PROBABLE MAXIMUM FLOOD DETERMINATION FOR THE RIVER MODAU

Peter Bakucz
PTB Braunschweig, Bundesallee 100, 38100 Braunschweig, bakucz@bakucz.de

Abstract: We present an alternative method to analyse PMF (Probable Maximum Flood) for the more than 40 years discharge data set for Modau river, Germany. We base our analysis on the study of the partition function. This function is used to investigate the Modau discharge values making use of the different information embedded at different scales and moments. Using the partition function in a likelihood analysis in two dimensions (*Q, n*) to define the PMF value, we find the best-fitting model to the best data available at present (the). By means of this analysis we find a PMF value $Q_{\max}=317\frac{m^3}{\sec}$ with the likelihood function for n=1.8 and $Q_{\max}=212\frac{m^3}{\sec}$ (95 % confidence level).

Keywords: PMF, partition function, likelihood analysis.

1. Introduction

This article is primarily intended to provide procedure for the development of the Probable Maximum Flood (PMF). For about last 30 years the PMF has received general acceptance as the design flood for dams in Germany, whose failure would pose a threat to public safety. More recently, the PMF has received acceptance as the design flood for large dams in many other countries as well.

The PMF is the flood that may be expected from the most severe combination of critical meteorological and hydrologic conditions that are reasonably possible in the drainage basin under study. A PMF could be generated by the probable maximum precipitation (PMP) which is theoretically the greatest depth of precipitation for a given duration that is a physically possible for a given size storm area at a particular geographic location at a certain time of year.

Developing a PMF hydrograph for a dam safety evaluation generally involves two steps, which are, respectively, hydrologic and hydraulic in nature:

- Modeling of runoff through the project drainage basin to produce an inflow PMF for the project reservoir.
- Routing of the inflow PMF through the project reservoir and dam outlet works to obtain the outflow PMF and the maximum reservoir elevation at the dam.

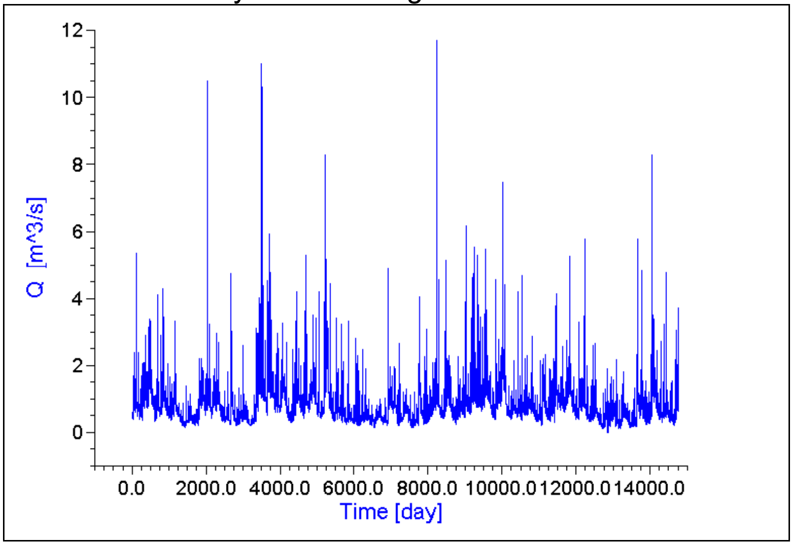
These steps involve considering several coincident or sequential events, each of which may have a strong effect on the resulting PMF.

Estimation of the PMF is less than a perfect science and efforts continue go into its refinement. A reason for these indistinctnesses could be that the methods used for modeling the runoff are based on standard linear statistical techniques which do not fit when the phenomenon to be analyzed is essentially non-linear.

One of the previous analysis of the nonlinear characteristics of the runoff was developed by Krasovskaia et al. (1999) through the study of dimensionality of Scandinavian river flow regimes. The dimension found was a low and non integer value, which indicates that the flow system may be described by a nonlinear function of just a few variables. The fact that the dimension found was fractal predicts that this function, also called map in nonlinear dynamics can display chaotic behavior under certain circumstances.

In this paper we present a known alternative method in the nonlinear mechanics to analyze PMF values based on the partition function. This function contains useful information about the discharge anisotropies at the different scales and moments. The method presented here is related to the one used by Diego et al. (1999) based on moments at different smoothing angles. However, our method is more general and powerful because it works with any moment, not only with positive and integer ones.

The structure of the paper is as follows. In section 2 we present and discuss the multifractal analysis and the partition function. In the same section the likelihood analysis based on that function is introduced. The likelihood analysis uses the partition function to search for the PMF values that best fit a given data set. In section 3 we apply the results of the previous section to the Modau 40 years discharge data set.



(Figure1).

2. The partition function

The general aim of the multifractal formalism is to determinate the $f(\alpha)$ singularity spectrum of a measure μ. It associates the Haussdorff dimension of each point with the singularity exponent α, which gives us an idea of the strength of the singularity:

$$N_{\alpha}(\varepsilon) \approx \varepsilon^{-f(\alpha)} \quad (1)$$

where $N_{\alpha}(\varepsilon)$ is the number of boxes needed to cover the measure and ε the size of each box.

A partition function Z can be defined from this spectrum:

$$Z(q,\varepsilon) = \sum_{i=1}^{N(\varepsilon)} \mu_i^q(\varepsilon) \approx \varepsilon^{\tau(q)} \text{ for } \varepsilon \to 0 \quad (2)$$

where $\tau(q)$ is a spectrum which arouses by Legendre transforming the $f(\alpha)$ singularity spectrum. The spectrum of generalized fractal dimensions is obtained from the spectrum $\tau(q)$:

$$D_q = \frac{\tau(q)}{(q-1)} \quad (3)$$

The capacity or box dimension of the support of the distribution is given by $D_0 = f(\alpha(0)) = -\tau(0)$. $D_0 = f(\alpha(1)) = -\tau(1)$ corresponds to the scaling behavior of the information and is called information dimension. (Grassberger and Procaccia, 1983; Peitgen et al., 1992)

In the partition function (Eq.2.) the quantity $\mu_i(\varepsilon)$ is the size or scale of the boxes used to cover the sample. The boxes are labeled by *i* and $N(\varepsilon)$ is the number of boxes (or cells) needed to cover the map when the grid with resolution ε is used. The exponent *q* is a continuous real parameter that plays the role of the order of the moment of the measure.

Let us consider a time series with *N* discharge values. Now the time series is divided in boxes of size $\varepsilon \times \varepsilon$ and the measure $\mu_i(\varepsilon)$ is computed in each one of the resulting boxes. Changing both, *q* and ε, one calculates the function $Z(q,\varepsilon)$. One is free to make any choice of the measure $\mu_i(\varepsilon)$ provided that several conditions are satisfied, the most restrictive being $\mu_i(\varepsilon) \geq 0$. There are no general rules to decide which is the best choice. For discharge time series, we use the most natural measure defined as follows:

$$\mu_i(\varepsilon) = \frac{1}{Q_*} \sum_{day_j \in box_i} Q_{day_j} \quad (4)$$

Thus the measure in the box *i* is the sum of the daily discharge *Q* of the discharge values inside the box in units of $\frac{m^3}{\sec}$. The measures are interpreted as probabilities and they have to be normalized, i.e $\sum_i \mu_i = 1$. So Q_* is simply the sum of the daily discharges and therefore it is a constant for all boxes and scales.

Using the measure proposed in this paper, the differences between two discharge data sets (the original and the modified one in order to determine the PMF value) appear when high values of the exponent *q* are considered. The method is able to differentiate between two very close models with *q* ranging between [0,50]. This range for *q* is in agreement with the level of inhomogeneity. We are using daily discharges, that is, we have inhomogeneities of order 10^{-1} with respect to the mean value. One can consider *q* as a powerful microscope, able to enhance the smallest differences of two very similar time series. Furthermore, *q* is a selective parameter. Choosing large values of *q* in the partition function, favors contributions from cells with relatively high values of $\mu_i(\varepsilon)$ since $\mu_i^q >> \mu_j^q$ for $\mu_i >> \mu_j$, if *q<<0*. Conversely, *q*>>0 favors the discharges with relatively low values of the measure. This is the role played by the moments, changing *q* one explores the different parts of the measure probability distribution. The other parameter ε, acts like a filter. Choosing big values of ε is similar to apply a large scale filter to the time series. One looks at different scales when the parameter ε is changed.

To summarize, *Z(q,* ε*)* contains information at different scales and moments. The multi-scale information gives an idea of the correlations in the discharge time series, meanwhile the moments are sensitive to possible asymmetries in the data, as well as some deviations from Gaussianity. In what follows we show the power of the partition function to extract useful information from daily discharge data.

2.1. Likelihood analyis

We shall use the partition function to encode the information of a given Modau discharge series. We compute it both for the measured data and for simulated ones. In this process we are comparing the data and the model at several scales and using different moments. If there are some differences at some scale or moment, then the partition function should make it evident. The likelihood function will have a maximum for the best-fitting model to the data. For the daily discharge analyses, we consider models corresponding to different values of the spectral index *n* and the normalization Q.

The likelihood is defined in the usual way (assuming a Gaussian distribution for ln *Z(q,* ε*)*). We work with **Z**=ln *Z(q,* ε*)* instead of *Z(q,* ε*)* because of the large values of *q* which make impossible to compute directly *Z(q,* ε*)*,

$$L(Q,n) = \frac{1}{(2\pi)^{n/2}(\det M)^{1/2}} \exp(-\frac{1}{2}\chi^2) \quad (5)$$

where,

$$\chi^2 = \sum_{i=1}^{N_p} \sum_{j=1}^{N_p} (\langle Z(i) \rangle - Z_D(i)) M_{ij}^{-1} \left(\langle Z(j) \rangle Z_D(j) \right) \quad (6)$$

and $\langle Z(i) \rangle$ is the average of the **Z** for the realizations of the model at bin *i*. Realization means: systematically changed discharge values at randomly chosen time positions. The index *i* defines pairs of values (q, ε). That is, *i* runs from 1 to the total number of points N_p where Z(q, ε) is defined. $Z_D(i)$ is the value of **Z** for the experimental data at bin *i*.

M_{ij} is the covariance matrix calculated with Monte Carlo realizations:

$$M_{ij} = \frac{1}{N_{rea}} \sum_{k=1}^{N_{rea}} \left(Z_k(i) - \langle Z(i) \rangle \right)\left(Z_k(j) - \langle Z(j) \rangle \right)$$

$Z_k(i)$ denotes the value of ***Z*** at bin *i* for the *k* realization (Falconer, 1990).

We have two possibilities to perform a best fit to the data. The first one is to minimize χ^2 and take the values of $Q_{\max}$ and *n* at the minimum of the χ^2 surface as the best-fitting values. The second possibility is to work with the likelihood *L* looking for the maximum. We tested the two possibilities using simulated discharge time series derived from a given pair of parameters *($Q_{\max}$, n)*. Due to daily discharge variance we obtain a set of maxima in the likelihood and of minima in the χ^2. The conclusion is that the likelihood is somewhat better than the χ^2 as expected.

3. Results

In order to determine which are the values of the quadrupole normalization *Q* and the spectral index *n* that best fit the Modau discharge data, we perform Monte Carlo simulations of the time-series for a scale-free model with an equal multifractal spectrum.

We consider different values for Q and the *n* ranging from Q=0.5 $\frac{m^3}{\sec}$ to Q=750 $\frac{m^3}{\sec}$ and from *n*=0.3 to *n*=4.3. We add instrumental noise based on the number of data collected by TU Darmstadt. Furthermore, there is another effect that must be taken into account, the seasonal variance. To treat conveniently this effect we perform a large number of simulations (more than 2000) for each pair of values *(Q, n)* and then we compare the average *Z(q, ε)* values of these simulations with the **Z** corresponding to the Modau discharge time series (the used values for *q* and ε were *q*=1,4,7,25 and ε =3,4,8,16). The size of the *Z(q, ε)* grid, is not critical and what is now relevant is the *q* values considered. In particular, high order moments (i.e large *q*) are very sensitive to the tail of the distribution and therefore the results obtained with those high values on the parameter estimates are not stable. The combination of *q* and ε values, was one of the combinations for which the recovered parameters Q and *n* were closer to the input parameters and with smaller error bars. As mentioned in section 2, *q* should take values of order 10^{-1}. The values of *q* where chosen to be asymmetric in an attempt to consider possible asymmetries. The range for ε runs from 3 pixels to 16 pixels which is the largest box size required to have at least 8 boxes. Using a maximum likelihood method one can determine which are the best-fitting parameter values of the simulations (signal + noise) to the Modau discharge data.

In Fig. 2. we show a contour plot of the likelihood obtained for the Modau discharge data. The maximum is at $Q_{\max}$ =317$\frac{m^3}{\sec}$ and *n*=1.8 (95% marginalised errors) and the contour level at 68% is compatible with the assumed standard value *Q*=188 with *n*=1. The various analysis of the 40 years Modau data combined give as the best-fitting parameters *Q*=212 and *n*=1.2. The result presented here predicts larger values of *n* and smaller values of $Q_{\max}$ than the result indicated above (although always inside the anticorrelation law for the two parameters).

This result is in agreement with the one found by Bakucz (1999), using an another approach. A possible explanation for the discrepancy between our results and those obtained with the another methods could be a bias present in the likelihood estimator. In the tests of our algorithm we found a systematic bias in the marginalized likelihood functions both for *Q* and *n* with typical values of $\Delta n = 0.2$ and $\Delta Q = 2\,03$ which could explain part of our discrepancy. The reason for this bias can be the difference between the assumed Gaussian form for the likelihood of the partition function in eq. (2) and the real non-Gaussian distribution. The probability distribution of the ***Z*** at each *(q, ε)* obtained from simulations is similar to a Gaussian probability distribution but with a longer tail for high values. We also think that maybe the noise can contribute to that bias. The high order moments (large *q*) of the partition function are very sensitive to the tails of the distribution of the seasonal fluctuations. A low signal to noise ratio (as is the case for the Modau data) could raise the parameter *n* that best fit the Modau data. We did some tests in this direction and apparently the noise can increase the value of *n* (and consequently can produce a lower value of *Q*).

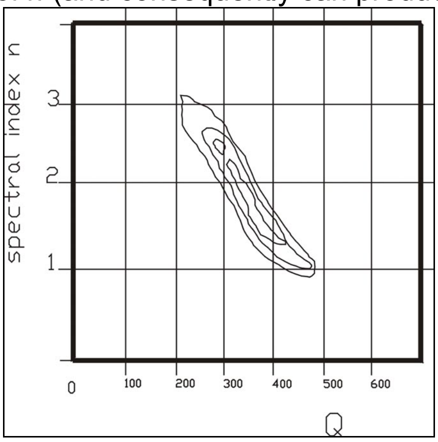


Figure.2.

5. Acknowledgement

P.Bakucz would like to thank the Alexander von Humboldt Foundation, Germany for supporting a grant.

6. References

Diego J.M. et al. Mon. Not. R. Astron. Soc. 306. 427. 1999.

Bakucz, P.: *Fractal geometry for water management*. (Proc. of 2nd Inter-regional Conf. on Environment-water, Lausanne 1999).

Thompson, K. , Stedinger J. , Heath. D. J. Water Resources Planing and Management. July/August. (1997).

Krasovskaia, I., Gottschalk, L. Kundzewicz, Z. : *Dimensionality of scandinavian river flow regimes*. Hydr.Sci.Jour. 44(5) oct.1999.

Haimes, Y. et al.: *Risk analysis of dam failure and extreme floods: application of the partitioned multiobjective risk method.* IWR Rep. No. 88-R-1. US.Army Corps of Engineers. (1988).

Davis, D.: *A risk and uncertainity based concept for sizing levee projects*. Proc. Hydrol. and Hydr. Workshop on Riverine Levee Freeboard SP-24. US Army Corps of Engs Hydr.Engg. Davis, Calif. (1991).

Falconer, K.: *Fractal geometry.* John Wiley and Sons. Chichester. 1990.

Grassberger, P., and Proccacia, I.: *Measuring the strangeness of strange attractors* Physica.,189 (1983).