Journal of Hydrology 559 (2018) 209–221

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Research papers

The weighted function method: A handy tool for flood frequency analysis or just a curiosity?

Ewa Bogdanowicz, Krzysztof Kochanek *, Witold G. Strupczewski

Department of Hydrology and Hydrodynamics, Institute of Geophysics, Polish Academy of Sciences, Ksiecia Janusza 64, 01-452 Warsaw, Poland

ARTICLE INFO

Article history:
Received 2 August 2017
Received in revised form 19 January 2018
Accepted 11 February 2018
Available online 15 February 2018
This manuscript was handled by A. Bardossy, Editor-in-Chief, with the assistance of Felix Frances, Associate Editor

Keywords:
Probability distribution
Skewness
Quantiles
Estimation
Shifted Inverse Gaussian
Pearson 3 type
Weighted functions

ABSTRACT

The idea of the Weighted Function (WF) method for estimation of Pearson type 3 (Pe3) distribution introduced by Ma in 1984 has been revised and successfully applied for shifted inverse Gaussian (IGa3) distribution. Also the conditions of WF applicability to a shifted distribution have been formulated.

The accuracy of WF flood quantiles for both Pe3 and IGa3 distributions was assessed by Monte Caro simulations under the true and false distribution assumption versus the maximum likelihood (MLM), moment (MOM) and L-moments (LMM) methods. Three datasets of annual peak flows of Polish catchments serve the case studies to compare the results of the WF, MOM, MLM and LMM performance for the real flood data. For the hundred-year flood the WF method revealed the explicit superiority only over the MLM surpassing the MOM and especially LMM both for the true and false distributional assumption with respect to relative bias and relative mean root square error values. Generally, the WF method performs well and for hydrological sample size and constitutes good alternative for the estimation of the flood upper quantiles.

1. Introduction and literature review

In flood frequency analysis (FFA) and water management policy there is a need to estimate the maximum discharge of the flood wave for the given return period (or probability of non-exceedance, F) in other words the estimation of the upper quantiles by means of the sample of maximum annual flows in the river (Griffis and Stedinger, 2007). The statistical inference concerning the upper quantiles generates the problems of both the statistical and hydrological nature that are rarely treated jointly. The statistical uncertainty of the upper quantiles estimators stems *inter alia* from short time series, erroneous datasets, wrong assumption of the model of annual flow maxima, simplifications (e.g. independent identically distributed elements in the sample) and assumption on the stationarity of the river regime. In order to improve the quality of the estimates of the flood quantiles, the indispensable attempts to correct the quality of the datasets by e.g. engaging better measurement technology and hydrological information from various non-systematic sources (Viglione et al., 2013) should be supported by the proper hydrological models and methods of estimation. Both the input data (in terms of their size and quality) and the procedures of estimation (understood as the distribution function and estimation method) should be of the highest quality to obtain authoritative estimators of the annual peak flows (Strupczewski, et al., 2005, Kochanek, et al., 2005).

In the flood frequency analysis, a probability density function (PDF) is fitted to the datasets by means of more or less subjective methods from among the positively skewed, mostly two- or three-parameter continuous distribution functions (e.g. Katz et al., 2002). Cunnane (1985, 1989) claims that some of these models were introduced to the hydrology because of their flexibility to adjust various shapes of peak discharge distributions. Often, the theoretical arguments supporting the choice of a certain type of the FFA model can be easily undermined, so more the empirical advantages of the model seem to be more important than the a priori argumentation.

In Poland, as in some other countries, the Pearson type 3 model (Pe3) was recommended in engineering applications for the estimation of annual peak flows in the Polish rivers. Nowadays different models can be used (e.g. Weibull, LogPearson, LogNormal, GEV, to name the most popular), and different methods (e.g. seasonal or regional approach estimated mainly by means of MLM), however Pe3 distribution still plays a role in the Polish FFA as the reference or alternative model. Our former research (Strupczewski et al.,

* Corresponding author.
E-mail addresses: ewabgd@o2.pl (E. Bogdanowicz), kochanek@igf.edu.pl (K. Kochanek).

https://doi.org/10.1016/j.jhydrol.2018.02.020
0022-1694/

2001, Strupczewski and Węglarczyk, 2002c, Strupczewski et al., 2003, Strupczewski et al., 2006, Markiewicz et al., 2015) indicate that the shifted Inverse Gaussian (IGa3) distribution function (aka the shifted Convective-Diffusion or the shifted Wald distribution) being an alternative to the shifted Log-normal distribution (for medium skewness samples, Strupczewski et al., 2002b) is increasingly used in the FFA and represents relatively good upper quantiles in Polish rivers, regardless the flood generation processes and hydro morphology of the river. One has to bear in mind, however, that the real model of the flood quantiles population is unknown, and even though it was known, it would have too many parameters to be estimated from short hydrological datasets.

All the more so, the perfectly fitted model is only a part of the estimation procedure which consists also of the method of the estimation of quantiles. The carefully selected method of estimation can help with the mitigation of the modelling and sampling errors as well as the measurement mistakes (Strupczewski et al., 2002a, b). Therefore, statisticians and hydrologists have still been developing methods of estimations that could address certain problems of the FFA. The historically first methods of estimation were based on graphical representation of the maximal annual flows and their plotting positions (Hazen, 1930). Although, some of these 'visual' concepts are still in use, soon after the estimation methods evolved into statistical parametric paradigm, and now the most popular methods used in at-site FFA are the maximum likelihood (e.g. Kaczmarek, 1977, Rao & Hamed, 2000, Coles, 2001), moments (Rao and Hamed, 2000) and L-moments (Greenwood et al., 1979, Hosking, 1986, Hosking and Wallis, 1997, Hosking, 2006). Recently, Bayesian methods and expected moments algorithm are of great interest especially when historical information is available (e.g. Paretti et al., 2014, Parkes and Demeritt, 2016).

The MLM is said to provide asymptotically unbiased and optimal (in the sense of minimum variance) estimators of the parameters when the assumed model is true, but it concentrates on the main mass of the probability far from the upper quantiles (Strupczewski, 2000). Its theoretical advantages fade when a wrong estimating model is fitted to the dataset (Strupczewski et al., 2002b). Moreover, the numerical algorithms of the MLM can fail for unknown reasons when the number of estimated parameter is large (even more than 2!). The main attractiveness of the LMM stems from the fact that the L-moments are the linear (of course) combinations of the sample elements which does not lead to the multiplication of errors and guarantees the existence of all the theoretical L-moments provided that the mean exists (Stedinger and Vogel, 1992). Also, the LMM can give equally good estimates of the upper quantiles as the MLM and is robust to the outliers, but unfortunately, it can be hardly applied to the PDFs that do not have the explicit mathematical form of the quantile, including the Pearson type 3 (see Hosking and Wallis, 1997) as well as the shifted Inverse Gaussian. Besides, the LMM requires the order, monotonically increasing samples, that ruins the temporal order of the floods events and impedes the use of this method in the non-stationary FFA.

As far as the MOM is concerned, the Monte-Carlo simulations showed that for short samples and false assumption of the model this method gives smaller bias of the upper quantile estimators than the MLM (Strupczewski et al., 2002a, b, Strupczewski et al., 2005, Kochanek et al. 2005). Moreover, the MOM estimates are characterised by relatively low mean square error. However, especially the systematic error rises with the order of the moments, because the sample elements are squared (as in the variance, *var*, and coefficient of variation, *CV*) and cubed (as in coefficient of skewness, *CS*) when the three-parameter PDF is involved. Additionally, the upper quantiles by MOM are undervalued (negative bias) which is unwelcome when flood security systems are designed. In the classical remarkable work 'Just a moment!' Wallis et al. (1974) analysed in detail the positive and negative features of the MOM for the most frequently used distribution functions, including Pearson type 3. Of course the trio MOM, MLM and LMM does not exhaust the range of the methods of parameters estimation used nowadays in hydrology, but their description falls far beyond the scope of this paper (Review applied, 2012).

The aforementioned features of the LMM and the MOM, their numerical simplicity encouraged the scientists to find a way to merge the advantages of both methods without their disadvantages, i.e. keep the powers of the sample elements as low as possible (as in the LMM) and no need to have the explicit mathematical form of the quantile and to order samples (MOM). One of such ideas, first proposed by Kartvelisvili as early as in 1963, was based on the generalisation of the moments by resigning from the traditional assumption that the moments are the functions of the sample elements raised to the positive integer power, i.e. moments' power = 1 (mean), 2 (variance), 3 (skewness), etc. (e.g. Ashkar & Bobee, 1987, Ashkar and Mahdi 2003, 2006, Kochanek & Feluch, 2016).

Despite the vast literature giving the impression that there is no room for other moment-based method of upper quantiles estimation, the two papers (Liang et al., 2014; Wang et al., 2015) referring to the article of Ma (1984) in Chinese presented the new approach to the estimation of the Pearson type 3 PDF (Pe3) where the *CS* is expressed as a function of the *CV* and two weighted central moments. In consequence, this new approach 'powered down' the moments from 3 to 2. This idea, known as the Weighted Function Method (WF) was modified and developed by Liang, et al. (2014) by further reduction of the power from 2 to 1. It resulted in the concept of the Modified Weighted Function Method (MWF). In the Liang at al.'s paper the Normal distribution plays the role of the weight function; however the authors suggest that the other probability density function can also serve as the weight functions in the MWF method. The Monte Carlo simulations showed that the WF and MWF for the Pe3 generally proved to generate smaller bias of upper quantiles (F = 0.99–0.9999) for hydrological sample size (N = 20–50) compared with the results by the LMM. Also the mean square error of the upper quantiles estimators by the WF is competitive, when the true, that is Pe3, model is assumed.

As far as the weight function is concerned, its choice is constrained to the family of unimodal probability density functions. However, for a finite sample the selection of the weight function can affect the estimator of skewness and thus the usability of the WF method. Therefore, we started our work from the assessment of the sensitivity of the estimate of the upper quantiles to the weight function choice for the Pe3 distribution. The six two-parameter probability density functions serve as the alternatives for weight functions, namely Normal, Exponential-Exponential (compare Liang, et al., 2014; Wang et al., 2015), Gumbel, Log-Normal, Gamma and Inverse Gaussian. Although the WF method derives from the MOM and is intended to improve its deficiencies, none of the two articles compares the accuracy of these methods. Also there was no comparison with the most popular MLM method. So, we compared of WF accuracy in respect to upper quantiles estimators with the outcomes of three routine FFA methods, that is MOM, MLM and LMM.

In recognition of attractiveness of the WF method with regard to the Pearson 3 distribution, we searched for other probability distribution functions used in the FFA for which WF method would be applicable. The conditions of WF applicability have been formulated and the WF has been successfully applied for the shifted