Journal of Hydrology 525 (2015) 658–675

Contents lists available at ScienceDirect

Journal of Hydrology

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

A framework for multivariate data-based at-site flood frequency analysis: Essentiality of the conjugal application of parametric and nonparametric approaches

H. Vittal [a], Jitendra Singh [a], Pankaj Kumar [b], Subhankar Karmakar [a,c,d,*]

[a] Centre for Environmental Science and Engineering, Indian Institute of Technology Bombay, Mumbai 400 076, India
[b] Applied Statistics and Informatics, Indian Institute of Technology Bombay, Mumbai 400 076, India
[c] Interdisciplinary Programme in Climate Studies, Indian Institute of Technology Bombay, Mumbai 400 076, India
[d] Centre for Urban Science and Engineering, Indian Institute of Technology Bombay, Mumbai 400 076, India

ARTICLE INFO

Article history:
Received 11 January 2015
Received in revised form 4 April 2015
Accepted 14 April 2015
Available online 20 April 2015
This manuscript was handled by Andras Bardossy, Editor-in-Chief, with the assistance of Attilio Castellarin, Associate Editor

Keywords:
Allegheny River
Copula
Flood frequency analysis
Joint return period
Multivariate analysis
Nonparametric

SUMMARY

In watershed management, flood frequency analysis (FFA) is performed to quantify the risk of flooding at different spatial locations and also to provide guidelines for determining the design periods of flood control structures. The traditional FFA was extensively performed by considering univariate scenario for both at-site and regional estimation of return periods. However, due to inherent mutual dependence of the flood variables or characteristics [i.e., peak flow (P), flood volume (V) and flood duration (D), which are random in nature], analysis has been further extended to multivariate scenario, with some restrictive assumptions. To overcome the assumption of same family of marginal density function for all flood variables, the concept of copula has been introduced. Although, the advancement from univariate to multivariate analyses drew formidable attention to the FFA research community, the basic limitation was that the analyses were performed with the implementation of only parametric family of distributions. The aim of the current study is to emphasize the importance of nonparametric approaches in the field of multivariate FFA; however, the nonparametric distribution may not always be a good-fit and capable of replacing well-implemented multivariate parametric and multivariate copula-based applications. Nevertheless, the potential of obtaining best-fit using nonparametric distributions might be improved because such distributions reproduce the sample's characteristics, resulting in more accurate estimations of the multivariate return period. Hence, the current study shows the importance of conjugating multivariate nonparametric approach with multivariate parametric and copula-based approaches, thereby results in a comprehensive framework for complete at-site FFA. Although the proposed framework is designed for at-site FFA, this approach can also be applied to regional FFA because regional estimations ideally include at-site estimations. The framework is based on the following steps: (i) comprehensive trend analysis to assess nonstationarity in the observed data; (ii) selection of the best-fit univariate marginal distribution with a comprehensive set of parametric and nonparametric distributions for the flood variables; (iii) multivariate frequency analyses with parametric, copula-based and nonparametric approaches; and (iv) estimation of joint and various conditional return periods. The proposed framework for frequency analysis is demonstrated using 110 years of observed data from Allegheny River at Salamanca, New York, USA. The results show that for both univariate and multivariate cases, the nonparametric Gaussian kernel provides the best estimate. Further, we perform FFA for twenty major rivers over continental USA, which shows for seven rivers, all the flood variables followed nonparametric Gaussian kernel; whereas for other rivers, parametric distributions provide the best-fit either for one or two flood variables. Thus the summary of results shows that the nonparametric method cannot substitute the parametric and copula-based approaches, but should be considered during any at-site FFA to provide the broadest choices for best estimation of the flood return periods.

1. Introduction

Flood frequency analysis (FFA) defines the severity of a flood event by summarizing the flood variables/characteristics [i.e., peak (P), volume (V) and duration (D)] and by estimating their mutual

* Corresponding author at: Centre for Environmental Science and Engineering, Indian Institute of Technology Bombay, Mumbai 400 076, India. Tel.: +91 22 2576 7857.
E-mail address: skarmakar@iitb.ac.in (S. Karmakar).

http://dx.doi.org/10.1016/j.jhydrol.2015.04.024
0022-1694/

dependence structure. FFA is often required for the planning (Stedinger and Griffis, 2008), design (Haddad and Rahman, 2012), and operation of hydraulic structures (De Michele et al., 2005) and for spatial mapping of the flood risk together with vulnerability and exposure for urban or riverine flood management (Karmakar et al., 2010). FFA approaches can be broadly classified into the following categories (Renard et al., 2013): at-site FFA, climate/weather-informed at-site FFA, historical and paleoflood analyses, regional FFA.

1.1. A brief review of univariate and multivariate approaches

A large body of literature is available for at-site FFA, beginning with Cunnane (1985) and extending to a recent study by Rahman et al. (2013), as shown in Fig. 1, wherein FFA is broadly categorized into three different classes. Initially, FFA was predominantly performed on peak flood flow intensities or flood peaks (Cunnane, 1985; Ahmad et al., 1988; Seckin et al., 2011), but in hydrologic planning and design, it is also important to determine the characteristics of the flood volume (area above the threshold value) and flood duration (length of the flood event) and their joint probabilistic behavior together with flood peak because these flood variables are correlated (Singh and Singh, 1991; Adamson et al., 1999; Yue, 1999, 2001; Karmakar and Simonovic, 2008). Hence, many researchers have performed multivariate FFA. For example, Yue (2001) selected a five-parameter bivariate gamma distribution to describe the joint probability distributions, conditional distributions, and associated joint return periods of two correlated variables. Additionally, Yue et al. (2001) provided a comprehensive review of the bivariate gamma distributions for hydrological applications. Furthermore, Yue and Wang (2004) provided a comparison of two bivariate extreme value distributions, i.e., Gumbel mixed and Gumbel logistic models, in applications to FFA. In the same study, the researchers concluded that these two distributions provided the same joint return periods and may be useful for representing the joint statistical properties of the two random variables. Yue and Rasmussen (2002) and Sandoval (2007) used the bivariate extreme value distribution to determine the joint probability and return period of the correlated flood variables. Apart from above mentioned, numerous studies were conducted using different bivariate distributions (Fig. 1), with restrictive assumptions (Zhang and Singh, 2006).

In most multivariate FFA, it is assumed that the marginal distribution of *P*, *V*, and *D* originates from the same parametric family of statistical distributions (e.g., exponential, gamma, extreme value, lognormal, log-Pearson, etc.), whose probability density function (PDF) is known; thus, the procedure is mathematically straightforward (Adamowski, 1989). Numerous studies have described parametric multivariate analyses (a few recent examples are shown in Fig. 1). However, all of these methods contain certain limitations: (i) the joint distributions among flood peak–volume (*P*–*V*), volume–duration (*V*–*D*), and peak–duration (*P*–*D*) originate from the same family of parametric distributions; (ii) the mathematical formulation becomes more complicated if the number of variables are increased; and (iii) it is not possible to distinguish between the marginal and joint behavior of the variables (Grimaldi and Serinaldi, 2006). However, for natural phenomena, the best-fitting marginal distribution may not follow the same family of distributions (Zhang and Singh, 2006). Hence, to overcome the limitations of traditional multivariate FFA, the concept of a copula is introduced, which relaxes the requirement to select from the same family of marginal distributions for the flood variables.

Copulas are functions that combine the best-fit marginals from the univariate distributions of the flood variables to form multivariate distributions (Salvadori and De Michele, 2004). Hence, this approach provides greater flexibility in choosing the univariate

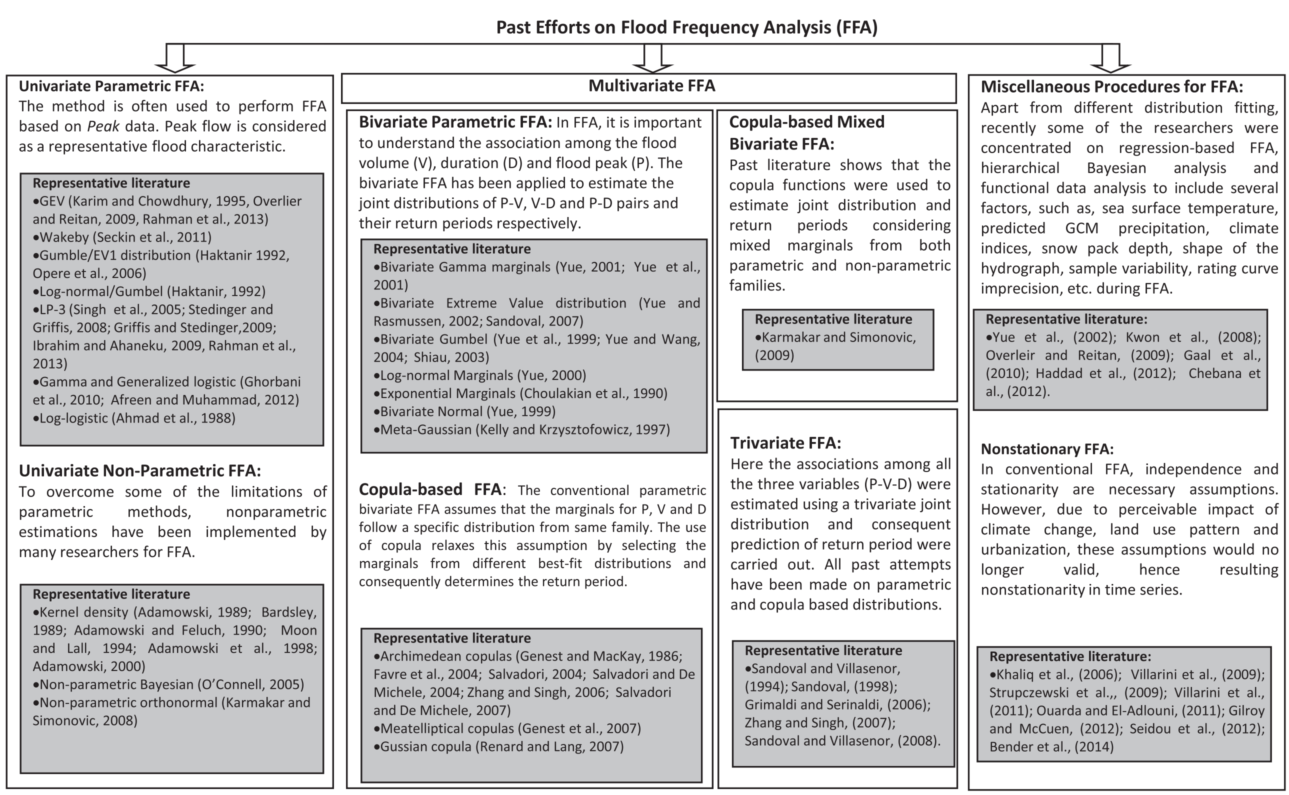

Fig. 1. Past efforts on flood frequency analysis (FFA), showing different classes of frequency analyses with representative literature. (See above-mentioned references for further information.)