Advances in Water Resources 84 (2015) 1–13

Contents lists available at ScienceDirect

Advances in Water Resources

journal homepage: www.elsevier.com/locate/advwatres

Modeling short duration extreme precipitation patterns using copula and generalized maximum pseudo-likelihood estimation with censoring

Zoubeida Kebaili Bargaoui [a,*], Andràs Bardossy [b]

[a] ENIT, Université de Tunis El Manar, Civil Engineering School, bp 37, 1002 Tunis, Tunisia
[b] IWS, Stuttgart University, Stuttgart, Germany

ARTICLE INFO

Article history:
Received 7 January 2015
Revised 27 June 2015
Accepted 8 July 2015
Available online 13 July 2015

Keywords:
Bivariate copula
Spatial copula
Dependence
Extremal coefficient
Rainfall

ABSTRACT

The paper aims to develop researches on the spatial variability of heavy rainfall events estimation using spatial copula analysis. To demonstrate the methodology, short time resolution rainfall time series from Stuttgart region are analyzed. They are constituted by rainfall observations on continuous 30 min time scale recorded over a network composed by 17 raingages for the period July 1989–July 2004. The analysis is performed aggregating the observations from 30 min up to 24 h. Two parametric bivariate extreme copula models, the Husler–Reiss model and the Gumbel model are investigated. Both involve a single parameter to be estimated. Thus, model fitting is operated for every pair of stations for a giving time resolution. A rainfall threshold value representing a fixed rainfall quantile is adopted for model inference. Generalized maximum pseudo-likelihood estimation is adopted with censoring by analogy with methods of univariate estimation combining historical and paleoflood information with systematic data. Only pairs of observations greater than the threshold are assumed as systematic data. Using the estimated copula parameter, a synthetic copula field is randomly generated and helps evaluating model adequacy which is achieved using Kolmogorov Smirnov distance test. In order to assess dependence or independence in the upper tail, the extremal coefficient which characterises the tail of the joint bivariate distribution is adopted. Hence, the extremal coefficient is reported as a function of the interdistance between stations. If it is less than 1.7, stations are interpreted as dependent in the extremes. The analysis of the fitted extremal coefficients with respect to stations inter distance highlights two regimes with different dependence structures: a short spatial extent regime linked to short duration intervals (from 30 min to 6 h) with an extent of about 8 km and a large spatial extent regime related to longer rainfall intervals (from 12 h to 24 h) with an extent of 34 to 38 km.

1. Introduction

The risk assessment of heavy rainfall for short time scales is quite important for flood risk and erosion risk evaluation, as well for the estimation of economic damages resulting from rainfall excess.

Because rainfall patterns are well organized structures in space and time, we need to adopt appropriate statistical methods that take into account the concomitant occurrence of rainfall in different geographic locations. Spatial variability assessment of environmental data is commonly based on the identification of the variogram function (Matheron [17]) which represents the mean quadratic deviation in a random field (here rainfall). However the assessment of rainfall spatial dependence attached to heavy rainfall events needs more specific approaches because variogram analysis is sensitive to outlying observations and assumes that the marginal distribution of the underlying random field is Normal. The hypothesis of normality is far of being realistic in the case of rainfall patterns for that the Gaussian type of dependence underestimates the dependence of extremes. Smith [21] proposed multivariate Student *t*-transformation as extreme value process. Yet, in the geostatistical approaches, the madogram which represents the mean absolute deviation in the random field has been suggested by Cooley [6] and Cooley et al. [7] instead of the variogram when studying the spatial dependence of extremes because it “conveniently links geostatistical ideas to measures of dependence for extremes”. In effect, it is related to the extremal coefficient, which characterizes dependence between extremes, as outlined by Gillou et al. [11]. It is worth noting that as a geostatistical tool, the madogram was proposed by Matheron [18] in linkage with the variogram of indicator variables.

Copula approaches have been recently investigated as alternative to analyze and interpolate environmental data spatial patterns. Copula is defined as a multivariate distribution function with marginal distributions uniforms on (0,1). Compared to more classical bivariate density approach, one great advantage in bivariate copula is that

* Corresponding author. Tel.: +216 21358495; fax: +216 71872729.
E-mail address: zoubeida.bargaoui@laposte.net (Z.K. Bargaoui).

http://dx.doi.org/10.1016/j.advwatres.2015.07.006

transformation of the random variable (here precipitation) does not change the copula. Also, as pointed out by De Michele and Salvadori [9] and Zhang and Singh [24] in their respective studies of rainfall intensity, depth and duration relationships, there is no need to assume that variables have the same marginal distributions with Copula approach while such an assumption is needed in order to simplify the computations in bivariate of trivariate analysis. . Because of the above cited merits of Copula approaches, bivariate copula was adopted by Cooley [6] in the domain of environmental data, to analyze air temperature data, as alternative to madogram analysis. Also, multivariate Chi-square copula was investigated by Bardossy [2] in analyzing groundwater quality data. Moreover, spatial interpolation based on Copula was proposed by Bardossy and Li [3] for groundwater quality parameters. Recently, Wasco et al. [23] used copula for spatial interpolation of rainfall data at the daily scale. What is particularly challenged is whether the high values of rainfall have a stronger dependence or the same dependence in comparison to the low values. In this perspective, Bardossy and Pegram [4] proposed a multidimensional copula model to simulate multi-site daily rainfall data and put into evidence the difference in dependence structure between zones of high and low rainfall. Wasco et al. [23] also adopted two-level approach: local and global estimations to improve rainfall estimation at non observed sites.

In Gaussian copula, the dependence structure (which is symmetrical) is carried out by the coefficient of correlation which acts as copula parameter.

Asymmetric bivariate copulas are more likely to represent spatial heavy rainfall structures or how does the correlation function varies with the interdistance. Thus, asymmetric spatial copula models have been constructed from the Gaussian copula using transformations such as the Chi-square distribution ([2]. However, Kazienka and Pilz [16] pointed out that the Chi-square copula is not appropriate to model extremes because it is asymptotically independent in the range of variation of the correlation coefficient. Thus, the paper aims to investigate other copulas belonging to extreme copula models in order to make an efficient estimation of the tail of the spatial extreme rainfall distribution and to investigate the spatial extent of extremes.

In the following, the paper presents in the next section the data available for the study as well as a first description of their spatial variability related to extreme rainfall values. Then, we present in Section 3 the methodology which mainly focuses on the method of generalized pseudo likelihood estimation to infer extreme copula parameters. Finally results are interpreted and discussed in Section 4 in the light of spatial extension of rainfall extreme patterns in linkage with the time resolution.

2. Data

Data are from the Stuttgart (Germany) region. The average annual temperature is 9.3 °C and the average mean annual precipitation is 674 mm. The climate is classified as warm and temperate (http://en.climate-data.org/). Fig. 1 shows the stations geographical locations. Available time series are 30-min time step observations of rainfall for the period July 1989- July 2004. Time series of 17 rainfall stations are analyzed. Series contain some gaps. After identifying gaps, the 30-min series were aggregated to the following time resolutions Δ = 30 mn, 60 mn, 3 h, 6 h, 8 h, 10 h, 12 h, 15 h, 18 h and 24 h. If for a given Δ, the interval contains a gap at 30min time resolution, this interval is removed from the time series corresponding to Δ. Thus, because of the presence of gaps, time series related to a given Δ have different sample sizes.

The probability of occurrence of zero is high. In the present case, it is around 0.78 for $\Delta = 30$ min, meaning that 78% of time intervals are not rainy. Conversely, a rainfall of 10 mm in 30 min has a probability 0.9999 of being exceeded.

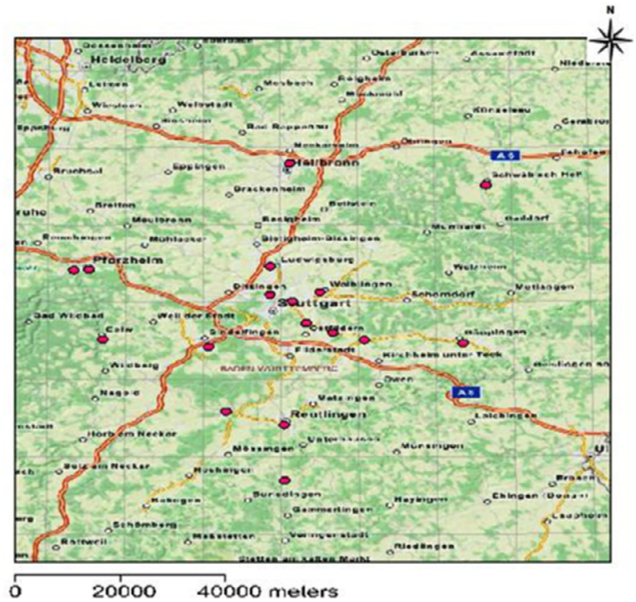

Fig. 1. The studied rainfall network; rainfall stations are in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

Many definitions are provided in the literature concerning the statement of heavy rainfall. It may correspond to 25 up to 60 mm/day. However, rainfall thresholds are generally associated with the time resolution Δ. Thus, it is found more appropriate to transform rainfall observations to unit less variables. To this end, in each station, observations are ranked and transformed using the sample cumulative distribution according to the Weibull formula $u_{\mathrm{i}} = \mathrm{rank}(x_{\mathrm{i},\Delta})/(N_0+1)$ where $x_{\mathrm{i},\Delta}$ $(i = 1,N_0)$ is rainfall observation at time $\mathrm{t_i}$ for timescale Δ, $\mathrm{rank}(x_{\mathrm{i},\Delta})$ is the rank of the observation $(x_{\mathrm{i},\Delta})$ when observations are sorted and $\mathrm{N_0}$ is the sample size. Thus, u is a uniform random variable.

Carlton et al. [5] adopted the 90th percentile per 24 h to define heavy rainfall events to investigate their effect on diarrhea in Ecuador. Here, independently from Δ, we considered the threshold $u_0 = 0.95$ for characterizing high rainfall values. We assumed that rare events correspond to $u_0 \geq 0.99$.

In the bivariate case, we consider two stations with two series x_{i} and y_{i} transformed to unit less uniform random variables u_{i} and v_{i}. The problem of interest is the concomitant occurrence of extremes in the two locations, for when extremes are reported in many locations, the flood risk is amplified.

Thus, it is important to estimate the joint probability of occurrence $\mathrm{Prob}(U = u,V = v)$, the copula itself which is $C(u,v) = \mathrm{Prob}(U \leq u,\mathrm{V} \leq v)$ as well as the conditional probability of occurrence $\mathrm{Prob}(U > u|V > v)$. To that purpose, it is helpful to adjust a parametric model of the Copula.

Stations geographical locations are provided in the coordinate's system projection of Gauss–Krueger. Euclidean distance is adopted. Fig. 2 reports rainfall scatter plot corresponding to Δ =30 mn for two stations (number 1 and 2) separated by an inter distance of about 20 km. The plot reports (u_i, v_i) such as $u_i \geq 0.95$ and or $v_i \geq 0.95$; each square corresponds to a pair (u_i, v_i). Despite the importance of inter-distance, the plot highlights the simultaneous occurrence of extremes (many realizations are reported for u_i and $v_i \geq 0.95$). However, it may be seen from the plot that it often happens that one station records an extreme while the other station does not (realizations where we have u_i or $v_i \geq 0.95$ but not for both).

Now, consider four subsets:

- S0: $u_i < u_0$ and $v_i < v_0$;
- S1: $u_i > u_0$ and $\mathrm{v}_i < v_0$;