Cautionary note on multi-component flood distributions for annual maxima

Associate Professor William Earl Bardsley
Faculty of Science and Engineering
University of Waikato
Private Bag 3105
Hamilton 3240
New Zealand

Phone (Office): +64-7-8585011

E-mail: e.bardsley@waikato.ac.nz

Abstract

Multi-component probability distributions such as the two-component Gumbel distribution are sometimes applied to annual flood maxima when individual floods are seen as belonging to different classes, depending on physical processes or time of year. However, hydrological inconsistencies may arise if only non-classified annual maxima are available to estimate the component distribution parameters. In particular, an unconstrained best fit to annual flood maxima may yield some component distributions with a high probability of simulating floods with negative discharge. In such situations multi-component distributions cannot be justified as an improved approximation to a local physical reality of mixed flood types, even though a good data fit is achieved. This effect usefully illustrates that a good match to data is no guarantee against degeneracy of hydrological models.

Key words: flood maxima; multi-component distribution; parameter error, two-component extreme value distribution

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/hyp.10886

This article is protected by copyright. All rights reserved.

Accepted Article

1. INTRODUCTION

It is often convenient to classify individual floods into different subsets to reflect different causal effects. For example, different flood types might be modelled by different probability distributions depending on runoff mechanisms. Whatever the classification, if it happens that *N* different flood types occur as independent events in every year, then the annual maxima may be viewed as the largest of *N* random variables, being themselves the largest magnitudes of each of the *N* flood types in a given year.

If $F_i(x)$ denotes the distribution function of the annual maxima of the *i* th flood type, then the distribution function of the annual maxima without regard to type can be written:

$$G(x) = \prod_{i=1}^{N} F_i(x) \tag{1}$$

Probably the best-known flood distribution of this form is the two-component distribution introduced by (Rossi et al., 1984). In this case the annual maxima were viewed as the larger of two maxima, which themselves were the largest members from two samples drawn from two different exponential distributions, with the sample sizes being Poisson random variables. In that particular model the Poisson component (with nonzero probability of zero floods) implies that some years might have only one flood type, and some years may have no floods.

When applying Equation (1), the different flood types may be identifiable by time of year or by observed runoff mechanism so the parameters of the component distributions can be estimated sequentially using the recorded discharges of the respective flood types (Waylen and Woo, 1982; Gioia et al., 2008, Strupczewski et al., 2012). However, in the event of non-identifiability it is necessary to estimate the component parameters simultaneously from the annual maxima record alone, even though it is recognised that different flood types do in fact occur within a given year.

The purpose of this brief communication is to illustrate via the two-Gumbel distribution that fitting *G(x)* from annual maxima alone may result in some component flood distributions having non-hydrological properties. This arises when the maxima cannot be categorized

Accepted Article

This article is protected by copyright. All rights reserved.

individually for whatever reason. The effect should be obvious normally but sometimes could be hidden to some degree if an alternative parameterisation is employed.

2. THE TWO-GUMBEL DISTRIBUTION

The two-Gumbel distribution is defined here as the distribution of the maxima of pairs of independent random variables, respectively arising from two different Gumbel distributions. From Equation (1), the two-Gumbel cumulative distribution function can be written:

$$G(x) = \prod_{i=1}^{2} \exp\{-\exp[-(x-\xi_i)/\theta_i]\} \qquad \theta > 0 \tag{2}$$

where ξ_i and θ_i are respectively the modal values and scale parameters of the two Gumbel component distributions. An alternative parameterisation is:

$$G(x) = \prod_{i=1}^{2} \exp[-\lambda_i \exp(-x/\theta_i)] \tag{3}$$

where $\lambda_i = \exp(\xi_i / \theta_i)$.

Two-Gumbel distribution can arise as limit distributions of the two-component model of Rossi et al., (1984), as defined by their Equation (15). A requirement here is that the two Poisson expected values are both sufficiently large. More generally, the two-Gumbel distribution will arise when individual floods of the two different types occur in every year and there must be sufficiently large numbers of both types per year to enable the asymptotic extreme value conditions to apply for their respective annual maxima. In addition, the probability distributions of the two flood types must both be in the extreme value domain of attraction of the Gumbel distribution.

Accepted Article

This article is protected by copyright. All rights reserved.

3. POTENTIAL ANOMALIES FROM DATA FITTING

Two-Gumbel parameter estimation procedures via maximum likelihood and least-squares are given respectively by Rossi et al., (1984) and Canfield (1979). It is conceptually possible, however, for flood models structured as per Equation (1) to yield good fits to data from parameter estimation while at the same time being inconsistent with hydrology.

The potential for inconsistency is illustrated in Figure 1a, showing a plot of two hypothetical Gumbel component distributions and the associated probability distribution of the complete two-component distribution. The two-component distribution in fact differs in this case only slightly from component distribution with the larger mean because the maxima of the pair of variables will in this case very often come from the component distribution with the larger mean. However, for the smaller-valued maxima the component distribution here with the smaller mean may sometimes provide the maxima.

In this hypothetical example the influence of the smaller-mean component distribution could result in a slightly better fit to the smallest recorded annual maxima. However, the price to be paid is that the modal value of the smaller-mean component flood distribution is located at zero discharge, which has no hydrological interpretation.

Figure 1b shows the corresponding cumulative distribution functions. There is evidently a relatively high probability (0.37) that a negative discharge value will arise from a random flood event generated from the component flood distribution with lower mean. However, the model as a whole has only a very small probability of generating a negative flood discharge and so might yield a good fit to annual maxima.

A similar effect is shown from a set of recorded flood data in Figures 2 and 3, as a result of least-squares fitting the two-Gumbel distribution of Equation (2) to a set of uncategorized annual flood maxima from the Orari River in New Zealand. Again, the modal value of the smaller-mean component flood distribution is close to zero discharge. The component distribution with the smaller mean provides some extra flexibility here through slightly better fitting the smaller annual maxima, but at the expense of loss of hydrological reality.

Accepted Article

This article is protected by copyright. All rights reserved.

This type of anomalous fitting result should normally be obvious from unexpected estimated parameter values. However, if the parameterisation of Equation (3) was employed in a fitting process then it might not be appreciated that a resulting parameter combination could imply one of the component distributions is inconsistent with hydrology. The present paper was motivated in fact by such a condition becoming evident in the course of a review a flood analysis report. The two-Gumbel distribution could still be employed in such circumstances as a flexible four-parameter empirical flood distribution. However, it would be incorrect to assume at the same that it had physical justification based on a supposed better representation of a two-component hydrological reality.

3. CONCLUSION

It would be desirable to check previous two-Gumbel fits applied to unclassified annual maxima to ensure the component flood distributions have hydrological consistency. In fact, this is unlikely to involve many past analyses. The message here is more to do with the philosophy of hydrological modelling generally. Any multi-component hydrological model is only valid if all its components are valid. If the process of fitting to data causes unrealistic parameterisation of a model component then the hydrological model as a whole fails, even if it gives excellent matching to data.

ACKNOWLEDGEMENTS

The Orari River annual maxima data record was kindly provided by the Environment Canterbury Regional Council, New Zealand. Thanks go to two anonymous reviewers for helpful modifications to the original manuscript.

Accepted Article

This article is protected by copyright. All rights reserved.

REFERENCES

Canfield RV. 1979. The distribution of the extreme of a mixture random variable with applications in hydrology, in Input for Risk Analysis in Water Systems, ed McBean EA, Hipel KW, Unny TE, p77-84, Water Resources Publications, Fort Collins, Colorado.

Gioia A, Iacobellis V, Manfredo S, Fiorentino M. 2008. Runoff thresholds in derived flood frequency distributions. Hydrology and Earth System Sciences 12: 1295–1307.

Rossi FR, Fiorentino M, Versace P. 1984. Two-component extreme value distribution for flood frequency analysis. Water Resources Research 20: 847–856.

Strupczewski WG, Kochanek K, Bogdanowicz E, Markiewicz, I. 2012. On seasonal approach to flood frequency modelling. Part I: Two-component distribution revisited. Hydrological Processes 26: 705–716.

Waylen P, Woo MK. 1982. Prediction of annual floods generated by mixed processes. Water Resources Research 18: 1283–1286.

Accepted Article

This article is protected by copyright. All rights reserved.

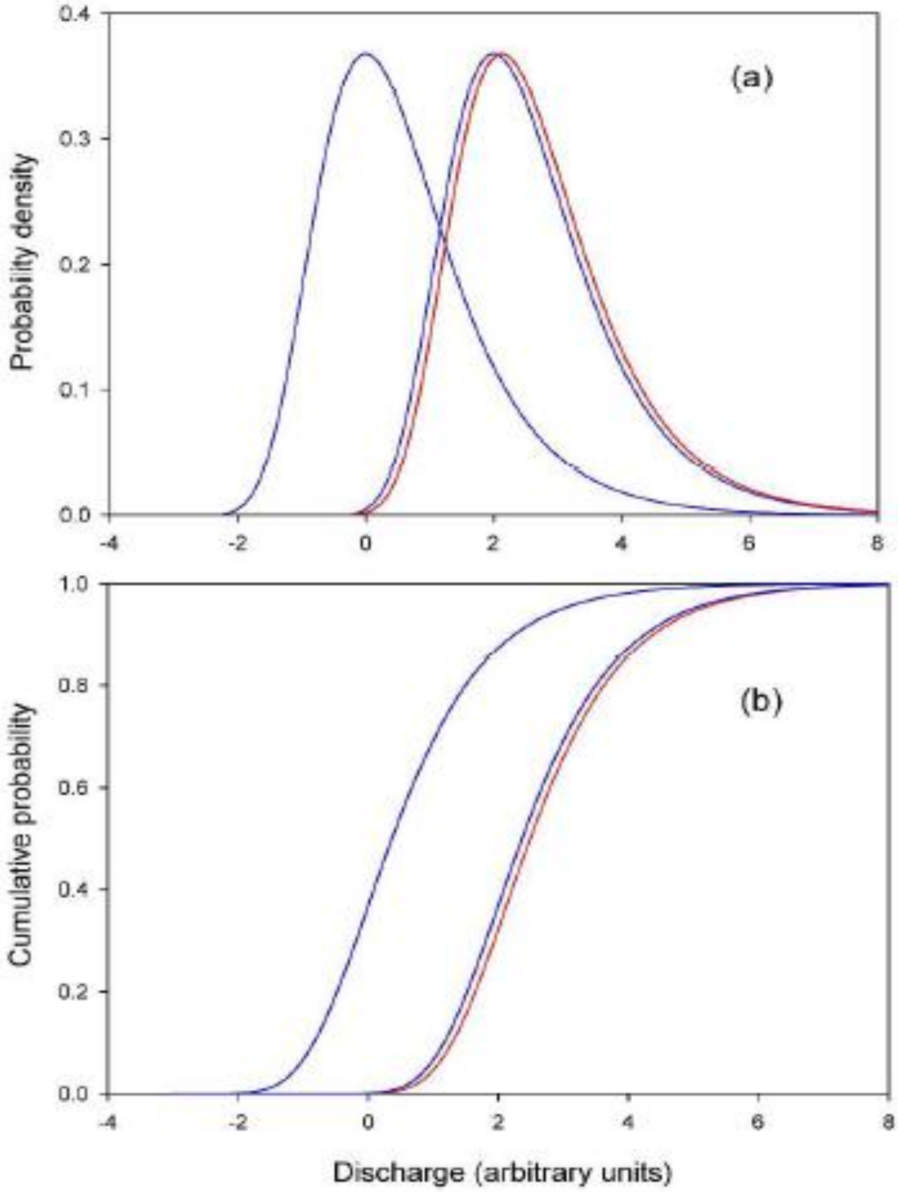

Figure 1. Plot of a two-Gumbel distribution, with component distribution parameters respectively $\xi_1 = 0$, $\theta_1 = 1$, and $\xi_2 = 2$, $\theta_2 = 1$, as defined in Equation (2). Probability density functions and cumulative distribution functions are shown in (a) and (b) respectively. The component distributions are graphed in black. The red plots denote the distribution for maxima of pairs if random variables from the two component distributions.

Accepted Article

This article is protected by copyright. All rights reserved.

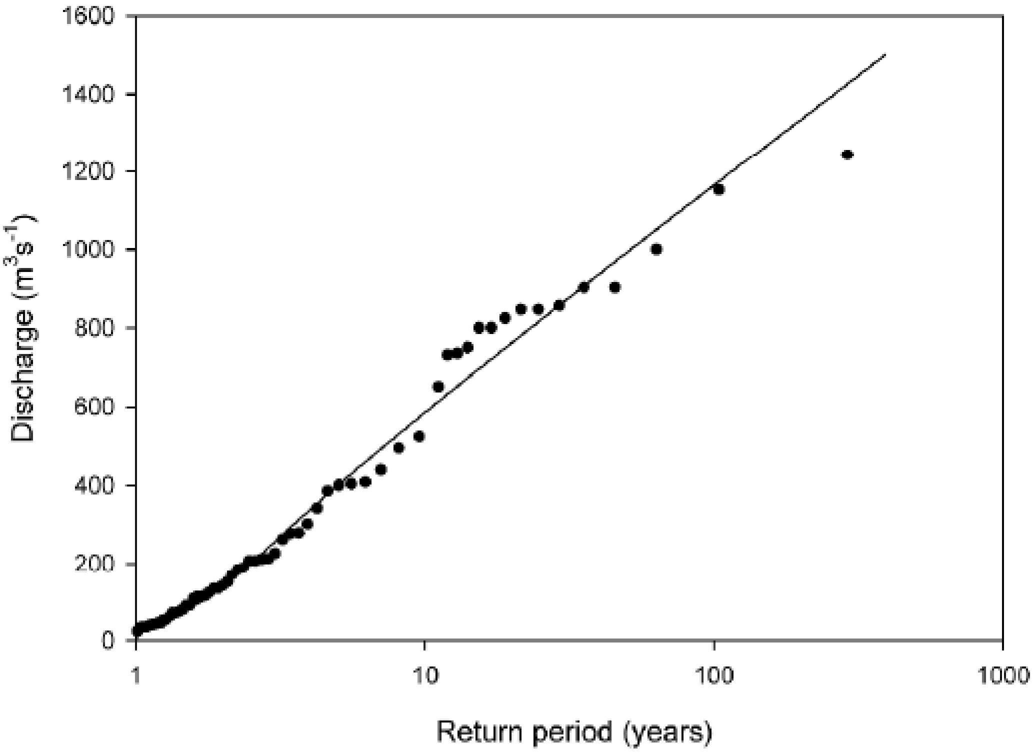

Figure 2. Least-squares fit of the two-Gumbel distribution to 63 annual maxima of the Orari River, Canterbury, New Zealand. The data includes both recorded and historic floods.

Accepted Article

This article is protected by copyright. All rights reserved.

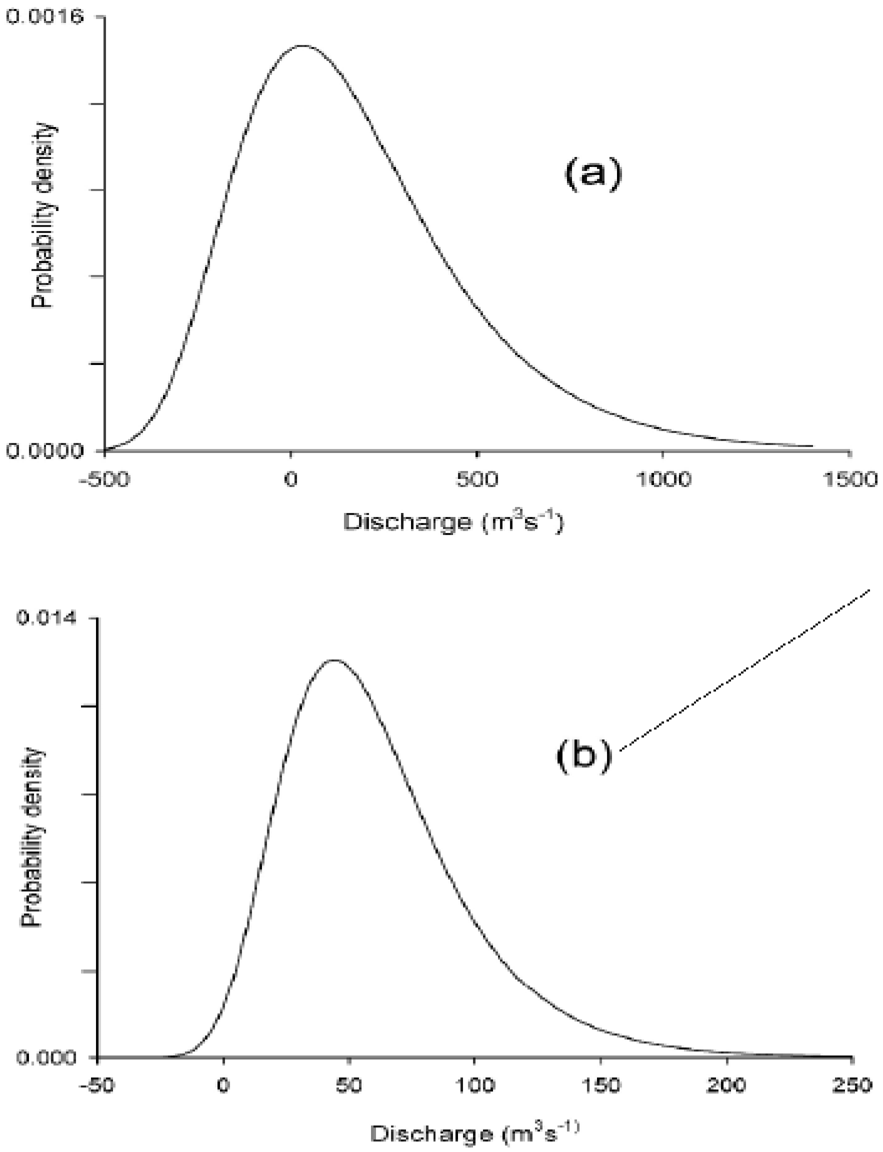

Figure 3. Plot of the two component Gumbel distributions from the component parameter values obtained from the fit to annual maxima shown in Figure 2.

Accepted Article

This article is protected by copyright. All rights reserved.