scientific reports

Check for updates

OPEN

A unified probability distribution of second order difference of global streamflow

Haiting Gu[1], Weiping Cheng[1]✉, Hao Chen[1], Li Liu[1], Jingkai Xie[2] & Yue-Ping Xu[1]✉

Hydrological processes, as part of a natural system, are highly complex and chaotic. By analyzing long time series of streamflow data from ~4800 hydrologic stations, it is interesting to find out that probability density functions in second order difference (SOD) of the streamflow data are fat-tailed and bell-shaped curves, and their cumulative distribution functions (CDFs) follow an S-shaped curve (S-curve). We found that t-distribution is a good approximation for S-curve, which uses the degree of freedom (DF) to control the tail thickness. We also found that DF of more than 80% of stations are gathered in the range between 5 and 8. Analysis of the S-curves in seven large river basins indicated that the S-curves can vary with time and space, which is regarded as a good indicator for identifying natural and anthropogenic changes. This study provides a symmetrical, identical and concise probability distribution to describe global streamflow under changing environment.

Keywords Streamflow, S-shaped curve, Second order difference, Probability distribution

Hydrosystem links the environment to human water needs, and understanding hydrological changes is a key to sustainable water management for drinking, sanitation, food, energy, and societal development[1]. Streamflow (Q) is the output of a complex dynamic system, namely hydrosystem, whose predictability is constrained by its inherent variability and sensitivity to numerous factors, often resulting in a limited predictability window. The components of a hydrosystem consist of internal state (S) and external pulse (P), which are assumed to be independent of each other (see Supplementary information). The measurements of S and P of the hydrosystem are sometimes difficult and it is even more difficult to estimate the errors in measurements. Q is an integrated signal, and its subtle feature may be obscured by the integration process. Therefore, we determine the changes of S and P through analyzing the probability changes of the first order ($\dot{Q}$) or second order ($\ddot{Q}$) of Q. There are many probability models for describing Q, including Gamma distribution and Pearson III distribution etc., which are commonly used in flood frequency analysis[2,3]. However, no unique probability exists for descriptions of Q. In order to investigate the probability distribution change of $\dot{Q}$ or $\ddot{Q}$, we need to know if there exists a unified probability function and how the probability function changes in the view of high order with natural and anthropogenic changes.

The streamflow data is of great importance for flood warning[4,5], water management[6,7] and ecosystem heath[8] etc. However, global streamflow is facing huge challenges due to climate change and human activities[9–11], which changes the key elements of the water cycle and causes the variation of the probability distribution of streamflow. It has been asserted that the stationarity assumption for streamflow does not hold anymore under changing environment[12]. Different regions may face different risks under changing environments. In many parts of the world, including Europe, climate-driven changes are already happening and supporting calls for the consideration of climate change in flood risk management[13]. However, natural U.S. surface water supply has increased without a concomitant increase in flooding[14]. To deal with the non-stationarity issue for the streamflow, some approaches are adopted, such as probability distribution with time-varying parameters[15,16] and conditional distribution[17,18].

We turned our attention to the first and higher order differences of the dynamic hydrosystem. A stable dynamic hydrosystem gets rid of the initial value, which is also called an initial-value-insensitivity dynamic system in terms of Lyapunov stability[19]. Such a dynamic system is likely to be symmetrical. We calculated the skewness coefficients (Cs) of the streamflow time series in the first order difference (FOD) and the second order difference (SOD) at ~4800 hydrological gauge stations around the world to verify the assumption. Cs is a measure of the asymmetry of the probability distribution of time series. A zero value of Cs means that the

[1]Institute of Water Science and Engineering, Zhejiang University, Hangzhou 310058, China. [2]School of Earth Sciences and Engineering, Hohai University, Nanjing 211100, Jiangsu, China. ✉email: weipingcheng@zju.edu.cn; yuepingxu@zju.edu.cn

series obey a symmetric distribution in most cases. Compared with *Cs* for FOD, more values of *Cs* for SOD are distributed around 0, indicating that the SOD series is more symmetric. Therefore, we assumed that the second-order-difference dynamic hydrosystem is an initial-value-insensitivity stable system.

In this study, we investigated the variation of the probability distribution of streamflow in the view of second order. We found that the SOD of global streamflow obeys a symmetrical, identical and concise probability distribution. Previous studies have proposed that the Pareto-Burr-Feller probability distribution family could be a good candidate for a unique probability distribution for the streamflow process[20–22]. However, our study focuses on the distribution of SOD of streamflow data. The use of a single-parameter distribution, such as the t-distribution, provides a simple and intuitive approach to capturing statistical patterns in the streamflow process. The t-distribution, with its single parameter (degree of freedom), allows us to explore changes in the distribution tail, offering insights into the natural and anthropogenic variations over time.

Materials and methods

Data and data processing

We used daily streamflow data provided by Global Runoff Data Centre (GRDC) (https://www.bafg.de/GRDC/). Since the streamflow data are insufficient in Asia, we have added our own data from the Yellow River and the Yangtze River. More streamflow data in the Mekong River were also added which were provided by the Mekong River Commission (https://portal.mrcmekong.org).

We selected ~ 4800 stations from GRDC, where the monitoring time of streamflow data is more than 20 years and the missing data is less than 25%. The station, Dresden, located on the Elbe River, has the longest streamflow data, lasting from 1806 to 2016 without missing data. The shortest streamflow data lasts from 1982 to 2001, with 24% missing data in the station Catarama on Zapotal River. The missing data in streamflow are noted as null values. Hence, their corresponding values in FOD and SOD are also null and they are eliminated in the analysis.

Definition of time series differences

Here, we defined the calculation formulas of FOD and SOD. There are three different definitions of the first order difference of a time series as follows:

$$\Delta f(i) = f(i) - f(i-1) \tag{1}$$

$$\Delta f_m(i) = \frac{f(i) - f(i-1)}{(f(i) + f(i-1))/2} \tag{2}$$

$$\Delta f_{\ln}(i) = \ln(f(i)) - \ln(f(i-1)) \tag{3}$$

where $f(i)$ is the value of time series at time *i*, $\Delta f(i)$ is the first order difference, $\Delta f_m(i)$ is the difference in the ratio of the time series to the mean value of the nearest two samples, and $\Delta f_{\ln}(i)$ is the logarithmic first order difference. Equation (1) can be used when the sample does not vary greatly; Eq. (2) is recommended for the time series with great changes and it also makes the data dimensionless so that different time series can be compared with each other; Eq. (3) can be used when a positive variable changes to a large degree. In our study, we compared the streamflow of different stations, so Eq. (2) is adopted. The formula of SOD is shown as Eq. (4):

$$\Delta^2 f_m(i) = \Delta f_m(i) - \Delta f_m(i-1) \tag{4}$$

where $\Delta^2 f_m(i)$ is SOD of the streamflow at Time *i*.

Optimized correlation coefficient-based knee point detecting method

The optimized correlation coefficient-based knee point detecting method is developed to detect knee points of an S-curve. In our study, we assumed that two knee points (break points) split the S-curve into three sections. A graphic sketch of the knee points of an S-curve is shown in Fig. S3. Three linear functions are used to fit the three curves.

$$\hat{y}_i = \begin{cases} a_1 x_i + b_1, & 1 \leqslant i < n_1 \\ a_2 x_i + b_2, & n_1 \leqslant i < n_2 \\ a_3 x_i + b_3, & n_2 \leqslant i < n \end{cases} \tag{5}$$

where $\hat{y}_i$ is the estimated value of y_i; n_1 is the first break point; n_2 is the second break point; n is the length of the series. Then we need to find two break points to satisfy Eq. (6).

$$\max r = \frac{\sum_{t=1}^{n} (y_t - \bar{y}_t)(\hat{y}_t - \bar{\hat{y}}_t)}{\sqrt{\sum_{t=1}^{n} (y_t - \bar{y}_t)^2} \sqrt{\sum_{t=1}^{n} (\hat{y}_t - \bar{\hat{y}}_t)^2}} \tag{6}$$

where $\bar{y}_i$ and $\bar{\hat{y}}_i$ are the mean value of y_i and $\hat{y}_i$, respectively. In our study, there are two knee points to be detected. The number of points on the S-curve is very large (the length of the series y_i is large). Hence, we used the global search algorithm[23] to search the points n_1 and n_2, where the optimization objective is maximizing the correlation coefficient shown in Eq. (6).

S-curve and its approximation

We calculated the empirical CDF of SOD of the streamflow at 4800 stations. Then the probability values in the y axis are converted into the quantiles of the Gaussian distribution. We found that the central section of the CDF curve matches well a Gaussian distribution, but both ends deviate substantially from the straight line (the Gaussian CDF curve), forming a symmetric S-curve. By definition of differential statistics, the mean value of SOD is zero or close to zero[24]. The probability density function (PDF) of SOD is a bell-shaped symmetric curve. The CDF curves are S-shaped, which shows the feature of a fat-tailed distribution. In our study, we used *t*-distribution to approximate the S-curves. The PDF and CDF of *t*-distribution are shown as Eqs. (7) and (8) respectively.

$$f(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi}\Gamma(\frac{\nu}{2})}(1+\frac{t^2}{\nu})^{-\frac{\nu+1}{2}} \tag{7}$$

$$F(t) = \int_{-\infty}^{t} f(x)dx \tag{8}$$

where ν is DF and $\Gamma()$ is the gamma function. It is clear that in *t*-distribution, there is only one parameter, DF, to control the shape of the PDF and CDF. With the increase of DF, the t-distribution gets closer to the Gaussian distribution.

Results and discussion

A unified distribution for S-curve of streamflow on a global scale

We calculated the cumulative distribution function (CDF) of these SOD series and drew these CDF curves in the Gaussian probability paper, where the y-axis was converted into probability values of the Gaussian distribution. Interestingly, we found that all curves follow a symmetric S-curve. Although the middle parts of the CDF curves match a normal distribution, the two tails of the CDF curves deviate substantially from the straight line indicated by the normal distribution (see Fig. 1d). To find the knee points of these S-curves, we used an optimized correlation coefficient-based knee point detecting method described in "Optimized correlation coefficient-based knee pointdetecting method". Figure 1a presented the knee points detected. In these S-curves, there are two knee points, noted as left knee point and right knee point. For more than 90% of stations, the left knee points are distributed in the probability of 15–25% and the right knee points are distributed in the probability of 75–85%. This confirms that S-curves of most stations are symmetric. Influenced by streamflow measurement errors or artificial flow regulation, the knee points of S-curves in some stations (less than 10%) are found around the probability of 40% and 60%. However, these curves are also S-shaped.

As the central parts of the curves coincide with the Gaussian distribution but both tails are fat, we made a hypothesis that the curves follow a *t*-distribution. We adopted the Gaussian distribution and the *t*-distribution to fit the CDF curves with maximum likelihood estimate with the log-likelihood function (MLE)[25] and calculated their coefficients of determination (r^2). Here, we used the parts of curves with a probability less than 20% and large than 80% as the tails of the S-curves. The curve approximation with *t*-distribution shows better performance than that with Gaussian distribution (Fig. 1b) in the two tails of the S-curves. The average value of r^2 with *t*-distribution is 0.92, higher than that with Gaussian distribution at the value of 0.87. The peak of r^2 with *t*-distribution ranges from 0.95 to 0.96, better than that with Gaussian distribution at the range from 0.89 to 0.90. This means that the *t*-distribution is a good approximation for the S-curve. The spatial distribution of r^2 for global streamflow data is displayed in Fig. 1c. Since more than 90% of r^2 are larger than 0.8, which indicates that the *t*-distribution approximation can be applied globally, we categorized r^2 ranged from 0.8 to 1 in more detail. The goodness of fit is partly related to station location, which is closely related to its climate and surface conditions.

Distribution features of S-curve on a global scale

To study the distribution features of well-identified S-curves, we used the *t*-distribution to describe the S-shape curve of the cumulative distribution function (CDF). The *t*-distribution is symmetric and bell-shaped. Compared with the Gaussian distribution, the *t*-distribution has fatter tails. In the *t*-distribution, the parameter, degrees of freedom (DF), controls the tail thickness and a smaller DF means a fatter tail. As the value of DF grows, the *t*-distribution approaches the Gaussian distribution. Figure 2 shows the relationship between DF and r^2. This figure shows that the DF values of more than 80% of stations are ranged from 5 to 8 (Fig. 2d). With the growth of DF, the goodness of fit (described with r^2) is also improved on the whole. As shown in Fig. 2, a smooth curve can be found as the upper bounds of r^2 against different DFs. This upper bound follows an inverse proportional function which is shown in Fig. S1. We marked three points in Fig. 2 to investigate the features of S-curves. As shown in Fig. 2a, the CDF curve of point a, which has a relatively large DF, is close to a Gaussian CDF curve. Figure 2b and c show that the *t*-distribution CDF curves have a good performance to fit the fat tails. The red dashes in Fig. 2 are the lines connecting the first and third quartiles, which follow Gaussian distribution and represent Gaussian CDF curves. From Fig. 2a-c, we can find that the central parts of the CDF curves coincide with the Gaussian CDF curve, but the two tails of the CDF curves deviate from the Gaussian CDF curve and accord with the *t*-distribution CDF curve. With the decrease of r^2, the number of zero SOD is increasing and the central parts of the S-curves are deviating from the t-distribution CDF curve. Such phenomenon is likely caused by streamflow measurement errors or artificial regulated flow. Anyway, the *t*-distribution can well describe the shape of S-curve.

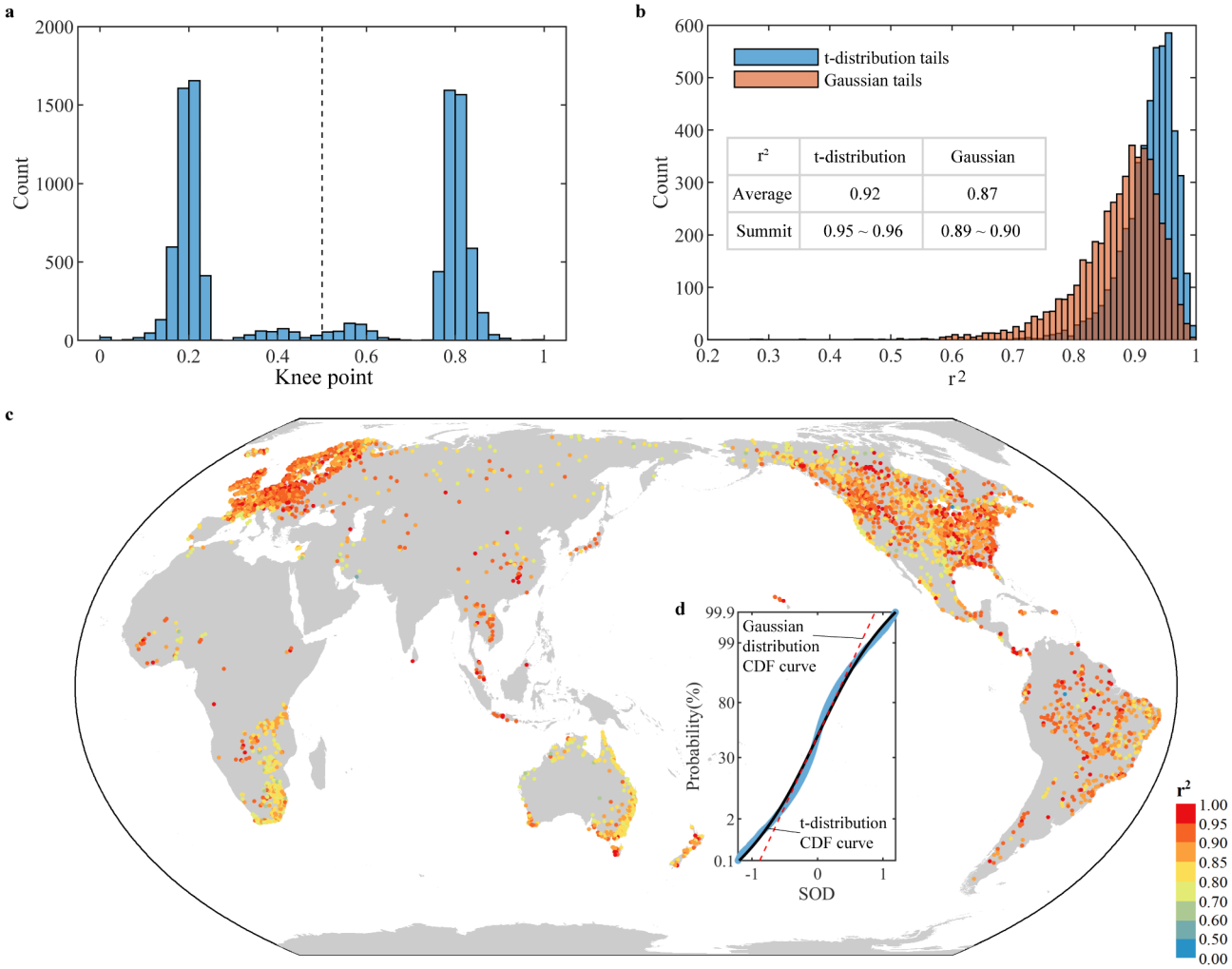

Fig. 1. Approximation of the S-curve at ~4800 stations. (**a**) Position of knee point. (**b**) Performance of the approximation with t-distribution and Gaussian distribution in the two tails of the curves. (**c**) Map of the performance of the approximation with t-distribution. (**d**) Sample of the S-curve and its approximations.

Role of S-curves in identifying natural and anthropogenic changes

We choose seven large well-known river basins to figure out how the S-curves change from upstream to downstream and from basin to basin (Fig. 3). All the hydrological stations we studied are located in the main streams of our study river basins and listed with an order from upstream to downstream in Table S1. For comparison purposes, we normalized the SOD series of all stations to scale the part between 25th and 75th percentiles into the range from −1 to 1. As a result, the S-curves of all the stations can be drawn in a unified coordinate system with the range of x-axis from −5 to 5. Figure 3 confirms previous results that the central parts of the curves coincide with the Gaussian distribution. Here, we defined the value of the 75th percentile (or the minus value of the 25th percentile) of SOD series as the base value of the SOD series.

Figure 3b shows the relationship between the standard deviations (SD) of normalized SOD (NSOD) and DF in the seven basins. We found that SD has an inverse relationship with DF. The scatter plot of SD of NSOD versus DF is divided into two parts (A and B in Fig. 3b) at the position, DF of 6 and SD of 3. In Part B, the values of SD are larger and the values of DF are smaller, which means that the two tails of these stations deviate from the base value more seriously. In other words, the S-shape curves of these stations have fatter tails. The stations in Part B are in more arid climate areas than those in Part A. According to the Köppen-Geiger climate classification[26], the stations in Part B are mainly classified into BWh (Arid-Desert-Hot climate), BSh (Arid-Steppe-Hot climate) and BSk (Arid-Steppe-Cold climate). The base values of SOD in the arid areas are usually small and the variance in the tails can be enlarged. This may lead to the phenomenon of fatter tails in Part B, but this conclusion needs to be drawn more carefully after further studies. In Fig. 3b, we also found that the values of DF in different basins are mixed. The characteristics of different basins are mainly presented in the form of the base values of SOD.

Figure 3c-i show the S-curves of seven large river basins and the plots of DF values from upstream to downstream in each basin can be seen in Fig. S2. In general, compared with the upstream stations, the downstream stations have larger DF values. This occurs particularly in the Yangtze River, Murray-Darling River, Amazon River, Mississippi River and Rhine River. The larger DF means a thinner tail and lower probability that extreme

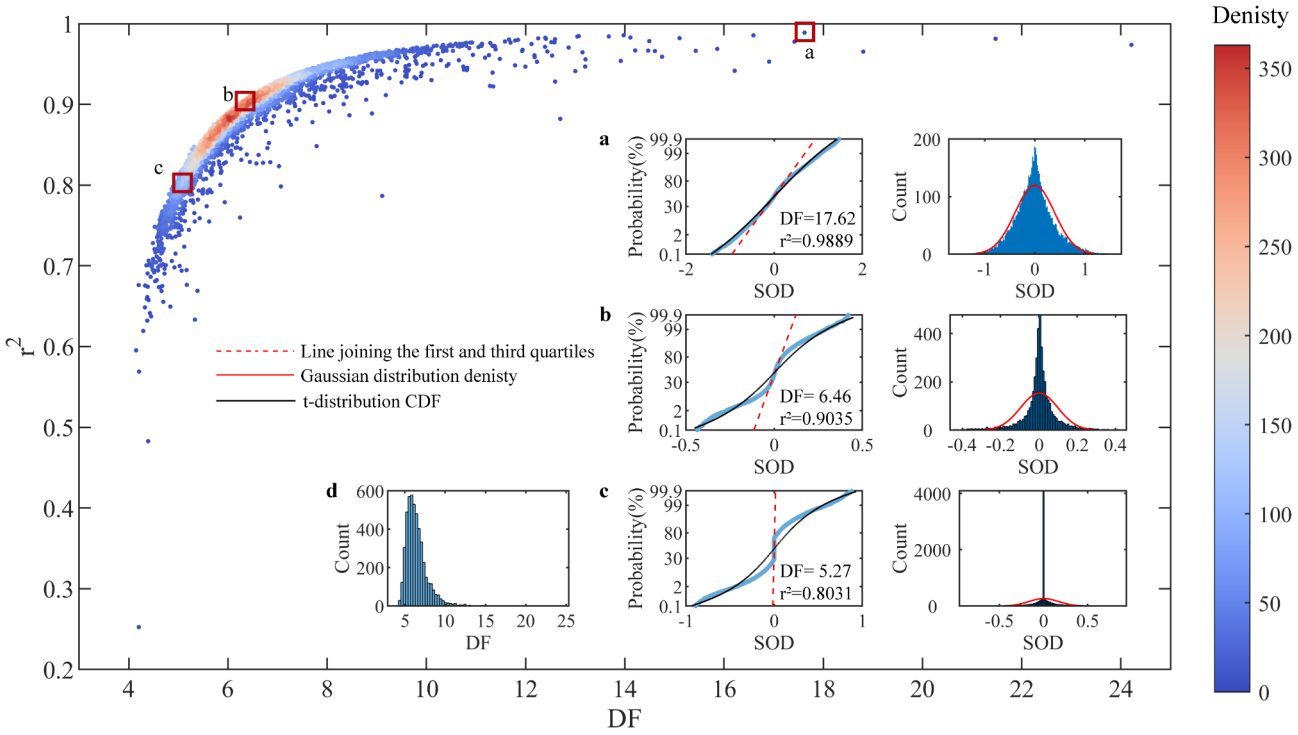

Fig. 2. Degree of freedom (DF) versus coefficient of determination (r^2). (**a**–**c**), CDF curves in the left panel and histograms of second order difference (SOD) in the right panel at three special points marked with red boxes. (**d**) Histogram of DF.

SOD occurs. This indicates that the streamflow at the upstream stations is easier to be influenced by natural and anthropogenic changes. In the Rhine River (RH), the value of DF decreases sharply from the 3rd station (Station 6935055) to the 4th station (Station 6935054). We found that there is a huge change in climate classification between the 2nd station (Station 6935500) and the 3rd station. Then the DF values have an increasing trend after the 4th station in RH. In the Niger River (NG), the DF values increase in the first four stations and then decrease. The annual mean streamflow from the 4th station (Station 1134700) to the 7th station (Station 1234150) also decreases, as the river goes through the Sahara Desert. This confirms that the streamflow of arid areas is easier to be influenced by natural and anthropogenic changes. In the Mekong River (MK), the DF values from upstream to downstream change slightly, which may contribute to the small change of environmental condition. The DF values in the Murray-Darling River (MD) change little. As seen from Fig. 3b, we found that the values of SD in MD have a large difference in the four stations. However, when the DF is less than 6, the sensitivity of DF to SD decreases. Hence, the values of DF at the four stations in MD are very close.

To investigate the temporal changes of S-curves in the recent 100 years, we chose three stations with no missing data from the Mississippi River, the Rhine River and the Mekong River. We adopted a one-year-lagged moving window to partition the whole studied period into different, but overlapping and sequential 20-year periods. The S-curves in these periods are shown in Fig. 4a–c. The S-curves of each station are changing with time, but the variation extent is not large. We used the first 20-year S-curve as a baseline. We found that S-curves drift away from the baselines in the first 50 years of the study period, and tend to converge toward the baselines in the second 50 years at all three stations. One interesting fact is that the Hurst exponent of the DF for all three stations is greater than 0.5, indicating long-term persistence in the DF variations[27]. There are abrupt changes of the DF series in all three stations. In the Mississippi River and the Rhine River, the DF values first increase and after 1970s the values of DF begin to decrease. Meanwhile, the DF values in the Mekong River are stable until 1940s. The DF values decrease sharply from 1950 to 1970, followed by a steady increase post-1970s. Such changes are caused by both natural and anthropogenic changes[28–31]. Table S3 presents the major water infrastructure projects along the Mississippi, Rhine, and Mekong Rivers. Our analysis reveals that during the construction phases of significant flood control measures (1880–1950 for the Mississippi, 1910–1950 for the Rhine, and 1990–2020 for the Mekong), the DF values increased, which aligns with the observed reduction of extreme values in the distribution tail as the DF increased. Conversely, in regions with integrated conservation projects, a decline in DF values was observed.

Figure 4d–f shows the daily SOD plots, where the red lines are the 5th and 95th percentiles of 1-year-length SOD. The range between the 5th and 95th percentiles is defined as a normal range of SOD. SOD out of the normal range is extreme SOD. We found that the Hurst exponent of the upper and lower boundaries of the normal range is greater than 0.5, indicating long-term persistence. In the Mississippi River, the normal range has an increasing trend and the range of extreme SOD is narrowed. In the Rhine River, the normal range of SOD changes slightly, but the extreme SOD is closer to 0 from 1900 to 2016. In the Mekong River, the normal range

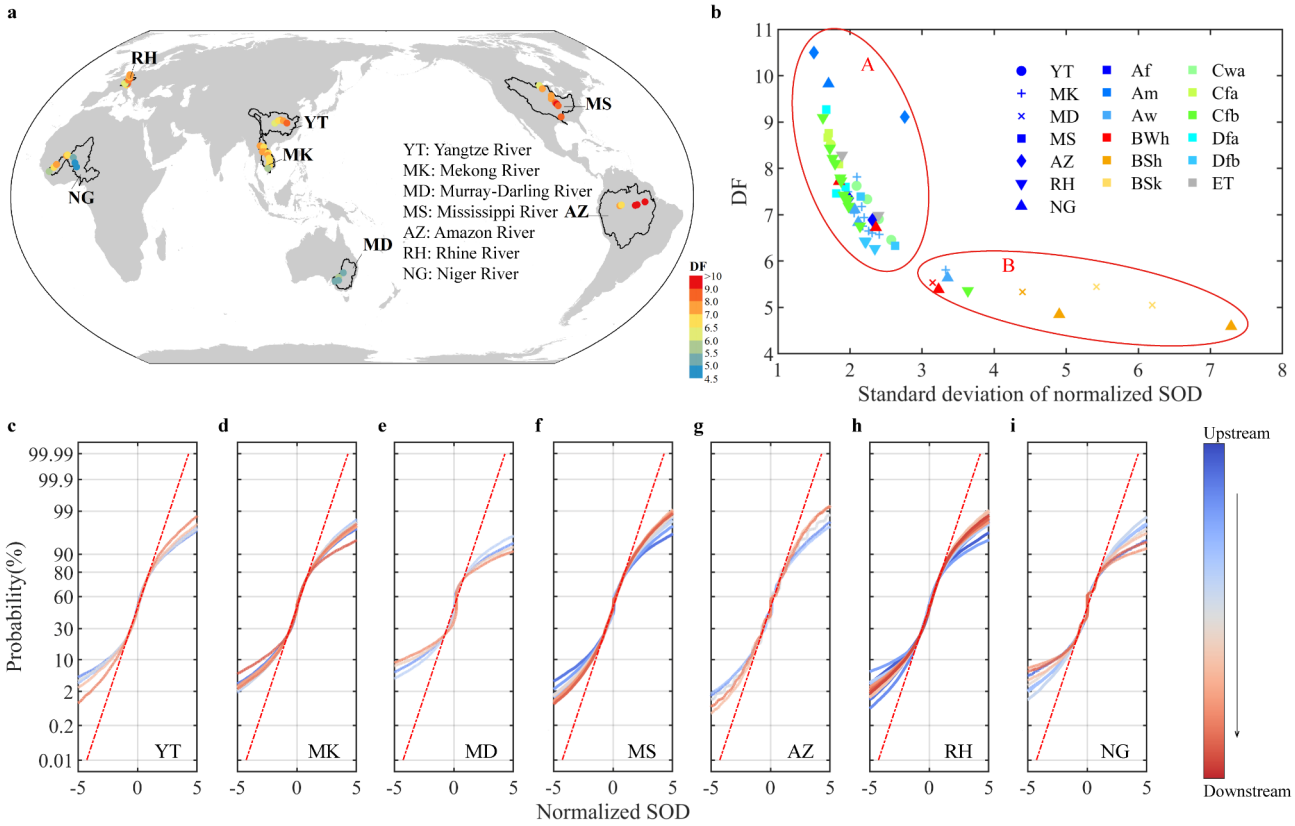

Fig. 3. (**a**) Locations of seven representative river basins. (**b**) Relationship between standard deviations (SD) of normalized SOD (NSOD) and degrees of freedom (DF). (**c**-**i**) CDF curves of NSOD. The colors in (**b**) represent different climate classifications according to Köppen-Geiger climate classification[26] and the meaning of these abbreviations can be seen in Table S2. The red dashes in (**c**-**i**) are the standard Gaussian distributions.

of SOD has a decreasing trend. The variations of the normal range and extreme SOD are consistent with the variations of DF. It makes sense that the normal range and extreme SOD decided the shape of S-curve, where a larger normal range and smaller extreme SOD means a larger DF. The analysis above shows that the shape of S-curves changes with natural and anthropogenic changes. In the Mississippi River, the DF value ranges from 7 to 11, which is broader than the other two river basins. This indicates the streamflow in the Mississippi River is more strongly influenced by natural and anthropogenic changes. However, the curves remain S-shaped, indicating the level of non-stationarity in the last 100 years.

Conclusion

In this study, we found that for the SOD of streamflow, the CDF curve has a good symmetry, which is presented as a S-curve. This S-curve exists widely in the streamflow of the global rivers. We used *t*-distribution to fit these S-curves and found that r^2 of most stations are above 0.8, suggesting that the *t*-distribution provides a suitable model for characterizing the S-shaped cumulative distribution curve. The *t*-distribution is a very concise function, which only uses one parameter DF to control the shape of the distribution. Since most stations we used in this study are distributed in America and Europe, this conclusion still needs to be further verified with more streamflow data in other regions. In the seven large river basins we picked out, the *t*-distribution also fits well. As stated before, the S-curves can be a good indicator to help identify natural and anthropogenic changes. Furthermore, we calculated the Hurst exponent of the 20-year-window DF series and normal range boundaries of SOD for three representative stations along the Mississippi, Rhine, and Mekong Rivers. The relatively high Hurst values indicate that strong long-term persistence is still present, even after second-order differencing. This suggests that the streamflow variation patterns are still influenced by past changes in the SOD perspective. The unified probability function to describe global streamflow provides a new way to understand the dynamic hydrosystem under environmental changes. Applications of such unified probability function can give very useful insights for hydrologic frequency analysis, flood risk assessment and sustainable water management.

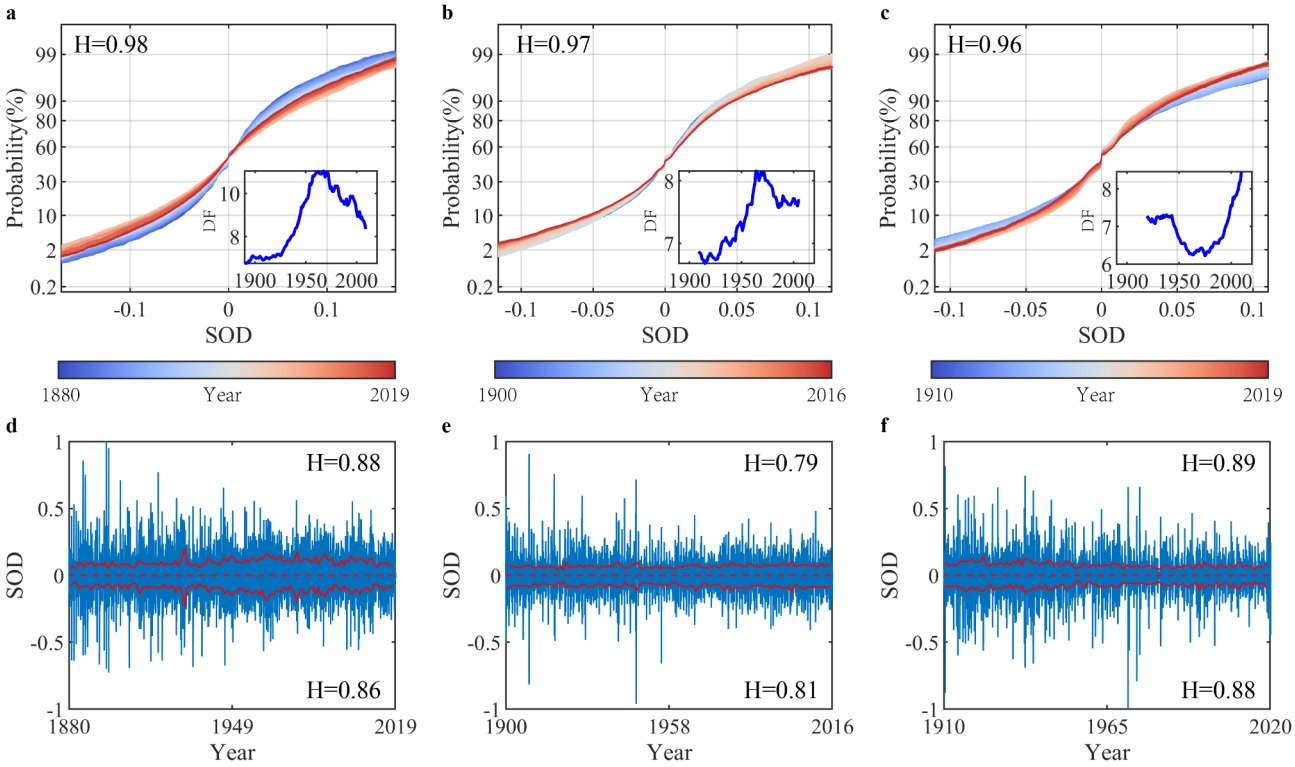

Fig. 4. CDF curves of the second order difference (SOD) and DF plots of SOD in the first line, temporal variability of SOD in the second line. (**a**,**d**) Station 4,127,503 in the Mississippi River; (**b**,**e**) Station 6,335,050 in the Rhine River. (**c**,**f**) Station 14,501 in the Mekong River. The red lines represent the 5% upper and lower boundaries of the SOD series. H presents the Hurst exponent of DF in (**a**–**c**) and the Hurst exponent of the boundaries in (**d**–**f**).

Data availability

The streamflow data produced by this study are available at Global Runoff Data Centre (https://www.bafg.de/GRDC/) and the Mekong River Commission (https://portal.mrcmekong.org).

Received: 29 September 2024; Accepted: 9 April 2025
Published online: 24 April 2025

References

1. Montanari, A. et al. Panta Rhei—Everything flows: change in hydrology and society—The IAHS scientific decade 2013–2022. *Hydrol. Sci. J.* **58**, 1256–1275 (2013).
2. Xie, P., Gu, H., Sang, Y., Wu, Z. & Singh, V. P. Comparison of different methods for detecting change points in hydroclimatic time series. *J. Hydrol.* **577**, 123973 (2019).
3. Yue, S., Ouarda, T. B. & Bob, E. E. A review of bivariate gamma distributions for hydrological application. *J. Hydrol.* **246**, 1–18 (2001).
4. Chai, Y. et al. Influence of climate variability and reservoir operation on streamflow in the Yangtze river. *Sci. Rep.* **9**, 1–10 (2019).
5. Tanoue, M., Hirabayashi, Y. & Ikeuchi, H. Global-scale river flood vulnerability in the last 50 years. *Sci. Rep.* **6**, 1–9 (2016).
6. Chen, W. & Olden, J. D. Designing flows to resolve human and environmental water needs in a dam-regulated river. *Nat. Commun.* **8**, 1–10 (2017).
7. Poff, N. L. et al. Sustainable water management under future uncertainty with eco-engineering decision scaling. *Nat. Clim. Chang.* **6**, 25–34 (2016).
8. Chen, J., Wang, Y., Li, F. & Liu, Z. Aquatic ecosystem health assessment of a typical sub-basin of the Liao river based on entropy weights and a fuzzy comprehensive evaluation method. *Sci. Rep.* **9**, 1–13 (2019).
9. Chai, Y. et al. Influence of climate variability and reservoir operation on streamflow in the Yangtze river. *Sci. Rep.* **9**, 5060 (2019).
10. Lv, X., Zuo, Z., Ni, Y., Sun, J. & Wang, H. The effects of climate and catchment characteristic change on streamflow in a typical tributary of the yellow river. *Sci. Rep.* **9**, 1–10 (2019).
11. Tan, X. & Gan, T. Y. Contribution of human and climate change impacts to changes in streamflow of Canada. *Sci. Rep.* **5**, 17767 (2015).
12. Milly, P. et al. Stationarity is dead: whither water management? *Earth* **4**, 20 (2008).
13. Blöschl, G. et al. Changing climate both increases and decreases European river floods. *Nature* **573**, 108–111 (2019).
14. Lins, H. F. & Slack, J. R. Seasonal and regional characteristics of U.S. Streamflow trends in the united States from 1940 to 1999. *Phys. Geogr.* **26**, 489–501 (2005).
15. Gilroy, K. L. & McCuen, R. H. A nonstationary flood frequency analysis method to adjust for future climate change and urbanization. *J. Hydrol.* **414**, 40–48 (2012).
16. López, J. & Francés, F. Non-stationary flood frequency analysis in continental Spanish rivers, using climate and reservoir indices as external covariates. *Hydrol. Earth Syst. Sci.* **17**, 3189–3203 (2013).

17. Singh, V. P., Wang, S. X. & Zhang, L. Frequency analysis of nonidentically distributed hydrologic flood data. *J. Hydrol.* **307**, 175–195 (2005).
18. Yan, L. et al. Reducing uncertainty of design floods of two-component mixture distributions by utilizing flood timescale to classify flood types in seasonally snow covered region. *J. Hydrol.* **574**, 588–608 (2019).
19. Lyapunov, A. M. The general problem of the stability of motion. *Int. J. Control.* **55**, 531–534 (1992).
20. Dimitriadis, P., Koutsoyiannis, D., Iliopoulou, T. & Papanicolaou, P. A global-scale investigation of stochastic similarities in marginal distribution and dependence structure of key hydrological-cycle processes. *Hydrology* **8**, 59 (2021).
21. Koutsoyiannis, D., Yao, H. & Georgakakos, A. Medium-range flow prediction for the nile: a comparison of stochastic and deterministic methods/prévision du débit du nil à Moyen Terme: Une comparaison de méthodes stochastiques et déterministes. *Hydrol. Sci. J.* **53**, 142–164 (2008).
22. Pizarro, A., Dimitriadis, P., Iliopoulou, T., Manfreda, S. & Koutsoyiannis, D. Stochastic analysis of the marginal and dependence structure of streamflows: from fine-scale records to multi-centennial paleoclimatic reconstructions. *Hydrology* **9**, 126 (2022).
23. Ugray, Z. et al. Scatter search and local NLP solvers: A multistart framework for global optimization. *Informs J. Comput.* **19**, 328–340 (2007).
24. Koutsoyiannis, D. Time's arrow in stochastic characterization and simulation of atmospheric and hydrological processes. *Hydrol. Sci. J.* **64**, 1013–1037 (2019).
25. Forbes, C., Evans, M., Hastings, N. & Peacock, B. *Statistical Distributions* (Wiley, 2011).
26. Peel, M. C., Finlayson, B. L. & McMahon, T. A. Updated world map of the Köppen-Geiger climate classification. *Hydrol. Earth Syst. Sci.* **11**, 1633–1644 (2007).
27. Wang, H., Song, S., Zhang, G., Ayantobo, O. O. & Guo, T. Stochastic volatility modeling of daily streamflow time series. *Water Resour. Res.* **59**, e2021WR031662 (2023).
28. Best, J. Anthropogenic stresses on the world's big rivers. *Nat. Geosci.* **12**, 7–21 (2019).
29. Chiang, F., Mazdiyasni, O. & AghaKouchak, A. Evidence of anthropogenic impacts on global drought frequency, duration, and intensity. *Nat. Commun.* **12**, 1–10 (2021).
30. Munoz, S. E. & Dee, S. G. El Niño increases the risk of lower Mississippi river flooding. *Sci. Rep.* **7**, 1–7 (2017).
31. Slater, L. J., Khouakhi, A. & Wilby, R. L. River channel conveyance capacity adjusts to modes of climate variability. *Sci. Rep.* **9**, 1–10 (2019).

Acknowledgements

The authors would like to thank China National Key Research for Development Plan (2021YFC3201105), Natural Science Foundation of Zhejiang Province (LMS25E090004), and National Natural Science Foundation of China (52479028; 52209036; 52109037) for the financial support. We are grateful to Global Runoff Data Centre (https://www.bafg.de/GRDC/) and the Mekong River Commission (https://portal.mrcmekong.org).

Author contributions

H.G. and Y.P.X. conceived the idea, designed the research, analyzed the data and drafted the manuscript. W.C. conceived the idea and helped data analysis. H.C. contributed to prepare the data. L.L. and J.X. contributed to analyze the data. All authors reviewed the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-98191-w.

Correspondence and requests for materials should be addressed to W.C. or Y.-P.X.