Multi-parameter auto-models with applications to cooperative systems and analysis of mixed-state data

By CÉCILE HARDOUIN
SAMOS/Université de Paris 1, 90 rue de Tolbiac, 75634 Paris Cedex, France
hardouin@univ-paris1.fr

And JIAN-FENG YAO
IRMAR/Université de Rennes 1, Campus de Beaulieu, 35042 Rennes Cedex, France
jian-feng.yao@univ-rennes1.fr

Summary

We propose in this paper an extension of Besag's auto-models to exponential families with multi-dimensional parameters. This extension is necessary for the treatment of spatial models like the ones where the conditional distributions are Beta-distributed. A family of cooperative auto-models is proposed. Moreover we apply the result to the analysis of lattice observations which are mixtures of a point mass at the origin and a continuous component.

Some key words: Auto-models; Multi-parameter exponential families; Cooperative systems; Beta conditionals; Mixed-state data.

1 Introduction

We consider a random field $X = \{X_i, i \in S\}$ on a finite set of sites $S = \{1, \ldots, n\}$. For a *site* i, let us note

$$p_i(x_i|\cdot) = p_i(x_i|x_j, j \neq i) ,$$

the probability density of X_i given the event $\{X_j = x_j, j \neq i\}$. An important approach in stochastic modelling consists in specifying the family of all these conditional distributions $\{p_i(x_i|\cdot)\}$, and then to determine a joint distribution P of the system, which is compatible with this family, i.e. the p_i's are exactly the conditional distributions associated to P. This problem was first investigated by Whittle, Bartlett and Besag since years 1960-1970. We refer to the seminal paper of Besag (1974) which presents general results including a useful summary of the previous results of Whittle and Bartlett about the "nearest neighbours systems" (Whittle (1963), Bartlett (1968)). Recently a Markovian approach is proposed in Kaiser & Cressie (2000) (see also Kaiser & al. (2002)) where the commonly used positivity condition on the joint probability distribution is relaxed.

In this paper, we focus on auto-models introduced by Besag (1974). Let us recall that if the joined distribution P is positive everywhere, the Hammersley-Clifford's Theorem gives a characterization of P by an energy $Q(x)$ given by a sum of potentials G deduced from a set of cliques. These auto-models are constructed under two assumptions: first, the dependence between sites is pairwise and secondly, the collection of conditional distributions from the sites belongs to a one-parameter exponential family.

Recent studies in spatial statistics indicate clearly a need for an enlargement of these one dimensional auto-models. An extension of the first condition is proposed in Lee & al. (2001) where the pairwise only dependence is replaced by a multiway dependence but still with one parameter exponential families. We propose in this work an extension to exponential families involving a multi-dimensional parameter, as for instance the Beta distribution, the Gaussian law with mean and variance depending on $(x_j, j \neq i)$ or mixture distributions considered in 4.

In section 2, we give a careful construction of multi-parameter auto-models. As a first application of this general approach, we address the particular problem of building *cooperative* spatial models. Indeed, the admissibility conditions for usual auto-models based on Exponential, Gamma or Poisson conditional distributions imply a spatial competition between the neighbour sites (*cf.* Besag (1974)). In Section 3, we provide a special class of auto-models where the family of conditional distributions are Beta-distributed. This class has the advantage to be able to exhibit spatial cooperation as well as spatial competition according to suitable choice of its parameter values. We notice that Kaiser & Cressie (2000) have considered a Markov field admitting Beta conditional distributions. However their approach is different and for the special class of auto-models under consideration, they have taken a much more constrained parametrization which implies in particular that the model can exhibit spatial cooperation behavior only.

In Section 4, we present auto-models for *mixed-state data*. In several application fields like daily pluviometry data, observations contain two components of different nature. A first part is made with discrete values accounting for some symbolic information and a second part records a continuous measurement. We call such type of observations "mixed states observations". We use the previous result to propose a specific multi-parameter auto-model modelling this kind of data.

Finally we provide in Section 5 some experiments on the analysis of motion measures from video image sequences. Indeed empirical histograms of these motion measurements indicate a composite picture: an important peak appears at the origin accounting for regions where no motion is present, while a large continuous component encompasses actual motion magnitudes in the images.

2 Multi-parameter auto-models

Let us consider a set of sites $S = \{1, \ldots, n\}$, a measurable state space $(E, \mathcal{E}, m)$ (usually a subset of $\mathbb{R}^d$). We denote the layouts space by $\Omega = E^S$, equipped with the σ-algebra and the product measure $(\mathcal{E}^{\otimes S}, \nu := m^{\otimes S})$. For simplicity, we shall consider $\Omega = E^S$, but all the following results hold equally with a more general configuration space $\Omega = \prod_{i \in S} E_i$, where each individual space $(E_i, \mathcal{E}_i)$ is equipped with some measure m_i. We will give such an example in section 2·1.

A random field is specified by a probability distribution μ on Ω, and we will assume that μ has an everywhere positive density P with respect to ν i.e.

$$\mu(dx) = P(x)\nu(dx) \ , \qquad P(x) = Z^{-1} \exp Q(x) \ , \qquad (2{\cdot}1)$$

where Z is a normalizing constant. The Hammersley-Clifford's Theorem gives a characterization of $Q(x)$ as a sum of potentials G deduced from a set of cliques. Moreover, the positivity condition implies that in each site i, the conditional distribution $(X_i | X_j = x_j, j \neq i)$ admits with respect to $m(dx_i)$, a density $p_i(x_i|\cdot)$ which is itself everywhere positive.

Let us set the pairwise-only dependence assumption.

[B1] the cliques involve at most two sites, i.e.,

$$Q(x) = \sum_{i \in S} G_i(x_i) + \sum_{\{i,j\}} G_{ij}(x_i, x_j) \ .$$

We fix a *reference configuration* $\tau = (\tau_i) \in \Omega$. In most cases, $\tau = (0, \ldots, 0)$ but the choice of this reference configuration essentially depends on the state space E ; for instance, if we look for Beta distributions (Section 3), $E = (0, 1)$ and we will take $\tau = (\frac{1}{2}, \ldots, \frac{1}{2})$.

Next we can always assume that, for all i, j and x it happens that

$$G_{ij}(\tau_i, x_j) = G_{ij}(x_i, \tau_j) = G_i(\tau_i) = 0 \ . \tag{2·2}$$

Actually if this condition was not naturally satisfied, we may substitute for $G_{ij}(x_i, x_j)$

$$G_{ij}(x_i, x_j) - G_{ij}(\tau_i, x_j) - G_{ij}(x_i, \tau_j) + G_{ij}(\tau_i, \tau_j) \ ,$$

and make a similar adjustment for $G_i(x_i)$. Note that we thus have $Q(\tau) = 0$ and $Z^{-1} = P(\tau)$.

The second assumption generalizes Besag's one-parameter setting to the multi-parameter case:

$$[\mathbf{B2}]: \quad \log p_i(x_i|\cdot) = \langle A_i(\cdot), B_i(x_i) \rangle + C_i(x_i) + D_i(\cdot) \ , \quad A_i(\cdot) \in \mathbb{R}^d, \quad B_i(x_i) \in \mathbb{R}^d.$$

The main result of the paper is the following.

Theorem 1 *Let us assume that the two conditions* ***[B1]****-****[B2]*** *are satisfied with the normalization* $B_i(\tau_i) = C_i(\tau_i) = 0$ *in* ***[B2]****, as well as the following condition*

[C]: *for all* $i \in S$, $\mathrm{Span}\{B_i(x_i), x_i \in E\} = \mathbb{R}^d$.

Then there exists for $i, j \in S, \ i \neq j$*, a family of* d*-dimensional vectors* $\{\alpha_i\}$ *and a family of* $d \times d$ *matrices* $\{\beta_{ij}\}$ *satisfying* $\beta_{ij}^T = \beta_{ji}$*, such that*

$$A_i(\cdot) = \alpha_i + \sum_{j \neq i} \beta_{ij} B_j(x_j) \ . \tag{2·3}$$

And the potentials are given by

$$G_i(x_i) = \langle \alpha_i, B_i(x_i) \rangle + C_i(x_i) \ , \tag{2·4}$$

$$G_{ij}(x_i, x_j) = B_i^T(x_i) \beta_{ij} B_j(x_j) \ . \tag{2·5}$$

A model satisfying the assumptions of the theorem is called a *multi-parameter auto-model.* The additional condition **[C]** eliminates singular local statistics $B_i(x_i)$. This condition does not exist for the one-parameter case, since it is automatically satisfied, meaning that the B_i's are not identically zero. We will see below that this condition is not restrictive, easily satisfied in most examples.

We give two useful notations. If $x \in \Omega$, for each i we note $\tau_i x$ the layout deduced from x replacing x_i by τ_i, and moreover $x^{(i)} = (x_j, j \neq i)$ the outer configuration of site i.

Proof of Theorem 1. For each i, we have:

$$\begin{aligned}
&Q(x)-Q(\tau_i x)\\
&=G_i(x_i)+\sum_{j\neq i}G_{ij}(x_i,x_j)=\log\frac{p_i(x_i|x^{(i)})}{p_i(\tau_i|x^{(i)})}\\
&=\langle A_i(x^{(i)}),B_i(x_i)\rangle+C_i(x_i)\ .
\end{aligned}$$

We choose $x^{(i)}=\tau^{(i)}=(\tau_j, j\neq i)$, which ensures:

$$G_i(x_i)=\langle A_i(\tau^{(i)}),B_i(x_i)\rangle+C_i(x_i)\ . \tag{2·6}$$

Then we set two indexes $i\neq j$, that is to say for simplification $i=1$ and $j=2$. The previous calculus also leads to

$$\begin{aligned}
&Q(x_1,x_2,\tau_3,\ldots,\tau_n)-Q(\tau_1,x_2,\tau_3,\ldots,\tau_n)\\
&=G_1(x_1)+G_{12}(x_1,x_2)=\langle A_1(x_2,\tau_3,\ldots,\tau_n),B_1(x_1)\rangle+C_1(x_1)\ .
\end{aligned}$$

Therefore

$$G_{12}(x_1,x_2)=\langle \mathcal{A}_1(x_2),B_1(x_1)\rangle\ ,$$

where we put

$$\mathcal{A}_1(x_2)=A_1(x_2,\tau_3,\ldots,\tau_n)-A_1(\tau_2,\tau_3,\ldots,\tau_n)\ .$$

In an analogous way and switching indexes 1 and 2, we finally obtain for all $x_1,x_2\in E$,

$$G_{12}(x_1,x_2)=\langle \mathcal{A}_1(x_2),B_1(x_1)\rangle=\langle \mathcal{A}_2(x_1),B_2(x_2)\rangle\ ,$$

with

$$\mathcal{A}_2(x_1)=A_2(x_1,\tau_3,\ldots,\tau_n)-A_2(\tau_1,\tau_3,\ldots,\tau_n)\ .$$

Since condition **[C]** there exists d vectors $\mathcal{X}_2=(x_2(1),x_2(2),\cdots,x_2(d))$ such that the $d\times d-$matrix $B_2(\mathcal{X}_2)=(B_2(x_2(1)),\ldots,B_2(x_2(d)))$ is regular. Let us note $\mathcal{A}_1(\mathcal{X}_2)=(\mathcal{A}_1(x_2(1)),\ldots,\mathcal{A}_1(x_2(d)))$. We still have:

$$\mathcal{A}_2(x_1)^T B_2(\mathcal{X}_2)=B_1(x_1)^T\mathcal{A}_1(\mathcal{X}_2)\ ,$$

that is

$$\mathcal{A}_2(x_1)^T=B_1(x_1)^T\beta_{12}(\mathcal{X}_2)\ ,\qquad \text{where}\quad \beta_{12}(\mathcal{X}_2)=\mathcal{A}_1(\mathcal{X}_2)[B_2(\mathcal{X}_2)]^{-1}\ .$$

Finally G_{12} can be written as

$$G_{12}(x_1,x_2)=B_1(x_1)^T\beta_{12}(\mathcal{X}_2)B_2(x_2)\ .$$

The left hand side of this equality does not depend on $\mathcal{X}_2$, so $\beta_{12}(\mathcal{X}_2)\equiv\beta_{12}$ is a constant matrix. Besides, as the potential G_{12} keeps invariant by permutation of its indexes, this implies that $\beta_{12}=\beta_{21}^T$. $\square$

In all the following, we will say that a measurable function $H(x)$ defined on Ω is *admissible* if

$$\int_\Omega \exp H(x)\nu(dx) < \infty \ . \tag{2·7}$$

The following proposition is useful, giving a converse to the previous theorem. It also provides a practical way to choose the parameters for a well-defined multi-parameter auto-model.

Proposition 1 *Assume that the energy function Q is defined by **[B1]** with potentials G_i, G_{ij} given in (2·4)-(2·5), and that it is moreover admissible. Then the family of conditional distributions $p_i(x_i|\cdot)$ belong to an exponential family of type **[B2]** whose sufficient statistics $A_i(\cdot)$ satisfy (2·3).*

Proof. We just have to check that the conditional laws of the field with potentials (2·4)-(2·5) are those given by **[B2]** and (2·3). This follows from:

$$\begin{aligned} Q(x) - Q(\tau_i x) &= G_i(x_i) + \sum_{j:j\neq i} G_{ij}(x_i, x_j) \\ &= \langle \alpha_i, B_i(x_i)\rangle + C_i(x_i) + \sum_{j\neq i} B_i(x_i)^T \beta_{ij} B_j(x_j) \\ &= \langle A_i(x^{(i)}), B_i(x_i)\rangle + C_i(x_i) = \log \frac{p_i(x_i|x^{(i)})}{p_i(\tau_i|x^{(i)})} \ . \ \square \end{aligned}$$

2·1 *An illustrative application to a pair of Gaussian and Gamma variables*

In section 3, we will examine multi-parameter auto-models on a lattice. We give here an illustration of Theorem 1 in the case of a couple of variables (X_1, X_2) which respectively take their values in $E_1 = \mathbb{R}^+$ and $E_2 = \mathbb{R}$. This example is interesting since the two state spaces E_1 et E_2 are different. We first require that the conditional distribution of X_1 given $X_2 = x_2$ is a Gamma distribution, and that X_2 given $X_1 = x_1$ is Gaussian. The reference configuration is $\tau = (1, 0)$. In other words, we have according to **[B2]**:

$$\log p_1(x_1|x_2) = \log f_{\theta_1(x_2)}(x_1) = \langle A_1(x_2), B_1(x_1)\rangle - D_1(x_2) \ ,$$
$$\log p_2(x_2|x_1) = \log g_{\theta_2(x_1)}(x_2) = \langle A_2(x_1), B_2(x_2)\rangle - D_2(x_1) \ .$$

Here the sufficient statistics are $B_1(x) = (-x+1, \log x)^T$ and $B_2(x) = (x, x^2)^T$.

Condition **[C]** is trivially satisfied here. Therefore by Theorem 1, there exists two vectors α_1, α_2 and one (2×2)-matrix β such that

$$A_1(x_2) = \alpha_1 + \beta B_2(x_2) \ , \quad A_2(x_1) = \alpha_2 + \beta^T B_1(x_1)$$

The joint density is $P(x_1, x_2) = P(\tau)\exp Q(x_1, x_2)$ with $Q(x_1, x_2) = \langle \alpha_1, B_1(x_1)\rangle + \langle \alpha_2, B_2(x_2)\rangle + B_1(x_1)^T \beta B_2(x_2)$.

The model contains 8 parameters. Explicit conditions on these parameters can be obtained in a straightforward way to ensure admissibility of the energy function Q and consequently the existence of a joint distribution. Therefore we have in a simple way recovered a known-result as presented in Arnold & *al.* (1999), §4.8.

3 A special class of auto-models with Beta conditionals

As pointed out in Besag (1974), several common one-parameter auto-models necessarily imply *spatial competition* between neighbouring sites. For instance, this is the case for the auto-exponential, auto-Poisson schemes as well as for the auto-Gamma family. This competition behaviour is clearly inadequate for many spatial systems where neighbouring sites are indeed cooperative. A common way to get rid of this drawback is to transform the variables into a bounded range. For instance a truncation or projection procedure could be used.

In this section, we propose another way to get cooperative auto-models by using Beta conditionals. This is made possible by the multi-parameter auto-models introduced previously. Moreover, the family of Beta distributions offers a large variety of densities on a determined interval $[a, b]$.

Let us write the density of a Beta distribution on [0,1] with parameters $p, q > 0$ as

$$f_\theta(x) = \kappa(p,q)x^{p-1}(1-x)^{q-1} = \exp\{\langle\theta, B(x)\rangle - \psi(\theta)\}\ ,0 < x < 1$$

with $\theta = (p-1, q-1)^T$, $B(x) = [\log(2x), \log(2(1-x))]^T$ and $\psi(\theta) = (p+q-2)\log 2 + \log\kappa(p,q)$. We recall that $\kappa(p,q) = \Gamma(p+q)/[\Gamma(p)\Gamma(q)]$. Here the reference state is $\tau = \frac{1}{2}$ ensuring $B(\tau) = 0$. Throughout this section we denote the two base functions by $u(x) = \log(2x)$ and $v(x) = \log[2(1-x)]$.

We now consider a random field X with such Beta conditional distributions. Clearly, Condition **[C]** is satisfied. From Theorem 1, there exists for $i, j \in S$ and $i \neq j$ some vectors $\alpha_i = (a_i, b_i)^T \in \mathbb{R}^2$ and (2×2)-matrices $\beta_{ij} = \begin{pmatrix} c_{ij} & d_{ij} \\ d^*_{ij} & e_{ij} \end{pmatrix}$ verifying $\beta_{ij} = \beta_{ji}^T$, such that

$$A_i(\cdot) = \alpha_i + \sum_{j\neq i}\beta_{ij}B(x_j) = \alpha_i + \sum_{j\neq i}\beta_{ij}\begin{pmatrix} u(x_j) \\ v(x_j)\end{pmatrix}\ .$$

The energy function Q can be written as

$$Q(x_1,\ldots,x_n) = \sum_{i\in S}\langle\alpha_i, B(x_i)\rangle + \sum_{\{i,j\}} B(x_i)^T\beta_{ij}B(x_j)\ .$$

Finally the reference configuration is $\tau = (\frac{1}{2},\ldots,\frac{1}{2})$ satisfying $Q(\tau) = 0$.

This model is well-defined as soon as its energy function Q is admissible. We first note that the canonical parameters of the conditional Beta distributions are given by

$$A_i(\cdot) = \begin{pmatrix} a_i + \sum_{j\neq i}\{c_{ij}u(x_j) + d_{ij}v(x_j)\} \\ b_i + \sum_{j\neq i}\left\{d^*_{ij}u(x_j) + e_{ij}v(x_j)\right\}\end{pmatrix}\ . \qquad (3\cdot1)$$

It is then necessary that for all i and all outer layout $x^i \in (0,1)^{n-1}$,

$$1 + a_i + \sum_{j\neq i}\{c_{ij}u(x_j) + d_{ij}v(x_j)\} > 0\ , \qquad (3\cdot2)$$

and

$$1 + b_i + \sum_{j\neq i}\left\{d^*_{ij}u(x_j) + e_{ij}v(x_j)\right\} > 0\ . \qquad (3\cdot3)$$

Let us consider the first inequality (3·2). If x_j tends to 0_+ or 1_-, necessarily it follows that $c_{ij} \leq 0$ and $d_{ij} \leq 0$. Assume for a moment $c_{ij} + d_{ij} \neq 0$. Then by studying the function defined in this inequality, we see that it is convex and reaching its minimum at $\xi = (\xi_j)$ where $\xi_j = c_{ij}/(c_{ij} + d_{ij})$ for $j \neq i$. Finally the first inequality holds as soon as

$$h_i := 1 + a_i + \sum_{j \neq i} \left\{ c_{ij} \log \frac{2c_{ij}}{c_{ij} + d_{ij}} + d_{ij} \log \frac{2d_{ij}}{c_{ij} + d_{ij}} \right\} > 0 \, . \tag{3·4}$$

If indeed $c_{ij} = d_{ij} = 0$, the situation is trivial and the same conclusion holds with the agreement $0 \log \frac{0}{0} = 0 \log 0 = 0$. The case of the second inequality (3·3) is similar by substituting (b_i, d^*_{ij}, e_{ij}) for (a_i, c_{ij}, d_{ij}) with this time $d^*_{ij} \leq 0$ and $e_{ij} \leq 0$, and we obtain the following condition similar to (3·4)

$$k_i := 1 + b_i + \sum_{j \neq i} \left\{ d^*_{ij} \log \frac{2d^*_{ij}}{d^*_{ij} + e_{ij}} + e_{ij} \log \frac{2e_{ij}}{d^*_{ij} + e_{ij}} \right\} > 0 \, . \tag{3·5}$$

Proposition 2 *With (h_i) and (k_i) defined in (3·4)-(3·5), the family of conditional distributions $\{p_i(x_i|\cdot)\, , i \in S\}$ is everywhere well-defined under the following assumptions*

*[**T1**] :*   *(i)*   *for all* $\{i,j\}$, $\beta_{ij} = \begin{pmatrix} c_{ij} & d_{ij} \\ d^*_{ij} & e_{ij} \end{pmatrix} \leq \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$, *element-wisely*

(ii)   *for all* i, $h_i > 0$ *and* $k_i > 0$.

Interestingly enough, these conditions also ensure the admissibility of the energy function Q.

Proposition 3 *Under Conditions [**T1**], the energy function Q is admissible and consequently, the auto-model with Beta conditionals is well-defined.*

Proof. We have

$$\begin{aligned} Q(x_1, \ldots, x_n) &= \sum_{i \in S} u(x_i) \left[a_i + \sum_{j \neq i} \{ c_{ij} u(x_j) + d_{ij} v(x_j) \} \right] \\ &+ \sum_{i \in S} v(x_i) \left[b_i + \sum_{j \neq i} \left\{ d^*_{ij} u(x_j) + e_{ij} v(x_j) \right\} \right] \\ &\leq \sum_{i \in S} \{ (h_i - 1) u(x_i) + (k_i - 1) v(x_i) \} \, . \end{aligned}$$

Clearly $\exp Q(x)$ is integrable on $(0,1)^n$. □

In practice one may, instead of the conditions [**T1**], prefer the following

[**T2**] :   (i)   for all $\{i,j\}$, $\beta_{ij} = \begin{pmatrix} c_{ij} & d_{ij} \\ d^*_{ij} & e_{ij} \end{pmatrix} \leq \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$,

(ii)   for all i, $1 + a_i + (\log 2) \sum_{j \neq i} \{ c_{ij} + d_{ij} \} > 0$, and

$$1 + b_i + (\log 2) \sum_{j \neq i} \left\{ d^*_{ij} + e_{ij} \right\} > 0 \, .$$

This set of stronger conditions has the merit to be much simpler to use.

3·1 *Spatial cooperation versus spatial competition*

We now examine the spatial competition or cooperation behavior of this model. At each site i, the mean of the conditional distribution $p_i(x_i|\cdot)$ is

$$\mathbb{E}(X_i|\cdot) = \frac{1 + A_{i,1}(\cdot)}{2 + A_{i,1}(\cdot) + A_{i,2}(\cdot)} \ .$$

This conditional mean increases with $A_{i,1}(\cdot)$ and decreases with $A_{i,2}(\cdot)$. Besides the model is spatially cooperative if at each i the above conditional mean increases with each neighbouring value $x_j, \ j \neq i$. This is possible by requiring for all i, j, $c_{ij} = e_{ij} = 0$.

Alternatively, if we adopt the constraints $d_{ij} = d^*_{ij} = 0$ for all pairs $i \neq j$, the above conditional mean becomes a decreasing function on any of its neighbouring value x_j. There is then a spatial competition between neighbouring sites.

To conclude the discussion about this special class, we compare the above results to those of Kaiser & Cressie (2000). Following a different approach the authors have also considered an auto-model with Beta conditional distributions. More precisely, the proposed model - Eq. (16) of the reference-, corresponds to the following special specification:

$$c_{ij} = e_{ij} = 0, \quad d_{ij} = d^*_{ij} \ .$$

This provides an auto-model with spatial cooperation as proved by the authors and as mentioned above. However the additional constraint $d_{ij} = d^*_{ij}$ is by no means necessary. It is also worth noticing that the assumed conditions for the admissibility of the energy function $Q(x)$ corresponds to the set of conditions **[T2]** above which is sufficient but not necessary.

3·2 *A cooperative model with the four-nearest-neighbours system*

Consider the four-nearest-neighbours system on a two-dimensional lattice $S = [1, M] \times [1, N]$: each site $i \in S$ has the four neighbours denoted as $\{i_e = i + (1, 0), i_o = i - (1, 0), i_n = i + (0, 1), i_s = i - (0, 1)\}$ (with obvious correction on the boundary). We next assume spatial stationarity but allow possible anisotropy between the horizontal and vertical directions. Also the systeme is required to be spatially cooperative so that we assume $c_{ij} = e_{ij} = 0$. Under all these considerations and by the previous results, there exist a vector $\alpha = (a, b)$ and two 2×2 matrices $\beta^{(1)}$ and $\beta^{(2)}$ such that $\forall i, \ \alpha_i = \alpha$, and for $\forall \{i, j\}, \ \beta_{ij} = 0$ unless i and j are neighbours where

$$\beta_{i,i_e} = \beta^{(1)} = \begin{pmatrix} 0 & d_1 \\ d^*_1 & 0 \end{pmatrix} = \beta^T_{i_o,i}, \quad \beta_{i,i_n} = \beta^{(2)} = \begin{pmatrix} 0 & d_2 \\ d^*_2 & 0 \end{pmatrix} = \beta^T_{i_s,i},$$

The model involves 6 parameters $(a, b, d_1, d^*_1, d_2, d^*_2)$. The admissibility conditions **[T1]** become

$$d_k \leq 0, \quad d^*_k \leq 0, \ k = 1, 2 \, ; \quad 1 + a + (d_1 + d_2 + d^*_1 + d^*_2) \log 2 > 0 \, ; \ 1 + b + (d_1 + d_2 + d^*_1 + d^*_2) \log 2 > 0 \, .$$

The associated local conditional distributions are Beta-distributed with canonical parameters

$$A_i(\cdot) = \begin{pmatrix} a + d_1 v(x_{i_e}) + d^*_1 v(x_{i_o}) + d_2 v(x_{i_n}) + d^*_2 v(x_{i_s}) \\ b + d^*_1 u(x_{i_e}) + d_1 u(x_{i_o}) + d^*_2 u(x_{i_n}) + d_2 u(x_{i_s}) \end{pmatrix} \ . \qquad (3·6)$$

Note that these parameters can be easily estimated by the well-known pseudo-likelihood method.

4 Auto-models for mixed-state data

It is frequent to get measurements that can present continuous values during some periods and discrete values at other times. For example, daily pluviometry time series at a given site records many zeros when the rain is absent, followed by periods with positive rainfall values (Allcroft and Glasbey (2003)). Another example arises in the motion analysis problem from image sequences (Section 5). It then raises the question to find accurate models for this type of data, where the variables belong to what we shall call a *mixed-state space.* If the data are collected from some lattice, then we will use the previous multi-parameter auto-models involving observations of both continuous and discrete type.

Let us first define a simple random variable on the mixed-state space $E = \{0\} \cup (0, \infty)$. This space 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*, and the associated distribution a *mixed-state distribution.* Such a variable arises from the following construction: with probability $\gamma \in (0, 1)$ we set $X = 0$, and with probability $1 - \gamma$, X is positive, continuous having on $(0, \infty)$ a density belonging to a s-dimensional exponential family

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

where the sufficient statistics T is defined with $T(0) = 0$ (a priori, T is not defined at the origin but in practice, usual sufficient statistics can be extended by continuity). Define the indicator function $\delta(x) = \mathbb{I}_{\{0\}}(x)$ and set $\delta^*(x) = 1 - \delta(x)$. The variable X has then the following density function w.r.t. $m(dx)$

$$\begin{aligned} f_\theta(x) &= \gamma\delta(x) + (1-\gamma)\delta^*(x) g_\xi(x) \\ &= \gamma \exp\{\delta^*(x) \ln \frac{(1-\gamma)H(\xi)}{\gamma} + \langle \xi, T(x)\rangle\} \\ &= Z^{-1}(\theta) \exp\langle \theta, B(x)\rangle \end{aligned} \tag{4·1}$$

where we have set

$$\theta = (\theta_1, \theta_2)^T = \left(\log \frac{(1-\gamma)H(\xi)}{\gamma}, \xi\right)^T \quad , \quad B(x) = (\delta^*(x), T(x)^T)^T \ .$$

In other words, X also belongs to an exponential family with parametric dimension $1 + s$. Note that the use of δ^* ensures the normalization equality $B(0) = 0$ used in Condition [**B2**]. Moreover the original parameters ξ and γ can be recovered from the natural parameter θ by

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

Coming back to spatial data on a lattice and by using the general theory of multi-parameter auto-models, we are able to construct auto-models for mixed-state variables. We start by assuming that the family of conditional distributions $p_i(x_i|\cdot)$ belongs to the family of mixed-state distribution given in (4·1). In other words, we assume that

$$\log p_i(x_i|\cdot) = \langle \theta_i(\cdot), B(x_i)\rangle - D_i(\cdot) \tag{4·2}$$

Based on Theorem 1, we know that there are a family of $(s+1)$-dimensional vectors $\{\alpha_i\}$ and a family of $(s+1)\times(s+1)$ matrices $\{\beta_{ij}\}$ satisfying $\beta_{ij} = \beta_{ji}^T$ such that

$$\theta_i(\cdot) = \alpha_i + \sum_{j\neq i} \beta_{ij} B(x_j) \ . \tag{4·3}$$

The associated energy function is given by

$$Q(x) = \sum_{i\in S} \langle \alpha_i, B(x_i)\rangle + + \sum_{\{i,j\}} B(x_i)^T \beta_{ij} B(x_j) \ . \tag{4·4}$$

Note that for each specific family g_ξ, we will have to look for the conditions ensuring the admissibility of the energy function $Q(x)$.

5 An application to motion analysis from video image sequences

5·1 *Motion measurements from video sequences*

Motion computation and analysis are of central importance in image analysis. Let $\{I_i(t)\}$ be an image sequence where $i = (i_1, i_2) \in S$ denotes the pixel locations and $t = 1, \ldots, T$ time instants in the sequence. Roughly speaking, a motion map at time t, $X(t) = \{X_i(t)\} = \{\|v_i(t)\|\}$ is defined as the norm of the underlying motion field $\{v_i(t)\}$ which is estimated by a "regularized" minimization of the sum of squares $\sum_i [I_{i+v_i(t)}(t+1) - I_i(t)]^2$. Usually some local smoothing procedures are needed to get a more robust motion map and we refer to Fablet and Bouthemy (2003) for details of these computations.

Here we consider video sequences of natural scenes. Figure 1 displays three sample images from each of two sequences involving a moving escalator and a tree under wind respectively. The corresponding motion maps are displayed in Figure 2. Next, sample histograms from these motion maps are presented in Figure 3. As a matter of fact, these histograms present a composite picture. An important peak appears at the origin accounting for regions where no motion is present, while a large continuous component encompasses actual motion magnitudes in the images.

5·2 *The auto-model with mixed positive Gaussian distributions*

We follow the general construction of mixed states auto-models. First, we call W a *mixed positive Gaussian variable* if g_ξ is the density of the modulus of a zero-mean Gaussian distribution with variance σ^2: $g_\xi(x) = 2(2\pi\sigma^2)^{-1/2} \exp\{-\frac{1}{2\sigma^2}x^2\}$. Then W has the following density w.r.t. $m(dx)$:

$$f_\theta(x) = \gamma\delta(x) + (1-\gamma)\delta^*(x) g_\xi(x) = \exp\left[\langle \theta, B(x)\rangle + \log\gamma\right] . \tag{5·1}$$

Here we have set $\xi = (2\sigma^2)^{-1}$ and consequently

$$\theta = (\theta_1, \theta_2)^T = \left(\log\frac{(1-\gamma)g_\xi(0)}{\gamma}, \xi\right)^T \ , \quad B(x) = (\delta^*(x), -x^2)^T \ . \tag{5·2}$$

To construct auto-models for the motion maps observations $\{X_i(t)\}$, we assume that the family of conditional distributions $p_i(x_i|\cdot)$ belongs to the family of mixed positive Gaussian distribution $f_{\theta_i(\cdot)}(x_i)$ given in (5·1). By Theorem 1, there are a family of vectors $\alpha_i = (a_i, b_i) \in \mathbb{R}^2$ and a family of 2×2 matrices $\{\beta_{ij}\}$ satisfying $\beta_{ij} = \beta_{ji}^T$, such that

$$\theta_i(\cdot) = \alpha_i + \sum_{j \neq i} \beta_{ij} B(x_j) \ . \tag{5·3}$$

Moreover the associated energy function is given by

$$Q(x_1, \ldots, x_n) = \sum_{i \in S} \left[a_i \delta^*(x_i) - b_i x_i^2 \right] + \sum_{\{i,j\}} (\delta^*(x_i), -x_i^2) \beta_{ij} (\delta^*(x_j), -x_j^2)^T \ . \tag{5·4}$$

We now consider the context of the four-nearest-neighbours system analoguous to the one defined in Section 3·2 with spatial stationarity and possible anisotropy between the horizontal and vertical directions. There exist then a vector $\alpha = (a, b)$ and two 2×2 matrices $\beta^{(1)}$ and $\beta^{(2)}$ such that $\forall i,\ \alpha_i = \alpha$, and for $\forall \{i,j\},\ \beta_{ij} = 0$ unless i and j are neighbours where

$$\beta_{i,i_e} = \beta^{(1)} = \begin{pmatrix} c_1 & d_1 \\ d_1^* & e_1 \end{pmatrix} = \beta_{i_o,i}^T, \quad \beta_{i,i_n} = \beta^{(2)} = \begin{pmatrix} c_2 & d_2 \\ d_2^* & e_2 \end{pmatrix} = \beta_{i_s,i}^T \ .$$

Moreover we ask for spatial cooperation and thus we need to further constrain the parameters $d_k,\ d_k^*$ and $e_k,\ k = 1, 2$ to be zero. The resulting auto-model has four parameters $\phi = (a, b, c_1, c_2)$ and is well-defined (admissible) under an unique condition: $b > 0$.

Furthermore the parameter ϕ can be estimated by maximizing the pseudo-likelihood (in fact its logarithm)

$$L(x; \phi) = \sum_{i \in S} \log p_i(x_i | x_j,\ j \neq i) \ . \tag{5·5}$$

This method has good consistency properties for classical one-parameter auto-models, see e.g. Besag (1977), Guyon (1995).

5·3 *Experiments*

The experiments are conducted in order to evaluate whether the above model can correctly account for a fundamental characteristic of an homogeneous texture, namely spatial isotropy or spatial anisotropy. According to the four-nearest-neighbours Gaussian mixed-state model of Section 5·2, the spatial isotropy occurs if (and only if) $c_1 = c_2$. We then have fitted this model to the motion maps displayed in Figure 2.

First we consider motions from trees (bottom row of the figure). A typical set of parameter estimates is $\hat{\phi} = (\hat{a}, \hat{b}, \hat{c}_1, \hat{c}_2) = (-5.805, 3.044, 3.057, 2.954)$. The parameters c_1 and c_2 are almost identical (with regard to standard deviations of these estimates computed at other time instants of a same tree sequence). Therefore, the believed spatial isotropy for these motions is well reflected here by the equality between the parameters $\{c_k\}$.

Next we consider the motion maps from a moving escalator (top row of Figure 2). Since the motion is a vertical one, we have clearly anisotropic motion. A typical set of parameter estimates is $\hat{\phi} = (\hat{a}, \hat{b}, \hat{c}_1, \hat{c}_2) = (-6.512, 0.320, 2.192, 3.598)$. Clearly, the difference between c_1 and c_2 is significant. Therefore, the fitted model seems able to reflect the spatial anisotropy of the considered motion.

6 Conclusion

In this paper we have proposed an extension of Besag's auto-models to the situations where the local conditional distributions belong to some multi-parameter exponential families. This extension is fundamental for the treatment of useful spatial models like the one exposed in Section 3 with Beta conditionals. This model in particular can exhibit spatial cooperation (as well as spatial competition). Another interesting application we have proposed is on the modelling of mixed-state data where the distributions are mixtures of discrete and continuous components. We look further to work on the development of a satisfactory estimation theory both for the pseudo-likelihood estimator and the more difficult likelihood estimator.

Acknowledgment. The authors thank Gwënaelle Piriou (INRIA/IRISA, Rennes) for providing experiments of Section 5.

References

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

B. C. Arnold, E. Castillo et J. M. Sarabia, 1999. *Conditional Specification of Statistical Models.* Springer-Verlag, New York

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

J. Besag, 1977. Efficiency of pseudolikelihood estimation for simple Gaussian fields. *Biometrika* **64**, no. 3, 616–618

M. S. Bartlett, 1968. A further note on nearest neighbour models. *J. Roy. Statist. Soc.* **A** 131, 579-580

Fablet, R. and P. Bouthemy, 2003. Motion recognition using non parametric image motion models estimated from temporal and multi-scale co-occurrence statistics. *IEEE Trans. on Pattern Analysis and Machine Intelligence* **25**(122), 1619–1624.

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

M. S. Kaiser and N. Cressie, 2000. The construction of multivariate distributions from Markov random fields. *J. Multivariate Anal.* **73**, 199-220

M. S. Kaiser, N. Cressie and J. Lee, 2002. Spatial mixture models based on exponential family conditional distributions. *Statistica Sinica* **12**, 449-474

J. Lee, M. S. Kaiser and N. Cressie, 2001. Multiway dependence in exponential family conditional distributions. *J. Multivariate Anal.* **79**, 171-190

P. Whittle, 1963. Stochastic processes in several dimensions. *Bull. Inst. Statist. Inst.* **40**, 974-994

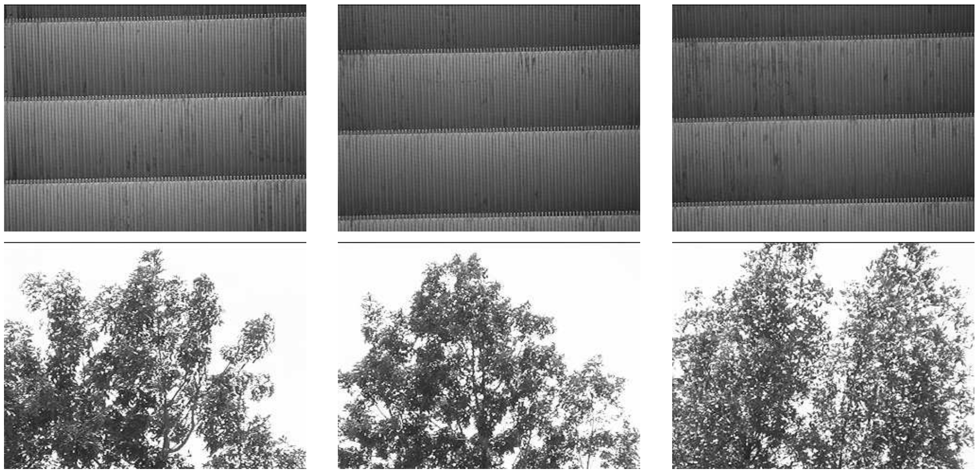

Figure 1: Sample images from two videos. Top row: a moving escalator; bottom row: trees.

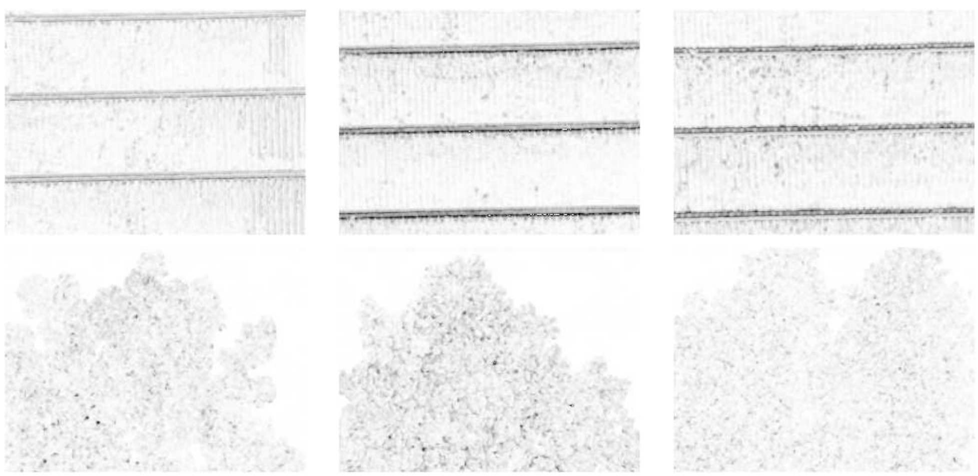

Figure 2: Sample motion measures $\{X_i(t)\}$ from the videos of Figure 1. Top row: a moving escalator; bottom row: a tree (white=0; black=maximum value).

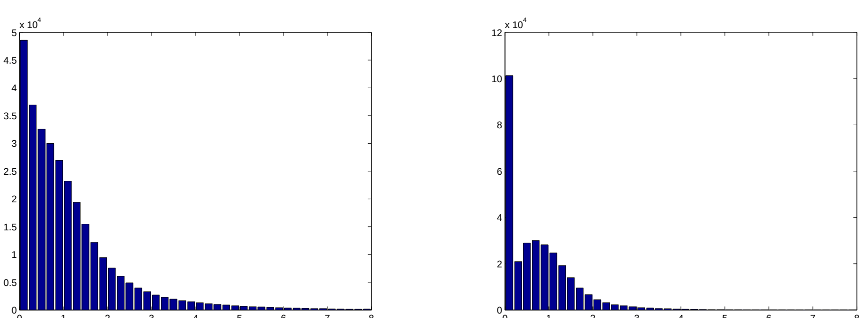

Figure 3: Sample histograms of motion measures $\{X_i(t)\}$ of Figure 2.