

# On a nonparametric resampling scheme for Markov random fields

Lionel Truquet

*UMR 6625 CNRS Institut de Recherche Mathématique de Rennes (IRMAR) Université de  
Rennes I, Campus de Beaulieu, F-35042 Rennes Cedex &*

*CREST - ENSAI, Timbre J380, Campus de Ker Lann, rue Blaise Pascal, 35170 Bruz,  
France*

*e-mail: [lionel.truquet@univ-rennes1.fr](mailto:lionel.truquet@univ-rennes1.fr)*

**Abstract:** We study an extension to general Markov random fields of the resampling scheme given in Bickel and Levina (2006) [4] for texture synthesis with stationary Markov mesh models. The procedure generates bootstrap replicates of a sample using kernel regression and the principle of Gibbs sampling. Consistency of the bootstrap distribution is investigated under the Dobrushin contraction condition. Some simulation examples are given, in particular for the texture synthesis, for which the multiscale algorithm of Paget and Longstaff (1998) [27] is revisited.

**AMS 2000 subject classifications:** Primary 62M40; secondary 62G08, 62G09.

**Keywords and phrases:** Random fields, image analysis, nonparametric regression, resampling methods.

Received May 2010.

## 1. Introduction

The aim of this paper is to study an extension of a resampling scheme investigated by Bickel & Levina [4] for a particular class of stationary Markov random fields. This resampling method is in turn an adaption to random fields of the local bootstrap for Markov processes introduced by Paparoditis & Politis [28] for bootstrapping some time series. The basic principle of this bootstrap procedure can be formulated as follows: given a sample of a process, generate a replicate of this sample whose characteristics are similar to the original one. Generally, a stationary Markov model is defined by an invariant measure with respect to some probability kernels. Under some conditions, it is possible to simulate the model, using these kernels and a convenient recursive algorithm. Then, in order to reproduce the behaviour of the original time series or random field, a natural idea is to use nonparametric estimators of those kernels and to simulate a replicate of the sample, using the same algorithm. The term “bootstrap” comes from the fact that the nonparametric estimators of the conditional distributions used in [28] and [4] are based on a weighted empirical measure of the sample. This empirical measure is constructed using the idea of the kernel regression estimation. For example, in the case of a  $p$ -Markovian time series, the purpose is to estimate the regression function  $\mathbb{E}(\mathbb{1}_{X_t \leq x} | X_{t-1}, \dots, X_{t-p})$ .

In [28], this method is shown to be applicable to a large class of Markovian time series in order to estimate some statistics of the original process. A variant of this method was investigated by Monbet & Marteau [25] for multivariate processes.

Bickel & Levina [4] studied an extension of this method to a class of Markov random fields in the texture synthesis framework. The goal of texture synthesis can be stated as follows: given a texture sample, synthesize a new texture that appears similar to a human observer. Texture mapping or image compression are frequent applications for such algorithms. The stochastic nature of texture variations makes it a particularly natural area for applying statistical methods but usually the consistency of those algorithms are not investigated. In [4], the motivation of the authors was to give a theoretical justification to the popular algorithm of Efros and Leung [13] used for texture synthesis. But a crucial assumption made on the random field is that its simulation can be done recursively, starting from a seed and ordering the pixels as for the simulation of a time series. In particular, if  $(X_s)_{s \in \mathbb{Z}^2}$  denotes a random field, the rectangular lattice  $R$  (the sites of the pixels) of  $\mathbb{Z}^2$  is ordered setting  $R = \{s_1, \dots, s_k\}$  and it is assumed that for  $i \in \{1, \dots, k\}$ , the conditional distribution of  $X_{s_i}$  given the “past”  $X_{s_{i-1}}, \dots, X_{s_1}$  depends only on a limited number of variables. The consequence is that a texture can be simulated quickly with a single sweep. A typical example of such random fields were introduced by Pickard [29]. Although these models do not have a physical interpretation for a spatial phenomenon, they remain numerically efficient when the conditional kernels used for the simulation depend on more and more variables of the “past”. In fact, these models have been introduced in the context of image analysis in order to approximate the general Markov-Gibbs model which is more difficult to handle: a recursive computation of joint distributions is generally impossible for the general model and Markov chains methods such as the Gibbs sampler or the Metropolis algorithm must be used for the simulation. However, this restrictive class of random fields approximates generally very well the textures for the statistics distinguishable by a human observer when large data sets are available (see in particular the simulation results given in [4]).

In this paper we will not make the same assumption as in [4] and we will study an extension of the results given in [4] for a general Markov-Gibbs random field satisfying some regularity conditions. We will consider a classical stochastic relaxation algorithm: the Gibbs sampler. Gibbs sampling is a Markov chain method used to simulate a sample of a random field, using its local conditional distribution (LCD), that is for  $s \in \mathbb{Z}^2$  the distribution of  $X_s$  given  $(X_t)_{t \neq s}$  (which only depends on a finite number of variables if  $X$  is a Markov random field). Then a natural idea is to build a nonparametric estimator of the LCD, using the same idea as Bickel & Levina [4] and to simulate a replicate of the random field. However, in general, it is not possible to define a joint distribution associated to the nonparametric estimator of the LCD. This point makes the problem difficult to study because our algorithm will simulate a strange non-Markovian distribution whose shape depends on the estimator but also on the size of the desired replicate. Nevertheless we give some conditions under which

our method has consistency properties (i.e convergence to the distribution of the original random field).

We believe that studying such an extension is interesting for several reasons. First, our approach enlightens the good behaviour of another algorithm for texture synthesis introduced by Paget & Longstaff [27] and the resampling scheme studied is in the spirit of their work, using the nonparametric regression estimation introduced in [28] and extended in [4] to the random fields case. Secondly, although Gibbs sampling leads to long simulation times, this stochastic relaxation method applies to general Markov random fields, taking into account all the directions in the restitution of their local interactions. Motivated by other statistical applications, we believe that bootstrapping the “true” model instead of an approximation is more natural. Of course, the simulation of an image using a single sweep is numerically a great advantage and the speed of the algorithm is a major problem for the applications of texture synthesis. But except simulation time, the general statistical model should give better results, provided of course finding a suitable algorithm. However we do not claim that the approach studied in this paper is a satisfying answer, especially for the texture synthesis problem because the general Gibbs model is very difficult to handle, as shown by the impressive multiscale algorithm of Paget and Longstaff for texture synthesis (see Section 5), necessary to speed up the algorithm and to avoid problems related to local minima of the energy. Moreover the texture synthesis problem has been extensively studied over the last decade and several complex algorithms give now better results for this particular problem (see in particular [22, 12] or [21]). But to our knowledge, a consistency result for bootstrapping the distribution of a general Markov field has not been investigated yet.

Finally, it could be interesting to consider other statistical applications. As pointed in [4], this kind of resampling scheme could also be used to approximate the distribution of some linear statistics, although its principle differs from some classical methods (using blocks) studied for stationary time series (see Künsch [20]) or random fields (see Politis & Romano [30]). This approach is used in [28] for time series. Moreover in [25], the simulation of missing data is investigated for time series. The so-called local grid bootstrap of [25] was successfully applied to wind data (see [2]) and it seems of interest to extend this approach to spatio-temporal phenomena. In the spatial context, one could use the observed sites for the estimation of the LCD and in a second step the Gibbs sampling for the simulation of the missing data. In this kind of application, ordering the different sites for the simulation seems not very natural. One can also argue that all the statistics are not distinguishable by a human observer and that the (visual) good results obtained in [4] for the texture synthesis should be extended to more natural random fields.

The paper is organized as follows. In Section 2, we recall the results of Bickel and Levina about their nonparametric resampling for a particular class of Markov random fields. In Section 3, we extend the resampling principle of Section 2 to general Markov random fields and some consistency results are proved for our extension. The example of a nearest-neighbor Markov random field with a pairwise potential is detailed in Section 4. Section 5 is devoted to

the texture synthesis problem: we recall the multiscale algorithm used by Paget and Longstaff, which has an independent interest and we give several simulation examples. Proofs of our investigations are postponed to the last section of the paper.

## 2. The Markov Mesh Models algorithm

### 2.1. Principle

We first recall the Markov Mesh Models (MMM in the sequel) algorithm introduced by Bickel and Levina [4] in the context of the texture synthesis problem. This algorithm is different from the original algorithm of Efros and Leung [13] by the order in which pixels are filled in the synthesized texture (raster instead of spiral), and the weights with which the pixels are resampled.

In the sequel we consider  $\{X_t, t \in \mathbb{N}^* \times \mathbb{N}^*\}$  a real-valued random field and a positive integer  $o \in \mathbb{N}^*$ . We will use the following notations:

- for  $A \subset \mathbb{N}^* \times \mathbb{N}^*$ ,  $X_A$  denote the family  $(X_t)_{t \in A}$ ;
- for  $A, B \subset \mathbb{N}^* \times \mathbb{N}^*$ ,  $A + B = \{t_A + t_B, (t_A, t_B) \in A \times B\}$  and  $A - B = \{t_A - t_B, (t_A, t_B) \in A \times B\}$ .

For  $t = (t_1, t_2) \in \mathbb{N}^* \times \mathbb{N}^*$  and  $s \in \mathbb{N}^* \times \mathbb{N}^*$ , define the index sets

- $U_t^{(o)} = \{u = (u_1, u_2) \in \mathbb{N}^* \times \mathbb{N}^*; \max(1, t_1 - o) \leq u_1 \leq t_1, \max(1, t_2 - o) \leq u_2 \leq t_2 \text{ and } u \neq t\}$ ;
- $U_t^{(o)}(s) = U_t^{(o)} - \{t\} + \{s\}$ ;
- $W_t = \{1, \dots, t_1\} \times \{1, \dots, t_2\} \setminus \{t\}$ .

The set  $U_t^{(o)}$  is always included in the square of size  $(o + 1) \times (o + 1)$  with  $t$  as the bottom right corner,  $t$  itself excluded, but there are  $(o + 1)^2 - 1$  possible shapes of  $U_t^{(o)}$ . Then,

**Definition 1.** A random field  $X = \{X_t, t \in \mathbb{N}^* \times \mathbb{N}^*\}$  is a Markov mesh model if there exists  $o \in \mathbb{N}^*$  such that for all  $t \in \mathbb{N}^* \times \mathbb{N}^*$ ,

$$\mathbb{P}(X_t | X_{W_t}) = \mathbb{P}(X_t | X_{U_t^{(o)}}). \quad (2.1)$$

Now, the MMM resampling algorithm of Bickel and Levina [4] can be presented. First assume that a MMM  $X$  is observed on rectangle  $J_T = \{1, \dots, T_1\} \times \{1, \dots, T_2\}$  with  $T_1, T_2 \geq o$ . Then consider a family of kernels  $(W^{(\ell)})_{\ell \in \mathbb{N}^*}$  that are Borelian functions  $W^{(\ell)} : \mathbb{R}^\ell \rightarrow [0, \infty)$  satisfying some general smoothness assumptions (see Assumption **(A4)** in the Appendix). Moreover, for a resampling width  $b > 0$  and all  $\ell \in \mathbb{N}^*$ , define

$$W_b^{(\ell)}(y) = b^{-\ell} W^{(\ell)}(y/b) \quad y \in \mathbb{R}^\ell.$$

In the sequel, for simplicity, we will omit the exponent  $o$  and  $\ell$  for respectively  $U_t^{(o)}$ ,  $U_t^{(o)}(s)$  and  $W_b^{(\ell)}$ .

**The MMM resampling algorithm**

The aim of the algorithm is to generate a texture  $X^* = \{X_t^*, t \in R_T\}$  from the observed field  $\{X_t, t \in J_T\}$ , where  $R_T = \{1, \dots, u_T\} \times \{1, \dots, v_T\}$ . There are 3 main steps in this algorithm.

1. Select a starting value for  $\{X_t^* : 1 \leq t_1 \leq o + 1, 1 \leq t_2 \leq o + 1\}$ , the top left  $(o + 1) \times (o + 1)$  square. Typically the starting value will be a  $(o + 1) \times (o + 1)$  square random chosen from the observed field  $(X_t)_{t \in I_T}$ .
2. Suppose that there exists  $(u, v) \in R_T$  such that  $X_t^*$  has been generated for  $t \in \{1, \dots, u - 1\} \times \{1, \dots, v_T\} \cup \{u\} \times \{1, \dots, v - 1\}$ , that is,  $u - 1$  rows are filled in completely, and the row  $u$  is filled up for the column  $v$ . To generate the next value  $X_t^* = X_{(u,v)}^*$ , let  $N_t$  be a discrete random variable with probability distribution

$$\mathbb{P}(N_t = s) = \frac{1}{Z_T} W_b(X_{U_t}^* - X_{U_t(s)})$$

for all  $s \in \mathbb{N}^* \times \mathbb{N}^*$  such that  $U_t(s) \subset J_T$  and where  $Z_T = \sum_s W_b(X_{U_t}^* - X_{U_t(s)})$  is a normalizing constant. Note that the set of all possible  $s$  is such that the conditioning neighborhood of  $s$  fits within the observed texture field.

3. Generate  $N_t$  and set  $X_t^* = X_{(u,v)}^* = X_{N_t}$ .

In Figure 1, we show two steps in the progress of the MMM algorithm, just after the choice of the seed (step 1 above) and when the neighborhood of the pixel is full (see the center of the picture in Figure 1). In Figure 2, we illustrate the difference with the original algorithm of Efros and Leung. Here, the seed is located at the center of the new texture and the synthesis uses a spiral ordering. At each step, a pixel  $X_t^*$  is chosen conditionally to some pixels previously simulated and located in a square window centered at  $t$ . Another difference with the MMM algorithm is the choice of uniform weights for the synthesis (see [4] for details).

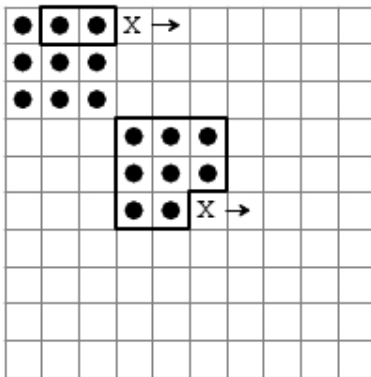


FIG 1. MMM algorithm,  $o = 2$

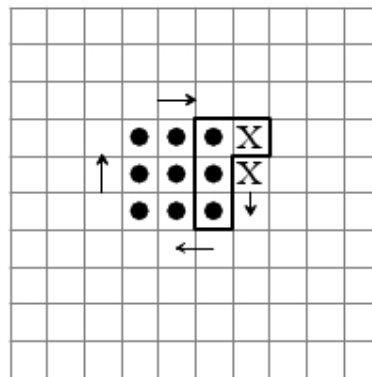


FIG 2. Efros and Leung algorithm,  $o = 2$

The MMM algorithm is formulated for a particular class of random fields, the Markov Mesh Models (also known as Pickard random fields [29]) which were introduced by Abend, Harley and Kanal [1]. These models have been developed for image applications and can be efficiently simulated. The resampling scheme described above is an adaption to random fields of a method proposed for bootstrapping Markovian time series ([31, 28]).

### 2.2. Consistency results

To state the consistency of the above algorithm, we need the following notations:

- for  $A \subset \mathbb{Z}^2$ ,  $|A|$  is the cardinal of  $A$ .
- for  $x = (x_1, x_2) \in \mathbb{Z}^2$ ,  $\|x\|_\infty = \max(|x_1|, |x_2|)$ .
- for  $A, B \subset \mathbb{Z}^2$ ,  $d(A, B) = \inf_{x \in A, y \in B} \{\|x - y\|_\infty\}$ .
- for  $y \in \mathbb{R}^\ell$  with  $\ell \in \mathbb{N}^*$ ,  $\|y\|$  is the usual Euclidean norm.
- for  $T = (T_1, T_2) \in \mathbb{N}^* \times \mathbb{N}^*$ , let  $[T] = T_1 T_2$  and  $T \rightarrow \infty$  means  $T_1 \wedge T_2 \rightarrow \infty$ .

Let  $A \subset \mathbb{Z}^2 \setminus \{0\}$  such that  $|A| < \infty$ . For  $t \in \mathbb{Z}^2$ , define

$$Y_t = (X_{t+j})_{j \in A}.$$

Moreover we define the following subsets of  $\mathbb{N}^* \times \mathbb{N}^*$ :

$$I_T = \left\{ t \in J_T, \{t\} + A \subset J_T \right\}.$$

To show the consistency of their algorithm, Bickel and Levina prove a general lemma about the estimation of the local conditional distribution function

$$F_{X|Y}(x|y) = \mathbb{P}(X_t \leq x | Y_t = y)$$

(see Theorem 2 in [4]). More precisely they establish the convergence of the following sample cumulative conditional distribution function, defined by

$$F_T(x|y) = \frac{1}{Z_T} \sum_{s \in I_T} \mathbb{1}_{X_s \leq x} W_{b_T}(y - Y_s), \tag{2.2}$$

for  $(x, y) \in S \times S^A$  and  $T \in \mathbb{N}^* \times \mathbb{N}^*$  such that  $I_T \neq \emptyset$ . Here  $Z_T = \sum_{s \in I_T} W_{b_T}(y - Y_s)$ . Note that  $F_T$  is a classical nonparametric regression estimator of the conditional expectation  $\mathbb{E}(\mathbb{1}_{X_t \leq x} | Y_t = y)$ . We recall the following theorem whose assumptions are reported to the Appendix.

**Theorem 1** (Theorem 2 [4]). *If  $X$  is a MMM satisfying assumptions (A1)–(A4) (see section 6.1), then for all  $A \in \mathbb{Z}^2$  such that  $|A| < \infty$ ,*

$$\sup_{(x,y) \in S \times S^A} |F_T(x|y) - F_{X|Y}(x|y)| \xrightarrow{T \rightarrow \infty} 0.$$

**Note.** Theorem 1 will also be used for our extension. Moreover an almost sure convergence rate for this approximation will be given (see Theorem 4).

Using Theorem 1, Bickel & Levina show the consistency of their MMM algorithm and also of the original spiral resampling algorithm of Efros and Leung, using appropriated conditional independence properties (we refer to [4] for details). This resampling scheme uses the kernel regression estimation and requires some regularity assumptions (see Assumptions (A1-4) in section 6.1). As it is pointed in [4], these assumptions are perfectly plausible for most real textures: the mixing property is natural for stochastic textures, the compactness assumption is always satisfied since the number of gray levels is finite, and this number is sufficiently high in most of real textures to make the smoothness assumptions plausible.

However, the random fields used for the two algorithms described above must satisfy suitable conditional independence properties. For example, when  $o = 1$ , the rows of the MMM form a Markov chain. This is also the case for the Efros & Leung algorithm for which the last simulated row on a square lattice depends (conditionally to the previous rows) only on the last row: then it is easily seen that the rows of a such stationary process form a Markov chain. Moreover if  $o > 1$ , the rows on a square lattice evolve as a multivariate  $o$ -markovian time series.

For a general Markov-Gibbs model, such properties do not hold in general and it is not possible to simulate a sample using a single sweep. For example, consider the simple example of the four nearest neighbor AR model with Gaussian white noise

$$X_t = \alpha \sum_{\|j\|_\infty=1} X_{t-j} + \xi_t, \quad t \in \mathbb{Z}^2,$$

with  $0 < |\alpha| < \frac{1}{4}$ . Then it is shown in Guyon [18] (see p. 13, Theorem 1.3.2 and its proof) that this Gaussian Markov random field does not admit a representation of the form

$$X_t = \sum_{j \in K} a_j X_{t-j} + e_t, \quad t \in \mathbb{Z}^2,$$

where  $e_t$  is independent of  $(X_{t-j})_{j \in J}$ ,  $J$  is a half space of  $\mathbb{Z}^2$  and  $K$  is a finite subset of  $J$ . However, such an “unilateral” representation is necessary in order to apply the Markov property at each step of the MMM algorithm or the Efros & Leung algorithm. For the simple class of linear gaussian random fields, the unilateral type representation writes generally as an infinite linear expansion (see [18], Example 1.3.3). Then for a general Markov random field, it would be necessary to take in account of all the pixels previously simulated in order to apply the MMM or the spiral algorithm. Of course, applications of some Markov properties using a large neighborhood size  $o$ , provide an approximation of the model (see [3]). A class of Markov random fields satisfying these unilateral type Markov assumptions were introduced in [1] and [29]. These models have been developed for image applications and they remain numerically efficient tools

which approximate very well the general model, as shown by the simulations given in [4]. The stationary models used in [4] satisfy the usual Markov property even for the sites located on the boundary of a square lattice. Such processes were characterized in Champagnat & al. [5] and as pointed by the authors, the class of such models is restrictive compared to general MRF's.

Having in mind other statistical applications than the texture synthesis problem, we believe that a resampling scheme that can apply to a general Markov random field is of interest. For example, the simulation of missing data in meteorology, using an ordering for the different sites, seems not very appropriated. This is why we investigate in the next section, an extension of the algorithms described above for the general model using Gibbs sampling.

### 3. Extension of the resampling algorithm for a general Markov-Gibbs model

The goal of this section is to study convergence properties of a resampling scheme that can apply to more natural random fields than MMM. In particular, the model is not supposed to satisfy assumption (2.1). We will use the non-parametric estimators (2.2) of the local conditional distributions of the random field. Basically, our extension of the algorithm given in [4] can be formulated as follows:

1. Choose some values for the pixels  $\{X_j^*/j \in R_T\}$  using the sample  $\{X_j/j \in J_T\}$ .
2. Choose randomly a site  $j$  of  $R_T$ , extract the neighborhood  $X_{j+\mathcal{N}_o}^*$  and choose a new pixel  $X_s$  with a probability given by the conditional distribution  $F_T$  (constructed with  $A = \mathcal{N}_o$ ).
3. Repeat point 2.

Here  $\mathcal{N}_o$  denotes a full square centered at 0. The algorithm just formulated above has similarities with the Gibbs sampler with a random visiting scheme used for the simulation of a Markov random field. However, as it will be explained in this section, this algorithm is not exactly a Gibbs sampler but just a Markov chain approach for the simulation of a bootstrap sample (theoretically, the step 2 must be repeated infinitely often). Note that at each step we use a conditional distribution with respect to a full neighborhood of a given site. Contrarily to the MMM algorithm, there is no initialization with a seed.

#### 3.1. An approximated Gibbs sampler for Markov random fields

In this paper we consider a Markov-Gibbs random field  $X$  over  $\mathbb{Z}^2$  with a continuous compact state space  $S \subset \mathbb{R}$  and we denote by  $\mu$  its distribution. More precisely, denoting by  $\lambda_S$  the Lebesgue measure on  $S$  and

$$\mathcal{N}_o = \{j \in \mathbb{Z}^2 : 0 < \|j\|_\infty \leq o\},$$



we assume that the probability measure  $\mu$  on  $S^{\mathbb{Z}^2}$  satisfies the following definition.

**Definition 2.** A probability measure  $\mu$  on  $S^{\mathbb{Z}^2}$  is said to be the distribution of a stationary Markov random field  $X$  with neighborhood system  $\mathcal{N}_o$  if there exists a family of applications (called a Gibbs potential)  $\Phi = \{\phi_A : S^A \rightarrow \mathbb{R}, A \in \mathcal{S}\}$  indexed by  $\mathcal{S}$ , the set of all finite subsets of  $\mathbb{Z}^2$  such that

$$\begin{aligned} \phi_A(x_A) &= \phi_{A+s}(x_A), \quad \forall (A, s) \in \mathcal{S} \times \mathbb{Z}^2, \\ \phi_A &\equiv 0, \text{ if } \mathcal{N}_o \subset A, \end{aligned}$$

and for  $V \in \mathcal{S}$ , the conditional distributions of  $\mu$  are given by:

$$\begin{aligned} \mu(dx_V|x_{V^c}) &= \mathbb{P}(X_V \in dx_V | X_{V^c} = x_{V^c}) \tag{3.1} \\ &= \frac{1}{Z_V(x_{V^c})} \exp\left(\sum_{A \in \mathcal{S}: A \cap V \neq \emptyset} \phi_A(x_A)\right) \lambda_S^{(V)}(dx_V), \tag{3.2} \end{aligned}$$

where  $Z_V(x_{V^c})$  is a normalization constant,  $\lambda_S^{(V)}$  denotes the product measure on  $S^V$  and  $V^c = \mathbb{Z}^2 \setminus V$  the complementary of  $V$ .

We refer to Guyon [18] for an introduction to Markov random fields. Using this formulation, it is easily seen that the following Markov properties are satisfied:

$$\mathbb{P}(X_t | X_{\mathbb{Z}^2 \setminus \{t\}}) = \mathbb{P}(X_t | X_{t+\mathcal{N}_o}), \quad t \in \mathbb{Z}^2. \tag{3.3}$$

The set  $t+\mathcal{N}_o$  is called the set of neighbors of the site  $t$ . Moreover  $\mathbb{P}(X_t | X_{\mathbb{Z}^2 \setminus \{t\}})$  is referred to the local conditional distribution (LCD). It is possible to show that any stationary random field satisfying (3.3) with a positive conditional density over  $S$  can be defined using a Gibbs potential  $\Phi$  as in Definition 2. This characterization is given by the Hammersley-Clifford theorem (see [18]) and has been studied by several authors (see for example [3]). In general, a Markov random field satisfies the condition

$$\mathbb{P}(X_t | X_C) = \mathbb{P}(X_t | X_{(t+\mathcal{N}_o) \cap C}), \quad t \in \mathbb{Z}^2,$$

only if  $t + \mathcal{N}_o \subset C$ . In contrast, the more restrictive class of MMM assumes additional conditional independence properties (see property (2.1)).

A difficulty which arises with Gibbs random fields is the phase transition problem. For a given potential  $\Phi$ , several random fields distributions can satisfy (3.1) and natural candidates are the cluster points of the probability kernels given by the right hand side of (3.1), when  $V \nearrow \mathbb{Z}^2$ . In the sequel, we denote by  $\mathcal{G}(\Phi)$  the set of probability measures whose conditional distributions are given by (3.1). We refer to Georgii [17] for an introduction to Gibbs random fields and the phase transition phenomenon.

In order to extend the consistency proof of Bickel and Levina to general Markov random fields, we use the one point conditional distributions (2.2) with  $A = \mathcal{N}_o$ . However, in this case, the one point distribution (2.2) cannot be in general the LCD of a random field.

**Proposition 1.** *Suppose that  $|I_T| \geq 2$ . Then when  $W$  is the Gaussian kernel, the one point conditional law  $F_T$  cannot be the LCD of a random field.*

In particular, no Gibbs potential can be associated to  $F_T$  when  $W$  is the Gaussian kernel. It is well known that for a given one point conditional distribution, several constraints are required to be the conditional distribution of a random field (see [18] p. 98). Those constraints are satisfied if the form of the one point conditional distribution is given by a Gibbs potential, as in (3.1). Thus, we could avoid this compatibility issue using a nonparametric estimator of the Gibbs potential. Unfortunately, this kind of estimation is not very well investigated in its full generality and its study is beyond the scope of this paper.

Nevertheless, we will use the conditional distributions  $F_T$  and the principle of the Gibbs sampler for the simulation of an approximated realization of the random field  $X$ . We refer to [16] for the use of the Gibbs sampler (see also [18] p211). Suppose that  $R_T = \{-u_T, \dots, u_T\} \times \{-v_T, \dots, v_T\}$  is a rectangle of  $\mathbb{Z}^2$  and define  $\partial R_T = (R_T + \mathcal{N}_o) \setminus R_T$  the boundary of  $R_T$ . Then, given a sequence  $z \in S^{\mathbb{Z}^2}$ , the Gibbs sampler gives an approximated simulation of a realization of the distribution  $\mu_{R_T}(\cdot|z) = \mathbb{P}(X_{R_T}|X_{\partial R_T} = z_{\partial R_T})$ . The principle is to simulate a  $S^{R_T}$ -valued Markov chain with invariant distribution  $\mu_{R_T}(\cdot|z)$ . When  $\mathcal{G}(\Phi) = \{\mu\}$ , the compactness assumption ensures that

$$\lim_{T \rightarrow \infty} \mu_{R_T}(\cdot|z) \times \delta_{z_{R_T^c}} = \mu$$

weakly, whatever the value of the boundary conditions  $z_{\partial R_T}$ . The transition  $P_{R_T}^{(z)}$  of this Markov chain is constructed using a visiting scheme for the sites of  $R_T$ : at each time a site is chosen, we replace the value at this site by a realization of the conditional distribution  $F_{X|Y}$ . A requirement is that each site has to be visited infinitely often. Here are two well known examples of transitions for the Gibbs sampler:

- The periodic visiting scheme. Here we denote by  $\prec$  the lexicographic order relation on  $R_T$ . If  $x, \tilde{x} \in S^{R_T}$  and  $s \in R_T$ , we define the vectors  $\tilde{x}x(s) \in S^{\mathcal{N}_o}$  such that  $\tilde{x}x(s)_j = \tilde{x}_{s+j}$  if  $s+j \prec s$  and  $\tilde{x}x(s) = x_{s+j}$  otherwise, completed with the boundary conditions  $x_{s+j} = z_{s+j}$  or  $\tilde{x}_{s+j} = z_{s+j}$  if the site  $(s, j) \in R_T \times \mathcal{N}_o$  is such that  $s+j \notin R_T$ . Then we define the following transition on  $S^{R_T}$ :

$$P_{R_T}^{(z)}(x, d\tilde{x}) = \otimes_{s \in R_T} F_{X|Y}(d\tilde{x}_s | \tilde{x}x(s)). \tag{3.4}$$

- The random visiting scheme with uniform law uses the transition:

$$P_{R_T}^{(z)}(x, d\tilde{x}) = \frac{1}{|R_T|} \sum_{s \in R_T} \prod_{u \in R_T, u \neq s} \delta_{x_u}(d\tilde{x}_u) \times F_{X|Y}(d\tilde{x}_s | x_{s+\mathcal{N}_o}). \tag{3.5}$$

From now on, we fix an arbitrary element  $z \in S^{\mathbb{Z}^2}$  for the boundary conditions of the Gibbs sampler. Then we just write  $P_{R_T}$  instead of  $P_{R_T}^{(z)}$  and the conditional distribution  $\mu_{R_T}(\cdot|z)$  is simply denoted by  $\mu_{R_T}$ .

Suppose that we use the conditional distributions  $F_T$  and the Gibbs sampler to synthesize a new texture on the rectangle  $R_T$ . We suppose here that assumptions of Theorem 1 hold. Though the conditional distributions  $F_T$  are not compatible with a Markov random field, one can use those distributions to simulate a Markov chain. The transition of this Markov chain can be constructed by replacing  $F_{X|Y}$  by  $F_T$  in (3.4) or (3.5) and we define

$$P_{T,R_T}^*(x, d\tilde{x}) = \otimes_{s \in R_T} F_T(d\tilde{x}_s | \tilde{x}(s)), \quad (3.6)$$

or

$$P_{T,R_T}^*(x, d\tilde{x}) = \frac{1}{|R_T|} \sum_{s \in R_T} \prod_{u \in R_T, u \neq s} \delta_{x_u}(d\tilde{x}_u) \times F_T(d\tilde{x}_s | x_{s+\mathcal{N}_o}). \quad (3.7)$$

Then if the marginal distribution of the Markov chain with transition  $P_{T,R_T}^*$  converges a.s to an invariant probability  $\mu_{T,R_T}^*$  on  $I_T^{R_T}$  satisfying

$$\mu_{T,R_T}^* = P_{T,R_T}^* \mu_{T,R_T}^*, \quad (3.8)$$

we hope that  $\mu_{T,R_T}^*$  is not too far from the unique invariant distribution  $\mu_{R_T}$  associated to the transition  $P_{R_T}$  (the unicity holds if the transition  $P_{R_T}$  is positive, see [24]).

**Note.** Compared to the method studied in [4], our extension uses the same nonparametric estimation but requires several sweeps on  $R_T$  for the simulation of one bootstrap sample and then a high computational cost. Numerically, a simulation procedure using a single sweep is of course a great advantage, especially for the texture synthesis problem. However, without additional assumptions of conditional independence, it is not possible to simulate a sample using a single sweep.

### 3.2. Asymptotic when $R_T = R$

Here we consider the case  $R_T = R$  for all  $T$  (i.e we consider a fixed rectangle  $R$  for the simulation and an increasing size for the original sample). We are going to study the convergence of the sequence  $(\mu_{T,R}^*)_T$ , defined in (3.8), to the distribution  $\mu_R$ . Usually, the size of the bootstrap sample must increase to infinity with the size of the original sample. This is the case for statistical applications (see [28] for applications to the distribution of some linear statistics) but also for the texture synthesis problem, where the bootstrap sample is generally larger than the original one. Then the asymptotic investigated here is not the most natural, but a consistency result can be stated using the assumptions of Theorem 1 only. We also point out that the family of distributions  $(\mu_{T,R}^*)_R$  is not ensured to be coherent, in the sense that

$$\mu_{T,R}^*(dx_U | z_{U^c}) \neq \mu_{T,U}^*(dx_U),$$

for a square  $U$  such that  $U \cup (U + \mathcal{N}_o) \subset R$ . For example if  $U = \{0\}$ , one have  $\mu_{T,U}^*(dx_U) = F_T(dx_0|z_{\mathcal{N}_o})$  which cannot be the conditional distribution of a measure defined on a product (see Proposition 1 and its proof). This problem suggests that the shape of the bootstrap distributions depends not only on the size of the sample  $T$ , but also on the size of  $R$ . It also emphasizes that a classical nonparametric estimation of the LCD is not enough to estimate the full Markov model: this point is an important difference with respect to the markovian time series studied in [28].

The following theorem states a result of continuity of the invariant measure  $\mu_{T,R}^*$  with respect to the transition  $P_{T,R}^*$ . Since the state space of the process  $X$  is compact, Theorem 1 and a tightness argument lead to the following result

**Theorem 2.** *For  $T \in \mathbb{N}^2$  such that  $I_T \neq \emptyset$ , let  $P_{T,R}^*$  be the transition defined in (3.4) or (3.5) replacing  $F_{X|Y}$  by  $F_T$ . Assume that the assumptions of Theorem 1 hold. Then  $P_{T,R}^*$  has a unique invariant distribution  $\mu_{T,R}^*$  and for all initial value  $x^{(0)} \in I_T^R$ , the approximated Gibbs sampler converges to  $\mu_{T,R}^*$  in distribution. Moreover we have almost surely:*

$$\lim_{T \rightarrow \infty} \mu_{T,R}^* = \mu_R = \mathbb{P}(X_R | X_{\partial R} = z_{\partial R}), \quad \text{in distribution.}$$

## Notes

1. It is easily seen that the regularity assumption **(A3)** of Theorem 1 is satisfied when the potential  $\Phi$  is a family of continuously differentiable applications.
2. Theorem 2 shows the consistency of the bootstrap distribution only for the simulation on a fixed rectangle  $R$ . Of course, it is more interesting to investigate the case of increasing bootstrap samples  $X_{R_T}^*$  with  $R_T \nearrow \mathbb{Z}^2$ . The difficulty for such an asymptotic is that no information seems available for the bootstrap distribution  $\mu_{T,R_T}^*$  except its invariance with respect to the Markov kernel  $P_{T,R_T}^*$ . In particular, the conditional distributions of  $\mu_{T,R_T}^*$  are unknown and it is difficult to derive properties of some marginal distributions. However, Theorem 2 suggests that such an asymptotic should be possible if the growth of the sequence  $(R_T)_T$  is moderated and  $\mathcal{G}(\Phi) = \{\mu\}$ . In the next section, we show that the sequence  $(R_T)_T$  can be chosen arbitrarily under a strong assumption on the LCD.

### 3.3. Asymptotic when $R_T \nearrow \mathbb{Z}^2$

The case  $R_T \nearrow \mathbb{Z}^2$  is the most natural for the bootstrap and its applications (e.g texture synthesis). We only consider the case of the Gibbs sampler with a random visiting scheme (3.5). Our aim here is to study the convergence of the sequence  $(\mu_{T,R_T}^*)_T$  to  $\mu$ , the distribution of the random field  $X$ . The difficulty is that the transition  $P_{T,R_T}^*(x, \cdot)$  defined in (3.5) converges weakly to the Dirac measure  $\delta_x$  and this limit is not informative to study the cluster points of the sequence  $(\mu_{T,R_T}^*)_T$ . On the other hand, the single site conditional distributions

associated to  $\mu_{T,R_T}^*$  are not  $F_T$  in general and are just defined through the implicit equation  $\mu_{T,R_T}^* = F_{T,R_T}^* \mu_{T,R_T}^*$ ; a tractable expression of the bootstrap distributions seems not available. Moreover, the sequence  $(\mu_{R_T})_T$  can have several cluster points if the random field exhibits a phase transition, i.e  $|\mathcal{G}(\Phi)| > 1$ . A sufficient condition ensuring the unicity of the Gibbs measure is the Dobrushin contraction condition (see [8]). A survey of the Dobrushin contraction technique and its application is given in [15]. As we will see, this kind of condition is also sufficient to give a positive answer for the convergence of the sequence  $(\mu_{T,R_T}^*)_T$ . Let us first recall the definition of the contraction coefficients in our setting. If  $|\cdot|$  denotes the usual distance on  $S$ , we recall that the Wasserstein metric of order 1 is defined for two distributions  $G$  and  $H$  on  $S$  by

$$d_W(G, H) = \sup \frac{|\int f dG - \int f dH|}{\delta(f)},$$

where the supremum is taken over all Lipschitz functions  $f$  on  $S$  with:

$$\delta(f) = \sup_{a \neq b} \frac{|f(a) - f(b)|}{|a - b|} < \infty.$$

Another expression is given by

$$d_W(G, H) = \int_S |G(t) - H(t)| dt.$$

Define for  $j \in \mathcal{N}_o$ ,

$$L_j = \sup \left\{ \frac{d_W(F_{X|Y}(\cdot|x_{\mathcal{N}_o}), F_{X|Y}(\cdot|\tilde{x}_{\mathcal{N}_o}))}{|x_j - \tilde{x}_j|} / x = \tilde{x} \text{ off } j \right\}, \tag{3.9}$$

Then it can be shown (see [15]) that under the condition

$$\alpha = \sum_{j \in \mathcal{N}_o} L_j < 1, \tag{3.10}$$

we have  $\mathcal{G}(\Phi) = \{\mu\}$ .

Also for  $R \subset \mathbb{Z}^2$ , denote by  $\mathcal{L}_R$  the class of Lipschitzian functions  $f$  on  $S^R$  satisfying

$$|f(u) - f(v)| \leq \sum_{i \in R} |u_i - v_i| \delta_i(f), \quad \sum_{i \in R} \delta_i(f) < \infty$$

where

$$\delta_i(f) = \sup \left\{ \frac{|f(u) - f(v)|}{|u_i - v_i|} / u = v \text{ off } i \right\}.$$

Finally, we denote by  $d_R$  the distance between probability measures on  $S^R$  defined by

$$d_R(\alpha, \beta) = \sup \left\{ \left| \int f d\alpha - \int f d\beta \right| : f \in \mathcal{L}_R, \quad L(f) = \sum_{j \in R} \delta_j(f) = 1 \right\}.$$

Then we have the following result.

**Theorem 3.** *Assume (3.10). Then we have almost surely*

$$d_R(\mu_{T,R}^*, \mu_R) \leq \frac{1}{1-\alpha} \sup_{y \in S^{\mathcal{N}_o}} \int_S |F_T(x|y) - F_{X|Y}(x|y)| dx.$$

Moreover, for an arbitrary sequence of rectangles  $R_T \nearrow \mathbb{Z}^2$ , we have almost surely

$$\lim_{T \rightarrow \infty} \mu_{T,R_T}^* \times \delta_{z_{R_T^c}} = \mu \text{ weakly.}$$

**Notes**

1. Condition (3.10) is a restrictive assumption because it assumes that the random field  $X$  has very weak interactions. Under this condition and using Theorem 4, a rate of convergence for the invariant measure of the approximated Gibbs sampler is available.

We did not prove the convergence of  $\mu_{T,R_T}^*$  to the distribution  $\mu$  of the sample assuming only that there is no phase transition. As mentioned before, the main difficulty is due to the lack of available local properties of the measure  $\mu_{T,R_T}^*$ , e.g local conditional distributions. This problem could be avoided using a nonparametric estimation of the Gibbs potential (although the corresponding simulation scheme will probably be no more linked to a resampling method). Indeed, in this case, an estimation of the local conditional distribution would be compatible with the existence of joint distributions, which is a failure in our context as shown in Proposition 1. Nonparametric estimators of the Gibbs potential seem to have been investigated only for particular cases. One can mention the work of Dachian [6] who considers  $\{0, 1\}$ -valued random fields or [7] for the nonparametric estimation of a pairwise-interaction Gibbs point process. On the other hand, an estimation of each function  $\Phi_C$  where  $C$  is a clique of  $\mathcal{N}_o$  and  $\Phi_C$  is defined in (3.1) could lead to a much more complicated procedure for the estimation of the local conditional distribution than the method studied here which is a simple extension of that of Bickel and Levina, using nonparametric regression estimation.

2. We mention that the more classical Dobrushin condition for unicity (i.e  $\mathcal{G}(\Phi) = \{\mu\}$ ) is formulated using the total variation distance. In this case the condition (3.10) becomes

$$\frac{1}{4} \sum_{j \in \mathcal{N}_o} \sup \left\{ \int |f_{X|Y}(x|y) - f_{X|Y}(x|y')| dx, \quad y = y' \text{ off } j \right\} < 1, \quad (3.11)$$

But contrarily to the Wasserstein metric, the total variation distance is not compatible with the convergence stated in Theorem 1 for the LCD.

3. In this paper, we study the Gibbs sampler with random scanning using the uniform distribution. We mention that one can also use different distributions  $\{c_{j,T} : j \in R_T\}$  for the visiting scheme, for which  $\lim_{T \rightarrow \infty} c_{T,j} = c_j > 0$  with  $\sum_{j \in \mathbb{Z}^2} c_j = 1$ . In this case, the cluster points of the sequence

$(\mu_{T,R_T}^*)_T$  can be shown to be invariant measures for an infinite volume Gibbs sampling. Additional developments are required in order to study the unicity of these cluster points; this problem will be investigated in a subsequent paper.

### 3.4. A convergence rate in Theorem 1

In this section, we establish an almost sure convergence rate for the estimators (2.2), which is a result of independent interest. Here we consider a random field  $X$  and a finite subset  $A$  of  $\mathbb{Z}^2 \setminus \{0\}$ . A convergence rate for the Theorem 2 of [4] can be established. In fact, as mentioned previously, the proof of this theorem does not use the MMM property (2.1) and applies to more general random fields. Following Bickel and Levina’s proof, we obtain the following result.

**Theorem 4.** *If  $X$  is a random field satisfying assumptions (A1-4), then for all  $A \in \mathbb{Z}^2 \setminus \{0\}$  such that  $v = |A| < \infty$ ,*

$$\sup_{(x,y) \in S \times S^A} |F_T(x|y) - F_{X|Y}(x|y)| = O([T]^{-\gamma}) \quad a.s$$

where  $0 < \gamma < \frac{\tau-2}{2(v+1)(\tau+v+2)}$  and  $b = b_T = O([T]^{-\delta})$  with  $\delta = \frac{\tau-2}{2(v+1)(\tau+v+2)}$ .

#### Notes

- The convergence rate in Theorem 4 is depending on a power law of  $[T]$  and the maximal exponent of convergence rate that we can obtain is  $\frac{1}{2(1+v)}$  (when  $\tau \rightarrow \infty$  for the mixing assumption). Note that if  $o = 0$  (corresponding to an independent random field) then  $v = 0$  and the convergence rate is arbitrary close to  $[T]^{1/2}$ .
- From this result, one can also obtain a convergence rate for joint distributions (see in particular Theorem 6.1). Such a rate could be interesting to obtain additional properties of the bootstrap distribution.

### 3.5. The choice of parameters

From the above results, we have a theoretical method to resample a Markov random field. It is enough to use the sample  $\{(X_s, Y_s)/s \in I_T\}$  and to run an approximated Gibbs sampler with the conditional distribution  $F_T$  evaluated with the help of a kernel  $W$ . But from a practical point of view, the choice of two parameters need to be discussed: the neighborhood size  $o$  in the Markov property and the value of the bandwidth  $b_T$ . In this paper, we focus on the bandwidth parameter  $b_T$  assuming that the neighborhood size  $o$  is known.

The quality of the bootstrap approximations depends heavily on how close the function  $F_T(\cdot|y)$  is to the LCD  $F_{X|Y}(\cdot|y)$ . Usually, for the nonparametric kernel regression, a value of  $b_T$  which is too small will give predictions with too high variance while a value of  $b_T$  which is too large will lead to an oversmooth

function with high bias. In [28], the authors propose to equilibrate the bias and the variance, fitting a Gaussian AR process to the data. If  $X$  denotes a  $p$ -Markovian time series, they assume that  $X$  satisfies the following equation

$$X_t = m + a_1 X_{t-1} + \dots + a_p X_{t-p} + \xi_t, \quad t \in \mathbb{Z},$$

where  $\xi$  is a Gaussian white noise ( $\xi_0 \sim \mathcal{N}(0, \sigma^2)$ ). In this case, assuming that the kernel  $W$  satisfies  $W(y) = \prod_{i=1}^p W(y_i)$  (e.g the Gaussian kernel), the number  $b_T = b_T(y)$  which minimizes the quantity

$$\int (\text{Bias } (F_T(x|y)))^2 + \text{Var } (F_T(x|y)) \, dx,$$

is given by (see [28])

$$b_T(y) = \left( \frac{\sigma^4 W_1}{T f_Y(y) W_2^2 (2\sigma^2 C_1(y)^2 + 0.25 C_2^2)} \right)^{\frac{1}{p+4}}, \tag{3.12}$$

where

$$C_1(y) = \mathbf{a}^T \mathbf{\Gamma}_p^{-1} (\mathbf{y} - \boldsymbol{\kappa}), \quad \mathbf{C}_2 = \mathbf{a}^T \mathbf{a},$$

with

$$\mathbf{a} = (\mathbf{a}_1, \dots, \mathbf{a}_p)', \quad \boldsymbol{\kappa} = (\mathbb{E}(\mathbf{X}_0), \dots, \mathbb{E}(\mathbf{X}_0))', \quad \mathbf{\Gamma}_p = (\text{Cov } (\mathbf{X}_{t-i}, \mathbf{X}_{t-j}))_{1 \leq i, j \leq p}.$$

Moreover  $W_1 = \int W^2(y) dy$  and  $W_2 = \int y_1^2 W(y_1) dy_1 < \infty$ . Note that an estimator  $\hat{b}_{AR}(y)$  of  $b_{AR}(y)$  can be derived substituting the unknown quantities in (3.12) by sample estimates or fitting the linear model to the data. The advantage of such an approximation is that the values of  $b_{AR}$  depend on the particular state  $y$  of the process. In fact this method can also be adapted to the case of random fields, assuming in our case

$$X_t | Y_t \sim \mathcal{N} \left( m + \sum_{j \in \mathcal{N}_o} a_j X_{t-j}, \sigma^2 \right)$$

and using a straightforward modification of (3.12).

One other approach is to use cross validation in the spirit of Härdle and Vieu [19]. For example, one can compute the kernel estimator using a part of the sample and then a prediction error using the second part of the sample. In [19], the case of the leave-out technique is studied for a two-dimensional time series.

We propose to combine the two approaches mentioned above. More precisely, for a particular state  $y$ , we define a bandwidth of the form

$$b_T(y) = C_T \times b_{AR}(y). \tag{3.13}$$

Then we first compute an estimator  $\hat{b}_{AR}$  as in [28] using the formula (3.12) and after the constant  $C_T$  is estimated using cross-validation for the kernel estimation of the mean  $\mathbb{E}(X_t | Y_t)$ , separating the sample into two parts of equal size.



In this case, the scale parameter  $C_T$  could correct the misspecification made by fitting the Gaussian model to the data. Of course, one can modify this approach using cross-validation for the kernel estimation of  $\mathbb{E}(\mathbb{1}_{X_t \leq x} | Y_t)$  and minimizing an integrated error over  $x$ . One can also use the leave-out method for the second step or just one step with cross validation only. However, more complicated methods will require more programming effort. In the next section, we use the two-step method described previously for the choice of the bandwidth parameter. We also mention that these methods also apply to the MMM algorithm, with appropriated modifications.

#### 4. The example of a pairwise-interaction Gibbs potential

In this section we consider the case of a Markov random field whose state space is  $S = [a, b]$ , the reference measure is the Lebesgue measure and the (translation invariant) Gibbs potential is given for  $A \subset \mathcal{N}_1$  by

$$\phi_A(x_A) = \begin{cases} \beta x_0, & \text{if } A = \{0\}, \\ \beta_j x_0 x_j, & \text{if } A = \{0, j\} \text{ and } \|j\|_\infty = 1, \\ 0, & \text{otherwise.} \end{cases} \tag{4.1}$$

A random field  $X$  associated with the previous Gibbs potential is a nearest neighbor Markov random field (i.e  $o = 1$ ), whose one point conditional densities are given for  $i \in \mathbb{Z}^d$  by

$$f_{X|Y}(x_i | y_i) = \frac{\alpha_i e^{x_i \alpha_i}}{e^{b \alpha_i} - e^{a \alpha_i}},$$

$$\alpha_i = \beta + \sum_{j \in \mathcal{N}_1} \beta_j x_{i+j}.$$

It is a particular example of a continuous spin Ising model. The following proposition makes condition (3.10) more precise in this context:

**Proposition 2.** *If  $\sum_{j \in \mathcal{N}_1} |\beta_j| < \frac{12}{(b-a)^2}$ , then condition (3.10) holds.*

#### Notes

- One can also extend this result for a neighborhood size  $o \geq 1$ .
- The more classical unicity condition of Dobrushin which is expressed in term of total variation distance (see 3.11) is less easy to precise. One can consider the stricter condition given by Simon [32] which writes

$$\sum_{A \ni i} (|A| - 1) \left( \sup_x \Phi_A(x_A) - \inf_x \Phi_A(x_A) \right) < 2.$$

For the potential defined in (4.1), when  $a = -b < 0$ , we obtain the condition  $\sum_{j \in \mathcal{N}_o} |\beta_j| < \frac{1}{b^2}$  which is more restrictive than that of Proposition 2.

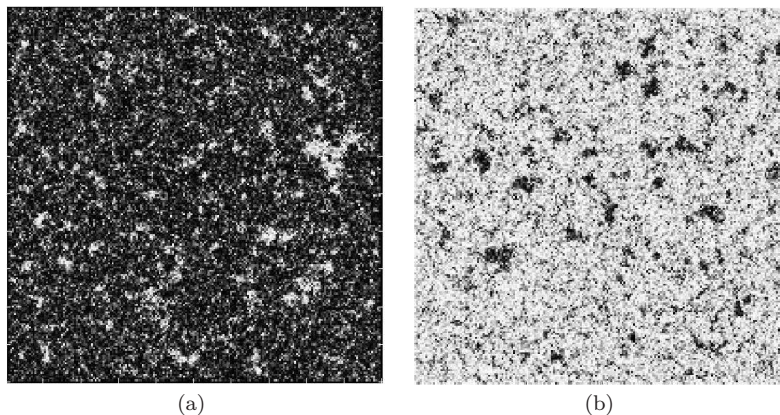


FIG 3. *Dependence with respect to boundary conditions (200 × 200 pixels).*

- For a potential given by (4.1), it is straightforward to verify the regularity assumptions (A2) and (A3). However the assumption (A1) is less clear. It is well known that mixing conditions with exponential decays hold under the usual Dobrushin unicity condition (see [17]). When for example  $b = -a$ , assumption (A1) is satisfied if  $\sum_{j \in \mathcal{N}_o} |\beta_j| < \frac{1}{b^2}$  (which implies the condition (3.11) by the remark above). Without condition (3.11), it is not easy to prove mixing properties of the random field. This problem occurs in particular if we assume only the weaker contraction condition given in Proposition 2. Nevertheless, the contraction condition (3.10) implies covariance inequalities between Lipschitzian functionals of some marginals of the unique Gibbs measure (see [15]). In this case, one can verify the Lipschitzian type mixing conditions introduced by Doukhan & Louichi [10] and an analogue of inequality (1) for the moment of partial sums is available (see [11]). Thus, with this slight modification of assumption (A1), Theorem 1 still holds. Details are omitted.

The problem of phase transition is illustrated in figure 3 when all the  $\beta_j$ 's equal to 0.5,  $\beta = 0$  and  $S = [-1, 1]$ . In this case, a simulation of a sample of size  $200 \times 200$  does not give the same distribution when the boundary conditions satisfy  $z \equiv -1$  or  $z \equiv 1$ , as it is usually the case in statistical mechanics when the inverse of the so-called temperature (here 0.5) is sufficiently high to ensure that  $|\mathcal{G}(\Phi)| > 1$ .

When  $S = [-1, 1]$ , Figure (4) shows a sample and its replicate in two cases. For the first case, we set  $\beta = 0$  and the  $\beta_j$ 's are given by the following array

-1.2	0	1.2
0	×	0
1.2	0	-1.2

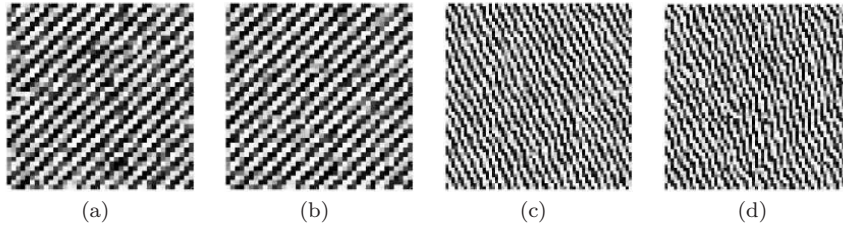


FIG 4. (a) sample  $40 \times 40$ , (b) replicate  $40 \times 40$ , (c) sample  $60 \times 60$ , (d) replicate  $60 \times 60$ .

while for the second case, we set  $\beta = 1$  and the  $\beta_j$ 's are given by the array

1.5	1.5	-1.5
-1.5	$\times$	-1.5
-1.5	1.5	1.5

Observe that the condition given in Proposition 2 does not hold, but we choose the coefficients  $\beta_j$  large enough in order to make the “attraction-repulsion” phenomenon in the Gibbs potential sufficient for a visual analysis of the interactions and of the quality of the resampling. Indeed, condition (3.10) is related to weak interactions for the random field and local characteristics of the sample are less distinguishable. In figure 3, in order to avoid a possible dependency of the simulation with respect to the boundary conditions, we use the same boundary for the sample and the bootstrap replicate.

Of course, one could also estimate the distribution of some linear statistics as in [28] and a comparison with the MMM algorithm studied in [4] or the block method [30] would be of interest. This can be done for model (4.1), under the condition of Proposition 2 which ensures the consistency of the bootstrap distributions. Those aspects will not be considered in the present paper: a statistical comparison between the MMM and the method of Gibbs sampling requires the simulation of several bootstrap samples for several original samples and then a high computational coast. This kind of comparison could be investigated more seriously in a subsequent paper.

## 5. The approach of Paget and Longstaff for texture synthesis and simulation examples

### 5.1. Paget and Longstaff method

For the texture synthesis problem, Paget and Longstaff [27] investigate a nonparametric method also based on the Gibbs sampler, using kernel density estimation, but the consistency of their algorithm was not considered. Their nonparametric estimator of the LCD is not directly linked to the regression estimation and resampling but all the results formulated in section 3 have their analogue. However, we choose to formulate our results in a resampling context and in the spirit of the existing literature ([4, 28]).

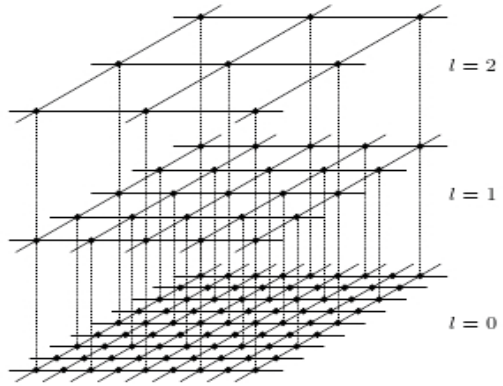


FIG 5. Grid organization via decimation (source [27]).

For highly structured textures, the convergence of the Gibbs sampler can lead to some numerical difficulties, especially if the neighborhood size  $o$  must be very large to capture the global characteristics of the texture and if the Gibbs potential of the underlying stochastic process has a large number of local minima. In this case, a huge number of iterations could be necessary in order to reach the equilibrium.

Then Paget and Longstaff proposed a multiscale algorithm, using a multigrid representation of an image, as shown in Figure 5 which is taken from [27]. If  $S_0 = [0, M_1] \times [0, M_2]$  represents the pixel's sites of an image  $x_0$ , the lower resolutions, or higher grid levels  $l > 0$ , are decimated versions of the image at level  $l = 0$ . For a grid level  $l > 0$ , the image  $x_l$  is defined on the lattice  $S_l \subset S$ , where

$$S_l = \{s = (2^l i, 2^l j) / 0 \leq i \leq M_1 / 2^l, 0 \leq j \leq M_2 / 2^l\}.$$

The set of sites  $S_l$  at level  $l$  represents a decimation of the previous set of sites  $S_{l-1}$  at the lower grid level  $l - 1$ . The neighborhood system is redefined for each grid level  $l > 0$ :

$$\mathcal{N}_t^l = \{s \in S_l / \|t - s\|_\infty \leq o\}.$$

For a level grid  $l$ , the Gibbs sampling is not applied to the sites  $s \in S^{l+1}$ . In fact, a multiscale representation of a Markov random field is used (see [23] for a precise definition). To better incorporate the multiscale relaxation described above, Paget and Longstaff have also introduced a pixel temperature function in order to determine when to stop the Gibbs sampler at one level and start it at the next level. Let  $l$  be a grid level. A pixel temperature is incorporated in equation (2.2) by modifying the form of the difference

$$d = y - Y_s. \tag{5.1}$$

In fact, at the beginning of the stochastic relaxation at a level  $l$ , they define for a site  $j \in S_l$  of the output texture the pixel temperature  $c_j$  as follows:  $c_j = 0$  if

$j \in S^{l+1}$  and  $c_j = 1$  otherwise. The difference  $d$  is replaced by  $d'$  such that:

$$d'_j = (1 - c_{t+j})(x_{t+j} - X_{s+j}), \quad j \in \mathcal{N}_o.$$

When a pixel  $x_t$  has been relaxed in the stochastic relaxation process, Paget and Longstaff set:

$$\tilde{T}_t = \max\left\{0, \frac{\xi + \sum_{j \in \mathcal{N}_o} c_{t+j}}{|\mathcal{N}_o|}\right\}$$

where  $\xi < 0$  is fixed by the user.

Here, the idea is to provide a total confidence to pixels coming from the preceding resolution and to progressively increase the confidence level of a pixel synthesized in the present resolution. When  $c_j = 0 \forall j \in S_l$ , the stochastic relaxation process is considered to have reached an equilibrium state, indicating that the image can be propagated to the next lower grid level. This notion of temperature is related to the global temperature used in stochastic annealing (see [16]). Although we have incorporated this pixel temperature function for texture synthesis, we will not study in this paper statistical properties of such an approximation.

## 5.2. Texture synthesis examples

In order to illustrate the behaviour of the resampling scheme for the texture synthesis problem, we use the multiscale algorithm with the pixel temperature function described above in order to simulate a new texture from a sample. The Gaussian kernel is chosen for the simulations.

- The neighborhood size is set at  $o = 3$  or  $o = 4$ .
- As in [26], we set  $\xi = -1$ . Moreover 4 or 5 grid levels are used for the synthesis.

Figure 6 shows a step of the synthesis at the highest resolution. The Gibbs sampler runs in the raster ordering (the periodic visiting scheme (3.4)) and

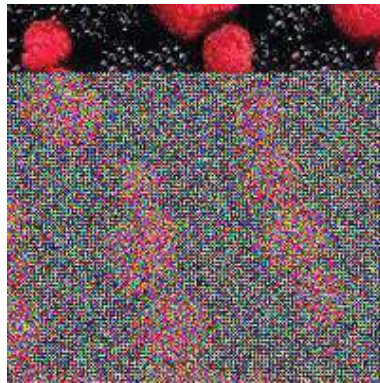


FIG 6. *The Gibbs sampler and the highest resolution.*

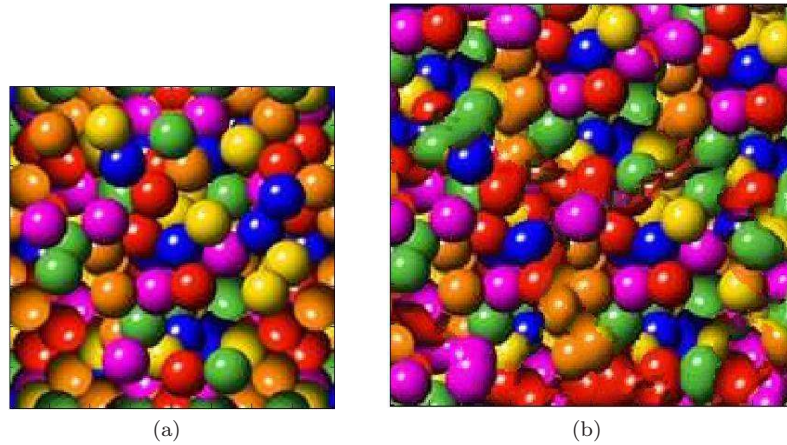


FIG 7. (a) Original texture  $160 \times 160$  pixels, (b) Synthesis with the multiresolution algorithm  $200 \times 200$  pixels.

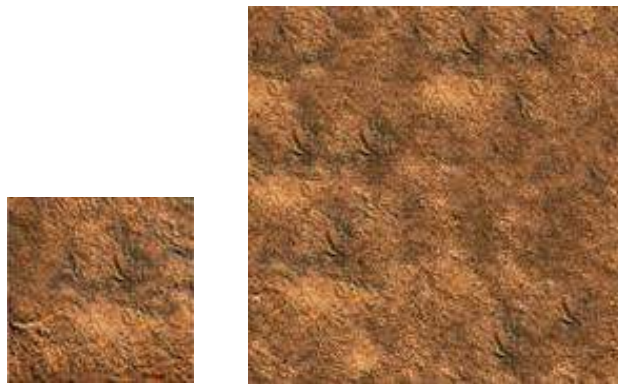


FIG 8. Original texture ( $128 \times 128$ ) and synthesis ( $200 \times 200$ ).

Figure 6 shows the first sweep. One can see that the lower resolutions give the shape of the texture. Moreover the pixel temperature function helps for a good initialization of the sampler. Simulation results are also given in Figure 8 and Figure 7 and the visual quality is comparable to the simulations given in [4]. Note that small discontinuities appear for the bootstrap sample in Figure 7. Those problems could be explained by the presence of several local maxima for the true probability distribution of the sample but also by additional local maxima due to the nonparametric estimation.

A comparison between the different algorithms in the texture synthesis setting is not the objective of the present paper. Generally, the MMM and the multiscale algorithm give good results and the computer graphics community has explored several variants and complex computational tools very well adapted for this

difficult problem over the last decade (see in particular [22, 12] or [21]), in particular for improving the spiral algorithm of Efros & Leung. Instead of a visual appreciation of the simulations, we believe more informative a comparison of some statistical properties of the different resampling schemes used in their simplest form, in order to bootstrap the distribution of a simple model (for example model (4.1)).

## 6. Appendix

### 6.1. Assumptions of Theorem 1

(A1) The random field  $X$  is strictly stationary and  $\alpha$ -mixing: for  $k, c \in \mathbb{N}^*$ , we assume that the strong mixing coefficients

$$\alpha_X(k, c) = \sup \{ |P(AB) - P(A)P(B)|, A \in \sigma(X_E), B \in \sigma(X_F) \},$$

where the supremum is taken over all the subsets  $E, F \subset \mathbb{Z}^2$  such that  $d(E, F) = k$ ,  $|E| + |F| \leq c$ , satisfy

$$\sum_{k=1}^{\infty} (k+1)^{2c-1} \alpha_X(k, c)^{\epsilon/(c+\epsilon)} < \infty,$$

for a given  $\epsilon > 0$  and a real number  $\tau > 2$  such that for  $c = 2 \lceil \tau/2 \rceil$ .

(A2) The distribution of  $X_t$  has a compact support  $S \subset \mathbb{R}$ .

(A3)  $F_{X,Y} = \mathbb{P}(X_t \leq \cdot, Y_t \leq \cdot)$ ,  $F_{X|Y}$  and  $F_Y = \mathbb{P}(Y_t \leq \cdot)$  have bounded continuous strictly positive densities (denoted  $f_{X,Y}$ ,  $f_{X|Y}$  and  $f_Y$  respectively) with respect to Lebesgue measure. Moreover, there exists  $L > 0$  such that for any  $y, y' \in S^A$ , any  $x \in S$ ,

$$\left| \int_{-\infty}^x f_{X,Y}(z, y) dz - \int_{-\infty}^x f_{X,Y}(z, y') dz \right| \leq L \|y - y'\|.$$

(A4) The family of kernels  $(W^{(\ell)})_{\ell \in \mathbb{N}^*}$  is such that  $W^{(\ell)} : \mathbb{R}^\ell \rightarrow (0, \infty)$  are bounded, symmetric and first-order Lipschitz continuous functions such that for all  $\ell \in \mathbb{N}^*$ ,

$$\int u W^{(\ell)}(u) d\lambda_\ell(u) = 0 \quad \text{and} \quad \int \|u\| W^{(\ell)}(u) d\lambda_\ell(u) < \infty.$$

Moreover, the width of  $W_b^{(\ell)}$  is supposed to be such that  $b = b_T = O([T]^{-\delta})$ , with  $\delta > 0$ .

### 6.2. Proof of Theorem 4

We follow the proof of theorem 2 of Bickel and Levina in order to compute a convergence rate. We first recall the following lemma which proof can be found in [9].

**Lemma 1.** (Moment inequality). Let  $F_t$  be a real-valued random field indexed by  $I \subset \mathbb{Z}^d$  satisfying conditions **(A1)**. If  $\mathbb{E}F_t = 0$ ,  $F_t \in \mathbb{L}^\tau$  then there exists a constant  $C$  depending only on  $\tau$  and mixing coefficients of  $F_t$  such that

$$\mathbb{E} \left| \sum_{t \in I} F_t \right|^\tau \leq C \max \left( L(\tau, \epsilon), L(2, \epsilon)^{\tau/2} \right),$$

where

$$L(\kappa, \epsilon) = \sum_{t \in I} \left( \mathbb{E} |F_t|^{\kappa+\epsilon} \right)^{\kappa/(\kappa+\epsilon)}.$$

It is easy to see that if  $\sup_t \|F_t\|_\infty \leq M$ , then we obtain:

$$\mathbb{E} \left| \sum_{t \in I} F_t \right|^\tau \leq CM^\tau |I|^{\tau/2}. \tag{6.1}$$

For  $(x, y) \in S \times S^A$ , we set:

$$\begin{aligned} r_T(x, y) &= [T]^{-1} \sum_{s \in I_T} \mathbb{1}_{(-\infty, x]}(X_s) W_b(y - Y_s), \\ r(x, y) &= \int \mathbb{1}_{(-\infty, x]}(z) f_{X,Y}(z, y) dz, \\ f_T(y) &= [T]^{-1} \sum_{s \in I_T} W_b(y - Y_s). \end{aligned}$$

We have:

$$F_T(x|y) = \frac{r_T(x, y)}{f_T(y)}, \quad F_{X|Y}(x|y) = \frac{r(x, y)}{f_Y(y)}. \tag{6.2}$$

Following the proof of lemma A2 in [4], we prove the following result

**Lemma 2.** Under assumptions (A1)–(A4), for any  $x \in \mathbb{R}$

$$\sup_{(x,y) \in S \times S^A} |r_T(x, y) - r(x, y)| = O([T]^{-\gamma})$$

for  $0 < \gamma < \frac{\tau-2}{2(v+1)(\tau+v+2)}$ .

**Proof of Lemma 2** In this proof, we will denote by  $C > 0$  a generic constant which does not depend on  $T$ .

Let  $\delta > 0$  such that  $b_T = O([T]^{-\delta})$ , then the proof of lemma A2 in [4] leads to

$$\sup_{(x,y) \in S \times S^A} |\mathbb{E}r_T(x, y) - r(x, y)| = O([T]^{-\delta}). \tag{6.3}$$

Then we need to bound  $\sup_{(x,y) \in S \times S^A} |r_T(x, y) - \mathbb{E}r_T(x, y)|$ .

As in [4], we define

$$Z_{t,T}(x, y) = \mathbb{1}_{(-\infty, x]}(X_t) W_{b_T}(y - Y_t) - \mathbb{E}(\mathbb{1}_{X_t \leq x} W_{b_T}(y - Y_t))$$

and we need to bound  $\sup_{(x,y) \in S \times S^A} \left| \frac{1}{[T]} \sum_{t \in I_T} Z_{t,T}(x, y) \right|$ .



As  $S \times S^A$  is compact, we can cover  $S \times S^A$  with  $N_T$  cubes  $I_{i,T}$  with centers  $(x_i, y_i)$  and sides  $L_T$  for the supremum norm. Without loss of generality, we suppose  $x_1 \leq \dots \leq x_{N_T}$  and we set  $x_0 = x_1 - L_T$  and  $x_{N_T} = x_{N_T} + L_T$ . Then

$$\begin{aligned} & \sup_{(x,y) \in S \times S^A} \left| [T]^{-1} \sum_{t \in I_T} Z_{t,T}(x, y) \right| \\ & \leq \max_{1 \leq i \leq N_T} \left| [T]^{-1} \sum_{t \in I_T} Z_{t,T}(x_i, y_i) \right| \\ & + \max_{1 \leq i \leq N_T} \sup_{(x,y) \in (S \times S^A) \cap I_{i,T}} \left| [T]^{-1} \sum_{t \in I_T} (Z_{t,T}(x, y) - Z_{t,T}(x_i, y_i)) \right| \\ & = I + II. \end{aligned}$$

- First let us deal with term II. Using assumption **(A4)** for the kernel, we have for  $t \in I_T$  and  $x \in (x_{i-1}, x_i]$ :

$$\begin{aligned} & |Z_{t,T}(x, y) - Z_{t,T}(x_i, y_i)| \\ & \leq C \left( b_T^{-(v+1)} \|y - y_i\| + b_T^{-v} (\mathbb{1}_{x_{i-1} < X_t \leq x_i} + \mathbb{P}(x_{i-1} < X_t \leq x_i)) \right) \\ & \leq C \left( b_T^{-(v+1)} L_T + b_T^{-v} (\mathbb{1}_{x_{i-1} < X_t \leq x_i} + \mathbb{P}(x_{i-1} < X_t \leq x_i)) \right). \end{aligned}$$

We choose  $L_T = [T]^{-\beta}$  and we set  $U_{i,t} = \mathbb{1}_{]x_{i-1}, x_i]}(X_t) - \mathbb{P}(x_{i-1} < X_t \leq x_{i+1})$ . Note that assumption **(A2)** about the existence of densities allows to derive the bound:

$$\mathbb{P}(x_{i-1} < X_0 \leq x_i) \leq CL_T.$$

We have:

$$\begin{aligned} & \sup_{(x,y) \in (S \times S^A) \cap I_{i,T}} \left| [T]^{-1} \sum_{t \in I_T} (Z_{t,T}(x, y) - Z_{t,T}(x_i, y_i)) \right| \\ & \leq C \left( [T]^{\delta(v+1)-\beta} + [T]^{v\delta-1} \left| \sum_{t \in I_T} U_{i,t} \right| + [T]^{v\delta} \mathbb{P}(x_{i-1} < X_0 \leq x_i) \right) \\ & \leq C \left( [T]^{\delta(v+1)-\beta} + \max_{1 \leq i \leq N_T} [T]^{v\delta-1} \left| \sum_{t \in I_T} U_{i,t} \right| \right). \end{aligned}$$

Now we consider a real number  $\gamma < \frac{\tau - 2 - 2v\delta\tau - 2\beta(v+1)}{2\tau}$ .

Since  $N_T = O(L_T^{-(v+1)}) = O([T]^{\beta(v+1)})$ , we obtain using (A1) and lemma 1

$$\begin{aligned} \mathbb{P} \left( \max_{1 \leq i \leq N_T} [T]^{v\delta-1} \left| \sum_{t \in I_T} U_{i,t} \right| > [T]^{-\gamma} \right) & \leq \sum_{i=1}^{N_T} [T]^{(\gamma+v\delta-1)\tau} \mathbb{E} \left| \sum_{t \in I_T} U_{i,t} \right|^\tau \\ & \leq C [T]^{\beta(v+1) + (\gamma+v\delta-1)\tau + \frac{\tau}{2}}. \end{aligned}$$

By the choice of  $\gamma$ , we have  $\beta(v+1) + (\gamma + v\delta - 1)\tau + \frac{\tau}{2} < -1$  and we deduce from the Borel Cantelli lemma that

$$\max_{1 \leq i \leq N_T} [T]^{v\delta-1} \left| \sum_{t \in I_T} U_{i,t} \right| = O([T]^{-\gamma}) \text{ a.s.}, \quad \gamma < \frac{\tau - 2 - 2v\delta\tau - 2\beta(v+1)}{2\tau}.$$

Now using the previous inequalities, we deduce that:

$$II \leq O\left([T]^{\delta(v+1)-\beta} + [T]^{-\gamma}\right), \quad \gamma < \frac{\tau - 2 - 2v\delta\tau - 2\beta(v+1)}{2\tau}. \quad (6.4)$$

- Now we turn on the term I. For a real number  $\tilde{\gamma} < \frac{\tau - 2 - 2v\delta\tau - 2\beta(v+1)}{2\tau}$ , we have using (A1) and lemma 1:

$$\begin{aligned} \mathbb{P}(I > [T]^{-\tilde{\gamma}}) &\leq \sum_{i=1}^{N_T} [T]^{(\tilde{\gamma}-1)\tau} \mathbb{E} \left| \sum_t Z_{t,T}(x_i, y_i) \right|^\tau \\ &\leq CN_T [T]^{(\tilde{\gamma}-1)\tau + \frac{\tau}{2}} b_T^{-v\tau} \\ &\leq C [T]^{\beta(v+1) + (\tilde{\gamma}-1)\tau + \frac{\tau}{2} + \delta v\tau}. \end{aligned}$$

By the choice of  $\tilde{\gamma}$ , we have

$$\beta(v+1) + (\tilde{\gamma} - 1)\tau + \frac{\tau}{2} + \delta v\tau < -1,$$

and by the Borel Cantelli lemma, we have

$$I = O\left(T^{-\tilde{\gamma}}\right) \text{ a.s.}, \quad \tilde{\gamma} < \frac{\tau - 2 - 2v\delta\tau - 2\beta(v+1)}{2\tau}.$$

Now we choose the number  $\beta$  such that:

$$\beta - \delta(v+1) = \frac{\tau - 2 - 2v\delta\tau - 2\beta(v+1)}{2\tau}.$$

This leads to  $\beta = \frac{\tau - 2 + 2\tau\delta}{2(v + \tau + 1)}$  and to the following rate:

$$I + II = O\left([T]^{-\gamma}\right), \quad \gamma < \frac{\tau - 2 - 2\delta((v+1)^2 + v\tau)}{2(v + \tau + 1)}.$$

Finally, we choose  $\delta$  for an equilibrium with the bound (6.3), solving the equation:

$$\delta = \frac{\tau - 2 - 2\delta((v+1)^2 + v\tau)}{2(v + \tau + 1)}.$$

This leads to:

$$\delta = \frac{\tau - 2}{2(v+1)(\tau + v + 2)} > 0,$$

which gives the rate given by the Lemma 2.  $\square$

**Proof of Theorem 4** We write:

$$\begin{aligned} & |F_T(x|y) - F_{X|Y}(x|y)| \\ &= \frac{1}{f_T(y)} |r_T(x, y) - r(x, y) + r(x, y) - F_{X|Y}(x|y)f_T(y)| \\ &\leq \frac{1}{f_T(y)} (|r_T(x, y) - r(x, y)| + F_{X|Y}(x|y) |f_T(y) - f_Y(y)|). \end{aligned}$$

By lemma 2, we have  $\sup_{x,y} |r_T(x, y) - r(x, y)|, \sup_y |f_T(y) - f_Y(y)| = O(T^{-\gamma})$  a.s with  $0 < \gamma < \frac{\tau-2}{2(v+1)(\tau+v+2)}$ . Since by (A2),  $\inf_{y \in S^A} f_Y(y) > 0$  and  $\sup_{(x,y) \in S \times S^A} F_{X|Y}(x|y) \leq 1$ , we get the result.  $\square$

### 6.3. Proof of Proposition 1

Suppose that  $Q$  is a probability measure on  $E = I_T^{\mathbb{Z}^2}$  whose one point conditional distributions are given by  $F_T$ . For  $h \in \mathbb{Z}^2$ , we denote by  $p_h : z \mapsto z_h$  the projection defined on  $E$ . We consider  $k \in (\mathbb{N}^*)^2$  and  $i \in \mathcal{N}_o$  such that  $\{k, k+i\} \subset I_T$ . We have:

$$\begin{aligned} & \frac{Q(p_0 = p_i = X_{i+k} | p_h = X_k, h \notin \{0, i\})}{Q(p_0 = p_i = X_k | p_h = X_k, k \notin \{0, i\})} \\ &= \frac{Q(p_0 = X_{i+k} | p_i = X_{i+k}, p_h = X_k, h \notin \{0, i\})}{Q(p_0 = X_k | p_i = X_{i+k}, p_h = X_k, h \notin \{0, i\})} \\ &\times \frac{Q(p_i = X_{i+k} | p_h = X_k, h \neq i)}{Q(p_i = X_k | p_h = X_k, h \neq i)} \\ &= \frac{Q(p_0 = X_{i+k} | p_i = X_{i+k}, p_h = X_k, h \notin \{0, i\})}{Q(p_0 = X_k | p_i = X_{i+k}, p_h = X_k, h \notin \{0, i\})} \\ &\times \frac{Q(p_0 = X_{i+k} | p_h = X_k, h \neq 0)}{Q(p_0 = X_k | p_h = X_k, h \neq 0)} \end{aligned}$$

On the other hand, we have also:

$$\begin{aligned} & \frac{Q(p_0 = p_i = X_{i+k} | p_h = X_k, h \notin \{0, i\})}{Q(p_0 = p_i = X_k | p_h = X_k, h \notin \{0, i\})} \\ &= \frac{Q(p_i = X_{i+k} | p_0 = X_{i+k}, p_h = X_k, h \notin \{0, i\})}{Q(p_i = X_k | p_0 = X_{i+k}, p_h = X_k, h \notin \{0, i\})} \\ &\times \frac{Q(p_0 = X_{i+k} | p_h = X_k, h \neq 0)}{Q(p_0 = X_k | p_h = X_k, h \neq 0)} \end{aligned}$$

Using that  $W(u) = (2\pi)^{-v/2} \exp(-\|u\|^2/2)$  as well as the form of the distributions  $F_T$ , the equality between the two expressions of  $\frac{Q(p_0=p_i=X_{i+k}/p_h=X_k, k \notin \{0, i\})}{Q(p_0=p_i=X_k/p_h=X_k, k \notin \{0, i\})}$  writes:

$$\begin{aligned}
 & |X_{2i+k} - X_{i+k}|^2 + \sum_{h \in \mathcal{N}_o \setminus \{i\}} |X_{i+k+h} - X_k|^2 - \sum_{h \in \mathcal{N}_o \setminus \{i\}} |X_{k+h} - X_k|^2 \\
 = & |X_k - X_{i+k}|^2 + \sum_{h \in \mathcal{N}_o \setminus \{-i\}} |X_{i+k+h} - X_k|^2 \\
 - & \sum_{h \in \mathcal{N}_o \setminus \{-i\}} |X_{k+h} - X_k|^2 - |X_{k-i} - X_{i+k}|^2, \quad \text{a.s.}
 \end{aligned}$$

Then we derive:

$$X_{i+k}^2 - X_{i+k} (X_{k-i} + X_{2i+k}) = X_k^2 - X_k (X_{k-i} + X_{2i+k}), \quad \text{a.s.}$$

This is a contradiction with the density assumption for the conditional law  $X_k/X_s, s \neq k$ . □

### 6.4. Proof of Theorem 2

In the theory of Markov processes, conditions ensuring the continuity of the invariant distribution with respect to the generator can be found in [14] (see theorem 9.10) and the following theorem is a particular case for discrete time Markov chains with a compact state space. For completeness of this work we give a proof of the following theorem.

**Theorem 5.** *Suppose that for  $T \in \mathbb{N}^2$  such that  $I_T \neq \emptyset$ ,  $Q_T$  is a transition on  $I_T^R \subset S^R$ . Assume that there exists a transition  $P$  on  $S^R$  satisfying for a given continuous and bounded real valued function  $g$  on  $S^R$ :*

1.  $\sup_{x \in I_T^R} \left| \int g(\tilde{x}) (Q_T(x, d\tilde{x}) - P(x, d\tilde{x})) \right| \rightarrow_{T \rightarrow \infty} 0$ .
2.  $x \rightarrow \int g(\tilde{x}) P(x, d\tilde{x})$  is continuous on  $S^R$ .

We suppose that  $Q_T$  (resp.  $P$ ) admits an unique invariant distribution  $\nu_T$  (resp.  $\mu$ ). Then we have almost surely:

$$\lim_{T \rightarrow \infty} \nu_T = \mu, \quad \text{in distribution.}$$

**Proof of Theorem 5** By definition of  $\nu_T$ , we have the following equality:

$$\nu_T(A) = \int Q_T(x, A) \nu_T(dx), \quad A \in \mathcal{B}(S^R),$$

where  $\mathcal{B}(S^R)$  denote the Borel  $\sigma$ -algebra on  $S^R$ . Since  $S^R$  is a compact metric space, the tightness of the sequence  $(\nu_T)_T$  implies the existence of a cluster point denoted by  $\nu$ . We are going to show that  $\nu = \mu$ . Then by uniqueness of the cluster point, we will conclude:

$$\text{Almost surely : } \lim_{T \rightarrow \infty} \nu_T = \mu \quad \text{in distribution.} \tag{6.5}$$

To show that  $\nu = \mu$ , it is sufficient to prove that  $\nu$  is an invariant probability of the transition  $P$  on  $S^R$ . Then we need to show that  $\nu(A) = \int P(x, A)\nu(dx)$ ,  $\forall A \in \mathcal{B}(S^R)$ .

Suppose that  $(T_n)_{n \in \mathbb{N}}$  is sequence in  $\mathbb{N}^* \times \mathbb{N}^*$  such that  $\lim_{n \rightarrow \infty} \nu_{T_n} = \nu$ . Then if  $g$  is a continuous and bounded real valued function on  $S^R$ , we have:

$$\begin{aligned} & \left| \int g(y)\nu_{T_n}(dy) - \int g(y)P(x, dy)\nu(dx) \right| \\ &= \left| \int g(y)Q_{T_n}(x, dy)\mu_{T_n}(dx) - \int g(y)P(x, dy)\nu(dx) \right| \\ &\leq \sup_{x \in I_T^R} \left| \int g(y)Q_{T_n}(x, dy) - \int g(y)P(x, dy) \right| \\ &+ \left| \int g(y)P(x, dy)(\nu_{T_n} - \nu)(dx) \right| \\ &= A + B. \end{aligned}$$

By assumption, the function  $x \rightarrow \int g(y)P(x, dy)$  is still bounded and continuous and the weak convergence of the sequence  $(\nu_{T_n})_n$  implies that  $B \rightarrow 0$  ( $n \rightarrow \infty$ ).

More,  $A \rightarrow 0$  ( $n \rightarrow \infty$ ) follows directly from the assumption 1. Then, by the uniqueness of the limit of the sequence  $(\nu_{T_n})_n$ , we conclude that  $\nu(dy) = \int P(x, dy)\nu(dx)$  and the conclusion of the theorem follows.  $\square$

Theorems 5 and 4 can now be applied to prove the convergence of the invariant distributions of the approximated Gibbs sampler for the two visiting schemes (3.4) and (3.5).

**Proof of Theorem 2** First, from the positivity assumption on the kernel  $W$  in Theorem 2.2,  $P_{T,R}^*$  is a.s a positive transition on the finite state space  $I_T^R$ . Then this Markov chain has a unique invariant distribution  $\mu_{T,R}^*$  and almost surely, we have for a given starting value  $x^{(0)} \in I_T^R$ :

$$\lim_{n \rightarrow \infty} (P_{T,R}^*)^n(x^{(0)}, dx) = \mu_{T,R}^*(dx), \quad \text{in distribution,}$$

and the approximated Gibbs sampler converges to  $\mu_{T,R}^*$ .

Now we are going to show that transitions  $P_{T,R}^*, P_R$  satisfy the assumptions of Theorem 5. The positivity of the transition  $P_R$  has been already discussed for both cases. As  $\mu = \mathbb{P}(X_R | X_{\partial R} = z_{\partial R})$  is an invariant distribution for the transition  $P_R$ , we know from theoretical results about Markov chains that  $\mu$  is the unique invariant distribution for  $P_R$  (see [24]). Assumption 2 in Theorem 5 holds from assumption (A2) on  $f_{X|Y}$  and the Lebesgue theorem. So we only need to verify assumption 1 in Theorem 5. To this end, we first prove that for a given continuous and bounded real valued function  $h$  on  $S^R \times S^R$ , we have  $\forall s \in R$ :

$$\sup_{(x, \tilde{x}) \in I_T^R \times I_T^R} \left| \int h(x, \tilde{x}) (F_T(d\tilde{x}_s | \tilde{x}x(s)) - F_{X|Y}(d\tilde{x}_s | \tilde{x}x(s))) \right| \rightarrow_{T \rightarrow \infty} 0, \quad \text{a.s.} \tag{6.6}$$

Indeed if  $\epsilon > 0$ , the uniform continuity of  $h$  on  $S^R \times S^R$  entails the existence of a number  $\delta > 0$  such that

$$\max(\|\rho - \rho'\|_\infty, \|\kappa - \kappa'\|_\infty) < \delta \Rightarrow |h(\rho, \kappa) - h(\rho', \kappa')| < \epsilon.$$

Now choose a subdivision  $a_1, \dots, a_k$  of  $S$  such that for  $i \in \{1, \dots, k-1\}$  either  $(a_i, a_{i+1}) \cap S$  is an empty set, either  $|a_i - a_{i+1}| < \delta$ . Then for  $s \in R$ , we define the function  $h_{k,s}$  on  $S^R \times S^R$  by

$$h_{k,s}(\rho, \kappa) = \sum_{\ell=1}^{k-1} \mathbb{1}_{[a_\ell, a_{\ell+1})}(\kappa_s) h(\rho, \kappa^{(\ell)}),$$

where for  $\ell \in \{1, \dots, k-1\}$ ,  $\kappa^{(\ell)}$  is the element of  $S^R$  such that  $\kappa_s^{(\ell)} = a_\ell$  and  $\kappa_t^{(\ell)} = \kappa_t$  if  $t \in R \setminus \{s\}$ . Then we have  $\|h - h_{k,s}\|_\infty < \epsilon$  and we deduce

$$\begin{aligned} & \sup_{(x, \tilde{x}) \in I_T^R \times I_T^R} \left| \int h(x, \tilde{x}) (F_T(d\tilde{x}_s | \tilde{x}x(s)) - F_{X|Y}(d\tilde{x}_s | \tilde{x}x(s))) \right| \\ & \leq 2\epsilon + (k-1) \|h\|_\infty \|F_T(\cdot) - F_{X|Y}(\cdot)\|_\infty \\ & \leq 3\epsilon, \end{aligned}$$

if  $T$  is large enough, using Theorem 1. Then the limit (6.6) is proved.

Now, we use (6.6) to verify assumption 1 in Theorem 5.

- In the case of the periodic visiting scheme (3.4), if  $u \in R$  is such that  $u \succ s, \forall s \in R \setminus \{u\}$ , we have:

$$\begin{aligned} & \left| \int g(\tilde{x}) P_{T,R}^*(x, d\tilde{x}) - \int g(\tilde{x}) P_R(x, d\tilde{x}) \right| \\ & \leq \left| \int \int g(\tilde{x}) F_{X|Y}(d\tilde{x}_u | \tilde{x}x(u)) \right. \\ & \quad \times \left( \otimes_{s \in R \setminus \{u\}} F_T(d\tilde{x}_s | \tilde{x}x(s)) - \otimes_{s \in R \setminus \{u\}} F_{X|Y}(d\tilde{x}_s | \tilde{x}x(s)) \right) \left. \right| \\ & \quad + \sup_{\tilde{x}_s \in S, s \neq u} \left| \int g(\tilde{x}) (F_T(d\tilde{x}_u | \tilde{x}x(u)) - F_{X|Y}(d\tilde{x}_u | \tilde{x}x(u))) \right|. \end{aligned} \tag{6.7}$$

Then if we iterate the bound (6.7), using a non increasing enumeration of the sites of  $R$ , we deduce the existence of a family  $\{h_s/s \in R\}$  of continuous and bounded functions on  $S^R \times S^R$  such that:

$$\begin{aligned} & \left| \int g(\tilde{x}) P_{T,R}^*(x, d\tilde{x}) - \int g(\tilde{x}) P_R(x, d\tilde{x}) \right| \\ & \leq \sum_{s \in R} \sup_{(x, \tilde{x}) \in I_T^R \times I_T^R} \left| \int h_s(x, \tilde{x}) (F_T(d\tilde{x}_s | \tilde{x}x(s)) - F_{X|Y}(d\tilde{x}_s | \tilde{x}x(s))) \right|. \end{aligned} \tag{6.8}$$

Note that the continuity of  $h_s$  follows from assumption (A2) and the Lebesgue theorem. Then applying (6.6) to the bound (6.8), we conclude that assumption 1 in Theorem 5 holds.

- In the case of the random visiting scheme (3.5), we have for  $u \in I_T^R$ :

$$\begin{aligned} & \left| \int g(w) P_{T,R}^*(u, dw) - \int g(w) P_R(u, dw) \right| \\ & \leq \frac{1}{|R|} \sum_{s \in R} \left| \int g(w_s u_{R \setminus \{s\}}) (F_T(dw_s | u_{s+\mathcal{N}_o}) - F_{X|Y}(dw_s | u_{s+\mathcal{N}_o})) \right|. \end{aligned} \tag{6.9}$$

By (6.6), the last expression converges to zero when  $T \rightarrow \infty$ . □

**Proof of Theorem 6.1** For the proof we reformulate the contraction method used by Dobrushin [8] (see also [15] (p 116-123)) in order to obtain a contraction for the transition  $P_R$ . From the two expressions of the Wasserstein metric  $d_W$ , we have the following inequality

$$\left| \int g dG - \int g dH \right| \leq L(g) \int_S |G(w) - H(w)| dw, \tag{6.10}$$

which holds for a Lipschitz function  $g : S \rightarrow \mathbb{R}$  and for two distribution functions  $H, G$  with support  $S$  ( $L(g)$  denotes the Lipschitz constant of  $g$ ).

We use the notation  $u_j^- = (u_i)_{i \in \mathbb{Z}^2 \setminus \{j\}}$ . For  $f \in \mathcal{L}_R$ , we have

$$\begin{aligned} & \left| \int f d\mu_{T,R}^* - \int f d\mu_R \right| \\ & \leq \left| \int P_R(f) d\mu_{T,R}^* - \int P_R(f) d\mu_R \right| + \left| \int (P_R(f) - P_{T,R}^*(f)) d\mu_{T,R}^* \right| \\ & \leq L(P_R(f)) \times d_R(\mu_{T,R}^*, \mu_R) \\ & + \frac{1}{|R|} \sum_{j \in R} \sup_{u_j^-} \left| \int f(u) F_T(du_j | u_{j+\mathcal{N}_o}) - \int f(u) F_{X|Y}(du_j | u_{j+\mathcal{N}_o}) \right| \\ & \leq L(P_R(f)) \times d_R(\mu_{T,R}^*, \mu_R) \\ & + \frac{1}{|R|} L(f) \sup_{y \in \mathcal{N}_o} \int_S |F_T(x|y) - F_{X|Y}(x|y)| dx, \end{aligned}$$

using (6.10). Moreover we have

$$L(P_R(f)) \leq \frac{1}{|R|} \sum_{(i,j) \in R^2} \delta_i \left( u \mapsto \int f(w_j, u_j^-) F_{X|Y}(dw_j | u_{j+\mathcal{N}_o}) \right)$$

Using (6.10) and setting  $L_j = 0$  if  $j \notin \mathcal{N}_o$ , we have

$$\delta_i \left( u \mapsto \int f(w_j, u_j^-) F_{X|Y}(dw_j | u_{j+\mathcal{N}_o}) \right) \leq \delta_i(f) \mathbb{1}_{i \neq j} + \delta_j(f) L_{i-j}.$$

This yields to the following contraction inequality

$$L(P_R(f)) \leq \left( 1 + \frac{\alpha - 1}{|R|} \right) L(f),$$

with  $\alpha = \sum_{j \in \mathcal{N}_o} L_j < 1$ , from (3.10). The first part of the theorem easily follows. Since the bound obtained for  $d_R(\mu_{T,R}^*, \mu_R)$  does not depend on  $R$ , the second part of theorem follows from the first part, the converge of  $\mu_{R_T}$  to  $\mu$  and the fact that Lipschitzian functions form a convergence-determinating class.  $\square$

**Proof of Proposition 2** The result will easily follow from the following inequality which states that for every  $(\alpha, \tilde{\alpha}) \in \mathbb{R}^2$ :

$$\int_a^b \left| \frac{e^{\alpha z} - e^{\alpha a}}{e^{\alpha b} - e^{\alpha a}} - \frac{e^{\tilde{\alpha} z} - e^{\tilde{\alpha} a}}{e^{\tilde{\alpha} b} - e^{\tilde{\alpha} a}} \right| dz \leq \frac{(b-a)^2}{12} |\alpha - \tilde{\alpha}|. \quad (6.11)$$

Assume that  $\alpha < \tilde{\alpha}$ . First, we observe that the left hand side of (6.11) can be written

$$(b-a) \int_0^1 \left| \frac{e^{\alpha(b-a)z} - 1}{e^{(b-a)\alpha} - 1} - \frac{e^{\tilde{\alpha}(b-a)z} - 1}{e^{\tilde{\alpha}(b-a)} - 1} \right| dz.$$

Using the concavity of the function  $y \rightarrow y^z$  on  $\mathbb{R}_+$ , it is easily seen that the absolute values can be suppressed. Then the left hand side of (6.11) writes

$$(b-a) \left( \frac{1}{(b-a)\alpha} - \frac{1}{e^{(b-a)\alpha} - 1} - \frac{1}{(b-a)\tilde{\alpha}} + \frac{1}{e^{(b-a)\tilde{\alpha}} - 1} \right)$$

and can be bounded by

$$(b-a)^2 |\alpha - \tilde{\alpha}| \sup_{x \in \mathbb{R}} \left| -\frac{1}{x^2} + \frac{1}{4sh^2(x)} \right|.$$

One can show that the odd function  $g : x \rightarrow -\frac{1}{x^2} + \frac{1}{4sh^2(x)}$  is increasing on  $[0, +\infty)$  with  $\lim_{x \rightarrow \infty} g(x) = 0$  and  $\lim_{x \rightarrow 0} g(x) = -\frac{1}{12}$ . Then  $\sup_{x \in \mathbb{R}} |g(x)| = \frac{1}{12}$ . This completes the proof of inequality (6.11).  $\square$

## References

- [1] ABEND, K., HARLEY, T. J. AND KANAL, L. N. (1965) *Classification of binary random patterns* IEEE Trans. Inform. Theory, IT-11, 538-544. [MR0191728](#)
- [2] AILLOT, P., MONBET, V., PREVOSTO, M. (2007) *Survey of stochastic models for wind data and sea state time series*. Probabilistic Engineering Mechanics, **22**, 113-126.
- [3] BESAG, J. (1974) *Spatial interaction and the statistical analysis of lattice systems*. J. R. Statis. Soc., Series B, Vol. 36, 192-236. [MR0373208](#)
- [4] BICKEL, P., LEVINA, E. (2006) *Texture synthesis and nonparametric resampling of random fields*. The Annals of Statistics, Vol. 34, 4, 1751-1773. [MR2283716](#)
- [5] CHAMPAGNAT, F. IDIER, J. GOUSSARD, Y. (1998) *Stationary Markov random fields on a finite rectangular lattice*. Information Theory, IEEE Transactions on, Vol. 44, Issue 7, 2901-2916 [MR1672047](#)



- [6] DACHIAN, S. (1998) *Nonparametric estimation for Gibbs random fields specified through one-point systems*. Statistical Inference for Stochastic Processes, 1, 245-264. [MR2797136](#)
- [7] DIGGLE, P.J., GATES, D.J., STIBBARD, A. (1987) *A nonparametric estimator for pairwise-interaction point process*. Biometrika, 74, 763-770. [MR0919844](#)
- [8] DOBRUSHIN, R., L. (1970) *Prescribing systems of random variables by conditional distributions*. Theory of Probability and its Applications, **15**(3), 458-486.
- [9] DOUKHAN, P. (1994) *Mixing: Properties and Examples*. Lecture Notes in Statist. 85. Springer, New York. [MR1312160](#)
- [10] DOUKHAN, P., LOUHICHI S. (1999) *A new weak dependence condition and applications to moment inequalities*. Stoch. Proc. Appl. 84, 313-342. [MR1719345](#)
- [11] DOUKHAN P., MAYO N., TRUQUET L. (2009) *Weak dependence, models and some applications*. Metrika 69, 2-3, 199-225. [MR2481921](#)
- [12] EFROS, A. A., FREEMAN, W. T. (2001) *Image quilting for texture synthesis and transfer*. In Proc. 28th Annual Conference on Computer Graphics and Interactive Techniques 341-346. ACM Press, New York.
- [13] EFROS, A. A., LEUNG, T. (1999) *Texture synthesis by non-parametric sampling*. In Proc. IEEE International Conference on Computer Vision 2, 1033-1038. IEEE Computer Soc., Washington.
- [14] ETHIER, S.N., KURTZ, T.Z. (1986) *Markov processes: Characterization and convergence*. Wiley, New York. [MR0838085](#)
- [15] FÖLLMER, H. (1988) *Random fields and diffusion processes*. Ecole d'Ete de Saint Flour XV XVII. Lecture Notes in Math. 1362 101 203. Springer, New York. [MR0983373](#)
- [16] GEMAN, S., GEMAN, D. (1984) *Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images*. IEEE Trans. Pattern Anal. Mach. Intell, 6, 721-741, 1984.
- [17] GEORGII, H.-O. (1988) *Gibbs measures and phase transitions*. de Gruyter, Berlin. [MR0956646](#)
- [18] GUYON, X. (1995) *Random fields on a network*. Springer, Berlin. [MR1344683](#)
- [19] HÄRDLE, W., VIEU, P. (1992) *Kernel regression smoothing of time series*. JTSA, Vol 13, No. 3, 209-232. [MR1168165](#)
- [20] KÜNSCH, H.R. (1989) *The jackknife and the bootstrap for general stationary observations*. Annals of Statistics, **17**, 1217-1241. [MR1015147](#)
- [21] KWATRA, V., ESSA, I., BOBICK, A., KWATRA, N. (2005) *Texture optimization for example-based synthesis*. ACM Transactions on Graphics, SIGGRAPH 2005, August 2005.
- [22] LIANG, L., LIU, C., XU, Y., GUO, B., SHUM, H.-Y. (2001) *Real-time texture synthesis by patch-based sampling*. Technical Report MSR-TR-2001-40, Microsoft Research.

- [23] LUETTGEN, M.R., KARL, W.C, WILSKY, A.S, TENNEY, R.R. (1993) *Multiscale Representations of Markov Random fields*. IEEE transactions on signal processing, vol. 41, 12, 3377-3396.
- [24] MEYN S.P., TWEEDIE R. (1993) *Markov Chains and Stochastic Stability*. 3rd edition, Springer, London. [MR1287609](#)
- [25] MONBET V., MARTEAU P.F. (2006) *Local Grid Bootstrap for Stationary Markov Processes*. J. Statistical Planning and Inference 136(10), 3319-3338. [MR2256271](#)
- [26] PAGET, R. (1999) *Nonparametric Markov random field models for natural texture images*. PhD Thesis, University of Queensland.
- [27] PAGET, R., LONGSTAFF, I. D. (1998) *Texture synthesis via a noncausal nonparametric multiscale Markov random field* IEEE Transactions on Image Processing, Vol. 7, 6, 925-931.
- [28] PAPANODITIS, E., POLITIS, D. (2002) *The local bootstrap for Markov processes*. J. Statist. Plann. Inference, 108, 301-328. [MR1947405](#)
- [29] PICKARD, D.K. (1980) *Unilateral Markov fields*. Adv. Appl. Proba., **12**, 655-671. [MR0578842](#)
- [30] POLITIS, D., ROMANO, J. (1993) *Nonparametric resampling for homogeneous strong mixing random fields*. Journal of Multivariate Analysis, **47**, 301-328. [MR1247380](#)
- [31] RAJARSHI, M. (1990) *Bootstrap in Markov sequences based on estimates of transition density*. Ann. Inst. Math. Statist., **42**, 253-268. [MR1064787](#)
- [32] SIMON, B. (1979) *A remark on Dobrushin's uniqueness theorem*. Comm. Math. Phys., **68**, 183-185. [MR0543198](#)