# Dempster–Shafer Fusion of Evidential Pairwise Markov Chains

Mohamed El Yazid Boudaren and Wojciech Pieczynski

***Abstract*—Hidden Markov models have been extended in many directions, leading to pairwise Markov models, triplet Markov models, or discriminative random fields, all of which have been successfully applied in many fields covering signal and image processing. The Dempster–Shafer theory of evidence has also shown its interest in a wide range of situations involving reasoning under uncertainty and/or information fusion. There are, however, only few works dealing with both of these modeling tools simultaneously. The aim of this paper, which falls under this category of works, is to propose a general evidential Markov model offering wide modeling and processing possibilities regarding information imprecision, sensor unreliability, and data fusion. The main interest of the proposed model relies in the possibility of achieving, easily, the Dempster–Shafer fusion without destroying the Markovianity.**

***Index Terms*—Dempster–Shafer (DS) fusion, hidden Markov chains (HMCs), theory of evidence, triplet Markov chains (TMCs).**

## Nomenclature

| Acronym | Designation |
|---|---|
| *bba* | Basic belief assignment |
| CEMMC | Conditional evidential marginal Markov chain |
| CEPMC | Conditional evidential pairwise Markov chain |
| DS | Dempster-Shafer |
| EM | Expectation-Maximization |
| EMC | Evidential Markov chain |
| EMM | Evidential marginal Markov |
| EMMC | Evidential marginal Markov chain |
| EPMC | Evidential pairwise Markov chain |
| HEMC | Hidden evidential Markov chain |
| HMC | Hidden Markov chain |
| HMM | Hidden Markov model |
| MPM | Marginal posterior mode |
| PMC | Pairwise Markov chain |
| TMC | Triplet Markov chain |

## I. Introduction

Let us consider two random sequences $X_1^N = (X_1, \ldots, X_N)$ and $Y_1^N = (Y_1, \ldots, Y_N)$, where $X_1^N$ takes its values in $\Omega = \{\omega_1, \ldots, \omega_K\}$, whereas $Y_1^N$ takes its values in $R$. The sequence $X_1^N$ is hidden, while $Y_1^N$ is observed. The aim then is to recover $X_1^N$ from $Y_1^N$. Realizations of such processes will be denoted by lowercase letters. Probabilistic links between $X_1^N$ and $Y_1^N$ are given by the distribution $p(x_1^N, y_1^N)$ of the couple $(X_1^N, Y_1^N)$, and the problem is to define $p(x_1^N, y_1^N)$ in such a way that the search of $X_1^N = x_1^N$ would be workable for large $N$. The estimator of $X_1^N = x_1^N$ from $Y_1^N = y_1^N$ considered in this paper will be the classic MPM [1] estimator defined by

$$[\hat{x}_1^N = (x_1, \ldots, x_N) = \hat{s}(y_1^N)] \Leftrightarrow [\forall n \in \{1, \ldots, N\}, \hat{x}_n = \arg\max_{\omega \in \Omega} p(x_n = \omega | y_1^N)]. \quad (1)$$

One possible choice for $p(x_1^N, y_1^N)$, which makes the computation of $p(x_n|y_1^N)$ feasible with a linear complexity in $N$, is the classic HMC model. HMCs turn out to be very robust and have been extensively used to solve various inverse problems occurring in a wide range of fields including signal and image processing [2], pattern recognition [3], time series [4], finance [5], [6], and biology [7], [8]. Let us also mention the pioneering papers [1], [9], [10] and the general books [11], [12].

In HMC, the distribution $p(x_1^N, y_1^N)$ is given by

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

The MPM estimation can then be achieved thanks to the possibility of recursive computations of forward probabilities $\alpha_n(\omega) = p(x_n{=}\omega, y_1^n)$ and backward probabilities $\beta_n(\omega) = p(y_{n+1}^N | x_n{=}\omega)$, where $y_1^n = (y_1, \ldots, y_n)$ and $y_{n+1}^N = (y_{n+1}, \ldots, y_N)$. Then, $p(x_n = \omega | y_1^N) \propto \alpha_n(\omega)\beta_n(\omega)$.

Among other extensions, the classic HMC has been generalized into the following directions.

1) The possibility of fast computation of $p(x_n|y_1^N)$ in HMC is due to the fact that $p(x_1^N|y_1^N)$ is a Markov distribution with computable transitions. The latter property remains valid once the couple $(x_1^N, y_1^N)$ is Markov, and thus, HMC can be extended. In fact, in HMC, both $X_1^N$ and $(X_1^N|y_1^N)$ are Markov, and thus, the Markovianity of $X_1^N$ is not necessary and can be relaxed. Doing so, we arrive at PMCs proposed in [13] and being able, as shown in [14], to noticeably improve the results obtained with HMCs;
2) When the transitions $p(x_n|x_{n-1})$ in the HMC of (2) are unknown and strongly vary with $n$, their estimation poses problem. It is then possible to manage it by extending the Markov chain $X_1^N$ to a so-called EMC, which gives a stationary model with possibilities of parameter estimation. As shown in [15], such an extension can improve results obtained with the classic HMC.
3) The theory of evidence has also been used to extend the Markov observation information. Indeed, when different

Manuscript received June 21, 2015; revised October 23, 2015 and January 6, 2016; accepted February 5, 2016. Date of publication March 18, 2016; date of current version December 22, 2016.

M. E. Y. Boudaren is with Ecole Militaire Polytechnique, Algiers 16111, Algeria (e-mail: boudaren@gmail.com).

W. Pieczynski is with SAMOVAR, Télécom SudParis, CNRS, Université Paris-Saclay, 91011 Evry Cedex, France (e-mail: wojciech.pieczynski@telecom-sudparis.eu).

Digital Object Identifier 10.1109/TFUZZ.2016.2543750

sources of information are available, even if the result of their fusion is no longer a Markov distribution, the computation of posterior margins $p(x_n|y_1^N)$ remains workable. Such a result has been applied in [16] and [17].

The aim of this paper is to propose a family of general models supporting all the extensions mentioned above, and being able to easily take into account new information, even partial (available only at some space/time points). Such models also generalize [18], [19] which proposed to integrate "partial knowledge" or "soft labels" on latent variables. For this purpose, the proposed formalism should support the fusion operator as an internal function to easily include new evidence when available. More precisely, the extension with respect to the previous evidential noisy models [15], [17] is the following. In previous models, one considers a sequence of masses $M_1^N = (M_1, \ldots, M_N)$, each $M_n$ taking thus its values in the power set $P(\Omega)$, a sequence of observations $Y_1^N = (Y_1, \ldots, Y_N)$, and the couple $(M_1^N, Y_1^N)$ is assumed Markov. Here, we introduce a finite set $\Lambda = \{\Lambda_1, \ldots, \Lambda_N\}$, an additional latent random chain $U_1^N = (U_1, \ldots, U_N)$, each $U_n$ taking its values in $\Lambda$, and the triplet $(M_1^N, U_1^N, Y_1^N)$ is assumed Markov. In both classic and new cases, the sequence of $Y_1^N$ can be continuous or discrete: one important thing is that the distribution of $M_1^N$—in the classic case—or the distribution of $(M_1^N, U_1^N)$—in the proposed extension—is Markov conditionally on observations. Thus, in the new case, the distribution of $U_1^N$ (conditional on observations) is a marginal distribution of the Markov distribution of $(M_1^N, U_1^N)$ (conditional on observations). For this reason, the new model is called the EMM model. As specified in the following, the interest of EMMs is that they are stable with respect to the DS fusion (see Proposition 4.1). More precisely, for two EMMs $(M^1, U^1, Y^1)$ and $(M^2, U^2, Y^2)$, there exists $U^3$ such that $(M^1 \oplus M^2, U^3, (Y^1, Y^2))$ is an EMM. In particular, this allows an easy integration of new pieces of information in the Markov context.

Let us also mention some works that dealt with theory of evidence within the framework of HMMs. In [16], DS fusion is used to fuse multisensor data in nonstationary Markovian context. Ramasso and Denoeux use belief functions to introduce partial knowledge about hidden states of an HMM [18]. In [20], the authors use evidential reasoning to relax Bayesian decisions given by a Markovian classification. The approach is applied to noisy images classification. In [21], a method is developed to prevent hazardous accidents due to operators' action slip in their use of a skill-assist. Other applications of evidential Markov models include data fusion and image classification [22], [23], power quality disturbance classification [24], particle filtering [25], prognostics [26], dynamical system analysis [27], and human action recognition [28]. Let us notice that theory of evidence has also been applied in the Markov random field context for image-related modeling problems [29]–[32]. Let us also underline the fact that the evidential Markov models considered in this paper are based on Markov masses, and thus, the associated computations strongly resemble to the classic computations in classic probabilistic HMCs. This is different from the model proposed in [28], where Markovianity is proposed through evidential conditioning of credibilities. However, both methods allow exact recursive computation of the final credibility (or, equivalently, the final *bba*) of interest.

Since the experiments considered at the end of this paper deal with nonstationary data modeling, it is worth pointing out that other Markov approaches have been successfully used to handle nonstationary or imprecise data. In [33], switching HMCs are introduced. For such models, the signal is considered stationary "per part," and an HMC model with a different set of parameters is devoted to each part. The same formalism is used in [34] to consider noise switches. Another family of models called "hidden semi-Markov models" [35] has also been used to model nonstationary data [36]. In [37], De Bock and De Cooman introduce an exact algorithm to estimate state sequences from outputs or observations in "imprecise" HMMs. The uncertainty linking one state to the next, and that linking a state to its observed output, is represented by a set of probability mass functions instead of a single such mass function. Other related works are presented in [38] and [39].

The paper is organized as follows. Section II summarizes PMCs and TMCs. In Section III, HEMCs are briefly recalled and commented. Section IV describes the proposed model. The application of such a model to unsupervised segmentation of nonstationary signals and the associated experiments are presented in Sections V and VI, respectively. Finally, a conclusion ends the paper.

## II. Pairwise and Triplet Markov Chains

The link between the classic HMCs and the HEMCs is made through the so-called TMCs. TMCs are extension of PMCs, which are themselves extensions of HMCs. This section is devoted to briefly recall what PMCs and TMCs are. Both models being probabilistic, there is no theory of evidence in this section.

### *A. Paiwise Markov Chains and Unsupervised Segmentation*

Let us consider two random sequences $X_1^N = (X_1, \ldots, X_N)$ and $Y_1^N = (Y_1, \ldots, Y_N)$ defined as above, and let $Z_1^N = (Z_1, \ldots, Z_N)$, with $Z_n = (X_n, Y_n)$ for each $n = 1, \ldots, N$. The couple $Z_1^N$ , which will also be denoted by $Z_1^N = (X_1^N, Y_1^N)$, is a classic HMC if its distribution is written:

$$p(x_1^N, y_1^N) = p(x_1)p(y_1|x_1) \prod_{n=2}^{N} p(x_n|x_{n-1})p(y_n|x_n). \quad (3)$$

The couple $Z_1^N$ is said to be a "PMC" if it is a Markov chain:

$$p(x_1^N, y_1^N) = p(x_1, y_1) \prod_{n=2}^{N} p(x_n, y_n|x_{n-1}, y_{n-1}) \quad (4)$$

which will also be equivalently written as

$$p(x_1^N, y_1^N) = p(z_1) \prod_{n=2}^{N} p(z_n|z_{n-1}). \quad (5)$$

As the transitions in (4) can be written as

$$\begin{aligned} &p(x_n, y_n|x_{n-1}, y_{n-1}) \\ &\quad = p(x_n|x_{n-1}, y_{n-1})p(y_n|x_n, x_{n-1}, y_{n-1}) \end{aligned} \quad (6)$$

we see that HMCs can be seen as particular PMCs in which $p(x_n|x_{n-1}, y_{n-1}) = p(x_n|x_{n-1})$ and $p(y_n|x_n, x_{n-1}, y_{n-1}) = p(y_n|x_n)$, which show locally how more general PMCs are with respect to HMCs. This greater generality can also be seen at the global level, through the following result. For stationary and invertible (which means that $p(z_2|z_1)$ and $p(z_1|z_2)$ are equal) PMCs, the chain $X_1^N$ is Markov if and only if for each $n = 1, 2, \ldots, N$, $p(y_n|x_1^N) = p(y_n|x_n)$: See the proof in [33]. Thus, the Markovianity of $X_1^N$ is not only useless, but it can even be inappropriate, as it imposes constraints on the noise distribution $p(y_1^N|x_1^N)$.

Then, forward probabilities denoted $\alpha_n(\omega) = p(y_1^n, x_n = \omega)$ and backward probabilities denoted $\beta_n(\omega) = p(y_{n+1}^N|x_n = \omega, y_n)$ are computable recursively, as in HMCs:

$$\begin{aligned}\alpha_1(x_1) &= p(x_1, y_1);\\ \alpha_{i+1}(x_{i+1}) &= \sum_{x_i \in \Omega} \alpha_i(x_i) p(x_{i+1}, y_{i+1}|x_i, y_i) \end{aligned}\tag{7}$$

$$\begin{aligned}\beta_1(x_N) &= 1;\\ \beta_i(x_i) &= \sum_{x_{i+1} \in \Omega} \beta_{i+1}(x_{i+1}) p(x_{i+1}, y_{i+1}|x_i, y_i).\end{aligned}\tag{8}$$

We have the following results, which extend those used in the classic HMCs:

$$p(x_{i+1}|x_i, y_1^N) = \frac{\beta_{i+1}(x_{i+1}) p(z_{i+1}|z_i)}{\beta_i(x_i)}\tag{9}$$

$$p(x_i|y_1^N) = \frac{\alpha_i(x_i)\beta_i(x_i)}{\sum_{x_i' \in \Omega} \alpha_i(x_i')\beta_i(x_i')}\tag{10}$$

$$p(x_i, x_{i+1}|y_1^N) = \frac{\alpha_i(x_i) p(z_{i+1}|z_i) \beta_{i+1}(x_{i+1})}{\sum_{x_i' \in \Omega} \alpha_i(x_i')\beta_i(x_i')}.\tag{11}$$

*Remark 2.1:* Let $Z_1^N = (Z_1, \ldots, Z_N)$ be a PMC. As $p(x_1, y_1) = p(x_1)p(y_1|x_1)$ and the transitions $p(z_n|z_{n-1})$ in (5) can be written as

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

(5) can be written as

$$p(x_1^N, y_1^N) = \underbrace{p(x_1) \prod_{n=2}^{N} \frac{p(x_{n-1}, x_n)}{p(x_{n-1})}}_{a(x_1^N)} \underbrace{p(y_1|x_1) \prod_{n=2}^{N} \frac{p(y_{n-1}, y_n|x_{n-1}, x_n)}{p(y_{n-1}|x_{n-1})}}_{b(x_1^N, y_1^N)}.\tag{12}$$

In $a(x_1^N)$, $p(x_{n-1}, x_n)$ is the distribution of $(X_{n-1}, X_n)$, $p(x_{n-1})$ is the distribution of $X_{n-1}$, but $a(x_1^N)$ is not necessarily the distribution of $X_1^N$. In fact, $a(x_1^N)$ is Markovian, and $p(x_1^N)$ is not necessarily Markovian. As a consequence, $b(x_1^N, y_1^N)$ is not necessarily the distribution of $p(y_1^N|x_1^N)$.

### *B. Parameter Estimation in the Gaussian Case*

Parameters estimation is identical in PMCs and TMCs, and thus, the method below is applicable to TMCs. More generally, it is applicable in some particular models among the new CEPMCs we propose. Hence, this paragraph is of importance for the unsupervised MPM segmentation with some of the proposed models. Let us consider a stationary PMC defined by $p(z_1, z_2) = p(x_1, y_1, x_2, y_2) = p(x_1, x_2)p(y_1, y_2|x_1, x_2)$. Let us consider the Gaussian case, which means that $p(y_1, y_2|x_1, x_2)$ are Gaussian. However, let us notice that, according to (12), $p(y_1^N|x_1^N)$ is not necessarily Gaussian. The problem we deal with is to estimate model parameters from $Y_1^N = y_1^N$. In the classic Gaussian HMC, the most known and used method is the EM algorithm [40], [41], and we will use in this paper its extension to PMC. For $K$ classes, we have, thus, to estimate $K^2$ parameters $p_{\rm ij} = p(x_1 = \omega_i, x_2 = \omega_j)$, and all parameters of $K^2$ Gaussian distributions $f_{\rm ij}(y_1, y_2) = p(y_1, y_2|x_1 = \omega_i, x_2 = \omega_j)$ in $R^2$ denoted with $N(\mu_{\rm ij}, \Gamma_{\rm ij})$. Denoting by $p_{\rm ij}^{(q)}$, $\mu_{\rm ij}^{(q)}$, and $\Gamma_{\rm ij}^{(q)}$ the current parameters, the next ones are given with (13)–(15), as shown at the bottom of the page, where $\psi_{n-1}^{(q)}(i, j) = p(x_{n-1} = \omega_i, x_n = \omega_j|y_1^N)$ is computed with (11) and the current parameters.

Finally, the EM runs as follows.

1) Find an initial value $\theta^{(0)}$ of the parameters.
2) Compute $\theta^{(q+1)}$ from $\theta^{(q)}$ and $y_1^N$ as follows:
   i) Step E: Use (7), (8), and (11) with the current parameters $\theta^{(q)}$ to compute $\psi_{n-1}^{(q)}(i, j)$.
   ii) Step M: Use (13)–(15) to compute $\theta^{(q+1)}$.

until an end criterion is reached.

$$p_{\rm ij}^{(q+1)} = \frac{1}{N-1} \sum_{n=1}^{N} \psi_{n-1}^{(q)}(i, j)\tag{13}$$

$$\mu_{\rm ij}^{(q+1)} = \frac{\sum_{n=2}^{N} (y_{n-1}, y_n)^t \psi_{n-1}^{(q)}(i, j)}{\sum_{n=2}^{N} \psi_{n-1}^{(q)}(i, j)}\tag{14}$$

$$\Gamma_{\rm ij}^{(q+1)} = \frac{\sum_{n=2}^{N} [(y_{n-1}, y_n)^t - \mu_{\rm ij}^{(q+1)}][(y_{n-1}, y_n)^t - \mu_{\rm ij}^{(q+1)}]^t \psi_{n-1}^{(q)}(i, j)}{\sum_{n=2}^{N} \psi_{n-1}^{(q)}(i, j)}\tag{15}$$

Let us notice that we use EM because it is well known that it is quite efficient in the Gaussian case considered in this paper. In a more complex situation, with non-Gaussian correlated noise, the "Iterative Conditional Estimation" method can be used, as proposed in [42].

### C. Triplet Markov Chains

Let us recall the TMC, which will be used to achieve DS fusion in the Markovian context. Let us consider three random sequences $X_1^N = (X_1, \ldots, X_N)$, $Y_1^N = (Y_1, \ldots, Y_N)$, and $U_1^N = (U_1, \ldots, U_N)$, taking their values in $\Omega = \{\omega_1, \ldots, \omega_K\}$, $R$, and $\Sigma = \{\sigma_1, \ldots, \sigma_M\}$, respectively. Considering the triplet $T_1^N = (U_1^N, X_1^N, Y_1^N)$, Markovian allows one to search $X_1^N = (X_1, \ldots, X_N)$ and $U_1^N = (U_1, \ldots, U_N)$. In fact, setting $V_1^N = (U_1^N, X_1^N)$, Markovianity of $T_1^N$ means that $T_1^N = (V_1^N, Y_1^N)$ is a PMC, and thus, $V_1^N$ can be searched from $Y_1^N$, as specified above. Let us notice that TMC is strictly more general than PMC; in fact, $(X_1^N, Y_1^N)$ is not necessarily Markov in a TMC $T_1^N = (U_1^N, X_1^N, Y_1^N)$. More precisely, none of the chains $(U_1^N, X_1^N)$, $(U_1^N, Y_1^N)$, $(X_1^N, Y_1^N)$, $U_1^N$, $X_1^N$, or $Y_1^N$ is necessarily Markov in a TMC. The auxiliary process $U_1^N$ may have many meaning according to the application [15], [33], [36], [43].

To summarize, we can say that in HMC, $p(x_1^N)$ is Markov, the noise distribution $p(y_1^N|x_1^N)$ is simple, $p(x_1^N, y_1^N)$ is Markov, and $p(x_1^N|y_1^N)$ is Markov, the latter property being important in MPM application. In PMC, $p(x_1^N)$ is not necessarily Markov, the noise distribution $p(y_1^N|x_1^N)$ is Markov and, thus, more complete than in HMC, $p(x_1^N, y_1^N)$ is Markov, and $p(x_1^N|y_1^N)$ is Markov, the latter property allowing the use of MPM in a similar way as in HMC. In TMC, none of the distributions $p(x_1^N)$, $p(y_1^N|x_1^N)$, $p(x_1^N, y_1^N)$, and $p(x_1^N|y_1^N)$ is necessarily Markov; however, as $p(u_n, x_n|y_1^N)$ are computable with complexity linear in $N$, $p(x_n|y_1^N) = \sum_{u_n} p(u_n, x_n|y_1^N)$ also are, and thus, the use of MPM is possible, with comparable computational complexity as in PMC and HMC [17].

## III. Hidden Evidential Markov Chains

### A. Theory of Evidence

Let us briefly recall some basic notions of the theory of evidence [44]–[47]. Let $\Omega = \{\omega_1, \ldots, \omega_K\}$, and let $P(\Omega) = \{A_1, \ldots, A_Q\}$ be its power set, with $Q = 2^K$. A function $M$ from $P(\Omega)$ to $[0, 1]$ is called a *bba* if $M(\emptyset) = 0$ and $\sum_{A \in P(\Omega)} M(A) = 1$. A *bba* $M$ defines, then, a "plausibility" function $Pl$ from $P(\Omega)$ to $[0, 1]$ by $Pl(A) = \sum_{A \cap B \neq \emptyset} M(B)$, and a "credibility" function $Cr$ from $P(\Omega)$ to $[0, 1]$ by $Cr(A) = \sum_{B \subset A} M(B)$. For a given *bba* $M$, the corresponding plausibility function $Pl$ and credibility function $Cr$ are linked by $Pl(A) + Cr(A^c) = 1$ so that each of them defines the other. Conversely, $Pl$ and $Cr$ can be defined by some axioms, and each of them defines then a unique corresponding *bba* $M$. More precisely, $Cr$ is a function from $P(\Omega)$ to $[0, 1]$ verifying $Cr(\emptyset) = 0$, $Cr(\Omega) = 1$, and $Cr(\bigcup_{j \in J} A_j) \geq \sum_{\emptyset \neq I \subset J} (-1)^{|I|+1} Cr(\bigcap_{j \in I} A_j)$, and $Pl$ is a function from $P(\Omega)$ to $[0, 1]$ verifying analogous conditions, with $\leq$ instead of $\geq$ in the third one. A credibility function $Cr$ verifying such conditions also is the credibility function defined by the *bba* $M(A) = \sum_{B \subset A} (-1)^{|A-B|} Cr(B)$.

Finally, each of the three functions $M$, $Pl$, and $Cr$ can be defined in an axiomatic way, and each of them defines the two others. Furthermore, a probability function $p$ can be seen as a particular case in which $Pl = Cr = p$.

When two *bba*s $M_1$ and $M_2$ represent two pieces of evidence, we can combine—or fuse—them using the so-called DS fusion, which gives $M = M_1 \oplus M_2$ defined by

$$M(A) = (M_1 \oplus M_2)(A) \propto \sum_{B_1 \cap B_2 = A} M_1(B_1) M_2(B_2). \tag{16}$$

We will say that a *bba* is "Bayesian" or "probabilistic" when, being null outside singletons, it defines a probability, and we will say that it is "evidential" otherwise. One can then see that when either $M_1$ or $M_2$ is probabilistic, the fusion result $M$ is also probabilistic.

### B. Dempster–Shafer Fusion and Posterior Distribution

The main link between classical Bayesian computations and the theory of evidence is that the computation of the posterior distribution can be seen as a DS fusion of two probabilities. Thus, extending the latter to belief functions, one extends the posterior probabilities, and thus, one extends the frames of Bayesian computation. This is the crux point, and the novelty of this paper is to propose using it in a wide Markov context, extending the different recent results.

Let us consider some examples showing the interest of extending classic posterior probabilities calculation to DS fusion. As a first step, we limit the frame to a simple context without Markovianity, but we will show in the next section that each of the examples below can be extended to the general Markov context introduced in this paper. Let us mention that similar examples may be found in [48] and [49]. In addition, it should be noticed that *Example* 3.1 uses the so-called linear-vacuous mixture, while *Example* 3.2 uses what is commonly known as "inner-measure" [48], [49].

Let us also mention that since Zadeh's criticism [50], many research works have discussed the consistency of theory of evidence [51], some of which concluded that DS theory remains quite limited when modeling probability sets or unknown probabilities [48], [49]. Furthermore, while Dempster's rule generalizes Bayes rule, it does not comply with a subjective interpretation of probability [52]. In addition, let us point out that while this paper employs theory of evidence, imprecise probabilities can also deal easily with such situations [52].

*Example 3.1:* Let $\Omega = \{\omega_1, \ldots, \omega_K\}$, and let us suppose that our knowledge about the distribution $p(x)$ is $p_1 = p(x = \omega_1) \geq \varepsilon_1, \ldots, p_K = p(x = \omega_K) \geq \varepsilon_K$ with $\varepsilon = \varepsilon_1 + \cdots + \varepsilon_K \leq 1$. We see that $\varepsilon$ measures the degree of knowledge of $p(x)$ in a "continuous" manner: For $\varepsilon = 1$, the distribution $p(x)$ is perfectly known, and for $\varepsilon = 0$, nothing is known about $p(x)$. Assume that $p(y|x = \omega_1), \ldots, p(y|x = \omega_K)$ are known, and let us consider the distribution $q^y = (q_1^y, \ldots, q_K^y)$

defined with

$$q_1^y = \frac{p(y|x=\omega_1)}{\sum_{i=1}^K p(y|x=\omega_i)}, \ldots, q_K^y = \frac{p(y|x=\omega_K)}{\sum_{i=1}^K p(y|x=\omega_i)}.$$

Using Bayesian classification to estimate $X = x$ from $Y = y$ requires the knowledge of $p(x|y) \propto p(x)p(y|x)$, which is thus only partly known. How could one use this partial knowledge to perform Bayesian classification? This is made possible by introducing the following *bba* $A$ on $P(\Omega)$: $A$ is null outside $\{\{\omega_1\}, \ldots, \{\omega_K\}, \Omega\}$ and $A[\{\omega_1\}] = \varepsilon_1, \ldots, A[\{\omega_K\}] = \varepsilon_K$, $A[\Omega] = 1 - (\varepsilon_1 + \cdots + \varepsilon_K) = 1 - \varepsilon$. The DS fusion of $A$ with $q^y = (q_1^y, \ldots, q_K^y)$ gives a probability $p^*$ defined on $\Omega$ by

$$p^*(\omega_i) = \frac{(\varepsilon_i + 1 - \varepsilon)q_i^y}{\sum_{j=1}^K (\varepsilon_j + 1 - \varepsilon)q_j^y}.$$

Then, using $p^*$ to perform the classification allows one to use the partial knowledge of $p(x)$ in a "continuous" manner: Perfect knowledge of $p(x)$ corresponds to $\varepsilon = 1$, and indeed, when $\varepsilon = 1$, we have $p^*(x) = p(x|y)$. The case $\varepsilon = 0$ corresponds to the case where $p(x)$ is not known at all, and indeed, this case implies $p^*(x) = q^y(x)$, and the corresponding classification rule is the maximum likelihood classification.

*Example 3.2:* Let us consider the following example [44]. Let $\Omega = \{\omega_1, \ldots, \omega_K\}$ with the distribution $p(x)$ known. There is a partition of $\Omega$ into $L \le K$ subsets $\Omega_1$, ... , $\Omega_L$ such that the elements of each subset $\Omega_i$ produce a same $q_i^y$. For example, in image processing, we can have three classes "river," "field," "houses," and an infrared sensor. As such sensor measures the temperature, it cannot make a difference between "field" and "river." Hence, there are only two subsets $\{\text{river}, \text{field}\}$ and $\{\text{houses}\}$. Then, we can consider a *bba* $B$ null outside $\{\Omega_1, \ldots, \Omega_L\}$ and defined for each $\Omega_i$ by

$$B^y(\Omega_i) = \frac{p(y|x=\omega_i)}{\sum_{j=1}^L p(y|x=\omega_j)}$$

where $\omega_1, \ldots, \omega_L$ are arbitrary elements such that $\omega_1 \in \Omega_1, \ldots, \omega_L \in \Omega_L$. The DS fusion of $p(x)$ with $B^y$ gives a probability $p^*$ defined on $\Omega$ by

$$p^*(\omega_i) = \frac{p(\omega_i)B^y(\Omega_i)}{\sum_{(\omega_j,\Omega_j)|\omega_j \in \Omega_j} p(\omega_j)B^y(\Omega_j)}.$$

As in *Example 3.1*, when $B^y$ is itself $q^y = (q_1^y, \ldots, q_K^y)$ from *Example 3.1*, $p^*$ is the classic posterior distribution.

*Example 3.3:* As in Example 3.2, let $\Omega = \{\omega_1, \ldots, \omega_K\}$ with the distribution $p(x)$ known. Assume that $p(y|x = \omega_1), \ldots, p(y|x = \omega_K)$ are known; however, there exists an additional class $\omega_{K+1}$ which "hides" the classes of interest forming $\Omega$, and which produces $p(y|x = \omega_{K+1})$. For example, in optical satellite image processing, there may be three classes of interest "forest," "water," and "houses," while a fourth class "clouds" can be of no interest. According to Yager's rule, one can transfer the mass associated with such an "additional" class to $\Omega$. Then, one can consider $B^y$ is null outside $\{\{\omega_1\}, \ldots, \{\omega_K\}, \Omega\}$ and

$$B^y[\{\omega_i\}] = \frac{p(y|x=\omega_i)}{\sum_{j=1}^{K+1} p(y|x=\omega_j)}, \text{for } i \in \{1, \ldots, K\} \text{ and}$$

$$B^y[\Omega] = \frac{p(y|x=\omega_{K+1})}{\sum_{j=1}^{K+1} p(y|x=\omega_j)}.$$

The DS fusion of $p(x)$ with $B^y$ gives a probability $p^*$ defined on $\Omega$ by

$$p^*(\omega_i) = \frac{p(\omega_i)[B^y(\{\omega_i\}) + B^y(\Omega)]}{\sum_{\omega_j \in \Omega} p(\omega_j)[B^y(\{\omega_j\}) + B^y(\Omega)]}.$$

This fusion is mathematically similar to that used in *Example 3.1*; however, it models a quite different situation. Such models have been successfully used in cloudy images segmentation in [32].

It is worth pointing out that the situation of the presence of an additional class, considered here from Yager's rule viewpoint, can be managed differently. In the transferable belief model [45], for instance, the associated mass is transferred to the empty set. For conflict management in general, see [53].

*Example 3.4:* Let $\Omega = \{\omega_1, \ldots, \omega_K\}$, and let us suppose that the distributions $p(x)$, $p(y|x = \omega_1), \ldots, p(y|x = \omega_K)$ are known. Let us assume, at a first step, that the knowledge of $p(x)$ is poor. This fact can be taken into account through the *bba* $A$ defined on $\{\{\omega_1\}, \ldots, \{\omega_K\}, \Omega\}$ by $A(\{\omega_i\}) = \varepsilon p(\omega_i)$ for $i = 1, \ldots, K$ and $A(\Omega) = 1 - \varepsilon$. Considering the distribution $q^y = (q_1^y, \ldots, q_K^y)$, where $q_i^y \propto p(x = \omega_i)$ (like in *Example 3.1*), the DS fusion of $A$ with $q^y = (q_1^y, \ldots, q_K^y)$ gives a probability $p^*$ defined on $\Omega$ by $p^*(\omega_i) \propto (A(\omega_i) + A(\Omega))q_i^y$. Let us now assume that the unreliability is related to the distributions $p(y|x = \omega_1), \ldots, p(y|x = \omega_K)$ as for the infrared sensor of *Example 3.2*. Hence, there is a partition of $\Omega$ into $L \le K$ subsets $\Omega_1, \ldots, \Omega_L$ such that the elements of each subset $\Omega_i$ produce a same $q_i^y$. Then, we can consider a *bba* $B$ null outside $\{\Omega_1, \ldots, \Omega_L\}$ and defined for each $\Omega_i$ by $B^y(\Omega_i) \propto p(y|x = \omega_i)$, where $\omega_i$ is an arbitrary element verifying $\omega_i \in \Omega_i$. If $p(x)$ is perfectly known, the DS fusion of $p(x)$ with $B^y$ gives a probability $p^*$ defined on $\Omega$ by $p^*(\omega_i) \propto p(\omega_i)B^y(\Omega_i)$. Finally, if $p(x)$ and $p(y|x = \omega_1), \ldots, p(y|x = \omega_K)$ are all unreliable, the DS fusion $A \oplus B^y$ is a *bba* defined on $\{\{\omega_1\}, \ldots, \{\omega_K\}, \Omega_1, \ldots, \Omega_L\}$ by $[A \oplus B^y](\{\omega_i\}) \propto A(\{\omega_i\})B^y(\Omega_j)1_{\omega_i \in \Omega_j}$ and, $[A \oplus B^y](\Omega_j) \propto A(\Omega)B^y(\Omega_j)$.

*Example 3.5:* Examples 3.1 and 3.2 can be blended; for example, let us consider a situation where the distribution $p(x)$ is not perfectly known, which is modeled with a *bba* $A$ on $\{\{\omega_1\}, \ldots, \{\omega_K\}, \Omega\}$ as specified in example 1, and $Y = y$ defines a *bba* $B^y$ on $\{\Omega_1, \ldots, \Omega_L\}$, as specified in example 2. Then, the DS fusion $A \oplus B^y$ is then a *bba* on $\{\{\omega_1\}, \ldots, \{\omega_K\}, \Omega_1, \ldots, \Omega_L\}$ defined by

$$[A \oplus B^y](\{\omega_i\}) = \frac{A(\{\omega_i\})B^y(\Omega_j)1_{\omega_i \in \Omega_j}}{c}$$

$$[A \oplus B^y](\Omega_j) = \frac{A(\Omega)B^y(\Omega_j)}{c}$$

with the normalizing constant

$$c = \sum_{(\omega_i, \Omega_j)|\omega_i \in \Omega_j} A(\{\omega_i\}) B^y(\Omega_j) + \sum_{j=1}^{L} A(\Omega) B^y(\Omega_j).$$

*Remark 3.1:* Let us notice that when DS fusion produces an evidential mass, it can still be used to perform a classification by means of "maximum of plausibility" by choosing the element in $\Omega$ whose plausibility is maximal.

### C. Hidden Evidential Markov Chains

Let us consider the classic HMC with the distribution given with (2).

Let

$$p_1(x_1^N) = p(x_1) \prod_{n=2}^{N} p(x_n|x_{n-1}) \tag{17}$$

and let

$$p^{y_1^N}(x_1^N) = \frac{\prod_{n=1}^{N} p(y_n|x_n)}{\prod_{n=1}^{N} \sum_{x_n} p(y_n|x_n)}. \tag{18}$$

Then, the posterior distribution $p(x_1^N|y_1^N)$ can be seen as the DS fusion of $p_1$ and $p^{y_1^N}$:

$$p(x_1^N|y_1^N) = (p_1 \oplus p^{y_1^N})(x_1^N). \tag{19}$$

That is of importance as it opens way to different extensions [17], [54]. More precisely, if either $p_1$ or $p^{y_1^N}$ is extended to an evidential mass, the fusion (19) remains a probability distribution, which can then be seen as an extension of the classic posterior probability $p(x_1^N|y_1^N)$. In addition, if the "evidential" extension of $p_1$ or $p^{y_1^N}$ is of a similar Markovian form, in spite of the fact that the fusion result is no longer a Markov distribution, the computation of posterior margins $p(x_n|y_1^N)$ remains feasible. In fact, the core point is to remark that the fusion (16) can be interpreted as the computation of some marginal distribution, which leads, in the Markov context we deal with in this paper, to consider the so-called TMCs [17] recalled in Section II. For example, if $p_1$ is extended to a *bba* of the form

$$M(A_1, \ldots, A_N) = M(A_1)M(A_2|A_1)\ldots M(A_N|A_{N-1}) \tag{20}$$

where for each $i = 2, \ldots, N$ and $A_i \in P(\Omega)$, $M(.|A_i)$ is a *bba* on $P(\Omega)$, then $M \oplus p^{y_1^N}$ is a workable distribution, in spite of being not Markovian.

*Remark 3.2:* Let us recall that (20) is not based on any considerations about evidential conditioning, as it is made in [28]. In this paper, we deal with *bba*s in a similar way as if they were probabilities, and this results in similar recursive formulas that compute different quantities of interest. However, at each step, *bba*s define credibilities and plausibilities, and thus, our model is not probabilistic but evidential.

The interest of such extension has been shown in hidden nonstationary Markov chain segmentation [15]: When the Markovian $p_1$ is nonstationary, it is to say when $p_1(x_n, x_{n+1})$ varies with $n$, replacing $p_1$ with stationary $M$ of (20) form gives better results in unsupervised segmentation than replacing it with any other stationary Markov distribution. This is of interest in unsupervised context, where all parameters have to be estimated from $y_1^N$. Using some estimation method like EM, leads, when keeping the classic model given by (2), to a stationary $\hat{p}_1$. When using a stationary extension (20), EM can still be used, and the estimated *bba*-based segmentation provides better results. Such a model is called HEMC.

## IV. Evidential Pairwise Markov Chains

In this section, we introduce two original models, called EPMC, and CEPMC. The latter simultaneously extends probabilistic HMCs, PMCs, and TMCs recalled in Sections I and II, conditionally Markov models [55], [56], and hidden or pairwise EMCs recently published [15], [57]. Let us notice that considering evidential sequences, no temporally independent opens very rich perspectives, and models we consider are somewhat particular. For a frame of discernment $\Omega = \{\omega_1, \ldots, \omega_K\}$ and a sequence of length $N$ to deal with, the most general case would be to consider the *bba*s defined on power set $P[\Omega^N]$ of $\Omega^N$. We limit our investigations to the *bba*s defined on $(P[\Omega])^N$, which is a subset of $P[\Omega^N]$. Then, every $A_1 \times \cdots \times A_N \in (P[\Omega])^N$ is assimilated to a sequence $(A_1, \ldots, A_N)$, and "Markov" *bba*s are defined in a similar way as the classic probabilistic Markov chains are, which results in "EMCs" [15], [17]. Such models are particular ones and other possibilities of defining *bbas* on sequences, using some kind of Markovianity stemming from evidential conditioning, exist [19], [27], [28]. Thus, here, we consider these particular models, and we extend them to new families, whose originality and interest lie in the introduction of an auxiliary set $\Lambda = \{\lambda_1, \ldots, \lambda_J\}$, which allows one, roughly speaking, to keep Markovian form of considered distributions after DS fusion. In other words, the main idea is to consider the *bba*s of interest as marginal distribution of a Markov chain rather than a Markov chain. Then, we show how the examples of Section III can be extended, using the proposed model, to take the spatial information into account.

*Definition 4.1:* Let us consider the following:

1) $\Omega = \{\omega_1, \ldots, \omega_K\}$ a set of classes (frame of discernment), $P(\Omega) = \{A_1, \ldots, A_Q\}$ its power set, and $\Lambda = \{\lambda_1, \ldots, \lambda_J\}$ a finite set.
2) $V_1^N = (M_1^N, U_1^N)$ a random chain, each $(M_n, U_n)$ taking its values in $P(\Omega) \times \Lambda$.

Let $I_{V_n} \subset P(\Omega) \times \Lambda$ be the image set common for all $V_n = (M_n, U_n)$, $n = 1, \ldots, N$.

Then, $V_1^N$ is called EPMC if there exist $q_1, \ldots, q_{N-1}$ functions from $I_{V_n} \times I_{V_n}$ to $R^+$ such that its distribution verifies

$$p(v_1^N) \propto q_1(v_1, v_2) q_2(v_2, v_3) \ldots q_{N-1}(v_{N-1}, v_N). \tag{21}$$

*Definition 4.2:* Let us consider the context of Definition 4.1. Let $Y_1^N = (Y_1, \ldots, Y_N)$ be a random chain, each $Y_n$ taking its values in $R^d$. $V_1^N$ is called CEPMC if its distribution conditional on $Y_1^N = y_1^N$ is an EPMC.

Let us notice that the meaning of the word "pairwise Markov" is somewhat different here from the one used in Section II. Indeed, here, we consider a pair in which one sequence is evidential and the other is probabilistic latent, while in

Section II, we consider a pair in which one is probabilistic hidden and the other is probabilistic observed. Anyway, in both cases, the related couples are Markov, and it will result from the context in which case one is.

*Remark 4.1:* It will be of importance to notice that the support $I_{V_n} \subset P(\Omega) \times \Lambda$ of the law of $V_1^N = (M_1^N, U_1^N)$ is not, in general, the whole $P(\Omega) \times \Lambda$, but only a part of it. It will be convenient to consider $I_{V_n}$ as being defined by:

1) $I_{M_n}$ the set of *bba*s such that there exists $\lambda_j \in \Lambda$ such that $(m_n, u_n = \lambda_j) \in I_{V_n}$ (the image set of $M_n$);
2) the function that associates with each $m_n \in I_{M_n}$ the set $\Lambda(m_n)$ of elements $\lambda_j$ in $\Lambda$ such that $(m_n, u_n = \lambda_j) \in I_{V_n}$.

*Definition 4.3:* Let $V_1^N = (M_1^N, U_1^N)$ be an EPMC (a CEPMC, respectively). The random chain $M_1^N$ will be said EMMC (a CEMMC, respectively).

We see how EMMCs and EPMCs (CEMMCs and CEPMCs, respectively) are linked to each other: $M_1^N$ is an EMMC if there exists $U_1^N$ such that $V_1^N = (M_1^N, U_1^N)$ is an EPMC ( $M_1^N$ is a CEMMC if there exists $U_1^N$ such that $V_1^N = (M_1^N, U_1^N)$ is a CEPMC, respectively).

The interest of formulation (21) rather than formulation (20) will appear in Proposition 4.1. In fact, expressing an EPMC through such a formulation makes it possible to prove that EPMCs are stable with respect to DS fusion; it is to say that the DS fusion result of two EPMCs is itself an EPMC. On the other hand, (21) is equivalent to Markovianity of $V_1^N = (M_1^N, U_1^N)$, and this Markovianity allows the computation of $p(m_n)$ (which is $p(m_n|y_1^N)$ in the CEMMC case). Indeed, if $V_1^N$ is Markov, one can take $q_1(v_1, v_2) = p(v_1, v_2), q_2(v_2, v_3) = p(v_3|v_2), \ldots, q_{N-1}(v_{N-1}, v_N) = p(v_N|v_{N-1})$. Conversely, (21) implies Markovianity of $V_1^N$ with transitions $p(v_n|v_{n-1})$ and $p(v_1)$ computable by the classical backward recursion. More precisely, setting

$$\begin{aligned} f_N(v_N) &= 1; \\ f_{n-1}(v_{n-1}) &= \sum_{v_n} q_{n-1}(v_{n-1}, v_n) f_n(v_n), \\ \text{for} &= N, \ldots, 2. \end{aligned} \tag{22}$$

We have

$$\begin{aligned} p(v_n|v_{n-1}) &= \frac{q_{N-1}(v_{n-1}, v_n) f_n(v_n)}{f_{n-1}(v_{n-1})}, \quad \text{for } n = N, \ldots, 2 \\ p(v_1) &= \frac{q_1(v_1, v_2) f_2(v_2)}{f_1(v_1)}. \end{aligned} \tag{23}$$

Having $p(v_1)$ and $p(v_n|v_{n-1})$ for $n = 2, \ldots, N$, we compute $p(v_n)$ for each $n = 2, \ldots, N$, by the classical forward recursion

$$\begin{aligned} &p(v_1) \text{ given };\\ &p(v_{n+1}) = \sum_{v_n} p(v_n) p(v_{n+1}|v_n), \quad \text{for } n = 1, \ldots, N-1. \end{aligned} \tag{24}$$

Finally, having $p(v_n) = p(m_n, u_n)$ gives $p(m_n)$ for each $n = 1, \ldots, N$:

$$p(m_n) = \sum_{u_n} p(v_n) = \sum_{u_n} p(m_n, u_n). \tag{25}$$

Let us now consider the CEMMC case: $p(m_n)$ calculated above is $p(m_n|y_1^N)$. There are two possibilities: it is a probability or a *bba*. In the first case, the classic Bayesian MPM method can be used to estimate the hidden class. In the second case, one can compute the plausibility $Pl(\omega_i) = \sum_{m_n|\omega_i \in m_n} p(m_n|y_1^N)$ and estimate the hidden class by the "maximum of plausibility." Finally, one important thing is that in the general CEPMC, the hidden classes can be searched by a method extending the classic MPM method, which is merely as simple as the latter.

*Remark 4.2:* It can be shown that the complexity of a CEPMC $V_1^N = (M_1^N, U_1^N)$ is equivalent to the complexity of a TMC $T_1^N = (U'^N_1, M'^N_1, Y_1^N)$ with $Card[\mathbf{D}(U'^N_1)] = Card[\mathbf{D}(U_1^N)]$ and $Card[\mathbf{D}(M'^N_1)] = Card[\mathbf{D}(M_1^N)]$, where $\mathbf{D}(A)$ denotes the domain of sequence $A$. For more details, see [17].

We provide different examples, showing how the EPMCs family includes different models proposed so far. Before, let us specify how the DS fusion is performed inside EPMCs family. We can state the following result (the proof is provided in the Appendix).

*Proposition 4.1:* Let $V^1$ and $V^2$ be two EPMCs defined on $I_{V_n^1} \subset P(\Omega) \times \Lambda^1$ and $I_{V_n^2} \subset P(\Omega) \times \Lambda^2$ with distributions given by (21), with $(q_1^1, \ldots, q_{N-1}^1)$ and $(q_1^2, \ldots, q_{N-1}^2)$, respectively. Then, the DS fusion $V = V^1 \oplus V^2$ is an EPMC defined on $I_{V_n} \subset P(\Omega) \times \Lambda^3$, with:

1) $\Lambda^3 = P(\Omega) \times P(\Omega) \times \Lambda^1 \times \Lambda^2$, $I_{V_n}$ defined by $I_{M_n} = \{A \in P(\Omega) | A = A^1 \cap A^2, A^1 \in I_{M_n^1}, A^2 \in I_{M_n^2}\}$ and $\Lambda^3(m_n) = \{(A^1, A^2, \lambda^1, \lambda^2) \in \Lambda^3 | A^1 \in I_{M_n^1}, A^2 \in I_{M_n^2}, \lambda^1 \in \Lambda^1(A^1), \lambda^2 \in \Lambda^2(A^2)\}$;
2) $(q_1^3, \ldots, q_{N-1}^3)$, where for each $n = 1, \ldots, N-1$, $q_n^3$ is defined on $I_{V_n} \times I_{V_n}$ by

$$\begin{aligned} &q_n^3(A_n^3, u_n^3, A_{n+1}^3, u_{n+1}^3) \\ &\quad = q_n^3(A_n^3, (u_n^1, u_n^2, A_n^1, A_n^2), A_{n+1}^3, \\ &\quad (u_{n+1}^1, u_{n+1}^2, A_{n+1}^1, A_{n+1}^2)) \\ &\quad = q_n^1(A_n^1, u_n^1, A_{n+1}^1, u_{n+1}^1) q_n^2(A_n^2, u_n^2, A_{n+1}^2, u_{n+1}^2). \end{aligned} \tag{26}$$

In practice, the DS fusion is performed as follows:

1) one searches for $I_{M_n}$ and $\Lambda^3(m_n)$, which gives $I_{V_n}$;
2) one computes $q_n^3(A_n^3, u_n^3, A_{n+1}^3, u_{n+1}^3)$ on $I_{V_n}$ with (26).

*Remark 4.3:* As stated above, the DS fusion is achieved by simple multiplication. While such a computation is not heavy from computational point of view, the complexity of the EPMC resulting from this fusion is higher than the one of each model involved in such fusion. Furthermore, when there are many EPMCs involved in the fusion, the model complexity increases quickly at each application of DS operator. This is due to the quick increase of the cardinality of the domain $\Lambda^3$ as specified in *Proposition* 4.1.

We have considered until now that there was an unique set $\Lambda$ for all $n = 1, \dots, N$; let us now assume that it can depend on $n$. Thus, the sets $I_{V_n}$ vary with $n$. This allows one to take into account additional "local" information. More precisely, let us consider a CEPMC with distribution defined by $q_1(v_1, v_2, y_1^N), \dots, q_{N-1}(v_{N-1}, v_N, y_1^N)$. Let us imagine that an additional information on the hidden class $\omega \in \Omega$ at point $n$, for example, provided by a new observation, has arrived and this new piece of information is modeled by a *bba* $M_n^{\text{New}}$. According to Proposition 4.1, we set $I_{M_n}^* = \{A \in P(\Omega) | A_n = A_n^1 \cap A_n^2, \text{ with } A_n^1 \in I_{M_n}, A_n^2 \in I_{M_n^{New}}\}$. Then, $q_{n-1}(v_{n-1}, v_n, y_1^N) = q_{n-1}(v_{n-1}, (A_n, \lambda_n), y_1^N)$ (or $q_n(v_n, v_{n+1}, y_1^N)$) is modified by setting $q_{n-1}^*(v_{n-1}, v_n^*, y_1^N) = q_{n-1}^* (v_{n-1}, (A_n, \lambda_n^*), y_1^N) = q_{n-1}(v_{n-1}, (A_n^1, \lambda_n), y_1^N) M_n(A_n^2)$, where $\lambda_n^*$ varies in $\Lambda_n^* \subset P(\Omega) \times P(\Omega) \times \Lambda_n$ defined by $(A_1, A_2, \lambda_n) \in \Lambda_n^*$ iff $A_n^1 \cap A_n^2 \in I_{M_n}^*$. Thus, this DS fusion modifies $\Lambda_n$, which becomes $\Lambda_n^*$.

Let us briefly summarize how CEPMCs extend different known models.

1) Let $V_1^N = (M_1^N, U_1^N)$ be a CEPMC defined on $(I_{V_1}, \dots, I_{V_N}) \subset [P(\Omega) \times \Lambda_1] \times \dots \times [P(\Omega) \times \Lambda_N]$, with $Y_1^N = y_1^N$ observed random sequence. Thus, its distribution is defined by $q_1(v_1, v_2, y_1^N), \dots, q_{N-1}(v_{N-1}, v_N, y_1^N)$. When $\Lambda_1, \dots, \Lambda_N$ are all reduced to singletons, all $P(\Omega) \times \Lambda_n$ can be seen as being reduced to $P(\Omega)$. When, in addition, $V_1^N = M_1^N$ so obtained is probabilistic, we find again the famous "discriminative random fields" (DRF) [58], [59]. When $V_1^N = M_1^N$ so obtained is not probabilistic, we obtain an original "evidential" extension of DRFs.
2) Let us return to the general case. If $q_1(v_1, v_2, y_1^N) = f_1(v_1, v_2) g_1(v_1, y_1)\ h_1\ (v_2, y_2), \dots, q_{N-1}(v_{N-1}, v_N, y_1^N) = f_{N-1}(v_{N-1}, v_N)\ g_{N-1}(v_{N-1}, y_{N-1})\ h_{N-1}(v_n, y_n)$ and $\Lambda_1, \dots, \Lambda_N$ are all reduced to singletons, one finds again some evidential models including the HEMCs [15]. When, in addition, $\Lambda_1, \dots, \Lambda_N$ are not all reduced to a singleton, we obtain original "triplet" EMCs with "independent noise." This last model becomes the classic probabilistic TMC "with independent" noise (TMC-IN) if, in addition, $M_1^N$ is probabilistic. Let us recall that such a TMC-IN is an extension of both "hidden semi-Markov chains" [35] and "hidden bivariate Markov chains" [60].
3) Let us return to the general case. If $q_1(v_1, v_2, y_1^N) = f_1(v_1, v_2, y_1, y_2), \dots, q_{N-1}(v_{N-1}, v_N, y_1^N) = f_{N-1} (v_{N-1}, v_N, y_{N-1}, y_N)$ and $\Lambda_1, \dots, \Lambda_N$ are all reduced to singletons, we find again the pairwise EMC [17], [57]. If $\Lambda_1, \dots, \Lambda_N$ are not all reduced to singletons, we obtain an original "triplet" EMC that extends the same model with "independent noise" of the previous point. On the other hand, it also extends the pairwise evidential chain introduced in [17] and [57].

We see that there are three kinds of factors influencing the degree of generality of different models. The kind of dependence of $q_n(v_n, v_{n+1}, y_1^N)$ on $y_1^N$ gives "independent noise," "pairwise" models, or "DRF" models. Probabilistic or evidential nature of $M_1^N$ gives "probabilistic" or "evidential" models. Finally, $\Lambda_1, \dots, \Lambda_N$ can all be reduced to singletons or not. These factors can be blended, which results in numerous possibilities some of which giving again known models, while some others providing their original extensions.

Let us resume the examples of Section III, to see what is modified when using the Markov context proposed in *Proposition 4.1*.

*Example 4.1:* Let us consider *Example 3.1*. Thus, $p(x)$ on $\Omega$ is replaced by a Markov distribution $p(x_1^N)$ on $\Omega^N$. In particular, the latter distribution can be considered as given by $p(x_1^N) \propto q_1(x_1, x_2) \dots q_{N-1}(x_{N-1}, x_N)$, with the functions $q_1(x_1, x_2) = p(x_1, x_2), q_2(x_2, x_3) = p(x_3|x_2), \dots, q_{N-1}(x_{N-1}, x_N) = p(x_N|x_{N-1})$. Let us assume that $p_1^{\text{ij}} = p(x_1 = \omega_i, x_2 = \omega_j) \geq \varepsilon_1^{\text{ij}}$, with $\sum_{1 \leq i,j \leq K} \varepsilon_1^{\text{ij}} = \varepsilon_1 \leq 1$ and for each $n = 2, \dots, N-1$, $p_n^{\text{ij}} = p(x_{n+1} = \omega_j | x_n = \omega_i) \geq \varepsilon_n^{\text{ij}}$, with $\sum_{1 \leq j \leq K} \varepsilon_n^{\text{ij}} = \varepsilon_n^i \leq 1$. This poor knowledge of $p(x_1^N)$ can then be modeled by $q_n^1(v_n, v_{n+1})$, with $(v_n, v_{n+1}) \in \{\{\omega_1\}, \dots, \{\omega_K\}, \Omega\}^2$, defined with $q_n^1(v_n, v_{n+1}) = \varepsilon_n^{\text{ij}}$ if $(v_n, v_{n+1}) = (\{\omega_i\}, \{\omega_j\})$, and $q_n^1(v_n, v_{n+1})$ defined on $\{\{\omega_1\}, \dots, \{\omega_K\}, \Omega\}^2 - \Omega^2$ in some way with the constraints $\sum_{v_1, v_2} q_1^1(v_1, v_2) = 1$ and for $n = 2, \dots, N-1$, $\sum_{v_{n+1}} q_n^1(v_n, v_{n+1}) = 1$. Let us then assume, as in *Example 3.1*, that $p(y_n|x_n)$ are perfectly known for each $n = 1, \dots, N$ and $x_n \in \Omega$. Let $q^{y_1, \dots, y_N}(x_1, \dots, x_N) \propto p(y_1|x_1) \dots p(y_N|x_N)$ be the corresponding probability on $\Omega^N$. This probability can be considered as a particular Markov chain defined by $q_1^2(x_1, x_2) = p(y_1|x_1)p(y_2|x_2)$, $q_2^2(x_2, x_3) = p(y_3|x_3), \dots, q_{N-1}^2(x_{N-1}, x_N) = p(y_N|x_N)$. Thus, we have an EMC given with $q_1^1, \dots, q_{N-1}^1$ (which is an EPMC with no $\Lambda^1$), and a probabilistic "Markov chain" (which is a product of margins and that can be seen as a particular EPMC with no $\Lambda^2$). According to *Proposition 4.1*, the DS fusion, which extends the well-known HMC, is an EPMC with $\Lambda^3 = \Omega$ ($q^{y_1, \dots, y_N}$ being a probability, simplifications within the operations specified in the proof of *Proposition 4.1* lead to $\Lambda^3 = \Omega$). The functions $q_n^3$ are then null when $A_1^3, \dots, A_N^3$ are not equal to $A_1^2, \dots, A_N^2$. For simplicity sake, we can write

$$
\begin{aligned}
& q_1^3(A_1^1, A_1^2, A_2^1, A_2^2) \\
& \quad = 1_{A_1^2 \in A_1^1} 1_{A_2^2 \in A_2^1} q_1^1(A_1^1, A_2^1) p(y_1|A_1^2) p(y_2|A_2^2); \\
& q_2^3(A_2^1, A_2^2, A_3^1, A_3^2) = 1_{A_3^2 \in A_3^1} q_2^1(A_2^1, A_3^1) p(y_3|A_3^2); \\
& \dots \\
& q_{N-1}^3(A_{N-1}^1, A_{N-1}^2, A_N^1, A_N^2) \\
& \quad = 1_{A_N^2 \in A_N^1} q_{N-1}^1(A_{N-1}^1, A_N^1) p(y_N|A_N^2).
\end{aligned}
$$

It is worth pointing out that the EPMC resulting from the above DS fusion is the "HEMC" described in Section III-C. Furthermore, setting $K = 2$, one finds the experimental examples considered in [15], where the EMC defined by $q_1^1, \dots, q_{N-1}^1$ has been used to model priors nonstationarity. Indeed, the core idea was to interpret this nonstationarity as a poor knowledge of priors. This same idea is adopted in the experiments conducted in this paper but in a more general context (the DS fusion result is more general than a HEMC since $q_1^2, \dots, q_{N-1}^2$ are no longer expressed simply as a product of margins).

*Example 4.2:* Let $\Omega = \{\omega_1, \ldots, \omega_K\}$ with the Markov distribution given by $p(x_1^N) = p(x_1)p(x_2|x_1)\ldots\ p(x_N|x_{N-1})$ known. There is a partition of $\Omega$ into $L \leq K$ subsets $\Omega_1, \ldots, \Omega_L$, such that for each subset $\Omega_i$ and for each $n = 1, \ldots, N$, the distributions $p(y_n|x_n)$ are all equal when $x_n$ varies in $\Omega_i$. Thus, we have for each $n = 1, \ldots, N$, a distribution on $\{\Omega_1, \ldots, \Omega_L\}$ defined by $p(y_n|\Omega_1), \ldots, p(y_n|\Omega_L)$. Accordingly, we can set $q_1^1(x_1, x_2) = p(x_1)p(x_2|x_1), q_2^1(x_2, x_3) = p(x_3|x_2), \ldots, q_{N-1}^1(x_{N-1}, x_N) = p(x_N|x_{N-1})$ and $q_1^2(A_1, A_2) = p(y_1|A_1)\, p(y_2|A_2)$, $q_2^2(A_2, A_3) = p(y_3|A_3), \ldots,\ q_{N-1}^2(A_{N-1}, A_N) = p(y_N|A_N)$. Then

$$\begin{aligned} & q_1^3(x_1, A_1, x_2, A_2) \\ & \quad = 1_{x_1 \in A_1} 1_{x_2 \in A_2} q_1^1(x_1, x_2) p(y_1|A_1) p(y_2|A_2); \\ & \quad q_2^3(x_2, A_2, x_3, A_3) = 1_{x_2 \in A_2} 1_{x_3 \in A_3} q_2^1(x_2, x_3) p(y_3|A_3); \\ & \quad \ldots \\ & \quad q_{N-1}^3(x_{N-1}, A_{N-1}, x_N, A_N) \\ & \quad = 1_{x_{N-1} \in A_{N-1}} 1_{x_N \in A_N} q_{N-1}^1(x_{N-1}, x_N) p(y_N|A_N). \end{aligned}$$

*Example 4.3:* Let $\Omega = \{\omega_1, \ldots, \omega_K\}$, and let us consider that the Markov distribution $p(x_1^N) = p(x_1)p(x_2|x_1)\ldots p(x_N|x_{N-1})$ is known. Let us assume that there exists an additional class $\omega_{K+1}$. As in *Example 4.2*, we set $q_1^1(x_1, x_2) = p(x_1)p(x_2|x_1), q_2^1(x_2, x_3) = p(x_3|x_2), \ldots, q_{N-1}^1(x_{N-1}, x_N) = p(x_N|x_{N-1})$ and $q^2$ can be defined by the same formulas as in the previous example, knowing that the sets $A_1, \ldots, A_N$ here vary in $\{\{\omega_1\}, \ldots, \{\omega_K\}, \Omega\}$.

*Example 4.4:* Let $\Omega = \{\omega_1, \ldots, \omega_K\}$, and let us suppose that the Markov distribution $p(x_1^N) = p(x_1)p(x_2|x_1)\ldots p(x_N|x_{N-1})$ and the distributions $p(y_n|x_n = \omega_1), \ldots, p(y_n|x_n = \omega_K)$ are known; however, the reliability of this knowledge is poor. The poor knowledge of $p(x_1^N)$ can be modeled by $q^1$ as in *Example 4.1*, and $q^2$ can be defined by the same formulas knowing that sets vary in $\{\{\omega_1\}, \ldots, \{\omega_K\}, \Omega\}$.

*Example 4.5:* To extend *Example 3.5*, let us consider a situation where the Markov distribution $p(x_1^N) = p(x_1)p(x_2|x_1)\ldots p(x_N|x_{N-1})$ is not perfectly known, which can be modeled with $q^1$ as in *Example 4.1*. Besides, there exists a partition $\Omega_1, \ldots, \Omega_L$ of $\Omega$ as specified in *Example 4.2*, which can be modeled by $q^2$ of the latter example. Then, the DS fusion is obtained by the general formula (26).

## V. Nonstationary Signals Segmentation Using Conditional Evidential Pairwise Markov Chains

To demonstrate the interest of the proposed EPMC, we choose unsupervised segmentation of nonstationary signals (with parameters estimated by EM) as an illustrative application. For this purpose, let $X_1^N = (X_1, \ldots, X_N)$ be a hidden random chain taking its values in $\Omega = \{\omega_1, \ldots, \omega_K\}$ and which is to be estimated from an observable random chain $Y_1^N = (Y_1, \ldots, Y_N)$ taking its values in $R$. Let us assume that the distributions $p(x_n, x_{n+1})$ depend on $n$. This fact can be interpreted as an unreliability of the knowledge of $p(x_n, x_{n+1})$, as done in *Example 4.1*. Considering the random chain $U_1^N = (U_1, \ldots, U_N)$ taking its values in $P(\Omega)$, let $M_1$ be an EMC defined on $P(\Omega)^N$ by (20), and let $M_2$ be the Bayesian distribution defined by the observation $y_1^N$ and given by

$$M_2(x_1, \ldots, x_N) \propto \frac{p(y_1, y_2|x_1, x_2)\ldots p(y_{N-1}, y_N|x_{N-1}, x_N)}{p(y_2|x_2)\ldots p(y_{N-1}|x_{N-1})}.$$

According to *Definition 4.1*, $M_1$ is a particular EPMC defined on $P(\Omega)$ with no $\Lambda^1$ (or by considering $\Lambda^1$ as a singleton) by $q_1^1(u_1, u_2) = p(u_1, u_2)$ and $q_{n-1}^1(u_{n-1}, u_n) = p(u_n|u_{n-1})$ for $n = 3, \ldots, N$. On the other hand, $M_2$ is a particular EPMC defined on $\Omega$ with no $\Lambda^2$ by $q_1^2(x_1, x_2) = p(y_1, y_2|x_1, x_2)$ and $q_{n-1}^2(x_{n-1}, x_n) = \frac{p(y_{n-1}, y_n|x_{n-1}, x_n)}{p(y_{n-1}|x_{n-1})}$ for $n = 3, \ldots, N$. In accordance with *Remark 2.1*, when $M_1$ is null outside $\Omega^N$, the DS fusion $M_1 \oplus M_2$ defines a classic PMC. On the other hand, when $M_2$ is replaced by $M_2(x_1, \ldots, x_N) \propto p(y_1|x_1)\ldots p(y_N|x_N)$, one finds the HEMC. This shows again how EPMCs generalize both PMCs and HEMCs.

According to *Proposition 4.1* and *Definition 4.2*, the DS fusion $M = [M_1 \oplus M_2]$ defines then a CEPMC $V_1^N = (X_1^N, U_1^N)$ taking its values in $\Omega \times \Lambda^3$ with $\Lambda^3 = P(\Omega) \times \Omega$, and the result of such a fusion is a probability. Accordingly, the CEPMC $V_1^N = (X_1^N, U_1^N)$ is given by

$$p(v_1^N) \propto q_1^3(x_1, u_1, x_2, u_2)\ldots q_{N-1}^3(x_{N-1}, u_{N-1}, x_N, u_N) \tag{27}$$

where

$$q_1^3(x_1, u_1, x_2, u_2) = 1_{x_1 \in u_1} 1_{x_2 \in u_2} m_1(u_1, u_2) p(y_1, y_2|x_1, x_2),$$

and for $3 \leq n \leq N$

$$\begin{aligned} & q_{n-1}^3(x_{n-1}, v_{n-1}, x_n, u_n) \\ & \quad = 1_{x_n \in u_n} m_1(u_{n-1}, u_n) \frac{p(y_{n-1}, y_n|x_{n-1}, x_n)}{p(y_{n-1}|x_{n-1})}. \end{aligned}$$

Hence, the distributions $p(x_n, v_n|y_1^N)$ are workable, and the complexity of their computation is linear in the size of the data $N$.

### A. Marginal Posterior Mode Restoration of Evidential Pairwise Markov Chains

Let us consider the CEPMC $V_1^N = (X_1^N, U_1^N)$ of (27). In the following, we use $q$ instead of $q^3$. One can, then, define the backward functions as follows:

$$\beta_N(v_N) = 1, \text{and } \beta_{n-1}(v_{n-1}) = \sum_{v_n} q_{n-1}(v_{n-1}, v_n)\beta_n(v_n). \tag{28}$$

Hence

$$\begin{aligned} p(v_1|y_1^N) &= \frac{q_1(v_1, v_2)\beta_2(v_2)}{\beta_1(v_1)} \\ p(v_n|v_{n-1}, y_1^N) &= \frac{q_{N-1}(v_{n-1}, v_n)\beta_n(v_n)}{\beta_{n-1}(v_{n-1})}, \quad n = N, \ldots, 2. \end{aligned} \tag{29}$$

$p(v_1|y_1^N)$ and $p(v_n|v_{n-1}, y_1^N)$ for $n = 2, \ldots, N$ being known, one can compute $p(v_n|y_1^N)$ for each $n = 2, \ldots, N$, in the following classical way:

$$p(v_{n+1}|y_1^N) = \sum_{v_n} p(v_n|y_1^N)p(v_{n+1}|v_n, y_1^N). \quad (30)$$

The distributions $p(v_n, v_{n+1}|y_1^N)$ required for the parameter estimation can also be computed as follows:

$$p(v_n, v_{n+1}|y_1^N) = p(v_n|y_1^N)p(v_{n+1}|v_n, y_1^N). \quad (31)$$

Finally, since $p(v_n|y_1^N) = p(x_n, u_n|y_1^N)$, one can evaluate $p(x_n|y_1^N)$ for each $n = 1, \ldots, N$ in the following simple manner:

$$p(x_n|y_1^N) = \sum_{u_n} p(v_n|y_1^N) = \sum_{u_n} p(m_n, u_n|y_1^N). \quad (32)$$

### B. Parameter Estimation in the Gaussian Case

Let us consider the Gaussian case where the distributions $p(y_i, y_{i+1}|x_i, x_{i+1})$ are Gaussian densities (as in Section II-B). For $K$ classes, we have, thus, to estimate $2^{2K}$ parameters $m_{\text{lm}} = p(u_1 = A_l, u_2 = A_m)$, and all parameters of $K^2$ Gaussian distributions $f_{\text{ij}}(y_1, y_2) = p(y_1, y_2|x_1 = \omega_i, x_2 = \omega_j)$ in $R^2$ denoted with $N(\mu_{\text{ij}}, \Gamma_{\text{ij}})$. Denoting by $m_{\text{lm}}^{(q)}$, $\mu_{\text{ij}}^{(q)}$, and $\Gamma_{\text{ij}}^{(q)}$ the current parameters during the execution of EM algorithm, and setting $\psi_n^{(q)}(i,j,l,m) = p(x_n = \omega_i, x_{n+1} = \omega_j, u_n = A_l, u_{n+1} = A_m|y_1^N)$ computed with (31), the next set of parameters is given by (33)–(35), as shown at the bottom of the page.

Accordingly, the EM runs as follows.

1) Find an initial value $\theta^{(0)}$ of the parameters.
2) Compute $\theta^{(q+1)}$ from $\theta^{(q)}$ and $y_1^N$ with
   a) Step E: use (28)–(31) with the current parameters $\theta^{(q)}$ to compute $\psi_{n-1}^{(q)}(i,j,l,m)$;
   b) Step M: use (33)–(35) to compute $\theta^{(q+1)}$

until and end criterion is reached.

## VI. Experiments

In this section, we present two series of experiments. In the first one, we deal with synthetic data sampled according to a nonstationary PMC. The aim is to compare the optimal results given by the true nonstationary model (called “TNS-PMC”) using the true parameters with those obtained with the four following methods: HMC, PMC, HEMC, and the proposed EPMC, all of them being based on stationary models with parameters estimated by EM. The main goal is to see whether EPMC gives good results with respect to the optimal ones provided by TNS-PMC. However, comparing EPMC with PMC, on one hand, and with HEMC, on the other hand, also is of importance to show the interest of the proposed method. In the second series, we consider a real two-class image that we corrupt in some random manner with a correlated noise. Then, the noisy image is segmented, it is to say $X_1^N = x_1^N$ is searched from $Y_1^N = y_1^N$, in an unsupervised manner through the four methods HMC, HEMC, PMC, and EPMC. The main interest is to study what happens when the data have complex distribution and follow none of the four models considered.

TABLE I
Densities Parameters of Sampled Data

| Densities | $\mu_{\text{ij}}^1$ | $\mu_{\text{ij}}^2$ | $\sigma_{\text{ij}}^1$ | $\sigma_{\text{ij}}^2$ | $\rho_{\text{ij}}$ |
|---|---|---|---|---|---|
| $f_{11}$ | −5 | −5 | 4 | 4 | $\rho_{11}$ |
| $f_{12}$ | −3 | 3 | 4 | 4 | $\rho_{12}$ |
| $f_{21}$ | 3 | −3 | 4 | 4 | $\rho_{21}$ |
| $f_{22}$ | 5 | 5 | 4 | 4 | $\rho_{22}$ |

### A. Unsupervised Segmentation of Nonstationary Pairwise Markov Chain

Let us consider a nonstationary PMC $Z_1^N = (X_1^N, Y_1^N)$ with $\Omega = \{\omega_1, \omega_2\}$ and $N = 4096$. The realization of $X_1$ is sampled uniformly from $\Omega$. Let us assume that the nonstationary distributions $p_n(i,j)$ are governed by two different matrices Q and L alternately each $s$ sites ($X_1^s, X_{2s+1}^{3s}, \ldots$ are sampled using Q, while $X_{s+1}^{2s}, X_{3s+1}^{4s}, \ldots$ are sampled using L):

$$\mathrm{Q} = \begin{pmatrix} 0.49 & 0.01 \\ 0.01 & 0.49 \end{pmatrix}, \quad \mathrm{L} = \begin{pmatrix} 0.25 & 0.25 \\ 0.25 & 0.25 \end{pmatrix}.$$

The data are sampled considering different values for $s$. For Gaussian noise densities, $\mu$ and $\sigma$ are provided in Table I, whereas the correlation coefficients $\rho$ are given in Table II (four sets of $\rho$ are considered).

MPM restoration is then achieved on one hand according to TNS-PMC with real parameters $\theta$ (used as a reference) and on the other hand according to HMC, PMC, HEMC, and the proposed EPMC using parameters estimated with EM (initialized by K-means). Average results obtained on 100 simulations are summarized in Table III.

$$m_{\text{lm}}^{(q+1)} = \frac{1}{N-1}\sum_{n=1}^{N}\sum_{\omega_i,\omega_j}\psi_{n-1}^{(q)}(i,j,l,m) \quad (33)$$

$$\mu_{\text{ij}}^{(q+1)} = \frac{\sum_{n=2}^{N}\sum_{A_l,A_m}(y_{n-1},y_n)^t\psi_{n-1}^{(q)}(i,j,l,m)}{\sum_{n=2}^{N}\sum_{A_l,A_m}\psi_{n-1}^{(q)}(i,j,l,m)} \quad (34)$$

$$\Gamma_{\text{ij}}^{(q+1)} = \frac{\sum_{n=2}^{N}\sum_{A_l,A_m}[(y_{n-1},y_n)^t - \mu_{\text{ij}}^{(q+1)}][(y_{n-1},y_n)^t - \mu_{\text{ij}}^{(q+1)}]^t\psi_{n-1}^{(q)}(i,j,l,m)}{\sum_{n=2}^{N}\sum_{A_l,A_m}\psi_{n-1}^{(q)}(i,j,l,m)}. \quad (35)$$

TABLE II
CORRELATION COEFFICIENTS OF GAUSSIAN DENSITIES

| Set | $\rho_{11}$ | $\rho_{12}$ | $\rho_{21}$ | $\rho_{22}$ |
|---|---|---|---|---|
| A | 0.5 | 0.5 | 0.5 | 0.5 |
| B | 0.9 | 0.5 | 0.5 | 0.9 |
| C | 0.5 | 0.9 | 0.9 | 0.5 |
| D | 0.9 | 0.9 | 0.9 | 0.9 |

TABLE III
SEGMENTATION ERROR RATIOS (%) OF SYNTHETIC DATA

| Set | $s$ | TNS-PMC | HMC | PMC | HEMC | EPMC |
|---|---|---|---|---|---|---|
| Set A | 16 | 8.7 | 16 | 15.4 | 15.9 | 13.5 |
| | 64 | 7.4 | 15.9 | 16 | 15 | 8.9 |
| | 256 | 8.3 | 16.2 | 17.4 | 14.5 | 8.9 |
| | 1024 | 7.5 | 15.8 | 15.9 | 14.3 | 7.6 |
| Set B | 16 | 7.4 | 21.2 | 23.5 | 21.1 | 9.2 |
| | 64 | 5.8 | 19.8 | 32.9 | 19.2 | 5.9 |
| | 256 | 5.5 | 19.9 | 30.8 | 18.5 | 5.9 |
| | 1024 | 5.6 | 19.7 | 37.4 | 18.7 | 5.7 |
| Set C | 16 | 5.6 | 17.7 | 14 | 17.7 | 10.9 |
| | 64 | 5.4 | 17.3 | 12.8 | 17.1 | 7.2 |
| | 256 | 5.5 | 17.9 | 14.2 | 17.3 | 5.9 |
| | 1024 | 5.7 | 17.4 | 13.9 | 17.3 | 5.8 |
| Set D | 16 | 2.8 | 22.3 | 6.8 | 22.1 | 4 |
| | 64 | 1.4 | 22.1 | 27.7 | 20.9 | 2.1 |
| | 256 | 2.7 | 21.3 | 44.6 | 20.8 | 2.9 |
| | 1024 | 3 | 22.2 | 44 | 21.7 | 3.1 |

The misclassification rates establish that the proposed EPMC outperforms the classic models. Moreover, the segmentation results based on the EPMC, with parameters estimated by EM, are comparable with those obtained with the TNS-PMC based on real parameters, especially for high values of $s$. It is worth pointing out that the proposed model, not only takes into account the data correlation, but rather benefits from the correlation as a feature to distinguish between the classes. In fact, for both TNS-PMC and EPMC, the best performance is observed for high values of $\rho$, whereas other models seem quite insensitive to such a parameter. Finally, notice that the HEMC, taking into account the nonstationary aspect of data, performs better than HMC and PMC in most cases.

### B. Unsupervised Segmentation of Nonstationary Images Corrupted With Correlated Noise

Let us consider the 128 × 128 "Nazca" nonstationary class image (see Fig. 1). We have then a realization of the hidden process $X$ with $\Omega = \{\omega_1, \omega_2\}$, where $\omega_1$ and $\omega_2$ correspond to black pixels and white ones, respectively. Then, the image is corrupted with a correlated noise. The observed process is $Y_s = \sigma_{x_s} W_s + \mu_{x_s} + a(\sum_{i=1}^{4} \sigma_{x_{s_i}} W_s + \mu_{x_{s_i}})$, where $W$ is a white Gaussian noise with variance 1 and $s_1, \ldots, s_4$ denote the four neighbors of pixel $s$. The bidimensional set of pixels is transformed into a mono-dimensional sequence via Hilbert–Peano scan, as done in [14]. The distribution of the random

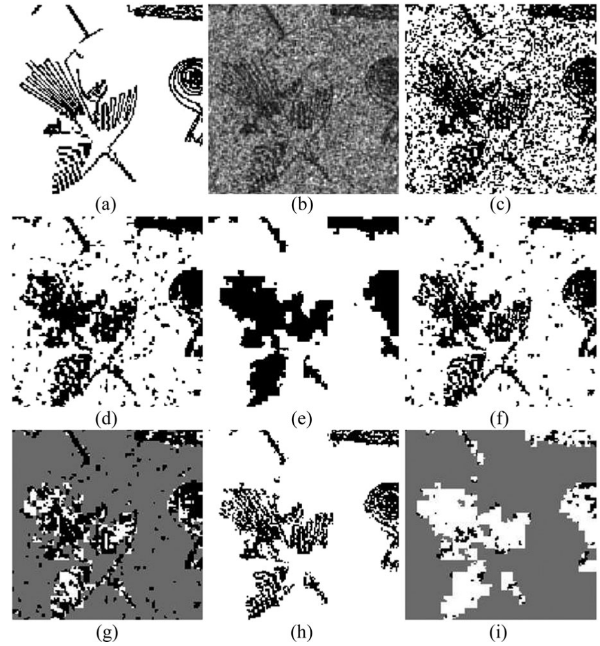

Fig. 1. Unsupervised segmentation of "Nazca" image. (a) Class image $X = x$. (b) Noised image $Y = y$. (c) K-means clustering, error ratio $\tau = 28\%$. (d) HMC-based segmentation, error ratio $\tau = 13.8\%$. (e) PMC-based segmentation, error ratio $\tau = 13.5\%$. (f) HEMC-based segmentation, error ratio $\tau = 9.9\%$. (g) HEMC-based estimate of $U$. (h) EPMC-based segmentation, error ratio $\tau = 6\%$. (i) EPMC-based estimate of $U$.

TABLE IV
SEGMENTATION ERROR RATIOS (%) OF NOISED "NAZCA" IMAGE

| $a$ | K-means | HMC | PMC | HEMC | EPMC | Gain |
|---|---|---|---|---|---|---|
| 0 | 35 | 8.2 | 8.2 | 5.8 | 5.8 | 0 |
| 0.1 | 32.1 | 9.4 | 8.8 | 6.1 | 5.5 | 10 |
| 0.25 | 28 | 13.8 | 13.5 | 9.9 | 6 | 40 |
| 0.5 | 26.7 | 14.4 | 15.2 | 13.2 | 9.1 | 31 |
| 0.75 | 26.3 | 15.2 | 15.6 | 14.4 | **11.1** | 23 |

pairwise chain $(X_1^N, Y_1^N)$ so obtained is very complex. In particular, it probably could not be considered as stationary.

Setting $\mu_{\omega_1} = 0$, $\mu_{\omega_2} = 3$, $\sigma_{\omega_1} = 1$, and $\sigma_{\omega_2} = 2$, and considering different values for $a$, we have performed numerous experiments. MPM restoration has been performed using the EM procedure according to standard HMC, PMC, HEMC, and EPMC. Average results computed on 100 simulations per each value of $a$ are reported in Table IV. Restoration results for $a = 0.25$ are also illustrated in Fig. 1.

As we can see in Fig. 1, the HMC-based segmentation is unsatisfactory. This is mainly due to the overregularization, while considering the prior distribution stationary. Indeed, the details within the bird wings and tail are blurred. Another difficulty relies in the noise correlation that cannot be handled through the conventional HMC. The repercussion of such a drawback can be checked through the salt and pepper effect in the image background. The PMC, on the other hand, takes the noise correlation into account and makes it possible to isolate the image background. However, assuming the data stationary, the details inside the image foreground are also blended due to the overregularization. The HEMC overcomes the problem of fluctuating *a priori* distribution but not the noise correlation one. The segmentation result is better, but the pepper and salt effect remains. Finally, the EPMC provides the best segmentation. In fact, the proposed model considers both noise correlation and nonstationary aspect of data. The image details are then preserved, and the background is "clean." In addition, the estimate of the auxiliary process $U$ shows clearly that the EPMC is able to distinguish between image background and foreground. The HEMC, on the other hand, cannot make such a differentiation because it does not take advantage of correlation information. Quantitatively, the proposed EPMC yields the best segmentation in all simulations, while the HEMC always performs better than both HMC and PMC. The gain in misclassification rate of the EPMC with respect to the HEMC reaches 40% for $a = 0.25$.

## VII. Conclusion

In this paper, we have introduced a general family of Markov models allowing, on one hand, to model information imprecision or unreliability and, on the other hand, to fuse such information when different sources are available. The main property of the proposed family of models was that it is closed with respect to the DS fusion. We have shown how different known models, purely probabilistic like HMCs, TMCs and DRFs, or evidential like EMCs, belong to this general family. Experimental results demonstrate the interest of the proposed extensions. For future work, it would be interesting to consider the same kind of extensions in the frame of general Bayesian networks [61]. Another promising direction would be to consider models based on Markovianity related to some "evidential conditioning," which would open new horizons. In particular, theoretical comparisons of the proposed family of models with evidential models based on Markov credibilities [28] seem to be an important topic for further studies.

## Appendix
## Proof of Proposition 4.1

We have

$$\begin{aligned}
V(v_1^N) = {} & [V^1 \oplus V^2](v_1^N) = [V^1 \oplus V^2](A_1^3, \lambda_1^3, \ldots, A_N^3, \lambda_N^3) \\
\propto {} & \sum_{(A_1^1,\ldots,A_N^1)(A_1^2,\ldots,A_N^2)} 1_{A_1^3 = A_1^1 \cap A_1^2, \ldots, A_N^3 = A_N^1 \cap A_N^2} \\
& \times V^1(A_1^1, \lambda_1^1, \ldots, A_N^1, \lambda_N^1) V^2(A_1^2, \lambda_1^2, \ldots, A_N^2, \lambda_N^2) \\
= {} & \sum_{(A_1^1,\ldots,A_N^1)(A_1^2,\ldots,A_N^2)} 1_{A_1^3 = A_1^1 \cap A_1^2, \ldots, A_N^3 = A_N^1 \cap A_N^2} \\
& \times q_1^1(A_1^1, \lambda_1^1, A_2^1, \lambda_2^1) \ldots q_{N-1}^1(A_{N-1}^1, \lambda_{N-1}^1, A_N^1, \lambda_N^1) \\
& \times q_1^2(A_1^2, \lambda_1^2, A_2^2, \lambda_2^2) \ldots q_{N-1}^2(A_{N-1}^2, \lambda_{N-1}^2, A_N^2, \lambda_N^2) \\
= {} & \sum_{(A_1^1,\ldots,A_N^1)(A_1^2,\ldots,A_N^2)} 1_{A_1^3 = A_1^1 \cap A_1^2, \ldots, A_N^3 = A_N^1 \cap A_N^2} \\
& \times q_1^1(A_1^1, \lambda_1^1, A_2^1, \lambda_2^1) q_1^2(A_1^2, \lambda_1^2, A_2^2, \lambda_2^2) \ldots \\
& \times q_{N-1}^1(A_{N-1}^1, \lambda_{N-1}^1, A_N^1, \lambda_N^1) \\
& \times q_{N-1}^2(A_{N-1}^2, \lambda_{N-1}^2, A_N^2, \lambda_N^2).
\end{aligned}$$

■

## References


[1] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains," *Annals Math. Statist.*, vol. 41, pp. 164–171, 1970.

[2] W. Qian and D. Titterington, "On the use of Gibbs Markov chain models in the analysis of images based on second-order pairwise interactive distributions," *J. Appl. Statist.*, vol. 16, no. 2, pp. 267–281, 1989.

[3] M.-Y. Chen, A. Kundu, and J. Zhou, "Off-line handwritten word recognition using a hidden Markov model type stochastic network," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 5, pp. 481–496, May 1994.

[4] I. L. MacDonald and W. Zucchini, *Hidden Markov and Other Models for Discrete-Valued Time Series*, vol. 110. Boca Raton, FL, USA: CRC Press, 1997.

[5] R. S. Mamon and R. J. Elliott, *Hidden Markov Models in Finance*, vol. 4. New York, NY, USA: Springer, 2007.

[6] R. Bhar and S. Hamori, *Hidden Markov Models: Applications to Financial Economics*, vol. 40. New York, NY, USA: Springer, 2004.

[7] T. Koski, *Hidden Markov Models for Bioinformatics*, vol. 2. New York, NY, USA: Springer, 2001.

[8] M. Vidyasagar, *Hidden Markov Processes: Theory and Applications to Biology*. Princeton, NJ, USA: Princeton Univ. Press, 2014.

[9] Y. Ephraim and N. Merhav, "Hidden Markov processes," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1518–1569, Jun. 2002.

[10] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.

[11] O. Cappé, E. Moulines, and T. Rydén, *Inference in Hidden Markov Models*, vol. 6. New York, NY, USA: Springer, 2005.

[12] R. J. Elliott, L. Aggoun, and J. B. Moore, *Hidden Markov Models*. New York, NY, USA: Springer, 1994.

[13] W. Pieczynski, "Pairwise Markov chains," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 5, pp. 634–639, May 2003.

[14] S. Derrode and W. Pieczynski, "Signal and image segmentation using pairwise Markov chains," *IEEE Trans. Signal Process*, vol. 52, no. 9, pp. 2477–2489, Sep. 2004.

[15] P. Lanchantin and W. Pieczynski, "Unsupervised restoration of hidden nonstationary Markov chains using evidential priors," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3091–3098, Aug. 2005.

[16] M. E. Y. Boudaren, E. Monfrini, W. Pieczynski, and A. Aïssani, "Dempster-Shafer fusion of multisensor signals in nonstationary Markovian context," *EURASIP J. Adv. Signal Process.*, vol. 2012, p. 134, 2012.

[17] W. Pieczynski, "Multisensor triplet Markov chains and theory of evidence," *Int. J. Approx. Reason.*, vol. 45, no. 1, pp. 1–16, 2007.

[18] E. Ramasso and T. Denoeux, "Making use of partial knowledge about hidden states in HMMs: An approach based on belief functions," *IEEE Trans. Fuzzy Syst.*, vol. 22, no. 2, pp. 395–405, Apr. 2014.

[19] E. Come, L. Oukhellou, T. Denoeux, and P. Aknin, "Mixture model estimation with soft labels," in *Soft Methods for Handling Variability and Imprecision*, D. Dubois, M. A. Lubiano, H. Prade, M. A. Gil, P. Grzegorzewski, and O. Hryniewicz, Eds. Berlin, Germany: Springer-Verlag, 2008, pp. 165–174.

[20] S. Foucher, M. Germain, J.-M. Boucher, and G. B. Benie, "Multisource classification using ICM and Dempster-Shafer theory," *IEEE Trans. Instrum. Meas.*, vol. 51, no. 2, pp. 277–281, Apr. 2002.

[21] Y. Yoji, M. Tetsuya, U. Yoji, and S. Yukitaka, "A method for preventing accidents due to human action slip utilizing HMM-based Dempster-Shafer theory," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2003, vol. 1, pp. 1490–1496.

[22] L. Fouque, A. Appriou, and W. Pieczynski, "An evidential Markovian model for data fusion and unsupervised image classification," in *Proc. 3rd Int. Conf. Inf. Fusion*, 2000, vol. 1, pp. TUB4-25–TUB4-32.

[23] M. E. Y. Boudaren and W. Pieczynski, "Unified representation of sets of heterogeneous Markov transition matrices," *IEEE Trans. Fuzzy Syst.*, vol. 24, no. 2, pp. 497–503, 2016.

[24] H. Dehghani, B. Vahidi, R. Naghizadeh, and S. Hosseinian, "Power quality disturbance classification using a statistical and wavelet-based hidden Markov model with Dempster–Shafer algorithm," *Int. J. Electr. Power Energy Syst.*, vol. 47, pp. 368–377, 2013.

[25] T. Reineking, "Particle filtering in the Dempster–Shafer theory," *Int. J. Approx. Reason.*, vol. 52, no. 8, pp. 1124–1135, 2011.

[26] E. Ramasso and R. Gouriveau, "Prognostics in switching systems: Evidential Markovian classification of real-time neuro-fuzzy predictions," in *Proc. Prognostics Health Manage. Conf.*, 2010, pp. 1–10.

[27] L. Serir, E. Ramasso, and N. Zerhouni, "Time-sliced temporal evidential networks: the case of evidential HMM with application to dynamical system analysis," in *Proc. IEEE Int. Conf. Prognostics Health Manage.*, Denver, CO, USA, Jun. 2011, pp. 1–10.

[28] E. Ramasso, M. Rombaut, and D. Pellerin, "Forward-backward-Viterbi procedures in the transferable belief model for state sequence analysis using belief functions," in *Proc. Eur. Conf. Symbolic Quantitative Approaches Reason. Uncertainty*, Hammamet, Tunisia, 2007, pp. 405–417.
[29] W. Pieczynski and D. Benboudjema, "Multisensor triplet Markov fields and theory of evidence," *Image Vision Comput.*, vol. 24, no. 1, pp. 61–69, 2006.
[30] S. Le Hégarat-Mascle, I. Bloch, and D. Vidal-Madjar, "Introduction of neighborhood information in evidence theory and application to data fusion of radar and optical images with partial cloud cover," *Pattern Recog.*, vol. 31, no. 11, pp. 1811–1823, 1998.
[31] F. Tupin, I. Bloch, and H. Maître, "A first step toward automatic interpretation of SAR images using evidential fusion of several structure detectors," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1327–1343, May 1999.
[32] A. Bendjebbour, Y. Delignon, L. Fouque, V. Samson, and W. Pieczynski, "Multisensor image segmentation using Dempster-Shafer fusion in Markov fields context," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 8, pp. 1789–1798, Aug. 2001.
[33] P. Lanchantin, J. Lapuyade-Lahorgue, and W. Pieczynski, "Unsupervised segmentation of randomly switching data hidden with non-Gaussian correlated noise," *Signal Process.*, vol. 91, no. 2, pp. 163–175, 2011.
[34] M. E. Y. Boudaren, E. Monfrini, and W. Pieczynski, "Unsupervised segmentation of random discrete data hidden with switching noise distributions," *IEEE Signal Process. Lett.*, vol. 19, no. 10, pp. 619–622, Oct. 2012.
[35] S.-Z. Yu, "Hidden semi-Markov models," *Artif. Intell.*, vol. 174, no. 2, pp. 215–243, 2010.
[36] J. Lapuyade-Lahorgue and W. Pieczynski, "Unsupervised segmentation of hidden semi-Markov non-stationary chains," *Signal Process.*, vol. 92, no. 1, pp. 29–42, 2012.
[37] J. D. Bock and G. de Cooman, "An efficient algorithm for estimating state sequences in imprecise hidden Markov models," *J. Artif. Intell. Res.*, vol. 50, pp. 189–233, 2014.
[38] G. de Cooman, J. D. Bock, and A. V. Camp, "Recent advances in imprecise-probabilistic graphical models," in *Proc. 20th Eur. Conf. Artif. Intell. Including Prestigious Appl. Artif. Intell. Syst. Demonstrations Track*, Montpellier, France, Aug. 27–31, 2012, pp. 27–32.
[39] C. D. Boom, J. D. Bock, A. V. Camp, and G. de Cooman, "Robustifying the viterbi algorithm," in *Proc. 7th Eur. Workshop Probab. Graph. Models*, Utrecht, The Netherlands, Sep. 17–19, 2014, pp. 160–175.
[40] G. McLachlan and T. Krishnan, *The EM Algorithm and Extensions*, vol. 382. Hoboken, NJ, USA: Wiley, 2007.
[41] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Royal Statist. Soc. Ser. B*, vol. 39, pp. 1–38, 1977.
[42] S. Derrode and W. Pieczynski, "Unsupervised data classification using pairwise Markov chains with automatic copulas selection," *Comput. Statist. Data Anal.*, vol. 63, pp. 81–98, 2013.
[43] M. E. Y. Boudaren, E. Monfrini, W. Pieczynski, and A. Aïssani, "Phasic triplet Markov chains," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 36, no. 11, pp. 2310–2316, Nov. 2014.
[44] G. Shafer, *A Mathematical Theory of Evidence*, vol. 1. Princeton, NJ, USA: Princeton Univ. Press, 1976.
[45] P. Smets and R. Kennes, "The transferable belief model," *Artif. Intell.*, vol. 66, no. 2, pp. 191–234, 1994.
[46] F. Janez and A. Appriou, "Theory of evidence and non-exhaustive frames of discernment: Plausibilities correction methods," *Int. J. Approx. Reason.*, vol. 18, no. 1, pp. 1–19, 1998.
[47] R. R. Yager, "Set measure directed multi-source information fusion," *IEEE Trans. Fuzzy Syst.*, vol. 19, no. 6, pp. 1031–1039, Dec. 2011.
[48] P. Walley, "Measures of uncertainty in expert systems," *Artif. Intell.*, vol. 83, no. 1, pp. 1–58, 1996.
[49] T. Augustin, F. P. Coolen, G. de Cooman, and M. C. Troffaes, "Special cases," in *Introduction to Imprecise Probabilities* (Wiley Series in Probability and Statistics.). Hoboken, NJ, USA: Wiley, 2014, ch. 4.
[50] L. A. Zadeh, *On the Validity of Dempster's Rule of Combination of Evidence*. Berkeley, CA, USA: Electron. Res. Lab., Univ. California, 1979.
[51] J. Dezert, P. Wang, and A. Tchamova, "On the validity of Dempster-Shafer theory," in *Proc. 15th Int. Conf. Inf. Fusion*, 2012, pp. 655–660.
[52] P. Walley, *Statistical Reasoning with Imprecise Probabilities*, vol. 42. London, U.K.: Chapman & Hall, 1991.
[53] E. Lefevre, O. Colot, and P. Vannoorenberghe, "Belief function combination and conflict management," *Inf. Fusion*, vol. 3, pp. 149–162, 2002.
[54] H. Soubaras, "On evidential Markov chains," in *Foundations of Reasoning Under Uncertainty*. New York, NY, USA: Springer, 2010, pp. 247–264.
[55] S. Kumar and M. Hebert, "Discriminative random fields," *Int. J. Comput. Vision*, vol. 68, no. 2, pp. 179–201, 2006.
[56] A. Quattoni, S. Wang, L.-P. Morency, M. Collins, and T. Darrell, "Hidden conditional random fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 10, pp. 1848–1852, Oct. 2007.
[57] M. E. Y. Boudaren, E. Monfrini, and W. Pieczynski, "Unsupervised segmentation of nonstationary pairwise Markov chains using evidential priors," in *Proc. 20th Eur. Signal Process. Conf.*, 2012, pp. 2243–2247.
[58] K. Sanjiv and M. Hebert, "Discriminative random fields," *Int. J. Comput. Vision*, vol. 68, no. 2, pp. 179–201, 2006.
[59] B. Settles, "Abner: an open source tool for automatically tagging genes, proteins and other entity names in text," *Bioinformatics*, vol. 21, no. 14, pp. 3191–3192, 2005.
[60] Y. Ephraim and B. L. Mark, "Causal recursive parameter estimation for discrete-time hidden bivariate Markov chains," *IEEE Trans. Signal Process.*, vol. 63, no. 8, pp. 2108–2117, Apr. 2015.
[61] F. V. Jensen, *An Introduction to Bayesian Networks*, vol. 210. London, U.K.: UCL Press, 1996.

**Mohamed El Yazid Boudaren** received the Doctor degree in computer science from both the Université des Sciences et de la Technologie Houari Boumediene (USTHB), Bab Ezzouar, Algeria, and the Université Pierre et Marie Curie, Paris, France, in 2014, and the Habilitation degree in computer science from USTHB in 2015.

He is currently an Associate Professor with Ecole Militaire Polytechnique, Algiers, Algeria. His research interests include stochastic processes and artificial intelligence.

**Wojciech Pieczynski** received the Doctorat d'Etat degree in mathematical statistics from the Université Pierre et Marie Curie, Paris, France, in 1986.

He is currently a Professor with the Telecom Sud-Paris, Evry Cedex, France, where he is the Head of the Communications, Images, and Information Processing (CITI) Department. His research interests include mathematical statistics, stochastic processes, and statistical image processing.