Journal of Hydrology 533 (2016) 523–532

Contents lists available at ScienceDirect

Journal of Hydrology

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

On the prediction of extreme flood quantiles at ungauged locations with spatial copula

Martin Durocher [a,*], Fateh Chebana [a,*], Taha B.M.J. Ouarda [b,a]

[a] Institut National de Recherche Scientifique (INRS-ETE), University of Québec, 490 de la Couronne, Québec G1K 9A9, Canada

[b] Institute Center for Water Advanced Technology and Environmental Research (iWater), Masdar Institute of Science and Technology, P.O. Box 54224, Abu Dhabi, United Arab Emirates

ARTICLE INFO

Article history:
Received 21 July 2015
Received in revised form 11 December 2015
Accepted 15 December 2015
Available online 21 December 2015
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Niko Verhoest, Associate Editor

Keywords:
Flood analysis
Physiographical-based kriging
Regional frequency analysis
Spatial copula
Ungauged basin
Interpolation

SUMMARY

The present study investigates the use of the spatial copula approach for predicting flood quantiles at ungauged basins. Spatial copulas are the formalization of traditional geostatistics by copulas. In regional flood frequency analysis (RFFA), the regression of flood quantiles is often carried out at the logarithmic scale. Consequently, traditional interpolation methods introduce a bias and provide suboptimal predictions. In this study, the copula framework is examined for offering proper corrections in this framework. Moreover, copula techniques separate the regional distribution of flood quantiles from spatial dependence. This provides a full probabilistic model that represents a more flexible framework where proper combinations of regional distribution and dependence can be adapted to various situations that are encountered in RFFA. The adequacy of the investigated methodology is evaluated on a real world case study involving hydrometric stations from southern Quebec, Canada. Results show that the spatial copula framework is able to deal with the problem of bias, is robust to the presence of problematic stations and may improve the quality of quantile predictions while reducing the level of complexity of the models used in RFFA.

1. Introduction

Knowing the likelihood of threatening events is of primary interest for water resource managers. In practice, a large number of years of record is required in order to collect adequate information for carrying out a reliable at-site frequency analysis. Regional flood frequency analysis relies on at-site quantile estimates to transfer hydrological information to an ungauged site. There are two main approaches in regional flood frequency analysis (RFFA) for predicting flood quantiles at ungauged locations: The quantile based regression which predicts directly at-site flood quantiles corresponding to a given return period (Chebana and Ouarda, 2008; Ouarda et al., 2001; Pandey and Nguyen, 1999; Tasker et al., 1996) and the parameter based regression that predicts the at-site distribution before calculating the desired flood quantiles (Chebana and Ouarda, 2009; Eng et al., 2007; Haddad and Rahman, 2012; Hosking and Wallis, 1997). Notice that the latter approach includes as a special case the popular index flood model (Hosking and Wallis, 1997). In the following, the quantile based regression approach is investigated.

Spatial modeling was shown to be useful in RFFA in a large number of publications (Archfield et al., 2013; Castiglioni et al., 2009; Chokmani and Ouarda, 2004; Hundecha et al., 2008; Nezhad et al., 2010; Skøien et al., 2006). The flood generating processes of a given basin may depend on several factors including the physiographical characteristics of the basin and the meteorological conditions in the region of study. Two rivers representing different basin characteristics can produce river discharges possessing very different characteristics even if they are very close geographically. Direct interpolation from the geographical coordinates is therefore inappropriate. This motivated the development of suitable spaces, called physiographical spaces, where hydrological variables can be treated as spatial data. On the other hand, it was found that hybrid techniques that combine both geographical distance and similarity between the basin characteristics can improve the quality of regional models (Eng et al., 2007; Haddad and Rahman, 2012). Geographical distance is often implicitly included in RFFA models through the use of the latitude and longitude as physiographical characteristics. Therefore, the term physiographical space is used in the present work to designate general spaces, in which spatial methods can be applied.

Kriging techniques are developed for predicting statistical variables, such as precipitation, at unknown locations as linear combinations of observed values. One property that may explain the

* Corresponding authors. Tel.: +1 418 654 4677/2542.
E-mail addresses: martin.durocher@ete.inrs.ca (M. Durocher), fateh.chebana@ete.inrs.ca (F. Chebana).

http://dx.doi.org/10.1016/j.jhydrol.2015.12.029
0022-1694/

popularity of kriging, is that it minimizes the predictive mean square errors (Schabenberger and Gotway, 2004). However, this property assumes that spatial data can be characterized by a multivariate normal distribution (MVN), which is usually not valid for flood quantiles. Conversely, the logarithm transformation is commonly used for describing the exponential relationship between flood quantiles and basin characteristics (Pandey and Nguyen, 1999; Eaton et al., 2002). At the original scale, this strategy creates bias and produces suboptimal predictions because of the nonlinear nature of the data (Schabenberger and Gotway, 2004).

Copulas have been introduced in the field of hydrology for providing a more realistic evaluation of risks associated to flood events by characterizing multiple aspects of the hydrographs (e.g. Chebana and Ouarda, 2009, 2007; Requena et al., 2013; Salvadori and De Michele, 2004; Shiau et al., 2006). In this framework, copulas offer a simple and efficient way to account for the dependence between different aspects of the hydrograph, such as peak, duration and volume. The usefulness of the copula framework emanates from the simplicity by which the model separates the marginal distribution from the dependence structure. This decomposition of the model allows combinations of proper components for fitting data that would otherwise be difficult to describe.

The motivation for using copulas to describe the dependence structure of spatial data was discussed by Bárdossy (2006). The objective was to develop a more general approach for characterizing spatial patterns that traditional methods fail to model. Building a copula that allows easy formulation of the spatial structure according to the separating distance can be a difficult task and not all copulas are suited for spatial analysis. The Gaussian copula is a particular case that ensures the continuity with the existing methodologies, while allowing marginals to belong to various classes of distributions. For predictive purposes, Bárdossy and Li (2008) showed that copulas provide a practical way of writing a predictive distribution at a new location. Adopting a copula framework for spatial modeling, therefore called spatial copula, is a way of generalizing traditional kriging techniques such as indicator kriging, trans-Gaussian kriging and normal rank kriging (Kazianka and Pilz, 2010), as well as providing additional flexibility.

This study proposes the recently developed spatial copula approach as an alternative geostatistical method for predicting flood quantiles at ungauged locations. One of the interests of considering spatial copulas is to provide a correction for the bias and to account properly for the nonnormal distribution of the at-site flood quantiles in order to define optimal predictors. To this end, spatial copulas provide a simple expression of the mean of the predictive distribution, which in terms of square errors corresponds to the best predictors at the original scale (Bárdossy and Li, 2008). Accordingly, the present work aims at evaluating the gain of accuracy made by adopting spatial copulas in RFFA. A second objective is to examine how the more general framework of spatial copulas may be used to provide predictions that are more robust and to adapt to problematic situations.

The present study is organized as follows. Section 2 provides a review of the methodology of spatial copulas and proposes some specific adaptations to RFFA. Section 3 illustrates a real world application to hydrometric stations in the southern region of the province of Quebec, Canada. Finally, discussions are provided in Section 4 and concluding remarks are drawn in Section 5.

2. Methodology

We only provide a short introduction on copula theory in this section for assuring the completeness of the document. For a more detailed treatment on the topic of copulas, the reader is referred to Nelsen (2006) and Salvadori et al., 2007. The present section also follows closely the general methodology developed in Bárdossy (2006) and Bárdossy and Li (2008), where further discussion on the technical aspect of spatial copulas can be found.

The methodology section is divided in several subsections that develop the steps to follow for performing RFFA in a spatial copula framework. Section 2.1 reviews the basics of the copula theory. These are necessary to develop the methodology of the present work. Section 2.2 discusses the particularity of the copulas adapted to spatial analysis and provides an adaptation of the general structure to RFFA. Section 2.3 presents common methods used to build the physiographical space and shows how to integrate this information in a regional model. Sections 2.4 and 2.5 describe respectively the dependence and the marginal components that characterize the regional model. Finally, Section 2.6 proposes estimation procedures and Section 2.7 explains how to predict the flood quantiles of the spatial copula model.

2.1. Basic concepts

Copulas aim at providing a framework where dependence is treated separately from marginal distributions. This section reviews the necessary elements to understand the methodology described in the following sections. For more information on copulas, the reader is referred to the general introduction provided by Nelsen (2006).

A d-dimensional copula

$$C : [0,1]^d \rightarrow [0,1] \quad (1)$$

is a multivariate distribution defined on the unit hypercube, which respects some regularity conditions (Nelsen, 2006). The main result concerning copulas is known as the Sklar's theorem that states that every multivariate distribution G can be expressed in function of a copula C and marginals $\{F_i\}_{i=1}^d$:

$$G(\mathbf{x}) = C[F_1(x_1), \ldots, F_d(x_d)] \quad (2)$$

where $\mathbf{x}' = (x_1, \ldots, x_n) \in \mathbb{R}^d$. Moreover, if each marginal F_i is continuous, then C is unique. Conversely, copulas can be constructed from existing distributions:

$$C(\mathbf{u}) = G\left[F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)\right]. \quad (3)$$

where $\mathbf{u}' = (u_1, \ldots, u_d) \in [0,1]^d$. The values u_i are pseudo-observations and correspond to uniform observations on the unit hypercube. An important copula built from Eq. (3) is the Gaussian copula for which $G = \Phi_\Sigma$ is a MVN with correlation matrix Σ and $F_i = \phi$ are standard univariate normal distributions. Both Φ_Σ and ϕ have a zero mean and a unit variance.

As suggested by Eq. (2), the modeling strategy consists of finding the best combination of copula C and marginals F_i, even if the resulting distribution G is not a member of a well-known distribution family. Another appealing property of copulas is their invariance to strictly monotonic transformations of the variables. This property removes the subjectivity of choosing the best transformation, which otherwise affects the shape of the dependence (Nelsen, 2006).

2.2. Spatial copulas

Spatial data may be seen as a single realization of a random field $Z(\mathbf{s}) \in \mathbb{R}$ defined as a set of random variables indexed by a location $\mathbf{s} \in \mathbb{R}^2$. Accordingly, let $\mathbf{z} = (z_1, \ldots, z_n)$ be a set of observations of the random field Z where every observation is simultaneously collected at respective location $\mathbf{s}_i$. For RFFA, $\mathbf{z}$ represents at-site flood quantiles estimated from gauged stations located at $\mathbf{s}_i$ in the