Design flood estimation using model selection criteria

Giuliano Di Baldassarre [a,*], Francesco Laio [b], Alberto Montanari [c]

[a] School of Geographical Sciences, University of Bristol, University Road, Bristol BS8 1SS, United Kingdom
[b] Dipartimento di Idraulica, Trasporti ed Infrastrutture Civili, Politecnico di Torino, Torino, Italy
[c] Dipartimento di Ingegneria delle Strutture, dei Trasporti, delle Acque, del Rilevamento, del Territorio, Università degli Studi di Bologna, Bologna, Italy

ARTICLE INFO

ABSTRACT

The design flood is defined as the discharge value corresponding to an assigned non-exceedance probability, which is usually expressed in terms of the return period. Estimation of the design flood is usually carried out by fitting observed data samples with a suitable probability distribution. The objective of this study is to evaluate if model selection criteria, which are seldom used in hydrological applications, can help identifying the best probability model for this purpose. The study analyzes the performance of three model selection criteria, namely, the Akaike information criterion, the Bayesian information criterion, and the Anderson–Darling criterion. The three methods are compared trough an extensive numerical analysis by using synthetic data samples. The study demonstrates that model selection criteria are a valuable tool for reducing the uncertainty of design flood estimation.

1. Introduction

The design flood (or design event) is the discharge value corresponding to an assigned non-exceedance probability, usually expressed in terms of the return period (e.g. Stedinger et al., 1992). The estimation of the design flood at gauged stations requires the choice and parameterization of an appropriate probabilistic model. Many probability distributions have been considered, in different situations, for this purpose. Classical examples include the Generalized Extreme Value, Gumbel and Frechet distributions, as well as the Gamma distribution and many others less frequently used (e.g. Kottegoda and Rosso, 1998).

Objective techniques for the choice of the most appropriate probabilistic model are needed to improve design flood estimation. The general issue of model selection was first studied by Akaike (1973), who introduced the principle of maximum entropy as the theoretical basis for model selection, and by Schwarz (1978) who, by developing a similar idea in a Bayesian context, proposed the Bayesian Information Criterion for model selection. Extensions of these methods include corrections to be used with small sample size (Hurvich and Tsai, 1989) and other generalizations (e.g., see Bozdogan, 1987; Konishi and Kitagawa, 1996; Wasserman, 2000). In the recent years, some applications of model selection criteria within the frequency analysis of hydrological extremes have been proposed. Mutua (1994), Hache et al. (1999), Strupczewski et al.

(2001, 2002) and Cahill (2003) applied the Akaike Information Criterion to single case studies, without however analyzing its properties for small samples. Moreover, a comparative evaluation with other criteria was not performed. Mitosek et al. (2006) applied three model discrimination procedures in order to identify the best fitting probability distribution. Nevertheless, the study considered only two-parameter distributions. Therefore, the ability of the model selection techniques to properly account for the principle of parsimony (Box and Jenkins, 1970) was not assessed.

In hydrology, the choice of the probabilistic model is often based on the use of L-moments plots (Vogel et al., 1993a,b; Hosking and Wallis, 1997) that can provide indications about the fitting capabilities of two-parameter (e.g., Di Baldassarre et al., 2006a) and three-parameter distributions (e.g., Onoz and Bayazit, 1995; Di Baldassarre et al., 2006b). Such approach, however, is not fully objective, because the goodness-of-fit provided by a distribution is often based only upon graphic visualization. Pandey et al. (2001) and Kroll and Vogel (2002) developed performance measures to overcome the problem induced by the subjective interpretation of the L-moment diagrams. However, how to compare the descriptive ability of distributions with different number of parameters still remains an open question. For example, two and three-parameter distributions are separately compared by Kroll and Vogel (2002).

Laio et al. (2008) carried out a systematic analysis of the performance of three model selection criteria, namely, the Akaike information criterion (AIC), the Bayesian information criterion (BIC) and the Anderson–Darling criterion (ADC), when applied to identify the best probabilistic model for fitting hydrological extremes. The

* Corresponding author.
E-mail addresses: g.dibaldassarre@bristol.ac.uk (G. Di Baldassarre), francesco.laio@polito.it (F. Laio), alberto.montanari@unibo.it (A. Montanari).

study pointed out that the criteria have a good capability to recognize the correct parent distribution that was used to generate the synthetic data samples, with a slight tendency towards the selection of two-parameter distributions.

This study aims at further developing the analysis carried out by Laio et al. (2008) by assessing how the estimation of the design flood can be effectively improved through the application of the three aforementioned model selection criteria (AIC, BIC, ADC). This is carried out through an extensive numerical analysis, where the performance of the three criteria is compared by using synthetic samples of data. In particular, the study makes reference to the typical conditions that are encountered within the flood frequency analysis, namely, small sample sizes and asymmetric distributions.

2. Model selection criteria

We consider a record D with sample size n of independent outcomes of the random variable x, $x_1 \leqslant x_2 \leqslant \ldots \leqslant x_n$. The record is sampled from an unknown parent distribution $f(x)$. The problem of model selection can be formalized as follows. N_m probability distributions, M_j, $j = 1, \ldots, N_m$, that can be written in the form $M_j = g_j(x, \vartheta)$, with parameters $\hat{\vartheta}$ estimated from the available data sample D, can be used to describe the probability distribution of x. The scope of model selection is to identify the model M_{opt} which is better suited to represent the data, i.e. the model which is closer in some sense to the parent distribution $f(x)$.

Model selection criteria implicitly employ some notion of the principle of parsimony (Box and Jenkins, 1970) that is based on the analysis of the trade-off between bias and variance of the parameter estimates. In fact, it is well known that the bias of estimation decreases and the variance increases as the number of model parameters increases (Fig. 1, Burnham and Anderson, 2002). In this paper we consider three model selection criteria, namely the Akaike information criterion (AIC), the Bayesian information criterion (BIC) and the Anderson-Darling criterion (ADC). We refer to Laio et al. (2008) for a detailed description of the theoretical bases behind the three criteria, and report here only the essential information for an operative use of the criteria.

Akaike (1973) defined an information criterion that, for the candidate model M_j, can be written in the form:

$$AIC_j = -2\ln(L_j(\hat{\theta})) + 2p_j, \tag{1}$$

where p_j is the number of estimated parameters, and $L_j(\hat{\theta})$ is the likelihood function, $L_j(\hat{\vartheta}) = \prod_{i=1}^{n} g_j(x_i, \hat{\vartheta})$, evaluated at $\vartheta = \hat{\vartheta}$, where $\hat{\vartheta}$ is the maximum likelihood estimator of the parameters and $g_j(x, \vartheta)$ the probability density function (PDF). In practice, after the computation of the AIC_j, for all of the operating models, one selects

the model with the minimum AIC value, $AIC_{\min}$. One can see in Eq. (1) that the first and the second term in the right-hand side tend to decrease and increase, respectively, for increasing number of model parameters (Burnham and Anderson, 2002). This is necessary to assure the aforementioned trade-off between bias and variance according to the principle of parsimony (Box and Jenkins, 1970, Fig. 1).

The Bayesian information criterion (BIC) is based on the discrepancy between the model and the parent distribution in a Bayesian framework (Schwarz, 1978). BIC can be computed according to the following relationship:

$$BIC_j = -2\ln(L_j(\hat{\theta})) + \ln(n)p_j. \tag{2}$$

Model selection is performed by looking for the minimum BIC value. It turns out that the final form of this criterion is rather similar to that of AIC (see Eq. (1)), but one can see that the penalty due to the number of model parameters p_j is multiplied by $0.5\ln(n)$. As a consequence, BIC leans more than AIC towards lower-dimensional models when there are at least eight available observations. Several attempts to extend and generalize BIC have been made in the literature (e.g., Wasserman, 2000; Konishi and Kitagawa, 1996), but none of these seems to be particularly attractive when dealing with small samples and highly asymmetrical distributions, which is the usual case in hydrological applications.

We use a third model selection criterion, which is based on the use of the Anderson–Darling test statistic (e.g. Laio, 2004; Laio et al., 2008). The Anderson–Darling (Anderson and Darling, 1952) test has demonstrated good skills when applied to hydrology (e.g., Onoz and Bayazit, 1995; Laio, 2004; Viglione et al., 2007). We define the Anderson–Darling Criterion (ADC) as

$$ADC_j = 0.0403 + 0.116\left(\frac{\Delta_{AD,j} - \xi_j}{\beta_j}\right)^{\frac{\eta_j}{0.861}}$$
$$\text{if} \quad 1.2\xi_j \leq \Delta_{AD,j} \tag{3a}$$

$$ADC_j = \left[0.0403 + 0.116\left(\frac{0.2\xi_j}{\beta_j}\right)^{\frac{\eta_j}{0.861}}\right]\frac{\Delta_{AD,j} - 0.2\xi_j}{\xi_j}$$
$$\text{if} \quad 1.2\xi_j > \Delta_{AD,j}, \tag{3b}$$

where:

$$\Delta_{AD,j} = -n - \frac{1}{n}\sum_{i=1}^{n}\left[(2i-1)\ln\left[G_j(x_i, \hat{\vartheta})\right]\right.$$
$$\left. + (2n+1-2i)\ln\left[1 - G_j(x_i, \hat{\vartheta})\right]\right] \tag{3c}$$

and ξ_j, β_j, and η_j are distribution-dependent coefficients that are tabled by Laio (2004, Tables 3 and 5), for a set of seven distributions commonly employed for the frequency analysis of extreme events. In practice, after the computation of the ADC_j, for all of the operating models, one selects the model with the minimum ADC value. The principle of parsimony is in this case preserved by the fact that the coefficients ξ_j, β_j, and η_j in Eq. (3a)–(3c) depend on the considered distribution (see Laio et al., 2008 for details).

3. Numerical analysis

The analysis is performed by means of Monte Carlo simulations by using as operational models M_j a total of seven probability models commonly used in the frequency analysis of extreme events: four of these models (Gumbel or Extreme Value 1 (EV1) distribution, Normal (NORM) distribution, Generalized Extreme Value (GEV) distribution, Gamma or Pearson Type III (GAM) distribution) are defined in Table 1 in terms of their cumulative distribution function (CDF), $G_j(x, \vartheta)$, or probability density function (PDF),
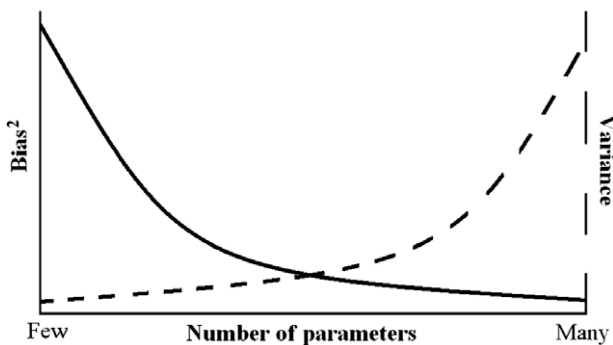

Fig. 1. The principle of parsimony: the conceptual trade off between squared bias (solid line) and variance (dashed line) versus the number of model parameters (Burnham and Anderson, 2002).

Table 1
Probability models considered in this study; the EV2, LN, and LP3 distributions are obtained as log-transforms of the EV1, NORM, and GAM distributions, respectively.

Distribution	Acronym (parameters)	CDF or PDF
Gumbel or extreme Value type I	EV1 $(\vartheta_1, \vartheta_2)$	$G(x,\vartheta) = \exp\left[-\exp\left(-\frac{x-\vartheta_1}{\vartheta_2}\right)\right]$
Normal or Gaussian	NORM $(\vartheta_1, \vartheta_2)$	$g(x,\vartheta) = \frac{1}{\sqrt{2\pi}\vartheta_2}\exp\left[-\frac{1}{2}\left(\frac{x-\vartheta_1}{\vartheta_2}\right)^2\right]$
Generalized extreme value	GEV $(\vartheta_1, \vartheta_2, \vartheta_3)$	$G(x,\vartheta) = \exp\left[-\left(1-\frac{\vartheta_3(x-\vartheta_1)}{\vartheta_2}\right)^{1/\theta_3}\right]$
Gamma or Pearson type III	GAM $(\vartheta_1, \vartheta_2, \vartheta_3)$	$g(x,\vartheta) = \frac{1}{|\vartheta_2|\Gamma(\vartheta_3)}\left(\frac{x-\vartheta_1}{\vartheta_2}\right)^{\vartheta_3-1}\exp\left(-\frac{x-\vartheta_1}{\vartheta_2}\right)$

Table 2
Probability models used as parent distribution in this study.

Distribution	Acronym (parameters)	CDF or Quantile function
Kappa	KAP $(\vartheta_1, \vartheta_2, \vartheta_3, \vartheta_4)$	$G(x,\vartheta) = \left\{1-\vartheta_4\left[1-\frac{\vartheta_3(x-\vartheta_1)}{\vartheta_2}\right]^{1/\vartheta_3}\right\}^{1/\vartheta_4}$
Wakeby	WAK $(\vartheta_1, \vartheta_2, \vartheta_3, \vartheta_4, \vartheta_5)$	$x(G) = \vartheta_5 + \frac{\vartheta_1}{\vartheta_2}\left[1-(1-G)^{\vartheta_2}\right] + \frac{\vartheta_3}{\vartheta_4}\left[1-(1-G)^{\vartheta_4}\right]$

$g_j(x,\vartheta)$. Three other distributions, namely the Frechet or Extreme Value 2 (EV2) distribution, the two-parameter lognormal (LN) distribution, and the log-Pearson type 3 (LP3) distribution, are converted to EV1, NORM and GAM distributions, respectively, when the data are preliminarily log transformed. Furthermore, two complex probabilistic models (four-parameter and five-parameter distributions) herein used as parent distribution (Kappa (KAP) distribution and Wakeby (WAK) distribution) are defined in Table 2.

In detail, the Monte Carlo experiment is structured as follows:

(a) One of the two complex probabilistic models (Table 2), with a specified set of parameters ϑ^*, is set as parent distribution, $f(x) = g_{j*}(x, \vartheta^*)$. The parameters are fixed and the "real" design flood value is estimated as the quantile corresponding to an assigned return period T.
(b) $f(x)$ is used for generating a sample of length n.
(c) The parameters of the seven approximating probabilistic models (Table 1) are estimated by using maximum likelihood (see Laio, 2004). For the GAM and GEV distributions, the maximum likelihood estimators do not exist or are not asymptotically efficient in few non-regular cases; in these case we use Smith's (1985) estimators instead of maximum likelihood estimators (see Laio (2004)Appendix A) for details).
(d) The seven probabilistic models are used to estimate the design flood as the quantile corresponding to the assigned return period T.
(e) The Akaike information criterion, Bayesian information criterion and Anderson–Darling criterion (see Section 2) are calculated for each of the seven models, therefore obtaining the measures AIC_j, BIC_j and ADC_j, $j = 1,..,7$. Each criterion will point out a potential optimal distribution, which is then used to estimate the design flood as the quantile corresponding to the assigned return period T.
(f) The relative error ε is evaluated by comparing the true design flood (step a) with (1) the design floods estimated by using the seven probabilistic models and (2) the design flood provided by the model selected by each criterion. Therefore 10 values for ε are obtained.
(g) Steps (b)–(f) are repeated m times.

In order to choose a specified set of parameters ϑ^* for the parent distribution that is used for generating the data we refer to the large database of annual peak discharges pertaining to catchments located in the United Kingdom. The catalogue is distributed with the Flood Estimation Handbook (Institute of Hydrology, 1999) and it contains 1000 annual floods time series of varying length, from a minimum of 5 years to a maximum of 112 years. Fig. 2 shows the two L-moment ratio diagrams (see e.g., Hosking and Wallis, 1997) for the UK database and a second-order polynomial interpolating curve. We use this curve to define four points (A, B, C and D, Fig. 2) characterized by an increasing L-skewness from 0.15 to 0.30 at steps of 0.05 (Fig. 2). By analysing Fig. 2 one can observe that there are many hydrometric stations characterised by a higher (or lower) L-skewness value. This was found to be mainly due to short sample lengths.

A first set of Monte Carlo experiments is characterized by taking $n = 50$, $m = 1000$, $T = 100$ and using KAP (Table 2) as the parent distribution. Therefore we focus on the estimation of the 100-year quantile using a sample size equal to 50. Each experiment is performed by using a set of parameters ϑ^* for the parent distribution such that the distribution L-moment ratios (Hosking and Wallis, 1997) match those of the four points A, B, C and D (Fig. 2).

Fig. 3 reports the boxplots of the relative errors while Table 3 reports the percentage of cases when the relative error lies in the ranges $[-0.15, +0.25]$ and $[-0.05, +0.15]$ (good estimation percentage and optimal estimation percentage, respectively). These goodness-of-fit measures were used as an overestimation of the design flood is commonly preferable in practical applications.

Table 3 and Fig. 3 show that the fitting capabilities of the probability distributions depend on the statistical properties of the generated data sample, with some distributions performing very well for low L-skewness values, and very badly for large L-skewness values (e.g., the EV1 and LN distributions). Also, it is interesting to note the poor performance of EV2 distribution for low L-skewness values (Fig. 3). This can be due to inefficiency of the maximum likelihood estimators for log transformed data. A general result is that the performance of the three model selection criteria is comparable and the estimation efficiency decreases as the L-skewness of the data sample increases. In particular, it is interesting to note that ADC performs better, in comparison to AIC and BIC, when the L-skewness increases (point D, L-skewness equal to 0.30). This result confirms some findings recently presented in the scientific literature (Laio et al., 2008). Table 3 and Fig. 3 demonstrate that the model selection criteria are more robust to changes in the parent distribution parameters compared to the probabilistic models: some distributions perform better than the criteria in each of the four points A, B, C, and D, but on average the criteria seem to perform better (e.g. row 6 of Table
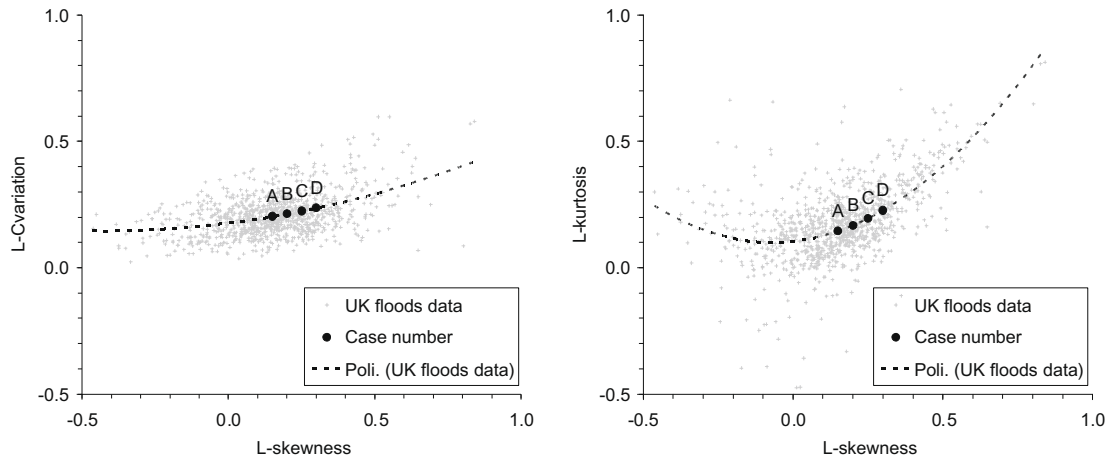
Fig. 2. L-moment ratio diagrams for the UK database: 2nd-order polynomial approximating the data and 4 cases considered here (L-skewness equal to 0.15, 0.20, 0.25, 0.30).
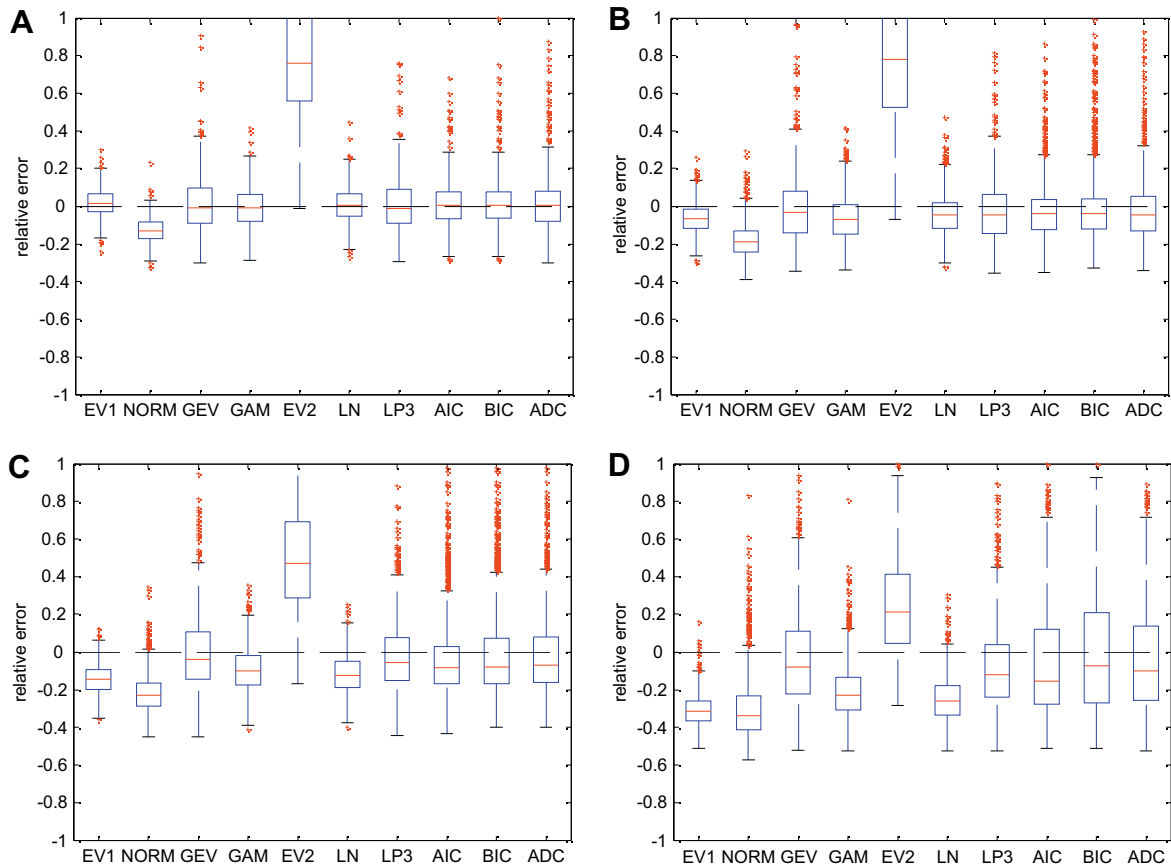

Fig. 3. Relative error distribution (boxplots); Parent distribution: Kappa. The four panels correspond to KAP distributions with L-moments ratios matching those of the points A, B, C, and D, in Fig. 2.

3, see also discussion below). An exception is the GEV distribution, which gives good results, providing similar performance with respect to the three model selection techniques. This can be explained by considering that the parent distribution, KAP, includes the GEV distribution as a special case. We will return in the next paragraph to this point.

A second set of Monte Carlo experiments was performed by taking $n = 50$, $m = 1000$, $T = 100$ and by using WAK as parent distribution (Table 2). Each experiment is performed by using a set of

parameters ϑ^* for the parent distribution such that the distribution L-moment ratios match those of the four points A, B, C and D (Fig. 1) and by assuming that $\vartheta_1 = 0$ (see Table 2). Table 4 reports the percentages of good and optimal estimation. It can be seen that the results obtained by using WAK as parent distribution (Table 4) are similar to the results obtained by using KAP (Table 3). However, the results obtained with the GEV distribution are now less satisfactory with respect to those obtained with the model selection techniques. The latter outcome confirms that the good perfor-

Table 3
Percentage of good (and optimal) estimation (%).

	EV1	NORM	GEV	GAM	EV2	LN	LP3	AIC	BIC	ADC
A	99 (74)	59 (12)	81 (47)	88 (56)	3 (1)	95 (63)	82 (47)	86 (58)	85 (59)	84 (55)
B	87 (38)	32 (8)	68 (38)	74 (39)	4 (1)	83 (43)	69 (38)	76 (39)	73 (37)	72 (38)
C	48 (11)	20 (4)	64 (35)	62 (28)	22 (9)	57 (21)	65 (35)	59 (26)	52 (20)	57 (25)
D	6 (2)	13 (5)	43 (21)	28 (10)	50 (31)	21 (7)	43 (21)	33 (18)	35 (20)	40 (21)
Mean	52 (27)							63 (35)		

Parent distribution: Kappa.

Table 4
Percentage of good (and optimal) estimation (%).

	EV1	NORM	GEV	GAM	EV2	LN	LP3	AIC	BIC	ADC
A	85 (45)	40 (9)	55 (32)	88 (54)	0 (0)	33 (15)	67 (38)	77 (44)	71 (39)	69 (40)
B	94 (66)	47 (9)	74 (39)	75 (34)	4 (2)	65 (38)	73 (38)	74 (45)	74 (47)	73 (46)
C	63 (24)	25 (7)	59 (29)	63 (29)	26 (14)	63 (29)	62 (30)	69 (35)	68 (35)	66 (32)
D	41 (11)	17 (4)	57 (30)	49 (16)	21 (11)	56 (24)	58 (30)	59 (24)	57 (22)	56 (24)
Mean	52 (25)							68 (36)		

Parent distribution: Wakeby.

mance of the GEV distribution during the first set of Monte Carlo experiments were partly due to using KAP as parent distribution.

The results of Monte Carlo experiments demonstrate that the use of model selection criteria improves design flood estimation compared to blindly using a fixed probabilistic model. For instance, one can observe that (see row 6 of Tables 3 and 4) the average percentage of good estimation (optimal estimation) obtained with the probabilistic models is equal to around 52% (25%), while the average percentage of good estimation (optimal estimation) obtained with the three criteria is equal to around 66% (35%).

Finally, Fig. 4 reports the minimum, mean and maximum values of the percentage of optimal estimation obtained in the two sets of Monte Carlo experiments (Tables 3 and 4). Fig. 4 points out that all criteria are more robust to changes in the parent distribution parameters. In particular, NORM and EV2 perform very poorly, EV1, GAM and LN perform very badly for large L-skewness values (see also, row 4 Tables 3 and 4). Only GEV and LP3 have minimum values of the percentage of optimal estimation comparable to the criteria (Fig. 4). Nevertheless, all criteria outperform GEV and LP3 in terms of mean and maximum value of the percentage of optimal estimation (Fig. 4).

A last numerical experiment was performed by referring to a real world river basin, namely the Samoggia River (Northern Italy), for which 1000 years of synthetic hourly river flows were gener-
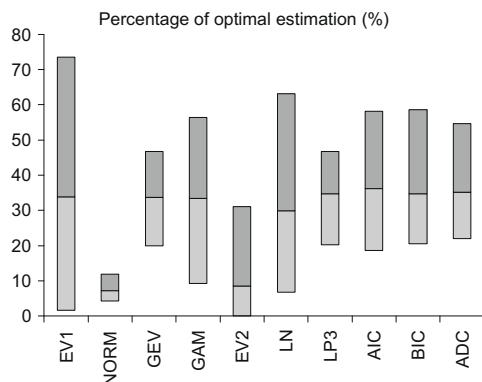
ated. The simulation of synthetic river flows was performed by first generating 1000 years of synthetic hourly temperature and rainfall data, by using stochastic models whose parameters have been previously calibrated (Brath et al., 2006). In detail, rainfall and temperature data were generated by using the multivariate Neyman–Scott model (Cowpertwait, 1996) and a linear stochastic process (Montanari et al., 1997), respectively. Therefore, using the above synthetic series as input to a rainfall–runoff model (Brath et al., 2006), 1000 years of river flows have been simulated, and the 1000 annual maxima have been extracted. The L-moment ratios of these synthetic flow data are equal to 0.459 (L-Cvariation), 0.271 (L-skewness), 0.121 (L-kurtosis).

The synthetic river flow series allowed us to evaluate the 100-year design flood Q_{100} by using the empirical distribution function of the 1000 maximum values. Then, we extracted sub-samples of size n, with n varying between 10 and 100 at steps of 10, from the series of the 1000 annual maximum peak flows. Finally, the three model selection criteria where applied to identify the best performing probability distribution. Fig. 5 shows the percentage
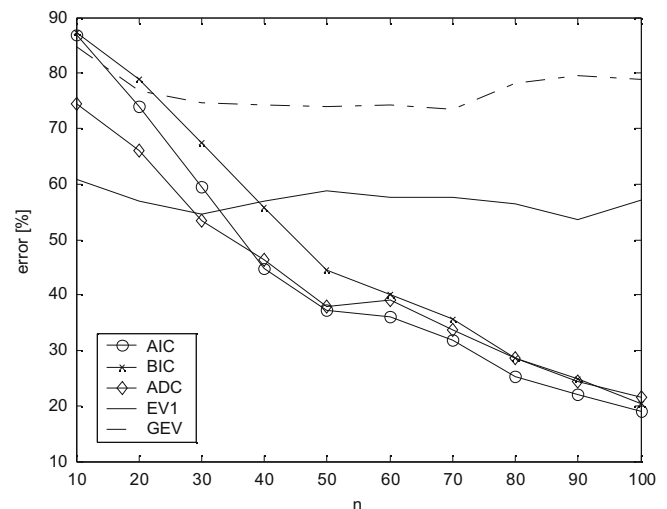

Fig. 4. Minimum, mean and maximum values of the percentage of optimal estimation for the seven probabilistic models and the three model selection criteria obtained in the two sets of Monte Carlo experiments (Tables 3 and 4).

Fig. 5. Samoggia synthetic river flows: percentage of cases when the estimated 100-year design flood lies out the ranges $[0.85Q_{100}–1.25Q_{100}]$ (poor estimation) as a function of the sample size.

of cases for which the estimated 100-year design flood is not included in the range $[0.85Q_{100}-1.25Q_{100}]$ (poor estimation percentage), depending on the sample size.

Fig. 5 shows that the three model selection criteria perform equally well, with an increasing efficiency with increasing sample size. Their performance is generally improved with respect to selecting always the EV1 and always the GEV, although one may note that better performance is obtained by selecting always the EV1 for very small sample sizes ($n = 10$ and $n = 20$). This result was expected and is due to the inefficiency of over-parameterized models (i.e., the GEV with respect to the EV1) in validation mode when dealing with very small sample sizes. Therefore Fig. 5 points out that applying a model selection criterion leads to a more efficient estimation of the 100-year quantile with respect to blindly using a candidate parent distribution, unless one suspects the potential presence of overfitting which may lead to prefer the choice of the more parsimonious model. It is interesting to note that if one uses a sample size equal to 100 year to estimate the 100-year design flood, by applying one of the model selection criteria the percentage of poor estimation is equal to around 20%, while using EV1 (or GEV) the percentage of poor estimation is equal to around 60% (or 80%).

4. Conclusions

The aim of the study was to investigate the capability of three model selection criteria, namely, the Akaike information criterion, the Bayesian information criterion, and the Anderson–Darling criterion, to improve the estimation of the so-called design flood. The criteria were tested through an extensive numerical analysis by referring to conditions that are frequently encountered within the flood frequency analysis; that is, small sample sizes and asymmetric distributions.

The results pointed out that all criteria give similar performance, although the Anderson–Darling criterion seems to perform better for increasing L-Skewness. Furthermore, it seems that the use of the criteria leads to a more efficient estimation of the flood quantile with respect to blindly using a candidate parent distribution among those considered within the present study.

Although the obtained results can not be considered completely conclusive, it turns out that model selection techniques are an interesting tool for flood frequency analysis as they are able to reduce the uncertainty in the estimation of the design flood.

Acknowledgement

References

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In: Petrov, B.N., Csaki, F. (Eds.), Second International Symposium on Information Theory. Academiai Kiado, Budapest, p. 281.

Anderson, T.W., Darling, D.A., 1952. Asymptotic theory of certain goodness-of-fit criteria based on stochastic processes. Ann. Mat. Stat. 23, 193–212.
Box, G.E.P., Jenkins, G.M., 1970. Time Series Analysis: Forecasting and Control. Holden Day Press, San Francisco, CA.
Bozdogan, H., 1987. Model selection and Akaike's information criterion (AIC): the general theory and its analytic extensions. Psychometrika 52, 270–345.
Brath, A., Montanari, A., Moretti, G., 2006. Assessing the effect on flood frequency of land use change via hydrological simulation (with uncertainty). J. Hydrol. 324, 141–153.
Burnham, K.P., Anderson, D.R., 2002. Model Selection and Multimodel Inference, second ed. Springer, New York.
Cahill, A.T., 2003. Significance of AIC differences for precipitation intensity distributions. Adv. Water Resour. 26, 457–464.
Cowpertwait, P.S.P., 1996. A generalized spatial-temporal model of rainfall based on a clustered point process. Proc. Royal Soc. London A450, 163–175.
Di Baldassarre G., Brath, A., Montanari, A., 2006a, Reliability of different depth–duration-frequency equations for estimating short-duration design storms, Water Resour. Res. 42, W12501, doi:10.1029/2006WR004911.
Di Baldassarre, G., Castellarin, A., Brath, A., 2006b. Relationships between statistics of rainfall extremes and mean annual precipitation: an application for design storm in northern central Italy. Hydrol. Earth Syst. Sci. 10, 589–601.
Hache, M., Perrault, L., Remillard, L., 1999. An approach for statistical model selection: application to the Saguenay-Lac-St-Jean hydrographic basin. Canad. J. Civil Eng. 26 (2), 216–225.
Hosking, J.R.M., Wallis, J.R., 1997. Regional Frequency Analysis. Cambridge University Press, Cambridge, UK.
Hurvich, C.M., Tsai, C., 1989. Regression and time series model selection in small samples. Biometrika 76, 297–307.
Institute of Hydrology (1999), Flood Estimation Handbook, Institute of Hydrology, Wallingford, Oxford.
Konishi, S., Kitagawa, G., 1996. Generalized information criteria in model selection. Biometrika 83 (4), 875–890.
Kottegoda, N.T., Rosso, R., 1998. Statistics Probability and Reliability for Civil and Environmental Engineers. McGrawHill, New York.
Kroll, C.N., Vogel, R.M., 2002. Probability distribution of low streamflow series in the United States. J. Hydrol. Eng. ASCE 7 (2), 137–146.
Laio, F., 2004. Cramer–von Mises and Anderson–Darling goodness of fit tests for extreme value distributions with unknown parameters, Water Resour. Res. 40, W09308, doi:10.1029/2004WR003204.
Laio, F., Di Baldassarre, G., Montanari, A., 2008. Model selection techniques for the frequency analysis of hydrological extremes, submitted for publication.
Mitosek, H.T., Strupczewski, W.G., Singh, V.P., 2006. Three procedures for selection of annual flood peak distribution. J. Hydrol. 323, 57–73.
Montanari, A., Rosso, R., Taqqu, M.S., 1997. Fractionally differenced ARIMA models applied to hydrologic time series: identification, estimation and simulation. Wat. Resour. Res. 33, 1035–1044.
Mutua, F.M., 1994. The use of the Akaike Information Criterion in the identification of an optimum flood frequency model. Hydrol. Sci. J. 39 (3), 235–244.
Onoz, B., Bayazit, M., 1995. Best-fit distribution of largest available flood samples. J. Hydrol. 167, 195–204.
Pandey, M.D., van Gelder, P., Vrijling, J.K., 2001. Assessment of an L-kurtosis-based criterion for quantile estimation. J. Hydrol. Eng. 6 (4), 284–292.
Schwarz, G., 1978. Estimating the dimension of a model. Annals Stat. 6, 461–464.
Smith, R.L., 1985. Maximum likelihood estimation in a class of nonregular cases. Biometrika 72 (1), 67–90.
Stedinger, J.R., Vogel, R.M., Foufula-Georgiou, E., 1992. Frequency analysis of extreme events. In: Maidment, R. (Ed.), Handbook of Hydrology. McGraw-Hill, New York (Chapter 18).
Strupczewski, W.G., Singh, V.P., Feluch, W., 2001. Non-stationary approach to at-site flood frequency modelling Maximum likelihood estimation. J. Hydrol. 248, 123–142.
Strupczewski, W.G., Singh, V.P., Weglarczyk, S., 2002. Asymptotic bias of estimation methods caused by the assumption of false probability distributions. J. Hydrol. 258, 122–148.
Viglione, A., Laio, F., Claps, P., 2007. A comparison of homogeneity tests for regional frequency analysis, Water Resour. Res. 43, W03428, doi:10.1029/2006WR005095.
Vogel, R.M., Wilbert, O., McMahon, T.A., 1993a. Floodflow frequency model selection in southwestern United States. J. Water Resour. Planning Manag. 119 (3), 353–366.
Vogel, R.M., McMahon, T.A., Chiew, F., 1993b. Floodflow frequency model selection in Australia. J. Hydrol. 146, 421–449.
Wasserman, L., 2000. Bayesian model selection and model averaging. J. Math. Psychol. 44, 92–107.