



HAL
open science

Modèles graphiques évidentiels

Mohamed El Yazid Boudaren

► **To cite this version:**

Mohamed El Yazid Boudaren. Modèles graphiques évidentiels. Autre. Institut National des Télécommunications; Université des Sciences et de la Technologie Houari-Boumediène (Alger), 2014. Français. NNT : 2014TELE0001 . tel-01004504

HAL Id: tel-01004504

<https://theses.hal.science/tel-01004504>

Submitted on 11 Jun 2014

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

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



THESE DE DOCTORAT CONJOINT TELECOM SUDPARIS et L'UNIVERSITE PIERRE ET MARIE CURIE

En cotutelle avec l'Université des Sciences et de la Technologie Houari Boumediene,
Alger, Algérie

Spécialité : Systèmes Intelligents et Ingénierie des Logiciels

Ecole doctorale : Informatique, Télécommunications et Electronique de Paris

Présentée par

Mohamed El Yazid BOUDAREN

**Pour obtenir le grade de
DOCTEUR DE TELECOM SUDPARIS**

MODELES GRAPHIQUES EVIDENTIELS

**Soutenu le 12 janvier 2014 à l'Université des Sciences et de la Technologie Houari Boumediene,
Alger, devant le jury composé de :**

Président	Slimane LARABI,	Professeur, USTHB, Alger
Directeur de thèse	Amar AÏSSANI,	Professeur, USTHB, Alger
Directeur de thèse	Wojciech PIECZYNSKI,	Professeur, Télécom SudParis, Evry
Rapporteur	Amina SERIR,	Professeure, USTHB, Alger
Rapporteur	Stéphane DERRODE,	MC HDR, ECM, Marseille
Examineur	Pr. Aziz BENLARBI-DELAÏ,	Professeur, UPMC Paris
Invité	Emmanuel MONFRINI,	MC, Télécom SudParis, Evry

Thèse n°2014TELE0001

Résumé

Les modélisations par chaînes de Markov cachées permettent de résoudre un grand nombre de problèmes inverses se posant en traitement d'images ou de signaux. En particulier, le problème de segmentation figure parmi les problèmes où ces modèles ont été le plus sollicités. Selon ces modèles, la donnée observable est considérée comme une version bruitée de la segmentation recherchée qui peut être modélisée à travers une chaîne de Markov à états finis. Des techniques bayésiennes permettent ensuite d'estimer cette segmentation même dans le contexte non-supervisé grâce à des algorithmes qui permettent d'estimer les paramètres du modèle à partir de l'observation seule. Les chaînes de Markov cachées ont été ultérieurement généralisées aux chaînes de Markov couples et triplets, lesquelles offrent plus de possibilités de modélisation tout en présentant des complexités de calcul comparables, permettant ainsi de relever certains défis que les modélisations classiques ne supportent pas. Un lien intéressant a également été établi entre les modèles de Markov triplets et la théorie de l'évidence de Dempster-Shafer, ce qui confère à ces modèles la possibilité de mieux modéliser les données multi-senseurs. Ainsi, dans cette thèse, nous abordons trois difficultés qui posent problèmes aux modèles classiques : la non-stationnarité du processus caché et/ou du bruit, la corrélation du bruit et la multitude de sources de données. Dans ce cadre, nous proposons des modélisations originales fondées sur la très riche théorie des chaînes de Markov triplets.

Dans un premier temps, nous introduisons les chaînes de Markov à bruit M-stationnaires qui tiennent compte de l'aspect hétérogène des distributions de bruit s'inspirant des chaînes de Markov cachées M-stationnaires. Les chaînes de Markov cachées ML-stationnaires, quant à elles, considèrent à la fois la loi *a priori* et les densités de bruit non-stationnaires.

Dans un second temps, nous définissons deux types de chaînes de Markov couples non-stationnaires. Dans le cadre bayésien, nous introduisons les chaînes de Markov couples M-stationnaires puis les chaînes de Markov couples MM-stationnaires qui considèrent la donnée stationnaire par morceau. Dans le cadre évidentiel, nous définissons les chaînes de Markov couples évidentielles modélisant l'hétérogénéité du processus caché par une fonction de masse.

Enfin, nous présentons les chaînes de Markov multi-senseurs non-stationnaires où la fusion de Dempster-Shafer est employée à la fois pour modéliser la non-stationnarité des données (à l'instar des chaînes de Markov évidentielles cachées) et pour fusionner les informations provenant des différents senseurs (comme dans les champs de Markov multi-senseurs).

Pour chacune des modélisations proposées, nous décrivons les techniques de segmentation et d'estimation des paramètres associées. L'intérêt de chacune des modélisations par rapport aux modélisations classiques est ensuite démontré à travers des expériences menées sur des données synthétiques et réelles.

Mots clés : Chaînes de Markov cachées, chaînes de Markov couples, chaînes de Markov triplets, segmentation non-supervisée, estimation bayésienne, maximum marginal *a posteriori*, estimation des paramètres, espérance-maximisation, données non-stationnaires, bruit corrélé, données multi-senseurs, théorie de l'évidence, fusion de Dempster-Shafer.

Abstract

Hidden Markov chains (HMCs) based approaches have been shown to be efficient to resolve a wide range of inverse problems occurring in image and signal processing. In particular, unsupervised segmentation of data is one of these problems where HMCs have been extensively applied. According to such models, the observed data are considered as a noised version of the requested segmentation that can be modeled through a finite Markov chain. Then, Bayesian techniques such as MPM can be applied to estimate this segmentation even in unsupervised way thanks to some algorithms that make it possible to estimate the model parameters from the only observed data. HMCs have then been generalized to pairwise Markov chains (PMCs) and triplet Markov chains (TMCs), which offer more modeling possibilities while showing comparable computational complexities, and thus, allow to consider some challenging situations that the conventional HMCs cannot support. An interesting link has also been established between the Dempster-Shafer theory of evidence and TMCs, which give to these latter the ability to handle multisensor data. Hence, in this thesis, we deal with three challenging difficulties that conventional HMCs cannot handle: nonstationarity of the *a priori* and/or noise distributions, noise correlation, multisensor information fusion. For this purpose, we propose some original models in accordance with the rich theory of TMCs.

First, we introduce the M-stationary noise- HMC (also called jumping noise- HMC) that takes into account the nonstationary aspect of the noise distributions in an analogous manner with the switching-HMCs. Afterward, ML-stationary HMC consider nonstationarity of both the *a priori* and/or noise distributions.

Second, we tackle the problem of non-stationary PMCs in two ways. In the Bayesian context, we define the M-stationary PMC and the MM-stationary PMC (also called switching PMCs) that partition the data into M stationary segments. In the evidential context, we propose the evidential PMC in which the realization of the hidden process is modeled through a mass function.

Finally, we introduce the multisensor nonstationary HMCs in which the Dempster-Shafer fusion has been used on one hand, to model the data nonstationarity (as done in the hidden evidential Markov chains) and on the other hand, to fuse the information provided by the different sensors (as in the multisensor hidden Markov fields context).

For each of the proposed models, we describe the associated segmentation and parameters estimation procedures. The interest of each model is also assessed, with respect to the former ones, through experiments conducted on synthetic and real data.

Keywords: Hidden Markov chains, pairwise and triplet Markov chains, unsupervised segmentation, Bayesian estimation, marginal posterior mode, parameters estimation, expectation-maximization, nonstationary data, correlated noise, multisensor data, theory of evidence, Dempster-Shafer fusion.

Sommaire

Notations	vi
Introduction générale	1
Chapitre 1. Généralités	4
1.1. Approche bayésienne pour le problème de segmentation.....	5
1.1.1. Modélisation probabiliste du problème de segmentation.....	5
1.1.2. Stratégie bayésienne et fonction de coût	5
1.1.3. Estimateurs bayésiens et algorithmes d'inférence	6
1.2. Estimation des paramètres.....	8
1.2.1. Algorithme Espérance-Maximisation (EM)	8
1.2.2. Algorithme EM stochastique (SEM)	10
1.2.3. Estimation conditionnelle itérative (ECI).....	10
1.3. Conclusion.....	11
Chapitre 2. Chaînes de Markov cachées, couples et triplets	12
2.1. Chaînes de Markov cachées	12
2.1.1. Définition du modèle	13
2.1.2. Simulation d'une chaîne de Markov cachée gaussienne.....	13
2.1.3. Inférence dans les chaînes de Markov cachées	14
2.2. Chaînes de Markov couples	15
2.2.1. Définition du modèle	15
2.2.2. Simulation d'une chaîne de Markov couple gaussienne.....	17
2.2.3. Inférence dans les chaînes de Markov couples	17
2.2.4. Estimation des paramètres d'une chaîne de Markov couple	18
2.3. Chaînes de Markov triplets	19
2.3.1. Définition du modèle général	19
2.3.2. Travaux antérieurs.....	19
2.3.3. Chaînes de Markov cachées M-stationnaires	20
2.4. Conclusion.....	23
Chapitre 3. Chaînes de Markov et théorie de l'évidence	24
3.1. Généralités sur la théorie de l'évidence.....	24
3.2. Chaînes de Markov évidentielles cachées.....	27
3.2.1. Chaînes de Markov évidentielles	28
3.2.2. Définition d'une chaîne de Markov cachée évidentielle	28
3.2.3. Estimation des paramètres d'une CMEC	30
3.3. Sur le choix de la modélisation dans le cas d'un signal monodimensionnel non-stationnaire.....	30
3.3.1. Segmentation non-supervisée de CMC non-stationnaires	31
3.3.2. Segmentation non-supervisée d'images non-stationnaires bruitées.....	39
3.4. Chaînes de Markov cachées multi-senseurs	44

3.4.1. Définition du modèle	44
3.4.2. Restauration MPM d'une chaîne de Markov multi-senseurs	45
3.4.3. Estimation des paramètres d'une chaîne de Markov multi-senseurs.....	46
3.5. Conclusion.....	47
Chapitre 4. Chaînes de Markov cachées à bruit M-stationnaire	48
4.1. Définition du modèle.....	48
4.2. Simulation d'une CMC BM-S	49
4.3. Restauration MPM d'une CMC BM-S	49
4.4. Estimation des paramètres d'une CMC BM-S	50
4.5. Résultats expérimentaux	51
4.5.1. Segmentation non-supervisée de chaînes de Markov cachées à bruit M-stationnaires.....	51
4.5.2. Segmentation non-supervisée d'images entachées de bruit M-stationnaire	52
4.5.3. Segmentation non-supervisée d'image réelles via les CMC BM-S.....	58
4.6. Chaînes de Markov cachées ML-stationnaire.....	61
4.6.1. Définition du modèle	61
4.6.2. Simulation d'une chaîne de Markov cachée ML-stationnaire	61
4.6.3. Restauration MPM d'une chaîne de Markov cachée ML-stationnaire.....	62
4.6.4. Estimation des paramètres d'une chaîne de Markov cachée ML-stationnaire	63
4.6.5. Résultats expérimentaux	64
4.7. Conclusion.....	67
Chapitre 5. Chaînes de Markov couples M-stationnaires	68
5.1. Chaîne de Markov couples M-stationnaire	68
5.1.1. Définition du modèle	68
5.1.2. Paramètres et simulation d'une CMCo M-S gaussienne	70
5.1.3. Restauration MPM d'une chaîne de Markov couple M-stationnaire.....	70
5.1.4. Estimation des paramètres d'une chaîne de Markov couple M-stationnaire	71
5.1.5. Etude expérimentale	72
5.2. Chaîne de Markov couples MM-stationnaire	77
5.2.1. Définition du modèle	77
5.2.2. Paramètres d'une CMCo MM-S	77
5.2.3. Restauration MPM d'une chaîne de Markov couple MM-stationnaire.....	78
5.2.4. Estimation des paramètres d'une chaîne de Markov couple MM-stationnaire	79
5.2.5. Application des CMCo MM-S au problème de segmentation des textures	79
5.3. Conclusion.....	85
Chapitre 6. Chaînes de Markov couples évidentielles	86
6.1. Définition du modèle.....	86
6.2. Inférence dans une chaîne de Markov couple évidentielle	87
6.2.1. Restauration MPM d'une chaîne de Markov couple évidentielle	88
6.2.2. Estimation des paramètres d'une chaîne de Markov couple évidentielle.....	88

6.3. Expériences.....	89
6.4. Segmentation non-supervisée de chaînes de Markov couples non-stationnaires	89
6.4.2. Segmentation non-supervisée d'images non-stationnaires entachées de bruit corrélé.....	91
6.5. Conclusion.....	94
Chapitre 7. Chaînes de Markov cachées multi-senseurs non-stationnaires ..	95
7.1. Chaînes de Markov cachées multi-senseurs non-stationnaires.....	95
7.1.1. Définition du modèle	96
7.2. Segmentation non-supervisée d'une chaîne de Markov cachée multi-senseurs non-stationnaire.....	97
7.2.1. Modélisation des données	97
7.2.2. Restauration MPM d'une chaîne de Markov multi-senseurs non-stationnaire	97
7.2.3. Estimation des paramètres d'une chaîne de Markov multi-senseurs non- stationnaire	99
7.3. Résultats expérimentaux	99
7.3.1. Segmentation non-supervisée de chaînes de Markov cachées multi-senseurs non- stationnaires.....	100
7.3.2. Segmentation non-supervisée d'images multi-senseurs non-stationnaire bruitées	102
7.4. Conclusion.....	108

Notations

Symboles

\mathcal{N}	Ensemble de sites
$n \in \mathcal{N}$	site
X, Y, U	Variabes aléatoires (en majuscule) des processus caché, observables et auxiliaire
x, y, u	Leurs réalisations respectives en minuscules
N	Taille (dimension) de la donnée
\mathcal{X}	Espace des états cachés
\mathcal{Y}	Espace des observations
Ω	Ensemble de classes (lorsque \mathcal{X} est fini)
$\omega_1, \dots, \omega_K$	Classes (états cachés)
K	Nombre de classes (cardinal de Ω)
p	Mesure de probabilité
$p(\cdot \cdot)$	Probabilité conditionnelle
κ	Mesure de comptage sur \mathcal{X}
μ	Mesure de référence sur \mathcal{Y}
\hat{s}	Stratégie de classification
L	Fonction de perte
$\mathcal{R}(\hat{s}, L)$	Risque bayésien associé à la stratégie \hat{s} et la fonction de perte L
\hat{s}_B	Décision bayésienne
\mathbb{E}	Espérance mathématique
$\mathbf{1}_{[\cdot]}$	Fonction indicatrice
$Z = (X, Y)$	Données complètes ou complétées (processus couple)
Θ	Espace des paramètres
θ	Jeu de paramètres
Z_θ	Modèle statistique paramétriques
$p(z; \theta)$	Loi du modèle paramétrique
$\mathcal{L}(y; \theta)$	Log-vraisemblance
$\mathcal{L}_c(z; \theta)$	Log-vraisemblance complétée
$p(x)$	$p(X = x)$

\mathbb{N}	Ensemble des nombres naturel
\mathbb{Z}	Ensemble des nombres entiers
\mathbb{R}	Ensemble des nombres réels
μ	Moyenne
σ	Ecart-type
ρ	Coefficient de corrélation
α	Fonction de probabilité progressive
β	Fonction de probabilité rétrograde
Λ	Ensemble de classes auxiliaires
$\lambda_1, \dots, \lambda_M$	Classes auxiliaires
M	Nombre de stationnarités du processus caché
L	Nombre de stationnarités de bruit
$T = (U, X, Y)$	Processus triplet
m	Fonction de masse
$\mathcal{P}(\cdot)$	Ensemble de sous ensemble
Cr	Fonction de crédibilité
Pl	Fonction de plausibilité
\oplus	Opérateur de fusion
\mathcal{K}	Inconsistance de la fusion
Cr^a, m^a	Affaiblissement d'une crédibilité, d'une masse
τ	Taux d'erreur
S_1, S_2, \dots, S_R	Senseurs
R	Nombre de senseurs
y^r	Observation relative au senseur S_r

Acronymes

MAP	Maximum A Posteriori
MPM	Maximum marginal <i>A Posteriori</i>
EMV	Estimateur de maximum de vraisemblance
CMC	Chaîne(s) de Markov cachée(s)
CMCo	Chaîne(s) de Markov couple(s)
CMT	Chaîne(s) de Markov triplet(s)
CMC M-S	Chaîne(s) de Markov cachée(s) M-stationnaire(s)
CME	Chaîne(s) de Markov évidentielle(s)
CMC MS	Chaîne(s) de Markov cachée(s) multi-senseurs
CMEC	Chaîne(s) de Markov évidentielle(s) cachée(s)
CMC BM-S	Chaîne(s) de Markov cachée(s) à bruit M-stationnaire
CMC ML-S	Chaîne(s) de Markov cachée(s) ML-stationnaire(s)
CMCo M-S	Chaîne(s) de Markov couple(s) M-stationnaire(s)
CMCo MM-S	Chaîne(s) de Markov couple(s) MM-stationnaire(s)
CMC MN	Chaîne(s) de Markov cachée(s) multi-senseurs non-stationnaire(s)
DS	Dempster-Shafer
EM	Espérance-Maximisation (en anglais <i>Expectation-Maximization</i>)
SEM	EM stochastique (en anglais <i>Stochastic EM</i>)
ECI	Estimation conditionnelle itérative (en anglais <i>Iterative conditional estimation ICE</i>)

Introduction générale

Les chaînes de Markov cachées (CMC) jouissent d'une grande notoriété dans le domaine du traitement du signal et de l'image. Elles servent, en effet, d'outil de régularisation permettant de prendre des décisions localement à partir de l'ensemble de l'échantillon tout en contournant de manière élégante les problèmes de complexité de calcul prohibitive. Leur succès est dû, principalement, à l'existence d'une panoplie de méthodes associées qui permettent d'effectuer les différentes opérations d'estimation en un temps raisonnable, même lorsque la taille des échantillons est relativement grande.

L'utilisation de ces modèles montre cependant ses limites dans un certain nombre de situations où des données de nature complexe sont concernées. Dans cette thèse, nous abordons trois sources de complexité : les données non-stationnaires, les données à bruit corrélé et les données multi-senseurs. Nous proposons ainsi des modélisations markoviennes plus générales pour pallier aux difficultés rencontrées par les modélisations « classiques » et liées à ces sources de complexité. Pour ce faire, la segmentation des données est choisie comme cadre illustratif pour les différentes modélisations proposées afin d'en exposer la nature et l'intérêt profond par rapport aux modélisations classiques.

Le problème de la segmentation est simple dans sa formulation. Il consiste à partitionner un signal en plusieurs classes présentant certaines propriétés d'homogénéité, à partir d'une estimation de ses caractéristiques cachées basée sur certaines de ses caractéristiques observables. Les approches statistiques bayésiennes, en particulier celles basées sur les CMC, sont parmi les techniques de segmentation les plus utilisées grâce à leur praticabilité pour la segmentation de données de grande dimension. Ces approches sont fondées sur une modélisation probabiliste du phénomène étudié grâce à l'introduction de la loi jointe des processus caché et observé associés au phénomène en question. Une stratégie de segmentation vise à déterminer l'estimateur optimal du signal caché en minimisant la perte moyenne, elle-même définie à partir d'une fonction de coût (ou fonction de perte) permettant de pénaliser les erreurs de classification. L'optimalité de la solution étant liée à la fonction de perte, les méthodes bayésiennes jouissent d'une grande souplesse grâce à la possibilité de choisir la fonction de coût selon des considérations particulières liées à l'application visée. De plus, lorsque les paramètres du modèle ne sont pas connus, des techniques permettant de les estimer à partir des caractéristiques observables du signal existent, ce qui confère aux approches bayésiennes la possibilité d'automatiser les différents traitements. Ces généralités constituant le contexte global de cette thèse seront présentées dans le chapitre 1.

Les approches bayésiennes étant fondées sur la modélisation probabiliste du phénomène d'intérêt, le choix de la loi jointe des processus caché et observable est donc primordiale. La définition de cette loi est liée à la structure de dépendance du signal à modéliser. D'une manière globale, plus cette structure est complexe, plus l'adéquation avec le phénomène étudié peut-être affinée, et par conséquent, la segmentation est de meilleure qualité, mais en contrepartie, les différents calculs, requis pour l'estimation des données cachées, peuvent devenir impraticables.

Un compromis s'impose donc entre richesse du formalisme et complexité du calcul. Par ailleurs, des hypothèses simplificatrices sont émises pour permettre d'alléger la complexité des calculs nécessaires aux différentes procédures d'estimation tout en maintenant une concordance appropriée entre le modèle et le phénomène d'intérêt. Dans une CMC, le processus caché est supposé de Markov, alors que les observations sont souvent supposées indépendantes conditionnellement au processus caché dans le but de maintenir la Markovianité du processus caché *a posteriori*. Bien que ces hypothèses fortes puissent se révéler inadaptées à certaines situations, il est primordial de préserver la markovianité *a posteriori* du processus caché nécessaire à la segmentation et qui n'est assurée que sous certaines conditions.

Afin de relâcher certaines hypothèses trop contraignantes, les CMC ont été généralisées aux chaînes de Markov couples (CMCo). Dans de tels modèles, la loi jointe est supposée de Markov, ce qui permet d'assurer la markovianité du processus caché *a posteriori* tout en maintenant la faisabilité des différents calculs d'estimation. La suprématie des CMCo sur les CMC est donc sans doute. Les CMCo ont, de leur côté, été étendues aux chaînes de Markov triplet (CMT). Dans ces dernières, un troisième processus adjacent est utilisé pour prendre en compte certaines propriétés de la donnée à modéliser. Ces différents modèles, ainsi que leurs traitements associés, seront décrits dans le chapitre 2. Dans ce même chapitre, le formalisme d'une CMT particulière appelée chaîne de Markov cachée M-stationnaire est également brièvement présentée.

Un lien intéressant a été établi entre les CMT et la théorie de l'évidence de Dempster-Shafer. En effet, la distribution *a posteriori* du processus caché peut être vue comme la fusion de l'information *a priori* et de l'observation. Toutefois, lorsque l'une de ces deux informations n'est pas bayésienne, cette fusion détruit la markovianité requise pour les différents calculs. Il a cependant été démontré que la distribution résultante d'une telle fusion est une distribution marginale d'une CMT, ce qui permet d'envisager les différentes méthodes d'estimation. Les mesures bayésiennes impliquées dans les modèles de Markov cachés ont donc été étendues aux mesures évidentielles et permettent ainsi de manipuler des données plus complexes. En particulier, les modèles de Markov ont été adaptés, grâce à la théorie de l'évidence, aux données non-stationnaires et aux données multi-senseurs. Dans, le chapitre 3, la théorie de l'évidence ainsi que son lien avec la théorie Markovienne sont discutés. L'utilisation des modèles de Markov évidentiels dans le cas de données non-stationnaires et multi-senseurs est également présentée.

Les CMC M-stationnaires peuvent être vues comme des CMC stationnaires par morceaux. Le processus auxiliaire modélise ainsi les changements (ou sauts) de régime. Cette situation peut se présenter par exemple dans le cas des images de textures où il devient possible d'allouer à chaque texture un modèle qui lui est propre. Dans cette étude, nous proposons d'étendre ces modèles dans deux directions différentes. La première extension consiste à considérer le cas de sauts de bruit pour modéliser les bruits non stationnaires. Le modèle correspondant, appelé CMC à bruit M-stationnaires, et dans lequel, le processus auxiliaire peut, par exemple, modéliser les conditions d'éclairage dans une image, est présenté dans le chapitre 4. La deuxième extension consiste à considérer le cas de bruit corrélé où les modèles de CMCo ont montré leur supériorité. Nous proposons de ce fait, d'étendre le modèle de CMC M-stationnaire au CMCo M-stationnaire.

Le formalisme de dernier modèle ainsi que les traitements qui lui sont associés sont décrits dans le chapitre 5.

A l'instar des CMC, les CMCo peuvent être étendues aux CMCo évidentielles pour une modélisation plus appropriées des données non-stationnaires à bruit corrélé. L'objectif de cette extension est de confirmer l'intérêt que présentent les modèles de CMCo par rapport aux CMC, déjà validées dans leurs versions classiques et M-stationnaires. Les fondements théoriques ainsi que les résultats expérimentaux des CMCo évidentielles sont présentés dans le chapitre 6.

L'introduction des mesures non-bayésiennes dans les CMC a permis, jusqu'ici, de modéliser deux types de données complexes : les données non-stationnaires ou les données multi-senseurs. Dans le chapitre 7 où les chaînes de Markov non-stationnaires multi-senseurs sont présentées, nous considérons des données qui présentent les deux propriétés en même temps. Ces modèles sont particulièrement bien adaptés au traitement d'images multi-capteurs.

Organisation du manuscrit Dans le chapitre 1, nous présentons le problème de segmentation qui constituera le cadre illustratif de notre étude. Nous décrivons également les principes de la segmentation bayésienne et les différents algorithmes d'estimation des paramètres.

Dans le chapitre 2, nous décrivons les aspects théoriques des différents modèles qui constituent la base de notre étude, et qui sont, par ordre de complexité croissante : les CMC, les CMCo et les CMT. Les CMC M-stationnaires, qui sont des CMT particulières, sont également brièvement présentées pour illustrer l'intérêt des CMT.

Le chapitre 3, quant à lui, sera consacré à l'introduction de la théorie de l'évidence au sein des modèles de Markov. Notre contribution dans ce chapitre porte, premièrement, sur la comparaison entre les CMC évidentielles et les CMC M-stationnaires décrites dans le chapitre 2 pour la modélisation des données non-stationnaires, et deuxièmement, sur l'adaptation des champs de Markov cachés multi-senseurs au contexte de chaînes de Markov cachées.

Les chapitres suivants sont intégralement originaux, et seront dédiés à la présentation de modèles inédits. Dans le chapitre 4, nous présentons le modèle de CMC à bruit M-stationnaire. Dans le chapitre 5, nous étendons les CMC M-stationnaire au contexte des CMCo. Le chapitre 6 présentera les CMCo évidentielles qui généralisent les CMCo en considérant des distributions *a priori* évidentielles. Dans le chapitre 7, nous présentons les chaînes de Markov non-stationnaire multi-senseurs.

Enfin, nous clôturons ce manuscrit par une conclusion générale, dans laquelle nous établissons un premier bilan de nos contributions et proposons quelques perspectives de poursuite de cette étude.

Chapitre 1. Généralités

La segmentation des données figure parmi les problèmes les plus importants en traitement d'informations. Dans certaines situations le résultat d'une segmentation fournit directement la solution à un problème posé ; dans d'autres, il sert de point de départ à d'autres traitements. Dans tous les cas, la qualité de la segmentation influe directement sur le processus de prise de décision.

Le principe général d'une segmentation est de partitionner la donnée en plusieurs sous-ensembles présentant certaines propriétés d'homogénéité. Parmi les exemples les plus rencontrés dans la littérature, nous pouvons citer la segmentation d'images où deux grandes familles d'approches sont adoptées : les approches basées régions, qui cherchent à regrouper des sites ayant des propriétés communes, et les approches basées contours, qui recherchent les frontières délimitant les régions d'intérêt. Le travail de recherche mené dans cette thèse s'inscrit dans la première famille d'approches et est fondé sur une approche bayésienne reposant sur la modélisation des phénomènes cachés et observés, le phénomène caché étant la partition recherchée. Le phénomène caché n'étant pas directement déductible à partir du phénomène observé, l'approche bayésienne de segmentation repose sur la modélisation du lien entre ces deux phénomènes à travers une loi de probabilité. Cette dernière constitue la connaissance avant toute expérience sur le lien entre les deux phénomènes concernés. Les estimateurs bayésiens permettent par la suite, d'estimer le phénomène caché à partir du phénomène observable. De ce fait, le bon choix de la loi jointe modélisant le lien entre ces deux phénomènes est primordial. Une modélisation fidèle du phénomène en question requiert en général d'adopter une distribution assez complexe. Toutefois, les calculs nécessaires pour l'estimation des distributions *a posteriori* du phénomène caché explosent quand la dimension de la donnée augmente. Un compromis s'impose donc entre richesse et simplicité de cette loi. En effet, des hypothèses simplificatrices sont souvent adoptées pour alléger la complexité de la modélisation.

L'un des points forts de l'approche bayésienne est la possibilité d'adapter la notion d'optimalité du résultat de la segmentation à des préoccupations particulières à travers la définition de la fonction de coût qui peut également être vue comme une règle de décision, ce qui confère aux approches bayésiennes une grande souplesse.

La loi jointe modélisant la donnée en question dépend, dans la plupart des cas, d'un jeu de paramètres. L'estimation du phénomène caché suppose ces paramètres connus. Cependant, lorsque tel n'est pas le cas, des algorithmes d'estimation permettent d'estimer ces derniers d'une manière itérative.

Ce chapitre comporte deux volets. Le premier volet est consacré aux fondements de l'inférence bayésienne dont l'objectif est de modéliser ce que l'on cherche à inférer en se donnant une loi *a priori* dans le cas général. Ces fondements seront étudiés dans le cas particulier du problème de segmentation. Le second volet, quant à lui, aborde le problème de l'estimation du jeu de paramètres correspondant à la loi de probabilité définissant le lien entre les phénomènes cachés et observés.

1.1. Approche bayésienne pour le problème de segmentation

L'étude probabiliste des phénomènes s'introduit de manière naturelle lorsqu'il existe une incertitude sur la mesure décrivant un phénomène. Dans le problème de segmentation de données, lorsqu'on cherche à étudier les liens entre les deux phénomènes observé et caché, il existe, en dehors de la théorie des probabilités, deux possibilités : soit un lien déterministe, soit aucun lien [Hil11]. Le calcul des probabilités permet d'introduire une infinité de liens intermédiaires, d'où l'intérêt de l'approche probabiliste pour la modélisation du problème de segmentation.

1.1.1. Modélisation probabiliste du problème de segmentation

La segmentation peut être vue comme un problème où il s'agit de rechercher des caractéristiques cachées à partir de données observables. En l'absence d'un lien déterministe entre les deux données, plusieurs segmentations possibles peuvent correspondre à une même observation. L'adoption d'une approche statistique semble donc un moyen adéquat. En effet, l'approche bayésienne de segmentation repose sur la modélisation du lien entre les deux phénomènes caché et observé.

Soit \mathcal{N} l'ensemble de sites (de cardinal N). En chaque site $n \in \mathcal{N}$, nous considérons x_n (resp. y_n) comme la réalisation d'une variable aléatoire X_n (resp. Y_n) prenant ses valeurs dans \mathcal{X} (resp. \mathcal{Y}). Par conséquent, nous définissons le processus aléatoire caché $X = \{X_n\}_{n \in \mathcal{N}}$, le processus aléatoire observable $Y = \{Y_n\}_{n \in \mathcal{N}}$ et le processus couple $Z = (X, Y)$ représentant la donnée complète. Le problème de segmentation se ramène donc à estimer la réalisation x appelée « réalité terrain » à partir de l'observation $Y = y$. Tout estimateur sera dit stratégie de classification. Si l'ensemble \mathcal{X} est fini, nous parlons de problème de classification. Un tel ensemble sera dit ensemble de classes et sera noté $\Omega = \{\omega_1, \dots, \omega_K\}$.

Notons κ la mesure de comptage sur \mathcal{X} et μ une mesure de référence sur \mathcal{Y} et supposons que le processus Z admet une densité de probabilité $p(z)$ par rapport à la mesure produit $(\kappa \times \mu)^N$. Cette densité modélise donc tous les liens probabilistes entre le processus X (phénomène caché) et le processus Y (phénomène observé) et il est possible d'associer à toute observation y une connaissance sur la vérité terrain x à travers la probabilité dite *a posteriori* définie sur \mathcal{X}^N et donnée par

$$p(x|y) = \frac{p(x,y)}{p(y)} \quad (1.1)$$

1.1.2. Stratégie bayésienne et fonction de coût

Sur la base de la densité de probabilité $p(z)$ relative à la mesure $(\kappa \times \mu)^N$, une stratégie de classification \hat{s} associe à toute réalisation y sur \mathcal{Y}^N , une réalisation $\hat{s}(y) = \hat{x}$ sur \mathcal{X}^N appelée classification. A une classification \hat{s} , nous associons une application L définie de $\mathcal{X}^N \times \mathcal{X}^N$ vers \mathbb{R}^+ appelée fonction de perte ou de coût. Cette fonction mesure l'impact d'une erreur de classification (perte ou coût éventuellement) d'une mauvaise estimation de la vérité terrain et associe à toute classification $x' \in \mathcal{X}^N$ la perte $L_{x,x'}$ qui lui est associée par rapport à la vérité terrain recherchée x . Une telle fonction doit vérifier

$$\forall (x, x') \in \mathcal{X}^N \times \mathcal{X}^N, [L(x', x) = 0] \Leftrightarrow [x' = x] \quad (1.2)$$

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

$$\mathcal{R}(\hat{s}, L) = \mathbb{E}[L(\hat{s}(Y), X)] \quad (1.3)$$

Ainsi, la loi $p(z)$ étant définie, adopter une fonction de perte L permet de quantifier la qualité d'une classification grâce à la mesure du risque bayésien de la formule (1.3).

Par ailleurs, la décision bayésienne \hat{s}_B associée à la fonction de perte L est la stratégie qui minimise la perte moyenne donnée par (1.3)

$$\mathcal{R}(\hat{s}_B, L) = \min_{\hat{s}} \mathcal{R}(\hat{s}, L) \quad (1.4)$$

Pour une observation y donnée, la décision bayésienne qui lui est associée est retrouvée en minimisant la perte moyenne conditionnelle à y .

$$\mathbb{E}[L(\hat{s}_B(Y), X)|Y = y] = \min_{\hat{s}} \mathbb{E}[L(\hat{s}(Y), X)|Y = y] \quad (1.5)$$

L'estimateur bayésien optimal associé à la fonction de perte L est donc obtenu par la formule

$$\hat{s}_B(y) = \arg \min_{x' \in \mathcal{X}^N} \sum_{x \in \mathcal{X}^N} L_{x', x} p(x|y) \quad (1.6)$$

La formule (1.6) illustre la souplesse des méthodes bayésiennes dans la mesure où devant une situation donnée, modélisée à travers la loi $p(z)$, l'optimalité d'une classification est liée à la définition de la fonction de perte dont le choix dépendra de l'application visée et des préoccupations de l'utilisateur. Dans le cas particulier d'un système de détection d'incendie, il est possible d'exprimer la préférence d'avoir une fausse alarme plutôt que de ne pas remarquer le déclenchement d'un incendie.

Dans la pratique, il arrive souvent que la fonction de perte soit exprimée à travers des fonctions de pertes définies localement au niveau de chaque site. Dans ce cas, un estimateur bayésien est associé à chaque site, toutefois, l'estimation de la classe associée à ce site se fait toujours en exploitant toute l'information (sauf simplification particulière).

1.1.3. Estimateurs bayésiens et algorithmes d'inférence

Dans ce qui suit, nous définissons deux estimateurs bayésiens fondés sur la fonction de perte $L_{0/1}$ définie comme suit

$$\forall (i, j) \in \mathcal{J}^2, L_{0/1}(i, j) = \mathbf{1}_{[i \neq j]} \quad (1.7)$$

1.1.3.1. Estimateur MAP

L'estimateur bayésien MAP est associé à la fonction de perte $L_{0/1}$ définie sur l'ensemble des configurations possibles des classifications \mathcal{X}^N :

$$\forall (x', x) \in \mathcal{X}^N \times \mathcal{X}^N, L_{MAP}(x', x) = \mathbf{1}_{[x' \neq x]} \quad (1.8)$$

Selon la formule (1.6), l'estimateur bayésien MAP est donc donné par la formule

$$\begin{aligned}
\hat{x}_{MAP}(y) &= \arg \min_{x' \in \mathcal{X}^N} \sum_{x \in \mathcal{X}^N} \mathbf{1}_{[x' \neq x]} p(x|y) \\
&= \arg \min_{x' \in \mathcal{X}^N} \sum_{x \in \mathcal{X}^N / \{x'\}} p(x|y) & (1.9) \\
&= \arg \min_{x' \in \mathcal{X}^N} \{1 - p(x'|y)\} \\
&= \arg \max_{x' \in \mathcal{X}^N} p(x'|y)
\end{aligned}$$

Cet estimateur est particulièrement sévère vu qu'il pénalise de la même façon une classification présentant une erreur sur un seul site et une classification quelconque.

L'estimateur MAP maximise donc la probabilité *a posteriori* $p(x|y)$. Il permet ainsi de retrouver la séquence x qui a le plus de chance d'avoir donné lieu à l'observation y .

1.1.3.2. Estimateur MPM

L'estimateur bayésien MPM est fondé sur une fonction de perte déterminée à partir de la somme de N fonctions de perte $L_{0/1}$ définies localement au niveau de chaque site $n \in \mathcal{N}$

$$\forall (x'_n, x_n) \in \mathcal{X} \times \mathcal{X}, L_{MPM}(x'_n, x_n) = L_{0/1}(x'_n, x_n) \quad (1.10)$$

L'estimateur bayésien MPM est donc donné par la formule

$$\begin{aligned}
(\hat{x}_n)_{MPM}(y) &= \arg \min_{x'_n \in \mathcal{X}} \sum_{x_n \in \mathcal{X}} \mathbf{1}_{[x'_n \neq x_n]} p(x_n|y) \\
&= \arg \min_{x'_n \in \mathcal{X}} \sum_{x_n \in \mathcal{X} / \{x'_n\}} p(x_n|y) & (1.11) \\
&= \arg \min_{x'_n \in \mathcal{X}} \{1 - p(x'_n|y)\} \\
&= \arg \max_{x'_n \in \mathcal{X}} p(x'_n|y)
\end{aligned}$$

L'estimateur MPM maximise donc la probabilité marginale *a posteriori* $p(x_n|y)$. Il permet ainsi de minimiser le nombre moyen de sites erronément classés.

La principale préoccupation pour le problème de segmentation considéré dans cette thèse étant de minimiser le taux d'erreur de segmentation, l'estimateur MPM sera donc retenu. Ainsi, la stratégie MPM sera développée par rapport à toutes les modélisations proposées dans ce manuscrit. Il convient de noter que la stratégie MAP peut être accomplie d'une manière analogue.

Le calcul des estimateurs MPM et MAP, entre autres estimateurs bayésiens, requiert l'évaluation des probabilités *a posteriori* $p(\cdot|y)$. Toutefois, l'espace de configurations possibles étant de très grande taille, il est impossible de calculer ces probabilités sans émettre des hypothèses simplificatrices. Un compromis s'impose souvent entre richesse de la loi $p(z)$ et simplicité des calculs liés à l'évaluation de ces probabilités.

1.2. Estimation des paramètres

Le cadre illustratif choisi dans cette thèse étant la segmentation à partir de seule l'observation y , il est indispensable d'estimer les paramètres θ régissant la loi $p(z)$ supposée jusqu'ici parfaitement connue. Dans cette section, nous présentons trois algorithmes populaires, souvent utilisés dans la littérature, notamment dans le cas des modèles de Markov sur lesquels porte la présente thèse : l'algorithme espérance-maximisation (EM) [Bau70, Dem77, McL97], sa version stochastique (EMS, en anglais SEM) [Bro83, Cel85, Cel96] et l'algorithme espérance conditionnelle itérative (ECI) [Pie92, Del97, Pie10a].

1.2.1. Algorithme Espérance-Maximisation (EM)

L'algorithme espérance-maximisation (en anglais *Expectation Maximization*) est parmi les techniques les plus utilisées pour l'estimation des paramètres des modèles à données latentes. Cet algorithme a été introduit par Baum *et al.* Dans le contexte du traitement de la parole [Bau70] puis repris par Dempster *et al.* [Dem77] pour l'estimation par maximum de vraisemblance à partir des données incomplètes.

Cet algorithme est fondé sur l'estimateur du maximum de vraisemblance (EMV) qui vise à maximiser la vraisemblance $p(y|\theta)$ que nous supposons dérivable par rapport au jeu de paramètre θ . Toutefois, le logarithme de la vraisemblance (Log-vraisemblance) noté $\mathcal{L}(y; \theta)$ est souvent considéré au lieu de la vraisemblance vu que cette dernière s'écrit souvent sous forme de produit de densité. L'estimateur est donc donné par la formule

$$\hat{\theta}(y) = \arg \max_{\theta} \mathcal{L}(y; \theta) \quad (1.12)$$

La résolution analytique de l'estimateur donné par (1.12) est souvent très difficile voire impossible. Ceci est dû au fait que la formule (1.12) fait souvent intervenir le logarithme d'une somme à cause de la présence des variables latentes dans la plupart des modèles concernés par notre étude. Par conséquent, le recours aux méthodes itératives telles que l'algorithme EM semble une solution intéressante.

L'algorithme EM est donc une technique itérative d'estimation de paramètre fondée sur l'estimateur EMV pour les modèles à variables latentes. Il part d'un jeu initial de paramètres θ^0 et l'itération $(q + 1)$ consiste à calculer le nouveau jeu de paramètres θ^{q+1} à partir de θ^q . Cet algorithme repose sur deux idées clés :

1. Il est généralement plus adéquat de calculer l'estimateur EMV sur la base des données complétées z que sur la base de l'observation seule y . Ainsi, la log-vraisemblance complétée est donnée par

$$\mathcal{L}_c(z; \theta) = \mathcal{L}(y; \theta) + \log p(x|y; \theta) \quad (1.13)$$

Etant à l'itération $(q + 1)$, l'accroissement de $\mathcal{L}(y; \theta)$ que l'on cherche à maximiser est donné par

$$\mathcal{L}(y; \theta) - \mathcal{L}(y; \theta^q) = \mathcal{L}_c(z; \theta) - \mathcal{L}_c(z; \theta^q) + \log \frac{p(x|y; \theta^q)}{p(x|y; \theta)} \quad (1.14)$$

2. La réalisation du processus caché x étant inconnue, la log-vraisemblance complétée est remplacée par son espérance conditionnellement aux observations. La formule (1.14) devient

$$\mathcal{L}(y; \theta) - \mathcal{L}(y; \theta^q) = \mathbb{E}_{\theta^q} \{ \mathcal{L}_c(Z; \theta) - \mathcal{L}_c(Z; \theta^q) | Y = y \} + \mathbb{E}_{\theta^q} \left\{ \log \frac{p(X|y; \theta^q)}{p(X|y; \theta)} \middle| Y = y \right\} \quad (1.15)$$

Le terme $\mathbb{E}_{\theta^q} \left\{ \log \frac{p(X|y; \theta^q)}{p(X|y; \theta)} \middle| Y = y \right\} \geq 0$ car il correspond à la divergence de Kullback-Leibler qui est toujours positive ou nulle. En posant $Q(\theta|\theta^q) = \mathbb{E}_{\theta^q} \{ \mathcal{L}_c(Z; \theta) | Y = y \}$, nous pouvons déduire

$$\mathcal{L}(y; \theta) \geq \mathcal{L}(y; \theta^q) + Q(\theta|\theta^q) - Q(\theta^q|\theta^q) = l(\theta|\theta^q) \quad (1.16)$$

L'inéquation (1.16) devient une égalité pour $\theta = \theta^q$. Sachant que $\mathcal{L}(y; \theta) \geq l(\theta|\theta^q)$, l'algorithme EM vise à rechercher le jeu de paramètres maximisant $l(\theta|\theta^q)$ afin de maximiser la borne inférieure de l'accroissement $\mathcal{L}(y; \theta) - \mathcal{L}(y; \theta^q)$. Le nouveau jeu de paramètres θ^{q+1} est donc obtenu par

$$\begin{aligned} \theta^{q+1}(y) &= \arg \max_{\theta} l(\theta|\theta^q) \\ &= \arg \max_{\theta} \{ Q(\theta|\theta^q) + \mathcal{L}(y; \theta^q) - Q(\theta^q|\theta^q) \} \\ &= \arg \max_{\theta} Q(\theta|\theta^q) \end{aligned} \quad (1.17)$$

Explicitement, l'algorithme EM se déroule comme suit :

0. Un jeu de paramètres θ^0 est choisi pour initialiser l'algorithme.
1. La procédure itérative suivante est répétée jusqu'à satisfaction d'un critère d'arrêt.
 - a. Etape E : Le terme $Q(\theta|\theta^q)$ est calculé ;
 - b. Etape M : Le nouveau jeu de paramètres est estimé selon (1.17).

Dans la pratique, le déroulement de l'algorithme est arrêté lorsqu'un nombre fixé au préalable d'itérations est atteint ou lorsque $\mathcal{L}(y; \theta^{q+1}) - \mathcal{L}(y; \theta^q) \leq \varepsilon$ pour une valeur ε fixée au préalable.

L'algorithme EM permet ainsi de construire, étant donné un jeu initial de paramètres θ^0 , une suite de jeux de paramètres $(\theta^q)_{q>0}$ assurant la croissance de $\{ \mathcal{L}(y; \theta^q) \}_{q>0}$ qui converge vers un maximum local de vraisemblance sous certaines conditions de régularité. Toutefois, la valeur de ce maximum local, qui peut être différent du maximum global, est fortement liée à l'initialisation. Dans ce cadre, des stratégies d'initialisation ont été proposées. L'une de ces stratégies repose sur l'exécution de l'algorithme EM pendant un petit nombre d'itérations en démarrant de plusieurs jeux de paramètres initiaux et de choisir celui dont la log-vraisemblance est maximale. Une autre stratégie consiste à produire une pré-classification selon une méthode aveugle et calculer le jeu de paramètres initial par l'estimateur de la vraisemblance complétée.

En sus des problèmes de convergence, d'autres limites de l'algorithme EM concernent l'éventuelle difficulté voire l'impossibilité d'évaluer le terme $Q(\theta|\theta^q)$ analytiquement. Pour cela, d'autres approches d'estimation de paramètres ont été proposées telles que SEM et ECI.

1.2.2. Algorithme EM stochastique (SEM)

L'algorithme EM stochastique (en anglais *Stochastic Expectation Maximization*) fut introduit par Broniatowski *et al.* [Bor83]. Il constitue une version stochastique de l'algorithme de référence EM comprenant une étape stochastique (étape S) précédant l'étape d'estimation (étape E). Cette étape consiste à échantillonner une réalisation du processus caché selon les lois *a posteriori*. A la base, l'algorithme SEM fut proposé pour pallier aux problèmes de convergence de l'algorithme EM, toutefois, il permet également de pallier aux éventuelles difficultés de l'évaluation du terme $Q(\theta|\theta^q)$ rencontré dans l'étape E de l'algorithme EM.

L'étape S de l'algorithme consiste à générer des échantillons pseudo-complets par tirages aléatoires des échantillons inobservables en se basant sur leur densité conditionnelle aux observations selon le principe SIP (*Stochastic Imputation Principle* [Cel88]). Par conséquent, contrairement à l'algorithme EM, il est possible de dériver à l'itération $(q + 1)$ un estimé θ^{q+1} pour lequel la log-vraisemblance $\mathcal{L}(y; \theta^{q+1})$ décroît par rapport à $\mathcal{L}(y; \theta^q)$, ce qui a l'avantage de contourner le problème de rester bloqué dans un maximum local non significatif. L'avantage principal donc de l'algorithme SEM par rapport à EM est d'être moins sensible à l'initialisation. Toutefois, l'algorithme EM converge plus rapidement que son homologue stochastique.

1.2.3. Estimation conditionnelle itérative (ECI)

L'estimation conditionnelle itérative (en anglais *Iterative Conditional Estimation* ou ICE) est une technique générale d'estimation des paramètres des modèles à données latentes introduite par Pieczynski [Pie92, Pie94, Ben95, Pen95]. Son principe est de considérer un estimateur $\hat{\theta} = \hat{\theta}(X, Y)$ défini sur la base des données complètes et de l'approcher par une fonction de la seule donnée observable Y .

Dans ce cadre, la meilleure approximation, au sens de l'erreur quadratique moyenne, est l'espérance conditionnelle. L'espérance conditionnelle de $\hat{\theta} = \hat{\theta}(X, Y)$ par rapport à Y notée $\mathbb{E}(\hat{\theta}(X, Y)|Y)$, dépend du jeu de paramètres θ . Pour l'évaluer, on a recours à la procédure itérative qui consiste à dériver θ^{q+1} à partir de θ^q selon

$$\theta^{q+1} = \mathbb{E}_{\theta^q}(\hat{\theta}(X, Y)|Y = y) \quad (1.18)$$

Lorsque $\mathbb{E}_{\theta^q}(\hat{\theta}(X, Y)|Y = y)$ n'est pas explicitement calculable mais la simulation des réalisations du processus caché X selon la loi conditionnelle à Y est possible, nous pouvons l'approcher de façon stochastique. En effet, l'espérance conditionnelle est aussi l'espérance selon la loi conditionnelle et peut donc être approchée, en vertu de la loi des grands nombres, par la moyenne empirique. Nous pouvons donc échantillonner Q réalisations x^1, \dots, x^I de X selon la loi conditionnelle à l'observation $Y = y$ donnée par θ^q et poser

$$\theta^{i+1} = \frac{1}{I} \sum_{i=1}^I \hat{\theta}(x^i, y) \quad (1.19)$$

L'algorithme ECI se déroule donc comme suit :

1. Un jeu de paramètres θ^0 est choisi pour initialiser l'algorithme.
2. La procédure itérative suivante est répétée jusqu'à satisfaction d'un critère d'arrêt.

- Poser (1.18) pour toutes les composantes $\hat{\theta}^i$ pour lesquelles $\mathbb{E}_{\hat{\theta}^i}(\hat{\theta}(X, Y) | Y = y)$ est calculable;
- Poser (1.19) pour toutes les composantes $\hat{\theta}^i$ pour lesquelles $\mathbb{E}_{\hat{\theta}^i}(\hat{\theta}(X, Y) | Y = y)$ n'est pas calculable.

1.3. Conclusion

Dans ce chapitre, nous avons présenté le problème de segmentation selon le paradigme des approches bayésiennes. Nous avons également défini la notion de fonction de perte qui permet à la fois d'attribuer une certaine souplesse aux approches bayésiennes dans la mesure où elle fait dépendre l'optimalité des segmentations des préoccupations particulières liées à l'application visée ; et de mesurer la qualité d'une classification. Ainsi, nous avons décrit les deux estimateurs bayésiens les plus employés dans la littérature, en l'occurrence MAP et MPM. Ce dernier a été retenu comme technique de segmentation selon les différentes approches proposées dans la suite de cette étude. Enfin, le problème de segmentation, considéré dans cette thèse, étant non-supervisé, nous avons décrit l'algorithme d'estimation des paramètres EM ainsi que ses deux concurrents EMS et ECI. L'algorithme EM sera choisi comme algorithme de référence pour les différents modèles proposés dans ce manuscrit étant donnée sa rapidité. Il a été établi dans des travaux antérieurs [Pen95] que les trois approches EM, SEM et ECI se valent pour certains types de bruit, notamment le bruit gaussien qui sera adopté lors de cette thèse.

Chapitre 2. Chaînes de Markov cachées, couples et triplets

Comme nous l'avons vu dans le chapitre précédent, la segmentation bayésienne est fondée sur une modélisation stochastique des phénomènes caché X et observé Y . Plus explicitement, le lien entre ces deux phénomènes est modélisé à travers la loi jointe du processus couple $Z = (X, Y)$. Cette loi doit être aussi riche que possible afin de refléter la réalité que l'on souhaite modéliser. En même temps, elle doit rester assez simple pour que les calculs associés aux différentes procédures d'estimation bayésiennes demeurent faisables. Un compromis s'impose ainsi entre richesse et simplicité lors de la définition de cette loi jointe. Vu la taille des configurations possibles des données, il est très coûteux voire impossible de calculer les probabilités *a posteriori* requises pour les différentes procédures d'estimation. Des hypothèses simplificatrices sont ainsi émises de façon à alléger les calculs tout en maintenant la loi jointe assez représentative. L'une de ces hypothèses porte sur l'indépendance des variables aléatoires Z_n associés aux différents sites $n \in \mathcal{N}$. La segmentation bayésienne est effectuée dans ce cas, site par site tout en ignorant l'information spatiale, ce qui attribue à ce genre d'approches le qualificatif « aveugle ». Les modèles de chaînes de Markov cachées (à bruit indépendant) réussissent un meilleur compromis en supposant les variables Y_n indépendantes conditionnellement à X , ce qui leur permet d'effectuer, en un temps raisonnable, les différentes opérations d'estimation à partir de toute l'observation $Y = y$. De leur côté, les CMC ont été étendues aux chaînes de Markov couples (CMCo) et triplet (CMT), qui sont de généralité strictement croissante. Dans ce chapitre, nous allons présenter ces modèles tout en détaillant leurs différentes techniques d'estimation associées. A la fin de ce chapitre, un modèle particulier du type CMT appelé chaîne de Markov M-stationnaire est présenté pour illustrer l'intérêt des CMT par rapport aux CMC classiques pour le problème de modélisation des données non-stationnaires qui constitue un volet important dans cette thèse.

2.1. Chaînes de Markov cachées

Les chaînes de Markov cachées figurent parmi les modèles stochastiques les plus utilisés dans la littérature pour la segmentation des données aléatoires. Cette notoriété est due à leur aptitude d'effectuer les différentes tâches d'estimation de manière holistique en utilisant toute la donnée grâce à des techniques bayésiennes telle que MPM et MAP. En plus, lorsque les paramètres du modèle ne sont pas connus, il est possible de les estimer grâce à des algorithmes itératifs à l'instar de l'algorithme EM [Bau70, Dem77]. Ainsi, les chaînes de Markov cachées ont été utilisées dans plusieurs domaines d'application, notamment en reconnaissance des formes et traitement du signal et de l'image : reconnaissance de la parole [Rab89, Dai94], segmentation d'image [Qia89, Fjo03, Hat07, Bou10], analyse d'images [Aas99], reconnaissance des mots manuscrits [Che94, Jac99], classification des tumeurs [Sol06] et détection de changement [Car06]. D'autres domaines d'application de ces modèles incluent la biologie [Chu92, Lia07], la biométrie [Sha03, Ben13a, Ben13b], l'écologie [Ber06], les communications [Kal94] et la finance [Tho02, Bha04]. Citons également les articles pionniers [Bau70, For73, Rab89, Eph02, Pie03b] et

les livres [Kos01, Cap05] où la théorie relative à ces modèles a été décrite exhaustivement. Dans cette section, nous décrivons ces modèles ainsi que leurs différentes techniques d'estimation associées.

2.1.1. Définition du modèle

Soit $X = (X_n)_{n=1}^N$ un processus caché qui prend ses valeurs à partir d'un ensemble fini de classes $\Omega = \{\omega_1, \dots, \omega_K\}$ et qui doit être estimé à partir d'un processus observable $Y = (Y_n)_{n=1}^N$ où $Y_n \in \mathbb{R}$. Le processus couple $Z = (X, Y)$ est dit chaîne de Markov cachée si X est une chaîne de Markov et si Z est de Markov.

La loi de X est donc donnée par

$$p(x) = p(x_1) \prod_{n=2}^N p(x_n | x_{n-1}) \quad (2.1)$$

Le processus observable Y étant perçu comme une version bruitée du processus caché X , la loi $p(y|x)$ est souvent appelée distribution de bruit.

Si dans une chaîne de Markov cachée, les observations sont indépendantes conditionnellement à $X = x$, c'est-à-dire

$$p(y|x) = \prod_{n=2}^N p(y_n | x_n) \quad (2.2)$$

La loi de Z sera donc

$$p(x, y) = p(x_1) p(y_1 | x_1) \prod_{n=2}^N p(x_n | x_{n-1}) p(y_n | x_n) \quad (2.3)$$

Z est appelé dans ce cas « chaîne de Markov cachée à bruit indépendant ». Tout le long de cette thèse, nous utilisons l'appellation « chaîne de Markov cachée » abrégée en CMC pour signifier chaîne de Markov cachée à bruit indépendant.

2.1.2. Simulation d'une chaîne de Markov cachée gaussienne

Avant de détailler la procédure de simulation d'une CMC, il convient de spécifier ses paramètres θ . Soit $Z = (X, Y)$ une CMC dont la loi est donnée par (2.3). Z est dite gaussienne si les distributions de bruit $p(y_n | x_n)$ sont gaussiennes. Une CMC gaussienne est donc entièrement définie dès lors que les paramètres suivants sont définis :

- La loi $p(x_1)$ donnée par le vecteur Π où $\pi_i = p(x_1 = \omega_i)$;
- Les K^2 probabilités de transition $p(x_n | x_{n-1})$ données par la matrice A où $a_{ij} = p(x_n = \omega_j | x_{n-1} = \omega_i)$;
- Les K moyennes μ_i et les K écarts-types σ_i définissant les densités gaussiennes de bruit $p(y_n | x_n = \omega_i)$.

La procédure de simulation de la CMC gaussienne Z_θ définie plus haut se fait de la manière suivante :

1. Simuler la réalisation x_1 par tirage aléatoire à partir du vecteur Π ;

2. Simuler les prochaines réalisations du processus caché X par tirages aléatoires à partir de la matrice de transition A ;
3. Simuler les réalisations du processus observable Y par tirages aléatoires à partir des densités gaussiennes associées aux réalisations du processus caché X .

2.1.3. Inférence dans les chaînes de Markov cachées

A présent, nous décrivons la technique de restauration MPM conformément au modèle de CMC. Soit $Z = (X, Y)$ une CMC donnée par (2.3). L'estimateur MPM est donné par

$$[\hat{x} = \hat{s}_{MPM}(y)] \Leftrightarrow \left[\hat{x}_n = \underset{\omega \in \Omega}{\operatorname{argmax}} p(x_n = \omega | y) \forall n \in \{1, \dots, N\} \right] \quad (2.4)$$

Pour le calcul des probabilités *a posteriori* d'intérêt $p(x_n | y)$, nous définissons les fonctions de probabilité progressive $\alpha_n(x_n) = p(y_1, \dots, y_n, x_n)$ et rétrograde $\beta_n(x_n) = p(y_{n+1}, \dots, y_N | x_n)$ [Bau70] et dont le calcul se fait de la manière itérative suivante :

Initialisation :

$$\alpha_1(x_1) = p(x_1)p(y_1|x_1) \quad (2.5)$$

$$\beta_N(x_N) = 1 \quad (2.6)$$

Induction :

$$\alpha_n(x_n) = \sum_{x_{n-1}} \alpha_{n-1}(x_{n-1})p(x_n|x_{n-1})p(y_n|x_n) \quad (2.7)$$

$$\beta_n(x_n) = \sum_{x_{n+1}} \beta_{n+1}(x_{n+1})p(x_{n+1}|x_n)p(y_{n+1}|x_{n+1}) \quad (2.8)$$

Les distributions *a posteriori* $p(x_n | y)$ peuvent être calculées selon les formules

$$p(x_n | y) \propto \alpha_n(x_n) \beta_n(x_n) \quad (2.9)$$

Enfin, nous définissons les probabilités *a posteriori* suivantes que nous allons utiliser au sein de l'algorithme d'estimation des paramètres.

$$\psi_n(x_n, x_{n+1}) = p(x_n, x_{n+1} | y) \propto \alpha_n(x_n)p(x_{n+1}|x_n)p(y_{n+1}|x_{n+1}) \beta_{n+1}(x_{n+1}) \quad (2.10)$$

$$\xi_n(x_n) = p(x_n | y) = \sum_{x_{n+1}} \psi_n(x_n, x_{n+1}) \quad (2.11)$$

2.1.1. Estimation des paramètres d'une chaîne de Markov cachée

Dans le cas où les paramètres θ du modèle ne pas sont connus, il est possible de les estimer en utilisant l'algorithme EM (espérance-maximisation) [Bau70, Dem77], sa version stochastique SEM [Bor83, Cel85] ou l'algorithme ECI (estimation conditionnelle itérative) [Pie92, Pie94, Ben95, Pen95]. L'algorithme itératif EM étant choisi tout le long de cette thèse, nous montrons à présent son déroulement

- Considérer un jeu initial de paramètres $\theta^0 = (\pi^0, A^0, \mu^0, \sigma^0)$;
- Pour chaque itération $q+1$:

- Estimation : Calculer les fonctions ψ_n^q and ξ_n^q à partir de y et θ^q .
- Maximisation : Estimer le nouveau jeu de paramètres θ^{q+1} , qui est calculable analytiquement, comme suit :

$$\pi_k^{q+1} = \xi_1^q(x_n = \omega_k) \quad (2.12)$$

$$a_{ij}^{q+1} = \frac{\sum_{n=1}^{N-1} \psi_n^q(x_n = \omega_i, x_{n+1} = \omega_j)}{\sum_{n=1}^{N-1} \xi_n^q(x_n = \omega_i)} \quad (2.13)$$

$$\mu_k^{q+1} = \frac{\sum_{n=1}^N \xi_n^q(x_n = \omega_k) y_n}{\sum_{n=1}^N \xi_n^q(x_n = \omega_k)} \quad (2.14)$$

$$\sigma_k^{q+1} = \frac{\sum_{n=1}^N \xi_n^q(x_n = \omega_k) (y_n - \mu_k^{q+1})^2}{\sum_{n=1}^N \xi_n^q(x_n = \omega_k)} \quad (2.15)$$

Selon l'application, il est possible de remplacer (3.12) la formule par

$$\pi_k^{q+1} = \frac{1}{N} \sum_{n=1}^N \xi_n^q(x_n = \omega_k) \quad (2.16)$$

2.2. Chaînes de Markov couples

Dans cette section, nous définissons le modèle de chaîne de Markov couple introduit dans [Pie03a, Der04] et qui généralise le modèle de CMC classique.

2.2.1. Définition du modèle

Soit $X = (X_n)_{n=1}^N$ un processus caché qui prend ses valeurs à partir d'un ensemble fini de classes $\Omega = \{\omega_1, \dots, \omega_K\}$ et qui doit être estimé à partir d'un processus observable $Y = (Y_n)_{n=1}^N$ où $Y_n \in \mathbb{R}$. Le processus couple $Z = (X, Y)$ est dit chaîne de Markov couple (CMCo) s'il est de Markov.

La loi de Z est donc donnée par

$$p(z) = p(z_1) \prod_{n=2}^N p(z_n | z_{n-1}) \quad (2.17)$$

Sachant que les transitions $p(z_n | z_{n-1})$ peuvent s'écrire

$$p(z_n | z_{n-1}) = p(x_n | x_{n-1}, y_{n-1}) p(y_n | x_n, x_{n-1}, y_{n-1}), \quad (2.18)$$

il est facile de remarquer que le modèle de CMC défini plus haut et dont la loi est donnée par (2.3) est une CMCo particulière vérifiant : $p(x_n | x_{n-1}, y_{n-1}) = p(x_n | x_{n-1})$ et $p(y_n | x_n, x_{n-1}, y_{n-1}) = p(y_n | x_n)$. Ceci illustre la plus grande généralité des modèles CMCo par rapport aux CMC.

Dans cette thèse, nous nous limitons à l'étude des CMCo stationnaires. Une CMCo est dite stationnaire si les distributions $p(z_n, z_{n+1})$ ne dépendent pas de n . Par ailleurs, une CMCo est dite réversible si $p(z_n, z_{n+1}) = p(z_{n+1}, z_n)$. Dans ce cadre, il convient de citer le résultat suivant [Pie10b] concernant la différence entre une CMC et une CMCo.

Soit Z une CMCo stationnaire et réversible, alors les trois conditions suivantes sont équivalentes

- 1) X est une chaîne de Markov ;
- 2) pour tout $2 \leq n \leq N$, $p(y_n|x_n, x_{n-1}) = p(y_n|x_n)$;
- 3) pour tout $1 \leq n \leq N$, $p(y_n|x) = p(y_n|x_n)$.

Ce résultat illustre la plus grande généralité des CMCo par rapport aux CMC quant à la richesse de la modélisation du bruit. En effet, supposer la markovianité du processus caché revient à restreindre la dépendance de la loi $p(y_n|x)$ à l'état x_n seul, alors que cette même loi dépend de toute la séquence x dans une CMCo.

Il convient de remarquer que dans une CMCo, les distributions conditionnelles $p(y|x)$ et $p(x|y)$ sont markoviennes sans qu'aucune des distributions marginales $p(x)$ et $p(y)$ ne le soit [Pie03a].

A présent, nous présentons la preuve [Pie03a] de cette proposition. Nous avons

$$p(y|x) = \frac{p(z)}{p(x)} \quad (2.19)$$

En exploitant (2.17), la formule (2.19) devient

$$\begin{aligned} p(y|x) &= \frac{p(z_1)p(z_2|z_1)\dots p(z_N|z_{N-1})}{p(x)} \\ &= \frac{p(x_1, y_1)p(x_2, y_2|x_1, y_1)\dots p(x_N, y_N|x_{N-1}, y_{N-1})}{p(x)} \end{aligned} \quad (2.20)$$

Comme x est constante dans (2.20), nous vérifions que $p(y_1, \dots, y_N|x)$ est de forme markovienne. Grâce à la symétrie de X et Y , nous pouvons démontrer de la même manière que $p(x|y)$ est de Markov.

Soit Z une CMCo stationnaire, la loi d'une CMCo stationnaire Z est donc entièrement définie par la distribution

$$p(z_1, z_2) = p(i, j) f_{i,j}(y_1, y_2) \quad (2.21)$$

Où $p(i, j)$ sont des probabilités définies sur Ω^2 et $f_{i,j}(y_1, y_2) = p(y_1, y_2|x_1 = \omega_i, x_2 = \omega_j)$ sont des densités de probabilité définies sur \mathbb{R}^2 .

La loi de peut être déterminée également de la manière traditionnelle à travers les probabilités initiales et les probabilités de transition données par (2.20) et (2.21) respectivement

$$p(z_1) = \sum_{\omega_k \in \Omega} p(i, k) f_{i,k}(y_1) \quad (2.22)$$

où $f_{i,k}(y_1) = \int f_{i,k}(y_1, y_2) dy_2$.

$$p(z_n|z_{n-1}) = \frac{p(z_{n-1}, z_n)}{p(z_{n-1})} = \frac{p(i, j) f_{i,j}(y_{n-1}, y_n)}{\sum_{\omega_k \in \Omega} p(i, k) f_{i,k}(y_{n-1})} \quad (2.23)$$

Il convient de noter que (2.23) n'est pas de forme gaussienne.

2.2.2. Simulation d'une chaîne de Markov couple gaussienne

Une CMCo est dite gaussienne si les distributions de bruit $f_{i,j}(y_{n-1}, y_n)$ sont des densités gaussiennes. Avant de détailler la procédure de simulation d'une CMCo gaussienne, nous spécifions ses paramètres θ . Un tel modèle est entièrement défini à travers les K^2 densités bidimensionnelles $p(i, j)$ données par la matrice P et les K^2 densités gaussiennes de bruit définies par les vecteurs moyennes $\mu_1^{i,j}, \mu_2^{i,j}$, écarts-type $\sigma_1^{i,j}, \sigma_2^{i,j}$ et les coefficients de corrélation $\rho^{i,j}$.

Soit $Z = (X, Y)$ une CMCo de paramètres θ . La procédure de simulation se déroule de la manière suivante

- La réalisation z_1 est simulée par tirages à partir des distributions

$$p(x_1 = \omega_k) = \sum_{\omega_i \in \Omega} p(k, i) \quad (2.24)$$

$$p(y_1 | x_1 = \omega_k) = \sum_{\omega_i \in \Omega} p(k, i) f_{k,i}(y_1) \quad (2.25)$$

Notons que $f_{k,i}(y_1)$ est une densité gaussienne de moyenne $\mu_1^{k,i}$ et d'écart-type $\sigma_1^{k,i}$. La réalisation y_1 est donc obtenue par tirage à partir d'un mélange gaussien.

- Simuler les prochaines réalisations du processus Z par tirages aléatoires à partir des distributions

$$p(x_n = \omega_j | x_{n-1} = \omega_i, y_{n-1}) = \frac{p(i, j) f_{i,j}(y_{n-1})}{\sum_{\omega_k \in \Omega} p(i, k) f_{i,k}(y_{n-1})} \quad (2.26)$$

$$p(y_n | x_n = \omega_j, x_{n-1} = \omega_i, y_{n-1}) = \frac{f_{i,j}(y_{n-1}, y_n)}{f_{i,j}(y_{n-1})} \quad (2.27)$$

Il est possible de démontrer que $p(y_n | x_n, z_{n-1})$ est une densité gaussienne de moyenne $\mu_2^{i,j} + \rho^{i,j}(\sigma_2^{i,j}/\sigma_1^{i,j})(y_{n-1} - \mu_1^{i,j})$ et d'écart-type $\sigma_2^{i,j} \sqrt{1 - (\rho^{i,j})^2}$ [Der04]. La simulation de y_n est donc obtenue par tirage aléatoire à partir d'un mélange de gaussiennes.

2.2.3. Inférence dans les chaînes de Markov couples

A présent, nous présentons la technique de restauration MPM correspondante au modèle CMCo. Soit $Z = (X, Y)$ une CMCo donnée par (2.19).

Pour le calcul de l'estimateur MPM donné par (2.4), nous définissons les fonctions de probabilité progressive $\alpha_n(x_n) = p(y_1, \dots, y_n, x_n)$ et rétrograde généralisée $\beta_n(x_n) = p(y_{n+1}, \dots, y_N | x_n, y_n)$ et dont le calcul se fait de la manière itérative suivante

Initialisation :

$$\alpha_1(x_1) = p(z_1) \quad (2.28)$$

$$\beta_N(x_N) = 1 \quad (2.29)$$

Induction :

$$\alpha_n(x_n) = \sum_{x_{n-1}} \alpha_{n-1}(x_{n-1}) p(z_n | z_{n-1}) \quad (2.30)$$

$$\beta_n(x_n) = \sum_{x_{n+1}} \beta_{n+1}(x_{n+1}) p(z_{n+1} | z_n) \quad (2.31)$$

Les distributions *a posteriori* $p(x_n|y)$ requises pour la restauration MPM peuvent être calculées selon la formule

$$p(x_n|y) \propto \alpha_n(x_n) \beta_n(x_n) \quad (2.32)$$

Enfin, les fonctions de probabilité *a posteriori* $\psi_n(x_n, x_{n+1}) = p(x_n, x_{n+1}|y)$ et $\xi_n(x_n) = p(x_n|y)$ nécessaires pour l'algorithme d'estimation des paramètres se calculent selon les formules suivantes

$$\psi_n(x_n, x_{n+1}) \propto \alpha_n(x_n) p(z_{n+1}|z_n) \beta_{n+1}(x_{n+1}) \quad (2.33)$$

$$\xi_n(x_n) = \sum_{x_{n+1}} \psi_n(x_n, x_{n+1}) \quad (2.34)$$

2.2.4. Estimation des paramètres d'une chaîne de Markov couple

Lorsque les paramètres θ de la CMCo Z ne sont pas connus, nous proposons de les estimer en utilisant l'algorithme EM qui se déroule de la manière itérative suivante

- Considérer un jeu initial de paramètres $\theta^0 = (P^0, \mu^0, \sigma^0, \rho^0)$;
- Pour chaque itération $q+1$:
 - o Estimation : Calculer les fonctions ψ_n^q and ξ_n^q à partir de y et θ^q .
 - o Maximisation : Estimer le nouveau jeu de paramètres θ^{q+1} comme suit :

$$p_{ij}^{q+1} = \frac{1}{N-1} \sum_{n=1}^{N-1} \psi_n^q(x_n = \omega_i, x_{n+1} = \omega_j) \quad (2.35)$$

$$(\mu_{ij}^1)^{q+1} = \frac{\sum_{n=1}^N \psi_n^q(x_n = \omega_i, x_{n+1} = \omega_j) y_n}{\sum_{n=1}^N \psi_n^q(x_n = \omega_i, x_{n+1} = \omega_j)} \quad (2.36)$$

$$(\mu_{ij}^2)^{q+1} = \frac{\sum_{n=1}^N \psi_n^q(x_n = \omega_i, x_{n+1} = \omega_j) y_{n+1}}{\sum_{n=1}^N \psi_n^q(x_n = \omega_i, x_{n+1} = \omega_j)} \quad (2.37)$$

$$(\sigma_{ij}^1)^{q+1} = \frac{\sum_{n=1}^N \psi_n^q(x_n = \omega_i, x_{n+1} = \omega_j) (y_n - (\mu_{ij}^1)^{q+1})^2}{\sum_{n=1}^N \psi_n^q(x_n = \omega_i, x_{n+1} = \omega_j)} \quad (2.38)$$

$$(\sigma_{ij}^2)^{q+1} = \frac{\sum_{n=1}^N \psi_n^q(x_n = \omega_i, x_{n+1} = \omega_j) (y_{n+1} - (\mu_{ij}^2)^{q+1})^2}{\sum_{n=1}^N \psi_n^q(x_n = \omega_i, x_{n+1} = \omega_j)} \quad (2.39)$$

$$\rho_{ij}^{q+1} = \frac{\sum_{n=1}^N \psi_n^q(x_n = \omega_i, x_{n+1} = \omega_j) (y_n - (\mu_{ij}^1)^{q+1}) (y_{n+1} - (\mu_{ij}^2)^{q+1})}{(\sigma_{ij}^1)^{q+1} (\sigma_{ij}^2)^{q+1} \sum_{n=1}^N \psi_n^q(x_n = \omega_i, x_{n+1} = \omega_j)} \quad (2.40)$$

Le lecteur peut se référer à [Der11] où des codes sources ainsi que des démos relatives aux CMCo sont disponibles.

2.3. Chaînes de Markov triplets

Dans cette section, nous présentons une famille de modèles appelés « chaînes de Markov triplets » [Pie02a, Pie02b] qui généralise les CMC et les CMCo décrites plus haut. Dans un premier temps, nous définissons le modèle général. Puis, nous passons en revue certains travaux ayant porté sur les modèles de Markov triplets. Enfin, et afin de comprendre le déroulement des différentes procédures d'estimation, nous présentons le modèle de CMC M-stationnaire introduit dans [Lan11] pour la modélisation des données non-stationnaires.

2.3.1. Définition du modèle général

Soit $X = (X_n)_{n=1}^N$ un processus caché qui prend ses valeurs à partir d'un ensemble fini de classes $\Omega = \{\omega_1, \dots, \omega_K\}$ et qui doit être estimé à partir d'un processus observable $Y = (Y_n)_{n=1}^N$ où $Y_n \in \mathbb{R}$. Soit $U = (U_n)_{n=1}^N$ un autre processus sous-jacent, appelé processus auxiliaire, qui prend ses valeurs à partir d'un ensemble fini de classes auxiliaires $\Lambda = \{\lambda_1, \dots, \lambda_M\}$. Le processus triplet $T = (U, X, Y)$ est dit chaîne de Markov triplet (CMT) s'il est de Markov. La loi $p(t)$ est donc donnée par

$$p(t) = p(t_1) \prod_{n=2}^N p(t_n | t_{n-1}) \quad (2.41)$$

Notons que les modèles d'arbres et champs de Markov cachés ont également été étendus aux arbres et champs de Markov triplets respectivement. Les techniques d'estimation en vigueur pour les modèles de Markov cachés classiques ont aussi été adaptées aux modèles de Markov triplets. En effet, il convient de remarquer que la loi donnée par (2.41) est la même que celle d'une CMCo donnée par (2.17). De ce fait, il suffit de poser $T = (U, X, Y) = (V, Y)$ et le processus T peut être considéré comme une CMCo associée aux processus V et Y , et par conséquent, les différentes techniques d'estimations utilisées dans le contexte des CMCo restent valables pour les CMT. Dans ce même cadre, mentionnons le théorème intéressant, énoncé dans [Pie07], qui donne une définition générale d'une chaîne de Markov et qui peut être exploité pour les différentes procédures d'inférence des chaînes de Markov en général, en particulier celles définies par (2.41).

2.3.2. Travaux antérieurs

L'intérêt majeur des CMT, et des modèles de Markov triplets en général, réside dans l'utilité du processus auxiliaire U pour la prise en compte de certaines situations difficiles que les modèles classiques ne parviennent pas à modéliser. A présent, nous mentionnons certains travaux antérieurs qui ont porté sur les modèles de Markov triplets. Dans [Lan06, Lan11], Lanchantin introduit le modèle de CMC à sauts, également appelé CMC M-stationnaire ou M-non stationnaire. Pour de tels modèles, la séquence de donnée est partitionnée en M portions, chacune étant modélisable via une matrice de transition différente. Le rôle du processus auxiliaire dans ce cas est de gérer les sauts entre ces matrices permettant ainsi de tenir compte de l'hétérogénéité de la loi *a priori* $p(x)$. L'une des applications potentielles de ce modèle est la segmentation d'image où des champs de Markov cachés à sauts définis de façon similaire ont été appliqués avec succès [Ben05, Ben07].

Un lien intéressant a également été établi entre les modèles de Markov triplets et la théorie de l'évidence [Dem67, Sha76, Yag94, Blo96, Cob03, Den97, Fou00, Jan98, Sme93, Sud03, Ram09, Sou10]. En effet, l'emploi de la fusion de Dempster-Shafer n'étant possible directement au sein des CMC vu que cette fusion détruit la markovianité. Toutefois, il a été démontré que le résultat d'une telle fusion est une distribution marginale d'un modèle de Markov triplet et par conséquent, les différentes techniques d'estimation restent faisables [Pie06, Pie07]. Ainsi, dans [Lan05], les auteurs proposent un modèle appelé « chaîne de Markov évidentielle cachée » pour la modélisation de données non-stationnaires. Dans ce contexte, la loi *a priori* non fiable a été remplacée par une fonction de croyance pour modéliser son incertitude. Dans le même cadre, la segmentation d'images non-stationnaires a été effectuée dans le contexte des champs de Markov [Pie06].

Blanchet [Bla07, Bla08] a appliqué les champs de Markov triplets au problème de classification pour contourner trois sources de complexité : le problème de la dimension des données observées, la dépendance entre ces observations et la nature générale du modèle de bruit. Dans [Lap12], les auteurs proposent une autre extension des CMC appelée « chaînes partiellement de Markov triplets » qui peut être utilisée pour la modélisation de la semi-markovianité et la non-stationnarité.

D'autres applications potentielles des modèles de Markov triplets incluent la segmentation des images IRM 3D [Bri06], le filtrage [Ait06, Pie11], la détection de changement en imagerie SAR [Wan13], la segmentation multi-classes en imagerie SAR [Gan12, Zha12a, Zha12b].

Il convient de noter que certains modèles existants ont été redéfinis selon le formalisme des modèles de Markov triplets. Citons par exemple les modèles semi-markoviens cachés. En effet, le processus auxiliaire U peut être utilisé pour définir la durée de séjour dans un état [Lap12]. Les modèles de mélanges gaussiens (en anglais *Gaussian mixture models* ou GMM [Ngu12]) peuvent aussi être vus comme des modèles de Markov triplet particuliers.

Le processus auxiliaire peut être un processus couple $U = (U^1, U^2)$. Citons le modèle de chaîne semi-markovienne non-stationnaire cachée où U^1 est utilisé pour la semi-markovianité alors que U^2 est employé pour modéliser la non-stationnarité du modèle [Lap12]. Enfin, mentionnons qu'un processus auxiliaire à états continus a été considéré pour modéliser les données non-stationnaires pour décrire la situation où la loi *a priori* varie graduellement [BeM07, BeM11].

2.3.3. Chaînes de Markov cachées M-stationnaires

Dans ce qui suit, nous présentons un modèle de CMT particulier appelé « CMC M-stationnaire » conçu pour la modélisation de données non-stationnaires. Notons que d'autres approches stochastiques ont abordé la non-stationnarité des données à travers les fonctions de durée [Sin95]. Les paramètres peuvent alors être estimés via des méthodes de type Monte Carlo (MCMC) [Dju02]. Des extensions ultérieures aux chaînes semi-markoviennes cachées ont également été proposées [You03].

2.3.3.1. Définition du modèle

Soit $X = (X_n)_{n=1}^N$ un processus inobservable à valeurs dans $\Omega = \{\omega_1, \dots, \omega_K\}$ et que l'on doit estimer à partir d'un processus observable $Y = (Y_n)_{n=1}^N$ à valeurs dans \mathbb{R} . $Z = (X, Y)$ sera dite une CMC M-stationnaire (à bruit indépendant) s'il existe un troisième processus $U = (U_n)_{n=1}^N$ à valeurs dans $\Lambda = \{\lambda_1, \dots, \lambda_M\}$ tel que le processus triplet $T = (U, X, Y)$ est une chaîne de Markov stationnaire.

Dans la suite, nous considérons un modèle de CMC M-stationnaire particulier où (i) le processus U est markovien, (ii) les transitions $p(x_n|x_{n-1})$ ne dépendent que de la réalisation courante de U_n et (iii) la distribution du bruit est donnée par $p(y|u, x) = p(y|x) = \prod_{n=1}^N p(y_n|x_n)$. Un tel modèle sera noté CMC M-S.

La loi $p(t)$ peut alors s'exprimer à travers les lois de probabilités initiales

$$p(t_1) = p(u_1)p(x_1|u_1)p(y_1|x_1) \quad (2.42)$$

et les probabilités de transition

$$p(t_n|t_{n-1}) = p(u_n|u_{n-1})p(x_n|x_{n-1}, u_n)p(y_n|x_n) \quad (2.43)$$

2.3.3.2. Simulation d'une chaîne de Markov cachée M-stationnaire

Une CMC M-S est dite gaussienne si les lois de bruit $p(y_n|x_n)$ sont des densités gaussiennes.

Spécifions à présent les paramètres θ d'un tel modèle. Une CMC M-S est entièrement défini à travers les M probabilités initiales données par le vecteur $\varpi = [\varpi_m]_{\lambda_m \in \Lambda}$ où $\varpi_m = p(u_1 = \lambda_m)$, les M^2 probabilités données par la matrice de transition $A = [a_{l,m}]_{\lambda_l, \lambda_m \in \Lambda}$ où $a_{l,m} = p(u_n = \lambda_m | u_{n-1} = \lambda_l)$, les MK probabilités initiales données par les vecteurs $(\pi_m)_{m=1}^M$ où $\pi_{m,k} = p(x_1 = \omega_k | u_1 = \lambda_m)$, les MK^2 probabilités données par les matrices de transition $(P_m)_{m=1}^M$ où $p_{m,i,j} = p(x_n = \omega_j | x_{n-1} = \omega_i, u_n = \lambda_m)$ et les K densités gaussiennes de bruit définies par les vecteurs moyenne $\mu = (\mu_k)_{k=1}^K$ et écart-type $\sigma = (\sigma_k)_{k=1}^K$.

Etant donnés les paramètres $\theta = (\varpi, A, \pi, P, \mu, \sigma)$ d'une CMCo M-S, nous pouvons simuler une réalisation de cette dernière comme suit :

- Simuler la réalisation du processus auxiliaire U selon le formalisme d'une chaîne de Markov classique par tirage aléatoire à partir de ϖ et A respectivement ;
- La réalisation x_1 est simulée par tirage à partir du vecteur π_m correspondant à la réalisation u_1 et les prochaines réalisations x_n du processus caché sont simulées conformément aux matrices de transitions P_m correspondantes aux réalisations u_n .
- La réalisation du processus observable Y est obtenue par tirages aléatoires à partir des densités gaussiennes correspondantes aux réalisations du processus caché.

2.3.3.3. Inférence dans les chaînes de Markov cachées M-stationnaires

Nous décrivons maintenant la technique de restauration MPM correspondante au modèle CMC M-S. Soit $T = (U, X, Y) = (V, Y)$ où $Z = (X, Y)$ est CMC M-S de paramètres $\theta = (\varpi, A, \pi, P, \mu, \sigma)$. Pour alléger l'écriture, nous utiliserons la notation v_n au lieu de (u_n, x_n) à chaque fois que cela est possible.

Pour le calcul de l'estimateur MPM donné par (2.4), nous définissons les fonctions de probabilité progressive $\alpha_n(v_n) = p(y_1, \dots, y_n, v_n)$ et rétrograde $\beta_n(v_n) = p(y_{n+1}, \dots, y_N | v_n)$ et dont le calcul se fait de la manière itérative suivante

Initialisation :

$$\alpha_1(v_1) = p(t_1) \quad (2.44)$$

$$\beta_N(v_N) = 1 \quad (2.45)$$

Induction :

$$\alpha_n(v_n) = \sum_{x_{n-1}} \alpha_{n-1}(v_{n-1}) p(t_n | t_{n-1}) \quad (2.46)$$

$$\beta_n(v_n) = \sum_{x_{n+1}} \beta_{n+1}(v_{n+1}) p(t_{n+1} | t_n) \quad (2.47)$$

où les probabilités $p(t_1)$ et $p(t_n | t_{n-1})$ sont données par (2.42) et (2.43) respectivement.

Les distributions *a posteriori* $p(v_n | y)$ se calculent selon la formule

$$p(v_n | y) \propto \alpha_n(v_n) \beta_n(v_n) \quad (2.48)$$

Les distributions *a posteriori* $p(x_n | y)$ requises pour calculer l'estimateur (2.4) sont déduites selon la formule

$$p(x_n | y) = \sum_{v_n} p(v_n | y) \quad (2.49)$$

Il est parfois utile de calculer les distributions *a posteriori* $p(u_n | y)$ qui se calculent de manière similaire

$$p(u_n | y) = \sum_{x_n} p(v_n | y) \quad (2.50)$$

Enfin, les fonctions de probabilité *a posteriori* $\psi_n(v_n, v_{n+1}) = p(v_n, v_{n+1} | y)$ et $\xi_n(v_n) = p(v_n | y)$ requises pour l'algorithme d'estimation des paramètres peuvent être calculées selon les formules suivantes

$$\psi_n(v_n, v_{n+1}) \propto \alpha_n(v_n) p(t_n | t_{n-1}) \beta_{n+1}(v_{n+1}) \quad (2.51)$$

$$\xi_n(v_n) = \sum_{v_{n+1}} \psi_n(v_n, v_{n+1}) \quad (2.52)$$

2.3.3.4. Estimation des paramètres d'une chaîne de Markov cachée M-stationnaire

Nous montrons à présent comment estimer le jeu de paramètres $\Theta = (\varpi, A, \pi, P, \mu, \sigma)$ à partir d'une séquence observable y à travers l'algorithme EM. La procédure d'estimation se déroule de la manière itérative suivante [Lan06]:

- Considérer un jeu initial de paramètres $\Theta^0 = (\varpi, A, \pi, P, \mu, \sigma)^0$.
- Pour chaque itération $q + 1$, calculer Θ^{q+1} à partir de Θ^q en deux étapes jusqu'à ce qu'un critère d'arrêt soit satisfait :
 - o Estimation : Calculer les fonctions ψ_n^q and ξ_n^q à partir de y et Θ^q .
 - o Maximisation : Calculer Θ^{q+1} selon les formules :

$$\bar{\omega}_m^{q+1} = \sum_{x_n} \xi_1(x_n, u_n = \lambda_m) \quad (3.53)$$

$$\pi_{m,k}^{q+1} = \frac{\xi_1(x_n = \omega_k, u_n = \lambda_m)}{\sum_{x_n} \xi_1(x_n, u_n = \lambda_m)} \quad (3.54)$$

$$a_{l,m}^{q+1} = \frac{\sum_{n=1}^{N-1} \sum_{x_n, x_{n+1}} \psi_n(u_n = \lambda_l, x_n, u_{n+1} = \lambda_m, x_{n+1})}{\sum_{n=1}^{N-1} \sum_{x_n} \xi_n(u_n = \lambda_l, x_n)} \quad (3.55)$$

$$p_{m,i,j}^{q+1} = \frac{\sum_{n=1}^{N-1} \sum_{u_n} \psi_n(u_n, x_n = \omega_i, u_{n+1} = \lambda_m, x_{n+1} = \omega_j)}{\sum_{n=1}^{N-1} \sum_{u_n, x_{n+1}} \psi_n(u_n, x_n = \omega_i, u_{n+1} = \lambda_m, x_{n+1})} \quad (3.56)$$

$$\mu_k^{q+1} = \frac{\sum_{n=1}^{N-1} \sum_{u_n} \xi_n(x_n = \omega_k, u_n) y_n}{\sum_{n=1}^{N-1} \sum_{u_n} \xi_n(x_n = \omega_k, u_n)} \quad (3.57)$$

$$\sigma_k^{q+1} = \frac{\sum_{n=1}^{N-1} \sum_{u_n} \xi_n(x_n = \omega_k, u_n) (y_n - \mu_k^{q+1})^2}{\sum_{n=1}^{N-1} \sum_{u_n} \xi_n(x_n = \omega_k, u_n)} \quad (3.58)$$

2.4. Conclusion

Dans ce chapitre, nous avons décrit trois modèles de généralité croissante. Pour chacun de ces modèles, nous avons présenté la technique d'estimation MPM et la procédure d'estimation des paramètres. La modélisation des données non-stationnaires étant un volet crucial dans la présente thèse, nous avons décrit un modèle du type CMT définis pour ce problème particulier. Ce modèle, qui s'appelle CMC M-S, sera comparé à un autre modèle du type CMT défini pour le même problème dans le prochain chapitre.

Chapitre 3. Chaînes de Markov et théorie de l'évidence

Dans le présent chapitre, nous présentons brièvement la théorie de l'évidence introduite par Dempster en 1967 [Dem67] et reformulée par Shafer en 1976 [Sha76]. Cette théorie constitue une généralisation de la théorie bayésienne dans la mesure où elle permet de prendre en compte l'imprécision ou l'incertitude entachant les données. Après un bref aperçu sur ses fondements, nous établissons le lien entre cette théorie et les modèles de Markov en considérant problème de segmentation, qui fait l'objet de cette étude. Ensuite, un modèle évidentiel appelé « chaîne de Markov évidentielle cachée » introduit dans [Fou00, Lan05] est décrit pour illustrer l'apport de ladite théorie dans le problème de segmentation de données non-stationnaires. Puis, ce même modèle est comparé avec un autre modèle bayésien du type CMT introduit dans [Lan06, Lan11] pour le problème de segmentation de données non-stationnaires. Enfin, nous définissons un modèle évidentiel multi-senseurs défini pour la fusion de données multi-senseurs dans un contexte markovien.

3.1. Généralités sur la théorie de l'évidence

Soit $\Omega = \{\omega_1, \dots, \omega_K\}$ un cadre de discernement contenant K hypothèses exclusives et soit $\mathcal{P}(\Omega) = \{\emptyset, \{\omega_1\}, \dots, \Omega\}$ l'ensemble de sous-ensembles de ce cadre de discernement. Les méthodes probabilistes restreignent le domaine de représentativité de tout type de connaissances préalables aux simples hypothèses exclusives. De ce fait, elles sont rapidement mises en défaut lorsque les conditions d'observation réelles ne s'adaptent pas à ce domaine de représentativité. A titre d'exemple, nous pouvons citer les informations fournies par des senseurs insensibles à l'une des hypothèses, ou encore la présence de classes parasites dans l'observation fournie par certains senseurs [Ben01]. Il convient alors de faire appel aux théories de l'incertain, propres à fournir un meilleur cadre théorique pour ce genre de problème [App91]. Parmi ces théories, la théorie de l'évidence de Dempster-Shafer semble bien adaptée dans la mesure où elle généralise l'inférence bayésienne au traitement de grandeurs incertaines issues de l'évaluation d'hypothèses non-exclusives grâce au concept de fonction de masse. Une telle fonction est une fonction de $\mathcal{P}(\Omega)$ dans \mathbb{R}^+ vérifiant :

$$m(\emptyset) = 0 \quad (3.1)$$

$$\sum_{A \in \mathcal{P}(\Omega)} m(A) = 1 \quad (3.2)$$

Les éléments focaux sont les éléments $A \in \mathcal{P}(\Omega)$ dont la masse $m(A)$ est non nulle. La quantité $m(A)$ représente la partie du degré de croyance placée sur une hypothèse A . Il est donc possible d'affecter une masse aux hypothèses exclusives, mais également aux hypothèses composites, ce qui a pour avantage de ne pas introduire d'information arbitraire pour forcer une répartition de masse sur des hypothèses qu'un senseur ne serait pas en mesure de différencier.

Notons par ailleurs que si les éléments focaux se réduisent aux singletons, la masse devient une probabilité, également appelée masse « bayésienne » ou « probabiliste » par opposition à

masse « évidentielle ». De ce fait, elle peut être vue comme une généralisation de la mesure de probabilité. L'apport de la théorie de l'évidence est donc de rendre possible l'évaluation des hypothèses non-exclusives. Ainsi, une masse $m(A)$ est représentative de la vraisemblance attribuable à l'un des éléments du sous-ensemble A sans aucun discernement entre les différents éléments du sous-ensemble A [App91]. En particulier, $m(\Omega)$ désigne le degré d'incertitude totale.

Une fonction de crédibilité Cr est une fonction définie sur $\mathcal{P}(\Omega)$ et vérifiant

$$Cr(\emptyset) = 0 \quad (3.3)$$

$$Cr(\Omega) = 1 \quad (3.4)$$

$$Cr(\cup_{j \in J} A_j) \leq \sum_{I \subset J, I \neq \emptyset} (-1)^{|I|+1} Cr(\cap_{j \in I} A_j) \quad (3.5)$$

Il est possible de vérifier que la fonction de probabilité est un cas particulier de fonction de crédibilité, obtenue dans le cas de l'égalité de (3.5) [App91].

Les fonctions de masse et de crédibilité étant définies indépendamment, il est possible d'associer à tout jeu de masses définies sur $\mathcal{P}(\Omega)$ un jeu de crédibilités sur le même ensemble et vice versa.

$$Cr(B) = \sum_{A \subset B} m(A) \quad (3.6)$$

$$m(A) = \sum_{B \subset A} (-1)^{|A-B|} Cr(B) \quad (3.7)$$

La fonction de plausibilité $Pl(B)$ peut être définie sur $\mathcal{P}(\Omega)$ à partir des fonctions de masses ou de crédibilités définies sur le même ensemble à travers les relations

$$Pl(B) = \sum_{A \cap B \neq \emptyset} m(A) \quad (3.8)$$

$$Pl(B) = 1 - Cr(\bar{B}) \quad (3.9)$$

où \bar{B} désigne le complément de B dans Ω .

Intuitivement, il est possible d'assimiler la crédibilité à une mesure de vraisemblance minimale et la plausibilité à une mesure de vraisemblance maximale. En effet, à partir des relations, nous pouvons déduire

$$Cr(B) \leq Pl(B) \quad (3.10)$$

L'intervalle $[Cr(B), Pl(B)]$ est appelée intervalle de croyance [App91]. Notons que l'égalité entre ces deux fonctions est obtenue lorsque la fonction de masse s'annule en dehors des singletons, auquel cas (3.10) est une égalité et les deux mesures valent la mesure de probabilité classique.

Lorsque les données à modéliser émanent de distinctes sources $(S_j)_{j=1}^R$, la théorie de l'évidence fournit un formalisme intéressant permettant de fusionner les différentes fonctions de masses élémentaires $(m_j)_{j=1}^R$ en une seule masse élémentaire m .

$$m(.) = m_1(.) \oplus m_2(.) \oplus \dots \oplus m_R(.) \quad (3.11)$$

La fusion de Dempster Shafer (DS) est donnée par la formule

$$m(A) = (1 - \mathcal{K})^{-1} \sum_{\cap_{i=1}^R A_i = A \neq \emptyset} \left\{ \prod_{i=1}^R m_i(A_i) \right\} \quad (3.12)$$

où \mathcal{K} est l'inconsistance de la fusion dont le calcul peut se faire selon

$$\mathcal{K} = \sum_{\cap_{i=1}^R A_i = \emptyset} \left\{ \prod_{i=1}^R m_i(A_i) \right\} \quad (3.13)$$

Il convient de noter que cette règle de combinaison est commutative et associative, deux propriétés primordiales en fusion de données.

Enfin, nous clôturons cet aperçu sur la théorie de l'évidence par la définition de l'affaiblissement, lequel figure parmi les modes opératoires cruciaux dans la pratique.

L'affaiblissement Cr^a d'une crédibilité Cr de coefficient a où $a \in [0,1]$ permet de pondérer la fiabilité de la source d'information lors de sa fusion avec d'autres sources. Le coefficient a est appelé coefficient d'affaiblissement ou de fiabilité. Ainsi, l'affaiblissement modifie le jeu de crédibilités selon les formules

$$Cr^a(\Omega) = 1 \quad (3.14)$$

$$Cr^a(A) = (1 - a) Cr(A), \quad \forall A \neq \Omega \quad (3.15)$$

Pareillement, l'affaiblissement modifie le jeu de masses comme suit

$$m^a(A) = (1 - a) m(A), \quad \forall A \neq \Omega \quad (3.16)$$

$$m^a(\Omega) = a + (1 - a) m(\Omega) \quad (3.17)$$

3.1. Chaînes de Markov cachées et théorie de l'évidence

Pour établir un lien entre la fusion de Dempster-Shafer et la présente thèse, considérons cet exemple de segmentation d'image satellitaire où le cadre de discernement est $\Omega = \{\omega_1, \omega_2\}$ avec ω_1 et ω_2 correspondant aux classes « forêt » et « eau » respectivement. Les hypothèses exclusives modélisent le fait qu'un pixel de l'image appartient soit à la classe « forêt » soit à la classe « eau ». En considérant des hypothèses composites, la théorie de l'évidence offre un formalisme particulièrement élégant pour modéliser l'incertitude, l'imprécision ou même le manque d'information grâce à la notion de fonction de masse. Comme vu plus haut, cette fonction de masse généralise la mesure de probabilité. Cette généralisation est une idée clé qui va être exploitée plus tard pour étendre les modèles bayésiens conventionnels.

Dans le problème de segmentation d'image satellitaire défini plus haut, considérons une classification aveugle pixel par pixel. L'information *a priori* des classes ω_1 et ω_2 peut être modélisée à travers une masse bayésienne m_1 définie sur $\{\{\omega_1\}, \{\omega_2\}\}$ par $m_1(\omega_k) = p(x_n = \omega_k)$. Par ailleurs, l'information « observation » peut être modélisée à travers la masse bayésienne m_2 définie sur Ω par $m_2(\omega_k) \propto p(y_n | x_n = \omega_k)$. La fusion DS des deux masses donne $m(x_n) =$

$(m_1 \oplus m_2)(x_n) = p(x_n|y_n)$. Supposons maintenant que l'image présente des nuages [Ben01, Heg98], nous avons donc trois classes observables « forêt », « eau » et « nuage ». Le concept de masse évidentielle peut être introduit ici pour modéliser le fait que nous ne pouvons pas voir à travers les nuages. Explicitement, nous pouvons utiliser une fonction de masse m_2 définie sur $\{\{\omega_1\}, \{\omega_2\}, \Omega\}$ par $m_2(\omega_1) \propto p(y_n|\omega_1)$, $m_2(\omega_2) \propto p(y_n|\omega_2)$ et $m_2(\Omega) \propto p(y_n|\Omega)$.

Pareillement, la fusion DS des deux masses $m(x_n) = (m_1 \oplus m_2)(x_n)$ donne une masse bayésienne qui généralise la probabilité *a posteriori* $p(x_n|y_n)$. Nous rappelons que la règle de combinaison DS d'un ensemble de fonctions de masse $m_{1..R}$ est donnée par la formule (3.12) qui peut être réécrite comme suit

$$m(A) = (m_1 \oplus \dots \oplus m_R)(A) \propto \sum_{\cap_{i=1}^R A_i = A \neq \emptyset} \left\{ \prod_{i=1}^R m_i(A_i) \right\} \quad (3.18)$$

Rappelons que lorsqu'au moins l'une des masses impliquées dans une fusion est bayésienne, le résultat de cette fusion, conformément à la formule (3.12), est également bayésien.

Revenons au problème de la segmentation de l'image satellitaire et supposons maintenant que la distribution *a priori* de la segmentation recherchée X est de forme markovienne. L'information *a priori* est donc modélisée par une fonction de masse m_1 définie sur Ω par $m_1(x) \propto p(x_1)p(x_2|x_1) \dots p(x_N|x_{N-1})$. Le résultat intéressant est que la fusion DS des deux masses est toujours la probabilité *a posteriori* $p(x|y)$.

$$m(x) = (m_1 \oplus m_2)(x) = \frac{p(x)p(y|x)}{\sum_{x' \in \Omega^N} p(x')p(y|x')} = p(x|y) \quad (3.19)$$

L'étape suivante est de tirer profit des avantages de la théorie markovienne et la théorie de l'évidence en généralisant les masses bayésiennes aux masses évidentielles et exploiter ainsi le résultat de la formule (3.19). Cependant, quand au moins l'une des masses impliquées dans la fusion DS est évidentielle, le résultat de cette fusion n'est plus une chaîne de Markov et par conséquent, les différents traitements d'estimation ne sont plus faisables directement. Toutefois, il a été établi que le résultat de cette fusion est une chaîne de Markov triplet, ce qui rend les calculs associés aux différentes procédures d'estimation réalisables [Pie07].

Ce résultat essentiel a été employé pour modéliser les données non-stationnaires dans les chaînes de Markov évidentielles cachées (CMEC). L'objectif de ce chapitre est de comparer le modèle évidentiel CMEC au modèle CMC M-S dans la tâche de segmentation des données non-stationnaires.

3.2. Chaînes de Markov évidentielles cachées

Dans cette section, nous rappelons les fondements théoriques des chaînes de Markov évidentielles cachées introduites par Lanchantin et Pieczynski dans [Lan05, Fou00] pour la modélisation des données non-stationnaires. Dans un premier temps, nous définissons le modèle de chaîne de Markov évidentielle. Puis, nous décrivons le modèle de chaîne de Markov

évidentielle cachée ainsi que les différentes procédures d'estimations d'intérêt qui lui sont associés.

3.2.1. Chaînes de Markov évidentielles

Toute fonction de masse m définie sur $\mathcal{P}(\Omega^N)$ et vérifiant ce qui suit :

- m s'annule en dehors de $\mathcal{P}(\Omega)^N$;
- m est d'une forme markovienne

$$m(u) = m(u_1)m(u_2|u_1) \dots m(u_N|u_{N-1}), \forall u \in \mathcal{P}(\Omega)^N \quad (3.20)$$

est appelée chaîne de Markov évidentielle (CME).

Le modèle de chaîne de Markov classique peut alors être vu comme un cas particulier d'une CME s'annulant partout sauf éventuellement sur $u = (u_1, \dots, u_N)$ où chacun des u_n est un singleton.

Soit $X = (X_n)_{n=1}^N$ un processus caché qui prend ses valeurs à partir d'un ensemble fini de classes $\Omega = \{\omega_1, \dots, \omega_K\}$. X est de Markov si sa loi est donnée par

$$p(x) = p(x_1)p(x_2|x_1) \dots p(x_N|x_{N-1}), \forall x \in \Omega^N \quad (3.21)$$

Les distributions $p(x_n|x_{n-1})$ sont généralement supposées indépendantes de n . En adoptant la matrice de transition A où $a_{x_{n-1}, x_n} = p(x_n|x_{n-1})$, la formule (3.21) devient

$$p(x) = p(x_1) \prod_{n=2}^N a_{x_{n-1}, x_n}, \forall x \in \Omega^N \quad (3.22)$$

Si tel n'est pas le cas, l'estimation des paramètres, en supposant la chaîne X stationnaire, fournit une matrice de transition commune A^* pour toutes les distributions $p(x_n|x_{n-1})$. Il est possible de remplacer la chaîne de Markov de la formule (3.22) par la CME m afin de prendre en compte les fluctuations de $p(x_n|x_{n-1})$.

Soit $Y = (Y_n)_{n=1}^N$ un processus observable $Y = (Y_n)_{n=1}^N$ où $Y_n \in \mathbb{R}$ qui peut être considéré comme une version bruitée de Y . Il est possible de fusionner la masse bayésienne m_2 définie sur Ω^N par

$$m_2(x) \propto p(y|x) = \prod_{n=1}^N p(y_n|x_n) \quad (3.23)$$

avec une CME m_1 , et conformément à (3.19), le résultat de cette fusion lorsque $m_1 = p^*$ est $p(x|y)$. Toutefois, lorsque la masse m_1 n'est pas bayésienne, le résultat de cette fusion ($m_1 \oplus m_2$)(x) n'est plus une chaîne de Markov. Il a été établi cependant que c'est une distribution marginale d'une CMT et par conséquent, les différents calculs d'intérêt restent faisables [Lan05].

3.2.2. Définition d'une chaîne de Markov cachée évidentielle

Soit $X = (X_n)_{n=1}^N$ un processus caché qui prend ses valeurs à partir d'un ensemble fini de classes $\Omega = \{\omega_1, \dots, \omega_K\}$ et qui doit être estimé à partir d'un processus observable $Y = (Y_n)_{n=1}^N$ où $Y_n \in \mathbb{R}$. Soit le processus auxiliaire $U = (U_n)_{n=1}^N$ prenant ses valeurs à partir de $\mathcal{P}(\Omega)$. Soit m_1

une CME définie sur $\mathcal{P}(\Omega)^N$ par la formule (3.20) et soit $m_2(x) \propto p(y|x) = \prod_{n=1}^N p(y_n|x_n)$ une fonction de masse bayésienne déterminée par l'observation y . Soit la CMT définie par

$$p(t) \propto q_1(t_1, t_2) q_2(t_2, t_3) \dots q_{N-1}(t_{N-1}, t_N) \quad (3.24)$$

avec

$$q_1(t_1, t_2) = \mathbf{1}_{[x_1 \in u_1]} \mathbf{1}_{[x_2 \in u_2]} m_1(u_1) m_1(u_2|u_1) m_2(x_1) m_2(x_2);$$

$$q_n(t_n, t_{n+1}) = \mathbf{1}_{[x_{n+1} \in u_{n+1}]} m_1(u_{n+1}|u_n) m_2(x_{n+1}).$$

Alors $(m_1 \oplus m_2)(x)$ est la distribution *a posteriori* $p(x|y)$ associée à $p(x, y)$ qui est une distribution marginale de la CMT définie par (3.24). Il s'en suit que les probabilités d'intérêt $p(u_n, x_n|y)$ sont calculables et ainsi le sont les probabilités $p(x_n|y)$.

Le processus couple $Z = (X, Y)$ est alors dit chaîne de Markov évidentielle cachée (à bruit indépendant).

3.1.1. Restauration MPM d'une CMEC

Soit $T = (U, X, Y) = (V, Y)$ où $Z = (X, Y)$ est une CMEC. Pour compacter l'écriture, nous utiliserons à chaque fois que cela sera possible v_n à la place de (u_n, x_n) . Afin d'accomplir la restauration bayésienne MPM, les fonctions de probabilités progressives $\alpha_n(v_n) = p(y_1, \dots, y_n, v_n)$ et rétrogrades $\beta_n(v_n) = p(y_{n+1}, \dots, y_N|v_n)$ adaptées au contexte des CMEC sont introduites. Ces dernières peuvent être calculées de la manière itérative suivante

Initialisation :

$$\alpha_1(v_1) = p(v_1, y_1) \quad (3.25)$$

$$\beta_N(v_N) = 1 \quad (3.26)$$

Induction :

$$\alpha_n(v_n) = \sum_{v_{n-1}} \alpha_{n-1}(v_{n-1}) p(v_n|v_{n-1}) m_2(x_n) \quad (3.27)$$

$$\beta_n(v_n) = \sum_{v_{n+1}} \beta_{n+1}(v_{n+1}) p(v_{n+1}|v_n) m_2(x_{n+1}) \quad (3.28)$$

avec

$$p(v_{n+1}|v_n) \propto \mathbf{1}_{[x_n \in u_n]} \mathbf{1}_{[x_{n+1} \in u_{n+1}]} m_1(u_{n+1}|u_n);$$

$$p(v_1, y_1) \propto \mathbf{1}_{[x_1 \in u_1]} m_1(u_1) m_2(x_1).$$

Par ailleurs, les probabilités marginales *a posteriori* $p(v_n|y)$ et $p(x_n|y)$ se calculent comme suit :

$$p(v_n|y) \propto \alpha_n(v_n) \beta_n(v_n) \quad (3.29)$$

$$p(x_n|y) = \sum_{u_n} p(u_n, x_n|y) \quad (3.30)$$

Enfin, les probabilités jointes et marginales *a posteriori* utiles pour l'algorithme d'estimation des paramètres sont calculées selon les formules

$$\psi(v_n, v_{n+1}) = p(v_n, v_{n+1}|y) = \alpha_n(v_n) p(v_{n+1}|v_n) m_2(x_{n+1}) \beta_{n+1}(v_{n+1}) \quad (3.31)$$

$$\xi(v_n) = p(v_n|y) = \sum_{v_{n+1}} \psi(v_n, v_{n+1}) \quad (3.32)$$

3.2.3. Estimation des paramètres d'une CMEC

Pour estimer les paramètres du modèle, nous proposons d'utiliser une version adaptée de l'algorithme EM. Pour ce faire, considérons la CMT $T = (X, U, Y) = (V, Y)$ définie plus haut. Supposons également que les densités de bruit $p(y_n|x_n)$ sont gaussiennes.

Par conséquent, nous devons estimer les paramètres suivants : les masses évidentielles $m_{ij} = m_1(u_{n+1} = \lambda_j | u_n = \lambda_i)$ définie sur $\mathcal{P}(\Omega)^2$, les K moyennes $(\mu_k)_{k=1}^K$ et K écarts-types $(\sigma_k)_{k=1}^K$ des densités gaussiennes régissant m_2 .

La procédure d'estimation se fait de la manière suivante :

- Considérer un jeu initial de paramètres $\theta^0 = (m_{ij}^0, \mu_k^0, \sigma_k^0)$.
- Pour chaque itération $q + 1$, calculer θ^{q+1} à partir de θ^q en deux étapes jusqu'à ce qu'un critère d'arrêt soit satisfait :
 - o Estimation : Calculer les fonctions ψ_n^q and ξ_n^q à partir de y et θ^q .
 - o Maximisation : Calculer θ^{q+1} selon les formules :

$$\mu_k^{q+1} = \frac{\sum_{n=1}^N \sum_{u_n} \xi^q(x_n = \omega_k, u_n) y_n}{\sum_{n=1}^N \sum_{u_n} \xi^q(x_n = \omega_k, u_n)} \quad (3.33)$$

$$\sigma_k^{q+1} = \frac{\sum_{n=1}^N \sum_{x_n} \xi^q(x_n, u_n = \lambda_k) (y_n - \mu_k^{q+1})^2}{\sum_{n=1}^N \sum_{x_n} \xi^q(x_n, u_n = \lambda_k)} \quad (3.34)$$

$$m_{ij}^{q+1} = \frac{\sum_{n=1}^N \sum_{x_n, x_{n+1}} \psi^q(u_n = \lambda_i, x_n, u_{n+1} = \lambda_j, x_{n+1})}{(N-1) \#i \#j} \quad (3.35)$$

où $\#i$ désigne le cardinal de λ_i .

3.3. Sur le choix de la modélisation dans le cas d'un signal monodimensionnel non-stationnaire

La modélisation de données non-stationnaires via les modèles de CMT a été réalisée dans un contexte bayésien à travers les CMC M-S et dans un contexte évidentiel à travers les CMEC. Dans cette section, nous comparons les performances des deux modèles dans le problème de segmentation de données non-stationnaires. Pour ce faire, nous menons deux séries d'expériences. Dans la première série, nous considérons des données synthétiques que nous simulons de différentes façons. Dans la seconde série, par ailleurs, une image de classes non-stationnaire est bruitée selon le formalisme des CMC M-S.

3.3.1. Segmentation non-supervisée de CMC non-stationnaires

Dans cette série, nous considérons des données synthétiques non-stationnaires que nous générons de différentes façons selon le formalisme des CMC M-S. Plus explicitement, les expériences sont menées sur trois jeux de données de caractéristiques différentes :

- Jeu A : Pour ce jeu, les données sont simulées selon le formalisme des CMC M-S. Soit $T = (U, X, Y)$ une CMC M-S avec $T = (T_n)_{n=1}^N$ où $N = 4096$, U_n prend ses valeurs à partir de l'ensemble $\Lambda = \{\lambda_1, \dots, \lambda_M\}$, X_n prend ses valeurs à partir de l'ensemble $\Omega = \{\omega_1, \omega_2\}$ et Y_n prend ses valeurs à partir de l'ensemble des réels \mathbb{R} . Les valeurs de M allant de 1 à 9 sont considérées. Ainsi, 9 expériences sont menées. Pour chaque valeur de M , les matrices M_1, \dots, M_M sont utilisées

$$M_1 = \begin{pmatrix} 0.99 & 0.01 \\ 0.01 & 0.99 \end{pmatrix}, M_2 = \begin{pmatrix} 0.88 & 0.12 \\ 0.12 & 0.88 \end{pmatrix}, M_3 = \begin{pmatrix} 0.77 & 0.23 \\ 0.23 & 0.77 \end{pmatrix},$$

$$M_4 = \begin{pmatrix} 0.66 & 0.34 \\ 0.34 & 0.66 \end{pmatrix}, M_5 = \begin{pmatrix} 0.55 & 0.45 \\ 0.45 & 0.55 \end{pmatrix}, M_6 = \begin{pmatrix} 0.44 & 0.56 \\ 0.56 & 0.44 \end{pmatrix},$$

$$M_7 = \begin{pmatrix} 0.33 & 0.67 \\ 0.67 & 0.33 \end{pmatrix}, M_8 = \begin{pmatrix} 0.22 & 0.78 \\ 0.78 & 0.22 \end{pmatrix}, M_9 = \begin{pmatrix} 0.11 & 0.89 \\ 0.89 & 0.11 \end{pmatrix}.$$

- La première réalisation u_1 du processus auxiliaire est simulée par un tirage aléatoire uniforme à partir de Λ . Les prochaines réalisations du même processus sont simulées via un tirage aléatoire à partir de la matrice de transition :

$$Q = \frac{1}{10^3} \begin{bmatrix} 995 & \dots & \frac{5}{M-1} \\ \vdots & \ddots & \vdots \\ \frac{5}{M-1} & \dots & 995 \end{bmatrix}.$$

- La première réalisation x_1 du processus caché est simulée via un tirage aléatoire uniforme à partir de l'ensemble des classes Ω . Les prochaines réalisations du même processus sont simulées à partir des tirages aléatoires selon la matrice de transition correspondante à la réalisation u_n .
- Les réalisations du processus observable Y sont simulées selon les densités gaussiennes associées à ω_1 et ω_2 de moyennes 0 et 2 respectivement et d'écart-types égaux à 1.

Pour ce jeu de données, la segmentation MPM est réalisée sur la base des paramètres estimés par EM selon les CMEC et les CMC M-S en considérant pour chaque expérience les valeurs de M allant de 1 à 9. Les résultats obtenus sont présentés dans TAB. 3.I.

Tableau 3.1. Taux d'erreur de la segmentation non-supervisée des données non-stationnaires synthétiques, Jeu A (%)

Modèle	CMCE	CMC M-S									
		CMC 1-S	CMC 2-S	CMC 3-S	CMC 4-S	CMC 5-S	CMC 6-S	CMC 7-S	CMC 8-S	CMC 9-S	CMC 10-S
Exp. A.1	0.66	0.66	0.66	0.68	0.68	0.71	0.68	0.68	0.71	0.68	0.66
Exp. A.2	4.25	4.35	4.13	4.13	4.13	4.15	4.20	4.13	4.22	4.22	4.22
Exp. A.3	7.59	7.80	7.08	7.10	7.04	7.04	7.04	7.02	7.03	7.04	7.03
Exp. A.4	9.46	9.92	8.88	8.81	8.76	8.78	8.79	8.79	8.80	8.79	8.78
Exp. A.5	10.36	11.68	10.39	9.99	9.88	9.85	9.86	9.85	9.85	9.86	9.86
Exp. A.6	11.73	13.70	11.73	11.36	11.26	11.28	11.29	11.30	11.32	11.30	11.30
Exp. A.7	12.85	15.11	12.79	12.40	12.28	12.24	12.25	12.26	12.26	12.27	12.27
Exp. A.8	13.13	15.48	13.04	12.26	12.14	12.05	12.03	12.05	12.03	12.02	12.02
Exp. A.9	14.27	15.81	12.96	12.09	11.73	11.64	11.61	11.61	11.63	11.63	11.64

- Jeu B : Soit $Z = (X, Y)$ une chaîne de Markov cachée non-homogène avec $Z = (Z_n)_{n=1}^N$ où $N = 4096$, X_n prend ses valeurs à partir de l'ensemble $\Omega = \{\omega_1, \omega_2\}$ et Y_n prend ses valeurs à partir de l'ensemble des réels \mathbb{R} . La loi de Z est donnée par

$$p(x, y) = p(x_1)p(y_1|x_1) \prod_{n=2}^N p(x_n|x_{n-1}) p(y_n|x_n)$$

où les distributions de bruit $p(y_n|x_n)$ associées aux classes ω_1 et ω_2 sont des densités gaussiennes $\mathcal{N}(0,1)$ et $\mathcal{N}(2,1)$ respectivement.

Les transitions $p(x_n|x_{n-1})$, quant à elles, sont données par la matrice A_n définie comme suit

$$A_n = \begin{pmatrix} \delta_n & 1 - \delta_n \\ 1 - \delta_n & \delta_n \end{pmatrix}$$

- Expérience B.1 : $\forall 1 \leq n \leq N: \delta_n = \frac{n}{N}$;
- Expérience B.2 : $\forall 1 \leq n \leq N: \delta_n = \frac{3}{4} + \frac{1}{4} \sin\left(\frac{n}{5}\right)$.
- Jeu C : Soit $Z = (X, Y)$ une chaîne de Markov cachée non-homogène avec $Z = (Z_n)_{n=1}^N$ où $N = 4096$, X_n prend ses valeurs à partir de l'ensemble $\Omega = \{\omega_1, \omega_2, \omega_3\}$ et Y_n prend ses valeurs à partir de l'ensemble des réels \mathbb{R} . La loi de Z est donnée par

$$p(x, y) = p(x_1)p(y_1|x_1) \prod_{n=2}^N p(x_n|x_{n-1}) p(y_n|x_n)$$

où les distributions de bruit $p(y_n|x_n)$ associées aux classes ω_1 , ω_2 et ω_3 sont les densités gaussiennes $\mathcal{N}(0,1)$, $\mathcal{N}(2,1)$ et $\mathcal{N}(4,1)$ respectivement.

Les transitions $p(x_n|x_{n-1})$, quant à elles, sont données par la matrice A_n définie comme suit

$$A_n = \begin{pmatrix} \delta_n & \frac{1 - \delta_n}{2} & \frac{1 - \delta_n}{2} \\ \frac{1 - \delta_n}{2} & \delta_n & \frac{1 - \delta_n}{2} \\ \frac{1 - \delta_n}{2} & \frac{1 - \delta_n}{2} & \delta_n \end{pmatrix}$$

- Expérience C.1 : $\forall 1 \leq n \leq N: \delta_n = \frac{n}{N}$;
- Expérience C.2 : $\forall 1 \leq n \leq N: \delta_n = \frac{3}{4} + \frac{1}{4} \sin\left(\frac{n}{5}\right)$.

Pour les jeux de données B et C, la restauration MPM est effectuée sur la base des paramètres estimés par EM selon les CMEC et les CMC M-S en considérant pour chaque expérience les valeurs de M allant de 1 à 5. Les résultats obtenus sont présentés dans TAB. 3.II. Les résultats obtenus ont également été convertis en images grâce au parcours d'Hilbert-Peano [Pea1890, Hil1891] et sont présentés sur les figures Fig.3.1, Fig. 3.2, Fig. 3.3 et Fig. 3.4.

Tableau 3.2. Taux d'erreur de la segmentation non-supervisée des données non-stationnaires synthétique, Jeux B et C (%)

Modèle	K-means	CMEC	CMC M-S				
			CMC 1-S	CMC 2-S	CMC 3-S	CMC 4-S	CMC 5-S
Exp. B.1	26.1	11.4	25.5	11.0	10.5	10.4	10.3
Exp. B.2	16.0	12.0	12.7	12.1	12.1	12.0	12.0
Exp. C.1	21.7	17.6	45.1	17.8	16.7	16.5	16.5
Exp. C.2	21.2	12.0	17.6	12.2	12.2	12.3	12.2

Pour le jeu de données A, et sur la base des résultats obtenus, nous pouvons émettre les commentaires suivants :

- les résultats obtenus via les CMC M-S sont meilleurs que ceux fournis par les CMEC, notamment pour les bonnes valeurs de M (et généralement pour les valeurs proches ou plus grandes). Ceci est justifié par le fait que les données ont été générées selon les CMC M-S.
- Les CMC M-S recherchent la meilleure régularisation possible qui s'adapte à chaque région de la donnée (pour un nombre de stationnarités M donné), alors que les CMEC adoptent une régularisation unique en préconisant un mécanisme (affaiblissement) qui consiste à pondérer l'importance de la loi *a priori* par rapport à la loi de vraisemblance.
- Dans les régions qui subissent un affaiblissement par le modèle CMEC, la fusion accordera plus d'importance aux observations que l'information contextuelle pour limiter l'effet de la non-stationnarité. En contre partie, les CMC M-S s'appuient sur la régularisation exacte associée à ces régions.
- En conclusion, pour les CMEC, la fusion se fait en considérant moins d'information suite à l'affaiblissement de l'information *a priori* qui manque de fiabilité. Les CMC M-S exploitent toute l'information en adoptant une fusion bayésienne des informations fiables.

A partir des résultats enregistrés pour les jeux de données B et C, il convient de noter ce qui suit :

- D'une manière générale, les deux modèles fournissent des résultats comparables.
- Pour les expériences B.1 et C.1, le paramètre δ_n augmente de 0 jusqu'à 1 le long de la donnée. Pour accomplir la segmentation MPM, le modèle CMC M-S subdivise la donnée en M parties homogènes (ou presque) en préconisant une valeur différente pour chaque partie et parvient ainsi à fournir des résultats meilleurs que ceux enregistrés via les CMEC. Il est possible de vérifier à partir des figures Fig. 3.1 et Fig. 3.3 comment le modèle partitionne la donnée en $M = 3$ parties (Fig. 3.1-h et Fig. 3.3-h), $M = 4$ parties (Fig. 3.1-j et Fig. 3.3-j), en $M = 5$ parties (Fig. 3.1-l et Fig. 3.3-l).
- Pour les expériences B.2 et C.2, le paramètre δ_n est d'une forme sinusoïdale. Pour ce type de matrice de transition, il est difficile de partitionner la donnée en M parties homogènes. En contre partie, le modèle CMEC parvient à bien modéliser ce genre de données grâce au mécanisme d'affaiblissement. En effet, il convient de remarquer que plus la valeur de δ_n est petite, l'affaiblissement est intense et vice versa (Fig. 3.2-g et 3.2-h). En effet, pour les régions à petites valeurs de δ_n , l'information d'observation l'emporte sur l'information contextuelle et par conséquent le mécanisme d'affaiblissement est appliqué aux sites de ces zones.
- Le modèle de CMC classique fournit des résultats comparables avec l'algorithme *K-means* (quelques fois pire) à cause de la forte non-stationnarité des données. La supériorité des résultats obtenus par les modèles de CMEC et CMC M-S démontre l'intérêt de la prise en compte de l'information contextuelle même lorsque les données sont fortement non-stationnaires.

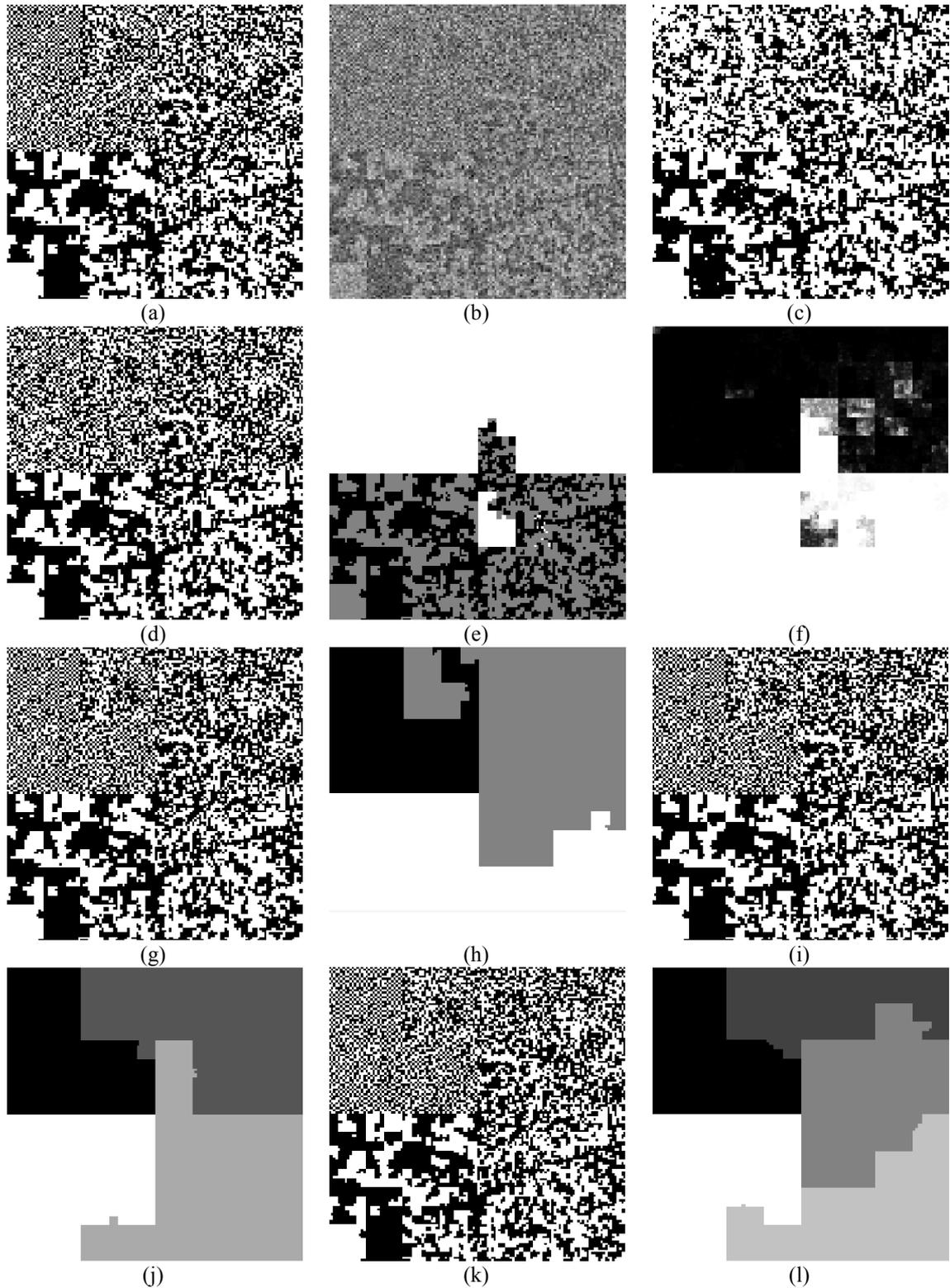


Figure 3.1. Segmentation non-supervisée de données synthétiques à l'aide des CMEC et les CMC M-S (Expérience B.1). (a) Image de classes d'origine $X = x$. (b) Image observée $Y = y$. (c) Segmentation de l'image selon les CMC, taux d'erreur $\tau = 25.5\%$. (d) Segmentation de l'image selon les CMEC, taux d'erreur $\tau = 12\%$. (e) Estimation de la réalisation du processus auxiliaire U selon une CMEC. (f) Valeur de coefficient d'affaiblissement conditionnellement aux observations. (g) Segmentation de l'image selon les CMC 3-S, taux d'erreur $\tau = 10.5\%$. (h) Estimation de la réalisation du processus auxiliaire U selon une CMC 3-S. (i) Segmentation de l'image selon les CMC 4-S, taux d'erreur $\tau = 10.4\%$. (j) Estimation de la réalisation du processus auxiliaire U selon une CMC 4-S. (k) Segmentation de l'image selon les CMC 5-S, taux d'erreur $\tau = 10.3\%$. (l) Estimation de la réalisation du processus auxiliaire U selon une CMC 5-S.

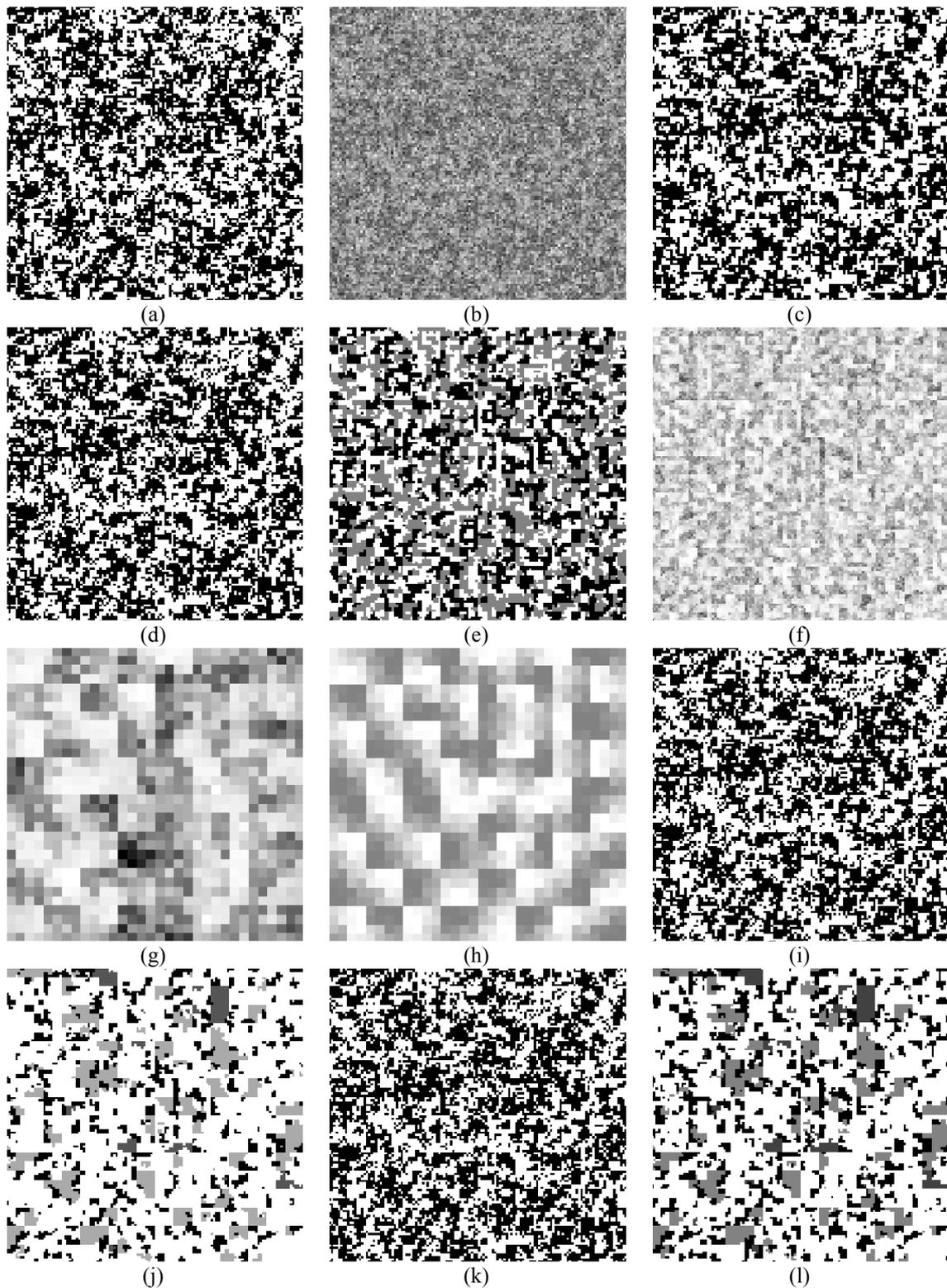


Figure 3.2. Segmentation non-supervisée de données synthétiques à l'aide des CMEC et les CMC M-S (Expérience B.2). (a) Image de classes d'origine $X = x$. (b) Image observée $Y = y$. (c) Segmentation de l'image selon les CMC, taux d'erreur $\tau = 12.7\%$. (d) Segmentation de l'image selon les CMEC, taux d'erreur $\tau = 12\%$. (e) Estimation de la réalisation du processus auxiliaire U selon une CMEC. (f) Valeur de coefficient d'affaiblissement conditionnellement aux observations. (g) 1024 première valeurs du coefficient d'affaiblissement conditionnellement aux observations. (h) 1024 premières valeurs du paramètre. (i) Segmentation de l'image selon les CMC 4-S, taux d'erreur $\tau = 12\%$. (j) Estimation de la réalisation du processus auxiliaire U selon une CMC 4-S. (k) Segmentation de l'image selon les CMC 5-S, taux d'erreur $\tau = 12\%$. (l) Estimation de la réalisation du processus auxiliaire U selon une CMC 5-S.

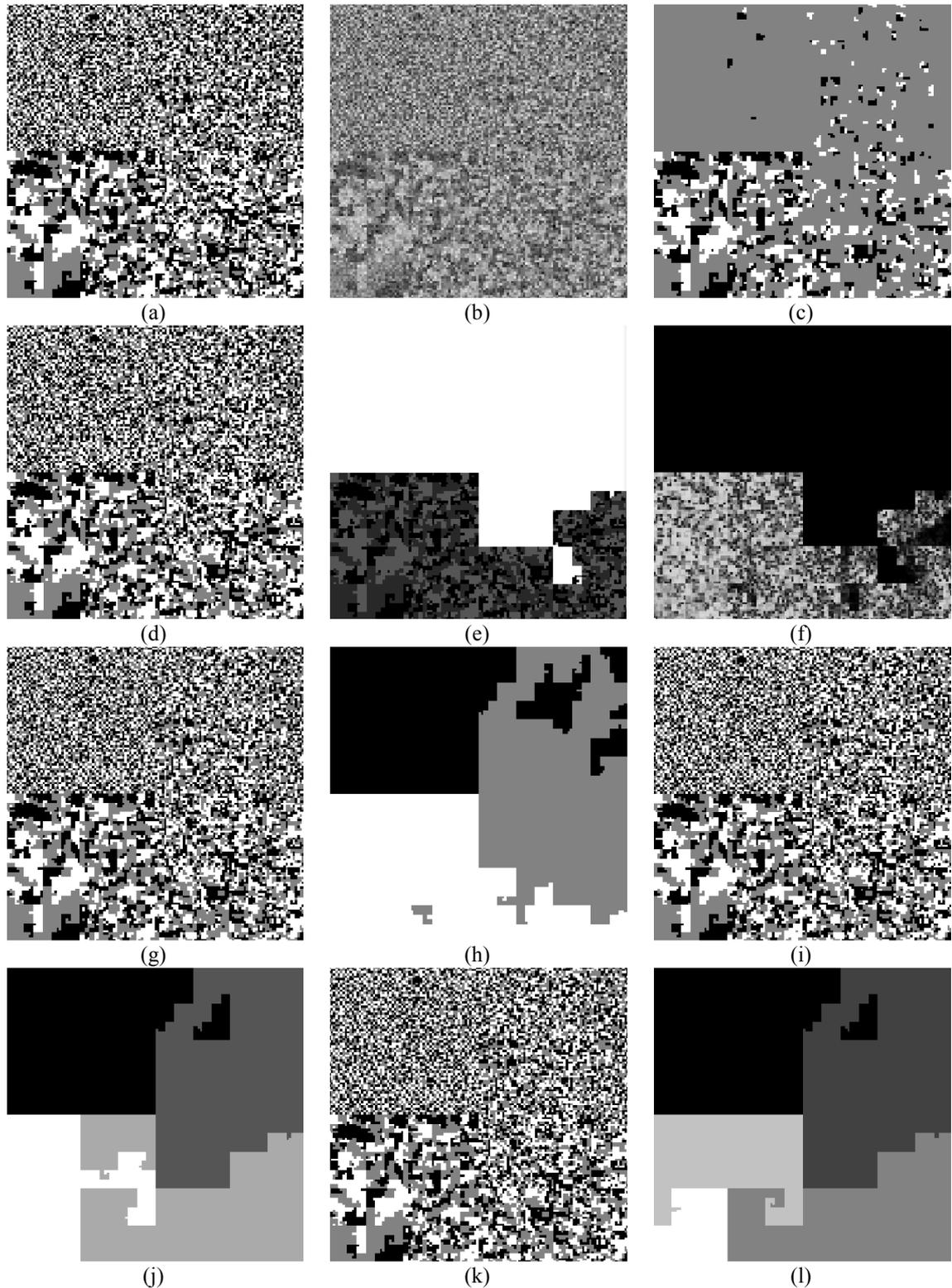


Figure 3.3. Segmentation non-supervisée de données synthétiques à l'aide des CMEC et les CMC M-S (Expérience C.1). (a) Image de classes d'origine $X = x$. (b) Image observée $Y = y$. (c) Segmentation de l'image selon les CMC, taux d'erreur $\tau = 45.1\%$. (d) Segmentation de l'image selon les CMEC, taux d'erreur $\tau = 17.6\%$. (e) Estimation de la réalisation du processus auxiliaire U selon une CMEC. (f) Valeur de coefficient d'affaiblissement conditionnellement aux observations. (g) Segmentation de l'image selon les CMC 3-S, taux d'erreur $\tau = 16.7\%$. (h) Estimation de la réalisation du processus auxiliaire U selon une CMC 3-S. (i) Segmentation de l'image selon les CMC 4-S, taux d'erreur $\tau = 16.5\%$. (j) Estimation de la réalisation du processus auxiliaire U selon une CMC 4-S. (k) Segmentation de l'image selon les CMC 5-S, taux d'erreur $\tau = 16.5\%$. (l) Estimation de la réalisation du processus auxiliaire U selon une CMC 5-S.

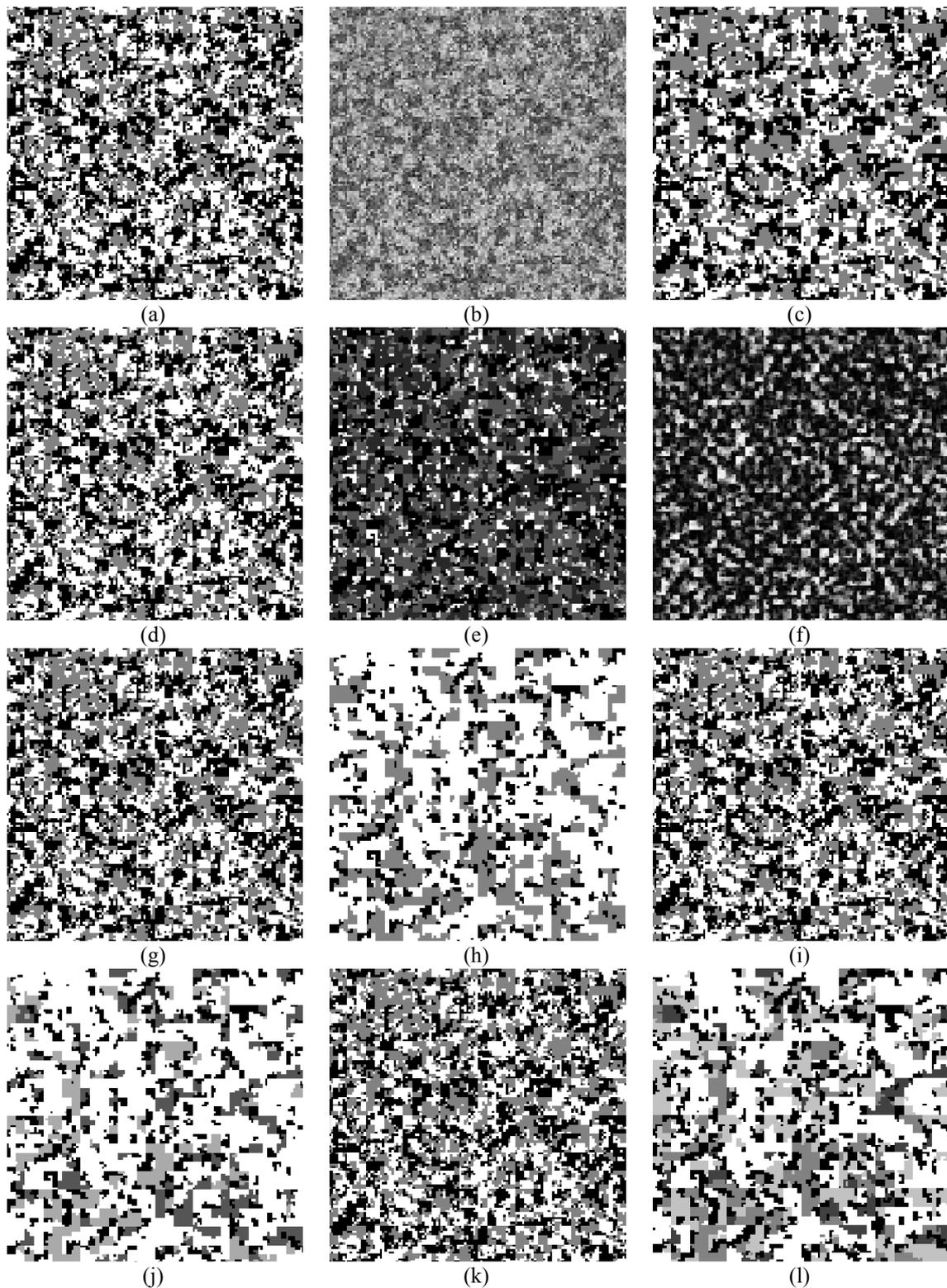


Figure 3.4. Segmentation non-supervisée de données synthétiques à l'aide des CMEC et les CMC M-S (Expérience C.2). (a) Image de classes d'origine $X = x$. (b) Image observée $Y = y$. (c) Segmentation de l'image selon les CMC, taux d'erreur $\tau = 17.6\%$. (d) Segmentation de l'image selon les CMEC, taux d'erreur $\tau = 12\%$. (e) Estimation de la réalisation du processus auxiliaire U selon une CMEC. (f) Valeur de coefficient d'affaiblissement conditionnellement aux observations. (g) Segmentation de l'image selon les CMC 3-S, taux d'erreur $\tau = 12.2\%$. (h) Estimation de la réalisation du processus auxiliaire U selon une CMC 3-S. (i) Segmentation de l'image selon les CMC 4-S, taux d'erreur $\tau = 12.3\%$. (j) Estimation de la réalisation du processus auxiliaire U selon une CMC 4-S. (k) Segmentation de l'image selon les CMC 5-S, taux d'erreur $\tau = 12.2\%$. (l) Estimation de la réalisation du processus auxiliaire U selon une CMC 5-S.

3.3.2. Segmentation non-supervisée d'images non-stationnaires bruitées

Dans cette section, nous considérons des images de classes non-stationnaires telles que les images 'Nazca bird' et 'Zebra' (Fig. 3.1 et Fig. 3.2 respectivement) de tailles 128×128 et 256×256 respectivement.

Soit $T = (U, X, Y)$ une CMC M-S avec $T = (T_n)_{n=1}^N$. Nous avons donc une réalisation du processus caché x où $\Omega = \{\omega_1, \omega_2\}$ avec ω_1 et ω_2 correspondent aux pixels noirs et blanc respectivement. Pour chacune des deux images, les observations sont ensuite générées de deux manières différentes :

- Expérience 1 : Les densités gaussiennes associées à ω_1 et ω_2 sont de moyennes 0 et 2 respectivement et d'écart-types égaux à 1.
- Expérience 2 : Les densités gaussiennes associées à ω_1 et ω_2 sont de moyennes 0 et 1 respectivement et d'écart-types égaux à 1.

Enfin, la segmentation non-supervisée MPM est accomplie selon les modèles CMEC et CMC M-S en considérant différentes valeurs de M . Les résultats obtenus sont illustrés sur le TAB. 3.III et les figures Fig. 3.5, Fig. 3.6, Fig. 3.7 et Fig. 3.8.

Tableau 3.3. Taux d'erreur de la segmentation non-supervisée des images non-stationnaires (%)

Modèle	CMCE	CMC M-S									
		CMC 1-S	CMC 2-S	CMC 3-S	CMC 4-S	CMC 5-S	CMC 6-S	CMC 7-S	CMC 8-S	CMC 9-S	CMC 10-S
<i>Nazca bird</i> , Exp. 1	5.0	13.3	5.2	4.9	5.2	4.9	5.0	4.8	4.8	7.0	5.1
<i>Nazca bird</i> , Exp. 2	11	15.4	16.4	14.6	14.1	11.9	12.2	12.2	12	11.7	11.7
<i>Zebra</i> , Exp. 1	3.7	6.0	3.7	3.3							
<i>Zebra</i> , Exp. 2	8.2	12.3	8.1	7.7	7.6	7.7	7.7	7.8	7.7	7.7	7.7

L'intérêt de cette série d'expérience réside dans le fait que les réalisations du processus caché n'ont pas été simulées et par conséquent, elles ne correspondent à aucun des deux modèles.

Sur la base des résultats obtenus, il est difficile de trancher en faveur de l'une des deux modélisations. Toutefois, il convient d'énoncer les remarques suivantes :

- En l'absence de toute connaissance sur le nombre de stationnarités, le modèle CMEC semble une bonne solution pour modéliser des données non-stationnaire grâce au mécanisme d'affaiblissement qui pondère l'information *a priori* pour pallier aux fluctuations et/ou non-fiabilité de cette dernière.
- Le modèle CMC M-S peut être utilisé pour différentes valeurs de M quand le nombre de stationnarités n'est pas connu. Dans ce cadre, il est possible de faire appel à certains critères tels que le critère d'information bayésien (BIC) pour le bon choix du modèle comme fait dans [Lan11].

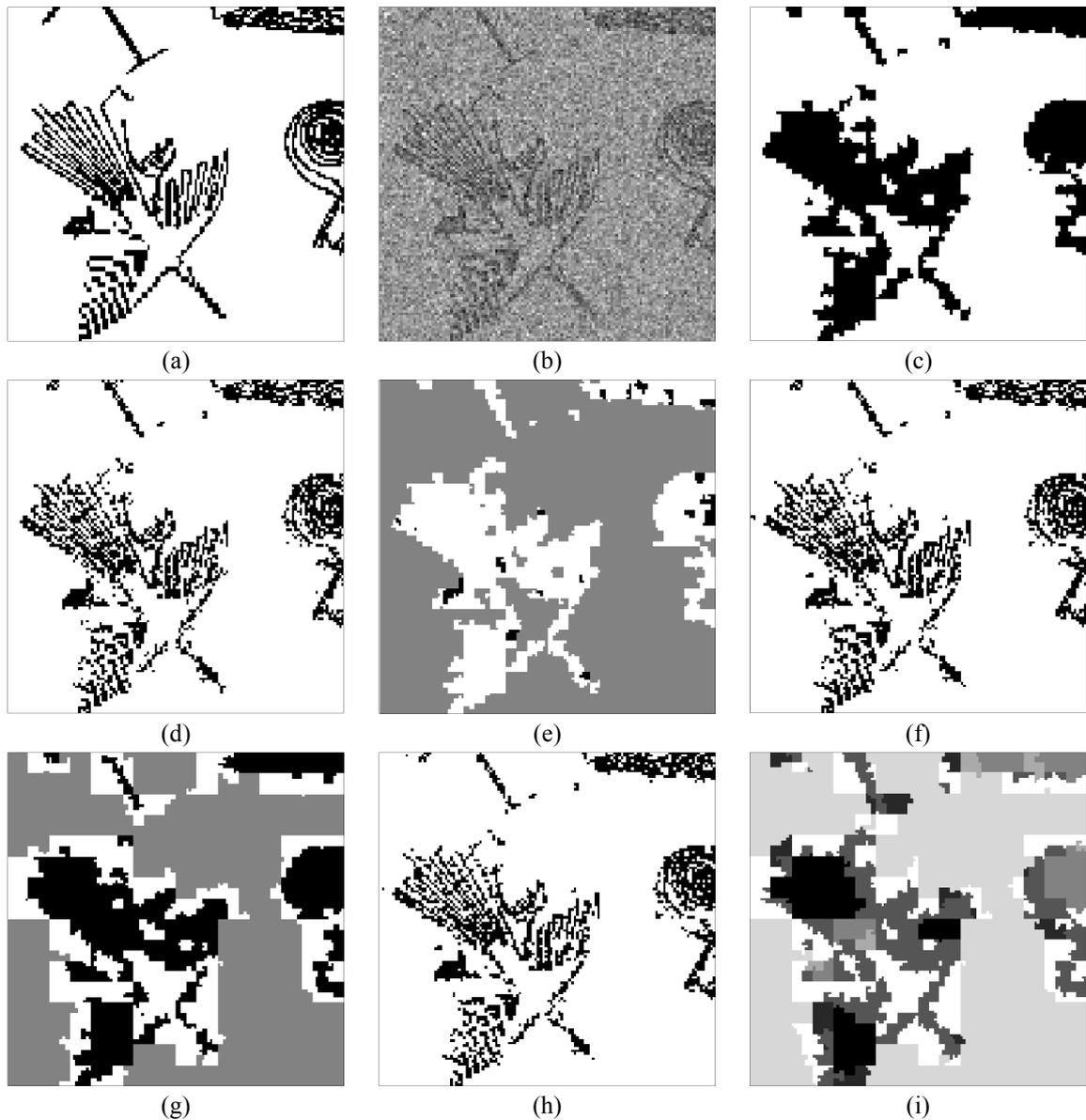


Figure 3.5. Segmentation non-supervisée d'une image non-stationnaire à l'aide des CMEC et les CMC M-S. (a) Image de classes d'origine $X = x$. (b) Image observée $Y = y$. (c) Segmentation de l'image selon les CMC, taux d'erreur $\tau = 13.3\%$. (d) Segmentation de l'image selon les CMEC, taux d'erreur $\tau = 5\%$. (e) Estimation de la réalisation du processus auxiliaire U selon une CMEC. (f) Segmentation de l'image selon les CMC 3-S, taux d'erreur $\tau = 4.9\%$. (g) Estimation de la réalisation du processus auxiliaire U selon une CMC 3-S. (h) Segmentation de l'image selon les CMC 7-S, taux d'erreur $\tau = 4.8\%$. (i) Estimation de la réalisation du processus auxiliaire U selon une CMC 7-S.

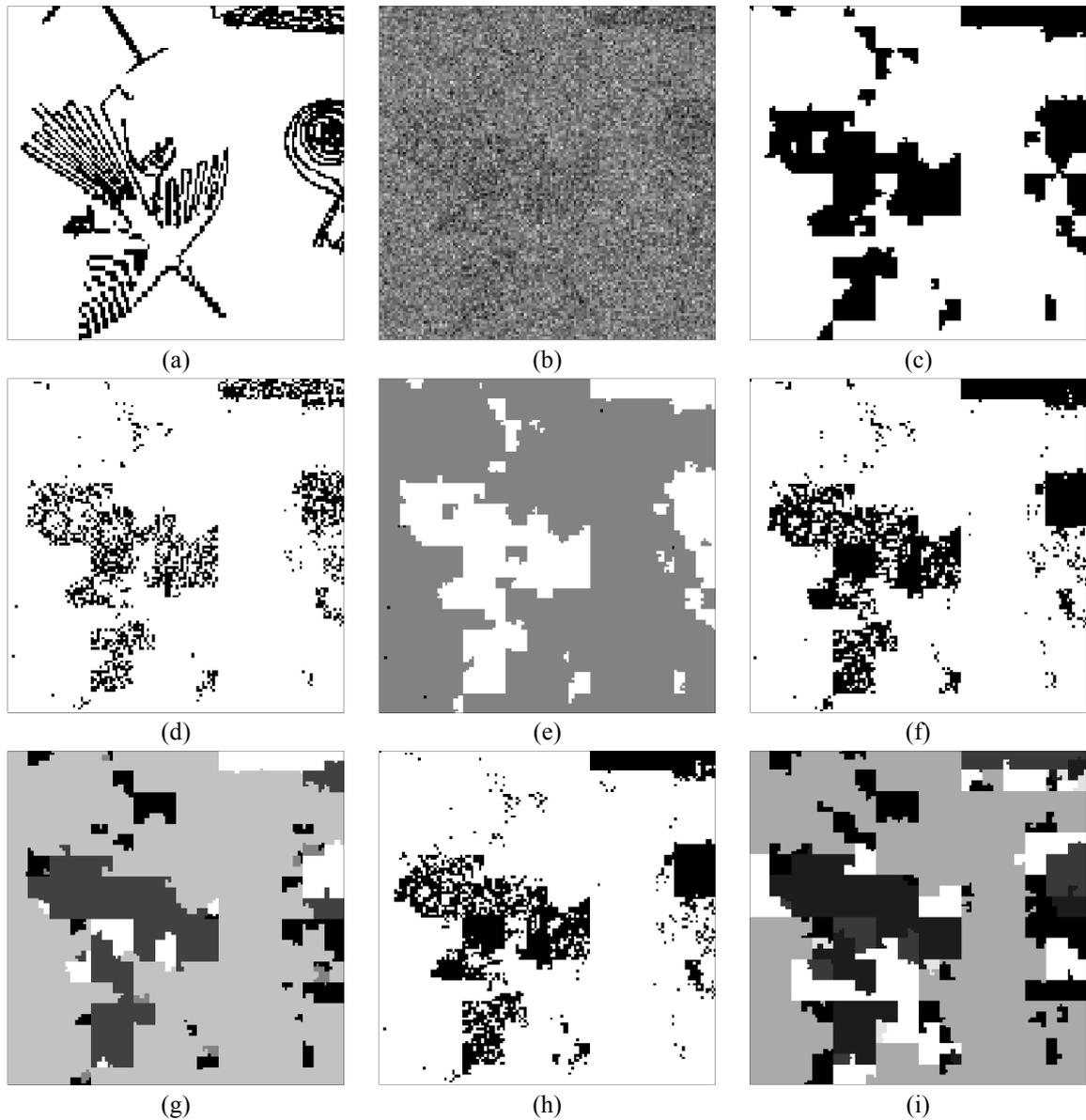


Figure 3.6. Segmentation non-supervisée d'une image non-stationnaire à l'aide des CMEC et les CMC M-S. (a) Image de classes d'origine $X = x$. (b) Image observée $Y = y$. (c) Segmentation de l'image selon les CMC, taux d'erreur $\tau = 15.4\%$. (d) Segmentation de l'image selon les CMEC, taux d'erreur $\tau = 11\%$. (e) Estimation de la réalisation du processus auxiliaire U selon une CMEC. (f) Segmentation de l'image selon les CMC 5-S, taux d'erreur $\tau = 11.9\%$. (g) Estimation de la réalisation du processus auxiliaire U selon une CMC 5-S. (h) Segmentation de l'image selon les CMC 10-S, taux d'erreur $\tau = 11.7\%$. (i) Estimation de la réalisation du processus auxiliaire U selon une CMC 10-S.

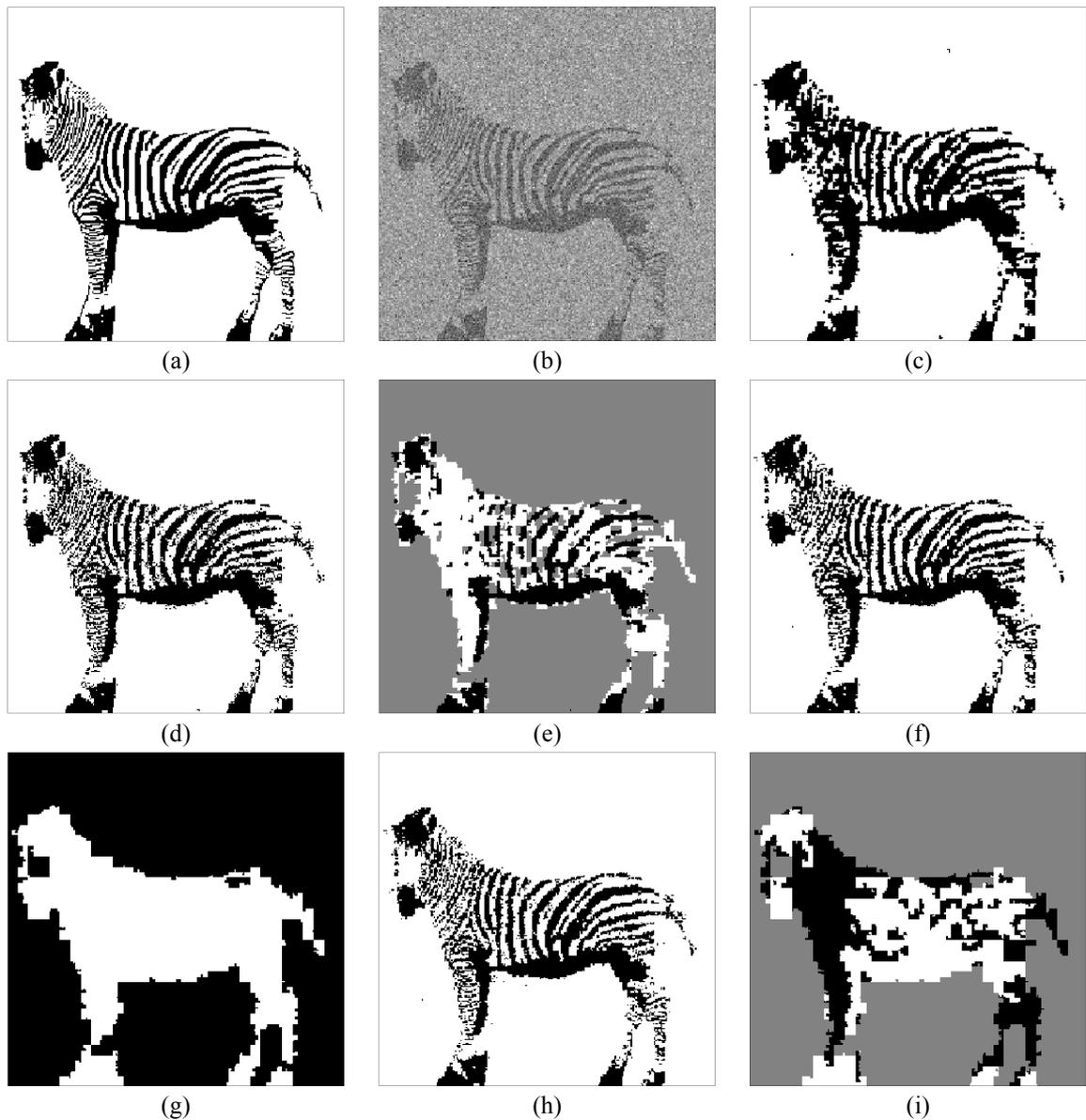


Figure 3.7. Segmentation non-supervisée d'une image non-stationnaire à l'aide des CMEC et les CMC M-S. (a) Image de classes d'origine $X = x$. (b) Image observée $Y = y$. (c) Segmentation de l'image selon les CMC, taux d'erreur $\tau = 6\%$. (d) Segmentation de l'image selon les CMEC, taux d'erreur $\tau = 3.7\%$. (e) Estimation de la réalisation du processus auxiliaire U selon une CMEC. (f) Segmentation de l'image selon les CMC 2-S, taux d'erreur $\tau = 3.7\%$. (g) Estimation de la réalisation du processus auxiliaire U selon une CMC 2-S. (h) Segmentation de l'image selon les CMC 3-S, taux d'erreur $\tau = 3.3\%$. (i) Estimation de la réalisation du processus auxiliaire U selon une CMC 3-S.

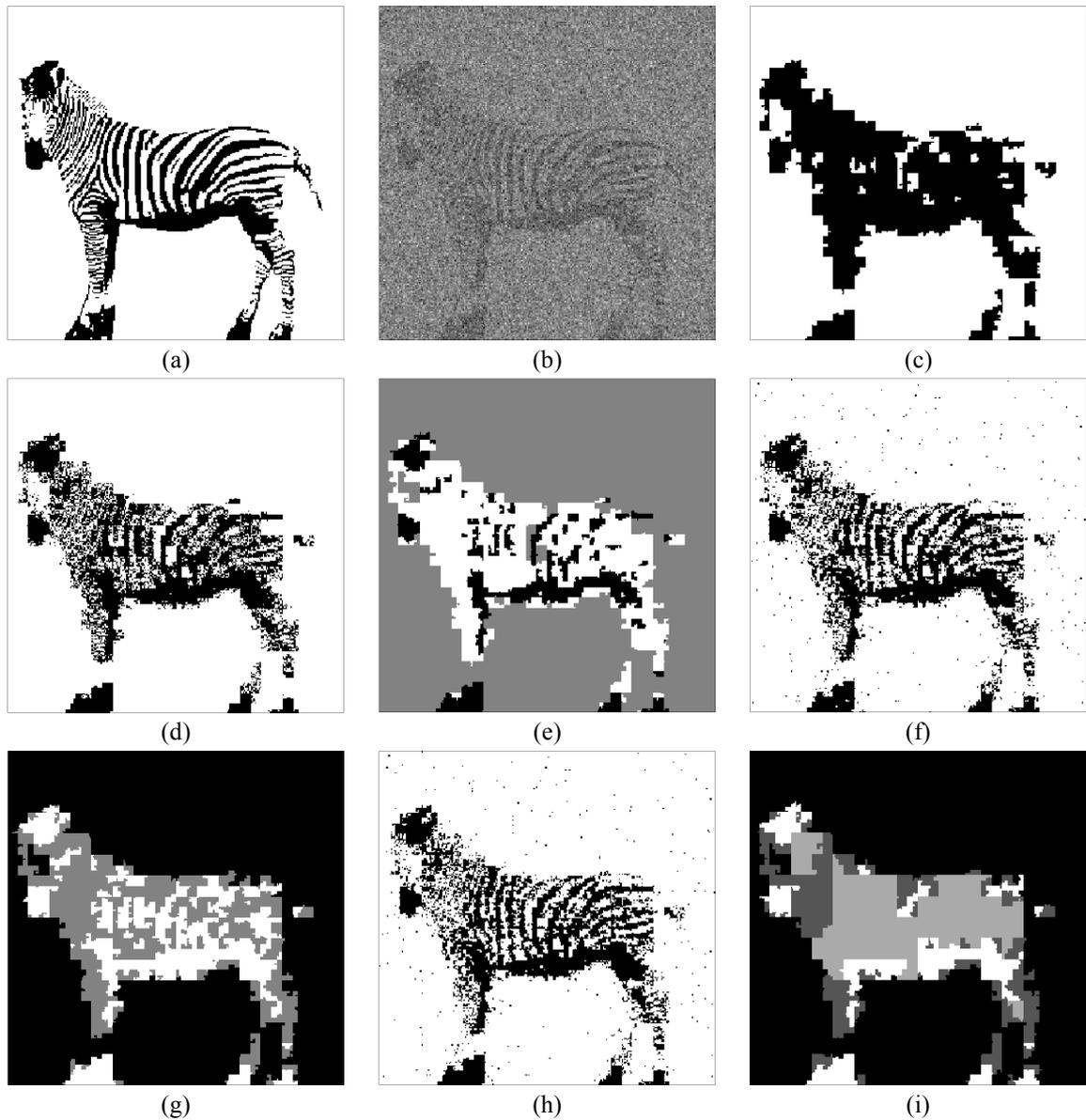


Figure 3.8. Segmentation non-supervisée d'une image non-stationnaire à l'aide des CMEC et les CMC M-S. (a) Image de classes d'origine $X = x$. (b) Image observée $Y = y$. (c) Segmentation de l'image selon les CMC, taux d'erreur $\tau = 12.3\%$. (d) Segmentation de l'image selon les CMEC, taux d'erreur $\tau = 8.2\%$. (e) Estimation de la réalisation du processus auxiliaire U selon une CMEC. (f) Segmentation de l'image selon les CMC 3-S, taux d'erreur $\tau = 7.7\%$. (g) Estimation de la réalisation du processus auxiliaire U selon une CMC 3-S. (h) Segmentation de l'image selon les CMC 4-S, taux d'erreur $\tau = 7.6\%$. (i) Estimation de la réalisation du processus auxiliaire U selon une CMC 4-S.

3.4. Chaînes de Markov cachées multi-senseurs

Dans cette section, nous définissons les chaînes de Markov cachées multi-senseurs classiques (CMC MS) inspirées des champs de Markov multi-senseur définis dans [Ben01]. Pour ce faire, considérons l'exemple suivant de segmentation d'image multi-senseurs.

Soit deux images fournies par deux senseurs indépendants d'une même scène dont nous recherchons la segmentation x à valeurs dans $\Omega = \{\omega_1, \omega_2\}$ où ω_1 et ω_2 correspondent à « forêt » et « eau » respectivement. Soit S_1 et S_2 ces deux senseurs :

- Le senseur S_1 est bayésien, son observation $y_{1..N}^1$ est reliée aux classes via les densités de bruit $p(y_n^1|x_n = \omega_1)$ et $p(y_n^1|x_n = \omega_2)$.
- Le senseur S_1 est évidentiel : l'image $y_{1..N}^2$ présente des nuages. Nous sommes donc en présence de trois classes : « forêt », « eau » et « nuage ». L'observation $y_{1..N}^2$ est reliée à ces classes à travers les densités de bruit $p(y_n^2|x_n = \omega_1)$ et $p(y_n^2|x_n = \omega_2)$ et $p(y_n^2|x_n = \Omega)$.

Le problème donc est d'estimer la réalisation $x_{1..N}$ en utilisant les images fournies par les deux senseurs tout en exploitant l'information contextuelle.

3.4.1. Définition du modèle

Soit $X = X_{1..N}$ un processus caché qui prend ses valeurs à partir d'un ensemble fini de classes $\Omega = \{\omega_1, \dots, \omega_K\}$ et qui doit être estimé à partir d'une famille de processus observables $Y = Y^{1..R}$ fournis par R senseurs indépendants $S_{1..R}$ où $Y^r = Y_{1..N}^r$ avec chaque Y_n^r dans \mathbb{R} . Supposons que la réalisation de X est gouvernée par une chaîne de Markov. Supposons aussi qu'au moins l'un de ces senseurs est bayésien. Soit S_1 un tel senseur. La restauration bayésienne MPM en utilisant toute l'observation $y = y^{1..R}$ est alors faisable et le nombre d'opérations nécessaires pour son calcul est linéaire par rapport à la taille des données N [Pie07]. Le modèle $Z = (X, Y)$ est alors appelé chaîne de Markov cachée multi-senseurs. La restauration MPM s'effectue alors grâce à la fusion des différentes masses impliquées dans le modèle.

$$m(x) = (m_0 \oplus \dots \oplus m_R)(x) = p(x|y) \quad (3.36)$$

Même si le résultat de cette fusion n'est pas nécessairement une chaîne de Markov, elle est une distribution marginale d'une CMT. Par conséquent, le calcul des probabilités *a posteriori* $p(x_n|y)$ est possible.

Revenons maintenant à l'exemple de segmentation considéré plus haut et montrons comment modéliser les données via le modèle CMC MS :

- D'abord, la distribution du processus caché X est modélisée par une masse bayésienne stationnaire m_0 définie sur Ω par $m_0(x) = m_0(x_1) \prod_{n=2}^N m_0(x_n|x_{n-1})$.
- Le senseur S_1 étant bayésien, la masse associée à son observation m_1 est alors définie sur Ω par $m_1(x) \propto \prod_{n=1}^N p(y_n^1|x_n)$.
- Pour le senseur S_2 , la présence des nuages peut être modélisée par une fonction de masse évidentielle définie sur $\mathcal{P}(\Omega) = \{\omega_1, \omega_2, \Omega\}$ par $m_2(u) \propto \prod_{n=1}^N m_2(u_n)$ où u est à

valeurs dans $\mathcal{P}(\Omega)$ et $m_2(\omega_1) \propto p(y_n^2|u_n = \omega_1)$, $m_2(\omega_2) \propto p(y_n^2|u_n = \omega_2)$ et $m_2(\Omega) \propto p(y_n^2|u_n = \Omega)$.

Les masses impliquées dans la fusion étant définies, la probabilité *a posteriori* est alors donnée par

$$m(x) = (m_0 \oplus m_1 \oplus m_2)(x) = p(x|y) \quad (3.37)$$

3.4.2. Restauration MPM d'une chaîne de Markov multi-senseurs

Les masses m_0 , m_1 et m_2 , représentant l'information que nous avons sur les données, étant définie, nous posons $T = (X, U, Y) = (V, Y)$ une CMT où $X = X_{1..N}$ avec chaque $x_n \in \Omega$, $X = X_{1..N}$ avec chaque $u_n \in \mathcal{P}(\Omega)$ et $Y = (Y^1, Y^2)$ avec chaque $y_n \in \mathbb{R}^2$. La distribution de T est alors donnée par

$$p(t_1, \dots, t_N) \propto q_1(t_1, t_2)q_2(t_2, t_3) \dots q_{N-1}(t_{N-1}, t_N) \quad (3.38)$$

avec

$$q_1(t_1, t_2) = \mathbf{1}_{[x_1 \in u_1]} \mathbf{1}_{[x_2 \in u_2]} m_0(x_1) m_0(x_2|x_1) m_1(x_1) m_1(x_2) m_2(u_1) m_2(u_2);$$

$$q_n(t_n, t_{n+1}) = \mathbf{1}_{[x_{n+1} \in u_{n+1}]} m_0(x_{n+1}|x_n) m_1(x_{n+1}) m_2(u_{n+1}).$$

Ainsi, le résultat de la fusion DS $m = (m_0 \oplus m_1 \oplus m_2)$ est la distribution *a posteriori* $p(x|y)$ associée à la loi jointe $p(x, y)$ qui est aussi la distribution marginale de la CMT définie par (3.). Par conséquent, les probabilités marginales $p(x_n, u_n|y)$ sont calculables et ainsi sont les probabilités $p(x_n|y)$ requises pour la restauration MPM.

Enfin, pour effectuer la restauration bayésienne MPM, nous introduisons les fonctions de probabilités progressives $\alpha_n(v_n) = p(y_1, \dots, y_n, v_n)$ et rétrogrades $\beta_n(v_n) = p(y_{n+1}, \dots, y_N|v_n)$ qui ont été adaptées au contexte des CMC MN. Ainsi, ces fonctions sont calculables de la manière itérative suivante

Initialisation :

$$\alpha_1(v_1) = p(v_1, y_1) \quad (3.39)$$

$$\beta_N(v_N) = 1 \quad (3.40)$$

Induction :

$$\alpha_n(v_n) = \sum_{v_{n-1}} \alpha_{n-1}(v_{n-1}) p(v_n|v_{n-1}) m_1(x_n) m_2(u_n) \quad (3.41)$$

$$\beta_n(v_n) = \sum_{v_{n+1}} \beta_{n+1}(v_{n+1}) p(v_{n+1}|v_n) m_1(x_{n+1}) m_2(u_{n+1}) \quad (3.42)$$

avec

$$p(v_{n+1}|v_n) \propto \mathbf{1}_{[x_n \in u_n]} \mathbf{1}_{[x_{n+1} \in u_{n+1}]} m_0(x_{n+1}|x_n);$$

$$p(v_1, y_1) \propto \mathbf{1}_{[x_1 \in u_1]} m_0(x_1) m_1(x_1) m_2(u_1).$$

Les probabilités marginales *a posteriori* $p(v_n|y)$ et $p(x_n|y)$ se calculent comme suit :

$$p(v_n|y) \propto \alpha_n(v_n) \beta_n(v_n) \quad (3.43)$$

$$p(x_n|y) = \sum_{u_n} p(u_n, x_n|y) \quad (3.44)$$

Par ailleurs, les probabilités jointes et marginales *a posteriori* nécessaires pour l'algorithme d'estimation des paramètres sont calculées selon les formules

$$\psi(v_n, v_{n+1}) = p(v_n, v_{n+1}|y) = \alpha_n(v_n) p(v_{n+1}|v_n) m_1(x_{n+1}) m_2(u_{n+1}) \beta_{n+1}(v_{n+1}) \quad (3.45)$$

$$\xi(v_n) = p(v_n|y) = \sum_{v_{n+1}} \psi(v_n, v_{n+1}) \quad (3.46)$$

3.4.3. Estimation des paramètres d'une chaîne de Markov multi-senseurs

Pour estimer les paramètres du modèle, nous proposons d'utiliser une version adaptée de l'algorithme EM. Pour ce faire, considérons la CMT $T = (X, U, Y) = (V, Y)$ définie plus haut. Supposons de plus que les densités de bruit $p(y_n^1|x_n)$ et $p(y_n^2|u_n^2)$ sont gaussiennes.

Conformément au problème de segmentation défini plus haut, nous devons estimer les paramètres suivants : les masses bayésiennes $m_{ij} = m_0(x_{n+1} = \lambda_j | x_n = \lambda_i)$ définie sur Ω^2 , les $K = 2$ moyennes $\mu_{1..K}^1$ et écarts-type $\sigma_{1..K}^1$ des densités gaussiennes régissant m_1 , les $K + 1 = 3$ moyennes $\mu_{1..K+1}^2$ et écarts-type $\sigma_{1..K+1}^2$ des densités gaussiennes régissant m_2 .

La procédure d'estimation se fait de la manière suivante :

- Considérer un jeu initial de paramètres $\theta^0 = (m_{ij}^0, (\mu_{1..K}^1, \mu_{1..K}^1, \sigma_{1..K}^2, \sigma_{1..K}^2)^0)$.
- Pour chaque itération $q + 1$, calculer θ^{q+1} à partir de θ^q en deux étapes jusqu'à ce qu'un critère d'arrêt soit satisfait :
 - o Estimation : Calculer les fonctions ψ_n^q and ξ_n^q à partir de y et θ^q .
 - o Maximisation : Calculer θ^{q+1} selon les formules :

$$(\mu_k^1)^{q+1} = \frac{\sum_{n=1}^N \sum_{u_n} \xi^q(x_n = \omega_k, u_n) y_n^1}{\sum_{n=1}^N \sum_{u_n} \xi^q(x_n = \omega_k, u_n)} \quad (3.47)$$

$$(\mu_k^2)^{q+1} = \frac{\sum_{n=1}^N \sum_{x_n} \xi^q(x_n, u_n = \lambda_k) y_n^2}{\sum_{n=1}^N \sum_{x_n} \xi^q(x_n, u_n = \lambda_k)} \quad (3.48)$$

$$\left[(\sigma_k^1)^{q+1} \right]^2 = \frac{\sum_{n=1}^N \sum_{u_n} \xi^q(x_n = \omega_k, u_n) \left(y_n^1 - (\mu_k^1)^{q+1} \right)^2}{\sum_{n=1}^N \sum_{u_n} \xi^q(x_n = \omega_k, u_n)} \quad (3.49)$$

$$\left[(\sigma_k^2)^{q+1} \right]^2 = \frac{\sum_{n=1}^N \sum_{x_n} \xi^q(x_n, u_n = \lambda_k) \left(y_n^2 - (\mu_k^2)^{q+1} \right)^2}{\sum_{n=1}^N \sum_{x_n} \xi^q(x_n, u_n = \lambda_k)} \quad (3.50)$$

$$m_{ij}^{q+1} = \frac{\sum_{n=1}^N \sum_{x_n, x_{n+1}} \psi^q(u_n = \lambda_i, x_n, u_{n+1} = \lambda_j, x_{n+1})}{\sum_{n=1}^N \sum_{x_n} \xi^q(u_n = \lambda_i, x_n)} \quad (3.51)$$

3.5. Conclusion

Dans ce chapitre, nous avons décrit brièvement la théorie de l'évidence de Dempster-Shafer et avons exposé son importance pour la modélisation de certains types de données telles que celles présentant un manque d'information ou un quelconque type d'imprécision.

Pour montrer l'intérêt de l'emploi de cette théorie dans le contexte markovien, nous avons décrit le modèle de CMEC introduit dans [Lan05] pour la modélisation de données non-stationnaire.

Notre contribution dans ce chapitre porte sur deux volets : d'abord la comparaison entre les modèles CMEC et CMC M-S dans la tâche de la segmentation non-supervisée de données non-stationnaire ; et la description du modèle CMC MS inspiré des champs de Markov caché multi-senseur introduit dans [Ben01] et qui va servir pour des fins de comparaison dans le chapitre 7.

Chapitre 4. Chaînes de Markov cachées à bruit M-stationnaire

Les CMC ont montré des limites pour le traitement de certaines données complexes telles que les données non-stationnaires. Les CMT ont permis de surmonter ces limites grâce à leur théorie très riche qui offre plus de possibilités de modélisation tout en maintenant la complexité des calculs raisonnable. En particulier, le modèle de chaîne de Markov M-stationnaire (ou M-non stationnaire [Lan06, Lan12]) permet de considérer la variabilité de la loi *a priori* du processus caché en la supposant osciller entre M régimes différents. Comme vu dans le chapitre 2, un tel modèle peut trouver ses applications en segmentation d'images texturées où chaque texture correspond à un régime différent.

Dans le présent chapitre, nous considérons la situation où ce changement de régimes concerne la loi de vraisemblance (ou les distributions de bruit). En effet, dans les CMC classiques, nous supposons que les symboles émis par un même état sont régis par la même distribution. Toutefois, il arrive qu'un phénomène extérieur influe sur cette distribution. En particulier, dans une image, les conditions d'éclairages ont tendance à donner des aspects différents aux sites appartenant à des états identiques. En finance, les revenus se comportent différemment en temps de crises. Pour cela, nous introduisons un processus auxiliaire sous-jacent conformément au formalisme des CMT pour prendre en compte la variabilité de la loi de vraisemblance. Le processus triplet est supposé donc de Markov et le modèle correspondant est appelé chaîne de Markov cachée à bruit M-stationnaire (CMC BM-S). Nous montrons que l'estimation des distributions *a posteriori* reste possible. Nous décrivons également les différents algorithmes d'estimation de paramètres et de restauration. Pour valider ce modèle, les expériences portent sur des données synthétiques et des images réelles. Les résultats obtenus démontrent l'intérêt du modèle CMC BM-S par rapport aux CMC conventionnels.

Par ailleurs, il est possible de considérer à la fois la variabilité de la loi *a priori* et des lois de bruit grâce à un processus auxiliaire composé conformément à la théorie des CMT. De ce fait, nous définissons le modèle de chaîne de Markov cachée ML-stationnaire (où M et L correspondent aux nombres des régimes de la loi *a priori* et des lois de bruit respectivement). Les différentes procédures d'inférence sont également décrites pour ce modèle et les expériences de validation sont menées sur des images synthétiques et réelles.

4.1. Définition du modèle

Soit $X = (X_n)_{n=1}^N$ un processus caché qui prend ses valeurs à partir d'un ensemble fini de classes $\Omega = \{\omega_1, \dots, \omega_K\}$ et qui doit être estimé à partir d'un processus observable $Y = (Y_n)_{n=1}^N$ où $Y_n \in \mathbb{R}$. Le processus couple $Z = (X, Y)$ est dit chaîne de Markov cachée à bruit M-stationnaire s'il existe un processus auxiliaire $U = (U_n)_{n=1}^N$ prenant ses valeurs dans l'ensemble des classes auxiliaires $\Lambda = \{\lambda_1, \dots, \lambda_M\}$ tel que le processus triplet $T = (U, X, Y)$ est une CMT stationnaire dont la distribution est donnée par :

$$p(t) = p(t_1) \prod_{n=2}^N p(t_n | t_{n-1}) \quad (4.1)$$

où $p(t_1) = p(u_1)p(x_1)p(y_1|u_1, x_1)$ et $p(t_n|t_{n-1}) = p(u_n|u_{n-1})p(x_n|x_{n-1})p(y_n|u_n, x_n)$.

L'estimation MPM du processus *a posteriori* est donc possible et sa complexité est linéaire par rapport à la taille de la donnée N [Bou12c].

Il convient de préciser la différence entre une CMC BM-S et une CMC à bruit non-stationnaire classique. Dans une CMC à bruit non-stationnaire, la loi de vraisemblance $p(y_n|x_n)$ dépend de n . Dans une CMC BM-S par contre, $p(y_n|x_n)$ ne dépend pas de n ; toutefois, étant donné une réalisation u du processus auxiliaire U , la loi $p(z|u)$ peut être vue comme une CMC à bruit non-stationnaire classique particulière (le nombre de distributions par état M étant fixe et connu à l'avance).

4.2. Simulation d'une CMC BM-S

Dans ce chapitre, nous limitons notre champ d'étude aux CMC BM-S gaussiennes. La distribution de bruit $p(y_n|u_n, x_n)$ est donc une densité gaussienne. Par conséquent, nous définissons les paramètres d'une CMC BM-S comme suit :

- Les distributions initiales des processus latents U et X sont données par ϖ et π respectivement. Ainsi, $p(u_1 = \lambda_m) = \varpi_m$ et $p(x_1 = \omega_k) = \pi_k$;
- Les distributions de transition $p(u_n|u_{n-1})$ et $p(x_n|x_{n-1})$ sont caractérisées par les matrices Q et A respectivement. Ainsi, $p(u_n = \lambda_{m'}|u_{n-1} = \lambda_m) = q_{m,m'}$ et $p(x_n = \omega_{k'}|x_{n-1} = \omega_k) = a_{k,k'}$ respectivement ;
- Les distributions de bruit $p(y_n|u_n = \lambda_m, x_n = \omega_k)$ étant gaussiennes, elles sont donc paramétrées par les moyennes $\mu_{k,m}$ et les variances $\sigma_{k,m}^2$. Comme nous avons K classes cachées et M classes auxiliaires, notre modèle comporte donc $M \times K$ densités gaussiennes et par conséquent, $M \times K$ moyennes et $M \times K$ variances.

Les paramètres θ du modèle CMC BM-S étant définis, nous montrons maintenant comment simuler une réalisation d'un tel modèle :

1. Simuler la réalisation du processus auxiliaire U selon le formalisme d'une chaîne de Markov classique par tirages à partir de ϖ et Q respectivement ;
2. Simuler la réalisation du processus caché X selon le formalisme d'une chaîne de Markov classique par tirages à partir de π et A respectivement ;
3. Simuler la réalisation du processus observable Y par tirage aléatoire à partir des densités gaussiennes correspondantes aux réalisations de U et X .

4.3. Restauration MPM d'une CMC BM-S

Soit une réalisation $Y = y$ et supposons que les paramètres $\theta = (\varpi, \pi, Q, A, \mu, \sigma)$ du modèle sont connus. Nous montrons dans cette section comment estimer les réalisations des processus caché $X = \hat{x}$ et auxiliaire $U = \hat{u}$ à partir de $Y = y$.

L'estimateur MPM est donné par la formule

$$[\hat{x} = \hat{s}_{MPM}(y)] \Leftrightarrow \left[\hat{x}_n = \underset{\omega \in \Omega}{\operatorname{argmax}} p(x_n = \omega | y) \right] \quad (4.2)$$

Considérons les fonctions progressives $\alpha_n(x_n, u_n) = p(y_1, \dots, y_n, x_n, u_n)$ et rétrogrades $\beta_n(x_n) = p(y_{n+1}, \dots, y_N | x_n, u_n)$ qui peuvent être calculées itérativement comme suit :

Initialisation :

$$\alpha_1(x_1, u_1) = p(u_1)p(x_1)p(y_1|u_1, x_1) \quad (4.3)$$

$$\beta_N(x_N, u_N) = 1 \quad (4.4)$$

Induction :

$$\alpha_n(x_n, u_n) = \sum_{x_{n-1}, u_{n-1}} \alpha_{n-1}(x_{n-1}, u_{n-1})p(u_n|u_{n-1})p(x_n|x_{n-1})p(y_n|u_n, x_n) \quad (4.5)$$

$$\beta_n(x_n, u_n) = \sum_{x_{n+1}, u_{n+1}} \beta_{n+1}(x_{n+1}, u_{n+1})p(u_{n+1}|u_n)p(x_{n+1}|x_n)p(y_{n+1}|u_{n+1}, x_{n+1}) \quad (4.6)$$

Les distributions *a posteriori* $p(x_n, u_n | y)$ peuvent alors être calculées selon les formules

$$p(x_n, u_n | y) \propto \alpha_n(x_n, u_n) \beta_n(x_n, u_n) \quad (4.7)$$

Les distributions marginales *a posteriori* des processus caché et auxiliaire sont respectivement dérivées comme suit

$$p(x_n | y) = \sum_{u_n} p(x_n, u_n | y) \quad (4.8)$$

$$p(u_n | y) = \sum_{x_n} p(x_n, u_n | y) \quad (4.9)$$

4.4. Estimation des paramètres d'une CMC BM-S

Dans le cas où les paramètres du modèle sont inconnus, il est possible de les estimer en utilisant par exemple une méthode inspirée de l'algorithme EM. Pour ce faire, nous définissons les distributions *a posteriori* suivantes

$$\begin{aligned} \psi_n(x_n, u_n, x_{n+1}, u_{n+1}) &= p(x_n, u_n, x_{n+1}, u_{n+1} | y) \\ &\propto \alpha_n(x_n, u_n)p(u_{n+1}|u_n)p(x_{n+1}|x_n)p(y_{n+1}|x_{n+1})\beta_{n+1}(x_{n+1}, u_{n+1}) \end{aligned} \quad (4.10)$$

$$\xi_n(x_n, u_n) = p(x_n, u_n | y) \quad (4.11)$$

La procédure d'estimation est exécutée de la manière itérative suivante :

- Considérer un jeu initial de paramètres $\theta^0 = (\varpi^0, \pi^0, Q^0, A^0, \mu^0, \sigma^0)$;
- Pour chaque itération $i+1$:
 - o Estimation : Calculer les fonctions ψ_n^i and ξ_n^i à partir de y et θ^i .
 - o Maximisation : Estimer le nouveau jeu de paramètres θ^{i+1} comme suit

$$\varpi_m^{i+1} = \sum_{x_n} \xi_1^i(x_n, u_n = \lambda_m) \quad (4.12)$$

$$\pi_k^{i+1} = \sum_{u_n} \xi_1^i(x_n = \omega_k, u_n) \quad (4.13)$$

$$q_{m,m'}^{i+1} = \frac{\sum_{n=1}^{N-1} \sum_{x_n, x_{n+1}} \psi_n^i(x_n, u_n = \lambda_m, x_{n+1}, u_{n+1} = \lambda_{m'})}{\sum_{n=1}^{N-1} \sum_{x_n, u_n = \lambda_m} \xi_n^i(x_n, u_n)} \quad (4.14)$$

$$a_{k,k'}^{i+1} = \frac{\sum_{n=1}^{N-1} \sum_{u_n, u_{n+1}} \psi_n^i(x_n = \omega_k, u_n, x_{n+1} = \omega_{k'}, u_{n+1})}{\sum_{n=1}^{N-1} \sum_{u_n, x_n = \omega_k} \xi_n^i(x_n, u_n)} \quad (4.15)$$

$$\mu_{k,m}^{i+1} = \frac{\sum_{n=1}^N \xi_n^i(x_n = \omega_k, u_n = \lambda_m) y_n}{\sum_{n=1}^N \xi_n^i(x_n = \omega_k, u_n = \lambda_m)} \quad (4.16)$$

$$\sigma_{k,m}^{i+1} = \frac{\sum_{n=1}^N \xi_n^i(x_n = \omega_k, u_n = \lambda_m) (y_n - \mu_{k,m}^{i+1})^2}{\sum_{n=1}^N \xi_n^i(x_n = \omega_k, u_n = \lambda_m)} \quad (4.17)$$

4.5. Résultats expérimentaux

Pour valider le modèle de CMC BM-S, nous proposons trois séries d'expériences. Dans la première série, nous considérons des données synthétiques simulées selon les formalismes des CMC et CMC BM-S respectivement. Dans les deux dernières séries, nous abordons le problème de la segmentation d'images bruitées et réelles. Pour rendre nos modèles monodimensionnels applicables aux données 2D, le parcours d'Hilbert-Peano [Pea1890, Hil1891] est adopté. Pour les trois séries, les paramètres seront initialisés via l'algorithme K-means, ce qui permet généralement d'obtenir un jeu initial acceptable de paramètres. La phase d'initialisation des paramètres est cruciale dans ce contexte et une mauvaise initialisation peut conduire à une mauvaise estimation de paramètres. La difficulté relative à cette initialisation existe pour tous les modèles de Markov, elle est plus prononcée quand le jeu de paramètres en question est grand.

4.5.1. Segmentation non-supervisée de chaînes de Markov cachées à bruit M-stationnaires

L'objectif de cette première d'expérience est de comparer les CMC et les CMC BM-S lorsque des données simulées selon l'un de ces deux modèles sont concernées. Pour ce faire, nous considérons deux jeux de données pour lesquels 100 chaînes de longueur $N = 15000$ sont simulées:

- Jeu A : Dans cette série, les données sont simulées selon le formalisme des CMC. Soit $Z = (X, Y)$ une CMC avec $Z = (Z_n)_{n=1}^N$ où $N = 15000$ et X_n prend ses valeurs dans de l'ensemble $\Omega = \{\omega_1, \omega_2\}$ et Y_n prend ses valeurs dans de l'ensemble des réels \mathbb{R} .
 - La première réalisation du processus caché x_1 est simulée via un tirage aléatoire uniforme sur l'ensemble des classes Ω .
 - Les réalisations suivantes du processus caché sont simulées via un tirage aléatoire selon la matrice de transition $A = \begin{pmatrix} 0.98 & 0.02 \\ 0.02 & 0.98 \end{pmatrix}$.
 - Les réalisations du processus observable sont simulées par des tirages aléatoires à partir des probabilités $p(y_n|\omega_1)$ et $p(y_n|\omega_2)$ données par les

densités gaussiennes de moyennes 0 et 1 respectivement et d'écart-types égaux à 1.

- Jeu B : Dans cette série, les données sont simulées selon le formalisme des CMC BM-S. Soit $T = (U, X, Y)$ une CMT avec $T = (T_n)_{n=1}^N$ où $N = 15000$ et U_n prend ses valeurs à partir de $\Lambda = \{\lambda_1, \lambda_2\}$, X_n prend ses valeurs à partir de $\Omega = \{\omega_1, \omega_2\}$ et Y_n prend ses valeurs à partir de l'ensemble des réels \mathbb{R} .
 - La première réalisation du processus caché x_1 est simulée via un tirage aléatoire uniforme sur l'ensemble des classes cachées Ω et les réalisations suivantes x_n sont simulées via un tirage aléatoire selon la matrice de transition $A = \begin{pmatrix} 0.98 & 0.02 \\ 0.02 & 0.98 \end{pmatrix}$.
 - La première réalisation du processus auxiliaire u_1 est simulée via un tirage aléatoire uniforme sur l'ensemble des classes auxiliaires Λ et les réalisations suivantes u_n sont simulées via un tirage aléatoire selon la distribution de matrice de transition $Q = \begin{pmatrix} 0.99 & 0.01 \\ 0.01 & 0.99 \end{pmatrix}$.
 - Les réalisations du processus observable sont simulées par des tirages aléatoires à partir des probabilités $p(y_n|\omega_1, \lambda_1)$, $p(y_n|\omega_1, \lambda_2)$, $p(y_n|\omega_2, \lambda_1)$ et $p(y_n|\omega_2, \lambda_2)$ données par les densités gaussiennes de moyennes 0, 1, 2 et 3 respectivement et d'écart-types égaux à 1.

Pour les deux jeux de données, la restauration est effectuée suivant trois méthodes : K-means, CMC ($K = 2$) et CMC BM-S ($K = 2$ et $M = 2$). Les taux moyens d'erreur sont présentés dans TAB. 4.1.

Tableau 4.1. Taux d'erreur moyens de segmentation des données synthétiques (%)

Jeu	$\tau_{K\text{-means}}(\hat{x})$	$\tau_{CMC}(\hat{x})$	$\tau_{CMC\text{ BM-S}}(\hat{x})$	$\tau_{CMC\text{ BM-S}}(\hat{u})$
1	30.9	5.4	5.4	-
2	18.8	4.2	2.0	4.2

Les résultats obtenus confirment la suprématie des CMC BM-S par rapport aux CMC. En effet, lorsque les données sont régies par une CMC conventionnelle, les taux d'erreurs obtenus par les deux modèles sont comparables. En revanche, appliqué aux données simulées par le modèle de CMC BM-S, ce dernier modèle donne des taux d'erreur nettement meilleurs que ceux fournis par le modèle de CMC classique. Ceci s'explique par le fait que la CMC BM-S modélise les fluctuations des distributions de bruit, contrairement à la CMC classique. Nous pouvons donc conclure que la CMC BM-S généralise effectivement la CMC.

4.5.2. Segmentation non-supervisée d'images entachées de bruit M-stationnaire

Dans cette série d'expériences, nous considérons l'image de classes 'zèbre' de taille 256×256 (Fig. 4.1-a) que nous bruyons grâce à un modèle de bruit non-stationnaire. L'objectif par la suite est de comparer les résultats de restauration de cette image selon les modèles de CMC classiques et les CMC BM-S proposés dans ce chapitre. La réalisation du processus caché $X = x$, qui n'est pas nécessairement de Markov, est connue à la donnée de l'image de classes. Soit $\Omega = \{\omega_1, \omega_2\}$ où ω_1 et ω_2 correspondent aux classes des pixels noirs et blancs

respectivement. S'agissant du processus auxiliaire U , modélisant les différentes stationnarités de bruit et où chaque U_n prend ses valeurs dans l'ensemble $\Lambda = \{\lambda_1, \lambda_2\}$, nous considérons des réalisations prédéfinies de la manière suivante : l'image est subdivisée en blocs de même taille s (Fig. 4.1-b), et ces derniers sont assignés alternativement à λ_1 (noir) et λ_2 (blanc). L'image bruitée $Y = y$ est ensuite obtenue en considérant les mêmes densités de bruit que celles utilisées dans l'expérience précédente. La même expérience est menée pour plusieurs valeurs de taille de bloc s . Enfin, l'image bruitée est segmentée selon les formalismes de K -means, CMC et CMC BM-S. Les taux d'erreur moyens obtenus sur 100 expériences sont récapitulés dans TAB. 4.II. Les résultats de restauration MPM pour $s = 64 \times 64$ sont également schématisés dans Fig. 4.1.

Tableau 4.2. Taux d'erreur de segmentation des images bruitées (%)

s	$\tau_{K\text{-means}}(\hat{x})$	$\tau_{CMC}(\hat{x})$	$\tau_{CMC\text{ BM-S}}(\hat{x})$	$\tau_{CMC\text{ BM-S}}(\hat{u})$
2×2	18.9	10.3	10.3	48.3
4×4	19.1	13.9	13.9	29.3
8×8	19.1	15.6	11.7	13.3
16×16	19.1	16.0	9.1	4.1
32×32	19.0	16.3	8.9	1.0
64×64	18.9	15.8	8.5	0.1
128×128	18.9	15.8	8.5	0.0

La lecture des résultats présentés dans TAB. 4.II permet de dégager les conclusions suivantes :

- Quand l'image est subdivisée en blocs de petite taille, les performances des CMC et CMC BM-S sont comparables. En effet, la forte fluctuation des stationnarités de bruit ne permet pas au modèle de CMC BM-S de distinguer entre les blocs sombres (λ_1) et les blocs clairs (λ_2). Par conséquent, la restauration MPM du processus caché ne bénéficie pas de la prise en compte de la variabilité des densités de bruit (le taux d'erreur de l'estimation du processus auxiliaire est proche de 50% pour $s = 2 \times 2$) et le taux d'erreur obtenu pour \hat{x} est équivalent à celui fourni par le modèle de CMC classique (le même taux d'erreur est observé pour $s = 2 \times 2$ et $s = 4 \times 4$).

- Plus la taille s des blocs est grande, plus l'estimation de la réalisation du processus auxiliaire est bonne (un taux d'erreur de 0% est même enregistré pour $s = 128 \times 128$) et par conséquent la restauration MPM de l'image est meilleure et l'écart entre les deux modèles est grand (le gain en taux d'erreur avoisine les 50% pour les grandes valeurs de s).

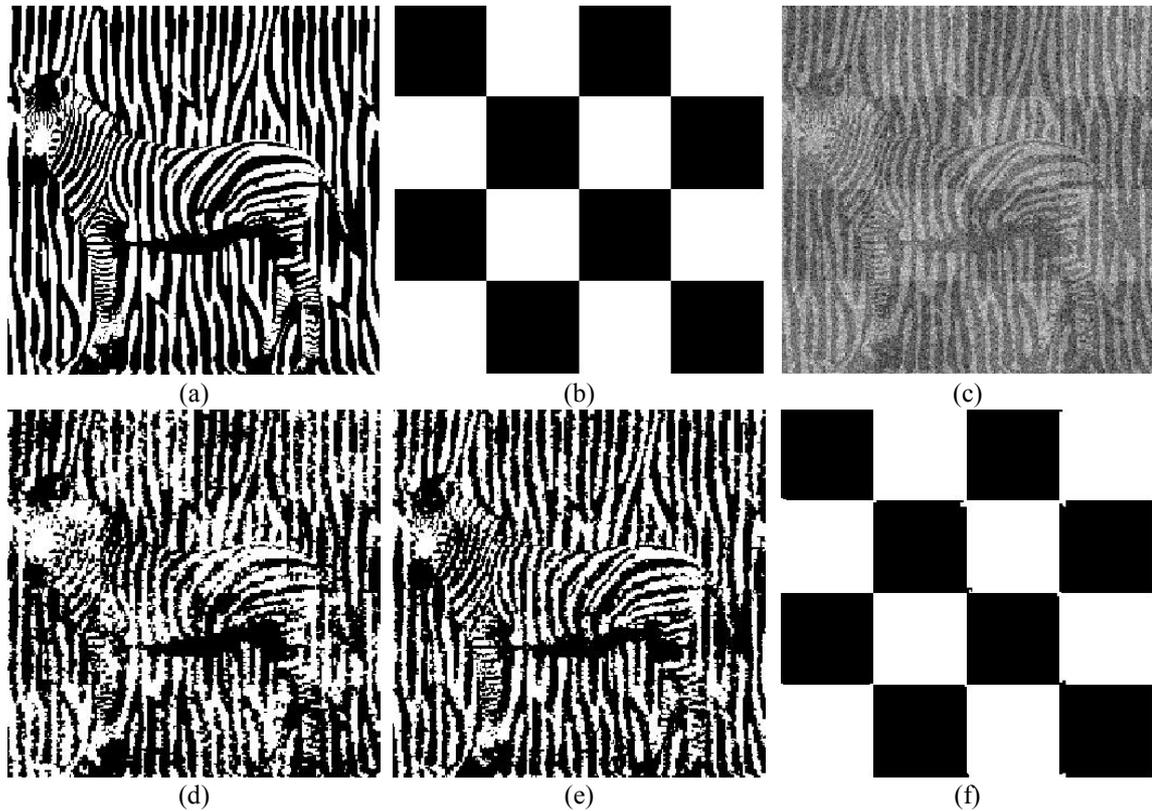


Figure 4.1. Restauration MPM de l'image de zèbre selon les modèles CMC et CMC BM-S. (a) Image de classe $X = x$. (b) $U = u$ correspondant à $s = 64 \times 64$. (c) Image bruitée $Y = y$. (d) Image restaurée selon le modèle de CMC, taux d'erreur $\tau = 15.8\%$. (e) Image restaurée selon le modèle de CMC BM-S, taux d'erreur $\tau = 8.5\%$. (f) Restauration du processus auxiliaire U , taux d'erreur $\tau = 0.1\%$.

Le parcours d'Hilbert-Peano semble particulièrement adapté à de telles formes de réalisations en blocs du processus auxiliaire $U = u$. Pour cela, Les mêmes simulations ont été réalisées sur la base de réalisations quelconques du processus auxiliaire. Les résultats de restauration sont illustrés par les figures Fig. 4.2, Fig. 4.3, ..., Fig. 4.7 où, les taux moyens d'erreur enregistrés sur 100 expériences sont indiqués.

Les résultats obtenus confirment la prépondérance du modèle proposé par rapport aux CMC classiques. En effet le gain en taux d'erreur par rapport au modèle conventionnel se situe entre 25% et 40%. Il convient de remarquer que le modèle proposé permet une bonne estimation de la réalisation du processus auxiliaire $U = u$ (les taux d'erreurs se situent entre 3% et 15%).

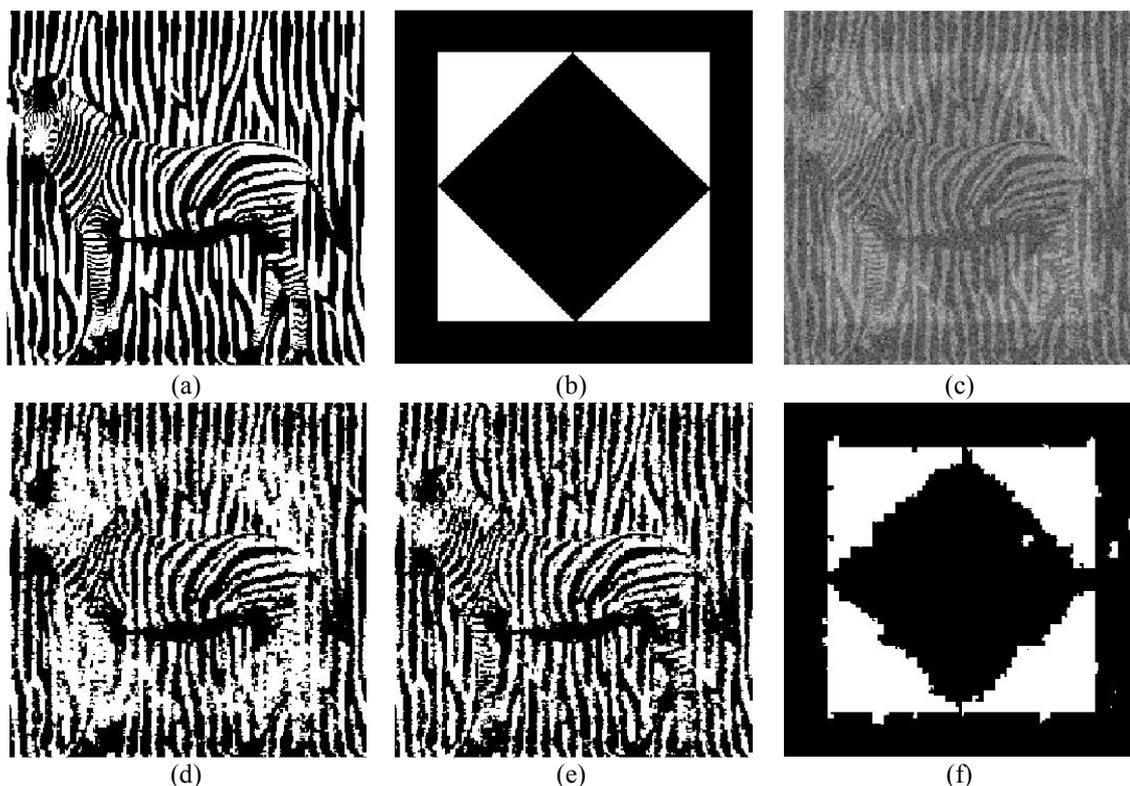


Figure 4.2. Restauration non-supervisée de l'image du zèbre selon les modèles CMC et CMC BM-S. (a) Image originale $X = x$. (b) Réalisation du processus auxiliaire $U = u$. (c) Image bruitée $Y = y$. (d) Image restaurée selon le modèle de CMC, taux d'erreur $\tau = 15.3\%$. (e) Image restaurée selon le modèle de CMC BM-S, taux d'erreur $\tau = 9.7\%$. (f) Restauration du processus auxiliaire U , taux d'erreur $\tau = 3.6\%$.

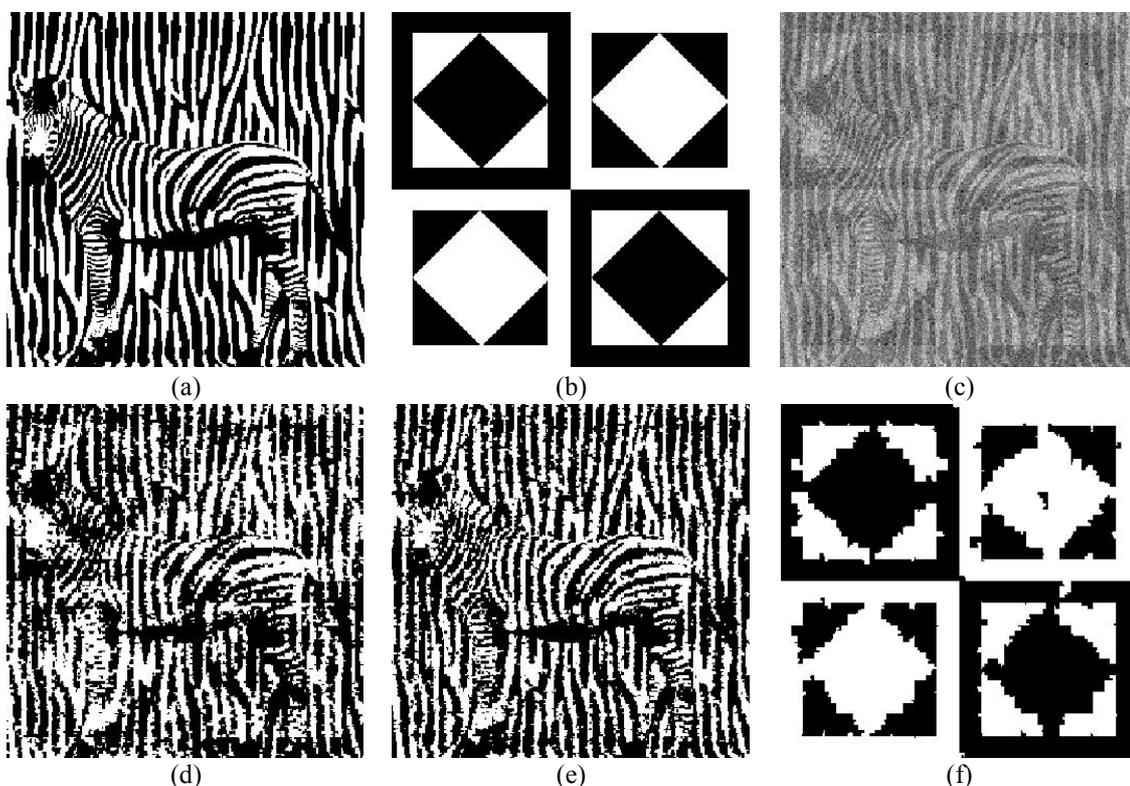


Figure 4.3. Restauration non-supervisée de l'image du zèbre selon les modèles CMC et CMC BM-S. (a) Image originale $X = x$. (b) Réalisation du processus auxiliaire $U = u$. (c) Image bruitée $Y = y$. (d) Image restaurée selon le modèle de CMC, taux d'erreur $\tau = 15.8\%$. (e) Image restaurée selon le modèle de CMC BM-S, taux d'erreur $\tau = 9.9\%$. (f) Restauration du processus auxiliaire U , taux d'erreur $\tau = 5.5\%$.

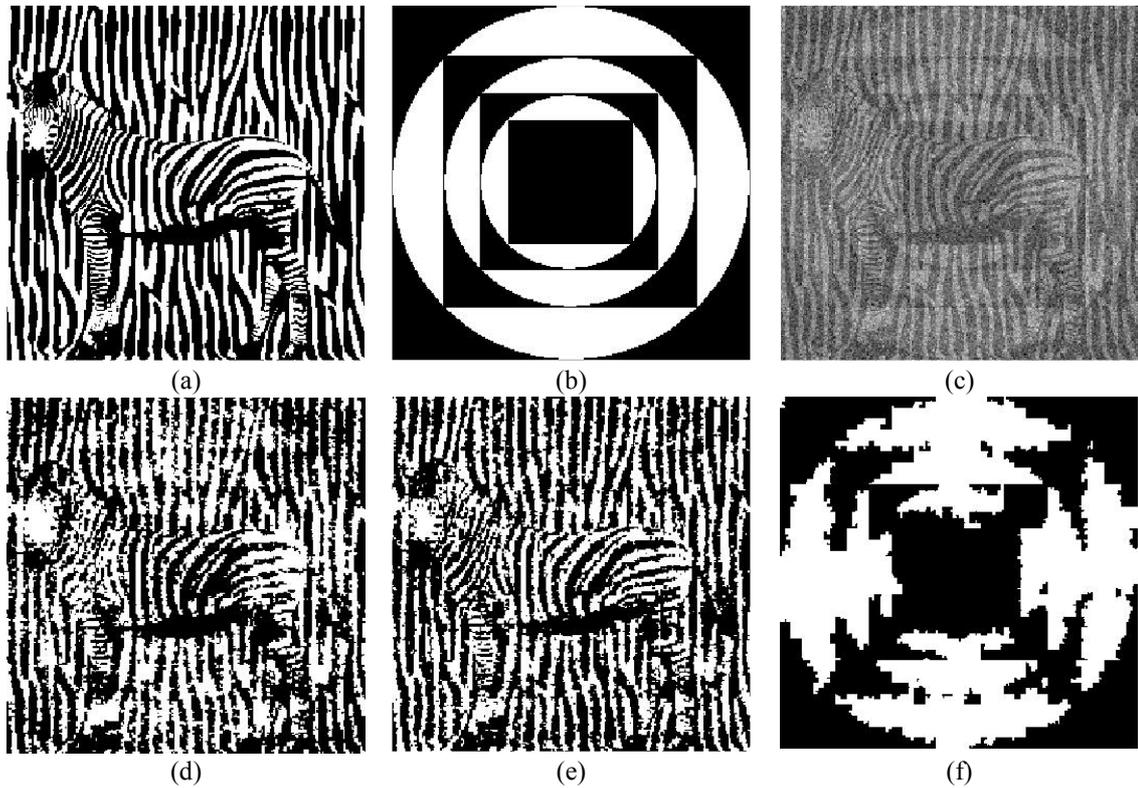


Figure 4.4. Restauration non-supervisée de l'image du zèbre selon les modèles CMC et CMC BM-S. (a) Image originale $X = x$. (b) Réalisation du processus auxiliaire $U = u$. (c) Image bruitée $Y = y$. (d) Image restaurée selon le modèle de CMC, taux d'erreur $\tau = 15.5\%$. (e) Image restaurée selon le modèle de CMC BM-S, taux d'erreur $\tau = 10.5\%$. (f) Restauration du processus auxiliaire U , taux d'erreur $\tau = 9.2\%$.

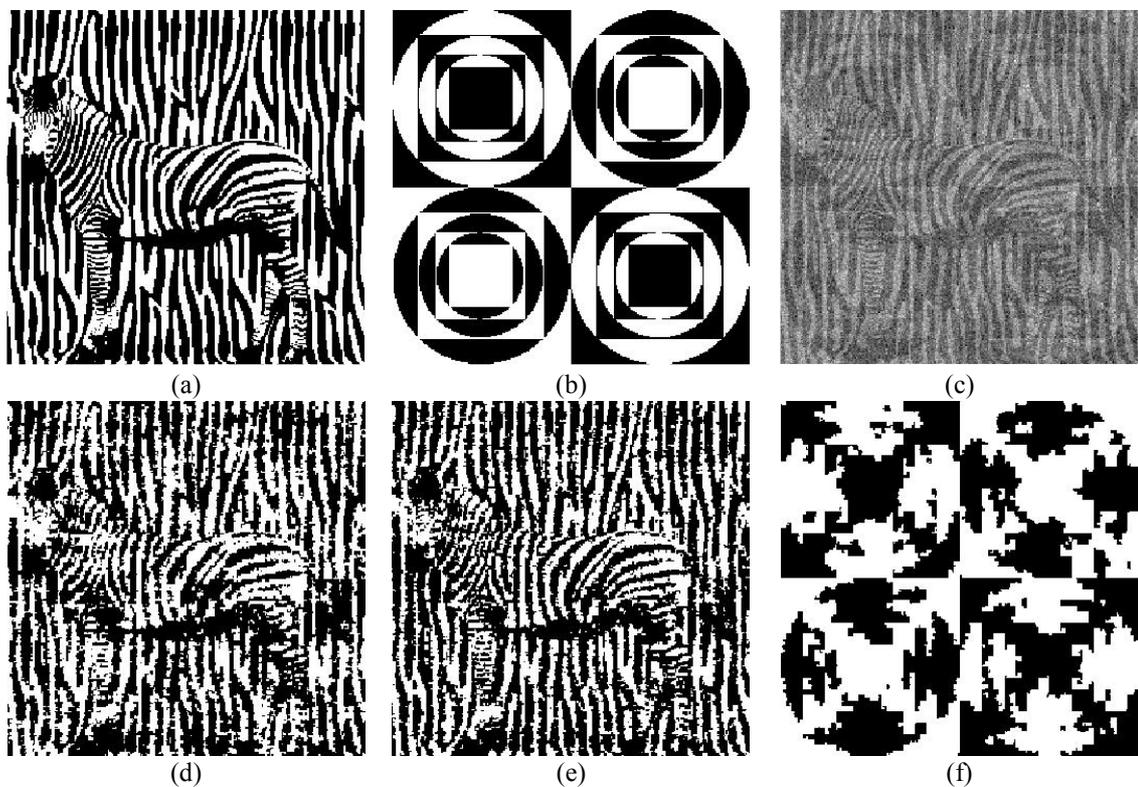


Figure 4.5. Restauration non-supervisée de l'image du zèbre selon les modèles CMC et CMC BM-S. (a) Image originale $X = x$. (b) Réalisation du processus auxiliaire $U = u$. (c) Image bruitée $Y = y$. (d) Image restaurée selon le modèle de CMC, taux d'erreur $\tau = 14.8\%$. (e) Image restaurée selon le modèle de CMC BM-S, taux d'erreur $\tau = 11.2\%$. (f) Restauration du processus auxiliaire U , taux d'erreur $\tau = 14.8\%$.

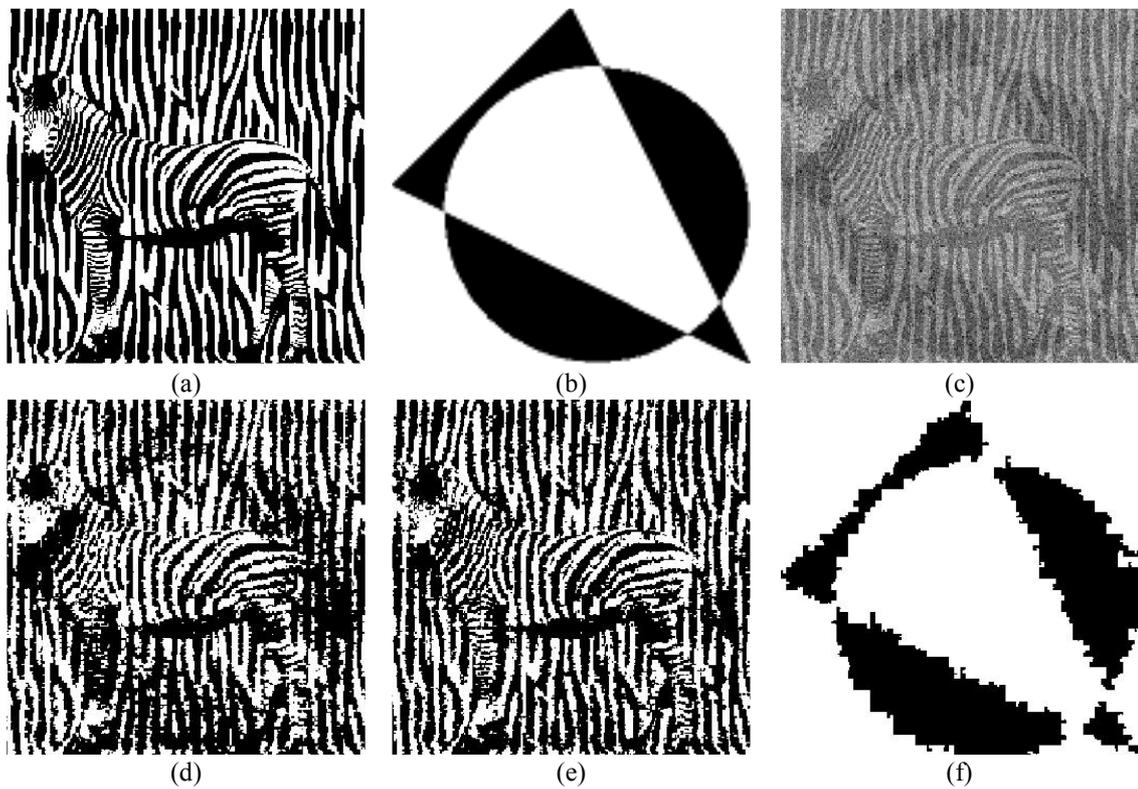


Figure 4.6. Restauration non-supervisée de l'image du zèbre selon les modèles CMC et CMC BM-S. (a) Image originale $X = x$. (b) Réalisation du processus auxiliaire $U = u$. (c) Image bruitée $Y = y$. (d) Image restaurée selon le modèle de CMC, taux d'erreur $\tau = 13.7\%$. (e) Image restaurée selon le modèle de CMC BM-S, taux d'erreur $\tau = 9.1\%$. (f) Restauration du processus auxiliaire U , taux d'erreur $\tau = 4.7\%$.

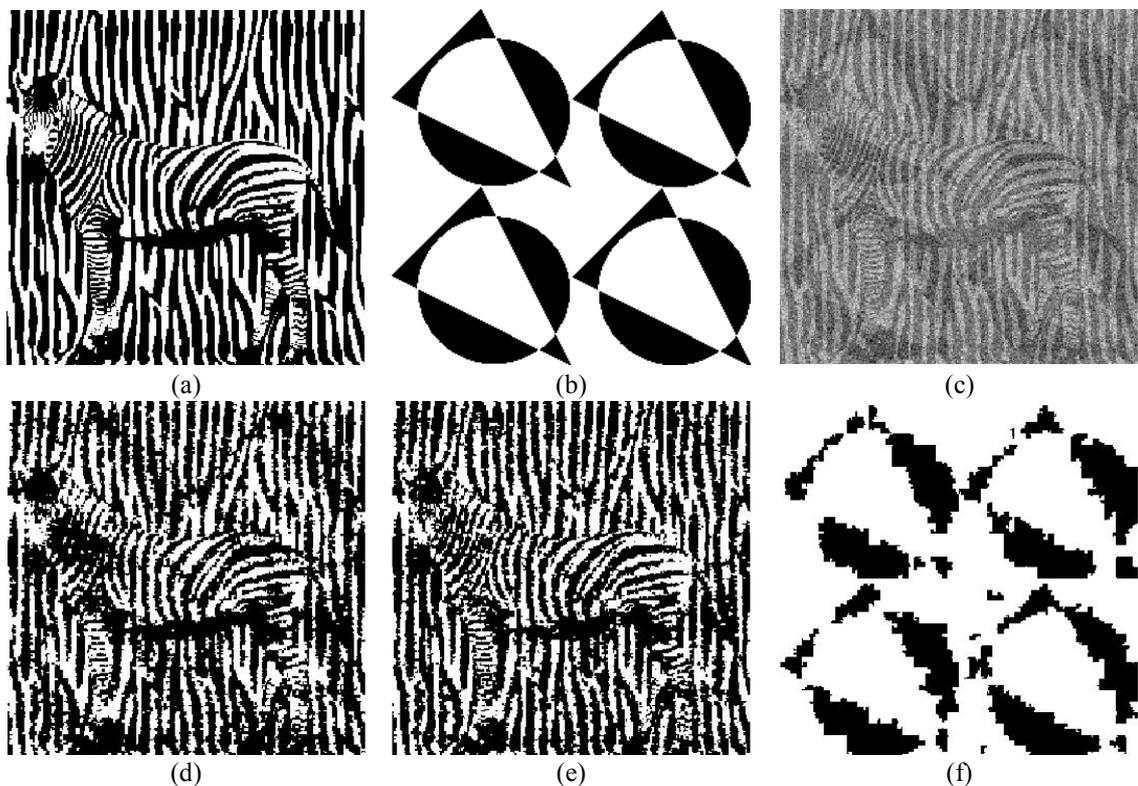


Figure 4.7. Restauration non-supervisée de l'image du zèbre selon les modèles CMC et CMC BM-S. (a) Image originale $X = x$. (b) Réalisation du processus auxiliaire $U = u$. (c) Image bruitée $Y = y$. (d) Image restaurée selon le modèle de CMC, taux d'erreur $\tau = 13.8\%$. (e) Image restaurée selon le modèle de CMC BM-S, taux d'erreur $\tau = 10.4\%$. (f) Restauration du processus auxiliaire U , taux d'erreur $\tau = 8.4\%$.

4.5.3. Segmentation non-supervisée d'image réelles via les CMC BM-S

Dans cette série d'expériences, nous évaluons la performance du modèle de CMC BM-S par rapport au modèle de CMC classique sur un jeu d'images en couleur (Fig. 4.8-a, Fig. 4.9-a,..., Fig. 4.13-a). Pour ce faire, nous avons produit manuellement les images de vérité terrain $X = x$ (Fig. 4.8-b, Fig. 4.9-b,..., Fig. 4.13-b). Considérons l'image en couleur 'Africa' dont nous allons commenter les différents résultats le long de cette section. L'objectif est donc de segmenter cette dernière en 3 classes : ω_1 (eau), ω_2 (terre) et ω_3 (nuages). Comme l'image présente certaines zones sombres, il serait intéressant de procéder à la segmentation de cette dernière selon le formalisme de CMC BM-S en considérant 2 classes auxiliaires selon l'illumination de l'image: λ_1 (zones sombres) et λ_2 (zones claires). Le résultat de la segmentation est par suite comparé à celui fourni par le modèle de CMC classique. Comme il s'agit d'images en couleurs, l'observation $Y = y$ est dans \mathbb{R}^3 . Par conséquent, les versions multi-variées des modèles de CMC et CMC BM-S sont adoptées. Plus explicitement, l'observation est considérée dans l'espace de couleur RGB et 3 distributions de bruit sont considérées pour chaque classe cachée. Ces densités sont supposées gaussiennes et indépendantes.

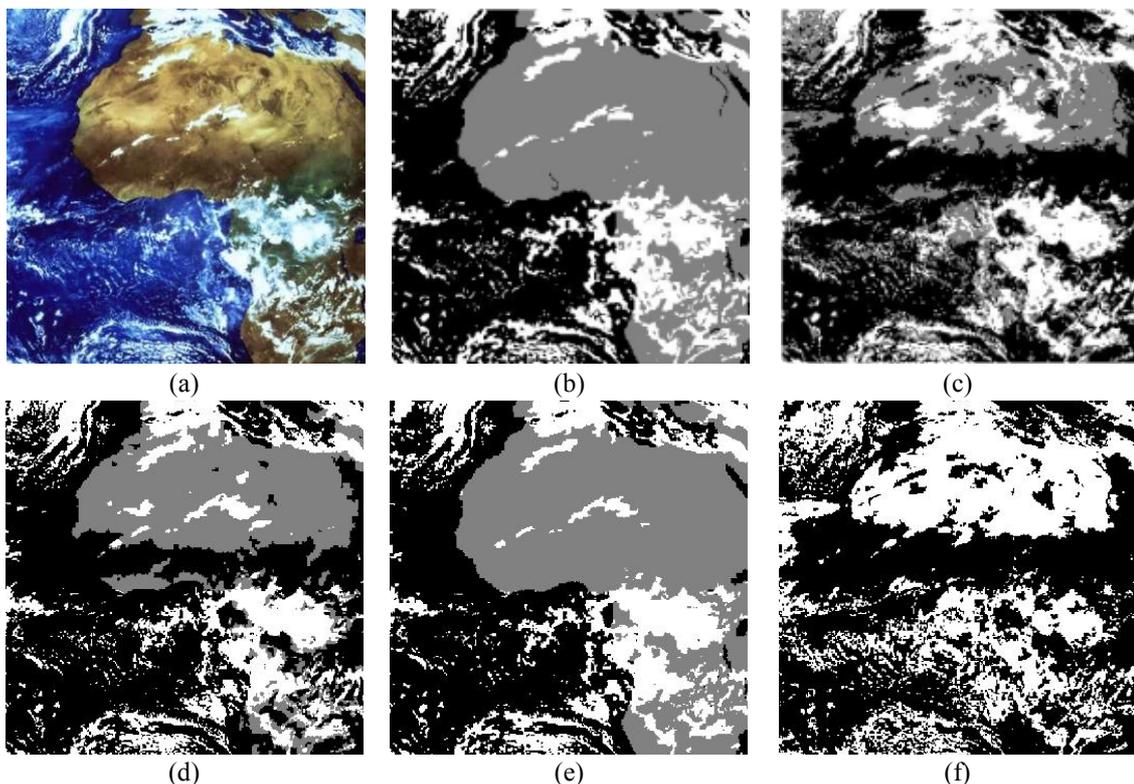


Figure 4.8. Segmentation non-supervisée de l'image 'Africa' selon les modèles CMC et CMC BM-S. (a) Image originale $Y = y$. (b) Vérité terrain établie manuellement $X = x$. (c) Segmentation en 3 classes via *K-means*, taux d'erreur $\tau = 41.7$ %. (d) Segmentation en 3 classes selon le formalisme de CMC, taux d'erreur $\tau = 27.3$ %. (e) Segmentation en 3 classes selon le formalisme de CMC BM-S, taux d'erreur $\tau = 3.8$ %. (f) Estimation de la réalisation du processus auxiliaire U .

Comme illustré dans Fig. 4.8-d, le modèle de CMC confond les parties sombres des classes 'Eau' et 'Terre' tout en enregistrant un taux d'erreur de 27.3%. Ceci est dû au fait que les pixels appartenant aux zones sombres de la classe 'Terre' ont un aspect visuel similaire à ceux appartenant à la classe 'Eau'. Par ailleurs, grâce au processus auxiliaire dont il est doté, le modèle de CMC BM-S peut faire une telle différenciation. Malgré la présence de zones sombres dans l'image, la CMC BM-S fournit un taux moyen de 3.8 %.

La réalisation estimée du processus auxiliaire (Fig. 4.8-f) peut être interprétée comme la carte d'éclairage de l'image. Nous pouvons vérifier visuellement que les pixels de la classe λ_1 appartiennent effectivement aux zones sombres de l'image et vice versa. Des résultats similaires ont été enregistrés pour les autres images (Fig. 4.9, ..., Fig. 4.13), ce qui démontre encore une fois la suprématie du modèle de CMC BM-S par rapport au modèle classique.

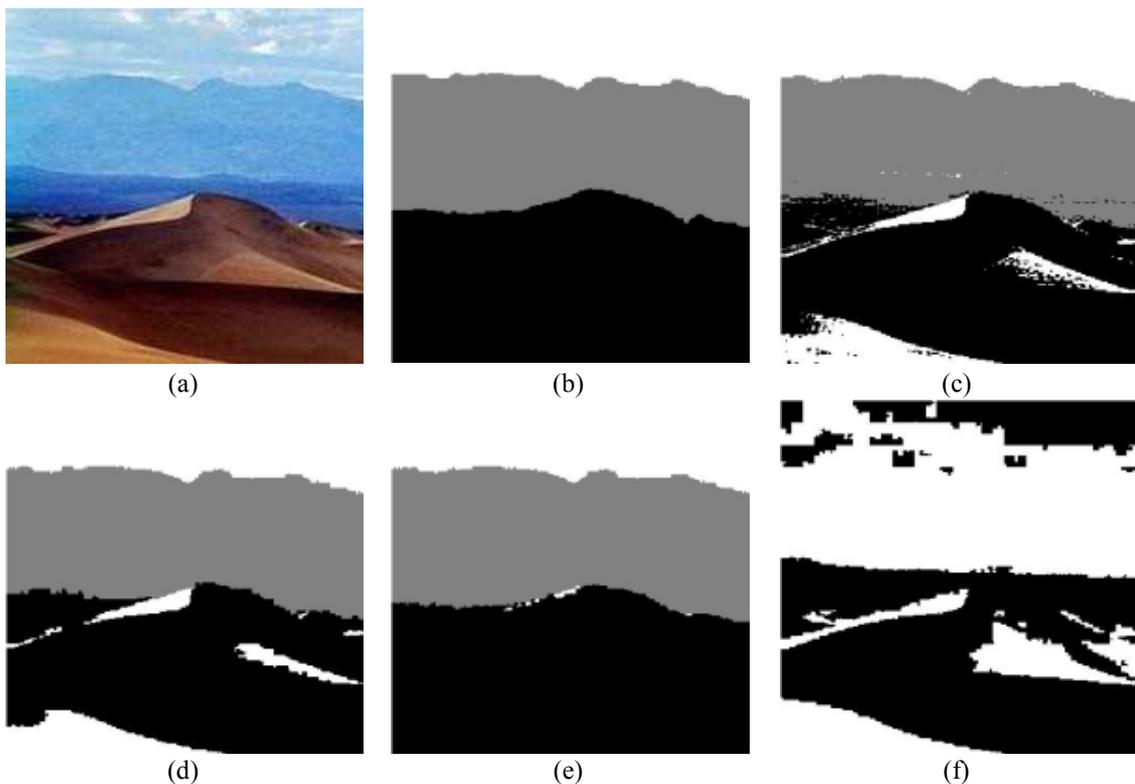


Figure 4.9. Segmentation non-supervisée de l'image 'Landscape' selon les modèles CMC et CMC BM-S. (a) Image originale $Y = y$. (b) Vérité terrain établie manuellement $X = x$. (c) Segmentation en 3 classes via *K-means*, taux d'erreur $\tau = 8.9\%$. (d) Segmentation en 3 classes selon le formalisme de CMC, taux d'erreur $\tau = 8.1\%$. (e) Segmentation en 3 classes selon le formalisme de CMC BM-S, taux d'erreur $\tau = 1.4\%$. (f) Estimation de la réalisation du processus auxiliaire U .

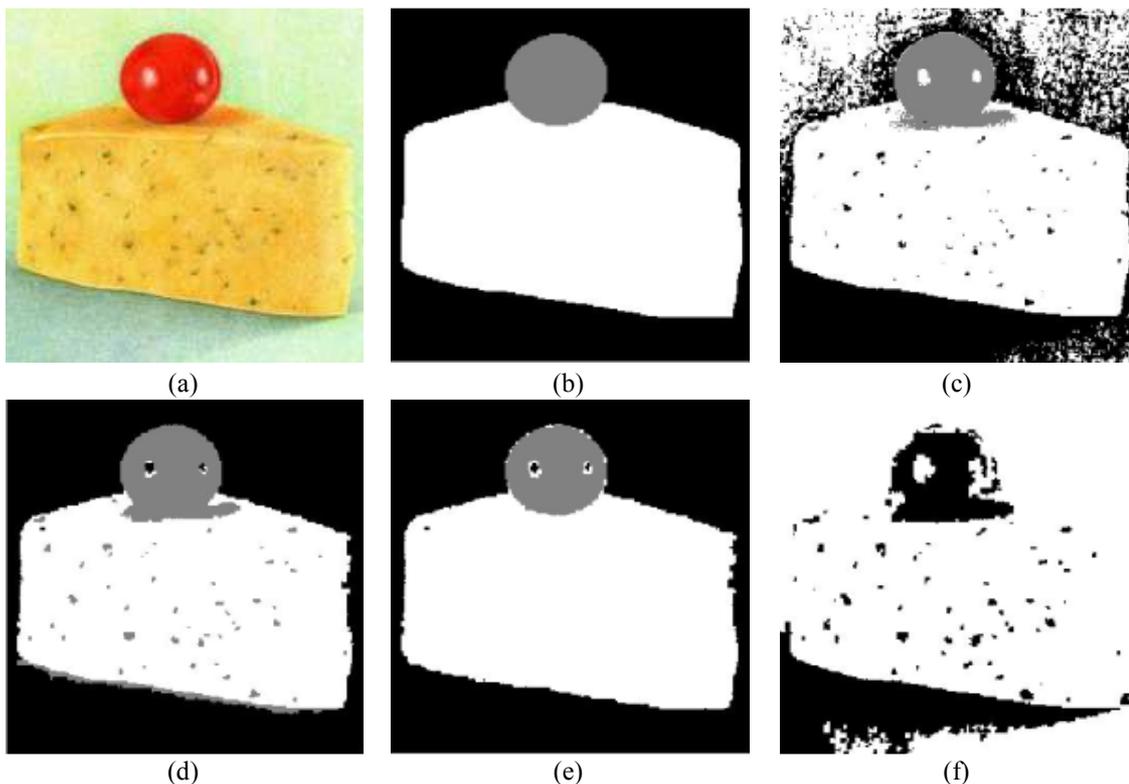


Figure 4.10. Segmentation non-supervisée de l'image 'Cake' selon les modèles CMC et CMC BM-S. (a) Image originale $Y = y$. (b) Vérité terrain établie manuellement $X = x$. (c) Segmentation en 3 classes via *K-means*, taux d'erreur $\tau = 18.9\%$. (d) Segmentation en 3 classes selon le formalisme de CMC, taux d'erreur $\tau = 4.1\%$. (e) Segmentation en 3 classes selon le formalisme de CMC BM-S, taux d'erreur $\tau = 2\%$. (f) Estimation de la réalisation du processus auxiliaire U .

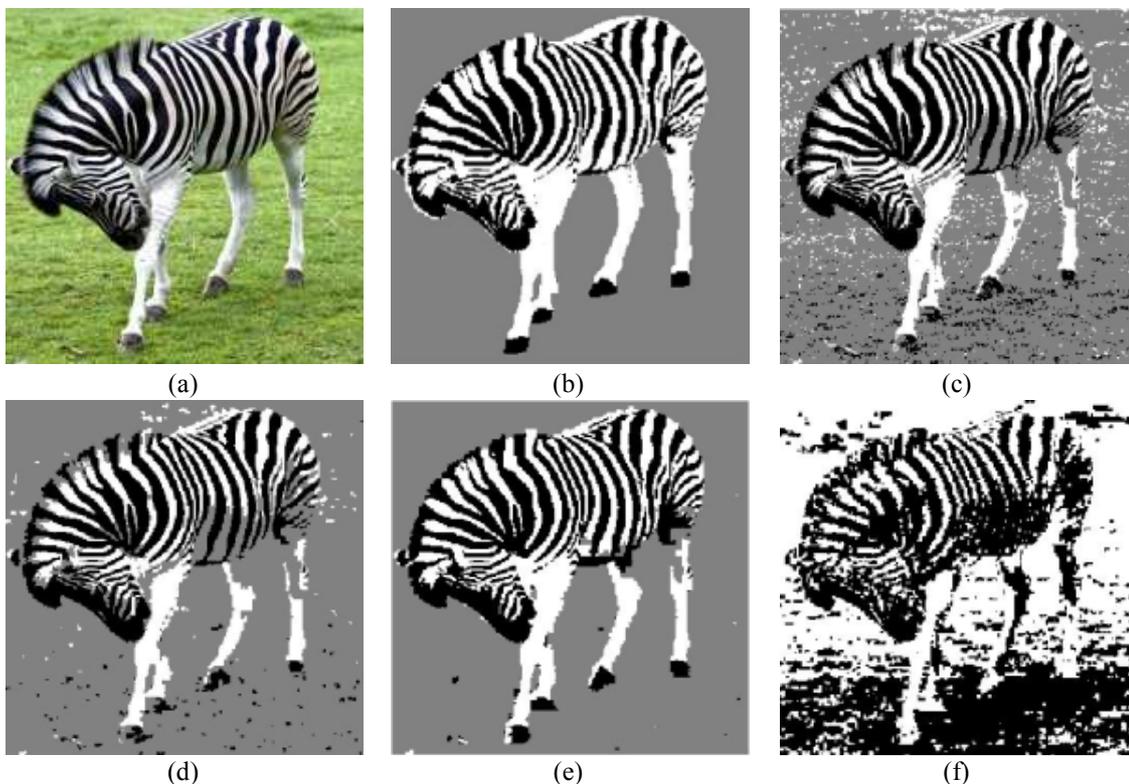


Figure 4.11. Segmentation non-supervisée de l'image 'Zebra' selon les modèles CMC et CMC BM-S. (a) Image originale $Y = y$. (b) Vérité terrain établie manuellement $X = x$. (c) Segmentation en 3 classes via *K-means*, taux d'erreur $\tau = 15.2\%$. (d) Segmentation en 3 classes selon le formalisme de CMC, taux d'erreur $\tau = 11.8\%$. (e) Segmentation en 3 classes selon le formalisme de CMC BM-S, taux d'erreur $\tau = 6.1\%$. (f) Estimation de la réalisation du processus auxiliaire U .

4.6. Chaînes de Markov cachées ML-stationnaire

Dans cette section, nous proposons un modèle original considérant à la fois les changements de régime de la loi *a priori* (tel qu'une CMC M-S) et des distributions de bruit (tel qu'une CMC BM-S). Pour cela, nous allons employer un processus auxiliaire à deux composantes $U = (U^1, U^2)$ où U^1 gère les fluctuations de $p(x)$ et U^2 gère celles des lois $p(y|x)$.

4.6.1. Définition du modèle

Soit $X = (X_n)_{n=1}^N$ un processus caché qui prend ses valeurs dans un ensemble fini de classes $\Omega = \{\omega_1, \dots, \omega_K\}$ et qui doit être estimé à partir d'un processus observable $Y = (Y_n)_{n=1}^N$ où $Y_n \in \mathbb{R}$. Le processus couple $Z = (X, Y)$ est dit chaîne de Markov cachée ML-stationnaire s'il existe un processus couple auxiliaire $U = (U_n)_{n=1}^N$ avec $U_n = (U_n^1, U_n^2)$ où U_n^1 et U_n^2 prennent leurs valeurs dans $\Lambda^1 = \{\lambda_1^1, \dots, \lambda_M^1\}$ et $\Lambda^2 = \{\lambda_1^2, \dots, \lambda_L^2\}$ respectivement, tel que le processus triplet $T = (U, X, Y)$ est une CMT homogène dont la distribution est donnée par :

$$p(t) = p(t_1) \prod_{n=2}^N p(t_n | t_{n-1}) \quad (4.18)$$

où $p(t_1) = p(u_1)p(x_1|u_1^1)p(y_1|x_1, u_1^2)$ et $p(t_n | t_{n-1}) = p(u_n|u_{n-1})p(x_n|x_{n-1}, u_n^1)p(y_n|x_n, u_n^2)$.

L'estimation MPM du processus *a posteriori* est donc possible et sa complexité est linéaire par rapport à la taille N de l'échantillon [Bou11a].

Dans ce qui suit, nous notons $\Lambda = \Lambda^1 \times \Lambda^2$. Ainsi, tout élément $\lambda \in \Lambda$ est un couple égal à (λ^1, λ^2) . De même, toute réalisation u du processus auxiliaire U est implicitement égale à (u^1, u^2) . Par ailleurs, les classes auxiliaires des processus U^1 et U^2 seront notées λ_m et λ_l respectivement, alors que les réalisations du processus couple U seront notées λ_c .

4.6.2. Simulation d'une chaîne de Markov cachée ML-stationnaire

Notre champ d'étude est limité aux CMC ML-S gaussiennes. La distribution de bruit $p(y_n|x_n, u_n^2)$ est donc une densité gaussienne et par conséquent, nous définissons les paramètres d'une CMC ML-S comme suit :

- Les distributions initiales des processus latents U et X sont données par ϖ et π respectivement. Ainsi, $p(u_1 = \lambda_c) = \varpi_c$ où $\lambda_c \in \Lambda$ et $p(x_1 = \omega_k | u_1^1 = \lambda_m) = \pi_{m,k}$;
- Les distributions de transition $p(u_n|u_{n-1})$ et $p(x_n|x_{n-1}, u_n^1)$ sont caractérisées par les matrices Q et A respectivement. Ainsi, $p(u_n = \lambda_{c'} | u_{n-1} = \lambda_c) = q_{c,c'}$ et $p(x_n = \omega_{k'} | x_{n-1} = \omega_k, u_n^1 = \lambda_m) = a_{m,k,k'}$ respectivement ;
- Les distributions de bruit $p(y_n|x_n = \omega_k, u_n^2 = \lambda_l)$ étant gaussiennes, sont donc paramétrées par les moyennes $\mu_{k,l}$ et les écarts-types $\sigma_{k,l}$. Comme nous avons K classes cachées et L classes auxiliaires, notre modèle comporte donc $L \times K$ densités gaussiennes et par conséquent, $L \times K$ moyennes et $L \times K$ écarts-types.

Les paramètres θ du modèle CMC ML-S étant définis, nous montrons maintenant comment simuler la réalisation d'un tel modèle :

1. Simuler la réalisation du processus auxiliaire U selon le formalisme d'une chaîne de Markov classique par tirages à partir de ϖ et Q respectivement ;
2. Simuler la réalisation du processus caché X selon le formalisme d'une CMC ML-S conformément à la réalisation du processus auxiliaire u^1 par tirages à partir de π et A respectivement ;
3. Simuler la réalisation du processus observable Y par tirage aléatoire à partir des densités gaussiennes correspondantes aux réalisations de u^2 et x .

4.6.3. Restauration MPM d'une chaîne de Markov cachée ML-stationnaire

Soit une réalisation $Y = y$ et supposons que les paramètres $\theta = (\varpi, \pi, Q, A, \mu, \sigma)$ du modèle sont connus. Nous présentons à présent la méthode bayésienne MPM permettant d'estimer les réalisations des processus caché $\hat{X} = \hat{x}$ et auxiliaire $\hat{U} = \hat{u}$ à partir de $Y = y$.

L'estimateur MPM est donné par la formule

$$[\hat{x} = \hat{s}_{MPM}(y)] \Leftrightarrow \left[\hat{x}_n = \underset{\omega}{\operatorname{argmax}} p(x_n = \omega | y) \right] \quad (4.19)$$

Nous définissons les fonctions progressives $\alpha_n(x_n, u_n) = p(y_1, \dots, y_n, x_n, u_n)$ et rétrogrades $\beta_n(x_n) = p(y_{n+1}, \dots, y_N | x_n, u_n)$ qui peuvent être calculées itérativement comme suit

Initialisation :

$$\alpha_1(x_1, u_1) = p(u_1)p(x_1|u_1^1)p(y_1|x_1, u_1^2) \quad (4.20)$$

$$\beta_N(x_N, u_N) = 1 \quad (4.21)$$

Induction :

$$\alpha_n(x_n, u_n) = \sum_{x_{n-1}, u_{n-1}} \alpha_{n-1}(x_{n-1}, u_{n-1})p(u_n|u_{n-1})p(x_n|x_{n-1}, u_n^1)p(y_n|x_n, u_n^2) \quad (4.22)$$

$$\beta_n(x_n, u_n) = \sum_{x_{n+1}, u_{n+1}} \beta_{n+1}(x_{n+1}, u_{n+1})p(u_{n+1}|u_n)p(x_{n+1}|x_n, u_{n+1}^1)p(y_{n+1}|x_{n+1}, u_{n+1}^2) \quad (4.23)$$

Les distributions *a posteriori* $p(x_n, u_n | y)$ peuvent alors être calculées selon les formules

$$p(x_n, u_n | y) \propto \alpha_n(x_n, u_n) \beta_n(x_n, u_n) \quad (4.24)$$

Les distributions marginales *a posteriori* du processus caché et auxiliaire sont dérivées comme suit

$$p(x_n | y) = \sum_{u_n} p(x_n, u_n | y) \quad (4.25)$$

$$p(u_n | y) = \sum_{x_n} p(x_n, u_n | y) \quad (4.26)$$

Bien entendu, l'estimation des réalisations de chacun des processus auxiliaires U^1 et U^2 est également effectuée par sommation selon des formules similaires.

4.6.4. Estimation des paramètres d'une chaîne de Markov cachée ML-stationnaire

Pour que le modèle de CMC ML-S soit applicable dans le contexte non-supervisé, il est indispensable d'estimer ses paramètres. Pour ce faire, nous proposons une procédure de type EM permettant d'accomplir cette tâche sous certaines conditions. Comme, les rôles des deux processus auxiliaires U^1 et U^2 sont interchangeable, nous supposons que les densités de bruits relatives à une classe cachée données sont plus proches les une des autres que n'importe quelle densité de bruit d'une autre classe cachée.

Avant de détailler le déroulement de l'algorithme d'estimation, nous définissons les distributions *a posteriori* $\psi_n(x_n, u_n, x_{n+1}, u_{n+1}) = p(x_n, u_n, x_{n+1}, u_{n+1} | y)$ et $\xi_n(x_n, u_n) = p(x_n, u_n | y)$ données par les formules

$$\psi_n(x_n, u_n, x_{n+1}, u_{n+1}) \propto \alpha_n(x_n, u_n) p(u_{n+1} | u_n) p(x_{n+1} | x_n, u_{n+1}^1) p(y_{n+1} | x_{n+1}, u_{n+1}^2) \beta_{n+1}(x_{n+1}, u_{n+1}) \quad (4.27)$$

$$\xi_n(x_n, u_n) = \sum_{x_{n+1}, u_{n+1}} \psi_n(x_n, u_n, x_{n+1}, u_{n+1}) \quad (4.28)$$

Enfin, la procédure de l'estimation est exécutée de la manière itérative suivante :

- Considérer un jeu initial de paramètres $\theta^0 = (\varpi^0, \pi^0, Q^0, A^0, \mu^0, \sigma^0)$;
- Pour chaque itération $i+1$:
 - o Estimation : Calculer les fonctions ψ_n^i and ξ_n^i à partir de y et θ^i .
 - o Maximisation : Estimer le nouveau jeu de paramètres θ^{i+1} comme suit

$$\varpi_c^{i+1} = \sum_{x_n} \xi_1^i(x_n, u_n = \lambda_c) \quad (4.29)$$

$$\pi_{m,k}^{i+1} = \frac{\xi_1^i(x_n = \omega_k, u_n^1 = \lambda_m, u_n^2)}{\sum_{x_n, u_n^2} \xi_1^i(x_n, u_n^1 = \lambda_m, u_n^2)} \quad (4.30)$$

$$q_{m,m'}^{i+1} = \frac{\sum_{n=1}^{N-1} \sum_{x_n, x_{n+1}, u_n = \lambda_m, u_{n+1} = \lambda_{m'}} \psi_n^i(x_n, u_n, x_{n+1}, u_{n+1})}{\sum_{n=1}^{N-1} \sum_{x_n, u_n = \lambda_m} \xi_n^i(x_n, u_n)} \quad (4.31)$$

$$a_{m,k,k'}^{i+1} = \frac{\sum_{n=1}^{N-1} \sum_{u_n^1 = \lambda_m, u_n^2, u_{n+1}, x_n = \omega_k, x_{n+1} = \omega_{k'}} \psi_n^i(x_n, u_n, x_{n+1}, u_{n+1})}{\sum_{n=1}^{N-1} \sum_{u_n^1 = \lambda_m, u_n^2, x_n = \omega_k} \xi_n^i(x_n, u_n)} \quad (4.32)$$

$$\mu_{k,m}^{i+1} = \frac{\sum_{n=1}^N \sum_{x_n = \omega_k, u_n^1, u_n^2 = \lambda_m} \xi_n^i(x_n, u_n) y_n}{\sum_{n=1}^N \sum_{x_n = \omega_k, u_n^1, u_n^2 = \lambda_m} \xi_n^i(x_n, u_n)} \quad (4.33)$$

$$\sigma_{k,m}^{i+1} = \frac{\sum_{n=1}^N \sum_{x_n = \omega_k, u_n^1, u_n^2 = \lambda_m} \xi_n^i(x_n, u_n) (y_n - \mu_{k,m}^{i+1})^2}{\sum_{n=1}^N \sum_{x_n = \omega_k, u_n^1, u_n^2 = \lambda_m} \xi_n^i(x_n, u_n)} \quad (4.34)$$

4.6.5. Résultats expérimentaux

Pour valider le modèle de CMC ML-S, nous l'appliquons au problème de restauration d'images. Pour cela, nous proposons deux expériences. Dans la première, nous générons des images synthétiques selon le formalisme des CMC ML-S. Dans la seconde, en revanche, une image de classes non-stationnaire est bruitée selon le formalisme des CMC ML-S. Pour les deux expériences, les images sont converties de et vers des séquences unidimensionnelles à l'aide du parcours d'Hilbert-Peano [Pea1890, Hil1891], et les paramètres sont initialisés via l'algorithme *K-means*.

- Expérience 1 : Dans cette expérience, les données sont simulées selon le formalisme des CMC ML-S. Soit $T = (U, X, Y)$ une CMC ML-S avec $T = (T_n)_{n=1}^N$ où $N = 16384$, U_n prend ses valeurs dans l'ensemble $\Lambda = \Lambda^1 \times \Lambda^2$ où $\Lambda^1 = \{\lambda_1^1, \lambda_2^1, \lambda_3^1\}$ et $\Lambda^2 = \{\lambda_1^2, \lambda_2^2\}$, X_n prend ses valeurs dans l'ensemble $\Omega = \{\omega_1, \omega_2\}$ et Y_n prend ses valeurs dans l'ensemble des réels \mathbb{R} . Les données sont générées de la manière suivante
 - La première réalisation u_1 du processus auxiliaire est simulée par un tirage aléatoire uniforme sur Λ . Les réalisations suivantes du même processus sont simulées de proche en proche via un tirage aléatoire à partir de la matrice de transition

$$Q = \frac{1}{10^3} \begin{bmatrix} 955 & 3 & 1 & 0 & 1 & 0 \\ 3 & 955 & 0 & 1 & 0 & 1 \\ 1 & 0 & 955 & 3 & 1 & 0 \\ 0 & 1 & 3 & 955 & 0 & 1 \\ 1 & 0 & 1 & 0 & 955 & 3 \\ 0 & 1 & 0 & 1 & 3 & 955 \end{bmatrix}.$$

- La première réalisation x_1 du processus caché est simulée via un tirage aléatoire uniforme sur l'ensemble des classes Ω . Les réalisations suivantes du même processus sont simulées à partir des tirages aléatoires conformément à la réalisation du processus auxiliaire U^1 selon les matrices de transition suivantes

$$A_1 = \begin{pmatrix} 0.99 & 0.01 \\ 0.01 & 0.99 \end{pmatrix}, A_2 = \begin{pmatrix} 0.77 & 0.23 \\ 0.23 & 0.77 \end{pmatrix}, A_3 = \begin{pmatrix} 0.55 & 0.45 \\ 0.45 & 0.55 \end{pmatrix}.$$

- Les réalisations du processus observable Y sont simulées via des tirages aléatoires conformément aux réalisations des processus X et U^1 selon les densités gaussiennes dont les paramètres sont

$$\mu = \begin{bmatrix} -1.25 & -0.75 \\ 0.75 & 1.25 \end{bmatrix}, \sigma_{i,j} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

La restauration MPM a été accomplie selon le formalisme des CMC ML-S sur la base des paramètres estimés grâce à l'algorithme EM et des paramètres réels (de simulation). La restauration MPM a également été effectuée selon le formalisme des CMC M-S sur la base des paramètres estimés à travers l'algorithme EM et également sur la base des paramètres (moyens) suivants en guise de paramètres réels

$$\mu = \begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix}, \sigma = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

Les résultats de restauration obtenus sont présentés dans TAB. 4.III. Certains résultats visuels sont également illustrés dans Fig. 4. 12.

- Expérience 2 : Dans cette expérience, nous considérons l'image de classes non stationnaire '*Nazca bird*' que nous bruitons selon au formalisme des CMC ML-S en utilisant les mêmes densités de bruit de la première expérience conformément aux réalisations des processus X et U^2 . La première réalisation de ce dernier est simulée par tirage uniforme à partir de l'ensemble $\Lambda^2 = \{\lambda_1^2, \lambda_2^2\}$ et les prochaines réalisations sont simulées selon la matrice de transition suivante

$$A_{u_n^2, u_{n+1}^2} = \begin{bmatrix} 0.998 & 0.002 \\ 0.002 & 0.998 \end{bmatrix}.$$

Notons que la réalisation du processus U^1 régissant les changements de régime du processus caché X est inconnue ici et doit être estimée. La restauration non-supervisée MPM a été réalisée selon les modèles de CMC M-S et CMC ML-S. Les résultats obtenus sont illustrés dans TAB. 4.III et Fig. 4.13.

Taux d'erreur moyens de segmentation des images synthétique bruitées (%)

Expérience	Expérience 1				Expérience 2	
	CMC ML-S θ	CMC M-S $E(\theta)$	CMC ML-S $\hat{\theta}$	CMC M-S $\hat{\theta}$	CMC ML-S $\hat{\theta}$	CMC M-S $\hat{\theta}$
τ_X (%)	8.44	9.20	8.41	9.23	4.84	5.36
τ_{U^1} (%)	2.86	3.70	3.78	4.33	-	-
τ_{U^2} (%)	5.76	-	7.60	-	5.51	-

A partir des résultats obtenus, nous constatons que la prise en compte des fluctuations des densités du bruit améliore la qualité de la restauration. Il est important de constater que la restitution du processus auxiliaire U^2 peut être utile pour certaines applications en traitement d'images.

La seconde expérience est d'une importance cruciale car nous considérons une réalisation non-simulée du processus caché X . Comme nous pouvons le constater dans Fig. 4.13, le modèle proposé parvient à bien estimer les réalisations de U^1 et U^2 . La première représente les trois stationnarités de l'image de classes : les pixels de l'arrière plan sont assignés à λ_1^1 (noir), les pixels appartenant au corps de l'oiseau sont assignés à λ_2^1 (gris) et les pixels appartenant aux ailes de l'oiseau sont assignés à λ_3^1 (blanc). Par ailleurs, la réalisation du second processus représente les deux stationnarités du bruit dans l'image : les pixels sombres sont assignés à λ_1^2 (noir), alors que les pixels clairs sont assignés à λ_2^2 (blanc).

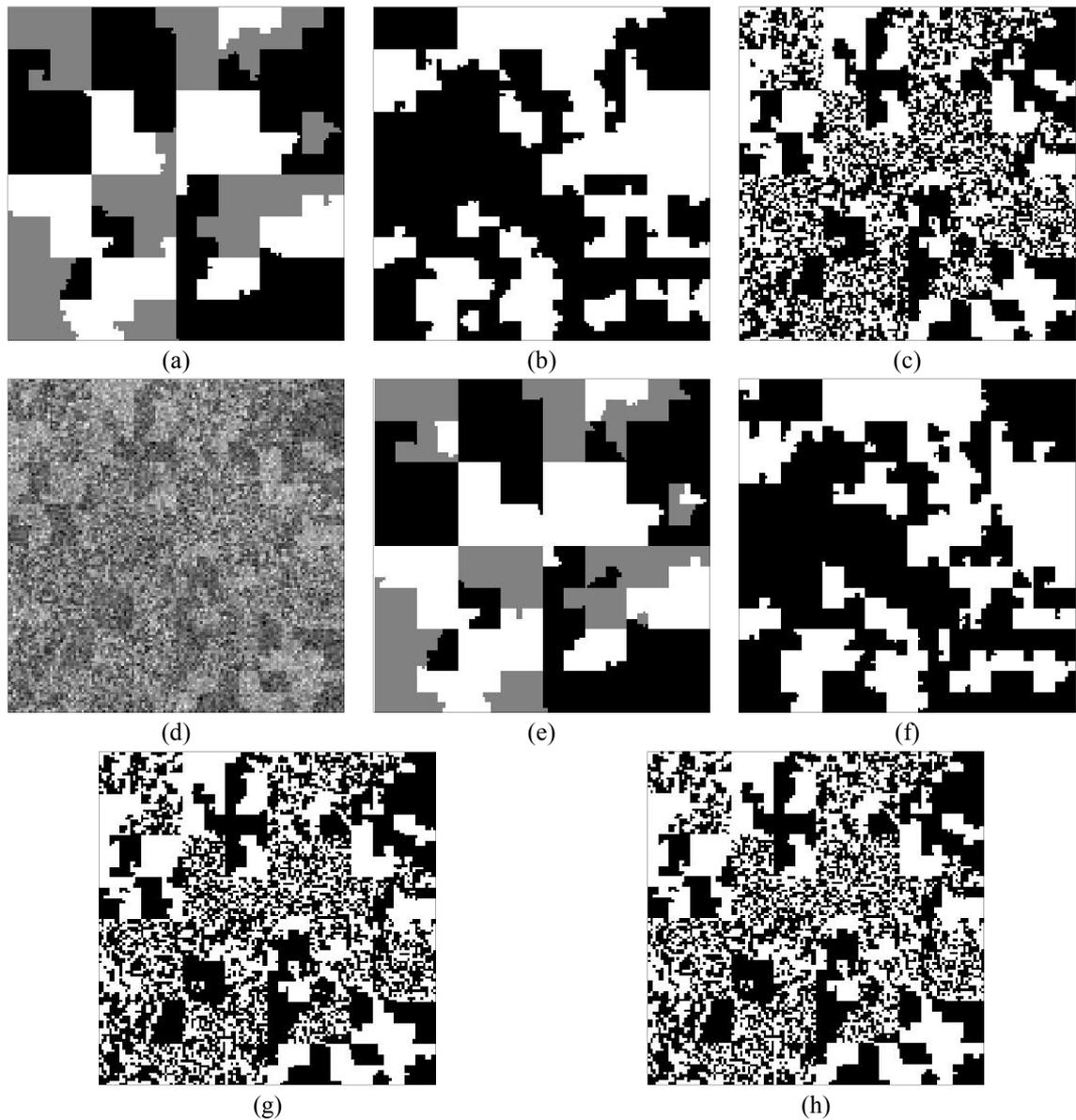


Figure 4.12. Exemple de segmentation non-supervisée d'une image synthétique non-stationnaire à bruit non-stationnaire selon les CMC ML-S. (a) Réalisation du processus auxiliaire $U^1 = u^1$. (b) Réalisation du processus auxiliaire $U^2 = u^2$. (c) Réalisation du processus caché $X = x$. (d) Image observée $Y = y$. (e) Estimation non-supervisée de la réalisation du processus auxiliaire U^1 , taux d'erreur $\tau = 3.8\%$. (f) Estimation non-supervisée de la réalisation du processus auxiliaire U^2 , taux d'erreur $\tau = 7.6\%$. (g) Estimation de la réalisation du processus caché X en utilisant les paramètres réels, taux d'erreur $\tau = 8.4\%$. (h) Estimation non-supervisée de la réalisation du processus caché X , taux d'erreur $\tau = 8.4\%$.

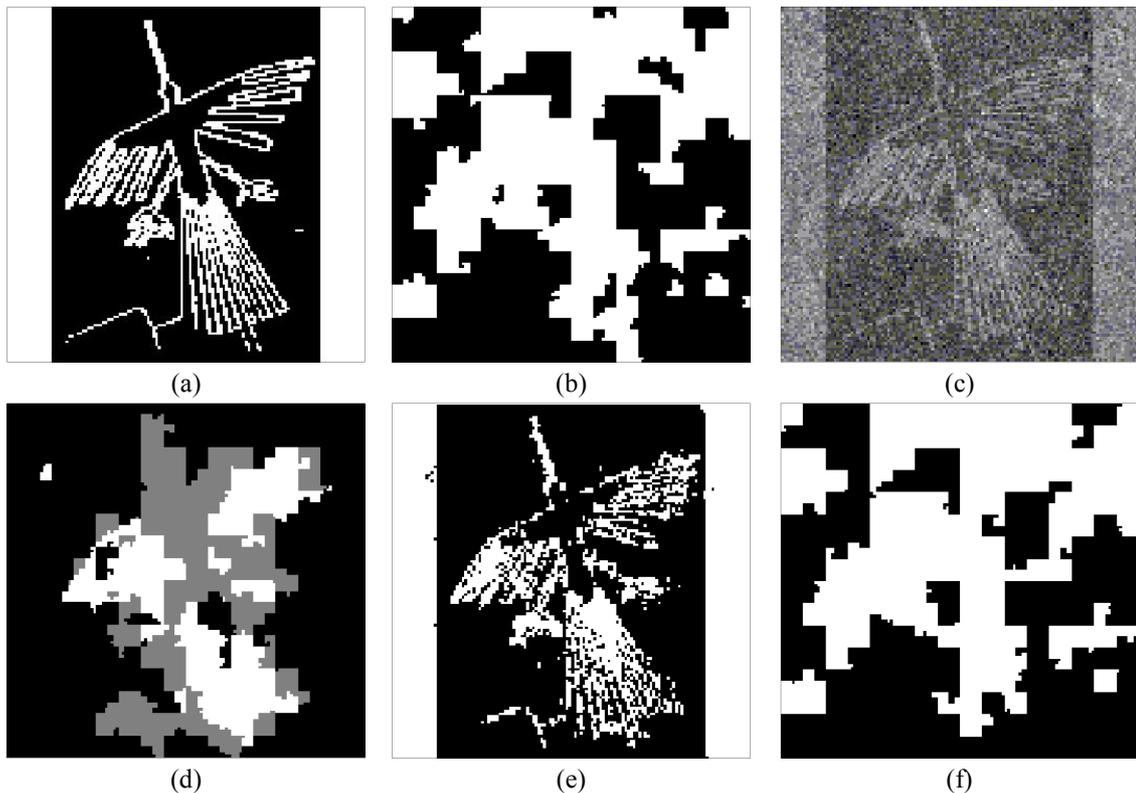


Figure 4.13. Segmentation non-supervisée de l'image ‘Nazca bird’ selon les CMC ML-S. (a) Image originale de classes $X = x$. (b) Réalisation du processus auxiliaire $U^1 = u^1$. (c) Image observée $Y = y$. (d) Estimation de la réalisation de u^1 . (e) Segmentation en 2 classes selon le formalisme de CMC ML-S, taux d’erreur $\tau = 4.8 \%$. (f) Estimation de la réalisation du processus auxiliaire $U^1 = u^1$, taux d’erreur $\tau = 5.5 \%$.

4.7. Conclusion

Dans ce chapitre, nous avons proposé deux nouveaux modèles de type CMT qui étendent le modèle de CMC classique dans la mesure où ils permettent de prendre en considération les sauts aléatoires des distributions de bruit dans contexte markovien. Ces modèles originaux, peuvent s’appliquer dans toutes les situations où les données observables dépendent, non seulement des classes cachées comme dans le cas de CMC habituelles, mais aussi d’un autre processus sous-jacent. Ce dernier peut correspondre, par exemple, aux conditions d’éclairage dans une image (ou présence d’une ombre). Nous avons démontré à travers l’étude expérimentale que les modèles de CMC BM-S et CMC ML-S proposés permettaient d’accomplir l’estimation MPM dans le contexte non-supervisé (en supposant le nombre de classes auxiliaires connu) grâce à un algorithme original du type EM.

Une piste intéressante de travail futur serait d’étendre ces modélisations au contexte de champs de Markov cachés. Ces modèles, qui seront définis de façon équivalente seront mieux adaptés aux images. En effet, beaucoup d’études ont confirmé la supériorité des modèles 2D par rapport aux modèles monodimensionnels pour les données de type image [Gem84, Bes86]. Une autre perspective serait d’appliquer le modèle de CMC BM-S aux problèmes de segmentation ou de reconnaissance de formes. A titre d’exemple, nous pouvons citer le problème de segmentation d’iris à partir d’images de visage [Ben13a, Ben13b].

Chapitre 5. Chaînes de Markov couples M-stationnaires

Grâce à la théorie solide dont elles sont dotées, les chaînes de Markov cachées ont été appliquées avec succès dans plusieurs domaines dont le traitement du signal et de l'image. Ces modèles ont été ensuite généralisés aux chaînes de Markov couples [Pie03a, Der04] qui offrent davantage de possibilités de modélisation tout en présentant une complexité de calcul comparable. L'un des points forts des deux modèles est la possibilité d'effectuer les différents traitements d'estimation dans le contexte non-supervisé grâce à des procédures itératives d'estimation tels que l'algorithme Espérance-Maximisation (EM). Cependant, de tels algorithmes supposent les différents paramètres indépendants de la position des variables aléatoires dans la séquence. D'ailleurs, le jeu final de paramètres fourni par ces algorithmes est constitué de valeurs moyennes des paramètres estimées localement au niveau de chaque site. Par conséquent, l'application de ces modèles aux données non-stationnaires fournit des résultats inadaptés. Pour pallier à ce problème, un modèle du type chaîne de Markov triplet, appelé chaîne de Markov caché M-stationnaire (ou M-non stationnaire), a été défini. Le principe de ce dernier est de supposer la donnée à modéliser stationnaire par partie et de préconiser un jeu de paramètres différent (distributions *a priori* notamment) pour chacune de ces M parties.

Dans ce chapitre, nous proposons d'exploiter la théorie riche des chaînes de Markov triplet pour permettre l'application des CMCo aux données non-stationnaires dans le contexte non-supervisé. Dans un premier temps, nous définissons le modèle de chaîne de Markov couple M-stationnaire où seules les distributions *a priori* des données cachées sont concernées par la non-stationnarité. Ensuite, nous définissons le modèle de chaîne de Markov couple MM-stationnaire où les distributions de bruit sont également M-stationnaires. Pour les deux modèles, nous détaillons la technique bayésienne de restauration MPM et une version adaptée de l'algorithme d'estimation des paramètres EM. Enfin, pour valider les modèles proposés, des expériences sont menées sur des données synthétiques et réelles.

5.1. Chaîne de Markov couples M-stationnaire

Dans ce premier volet du chapitre, nous définissons le modèle de chaîne de Markov couple M-stationnaire qui permet à la fois d'améliorer les performances des CMCo et des CMC M-stationnaires dans le contexte non-supervisé. Dans ce cadre, nous explicitons la méthode d'estimation bayésienne MPM ainsi qu'un algorithme originale d'estimation des paramètres du type EM. Enfin, une étude expérimentale est menée dans le but de confirmer l'intérêt du modèle proposé par rapport à ses prédécesseurs.

5.1.1. Définition du modèle

Soit $X = (X_n)_{n=1}^N$ un processus inobservable à valeurs dans $\Omega = \{\omega_1, \dots, \omega_K\}$ et que l'on doit estimer à partir d'un processus observable $Y = (Y_n)_{n=1}^N$ à valeurs dans \mathbb{R} et qui peut être vu comme une version bruitée de X . Nous nous intéressons à la situation où $Z = (X, Y)$ est une

chaîne de Markov couple dont les paramètres dépendent de la réalisation d'un troisième processus $U = (U_n)_{n=1}^N$ à valeurs dans $\Lambda = \{\lambda_1, \dots, \lambda_M\}$ selon la formule :

$$p(t_n|t_{n-1}) = p(u_n|u_{n-1}, z_{n-1})p(z_n|u_n, u_{n-1}, z_{n-1}) \quad (5.1)$$

$T = (U, X, Y)$ est donc une chaîne de Markov triplet particulière qui étend le modèle de chaîne de Markov cachée M-stationnaire (CMC M-S) [Lan06, Lan12] dont la probabilité de transition est donnée par :

$$p(t_n|t_{n-1}) = p(u_n|u_{n-1}, x_{n-1})p(x_n|u_n, u_{n-1}, x_{n-1})p(y_n|x_n) \quad (5.2)$$

Tout le long de cette étude, nous supposerons que le processus U est markovien et que les transitions $p(z_n|z_{n-1})$ ne dépendent que de la réalisation courante de U_n . L'équation (5.1) devient donc

$$p(t_n|t_{n-1}) = p(u_n|u_{n-1})p(z_n|z_{n-1}, u_n) \quad (5.3)$$

Plus explicitement la probabilité de transition du processus T peut s'écrire comme suit :

$$p(t_n|t_{n-1}) = p(u_n|u_{n-1})p(x_n|x_{n-1}, y_{n-1}, u_n)p(y_n|x_{n-1}, y_{n-1}, x_n, u_n) \quad (5.4)$$

Dans un premier temps, nous simplifions l'équation (5.4) en posant

$$p(y_n|x_{n-1}, y_{n-1}, x_n, u_n) = p(y_n|x_{n-1}, y_{n-1}, x_n) \quad (5.5)$$

L'équation (5.4) devient

$$p(t_n|t_{n-1}) = p(u_n|u_{n-1})p(x_n|u_n, x_{n-1}, y_{n-1})p(y_n|x_n, x_{n-1}, y_{n-1}) \quad (5.6)$$

L'équation (5.6) définit une chaîne de Markov triplet particulière que nous appellerons chaîne de Markov couple M-stationnaire (CMCo M-S). Nous reviendrons plus bas à l'équation (5.4) qui définit un modèle plus général.

La prédominance des modèles de CMCo par rapport à leurs devanciers CMC a été prouvée dans plusieurs travaux [Pie03a, Der04]. Nous allons voir que cette propriété reste valide dans le cas des modèles M-stationnaires puisque Z demeure bel et bien une CMCo, seule ses paramètres changent le long de la séquence de donnée.

Dans ce qui suit, nous adoptons les notations

$$p_m(i, j) = p(x_{n-1} = \omega_i, x_n = \omega_j | u_n = \lambda_m) \quad (5.7)$$

$$f_{i,j}(y_{n-1}, y_n) = p(y_{n-1}, y_n | x_{n-1} = \omega_i, x_n = \omega_j) \quad (5.8)$$

La distribution $p(z|u)$ peut s'écrire à travers :

$$p(z_{n-1}, z_n | u_n = \lambda_m) = p_m(i, j) f_{i,j}(y_{n-1}, y_n) \quad (5.9)$$

Cette même distribution peut être déterminée également de la manière classique, en l'occurrence, à travers les probabilités initiales (5.10) et de transition (5.11).

$$p(z_1|u_2 = \lambda_m) = \sum_{\omega_k \in \Omega} p_m(i, k) f_{i,k}(y_1) \quad (5.10)$$

où $f_{i,k}(y_1) = \int f_{i,k}(y_1, y_2) dy_2$.

$$p(z_n | z_{n-1}, u_n = \lambda_m) = \frac{p(z_{n-1}, z_n | u_n)}{p(z_{n-1} | u_n)} = \frac{p_m(i,j) f_{i,j}(y_{n-1}, y_n)}{\sum_{\omega_k \in \Omega} p_m(i,k) f_{i,k}(y_{n-1})} \quad (5.11)$$

5.1.2. Paramètres et simulation d'une CMCo M-S gaussienne

Une CMCo M-S est dite gaussienne si les densités de bruit $f_{i,j}(y_{n-1}, y_n)$ le sont. Un tel modèle est entièrement défini à travers les M probabilités initiales données par le vecteur $\varpi = [\varpi_m]_{\lambda_m \in \Lambda}$ où $\varpi_m = p(u_1 = \lambda_m)$, les M^2 probabilités données par la matrice de transition $A = [a_{\lambda, \lambda'}]_{\lambda, \lambda' \in \Lambda}$ où $a_{\lambda, \lambda'} = p(u_n = \lambda' | u_{n-1} = \lambda)$, les MK^2 densités bidimensionnelles $p_{\lambda_m}(i, j)$ données par les matrices $(P_m)_{m=1}^M$ et les K^2 densités gaussiennes de bruit définies par les vecteurs moyennes $\mu_1^{i,j}, \mu_2^{i,j}$, écarts-types $\sigma_1^{i,j}, \sigma_2^{i,j}$ et les coefficients de corrélation $\rho^{i,j}$.

Etant donnés les paramètres $\Theta = (\varpi, A, P, \mu, \sigma, \rho)$ d'une CMCo M-S, nous pouvons simuler une réalisation de cette dernière comme suit :

1. Simuler la réalisation du processus auxiliaire U , selon le formalisme d'une chaîne de Markov classique, par tirage aléatoire à partir de ϖ et A respectivement ;
2. La réalisation z_1 est simulée par tirages à partir des distributions

$$p(x_1 = \omega_k | u_1 = \lambda_m) = \sum_{\omega_i \in \Omega} p_m(k, i) \quad (5.12)$$

$$p(y_1 | x_1 = \omega_k, u_1 = \lambda_m) = \sum_{\omega_i \in \Omega} p_m(k, i) f_{k,i}(y_1) \quad (5.13)$$

Notons que $f_{k,i}(y_1)$ est une densité gaussienne de moyenne $\mu_1^{k,i}$ et d'écart-type $\sigma_1^{k,i}$. La réalisation y_1 est donc obtenue par tirage à partir d'un mélange gaussien.

3. Simuler les prochaines réalisations du processus Z par tirages aléatoires à partir des distributions

$$p(x_n = \omega_j | u_n = \lambda_m, x_{n-1} = \omega_i, y_{n-1}) = \frac{p_m(i,j) f_{i,j}(y_{n-1})}{\sum_{\omega_k \in \Omega} p_m(i,k) f_{i,k}(y_{n-1})} \quad (5.14)$$

$$p(y_n | x_n = \omega_j, x_{n-1} = \omega_i, y_{n-1}) = \frac{f_{i,j}(y_{n-1}, y_n)}{f_{i,j}(y_{n-1})} \quad (5.15)$$

Il est possible de démontrer que $p(y_n | x_n, z_{n-1})$ est une densité gaussienne de moyenne $\mu_2^{i,j} + \rho^{i,j}(\sigma_2^{i,j}/\sigma_1^{i,j})(y_{n-1} - \mu_1^{i,j})$ et d'écart-type $\sigma_2^{i,j} \sqrt{1 - (\rho^{i,j})^2}$ [Der04]. La simulation de y_n est donc obtenue par tirage aléatoire à partir d'un mélange de gaussiennes.

5.1.3. Restauration MPM d'une chaîne de Markov couple M-stationnaire

Soit $T = (U, X, Y)$ une CMCo M-S définie par la formule (5.6). Notons $V = (U, X)$. $T = (V, Y)$ est donc une CMCo. Par conséquent, les probabilités marginales *a posteriori* $p(v_n | y) = p(u_n, x_n | y)$ nécessaires pour la restauration MPM sont calculables.

Lorsque les paramètres Θ du modèle sont connus, les distributions *a posteriori* $p(u_n, x_n | y)$ sont calculables à travers les fonctions progressives généralisées $\alpha_n(v_n) = p(y_1, \dots, y_n, v_n)$ et les

fonctions rétrogrades $\beta_n(x_n) = p(y_{n+1}, \dots, y_N | v_n, y_n)$ et dont le calcul se fait de la manière itérative suivante

Initialisation :

$$\alpha_1(v_1) = p(y_1, v_1) \quad (5.16)$$

$$\beta_N(v_N) = 1 \quad (5.17)$$

Induction :

$$\alpha_n(v_n) = \sum_{v_{n-1} \in \Lambda \times \Omega} \alpha_{n-1}(v_{n-1}) p(u_n | u_{n-1}) p(z_n | z_{n-1}, u_n) \quad (5.18)$$

$$\beta_n(v_n) = \sum_{v_{n+1} \in \Lambda \times \Omega} \beta_{n+1}(v_{n+1}) p(u_n | u_{n-1}) p(z_{n+1} | z_n, u_{n+1}) \quad (5.19)$$

Les distributions *a posteriori* $p(v_n | y)$ sont ensuite calculées comme suit

$$p(v_n | y) \propto \alpha_n(v_n) \beta_n(v_n) \quad (5.20)$$

Puis, les distributions marginales *a posteriori* des processus X et U sont dérivées selon les formules

$$p(x_n | y) = \sum_{u_n \in \Lambda} p(v_n | y) \quad (5.21)$$

$$p(u_n | y) = \sum_{x_n \in \Omega} p(v_n | y) \quad (5.22)$$

Enfin, les distributions jointes *a posteriori* $\psi_n(v_n, v_{n+1}) = p(v_n, v_{n+1} | y)$ et $\xi_n(v_n) = p(v_n | y)$ nécessaires pour la phase d'estimation des paramètres sont données par les formules

$$\psi_n(v_n, v_{n+1}) \propto \alpha_n(v_n) a_{u_n, u_{n+1}} p(z_{n+1} | z_n, u_{n+1}) \beta_{n+1}(v_{n+1}) \quad (5.23)$$

$$\xi_n(v_n) = \sum_{v_{n+1}} \psi_n(v_n, v_{n+1}) \quad (5.24)$$

5.1.4. Estimation des paramètres d'une chaîne de Markov couple M-stationnaire

Lorsque les paramètres du modèle $\theta = (\varpi, A, P, \mu, \sigma, \rho)$ ne sont pas connus, il est possible de les estimer en faisant appel à une version adaptée de l'un de ces algorithmes : EM, SEM ou ECI. Dans ce qui suit, nous proposons l'algorithme EM adapté au modèle de CMCo M-S.

La procédure d'estimation se fait de la manière suivante :

- Considérer un jeu initial de paramètres $\theta^0 = (\varpi, A, P, \mu, \sigma, \rho)^0$.
- Pour chaque itération $q + 1$, calculer θ^{q+1} à partir de θ^q en deux étapes jusqu'à ce qu'un critère d'arrêt soit satisfait :
 - o Estimation : Calculer les fonctions ψ_n^q and ξ_n^q à partir de y et θ^q .
 - o Maximisation : Calculer θ^{q+1} selon les formules :

$$\varpi_m^{q+1} = \sum_{x_n} \xi_1^q(v_n) \mathbf{1}_{[u_n = \lambda_m]} \quad (5.24)$$

$$a_{l,m}^{q+1} = \frac{\sum_{n=1}^{N-1} \sum_{v_n, v_{n+1}} \psi_n(v_n, v_{n+1}) \mathbf{1}_{[(u_n, u_{n+1})=(\lambda_l, \lambda_m)]}}{\sum_{n=1}^{N-1} \sum_{v_n} \xi_n(v_n) \mathbf{1}_{[u_n=\lambda_l]}} \quad (5.25)$$

$$(p_m^{i,j})^{q+1} = \frac{\sum_{n=1}^{N-1} \sum_{v_n} \psi_n(v_n, v_{n+1}) \mathbf{1}_{[(u_{n+1}, x_n, x_{n+1})=(\lambda_m, \omega_i, \omega_j)]}}{\sum_{n=1}^{N-1} \sum_{v_n, v_{n+1}} \psi_n(v_n, v_{n+1}) \mathbf{1}_{[(u_{n+1}, x_n)=(\lambda_m, \omega_i)]}} \quad (5.26)$$

$$(\mu_1^{i,j})^{q+1} = \frac{\sum_{n=1}^{N-1} \sum_{v_n, v_{n+1}} \psi_n(v_n, v_{n+1}) y_n \mathbf{1}_{[(x_n, x_{n+1})=(\omega_i, \omega_j)]}}{\sum_{n=1}^{N-1} \sum_{v_n, v_{n+1}} \psi_n(v_n, v_{n+1}) \mathbf{1}_{[(x_n, x_{n+1})=(\omega_i, \omega_j)]}} \quad (5.27)$$

$$(\mu_2^{i,j})^{q+1} = \frac{\sum_{n=1}^{N-1} \sum_{v_n, v_{n+1}} \psi_n(v_n, v_{n+1}) y_{n+1} \mathbf{1}_{[(x_n, x_{n+1})=(\omega_i, \omega_j)]}}{\sum_{n=1}^{N-1} \sum_{v_n, v_{n+1}} \psi_n(v_n, v_{n+1}) \mathbf{1}_{[(x_n, x_{n+1})=(\omega_i, \omega_j)]}} \quad (5.28)$$

$$(\sigma_1^{i,j})^{q+1} = \frac{\sum_{n=1}^{N-1} \sum_{v_n, v_{n+1}} \psi_n(v_n, v_{n+1}) (y_n - \mu_1^{i,j})^2 \mathbf{1}_{[(x_n, x_{n+1})=(\omega_i, \omega_j)]}}{\sum_{n=1}^{N-1} \sum_{v_n, v_{n+1}} \psi_n(v_n, v_{n+1}) \mathbf{1}_{[(x_n, x_{n+1})=(\omega_i, \omega_j)]}} \quad (5.29)$$

$$(\sigma_2^{i,j})^{q+1} = \frac{\sum_{n=1}^{N-1} \sum_{v_n, v_{n+1}} \psi_n(v_n, v_{n+1}) (y_{n+1} - \mu_2^{i,j})^2 \mathbf{1}_{[(x_n, x_{n+1})=(\omega_i, \omega_j)]}}{\sum_{n=1}^{N-1} \sum_{v_n, v_{n+1}} \psi_n(v_n, v_{n+1}) \mathbf{1}_{[(x_n, x_{n+1})=(\omega_i, \omega_j)]}} \quad (5.30)$$

$$(\rho^{i,j})^{q+1} = \frac{\sum_{n=1}^{N-1} \sum_{v_n, v_{n+1}} \psi_n(v_n, v_{n+1}) (y_n - \mu_1^{i,j}) (y_{n+1} - \mu_2^{i,j}) \mathbf{1}_{[(x_n, x_{n+1})=(\omega_i, \omega_j)]}}{(\sigma_1^{i,j})^{q+1} (\sigma_2^{i,j})^{q+1} \sum_{n=1}^{N-1} \sum_{v_n, v_{n+1}} \psi_n(v_n, v_{n+1}) \mathbf{1}_{[(x_n, x_{n+1})=(\omega_i, \omega_j)]}} \quad (5.31)$$

5.1.5. Etude expérimentale

A présent, nous évaluons les performances du modèle proposé à travers des expériences menées sur des images synthétiques et réelles. Plus explicitement, nous proposons deux séries d'expériences. La première série porte sur des données simulées selon le formalisme des CMCo M-S et a pour objectif de comparer les performances de la segmentation MPM accomplie sur la base de paramètres estimés par l'algorithme EM décrit plus haut par rapport à celle basée sur les paramètres réels de simulation. Dans la seconde série d'expérience, nous considérons l'image de classe non-stationnaire du zèbre que l'on bruitera de différentes façons. Ensuite, la restauration MPM est réalisée selon les formalismes des modèles de CMC M-S et CMCo M-S pour évaluer les performances de ce dernier modèle par rapport à son prédécesseur. Rappelons que les images sont converties à partir et vers les séquences monodimensionnelles grâce au parcours d'Hilbert-Peano [Pea1890, Hil1891].

5.1.5.1. Segmentation non-supervisée d'images synthétiques non-stationnaires entachées de bruit corrélé

Soit $T = (U, X, Y)$ avec U à valeurs dans $\Lambda = \{\lambda_1, \lambda_2, \lambda_3\}$ où λ_1, λ_2 et λ_3 correspondent aux couleurs noire, grise et blanche respectivement ; et X à valeurs dans $\Omega = \{\omega_1, \omega_2\}$ où ω_1 et ω_2 correspondent aux couleurs noire et blanches respectivement. Ainsi, nous générons des images de taille 256×256 ($N = 65536$) selon le formalisme des CMCo M-S en utilisant les paramètres suivants :

- Jeu A :

$$A = \frac{1}{10^3} \begin{bmatrix} 998 & 1 & 1 \\ 1 & 998 & 1 \\ 1 & 1 & 998 \end{bmatrix};$$

$$P_1 = \begin{bmatrix} 0.49 & 0.01 \\ 0.01 & 0.49 \end{bmatrix}, P_2 = \begin{bmatrix} 0.39 & 0.11 \\ 0.11 & 0.39 \end{bmatrix}, P_3 = \begin{bmatrix} 0.29 & 0.21 \\ 0.21 & 0.29 \end{bmatrix};$$

$$\mu_1 = \begin{bmatrix} -1.25 & -0.75 \\ 0.75 & 1.25 \end{bmatrix}, \mu_2 = \begin{bmatrix} -1.25 & 0.75 \\ -0.75 & 1.25 \end{bmatrix}, \sigma_1 = \sigma_2 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

Un vecteur de coefficients de corrélation différent a été considéré pour chaque expérience :

- Expérience A.1 : $\rho = \begin{bmatrix} 0.1 & 0.1 \\ 0.1 & 0.1 \end{bmatrix};$

- Expérience A.2 : $\rho = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix};$

- Expérience A.3 : $\rho = \begin{bmatrix} 0.9 & 0.9 \\ 0.9 & 0.9 \end{bmatrix}.$

- Jeu B :

$$A = \frac{1}{10^3} \begin{bmatrix} 998 & 1 & 1 \\ 1 & 998 & 1 \\ 1 & 1 & 998 \end{bmatrix};$$

$$P_1 = \begin{bmatrix} 0.49 & 0.01 \\ 0.01 & 0.49 \end{bmatrix}, P_2 = \begin{bmatrix} 0.39 & 0.11 \\ 0.11 & 0.39 \end{bmatrix}, P_3 = \begin{bmatrix} 0.29 & 0.21 \\ 0.21 & 0.29 \end{bmatrix}.$$

Un jeu de paramètres de bruit différent a été considéré pour chaque expérience :

- Expérience B.1 :

$$\mu_1 = \begin{bmatrix} -5 & -3 \\ 3 & 5 \end{bmatrix}, \mu_2 = \begin{bmatrix} -5 & 3 \\ -3 & 5 \end{bmatrix}, \sigma_1 = \begin{bmatrix} 4 & 4 \\ 8 & 8 \end{bmatrix}, \sigma_2 = \begin{bmatrix} 4 & 8 \\ 4 & 8 \end{bmatrix}, \rho = \begin{bmatrix} 0.99 & 0.99 \\ 0.99 & 0.99 \end{bmatrix}.$$

- Expérience B.2 :

$$\mu_1 = \begin{bmatrix} -5 & -3 \\ 3 & 5 \end{bmatrix}, \mu_2 = \begin{bmatrix} -5 & 3 \\ -3 & 5 \end{bmatrix}, \sigma_1 = \begin{bmatrix} 14 & 7 \\ 9 & 20 \end{bmatrix}, \sigma_2 = \begin{bmatrix} 14 & 9 \\ 7 & 20 \end{bmatrix}, \rho = \begin{bmatrix} 0.99 & 0.99 \\ 0.99 & 0.99 \end{bmatrix}.$$

Pour toutes ces expériences, la restauration est effectuée selon le formalisme des CMCo M-S sur la base des paramètres réels θ utilisés lors de la génération des images et sur la base des paramètres $\hat{\theta}$ estimés via l'algorithme EM. L'objectif est donc d'évaluer l'aptitude du modèle à modéliser les données dans le contexte non-supervisé.

Les différents taux d'erreur moyens obtenus sur 100 expériences sont présentés dans TAB. 5.1. Les résultats de la segmentation de l'expérience A.3 sont également illustrés dans Fig. 5.1.

Tableau 5.1. Taux d'erreur de segmentation des images synthétiques (%)

Expérience	Jeu A			Jeu B	
	Exp. 1	Exp. 2	Exp. 3	Exp. 1	Exp. 2
$\tau_{\theta}(x)$	9.72	9.07	2.69	14.38	6.67
$\tau_{\hat{\theta}}(x)$	9.73	9.11	2.73	14.42	6.70
$\tau_{\theta}(u)$	5.19	6.05	1.77	7.30	5.84
$\tau_{\hat{\theta}}(u)$	5.45	8.40	1.72	8.54	6.05

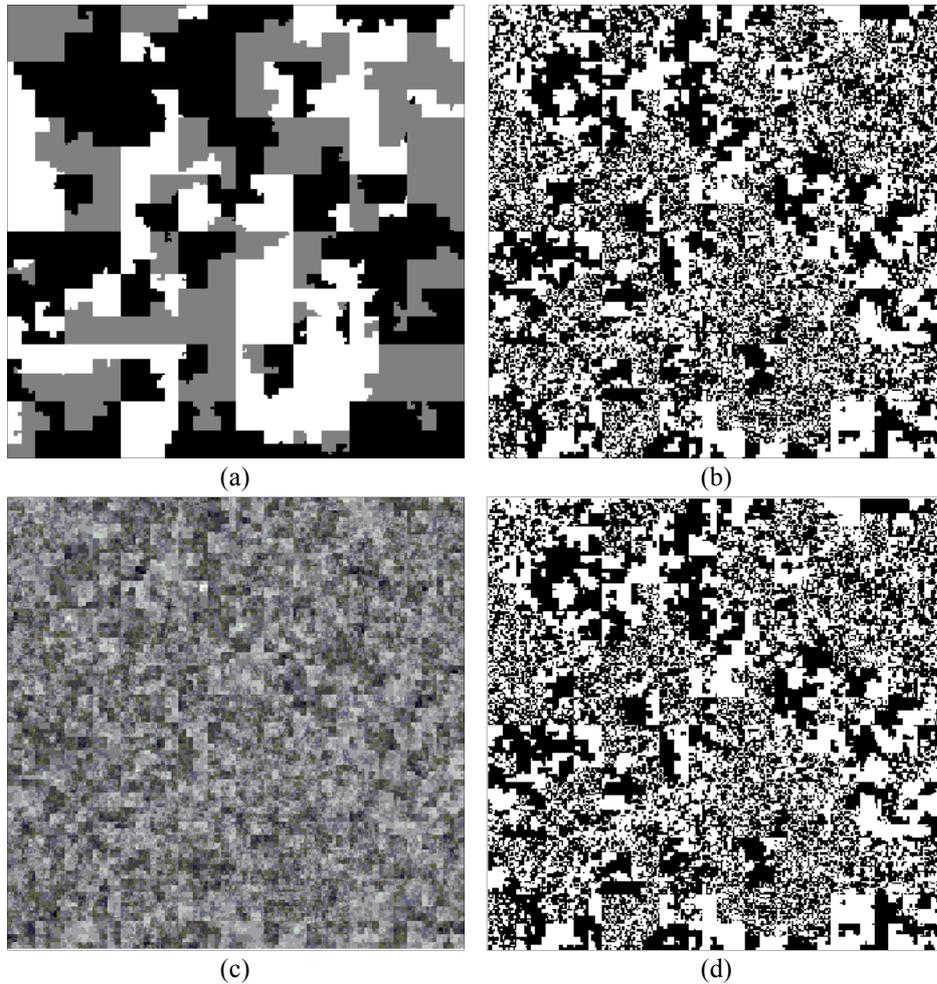


Figure 5.1. Restauration MPM d'une image synthétique non-stationnaire selon le modèle de CMCo M-S. (a) Réalisation du processus auxiliaire $U = u$. (b) Réalisation du processus caché $X = x$. (c) Réalisation du processus observable $Y = y$. (d) Image restaurée selon le modèle de CMCo M-S, taux d'erreur $\tau = 2.7\%$.

A partir des résultats obtenus, nous pouvons émettre les commentaires suivants :

- Conformément à la théorie bayésienne, les taux d'erreur obtenus sur la base des paramètres réels θ sont, en moyenne, les meilleurs possibles. Comme illustré dans TAB.5.I, les taux d'erreur enregistrés par la segmentation non-supervisée sont très proches de ceux obtenus sur la base des paramètres de simulation. Nous pouvons conclure que le modèle de CMCo M-S proposé dans ce premier volet du chapitre est bien adapté pour la modélisation des données dans le contexte non-supervisé. Précisons encore une fois que les paramètres du même modèle peuvent être estimés à travers d'autres procédures que EM telles que SEM ou ECI.

- Une autre remarque très importante est que la corrélation de bruit favorise les modèles couples parce qu'elle est prise en compte par ces derniers. Mieux encore, elle est exploitée comme caractéristique discriminante pour mieux distinguer les différentes classes. En effet, nous pouvons constater que pour les expériences du jeu A où les paramètres du bruit sont identiques sauf pour les paramètres de la corrélation du bruit, les qualités des restaurations sont proportionnelles à la corrélation de bruit.

5.1.5.2. Restauration non-supervisée d'images non-stationnaire entachée de bruit corrélé

Pour cette série d'expérience, nous considérons l'image à deux classes non-stationnaire du zèbre (Fig. 5.2) que nous bruitons de trois manières différentes :

- Expérience 1: l'image est bruitée selon un HMC gaussien classique avec $\mu = [1 \ 3]$ et $\sigma = [1 \ 1]$.
- Expérience 2: un bruit corrélé gaussien est introduit selon les paramètres suivants

$$\mu_1 = \begin{bmatrix} -2 & -1 \\ 1 & 2 \end{bmatrix}, \mu_2 = \begin{bmatrix} -2 & 1 \\ -1 & 2 \end{bmatrix}, \sigma_1 = \sigma_2 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \rho = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}.$$

- Expérience 3: un bruit corrélé gaussien est introduit selon les paramètres suivants

$$\mu_1 = \begin{bmatrix} -2 & -1 \\ 1 & 2 \end{bmatrix}, \mu_2 = \begin{bmatrix} -2 & 1 \\ -1 & 2 \end{bmatrix}, \sigma_1 = \sigma_2 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \rho = \begin{bmatrix} 0.99 & 0.99 \\ 0.99 & 0.99 \end{bmatrix}.$$

Ensuite, la restauration bayésienne MPM est effectuée selon les deux modèles : CMC 3-S et CMCo 3-S. L'objectif de cette série d'expérience est de montrer l'intérêt du modèle proposé par rapport à son devancier. En particulier, dans la première expérience, le bruitage semble s'adapter au modèle classique de CMC M-S ; pour cette expérience, il s'agit de vérifier si le modèle proposé ici est en mesure de rivaliser avec son prédécesseur. Par contre, pour les deux dernières expériences, le bruitage semble favoriser le modèle de CMCo M-S grâce à la corrélation de bruit injectée dans l'image, notamment lors de la dernière expérience.

Les taux d'erreurs de restauration des trois expériences sont présentés dans TAB.5.II. Les résultats de restauration des expériences 2 et 3 sont également schématisés dans Fig. 5.2 et 5.3 respectivement.

Tableau 5.2. Taux d'erreur de segmentation des images bruitées (%)

$\tau_{\hat{\theta}}(x)$	Jeu C		
	Exp. 1	Exp. 2	Exp. 3
CMC M-S	3.42	3.39	22.68
CMCo M-S	3.43	2.23	0

Sur la base des résultats obtenus, nous pouvons conclure ce qui suit :

- Lorsque le bruitage est effectué conformément à la théorie des CMC M-S (expérience 1), les taux d'erreurs sont équivalents. Autrement, lorsqu'un bruit corrélé est injecté dans la donnée en question (expérience 2 et 3), le modèle de CMCo M-S est plus performant que son devancier. Nous pouvons conclure que le modèle proposé généralise effectivement le modèle de CMC M-S.

- Les résultats de l'expérience 3 où l'image d'origine $X = x$ a parfaitement été restaurée (avec un taux d'erreur strictement nul) confirment que la corrélation du bruit peut être vue comme un atout surprenant pour les modèles couples en général et le modèle de CMCo M-S en particulier. Ce même constat a été fait dans la première série d'expérience (jeu A).

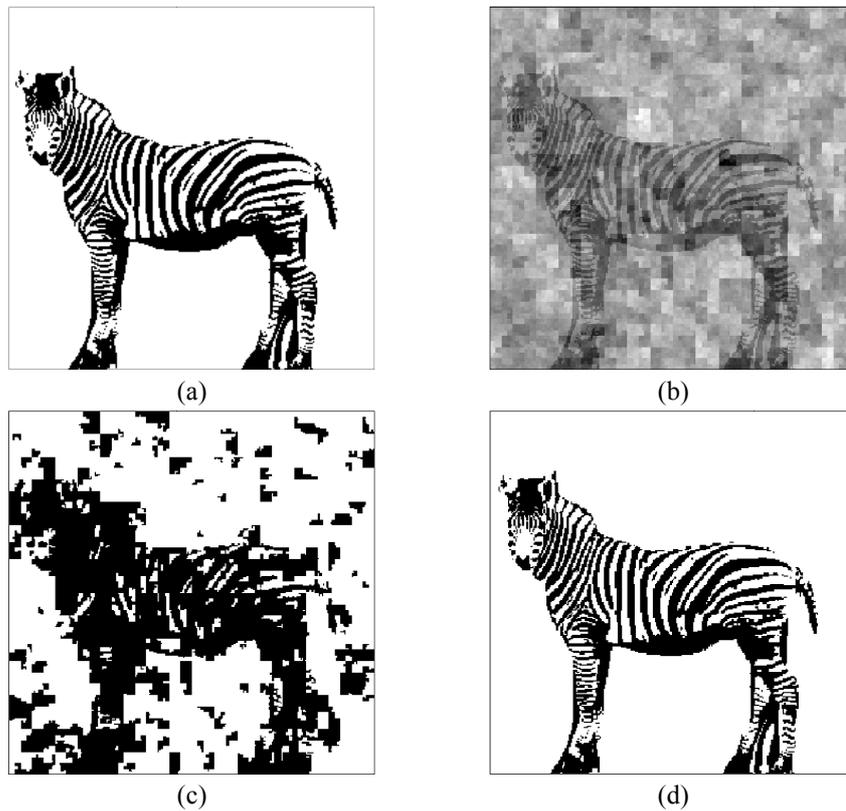


Figure 5.2. Restauration MPM de l'image de classe non-stationnaire du zèbre selon les modèles CMC M-S et CMC Co M-S. (a) Image de classe $X = x$. (b) Image bruitée $Y = y$ avec un bruit fortement corrélé. (c) Image restaurée selon le modèle de CMC M-S, taux d'erreur $\tau = 22.7\%$. (d) Image restaurée selon le modèle de CMC Co M-S, taux d'erreur $\tau = 0\%$.

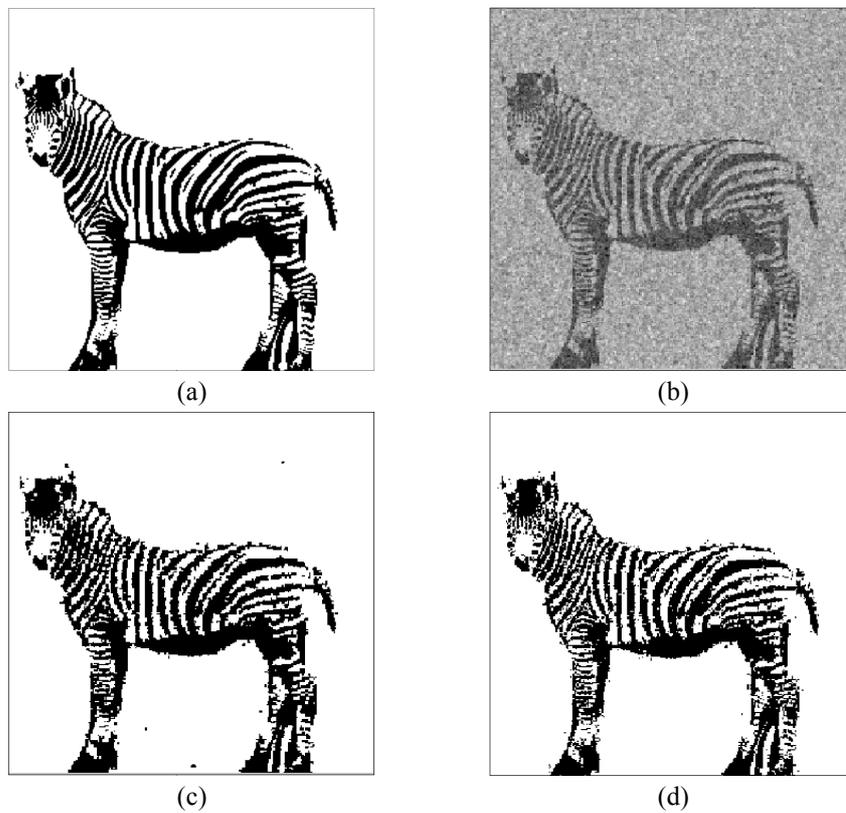


Figure 5.3. Restauration MPM de l'image de classe non-stationnaire du zèbre selon les modèles CMC M-S et CMC Co M-S. (a) Image de classe $X = x$. (b) Image bruitée $Y = y$ avec un bruit faiblement corrélé. (c) Image restaurée selon le modèle de CMC M-S, taux d'erreur $\tau = 3.39\%$. (d) Image restaurée selon le modèle de CMC Co M-S, taux d'erreur $\tau = 2.23\%$.

5.2. Chaîne de Markov couples MM-stationnaire

A présent, nous définissons un autre modèle de CMCo non-stationnaire qui diffère du précédent par le fait que les paramètres de bruit sont également liés à la réalisation du processus auxiliaire. Une telle propriété est très utile pour modéliser par exemples les images texturées comme nous allons le voir plus loin.

5.2.1. Définition du modèle

Soit $Y = (Y_n)_{n=1}^N$ une séquence observable que l'on doit indexer en $X = (X_n)_{n=1}^N$ où chaque Y_n est dans \mathbb{R} et X_n prend ses valeurs à partir de l'ensemble des classes $\Omega = \{\omega_1, \dots, \omega_K\}$. Nous nous intéressons à la situation où $Z = (X, Y)$ est une chaîne de Markov couple dont les paramètres dépendent de la réalisation d'un troisième processus markovien $U = (U_n)_{n=1}^N$ où chaque U_n prend ses valeurs à partir de l'ensemble fini $\Lambda = \{\lambda_1, \dots, \lambda_M\}$.

Soit $T = (U, X, Y) = (U, Z)$ un processus triplet dont la probabilité de transition est donnée par

$$p(z_n, u_n | z_{n-1}, u_{n-1}) = p(u_n | u_{n-1})p(x_n | x_{n-1}, y_{n-1}, u_n)p(y_n | x_{n-1}, y_{n-1}, x_n, u_n) \quad (5.32)$$

Cette probabilité définit un modèle de chaîne de Markov triplet que nous appellerons chaîne de Markov couple MM-stationnaire (CMCo MM-S) qui étend ainsi le modèle défini plus haut dans la mesure où il lie les distributions de bruit à la réalisation du processus auxiliaire. Ce modèle peut être appliqué à toutes les situations où le signal présente plusieurs portions hétérogènes (régimes) pourvu que chacune d'elle puisse être modélisée via une CMCo. L'intérêt du modèle proposé réside dans le fait que les transitions entre ces régimes sont modélisées via une chaîne de Markov (loi du processus U). Le modèle de CMCo peut être vu comme un CMCo MM-S où $M = 1$.

5.2.2. Paramètres d'une CMCo MM-S

Selon le formalisme des CMCo MM-S, pour $\lambda_m \in \Lambda$ fixé, la distribution $p(z_{n-1}, z_n | U_n = \lambda_m)$ est régie par une CMCo homogène (dont les paramètres ne dépendent pas de la position du site n).

Nous adoptons les notations suivantes

$$p_m(i, j) = p(X_{n-1} = \omega_i, X_n = \omega_j | U_n = \lambda_m) \quad (5.33)$$

$$f_m^{i,j}(y_{n-1}, y_n) = p(y_{n-1}, y_n | X_{n-1} = \omega_i, X_n = \omega_j, U_n = \lambda_m) \quad (5.34)$$

La loi de Z est donc donnée par

$$p(z_{n-1}, z_n | U_n = \lambda_m) = p_m(i, j) f_m^{i,j}(y_{n-1}, y_n) \quad (5.35)$$

La loi de Z peut s'exprimer de façon équivalente par les lois de probabilités initiales $p(z_1 | U_2 = \lambda_m)$ données par

$$p(z_1 | U_2 = \lambda_m) = \sum_{\omega_j \in \Omega} p_m(i, j) \int_{\mathbb{R}} f_m^{i,j}(y_1, y_2) dy_2 = \sum_{\omega_j \in \Omega} p_m(i, j) f_m^{i,j}(y_1) \quad (5.36)$$

et la matrice de transition donnée par

$$p(z_n|z_{n-1}, U_n = \lambda_m) = \frac{p(z_{n-1}, z_n|U_n = \lambda_m)}{p(z_{n-1}|U_n = \lambda_m)} = \frac{p_m(i,j)f_m^{i,j}(y_{n-1}, y_n)}{\sum_{\omega_j \in \Omega} p_m(i,j)f_{i,j}(y_{n-1})} \quad (5.37)$$

Dans cette étude, seules les CMCo MM-S gaussiennes seront considérées. Une CMCo MM-S est qualifiée de gaussiennes si toutes ses densités $f_m^{i,j}$ sont gaussiennes. Ainsi, une CMCo MM-S gaussienne est entièrement définie à travers les paramètres suivants : le vecteur de probabilités initiales $\varpi = [\varpi_m]_{\lambda_m \in \Lambda}$ où $\varpi_m = p(u_n = \lambda_m)$, la matrice de transition $A = (a_{l,m})$ où $a_{\lambda_l, \lambda_m} = p(U_n = \lambda_m | U_{n-1} = \lambda_l)$, les M matrices $P_m = (p_m^{i,j})$ où $p_m^{i,j} = p_m(i,j)$, les moyennes $\mu_{m,1}^{i,j}, \mu_{m,2}^{i,j}$, les écarts-type $\sigma_{m,1}^{i,j}, \sigma_{m,2}^{i,j}$ et les coefficients de corrélation $\rho_m^{i,j}$ des $M \times K^2$ densités bidimensionnelles $f_{i,j}$.

5.2.3. Restauration MPM d'une chaîne de Markov couple MM-stationnaire

Soit $T = (U, X, Y)$ une CMCo MM-S définie par la formule (5.32). Notons $V = (U, X)$. $T = (V, Y)$ est donc une CMCo. Par conséquent, les probabilités marginales *a posteriori* $p(v_n|y) = p(u_n, x_n|y)$ nécessaires pour la restauration MPM sont calculables.

Lorsque les paramètres du modèle sont connus, les distributions *a posteriori* $p(u_n, x_n|y)$ sont calculables à travers les fonctions progressives généralisées $\alpha_n(v_n) = p(y_1, \dots, y_n, v_n)$ et les fonctions rétrogrades $\beta_n(x_n) = p(y_{n+1}, \dots, y_N | v_n, y_n)$ et dont le calcul se fait de la manière itérative suivante :

Initialisation :

$$\alpha_1(v_1) = p(v_1, y_1) \quad (5.38)$$

$$\beta_N(v_N) = 1 \quad (5.39)$$

Induction :

$$\alpha_n(v_n) = \sum_{v_{n-1} \in \Lambda \times \Omega} \alpha_{n-1}(v_{n-1}) p(u_n | u_{n-1}) p(z_n | z_{n-1}, u_n) \quad (5.40)$$

$$\beta_n(v_n) = \sum_{v_{n+1} \in \Lambda \times \Omega} \beta_{n+1}(v_{n+1}) p(u_n | u_{n-1}) p(z_{n+1} | z_n, u_{n+1}) \quad (5.41)$$

Les distributions *a posteriori* $p(v_n|y)$ sont ensuite calculées comme suit

$$p(v_n|y) \propto \alpha_n(v_n) \beta_n(v_n) \quad (5.42)$$

Puis, les distributions marginales *a posteriori* des processus X et U sont dérivées selon les formules

$$p(x_n|y) = \sum_{u_n \in \Lambda} p(v_n|y) \quad (5.42)$$

$$p(u_n|y) = \sum_{x_n \in \Omega} p(v_n|y) \quad (5.43)$$

Enfin, les distributions jointes *a posteriori* $\psi_n(v_n, v_{n+1}) = p(v_n, v_{n+1}|y)$ et $\xi_n(v_n) = p(v_n|y)$ nécessaires pour la phase d'estimation des paramètres sont données par les formules

$$\psi_n(v_n, v_{n+1}) \propto \alpha_n(v_n) a_{u_n, u_{n+1}} p(z_{n+1} | z_n, u_{n+1}) \beta_{n+1}(v_{n+1}) \quad (5.44)$$

$$\xi_n(v_n) = \sum_{v_{n+1}} \psi_n(v_n, v_{n+1}) \quad (5.45)$$

5.2.4. Estimation des paramètres d'une chaîne de Markov couple MM-stationnaire

Lorsque les paramètres du modèle $\Theta = (\varpi, A, P, \mu, \sigma, \rho)$ ne sont pas connus, nous proposons de les estimer en faisant appel à un algorithme du type EM adapté au modèle de CMCo MM-S.

La procédure d'estimation se fait de la manière suivante :

- Considérer un jeu initial de paramètres $\Theta^0 = (\varpi, A, P, \mu, \sigma, \rho)^0$.
- Pour chaque itération $q + 1$, calculer Θ^{q+1} à partir de Θ^q en deux étapes jusqu'à ce qu'un critère d'arrêt soit satisfait :
 - o Estimation : Calculer les fonctions ψ_n^q and ξ_n^q à partir de y et Θ^q .
 - o Maximisation : Calculer Θ^{q+1} selon les formules

$$\varpi_m^{q+1} = \sum_{x_n} \xi_1(v_n) 1_{[u_n=\lambda_m]} \quad (4.45)$$

$$a_{l,m}^{q+1} = \frac{\sum_{n=1}^{N-1} \sum_{v_n} \psi_n(v_n, v_{n+1}) 1_{[(u_n, u_{n+1})=(\lambda_l, \lambda_m)]}}{\sum_{n=1}^{N-1} \sum_{v_n} \xi_n(v_n) 1_{[u_n=\lambda_l]}} \quad (4.46)$$

$$(p_m^{i,j})^{q+1} = \frac{\sum_{n=1}^{N-1} \sum_{v_n} \psi_n(v_n, v_{n+1}) 1_{[(u_{n+1}, x_n, x_{n+1})=(\lambda_m, \omega_i, \omega_j)]}}{\sum_{n=1}^{N-1} \sum_{v_n} \xi_n(v_n) 1_{[u_{n+1}=\lambda_m]}} \quad (4.46)$$

$$(\mu_{m,1}^{i,j})^{q+1} = \frac{\sum_{n=1}^{N-1} \sum_{v_n, v_{n+1}} \psi_n(v_n, v_{n+1}) y_n 1_{[(u_n, x_n, x_{n+1})=(\lambda_m, \omega_i, \omega_j)]}}{\sum_{n=1}^{N-1} \sum_{v_n, v_{n+1}} \psi_n(v_n, v_{n+1}) 1_{[(u_n, x_n, x_{n+1})=(\lambda_m, \omega_i, \omega_j)]}} \quad (4.47)$$

$$(\mu_{m,2}^{i,j})^{q+1} = \frac{\sum_{n=1}^{N-1} \sum_{v_n, v_{n+1}} \psi_n(v_n, v_{n+1}) y_{n+1} 1_{[(u_n, x_n, x_{n+1})=(\lambda_m, \omega_i, \omega_j)]}}{\sum_{n=1}^{N-1} \sum_{v_n, v_{n+1}} \psi_n(v_n, v_{n+1}) 1_{[(u_n, x_n, x_{n+1})=(\lambda_m, \omega_i, \omega_j)]}} \quad (4.48)$$

$$(\sigma_{m,1}^{i,j})^{q+1} = \frac{\sum_{n=1}^{N-1} \sum_{v_n, v_{n+1}} \psi_n(v_n, v_{n+1}) (y_n - (\mu_{m,1}^{i,j})^{q+1})^2 1_{[(u_n, x_n, x_{n+1})=(\lambda_m, \omega_i, \omega_j)]}}{\sum_{n=1}^{N-1} \sum_{v_n, v_{n+1}} \psi_n(v_n, v_{n+1}) 1_{[(u_n, x_n, x_{n+1})=(\lambda_m, \omega_i, \omega_j)]}} \quad (4.49)$$

$$(\sigma_{m,2}^{i,j})^{q+1} = \frac{\sum_{n=1}^{N-1} \sum_{v_n, v_{n+1}} \psi_n(v_n, v_{n+1}) (y_{n+1} - (\mu_{m,2}^{i,j})^{q+1})^2 1_{[(u_n, x_n, x_{n+1})=(\lambda_m, \omega_i, \omega_j)]}}{\sum_{n=1}^{N-1} \sum_{v_n, v_{n+1}} \psi_n(v_n, v_{n+1}) 1_{[(u_n, x_n, x_{n+1})=(\lambda_m, \omega_i, \omega_j)]}} \quad (4.50)$$

$$(\rho_m^{i,j})^{q+1} = \frac{\sum_{n=1}^{N-1} \sum_{v_n, v_{n+1}} \psi_n(v_n, v_{n+1}) (y_n - (\mu_{m,1}^{i,j})^{q+1}) (y_{n+1} - (\mu_{m,2}^{i,j})^{q+1}) 1_{[(u_n, x_n, x_{n+1})=(\lambda_m, \omega_i, \omega_j)]}}{\sum_{n=1}^{N-1} \sum_{v_n, v_{n+1}} \psi_n(v_n, v_{n+1}) 1_{[(u_n, x_n, x_{n+1})=(\lambda_m, \omega_i, \omega_j)]}} \quad (4.51)$$

5.2.5. Application des CMCo MM-S au problème de segmentation des textures

Dans cette section, nous appliquons le modèle de CMCo MM-S au problème de la segmentation des images de textures. Pour ce faire, nous considérons le cas le plus général soit quand l'image comporte plus d'une texture. Soit M le nombre de classes de textures présentes dans l'image. Conformément au principe des modèles M-stationnaire, nous préconiserons un

modèle homogène (caché ou couple) pour chaque classe de texture. En effet, pour de tels modèles, toute portion de l'image (présentant la même texture) est considérée comme une donnée stationnaire pour laquelle un seul modèle sera assigné. Le passage d'une texture à une autre est régi, quant à lui, par une chaîne de Markov.

La présente étude expérimentale comporte trois séries d'expériences. La première série porte sur des images de textures simulées. Dans la seconde série, nous appliquons un filtre moyen sur une portion d'une image texturée afin d'y injecter un bruit corrélé pour évaluer l'aptitude du modèle de CMCo MM-S à détecter la différence entre les deux portions de l'image (présentant la même texture et ne différant que par la corrélation entre les pixels). La dernière série, quant à elle, portera sur une image assemblée de quatre textures différentes. Rappelons que les données bidimensionnelles sont converties à partir et vers les séquences monodimensionnelles grâce au parcours d'Hilbert-Peano [Pea1890, Hil1891].

5.2.5.1. Segmentation non-supervisée des images synthétiques de textures

Dans cette série d'expériences portant sur des images synthétiques de textures, la segmentation est effectuée à travers la stratégie MPM d'une manière non-supervisée via l'algorithme EM selon les différents modèles markoviens : CMC, CMCo, CMC M-S, CMCo MM-S.

Cette série comporte trois expériences. Pour les différentes simulations, nous considérons deux classes réelles $\Omega = \{\omega_1, \omega_2\}$ et trois classes auxiliaires $\Lambda = \{\lambda_1, \lambda_2, \lambda_3\}$. Les images de textures de taille 128×128 sont générées comme suit :

- La réalisation du processus auxiliaire $U = u$ (représentant la vérité terrain) est déterministe : les pixels du quart gauche supérieur sont assignés à λ_1 (noir), les pixels du quart droit supérieur sont assignés à λ_2 (gris), les pixels de la moitié inférieure sont assignés à λ_3 (blanc).
- Les réalisations $X = x$ et $Y = y$ sont simulées à partir de $\Omega = \{\omega_1, \omega_2\}$ et de \mathbb{R} selon les matrices $(\Gamma_m)_{m=1}^3$ et les paramètres de bruit illustrés dans TAB. 5.III.

Pour chaque expérience, 100 images sont générées. Les résultats de segmentation sont résumés dans TAB. 5.IV. Les résultats de l'expérience 2 sont également illustrés dans Fig. 5.4.

Tableau 5.3. Paramètres des modèles de CMCo MM-S utilisés dans la simulation des images synthétiques de textures

Expérience	Γ	μ_1	μ_2	σ_1	σ_2	ρ
λ_1	[0.49 0.01; 0.01 0.49]	[1 1; 3 3]	[1 3; 1 3]	[1 1; 1 1]	[1 1; 1 1]	[0 0; 0 0]
λ_2	[0.25 0.25; 0.25 0.25]	[1 1; 3 3]	[1 3; 1 3]	[1 1; 1 1]	[1 1; 1 1]	[0 0; 0 0]
λ_3	[0.01 0.49; 0.49 0.01]	[1 1; 3 3]	[1 3; 1 3]	[1 1; 1 1]	[1 1; 1 1]	[0 0; 0 0]
Expérience :	Γ	μ_1	μ_2	σ_1	σ_2	ρ
λ_1	[0.49 0.01; 0.01 0.49]	[1 1; 3 3]	[1 3; 1 3]	[1 1; 1 1]	[1 1; 1 1]	[0.1 0.1; 0.1 0.1]
λ_2	[0.25 0.25; 0.25 0.25]	[1 1; 3 3]	[1 3; 1 3]	[1 1; 1 1]	[1 1; 1 1]	[0.5 0.5; 0.5 0.5]
λ_3	[0.01 0.49; 0.49 0.01]	[1 1; 3 3]	[1 3; 1 3]	[1 1; 1 1]	[1 1; 1 1]	[0.9 0.9; 0.9 0.9]
Expérience :	Γ	μ_1	μ_2	σ_1	σ_2	ρ
λ_1	[0.49 0.01; 0.01 0.49]	[1 1; 3 3]	[1 3; 1 3]	[1 1; 1 1]	[1 1; 1 1]	[0.5 0.5; 0.5 0.5]
λ_2	[0.25 0.25; 0.25 0.25]	[2 2; 4 4]	[2 4; 2 4]	[1 1; 1 1]	[1 1; 1 1]	[0.5 0.5; 0.5 0.5]
λ_3	[0.01 0.49; 0.49 0.01]	[0 0; 2 2]	[0 2; 0 2]	[1 1; 1 1]	[1 1; 1 1]	[0.5 0.5; 0.5 0.5]

Tableau 5.4. Taux moyens d'erreurs de la segmentation des images synthétiques de textures selon les différents modèles

Expérience		K-means	CMC	CMCo	CMC M-S	CMCo MM-S
1	T _X	15.26	22.22	37.81	4.58	4.61
	T _U	-	-	-	0.00	0.00
2	T _X	15.93	34.24	22.11	10.95	3.58
	T _U	-	-	-	10.96	0.00
3	T _X	22.75	38.51	40.34	17.03	13.25
	T _U	-	-	-	3.83	0.00

L'intérêt de la première expérience est de vérifier si le modèle de CMCo MM-S parvient à rivaliser avec le modèle de CMC M-S lorsque les données sont simulées suivant un processus favorisant ce dernier.

Dans la deuxième expérience, les textures diffèrent de par les matrices de transitions et les coefficients de corrélation considérés. L'intérêt de cette expérience est de mesurer l'impact de la prise en compte de la corrélation sur la qualité de la segmentation. En effet, seule la corrélation au sein de chaque texture pourrait privilégier le modèle proposé par rapport à son devancier. Il convient de noter que les images ici sont générées suivant la méthode de simulation des CMCo M-S définie dans la section 5.1.

Dans la troisième expérience, les paramètres de bruit des trois textures sont différents. Dans ce cas, les données sont générées selon une CMCo MM-S et ce dernier est donc censé fournir les meilleurs résultats.

L'analyse des résultats présentés dans TAB. 5.IV nous permet de dégager les conclusions suivantes :

- Pour la première expérience, où les données favorables aux CMCo M-S classiques, les taux d'erreur de segmentation sont équivalents. Pour les autres expériences, le nouveau modèle fournit de meilleurs résultats grâce à la prise en compte de la corrélation (expérience 2) et les différences des paramètres de bruit entre les différentes textures (expérience 3). Ceci affirme que le nouveau modèle étend effectivement son prédécesseur.

- Pour les trois expériences, le modèle proposé fournit des segmentations quasi parfaites (avec un taux d'erreur proche de 0 %) surpassant son prédécesseur qui échoue en présence de bruit corrélé (expérience 2) ou lorsque les densités de bruit des différentes textures ne sont pas identiques (expérience 3).

- Pour ce type particulier de données (non-stationnaire), le *clustering* K-means peut s'avérer plus performant que les CMC et les CMCo classiques. Ces derniers étant des outils de régularisation, cette régularisation peut dérouter les différents traitements en considérant les données stationnaires.

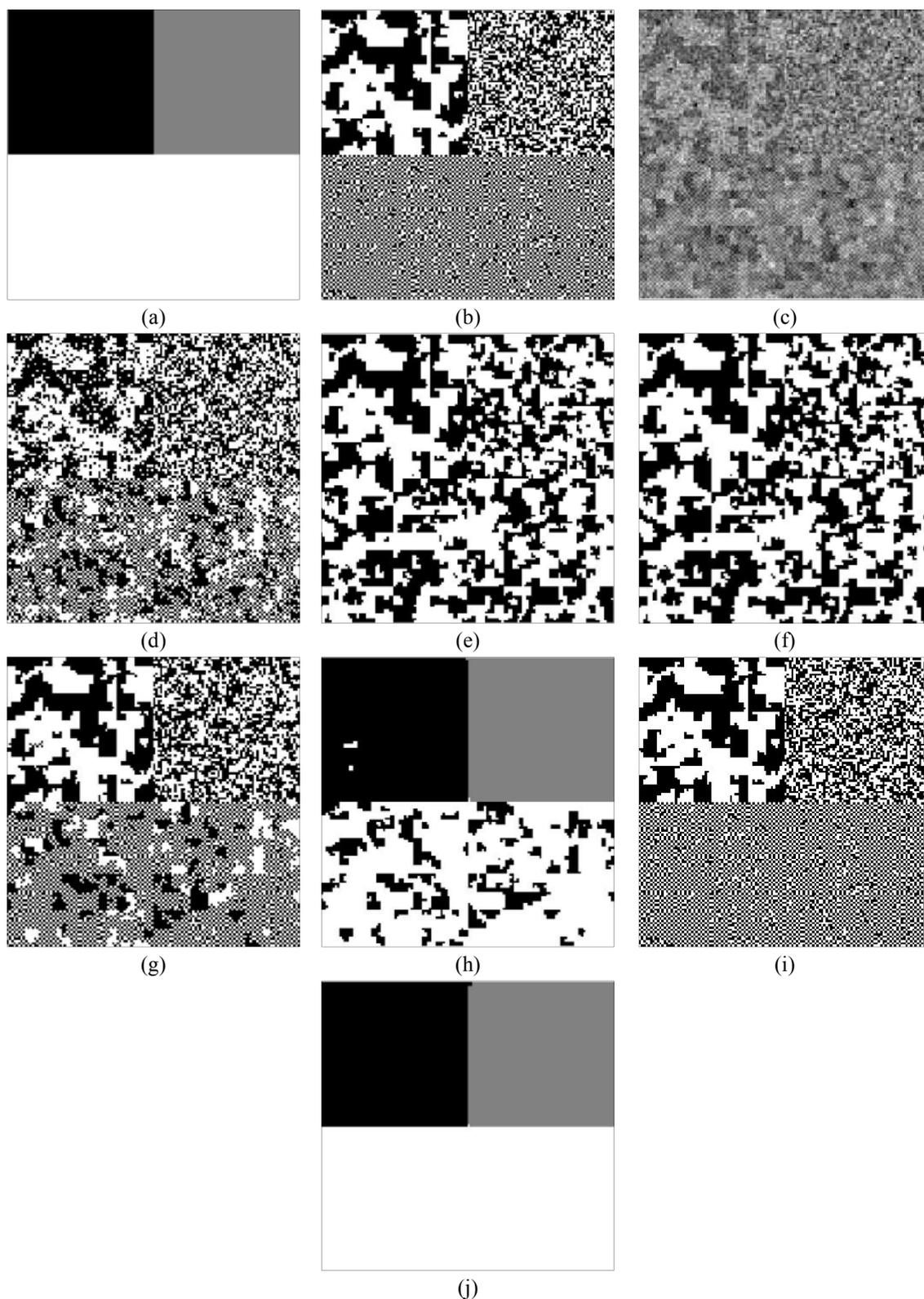


Figure 5.4. Segmentation non-supervisée de l'image texturée synthétique de l'expérience 2. (a) Vérité terrain $U = u$. (b) Image de classe d'origine $X = x$. (c) Image observée $Y = y$. (d) Segmentation en 2 classes via *K-means*, taux d'erreur $\tau = 16\%$. (e) Segmentation en 2 classes selon une CMC, taux d'erreur $\tau = 34\%$. (f) Segmentation en 3 classe selon une CMC, taux d'erreur $\tau = 22\%$. (g) Segmentation en 2 classes selon une CMC M-S, taux d'erreur $\tau = 11\%$. (h) Estimation de la réalisation du processus auxiliaire U selon une CMC MM-S, taux d'erreur $\tau = 11\%$. (i) Segmentation en 2 classes selon une CMC MM-S, taux d'erreur $\tau = 3.5\%$. (j) Estimation de la réalisation du processus auxiliaire U selon une CMC MM-S, taux d'erreur $\tau \approx 0\%$.

5.2.5.2. Segmentation non-supervisée des images de textures entachées d'un filtre moyen

Dans cette expérience, nous effectuons une comparaison entre les CMCo MM-S et leurs devancières les CMC M-S. Pour cela, considérons l'image de texture D22 tirée de la très populaire base d'images texturées Brodatz [Bro66]. Une sous-image de taille 256×256 est extraite de cette image. Puis, nous appliquons un filtre moyen à sa moitié supérieure, ce qui fournit deux parties homogènes (Fig. 5.5).

Ensuite, nous effectuons la segmentation non-supervisée basée sur la stratégie bayésienne MPM à partir des paramètres estimés via l'algorithme EM conformément aux formalismes de CMC M-S et CMCo MM-S. Pour les deux modèles, nous posons $\Omega = \{\omega_1, \omega_2\}$ et $\Lambda = \{\lambda_1, \lambda_2\}$. Les paramètres des deux modèles sont initialisés via l'algorithme *K-means*. Les résultats obtenus sont illustrés dans Fig. 5.5.

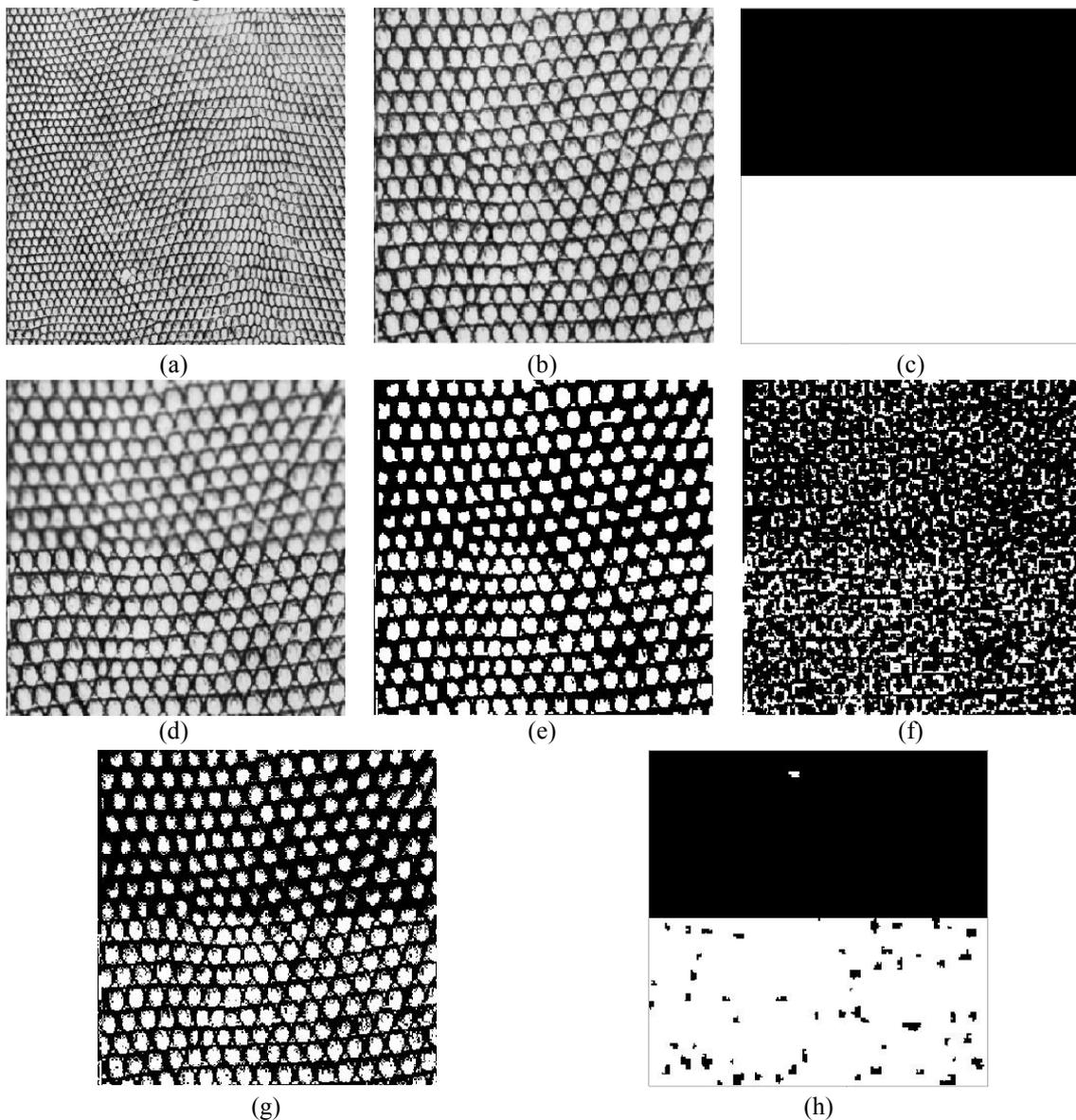


Figure 5.5. Segmentation non-supervisée de l'image texturée D22 selon les formalismes de CMC M-S et CMCo MM-S. (a) Image réelle D22. (b) Image extraite de taille 256×256 . (c) vérité terrain $U = u$. (d) Image observée $Y = y$. (e) Segmentation en 2 classes selon une CMC M-S. (f) Estimation de la réalisation du processus auxiliaire U selon une CMC M-S, taux d'erreur $\tau = 45\%$. (g) Segmentation en 2 classes selon une CMCo MM-S. (h) Estimation de la réalisation du processus auxiliaire U selon une CMCo MM-S, taux d'erreur $\tau = 2\%$.

Sur la base des résultats obtenus, nous pouvons conclure :

- Pour le modèle de CMC M-S, les deux textures sont identiques. En effet, ne prenant pas en compte la corrélation entre les pixels de l'image, ce modèle ne parvient pas à distinguer les deux textures qui ne diffèrent que par la corrélation due au filtre moyen appliqué à la partie supérieure de l'image.

- La corrélation au sein des données est une caractéristique discriminante qui peut se révéler cruciale. Effectivement, cette particularité a permis au modèle de CMCo MM-S de différencier deux textures très semblables en enregistrant un taux d'erreur de seulement 2%.

5.2.5.3. Segmentation non-supervisée d'images réelles de textures

Dans cette dernière expérience, nous considérons une image assemblée (Fig. 5.6) en combinant quatre textures différentes (avec quelques similitudes) tirées de la base CGT [CGT11]. La segmentation est effectuée selon la stratégie bayésienne MPM sur la base des paramètres estimés via l'algorithme EM selon les formalismes des modèles CMC M-S et CMCo MM-S avec $\Omega = \{\omega_1, \omega_2, \omega_3\}$ et $\Lambda = \{\lambda_1, \lambda_2, \lambda_3, \lambda_4\}$. Cette expérience constitue un test réel pour les deux modèles dans la mesure où l'on ignore complètement la nature des données (ne suivant aucun modèle particulier).

Les résultats obtenus sont schématisés dans Fig. 5.6.

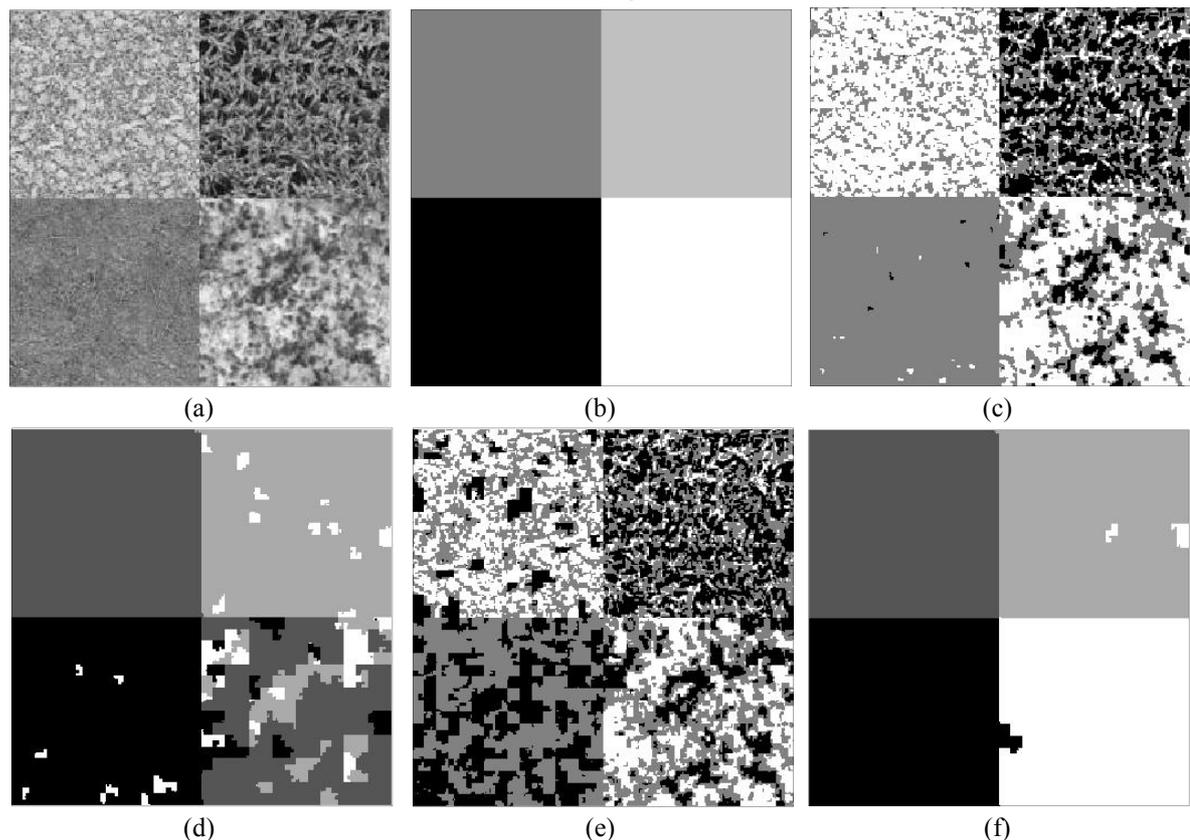


Figure 5.6. Segmentation non-supervisée d'une image assemblée de 4 textures réelles selon les formalismes de CMC M-S et CMCo MM-S. (a) Image réelle D22. (b) vérité terrain $U = u$. (c) Estimation de la réalisation du processus caché X selon une CMC M-S. (d) Segmentation de l'image en 4 textures selon une CMCo MM-S, taux d'erreur $\tau = 24\%$. (e) Estimation de la réalisation du processus caché X selon une CMCo MM-S. (f) Segmentation de l'image en 4 textures selon une CMCo MM-S, taux d'erreur $\tau = 0.65\%$.

A partir de ces derniers, nous pouvons émettre les remarques suivantes :

- Le modèle de CMCo MM-S permet de distinguer les quatre textures en enregistrant un taux d'erreur de 2%.
- Le modèle de CMC M-S, quant à lui, confond les textures λ_2 (gris foncé) et λ_4 (blanc). Ceci est dû au fait que ces deux dernières présentent des aspects visuels similaires. Ceci démontre encore une fois l'importance de la corrélation du bruit qui constitue une caractéristique complémentaire primordiale pour discriminer les différentes classes auxiliaires.

5.3. Conclusion

Dans ce chapitre, nous avons proposé deux modèles : le modèle de CMCo M-stationnaire et le modèle de CMCo MM-stationnaire qui étendent, tous les deux, le modèle de CMC M-stationnaire. Pour chacun des deux modèles, nous avons présenté les différentes procédures d'inférence. L'étude expérimentale a établi l'intérêt des modèles proposés par rapport aux modèles courants.

Une piste prometteuse de travail futur consiste à étendre les présentes modélisations au contexte des champs de Markov, ce qui permettrait de mieux prendre en compte l'aspect bidimensionnel des images.

Chapitre 6. Chaînes de Markov couples évidentielles

Les modèles de chaînes de Markov cachés ont été largement utilisés pour la résolution de problèmes inverses survenant en traitement du signal et de l'image. Comme nous l'avons vu dans le chapitre 2, ces modèles ont été généralisés aux chaînes de Markov couples, lesquelles présentent une complexité de calcul comparable, avec plus de potentiel de modélisation, notamment pour les données à bruit corrélé. Pour être applicables dans le contexte de segmentation non-supervisée, les deux modèles supposent les données en question stationnaires. Lorsque ces dernières le sont effectivement, les résultats obtenus sont satisfaisants grâce aux techniques bayésiennes telles que MPM et MAP. Toutefois, quand les données concernées par les traitements sont non-stationnaires, les modélisations classiques ne parviennent pas à établir un lien convenable avec ces données et, par conséquent, les résultats correspondants, sont moyens voire inadmissibles. Une manière intéressante de remédier à ce problème est l'emploi de la théorie de l'évidence de Dempster-Shafer en introduisant une fonction de masse pour modéliser le manque de précision (fluctuation dans ce cas) des probabilités *a priori* des données cachées. Le recours à cette théorie a fait ses preuves dans le cadre des chaînes de Markov cachées. En effet, il a été démontré que le gain en termes de performances des chaînes de Markov évidentielles est significatif par rapport aux modèles conventionnels [Lan05]. L'objectif de ce chapitre est d'introduire la théorie de Dempster-Shafer dans le contexte des chaînes de Markov couples afin de bénéficier de l'apport de ces modélisations plus générales pour le traitement des données non-stationnaires entachées de bruit corrélé. Nous montrons que la restauration bayésienne demeure faisable grâce au formalisme des modèles de Markov triplets. Nous présentons également les algorithmes d'estimation des paramètres correspondant au cas non-supervisé. Enfin, le nouveau modèle évidentiel est validé sur des données synthétiques et des images réelles bruitées.

6.1. Définition du modèle

Soit $Z = (X, Y)$ une chaîne de Markov couple. Nous avons

$$p(z_n, z_{n+1}) = p(x_n, x_{n+1})p(y_n, y_{n+1}|x_n, x_{n+1}) \quad (6.1)$$

Le modèle de chaîne de Markov couple stationnaire, défini dans le chapitre 3, suppose que les distributions *a priori* ne dépendent pas de n , et par conséquent, le modèle est entièrement défini par la distribution

$$p(z_1, z_2) = p(i, j)f_{i,j}(y_1, y_2) \quad (6.2)$$

où $p(i, j)$ est une probabilité sur Ω^2 et $f_{i,j}(y_1, y_2)$ est une distribution dans \mathbb{R}^2 .

Dans le cas où cette hypothèse n'est pas vérifiée, soit lorsque les distributions $p(x_n, x_{n+1})$ dépendent de n , la formule (6.2) devient :

$$p(z_n, z_{n+1}) = p_n(i, j)f_{i,j}(y_n, y_{n+1}) \quad (6.3)$$

L'application du modèle de chaîne de Markov couple conventionnel à de telles données, en les considérant stationnaires conduit, dans le contexte non-supervisé à l'estimation d'un paramètre commun $p^*(i, j)$ pour les distributions $p_n(i, j)$.

Comme nous l'avons signalé en introduction, une autre alternative est de faire appel à la théorie de l'évidence de Dempster-Shafer comme dans le contexte des chaînes de Markov cachées [Lan05].

Soit m_1 une chaîne de Markov évidentielle définie sur $\mathcal{P}(\Omega)^N$ par

$$m_1(A_1, \dots, A_N) = m_1(A_1) m_1(A_2|A_1) \dots m_1(A_N|A_{N-1}) \quad (6.4)$$

et soit m_2 une masse bayésienne définie par l'observation $y = (y_1, \dots, y_N)$ comme suit

$$m_2(x_1, \dots, x_N) = \frac{p(y_1, y_2|x_1, x_2) \dots p(y_{N-1}, y_N|x_{N-1}, x_N)}{p(y_2|x_2) \dots p(y_N|x_{N-1})} \quad (6.5)$$

Nous considérons le triplet $T = (U, X, Y)$ défini sur $(\mathcal{P}(\Omega) \times \Omega \times \mathbb{R})^N$ et dont la loi est donnée par :

$$p(t_1, \dots, t_N) \propto q_1(t_1, t_2) \dots q_{N-1}(t_{N-1}, t_N) \quad (6.6)$$

où pour $3 \leq n \leq N$, nous avons

$$q_1(t_1, t_2) = 1_{x_1 \in u_1} 1_{x_2 \in u_2} m_1(u_1) m_1(u_2|u_1) \frac{p(y_1, y_2|x_1, x_2)}{p(y_2|x_2)} \quad (6.7)$$

$$q_n(t_n, t_{n+1}) = 1_{x_n \in u_n} m_1(u_n|u_{n-1}) \frac{p(y_{n-1}, y_n|x_{n-1}, x_n)}{p(y_n|x_n)} \quad (6.8)$$

Ce modèle est donc un modèle de chaîne de Markov triplet (CMT). La fusion des deux masses m_1 et m_2 fournit la distribution *a posteriori* $p(x|y)$ définie à partir de la loi jointe $p(x, y)$, laquelle est la loi marginale de la loi $p(u, x, y)$ du triplet T . En conséquence, les probabilités $p(u_n, x_n|y)$ sont calculables et la complexité des calculs correspondants est linéaire par rapport à la taille de la donnée N .

6.2. Inférence dans une chaîne de Markov couple évidentielle

Soit le modèle triplet $T = (U, X, Y)$ défini précédemment. Nous définissons les probabilités progressive $\alpha_n(x_n, u_n) = p(y_1, \dots, y_n, x_n, u_n)$ et rétrograde $\beta_n(x_n, u_n) = p(y_{n+1}, \dots, y_N|y_n, x_n, u_n)$ généralisées et dont le calcul se fait de la manière itérative suivante :

Initialisation :

$$\alpha_1(x_1, u_1) = p(y_1, x_1, u_1) \quad (6.9)$$

$$\beta_N(x_N, u_N) = 1 \quad (6.10)$$

Induction :

$$\alpha_n(x_n, u_n) = \sum_{x_{n-1}, u_{n-1}} \alpha_{n-1}(x_{n-1}, u_{n-1}) p(t_n|t_{n-1}) \quad (6.11)$$

$$\beta_n(x_n, u_n) = \sum_{x_{n+1}, u_{n+1}} \alpha_{n+1}(x_{n+1}, u_{n+1}) p(t_{n+1}|t_n) \quad (6.12)$$

où $p(t_n, t_{n+1}) = 1_{x_n \in u_n} 1_{x_{n+1} \in u_{n+1}} m(u_n, u_{n+1}) p(y_n, y_{n+1}|x_n, x_{n+1})$.

6.2.1. Restauration MPM d'une chaîne de Markov couple évidentielle

Les probabilités progressive et rétrograde étant définies, nous pouvons calculer les probabilités a posteriori comme suit

$$p(x_n, u_n | y) \propto \alpha_n(x_n, u_n) \beta_n(x_n, u_n) \quad (6.13)$$

Les distributions marginales a posteriori sont ensuite déduites par de simples sommations comme suit

$$p(x_n | y) = \sum_{u_n} p(x_n, u_n | y) \quad (6.14)$$

$$p(u_n | y) = \sum_{x_n} p(x_n, u_n | y) \quad (6.15)$$

Par ailleurs, les distributions a posteriori $\psi(x_n, u_n, x_{n+1}, u_{n+1}) = p(x_n, u_n, x_{n+1}, u_{n+1} | y)$ requises pour l'estimation des paramètres sont calculables selon les formules

$$\psi(x_n, u_n, x_{n+1}, u_{n+1}) \propto \alpha_n(x_n, u_n) p(t_{n+1} | t_n) \beta_{n+1}(x_{n+1}, u_{n+1}) \quad (6.16)$$

6.2.2. Estimation des paramètres d'une chaîne de Markov couple évidentielle

Dans le cas où les paramètres du modèle ne sont pas connus, il est possible de les estimer en faisant appel à des algorithmes du type EM, SEM ou ICE. Comme dans cette étude, l'algorithme EM est adopté, nous adaptons ce dernier à notre nouveau modèle. Pour ce faire, nous considérons un modèle de CMCo évidentielle $T = (U, X, Y)$ comme défini dans (6.6). Un tel modèle est dit gaussien si de plus, les distributions $p(y_{n-1}, y_n | x_{n-1}, x_n)$ sont gaussiennes.

En considérant un ensemble de K classes $\Omega = \{\omega_1, \dots, \omega_K\}$, nous devons estimer les paramètres suivants : $2K^2$ moyennes $\mu_{kk'}^1, \mu_{kk'}^2$, $2K^2$ écarts-types $\sigma_{kk'}^1, \sigma_{kk'}^2$ et K^2 coefficients de corrélation $\rho_{kk'}$ avec $1 \leq k, k' \leq K$ définissant les distributions $p(y_{n-1}, y_n | x_{n-1}, x_n)$ et les $(2^K - 1)^2$ éléments $m_{ij} = m(u_n, u_{n+1})$ de la masse évidentielle stationnaire définie sur $\mathcal{P}(\Omega)^2$.

La procédure d'estimation se fait de la manière suivante :

- Considérer un jeu initial de paramètres $\Theta^0 = (\mu, \sigma, \rho, m)^0$.
- Pour chaque itération $q + 1$, calculer Θ^{q+1} à partir de Θ^q en deux étapes jusqu'à ce qu'un critère d'arrêt soit satisfait:
 - o Estimation : Calculer $\alpha^q(x_n, u_n)$ et $\beta^q(x_n, u_n)$ puis en dériver $\psi^q(x_n, u_n, x_{n+1}, u_{n+1})$.
 - o Maximisation : Calculer Θ^{q+1} selon les formules

$$(\mu_{kk'}^1)^{q+1} = \frac{\sum_{n=1}^{N-1} \sum_{u_n, u_{n+1}} \psi_n^q(x_n = \omega_k, u_n, x_{n+1} = \omega_{k'}, u_{n+1}) y_n}{\sum_{n=1}^{N-1} \sum_{u_n, u_{n+1}} \psi_n^q(x_n, u_n = \lambda_i, x_{n+1}, u_{n+1} = \lambda_j)} \quad (6.17)$$

$$(\mu_{kk'}^2)^{q+1} = \frac{\sum_{n=1}^{N-1} \sum_{u_n, u_{n+1}} \psi_n^q(x_n = \omega_k, u_n, x_{n+1} = \omega_{k'}, u_{n+1}) y_{n+1}}{\sum_{n=1}^{N-1} \sum_{u_n, u_{n+1}} \psi_n^q(x_n, u_n = \lambda_i, x_{n+1}, u_{n+1} = \lambda_j)} \quad (6.18)$$

$$[(\sigma_{kk'}^1)^{q+1}]^2 = \frac{\sum_{n=1}^{N-1} \sum_{u_n, u_{n+1}} \psi_n^q(x_n = \omega_k, u_n, x_{n+1} = \omega_{k'}, u_{n+1}) (y_n - (\mu_{kk'}^1)^{q+1})^2}{\sum_{n=1}^{N-1} \sum_{u_n, u_{n+1}} \psi_n^q(x_n, u_n = \lambda_i, x_{n+1}, u_{n+1} = \lambda_j)} \quad (6.19)$$

$$[(\sigma_{kk'}^2)^{q+1}]^2 = \frac{\sum_{n=1}^{N-1} \sum_{u_n, u_{n+1}} \psi_n^q(x_n = \omega_k, u_n, x_{n+1} = \omega_{k'}, u_{n+1}) (y_{n+1} - (\mu_{kk'}^2)^{q+1})^2}{\sum_{n=1}^{N-1} \sum_{u_n, u_{n+1}} \psi_n^q(x_n, u_n = \lambda_i, x_{n+1}, u_{n+1} = \lambda_j)} \quad (6.20)$$

$$[(\rho_{kk'})^{q+1}]^2 = \frac{\sum_{n=1}^{N-1} \sum_{u_n, u_{n+1}} \psi_n^q(x_n = \omega_k, u_n, x_{n+1} = \omega_{k'}, u_{n+1}) (y_n - (\mu_{kk'}^1)^{q+1}) (y_{n+1} - (\mu_{kk'}^2)^{q+1})}{(\sigma_{kk'}^1)^{q+1} (\sigma_{kk'}^2)^{q+1} \sum_{n=1}^{N-1} \sum_{u_n, u_{n+1}} \psi_n^q(x_n, u_n = \lambda_i, x_{n+1}, u_{n+1} = \lambda_j)} \quad (6.21)$$

$$m_{ij}^{q+1} = \frac{1}{\#i\#j} \sum_{n=1}^{N-1} \sum_{x_n, x_{n+1}} \psi_n^q(x_n, u_n = \lambda_i, x_{n+1}, u_{n+1} = \lambda_j) \quad (6.22)$$

Où $\#i$ désigne le cardinal de l'ensemble λ_i .

6.3. Expériences

Dans cette section, nous évaluons les performances du modèle de chaîne de Markov couple évidentielle par rapport à son homologue bayésien. Pour ce faire, nous présentons deux séries d'expériences. Dans la première, nous considérons des données synthétiques, générée aléatoirement. Dans la deuxième série, nous considérons une image de classes à forte non-stationnarité que nous bruyons de manière à y introduire un bruit corrélé. Bien entendu, l'image est convertie en donnée unidimensionnelle grâce au parcours Hilbert-Peano [Pea1890, Hil1891].

6.4. Segmentation non-supervisée de chaînes de Markov couples non-stationnaires

Soit le modèle de CMCo non-stationnaire définis par la formule (6.3) avec $\Omega = \{\omega_1, \omega_2\}$ et $N = 16384$. Soit les deux matrices Q et L

$$Q = \begin{pmatrix} 0.48 & 0.02 \\ 0.02 & 0.48 \end{pmatrix}, L = \begin{pmatrix} 0.25 & 0.25 \\ 0.25 & 0.25 \end{pmatrix}$$

Pour une valeur donnée de s , les distributions non-stationnaires sont régies par ces deux matrices selon les formules suivantes :

- $p_n(i, j) = q_{ij}$ pour $2rs + 1 \leq n \leq (2r + 1)s$ avec $r \in \mathbb{N}$;
- $p_n(i, j) = l_{ij}$ pour $(2r + 1)s \leq n \leq (2r + 2)s$ avec $r \in \mathbb{N}$.

Les données sont simulées en considérant différentes valeurs de s en utilisant un bruit gaussien dont les paramètres sont les suivants

$$\mu^1 = [\mu^2]^T = \begin{pmatrix} -5 & -3 \\ 3 & 5 \end{pmatrix}, \sigma^1 = [\sigma^2]^T = \begin{pmatrix} 4 & 4 \\ 4 & 4 \end{pmatrix}.$$

Pour le coefficient de corrélation ρ , nous avons considéré trois valeurs, définissant chacune un jeu de données :

- Jeu A : $\rho = \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}$;
- Jeu B : $\rho = \begin{pmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{pmatrix}$;
- Jeu C : $\rho = \begin{pmatrix} 0.9 & 0.9 \\ 0.9 & 0.9 \end{pmatrix}$.

La restauration MPM est ensuite effectuée selon le formalisme des chaînes de Markov couple non-stationnaires en considérant le jeu de paramètres réels Θ . Les résultats obtenus serviront de référence pour les autres modèles. La restauration MPM est réalisée ensuite en utilisant les paramètres estimés via l'algorithme EM correspondant aux modèles de CMC, CMC évidentielles et CMCo évidentielles. La phase d'initialisation des paramètres est primordiale et influe toujours sur la qualité des résultats finaux. Dans cette série d'expériences, les paramètres sont initialisés en utilisant les paramètres estimés selon le formalisme des CMC évidentielles. Les résultats obtenus sont présentés dans TAB. 6.I.

Tableau 6.1. Taux d'erreur moyens de segmentation des données synthétiques (%)

Jeu	s	$\tau_{\Theta-PMC}$	τ_{EM-HMC}	τ_{EM-EHM}	$\tau_{EM-EPMC}$
A	16	9.1	17.2	17.1	12.8
	64	9.2	16.8	16.2	10.6
	256	9.2	16.6	15.9	9.7
	1024	9.1	16.7	15.7	9.2
	4096	9	16.5	15.5	9
B	16	7.6	20.5	20.4	8.9
	64	7.3	20.7	19.9	7.7
	256	7.3	21	20.5	7.7
	1024	7.6	20.4	19.7	7.6
	4096	7.8	21.2	20.4	7.8
C	16	3	24.2	24.2	3.7
	64	2.9	22.9	22.6	3.2
	256	2.6	23	22.8	2.8
	1024	2.4	22.4	21.8	2.6
	4096	2.7	23	22	2.7

Les résultats de restauration sont également illustrés pour le jeu de données C pour $s = 4096$ dans Fig. 6.1.

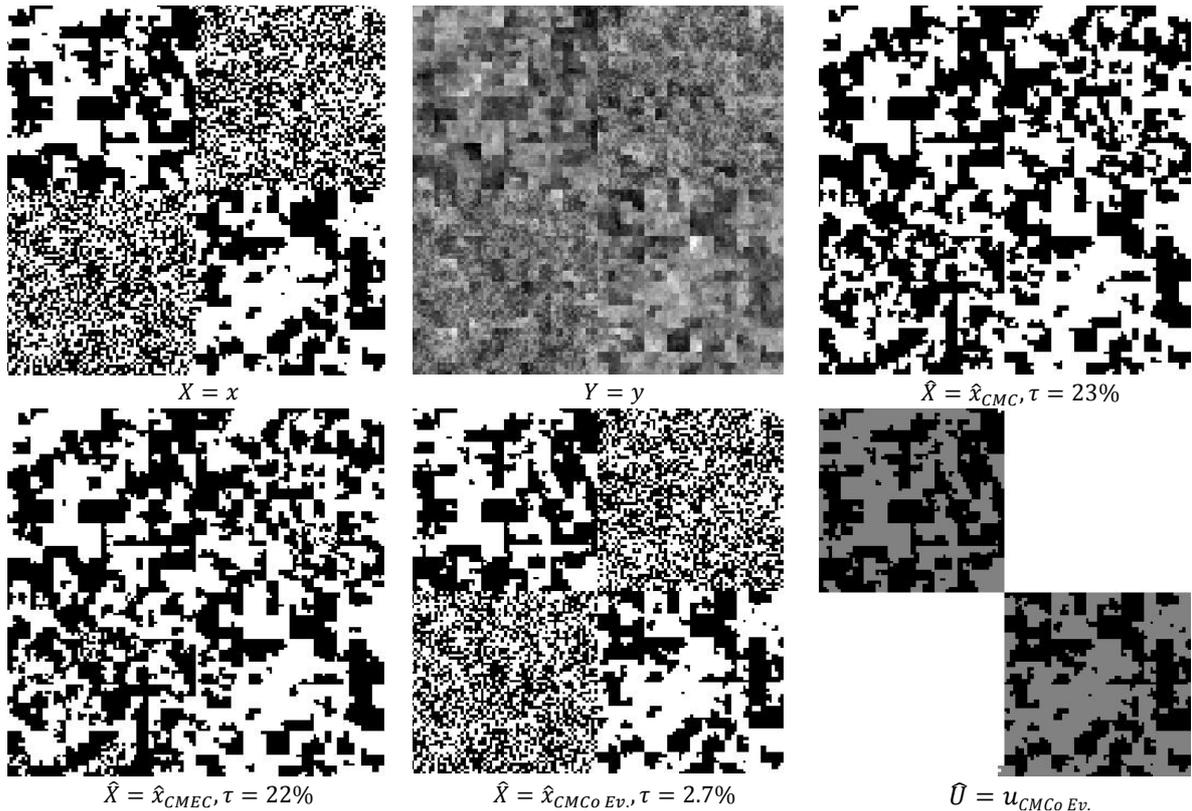


Figure 6.1. Résultats de restauration des données synthétiques simulées selon une CMCo non-stationnaire

D'une manière générale, les résultats de segmentation démontrent que le modèle de CMCo évidentielle surpasse les modèles classiques.

La lecture de TAB. 6.I permet de formuler les constats suivants :

- La distorsion entre les modélisations classiques et les données est d'autant plus grande que la corrélation de bruit est grande. En effet, indépendamment de la valeur de s , les pires taux d'erreurs sont obtenus pour le jeu C caractérisé par un fort coefficient de corrélation.
- Par ailleurs, il convient de remarquer que le modèle de CMCo évidentielle prend, non seulement en compte la corrélation du bruit, mais aussi en tire profit comme étant une caractéristique discriminante. Effectivement, les meilleurs résultats de restauration sont obtenus pour le jeu de données C.
- Une autre remarque très importante est le fait que les résultats de restauration MPM obtenus sur la base des paramètres estimés selon le modèle de CMCo évidentielle sont comparables à ceux obtenus en utilisant le jeu de paramètre réels, notamment pour les grandes valeurs de s (1024 et 4096). Ceci confirme l'intérêt du modèle proposé dans la mesure où il permet d'obtenir des résultats satisfaisants dans le contexte non-supervisé.

6.4.2. Segmentation non-supervisée d'images non-stationnaires entachées de bruit corrélé

Soit l'image de classes non-stationnaire 'Nazca bird' de taille 128×128 (Fig. 6.2). Nous disposons donc de la réalisation du processus caché X à valeurs dans $\Omega = \{\omega_1, \omega_2\}$ où ω_1 et ω_2

correspondent physiquement aux pixels noirs et blancs respectivement. Ensuite, nous introduisons un bruit corrélé au sein de l'image. Le processus observable est donné par

$$Y_s = \sigma_{x_s} W_s + \mu_{x_s} + a \sum_{i=1}^4 (\sigma_{x_{s_i}} W_{s_i} + \mu_{x_{s_i}}) \quad (6.23)$$

où W est un bruit blanc gaussien de variance 1 et s_1, s_2, s_3 et s_4 correspondent aux quatre voisins du pixel s . Il convient de noter que le paramètre a reflète le degré d'implication du voisinage d'un site dans sa valeur observée, ce qui peut être interprété comme un indicateur de corrélation entre sites voisins.

Les réalisations des deux processus X et Y sont ensuite converties en deux séquences unidimensionnelles grâce au parcours Hilbert-Peano [Pea1890, Hil1891]. Nous sommes donc en présence d'une donnée $z = (x, y)$ de nature très complexe et ne correspondant à aucun des modèles markoviens étant donnée la méthode de bruitage (basée sur une convolution impliquant les quatre plus proches voisins) et le parcours d'Hilbert-Peano.

Les simulations ont été effectuées pour différentes valeurs du paramètre a, μ et σ . La restauration MPM a été accomplie en utilisant les paramètres estimés par l'algorithme EM conformément aux modèles de chaîne de Markov cachée, chaîne de Markov couple, chaîne de Markov cachée évidentielle et chaîne de Markov couple évidentielle.

Comme nous pouvons le constater dans Fig. 6.2, où les réalisations des différents processus simulés et estimés sont illustrées pour les valeurs $a = 0.25, \mu_1 = 0, \mu_1 = 3, \sigma_1 = 1$ et $\sigma_2 = 2$, le gain en taux d'erreur par rapport au modèle de chaîne de Markov cachée évidentielle avoisine les 50%. L'estimation de la réalisation du processus auxiliaire U démontre clairement comment le nouveau modèle est capable de distinguer entre les régions de l'image où il y a beaucoup de détails (ailes et queue de l'oiseau) et celles où il y a très peu d'échange (transition) entre les deux classes. Le modèle de chaîne de Markov cachée évidentielle, quant à lui, ne permet pas de faire une telle différenciation, car il ne tire pas profit de l'information de la corrélation. Par ailleurs, les restaurations basées sur les modèles de chaînes de Markov cachée et couple conventionnelles sont de mauvaise qualité.

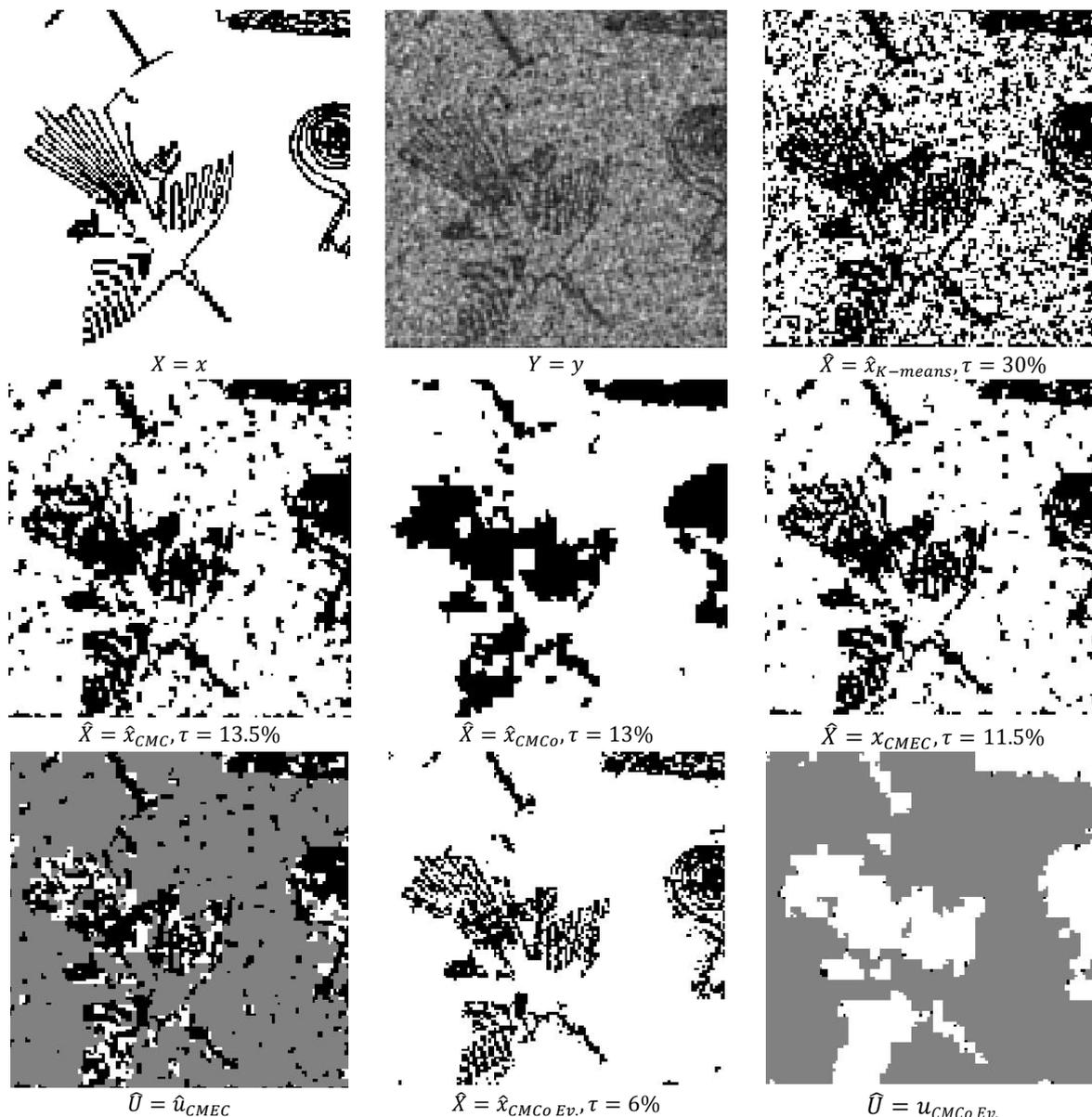


Figure 6.2. Exemple de restauration d'une image non-stationnaire entachée de bruit corrélé

Plus explicitement, nous pouvons tirer les conclusions suivantes :

- Le modèle de CMC, ne tenant compte ni de la fluctuation des distributions *a priori* ni de la corrélation du bruit, fournit un taux d'erreur médiocre ($\tau = 13.5\%$). L'imprécision s'exprime notamment dans l'arrière-plan composé exclusivement de pixels blancs et où des sites ont été erronément attribués à la classe des pixels noirs.
- Le modèle de CMCo, contrairement au modèle de CMC, permet de distinguer l'arrière plan de l'image grâce à son aptitude à modéliser les bruits corrélés. Toutefois, les zones du dessin caractérisées par un fort échange entre les deux classes, ont été assignées à une même classe. Ceci est dû à la non prise en compte de la variation des distributions *a priori*. Par conséquent, le taux d'erreur pour ce modèle est également assez médiocre ($\tau = 13\%$).
- Le modèle de CMC évidentielle fournit une segmentation relativement meilleure dans les zones détaillées du dessin, mais ne parvient pas à bien classifier les pixels de l'arrière plan.
- Enfin, le modèle de CMCo évidentielle, doté à la fois des avantages de la théorie de l'évidence et des modèles couples fournit une segmentation satisfaisante malgré le bruitage

particulier. L'estimation du processus auxiliaire confirme la performance de ce modèle ; les zones à fort échange entre les deux classes subissent un affaiblissement plus conséquent à cause de l'incertitude des fonctions *a priori*. Pour les sites appartenant à de telles zones, l'information sur l'observation est prépondérante par rapport à l'information contextuelle.

- Les résultats de restauration confirment que les sites de l'arrière plan sont mieux classifiés par les modèles couples. En effet, la distribution *a priori* commune (stationnaire) estimée privilégie les zones à faible échange. De ce fait, seules les zones à fort échange posent problème pour ces modèles. En revanche, les modèles de CMC évidentiels échouent surtout dans l'arrière plan de l'image à cause de leur faiblesse face au bruit corrélé. Le modèle proposé dans ce chapitre, quant à lui, réunit les avantages des deux modèles et permet ainsi de fournir un résultat convenable.

6.5. Conclusion

Dans ce chapitre, nous avons défini un nouveau modèle appelé chaîne de Markov couple évidentielle. Nous avons ainsi montré comment la théorie de l'évidence peut être exploitée au sein des modèles de CMCo pour modéliser les données non-stationnaires entachées de bruit corrélé. Nous avons également proposé un algorithme de type EM pour l'estimation des paramètres. La stratégie bayésienne MPM a aussi été présentée dans ce contexte. Les expériences menées sur des images et des données synthétiques ont démontré la supériorité de la modélisation proposée par rapport aux modélisations classiques.

Chapitre 7. Chaînes de Markov cachées multi-senseurs non-stationnaires

Les dernières avancées dans la théorie des modèles de Markov et leurs techniques computationnelles associées ont ouvert de nouvelles perspectives pour le traitement de l'image et du signal. En particulier, l'utilisation de la théorie de l'évidence a permis de relever certains défis que les modélisations classiques ne permettent pas d'aborder de manière fiable. Ces challenges portent essentiellement sur deux situations : les données multi-senseurs où l'utilisation de la fusion de Dempster-Shafer (DS) détruit la markovianité, et les données non-stationnaires où les traitements statistiques associés ne sont pas faisables dans le contexte non-supervisé. Pour chacune de ces deux problématiques, la théorie de l'évidence de Dempster-Shafer a été employée avec succès, conformément au formalisme des modèles de Markov triplets, pour pallier aux limites présentés par les modèles bayésiens. Cependant, jusqu'ici, les deux difficultés n'ont pas été considérées simultanément. Ainsi, dans ce chapitre, nous proposons d'utiliser la fusion de Dempster-Shafer pour prendre en compte les informations contextuelles lors de la segmentation des signaux multi-senseurs. Dans ce cadre, nous décrivons les différentes techniques de restauration et d'estimation de paramètres. Enfin, une étude expérimentale est menée sur des données multi-senseurs non-stationnaires pour montrer l'intérêt de la modélisation proposée par rapport à ses devancières.

7.1. Chaînes de Markov cachées multi-senseurs non-stationnaires

Soit $\Omega = \{\omega_1, \dots, \omega_K\}$ un cadre de discernement contenant K hypothèses exclusives et soit $\mathcal{P}(\Omega) = \{\emptyset, \{\omega_1\}, \dots, \Omega\}$ l'ensemble de sous-ensembles qui lui est associé. Contrairement aux approches bayésiennes qui considèrent uniquement les hypothèses exclusives, les méthodes employant la théorie de l'évidence adoptent des hypothèses composées à travers la notion de masse qui généralise la mesure de probabilité offrant ainsi un formalisme élégant pour modéliser l'incertitude, l'imprécision ou même le manque d'information.

Dans le cadre des modèles de Markov, la théorie de l'évidence a été utilisée pour mieux modéliser l'aspect non-stationnaire des données et également pour permettre la fusion de différents senseurs hétérogènes tout en restant dans le cadre markovien. La faisabilité de ces différentes techniques est due aux modèles de Markov triplets. En effet, il a été démontré que le résultat de la fusion de différentes masses, dont des masses évidentielles, représentant les données est une distribution marginale d'une CMT. Ce même résultat étant la loi *a posteriori* d'intérêt, les différents calculs d'estimation sont faisables.

Cette propriété a été employée pour modéliser l'aspect non-stationnaire des données dans le contexte des CMC et CMCo. Elle a également été exploitée pour rendre possible la fusion de différents senseurs dans le contexte markovien à travers les CMC MS et les champs de Markov cachés multi-senseurs [Ben01]. L'objectif de ce chapitre est d'exploiter ce résultat pour considérer les deux situations en même temps. Plus explicitement, la masse *a priori* sera étendue à une masse évidentielle pour prendre en considération l'aspect non-stationnaire de $p(x)$. Par ailleurs,

la masse modélisant l'observation du senseur présentant une imprécision (présence de nuage dans l'image satellitaire par exemple [Heg98, Ben01]) sera également étendue à une masse évidentielle. Nous définissons ainsi le modèle de « chaîne de Markov cachée multi-senseurs non-stationnaire » qui étend à la fois les modèles de CMEC et CMC MS décrit dans le chapitre 3.

7.1.1. Définition du modèle

Dans cette section, nous décrivons le modèle de chaîne de Markov cachée multi-senseurs non-stationnaire (CMC MN). Soit $X = X_{1..N}$ un processus caché qui prend ses valeurs à partir d'un ensemble de classes $\Omega = \{\omega_1, \dots, \omega_K\}$ et qui doit être estimé à partir d'une famille de processus observables $Y = Y^{1..R}$ fournis par R senseurs indépendants $S_{1..R}$ où $Y^r = Y_{1..N}^r$ avec chaque Y_n^r dans \mathbb{R} . Supposons maintenant que la réalisation de X est régie par une chaîne de Markov non-stationnaire dans le sens où les distributions $p(x_{n+1} = \omega | x_n = \omega')$ dépendent de n . Supposons également qu'au moins l'un des senseurs $S_{1..R}$ est bayésien. Sans perte de généralité, soit S_1 un tel senseur. La restauration bayésienne MPM en utilisant toute l'observation $y = y^{1..R}$ est faisable et le nombre d'opérations nécessaires pour son calcul est linéaire par rapport à la taille des données N [Pie07]. Le modèle $Z = (X, Y)$ est alors appelé chaîne de Markov cachée multi-senseurs non-stationnaire. La restauration MPM s'effectue alors grâce à la fusion des différentes masses impliquées dans le modèle.

$$m(x) = (m_0 \oplus \dots \oplus m_R)(x) = p(x|y) \quad (7.4)$$

Bien que la distribution résultante de cette fusion ne soit pas forcément une chaîne de Markov, elle est une distribution marginale d'une chaîne de Markov triplet. Par conséquent, les probabilités *a posteriori* $p(x_n|y)$ sont calculables.

Nous montrons à présent comment modéliser les données via le modèle CMC MN :

- D'abord, la chaîne de Markov non-stationnaire régissant le processus caché X est remplacée par une masse évidentielle stationnaire m_0 . Pour cela, nous définissons le processus sous-jacent $U^0 = U_{1..N}^0$ qui prend ses valeurs dans l'ensemble $\mathcal{P}(\Omega) = \{\emptyset, \{\omega_1\}, \dots, \Omega\}$. m_0 est donc définie sur $\mathcal{P}(\Omega)$.

- Le senseur S_1 étant bayésien, la masse relative à son observation m_1 est alors définie sur Ω par $m_1(x) \propto p(y_1^1|x_1)p(y_2^1|x_2) \dots p(y_N^1|x_N)$.

- Pour chaque senseur S_r , nous préconisons une fonction de masse associée à son observation m_r . A titre d'exemple, si S_r est sensible uniquement à la classe ω_1 , la masse correspondante m_r sera définie sur $\Lambda_r = \{\{\omega_1\}, \{\omega_2, \dots, \omega_K\}\}$. Pour cela, nous définissons le processus sous-jacent $U^r = U_{1..N}^r$ qui prend ses valeurs dans l'ensemble Λ_r .

- Enfin, pour certains senseurs, il est possible de considérer plus d'une fonction de masse. Considérons par exemple le senseur optique vu plus haut où l'image fournie présente des nuages. La présence des nuages peut être modélisée par une fonction de masse définie sur $\mathcal{P}(\Omega) = \{\omega_1, \omega_2, \Omega\}$. Par ailleurs, nous pouvons préconiser une autre fonction de masse, pour le même senseur, modélisant l'information *a priori* sur la présence des nuages dans l'image (en faisant abstraction de l'observation du senseur).

7.2. Segmentation non-supervisée d'une chaîne de Markov cachée multi-senseurs non-stationnaire

Pour détailler le déroulement des différents traitements d'estimation, nous considérons le problème de segmentation suivant. Soit deux images fournies par deux senseurs indépendants d'une même scène non-stationnaire. L'objectif est de dériver la carte de la scène x à valeurs dans $\Omega = \{\omega_1, \omega_2\}$ où ω_1 et ω_2 correspondent à « forêt » et « eau » respectivement. Soit S_1 et S_2 deux senseurs avec les particularités suivantes :

- Le senseur S_1 est bayésien, son observation $y_{1..N}^1$ est reliée aux classes à travers les densités de bruit $p(y_n^1|x_n = \omega_1)$ et $p(y_n^1|x_n = \omega_2)$.
- Le senseur S_2 est évidentiel : l'image $y_{1..N}^2$ présente des nuages. Nous avons donc trois possibilités : « forêt », « eau » et « nuage ». L'observation $y_{1..N}^2$ est reliée à ces classes à travers les densités de bruit $p(y_n^2|x_n = \omega_1)$ et $p(y_n^2|x_n = \omega_2)$ et $p(y_n^2|x_n = \Omega)$.

Le problème est donc d'estimer la réalisation $x_{1..N}$ en utilisant les images fournies par les deux senseurs dans un contexte markovien de façon à ce que l'aspect non-stationnaire de $p(x)$ soit pris en compte. Pour ce faire, nous utilisons le modèle CMC MN.

7.2.1. Modélisation des données

La première étape consiste à regrouper toutes les informations qui peuvent être fusionnées plus tard pour produire la segmentation recherchée.

- Tout d'abord, la CMC non-stationnaire gouvernant le processus X est remplacée par une chaîne de Markov évidentielle m_0 . Soit le processus auxiliaire $U^1 = U_{1..N}^1$ à valeurs dans $\mathcal{P}(\Omega) = \{\emptyset, \omega_1, \omega_2, \Omega\}$ et modélisant la non-stationnarité de X . La masse m_0 est alors définie sur $\mathcal{P}(\Omega)^N$ par $m_0(u^1) = m_0(u_1^1) \prod_{n=2}^N m_0(u_n^1|u_{n-1}^1)$.

- Le senseur S_1 étant bayésien, la masse relative à son observation m_1 est alors définie sur Ω^N par $m_1(x) = \prod_{n=1}^N m_1(x_n) \propto \prod_{n=1}^N p(y_n^1|x_n)$. Par conséquent, la restauration MPM peut être effectuée sur la base de l'observation de ce senseur pris seul.

- La présence des nuages dans l'image fournie par le second senseur S_2 peut être modélisée par une mesure de probabilité sur $\mathcal{P}(\Omega) = \{\emptyset, \omega_1, \omega_2, \Omega\}$ qui est une fonction de masse définie par $m_2(u^2) \propto \prod_{n=1}^N m_2(u_n^2)$ où u^2 est à valeurs dans $\mathcal{P}(\Omega)$ et $m_2(\omega_1) \propto p(y_n^2|u_n^2 = \omega_1)$, $m_2(\omega_2) \propto p(y_n^2|u_n^2 = \omega_2)$ et $m_2(\Omega) \propto p(y_n^2|u_n^2 = \Omega)$.

- Enfin, il semble intuitif de considérer une masse évidentielle markovienne m_3 modélisant l'information *a priori* sur la présence des nuages dans l'image de S_2 et définie sur $\mathcal{P}(\Omega)$ par $m_3(u^2) = m_3(u_1^2) \prod_{n=2}^N m_3(u_n^2|u_{n-1}^2)$.

7.2.2. Restauration MPM d'une chaîne de Markov multi-senseurs non-stationnaire

Nous avons défini une famille de quatre masses m_0 , m_1 , m_2 et m_3 qui représentent toute l'information que nous avons sur les données. Soit $T = (X, U, Y) = (V, Y)$ une CMT où $X = X_{1..N}$ avec chaque $x_n \in \Omega$, $U = (U^1, U^2)$ avec chaque $u_n \in [\mathcal{P}(\Omega)]^2$ et $Y = (Y^1, Y^2)$ avec chaque $y_n \in \mathbb{R}^2$. La distribution de T est alors donnée par

$$p(t_1, \dots, t_N) \propto q_1(t_1, t_2)q_2(t_2, t_3) \dots q_{N-1}(t_{N-1}, t_N) \quad (7.4)$$

avec

$$q_1(t_1, t_2) = \mathbf{1}_{[x_1 \in u_1^1 \cap u_1^2]} \mathbf{1}_{[x_2 \in u_2^1 \cap u_2^2]} m_0(u_1^1) m_0(u_2^1 | u_1^1) m_1(x_1) m_1(x_2) m_2(u_1^2) m_2(u_2^2) m_3(u_1^2) m_3(u_2^2 | u_1^2);$$

$$q_n(t_n, t_{n+1}) = \mathbf{1}_{[x_{n+1} \in u_{n+1}^1 \cap u_{n+1}^2]} m_0(u_{n+1}^1 | u_n^1) m_1(x_{n+1}) m_2(u_{n+1}^2) m_3(u_{n+1}^2 | u_n^2).$$

Subséquentement, le résultat de la fusion DS $m = (m_0 \oplus m_1 \oplus m_2 \oplus m_3)$ est la distribution *a posteriori* $p(x|y)$ définie par la loi jointe $p(x, y)$ et qui est également la distribution marginale de la CMT définie par (7.4). Par conséquent, les probabilités marginales $p(x_n, u_n | y)$ sont calculables et ainsi sont les probabilités $p(x_n | y)$ requises pour la restauration MPM.

Enfin, pour accomplir la restauration bayésienne MPM, nous pouvons soit utiliser la définition générale d'une chaîne de Markov [Pie07] soit faire appel aux fonctions de probabilités progressives $\alpha_n(v_n) = p(y_1, \dots, y_n, v_n)$ et rétrogrades $\beta_n(v_n) = p(y_{n+1}, \dots, y_N | v_n)$ qui ont été adaptées au contexte des CMC MN [Bou12a]. Dans cette étude, nous adoptons cette dernière option. Ainsi, ces fonctions sont calculables de la manière itérative suivante

Initialisation :

$$\alpha_1(v_1) = p(v_1, y_1) \quad (7.5)$$

$$\beta_N(v_N) = 1 \quad (7.6)$$

Induction :

$$\alpha_n(v_n) = \sum_{v_{n-1}} \alpha_{n-1}(v_{n-1}) p(v_n | v_{n-1}) m_1(x_n) m_2(u_n^2) \quad (7.7)$$

$$\beta_n(v_n) = \sum_{v_{n+1}} \beta_{n+1}(v_{n+1}) p(v_{n+1} | v_n) m_1(x_{n+1}) m_2(u_{n+1}^2) \quad (7.8)$$

avec

$$p(v_{n+1} | v_n) \propto \mathbf{1}_{[x_n \in u_n^1 \cap u_n^2]} \mathbf{1}_{[x_{n+1} \in u_{n+1}^1 \cap u_{n+1}^2]} m_0(u_{n+1}^1 | u_n^1) m_3(u_{n+1}^2 | u_n^2);$$

$$p(v_1, y_1) \propto \mathbf{1}_{[x_1 \in u_1^1 \cap u_1^2]} m_0(u_1^1) m_1(x_1) m_2(u_1^2) m_3(u_1^2).$$

Les distributions marginales *a posteriori* $p(v_n | y)$ et $p(x_n | y)$ sont calculées comme suit

$$p(v_n | y) \propto \alpha_n(v_n) \beta_n(v_n) \quad (7.9)$$

$$p(x_n | y) = \sum_{v_n \ni x_n} p(v_n | y) \quad (7.10)$$

Par ailleurs, les probabilités jointes et marginales *a posteriori* nécessaires pour l'algorithme d'estimation des paramètres sont calculées selon les formules

$$\psi(v_n, v_{n+1}) = p(v_n, v_{n+1} | y) = \alpha_n(v_n) p(v_n | v_{n-1}) m_1(x_{n+1}) m_2(u_{n+1}^2) \beta_n(v_{n+1}) \quad (7.11)$$

$$\xi(v_n) = p(v_n | y) = \sum_{v_{n-1}} \psi(v_{n-1}, v_n) \quad (7.12)$$

7.2.3. Estimation des paramètres d'une chaîne de Markov multi-senseurs non-stationnaire

Pour estimer les paramètres du modèle, nous proposons d'utiliser une version adaptée de l'algorithme EM. Notons également que les algorithmes SEM et ECI restent applicables sous réserve d'adaptation au contexte des CMC MN.

Considérons la CMT $T = (X, U, Y) = (V, Y)$ définie plus haut. Supposons de plus que les densités de bruit $p(y_n^1|x_n)$ et $p(y_n^2|u_n^2)$ sont gaussiennes.

Conformément au problème de segmentation défini plus haut, nous devons estimer les paramètres suivants : les masses évidentielles $m_{ij} = m_0(u_n^1 = \lambda_i, u_{n+1}^1 = \lambda_j)$ définie sur $[P(\Omega)]^2$, les $K = 2$ moyennes $\mu_{1..K}^1$ et écarts-types $\sigma_{1..K}^1$ des densités gaussiennes régissant m_1 , les $K + 1 = 3$ moyennes $\mu_{1..K+1}^2$ et écarts-types $\sigma_{1..K+1}^2$ des densités gaussiennes régissant m_2 et les masses de transition $c_{ij} = m_3(u_n^2 = \lambda'_i | u_{n+1}^2 = \lambda'_j)$ définies sur $[P(\Omega)]^2$.

La procédure d'estimation se fait de la manière suivante :

- Considérer un jeu initial de paramètres $\Theta^0 = (m_{ij}^0, c_{ij}^0, (\mu_{1..K}^1, \mu_{1..K}^1, \sigma_{1..K}^1, \sigma_{1..K}^1)^0)$.
- Pour chaque itération $q + 1$, calculer Θ^{q+1} à partir de Θ^q en deux étapes jusqu'à ce qu'un critère d'arrêt soit satisfait :
 - o Estimation : Calculer les fonctions ψ_n^q and ξ_n^q à partir de y et Θ^q .
 - o Maximisation : Calculer Θ^{q+1} selon les formules

$$(\mu_k^1)^{q+1} = \frac{\sum_{n=1}^N \sum_{u_n} \xi^q(x_n = \omega_k, u_n) y_n^1}{\sum_{n=1}^N \sum_{u_n} \xi^q(x_n = \omega_k, u_n)} \quad (7.13)$$

$$(\mu_k^2)^{q+1} = \frac{\sum_{n=1}^N \sum_{v_n} \mathbf{1}_{[u_n^2 = \lambda'_k]} \xi^q(v_n) y_n^2}{\sum_{n=1}^N \sum_{v_n} \mathbf{1}_{[u_n^2 = \lambda'_k]} \xi^q(v_n)} \quad (7.14)$$

$$[(\sigma_k^1)^{q+1}]^2 = \frac{\sum_{n=1}^N \sum_{u_n} \xi^q(x_n = \omega_k, u_n) (y_n^1 - (\mu_k^1)^{q+1})^2}{\sum_{n=1}^N \sum_{u_n} \xi^q(x_n = \omega_k, u_n)} \quad (7.15)$$

$$[(\sigma_k^2)^{q+1}]^2 = \frac{\sum_{n=1}^N \sum_{v_n} \mathbf{1}_{[u_n^2 = \lambda'_k]} \xi^q(v_n) (y_n^2 - (\mu_k^2)^{q+1})^2}{\sum_{n=1}^N \sum_{v_n} \mathbf{1}_{[u_n^2 = \lambda'_k]} \xi^q(v_n)} \quad (7.16)$$

$$m_{ij}^{q+1} = \frac{1}{(N-1)\#i\#j} \sum_{n=1}^N \sum_{v_n, v_{n+1}} \mathbf{1}_{[u_n^1 = \lambda_i]} \mathbf{1}_{[u_{n+1}^1 = \lambda_j]} \psi^q(v_n, v_{n+1}) \quad (7.17)$$

$$c_{ij}^{q+1} = \frac{\sum_{n=1}^N \sum_{v_n, v_{n+1}} \mathbf{1}_{[u_n^2 = \lambda'_i]} \mathbf{1}_{[u_{n+1}^2 = \lambda'_j]} \psi^q(v_n, v_{n+1})}{\sum_{n=1}^N \sum_{v_n} \mathbf{1}_{[u_n^2 = \lambda'_i]} \xi^q(v_n)} \quad (7.18)$$

avec $\#i$ représente le cardinal de l'ensemble λ_i .

7.3. Résultats expérimentaux

Dans cette section, nous évaluons les performances du modèle de CMC MN, proposé dans ce chapitre, pour le problème de segmentation non-supervisée de signaux non-stationnaires multi-senseurs par rapport aux autres modélisations markoviennes.

Pour cela, nous considérons la situation suivante :

Soit S_1 et S_2 deux senseurs fournissant deux données observables $y^1 = y_{1..N}^1$ et $y^2 = y_{1..N}^2$ qui peuvent être vues comme des versions bruitées d'une même vérité terrain $x = x_{1..N}$ présentant les difficultés suivantes :

- le processus X , qui est caché et doit être estimé d'une certaine manière à partir de l'observation $Y = (Y^1, Y^2)$, peut être une chaîne de Markov fortement non-stationnaire.
- La séquence d'observation Y^2 présente, visuellement, une classe supplémentaire. Cette situation peut correspondre à la présence de nuages dans une image aérienne ; voire l'absence d'observation au niveau de certains sites. Ainsi, Cette classe, de trop, représente l'ignorance associée au fait qu'il n'est pas possible de décider à quelles classes appartiennent les sites concernés. Le long de cette étude expérimentale, nous supposons le processus associé à la présence d'une telle extra-classe markovien. Ce processus sera noté $B = B_{1..N}$.

Conformément à cette situation, nous proposons deux séries d'expériences. Dans la première série, nous considérons des données synthétiques multi-senseurs non-stationnaires. La seconde série d'expériences, quant à elle, portera sur des images de classes non-stationnaires que nous bruyons conformément au contexte multi-senseurs non-stationnaire. Pour les deux séries, la restauration non-supervisée MPM est accomplie conformément aux différentes modélisations markoviennes.

7.3.1. Segmentation non-supervisée de chaînes de Markov cachées multi-senseurs non-stationnaires

Dans cette série d'expérience, nous considérons des CMC MN. Soit $T = (X, U^2, Y)$ un tel model avec $\Omega = \{\omega_1, \omega_2\}$ et $N = 5000$ et soient les matrices suivantes :

$$M_1 = \begin{bmatrix} 0.98 & 0.02 \\ 0.02 & 0.98 \end{bmatrix}, M_2 = \begin{bmatrix} 0.6 & 0.4 \\ 0.4 & 0.6 \end{bmatrix}, J = \begin{bmatrix} 0.98 & 0.02 \\ 0.01 & 0.99 \end{bmatrix}.$$

Le processus caché X est non-stationnaire de la manière suivante : étant donnée une valeur $s \in \mathbb{N}$, nous définissons $X^i = (X_{(i-1)s+1}, X_{(i-1)s+2}, \dots, X_{is})$. La réalisation x est simulée comme suit :

- La réalisation x_1 est simulée par tirage aléatoire uniforme à partir de Ω .
- Les prochaines réalisations de x sont simulées selon la matrice de transition M_1 pour X^1, X^3, \dots et M_2 pour X^2, X^4, \dots

Par ailleurs, la réalisation du processus B associé à la présence de la classe supplémentaire dans Y^2 est simulée selon une chaîne de Markov de matrice de transition J .

Par conséquent, la réalisation du processus auxiliaire U^2 est dérivée comme suit : pour tout $1 \leq n \leq N$, $u_n^2 = \Omega$ si $b_n = 1$ et $u_n^2 = x_n$ sinon.

Etant données les réalisations de X et U^2 , les deux séquences observables sont simulées comme suit :

- Senseur 1 : La réalisation $y^1 = y_{1..N}^1$ est simulée conformément à $p(y^1|x) = p(y_2^1|x_2) \dots p(y_N^1|x_N)$ où pour tout $1 \leq n \leq N$, $p(y_n^1|x_n = \omega_1)$ et $p(y_n^1|x_n = \omega_2)$ sont deux densités gaussiennes de moyennes 0 et 1 respectivement et d'écart-type égaux à 1.
- Senseur 2 : La réalisation $y^2 = y_{1..N}^2$ est simulée conformément à $p(y^2|u^2) = p(y_1^2|u_1^2)p(y_2^2|u_2^2) \dots p(y_N^2|u_N^2)$ où pour tout $1 \leq n \leq N$, $p(y_n^2|u_n^2 = \{\omega_1\})$, $p(y_n^2|u_n^2 = \{\omega_2\})$ et $p(y_n^2|u_n^2 = \{\omega_1, \omega_2\})$ sont des densités gaussiennes de moyennes 0, 1 et 2 respectivement et d'écart-type égaux à 1.

Enfin, la segmentation non-supervisée est effectuée selon les approches suivantes (de généralité croissante):

- Sur la base de l'observation y^1 seule, la segmentation non-supervisée est accomplie selon l'algorithme K-means, le modèle de CMC et le modèle de CMEC.
- Sur la base de toute l'observation y , la segmentation non-supervisée MPM est accomplie selon le modèle de CMC MS et le modèle de CMC MN.

Pour chaque valeur de s , 100 simulations ont été effectuées. Les taux d'erreur moyens de segmentation sont illustrés dans TAB.7.I.

Tableau 7.1. Taux d'erreur de la segmentation non-supervisée des données multi-senseurs non-stationnaires synthétiques (%)

s	$T_{K-means}$	T_{CMC-EM}^{MPM}	$T_{CMEC-EM}^{MPM}$	$T_{CMC\ MS-EM}^{MPM}$	$T_{CMC\ MN-EM}^{MPM}$
2	30.6	26.7	26.7	22.0	22.0
4	32.3	25.7	25.6	19.7	19.7
8	31.5	23.6	24.0	20.0	19.7
16	30.3	22.5	22.3	19.2	17.7
32	31.6	24.5	22.2	20.4	17.5
64	30.9	23.1	21.6	19.7	17.2
128	30.7	22.4	19.6	18.4	15.5
256	29.9	25.7	19.2	19.1	15.4
512	32.4	23.5	18.2	21.5	15.3
1024	31.6	22.9	17.7	20.7	15.3
2048	31.0	25.3	17.1	17.9	14.1

A partir des résultats obtenus, nous pouvons émettre les remarques suivantes :

- Les taux d'erreurs de segmentation relatifs à l'algorithme *K-means* et aux modèles de CMC classique ne dépendent pas de la valeur de s .
- Plus la valeur de s augmente, plus la non-stationnarité des données est forte (la disparité entre les différentes parties des données est plus prononcée). Comme il s'agit ici de segmentation non-supervisée, les performances des CMEC devancent celles des CMC. Ceci est dû au fait que les CMC évidentielles prennent en considération l'aspect non-stationnaire des données. Pour la même raison, les performances des CMC MN proposées dans ce chapitre sont meilleures que celles des CMC MS.
- Les taux d'erreur des modèles multi-senseurs sont meilleurs que ceux enregistrés par les CMC standards. Cette différence est justifiée par le fait que les premiers utilisent plus d'informations que les seconds.
- L'importance de la prise en compte de la non-stationnarité des données peut être mesurée par rapport à celle de la quantité d'information utilisée (mono-senseurs ou multi-senseurs) à travers la comparaison entre les taux d'erreurs enregistrés par les CMEC et ceux obtenus par les CMC MS. Ainsi, lorsque les données sont fortement non-stationnaires (grandes valeurs de s), le modèle de CMEC fournis de meilleurs résultats que le modèle de CMC MS. En contrepartie, pour les petites valeurs de s où la non-stationnarité des données est modérée, le modèle multi-senseurs CMC MS fournit les meilleurs résultats.
- D'une manière générale, le modèle de CMC MN proposé présente les meilleures performances par rapport à tous ses concurrents. En effet, le modèle proposé utilise toute l'information disponible tout en tenant compte de l'aspect non-stationnaire des données combinant ainsi les avantages du modèles CMEC avec ceux du modèle CMC MS.

7.3.2. Segmentation non-supervisée d'images multi-senseurs non-stationnaire bruitées

Dans cette série d'expériences, nous proposons d'appliquer le modèle de CMC MN au problème de restauration d'images non-stationnaires multi-senseurs. Pour cela nous considérons

les deux images de classes non-stationnaires ‘Nazca bird’ (Fig. 7.1) et ‘Squares’ (Fig. 7.2). Pour rendre notre modèle applicable aux images, ces dernières sont converties de et vers les séquences unidimensionnelles à travers le parcours d’Hilbert-Peano [Pea1890, Hil1891].

Considérons l’image non-stationnaire ‘Nazca bird’ de taille 128×128 ($N = 16384$) qui est une image de $K = 2$ classes ($\Omega = \{\omega_1, \omega_2\}$ où ω_1 et ω_2 correspondent aux pixels noirs et blancs respectivement) que nous bruitons de deux manières différentes de façon à avoir deux observations que nous pourrions fusionner plus tard selon la théorie du modèle CMC MN. Plus explicitement, nous simulons y de la manière suivante :

- Senseur 1 : La réalisation $y^1 = y_{1..N}^1$ est simulée conformément à $p(y^1|x) = p(y_2^1|x_2) \dots p(y_N^1|x_N)$ où pour tout $1 \leq n \leq N$, $p(y_n^1|x_n = \omega_1)$ et $p(y_n^1|x_n = \omega_2)$ sont deux densités gaussiennes de moyennes 0 et 1 respectivement et d’écart-type égaux à 1.
- Senseur 2 : L’image y^2 fournie par ce senseur présente une classe additionnelle dont la présence est régie par le processus B . La réalisation de ce dernier est simulée selon une chaîne de Markov de matrice de transition $J = \begin{pmatrix} 0.998 & 0.002 \\ 0.001 & 0.999 \end{pmatrix}$. Par conséquent, la réalisation du processus auxiliaire U^2 est dérivée comme suit : pour tout $1 \leq n \leq N$, $u_n^2 = \Omega$ si $b_n = 1$ et $u_n^2 = x_n$ sinon. Enfin, l’observation $y^2 = y_{1..N}^2$ est simulée selon $p(y^2|u^2) = p(y_1^2|u_1^2)p(y_2^2|u_2^2) \dots p(y_N^2|u_N^2)$ où pour tout $1 \leq n \leq N$, $p(y_n^2|u_n^2 = \{\omega_1\})$, $p(y_n^2|u_n^2 = \{\omega_2\})$ et $p(y_n^2|u_n^2 = \{\omega_1, \omega_2\})$ sont des densités gaussiennes de moyennes 0, 2 et 4 respectivement et d’écart-type égaux à 1.

Notons que le contraste entre les deux classes ω_1 et ω_2 est plus grand dans l’image y^2 . Toutefois, la non-fiabilité du senseur la fournissant (La présence d’une classe supplémentaire) rend l’application directe des modèles de Markov conventionnels inadaptée. La même situation se présente lorsque l’image est partiellement observable (manque d’observation au niveau de certains sites). Cette difficulté peut être surmontée grâce au modèles évidentiels CMC MS et CMC MN.

La segmentation non-supervisée est ensuite accomplie selon les approches suivantes :

- Sur la base de l’observation y^1 seule, la segmentation non-supervisée est accomplie selon l’algorithme *K-means*, le modèle de CMC et le modèle de CMEC.
- Sur la base de toute l’observation y , la segmentation non-supervisée MPM est accomplie selon le modèle de CMC MS et le modèle de CMC MN.

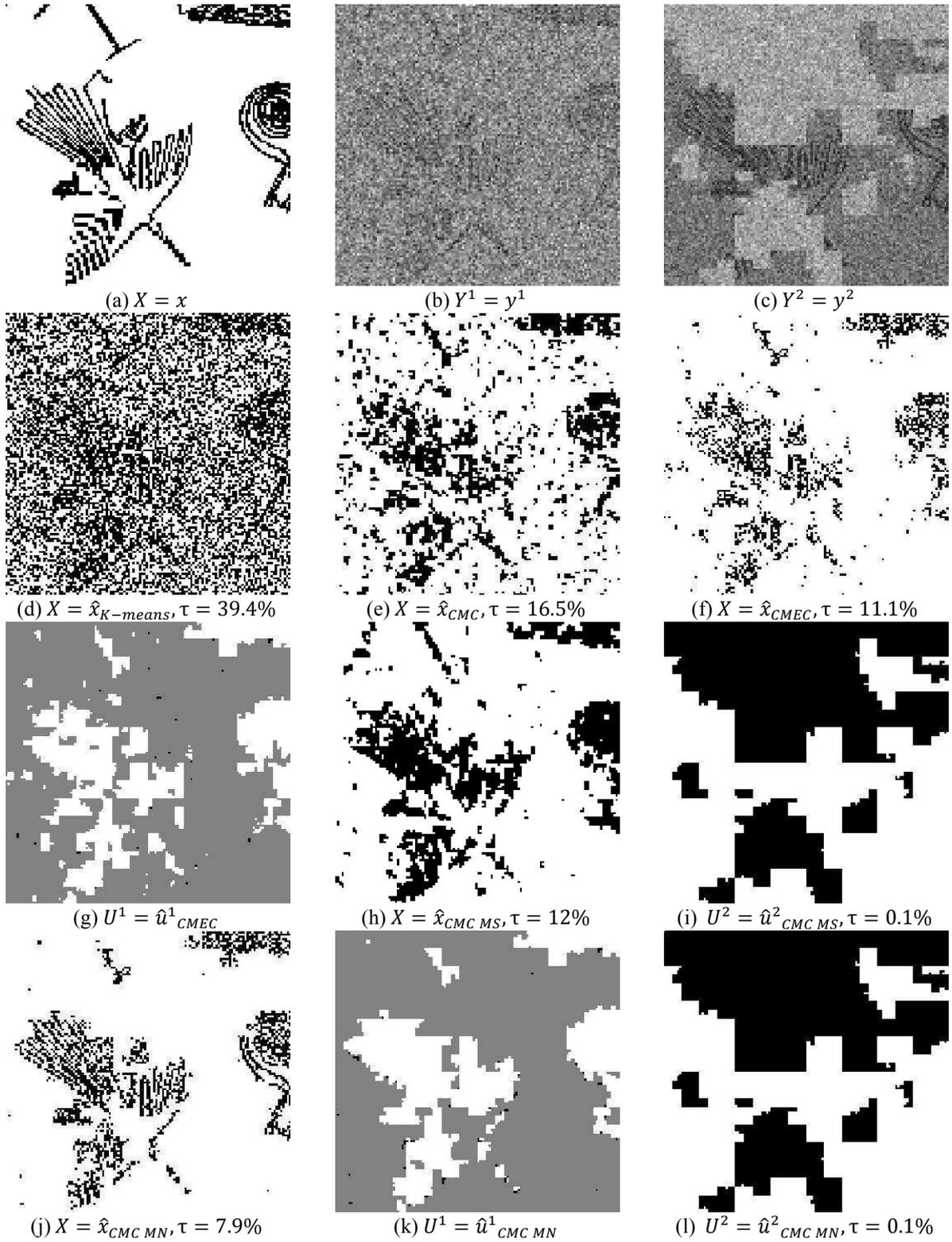


Figure 7.1. Segmentation non-supervisée d'une image multi-senseurs non-stationnaire. (a) Image de classes d'origine $X = x$. (b) Image observée $Y^1 = y^1$. (c) Image observée $Y^2 = y^2$. (d) Segmentation de l'image Y^1 selon K -means, taux d'erreur $\tau = 39.4\%$. (e) Segmentation de l'image Y^1 selon les CMC, taux d'erreur $\tau = 16.5\%$. (f) Segmentation de l'image Y^1 selon les CMEC, taux d'erreur $\tau = 11.1\%$. (g) Estimation de la réalisation du processus auxiliaire U^1 selon une CMEC. (h) Segmentation de l'image selon les CMC MS, taux d'erreur $\tau = 12\%$. (i) Estimation de la réalisation du processus auxiliaire U^2 selon une CMC MS, taux d'erreur $\tau = 0.1\%$. (j) Segmentation de l'image selon les CMC MN, taux d'erreur $\tau = 7.9\%$. (k) Estimation de la réalisation du processus caché U^1 selon une CMC MN. (l) Estimation de la réalisation du processus caché U^2 selon une CMC MN, taux d'erreur $\tau = 0.1\%$.

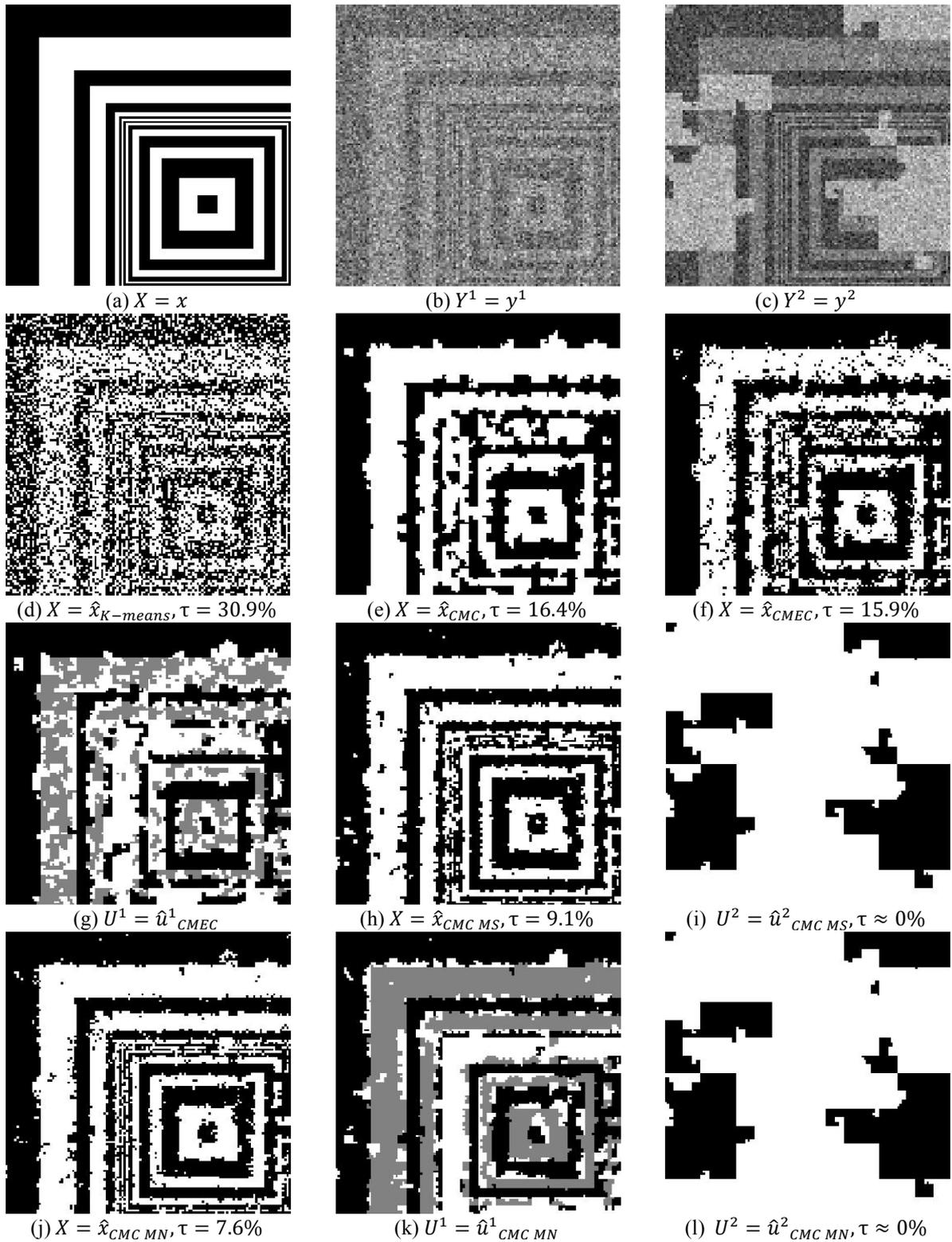


Figure 7.2. Segmentation non-supervisée d'une image multi-senseurs non-stationnaire. (a) Image de classes d'origine $X = x$. (b) Image observée $Y^1 = y^1$. (c) Image observée $Y^2 = y^2$. (d) Segmentation de l'image Y^1 selon *K-means*, taux d'erreur $\tau = 30.9\%$. (e) Segmentation de l'image Y^1 selon les CMC, taux d'erreur $\tau = 16.4\%$. (f) Segmentation de l'image Y^1 selon les CMEC, taux d'erreur $\tau = 15.9\%$. (g) Estimation de la réalisation du processus auxiliaire U^1 selon une CMEC. (h) Segmentation de l'image selon les CMC MS, taux d'erreur $\tau = 9.1\%$. (i) Estimation de la réalisation du processus auxiliaire U^2 selon une CMC MS, taux d'erreur $\tau \approx 0\%$. (j) Segmentation de l'image selon les CMC MN, taux d'erreur $\tau = 7.6\%$. (k) Estimation de la réalisation du processus caché U^1 selon une CMC MN. (l) Estimation de la réalisation du processus caché U^2 selon une CMC MN, taux d'erreur $\tau \approx 0\%$.

Des exemples de segmentation sont illustrés dans Fig. 7.1. A partir des taux moyens correspondants à 20 simulations, nous pouvons constater :

- Lorsque l'algorithme *K-means* est appliqué, seule l'information « observation » est considérée. De ce fait, aucune information contextuelle a priori n'est prise en compte, ce qui justifie le taux d'erreur élevé ($\tau_{K-Means} = 39.4\%$). En effet, toute modélisation omettant l'information contextuelle est très sensible au bruit.

- Pour les CMC classiques, le voisinage de tout site est pris en considération. Toutefois, de tels modèles sont incompatibles avec l'aspect non-stationnaire de l'image ($\tau_{CMC} = 16.5\%$).

- L'application des CMEC permet de remédier au problème cité plus haut ($\tau_{CMEC} = 11.1\%$) à travers l'introduction d'une fonction de masse qui généralise les probabilités bayésiennes *a priori* et prend ainsi en considération la l'imprécision associée à la loi de X . Cependant, il serait intéressant d'exploiter l'image fournie par le second senseur même si ce dernier est moins fiable que le premier.

- Le modèle de CMC MS exploite toute l'observation y , et fournit par conséquent, de meilleurs résultats que le modèle de CMC. Toutefois, ne considérant pas l'aspect non-stationnaire de la loi de X , il fournit un taux d'erreur moyen comparable avec celui enregistré par le modèle de CMEC ($\tau_{CMC MS} = 12\%$).

- Enfin, le modèle proposé CMC MN apporte les meilleurs résultats de tous ($\tau_{CMC MN} = 7.9\%$). Ceci se justifie par le fait qu'il tire profit de toute l'observation y tout en prenant en compte l'aspect non-stationnaire de l'image x . De ce fait, le modèle CMEC peut être vu comme une CMC MN particulière où seul un senseur est disponible, alors que le modèle CMC MS est équivalent à une CMC MN lorsque les données en question sont stationnaires.

- Dans l'image correspondant à la réalisation du processus auxiliaire U^1 (Fig. 7.1-k), la région en blanc correspond au sous-ensemble Ω où la confusion entre les deux classes ω_1 et ω_2 est très grande. En effet, cette partie de l'image se caractérise par une abondance de détails. Rappelons que pour une telle région, l'algorithme *K-means* peut produire de meilleurs résultats que les CMC et les CMC MS. Ceci s'explique par le fait que la régularisation (en ignorant la non-stationnarité de x) dans ces deux modèles dévie le processus de classification. L'intérêt du modèle CMEC réside justement dans l'affaiblissement de l'information « *a priori* » par rapport à l'information « observation » dans de telles régions à forte présence de détails. Pour ces régions-là, notre modèle CMC MN fournit également les meilleurs résultats car, en plus du fait qu'il fait appel à l'affaiblissement de l'information *a priori*, il exploite toute l'observation fournie par les deux senseurs.

Les résultats de segmentation de l'image non-stationnaire '*Squares*', obtenus dans les mêmes conditions que précédemment, confirment ceux enregistrés pour l'image '*Nazca Bird*' avec la différence suivante : le taux d'erreur moyen enregistré par le modèle CMEC ($\tau_{CMEC} = 15.9\%$) est significativement plus élevé que celui du modèle CMC MS ($\tau_{CMC MS} = 9.1\%$). Ceci s'explique par le fait que la non-stationnarité de l'image '*Squares*' est plus modérée que celle de l'image '*Nazca Bird*'. Dans de tels cas, l'importance de la quantité d'information utilisée est plus importante que celle de la prise en compte de la non-stationnarité de l'image. La même

conclusion a été dégagée pour les données synthétiques pour les petites valeurs de s (non-stationnarité modérée). Notons toutefois, que même dans ce cas là, l'application du nouveau modèle est toujours avantageuse ($\tau_{CMC MN} = 7.6\%$ contre $\tau_{CMC MS} = 9.1\%$).

Les paramètres estimés par l'algorithme EM associé à chacun des modèles markovien sont également illustrés dans TAB. 7.II. Les paramètres réels des densités de bruit utilisés lors de la simulation étant connus ($\mu_1^1 = 0, \mu_2^1 = 1, \mu_1^2 = 0, \mu_2^2 = 2, \mu_{1,2}^2 = 4$ et tous les $\sigma_k^s = 1$), nous pouvons vérifier ainsi si les paramètres estimés par le modèle CMC MN sont plus proches des paramètres réels que ceux estimés par les autres modèles.

Tableau 7.2. Paramètres estimés par l'algorithme EM selon les différents modèles de chaînes de Markov de l'exemple de la Fig. 7.1

Paramètres de CMC estimés par l'algorithme EM					
Paramètres des densités de bruit	μ		σ		
	ω_1	0.3593		1.0331	
ω_2	1.1189		0.9526		
Matrice de transition	ω_1		ω_2		
	ω_1	0.7328		0.2672	
ω_2	0.1525		0.8475		
Paramètres de CMEC estimés par l'algorithme EM					
Paramètres des densités de bruit	μ		σ		
	ω_1	-0.0652		0.9498	
ω_2	1.0284		0.9676		
Masse de transition	ω_1		Ω		
	ω_1	0.0105	0.0403	0.0008	
	ω_2	0.0402	0.6726	0.0008	
Ω	0.0008	0.0008	0.2332		
Paramètres de CMC MS estimés par l'algorithme EM					
Paramètres des densités de bruit	μ^1	μ^2	σ^1	σ^2	
	ω_1	0.2964	0.6025	1.0745	1.3182
	ω_2	1.0061	1.9891	0.9831	1.0092
Ω	-	4.0099	-	0.9927	
Masse de transition	ω_1		ω_2		
	ω_1	0.2081		0.0219	
ω_2	0.0219		0.7481		
Paramètres de CMC MN estimés par l'algorithme EM					
Paramètres des densités de bruit	μ^1	μ^2	σ^1	σ^2	
	ω_1	0.0046	-0.0356	0.9994	0.9644
	ω_2	0.9928	1.9893	0.9843	0.9941
Ω	-	4.0099	-	0.9927	
Masse de transition	ω_1		Ω		
	ω_1	0.0074	0.0108	0.0022	
	ω_2	0.0106	0.7096	0.0016	
Ω	0.0024	0.0014	0.2540		

Considérons les différents jeux de paramètres présentés dans TAB. 7.II :

- Selon les CMC, les données fournies par le premier senseur sont supposées stationnaires. Ainsi, l'excès de régularisation dévie le processus d'estimation des paramètres. Le même problème survient avec les CMC MS. En effet, même si ces derniers exploitent les deux

images observées, l'incompatibilité entre le modèle, supposé stationnaire, et les données conduit à l'estimation d'un jeu de paramètres inapproprié.

- Le modèle CMEC prend en compte l'aspect non-stationnaire de l'image. Toutefois, il ne considère que l'image fournie par le premier senseur. De ce fait, même si les paramètres estimés sont proches des paramètres réels, les résultats de segmentation sont limités à cause du bruitage très fort de l'image y^1 .

- Les paramètres estimés par le modèle proposé dans ce chapitre sont les plus proches des paramètres réels. Par ailleurs, nous pouvons mesurer la différence entre les deux modélisations multi-senseurs (CMC MS et CMC MN) à travers les deux jeux de paramètres estimés. Plus explicitement, vu que le modèle CMC MS est un CMC MN particulier où la masse m s'annule en dehors des singletons $\{\omega_1\}$ et $\{\omega_2\}$, la relativement grande probabilité associée à $m(\{\omega_1, \omega_2\}, \{\omega_1, \omega_2\})$ peut être vue comme une mesure d'inadéquation du modèle CMC MS par rapport à l'image considérée ici.

7.4. Conclusion

Dans ce chapitre, nous avons proposé une nouvelle approche pour la modélisation des données multi-senseurs non-stationnaires dans le contexte markovien. Cette approche permet de bénéficier simultanément des avantages de la théorie markovienne et de la théorie de l'évidence. Ainsi, la règle de combinaison de Dempster-Shafer a été utilisée à deux fins différentes : prendre en considération la non-stationnarité de la loi *a priori* du processus caché, et fusionner les données fournies par les différents senseurs dans le contexte markovien permettant ainsi d'améliorer les performances des modélisations classiques. L'étude expérimentale menée sur des données synthétiques et des images bruitées a confirmé l'intérêt du modèle proposé par rapport aux modèles existants dans la littérature. Comme perspective, nous pouvons considérer des modèles graphiques plus complexes tels que les arbres de Markov pour la modélisation des images multi-résolution [Bou94, Wil02, Mon03, Mon11].

Conclusion Générale

Le contexte général de cette étude était de proposer des extensions aux modèles de chaînes de Markov cachées (CMC) mieux adaptées aux données de nature complexe. Le cadre illustratif choisi fut la segmentation des données. De ce fait, trois sources de complexité ont été abordées : la non-stationnarité du processus caché et/ou du bruit, la corrélation du bruit et la multitude de sources de données. Pour contourner ces difficultés, nous avons fait appel à la très riche théorie des modèles de Markov triplet (MMT). En notant les processus caché et observables X et Y respectivement, nous pouvons énumérer les différentes contributions de cette thèse comme suit :

Pour les données non-stationnaires, nous avons proposé des modélisations originales considérant la non-stationnarité du bruit. Ainsi, la non-stationnarité de la loi $p(y|x)$ a été traitée d'une manière analogue à celle relative à la distribution *a priori* $p(x)$ dans les modèles de chaînes de Markov caché M-stationnaire (CMC M-S). En conséquence, nous avons défini les chaînes de Markov cachées à bruit M-stationnaire [Bou12c] et les chaînes de Markov cachées ML-stationnaire [Bou11a].

Pour le bruit corrélé, nous avons étendu les CMC M-S et les chaînes de Markov évidentielles cachées (CMEC) au contexte des chaînes de Markov couples (CMCo). Ainsi, les CMC M-S ont été étendues aux CMCo M-S [Bou11c] et CMCo MM-S [Bou11b]. Par ailleurs, les CMEC ont été étendues aux chaînes de Markov couples évidentielles [Bou12b].

Pour les données multi-senseurs, nous avons proposé une modélisation originale [Bou12a] exploitant la théorie de l'évidence à deux fins différentes : la prise en compte de la non-stationnarité de la loi *a priori* $p(x)$ et la fusion Dempster-Shafer des différents senseurs.

Pour chacune des modélisations proposées, une étude expérimentale a été menée sur des données synthétiques et/ou réelles pour démontrer son intérêt par rapport à ses devancières.

Cette thèse a également constitué une occasion pour essayer d'analyser finement des modélisations précédentes. En sus des apports cités plus haut et la compréhension des travaux antérieurs, une étude comparative entre les CMC M-S et les CMEC a été menée dans le contexte des données non-stationnaires. Cette étude nous a permis de mieux cerner l'intérêt que suscite chacune des deux modélisations. Les CMC M-S sont plus utiles lorsque les données sont régies par un système fluctuant entre plusieurs régimes fixes et dénombrables. De ce fait la connaissance *a priori* du nombre de ces régimes est primordiale pour les différents traitements. En absence de toute information, le recours aux CMEC semble être une bonne solution, notamment si la distribution *a priori* fluctue d'une manière continue et imprévisible.

Enfin, la richesse de la théorie des MMT perçue le long de l'évolution des différents travaux de recherches menés dans le cadre de cette thèse ouvre différentes perspectives pour des travaux futurs.

En premier lieu, il semble très naturel et intuitif d'étendre les différentes modélisations, proposées dans cette thèse dans le contexte des CMC, à des modèles graphiques plus généraux tels que les arbres et les champs de Markov.

Une autre perspective consisterait à considérer d'autres types de données complexes. A titre d'exemple, nous pouvons citer les séquences biologiques telles que celles de l'ADN qui sont régies par des normes qui relèvent du traitement de langage et pour lesquelles les premiers résultats obtenus par l'application des CMT semblent très prometteurs.

Enfin, l'application des modélisations s'appuyant sur des modèles de Markov triplets à des problèmes réels semble être une continuité naturelle de cette thèse. Notons que de telles modélisations ont apporté des résultats prometteurs dans les domaines de segmentation [Zha12a, Zha12b] et de détection de changement [Wan13] en imagerie SAR et la classification d'images médicales [Bri06].

Bibliographie

- [Aas99] K. Aas, L. Eikvil, R.B. Huseby, "Applications of Hidden Markov Chains in Image Analysis," *Pattern Recognition*, vol. 32, no. 4, pp. 703-713, Apr. 1999.
- [Ait06] B. Ait-el-Fquih, F. Desbouvries, "Kalman filtering in triplet Markov chains," *IEEE Trans. Signal Processing*, Vol. 54, No. 8, pp. 2957-63, Aug. 2006.
- [App91] A. Appriou, "Probabilités et incertitude en fusion de données multi-senseurs," *Revue Scientifique et Technique de La Défense*, 1^{er} trimestre, 1991.
- [Bau70] L.E. Baum, T. Petrie, G. Soules, N. Weiss, "A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains," *Ann. Math. Statistics*, Vol. 41, No. 1, pp. 164-171, Feb. 1970.
- [BeM07] M. Ben Mabrouk, W. Pieczynski, "Unsupervised segmentation of random discrete data using triplet Markov chains," *International Symposium on Applied Stochastic Models and Data Analysis*, ASMDA, Crete, Greece, May 2007.
- [BeM11] Mohamed Ben Mabrouk, Modèles de Markov triplets en restauration des signaux, thèse de Telecom SudParis et de l'Université Pierre et Marie Curie, 2011.
- [Ben95] B. Benmiloud, W. Pieczynski, "Estimation des paramètres dans les chaînes de Markov cachées et segmentation d'images," *Traitement du Signal*, Vol. 12, No. 5, pp.433-454, 1995.
- [Ben01] A. Bendjebbour, Y. Delignon, L. Fouque, V. Samson, W. Pieczynski, "Multisensor Images Segmentation Using Dempster-Shafer Fusion in Markov Fields Context," *IEEE Trans. Geoscience and Remote Sensing*, Vol. 39, No. 8, pp. 1789-1798, 2001
- [Ben05] D. Benboudjema, W. Pieczynski, "Unsupervised image segmentation using triplet Markov fields," *Computer Vision and Image Understanding*, Vol.99, No. 3, pp. 476-498, 2005.
- [Ben07] D. Benboudjema, W. Pieczynski, "Unsupervised statistical segmentation of non stationary images using triplet Markov fields," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 29, No. 8, pp. 1367-1378, Aug. 2007.
- [Ben13a] D. Benboudjema, N. Othman, B. Dorizzi, W. Pieczynski, "Challenging eye segmentation using triplet Markov spatial models," *International Conference on Acoustics, Speech and Signal Processing (ICASSP 2013)*, 26-31 May, Vancouver, Canada, 2013.
- [Ben13b] D. Benboudjema, N. Othman, B. Dorizzi, W. Pieczynski, "Segmentation d'images des yeux par champs de Markov triplets: Applications à la biométrie," *GRETSI 2013*, 3-6 septembre, Brest, France, 2013.
- [Ber06] F. Le Ber, M. Benoît, C. Schott, J.-F. Mari, C. Mignolet, "Studying crop sequences with Carrot Age, a HMM based data mining software", *Ecological Modelling*, Vol. 191, No. 1, 2006.
- [Bes86] J. Besag, "On the Statistical Analysis of Dirty Pictures," *J. Royal Statistical Soc., Series B*, vol. 48, pp. 259-302, 1986.
- [Bha04] R. Bhar, S. Hamori, Hidden Markov models: applications to financial economics, Springer, 2004.
- [Bla07] J. Blanchet, Modèles Markoviens et extensions pour la classification de données complexes, Thèse en Mathématiques appliquées, Université Joseph Fourier, 2007.
- [Bla08] J. Blanchet, F. Forbes, "Triplet Markov Fields for the Classification of Complex Structure Data," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 30, No. 6, pp. 1055-1067, Jun. 2008.

- [Blo96] I. Bloch, "Some aspects of Dempster–Shafer evidence theory for classification of multi-modality medical images taking partial volume effect into account," *Pattern Recognition Letters*, Vol. 17, No. 8, 1996.
- [Bou94] Bouman, C., Shapiro, M., "A Multiscale Random Field Model for Bayesian Image Segmentation," *IEEE Transactions on Image Processing*, Vol. 3, No. 2, pp. 162- 177, 1994.
- [Bou10] M. Y. Boudaren, A. Belaïd, "A New Scheme for Land Cover Classification in Aerial Images", *Electronic Engineering and Computing Technology, Lecture Notes in Electrical Engineering*, Chapter 40, Volume 60, pp. 471-482, Springer 2010.
- [Bou11a] M. Y. Boudaren, W. Pieczynski, E. Monfrini, "Unsupervised segmentation of non stationary data hidden with non stationary noise," *IEEE International Workshop on Systems, Signal Processing and their Applications*, Tipasa, Algeria, May 2011.
- [Bou11b] M. Y. Boudaren, E. Monfrini, W. Pieczynski, "Switching pairwise Markov chains for non stationary textured images segmentation," in *Proceedings of IADIS Int. Conf. Computer Graphics, Visualization, Computer Vision and Image Processing*, pp. 11-18, Rome, Jul. 2011.
- [Bou11c] M. Y. Boudaren, E. Monfrini, W. Pieczynski, "Unsupervised segmentation of switching pairwise Markov chains," in *Proceedings of IEEE Int. Sym. On Image and Signal Processing and Analysis*, Dubrovnik, Croatia, 4-6 Sep. 2011.
- [Bou12a] M. Y. Boudaren, E. Monfrini, W. Pieczynski, A. Aïssani, "Dempster-Shafer fusion of multisensor signals in nonstationary Markovian context," *EURASIP J. Advances in Signal Processing*, No. 134, Jul. 2012.
- [Bou12b] M. Y. Boudaren, E. Monfrini, W. Pieczynski, "Unsupervised segmentation of nonstationary pairwise Markov chains using evidential priors," *European Signal Processing Conf. (EUSIPCO 2012)*, pp. 2243-2247, Aug. 2012.
- [Bou12c] M. Y. Boudaren, E. Monfrini, W. Pieczynski, "Unsupervised segmentation of random discrete data hidden with switching noise distributions," *IEEE Signal Processing Letters*, Vol. 19, No. 10, pp. 619-622, Oct. 2012.
- [Bri06] S. Bricq, C. Collet, J.-P. Armspach, "Triplet Markov chain for 3D MRI brain segmentation using a probabilistic atlas," *Proc. IEEE Int. Symp. Biomedical Imaging*, Apr. 2006.
- [Bro66] Brodatz, P., *A Photographic Album for Artists and Designers, Textures*, Dover Publications, New York 1966.
- [Bro83] M. Broniatowski, G. Celeux, J. Diebolt, Reconnaissance de mélanges de densités par un algorithme d'apprentissage probabilistes, *In Data Analysis and Informatics 3, E. Diday (Ed.)*, North Holland, Amsterdam, pp. 359-373, 1983.
- [Cap05] O. Cappé, E. Moulines, T. Ryden, *Inference in Hidden Markov Models*, Springer, Series in Statistics, 2005.
- [Car06] C. Carincotte, S. Derrode, S. Bourennane, "Unsupervised change detection on SAR images using fuzzy hidden Markov chains," *IEEE Trans. Geosciences and Remote Sensing*, Vol. 44, No.2, pp. 432–441, Feb. 2006.
- [Cel85] G. Celeux, D. Diebolt, "The SEM algorithm: A probabilistic teacher algorithm derived from the EM algorithm for the mixture problem," *Computational Statistic Quarterly*, Vol. 2, pp.73-82, 1985.
- [Cel88] G. Celeux, D. Diebolt, A random imputation principle: the stochastic EM algorithm, Technical report 901, INRIA Roquencourt, Sep. 1988.

- [Cel96] G. Celeux, D. Chauveau, J. Diebolt, "Stochastic versions of the EM algorithm: an experimental study in the mixture case," *J. Statistical Computation and Simulation*, Vol. 55, No. 4, pp. 287–314, 1996.
- [CGT11] <http://www.cgtextures.com>, 2011.
- [Che94] M.-Y. Chen, A. Kundu, J. Zhou, "Off-Line Handwritten Word Recognition Using a Hidden Markov Model Type Stochastic Network," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 16, no. 5, pp. 481- 496, May 1994.
- [Chu92] G.A. Churchill, "Hidden Markov Chains and the Analysis of Genome Structure," *Computers and Chemistry*, vol. 16, no. 2, pp. 107-115, Apr. 1992.
- [Cob03] B. R. Cobb, P. P. Shenoy, *A Comparison of Methods for Transforming Belief Function Models to Probability Models*, Springer- Verlag, 2003.
- [Dai94] J. Dai, "Hybrid Approach to Speech Recognition Using Hidden Markov Models and Markov Chains," *IEE Proc. Vision, Image, and Signal Processing*, Vol. 141, no. 5, pp. 273-279, 1994.
- [Del97] J. -P. Delmas, "An equivalence of the EM and ICE algorithm for exponential family," *IEEE Trans. on Signal Processing*, Vol.45, No.10, pp. 2613–2615, 1997.
- [Dem67] A. P. Dempster, "Upper and lower probabilities induced by a multivalued mapping," *The Annals of Mathematical Statistics*, Vol. 38, No. 2, pp. 325–339. 1967.
- [Dem77] A. P. Dempster, N. M. Laird, D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm (with discussion)," *Journal of the Royal Statistical Society B*, Vol. 39, pp. 1–38, 1977.
- [Den97] T. Denoeux, "Analysis of evidence-theoretic decision rules for pattern classification," *Pattern Recognition*, Vol. 30, No.7, 1997.
- [Der04] S. Derrode, W. Pieczynski, "Signal and image segmentation using Pairwise Markov chains," *IEEE Trans. Signal Processing*, Vol. 52, No. 9, pp. 2477-2489, Sep. 2004.
- [Der11] S. Derrode, "On-line demos for recent extensions of HMC model," <http://www.fresnel.fr/perso/hmcext/>. 2011.
- [Dju02] M. Djuric, J.-H. Chun, "An MCMC sampling approach to estimation of nonstationary hidden Markov models," *IEEE Trans. Signal Processing*, vol. 50, no. 5, pp. 1113–1123, May 2002.
- [Eph02] Y. Ephraim, "Hidden Markov processes," *IEEE Trans. Information Theory*, Vol. 48, No. 6, pp.1518–1569, Jun. 2002.
- [Fjo03]. R. Fjortoft, Y. Delignon, W. Pieczynski, M. Sigelle, F. Tupin, "Unsupervised classification of radar images using hidden Markov chains and hidden Markov random fields," *IEEE Trans. Geoscience Remote Sensing*, Vol. 41, No. 3, pp. 675- 686, 2003.
- [For73]G. D. Forney Jr., "The Viterbi Algorithm," *Proc. IEEE*, vol. 61, no. 3, pp. 268- 277, Mar. 1973.
- [Fou00]. L. Fouque, A. Appriou, W. Pieczynski, "An evidential markovian model for data fusion and unsupervised image classification," in *Proceedings of 3rd International Conference on Information Fusion, FUSION 2000*, Paris, France, July 2000.
- [Gan12] L. Gan, Y. Wu, M. Liu, P. Zhang, H. Ji, F. Wang, "Triplet Markov fields with edge location for fast unsupervised multi-class segmentation of synthetic aperture radar images," *IET Image Processing*, Vol. 6, No. 7, pp. 831 – 838, Oct. 2012.
- [Gem84] S. Geman, D. Geman, "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 6, pp. 721-741, 1984.

- [Hat07] M. Hatt, F. Lamare, N. BouSSION, A. Turzo, C. Collet, F. Salzenstein, C. Roux, P. Jarritt, K. Carson, C. Cheze-Le Rest and D. Visvikis, "Fuzzy hidden Markov chains segmentation for volume determination and quantitation in PET," *Physics in Medicine and Biology*, Vol. 52, No. 12, pp. 3467-3491, Jun. 2007.
- [Heg98] S. L. Hégarat-Masclé, I. Bloch, 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*, Vol. 31, No. 11, pp. 1811-1823, 1998.
- [Hil1891] D. Hilbert, "Über die stetige Abbildung einer Linie auf ein Flächenstück," *Math. Ann.*, Vol. 38, pp. 459-460, 1891.
- [Hil11] A. Hillion, W. Pieczynski, *Taritement statistique des images*, Cours Master 2, Université Pierre et Marie Curie, 2011.
- [Jac99] A. El-Jacoubi, M. Gilloux, R. Sabourin, C.Y. Suen, "An HMM-Based Approach for Off-Line Unconstrained Handwritten Word Modeling and Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 21, no. 8, pp. 752-760, Aug. 1999.
- [Jan98]. F. Janez, A. Appriou, "Theory of evidence and non-exhaustive frames of discernment: plausibilities correction methods," *Intern. J Approximate Reasoning*, Vol. 18, No. 1, 1998.
- [Kal94] G. K. Kaleh, R. Vallet, "Joint Parameter Estimation and Symbol Detection for Linear or Nonlinear Unknown Channels," *IEEE Trans. Communications*, vol. 42, no. 7, pp. 2406-2413, Jul. 1994.
- [Kos01] T. Koski, *Hidden Markov Models for Bioinformatics*, Kluwer Academic Publishers, 2001.
- [Lan05] P. Lanchantin, W. Pieczynski, "Unsupervised restoration of hidden nonstationary Markov chain using evidential priors," *IEEE Trans. on Signal Processing*, Vol. 53, No. 8, pp. 3091-3098, Aug. 2005.
- [Lan06] P. Lanchantin, *Chaînes de Markov triplets et segmentation non-supervisée de signaux*, Thèse de l'Institut national de Télécommunications, Evry 2006.
- [Lan11] P. Lanchantin, J. Lapuyade-Lahorgue, W. Pieczynski, "Unsupervised segmentation of randomly switching data hidden with non-Gaussian correlated noise," *Signal Processing*, Vol. 91, No. 2, pp. 163-175, Feb. 2011.
- [Lap12] J. Lapuyade-Lahorgue, W. Pieczynski, "Unsupervised segmentation of hidden semi-Markov non stationary chains," *Signal Processing*, Vol. 92, No. 1, pp. 29-42, Jan. 2012.
- [Lia07] K. Liang, X. Wang, D. Anastassiou, "Bayesian Basecalling for DNA Sequence Analysis Using Hidden Markov Models," *IEEE/ACM Trans. Computational Biology and Bioinformatics*, Vol. 4, No. 3, pp. 430 - 440, Jul/Sep 2007.
- [McL97] G. J. McLachlan, T. Krishnan, *EM Algorithm and Extensions*, Wiley, Series in Probability and Statistics, 1997.
- [Mon03] E. Monfrini, J. Lecomte, F. Desbouvries, W. Pieczynski, "Image and Signal Restoration using Pairwise Markov Trees," *IEEE Workshop on Statistical Signal Processing (SSP 2003)*, Saint Louis, Missouri, Sep.-Oct., 2003.
- [Mon11] E. Monfrini, W. Pieczynski, "Arbres de Markov couple et segmentation non supervisée d'images: cas de la mono-résolution," *GRETSI 2011*, 5-8 septembre 2011, Bordeaux, France, 2011.
- [Ngu12] T. M. Nguyen, Q. M. J. Wu, "Gaussian-Mixture-Model-Based Spatial Neighborhood Relationships for Pixel Labeling Problem," *IEEE. Trans. Systems, Man, and Cybernetics, Part B: Cybernetics*, Vol. 42, No. 1, pp. 193-202, Feb. 2012.

- [Pea1890] G. Peano, "Sur une courbe, qui remplit une aire plane," *Math. Ann.*, Vol. 36, pp. 157-160, 1890.
- [Pen95] A. Peng, W. Pieczynski, "Adaptive mixture estimation and unsupervised local bayesian image segmentation," *Graphical models and Image Processing*, Vol. 57, No. 5, pp. 389-399, 1995.
- [Pie92] W. Pieczynski, "Statistical image segmentation," *Machine Graphics and Vision*, Vol. 1, No. 1/2 (Proceedings of GKIP0'92), pp. 261-268, 1992.
- [Pie94] W. Pieczynski, "Champs de Markov cachés et estimation conditionnelle itérative," *Traitement du Signal*, Vol. 11, No. 2, pp. 141-153, 1994.
- [Pie02a] W. Pieczynski, "Chaînes de Markov Triplet, Triplet Markov Chains," *Comptes Rendus de l'Académie des Sciences – Mathématique*, Série I, Vol. 335, No. 3, pp. 275-278, 2002.
- [Pie02b] W. Pieczynski, C. Hulard, T. Veit, "Triplet Markov chains in hidden signal restoration," *Proc. SPIE's Int. Sym.. Remote Sensing*, Vol. 4885, pp. 58-68, Sep. 2002.
- [Pie03a] W. Pieczynski, "Pairwise Markov chains," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 25, No. 5, pp. 634-639, 2003.
- [Pie03b] W. Pieczynski, "Modèles de Markov en traitement d'images," *Traitement du Signal*, Vol. 20, No. 3, pp. 255-278, 2003.
- [Pie06] W. Pieczynski, D. Benboudjema, "Multisensor triplet Markov fields and theory of evidence", *Image Vision Computing*, Vol. 24, No. 1, pp. 61-69, 2006.
- [Pie07] W. Pieczynski, "Multisensor triplet Markov chains and theory of evidence", *Int. J. Approximate Reasoning*, Vol. 45, No. 1, pp. 1-16, May 2007.
- [Pie10a] W. Pieczynski, "EM and ICE in hidden and triplet Markov models", in *Proceedings of Stochastic Modeling and Data Analysis International Conference*, Chania, 8-11 June, 2010.
- [Pie10b] Pieczynski, W., Triplet Markov chains and image segmentation, *Inverse problems in Vision and 3D Tomography*, Chapter 4, A. Mohammed-Djafari Ed., Wiley, 2010.
- [Pie11] W. Pieczynski, "Exact filtering in conditionally Markov switching hidden linear models", *Comptes Rendus Mathématiques*, Vol. 349, No. 9-10, pp. 587-590, 2011.
- [Qia89] W. Qian, D.M. Titterton, "On the Use of Gibbs Markov Chain Models in the Analysis of Images Based on Second-Order Pairwise Interactive Distributions," *J. Applied Statistics*, vol. 16, no. 2, pp. 267-282, 1989.
- [Rab89] L. R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257- 286, Feb. 1989.
- [Ram09] E. Ramasso, "Contribution of belief functions to HMM with an application to fault diagnosis", *IEEE International Workshop on Machine Learning and Signal Processing*, Grenoble, France, Sept. 2-4, pp. 1-6, 2009.
- [Sin95] B. Sin, J. H. Kim, "Nonstationary hidden Markov model," *Signal Processing*, Vol. 46, pp. 31-46, 1995.
- [Sha76] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University Press, Princeton NJ, 1976.
- [Sha03] M. M. Shafiei, H. R. Rabiee, "A new online signature verification algorithm using variable length segmentation and hidden Markov models," *Seventh International Conference on Document Analysis and Recognition*, Vol. 1, pp. 443-446, Aug. 2003.

- [Sme93]. P. Smets, "Belief functions: the disjunctive rule of combination and the generalized Bayesian theorem", *Intern. J Approximate Reasoning*, Vol. 9, No. 1, 1993.
- [Sou10] H. Soubaras, "On Evidential Markov Chains, Foundations of Reasoning under Uncertainty," *Studies in Fuzziness and Soft Computing*, Springer, Vol. 249, pp. 247-264, 2010.
- [Sol06] J. Solomon, J. A. Butman, A. Sood, "Segmentation of brain tumors in 4D MR images using the hidden Markov model," *Computer Methods and Programs in Biomedicine*, Vol. 84, No. 2, pp. 76-85, Sep. 2006.
- [Sud03]. J. J. Sudano, "Equivalence between belief theories and naive Bayesian fusion for systems with independent evidential data", in *Proceedings of the Sixth International Conference on Information Fusion*, 2003.
- [Tho02] L.C. Thomas, D.E. Allen, N. Morkel-Kingsbury, "A hidden Markov chain model for the term structure of bond credit risk spreads", *Intern. Review Financial Analysis*, Vol. 11, No. 3, 2002.
- [Wan13] F. Wang, Y. Wu, Q. Zhang, P. Zhang, M. Li, Y. Lu, "Unsupervised Change Detection on SAR Images Using Triplet Markov Field Model," *IEEE Geoscience and Remote Sensing Letters*, Vol. 10, No. 4, pp. 697–701, Jul. 2013.
- [Wil02]. A. S. Willsky, "Multiresolution Markov models for signal and image processing", *Proceedings of IEEE*, Vol. 90, No. 8, pp. 1396-1458, 2002.
- [Yag94]. R. R. Yager, J. Kacprzyk, M. Fedrizzi. *Advances in the Dempster- Shafer theory of evidence*, Wiley, 1994.
- [You03] S.-Z. You, H. Kobayashi, "A hidden semi-Markov model with missing data and multiple observation sequences for mobility tracking," *Signal Processing*, vol. 83, pp. 235–250, 2003.
- [Zha12a] P. Zhang, M. Li, Y. Wu, M. Liu, F. Wang, L. Gan , "SAR Image Multiclass Segmentation Using a Multiscale TMF Model in Wavelet Domain," *IEEE Trans. on Geoscience and Remote Sensing Letters*, Vol. 9 , No. 6, pp. 1099 –1103, 2012.
- [Zha12b] P. Zhang, M. Li, Y. Wu, L. Gan, M. Liu, F. Wang, G. Liu, "Unsupervised multi-class segmentation of SAR images using fuzzy triplet Markov fields model," *Pattern Recognition*, Vol. 45, No. 11, pp. 4018-4033, 2012.