Hydrol. Earth Syst. Sci. Discuss., 10, 4597–4626, 2013
www.hydrol-earth-syst-sci-discuss.net/10/4597/2013/
doi:10.5194/hessd-10-4597-2013

Hydrology and Earth System Sciences
Open Access
Discussions

Probability distributions for explaining hydrological losses in South Australian catchments

S. H. P. W. Gamage[1], **G. A. Hewa**[1,2], **and S. Beecham**[1,2]

[1]School of Natural and Built Environments, University of South Australia, Adelaide, Australia
[2]Centre for Water Management and Reuse (CWMR), University of South Australia, Adelaide, Australia

Received: 20 March 2013 – Accepted: 20 March 2013 – Published: 10 April 2013

Correspondence to: S. H. P. W. Gamage (sithara.walpitagamage@unisa.edu.au)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

The wide variability of hydrological losses in catchments is due to multiple variables that affect the rainfall-runoff process. Accurate estimation of hydrological losses is required for making vital decisions in design applications that are based on design rainfall models and rainfall-runoff models. Using representative single values of losses, despite their wide variability, is common practice, especially in Australian studies. This practice leads to issues such as over or under estimation of design floods. Probability distributions can be used as a better representation of losses. In particular, using joint probability approaches (JPA), probability distributions can be incorporated into hydrological loss parameters in design models. However, lack of understanding of loss distributions limits the benefit of using JPA.

The aim of this paper is to identify a probability distribution function that can successfully describe hydrological losses in South Australian (SA) catchments. This paper describes suitable parametric and non-parametric distributions that can successfully describe observed loss data. The goodness-of-fit of the fitted distributions and quantification of the errors associated with quantile estimation are also discussed a two-parameter Gamma distribution was identified as one that successfully described initial loss (IL) data of the selected catchments. Also, a non-parametric standardised distribution of losses that describes both IL and continuing loss (CL) data were identified. The results obtained for the non-parametric methods were compared with similar studies carried out in other parts of Australia and a remarkable degree of consistency was observed. The results will be helpful in improving design flood applications.

1 Introduction

Hydrological loss, which has wide temporal and spatial variability, is a crucial parameter in rainfall-runoff (RR) models. In Australia, despite its variability, current practice (Institution of Engineers, 2001) adopts a representative single value of losses as an input parameter for design applications. However, recent studies (Hill, 2010) have highlighted possible errors that can result from using a representative single value in design applications. A Joint Probability Approach (JPA), which provides a joint response of hydrological variables, can be used to overcome the problem associated with models that use representative single values of input parameters. JPAs require a distribution of random variable inputs rather than a measure of central tendency (Rahman et al., 2000, 2002b; Nathan et al., 2003; Kuczera et al., 2006). JPAs that incorporate probabilistic behaviours of the input variables can improve RR simulations (Golian et al., 2012), as well as improve estimation of major flood flows that are required for the design and operation of large water infrastructure (Haddad et al., 2010). In addition, JPAs support more accurate design flood estimations (Haddad and Rahman, 2005; Caballero et al., 2011), streamflow forecasting (Wang et al., 2011) and runoff-yield accounting (Liang et al., 2008).

The most common parameters included in JPAs include initial soil moisture content, rainfall duration and rainfall intensity, and surface runoff (Golian et al., 2012; Singh et al., 2012; Wang et al., 2011; Liang et al., 2008). However, it is useful to incorporate the joint response of initial losses and total losses into rainfall runoff simulation to improve model accuracy (Haddad and Rahman, 2005). Therefore, the probability distributions of these losses need to be identified. It is also useful to present the cumulative frequency distribution of losses as a continuous mathematical equation instead of as a discrete set of data. In order to do that, it is necessary to fit the cumulative frequency distribution of losses to a known cumulative probability distribution function.

In recent years, there has been significant research in Australia on the development and application of the JPAs combined with Monte Carlo Simulation Technique (MCST) for design flood estimation (Haddad and Rahman, 2005). However, the application of a JPA and MCST approach for design flood estimation has so far been limited to gauged catchments with reasonably long rainfall and streamflow records. In practical situations, many catchments are ungauged where there is no or limited data available to identify the probability distributions of various input variables. To apply the

JPA to ungauged catchments, it is necessary to regionalise the distributions of the input variables.

The distribution of losses can be estimated by either parametric or non-parametric methods. In the parametric method, parameters are estimated by equating theoretical moments of the distribution (location, scale and shape) to sample estimated moments such as mean, standard deviation and skewness. Recent studies have used methods such as Method of Moments (MOM), Maximum Likelihood (ML) and Probability Weighted Moments (PWM) to estimate sample estimated moments of a random variable (Trefry et al., 2004; Haktanir et al., 2010). Non-parametric methods, on the other hand do not require a distributional assumption. Nonparametric methods are accurate, uniform, and in particular can provide improved estimates of the distribution tail (Adamowski, 1989).

Some Australian studies report that the parametric method can be used successfully for describing the distribution of hydrological losses. For instance, a four-parameter Beta distribution was used for interpreting initial losses for 10 Victorian catchments (Rahman et al., 2002a). A similar approach was used for describing initial losses (IL) for 15 Queensland catchments (Tularam and Ilahee, 2007). Three probability distributions (Exponential, two-parameter Gamma and four-parameter Beta) were fitted to the observed continuing loss (CL) values for four Victorian catchments (Ishak and Rahman, 2006). A case study which covered three Queensland catchments found that the CL values could be approximated by an exponential function (Ilahee and Rahman, 2003). Recently, a two-parameter Gamma distribution was used to describe the observed IL and CL distributions for five NSW catchments (El-Kafagee and Rahman, 2011). Even though all these studies concluded that the statistics generated using the selected distributions provided a good match to the observed data, none of these distributions can be used with confidence to model hydrological losses in South Australian (SA) catchments unless the catchment of interest is hydrologically similar to the modelled catchments. Further, it is hard to generalise these published results as the case study catchments are not representative of particular climatic zones in Australia.

Non-parametric distributions of loss values are available for certain Australian catchments. For example, IL and CL values of 22 selected Victorian catchments were expressed as a proportion of the median loss value (Nathan et al., 2003). A similar approach was used for 48 rural catchments in Queensland (Ilahee, 2005) and for five catchments in the Darling Ranges in Western Australia (Waugh, 1991). Distributions of the standardised losses of these three studies were found to be largely consistent. This indicates that the shape of the standardised distribution (by median) is the same despite the data being derived from very different hydro-climatic regions from across Australia (Hill, 2010). Therefore, if the median loss rate can be estimated accurately then the standardised distribution can be applied to estimate the distribution of losses for any given catchment. In this study, a similar non-parametric method was tested for modelling hydrological losses of selected SA catchments.

The aim of this paper is to demonstrate the selection of a suitable distribution function for four selected SA catchments using both parametric and non-parametric methods. Identifying suitable parametric distributions for the selected catchments requires following a systematic process. The steps in this process include: (1) calculating IL and CL; (2) Identifying suitable distributions that fit the data; (3) fitting observed data to the selected distribution; (4) simulating data and parameter estimation; and (5) evaluating parameters by calculating bias and mean square error. The non-parametric method used in this study involves standardising losses by the median value. The results of the non-parametric method were compared against other Australian studies.

2 Catchment selection and data

The characteristics considered in selecting the catchments for this study are catchment regulation, size, land-use type and record lengths for available rainfall and streamflow data. The selected catchments were unregulated and had no major land-use changes during the period of their gauge record lengths. As the selected catchments were within the small to medium size range, it can be assumed that the temporal patterns of the

HESSD
10, 4597–4626, 2013

pluviograph data provide representative temporal patterns for the whole catchment (Ilahee, 2005). According to several studies (Boni et al., 2007; Jingyi and Hall, 2004; Kumar and Chatterjee, 2005), the record length of data should be at least 10 years for adequate empirical analysis. The four selected catchments: Scott Bottom (A5030502), Mt Pleasant (A5040512), Yaldara (A5050502) and Penrice (A5050517) all satisfied these conditions. A location map of the selected catchments is given in Fig. 1 and summary details of the geographic, climatic and meteorological data for each catchment are provided in Table 1. The catchment rainfall and streamflow data were collected from the Department for Water, South Australia. After initial quality screening of the time series data, the four stations: A5030502, A5040512, A5050502 and A5050517 were selected because of their high quality data for 37, 35, 47 and 30 yr, respectively.

3 Methodology

3.1 Loss calculation

For each of the selected catchments, rainfall events that produce a reasonable amount of runoff were extracted for this study using the HYDSTRA (KISTERS, 2008) program. The IL, which is defined as the amount of rainfall that occurs before the start of runoff, was calculated using Eq. (1).

$$\mathrm{IL} = \sum_{i=1}^{n} I_i \tag{1}$$

where n is the duration in hours from the start of the storm burst to the start of the surface runoff (rainfall excess) and I_i is rainfall in mm in the ith hour.

Measured streamflow data at a gauged station usually comprises Quickflow (QF) (rainfall excess) and Baseflow (BF) components. For hydrological loss estimations, only the QF is of interest. Therefore, the BF needs to be separated from the original total streamflow data prior to loss calculation. Nathan and McMahon (1990) compared the Lyne and Hollick method of BF separation with several other rigorous algorithms and concluded that it was simple to use, yet produced as good results as the alternatives. Hence, in this study, the Lyne and Hollick algorithm, which is in-built in the HYDSTRA (KISTERS, 2008) program, was used for BF separation.

The Total Rainfall (TR) resulting from a rainfall event can be expressed by Eq. (2) and hence this can be rearranged as in Eq. (3) to calculate the CL, which is defined as the average loss in mm h^{-1} over the remaining duration of the rainfall event.

$$\mathrm{TR} = \mathrm{IL} + \mathrm{CL} \times t + \mathrm{QF} \tag{2}$$

$$\mathrm{CL} = \frac{(\mathrm{TR} - \mathrm{IL} - \mathrm{QF})}{t} \tag{3}$$

where TR, IL and QF are in mm, CL is in mm h^{-1} and t is the time (in h) elapsed between the start of the surface runoff and the end of the rainfall event.

3.2 Parametric method for describing IL

In the parametric approach, sample estimated moments of a random variable are approximated to moments of a known theoretical distribution. In this study, the observed IL data series, $X(1),\ldots,X(n)$ is assumed as the sample of a random variable of interest. Finding a theoretical distribution that can reasonably describe the observed loss data series is now described.

3.2.1 Identifying theoretical distributions

A suitable distribution function that can describe the observed loss data was determined using Quantile–Quantile (Q–Q) plots. A Q–Q plot is a probability plot that compares the quantiles of the observed series against the expected quantiles of a known theoretical distribution. Q–Q plots provide a graphical assessment of "goodness-of-fit" and indicate whether or not the selected sample could have come from the selected target distribution. In this study, Q–Q plots were drawn for the observed IL and CL data series considering certain theoretical distributions, namely Normal, Lognormal, Pareto, Weibull, Gamma and Exponential.

The construction of Q–Q plots consists of a number of steps. The first step is to estimate the quantiles to be plotted. Then the plot has to be constructed in a way that a point (x,y) on the plot corresponds to one of the quantiles of the second distribution (y-coordinate) plotted against the same quantile of the first distribution (x-coordinate) (Ledolter and Hogg, 2010). Thus the line is a parametric curve with the parameter as the interval for the quantile. If one or both of the axes in a Q–Q plot is based on a theoretical distribution with a continuous cumulative distribution function (CDF), all quantiles are uniquely defined and can be obtained by inverting the CDF (Ledolter and Hogg, 2010). However, in this study, the observed data with an unknown distribution is fitted to a theoretical distribution which has potential to fit the data. Therefore, the quantile estimation was done when constructing the Q–Q plots. The quantile estimations for constructing the Q–Q plots was undertaken using the plotting position formula given in Eq. (4).

$$\frac{i-0.5}{n} = f(q_i) \tag{4}$$

where n is the sample size, i is the particular sample and $f(q_i)$ is the quantile of the observed data.

The x-axis of the Q–Q plots consists of order statistics, $x(1) \leq x(1) \leq \ldots \leq x(n)$ with a theoretical CDF, $\mathrm{F}(x) = \mathrm{P}(X \leq x)$. The y-axis of the Q–Q plots consists of quantiles of observed data.

After constructing the Q–Q plots, the theoretical distribution that provides a reasonable fit can be selected. The selected distribution then needs to be further investigated to assess possible estimation errors.

HESSD
10, 4597–4626, 2013

The Q–Q plot analyses confirms that of the five distributions investigated, the Gamma distribution is the most appropriate to describe loss data out. The methodology of the parametric modelling is now described with particular reference to the Gamma distribution. The sequential steps involved in the parametric modelling are presented in Fig. 2.

3.2.2 Fitting and testing the gamma distribution

Gamma distribution

This is a two-parameter continuous probability distribution function (PDF) with a shape parameter (k) and a scale parameter (θ). The shape parameter can also be denoted by $\alpha = k$ and an inverse scale parameter denoted by $\beta = 1/\theta$ (Freund and Johnson, 2010).

The PDF of the Gamma distribution is given in Eq. (5) and the Gamma function $\Gamma(\alpha)$ is given in Eq. (6).

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta} \tag{5}$$

where $0 \le x \le \infty$, and parameters $\alpha > 0$ and $\beta > 0$.

$$\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx \tag{6}$$

$\Gamma(\alpha)$ is a generalized factorial that can be shown as $\Gamma(\alpha) = (\alpha-1)!$, if α is a positive integer. The Gamma function for the arguments (α) between 0 and 1 can be found from standard mathematical tables.

HESSD
10, 4597–4626, 2013

Inverse Gamma distribution

The inverse gamma distribution is a two-parameter family of continuous probability distributions on the positive real line, which is the distribution of the reciprocal of a variable distributed according to the Gamma distribution. The inverse Gamma distribution's PDF is given in Eq. (7).

$$f(x;\alpha,\beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)}(x)^{-\alpha-1}\exp\left(-\frac{\beta}{x}\right) \tag{7}$$

where $x > 0$ and α and β are the shape and scale parameters, respectively.

Parameter estimation

Once the distribution is selected, parameters have to be estimated. The most commonly used methods for determining parameters of the PDFs include Method of Moments (MOM), Maximum Likelihood (ML) and Probability Weighted Moments (PWM). In this study, MOM was adopted to estimate parameters because of its simplicity and ease of use. Estimation of distribution parameters involves equating theoretical moments of the distribution to the sample estimated moments. For the Gamma distribution, the first two theoretical moments are given in Eqs. (8) and (9) (Freund and Johnson, 2010).

$$\mu = \alpha\beta \tag{8}$$

$$\sigma^2 = \sigma\beta^2 \tag{9}$$

HESSD
10, 4597–4626, 2013

Replacing μ and σ^2 in Eqs. (8) and (9) by sample estimated $\bar{x}$ and s^2, the distribution expressions for the Gamma parameters are given in Eqs. (10) and (11).

$$\hat{\alpha} = \frac{\bar{x}^2}{s^2} \tag{10}$$

$$\hat{\beta} = \frac{s^2}{\bar{x}} \tag{11}$$

3.2.3 Simulating data

Although the true distribution of loss data is not known, it is still interesting to understand the errors associated with the estimated quantiles, when data are derived from a Gamma distribution. Hence, in this study, the estimated Gamma distribution using the observed data series was assumed to be the true distribution, and quantiles estimated from the observed data were assumed to be the true quantiles in estimating bias and MSE associated with the estimated quantiles. The estimated quantiles in this study correspond to non-exceedance probabilities of 0.001, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 and 0.999. Probability values of 0 and 1 were intentionally avoided to eliminate mathematical errors when applying the Gamma inverse function. With the use of a uniform random number generator (discussed in the next section), 500 simulated Gamma samples of losses of the same size as the observed series were generated. For each of the generated samples, quantiles at 10 selected non-exceedance probabilities were calculated. Hence, the simulated quantile estimation involved: (1) calculating Gamma parameters for each of the 500 simulated series; (2) applying a Gamma inverse function for the series in step 1 to obtain simulated loss quantiles at each selected non-exceedance probability; and (3) estimating 500 simulated loss quantiles for each non-exceedance probability. The sequence of these steps are shown in Fig. 2.

HESSD
10, 4597–4626, 2013

Random number generation

A reliable source of random numbers, and a means of transforming them into prescribed distributions, is essential for the success of the simulation approach. Generated random numbers, x, should belong to a domain, $x \in [x_{\min}, x_{\max}]$, in such a way that the frequency of occurrence (or probability density) will depend upon the value of x in a prescribed functional form $f(x)$ (Saucier, 2000). Among the various techniques available for generating random numbers, most of the methods presume that a supply of uniformly distributed random numbers are in the half-closed unit interval [0, 1) (Saucier, 2000). The methods for random number generation include Inverse Transformation, Composition, Convolution, Acceptance–Rejection, Sampling and Data–Driven Techniques, Techniques Based on Number Theory and Monte Carlo Simulation (Saucier, 2000). If the inverse form of a distribution function (F^{-1}) is not available, then the inverse transformation technique is not feasible and other techniques need to be considered. However, in this study, the inverse transformation technique can be used because the inverse Gamma function is available. Also, the inverse transformation is a simple, efficient and commonly used technique (Saucier, 2000).

3.2.4 Evaluating quantiles

As mentioned earlier, the estimated quantiles were evaluated by comparing bias and mean square error (MSE). The variance of the estimator was also calculated as the difference between MSE and bias estimation. In addition, probability plots were investigated for evaluating the estimated quantiles and to determine the validity of the Gamma distribution for fitting to extreme loss values.

Bias

The bias (or bias function) of an estimator is the difference between expected value of the estimator and the true value of the parameter being estimated, and is given in

HESSD
10, 4597–4626, 2013

Eq. (12).

$$\text{Bias} = E\left[\hat{\theta} - \theta\right] = E\left[\hat{\theta}\right] - \theta \tag{12}$$

where $E[]$ denotes the expected value over the distribution $P(x|\theta)$ (Lebanon, 2010).

$P(x|\theta)$ is a probability distribution for observed data x, with parameter θ. Statistic $\hat{\theta}$ serves as an estimator of θ based on any observed data x. In other words, it is assumed that the data follows some unknown distribution $P(x|\theta)$ (where θ is a fixed constant, which is part of this distribution, but is unknown). And then an estimator $\hat{\theta}$, which maps observed data to values that are expected to be close to θ, is constructed. An estimator is said to be unbiased if its bias is equal to zero for all the values of parameter θ (Lebanon, 2010).

Mean square error (MSE)

There are more important performance characterizations for an estimator than just being unbiased. The mean squared error, which captures the error that the estimator makes, is perhaps the most important of these (Lebanon, 2010). The MSE is a measure of the variance of error in the quantile estimator and is used to give an overall measure of accuracy. The MSE thus assesses the quality of an estimator in terms of its variation and unbiasedness.

Since MSE is an expectation, it is not a random variable. It may be a function of the unknown parameter θ, but it does not depend on any random quantities. However, when MSE is computed for a particular estimator of θ, the true value of which is not known, it will be subjected to estimation error (Lebanon, 2010). If there are many repeated samplings of $X(1), \ldots, X(n)$, it is necessary to average over the distribution thus capturing the average performance. The mean square error (MSE) of an estimator is

HESSD
10, 4597–4626, 2013

Title Page
Abstract | Introduction
Conclusions | References
Tables | Figures
◂◂ | ▸▸
◂ | ▸
Back | Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

given in Eq. (13).

$$E\left(\left\|\hat{\theta}-\theta\right\|^2\right) = E\left(\sum_{j=1}^{d}\left(\hat{\theta}_j-\theta_j\right)^2\right) = \left[\text{bias}\left(\hat{\theta}\right)\right]^2 + \text{var}\left(\hat{\theta}\right) \tag{13}$$

Confidence interval

In this study, the level of confidence is set as 95 % which reflects a significance level of 0.05. For the Gamma distribution, exact confidence intervals are difficult to construct and the available methods for finding confidence interval of the Gamma distribution are very complex (Fay and Feuer, 1997; Banneheka, 2012). Therefore, in this study, a simple approximation was undertaken to determine the CI. The Upper Confidence Level (UCL) and Lower Confidence Level (LCL) were calculated as UCL $= f(x_\mathrm{u})$ and LCL $= f(x_\mathrm{l})$, respectively where $x_\mathrm{u} = 0.95n$, $x_\mathrm{l} = 0.05n$ and n is the sample size. As the study generated 500 simulated series, $n = 500$.

3.3 Non-parametric method

A non-parametric method, as the name implies, has no dependency on parameters. Non-parametric plots were drawn for all the catchments with the y-axis as the standard storm IL and CL (fraction of the median) and the x-axis as the proportion of the sample value exceeded (%). The plots were also compared with similar studies carried out for other parts of Australia.

4 Results and discussion

4.1 Parametric method

One objective of this study is to identify a suitable parametric distribution that can describe IL and CL data of the selected SA catchments. The two-parameter Gamma

HESSD
10, 4597–4626, 2013

distribution was selected from six theoretical distributions using Q–Q plots. Figure 3 shows the Gamma Q–Q plots for the observed IL data for the four selected catchments. Parameters of the Gamma distribution that were calculated for the observed set of IL data are presented in Table 2.

In all four catchments, the higher values of the losses deviate from the $y = x$ line, which means that extreme loss values deviate from the Gamma distribution. It is necessary to determine how much these extreme values deviate from the Gamma distribution and this will be explained by analysing probability plots, later in this discussion.

Although the observed CL data were also checked for the same distributions, none of the Q–Q plots for the four catchments followed the $y = x$ line.

The IL quantiles that were calculated for the selected non-exceedence probabilities are presented in Table 3. It can be seen from Table 3 that the observed and simulated IL values for each non-exceedence probability are within the 95 % confidence intervals. The last column of the table, which provides the range of IL values, was generated considering both the observed and simulated data. These ranges with their non-exceedence probabilities are useful for design applications.

Probability plots (PDFs for the observed, fitted and simulated IL data for each catchment) were used to demonstrate that the data simulated using the two-parameter Gamma distribution match well with the observed data. The probability plots for the selected catchment are shown in Fig. 4 with a randomly selected simulated sample. Evaluating the fitted curves and observed data, it can be concluded that the IL data follow the two-parameter Gamma distribution very well. In addition, these probability plots show that the simulated and the observed data are very close. Although there is a deviation of fitted, observed and simulated series in their high values, the difference is very small.

As mentioned earlier, the Q–Q plots indicate that extreme loss values do not closely follow Gamma distribution as other data values. However, the probability plots indicate that the gap between the fitted (theoretical line) and simulated values is very small. Therefore the two-parameter Gamma distribution is not limited to low and medium ranges of loss values, but can also be used to represent the high values. In addition, the error caused by the quantile estimation and simulation should also be quantified. Although the probability plots provide certain estimation of the error associated with the simulation, the errors have been further quantified using other methods, namely bias and MSE of the estimator.

The bias and MSE for the four selected catchments are presented in Fig. 5. Because all the values are close to zero, it can be concluded that the IL simulated using a two-parameter Gamma distribution is accurate and very close to observed data. However, the values of bias and MSE can change slightly according to the different simulated samples selected.

After comparing a number of samples (observed data series with different randomly selected simulated data series), it was observed that the bias and MSE are comparatively lower in low non-exceedence probabilities. The bias and MSE have a relatively wide range in the higher non-exceedence probabilities relative to low non-exceedence probabilities. Despite this, the bias and MSE values are always close to zero. Therefore it can be assumed that the range of IL and its probability of occurrence as shown in Table 3, can be used in design applications. However those values should only be used for hydrologically similar catchments. In order to use the results for all the catchments in SA, the parameters need to be generalised. Generalised parameters can be determined by fitting samples of loss values derived from all the catchments in the region. However, parameter generalisation is not within the scope of this paper.

4.2 Non-parametric method

A non-parametric method of describing losses was also developed. The non-parametric method used in this study involves standardising both IL and CL with their median values. Median values have been used for standardisation in similar studies conducted in other parts of the Australia including South Eastern Australian Catchments (Hill et al., 1996), Queensland catchments (Ilahee 2005) and South Western

HESSD
10, 4597–4626, 2013

Australian catchments (Waugh, 1990, 1991). Using the median values for the standardisation in this study allows the distributions of losses across the different catchments to be directly compared, as shown in Fig. 6. Figure 6a shows the distribution of standardised IL while Fig. 6b shows the distribution of standardised CL. Both Fig. 6a and b indicate that both IL and CL have consistency when standardized by median values. They are also consistent with similar studies conducted for other regions of Australia (Nathan and Weinmann, 2004; Ilahee, 2005; Nathan et al., 2003; Waugh, 1991). In particular, Nathan et al. (2003) show that the shape of the standardised IL distribution does not change with location. Thus despite the data being derived from very different hydro-climatic regions across Australia, the results clearly show that while the magnitude of losses may vary between different catchments, the shape of the distribution does not.

In addition, it can be concluded that the variation of IL and CL values across the four catchments are higher for the proportion of sample exceedence that are less than 10 %. This variation is higher for the CL than the IL values. In this study, the standardised CL varied from 2 to 45, and IL values varied from 0 to 8 while in other similar studies (Nathan and Weinmann, 2004; Ilahee, 2005; Nathan et al., 2003; Waugh, 1991) CL varied from 1 to 14 and IL varied from 0 to 8 (Nathan and Weinmann, 2004; Ilahee, 2005; Nathan et al., 2003; Waugh, 1991). The variance of CL can be reduced by excluding the high outliers, however this can be very subjective.

5 Conclusions

This paper investigates both parametric and non-parametric methods to describe hydrological losses. The two-parameter Gamma distribution was successfully fitted for observed IL data. For each catchment, the parameters were estimated and the IL values were simulated from the two-parameter Gamma distribution. The simulated data compare very well with the observed data, with some tendency to overestimate the

HESSD
10, 4597–4626, 2013

occurrence of higher losses. The parameters and CDF of the Gamma distribution can be used to find the frequency distribution and can be used to estimate the probability of occurrence of IL in design applications. This can particularly improve the joint probability approaches of design flood applications. However, for the CL component, none of the parametric distributions seems to fit the observed data satisfactorily. The non-parametric method tested, which is the standardised distribution of both IL and CL over median values, exhibits a remarkable degree of consistency with other studies. These standardized values can therefore be used in design applications.

References

Adamowski, K.: Monte Carlo comparison of parametric and nonparametric estimation of flood frequencies, J. Hydrol., 108, 295–308, doi:10.1016/0022-1694(89)90290-4, 1989.

Banneheka, B.: Confidence intervals for the median of a gamma distribution, Vidyodaya Journal of Science, 15, 123–129, 2012.

Bias, Variance, and MSE of Estimators: availavle at: http://www.cc.gatech.edu/~lebanon/notes/estimators1.pdf (last access: 1 April 2013), 2010.

Boni, G., Ferraris, L., Giannoni, F., Roth, G., and Rudari, R.: Flood probability analysis for ungauged watersheds by means of a simple distributed hydrologic model, Adv. Water Resour., 30, 2135–2144, 2007.

Caballero, W. L., Taylor, M., Rahman, A., and Shrestha, S.: Regionalisation of intensity-frequency-duration data: A case study for New South Wales, in: MODSIM 2011 – 19th International Congress on Modelling and Simulation – Sustaining Our Future: Understanding and Living with Uncertainty, MODSIM2011, 12–16 December 2011, Perth, WA, Australia, 3775–3781, 2011.

El-Kafagee, M. and Rahman, A.: A study on initial and continuing losses for design flood estimation in New South Wales, in: MODSIM 2011 – 19th International Congress on Modelling and Simulation – Sustaining Our Future: Understanding and Living with Uncertainty, MODSIM2011, 12–16 December 2011, Perth, WA, Australia, 3782–3787, 2011.

Fay, M. P. and Feuer, E. J.: Confidence intervals for directly standardized rates: a method based on the gamma distribution, Stat. Med., 16, 791–801, 1997.

HESSD

10, 4597–4626, 2013

Probability distributions for explaining hydrological losses

S. H. P. W. Gamage et al.

Freund, J. and Johnson, R. A.: Miller and Freund's probability and statistics for engineers, Pearson Education (US), Upper Saddle River, NJ 07458, 2010.
Golian, S., Saghafian, B., and Farokhnia, A.: Copula-based interpretation of continuous rainfall-runoff simulations of a watershed in Northern Iran, Can. J. Earth Sci., 49, 681–691, doi:10.1139/e2012-011, 2012.
Haddad, K. and Rahman, A.: Regionalisation of rainfall duration in Victoria for design flood estimation using Monte Carlo simulation, in: Proceedings of MODSIM05 – International Congress on Modelling and Simulation: Advances and Applications for Management and Decision Making, MODSIM05, 12–15 December 2005, Melbourne, VIC, Australia, 1827–1833, 2005
Haddad, K., Rahman, A., and Weinmann, P. E.: Estimation of major floods: applicability of a simple probabilistic model, Aust. J. Water Res., 14, 117–126, 2010.
Haktanir, T., Cobaner, M., and Kisi, O.: Frequency analyses of annual extreme rainfall series from 5 min to 24 h, Hydrol. Process., 24, 3574–3588, doi:10.1002/hyp.7759, 2010.
Hill, P. I.: Project 6: Loss models for catchment simulation, Engineers Australia, ACT, Project report, 11 National Circuit Barton ACT 2600, 2010.
Institution of Engineers: Australian rainfall and runoff: a guide to flood estimation, Rev. ed. 1987, 3rd Edn., edited by: Pilgrim, D. H., Institution of Engineers, Australia, Barton, ACT, 2001.
Ilahee, M.: Modelling Losses in Flood Estimation, Ph.D. thesis, School of Urban Development Queensland University of Technology, Queensland University of Technology, Queensland, 2005.
Ilahee, M. and Rahman, A.: Investigation of Continuing Losses for Design Flood Estimation: A Case Study in Queensland, in: Proceedings of the 28th International Hydrology and Water Resources Symposium: About Water, 10–13 November 2003, Wollongong, NSW, Australia, 125–130, 2003.
Ishak, E. and Rahman, A.: Investigation into probabilistic nature of continuing loss in four catchments in Victoria, 30 Hydrology & Water Resources Symposium: Past, Present & Future, Hotel Grand Chancellor, Launceston, 4–7 December 2006, 6–1, 2006.
Jingyi, Z. and Hall, M.: Regional flood frequency analysis for the Gan-Ming River basin in China, J. Hydrol., 296, 98–117, 2004.
KISTERS: HYDSTRA, KISTERS North America, Inc, available at: http://www.kisters.com.au/english/html/au/homepage.html (last access: 2 April 2013) 2008.

Kuczera, G., Kavetski, D., Franks, S., and Thyer, M.: Towards a Bayesian total error analysis of conceptual rainfall-runoff models: characterising model error using storm-dependent parameters, J. Hydrol., 331, 161–177, 2006.
Kumar, R. and Chatterjee, C.: Regional flood frequency analysis using L-moments for North Brahmaputra region of India, J. Hydrol. Eng., 10, 1–7, 2005.
Ledolter, J. and Hogg, R. V.: Applied Statistics for Engineers and Physical Scientists, Prentice Hall, Pearson, 2010.
Liang, Z., Wang, J., Lei, Y., and Shi, Y.: A Statistically-Based Runoff-Yield Model, in: IAHS-AISH Publication, Proceedings of the Chinese Prediction in Ungauged Basins (PUB) International Symposium 2006, 1 September 2008–1 September 2008, Beijing, China, 88–94,2008.
Nathan, R. and McMahon, T.: Evaluation of automated techniques for base flow and recession analyses, Water Resour. Res., 26, 1465–1473, 1990.
Nathan, R. and Weinmann, P.: An improved framework for the characterisation of extreme floods and for the assessment of dam safety, in: Proceedings of the British Hydrological Society International Conference, Imperial College, London, July, 186–193, 2004.
Nathan, R., Weinmann, E., and Hill, P.: Use of Monte Carlo Simulation to Estimate the Expected Probability of Large to Extreme Floods, in: Proceedings of the 28th International Hydrology and Water Resources Symposium: About Water, 10–13 November, Novotel North beach, Wollongong, NSW, Australia, 1.105–1.112, 2003.
Rahman, A., Weinmann, P. E., and Mein, R. G.: The probabilistic nature of initial losses and its impact on design flood estimation, Hydrology and Water Resources Symposium, Perth, Australia, 20–23 November, Sheraton Perth Hotel, Perth, Western Australia, 71–75, 2000.
Rahman, A., Weinmann, E., and Mein, R.: The use of probability-distributed initial losses in design flood estimation, Aust. J. Water Res., 6, 17–29, 2002a.
Rahman, A., Weinmann, P., Hoang, T., and Laurenson, E.: Monte Carlo simulation of flood frequency curves from rainfall, J. Hydrol., 256, 196–210, 2002b.
Saucier, R.: Computer generation of statistical distributions, DTIC Document, Army Research Laboratory, available at: http://ftp.arl.mil/random/random.pdf (last access: 2 April 2013), 2000.
Singh, V. P., Zhang, L., and Rahimi, A.: Probability distribution of rainfall-runoff using entropy theory, T. ASABE, 55, 1733–1744, 2012.
Trefry, C. M., Watkins, D. W., and Johnson, D. L.: Development of regional rainfall intensity-duration-frequency estimates for the state of Michigan, in: Joint Conference on Water

HESSD
10, 4597–4626, 2013

Resource Engineering and Water Resources Planning and Management 2000: Building Partnerships, 30 July–2 August 2000, Minneapolis, MN, United States, 2004.
Tularam, G. A. and Ilahee, M.: Initial loss estimates for tropical catchments of Australia, Environ. Impact Asses., 27, 493–504, doi:10.1016/j.eiar.2006.12.006, 2007.
Wang, E., Zhang, Y., Luo, J., Chiew, F. H. S., and Wang, Q. J.: Monthly and seasonal streamflow forecasts using rainfall-runoff modeling and historical weather data, Water Resour. Res., 47, W05516, doi:10.1029/2010wr009922, 2011.
Waugh, A.: Design losses in flood estimation, International Hydrology and Water Resources Symposium, Perth, October 1991, 629–630, 1991.

HESSD
10, 4597–4626, 2013

Table 1. Geographical, climatic and meteorological data.

Characteristics	Scott Bottom (A5030502)	Mt Pleasant (A5040512)	Yaldara (A5050502)	Penrice (A5050517)
River	Scott Creek	Torrens	North Para	North Para
Area (km^2)	27	26	384	118
Annual Rainfall (mm)	69–74	80–103	93–102	63–67
Elevation at gauging station (m)	205	415	145	285
Evaporation ($mm\,day^{-1}$)	1.38	1.69	1.69	1.69

HESSD
10, 4597–4626, 2013

Table 2. Estimated Gamma distribution parameters.

Station no	A5040512	A5030502	A5050502	A5050517
Shape parameter (k)	1.798	2.714	2.015	1.450
Scale parameter (θ)	0.142	0.153	0.137	0.094

HESSD
10, 4597–4626, 2013

Table 3. Estimated IL quantiles for selected non-exceedance probabilities.

$P(x)$	A5030502 obs(x)	A5030502 sim(x) LCL-UCL	A5040512 obs(X)	A5040512 sim(X) LCL-UCL	A5050502 obs(x)	A5050502 sim(x) LCL-UCL	A5050517 obs(x)	A5050517 sim(x) LCL-UCL	Range
0.01	2.53	1.52–3.50	0.85	0.46–1.21	1.39	0.78–2.11	0.56	0.46–0.96	1–3
0.1	6.53	4.49–7.90	3.24	3.03–5.31	4.48	3.43–5.92	2.94	2.02–3.29	3–7
0.2	9.15	7.70–10.01	5.11	4.09–5.67	6.73	6.17–7.92	5.12	4.07–5.71	5–9
0.3	11.46	10.00–12.65	6.88	6.05–9.27	8.79	8.02–9.81	7.31	6.24–7.89	7–11
0.4	13.72	13.66–16.00	8.71	7.67–11.87	10.87	10.79–11.82	9.64	7.56–9.92	9–14
0.5	16.10	15.03–16.90	10.69	9.64–11.34	13.10	14.01–16.92	12.26	9.16–13.01	11–16
0.6	18.73	17.67–19.61	12.96	11.90–13.82	15.61	15.53–20.32	15.32	11.21–16.92	13–19
0.7	21.86	20.80–23.59	15.73	14.66–17.09	18.64	20.56–25.01	19.12	14.00–21.72	16–22
0.8	25.93	25.88–27.01	19.42	17.34–21.03	22.65	25.57–33.21	24.29	19.17–26.37	19–26
0.9	32.34	29.31–33.46	25.40	23.32–28.71	29.07	29.01–33.21	32.85	30.72–37.84	25–33
0.99	51.34	45.37–58.32	43.87	40.77–50.92	48.57	40.59–55.92	59.99	49.90–72.13	44–60

HESSD
10, 4597–4626, 2013

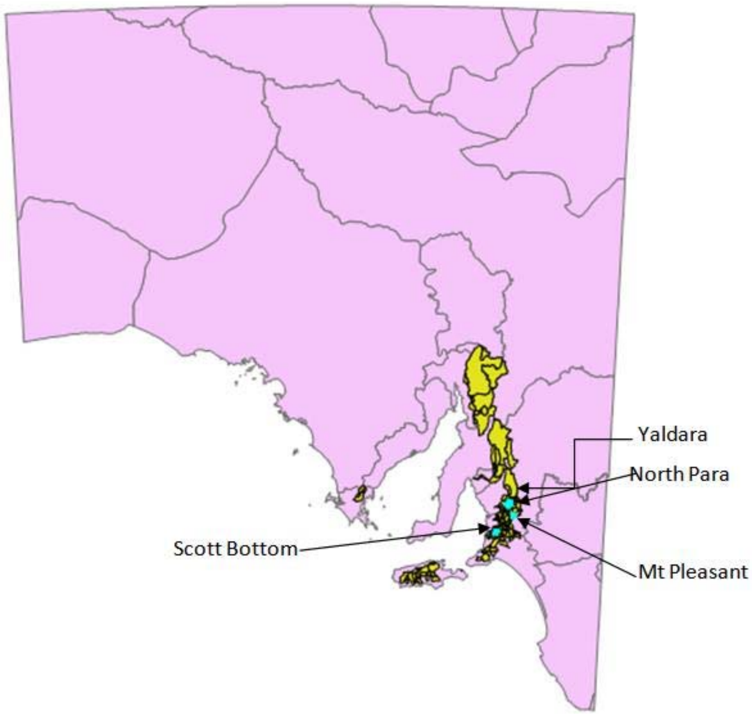

Fig. 1. Location map of the study area.

HESSD
10, 4597–4626, 2013

Probability distributions for explaining hydrological losses

S. H. P. W. Gamage et al.

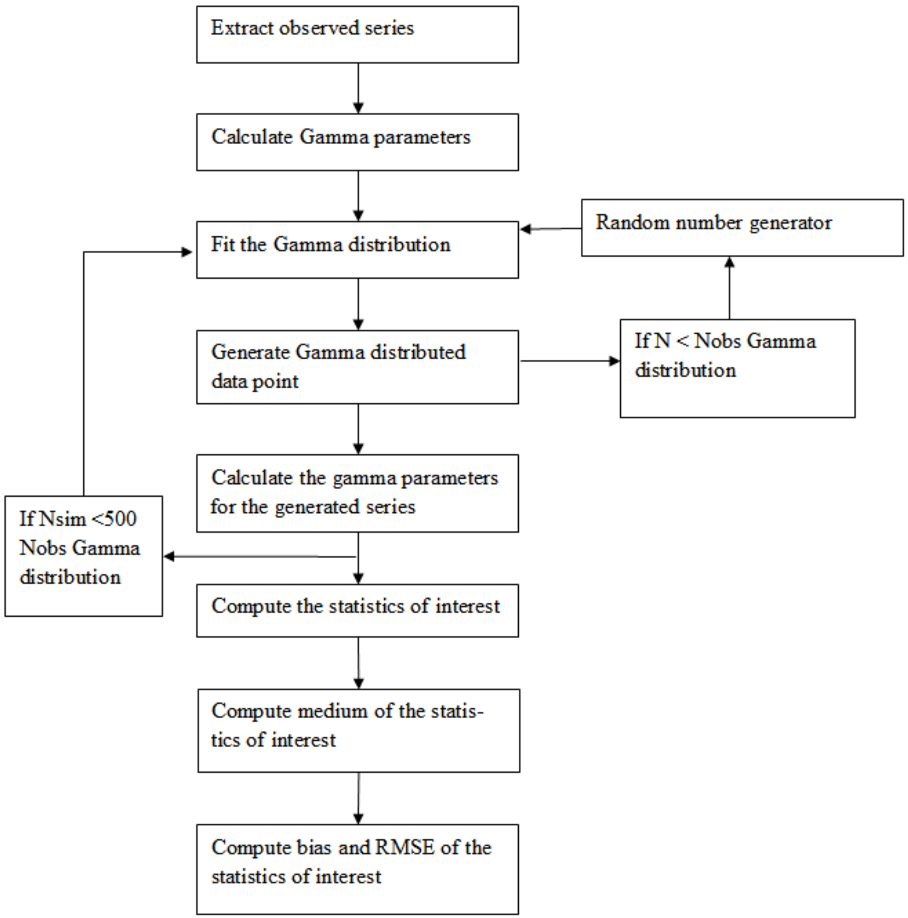

Fig. 2. Flowchart for data fitting, simulation and evaluation. Nobs – Number of observations; Nsim – Number of simulations; MSE – Mean square error.

HESSD
10, 4597–4626, 2013

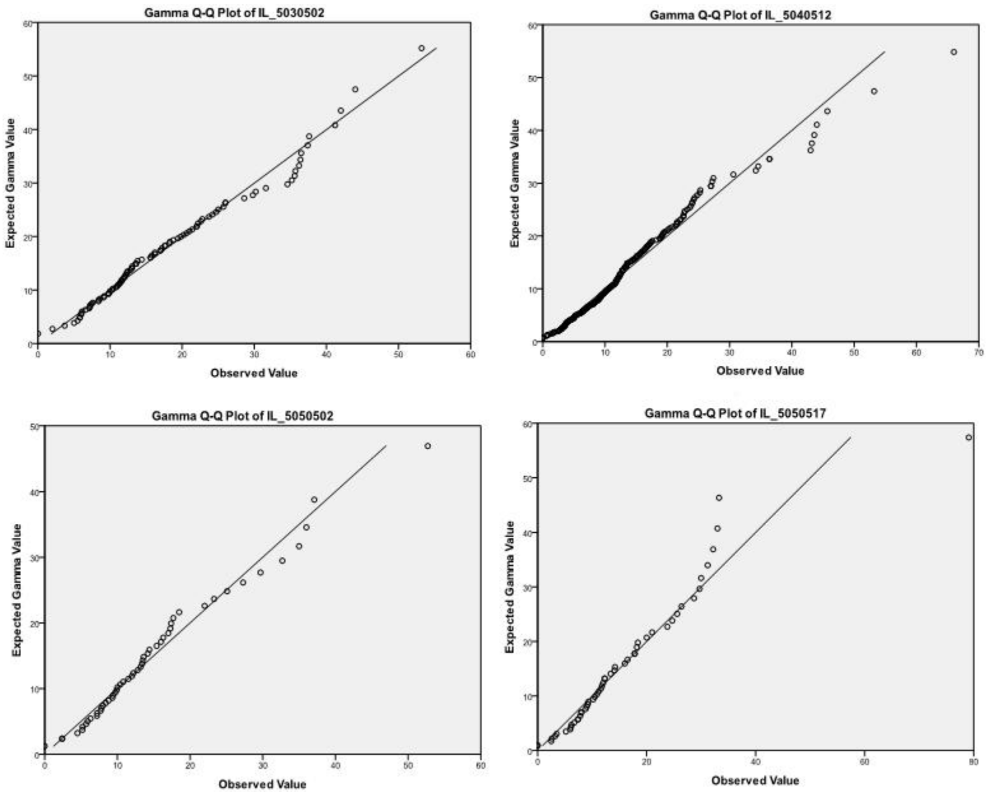

Fig. 3. Gamma Q–Q plots for the observed IL data.

HESSD
10, 4597–4626, 2013

Probability distributions for explaining hydrological losses

S. H. P. W. Gamage et al.

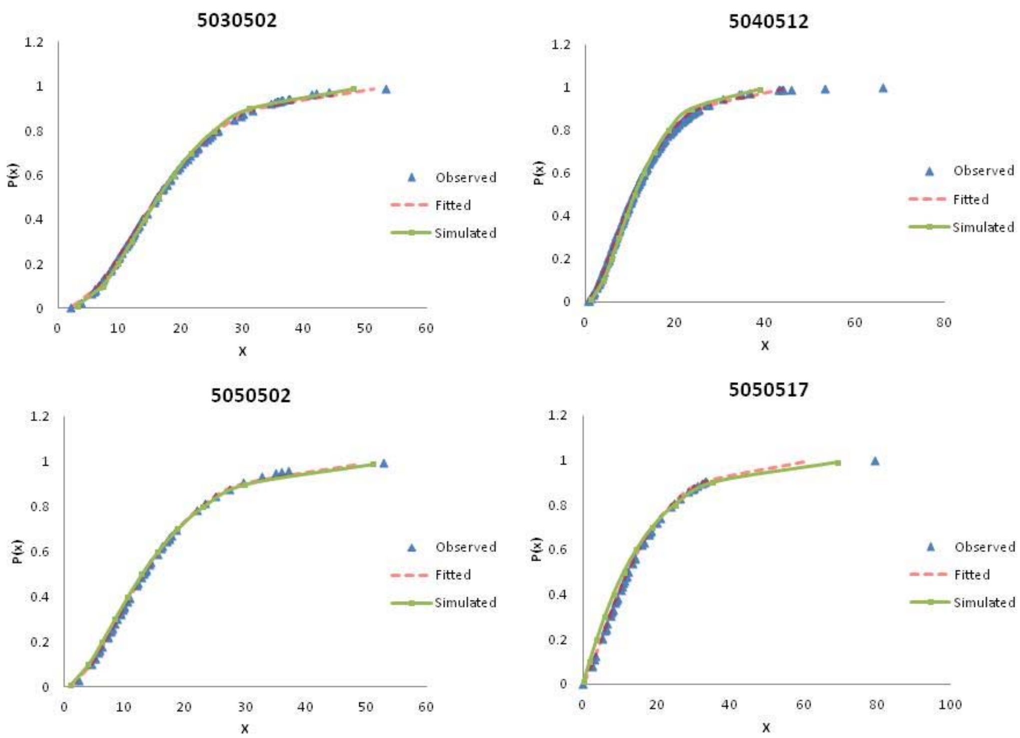

Fig. 4. Probability plots of the study catchments.

HESSD
10, 4597–4626, 2013

Probability distributions for explaining hydrological losses

S. H. P. W. Gamage et al.

Title Page
Abstract | Introduction
Conclusions | References
Tables | Figures
◀| |▶
◀ ▶
Back | Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

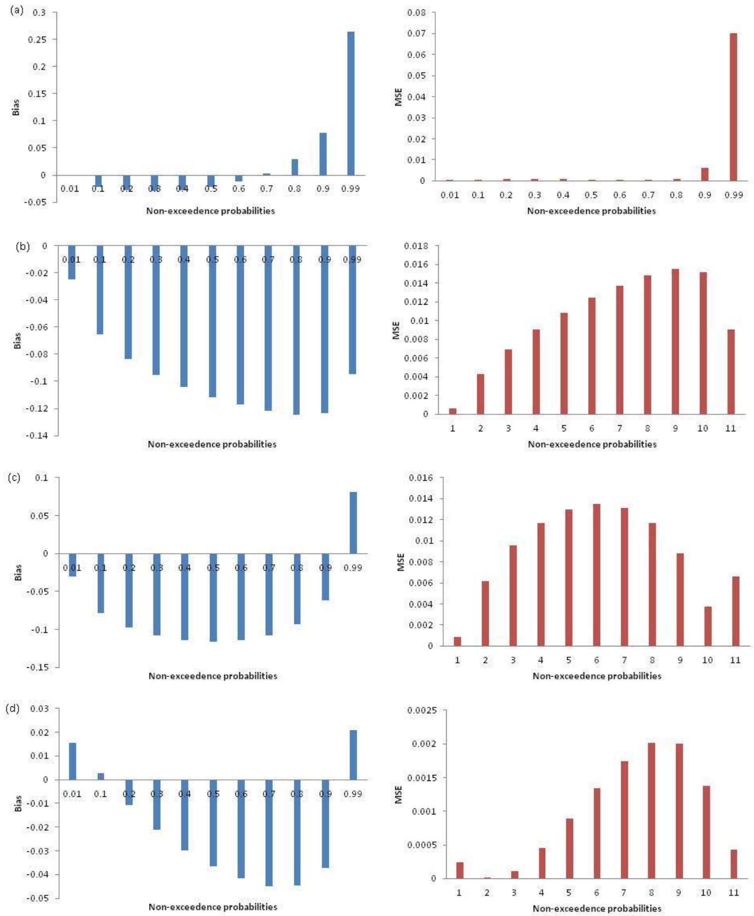

Fig. 5. Bias and MSE for simulated samples for catchments **(a)** A5040512 **(b)** A5030502 **(c)** A5050502 **(d)** A5050517.

HESSD
10, 4597–4626, 2013

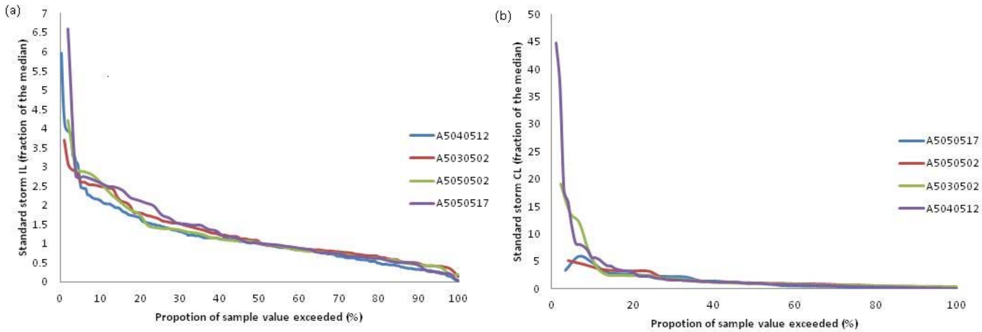

Fig. 6. IL and CL standardised by median values

HESSD
10, 4597–4626, 2013