

# CONDITIONAL SAMPLING FOR SPECTRALLY DISCRETE MAX-STABLE RANDOM FIELDS

YIZAO WANG\* \*\* AND

STILIAN A. STOEV,\* \*\*\* *University of Michigan*

## Abstract

Max-stable random fields play a central role in modeling extreme value phenomena. We obtain an explicit formula for the conditional probability in general max-linear models, which include a large class of max-stable random fields. As a consequence, we develop an algorithm for efficient and exact sampling from the conditional distributions. Our method provides a computational solution to the prediction problem for spectrally discrete max-stable random fields. This work offers new tools and a new perspective to many statistical inference problems for spatial extremes, arising, for example, in meteorology, geology, and environmental applications.

*Keywords:* Max-stable process; spectrally discrete representation; regular conditional probability; MARMA process; Smith model; set cover problem

2010 Mathematics Subject Classification: Primary 60G70

Secondary 65C50

## 1. Introduction

### 1.1. Motivation

Max-stable stochastic processes and random fields are fundamental statistical models for the dependence of extremes. This is because they arise in the limit of rescaled maxima. Indeed, consider the componentwise maxima

$$M_t^{(n)} = \max_{j=1, \dots, n} \xi_t^{(j)}, \quad t \in T,$$

of independent realizations  $\{\xi_t^{(j)}\}_{t \in T, j = 1, \dots, n}$ , of a random field  $\xi = \{\xi_t\}_{t \in T}$ . If the random field  $\{M_t^{(n)}\}_{t \in T}$  converges in law, as  $n \rightarrow \infty$ , under judicious normalization, then its limit  $X = \{X_t\}_{t \in T}$  is necessarily *max-stable* (see, e.g. [11] and [19]).

Max-stable processes (including random fields) are therefore as important to extreme value theory as Gaussian processes are to the classical statistical theory based on the central limit theorem. The multivariate max-stable laws and processes have been studied extensively in the past 30 years. See, e.g. [1], [6], [7], [9], [10], [13], [15], [16], [21], [24], [26], and [28], among many others.

The modeling and parameter estimation of the univariate marginal distributions of the extremes have been studied extensively (see, e.g. [8], [11], [20], and the references therein). Many of the recent developments in this domain focus on the characterization, modeling, and

Received 17 August 2010; revision received 24 November 2010.

\* Postal address: Department of Statistics, University of Michigan, 439 West Hall, 1085 South University, Ann Arbor, MI 48109-1107, USA.

\*\* Email address: yizwang@umich.edu

\*\*\* Email address: sstoev@umich.edu

estimation of the dependence for multivariate extremes. In this context, building adequate max-stable processes and random fields plays a key role. See, e.g. [2], [4], [12], [17], [22], and [23].

Our present work is motivated by an important and long-standing challenge, namely, the prediction for max-stable random processes and fields. Suppose that we already have a suitable max-stable model for the dependence structure of a random field  $\{X_t\}_{t \in T}$ . The field is observed at several locations  $t_1, \dots, t_n \in T$  and we want to *predict* the values of the field  $X_{s_1}, \dots, X_{s_m}$  at some other locations. The optimal predictors involve the *conditional distribution* of  $\{X_t\}_{t \in T}$ , given the data. Even if the finite-dimensional distributions of the field  $\{X_t\}_{t \in T}$  are available in analytic form, it is typically impossible to obtain a closed-form solution for the conditional distribution. Naïve Monte Carlo approximations are not practical either, since they involve conditioning on events of infinitesimal probability, which leads to mounting errors and computational costs.

Prior studies by Davis and Resnick [6], [7] and Cooley *et al.* [4], among others, have shown that the prediction problem in the max-stable context is challenging, and it does not have an elegant analytical solution. On the other hand, the growing popularity and the use of max-stable processes in various applications, make this an important problem. This motivated us to seek a *computational* solution.

In this work we develop theory and methodology for sampling from the conditional distributions of spectrally discrete max-stable models. More precisely, we provide an algorithm that can generate efficiently *exact* independent samples from the *regular conditional probability* of  $(X_{s_1}, \dots, X_{s_m})$ , given the values  $(X_{t_1}, \dots, X_{t_n})$ . For the sake of simplicity, we write  $\mathbf{X} = (X_1, \dots, X_n) \equiv (X_{t_1}, \dots, X_{t_n})$ . The algorithm applies to the general *max-linear model*:

$$X_i = \max_{j=1, \dots, p} a_{i,j} Z_j \equiv \bigvee_{j=1}^p a_{i,j} Z_j, \quad i = 1, \dots, n. \quad (1)$$

Here the  $a_{i,j}$ s are known nonnegative constants and the  $Z_j$ s are independent continuous nonnegative random variables. Any multivariate max-stable distribution can be approximated arbitrarily well via a max-linear model with sufficiently large  $p$ .

The main idea is to first generate samples from the *regular conditional probability* distribution of  $\mathbf{Z} \mid \mathbf{X} = \mathbf{x}$ , where  $\mathbf{Z} = (Z_j)_{j=1, \dots, p}$ . Then, the conditional distributions of

$$X_{s_k} = \bigvee_{j=1}^p b_{k,j} Z_j, \quad 1 \leq k \leq m,$$

given  $\mathbf{X} = \mathbf{x}$ , can be readily obtained, for any given  $b_{k,j}$ s. In this paper we assume that the model is completely known, i.e. the parameters  $\{a_{i,j}\}$  and  $\{b_{k,j}\}$  are given. The statistical inference for these parameters is beyond the scope of this paper.

Observe that if  $\mathbf{X} = \mathbf{x}$  then (1) implies natural equality and inequality constraints on the  $Z_j$ s. More precisely, (1) gives rise to a set of so-called *hitting scenarios*. In each hitting scenario, a subset of the  $Z_j$ s equal, in other words *hit*, their upper bounds and the rest of the  $Z_j$ s can take arbitrary values in certain open intervals. We will show that the regular conditional probability of  $\mathbf{Z} \mid \mathbf{X} = \mathbf{x}$  is a weighted mixture of the various distributions of the vector  $\mathbf{Z}$ , under all possible hitting scenarios corresponding to  $\mathbf{X} = \mathbf{x}$ .

The resulting formula, however, involves determining *all* hitting scenarios, which becomes computationally prohibitive for large and even moderate values of  $p$ . This issue is closely related to the NP-hard *set-covering problem* in computer science (see, e.g. [3]).

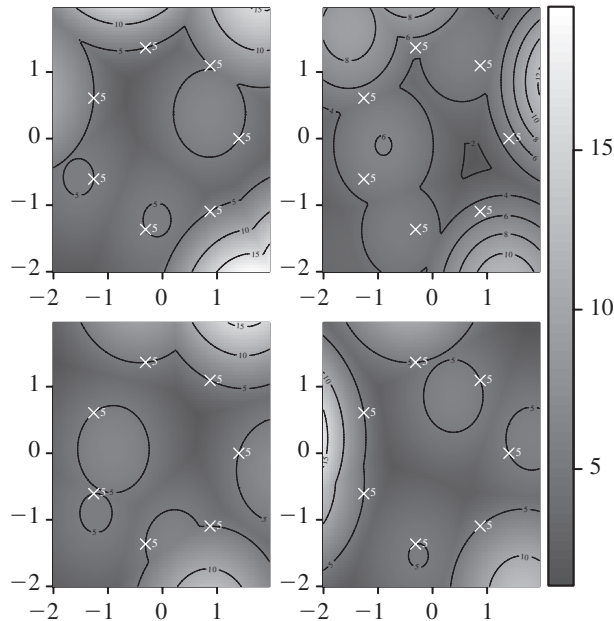


FIGURE 1: Four samples from the conditional distribution of the discrete Smith model (see Section 3.2), given the *observed* values (all equal to 5) at the locations marked by crosses. The values of the parameters are  $\rho = 0$ ,  $\beta_1 = 1$ , and  $\beta_2 = 1$ .

Fortunately, further detailed analysis of the probabilistic structure of the max-linear models allows us to obtain a different formula of the regular conditional probability (Theorem 2). It yields an exact and computationally efficient algorithm, which in practice can handle complex max-linear models with  $p$  in the order of thousands, on a conventional desktop computer. The algorithm is implemented in the R (see [18]) package `maxLinear` [27], with the core part written in C/C++. We also used the R package `fields` (see [14]) to generate some of the figures in this paper.

We illustrate the performance of our algorithm over two classes of processes: the max-autoregressive moving average (MARMA) time series (see [6]), and the Smith model (see [24]) for spatial extremes. The MARMA processes are spectrally discrete max-stable processes, and our algorithm applies directly. In Section 3.1, we demonstrate the prediction of MARMA processes by conditional sampling and compare our result to the projection predictors proposed in [6]. To apply our algorithm to the Smith model, on the other hand, we first need to discretize the (spectrally continuous) model. Section 3.2 is devoted to conditional sampling for the discretized Smith model. Thanks to the computational efficiency of our algorithm, we can choose a mesh fine enough to obtain a satisfactory discretization. Figure 1 shows four realizations from such a discretized Smith model, conditioning only on seven observations (with assumed value 5). The algorithm applies in the same way to more complex models.

### 1.2. Multivariate max-stable distributions: a brief review

Consider a general max-stable process  $X = \{X_t\}_{t \in T}$ , indexed by a set  $T$  (e.g.  $T = [0, 1]$ ,  $\mathbb{R}$ ,  $\mathbb{R}^d$ , or  $\mathbb{Z}^d$ ). We will assume that the finite-dimensional distributions of  $X$  are known

and the ultimate goal is to study the conditional distributions of  $X$ . For convenience and without loss of generality, we focus on max-stable processes  $X$  with  $\alpha$ -Fréchet marginals ( $\alpha > 0$ ) such that all max-linear combinations

$$\xi = \max_{j=1, \dots, n} a_j X_{t_j} \equiv \bigvee_{i=1}^n a_i X_{t_i}, \quad a_j > 0, t_j \in T,$$

have the  $\alpha$ -Fréchet distribution:

$$P(\xi \leq x) = \exp\{-\sigma_\xi^\alpha x^{-\alpha}\}, \quad x \in (0, \infty),$$

with scale coefficient  $\sigma_\xi > 0$ . Any max-stable process can be related to such an  $\alpha$ -Fréchet process by simple transformation of the marginals (see, e.g. [19, Chapter 5]).

Essentially all max-stable processes  $\{X_t\}_{t \in T}$  admit the following *extremal integral representation*:

$$\{X_t\}_{t \in T} \stackrel{D}{=} \left\{ \int_S f_t(s) M_\alpha(ds) \right\}_{t \in T}.$$

Here the  $f_t$ s are nonnegative, measurable deterministic functions defined on a suitable measure space  $(S, \mu)$  and such that  $\int_S f_t^\alpha(s) \mu(ds) < \infty$ , and  $M_\alpha$  is an  $\alpha$ -Fréchet random sup-measure with control measure  $\mu$ . The extremal integral  $\int_S f dM_\alpha$  can be defined for all  $f \in L^\alpha(S, \mu)$ ,  $f \geq 0$ , as the limit in probability of extremal integrals of simple functions. For more details, see [26] and the seminal work [10] for an alternative treatment.

The functions  $\{f_t\}_{t \in T}$  are called the *spectral functions* of the process  $\{X_t\}_{t \in T}$ . They determine the finite-dimensional distributions of  $\{X_t\}_{t \in T}$ , i.e.

$$P(X_{t_1} \leq x_1, \dots, X_{t_n} \leq x_n) = \exp\left\{- \int_S \left(\bigvee_{i=1}^n \frac{f_i(s)}{x_i}\right)^\alpha \mu(ds)\right\}$$

for all  $t_i \in T$  and  $x_i \in \mathbb{R}_+ := (0, \infty)$ ,  $i = 1, \dots, n$ . A popular equivalent representation of multivariate max-stable laws is as follows:

$$P(X_{t_1} \leq x_1, \dots, X_{t_n} \leq x_n) = \exp\left\{- \int_{\mathbb{S}_+^{n-1}} \left(\bigvee_{i=1}^n \frac{w_i}{x_i}\right)^\alpha \Gamma(d\mathbf{w})\right\}.$$

Here  $\mathbb{S}_+^{n-1} = \{\mathbf{w} = (w_i)_{i=1}^n \in \mathbb{R}^n : 0 \leq w_j \leq \max_{i=1, \dots, n} w_i = 1\}$  is the positive unit sphere in the sup-norm, and  $\Gamma$  is a unique finite measure on  $\mathbb{S}_+^{n-1}$  called the *spectral measure* of the distribution (see, e.g. [11, Chapter 6] and [19, Chapter 5]).

Any multivariate max-stable vector  $(X_{t_j})_{j=1}^n$  can be approximated arbitrarily well in probability, by discretizing the extremal integral, i.e.

$$X_{t_i} = \int_S f_{t_i}(s) M_\alpha(ds) \approx \bigvee_{i=1}^p a_{i,j} Z_j,$$

where the  $Z_j$ ,  $j = 1, \dots, p$ , are independent standard  $\alpha$ -Fréchet variables and  $a_{i,j} \geq 0$ . This is equivalent to considering multivariate max-stable vectors with discrete spectral measures concentrated on at most  $p$  points on the unit sphere  $\mathbb{S}_+^{n-1}$ . The error of approximation, moreover, can be controlled explicitly through convenient probability metrics (see, e.g. [26]).

In this paper we will focus on the class of max-stable processes, i.e.

$$X_t := \bigvee_{j=1}^p \phi_j(t)Z_j, \quad t \in T,$$

where the  $\phi_j(t)$ s are nonnegative deterministic functions. These processes are called *spectrally discrete*, since their spectral measures  $\Gamma$  are discrete. By taking sufficiently large  $p$ s and with judicious  $\phi_j(t)$ s, we can build flexible models that can replicate the behavior of an arbitrary max-stable process. From this point of view, a satisfactory computational solution must be able to deal with max-linear models with large  $p$ s.

The treatment of the exact conditional distributions of general spectrally continuous max-stable processes requires different tools and still remains an open problem, to the best of the authors' knowledge. As we will see, the solution in the discrete case, although complete, is already quite involved.

## 2. Conditional probability in max-linear models

### 2.1. Intuition and basic theory

Consider the max-linear model in (1). We will denote this model by

$$\mathbf{X} = \mathbf{A} \odot \mathbf{Z}, \tag{2}$$

where  $\mathbf{A} = (a_{i,j})_{n \times p}$  is a matrix with nonnegative entries, and  $\mathbf{X} = (X_1, \dots, X_n)$  and  $\mathbf{Z} = (Z_1, \dots, Z_p)$  are column vectors. We assume that the  $Z_j$ s,  $j = 1, \dots, p$ , are independent nonnegative random variables having probability densities.

In this section we provide an explicit formula for the *regular conditional probability* of  $\mathbf{Z}$  with respect to  $\mathbf{X}$  (see Theorem 1 below and Appendix A for a precise definition). We start with some intuition and notation. Throughout this paper, we assume that the matrix  $\mathbf{A}$  has at least one nonzero entry in each of its rows and columns. This will be referred to as *Assumption A*.

Observe that if  $\mathbf{x} = \mathbf{A} \odot \mathbf{z}$  with  $\mathbf{x} \in \mathbb{R}_+^n$  and  $\mathbf{z} \in \mathbb{R}_+^p$ , then

$$0 \leq z_j \leq \widehat{z}_j \equiv \widehat{z}_j(\mathbf{A}, \mathbf{x}) := \min_{1 \leq i \leq n} \frac{x_i}{a_{i,j}}, \quad j = 1, \dots, p. \tag{3}$$

That is, the max-linear model (2) imposes certain inequality and equality constraints on the  $Z_j$ s, given a set of *observed*  $X_i$ s. Namely, some of the upper bounds  $\widehat{z}_j(\mathbf{A}, \mathbf{x})$  in (3) must be attained, or *hit*, i.e.  $z_j = \widehat{z}_j(\mathbf{A}, \mathbf{x})$  in such a way that

$$x_i = a_{i,j(i)}z_{j(i)}, \quad i = 1, \dots, n,$$

with judicious  $j(i) \in \{1, \dots, p\}$ . The next example helps to explain the inequality and equality constraints.

**Example 1.** Suppose that  $n = p = 3$  and

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}.$$

Let  $\mathbf{x} = \mathbf{A} \odot \mathbf{z}$  for some  $\mathbf{z} \in \mathbb{R}_+^3$ . In this case, it necessarily follows that  $x_1 \leq x_2 \leq x_3$ . Moreover, (3) yields  $\widehat{\mathbf{z}} = \mathbf{x}$ .

- (i) If  $\mathbf{x} = (1, 2, 3)$  then it trivially follows that  $\mathbf{z} = \widehat{\mathbf{z}} = (1, 2, 3)$ , which is an equality constraint on  $\mathbf{z}$ .

- (ii) If  $\mathbf{x} = (1, 1, 3)$  then it follows that  $z_1 = \widehat{z}_1 = 1$ ,  $z_2 \leq \widehat{z}_2 = 1$ , and  $z_3 = \widehat{z}_3 = 3$ . Here, the ‘equality constraints’ must hold for  $z_1 = \widehat{z}_1$  and  $z_3 = \widehat{z}_3$ , while  $z_2$  needs to only satisfy the ‘inequality constraint’  $0 \leq z_2 \leq \widehat{z}_2$ .

Write

$$C(\mathbf{A}, \mathbf{x}) := \{\mathbf{z} \in \mathbb{R}_+^p : \mathbf{x} = \mathbf{A} \odot \mathbf{z}\},$$

and note that the conditional distribution of  $\mathbf{Z} \mid \mathbf{X} = \mathbf{x}$  concentrates on the set  $C(\mathbf{A}, \mathbf{x})$ . The observation in Example 1 can be generalized and formulated as follows.

- Every  $\mathbf{z} \in C(\mathbf{A}, \mathbf{x})$  corresponds to a set of *active* (equality) constraints  $J \subset \{1, \dots, p\}$ , which we refer to as a *hitting scenario* of  $(\mathbf{A}, \mathbf{x})$ , such that

$$z_j = \widehat{z}_j(\mathbf{A}, \mathbf{x}) \quad \text{for } j \in J \quad \text{and} \quad z_j < \widehat{z}_j(\mathbf{A}, \mathbf{x}) \quad \text{for } j \in J^c := \{1, \dots, p\} \setminus J. \quad (4)$$

Observe that if  $j \notin J$  then there are no further constraints and  $z_j$  can take any value in  $[0, \widehat{z}_j]$ , regardless of the values of the other components of the vector  $\mathbf{z} \in C(\mathbf{A}, \mathbf{x})$ .

- Every value  $\mathbf{x}$  may give rise to many different hitting scenarios  $J \subset \{1, \dots, p\}$ . Let  $\mathcal{J}(\mathbf{A}, \mathbf{x})$  denote the collection of all such  $J$ s. We refer to  $\mathcal{J}(\mathbf{A}, \mathbf{x})$  as to the *hitting distribution* of  $\mathbf{x}$  with respect to  $\mathbf{A}$ :

$$\mathcal{J}(\mathbf{A}, \mathbf{x}) \equiv \{J \subset \{1, \dots, p\} : \text{exist } \mathbf{z} \in C(\mathbf{A}, \mathbf{x}), \text{ such that (4) holds}\}.$$

To illustrate the notions of hitting scenario and hitting distribution, consider again Example 1. Therein, we have  $\mathcal{J}(\mathbf{A}, \mathbf{x}) = \{\{1, 2, 3\}\}$  in case (i), and  $\mathcal{J}(\mathbf{A}, \mathbf{x}) = \{\{1, 3\}, \{1, 2, 3\}\}$  in case (ii).

The hitting distribution  $\mathcal{J}(\mathbf{A}, \mathbf{x})$  is a finite set and, thus, can always be identified. However, the identification procedure is the key difficulty in providing an efficient algorithm for conditional sampling in practice. This issue is addressed in Section 2.2. In the rest of this section, suppose that  $\mathcal{J}(\mathbf{A}, \mathbf{x})$  is given. Then, we can partition  $C(\mathbf{A}, \mathbf{x})$  as

$$C(\mathbf{A}, \mathbf{x}) = \bigcup_{J \in \mathcal{J}(\mathbf{A}, \mathbf{x})} C_J(\mathbf{A}, \mathbf{x}),$$

where

$$C_J(\mathbf{A}, \mathbf{x}) = \{\mathbf{z} \in \mathbb{R}_+^p : z_j = \widehat{z}_j, j \in J, \text{ and } z_j < \widehat{z}_j, j \notin J\}.$$

The sets  $C_J(\mathbf{A}, \mathbf{x})$  and  $J \in \mathcal{J}(\mathbf{A}, \mathbf{x})$  are disjoint since they correspond to different *hitting scenarios* in  $\mathcal{J}(\mathbf{A}, \mathbf{x})$ . Let

$$r(\mathcal{J}(\mathbf{A}, \mathbf{x})) = \min_{J \in \mathcal{J}(\mathbf{A}, \mathbf{x})} |J|,$$

where  $|J|$  is the number of elements in  $J$ . We call  $r(\mathcal{J}(\mathbf{A}, \mathbf{x}))$  the *rank* of the hitting distribution  $\mathcal{J}(\mathbf{A}, \mathbf{x})$ . It equals the minimal number of equality constraints among the hitting scenarios in  $\mathcal{J}(\mathbf{A}, \mathbf{x})$ . It will turn out that the hitting scenarios  $J \subset \mathcal{J}(\mathbf{A}, \mathbf{x})$  with  $|J| > r(\mathcal{J}(\mathbf{A}, \mathbf{x}))$  occur with (conditional) probability 0 and can be ignored. We therefore focus on the set of all *relevant* hitting scenarios:

$$\mathcal{J}_r(\mathbf{A}, \mathbf{x}) = \{J \in \mathcal{J}(\mathbf{A}, \mathbf{x}) : |J| = r(\mathcal{J}(\mathbf{A}, \mathbf{x}))\}.$$

**Theorem 1.** Consider the max-linear model in (2), where the  $Z_j$ s are independent random variables with densities  $f_{Z_j}$  and distribution functions  $F_{Z_j}$ ,  $j = 1, \dots, p$ . Let  $\mathbf{A} = (a_{i,j})_{n \times p}$  have nonnegative entries satisfying Assumption A, and let  $\mathcal{R}_{\mathbb{R}_+^p}$  be the class of all rectangles  $\{(\mathbf{e}, \mathbf{f}], \mathbf{e}, \mathbf{f} \in \mathbb{R}_+^p\}$  in  $\mathbb{R}_+^p$ .

For all  $J \in \mathcal{J}(\mathbf{A}, \mathbf{x})$ ,  $E \in \mathcal{R}_{\mathbb{R}_+^p}$ , and  $\mathbf{x} \in \mathbb{R}_+^p$ , define

$$v_J(\mathbf{x}, E) := \prod_{j \in J} \delta_{\widehat{z}_j}(\pi_j(E)) \prod_{j \in J^c} P(Z_j \in \pi_j(E) \mid Z_j < \widehat{z}_j), \tag{5}$$

where  $\pi_j(z_1, \dots, z_p) = z_j$  and  $\delta_a$  is a unit point mass at  $a$ .

Then, the regular conditional probability  $v(\mathbf{x}, E)$  of  $\mathbf{Z}$  with respect to  $\mathbf{X}$  equals

$$v(\mathbf{x}, E) = \sum_{J \in \mathcal{J}_r(\mathbf{A}, \mathbf{x})} p_J(\mathbf{A}, \mathbf{x}) v_J(\mathbf{x}, E), \quad E \in \mathcal{R}_{\mathbb{R}_+^p}, \tag{6}$$

for  $P^{\mathbf{X}}$ -almost all  $\mathbf{x} \in \mathbf{A} \odot (\mathbb{R}_+^p)$ , where, for all  $J \in \mathcal{J}_r(\mathbf{A}, \mathbf{x})$ ,

$$p_J(\mathbf{A}, \mathbf{x}) = \frac{w_J}{\sum_{K \in \mathcal{J}_r(\mathbf{A}, \mathbf{x})} w_K} \quad \text{with} \quad w_J = \prod_{j \in J} \widehat{z}_j f_{Z_j}(\widehat{z}_j) \prod_{j \in J^c} F_{Z_j}(\widehat{z}_j). \tag{7}$$

In the special case when the  $Z_j$ s are  $\alpha$ -Fréchet with scale coefficient 1, we have  $w_J = \prod_{j \in J} (\widehat{z}_j)^{-\alpha}$ .

**Remark 1.** We state (6) *only* for rectangle sets  $E$  because the projections  $\pi_j(B)$  of an arbitrary Borel set  $B \subset \mathbb{R}_+^p$  are not always Borel (see, e.g. [25, Chapter 4]). Nevertheless, the extension of measure theorem ensures that (6) specifies completely the regular conditional probability.

We do not provide a proof of Theorem 1 directly. Instead, we will first provide an equivalent formula for  $v(\mathbf{x}, E)$  in Theorem 2 in Section 2.2, and then prove that  $v(\mathbf{x}, E)$  is the desired regular conditional probability. All the proofs are deferred to Section 4. The next example gives the intuition behind (6).

**Example 2.** (Example 1, continued.) (i) If  $\mathbf{X} = \mathbf{x} = (1, 2, 3)$  then  $\widehat{\mathbf{z}} = \mathbf{x}$  and  $\mathcal{J}(\mathbf{A}, \mathbf{x}) = \{\{1, 2, 3\}\}$ . Therefore,  $r(\mathcal{J}(\mathbf{A}, \mathbf{x})) = 3$  and (6) yields

$$v(\mathbf{x}, E) = v_J(\mathbf{x}, E) = \delta_{\widehat{z}_1}(\pi_1(E)) \delta_{\widehat{z}_2}(\pi_2(E)) \delta_{\widehat{z}_3}(\pi_3(E)) \equiv \delta_{\widehat{\mathbf{z}}}(E),$$

a degenerate distribution with single unit point mass at  $\widehat{\mathbf{z}}$ .

(ii) If  $\mathbf{X} = \mathbf{x} = (1, 1, 3)$  then  $\widehat{\mathbf{z}} = \mathbf{x}$ ,  $\mathcal{J}(\mathbf{A}, \mathbf{x}) = \{\{1, 3\}, \{1, 2, 3\}\}$ , and  $r(\mathcal{J}(\mathbf{A}, \mathbf{x})) = 2$ . Therefore,  $\mathcal{J}_r(\mathbf{A}, \mathbf{x}) = \{\{1, 3\}\}$  and (6) yields

$$v(\mathbf{x}, E) = v_{\{1,3\}}(\mathbf{x}, E) = \delta_{\widehat{z}_1}(\pi_1(E)) P(Z_2 \in \pi_2(E) \mid Z_2 < \widehat{z}_2) \delta_{\widehat{z}_3}(\pi_3(E)).$$

In this case, the conditional distribution concentrates on the one-dimensional set  $\{1\} \times (0, 1) \times \{3\}$ .

(iii) Finally, if  $\mathbf{X} = \mathbf{x} = (1, 1, 1)$  then  $\widehat{\mathbf{z}} = \mathbf{x}$  and  $\mathcal{J}(\mathbf{A}, \mathbf{x}) = \{\{1\}, \{1, 2\}, \{1, 2, 3\}\}$ . Then,  $\mathcal{J}_r(\mathbf{A}, \mathbf{x}) = \{\{1\}\}$  and

$$v(\mathbf{x}, E) = v_{\{1\}}(\mathbf{x}, E) = \delta_{\widehat{z}_1}(\pi_1(E)) \prod_{j=2}^3 P(Z_j \in \pi_j(E) \mid Z_j < \widehat{z}_j).$$

The conditional distribution concentrates on the set  $\{1\} \times (0, 1) \times (0, 1)$ .

We conclude this section by showing that the conditional distributions (6) arise as suitable limits. This result can be viewed as a heuristic justification of Theorem 1. Let  $\varepsilon > 0$ , consider

$$C_j^\varepsilon(\mathbf{A}, \mathbf{x}) := \{\mathbf{z} \in \mathbb{R}_+^p : z_j \in [\widehat{z}_j(1 - \varepsilon), \widehat{z}_j(1 + \varepsilon)], j \in J, z_k < \widehat{z}_k(1 - \varepsilon), k \in J^c\}, \tag{8}$$

and set

$$C^\varepsilon(\mathbf{A}, \mathbf{x}) := \bigcup_{J \in \mathcal{J}(\mathbf{A}, \mathbf{x})} C_J^\varepsilon(\mathbf{A}, \mathbf{x}). \tag{9}$$

Note that the sets  $\mathbf{A} \odot (C^\varepsilon(\mathbf{A}, \mathbf{x}))$  shrink to the point  $\mathbf{x}$ , as  $\varepsilon \downarrow 0$ .

**Proposition 1.** *Under the assumptions of Theorem 1, for all  $\mathbf{x} \in \mathbf{A} \odot (\mathbb{R}_+^p)$ , we have, as  $\varepsilon \downarrow 0$ ,*

$$P(\mathbf{Z} \in E \mid \mathbf{Z} \in C^\varepsilon(\mathbf{A}, \mathbf{x})) \rightarrow \nu(\mathbf{x}, E), \quad E \in \mathcal{R}_{\mathbb{R}_+^p}.$$

*Proof.* Recall the definition of  $C_J^\varepsilon$  in (8). Note that, for all  $\varepsilon > 0$ , the sets  $\{C_J^\varepsilon(\mathbf{A}, \mathbf{x})\}_{J \in \mathcal{J}(\mathbf{A}, \mathbf{x})}$  are mutually disjoint. Thus, writing  $C^\varepsilon \equiv C^\varepsilon(\mathbf{A}, \mathbf{x})$  and  $C_J^\varepsilon \equiv C_J^\varepsilon(\mathbf{A}, \mathbf{x})$ , by (9) we have

$$\begin{aligned} P(\mathbf{Z} \in E \mid \mathbf{Z} \in C^\varepsilon) &= \sum_{J \in \mathcal{J}} P(\mathbf{Z} \in E \mid \mathbf{Z} \in C_J^\varepsilon) P(\mathbf{Z} \in C_J^\varepsilon \mid \mathbf{Z} \in C^\varepsilon) \\ &= \sum_{J \in \mathcal{J}} P(\mathbf{Z} \in E \mid \mathbf{Z} \in C_J^\varepsilon) \frac{P(\mathbf{Z} \in C_J^\varepsilon)}{\sum_{K \in \mathcal{J}} P(\mathbf{Z} \in C_K^\varepsilon)}, \end{aligned}$$

where the terms with  $P(\mathbf{Z} \in C_J^\varepsilon) = 0$  are ignored. We can see that  $P(\mathbf{Z} \in E \mid \mathbf{Z} \in C_J^\varepsilon)$  converges to  $\nu_J(E, \mathbf{x})$  in (5), as  $\varepsilon \downarrow 0$ . The independence of the  $Z_j$ s also implies that

$$\begin{aligned} P(\mathbf{Z} \in C_J^\varepsilon) &= \prod_{j \in J} P(Z_j \in [\widehat{z}_j(1 - \varepsilon), \widehat{z}_j(1 + \varepsilon)]) \prod_{k \in J^c} P(Z_k \leq \widehat{z}_k(1 - \varepsilon)) \\ &= \prod_{j \in J} (f_{Z_j}(\widehat{z}_j) \widehat{z}_j \cdot 2\varepsilon + o(\varepsilon)) \prod_{k \in J^c} (F_{Z_k}(\widehat{z}_k) + o(\varepsilon)). \end{aligned} \tag{10}$$

Observe that, for  $J \in \mathcal{J}_r(\mathbf{A}, \mathbf{x})$ , the latter expression equals  $2w_J \varepsilon^{|J|} (1 + o(1))$ ,  $\varepsilon \downarrow 0$ , and the terms with  $|J| > r$  will become negligible since they are of smaller order. Therefore, (10) yields (6), and the proof is thus complete.

The proof of Proposition 1 provides an insight to the expressions of the weights  $w_J$  in (7) and the components  $\nu_J$  in (5). In particular, it explains why only hitting scenarios of rank  $r$  are involved in the expression of the conditional probability. The formal proof of Theorem 1, however, requires a different argument.

**2.2. Conditional sampling: computational efficiency**

We discuss here important computational issues related to sampling from the regular conditional probability in (6). It turns out that identifying all hitting scenarios amounts to solving the *set covering problem*, which is NP-hard (see, e.g. [3]). The probabilistic structure of the max-linear models, however, will lead us to an alternative efficient solution, valid with probability 1. In particular, we will provide a new formula for the regular conditional probability, showing that  $\mathbf{Z}$  can be decomposed into conditionally independent vectors, given  $\mathbf{X} = \mathbf{x}$ . As a consequence, with probability 1, we are not in the ‘bad’ situation that the corresponding set covering problem requires exponential time to solve. Indeed, this will lead us to an efficient and linearly scalable algorithm for conditional sampling, which works well for max-linear models with large dimensions  $n \times p$  arising in applications.

To fix ideas, observe that Theorem 1 implies the following simple algorithm.



**Algorithm I.**

1. Compute  $\widehat{z}_j$  for  $j = 1, \dots, p$ .
2. Identify  $\mathcal{J}(\mathbf{A}, \mathbf{x})$ , compute  $r = r(\mathcal{J}(\mathbf{A}, \mathbf{x}))$ , and focus on the set of *relevant hitting scenarios*  $\mathcal{J}_r = \mathcal{J}_r(\mathbf{A}, \mathbf{x})$ .
3. Compute  $\{w_J\}_{J \in \mathcal{J}_r}$  and  $\{p_J\}_{J \in \mathcal{J}_r}$ .
4. Sample  $\mathbf{Z} \sim \nu(\mathbf{x}, \cdot)$  according to (6).

Step 1 is immediate. Provided that step 2 is done, step 3 is trivial, and step 4 can be carried out by first picking a *hitting scenario*  $J \in \mathcal{J}_r(\mathbf{A}, \mathbf{x})$  (with probability  $p_J(\mathbf{A}, \mathbf{x})$ ), setting  $Z_j = \widehat{z}_j$  for  $j \in J$ , and then resampling independently the remaining  $Z_j$ s from the truncated distributions:  $Z_j \mid \{Z_j < \widehat{z}_j\}$  for all  $j \in \{1, \dots, p\} \setminus J$ .

The most computationally intensive aspect of this algorithm is to identify the set of all relevant hitting scenarios  $\mathcal{J}_r(\mathbf{A}, \mathbf{x})$  in step 2. This is closely related to the NP-hard *set covering problem* in theoretical computer science (see, e.g. [3]), which is formulated next. Let  $\mathbf{H} = (h_{i,j})_{n \times p}$  be a matrix of 0s and 1s, and let  $\mathbf{c} = (c_j)_{j=1}^p \in \mathbb{Z}_+^p$  be a  $p$ -dimensional *cost vector*. For simplicity, introduce the notation:

$$\langle m \rangle \equiv \{1, 2, \dots, m\}, \quad m \in \mathbb{N}.$$

For the matrix  $\mathbf{H}$ , we say that the column  $j \in \langle p \rangle$  covers the row  $i \in \langle n \rangle$  if  $h_{i,j} = 1$ . The goal of the set-covering problem is to find a minimum-cost subset  $J \subset \langle p \rangle$  such that every row is covered by at least one column  $j \in J$ . This is equivalent to solving

$$\min_{\substack{\delta_j \in \{0,1\} \\ j \in \langle p \rangle}} \sum_{j \in \langle p \rangle} c_j \delta_j \quad \text{subject to} \quad \sum_{j \in \langle p \rangle} h_{i,j} \delta_j \geq 1, \quad i \in \langle n \rangle. \tag{11}$$

We can relate the problem of identifying  $\mathcal{J}_r(\mathbf{A}, \mathbf{x})$  to the set covering problem by defining

$$h_{i,j} = \mathbf{1}\{a_{i,j} \widehat{z}_j = x_i\}, \tag{12}$$

where  $\mathbf{A} = (a_{i,j})_{n \times p}$  and  $\mathbf{x} = (x_i)_{i=1}^n$  are as in (2), and  $c_j = 1, j \in \langle p \rangle$ . It is easy to see that every  $J \in \mathcal{J}_r(\mathbf{A}, \mathbf{x})$  corresponds to a solution of (11), and vice versa. Namely, for  $\{\delta_j\}_{j \in \langle p \rangle}$  minimizing (11), we have  $J = \{j \in \langle p \rangle : \delta_j = 1\} \in \mathcal{J}_r(\mathbf{A}, \mathbf{x})$ .

The set  $\mathcal{J}_r(\mathbf{A}, \mathbf{x})$  corresponds to the set of *all* solutions of (11), which depends only on the matrix  $\mathbf{H}$ . Therefore, in the sequel we write  $\mathcal{J}_r(\mathbf{H})$  for  $\mathcal{J}_r(\mathbf{A}, \mathbf{x})$ , and

$$\mathbf{H} = (h_{i,j})_{n \times p} \equiv \mathbb{H}(\mathbf{A}, \mathbf{x}), \tag{13}$$

where  $h_{i,j}$  as in (12) will be referred to as the *hitting matrix*.

**Example 3.** Recall Example 2. The following hitting matrices correspond to the three cases of  $\mathbf{x}$  discussed therein:

$$\mathbf{H}^{(i)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{H}^{(ii)} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \text{and} \quad \mathbf{H}^{(iii)} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}.$$

Observe that solving for  $\mathcal{J}_r(\mathbf{H})$  is even more challenging than solving the set covering problem (11), where only one minimum-cost subset  $J$  is needed, and often an approximation of the optimal solution is acceptable. Here, we need to identify exhaustively all  $J$ s such that (11)

holds. Fortunately, this problem can be substantially simplified, thanks to the probabilistic structure of the max-linear model.

We first study the distribution of  $\mathbf{H}$ . In view of (13), it follows that  $\mathbf{H} = \mathbb{H}(\mathbf{A}, \mathbf{X})$ , with  $\mathbf{X} = \mathbf{A} \odot \mathbf{Z}$ , is a random matrix. It will turn out that, with probability 1,  $\mathbf{H}$  has a nice structure, leading to an efficient conditional sampling algorithm.

For any hitting matrix  $\mathbf{H}$ , we will decompose the set  $\langle p \rangle \equiv \{1, \dots, p\}$  into a certain disjoint union  $\langle p \rangle = \bigcup_{s=1}^r \bar{J}^{(s)}$ . The vectors  $(Z_j)_{j \in \bar{J}^{(s)}}$ ,  $s = 1, \dots, r$ , will turn out to be conditionally independent (in  $s$ ), given  $\mathbf{X} = \mathbf{x}$ . Therefore,  $\nu(\mathbf{x}, E)$  will be expressed as a product of (conditional) probabilities.

We start by decomposing the set  $\langle n \rangle \equiv \{1, \dots, n\}$ . First, for all  $i_1, i_2 \in \langle n \rangle$  and  $j \in \langle p \rangle$ , we write  $i_1 \overset{j}{\sim} i_2$  if  $h_{i_1, j} = h_{i_2, j} = 1$ . Then, we define an equivalence relation on  $\langle n \rangle$ , i.e.

$$i_1 \sim i_2 \quad \text{if } i_1 = \overset{j_1}{\sim} i_0 \overset{j_2}{\sim} \dots \overset{j_m}{\sim} i_m = i_2, \tag{14}$$

with some  $m \leq n$ ,  $i_1 = \overset{j_1}{\sim} i_0, \dots, \overset{j_m}{\sim} i_m = i_2 \in \langle n \rangle$ ,  $j_1, \dots, j_m \in \langle p \rangle$ . That is, ‘ $\sim$ ’ is the transitive closure of ‘ $\overset{j}{\sim}$ ’. Consequently, we obtain a partition of  $\langle n \rangle$ , denoted by

$$\langle n \rangle = \bigcup_{s=1}^r I_s, \tag{15}$$

where the  $I_s$ ,  $s = 1, \dots, r$ , are the equivalence classes with respect to (14). Based on (15), we define further

$$J^{(s)} = \{j \in \langle p \rangle : h_{i, j} = 1 \text{ for all } i \in I_s\}, \tag{16}$$

$$\bar{J}^{(s)} = \{j \in \langle p \rangle : h_{i, j} = 1 \text{ for some } i \in I_s\}. \tag{17}$$

The sets  $\{J^{(s)}, \bar{J}^{(s)}\}_{s \in \langle r \rangle}$  will determine the factorization form of  $\nu(\mathbf{x}, E)$ .

**Theorem 2.** *Let  $\mathbf{Z}$  be as in Theorem 1. Furthermore, let  $\mathbf{H}$  be the hitting matrix corresponding to  $(\mathbf{A}, \mathbf{X})$  with  $\mathbf{X} = \mathbf{A} \odot \mathbf{Z}$ , and  $\{J^{(s)}, \bar{J}^{(s)}\}_{s \in \langle r \rangle}$  be the sets defined in (16) and (17). Then, with probability 1, we have*

(i)  $r = r(\mathcal{J}(\mathbf{A}, \mathbf{X}))$ ,

(ii) for all  $J \subset \langle p \rangle$ ,  $J \in \mathcal{J}_r(\mathbf{A}, \mathbf{A} \odot \mathbf{Z})$  if and only if  $J$  can be written as

$$J = \{j_1, \dots, j_r\} \quad \text{with } j_s \in J^{(s)}, \quad s \in \langle r \rangle, \tag{18}$$

(iii) for  $\nu(\mathbf{x}, E)$  defined in (6),

$$\nu(\mathbf{X}, E) = \prod_{s=1}^r \nu^{(s)}(\mathbf{X}, E) \quad \text{with} \quad \nu^{(s)}(\mathbf{X}, E) = \frac{\sum_{j \in J^{(s)}} w_j^{(s)}(\mathbf{X}) \nu_j^{(s)}(\mathbf{X}, E)}{\sum_{j \in \bar{J}^{(s)}} w_j^{(s)}(\mathbf{X})}, \tag{19}$$

where, for all  $j \in J^{(s)}$ ,

$$w_j^{(s)}(\mathbf{x}) := \widehat{z}_j f_{Z_j}(\widehat{z}_j) \prod_{k \in \bar{J}^{(s)} \setminus \{j\}} F_{Z_k}(\widehat{z}_k), \tag{20}$$

$$\nu_j^{(s)}(\mathbf{x}, E) := \delta_{\pi_j(E)}(\widehat{z}_j) \prod_{k \in \bar{J}^{(s)} \setminus \{j\}} \mathbb{P}(Z_k \in \pi_k(E) \mid Z_k < \widehat{z}_k), \tag{21}$$

with  $\widehat{z}_j = \widehat{z}_j(\mathbf{x})$  as in (3).

The proof of Theorem 2 is given in Section 4.

TABLE 1: Means and standard deviations (in parentheses) of the running times (in seconds) for the decomposition of the hitting matrix  $\mathbf{H}$ , based on 100 independent observations  $\mathbf{X} = \mathbf{A} \odot \mathbf{Z}$ , where  $\mathbf{A}$  is an  $n \times p$  matrix corresponding to a discretized Smith model.

$p$	$n$			
	1	5	10	50
2500	0.03 (0.02)	0.13 (0.03)	0.24 (0.04)	1.25 (0.09)
10000	0.11 (0.04)	0.50 (0.05)	1.00 (0.08)	4.98 (0.33)

**Remark 2.** Note that this result does not claim that  $\nu(\mathbf{x}, E)$  in (19) is the *regular conditional probability*. It merely provides an equivalent expression for (6), which is valid with probability 1. We still need to show that (6), or, equivalently, (19), is indeed the regular conditional probability.

From (20) and (21), we can see that  $\nu^{(s)}$  is the conditional distribution of  $(Z_j)_{j \in \bar{J}^{(s)}}$ . Therefore, relation (19) implies that  $\{(Z_j)_{j \in \bar{J}^{(s)}}\}_{s \in \langle r \rangle}$ , as vectors indexed by  $s$ , are *conditionally independent*, given  $\mathbf{X} = \mathbf{x}$ . This leads to the following improved conditional sampling algorithm.

**Algorithm II.**

1. Compute  $\widehat{z}_j$  for  $j = 1, \dots, p$  and the hitting matrix  $\mathbf{H} = \mathbb{H}(\mathbf{A}, \mathbf{x})$ .
2. Identify  $\{J^{(s)}, \bar{J}^{(s)}\}_{s \in \langle r \rangle}$  by (16) and (17).
3. Compute  $\{w_j^{(s)}\}_{j \in J^{(s)}}$  for all  $s \in \langle r \rangle$  by (20).
4. Sample  $(Z_j)_{j \in \bar{J}^{(s)}} \mid \mathbf{X} = \mathbf{x} \sim \nu^{(s)}(\mathbf{x}, \cdot)$  independently for  $s = 1, \dots, r$ .
5. Combine the sampled  $(Z_j)_{j \in \bar{J}^{(s)}}$ ,  $s = 1, \dots, r$ , to obtain a sample  $\mathbf{Z}$ .

This algorithm identifies all hitting scenarios in an efficient way. To illustrate its efficiency compared to Algorithm I, consider that  $r = 10$  and  $|J^{(s)}| = 10$  for all  $s \in \langle 10 \rangle$ . Then, applying (6) in Algorithm I requires storing in memory the weights of all  $10^{10}$  hitting scenarios. In contrast, the implementation of (19) requires saving only  $10 \times 10$  weights. This improvement is critical in practice since it allows us to handle large, realistic models.

Table 1 demonstrates the running times of Algorithm II as a function of the dimensions  $n \times p$  of the matrix  $\mathbf{A}$ . It is based on a discretized two-dimensional Smith model (Section 3.2) and measured on an Intel<sup>®</sup> Core<sup>™</sup> Duo CPU E4400 2.00 GHz with 2 GB RAM. It is remarkable that the times scale linearly in both  $n$  and  $p$ .

**3. Examples**

**3.1. MARMA processes**

In this section we apply our result to the MARMA processes studied by Davis and Resnick [6]. A stationary process  $\{X_t\}_{t \in \mathbb{Z}}$  is a MARMA( $m, q$ ) process if it satisfies the MARMA recursion, i.e.

$$X_t = \phi_1 X_{t-1} \vee \dots \vee \phi_m X_{t-m} \vee Z_t \vee \theta_1 Z_{t-1} \vee \dots \vee \theta_q Z_{t-q} \tag{22}$$

for all  $t \in \mathbb{Z}$ , where  $\phi_i \geq 0$  and  $\theta_j \geq 0$ ,  $i = 1, \dots, m$ ,  $j = 1, \dots, q$ , are the parameters, and  $\{Z_t\}_{t \in \mathbb{Z}}$  are independent and identically distributed 1-Fréchet random variables. Proposition 2.2

of [6] shows that (22) has a unique solution of the form

$$X_t = \bigvee_{j=0}^{\infty} \psi_j Z_{t-j} < \infty \quad \text{almost surely,} \tag{23}$$

with  $\psi_j \geq 0, j \geq 0$ , and  $\sum_{j=0}^{\infty} \psi_j < \infty$ , if and only if  $\phi^* = \bigvee_{i=1}^m \phi_i < 1$ . In this case,

$$\psi_j = \bigvee_{k=0}^{j \wedge q} \alpha_{j-k} \theta_k,$$

where the  $\{\alpha_j\}_{j \in \mathbb{Z}}$  are determined recursively by  $\alpha_j = 0$  for all  $j < 0, \alpha_0 = 1$ , and

$$\alpha_j = \phi_1 \alpha_{j-1} \vee \phi_2 \alpha_{j-2} \vee \dots \vee \phi_m \alpha_{j-m} \quad \text{for all } j \geq 1. \tag{24}$$

In the sequel we will focus on the MARMA process (22) with unique stationary solution (23). In this case, the MARMA process is a spectrally discrete max-stable process. Without loss of generality, we also assume  $\{Z_k\}_{k \in \mathbb{Z}}$  to be standard 1-Fréchet.

We consider the prediction of the MARMA process in the following framework: suppose that at each time  $t \in \{1, \dots, n\}$  we observe the value  $X_t$  of the process, and the goal is to predict  $\{X_s\}_{n < s \leq n+N}$ . We do so by generating independent and identically distributed samples from the conditional distribution  $\{X_s\}_{n < s \leq n+N} \mid \{X_t\}_{t=1, \dots, n}$ . To apply our result, it suffices to provide a max-linear representation of this model. We will truncate (23) to obtain

$$\tilde{X}_t = \bigvee_{j=0}^p \psi_j Z_{t-j} \quad \text{for all } t = 1, \dots, n + N. \tag{25}$$

The truncated process can approximate the original process arbitrarily well, if we take  $p$  large enough. Indeed, by using the independence and max-stability of the  $Z_t$ s, we can show that

$$P(\tilde{X}_t = X_t) = P\left(\bigvee_{j=0}^p \psi_j Z_{t-j} \geq \bigvee_{j=p+1}^{\infty} \psi_j Z_{t-j}\right) = 1 - \frac{\sum_{j=p+1}^{\infty} \psi_j}{\sum_{j=0}^{\infty} \psi_j} \rightarrow 1 \tag{26}$$

as  $p \rightarrow \infty$ . Moreover, by induction on  $\alpha_j$  in (24), we can show that  $\alpha_j \leq (\phi^*)^{\lceil j/m \rceil}$  for all  $j \in \mathbb{N}$ , and, thus, convergence (26) above is geometrically fast.

Now, we reformulate the prediction problem with the model (25) as follows:

$$\text{observe } \mathbf{X}_{[1,n]} = \mathbf{A} \odot \mathbf{Z} \quad \text{and} \quad \text{predict } \mathbf{Y}_{[1,N]} = \mathbf{B} \odot \mathbf{Z} \mid \mathbf{X}_{[1,n]}.$$

Here  $\mathbf{X}_{[1,n]} = (\tilde{X}_1, \dots, \tilde{X}_n)$ ,  $\mathbf{Y}_{[1,N]} = (\tilde{X}_{n+1}, \dots, \tilde{X}_{n+N})$ ,  $\mathbf{Z} = (Z_{1-p}, Z_{2-p}, \dots, Z_{n+N})$ , and  $\mathbf{A} \in \mathbb{R}_+^{n \times (p+n+N)}$  and  $\mathbf{B} \in \mathbb{R}_+^{N \times (p+n+N)}$  are determined by (25). In particular,

$$\begin{pmatrix} \mathbf{A} \\ \mathbf{B} \end{pmatrix} = \begin{pmatrix} \psi_p & \psi_{p-1} & \dots & \psi_0 & 0 & 0 & \dots & 0 \\ 0 & \psi_p & \psi_{p-1} & \dots & \psi_0 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & \psi_p & \psi_{p-1} & \dots & \psi_0 & 0 \\ 0 & \dots & 0 & 0 & \psi_p & \psi_{p-1} & \dots & \psi_0 \end{pmatrix}. \tag{27}$$

In practice, given the observations  $X_{[1,n]}$ , we use our algorithm to sample from the conditional distribution  $Z \mid X_{[1,n]}$ . Therefore, we can sample

$$Y_{[1,N]} \mid X_{[1,n]} \stackrel{D}{=} B \odot Z \mid X_{[1,n]}. \tag{28}$$

Our approach is different from the prediction considered in [6], which we will briefly review. Davis and Resnick took the classic time series point of view and investigated how to approximate  $X_s$  by a max-linear combination of  $\{X_t\}_{t=1,\dots,n}$ , with respect to a certain metric  $d$ . Namely, for all  $Y \in \mathcal{H}$  with

$$\mathcal{H} = \left\{ \bigvee_{j=-\infty}^{\infty} \alpha_j Z_j : \alpha_j \geq 0, \sum_{j=-\infty}^{\infty} \alpha_j < \infty \right\},$$

they considered a projection of  $Y$  onto the space  $\mathcal{F}_n$ , max-linearly spanned by  $\{X_t\}_{t=1,\dots,n}$ :  $\mathcal{F}_n = \{ \bigvee_{j=0}^{\infty} b_j X_{n-j} : b_j \geq 0, \sum_{j=0}^{\infty} b_j < \infty \}$ . That is, consider the projection  $\mathcal{P}_n Y$  defined by

$$\mathcal{P}_n Y = \operatorname{argmin}_{\tilde{Y} \in \mathcal{F}_n} d(\tilde{Y}, Y) \tag{29}$$

with the metric  $d$  induced by  $d(\bigvee_j \alpha_j Z_j, \bigvee_j \beta_j Z_j) = \sum_j |\alpha_j - \beta_j|$ . For specific MARMA processes, Davis and Resnick [6] provided predictors based on the projection (29). We will refer to these predictors as the *projection predictors*.

In general, the conditional samplings reflect the conditional distribution (28), and they provide more information than the projection predictors. Sampling multiple times from (28), we can calculate, e.g. conditional medians, conditional means, quantiles, etc., which are optimal predictors with respect to various loss functions.

**Example 4.** (MAR( $m$ ) processes.) Consider the MAR( $m$ )  $\equiv$  MARMA( $m, 0$ ) process with

$$X_t = \phi_1 X_{t-1} \vee \dots \vee \phi_m X_{t-m} \vee Z_t. \tag{30}$$

The projection predictor for this model can be obtained recursively by

$$\widehat{X}_{t+k} = \phi_1 \widehat{X}_{t+k-1} \vee \dots \vee \phi_m \widehat{X}_{t+k-m}, \tag{31}$$

with  $\widehat{X}_t = X_t$ ,  $t = 1, \dots, n$  (see [6, p. 799]).

Figure 2 illustrates an application of our conditional sampling algorithm in this case. Consider an MAR(3) process  $\{X_t\}_{t=1}^{150}$  with  $\phi_1 = 0.7$ ,  $\phi_2 = 0.5$ , and  $\phi_3 = 0.3$ . In effect, we use the truncated model  $\{\tilde{X}_t\}_{t \in \mathbb{N}}$  in (25) with  $p = 500$ , but we still write  $X_t$  for the sake of simplicity. Treating the first 100 values as observed, we plot the projection predictor, conditional upper 95% quantiles and the conditional medians of  $\{X_s\}_{s=101}^{150}$  based on 500 independent samples from the conditional distribution.

Observe that the value of the projection predictor in Figure 2 is always below the conditional median. This ‘underestimation’ phenomenon was typical in all the simulations we performed. It can be explained by the fact that, the projection predictor in (31) does not account for the jumps of the process caused by new arrivals  $\{Z_t\}_{t>100}$ . Indeed, a large new arrival  $Z_t$  will cause the process to jump immediately to  $Z_t$  at time  $t$ , but this will never occur for the projection predictor  $\widehat{X}_t$ .

Next, we apply our algorithm to examine the bias of the projection predictor. To do this, for each generated MARMA process, we calculated the cumulative probability that the projection predictor corresponds to, for each location  $s = 101, \dots, 150$ . Namely, using 500 independent

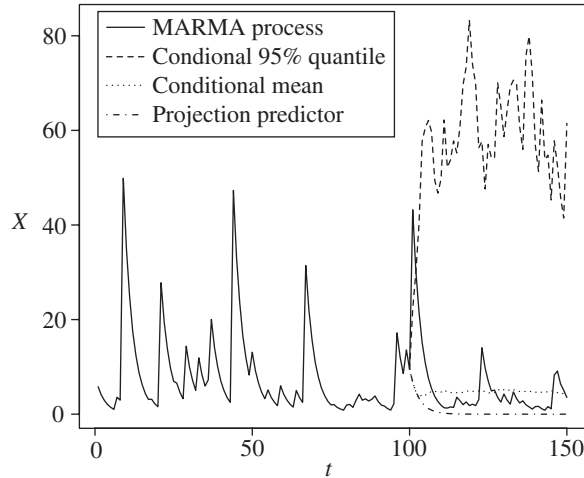


FIGURE 2: Prediction of a MARMA(3,0) process with  $\phi_1 = 0.7, \phi_2 = 0.5,$  and  $\phi_3 = 0.3,$  based on the observation of the first 100 values of the process.

samples  $\{X_s^{(k)}\}_{s=101}^{150}, k = 1, \dots, 500,$  from the conditional distribution, we calculated

$$P(X_s \leq \widehat{X}_s \mid \{X_t\}_{t=1}^{100}) \approx \frac{1}{500} \sum_{k=1}^{500} \mathbf{1}\{X_s^{(k)} \leq \widehat{X}_s\} \quad \text{for all } s > 100, \tag{32}$$

where  $\widehat{X}_s$  is the projection predictor in (31). This procedure was repeated 1000 times for independent realizations of  $\{X_t\}_{t=1}^{100}$  and the means of the (estimated) probability in (32) are reported in Table 2. Note that as the time lag increases, the conditional quantiles of the projection predictors decrease. In this way, our conditional sampling algorithm helps quantify numerically the observed underestimation phenomenon in Figure 2.

Finally, we compare the generated conditional samples to the true process values at times  $s = 101, \dots, 150.$  Our goal is to demonstrate the validity of our conditional sampling algorithm. The idea is that, at each location  $s = 101, \dots, 150,$  the true process should lie below the

TABLE 2: Cumulative probabilities that the projection predictors correspond to at time  $100 + t,$  based on 1000 simulations.

$t$	Mean (%)
1	70.6
2	50.3
3	35.6
4	25.3
5	17.8
10	2.9
20	0.1
30	0.0
40	0.0

TABLE 3: Coverage rates (CRs) and the widths of the upper 95% confidence intervals at time  $100 + t$ , based on 1000 simulations.

$t$	CR	Width
1	0.956	13.06
2	0.952	26.6
3	0.954	37.8
4	0.957	45.6
5	0.966	51.2
10	0.947	62.8
20	0.943	66.0
30	0.951	66.2
40	0.955	65.4

predicted 95% upper confidence bound of  $X_s \mid \{X_t\}_{t=1}^{100}$ , with probability at least 95%. (Note that due to the presence of atoms in the conditional distributions, the coverage probability may in principle be higher than 95%.) Motivated by this, we repeat the procedure in the previous paragraph and record the proportion of the times that  $X_s$  is below the predicted confidence quantile, for each  $s$ . We refer to these values as the *coverage rates*. As discussed, the coverage rates should be close to 95%. This is supported by our simulation result, shown in Table 3.

Table 3 also shows the widths of the upper 95% confidence intervals. Note that these widths are not equal to the upper confidence bounds, given by the conditional 95% quantiles, since the left endpoint of the conditional distributions are greater than 0. When the time lag is small, the left endpoint is large and the widths are small, due to the strong influence of the past of the process  $\{X_t\}_{t=1}^{100}$ . On the other hand, because of the weak temporal dependence of the MAR(3) processes, this influence decreases fast as the lags increase. Consequently, the conditional distribution converges to the unconditional distribution, and the conditional quantile to the unconditional quantile. Note that the (unconditional) 95% quantile of  $X_s$  for the MARMA process (23) can be calculated via the formula  $0.95 = P(\sigma Z \leq u) = \exp\{-\sigma u^{-1}\}$ , with  $\sigma = \sum_{j=0}^p \psi_j$ . For the MAR(3) process, we have  $\sigma = 3.4$  and the 95% quantile of  $X_s$  equals 66.29. This is consistent with the widths in Table 3 for large lags.

**Remark 3.** As pointed out by an anonymous referee, in this case we can directly generate samples from  $\{X_s\}_{s=n+1}^N \mid \{X_t\}_{t=1}^n$ , by generating independent Fréchet random variables and iterating (30). We selected this example only for illustrative purposes and to be able to compare with the projection predictors in [6]. We can modify slightly the prediction problem, such that our algorithm still applies by accordingly adjusting (27), while both the projection predictor and the direct method by using (30) do not apply. For example, consider the prediction problem with respect to the conditional distribution

$$P(\{X_s\}_{s=2n+1}^{2n+N} \in \cdot \mid \{X_t : t = 1, 3, \dots, 2n - 1\})$$

(prediction with only partial history observed) or

$$P(\{X_s\}_{s=2}^{n-1} \in \cdot \mid X_1, X_n)$$

prediction of the middle path with the beginning and the endpoint (in the future) given). In other words, our algorithm has no restriction on the locations of observations. This feature is of great importance in spatial prediction problems.

### 3.2. The discrete Smith model

Consider the following *moving maxima* random field model in  $\mathbb{R}^2$ :

$$X_t = \int_{\mathbb{R}^2}^e \phi(\mathbf{t} - \mathbf{u}) M_\alpha(d\mathbf{u}), \quad \mathbf{t} = (t_1, t_2) \in \mathbb{R}^2. \tag{33}$$

Here  $M_\alpha$  is an  $\alpha$ -Fréchet random sup-measure on  $\mathbb{R}^2$  with the Lebesgue control measure. Smith [24] proposed to use for  $\phi$  the bivariate Gaussian density,

$$\phi(t_1, t_2) := \frac{\beta_1 \beta_2}{2\pi \sqrt{1 - \rho^2}} \exp \left\{ -\frac{1}{2(1 - \rho^2)} [\beta_1^2 t_1^2 - 2\rho \beta_1 \beta_2 t_1 t_2 + \beta_2^2 t_2^2] \right\},$$

with correlation  $\rho \in (-1, 1)$  and variances  $\sigma_i^2 = 1/\beta_i^2$ ,  $i = 1, 2$ . Consistent and asymptotically normal estimators for the parameters  $\rho$ ,  $\beta_1$ , and  $\beta_2$  were obtained by de Haan and Pereira [12]. Here, we will assume that these parameters are known and will illustrate the conditional sampling methodology over a discretized version of the random field (33). Namely, we truncate the extremal integral in (33) to the square region  $[-M, M]^2$  and consider a uniform mesh of size  $h := M/q$ ,  $q \in \mathbb{N}$ . We then set

$$X_t := \bigvee_{-q \leq j_1, j_2 \leq q-1} h^{2/\alpha} \phi(\mathbf{t} - \mathbf{u}_{j_1 j_2}) Z_{j_1 j_2}, \tag{34}$$

where

$$\mathbf{u}_{j_1 j_2} = \left( \left( j_1 + \frac{1}{2} \right) h, \left( j_2 + \frac{1}{2} \right) h \right)$$

and

$$h^{2/\alpha} Z_{j_1 j_2} \stackrel{D}{=} M_\alpha((j_1 h, (j_1 + 1)h] \times (j_2 h, (j_2 + 1)h]).$$

This discretized model (34) can be made arbitrarily close to the spectrally continuous model in (33) by taking a fine mesh  $h$  and sufficiently large  $M$  (see, e.g. [26]).

Suppose that the random field  $X$  in (34) is observed at  $n$  locations  $X_{t_i} = x_i$ ,  $t_i \in [-M, M]^2$ ,  $i = 1, \dots, n$ . In view of (34), we have the max-linear model  $\mathbf{X} = \mathbf{A} \odot \mathbf{Z}$ , with  $\mathbf{X} = (X_{t_i})_{i=1}^n$  and  $\mathbf{Z} = (Z_j)_{j=1}^p$ ,  $p = q^2$ . By sampling from the conditional distribution of  $\mathbf{Z} \mid \mathbf{X} = \mathbf{x}$ , we can predict the random field  $X_s$  at arbitrary locations  $s \in \mathbb{R}^2$ .

To illustrate our algorithm, we used model (34) with parameter values  $\rho = 0$ ,  $\beta_1 = \beta_2 = 1$ ,  $M = 4$ ,  $p = q^2 = 2500$ , and  $n = 7$  observed locations. We generated  $N = 500$  independent samples from the conditional distribution of the random field  $\{X_s\}$ , where  $s$  takes values on a uniform  $100 \times 100$  grid, in the region  $[-2, 2] \times [-2, 2]$ . We have already seen four of these realizations in Figure 1. Figure 3 illustrates the median and 0.95th quantile of the conditional distribution. The former provides the optimal predictor for the values of the random field given the observed data, with respect to the absolute deviation loss. The marginal quantiles, on the other hand, provide important confidence regions for the random field, given the data.

Certainly, conditional sampling may be used to address more complex *functional* prediction problems. In particular, given a two-dimensional threshold surface, we can readily obtain the *correct* probability that the random field exceeds or stays below this surface, conditionally on the observed values. This is much more than what marginal conditional distributions can provide.



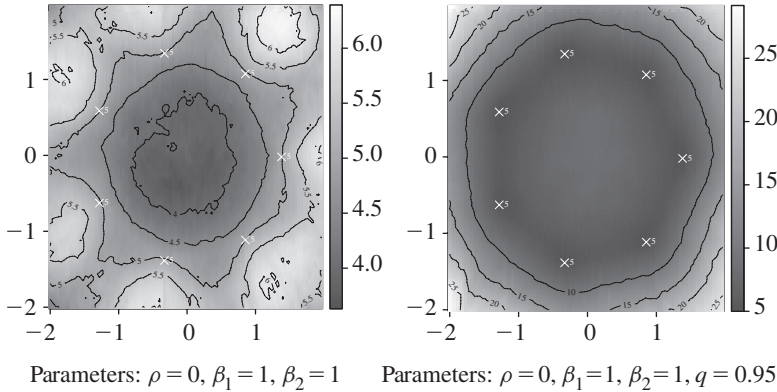


FIGURE 3: Conditional medians (left) and 0.95th conditional marginal quantiles (right) of the Smith model. Each cross indicates an observed location of the random field, with the observed value at the right.

#### 4. Proofs of Theorems 1 and 2

In this section we prove Theorems 1 and 2. We will first prove Theorem 2, which simplifies the regular conditional probability formula (6) in Theorem 1. Then, we show that the simplified new formula is the desired regular conditional probability, which completes the proof of Theorem 1. The key step to prove Theorem 2 is the following lemma. Write  $H_{\cdot j} = \{i \in \langle r \rangle : h_{i,j} = 1\}$ .

**Lemma 1.** *Under the assumptions of Theorem 2, with probability 1,*

- (i)  $J^{(s)}$  is nonempty for all  $s \in \langle r \rangle$ , and
- (ii) for all  $j \in J^{(s)}$ ,  $H_{\cdot j} \cap I_s \neq \emptyset$  implies that  $H_{\cdot j} \subset I_s$ .

*Proof.* Note that to show part (ii), it suffices to observe that since  $I_s$  is an equivalence class with respect to relation (14),  $H_{\cdot j} \setminus I_s$  and  $H_{\cdot j} \cap I_s$  cannot be both nonempty. Thus, it remains to show part (i). We proceed by excluding several P-measure zero sets, on which the desired results may not hold.

First, observe that, for all  $i \in \langle n \rangle$ , the maximum value of  $\{a_{i,j} Z_j\}_{j \in \langle r \rangle}$  is achieved for unique  $j \in \langle p \rangle$  with probability 1, since the  $Z_j$ s are independent and have continuous distributions. Thus, the set

$$\mathcal{N}_1 := \bigcup_{i \in \langle n \rangle, j_1, j_2 \in \langle p \rangle, j_1 \neq j_2} \{a_{i,j_1} Z_{j_1} = a_{i,j_2} Z_{j_2} = \max_{j \in \langle p \rangle} a_{i,j} Z_j\}$$

has P-measure 0. From now on, we focus on the event  $\mathcal{N}_1^c$  and set  $j(i) = \operatorname{argmax}_{j \in \langle p \rangle} a_{i,j} Z_j$  for all  $i \in \langle n \rangle$ .

Next, we show that, with probability 1,  $i_1 \overset{j}{\sim} i_2$  implies that  $j(i_1) = j(i_2)$ . That is, the set

$$\mathcal{N}_2 := \bigcup_{j \in \langle p \rangle, i_1, i_2 \in \langle n \rangle, i_1 \neq i_2} \mathcal{N}_{j,i_1,i_2} \quad \text{with} \quad \mathcal{N}_{j,i_1,i_2} := \{j(i_1) \neq j(i_2), i_1 \overset{j}{\sim} i_2\}$$

has P-measure 0. It suffices to show that  $P(\mathcal{N}_{j,i_1,i_2}) = 0$  for all  $i_1 \neq i_2$ . If not, since  $\langle p \rangle$  and  $\langle n \rangle$  are finite sets, there exists  $\mathcal{N}_0 \subset \mathcal{N}_{j,i_1,i_2}$  such that  $j(i_1) = j_1 \neq j(i_2) = j_2$  on  $\mathcal{N}_0$ , and

$P(\mathcal{N}_0) > 0$ . At the same time, however, observe that  $i_1 \stackrel{j}{\sim} i_2$  implies that  $h_{i_1,j} = h_{i_2,j} = 1$ , which yields

$$a_{i_k,j} \widehat{z}_j = x_{i_k} = a_{i_k,j(i_k)} Z_{j(i_k)} = a_{i_k,j_k} Z_{j_k}, \quad k = 1, 2.$$

It then follows that on  $\mathcal{N}_0$ ,  $Z_{j_1}/Z_{j_2} = a_{i_1,j} a_{i_2,j_2} / (a_{i_2,j} a_{i_1,j_1})$ , which is a constant. This constant is strictly positive and finite. Indeed, this is because on  $\mathcal{N}_1^c$ ,  $a_{i,j(i)} > 0$  by Assumption A and  $h_{i,j} = 1$  implies that  $a_{i,j} > 0$ . Since  $Z_{j_1}$  and  $Z_{j_2}$  are independent continuous random variables, it then follows that  $P(\mathcal{N}_0) = 0$ .

Finally, we focus on the event  $(\mathcal{N}_1 \cup \mathcal{N}_2)^c$ . Then, for any  $i_1, i_2 \in I_s$ , we have  $\tilde{i}_1 \sim i_2$  and let  $\tilde{i}_0, \dots, \tilde{i}_n$  be as in (14). It then follows that  $j(i_1) = j(\tilde{i}_0) = j(\tilde{i}_1) = \dots = j(\tilde{i}_n) = j(i_2)$ . Note that, for all  $i \in \langle n \rangle$ ,  $h_{i,j(i)} = 1$  by the definition of  $j(i)$ . Hence,  $j(i_1) = j(i_2) \in J^{(s)}$ . We have thus completed the proof.

*Proof of Theorem 2.* Since the  $\{I_s\}_{s \in \langle r \rangle}$  are disjoint with  $\bigcup_{s \in \langle r \rangle} I_s = \langle n \rangle$ , in the language of the set-covering problem, to cover  $\langle n \rangle$ , we need to cover each  $I_s$ . By part (ii) of Lemma 1, any two different  $I_{s_1}$  and  $I_{s_2}$  cannot be covered by a single set  $H_j$ . Thus, we need at least  $r$  sets to cover  $\langle n \rangle$ . On the other hand, with probability 1, we can select one  $j_s$  from each  $J^{(s)}$  (by part (i) of Lemma 1), which yields a valid cover. That is, with probability 1,  $r = r(\mathcal{J}(\mathbf{H}))$  and any valid minimum-cost cover of  $\langle n \rangle$  must be as in (18), and vice versa. We have thus proved parts (i) and (ii).

To show (iii), by a straightforward calculation, we have, with probability 1,

$$\begin{aligned} \sum_{J \in \mathcal{J}_r(\mathbf{A}, \mathbf{x})} w_J &= \sum_{j_1 \in J^{(1)}} \cdots \sum_{j_r \in J^{(r)}} w_{j_1, \dots, j_r} \\ &= \sum_{j_1 \in J^{(1)}} \cdots \sum_{j_{r-1} \in J^{(r-1)}} \left[ \prod_{s=1}^{r-1} \widehat{z}_{j_s} f_{Z_{j_s}}(\widehat{z}_{j_s}) \prod_{\substack{j \in \overline{J}^{(r)} \\ j \neq j_1, \dots, j_{r-1}}} F_{Z_j}(\widehat{z}_j) \right] \\ &\quad \times \left\{ \sum_{j \in J^{(r)}} \left( \widehat{z}_j f_{Z_j}(\widehat{z}_j) \prod_{k \in \overline{J}^{(r)} \setminus \{j\}} F_{Z_k}(\widehat{z}_k) \right) \right\} \\ &= \prod_{s=1}^r \sum_{j \in J^{(s)}} \left( \widehat{z}_j f_{Z_j}(\widehat{z}_j) \prod_{k \in \overline{J}^{(s)} \setminus \{j\}} F_{Z_k}(\widehat{z}_k) \right) \\ &= \prod_{s=1}^r \sum_{j \in J^{(s)}} w_j^{(s)}. \end{aligned} \tag{35}$$

Similarly, we have

$$\sum_{J \in \mathcal{J}_r(\mathbf{A}, \mathbf{x})} w_J v_J(\mathbf{x}, E) = \prod_{s=1}^r \left( \sum_{j \in J^{(s)}} w_j^{(s)} v_j^{(s)}(\mathbf{x}, E) \right). \tag{36}$$

By substituting (35) and (36) into (6), we obtain the desired result and complete the proof.

*Proof of Theorem 1.* To prove that  $\nu$  in (6) yields the regular conditional probability of  $\mathbf{Z}$  given  $\mathbf{X}$ , it is enough to show that

$$P(\mathbf{X} \in D, \mathbf{Z} \in E) = \int_D \nu(\mathbf{x}, E) P^{\mathbf{X}}(d\mathbf{x}) \tag{37}$$

for all rectangles  $D \in \mathcal{R}_{\mathbb{R}_+^n}$  and  $E \in \mathcal{R}_{\mathbb{R}_+^p}$ . In view of Theorem 2, it is enough to work with  $\nu(\mathbf{x}, E)$  given by (19).

We will prove (37) by breaking the integration into a suitable sum of integrals over regions corresponding to all hitting matrices  $\mathbf{H}$  for the max-linear model  $\mathbf{X} = \mathbf{A} \odot \mathbf{Z}$ . We say such a hitting matrix  $\mathbf{H}$  is *nice* if  $J^{(s)}$  defined in (16) is nonempty for all  $s \in \langle r \rangle$ . In view of Lemma 1, it suffices to focus on the set  $\mathcal{H}(\mathbf{A})$  of *nice* hitting matrices  $\mathbf{H}$ . Note that the set  $\mathcal{H}(\mathbf{A})$  is finite since the elements of the hitting matrices are 0s and 1s.

For all rectangles  $D \in \mathcal{R}_{\mathbb{R}_+^n}$ , let

$$D_{\mathbf{H}} = \{\mathbf{x} = \mathbf{A} \odot \mathbf{z} : \mathbb{H}(\mathbf{A}, \mathbf{x}) = \mathbf{H}, \mathbf{x} \in D\}$$

be the set of all  $\mathbf{x} \in \mathbb{R}_+^n$  that give rise to the hitting matrix  $\mathbf{H}$ . By Lemma 1(i), for the random vector  $\mathbf{X} = \mathbf{A} \odot \mathbf{Z}$ , with probability 1, we have

$$\mathbf{X} = \sum_{\mathbf{H} \in \mathcal{H}(\mathbf{A})} \mathbf{X} \mathbf{1}_{D_{\mathbf{H}}}(\mathbf{X})$$

and, hence,

$$\int_D \nu(\mathbf{x}, E) P^{\mathbf{X}}(d\mathbf{x}) = \sum_{\mathbf{H} \in \mathcal{H}(\mathbf{A})} \int_{D_{\mathbf{H}}} \nu(\mathbf{x}, E) P^{\mathbf{X}}(d\mathbf{x}). \tag{38}$$

Now fix an arbitrary and *nonrandom* nice hitting matrix  $\mathbf{H} \in \mathcal{H}(\mathbf{A})$ . Let  $\{I_s\}_{s \in \langle r \rangle}$  denote the partition of  $\langle n \rangle$  determined by (14), and let  $J^{(s)}, \bar{J}^{(s)}, s = 1, \dots, r$ , be as in (16). Recall that  $J^{(s)} \subset \bar{J}^{(s)}$  and the sets  $\bar{J}^{(s)}, s = 1, \dots, r$ , are disjoint.

Focus on the set  $D_{\mathbf{H}} \subset \mathbb{R}_+^n$ . Without loss of generality, and for notational convenience, suppose that  $s \in I_s$  for all  $s = 1, \dots, r$ . That is,

$$I_1 = \{1, i_{1,2}, \dots, i_{1,k_1}\}, \quad I_2 = \{2, i_{2,2}, \dots, i_{2,k_2}\}, \dots, I_r = \{r, i_{r,2}, \dots, i_{r,k_r}\}.$$

Define the projection mapping  $\mathcal{P}_{\mathbf{H}} : D_{\mathbf{H}} \rightarrow \mathbb{R}_+^r$  onto the first  $r$  coordinates:

$$\mathcal{P}_{\mathbf{H}}(x_1, \dots, x_n) = (x_1, \dots, x_r) \equiv \mathbf{x}_r.$$

Note that  $\mathcal{P}_{\mathbf{H}}$  restricted to  $D_{\mathbf{H}}$  is one-to-one. Indeed, for all  $i \in I_s$ , we have  $x_i = a_{i,j} \widehat{z}_j$  and  $x_s = a_{s,j} \widehat{z}_j$  for all  $j \in J^{(s)}$  (recall (16)). This implies that  $x_i = (a_{i,j}/a_{s,j})x_s$  for all  $i \in I_s$  and all  $s = 1, \dots, r$ . Hence,  $\mathcal{P}_{\mathbf{H}}(\tilde{\mathbf{x}}) = \mathcal{P}_{\mathbf{H}}(\mathbf{x})$  implies that  $\tilde{\mathbf{x}} = \mathbf{x}$ .

Consequently, can write  $\mathbf{x} = \mathcal{P}^{-1}(\mathbf{x}_r), \mathbf{x}_r \in \mathcal{P}(D_{\mathbf{H}})$ , and

$$\int_{D_{\mathbf{H}}} \nu(\mathbf{x}, E) P^{\mathbf{X}}(d\mathbf{x}) = \int_{\mathcal{P}_{\mathbf{H}}(D_{\mathbf{H}})} \nu(\mathbf{x}, E) \mathbb{Q}_{\mathbf{H}}^{\mathbf{X}_r}(dx_1 \cdots dx_r),$$

where  $\mathbb{Q}_{\mathbf{H}}^{\mathbf{X}_r} := P^{\mathbf{X}} \circ \mathcal{P}_{\mathbf{H}}^{-1}$  is the induced measure on the set  $\mathcal{P}_{\mathbf{H}}(D_{\mathbf{H}})$ .

**Lemma 2.** *The measure  $\mathbb{Q}_H^{X_r}$  has a density with respect to the Lebesgue measure on the set  $\mathcal{P}_H(D_H)$ . The density is given by*

$$\mathbb{Q}_H^{X_r}(\mathbf{d}\mathbf{x}_r) = \mathbf{1}_{\mathcal{P}_H(D_H)}(\mathbf{x}_r) \prod_{s=1}^r \sum_{j \in J^{(s)}} w_j^{(s)}(\mathbf{x}) \frac{dx_1}{x_1} \dots \frac{dx_r}{x_r}. \tag{39}$$

The proof of this result is given below. In view of (19) and (39), we obtain

$$\begin{aligned} & \int_{\mathcal{P}_H(D_H)} v(\mathbf{x}, E) \mathbb{Q}_H^{X_r}(\mathbf{d}\mathbf{x}_r) \\ &= \int_{\mathcal{P}_H(D_H)} \underbrace{\prod_{s=1}^r \left( \frac{\sum_{j \in J^{(s)}} w_j^{(s)}(\mathbf{x}) v_j^{(s)}(\mathbf{x}, E)}{\sum_{k \in J^{(s)}} w_k^{(s)}(\mathbf{x})} \right)}_{=v(\mathbf{x}, E)} \times \underbrace{\prod_{s=1}^r \sum_{j \in J^{(s)}} w_j^{(s)}(\mathbf{x}) \frac{dx_1}{x_1} \dots \frac{dx_r}{x_r}}_{=\mathbb{Q}_H^{X_r}(\mathbf{d}\mathbf{x}_r)} \\ &= \int_{\mathcal{P}_H(D_H)} \prod_{s=1}^r \sum_{j \in J^{(s)}} w_j^{(s)}(\mathbf{x}) v_j^{(s)}(\mathbf{x}, E) \frac{dx_1}{x_1} \dots \frac{dx_r}{x_r}, \end{aligned}$$

which equals

$$\sum_{j_1 \in J^{(1)}, \dots, j_r \in J^{(r)}} \underbrace{\int_{\mathcal{P}_H(D_H)} \prod_{s=1}^r w_{j_s}^{(s)}(\mathbf{x}) v_{j_s}^{(s)}(\mathbf{x}, E) \frac{dx_1}{x_1} \dots \frac{dx_r}{x_r}}_{=:I(j_1, \dots, j_r)}. \tag{40}$$

Fix  $j_1 \in J^{(1)}, \dots, j_r \in J^{(r)}$  and focus on the integral  $I(j_1, \dots, j_r)$ . Define

$$\Omega_H^r(D_H) := \left\{ (z_{j_1}, \dots, z_{j_r}) : z_{j_s} = \frac{x_s}{a_{s, j_s}}, s = 1, \dots, r, \mathbf{x}_r = (x_s)_{s=1}^r \in \mathcal{P}_H(D_H) \right\}.$$

We have, by (20), (21), and replacing  $x_s$  with  $a_{s, j_s} z_{j_s}$ ,  $s = 1, \dots, r$  (simple change of variables),

$$\begin{aligned} & I(j_1, \dots, j_r) \\ &= \int_{\Omega_H^r(D_H)} \prod_{s=1}^r \left( f_{Z_{j_s}}(z_{j_s}) \prod_{k \in \overline{J}^{(s)} \setminus \{j_s\}} F_{Z_k}(\widehat{z}_k) \right. \\ & \quad \left. \times \delta_{\pi_{j_s}(E)}(z_{j_s}) \prod_{k \in \overline{J}^{(s)} \setminus \{j_s\}} \mathbb{P}(Z_k \in \pi_k(E) \mid Z_k < \widehat{z}_k) \right) \frac{dz_{j_1}}{z_{j_1}} \dots \frac{dz_{j_r}}{z_{j_r}} \\ &= \int_{\Omega_H^r(D_H)} \prod_{s=1}^r f_{Z_{j_s}}(z_{j_s}) \delta_{\pi_{j_s}(E)}(z_{j_s}) \\ & \quad \times \prod_{k \in \langle p \rangle \setminus \{j_1, \dots, j_r\}} \mathbb{P}(Z_k \in \pi_k(E), Z_k < \widehat{z}_k) dz_{j_1} \dots dz_{j_r}. \tag{41} \end{aligned}$$

Define

$$\Omega_{\mathbf{H};j_1,\dots,j_r}(D_{\mathbf{H}}) = \left\{ \mathbf{z} \in \mathbb{R}_+^p : \mathbf{x} = \mathbf{A} \odot \mathbf{z} \in D_{\mathbf{H}}, \right. \\ \left. z_{j_s} = \frac{x_s}{a_{s,j_s}}, s = 1, \dots, r, z_k < \widehat{z}_k(\mathbf{x}), k \in \langle p \rangle \setminus \{j_1, \dots, j_r\} \right\}.$$

By the independence of the  $Z_k$ s, (41) becomes

$$I(j_1, \dots, j_r) = \mathbb{P}(\mathbf{Z} \in \Omega_{\mathbf{H};j_1,\dots,j_r}(D_{\mathbf{H}}) \cap E). \tag{42}$$

By substituting (42) into (40), we obtain

$$\int_{D_{\mathbf{H}}} v(\mathbf{x}, E) \mathbb{P}^{\mathbf{X}}(d\mathbf{x}) = \int_{\mathcal{P}_{\mathbf{H}}(D_{\mathbf{H}})} v(\mathbf{x}, E) \mathbb{Q}_{\mathbf{H}}^{\mathbf{X}_r}(d\mathbf{x}_r) \\ = \sum_{j_1 \in J^{(1)}, \dots, j_r \in J^{(r)}} \mathbb{P}(\mathbf{Z} \in \Omega_{\mathbf{H};j_1,\dots,j_r}(D_{\mathbf{H}}) \cap E) \\ = \mathbb{P}(\mathbf{A} \odot \mathbf{Z} \in D_{\mathbf{H}}, \mathbf{Z} \in E), \tag{43}$$

because the summation over  $(j_1, \dots, j_r)$  accounts for all relevant hitting scenarios corresponding to the matrix  $\mathbf{H}$ . Substituting (43) into (38), we have

$$\int_D v(\mathbf{x}, E) \mathbb{P}^{\mathbf{X}}(d\mathbf{x}) = \sum_{\mathbf{H} \in \mathcal{H}(A)} \mathbb{P}(\mathbf{X} \equiv \mathbf{A} \odot \mathbf{Z} \in D_{\mathbf{H}}, \mathbf{Z} \in E) = \mathbb{P}(\mathbf{X} \in D, \mathbf{Z} \in E).$$

This completes the proof of Theorem 1.

*Proof of Lemma 2.* Consider the random vector  $\mathbf{X}_r = (X_1, \dots, X_r)$ . Observe that, by the definition of the set  $\mathcal{P}_{\mathbf{H}}(D_{\mathbf{H}})$ , on the event  $\{\mathbf{X}_r \in \mathcal{P}_{\mathbf{H}}(D_{\mathbf{H}})\}$ , we have

$$\mathbf{X}_r = \sum_{j_1 \in J^{(1)}, \dots, j_r \in J^{(r)}} \begin{pmatrix} a_{1,j_1} Z_{j_1} \\ \vdots \\ a_{r,j_r} Z_{j_r} \end{pmatrix} \prod_{s=1}^r \underbrace{\mathbf{1} \left\{ \bigvee_{k \in \overline{J}^{(s)} \setminus \{j_s\}} a_{s,k} Z_k < a_{s,j_s} Z_{j_s} \right\}}_{=:\mathbf{1}\{C_{s,j_s}\}}. \tag{44}$$

Note that, since  $J(s) \subset \overline{J}^{(s)}$ ,  $s = 1, \dots, r$ , the events  $\bigcap_{s=1}^r C_{s,j_s}$  are disjoint for all  $r$ -tuples  $(j_1, \dots, j_r) \in J^{(1)} \times \dots \times J^{(r)}$ .

Recall that our goal is to establish (39). By the fact that the sum in (44) involves only one nonzero term for some  $(j_1, \dots, j_r)$ , with probability 1, we have, for all measurable sets  $\Delta \subset \mathcal{P}_{\mathbf{H}}(D_{\mathbf{H}})$ , writing  $\xi_{j_s} = a_{s,j_s} Z_{j_s}$ ,

$$\mathbb{Q}_{\mathbf{H}}^{\mathbf{X}_r}(\Delta) \equiv \mathbb{P}(\mathbf{X}_r \in \Delta) \\ = \sum_{j_1 \in J^{(1)}, \dots, j_r \in J^{(r)}} \mathbb{P} \left( \{(\xi_{j_1}, \dots, \xi_{j_r}) \in \Delta\} \cap \left( \bigcap_{s=1}^r C_{s,j_s} \right) \right). \tag{45}$$

Now, consider the last probability, for fixed  $(j_1, \dots, j_r)$ . The random variables  $\xi_{j_s}$ ,  $s = 1, \dots, r$ , are independent and they have densities  $f_{Z_{j_s}}(x_s/a_{s,j_s})/a_{s,j_s}$ ,  $x_s \in \mathbb{R}_+$ . We also know

that the events  $C_{s,j_s}, s = 1, \dots, r$ , are mutually independent, since their definitions involve  $Z_k$ s indexed by disjoint sets  $\bar{J}^{(s)}, s = 1, \dots, r$ . By conditioning on the  $\xi_{j_s} s$ , we find that the probability on the right-hand side of (45) equals

$$\begin{aligned} & \int_{\Delta} \left( \prod_{s=1}^r \frac{1}{a_{s,j_s}} f\left(\frac{x_s}{a_{s,j_s}}\right) \right) \prod_{s=1}^r \mathbb{P}\left( \bigvee_{k \in \bar{J}^{(s)} \setminus \{j_s\}} a_{s,k} Z_k < x_s \right) dx_1 \cdots dx_r \\ &= \int_{\Delta} \prod_{s=1}^r \left( \frac{1}{a_{s,j_s}} f\left(\frac{x_s}{a_{s,j_s}}\right) \prod_{k \in \bar{J}^{(s)} \setminus \{j_s\}} F_{Z_k}\left(\frac{x_s}{a_{s,k}}\right) \right) dx_1 \cdots dx_r. \end{aligned}$$

In view of expressions (20) and (44), replacing

$$\sum_{j_1 \in J^{(1)}, \dots, j_r \in J^{(r)}} \prod_{s=1}^r \dots \quad \text{by} \quad \prod_{s=1}^r \left( \sum_{j \in J^{(s)}} \dots \right),$$

we find that the measure  $\mathbb{Q}_H^{X_r}$  has a density on  $\mathcal{P}(D_H)$ , given by (39).

### Appendix A. Regular conditional probability

We recall here the notion of regular conditional probability. Let  $\mathbf{Z} = (Z_1, \dots, Z_p)$  and  $\mathbf{X} = (X_1, \dots, X_n)$ , and let  $\mathcal{B}_{\mathbb{R}_+^p}$  denote the Borel  $\sigma$ -algebra on  $\mathbb{R}_+^p$ . The *regular conditional probability*  $\nu$  of  $\mathbf{Z}$  given  $\sigma(\mathbf{X})$  is a function from  $\mathcal{B}_{\mathbb{R}_+^p} \times \mathbb{R}^n$  to  $[0, 1]$  such that

- (i)  $\nu(\mathbf{x}, \cdot)$  is a probability measure for all  $\mathbf{x} \in \mathbb{R}^n$ ,
- (ii) the function  $\nu(\cdot, E)$  is measurable for all Borel sets  $E \in \mathcal{B}_{\mathbb{R}_+^p}$ ,
- (iii)  $\mathbb{P}(\mathbf{Z} \in E, \mathbf{X} \in D) = \int_D \nu(\mathbf{x}, E) \mathbb{P}^{\mathbf{X}}(d\mathbf{x})$  for all  $E \in \mathcal{B}_{\mathbb{R}_+^p}$  and  $D \in \mathcal{B}_{\mathbb{R}^n}$ , where  $\mathbb{P}^{\mathbf{X}}(\cdot) := \mathbb{P}(\mathbf{X} \in \cdot)$ .

See, e.g. Proposition A 1.5.III of [5] for more details.

In Section 2 we provided an expression for the regular conditional probability in the max-linear model (2):

$$\nu(\mathbf{x}, E) := \mathbb{P}(\mathbf{Z} \in E \mid \mathbf{X} = \mathbf{x}), \quad E \in \mathcal{B}_{\mathbb{R}_+^p}, \mathbf{x} \in \mathbb{R}_+^n.$$

The definition of  $\nu$  implies that

$$\int_{\mathbb{R}^p} g(\mathbf{z}) \nu(\mathbf{X}, d\mathbf{z}) = \mathbb{E}(g(\mathbf{Z}) \mid \sigma(\mathbf{X})) \quad \mathbb{P}^{\mathbf{X}}\text{-almost surely}$$

for all Borel functions  $g: \mathbb{R}^p \rightarrow \mathbb{R}$  with  $\mathbb{E}|g(\mathbf{Z})| < \infty$ . By the strong law of large numbers, the latter conditional expectations are readily approximated by  $N^{-1} \sum_{i=1}^N g(\mathbf{Z}^{(i)})$ , where the  $\mathbf{Z}^{(i)}$ ,  $i = 1, \dots, N$ , are independent samples from the regular conditional probability  $\nu(\mathbf{X}, d\mathbf{z})$ . Thus,  $\nu$  is the *right* distribution to sample from when performing prediction, given prior observed data.

### Acknowledgement

The authors were partially supported by NSF grant DMS-0806094 at the University of Michigan.

## References

- [1] BALKEMA, A. A. AND RESNICK, S. I. (1977). Max-infinite divisibility. *J. Appl. Prob.* **14**, 309–319.
- [2] BUISSHAND, T. A., DE HAAN, L. AND ZHOU, C. (2008). On spatial extremes: with application to a rainfall problem. *Ann. Appl. Statist.* **2**, 624–642.
- [3] CAPRARA, A., TOTH, P. AND FISCHETTI, M. (2000). Algorithms for the set covering problem. *Ann. Operat. Res.* **98**, 353–371.
- [4] COOLEY, D., NYCHKA, D. AND NAVEAU, P. (2007). Bayesian spatial modeling of extreme precipitation return levels. *J. Amer. Statist. Assoc.* **102**, 824–840.
- [5] DALEY, D. J. AND VERE-JONES, D. (2008). *An Introduction to the Theory of Point Processes*, Vol. II, 2nd edn. Springer, New York.
- [6] DAVIS, R. A. AND RESNICK, S. I. (1989). Basic properties and prediction of max-ARMA processes. *Adv. Appl. Prob.* **21**, 781–803.
- [7] DAVIS, R. A. AND RESNICK, S. I. (1993). Prediction of stationary max-stable processes. *Ann. Appl. Prob.* **3**, 497–525.
- [8] DAVISON, A. C. AND SMITH, R. L. (1990). Models for exceedances over high thresholds. *J. R. Statist. Soc. B* **52**, 393–442.
- [9] DE HAAN, L. (1978). A characterization of multidimensional extreme-value distributions. *Sankhyā A* **40**, 85–88.
- [10] DE HAAN, L. (1984). A spectral representation for max-stable processes. *Ann. Prob.* **12**, 1194–1204.
- [11] DE HAAN, L. AND FERREIRA, A. (2006). *Extreme Value Theory*. Springer, New York.
- [12] DE HAAN, L. AND PEREIRA, T. T. (2006). Spatial extremes: models for the stationary case. *Ann. Statist.* **34**, 146–168.
- [13] DE HAAN, L. AND PICKANDS, J., III (1986). Stationary min-stable stochastic processes. *Prob. Theory Related Fields* **72**, 477–492.
- [14] FURRER, R., NYCHKA, D. AND SAIN, S. (2009). *Fields: Tools for Spatial Data*. R package version 6.01.
- [15] GINÉ, E., HAHN, M. G. AND VATAN, P. (1990). Max-infinitely divisible and max-stable sample continuous processes. *Prob. Theory Related Fields* **87**, 139–165.
- [16] KABLUCHKO, Z., SCHLATHER, M. AND DE HAAN, L. (2009). Stationary max-stable fields associated to negative definite functions. *Ann. Prob.* **37**, 2042–2065.
- [17] NAVEAU, P., GUILLOU, A., COOLEY, D. AND DIEBOLT, J. (2009). Modelling pairwise dependence of maxima in space. *Biometrika* **96**, 1–17.
- [18] R DEVELOPMENT CORE TEAM (2009). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing Vienna, Austria.
- [19] RESNICK, S. I. (1987). *Extreme Values, Regular Variation, and Point Processes* (Appl. Prob. Trust **4**). Springer, New York.
- [20] RESNICK, S. I. (2007). *Heavy-Tail Phenomena*. Springer, New York.
- [21] RESNICK, S. I. AND ROY, R. (1991). Random USC functions, max-stable processes and continuous choice. *Ann. Appl. Prob.* **1**, 267–292.
- [22] SCHLATHER, M. (2002). Models for stationary max-stable random fields. *Extremes* **5**, 33–44.
- [23] SCHLATHER, M. AND TAWN, J. A. (2003). A dependence measure for multivariate and spatial extreme values: properties and inference. *Biometrika* **90**, 139–156.
- [24] SMITH, R. L. (1990). Max-stable processes and spatial extremes. Unpublished manuscript.
- [25] SRIVASTAVA, S. M. (1998). *A Course on Borel Sets* (Graduate Texts Math. **180**). Springer, New York.
- [26] STOEV, S. A. AND TAQQU, M. S. (2005). Extremal stochastic integrals: a parallel between max-stable and alpha-stable processes. *Extremes* **8**, 237–266.
- [27] WANG, Y. (2010). *Maxlinear: Conditional Sampling for Max-Linear Models*. R package version 1.0.
- [28] WANG, Y. AND STOEV, S. A. (2010). On the structure and representations of max-stable processes. *Adv. Appl. Prob.* **42**, 855–877.