AGU PUBLICATIONS

Geophysical Research Letters

Supporting Information for

A hydrogeomorphological index of heavy-tailed flood behavior

H. -J. Wang[1], R. Merz[1,2], S. Yang[3], and S. Basso[1,4]

[1]Department of Catchment Hydrology, Helmholtz Centre for Environmental Research – UFZ, Halle (Saale), Germany,

[2]Institute of Geosciences and Geography, Martin-Luther University Halle-Wittenberg, Halle (Saale), Germany,

[3]Department of Aquatic Ecosystem Analysis, Helmholtz Centre for Environmental Research – UFZ, Magdeburg, Germany

[4]Norwegian Institute for Water Research (NIVA), Oslo, Norway

Contents of this file

Introduction

This supporting information contains two supplementary methods and one figure. Text S1 is the theory of identifying tail behavior for distributions of peak flows and flow maxima from hydrological dynamics. Text S2 is the method we used to test the power law hypothesis. Figure S1 is a reference map of the analyzed basins.

Text S1. Identifying tail behavior for distributions of peak flows and flow maxima from hydrological dynamics

The probability distribution of ordinary peak flows (i.e., local flow peaks generated by streamflow-producing rainfall events (Zorzetto et al., 2016)) and flow maxima (i.e. maximum values in a specified time frame) can be analytically expressed as $p_j(q)$ and $p_M(q)$, respectively (Basso et al., 2016):

$$p_j(q) = C_2 \cdot q^{1-a} \cdot e^{-\frac{q^{2-a}}{\alpha K(2-a)}} \cdot e^{\frac{q^{1-a}}{K(1-a)}} \tag{S1}$$

$$p_M(q) = p_j(q) \cdot \lambda\tau \cdot e^{-\lambda\tau \cdot D_j(q)}, \qquad D_j(q) = \int_q^{\infty} p_j(q)\, dq \tag{S2}$$

where $\tau[day]$ is the duration of the specified time frame, C_2 is normalization constants, and all the other notations have been listed in the main context.

To analyze the tail behavior of these distributions, we take the limit of $q \longrightarrow +\infty$ for both Equations S1 and S2. Because $\lim_{q\to\infty} D_j(q) = \int_{\infty}^{\infty} p_j(q)\, dq = 0$, the Equations S1 and S2 can be transformed into: (set $C_3 = \lambda\tau C_2$)

$$\lim_{q\to\infty} p_j(q) = \begin{cases} C_2 \cdot q^{1-a} \cdot e^{-\frac{q^{2-a}}{\alpha K(2-a)}}, & 1 < a < 2 \\ C_2 \cdot q^{1-a}, & a > 2 \end{cases} \tag{S3}$$

$$\lim_{q\to\infty} p_M(q) = \begin{cases} C_3 \cdot q^{1-a} \cdot e^{-\frac{q^{2-a}}{\alpha K(2-a)}}, & 1 < a < 2 \\ C_3 \cdot q^{1-a}, & a > 2 \end{cases} \tag{S4}$$

For both of the cases, the tail behavior is determined by a power law term and an exponential term when $1 < a < 2$, which indicates the tail decreases slower than the exponential but faster than the power law tail; while the tail behavior is solely determined by a power law function, representing heavy-tailed flow distribution when $a > 2$. Therefore, the hydrograph recession exponent ($a > 2$) is shown as an indicator of the heavy-tailed flood behavior.

Text S2. Testing the power law hypothesis

Every empirical data distribution can be fitted by a power law model no matter what is the true distribution from which the data is drawn. To identify case studies for which the power law is a plausible distribution of the observed data, we test the power law hypothesis by means of the method of Clauset et al. (2009), which statistically confirms whether the power law distribution fitted on the empirical data provides a reliable description of those data. We compute this goodness-of-fit framework via the function test_pl in the python package plfit 1.0.3 (https://pypi.org/project/plfit/).

The challenge here is to discern the errors caused by the sampling randomness from those arising because the data might be actually drawn from another distribution rather than the power law. The principle of the approach is to first measure the error distance ε_d between the empirical data and the optimized power law model, which is the distance need to be tested. Secondly, we generate a number of synthetic data samples by randomly sampling from the optimized power law model. The error distance ε_s between the synthetic data and the optimized power law model is measured, indicating the fluctuation caused by randomness only. A power law hypothesis is accepted if $\varepsilon_d < \varepsilon_s$ but rejected if $\varepsilon_d > \varepsilon_s$.

However, it is possible that non-power-law empirical data also has a smaller ε_d than ε_s. To address this issue, a great number n of iterations via the Monte-Carlo test for this approach is needed.

The Kolmogorov-Smirnov statistic is used to measure the error distance with $n = 1000$ (as suggested by Clauset et al. (2009)). In the meanwhile, the p-value is defined as the frequency of $\varepsilon_s > \varepsilon_d$. The power law hypothesis is ruled out if $p \leq 0.1$ whereas it is confirmed as plausible if $p > 0.1$. We, therefore, term all the qualified cases (i.e., $p > 0.1$) ‘confirmed heavy-tailed cases’ to indicate their empirical power law distributions are convincingly supported by the data, whereas the others are not.

It is worth mentioning that, statistically, we cannot say those who does not qualify ‘are not’ power law distributions. There are at least two potential reasons for this result: (1) they are indeed not power law functions, or (2) The empirical data do not represent well the actual underlying distribution, often due to small sample sizes.

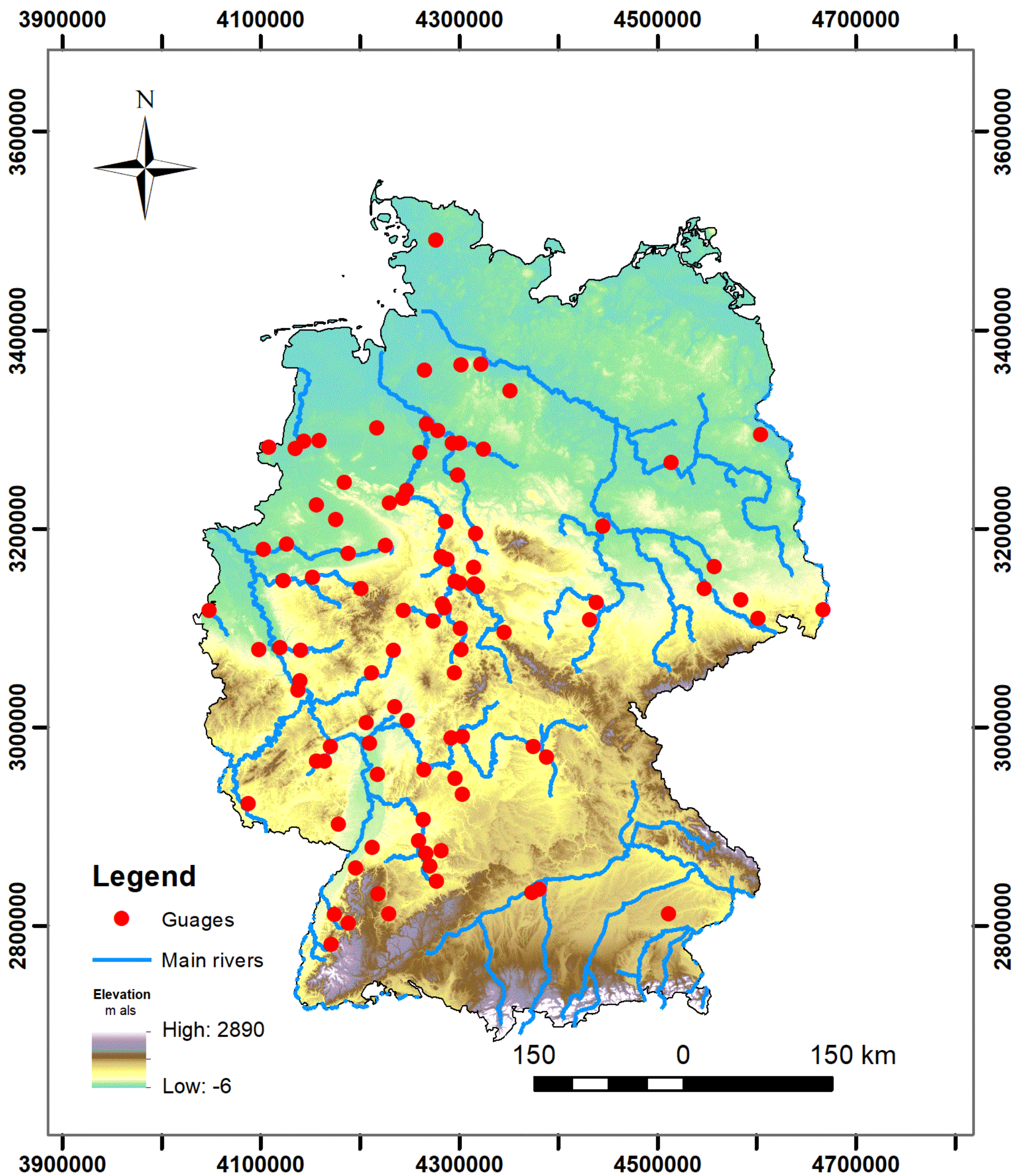

Figure S1. A reference map of 98 streamflow gauges across Germany. These river basins encompass a variety of climate and physiographic settings, without strong impact from snow dynamics. Their areas range from 110 to 23,843 km^2 with a median value of 1,195 km^2. The minimum, median, and maximum lengths of the daily streamflow records are 35, 58, and 63 years (inbetween 1951 – 2013).