



**HAL**  
open science

## Chaînes de Markov cachées à bruit généralisé

Hugo Gangloff, Katherine Morales, Yohan Petetin

► **To cite this version:**

Hugo Gangloff, Katherine Morales, Yohan Petetin. Chaînes de Markov cachées à bruit généralisé. 2022. hal-03698101

**HAL Id: hal-03698101**

**<https://hal.science/hal-03698101v1>**

Preprint submitted on 17 Jun 2022

**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.

# Chaînes de Markov cachées à bruit généralisé

Hugo GANGLOFF<sup>1</sup>, Katherine MORALES<sup>2</sup>, Yohan PETETIN<sup>2</sup>

<sup>1</sup>IRISA, Université Bretagne Sud, UMR 6074, 56000 Vannes, France

<sup>2</sup>Samovar, Telecom Sudparis, Institut Polytechnique de Paris, 91011 Évry, France

hugo.gangloff@irisa.fr,

katherine.morales\_quinga@telecom-sudparis.eu, yohan.petetin@telecom-sudparis.eu

**Résumé** – Les chaînes de Markov cachées sont des modèles très populaires pour le traitement non-supervisé du signal dans un contexte bayésien. Nous nous intéressons à la complexification du terme d’attache aux données afin de pouvoir apprendre automatiquement des modèles de bruits très généraux. Pour cela, nous introduisons un processus auxiliaire continu et nous développons une approche d’inférence variationnelle efficace, préservant l’interprétabilité des variables cachées d’intérêt. Les résultats d’expériences de classification non-supervisée montrent que ces nouveaux modèles à variable auxiliaire continue donnent des meilleurs résultats que les modèles classiques de la littérature.

**Abstract** – Hidden Markov chains are popular models for unsupervised signal processing in a Bayesian context. We aim at complexifying the conditional likelihood term to automatically learn complex noises. We introduce a third auxiliary continuous process and we develop a variational inference approach which preserves the interpretability of the hidden process of interest. Our results in unsupervised classification experiments suggest that using a third continuous auxiliary random variable gives better results than the classical approaches from the literature.

## 1 Introduction

Dans les modèles à variables cachées, dans un cadre de travail bayésien, l’objectif de la classification bayésienne est d’estimer des réalisations de variables aléatoires (v.a.) cachées  $h_k$ , pour  $k$  dans  $\{0, \dots, K\}$ , à partir des réalisations des v.a. observées  $x_k$ . Les vraies distributions *a posteriori*  $p(h_k | \mathbf{x}_K)$  doivent donc être estimées à partir d’une modélisation de la loi jointe du couple  $(\mathbf{h}_K, \mathbf{x}_K)$ , paramétrée par un ensemble de paramètres  $\theta$ , que nous notons  $p_\theta(\mathbf{h}_K, \mathbf{x}_K)$ . Une fois le modèle choisi, la question de l’estimation de  $\theta$  se pose. Un estimateur populaire est celui du Maximum de Vraisemblance (MV) qui est tel que  $\hat{\theta}^{\text{MV}} = \arg \max_{\theta} p_\theta(\mathbf{x}_K)$  [12]. L’estimation des états cachés (ou inférence) s’obtient alors classiquement à l’aide du critère du Maximum A Posteriori (MAP), *i.e.*,  $\hat{h}_k^{\text{MAP}} = \arg \max_{h_k} p_{\hat{\theta}^{\text{MV}}}(h_k | \mathbf{x}_K)$ ,  $\forall k$ .

Les chaînes de Markov sont des modèles à données cachées très populaires pour la modélisation de données séquentielles dans un cadre de travail bayésien non-supervisé. Ce modèle a été utilisé dans de nombreux domaines comme le traitement du langage, d’images, de données génétiques, etc. [2].

Dans le reste de cet article, nous considérons des chaînes de Markov cachées à états cachés discrets (resp. observés continus) représentés par des réalisations d’une suite de v.a.  $\mathbf{h}_K = (h_0, \dots, h_K)$  où  $h_k \in \Omega = \{\omega_c\}_{c=1}^C$  (resp.  $\mathbf{x}_K = (x_0, \dots, x_K)$  où  $x_k \in \mathbb{R}^{d_x}$ ). Dans le modèle de Chaîne de Markov Cachée (CMC) [10], la loi jointe  $(\mathbf{h}_K, \mathbf{x}_K)$  est modélisée par

$$p_\theta(\mathbf{h}_K, \mathbf{x}_K) = p_\theta(h_0) \prod_{k=1}^K p_\theta(h_k | h_{k-1}) \prod_{k=0}^K p_\theta(x_k | h_k). \quad (1)$$

Les distributions *a posteriori* pour l’estimation des états cachés  $\mathbf{h}_K$  sont obtenues par des calculs exacts via l’algorithme Forward Backward [10]. Nous avons que

$$p_\theta(\mathbf{h}_{k-1:k} | \mathbf{x}_K) \propto \alpha_{\theta,k-1}(h_{k-1}) \beta_{\theta,k}(h_k) p_\theta(h_k | h_{k-1}) p_\theta(x_k | h_k), \quad (2)$$

où les  $\alpha_{\theta,k}(h_k) = p_\theta(h_k, \mathbf{x}_{0:k})$  sont obtenus via

$$\alpha_{\theta,k}(h_k) = \sum_{h_{k-1}} \alpha_{\theta,k-1}(h_{k-1}) p_\theta(h_k | h_{k-1}) p_\theta(x_k | h_k), \quad (3)$$

et les  $\beta_{\theta,k}(h_k) = p_\theta(\mathbf{x}_{k+1:K} | h_k, x_k)$  par

$$\beta_{\theta,k}(h_k) = \sum_{h_{k+1}} \beta_{\theta,k+1}(h_{k+1}) p_\theta(h_{k+1} | h_k) p_\theta(x_{k+1} | h_{k+1}). \quad (4)$$

Notons que la vraisemblance est aussi calculable via des calculs exacts avec

$$p_\theta(\mathbf{x}_K) = \sum_{\mathbf{h}_K} \alpha_{\theta,K}(h_K). \quad (5)$$

La complexification du modèle des CMC est l’objet de nombreuses études afin d’améliorer les dépendances directes disponibles dans le modèle pour, *in fine*, mieux modéliser les données et améliorer l’estimation des états cachés. Nous nous intéressons à la complexification de la vraisemblance conditionnelle,  $p_\theta(x_k | h_k)$ , aussi appelé terme d’attache aux données (c’est la relation entre l’observation et l’état caché à un instant  $k$ ). Dans cette optique, l’introduction d’une troisième v.a. auxiliaire et cachée  $\mathbf{z}_K = (z_0, \dots, z_K)$  est souvent proposée dans la littérature. Nous définissons maintenant un modèle de CMC incluant une troisième suite de v.a.  $\mathbf{z}_K$ .

Le modèle de CMC à Bruit Généralisé (CMC-BG) a pour but la complexification du terme d'attache aux données via l'introduction d'une dépendance sur la réalisation d'une v.a.  $z_k$ . Dans une CMC-BG, nous définissons la loi jointe d'un triplet  $\mathbf{t}_K = (\mathbf{h}_K, \mathbf{z}_K, \mathbf{x}_K)$ ,

$$p_{\boldsymbol{\theta}}(\mathbf{t}_K) = p_{\boldsymbol{\theta}}(h_0) \prod_{k=1}^K p_{\boldsymbol{\theta}}(h_k|h_{k-1}) \prod_{k=0}^K p_{\boldsymbol{\theta}}(z_k)p_{\boldsymbol{\theta}}(x_k|h_k, z_k), \quad (6)$$

où les  $z_k$  sont indépendants,  $\forall k$ . Deux cas sont à envisager.

Si  $\mathbf{z}_K$  est discret à valeurs dans  $\Lambda$ , alors les rappels précédents s'appliquent, il suffit de considérer un espace discret caché étendu dont les valeurs sont dans  $\Omega \times \Lambda$ . Ce cas est déjà connu et permet, par exemple, de modéliser des stationnarités différentes dans le processus de bruit du modèle [4] [9]. Lorsque  $\mathbf{z}_K$  est discret nous notons le modèle CMC-BG-di.

Si  $\mathbf{z}_K$  est continu, à valeurs dans  $\mathbb{R}^{d_z}$ , les éléments vu auparavant ne s'appliquent plus. En particulier les récursions Forward Backward ne sont plus calculables exactement, les lois d'intérêt ne peuvent plus être calculées et l'estimation MV n'est plus faisable. Dans cet article nous proposons une méthodologie inédite, basée sur l'inférence variationnelle, pour l'estimation des paramètres et l'inférence dans les CMC-BG avec  $\mathbf{z}_K$  continu (modèle noté CMC-BG quand il n'y a pas de confusion possible). Nous montrons l'intérêt d'une troisième v.a. continue qui peut, en théorie, mener à un modèle de bruit arbitrairement complexe. Nous proposons également des éléments de réponse au problème de l'interprétabilité des  $h_k$  dans les CMC-BG.

L'article est organisé comme suit. Dans les sections suivantes nous effectuons des rappels sur l'inférence variationnelle et développons son adaptation pour l'estimation des paramètres dans les CMC-BG. Nous présentons ensuite l'inférence dans ce modèle et concluons par une section d'expériences.

## 2 Rappels sur l'inférence variationnelle

Dans un contexte général, notons  $z$  (resp.  $x$ ) une v.a. cachée (resp. observée) à valeurs  $\mathbb{R}$ . L'inférence variationnelle permet l'estimation des paramètres  $\boldsymbol{\theta}$  du modèle lorsque les lois *a posteriori*,  $p_{\boldsymbol{\theta}}(z|x)$ , ne sont pas connues et doivent être approximées par une loi *variationnelle*,  $q_{\boldsymbol{\varphi}}(z|x)$ , paramétrée par  $\boldsymbol{\varphi}$ , un ensemble de paramètres devant lui-même être estimé. Nous maximisons pour cela une borne inférieure de la vraisemblance, notée  $F(\boldsymbol{\theta}, \boldsymbol{\varphi})$  et nommée ELBO [6],

$$\log p_{\boldsymbol{\theta}}(x) \geq \int q_{\boldsymbol{\varphi}}(z|x) \log \frac{p_{\boldsymbol{\theta}}(z, x)}{q_{\boldsymbol{\varphi}}(z|x)} dz = F(\boldsymbol{\theta}, \boldsymbol{\varphi}). \quad (7)$$

Il y a égalité dans l'Éq. (7) lorsque  $q_{\boldsymbol{\varphi}}(z|x) = p_{\boldsymbol{\theta}}(z|x)$ . La maximisation alternée de l'ELBO par rapport à  $\boldsymbol{\varphi}$  et  $\boldsymbol{\theta}$  est également connue comme l'algorithme Variational EM [11]. On peut écrire

$$F(\boldsymbol{\theta}, \boldsymbol{\varphi}) = \mathbb{E}_{q_{\boldsymbol{\varphi}}(z|x)}(p_{\boldsymbol{\theta}}(x|z)) - \text{D}_{\text{KL}}(q_{\boldsymbol{\varphi}}(z|x)||p_{\boldsymbol{\theta}}(z)), \quad (8)$$

l'ELBO est ainsi la somme d'un terme dit *de reconstruction*, qui favorise des variables cachées permettant de reconstruire

fidèlement les variables observées, et d'un terme de divergence de Kullback Leibler forçant  $q_{\boldsymbol{\varphi}}(z|x)$  à être proche de la loi *a priori*  $p_{\boldsymbol{\theta}}(z)$ .

## 3 Inférence variationnelle dans les CMC-BG

Nous proposons donc d'effectuer l'estimation des paramètres dans les CMC-BG par inférence variationnelle. En effet, les lois *a posteriori*  $p_{\boldsymbol{\theta}}(\mathbf{h}_K, \mathbf{z}_K|\mathbf{x}_K)$  doivent ici être approximées. Cependant, trois adaptations importantes vont apparaître par rapport à l'inférence variationnelle classique. D'une part nous avons des suites de v.a. et de plus, les v.a. cachées se divisent en une partie discrète  $\mathbf{h}_K$  et une partie continue  $\mathbf{z}_K$ . Enfin, les problèmes d'interprétabilité des  $\hat{\mathbf{h}}_K$  doivent être étudiés.

### 3.1 Une ELBO optimale

Introduisons  $q_{\boldsymbol{\varphi}}(\mathbf{h}_K, \mathbf{z}_K|\mathbf{x}_K)$  comme loi variationnelle approximant  $p_{\boldsymbol{\theta}}(\mathbf{h}_K, \mathbf{z}_K|\mathbf{x}_K)$ . Nous avons

$$q_{\boldsymbol{\varphi}}(\mathbf{h}_K, \mathbf{z}_K|\mathbf{x}_K) = q_{\boldsymbol{\varphi}}(\mathbf{h}_K|\mathbf{z}_K, \mathbf{x}_K)q_{\boldsymbol{\varphi}}(\mathbf{z}_K|\mathbf{x}_K), \quad (9)$$

ou  $q_{\boldsymbol{\varphi}}(\mathbf{h}_K|\mathbf{z}_K, \mathbf{x}_K)$  est calculable exactement par l'algorithme Forward Backward, avec un espace des observations étendu  $(\mathbf{z}_K, \mathbf{x}_K)$ . En effet, il nous est possible d'obtenir des réalisations de  $\mathbf{z}_K$  par tirage selon  $q_{\boldsymbol{\varphi}}(\mathbf{z}_K|\mathbf{x}_K)$ . Dans ce cas le choix de la loi variationnelle se réduit au choix de  $q_{\boldsymbol{\varphi}}(\mathbf{z}_K|\mathbf{x}_K)$  et nous utilisons une loi variationnelle optimale

$$q_{\boldsymbol{\varphi}}^{\text{opt}}(\mathbf{h}_K, \mathbf{z}_K|\mathbf{x}_K) = p_{\boldsymbol{\theta}}(\mathbf{h}_K|\mathbf{z}_K, \mathbf{x}_K)q_{\boldsymbol{\varphi}}(\mathbf{z}_K|\mathbf{x}_K), \quad (10)$$

associée à l'ELBO optimale  $F^{\text{opt}}(\boldsymbol{\theta}, \boldsymbol{\varphi})$ <sup>1</sup>. Finalement, dans le cas séquentiel des CMC-BG, on peut montrer que l'ELBO s'écrit

$$\begin{aligned} F^{\text{opt}}(\boldsymbol{\theta}, \boldsymbol{\varphi}) &= \mathbb{E}_{q_{\boldsymbol{\varphi}}^{\text{opt}}(\mathbf{h}_K, \mathbf{z}_K|\mathbf{x}_K)}(\log \prod_{k=0}^K p_{\boldsymbol{\theta}}(x_k|h_k, z_k)) \\ &\quad - \text{D}_{\text{KL}}(q_{\boldsymbol{\varphi}}^{\text{opt}}(\mathbf{h}_K, \mathbf{z}_K|\mathbf{x}_K)||p_{\boldsymbol{\theta}}(h_0)p_{\boldsymbol{\theta}}(z_0)\prod_{k=1}^K p_{\boldsymbol{\theta}}(z_k)p_{\boldsymbol{\theta}}(h_k|h_{k-1})). \end{aligned} \quad (11)$$

### 3.2 Pénalisation de l'ELBO

Nous proposons enfin deux modifications de  $F^{\text{opt}}(\boldsymbol{\theta}, \boldsymbol{\varphi})$  afin de pouvoir gérer les problèmes d'interprétabilité de l'estimation de  $\mathbf{h}_K$ , en offrant plus de flexibilité dans la fonction à maximiser.

Tout d'abord nous ajoutons un terme basé sur une entropie-croisée qui force la prise en compte d'une pré-segmentation, obtenue par un modèle plus simple, à moindre coût, et que l'on sait interprétable. Cette approche est similaire à [7] [8]. Ensuite, nous pénalisons des termes de l'ELBO par des scalaires  $\beta_1$  et  $\beta_2$  qui permettent la prise en compte, plus ou moins forte,

1. Optimale dans le sens où nous pouvons montrer que :

$$\log p_{\boldsymbol{\theta}}(\mathbf{x}_K) \geq F^{\text{opt}}(\boldsymbol{\theta}, \boldsymbol{\varphi}) \geq F(\boldsymbol{\theta}, \boldsymbol{\varphi}).$$

de la loi *a priori* ou du terme d'entropie croisée [5]. Finalement, la fonction de coût maximisée est

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\varphi}) = & \mathbb{E}_{q_{\boldsymbol{\varphi}}^{\text{opt}}(\mathbf{h}_K, \mathbf{z}_K | \mathbf{x}_K)} \left( \log \prod_{k=0}^K p_{\boldsymbol{\theta}}(x_k | h_k, z_k) \right) \\ & - \beta_1 \text{D}_{\text{KL}}(q_{\boldsymbol{\varphi}}^{\text{opt}}(\mathbf{h}_K, \mathbf{z}_K | \mathbf{x}_K) || p_{\boldsymbol{\theta}}(h_0) p_{\boldsymbol{\theta}}(z_0) \prod_{k=1}^K p_{\boldsymbol{\theta}}(z_k) p_{\boldsymbol{\theta}}(h_k | h_{k-1})) \\ & + \beta_2 \mathbb{E}_{p^{\text{emp}}(\mathbf{h}_K)} (\log q_{\boldsymbol{\varphi}}(\mathbf{h}_K | \mathbf{x}_K)), \end{aligned} \quad (12)$$

En pratique les différents termes de  $\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\varphi})$  sont approxi- més par des tirages avec *reparametrization trick* [6] selon la loi  $q_{\boldsymbol{\varphi}}(\mathbf{z}_K | \mathbf{x}_K)$ , en effet,  $q_{\boldsymbol{\varphi}}^{\text{opt}}(\mathbf{h}_K, \mathbf{z}_K | \mathbf{x}_K)$  s'écrit selon l'Éq. (10) et les sommations suivant  $p_{\boldsymbol{\theta}}(h_k | \mathbf{z}_K, \mathbf{x}_K)$  sont calculables exacte- ment. L'Algo. 1 résume la maximisation de l'ELBO pour l'estimation des paramètres dans les CMC-BG.

**Entrée :**  $\mathbf{x}_K$ , les observations  
**Résultat :**  $(\boldsymbol{\theta}^*, \boldsymbol{\varphi}^*)$ , les paramètres estimés

- 1 Initialiser  $(\boldsymbol{\theta}^{(0)}, \boldsymbol{\varphi}^{(0)})$
- 2  $t \leftarrow 0$
- 3 **tant que** convergence de  $\mathcal{L}(\boldsymbol{\theta}^{(t)}, \boldsymbol{\varphi}^{(t)})$  pas atteinte **faire**
- 4     Tirer  $\mathbf{z}_K^{(n)} \sim q_{\boldsymbol{\varphi}}(\mathbf{z}_K | \mathbf{x}_K)$ , pour tout  $1 \leq n \leq N$
- 5     Calculer  $p_{\boldsymbol{\theta}}(h_{k-1} | h_k, \mathbf{z}_K^{(n)}, \mathbf{x}_K)$ , pour tout  $h_{k-1:k} \in \Omega \times \Omega$ , pour tout  $1 \leq n \leq N$ , pour tout  $1 \leq k \leq K$
- 6     Calculer la dérivée  $\nabla_{(\boldsymbol{\theta}, \boldsymbol{\varphi})} \mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\varphi})$
- 7     Mettre à jour les paramètres
- 8      $\begin{pmatrix} \boldsymbol{\theta}^{(t+1)} \\ \boldsymbol{\varphi}^{(t+1)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\theta}^{(t)} \\ \boldsymbol{\varphi}^{(t)} \end{pmatrix} + \epsilon \nabla_{(\boldsymbol{\theta}, \boldsymbol{\varphi})} \mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\varphi}) \Big|_{(\boldsymbol{\theta}^{(t)}, \boldsymbol{\varphi}^{(t)})}$
- 9      $t \leftarrow t + 1$
- 10 **fin**
- 11  $\boldsymbol{\theta}^* \leftarrow \boldsymbol{\theta}^{(t)}$
- 12  $\boldsymbol{\varphi}^* \leftarrow \boldsymbol{\varphi}^{(t)}$

**Algorithme 1 :** Estimation des paramètres dans les CMC-BG

### 3.3 Inférence des $h_K$ dans les CMC-BG

Une fois les paramètres  $(\boldsymbol{\theta}^*, \boldsymbol{\varphi}^*)$  estimés il faut estimer les états cachés. Pour cela nous pouvons observer que

$$p_{\boldsymbol{\theta}^*}(h_k | \mathbf{x}_K) = \frac{\mathbb{E}_{q_{\boldsymbol{\varphi}^*}(\mathbf{z}_K | \mathbf{x}_K)} \left( \frac{p_{\boldsymbol{\theta}^*}(h_k | \mathbf{z}_K, \mathbf{x}_K) p_{\boldsymbol{\theta}^*}(\mathbf{z}_K | \mathbf{x}_K)}{q_{\boldsymbol{\varphi}^*}(\mathbf{z}_K | \mathbf{x}_K)} \right)}{\mathbb{E}_{q_{\boldsymbol{\varphi}^*}(\mathbf{z}_K | \mathbf{x}_K)} \left( \frac{p_{\boldsymbol{\theta}^*}(\mathbf{x}_K)}{q_{\boldsymbol{\varphi}^*}(\mathbf{z}_K | \mathbf{x}_K)} \right)}, \quad (13)$$

qui est calculable par un algorithme d'échantillonnage d'im- portance séquentiel [1]. Dans cette approche  $q_{\boldsymbol{\varphi}^*}(\mathbf{z}_K | \mathbf{x}_K)$  est la loi d'importance estimée avec l'Algo. 1. Notons que  $p_{\boldsymbol{\theta}^*}(\mathbf{z}_K | \mathbf{x}_K)$  est calculable directement avec Éq. (5) (en considérant une vraisemblance augmentée).

## 4 Segmentation non-supervisée d'images

Dans cette section, nous considérons deux expériences de segmentation binaire ( $C = 2$ ) d'images dans un contexte non-supervisé. Les images proviennent de la Binary Shape Data- base<sup>2</sup> et seront artificiellement bruitées par des bruits complexes. Pour ce faire, les images sont transformées en des signaux 1D grâce à un parcours de Hilbert-Peano. Nous commençons par présenter les modèles utilisés dans les expériences.

### 4.1 Modèles

Dans tous les modèles qui suivront nous avons la transi- tion des états cachés donnée par  $p_{\boldsymbol{\theta}}(h_k = \omega_1 | h_{k-1} = \omega_i) = \text{sigm}(b_{\omega_i})$  où  $\text{sigm}$  est la fonction sigmoïde et  $b_{\omega_i} \in \mathbb{R}$ . Dans le modèle CMC,  $p_{\boldsymbol{\theta}}(x_k | h_k = \omega_j) = \mathcal{N}(x_k; \mu_{\omega_j}; \sigma_{\omega_j}^2)$ , où  $\mathcal{N}$  désigne la distribution gaussienne de moyenne  $\mu_{\omega_j}$  et variance  $\sigma_{\omega_j}^2$ . Cette loi est complexifiée dans les CMC-BG. Dans les CMC-BG-di,  $z_k \in \{\lambda_1, \lambda_2\}$  et  $p_{\boldsymbol{\theta}}(x_k | h_k = \omega_j, z_k = \lambda_i) = \mathcal{N}(x_k; \mu_{\omega_j, \lambda_i}; \sigma_{\omega_j, \lambda_i}^2)$ . Dans les CMC-BG,  $z_k \sim \mathcal{N}(0, 1)$ , et la loi variationnelle est du type

$$q_{\boldsymbol{\varphi}}(\mathbf{z}_K | \mathbf{x}_K) = \prod_{k=1}^K q_{\boldsymbol{\varphi}}(z_k | z_{k-1}, x_k) = \prod_{k=1}^K \mathcal{N}(z_k; \nu_{\boldsymbol{\varphi}}(z_{k-1}, x_k)). \quad (14)$$

où la fonction  $\nu_{\boldsymbol{\varphi}}(z_{k-1}, x_k)$  est paramétrée par un réseau de neurones à une couche cachée de 100 neurones et à fonction d'activation ReLU. Enfin, dans les CMC-BG,  $p_{\boldsymbol{\theta}}(x_k | h_k = \omega_j, z_k) = \mathcal{N}(x_k; \mu_{\omega_j} * z_k + d_{\omega_j}; \sigma_{\omega_j}^2)$ ,  $d_{\omega_j} \in \mathbb{R}$ , tandis que dans les CMC- BG profondes [3] (D-CMC-BG), la moyenne de  $p_{\boldsymbol{\theta}}(x_k | h_k = \omega_j, z_k)$  est paramétrée par un réseau de neurones à une couche cachée de 100 neurones et à fonction d'activation ReLU.

### 4.2 Expériences

**Bruit multiplicatif** - Dans ce premier scénario, les images de classes sont celles du type *camel* du jeu de données et sont bruitées selon un bruit multiplicatif gaussien,

$$x_k | h_k, z_k \sim \mathcal{N}(a_{h_k}; b_{h_k}^2) * z_k, \quad (15)$$

où  $z_k \sim \mathcal{N}(0, 1)$ ,  $a_{\omega_1} = 0$ ,  $a_{\omega_2}$  varie et  $b_{\omega_1} = b_{\omega_2} = 0.2$ . La Fig. 1 illustre les résultats des modèles. Dans notre approche variationnelle nous avons  $\beta_1 = 5$ ,  $\beta_2 = 1$ .  $\beta_1$  permet de renfor- cer l'influence de la loi *a priori* ce qui semble être favorable en vertu de la génération des  $\mathbf{z}_K$ .  $\beta_2$  est aussi requis ici et semble guider  $\hat{\mathbf{h}}_K$  vers la segmentation d'intérêt (interprétabilité). Des illustrations sont données dans la Fig. 2. Nous notons que les modèles de CMC-BG améliorent le modèle de CMC (jusqu'à 7 points). Cela atteste de l'intérêt d'un processus auxiliaire  $\mathbf{z}_K$ . De plus un léger avantage semble aller en faveur des CMC-BG avec  $\mathbf{z}_K$  continu.

**Bruit non-stationnaire** - Dans cette deuxième expérience, les images de classes sont celles du type *dog* du jeu de données

2. <http://vision.lerns.brown.edu/content/available-software-and-databases>

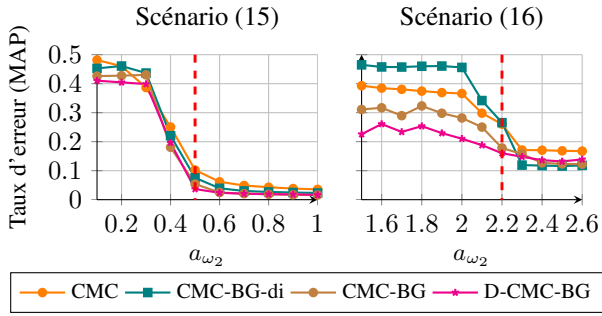


FIGURE 1 – Taux d’erreur dans les expériences de segmentation non-supervisée des Scénarios (15) et (16). Les résultats sont moyennés sur toutes les images de la catégorie.

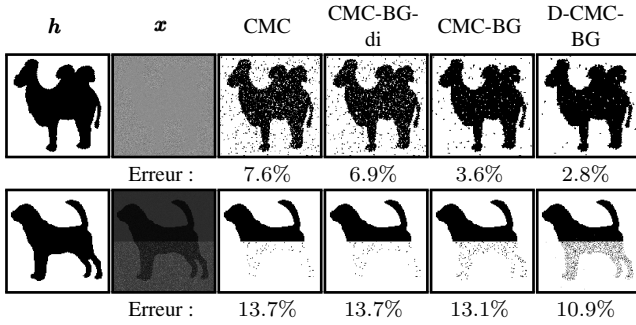


FIGURE 2 – Haut (resp. bas) : illustrations du Scénario (15) (resp. Scénario (16)) pour  $a_{\omega_2} = 0.5$  (resp.  $a_{\omega_2} = 2.2$ ) (correspond aux lignes pointillées rouges dans la Fig. 1).

et sont bruitées selon un bruit non-stationnaire,

$$\begin{cases} x_k | h_k \sim \mathcal{N}(a_{h_k}; \sigma^2), & \text{if } k \in \{1, \dots, \lfloor K/2 \rfloor\}, \\ x_k | h_k \sim a_{h_k} + \mathcal{E}(\lambda), & \text{if } k \in \{\lfloor K/2 \rfloor + 1, \dots, K\}, \end{cases} \quad (16)$$

où  $\mathcal{E}(\lambda)$  est la distribution de probabilité exponentielle de paramètre  $\lambda = 1.4$ ,  $a_{\omega_1} = 0$ ,  $a_{\omega_2}$  varie et  $\sigma = 0.2$ . La difficulté dans la segmentation de ces images est qu’un même modèle doit être capable de modéliser deux distributions d’attache aux données différentes. Ici nous choisissons  $\beta_1 = 0.1$  et  $\beta_2 = 0$ . Une faible valeur de  $\beta_1 = 0.1$  semble permettre plus de flexibilité pour coller aux observations et apprendre le bruit complexe.  $\beta_2 = 0$  car nous ne possédons pas de présegmentation exploitable issue d’un modèle plus simple. Les résultats sont donnés dans la Fig. 1 tandis qu’une illustration est donnée dans Fig. 2. Il est clair qu’en moyenne, les modèles CMC-BG (simple et profond) donnent de meilleurs résultats. CMC et CMC-BG-di deviennent inopérant lorsque  $a_{\omega_2} < 2$ . Cette expérience illustre à nouveau l’utilité d’un processus auxiliaire continu pour apprendre des modèles de bruits complexes.

## 5 Conclusion

Dans cette article nous proposons une nouvelle approche pour la modélisation de bruits très généraux dans les CMC, via l’utilisation d’un processus auxiliaire continu. Nous dévelop-

pons une approche variationnelle qui préserve l’interprétabilité des variables cachées d’intérêt estimées et permet des calculs efficaces. Nous montrons que les nouveaux modèles à v.a. auxiliaires continus donnent de meilleurs résultats que les modèles de l’état de l’art à v.a. auxiliaires discrètes.

## Références

- [1] A. DOUCET et A. M. JOHANSEN. “A tutorial on particle filtering and smoothing : Fifteen years later”. In : *Handbook of nonlinear filtering* 12.656-704 (2009), p. 3.
- [2] Przemyslaw DYMARSKI. *Hidden Markov Models*. IntechOpen, 2011.
- [3] H. GANGLOFF et al. “Unsupervised image segmentation with Gaussian Pairwise Markov Fields”. In : *Computational Statistics & Data Analysis* 158 (2021), p. 107178.
- [4] I. GORYNIN et al. “Assessing the segmentation performance of pairwise and triplet Markov Models”. In : *Signal Processing* 145 (2018), p. 183-192.
- [5] I. HIGGINS et al. “Beta-VAE : Learning Basic Visual Concepts with a Constrained Variational Framework”. In : *5th International Conference on Learning Representations, ICLR*. OpenReview.net, 2017.
- [6] D. P. KINGMA et M. WELLING. “Auto-encoding variational Bayes”. In : *2nd International Conference on Learning Representations, ICLR*. 2014.
- [7] J. KLYS, J. SNELL et R. ZEMEL. “Learning Latent Subspaces in Variational Autoencoders”. In : *Advances in Neural Information Processing Systems 31 : Annual Conference on Neural Information Processing Systems*. 2018, p. 6445-6455.
- [8] S. KUMAR, J. PRADEEP et H. ZAIDI. “Learning Robust Latent Representations for Controllable Speech Synthesis”. In : *Findings of the Association for Computational Linguistics : ACL/IJCNLP*. T. ACL/IJCNLP 2021. Findings of ACL. Association for Computational Linguistics, 2021, p. 3562-3575.
- [9] Pierre LANCHANTIN, Jérôme LAPUYADE-LAHORGUE et Wojciech PIECZYNSKI. “Unsupervised segmentation of randomly switching data hidden with non-Gaussian correlated noise”. In : *Signal Processing* 91.2 (2011), p. 163-175.
- [10] L. R. RABINER. “A tutorial on Hidden Markov Models and selected applications in speech recognition”. In : *Proceedings of the IEEE* 77.2 (1989), p. 257-286.
- [11] D. G. TZIKAS, A. C. LIKAS et N. P. GALATSANOS. “The variational approximation for Bayesian inference”. In : *IEEE Signal Processing Magazine* 25.6 (2008), p. 131-146.
- [12] H. WHITE. “Maximum Likelihood Estimation of Misspecified Models”. In : *Econometrica* 50.1 (1982), p. 1-25.