

# Chaînes et arbres de Markov évidentiels avec applications à la segmentation des processus non stationnaires

Evidential Markov chains and trees  
with applications to non stationary processes segmentation

**Pierre Lanchantin, Wojciech Pieczynski**

GET/INT, Département CITI, UMR-CNRS 5157, Institut National des Télécommunications, 9 rue Charles Fourier, 91011 Evry, France,  
Pierre.Lanchantin(Wojciech.Pieczynski)@int-evry.fr

Manuscrit reçu le 11 février 2004

Résumé et mots clés

Les chaînes de Markov Triplet (CMT) généralisent les chaînes de Markov Couple (CMCouple), ces dernières généralisant les chaînes de Markov cachées (CMC). Par ailleurs, dans une CMC la loi *a posteriori* du processus caché, qui est de Markov, peut être vue comme une combinaison de Dempster de sa loi *a priori*  $p$  avec une probabilité  $q$  définie à partir des observations. Lorsque l'on se place dans le contexte de la théorie de l'évidence en remplaçant  $p$  par une fonction de masse  $m$ , sa combinaison de Dempster avec  $q$  généralise ainsi la probabilité *a posteriori*. Bien que le résultat de cette fusion ne soit pas nécessairement une chaîne de Markov, il a été récemment établi qu'il est une CMT, ce qui autorise les divers traitements d'intérêt. De plus, les résultats analogues restent valables lorsque l'on généralise les différentes chaînes de Markov aux arbres de Markov. Nous proposons d'étendre ces résultats aux arbres de Markov Couple, dans lesquels la loi du processus caché n'est pas nécessairement de Markov. Nous montrons également l'intérêt pratique de ce type de fusion dans la segmentation non supervisée des chaînes de Markov non stationnaires, avec application à la segmentation d'images.

Chaînes de Markov évidentielles, arbres de Markov évidentiels, segmentation bayésienne d'images, Espérance-Maximisation, règle de combinaison de Dempster, théorie de l'évidence, chaînes de Markov couple, chaînes de Markov triplet, processus de Markov cachés non stationnaires.

Abstract and key words

The triplet Markov chains (TMC) generalize the pairwise Markov chains (PMC), and the latter generalize the hidden Markov chains (HMC). Otherwise, in an HMC the posterior distribution of the hidden process can be viewed as a particular case of the so called "Dempster's combination rule" of its prior Markov distribution  $p$  with a probability  $q$  defined from the observations. When we place ourselves in the theory of evidence context by replacing  $p$  by a mass function  $m$ , the result of the Dempster's combination of  $m$  with  $q$  generalizes the conventional posterior distribution of the hidden process. Although this result is not necessarily a Markov distribution, it has been recently shown that it is a TMC, which renders traditional restoration methods applicable. Further, these results remain valid when replacing the Markov chains with Markov trees. We propose to extend these results to Pairwise Markov trees. Further, we show the practical interest of such combination in the unsupervised segmentation of non stationary hidden Markov chains, with application to unsupervised image segmentation.

Evidential Markov chains, evidential Markov trees, Bayesian image segmentation, Expectation-Maximization, Dempster's combination rule, theory of evidence, pairwise Markov chains, triplet Markov chains, non stationary hidden Markov processes.

# 1. Introduction

Considérons  $X = (X_1, \dots, X_N)$  un processus inobservable, chaque  $X_n$  prenant ses valeurs dans un ensemble fini de classes  $\Omega = \{\omega_1, \dots, \omega_K\}$ , et  $Y = (Y_1, \dots, Y_N)$  un processus observé, chaque  $Y_n$  prenant ses valeurs dans  $\mathbb{R}$ . Le problème consiste à estimer le processus « caché »  $X$  à partir du processus observé  $Y$ . Nous traitons dans cet article de deux types de modélisation de la loi de  $(X, Y)$  permettant la mise en œuvre des estimations bayésiennes : modèle de type « chaîne de Markov cachée » et modèle de type « arbre de Markov caché ». Les chaînes de Markov cachées (CMC) sont couramment utilisées dans des problèmes de restauration ou de segmentation de signaux. En particulier, on peut les appliquer à la segmentation d'images [3], [15] après transformation de l'image 2D en signal 1D. Bien que les modèles par « champs de Markov cachés » [24], [31], [44] soient, *a priori*, mieux adaptés à l'imagerie, les chaînes procurent des méthodes plus rapides et peuvent, dans certaines situations, donner des résultats comparables [38]. L'utilisation des arbres de Markov cachés (AMC) est plus récente en imagerie [7], [19], [25], [26], [35] et peut également, de par la rapidité des méthodes de traitements correspondantes, concurrencer l'utilisation des champs de Markov (voir la synthèse récente [43] avec une riche bibliographie).

Les CMC et AMC ont été récemment généralisés aux modèles de Markov « Couple » [12], [28], [32] et « Triplet » [11], [21], [29], [30] (voir également [33], [34] pour des généralisations analogues concernant les champs de Markov). Ces deux généralisations successives sont strictes et leur intérêt est de présenter les mêmes possibilités de traitements dans des contextes de généralité, et donc de complexité, croissante des bruits. Les premières applications des modèles Couple et Triplet, qui permettent de tenir compte de bruits complexes et d'utiliser, en particulier, des copules, donnent des résultats encourageants [4], [10], [23].

Par ailleurs, la théorie de l'évidence [1], [6], [9], [14], [17], [39], [40], [41], [45] offre un cadre pouvant servir, dans de nombreuses situations et de manière particulièrement élégante, à certaines extensions du cadre classique du calcul des probabilités. En particulier, cette théorie s'avère utile dans le cadre des champs de Markov cachés, notamment dans le cas multisenseur [2], [13], [16]. En effet, on peut alors facilement tenir compte du fait que certains senseurs sont insensibles à certaines classes, ou encore on peut modéliser la présence des classes parasites, comme les nuages dans les images optiques.

Une autre possibilité offerte par cette théorie est de remplacer, lorsqu'elles sont connues avec insuffisamment de précision, des mesures de probabilité par des fonctions de masse. Ce type de remplacement, dont l'objectif général est l'amélioration de la qualité des divers traitements, est également possible dans le contexte des CMC classiques ; en effet, il a été montré que le résultat de la combinaison de Dempster, qui peut être interprétée comme une généralisation de la loi *a posteriori*, d'une « masse évidentielle » (qui étend la loi *a priori*) avec une pro-

babilité définie par les observations est une CMT [22], [29]. Par la suite, ce résultat a été généralisé aux cas où la CMC classique est remplacée par CM Couple ou Triplet [27]. Notons qu'il s'agit d'une utilisation de CMT très différente de celle mentionnée dans le paragraphe précédent. En effet, cette dernière est une extension, dans le cadre probabiliste, des CMC, ce qui permet de considérer des situations plus générales et, en particulier, de prendre en compte des bruitages complexes. Le fait que la combinaison de Dempster soit formellement une CMT est un résultat indépendant du premier, avec des applications différentes dont l'une, traitée dans cet article, est la gestion de la non stationnarité du processus caché. Cependant, le fait qu'une CMT permette des segmentations non supervisées du processus caché est, bien entendu, d'une importance déterminante dans les deux cas.

L'objet de cet article, qui est une extension de l'article « Arbres de Markov triplet et théorie de l'évidence » [20] paru dans les Actes du Colloque GRETSI2003, est double. D'une part, nous présentons un apport théorique, relatif à l'introduction d'une « masse évidentielle » dans les arbres de Markov couple. D'autre part, nous continuons l'exploration des chaînes de Markov évidentielles et présentons des résultats d'expérimentations originales ; en particulier, le modèle est appliqué à la segmentation non supervisée d'une image réelle.

L'organisation de l'article est la suivante. Dans la section suivante nous présentons brièvement les outils généraux de la théorie de l'évidence et introduisons le modèle de l'affaiblissement. La section 3 est consacrée aux rappels du modèle récent de chaîne de Markov cachée évidentielle, qui sont accompagnés de quelques résultats originaux de simulation. Le modèle original d'arbre de Markov couple évidentiel, généralisant l'arbre de Markov caché évidentiel, est présenté dans la section 4. Quelques résultats de simulations et de segmentations non supervisées d'images occupent la section 5, et la dernière section 6 contient les conclusions et les perspectives.

## 2. La théorie de l'évidence

### 2.1. Généralités

La théorie de l'évidence a été introduite par Dempster dans les années 60 puis reformulée par Shafer dans les années 1970 [39]. Soit  $\Omega = \{\omega_1, \dots, \omega_K\}$  le cadre de discernement représentant l'ensemble des hypothèses en faveur desquelles une décision peut-être prise. Nous supposons que celles-ci sont exhaustives et exclusives. Dans la modélisation probabiliste, les probabilités sont définies sur cet ensemble en fonction de la connaissance que l'on possède de la source d'information. On modélise de cette façon l'incertitude sur les événements, l'imprécision étant considérée comme nulle. Ceci peut-être un inconvénient en fusion de données, plus particulièrement en présence de données imparfaites. Dans le cadre de la théorie de l'évidence, les

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

Une fonction de masse élémentaire  $m$  est une fonction de  $P(\Omega)$  dans  $\mathbb{R}^+$  vérifiant  $m(\emptyset) = 0$  et  $\sum_{A \in P(\Omega)} m(A) = 1$ . Celle-ci représente la partie du degré de croyance placée sur une proposition  $A \subset \Omega$ . On peut alors affecter une masse aux hypothèses singletons, mais également aux hypothèses composites, ce qui a pour avantage de ne pas introduire d'information arbitraire pour forcer une répartition de masse sur des hypothèses qu'un capteur serait incapable de différencier. On obtient alors une modélisation plus flexible permettant de représenter de multiples situations comme l'absence d'information, l'ambiguïté ou encore l'ignorance [16].

Une fonction de masse élémentaire  $m$  permet de définir les fonctions de crédibilité  $Bel(A) = \sum_{B \subset A} m(B)$  et de plausibilité  $Pls(A) = \sum_{A \cap B \neq \emptyset} m(B)$ . La crédibilité représente la croyance totale placée sur une proposition  $A$ , la plausibilité représentant la masse totale de croyance que l'on pourrait potentiellement transférer sur cette proposition. Ces deux fonctions représentent respectivement la valeur minimum et maximum de l'incertitude sur une proposition  $A$ . La longueur de l'intervalle  $[Bel(A), Pls(A)]$ , appelé intervalle de croyance, donne alors directement une mesure de l'imprécision sur l'incertitude sur une proposition  $A$ . Notons que la probabilité apparaît bien comme un cas particulier; en effet, lorsqu'une masse est nulle partout sauf, éventuellement, sur les singletons, la crédibilité est alors égale à la plausibilité et c'est une probabilité. Une telle masse sera appelée dans la suite « bayésienne ».

On peut imaginer qu'une masse modélise l'information fournie par une « source », qui peut être, à titre d'exemple, un capteur en imagerie. Lorsque l'on dispose de plusieurs masses  $m_1, \dots, m_l$  fournies par plusieurs sources, il se pose alors le problème de fusion des informations qu'elles modélisent. Une des possibilités que nous retenons pour la suite est la « combinaison de Dempster », dont la règle est la suivante :

$$m(A) = (m_1 \oplus m_2 \oplus \dots \oplus m_l)(A) \quad (1)$$

$$\propto \sum_{B_1 \cap \dots \cap B_l = A} m_1(B_1) m_2(B_2) \dots m_l(B_l)$$

Une des propriétés importante de la combinaison de Dempster est que lorsque au moins l'une des masses impliquées dans la combinaison est bayésienne, le résultat de la combinaison est une masse bayésienne. Cette propriété permet de généraliser le calcul de la loi *a posteriori* dans la sous-section suivante.

## 2.2. Généralisation de la loi *a posteriori*

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

combinaison de Dempster. En effet, on a

$$(m \oplus q)(x) = \frac{p(x)p(y|x)}{\sum_{x' \in \Omega} p(x')p(y|x')} = p(x|y) \quad (2)$$

On peut alors généraliser le calcul de la loi *a posteriori* en remplaçant la masse bayésienne *a priori*  $p(x)$  par une masse évidentielle  $m(u)$ ,  $u \in P(\Omega)$ . En effet, l'une des masses étant bayésienne (masse relative aux observations), le résultat de la combinaison de Dempster est une probabilité. Plus précisément, la combinaison de Dempster des masses  $m(u)$  et  $q(x)$  s'écrit (la somme  $\sum_{x \in u}$  est prise sur les  $u$  tels que  $x \in u$ ) :

$$(m \oplus q)(x) = \frac{\sum_{x \in u} m(u)p(y|x)}{\sum_{x' \in \Omega} \sum_{x' \in u'} m(u')p(y|x')} \quad (3)$$

## 2.3. Affaiblissement

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

$$m(u = \{\omega_i\}) = \alpha p(x = \{\omega_i\}), \quad m(\Omega) = 1 - \alpha \quad (4)$$

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

*Exemple 2.1.* Considérons une image satellite contenant deux classes « eau » et « forêt ». Nous avons  $\Omega = \{E, F\}$  et la connaissance *a priori* sur la région où a été prise l'image peut être modélisée par la masse bayésienne *a priori*  $p(E)$ ,  $p(F)$ . La masse modélisant l'information issue de l'observation  $y \in \mathbb{R}$  est la masse bayésienne définie par  $q(E) = f_E(y)/(f_E(y) + f_F(y))$  et  $q(F) = f_F(y)/(f_E(y) + f_F(y))$  où  $f_E(y) = p(y|x = E)$  et  $f_F(y) = p(y|x = F)$ . En combinant ces deux masses par la règle de combinaison de Dempster, on obtient la masse bayésienne  $m' = m \oplus q$ , qui correspond dans ce cas à la probabilité *a posteriori* :

$$m'(E) = \frac{p(E)f_E(y)}{\sum_{x' \in \{E, F\}} p(x')f_{x'}(y)} \quad (5)$$

$$m'(F) = \frac{p(F)f_F(y)}{\sum_{x' \in \{E, F\}} p(x')f_{x'}(y)}$$

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

On souhaite à présent modéliser la fiabilité de la source d'information *a priori*. Pour cela, nous introduisons le coefficient  $\alpha$  et nous remplaçons la masse bayésienne *a priori* par le nouveau jeu de masses  $m(E) = \alpha p(E)$ ,  $m(F) = \alpha p(F)$ ,  $m(\Omega) = 1 - \alpha$ . Après calcul, la masse bayésienne  $m' = m \oplus q$  généralisant la loi *a posteriori* s'écrit :

$$m'(E) = \frac{[\alpha p(E) + (1 - \alpha)] f_E(y)}{\sum_{x' \in \{E, F\}} [\alpha p(x') + (1 - \alpha)] f_{x'}(y)} \quad (6)$$

$$m'(F) = \frac{[\alpha p(F) + (1 - \alpha)] f_F(y)}{\sum_{x' \in \{E, F\}} [\alpha p(x') + (1 - \alpha)] f_{x'}(y)}$$

Cette masse peut alors être utilisée pour effectuer une classification bayésienne : en adoptant la fonction de coût « 0-1 », on obtient la règle de décision bayésienne  $\hat{s}_B$  définie par  $\hat{s}_B(y) = E$  si  $m'(E) \geq m'(F)$  et  $\hat{s}_B(y) = F$  si  $m'(E) \leq m'(F)$ .

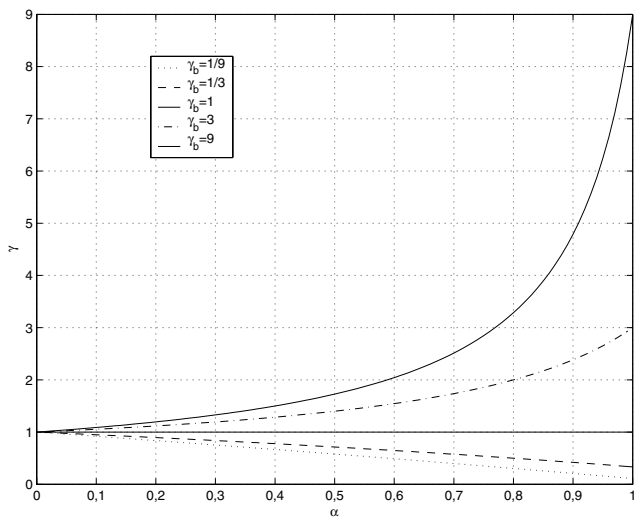
$$\hat{s}_B(y) = \begin{cases} F & \text{si } f_F(y)/f_E(y) \geq \gamma \\ E & \text{si } f_E(y)/f_F(y) \leq \gamma \end{cases} \quad (7)$$

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

$$\gamma = \frac{\alpha p(E) + (1 - \alpha)}{\alpha p(F) + (1 - \alpha)} \quad (8)$$

Sur le graphique (a) de la Figure 1, nous représentons la valeur du seuil de décision  $\gamma$  en fonction du coefficient  $\alpha$  pour différentes valeurs du seuil de décision avant affaiblissement  $\gamma_b = p(E)/p(F)$ .

- Lorsque  $\alpha = 1$ , la source est fiable et la valeur du seuil de décision n'est pas modifiée.
- Lorsque  $\alpha$  diminue, la source devient de moins en moins fiable et la valeur du seuil de décision tend vers 1.



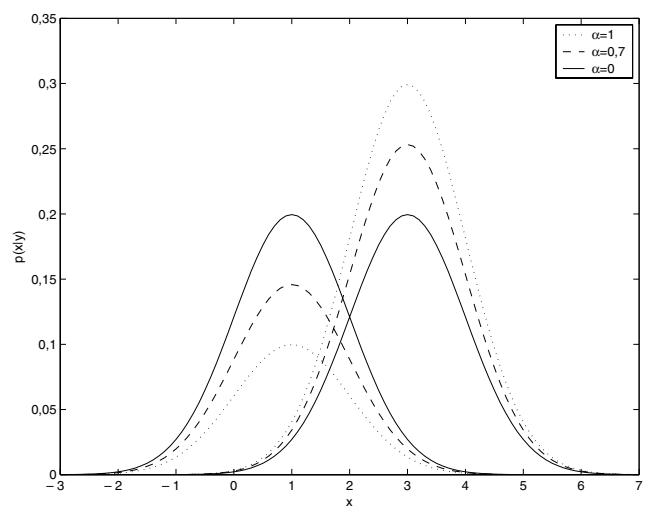
(a)

L'importance de la source d'information *a priori* dans la combinaison diminue.

- Lorsque  $\alpha = 0$ , la source n'est plus fiable et la valeur du seuil de décision est égale à 1. La source d'information *a priori* n'a donc plus d'influence sur le résultat de la classification.

Prenons le cas pour lequel  $p(E) = 1/4$ ,  $p(F) = 3/4$ ,  $f_E(y)$  étant une gaussienne de moyenne 1 et de variance 1 et  $f_F(y)$  étant une gaussienne de moyenne 3 et de variance 1. La valeur du seuil avant affaiblissement est  $\gamma_b = 1/3$ . Pour une valeur  $\alpha = 0,7$ , on décidera  $E$  si le rapport des vraisemblances est inférieur à la courbe correspondante à  $\gamma_b = 1/3$ , sinon on décidera  $F$ . Sur le graphique (b) de la Figure 1, nous avons représenté l'allure des lois *a posteriori*  $p(x|y) = m'(E)$  pour le cas considéré ainsi que pour les cas  $\alpha = 1$  et  $\alpha = 0$ . Pour  $\alpha = 0$ , l'information *a priori* n'a plus d'influence dans la combinaison des sources d'informations.

L'opération d'affaiblissement a donc pour effet de pondérer l'importance de la source d'information *a priori*, lorsque cette dernière n'est pas entièrement fiable. L'utilisation de ce fait dans un cadre markovien est le point crucial de l'article ; plus précisément, nous montrons par la suite deux points. D'abord, bien que la combinaison de Dempster détruit la markovianité, la segmentation bayésienne est possible en utilisant les chaînes de Markov triplet. Ensuite, nous montrons par des simulations que dans le cas des chaînes de Markov cachées non stationnaires l'introduction d'un affaiblissement particulier permet d'améliorer, de manière automatique, la segmentation non supervisée (fondée sur l'estimation des paramètres par EM) classique. En d'autres termes, il est montré que plonger le modèle non stationnaire dans un modèle triplet stationnaire permet d'améliorer les résultats de segmentation non supervisée.



(b)

Figure 1. (a) Valeur du seuil de décision  $\gamma$  en fonction du coefficient d'affaiblissement  $\alpha$  pour différentes valeurs du seuil de décision  $\gamma_b$  initial. (b) Représentations des lois *a posteriori* pour  $\gamma_b = 1/3$  et différentes valeurs du coefficient d'affaiblissement  $\alpha$  avec  $f_E(y) = \mathcal{N}(1, 1)$  et  $f_F(y) = \mathcal{N}(3, 1)$ .

## 3. Introduction de l'information spatiale

L'information spatiale peut s'avérer très utile, notamment pour des problèmes de segmentation d'images ou de signaux. Les modèles de Markov sont bien adaptés au problème en raison de l'existence d'algorithmes d'échantillonnage et d'estimation itérative. En particulier, les chaînes de Markov cachées offrent des méthodes d'estimation et de segmentation rapides basées sur l'utilisation des récurrences forward-backward, possible grâce à la markovianité de la loi *a posteriori*. La méthode que nous allons présenter consiste à remplacer la loi *a priori* par une masse évidentielle markovienne, ce qui permettra, d'après ce que nous avons vu précédemment, de gérer l'insuffisance de la connaissance de la loi *a priori* non stationnaire en introduisant un affaiblissement. Nous rappelons dans un premier temps le modèle de chaîne de Markov cachée avec bruit indépendant ainsi que ses généralisations successives aux chaînes de Markov couple et triplet. Puis nous introduisons le concept de chaîne de Markov évidentielle et généraliserons le modèle de chaîne de Markov cachée au modèle de chaîne de Markov cachée évidentielle. Bien que la combinaison de Dempster détruise la markovianité de la loi *a posteriori*, les méthodes de segmentation et d'estimation restent possibles, la combinaison étant un cas particulier de chaîne de Markov triplet. L'intérêt de l'introduction de l'affaiblissement via une masse évidentielle markovienne est illustrée par un exemple original.

### 3.1. Chaîne de Markov cachée

Une modélisation classique de la loi du processus  $Z = (X, Y)$ , qui sera dite « chaîne de Markov cachée avec bruit indépendant » (CMC-BI), est la suivante

$$p(x, y) = p(x_1)p(x_2|x_1) \dots p(x_N|x_{N-1}) \times p(y_1|x_1)p(y_2|x_2) \dots p(y_N|x_N) \quad (9)$$

Il existe trois généralisations successives du modèle classique CMC-BI. Dans la première, qui aboutit au modèle dit « chaîne de Markov cachée » (CMC), les deux processus  $X$  et  $Z$  sont de Markov (le bruit n'est plus nécessairement « indépendant » dans le sens que les variables aléatoires  $Y_1, \dots, Y_N$  ne sont plus nécessairement indépendantes conditionnellement à  $X$ ). Dans la deuxième, on aboutit au modèle dit « chaîne de Markov couple » (CMCouple), dans lequel on suppose  $Z$  de Markov. Notons que CMCouple est strictement plus général que CMC; en d'autres termes, dans un CMCouple le processus  $X$  n'est pas nécessairement de Markov. Enfin, le processus  $Z$  sera dit « chaîne de Markov triplet » (CMT) s'il existe un processus  $U = (U_1, \dots, U_N)$  tel que  $T = (X, U, Y)$  soit une chaîne de Markov. Le modèle CMT est alors strictement plus général que le modèle CMCouple car dans une CMT le processus  $Z$  n'est

pas nécessairement de Markov. Le modèle CMC-BI est le plus ancien et bien connu, l'utilisation des CMC qui ne soient pas des CMC-BI étant plus rare. L'utilisation des CMCouple qui ne soient pas des CMC est récente et leur potentiel, qui semble prometteur, commence seulement à être exploré [10]. Par ailleurs, notons l'existence des champs de Markov Couple et Triplet [33], [34]. Enfin, les processus de Markov Couple et Triplet peuvent être également considérés dans le cas des processus cachés à états continus [11], [12], [23].

### 3.2. Chaîne de Markov évidentielle Cachée

On dira qu'une fonction de masse  $m$  définie sur  $P(\Omega^N)$  est une chaîne de Markov évidentielle (CME) si elle est nulle en dehors de  $[P(\Omega)]^N$  et si  $m(u) = m(u_1, \dots, u_N) = m(u_1)m(u_2|u_1) \dots m(u_N|u_{N-1})$ ; l'écriture  $m(u_{n+1}|u_n)$  signifiant que l'on associe à chaque  $u_n \subset \Omega$  une masse  $m(\cdot|u_n)$ . Ainsi une chaîne de Markov (CM) classique est un cas particulier d'une CME, dans laquelle  $m(u_1)$  et toutes les  $m(u_{n+1}|u_n)$  sont nulles en dehors des singletons. De manière analogue à la démarche de la sous-section 2.2, notons qu'en remplaçant dans (9)  $p(x) = p(x_1)p(x_2|x_1) \dots p(x_N|x_{N-1})$  par  $m(u) = m(u_1)m(u_2|u_1) \dots m(u_N|u_{N-1})$  la combinaison de Dempster de  $m$  avec la loi de probabilité  $q(x) = q(x_1) \dots q(x_N) \propto p(y_1|x_1) \dots p(y_N|x_N)$  ( $y$  est fixé et  $q(x_n) = p(y_n|x_n) / \sum_{x'_n \in \Omega} p(y_n|x'_n)$ ) est une probabilité qui généralise la loi *a posteriori* (cette dernière étant obtenue lorsque  $m(u) = p(x) = p(x_1)p(x_2|x_1) \dots p(x_N|x_{N-1})$ ). Cette fusion s'écrit

$$(m \oplus q)(x) = \frac{\sum_{u \in [P(\Omega)]^N} \mathbb{1}_{[x \in u]} m(u) q(x)}{\sum_{x' \in \Omega^N} \sum_{u' \in [P(\Omega)]^N} \mathbb{1}_{[x' \in u']} m(u') q(x')} \quad (10)$$

Afin de montrer que  $m \oplus q$  est une marginale d'une chaîne de Markov, énonçons le résultat classique suivant :

**Lemme 3.1.** Soit  $V = (V_1, \dots, V_N)$  un processus stochastique, chaque  $V_n$  prenant ses valeurs dans un même ensemble fini. Alors  $V$  est une chaîne de Markov si et seulement si il existe  $N - 1$  fonctions positives sommables  $r_1, \dots, r_{N-1}$ , telles que la loi de  $V$  est proportionnelle à  $r_1(v_1, v_2) \dots r_{N-1}(v_{N-1}, v_N)$ . De plus, la loi de  $V$  est alors donnée par  $p(v_1) = \beta_1(v_1) / \sum_{v'} \beta_1(v')$  et les transitions  $p(v_n|v_{n-1}) = r_{n-1}(v_{n-1}, v_n) \beta_n(v_n) / \beta_{n-1}(v_{n-1})$ , où les fonctions  $\beta_1, \dots, \beta_N$ , sont définies par  $\beta_N(v_N) = 1$ , et  $\beta_{n-1}(v_{n-1}) = \sum_{v_n} r_{n-1}(v_{n-1}, v_n) \beta_n(v_n)$ .

En considérant  $V = (X, U)$ , avec  $V_n = (X_n, U_n)$ , posons  $r_1(v_1, v_2) = \mathbb{1}_{[x_1 \in u_1]} \mathbb{1}_{[x_2 \in u_2]} m(u_1, u_2) q(x_1) q(x_2)$ , et  $r_{n-1}(v_{n-1}, v_n) = \mathbb{1}_{[x_n \in u_n]} m(u_n|u_{n-1}) q(x_n)$  pour  $3 \leq n \leq N$ . La quantité  $\mathbb{1}_{[x \in u]} m(u) q(x)$  est de la forme  $r_1(v_1, v_2) \dots r_{N-1}(v_{N-1}, v_N)$  et donc  $p(v) = p(x, u)$

$= \mathbb{1}_{[x \in u]} m(u) q(x) / \sum_{x' \in \Omega^N} \sum_{u' \in [P(\Omega)]^N} \mathbb{1}_{[x' \in u']} m(u') q(x')$  est, en vertu du Lemme 3.1, une chaîne de Markov avec les transitions calculables. Par ailleurs,  $(m \oplus q)(x) = \sum_{u \in [P(\Omega)]^N} p(x, u)$

d'après (10), ce qui montre que  $m \oplus q$  est une marginale de cette même chaîne de Markov. Sachant que  $(m \oplus q)(x)$  généralise la loi *a posteriori* de  $X$  classique ( $y$  figure dans  $(m \oplus q)(x)$ ), on a  $p(x_n | y) = (m \oplus q)(x_n) = \sum_{u_n \in P(\Omega)} p(x_n, u_n)$  qui est ainsi calculable, ce qui permet d'appliquer la méthode de segmentation Bayésienne MPM donnée par  $\hat{s}_{MPM}(y_1, \dots, y_N) = (\hat{x}_1, \dots, \hat{x}_N)$ , avec  $\hat{x}_n = \arg \max_{x_n \in \Omega} [p(x_n | y)]$ .

Concrètement, les quantités  $p(x_n | y)$  (utilisées dans la segmentation par MPM) et  $p(v_n, v_{n+1} | y)$  (utilisées dans l'estimation des paramètres précisée ci-après) se calculent à partir des  $r_{n-1}(v_{n-1}, v_n) = \mathbb{1}_{[x_n \in u_n]} m(u_n | u_{n-1}) q(x_n)$  (dans lesquels  $y$  figure implicitement, sachant que  $q(x_n) \propto p(y_n | x_n)$ ) de la manière suivante. On calcule la suite  $\beta_1, \dots, \beta_N$ , et les transitions  $p(v_n | v_{n-1})$  par les formules du Lemme 3.1. Ensuite, les marginales  $p(v_n)$  sont calculées récursivement par  $p(v_1) = \beta_1(v_1)$ ,  $p(v_n) = \sum_{v_{n-1}} p(v_n | v_{n-1}) p(v_{n-1})$ , ce qui donne  $p(x_n | y) = \sum_{u_n \in P(\Omega)} p(x_n, u_n)$  d'une part, et  $p(v_n, v_{n+1} | y) = p(v_n | v_{n-1}) p(v_{n-1})$ , qui seront utilisées dans la suite pour l'estimation des paramètres, d'autre part.

Par ailleurs, considérer  $V = (X, U)$  comme étant le processus caché dans  $T = (X, U, Y)$  (qui devient donc CMC-BI classique) permet d'appliquer l'une des méthodes classiques d'estimation des paramètres, comme EM, SEM, ou ICE. L'efficacité de ces trois méthodes étant comparable dans le cas Gaussien avec un léger avantage de la rapidité revenant à la méthode EM [3], nous avons choisi cette dernière dans les applications présentées ci-après. Remarquons que l'estimation des paramètres de la CMC-BI  $T = (X, U, Y)$  résout l'important problème de « l'estimation de l'affaiblissement ».

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

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

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

$$m_{ij}^{q+1} = \frac{1}{A} \sum_{n=1}^{N-1} \sum_{(x_n, x_{n+1}) \in \Omega^2} \xi^q(x_n, u_n = A_i, x_{n+1}, u_{n+1} = A_j) \quad (11)$$

$$\mu_k^{q+1} = \frac{1}{N} \sum_{n=1}^N \sum_{u_n \in P(\Omega)} \gamma^q(x_n = k, u_n) y_n \quad (12)$$

$$(\sigma_k^{q+1})^2 = \frac{1}{N} \sum_{n=1}^N \sum_{u_n \in P(\Omega)} \gamma^q(x_n = k, u_n) \times (y_n - \mu_k^{q+1})^2 \quad (13)$$

avec  $A = (N - 1) \#(A_i) \#(A_j)$  et  $\#(A_i)$  le cardinal de  $A_i$ . Notons que la méthode utilisée donne l'algorithme EM classique lorsque la CME  $U$  ne charge que les singletons (on a alors  $U = X$  et le modèle est une CMC-BI classique).

Revenons au problème de la segmentation non supervisée des chaînes de Markov cachées fortement non stationnaires. Nous avons deux manières de traiter le problème et l'objectif des résultats présentés à la Figure 2 est de montrer l'intérêt de la seconde.

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

La réalisation  $X = x$  du processus caché à deux classes, représentée sur le graphique (a) de la Figure 2, est fortement non stationnaire. Le bruitage classique consiste à considérer deux densités gaussiennes de variances égales à 1, et de moyennes 1 et 3 respectivement, ce qui donne une réalisation  $Y = y$  représentée sur le graphique (b). La réalisation du processus caché estimée par la première méthode est présentée sur le graphique (c); on constate que cette méthode est incapable d'estimer correctement l'endroit où les probabilités de transition d'un état à un état différent sont très fortes (les  $x_n$  pour  $n$  variant de 70 à 130). Par contre, la deuxième méthode, dont la trajectoire estimée est présentée sur le graphique (e), présente une estimation convenable de cette partie du processus caché. Afin de mieux cerner ce phénomène nous avons présenté sur le graphique (d) les coefficients d'affaiblissement appliqués, pour chaque point  $n$ , à la loi marginale *a posteriori* correspondante. Il s'avère que ces marginales ne sont pratiquement pas modifiées aux points où la première méthode donne des bons résultats (les points en dehors des  $n$  variant de 70 à 130); par contre, l'affaiblissement est maximal pour une bonne partie des points où la première méthode faillit (les  $n$  variant de 70 à 130). Ce phénomène, dont l'explication mathématiquement rigoureuse ne nous apparaît pas avec évi-

dence, est intéressant pour les applications pratiques. En particulier, il réapparaît dans les segmentations non supervisées d'images présentées dans la Section 5.

*Remarque 3.1.* Le modèle de la CME cachée, où la CME  $m(u) = m(u_1)m(u_2|u_1) \dots m(u_N|u_{N-1})$  est fusionnée avec la loi de probabilité  $q(x) = q(x_1) \dots q(x_N) \propto p(y_1|x_1) \dots p(y_N|x_N)$  ( $y$  est fixé et  $q(x_n) \propto p(y_n|x_n)$  peut être généralisé au cas de plusieurs capteurs indépendants  $c_1, \dots, c_r$  de la façon suivante. Si tous les capteurs produisent des lois de probabilité, la généralisation est immédiate en remplaçant  $q(x_n) \propto p(y_n|x_n)$  par  $q(x_n) \propto p(y_n^1|x_n) \dots p(y_n^r|x_n)$ , qui est une combinaison de Dempster. Si certains des capteurs  $c_1, \dots, c_s$ , (par exemple les  $s$  premiers) produisent des masses (de telles situations où, par exemple, un capteur optique est confronté à la présence des nuages, sont étudiées dans [2], [16]), ces dernières sont fusionnées avec les  $r - s$  probabilités données par  $p(y_n^{s+1}|x_n), \dots, p(y_n^r|x_n)$  avec, pour résultat, une probabilité  $q(x_n)$ . Cette dernière est alors fusionnée avec  $m(u) = m(u_1)m(u_2|u_1) \dots m(u_N|u_{N-1})$  et la suite de la démarche est identique à celle correspondant au cas d'un seul capteur probabiliste, traité dans cet article.

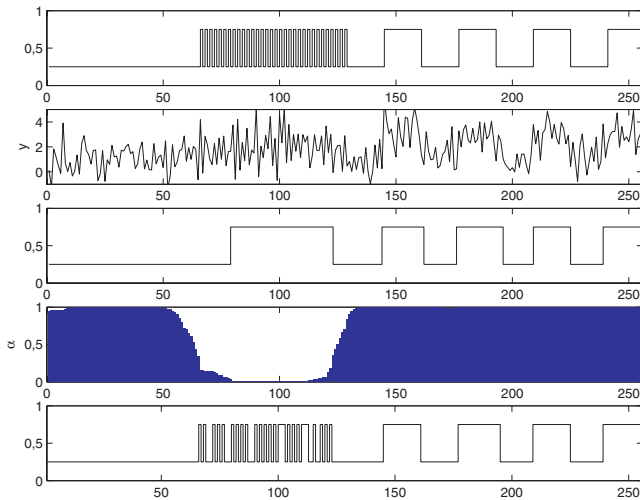


Figure 2. De haut en bas : (a) Processus caché, (b) Processus observé, (c) Résultat de la segmentation non-supervisée de la CMC, (d) Variation du coefficient d'affaiblissement  $\alpha$  conditionnellement aux observations, (e) Résultat de la segmentation non-supervisée de la CMEC.

## 4. Fusion DS dans CMCCouple et AMCCouple

Nous avons vu dans ce qui précède que lorsque dans une CMC-BI  $(X, Y)$  on remplace la loi *a priori* de  $X$  par une « masse évidentielle Markovienne », la fusion de cette dernière avec une

probabilité définie par les observations  $Y = y$  est formellement la loi de  $X$  conditionnelle à  $Y = y$  déduite d'une CMT  $T = (X, U, Y)$ , ce qui permet le calcul des marginales  $p(x_n|y)$ . Des résultats analogues peuvent être montrés lorsque l'on considère les arbres de Markov cachés avec du bruit indépendant (AMC-BI). Sachant que les CMC-BI peuvent être étendues aux CMC, CMCCouple, et CMT (et il en est de même des AMC-BI), la question qui se pose est celle de savoir s'il est encore possible, dans ces extensions, de remplacer la loi *a priori* de  $X$  par une « masse évidentielle Markovienne » de manière à ce que le résultat de la combinaison de Dempster avec une probabilité définie par les observations  $Y = y$  étende, d'une part, la loi *a posteriori* de  $X$  et aboutisse, d'autre part, à un résultat exploitable comme dans le cas des CMC-BI. L'objet de cette section est de montrer que la réponse est positive.

### 4.1. Fusion dans CMCCouple

Considérons une CMCCouple  $Z = (X, Y)$ , avec  $p(z) = p(z_1)p(z_2|z_1) \dots p(z_N|z_{N-1}) = \frac{p(z_1, z_2) \dots p(z_{N-1}, z_N)}{p(z_2) \dots p(z_{N-1})}$ . Sachant que  $p(z_n, z_{n+1}) = p(x_n, x_{n+1})p(y_n, y_{n+1}|x_n, x_{n+1})$  et  $p(z_n) = p(x_n)p(y_n|x_n)$ , la loi de  $Z$  s'écrit également

$$p(z) = \underbrace{\frac{p(x_1, x_2) \dots p(x_{N-1}, x_N)}{p(x_1) \dots p(x_{N-1})}}_{a(x)} \times \underbrace{\frac{p(y_1, y_2|x_1, x_2) \dots p(y_{N-1}, y_N|x_{N-1}, x_N)}{p(y_2|x_2) \dots p(y_{N-1}|x_{N-1})}}_{b(x, y)} \quad (14)$$

L'écriture (14) est ainsi valable dans toute CMCCouple, avec trois cas particuliers :

- 1) les  $p(y_n, y_{n+1}|x_n, x_{n+1})$  dans  $b(x, y)$  sont de la forme  $p(y_n, y_{n+1}|x_n, x_{n+1}) = p(y_n|x_n)p(y_{n+1}|x_{n+1})$  ; on retrouve le cas de CMC-BI traité dans la section précédente ;
- 2) l'égalité  $p(y_n, y_{n+1}|x_n, x_{n+1}) = p(y_n|x_n)p(y_{n+1}|x_{n+1})$  n'a pas nécessairement lieu mais l'intégrale de  $b(x, y)$  par rapport à  $y$  vaut 1 : alors le modèle est une CMC et  $a(x)$  est la loi de  $X$  ;
- 3) l'intégrale de  $b(x, y)$  par rapport à  $y$  ne vaut pas 1 (ce qui revient à dire que  $a(x)$  n'est pas la loi de  $X$ , ou encore que  $X$  n'est pas de Markov et donc  $(X, Y)$  est une CMCCouple sans être une CMC). Notons que dans ce dernier cas  $p(x_n, x_{n+1})$  figurant dans (14) est bien la loi de  $(X_n, X_{n+1})$  pour tout  $1 \leq n \leq N - 1$ , mais  $a(x)$  n'est pas la loi de  $X$ .

Considérons le cas le plus général où la loi de  $Z = (X, Y)$  est donnée par (14). Ainsi  $a$  n'est plus nécessairement la loi de  $X$  et, de manière équivalente,  $b(x, y)$  n'est plus nécessairement la loi de  $Y$  conditionnelle à  $X = x$ . Cependant, en notant  $q$  la loi de probabilité obtenue à partir de  $b(x, y)$  par normalisation par rapport à  $x$  et à  $y$  fixés, on peut affirmer que la loi *a posteriori*  $p(x|y)$  est le résultat de combinaison de Dempster de  $a$  par  $q$ . En suivant la même démarche que celle appliquée au cas (1)

dans la section précédente, on peut donc envisager de remplacer  $a$  par une chaîne évidentielle.

**Proposition 4.1.** Soit  $Z = (Z_1, \dots, Z_N)$  une CM Couple dont la loi est définie par (14). Soient  $q$  la loi de probabilité définie par  $b(x, y)$  (après normalisation par rapport à  $x$  et à  $y$  fixé) et  $m$  une CME définie par  $m(u) = \frac{m(u_1, u_2) \dots m(u_{N-1}, u_N)}{m(u_2) \dots m(u_{N-1})}$  ( $m$  étend la probabilité  $a$  de (14)).

Alors la combinaison de Dempster  $m \oplus q$ , qui généralise la loi a posteriori  $p(x|y)$  du cas classique où  $m = a$ , est la loi marginale de la chaîne de Markov  $p(x, u|y)$  définie par les fonctions

$$r_{n-1}(v_{n-1}, v_n) = r_{n-1}(x_{n-1}, u_{n-1}, x_n, u_n) \\ = \mathbb{1}_{[x_{n-1} \in u_{n-1}]} \mathbb{1}_{[x_n \in u_n]} m(u_{n-1}, u_n) q_n(x_{n-1}, x_n)$$

$$\text{avec } q_2(x_1, x_2) \propto p(y_1, y_2 | x_1, x_2)$$

$$\text{et } q_n(x_{n-1}, x_n) \propto p(y_{n-1}, y_n | x_{n-1}, x_n) / p(y_{n-1} | x_{n-1})$$

pour  $3 \leq n \leq N$  ( $y$  est fixé).

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

La preuve découle de la définition de la combinaison de Dempster et du Lemme 3.1

Remarque 4.1

- 1) Une étude analogue à celle faite dans la section précédente pourrait ainsi être effectuée dans le cas des CM Couple non stationnaires dans le sens où  $p(x_n, x_{n+1})$  varieraient avec  $n$ . On pourrait alors introduire l'affaiblissement et utiliser EM pour en estimer automatiquement l'importance. Mentionnons également des méthodes d'estimation plus récentes, fondées sur « Iterative Conditional Estimation » (ICE), qui permettent de sortir du cas Gaussien et d'estimer automatiquement la forme du bruit, cette dernière pouvant varier avec la classe [10].
- 2) L'extension de la combinaison de Dempster aux chaînes de Markov triplet (CMT), modèle strictement plus général que les CM Couple, est la suivante. Considérons une CMT  $T = (X, W, Y)$ , avec  $X$  processus caché,  $Y$  processus observé, et chaque  $W_n$  prenant ses valeurs dans un ensemble fini  $\Delta$ . En posant  $V = (X, W)$ , le processus  $T = (V, Y)$  est une CM Couple, dont la loi peut s'écrire sous la forme (14), avec  $V$  à la place de  $X$ . En appliquant ce qui précède à  $T = (V, Y)$ , on peut affirmer que lorsque l'on remplace  $a(v)$  par une masse évidentielle  $m(v)$ , la combinaison de Dempster  $m \oplus q$  est la loi a posteriori  $p(v|y)$  définie par la CMT  $T^* = (V, U, Y)$ , où chaque  $U_n$  prend ses valeurs dans  $\Lambda = P(\Omega \times \Delta)$ . Finalement, nous avons deux chaînes triplet: la première  $T = (X, W, Y)$  est purement probabiliste, et la deuxième  $T^* = (V, U, Y)$  (avec  $V = (X, W)$ ) est obtenue à partir de la CM Couple  $T = (V, Y)$  lorsque l'on remplace la loi  $a(v)$  par une masse évidentielle  $m(v)$ . En appliquant les résultats précédents il est ainsi possible de considérer l'introduction de l'affaiblissement dans une CMT non stationnaire et l'utilisation d'un tel modèle à des fins de segmentation non supervisée.

## 4.2. Fusion dans Arbre de Markov Couple

Soit  $S$  un ensemble fini d'indices,  $X = (X_s)_{s \in S}$ ,  $Y = (Y_s)_{s \in S}$  deux processus, avec chaque  $Z_s = (X_s, Y_s)$  à valeurs dans  $\Omega \times \mathbb{R}$ . Soit  $S^1, \dots, S^N$  une partition de  $S$ . Posons  $M = \text{Card}(S)$  et, pour  $1 \leq n \leq N$ ,  $M^n = \text{Card}(S^n)$ . Supposons que  $M^1 \leq M^2 \leq \dots \leq M^N$  (avec  $M^1 = 1$ ) et pour tout  $1 \leq n \leq N-1$ , associons à chaque  $s \in S^n$  un sous-ensemble  $s^+$  de  $S^{n+1}$ , appelé « enfants de  $s$  », de manière à ce que  $(s^+)_{s \in S^n}$  soit une partition de  $S^{n+1}$  (on suppose  $S^1 = \{s_1\}$ ). Pour tout  $s \in S - S^1$ , son unique « père » sera noté  $s^-$ . Sous toutes ces conditions on dira que  $S$  a une structure d'« arbre ».

Le processus  $Z$  sera alors appelé Arbre de Markov Couple (AM Couple) si sa loi s'écrit  $p(z) = p(z_{s_1}) \prod_{n=2}^N \prod_{s \in S^n} p(z_s | z_{s^-})$ .

Une telle structure permet divers traitements; en particulier,  $p(x_s | y)$ , et  $p(x_s | x_{s^-}, y)$  sont calculables, ce qui rend possible le calcul de la segmentation bayésienne MPM et la simulation de  $X$  selon  $p(x|y)$  [28]. Par ailleurs, les AM Couple généralisent les CM Couple, ces dernières étant obtenues lorsque tous les  $s^+$  sont des singletons.

De manière analogue au lemme 3.1, on peut énoncer

**Lemme 4.1.** Soit  $V = (V_s)_{s \in S}$  un processus stochastique, chaque  $V_s$  à valeurs dans un ensemble fini, avec  $S$  ensemble de noeux muni de la structure d'arbre précisée ci-dessus. Alors  $V$  est un arbre de Markov si et seulement si il existe des fonctions positives sommables  $r_{s^-, s}$  telles que

$$p(v) \propto r_{s_1}(v_{s_1}) \prod_{n=2}^N \prod_{s \in S^n} r_{s^-, s}(v_{s^-}, v_s). \text{ De plus, les transitions } \\ p(v_s | v_{s^-}) \text{ de l'arbre sont données par } p(v_s | v_{s^-}) = \\ r_{s^-, s}(v_{s^-}, v_s) \beta_s(v_s) / \beta_{s^-}(v_{s^-}), \text{ où les } \beta_s \text{ sont définies par } \\ \beta_s(v_s) = 1 \text{ pour tout } s \in S^N, \text{ et } \beta_{s^-}(v_{s^-}) = \\ \prod_{u \in (s^-)^+} \sum_{v_u} r_{s^-, u}(v_{s^-}, v_u) \beta_u(v_u).$$

De manière analogue à celle du paragraphe précédent, la loi d'un AM Couple peut s'écrire

$$p(z) = \underbrace{\left[ p(x_{s_1}) \prod_{n=2}^N \prod_{s \in S^n} p(x_s | x_{s^-}) \right]}_{a(x)} \times \underbrace{\left[ p(y_{s_1} | x_{s_1}) \prod_{n=2}^N \prod_{s \in S^n} \frac{p(y_{s^-}, y_s | x_{s^-}, x_s)}{p(y_{s^-} | x_{s^-})} \right]}_{b(x, y)} \quad (15)$$

Comme précédemment,  $a(x)$  n'est pas nécessairement la loi de  $X$  et  $b(x, y)$  n'est pas nécessairement la loi de  $Y$  conditionnelle à  $X = x$ ; cependant, on note que  $p(x|y)$  définie par (15) est le résultat de la combinaison de Dempster de la probabilité  $a(x)$  avec la loi de probabilité  $q$  obtenue à partir de  $b(x, y)$  par normalisation par rapport à  $x$  et à  $y$  fixés. De manière analogue à ce qui a été fait dans le cas des chaînes Couple ci-dessus, il est alors possible d'envisager d'étendre la probabilité  $a(x)$  à un



« arbre de Markov évidentiel » et de fusionner ce dernier avec la loi de probabilité  $q$ . Nous avons le résultat suivant :

**Proposition 4.2.** Soient  $Z = (Z_s)_{s \in S}$  un AMCouple dont la loi est définie par (15),  $q$  définie par  $b(x, y)$  dans (15) par normalisation, à  $y$  fixé, par rapport à  $x$ , et  $m$  un AME défini par  $m(u_{s_1}) \prod_{n=2}^N \prod_{s \in S^n} m(u_s | u_{s^-})$ .

Alors la combinaison de Dempster  $m \oplus q$ , qui généralise la loi **a posteriori**  $p(x|y)$  du cas  $m = a$ , est la loi marginale de l'arbre de Markov  $p(x, u|y)$  définie par les fonctions  $r_{s^-, s}(v_{s^-}, v_s) = r_{s^-, s}(x_{s^-}, u_{s^-}, x_s, u_s)$

$$= \mathbb{1}_{[x_{s^-} \in u_{s^-}]} \mathbb{1}_{[x_s \in u_s]} m(u_{s^-}, u_s) q_s(x_{s^-}, x_s),$$

(avec  $q_s(x_{s^-}, x_s) = p(y_{s^-}, y_s | x_{s^-}, x_s) / p(y_{s^-} | x_{s^-})$  pour  $s^- \neq s_1$  et  $q_s(x_{s_1}, x_s) = p(y_{s_1}, y_s | x_{s_1}, x_s)$ ).

Comme précédemment, la preuve découle de la définition de la fusion de Dempster et du Lemme 4.1.

Remarque 4.2.

- 1) Introduire des masses évidentielles dans les modèles CMCouple et AMCouple est ainsi moins « naturel » que dans les modèles CMC et AMC. En effet, on étend aux masses les probabilités  $a(x)$  qui sont des CM et AM respectivement, mais qui ne sont pas nécessairement les lois des processus cachés. Cependant, cette extension se fait de manière « continue » dans la mesure où l'on retrouve CMCouple et AMCouple lorsque les masses considérées ne chargent que des singletons.
- 2) De manière analogue à la démarche évoquée dans le deuxième point des remarques 4.1, il est possible d'étendre les AMCouple  $Z = (X, Y)$  aux arbres de Markov triplet (AMT)  $T = (X, W, Y)$ .

## 5. Segmentation non supervisée des images non-stationnaires

L'image  $128 \times 128$  de la figure 3 est bruitée par un bruit gaussien indépendant de variance 1 et de moyennes 1 et 3. L'ensemble bi-dimensionnel de pixels est converti, via un parcours de Hilbert-Peano, en un ensemble mono-dimensionnel  $1, 2, \dots, N = 128 \times 128$ , comme décrit dans [3], [10], [15]. Les résultats présentés sur la figure 5 montrent que la restauration MPM fondée sur les paramètres estimés par CMC-EM (b) ne permet pas de trouver les détails des ailes et de la queue de l'oiseau, dû à un trop fort taux de régularisation spatiale. Sur l'image (c), nous avons représenté le taux d'affaiblissement  $a$

*posteriori* par pixel. On constate que pour les zones comprenant des détails, celui-ci tend vers 0. Le résultat de la restauration MPM basée sur le modèle CMEC représenté sur l'image (d) montre bien l'amélioration obtenue. Les estimés des différents paramètres en considérant le modèle CMC puis le modèle CMEC sont présentés sur le tableau 1. Nous constatons que l'estimation des paramètres du bruit est plus efficace en considérant le modèle CMEC.

Tableau 1. Estimés obtenus par EM en considérant les cas CMC et CMEC.

CMC			$N(0, 1)$	$N(2, 1)$
$p(\omega_1)$	0.2789	$\mu$	-0.002	1.063
$p(\omega_2)$	0.7213	$\sigma$	1.0150	2.075
$p(x_1, x_2)$		$\omega_1$	$\omega_2$	
$\omega_1$		0.2730	0.0015	
$\omega_2$		0.0056	0.7156	
Paramètres de la CMC estimés par EM				
CMEC			$N(0, 1)$	$N(2, 1)$
$m(\{\omega_1\})$	0.0234	$\mu$	-0.0065	2.0815
$m(\{\omega_2\})$	0.7119	$\sigma$	1.0128	0.9252
$m(\{\omega_1, \omega_2\})$	0.2648			
$m(u_1, u_2)$		$\{\omega_1\}$	$\{\omega_2\}$	$\{\omega_1, \omega_2\}$
$\{\omega_1\}$		0.0205	0.0000	0.0028
$\{\omega_2\}$		0.0000	0.7074	0.0045
$\{\omega_1 \omega_2\}$		0.0028	0.0045	0.2573
Paramètres de la CMEC estimés par EM				

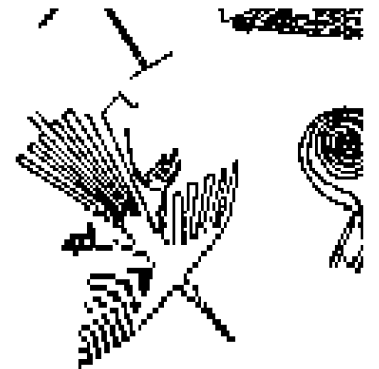


Figure 3. Image originale à deux classes.

### 5.1. Image réelle

Nous réeffectuons la même expérimentation à partir d'une image réelle. Les résultats sont présentés sur la figure 5.

Il est difficile, en absence de la vérité terrain, de comparer quantitativement les résultats obtenus. Cependant, certains détails semblent mieux conservés par la nouvelle méthode. Par ailleurs, la carte des valeurs du coefficient d'affaiblissement *a posteriori* semble compatible avec les zones difficiles à segmenter.



Figure 4. Segmentations non supervisées utilisant les chaînes classiques et les chaînes évidentielles. (a) Image bruitée par deux gaussiennes  $\mathcal{N}(0,1)$  et  $\mathcal{N}(2,1)$ . (b) Restauration non supervisée CMC-EM-MPM, taux d'erreur: 14,62 %. (c) Valeur du coefficient d'affaiblissement conditionnellement aux observations. (d) Restauration non supervisée CMEC-EM-MPM, taux d'erreur: 5,05 %.

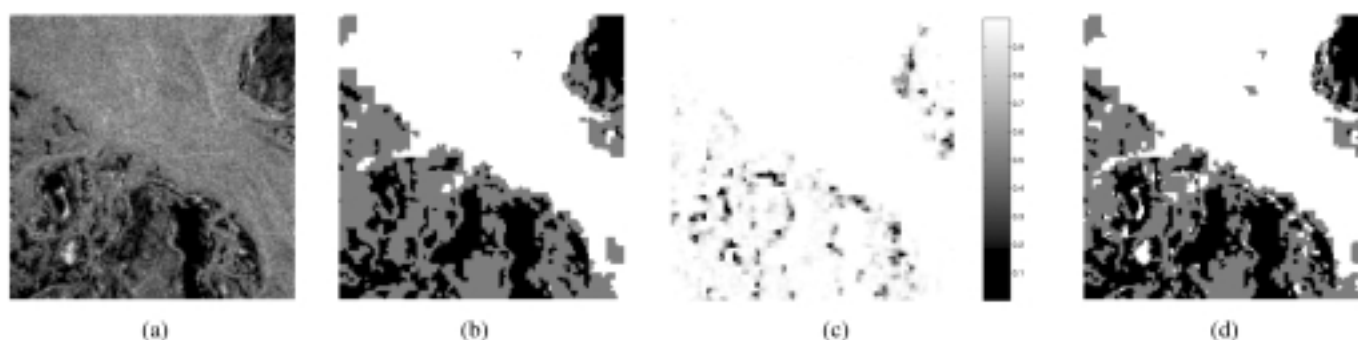


Figure 5. Image réelle et résultats de la segmentation à trois classes en considérant le modèle CMC puis le modèle CMEC. (a) Image réelle. (b) Segmentation non supervisée CMC-EM-MPM. (c) Valeur du coefficient d'affaiblissement conditionnellement aux observations. (d) Segmentation non supervisée CMEC-EM-MPM.

## 6. Conclusions et perspectives

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

Par ailleurs, nous avons précisé comment étendre la démarche aux arbres de Markov couple, modèles strictement plus généraux que les arbres de Markov cachés.

Mentionnons quelques perspectives de poursuite des recherches

- 1) la prise en compte simultanée de plusieurs capteurs, discriminant éventuellement des classes différentes, d'une part, et la gestion des non stationnarités comme décrit dans cet article, d'autre part, est possible et son étude dans le contexte des chaînes ou arbres de Markov couple constitue un prolongement naturel des résultats présentés;
- 2) l'extension aux réseaux bayésiens [18], [25], [42] constitue une autre perspective naturelle pour la poursuite du travail présenté;
- 3) les approches floues de classification peuvent être envisagées sous un angle de la théorie de l'évidence. Par ailleurs, la superposition du flou et des probabilités ont déjà été étudiés dans le contexte des champs ou chaînes de Markov cachés [5], [8], [36], [37]. Une étude des possibilités d'unification de ces formalismes dans le cadre Markovien constitue un autre axe éventuel des recherches.

# Références

- [1] A. APPRIOU, Probabilités et incertitude en fusion de données multi-senseurs, *Revue Scientifique et Technique de la Défense*, (11):27-40, 1991.
- [2] A. BENDJEBBOUR, Y. DELIGNON, L. FOUQUE, V. SAMSON, and W. PIECZYNSKI, Multisensor images segmentation using Dempster-Shafer fusion in Markov fields context, *IEEE Trans. on Geoscience and Remote Sensing*, 39(8):1789-1798, 2001.
- [3] B. BENMILLOUD and W. PIECZYNSKI, Estimation des paramètres dans les chaînes de Markov cachées et segmentation d'images, *Traitement du Signal*, 12(5):433-454, 1995.
- [4] N. BRUNEL and W. PIECZYNSKI, Unsupervised signal restoration using copulas and pairwise Markov chains, In *IEEE Workshop on Statistical Signal Processing (SSP 2003)*, Saint Louis, Missouri, September 28-October 1 2003.
- [5] C. CARINCOTTE, S. DERRODE, G. SICOT, and J.-M. BOUCHER, Unsupervised image segmentation based on a new fuzzy HMC model, In *ICASSP'04*, Montreal, Canada, May 17-21 2004.
- [6] B. R. COBB and P. P. SHENOY, *A Comparison of Methods for Transforming Belief Function Models to Probability Models*, volume 2711, Symbolic and Quantitative Approaches to reasoning with uncertainty – Lecture Notes in Artificial Intelligence, Springer-Verlag, 2003.
- [7] C. COLLET, J.-N. PROVOST, P. ROSTAING, P. PÉREZ, and P. BOUTHEMY, Segmentation bathymétrique d'images multispectrales SPOT, *Traitement du Signal*, 18(1):1-16, 2001.
- [8] M. DANG and G. GOVAERT, Spatial fuzzy clustering using EM and Markov random fields, *International Journal of System Research and Information Science*, 8(4):183-202, 1998.
- [9] T. DENOEU, Analysis of evidence-theoretic decision rules for pattern classification, *Pattern Recognition*, 30(7), 1997.
- [10] S. DERRODE and W. PIECZYNSKI, Signal and image segmentation using pairwise Markov chains, *IEEE Trans. on Signal Processing*, 52(9):2477-2489, 2004.
- [11] F. DESBOUVRIES and W. PIECZYNSKI, Modèles de Markov triplet et filtrage de Kalman, *CRAS*, 336(8):667-670, 2003, Serie I.
- [12] F. DESBOUVRIES and W. PIECZYNSKI, Particle filtering with pairwise Markov processes, In *ICASSP'03*, Hong-Kong, 2003.
- [13] S. FOUCHER, M. GERMAIN, J.-M. BOUCHER, and G. B. BENIÉ, Multisource classification using ICM and Dempster-Shafer theory, *IEEE Trans. on IM*, 51(2):277-281, 2002.
- [14] L. FOUQUE, A. APPRIOU, and W. PIECZYNSKI, An evidential markovian model for data fusion and unsupervised image classification, Proceedings of 3rd International conference on Information Fusion, FUSION 2000 volume 1, pages TuB4-25 - TuB4-31, Paris, France, July 10th-13th 2000.
- [15] N. GIORDANA and W. PIECZYNSKI, Estimation of generalized multisensor hidden Markov chains and unsupervised image segmentation, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(5):465-475, 1997.
- [16] S. L. HÉGARAT-MASCLE, I. BLOCH, and D. VIDAL-MADJAR, Introduction of neighborhood information on evidence theory and application to data fusion of radar and optical images with partial cloud cover, *Pattern Recognition*, 31(11):1811-1823, 1998.
- [17] F. JANEZ and A. APPRIOU, Theory of evidence and non-exhaustive frames of discernment : plausibilities correction methods, *International Journal of Approximate Reasoning*, (18):1-19, 1998.
- [18] F. V. JENSEN, *An Introduction to Bayesian Networks*, UCL Press, 2000.
- [19] J.-M. LAFERTÉ, P. PÉREZ, and F. HEITZ, Discrete Markov image modeling and inference on the quadtree, *IEEE Trans. on Image Processing*, 9(3):390-404, 2000.
- [20] P. LANCHANTIN and W. PIECZYNSKI, Arbres de Markov triplet et théorie de l'évidence, In *Actes du colloque GRETSI'03*, Paris, France, 8-11 Septembre 2003.
- [21] P. LANCHANTIN and W. PIECZYNSKI, Unsupervised non stationary image segmentation using triplet Markov chain, In *Advanced Concepts for Intelligent Vision Systems (ACIVS 2004)*, Brussels, Belgium, August 31-September 3 2004.
- [22] P. LANCHANTIN and W. PIECZYNSKI, Unsupervised restoration of non stationary Markov chain using evidential priors, *IEEE Trans. on Signal Processing*, September 2004, accepted.
- [23] E. MONFRINI, J. LECOMTE, F. DESBOUVRIES, and W. PIECZYNSKI, Image and signal restoration using pairwise Markov trees, In *IEEE Workshop on Statistical Signal Processing (SSP 2003)*, Saint Louis, Missouri, September 28-October 1 2003.
- [24] P. PÉREZ, Markov random fields and images, *CWI Quarterly*, 11(4):413-437, 1998.
- [25] P. PÉREZ, Modèles et algorithmes pour l'analyse probabiliste des images, *Habilitation à diriger des recherches de l'Université de Rennes 1*, Décembre 2003.
- [26] P. PÉREZ, A. CHARDIN, and J.-M. LAFERTÉ, Noniterative manipulation of discrete energy-based models for image analysis, *Pattern Recognition*, 33(4):573-586, 2000.
- [27] W. PIECZYNSKI, Triplet Markov chains and theory of evidence, submitted.
- [28] W. PIECZYNSKI, Arbres de Markov couple, *CRAS*, 335:79-82, 2002.
- [29] W. PIECZYNSKI, Chaînes de Markov triplet, *CRAS*, 335(1):275-278, 2002.
- [30] W. PIECZYNSKI, Arbres de Markov triplet et fusion de Dempster-Shafer, *CRAS*, 336:869-872, 2003, Serie I, ISSUE 10.
- [31] W. PIECZYNSKI, Modèles de Markov en traitement d'images, *Traitement du Signal*, 20(3):255-278, 2003.
- [32] W. PIECZYNSKI, Pairwise Markov chains, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25(5):634-639, 2003.
- [33] W. PIECZYNSKI, D. BENBOUDJEMA, and P. LANCHANTIN, Statistical image segmentation using triplet Markov fields, In *SPIE's International Symposium on Remote Sensing*, pages 92-101, Crete, Greece, September 22-27 2002.
- [34] W. PIECZYNSKI and A.-N. TEBBACHE, Pairwise Markov random fields and segmentation of textured images, *Machine Graphics and Vision*, 9(3):705-718, 2000.
- [35] J.-N. PROVOST, C. COLLET, P. ROSTAING, P. PÉREZ, and P. BOUTHEMY, Hierarchical Markovian segmentation of multispectral images for reconstruction of water depth maps, *Computer Vision and Image Understanding*, 93(2):155-174, 2004.
- [36] S. RUAN, B. MORETTI, J. FADILI, and D. BLOYET, Fuzzy Markovian segmentation in application of magnetic resonance images, *Computer Vision and Image Understanding*, 85(1):54-69, 2002.
- [37] F. SALZENSTEIN and W. PIECZYNSKI, Parameter estimation in hidden fuzzy Markov random fields and image segmentation, *Graphical Models and Image Processing*, 59(4):205-220, 1997.
- [38] F. SALZENSTEIN and W. PIECZYNSKI, Sur le choix de méthode de segmentation statistique d'images, *Traitement du Signal*, 15(2):119-128, 1998.
- [39] G. SHAFER, *A mathematical theory of evidence*, Princeton University Press, Princeton, 1976.
- [40] P. SMETS, Belief functions: the disjunctive rule of combination and the generalized bayesian theorem, *International Journal of Approximate Reasoning*, 9:1-35, 1993.
- [41] J. J. SUDANO, Equivalence between belief theories and naïve Bayesian fusion for systems with independent evidential data, In *Proceeding of The Sixth International Conference on Information Fusion*, 2003.
- [42] J. WHITTAKER, *Graphical models in applied multivariate statistics*, Wiley, 1990.
- [43] A. S. WILLSKY, Multiresolution Markov models for signal and image processing, *Proceedings of IEEE*, 90(8):1396-1458, 2002.
- [44] G. WINKLER, *Image analysis, random fields and Markov Chain Monte Carlo Methods: a mathematical introduction*, Springer, 2003.
- [45] R. R. YAGER, J. KACPRZYK, and M. FEDRIZZI, *Advances in the Dempster-Shafer theory of evidence*, Wiley, 1994.



Pierre **Lanchantin**

Pierre Lanchantin est né à Paris en 1976. Il est titulaire du DEA d'Acoustique, Traitement du signal, Informatique appliqués à la Musique (ATIAM) de l'Université Paris VI (2001), du Diplôme d'Ingénieur en Télécommunications de l'Institut National des Télécommunications (INT), Evry, et du DEA d'Optimisation et Sécurité des Systèmes (OSS) de l'Université de Technologie de Troyes (UTT)(2002). Depuis octobre 2002, il est doctorant au département Communications, Images et Traitement de l'Information (CITI), INT. Son domaine d'intérêt est le traitement statistique du signal et de l'image par modèles de Markov cachés.



Wojciech **Pieczynski**

Wojciech Pieczynski est titulaire d'un Doctorat d'Etat en Mathématiques obtenu à l'Université Paris VI en 1986. Il a enseigné au Centre Universitaire de Tiaret, Algérie, à l'Université de Brazzaville, Congo, et à l'Ecole Nationale Supérieure des Télécommunications de Bretagne, Brest. Actuellement Professeur et responsable du département Communications, Images et Traitement de l'Information (CITI) à l'Institut National des Télécommunications, Evry. Ses recherches portent sur les modélisations probabilistes et les traitements statistiques d'images.

