AGU PUBLICATIONS

[Water Resources Research]

Supporting Information for

Emergence of Heavy Tails in Streamflow Distributions: the Role of Spatial Rainfall Variability

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

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

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

Contents of this file

Text S1 to S2
Figures S1 to S3
Tables S1 to S2

Introduction

This supporting information contains two supplementary methods, three figure, and three tables. Text S1 is the method we used for the identification of suitable probability distribution for REGNIE rainfall fields; Text S2 is the method we applied to identify an effective threshold of the increasing spatial rainfall variability from our results; Figure S1 is together with text S1, which is the results of goodness-of-fit for probability distributions; Figure S2 is the schematic diagram of the adopted hydrological model; Figure S3 represents the definition of the index of relative tail heaviness; Table S1 lists the information of five select catchments for simulation scenarios; Table S2 is the imposed parameters for synthetic rainfall generation in stationary scenarios; Table S3 is the imposed parameters for synthetic rainfall generation in non-stationary scenarios.

Text S1. Identification of suitable probability distribution for REGNIE rainfall fields

We computed the chi-square statistic (χ^2) for the goodness-of-fit test to identify the suitable probability distribution for REGNIE rainfall fields in 5-tested catchments (see section 2). REGNIE is a daily data set which is an 1-km^2 interpolated rainfall field estimated from point observations through multiple regression provided by the German Weather Service (Rauthe et al., 2013). Ten probability distributions (see figure S1) which were often applied in climatology or environmental researches were tested (e.g., Allard & Soubeyrand, 2012; Ayalew et al., 2014; Ben-Gai et al., 1998; Clauset et al., 2009; Gaitan et al., 2007; Hu et al., 2019; Li et al., 2014; Li & Shi, 2019; Ng et al., 2019; Tabari, 2020; Ye et al., 2018). To evaluate the goodness-of-fit between distributions in the same scale we normalized the χ^2 of each distribution by the total χ^2 of all tested distributions for each rainfall day (i.e., re-scale each χ^2 into [0, 1]). To identify the general expression of rainfall-field distributions from all days in the time series we computed the summation of all the normalized χ^2 (i.e., all rainfall days) as the total normalized χ^2 which showed the overall size of the discrepancies between the rainfall fields and the tested probability distribution. Finally, the most suitable probability distribution was determined by the lowest summation of the normalized χ^2 (i.e., the distribution has least discrepancies to the rainfall fields). The computing procedure for each catchment is listed as follows:

1. Compute χ^2 for 10-tested distributions day by day (rainfall days).
2. Normalized each χ^2 by the total χ^2 for 10-tested distributions of each day.
3. Sum up the normalized χ^2 of all days for each distribution.
4. Select the distribution which has the least summation in the step 3.

We estimated the suitable probability distribution for the rainfall fields of the summer data of the daily time series from 1980 to 2002 in 5-tested catchments. Figure S1 shows the total normalized χ^2 of 5-tested catchments for 10 distributions. Gamma, Exponential-Normal, Generalized Extreme Value (GEV), Skew-Normal distributions appear well goodness-of-fit to the rainfall fields. We finally selected the Gamma distribution as the most suitable one for our simulation because it is a 2-parameter distribution while the others are 3-parameter distribution, which has less uncertainty.

Text S2. Identification of an effective threshold of the increasing spatial rainfall variability

We proposed a statistical-cascade criteria to test if an effective threshold exists in the results of the scatter plot of spatial rainfall variability (i.e., CV_{in}) and the streamflow-distribution tail heaviness (i.e., H). In this criteria, we separately analyzed the trend between CV_{in} and H below and beyond a certain CV_{in} value by computing their linear regressions. All the value of CV_{in} from the results are tested as potential thresholds ($\mathrm{CV}_{\mathrm{thre}}$). We call the sequence below a potential threshold as sequence α whereas the sequence beyond it as sequence β. The cascade criteria is structured as follows:

Criteria 1 –Significance Test: the slope of the tendency line of α is not significant ($p \geq 0.05$) different to zero while the slope of β is tested to be significantly ($p<0.05$) larger than zero.

Criteria 2 –Optimization of Correlation: we select the best value of CV_{thre} among all the qualified CV_{thre} from criteria 1, which is determined by maximizing the optimization matrix: $(1 - r_{\alpha}) + r_{\beta}$, where r_{α} is the correlation coefficient of α and r_{β} is the correlation coefficient of β. We ensure that CV_{in} and H correlated to each other as strong as possible in sequence β (i.e., beyond the threshold) while they correlated to each other as weak as possible in sequence α (i.e., below the threshold).

Criteria 3 –Increasing Trend: the maximum value of *H in sequence* α have to be smaller than the maximum value of *H in sequence* β, by which we ensure the general increasing trend of H toward CV_{in}.

The unique effective threshold is identified as the CV_{thre} if and only if it is qualified by criteria 1 ⟶ criteria 2 ⟶ criteria 3. The Pearson correlation coefficient and the Wald Test with t-distribution were computed by means of the scipy.stats.linregress tool of SciPy v1.7.1. for the linear regression.

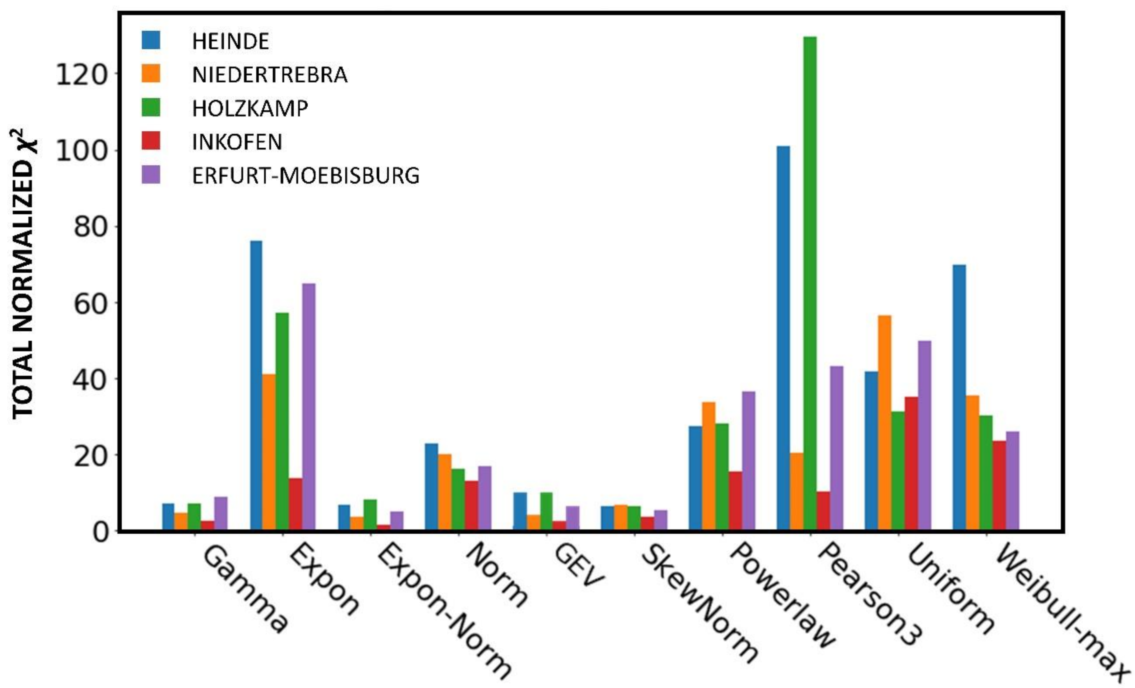

Figure S1. Overall goodness-of-fit of probability distributions for the rainfall fields in REGNIE data set. The total normalized chi-square statistic (χ^2) is the summation of normalized χ^2 (which is normalized by the total χ^2 of 10-tested distributions for each rainfall day) of all rainfall days from the summer data of the daily time series from 1980 to 2002 in five-tested catchments.

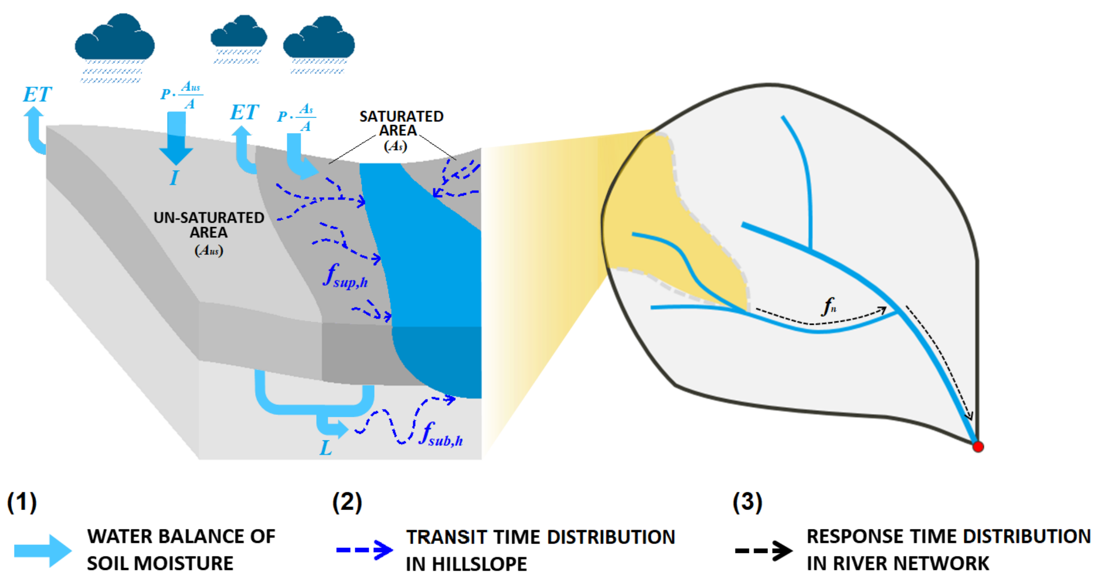

Figure S2. Schematic diagram of the adopted hydrological model. Three key components were accounted for in the model: (1) the soil water balance in hillslopes, with main fluxes depicted through blue solid arrows, (2) the transit time distributions of surface and subsurface flows in hillslopes, graphically represented through blue dashed arrows, and (3) the response time distribution of each catchment unit in the river network, displayed by means of black dashed arrows. Symbols used in the Figure are reported in section 3.2.

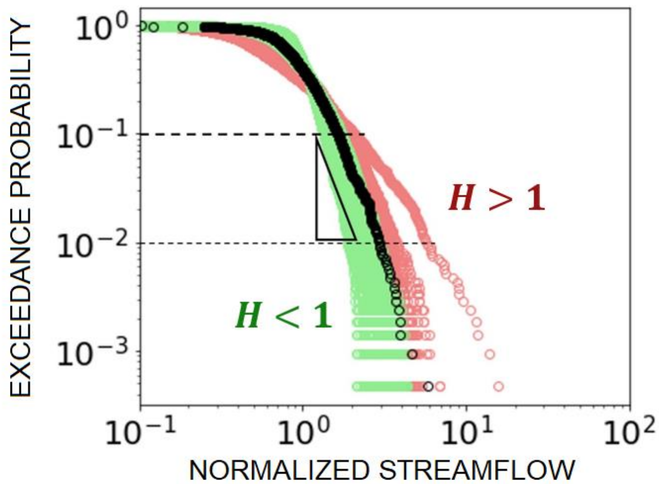

Figure S3. Definition of the index of relative tail heaviness. The exceedance probability distribution of the normalized streamflow (i.e., divided by the long-term mean daily flow) is displayed in a double logarithmic plot. Black dots represent the case with uniform rainfall. Exceedance probability distributions resulting from spatially variable rainfall are marked with either red or green dots depending on their values of the index of relative tail heaviness *H*. All the cases marked in red have $H > 1$, which signifies heavier tails than in the case of uniform rainfall; all the cases marked in green have $H < 1$, which indicates lighter tails than in the case of uniform rainfall.

Table S1. Study Catchments Used for Scenario Simulation

Group[a]	Catchment	River	Area [km^2]	Elongation Ratio, R_e [−][b]	Drainage Density, D_d [km/km^2]	Mean Elevation [m]
1	Holzkamp	Delme	98	0.45	0.74	41
1, 2	Niedertrebra	Ilm	887	0.45	0.64	394
1	Inkofen	Amper	2841	0.47	0.88	619
2	Heinde	Innerste	898	0.78	0.69	243
2	Erfurt-Moebisburg	Unstrut	847	0.90	0.71	441

[a]Group 1 basins are used for the investigation of the effects of catchment size and group 2 basins are used to study the effects of catchment shape. [b]The elongation ratio closed to 1 indicates more circularity.

Table S2. Parameter for Stationary Scenarios: Shape Parameter (*k*)

#	*k* [−]	#	*k* [−]	#	*k* [−]	#	*k* [−]
1	1000	14	1.3	27	0.09	40	0.005
2	500	15	1.2	28	0.08	41	0.004
3	100	16	1.1	29	0.07	42	0.003
4	50	17	1	30	0.06	43	0.002
5	10	18	0.9	31	0.05	44	0.001
6	5	19	0.8	32	0.04	45	0.0009
7	2	20	0.7	33	0.03	46	0.0008
8	1.9	21	0.6	34	0.02	47	0.0007
9	1.8	22	0.5	35	0.01	48	0.0006
10	1.7	23	0.4	36	0.009	49	0.0005
11	1.6	24	0.3	37	0.008	50	0.0004
12	1.5	25	0.2	38	0.007	51	0.0003
13	1.4	26	0.1	39	0.006	52	0.0001

Table S3. Parameters for Non-stationary Scenarios: Shape Parameter (k), Central Shape Parameter (k_0), and Random Fluctuation Range ($\pm$b):

#	k_0 [−]	$\pm b$ [−]	*k* [−]	#	k_0 [−]	$\pm b$ [−]	*k* [−]
1	10^{-1}	0.05×10^{-1}	(0.095, 0.105)	11	10^{-1}	0.55×10^{-1}	(0.045, 0.155)
2	10^{-1}	0.10×10^{-1}	(0.090, 0.110)	12	10^{-1}	0.60×10^{-1}	(0.040, 0.160)
3	10^{-1}	0.15×10^{-1}	(0.085, 0.115)	13	10^{-1}	0.65×10^{-1}	(0.035, 0.165)
4	10^{-1}	0.20×10^{-1}	(0.080, 0.120)	14	10^{-1}	0.70×10^{-1}	(0.030, 0.170)
5	10^{-1}	0.25×10^{-1}	(0.075, 0.125)	15	10^{-1}	0.75×10^{-1}	(0.025, 0.175)
6	10^{-1}	0.30×10^{-1}	(0.070, 0.130)	16	10^{-1}	0.80×10^{-1}	(0.020, 0.180)
7	10^{-1}	0.35×10^{-1}	(0.065, 0.135)	17	10^{-1}	0.85×10^{-1}	(0.015, 0.185)
8	10^{-1}	0.40×10^{-1}	(0.060, 0.140)	18	10^{-1}	0.90×10^{-1}	(0.010, 0.190)
9	10^{-1}	0.45×10^{-1}	(0.055, 0.145)	19	10^{-1}	0.95×10^{-1}	(0.005, 0.195)
10	10^{-1}	0.50×10^{-1}	(0.050, 0.150)	20	10^{-1}	0.10×10^{-1}	(0.000, 0.200)