Asymptotic models and inference for extremes of spatio-temporal models

K.F. Turkman
CEAUL and DEIO, FCUL, University of Lisbon

1 Introduction

Suppose we have observations $x(s_i,t)$, $s_i \in A \subset \mathcal{R}^2$, $t = 0,1,2,...T$, $i = 1,...,n_t$ from a non-stationary space-time process $X(s,t)$, continuous in space and discrete in time. Based on such data set, how can we make inference on the extremal properties of $X(s,t)$? Specifically, we may want to estimate probabilities such as

- $$P(\max_{s\in A} X(s,T+1) > u),$$
- $$P(\max_{1\le t\le T}\max_{s\in A} X(s,t) > u),$$
- $$P(X(s_1,T+1) > u_1,...,X(s_n,T+1) > u_n),$$

for some high thresholds u_1,...,u_n and for any locations s_1,...,s_n.

Let us start with the simpler case of extremes of a spatial process $X(s)$, $s \in A \subset \mathcal{R}^2$. As in the univariate and multivariate extreme value theory, it makes sense to base our inference on asymptotic models. However, On the account that in most applications, A is limited, it may not make sense to look for asymptotics by expanding the space indefinitely, looking for similar asymptotics as in the case of temporal extremes. Hence there are two alternatives to look for aymptotics in space:

1. Start with iid replicates $X_i(s)$ and look at the convergence of the continuous process $\max_{1\le i\le n} X_i(s)$, as $n \to \infty$, for all $s \in A$ upon suitable normalization

$$Y_n(s) = a_n(s)^{-1}(\max_{1\le i\le n} X_i(s) - b_n(s)) \to^{\mathcal{D}} Y(s). \quad (1)$$

We will call this asymptotics, the de Haan approach.

2. Start with a sufficiently fine grid(lattice) over A. characterize the extremal properties of the process $X(s)$ over this fine grid, and obtain asymptotics by letting the grid sizes go to 0. This method is often called the Double sum method of Pickand and used extensively by Piterbarg(1996) to obtain limiting results for Gaussian random fields. However, I will call this method Leadbetter-Albin method due to their extensive use of this techniques in obtaining limiting results for the extremes of continuous time processes.

Let us look closely the first method and the corresponding asymptotic models for spatial extremes.

2 de Haan asymptotics

For a fixed s_0 or for a finite collection of locations $s_1, .., s_k$, the conditions for the convergence of (1), as well as the possible asymptotic models are well known and are part of the univariate and multivariate extreme value theory. The convergence of the stochastic process in (1) is more complicated but the results are also complete. See for example, de Haan and Pereira(2006). Convergence is assumed to occur with non-degenerate margins and when (1) converges weakly or in distribution to a stochastic process $Y(s)$, $s \in A$, (or any compact subset of an Euclidean space) then the limiting process $V(s)$ is a max-stable process, having non-degenerate,univariate marginal GEV distributions. In order to obtain mathematically tractable results and representations, the process in (1) is transformed to have standard marginal distributions: If the functions $a_n(s)$, $b_n(s)$ are chosen in such a manner that for every s, the marginal limiting distribution have the form

$$\lim_{n\to\infty} P(Y_n(s) \leq y) = \exp(-(1 + \frac{k(s)}{\sigma(s)}(y - \mu(s)))^{-1/k(s)},$$

then $Y(s)$ is transformed to

$$V(s) = (1 + \frac{k(s)}{\sigma(s)}(Y(s) - \mu(s)))^{1/k(s)}, \tag{2}$$

so that the limiting max stable process has unit Frechet margins given by

$$P(V(s) \leq v) = \exp(-v^{-1}).$$

Here, $k(s)$ is called the index (shape) function, $\sigma(s)$ and $\mu(s)$ can be called scale and location functions. In statistical applications, either all these functions are assumed to be known or if estimated, their sampling variations are ignored in order to get tractable representations.

When the limit exists, the max-stable process $V(s)$ with unit Frechet margins have the following very useful representation:

$$V(s) =^d \vee_{i=1}^{\infty} Y_i W_i(s), \tag{3}$$

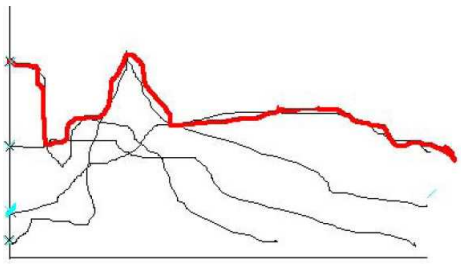

Figure 1: Realization of a max-stable process

where, Y_i are the points of a point process defined on $(0, \infty)$ with mean measure $r^{-2}dr$ and $W_i(s)$ are i.i.d replicates of an arbitrary positive spatial process $W(s)$, $s \in A$, satisfying some conditions to ensure that the marginal distribution of $V(s)$ for each s is a unit Frechet distribution. Hence, $W(s)$ is a process such that $E(W(s)) = ||A||$, for every $s \in A$ and $E(\sup_{s \in A} V(s)) < \infty$. Here, the mean measure $r^{-2}dr$ corresponds to the exponent measure of the unit Frechet extreme value distribution.

Transformation (2) is crucial for the representation (3). To our knowledge, there is no mathematically tractable representation for max stable processes with non-unit Frechet margins (which i call non-stationary version, assuming that shape, scale and location parameters are spatially varying functions) Non-stationary version of this representation can be quite complicated. The point process Y_i in the representation (3) has to replaced by a non-homogenous point process on $R^2 \times (0, \infty)$ with mean measure depending on the spatially varying index, scale and location functions and the necessary conditions on $W(s)$ to render a Frechet marginal distribution to $V(s)$ with parameters $k(s), \sigma(s), \mu(s)$ at every point $s \in A$ would be intractable. Hence, in order to use the representation (3) as an asymptotic model for extremes of non-stationary data, Buishand et all (2007) suggest using the following general steps:

1. Fit $k(s)$, $\sigma(s)$ and $\mu(s)$ locally using (temporal) block maxima data at each location s_i (assuming that such temporal data at each location exist)

2. transform the data using (2), using these estimated functions.

3. In the representation (3), choose a specific parametric model for $W(s)$. Buishand et al(2007) suggest using exponential martingale, having a single parameter which represents the degree of dependence of large values of the process at two spatial locations . (Other alternative is to use a Gaussian process, see Schlather(2002) for details.)

4. Estimate the parameter β of this latent process (see de Haan and Pereira, 2006)

5. $1/Y_i$ is a Poisson point process on $(0, \infty)$ with mean measure dr, and the points of this process can be constructed by sums of unit exponential

random variables E_1, $E_1 + E_2$,...$E_1 + E_2 + ... + E_n$... ,

6. Approximate $V(s)$

$$V(s) = \max_{i=1}^{n} \frac{1}{E_1 + E_2 + ... + E_n} W(s),$$

for some finite n. Buishand et al (2004) suggest that using n as low as 4 would be sufficient.

7. The transformation

$$\eta(s) = (1 - \exp(-W^{-1}(s)))^{-1}$$

results in a process whose marginal distributions are unit Pareto

8. Finally, the process

$$X(s) = \hat{\sigma}(s) \left(\frac{\eta(s)^{\hat{k}(s)} - 1}{\hat{k}(s)} \right) + \hat{\mu}(s),$$

can be used for simulating extreme values of the spatial process.

The representation (3) and the consequent statistical inference is not flexible enough to model extremes of non stationary spatial processes. First of all, temporal data at each spatial location is not available to estimate these parameters locally. Second, even if temporal data had exists at each location, these parameters often show spatial dependence and without taking into consideration these spatial dependence, these estimators would show high bias. For example, it is not possible to introduce other information into the inference in the form of spatially varying explanatory variables. third, sampling variation that exist in the estimators are not taken into consideration and hence the model given in (3) is specified conditionally on the estimated parameters. Also, joint distributions of spatial extremes beyond bivariate distributions are very difficult to calculate. Ideally, we would like to use more flexible models which over come at least partially these difficulties.

Bayesian Hierachical modeling and simulation Based inference techniques are now accepted as the main principal tools for modeling spatial, non-stationary data. See for example Banarjee et all(2004) for the general introduction to Bayesian Hierarchical modeling. Typically, if we have block maxima data $v(s_i)$, $i = 1, ..., n$, where each $v(s_i)$ represent the maximum of observations at location s_i over a time interval, we may want to model $v(s_i)$, $i = 1, ..., n$ by a Bayesian hierachical model assuming that conditional on the realization of a latent spatial process $W(s)$ and a set of explanatory variables chosen in accordance with the physical process governing the extremes, the data are independent having $GEV(k(s), \sigma(s), \mu(s)$ distribution, where the spatially varying (random) model parameters depend on the latent process $W(s)$ and explanatory variables through properly chosen link functions. Once a proper parametric

model for this latent process is chosen(often a Gaussian process) and prior distributions are assigned to the model parameters and hyper-parameters, then simulation based inference techniques can be applied to make inference on the posterior distributions of the model parameters, including the latent process, as well as the joint predictive distributions of $(V(s_1^*), ..., V(S_p^*)$ at any (unobserved)locations $(s_1^*, ..., s_p^*)$. Such models are very flexible and recently their use has been increasing. See for example, Cooley(2006, 2007). Methodological study of strategies for Bayesian hierarchical models for block maxima data, can be found in Sang(2008). The principal point of start for using these modeling strategies is that we want to use as model a spatial process whose univariate marginal distributions are generalized extreme value distributions with space dependent parameters, but have an arbitrary dependence structure. This dependence structure is then introduced by a latent spatial process through the model parameters and the observations, conditional on the realizations of this latent process as well as other explanatory variables are independent, but not identical. Such modeling strategy also bring the benefit of introducing meaningful,important characteristics of the process under study into the model through the model parameters, which may help justify the assumption of conditional independence.

There is however, the question if these models are compatible with the asymptotic theory, namely if such conditionally independent processes are max-stable. At first look, such modeling strategies are hardly compatible with the max-stable representation given in (3). However, we now give an justification to show that these models indeed are compatible with the max stable processes and hence with the asymptotic theory. Consider the spatial process $V(s)$ defined over $s \in A \subset R^2$, given by

$$V(s) = Y(s)W(s), \tag{4}$$

where for each s, $Y(s)$ have marginal $GEV(k(s), sigma(s), \mu(s))$ distribution and for any $s_1, s_2, ..., s_p$, $Y(s_1), ..., Y(s_i)$ are independent and $W(s)$ is the positive spatial process given inthe representation (3) Note that, conditional on the realizations of the latent process $W(s)$,

$$\begin{aligned} & P(V(s_1) \leq v_1, ..., V(s_n) \leq v_n | W(s_1) = w_1, ..., W(s_n) = w_n) \\ = & \prod_{i=1}^{n} P(V(s_i) \leq v_i | W(s_n) = w_i) \\ = & \prod_{i=1}^{n} GEV(k(s_i), \sigma(s_i)w_i, \mu(s_i)w_i). \end{aligned} \tag{5}$$

Thus, the $V(s)$ is a conditionally independent and have spatially varying random scale and location parameters, having the same (non-random) shape (index) function. Note also that the dependence structure to $V(s)$ is introduced by the latent process $W(s)$ through the scale and location parameters. Hence, conceptually, $V(s)$ represents the conditionally independent, marginally

GEV spatial process which is suitable for hierarchical modelling and we would like to use as model for our block maxima data However, the process $V(s)$ is not max stable. For simplicity, take $Y(s) \sim GEV(1,1,1)$. If for any k, $V_1(s), V_2(s), ..., V_k(s)$ are independent, identical replicates of $V(s)$ then

$$\begin{aligned}
& P(\frac{1}{k} \vee_{i=1}^{k} V_i(s_0) \leq v) \\
= \; & P(Y_1(s_0)W_1(s_0) \leq kv, ..., Y_k(s_0)W_k(s_0) \leq kv) \\
= \; & \int P(Y_1(s_0)W_1(s_0) \leq kv, ..., Y_k(s_0)W_k(s_0) \leq kv | W_1(s_0) = w_1, ..., W_k(s_0) = w_k) \\
\times \; & dP_{(W_1(s_0),..,W_k(s_0)}(w_1, ..., w_k) \\
= \; & \int P(Y_1(s_0) \leq kv/w_1, ..., Y_k(s_0) \leq kv/w_k) dP_{(W_1(s_0),..,W_k(s_0)}(w_1, ..., w_k) \\
= \; & \prod_{i=1}^{k} \int P(Y(s_0) \leq kv/w_i) dP_{W(s_0)}(w_i) \\
= \; & \prod_{i=1}^{k} \int \exp(-(kv/w_i)^{-1}) dP_{W(s_0)}(w_i) \\
= \; & E_W^k(e^{-(kv/W)^{-1}}) \\
\neq \; & E_W(e^{-(v/W)^{-1}}).
\end{aligned}$$

we can similarly show that for any collection $s_1, ..., s_n$,

$$(\frac{1}{k} \vee_{i=1}^{k} V_i(s_1), ..., \frac{1}{k} \vee_{i=1}^{k} V_i(s_n) \neq (V(s_1), ..., V(s_n)$$

. Hence the process is not a max-stable process.

However, $V(s)|W(s) = w$ is conditional max-stable, and hence $V(s)$ is a mixture of max-stable processes. Also, the process $Y(s)W(s)$ is the max domain of attraction of the simple max stable process (see arguments given in de Haan and Pereira (2006), page 326) hence asymptotic dependence structure of the two processes is the same and either process can be used for inference on the dependence structure. It is easy to verify that when $Y(s)$ are independent but non-identical with marginal $G(k(s), \sigma(s), \mu(s)$ distributions, the process $V(s)$ is conditionally max stable, and hence is a mixture of max-stable processes and is in the domain of attraction of the max stable process

$$V_1(s) = \frac{k(s)}{\sigma(s)}(\nu(s)^{k(s)} - 1) + \mu(s),$$

where $\nu(s)$ is the simple max stable process given in (3).

Therefore, it is our belief that conditionally specified, conditionally independent Bayesian hierarchical GEV models are compatible with the asymptotic theory, although they are not max stable processes.

We also note that the strictly speaking, the model suggested by Buishand et al(2007) is max-stable process conditional on the values of the estimated parameters: Once the parameters are estimated, their model results in a conditionally specified model given by

$$V(s)|\hat{k}(s), \hat{\sigma}(s), \hat{\mu}(s), \hat{\mathbf{\Theta}}$$

where $\hat{k}(s), \hat{\sigma}(s), \hat{\mu}(s)$, are the model parameters, estimated marginally at each location and $\mathbf{\Theta}$ are the estimated parameters of the latent process $W(s)$. The "unconditional" distribution of $V(s)$ is obtained by the integration

$$F_{V(s)}(v|\hat{k}(s), \hat{\sigma}(s), \hat{\mu}(s), \hat{\mathbf{\Theta}}) = \int_w F_{V(s)}(v|\hat{k}(s), \hat{\sigma}(s), \hat{\mu}(s), w) dP_{W(s)}(w|\hat{\mathbf{\Theta}}) dw.$$

Here, the process $W(s)$ satisfy conditions so that for each $s \in A$,

$$F_{V(s)}(v|\hat{k}(s), \hat{\sigma}(s), \hat{\mu}(s), \hat{\mathbf{\Theta}})$$

are respectively unit Frechet distributions. ((strictly speaking, the intensity measure of the poisson process on $A \times (0, \infty)$, or

$$-log(F_{V(s)}(v|\hat{k}(s), \hat{\sigma}(s), \hat{\mu}(s), \hat{\mathbf{\Theta}}))$$

is integrated with respect to $P_{W(s)}(w|\hat{\mathbf{\Theta}})$, and conditions on $W(s)$ guarantee that this integral is equal to v^{-1}, resulting in unit Frechet marginals. See de Haan and Fereira(2007) for details.) The unconditional distributions $F_{V(s)}(v)$ which are obtained by integrating out the sampling variation of these estimators ofcourse need not have GEV margins and hence can not be max-stable.

Fundamental drawback of using conditionally specified GEV model is that rarely there is time-block maxima data. Often, we do not have sufficient time replicate data at each location to construct block maxima data for inference on the GEV model. However, there is a very useful duality between modeling peaks over a high threshold and block maxima, equivalently between the generalized Pareto distribution (GPD) and the Generalized extreme value distribution (GEV) in classical extreme value theory, which also exist for spatial extremes, permitting more efficient use of spatial data.

Let u be a high threshold. For any $x > 0$, we can write

$$P(X > u + x) = P(X > u + x|X > u)P(X > u), \tag{6}$$

and if M_n, the maximum of n iid random variables with distribution function F satisfies

$$P(M_n \leq z) \sim G(z),$$

where

$$G(z) = \exp[-(1 + k(\frac{z - \mu}{\sigma})^{-1/k}],$$

for some $\mu, \sigma > 0$ and $-\infty < k < \infty$, then for sufficiently large enough threshold u,

$$\begin{aligned} H(x) &= P(X \leq u + x | X > u) \\ &\sim 1 - (1 + \frac{k}{\hat{\sigma}} x)^{-1/k}, \end{aligned} \tag{7}$$

for x such that $1 + k/\sigma x > 0$ and $\hat{\sigma} = \sigma + k(u - \mu)$. $H(x)$ is called the Generalized Pareto distribution (GPD) and the parameters of the GEV model $G(z)$ can be uniquely determined by the parameters of the corresponding GPD model $H(x)$. See for example Coles (2001). This duality between the GEV and the GPD models is often used in statistical inference for extremes, since fitting the GPD models is more efficient due to the fact that all observations above a high threshold are used in the inference, as compared to the block maxima needed in fitting the GEV model. However, $P(X > u)$ still needs to be estimated from the data. This can either be done empirically, or as Coles (2001) suggests by adding one more parameter to the model

$$H(x) = 1 - \eta(1 + \frac{k}{\sigma}(x - u)^{-1/k} \quad , \quad x > u$$

to estimate the tail probability

$$P(X > u + x).$$

Here, the parameter η is used for estimating the (fixed) probability $P(X > u)$ in the representation (6). See also Coles(2001), inference for extremes using the Poisson model and its relation to the GPD model.

The duality between the GEV and GPD model extends to max-stable processes as Buishand et all (2007) suggest: Consider the following GPD-process

$$V_1(s) = YW(s), \mathbf{??} \tag{8}$$

where $Y(s)$ be is a process with marginal Pareto distribution function

$$P(Y \leq y) + 1 - \frac{1}{y} \quad , y > 1.$$

Writing this distribution in the form

$$P(Y \leq y) = 1 - (1 - (y - 1))^{-1},$$

we see that this is GPD distribution with $k = 1$, $\sigma = 1$ and $u = 1$. Buishand et all(2007) call this random variable, the unit GPD random variable. Now, let $W(s)$ be the latent spatial process in the representation (3). $V_1(s)$ is in the domain of attraction of the max stable process given in (3), hence the asymptotic dependence structure of the max stable processes $V(s)$ in (3) and $V_1(s)$ are identical and inference on the extremal properties of the process can equivalently be made on the $V_1(s)$ process. Note that this is similar to the

duality that exist between the GEV and the GPD models in the univariate case.

Note that the distribution of $V_1(s)$ conditional on $W_1(s) = w(s)$, is given by

$$P(V_1(s) \leq v | W_1(s) = w(s)) = 1 - (\frac{1}{w(s)} y)^{-1} \quad , \quad y > w(s)$$

is a GPD with $k = 1$, $\sigma = w(s)$ and $u = w(s)$. Hence, $W_1(s)$ introduces independence through the scale and shape parameter of the GPD.

Note also that, for any s_1 and s_2 in A,

$$P(V_1(s_1) \leq v_1, V_2(s_2) \leq v_2 | W_1(s_1) = w_1, W_1(s_2) = w_2) = P(Y \leq \min(v_1/w_1, v_2/w_2))$$

collapsing to the total dependence case. The dependence structure to the bivariate random variable $(V_1(s_1), V_1(s_2)$ is then introduced through integration of this degenerate distribution with respect to the joint distribution of $W(s_1), W(s_2)$ by

$$P(V_1(s_1) \leq v_1, V_2(s_2) \leq v_2) = \int P(Y \leq \min(v_1/w_1, v_2/w_2)) dP_{W_1(s_1), W_1(s_2)}(w_1, w_2).$$

Based on these facts, we suggest the model

$$V_1(s) = Y(s)W(s), \tag{9}$$

where $W(s)$ is a (latent) spatial process and $Y(s)$ is a process with marginal distribution function $GPD(k(s), \sigma(s), 0)$, representing the excesses over a fixed but high threshold and for any $s_1, ..., s_n$, $Y(s_1), ..., Y(s_n)$ are independent. The marginal distribution of $V_1(s)$ conditional on $w(s) = w$, $k(s)$ and $\sigma(s)$ is given by

$$P(Y(s) \leq y/w(s), k(s), \sigma(s)) = 1 - (1 + \frac{k(s)}{\sigma(s)w(s)}(y))^{1/k(s)}, \tag{10}$$

which is $GPD(k(s), \sigma(s)w(s), 0)$. Note that for any s_1,s_2,...,s_n, $V_1(s), ..., V_1(s_n)$ conditional on $W(s_1) = w(s_1), ..., W(s_n) = w(s_n)$ are independent with

$$P(V_1(s_1) \leq v_1, ..., V_1(s_n) \leq v_n | w(s_1), ..., w(s_n)) = \prod_{i=1}^{n} GPD(k(s_i), \sigma(s_i)w(s_i), 0).$$

Hence, $W(s)$ can be seen as a latent spatial process which introduces dependence to the GPD process $V_1(s)$ through its scale parameter, and the GPD process $V_1(s)$ is conditionally independent given the specification of its spatially varying, random location scale parameter. Hence we suggest the following strategies for modeling spatial extremal data:

- Model excesses over a high fixed threshold.
- Assume that the observations are independent having a GPD, conditional on the realizations of a spatially varying random scale function, as well as a random shape parameter which is spatially static.

or

- if there is temporal data on each observational location, model the time block maximal data by the the process (4).
- Assume that the observations are independent having a GEV distribution, conditional on the realizations of a spatially varying random scale and location functions, as well as a random shape parameter which is spatially static.

As explained, such assumptions are clearly compatible with the de Haan asymptotics.

We give the basic hierarchical model for excess data over a fixed high threshold. Full analysis of the GEV model as well as modelling strategies are given in Sang(2008).

Suppose that $X(s)$ is a spatial process and $\mathbf{v} = (v_1(s_i), i = 1, ..., n)$ are observed excesses of $X(s)$ over a sufficiently high but fixed threshold u, so that we assume them to be a realization of the process $V_1(s)$ given in (9). Assuming a Bayesian set up, we assume that the shape parameter is random but spatially static, whereas scale and location parameters are spatially varying random functions. Based on this set up, we suggest the following hierarchical model: Then

1.
$$p(v(s_1), ..., v(s_n)|k, \sigma(s_i), i = 1, ..., n) = \prod_{i=1}^{n} \frac{1}{\sigma(s)}(1 + \frac{k}{\sigma(s)}(v(s_i))^{-1/k-1}, \tag{11}$$

2. Assume that we have a prior information for k, given in the form of the density function $p(k|\boldsymbol{\beta})$ for $-\infty < k < \infty$, where $\boldsymbol{\beta}$ are the hyper parameters of this prior distribution.

3. $\sigma(s_i) > 0$ for each $s \in A$, hence we model $\sigma(s_i)$ by

$$\log \sigma(s_) = \sigma_0 + \sigma_1 W_0(s)$$

$W_0(s_i)$ is a Gaussian process $N(\mu(s_i), \Sigma)$. Often, a parametric representation for $\mu(s_i, t)$ is given as

$$\mu(s_i) = \beta_0 + \mathbf{x}(s_i)\boldsymbol{\beta}',$$

where $\mathbf{x}(s_i)$ are covariates and $\boldsymbol{\beta}$ are random regression coefficients having prior distributions. Σ is a covariance matrix based on a valid, parametric covariance structure. See for example, Banarjee et all(2004) for a list of these valid covariance structures. Often it is preferential to write

$$\log \sigma(s_i) = \beta_0 + \mathbf{x}(s_i)\boldsymbol{\beta}' + W_0(s_i),$$

where $W(s)$ is stationary Gaussian process with covariance Σ.

4. Once prior distributions for the hyperparameters of the model are specified, the above model becomes a fully specified hierarchical model and well established simulation based inference techniques can be used to estimate the model parameters.

5. It is possible to introduce spatial dependence and heterogeneity to the model through the shape parameter, by adding another link function of the form
$$k(s_i) = k_0 + \boldsymbol{\alpha}\mathbf{x}' + W_1(s_i),$$
where $W_1(s)$ is another latent, Gaussian spatial process which may or may not be independent of the latent process $W(s_i)$. One possible dependence structure for $W_0(s)$ and $W_1(s)$ is to start with two independent Gaussian processes $G_1(s)$ and $G_2(s)$ and write
$$W(s) = G_1(s),$$
and
$$W_1(S) = a_1 G_1(s) + a_2 G_2(s),$$
for some parameters a_1 and a_2.

6. it is possible to introduce a spatially varying random location parameter in the model. However, since the location parameter in the GPD is in fact a truncation parameter for the part of the data which is used for inference, implemetation of a Bayesian Hierarchical model with random location parameter is difficult. Similar strategies such as those given by Tancredi et al (2001) and de Zea Bermudez et all(2001) can be adopted for spatial data.

7. Note that we model the excesses over a high threshold, that is, we model
$$P(X(s_1) > x_1 + u, ..., X(s_n) > x_n + u | X(s_1) > u, ..., X(s_n) > u),$$
for any collection of locations $s_1, ..., s_n$ and for some high, fixed threshold u In principle, we may want to make the inference based on at least one component exceeding the fixed threshold. However, implementation of the hierarchical model, due to the complications in the likelihood (see Coles(2001), page 155 for complications in the bivariate distributions) is very difficult. One alternative is to start with conditional approach to multivariate extremes suggested by Heffernan and Tawn (2004) and adopt these methods to spatial extremes.

The hierarchical model above and its numerical implementation, will report on the joint predictive density of excesses at any location s_1,...,s_p,
$$p(v(s_1), .., v(s_p)|\mathbf{v}),$$
in the form of simulated samples from this density. As a by product, joint posterior densities of all model parameters including the latent random effect

$W_0(s)$ can also be reported. Posterior expected values of the model parameters as well as credible intervals can be calculated from the simulated samples from the marginal posterior densities. The estimated posterior joint density of $W(s)$ is particularly useful, as it gives a very good picture of the dependence structure as well as the spatial heterogeneity of the data.

Often, one needs to make inference on the joint probability $P(X(s_1) > u+x_1, .., X(s_p) > u+x_p)$ for arbitrary locations s_1,...,s_p and for a high threshold u. Since

$$\begin{aligned} & P(X(s_1) > u + x_1, .., X(s_p) > u + x_p) \\ = \; & P(V_1(s_1) > x_1, ..., V_1(s_p) > x_p) P(X(s_1) > u, ..., X(s_p) > u), \end{aligned} \quad (12)$$

apart from the joint distribution of the excesses, one will need an estimate of the joint probability

$$P(X(s_1) > u, ..., X(s_p) > u).$$

Empirical methods for estimating this probability may not be straightforward. We suggest adopting and extending Cole(2001) suggestion:

Assume that there is an latent process $W(s)$ such that conditional on $W(s)$ large values of $X(s)$ over the threshold are independent, so that

$$\begin{aligned} & P(X(s_1) > u + x_1, .., X(s_p) > u + x_p | W(s_1) = w(s_1), ..., W(s_n) = w(s_n) \\ = \; & P(V_1(s_1) > x_1, ..., V_1(s_p) > x_p | W(s_1) = w(s_1), ..., W(s_n) = w(s_n)) \\ \times \; & P(X(s_1) > u, ..., X(s_p) > u | W(s_1) = w(s_1), ..., W(s_n) = w(s_n)) \\ = \; & \prod_{i=1}^{n} P(V_1(s_i) > x_i | W(s_i) = w(s_i)) P(X(s_i) > u | W(s_i) = w(s_i)) \end{aligned} \quad (13)$$

Writing $\eta(s_i) = P(X(s_i) > u | W(s_i) = w(s_i)$ as a spatially varying random parameter, we get the following hierarchical model for making inference on the joint posterior distribution of $X(s_1) > u + x_1, ..., X(s_p) > u + x_p$:

1. likelihood:

2. $$p(x(s_1), ..., x(s_n) | k, \sigma(s_i), \eta(s_i) i = 1, ..., n) = \prod_{i=1}^{n} \eta(s_i) \frac{1}{\sigma(s)} (1 + \frac{k}{\sigma(s)} v(s_i))^{-1/k-1}, \quad (14)$$

3. Since $\eta(s_i) \in (0, 1)$, $\eta(s_i)$ can be connected to the latent process $W(s_i)$ through the logit function:

   $$\log \frac{\eta(s_i)}{1 - \eta(s_i)} = \eta_0 + W(s_i).$$

   specification of the other parameters would be same as in the previous model. Note that for any s_i, in principle, $\eta(s_i)$ is a very small positive

number, hence prior distributions of the parameters in the link function need to be finely tuned, otherwise it is very unlikely that the chain would converge. In general, we would recommend simplification by assuming that $\eta(s_i) = \eta$ and $\eta \sim Beta(0.1, 10)$.

Conditional independence assumption by nature produce non-smooth response surfaces. In order to remedy this situation and to relax conditional independence assumption, Sang(2008) suggests using the following structure for the first stage of hierarchy to obtain smoother response surface:

$$V_1(s) = \left(\frac{k(s)}{\sigma(s)} Z(s) + 1\right)^{1/k(s)}, \tag{15}$$

Here, conditional independence assumption implies that $Z(s)$ are iid unit GPD random variables. Sang(2008) suggests introducing extra independence through the $Z(s)$ by assuming that

$$Z(s) = G^{-1}(\Phi(Z^*(s)),$$

where G is the unit GPD, ϕ unit Normal distributions and $Z^*(s)$ is a spatio-temporal Gaussian process. However, we will not pursuit this generalization. However, Introduction of such extra dependence structure directly on the data (apart from dependence introduced through model parameters) evidently may make the model more flexible with smoother response surfaces, provided that the data have enough information to capture such detailed latent dependence structure.

Finally, above models can be extended to spatio-temporal data sets. If $V_1(s_i, t), i = 1, ..., n(t), t = 1, ..., T$ are the excesses over a fixed threshold, then a hierarchical model similar to (14) can be written, except that the latent 0 mean gaussian process $W(s)$ is now replaced by a stationary 0 mean spatio-temporal Gaussian process $W(s, t)$. Valid stationary covariance structures for such processes are given by Gneiding(2002). These covariance structures have few parameters and conceptually, there are no problems in implementing MCMC methods for these models. However, there are quite a few technical difficulties in implementing these models in practice. The choice of prior distributions may have big impact on convergence and it may not be possible to use vague priors for all the parameters and hyper-parameters. Other possible technical problem is the large dimension, although, in principle, this should not be a problem for data sets appearing in extreme value problems. However, data sets having more that 150 observations, will involve inversion of matrices of 150×150 in each updating iteration, which often results in failure in convergence. Banarjee et all(2006) discuss ways of overcoming this dimension problem.

Alternatively, one can use conditionaly independent GEV model as discussed previously. See for example Cooley et (2007), or more recently H. Sang(2008) on the strategies of implementing these models: Let $Y(s_i, t)$ denote the block maximum over some time period (say, annual) at some areal units or grid cells

(or observation stations) s_i. Then, using conditional independence assumption, the data can be modeled by

$$y(s_i, t) \sim GEV(\mu_{s_i,t}, \sigma_{s_i,t}, k_{s_i,t}).$$

Sang(2008) assumes

- $k(s_i, t) = k$,
- $\sigma(s_i, t) = \sigma_i$ are spatial random effects,
- $P(\mu(s_i, t)|\boldsymbol{\beta}, W(s_i, t), \tau^2) \sim N(\mathbf{x}_i^T\boldsymbol{\beta}, +W(s_i, t))$, where $\mathbf{x}_i$ are site specific explanatory variables, $\boldsymbol{\beta}$ regression coefficients and $W(s_i, t)$ is a spatial-temporal random effect.

There are many alternative ways of modeling the spatial, temporal variation in $W(s_i, t)$. One possibility is to assume

$$W(s_i, t) = \phi_i + \delta_t,$$

$$\delta_t = \psi\delta_{t-1} + w_t,$$

$$w_t \sim N(0, W_0^2)$$

Sang(2008) suggests modelling δ_i and ϕ_i jointly by using coregionalization CAR model:

- $\log \sigma_i = \sigma_0 + \lambda_i$,
  $$(\lambda_i, \phi_i)^t = (\mathbf{A})(V_{1,i}, V_{2,i})$$
- $(V_{1,i}, V_{2,i})$ *are two independent CAR models*
- $$\mathbf{A} = \begin{pmatrix} a_{11} & 0 \\ a_{12} & a_{22} \end{pmatrix} \tag{16}$$

More complicated space-time structures for the scale and location parameters can be given (Sang, 2008). However, it is doubthful that the extremal data will be rich enough to give support for such complicated structures. The choice of prior distributions are discussed in Sang(2008). The choice of these prior distributions is crucial. Choosing non-informative priors for all the hyper-parameters may not work, resulting in non-convergence of the chain.

If

$$\boldsymbol{y}_{T+1} = Y(s_1, T+1), ..., Y(s_n, T+1)$$

represent the annual maxima of the process at the grid cells for next time period, then they can be obtained by updating samples from the predictive density

$$p(\boldsymbol{y}_{T+1}|\text{Data}) = \int p(\boldsymbol{y}_{T+1}|\boldsymbol{\Theta})p(\boldsymbol{\Theta}|\text{Data})d\boldsymbol{\Theta},$$

where $\mathbf{\Theta}$ are all the model parameters. Implementation of these models, using metropolis-hasting algorithm is given in Sang(2008).

Conditional independence assumption by nature produce non-smooth response surfaces. In order to remedy this situation and to relax conditional independence assumption, Sang(2008) suggests using the following equation for the first stage of hierarchy to obtain smoother response surface:

$$Y(s,t) = \mu(s,t) + \frac{\sigma(s,t)}{k(s,t)}(z(s,t)^{k(s,t)} - 1). \tag{17}$$

Here, conditional independence assumption implies that $Z(s,t)$ are iid unit Frechet random variables. Sang(2008) suggests introducing extra independence through the $Z(s,t)$ by assuming that

$$Z(s,t) = G^{-1}(\Phi(Z^*(s,t)),$$

where G is the unit Frechet, ϕ unit Normal distributions and $Z^*(s,t)$ is a spatio-temporal Gaussian process. Here, $\mu(s,t)$,$\sigma(s,t)$, $k(s,t)$ are then modeled in the usual manner explained above. .

Conditional independence assumption, or its relaxed forms (4.1) and (15)are flexible ways of introducing dependence through the model parameters. Introducing structural dependence directly on the high dimensioned spatial data in general is not efficient, due to the fact that one has to model and work with high dimensioned non-Gaussian joint distributions. However, transformed Gaussian processes and the respective Gaussian copula, as suggested by Sang(2008) is particularly suited to hierarchical modeling: Let $z^*(s)$ be a 0 mean, unit variance Gaussian process. A transformed process based on this gaussian process can be constructed as

$$z(s) = G^{-1}\Phi(z^*(s)), \tag{18}$$

where $\Phi(.)$ is the standard normal distribution and $G(.)$ is the unit Frechet distribution. Given the correlation function of the $z^*(s)$ process the corresponding Gaussian copula function

$$C_{z^*}(u_1, ..., u_n) = F_{\mathbf{z}^*}(\phi^{-1}(u_1), ..., \phi^{-1}(u_n)),$$

where $u_i \in [0,1]$, and F_{z^*} is the joint distribution of $z^*(s_1), ..., z^*(s_1))$ can be constructed. If $F(\mathbf{z})$ is the joint distribution of $\mathbf{z}$, then

$$F(\mathbf{z}) = C_{\mathbf{z}^*}(\Phi^{-1}(G(z_1), ..., \Phi^{-1}(G(z_1)), \tag{19}$$

and it can be shown that the marginal distributions of (19) are unit Frechet. Hence, this construction results in a process with unit Frechet margins, whose dependence structure is completely determined by the Gaussian Copula, which in return is completely determined by the correlation structure of the $\mathbf{z}^*$ process. Although, this model can be adopted with some relative computational ease, introduced dependence structure is arbitrary, fully specified by a bivariate dependence specification and hence may not represent a flexible enough dependence structure for extremes.

Should one use the spatial GEV model or the spatial GPD model? Both models are justified by the asymptotic theory and they infer equally on the same dependence structure among extreme observations. If there are no long, time series of observations at each observational locations, it is difficult to define meaningful blocks and consequently block maxima data to fit the GEV model. Defining spatial blocks to create block-maxima data without taking into consideration strong local spatial dependence, is inconsistent with the global aim of quantifying the effect of spatial dependence on the extremes. Therefore GPD model seems to suit the spatial extremal data better. However, GPD model depends on the proper choice of the high fixed threshold. There are difficulties of introducing threshold uncertainty into the model. See For example, Tancredi et all (2002) or de Zea Bermudez et al (2001) for strategies of introducing threshold uncertainity into the model. Even when the fixed threshold is chosen successfuly, there is still the problem of estimation of the joint tail above the fixed threshold $P(X(s_1) > u, ..., X(s_p) > u)$, which can not be handled without assuming the extra conditions leading to model (14).

We now apply these hierarchical modeling ideas to a excess wildfire data set.

2.1 A case study

In many fire regimes, a small number of very large fires is responsible for the vast majority of the area burned and of the social and environmental damage caused. Therefore, large wildfires are a relevant public policy issue , especially considering that the frequency of occurrence of extremely severe fire weather may increase, as a consequence of global warming Large fires are, from various perspectives, qualitatively different from small fires. Large fires tend to occur under specific, relatively uncommon synoptic meteorology conditions, typically involving variable combinations of high temperatures, prolonged drought and strong winds. They display extreme fire behavior patterns and spread mechanisms not observed in small fires. Fighting very large fires is qualitatively different from fighting small fires, because the range of options for controlling large fires is drastically reduced, in comparison with the diversity of options available for initial attack. The degree of organizational complexity required for fighting large fires scales up nonlinearly and logistical aspects take over tactical considerations. The ecological effects of large fires also may differ qualitatively from those of smaller events. It is therefore important to study extreme wildfires regimes separately. Evident existence of Spatio-temporal variations and dependencies, as well as the strong effect of meteorological, ecological and topological covariates demand very flexible, highly parameterized models for the extremes of wild fires. A review of the study of wildfire extremes can be found in Bermudez et al (2007).

The study area corresponds to mainland Portugal (see Figure 2), located between 37N and 42N latitude and between 6W and 10W longitude. Geographical stratification of the study area represents a compromise between the Portuguese Forest Service standard procedure of organizing statistical fire data by administrative units, and the natural regions classification of Portugal. The 18

administrative districts of Portugal were grouped into eight geographical units that closely approximate the 12 natural regions, based on similarities of climate, topography, vegetation, land use, population density, and fire incidence, namely number of fires and area burned.

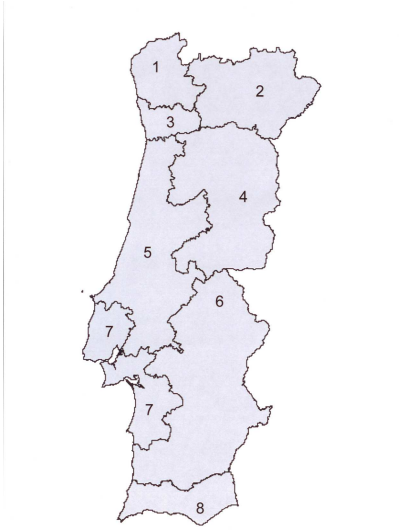

Figure 2: Geographical Regions

The data consist of 36180 size records of wildfires larger than 5ha, observed in Portugal between 1984 and 2004. Fire perimeters were mapped from Landsat 5 Thematic Mapper and Landsat 7 Enhanced Thematic Mapper satellite imagery, with 30m spatial resolution, to a geographical scale of 1:100.000. About 170 satellite images, acquired annually after the end of the summer fire season, were analyzed over the 20-year period. During the exceptional fire season of 2003, several large fires coalesced into huge, continuous fire scars, corresponding to multiple individual fire events. It is unfeasible, using post-fire season Landsat imagery to split these composite fire scars into their individual components. Field data, from both the Forest Service and the Civil Protection Service were also found inadequate to accomplish this task, due to geographical and or temporal inaccuracies and missing information. Thus, work is under progress using daily NOAA/AVHRR imagery, at 1km spatial resolution to unravel these cases. These possible outliers are kept as part of the data, but they are highly influential and introduce bias to inference, making the data more heavy tailed that possibly is.

In Figure 3, the spatial locations of all recorded fires above 5 hectares as well as the location of fires above 250 hectares are given for the years 2000 and 2004.

At present, no data on climatological, ecological as well as topological co-

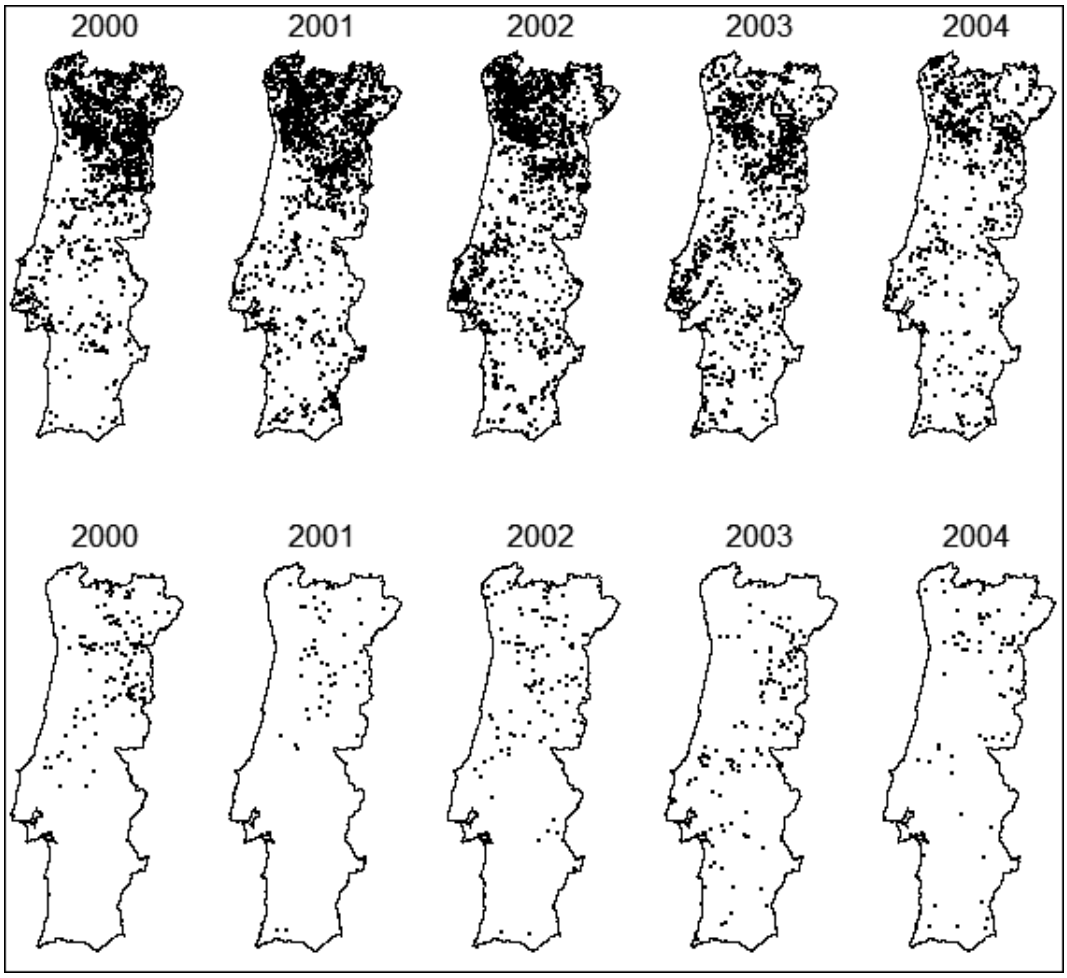

Figure 3: Locations of fires above 5 hectares (top) and locations of fires above 250 hectares (bottom)

Figure 4: Administrative regions at district level, in Portugal

variates are available for the analysis. A latent spatially and temporally colored random effect will be used to describe the spatial-temporal variations due to these covariates.

Hence our data set is of the form

$$\{x(\mathbf{s}_i, t), i = 1, ..., n_t, t = 1, 2, ..., 30\}, \tag{20}$$

where $\mathbf{s} = (s_1, s_2)$ are the spatial coordinates of the centroids of the recorded fires, n_t is the number of wildfires with sizes above 35 hectares in year t, $t = 1975, ..., 2004$.

2.2 Models

In order to account for the spatial variation of the large fire sizes, we divide up the data into 18 administrative regions $\mathbf{D}_j, j = 1, 2, ..., 18$, called *districts* (see Figure 4). This geographical stratification of the study area represents a compromise between the Portuguese Forest Service standard procedure of organizing statistical fire data by administrative units, and the natural regions classification of Portugal.

Let $N_j(t)$ denote the number of fires with sizes over 250 hectares, observed in region j, during year t. We suggest the following model for the excesses over the threshold of 250 hectares conditional on $(n_j(t), j = 1, ..., 18, t = 1, ..., 30)$, the observed number of fires in excess of 250 hectares in each region and year to account for the spatial variation in the excess data:

1. **Level 1: likelihood**

$$p(z(\mathbf{s}_{ij},t), i=1,2,...,n_j(t), t=1,...,30; j=1,...,18|(n_j(t), j=1,...,18, t=1,...,30), \boldsymbol{\eta}_\sigma, \boldsymbol{\Theta})$$
$$= \prod_{t=1}^{30}\prod_{j=1}^{18}\prod_{i=1}^{n_j(t)} p(z(\mathbf{s}_{ij},t)|k,\sigma(\mathbf{s}_{ij},t),u(s_{ij},t)), \tag{21}$$

where,

$$\begin{aligned} p(z(\mathbf{s}_{ij},t)) &= GPD(k,\sigma(\mathbf{s}_{ij},t)) \\ &= \frac{1}{\sigma(\mathbf{s}_{ij},t)}\left(1+\frac{k}{\sigma(\mathbf{s}_{ij},t)}z\right)^{-1-1/k} \end{aligned} \tag{22}$$

in year t. Here the location $\mathbf{s}_{ij}$ belongs to region j, indicating the ith location belonging to the jth region.

2. **Level 2: link Functions**

   (a)

   $$\log(\sigma(s_{ij},t)) = \sigma_0 + W(j) + W_j(s_{ij}) + \alpha_1\mathbf{N}_1(j,t-1) + \alpha_2\mathbf{N}_2(j,t-1) \tag{23}$$

   Here $\mathbf{N}_1(j,t-1)$ and $\mathbf{N}_2(j,t-1)$ are exploratory variables, denoting respectively the (normalized) average burned area and (normalized) average number of fires in $t-3$, $t-2$ and $t-1$ in *district* j, whereas α_1, α_2 are scalar regression parameters. Here, $W(j)$ is a latent discrete process modeled by a nearest neighbor CAR model (see Banarjee et al,2004) representing low resolution dependencies between the regions, and $Wj(s_{ij})$ is another latent, region specific Gaussian process representing high resolution dependencies within each region. Here, $W(i)$ represent the unobserved explanatory variables acting on larger distances, such as temperature and other global atmospheric conditions, whereas $W_j(s)$ represent unobservent local explanatory variables such as vegetation, wind speed, wind direction,local topological conditions etc.

   (b) For mathematical simplicity, we will assume that the shape parameter has no spatial structure, although this is not realistic based on the information we have from the historical data. Theoretically, the shape parameter k can take values in $(-\infty,+\infty)$. However, preliminary data analysis of historical data indicated that the shape parameter has a mode around 0.47. Hence we use as prior for k a vague beta distribution, restricting it to the interval $[0,1]$. This prior distribution matches with the expert opinion that the fire size distribution should have finite mean but heavy tailed. Hence, we assume

   $$k \sim \text{beta}(0.5, 0.5). \tag{24}$$

3. **Level 3: parameters, priors, hyper-parameters and hyper-priors** We will not discuss the in detail the prior distributions for the hyper parameters here. Ideally, all prior distributions should be vague, but convergence without being specific for some of the hyperparameters can not be guaranteed. Therefore priors for some of the precision parameters need to be informative.

Note that the likelihood is conditional on the fixed values of the The counting processes $N_j(t)$, $j = 1, ..., 18$, $t = 1, ..., 30$. These processes can also be modeled by BHM, but we will not discuss this problem here. See Mendes et all (2008) for the details of these models and the model fitting details. We only remark that such models can be implemented by using BUGS. More complicated models where the shape parameter is assumed to vary spatially are also implimented using the BUGS software.

3 Leadbetter-Albin Asymptotics

We start with a spatial process $X(\mathbf{s})$, $\mathbf{s} \in S \subset R^2$, continuous in space and give an asymptotic expression for

$$P(\sup_{\mathbf{s}\in S} X(\mathbf{s}) > u), \tag{25}$$

as $u \to \infty$ for fixed intervals $S = [0, h_1] \times [0, h_2]$.

This asymptotic expression will be given in terms of the stationary distribution $P(X(0,0) > u)$, as well as in terms of the local clustering of large values of the process. This characterization is based on starting with a sufficiently accurate discrete version of the spatial process over a discrete grid, deriving the distribution of the maximum over that discrete grid, then using limiting arguments to tie the results to the maximum of the continuous process. Albin(1987, 1990) gives the sufficient conditions and the characterization of the limiting distribution for

$$P(\sup_{t\in[0,h]} X(t) > u),$$

as $u \to \infty$, for fixed intervals $[0, h]$, where $X(t)$ is a stationary process. Albin(1987, 1990) also extend the results to increasing intervals. See also Leadbetter et al (1983). There are basically 4 sets of conditions to achieve such results:

- Existence of a sufficiently fine grid over which the maxima of continuous process and its discrete version over this grid asymptotically match. Note that when $X(t)$ is a separable process, then there is always a countable version, and further, if the process is continuous in probability then the dyadic numbers are a separant. However, for many processes, the minimal grid need not be so dense.
- Behavior of the process in a very small neighborhood of a upcrossing, which defines the formation of clusters of large values.

- A condition on the local dependence behavior of the process, restricting the size of these clusters by assuming that the exceedances are short lasting. This condition in many ways resemble and is in the same spirit as the condition $D'(u_n)$ condition often assumed for obtaining asymptotic results for the extremes of discrete random sequences.

- And finally, for extending the results to increasing intervals, two further dependence assumptions are needed,one being a mild mixing condition, other restricting the formation of clusters of clusters.

We assume similar set of conditions for the stationary spatial process and by extending the Albin(1987) techniques, we obtain an asymptotic expression for (25), when the process is heavy tailed, although results extend to other types of tail behaviour. Here, we give a brief summary of the results. Detailed calculations can be found in Anderson and Turkman(2008)

4 Main results

We will assume that the stationary field $X(t_1, t_2)$ is also isotropic, satisfying the following conditions:

- (1) *Domains of attraction criteria:*

  Assume that there exists a strictly positive constant c such that

  $$\lim_{u\to\infty} \frac{1-F(ux)}{1-F(u)} = x^{-c}, \tag{26}$$

  for all $x > 0$, so that F belongs to the Frechet domain of attraction.

- 2 *Minimal discrete approximation:*

  Assume that there is a positive, non-increasing function $q = q(u)$, with $\lim_{u\to\infty} q(u) = 0$, such that as $a \to 0$,

  $$\frac{q^2(u)}{1-F(u)} P(M(h_1, h_2) > u, \max_{\substack{0\le iaq\le h_1,\\ 0\le jag\le h_2}} X(iaq, jag) \le u) = 0, \tag{27}$$

  where $M(h_1, h_2) = \max_{(t_1,t_2)\in[0,h_1]\times[0,h_2]} X(t_1, t_2)$. We call the grid

  $$\delta = \{(iaq, jaq), i = 0, 1, ..., [h_1/aq], j = 0, 1, ..., [h_2/aq]\}$$

  satisfying (27), the Pickand grid. For ease in notation, let

  $$M_\delta(h_1, h_2) = \max_{\substack{0\le iaq\le h_1,\\ 0\le jag\le h_2}} X(iaq, jag),$$

  the maximum of the process over the Pickand grid.

- 3 *Local clusters:*

  For any $t_1 \in [0, h_1]$, let

  $$Y(t) = \max_{0 \le t_2 \le h_2} X(t_1, t_2),$$

  be column-wise maximal process traveling along the x-direction and

  $$Y(iaq) = \max_{0 \le j \le [h_2/aq]} X(iaq, jaq),$$

  for $i = 0, 1, 2..., [h_1/aq]$ be the discrete version of $Y(t)$ over the pickand grid. For any integer $N > 0$, and fixed i, let

  $$Y_N(iaq) = \max_{0 \le j \le N} X(iaq, jaq),$$

  and assume that

  there exists random variables $\{\eta_a(k)\}_{k=1}^N$, such that for any $a > 0$,

  $$\left(\frac{1}{u} X(0, jaq), i = 1, .., N | \frac{1}{u} X(0,0) > 1 \right) \to^D \{\eta_a(k)\}_{k=1}^N, \qquad (28)$$

  and further there exists random variables $\{\zeta_a(k)\}_{k=1}^N$ such that for any $a > 0$,

  $$\left(\frac{1}{u} Y_N(iaq), i = 1, .., N | \frac{1}{u} Y_N(0)) > 1 \right) \to^D \{\zeta_a(k)\}_{k=1}^N. \qquad (29)$$

- 4 *Short lasting exceedances:*

  as $N \to \infty$, for every $a > 0$,

  $$\lim_{u \to \infty} \frac{1}{1 - F(u)} \sum_{i=0}^{h_1/aq} \sum_{j=0}^{h_2/aq} \sum_{\sqrt{i^2+j^2} \ge N} P(X(0,0) > u, X(iaq, jaq) > u) = 0. \qquad (30)$$

Albin(1987) assume the condition (28) for the characterization of local clusters of large values of a continuous process. (29) is an extra condition needed to extend the formation of clusters in the case of random fields. Note that the process $Y(t)$ is the y-column maxima of the field traveling along the x-direction, and condition (28) explains how this process clusters along the x-direction in a similar fashion to (28). The characterization of local clusters over a high threshold using the characterization of local clusters in terms of coordinate-wise processes is mathematically very convenient. However, it is may not be easy to verify condition (29). Hence, one may want to characterize the formation of local clusters by assuming a similar condition to (28), directly on the random field, such as for example, the existence of the limit

$$\left(\frac{1}{u} X(iaq, jaq), i, j = 1, ..., N | X(0,0) > u \right) \to^D \{\theta_a(k_1, k_2)\}_{i,j=1}^{N,N} \qquad (31)$$

However, this seems to be difficult, if not impossible.

Let $M(h_1, h_2) = \max_{\substack{0 \le t_1 \le h_1, \\ 0 \le t_2 \le h_2}} X(t_1, t_2)$.

Theorem 1

Under the assumptions (26)-(30), the limits

$$H_1 = \lim_{a \to 0} \frac{1}{a} P(\max_{k \ge 1} \eta_a(k) < 1), \tag{32}$$

and

$$H_2 = \lim_{a \to 0} \frac{1}{a} P(\max_{k \ge 1} \zeta_a(k) < 1), \tag{33}$$

exist with $H_1 \in (0, \infty)$, $H_2 \in (0, \infty)$ and

$$\lim_{u \to \infty} \frac{q^2(u)}{1 - F(u)} P(M(h_1, h_2) > u) = h_1 h_2 H_1 H_2, \tag{34}$$

4.1 Alternative conditioning

An alternative form of theorem 1 can be given if we condition our probabilities on the event $\{X(0,0) = ux\}$ and $\{Y_N(0) = ux\}$ for some $x > 1$, rather than on $\{X(0,0) > u\}$ and $\{Y_N(0) > u\}$ as is given in section 2.

Assume that

- $F(x)$ and $G(x)$have densities f_1 and f_2 such that

$$\lim_{u \to \infty} \frac{u f_1(u)}{1 - F(u)} = c > 0$$

and

$$\lim_{u \to \infty} \frac{u f_2(u)}{1 - G(u)} = d > 0.$$

Further assume that there exists variables $\{\eta_{a,x}\}_{k=1}^N$,$\{\zeta_{a,x}\}_{k=1}^N$ such that for every $a > 0$ and $x > 1$,

$$\left(\frac{1}{u} X(0, jaq), i = 1, .., N | \frac{1}{u} X(0,0) = x \right) \to^D \{\eta_{a,x}(k)\}_{k=1}^N,$$

and

$$\left(\frac{1}{u} Y_N(iaq), i = 1, .., N | \frac{1}{u} Y_N(0)) = x \right) \to^D \{\zeta_{a,x}(k)$$

Assume further that the approximation

$$\limsup_{u \to \infty} \frac{1}{a^2(1 - F(u))} P(M(aq, aq) > ux(1 + \delta), X(0,0) \le ux) = 0, \tag{35}$$

holds for every $\delta > 0$ and $x \ge 1$.

Then

Theorem 4

$$\lim_{u\to\infty} \frac{q^2(u))}{1-F(ux)} P(M(h_1,h_2) > ux) = h_1 h_2 H_1(x) H_2(x), \tag{36}$$

where

$$H_1(x) = \lim_{a\to 0} \frac{1}{a} P \int_1^\infty P((\max_{k\geq 1} \eta_{a,xy}(k) < 1) c y^{-(c+1)} dy, \tag{37}$$

and

$$H_2(x) = \lim_{a\to 0} \frac{1}{a} P \int_1^\infty P((\max_{k\geq 1} \zeta_{a,xy}(k) < 1) dy^{-(d+1)} dy. \tag{38}$$

With this conditioning in hand, we can state the following Theorem:

Theorem

For any $x > 0$ fixed,

$$\lim_{u\to\infty} \frac{q^2(u)}{1-F(u)} P(M(h_1,h_2) > ux) = x^{-c} H_1(x) H_2(x).$$

At first sight, this result suggest that the limiting distribution may not be in the domain of attraction of the GEV. However, it can be shown that the function $q(u)$ is regularly varying and $H_1(x)H_2(x) \sim x^{-d}$ for some $d < c$ so that

$$\lim_{u\to\infty} \frac{q^2(u)}{1-F(u)} P(M(h_1,h_2) > ux) = x^{-c^*} H_1 H_2,$$

for some $c^* < c$.

How can these asymptotics be used for finding meaningful models for spatial extremes? For continuous time series, Albin(1990) looks at some special non-Gaussian processes (rayleigh type process, ie squares of Gaussian processes), verify conditions and derive the local cluster parameter H. Such exercise would be quite difficult for spatial processes, as one needs to know the sample path properties of the column maxima process $Y(t)$. However, since H_1 and H_2 are coordinate wise clustering of two processes along the coordinates, discrete approximations to such quantities can be made by using the known methods of extremal index estimation. However, at present, such asymptotical results are not particularly useful in modelling spatial extremes.

5 References

Anderson C.W. and K.F. Turkman(2008) Maximum of a Stationary Random Field. In preparation.

Albin, P. (1987) On Extremal theory for non differentiable stationary processes. P.hD. Theses, Dept of Mathematical Statistics, University of Lund

Albin, P. (1990) On extremal theory for stationary processes. *The Ann. of Probab.* 18, 92-128.

Buishand, A. de Haan, L. and C. Zhou(2007) On spatial extremes: with application to a rainfall problem. J. Applied Statistics(?).

Beirland,J. Goegebeur,Y. Segersd,Y. and J. Teugels(2004) Statistics of Extremes. Wiley.

De Haan, L and T. Pereira (2006) spatial extremes: Models for the stationary case. *The annals of Statistics,* 34, 146-168.

Banarjee, S., Carlin, B. P. and Gelfand, A. (2004). Hierarchical modelling and Analysis for spatial data. Chapman and Hall.

P. de Zea Bermudez, J. Mendes , J.M.C. Pereira, K.F. Turkman, and M.J.P. Vasconcelos (2008) Spatial and temporal extremes of wildfire sizes in Portugal(1984-2004). Submitted to International journal of wildland fires.

Casson, E. and S. Coles (1999) Spatial regression models for extremes. *Extremes*, **1**, 449-468.

Coles,S. (2001) An Introduction to Statistical Modelling of Extreme values. Springer.

Coles, S. and Tawn, J. (1996) Modelling Extremes of the areal rainfall process. *Journal of Royal Statistical Society, B* **58:2**, 329-347

Cooley, D., Nychka, D. and Naveau, P. (2006). Bayesian spatial model of extreme precipitation return levels. JASA.

Cooley, D., Neveau, P., Joneli, V. Rababtel, A. and D. Grancher (2005). A Bayesian hierarchical extreme value model for lichenometry. *Environmetrics.* In press.

Gneiding,T. (2002) Nonseparable, stationary covariance functions for space-time data. JASA 97, 590-600.

Heffernan,J.E. and J. Tawn(2004) A conditional approach for multivariate extreme values. J.R. Statist. Soc. B, 66, 497-546.

Leadbetter, M.R., Lindgren,G. and H. Rootzen (1983) Extremes and Related properties of random sequences and processes. Springer-Verlag

Mendes,J. de Zea Bermudez, Pereira,J. and K.F. Turkman(2008)Spatial extremes of wildfire sizes: Bayesian hierarchical models for extremes. Submitted to EES.

Piterbarg, V. (1996) Asymptotic methods in the theory of Gaussian Processes and fields. AMS monographs.

Resnick,S. (1987) Extreme values, Regular variation and point processes. Springer-Verlag.

Huiyan Sang, 2008 "Extreme Value Modeling for Space-time Data with Meteorological Applications". Phd theses, Duke University.

Tancredi, A. Anderson, C.W. and A. O'Hagan (2001) Accounting for threshold uncertainity in extreme value estimation Research report, University of Sheffield

de Zae Bermudez, P. Amaral Turkman M.A and K.F. Turkman(2001) A predictive Approach to Tail Probability Estimation. Extremes, 4, 295-314.