HAL
open science

Markov chain Markov field dynamics: models and statistics

Xavier Guyon, Cécile Hardouin

To cite this version:

Xavier Guyon, Cécile Hardouin. Markov chain Markov field dynamics: models and statistics. Statistics A Journal of Theoretical and Applied Statistics, 2002, 36 (4), pp.339 - 363. 10.1080/02331880213192 . hal-00110370

HAL Id: hal-00110370

https://hal.science/hal-00110370v1

Submitted on 5 Oct 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Markov Chain Markov Field Dynamics: Models and Statistics

X. GUYON and C. HARDOUIN

SAMOS - Université Paris 1[1]

Abstract

This study deals with time dynamics of Markov fields defined on a finite set of sites with state space E, focussing on Markov Chain Markov Field (MCMF) evolution. Such a model is characterized by two families of potentials: the instantaneous interaction potentials, and the time delay potentials. Four models are specified: auto-exponential dynamics ($E = \mathbb{R}^+$), auto-normal dynamics ($E = \mathbb{R}$), auto-Poissonian dynamics ($E = \mathbb{N}$) and auto-logistic dynamics (E qualitative and finite). Sufficient conditions ensuring ergodicity and strong law of large numbers are given by using a Lyapunov criterion of stability, and the conditional pseudo-likelihood statistics are summarized. We discuss the identification procedure of the two Markovian graphs and look for validation tests using martingale central limit theorems. An application to meteorological data illustrates such a modelling.

Key words. Markov Field; Markov Chain dynamics; Auto-model; Lyapunov stability criterion; Martingale CLT theorem; Model diagnostic.

AMS Classification Numbers: 62M40, 62M05, 62E20.

1 Introduction

The purpose of this paper is to study time dynamics of Markov fields defined on a measurable state space E and a **finite** set of sites S. We present a *semi causal* parametric model called **MCMF**, Markov Chain of Markov Field, defined as follows: $X = (X(t),\ t \in \mathbb{N}^*)$ is a Markov

[1] **e-mail**: hardouin[guyon]@univ-paris1.fr

Running head : Markov chain of Markov field

Mail address for correspondence : Cécile Hardouin, SAMOS, Université Paris 1, 90 rue de Tolbiac, 75634-Paris Cedex 13, France

chain on E^S and $X(t) = (X_i(t),\ i \in S)$ is a Markov field on E^S conditionally to the past. We study the properties of this model characterized by two families of potentials: instantaneous interaction potentials and time-delay potentials.

Space time modelling has been considered in the literature by Preston (1974; [28]) for birth and death processes, Durrett (1995; [14]) for particles systems, Künsch (1984; [25]) and Koslov and Vasilyev (1980; [24]) for the study of reversible and synchronous dynamics, Pfeifer and Deutsch (1980; [26], [27]) for ARMA models in space. Among statistical studies and their applications, some important contributions are those of Pfeifer and Deutsch (1980; [26], [27]), Besag (1974, 1977; [8], [9]), Keiding (1975; [23]) for birth and death process, Bennett and Haining (1985; [7]) for geographical data, Chadoeuf *et al.* (1992; [11]) for plant epidemiology. The asymptotic properties of these methods are obtained in a standard way (see Amemiya (1985; [1]), Dacunha-Castelle and Duflo (1986; [12]), Guyon (1995; [18]), Bayomog *et al.* (1994; [6]).

The aim of this paper is to specify the structure of conditional Gibbs models according to the kind of state space considered as well as to study ergodicity, identification, estimation and validation of such models. For simplicity, we consider only time homogeneous dynamics (but spatial stationarity is not assumed) and the one step Markovian property. Our results can be generalized to a larger dependency with respect to the past as well as to time inhomogeneous chains.

After the description of the probability transition $P(x, y)$ of the model in Section 2, we study in Section 3 some properties about time reversibility, invariant probability measure, and marginal distributions and we show that the MCMF dynamics is equivalent to a time homogeneous space$\times$time non causal Markov field representation.

Four examples are depicted in Section 4: the auto-normal dynamics ($E = \mathbb{R}$), the auto-exponential dynamics ($E = \mathbb{R}_+$), the auto-Poissonian dynamics ($E = \mathbb{N}$), and finally, the auto-discrete dynamics (E qualitative and finite). For each specification, we give sufficient conditions ensuring the ergodicity of the chain with the help of a Lyapunov criterion. Further, we specify the results for weak-reversible models, *i.e.* models such that the reverse (in time) transition $Q(x, y)$ belongs to the same auto-model family as $P(x, y)$.

In Section 5, we summarize the main results on Conditional Pseudo-Likelihood (CPL) estima-

tion: consistency and asymptotic (in time) normality of the estimator, tests of nested hypotheses based on the CPL-ratio.

Section 6 is devoted to the identification problem *i.e.* to the determination of the dependency graphs $\mathbf{G} = \{\mathcal{G}, \mathcal{G}_-\}$ related to the MCMF dynamics. Graph $\mathcal{G}$ is undirected and determines the instantaneous dependency, while graph $\mathcal{G}_-$ is directed and associated to the time-delay dependency. In the Gaussian case, the characterization of $\mathbf{G}$ is given through the partial-autocorrelations; for non-Gaussian models, we suggest a stepwise procedure based on the CPL.

In Section 7, we propose two validation tests based on CLT for martingales. We conclude this paper in Section 8 with the study of a real set of meteorological data to which we fit an auto-logistic MCMF model. The data consist of daily pluviometric measures on a network of 16 stations in the Mekong Delta (Vietnam) during a three-month period (Tang 1991; [31]). The auto-logistic model allows for information to be gathered on spatial and temporal dependencies, and the forecasting is relatively accurate for some sites. This study has to be considered as a first illustrative attempt; there is no doubt that it will have to be refined in a precise investigation and then compared to other space time models, like hidden Markov chains or threshold models. It will, of course, be interesting to take into account the dual feature of such data (it is not raining or the precipitation level is observed in $\mathbb{R}^{+*}$) and integrate it in a Gibbs model. Such a generalization and comparisons to other models are subjected to another study actually in progress.

2 Notations and description of the MCMF model

Let $S = \{1, 2, ..., n\}$ be a finite set of sites. We call X an MCMF model if X is a Markov Chain of Markov Fields (conditionally to the past). The latter are defined on a measurable state space $(E, \mathcal{E})$ equipped with a positive σ-finite measure ν (ν is usually the counting measure if E is discrete, and the Lebesgue one if $E \subseteq \mathbb{R}^p$). The product space is $(\Omega, \mathcal{O}) = (E^S, \mathcal{E}^{\otimes S})$ with the product measure $\nu^S = \nu^{\otimes S}$. We shall consider the case $\Omega = E^S$ but all the following still holds for $\Omega = \Pi_{i \in S} E_i$, with measures ν_i on measurable spaces $(E_i, \mathcal{E}_i)$, $i \in S$.

We use the following notations. Let A be a subset of S; we denote $x_A = (x_i, i \in A)$ and $x^A = (x_j, j \in S \setminus A)$. Let $\mathcal{G}$ be a symmetric graph on S without loop: $\langle i, j \rangle$ denotes that i and

j are neighbours (in particular, $i \neq j$). The neighbourhood of A is $\partial A = \{j \in S : j \notin A \ s.t. \ \exists \ i \in A \ with \ \langle i, j\rangle\}$. We shall write $x^i = x^{\{i\}}$ and $\partial i = \partial\{i\}$. With these notations, our MCMF model X is the following:

•$X = (X(t),\, t \in \mathbb{N}^*)$ is a homogeneous Markov chain on Ω,

• • $X(t) = (X_i(t),\, i \in S)$ is, conditionally to $X(t-1)$, a Markov field on S, with $X_i(t) \in E$.

We suppose that the transition probability measure of X has a density $P(x, y)$ w.r.t ν^S.

If E is discrete, $P(x, y) = P(X(t) = y|X(t-1) = x)$. We look at models for which $P(x, y)$ is defined by an $(x) - a.s.$ admissible conditional energy $U(y|x)$, *i.e.* such that

$$P(x, y) = Z^{-1}(x) \exp U(y|x) \quad \text{and} \quad Z(x) = \int_\Omega \exp U(y|x)\nu^S(dy) < \infty \tag{1}$$

($Z(x) = \sum_\Omega \exp U(y|x) < \infty$ if E is discrete). Let $X(t-1) = x$. From the Moebius inversion formula (Besag 1974, [8]; Prum 1986, [29]; Guyon 1995, [18]), there exists a minimal family $\mathcal{C}_x$ of non empty subsets of S, and conditional potentials $\Phi^*_W(.|x)$ for each $W \in \mathcal{C}_x$ such that $U(y|x) = \sum_{W \in \mathcal{C}_x} \Phi^*_W(y|x)$.

Throughout the paper, we suppose that $\mathcal{C}_x = \mathcal{C}$ does not depend on x. If we denote $\mathbf{0}$ a reference state in Ω ($\mathbf{0}$ is 0 when $E = \mathbb{N}$, $\mathbb{R}$ or $\mathbb{R}_+$), then, almost surely in x, we can choose (in a unique way) potentials according to the "identifiability conditions": $\Phi^*_W(y|x) = 0$ if for some $i \in W,\ \ y_i = 0$. Besides, for each W in $\mathcal{C}$, there exists a family $\mathcal{C}_W$ of non-empty parts of S such that the potential Φ^*_W can be written as:

$$\Phi^*_W(y|x) = \Phi_W(y) + \sum_{W' \in \mathcal{C}_W} \Phi_{W',W}(x, y)$$

This means that the energy U is linked to two families of potentials: instantaneous interaction potentials $\{\Phi_W,\ W \in \mathcal{C}\}$ and "conditional" interaction potentials $\{\Phi_{W',W},\ W \in \mathcal{C},\ W' \in \mathcal{C}_W\}$.

The semi-causal representation is associated to $\mathbf{G} = \{\mathcal{G}, \mathcal{G}_-\}$ where $\mathcal{G}$ and $\mathcal{G}_-$ are defined respectively by the instantaneous and time-delay dependencies:

$$\langle i, j\rangle_{\mathcal{G}} \Leftrightarrow \langle (t, i), (t, j)\rangle_{\mathbf{G}} \Leftrightarrow \exists W \in \mathcal{C} \text{ s.t. } \{i, j\} \subseteq W \text{ and } \Phi_W \neq 0$$

$$\langle j, i\rangle_{\mathcal{G}_-} \Leftrightarrow \langle (t-1, j), (t, i)\rangle_{\mathbf{G}} \Leftrightarrow \exists W \in \mathcal{C}, W' \in \mathcal{C}_W \text{ s.t. } \Phi_{W',W} \neq 0,\ i \in W,\ j \in W'$$

Note that $\mathcal{G}_-$ is a directed graph while $\mathcal{G}$ is not. Let us define $\mathcal{C}^- = \underset{W \in \mathcal{C}}{\cup} \mathcal{C}_W$; then $U(y|x) = \sum_{W \in \mathcal{C}} \Phi_W(y) + \sum_{W \in \mathcal{C}, W' \in \mathcal{C}^-} \Phi_{W',W}(x, y)$ with the understanding that $\Phi_{W',W} \equiv 0$ if $W' \notin \mathcal{C}_W$.

Thus we finally write:

$$P(x,y) = Z^{-1}(x)\exp\{\sum_{W}\Phi_W(y) + \sum_{W_1,W_2}\Phi_{W_1,W_2}(x,y)\} \tag{2}$$

where we have put: $\sum_W$ for $\sum_{W\in\mathcal{C}}$ and $\sum_{W_1,W_2}$ for $\sum_{W_1\in\mathcal{C}^-,W_2\in\mathcal{C}}$, with $\Phi_W(y) = 0$ (resp. $\Phi_{W_1,W_2}(x,y) = 0$) if for some $i \in W$ (resp. $i \in W_2$), $y_i = 0$.

There are three components in the neighbourhood of a site i:

•$\partial i = \{j \in S \setminus \{i\},\ \langle i,j\rangle_{\mathcal{G}}\}$: the t-instantaneous neighbourhood of i

•$\partial i^- = \{j \in S,\ \langle j,i\rangle_{\mathcal{G}_-}\}$: the $(t-1)$-antecedent neighbourhood of i

•$\partial i^+ = \{j \in S,\ \langle i,j\rangle_{\mathcal{G}_-}\}$: the $(t+1)$-successor neighbourhood of i

The semi-causal representation is related to ∂i and ∂i^-, while the non causal representation that we will present in the next section depends on ∂i, ∂i^- and ∂i^+.

Therefore, for each $t \geq 1$ and $A \subset S$ ($A \neq \emptyset$), the conditional distribution of $X_A(t)$ given the past and $X^A(t) = y^A$ depends on $X_{\partial A^-}(t-1) = x_{\partial A^-}$ and $X_{\partial A}(t) = y_{\partial A}$ only, where $\partial A^- = \{i \in S : \exists j \in A \text{ s.t. } \langle i,j\rangle_{\mathcal{G}^-}\}$. The corresponding conditional energy is:

$$U_A(y_A|y^A,x) = \sum_{W:W\cap A\neq\varnothing}\Phi_W(y) + \sum_{W_2:W_2\cap A\neq\varnothing}\{\sum_{W_1}\Phi_{W_1,W_2}(x,y)\}$$

3 Some properties of an MCMF

3.1 Time reversibility, invariant and marginal distributions

In this section we only consider potentials Φ_{W_1,W_2} such that $\Phi_{W_1,W_2}(x,y) = 0$ if for an $i \in W_1, x_i = 0$ or for a $j \in W_2, y_j = 0$. When there is no ambiguity, 0 denotes also the layout with 0 in any site of S.

The transition P is synchronous if we can write $P(x,y) = \prod_{s\in S} p_s(x,y_s)$: the values at all sites s are independently and synchronously relaxed with distributions $p_s(x,.)$ in s.

Proposition 1 *(i) The chain is time-reversible if and only if for all $W_1, W_2, x, y : \Phi_{W_1,W_2}(x,y) = \Phi_{W_2,W_1}(y,x)$. In this case, P has an unique invariant probability measure given by:*

$$\pi(y) = \pi(0)\frac{P(0,y)}{P(y,0)} = \pi(0)Z^{-1}(0)Z(y)\exp\sum_{W}\Phi_W(y)$$

(ii) This invariant measure π is usually not a Markov field. If the transition P is synchronous and reversible, then π has a Markov property.

Proof:

(i) This can be derived from Künsch (1984), and Guyon (1995, Theorem 2.2.3.).

(ii) We give in Appendix 1 two examples of non Markovian π.

If P is synchronous, only singletons occur for W in Φ_W and for W_2 in Φ_{W_1,W_2}. As the chain is reversible, $\Phi_{W_1,W_2} \equiv 0$ if either $|W_1| > 1$ or $|W_2| > 1$. This means that $P(x_i(t)|x^i(t), x(t-1)) = P(x_i(t)|x_{\partial i^-}(t-1))$, so that,

$$P(x,y) = Z^{-1}(x)\exp\{\sum_{s\in S}\Phi_s(y_s) + \sum_{\langle s',s\rangle_{\mathcal{G}^-}}\Phi_{\{s'\},\{s\}}(x_{s'},y_s)\} = \prod_{s\in S} p_s(x,y_s)$$

with $p_s(x,y_s) = Z^{-1}(x)\exp\{\Phi_s(y_s) + \sum_{s'}\Phi_{\{s'\},\{s\}}(x_{s'},y_s)\}$. Then we have:

$$\pi(y) = \pi(0)Z^{-1}(0)Z(y)\prod_{s\in S}\exp\Phi_s(y_s)$$

Besides, if $U_s(z_s,y) = \Phi_s(z_s) + \sum_{s'\in\partial s^-}\Phi_{\{s'\},\{s\}}(y_{s'},z_s)$, $Z(y) = \int\exp\{\sum_{s\in S}U(z_s,y)\}\nu^S(dz) = \prod_{s\in S}\int\exp U(z_s,y)\nu(dz_s) = \exp\sum_{s\in S}\Psi_{\partial s^-}(y_{\partial s^-})$ with $\Psi_{\partial s^-}(y_{\partial s^-}) = \ln\{\sum_{z_s}\exp U(z_s,y)\}$.□

Example 1 *A synchronous and reversible transition.*

Let $E = \{0,1\}$, S is the one dimensional torus; the transition $P(x,y) = Z^{-1}(x)\exp\{\alpha\sum_{i\in S}y_i(x_{i-1} + x_{i+1})\}$ is reversible with invariant law $\pi(y) = \pi(0)Z^{-1}(0)\exp\sum_{i\in S}\Phi_{i-1,i+1}(y)$, where $\Phi_{i-1,i+1}(y) = \ln\{1+\exp\alpha(y_{i-1}+y_{i+1})\}$. The conditional distribution at site l depends on y_{l-2} and y_{l+2}.

Marginal distributions

For $A \subseteq S$, $A \neq S$, the marginal distributions $(y_A|x)$, conditionally to x, are generally not local in x, and not explicit and not local in y, except in specific cases as the Gaussian case. This is illustrated in Appendix 2.

3.2 Non causal Markov Field representation

Let X be an MCMF with the semi-causal representation (2). We are going to show that there is a unique equivalent time homogeneous space×time non-causal Markov field representation given

by the bilateral transitions: for $X(t-1)=x$ and $X(t+1)=z$,

$$P(y|x,z)=Z^{-1}(x,z)\exp\{\sum_{W_1,W_2}\{\Phi_{W_1,W_2}(x,y)+\Phi_{W_1,W_2}(y,z)\}+\sum_{W}\Phi_W(y)\} \tag{3}$$

where the normalizing constant $Z(x,z)$ is finite $a.s$ in (x,z). The time translation invariant potentials on $\mathcal{S}=S\times\mathbb{Z}$ are $\widetilde{\Phi}_{W\times\{0\}}(y,0)=\Phi_W(y)$ and $\widetilde{\Phi}_{W_1\times\{0\},W_2\times\{1\}}((x,0),(y,1))=\Phi_{W_1,W_2}(x,y)$. The non-causal representation depends on all the three neighbourhoods ∂i, ∂i^- and ∂i^+.

Proposition 2 *The representation (2) of MCMF dynamics with the neighbourhood system $\{\partial i,\partial i^-,i\in S\}$ is equivalent to the $\mathcal{S}$ Space$\times$Time Markov-Field representation (3) with the neighbourhood system $\{\partial i,\partial i^-,\partial i^+,i\in S\}$.*

Proof:

(i) It is easy to see that the chain is also a two-nearest neighbours Markov field in time. Let τ be the density of $X(0)$; the likelihood of $(x(0),x(1),...x(T))$ is $\tau(x(0))\prod_{t=1,T}P(x(t-1),x(t))$. For $1\leq t\leq T-1$, the conditional density is:

$$P\left(x(t)|x(t-1),x(t+1)\right)=\frac{P\left(x(t-1),x(t)\right)P\left(x(t),x(t+1)\right)}{\int_\Omega P\left(x(t-1),a(t)\right)P\left(a(t),x(t+1)\right)\nu^S(da)}.$$

Let us denote $x=x(t-1)$, $y=x(t)$, $z=x(t+1)$. As $(X(t+1)\mid X(t-1)=x)$ admits an $a.s.$ finite density, $Z(x,z)=P(X(t+1)=z\mid X(t-1)=x)=\int_\Omega P\left(x,a\right)P\left(a,z\right)\nu^S(da)$, is finite. We obtain from (2):

$$P(y|x,z)=\frac{\exp\left\{\sum_{W_1,W_2}\{\Phi_{W_1,W_2}(x,y)+\Phi_{W_1,W_2}(y,z)\}+\sum_W\{\Phi_W(y)+\Phi_W(z)\}\right\}}{\int_\Omega\exp\left\{\sum_{W_1,W_2}\{\Phi_{W_1,W_2}(x,a)+\Phi_{W_1,W_2}(a,z)\}+\sum_W\{\Phi_W(a)+\Phi_W(z)\}\right\}\nu^S(da)}$$

This is nothing other than (3).

(ii) Conversely, let X be the space$\times$time Markov field on $\mathcal{S}$ (3) with the neighbourhood system $\{\partial i,\ \partial i^-,\ \partial i^+\}$. The field being time-homogeneous, a direct computation shows that its semi-causal representation is (2). $\square$

We can derive easily from (2) or (3) the semi-causal or non-causal conditional distributions at any point (i,t). Note that we have for each $i,j:\ \ i\in\partial j^-\Longleftrightarrow j\in\partial i^+$.

Figure 1 shows an example of both representations. For the causal representation, we have $\partial i = \{j,k\}$, $\partial i^- = \{i,j\}$ (note that $i \in \partial l^-$); while for the non-causal representation, $\partial i = \{j,k\}$, $\partial i^- = \{i,j\}$, $\partial i^+ = \{i,j,l\}$ and now $i \in \partial l^-$ and $l \in \partial i^+$.

(include here Figure 1)

The semi-causal conditional distribution at (i,t) is:

$$P\left(y_i | y_{\partial i}, x_{\partial i^-}\right) = Z_i^{-1}\left(y_{\partial i}, x_{\partial i^-}\right) \exp\{\sum_{W_1,W_2 \ni i} \Phi_{W_1,W_2}(x,y) + \sum_{W \ni i} \Phi_W(y)\} \tag{4}$$

The non-causal conditional distribution $P\left(y_i | y_{\partial i}, x_{\partial i^-}, z_{\partial i^+}\right)$ at (i,t) has conditional energy:

$$\sum_{W_1,W_2 \ni i} \{\Phi_{W_1,W_2}(x,y) + \Phi_{W_1,W_2}(y,z)\} + \sum_{W \ni i} \Phi_W(y)$$

A time inhomogeneous Markov field on $\mathcal{S} = S \times T$ is not reducible to a semi-causal representation because ∂i^- and ∂i^+ are strongly related: the Markov field (3) is very specific. For the example considered in Figure 1, the non causal representation without the dotted arrow $(i,t-1) \to (l,t)$ cannot be reducible to a causal representation. In all the following, we consider the time homogeneous framework.

Reversed dynamics

We give in Appendix 3 an example where the time reversed process of an MCMF is no more an MCMF.

4 Ergodicity of automodels

Here we examine several examples of auto-models (see Besag 1974; Guyon 1995), and we give conditions ensuring their ergodicity. We will use the *Lyapunov Stability Criterion* (see e.g. Duflo 1997; [13], 6.2.2). For $n \geq 0$, let $\mathcal{F}_n = \sigma\left(X(s),\, s \leq n\right)$ be the $\sigma-$algebra generated by the $X(s)$, $s \leq n$. The Lyapunov Stability Criterion is the following. Let us assume that a Markov chain defined on a closed subset of $\mathbb{R}^d$ is strongly Feller, and that there exists a *Lyapounov function* V such that, for $n > 0$, and for some $0 \leq \alpha < 1$ and $\beta < \infty$

$$\mathbb{E}[V(X_n)|\mathcal{F}_{n-1}] \leq \alpha V(X_{n-1}) + \beta$$

Then if there is at most one invariant probability measure, the chain is positive recurrent.

Besides, the following law of large numbers applies: $\frac{1}{n+1}\sum_{k=0,n}\varphi(X_k) \overset{a.s.}{\to} \mu(\varphi)$ for any $\mu-$integrable function φ, where μ is the invariant measure of the chain. A difficulty is to check whether φ is $\mu-$integrable. Another useful result is available: we get the same law of large numbers for any function φ which is μ-a.s. continuous and s.t. $|\varphi| \leq aV + b$ for some constants a, b.

4.1 The autoexponential dynamics

Let us consider $S = \{1, 2, \cdots, n\}$, and $E = \mathbb{R}^+$. We suppose that for each i, conditionally to $\big(X^i(t) = y^i,\ X(t-1) = x\big)$ (later, we shall note this condition (y^i, x)), the distribution of $X_i(t)$ is exponential with parameter $\lambda_i(y^i, x)$ (in fact $\lambda_i(y_{\partial i}, x_{\partial i^-})$). From Arnold and Strauss (1988; [3]) we can write:

$$-U(y|x) = \sum_{i\in S}\alpha_i(x)y_i + \sum_{W:|W|\geq 2}\lambda_W(x)y_W$$

with $y_W = \prod_{i\in W} y_i$, $\alpha_i(x) > 0$ for any x and $\lambda_W(x) \geq 0$. Then, the parameters are equal to $\lambda_i(y^i, x) = \alpha_i(x) + \sum_{W\ni i}\lambda_W(x)y_{W\setminus\{i\}}$. The ergodicity of the chain is obtained through the following assumption:

E1 *(i)* $\forall i \in S$, $\forall W$, $x \to \alpha_i(x)$ *and* $x \to \lambda_W(x)$ *are continuous.*

(ii) $\exists a \in (0, \infty)$, *s.t.* $\forall x \in (\mathbb{R}^+)^S$, $\forall i \in S : \alpha_i(x) \geq a$.

Proposition 3 *Under assumption **E1**, the autoexponential dynamics is positive recurrent. The strong law of large numbers holds for any integrable function and particularly for functions $x \mapsto f(x)$ such that $|f(x)| \leq \alpha V_r(x) + \beta$ (for some finite constants α and β) with $V_r(x) = \sum_{i\in S} x_i^r$, r being any positive integer.*

Proof: The proof uses the Lyapounov stability criterion.

• The lower bound condition **E1** (ii) says that $\exp U(y|x) \leq \exp -a\sum_{i\in S} y_i$, which is Lebesgue integrable. Then, the chain is strongly Feller.

• As P is strictly positive, the chain is irreducible and there exists no more than one invariant distribution (see Duflo (1997), Proposition 6.1.9).

- On the other hand, for any positive integer r, we have for all $x, y \in E$, $i \in S$,

$$E\left[\{X_i(t)\}^r | x, y^i\right] = \frac{\Gamma(r+1)}{\lambda_i(y^i, x)^r} \leq \frac{\Gamma(r+1)}{a^r}$$

It follows:

$$E\left[V_r(X(t)) | X(t-1) = x\right] \leq \frac{\Gamma(r+1)}{a^r} |S| < \infty \tag{5}$$

where $|S|$ is the cardinal of S. $\square$

Example 2 *A case of "weak-reversibility"*

We assume that, conditionally to $\left(X^i(t-1) = x^i,\ X(t) = y\right)$, the reversed transition $Q(y, x)$ is also exponential with parameter $\mu_i(x^i, y)$. Then, from Arnold and Strauss (1988), the joint density for $(X(t-1) = x, X(t) = y)$ must be:

$$f(x, y) = C \exp U(x, y), \text{ with } U(x, y) = -\sum_{W = W_1 \times W_2 \subset S^2, W \neq \emptyset} \lambda_W\, x_{W_1} y_{W_2}$$

with $\lambda_W > 0$ if $|W| = |W_1| + |W_2| = 1$ and $\lambda_W \geq 0$ if $|W| \geq 2$. So **E1** is satisfied.

Example 3 *Besag's conditional auto-models*

We consider the case of conditional auto-models, i.e. for $W \subseteq S$, $\lambda_W = 0$ if $|W| > 2$. Therefore:

$$-U(y|x) = \sum_{i \in S} \alpha_i(x) y_i + \sum_{\langle i,j \rangle} \beta_{ij}(x) y_i y_j.$$

E1-(ii) is satisfied if $\alpha_i(x) \geq a$ and $\beta_{ij}(x) \geq 0$ for all x, i, j. For example, if

$$-U(y|x) = \sum_{i \in S} \delta_i y_i \ + \sum_{\langle i,j \rangle_{\mathcal{G}}} \beta_{ij} y_i y_j + \sum_{\langle j,i \rangle_{\mathcal{G}_-}} \alpha_{ji} x_j y_i \tag{6}$$

with $\delta_i > 0$, β_{ij} and $\alpha_{ij} \geq 0$, the distribution of $X_i(t)$ conditionally to (y^i, x) is exponential with parameter $\lambda_i(y^i, x) = \delta_i + \sum_{j \in \partial i} \beta_{ij} y_j + \sum_{l \in \partial i^-} \alpha_{li} x_l$. The condition **E1** is fulfilled .

4.2 The autonormal dynamics

Let $E = \mathbb{R}$. We assume that the conditional distribution of $X_i(t)$ given $X^i(t) = y^i$ and $X(t-1) = x$ is Gaussian with mean $\mu_i(y^i, x)$ and variance $\sigma_i^2(y^i, x)$. The principle of compatibility requires (see Arnold and Press (1989; [2]), Arnold, Castillo and Sarabia (1991; [4])) that the conditional energy is of the following feature:

$$-U(y|x) = \sum_{i \in S} \{\alpha_i(x) y_i + \beta_i(x) y_i^2\} + \sum_{W \, s.t. \, |W| \geq 2} \gamma_W(x) y_W^{l_W}$$

where $y_W^{l_W} = \prod_{i \in W} y_i^{l_i}$, $l_i = 1, 2$, and the functions α, β, γ ensuring that all the conditional variances are positive and $U(.|x)$ is admissible.

Let us consider now Besag's automodels. Then $-U(y|x) = -\sum_{i \in S} \{\alpha_i(x) y_i + \beta_i(x) y_i^2\} + \sum_{\langle i,j \rangle_{\mathcal{G}}} \{\gamma_{ij}(x) y_i y_j + \delta_{ij}(x) y_i^2 y_j + \nu_{ij}(x) y_i y_j^2 + \chi_{ij}(x) y_i^2 y_j^2\}$. A typical example is:

$$P(x, y) = Z^{-1}(x) \exp -\{\sum_{i \in S} y_i (\delta_i + \sum_{l \in \partial i^-} \alpha_{li} x_l + \gamma_i y_i) + \sum_{\langle i,j \rangle} \beta_{ij} y_i y_j\} \tag{7}$$

with $\gamma_i > 0$, $\beta_{ij} = \beta_{ji}$ such that the matrix $Q = (Q_{ij})$ defined by $Q_{ii} = 2\gamma_i$ and $Q_{ij} = \beta_{ij}$ is definite positive. The conditional distribution of $X_i(t)$ given (y^i, x) is Gaussian with mean $-(\delta_i + \sum_{l \in \partial i^-} \alpha_{li} x_l + \sum_{j \in \partial i} \beta_{ij} y_j)/2\gamma_i$ and variance $(2\gamma_i)^{-1}$.

Ergodicity: We consider the model (7). Let us note $\Delta = (\Delta_{il})$, $\Delta_{il} = -\alpha_{li}, i, l \in S$, and $\delta = (\delta_i)$; a direct identification of the distribution of $(Y \mid x) \doteq (X(t) \mid X(t-1) = x)$ gives

$$(Y \mid x) \sim \mathcal{N}_S(m + Ax, \Gamma), \quad \text{with } m = -Q^{-1}\delta, A = Q^{-1}\Delta, \text{ and } \Gamma = Q^{-1}.$$

This can be written as an AR(1) process $X(t) = m + AX(t-1) + \varepsilon(t)$ with a Gaussian white noise ε having covariance matrix Γ. Define $\tau = (I - A)^{-1} m$ when $(I - A)$ is regular. Then the zero-mean variable $X^*(t) = X(t) - \tau$ verifies $X^*(t) = AX^*(t-1) + \varepsilon(t)$. Let $\rho(A)$ be the spectral radius of A (i.e. the greatest modulus of its eigen values).

Proposition 4 *If $\rho(A) < 1$, then $(I - A)$ is regular and the chain is ergodic with a Gaussian stationary measure μ.*

Proof: The result is classic, given for example in Duflo (1997), Theorem 2.3.18. As ε is Gaussian, the stationary distribution for X^* is the one of $\sum_{k \geq 0} A^k \varepsilon(k)$, i.e. $\mathcal{N}_S(0, \Sigma)$, Σ being

the unique solution of $\Gamma = \Sigma - A\Sigma({}^tA)$, that is $\Sigma = \sum_{k\geq 0} A^k \Gamma({}^tA)^k$. For $X(t)$, we add the mean τ. As ε has finite moments of all orders, we get the strong law of large numbers for any continuous integrable function. $\square$

4.3 The Auto-Poissonian dynamics

Now $E = \mathbb{N}$ (ν is the counting measure) and we consider the dynamics associated with the following conditional energy:

$$U(y|x) = \sum_{i\in S} \{\alpha_i(x) y_i - \ln(y_i!)\} + \sum_{i\neq j} \beta_{ij}(x) y_i y_j$$

where $\beta_{ij}(x) \leq 0$ for all i, j in order to make U admissible. Conditionally to $(y^i, x), X_i(t)$ has a Poisson distribution with parameter $\lambda_i(y^i, x) = \exp\{\alpha_i(x) + \sum_{j:j\neq i} \beta_{ij}(x) y_j\}$. The ergodicity of the chain is obtained through the following hypothesis:

P1: $\exists\ M < \infty$ *such that* $\forall i \in S$, $\sup_x \alpha_i(x) \leq M$

Proposition 5 *Under **P1**, the auto-Poissonian chain is positive recurrent. Besides, the strong law of large numbers holds for any μ-integrable function and for any functions f such that $|f(x)| \leq \alpha G_u(x) + \beta$ (for some finite constants α and β) where $G_u(x) = \prod_{i\in S} e^{u_i x_i}$, for any fixed $u \in (\mathbb{R}^+)^S$.*

Proof: The proof follows the same lines as the one given for the auto-exponential dynamics. **P1** implies that $\lambda_i(y^i, x) \leq e^M$. Then the transition is strongly Feller. As this transition is strictly positive, the invariant measure is unique.

For the conditional moment generating function, we have, for all $s > 0$,
$\Psi_{X_i}(s) = E\left[e^{sX_i(t)}|y^i, x\right] = e^{\lambda_i(y^i,x)(e^s-1)} \leq \exp(e^M(e^s - 1))$. Let us set $u = (u_i, i \in S) \in (\mathbb{R}^+)^S$, $K_u = \max_{\ i\in S} \exp(e^M(e^{u_i} - 1))$ and $G_u(x) = \prod_{i\in S} e^{u_i x_i}$. Then, conditionally to $X(t-1) = x$,

$$E\left[G_u(X(t))|x\right] = E\left[\prod_{i\in S\setminus\{j\}} e^{u_i X_i(t)}\ E\left[e^{u_j X_j(t)}|y^j, x\right]|x\right] \ \leq K_u E\left[\prod_{i\in S\setminus\{j\}} e^{u_i X_i(t)}|x\right]$$

Taking successive conditional expectations, one obtains:

$$E\left[\prod_{i\in S} e^{u_i X_i(t)}\,|\,X(t-1)\right] \leq K_u^{|S|} < \infty. \qquad \square$$

For example, **P1** is satisfied for

$$\alpha_i(x) = \delta_i + \sum_{l \in \partial i^-} \alpha_{li} x_l \text{ and } \beta_{ij}(x) = \beta_{ij} = \beta_{ji} \tag{8}$$

with $\alpha_{li} \leq 0$ for any i, l, and $\beta_{ij} < 0$ if $\langle i, j \rangle$ and 0 else.

4.4 The auto-discrete dynamics

Let E be a finite qualitative set. The conditional energy of the auto-model is:

$$U(y|x) = \sum_{i \in S} \alpha_i(y_i, x) + \sum_{\langle i,j \rangle} \beta_{ij}(y_i, y_j, x).$$

As $\alpha_i(.)$ and $\beta_{ij}(.)$ are finite conditional potentials, the ergodicity is ensured without any restrictions on the parameters. For instance, the autologistic dynamics is defined for $E = \{0, 1\}$, and $U(y|x) = \sum_{i \in S} \alpha_i(x) y_i + \sum_{\langle i,j \rangle} \beta_{ij} y_i y_j$; $X_i(t)$ has a conditional Bernouilli distribution with parameter

$$p_i(y^i, x) = \frac{\exp \delta_i(y^i, x)}{1 + \exp \delta_i(y^i, x)} \text{ with } \delta_i(y^i, x) = \alpha_i(x) + \sum_{j \in \partial i} \beta_{ij} y_j. \tag{9}$$

For $E = \{0, 1, .., m\}$, the autobinomial dynamics is given by $U(y|x) = \sum_{i \in S} \alpha_i(x) y_i + \sum_{\langle i,j \rangle} \beta_{ij} y_i y_j$. Conditionally to (y^i, x), $X_i(t)$ has a Binomial distribution $\mathcal{B}(m, \theta_i(y^i, x))$ with parameter $\theta_i(y^i, x) = (1 + \exp\{\alpha_i(x) + \sum_{j \in \partial i} \beta_{ij} y_j\})^{-1}$ if we take the counting measure with density $\binom{m}{y_i}$ as reference measure.

5 Conditional Pseudo-Likelihood Statistics

5.1 Parametric estimation

We suppose that the transition probabilities of the MCMF depend on an unknown parameter θ, θ lying in the interior of Θ, a compact subset of $\mathbb{R}^d$. When ergodicity holds, we can obtain the asymptotic properties of the estimators derived from the classic estimation methods (maximum likelihood, pseudo maximum likelihood, maximum of another "nice" objective function) in a standard way. An analytical and numerical difficulty inherent to the maximum likelihood procedure is the complexity of the normalizing constant $Z_\theta(x)$ in the likelihood; one can use

stochastic gradient algorithms to solve the problem (see Younes (1988), [33]); another numerical option is to compute the log likelihood and its gradient by simulations via a Monte Carlo algorithm (see [16] Chapter 3 and [17]). A third alternative (see Besag (1974)) is to consider the Conditional Pseudo-Likelihood (CPL); in the presence of strong spatial autocorrelation this method performs poorly, and we then have to use the previous procedures. In the absence of strong dependency, it has good asymptotic properties, the same rate of convergence as the maximum likelihood estimator with a limited loss in efficiency (see Besag 1977 [10], Guyon 1995 [18], Guyon and Künsch 1992 [20]). The asymptotic behaviour follows in a standard way (see Amemiya (1985) and Dacunha-Castelle and Duflo (1986) for general theory; Besag (1984,[10]), Guyon and Hardouin (1992; [19]), Guyon (1995) for Markov field estimation; Bayomog (1994; [5]), and Bayomog *et al.* (1996) for field dynamics estimation). We briefly recall the main results.

We assume that the chain is homogeneous and that for all $i \in S$, $x, y \in E$, $\theta \in \Theta$, the conditional distribution of $X_i(t)$ given $X^i(t) = y^i$ and $X(t-1) = x$ is absolutely continuous with respect to ν, with positive conditional density $f_i(y_i, y^i, x; \theta)$ (which is in fact $f_i(y_i, y_{\partial i}, x_{\partial i^-}; \theta)$).

Let θ_0 be the true value of the parameter, and P_0 be the associated transition. The process is observed at times $t = 0, \cdots, T$. Let us denote $\hat{\theta}_T = \arg\min_{\theta \in \Theta} U_T(\theta)$ the conditional pseudo-conditional likelihood estimator (CPLE) of θ, a value minimizing the opposite of the Log-CPL:

$$U_T(\theta) = -\frac{1}{T} \sum_{t=1}^{T} \sum_{i \in S} \ln f_i(x_i(t), x^i(t), x(t-1); \theta)$$

The following conditions **C** and **N** ensure the consistency and the asymptotic normality of $\hat{\theta}_T$ respectively.

Conditions for consistency (C):

C1: For $\theta = \theta_0$, the chain X is ergodic with a unique stationary measure μ_0.

C2: (i) For all $i \in S$, $x, y \in E$, $\theta \mapsto f_i(y_i, y^i, x; \theta)$ is continuous..

(ii) There exists a measurable $\mu_0 \otimes P_0$-integrable function h on $E \times E$ such that for all $i \in S$, $\theta \in \Theta$, x, $y \in E$, $|\ln f_i(y_i, y^i, x; \theta) - \ln f_i(y_i, y^i, x; \theta_0)| \leq h(y, x)$.

C3: Identifiability: if $\theta \neq \theta_0$ then $\sum_{i \in S} \mu_0(\{x \ \ s.t. \ f_i(., ., x; \theta_0) \neq f_i(., ., x; \theta)\}) > 0$.

Let $f_i^{(1)}(\theta)$ and $f_i^{(2)}(\theta)$ stand for the gradient and the Hessian matrix of $f_i(y_i, y^i, x; \theta)$ with respect to θ. We define the following conditional (pseudo) information matrices: for $x, y \in E$, $i, j \in S, i \neq j$, $\theta \in V_0$, a neighbourhood of θ_0, $V_0 \subset \Theta$:

$$I_{ij}(y^{\{i,j\}},x,\theta)=\mathbf{E}_{\theta_0}[\tfrac{f_i^{(1)}(\theta)f_j^{(1)}(\theta)'}{f_i(\theta)f_j(\theta)}\mid X^{\{i,j\}}(t)=y^{\{i,j\}},X(t-1)=x],$$
$$I_{ij}(x,\theta)=\mathbf{E}_{\theta_0}\left[I_{ij}(X^{\{i,j\}}(t),X(t-1),\theta)\mid X(t-1)=x\right].$$

If $Z_i=\frac{\partial}{\partial\theta}\ln f_i(X_i(t),X^i(t),X(t-1);\theta)|_{\theta=\theta_0}$, then $I_{ij}(y^{\{i,j\}},x,\theta)$ is the covariance matrix of (Z_i,Z_j) given $X^{\{i,j\}}=y^{\{i,j\}}$ and $X(t-1)=x$.

Conditions for asymptotic normality (N):

N1: For some $V_0\subset\overset{\circ}{\Theta}$, a neighbourhood of θ_0, $\theta\mapsto f_i(y_i,y^i,x;\theta)$ is two times continuously differentiable on V_0 and there exists a measurable, $\mu_0\otimes P_0$-square integrable function H on $E\times E$ such that for all $\theta\in V_0$, $x,y\in E$, $1\leq u,v\leq d$:

$$|\frac{1}{f_i}\frac{\partial}{\partial\theta_u}f_i(y_i,y^i,x;\theta)|\text{ and }|\frac{1}{f_i}\frac{\partial^2}{\partial\theta_u\partial\theta_v}f_i(y_i,y^i,x;\theta)|\leq H(y,x)$$

N2: $I_0=\sum_{i\in S}E_{\mu_0}\left[I_{ii}(X(t-1),\theta_0)\right]$ and $J_0=\sum_{i,j\in S}E_{\mu_0}\left[I_{ij}(X(t-1),\theta_0)\right]$ are positive definite.

Proposition 6 *(Bayomog et al. (1996), Guyon and Hardouin (1992), Guyon (1995))*

(i) Under assumptions ***C****,* $\hat\theta_T\xrightarrow[T\to\infty]{P_0}\theta_0$

(ii) Under assumptions ***C*** *and* ***N****,* $\sqrt{T}(\hat\theta_T-\theta_0)\xrightarrow[T\to\infty]{\mathcal{D}}\mathcal{N}_d(0,{I_0}^{-1}J_0{I_0}^{-1})$

Identifiability of the model and regularity of $\mathbf{I_0}$: We give here sufficient conditions ensuring both **(C3)** and the regularity of the matrix I_0. We suppose that each conditional density belongs to an exponential family

$$f_i(y_i,y^i,x;\theta)=\ K_i(y_i)\exp\{{}^t\theta g_i(y_i,y^i,x)-\Psi_i(\theta,y^i,x)\},\tag{10}$$

and set the hypotheses:

(H) : there exists $(i(k),x(k),y(k),k=1,d)$ s.t. $g=(g_1,g_2,\cdots,g_d)$ is of rank d,

where we denote $g_k=g_{i(k)}(y_{i(k)}(k),y^{i(k)}(k),x(k))$. We strengthen **(H)** in **(H')**:

(H'1) : for each $i\in S$, $g_i(y_i,y^i,x)=h_i(y_i)G_i(y^i,x)$ with $h_i(.)\in\mathbb{R}$

(H'2) : $\exists a>0$ s.t. for each i,x,y^i, $Var\{h_i(Y_i)\mid y^i,x)\}\geq a\ >0$

(H'3) : $\exists\{(i(k),x(k),y^{i(k)}(k)),k=1,d\}$ s.t. $G=(G_1,G_2,\cdots,G_d)$ is of rank d

where $G_k=G_k(y^{i(k)}(k),x(k))$. Obviously, **(H')** implies **(H)**.

Proposition 7 *We suppose that the conditional densities* $\{f_i(., y^i, x; \theta), i \in S\}$ *belong to the exponential family (10).*

(i) Under ***(H)****, the model related to this family of conditional densities is identifiable.*

(ii) I_0 *is regular under* ***(H')****.*

The proof is given in Appendix 4. A sufficient condition ensuring that J_0 is positive definite can also be obtained using a strong coding set; this idea is developed in Jensen and Künsch (1994; [22]). It is sketched in the same Appendix 4. Many models fulfil **(H)** or **(H')**, for instance log-linear models or automodels. We give explicit conditions (**C**) and (**N**) for the autoexponential dynamics in section 5.3.

5.2 Testing submodels

It is now possible to test the submodel $(H_q) : \theta = \varphi(\alpha),\ \alpha \in \mathbb{R}^q,\ q < d,\ \varphi : \mathbb{R}^q \to \mathbb{R}^d$ such that:

• φ is twice continuously differentiable in a bounded open set Λ of $\mathbb{R}^q$, with $\varphi(\Lambda) \subset \Theta$, and there exists $\alpha_0 \in \Lambda$ such that $\varphi(\alpha_0) = \theta_0$.

•• $R = \frac{\partial \varphi}{\partial \alpha} |_{\alpha=\alpha_0}$ is of rank q.

Let $\bar{\theta}_T = \arg\min_{\alpha \in \Lambda} U_T(\varphi(\alpha))$ be the CPLE of θ under (H_q), and $\overline{I}_0 = \overline{I}(\alpha_0)$ the associated information matrix I. If A is a positive definite matrix, with spectral decomposition $A = PDP'$, P orthogonal, we take $A^{\frac{1}{2}} = PD^{\frac{1}{2}}P'$ as a square root of A.

Proposition 8 *CPL ratio test (Bayomog (1994), Bayomog et al. (1996), Guyon(1995))*

If $U_T(\theta)$ *and* $U_T(\varphi(\alpha))$ *satisfy assumptions* ***C*** *and* ***N****, then, under* (H_q)*, we have as* $T \to \infty$*:*

$$\Delta_T = 2T\left(U_T(\bar{\theta}_T) - U_T(\hat{\theta}_T)\right) \xrightarrow{\mathcal{D}} \sum_{i=1}^{d-q} \lambda_i \chi^2_{1,i}$$

where the $\chi^2_{1,i}$*'s are independent* χ^2_1 *variables and the* λ_i*,* $i = 1, d-q$*, are the* $(d-q)$ *strictly positive eigenvalues of* $\Gamma_0 = {J_0}^{1/2}\left[I_0^{-1} - R\bar{I}_0^{-1}R'\right]J_0^{1/2}$.

Let $\mathcal{C}$ be a coding subset of S, i.e., for all $i, j \in \mathcal{C}$, $i \neq j$, i and j are not (instantaneous) neighbour sites. We can define coding estimators as previously, but in the definition of the coding contrast $U_T^{\mathcal{C}}(\theta)$, the summation in i is then restricted to $\mathcal{C}$. For those estimators, we

have $I_0^{\mathcal{C}} = J_0^{\mathcal{C}}$; so the asymptotic variance for $\sqrt{T}\left(\widehat{\theta}_T^{\mathcal{C}} - \theta_0\right)$ is $(I_0^{\mathcal{C}})^{-1}$, and the former statistic has a χ^2_{d-q} asymptotic distribution.

5.3 An example: The autoexponential model

Here we look at these various conditions for the model given in section 4.1. The assumption **E1** ensures the ergodicity. For a positive integer r and $V_r(x) = \sum_{i\in S} x_i^r$, we define the following property for a function f:

There exist two finite constants α and β such that

$$f(x) \leq \alpha V_r(x) + \beta \tag{11}$$

We add the following hypotheses to obtain the consistency and the asymptotic normality of the CPL estimator.

E2 *For all $i \in S$, $W \in \mathcal{C}, \theta \in \Theta$ and $x \in E$, $\alpha_i(x,\theta)$ and $\lambda_W(x,\theta)$ satisfy (11).*

E3 *If $\theta \neq \theta_0$, then there exists $A \subseteq (\mathbb{R}^+)^S$, $\lambda(A) > 0$, such that for one $i \in S$* and *all $x \in A$, $\lambda\left(\left\{y \in E \,|\, \lambda_i(y^i, x, \theta) \neq \lambda_i(y^i, x, \theta_0)\right\}\right) > 0$*

E4 *The functions $\theta \to \alpha_i(x,\theta)$, $\theta \to \lambda_W(x,\theta)$ are twice continuously differentiable for all x, i, j, W, and for all $1 \leq u, v \leq d$, and the absolute values of their first and second order derivatives satisfy (11).*

E5 *I_0 and J_0 are positive definite.*

E6 (H') *is satisfied and J_0 is positive definite.*

Proposition 9 *Let $\widehat{\theta}_T$ be the CPLE of θ for the autoexponential model. Then:*

*(i) under assumptions **E1**, **E2**, and **E3**, $\widehat{\theta}_T$ is consistent. If we add **E4** and **E5**, $\widehat{\theta}_T$ is asymptotically normal.*

*(ii) under assumptions **E1**, **E2**, **E4** and **E6**, $\widehat{\theta}_T$ is asymptotically normal.*

Proof: **E1** implies **C1** and **E4** implies **C2-**(i). **E3** implies **C3**. Then we just have to show that **E2** implies **C2-**(ii). The conditional density of $X_i(t)$ is given here by $\ln f_i(y_i, y^i, x; \theta) = \ln \lambda_i(y^i, x; \theta) - y_i \lambda_i(y^i, x; \theta)$. Then $\left|\ln f_i(y_i, y^i, x; \theta) - \ln f_i(y_i, y^i, x; \theta_0)\right| \leq \left|\ln \frac{\lambda_i(y^i,x;\theta)}{\lambda_i(y^i,x;\theta_0)}\right| + y_i \left|\lambda_i(y^i, x; \theta) - \lambda$

As $\left|\ln \frac{x}{y}\right| \leq |x-y|\,(x^{-1}+y^{-1})$ for all $x, y > 0$, we obtain $\left|\ln f_i(y_i, y^i, x; \theta) - \ln f_i(y_i, y^i, x; \theta_0)\right| \leq (\frac{2}{a} + y_i)\left|\lambda_i(y^i, x; \theta) - \lambda_i(y^i, x; \theta_0)\right|$. Finally, there exist two integers r and r' and four constants a_1, a_2, a_3, a_4 such that we can take $h(y, x) = \sum_{i \in S} 2(\frac{2}{a} + y_i)\{a_1 V_r(x) + a_2 + (a_3 V_{r'}(x) + a_4) \sum_{j \neq i} y_j\}$.

On the other hand, **E4** implies that the modulus of the first and second derivatives of $\lambda_i(y^i, x; \theta)$ with respect to θ are bounded by a square integrable function of x and y, and this ensures **N1**. Finally, **E5** is **N2**. In another hand, **E6** ensures **E3** and **E5** (see Proposition 7). $\square$

Example 4 *The conditional-exponential dynamics.*

For the particular model (6), all conditions **E** are fulfilled without any assumption on the parameters. Besides, **E6** is satisfied with $d = n + \frac{1}{2}\sum_{i \in S} |\partial i| + \sum_{i \in S} |\partial i^-|$, $\lambda_i(y^i, x, \theta) = {}^t\theta g_i(y^i, x)$, ${}^t\theta = {}^t\left((\delta_i)_{i \in S}, (\beta_{ij})_{\langle i,j \rangle_{\mathcal{G}}}, (\alpha_{ji})_{\langle j,i \rangle_{\mathcal{G}^-}}\right) \in \mathbb{R}^d$.

6 Model identification

We suppose that we want to fit a semi-causal MCMF dynamics model; first, we have to determine the graphs $\mathcal{G}$ and $\mathcal{G}^-$; then, we will estimate the parameters, and lastly validate the model. There are two possible strategies for the identification procedure, the choice depending on the complexity of the problem and on the number of sites. The first one is *global*: we could try to determine globally the graphs by CPL maximization, joined with a convenient "AIC" penalization criterion.

On the other hand, a less expensive procedure is to work *locally* site by site, providing estimations for ∂i and ∂i^- for each site $i \in S$. For this, we maximize the likelihood of $X_i(t)$, $t = 1, T$, conditionally to $X^i(t) = y^i$ ($|S| - 1$ sites) and $X(t-1) = x$ ($|S|$ sites) (the conditional distribution of $X_i(t)$ depends on $(2|S| - 1)$ sites) using a backward procedure; we choose a small signification level for the adequate statistics in order to keep only the most significant variables. This can be associated with a forward procedure if we want to take into account a particular information on the geometry of S. Further, we have to harmonize the instantaneous neighbourhood relation to get a symmetric graph $\mathcal{G}$: if $j \in \widehat{\partial i}$, we decide $i \in \widehat{\partial j}$. Thus we

generally get an overfitted model, and the next step is to reduce it by progressive elimination using a descending stepwise procedure. Finally, if we have got two or more models, we choose the one which minimizes the AIC (or BIC) criterion.

In a Gaussian context, the computation is particularly easy and fast, because it is linear and explicit. Using the partial correlation ρ_P, we have the characterizations (see Garber (1981; [15]); Guyon (1995), §1.4.): $\rho_P(X_j\ ,\ X_i\,|X_{L_i}) = 0 \Longleftrightarrow j \notin L_i = \partial i \cup \partial i^-$. This equation allows us to determine L_i by a fast linear stepwise procedure of the regression of $X_i(t)$ on the $(2|S|-1)$ other variables.

For other log linear models, the conditional log likelihood is concave so we can get the CPLE using a gradient algorithm. We then consider the general procedure given previously. An alternative is to follow the Gaussian approach, even if the model is not Gaussian. This can be done when the dimension of the parameter becomes large enough to make the results suspicious and slow progressing. This procedure is used in section 8.

7 Model validation

7.1 Some Central Limit Theorems

In the case $E \subseteq \mathbb{R}$, validation tests are based on the estimated conditional residuals. Let us denote θ_0 and $\widehat{\theta}_T$ the true value of the parameter and its CPLE respectively. Let us set:

$$\varepsilon_{it} = X_i(t) - \mu_{it} \quad \text{and} \quad \widehat{\varepsilon}_{it} = X_i(t) - \widehat{\mu}_{it} \tag{12}$$

where $\mu_{it}(\theta) = E[X_i(t)|X^i(t), X(t-1); \theta]$, $\mu_{it} = \mu_{it}(\theta_0)$. $\widehat{\mu}_{it} = \mu_{it}(\widehat{\theta}_T)$ is explicit in $y_{\partial i}$, $x_{\partial i^-}$, and $\widehat{\theta}_T$.

For an autodiscrete dynamics on a K-states space E, we will have an expression equivalent to (12), but with the $(K-1)$-dimensional encoded variable $Z_i(t)$ related to $X_i(t)$ (see section 7.2.4).

From (12), we propose a validation statistic; we derive its limit distribution using a Central Limit Theorem for martingales (Duflo (1997), Hall and Heyde (1980; [21])).

We build up two tests, based on the estimated residuals $\widehat{\varepsilon}_{it}$ and on the squared residuals $\widehat{\varepsilon}^2_{it}$.

The latter allows for the detection of possible variance deviation. Anyway, those tests are more useful to reject a model than to select the best one.

7.1.1 CLT for the residuals ε_{it}

We denote $A \perp B$ if the variables A and B are independent. Let $\mathcal{C}$ be a coding subset of S (for $\mathcal{G}$), $e_{it} = \frac{\varepsilon_{it}}{\sigma(\varepsilon_{it})}$ and $\widetilde{e}_t = \sum_{i \in \mathcal{C}} e_{it}$:
(i) e_{it} is zero mean and of variance 1;
(ii) $\varepsilon_{it} \perp X_j(t)$ if $i \neq j$ and $\varepsilon_{it} \perp X_j(t-1)$ for all $j \in S$.

Besides, for all i and j which are not neighbour sites, $X_i(t)$ and $X_j(t)$ are independent, conditionally to the past and the neighbourhood values on ∂i, ∂j. Then $e_{it} \perp e_{js}$ if $t \neq s$, for all $i, j \in S$, and conditionally to $X(t-1)$ and $X_{\overline{\mathcal{C}}}(t)$, $e_{it} \perp e_{jt}$ for all $i, j \in \mathcal{C}$, $i \neq j$;
(iii) as $E\left[e_{it}|\mathcal{F}_{t-1}\right] = E\left[E\left[e_{it} \mid y^i, x\right] \mathcal{F}_{t-1}\right] = 0$, $(e_{it})_{t \geq 0}$ is a square integrable martingale difference sequence w.r.t. the filtration $(\mathcal{F}_t = \sigma(X(s), s \leq t), t \geq 0)$. Then $\widetilde{e}_t = \sum_{i \in \mathcal{C}} e_{it}$ is also a square integrable martingale difference sequence.

We can thus apply a martingale's CLT (Duflo (1996), Corollary 2.1.10). For a square integrable martingale (M_t), let $(\langle M \rangle_t)$ be its increasing process defined by:

$$\langle M \rangle_t = \langle M \rangle_{t-1} + E[||M_t - M_{t-1}||^2 | \mathcal{F}_{t-1}] \text{ for } t \geq 1, \text{ and } \langle M \rangle_0 = 0.$$

The first condition to check is:

$$\frac{\langle M \rangle_T}{T} \underset{T \to \infty}{\overset{P}{\longrightarrow}} \mathbf{\Gamma}, \quad \text{for some positive } \Gamma. \tag{13}$$

We set $M_s = \sum_{t=1}^{s} \widetilde{e}_t$; we then have:

$$\langle M \rangle_t - \langle M \rangle_{t-1} = \sum_{i \in \mathcal{C}} E\left[e_{it}^2 | \mathcal{F}_{t-1}\right] = \sum_{i \in \mathcal{C}} E\left[E\left[e_{it}^2 | y^i, x\right] | \mathcal{F}_{t-1}\right] = |\mathcal{C}|$$

The second condition is the Lindeberg condition. It is implied by the following one:

$$\exists\ \alpha > 2 \text{ such that } E\left[|\widetilde{e}_t|^\alpha | \mathcal{F}_{t-1}\right] \text{ is bounded} \tag{14}$$

Therefore, (13) is fulfilled with $\Gamma = |\mathcal{C}|$,whatever the model, while (14) has to be checked in each particular case.

Proposition 10 *(Duflo (1997), Hall and Heyde (1980))*

Let ε_{it} be defined by (12), $\mathcal{C}$ a coding subset such that $\widetilde{e}_t$ fulfills (14). Then,

$$\frac{1}{\sqrt{|\mathcal{C}|T}} \sum_{t=1}^{T} \widetilde{e}_t \xrightarrow[T\to\infty]{\mathcal{D}} \mathcal{N}(0,1)$$

7.1.2 CLT for the squared residuals

Let us define

$$w_{it} = \frac{e_{it}^2 - 1}{\sigma(e_{it}^2)} \tag{15}$$

The w_{it} 's have the same properties of independency as the e_{it}'s. Hence (w_{it}) is a martingale difference sequence in t, and so is $\widetilde{w}_t = \sum_{i\in\mathcal{C}} w_{it}$, for $\mathcal{C}$ a coding subset of S. Let $N_T = \sum_{t=1,T} \widetilde{w}_t$. As $\langle N\rangle_t - \langle N\rangle_{t-1} = \sum_{i\in\mathcal{C}} E\left[w_{it}^2|\mathcal{F}_{t-1}\right] = |\mathcal{C}|$, the first condition (13) for the CLT is fulfilled. The second condition will be ensured under:

$$\exists\ \alpha > 2 \text{ such that } E\left[|\widetilde{w}_t|^{\alpha}\,|\mathcal{F}_{t-1}\right] \text{ is bounded} \tag{16}$$

Proposition 11 *Let $\mathcal{C}$ be a coding subset of S , w_{it} be defined by (15) s.t. $\widetilde{w}_t$ fulfills (16). Then, $\frac{1}{\sqrt{|\mathcal{C}|T}} \sum_{i\in\mathcal{C}} \sum_{t=1}^{T} w_{it} \xrightarrow[T\to\infty]{\mathcal{D}} \mathcal{N}(0,1)$*

As we do not know θ_0, we apply the previous results to the residuals calculated with $\widehat{\theta}_T$. When CPLE is convergent, standard manipulations show that Propositions 10 and 11 still remain valid for the estimated residuals.

7.2 Applications

We give the explicit results for the auto-models studied in section 4. The proofs of the conditions (14) or (16) are given in Appendix 5.

7.2.1 The autoexponential dynamics

We suppose that we are in the framework of (6), assuming **E1**, **E2**, **E3**. Then $\widehat{\theta}_T$ is convergent and we have:

$$\frac{1}{\sqrt{|\mathcal{C}|T}} \sum_{i\in C} \sum_{t=1}^{T} \widehat{e}_{it} \xrightarrow[T\to\infty]{\mathcal{D}} \mathcal{N}(0,1) \text{ and } \frac{1}{2\sqrt{2|\mathcal{C}|T}} \sum_{i\in C} \sum_{t=1}^{T} (\widehat{e}_{it}^2 - 1) \xrightarrow[T\to\infty]{\mathcal{D}} \mathcal{N}(0,1)$$

7.2.2 The autopoissonian dynamics

We consider the framework of (8) and we suppose that the CPLE $\widehat{\theta}_T$ is consistent. Then for $\lambda_i = \lambda_i(X^i(t), X(t-1))$, $\gamma^2 = \sum_{i\in\mathcal{C}} E[\lambda_i]$ and $\sigma^2 = \sum_{i\in\mathcal{C}} E[\lambda_i\,(\lambda_i+1)\left({\lambda_i}^2+5\lambda_i+1\right)]$, we have:

$$\frac{1}{\sqrt{T}}\sum_{i\in C}\sum_{t=1}^{T}\widehat{\varepsilon}_{it} \xrightarrow[T\to\infty]{\mathcal{D}} \mathcal{N}(0,\gamma^2) \text{ and } \frac{1}{\sqrt{T}}\sum_{i\in C}\sum_{t=1}^{T}\left(\widehat{\varepsilon}_{it}^2-\lambda_i(y^i,x)\right) \xrightarrow[T\to\infty]{\mathcal{D}} \mathcal{N}(0,\sigma^2)$$

7.2.3 The autologistic dynamics

For E=$\{0,1\}$, we consider the framework of (9) and we assume:

B1 : For $\theta\neq\theta_0$, $\exists\ x,\ y\in E$ s.t., for an $i\in S$, $p_i(y^i,x,\theta)\neq p_i(y^i,x,\theta_0)$.

Under assumption **B1**, we have:

$$\frac{1}{\sqrt{|\mathcal{C}|T}}\sum_{i\in\mathcal{C}}\sum_{t=1}^{T}\widehat{e}_{it} \xrightarrow[T\to\infty]{\mathcal{D}} \mathcal{N}(0,1) \text{ and } \frac{1}{\sqrt{|\mathcal{C}|T}}\sum_{i\in\mathcal{C}}\sum_{t=1}^{T}\widehat{e}_{it}.sign(\widehat{e}_{it}) \xrightarrow[T\to\infty]{\mathcal{D}} \mathcal{N}(0,1)$$

7.2.4 The autodiscrete dynamics

More generally, if E is qualitative, $E=\{a_0,a_1,\cdots,a_{K-1}\}$, we consider the encoded $(K-1)$-dimensional variable $Z_{it}\in\{0,1\}^{K-1}$ linked to the $X_i(t)$ by:

$$Z_{itl}=1 \text{ if } X_i(t)=a_l,\ \ Z_{itl}=0 \text{ elsewise},\ 1\leq l\leq K-1.$$

We suppose that the conditional energy is given by $U(z_t|z_{t-1}) = \sum_{i\in S}\sum_{l=1,K-1} z_{itl}(\delta_{i,l} + \sum_{k=1,K-1}\alpha_{i,lk}z_{i(t-1)k}) + \sum_{\langle i,j\rangle}\sum_{l=1,K-1}\sum_{k=1,K-1}\beta_{ij,kl}z_{itl}z_{jtk}$. We have:

$$\frac{1}{\sqrt{(K-1)|\mathcal{C}|T}}\sum_{i\in\mathcal{C}}\sum_{t=1}^{T}\left({}^t\varepsilon_{it}\,C_{it}^{-1}\,\varepsilon_{it}-(K-1)\right) \xrightarrow[T\to\infty]{\mathcal{D}} \mathcal{N}(0,1)$$

where $C_{it}=(C_{it})_{kl}$, $1\leq k,l\leq K-1$, with $(C_{it})_{kk}=p_{ik}(z_t^i,z_{t-1})(1-p_{ik}(z_t^i,z_{t-1}))$, $(C_{it})_{kl}=(C_{it})_{lk}=-p_{ik}(z_t^i,z_{t-1})p_{il}(z_t^i,z_{t-1})$ if $l\neq k$, and

$$p_{il}(z_t^i,z_{t-1})=\frac{\exp y_{itl}}{1+\sum_{l=1,K-1}\exp z_{itl}\,y_{itl}} \tag{17}$$

where $y_{itl}=\delta_{i,l}+\sum_{k=1,K-1}\alpha_{i,lk}z_{i(t-1)k}+\sum_{j\in\partial i}\sum_{k=1,K-1}\beta_{ij,kl}z_{jtk}$.

7.2.5 The autonormal dynamics

In the framework of (7), we suppose that the ergodicity condition is fulfilled. We have:

$$\frac{1}{\sqrt{|\mathcal{C}|T}} \sum_{i\in C} \sum_{t=1}^{T} e_{it} \underset{T\to\infty}{\overset{\mathcal{D}}{\longrightarrow}} \mathcal{N}(0,1) \text{ and } \sum_{i\in\mathcal{C}} \sum_{t=1}^{T} 2\gamma_i \varepsilon_{it}^2 \underset{T\to\infty}{\overset{\mathcal{D}}{\longrightarrow}} \chi^2_{|C|T}$$

and so,

$$\frac{1}{2\sqrt{|\mathcal{C}|T}} \sum_{i\in\mathcal{C}} \sum_{t=1}^{T} \left(2\gamma_i \varepsilon_{it}^2 - 1\right) \underset{T\to\infty}{\overset{\mathcal{D}}{\longrightarrow}} \mathcal{N}(0,1)$$

A third result holds, which uses the joint distribution on S of all the residuals. The ε_{it}'s covariances are $\mathrm{Cov}(\varepsilon_{it}, \varepsilon_{jt}) = \frac{1}{2\gamma_i}$ if $i = j$, $= -\frac{\beta_{ij}}{2\gamma_i\gamma_j}$ if $j \in \partial i$, and 0 else. Let C_ε be the $n \times n$ covariance matrix of $\varepsilon_t = (\varepsilon_{it},\ i \in S)$. Then $\sum_{t=1,T} {}^t\varepsilon_t\, C_\varepsilon^{-1}\, \varepsilon_t \sim \chi^2_{nT}$ and thus

$$\frac{1}{\sqrt{nT}} \sum_{t=1}^{T} {}^t\varepsilon_t\, C_\varepsilon^{-1}\, \varepsilon_t \underset{T\to\infty}{\overset{\mathcal{D}}{\longrightarrow}} \mathcal{N}(0,1)$$

Let $\widehat{\theta}_T$ be a consistent estimator of θ_0 and $\widehat{C}_\varepsilon$ is the estimate of C_ε obtained by replacing the parameters by their estimates. Then, the three statistics T_1, T_2, T_3 defined below are asymptotically Gaussian $\mathcal{N}(0,1)$ and can be used for validation tests of the model:

$$T_1 = \frac{1}{\sqrt{|\mathcal{C}|T}} \sum_{i\in\mathcal{C}} \sum_{t=1}^{T} \widehat{e}_{it}\ ;\ T_2 = \frac{1}{2\sqrt{|\mathcal{C}|T}} \sum_{i\in\mathcal{C}} \sum_{t=1}^{T} \left(2\widehat{\gamma}_i \widehat{\varepsilon}_{it}^2 - 1\right)\ ;\ T_3 = \frac{1}{\sqrt{nT}} \sum_{t=1}^{T} {}^t\widehat{\varepsilon}_t\, \widehat{C}_\varepsilon^{-1}\, \widehat{\varepsilon}_t$$

8 MCMF modelling of meteorological data

We have tried an MCMF modelling on a real set of meteorological data. The data comes from the study of Tang (1991) and consists of daily pluviometric measures on a network of 16 sites in the Mekong Delta (Vietnam). We have retained a period of 123 consecutive days from July to October 1983. Geographically, the 16 meteorological stations, situated at latitude 9.17 to 10.8 and longitude 104.48 to 106.67, are not regularly located (see table 1).

The data have been previously studied by Tang (1991); he proposed different models for each site, each being free of the other sites; this seems to be insufficient and hardly satisfactory.

We study an MCMF autologistic model; we consider the binary {0,1} data where 1 stands for rain and 0 for no rain. The results lead to an average of exact prediction of about 77%,

which is rather satisfactory noting that the random part is large in this kind of meteorological phenomenon. We compare this model with a site-by-site Markov Chain, for which the prediction results are bad.

Of course, we do not pretend that our models are definitive and we know they need to be refined for effective forecasting. An interesting study would be to fit other competitive models of more or less the same dimension, as an Hidden Markov Chain (see e.g. .Zucchini and Guttorp 1991 [34], MacDonald and Zucchini 1997, [35]), the hidden states standing for the underlying climatic stage, or threshold models, the dynamics at the time t depending on the position of the state at the time $t-1$ with respect to thresholds (see [32] and references herein). Also, these models should take into account the dual character of such a data, i.e. the binary feature $\{0\} \cup \mathbb{R}^{+*}$ of the space state. This is work in progress.

The first task is to identify the two dependency graphs. Considering the large number of parameters involved in the CPL procedure, we use the Gaussian linear procedure with the original data (see section 6); we fit a regression model on each site, with respect to the 15 other variables at the same time and all the 16 variables at the previous time. We then select the neighbours as the variables giving the best multiple correlation coefficient R^2 in a stepwise procedure, taking into account the symmetry of the instantaneous graph together with a principle of parsimony. We give in table 2 the neighbourhoods and the R^2 obtained from all the 31 variables (denoted R^2_{31}) and the R^2 calculated on the selected neighbours (denoted $R^2_{\partial i, \partial i^-}$). The instantaneous graph has 25 links and the time-delay graph 17 links. We see that for each site, the instantaneous neighbourhood may contain few sites, while there are at most 3 sites making up the time delay neighbourhood. We note that if we draw the directed graph $\mathcal{G}_-$ on a geographical map, there is a main direction of the arrows, which is roughly S-N.

(include here table 1 and table 2)

Going back to the binary data, we estimate the autologistic model (9): for each site i, $X_i(t)$ has a Bernouilli distribution of parameter $p_i(t)$, conditionally to $X_{\partial i}(t)$ and $X_{\partial i^-}(t)$, with $p_i(t) = (1+\exp -\lambda_i(t))^{-1}$ and $\lambda_i(t) = \delta_i + \sum_{j\in\partial i} \beta_{ij} X_j(t) + \sum_{i\in\partial i^-} \alpha_{li} X_l(t-1)$. The dimension of the model is 16+25+17=58. We first estimate the parameters site-by-site, maximizing the log

conditional pseudo-likelihoods; then, we calculate the validation statistics based on the square residuals (see section 7.2): $V_i = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \hat{e}_{it}.sign(\hat{e}_{it})$ for each site i. Secondly, we proceed to the global estimation on the basis of the global pseudo-likelihood; we take as initial values the first site-by-site estimates for the parameters δ_i and α_{li}, and the means of the two previous estimations β_{ij} and β_{ji} (theoretically equal) for the instantaneous interaction parameters. We do not give here the results for the 58 parameters for sake of place, but we summarize the results: concerning the statistics V_i, each local model is accepted. The final (global) estimation provides results close to the individual ones. The validation statistic $V = \frac{1}{\sqrt{\mathcal{C}T}} \sum_{i \in \mathcal{C}} \sum_{t=1}^{T} \hat{e}_{it}\, sign(\hat{e}_{it})$ computed for the coding set $\mathcal{C} = \{3, 4, 5, 7, 8, 10, 11, 14\}$ (of maximal size 8) is equal to -0.1186 and we do not reject the model.

Another way to validate the model is to compare the true data with what is predicted. So we compute the predicted values (maximizing the local conditional probabilities), first site by site (with the local estimates) and then globally (with the global estimates). Table 3 gives the percentages of similarity which spread out from 68.03% (site 10) to 84.43% (site 2), with a mean of 77.36% (for the global parameter's estimate).

(include here table 3)

We compare our model with a site by site Markov chain. Of course, this alternative is very poor; as expected, the predictions are bad, spreading out between 35.25% and 59.02% with a mean equal to 51.84% (see Table 3). In conclusion, the autologistic model leads to a relatively correct forecasting; it could be better if we increase the number of links or the dependence in time, at the expense of the rising parametric dimension.

9 Appendices

9.1 Appendix 1: the invariant law π of an MCMF is generally not Markovian

We illustrate this with two examples.

(1) *The one dimensional marginal of a two dimensional Gaussian Markov field is not anymore a Markov process.* Let us consider a centered Gaussian Markov isotropic field with respect

to the four-nearest neighbours over $\mathbb{Z}^2$ (see Besag, 1974 [10], Ripley, 1981 [30] and Guyon,1995 [18] §1.3.4.)

$$X_{st} = \alpha(X_{s-1,t} + X_{s+1,t} + X_{s,t-1} + X_{s,t+1}) + e_{st}, \quad |\alpha| < \frac{1}{4}$$

with $E[X_{st}e_{s't'}] = 0$ if $(s,t) \neq (s',t')$. The spectral density of X is $f(\lambda,\mu) = \sigma_e^2(1 - 2\alpha(\cos\lambda + \cos\mu))^{-1}$. The one index field $(X_{s,0},\ s \in \mathbb{Z})$ has spectral density

$$F(\lambda) = 2\int_0^\pi f(\lambda,\mu)d\mu = 2\pi\sigma_e^2[(1-2\alpha\cos\lambda)^2 - 4\alpha^2]^{-\frac{1}{2}}$$

which cannot be written $\frac{1}{Q(e^{i\lambda})}$ or $\frac{1}{|P(e^{i\lambda})|^2}$ as for a Markovian or an AR process. The distribution π of $(X_{s,0},\ s \in \mathbb{Z})$ is not Markovian.

(2) *Conditional Ising model*: $\Omega = \{0,1\}^S$, $S = \{1,2,\cdots,n\}$ is the one dimensional torus with the agreement $n+1 \equiv 1$, and $P(x,y) = Z^{-1}(x)\exp\{\alpha\sum_{i\in S} x_iy_i + \beta\sum_{i\in S} y_iy_{i+1}\}$. This chain is reversible and the invariant distribution can be written $\pi(y) = CZ(y)\exp\{\gamma\sum_{i\in S} y_i + \beta\sum_{i\in S} y_iy_{i+1}\}$ where $Z(y)$ is the normalizing constant of $P(.\mid y)$. It is easy to see that $\log\pi(y)$ has a non zero potential Φ_S on the whole set S.

9.2 Appendix 2: Two examples where marginals in y are not local in x

(1) *Binary state space.* Let us consider a binary chain on the one dimensional torus $S = \{1,2,\cdots,n\}$ (with $n+1 \equiv 1$) with the transition

$$P(x,y) = Z^{-1}(x)\exp\{\alpha\sum_{i=1}^n x_iy_i + \beta\sum_{i=1}^n y_iy_{i+1} + \gamma\sum_{i=1}^n y_i\}$$

Then $P_i(y_i|x) = \sum_{y^i} P(x,y_i,y^i) = \frac{e^{y_i(\alpha x_i+\gamma)}\sum_{y_j,j\neq i} e^{\beta y_i(y_{j-1}+y_{j+1})+A_i}}{\sum_{y_j,j\neq i} e^{A_i} + e^{(\alpha x_i+\gamma)}\sum_{y_j,j\neq i} e^{\beta(y_{j-1}+y_{j+1})+A_i}}$, with $A_i = \beta\sum_{j\neq i,i-1} y_jy_{j+1} + \sum_{j\neq i}(\alpha x_j+\gamma)y_j$. This distribution depends on x on all sites.

(2) *A Gaussian example.*

We take $P(x,y) = Z^{-1}(x)\exp U(y|x)$ with $U(y|x) = (y-m(x))'Q(y-m(x))$, where Q is symmetric, definite positive, and $E[Y|x] = m(x)$. For $-U(y\mid x) = \sum_{i=1}^n(\gamma_iy_i^2+\delta_iy_i) + \sum_{i=1}^n \alpha_ix_iy_i + \sum_{i,j=1,<i,j>}^n \beta_{ij}y_iy_j$, $m(x) = \frac{Q^{-1}}{2}(\delta+\alpha x)$ where $\delta = (\delta_i)_{i\in S}$, $\alpha x = (\alpha_ix_i)_{i\in S}$, and $Var[Y|x] = \frac{1}{2}Q^{-1}$: $V(Y\mid x)^{-1} = 2Q$ is local in x but it is easy to find parameters for which each component $m_l(x)$ of $m(x)$ depends on all $x_i, i \in S$.

9.3 Appendix 3: An example of a reverse MCMF non MCMF

We consider $S = \{1, 2, \cdots, n\}$, $X = X(0)$, $Y = X(1)$ and we suppose that $Z = (X, Y)$ is a $2n$-dimensional centered Gaussian variable with covariance $\Sigma = \begin{pmatrix} \Delta & B \\ {}^tB & \Delta \end{pmatrix}$. Then, $(Y \mid x) \sim \mathcal{N}_n(Ax, Q)$, where $A = {}^tB\Delta^{-1}$ and $Q = \Delta - {}^tB\Delta^{-1}B$. We fix Δ and look at B as a parameter (s.t. Σ is definite positive).

If we want conditional independence for $\{(X_i(1) \mid X(0) = x), i = 1, n\}$, we force Q to be diagonal; this involves $\frac{n(n-1)}{2}$ conditions on B, for n^2 degrees of freedom.

On the other hand, $(X \mid y) \sim \mathcal{N}_n(Ty, R)$ where $T = B\Delta^{-1}$ and $R = \Delta - B\Delta^{-1}{}^tB$. We can chose B such that for some i, $R_{ij} \neq 0$ for any j. In such a case, $(X_i(0) \mid x^i, y)$ depends on x^i on whole $S \backslash \{i\}$ and the reverse chain is not an MCMF.

For example, this happens for $n = 2$ if we take $\Delta = \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix}$, $B = \begin{pmatrix} \frac{1}{4} & \frac{3}{4} \\ -\frac{1}{2} & 0 \end{pmatrix}$.

9.4 Appendix 4: Parametric identifiability; regularity of I_0 and J_0

First we take $\nu^{*S} = \otimes_{i \in S} K_i \nu$ as the reference measure, where K_i is defined in (10).

Identifiability. We suppose that the conditional densities are equal for θ and θ'. Then, for each $i \in S, x$ and y^i we have

$$^t(\theta - \theta')g_i(y_i, y^i, x) = \Psi(\theta, y^i, x) - \Psi(\theta', y^i, x)$$

But the right member is 0 as $g_i(0, y^i, x) = 0$. Identifiability follows because the $d \times d$ matrix g is regular.

Regularity of I_0. First we note that for the stationary distribution

$$I_0 = E_{X(t-1), X^i(t)}[\sum_{i \in S} Var_{X_i(t)}\{g_i(X_i(t), X^i(t), X(t-1))\}].$$

Then, under (H'), we have

$$\sum_{i \in S} Var[g_i(Y_i, y^i, x) \mid y^i, x)] \geq a \sum_{i \in S} G_i(x, y^i)\, {}^tG_i(x, y^i).$$

As the density of $(X(t-1), X(t))$ is strictly positive anywhere under **C1** (see Duflo (1997), proposition 6.1.9), I_0 is regular.

Regularity of J_0. Let us recall that $J_0 = Var_\theta\{g(X,Y)\}$ where $(X,Y) = (Y(0), Y(1))$, $g(x,y) = \sum_{i\in S} g_i(x,y)$, $g_i(x,y) = [\log f_i(y_i, y^i, x, \theta)]_\theta^{(1)}$. We follow an idea given by Jensen and Künsch (1994). We suppose that there exists a "*strong coding subset*" $\mathcal{C} \subseteq S$ in the following sense:

(i) there exists a partition $\{S_j, j \in \mathcal{C}\}$ of S s.t. $j \in S_j$.

(ii) For $j \in \mathcal{C}$, let us define $G_j = \sum_{i\in S_j} g_i$ and let $\mathcal{F}$ be the σ-field generated by $X = Y(0)$ and $\{Y_i(1), i \notin \mathcal{C}\}$. Then, conditionally to $\mathcal{F}$, the variables $\{G_j,\ j \in \mathcal{C}\}$ are independent.

A sufficient condition for (ii) is that for each $j, l \in \mathcal{C}$, $j \neq l$, $l \notin S_j \cup \partial S_j$. For example, for the 2-dimensional torus $T = \{1, 2, \cdots, 3K\}^2$ and the 4-nearest neighbours vicinity, $\mathcal{C} = \{3(m,n), m, n = 1, K\}$ is a strong coding subset and $S_{m,n} = \{(i,k) : |i-m| + |k-n| \leq 2\}$.

As $g = \sum_{j\in\mathcal{C}} G_j$, and $J_0 = Var\, g(X,Y) \geq E_{\mathcal{F}}(Var\{g(X,Y) \mid \mathcal{F}\})$, we have, as a consequence of (ii):

$$J_0 \geq \sum_{j\in\mathcal{C}} E_{\mathcal{F}}(Var\{G_j(X,Y) \mid \mathcal{F}\}) = G_0$$

Then a sufficient condition ensuring that J_0 is p.d. is that G_0 is p.d.. Such a verification has to be done. For example, if S is the 1-dimensional torus with $n = 3K$ sites, with energy $U(y|x) = \alpha \sum_{i\in S} y_i v_i,\ v_i = x_i + (y_{i-1} + y_{i+1}), y_i \in \{0,1\}$, we can take $\mathcal{C} = \{3j,\ j = 1, K\}$ and $S_j = \{3j-1, 3j, 3j+1\}$. As the model is homogeneous in the space, and as $(g_2 + g_3 + g_4 \mid \mathcal{F})$ is never constant, $G_0 \geq E_{\mathcal{F}}(Var(g_2 + g_3 + g_4) \mid \mathcal{F}) > 0$.

9.5 Appendix 5: Validation tests for MCMF models

9.5.1 The autoexponential dynamics (6)

$e_{it} = \lambda_i(y^i, x)X_i(t) - 1$, and $w_{it} = 2^{-\frac{3}{2}}(e_{it}^2 - 1)$. Next, for any $\alpha > 2$, as $(a+b)^\alpha \leq 2^\alpha(a^\alpha + b^\alpha)$, we have

$$E\left[|e_{it}|^\alpha | y^i, x\right] \leq 2^\alpha \lambda_i(y^i, x)^\alpha \left\{ \frac{\Gamma(\alpha+1)}{\lambda_i(y^i,x)^\alpha} + \lambda_i(y^i,x)^{-\alpha} \right\} \leq 2^\alpha(\Gamma(\alpha+1) + 1)$$

and (14) is fulfilled.

Besides, $E\left[|w_{it}|^{\alpha}|y^i,x\right] \leq 2^{-\frac{\alpha}{2}}\left\{2^{2\alpha}(\Gamma(2\alpha+1)+1)+1\right\}$ and we get (16).

9.5.2 The autopoissonian dynamics (8)

The conditional distribution of $X_i(t)$ is Poisson with mean $\lambda_i(y^i,x) = \exp\{\delta_i + \sum_{l\in\partial i^-} \alpha_{li}x_l + \sum_{j\in\partial i} \beta_{ij}y_j\}$. We look at conditions (13) and (14) for the ε_{it}. First we have $\langle M\rangle_t - \langle M\rangle_{t-1} = \sum_{i\in\mathcal{C}} \lambda_i(X^i(t), X(t-1)) \leq e^M|\mathcal{C}|$. Next, for any integer α, and some constants

$$C_{k,\alpha}, E\left[\varepsilon_{it}^{\alpha}|y^i,x\right] \leq 2^{\alpha}(\sum_{k=1,\alpha} C_{k,\alpha}\lambda_i(y^i,x)^k + \lambda_i(y^i,x)^{\alpha}) \leq 2^{\alpha}(\sum_{k=1,\alpha} C_{k,\alpha}e^{Mk} + e^{M\alpha}).$$

Next, we consider the squared residuals $v_{it} = \varepsilon_{it}^2 - \lambda_i(y^i,x)$ and $\widetilde{v}_t = \sum_{i\in\mathcal{C}} v_{it}$. As previously, we can bound $E\left[|\varepsilon_{it}^2 - \lambda_i(y^i,x)|^{\alpha}|y^i,x\right]$ for any $\alpha > 2$, which implies (14). And we have (16) with $\langle N\rangle_t - \langle N\rangle_{t-1} = E\left[|\widetilde{v}_t|^2|\mathcal{F}_{t-1}\right] \leq |\mathcal{C}|(e^{4M} + 6e^{3M} + 6e^{2M} + e^M)$. We get the announced result.

9.5.3 The autologistic dynamics (9)

$e_{it} = \frac{X_i(t)-p_i(y^i,x)}{\sqrt{p_i(y^i,x)(1-p_i(y^i,x))}}$, and $w_{it} = e_{it}.sign(e_{it})$. We can easily check that $E\left[|e_{it}|^{\alpha}\,|\,y^i,x\right]$ as $E\left[|w_{it}|^{\alpha}\,|\,y^i,x\right]$ are bounded for any $\alpha > 2$. For example, if the conditional energy is $U(y|x) = \sum_{i\in S} y_i(\delta_i + \alpha_i x_i) + \sum_{\langle i,j\rangle} \beta_{ij}y_iy_j$, then $p_i(y^i,x) = (1+\exp-[\delta_i + \alpha_i x_i + \sum_{j\in\partial i} \beta_{ij}y_j])^{-1}$ and **B1** is satisfied.

9.5.4 The autodiscrete dynamics

For each l, Z_{itl} has a conditional Bernouilli distribution of parameter $p_{il}(y^i,x)$ given by (17). So, $\varepsilon_{it} = Z_{it} - E\left[Z_{it}\,|y^i,x;\theta_0\right]$ has conditional variance C_{it}. Conditionally to the past and $X_{\overline{\mathcal{C}}}(t)$, $\varepsilon_{it} \perp \varepsilon_{jt'}$ for all $i,j \in S$ if $t \neq t'$; and $\varepsilon_{it} \perp \varepsilon_{jt}$ if i,j are elements of a coding subset $\mathcal{C}$ of S; thus $\{\varepsilon_{it}\}_t$ is a martingale's increment. Besides, $E\left[{}^t\varepsilon_{it}\,C_{it}^{-1}\,\varepsilon_{it}\,|\,y^i,x\right] = K-1$.

Finally, the state space E being finite and ε_{it} belonging to $\{0,1\}^{K-1}$, all the moments of ε_{it} are bounded and the condition on the increasing process is fulfilled.

9.5.5 The autonormal dynamics

We consider (7). The conditional distribution of $X_{it}(t)$ is Gaussian $\mathcal{N}(\mu_{it}, \frac{1}{2\gamma_i})$ with $\mu_{it} = \frac{1}{2\gamma_i}(\delta_i + \sum_{l \in \partial i^-} \alpha_{li} x_l + \sum_{j \in \partial i} \beta_{ij} y_j)$. We immediately deduce: $e_{it} = \frac{X_i(t) - \mu_{it}}{\sqrt{2\gamma_i}} \sim \mathcal{N}(0,1)$, and $e_{it}^2 = 2\gamma_i \varepsilon_{it}^2 \sim \chi_1^2$.

References

[1] Amemiya T. (1985) *Advanced econometrics*, Blackwell.

[2] Arnold B.C. and Press J. (1989), Compatible conditional distributions. *JASA, **84**, 405, 152-156.*

[3] Arnold B.C. and Strauss D. (1988), Bivariate distribution with exponential conditionals. *JASA, **83**, 402, 522-527.*

[4] Arnold B.C., Castillo E. and Sarabia J.M. (1991), Conditionally specified distributions. *L.N.S. **74**, Springer.*

[5] Bayomog S. (1994), *Modélisation et analyse des données spatio-temporelles.* PhD Thesis, Univ. d'Orsay, France.

[6] Bayomog S., Guyon X., Hardouin C. et Yao J.F. (1996), Test de différence de contraste et somme pondérée de khi-deux. *Can. J. Stat., **24**, 1, 115-130.*

[7] Bennett R.J. and Haining H. (1985), Spatial structure and spatial interaction: modelling approaches to the statistical analysis of geographical data. *JRSS **148**, 1-36.*

[8] Besag J. (1974), Spatial interaction and the statistical analysis of lattice systems. *JRSS B, **36**, 192-236.*

[9] Besag J. (1974), On spatial temporal models and Markov fields. *7th Prague conference and of European Meeting of Statisticians, 47-55.*

[10] Besag J. (1977), Efficiency of pseudo likelihood estimation for simple Gaussian fields. *Biometrika, **64**, 616-618.*

[11] Chadoeuf J., Nandris D., Geiger J.P., Nicole M., Piarrat J.C. (1992), Modélisation spatio temporelle d'une épidémie par un processus de Gibbs: estimation et tests. *Biometrics, **48**, 1165-1175.*

[12] Dacunha-Castelle D. and Duflo M. (1986), *Probability and statistics*, *Vol. 2.* Springer Verlag.

[13] Duflo M. (1997), *Random iterative models.* Springer.

[14] Durrett R. (1995), Ten lectures on particles systems, Saint-Flour 1993. *L.N.M. 1608, 97-201,* Springer.

[15] Garber D.D. (1981), *Computational models for texture analysis and texture synthesis.* PhD Thesis, Univ. South California.

[16] Geyer C. (1999), *Likelihood inference for spatial point processes, in Stochastic geometry, likelihood and computation (Chapter 3).* Barndorff-Nielsen, Kendall and Van Lieshout Eds. Chapman and Hall.

[17] Geyer C.J. and Thompson E.A. (1992), Constrained Monte Carlo maximum likelihood for dependent data (with discussion). *JRSS,B* **54**, 657-699.

[18] Guyon X. (1995), *Random Fields on a Network: modeling, statistics and applications.* Springer.

[19] Guyon X. and Hardouin C. (1992), The chi2 difference of coding test for testing Markov Random Field hypothesis. *L.N.S. 74,* Springer*, 165-167.*

[20] Guyon X. and Künsch H.R. (1992), Asymptotic comparison of estimators in the Ising model. *L.N.S.* 74, Springer, 177-198.

[21] Hall P. and Heyde C.C. (1980), *Martingale limit theory and its applications.* Acad. Press.

[22] Jensen J.L. and Künsch H.R. (1994), On asymptotic normality of pseudo-likelihood estimate for pairwise interaction processes. *Ann. Inst. Statist. Math.*, **46**, 475-486.

[23] Keiding N. (1975), Maximum likelihood estimation in the Birth and death process. *Ann. of Stats. **3**, 363-372.*

[24] Koslov O. and Vasilyev N. (1980), Reversible Markov chain with local interaction*, in Multicomponent Random Systems.* Eds. Dobrushin and Sinai, Dekker.

[25] Künsch H.R. (1984), Time reversal and stationary Gibbs measures. *Stoch. Proc. and Applications,* ***17****, 159-166.*

[26] Pfeifer P.E. and Deutsch S.J. (1980), Identification and interpretation of first order for Space-time ARMA models. *Technometrics,* ***22****, 397-408.*

[27] Pfeifer P.E. and Deutsch S.J. (1980), A three stage iterative procedure for space time modelling. *Technometrics,* ***22****, 35-47.*

[28] Preston C. (1974), Gibbs states on countable set. *Cambridge tracts in math 68.*

[29] Prum B. (1986), *Processus sur un reseau et mesure de Gibbs. Applications.* Masson.

[30] Ripley B. (1981), *Spatial Statistics.* Wiley.

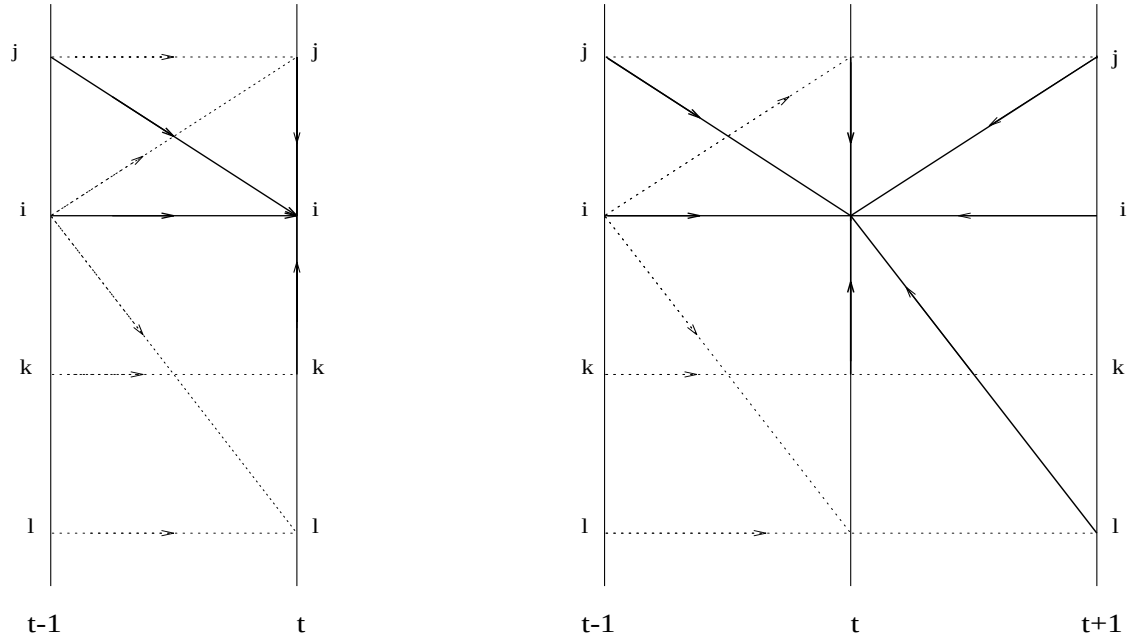
[31] Tang van Tham (1991), *Rainfall greenhouse effect in the delta of Mekong.* PhD Thesis. Univ. of Cantho. Vietnam.

[32] Tong H. (1990), *Non linear time series. A dynamical system approach.* Clarendon Press Oxford.

[33] Younes L. (1988), Estimation and Annealing for Gibssian fields. *Ann. I.H.P.* ***2****, 269-294.*

[34] Zucchini W. and Guttorp P. (1991), A hidden Markov model for space time precipitation. *Water resour. Res.* **27**, 1917-1923.

[35] MacDonald I. and Zucchini W. (1997), *Hidden markov and other models for discrete valued time series.* Monographs on statistic and applied probability. Chapman and Hall.



Semi-causal representation of a MC

The equivalent non causal representation of MF

Figure 1:

Site number	Name	Latitude	Longitude
1	Tan Chau	10.80	105.25
2	My Tho	10.35	106.37
3	Chau Doc	10.70	105.01
4	Can Tho	10.03	105.78
5	Soc Tran	09.60	105.97
6	Vinh Long	10.25	105.97
7	Sa Dec	10.30	105.75
8	Go Cong	10.35	106.67
9	Ca Mau	09.17	105.17
10	Long Xuyen	10.40	105.42
11	Rach Gia	10.00	105.08
12	Ha Tien	10.38	104.48
13	Cao Lanh	10.47	105.63
14	Moc Hoa	10.75	105.92
15	Vi Thanh	09.77	105.45
16	Tan An	10.53	106.40

Table 1: 16 meteorological stations in the Mekong Delta.

Site i	∂i	∂i^-	R^2_{31}	$R^2_{\partial i,\partial i^-}$
1	$\{2,3,8,14\}$		0.611	0.496
2	{1,5,6,7,8,15,16}	{14}	0.648	0.574
3	{1,12}	{1}	0.741	0.638
4	{6}	{7,10}	0.319	0.169
5	{2,9}	{4}	0.297	0.139
6	{2,4,7,10,14}	{7}	0.589	0.484
7	{2,6,13,15,16}	{6}	0.633	0.531
8	{1,2,15}		0.592	0.370
9	{5,11}	{4,9,12}	0.589	0.481
10	{6,12,13}	{4,5}	0.463	0.348
11	{9,15}	{11}	0.596	0.451
12	{3,10}	{11}	0.648	0.528
13	{7,10}	{4}	0.463	0.321
14	{1,6,16}	{4}	0.441	0.264
15	{2,7,8,11}	{10}	0.568	0.476
16	{2,7,14}		0.527	0.279

Table 2: neighbourhoods and R^2 statistics.

R^2_{31} : regression on all variables

$R^2_{\partial i,\partial i^-}$: regression on the neighbourhoods variables

Site	1	2	3	4	5	6	7	8
autologistic model from global estimates	74.59	84.43	72.95	84.43	82.79	81.97	82.79	71.31
Markov Chain	45.90	35.25	49.18	52.46	53.28	47.54	59.02	56.56
Site	9	10	11	12	13	14	15	16
autologistic model from global estimates	77.05	68.03	78.69	72.13	81.15	78.69	72.13	74.59
Markov Chain	57.38	54.92	51.64	57.38	50.00	51.64	50.00	57.38

Table 3: percentages of similarity between the real data and the predicted values for the autologistic model and the site by site Markov chain.