



An Interactive Graphic for Visually Fitting the Generalized Extreme Value Distribution

Software Introduction

Viacheslav Paramonov¹, Earl Bardsley²

¹*Matrosov Institute for System Dynamics and Control Theory, Russian Academy of Sciences (Siberian Branch), p/b 292, 134 Lermontov str., Irkutsk, 664033 Russia, slv@icc.ru*

²*Faculty of Science and Engineering, University of Waikato, Private Bag 3105, Hamilton 3240, New Zealand, earl.bardsley@waikato.ac.nz*

Abstract

A utility developed for the Windows operating system enables easy visual curve fitting of the generalized extreme value distribution (GEV), from a screen display of annual maxima in a Gumbel plot. This gives a useful indication of how positioning of curves passing through the data points may result in similar or different return periods for large events. The utility employs a general risk-based time axis, with the usual return period being a special case option.

Highlights

- An interactive graphic, applicable to teaching, is provided for visual GEV curve fitting
- GEV parameters are obtained from visual curve fitting to data
- An adjustable risk-based time scale is incorporated, with return period as a special case

Keywords: Visual fitting; generalized extreme value distribution; generalized time scale

1.0 Introduction

Fitting probability distributions to data histograms is perhaps the most basic form of annual maxima modelling. In essence, those few parameters that define the probability distribution serve as a substitute for the larger number of data points. Extrapolating the fitted distribution beyond the data range then allows some indication of the probability of large events, to a greater or lesser degree of accuracy.

The usual approach to data fitting is by way of using formal statistical methods to objectively estimate the unknown parameters of the distribution concerned. However, there is also value in a process of visual curve fitting

because the user may be familiar with individual data point reliability and would like the freedom of subjective weighting in the curve-fitting process.

The purpose of this paper is to draw attention to a graph-based software utility which avoids mathematical estimation in distribution fitting. It is deliberately subjective, enabling decisions in the fitting process as to which data points may be less reliable. The approach also gives an intuitive feel for the magnitude of estimation error without needing the intermediary step of considering parameter error.

The GEV interactive graphic utility for visual curve fitting was developed by the senior author during a visit to the University of Waikato. It is written as a Java application for Windows operating systems supporting the Java virtual machine version 1.8 or higher. This short communication gives a brief description of the utility. Setting-up instructions and details of use are given in the link, along with the software itself.

There are presently many formal GEV-specific estimation methods and the interested reader is referred to Hosking et al. (1985), Wang (1990), Wang (1997), Moisélo (2007), Mínguez et al. (2010), Ailliot et al. (2011), Raynal-Villasenor (2012), and He et al. (2015). An overview of extreme value distribution estimation methods is included in Gomes and Guillou (2015). Some related software is available via Gilleland et al. (2013).

2.0 Symbolism

For the purposes of designing the computer graphic, it is convenient to express the GEV in terms of the well-known three extreme value distribution Types, as separate entities. The cumulative distribution functions of the Type 2 and Type 3 extreme value distributions are symbolized:

$$F(x) = \exp \left\{ - \left(\frac{x - \omega}{\delta} \right)^{1/k} \right\} \quad k < 0, \delta > 0, x \geq \omega \quad \text{Type 2} \quad (1)$$

$$F(x) = \exp \left\{ - \left(\frac{\xi - x}{\sigma} \right)^{1/k} \right\} \quad k > 0, \sigma > 0, x \leq \omega \quad \text{Type 3} \quad (2)$$

The cumulative distribution function of the Type 1 (Gumbel) distribution is given by:

$$F(x) = \exp \{ - \exp [- (x - \xi) / \delta] \} \quad \delta > 0 \quad \text{Type 1} \quad (3)$$

In Eqs. (1)–(3), ω and ξ are location parameters, k is a shape parameter, and δ and σ are scale parameters.

The graphic is utilized as an interactive Gumbel plot. Fitting an EV1 is achieved by screen-clicking two points to fit the data. Similarly, screen-clicking three points enables either an EV2 or EV3 curve to be fitted to the data. An EV3 distribution is defined if three screen-clicked points produce a curve with decreasing slope, while an increasing slope defines an EV2 distribution. In each case, the graphic displays the appropriate parameter symbols as defined in Eqs (1)–(3). The relationship of the respective curves to the parameter values is discussed in Bardsley (1989).

The screen-clicked points are amenable to relocation by clicking nearby, allowing the curves through the data to be adjusted by the user.

3.0 Time scale

The graphic has a time scale of years and uses a risk-based scale determined by user-specification of the probability \mathbf{p} , which is the probability of at least one exceedance in a T -year period (Bardsley, 1990). This is a more general time scale than the return period and has the advantage of having the scale starting point as $T = 1$. The default setting in the graphic is $\mathbf{p} = 0.632$ which, for practical purposes, gives the return period time scale, although differing slightly for small values of T .

As noted by Serinaldi (2015), a built structure designed against a T -year return period event has the rather high and apparently arbitrary probability of 0.632 of being damaged within T years. The more flexible time scale gives users the option to shift the time axis to the right by decreasing \mathbf{p} to a sufficiently small value that better reflects risk.

The graphic enables user-specification of the tick-points which define the time scale. For example, it might be decided to populate the scale with tick-points at 1, 30, 50, and 300 years. The tick-points are not actually drawn on the horizontal Gumbel y axis but appear instead on a separate time axis, which can be moved up or down via a slide bar.

4.0 Interactive graphic

As a starting point, the user copies from a spreadsheet a set of already ranked annual maxima and associated estimated probabilities of a smaller value as obtained from a plotting position expression (left and right columns in a spreadsheet respectively). After pasting into the graphic, the data appear in the *Magnitude* and q columns, as can be seen to the left in the Fig. 1 example. The user then selects either “GEV” or “Gumbel” options depending on whether the data displays any curvature on the plot.

The main features of the interactive graphic can be seen in the screenshot shown in Fig. 1. If desired, this particular plot can be displayed as an interactive graphic using the pre-saved file which is downloaded with the software from the link.

In the Fig. 1 example the “data” are just a set of Type 3 extreme value random values giving the typical curved plot with its slope decreasing to the right. The display shown in Fig. 1 was created by first choosing the GEV option and then screen-clicking to create the three green points indicated. The Type 3 form of the plotted curve was detected by the graphic and the text “Type 3” appears beneath the horizontal axis.

The left and right screen-clicked point locations determine the left and right extent of the screen display of the curve. This is simply an artistic device and there is no suggestion that these left and right point locations are known with any accuracy, in general. It is likely in fact that many different left and right points all lead to curves that have an approximately equal visual match to a set of data points.

The Type 3 parameter values corresponding to the Fig. 1 curve appear at the lower left of the plot. Some indication is gained of parameter estimation error, in this particular case by comparison with the true parameter values listed in the Fig. 1 caption. In practice, Type 3 distribution upper bounds are likely to be poorly defined (Papalexiou and Koutsoyiannis, 2006).

The tick-point values for the time axis in Fig. 1 were nominated as 1, 10, 20, and 100 years and then entered onto the scale using the *Set time scale* button. In this case the default \mathbf{p} value of 0.632 was kept. The slide bar on the right enables the time axis to be shifted vertically to avoid any overlap with data points. Given a defined time scale, a magnitude can be found for a given time value or vice versa by entering a number in the appropriate input box on the bottom left of the graphic. In Fig. 1, it can be seen that inserting 100 into the *Time value* box resulted in 88.5 being displayed as the 100-year return period magnitude from the user-defined curve.

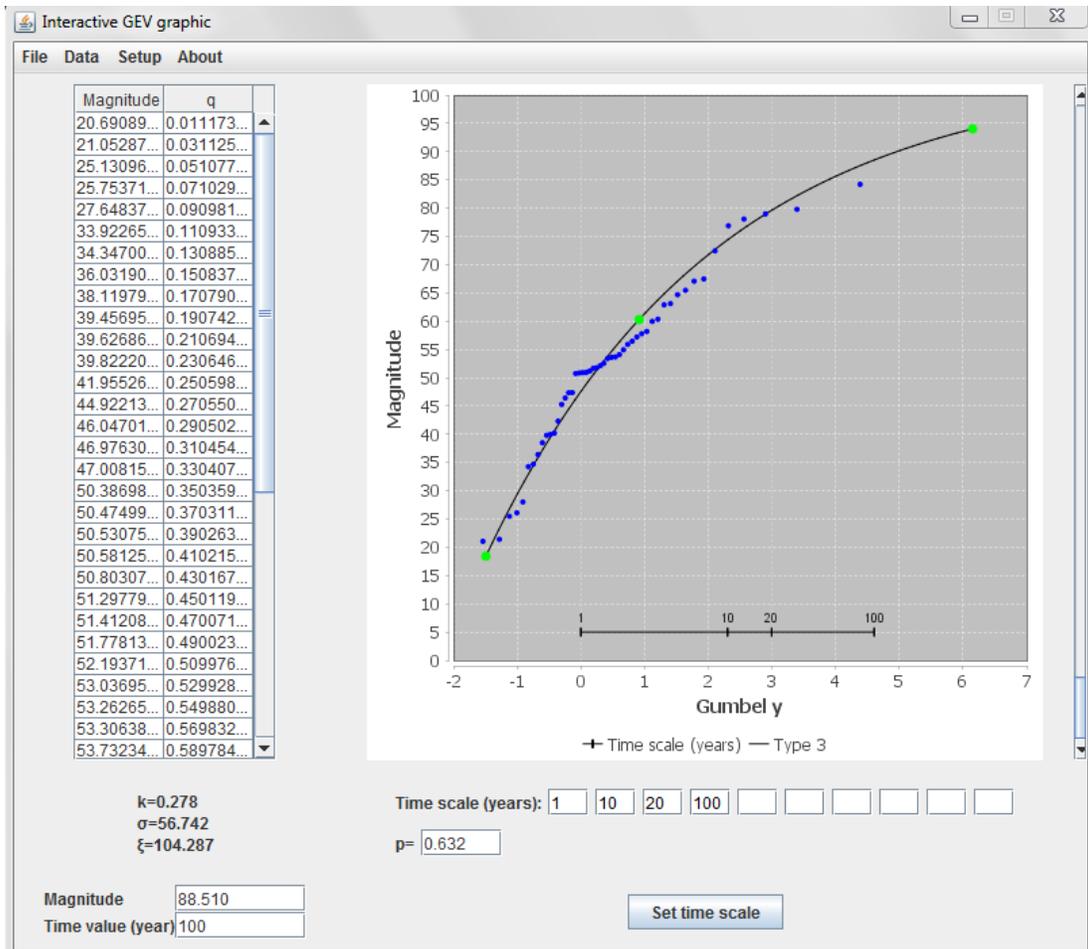


Fig. 1. Screenshot of a user-fitted 3-point curve passing through a set of 50 random values generated from a Type 3 extreme value distribution ($k = 0.41$, $\sigma = 45$, $\xi = 90$). The probabilities q of a smaller value were obtained previously from the Gringorten plotting position expression.

5.0 Conclusion

An interactive graphic was presented which enables easy visual fitting of GEV prediction functions to annual maxima, effectively allowing personal fitting weightings to be applied to data points. The graphic offers a broader perspective on fitting distributions to stationary time series of annual maxima and gives an intuitive feel for estimation error without mathematical distraction. The GEV distribution was utilized in this case but the graphic could be extended to include many other univariate distributions as well.

Software Availability

<https://sourceforge.net/projects/gev/files/>

References

- Ailliot, P., Thompson, C., Thomson, P. 2011. Mixed methods for fitting the GEV distribution. *Water Resour. Res.* 47 Article Number: W05551.
- Bardsley, W.E. 1989. Graphical estimation of extreme value prediction functions. *J. Hydrol.* 110, 315–321.
- Bardsley, W.E. 1990. Reliability based time axes for flood data presentation. *J. Hydrol.* 119, 389–391. Correction *J. Hydrol.* 123, 209.
- Gilleland, E., Ribatet, M., Stephenson, A.G. 2013. A software review for extreme value analysis. *Extremes* 16, 103–119.
- Gomes, M.I., Guillou, A. 2015. Extreme value theory and statistics of univariate extremes: a review. *International Statistical Review* 83, 263–292.
- He, J., Anderson, A., Valeo, C. 2015. Bias compensation in flood frequency analysis. *Hydrol. Sci. J.* 60, 381–401.
- Hosking, J., Wallis, J., Wood, E. 1985. Estimation of the generalized extreme value distribution by the method of the probability weighted moments. *Technometrics* 27, 251–261.
- Klemeš, V. 2000. Tall tales about tails of hydrological distributions. *J. Hydrol. Eng.* 5, 227–231 & 232–239.
- Mínguez, R., Méndez, F.J., Izaguirre, C., Menéndez, M., Losada I.J. 2010. Pseudo-optimal parameter selection of non-stationary generalized extreme value models for environmental variables. *Environ. Model. & Softw.* 25, 1592–1607.
- Moisello, U. 2007. On the use of partial probability weighted moments in the analysis of hydrological extremes. *Hydrol. Process.* 21, 1265–1279.
- Papalexiou, S.M., Koutsoyiannis, D. 2006. A probabilistic approach to the concept of probable maximum precipitation. *Advances in Geosciences* 7, 51–54.
- Raynal-Villasenor, J.A., 2012. Estimation procedures for the general extreme value distribution for the maxima: an alternate PWM method. *J. Hydrol. Eng.* 17, 909–922.
- Serinaldi, F. 2015. Dismissing return periods! *Stoch. Environ. Res. Risk Assess.* 29, 1179–1189.
- Wang, Q.J. 1990. Unbiased estimation of probability weighted moments and partial probability weighted moments from systematic and historical flood information and their application to estimating the GEV distribution. *J. Hydrol.* 120, 115–124.
- Wang, Q.J. 1997. LH moments for statistical analysis of extreme events. *Water Resour. Res.* 33, 2841–2848.