BAYESIAN INFERENCE FOR THE BROWN–RESNICK PROCESS, WITH AN APPLICATION TO EXTREME LOW TEMPERATURES[1]

By Emeric Thibaud[*], Juha Aalto[†], Daniel S. Cooley[*],
Anthony C. Davison[‡] and Juha Heikkinen[§]

*Colorado State University[*], Finnish Meteorological Institute[†],
Ecole Polytechnique Fédérale de Lausanne[‡] and
Natural Resources Institute Finland[§]*

The Brown–Resnick max-stable process has proven to be well suited for modeling extremes of complex environmental processes, but in many applications its likelihood function is intractable and inference must be based on a composite likelihood, thereby preventing the use of classical Bayesian techniques. In this paper we exploit a case in which the full likelihood of a Brown–Resnick process can be calculated, using componentwise maxima and their partitions in terms of individual events, and we propose two new approaches to inference. The first estimates the partitions using declustering, while the second uses random partitions in a Markov chain Monte Carlo algorithm. We use these approaches to construct a Bayesian hierarchical model for extreme low temperatures in northern Fennoscandia.

1. Introduction. Extreme events such as heat or cold waves have major impacts on human health and ecological sustainability. Motivated by the pressing need for better risk assessment, inference for spatial extremes has developed rapidly during the last decade [e.g., Davison, Huser and Thibaud (2013), Davison, Padoan and Ribatet (2012), de Haan and Ferreira (2006), Ribatet (2013)]. Max-stable processes are the only possible nondegenerate limits for rescaled pointwise maxima of random processes, thus motivating their use in modeling spatial extremes.

Flexible max-stable models have been developed for inference on spatial extremes. Over a variety of applications, the Brown–Resnick process [Brown and Resnick (1977), Kabluchko, Schlather and de Haan (2009)] has proven to be a valuable model for extremes of environmental variables, such as rainfall, temperature, wind speed and river discharge [see, e.g., Asadi, Davison and Engelke (2015), Buhl and Klüppelberg (2016), Davison, Huser and Thibaud (2013), Engelke et al.

(2015)]. It is built using Gaussian processes and its dependence structure is determined by a variogram, and so well-established ideas from classical geostatistics may be exported to the extremal context. The greater challenge is inference, since the full likelihood of the Brown–Resnick process is tractable only in relatively low-dimensional settings [Castruccio, Huser and Genton (2015), Huser and Davison (2013)]. Composite likelihood techniques have been successfully used to fit max-stable models to block maxima or threshold exceedances, but with reduced estimation efficiency. More efficient procedures based on full, possibly censored, likelihoods [Engelke et al. (2015), Thibaud and Opitz (2015), Wadsworth and Tawn (2014)] have been developed for threshold exceedances, allowing the construction of more elaborate models for spatial extremes.

Bayesian hierarchical models are appealing in spatial statistics because of the flexibility with which they can track the spatial variation of marginal distributions. Casson and Coles (1999), Cooley, Nychka and Naveau (2007) and Sang and Gelfand (2009) constructed hierarchical models for extremes assuming conditional independence of extreme occurrences. Such models may have flexible margins but cannot produce realistic simulations of extreme events [Davison, Padoan and Ribatet (2012)]. Sang and Gelfand (2010) and Fuentes, Henry and Reich (2013) used hierarchical models that model spatial dependence using Gaussian processes, a somewhat restrictive class of models that are asymptotically independent and thus may underestimate potential dependence at the most extreme levels [Davison, Huser and Thibaud (2013)]. Ribatet, Cooley and Davison (2012) and Shaby (2014) proposed the use of composite likelihoods in Markov chain Monte Carlo (MCMC) algorithms and applied them for Bayesian inference on max-stable models, but these methods present computational challenges and do not yield exact simulation from the posterior distribution. Shaby and Reich (2012) constructed a max-stable spatial model based on a hierarchical representation of logistic extreme-value distributions [Stephenson (2009)].

In this paper we propose a Bayesian hierarchical model based on the Brown–Resnick process, which exploits a special case for which the full likelihood of such a process can be calculated, using the joint distribution of componentwise maxima and their partitions in terms of individual events [Stephenson and Tawn (2005)]. We discuss the difficulties of using the partitions in practice, and we propose a new Bayesian approach to fitting max-stable processes without fixed partitions.

Our work is motivated by the changing climate of northern Fennoscandia. In northern regions of Norway, Finland, Russia and Sweden, extreme low temperatures play a central role in ecological sustainability by limiting the extent of outbreaks of forest pests such as the autumnal moth *Epirrita autumnata*, whose larvae have caused serious damage by defoliating mountain birch. Forests in the region have mostly been protected from this pest by extreme low winter temperatures, since its eggs cannot survive below $-36°C$ [Virtanen, Neuvonen and Nikula (1998)]. Winter minimum temperatures are often below this threshold, but in recent years they have increased and projections suggest they will increase even more in

the future [IPCC (2013), Section E.1]. Although temperature is not the only factor limiting the spread of the moth, it is one of the most important, and thus plays a central role in the ecology of northern Fennoscandia. In Section 4.5 we discuss the possible changes in winter minimum temperatures in the region, focusing on the probability that they will be less than $-36°C$.

Motivated by recent disastrous heat wave events over the world, max-stable processes and related methods have been used to model temperature extremes and their potential changes in time. Davison and Gholamrezaee (2012) used the so-called Schlather (2002) process to model annual maximum temperatures in Switzerland. They used latitude, longitude and elevation to estimate response surfaces for the marginal parameters of extremes. Fuentes, Henry and Reich (2013) describe a Bayesian spatial model for annual maximum temperatures in the US based on mixtures of Gaussian distributions. Shaby and Reich (2012) used a Bayesian max-stable model to model extreme temperature data over Europe.

Section 2 describes the data we study. Section 3 summarizes some extreme value ideas and introduces the inference methods for the Brown–Resnick model. The hierarchical model and the application are discussed in Section 4.

2. Extreme temperatures in northern Fennoscandia.

2.1. *Dataset.* We use temperature data from 20 meteorological stations located in northern regions of Finland, Norway and Sweden; see Figure 1. The smallest and largest distances between two stations are 26 and 471 km. Daily minimum temperatures are available from 1960 to 2013, but with data missing. In view of the motivation for our application, inference is needed only for the winter minima. We focus on modeling minimum temperatures in the months December–March, and we use the daily data to investigate their co-occurrences.

Missing daily data censor the winter minima. Winter minimum values for the 20 stations are shown in Figure 2. Our dataset has at least 26 winter minima for each station, and each year has values of winter minimum for at least 15 of the 20 stations.

Mean winter minimum temperatures depend on local geography and topography, and we shall model this nonstationarity in their marginal distributions using covariates and random effects, which will also allow us to extrapolate results to ungauged sites.

We derived covariates that are known to influence minimum temperature in northern Fennoscandia [Aalto, le Roux and Luoto (2014)]. We supplemented the coordinates of the stations with their absolute elevation above sea level and elevation relative to the surrounding areas (calculated on a 2 km × 2 km spatial moving average), a proximity index to the Arctic Ocean, calculated with a 160 km × 160 km moving window assuming that the effect of the ocean reaches 80 km inland, and a lake cover index, which expresses the proportion of area covered by lakes in a 3 km × 3 km window [see Aalto, le Roux and Luoto (2014)].
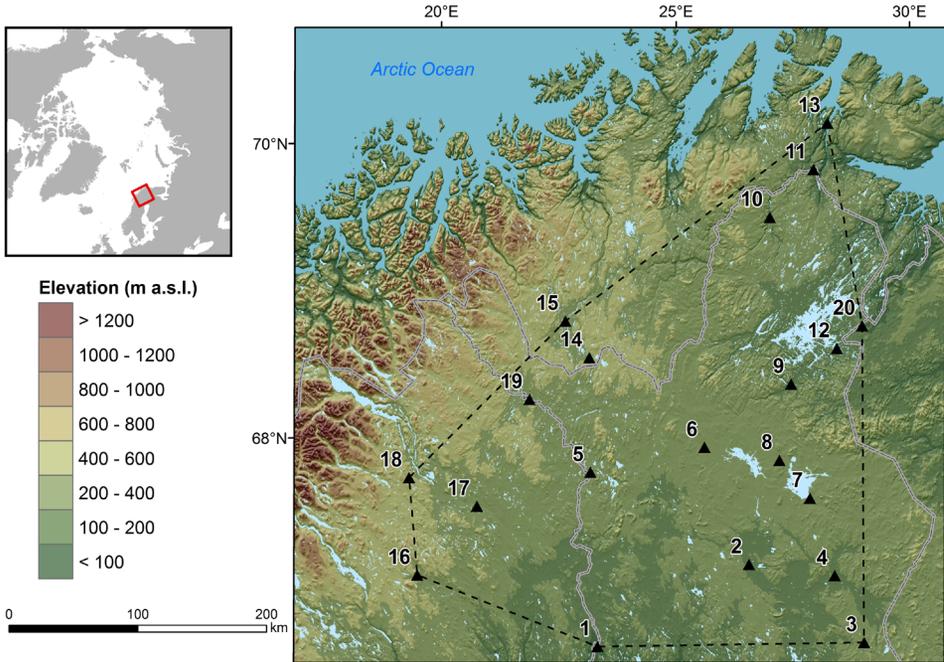
FIG. 1. *Locations of the* 20 *weather stations in the northern regions of Finland, Norway and Sweden. The dashed lines shows the convex hull of the stations, which will be our region of interest for simulation.*

2.2. *Temporal stationarity.* In Figure 2 it is difficult to see the presence of a temporal trend with only 53 years of noisy data, but the figure suggests that minima may generally be slightly less extreme from 1990 onward. The inclusion of a temporal effect in the model will help to assess this.

2.3. *Serial dependence.* Since we focus on modeling the winter minima, temporal dependence in the daily temperatures is not of interest in itself. But our approach requires us to partition minima in terms of independent generating events, which is complicated by the presence of temporal dependence in the daily data; see Section 4.2.3. Cold waves stay over the region for several days, and so minimum temperatures tend to be very dependent from day to day. Preliminary analysis [see the Supplementary Material, Thibaud et al. (2016)] showed that the dependence is strong for the first few days and vanishes after approximately five to ten days at each location.

2.4. *Spatial dependence.* We expect winter minimum temperatures to be dependent over large regions, and this is the case in our data. Figure 2 suggests that the minima for the 20 stations behave similarly for the 53 years: the years 1966 and 1999 were very cold at all 20 locations, whereas 1973, 1993 and 2007 are
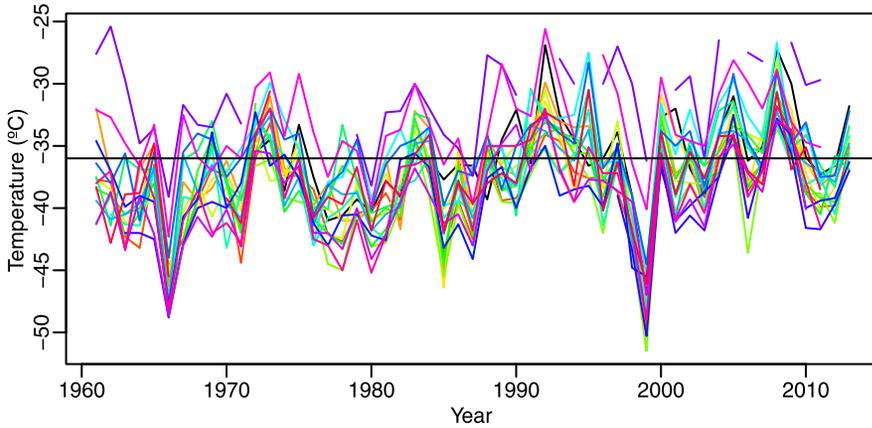
FIG. 2. *Winter minimum temperature (°C) at the 20 stations (a different color for each station).
The horizontal line shows the −36°C threshold.*

warmer at all locations. It appears that most of the minima occurred on the same day or only a few days apart, and so they may correspond to the same event, which suggests strong spatial dependence.

Results from a preliminary analysis [see Thibaud et al. (2016)] suggest that pairwise distributions are asymptotically dependent and max-stable. After accounting for differences in the marginal distributions, we will use a stationary isotropic max-stable model for modeling spatial dependence.

3. Max-stable processes and inference for spatial extremes.

3.1. *Generalities.* Let $Y = \{Y(s)\}_{s \in S}$ be a random process with continuous sample paths defined over a compact subset S of $\mathbb{R}^m$, and let $Y_1, Y_2, \ldots$ be independent replicates of Y. If there exist sequences of continuous functions $\{a_n(s)\}_{s \in S} > 0$ and $\{b_n(s)\}_{s \in S}$ such that we have the distributional convergence

$$\{a_n(s)^{-1}[\max\{Y_1(s), \ldots, Y_n(s)\} - b_n(s)]\}_{s \in S} \to \{Z(s)\}_{s \in S}, \qquad n \to \infty,$$

where $Z = \{Z(s)\}_{s \in S}$ is a nondegenerate random process, we say that Y lies in the max-domain of attraction of the process Z. Then the process Z must be max-stable [de Haan and Ferreira (2006)] and its marginals are of generalized extreme-value (GEV) form,

$$\Pr\{Z(s) \le z\} = \exp\left(-\left[1 + \xi(s)\sigma(s)^{-1}\{z - \mu(s)\}\right]_+^{-1/\xi(s)}\right), \qquad s \in S,$$

denoted $Z(s) \sim \text{GEV}\{\mu(s), \sigma(s), \xi(s)\}$, with $x_+ = \max(0, x)$, and real location $\mu(s)$, scale $\sigma(s) > 0$ and shape $\xi(s)$ parameters; throughout the paper, the case $\xi(s) = 0$ is defined as the limit for $\xi(s) \to 0$. Every max-stable process with marginal parameters $\mu(s)$, $\sigma(s)$ and $\xi(s)$ can be written as $\mu(s) +$

$\sigma(s)\xi(s)^{-1}\{Z(s)^{\xi(s)} - 1\}$, where Z is a simple max-stable process with unit Fréchet margins, that is, $\mu(s) = \sigma(s) = \xi(s) = 1$. Thus, it suffices to consider simple max-stable processes, henceforth denoted by Z.

As a consequence of max-stability, for every set of locations $s_1, \ldots, s_D \in S$,

$$
(1) \qquad
\begin{aligned}
&\Pr\{Z(s_1) \le z_1, \ldots, Z(s_D) \le z_D\} \\
&\quad = \exp\{-V(z_1, \ldots, z_D)\}, \qquad z_1, \ldots, z_D > 0,
\end{aligned}
$$

where the so-called exponent function V is homogeneous of order -1, that is, $V(tz_1, \ldots, tz_D) = t^{-1}V(z_1, \ldots, z_D)$, $t > 0$, and satisfies $V(z, \infty, \ldots, \infty) = z^{-1}$ for any permutation of its arguments. In particular,

$$
\Pr\{Z(s_1) \le z, \ldots, Z(s_D) \le z\} = \exp(-\theta/z), \qquad z > 0,
$$

where $\theta = V(1, \ldots, 1) \in [1, D]$ is known as the extremal coefficient of Z relative to $s_1, \ldots, s_D$. The extremal coefficient can be interpreted as the effective number of independent variables among $Z(s_1), \ldots, Z(s_D)$ and is useful for model checking, since it can be estimated nonparametrically.

As they appear as limits for rescaled pointwise maxima of random processes, max-stable processes are natural models for observed block maxima. Moreover, the equality $\min\{Y_1(s), \ldots, Y_n(s)\} = -\max\{-Y_1(s), \ldots, -Y_n(s)\}$ shows that such processes also provide natural models for block minima.

Several parametric max-stable models have been proposed. The Smith (1990) process was the first to be considered, but its realizations are too smooth for realistic modeling of complex extremes. The Schlather (2002) process is constructed from stationary Gaussian processes and provides more realistic realizations controlled by different correlation functions. The extremal-t process extends the Schlather process by introducing an additional shape parameter [Opitz (2013)]. Here we focus on the Brown–Resnick process [Brown and Resnick (1977), Kabluchko, Schlather and de Haan (2009)], which appears to be a good model for extremes of environmental processes.

3.2. *Brown–Resnick process*.   Kabluchko, Schlather and de Haan (2009) constructed a stationary max-stable process, called the Brown–Resnick process, as

$$
(2) \quad Z(s) = \max_{i=1,2,\ldots} Q_i^{-1} W_i(s), \qquad W_i(s) = \exp\{\varepsilon_i(s) - \sigma(s)^2/2\}, s \in S,
$$

where the Q_i's are the points of a unit rate Poisson process on $(0, \infty)$ and the ε_i are independent replicates of an intrinsically stationary centered Gaussian process ε with variance function $\sigma(s)^2 = \mathrm{E}\{\varepsilon(s)^2\}$. The dependence structure of Z is inherited from that of ε, which is controlled by the variogram $2\gamma(h) = \mathrm{E}[\{\varepsilon(s+h) - \varepsilon(s)\}^2]$. Without loss of generality, we suppose that $\varepsilon(0) = 0$ with probability one, and so $\sigma(s)^2 = 2\gamma(s)$. A large class of models may be obtained by choosing different variograms. Later we use the stable variogram $2\gamma(h) = 2(\|h\|/\lambda)^\kappa$, with $\lambda > 0$ and $\kappa \in (0, 2]$.

The exponent function of a Brown–Resnick process is [Nikoloulopoulos, Joe and Li (2009)]

$$
(3) \quad V(z_1, \ldots, z_D) = \sum_{j=1}^{D} z_j^{-1} \Phi_{D-1} \{ \log(\mathbf{z}_{-j}/z_j);
$$

$$
\mathbf{\Gamma}_{-j,j}, \mathbf{\Gamma}_{-j,j} \mathbf{1}'_{D-1} + \mathbf{\Gamma}_{j,-j} \mathbf{1}_{D-1} - \mathbf{\Gamma}_{-j,-j} \},
$$

where $\Phi_p(\cdot; \boldsymbol{\mu}, \boldsymbol{\Omega})$ is the cumulative distribution function of a p-dimensional Gaussian distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Omega}$, $\mathbf{1}_{D-1} = (1, \ldots, 1)' \in \mathbb{R}^{D-1}$, $\mathbf{\Gamma} = \{\gamma(\boldsymbol{s}_{j_1} - \boldsymbol{s}_{j_2})\}_{1 \le j_1, j_2 \le D}$, $\mathbf{\Gamma}_{-j,j}$ denotes the $(D-1) \times 1$ vector consisting of the jth column of $\mathbf{\Gamma}$ without its jth component, $\mathbf{\Gamma}_{-j,-j}$ denotes the matrix $\mathbf{\Gamma}$ without its jth row and jth column, and so forth.

Let $\mathbf{Z} = \{Z(\boldsymbol{s}_1), \ldots, Z(\boldsymbol{s}_D)\}$ be a vector stemming from a simple max-stable process Z. Differentiation of (1) shows that the full likelihood contribution for one realization $(z_1, \ldots, z_D)$ of $\mathbf{Z}$ may be written as

$$
(4) \quad \sum_{\pi = (\pi_1, \ldots, \pi_K) \in \mathcal{P}} \left[\prod_{k=1}^{K} \{ -V_{\pi_k}(z_1, \ldots, z_D) \} \right] \exp\{ -V(z_1, \ldots, z_D) \},
$$

where $\mathcal{P}$ denotes the set of all partitions of $\{1, \ldots, D\}$ and the subscript π_k on V_{π_k} denotes the partial derivative of the function V with respect to the components in the subset π_k. Although the likelihood has a closed-form expression for the Brown–Resnick process [Wadsworth and Tawn (2014)], the size of $\mathcal{P}$ equals the Bell number and is larger than 10^5 for $D \ge 10$, making full likelihood inference computationally unattainable in many situations. Hence, inference for max-stable processes has typically been based on composite likelihoods using lower-order densities [Castruccio, Huser and Genton (2015), Padoan, Ribatet and Sisson (2010)]. Not only do composite likelihoods result in a loss of efficiency compared to full likelihood estimation, they also increase the computational burden when employed in a Bayesian analysis because the composite likelihood does not directly yield a posterior which appropriately accounts for parameter uncertainty [Ribatet, Cooley and Davison (2012), Shaby (2014)]. Engelke et al. (2015), Wadsworth and Tawn (2014) and Thibaud and Opitz (2015) developed full likelihood inferences for threshold exceedances of processes in the max-domain of attraction of a Brown–Resnick process, which, while challenging, have lower computational burdens than does (4).

For our likelihood, we employ the result of Wadsworth and Tawn (2014), who showed that, for the Brown–Resnick process, partial derivatives of V with respect

to the first d indices, denoted $V_{1:d}$, satisfy

$$
-V_{1:d}(z) = \frac{\Phi_{D-d}(\log z_{(d+1):D}; \tilde{\tau}, \tilde{\Gamma})}{(2\pi)^{(d-1)/2}|\Sigma_{1:d}|^{1/2}(\mathbf{1}_d'\mathbf{q}_d)^{1/2}\prod_{j=1}^d z_j}
$$

$$
\times \exp\left\{-\frac{1}{2}\left(\frac{1}{4}\sigma_d'\Sigma_{1:d}^{-1}\sigma_d - \frac{1}{4}\frac{\sigma_d'\mathbf{q}_d\mathbf{q}_d'\sigma_d}{\mathbf{1}_d'\mathbf{q}_d} + \frac{\sigma_d'\mathbf{q}_d}{\mathbf{1}_d'\mathbf{q}_d} - \frac{1}{\mathbf{1}_d'\mathbf{q}_d}\right)\right\}
$$

(5)

$$
\times \exp\left[-\frac{1}{2}\left\{\log z_{1:d}'\mathbf{A}_{1:d}\log z_{1:d} + \log z_{1:d}'\left(\frac{2\mathbf{q}_d}{\mathbf{1}_d'\mathbf{q}_d}\right)\right.\right.
$$

$$
\left.\left. + \Sigma_{1:d}^{-1}\sigma_d - \frac{\mathbf{q}_d\mathbf{q}_d'\sigma_d}{\mathbf{1}_d'\mathbf{q}_d}\right\}\right],
$$

with $\tilde{\Gamma} = (\mathbf{K}_{01}'\mathbf{A}\mathbf{K}_{01})^{-1}$, $\tilde{\tau} = -\tilde{\Gamma}(\mathbf{K}_{01}'\mathbf{A}\mathbf{K}_{01}\log z_{1:d} + \mathbf{K}_{01}'\Sigma^{-1}\mathbf{1}_D/\mathbf{1}_D'\Sigma^{-1}\mathbf{1}_D)$, $\Sigma = \{\gamma(\mathbf{s}_{j_1}) + \gamma(\mathbf{s}_{j_2}) - \gamma(\mathbf{s}_{j_1} - \mathbf{s}_{j_2})\}_{1 \le j_1, j_2 \le D}$, $\Sigma_{1:d}$ the $d \times d$ matrix consisting of the first d rows and columns of Σ, $\mathbf{A} = \Sigma^{-1} - \mathbf{q}\mathbf{q}'/\mathbf{1}_D'\mathbf{q}$, $\mathbf{A}_{1:d} = \Sigma_{1:d}^{-1} - \mathbf{q}_d\mathbf{q}_d'/\mathbf{1}_d'\mathbf{q}_d$, $\mathbf{q} = \Sigma^{-1}\mathbf{1}_D$, $\mathbf{q}_d = \Sigma_{1:d}^{-1}\mathbf{1}_d$, $\sigma = \mathrm{diag}(\Sigma)$ and $\sigma_d = \sigma_{1:d}$, where

$$
\mathbf{K}_{10} = \begin{pmatrix} \mathbf{I}_d \\ \mathbf{0}_{D-d,d} \end{pmatrix}, \qquad \mathbf{K}_{01} = \begin{pmatrix} \mathbf{0}_{d,D-d} \\ \mathbf{I}_{D-d} \end{pmatrix}.
$$

3.3. *The Stephenson–Tawn model.* Stephenson and Tawn (2005) derived a likelihood based on componentwise block maxima and their occurrence times that avoids direct differentiation of (1). Let $\mathbf{Y}_1, \mathbf{Y}_2, \ldots$ denote independent replicates of a random vector $\mathbf{Y} = \{Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_D)\} \in \mathbb{R}^D$. Without loss of generality, suppose that the process Y has unit Fréchet margins and lies in the max-domain of attraction of a simple max-stable process Z. Let $\mathbf{M}_n = \max_{i=1,\ldots,n} \mathbf{Y}_i/n$ denote the componentwise maxima, and let $\Pi_n = (\pi_1, \ldots, \pi_K)$ denote the unique partition of $\{1, \ldots, D\}$ such that $\{M_{n,j}\}_{j \in \pi_k}$ comes from the same individual observation, that is, for each $k \in \{1, \ldots, K\}$, there exists i_k such that $\{M_{n,j}\}_{j \in \pi_k} = \{Y_{i_k,j}/n\}_{j \in \pi_k}$. Then Stephenson and Tawn (2005) showed that, as $n \to \infty$, the block maximum and its associated partition $(\mathbf{M}_n, \Pi_n)$ converge in distribution to $(\mathbf{Z}, \Pi)$, whose joint density is

(6) $$f(\mathbf{Z}, \Pi) = \exp\{-V(\mathbf{Z})\}\prod_{\pi_k \in \Pi}\{-V_{\pi_k}(\mathbf{Z})\}.$$

This suggests using (6) as a likelihood for inference on the joint distribution of block maxima and their occurrence times.

This approach only requires the maximum and the partition to be observed, and thus represents a compromise between the block maximum and threshold exceedance approaches. The relative efficiency of Stephenson–Tawn and pairwise maximum likelihood estimators has been investigated by Wadsworth and Tawn (2014). Because the partition contains information about the dependence,

the Stephenson–Tawn estimator has smaller variance than that based only on the block maxima. However, for finite block length, the former may be more biased because the model $(\boldsymbol{Z}, \Pi)$ is misspecified for $(\boldsymbol{M}_n, \Pi_n)$, that is, both for the maximum and the partition [Wadsworth (2015)]. Use of the limiting distribution for $(\boldsymbol{M}_n, \Pi_n)$ represents a strong assumption that can lead to poor fit, whereas approaches based on the maximum $\boldsymbol{M}_n$ alone are more robust.

A further difficulty arises from the definition of the partition for serially dependent data. The partition in the Stephenson–Tawn model is defined in terms of independent generating events. While it is natural to define the partition in terms of the days of occurrence of the maxima when the daily observations are independent, its estimation is harder when data show temporal dependence. In Section 4.2.3 we propose a space-time declustering procedure to estimate the partition.

3.4. *Random partition model.* Although the use of the estimated fixed partition is theoretically appealing, it represents a practical challenge. First, the model is misspecified, which affects the estimation of the max-stable model by the introduction of potentially biased information. Second, the partition must be estimated in a preliminary step, and inference based on the Stephenson–Tawn model is conditional on the fixed partition and ignores the uncertainty associated to its estimation. The partition is not always very well defined (see Section 4.2.3), and its estimation is sensitive to measurement error. This motivates us to propose an approach to estimating the parameters of the Brown–Resnick process, without fixing the partition, which uses only the values of the maxima and can be applied when the occurrence dates are unavailable.

Let $\boldsymbol{Z} = \{Z(\boldsymbol{s}_1), \ldots, Z(\boldsymbol{s}_D)\}$ be a vector stemming from a Brown–Resnick process Z. To overcome the intractability of the likelihood (4), we propose to augment the data with the unknown partition Π that generates $\boldsymbol{Z}$ in terms of the functions $P_i = Q_i^{-1} W_i(\boldsymbol{s})$ in (2), that is, $\Pi = (\pi_1, \ldots, \pi_K)$ is defined such that the points $Z(\boldsymbol{s}_j)$ in the same set π_k come from the same function P_i. The joint distribution of $(\boldsymbol{Z}, \Pi)$ is (6).

We impute the unknown partition Π in an MCMC algorithm, using ideas similar to Dombry, Éyi-Minko and Ribatet (2013), who considered the partition in the context of conditional simulation from max-stable processes. The partition is imputed from its conditional distribution given the observed value of $\boldsymbol{Z}$. From (6) we have

$$(7) \qquad \Pr(\Pi = \pi \mid \boldsymbol{Z} = z) = \frac{f(z, \pi)}{\sum_{\tilde{\pi} \in \mathcal{P}} f(z, \tilde{\pi})}.$$

However, the number of partitions of the set $\{1, \ldots, D\}$ is the Dth Bell number, and so (7) is intractable for large D. We use a Gibbs sampler to update the partition. For a partition $\pi = (\pi_1, \ldots, \pi_K)$, we choose a location $\boldsymbol{s}_j$ at random, and let π_{-j} denote the restriction of π to $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_D\} \setminus \{\boldsymbol{s}_j\}$. The size of π_{-j} is K, or $(K-1)$ if $\{\boldsymbol{s}_j\}$ is itself a partitioning set of π. Given π_{-j}, the partition is updated by

attributing the location s_j either to an existing set of π_{-j} or to a new set, giving at most $K+1$ choices. The conditional probability for a new partition π^* with $\pi^*_{-j} = \pi_{-j}$ is

$$(8) \qquad \Pr\big(\Pi = \pi^* \mid \Pi_{-j} = \pi_{-j}, \mathbf{Z} = z\big) = \frac{f(z, \pi^*)}{\sum_{\tilde{\pi} \in \mathcal{P}:\tilde{\pi}_{-j} = \pi_{-j}} f(z, \tilde{\pi})}.$$

The discrete distribution (8) has support on a set of size at most $K+1$, and so a new partition π^* can easily be simulated.

In practice, inference is based on independent replicates $\mathbf{Z}_1, \ldots, \mathbf{Z}_N$, and the MCMC algorithm independently updates the corresponding latent partitions $\Pi_1, \ldots, \Pi_N$. Given the observed values of the $\mathbf{Z}_i$ and the simulated Π_i, the marginal and dependence parameters of the model are updated from the joint distribution of the $(\mathbf{Z}_i, \Pi_i)$, which is a product of terms of the form (6).

4. Modeling extreme temperatures in northern Fennoscandia.

4.1. *Exploratory analysis.* We first fitted GEV distributions separately at each of the 20 stations; see Thibaud et al. (2016). The estimates for the scale and shape parameters were not significantly different from site to site, but variation in the location parameter estimates suggested use of a nonstationary GEV model with spatially varying location parameter and constant scale and shape parameters.

We then fitted nonstationary marginal models at the 20 locations by maximizing the independence likelihood, which assumes the independence of the temperatures from site to site. Nonstationarity was modeled using trend surfaces defined from the covariates mentioned in Section 2.1 and temporal trends in the location and scale parameters. We used the R package SpatialExtremes [Ribatet (2015)] to fit these models, which were compared using the Takeuchi information criterion, an equivalent of the Akaike information criterion for misspecified models. Our best model had all six covariates for the location surface, constant scale and shape parameters, and a temporal trend in the location only. Its fit was assessed using marginal QQ-plots: departure from the 95% confidence intervals was observed for five stations. This suggests that the trend surface model is not flexible enough to model the variability in the location parameters over the 20 stations. Inclusion of additional covariates did not improve the fit, motivating the use of a more flexible approach using random effects in a Bayesian hierarchical model.

4.2. *Hierarchical model.*

4.2.1. *Definition.* We consider two Bayesian hierarchical models based on the Brown–Resnick process. The first uses the Stephenson–Tawn model and estimates the partitions in a preliminary step through space-time declustering; see Sections 3.3 and 4.2.3. The second imputes the partitions in the MCMC algorithm;

see Section 3.4. Here we describe the part of the hierarchical model for the data conditional on the partitions, which is common to both approaches.

Let $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{iD})$ $(i = 1, \ldots, N)$ denote the observed minimum temperatures for the N years of data, let $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_D$ denote the $D = 20$ locations of the stations such that $y_{ij} = y_i(\boldsymbol{s}_j)$, let $t_{ij} \in \mathbb{R}$ denote the covariate for the day of occurrence of the winter minimum y_{ij}, defined as the number of years from 1 January 2000, and, finally, let $\pi_i = (\pi_{i1}, \ldots, \pi_{iK_i})$ denote the partitions of $\{1, \ldots, D\}$ for each year. When data are missing, the dimension of $\boldsymbol{y}_i$ is lower than D and so is the size of the corresponding partition. The first level of the hierarchical model gives the probability density of the observed minima and the partitions $\pi_1, \ldots, \pi_N$, conditional on the marginal and dependence parameters $\boldsymbol{\psi}$, as

(9)
$$
\begin{aligned}
&L(\{\boldsymbol{y}_i\}, \{\pi_i\} \mid \boldsymbol{\psi}) \\
&\quad = \prod_{i=1}^{N} \Bigg(\exp\big[-V\{f_{i1}(-y_{i1}), \ldots, f_{iD}(-y_{iD})\}\big] \\
&\qquad\qquad \times \prod_{k=1}^{K_i} \big[-V_{\pi_{ik}}\{f_{i1}(-y_{i1}), \ldots, f_{iD}(-y_{iD})\}\big] \prod_{j=1}^{D} f'_{ij}(-y_{ij}) \Bigg),
\end{aligned}
$$

where
$$
f_{ij}(y) = \big\{1 + \xi_{ij}\sigma_{ij}^{-1}(y - \mu_{ij})\big\}_+^{1/\xi_{ij}}
$$
denotes the marginal transformation to the standard Fréchet scale for $\xi_{ij} \neq 0$ and the prime denotes the derivative with respect to y; the negative signs for the y_{ij}'s in (9) correspond to modeling negated winter minima.

The second level of the hierarchical model specifies the models for the exponent measure V and the GEV parameters. We define a space-time model for μ with the covariates discussed in Section 2.1, with a linear temporal trend, and with a Gaussian prior distribution as
$$
\mu(\boldsymbol{s}, t) \sim \mathrm{GP}\{X(\boldsymbol{s})\boldsymbol{\beta}, \tau^2 \rho\} + \alpha t,
$$
where $X(\boldsymbol{s})$ denotes the vector of the intercept and covariates at location $\boldsymbol{s}$, and $\mathrm{GP}\{x(\boldsymbol{s}), \tau^2 \rho(\cdot)\}$ denotes a Gaussian process with mean $x(\boldsymbol{s})$, variance τ^2 and correlation function $\rho(\boldsymbol{h}) = \exp(-\|\boldsymbol{h}\|/\delta)$. The random effect allows the location parameters to vary smoothly over space, in addition to any effect of the covariates. We kept σ and ξ constant over space and time because this appeared to suffice for a good fit. Thus, $\mu_{ij} = \mu(\boldsymbol{s}_j, t_{ij})$, $\sigma_{ij} = \sigma$, and $\xi_{ij} = \xi$ for each pair of replicate and location indices (i, j), and the marginal GEV distributions of the model at the observed data can be written as
$$
-y_{ij} \sim \mathrm{GEV}(U_j + \alpha t_{ij}, \sigma, \xi),
$$
where $\boldsymbol{U} = (U_1, \ldots, U_D)$ is a Gaussian vector with mean $X\boldsymbol{\beta}$, where X is a $D \times 7$ matrix with rows $X(\boldsymbol{s}_j)$, and covariance matrix $\{\tau^2 \rho(\boldsymbol{s}_{j_1} - \boldsymbol{s}_{j_2})\}_{1 \le j_1, j_2 \le D}$. We use a Brown–Resnick process with variogram $2\gamma(\boldsymbol{h}) = 2(\|\boldsymbol{h}\|/\lambda)^\kappa$ as a model for the exponent function V.

4.2.2. *Prior distributions.* Prior distributions are assigned to the variables at the last level of a hierarchical model. We used vague priors for all parameters except those of the latent Gaussian field; see Thibaud et al. (2016). The prior distributions for $\boldsymbol{\beta}$ and τ^2 were chosen to obtain conjugate full conditional distributions for these parameters.

4.2.3. *Space-time declustering.* The use of the Stephenson–Tawn model involves choosing the partitions in (9), whose sets should correspond to independent individual events. Several approaches have been proposed to identify clusters in univariate time series [e.g., Beirlant et al. (2004), Chapter 10, Chavez-Demoulin and Davison (2012)]. In the multivariate case, declustering techniques have been proposed by Coles and Tawn (1991) and Nadarajah (2001). Here we follow Coles and Tawn (1991), who identified clusters of storm surges by assuming that events separated by a fixed time lag are independent.

In our application, a difficulty arises from the use of the dates of occurrences. Because of the measurement precision of the temperatures (0.1°C), the same annual minimum may occur on different days, affecting the estimation of the partition; this is the case for 7% of the minima, with dates of occurrences separated by more than two days in 3% of cases. When a minimum has multiple days of occurrences, we choose one of these days at random to define our partition.

For our temperature data, preliminary analysis suggests choosing a time lag of five days: two observations at the same location and separated by more than five days are assumed to be independent. To identify clusters for the 20 time series, we use a single linkage (or friend-to-friend) method whereby a cluster is composed of data that are dependent in pairs. Thus, if the minima occurred at time 1 for the first series, time 4 for the second and time 7 for the third, these three minima are considered to lie in the same cluster. With this procedure, the size of the partition for each year is at most five, and winter temperature minima usually occurred from widespread events that lasted for between five and ten days.

4.3. *MCMC algorithm.* The MCMC algorithm for our Brown–Resnick model is nonstandard because the likelihood (9) cannot be calculated exactly but must be estimated. The likelihood involves the function V and its partial derivatives, which themselves involve the Gaussian cumulative distribution function in their expressions, and thus can only be estimated. We used the approach of Andrieu and Roberts (2009) to use a likelihood estimator in the MCMC algorithm. Under the assumption that the likelihood estimator is unbiased, that algorithm generates a Markov chain that has the targeted posterior distribution as stationary distribution.

We estimated the likelihood (9) by replacing the Gaussian cumulative distribution functions in its expression by Monte Carlo estimators. Our estimator is not unbiased, but we constructed the estimator to minimize the potential bias in the estimation of the posterior and obtain good mixing properties for the Markov chain

in a reasonable time; see Thibaud et al. (2016). For our dataset and the fixed partitions estimated in Section 4.2.3, a single estimation of the likelihood (9) takes about one second on a single CPU.

We used a Gibbs sampler with Metropolis–Hastings steps to update the marginal and dependence parameters of the model. To minimize the number of likelihood estimations in the MCMC algorithm, and because some parameters were found to be very dependent, we updated the parameters of the model in blocks; see Thibaud et al. (2016). With our choice of blocks, one iteration of the MCMC algorithm requires four likelihood estimations for the Stephenson–Tawn model. For the random partition model, the Gibbs sampler for the partitions requires estimating the probabilities (8), which increases the computational burden a bit more.

Since one iteration of the MCMC algorithm takes several seconds on a single CPU, we used parallel computing, fitting 50 independent Markov chains each of length 15,000 and discarding the first 5000 iterations of each chain. The resulting chains were merged to obtain a final chain of length 500,000. The MCMC diagnostics are discussed in Thibaud et al. (2016). The mean of the posterior distributions were used as pointwise estimators. Credible intervals were calculated from the quantiles of the Markov chains.

The results of a simulation study suggest that the MCMC algorithm provides sensible inferences on the model parameters: the posterior means were close to the true values and the empirical coverages of the credible intervals were close to the nominal values; see Thibaud et al. (2016).

4.4. *Results*. We compare three nonstationary Brown–Resnick models. The first, M1, uses the same trend surfaces as does the best marginal model discussed in Section 4.1 and is fitted using pairwise likelihood as implemented in the R package SpatialExtremes; this is our baseline model, to which we compare our new approaches. The second, M2, is the Bayesian hierarchical model that uses the fixed partitions from the declustering procedure. The third, M3, is the Bayesian hierarchical model with the random partitions.

We discuss the estimates in terms of the negated location parameter μ and negated trend α, which correspond to the original scale of minimum temperatures. The estimates for the three models are given in Table 1. We assessed the fit of the marginal distributions using QQ-plots [see Thibaud et al. (2016)]. We assessed the fit of the dependence model in terms of empirical pairwise extremal coefficients, computed using the empirical marginal distributions of the data to avoid potential bias due to the estimation of the marginals [Ribatet (2013)]; see Figure 3. As a global assessment of the fit in terms of the marginals and dependence, we computed the maximum of winter minimum temperatures over four different sets of stations, for the 53 years of the dataset, and compared the observed values with the predictions from the three Brown–Resnick models, computed by simulation using the algorithm of Dieker and Mikosch (2015); see Figure 4. Further diagnostics for

TABLE 1
Estimated posterior mean parameters and 95% *confidence/credible intervals* (*in parentheses*) *from the trend surface model M*1, *the Bayesian model based on declustering M*2 *and the random partition Bayesian model M*3. *The second column is the estimated average location parameter*
$\mu_j = \mu(s_j, 0)$ *over the* 20 *stations*

Model	$-\overline{\hat{\mu}_j}$	$-\hat{\alpha}$	$\hat{\sigma}$	$\hat{\xi}$	$\hat{\lambda}$	$\hat{\kappa}$
M1	−34.9	0.07	3.4	−0.15	371	0.40
	(−35.8, −34.1)	(0.03, 0.11)	(3.0, 3.7)	(−0.19, −0.10)	(40, 702)	(0.34, 0.47)
M2	−34.6	0.09	4.0	−0.11	1811	0.54
	(−35.6, −33.7)	(0.05, 0.13)	(3.6, 4.6)	(−0.15, −0.06)	(977, 3615)	(0.46, 0.63)
M3	−35.0	0.06	3.5	−0.10	1086	0.53
	(−36.0, −33.8)	(0.00, 0.11)	(3.0, 4.1)	(−0.14, −0.06)	(466, 3143)	(0.43, 0.65)

the prediction of the minimum and the averages of winter minimum temperatures are given in Thibaud et al. (2016).

For the trend surface model, M1, the marginal QQ-plots show poor fits for six stations. Moreover, the model severely underestimates the dependence in terms of the pairwise extremal coefficient. This can be explained by the inadequacy of the fitted marginal distributions, which bias the dependence estimates; Ribatet (2013) also observed this when modeling wind gust data. In terms of predicting the maxima over groups of stations, the model performs poorly: many points lie outside the 95% confidence bounds. In general, this model does not predict extreme temperatures well.
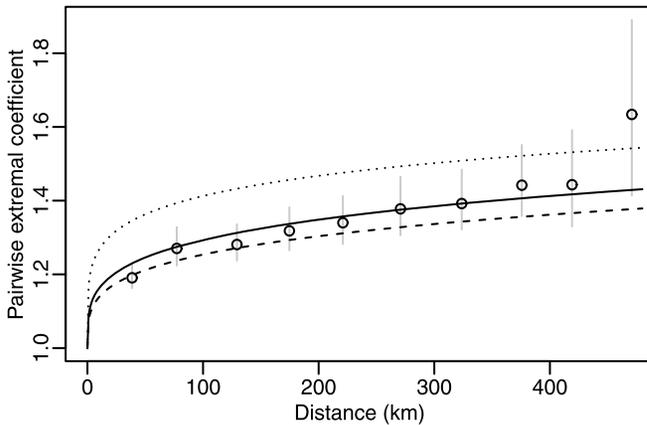

FIG. 3. *Binned estimates of empirical pairwise extremal coefficients* (*black, with their* 95% *confidence intervals in gray*). *Curves are from the fitted Brown–Resnick models* (*see Table* 1): *M*1 (*dotted*), *M*2 (*dashed*) *and M*3 (*plain*).
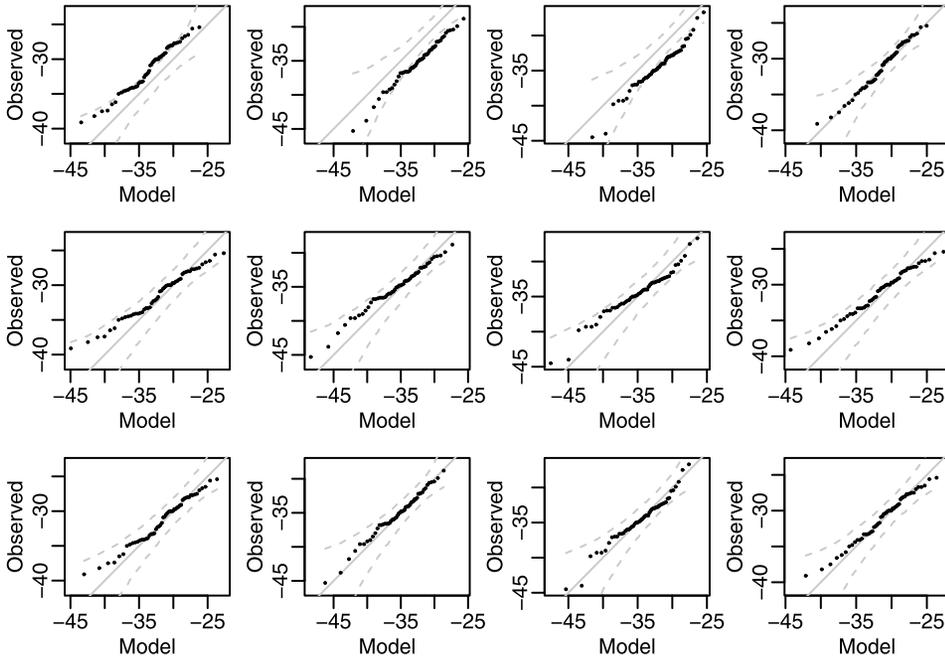
FIG. 4.   *Model checking for the three Brown–Resnick models (see Table* 1): *M*1 (*top row*), *M*2 (*middle row*) *and M*3 (*bottom row*). *The QQ-plots compare the predictions and the observed maximum of winter minimum temperatures* (°*C*) *over some groups of stations,* $m_i = \max_{j \in J} y_{ij}$ ($i = 1, \ldots, 53$). *The columns, from left to right, correspond to the group of stations* $J = \{5, 16, 17, 18, 19\}$ (*western region*), $J = \{2, 3, 4, 6, 7, 8, 9, 12, 20\}$ (*eastern region*), $J = \{5, 6, 8, 9, 10, 11, 12, 13, 14, 15, 19, 20\}$ (*northern region*) *and all* 20 *stations. Grey dashed lines are* 95% *overall confidence bands, and the solid grey line indicates a perfect fit.*

For the Bayesian model based on the declustering, M2, the marginal QQ-plots show that the sample quantiles are outside the 95% confidence bands for three stations, and underestimate them systematically at all stations; the estimated scale parameter σ seems too large. The model overestimates the pairwise dependence somewhat, and tends to overestimate return levels for the maxima over different groups of stations.

For the random partition model, M3, the marginal QQ-plots show a good fit of the GEV distributions: the empirical quantiles are in the 95% confidence bands for all the stations. The empirical pairwise extremal coefficients agree adequately with the curve from the model. In terms of the multivariate predictions, M3 can predict the maximum, minimum and average of the minima better than does M2, with all the observations within the 95% confidence bounds; see Figure 4 and Thibaud et al. (2016). This model appears to represent both the marginal and spatial behavior of extreme low temperatures adequately.

We further discuss the results for the random partition Bayesian model M3. Figure 5(a) shows the posterior mean of the random effects for the location parameter
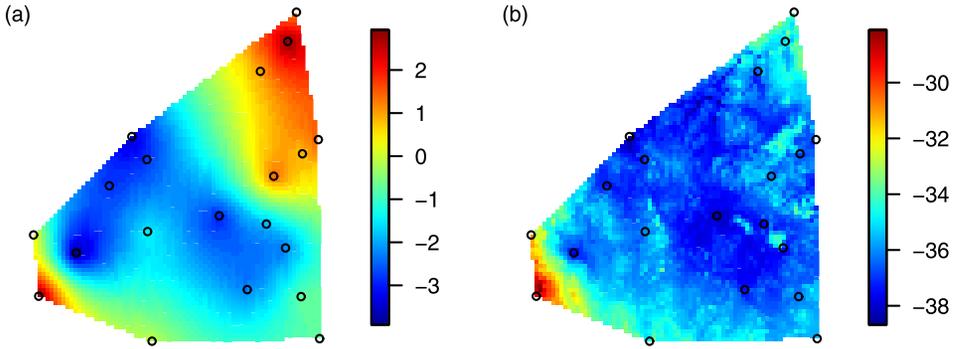
(a)

(b)

FIG. 5.   *Posterior mean of the random effect in the location parameter* (a) *and the mean of the GEV distribution for the year* 2016 (b). *The scale corresponds to temperature minima* (°C).

over the convex hull of the meteorological stations (see Figure 1), and the mean of the GEV distribution for the year 2016. We restrict our analysis to the convex hull of the stations because extrapolation outside this region is too uncertain. The posterior mean of the random effect in the location parameters shows the variability in this parameter that cannot be explained by the covariates. The posterior mean for the fitted GEV distributions shows lower temperatures in the inland region. The minima decrease as latitude, longitude, relative elevation and distance from the Arctic ocean increase, and increase as lake cover and elevation increase. Similar covariate effects were found by Aalto, le Roux and Luoto (2014) for extreme low temperatures in northern Fennoscandia. The estimated shape parameter, corresponding to Weibull—so-called light-tailed—distributions, agrees with previous studies on extreme temperatures. The dependence parameters correspond to strong dependence, with pairwise extremal coefficient $\hat{\theta} = 1.42$ (1.34, 1.50) at 400 km, while $\hat{\kappa} = 0.53$ (0.43, 0.65) implies that the realizations from the Brown–Resnick process are continuous but nowhere differentiable.

4.5. *Forecasts for extreme low temperatures for* 2016–2030.   We base our forecasts on the random partition Bayesian model M3. Changes in extremes are entirely based on the linear trend α in the location parameter. This translates to changes in the mean of the GEV distribution at location s and year t, which is

$$\mu(s, 0) + \alpha t + \sigma \xi^{-1} \{\Gamma(1 - \xi) - 1\}, \qquad \xi < 1, \xi \neq 0.$$

The estimate $-\hat{\alpha} = 0.06$ (0.00, 0.11) corresponds to an increase of 0.6°C per decade in the mean winter minimum temperature. Figure 5(b) shows the predicted mean winter minimum temperature for 2016. Figure 6 shows the probability that winter minimum temperatures are higher than -36°C, the threshold needed for protecting northern Fennoscandia forests from outbreaks of *Epirrita autumnata*, for 1980, 2016 and 2030. There is a strong change in these probabilities. Over the
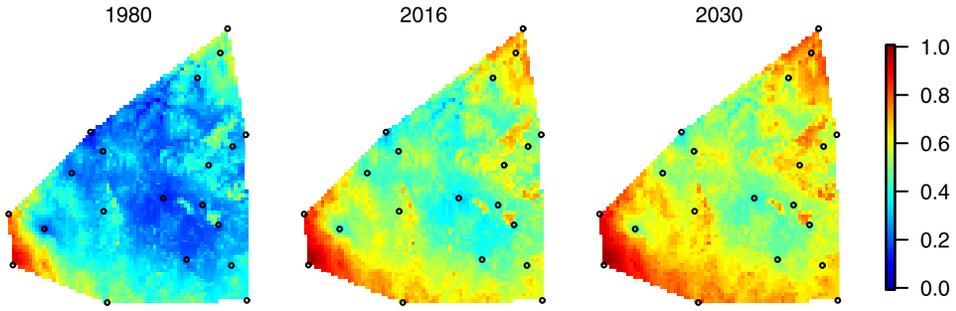
FIG. 6. *Probability that the winter minimum temperature is higher than* $-36°C$ *for* 1980, 2016 *and* 2030.

region defined by the stations, the winter minimum temperatures had low probability of exceeding $-36°C$ before 1980, rising to about 0.54 for 2016, and then to 0.61 for 2030, according to the fitted model. The predictions must be interpreted with care, owing to the large uncertainty in the estimation of α.

Spatial simulation of winter minimum temperatures is useful for estimating joint return levels over certain regions, which could be useful for assessing the risks of pest outbreaks over a region, for example. Figure 7 shows three simulations of winter minima fields from M3, which show how the strong dependence in the Brown–Resnick model results in winter minima that are either low or high throughout the region. The spatial heterogeneity is mostly due to nonstationarity in the marginal distributions. Conditional simulations are possible [Dombry, Éyi-Minko and Ribatet (2013)] and could be useful as input to models to link past pest outbreaks with extreme temperatures, for example.

5. Discussion. Although the Stephenson–Tawn approach was found to perform well on simulated data when the true partitions are known, the results of our application show that it can be inappropriate in practice. The hierarchical model for
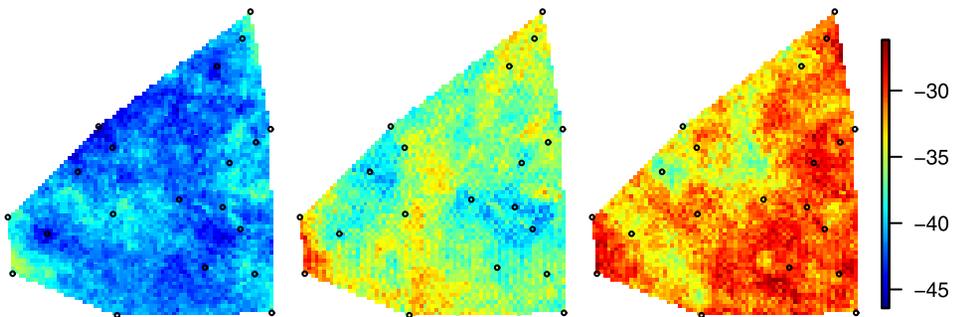

FIG. 7. *Three simulated minimum temperature fields* ($°C$) *for* 2016.

the marginals does not give a good fit to the marginal GEV distributions, notwith-
standing the additional flexibility provided by the random effects.

We explain this by the use of the joint model for the minima and the partitions,
and the declustering. The first of these requires good agreement of the data with the
distribution of the block maxima and the partitions, and this is a strong assumption
in practice. Although the Brown–Resnick dependence structure is quite flexible
for modeling stationary isotropic dependence structures, the distribution of the un-
derlying partitions can be very different from the partitions obtained through the
declustering. In the univariate case, previous studies have shown that declustering
can bias estimators of the marginal parameters if the clusters are poorly defined,
and we can expect similar problems in the multivariate case. In our application,
assuming that data separated by more than five days are independent in pairs re-
sults in partitions of small sizes each year; see Figure 8. A similar size for the
true partitions for a Brown–Resnick process can only be obtained if the depen-
dence in the process is very strong, and this is incompatible with the empirical
extremal coefficients shown in Figure 3. Figure 8 also shows the sizes of the ran-
dom partitions in the MCMC algorithm for model M3; in several cases the size of
the partitions from the declustering procedure have low probability for the random
partition model. To further compare the partitions from the MCMC algorithm with
that from the declustering procedure, we computed the Rand index [Rand (1971)],
which is, for each year, the proportion of pairs of locations that are simultaneously
either together or separate in both partitions; we then averaged the Rand index over
the years. It varies from 0.52 to 0.63, indicating that pairwise classification agrees
for less than 63% of the pairs between M2 and M3. The size of the partitions is
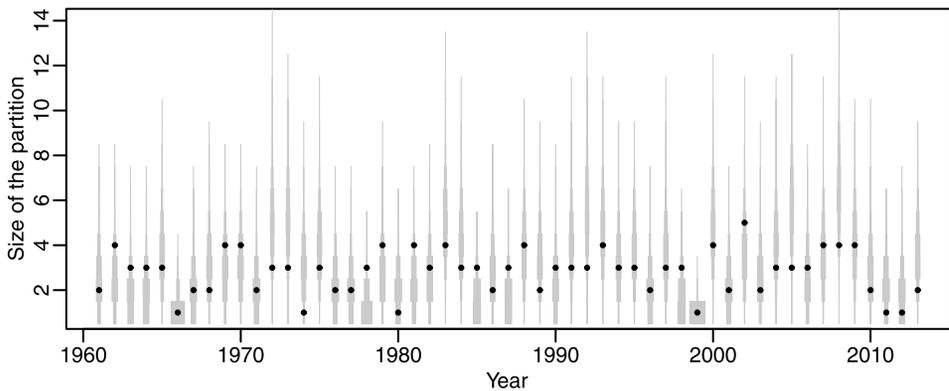typically larger in the random partitions, and the relation with the occurrence dates

FIG. 8.   *Estimated distributions of the partition sizes for each year. Grey boxes show the relative
probability for each size* (*a wider box corresponds to a higher probability*), *estimated from the Gibbs
sampler. The black dots show the sizes of the partitions obtained using the declustering procedure
discussed in Section* 4.2.3.

is unclear. For example, in the winter 1973–1974, the winter minima at the 16 stations with no missing data occurred on four consecutive days and the declustering procedure finds a unique partitioning set, whereas just 7% of the partitions visited in the MCMC run have a unique set and 37% have four partitioning sets or more.

These results suggest that a different declustering procedure must be used. We varied the number of days used in our declustering scheme, and also considered another scheme that forces the stations that are at least 150 km apart to be in different clusters. The parameter estimates changed with the scheme, but the diagnostic plots were never very satisfactory.

6. Conclusion.   We have shown how Bayesian inference can be conducted for the Brown–Resnick process, either using the occurrence times of maxima and a declustering procedure or by including the partitions in an MCMC algorithm.

The partitions add powerful information into the model about the dependence structure. However, our application shows that using fixed partitions can badly bias the estimation of the marginal and dependence parameters of a max-stable process. In a second approach we included the partitions in the MCMC algorithm without using information on the occurrence times, and this produced superior results. A more complex approach could use the occurrence times to construct a prior distribution on the space of the partitions to add more information into the model and improve the efficiency of the estimation, while accounting for the uncertainty on the partitions, but this might be rather complex. Wadsworth (2015) proposed using second-order terms in the Stephenson–Tawn model to reduce potential bias due to the partitions, but this seems difficult in our case because the number of terms added to the likelihood would be huge and the MCMC algorithm would become even slower.

Our Bayesian model can be generalized in several ways. We focused on a stationary isotropic Brown–Resnick process for simplicity, and it appears to be sufficient for our application. A similar Bayesian hierarchical model can be constructed for the extremal-t process using ideas of Thibaud and Opitz (2015), anisotropy would be easily incorporated using the idea of Engelke et al. (2015), and so perhaps could the work of Huser and Genton (2016) on nonstationary dependence models for extremes. However, the generalization of the present method to more complex models and larger datasets may be limited by computational feasibility. Monte Carlo estimation of the likelihood is critical to keep the running times reasonable, and is relatively fast and accurate in low dimensions, but with more sites, more simulations may be needed to obtain reliable inference.

The model predicts an increase of 0.6°C per decade in the mean winter minimum temperatures in northern Fennoscandia. Further work to validate the predictions from our model and to quantify the possible effects of global change on boreal forests would be valuable. The linear trend in time provides sensible short-term predictions, but for longer periods more complex approaches would be necessary, though limited by the availability of reliable data. One possibility would be to

link the changes in extreme temperatures to mean temperatures and use predictions from global climate models under different scenarios as inputs to forecasting.

Acknowledgments. The temperature data are from the Finnish Meteorological Institute database for Finland, and from the European Climate Assessment & Dataset database (http://www.ecad.eu) for Sweden and Norway. This work was motivated by discussions with Seppo Neuvonen, who pointed out to us the relevance of winter minimum temperatures to insect outbreaks.

SUPPLEMENTARY MATERIAL

Supplementary Material for "Bayesian inference for the Brown–Resnick process, with an application to extreme low temperatures" (DOI: 10.1214/16-AOAS980SUPP; .pdf). The online supplement provides additional results for the exploratory analysis and the MCMC algorithm used to fit the two Brown–Resnick models, and reports additional diagnostics for the fit of the three models considered in the manuscript.

REFERENCES

AALTO, J., LE ROUX, P. C. and LUOTO, M. (2014). The meso-scale drivers of temperature extremes in high-latitude Fennoscandia. *Climate Dynamics* **42** 237–252.

ANDRIEU, C. and ROBERTS, G. O. (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *Ann. Statist.* **37** 697–725. MR2502648

ASADI, P., DAVISON, A. C. and ENGELKE, S. (2015). Extremes on river networks. *Ann. Appl. Stat.* **9** 2023–2050. MR3456363

BEIRLANT, J., GOEGEBEUR, Y., TEUGELS, J. and SEGERS, J. (2004). *Statistics of Extremes: Theory and Applications*. Wiley, Chichester. With contributions from Daniel De Waal and Chris Ferro. MR2108013

BROWN, B. M. and RESNICK, S. I. (1977). Extreme values of independent stochastic processes. *J. Appl. Probab.* **14** 732–739. MR0517438

BUHL, S. and KLÜPPELBERG, C. (2016). Anisotropic Brown–Resnick space-time processes: Estimation and model assessment. *Extremes* **19** 627–660. MR3558348

CASSON, E. and COLES, S. (1999). Spatial regression models for extremes. *Extremes* **1** 449–468.

CASTRUCCIO, S., HUSER, R. and GENTON, M. G. (2015). High-order composite likelihood inference for max-stable distributions and processes. *J. Comput. Graph. Statist.* To appear. DOI:10.1080/10618600.2015.1086656.

CHAVEZ-DEMOULIN, V. and DAVISON, A. C. (2012). Modelling time series extremes. *REVSTAT* **10** 109–133. MR2912373

COLES, S. G. and TAWN, J. A. (1991). Modelling extreme multivariate events. *J. Roy. Statist. Soc. Ser. B* **53** 377–392. MR1108334

COOLEY, D., NYCHKA, D. and NAVEAU, P. (2007). Bayesian spatial modeling of extreme precipitation return levels. *J. Amer. Statist. Assoc.* **102** 824–840. MR2411647

DAVISON, A. C. and GHOLAMREZAEE, M. M. (2012). Geostatistics of extremes. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **468** 581–608. MR2874052

DAVISON, A. C., HUSER, R. and THIBAUD, E. (2013). Geostatistics of dependent and asymptotically independent extremes. *Math. Geosci.* **45** 511–529. MR3079649

DAVISON, A. C., PADOAN, S. A. and RIBATET, M. (2012). Statistical modeling of spatial extremes. *Statist. Sci.* **27** 161–186. MR2963980

DE HAAN, L. and FERREIRA, A. (2006). *Extreme Value Theory*: *An Introduction*. Springer, New York. MR2234156

DIEKER, A. B. and MIKOSCH, T. (2015). Exact simulation of Brown–Resnick random fields at a finite number of locations. *Extremes* **18** 301–314. MR3351818

DOMBRY, C., ÉYI-MINKO, F. and RIBATET, M. (2013). Conditional simulation of max-stable processes. *Biometrika* **100** 111–124. MR3034327

ENGELKE, S., MALINOWSKI, A., KABLUCHKO, Z. and SCHLATHER, M. (2015). Estimation of Hüsler–Reiss distributions and Brown–Resnick processes. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **77** 239–265. MR3299407

FUENTES, M., HENRY, J. and REICH, B. (2013). Nonparametric spatial models for extremes: Application to extreme temperature data. *Extremes* **16** 75–101. MR3020178

HUSER, R. and DAVISON, A. C. (2013). Composite likelihood estimation for the Brown–Resnick process. *Biometrika* **100** 511–518. MR3068451

HUSER, R. and GENTON, M. G. (2016). Non-stationary dependence structures for spatial extremes. *J. Agric. Biol. Environ. Stat.* **21** 470–491. MR3542082

IPCC (2013). Summary for policymakers. In *Climate Change* 2013: *The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (T. F. STOCKER, D. QIN, G. K. PLATTNER, M. TIGNOR, S. K. ALLEN, J. BOSCHUNG, A. NAUELS, Y. XIA, V. BEX and P. M. MIDGLEY, eds.) 3–29. Cambridge Univ. Press, New York.

KABLUCHKO, Z., SCHLATHER, M. and DE HAAN, L. (2009). Stationary max-stable fields associated to negative definite functions. *Ann. Probab.* **37** 2042–2065. MR2561440

NADARAJAH, S. (2001). Multivariate declustering techniques. *Environmetrics* **12** 357–365.

NIKOLOULOPOULOS, A. K., JOE, H. and LI, H. (2009). Extreme value properties of multivariate t copulas. *Extremes* **12** 129–148. MR2515644

OPITZ, T. (2013). Extremal t processes: Elliptical domain of attraction and a spectral representation. *J. Multivariate Anal.* **122** 409–413. MR3189331

PADOAN, S. A., RIBATET, M. and SISSON, S. A. (2010). Likelihood-based inference for max-stable processes. *J. Amer. Statist. Assoc.* **105** 263–277. MR2757202

RAND, W. M. (1971). Objective criteria for the evaluation of clustering methods. *J. Amer. Statist. Assoc.* **66** 846–850.

RIBATET, M. (2013). Spatial extremes: Max-stable processes at work. *J. SFdS* **154** 156–177. MR3120441

RIBATET, M. (2015). SpatialExtremes: Modelling spatial extremes. R package version 2.0-2.

RIBATET, M., COOLEY, D. and DAVISON, A. C. (2012). Bayesian inference from composite likelihoods, with an application to spatial extremes. *Statist. Sinica* **22** 813–845. MR2954363

SANG, H. and GELFAND, A. E. (2009). Hierarchical modeling for extreme values observed over space and time. *Environ. Ecol. Stat.* **16** 407–426. MR2749848

SANG, H. and GELFAND, A. E. (2010). Continuous spatial process models for spatial extreme values. *J. Agric. Biol. Environ. Stat.* **15** 49–65. MR2755384

SCHLATHER, M. (2002). Models for stationary max-stable random fields. *Extremes* **5** 33–44. MR1947786

SHABY, B. A. (2014). The open-faced sandwich adjustment for MCMC using estimating functions. *J. Comput. Graph. Statist.* **23** 853–876. MR3224659

SHABY, B. A. and REICH, B. J. (2012). Bayesian spatial extreme value analysis to assess the changing risk of concurrent high temperatures across large portions of European cropland. *Environmetrics* **23** 638–648. MR3019056

SMITH, R. L. (1990). Max-stable processes and spatial extremes. Unpublished manuscript, Univ. Surrey. Available at http://www.stat.unc.edu/postscript/rs/spatex.pdf.

STEPHENSON, A. G. (2009). High-dimensional parametric modelling of multivariate extreme events. *Aust. N. Z. J. Stat.* **51** 77–88. MR2504104

STEPHENSON, A. and TAWN, J. (2005). Exploiting occurrence times in likelihood inference for componentwise maxima. *Biometrika* **92** 213–227. MR2158621

THIBAUD, E. and OPITZ, T. (2015). Efficient inference and simulation for elliptical Pareto processes. *Biometrika* **102** 855–870. MR3431558

THIBAUD, E., AALTO, J., COOLEY, D. S., DAVISON, A. C. and HEIKKINEN, J. (2016). Supplement to "Bayesian inference for the Brown–Resnick process, with an application to extreme low temperatures." DOI:10.1214/16-AOAS980SUPP.

VIRTANEN, T., NEUVONEN, S. and NIKULA, A. (1998). Modelling topoclimatic patterns of egg mortality of *Epirrita autumnata* (Lepidoptera: Geometridae) with a Geographical Information System: Predictions for current climate and warmer climate scenarios. *Journal of Applied Ecology* **35** 311–322.

WADSWORTH, J. L. (2015). On the occurrence times of componentwise maxima and bias in likelihood inference for multivariate max-stable distributions. *Biometrika* **102** 705–711. MR3394285

WADSWORTH, J. L. and TAWN, J. A. (2014). Efficient inference for spatial extreme value processes associated to log-Gaussian random functions. *Biometrika* **101** 1–15. MR3180654

E. THIBAUD
D. S. COOLEY
DEPARTMENT OF STATISTICS
COLORADO STATE UNIVERSITY
FORT COLLINS, COLORADO 80523-1877
USA
E-MAIL: emeric.thibaud@colostate.edu
        cooleyd@stat.colostate.edu

J. AALTO
FINNISH METEOROLOGICAL INSTITUTE
P.O BOX 503
00101 HELSINKI
FINLAND
E-MAIL: juha.aalto@fmi.fi

A. C. DAVISON
ECOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
EPFL-FSB-MATHAA-STAT, STATION 8
1015 LAUSANNE
SWITZERLAND
E-MAIL: anthony.davison@epfl.ch

J. HEIKKINEN
NATURAL RESOURCES INSTITUTE FINLAND
P.O. BOX 18
01370 VANTAA
FINLAND
E-MAIL: juha.heikkinen@luke.fi