Hydrol. Earth Syst. Sci. Discuss., 11, 12731–12764, 2014 www.hydrol-earth-syst-sci-discuss.net/11/12731/2014/ doi:10.5194/hessd-11-12731-2014 © Author(s) 2014. CC Attribution 3.0 License.



This discussion paper is/has been under review for the journal Hydrology and Earth System Sciences (HESS). Please refer to the corresponding final paper in HESS if available.

Spatial extremes modeling applied to extreme precipitation data in the state of Paraná

R. A. Olinda¹, J. Blanchet², C. A. C. dos Santos³, V. A. Ozaki⁴, and P. J. Ribeiro Jr. 5

¹Department of Statistics, Center for Science and Technology, Paraíba State University – UEPB, Campina Grande, Paraíba, Brazil

²Ecole Polytechnique Fédérale de Lausanne, EPFL-FSB-MATHAA-STAT, Station 8, 1015 Lausanne, Switzerland

³Academic Unit Of Atmospheric Sciences – UACA/CTRN, Federal University of Campina Grande – UFCG, Campina Grande, Paraíba, Brazil

⁴Department of Exact Sciences, University of São Paulo, School of Agriculture Luiz de Queiroz, 13418-900, Piracicaba, São Paulo, Brazil

⁵Department of Statistics, Federal University of Paraná, Polytechnic Garden Center of the Americas, 81531-990, Curitiba, Paraná, Brazil



Received: 7 October 2014 – Accepted: 27 October 2014 – Published: 17 November 2014 Correspondence to: R. A. Olinda (prof_ricardo@cct.uepb.edu.br)

Published by Copernicus Publications on behalf of the European Geosciences Union.



Abstract

Most of the mathematical models developed for rare events are based on probabilistic models for extremes. Although the tools for statistical modeling of univariate and multivariate extremes are well developed, the extension of these tools to model spatial

- ⁵ extremes includes an area of very active research nowadays. A natural approach to such a modeling is the theory of extreme spatial and the max-stable process, characterized by the extension of infinite dimensions of multivariate extreme value theory, and making it possible then to incorporate the existing correlation functions in geostatistics and therefore verify the extremal dependence by means of the extreme coefficient and
- the Madogram. This work describes the application of such processes in modeling the spatial maximum dependence of maximum monthly rainfall from the state of Paraná, based on historical series observed in weather stations. The proposed models consider the Euclidean space and a transformation referred to as space weather, which may explain the presence of directional effects resulting from synoptic weather pat-
- terns. This method is based on the theorem proposed for de Haan and on the models of Smith and Schlather. The isotropic and anisotropic behavior of these models is also verified via Monte Carlo simulation. Estimates are made through pairwise likelihood maximum and the models are compared using the Takeuchi Information Criterion. By modeling the dependence of spatial maxima, applied to maximum monthly rainfall data
- from the state of Paraná, it was possible to identify directional effects resulting from meteorological phenomena, which, in turn, are important for proper management of risks and environmental disasters in countries with its economy heavily dependent on agribusiness.



1 Introduction

Statistical modeling of spatial extremes can be used to solve a number of real problems, both in the urban and rural environment. In many regions of Brazil, where long climatic records are available, it has been observed an increased frequency of extreme events that partly explains the growing number of natural disasters such as landslides and floods, which are responsible for an alarming number of deaths in large cities and

disasters in the agricultural sector.

An example is the Southern region of Brazil, which has more favorable atmospheric conditions for storm formation in the Southern Hemisphere (Brooks et al., 2006). There has been a large increase of 109.5 extreme events year⁻¹ during the period 1984 to

- ¹⁰ has been a large increase of 109.5 extreme events year ⁻¹ during the period 1984 to 1993, to 127.4 extreme events year⁻¹ for 1994–2003, for the State of Paraná (Marcelino et al., 2006). In January 2011, about 83 municipalities in the state of Paraná enacted emergency because of heavy rains. The civil defense recorded six deaths and about 930 000 people were affected. More than 26 000 people had to leave their homes
- throughout the state; rains caused damage estimated at USD 160 million. In the agricultural context, the study and quantification of the impacts of climate change on agriculture is of fundamental importance to the world, and particularly for Brazil and Latin America, which has its economy deeply dependent on agribusiness.

Most mathematical models developed for rare events are based on probabilistic ²⁰ models for extremes, adjusting the data through statistical techniques. Studies have shown that these techniques are insufficient to deal with the task of modeling georeferenced extremes or spatial extreme (Ribatet et al., 2011). Spatial extremes modeling attempts to describe the patterns existing in spatial data as well as establish, preferably in a quantitative manner, the relationship between different geographical variables and

their extremes (Sang and Gelfand, 2010; Kunihama et al., 2012; Kojadinovic and Yan, 2012). Max-stable processes are the well-founded methodology for that Schlather and Tawn (2003), Hashorva (2006), Smith and Stephenson (2009), Coles and Tawn (1991) and Coles (1993).



The max-stable process arises from an infinite-dimensional generalization of Extreme Value Theory or Extreme Value Analysis (for short, EVA), which, in turn, provides a natural generalization of the extremal dependence structures in continuous spaces. This process was developed initially by De Haan (1984). Since then, several contributions such as Smith (1990), Schlather (2002), Kabluchko et al. (2009), Ribatet et al. (2011), Davison and Gholamrezaee (2012) and Blanchet and Davison (2011) made it more suitable for real data. Recent applications for rainfall data can be found in Buishand et al. (2008), Smith and Stephenson (2009), Padoan et al. (2010), Davison et al. (2012), Sunyer et al. (2013), Uboldi et al. (2013), Gräler et al. (2013), Gilleland et al.
(2013), snow extremes Blanchet and Lehning (2010), Blanchet and Davison (2011) and maximum temperature data Davison and Gholamrezaee (2012).

The statistical modeling of spatial extremes can be used both in urban and rural environments. In many regions of Brazil, there has been an increase in the frequency of extreme events which, to some extent, explain the increasing number of natural

- ¹⁵ disasters, such as landslides and floods, which, in turn, are responsible for an alarming number of deaths in large cities, as well as for disasters in the agricultural sector. Frosts, heat and cold waves, pouring rains, floods, dry spells, and other extreme events affect the crops, making the financial segment and the rural insurance segment sensitive to climate change.
- This being the case, the research aims at: (1) modeling spatial extremes in order to contribute to the prediction of these, (2) evaluating the behavior of the extreme spatial models through cross-validation, together with the pairwise maximum likelihood estimation, (3) and quantifying, through maps, the risks associated with extreme weather data related to the State of Paraná.

25 2 Max-stable process

Max-stable processes are an extension of multivariate extreme value theory in infinite dimensions (Hsing et al., 2004). Thus, when trying to fit such processes to data, the



same difficulties arise as regards the finite dimensional case. Therefore we provide below a brief review of univariate and multivariate extreme value theory, and subsequently we present max-stable processes according to the models proposed in Smith (1990) and Schlather (2002).

5 2.1 Extreme value theory for univariate random variables

Let $Y_1, ..., Y_n$ be a sequence of independent and identically distributed (*i.i.d.*) random variables of the maximum monthly rainfall at a site. Let $M_n = \max\{Y_1, Y_2, ..., Y_n\}$ be the maximum value of *n* monthly maximum rainfall values. In studying extreme rainfall values, we are interested in studying the distribution of M_n , not the distribution of y_i . If a sequence of pairs of real numbers (a_n, b_n) exists such that each $a_n > 0$ and

$$\lim_{n \to \infty} P\left[\frac{M_n - b_n}{a_n} \le y\right] = F(y),\tag{1}$$

where *F* is a non degenerate distribution function, then the limit distribution *F* belongs to either the Gumbel, the Fréchet or the Weibull family (Coles, 2001). These distributions can be grouped into the generalized extreme value (GEV) distribution. By means of Eq. (1), it is possible, then, to estimate the asymptotic distribution of $\frac{M_n - b_n}{a_n}$ directly from the GEV distribution without making reference to the distribution of *Y*. The following probability density function, that includes the three types of distribution of extreme values, can be used to describe analytically a series of maximum monthly rainfall of a meteorological station

²⁰
$$F(y;\mu,\sigma,\xi) = \exp\left\{-\left[1-\frac{\xi(y-\mu)}{\sigma}\right]^{1/\xi}\right\},$$

where $1 - \xi(y - \mu) / \sigma > 0, \sigma > 0$.

10

Shang et al. (2011), with GEV models presented an analysis of extreme values of the annual maximum rainfall in Debre Markos, Ethiopia.



(2)

2.2 Extreme value theory for multivariate random variables

Let { $Y_i = (Y_{i1}, ..., Y_{iK})$, i = 1, 2, ...} be a sequence of *i.i.d.* random variables of the maximum monthly rainfall at *K*-dimensional observation from a distribution function F, in our study, *K* represents the number of meteorological stations, i.e., K = 232. The lack of a natural order in a *K*-dimensional space results in the fundamental difficulty of defining a multivariate extreme observation (Cooley et al., 2012). Similarly to the univariate case, to avoid degeneracy of the limiting distribution of M_n , one looks for normalizing sequences, a_n and b_n , with components $a_{ni} > 0$ and $b_{ni} \in \mathbb{R}$, j = 1, ..., K, such that,

$$\lim_{n \to \infty} P\left[\frac{\boldsymbol{M}_n - \boldsymbol{b}_n}{\boldsymbol{a}_n} \le \boldsymbol{y}\right] = F^n(\boldsymbol{a}_n \boldsymbol{y} + \boldsymbol{b}_n) \to Z(\boldsymbol{y}), \tag{3}$$

for a *K*-dimensional distribution function *Z* with non-degenerate margins. Here and throughout all vector operations are intended componentwise. If (3) holds, *F* is said to belong to the *domain of attraction* of *Z*, denoted by $F \in D(Z)$, and *Z* is called a *multivariate extreme value* distribution function.

2.3 Max-stable processes for spatial random variables

¹⁵ Besides taking into consideration many stations, our study focuses on the extreme spatial dependence, for this, we take into account the geographic location of the meteorological stations, i.e., *S* be an arbitrary set $\{Y_i(s)\}_{s\in S}$, i = 1, ..., n be *n* repetitions of a continuous random process. Suppose there are two numerical sequences $a_n(s) > 0$ and $b_n(s) \in \mathbb{R}$ such that

²⁰
$$\lim_{n\to\infty} \frac{\max_{i=1}^{n} Y_i(s) - b_n(s)}{a_n(s)} \to Z(s), s \in \mathcal{S}.$$

If Z is non degenerate, then, Z(s) is a max-stable process.

(4)

The Eq. (4) shows that the class of limiting processes corresponds to the class of max-stable processes, and therefore justify their use for modeling extreme spatial rainfall maxima, for example. De Haan (1984) proposed a general way of constructing such max-stable processes, from which Smith (1990) and Schlather (2002) derived particular cases that may be used in practice. We detail such representations in the next two sections.

2.4 Smith's Model

Smith's Storm Model: let $\{(\eta_i, w_i), i \in \mathbb{N}\}$ denote the points of a Poisson process on $(0, \infty) \times \mathbb{R}^2$ with intensity $\eta^{-2} d\eta \times v(dw)$, where v is a positive measure in \mathbb{R} . Let $\{f(w, s), w \in \mathbb{R}^d, s \in S\}$ denote a non-negative function for which, for all $s \in S$,

 $\int_{\mathbb{R}^d} f(w,s)v(\mathrm{d}w) = 1.$

Then the random process.

$$Z^* = \{\max_{i \in \mathbb{N}} \{\eta_i f(w_i, s)\}, s \in \mathcal{S}\}$$

is max-stable process.

Smith (1990) gives the following physical interpretation to the process defined in Eq. (5):

- 1. the space \mathbb{R}^d can be interpreted as a space of "storm centers", for example, "storm locations";
- 2. the function $f(w_i, \cdot)$ represents the shape of a storm centered at w_i , whereas η refers to a storm magnitude;



(5)

20

15

3. finally, $\eta_i f(w_i, s)$ represents the amount of rainfall received at location *s* for a storm of magnitude η_i centered at *s* and $Z^*(\cdot)$ in Eq. (5) is the maximum rainfall received at *s* over an infinite number of independent storms.

While for $K \ge 2$ the general *K*-dimensional distribution function under the max-stable ⁵ process representation Eq. (5) has no analytical tractable form, a class of spatial model for which the bivariate distribution is tractable is obtained by considering *f* to be a Gaussian density in R^d and *v* the Lebesgue measure (Padoan et al., 2010), *f* is equation is a bivariate gaussian density and *v* is a Lebesgue measure. In this case, for locations s_1 and s_2 the bivariate distribution function is defined by

 $P[Z(s_i) \le z_i, Z(s_j) \le z_j] =$

20

$$\exp\left[-\frac{1}{z_i}\Phi\left(\frac{a(\boldsymbol{h})}{2}+\frac{1}{a(\boldsymbol{h})}\log\frac{z_j}{z_i}\right)-\frac{1}{z_j}\Phi\left(\frac{a(\boldsymbol{h})}{2}+\frac{1}{a(\boldsymbol{h})}\log\frac{z_i}{z_j}\right)\right],$$

where $h = (s_1 - s_2)^T$, Φ is the standard gaussian distribution function, $a(h) = (h^T \Sigma^{-1} h)^{1/2}$ is the Mahalanobis distance and Σ is the *d*-dimensional covariance matrix of *f*, with, when d = 2, covariance σ_{12} and SD $\sigma_1, \sigma_2 > 0$.

2.5 Schlather's Model

¹⁵ Schlather's Storm Model: let $W(\cdot)$ denote a stationary process in \mathbb{R}^d so that $E[\max\{0, W(s)\}] = 1$ and $\{\eta_i, i \in \mathbb{N}\}$ denotes points of a Poisson process on \mathbb{R}_+ with intensity measure $\eta^{-2} d\eta$. Then the process defined by

$$Z^* = \max_{i \in \mathbb{N}} \left\{ \eta_i \right\} \times \max \left\{ 0, W_i(s) \right\},$$

where $W_i(\cdot)$ are *i.i.d.* copies of the random process $W(\cdot)$ is a stationary max-stable process with unit Fréchet margins.

Like Smith's model, the process defined in Eq. (7) also has a simple interpretation. Imagine $\eta_i W_i(\cdot)$ as daily spatial rainfall events, differing only in their magnitudes η_i .



(6)

(7)

The difference with Smith's model is that the shape of the storms is driven by a random process $W(\cdot)$ in Eq. (7) instead of being given by a deterministic function f in Eq. (5). Actually, Smith's model can be seen as a particular case of Schlather's model when $W_i(s) = f_0(w - s_i)$ where f_0 is a probability density function and s_i is a homoge-⁵ neous Poisson process on \mathbb{R}^d .

Additional assumptions are again needed to obtain useful models from Eq. (7). Schlather (2002) proposes to adopt $W_i(\cdot)$ as the positive part of a stationary Gaussian process with correlation function $\rho(h)$ scaled so that $E[\max\{0, W_i(s)\}] = 1$. In this work, nine correlation functions were adopted. Based on these new assumptions, it can be shown that the bivariate cumulative distribution function is defined by

$$P[Z^*(s_1) \le z_1, Z^*(s_2) \le z_2] = \exp\left[-\frac{1}{2}\left(\frac{1}{z_1} + \frac{1}{z_2}\right)\left(1 + \sqrt{1 - 2(\rho(h) + 1)\frac{z_1 z_2}{(z_1 + z_2)^2}}\right)\right],$$
(8)

where $h \in \mathbb{R}_+$ is a vector of Euclidean distances $||s_1 - s_2||$ between two stations.

2.6 Spatial dependence and the extremal coefficient

20

In literature there are many related metrics to quantify the dependence of the tail when the random vector exhibits asymptotic dependence. In this work we use two metrics: 15 the extremal coefficient (Schlather and Tawn, 2003), which is easily interpretable and madogram (Cooley et al., 2006), which has ties to the semivariogram.

Let $Z(\cdot)$ be a stationary max-stable random field with unit Fréchet margins. The extremal dependence at locations K fixed in S can be summarized by the extremal coef*ficient* (θ_{k}), defined as

$$P[Z(s_1) \le z, \dots, Z(s_K) \le z] = \exp\left(-\frac{\theta_K}{z}\right),$$

and ranging in [1, K] with the lower and upper limits corresponding to complete dependence and independence. Thereby θ_{κ} provides a measure of the degree of spatial



(9)

dependence between stations. It can be conveniently handled as the effective number of independent stations.

A special case of Eq. (9) is to consider the two-dimensional case of extremal coefficient defined by

5
$$P[Z(s_1) \le z, Z(s_2) \le z] = \exp\left\{-\frac{\theta_{s_1 s_2}}{z}\right\}.$$
 (10)

The bivariate extremal coefficient $\theta_{s_1s_2}$ provides sufficient information about the extremal dependence in many practical problems, although it does not characterize the full distribution.

The extremal coefficient functions for Smith's and Schlather's max-stable models are obtained straightforwardly by substiting z_1 and z_2 by z in the bivariate distributions (7) and (10). For Smith's model, this gives according to Eq. (6),

$$\theta_{s_1 s_2} = 2\Phi(a(||s_1 - s_2||)/2). \tag{11}$$

For Schlather's model, we obtain

$$\theta_{s_1 s_2} = 1 + \left\{ \frac{1 - \rho(\|s_1 - s_2\|)}{2} \right\}^{1/2}.$$
(12)

¹⁵ A standard tool, similar to the variogram, is the Madogram, (Neves and Gomes, 2011), defined as $v(s_1 - s_2) = E[||Z(s_1) - Z(s_2)||]$. As estimators for Madogram, there are pairwise estimator, the binned pairwise estimator (if isotropy) and, by relation between pairwise extremal coefficient and the Madogram, the plug-in estimator for pairwise extremal coefficient.

20 2.7 Estimation and selection procedure

Statistical models are usually estimated by maximizing the full likelihood for all observations. For spatial models, observations are usually available at a large number of



sites, e.g. 232 weather stations in the application of Sect. 3. However only the bivariate distributions are available both in Schathler's and Smith's model models. Consequently the full likelihood for all the available sites is unreachable in real applications. Therefore Padoan et al. (2010) proposed to replace the unreachable full likelihood by the pairwise likelihood of Varin (2008). This idea was latter used by Blanchet and Davison (2011), Davison et al. (2012) and Davison and Gholamrezaee (2012).

Let $\mathcal{K} = \{s_1, \dots, s_K\} \subset S$ denote the K observed sites used to fit the max-stable model in the region if interest S. Let $z_d^{(i)}$ denote the dth observed maximum for the *i*th station, transformed so that time-series $(z_1^{(i)}, \dots, z_D^{(i)})$ at each station have unit Fréchet distributions. In the application of Sect. 3 we will have D = 36 years and K = 232 weather stations. Let $f(z_i, z_j; \boldsymbol{\psi})$ be the bivariate density for the *i*th and *j*th locations, parametrized by some parameter vector $\boldsymbol{\psi}$. Then the pairwise log-likelihood function can be written

$$I_p(\boldsymbol{\psi}; \boldsymbol{z}) = \sum_{i < j} \sum_{d=1}^{D} \log f\left(\boldsymbol{z}_d^{(i)}, \boldsymbol{z}_d^{(j)}; \boldsymbol{\psi}\right).$$

as

¹⁵ Under suitable regularity conditions, the maximum pairwise maximum likelihood estimator $\hat{\psi}$ has a limiting normal distribution as $D \to \infty$, with mean ψ and covariance matrix of sandwich form estimable by

 $\hat{\boldsymbol{\psi}} \stackrel{:}{\sim} N(\boldsymbol{\psi}, H(\boldsymbol{\psi})^{-1} J(\boldsymbol{\psi}) H(\boldsymbol{\psi})^{-1}),$

where $H(\psi) = E$ is Fisher Information matrix and $J(\psi) = Var$. In practice, in order to estimate ψ and its uncertainty, we first maximize the pairwise likelihood Eq. (13), which gives an estimate $\hat{\psi}$. We used for this the sequential algorithm of Blanchet and Davison



(13)

(2011). Then estimates of H and J are obtained by:

$$\hat{H}(\hat{\boldsymbol{\psi}}) = -\sum_{i < j} \sum_{d=1}^{D} \frac{\partial^2 \log f\left(y_d^{(i)}, y_d^{(j)}; \boldsymbol{\psi}\right)}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}^{\mathsf{T}}};$$
$$\hat{J}(\hat{\boldsymbol{\psi}}) = \sum_{i < j} \sum_{d=1}^{D} \frac{\partial \log f\left(y_d^{(i)}, y_d^{(j)}; \boldsymbol{\psi}\right)}{\partial \boldsymbol{\psi}} \frac{\partial \log f\left(y_d^{(i)}, y_d^{(j)}; \boldsymbol{\psi}\right)}{\partial \boldsymbol{\psi}^{\mathsf{T}}},$$

and standard errors of $\boldsymbol{\psi}$ are obtained as the diagonal elements of the covariance matrix $\hat{H}^{-1}(\hat{\boldsymbol{\psi}})\hat{J}(\hat{\boldsymbol{\psi}})\hat{H}^{-1}(\hat{\boldsymbol{\psi}})$. Furthermore model selection can be performed using Takeuchi Information Criterion (TIC) extended in the composite likelihood setting, which is given by:

$$\mathsf{TIC} = -2\ell\left(\hat{\boldsymbol{\psi}}\right) + 2Tr\left\{\hat{J}(\hat{\boldsymbol{\psi}})\hat{H}^{-1}(\hat{\boldsymbol{\psi}})\right\}.$$

In accordance to AIC, the best model corresponds to that minimizing Eq. (14).

3 Application to rainfall data from the state of Paraná

The data used in this study are derived from historical data provided by the National Water Agency (referred to, in Portuguese, as ANA), an entity linked to the Ministry of Environment (referred to, in Portuguese, as MMA). The data were organized into two matrices, as follows: The first matrix of dimensions 136 × 232 is the maximum monthly

¹⁵ amount of rainfall in millimeters (mm) of the state of Paraná, each column corresponding to a weather station. The monthly maxima are extracted from daily measurements of precipitation that are read at about 7.30 a.m. from the months of January to April, which refers to the period of greatest concentration of grain harvest in the state. The original data set consists of 262 weather stations, although 30 of these stations were used for model validation.



(14)

The second matrix of dimensions 232 × 4 represents the coordinates of the weather stations and the covariates that may possibly influence the dependence of extreme rainfall data from the state of Paraná, each column corresponding to respectively the latitude, longitude, altitude and mean rainfall from the 232 stations; the covariate altitude was measured in kilometers, which is the same scale as the latitude and longitude.

In order to maintain the geographical structure with both economic and social similarities as regards the state of Paraná, three stations from each mesoregion were randomly selected for the models validation. Thus, these stations were spatially well distributed throughout the state of Paraná, as can be seen in Fig. 2. The distances

- between stations were measured in kilometers. The topography of the state of Paraná, as seen in Fig. 1, shows the importance of the climatological covariates altitude, since about 52 % of the territory of Paraná is above 600 m and 89 % above 300 m; only three percent are below 200 m. The morphologic picture is dominated by flat surfaces arranged in high altitude, forming plateaus forming the Serra do Mar and General (moun-
- tain ranges). Five units of land relief expand from east to west, in the following order: Coastal Lowlands, Serra do Mar, Crystalline Plateau, Paleozoic Plateau and Basalt Plateau. In this paper the Smith's and Schalther's models for maximum monthly rainfall data will be used, also used by Blanchet and Davison (2011). Four possible coordinates are considered for spatial modeling: longitude, latitude, altitude, and mean rainfall.
- It is important to note that the assumptions associated with the GEV distribution were checked punctually for each of the 262 weather stations. The Fig. 3 shows the fitted distributions for some regions. Most of the distributions are asymmetric to the right. In this context, the GEV distribution has good fits because it is a family of asymmetric distributions as described in Sect. 2.1. The Fig. 4 shows the diagnostics of the fitted distribution through the QQ-plot.

For Schlather's model, the following correlation functions were used: *Cauchy, Exponential Powered, Matérn, Circular, Cubic, Gneiting, Exponential, Spherical* and *Gaussian*, each of these correlation functions have one or two parameters, although correlation functions of the types *Spherical, Circular, Cubic* and *Gneiting* have upper limit. All



these correlation functions are such that $\rho(h) \rightarrow 1$ when $h \rightarrow 0^+$ and $\rho(h) \rightarrow 0$ when $h \rightarrow +\infty$. As mentioned at the end of Sect. 2.6, this limits the Schlather's model extremal coefficient to correspond to dependent data. However, Blanchet and Davison (2011) claim that such assumptions are justified in modeling spatial dependence of extreme snow depth in Switzerland.

Different combinations of the coordinates previously mentioned are considered. In all cases, longitude and latitude are used. The isotropic and anisotropic processes for Smith's and Schlather's models are also considered. In total, 60 types of models were considered. Other variables could be considered in this study, such as tempera-

- ture, humidity and maximum wind speed, which are also collected at weather stations. However, these relatively poor quality, with many missing data and, in many cases, presenting negative values, decided not to use them. Cooley et al. (2012), Fuentes et al. (2013), Blanchet and Davison (2011) and De Haan and Pereira (2006), are faced with the same problem.
- ¹⁵ Continuing the analysis, a comparison was made among the models using the TIC criterion (see Sect. 2.7). The values of the TIC profile are shown in Fig. 5. There are relatively small differences between them, although *Matérn, Powered Exponential* and *Cauchy* correlation functions have obtained the lowest values of TIC; these correlation functions are governed by two parameters. It can also be observed (Fig. 5) that, as the
- climatological covariates mean rainfall and altitude are added in different models, there is a considerable decrease in the values of the TIC for *Matérn*, *Powered Exponential* and *Cauchy* correlation functions.

These results are in agreement with Diggle et al. (1998), when referring to the flexibility of these functions which shape processes more or less distinguishable. It can be

observed in Fig. 5 that the *Gneiting, Cubic, Circular, Spherical* and *Gaussian* correlation function obtained a higher value of TIC. These correlation functions are governed by only one parameter and these results corroborate Webster and Oliver (2007), when they state that in practice most of these functions have some kind of restriction.



In general, Schlather's model performed better than Smith's model, whatever the chosen correlation function. In particular, the models that performed poorly correspond to models in Euclidean space. These results corroborate Blanchet and Davison (2011), when they defend the importance of working in a transformed space in order to allow

for anisotropy. Models where neither the altitude nor the mean rainfall are considered also obtained poor results. Therefore, in this study, it can be stated that these climate coordinates are very important in modeling extreme rainfall and thereby that models which consider them should be preferred.

According to the TIC value, the best fit is given by Schlather's model with *Powered Exponential* correlation function, and a 4-dimensional transformed space (S) with coordinates (longitude, latitude, altitude, mean rainfall). This means that a matrix is estimated in four dimensions and thus it has four parameters: the directional effect (ϑ), latitude (c_2), altitude (c_3) and mean rainfall (c_4) (Blanchet and Davison, 2011). The range (ϕ) and smoothness (ν) parameters of the Powered Exponential correlation function 15 are also estimated, for a total of six parameters, whose estimates and standard errors

¹⁵ are also estimated, for a total of six parameters, whose estimates and standard errors are shown in Table 1. The second best fit model is given by Schlather's model with *Powered Exponential* correlation function and a three-dimensional transformed space (S) with coordinates (longitude, latitude, mean rainfall). In this case, a three-dimensional matrix is estimated (see Blanchet and Davison, 2011). Thus, the second best fit is ²⁰ similar to the above but without the covariate altitude (c_3), for a total of five parameters.

²⁰ Similar to the above but without the covariate altitude (*C*₃), for a total of five parameters. The third best model is similar to the first but with another correlation function, that is, *Cauchy* correlation function. These two models act similarly, given that the covariates mean rainfall and altitude provide information about the local variability of rainfall. According to Fig. 5 it is possible to perceive the importance of using both covariates;
 ²⁵ however, there is a small increase in TIC when using only one of covariates. The worst

performances are assigned to Smith models, both in the isotropic and in the anisotropic cases, which, in turn, confirm the results obtained by Blanchet and Davison (2011).

It is no surprise that in Table 1, for the model with the lowest value of TIC, the mean rainfall is the most influential covariate in the climate distance and thus in the extremal



dependence function. This means that dependence between two stations at the same altitude is the same at a distance of 10 km from one another along the main direction of dependence, that is, an angle of $\vartheta = -0.20$ radians in the sense of an Argand diagram, at the same elevation but 2 km apart perpendicularly to the main direction of dependence, and at the same latitude and longitude but 322 m apart in elevation, are all equal.

It is observed that the model with the lowest TIC available in Table 1, takes into account the covariates altitude and mean rainfall. Since these covariates have an influence on the model, it is necessary to interpolate these two covariates in order to obtain the pairwise extremal coefficient everywhere in the region for the covariate elevation. The spatial kriging is performed with a spatial resolution of $90 \text{ m} \times 90 \text{ m}$, giving 5.935×8.305 gridded points on the state of Paraná. The same process was successfully used by Blanchet and Davison (2011), in the modeling of extreme snow in Switzerland.

- It can be observed in Fig. 6 the estimated maps of pairwise extremal dependence under the max-stable model with the correlation function exponential described in Table 1. Figure 6 clearly shows the effect of covariates elevation; one can also observe a weak extremal dependence. The results suggest that the elevation may influence the extreme value dependence. In other words, the flatter regions have a greater spatial
- ²⁰ dependence of rainfall than the regions with higher elevation. In this context, two points at same altitude will have exactly the same dependence. So if the difference in altitude is 0, then this term is zero. But what is true is that \hat{c}_3 is very large, so small difference in altitude will impact a lot on the extremal coefficients. This is mainly due to the fact that precipitation occurs when the air mass rises along with the set of mountains and
- ²⁵ finds lower temperatures, but from a certain altitude, when the air humidity decreases, precipitation also decreases.

Therefore, in the case of dependence of rainfall maxima, there are indications that the altitude has a negative effect on dependence. There are some practical implications related to this this fact. The first refers to risk management in agriculture. When



a high amount of rainfall occurs in the harvest seasons, there may be losses due to the inability to use the machinery to harvest. The results then may support the decision of deploying weather stations when the variable of interest refers to maximum rainfall. Regions with high spatial dependence would need much fewer stations than regions

⁵ of low dependence. The idea is to use a few stations that have high representation of maximum rainfall in certain regions, and vice versa. This reduces the cost of obtaining, installing and maintaining manual and automatic weather stations.

In Fig. 7 is shown the spatiotemporal annual extreme precipitation behavior. It can be observed that with the exception of 1970's, the annual precipitation extremes occur

- predominantly in the southwest region of the state of Paraná. This behavior can be associated with the distribution of terrain (see Fig. 2) which can promote the process of moist air vertical ascent, causing adiabatic cooling and, consequently, the formation of clouds and precipitation. Additionally, studies have identified the influence of the El Niño Southern Oscillation (ENSO) on the extreme precipitation events in southern areas of
- ¹⁵ Brazil (Grimm and Tedeschi, 2009; Pscheidt and Grimm, 2008). The South of Brazil is a region where ENSO episodes have strong and consistent impacts on rainfall extreme events. Pscheidt and Grimm (2008) showed that the variability of monthly rainfall and the number of severe events in the South of Brazil is mainly modulated by Sea Surface Temperature (SST) variability associated with ENSO. As the last decades presented
- strong noticed historical El Niños (1986–1987, 1991–1992, 1993, 1994 and 1997) they influenced the anual precipitation extreme events. As described above, the southern Brazil is impacted by El Niño/La Niña, with intense rain/drought during those events which results in positive or negative impact on agriculture (Grimm et al., 1998).

3.1 Validation of the models

For an initial verification of the quality of the selected model, we compare the prediction of extremal coefficient obtained by substituting the parameters involved in Eq. (10) for their respective estimates found in Table 1, with the estimated madograma based on Cooley et al. (2006). Since the extremal coefficient defined in Eq. (10) is based on



the distance between weather stations, we graphically evaluate extremal coefficient in terms of distance. Figure 8 presents these comparisons for the model that had the best performance as compared to TIC. Estimators Schlather and Tawn (2003) produce essentially the same image, but with a slightly higher variability. Model fitting performs

- well to a climate distance of 1.300, and then it underestimates. It was expected a madogram limit of about 1.8. However, it cannot be achieved with the Schlather's model; (see Sect. 2.6). This result confirms Davison et al. (2012) in a study of maximum rainfall in a historical series from 1962 to 2008 in the Swiss Plateau and Blanchet and Davison (2011), in modeling the spatial dependence of extreme snow depth in Switzerland.
- ¹⁰ Continuing the analysis, 30 weather stations are used to validate the model (approximately 11 % of the weather stations), as shown in Fig. 9. The model validity is verified and the empirical distribution of the maxima is compared to subsets of stations, i.e., $Z_{\mathcal{A}}^* = \max{\{Z^*(s_i), s_i \in \mathcal{A}\}}$, with maxima predicted by the selected model. The distribution of $Z_{\mathcal{A}}^*$ under the selected model is known analytically only when \mathcal{A} includes two stations. However, the samples of $Z_{\mathcal{A}}^*$ can be simulated for any given \mathcal{A} .
 - Since realizations z_A^* of Z_A^* are available for D = 36 years, one can compare the empirical quantiles of Z_A^* with simulated ones. More precisely, given a subset A, we simulate \mathcal{T} independent series $z_A^{*(\mathcal{T})}$ of length D, and thereby obtain \mathcal{T} replicates of the observed Fréchet series. Ordered values of observed z_A^* , can then be compared
- with ordered values of the $z_{A}^{*(\mathcal{I})}$. Pointwise and overall confidence bands can also be derived from these simulations (Davison, 1997). Validation test can also be found in Gagnon and Rousseau (2014), the authors validated a regional transformed model and to evaluate climate change impacts over a watershed in the subwatershed of the Yamaska River, located south of the St. Lawrence River, Québec, Canada.
- The Fig. 9 uses this approach in order to compare theoretical and empirical distributions for different groups of two, three or four weather stations taken from the 30 not used to fit the model, some groups being tightly clustered, and others being dispersed. The fit seems to be reasonably satisfactory in all cases. Even when there is a considerable distance between the stations, the data seem to be well modeled, despite the





mismatch between fitted and empirical pairwise extremal coefficients at such distances seen in Fig. 8.

4 Conclusions

20

25

The models and methods described here are considered as a step towards realistic and flexible modeling of maximum space, in agreement with the classical paradigm for max-stable processes and spatial extreme distributions. They are based on the models proposed by Smith (1990) and Schlather (2002).

Our application involved broader classes of functions, such as splines that, in principle, present no difficulty in their use, provided that the data are reliable. In many works of spatial extremes, simpler functions are commonly used, such as the application of a polynomial response surface for marginal parameters, which in turn does not take into account the max-stable process. In particular, the model can be represented in more flexible manner even in regions weakly dependent, involving a change in the Euclidean distance. Consequently, it enables the modeling of the directional effects resulting from environmental phenomena.

The impacts of climate changes on agriculture involve economic, environmental and social issues for the agricultural sector, related to identifying opportunities and managing risks. The study of spatial extremes provides the financial and the rural insurance segments with information to assist them in their decision-making processes in risk management, in a scenario of increased frequency and intensity of extreme events, as well as variation, considering the long-term meteorological phenomena.

Acknowledgements. The first author thanks the National School of Insurance – FUNENSEG and the Foundation for Agrarian Studies Luiz de Queiroz – FEALQ, for the financial support essential to this work. The authors thank the Study Group on Risk and Insurance – GESER, for the organization of data.

References

- Buishand, T. A., de Haan, L., and Zhou, C.: On spatial extremes: with application to a rainfall problem, Ann. Appl. Stat., 624–642, doi:10.1214/08-AOAS159, 2008. 12735
- Blanchet, J. and Davison, A. C.: Spatial modeling of extreme snow depth, Ann. Appl. Stat., 5,
- ⁵ 1699–1725, doi:10.1214/11-AOAS464, 2011. 12735, 12742, 12744, 12745, 12746, 12747, 12749
 - Blanchet, J. and Lehning, M.: Mapping snow depth return levels: smooth spatial modeling versus station interpolation, Hydrol. Earth Syst. Sci., 14, 2527–2544, doi:10.5194/hess-14-2527-2010, 2010. 12735
- Brooks, T. M., Mittermeier, R. A., da Fonseca, G. A. B., Gerlach, J., Hoffmann, M., Lamoreux, J. F., Mittermeier, C. G., Pilgrim, J. D., and Rodrigues, A. S. L.: Global biodiversity conservation priorities, Science, 313, 58–61, doi:10.1126/science.1127609, 2006 12734
 - Coles, S. An Introduction to Statistical Modeling of Extreme Values. Springer Series in Statistics, Springer, 2001, Germany, XIV, 209 p., 2001. 12736
- ¹⁵ Coles, S. G.: Regional modelling of extreme storms via max-stable processes, J. Roy. Stat. Soc. B, 55, 797–816, 1993. 12734
 - Coles, S. G. and Tawn, J. A.: Modelling extreme multivariate events, J. Roy. Stat. Soc. B, 53, 377–392, 1991. 12734
- Cooley, D., Naveau, P., and Poncet, P.: Variograms for spatial max-stable random fields, 373–390, doi:10.1007/0-387-36062-X_17, 2006. 12740, 12748
 - Cooley, D., Davis, R. A., and Naveau, P.: Approximating the conditional density given large observed values via a multivariate extremes framework, with application to environmental data, Ann. Appl. Stat., 6, 1406–1429, doi:10.1214/12-AOAS554, 2012. 12737, 12745
 Davison, A. and Gholamrezaee, M. M.: Geostatistics of extremes, P. Roy. Soc. A-Math. Phy.,
- ²⁵ 468, 581–608, doi:10.1098/rspa.2011.0412, 2012. 12735, 12742
- Davison, A. C.: Bootstrap Methods and Their Application, Vol. 1, Cambridge University Press, Cambridge, United Kingdom, 1997. 12749
 - Davison, A. C., Padoan, S., and Ribatet, M.: Statistical modeling of spatial extremes, Stat. Sci., 27, 161–186, doi:10.1214/11-STS376, 2012. 12735, 12742, 12749
- ³⁰ De Haan, L. and Pereira, T. T.: Spatial extremes: models for the stationary case, Ann. Stat., 34, 146–168, doi:10.1214/00905360500000886, 2006. 12745



- De Haan, L.: A spectral representation for max-stable processes, Ann. Probab., 12, 1194–1204, 1984. 12735, 12738
- Diggle, P. J., Tawn, J., and Moyeed, R.: Model-based geostatistics, J. Roy. Stat. Soc. C-App., 47, 299–350, doi:10.1111/1467-9876.00113, 1998. 12745
- ⁵ Fuentes, M., Henry, J., and Reich, B.: Nonparametric spatial models for extremes: application to extreme temperature data, Extremes, 16, 75–101, doi:10.1007/s10687-012-0154-1, 2013. 12745
 - Gagnon, P. and Rousseau, A. N.: Stochastic spatial disaggregation of extreme precipitation to validate a regional climate model and to evaluate climate change impacts over a small
- watershed, Hydrol. Earth Syst. Sci., 18, 1695–1704, doi:10.5194/hess-18-1695-2014, 2014. 12749
 - Gilleland, E., Brown, B. G., and Ammann, C. M.: Spatial extreme value analysis to project extremes of large-scale indicators for severe weather, Environmetrics, 24, 418–432, doi:10.1002/env.2234, 2013. 12735
- ¹⁵ Gräler, B., van den Berg, M. J., Vandenberghe, S., Petroselli, A., Grimaldi, S., De Baets, B., and Verhoest, N. E. C.: Multivariate return periods in hydrology: a critical and practical review focusing on synthetic design hydrograph estimation, Hydrol. Earth Syst. Sci., 17, 1281–1296, doi:10.5194/hess-17-1281-2013, 2013. 12735

Grimm, A. M. and Tedeschi, R. G.: ENSO and extreme rainfall events in South America, Am. Meteorol. Soc., 22, 1589–1609, doi:10.1175/2008JCLI2429.1, 2009. 12748

- Meteorol. Soc., 22, 1589–1609, doi:10.1175/2008JCLI2429.1, 2009. 12748
 Grimm, A. M., Ferraz, S. E. T., and Gomes, J.: Precipitation anomalies in Southern Brazil associated with El Niño and La Niña events, Am. Meteorol. Soc., 11, 2863–2880, doi:10.1175/1520-0442(1998)011<2863:PAISBA>2.0.CO;2, 1998. 12748
- Hashorva, E.: A novel class of bivariate max-stable distributions, Stat. Probabil. Lett., 76, 1047–1055, doi:10.1016/j.spl.2005.12.013, 2006. 12734
 - Hsing, T., Klüppelberg, C., and Kuhn, G.: Dependence estimation and visualization in multivariate extremes with applications to financial data, Extremes, 7, 99–121, 2004. 12735
 Kabluchko, Z., Schlather, M., and De Haan, L.: Stationary max-stable fields associated to negative definite functions, Ann. Probab., 37, 2042–2065, 2009. 12735
- ³⁰ Kojadinovic, I. and Yan, J. Goodness-of-fit testing based on a weighted bootstrap: A fast large-sample alternative to the parametric bootstrap, The Can. J. Stat., 40, 480–500, arXiv:1202.5682, 2012. 12734



- Kunihama, T., Omori, Y., and Zhang, Z.: Efficient estimation and particle filter for max-stable processes, J. Time Ser. Anal., 33, 61–80, doi:10.1111/j.1467-9892.2011.00740.x, 2012. 12734
- Marcelino, E. V., Nunes, L. H., and Kobiyama, M.: Banco de dados de desastres naturais: análise de dados globais e regionais, Caminhos Geogr., 6, 130–149, 2006 (in Portuguese) 12734

5

25

- Neves, M. and Gomes, D. P.: Geostatistics for spatial extremes. A case study of maximum annual rainfall in Portugal, Procedia Environ. Sci., 7, 246–251, doi:10.1016/j.proenv.2011.07.043, 2011. 12741
- Padoan, S. A., Ribatet, M., and Sisson, S. A.: Likelihood-based inference for max-stable pro-
- ¹⁰ cesses, J. Am. Stat. Assoc., 105, 263–277, doi:10.1198/jasa.2009.tm08577, 2010. 12735, 12739, 12742
 - Pscheidt, I. and Grimm, A. M.: Frequency of extreme rainfall events in Southern Brazil modulated by interannual and interdecadal variability, Royal Meteorol. Soc., 29, 1988–2011, doi:10.1002/joc.1799, 2008. 12748
- Ribatet, M., Singleton, R., and R. Core team: SpatialExtremes: modelling spatial extremes, R package version, 1–8, available at: http://cran.r-project.org/web/packages/SpatialExtremes/ SpatialExtremes.pdf (last access: 20 July 2014), 2011. 12734, 12735
 - Sang, H. and Gelfand, A. E.: Continuous spatial process models for spatial extreme values, J. Agr Biol. Envir. St., 15, 49–65, doi:10.1007/s13253-009-0010-1, 2010. 12734
- ²⁰ Sunyer, M. A., Sørup, H. J. D., Christensen, O. B., Madsen, H., Rosbjerg, D., Mikkelsen, P. S., and Arnbjerg-Nielsen, K.: On the importance of observational data properties when assessing regional climate model performance of extreme precipitation, Hydrol. Earth Syst. Sci., 17, 4323–4337, doi:10.5194/hess-17-4323-2013, 2013. 12735

Schlather, M.: Models for stationary max-stable random fields, Extremes, 5, 33–44, 2002. 12735, 12736

Schlather, M. and Tawn, J. A.: A dependence measure for multivariate and spatial extreme values: Properties and inference, Biometrika, 90, 139–156, doi:10.1093/biomet/90.1.139, 2003. 12734, 12740, 12749

Shang, H., Yan, J., Gebremichael, M., and Ayalew, S. M.: Trend analysis of extreme precipitation

³⁰ in the Northwestern Highlands of Ethiopia with a case study of Debre Markos, Hydrol. Earth Syst. Sci., 15, 1937–1944, doi:10.5194/hess-15-1937-2011, 2011. 12736



- Smith, R. L.: Max-stable processes and spatial extremes, Unpublished manuscript, Univer, available at: http://www.stat.unc.edu/postscript/rs/spatex.pdf (last access: 15 June 2014), 1990. 12735, 12736
- Smith, E. L. and Stephenson, A. G.: An extended Gaussian max-stable process model for
- spatial extremes, J. Stat. Plan. Infer., 139, 1266–1275, doi:10.1016/j.jspi.2008.08.003, 2009. 12734, 12735
 - Uboldi, F., Sulis, A. N., Lussana, C., Cislaghi, M., and Russo, M.: A spatial bootstrap technique for parameter estimation of rainfall annual maxima distribution, Hydrol. Earth Syst. Sci., 18, 981–995, doi:10.5194/hess-18-981-2014, 2014. 12735
- ¹⁰ Webster, R. and Oliver, M. A.: Geostatistics for Environmental Scientists, John Wiley & Sons, Chichester, England, 2007. 12745
 - Varin, C.: On composite marginal likelihoods, ASTA-Adv. Stat. Anal., 92, 1–28, doi:10.1007/s10182-008-0060-7, 2008. 12742



Table 1. Parameter estimates (with standard errors) and TIC of Schlather's isotropic model for the maximum monthly precipitation, considering the coordinates latitude, longitude covariate altitude in kilometers added to the rainfall covariate mean.

	Anisotropic Schlather Model		
	Ô	ĉ ₂	\hat{c}_3
Cauchy	-0.179(0.017)	6.103(0.714) $\hat{\phi}$	316.911(18.320) <i>v</i> ̂
	684.586(26.512)	0.495(3.397)	0.399(0.109) TIC
			30791.37
Powered Exponential	Ô	Ĉ ₂	ĉ ₃
	-0.200(0.016)	4.390(0.776) $\hat{\phi}$	31.009(12.968) <i>v</i>
	457.278(14.280)	168.276(3.758)	0.790(0.086) TIC
			30 783.81
Matérn	Ô	Ĉ ₂	ĉ ₃
	-0.181(0.018)	4.818(1.921) $\hat{\phi}$	20.394(4.883) <i>v</i>
	37.834(4.818)	601.627(3.448)	0.128 (0.078) TIC
			30 792.82

HESSD 11, 12731-12764, 2014 **Extreme precipitation** in the Paraná R. A. Olinda et al. **Title Page** Abstract Introduction Conclusions References Tables Figures Þ١ Back Close Full Screen / Esc Printer-friendly Version Interactive Discussion

Discussion Paper

Discussion Paper

Discussion Paper

Discussion Paper



Figure 1. Topographical of the state of Paraná for which daily precipitation data are available at meteorological stations. Location of the weather stations used in the modeling of spatial extremes (green) and in the model validation (brown).





Figures

Þ١

Close

Interactive Discussion

Figure 2. Location of the weather stations used in the modeling of spatial extremes (green) and in the model validation (brown).

Density Plot: Castro

Density Plot: Paranavaí





Density Plot: Cascavel





Discussion Paper **HESSD** 11, 12731-12764, 2014 **Extreme precipitation** in the Paraná R. A. Olinda et al. **Discussion** Paper Title Page Abstract Introduction References Tables **Figures Discussion Paper** 14 Back Close Full Screen / Esc **Discussion** Paper Printer-friendly Version Interactive Discussion

Figure 3. Fitting the generalized extreme value (GEV) distribution to the maximum monthly precipitation data for some counties in the state of Paraná.





Figure 4. QQ-plot of the Generalized Extreme Value distribution.



Number of Models

Figure 5. Scale values of TIC for all 6 × 9 fitted Schlather models. Ranges from 1 to 6 wherein: the nine Schlather isotropic models with covariates latitude and longitude (scale 1); Schlather nine isotropic models with covariates latitude, longitude and altitude (scale 2); Schlather nine anisotropic models with covariates latitude and longitude (scale 3); Schlather nine anisotropic models with covariates latitude, longitude and altitude (scale 4); Schlather nine anisotropic models with covariates latitude, longitude and precipitation (scale 5); Schlather nine anisotropic models with covariates latitude, longitude and precipitation (scale 5); Schlather nine anisotropic models with covariates latitude, longitude, altitude and rainfall (scale 6).





Figure 6. Pairwise extremal coefficient of Barbosa Ferraz, Paranavaí, Cascavel and Castro (white circles) predicted by max-stable process under the selected model.





Figure 7. Annual maximum interpolation of the last four decades of the state of Paraná.



Figure 8. Extremal coefficient for pairs of weather stations as a function of the distance between them, in Euclidean space (left plot) or transformed space (right). The red curve is the extremal coefficient curve adjusted to the max-stable model corresponding to Schlather's anisotropic model.





Figure 9. Comparison of theoretical and empirical quantiles for monthly maxima of groups of stations not used in the fitting. The stations used for each panel are shown in its map, and the envelopes are 95% pointwise and overall confidence bands obtained from $\mathcal{T} = 5.000$ simulations.

