Mixed state spatio-temporal auto-models

Cécile Hardouin, Tomas Crivelli

Mixed-state spatio-temporal auto-models

C. Hardouin°, T. Crivelli•

Abstract

We consider in this paper a general modelling for mixed-state data. Such data consist of two components of different types: the observations record many zeros, together with continuous real values. They occur in many application fields, like rainfall measures, or motion analysis from image sequences. The aim of this work is to present ad hoc spatio-temporal models for these kinds of data. We present a Markov Chain of Markov fields modelling, the Markovian fields being defined as mixed-state auto-models, whose local conditional distributions belong to an exponential family and the observations derive from mixed-states variables. Some specific examples are given as well as some preliminary experiments.

° SAMM, Université Paris 1, 90 rue de Tolbiac, 75634 Paris Cedex 13, France
• Faculty of Engeneering, University of Buenos Aires, Argentina.

Keywords: mixed-state variable, auto-models, Markov Chain, Markov fields

1 Introduction

This paper is devoted to the study of spatio-temporal data of dual nature, the data being made of both continuous and discrete values. There is a large variety of domains where we can observe such phenomenon. For instance, pluviometry series often consist of rainfall values during some periods, followed by zeros when the rain is absent ([1]). In the epidemiological context, we collect positive values when the disease spreads out, and then zeros when it disappears. We can think also of differences of some data, temperatures for instance, or motion measurements from video sequences. We are interested in the modelling of such temporal data, collected on a lattice, and whose main characteristic is the double feature (discrete/continuous) of its values.

We can find in the literature various attempts involving hierarchical construction. They consider a latent process, which is inherently unobservable, and needs computational supplementary efforts to be restored. Our approach is different; we propose a direct modelling, including simultaneously the discrete and continuous components of the data. The main idea is to consider a temporal Markov chain of spatial Markov fields, the latter accounting for the mixed feature of the data. More precisely, we take for those random fields mixed-state auto-models presented in [11].

In section 2 we define mixed-state variables and then turn out to the temporal dynamics on mixed-state Markovian fields; we crucially use some results given in [12] and [11], that concern the building of spatial multi-parameter auto-models. The general Markov Chain of Markov Fields dynamics is thoroughly described in [10].

We consider in section 3 the example of a mixed auto-exponential dynamics, the observations belonging to a mixed-state space $E = \{0\} \cup]0, \infty[$, with an exponential distribution on the positive part; we explore the properties of this model, especially ergodicity. A mixed auto-normal dynamics is studied in section 4; after the study of its general properties, we apply the model to motion texture modelling; the method achieves good performances for further recognition, segmentation and tracking.

2 Mixed-state spatio-temporal modelling

Let us define what we call "mixed-state" random variables, defined on a "mixed-state space" E. In the pure spatial context, Besag ([2]) defined auto-models as particular Markovian random fields on a finite lattice; roughly speaking, they rely on two main assumptions, that is, each local conditional density belongs to an exponential family, and interactions between sites rely on pairwise dependence. The concept has been extended to the multivariate case by [12]. Then it made possible to define auto-models on a lattice for mixed-state observations, see [11]. We consider now the spatio-temporal framework, adding a temporal dynamics over the spatial scheme. We consider a Markov chain in time, of Markov fields in space, those fields being defined in an analogous way to the auto-models for mixed-state observations of [11].

2.1 Random variables with mixed states

Random variables with mixed states have been first introduced in [12] and developed in [11]. We present the results in the context of the mixed-state space $E = \{0\} \cup (0, \infty)$, but they hold equally for a general state space $E = F \cup G$, where $F = \{e_1, ..., e_M\}$ and G is a subset of $\mathbb{R}^p$ (see [11]). Of course, E can be written $E = [0, \infty[$ but the previous writing points out the dual feature of the measurements. E is equipped with a "mixed" reference measure

$$m(dx) = \delta_0(dx) + \lambda(dx) ,$$

where δ_0 is the Dirac measure at 0 and λ the Lebesgue measure on $(0, \infty)$.

Any random variable X taking its values in E is called a mixed-state random variable; such a variable arises from the following construction: with probability $\gamma \in]0, 1[$ we set $X = 0$; otherwise X is distributed with a continuous probability density function g on $]0,\infty[$. We now make an essential assumption on this density, that is it belongs to an $s-$dimensional exponential family:

$$g_\xi(x) = H(\xi)L(x)\exp\langle\xi, T(x)\rangle \ , \qquad \xi \in \mathbb{R}^s \ , \quad T(x) \in \mathbb{R}^s$$

where $\langle,\rangle$ denotes the scalar product in $\mathbb{R}^l$.

We define $\delta^*(x) = 1 - \mathbf{1}_{\{0\}}(x)$. Then we write X's probability density function as follows, with respect to $m(dx)$:

$$\begin{aligned} f_\theta(x) &= \gamma\delta(x) + (1-\gamma)\delta^*(x)g_\xi(x) \\ &= \exp\{\delta^*(x)\ln\frac{(1-\gamma)H(\xi)}{\gamma} + \langle\xi, T(x)\delta^*(x)\rangle + \ln\gamma + \delta^*(x)\ln L(x)\} \\ &= H'(\theta)L'(x)\exp\langle\theta, B(x)\rangle \qquad (1) \end{aligned}$$

with $H'(\theta) = \gamma, \ \ L'(x) = \exp\{\delta^*(x)\ln L(x)\}$.

From 1, f_θ also belongs to an exponential family, of dimension $s+1$, where the canonical parameter and the sufficient statistics are respectively defined by

$$\theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} = \begin{pmatrix} \ln\frac{(1-\gamma)H(\xi)}{\gamma} \\ \xi \end{pmatrix} \ , \quad B(x) = \begin{pmatrix} \delta^*(x) \\ \delta^*(x)T(x) \end{pmatrix} \ .$$

Moreover the original parameters ξ and γ can be recovered from θ by

$$\xi = \theta_2 \ , \quad \gamma = \frac{H(\theta_2)}{H(\theta_2) + e^{\theta_1}} \ .$$

Let us note that the use of δ^* ensures the normalization equality $B(0) = 0$, which will be of interest subsequently.

As a simple example, if we consider the exponential density function on $]0,\infty[$, $g_\xi(x) = g_\lambda(x) = \lambda\exp(-\lambda x)$, with $\lambda > 0$, then (1) reduces to $f_\theta(x) = H'(\theta)\exp\langle\theta, B(x)\rangle$ with

$$\theta = \begin{pmatrix} \ln\frac{(1-\gamma)\lambda}{\gamma} \\ \lambda \end{pmatrix} \ , \quad B(x) = \begin{pmatrix} \delta^*(x) \\ -x\delta^*(x) \end{pmatrix} \ .$$

Let us now consider the Gaussian case on $E = \{0\} \cup \mathbb{R}$, $g_\xi(x) = g_{m,\sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp -\frac{(x-m)^2}{2\sigma^2}$, then $f_\theta(x) = H'(\theta)\exp\langle\theta, B(x)\rangle$ with

$$\theta = \begin{pmatrix} \ln\frac{(1-\gamma)g_{\mu,\sigma}(0)}{\gamma} \\ \frac{1}{2\sigma^2} \\ \frac{m}{\sigma^2} \end{pmatrix} \ , \quad B(x) = \begin{pmatrix} \delta^*(x) \\ -x^2\delta^*(x) \\ x\delta^*(x) \end{pmatrix} \ .$$

For other examples and general construction of mixed-state random variables, we refer the reader to [11].

2.2 Markov chain dynamics for mixed-state auto-models

Let us consider the spatio-temporal scheme. The idea is to build a temporal dynamics on spatial fields. We use the general Markov Chain of Markov Field modelling given in [10] on one hand, and suited spatial auto-models for mixed-state observations described in [11] on the other hand.

Let us precise the settings. Let S be a finite set of sites. $X = \{X(t),\ t \in \mathbb{N}^*\}$ is a Markov chain on E^S (equipped with the product measure $m^{\otimes S}$). Each $X(t) = \{X_i(t),\ i \in S\}$ is, conditionally to the past, a Markov random field on S; more precisely, $X(t)$ is a mixed-state auto-model,with $X_i(t) \in E$.

Let us introduce the following notations. We denote $x_A = \{x_i,\ i \in A\}$ and $x^A = \{x_j,\ j \notin A\}$. We shall write x_i (resp. x^i) for $x_{\{i\}}$ (resp. $x^{\{i\}}$). Finally, we use the notation $X_i(t) = y_i$ for the variable at site i at the present time t, and $X_i(t-1) = x_i$ for the variable at site i at the past time $t-1$.

Our model is defined as follows:

[B1] $X = \{X(t),\ t \in \mathbb{N}^*\}$ is an homogeneous Markov chain (of order one) on E^S.

We assume that the transition probability measure of X has a positive density $P(x,y)$ with respect to $m^{\otimes S}$, which verifies:

$$P(x,y) = Z^{-1}(x)\ \exp Q(y \mid x) \tag{2}$$

where $Z(x) = \int_{E^S} \exp Q(y \mid x)\ m^{\otimes S}(dy) < \infty$.

[B2] Conditionally to the past, the instantaneous dependence is pairwise only:

$$Q(y \mid x) = \sum_{i \in S} G_i(y_i \mid x) + \sum_{\{i,j\}} G_{ij}(y_i, y_j \mid x)\ .$$

with, almost surely in x, $G_i(0 \mid x) = G_{ij}(0, y_j \mid x) = G_{ij}(y_i, 0 \mid x) = 0$ for all $i \in S$, $j \in S$.

[B3] For each i, conditionally to $(X^i(t) = y^i\ ,\ X(t-1) = x)$, the distribution of $X_i(t)$ is a mixed-state distribution defined in (1).

$$\ln f_i(y_i \mid y^i, x) = \langle \theta_i(y^i, x), B_i(y_i) \rangle + C_i(y_i) + D_i(y^i, x) \tag{3}$$

with $\theta_i(y^i, x) \in \mathbb{R}^d$, $B_i(y_i) \in \mathbb{R}^d$ and the identifiability conditions $B_i(0) = C_i(0) = 0$ for all $i \in S$.

[B4] For all $i \in S$, Span $\{B_i(y_i),\ y_i \in E\} = \mathbb{R}^d$.

Equation (2) in assumption [B1] defines the temporal dynamics, and states also the so called positivity condition; the Hammersley-Clifford's Theorem gives then a characterization of P by an energy Q equal to a sum of potentials G defined on a set of cliques.

Assumptions [B2] and [B3] set the spatial auto-model, i.e. the pairwise only dependence (with identifiability conditions), and distribution relying in an exponential family. Assumption [B4] is a regularity condition necessary to define multivariate auto-models (see [12]). Thus, [B2], [B3] and [B4] define the spatial mixed-state auto-model, conditionally to the past; adding [B1] completes the spatio-temporal modelling. We refer the reader to the references for details.

We derive the explicit form of the joint distribution, conditionally to the past.

Proposition 1 *Let us assume that the Markov chain X satisfies assumption [B1], and the random field probability distribution and its energy $Q(y \mid x)$ satisfy [B2], [B3], and [B4]. Then, conditionally to $X(t-1) = x$, there exists a family of d-dimensional vectors $\{\alpha_i(x),\ i \in S\}$ and a family of $d \times d$ matrices $\{\beta_{ij}(x),\ i, j \in S,\ i \neq j\}$ verifying $\beta_{ij}(x) = \beta_{ij}(x)^T$ such that*

$$\theta_i(y^i, x) = \alpha_i(x) + \sum_{j:\{i,j\}} \beta_{ij}(x) B_j(y_j)$$

Consequently the set of potentials is given by

$$G_i(y_i \mid x) = \langle \alpha_i(x), B_i(y_i) \rangle + C_i(y_i)$$
$$G_{ij}(y_i, y_j \mid x) = B_i(y_i)^T \beta_{ij}(x) B_j(y_j)$$

This result ensures directly from the spatial auto-model feature (see [11] and [10]). It holds for all conditional distribution lying in the exponential family, and is not specific for mixed-state distributions. It gives the necessary form of the local canonical parameters and the potentials to ensure the compatibility of the conditional distributions in order to reconstruct the (conditional to the past) global energy.

We can see from the writing above that the conditional energy Q is linked to two families of potentials; here we have a semi-causal representation associated to a double graph $G = \{\mathcal{G}, \mathcal{G}^-\}$ where $\mathcal{G}$ and $\mathcal{G}^-$ depict respectively the time-instantaneous and time-delay dependencies. $\langle i, j \rangle_{\mathcal{G}}$ holds for $\langle (t, i), (t, j) \rangle_G$, i.e. the sites i and j are neighbours at the same time, while $\langle i, j \rangle_{\mathcal{G}-}$ means $\langle (t-1, i), (t, j) \rangle_G$, that is i belongs to the past neighbourhood of j. The instantaneous graph $\mathcal{G}$ is symmetric while $\mathcal{G}^-$ is a directed graph. Then, a site i has two neighbourhoods which we denote by $\partial i = \{j \in S \setminus \{i\} : \langle i, j \rangle_{\mathcal{G}}\}$ and $\partial i^- = \{j \in S : \langle j, i \rangle_{\mathcal{G}^-}\}$.

The general framework above contains a large variety of possible models. Indeed the different models are depicted by the choices for the functions α and β; these functions describe both past interactions and instantaneous dependencies; their form is free, subject to the integrability of $\exp Q(y \mid x)$.

We turn to the examination of two examples; first, we consider a dynamics associated to a mixed exponential auto-model on $E = \{0\} \cup (0, \infty)$. We next look for the Gaussian case on $E = \{0\} \cup \mathbb{R}$. For each example, we give conditions for the well-definiteness of the model, and investigate their properties, especially ergodicity.

3 The mixed auto-exponential dynamics

Let us now specify assumption [B3]. The state space is $E = \{0\} \cup (0, \infty)$. For each i, conditionally to $(X^i(t) = y^i\ ,\ X(t-1) = x)$, the distribution of $X_i(t)$ is a mixed exponential distribution defined as follows; $X_i(t) = 0$ with probability $\gamma_i(y^i, x)$, else $X_i(t)$ follows an exponential distribution with parameter $\lambda_i(y^i, x) > 0$. The probability density function of $X_i(t)$ (given $X(t-1) = x$) is

$$\ln f_i(y_i \mid y^i, x) = \langle \theta_i(y^i, x), B_i(y_i) \rangle + D_i(y^i, x) \tag{4}$$

where

$$\theta_i(y^i, x) = \begin{pmatrix} \ln \frac{(1-\gamma_i(y^i,x))\lambda_i(y^i,x)}{\gamma_i(y^i,x)} \\ \lambda_i(y^i, x) \end{pmatrix} \ , \quad B(y) = \begin{pmatrix} \delta^*(y) \\ -y\delta^*(y) \end{pmatrix} .$$

We verify that $B_i(0) = 0$ for all $i \in S$. Let us note that the contributions of y^i and x in the expressions above are in fact restricted to $y_{\partial i}$ and $x_{\partial i^-}$.

Then, from Proposition 1, and conditionally to $X(t-1) = x$, there exist a family $\alpha_i(x) = (a_i(x), b_i(x))^t$, and a family of matrices $\beta_{ij}(x) = \begin{pmatrix} c_{ij}(x) & d_{ij}(x) \\ f_{ij}(x) & e_{ij}(x) \end{pmatrix}$ verifying $c_{ij}(\bullet) = c_{ji}(\bullet)$, $e_{ij}(\bullet) = e_{ji}(\bullet)$ and $f_{ij}(\bullet) = d_{ji}(\bullet)$ such that we write the global energy as:

$$\begin{aligned} Q(y \mid x) &= \sum_{i \in S} \{a_i(x)\delta^*(y_i) - b_i(x)y_i\} \qquad (5) \\ &+ \sum_{(i,j):\langle i,j \rangle} \{c_{ij}(x)\delta^*(y_i)\delta^*(y_j) - d_{ij}(x)y_j\delta^*(y_i) - f_{ij}(x)y_i\delta^*(y_j) + e_{ij}(x)y_iy_j\} \end{aligned}$$

We also have the writing of the canonical parameters:

$$\theta_{1,i}(y^i, x) = a_i(x) + \sum_{j:\{i,j\}} \{c_{ij}(x)\delta^*(y_j) - d_{ij}(x)y_j\}$$
$$\theta_{2,i}(y^i, x) = b_i(x) + \sum_{j:\{i,j\}} \{f_{ij}(x)\delta^*(y_j) - e_{ij}(x)y_j\}$$

with the one to one correspondence:

$$\lambda_i(y^i, x) = \theta_{2,i}(y^i, x) \ , \qquad \gamma_i(y^i, x) = \frac{\theta_{2,i}(y^i, x)}{\theta_{2,i}(y^i, x) + e^{\theta_{1,i}(y^i,x)}} \tag{6}$$

The model is well defined if for all $i,\ x,\ y,\ \lambda_i(y^i, x) > 0,\ \gamma_i(y^i, x) \in]0, 1[$ and if the energy is integrable: $\int_{E^S} \exp Q(y \mid x)\ m^{\otimes S}(dy) < \infty$. We set the following conditions:

(A) (i) For all $i \in S$, for any subset $A \subset S$ and any x, $b_i(x) + \sum_{j \in A} f_{ij}(x) > 0$.

(ii) For all $\{i, j\} \in S$, and any x, $e_{ij}(x) \leq 0$.

Under the set of assumptions (A), any exponential parameter $\lambda_i(y^i, x)$ is strictly positive (ensuring $\gamma_i(y^i, x) \in]0, 1[$); and this also ensures the integrability of the energy (see Proposition 1 of [11]).

Example

Let us specify more precisely the model, i.e. we explicit the families $\alpha_i(x)$ and $\beta_{ij}(x)$. A simple model is to take linear functions for the $\alpha_i(\bullet)$. In order to keep the model parsimonious (minding about future estimation), we choose constants for the $\beta_{ij}(\bullet)$. Moreover, we choose to reflect the dual nature of the variables in the past featuring, introducing both x_l and $\delta^\star(x_l)$ for $x_l \in \partial i^-$, and a kind of pairwise dependence, but keeping the interactions only on pairs of the same feature, and we eliminate the interactions of type $y_i \delta^*(x_l)$. This leads us to consider $\alpha_i(x) = (a_i(x), b_i(x))^t$ with

$$\begin{array}{rl} a_i(x) = & a_i + \sum_{l \in \partial i^-} \alpha_{li} \delta^*(x_l) \\ b_i(x) = & b_i + \sum_{l \in \partial i^-} \varepsilon_{li} x_l \end{array}$$

and $\beta_{ij}(x) = \beta_{ij} = \left(\begin{array}{cc} c_{ij} & 0 \\ 0 & e_{ij} \end{array} \right)$.

Conditionally to the past, the energy's writing becomes
$Q(y \mid x) = \sum_{i \in S} \left(a_i + \sum_{l \in \partial i^-} \alpha_{li} \delta^*(x_l) \right) \delta^*(y_i) - \sum_{i \in S} \left(b_i + \sum_{l \in \partial i^-} \varepsilon_{li} x_l \right) y_i + \sum_{(i,j): \langle i,j \rangle} \left\{ c_{ij} \delta^*(y_i) \delta^*(y_j) + e_{ij} y_i y_j \right\}$.

Conditions (A) become: for all $i \in S$, and any subset $A \subset S$ and any x, $b_i + \sum_{l \in A} \varepsilon_{li} x_l > 0$ and for all $\{i, j\} \in S$, $e_{ij} \leq 0$.

Ergodicity of the dynamics

Let us come back to the general mixed exponential model. We add the following conditions:

(E) (i) For all $i, j \in S$, the functions a_i, b_i, c_{ij}, d_{ij}, e_{ij} are continuous.

(ii) There exists $\eta > 0$, such that for all $x \in E^S$, for all $i, j \in S$, $b_i(x) + \sum_{j \in A} f_{ij}(x) > \eta$.

Proposition 2 *Under assumptions (A) and (E), the mixed auto-exponential dynamics is positive recurrent. Furthermore, we have the following law of large numbers: denoting by μ the chain's invariant measure, for any function φ which is $\mu-$a.s. continuous, and such that $|\varphi| \leq aV_r + b$ with $V_r(y) = \sum_{i \in S} y_i^r$, a and b some constants, and r any positive integer, we have the convergence $\frac{1}{n+1} \sum_{k=0}^{n} \varphi(X_k) \longrightarrow \mu(\varphi)$ a.s.*

Proof :

We follow the proof of Proposition 3 of [10] based on the Lyapunov Stability Criterion, see [7], 6.2.2.

- From assumption (E)(ii), we can bound the energy $Q(y \mid x) \leq \sum_{i \in S} a_i(x) + \sum_{i,j \in S} c_{ij}(x) - \eta \sum_{i \in S} y_i$; thus $\exp Q(y \mid x)$ is $y-$integrable and the chain is strongly Feller.

- Since we assumed that P is strictly positive, the chain is irreducible and there exists at most one invariant probability distribution (see [7], Proposition 6.1.9).

- The third and last step is to provide the Lyapunov function.

If Z is a mixed exponential variable with a probability density function density $f(z) = \gamma\delta(z) + (1-\gamma)\delta^*(z)\lambda \exp(-\lambda z)$ (with $\lambda > 0$), then $E[Z] = E[Z \ \mathbf{1}_{Z>0}] = (1-\gamma)\int_0^\infty \lambda \exp(-\lambda z) dz = (1-\gamma)\lambda^{-1}$.

In a same way, for any positive integer r, $E[Z^r] = E[Z^r \ \mathbf{1}_{Z>0}] = (1-\gamma)\frac{\Gamma(r+1)}{\lambda^r}$. We deduce that $E[X_i(t)^r \mid y^i, x] = (1-\gamma_i(y^i, x))\frac{\Gamma(r+1)}{\theta_{2,i}(y^i,x)^r} \leq \frac{\Gamma(r+1)}{\eta^r}$.

Finally, we take $V(y) = V_r(y) = \sum_{i \in S} y_i^r$ as for the Lyapunov function and we get

$$E[V(X(t)) \mid X(t-1)] \leq \frac{\Gamma(r+1)}{\eta^r} \times n$$

which is finite. Applying the Lyapunov stability criterion achieves the proof.

Coming back to the example, conditions (A)+(E) reduce to: there exists $\eta > 0$, such that for all $x \in E^S$, any subset $A \subset S$, for all $i, j \in S$, $b_i + \sum_{l \in A} \varepsilon_{li} x_l > \eta$ and $e_{ij} \leq 0$.

4 The mixed auto-normal dynamics

4.1 The general case

Let us consider now the Gaussian case; for each i, conditionally to $(X^i(t) = y^i \ , \ X(t-1) = x)$, the distribution of $X_i(t) \in E = \{0\} \cup \mathbb{R}$ is a mixed Gaussian distribution with mean $m_i(y^i, x)$ and variance $\sigma_i^2(y^i, x)$ depending on the past and instantaneous neighbouring values. Then its probability density function with respect to $m(dx) = \delta_0(dx) + \lambda(dx)$ (where λ is the Lebesgue measure on $\mathbb{R}$ now) verifies:

$$\ln f_i(y_i \mid y^i, x) = \langle \theta_i(y^i, x), B_i(y_i) \rangle + D_i(y^i, x) \tag{7}$$

where

$$\theta_i(y^i, x) = \begin{pmatrix} \ln \frac{(1-\gamma_i(y^i,x))g_i(0;y^i,x)}{\gamma_i(y^i,x)} \\ \frac{1}{2\sigma_i^2(y^i,x)} \\ \frac{m_i(y_i,x)}{\sigma_i^2(y^i,x)} \end{pmatrix} \quad , \quad B(y) = \begin{pmatrix} \delta^*(y) \\ -y^2\delta^*(y) \\ y\delta^*(y) \end{pmatrix} \ .$$

with $g(0; \bullet) = \frac{1}{\sqrt{2\pi}\sigma(\bullet)} \exp -\frac{m^2(\bullet)}{2\sigma^2(\bullet)}$.

As previously, $B_i(0) = 0$ for all $i \in S$, and y^i and x hint at $y_{\partial i}$ and $x_{\partial i^-}$.

Then, from Proposition 1, conditionally to $X(t-1) = x$, there exist a family $\alpha_i(x) = (a_i(x), b_i(x), c_i(x))^t$, and a family of matrices $\beta_{ij}(x) = \begin{pmatrix} e_{ij}(x) & d^1_{ij}(x) & d^2_{ij}(x) \\ f^1_{ij}(x) & h_{ij}(x) & d^3_{ij}(x) \\ f^2_{ij}(x) & f^3_{ij}(x) & k_{ij}(x) \end{pmatrix}$

verifying $e_{ij}(\bullet) = e_{ji}(\bullet)$, $h_{ij}(\bullet) = h_{ji}(\bullet)$, $k_{ij}(\bullet) = k_{ji}(\bullet)$,and $f^q_{ij}(\bullet) = d^q_{ji}(\bullet)$ for $q = 1,\ 2,\ 3$ such that we can write the global energy as:

$$Q(y \mid x) = \sum_{i\in S} \left\{a_i(x)\delta^*(y_i) - b_i(x)y_i^2 + c_i(x)y_i\right\} + \sum_{(i,j):\langle i,j\rangle} B_i(y_i)^T \beta_{ij}(x)B_j(y_j)$$

We also have the writing of the canonical parameters:

$$\theta_{1,i}(y^i, x) = a_i(x) + \sum_{j:\{i,j\}} \left\{e_{ij}(x)\delta^*(y_j) - d^1_{ij}(x)y_j^2 + d^2_{ij}(x)y_j\right\}$$
$$\theta_{2,i}(y^i, x) = b_i(x) + \sum_{j:\{i,j\}} \left\{f^1_{ij}(x)\delta^*(y_j) - h_{ij}(x)y_j^2 + d^3_{ij}(x)y_j)\right\}$$
$$\theta_{3,i}(y^i, x) = c_i(x) + \sum_{j:\{i,j\}} \left\{f^2_{ij}(x)\delta^*(y_j) - f^3_{ij}(x)y_j^2 + k_{ij}(x)y_j\right\}$$

and the one to one correspondence:

$$m_i(y^i, x) = \frac{\theta_{3,i}(y^i, x)}{2\theta_{2,i}(y^i, x)}\ ,\ \ \sigma_i^2(y^i, x) = \frac{1}{2\theta_{2,i}(y^i, x)}\ , \tag{8}$$

$$\text{and } \gamma_i(y^i, x) = \left[1 + \sqrt{\pi/\theta_{2,i}(y^i, x)}\ \exp\left\{\frac{\theta^2_{3,i}(y^i, x)}{4\theta_{2,i}(y^i, x)} + \theta_{1,i}(y^i, x)\right\}\right]^{-1} \tag{9}$$

This model is well defined provided that $\theta_{2,i}(y^i, x) > 0$ and the energy is integrable, that is $\int_{E^S} \exp Q(y \mid x)\ m^{\otimes S}(dy) < \infty$. Under these assumptions, it is ergodic.

Proposition 3 *Let us assume that the mixed auto-normal dynamics is well defined. It is positive recurrent and we have the following law of large numbers: denoting by μ the chain's invariant measure, for any function φ which is $\mu-$a.s. continuous, and such that $|\varphi| \leq aV_r + b$ with $V_r(y) = \sum_{i\in S} y_i^r$, a and b some constants, and r any positive integer, we have the convergence $\frac{1}{n+1}\sum_{k=0}^n \varphi(X_k) \longrightarrow \mu(\varphi)$ a.s.*

Proof: we follow the proof of Proposition 2. For Z a mixed Gaussian variable, $E[Z] = E[Z\ \mathbf{1}_{Z\neq 0}] = (1-\gamma)m$ where m is the mean of the Gaussian part of the variable. Analogously, since a Gaussian variable has moments of any order, all the moments $E[Z^r]$ exist, for any positive integer r. Then the mixed auto-normal dynamics is positive recurrent.

Due to the complexity of the general model (8) above, one has to specify his model according to the data, and derive particular conditions to ensure the well definiteness.

4.2 An example

We suppose that the mean is a weighted average of the neighbours and the variance is considered constant, at least not depending on the neighbours. This is a reasonable assumption; for instance, in the context of homogeneous textures, the variability does not depend on the neighbours but is constant.

Furthermore, as for the auto-exponential scheme, we think of considering pairwise interactions between "similar" states, and eliminate interactions such as $x_l^k y_i^2$ or $y_i^k y_j^2$, $k = 1, 2$. Then we consider vectors $\alpha_i(x) = (a_i(x), b_i, c_i(x))^t$ with $a_i(x) = a_i + \sum_{l \in \partial i^-} \alpha_{li}\delta^*(x_l)$ and $c_i(x) = c_i + \sum_{l \in \partial i^-} \varepsilon_{li} x_l$, and matrices $\beta_{ij}(x) = \begin{pmatrix} e_{ij} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & k_{ij} \end{pmatrix}$ which determine the following writings:

$$\begin{aligned} \theta_{1,i}(y^i, x) &= a_i + \sum_{l \in \partial i^-} \alpha_{li}\delta^*(x_l) + \sum_{j:\{i,j\}} e_{ij}\delta^*(y_j) \\ \theta_{2,i}(y^i, x) &= b_i \\ \theta_{3,i}(y^i, x) &= c_i + \sum_{l \in \partial i^-} \varepsilon_{li} x_l + \sum_{j \in \partial i} k_{ij} y_j \end{aligned}$$

and lead to

$$m_i(y^i, x) = \frac{1}{2b_i}\left[c_i + \sum_{l \in \partial i^-} \varepsilon_{li} x_l + \sum_{j \in \partial i} k_{ij} y_j\right] \ , \ \ \sigma_i^2(y^i, x) = \sigma_i^2 = \frac{1}{2b_i} \ , \quad (10)$$

$$\gamma_i(y^i, x) = \left[1 + \sigma_i\sqrt{2\pi}\ \exp\left\{a_i + \sum_{l \in \partial i^-} \alpha_{li}\delta^*(x_l) + \sum_{j:\{i,j\}} e_{ij}\delta^*(y_j) + \frac{m_i^2(y^i, x)}{2\sigma_i^2}\right\}\right]^{-1} \quad (11)$$

Finally, we can write the conditional energy as:

$$\begin{aligned} Q(y \mid x) = \sum_{i \in S} &\left((a_i + \sum_{l \in \partial i^-} \alpha_{li}\delta^*(x_l))\delta^*(y_i) - b_i y_i^2 + (c_i + \sum_{l \in \partial i^-} \varepsilon_{li} x_l) y_i\right) \quad (12) \\ &+ \sum_{(i,j):\langle i,j \rangle} (e_{ij}\delta^*(y_i)\delta^*(y_j) + k_{ij} y_i y_j) \end{aligned}$$

Admissibility and ergodicity
The model (12) is well defined if each variance is strictly positive and if the conditional energy is integrable. The following condition (AG) ensures the well definiteness of the model.

(AG) For all $i \in S,\ b_i > \frac{1}{2}\sum_{j:j \neq i} |k_{ij}|$.

Proposition 4 *Under assumption (AG), the mixed auto-normal dynamics is well defined (and ergodic).*

Proof: First, we note that (AG) trivially implies $b_i > 0$ and each variance is therefore positive.

Let us write the energy as a sum of a discrete and a continuous components: $Q(y \mid x) = Q^*(y \mid x) + Q^c(y \mid x)$ with

$$Q^*(y \mid x) = \sum_{i \in S} \left((a_i + \sum_{l \in \partial i^-} \alpha_{li} \delta^*(x_l)) \delta^*(y_i) \right) + \sum_{(i,j):\langle i,j \rangle} e_{ij} \delta^*(y_i) \delta^*(y_j)$$

$$Q^c(y \mid x) = \sum_{i \in S} \left(-b_i y_i^2 + (c_i + \sum_{l \in \partial i^-} \varepsilon_{li} x_l) y_i \right) + \sum_{(i,j):\langle i,j \rangle} k_{ij} y_i y_j. \quad (13)$$

Obviously, the energy Q is integrable if Q^c is also. But $\exp Q^c$ is proportional to a transition probability $P^c(z_{t-1}, z_t)$ for which $Z_i(t)$ given $(Z(t-1) = z_{t-1}, Z^i(t) = z_t^i)$ is normally distributed with mean and variance

$m_i(y^i, x) = \left(c_i + \sum_{l \in \partial i^-} \varepsilon_{li} z_{t-1,l} + \sum_{,j:\langle i,j \rangle} k_{ij} \right) / 2b_i$, $\sigma_i^2(y^i, x) = \frac{1}{2b_i}$.

Hence the model is well defined if for all $i, j \in S$, $b_i > 0$, $k_{ij} = k_{ji}$ such that the matrix $\Delta = (\Delta_{ij})$ defined by $\Delta_{ii} = 2b_i$ and $\Delta_{ij} = k_{ij}$ is definite positive. From the Geršgorin theorem (see for example [8]), the eigenvalues of Δ belong to $\cup_i D_i$ where $D_i = \{z \in \mathbb{C} : \ |z - 2b_i| \leq \sum_{j:j \neq i} |k_{ij}|\}$. Then condition (AG) implies that Δ's eigenvalues are positive.

Cooperation and parameters influence

We say that a model is spatially cooperative (resp. competitive) if, at each i, $E[X_i(t) \mid y^i, x]$ is non decreasing (resp. non increasing) in each neighbouring value y_j or x_l, and is increasing (resp. decreasing) in at least one. Therefore, we study $E[X_i(t) \mid y^i, x] = (1 - \gamma_i(y^i, x)) E[X_i(t) \mid X_i(t) \neq 0, y^i, x] = (1 - \gamma_i(y^i, x)) \times m_i(y^i, x)$.

Then, the model is spatially cooperative if for all i, j, l, $\varepsilon_{li} \geq 0$, $k_{ij} \geq 0$;

Moreover, if for all i, j, l, $\alpha_{li} \geq 0$, $e_{ij} \geq 0$, then $\gamma_i(y^i, x)$ is decreasing with respect to these parameters.

5 Application

We consider temporal textures. We think of dynamic video contents displayed by natural scenes such as rivers, smoke, or trees. We consider them as textures in motion. Their study is important for many applications, fire or smoke detection for instance. Such motion maps exhibit values of two types, zeros when the motion is absent together with non null continuous values. Therefore, mixed-state auto-models dynamics are good candidates for the modelling. Let us define

$$\mathrm{V}_i^{\perp}(t) = -\frac{\frac{\partial I_i(t)}{\partial t}}{\|\nabla I_i(t)\|} \frac{\nabla I_i(t)}{\|\nabla I_i(t)\|} \quad (14)$$

where $I_i(t)$ is the pixel intensity at time t and location i, $\frac{\partial I_i(t)}{\partial t}$ is approximated by $I_i(t) - I_i(t-1)$, and $\nabla I_i(t)$ the spatial intensity gradient at location $i = (x_i, y_i)$

on the grid (the $i-$th coordinate of the vector $(\frac{\partial I(t)}{\partial x}, \frac{\partial I(t)}{\partial y}))$. First we compute (14) for each image point, using two consecutive frames of the sequence. Then, we take a weighted average of $\mathrm{V}_i^{\perp}(t)$ on a small window centered in i; finally we search for a model for the scalar motion observation at site i and time t described by

$$x_i(t) = \frac{\sum_{j \in W} \mathrm{V}_i^{\perp}(t) \ \|\nabla I_j(t)\|^2}{\max\left(\sum_{j \in W} \|\nabla I_j(t)\|^2, \eta^2\right)} \times \frac{\nabla I_i(t)}{\|\nabla I_i(t)\|} \tag{15}$$

where η^2 is a regularization constant related to noise. We refer the reader to [5] for more details about the construction of (15). Let us note that $x_i(t) \in \mathbb{R}$.

We present in Figure 1 some examples of issued maps and mixed histograms obtained for such dynamic textures. The top row present some natural scenes: grass, steam and river. Below are the corresponding motion textures which are to be modelled; these motion maps are based on normal flow computation (14) and obtained using two consecutive frames of the sequence. The bottom row displays the corresponding histograms. We clearly see the two components of the data, with a peak on zero featuring the discrete null value together with a continuous distribution.

Looking at the histograms of Figure 1, we are interested to fit the mixed auto-normal dynamics model (12). Precisely, $X(t) = \{X_i(t), i \in S\}$ represents the motion texture at time t on $S = \{1, ...N\}$, the spatial lattice of pixel locations, and each $X_i(t)$ is assumed to be a mixed Gaussian variable. We study a causal temporal model, and we assume spatial conditional independence within a motion texture at time t; this means that we do not involve the instantaneous neighbourhood in the model (but $x_i(t)$ and $x_j(t)$ are still spatially correlated). Next, we assume that the temporal neighbourhood of a site i is a nine-point set which consists of i (at time $t-1$), and its 8 nearest neighbours on the grid. Finally we assume translation invariance, and spatial symmetry ($\beta_{ij} = \beta_{ji}^T$) but possible anisotropy between the four directions (horizontal, vertical, diagonal and anti diagonal).

From these assumptions, the mixed Gaussian distribution is characterized by its mean $m_i(y^i, x)$, variance $\sigma_i^2(y^i, x)$, and probability of null value given by

$$m_i(y^i, x) = m_i(x) = \frac{1}{2b}\left[c + \sum_{l \in \partial i^-} \varepsilon_l x_l\right] \ , \ \ \sigma_i^2(y^i, x) = \sigma^2 = \frac{1}{2b} \ \ , \tag{16}$$

$$\gamma_i(y^i, x) = \gamma_i(x) = \left[1 + \sigma\sqrt{2\pi} \ \exp\left\{a + \sum_{l \in \partial i^-} \alpha_l \delta^*(x_l) + \frac{m_i^2(x)}{2\sigma^2}\right\}\right]^{-1} \tag{17}$$

The conditional energy is:

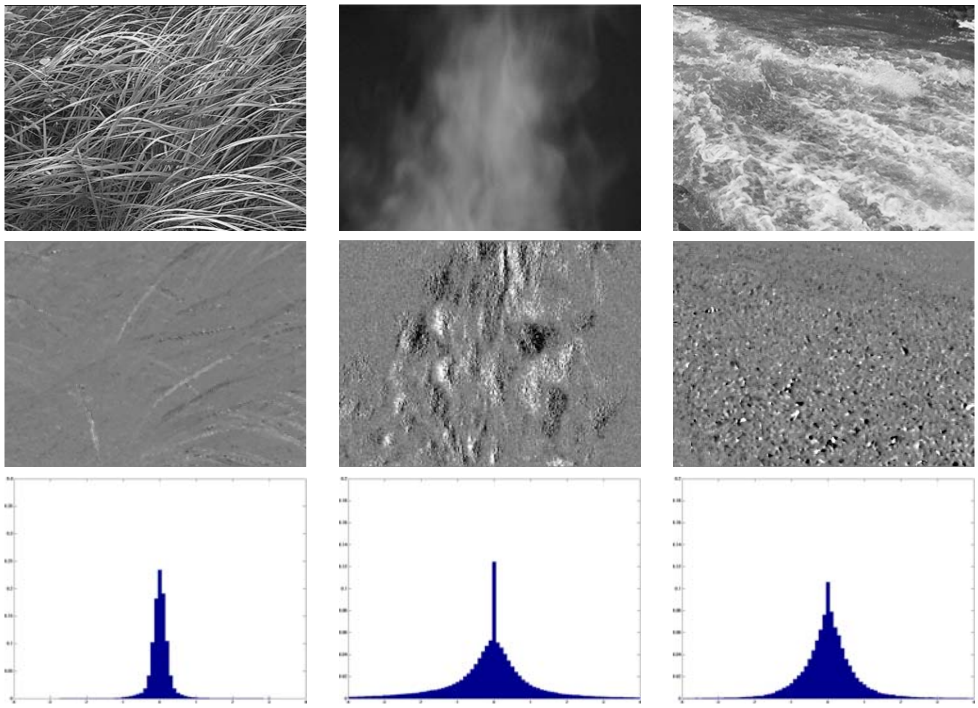

Figure 1: Top row: Sample images from dynamics textures (grass, steam, river). Middle: motion textures. Bottom: motion histograms

$$Q(y \mid x) = \sum_{i \in S} \left((a + \sum_{l \in \partial i^-} \alpha_l \delta^*(x_l))\delta^*(y_i) - by_i^2 + (c + \sum_{l \in \partial i^-} \varepsilon_l x_l) y_i \right). \quad (18)$$

This leads to estimate a set of 13 parameters $\phi = (a, b, c, \{\varepsilon_l\}, \{\alpha_l\})$.

As usual in spatial statistics, since the normalizing constant of the likelihood is intractable, we consider the conditional pseudo-likelihood introduced by Besag ([2]). In the absence of strong dependency, this method has good asymptotic properties, with the same rate of convergence as the maximum likelihood estimator. Let ϕ_0 be the true value of the parameter, and P_0 the associated transition. We suppose ϕ_0 belongs to the interior of Φ, a compact subset of $\mathbb{R}^d$. We define the maximum pseudo-likelihood estimator by

$$\hat{\phi}_T = \arg\min_{\phi \in \Phi} U_T(\phi)$$

$$\text{where } U_T(\phi) = -\frac{1}{T} \sum_{t=1}^{T} \sum_{i \in S} \ln f_i(y_i \mid y^i, x; \phi)$$

General conditions for the consistency and asymptotic normality for this estimator can be found in [10]. Roughly speaking, the goodness of the estimation follows from the temporal ergodicity, joined with the weak dependency of the Markov chain, and the fact that the pseudo-likelihood is a nice and regular functional to identify the model.

Practically, we use a gradient descent technique for the optimization as the derivatives of U are known in closed form. We give in Table 1 estimated parameters for the three motion textures presented in Figure 1. The estimation is made over a temporal window of five motion textures (i.e T=5, obtained from 6 images).

	a	b	c	α_C	ε_c	α_H	ε_H
Grass	-2.3800	3.4294	-0.0336	0.3000	0.0816	0.2670	0.1523
Steam	-1.8900	0.0598	-0.0226	0.1810	-0.000813	0.2510	-0.00288
River	-2.1000	1.0200	-0.0708	0.1270	0.0586	0.2480	0.1740

	α_V	ε_V	α_D	ε_D	α_{AD}	ε_{AD}
Grass	0.3260	0.5707	0.4720	0.5158	0.4790	0.0409
Steam	0.1290	-0.00169	0.1940	-0.00295	0.2080	-0.00281
River	0.2430	0.1230	0.3160	0.1561	0.4350	0.3240

Table 1: Parameter estimates ϕ for three different types of motion textures; for parameters α and ε, the subscripts stand for the position of the neighbour sites (in the past) C: center, H: horizontal, V: vertical, D: diagonal, AD: anti-diagonal.

Let us make some comments on these results. Coming back to the cooperation or competition behaviour, we can see here that for the grass and river motion textures, the mean $m_i(x)$ is an increasing function of x, while for the steam motion texture, the parameters ε are negative, but very close to zero. On the other hand, the probability of null value $\gamma_i(x)$ is in all cases decreasing, since the parameters α are all positive. Therefore, there is a cooperative system influence for grass and river motion textures. It is more difficult to state a cooperative or competitive influence in the case of the steam motion texture, though, since parameters ε are hardly non null, maybe there is an advantage to cooperation.

This model has been used successfully for motion textures recognition, and tracking, see [6] for the complete study; first, a motion texture class (fire, smoke, crowd, grass) is learned and defined through parameter estimation. For instance we give in Table 2 estimates for two different sets of frames of the same motion texture, obtained from grass. This is useful for textures classification. This classification is achieved using a Kullback-Leibler divergence as a distance between two densities, see [6] for full details. For 10 different classes, motion textures were well identified with a rate of about 90. Furthermore, motion tracking can be done, still using the Kullback-Leibler divergence to identify the next location of the motion texture. The results obtained are very encouraging, showing a good performance of the method.

	a	b	c	α_C	ε_c	α_H	ε_H
Grass 1	-2.3800	3.4294	-0.0336	0.3000	0.0816	0.2670	0.1523
Grass 2	-2.2500	3.3300	-0.106	0.382	0.197	0.292	0.0615

	α_V	ε_V	α_D	ε_D	α_{AD}	ε_{AD}
Grass 1	0.3260	0.5707	0.4720	0.5158	0.4790	0.0409
Grass 2	0.3230	0.4050	0.3930	0.3230	0.3940	0.1530

Table 2: Parameter estimates ϕ for two grass motion textures.

References

[1] Allcroft, D. J. and C. A. Glasbey, 2003. A latent Gaussian Markov random-field model for spatiotemporal rainfall disaggregation. *J. Roy. Statist. Soc.* **C** 52 (4), 487–498.

[2] Besag J., 1974. Spatial interactions and the statistical analysis of lattice systems. *J. Roy. Statist. Soc.* **B** 148, 1–36

[3] Bouthemy P., Hardouin C., Piriou G., Yao J.F., 2006. Mixed-states auto-models and motion textures modeling. *Journal of Mathematical Imaging and Vision, **25**(3), 387–402.*

[4] Cressie N., 1991. *Statistics for spatial Data.* New York, Wiley.

[5] Crivelli T., Cernuschi-Frias B., Bouthemy P., Yao J.F., 2006. Mixed-state Markov random fields for motion texture modelling and segmentation. *In proc. of the IEEE Int. Conf. on Image processing, ICIP'06,* 1857–1860.

[6] Crivelli T., Bouthemy P., Cernuschi-Frias B., Yao J.F., 2009. Learning mixed-state Markov models for statistical motion texture tracking. *2nd IEEE International Workshop on Machine Learning for Vision-based Motion Analysis (MLVMA09). Kyoto, Japan. September 2009.*

[7] Duflo M., 1997. *Random iterative models.* Springer.

[8] Golub G.H., Van Loan C.F., 1996. *Matrix computations.* The Johns Hopkins University Press.

[9] Guyon X., 1995. *Random Fields on a Network: Modeling, Statistics, and Applications.* Springer-Verlag, New York

[10] Guyon X., Hardouin C., 2002. Markov chain markov field dynamics: models and statistics. *Statistics*, 2002, Vol. 13, pp. 339–363.

[11] Hardouin C., Yao J.F., 2008. Spatial modelling for mixed-state observations. *Electronic Journal of Statistics*. Vol. 2, 213–233.

[12] Hardouin C. , Yao J.F., 2008. Multi-parameter auto-models with applications to cooperative systems and analysis of mixed-state data. *Biometrika 95(2), 335–349.*

[13] Jagger T., Niu X., 2005. Asymptotic properties of ESTAR models. *Statistica Sinica* 15, 569–595.