Conditional simulation of a positive random vector subject to max-linear contraints. A geometric perspective

Christian Lantuéjoul, Francis Maisonneuve, Jean-Noël Bacro, and Liliane Bel

Abstract Predicting natural phenomena modeled by max-stable random fields with Fréchet margins is not simple because these models do not possess finite first and second order moments. In such situations, a Monte Carlo approach based on conditional simulations can be considered. In this paper we examine a recent algorithm set up by Wang and Stoev to conditionally simulate a max-stable random field with discrete spectrum. Besides presenting this algorithm, we provide it with a geometric interpretation and put emphasis on several implementation details to obviate its combinatorial complexity. Along the way, a number of other critical issues are mentioned that are not often addressed in the current practice of conditional simulations. An illustrative example is given.

1 Introduction

Because they are infinitely divisible for the maximum, max-stable random fields are often used to model phenomena where extreme situations can occur. Although their statistical inference has been extensively studied (see, for instance, [1] and references therein), their prediction remains a challenging problem, and is a topic of ongoing research [3, 4]. Indeed, a standard approach like kriging is not applicable as these models may not possess finite first and second order moments. In such a situation, a Monte Carlo approach based on conditional simulations is a possible recourse. This paper examines an algorithm designed by Wang and Stoev [6] to conditionally simulate a prototype of max-stable random field. This model is defined as

$$Z(s) = \max_{j=1,\ldots,p} \varphi_j(s) X_j \qquad s \in \mathbb{R}^d \tag{1}$$

Christian Lantuéjoul: MinesParisTech, 35 rue Saint-Honoré, 77300 Fontainebleau, France
Francis Maisonneuve: MinesParisTech, 60 boulevard Saint-Michel, 75272 Paris Cedex 06, France
Jean-Noël Bacro: Université Montpellier II, 4 place Eugène Bataillon, 34095 Montpellier, France
Liliane Bel: AgroParisTech, 16 rue Claude Bernard, 75231 Paris, France

Ninth International Geostatistics Congress, Oslo, Norway, June 11. – 15., 2012

where the φ_j's are p deterministic, non-negative functions and the X_j's are positive and independent random variables with c.d.f. F_j and p.d.f. f_j. Of particular interest is the case where the F_j's are standard unit Fréchet distributions, i.e. $F_j(z) = \exp(-1/z)$, because formula (1) can be used to approximate arbitrarily closely any max-stable random field with unit Fréchet margins [2]. The constraints are $Z(s_i) = z_i, i = 1,...,n$.

In their paper, Wang and Stoev give an algebraic presentation of their algorithm. Our presentation is more geometric and provides complementary results. Although unchanged, the stochastic part of the algorithm is also described because of the important concept of *regular conditional distributions*. Finally, the practical implementation of the algorithm raises a number of combinatorial issues that can be more easily handled using our graph approach.

In the following, we use $\wedge$ and $\vee$ for min and max, and $[\![1,n]\!]$ for $\{1,...,n\}$.

2 Geometric aspects

Let us put $a_{ij} = \varphi_j(s_i)$ for short. The problem addressed is how to simulate the X_j's satisfying the max-linear set of equalities

$$\bigvee_{j=1}^{p} a_{ij} X_j = z_i \qquad i \in [\![1,n]\!].$$

For conditional simulations to exist, there must be solutions to the system of equations

$$\bigvee_{j=1}^{p} a_{ij} x_j = z_i \qquad i \in [\![1,n]\!] \tag{2}$$

If $x = (x_1,...,x_p)$ is a solution, then we clearly have

$$x_j \leq \hat{x}_j = \bigwedge_{i=1}^{n} \frac{z_i}{a_{ij}} \qquad j \in [\![1,p]\!]$$

Note that equation i in (2) is satisfied by x iff there exists an index j for which we have $a_{ij}x_j = z_i$. Because $a_{ij}\hat{x}_j \leq z_i$ and $x_j \leq \hat{x}_j$, it immediately follows $x_j = \hat{x}_j$. Then consider $J_i = \{ j \in [\![1,p]\!] : a_{ij}\hat{x}_j = z_i \}$ as well as $J = \{ j \in [\![1,p]\!] : x_j = \hat{x}_j \}$. x is a solution of (2) iff $J \cap J_i \neq \emptyset$ for each $i \in [\![1,n]\!]$. J is called a *hitting family*.

Let $\mathscr{J}$ be the set of hitting families of (2). Then it can be shown [5] that

$$\forall i \in [\![1,n]\!] \ \bigvee_{j=1}^{p} a_{ij} x_j = z_i \iff x \in \bigcup_{J \in \mathscr{J}} C_J,$$

C_J denoting the face (or cell) of the polytope $\{o \le x \le \hat{x}\}$ of equation $x_j = \hat{x}_j$ for $j \in J$ and $0 \le x_k \le \hat{x}_k$ for $k \in J^c$. In particular, the system (2) admits solutions iff $\hat{x}$ is itself a solution.

3 Stochastic aspects

Here the second member z of (2) is interpreted as a realization of random vector, say Z. Its distribution is specified by that of X:

$$F_Z(z) = P\left\{\bigvee_{i=1}^{n} \frac{Z_i}{z_i} < 1\right\} = \prod_{j=1}^{p} F_j\left(\bigwedge_{i=1}^{n} \frac{z_i}{a_{ij}}\right)$$

One of the important results established by Wang and Stoev [6] is the fact that each of the n equations is F_Z-a.s. satisfied by a single component of X. A consequence is that the conditional realisations of X cannot be found in all faces C_J of $\mathscr{J}$, but only in those of maximal dimension, corresponding to the set $\mathscr{J}_m$ of hitting families with minimal cardinality. Explicitly, the conditional distribution of X given $Z = z$ can be written as

$$dF_{X|Z=z}(x) = \sum_{J \in \mathscr{J}_m} w_J \, dF_J(x) \tag{3}$$

where

$$w_J = \frac{\prod_{j\in J} \hat{x}_j f_j(\hat{x}_j) \prod_{k\in J^c} F_k(\hat{x}_k)}{\sum_{J\in\mathscr{J}_m} \prod_{j\in J} \hat{x}_j f_j(\hat{x}_j) \prod_{k\in J^c} F_k(\hat{x}_k)} \qquad J \in \mathscr{J}_m \tag{4}$$

is the conditional probability that X belongs to F_J, and

$$dF_J(x) = \prod_{j\in J} d\delta_{\hat{x}_j}(x_j) \prod_{k\in J^c} \frac{f_k(x_k)}{F_k(\hat{x}_k)} 1_{x_k<\hat{x}_k} \qquad x \in F_J \tag{5}$$

is the distribution of X given that it belongs to F_J. Of course, the coefficients $\hat{x}_j$ depend on z.

The reader may be surprised by the presence of the products $\hat{x}_j f_j(\hat{x}_j)$ in the expression of w_J. In fact, the conditional distributions $dF_{X|Z=z}$ are supported by negligible sets, so their definition is a bit conventional. Nonetheless, they cannot be arbitrary. In particular, they must allow retrieving the distribution of X when Z is randomized:

$$\int_0^\infty P\{X \in A \mid Z = z\}\, dF_z(z) = P\{X \in A\} \tag{6}$$

Conditional distributions that satisfy (6) are said to be *regular*. The expression of regular conditional distribution may be sometimes surprising. Consider for instance the equation $Z = a_1X_1 \vee a_2X_2$, where X_1 and X_2 are uniformly distributed on $]0,1[$. Suppose $Z = z < a_1 \wedge a_2$, so that $\hat{x}_1 = z/a_1 < 1$ and $\hat{x}_2 = z/a_2 < 1$. Then the support of $X \mid Z = z$ is the union of two segments, namely $\{\hat{x}_1\} \times [0,\hat{x}_2]$ and $[0,\hat{x}_1] \times \{\hat{x}_2\}$. We could imagine that the conditional distribution of X would be uniform on the union of both segments, in which case each segment would have a chance of containing X that is proportional to its length. In fact the regular conditional distribution specified by (4) recommends that the same chance should be assigned to both segments.

4 Combinatorial aspects

Following the explicit formula for the conditional distribution of X given $Z = z$, the algorithm is as follows:

(i) generate $J \sim w_J$;
(ii) generate $x \sim dF_j$.

In this algorithm, the difficult part is the generation of J which requires $\mathscr{J}_m$ to be fully identified. In their paper [6], Wang and Stoev turn this problem into a *set covering problem* that is known to be NP-hard. Using a graph approach [5], this problem can be notably simplified.

Consider the following *bipartite graph*. It has $n+p$ vertices, one per equation and one per variable. The vertex $i \in [\![1,n]\!]$ is connected to the vertex $j \in [\![1,p]\!]$ if $j \in J_i$. This bipartite graph gives rise to a new graph G whose vertices are $[\![1,n]\!]$. Two vertices i and i' are connected in G if $J_i \cap J_{i'} \neq \emptyset$. It can be shown that the graph G has several connected components $G_1,...,G_r$ that are *cliques* (subgraphs saturated with edges). Their number r is nothing but the minimal cardinality of the hitting families. Moreover, each minimal hitting family $\{j_1,...,j_r\}$ is obtained by picking j_1 connected to all vertices of G_1, j_2 connected to all vertices of G_2 etc. in the bipartite graph.

As an example, consider the max-linear system

$$\begin{cases} x_1 \vee x_3 = z_1 \\ x_2 \vee x_3 = z_2 \end{cases}$$

– If $z_1 < z_2$, then $\hat{x}_1 = \hat{x}_3 = z_1$ and $\hat{x}_2 = z_2$. Thus $J_1 = \{1,3\}$ and $J_2 = \{2\}$. The hitting families are $\{1,2\},\{2,3\}$ and $\{1,2,3\}$, so that the minimal families are $\{1,2\}$ and $\{2,3\}$ with cardinality $r = 2$. Accordingly $G_1 = \{1\}$ and $G_2 = \{2\}$.
– Similar results are obtained when $z_2 < z_1$ by swapping the indices 1 and 2.
– If $z_1 = z_2 = z$, then $\hat{x}_1 = \hat{x}_2 = \hat{x}_3 = z$. This implies $J_1 = \{1,3\}$ and $J_2 = \{2,3\}$. The hitting families are $\{3\},\{1,2\},\{1,3\},\{2,3\}$ and $\{1,2,3\}$. There is only one minimal hitting family, namely $\{3\}$, and its cardinality is $r = 1$. Hence $G_1 = \{1,2\}$.

As the cardinality of $\mathscr{J}_m$ may be large, a direct generation of J using (4) is not always possible. In this case, w can be easily simulated as the limit distribution of a Markov chain using the independent sampler variation of Metropolis algorithm.

Here is the corresponding algorithm, G_k denoting the k^{th} connected component of G:

(i) initialize $J \in \mathscr{J}_m$;
(ii) for each $k \in [\![1,r]\!]$ *generate* $j'_k \sim \mathscr{U}\left(\cap_{i \in G_k} J_i\right)$, *and put* $J' = (j'_1, \ldots, j'_r)$;
(iii) generate $u \sim \mathscr{U}$ *and put* $J = J'$ *if* $u < \prod_{j \in J} \hat{x}_j / \prod_{j' \in J'} \hat{x}_{j'}$;
(iv) goto (ii).

5 Example

In this exercise, the model is built using $p = 1000$ basic functions. It takes the form

$$Z(s) = \bigvee_{j=1}^{p} \frac{X_j}{1 + |s - s_j|^2} \qquad s \in \mathbb{R}^2$$

where the X_j's are independent standard unit Fréchet variables, and the s_j's are

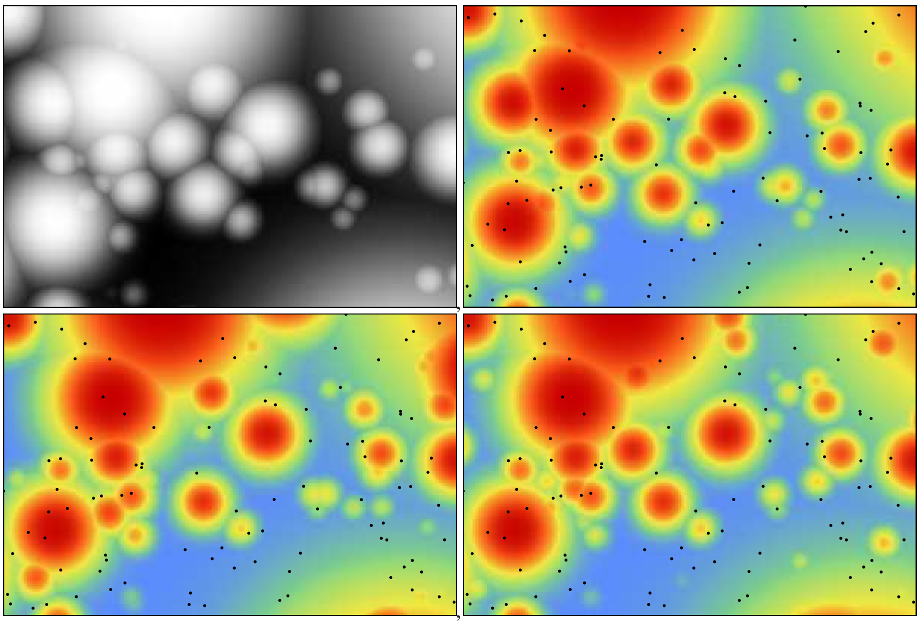

Fig. 1 Top left, a (non conditional) simulation. Top right, the same simulation with conditioning data points. Bottom, two conditional simulations

known points, generated at random in a domain 50×40 containing the 30×20 simulation field. Top left of figure 1 shows a (non conditional) simulation of the model. It is displayed using a uniform anamorphosis to emphasize the contrasts. From this simulation, 100 points have been selected at random to serve as conditioning data points for the exercise (cf. top right of figure 1). Running the algorithm, it appears that the graph G has 25 connected components. Their sizes range from 1 to 9. Overall, the cardinality of $\mathscr{J}_m$ exceeds 220 000. Two conditional simulations are depicted at the bottom of figure 1.

Despite its iterative component, the algorithm turns out to be very fast (a few seconds are required to produce each conditional simulation). Accordingly, 100 conditional simulations have been carried out to derive estimates of the conditional mean and the conditional standard deviation. The results are reproduced on figure 2.

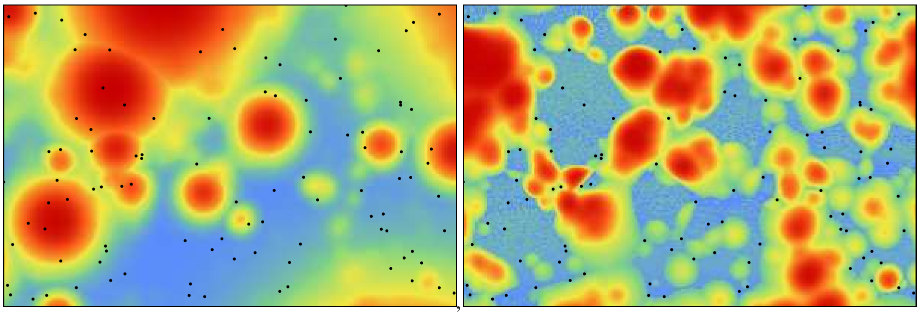

Fig. 2 Estimates of the conditional mean (left) and the conditional standard deviation (right)

Acknowledgements: This work was supported by the ANR project McSim. We are grateful to Marco Oesting and Yizao Wang for helpful comments.

References

1. Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J.: Statistics of extremes – Theory and applications. Wiley, Chichester (2004)
2. de Haan, L.: A spectral representation for max-stable processes. The Annals of Prob., **12-4**, 1194-2004 (1984)
3. Dombry, C., Eyi-Minko, F.: Regular conditional distributions of max infinitely divisible processes. arXiv:1109.6492v2 (2011)
4. Dombry, C., Eyi-Minko, F., Ribatet, M.: Conditional simulations of Brown-Resnick processes. arXiv:1112.3891v1 (2011)
5. Lantuéjoul, C., Maisonneuve, F.: Simulation conditionnelle d'un modèle max-stable. Unpublished manuscrit, MinesParisTech (2011)
6. Wang Y., Stoev, S.A.: Conditional sampling for spectrally discrete max-stable random fields. Adv. in Appl. Prob., **43-2**, 461-483 (2011)