Markov connected component fields

by

Jesper Møller
Department of Mathematics and Computer Science, Institute of Electronic Systems, University of Aalborg, Fredrik Bajers Vej 7E DK-9220 Aalborg.

and

Rasmus Plenge Waagepetersen
Department of Theoretical Statistics and Operations Research, Departments of Mathematical Sciences, Ny Munkegade DK-8000 C Århus.

Markov connected component fields

Jesper Møller[1] and Rasmus Plenge Waagepetersen[2]

[1]*Department of Mathematics and Computer Science, Institute of Electronic Systems, University of Aalborg, F. Bajers Vej 7E, DK-9220 Aalborg. Email:* jm@math.auc.dk.
[2]*Department of Theoretical Statistics and Operations Research, Departments of Mathematical Sciences, Ny Munkegade, DK-8000 Århus C. Email:* rasmus@mi.aau.dk.

Abstract

A new class of Gibbsian models with potentials associated to the connected components or homogeneous parts of images is introduced. For these models the neighbourhood of a pixel is not fixed as for Markov random fields, but given by the components which are adjacent to the pixel. The relationship to Markov random fields and marked point processes is explored and spatial Markov properties are established. Also extensions to infinite lattices are studied, and statistical inference problems including geostatistical applications and statistical image analysis are discussed. Finally, simulation studies are presented which show that the models may be appropiate for a variety of interesting patterns including images exhibiting intermediate degrees of spatial continuity and images of objects against background.

Key words: Bayesian image analysis; Exponential family models; Geostatistics; Infinite lattice processes; Markov chain Monte Carlo; Markov partitions; Markov random fields; Maximum likelihood; Nearest-neighbour Markov point processes;

AMS 1991 Subject classification: Primary 60G60. Secondary 60D05, 60J10.

1. Introduction

In this paper we study a new class of models for digitized images. The models are especially applicable for images where a relatively few number of colours/grey levels occur, and where some prior knowledge is available about e.g. size and shape characteristics of the connected components of constant colour in the image. The models are lattice processes of a Gibbsian type like Markov random fields (MRF's), but with potentials associated to each of the (maximal) connected components in the image so that 'global' information about the components can be more easily incorporated than for MRF's with bounded neighbourhoods. If there is a background colour, i.e. when the potentials are zero for the connected components of a natural reference colour, these models become a special case of the nearest-neighbour Markov point processes introduced in Baddeley & Møller (1989). We shall therefore name these models Markov connected component fields (MCCF's). MCCF's are also examples of the semi-Markov random fields considered in Tjelmeland & Holden (1993). The class of MCCF's contains on the other hand many widely used image models including MRF's like the Potts model. The closely connected random cluster model (Fortuin & Kasteleyn, 1972; Grimmett,

1995) is a MCCF too. The models used in Johnson (1994) for segmentation of noisy images and the Gibbsian prior boundary models recently studied in Helterbrand, Cressie & Davidson (1994) are essentially particular examples of MCCF's.

MRF's have been extensively used as prior models in Bayesian analysis of various kinds of images (see e.g. Geman & Geman, 1984; Besag, 1986) and as models for textures (e.g. in Cross & Jain, 1983; Derin & Elliot, 1987). In Bayesian image analysis MRF's usually serve as penalizing terms rather than as realistic models for the images which are to be restored, c.f. Aykroyd & Green (1991). If the task is only to find maximum a posteriori point estimates of the image and the noise level is moderate, it is not crucial that the prior is a realistic representation of the variability of the true image. It then often suffices to use simple pairwise interaction MRF's like for example the Potts model. However, if the level of degradation is high it is of course important to use more informative and realistic priors.

There are also other situations where it is important that more realistic prior models are used. This is for instance the case when Bayesian methods are used for exploration of properties of oil-reservoirs, see e.g. Haldorsen & Damsleth (1990) and Hjort & Omre (1995). In such applications data are typically very sparse and obtained by indirect observations. In order to ensure that the posterior correctly represents the uncertainty due to the incomplete observations it is necessary to use priors which in a realistic way reflects the properties of the reservoir and where information about larger scale structures in the reservoir can be incorporated. Such global information is difficult to model by the local characteristics of MRF's with small neighbourhoods like e.g. first order MRF's. Ji & Seymour (1994) study methods for choosing appropiate larger neighbourhoods, and Tjelmeland & Besag (1996) use second and fourth order MRF's defined on hexagonal lattices and succesfully models images of objects against background. Since the number of parameters for the general parametrization of a MRF grows rapidly as the neighbourhoods increase an essential problem is to find parsimonious parametrizations which maintain the flexibility of the model. An interesting alternative approach to modelling of global structures in images using hidden first order Markov random fields have recently been studied in Künsch, Geman & Kehagias (1995) but the application of hidden Markov random fields in practice is encumbered with problems which remains to be solved.

We believe that MCCF's may be appropiate for applications in image analysis and geostatistical applications where the structure is given by the components. For a MCCF the size of the neighbourhood of a pixel is not fixed, but given by the components which are adjacent to the pixel.

In practice images are represented digitally on a square lattice of pixels, so the statespace is finite. This is one reason for using a discrete setup like in the present paper. The discrete setup is also advantageous because conditional simulation is usually straightforward – it is e.g. often required to condition on available well-data when oil reservoirs are explored by simulation.

An alternative approach to the modelling of images is to use a continuous setup based on marked point processes. Consider for simplicity a binary image consisting

of 'white' regions of background and a 'black' part B which is the object of primary interest. In stochastic geometry B would typically be modelled as a germ-grain model generated by a marked point process $\Phi = \{(x_i, K_i)\}$ (see e.g. Stoyan, Kendall & Mecke, 1995). Here the grains or marks K_i are compact random sets and $B = \cup\{x_i + K_i\}$ is the union of these sets translated by the points x_i (in the non-binary case a label indicating the colour could be added to each grain). If an image in a natural way can be described as a union of simple geometrical objects like discs or deformations of such objects, this approach may be useful, see e.g. Baddeley & Van Lieshout (1993) and Grenander & Miller (1994). In some cases it is natural to have overlapping grains, but this may also be necessary in order to obtain sufficiently flexible and tractable models, see e.g. Omre & Syversveen (1995). There may then not be a one-to-one correspondence between B and Φ, and the germ-grain approach is moreover not so natural if the white part can not be interpreted as background. One way to get around these problems and to incorporate knowledge about the connected components in the image is to use MCCF's.

The paper is organized as follows. In Section 2 MCCF's are defined and compared with MRF's. MCCF's possesses certain appealing Markov properties established in Section 3. In Section 4 extensions of MCCF's to infinite lattices are studied. Aspects of statistical inference for parametric models of MCCF's are briefly discussed in Section 5. Finally, in Section 6 we fit certain MCCF models to two examples of real data and we use a MCCF prior in an example of Bayesian image analysis. The simulations presented in Section 6 confirm that MCCF's may be appropiate models for a wide range of spatial patterns.

2. Definition of Markov connected component fields and comparison with Markov random fields

In this section we introduce Markov connected component fields with and without a 'background colour'. The first case is of interest when a natural reference colour is given. Moreover, we show that the class of MRF's is not contained in the class of MCCF's and vice versa. A natural question is therefore how the intersection of the two classes is characterized. The answer is given in Theorem 1 under a positivity condition, and we use this to characterize motion invariant MRF's which are MCCF's. Finally, we consider some examples and show that MCCF's may also appear as posterior distributions in a Bayesian approach to image analysis.

2.1 Definition

Let I and V be the finite sets of pixels and colours/grey levels, respectively, so $S = V^I$ is the statespace of images. Assume that a symmetric and reflexive relation $\sim$ on I is given, i.e. $i \sim i$, and $i \sim j$ implies $j \sim i$ for $i, j \in V$ (typically I is a rectangular lattice and $\sim$ is the first or second order nearest neighbour relation). A connected component K is a nonempty subset of I so that for all $i, j \in K$ there exists $i_1, \ldots, i_n \in K$ with $i = i_1 \sim i_2 \sim \cdots \sim i_n = j$. We let $\mathcal{K}$ denote the class of connected components.

For each $x \in S$ and any nonempty subset of pixels $A \subseteq I$ we define the *connected component relation* $\underset{x_A}{\sim}$ on A by

$$i \underset{x_A}{\sim} j \Leftrightarrow \exists\, i_1, \ldots, i_n \in A:\ i = i_1 \sim i_2 \sim \cdots \sim i_n = j \text{ and } x_{i_1} = x_{i_2} = \cdots = x_{i_n}$$

where $x_A = (x_i)_{i \in A}$ denotes the restriction of x to A. Hence the set $\mathcal{K}(x_A) \subseteq \mathcal{K}$ of maximal cliques in A with respect to $\underset{x_A}{\sim}$ constitutes the *maximal connected components* of pixels with the same colour in the image x_A.

Now, a random field $X = (X_i)_{i \in I}$ with values in S is called a *Markov connected component field* if the probability density function of X is of the form

$$p(x) \propto \prod_{K \in \mathcal{K}(x)} \Psi_K(l(x_K)) \tag{1}$$

where '$\propto$' means 'proportional to', $l(x_K) = x_i$ is the common colour of the pixels $i \in K$ in the connected component $K \in \mathcal{K}(x)$, and $\Psi_\cdot(\cdot)$ is some nonnegative function on $\mathcal{K} \times V$ so that $-\log(\Psi_K(l))$ may be considered as the potential associated to a component K with colour l. Suppose that $0 \in V$. If we can replace $\mathcal{K}(x)$ in (1) by $\mathcal{K}_0(x) = \{K \in \mathcal{K}(x) | l(x_k) \neq 0\}$, i.e. $\Psi_K(0) = 1$ for all $K \in \mathcal{K}$, then X is said to be a *MCCF with background colour* 0. The class of MCCF's with background is strictly contained in the general class of MCCF's, see Example A1 in Appendix A.

MCCF's with a background colour are shown in Section 3 to be special cases of the nearest-neighbour Markov point processes introduced in Baddeley & Møller (1989) and further studied in Kendall (1990), Møller (1994b), Baddeley, Van Lieshout & Møller (1996), and Häggström, Van Lieshout & Møller (1996). On the other hand the particular nearest-neighbour Markov point processes studied in Baddeley, Van Lieshout & Møller (1996) appear as weak limits of MCCF's if a certain continuity condition is fulfilled (see Example 1 in Jensen & Møller, 1991; see also Besag, Milne & Zachary, 1982).

Henceforth, for $\emptyset \neq A \subseteq I$ and $K \in \mathcal{K}$, we let $\partial A = \{i \in A^c \mid \exists\, j \in A : i \sim j\}$ denote the neighbourhood of A with respect to $\sim$. Let $\bar{A} = A \cup \partial A$ and write $K \uparrow \bar{A}$ if $K \cap \bar{A} \neq \emptyset$. By the definition (1) of a MCCF X the conditional density of X_A given $X_{A^c} = x_{A^c}$ is

$$p_A(x_A \mid x_{A^c}) \propto \prod_{\substack{K \in \mathcal{K}(x): \\ K \uparrow \bar{A}}} \Psi_K(l(x_K)) \tag{2}$$

for all $x \in S$ with $P(X_{A^c} = x_{A^c}) > 0$. For $i \in I$ the conditional distribution of X_i given $X_{\{i\}^c} = x_{\{i\}^c}$ thus depends on $x_{\{i\}^c}$ only through x_{N_i} where $N_i = \cup\{K \in \mathcal{K}(x_{\{i\}^c}) : i \in \partial K\}$. This property characterizes MCCF's with a background colour in the binary case as verified in Theorem 3, Section 3.2.

The model (1) may be extended to *pairwise interaction MCCF's*, i.e. when

$$p(x) \propto \prod_{K \in \mathcal{K}(x)} \Psi_K(l(x_K)) \prod_{\{K, \acute{K}\} \subseteq \mathcal{K}(x)} \Psi_{K,\acute{K}}\big(l(x_K), l\big(x_{\acute{K}}\big)\big) \tag{3}$$

where $\Psi_{K,\acute{K}}\left(l,\acute{l}\right)$ is a nonnegative function which is symmetric in (K,l) and $\left(\acute{K},\acute{l}\right)$. For example, in the case where 0 is the background colour, a model for isolated components is obtained if $\Psi_{K,\acute{K}}\left(l,\acute{l}\right) = \Psi_{\acute{K},K}\left(\acute{l},l\right) = 0$ whenever $l,\acute{l} \neq 0$ and $K,\acute{K} \in \mathcal{K}$ are neighbouring components, i.e. $K \neq \acute{K}$ and $\partial K \cap \partial K' \neq \emptyset$. Tjelmeland & Holden (1993) call models of the form (3) semi-Markov random fields, and they propose these as useful models for e.g. the facies distribution in petroleum reservoirs. However, in this paper we shall restrict attention to MCCF's with or without a background colour.

2.2 Comparison with MRF's

If X is a MRF with respect to the relation $\sim$ and the density $p(\cdot)$ of X is positive, then by the Hammersley-Clifford theorem (see the historical account in Clifford, 1990, and the references therein) we have that

$$p(x) \propto \prod_{C \in \mathcal{C}} \phi_C(x_C) \tag{4}$$

for some nonnegative functions $\phi_C(\cdot)$ with $\phi_C(x_C) = 1$ whenever $x_i = 0$ for some $i \in C$. The set $\mathcal{C}$ is the set of $\sim$-cliques, i.e. $C \in \mathcal{C}$ when $\emptyset \neq C \subseteq I$ and $i \sim j$ for all $i,j \in C$. Hence, in the binary case $V = \{0,1\}$,

$$p(x) \propto \prod_{K \in \mathcal{K}(x)\,:\, l(x_K)=1} \left\{ \prod_{C \in \mathcal{C}\,:\, C \subseteq K} \phi_C(x_C) \right\}$$

becomes a MCCF with background colour 0. However, there exist simple examples of nonbinary MRF's which are not MCCF's (see Example A2 in Appendix A). Conversely, the class of MCCF's (with or without background) is not contained in the class of MRF's (Example A3). Those random fields which are both MRF's and MCCF's are characterized by the following theorem (for simplicity we consider only random fields whose density is positive everywhere).

Theorem 1 Suppose that $p(\cdot) > 0$ is the density of a random field. This is both a MRF and a MCCF with respect to the same symmetric and reflexive relation $\sim$ if and only if the density is of the form

$$p(x) \propto \prod_{K \in \mathcal{K}(x)} \prod_{\substack{C \in \mathcal{C}\,:\\ C \subseteq K}} \psi_C(l(x_K)) \tag{5}$$

for all $x \in S$ where ψ is a positive function defined on $\mathcal{C} \times V$ and $\mathcal{C}$ is the set of $\sim$-cliques.

Proof: See Appendix B. □

For example, suppose $I \subset \mathrm{Z}^2$ and $\sim = \sim_k$ is the k'th order relation, i.e. $i \sim_k j$ if and only if $||i-j||^2 \leq k$, $k = 1, 2, \ldots$. Consider a MCCF which is also a homogeneous MRF in the sense that (5) holds and $\psi_C(\cdot) = \psi_{\acute{C}}(\cdot)$ whenever C and $\acute{C}$ are cliques which are isomorphic with respect to motions. Let $|A| = \#A$ denote the cardinality of a finite set A and let $\mathcal{C}_k$ denote the set of $\sim_k$ cliques contained in Z^2. Then the density is of an exponential family type with a sufficient statistic given by certain geometrical characteristics:

Corollary Let the situation be as described above.

If $\sim = \sim_1$ is the first order relation, then

$$p(x) \propto \prod_{K \in \mathcal{K}(x)} \exp\left(\alpha_{l(x_K)} a(K) + \beta_{l(x_K)} u(K)\right) \tag{6}$$

for some real parameters α_l and $\beta_l,\ l \in V$, and where

$$a(K) = |K| \text{ and } u(K) = \#\left\{(i,j) \in K \times K^c \,|\, i \underset{1}{\sim} j\right\}$$

are the area and perimeter of K (here we define $K^c = \mathbf{Z}^2 \setminus K$).

If $\sim = \sim_2$ is the second order relation, then

$$p(x) \propto \prod_{K \in \mathcal{K}(x)} \exp\{\alpha_{l(x_K)} a(K) + \beta_{l(x_K)} u(K) + \gamma_{l(x_K)} k_+(K) + \delta_{l(x_K)} k_-(K) + \epsilon_{l(x_K)} d(K)\}$$

for some real parameters $\alpha_l,\ \beta_l,\ \gamma_l,\ \delta_l,\ \epsilon_l,\ l \in V$, where

$$d(K) = \#\{C \in \mathcal{C}_2 \mid |C| = 4, |C \cap K| = 2, C \cap K \notin \mathcal{C}_1\}$$

is the number of $\sim_1$-discontinuities, and

$$k_+(K) = \#\{C \in \mathcal{C}_2 | \, |C| = 4,\ |C \cap K| = 1\}$$

and

$$k_-(K) = \#\{C \in \mathcal{C}_2 \,|\, |C| = 4, |C \cap K| = 3\}$$

denote the number of concave and convex corners, respectively.

Proof: For $K \in \mathcal{K}$ let

$$n_{mk}(K) = \#\{C \in \mathcal{C}_k | \, C \subseteq K, |C| = m\},\ k, m = 1, 2, \ldots$$

The area $a = n_{11}$ together with the perimeter $u = 4n_{11} - 2n_{21}$ is in a linear one-to-one correspondence with (n_{11}, n_{21}), and so (5) implies (6) if $\sim = \sim_1$. Similarly, if $\sim = \sim_2$, then $a = n_{12}$ and by induction on $|K|$ it is straightforwardly verified that

$$d = n_{22} - n_{21} - n_{32} + 2n_{42}$$

$$k_+ = 4n_{12} - 2n_{21} - 2n_{22} + 3n_{32} - 4n_{42} \text{ and } k_- = n_{32} - 4n_{42}$$

so the numbers $(n_{12}, n_{21}, n_{22} - n_{21}, n_{32}, n_{42})$ of isomorphic $\sim_2$-cliques contained in an arbitrary connected component are in linear and injective correspondence with the geometrical characteristics (a, u, k_+, k_-, d) of that component. □

The geometrical characteristics (k_+, k_-, d) are in linear one-to-one correspondence with (k_+, χ, d) (or (k_-, χ, d)) where χ is the Euler-Poincaré characteristic given in terms of the $\sim_2$-cliques as

$$\chi = n_{12} - n_{22} + n_{32} - n_{42}.$$

If $\sim = \sim_2$ in the corollary, we have five geometrical characteristics a, u, k_+, k_-, d or equivalently, a, u, k_+, χ, d. Baddeley, Kendall & Van Lieshout (1996) consider a class of Gibbsian processes of geometrical objects in continuous space where the potential in the planar case is given as a linear combination of the geometrical characteristics a, u, and χ of the random set generated by the union of the objects.

Notice that if a pixel is added to or removed from a component for which the geometrical characteristics are known, then because of the linear relationships stated in the proof above, the calculation of the geometrical characteristics for the modified component requires only "local information". When single site updating Metropolis-Hastings simulation algorithms (Hastings, 1970) are employed it is therefore advantageous to use MCCF's whose sufficient statistic is a function of the five geometrical characteristics. Simulated examples of such MCCF's are considered in Section 6.

Example 1 For a component K, $\chi(K)$ is one minus the number of "holes" in the component. Consider the MCCF given by

$$p(x) \propto \prod_{K \in \mathcal{K}(x)} \exp\left(\gamma(\chi(K) - 1)\right) \tag{7}$$

where γ is a parameter which controls the number of holes in the components. This is not a MRF when $\gamma \neq 0$. □

Example 2 The perhaps best known example of a MRF is the Potts model given by (6) with $\alpha_{l(x_K)} = 0$ and $\beta_{l(x_K)} = \beta$. This model is closely related to the random cluster model introduced in Fortuin & Kasteleyn (1972) and which has been recognized as being of considerable intrinsic interest, cf. Grimmett (1995). The random cluster model is not a MRF but a MCCF: Let $\mathbf{\Gamma} = (\mathbf{V}, \mathbf{E})$ be an undirected graph where $\mathbf{V}$ is the set of vertices and $\mathbf{E}$ the set of edges. The random cluster model takes values in $S' = \{0, 1\}^{\mathbf{E}}$ and the density is given by

$$p(z) \propto q^{|\mathcal{C}(z)|} \prod_{e \in E} p_e^{z_e} (1 - p_e)^{1 - z_e}, \quad z \in S',$$

where $q > 0$ and $0 \leq p_e \leq 1$ are parameters and $\mathcal{C}(z)$ is the set of maximal connected components of the subgraph $(\mathbf{V}, \{e \in \mathbf{E} \mid z_e = 1\})$. Define the relation $\sim$ on $\mathbf{E}$ by

$$e_1 \sim e_2 \Leftrightarrow \exists\, s, t, u \in \mathbf{V} :\ e_1 = \{s, t\} \text{ and } e_2 = \{u, t\}$$

and let $a_s = \#\{u \in \mathbf{V} \setminus \{s\} | \{s,u\} \in \mathbf{E}\}$ for $s \in \mathbf{V}$. Then the random cluster model can be written as

$$p(z) \propto q^{|\mathcal{K}_0(z)| + \sum_{K \in \mathcal{K}(z) \setminus \mathcal{K}_0(z)} \sum_{s \in \mathbf{V}} 1(a_s = \#\{u \in \mathbf{V} \setminus \{s\} | \{u,s\} \in K\})} \prod_{e \in \mathbf{E}} p_e^{z_e} (1-p_e)^{1-z_e}$$

for every $z \in S'$, so it is a MCCF but clearly not a MRF with respect to $\sim$ (the last term in the exponent of q counts the number of maximal connected components of $(\mathbf{V}, \{e \in \mathbf{E} \mid z_e = 1\})$ which consists of just one vertex). We return to this model in Section 4. □

Example 3 MCCF's may appear as posterior distributions in a Bayesian approach to image analysis. The framework described in the following is similar to the one proposed in Johnson (1994).

Let I be a rectangular lattice and $\sim$ the first or second order relation. Assume further that $\Theta = (\Theta_i)_{i \in I}$ is a vector of labels corresponding to different types of landscape, and that a MCCF $p(\cdot)$ is chosen as a prior model for Θ. The prior could for instance be used to model the number and shape of the components in $\mathcal{K}(\Theta)$. The observed image could then conditionally on $\Theta = \theta$ be modelled as

$$Y_i = \mu_{l(\theta_K)} + \eta_K + \epsilon_i, \;\; i \in K \in \mathcal{K}(\theta).$$

Here $\mu_{l(\theta_K)}$ is the mean intensity corresponding to landscape type $l(\theta_K)$, η_K represents a Gaussian departure from this mean intensity (such that e.g. landscape type "field" may appear with varying shades of yellow in different components of the image), and ϵ_i represents iid Gausssian noise. The model is thus specified by

$$\begin{gathered}
\Theta \sim p \\
(\eta_K)_{K \in \mathcal{K}(\Theta)} \mid \Theta \quad \text{(conditionally) independent with } \eta_K \sim N\big(0, \sigma_K^2\big) \\
(Y_i)_{i \in I} \mid \Theta, (\eta_K)_{K \in \mathcal{K}(\Theta)} \quad \text{(conditionally) independent with} \\
Y_i \mid \Theta, (\eta_K)_{K \in \mathcal{K}(\Theta)} \sim N\big(\mu_{l(\Theta_K)} + \eta_K, \sigma^2\big) \text{ for } i \in K.
\end{gathered}$$

The posterior density of $(\Theta, (\eta_K)_{K \in \mathcal{K}(\Theta)})$ given $Y = y$ is then the MCCF

$$p\Big(\Big(\theta, (\eta_K)_{K \in \mathcal{K}(\theta)}\Big) \mid y\Big) \propto$$

$$p(\theta) \prod_{K \in \mathcal{K}(\theta)} \frac{\exp\left(-\frac{\eta_K^2}{2\sigma_K^2}\right)}{\sqrt{2\pi\sigma_K^2}} \prod_{i \in K} \frac{\exp\left(-\frac{\left(y_i - \left(\mu_{l(\theta_K)} + \eta_K\right)\right)^2}{2\sigma^2}\right)}{\sqrt{2\pi\sigma^2}}$$

which can be explored by Markov chain Monte Carlo methods. □

3. Markov properties and characterization results for Markov connected component fields

In this section we study various appealing Markov properties and characterization results for MCCF's which are important for statistical modelling and inference, and which simplify the computations involved in using Markov chain Monte Carlo methods. We will first describe a spatial Markov property of MCCF's in terms of certain random subdivisions of the images called random Markov partitions; we roughly speaking need to condition on more than in the case for MRF's in order to obtain conditional independence results since the neighbourhood relation $\sim_\cdot$ depends on the realizations. Secondly we consider some particular examples. Finally, we show that the class of MCCF's with a background colour is characterized by a local Markov property.

3.1 Global Markov property

The following is a modification of the definition of a Markov partition used in Kendall (1990).

A mapping $M(\cdot) = (A(\cdot), B(\cdot), C(\cdot))$ defined on S and taking values in

$$\Delta = \{(A, B, C) | A, B, C \text{ are disjoint sets with } I = A \cup B \cup C\}$$

is said to be a *Markov partition* if the conditions (M1)-(M3) below are satisfied. Using a notation as in Section 2.1, the two first conditions are

$$\forall\, x \in S\,,\; i \in A(x)\,,\; j \in \bar{C}(x)\;:\; i \underset{x}{\not\sim} j \tag{M1}$$

and

$$\forall\, x, y \in S\;:\;\; B(x) = B(y) \;\Leftrightarrow\; M(x) = M(y)\,. \tag{M2}$$

If $M(x) = (A, B, C)$, then according to (M1), B is a $splitting\ set$ as the set of x-components splits into two classes consisting of those contained in $A \cup B$ and not hitting ∂C and those contained in $B \cup C$ and hitting $\bar{C}$:

$$\mathcal{K}(x) = \{K \in \mathcal{K}(x_{A\cup B}) | K \not\uparrow \partial C\} \cup \left\{K \in \mathcal{K}(x_{B\cup C}) | K \uparrow \bar{C}\right\}$$

where the union is disjoint. Condition (M2) states that the Markov partition is determined by the splitting set. Finally, assuming (M2) and defining for $B \subseteq I$ and $x_B \in V^B$ the sets

$$\mathcal{A}(B, x_B) = \left\{y_A \in V^A \mid \exists\, y_C \in V^C : B(y_A, x_B, y_C) = B\right\}$$
$$\mathcal{C}(B, x_B) = \left\{y_C \in V^C \mid \exists\, y_A \in V^A : B(y_A, x_B, y_C) = B\right\}$$

and

$$\mathcal{D}(B, x_B) = \left\{(y_A, y_C) \in V^A \times V^C \mid B(y_A, x_B, y_C) = B\right\}$$

the last condition is

$$\forall\, x \in S:\; \mathcal{D}(B, x_B) = \mathcal{A}(B, x_B) \times \mathcal{C}(B, x_B) \text{ with } B = B(x)\,. \tag{M3}$$

Here and elsewhere in the following we have identified (y_A, x_B, y_C) with $z = (z_i)_{i\in I}$ where $z_i = y_i$ for $i \in A \cup C$ and $z_i = x_i$ for $i \in B$. It becomes clear in a moment why (M3) is needed.

Any random field X on S and any Markov partition $M(\cdot) = (A(\cdot), B(\cdot), C(\cdot))$ induce a random Markov partition $\mathbb{M} = (\mathbb{A}, \mathbb{B}, \mathbb{C}) = M(X) = (A(X), B(X), C(X))$ and random fields $X_{\mathbb{A}}$, $X_{\mathbb{B}}$, and $X_{\mathbb{C}}$. The condition (M1) together with (M3) ensures that the space of possible states of $(X_{\mathbb{A}}, X_{\mathbb{C}})$ given that $\mathbb{B} = B$ and $X_{\mathbb{B}} = x_B$ is a productspace. Hence, for a MCCF (1) we get by (M1)-(M3) that the conditional density of $X_{\mathbb{A}}$ and $X_{\mathbb{C}}$ given $\mathbb{B} = B$ and $X_B = x_B$ factorizes as follows: Suppose

$$p(B, x_B) = P(\mathbb{B} = B, X_B = x_B) > 0$$

so that the conditional density becomes well-defined. Then

$$\begin{aligned} p(x_A, x_C \mid B, x_B) \propto & \Big\{ 1(x_A \in \mathcal{A}(B, x_B)) \prod_{\substack{K\in\mathcal{K}(x_{A\cup B}) \\ K\not\gamma\partial C}} \Psi_K(l(x_K)) \Big\} \\ & \times \Big\{ 1(x_C \in \mathcal{C}(B, x_B)) \prod_{\substack{K\in\mathcal{K}(x_{B\cup C}) \\ K\uparrow\bar{C}}} \Psi_K(l(x_K)) \Big\} \end{aligned}$$

where the constant of proportionality depends only on (B, x_B). Here we are using a notation as in (2) and $1(\cdot)$ denotes indicator function. We have thus established the following spatial Markov property.

Theorem 2 For a Markov partition $M(\cdot) = (A(\cdot), B(\cdot), C(\cdot))$ and a MCCF X we have that $X_{\mathbb{A}}$ and $X_{\mathbb{C}}$ are conditionally independent given $\mathbb{B}$ and $X_{\mathbb{B}}$. Suppose for $B \subseteq I$ and x_B that $p(B, x_B) > 0$ or equivalently

$$\exists\, (y_A, y_C) \in \mathcal{D}(B, x_B) : \prod_{K\in\mathcal{K}(y)} \Psi_K(l(y_K)) > 0$$

where $y_B = x_B$. Then the conditional distribution of $X_{\mathbb{A}}$ given that $\mathbb{B} = B$ and $X_{\mathbb{B}} = x_B$ is concentrated on the set $\mathcal{A}(B, x_B)$ and its conditional density is

$$p(x_A|B, x_B) = c(B, x_B) \prod_{K\in\mathcal{K}(x_{A\cup B}) \,:\, K\not\gamma\partial C} \Psi_K(l(x_K)) \ ,\ x_A \in \mathcal{A}(B, x_B) \tag{9}$$

where $c(B, x_B)$ is a normalizing constant. □

Example 4 In applications it is often the case that we can observe only X_W where W is a subset of I. As a consequence of this edge effects must be taken into consideration when inference is performed. One way to handle edge effects is to use the spatial Markov property. We can consider the Markov partition with splitting set

$$B(x) = B(x_W) = \cup\{K \in \mathcal{K}(x_W) \,|\, K \uparrow \partial\{I\backslash W\}\}$$

and $A(x) = A(x_W) = W \setminus B(x_W)$, and then use the conditional density (9) where the product is now over all $K \in \mathcal{K}(x_A)$ and

$$\mathcal{A}(B, x_B) = \left\{ x_A \in V^A \,|\, \forall K \in \mathcal{K}(x_A),\, \acute{K} \in \mathcal{K}(x_B) :\, K \uparrow \partial \acute{K} \Rightarrow l(x_K) \neq l\left(x_{\acute{K}}\right) \right\}.$$

The splitting set of this Markov partition is clearly the minimal splitting set contained in W and the conditional density

$$p(x_A \mid B, x_B) = \\ c(B, x_B) \prod_{K \in \mathcal{K}(x_A)} \left\{ 1\left(\forall K' \in \mathcal{K}(x_B) \,:\, K \uparrow \partial K' \Rightarrow l(x_K) \neq l(x_{K'})\right) \Psi_K(l(x_K)) \right\}$$

is a MCCF too. □

Example 5 Another construction of a Markov partition is the following. Let $i \in I$ be a fixed pixel and $A(x)$ the connected component $K_i(x) \in \mathcal{K}(x)$ which contains i. Then the minimal splitting set $B_i(x)$ is trivially $B_i(x) = \cup\{K \in \mathcal{K}(x) | K_i(x) \uparrow \partial K\}$, whereby the conditional probability (9) simply becomes

$$p(x_A | B, x_B) = c(B, x_B) \Psi_A(l(x_A))\,,\; x_A \in \mathcal{A}(B, x_B)$$

with

$$\mathcal{A}(B, x_B) = \left\{ x_A = (l, \ldots, l) \in V^A \,|\, \forall K \in \mathcal{K}(x_B) :\, l \neq l(x_K) \right\}.$$

In other words, the conditional distribution of the colouring of any "point sampled" maximal connected component of the image given the rest of the image depends only on that component and the colouring of its neighbouring connected components in the image. □

3.2 Local Markov property

We turn now to a local Markov property of MCCF's. For $x = (x_i)_{i \in I} \in S$ and $i \in I$ define $x^i = (x^i_j)_{j \in I}$ by $x^i_i = 0$ and $x^i_j = x_j$ if $j \neq i$. Consider any nonnegative function $p(\cdot)$ on S which satisfies the following hereditary condition:

$$p(x) > 0 \Rightarrow p\left(x^i\right) > 0 \text{ for all } i \in I \text{ and } x \in S. \tag{10}$$

Then for $i \in I$ and $x \in S$ with $p(x) > 0$ we can define

$$\lambda_i(x) = p(x)/p\left(x^i\right)$$

so in the case where $p(\cdot)$ is a probability density function of a random field X, $\lambda_i(x)$ is the ratio of 'local characteristics'

$$\lambda_i(x) = p(x_i | x_{-i})/p(0 | x_{-i})$$

where $p(x_i|x_{-i})$ denotes the conditional probability of $X_i = x_i$ given that the rest is $X_{-i} = x_{-i}$. Moreover, in the case of a MCCF with background colour 0 and for $x \in S$ with $p(x) > 0$, we get a local Markov property:

$$\lambda_i(x) = \Psi_{K_i(x)}(x_i) / \prod_{K \in \mathcal{K}\left(x^i_{K_i(x)}\right)} \Psi_K(x_i)$$

depends only on x through $x_{K_i(x)}$, where $K_i(x)$ is defined in Example 5. This is useful when applying single-site updating Metropolis-Hastings algorithms.

In the next theorem we establish that the class of hereditary MCCF's with background colour 0 is uniquely characterized by the local Markov property. This Hammersley-Clifford type theorem may be useful for modelling purposes.

Theorem 3 For a nonnegative function $p(\cdot)$ on S the following statements (a) and (b) are equivalent:

(a) $p(\cdot)$ is of the form

$$p(x) \propto \prod_{K \in \mathcal{K}_0(x)} \Psi_K(l(x_K))$$

where the $\Psi_\cdot(\cdot)$ are nonnegative functions so that for all $l \in V \setminus \{0\}$ and $K, \acute{K} \in \mathcal{K}$ with $\acute{K} \subseteq K$,

$$\Psi_K(l) > 0 \Rightarrow \Psi_{\acute{K}}(l) > 0 \,.$$

(b) The hereditary condition (10) is satisfied and for all $x \in S$ with $p(x) > 0$ we have that $\lambda_i(x)$ depends only on $(i, x_i, K_i(x))$.

Remarks: We have not succeeded to establish a similar characterization result for MCCF's without a background colour. Note that in the binary case and if $p(\cdot)$ is a probability density function, then $\lambda_i(x)$ in (b) can be replaced by $p_i(x_i \mid x_{-i})$.

Proof: That (a) implies (b) is trivial. To prove the converse, represent each $x = (x_i)_{i \in I} \in S$ as a marked point configuration $\{(i, x_i) | x_i \neq 0\}$ and construct the marked point process analogue $f(\cdot)$ to $p(\cdot)$: Consider any finite marked point configuration $\{(i_1, y_{i_1}), \ldots, (i_n, y_{i_n})\} \subset I \times (V \setminus \{0\})$. If $i_j \neq i_k$ for all $j \neq k$ then let

$$x = (x_i)_{i \in I} \text{ with } x_i = \begin{cases} y_i & \text{if } i \in \{i_1, \ldots, i_n\} \\ 0 & \text{else.} \end{cases}$$

Let further

$$f(\{(i_1, y_{i_1}), \ldots, (i_n, y_{i_n})\}) = \begin{cases} p(x) & \text{if } i_j \neq i_k \text{ for all } j \neq k \\ 0 & \text{else.} \end{cases}$$

Then Lemma 1 in Baddeley, Van Lieshout & Møller (1996) shows that (b) implies (a). □

Example 6 Theorem 3 can e.g. be used to verify that $p(\cdot)$ is of the form

$$p(x) \propto \beta^{\#\{i\in I|x_i>0\}}\gamma^{\#\mathcal{K}_0(x)} \tag{11}$$

for some $\beta > 0$ and $\gamma > 0$ if and only if $p(\cdot) > 0$ and for any $i \in I$ and $x \in S$ with $x_i \neq 0$, $\lambda_i(x)$ depends only on x_i and $\#\{K \in \mathcal{K}_0(x^i)|i \in \partial K\,,\, l(x_K) = x_i\}$; that is, the number of maximal connected components in the image x^i which have the colour x_i and contain i in their neighbourhoods. The proof of this result follows the same lines as in Appendix C in Møller (1994b). In Example A3, Appendix A, we show that (11) is not a MRF. □

4. Existence of MCCF's on infinite lattices

In this section the question of existence of (stationary) MCCF's defined on a countably infinite lattice or index set I will be studied. This may be of use for establishing asymptotics for estimators and summary statistics. We shall also briefly discuss the closely related question of phase transition behaviour at the end of Subsection 4.2.

For a symmetric and reflexive relation $\sim$ defined on I, and V a finite set of labels/colours, the sets $\mathcal{K}$ and $\mathcal{K}(x_B)$ for $x \in S = V^I$ and $B \subseteq I$ are defined as in Section 2.1. Since the subfields $X_K,\ K \in \mathcal{K}$ maximal, of a MCCF X defined on I may naturally be considered as being mutually independent, we shall without loss of generality assume that $I \in \mathcal{K}$. Further, $\mathcal{G} = \{\Lambda \subseteq I \mid 0 < |\Lambda| < \infty\}$ is the set of finite nonempty subsets of I. We will assume that $\sim$ is of finite range, so $\bar{\Lambda} \in \mathcal{G}$ for all $\Lambda \in \mathcal{G}$. Furthermore, we suppose that $\{\Lambda_n\}_{n\geq 1}$ is an increasing sequence of subsets in $\mathcal{G} \cap \mathcal{K}$ such that any $\Lambda \in \mathcal{G}$ is contained in some $\Lambda_n \setminus \partial(\Lambda_n^c)$ for all sufficiently large n. For example, if $I = \mathbf{Z}^d$ and $\sim$ is defined by $i \sim j \Leftrightarrow ||i - j|| < \text{const}$, we can take $\Lambda_n = \mathbf{Z}^d \cap [-n, n]^d$.

We discuss below how to specify for each $x \in S$ a family of probability density functions $\{p_\Lambda(\cdot \,|\, x_{\Lambda^c})\}_{\Lambda\in\mathcal{G}}$ in such a way that for large n, $p_\Lambda(\cdot \,|\, x_{\Lambda^c})$ is related to the conditional probabilities of a MCCF restricted to the finite sublattice Λ_n and given the state $x_{\Lambda_n\setminus\Lambda}$ outside Λ. In the terminology of Preston (1976) we want $\{p_\Lambda\}_{\Lambda\in\mathcal{G}}$ to be a specification, i.e. the consistency condition

$$p_{\acute{\Lambda}}\left(x_{\acute{\Lambda}}|x_{\acute{\Lambda}^c}\right) = p_\Lambda(x_\Lambda|x_{\Lambda^c}) \sum_{y_\Lambda} p_{\acute{\Lambda}}\left(y_\Lambda, x_{\acute{\Lambda}\setminus\Lambda}|x_{\acute{\Lambda}^c}\right)$$

should be satisfied whenever $\Lambda \subset \acute{\Lambda}$. We shall also impose a condition on this specification which allow us to show the existence of a random field $X = (X_i)_{i\in I}$ where the conditional probabilities of X_Λ given $X_{\Lambda^c} = x_{\Lambda^c}$ are defined by $p_\Lambda(\cdot \mid x_{\Lambda^c})$ for all $x_{\Lambda^c} \in V^{\Lambda^c}$ with $P(X_{\Lambda^c} = x_{\Lambda^c}) > 0$. Then X will be considered as a MCCF extended to the infinite lattice case.

4.1 Specifications

A fundamental property for establishing existence of random fields corresponding to a specification is quasilocality, cf. Preston (1976) and Georgii (1988). Recent results (Grimmett, 1995) for the particular case of the random cluster model (Example 2 with all p_e equal) indicate that other properties than quasilocality may also be used – this is briefly discussed below and at the end of Section 4.2.

A function defined on $A \subseteq S$ is local if it for some $C \in \mathcal{G}$ and each $x \in A$ depends on x only through x_C, and a function f defined on A is quasilocal if it is the uniform limit of a sequence of local functions. Equivalently, f is quasilocal if

$$\lim_{n \longrightarrow \infty} \sup_{\substack{x,y \in A: \\ x_{\Lambda_n} = y_{\Lambda_n}}} |f(x) - f(y)| = 0$$

(Georgii, 1988, p. 32). The existence results in Preston (1976) and Georgii (1988) require that p_Λ is quasilocal for all $\Lambda \in \mathcal{G}$.

We suppose that $\Psi_K,\ K \in \mathcal{K}$, are nonnegative functions from which the specification should be constructed. Recalling (2) a first proposal would be to define $p_\Lambda(\cdot \mid x_{\Lambda^c})$ by

$$p_\Lambda(x_\Lambda \mid x_{\Lambda^c}) \propto \prod_{\substack{K \in \mathcal{K}(x): \\ K \uparrow \bar{\Lambda}}} \Psi_K(l(x_K)). \tag{12}$$

If Ψ_K factorizes like in the case of a MRF (Theorem 1), i.e.

$$\Psi_K(l) = \prod_{\substack{C \in \mathcal{C}: \\ C \subseteq K}} \psi_C(l)$$

then (12) reduces to

$$p_\Lambda(x_\Lambda \mid x_{\Lambda^c}) \propto \prod_{\substack{C \in \mathcal{C}: \\ C \uparrow \Lambda}} \psi_C(l(x_C))$$

which is clearly local. But otherwise it is easy to find examples of functions Ψ such that (12) is not quasilocal; one example is given by the model (11) with $\Psi_K(l)$ equal to a positive constant different from one when $l \neq 0$ (we return to this model at the end of Section 4.2 and show that this particular MCCF may be extended to the infinite lattice via a MRF which induces (11)). Moreover, it is not always obvious how to define Ψ_K for infinite components K. One example is the function $\Psi_K(l) = \exp\left(\gamma a(K)^2\right)$ which tends to infinity when $\gamma > 0$ and $a(K) \to \infty$. However, it should be noticed that Grimmett (1995) succesfully uses (12) as a natural specification for the random cluster model extended to the infinite lattice $I = \{0,1\}^{\mathbf{E}}$ where the edge set $\mathbf{E}$ consists of all first order neighbours of the vertex set $\mathbf{V} = \mathrm{Z}^d$.

Consequently, we will instead consider conditional probabilitites $p_{\Lambda,n}$ for MCCF's X_{Λ_n} on bounded windows Λ_n, and then define p_Λ as the limit of $p_{\Lambda,n}$ as n tends to infinity. More specifically, let

$$R_{\Lambda,n} = \left\{ x \in S \mid \exists y \in S : x_{\Lambda^c} = y_{\Lambda^c} \text{ and } \prod_{\substack{K \in \mathcal{K}\left(y_{\Lambda_n}\right): \\ K \uparrow \bar{\Lambda}}} \Psi_K(l(y_K)) > 0 \right\}$$

be the subset of S which contains those x for which the function

$$z_\Lambda \to \prod_{\substack{K\in\mathcal{K}\left(z_{\Lambda\cap\Lambda_n}, x_{\Lambda^c\cap\Lambda_n}\right):\\ K\uparrow\bar{\Lambda}}} \Psi_K(l(z_{K\cap\Lambda}, x_{K\cap\Lambda^c}))$$

can be normalized to a probability density. Then

$$p_{\Lambda,n}\left(x_\Lambda \mid x_{\Lambda_n\setminus\Lambda}\right) \propto \prod_{\substack{K\in\mathcal{K}\left(x_{\Lambda_n}\right):\\ K\uparrow\bar{\Lambda}}} \Psi_K(l(x_K)) \tag{13}$$

for all n such that $\Lambda \subseteq \Lambda_n$ and $x \in R_{\Lambda,n}$. Let further

$$R_\Lambda = \bigcap_{n\geq N(\Lambda)} R_{\Lambda,n}$$

where $N(\Lambda)$ is a sufficiently large number defined by N_1 in the proof of Lemma 1 in Appendix C using a certain condition on functions Ψ defined on $(\mathcal{K}\cap\mathcal{G})\times V$. This condition is given in (15) below and it ensures that

(i) for all $\Lambda \in \mathcal{G}$ and $x \in R_\Lambda$, the limit

$$p_\Lambda(x_\Lambda \mid x_{\Lambda^c}) = \lim_{n\to\infty} p_{\Lambda,n}(x_\Lambda \mid x_{\Lambda^c}) \tag{14}$$

exists and is given by an explicit expression,

(ii) p_Λ is quasilocal and $\{p_\Lambda\}_{\Lambda\in\mathcal{G}}$ is a well-defined specification,

(iii) there exists a random field X on I defined by this specification such that $P(X \in R_\Lambda) = 1$.

Hence, $R = \cap_{\Lambda\in\mathcal{G}} R_\Lambda$ becomes the space of 'possible states' for X. Note that if $|K| < \infty$ for all $K \in \mathcal{K}(x_{\Lambda^c})$ with $K \uparrow \bar{\Lambda}$, then the conditional probabilities defined by (12) and (14) are identical, but otherwise (14) will depend on extra terms due to infinite components as shown below.

4.2 Results and discussion

The condition mentioned above is roughly speaking that for any large but finite component K, the function Ψ_K can be approximated by a product like in the MRF case, i.e. $\Psi_K \approx \prod_{C\in\mathcal{C}:C\subseteq K} \phi_C$ where ϕ is a strictly positive function defined on $\mathcal{C}\times V$. More precisely,

$$\begin{aligned} &\forall\Lambda\in\mathcal{G},\ \forall\epsilon>0\ \exists N\geq 1: \\ &\quad n\geq N,\ l\in V,\ \text{ and } K\in\mathcal{K}\cap\mathcal{G} \text{ with } K\uparrow\bar{\Lambda} \text{ and } K\uparrow\{\Lambda_n^c\cup\partial\Lambda_n^c\} \\ &\quad\Rightarrow\left|\frac{\Psi_K(l)}{\prod_{C\in\mathcal{C}:C\subseteq K}\phi_C(l)} - 1\right| < \epsilon\,. \end{aligned} \tag{15}$$

Functions Ψ for which (15) hold could e.g. (in the notation of Section 2.2) be of the form

$$\Psi_K(l) = \exp\left(\alpha_l a(K) + \beta_l u(K) + \gamma g(a(K), u(K))\right),$$

where $g(a, u)$ tends to zero as a (or u) tends to infinity. For example, in the case $\sim\!=\!\sim_1$ we can take $\phi_C(l) = \exp(\alpha_1 + 4\beta_l)$ if $|C| = 1$ and $\phi_C(l) = \exp(-\beta_l)$ if $|C| = 2$. The function g could be of the form $u(K)/a(K)^\delta$ where $\delta > 1$, or equal to $1/a(K)$ as in the second example (b) in Section 6. Other examples include cases where $\Psi_K(\cdot) \to 1$ as $|K| \to \infty$, since we can then let $\phi_C = 1$.

We turn now to the statements (i)-(iii) above. The proofs of the following Lemma 1, Lemma 2, and Theorem 4 can be found in Appendix C. The existence result is Theorem 4.

Lemma 1 Let Ψ be a nonnegative function defined on $(\mathcal{K} \cap \mathcal{G}) \times V$ which satisfies (15). Then the limit in (14) exists for all $x \in R_\Lambda$ and $\Lambda \in \mathcal{G}$, and it is given by

$$\begin{aligned} p_\Lambda(x_\Lambda \mid x_{\Lambda^c}) = c(x_{\Lambda^c}) & \prod_{\substack{K \in \mathcal{K}(x): K \uparrow \bar{\Lambda}, \\ |K| < \infty}} \Psi_K(l(x_K)) \\ & \prod_{\substack{K \in \mathcal{K}(x): K \uparrow \bar{\Lambda}, \\ |K| = \infty}} \prod_{\substack{C \in \mathcal{C}: C \subseteq K, \\ C \uparrow K \setminus V\left(\Lambda, x_{\Lambda^c}\right)}} \phi_C(l(x_K)) \end{aligned} \tag{16}$$

where $c(x_{\Lambda^c})$ is a normalizing constant so that $p_\Lambda(\cdot|x_{\Lambda^c})$ becomes a probability density function and

$$V(\Lambda, x_{\Lambda^c}) = \cup_{\substack{K \in \mathcal{K}\left(x_{\Lambda^c}\right): \\ K \uparrow \bar{\Lambda}, |K| = \infty}} K.$$

□

Note that in (16), different types of functions appear for finite and infinite x-components which hit $\bar{\Lambda}$. As shown in the proof of Lemma 1 the ϕ_C-terms in (16) are those which are left after approximation and cancellation due to (15) when n tends to infinity in (13).

Lemma 2 Suppose that Ψ is a nonnegative function defined on $\mathcal{K} \cap \mathcal{G} \times V$ satisfying (15). Then for any $\Lambda \in \mathcal{G}$ the function $x \to p_\Lambda(x_\Lambda|x_{\Lambda^c})$ given by (16) is quasilocal. □

Theorem 4 Let Ψ be a nonnegative function on $(\mathcal{K} \cap \mathcal{G}) \times V$ for which (15) holds and let p_Λ be given by (14) for each $\Lambda \in \mathcal{G}$. Then there exists a random field X taking values in S such that for all $\Lambda \in \mathcal{G}$,

$$P(X \in R_\Lambda) = 1 \tag{17}$$

and for any $x \in R_\Lambda$,

$$P(X_\Lambda = x_\Lambda \mid X_{\Lambda^c} = x_{\Lambda^c}) = p_\Lambda(x_\Lambda \mid x_{\Lambda^c}). \tag{18}$$

Further, if $I = \mathbf{Z}^d,\ d \geq 1$, and Ψ is translation invariant, i.e.

$$\forall\, t \in \mathbf{Z}^d,\, K \in \mathcal{K}:\ \Psi_K(\cdot) = \Psi_{\{t+s|s\in K\}}(\cdot), \tag{19}$$

then there exists a stationary random field X for which (16) and (17) hold (here stationarity means that $X = (X_i)_{i\in I}$ is distributed as the translated field $(X_{i+t})_{i\in I}$ for any $t \in \mathbf{Z}^d$). □

For establishing the above existence result we have used results from the general theory of Gibbsian fields (Preston, 1976; Georgii, 1988). Sufficient conditions for unique existence of MCCF's on infinite lattices may be stated as well using Dobrushin's uniqueness criterion (Georgii, 1988, p. 142) or other techniques (Grimmett, 1995). There is a number of interesting models like e.g. the model (25) studied in Section 6 where the condition (15) does not hold and where it is therefore difficult to use the results in Preston (1976) or Georgii (1988). It seems to be an open problem whether such models in general can be extended to infinite lattices, but Grimmett's (1995) results for the random cluster model seem promising.

Another possibility is to consider a weak limit of MCCF's defined on the finite regions Λ_n. Grimmett (1995) studies this in detail for the random cluster model and shows that it leads to the same as using the 'specification' (12) in the case of translation-invariance or uniqueness when one extends the set of vertices $\mathbf{V}$ in Example 2, say $\mathbf{V} = \mathbf{V}_n = [-n, n]^d \cap \mathbf{Z}^d$, to $\mathbf{Z}^d$.

The weak limit approach seems natural for another special example of a MCCF where (15) is not fulfilled. This is the model (11) for values of $\gamma = 1, 2, 3, \ldots$: Let $I' \subseteq I$ be a finite index set, and let $k = 1, 2, \ldots$. Let further $Y_{I'} = (Y_i)_{i\in I'}$ be the MRF on $\{0, \ldots, k\}^{I'}$ given by the density

$$p(y) \propto \beta^{\#\{i\in I'|y_i>0\}} \prod_{\substack{i,j\in I':\\ i\sim j}} 1(y_i > 0, y_j > 0 \Rightarrow y_i = y_j),\ y \in \{0, \ldots k\}^{I'}.$$

where $\beta > 0$. That is, $Y_{I'}$-components which are not of the background colour 0 are separated by the background colour. Define the binary random field $X_{I'} = (X_i)_{i\in I'}$ by $X_i = 1(Y_i > 0),\ i \in I'$. Then the number of nonzero $Y_{I'}$-components equals the number of nonzero $X_{I'}$-components, so

$$P(X_{I'} = x_{I'}) \propto \beta^{\#\{i\in I'|x_i=1\}} k^{\mathcal{K}_0(x_{I'})}$$

which is the model (11) for $\gamma = k$. This construction is very similar to the connection between a bivariate point process model due to Widow & Rowlinson (1970) and 'the continuum random cluster model' which is a nearest-neighbour Markov point process model introduced independently in Klein (1982), Møller (1994a,b) and Chayes, Chayes & Kotecky (1995). Since $Y_{I'}$ is a simple pairwise interaction MRF it is as discussed in Section 4.1 obvious how to define a specification and it is easy to show that $Y_{I'}$ can be extended to a stationary random field $Y = (Y_i)_{i\in I}$ on the infinite lattice. The corresponding random field $X = (1(Y_i > 0))_{i\in I}$ is then an extension to the infinite

lattice of the model (11), and it would be interesting to see if this agrees with the weak limit approach.

A closely related property to uniqueness is phase transition, see e.g. Grimmett (1995) and the references therein. Lebowitz & Gallawotti (1971) shows that this occurs if β is sufficiently large and the model (11) is modified by including a strictly positive hard-core between non-zero connected components of different colours, i.e. such components are separated by the background colour. Similar results for the continuum random cluster model and the related Widow and Rowlinson model have been established, see Ruelle (1971), Chayes, Chayes & Kotecky (1995) and Georgii & Häggström (1996). In Section 6, we discuss practical implications of phase transition behaviour for the Ising model and a certain MCCF.

5. Estimation

In this section we discuss very briefly how to estimate parameters in a parametric model for a MCCF.

Suppose we have chosen a particular parametric family of densities $\{p_\theta : \theta \in \Theta\}$ for a MCCF defined on a finite lattice I. For example, this may be an exponential family

$$p_\theta(x) \propto \exp\left\{ \sum_{K \in \mathcal{K}(x)} [\theta \bullet t(K, l(x_K)) + w(K, l(x_K))] \right\} \tag{20}$$

where $\Theta \subseteq \mathbb{R}^k$, $\bullet$ denotes the usual inner product in $\mathbb{R}^k$, and $t : \mathcal{K} \times V \to \mathbb{R}^k$ and $w : \mathcal{K} \times V \to [-\infty, +\infty)$ are statistics which are functions of the size, shape, orientation, and colouring of the components (in fact all specific examples of MCCF's considered in this paper are of the form (20)). The choice of the statistics t and u may be based on factual knowledge, intuition and simulation studies (see Section 6).

The unknown parameter θ may be estimated using general techniques such as maximum likelihood and maximum pseudolikelihood (Besag, 1975). Due to intractibility of the normalizing constant in the likelihood usually only approximate maximum likelihood estimates can be obtained using for example Markov chain Monte Carlo methods (see e.g. Geyer, 1994). Maximum pseudolikehood may be easier. Notice that likelihood functions as in (20) and their corresponding pseudolikehoods are log-concave and of a form studied in Jacobsen (1989) who gives sufficient and necessary conditions for strictly log-concavity and unimodality of such functions.

Finally, we address some open problems. Consistency and asymptotic normality of maximum (pseudo)likelihood estimates for certain parametric models of random fields have been established in various papers. Geman & Graffigne (1986) consider the finite state space case of lattice variables as in the present paper, but they use a coding technique based on a spatial Markov property for MRF's which is not in general satisfied for MCCF's with unbounded components. Alternatively, the techniques presented in Guyon (1986), Gidas (1988), and Comets (1989) may be adapted; in particular the large deviation bounds used in the paper by Comets seem useful. It would also be interesting

to compare the efficiency of maximum likelihood and maximum pseudolikelihood. It is well-known that maximum likelihood is more efficient than maximum pseudolikelihood for MRF's with strong interactions like in the Ising or Potts model at or above the critical point where phase transitions occur (see Section 6). In Section 6 it is demonstrated that MCCF's which are simple modifications of the Ising model provide appealing models with softer interactions, so it would be interesting to compare the efficiency of the estimation methods for such models.

6. MCCF modelling of images

We will now turn to some specific examples of MCCF's which will be used to model the two binary images shown in Figure 1. The first 62×67 image is obtained by thresholding a microscopic image of a rock sample. Here black and white corresponds to presence of carbonate and anhydrite, respectively. The second 62×60 image is a thresholded section of the handmade paper texture D57 from Brodatz (1966). For the rock sample image we use either (a) the Ising model or (b) consider modifications of the Ising model where presence of small components are penalized. MCCF's where white is the background colour are used in (c) to model the second image. All models in (b) and (c) are MCCF's which are not MRF's. We shall also in (d) compare a Bayesian image analysis of a noisy version of the rock sample image when either the Ising model or the MCCF model from (b) is used as the prior.

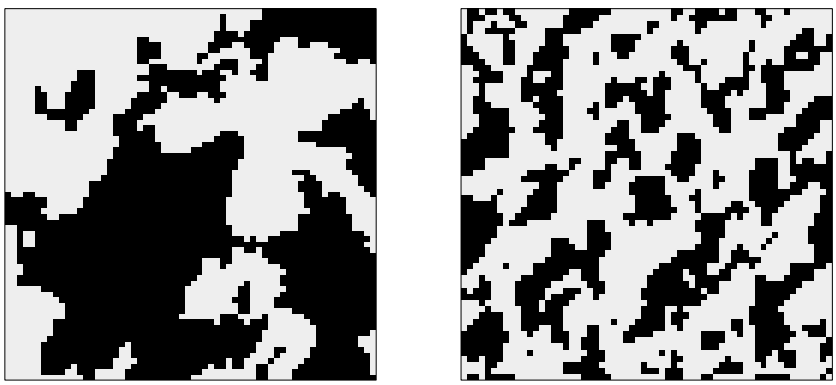

Figure 1: Data. Left: Rock sample image. Right: Handmade paper texture.

Rough estimates of the parameters in the models were obtained by stochastic approximation (Younes, 1988) and these estimates were then used as starting points for Markov chain Monte Carlo maximum likelihood estimation (MCMC mle) of the parameters, see Geyer & Thompson (1992). For the Ising model the Markov chain Monte Carlo samples were obtained by using the Swendsen & Wang (1987) algorithm, while in (b) and (c) it sufficed to use an ordinary single site updating Metropolis algorithm. For all the models considered, the sufficient statistic is a function of the geometrical characteristics introduced in the corollary and subsequent discussion in

Section 2.2. As mentioned in Section 2.2, this is convenient when single site MCMC algorithms are used.

It should be noted that specification of image models is not straightforward. One approach which is often used in Bayesian image analysis is to initially specify a class of potentials and then use the corresponding Gibbs distribution. Such an approach may be motivated by the maximum entropy principle (see e.g. the discussion in Geyer & Thompson, 1992) but does not necessarily lead to realistic image models. For MCCF models it may seem reasonable to use potentials of the form $\delta d(C(K), C)$, $\delta > 0$, $K \in \mathcal{K}$, where $C(\cdot)$ measures some characteristic of a component, and $d(\cdot, \cdot)$ measures deviations between $C(\cdot)$ and a fixed value C; potentials in this spirit have been proposed in Green (1986) and Tjelmeland & Holden (1993). This approach may work well for penalizing purposes in maximum a posteriori (MAP) estimation but can be misleading for construction of realistic models for images. Especially the expected value of $C(\cdot)$ do not need at all to be close to C. Other complications may appear as well. Consider for example the potential $\gamma u(K)/\sqrt{a(K)}$, $\gamma > 0$, $K \in \mathcal{K}$, which penalizes "noncompactness" of a component. Since the value of the statistic $u(K)/\sqrt{a(K)}$ is always greater than four, also the total number of components will be greatly influenced by this potential.

6.1 Examples and simulations

(a) *The Ising model*. The Ising model without an external field is the model (6) with $V = \{0, 1\}$, $\alpha_l = 0$, and $\beta_l = \beta \in \mathbf{R}$, $l \in V$, so that it is given by

$$p(x) \propto \exp\left(\beta \sum_{K \in \mathcal{K}(x)} u(K)\right), \; x \in V^I \tag{21}$$

where u is the perimeter as defined in Section 2.2. We will here focus on the attractive case where $\beta \leq 0$. For values of $\beta < \beta_c \approx -0.44$ phase transition occurs while there is no phase transition when $0 \geq \beta > \beta_c$. A description of phase transition for the Ising model can be found in e.g. Kindermann & Snell (1980). The MCMC mle of β obtained from the left image in Figure 1 is $\hat{\beta} = -0.481$ and thus supercritical. For the Ising model, the distribution of the statistic $\sum_{i \in I} X_i$ (known as the total magnetization among physicists) is unimodal when $\beta > \beta_c$, while for values of $\beta < \beta_c$ where phase transition occurs, the distribution is bimodal. This behaviour is exemplified in Figure 2. In this figure, the distribution of $\sum_{i \in I} X_i$ under the Ising model is estimated by probability scaled histograms where β is -0.43 (i.e. less than β_c) and equal to $\hat{\beta}$, respectively. When phase transition occurs, the bimodality of the distribution of $\sum_{i \in I} X_i$ implies that realizations of the Ising-model will be dominated by one of the colours. This behaviour is apparent from the simulations in Figure 3 where β takes the same values as in Figure 2.

When a noisy or blurred version of an image like the rock sample image in Figure 1 is observed, the Ising model works quite well as a prior for MAP estimation of the true unobserved image. By considering the simulations in Figure 3 it is however clear, that

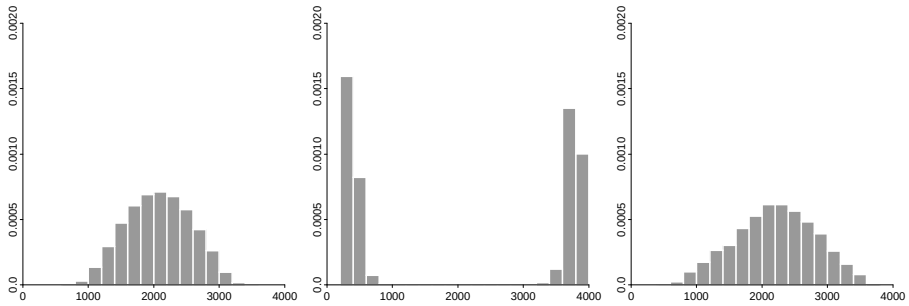

Figure 2: Histograms of the number of black pixels. Left: Ising model with $\beta = -0.43$. Middle: Ising model with $\beta = -0.481$ (MCMC mle). Right: The model (22) with $\beta = -0.483$ and $\gamma = -17.643$ (MCMC mle).

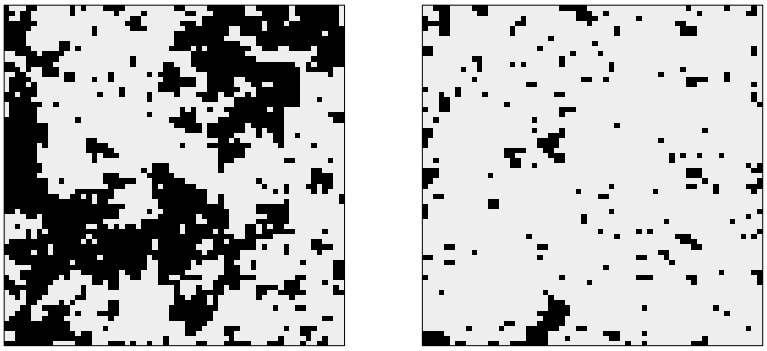

Figure 3: Simulations of the Ising model. Left: $\beta = -0.43$. Right: $\beta = -0.481$ (MCMC mle).

the Ising model is a poor representative of the variability present in the rock sample image because of the phase transition behaviour and the "noisy" appearance (due to small components) of the realizations of the Ising model. It is therefore questionable whether meaningful probalistic statements can be based on the posterior distribution when the Ising model is used as the prior; see also (d) and the discussion in Tjelmeland & Besag (1996).

(b) *A MCCF obtained from the Ising model by penalizing small components*. The model in this example is given by

$$p(x) \propto \exp\left(\sum_{K \in \mathcal{K}(x)} \beta u(K) + \frac{\gamma}{a(K)}\right), \; x \in V^I \tag{22}$$

where γ and β are real parameters. Compared to the Ising model, realizations where small components occur are downweighted when $\gamma < 0$. The condition (15) for extensions to infinite lattices holds for the function Ψ corresponding to this model and it can in fact easily be shown that for any negative value of γ there exists a value of β for which phase transition occur (see Waagepetersen, 1996). The MCMC mle's of β and γ obtained from the rock sample image are $\hat{\beta} = -0.483$ and $\hat{\gamma} = -17.643$. Simulations

of the estimated model in Figure 4 show that the model captures the variability in the true image rather well. The distribution of the number of black pixels is estimated by the right histogram in Figure 2 and since $\hat{\beta}$ is supercritical for the Ising model and close to the MCMC mle for the Ising model, it is interesting to notice that the phase transition behaviour of the Ising model does not seem to be present. This is because the supercritical Ising model is concentrated on states with one dominating component and many small components (like the right image in Figure 3) and such states are heavily downweighted by the estimated model (22). The absence of phase transition in the estimated model suggests that the model (22) is more flexible than the Ising model for modelling of images exhibiting varying degrees of spatial continuity.

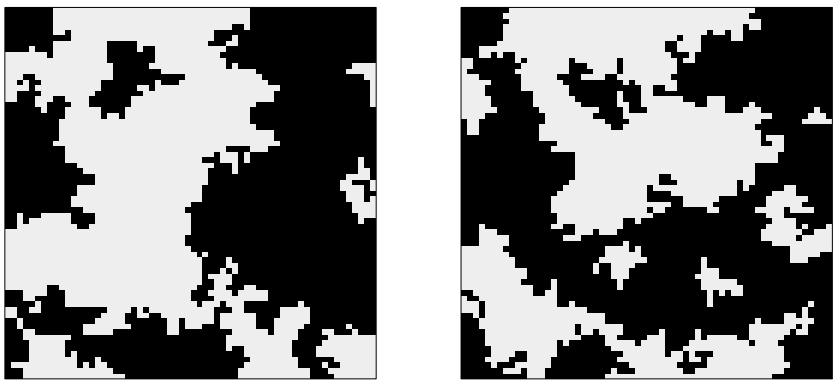

Figure 4: Realizations of the estimated model (22).

In order to study whether a more careful modelling of the boundaries would improve the results we have also tried to add the geometrical characteristics d and k_+ (see Section 2.2). Thereby the model

$$p(x) \propto \exp\left(\sum_{K\in\mathcal{K}(x)} \beta u(K) + \delta d(K) + \kappa k_+(K) + \frac{\gamma}{a(K)}\right),\ x \in V^I \qquad (23)$$

is obtained. Note that, except for boundary effects, we obtain the same model if we add an extra term ϵk_- in the exponent since the images are binary and the model is without background. The MCMC mle's are $\hat{\beta} = -0.453$, $\hat{\delta} = -0.378$, $\hat{\kappa} = -0.049$, and $\hat{\gamma} = -17.108$. Simulated realizations of the estimated model are shown in Figure 5. The model (23) fits the data better than (22) in the sense that the mean values of the statistics $\sum_{K\in\mathcal{K}(X)} k_+(K)$ and $\sum_{K\in\mathcal{K}(X)} d(K)$ are in accordance with the observed values. From a visual comparison of Figure 4 and 5 it however seems that the model (22) does better than (23) because the boundaries in Figure (22) appear more smoothly curved than the boundaries in Figure (23). Under both of the estimated models (22) and (23) the boundaries have a somewhat more irregular appearance than in the original image.

(c) *A MCCF model for black components on a white background.* The number and size of black components are important features of the handmade paper image in

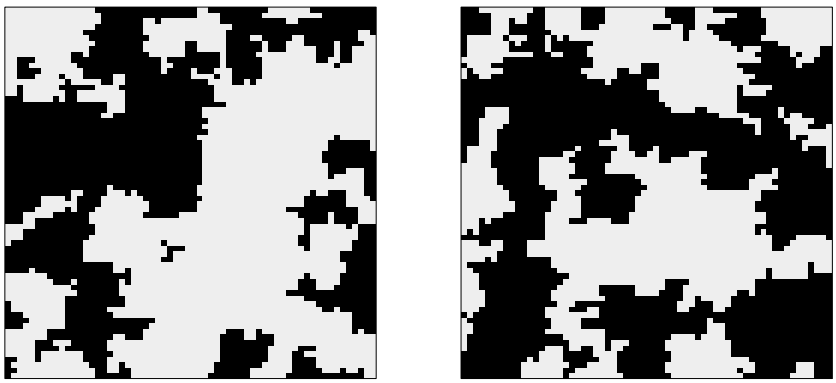

Figure 5: Realizations of the estimated model (23).

Figure 1. In order to model this image we tried first to use a binary second order MRF given in terms of the geometrical characteristics a, u, d, k_+, and χ. As in the proof of the corollary in Section 2.2 it can be seen that the numbers of the five types of isomorphic $\sim_2$-cliques contained in the black part of the image is in linear one-to-one correspondence with the five geometrical characteristics of the black part of the image. We have thus used the general parametrization (4) of a binary second order MRF with motion invariant clique potentials. This MRF model was not able to capture the variability displayed by the data.

We therefore considered the MCCF given by

$$p(x) \propto \exp\left(\sum_{K\in\mathcal{K}_0(x)} \alpha a(K) + \beta u(K) + \phi + \gamma a(K)^2\right),\ x \in V^I. \tag{24}$$

The parameter γ controls the size of the components since the squared area of the union of two components is greater than the sum of the squared areas of each component. The size of the components is however also influenced by the parameter α together with the parameter ϕ which controls the number of components. The parameter β controls the perimeter of the black part of the image and thereby also the shape (compactness) of the components. The MCMC mle's obtained from the handmade paper texture are $\hat{\alpha} = 0.083$, $\hat{\beta} = -0.849$, $\hat{\phi} = -0.986$, and $\hat{\gamma} = -2.490 \times 10^{-4}$. Figure 6 shows, that the variability of size and number of the black components in the original image is quite well represented by the model, but the black components in the original image appear somewhat more "compact" than the simulated black components. Especially, very few holes are present in the black components of the original image.

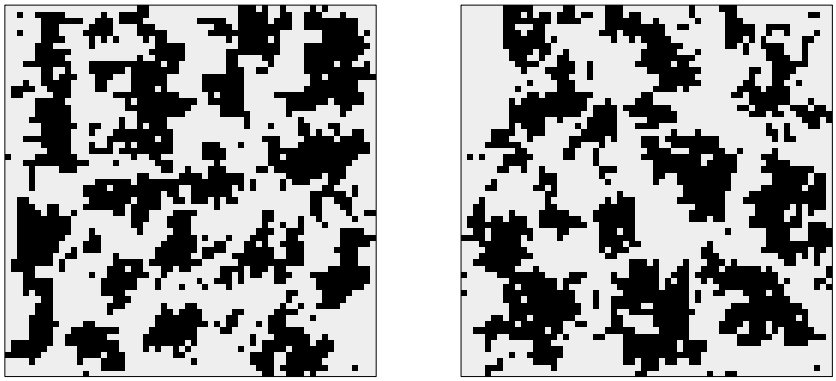

Figure 6: Realizations of the estimated model (24).

In order to improve the model we have added the geometrical characteristic $\chi - 1$, where $1 - \chi(K)$ is the number of holes in a component K (see Section 2.2). The extended model is thus

$$p(x) \propto \exp\Bigg(\sum_{K \in \mathcal{K}_0(x)} \alpha a(K) + \beta u(K) + \epsilon(\chi(K) - 1) + \phi + \gamma a(K)^2 \Bigg), \; x \in V^I \quad (25)$$

and the MCMC mle's of the parameters are $\hat{\alpha} = 0.087$, $\hat{\beta} = -0.834$, $\hat{\epsilon} = 1.550$, $\hat{\phi} = -1.045$, and $\hat{\gamma} = 1.923 \times 10^{-4}$. Simulations (see Figure 7) of the estimated model show that (25), compared to (24), is a better representative of the variability of the components in the handmade paper image since the number of holes in the black components has decreased. In order to judge the performance of the model not only by eye we have also compared the empirical distributions of the component characteristics $a(K)$, $u(K)$, and $u(K)/\sqrt{a(K)}$ for black components K with the distributions obtained by simulation under the estimated model (25), see Figure 8. Especially with respect to $a(K)$ and $u(K)$ the fit of the estimated model (25) seems quite good.

In the original image there is some indication of directional effects and the mutual placement of the black components seems quite structured. This is not apparent in the simulations of the "isotropic" model (25) where interactions between the components are not (explicitly) modelled.

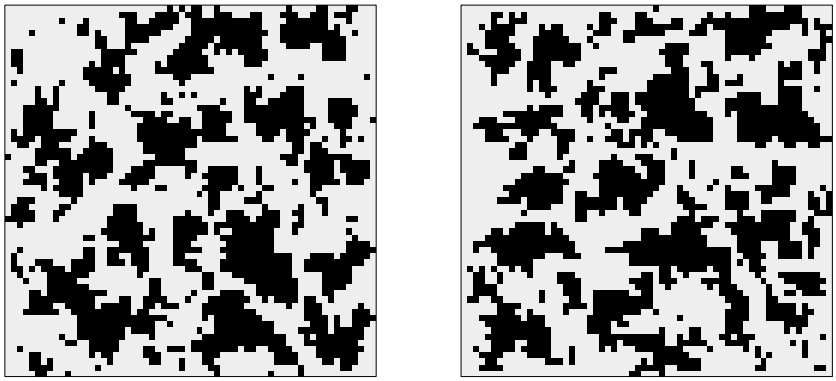

Figure 7: Realizations of the estimated model (25).

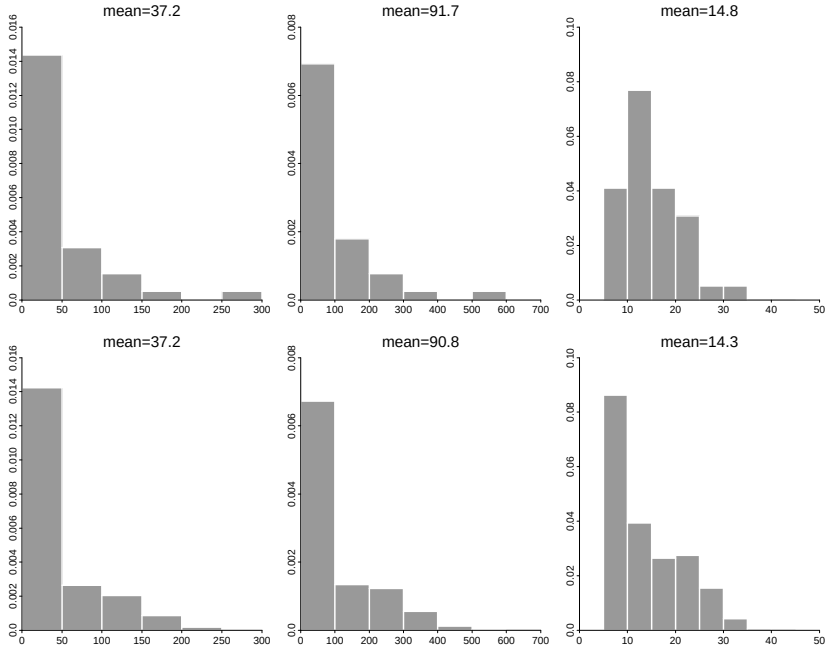

Figure 8: Upper row: empirical distribution of $a(K)$, $u(K)$, and $u(K)/\sqrt{a(K)}$, $K \in \mathcal{K}_0(x)$. Lower row: simulated distributions of these statistics under the estimated model (25).

(d) *An image analysis experiment.* This example demonstrates how the application of MCCF priors instead of simple pairwise interaction priors may improve the results of a Bayesian image analysis. We consider simulated data (see Figure 9) obtained by adding a Gaussian white noise with variance 1.25 to the rock sample image in Figure 1. This is a rather high noise level for which the expected rate of misclassified pixels in the maximum likelihood estimate of the true image is 32.9%. We shall only briefly discuss MAP estimation and instead mainly focus on a more genuine Bayesian analysis where we calculate posterior distributions of various statistics related to the true image.

When the Ising model is used as a prior for MAP estimation of the true image best results (visually and in terms of rates of misclassified pixels) are obtained with a

value of β around -0.25. The MAP estimate obtained with the MCCF model (22) with $\beta = -0.25$ and $\gamma = -8$ is somewhat better, and the improvement is mainly due to removal of certain small components which are not smoothed away by the Ising prior unless higher values of $-\beta$ is used. Such values of β however lead to oversmoothing in other parts of the image when the Ising prior is used. Note that for both models the optimal smoothing parameters are far from the maximum likelihood estimates obtained in (a) and (b). This is because the optimal values of the smoothing parameters depend much on the noise level. When the noise level is high and the likelihood accordingly weak a sufficient degree of smoothing can be obtained with smaller values of $-\beta$ than when the noise level is low.

In a "true" Bayesian analysis the prior is not just a convenient smoothing term but ideally should be a realistic representation of the prior uncertainty so that inference based on the full posterior is justified by the Bayesian rationale. In the following we use the maximum likelihood estimates obtained in (a) and (b) as prior parameters. This corresponds to the case where training data is available for estimation of prior parameters. The simulations in (b) show that this approach leads to a quite realistic prior modelling when the model (22) is used.

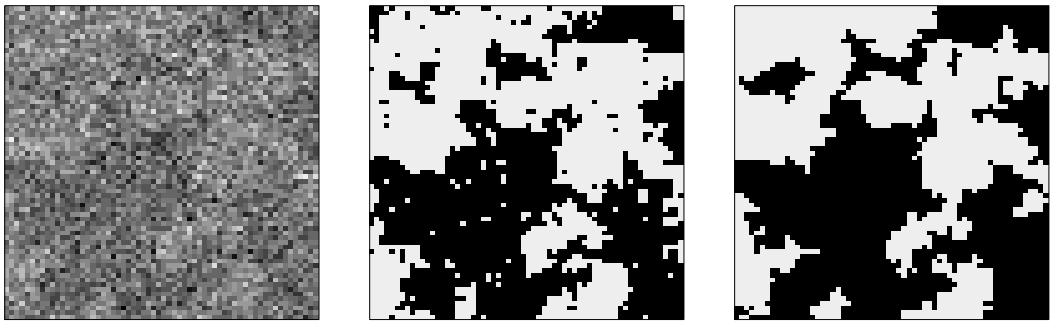

Figure 9: Left: simulated data, middle: realization of posterior with Ising prior, right: realization of posterior with prior given by (22).

We have used MCMC to estimate the posterior distributions of the number of black pixels, the number of components, and the marginal posterior probabilities of observing a black pixel. The results obtained with the two priors are shown in Figure 10 and the true numbers of components and black pixels together with posterior means and standard deviations are given in Table 1. The posterior distribution obtained with the Ising prior is clearly a misleading representation of the knowledge concerning the number of components in the true image. This is not surprising when considering realizations of this posterior where a large number of small components are present (see the middle image in Figure 9). The posterior distribution of the number of components is more correctly centered around the true value when the prior model (22) is used and note that this is also the case for the posterior distribution of the number of black pixels. Furthermore, the marginal posterior probabilities of observing a black pixel are in much better accordance with the pixel values of the true image when the model (22) is used as prior.

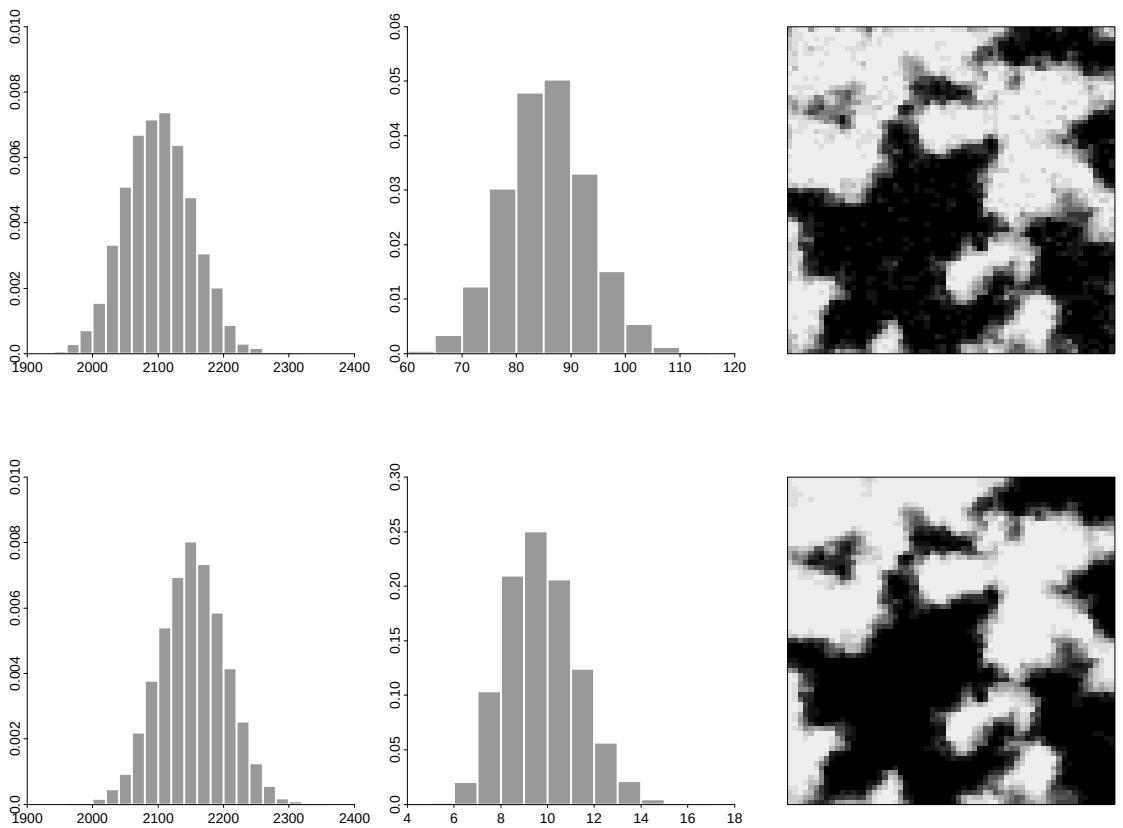

Figure 10: Upper row: posterior distribution with Ising prior of number of black pixels (left), number of components (middle), and the marginal posterior probabilities of observing a black pixel (where white in the gray level image corresponds to zero probability). Lower row: same as upper row but with prior given by (22).

True values	2157	11
Ising prior	2101.4 (51.8)	86 (7.7)
Prior given by (22)	2153.1 (50.7)	10.3 (1.6)

Table 1 True values and posterior means and standard deviations (in parenthesis) of number of black pixels and number of components.

This example stresses the importance of a realistic prior modelling when the full posterior distribution is used for inference. Quite noninformative priors were used and it is of course to be expected that the advantages of MCCF priors will become even more apparent when more prior knowledge can be included in the prior model. One may object that better results might be obtained with the Ising prior if another prior parameter was used instead of the maximum likelihood estimate. It is on the other hand not at all clear according to which criteria such an alternative parameter value should be chosen. This reveals another advantage of using realistic priors: when training data are available, estimation of prior parameters from such data is a natural and useful method for choosing prior parameters.

6.2 Final remarks

In this section we have discussed modelling of two binary images. The application of first or second order MRF's did not lead to satisfactory results, but quite good results were achieved by using rather simple MCCF models. The application of MCCF's is somewhat more computationally demanding than the application of MRF's. By the local Markov property the calculations for an update of a site in a single site updating algorithm only involves the components adjacent to this site. Still, if these components are large the extra computational expense due to the problem of deciding whether a pixel update leads to a division of a component or to a unification of a number of components may be considerable. Also the geometrical characteristics for the new components which appear after a division or unification must be calculated; as mentioned after the Corollary in Subsection 2.2 this requires only "local" information and is not computationally demanding. Sweeny (1983) gives a very efficient algorithm for the random cluster model (see Example 2, Section 2.2) but it does not seem applicable for MCCF's in general. The implementation used here works for any MCCF and is based on standard graph algorithms (see e.g. Cormen, Leiserson & Rivest, 1989).

MCMC maximum likelihood worked well, but (at least with the simulation algorithm actually used here) the computing time required to obtain sufficiently large MCMC samples is quite high (several hours of CPU time on a HP 9000 735/125).

Acknowledgement

This research was initiated while both authors were at the Department of Theoretical Statistics, University of Aarhus, and will be a part of the second authors Ph.D. Dissertation. The second author was supported by the Geological Survey of Denmark and Greenland. The rock sample image considered in Section 6 was kindly provided by Peter Frykman, the Geological Survey of Denmark and Greenland. We thank Olle Häggström, Wilfrid S. Kendall and Peter Frykman for helpful comments.

Appendix A: Examples A1–A3

Example A1 *The class of MCCF's is not contained in the class of MCCF's with background*: A counterexample is the three-state Potts model given by (6) with $\alpha_{l(x_K)} = 0$, $\beta_{l(x_K)} = \beta$, and $V = \{0, 1, 2\}$. Suppose this is of the form (1) with $\mathcal{K}(x)$ replaced by $\mathcal{K}_0(x)$. Hence $\Psi_K(l) > 0$ for $l = 1, 2$ and any $K \in \mathcal{K}$. Now, consider any fixed $K^* \in \mathcal{K}\backslash\{I\}$ and let $x, x' \in S$ be given by $x_i = 0$ and $x'_i = 2$ if $i \in K^*$ and $x_i = x'_i = 1$ otherwise. Then $\mathcal{K}_0(x') = \mathcal{K}_0(x) \cup \{K^*\}$. Since $p(\cdot)$ is a Potts-model we have that $p(x) = p(x')$ or equivalently

$$\prod_{K \in \mathcal{K}_0(x)} \Psi_K(l(x_K)) = \Psi_{K^*}(2) \prod_{K \in \mathcal{K}_0(x)} \Psi_K\big(l\big(x'_K\big)\big)$$

so $\Psi_{K^*}(2) = 1$ and by symmetry $\Psi_{K^*}(1) = 1$. Hence $p(x)$ is constant for all x with $x_i \neq x_j$ for some $i, j \in I$. This is a contradiction when $\beta \neq 0$. □

Example A2 *A* MRF *which is not a* MCCF: Let $I = \{1, 2\}$, $V = \{0, 1, 2, 3\}$, and $1 \sim 2$. Then

$$p(x) \propto |x_1 - x_2|$$

defines a MRF. Assume that this is of the form (1). It then follows that $\Psi_{\{1\}}(0) = \Psi_{\{1\}}(2)$ and $\Psi_{\{2\}}(3) = \Psi_{\{2\}}(1)$ as $p((0, 1)) = p((2, 1)) = p((2, 3))$. We thus get $p((0, 3)) = p((2, 1))$ which is a contradiction. □

Example A3 *A* MCCF *with background which is not a* MRF: Identify $I = \{1, 2, 3, 4\}$ with a square 2×2 lattice. Let $V = \{0, 1\}$ and $\sim$ be the first order relation so that $1 \sim 2 \sim 3 \sim 4 \sim 1$, but $1 \not\sim 3$ and $2 \not\sim 4$. Now, consider the MCCF with background colour given by

$$p(x) = \left\{1 + 13\gamma + 2\gamma^2\right\}^{-1} \gamma^{\#\mathcal{K}_0(x)} \tag{A1}$$

where $\gamma \neq 1$ is a positive parameter. Suppose this is of the form (4), i.e.

$$p(x) = c \prod_{i=1}^{4} \phi_{\{i\}}(x_i)\phi_{\{i,i+1\}}(x_i, x_{i+1}) \tag{A2}$$

where $\phi_{\{4,5\}} = \phi_{\{0,4\}}$, $x_5 = x_0$, and c is a normalizing constant. Since $\phi_{\{i\}}(0) = \phi_{\{i,i+1\}}(0, \cdot) = \phi_{\{i,i+1\}}(\cdot, 0) = 1$, it is easily seen that $\phi_{\{i\}}(1) = \alpha$ and $\phi_{\{i,i+1\}}(1, 1) = 1/\alpha$ for all $i \in I$, where α is a positive constant. Hence, by (A2),

$$p(x) = \left\{2 + 12\alpha + 2\alpha^2\right\}^{-1} \alpha^{\sum_{i=1}^{4} x_i(1-x_{i+1})}$$

which is clearly not identical to (A1). We thus have a contradiction, so (A1) can not be a MRF. □

Appendix B: Proof of Theorem 1

The following lemma will be needed in the proof of Theorem 1.

Lemma B Assume that $p(\cdot) > 0$ is a density defined on S and that $\sim$ is a symmetric and reflexive relation such that

$$\forall\, x \in S:\ p(x) = \psi_\emptyset \prod_{C\in\mathcal{C}} \psi_C(x_C) = \phi_\emptyset \prod_{C\in\mathcal{C}} \phi_C(x_C)$$

where $\mathcal{C}$ is the set of $\sim$-cliques and for each $C \in \mathcal{C}$, ψ_C and ϕ_C are positive functions defined on V^C with $\phi_C(x_C) = 1$ whenever there exists an $i \in C$ such that $x_i = 0$. Then for every $C \in \mathcal{C} \cup \{\emptyset\}$ and $x_C \in V^C$ we have that

$$\phi_C(x_C) = \prod_{\substack{B\in\mathcal{C}\cup\{\emptyset\}:\\ C\subseteq B}} \prod_{A\subseteq C} \psi_B\left(x_B^A\right)^{(-1)^{|C\setminus A|}}$$

with $x_B^A = (x_i^A)_{i\in B}$ given by

$$x_i^A = \begin{cases} x_i & i \in A \\ 0 & \text{otherwise.} \end{cases}$$

Proof: By Möbius inversion we have for every $C \in \mathcal{C} \cup \{\emptyset\}$ that

$$\begin{aligned}\phi_C(x_C) &= \prod_{A\subseteq C} p\left(x^A\right)^{(-1)^{|C\setminus A|}} = \prod_{A\subseteq C} \left(\psi_\emptyset \prod_{B\in\mathcal{C}} \psi_B\left(x_B^A\right)\right)^{(-1)^{|C\setminus A|}} \\ &= \prod_{A\subseteq C} \prod_{\substack{B\in\mathcal{C}\cup\{\emptyset\}:\\ C\subseteq B}} \psi_B\left(x_B^A\right)^{(-1)^{|C\setminus A|}} \prod_{A\subseteq C} \prod_{\substack{B\in\mathcal{C}\cup\{\emptyset\}:\\ C\not\subseteq B}} \psi_B\left(x_B^A\right)^{(-1)^{|C\setminus A|}}.\end{aligned}$$

Further,

$$\prod_{A\subseteq C} \prod_{\substack{B\in\mathcal{C}\cup\{\emptyset\}:\\ C\not\subseteq B}} \psi_B\left(x_B^A\right)^{(-1)^{|C\setminus A|}} = \prod_{\substack{B\in\mathcal{C}\cup\{\emptyset\}:\\ C\not\subseteq B}} \prod_{A\subseteq C\cap B} \prod_{A'\subseteq C\setminus B} \psi_B\left(x_B^A\right)^{(-1)^{|C\setminus(A\cup A')|}}$$

where

$$\prod_{A'\subseteq C\setminus B} \psi_B\left(x_B^A\right)^{(-1)^{|C\setminus(A\cup A')|}} = \psi_B\left(x_B^A\right)^{(-1)^{|C\setminus A|}(-1)^{|C\setminus B|}\sum_{A'\subseteq C\setminus B}(-1)^{|C\setminus B|-|A'|}}$$

for $A \subseteq B \cap C$. Using the binomial formula

$$\sum_{A'\subseteq C\setminus B} (-1)^{|C\setminus B|-|A'|} = 0$$

which inserted in the previous expressions concludes the proof. □

Proof of Theorem 1: Assume that $p(\cdot)$ is a positive density of a random field which is both a MRF and a MCCF. By assumption

$$p(x) = \phi_\emptyset \prod_{C \in \mathcal{C}} \phi_C(x_C) = c \prod_{K \in \mathcal{K}(x)} \Psi_K(l(x_K))$$

for normalizing constants $\phi_\emptyset$ and c, a positive function Ψ defined on $\mathcal{K} \times \mathcal{V}$, and positive functions $\phi_C(\cdot)$ with $\phi_C(x_C) = 1$ whenever $x_i = 0$ for some $i \in C$. The first step of the proof is to prove that $\phi_C(x_C)$ is equal to a constant c_C whenever $x_i \neq 0$ for all $i \in C$ and when x_C is not constant (i.e. there exists $i, j \in C$ with $x_i \neq x_j$). Let for any $A \subseteq I$,

$$k_A = \frac{c}{\phi_\emptyset} \prod_{K \in \mathcal{K}(0_{A^c})} \Psi_K(0)$$

where 0_{A^c} is the element in V^{A^c} whose coordinates are all zero. Consider any $C \in \mathcal{C}$. Then for $l \in V$,

$$\prod_{\substack{A \subseteq C: \\ A \neq \emptyset}} \phi_A(l, \ldots, l) = k_C \Psi_C(l).$$

Thus for $C \in \mathcal{C} \setminus I$ and $x \in S$ such that x_C with $x_i \neq 0$ if $i \in C$ and $x_i = 0$ if $i \notin C$,

$$p(x) = \phi_\emptyset \prod_{\substack{A \subseteq C \\ A \neq \emptyset}} \phi_A(x_A) = \phi_\emptyset k_C \prod_{K \in \mathcal{K}(x_C)} \Psi_K(l(x_K))$$

so

$$\begin{aligned} \left(\prod_{K \in \mathcal{K}(x_C)} k_K \right) \prod_{\substack{A \subseteq C \\ A \neq \emptyset}} \phi_A(x_A) &= k_C \prod_{K \in \mathcal{K}(x_C)} k_K \psi_K(l(x_K)) \\ &= k_C \prod_{K \in \mathcal{K}(x_C)} \prod_{\substack{A \subseteq K \\ A \neq \emptyset}} \phi_A(l(x_K), \ldots, l(x_K)) \end{aligned}$$

and hence

$$\phi_C(x_C) = \frac{k_C}{\prod_{K \in \mathcal{K}(x_C)} k_K} \Big(\prod_{\substack{A \subset C: A \neq \emptyset \\ x_A \text{ not constant}}} \phi_A(x_A) \Big)^{-1}. \tag{B1}$$

We now prove by induction that if $C \in \mathcal{C}$, x_C is not constant, and $x_i \neq 0$ for all $i \in C$, then

$$\phi_C(x_C) = \prod_{\substack{A \subseteq C: \\ A \neq \emptyset}} k_A^{(-1)^{|C \setminus A|}} = c_C \tag{B2}$$

is constant. Assume first that $C = \{i, j\} \in \mathcal{C}$ and $x_{\{i,j\}} = (l, l')$ where $l, l' \neq 0$ and $l \neq l'$. Then (B1) implies (B2). Consider next a $C \in \mathcal{C}$ where $|C| > 2$. Then by (B1)

and the induction hypothesis

$$\phi_C(x_C) = \frac{k_C}{\prod\limits_{K\in\mathcal{K}(x_C)} k_K}\Bigg(\prod_{\substack{B\subset C:\\ \forall K\in\mathcal{K}(x_C):B\not\subseteq K}}\prod_{\substack{A\subseteq B:\\ A\neq\emptyset}} k_A^{(-1)^{|B\setminus A|}}\Bigg)^{-1}$$
$$= \frac{k_C}{\prod\limits_{K\in\mathcal{K}(x_C)} k_K}\Bigg(\prod_{\substack{A\subset C:\\ A\neq\emptyset}} (k_A)^{\sum\limits_{\substack{A\subseteq B\subset C:\\ \forall K\in\mathcal{K}(x_C):B\not\subseteq K}}(-1)^{|B\setminus A|}}\Bigg)^{-1}.$$

Since

$$\sum_{B\colon A\subseteq B\subseteq C}(-1)^{|B\setminus A|} = \sum_{B\colon B\subseteq C\setminus A}(-1)^{|B|} = 0$$

we get that

$$\sum_{\substack{B\colon A\subseteq B\subset C,\\ \forall K\in\mathcal{K}(x_C):B\not\subseteq K}}(-1)^{|B\setminus A|} = \sum_{B\colon A\subseteq B\subset C}(-1)^{|B\setminus A|} - \sum_{\substack{B\colon A\subseteq B\subset C,\\ \exists K\in\mathcal{K}(x_C):B\subseteq K}}(-1)^{|B\setminus A|}$$
$$= \begin{cases} -(-1)^{|C\setminus A|} & \text{if } \forall K\in\mathcal{K}(x_C): A\not\subseteq K\\ -(-1)^{|C\setminus A|}-1 & \text{if } A\in\mathcal{K}(x_C)\\ -(-1)^{|C\setminus A|} & \text{if } \exists K\in\mathcal{K}(x_C): A\subset K\end{cases}$$

where the last case follows because

$$\sum_{B\colon A\subseteq B\subseteq K}(-1)^{|B\setminus A|} = \sum_{B\colon B\subseteq K\setminus A}(-1)^{|B|} = 0.$$

Thus,

$$\phi_C(x_C) = \frac{k_C}{\prod\limits_{K\in\mathcal{K}(x_C)} k_K}\Big(\prod_{\substack{A\subset C:\\ A\neq\emptyset}} k_A^{(-1)^{|C\setminus A|}}\Big)\Big(\prod_{K\in\mathcal{K}(x_C)} k_K\Big)$$

and (B2) is proved.

Finally, we verify (5) by defining a function $\psi_B : V^B \to (0,\infty)$ for each $B\subset I$ as follows. Set

$$\psi_I(0_I) = \begin{cases} c_I^{(-1)^{|I|}} & \text{if } I\in\mathcal{C},\\ 1 & \text{else}\end{cases}$$

and define $\psi_B(0_B)$ inductively on $|B| = |I|-1, |I|-2,\dots,1$ by

$$\psi_B(0_B) = \begin{cases} c_B^{(-1)^{|B|}} \Big/ \prod\limits_{A\in\mathcal{C}:\, B\subset A}\psi_A(0_A) & \text{if } B\in\mathcal{C}\\ 1 & \text{else}\end{cases}$$

while we set

$$\psi_\emptyset = \phi_\emptyset \Big/ \prod_{C\in\mathcal{C}}\psi_C(0_C).$$

Further, set

$$\psi_B(x_B) = \begin{cases} \phi_B(x_B)/c_B & \text{if } x_B \text{ is constant and } \neq 0 \text{ and } B \in \mathcal{C} \\ 1 & \text{if } x_B \text{ is not constant or } B \notin \mathcal{C}. \end{cases}$$

Since $\{\phi_C\}_{C\in\mathcal{C}}$ is the unique representation of $p(x) = \phi_\emptyset \prod_{C\in\mathcal{C}} \phi_C(x_C)$ satisfying $\phi(x_C) = 1$ if $x_i = 0$ for some $i \in C$, it suffices by Lemma B to show that

$$\phi_C(x_C) = \prod_{\substack{B\in\mathcal{C}\cup\{\emptyset\} \\ C\subseteq B}} \prod_{A\subseteq C} \psi_B\left(x_B^A\right)^{(-1)^{|C\setminus A|}} \tag{B3}$$

for every $C \in \mathcal{C} \cup \{\emptyset\}$. By definition of ϕ_B the righthand side of (B3) is one if $x_i = 0$ for some $i \in C$ and (B3) therefore holds in this case. If $C = \emptyset$ then (B3) is also easily seen to hold, so assume that $|C| > 0$ and that $x_i \neq 0$ for all $i \in C$. Then by definition of $\psi_B(\cdot)$,

$$\begin{aligned}
&\prod_{\substack{B\in\mathcal{C}\cup\{\emptyset\}: \\ C\subseteq B}} \prod_{A:\, A\subseteq C} \psi_B\left(x_B^A\right)^{(-1)^{|C\setminus A|}} \\
&= \psi_C(x_C)^{(-1)^0} \left\{ \prod_{B\in\mathcal{C}\cup\{\emptyset\}:\, C\subseteq B} \psi_B(0_B)^{(-1)^{|C|}} \right\} \\
&\left\{ \prod_{\substack{B:\, B\in\mathcal{C}\cup\{\emptyset\}, \\ C\subset B}} \prod_{A:\, \emptyset\neq A\subseteq C} \psi_B\left(x_B^A\right)^{(-1)^{|C\setminus A|}} \right\} \left\{ \prod_{A:\, \emptyset\neq A\subset C} \psi_B\left(x_B^A\right)^{(-1)^{|C\setminus A|}} \right\} \\
&= \psi_C(x_C)^{(-1)^0} \left\{ \prod_{B\in\mathcal{C}\cup\{\emptyset\}:\, C\subseteq B} \psi_B(0_B)^{(-1)^{|C|}} \right\} \\
&= c_C \left(\frac{\phi_C(x_C)}{c_C} \right)^{1(x_C \text{ constant})} \\
&= \phi_C(x_C)
\end{aligned}$$

where the last equality follows from (B2). □

Appendix C: Proofs of Lemma 1, Lemma 2, and Theorem 4

Observe that (15) implies that for every $\Lambda \in \mathcal{G}$ there is a constant c_Λ such that

$$\Psi_K(\cdot) = 0 \Rightarrow |K| \leq c_\Lambda \ \forall K \in \mathcal{K} \text{ with } K \uparrow \bar{\Lambda}. \tag{C1}$$

We need also the following result:

Lemma C For any $K \in \mathcal{K}$ and $\Lambda \in \mathcal{G}$ such that $K \uparrow \bar{\Lambda}$ let

$$K_n(K, \Lambda) = \cup_{\substack{K'\in\mathcal{K}:\mathcal{K}'\uparrow\bar{\Lambda}, \\ K'\subseteq K\cap\Lambda_n}} K'.$$

Then there exists an $N = N(K,\Lambda) \geq 1$ such that $K_n(K,\Lambda) \in \mathcal{K}$ for all $n \geq N$.

Proof: This easily follows by using that $\sim$ is of finite range and that $\bar{\Lambda}$ is finite. □

Proof of Lemma 1: Let $\Lambda \in \mathcal{G}$ and $x \in R_\Lambda$ be given. By Lemma C and since $\left|\bar{\Lambda}\right| < \infty$, we can choose $N(x,\Lambda)$ such that for all $n \geq N(x,\Lambda)$ and $K \in \mathcal{K}(x)$ with $K \uparrow \bar{\Lambda}$ either $K_n(K,\Lambda) \in \mathcal{K}$ if K is infinite or $K \subseteq \Lambda_n$ if K is finite. Choose further N_1 such that $K \subseteq \Lambda_{N_1} \setminus \partial\Lambda_{N_1}^c$ for all $K \in \mathcal{K}$ with both $K \uparrow \bar{\Lambda}$ and $|K| \leq c_\Lambda$ (here we have used again that $\left|\bar{\Lambda}\right| < \infty$ together with the assumption that the sequence $\{\Lambda_n \setminus \partial(\Lambda_n^c)\}$ contains any finite subset of I for n sufficiently large). Note that for $n \geq N_1$ and $K \in \mathcal{K}$ with $K \uparrow \bar{\Lambda}$ and $|K| \leq c_\Lambda$, we have by definition of N_1 and since $\{\Lambda_n\}$ is increasing, that $K \in \mathcal{K}(y_{\Lambda_n})$ if and only if $K \in \mathcal{K}(y_{\Lambda_{N_1}})$. Hence by (C1), for all $n \geq N_1$,

$$\begin{aligned} R_{\Lambda,n} &= \Big\{ z \in S \mid \exists\, y \in S : z_{\Lambda^c} = y_{\Lambda^c} \text{ and } \prod_{\substack{K \in \mathcal{K}\left(y_{\Lambda_n}\right): \\ K \uparrow \bar{\Lambda}, |K| \leq c_\Lambda}} \Psi_K(l(y_K)) > 0 \Big\} \\ &= \Big\{ z \in S \mid \exists\, y \in S : z_{\Lambda_{N_1}\setminus\Lambda} = y_{\Lambda_{N_1}\setminus\Lambda} \text{ and } \prod_{\substack{K \in \mathcal{K}\left(y_{\Lambda_{N_1}}\right): \\ K \uparrow \bar{\Lambda}, |K| \leq c_\Lambda}} \Psi_K(l(y_K)) > 0 \Big\} \end{aligned}$$

so that also

$$R_\Lambda = \Big\{ z \in S \mid \exists\, y \in S : z_{\Lambda_{N_1}\setminus\Lambda} = y_{\Lambda_{N_1}\setminus\Lambda} \text{ and } \prod_{\substack{K \in \mathcal{K}\left(y_{\Lambda_{N_1}}\right): \\ K \uparrow \bar{\Lambda}, |K| \leq c_\Lambda}} \Psi_K(l(y_K)) > 0 \Big\}. \quad \text{(C2)}$$

Let

$$N_2 = \max\{\max\{N(y,\Lambda) \mid y \in S : y_{\Lambda^c} = x_{\Lambda^c}\}, N_1\}$$

and y be a state in S with $y_{\Lambda^c} = x_{\Lambda^c}$. Then $K \cap \Lambda_n \in \mathcal{K}$ for all $n \geq N_2$ and $K \in \mathcal{K}(y)$ with $K \uparrow \bar{\Lambda}$. We can therefore define for $n \geq N_2$,

$$q_{\Lambda,n}(y) = \prod_{\substack{K \in \mathcal{K}(y): \\ K \uparrow \bar{\Lambda}}} \Psi_{K\cap\Lambda_n}(l(y_{K\cap\Lambda_n})). \quad \text{(C3)}$$

Hence (13) gives that

$$p_{\Lambda,n}(x) = \frac{q_{\Lambda,n}(x)}{\sum_{y : y_{\Lambda^c} = x_{\Lambda^c}} q_{\Lambda,n}(y)}$$

for all $n \geq N_2$; here the denumerator is nonzero as $x \in R_\Lambda$.

Consider now the finite set

$$U(\Lambda, x_{\Lambda^c}) = \big\{ K \in \mathcal{K}(x_{\Lambda^c}) \mid K \uparrow \bar{\Lambda} \text{ and } |K| = \infty \big\}.$$

If $U(\Lambda, x_{\Lambda^c})$ is empty, then in (C3) we have for all sufficiently large n and $K \in \mathcal{K}(x_{\Lambda^c})$ with $K \uparrow \bar{\Lambda}$ that $\Psi_{K\cap\Lambda_n}(l(y_{K\cap\Lambda_n})) = \Psi_K(l(y_K))$, and so we can immediately obtain

$p_\Lambda(x_\Lambda|x_{\Lambda^c})$ as $\lim_{n\to\infty} p_{\Lambda,n}(x)$ whereby (16) is verified. Assume instead that $U(\Lambda, x_{\Lambda^c})$ is nonempty and consider any $y \in S$ with $y_{\Lambda^c} = x_{\Lambda^c}$. Let

$$\big\{K \in \mathcal{K}(y) \mid K \uparrow \bar{\Lambda} \text{ and } |K| = \infty\big\} = \{K_1, \dots, K_m\}$$

where $1 \le m < \infty$ as $|\bar{\Lambda}| < \infty$. Then $U(\Lambda, x_{\Lambda^c})$ is the disjoint union

$$U(\Lambda, x_{\Lambda^c}) = \bigcup_{i=1}^{m} U_i(\Lambda, x_{\Lambda^c})$$

where $U_i(\Lambda, x_{\Lambda^c})$ consists of those $K \in U(\Lambda, x_{\Lambda^c})$ with $K \subseteq K_i$. Let $U_i(\Lambda, x_{\Lambda^c}) = \{K_{i1}, \dots, K_{im_i-1}\}$, $2 \le m_i < \infty$. Then $K_i = \cup_{j=1}^{m_i} K_{ij}$ where $K_{im_i} = K_i \setminus \cup_{j=1}^{m_i-1} K_{ij} \subseteq \Lambda$, so $|K_{im_i}| < \infty$. By Lemma C and (C1) we find for all sufficiently large n that $K_i \cap \Lambda_n \in \mathcal{K}$ and $K_{ij} \cap \Lambda_n \in \mathcal{K}$, $j = 1, \dots, m_i - 1$, $i = 1, \dots, m$. Further, if we consider so large n that $|K_i \cap \Lambda_n| > c_\Lambda$ and $|K_{ij} \cap \Lambda_n| > c_\Lambda$ for $j = 1, \dots, m_i - 1$, then by (C1), $\Psi_{K_i \cap \Lambda_n}(\cdot) > 0$ and $\Psi_{K_{ij} \cap \Lambda_n}(\cdot) > 0$, so the ratio

$$\frac{\Psi_{K_i \cap \Lambda_n}(l)}{\prod_{j=1}^{m_i-1} \Psi_{K_{ij} \cap \Lambda_n}(l)}$$

becomes well-defined for all sufficiently large n and any $l \in V$. Since the same relation $\sim$ is used to define cliques and connected components we see that

$$\begin{aligned}
&\lim_{n\to\infty} \frac{\Psi_{K_i \cap \Lambda_n}(l)}{\prod_{j=1}^{m_i-1} \Psi_{K_{ij} \cap \Lambda_n}(l)} \\
&= \lim_{n\to\infty} \frac{\Psi_{K_i \cap \Lambda_n}(l)}{\prod_{\substack{C \in \mathcal{C}: \\ C \subseteq K_i \cap \Lambda_n}} \phi_C(l)} \prod_{j=1}^{m_i-1} \frac{\prod_{C \in \mathcal{C}: C \subseteq K_{ij} \cap \Lambda_n} \phi_C(l)}{\Psi_{K_{ij} \cap \Lambda_n}(l)} \prod_{\substack{C \in \mathcal{C}: C \subseteq K_i \cap \Lambda_n, \\ C \uparrow K_{im_i} \cap \Lambda_n}} \phi_C(l) \\
&= \prod_{\substack{C \in \mathcal{C}: \\ C \uparrow K_{im_i}}} \phi_C(l)
\end{aligned}$$

where we have used (15) and the fact that $|K_{im_i}| < \infty$ to obtain the last identity. Using this result we can define q_Λ by

$$\begin{aligned}
q_\Lambda(y) &= \lim_{n\to\infty} \frac{q_{\Lambda,n}(y)}{\prod_{K' \in U(\Lambda, x_{\Lambda^c})} \Psi_{K' \cap \Lambda_n}(l(x_{K' \cap \Lambda_n}))} \\
&= \lim_{n\to\infty} \prod_{\substack{K \in \mathcal{K}(y): K \uparrow \bar{\Lambda}, \\ |K| < \infty}} \Psi_{K \cap \Lambda_n}(l(y_{K \cap \Lambda_n})) \prod_{i=1}^{m} \frac{\Psi_{K_i \cap \Lambda_n}(l(y_{K_i \cap \Lambda_n}))}{\prod_{j=1}^{m_i-1} \Psi_{K_{ij} \cap \Lambda_n}\big(l\big(y_{K_{ij} \cap \Lambda_n}\big)\big)} \qquad \text{(C4)} \\
&= \prod_{\substack{K \in \mathcal{K}(y): K \uparrow \bar{\Lambda}, \\ |K| < \infty}} \Psi_K(l(y_K)) \prod_{\substack{K \in \mathcal{K}(y): K \uparrow \bar{\Lambda}, \\ |K| = \infty}} \prod_{\substack{C \in \mathcal{C}: C \subseteq K, \\ C \uparrow K \setminus \cup_{K' \in U(\Lambda, y_{\Lambda^c})} K'}} \phi_C(l(y_K))
\end{aligned}$$

and then finally obtain (16) as

$$\begin{aligned}\lim_{n\to\infty} p_{\Lambda,n}(x) &= \lim_{n\to\infty} \frac{q_{\Lambda,n}(x)}{\sum\limits_{y:\, y_{\Lambda^c}=x_{\Lambda^c}} q_{\Lambda,n}(y)} \\ &= \lim_{n\to\infty} \frac{q_{\Lambda,n}(x) / \prod\limits_{K'\in U(\Lambda,x_{\Lambda^c})} \Psi_{K'\cap\Lambda_n}(l(x_{K'\cap\Lambda_n}))}{\sum\limits_{y:\, y_{\Lambda^c}=x_{\Lambda^c}} q_{\Lambda,n}(y) / \prod\limits_{K'\in U(\Lambda,x_{\Lambda^c})} \Psi_{K'\cap\Lambda_n}(l(x_{K'\cap\Lambda_n}))} \\ &= \frac{q_\Lambda(x)}{\sum\limits_{y:y_{\Lambda^c}=x_{\Lambda^c}} q_\Lambda(y)}.\end{aligned}$$

□

Proof of Lemma 2: Let $\Lambda \in \mathcal{G}$ be given and define for all $y \in S$ and all n such that $\bar{\Lambda} \subseteq \Lambda_n$,

$$\begin{aligned}\tilde{q}_{\Lambda,n}(y) = &\left\{ \prod_{\substack{K\in\mathcal{K}(y):|K|<\infty \\ K\uparrow\bar{\Lambda}, K\gamma\{\Lambda_n^c\cup\partial(\Lambda_n^c)\}}} \Psi_K(l(y_K)) \right\} \left\{ \prod_{\substack{K\in\mathcal{K}(y):|K|<\infty \\ K\uparrow\bar{\Lambda}, K\uparrow\{\Lambda_n^c\cup(\partial\Lambda_n^c)\}}} \prod_{\substack{C\in\mathcal{C}: \\ C\subseteq K}} \phi_C(l(y_K)) \right\} \\ &\times \left\{ \prod_{\substack{K\in\mathcal{K}(y):K\uparrow\bar{\Lambda}, \\ |K|=\infty}} \prod_{\substack{C\in\mathcal{C}:C\subseteq K, \\ C\uparrow K\setminus V(\Lambda,y_{\Lambda^c})}} \phi_C(l(y_K)) \right\}\end{aligned} \tag{C5}$$

where $V(\Lambda, y_{\Lambda^c}) = \cup\{K \in U(\Lambda, y_{\Lambda^c})\}$ and $U(\Lambda, y_{\Lambda^c})$ is defined as in the proof of Lemma 1. We shall later on use $\tilde{q}_{\Lambda,n}$ to approximate q_Λ in (C4). If we choose $\tilde{N}$ such that for all $n \geq \tilde{N}$ and all $K \in \mathcal{K}$ with $K \uparrow \bar{\Lambda}$ and $|K| \leq c_\Lambda$ we have that $K \subseteq \Lambda_n \setminus \partial(\Lambda_n^c)$, then by similar arguments as in the proof of Lemma 1 the function $r_{\Lambda,n}$ given by

$$r_{\Lambda,n}(x_\Lambda|x_{\Lambda^c}) = \frac{\tilde{q}_{\Lambda,n}(x)}{\sum\limits_{y:y_{\Lambda^c}=x_{\Lambda^c}} \tilde{q}_{\Lambda,n}(y)},\ x \in R_\Lambda \tag{C6}$$

is seen to be well-defined for all $n \geq \tilde{N}$.

Let $n \geq \tilde{N}$ in the following. We will show first that $r_{\Lambda,n}$ is local, and secondly that $r_{\Lambda,n}$ converges uniformly to p_Λ as $n \to \infty$.

Consider any $x \in R_\Lambda$ and assume that there is some $K \in \mathcal{K}(x)$ with $K \uparrow \bar{\Lambda}$ and some $C \in \mathcal{C}$ with $C \subseteq K$ and $C \not\subseteq \Lambda_n$ such that the expression for $\tilde{q}_{\Lambda,n}(x)$ given by (C5) includes the term $\phi_C(l(x_K))$. We will now show that this term occurs in the expression for $\tilde{q}_{\Lambda,n}(y)$ for all $y \in S$ with $y_{\Lambda^c} = x_{\Lambda^c}$ so that $\phi_C(l(x_K))$ cancels out in (C6). Since $C \in \mathcal{C}$, $C \not\subseteq \Lambda_n$, and $\bar{\Lambda} \subseteq \Lambda_n$ it follows that $C \not\uparrow \Lambda$. There exists therefore a $K_1 \in \mathcal{K}(x_{\Lambda^c})$ with $K_1 \uparrow \bar{\Lambda}$ such that $C \subseteq K_1 \subseteq K \setminus \Lambda$. Since $C \not\subseteq \Lambda_n$ it is clear that $K_1 \uparrow \{\Lambda_n^c \cup \partial(\Lambda_n^c)\}$. Consider further any $y \in S$ with $y_\Lambda^c = x_\Lambda^c$. Since $K_1 \uparrow \bar{\Lambda}$ there

is a $K_2 \in \mathcal{K}(y)$ with $K_1 \subseteq K_2$, and so $K_2 \uparrow \bar{\Lambda}$ and $K_2 \uparrow \{\Lambda_n^c \cup \partial(\Lambda_n^c)\}$. Clearly, if $|K_2| < \infty$, the term $\phi_C(l(x_K))$ will appear in the expression (C5) for $\tilde{q}_{\Lambda,n}(y)$. Note that $C \not\subseteq V(\Lambda, x_{\Lambda^c})$ because otherwise $\phi_C(l(x_K))$ could not appear in the expression (C5) for $\tilde{q}_{\Lambda,n}(x)$. This means that $C \uparrow K_2 \setminus V(\Lambda, x_{\Lambda^c})$ so that also in the case $|K_2| = \infty$, the term $\phi_C(l(x_K))$ appears in the expression (C5) for $\tilde{q}_{\Lambda,n}(y)$.

We can therefore replace $\tilde{q}_{\Lambda,n}(x)$ and $\tilde{q}_{\Lambda,n}(y)$ in (C6) with $q'_{\Lambda,n}(x)$ and $q'_{\Lambda,n}(y)$, respectively, where

$$q'_{\Lambda,n}(y) = \prod_{\substack{K \in \mathcal{K}(y):|K|<\infty \\ K \uparrow \bar{\Lambda}, K \not\uparrow \{\Lambda_n^c \cup \partial(\Lambda_n^c)\}}} \Psi_K(l(y_K)) \prod_{\substack{K \in \mathcal{K}(y):|K|<\infty \\ K \uparrow \bar{\Lambda}, K \uparrow \{\Lambda_n^c \cup \partial(\Lambda_n^c)\}}} \prod_{\substack{C \in \mathcal{C}: \\ C \subseteq K \cap \Lambda_n}} \phi_C(l(y_K)) \prod_{\substack{K \in \mathcal{K}(y): K \uparrow \bar{\Lambda}, \\ |K| = \infty}} \prod_{\substack{C \in \mathcal{C}: C \subseteq K \cap \Lambda_n, \\ C \uparrow K \setminus V(\Lambda, y_{\Lambda^c})}} \phi_C(l(y_K)).$$

For $x_\Lambda^c = y_\Lambda^c$ we get

$$\begin{aligned} & q'_{\Lambda,n}(y) \prod_{K \in U(\Lambda, x_{\Lambda^c})} \prod_{\substack{C \in \mathcal{C}: \\ C \subseteq K \cap \Lambda_n}} \phi_C(l(x_K)) \\ = & \prod_{\substack{K \in \mathcal{K}(y): \\ K \uparrow \bar{\Lambda}, K \not\uparrow \{\Lambda_n^c \cup \partial(\Lambda_n^c)\}}} \Psi_K(l(y_K)) \prod_{\substack{K \in \mathcal{K}(y): \\ K \uparrow \bar{\Lambda}, K \uparrow \{\Lambda_n^c \cup \partial(\Lambda_n^c)\}}} \prod_{\substack{C \in \mathcal{C}: \\ C \subseteq K \cap \Lambda_n}} \phi_C(l(y_K)) \end{aligned}$$

which depends on y only through y_{Λ_n}. Consequently, as the terms $\phi_C > 0$,

$$\begin{aligned} r_{\Lambda,n}(x_\Lambda \mid x_{\Lambda^c}) &= \frac{q'_{\Lambda,n}(x)}{\sum_{y \in S: y_{\Lambda^c} = x_{\Lambda^c}} q'_{\Lambda,n}(y)} \\ &= \frac{q'_{\Lambda,n}(x) \prod_{K \in U(\Lambda, x_{\Lambda^c})} \prod_{\substack{C \in \mathcal{C}: \\ C \subseteq K \cap \Lambda_n}} \phi_C(l(x_K))}{\sum_{y_{\Lambda_n} \in V^{\Lambda_n}: y_{\Lambda_n \setminus \Lambda} = x_{\Lambda_n \setminus \Lambda}} q'_{\Lambda,n}(y) \prod_{K \in U(\Lambda, x_{\Lambda^c})} \prod_{\substack{C \in \mathcal{C}: \\ C \subseteq K \cap \Lambda_n}} \phi_C(l(x_K))} \end{aligned}$$

depends on $x \in S$ only through x_{Λ_n}, and it is thus established that $r_{\Lambda,n}$ is a local function.

Consider now the signed measure $\mu_{\Lambda,n}(\cdot \mid x_{\Lambda^c}) = p_\Lambda(\cdot \mid x_{\Lambda^c}) - r_{\Lambda,n}(\cdot \mid x_{\Lambda^c})$ for $x \in R_\Lambda$. The total variation norm of this measure is

$$\left\| \mu_{\Lambda,n}(\cdot \mid x_{\Lambda^c}) \right\| = \frac{1}{2} \sum_{z_\Lambda \in V^\Lambda} \left| p_\Lambda(z_\Lambda \mid x_{\Lambda^c}) - r_{\Lambda,n}(z_\Lambda \mid x_{\Lambda^c}) \right|$$

and since V^Λ is finite, $r_{n,\Lambda}$ converges uniformly to p_Λ if and only if

$$\lim_{n \to \infty} \sup_{x \in R_\Lambda} \left\| \mu_{\Lambda,n}(\cdot \mid x_{\Lambda^c}) \right\| = 0.$$

Since $n \geq \tilde{N}$ and $\phi_C(\cdot) > 0$ we get by (C1) and (C4) that for all $x \in R_\Lambda$,

$$\begin{aligned} q_\Lambda(x) = 0 &\Leftrightarrow \exists K \in \mathcal{K}(x) : K \uparrow \bar{\Lambda},\ |K| < c_\Lambda,\ \Psi_K(l(x_K)) = 0 \\ &\Leftrightarrow \tilde{q}_{\Lambda,n}(x) = 0. \end{aligned}$$

We can then define

$$v_{\Lambda,n}(y) = \ln \frac{\tilde{q}_{\Lambda,n}(x)}{q_\Lambda(x)}$$

taking $\frac{0}{0} = 1$. Then, following the second part of the proof of Proposition 8.8 in Georgii (1988) (with $u_o(x) = \ln\left(q_\Lambda(x)\right)$ and $u_1(x) = \ln\left(\tilde{q}_{\Lambda,n}(x)\right)$ if $q_\Lambda(x) > 0$ and $u_0(x) = u_1(x) = 0$ else), we get that

$$2\left\|\mu_{\Lambda,n}(\cdot \mid x_{\Lambda^c})\right\| \le \delta\left(v_{\Lambda,n}\right)/2$$

where

$$\delta\left(v_{\Lambda,n}\right) = \sup_{x,y \in R_\Lambda} \left|v_{\Lambda,n}(x) - v_{\Lambda,n}(y)\right|$$

is the oscillation of $v_{\Lambda,n}$. Note that by (C4) and (C5),

$$v_{\Lambda,n}(x) = \ln\Bigg(\prod_{\substack{K \in \mathcal{K}(x):|K|<\infty \\ K \uparrow \bar{\Lambda},\, K \uparrow \{\Lambda_n^c \cup \partial(\Lambda_n^c)\}}} \frac{\prod\limits_{\substack{C \in \mathcal{C}: \\ C \subseteq K}} \phi_C(l(x_C))}{\Psi_K(l(x_K))} \Bigg).$$

Using (15) we can for any $\epsilon > 0$ find a $N \ge \tilde{N}$ such that for all $n \ge N$ and all $x \in R_\Lambda$,

$$(1-\epsilon)^{|\bar{\Lambda}|} \le \Bigg| \prod_{\substack{K \in \mathcal{K}(x):|K|<\infty \\ K \uparrow \bar{\Lambda},\, K \uparrow \{\Lambda_n^c \cup \partial(\Lambda_n^c)\}}} \frac{\Psi_K(l(x_K))}{\prod\limits_{\substack{C \in \mathcal{C}: \\ C \subseteq K}} \phi_C(x_C)} \Bigg| \le (1+\epsilon)^{|\bar{\Lambda}|}.$$

This implies that for $n \ge N$,

$$\delta\left(v_{\Lambda,n}\right) \le \left|\bar{\Lambda}\right|\left[\ln(1+\epsilon) - \ln(1-\epsilon)\right]$$

and thereby p_Λ is shown to be quasilocal. □

Proof of Theorem 4: The family of Markov kernels $\{\pi_\Lambda\}_{\Lambda \in \mathcal{G}}$ defined by

$$\pi_\Lambda(x, F) = \sum_{y:\, y_{\Lambda^c} = x_{\Lambda^c},\, y \in F} p_\Lambda(y_\Lambda \mid y_{\Lambda^c})$$

for all $x \in R_\Lambda$ and $F \subseteq S$ is seen to be a specification as defined in Preston (1976, p. 16–17), since Preston's consistency condition (2.14) is easily verified by using his formulae (6.11) and (6.12). Let $\Lambda \in \mathcal{G}$ be given and choose $N_1 = N(\Lambda)$ as in the proof of Lemma 1. By (C2),

$$R_\Lambda = \Big\{x \in S \mid \exists\, y \in S \text{ with } y_{\Lambda_{N_1}\setminus\Lambda} = x_{\Lambda_{N_1}\setminus\Lambda} \text{ and } \prod_{\substack{K \in \mathcal{K}\left(y_{\Lambda_{N_1}}\right): \\ K \uparrow \bar{\Lambda},\, |K| \le c_\Lambda}} \Psi_K(l(y_K)) > 0\Big\} \tag{C5}$$

is a cylinder set (i.e. it is of the form $V^A \times B$ with $A \in \mathcal{G}$ and $B \subseteq V^{A^c}$), so Preston's condition (3.6) is satisfied. Let $\tilde{x} \in S$ be given by $\tilde{x}_i = \text{const} \in V$ for all $i \in I$.

Then $\tilde{x} \in R_\Lambda$ because $|\Lambda_{N_1}| > c_\Lambda$ by definition of N_1 and $\mathcal{K}(\tilde{x}_{\Lambda_{N_1}}) = \{\Lambda_{N_1}\}$ as all $\Lambda_n \in \mathcal{K}$. Since $x \in R_\Lambda$ for every $\Lambda \in \mathcal{G}$, the intersection $\cap_{\Lambda\in\mathcal{G}} R_\Lambda$ is nonempty, and so by verifying conditions (3.8) and (3.10) in Preston (1976) the first part of Theorem 4 follows from Theorem 3.2 in Preston (1976). But Preston's condition (3.8) is just quasilocality as verified in Lemma 2. Furthermore, his condition (3.10) is easily verified because $|V^\Lambda|$ is finite. Using the additional assumption (19) the last part of the theorem follows immediately from Theorem 4.3 in Preston (1976). □

References

Aykroyd, R. G. & Green, P. J. (1991). "Global and local priors, and the location of lesions using gamma camera imagery". *Phil. Trans. R. Soc. London.*, A 337:323–342.

Baddeley, A. J.; Kendall, W. S. & Van Lieshout, M. N. M. (1996). "Quermass-integral interaction processes". Warwick Statistics Research Report 293.

Baddeley, A. J. & Møller, J. (1989). "Nearest-neighbour Markov point processes and random sets". *Int. Statist. Rev.*, 57:89–121.

Baddeley, A. J. & Van Lieshout, M. N. M. (1993). "Stochastic geometry models in highlevel vision". In *Statistics and images, Advances in Applied Statistics, a supplement to the Journal of Applied Statistics*, volume 20, pages 231–256.

Baddeley, A. J.; Van Lieshout, M. N. M. & Møller, J. (1996). "Markov properties of cluster processes". *Adv. Appl. Prob. (SGSA)*, 28:346–355.

Besag, J. E. (1975). "Statistical analysis of non-lattice data". *The Statistician*, 24:179–195.

Besag, J. E. (1986). "On the Statistical Analysis of dirty pictures". *J. Roy. Statist. Soc.*, B 48:259–302.

Besag, J. E.; Milne, R. & Zachary, S. (1982). "Point process limits of lattice processes". *J. Appl. Prob.*, 19:210–216.

Brodatz, P. (1966). *Textures - a Photographic Album for Artists and Designers*. Dover, New York.

Chayes, J. T.; Chayes, L. & Kotecky, R. (1995). "The analysis of the Widow-Rowlinson model by stochastic geometric methods". *Commun. Math. Phys.*, 172:551–569.

Clifford, P. (1990). "Markov random fields in statistics". In Grimmett, G. R. & Welsh, D. J. A., editors, *Disorder in physical systems, A volume in honour of J. M. Hammersley*. Clarendon Press, Oxford.

Comets, F. (1989). "On consistency of a class of estimators for exponential families of Markov random fields on the lattice". Preprint, Université Paris- X et Laboratoire de Statistique Appliqué, Univ. Paris-Sud.

Cormen, T. H.; Leiserson, C. E. & Rivest, R. L. (1989). *Introduction to algorithms*. MacGraw-Hill.

Cross, G. & Jain, A. (1983). “Markov random field texture models”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 5:25–39.

Derin, H. & Elliot, H. (1987). “Modelling and segmentation of noisy and textured images using Gibbs random fields”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9:39–55.

Fortuin, C. M. & Kasteleyn, P. W. (1972). “On the random cluster model.I. Introduction and relation to other models”. *Physica*, 57(4):536–564.

Geman, S. & Geman, D. (1984). “Stochastic Relaxation, Gibbs distributions and the Bayesian restoration of images”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741.

Geman, S. & Graffigne, C. (1986). “Markov random fields and their applications to computer vision”. In Gleason, M., editor, *Proc. Internat. Congress of Mathematics 1986*, Providence, RI. Amer. Math. Soc.

Georgii, H.-O. (1988). *Gibbs Measures and Phase Transitions*. Walter de Gruyter, Berlin.

Georgii, H.-O. & Häggström, O. (1996). “Phase transition in continuum Potts models”. To appear.

Geyer, C. J. (1994). “On the convergence of Monte Carlo maximum likelihood calculations”. *J. Roy. Statist. Soc.*, B 56:261–274.

Geyer, C. J. & Thompson, E. A. (1992). “Constrained Monte Carlo maximum likelihood for dependent data”. *J. Roy. Statist. Soc.*, B 54:657–699.

Gidas, B. (1988). “Consistency of maximum likelihood and pseudolikelihood estimators for Gibbs distributions”. In *Stochastic Differential Systems, Stochastic Control Theory and Applications*, volume IMA 10. Springer, New York.

Green, P. J. (1986). “Contribution to the discussion of the paper by J. Besag”. *J. Roy. Statist. Soc.*, 48 B:284–285.

Grenander, U. & Miller, M. (1994). “Representations of knowledge in complex systems”. *J. Roy. Statist. Soc.*, B 56:549–603.

Grimmett, G. (1995). “The stochastic random-cluster process, and the uniqueness of random-cluster measures”. *Annals of Probability*. To appear.

Guyon, X. (1986). “Estimation d’un champ de Gibbs”. Preprint, Univ. Paris-I.

Häggström, O.; van Lieshout, M. N. M. & Møller, J. (1996). “Characterisation results and Markov chain Monte Carlo algorithms including exact simulation for some spatial point processes”. In preparation.

Haldorsen, H. H. & Damsleth, E. (1990). “Stochastic Modeling”. *Journal of Petroleum Technology*, pages 404–412.

Hastings, W. (1970). “Monte Carlo sampling methods using Markov chains and their applications”. *Biometrika*, 57:97–109.

Helterbrand, J. D.; Cressie, N. & Davidson, J. L. (1994). “A statistical approach to identifying closed object boundaries in images”. *Adv. Appl. Prob.*, 26:831–854.

Hjort, N. L. & Omre, H. (1995). “Topics in spatial statistics (with discussion)”. *Scand. J. Statist.*, 21:289–357.

Jacobsen, M. (1989). "Existence and unicity of MLE in discrete exponential distributions". *Scand. J. Statist.*, 16:335–349.

Ji, C. & Seymour, L. (1994). "Approximate Bayes model selection procedures for Gibbs-Markov fields". *Journal of Statistical Planning and inference*. To appear.

Johnson, V. E. (1994). "A model for segmentation and analysis of noisy images". *J. Amer. Statist. Assoc.*, 89:230–241.

Kendall, W. S. (1990). "A spatial Markov property for nearest-neighbour Markov point processes". *J. Appl. Prob.*, 28:767–778.

Kindermann, R. & Snell, J. L. (1980). *Markov Random Fields and their Applications*. American Mathematical Society, Providence, RI.

Klein, W. (1982). "Potts-model formulation of continuum percolation". *Physical Review B*, 26:2677–2678.

Künsch, H.; Geman, S. & Kehagias, A. (1995). "Hidden Markov random fields". *Ann. Appl. Prob.*, 5. To appear.

Lebowitz, J. L. & Gallawotti, G. (1971). "Phase transitions in binary lattice gases". *J. Math. Phys.*, 12:1129–1133.

Møller, J. (1994a). "Discussion contribution". *Scand. J. Statist.*, 21:346–349.

Møller, J. (1994b). "Markov chain Monte Carlo and spatial point processes". Research Report 293, Department of Theoretical Statistics, University of Aarhus. To appear in Barndorff-Nielsen O. E. et al., editors, *Proc. Seminaire Européen de Statistique Toulouse, Current trends in stochastic geometry with applications", 1996*.

Omre, H. & Syversveen, A.-R. (1995). "Marked point models with complex conditioning in reservoir characterization". Submitted for publication.

Preston, C. (1976). *Random Fields*. Lecture Notes in Mathematics, 534. Springer-Verlag, Berlin-Heidelberg.

Ruelle, D. (1971). "Existence of a phase transition in a continous classical system". *Physical Review Letters*, 27:1040–1041.

Stoyan, D.; Kendall, W. S. & Mecke, J. (1995). *Stochastic Geometry and its Applications*. Wiley, New York, second edition.

Sweeny, M. (1983). "Monte Carlo study of weighted percolation clusters relevant to the Potts models". *Physical Review* B, 27:4445–4455.

Swendsen, R. H. & Wang, J.-S. (1987). "Nonuniversal critical dynamics in Monte Carlo simulations". *Phys. Rev. Letters*, 58:86–88.

Tjelmeland, H. & Besag, J. E. (1996). "Markov random field models for objects against background". Technical Report Statistics no. 1/1996, Department of Mathematical Sciences, Norwegian University of Science and Technology.

Tjelmeland, H. & Holden, L. (1993). "Semi-Markov random fields". In Soares, A., editor, *Geostatistics Troia '92*, volume 1, pages 479–492. Kluwer Academic, Amsterdam.

Waagepetersen, R. P. (1996). "A note on a particular Markov connected component field model". In preparation.

Widow, B. & Rowlinson, J. S. (1970). “New model for the study of liquid-vapor phase transitions”. *J. Chem. Physics*, 52:1670–1684.

Younes, L. (1988). “Estimation and annealing for Gibbsian fields”. *Annales de l’institut Henri Poincaré*, 24:269–294.