Space-time extreme rainfall simulation under a geostatistical approach

Simulazione spazio-temporale di precipitazioni estreme tramite un approccio geo-statistico

Gianmarco Callegher, Carlo Gaetan, Noemie Le Carrer, Ilaria Prosdocimi

Abstract In this work we illustrate an approach to simulate extreme events with high resolution in space. First we model spatio-temporal variability in the marginal distributions with a flexible semi-parametric specification. Then the Gaussian copula is used to model locally in time and space the extremal dependence. The methods are showcased with an application to daily precipitations in the Venice lagoon catchment.

Abstract *In questo lavoro illustriamo un approccio per simulare eventi con alta risoluzione nello spazio. Per prima cosa modelliamo la variabilità spazio-temporale delle distribuzioni marginali con un approccio semi-parametrico. Quindi la copula gaussiana viene utilizzata per modellare localmente la dipendenza estrema nel tempo e nello spazio. Come esempio mostriamo un'applicazione alle precipitazioni nella laguna di Venezia.*

Key words: Copula, Peak-over-threshold, quantiles, rainfall, Venice lagoon

1 Introduction

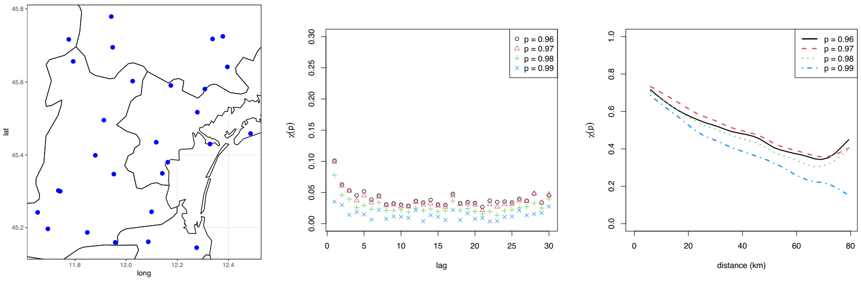
Observational studies have found that extreme precipitation can have heavy-tailed behaviour, i.e. the tail of the distribution of the magnitude of extreme events decays slower than an exponential. In the literature, we can find examples in which climate models of sufficiently high resolution may be capable of simulating precipitation extremes of comparable intensity to observed extremes. However it is not clear that they simulate daily intensities that are as heavy-tailed as observed, nor is it clear that they do so given the different scales in the observations at distinct points and simulated grid-box values. Moreover averaging in space and time smooths the tail

Gianmarco Callegher, Carlo Gaetan, Noemie Le Carrer, Ilaria Prosdocimi
Ca' Foscari University of Venice, DAIS (Italy)
e-mail: [giammarco.callegher,gaetan,noemie.lecarrer,ilaria.prosdocimi]@unive.it

behaviour recorded at weather stations, reducing the usefulness of simulated outputs for impact studies.

In this note we present a two-stage framework in which we couple models for extreme values and geostatistical space-time models. In the first stage, described in Section 2, we focus on the tail of the distribution of the rainfall amount by means of the the so-called peaks-over-threshold (POT) approach and we capture marginal spatio-temporal variation using regression splines (Youngman, 2019). Moreover, we use a Gaussian copula model to capture the short-range spatial and time dependence of the observed data (Section 3). This will allow to simulate extreme events which are consistent with the observed local variability in space and in time.

As a motivating example we consider daily rainfall records from long-term gauging stations in the Venice lagoon catchment from 1956 to 2018. The 28 locations of the stations are plotted in Figure 1-(a).



(a) (b) (c)

Fig. 1 (a) Locations of the stations in Venice lagoon catchment. Empirical estimates of $\chi(p)$ for pairs of observations at increasing temporal lags in (b) and spatial distances (in kilometers) in (c).

2 Extreme value model for a single site and time

We denote with $X(s,t)$ the daily rainfall accumulation at location s and time t. We consider a fixed high threshold $u(s,t)$, and we look at the distribution of the exceedances $(X(s,t)-u(s,t))$, conditional on $X(s,t)$ being larger than $u(s,t)$. Extreme value theory argues that it is possible to approximate this conditional distribution by a Generalized Pareto (GP) distribution. More precisely the distribution of threshold exceedances $Y(s,t) = (X(s,t)-u(s,t))$, given that $X(s,t) > u(s,t)$ has cumulative distribution function (cdf)

$$GPD(x;\xi,\sigma,u) = 1 - \left(1 + \xi \frac{x-u}{\sigma}\right)^{-1/\xi}, \quad (1)$$

where $\sigma > 0$ and $\xi \in \mathbf{R}$ are the scale and shape parameter of the distribution for $\{x > u : \left(1+\xi \frac{x-u}{\sigma}\right) > 0\}$. The threshold can either be chosen or estimated, but must be sufficiently high that the GPD assumption is valid.

For simulations to represent rainfall at given locations and time periods, the rate, i.e. the probability of exceeding the threshold must be taken into account: we denote this probability as $\zeta(s,t) = \Pr(Y(s,t) > u(s,t))$. For now, we assume that the threshold is known and we further simplify the modeling procedure by making the assumption that we are interested in a rate of exceedances that is constant at every site and at every time step, i.e. $\zeta(s,t) = \zeta$.

Then the unconditional distribution for $X(s,t)$ is defined as

$$F(x;\xi(s,t),\sigma(s,t),u(s,t)) = \begin{cases} 1-\zeta+\zeta\left(1+\xi(s,t)\frac{x-u(s,t)}{\sigma(s,t)}\right)^{-1/\xi(s,t)} & x > u(s,t), \\ 1-\zeta & x \leq u(s,t) \end{cases} \tag{2}$$

For the scale parameter $\sigma(s,t)$ we adopt an additive form of the log-link function

$$\log \sigma(s,t) = \beta^{\sigma} + f_1^{\sigma}(lon(s),lat(s)) + f_2^{\sigma}(t) \tag{3}$$

Here f_1^{σ} is thin plate regression spline where $lon(s)$ and $lat(s)$ represent longitude and latitude and f_2^{σ} is a cyclic cubic regression spline of period 365.25 to account for the leap years. Under this setup (3) can be written as

$$\log \sigma(s,t) = \beta^{\sigma} + \sum_{k=1}^{b_1} \beta_{1,k} B_{0,k}(lon(s),lat(s)) + \sum_{k=1}^{b_2} \beta_{2,k} B_{1,k}(t)$$

where $B_{k,i}(\cdot)$ are basis functions and $\beta_{i,k}$ the coefficient multiplying the spline basis. A similar specification is adopted for the shape parameter, namely

$$\xi(s,t) = \beta^{\xi} + f_1^{\xi}(lon(s),lat(s)) + f_2^{\xi}(t) \tag{4}$$

The model (2), (3) and (4) with parameters in the spline forms can be fitted using an approach that maximizes an independence likelihood (Chandler and Bate, 2007). More precisely, let $x(s_j,t)$ be realizations of $X(s_j,t)$ for s_j, $j = 1,\ldots,n$ sites and $t = \ldots,T$ times. By pretending that the observations are independent, the independence likelihood of the model (1) takes the form

$$L(\theta) = \prod_{j=1}^{n}\prod_{t=1}^{T} \frac{1}{\sigma(s_j,t)} \left(1+\xi(s_j,t)\frac{x(s_j,t)-u(s_j,t)}{\sigma(s_j,t)}\right)^{-1/\xi(s_j,t)-1} \tag{5}$$

where θ contains the unknown parameters in (3) and (4).

Maximization of (5) requires the knowledge of the space-time varying threshold $u(s)$. We follow Northrop and Jonathan (2011) and we estimate it by quantile regression (Koenker and Bassett, 1978) assuming that the threshold $u(s,t) = u_1(s) + u_2(t)$

can be splitted in two components: one ($u_1(s)$) which depends on the geographical coordinates and the other ($u_2(t)$) on the season. The effect of the season is modelled by a harmonic regression term. We have

$$\begin{aligned} u(s,t) &= u_1(s) + u_2(t) \\ &= \delta_0 + \sum_{i=1}^{d_0} \delta_{0,i} B_{1i}(lon(s), lat(s)) + \sum_{k=1}^{d_1} \delta_{1,k} \cos(\omega_k t) + \sum_{k=1}^{d_1} \delta_{2,k} \sin(\omega_k t) \end{aligned}$$

where $\omega_k = 2\pi k/365.25$.

3 Copula and extremal dependence

In the previous section we have described how to specify a model for the distribution of the extreme rainfall in one site s and at time t. However, this model can only reproduce the variability of the data at a low resolution (i.e at the point scale), and we need a model for efficient simulations of high-resolution extreme events. For this reason we couple the marginal model with a model for the local variation on space and time under a copula approach (Joe, 2014). It can be shown that every continuous multivariate distribution can be represented in terms of a copula which couples the univariate marginal distributions. More precisely, for a n-variate cdf $F(x_1,\dots,x_n) := \Pr(X_1 \le x_1,\dots,X_n \le x_n)$ with i-th univariate margin $F_i(x_i) := \Pr(X_i \le x_i)$, the copula associated with F is a cdf function $C_n : [0,1]^n \to [0,1]$ with $\mathscr{U}(0,1)$ margins that satisfies $F_n(x_1,\dots,x_n) = C_n(F_1(x_1),\dots,F_n(x_n))$. Note that the copula does not depend on the marginal distributions. For this reason, it is possible to characterize the extremal dependence through the copula function and distinguish between asymptotic independence and asymptotic dependence (Coles et al, 1999). Formally, let X_1 and X_2 be continuous random variables with distribution functions F_1 and F_2, respectively, and let

$$\chi(p) = \Pr(F_2(X_2) > p | F_1(X_1) > p) = \frac{1 - 2p - C_2(p,p)}{1-p}, \qquad 0 \le p < 1. \tag{6}$$

X_1 and X_2 are then said to be asymptotically independent if the limit $\chi := \lim_{p\to 1^-} \chi(p)$ is zero and asymptotically dependent if $\chi > 0$. Broadly speaking, under asymptotic independence the conditional probability of observing an exceedance in one variable given that the other variable has produced an exceedance converges to 0 as the threshold increases.

Copula based on Gaussian process can represent pairs of random variable which are asymptotically independent (Bortot et al, 2000). They play an important role since they can accommodate a variety of spatio-temporal dependence.

Assuming that the estimated marginal model (2,3,4) is the "true" generating model, we calculate uniformly distributed residuals on $[1-\zeta, 1]$:

$$R^*(s,t) = 1 - \zeta \left[1 + \xi(s,t)\frac{X(s,t) - u(s,t)}{\sigma(s,t)}\right]^{-1/\xi(s,t)}, \qquad \text{if } X(s,t) > u(s,t)$$

Figure 1 displays estimates of $\chi(p)$ for probabilities $p = 0.96, 0.97, 0.98, 0.99$ for pairs $R^*(s,t), R^*(s,t+h)$ with only temporal lag, and for pairs $R^*(s,t), R^*(s',t)$ with only spatial lag. The curves for spatial lags are the result of a smoothing procedure. These plots support the assumption of asymptotic independence at all positive distances and at all positive temporal lags.

Finally, the $R^*(s,t)$ random variable is transformed on a Gaussian scale by $R(s,t) = \Phi^{-1}(V(s,t))$ where $\Phi^{-1}(u)$ is the inverse of the cumulative distribution function of a standardized Gaussian random variable. We model $R(s,t)$ as a space-time zero mean Gaussian process with $\rho(s,s',t,t',\phi) = \text{cor}(R(s,t),R(s',t'))$, a correlation function that depends on an unknown parameter ϕ. Since for large data sets the evaluation of the censored likelihood becomes unfeasible the correlation can be estimated by maximizing the censored composite log-likelihood (see Bacro et al, 2020, for an example).

4 Results

Simulations are based on a stationary isotropic separable covariance function. The limited size of the area under analysis and daily temporal lags suggest that they will not have any impact on the recorded values. We consider an exponential-exponential separable correlation function, $\rho(s,s',t,t',\phi_1,\phi_2) = \exp(-\|s-s'\|/\phi_1) \times \exp(-|t-t'\|/\phi_2)$, $\phi_1, \phi_2 > 0$. The resulting estimates are $\hat{\phi}_1 = 127.41$ and $\hat{\phi}_2 = 1.04$, respectively. The low value of $\hat{\phi}_2$ indicate weak temporal dependence in the rainfall phenomena. As expected, an high spatial dependence is estimated. A fifty-years simulation is performed over a 700 evenly-spaced points grid. This results in a distance of $\simeq 2.43$ Km between two neighbouring locations. Figure 2 shows three different randomly selected events. In each row we report the spatial pattern of the day d, over an year, during which we simulated the maximum-precipitation in one site day, with the previous $(d-1)$ and following $d+1$ days. Extreme events can occur in different seasons, but their magnitude is strongly time-dependent by construction.

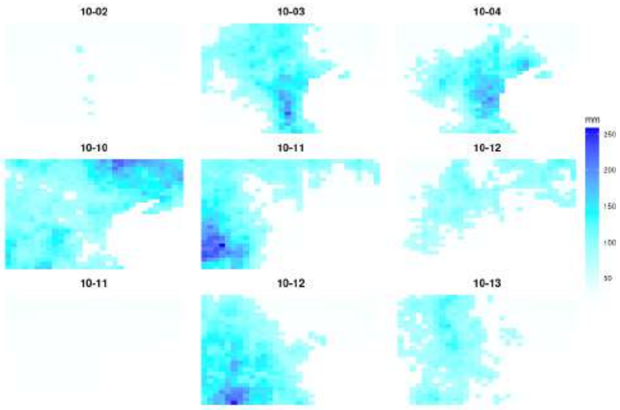


Fig. 2 Three examples of extreme precipitations simulations in three different period of the year. The central plot represents the day with highest precipitations

Acknowledgements
The research presented was funded by CORILA - research program “VENEZIA 2021”. The authors thank ARPAV for providing the rainfall data.

References

Bacro JN, Gaetan C, Opitz T, Toulemonde G (2020) Hierarchical space-time modeling of exceedances with an application to rainfall data. Journal of the American Statistical Association 115:555–569

Bortot P, Coles S, Tawn J (2000) The multivariate gaussian tail model: An application to oceanographic data. Journal of the Royal Statistical Society Series C 49:31–49

Chandler RE, Bate S (2007) Inference for clustered data using the independence loglikelihood. Biometrika 94:167–183

Coles S, Heffernan J, Tawn JA (1999) Dependence measures for extreme value analyses. Extremes 2:339–365

Joe H (2014) Dependence Modeling with Copulas. Chapman and Hall/CRC, Boca Raton, FL

Koenker R, Bassett G (1978) Regression quantiles. Econometrica 46:33 – 50

Northrop PJ, Jonathan P (2011) Threshold modelling of spatially dependent non-stationary extremes with application to hurricane-induced wave heights. Environmetrics 22:799–809

Youngman BD (2019) Generalized additive models for exceedances of high thresholds with an application to return level estimation for u.s. wind gusts. Journal of the American Statistical Association 114:1865–1879