Journal of Hydrology 540 (2016) 469–487

Contents lists available at ScienceDirect

Journal of Hydrology

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

Probabilistic modeling of flood characterizations with parametric and minimum information pair-copula model

Alireza Daneshkhah [a], Renji Remesan [b,d], Omid Chatrabgoun [c], Ian P. Holman [b,*]

[a] Warwick Centre for Predictive Modelling, School of Engineering, The University of Warwick, CV4 7AL, UK
[b] Cranfield Water Science Institute, Cranfield University, Cranfield MK43 0AL, UK
[c] Department of Statistics, Faculty of Mathematical Sciences and Statistics, Malayer University, Malayer, Iran
[d] Centre for Ecology and Hydrology, UK

ARTICLE INFO

Article history:
Received 19 December 2015
Received in revised form 26 April 2016
Accepted 20 June 2016
Available online 21 June 2016
This manuscript was handled by A. Bardossy, Editor-in-Chief, with the assistance of Fateh Chebana, Associate Editor

Keywords:
Flood frequency analysis
Flood hazard characterization
Return period
D-vine model
Minimum information pair-copula model
Himalaya (India)

ABSTRACT

This paper highlights the usefulness of the minimum information and parametric pair-copula construction (PCC) to model the joint distribution of flood event properties. Both of these models outperform other standard multivariate copula in modeling multivariate flood data that exhibiting complex patterns of dependence, particularly in the tails. In particular, the minimum information pair-copula model shows greater flexibility and produces better approximation of the joint probability density and corresponding measures have capability for effective hazard assessments. The study demonstrates that any multivariate density can be approximated to any degree of desired precision using minimum information pair-copula model and can be practically used for probabilistic flood hazard assessment.

1. Introduction

Operational planning and design of flood defence systems, irrigation water management systems and hydroelectric schemes requires accurate estimation of flood hazard and/or specified exceedance probabilities of river flow. Flood frequency analysis (FFA) is traditionally used to assess flood hazard with an assumption that annual maximum floods are a stationary, independent, identically distributed random process (Kidson and Richards, 2005). Conventionally, FFA is performed using either 'Block (annual) maxima' or 'peaks over threshold (POT)' methods on partial series of data (Hosking et al., 1985). Although, the univariate FFA is widely used in hydrology, many studies have highlighted its unreliability and suggested that univariate frequency analysis methods cannot sufficiently characterize inflow hydrographs or reduce uncertainty in flood analysis (Cunnane, 1988; Bobee and Rasmussen, 1994). Indeed, most hydrologic events are multivariate in nature and defined by a group of correlated random variables (e.g. flood peak, volume, and duration). Therefore, multivariate FFA would be more suitable to describe the uncertainties associated with these events.

By recognizing the limitations of univariate FFA, multivariate flood frequency analysis methods were developed. Many early multivariate studies focused on bivariate normal distribution to perform flood analysis with later researchers considering multivariate Gaussian (Krstanovic and Singh, 1987), gamma (Yue et al., 2001; Nadarajah and Gupta, 2006), exponential (Choulakian et al., 1990), Gumbel (Bacchi et al., 1994) and other distributions. Durrans et al. (2003) applied Pearson Type III distribution to perform joint frequency analysis. Yue and Wang (2004) developed Gumbel mixed and Gumbel logistic models; and compared their performances in flood analysis. However, distribution-based traditional univariate and multivariate analysis methods have mathematical weaknesses that limit their potential for practical applications. These flaws include that (a) the mathematical formulation is complicated when the number of variables are high (b) it is not possible to distinguish marginal and joint behavior of studied variables, (c) marginal distributions are of same type, or normal, or independent and (d) joint distributions hold validity in limited space (Song and Singh, 2010).

* Corresponding author.
E-mail address: I.holman@cranfield.ac.uk (I.P. Holman).

http://dx.doi.org/10.1016/j.jhydrol.2016.06.044
0022-1694/

Recently, the application of copulas in hydrology, as well as in other earth and environmental sciences, has received increasing attention. Copulas are efficient mathematical tools which are capable of combining several univariate marginal cumulative distribution functions into their joint cumulative distribution function (Sklar, 1959). The copula application in hydrology largely began after De Michele and Salvadori (2003) highlighted the suitability of the Frank copula for the joint distribution of negatively associated storm intensity and storm duration data, whilst Grimaldi and Serinaldi (2006a,b) applied several trivariate copulas for determining joint and conditional distributions among design hyetograph variables. Recent works on analysis of multivariate hydrological extreme events (Salvadori and De Michele, 2006, 2007, 2010) have popularized copulas as a tool for extreme value applications in rainfall (Évin and Favre, 2008; Wang et al., 2010; Zhang et al., 2012), floods (Zhang and Singh, 2007; Chowdhary et al., 2011), and droughts (Shiau, 2006; Song and Singh, 2010; Zhang et al., 2012; Ma et al., 2013). A brief review of the application of copula in various engineering and science fields can found in Genest and Favre (2007). They have also identified plausible Copula candidates for flood peak flow and volume data in FFA. Dupuis (2007) used 5 copulas (Normal, Student-*t*, Frank, Clayton, Gumbel, and associated Clayton) and warned about ignoring the tail dependence characteristics of flood data. Their analysis showed that the Frank copula performed relatively well in comparison to other approaches. Karmakar and Simonovic (2009) identified that the generalized hyperbolic copula is better at obtaining pair-wise joint distributions among flood peak flow, volume and duration. Leonard et al. (2008) used copula for bivariate analysis of rainfall and stream flow extremes accounting for seasonal and climatic partitions. Huard et al. (2006) and Silva and Lopes (2008) used Bayesian based copula selection method for estimating marginal and dependence parameters.

The set of higher dimensional copulas proposed in the literature is limited and is not rich enough to model all possible mutual dependencies among all variables (see Kurowicka and Cooke, 2006 for details). In addition, Aas et al. (2009) show that the multivariate copulas (in particular, multivariate *t*-copula) cannot efficiently be used to model multivariate data exhibiting complex patterns of dependence in the tails (which are common in analyzing the extreme events). These limits of the multivariate copula motivate Joe (1997) and Bedford and Cooke (2001, 2002), to propose a far efficient new way of constructing complex multivariate highly dependent models called vine or pair-copula (Aas et al., 2009). The principle behind this method is to model dependency using simple local building blocks based on conditional independence, known as the pair-copulae. The modeling scheme is then based on a decomposition of a multivariate density into a cascade of pair copulae, applied on the original variables and on their conditional and unconditional distribution functions. There is a growing literature of using the pair-copula models in the different real world applications including finance, economic and insurance studies (Aas et al., 2009; Czado and Min, 2010; Min and Czado, 2010; Bauer et al., 2012; Dissmann et al., 2013; Brechmann et al., 2014), risk management (Brechmann and Czado, 2013; Brechmann et al., 2014), energy (Czado et al., 2011), hydrological drought frequency analysis (Song and Singh, 2010). In addition to the above references which give an idea of recent advancements happening on pair-copula applications in the different fields. Recently, Gyasi-Agyei and Melching (2012) have used PCC to model the dependence structure of storm event properties using hourly rainfall data from Cook County, Illinois, USA. Song and Kang (2011) demonstrated pair-copula based trivariate discharge modeling considering variables like flood duration, severity, and severity peak. Vernieuwe et al. (2015) constructed a continuous rainfall model based on vine copulas and they compared the vine model with ensemble synthetic rainfall series. In a similar study, Xiong et al. (2014) have developed an annual rainfall-runoff model using the canonical vine copula derivation approach and employed in 40 watersheds in two large basins in China.

The multivariate copula models have also been used in different applications in the domain of spatial statistics. Bárdossy (2006) was one of the first who applied copulas in a geostatistical context. Gräler and Pebesma (2011) propose a more efficient approach for modeling spatial data (including extremes) using the vine copula model. One of the advantages of their approach was its flexibility in choosing appropriate parametric copula families through bivariate spatial copulas. Gräler (2014) extends this methodology further by adding several spatial trees at the foundation of the selected vine. These additional spatial trees add valuable information on the dependence of the higher order neighbors leading to an improved model of the spatial data. The predictive accuracy of the spatial vine copula outperforms other spatial multivariate copulas, including spatial Gaussian copula which used to be a very common method (as suggested by Bárdossy, 2006).

In a more relevant study, Gräler et al. (2013) use the vine copula model to construct a joint probability distribution for the flood variables, including peak discharge, duration, and volume. However, their main purpose of modeling the dependencies between the flood variables using the vine copula model and other multivariate copula models was to estimate design events for a given return period and to discuss their differences in a practical application. They concluded that the vine copula approach is the way to go for constructing flexible multivariate distribution functions for the same reasons mentioned above and discussed in further details in the next section.

It should be noticed that the use of a copula to model dependency is simply a translation of one difficult problem into another. By using (parametric) copula, the difficulty of specifying the full joint distribution will be reduced to the difficulty of specifying the copula. The advantage is the technical one that copulas are normalized to have support on the unit square and uniform marginals. As many authors restrict the copulas to a particular parametric class (Gaussian, multivariate t, etc.) the potential flexibility of the copula approach is not realized in practice. Bedford et al. (2015) proposed a so-called minimum informative pair-copula using the vine structure to approximate any given multivariate copula to any required degree of approximation, and to show how this can be operationalized for use in practice. The only technical assumptions required are that the multivariate copula density under study is continuous and is non-zero. This approach, by contrast to the parametric methods mentioned above, allows a lot of flexibility in copula specification. This new approach involves the use of minimum information copulas that can be specified to any required degree of precision based on the data available and are then stacked together to produce the multivariate copula and density function.

Based on the above discussion, we extend the parametric vine copula model (Gräler et al., 2013) in modeling flood characterizations with the minimum information pair-copula model. This model shows greater flexibility and produces better approximation of the joint probability density and corresponding measures have better capability for effective hazard assessments. We also present an approximation method at which any multivariate density can be approximated to any degree of desired precision using minimum information pair-copula model and practically be applied for assessing probabilistic flood hazard. We finally illustrate the methods described above by modeling the flood event properties of the Himalayan River Beas. Himalayan rivers in north India are highly influenced by both the monsoon and intra-annual release of stored water in the snow cover and glacier ice of the Himalayas and its nearby foothills. The response of Himalayan rivers to precipitation and temperature is highly variable as it depends on the extent of snow cover and volume of snowpack in their catchment,