Check for updates

A nonparametric statistical framework using a kernel density estimator to approximate flood marginal distributions – a case study for the Kelantan River Basin in Malaysia

Shahid Latif and Firuza Mustafa

ABSTRACT

Floods are becoming the most challenging hydrologic issue in the Kelantan River basin in Malaysia. All three flood characteristics, i.e. peak flow, flood volume and flood duration, are important when formulating actions and measures to manage flood risk. Therefore, estimating the multivariate designs and their associated return periods is an essential element of making informed risk-based decisions in this river basin. In this paper, the efficacy of a kernel density estimator is tested by assessing the adequacy of kernel functions for capturing flood marginal density of 50 years (from 1961 to 2016) of daily streamflow data collected at Gulliemard Bridge gauge station in the Kelantan River basin. Tests for stationarity or the existence of serial correlation within the flood series is often a pre-requisite before introducing the random samples into a univariate or a multivariate framework. It was found that homogeneity existed within the flood vector series. It was concluded therefore that time series of the flood vectors do not exhibit any significant trend. Based on analytically based fitness measures, it was concluded that it is likely that Triweight kernel function is the best-fitted distribution for defining the marginal distribution of peak flows, flood volumes and flood durations in the Kelantan River basin.

Key words | flood, goodness-of-fit statistics, marginal distribution, univariate kernel density estimator

Shahid Latif (corresponding author)
Firuza Mustafa
Department of Geography, Faculty of Arts & Social Sciences,
University of Malaya,
Kuala Lumpur 50603,
Malaysia
E-mail: *macet.shahid@gmail.com*

INTRODUCTION

From the perspective of water resources operational planning in a river basin, the design and operation of flood management infrastructure often demands an accurate estimate of the flow exceedance probability for assessing hydrologic risk (i.e. Salvadori 2004; Xu *et al.* 2015). Flood frequency analysis (FFA) establishes an association between extreme event quantiles and their non-exceedance probabilities by fitting a univariate or multivariate probability distribution function (Cunnane 1988; Rao & Hameed 2000; Zhang 2005). A flood can be defined through three inter-correlated random vectors; that is, peak flow, flood volume and flood duration, which can lead univariate distribution analysis of return periods to underestimate or overestimate floods (Yue *et al.* 2001; Zhang & Singh 2006; Reddy & Ganguli 2012). This has led researchers to explore multivariate distribution analysis for the estimation of flood design quantiles under different notations of return periods based on joint distribution, conditional joint distribution or Kendall's distribution and the establishment of joint probability density functions (or PDFs) and joint cumulative distribution functions (or CDFs) for various possible combinations of the flood vectors (Yue 1999; Zhang & Singh 2006; Daneshkhan *et al.* 2016). Most of the earlier research applied traditional multivariate functions such as bivariate normal, gamma and lognormal distributions (i.e. Yue 1999, 2000, 2001). However, several modelling constraints and

doi: 10.2166/ws.2020.081

Downloaded from http://iwa.silverchair.com/ws/article-pdf/20/4/1509/705551/ws020041509.pdf
by guest

limitations have prompted the incorporation of multivariate copulas distribution analysis (i.e. De Michele & Salvadori 2003; Favre *et al.* 2004) which is frequently undertaken via a parametric framework, where both the marginal distribution of each targeted flood vector and their joint dependence structure are modeled under univariate parametric functions or copulas distribution (i.e. Favre *et al.* 2004; Zhang 2005; Veronika & Halmova 2014; Fan & Zheng 2016 and references therein). While some researchers have implemented copulas under semiparametric settings where marginal functions are approximated via nonparametric; that is, kernel density functions (KDE) or orthonormal series, their joint structures are still modelled under parametric settings (i.e. Karmakar & Simonovic 2008, 2009; Reddy & Ganguli 2012).

Selecting the most parsimonious univariate functions for defining flood marginal distributions is often a mandatory pre-requisite before establishing their joint dependence structure. Parametric functions have always imposed an assumption that the random samples are coming from known populations whose PDFs are predefined; that is, marginal distributions are assumed to follow a specific family of parametric probability functions. In reality, however, the best fitted marginal distributions may not be from the same probability distribution family (Adamowski 1985; Silverman 1986; Kim & Heo 2002; Botev *et al.* 2010). Dooge (1986) has already pointed out that no amount of statistical refinement can overcome the consequences of a lack of prior probability distribution information of the observed random samples. More especially, in the case of multimodal or skewed distributions, parametric distributions might be incompatible and lead to inconsistencies in the estimated quantiles. According to Bardsley (1988) and Bardsley & Manly (1987), the approximation of any distribution tail beyond the largest value under a parameter distribution can be a difficult task. In the last few decades, in the field of hydrologic or flood frequency analysis an attempt has been made to define bona fide density functions via kernel density estimators, or KDE, which are recognized as a flexible and very stable nonparametric data smoothing procedure for approximating or inferencing populations based on finite observational (Adamowski 1989, 2000; Guo 1991; Lall *et al.* 1993; Bowman & Azzalini 1997; Efromovich 1999; Kim *et al.* 2003; Ghosh & Mujumdar 2007). Alternate theoretical overviews for nonparametric setting have been described in the earlier literature such as Rosenblatt (1956), Parzen (1962) and Bartlett (1963). Nonparametric distributions do not require any prior distribution assumptions and have a higher level of flexibility in their univariate function in comparison with parametric density estimators (Adamowski 1989; Moon & Lall 1994).

Adamowski (1985) retrieved the flood frequencies curve using nonparametric kernel estimators while Adamowski & Labatiuk (1987) compared the difference between real and synthetic data (derived from a Monte Carlo procedure) and identified the modelling efficiency of nonparametric density simulations. Similarly, Adamowski (1989) performed comparative assessments between parametric functions (i.e. Weibull and log-Pearson type III) and nonparametric density functions based on the Monte Carlo simulation statistics and found nonparametric density functions demonstrations performed better than the parametric functions. Adamowski & Feluch (1990) were the first to identify the efficiency of kernel estimators through selecting an appropriate kernel function for the extrapolation of distribution samples beyond the available record. Adamowski (1996) applied the nonparametric procedure to drought modelling while Kim & Heo (2002) compared nonparametric distributions from parametric functions for annual maximum flood samples. Lall *et al.* (1993) postulated the different orientations required to select an appropriate shape representation for kernel functions and their bandwidth for different distribution scenarios such as symmetrical, asymmetrical or skewed and mixed data structures. Kim *et al.* (2003) estimated the return periods of low-flow or drought characteristics using a nonparametric distribution framework. Kim *et al.* (2006) established bivariate drought characteristics using the Palmer drought index in a nonparametric procedure. Karmakar & Simonovic (2008) approximated the flood marginal densities by introducing both parametric and nonparametric functions. In this experiment, nonparametric based kernel density functions as well as orthonormal series functions, which are demonstrated earlier in the literature; for example, Bowman & Azzalini (1997) and Efromovich (1999), were used to establish univariate density functions. Santhosh & Srinivas (2013) demonstrated a flood frequency relationship by employing

Downloaded from http://iwa.silverchair.com/ws/article-pdf/20/4/1509/705551/ws020041509.pdf
by guest

a bivariate diffusion process based on the adaptive kernel functions; that is, D-kernel function, and compared this against parametric copulas functions. This demonstration revealed that the nonparametric methodology is a data-driven approach that poses the higher degree of flexibility when establishing probability density functions. Lall *et al.* (1996), Zhang & Karunamuni (1998), Kim *et al.* (2003), Karmakar & Simonovic (2008) and Santhosh & Srinivas (2013) all pointed out the flexibility of an interactive sets of kernel functions for modelling extreme hydrological episodes such as Gaussian or Normal, Epanechnikov, Triangular and Quadratic functions.

Floods are becoming the most challenging hydrologic issue in the Kelantan River basin in Malaysia, and particularly during the period of wet monsoons (DID 2004; MMD 2007). Recent extreme weather across the basin includes the intense and prolonged precipitation in the year 2002, which caused flooding and affected a total area of 1,640 km^2 with a total population of 714,287, while in the year 2014 the worst flood ever recorded in history affected more than 200,000 people in the several parts of the east coast of the Kelantan River basin. The river is about 248 km long and drains a catchment area of 13,100 km^2 The catchment is 150 km long and 140 km wide. During the wet season, between mid-October and mid-January, this basin receives a rainfall of about 2,500 mm. According to the studies of Hussain & Ismail (2013), the Gulliemard Bridge, Lebir and Galas gauge stations have higher flood frequency than the Nenggiri gauge station. In 2018, Alamgir *et al.* (2018) performed a multivariate analysis of floods under a parametric copulas distribution framework for the different gauge stations of the river basin, while Nashwan *et al.* (2018) performed flood susceptibility assessments at the different gauge stations, which revealed that the downstream area is under the highest risk of devastating floods. A number of studies have also identified the negative consequences of land use changes on the catchment's response (i.e. Wan 1996; Jamaliah 2007).

All three flood characteristics; that is, peak flow, flood volume and flood duration, are important when formulating actions and measures to manage flood risk. For example, when designing retention basins or the spillways for reservoirs or any other infrastructures designs where flood storage is involved, the estimation of flood volume is required as well as the peak discharge in order to calculate the impact of inflow on the storage (Gaál *et al.* 2015). Similarly, estimating the joint behaviour of peak flow-flood volume and flood volume-flood duration when assessing flood control or diversion options (i.e. Xu *et al.* 2015). Therefore, estimating the multivariate designs and their associated return periods is an essential element of making informed risk-based decisions in this river basin.

The case study is divided into two parts, with each part covered in a full paper. The objective of this part of the case study is to identify the best marginal distribution of the flood characteristics. In this paper, the efficacy of a kernel density estimator is tested by assessing the adequacy of an interactive set of kernel functions for capturing the flood marginal density. Demonstration of the efficacy of copulas functions for establishing joint distribution functions whose marginal distributions are approximated with nonparametric distribution or kernel density functions (which are also called semiparametric copulas distribution settings) is introduced separately in the companion paper. An at-site event-based or block (annual) maxima methodology based on 50 years (from 1961 to 2016) of daily stream flow data collected at Gulliemard Bridge gauge station in the Kelantan River basin is described. A brief mathematical overview of the nonparametric procedure in estimating univariate margins via kernel density function is provided below. The sampling procedure for estimating multivariate flood characteristics, estimating marginal distribution functions by employing a variety of univariate kernel functions and also via parametric functions are discussed in the third section of this paper.

NONPARAMETRIC PROCEDURE FOR MARGINAL DISTRIBUTIONS

Univariate kernel density estimators, or KDE

Rosenblatt (1956) introduced the concept of kernel estimators through smoothing kernel weights on each of the random observations. The univariate kernel density estimation (or KDE) is a nonparametric approach to approximate the PDFs, say $f(x)$, of a given random

Downloaded from http://iwa.silverchair.com/ws/article-pdf/20/4/1509/705551/ws020041509.pdf
by guest

observations of $'x'$ or flood characteristics such as peak flow, flood volume or flood duration, where inference about the flood populations is made based on finite samples. Therefore, univariate kernel functions are used to estimate the probability density of the random observations having the following statistical property.

Mathematically,

$$\int_{-\infty}^{+\infty} \mathrm{K(x)dx} = 1 \tag{1}$$

where $K(x)$ defines the univariate kernel function and can be used as a PDF (Karmakar & Simonovic 2008). According to Hardle (1991), the kernel functions can be approximated through a general equation as:

$$\mathrm{K_h(x)} = \frac{1}{\mathrm{h}}\mathrm{K}\left(\frac{\mathrm{x}}{\mathrm{h}}\right) \tag{2}$$

where 'h' is the smoothing parameter known as the 'bandwidth of the kernel function' for regulating the level of smoothness and roughness of the shape of the estimated PDF (Moon & Lall 1994). According to Miladinovic (2008) and Kim & Heo (2002), if $\mathrm{X_1, X_2, X_3, \ldots .X_n}$, are independent and identically distributed (or i.i.d) random observations having the PDF $f(x)$, then the univariate kernel density estimates of $f(x)$ are obtained by averaging Equation (2) in the given random observations, as given below:

$$\widehat{\mathrm{f}}_{\mathrm{h}}\mathrm{(x)} = \frac{1}{\mathrm{nh}}\sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{K_h}\left(\frac{\mathrm{x - X_i}}{\mathrm{h}}\right) \tag{3}$$

where 'n' is the number of random observations; X_i is the i$^{\mathrm{th}}$ observation and $\widehat{f_h}(x)$ is the kernel density estimate. Table 1 lists five standard univariate kernel functions, which have been reported previously when determining the PDFs and CDFs of hydrologic or flood vectors (i.e. Moon & Lall 1994; Adamowski 2000; Miladinovic 2008; Karmakar & Simonovic 2008).

The efficiency of the estimated kernel density depends upon two factors: (1) an appropriate choice of the kernel bandwidth and (2) selection of the kernel function. Bandwidth estimation procedures are discussed below. The kernel functions listed in Table 1 are ideally suited to unbounded observations but in actuality most hydrologic data including rainfall, streamflow or humidity and so on, are either upper or lower bounded. This can lead to boundary leakage problems when applying standard kernel functions in the hydrological data domain. This issue has been tackled in a number of ways including using a Method of reflection (Silverman 1986), Methods of transformations (Marron & Ruppert 1994), Beta kernels (Chen 2000) and Adaptive kernels (Botev *et al.* 2010). Such boundary leakage issues during kernel estimation are beyond the scope of this paper.

Table 1 | Some standard univariate kernel functions

Sl. no.	Kernel function	K(x)
1	Epanechnikov	$= 0.75(1 - x^2),\ \lvert x\rvert \le 1$ $= 0,\quad$ otherwise
2	Triangular	$= 1 - \lvert x\rvert,\ \lvert x\rvert \le 1$ $= 0,\quad$ otherwise
3	Bi-weight or Quartic	$= 0.9375(1 - x^2)^2,\ \lvert x\rvert \le 1$ $= 0,\quad$ otherwise
4	Tri-weight	$= 1.09375(1 - x^2)^3,\ \lvert x\rvert \le 1$ $= 0,\quad$ otherwise
5	Cosine	$= \frac{\pi}{4}\cos(\pi x/2),\ \lvert x\rvert \le 1$ $= 0,\quad$ otherwise

The appropriate choice of the kernel smoothing parameter or bandwidth 'h' is often a critical concern in KDE because the shape of kernel density estimates can be affected by the estimated bandwidth (Moon & Lall 1993; Efromovich 1999; Shabri 2002; Kim & Heo 2002). Insufficient smoothing can result in a rough density while over-smoothing can lead to bypass or smoothing of important features (Santhosh & Srinivas 2013). Table 2 summaries eight different bandwidth selection algorithms. Jones *et al.* (1996) and Sharma *et al.* (1998) provide an extended overview of these algorithms and their approaches. Several bandwidth estimation procedures are solely based on minimizing the estimates of the Mean square error or MSE (Shabri 2002). According to Miladinovic (2008), the asymptotic mean integrated square

Downloaded from http://iwa.silverchair.com/ws/article-pdf/20/4/1509/705551/ws020041509.pdf
by guest

Table 2 | Kernel bandwidth selection algorithms

Algorithms	Literature
Rule of Thumb (ROT)	Silverman (1986) & Azzalini (1981)
Least-squares cross-validation (LSCV)	Tarboton *et al.*, (1998)
Bandwidth factorized cross-validation	Kim & Heo (2002)
Smoothed cross-validation	Hall (1992)
Biased cross-validation	Scott & Terrell (1987)
Maximum likelihood cross-validation (MLCV)	Duin (1976)
Plug-in estimates	Wand & Jones (1995)
Asymptotic mean integrated squared error (AMISE)	Ghosh & Huang (1992)

error or AMISE depends on four factors; that is, the kernel bandwidth, the sample size, the kernel function and targeted density function. Therefore, it could be possible to minimize the AMISE value by selecting justifiable kernel functions and their smoothing parameters ‘h’. According to Bowman & Azzalini (1997) and Kim *et al.* (2003), the optimal bandwidth is typically estimated based on estimates of the integrated square error or ISE (Kim & Heo 2002). In other words, mean integrated square error or MISE, which is the expected value of ISE, determines the overall effectiveness of the estimated kernel density estimators during the asymptotically optimal choice of kernel bandwidth ‘h’ and can be derived as:

$$\mathrm{ISE} = \int [\widehat{f}_h(x) - f(x)]^2\, dx \tag{4}$$

$$\begin{aligned}\mathrm{MISE} &= E\left(\int [\widehat{f}_h(x) - f(x)]^2\right) dx \\ &= \int \mathrm{Var}\,\widehat{f}_h(x)dx + \int \mathrm{bias}^2\,\widehat{f}_h(x)dx \end{aligned}\tag{5}$$

where the terms $\int \mathrm{Var}\,\widehat{f}_h(x)dx$ and $\int \mathrm{bias}^2\,\widehat{f}_h(x)dx$ represent the integrated variance and integrated squared bias (Kim & Heo 2002). From Equation (4), it is clear that the second order derivatives of the density functions are required to estimate MISE, and are not defined or unknown. According to Silverman (1986), the rule of thumb or ROT was proposed to minimize the asymptotic MISE value. Therefore, Azzalini (1981) and Silverman (1986) estimated the optimal bandwidth h_0 based on a final distribution being Gaussian or symmetrical and can be formulated as:

$$\text{optimal bandwidth } = h_0 = (1.587)\sigma n^{-1/3} \tag{6}$$

where σ = minimum {Sample standard deviation, (Interquartile range or IQR/1.349)}. Thus, in this paper, the kernel bandwidth is estimated by minimizing the mean integrated square error (MISE) via the optimal bandwidth (h_0) algorithm. Plug-in bandwidth estimators that target the AMISE as the distance to be minimized are also very simple and promising (Wand & Jones 1995). Similarly, the performance of smoothed cross-validation becomes superior only for large sample size (Hall 1992). Beside this, readers are advised to follow the respective papers for visualizing the statistical significance of different bandwidth estimators, as listed in Table 2.

APPLICATION TO ANNUAL MAXIMUM FLOOD SERIES: A CASE STUDY

Data pre-processing stages: extraction of flood characteristics

Monsoonal floods seem to have increased in the Kelantan River basin in Malaysia in the last few decades in terms of both frequency and magnitude (DID 2004; MMD 2007). It is the longest river of Kelantan state, which rises in the Tahan mountain range and flows to the South China Sea in the north-eastern part of Peninsular Malaysia between the geographical location of Lat 4° 30′ N to 6° 15′ N and Long 101°E to 102° 45′ E. The Galas River and the Lebir River are the two major tributaries of the Kelantan River. The land use in the upper catchment is forest while agriculture including paddy farms, rubber and oil-palm plantations are the major land-use activities in the middle and lower areas of the catchment. The precipitation for this region typically varies between 0 mm (in the dry period) to 1,750 mm (in the wet or north-eastern monsoonal period) (DID 2004).

For data analysis, a partial data series (block (annual) maxima-based flood sampling procedure) was adopted to characterise the streamflow at the Gulliemard bridge gauge station in the Kelantan River basin in Malaysia.

Downloaded from http://iwa.silverchair.com/ws/article-pdf/20/4/1509/705551/ws020041509.pdf
by guest

Flood probability distributions based on partial data series only focus on the extreme portion of the hydrograph; that is, either high flow (for floods) or low flow (for droughts) instead of visualizing the entire hydrograph (Hosking *et al.* 1985). Daily streamflows were recorded by the Drainage and Irrigation Department, Malaysia, for the period 1961–2016. Peak flows were selected for each year based on the maximum flow record using Equation (7), while the flood volume and flood duration corresponding to each peak flow were estimated using the methodology described by Yue & Rasmussen (2002), Eckhardt (2005), Gonzales *et al.* (2009), Xu *et al.* (2015) and given in Equations (8) and (9) and illustrated in Figure 1.

$$\begin{aligned} P_i &= \max\{Q_{ij},\ j = SD_i + SD_i + 1,\ \ldots\ldots.,\ ED_i\} \\ &= \text{Annual flood peak series for the ith year} \end{aligned} \tag{7}$$

$$\begin{aligned} V_i &= V_i^{\text{total}} - V_i^{\text{Baseflow}} = \sum_{j=SD_i}^{ED} Q_{ij} - \frac{(1 + D_i)(Q_{is} + Q_{ie})}{2} \\ &= \text{ hydrograph volume series} \end{aligned} \tag{8}$$

$$\begin{aligned} D_i &= ED_i - SD_i \\ &= \text{Hydrograph durations for ith year} \end{aligned} \tag{9}$$

where 'Q_{ij}' is the j^{th} day streamflow magnitude in the i^{th} year; 'Q_{is}' *and* 'Q_{ie}' are the streamflow magnitude at the start date 'SD_i' and end date 'ED_i' of the flood. The flood volume were determined after separating the base flow (i.e. low frequency component) from the direct runoff (i.e. high frequency component). The flood duration extraction was based on the time difference between the rising (SDi) and recession (EDi) limb of the target flood peak flow. A recursive digital filtering procedure in the form of either a single parameter digital filter (i.e. Eckhardt 2004) or a recursive filtering algorithm (Eckhardt 2005) are the two different ways of extracting low-frequency components or base flow separation. In this demonstration, we adopted the Eckhardt (2005) algorithm, which usually provides an effective way to discriminate base flow from direct-surface runoff and is significant for the wider verification of catchments to reveal consistent measures (i.e. Zhang *et al.* 2013). Flood peak flow often attains the maximum value but it is not necessary for flood volume and duration observations (Xu *et al.* 2015). Figure 2 illustrates the time series of annual flood characteristics for the period 1961–2016. The descriptive statistics of the flood characteristics are given in Table 3 and reveal that each flood vector exhibits a positively skewed distribution; that is, asymmetrical behaviour, which is also indicated from the histogram plots given in Figure 3. Figure 4(a) and 4(b) provide the normal

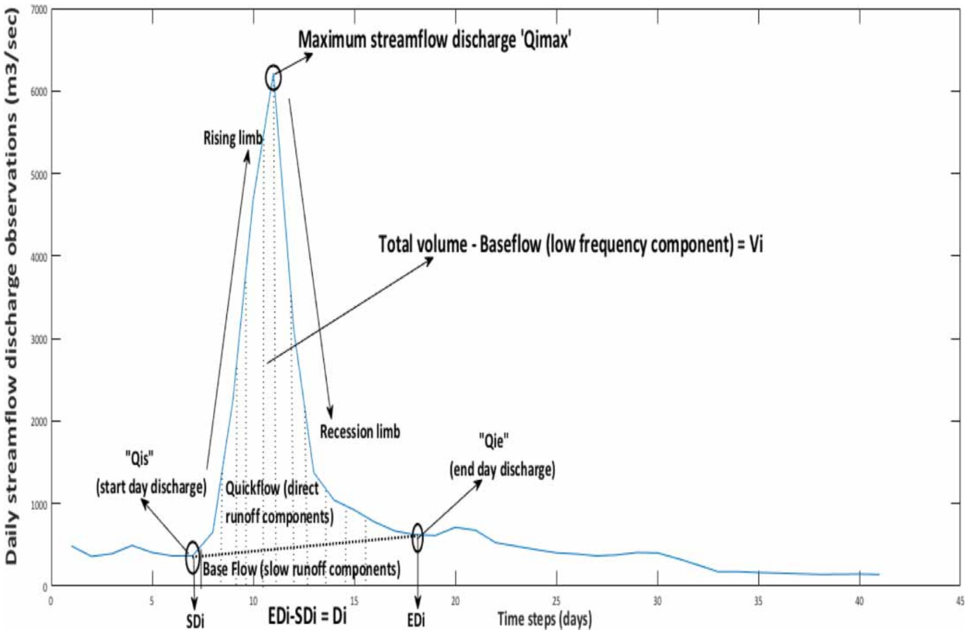


Figure 1 | A typical hydrograph characteristic for the i^{th} flood event.

Downloaded from http://iwa.silverchair.com/ws/article-pdf/20/4/1509/705551/ws020041509.pdf
by guest

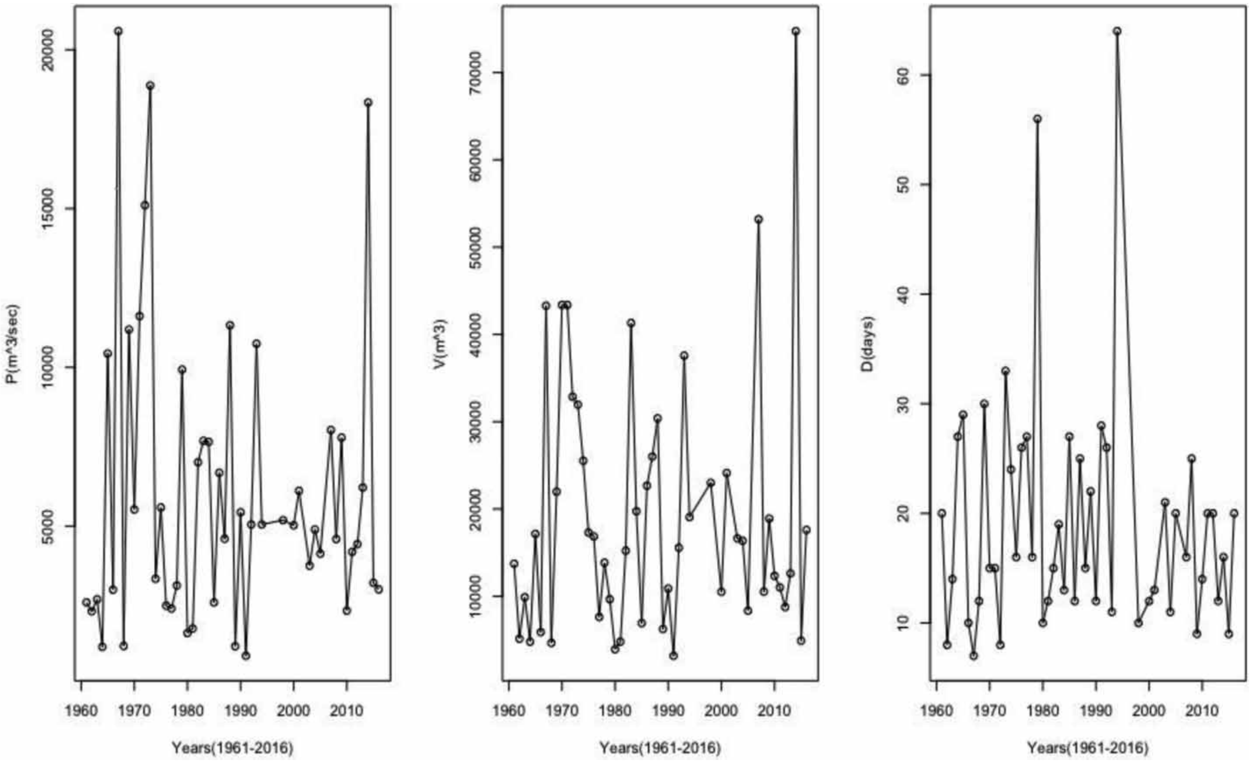


Figure 2 | Time series distribution of block (annual) maxima based flood characteristics between 1961–2016 at the Gulliemard bridge gauge station for Kelantan River basin in Malaysia.

Table 3 | Basic descriptive statistics of annual flood characteristics

Descriptive measure	P(m^3/sec)	V(m^3)	D(days)	Percentile	P(m^3/sec)	V(m^3)	D(days)
Sample size	50	50	50	Min	916.3	3,182.3	7
Range	19,670	71,558	57	5%	1,209.1	4,334.7	8
Mean	6,078	19,122	19.04	10%	1,647.1	4,811.7	9.1
Variance	2.15E + 07	2.14E + 08	117.75	25% (Q1)	2,671.8	8,668.5	12
Standard deviation	4,639	14,623	10.851	50% (Median)	4,961	15,959	16
Coefficient of variation	0.76324	0.76473	0.56993	75% (Q3)	7,711.7	24,476	25
Standard error	656.05	2,068.1	1.5346	90%	11,584	43,077	28.9
Skewness (Fisher)	1.5532	1.6392	2.2793	95%	18,581	47,790	43.35
Skewness (Pearson)	1.506	1.590	2.210	Max	20,586	74,740	64
Kurtosis (Pearson)	1.883	2.864	6.252				
Excess Kurtosis (Fisher)	2.2158	3.3029	7.0557				
Standard error of the mean	656.050	2,068.071	1.535				
Lower bound on mean (95%)	4,759.628	14,966.495	15.956				
Upper bound on mean (95%)	7,396.392	23,278.381	22.124				
Standard error of the variance	4,347,713.616	43,203,375.975	23.790				

Downloaded from http://iwa.silverchair.com/ws/article-pdf/20/4/1509/705551/ws020041509.pdf
by guest

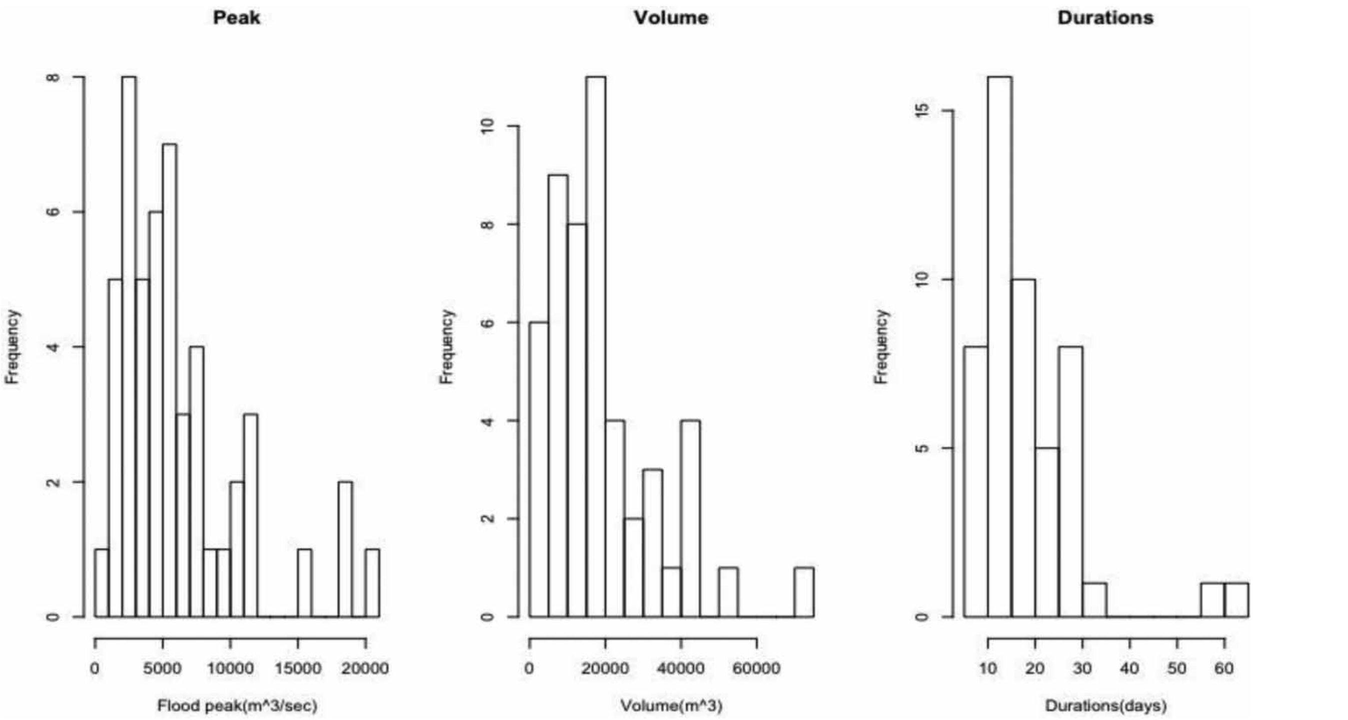


Figure 3 | Histogram plot of flood characteristics.

quantile-quantile (or q-q) plot and box-whisker plot of the annual flood characteristics.

Testing for stationarity within the flood characteristics

Tests for stationarity or the existence of serial correlation (or autocorrelation) within the flood series is often a prerequisite before introducing random samples into a univariate or multivariate framework (Daneshkhan *et al.* 2016). Ljung & Box (1978) based hypothesis testing, which is also known as Q-statistics, were performed on each time series of observations. As indicated in Table 4(a), the tests found negligible or zero first-order autocorrelations for each of the flood vector series for different lag sizes (i.e. lag 20, lag 10, lag 5). A nonparametric rank-based Mann-Kendall (or M-K) test (Mann 1945; Kendall 1975) was also performed to test for the existence of any monotonic trend within the historical flood series. As indicated in Table 4(b), the test found zero monotonic trend at the 5% or 0.05 level of significance within the flood vector series. Testing for the existence of a homogenous environment between any two given time points was also investigated for each flood vector through application of a Pettit test (Pettitt 1979), a Buishand test (Buishand 1982), von Neumann's test (Jaiswal *et al.* 2015) and by undertaking Alexanderson's SNHT based hypothesis testing (Alexandersson 1986). As demonstrated in Table 4(c), it was found that homogeneity existed within the flood vector series. It was concluded therefore that the time series of the flood vector do not exhibit any significant trend.

Nonparametric estimations

Table 1 identified some standard univariate kernel functions, which are adopted when defining a best-fit flood marginal distribution. The bandwidth of the candidate kernel functions was estimated using the optimal bandwidth algorithm given in Equation (6) (Azzalini 1981; Silverman 1986). According to Kim *et al.* (2006), the nonparametric density approximations do not facilitate a closed form of the PDF and CDF, thus CDFs were estimated through an empirical procedure that is based on numerical integration (Kim & Heo 2002).

Some frequently applied parametric family functions such as the Log Pearson type III distribution (Bobee 1974), the Log-normal-2P function (Yue 2000), the Weibull-3P distribution

Downloaded from http://iwa.silverchair.com/ws/article-pdf/20/4/1509/705551/ws020041509.pdf
by guest

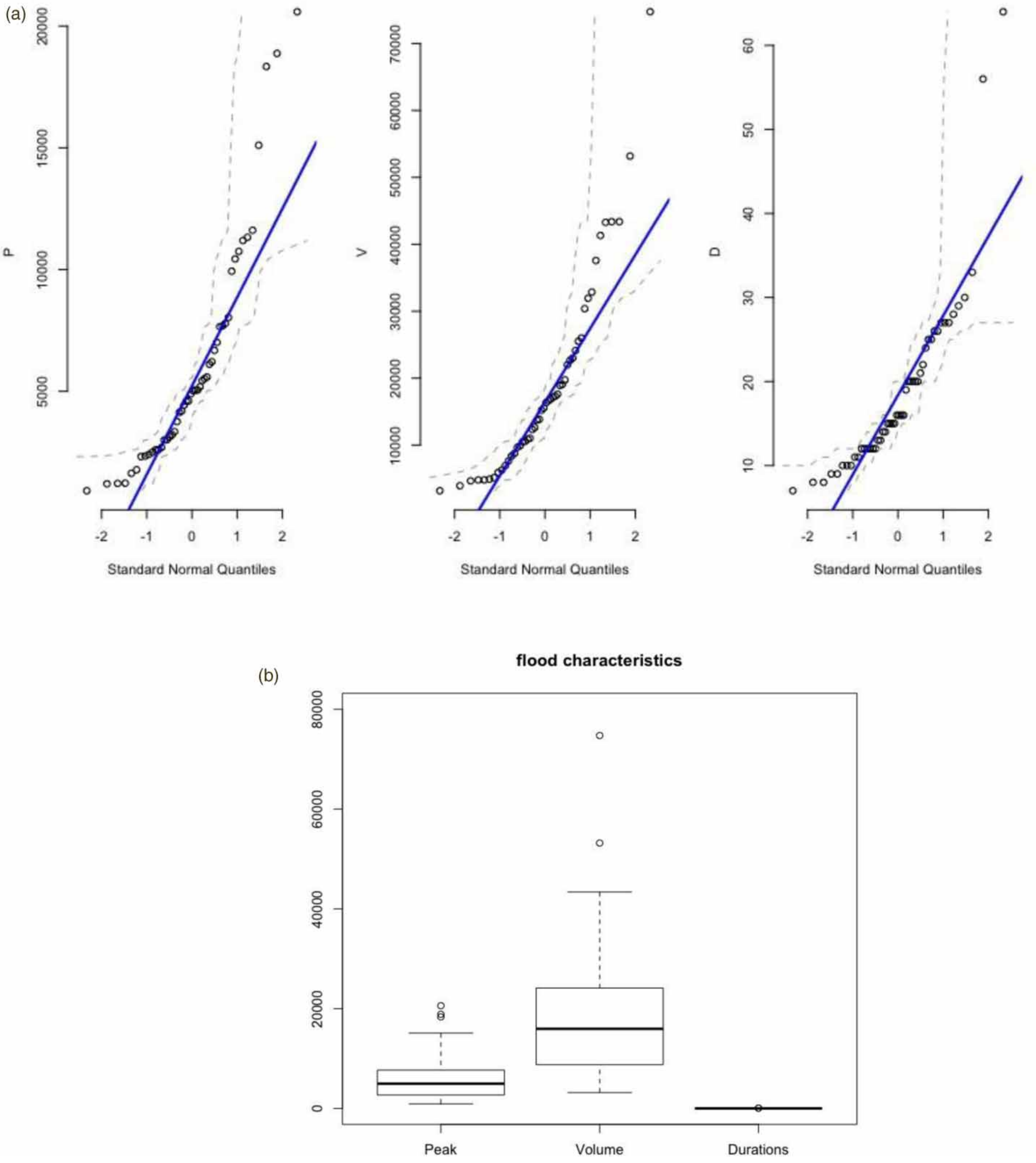


Figure 4 | (a) Normal quantile-quantile (Q-Q) plot, (b) Box-whisker plot of annual flood characteristics.

(Heo *et al.* 2001), the Johnson SB-4P distribution (Keshtkaran & Torabihaghighi 2011), the Gamma-3P (Xu *et al.* 2015) and the Inverse Gaussian-3P functions (Daneshkhan *et al.* 2016) were also tested and compared to the nonparametric estimates. The vectors of unknown statistical parameters or model parameters were estimated using the Maximum

Downloaded from http://iwa.silverchair.com/ws/article-pdf/20/4/1509/705551/ws020041509.pdf
by guest

Table 4 | (a) Q-statistics and their corresponding *p*-value, (b) M-K test for annual flood characteristics between year 1961 and 2016, (c) homogeneity test statistics

(a)

Flood vectors		Lag 20	Lag 10	Lag 5
P	*p*-value	0.78	0.466	0.43
	Q- statistics	14.9	9.719	4.89
V	*p*-value	0.92	0.678	0.99
	Q- statistics	12	7.497	0.5
D	*p*-value	0.73	0.801	0.69
	Q- statistics	15.7	6.171	3.04
Note:	Critical value	31.4104	18.307	11.070

(b)

Series/test	P	V	D
Kendall's tau	0.007	0.006533	−0.041
S	9.000	8	−49.000
Var (S)	7,541.387	8,474.018	8,985.326
p-value (two-tailed)	0.917	0.939386	0.613
Alpha	0.05	0.05	0.05
Sen's slope	2.100	3.954	0.000
Risk for rejecting null hypothesis Ho	91.75%	93.94%	61.26%

(c)

Test	Statistics	P	V	D	Overall conclusion
Pettitt	K	138.000	140	128	
	T	4	8	34	
	p-value (two-tailed)	0.715	0.744	0.555	Homogenous
	Confidence interval@99% on *p*-value	]0.704, 0.727 [	] 0.591, 0.616 [	] 0.542, 0.568 [	
SNHT	T0	3.614	2.992	2.504	
	T	13	6	34	
	p-value (two-tailed)	0.501	0.603	0.697	Homogenous
	Confidence interval @99% on *p*-value	]0.488, 0.513 [	4.051	] 0.685, 0.708 [	
Buishand's	Q	5.956	4.015	5.273	
	T	13	6	34	Homogenous
	p-value (two-tailed)	0.363	0.817	0.519	
	Confidence interval @99% on *p*-value	] 0.351, 0.376 [	] 807, 0.827 [	] 0.506, 532 [	
Von Nuemann's	N	2.080	2.015	2.441	
	p-value (two-tailed)	0.592	0.501	0.970	Homogenous
	Confidence interval @99% on *p*-value	] 0.580, 0.605 [	] 0.488, 0.513 [	] 0.965, 974 [	

Note: *p*-values are computed using 10,000 Monte Carlo simulations.

Downloaded from http://iwa.silverchair.com/ws/article-pdf/20/4/1509/705551/ws020041509.pdf
by guest

likelihood estimators (MLE) and Methods of Moments estimators (MOM) and their estimated values are listed in Table 5.

Goodness-of-fit test

The theoretical cumulative density of each flood series was estimated through the nonparametric procedure and compared against empirical non-exceedance probabilities to assess the data reproducing capabilities and fitness consistency with observational samples. Empirical observations were estimated using the Gringorten plotting position formulae (Gringorten 1963), expressed as:

$$P_i = \frac{i-1}{N+0.12} \tag{10}$$

where 'i' stands for the smallest observations within the data sets of N observations when the data are arranged in ascending order. Several fitness test statistics are incorporated such as using error indices statistics called the Mean Square Error (MSE) and the Root Mean Square Error (RMSE) (Moriasi *et al.* 2007), the Kullback-Leibler information measures (Kullback & Leibler 1951), statistics called the Akaike information criteria (AIC) (Akaike 1974), Schwartz's Bayesian information criteria (BIC) (Schwarz 1978) and the Hannan-Quinn criteria (HQC) (Hannan & Quinn 1979; Burnham & Anderson 2002). The lowest value of RMSE, MSE, AIC, BIC and the HQC statistics indicate the best fit. The AIC statistics include the lack of the fit of the model on one hand and the unreliability of the model due to the number of model parameters on the other hand (Zhang & Singh 2006). Therefore, maximizing the likelihood of fitted distributions or in the context of the maximized value of the likelihood functions, it can be mathematically estimated as:

$$\begin{aligned}\text{AIC} = &-2\log(\text{maximized likehood for fitted model})\\ &+2(\text{number of fitted model parameters})\end{aligned} \tag{11}$$

Also,

$$\begin{aligned}\text{AIC} = &-2\log(\text{MSE})\\ &+2(\text{number of fitted model parameters})\end{aligned} \tag{12}$$

Similarly, the BIC statistics can be formulated as:

$$\begin{aligned}\text{BIC} = &-(\text{sample size})\log(\text{maximized likehood for fitted}\\ &\text{distirbutions}) + [\text{number of fitted model parameters}\\ &\text{Log(sample size)}\end{aligned} \tag{13}$$

Also,

$$\begin{aligned}\text{BIC} = &-(\text{sample size})\log(\text{MSE})\\ &+[\text{number of fitted model parameters Log(sample size)}\end{aligned} \tag{14}$$

The HQC based model selection criteria, which is another alternative to the AIC and BIC statistics (Hannan & Quinn 1979; Burnham & Anderson 2002), can be

Table 5 | Estimated parameters of parametric probability distribution functions

Parametric functions	Flood peak (P)	Volume (V)	Durations (D)
Log-Pearson-3P	a = 663.54, b = −0.02887, g = 27.608	a = 1,781.0, b = −0.01771, g = 41.234	a = 14.523, b = 0.12506, g = 1.0099
Lognormal-2P	s = 0.7362, m = 8.4513	s = 0.74093, m = 9.594	s = 0.4717, m = 2.826
Weibull-3P	a = 1.1175, b = 5,389.8, g = 899.42	a = 1.0689, b = 16,369.0, g = 3,155.6	a = 1.1951, b = 12.878, g = 6.9279
Johnson SB-4P	g = 1.5161, d = 0.74495, l = 27,319.0, x = 1,304.2	g = 2.2027, d = 1.0357, l = 1.3052E + 5, x = 961.8	g = 2.5314, d = 0.92215, l = 118.81, x = 8.2791
Gamma-3P	a = 1.2106, b = 4,290.0, g = 884.47	a = 1.0848, b = 14,723.0, g = 3,150.8	a = 1.4696, b = 8.3319, g = 6.7958
Inverse Gaussian-3P	l = 10,556.0, m = 6,320.9, g = −242.85	l = 26,884.0, m = 19,086.0, g = 36.267	l = 28.913, m = 14.81, g = 4.2297

Downloaded from http://iwa.silverchair.com/ws/article-pdf/20/4/1509/705551/ws020041509.pdf
by guest

formulated as:

$$\mathrm{HQC} = -2\mathrm{L}_{\max} + 2\mathrm{k}\log(\log(\mathrm{N})) \tag{15}$$

where $\mathrm{L}_{\max}$ signifies the model log-likelihood of the total number of fitted parameters 'k' for the 'N' sample size. The HQC statistics are not an estimator of Kullback-Leibler divergence (Burnham & Anderson 2002) and are not an asymptotically efficient criterion (Claeskens & Hjort 2008; Haggag 2014). Such characteristics are identical to the BIC statistics; however, the HQC statistics exhibited a higher level of consistency. Similarly, the MSE and RMSE are estimated as:

$$\mathrm{MSE} = \sum_{i=1}^{N} (x_i^{\mathrm{Model}} - x_i^{\mathrm{Empirical}})^2/\mathrm{N} \tag{16}$$

and,

$$\mathrm{RMSE} = \sqrt{\mathrm{MSE}} = \sqrt{\sum_{i=1}^{N} (x_i^{\mathrm{Model}} - x_i^{\mathrm{Empirical}})^2/\mathrm{N}} \tag{17}$$

where 'xi' indicating the ith series of sample size N.

Table 6 | Analytical comparison for different probability functions

		Error indices statistics		**Information criteria statistics**		
Flood vector	**F(X)**	**MSE (or mean square error)**	**RMSE (or root mean square error)**	**AIC (or Akaike information criteria)**	**BIC (or Bayesian information criteria)**	**HQC (or Hannan-Quinn information criteria)**
P	Epanechnikov	0.00038	0.01957	−391.37	−389.45	−390.64
	Bi-weight or quartic	0.00026	0.01620	−410.25	−408.34	−409.52
	Triweight	0.00022	0.01483	−419.07	−417.16	−418.34
	Triangular	0.00028	0.01686	−406.26	−404.35	−405.54
	Cosine	0.00032	0.01800	−399.98	−398.07	−399.25
	LogPearson-3P	0.00045	0.02127	−379.03	−373.30	−376.85
	Lognormal-2P	0.00046	0.02163	−379.34	−375.52	−377.89
	Weibull-3P	0.01612	0.12699	−200.89	−194.62	−200.90
	Johnson SB-4P	0.00093	0.03053	−340.89	−333.25	−337.99
	Gamma-3P	0.01173	0.10828	−216.30	−210.56	−214.12
	Inverse Gaussia-3P	0.01636	0.12792	−199.63	−193.89	−197.45
V	Epanechnikov	0.00093	0.03060	−346.66	−344.75	−345.93
	Bi-weight or quartic	0.00018	0.01350	−428.44	−426.53	−427.71
	Triweight	0.00016	0.01287	−433.27	−431.36	−432.55
	Triangular	0.00020	0.01426	−423.01	−421.10	−422.29
	Cosine	0.00022	0.01514	−417.02	−415.11	−416.30
	Log-Pearson-3P	0.00048	0.02207	−375.31	−369.58	−373.13
	Lognormal-2P	0.00055	0.02351	−371.02	−367.20	−369.57
	Weibull-3P	0.00047	0.02182	−376.47	−370.74	−374.29
	Johnson SB-4P	0.00041	0.02027	−381.82	−374.17	−378.91
	Gamma-3P	0.01327	0.11520	−210.10	−204.37	−207.92
	Inverse Gaussian-3P	0.00077	0.02783	−352.16	−346.42	−349.98
D	Epanechnikov	0.00059	0.02430	−369.69	−367.77	−368.96
	Bi-weight or quartic	0.00051	0.02265	−376.71	−374.80	−375.99
	Triweight	0.00048	0.02208	−379.27	−377.36	−378.54
	Triangular	0.00055	0.02357	−372.74	−370.83	−372.01
	Cosine	0.00062	0.02496	−367.03	−365.12	−366.30
	Log-Pearson-3P	0.00100	0.03171	−339.08	−333.34	−336.89
	Lognormal-2P	0.00132	0.03635	−327.46	−323.63	−326.00
	Weibull-3P	0.01619	0.12726	−200.15	−194.41	−197.97
	Johnson-4P	0.00972	0.09861	−223.65	−216.00	−220.74
	Gamma-3P	0.00091	0.03012	−343.62	−337.88	−341.43
	Inverse Gaussian-3P	0.00091	0.03027	−343.74	−338.00	−341.55

Downloaded from http://iwa.silverchair.com/ws/article-pdf/20/4/1509/705551/ws020041509.pdf
by guest

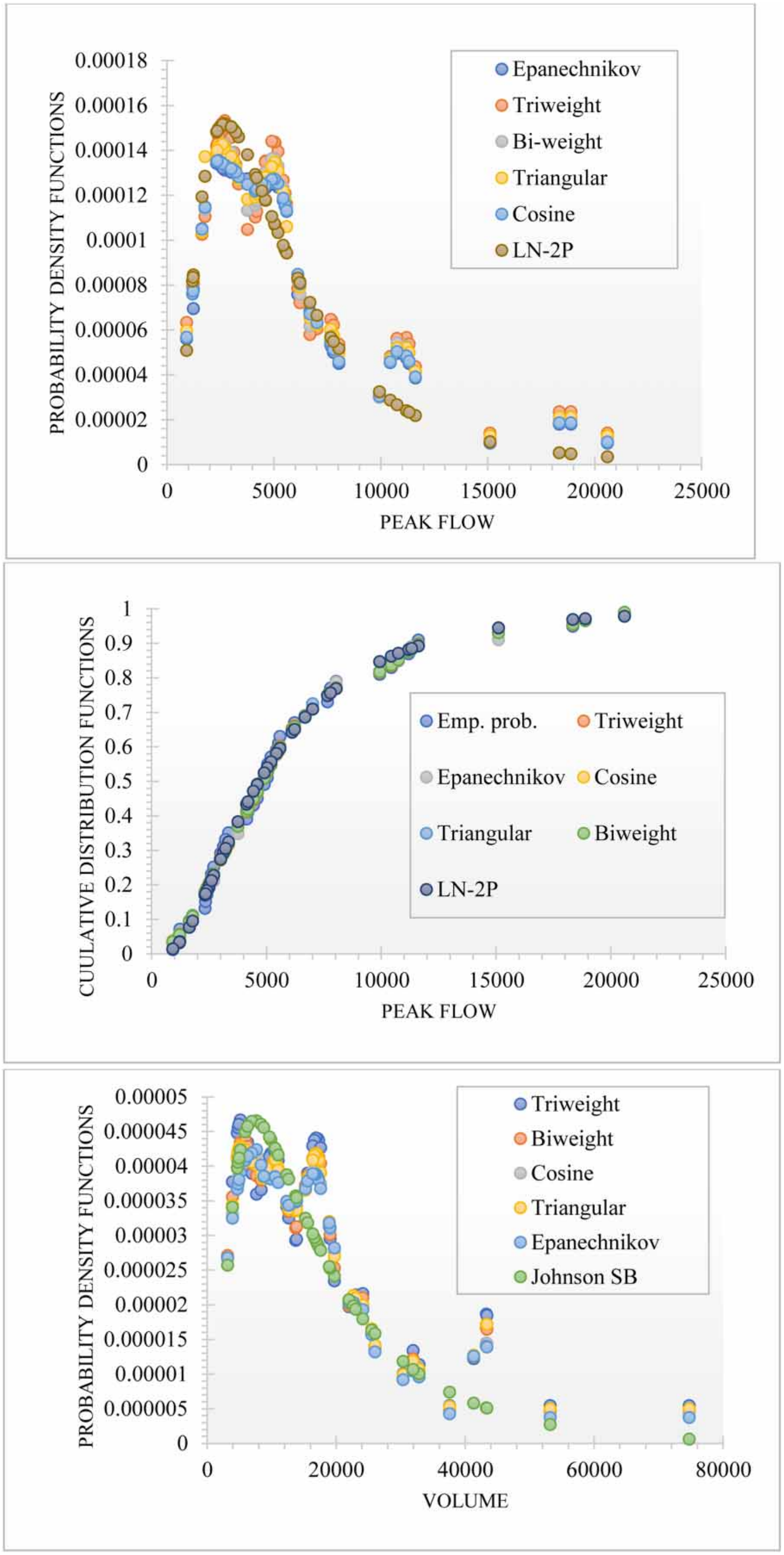


Figure 5 | Probability density functions (or PDFs), cumulative distribution functions (or CDFs) and probability-probability (p-p) plot of annual flood series. *(Continued.)*

Downloaded from http://iwa.silverchair.com/ws/article-pdf/20/4/1509/705551/ws020041509.pdf
by guest

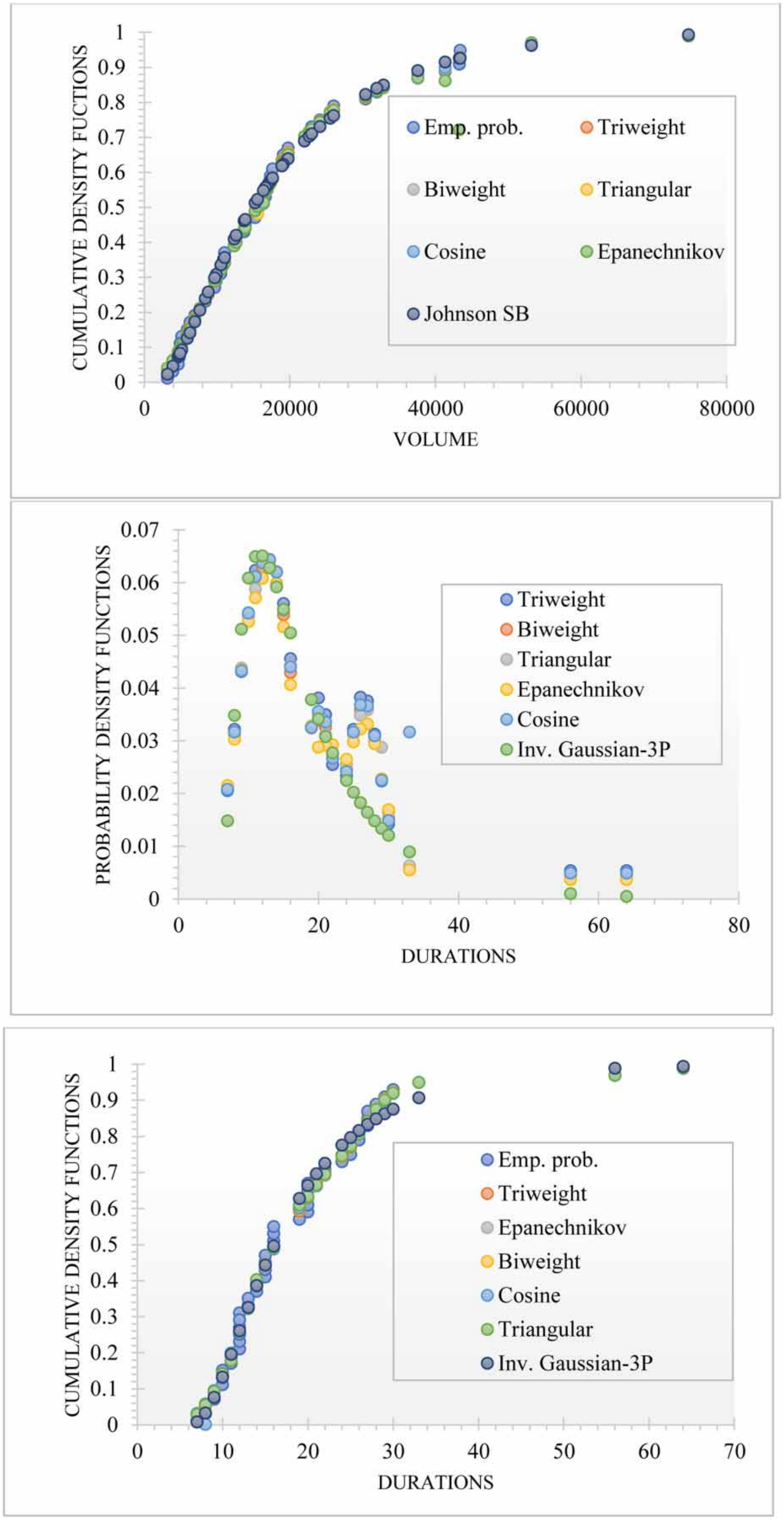


Figure 5 | Continued.

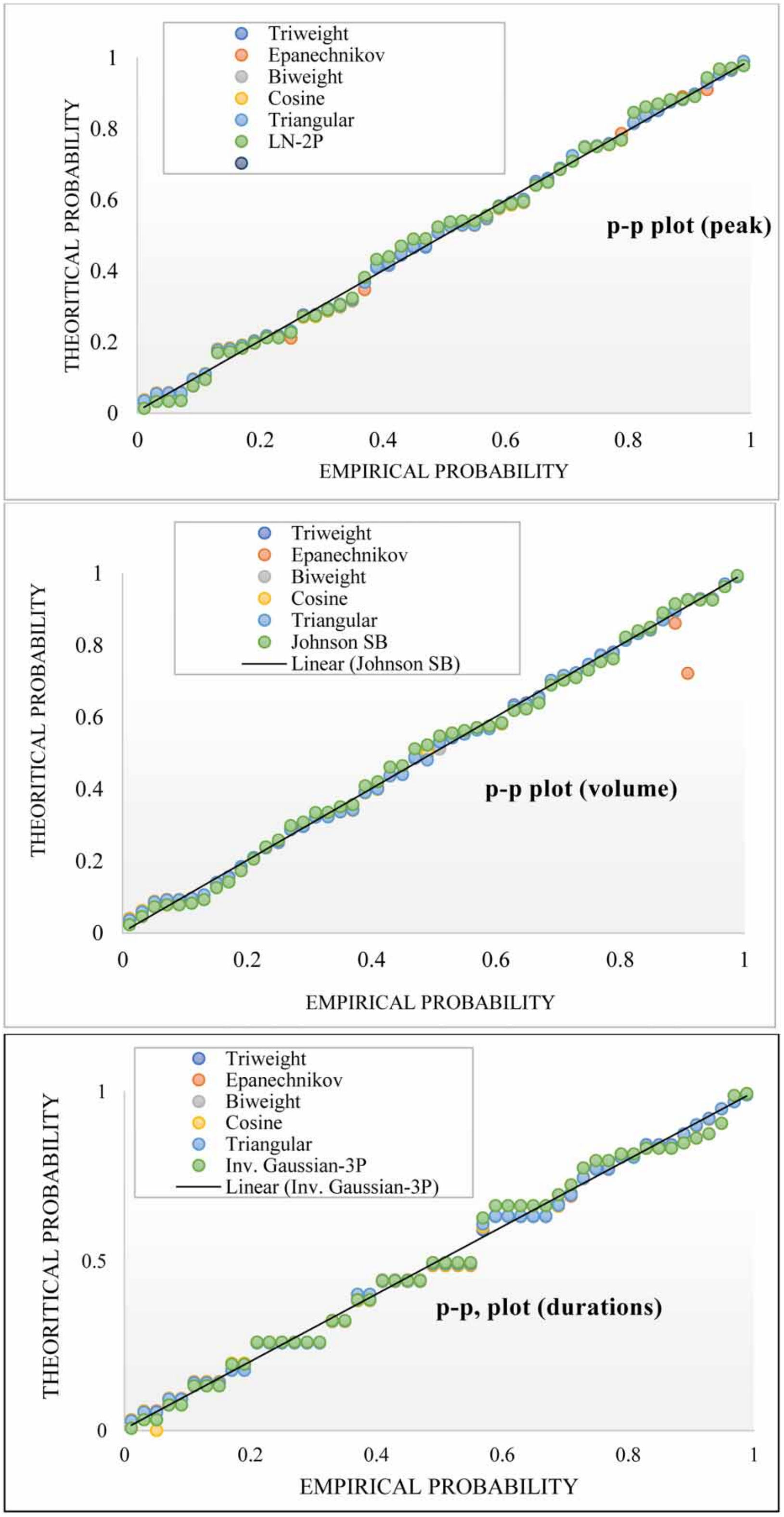


Figure 5 | Continued.

Downloaded from http://iwa.silverchair.com/ws/article-pdf/20/4/1509/705551/ws020041509.pdf
by guest

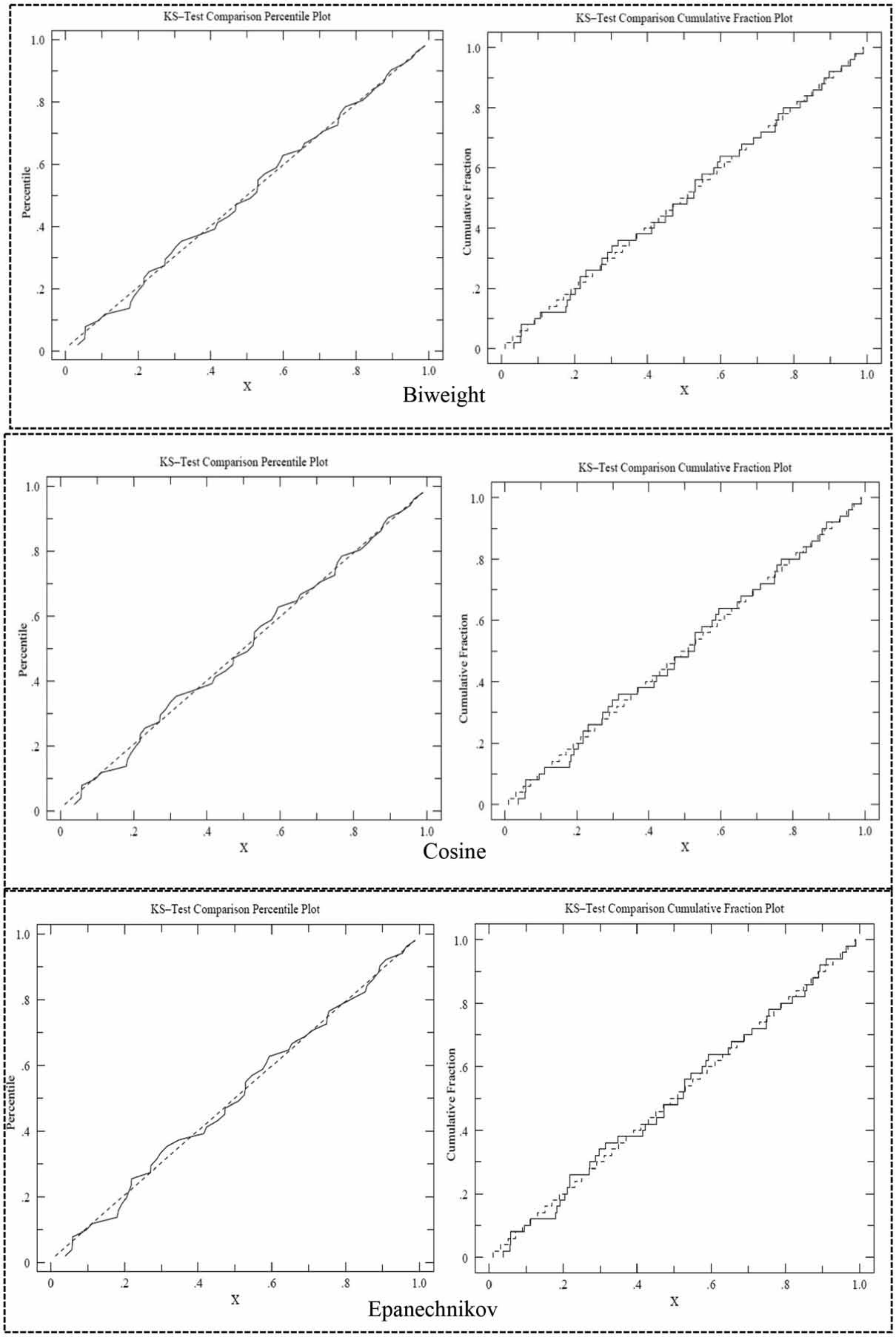


Figure 6 | K-S test comparison cumulative and percentile plot of fitted distributions for (a) flood peak discharge series, (b) volume series, (c) durations series. *(Continued.)*

Downloaded from http://iwa.silverchair.com/ws/article-pdf/20/4/1509/705551/ws020041509.pdf
by guest

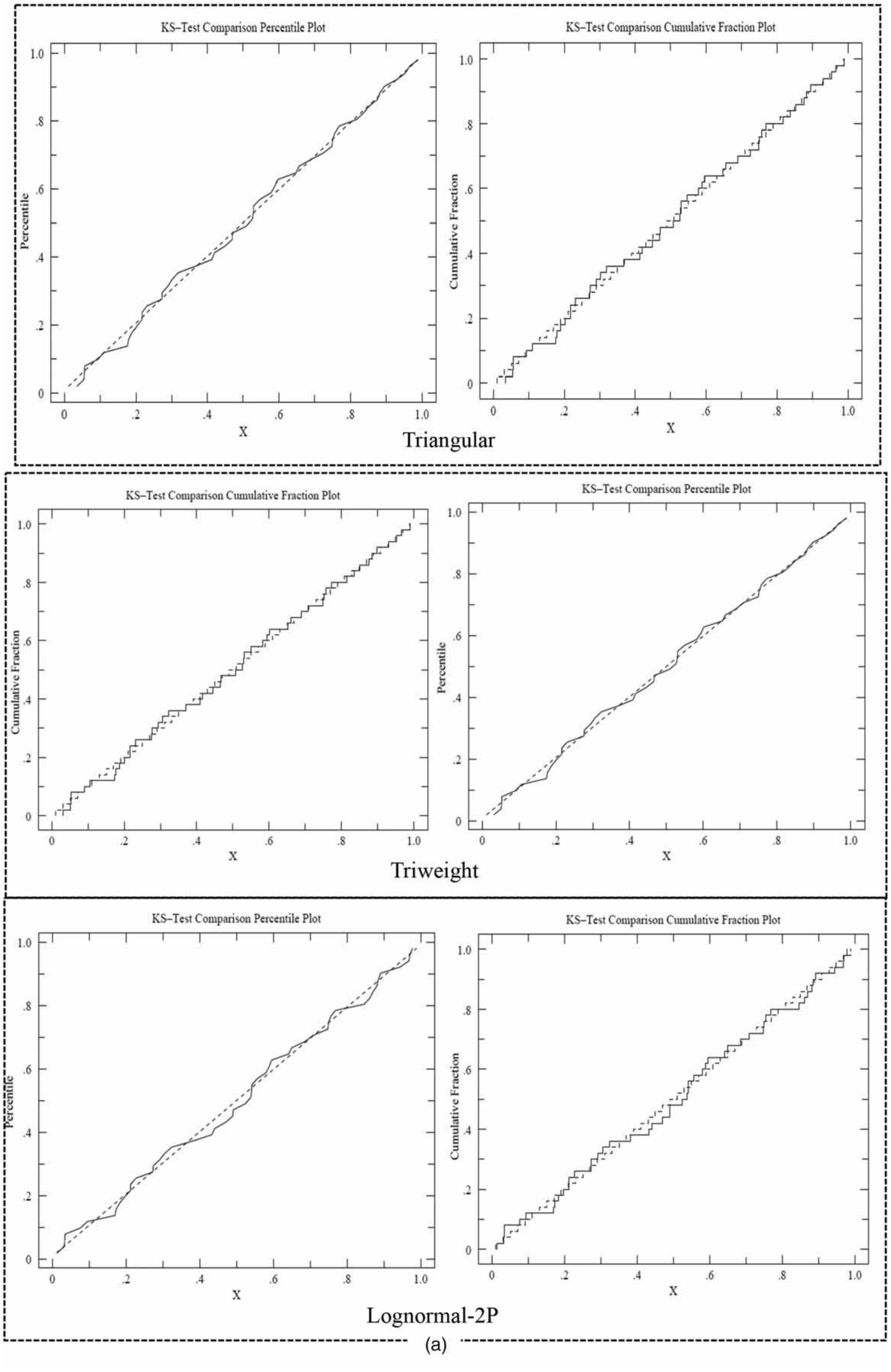


Figure 6 | Continued.

Downloaded from http://iwa.silverchair.com/ws/article-pdf/20/4/1509/705551/ws020041509.pdf
by guest

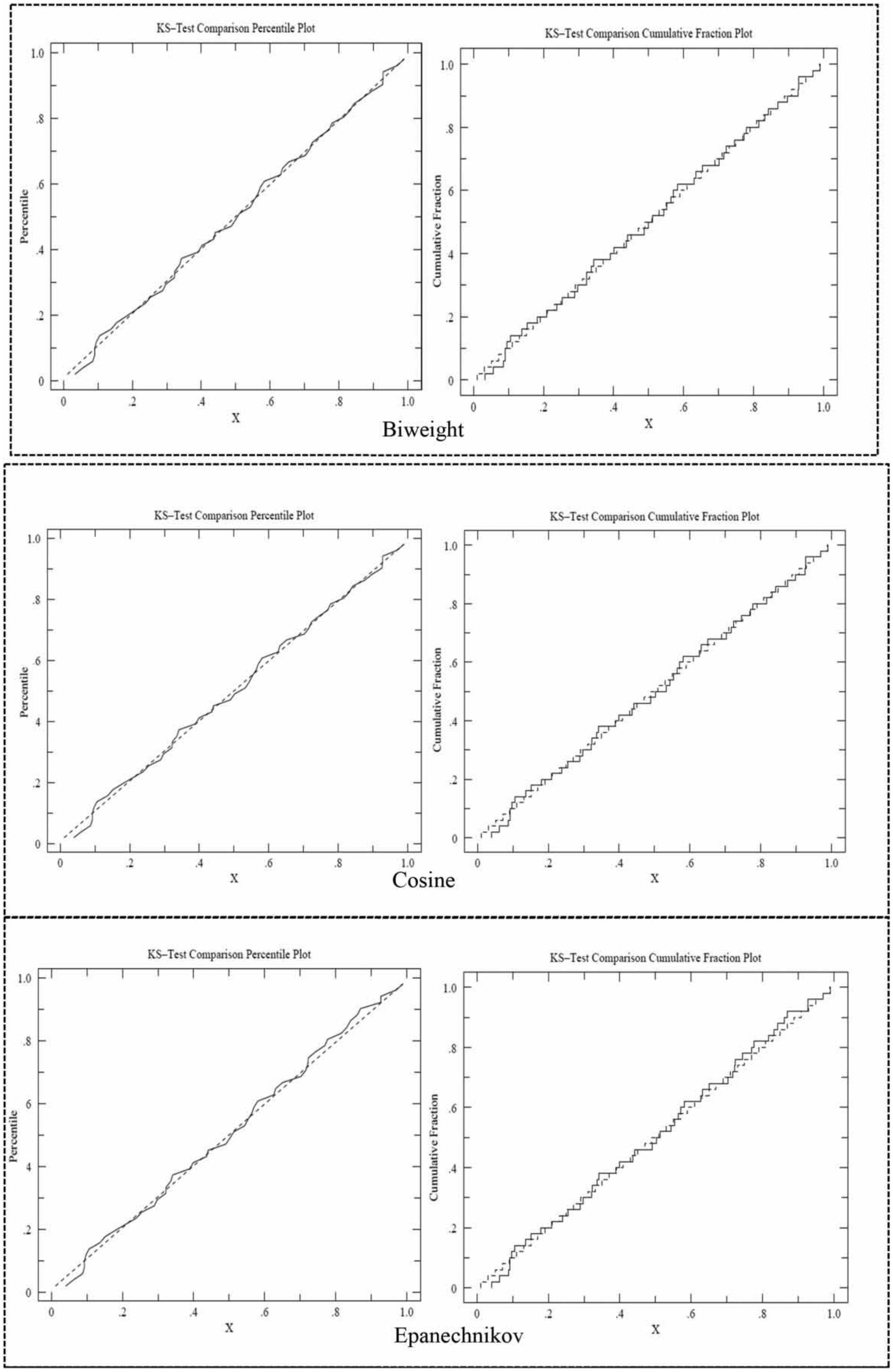
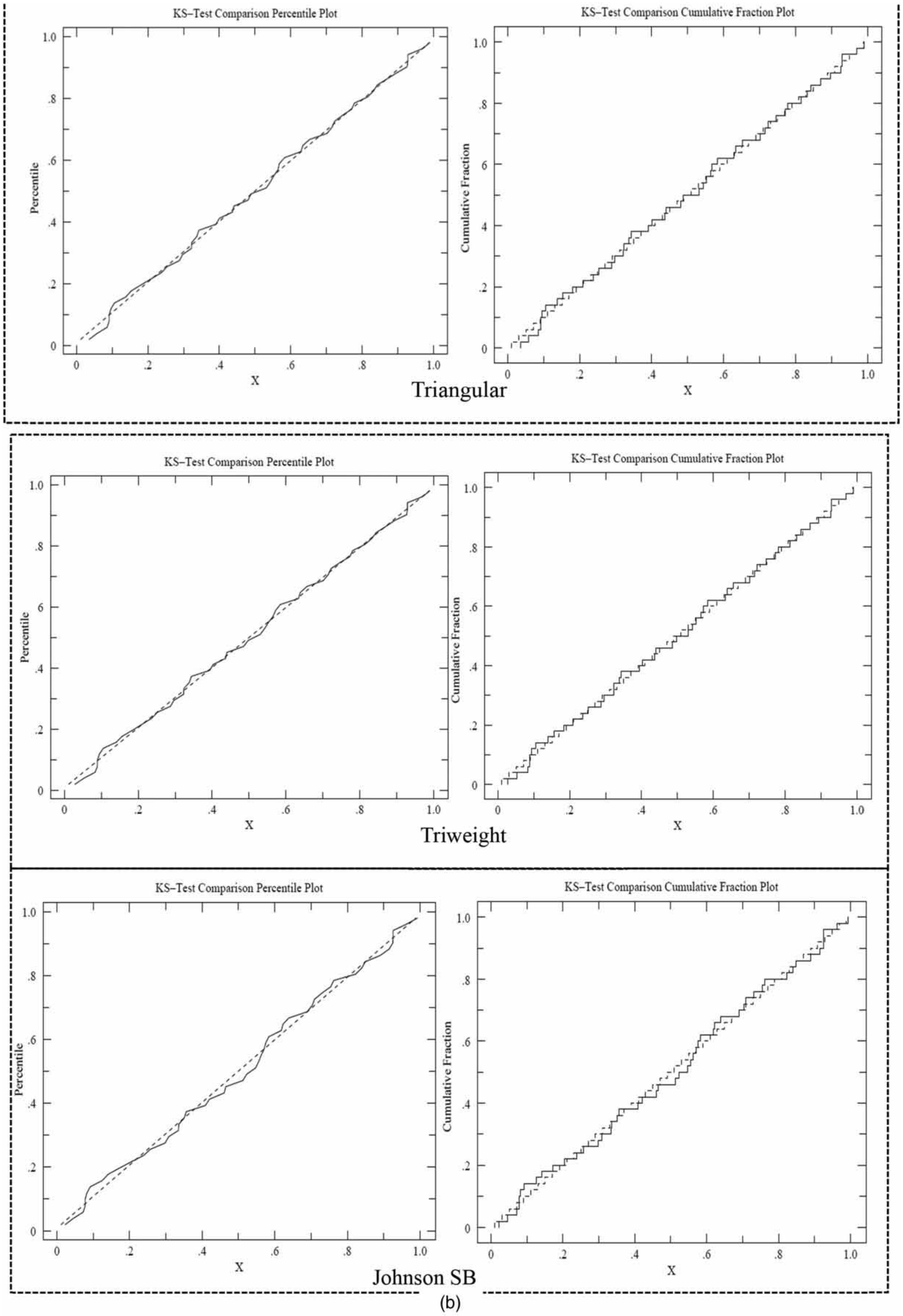


Figure 6 | Continued.

Downloaded from http://iwa.silverchair.com/ws/article-pdf/20/4/1509/705551/ws020041509.pdf
by guest



(b)

Figure 6 | Continued.

Downloaded from http://iwa.silverchair.com/ws/article-pdf/20/4/1509/705551/ws020041509.pdf
by guest

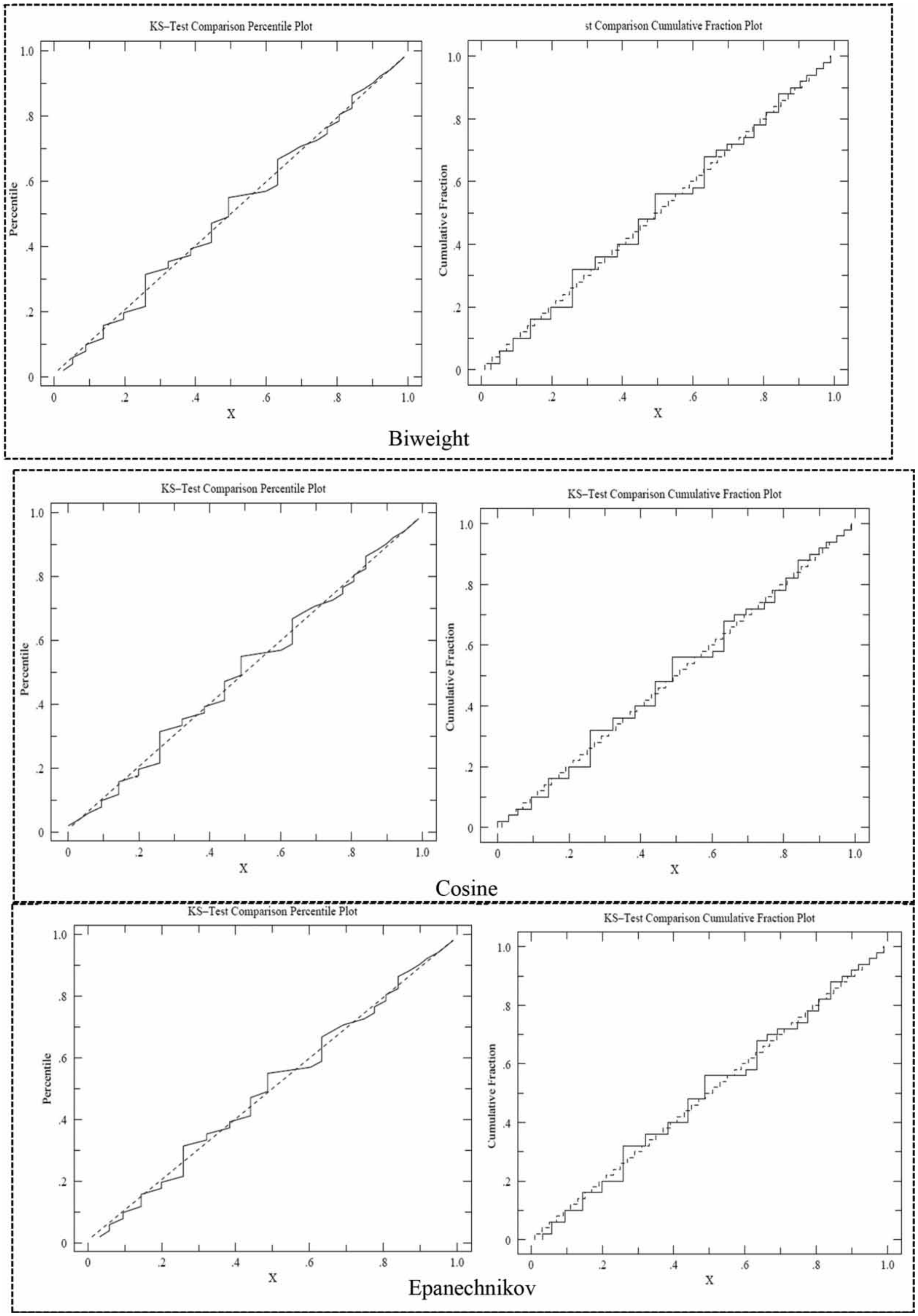


Figure 6 | Continued.

Downloaded from http://iwa.silverchair.com/ws/article-pdf/20/4/1509/705551/ws020041509.pdf
by guest

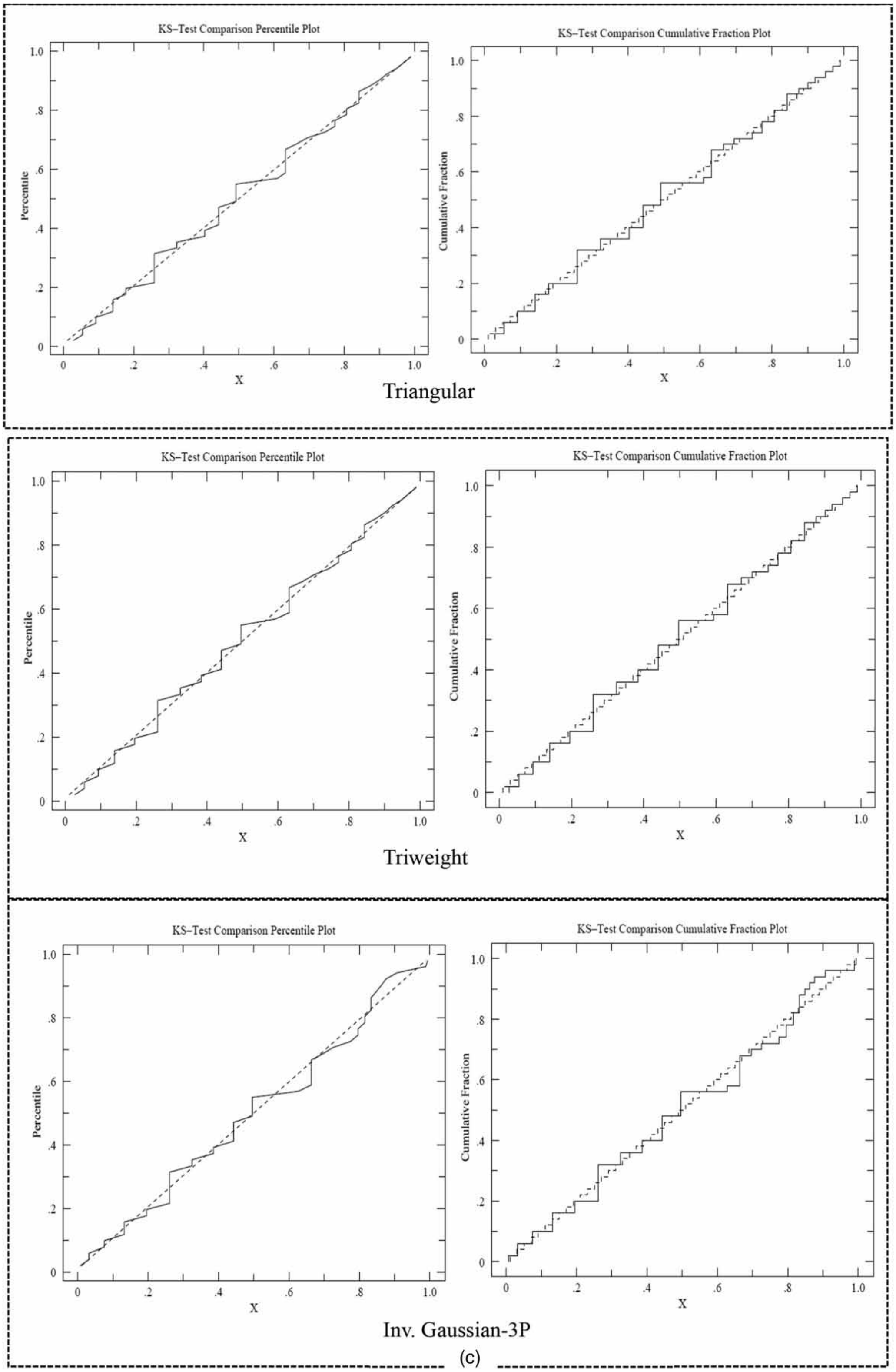


Figure 6 | Continued.

Table 6 presents the estimated values of MSE, RMSE, AIC, BIC and the HQC statistics of all the nonparametric kernel distribution functions fitted to the flood vectors. It was found that the Triweight kernel outperformed other functions as it gave the lowest values of the fitness test statistics for all three flood distribution series; that is, MSE (0.00022), RMSE (0.01483), AIC (−419.0760), BIC (−417.164) and HQC (−418.348) for peak flow, MSE (0.00016), RMSE (0.01287), AIC (−433.27), BIC (−431.36) and HQC (−432.55) for flood volume series and MSE (0.00048), RMSE (0.02208), AIC (−379.27), BIC (−377.26) and HQC (−378.54) for flood duration. The performance of the Biweight and Triangular functions was also more effective than the targeted parametric functions. While the Epanechnikov kernel was less effective than the other candidate kernel functions, it was still better than the parametric functions as revealed in Table 6. Based on analytically based fitness measures, it was concluded that it is likely that Triweight kernel function is the best-fitted distribution for defining the marginal distribution of peak flows, flood volumes and flood durations in the Kelantan River basin.

A qualitative approach based on a graphically based visual inspection was also conducted for each flood vector for the probability density plot, the cumulative density plot, the probability- probability (or p-p) plot, the K-S test comparison cumulative fraction plot and the K-S test comparison percentile plot as illustrated in Figures 5 and 6(a)–6(c). It is noted that the Kolmogorov-Smirnov test (K-S) is a nonparametric distribution-free test that seeks to investigate the largest vertical gap between cumulative empirical and theoretical probabilities and also has the advantage of not assuming the distribution of data (Xu *et al.* 2015).

It was concluded that these plots clearly indicate the effectiveness of a nonparametric kernel structure and support the adoption of a Triweight kernel function for defining univariate flood marginal distributions.

CONCLUSIONS

Floods are becoming the most challenging hydrologic issue in the Kelantan River basin in Malaysia, and particularly during the period of wet monsoons. All three flood characteristics; that is, peak flow, flood volume and flood duration, are important when formulating actions and measures to manage flood risk. Therefore, estimating the multivariate designs and their associated return periods is an essential element of making informed risk-based decisions in this river basin.

In this paper, the efficacy of a kernel density estimator is tested by assessing the adequacy of an interactive set of kernel functions for capturing the flood marginal density of 50 years (from 1961 to 2016) of daily stream flow data collected at Gulliemard Bridge gauge station in the Kelantan River basin.

Tests for stationarity or existence of serial correlation (or autocorrelation) within the flood series is often a pre-requisite before introducing the random samples into a univariate or a multivariate framework. It was found that homogeneity existed within the flood vector series. It was concluded therefore that the time series of the flood vectors do not exhibit any significant trend.

Based on analytically based fitness measures, it was concluded that it is likely that Triweight kernel function is the best-fitted distribution for defining the marginal distribution of peak flows, flood volumes and flood durations in the Kelantan River basin.

ACKNOWLEDGEMENTS

Special thanks are extended to the Drainage and Irrigation Department (DID), Malaysia, for supplying streamflow data of the Kelantan river basin.

REFERENCES

Adamowski, K. 1985 Nonparametric kernel estimation of flood frequencies. *Water Resour. Res.* **21** (11), 1885–1890.

Adamowski, K. 1989 A Monte Carlo comparison of parametric and nonparametric estimations of flood frequencies. *J. Hydrol.* **108**, 295–308.

Adamowski, K. 1996 Nonparametric estimations of low-flow frequencies. *J. Hydraul. Eng.* **122** (1), 46–49.

Adamowski, K. 2000 Regional analysis of annual maximum and partial duration flood data by nonparametric and L-moment methods. *J. Hydrol.* **229** (3), 219–231.

Adamowski, K. & Feluch, W. 1990 Nonparametric flood-frequency analysis with historical information. *J. Hydraul. Eng.* **116** (8), 1035–1047.

Downloaded from http://iwa.silverchair.com/ws/article-pdf/20/4/1509/705551/ws020041509.pdf
by guest

Adamowski, K. & Labatiuk, C. 1987 Estimation of flood frequencies by a nonparametric density procedure. In: Singh V. P. (ed.), *Hydrologic Frequency Modeling*. Springer, Dordrecht, The Netherlands, pp. 97–106.

Akaike, H. 1974 A new look at the statistical model identification. *IEEE Trans. Automat. Contr.* **19** (6), 716–723.

Alamgir, M., Ismail, T. & Noor, M. 2018 Bivariate frequency analysis of flood variables using copula in Kelantan River Basin. *Malaysian J. Civ. Eng.* **30** (3), 395–404.

Alexandersson, H. 1986 A homogeneity test applied to precipitation test. *J. Climatol.* **6**, 661–675.

Azzalini, A. 1981 A note on the estimation of a distribution function and quantiles by a kernel method. *Biometrika* **68**, 326–328.

Bardsley, W. E. 1988 Toward a general procedure for analysis of extreme random events in the earth sciences. *Math. Geol.* **20** (5), 513–528.

Bardsley, W. E. & Manly, B. F. J. 1987 Transformations for improved convergence of distributions of flood maxima to a Gumbel limit. *J. Hydrol.* **91**, 137–152.

Bartlett, M. S. 1963 Statistical estimation of density functions. *Sankhya, Indian J. Stat., Ser. A* **25** (3), 245–254.

Bobee, B. 1974 The log Pearson type 3 distribution and its application in hydrology. *Water Resour. Res.* **11** (5), 681–689.

Botev, Z. I., Grotowski, J. F. & Kroese, D. P. 2010 Kernel density estimation via diffusion. *Ann. Stat.* **38** (5), 2916–2957. doi:10.1214/10-AOS799.

Bowman, A. & Azzalini, A. 1997 *Applied Smoothing Techniques for Data Analysis: the Kernel Approach with S-Plus Illustrations*. Oxford University Press, New York, NY.

Buishand, T. A. 1982 Some methods for testing the homogeneity of rainfall records. *J. Hydrol.* **58** (1–2), 11–12.

Burnham, K. P. & Anderson, D. R. 2002 *Model Selection and Inference: A Practical Information-Theoretic Approach*, 2nd edn. Springer-Verlag, New York. http://dx.doi.org/10.1007/b97636.

Chen, S. 2000 Beta kernel smoothers for regression curves. *Stat. Sin.* **10**, 73–91.

Claeskens, G. & Hjort, N. L. 2008 *Model Selection and Model Averaging*. Cambridge, University Press, Cambridge, UK.

Cunnane, C. 1988 Methods and merits of regional flood frequency analysis. *J. Hydrol.* **100**, 269–290.

Daneshkhan, A., Remesan, R., Omid, C. & Holman, I. P. 2016 Probabilistic modelling of flood characteristics with parametric and minimum information pair-copula model. *J. Hydrol.* **540**, 469–487.

De Michele, C. & Salvadori, G. 2003 A generalized pareto intensity-duration model of storm rainfall exploiting 2-copulas. *J. Geophys. Res.* **108** (D2), 4067.

DID (Drainage and Irrigation Department Malaysia) 2004 *Annual Flood Report of DID for Peninsular Malaysia*. Unpublished report. DID, Kuala Lumpur.

Dooge, J. C. E. 1986 Looking for hydrologic laws. *Water Resour. Res.* **22** (9), 465–485.

Duin, R. P. W. 1976 On the choice of smoothing parameters for Parzen estimators of probability density functions. *IEEE T. Comput.* **C-25** (11), 1175–1179.

Eckhardt, K. 2004 How to construct recursive digital filters for baseflow separation. *Hydrol. Process.* **19** (2), 507–515

Eckhardt, K. 2005 How to construct recursive digital filters for baseflow separation. *Hydrol. Process.* **19**, 507–515.

Efromovich, S. 1999 *Nonparametric Curve Estimation: Methods, Theory and Applications*. Springer-Verlag, New York, NY.

Fan, L. & Zheng, Q. 2016 Probabilistic modelling of flood events using the entropy copula. *Adv. Water Resour.* **97**, 233–240.

Favre, A. C., Adlouni, S. E., Pereault, L., Thiemonge, N. & Bobee, B. 2004 Multivariate hydrological frequency analysis using copulas. *Water Resour. Res.* **40** (1), WR002456.

Gaál, L., Szolgay, J., Kohnová, S., Hlavčová, K., Parajka, J., Viglione, A. & Blöschl, G. 2015 Dependence between flood peaks and volumes: a case study on climate and hydrological controls. *Hydrolog. Sci. J.* **60** (6), 968–984.

Ghosh, B. K. & Huang, W. M. 1992 Optimum bandwidths and kernels for estimating certain discontinuous densities. *Ann. Inst. Statist. Math.* **44** (3), 563–577.

Ghosh, S. & Mujumdar, P. P. 2007 Nonparametric methods for modeling GCM and scenario uncertainty in drought assessments. *Water Resour. Res.* **43**, W07405.

Gonzales, A. L., Nonner, J., Heijkers, J. & Uhlenbrook, S. 2009 Comparison of different base flow separation methods in a lowland catchment. *Hydrol. Earth Syst. Sci.* **13**, 2055–2068.

Gringorten, I. I. 1963 A plotting rule of extreme probability paper. *J. Geophys. Res.* **68** (3), 813–814.

Guo, S. L. 1991 Nonparametric variable kernel estimation with historical floods and paleoflood information. *Water Resour. Res.* **27** (1), 91–98.

Haggag, M. M. M. 2014 New criteria of model selection and model averaging in linear regression models. *Am. J. Theor. Appl. Stat.* **3** (5), 148–166.

Hall, P. 1992 On global properties of variable bandwidth density estimators. *Ann. Statist.* **20** (2) 762–778.

Hannan, E. J. & Quinn, B. G. 1979 The determination of the order of an autoregression. *J. R. Stat. Soc. Series B*. **41**, 190–195.

Hardle, W. 1991 *Smoothing Technique with Implementation in S*. Springer, New York, NY.

Heo, J., Salas, J. D. & Boes, D. C. 2001 Regional flood frequency analysis based on a Weibull model: part 2. simulations and applications. *J. Hydrol.* **242**, 171–182.

Hosking, J. R. M., Wallis, J. R. & Wood, E. F. 1985 Estimation of the general extreme value distribution be the method of probability weighted moments. *Technometrics* **27** (3), 251–261.

Hussain, S. T. P. R. & Ismail, H. 2013 Flood frequency analysis of Kelantan River Basin, Malaysia. *World Appl. Sci. J.* **28** (12), 1989–1995. doi:10.5829/idosi.wasj.2013.28.12.1559.

Jaiswal, R. K., Lohani, A. K. & Tiwari, H. L. 2015 Statistical analysis for change detection and trend assessment in climatological parameters. *Environ. Process.* **2**, 729–749.

Jamaliah, J. 2007 *Emerging Trends of Urbanization in Malaysia*. Acessed from: http://www.statistics.gov.my/eng/images/stories/files/journalDOSM/V104 Article Jamaliah.pdf.

Downloaded from http://iwa.silverchair.com/ws/article-pdf/20/4/1509/705551/ws020041509.pdf
by guest

Jones, M. C., Marron, J. S. & Sheather, S. J. 1996 A brief survey of bandwidth selection for density estimation. *J. Am. Stat. Assoc.* **91**, 401–407.

Karmakar, S. & Simonovic, S. P. 2008 Bivariate flood frequency analysis. Part-1: determination of marginal by parametric and non-parametric techniques. *J. Flood Risk Manage.* **1**, 190–200.

Karmakar, S. & Simonovic, S. P. 2009 Bivariate flood frequency analysis. Part-2: a copula-based approach with mixed marginal distributions. *J. Flood Risk Manage.* **2** (1), 1–13.

Kendall, M. G. 1975 *Rank Correlation Methods*, 4th edn. Charles Griffin, London.

Keshtkaran, P. & Torabihaghighi, A. 2011 Regional flood frequency analysis of Fars Rivers in Iran using new statistical distributions (Case study for Ghareaghaj and Kor Rivers). *Geophys. Res. Abstr.* **13**, 161.

Kim, K. D. & Heo, J. H. 2002 Comparative study of flood quantiles estimation by nonparametric models. *J. Hydrol.* **260**, 176–193.

Kim, T. W., Valdes, J. B. & Yoo, C. 2003 Nonparametric approach for estimating return periods of droughts in arid regions. *J. Hydrol. Eng. ASCE* **8** (5), 237–246.

Kim, T. W., Valdes, J. B. & Yoo, C. 2006 Nonparametric approach for bivariate drought characterisation using Palmer drought index. *J. Hydrol. Eng.* **11** (2), 134–143.

Kullback, S. & Leibler, R. A. 1951 On information and sufficiency. *Ann. Math. Stat.* **22**, 79–86.

Lall, U., Moon, Y.-I. & Khalil, A. F. 1993 Kernel flood frequency estimators: bandwidth selection and kernel choice. *Water Resour. Res.* **29** (4), 1003–1015.

Lall, U., Rajagopalan, B. & Tarboton, D. G. 1996 A nonparametric wet/dry spell model for resampling daily precipitation. *Water Resour. Res.* **32** (9), 2803–2823.

Ljung, G. M. & Box, G. E. P. 1978 On a measure of lack of fit in time series models. *Biometrika* **65**, 297–303.

Mann, H. B. 1945 Nonparametric test against trend. *Econometrics* **13**, 245–259.

Marron, J. S. & Ruppert, T. D. 1994 Transformations to reduce boundary bias in kernel density estimation. *J. R. Stat. Soc., Ser. B* **56**, 653–671.

Miladinovic, B. 2008 *Kernel Density Estimation of Reliability with Applications to Extreme Value Distribution*. Graduate Theses and Dissertations. https://scholarcommons.usf.edu/etd/408.

MMD 2007 *Malaysian Meteorological Department (MMD). Report on Heavy Rainfall That Caused Floods in Kelantan and Terengganu*. Unpublished report. MMD, Kuala Lumpur.

Moon, Y.-I. & Lall, U. 1993 *A Kernel Quantile Function Estimator For Flood Frequency Analysis*. Reports. Paper 194. https://digitalcommons.usu.edu/water_rep/194.

Moon, Y.-I. & Lall, U. 1994 Kernel function estimator for flood frequency analysis. *Water Resour. Res.* **30** (11), 3095–3103.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D. & Veith, T. L. 2007 Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Trans. ASABE* **50** (3), 885–900.

Nashwan, M. S., Ismail, T. & Ahmed, K. 2018 Flood susceptibility assessment in Kelantan river basin using copula. *Int. J. Eng. Technol.* **7** (2), 584–590.

Parzen, E. 1962 On the estimation of a probability density function and mode. *Ann. Math. Stat.* **33**, 1065–1076.

Pettitt, A. N. 1979 A non-parametric approach to the change-point problem. *Appl. Statist.* **28**, 126–135.

Rao, A. R. & Hameed, K. H. 2000 *Flood Frequency Analysis*. CRC Press, Boca Raton, FL.

Reddy, M. J. & Ganguli, P. 2012 Bivariate flood frequency analysis of upper Godavari river flows using Archimedean copulas. *Water Resour. Manage.* **26**, 3995–4018.

Rosenblatt, M. 1956 Remarks on some nonparametric estimates of a density function. *Ann. Math. Stat.* **27**, 832–837.

Salvadori, G. 2004 Bivariate return periods via-2 copulas. *J. Roy. Stat. Soc. Ser. B* **1**, 129–144.

Santhosh, D. & Srinivas, V. 2013 Bivariate frequency analysis of flood using a diffusion kernel density estimator. *Water Resour. Res.* **49**, 8328–8343.

Schwarz, G. E. 1978 Estimating the dimension of a model. *Ann. Stat.* **6** (2), 461–464.

Scott, D. W. & Terrell, G. R. 1987 Biased and unbiased cross-validation in density estimation. *J. Am. Stat. Assoc.* **82** (400), 1131–114.

Shabri, A. 2002 Nonparametric kernel estimation of annual maximum stream flow quantiles. *Matematika* **18** (2), 99–107. *Jabatan Matematik, UTM*.

Sharma, A., Lall, U. & Tarboton, D. G. 1998 Kernel bandwidth selection for a first order nonparametric streamflow simulation model. *Stoch. Hydrol. Hydraul.* **12**, 33–52.

Silverman, B. W. 1986 *Density Estimation for Statistics and Data Analysis*, 1st edn. Chapman and Hall, London.

Tarboton, D. G., Sharma, A. & Lall, U. 1998 Disaggregation procedures for stochastic hydrology based on nonparametric density estimation. *Water Resour. Res.* **34** (1), 107–119.

Veronika, B. M. & Halmova, D. 2014 Joint modelling of flood peak discharges, volume and duration: a case study of the Danube River in Bratislava. *J. Hydrol. Hydromech.* **62** (3), 186–196.

Wan, I. 1996 Urban growth determinants for the state of Kelantano of the state's policy makers. *Penerbitan Akademik Fakulti Kejuruteraan dan Sains Geoinformasi. Buletin Ukur* **7**, 176–189.

Wand, M. P. & Jones, M. C. 1995 *Kernel Smoothing*. Chapman and Hall, London, UK.

Xu, Y., Huang, G. & Fan, Y. 2015 Multivariate flood risk analysis for Wei River. *Stoch. Environ. Res. Risk Assess.* **31** (1), doi:10.1007/s00477-015-1196-0.

Yue, S. 1999 Applying the bivariate normal distribution to flood frequency analysis. *Water Int.* **24** (3), 248–252.

Yue, S. 2000 The bivariate lognormal distribution to model a multivariate flood episode. *Hydrol. Process.* **14**, 2575–2588.

Yue, S. 2001 A bivariate gamma distribution for use in multivariate flood frequency analysis. *Hydrol. Process.* **15**, 1033–1045.

Downloaded from http://iwa.silverchair.com/ws/article-pdf/20/4/1509/705551/ws020041509.pdf
by guest

Yue, S. & Rasmussen, P. 2002 Bivariate frequency analysis: discussion of some useful concepts in hydrological applications. *Hydrol. Process.* **16**, 2881–2898.

Yue, S., Ouarda, T. B. M. J. & Bobee, B. 2001 A review of bivariate gamma distributions in hydrological application. *J. Hydrol.* **246**, 1–18.

Zhang, L. 2005 *Multivariate Hydrological Frequency Analysis and Risk Mapping*. Doctoral dissertation, Beijing Normal University, Beijing, China.

Zhang, S. & Karunamuni, R. J. 1998 On kernel density estimation near endpoints. *J. Stat. Plan. Infer.* **70**, 301–316.

Zhang, L. & Singh, V. P. 2006 Bivariate flood frequency analysis using copula method. *J. Hydrol. Eng.* **11** (2), 150.

Zhang, R., Li, Q., Chow, T. T., Li, S. & Danielescu, S. 2013 Baseflow separation in a small watershed in New Brunswick, Canada, using a recursive digital filter calibrated with the conductivity mass balance method. *Hydrol. Process.* **27**, 2659–2665.

First received 21 November 2019; accepted in revised form 20 April 2020. Available online 5 May 2020

Downloaded from http://iwa.silverchair.com/ws/article-pdf/20/4/1509/705551/ws020041509.pdf
by guest