzenstein, C. Collet, and M. Petremand. Champs de Markov flous pour imagerie multispectrale - fuzzy Markov random fields for multispectral images. *Traitement du Signal*, 21(1) :37–55, 2004.
- [108] F. Salzenstein and W. Pieczynski. Parameter estimation in hidden fuzzy Markov random fields and image segmentation. *Graphical Models and Image Processing*, 59(4) :205–220, 1997.
- [109] F. Salzenstein and W. Pieczynski. Sur le choix de méthode de segmentation statistique d’images. *Traitement du Signal*, 15(2) :119–128, 1998.
- [110] G. Shafer. *A mathematical theory of evidence*. Princeton, 1976.
- [111] P. Smets and R. Kennes. The transferable belief model. *Artificial Intelligence*, 66 :191–234, 1994.
- [112] S. Solimini and J.-M. Morel. Variational methods in image segmentation. In *Progress in Nonlinear Differential Equations and their Applications*, volume 14. Birkhauser, Boston, 1995.
- [113] L. C. Thomas, D. E. Allen, and N. Morkel-Kingsbury. A hidden Markov chain model for the term structure of bond credit risk spreads. *International Review of Financial Analysis*, 11(3) :311–329, 2002.
- [114] M. Thyer and G. Kuczera. Modeling long-term persistence in hydroclimatic time series using a hidden state markov model. *Water Resources Research*, 36 :3301, 2000.



- [115] F. Tupin, I. Bloch, and H. Maître. A first step towards automatic interpretation of SAR images using evidential fusion of several structure detectors. *IEEE Transactions on Geoscience and Remote Sensing*, 37(3) :1327–1343, 1999.
- [116] A. J. Viterbi. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, 13 :260 – 269, April 1967.
- [117] G. Wei and M. Tanner. A Monte-carlo implementation of the EM algorithm and the poor man’s data augmentation algorithms. *J. American Stat. Assoc.*, 85 :699–704, 1990.
- [118] C.J. Wellekens. Explicit time correlation in hidden Markov models for speech recognition. In *Proceedings ICASSP’87*, pages 384–386, 1987.
- [119] J. Whittaker. *Graphical models in applied multivariate statistics*. Wiley, 1990.
- [120] C.F.J. Wu. On the convergence properties of the EM algorithm. *Annals of Statistics*, 11 :95–103, 1983.
- [121] L.A. Zadeh. Fuzzy sets. *Information and control*, 8 :338–352, 1965.





## Résumé

L'objectif de cette thèse est de proposer des méthodes originales de segmentation non supervisées de signaux et d'images fondées sur les modèles de Markov triplets et couples partiellement de Markov. Nous décrivons dans un premier temps les différents modèles, de généralités croissantes et nous développons les algorithmes d'inférence et d'estimation des paramètres dans le cas de processus monodimensionnels (chaînes). Dans un deuxième temps, nous proposons et étudions des cas particuliers de chaînes triplets partiellement de Markov en commençant par un modèle de chaînes couples partiellement de Markov pour la segmentation de processus gaussiens centrés à bruit à corrélation longue.

Puis, nous proposons un modèle de chaînes triplets adapté à la segmentation de processus cachés non stationnaires. Nous étudions également les possibilités d'extension de modèles probabilistes classiques (chaînes et arbres) à un modèle « évidentiel », avec la loi a posteriori du processus caché donné par la fusion de Dempster-Shafer et à un modèle « flou » dans lequel la fonction d'appartenance aux classes est floue.

**Mots clés** Estimation bayésienne, modèles de Markov cachés (chaînes et arbres), modèles de Markov couples et triplets, Espérance-Maximisation (EM), estimation conditionnelle itérative (ECI), théorie des croyances, fusion de Dempster-Shafer, théorie des ensembles flous, segmentation de processus non stationnaires, segmentation de processus gaussien centrés à corrélation longue.

\*

## Abstract

The aim of this thesis is to propose original methods of unsupervised signal and image segmentation, based on triplet and pairwise partially Markov models. We first describe different models with increasing generality and develop inference and parameters estimation algorithms in the monodimensional case (chains). Then we propose and study particular cases of triplet partially Markov chains, starting with a model of pairwise partially Markov chains for the segmentation of centered gaussian processes with long correlation noise.

Finally, we propose a triplet Markov chains model adapted to the segmentation of non stationary hidden processes. We also study the extension possibilities of classical probabilistic models (chains and trees) in an evidential model, where the posterior hidden process distribution is given by the Dempster-Shafer fusion and in a “fuzzy” model in which the membership function is fuzzy.

**Key Words** Bayesian estimation, Hidden Markov Models (chains and trees), Pairwise and Triplet Markov Models, Esperance-Maximisation (EM), Iterative Conditional Estimation (ICE), Evidence Theory, Dempster-Shafer fusion, Fuzzy set theory, Non-stationary processes segmentation, Long-correlation centered Gaussian processes segmentation.