

NOTE

The Weibull distribution as an extreme value model for transformed annual maxima

Earl Bardsley

*University of Waikato,
Hamilton, New Zealand.
Earl.bardsley@waikato.ac.nz*

Abstract

There is no mathematical difference between theoretical extreme value models of maxima and minima because they link to each other by a simple sign reversal transformation. However, other transformations that change sample maxima to sample minima raise the possibility of alternatives to the generalised extreme value (GEV) distribution for annual maxima, while still maintaining extreme value justification. A general class of transformation is proposed that converts positive annual maxima to lower-bounded minima, then amenable to Weibull extreme value analysis for sufficiently large sample sizes. That is, the Weibull distribution of smallest extremes provides a theoretical extreme value model for the transformed maxima, which holds irrespective of any GEV form of the original maxima. This would apply, for example, to the analysis of reciprocals of discharge or rainfall annual maxima.

A useful feature of the transformation approach is that alternative prediction expressions with extreme value justification can arise for defining event magnitude as a function of return period. These expressions

may result in different hydrological conclusions than from a GEV analysis of maxima. A double exponential transformation is introduced and its prediction function for maxima is noted to have capability to mimic equivalent Type 3 extreme value (EV3) distribution expressions, but without having to introduce an upper bound parameter. The new function gives a good fit to apparent EV3 annual flood maxima recorded from two very different catchments: the Yangtze River in China and the upper Whanganui River in New Zealand.

Keywords

Weibull distribution; extreme value distribution; transformation of maxima; exceedance estimation; Yangtze River; Whanganui River

Introduction

The generalized extreme value (GEV) distribution has found widespread application for annual discharge maxima and exceedance probabilities since its introduction into the environmental literature by Jenkinson (1955). The main elements of extreme value theory originate from a number of fundamental studies including Fisher and Tippett (1928), von Mises (1936), and Gnedenko (1943). Kotz and Nadarajah (2000) give an historic overview. Coles (2001) describes a range of applications of extreme value models.

The present paper does not introduce any new theory but draws attention to the possibility of using existing extreme value theory in an alternative way. The alternative may sometimes offer an advantage over the GEV distribution for application to annual maxima time series.

The approach is concerned with utilising the Weibull distribution, which is the extreme value limit distribution of large-sample minima in the presence of a lower bound. If the lower bound is at zero then the limit distribution is the two-parameter Weibull distribution, which can be parameterised by its cumulative distribution function as:

$$F(v) = 1 - \exp[-(v/\alpha)^c] \quad c > 0, \alpha > 0, v \geq 0 \quad (1)$$

where c and α are shape and scale parameters, respectively.

The Weibull distribution is not to be confused with the Type 3 extreme value (EV3) distribution of largest sample maxima, arising when an upper bound is present. The EV3 cumulative distribution function can be parameterised:

$$H(x) = \exp\{-[(\xi - x)/\sigma]^{1/k}\} \quad k > 0, \sigma > 0, x \leq \xi \quad (2)$$

where ξ , k , and σ are location, shape, and scale parameters, respectively. EV3 random variables are equivalent to sign-reversed random variables from 3-parameter Weibull distributions. The distinction is made here because EV3 distributions are sometimes referenced in the literature as ‘Weibull distributions’, for example by Wang *et al.* (2017).

Because it is the extreme value limit distribution of lower-bounded minima from large samples, it is natural for the Weibull distribution to be utilised for describing river flow minima (see, for example, Gottschalk *et al.*, 2013).

This paper introduces the use of the Weibull distribution as an extreme value model for application to minima that have been created by applying a transformation to maxima. A general class of transformation is defined which enables Weibull distribution parameters to be incorporated into prediction functions for annual maxima. This may result in new prediction expressions with different physical implications than would

be obtained from a GEV analysis. The transformation class is defined in the next section, followed by an illustrative double exponential transformation applied to apparent EV3 annual discharge maxima from the Whanganui River in New Zealand and the Yangtze River in China.

General transformation expression

Define X_1, X_2, \dots, X_N to be a sequence of recorded annual maxima, such as discharge maxima, assumed to be positive-valued, independent random variables. Define the general transformation:

$$V_i = g(X_i) \quad (3)$$

where $g(X)$ is a continuous positive decreasing function of X . The transformation definition given would include, for example, $V_i = 1/X_i$ but not $V_i = -X_i$. Because the X_i are independent random variables then the V_i must be independent random variables also. If the X_i are unbounded above then the V_i variables are bounded below at zero.

The transformation has the effect of enabling extreme value analyses of positive annual maxima in terms of lower-bounded minima. Specifically, if the X_i are unbounded above then for sufficiently large sample sizes the distribution of the V_i will be approximated by a two-parameter Weibull distribution, under general conditions. Such limit Weibull distributions hold independently of the extreme value domain of attraction of the upper tail of the distribution of the original maxima. This independence property avoids consideration of whether maxima might be better described by Gumbel or Type 2 (EV2) extreme value distributions.

Users will prefer to fit prediction functions to recorded annual maxima, in a Gumbel plot for example, as opposed to working with the transformed equivalents. The prediction functions here are therefore expressed in

terms of fitting to annual maxima, while still incorporating the Weibull parameters. The approach is illustrated using a double exponential transformation, which is applied to two apparent EV3 flood maxima data sets.

Double exponential transformation and Weibull prediction function

One instance of Equation 3 is the double exponential transformation:

$$V = \exp[-\exp(X/\beta)] \quad X \geq 0 \quad \beta > 0 \quad (4)$$

where β is a scaling parameter to be estimated as part of fitting a prediction function to annual maxima.

Weibull distributions have no upper bound but Equation 4 defines V within the range $0 < V \leq e^{-1}$. This implies that if a Weibull-based prediction expression gives a good fit to annual maxima then β must be small enough to create a sufficiently small value of α , such that the e^{-1} upper bound to V does not influence the Weibull form of the distribution of the V variables.

The transformation given by Equation 4 is of interest because it has possibilities for alternative extreme value analysis of annual maxima that might otherwise have been fitted with EV3 curves on a Gumbel plot, requiring the introduction of an upper bound parameter.

On a Gumbel plot, the prediction function arising from the Equation 4 transformation can be written in terms of the Weibull distribution parameters of Equation 1 as:

$$x = \beta \ln[-\ln(\alpha) + y/c] \quad y > c[\ln(\alpha) + 1] \quad (5)$$

where y is the Gumbel reduced variate. The transformation causes y to be truncated at the left because the distribution of X is bounded below at zero.

The equivalent EV3 prediction function expression in a Gumbel plot is given by:

$$x = \xi - \sigma \exp(-ky) \quad (6)$$

where ξ , σ and k are defined in Equation 2. The EV3 prediction function increases to the right at a decreasing rate and approaches the upper magnitude limit ξ as y increases.

If required, any prediction expression as a function of y can be written alternatively as a function of return period, R , by using the relation $y = -\ln[-\ln(1-R^{-1})]$.

The prediction functions given by Equations 5 and 6 can appear to be quite similar in Gumbel plots, but with the important difference that the EV3 prediction function requires an upper bound to event magnitude while Equation 5 does not.

With parameters restricted to be non-negative, constrained least-squares data fitting of these two nonlinear functions means that there may be local minima in the fitting space for Equations 5 and 6. This reduces the chance of finding the optimal fit if starting from poor initial parameter estimates. Fortunately, a simple graphical fitting approach can provide good initial estimates for input into constrained least squares fitting routines (see Appendix).

The fitting procedure relating to the transformation given by Equation 4 is only workable for apparent EV3 data that indicate a decreasing rising gradient on a Gumbel plot. Otherwise, least squares estimates may not be possible and would give a bad fit in any case.

It is convenient on prediction function graphical plots to also display an indication of the variability of future magnitudes anticipated over the coming N years, assuming environmental conditions remain constant. The upper 0.95 quantile for given N is obtained for the EV3 case by solving for x in the expression:

$$H(x)^N = 0.95 \quad (7)$$

where $H(x)$ is defined by Equation 2.

The equivalent upper quantiles expressed in terms of Weibull parameters are also easily obtained by noting that transformations as defined by Equation 3 reverse the magnitude order. An upper quantile is derived by applying the transformation inverse to the corresponding Weibull distribution lower quantile for the distribution of the smallest value in samples of size N . If a Weibull distribution has scale parameter α and shape parameter c , then the distribution of the smallest value in samples of size N is also a Weibull distribution, with scale parameter $\alpha N^{-1/c}$ and shape parameter c . Therefore, for samples of size N , a Weibull lower p -quantile v for distributions of sample minima is obtained from:

$$v = \alpha N^{-1/c} [-\ln(1-p)]^{1/c} \quad (8)$$

With respect to the inverse of the specific transformation given by Equation 4, the corresponding upper quantile x with respect to the largest event in N years is then obtained by inserting v into:

$$x = \beta \ln[-\ln(v)] \quad (9)$$

For example, the upper 0.95 quantile x would be calculated by setting p to 0.05 in Equation 8. Such quantiles are only indicative of the variability of the largest events in N years, because it is assumed that the distribution of annual maxima is known exactly. The upper quantiles would be somewhat higher if parameter estimation error is also taken into account.

Examples

Figure 1 plots annual discharge maxima from the upper Whanganui River, New Zealand. Figures 1a and 1b show the least-squares fits to this data by the Equation 5 and 6 prediction functions, respectively. The resulting Equation 5 parameter estimates are $\beta = 48.00 \text{ m}^3\text{s}^{-1}$, $\alpha = 0.162$, and $c = 2.305$.

The Equation 6 parameter estimates are $\sigma = 53.12 \text{ m}^3\text{s}^{-1}$, $k = 0.221$ and $\xi = 81.77 \text{ m}^3\text{s}^{-1}$.

The two prediction functions shown in Figure 1 give very similar fits to the Whanganui annual maxima, to the extent that it would not be possible to give a data-based rejection of one in favour of the other. There are, nonetheless, differences in the forms of the 0.95 upper quantiles. The EV3 prediction function upper quantile is smaller and shows the anticipated convergence

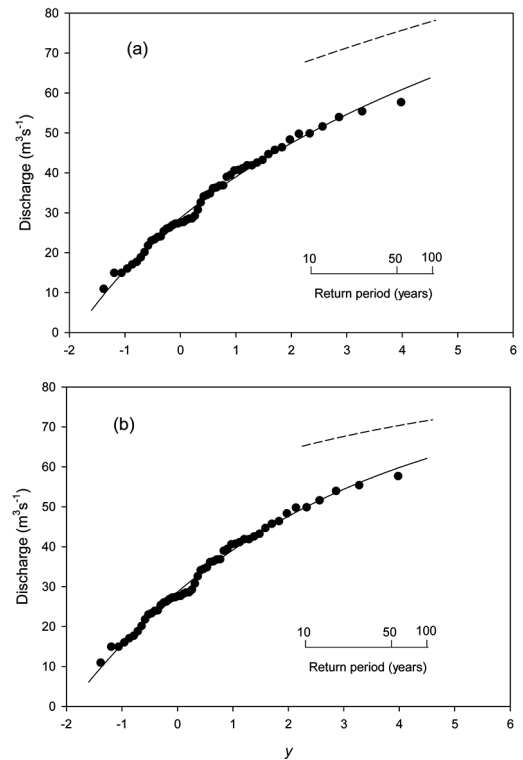


Figure 1 – Gumbel plots of least squares fits of prediction functions to annual discharge maxima recorded from the Whanganui River at Te Porere, 1966–2018: (a) Fitted Weibull prediction function as given by Equation 5; (b) Fitted EV3 prediction function as given by Equation 6. Dashed lines give the 0.95 upper discharge quantiles for return periods 10–100 years (see text). The Weibull plotting position formula is used here and in subsequent figures.

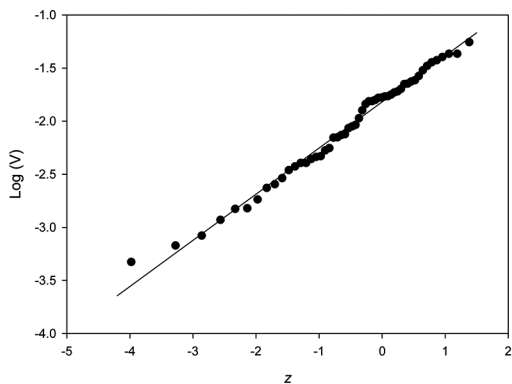


Figure 2 – Weibull plot of the V values obtained from transforming the Whanganui annual maxima of Figure 1, using Equation 4 with $\beta = 48 \text{ m}^3\text{s}^{-1}$. The z axis is the Weibull variate $\ln\{-\ln[1-F(z)]\}$. The straight line is a plot of $\ln(v) = \ln(\alpha) + c^{-1}z$, using $\alpha = 0.162$ and $c = 2.305$. All parameters are obtained from the least squares fit shown in Figure 1a. The largest annual maximum corresponds to the leftmost plotted value.

toward the $81.77 \text{ m}^3\text{s}^{-1}$ upper bound as y increases.

Figure 2 plots the transformed V data, giving a graphical linearity check of the goodness of fit of the Weibull model. As expected from the fit in Figure 1a, the Weibull distribution gives a reasonable approximation. However, the transformed largest flood plots a little away from the predicted Weibull line.

The second example uses annual discharge maxima from the Yangtze River at Yichang (China), the same data set used by Sutcliffe (1987). Previous analyses have established good fits of the EV3 distribution to this data set, most recently Wang *et al.* (2017).

The least squares fit of Equation 5 to the Yangtze data gives the prediction curve shown in Figure 3a, with parameter estimates $\beta = 24.13 \times 10^3 \text{ m}^3\text{s}^{-1}$, $c = 0.36$, $\alpha = 0.0005$. The transformation in this instance gives a Weibull shape parameter indicating the Weibull distribution of V has a mode at zero. In contrast, the transformed Whanganui

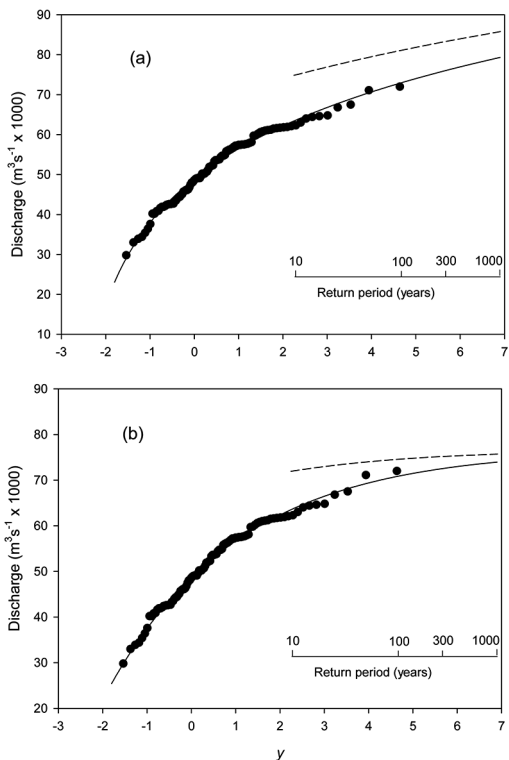


Figure 3 – Gumbel plot of least squares fits of prediction functions to Yangtze River at Yichang annual discharge maxima 1877–1981: (a) Fitted Weibull prediction function as given by Equation 5; (b) Fitted EV3 prediction function as given by Equation 6. Dashed lines give the 0.95 upper discharge quantiles for return periods 10–1000 years (see text).

maxima distribution mode was greater than zero ($c > 1$). Analogous to Figure 2, the plot of the transformed data in Figure 4 indicates a good approximation by the Weibull distribution.

Figure 3b shows the least squares EV3 fit of the Equation 6 prediction function, giving the EV3 parameter estimates as $\sigma = 28.06 \times 10^3 \text{ m}^3\text{s}^{-1}$, $k = 0.221$, and $\xi = 76.73 \times 10^3 \text{ m}^3\text{s}^{-1}$. Wang *et al.* (2017) used a formal GEV estimation method for the Yichang site data, which included some more recent annual maxima not available to the wider scientific community. As it happened,

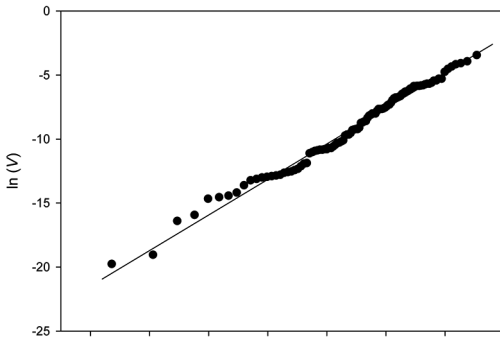


Figure 4 – Weibull plot of the V values obtained from transforming the Yangtze annual maxima of Figure 3 using Equation 4 with $\beta = 24.13 \times 10^3 \text{ m}^3\text{s}^{-1}$. The z axis is the Weibull variate $\ln\{-\ln[1-F(v)]\}$. The straight line is a plot of $\ln(v) = \ln(\alpha) + c^{-1}z$, using $\alpha = 0.0005$ and $c = 0.36$. All parameters are obtained from the least squares fit shown in Figure 3a. The largest annual maximum corresponds to the leftmost plotted value.

the upper bound obtained from their data fit is $76 \times 10^3 \text{ m}^3\text{s}^{-1}$, essentially identical to the ξ estimate from the least squares fit of Equation 6.

As with the Whanganui data, there is little difference between the two prediction functions within the data range (Fig. 3a and 3b). However, there is an evident difference when extrapolating to the 1000-year return period. This is because the Weibull function is unbounded above, giving a return period of 286 years to the supposed EV3 discharge upper bound of $76 \times 10^3 \text{ m}^3\text{s}^{-1}$. There is an even greater difference between the respective 0.95 quantiles because the EV3 quantile, in this case, is constrained by the proximity of the discharge upper bound.

Discussion

For the general class of transformation defined by Equation 3, the question arises as to the interpretation of a set of V values which are approximated by a Weibull distribution. It is known from extreme value

theory that for sufficiently large samples from lower-bounded variables, the sample minima will be distributed approximately as Weibull random variables. Matching the V values to a Weibull distribution might then be taken to mean that the transformation has enabled convergence to the true limit Weibull distribution of minima. On the other hand, fitting a Weibull-based prediction function to annual maxima corresponds to seeking to convert the transformed annual maxima to Weibull random variables. A matching of the V set to a Weibull distribution could therefore be seen as only a fitting exercise.

However, the same situation holds when fitting the GEV distribution to annual maxima. That is, the true asymptotic limit for maxima may have been achieved, or there may just be a fortuitous data fit by the flexible GEV distribution. For example, fitting the Yangtze annual maxima with the Weibull-based prediction function of Equation 5 involves optimising three parameters, as does fitting the EV3 prediction function given by Equation 6.

Figure 5 illustrates that matching a distribution to data, transformed or otherwise, need not imply achieving the theoretical requirements for the distribution concerned. The two example sets of sample maxima are well fitted respectively by extreme value distributions, to the extent that both would pass any goodness-of-fit test. As it happens, the EV2 and EV3 fitted distributions are both incorrect because the true limit distribution of maxima is a Gumbel distribution in both cases.

The illusionary EV2 and EV3 fits arise because maxima of samples from normal distributions, and from Weibull distributions with small shape parameter values, converge only slowly to the true Gumbel limit as sample size increases. Sub-asymptotic distributions of maxima arising from real world moderate sample sizes might then be fortuitously, and incorrectly, matched

by EV2 or EV3 distributions. In the same way, apparent EV2 forms for annual rainfall maxima might be sub-asymptotic forms on the way to a different limit distribution of extremes. Further discussion is given by Furrer and Katz (2008).

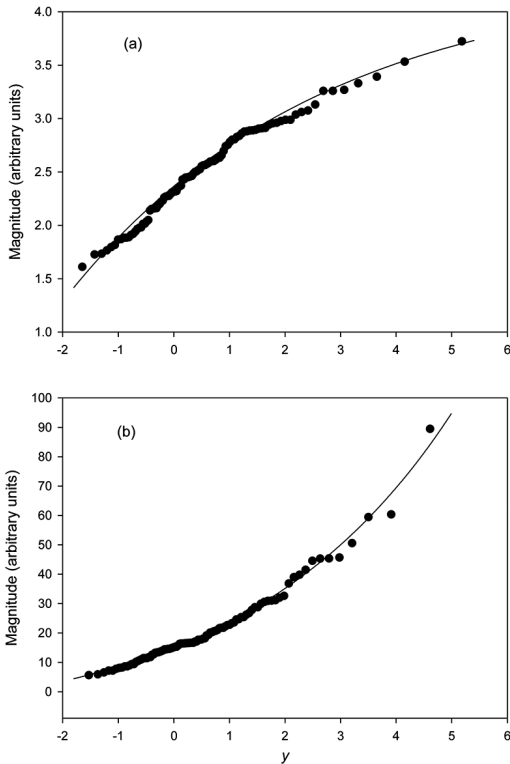


Figure 5 – Gumbel plots of sample maxima from 100 simulated samples of size 50: (a) Maxima of samples simulated from the standard half-normal distribution, with a fitted EV3 curve; (b) Maxima of samples simulated from a Weibull distribution ($\alpha = 1$, $c = 0.5$), with a fitted EV2 curve.

For the situation of having to choose between the Weibull and EV3 prediction functions given by Equations 5 and 6, respectively, Equation 5 is to be preferred when equally good fits are achieved for both. From an engineering viewpoint, this is for the empirical reason that Equation 5 gives a greater factor of safety when estimating design magnitudes. In terms of hydrological science,

the use of Equation 5 avoids the awkward issue of requiring an exact physical upper bound with zero exceedance probability.

There are any number of transformation expressions consistent with Equation 3. The reciprocal transformation $V = 1/X$ is of theoretical interest because if X is a 2-parameter EV2 random variable then $1/X$ is distributed as a 2-parameter Weibull random variable. In the context of the present paper, this could be interpreted that an EV2 distribution of sample maxima implies sample sizes large enough for the minima of the reciprocals of the sample members to follow a 2-parameter Weibull distribution.

The focus here has been on the double exponential transformation to illustrate the method and because it could find use as an EV3 substitute. Other transformations may find application in some data situations where the GEV does not provide a good fit. As noted earlier, there is no new extreme value theory involved in the present paper, but rather application of standard extreme value theory in a different way.

Conclusion

Applications of univariate extreme value theory in environmental sciences and engineering design have been typically with reference to data in the original measurement scale. Outside of the uninformative sign change transformation, transforming positive-valued sample members such that the resulting sample minima approximate the Weibull stable distribution of minima appears not to have been employed previously in practical situations. In reality, such transformations may or may not indicate convergence to the limit Weibull distribution of smallest extremes. However, the method could find application as an extreme value model to give a different interpretation to apparent EV3 data, or to provide alternative extreme value predictive functions for

situations where the GEV does not provide a good data fit.

Acknowledgements

The Whanganui data of Figure 1 was provided by New Zealand National Institute of Water and Atmospheric Research. The data can be obtained on request from Enquiries@niwa.co.nz. This paper has had a long evolution to its present form and useful reviewer comments on earlier versions are noted with gratitude. Interested readers may wish to see prior versions in HESSD discussions.

References

- Bardsley, W.E. 1989: Graphical estimation of extreme value prediction functions. *Journal of Hydrology* 110: 315–321.
- Coles, S.G. 2001: An introduction to statistical modeling of extreme values. Springer, London.
- Fisher, R.A.; Tippett, L.H.C. 1928: Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society* 24: 180–190.
- Furrer, E.M.; Katz, R.W. 2008: Improving the simulation of extreme precipitation events by stochastic weather generators. *Water Resources Research* 44: W12439.
- Gnedenko, B.V. 1943: Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of Mathematics* 44: 423–53.
- Gottschalk, L.; Yu K-x.; Leblois E.; Xiong, L. 2013: Statistics of low flow: Theoretical derivation of the distribution of minimum streamflow series. *Journal of Hydrology* 481: 204–219.
- Jenkinson, A.F. 1955: The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society* 81: 158–71.
- Kotz. S.; Nadarajah, S. 2000: Extreme value distributions: theory and applications. Imperial College Press, London.
- Sutcliffe, J.V. 1987: The use of historical records in flood frequency analysis. *Journal of Hydrology* 96: 159-171.
- von Mises, R. 1936: La distribution de la plus grande de n valeurs. [Reprinted in Selected Papers II: 271–294, *American Mathematical Society*, Providence, RI, 1954]
- Wang, J.; Lu, F.; Lin, K.; Xiao, W.; Song, X.; He, Y. 2017: Comparison and evaluation of uncertainties in extreme flood estimations of the upper Yangtze River by the Delta and profile likelihood function methods. *Stochastic Environmental Research and Risk Assessment* 31: 2281–2296.

Appendix: Obtaining initial parameter estimates by 3-point graphical fitting

It is helpful to obtain initial parameter estimates not too far removed from the least squares estimates, to reduce the possibility of least squares routines converging to local minima.

For the EV3 predictive function given by Equation 6, a simple three-point subjective fitting process can be applied (Bardsley, 1989).

For the Weibull predictive function given by Equation 5, an equivalent 3-point procedure is described here. Given a plotting position expression, the annual maxima data is displayed on a Gumbel plot (Gumbel reduced variate y on the horizontal axis and magnitude x on the vertical). Define three x, y points on the plot such that $x_1 < x_2 < x_3$ and $y_1 < y_2 < y_3$. Any three points on the plot define a prediction curve from Equation 5 because the curve must pass through all three points, which do not need to be coincident with data points. Some trial variations of the selected three points will be needed to give apparent best visual match between the curve and the plotted annual maxima data points.

Given a curve with satisfactory fit, β is estimated as that value β^* which gives the solution of:

$$\exp(x_2 / \beta^*) + [\exp(x_2 / \beta^*) - \exp(x_1 / \beta^*)](y_3 - y_2) / (y_2 - y_1) - \exp(x_3 / \beta^*) = 0 \quad (A1)$$

The c and α estimates are then obtained from:

$$c^* = (y_2 - y_1) / [\exp(x_2 / \beta^*) - \exp(x_1 / \beta^*)] \quad (A2)$$

$$\alpha^* = \exp[y_1 / c^* - \exp(x_1 / \beta^*)] \quad (A3)$$

For example, in Figure 1a, the utilised three points had x -coordinates 17, 47, and 59 m^3s^{-1} , for x_1, x_2 , and x_3 , respectively. The y coordinates were -1, 2, and 4, for y_1, y_2 , and y_3 , respectively. From Equations A1 to A3, these three points give the Equation 5 initial parameter estimates as $\beta^* = 39.6 \text{ m}^3\text{s}^{-1}$, $c^* = 1.72$, and $\alpha^* = 0.120$. Starting with these initial estimates, the Matlab constrained least squares routine *lsqcurvefit* gave the least squares estimates: $\beta = 48.00 \text{ m}^3\text{s}^{-1}$, $c = 2.305$, $\alpha = 0.162$.

The curve fitting approach considered will inevitably have some sensitivity to the choice of plotting position expression. However, this does not appear to be a major factor for Equation 5 fits. For example, with Weibull plotting positions the Whanganui 100-year flood magnitude is estimated via least squares as $64 \text{ m}^3\text{s}^{-1}$, compared to $62 \text{ m}^3\text{s}^{-1}$ for Gringorten plotting. For the Yangtze data, the return period flood magnitudes for 10, 50, 100 and 1000 years were estimated from least squares with Weibull plotting as 63, 70, 73, and $79 \times 10^3 \text{ m}^3\text{s}^{-1}$, respectively. The corresponding magnitudes with Gringorten plotting were 63, 70, 72, and $79 \times 10^3 \text{ m}^3\text{s}^{-1}$.