# Cramer-von Mises and Anderson-Darling goodness of fit tests for extreme value distributions with unknown parameters

# Cramer–von Mises and Anderson-Darling goodness of fit tests for extreme value distributions with unknown parameters

Francesco Laio

Dipartimento di Idraulica, Trasporti ed Infrastrutture Civili, Politecnico di Torino, Torino, Italy

Received 23 March 2004; revised 15 June 2004; accepted 1 July 2004; published 28 September 2004.

[1] The use of goodness of fit tests based on Cramer–von Mises and Anderson-Darling statistics is discussed, with reference to the composite hypothesis that a sample of observations comes from a distribution, $F_H$, whose parameters are unspecified. When this is the case, the critical region of the test has to be redetermined for each hypothetical distribution $F_H$. To avoid this difficulty, a transformation is proposed that produces a new test statistic which is independent of $F_H$. This transformation involves three coefficients that are determined using the asymptotic theory of tests based on the empirical distribution function. A single table of coefficients is thus sufficient for carrying out the test with different hypothetical distributions; a set of probability models of common use in extreme value analysis is considered here, including the following: extreme value 1 and 2, normal and lognormal, generalized extreme value, three-parameter gamma, and log-Pearson type 3, in all cases with parameters estimated using maximum likelihood. Monte Carlo simulations are used to determine small sample corrections and to assess the power of the tests compared to alternative approaches. *INDEX TERMS:* 1894 Hydrology: Instruments and techniques; 1821 Hydrology: Floods; 1854 Hydrology: Precipitation (3354); *KEYWORDS:* Anderson-Darling test, extreme value analysis, floods, tests of fit

**Citation:** 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.

## 1. Introduction

[2] Model testing and verification are basic steps of statistical inference, but relatively little attention has been devoted to this crucial issue in the hydrologic literature. This is probably due to a widespread distrust of classical goodness of fit tests, when applied to small samples, and to the lack of a valid alternative for the case when the parameters of the hypothesized distribution are estimated using the same sample that is being tested. Important testings techniques, borrowed from applied statistics, have been proposed in the hydrologic field by *Vogel* [1986] and *Vogel and McMartin* [1991], with an approach based on the probability plot correlation coefficient (PP tests), by *Ahmad et al.* [1988], using empirical distribution function statistics (EDF tests), and by *Chowdhury et al.* [1991], *Fill and Stedinger* [1995], and *Wang* [1998], with techniques based on the comparison of empirical and hypothetical L moments ratios (LM tests).

[3] However, none of these tests has reached a broad consensus in the hydrologic community, possibly due to some complications that inevitably arise when the parameters of the hypothetical distribution are unknown. This is the most common case in hydrology; in the following, it will be referred to as "case p," in analogy to the *Stephens*' [1986] use of "case 0" for the case when the parameters are fully specified a priori. In case p, the distributions of the PP, EDF, and LM test statistics depend on the so-called null hypothesis $H_0$, i.e., on the probability distribution that is being tested [e.g., *Stephens*, 1986]. This means that the percentage points, i.e., the 100 $(1 - \alpha)$ percentiles of the distributions of the test statistics ($\alpha$ is the significance level of the test), have to be recalculated for any $H_0$. The method of parameter estimation, the presence of a shape parameter, and the sample size also have an influence on percentage points, and this further complicates the analysis. Moreover, only for the LM tests the approximate test statistic distribution is known in analytical form, while for all other cases one should refer to tables of percentage points. There is thus the necessity to have a different table for each distribution, and tables of percentage values for some distributional families are still lacking.

[4] The scope of this paper is to overcome some of the above difficulties for tests based on quadratic EDF statistics (Cramer–von Mises and Anderson-Darling tests, see section 2). The main result of the paper is given by equation (11) in section 3, a transformation which converts the test statistic in case p to a case 0 statistic, with well known percentage points. A single table is sufficient for carrying out the test, containing the values of the three coefficients of the transformation for any distribution of interest. An application to extreme value distributions is presented in section 4, and a power study in section 5 demonstrates the good performances of this method in comparison to PP, LM, chi-square and Kolmogorov-Smirnov tests. The whole procedure is summarized in section 6, where the test is applied to the annual maxima of hourly rainfall recorded in Genoa (Italy) from 1931 to 1988.

## 2. EDF Tests

[5] Suppose that $x_1 \leq \ldots \leq x_n$ is an ordered sample of $n$ independent observations from a distribution with cumula-


0043-1397/04/2004WR003204$09.00

tive distribution function (CDF) $F_R(x)$, and that one wishes to test the null hypothesis $H_0: F_R(x) = F(x, \theta)$, where $F$ defines the family of distributions (such as normal or gamma) and $\theta$ is a vector of parameters. EDF tests are based on the comparison between the hypothetical and empirical distribution function, $F(x, \theta)$ and $F_n(x)$ respectively, where

$$
\begin{aligned}
&F_n(x) = 0, && x < x_1 \\
&F_n(x) = \frac{i}{n}, && x_i \le x < x_{i+1}, \quad i = 1, \dots, n-1 \\
&F_n(x) = 1, && x_n \le x.
\end{aligned} \tag{1}
$$

The discrepancy between the two distributions can be measured either with statistics of the form $\max_x |F_n(x) - F(x, \theta)|$ (Kolmogorov-Smirnov test), or using quadratic statistics,

$$
Q^2 = n \int_{\text{all}x} [F_n(x) - F(x, \theta)]^2 \Psi(x) \mathrm{d}F(x) \tag{2}
$$

where $\Psi(x)$ is a weight function. When $\Psi(x) = 1$, one has the Cramer–von Mises statistic, usually called $W^2$, that is a measure of the mean squared difference between the empirical and hypothetical CDF; when $\Psi(x) = [F(x, \theta)(1 - F(x, \theta))]^{-1}$, the tails of the distribution are weighted more than the central part, and one has the Anderson-Darling statistic, called $A^2$. $W^2$ and $A^2$ are estimated in practice as [e.g., *Stephens*, 1986],

$$
W^2 = \sum_{i=1}^{n} \left[ F(x_i, \theta) - \frac{2i-1}{2n} \right]^2 + \frac{1}{12n}, \tag{3}
$$

$$
\begin{aligned}
A^2 = &- n - \frac{1}{n} \sum_{i=1}^{n} [(2i-1) \ln[F(x_i, \theta)] \\
&+ (2n + 1 - 2i) \ln[1 - F(x_i, \theta)]].
\end{aligned} \tag{4}
$$

When $\theta$ is completely or partially unspecified (case p), the parameters need be replaced by estimates $\hat{\theta}$. With the transformation $z = F(x, \theta)$, the general quadratic statistic $Q^2$ in equation (2) becomes

$$
Q^2 = n \int_0^1 [F_n(z) - z]^2 \Psi(z) \mathrm{d}z = \int_0^1 y_n^2(z) \mathrm{d}z, \tag{5}
$$

where $F_n(z)$ is the EDF of the variable $z$, calculated as in equation (1). The asymptotic distribution of $Q^2$ is the same as the distribution of $\int_0^1 y_n^2(z)\mathrm{d}z$, where $y_n(z) = \sqrt{n}(F_n(z) - z)\sqrt{\Psi(z)}$. It follows that the limiting form of the stochastic process $y_n(z)$ is of central importance: for large $n$, it has been demonstrated by *Darling* [1955] and *Durbin* [1973] that $y_n(z)$ converges to a Gaussian process $y(z)$ with mean 0 and covariance

$$
\rho(z, s) = \left[ \min(z, s) - zs - g(z)^T \Sigma^{-1} g(s) \right] \sqrt{\Psi(s)\Psi(z)}, \quad 0 \le s, z \le 1. \tag{6}
$$

$g(z)^T$ in equation (6) is the transpose of the vector $g(z) = \frac{\partial z}{\partial \theta}\big|_{\theta=\hat{\theta}}$, and $\Sigma^{-1}$ is the inverse of Fisher information matrix (divided by $n$),

$$
\Sigma = E\left( \frac{\partial \ln f(x, \theta)}{\partial \theta} \frac{\partial \ln f(x, \theta)}{\partial \theta^T} \right), \tag{7}
$$

where $E(\cdot)$ indicates expectation, the derivatives are calculated at $\theta = \hat{\theta}$, and $f(x, \theta)$ is the probability density function (PDF) correspondent to $F(x, \theta)$. Equation (6) is valid when the estimator $\hat{\theta}$ is asymptotically efficient, i.e., unbiased and with minimum variance for $n \to \infty$, in which case $n\Sigma^{-1}$ is the asymptotic dispersion matrix of the estimated parameters [e.g., *Kendall and Stuart*, 1977, p. 59]. Maximum likelihood (ML) estimators are usually asymptotically efficient [*Kendall and Stuart*, 1977, p. 47], while other estimators commonly used in hydrology (e.g., moments or L moments estimators) are not. The term $g(z)^T \Sigma^{-1} g(s)$ in equation (6) vanishes when the parameters are known a priori (case 0), while it depends on the hypothetical distribution in case p, showing that the tests are not distribution-free in this case.

[6] Provided $y(z)$ is Gaussian, $Q^2 = \int_0^1 y^2(z)\mathrm{d}z$ is distributed as a weighted sum of independent $\chi_1^2$ random variables, where the weights are the inverse of the roots of the Fredholm determinant associated with $\rho(z, s)$ [*Stephens*, 1976]. The numerical evaluation of percentage points for $W^2$ and $A^2$ therefore requires (1) to determine the covariance function (6), (2) to calculate the eigenvalues of the Fredholm homogeneous integral equation of the second type with $\rho(z, s)$ as a kernel, and (3) to compute the distribution of a linear combination of independent $\chi_1^2$ variables [see *Johnson et al.*, 1994, pp. 444–450]. When $\theta$ is fully specified a priori (case 0), the results are well known since the seminal paper by *Anderson and Darling* [1952], with percentage points of $W^2$ and $A^2$ tabled in textbooks. Results for case p are also available for a limited number of cases [see *Stephens*, 1986], but the computational effort for distributions with more than 2 parameters is still formidable. Moreover, in hydrological applications $n$ is usually small, and this diminishes the importance of having very accurate asymptotic percentage points, since in any case these need be corrected using Monte Carlo simulations. In section 3 an alternative procedure for determining approximate percentage points in the parametric case is proposed.

## 3. A Procedure for Using Case 0 Percentage Points in the Parametric Case

[7] It is apparent from the discussion in the previous section that the distributions of $W^2$ and $A^2$ in case 0 are different from those obtained in case p. Nevertheless, the shape of the distributions is similar, in particular in the right tail. In fact, in both cases the test statistics are distributed as the integral of a squared Gaussian process (see equation (5)), while only the covariance function (6) is different. An example is given in Figure 1, where the case 0 and case p percentage points for the Cramer–von Mises test applied to a Gaussian distribution are plotted one against the other, for significance levels between 0.01 and 0.99.

[8] The relation between corresponding percentage points in Figure 1 is clearly monotonic and not far from being

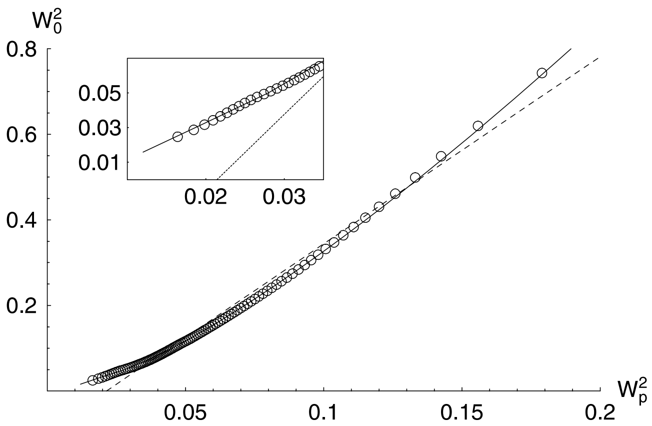


**Figure 1.** Relation between case 0 and case p percentage points for the Cramer–von Mises test for normality. The open circles represent "literature" data from *Anderson and Darling* [1952, equation (4.35)] and *Stephens* [1986, Table 4.9], the solid line is a plot of equation (11), and the dashed line represents equation (8). The inset shows an enlargement of the lower part of the graph.

linear; similar results are found for distributions other than the Gaussian. This suggests to find a mapping, $\omega = \phi(Q_p^2)$, that transforms the test statistic in the parametric case, $Q_p^2$, into a variable $\omega$ whose distribution is close to that of the case 0 Cramer–von Mises statistic, $W_0^2$. The hypothesis that the two distributions are approximately the same poses on empirical grounding, and appropriate verifications will be given in the following and in section 5. As a first attempt, the transformation $\phi$ can be supposed to be linear, which corresponds to imposing the identity of the first two moments of $\omega$ and $W_0^2$:

$$\omega = \frac{\sigma_0}{\sigma_p}\left(Q_p^2 - \mu_p\right) + \mu_0, \tag{8}$$

where $\mu$ and $\sigma$ are the mean and the standard deviation of $Q^2$, calculated as by *Stephens* [1976]:

$$\begin{aligned} \mu &= \int_0^1 \rho(z,z)\mathrm{d}z \\ \sigma &= \left(2\int_0^1\int_0^1 \rho(z,s)\rho(s,z)\mathrm{d}s\mathrm{d}z\right)^{0.5}, \end{aligned} \tag{9}$$

with the appropriate case 0 or case p covariance as argument of the integrals. However, the approximation provided by equation (8) is rather poor, in particular in the tails of the distribution, as shown for example by the dashed line in Figure 1.

[9] To improve the accuracy of the approximation, one can include also the third central moment,

$$M_3 = 8\int_0^1\int_0^1\int_0^1 \rho(z,s)\rho(s,t)\rho(z,t)\mathrm{d}t\mathrm{d}s\mathrm{d}z, \tag{10}$$

and find a transformation which entails the identity of the first three moments of $\omega$ and $W_0^2$. A transformation based on the identity $\left(\frac{\omega-\xi_0}{\beta_0}\right)^{\eta_0} = \left(\frac{Q_p^2-\xi_p}{\beta_p}\right)^{\eta_p}$, where $\xi$, $\beta$ and $\eta$ are location, scale and shape parameters, respectively, is used here:

$$\begin{aligned} \omega &= \beta_0\left(\frac{Q_p^2 - \xi_p}{\beta_p}\right)^{\frac{\eta_p}{\eta_0}} + \xi_0, \quad 1.2\xi_p \le Q_p^2 \\ \omega &= \left[\beta_0\left(\frac{0.2\xi_p}{\beta_p}\right)^{\frac{\eta_p}{\eta_0}} + \xi_0\right]\frac{Q_p^2 - 0.2\xi_p}{\xi_p}, \quad Q_p^2 \end{aligned} \tag{11}$$

The second equation in equation (11) is necessary for keeping $\omega$ on the real positive axis also when $Q_p^2 < \xi_p$, and to improve the accuracy of the transformation in the very low part of the distribution. This correction is seldom required, since the probability of having $Q_p^2 < 1.2\xi_p$ is below 0.12 even when $H_0$ is true.

[10] The basic hypothesis behind equation (11) is that the first three moments of the rescaled case 0 and case p statistics, $t_0 = \left(\frac{\omega-\xi_0}{\beta_0}\right)^{\eta_0}$ and $t_p = \left(\frac{Q_p^2-\xi_p}{\beta_p}\right)^{\eta_p}$, are identical; the problem is to find two sets of parameters $\mathcal{S}_0 : [\xi_0, \beta_0, \eta_0]$ and $\mathcal{S}_p : [\xi_p, \beta_p, \eta_p]$ in equation (11) such that this hypothesis is met. A distributional form for $t_0$ and $t_p$ need be specified for this scope: $t$ is supposed here to have a standardized exponential distribution, $f(t) = e^{-t}$, leading to the following relations between moments and parameters [e.g., *Johnson et al.*, 1994, pp. 629–632]

$$\begin{aligned} \mu &= \xi + \beta\Gamma\left[1 + \frac{1}{\eta}\right] \\ \sigma^2 &= \beta^2\left(\Gamma\left[1 + \frac{2}{\eta}\right] - \Gamma^2\left[1 + \frac{1}{\eta}\right]\right) \\ M_3 &= \beta^3\left(\Gamma\left[1 + \frac{3}{\eta}\right] - 3\Gamma\left[1 + \frac{1}{\eta}\right]\Gamma\left[1 + \frac{2}{\eta}\right] + 2\Gamma^3\left[1 + \frac{1}{\eta}\right]\right), \end{aligned} \tag{12}$$

where $\Gamma[\cdot]$ is the gamma function. The hypothesis that $t$ is exponentially distributed is ancillary to the determination of the relations (12) and does not pretend to be unique: the choice of the distribution, similarly to that of the kernel of the transformation, is inevitably subjective and empirical and as such needs be accurately verified. A first validation is given in Figure 1, where it is seen that the three-parameter equation (11) provides a very good fit to the "exact" points for the Gaussian distribution, even in the tails: the error in evaluating the significance level $\alpha$ is almost always below 1%, and it is lower than 0.25% for $\alpha < 0.1$, where the accuracy is mostly relevant for testing purposes. More validations are given in section 5 and Appendix B.

[11] Adopting the described methods, the case 0 coefficients are found as follows: the first 3 moments are determined by setting $\rho(s, t) = [\min(s, t) - st]$ in equations (9) and (10), obtaining $\mu = 1/6$, $\sigma^2 = 1/45$ and $M_3 = 8/945$, and the corresponding coefficients $\mathcal{S}_0$ from (12) are

$$\xi_0 = 0.0403 \qquad \beta_0 = 0.116 \qquad \eta_0 = 0.851. \tag{13}$$

The case p coefficients depend both on the hypothetical distribution and on the statistics that is used ($W^2$ or $A^2$), since the covariance function (6) changes from case to case. Examples of the determination of the $\mathcal{S}_p$ coefficients for extreme value distributions are given in section 4. Once the transformation (11) is adopted, $\omega$ becomes the new test

**Table 1.** Probability Models Considered in This Paper

| Distribution | Acronym | CDF or PDF | Range |
|---|---|---|---|
| Gumbel or extreme value type I | EV1 | $F(x,\theta) = \exp\left[-\exp\left[-\frac{x-\theta_1}{\theta_2}\right]\right]$ | $-\infty < x < \infty$ |
| Normal or Gaussian | NORM | $f(x,\theta) = \frac{1}{\sqrt{2\pi}\theta_2}\exp\left[-\frac{1}{2}\left(\frac{x-\theta_1}{\theta_2}\right)^2\right]$ | $-\infty < x < \infty$ |
| Generalized extreme value | GEV | $F(x,\theta) = \exp\left[-\left(1-\frac{\theta_3(x-\theta_1)}{\theta_2}\right)^{1/\theta_3}\right]$ | $\frac{\theta_3(x-\theta_1)}{\theta_2} < 1$ |
| Gamma or Pearson type 3 | GAM | $f(x,\theta) = \frac{1}{\lvert\theta_2\rvert\Gamma[\theta_3]}\left(\frac{x-\theta_1}{\theta_2}\right)^{\theta_3-1}\exp\left[-\frac{x-\theta_1}{\theta_2}\right]$ | $\frac{x-\theta_1}{\theta_2} > 0$ |

statistic and the critical region of the test can be taken to be the same as that of the case 0 Cramer–von Mises statistic, $W_0^2$. For example, the null hypothesis is rejected if $\omega$ is greater than 0.347, 0.461, and 0.743 for significance levels $\alpha$ = 0.10, 0.05, and 0.01, respectively. The probability distribution of $\omega$ when $H_0$ is true is approximately the same as that of $W_0^2$, which has been obtained in analytical form by *Anderson and Darling* [1952]; the first two terms of their infinite series solution are

$$F^*(\omega) = \frac{1}{\pi\sqrt{\omega}}\left\{e^{-\frac{1}{16\omega}}\mathrm{K}_{\frac{1}{4}}\left[\frac{1}{16\omega}\right] + 1.118e^{-\frac{25}{16\omega}}\mathrm{K}_{\frac{1}{4}}\left[\frac{25}{16\omega}\right]\right\}, \quad (14)$$

where $\mathrm{K}_{\frac{1}{4}}[\cdot]$ is the modified Bessel function of the second kind. These two terms alone provide a very accurate approximation of the probability distribution when $\omega < 1.2$; when $\omega$ exceeds 1.2, $F^*(\omega)$ can be set to 1. Equation (14) allows a fast calculation of the probability level corresponding to a given value of the test statistic, which is very important for a clear assessment of the goodness of fit of the hypothesized distribution and for combining the results of the test from multiple samples.

[12] The theory heretofore developed is valid in asymptotic conditions, i.e for sample size $n \to \infty$. The convergence of the $Q^2$ distribution to its asymptotic values is rather fast, but the deviations can be relevant when the sample size is very small. It is supposed here that also the small sample distribution can be characterized in terms of its first 3 central moments, and that the transformation from a case p small sample distribution to a case 0 asymptotic distribution can be carried out with equation (11) as before, but with modified parameters $\mathcal{S}_p(n)$,

$$\xi_p(n) = \xi_p\left(1 + \frac{k_1}{n} + \frac{\nu_1}{\sqrt{n}}\right) \quad \beta_p(n) = \beta_p\left(1 + \frac{k_2}{n} + \frac{\nu_2}{\sqrt{n}}\right)$$
$$\eta_p(n) = \eta_p\left(1 + \frac{k_3}{n} + \frac{\nu_3}{\sqrt{n}}\right). \quad (15)$$

The coefficients $k_i$ and $\nu_i$ in equation (15) are estimated by employing Monte Carlo simulations involving 100000 samples for each of many values of $n$ (see section 4 for details). A least squares regression of, say, $\left(\frac{\xi_p(n)}{\xi_p} - 1\right)$ on $\{1/n, 1/\sqrt{n}\}$ then provides the appropriate coefficients $k_1$ and $\nu_1$.

## 4. An Application to Extreme Value Distributions

[13] The procedure delineated in the previous section is general, and can be applied to find a transformation for any distributional family. This section is devoted to the description of this procedure for a group of probability distributions of common use in the frequency analysis of extreme events, defined in Table 1 in terms of their CDF, $F(x, \theta)$, or PDF, $f(x, \theta)$. 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. For all distributions in Table 1 the test is applied with parameters $\theta$ estimated using the maximum likelihood (ML) method, except in special cases described below. The procedure for determining the $\mathcal{S}_p(n)$ coefficients for the EV1 distribution is taken as a reference and discussed in detail in section 4.1.

### 4.1. An Example: The EV1 Distribution

[14] For the EV1 distribution the ML method is asymptotically efficient for estimating the parameter vector $\theta = [\theta_1, \theta_2]$; thus equation (6) holds, with

$$g(z) = \begin{pmatrix} \frac{\partial z}{\partial\theta_1} \\ \frac{\partial z}{\partial\theta_2} \end{pmatrix} = \begin{pmatrix} \frac{\partial F(x,\theta)}{\partial\theta_1} \\ \frac{\partial F(x,\theta)}{\partial\theta_2} \end{pmatrix} = \begin{pmatrix} -\frac{1}{\theta_2}e^{-e^{\frac{-x+\theta_1}{\theta_2}}+\frac{-x+\theta_1}{\theta_2}} \\ \frac{-x+\theta_1}{\theta_2{}^2}e^{-e^{\frac{-x+\theta_1}{\theta_2}}+\frac{-x+\theta_1}{\theta_2}} \end{pmatrix}$$
$$= \frac{1}{\theta_2}\begin{pmatrix} z\ln[z] \\ -z\ln[z]\ln[-\ln[z]] \end{pmatrix} \quad (16)$$

and (see equation (7))

$$\Sigma^{-1} = \frac{6\theta_2^2}{\pi^2}\begin{pmatrix} (1-\gamma)^2 + \frac{\pi^2}{6} & 1-\gamma \\ 1-\gamma & 1 \end{pmatrix} \quad (17)$$

where $\gamma \simeq 0.5772$ is Euler's constant. When equations (16) and (17) are set in equation (6) one obtains a covariance function that is independent of the parameter vector, as expected since $\theta_1$ and $\theta_2$ are location and scale parameters [see *Stephens*, 1986]. The first 3 central moments of the case p distribution are thus calculated by numerically solving the integrals in equations (9) and (10): $\mu$ = 0.0587, $\sigma^2 = 1.116 \times 10^{-3}$ and $M_3 = 6.553 \times 10^{-5}$ are obtained for the Cramer–von Mises statistic, and $\mu$ = 0.387, $\sigma^2 = 3.678 \times 10^{-2}$ and $M_3 = 1.158 \times 10^{-2}$ for the Anderson-Darling statistic. The coefficients $\xi_p$, $\beta_p$ and $\eta_p$ are then determined from equation (12), obtaining the results in the first row of Tables 2 and 3. The small sample corrections (15) are found by means of Monte Carlo simulations of size $N$ = 100000, with $n$ from 10 to 100 in steps of 10. For the EV1 distribution, it is found that the convergence is very rapid

**Table 2.** Coefficients to Be Set in Equation (11) for the Cramer–von Mises Statistic, Asymptotic Case[a]

| Distribution[b] | $\xi_p$ | $\beta_p$ | $\eta_p$ |
|---|---|---|---|
| EV1 and EV2 | 0.0223 | 0.0376 | 1.090 |
| NORM and LN | 0.0226 | 0.0380 | 1.081 |
| GEV[c] | $0.0200\ (1 + 0.11\hat{\theta}_3 + 0.20\ \hat{\theta}_3^2 + 0.08\ \hat{\theta}_3^3)$ | $0.0314\ (1 + 0.20\ \hat{\theta}_3 + 0.32\ \hat{\theta}_3^2 + 0.13\ \hat{\theta}_3^3)$ | $1.114(1 - 0.02\ \hat{\theta}_3 - 0.04\ \hat{\theta}_3^2 - 0.01\ \hat{\theta}_3^3)$ |
| GAM and LP3[d] | $0.0197\ (1 + 0.23\ \hat{\theta}_3^{-1} + 0.26\ \hat{\theta}_3^{-2})$ | $0.0308\ (1 + 0.35\ \hat{\theta}_3^{-1} + 0.36\hat{\theta}_3^{-2})$ | $1.119\ (1 - 0.05\hat{\theta}_3^{-1} - 0.09\hat{\theta}_3^{-2})$ |

[a]Here $\hat{\theta}_3$ is an asymptotic efficient estimator (usually maximum likelihood) of the shape parameter of the distribution.
[b]For tests of the EV2, LN, and LP3 distributions the data must be preliminarily log transformed.
[c]For the GEV distribution, if $\hat{\theta}_3 > 0.5$, $\hat{\theta}_3 = 0.5$ must be set in the regressions.
[d]For the GAM and LP3 distributions, if $\hat{\theta}_3 < 2$, $\hat{\theta}_3 = 2$ must be set in the regressions.

[see also *Stephens*, 1977], and the dependence on $1/\sqrt{n}$ is negligible (Tables 4 and 5, first row).

### 4.2. Procedure for Three-Parameter Distributions

[15] For any probability distribution with only location and scale parameters, the appropriate coefficients are determined in the same manner as for the EV1 distribution. This is the case of the Gaussian distribution (see coefficients in the second row of Tables 2–5). When instead one has a GEV or a GAM distribution, two further problems complicate the procedure: first, ML estimators are no more necessarily asymptotically efficient, since the range of the distribution depends on θ (see Table 1); second, a shape parameter $\theta_3$ is present.

[16] Asymptotic efficiency is considered first, since it affects the validity of equation (6), and the whole procedure as a consequence. *Smith* [1985] demonstrates that the classical regularity conditions for asymptotic efficiency are satisfied by ML estimators when $\theta_3 > 2$ for the GAM distribution, or $\theta_3 < 0.5$ for the GEV distribution. In all other cases, one should modify ML estimators as proposed by *Smith* [1985] in order to obtain asymptotically efficient estimators (see Appendix A).

[17] The second problem with three-parameter distributions relies on the presence of a shape parameter $\theta_3$: the percentage points of EDF tests in case p depend on $\theta_3$ [e.g., *Stephens*, 1986], and this should be reflected by the coefficients $\mathcal{S}_p$. The form of the functional dependence of $\mathcal{S}_p$ upon $\theta_3$ cannot in general be found analytically; this requires that (1) the $\mathcal{S}_p$ coefficients are determined for a set of $\theta_3$ values covering the presumed variability of the shape coefficient in real world samples and (2) a simple function of $\theta_3$ is found that approximates the obtained points. Details on the specific procedure followed for the GAM and GEV distributions are given in Appendix A, while the obtained coefficients are reported in Tables 2 and 3.

[18] A final point regards the small sample corrections: for the GEV and GAM distributions, the size of the Monte Carlo simulation is still $N = 100000$, and 35 combinations of $n$ and $\theta_3$ values are used ($n$ = [20, 35, 50, 75, 100]; $\theta_3$ = [−1, −0.75, −0.5, −0.25, 0, 0.25, 0.5] for the GEV distribution; $\theta_3$ = [2, 2.5, 3, 4, 6, 10, 100] for the GAM distribution). An evident dependence of the corrections on the shape coefficient is found for the gamma distribution, and this is reflected by the $\nu_i$ coefficients in equation (15) that are functions of $\theta_3$ (fourth row in Tables 4 and 5). No such dependence is found for the GEV: probably this is due to occasional failures of the maximization algorithm for finding ML estimators [see also *Madsen et al.*, 1997, p. 751], which in turn increase the "noise" in the Monte Carlo simulation and conceal the possible dependence on $\theta_3$.

## 5. Accuracy and Power of the Test Statistics

[19] The application of EDF tests with the proposed approach is simple, but it is important to further investigate if the approximation of ω to $W_0^2$ is accurate, and if the obtained tests are powerful. As for the accuracy, it is shown in Appendix B that the distribution of ω and $W_0^2$ are very close one to the other, with differences in percentage levels of the order of 1% (see Figure 2). The accuracy of the procedure is thus appropriate for goodness of fit evaluation, at least in the hydrologic field; as a term of comparison, consider that using the case 0 percentage points in the parametric case can correspond to an error as large as 40% in evaluating the percentage level of the test statistic, with a complete falsification of the test's results.

[20] The second open point regards the power of the test, i.e., its ability to reject the null hypothesis when it is false. Power studies are usually carried out by means of Monte Carlo simulations: a large number of samples is generated from an assumed parent distribution, $F_R$, and the power is determined as the percentage of samples that are rejected in a test for the distribution $F_H$. If $F_H \equiv F_R$, the power should approximate the significance level of the test, while it should be higher when $F_H \neq F_R$. The procedure is repeated with different tests, that are ranked for their power in recognizing deviations from $F_H$. Note that the choice of $F_R$ is subjective,

**Table 3.** Coefficients to Be Set in Equation (11) for the Anderson-Darling Statistic, Asymptotic Case[a]

| Distribution[b] | $\xi_p$ | $\beta_p$ | $\eta_p$ |
|---|---|---|---|
| EV1 and EV2 | 0.169 | 0.229 | 1.141 |
| NORM and LN | 0.167 | 0.229 | 1.147 |
| GEV[c] | $0.147\ (1 + 0.13\ \hat{\theta}_3 + 0.21\ \hat{\theta}_3^2 + 0.09\ \hat{\theta}_3^3)$ | $0.189\ (1 + 0.20\ \hat{\theta}_3 + 0.37\ \hat{\theta}_3^2 + 0.17\ \hat{\theta}_3^3)$ | $1.186\ (1 - 0.04\ \hat{\theta}_3 - 0.04\ \hat{\theta}_3^2 - 0.01\ \hat{\theta}_3^3)$ |
| GAM and LP3[d] | $0.145\ (1 + 0.17\ \hat{\theta}_3^{-1} + 0.33\ \hat{\theta}_3^{-2})$ | $0.186\ (1 + 0.34\ \hat{\theta}_3^{-1} + 0.30\ \hat{\theta}_3^{-2})$ | $1.194\ (1 - 0.04\ \hat{\theta}_3^{-1} - 0.12\ \hat{\theta}_3^{-2})$ |

[a]Here $\hat{\theta}_3$ is an asymptotic efficient estimator (usually maximum likelihood) of the shape parameter of the distribution.
[b]For tests of the EV2, LN, and LP3 distributions the data must be preliminarily log transformed.
[c]For the GEV distribution, if $\hat{\theta}_3 > 0.5$, $\hat{\theta}_3 = 0.5$ must be set in the regressions.
[d]For the GAM and LP3 distributions, if $\hat{\theta}_3 < 2$, $\hat{\theta}_3 = 2$ must be set in the regressions.

**Table 4.** Coefficients to Be Set in Equation (11) for the Cramer–von Mises Statistic, Small Sample Case[a]

| Distribution[b] | $\xi_p(n)$ | $\beta_p(n)$ | $\eta_p(n)$ |
|---|---|---|---|
| EV1 and EV2 | $\xi_p\left(1+\frac{0.2}{n}\right)$ | $\beta_p\left(1+\frac{0.2}{n}\right)$ | $\eta_p\left(1+\frac{0.7}{n}\right)$ |
| NORM and LN | $\xi_p\left(1+\frac{0.3}{n}\right)$ | $\beta_p\left(1+\frac{0.2}{n}\right)$ | $\eta_p\left(1+\frac{0.6}{n}\right)$ |
| GEV | $\xi_p\left(1+\frac{1.5}{n}-\frac{0.3}{\sqrt{n}}\right)$ | $\beta_p\left(1-\frac{1.3}{n}\right)$ | $\eta_p\left(1-\frac{0.7}{n}+\frac{0.3}{\sqrt{n}}\right)$ |
| GAM and LP3[c] | $\xi_p\left(1+\frac{2.1}{n}-\frac{0.3}{\sqrt{n}}-\frac{0.5}{\sqrt{n}\theta_3}\right)$ | $\beta_p\left(1-\frac{0.3}{n}-\frac{0.1}{\sqrt{n}}+\frac{0.1}{\sqrt{n}\theta_3}\right)$ | $\eta_p\left(1-\frac{1.6}{n}+\frac{0.1}{\sqrt{n}}+\frac{0.4}{\sqrt{n}\theta_3}\right)$ |

[a]The values of $\xi_p$, $\beta_p$, and $\eta_p$ should be taken from the corresponding rows in Table 2.
[b]For tests of the EV2, LN, and LP3 distributions the data must be preliminarily log transformed.
[c]For the GAM and LP3 distributions, if $\hat{\theta}_3 < 2$, $\hat{\theta}_3 = 2$ must be set in the regressions.

but it is crucial for the comparison, since different choices usually lead to different rankings of the tests.

[21] To limit the degree of subjectivity, the same set of $F_R$ distributions already utilized by *Wang* [1998] is used here: this is constituted by six Wakeby distributions (WA1, WA2, WA3, WA4, WA5, and WA6), parameterized as by *Landwehr et al.* [1980]. This set of $F_R$ distributions is appealing because, in the intentions of *Landwehr et al.* [1980], the distributions should be able to mimic ''floodlike'' behaviors; moreover, using a Wakeby distribution one is in the case $F_H \neq F_R$ for all of the seven hypothetical distributions (EV1, NORM, GEV, GAM, EV2, LN, and LP3) being tested, thus avoiding the case when the power is trivially equal to the significance level of the test. Kolmogorov-Smirnov (KS), probability plot (PP), L moments (LM), and chi-square ($\chi^2$) tests are considered in the comparison, in addition to the Anderson-Darling and Cramer–von Mises statistics (see Appendix C for details).

[22] The results of the power comparison for significance level $\alpha = 0.05$ and sample size $n = 50$ are reported in Table 6 for all of the 42 combinations of real and hypothesized distributions. Table 6 is constructed using Monte Carlo simulations with 10000 replicates for each of the six Wakeby distributions. The larger value in each row, corresponding to the test achieving best results for each couple ($F_R - F_H$), is highlighted in bold. To facilitate the comparison, the results for the six Wakeby distributions are averaged in Figure 3 in order to obtain a rough indicator of the power of each test against a generic floodlike distribution.

[23] The results of Table 6 and Figure 3 are not easy to interpret, but some points are apparent.

[24] 1. Among EDF statistics, $A^2$ is almost always slightly better than $W^2$, while the commonly used KS statistic tends to be rather weak in power. These results are in good agreement with other power studies, summarized by *Stephens* [1986].

[25] 2. LM statistics show the greater variability in their power results. This is expected, since the power of the test depends on the distance between the L moments ratios of the real and hypothetical distributions: when these are similar, the test cannot be effective. LM statistics can thus be very useful as directional test, e.g., for testing normal against skewed alternatives, or EV1 against GEV [*Fill and Stedinger*, 1995], while other tests should be preferred when the real parent distribution is unknown.

[26] 3. The PP test is on average the most effective in detecting deviations from normality, also with log transformed data, in which it slightly outperforms the $A^2$ test; this is in good agreement with the findings of several other power studies on tests for normality, reviewed by *D'Agostino* [1986]. The situation is reversed ($A^2$ does slightly better than PP) when the hypothetical distribution is EV2 or LP3, while for the EV1 and GAM distributions the advantage of using $A^2$ is substantial. The weakness of the PP test for the GAM distribution was recognized also by *Vogel and McMartin* [1991], and can be attributed to the difficulty of extending to three-parameter distributions a test statistic which arises naturally for distributions with only location and scale parameters.

[27] 4. The $\chi^2$ test is almost always the weaker among the considered statistics, and it also bears the drawback that the percentage points are defined with a large degree of uncertainty (see Appendix C), in particular for small samples and distributions with many parameters (note the differences between the two overlapping white bars in Figure 3). The use of $\chi^2$ statistics should therefore be discouraged when the sample size is small and parameters are estimated, as it is usually the case in the hydrologic field.

[28] 5. When changing the significance level $\alpha$, or the sample size $n$, the power of all tests changes as a consequence, but the overall picture remains rather stable in terms of the relative performances of different tests.

[29] As a final comment, the present power study suggests to use PP tests when the hypothesized distribution is normal or lognormal, and $A^2$ tests in all other cases. However, it must be recalled that a much wider range of parent distributions should be used for drawing exhaustive conclusions from a power study. As for the simplicity of

**Table 5.** Coefficients to Be Set in Equation (11) for the Anderson-Darling Statistic, Small Sample Case[a]

| Distribution[b] | $\xi_p(n)$ | $\beta_p(n)$ | $\eta_p(n)$ |
|---|---|---|---|
| EV1 and EV2 | $\xi_p\left(1+\frac{0.1}{n}\right)$ | $\beta_p\left(1-\frac{0.2}{n}\right)$ | $\eta_p\left(1+\frac{0.5}{n}\right)$ |
| NORM and LN | $\xi_p\left(1+\frac{0.3}{n}\right)$ | $\beta_p\left(1-\frac{0.2}{n}\right)$ | $\eta_p\left(1+\frac{0.5}{n}\right)$ |
| GEV | $\xi_p\left(1+\frac{0.9}{n}-\frac{0.2}{\sqrt{n}}\right)$ | $\beta_p\left(1-\frac{1.8}{n}\right)$ | $\eta_p\left(1-\frac{0.7}{n}+\frac{0.2}{\sqrt{n}}\right)$ |
| GAM and LP3[c] | $\xi_p\left(1+\frac{2.0}{n}-\frac{0.3}{\sqrt{n}}-\frac{0.4}{\sqrt{n}\theta_3}\right)$ | $\beta_p\left(1-\frac{0.5}{n}-\frac{0.3}{\sqrt{n}}+\frac{0.3}{\sqrt{n}\theta_3}\right)$ | $\eta_p\left(1-\frac{1.8}{n}+\frac{0.1}{\sqrt{n}}+\frac{0.5}{\sqrt{n}\theta_3}\right)$ |

[a]The values of $\xi_p$, $\beta_p$, and $\eta_p$ should be taken from the corresponding rows in Table 3.
[b]For tests of the EV2, LN, and LP3 distributions the data must be preliminarily log transformed.
[c]For the GAM and LP3 distributions, if $\hat{\theta}_3 < 2$, $\hat{\theta}_3 = 2$ must be set in the regressions.

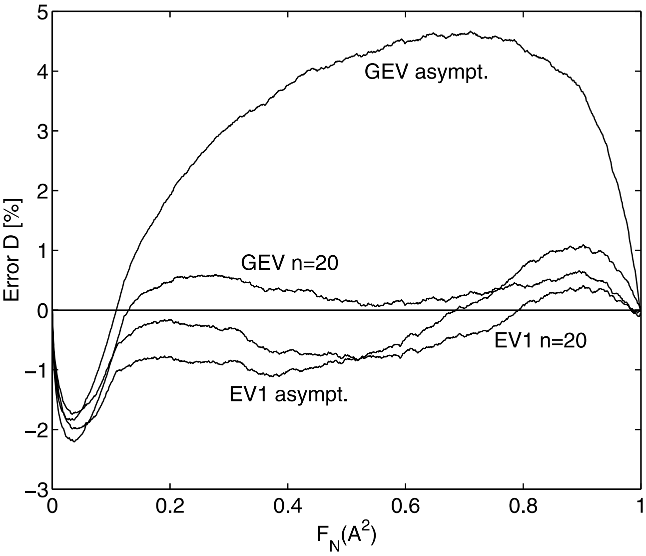


**Figure 2.** Distance $D$, in percent, between the Monte-Carlo, $F_N(A^2)$, and theoretical, $F^*(\omega)$, CDFs of the Anderson-Darling statistics, when the hypothetical distribution is EV1 or GEV with $\theta_3 = -0.2$. $F^*(\omega)$ is calculated either using the asymptotic coefficients from Table 3 or the small sample corrected coefficients from Table 5, considering that the sample size is $n = 20$.

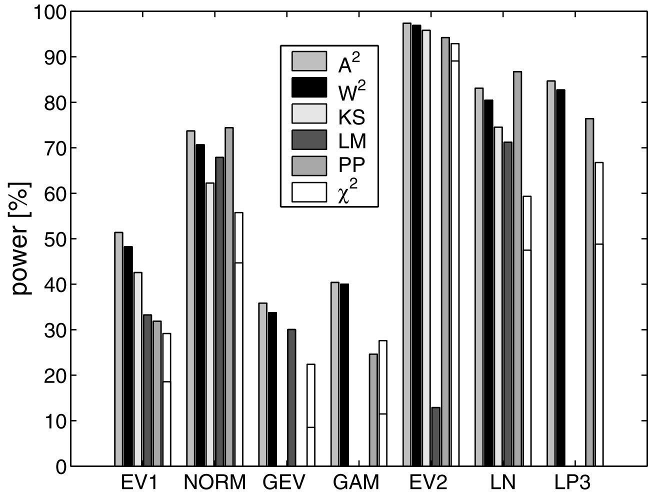


**Figure 3.** Power comparison of goodness of fit tests for extreme value distributions at 5% significance level for a sample size $n = 50$. Each bar represents the average of the powers obtained for the six Wakeby parent distributions in Table 6. The two overlapping white bars correspond to the lower and upper bounds for the power of the $\chi^2$ test.

application of the tests, an advantage of the PP test for two-parameter distributions is that the percentage points are the same for any estimation method that is used [*Vogel*, 1986], while for $A^2$ there can be relevant differences. On the other hand, with PP tests a new table of percentage points is necessary for any hypothetical distribution. Moreover, the transformed $A^2$ statistic, $\omega$, can be used for a direct comparison of the goodness of fit of different distributions, providing an effective method for selecting the most appropriate distribution (the lower $\omega$, the better the fit), while this cannot be done with PP statistics. This point is very important [e.g., *Bobée et al.*, 1993], and a thorough analysis would be needed to determine if $\omega$ is accurate in recognizing the real parent distribution. For lack of space, a detailed analysis of this point is postponed to future works.

## 6. An Example of Application and Final Remarks

[30] In order to summarize the testing procedure proposed in this paper, the application to the annual maxima series of hourly rainfall in Genoa (Italy) is described in full detail. The same data set is used by *Kottegoda and Rosso* [1997, p. 466] for the comparison of testing techniques. The

**Table 6.** Power of Goodness of Fit Tests for Extreme Value Distributions at 5% Significance Level When the True Population is Wakeby and the Sample Size is $n = 50$[a]

| | Real Distribution $F_R$: WA1 | | | | | | Real Distribution $F_R$: WA2 | | | | | | Real Distribution $F_R$: WA3 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $F_H$ | $A^2$ | $W^2$ | KS | LM | PP | $\chi^2$ | $A^2$ | $W^2$ | KS | LM | PP | $\chi^2$ | $A^2$ | $W^2$ | KS | LM | PP | $\chi^2$ |
| EV1 | **53** | 51 | 46 | 50 | 47 | 16–27 | **30** | **30** | 27 | 13 | 22 | 8–16 | **42** | 32 | 26 | 26 | 16 | 8–15 |
| NORM | **97** | 96 | 93 | **97** | 96 | 74–82 | 70 | 68 | 63 | 71 | **76** | 28–39 | 95 | 90 | 79 | **97** | **97** | 73–85 |
| GEV | **41** | 39 | | 17 | | 8–25 | 31 | 31 | | **34** | | 6–21 | **24** | 20 | | 22 | | 3–12 |
| GAM | **64** | 63 | | | 13 | 21–43 | 37 | **38** | | | 23 | 9–25 | **10** | **10** | | | 6 | 2–11 |
| EV2 | **89** | 87 | 85 | 7 | 77 | 70–77 | **99** | **99** | 98 | 6 | 97 | 92–95 | **100** | 99 | 97 | 16 | **100** | 89–95 |
| LN | 67 | 63 | 57 | **88** | 74 | 28–40 | 84 | 83 | 78 | 78 | **88** | 45–57 | 75 | 70 | 60 | 57 | **80** | 26–40 |
| LP3 | 71 | 68 | | | **78** | 27–46 | **89** | 88 | | | **89** | 45–65 | **75** | 72 | | | 53 | 31–53 |

| | Real Distribution $F_R$: WA4 | | | | | | Real Distribution $F_R$: WA5 | | | | | | Real Distribution $F_R$: WA6 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $F_H$ | $A^2$ | $W^2$ | KS | LM | PP | $\chi^2$ | $A^2$ | $W^2$ | KS | LM | PP | $\chi^2$ | $A^2$ | $W^2$ | KS | LM | PP | $\chi^2$ |
| EV1 | **81** | 80 | 76 | 20 | 36 | 40–55 | **26** | 21 | 16 | 5 | 4 | 4–10 | 77 | 75 | 64 | **86** | 65 | 35–52 |
| NORM | **77** | **77** | 68 | 49 | **77** | 40–54 | 81 | 71 | 56 | 86 | **87** | 48–61 | **23** | 21 | 16 | 8 | 14 | 6–13 |
| GEV | **74** | **74** | | 67 | | 27–51 | **27** | 22 | | **27** | | 3–11 | **19** | 17 | | 12 | | 4–16 |
| GAM | **74** | **74** | | | 63 | 26–51 | 18 | 18 | | | 7 | 4–14 | **39** | 37 | | | 35 | 7–22 |
| EV2 | **97** | **97** | 96 | 9 | 92 | 90–93 | **100** | **100** | 99 | 22 | **100** | 94–98 | **100** | **100** | **100** | 17 | **100** | **100** |
| LN | 87 | 85 | 83 | **91** | 90 | 58–68 | 86 | 83 | 71 | 56 | **89** | 40–56 | **100** | 99 | 98 | 57 | **100** | 89–95 |
| LP3 | 90 | 88 | | | **93** | 57–74 | **84** | 81 | | | 58 | 43–66 | **99** | **99** | | | 87 | 89–97 |

[a]Values are in percents. Bold values highlight the test with higher power for each couple of real and hypothetical distributions. Two numbers appear in the $\chi^2$ column since the power of the chi-square test cannot be obtained in an exact manner (see Appendix C for details).

**Table 7.** Example of Application of the Anderson-Darling Test to the Annual Maxima of Hourly Rainfall in Genoa (Italy)

| Distribution | $\hat{\theta}_1$ | $\hat{\theta}_2$ | $\hat{\theta}_3$ | $A^2$ | $\xi_p(n)$ | $\beta_p(n)$ | $\eta_p(n)$ | $\omega$ | $F^*(\omega)$ |
|---|---|---|---|---|---|---|---|---|---|
| EV1 | 38.2 | 15.4 | | 1.09 | 0.169 | 0.228 | 1.151 | 0.80 | 0.99 |
| NORM | 48.2 | 23.6 | | 2.46 | 0.168 | 0.228 | 1.157 | 2.71 | 1 |
| GEV | 35.1 | 12.0 | −0.43 | 0.41 | 0.142 | 0.177 | 1.216 | 0.25 | 0.81 |
| GAM | 22.8 | 21.8 | 1.18 | 0.18 | 0.163 | 0.226 | 1.145 | 0.04 | 0.08 |
| EV2 | 3.57 | 0.35 | | 0.42 | 0.169 | 0.228 | 1.151 | 0.17 | 0.68 |
| LN | 3.77 | 0.43 | | 0.77 | 0.168 | 0.228 | 1.157 | 0.47 | 0.95 |
| LP3 | 3.07 | 0.32 | 2.24 | 0.36 | 0.160 | 0.219 | 1.151 | 0.14 | 0.58 |

measured data from 1931 to 1988 are reported by *Kottegoda and Rosso* [1997, p. 717]. The Anderson-Darling test for the four distributions in Table 1 can proceed as follows: (1) Estimate the parameters from the sample data using ML (columns 2–4 in Table 7) or *Smith*'s [1985] estimators when necessary, as is the case for the GAM distribution in Table 7. (2) Sort the data in ascending order, find $z_i = F(x_i, \hat{\theta})$ for any distribution, and calculate $A^2$ from equation (4) (column 5 in Table 7). (3) Determine the $\mathcal{S}_p(n)$ coefficients using Tables 3 and 5, with the appropriate $n$ and $\hat{\theta}_3$ values (columns 6–8 in Table 7). (4) Find $\omega$ from equation (11), using the $\mathcal{S}_0$ coefficients from equation (13) (column 9 in Table 7). (5) Compare $\omega$ to the appropriate percentage points for the selected significance level (e.g., 0.461 for $\alpha = 0.05$), or use equation (14) to find the probability level $F^*(\omega)$ correspondent to each $\omega$ value (Column 10 in Table 7); if $F^*(\omega) < (1 - \alpha)$, the test is passed. For the LN, EV2, and LP3 distributions, one should preliminarily log transform the data, and then repeat steps 1–5 above. The results in Table 7 show that, in this particular case, four distributions out of seven pass the 5% Anderson-Darling test, with the GAM distribution achieving the lower probability level (best fit). Note that if a case 0 test were erroneously applied to the $A^2$ results in column 5, all distributions had passed the 5% test (the limit value is 2.462).

[31] In conclusion, the Anderson-Darling test for extreme value distributions combines good power properties and simplicity of application, and its use in the hydrologic field can give important indications for model verification and selection. The method proposed here to convert a case p to a case 0 statistic is flexible: it can be applied to other probability distributions of interest; to the case when some of the parameters are known, setting to zero the appropriate elements of the asymptotic dispersion matrix $n\Sigma^{-1}$ in equation (7); or to other quadratic EDF statistics, obtained using different weight functions in equation (2). The major disadvantage of the method is that it requires asymptotically efficient estimators, inhibiting the use of moments and L moments estimators. Note that this problem is not due to the transformation (11), but it stems from the use of equation (6), that is the base for the determination of asymptotic percentage points even in the "classical" approach [*Stephens*, 1986]. A possible solution is currently under investigation, based on a modification of equation (6), mentioned by *Darling* [1955] and *Durbin* [1973], allowing to extend the present method to non-asymptotically efficient estimators.

## Appendix A: The $S_p$ Coefficients for the GAM and GEV Distributions

[32] As mentioned in section 4.2, two problems complicate the determination of the coefficients $S_p$ to be used in equation (11) for the GAM and GEV distributions. First, the asymptotic efficiency of ML estimators is lost when when $\theta_3 < 2$ for the GAM distribution, or $\theta_3 > 0.5$ for the GEV distribution. In these cases one should use *Smith*'s [1985] estimators, defined as follows: for a positively skewed GAM distribution, estimate the location parameter as $\hat{\theta}_1 = x_1$, drop $x_1$ from the sample, define the modified likelihood function $\tilde{L}\,(\theta_2, \theta_3|x, \hat{\theta}_1) = \sum_{i=2}^{n} \ln[f(x_i, \theta|\hat{\theta}_1)]$, and find $\hat{\theta}_2$ and $\hat{\theta}_3$ at a local maximum of $\tilde{L}$. These modified estimators are asymptotically efficient and should be used when classical ML estimators do not exist, or when they yield $\hat{\theta}_3 < 2$.

[33] The case of a negatively skewed gamma distribution ($\theta_2 < 0$) is similar, but there is now an upper bound: the *Smith*'s [1985] estimator of $\theta_1$ is thus $\hat{\theta}_1 = x_n$, and the scale and shape parameters are estimated at the local maximum of $\tilde{L}(\theta_2, \theta_3|x, \hat{\theta}_1) = \sum_{i=1}^{n-1} \ln[f(x_i, \theta|\hat{\theta}_1)]$. For the GEV distribution, ML estimators are still regular when there is a lower bound ($\theta_3 < 0$), since the distribution and its first derivative are null at the bound [*Kendall and Stuart*, 1977, p. 35]; in contrast, when $\theta_3 > 0.5$ (negatively skewed distribution with an upper bound) the asymptotic efficiency of ML estimators is lost. An analogous of *Smith*'s [1985] estimator in this case is obtained by setting $\hat{\theta}_1 = x_n - \frac{\theta_2}{\theta_3}$, and then maximizing the modified likelihood function $\tilde{L}(\theta_2, \theta_3|x, \hat{\theta}_1) = \sum_{i=1}^{n-1} \ln[f(x_i, \theta|\hat{\theta}_1)]$ with respect to $\theta_2$ and $\theta_3$. In all these nonregular cases, one should drop the smallest (or largest) observation from the sample, and calculate $A^2$ from (4) on the remaining $n-1$ points, in order to avoid problems with the logarithms of null quantities.

[34] The second problem with the GAM and GEV distributions is that the $S_p$ coefficients depend on the shape parameter $\theta_3$. To account for this dependence, the case of the GEV distribution is treated as follows: the values of $\xi_p$, $\beta_p$ and $\eta_p$ are determined for any $\theta_3$ in the range [−1; 0.5], in steps of 0.05, and a polynomial of degree 3 is fitted to the data, obtaining the approximate equations in the third line of Tables 2 and 3, with coefficients of determination greater than 0.995. The minimum value of the shape parameter is taken at $\theta_3 = -1$, since for lower values the mean of the distribution diverges; the maximum is set at 0.5 because for $\theta_3 > 0.5$ the estimation of the location parameter becomes superefficient [*Darling*, 1955]: ignorance about $\theta_1$ makes no difference for the estimation of $\theta_2$ and $\theta_3$ [*Smith*, 1985], and for any $\theta_3 > 0.5$ one can use the percentage points (and the $\mathcal{S}_p$ coefficients) that are found for $\theta_3 = 0.5$ [see also *Choulakian and Stephens*, 2001]. Similarly, for the gamma distribution the coefficients in Tables 2 and 3 are found to depend on the powers of $\frac{1}{\theta_3}$, and the range of acceptable values is $0 \leq \theta_3 < \infty$, with the artifice to use the coefficients that are found for $\theta_3 = 2$ in the superefficient case $\theta_3 < 2$. Note that in case p the real value of the shape parameter of

the parent distribution is not known, and an estimate $\hat{\theta}_3$ should be used in Tables 2 and 3, with an unavoidable error due to the $\theta_3$ sample variability.

## Appendix B: Accuracy of the Approximation $\omega = W_0^2$

[35] Two different approaches can be used to investigate the accuracy of the approximation of $\omega$ to $W_0^2$: the first requires to refer to existing tables of percentage points, the second to use Monte Carlo simulations. As an example of the first approach, consider the $A^2$ test for the EV1 distribution: the "exact" 95% asymptotic percentage point is 0.757 [*Stephens*, 1986]. Setting this value in equation (11) with the appropriate coefficients from Table 3 (first row), one obtains $\omega = 0.451$, corresponding to a probability in equation (14) of 94.7%. The 0.3% error in significance level is negligible. Analogous results are obtained for other tabled percentage points (EV1 and NORM distributions [see *Stephens*, 1986]) with maximum errors of 0.5–0.6% in the central part of the test statistics distributions. No tabled results exist for the GAM distribution, while for the GEV a comparison is possible with the results of *Ahmad et al.* [1988] ($A^2$ statistic only). However, *Ahmad et al.*'s [1988] results are less rigorous and general than those for the EV1 and NORM distributions: in fact, their percentage points are obtained by small-size Monte Carlo simulations (with $N$ = 5000), without reference to asymptotic theoretical results; the variability of the percentage points with the shape parameter $\theta_3$ is not accounted for; the range of $\theta_3$ considered in the simulations is limited, [−0.2, 0.2]; L moments rather then ML are used for parameters estimation.

[36] In order to have a convincing verification for the three-parameter distributions, Monte Carlo simulations are again employed: samples of $N$ = 100000 EDF statistics are generated for any distribution, sample size, and shape parameter of interest: n = 10:10:100 for the EV1 and NORM distributions; $n$ = 20, 35, 50, 75, 100 and $\theta_3$ = −1:0.25:1 for the GEV distribution; $n$ = 20, 35, 50, 75, 100 and $\theta_3$ = 1, 1.5, 2, 2.5, 3, 4, 6, 10, 100 for the GAM distribution. For each of these 110 Monte Carlo experiments the empirical CDF of the test statistics, $F_N(Q^2)$, is calculated using (1). The standard error of estimation of $F_N(Q^2)$ is $[F_N(Q^2)(1 - F_N(Q^2))/N]^{0.5}$, i.e., lower than 0.2% when $N$ = 100000. This empirical CDF can be compared with $F^*(\omega)$, obtained from equations (11) and (14), and their distance in % can be calculated as $D = (F_N(Q^2) - F^*(\omega))\ 100$. As an example, Figure 2 shows the results for the Anderson-Darling statistic when the null hypothesis is EV1 or GEV (with $\theta_3 = -0.2$), in both cases with $n$ = 20.

[37] The modified statistic $\omega$ is calculated either using in equation (11) the asymptotic coefficients from Table 3, or the small sample corrected coefficients from Table 5. When the asymptotic coefficients are used, the error in evaluating the probability level can be as large as 4.5% for three-parameter distributions, while it is lower for two-parameter distributions. When instead the small sample corrections are accounted for, the error remains in a band of ±1%, except in the very low part of the distributions, where the distortion is stronger. Figure 2 is representative of the typical behavior that is found when other distributions, shape parameters and sample sizes are considered. Larger errors (maximum 2–2.5% in the central and lower parts of the distribution) are found when the shape parameter falls in the range where estimation is nonregular.

## Appendix C: Goodness of Fit Tests Used in the Power Comparison

[38] The following goodness of fit statistics are considered in the power comparison of section 5.

[39] 1. First is the Kolmogorov-Smirnov (KS) test, mentioned in section 2. Suitable formulas for calculating the test statistic and tables of percentage points are given by *Stephens* [1986] for the EV1 and normal distributions. For the GEV and GAM distributions with all the parameters estimated, the appropriate percentage points are instead not known.

[40] 2. Second are tests based on the comparison of empirical and hypothetical L moments ratios (LM in brief); the appropriate testing procedure and percentage points are given by *Fill and Stedinger* [1995] for the EV1 distribution, by *Stedinger et al.* [1993] for the normal distribution, and by *Wang* [1998] for the GEV distribution. The test is not available for the GAM distribution.

[41] 3. Third are tests using the linearity of the probability plot for measuring the goodness of fit (PP tests). Appropriate critical values for the EV1 and normal distributions are tabled by *Stedinger et al.* [1993]. No such tables exist for the GEV with the three parameters estimated from the sample. For the GAM distribution the testing procedure is described by *Vogel and McMartin* [1991].

[42] 4. Last are tests of chi-square type ($\chi^2$ test). The use of the classical Pearson test requires that the range of $x$ is partitioned in classes; a convenient procedure to avoid arbitrariness and maximize the power entails the choice of $k$ equiprobable classes under the hypothesized distribution, with $k = 2n^{0.4}$ [*Moore*, 1986]. This relation should be used even with small samples, since it is not advisable to set a lower bound on the number of elements in each class (e.g., $n/k \geq 5$) when equiprobable cells are used [*Moore*, 1986, p. 71]. The test statistic distribution in case p is not completely known, since there is a partial recovery of degrees of freedom of the chi-square distribution with respect to the commonly recommended value of $k - p - 1$, when efficient estimators are used [e.g., *Kendall and Stuart*, 1977, p. 455; *Moore*, 1986]. When using maximum likelihood, the critical points fall between those of $\chi^2(k - p - 1)$ and those of $\chi^2(k - 1)$, and not even this can be said when moments or L moments estimators are employed. As a consequence, the power of the test does not take a precise value, and only its lower and upper bounds can be determined from the $\chi^2(k - 1)$ and $\chi^2(k - p - 1)$ distributions.

[43] **Acknowledgments.** This work was carried out under CNR-GNDCI grant 03.00022.GN42. Thanks are owed to P. Claps and L. Ridolfi for helpful suggestions and comments during the course of this research.

## References

Ahmad, M. I., C. D. Sinclair, and B. D. Spurr (1988), Assessment of flood frequency models using empirical distribution function statistics, *Water Resour. Res.*, *24*(8), 1323–1328.

Anderson, T. W., and D. A. Darling (1952), Asymptotic theory of certain goodness-of-fit criteria based on stochastic processes, *Ann. Math. Stat.*, *23*, 193–212.

Bobée, G., G. Cavadias, F. Ashkar, J. Bernier, and P. Rasmussen (1993), Towards a systematic approach to comparing distribution used in flood frequency analysis, *J. Hydrol.*, *142*, 121–136.

Choulakian, V., and M. A. Stephens (2001), Goodness-of-fit tests for the generalized Pareto distribution, *Technometrics*, *43*(4), 478–484.

Chowdhury, J. U., J. R. Stedinger, and L. Lu (1991), Goodness-of-fit tests for regional generalized extreme value flood distributions, *Water Resour. Res.*, *27*(7), 1765–1776.

D'Agostino, R. B. (1986), Tests for the normal distribution, in *Goodness-of-Fit Techniques*, edited by R. B. D'Agostino and A. M. Stephens, pp. 367–420, Marcel Dekker, New York.

Darling, D. A. (1955), The Cramer-Smirnov test in the parametric case, *Ann. Math. Stat.*, *26*, 1–20.

Durbin, J. (1973), Weak convergence of the sample distribution function when parameters are estimated, *Ann. Stat.*, *1*(2), 279–290.

Fill, H. D., and J. R. Stedinger (1995), L-moment and probability plot correlation coefficient goodness-of-fit tests for the Gumbel distribution and impact of autocorrelation, *Water Resour. Res.*, *31*(1), 225–229.

Johnson, N. L., S. Kotz, and N. Balakrishnan (1994), *Continuous Univariate Distributions*, vol. 1, 2nd ed., John Wiley, Hoboken, N. J.

Kendall, M. G., and A. Stuart (1977), *The Advanced Theory of Statistics*, vol. 2, *Inference and Relationship*, 4th ed., Griffin, London.

Kottegoda, N. T., and R. Rosso (1997), *Statistics, probability and reliability for civil and environmental engineers*, McGraw-Hill, New York.

Landwehr, J. M., N. C. Matalas, and J. R. Wallis (1980), Quantile estimation with more or less floodlike distributions, *Water Resour. Res.*, *16*(3), 547–555.

Madsen, H., P. F. Rasmussen, and D. Rosbjerg (1997), Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events: 1. At-site modeling, *Water Resour. Res.*, *33*(4), 747–757.

Moore, D. S. (1986), Tests of chi-squared type, in *Goodness-of-Fit Techniques*, edited by R. B. D'Agostino and A. M. Stephens, pp. 63–96, Marcel Dekker, New York.

Smith, R. L. (1985), Maximum likelihood estimation in a class of nonregular cases, *Biometrika*, *72*(1), 67–90.

Stedinger, J. R., R. M. Vogel, and E. Foufula-Georgiou (1993), Frequency analysis of extreme events, in *Handbook of Hydrology*, edited by D. R. Maidment, chap. 17, McGraw-Hill, New York.

Stephens, M. A. (1976), Asymptotic results for goodness-of-fit statistics with unknown parameters, *Ann. Stat.*, *4*(2), 357–369.

Stephens, M. A. (1977), Goodness-of-fit for the extreme value distribution, *Biometrika*, *64*(3), 583–588.

Stephens, M. A. (1986), Tests based on EDF statistics, in *Goodness-of-Fit Techniques*, edited by R. B. D'Agostino and A. M. Stephens, pp. 97–194, Marcel Dekker, New York.

Vogel, R. M. (1986), The probability plot correlation coefficient test for the normal, lognormal, and Gumbel distributional hypotheses, *Water Resour. Res.*, *22*(4), 587–590.

Vogel, R. M., and D. E. McMartin (1991), Probability plot goodness-of-fit and skewness estimation procedures for the Pearson type 3 distribution, *Water Resour. Res.*, *27*(12), 3149–3158.

Wang, Q. J. (1998), Approximate goodness-of-fit tests of fitted generalized extreme value distributions using LH moments, *Water Resour. Res.*, *34*(12), 3497–3502.

F. Laio, Dipartimento di Idraulica, Trasporti ed Infrastrutture Civili, Politecnico di Torino, Corso Duca degli Abruzzi, 24, 10129 Torino, Italy. (francesco.laio@polito.it)