IASC 2008: December 5-8, Yokohama, Japan

Exact calculation of optimal filter in semi-Markov switching model

Wojciech Pieczynski[1]
[1]Telecom SudParis; 91 000 Evry, FRANCE
E-mail: Wojciech.Pieczynski@it-sudparis.eu

Keywords: Kalman filter, Markov switching system, semi-Markov switching system, triplet Markov chains

1 Introduction

Let $X = (X_1, ..., X_n, ...)$ be a "hidden" random sequence, each X_i taking its values in R^q, $Y = (Y_1, ..., Y_n, ...)$ an "observed" random sequence, each Y_i taking its values in R^m, and $R = (R_1, ..., R_n, ...)$ a discrete random sequence, each R_i taking its values in a finite set $S = \{1, ..., s\}$. For each $n = 1, 2, 3, ...$ we will set $X_1^n = (X_1, ..., X_n)$, $Y_1^n = (Y_1, ..., Y_n)$, and $R_1^n = (R_1, ..., R_n)$, respectively. The three chains are linked via some probability distribution denoted by $p(x_1^n, r_1^n, y_1^n)$. The problem considered in this paper, which is called the "filtering" problem, is to sequentially calculate $E[X_{n+1} | Y_1^{n+1} = y_1^{n+1}]$ from $E[X_n | Y_1^n = y_1^n]$. We will call such a computation "Jumping Filter", which comes from the fact that the discrete process R models the random changes of regime, or "jumps" of the distribution of (X_n, Y_n). Such models are vital in non stationary cases, and find numerous applications in different classical problems, like those needing the use of the classical Kalman filter. Otherwise, such filter is an "optimal" one as the conditional expectation classically is the best approximation when the means squared error is considered. The contribution of the paper is to consider a family of distributions $p(x_1^n, r_1^n, y_1^n)$ which are very different from the classical models, and which make possible the exact computation of the filter with linear complexity in number of observations. The ideas introduced here are inspired by the general ideas according to which finding X from Y is possible once the triplet (X, R, Y) - with R some auxiliary chain – is a Markov chain, and none one of the six marginal distributions of X, R, Y, (X, R), (X, Y), (R, Y) needs to be Markovian. These ideas lead to numerous possibilities of defining particular triplet Markov chains (X, R, Y), some of which are not more complex that the classical hidden Markov chains and though very different from them (Pieczynski and Desbouvries (2005), Ait-El-Fquih and Desbouvries (2006), Pieczynski (2007)).

Let us consider the classical Gaussian model, which consists of considering that R is a Markov chain and, roughly speaking, (X, Y) is the classical linear system conditionally on R. This is summarized in the following:

$$R \text{ is a Markov chain;} \tag{1}$$

$$X_{n+1} = F_n(R_n) X_n + W_n \; ; \tag{2}$$

$$Y_n = H_n(R_n) X_n + Z_n , \tag{3}$$

where W_1, …, W_n, … are independent Gaussian vectors in R^q, and Z_1, …, Z_n, … are independent

Gaussian vectors in R^m. For fixed $R_1 = r_1$, …, $R_n = r_n$, … the calculation of $E[X_{n+1} | Y_1^{n+1} = y_1^{n+1}]$ from $E[X_n | Y_1^n = y_1^n]$ is obtained by the well known Kalman filter, which has linear complexity in number of observations. However, such kind of calculation is no longer possible when R is random and different approximations, like "particle filtering", must be used ((Andrieu *et al.* (2003), Costa *et al.* (2005)). In fact, setting $V = (X, R)$, we have classically

$$p(v_{n+1} | y_1^{n+1}) = \frac{p(y_{n+1} | v_{n+1}) \int_{S \times R^s} p(v_n | y_1^n) p(v_{n+1} | v_n) dv_n}{p(y_{n+1} | y_1^n)} \quad (4)$$

but the computation of $p(y_{n+1} | y_1^{n+1})$ is not feasible with linear complexity in number of observations. The dependence graph of the classical model (1)-(3) is presented in Fig.1, (a).

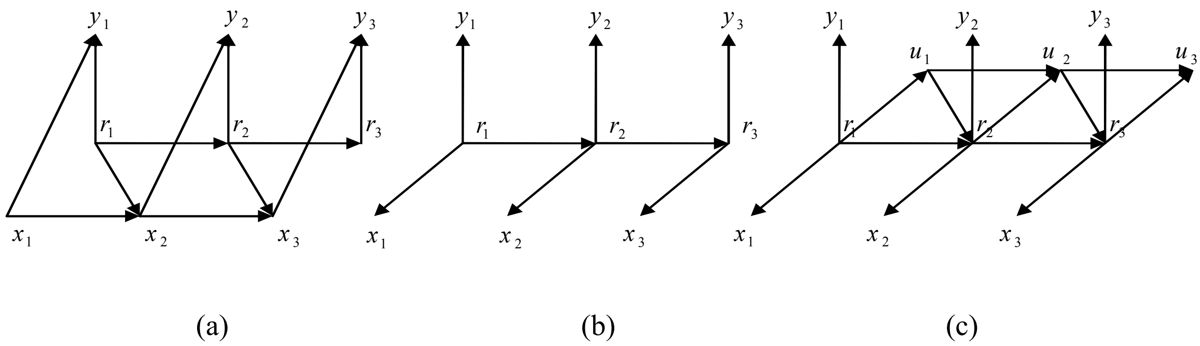

Figure 1: (a) Classical model; (b) Simplified classical model; (c) Model (b) with semi-Markov jumps

As stated above, the aim of this paper is to propose some different models allowing the exact calculation of $E[X_{n+1} | Y_1^{n+1} = y_1^{n+1})$ from $E[X_n | Y_1^n = y_1^n)$ with linear complexity in number of observations. The general idea is to leave the framework of the classical linear model conditionally on the jumps. In fact, we are faced with three random chains, such that the triplet $T = (X, R, Y)$ is a Markov chain and such "triplet" Markov chains (TMC) have been recently studied in different kind of situations (Pieczynski and Desbouvries (2005)). Although the situation here is different from different cases previously studied in that the searched chains X and R are here of different kind (X is continuous and R is discrete, when they were either both discrete as in (Pieczynski *et al.* (2002), Lanchantin and Pieczynski (2004), Le Cam *et al.* (2008)) , or both continuous as in (Desbouvries and Pieczynski (2003), Ait-El-Fquih and Desbouvries (2006)). However, the general idea is inspired from these different TMC.

As we are going to see, the main hypothesis allowing one to compute $E[X_{n+1} | Y_1^{n+1} = y_1^{n+1}]$ from $E[X_n | Y_1^n = y_1^n]$, which makes possible the calculation of $E[X_{n+1} | Y_1^{n+1} = y_1^{n+1}]$ with linear complexity in number of observations, is to assume, roughly speaking, that X and Y are independent conditionally on R. Then, partly exploiting different recent ideas relative to TMC, we will see that it is possible to consider Markov jumps, semi-Markov jumps, or even still more general random jumps models.

The contents of the paper is following. In the next section we provide the general theorem. The section three briefly recalls the first results concerning the problem proposed in (Abbassi and Pieczynski (2008)), and it is showed how these results are particular cases of the general theorem in section 2. The fourth section is devoted to two original models in which the jump chain is a Markov one or not, and the fifth section provides some conclusions and perspectives.

2 General theorem

Let us consider the three random chains $X=(X_1,...,X_n,...)$, $Y=(Y_1,...,Y_n,...)$, and $R=(R_1,...,R_n,...)$ as above. We can state the following result.

Theorem 1

For each $n \ge 1$, let $M(r_n, y_1^n)=\int_{R^q} x_n p(x_n, r_n | y_1^n) dx_n$, and $E[X_n | Y_1^n = y_1^n] = \sum_{r_n} M(r_n, y_1^n)$. Under the following hypotheses

(i) (R,Y) is a Markov chain with given transitions $p(r_{n+1}, y_{n+1} | r_n, y_n)$;

(ii) the chains X and Y are independent conditionally on R;

(iii) there exist W_1, …, W_n, … independent random centered vectors in R^q such that $X_{n+1}=F_n(R_n)X_n+W_n$ for each $n \ge 1$,

we have

$$M(r_{n+1}, y_1^{n+1}) = \sum_{r_n} \frac{p(r_{n+1}, y_{n+1} | r_n, y_n)}{p(y_{n+1} | y_1^n)} F_n(r_n) M(r_n, y_1^n), \tag{5}$$

and

$$p(r_{n+1}, y_1^{n+1}) = \sum_{r_n} p(r_n, r_{n+1}, y_1^n, y_{n+1}) = \sum_{r_n} p(r_n, y_1^n) p(r_{n+1}, y_{n+1} | r_n, y_n). \tag{6}$$

As a result, $p(y_{n+1} | y_1^n)$ in (5) is computable from $p(y_1^{n+1}) = \sum_{r_{n+1}} p(r_{n+1}, y_1^{n+1})$ given by (6), and thus the expectation $E[X_{n+1} | Y_1^{n+1} = y_1^{n+1}]$ is calculable from $M(r_n, y_1^n)$.

Consequently, the optimal filter $E[X_{n+1} | Y_1^{n+1} = y_1^{n+1}]$ is computable with linear complexity in number of observations.

In addition, the transitions and the marginal distributions of the Markov distribution $p(r|y)$ are also classically computable with linear complexity in number of observations, which make possible the estimation of $R=r$.

Proof.

According to the hypotheses $T=(X,R,Y)$ is a Markov chain; moreover, we have

$$p(x_{n+1}, r_{n+1}, y_{n+1} | x_n, r_n, y_n) = p(r_{n+1}, y_{n+1} | x_n, r_n, y_n) p(x_{n+1} | r_{n+1}, y_{n+1}, x_n, r_n, y_n) =$$

$$= p(r_{n+1}, y_{n+1} | r_n, y_n) p(x_{n+1} | x_n, r_n) \tag{7}$$

Thus we can write

$$p(x_n, r_n, x_{n+1}, r_{n+1} | y_1^{n+1}) = \frac{p(x_n, r_n | y_1^n) p(x_{n+1}, r_{n+1}, y_{n+1} | x_n, r_n, y_1^n)}{p(y_{n+1} | y_1^n)} =$$

$$\frac{p(x_n, r_n | y_1^n) p(x_{n+1}, r_{n+1}, y_{n+1} | x_n, r_n, y_n)}{p(y_{n+1} | y_1^n)} = [\frac{p(x_n, r_n | y_1^n) p(r_{n+1}, y_{n+1} | r_n, y_n)}{p(y_{n+1} | y_1^n)}][p(x_{n+1} | x_n, r_n)] \tag{8}$$

Then we use $p(x_n, r_n, x_{n+1}, r_{n+1} | y_1^{n+1})$ expressed by (8) to compute $M(r_{n+1}, y_1^{n+1})$:

$$M(r_{n+1}, y_1^{n+1}) = \int_{R^q} x_{n+1} p(x_{n+1}, r_{n+1} | y_1^n) dx_{n+1} = \int_{R^q} x_{n+1} [\sum_{r_n} \int_{R^q} p(x_n, r_n, x_{n+1}, r_{n+1} | y_1^n) dx_n] dx_{n+1} =$$

$$\int_{R^q} [\frac{1}{p(y_{n+1} | y_1^n)}][\sum_{r_n} p(x_n, r_n | y_1^n) p(r_{n+1}, y_{n+1} | r_n, y_n) \int_{R^q} x_{n+1} p(x_{n+1} | x_n, r_n) dx_{n+1}] dx_n$$

According to (iii) we have $\int_{R^q} x_{n+1} p(x_{n+1} | x_n, r_n) dx_{n+1} = F_n(r_n) x_n$, and thus

$$M(r_{n+1}, y_1^{n+1}) = \int_{R^q} [\frac{1}{p(y_{n+1} | y_1^n)}][\sum_{r_n} p(x_n, r_n | y_1^n) p(r_{n+1}, y_{n+1} | r_n, y_n) F_n(r_n) x_n] dx_n =$$

$$= \sum_{r_n} \frac{p(r_{n+1}, y_{n+1} | r_n, y_n)}{p(y_{n+1} | y_1^n)} F_n(r_n) \int_{R^q} x_n p(x_n, r_n | y_1^n) dx_n = \sum_{r_n} \frac{p(r_{n+1}, y_{n+1} | r_n, y_n)}{p(y_{n+1} | y_1^n)} F_n(r_n) M(r_n, y_1^n),$$

which is (5).

(6) classically comes from the fact that (R, Y) is a Markov chain; in fact, we have $p(r_n, r_{n+1}, y_1^n, y_{n+1}) = p(r_n, y_1^n) p(r_{n+1}, y_{n+1} | r_n, y_1^n) = p(r_n, y_1^n) p(r_{n+1}, y_{n+1} | r_n, y_n)$.

Finally, (R, Y) is a "pairwise" Markov chain (which simply means that (R, Y) is a Markov chain) according to (i) and thus the same calculations as in the classical hidden Markov chains are workable (Derrode and Pieczynski (2004)), which ends the proof.

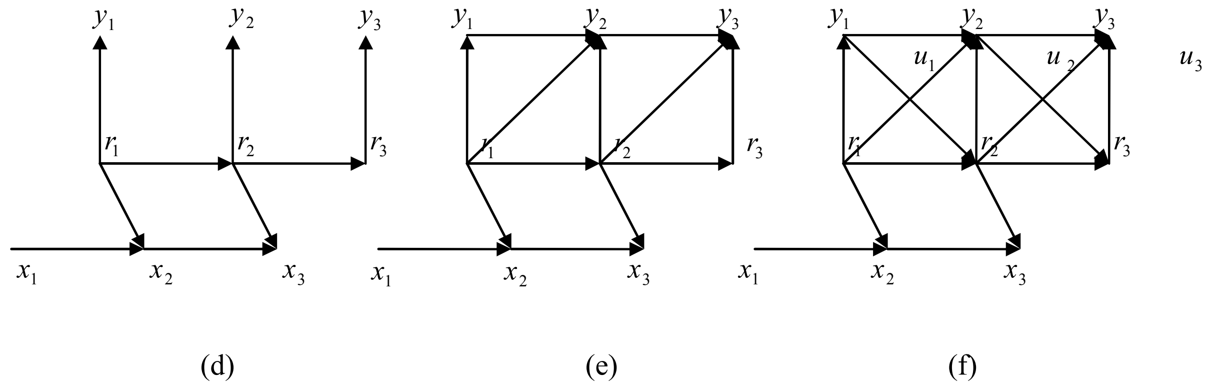

Figure 2: (d) Second simplified classical model: Markov jumps with independent noise; (e) Simplified classical model: Markov jumps with correlated noise; (f) General new model, with non necessarily Markov jumps

Remark 1

Let us insist on the fact that the distribution of the jumping chain R is not necessarily a Markov one. For example, for stationary and reversible chain (R,Y) it has been showed in (Pieczynski (2007)) that R is a Markov chain if and only if for each $n \geq 1$ and $1 \leq i \leq n$ one has $p(y_i | r_1^n) = p(y_i | r_i)$. Thus the Markovianity of (R,Y) is a rather mild condition, and the model (i)-(iii) contains a number of particular simpler models. The dependence graphs of some models are presented in Fig. 2: the very classical hidden Markov chain with independent noise corresponds to (d), a hidden Markov chain model with correlated noise corresponds to (e), while the graph corresponding to the general model is given in (f). In addition, different experiments presented in (Derrode and Pieczynski (2004)) show that the pairwise Markov chains which are not the classical hidden Markov chains (i.e. in which the hidden chain R is not a Markov one) can be much more efficient. Therefore using pairwise Markov chains instead of the classical hidden Markov ones should also be of interest in the context of the present paper.

Remark 2

The independence of X and Y conditionally on R - hypothesis (ii) – is the key point in the model and one could wonder whether such hypothesis can turn out to be penalizing in real situations. Important is to notice that the "conditional independence" is a very different notion from the "independence" notion; in other words, "conditional independence" can be very close, or event identical, to the "dependence". For example, having two dependent Gaussian real random variables A and B, one can always find a third real Gaussian random variable C such that A and B are independent conditionally on C. In other words, the "dependence" and the "conditional independence" are identical notions in the case of Gaussian real variables. In addition, such kind of triplet Markov models has already been successfully used in (Lanchantin and Pieczynski (2004)). There are a searched discrete X, an auxiliary discrete U (which models different stationarities of the model, and thus whose meaning is similar to the meaning of R in the present paper), and an observed continuous chain Y. The chain (X,U,Y) is assumed to be a Markov one and both X and U are searched from Y. Although the chains U and Y are independent conditionally on X, the estimation of U from Y is not worse that the estimation of X.

3. General model and semi-Markovian jumps

We have seen in the previous section that the general model of the Theorem 1 can produce either a Markov jumps process (Fig. 2, (d) and (e)), or more general one, whose distribution is the marginal distribution of a Markov one (Fig. 2, (f)). The aim of this section is to present a new model, where the jump process is a semi-Markov one. The general idea is the following. As already specified in (Lapuyade-Lahorgue and Pieczynski (2006)), a semi-Markov distribution of R can be seen as the marginal distribution of a particular pairwise Markov chain (R, U). Then the hidden semi-Markov chain (R, Y) can be seen as a triplet Markov chain (R, U, Y). In other words, we are going to specify the same kind of results as in the previous section, replacing the pairwise Markov chain (R, Y) by the triplet Markov chain (R, U, Y). Finally, we will have a quadruplet random chain (X, R, U, Y), where X is the searched chain as above, R is the jump process as above, U is an auxiliary process which will make R be semi-Markov, and Y is the observed process as above. Such kind of models could appear as somewhat complex; however, a similar model has been used in (Lapuyade-Lahorgue and Pieczynski (2006)). In fact, in the latter paper there are a hidden discrete searched chain X, an observed continuous chain Y, a jump chain R which models the different stationarities, and an auxiliary chain U, which makes X semi-Markov. We see that there are two differences between the model in (Lapuyade-Lahorgue and Pieczynski (2006)) and the model we are going to develop: (i) in this paper X is continuous instead of being discrete, and (ii) U is used to make semi-Markov R, instead of X. However, the results produced in (Lapuyade-Lahorgue and Pieczynski (2006)) show that such models can be of real interest, even in the unsupervised context, where all the model parameters are estimated from the only observed data Y.

Let us first recall how to introduce the distribution of (R, U) to make the distribution of R semi-Markov. Following the model proposed in (Lapuyade-Lahorgue and Pieczynski (2006)), we will assume that each U_i takes its values in $\Upsilon = \{0, 1, ..., m\}$, so that (R, U) is a finite Markov chain. For $(R_n, U_n) = (r_n, u_n)$, the number u_n denotes the minimal sojourn time of the next R_{n+1}, …. in r_n. Therefore, if $u_n = j > 0$, we have $(r_{n+1}, u_{n+1}) = (r_n, u_n - 1)$, …, $(r_{n+j}, u_{n+j}) = (r_n, 0)$. If $u_n = 0$, the distribution of R_{n+1} is a given transition $p(r_{n+1} | r_n, u_n = 0)$. Finally, the transition $p(r_{n+1}, u_{n+1} | r_n, u_n) = p(r_{n+1} | r_n, u_n) p(u_{n+1} | r_{n+1}, r_n, u_n)$ of the Markov chain (R, U) is defined by

$$p(r_{n+1} | r_n, u_n) = \delta_{r_n}(r_{n+1}) \text{ if } u_n > 0 \text{, and } p(r_{n+1} | r_n) \text{ if } u_n = 0 \; ; \tag{9}$$

$$p(u_{n+1} | r_{n+1}, r_n, u_n) = \delta_{u_n - 1}(u_{n+1}) \text{ if } u_n > 0 \text{, and } p(u_{n+1} | r_{n+1}) \text{ if } u_n = 0 \; ; \tag{10}$$

with $\delta_a(b) = 1$ for $a = b$, and $\delta_a(b) = 0$ otherwise.

Therefore we have four chains X, R, U, and Y and the problem is the same as in the first section : calculate $E[X_{n+1} | Y_1^{n+1} = y_1^{n+1})$ from $E[X_n | Y_1^n = y_1^n)$. Roughly speaking, we will state a result similar to the result in Theorem 1, with R replaced by (R, U).

Theorem 2
Let us consider four chains X, R, U, and Y verifying following hypotheses:

(1) $V = (R,U)$ is a Markov chain verifying (9)-(10) (R is a semi-Markov chain);
(2) (V,Y) is a Markov chain with computable transitions $p(v_{n+1}, y_{n+1} | v_n, y_n)$;
(3) the chains X and Y are independent conditionally on V;
(4) there exist W_1, …, W_n, … independent random centered vectors in R^q such that $X_{n+1} = F_n(V_n)X_n + W_n$ for each $n \geq 1$;

Then the optimal filter $E[X_{n+1} | Y_1^{n+1} = y_1^{n+1}]$ is computable with linear complexity in number of observations. In addition, the transitions and the marginal distributions of the Markov distribution $p(r,u|y)$ are also classically computable with linear complexity in number of observations, which make possible the estimation of $R = r$.

The proof is immediate, as the hypotheses (2)-(4) in the Theorem 2 are identical to the hypotheses (i)-(iii) in the Theorem 1.
Finally, we can say that the model specified in Theorem 2 is a particular case of the model specified in Theorem 1.

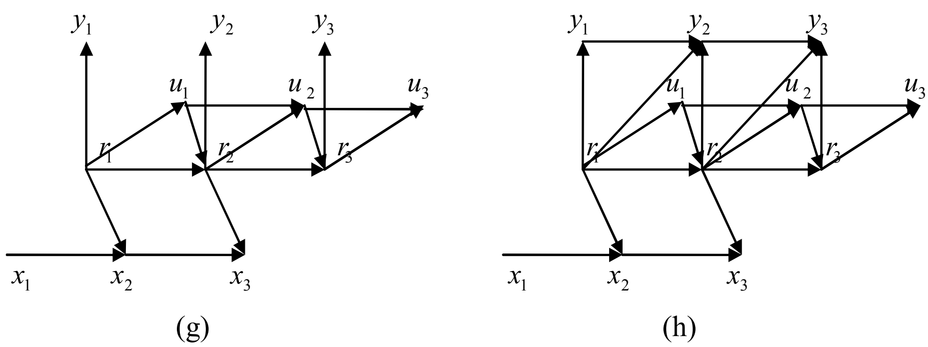

Figure 3: (g) new model with semi-Markov jumps and independent noise, (h) new model with semi-Markov jumps and correlated noise.

Let us consider two examples of such a semi-Markov jumps models whose dependence graphs are presented in Fig. 3. The model (g) is similar to the model (d) in Figure 2, except the fact that the jumps chain is semi-Markovian instead of being Markovian. The model (h) extends the model (g) in that the noise is correlated and $p(y_{n+1} | r_n, r_{n+1})$ can depend on both r_n and r_{n+1}.

4 Conclusion

We considered in this paper the problem of filtering continuous multivariate data in the presence of random jumps. When the jump process is assumed to be Markov in the classical linear systems, the exact computation with linear complexity in number of observations is not possible and different approximation techniques must be used. Using different general ideas inspired by different recent works on triplet Markov chains, we proposed different general families of models in which the filtering with linear complexity in number of observations is feasible.

Validation in real situations will be a natural perspective for further studies on the subject.

References

Abbassi*, N.* and Pieczynski, W. (2008). Exact Filtering In Semi-Markov Jumping System, Sixth International Conference of Computational Methods in Sciences and Engineering, September 25-30, Hersonissos, Crete, Greece.

Ait-El-Fquih, B., and Desbouvries, F. (2006). Kalman Filtering in Triplet Markov Chains, *IEEE Trans. on Signal Processing*, 54(8): 2957-2963, IEEE Computer Society.

Andrieu, C., Davy, C. M., and Doucet, A. (2003). Efficient particle filtering for jump Markov systems. Application to time-varying autoregressions, *IEEE Trans on Signal Processing*, 51(7): 1762-1770, IEEE Computer Society..

Costa, O. L. V., Fragoso, M. D., and Marques, R. P. (2005). Discrete time Markov jump linear systems, New York, Springer-Verlag.

Derrode S. and Pieczynski, W. (2004). Signal and Image Segmentation using Pairwise Markov Chains, *IEEE Trans. on Signal Processing*, 52 (9) : 2477-2489, IEEE Computer Society.

Desbouvries F. et Pieczynski, W. (2003). Modèles de Markov Triplet et filtrage de Kalman, *Comptes Rendus de l'Académie des Sciences – Mathématique*, Série I, 336(8) : 667-670, Elsevier.

Hirayama, J. I. Maeda, S. I., and Ishii, S. (2007), Markov and semi-Markov switching of source appearances for nonstationary independent component analysis, *IEEE Trans. on Neural Networks*, 18(5) : 1326-1342, IEEE Computer Society.

Le Cam, S., Collet, C. Salzenstein, F. (2008). Acoustical respiratory signal analysis and phase detection, IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP'08, March 30 - April 4, Las Vegas, Nevada, USA, IEEE Computer Society.

Lanchantin, P. and Pieczynski, W. (2004). Unsupervised non stationary image segmentation using triplet Markov chains, Advanced Concepts for Intelligent Vision Systems (ACVIS 04), August 31-September 3, Brussels, Belgium.

Lapuyade-Lahorgue, J. and Pieczynski, (2006). W. Unsupervised segmentation of hidden semi-Markov non stationary chains, Twenty six International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, MaxEnt 2006, Paris, France, July 8-13.

Pieczynski, W., Hulard, C. and Veit, T. (2002). Triplet Markov Chains in hidden signal restoration, SPIE's International Symposium on Remote Sensing, September 22-27, Crete, Greece.

Pieczynski W., and Desbouvries, F. (2005). On triplet Markov chains, International Symposium on Applied Stochastic Models and Data Analysis, (ASMDA), Brest, France.

Pieczynski, W. (2007). Multisensor triplet Markov chains and theory of evidence, *International Journal of Approximate Reasoning*, 45(1):1-16, Elsevier.

Salzenstein, F., Collet, C., Le Cam, S., and Hatt, M. (2007). Non stationary fuzzy Markov chains, *Pattern Recognition Letters*, 28(16): 2201-2208, Elsevier.