water

Editorial

Statistics in Hydrology

Yuanfang Chen [1], Dong Wang [2], Dedi Liu [3], Binquan Li [4,5,*] and Ashish Sharma [6]

1 College of Hydrology and Water Resources, Hohai University, Nanjing 210098, China; chenyuanfang@hhu.edu.cn
2 Department of Hydrosciences, School of Earth Sciences and Engineering, State Key Laboratory of Pollution Control and Resource Reuse, Nanjing University, Nanjing 210046, China; wangdong@nju.edu.cn
3 State Key Laboratory of Water Resources and Hydropower Engineering Science, Wuhan University, Wuhan 430072, China; dediliu@whu.edu.cn
4 State Key Laboratory of Simulation and Regulation of Water Cycle in River Basin, China Institute of Water Resources and Hydropower Research, Beijing 100038, China
5 State Key Laboratory of Hydrology-Water Resources and Hydraulic Engineering, Hohai University, Nanjing 210098, China
6 School of Civil and Environmental Engineering, University of New South Wales, Sydney, NSW 2052, Australia; a.sharma@unsw.edu.au
* Correspondence: libinquan@hhu.edu.cn

1. Introduction

Statistical methods have a long history in the analysis of hydrological data for designing, planning, infilling, forecasting, and specifying better models to assess scenarios of land use and climate change in catchments. The effectiveness of statistical descriptions of hydrological processes reflects the enormous complexity of hydrological systems, which makes a purely deterministic description ineffective. Many different statistically oriented methodologies are used in hydrological studies, with multiple examples in the literature where the statistical aspect of the work is incomplete, unreasonable, or even unacceptable.

On 19–20 October 2019, the 10th International Workshop on Statistical Hydrology (STAHY 2019), which was organized by the International Commission on Statistical Hydrology, International Association of Hydrological Sciences (ICSH-IAHS), took place in Nanjing, China. A total of 132 participants from 13 countries including Australia, Belgium, Canada, China, Germany, Italy, Poland, Spain, Switzerland, the United States, etc., registered to participate in the conference, and more than 300 graduate students and young scholars from universities in Nanjing also attended the conference. In addition, 28 early career scholars from eight countries participated in the first Early Career Course (ECC) of ICSH held in Nanjing. The ECC was held in advance of the STAHY 2019 workshop for one day (October 18). Multiple new academic innovations and achievements were inspired during and after the conference. The authors of the presentations proposed to publish their research results in the Special Issue of "Statistics in Hydrology". Thus, the main purpose of this Special Issue is to share the latest research in the field of statistics in hydrology with reference to the discussions held during STAHY 2019.

check for updates

Citation: Chen, Y.; Wang, D.; Liu, D.; Li, B.; Sharma, A. Statistics in Hydrology. *Water* **2022**, *14*, 1571. https://doi.org/10.3390/w14101571

Received: 9 May 2022
Accepted: 11 May 2022
Published: 13 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

2. Overview of the Contributions

The call for papers was announced in April 2021, and after a rigorous peer-review process, a total of 11 papers have been published [1–11]. To gain a better insight into the essence of the Special Issue, a brief overview of these papers is shown below.

Statistical methods provide effective tools for the analysis of changes in hydrometeorological variables and extreme hydrological/meteorological events, as well as the correlation between different variables. Ahemaitihali and Dong [1] analyzed the spatiotemporal characteristics and driving forces of flash floods in the Altay Prefecture, China. They examined the kernel density, standard deviational ellipse, and spatial gravity center of flash floods and analyzed the temporal and spatial variations. Several statistical methods including multiple linear regression, principal component analysis and random forest were used to analyze the driving force of flash floods. This study uses a variety of statistical methods to provide insights into the spatial–temporal dynamics of flash floods and a reference case for similar studies. Dong et al. [2] applied multiresolution analysis and continuous wavelet analysis approaches to evaluate the influences of precipitation and river level fluctuations on groundwater. Their results showed that the wavelet technique was more effective than spectral analysis in detecting the correlations among precipitation, river level, and aquifer level in the temporal domain. Lang et al. [3] explored the characteristics of extreme precipitation events over the Henan province, China using the latest ERA5 dataset. Their results showed that precipitable water, wind, and relative humidity were the most common drivers for extreme precipitation events over the Henan province for the 1981–2021 period. This study provides insights for further flood estimations and forecasts.

The forecasting and prediction of hydrological elements such as precipitation and runoff is one of the main topics in the hydrological field. The statistical method is central to this. He et al. [4] developed statistical forecasting models for summer rainfall and streamflow over the Yangtze River valley using a relatively small number of samples. The logistic regression was used to make probability forecasts for the binary classification, and then three testing procedures, i.e., predictability assessment (PA), model output statistics (MOS), and the reanalysis-based (RAN) approach, were explored. Their results showed that the RAN approach was better at generating exceedance probability forecasts than MOS, as RAN can utilize more samples. The findings provide a useful reference for these statistical methods in long-term hydrological forecasts. Data-driven machine learning (ML) methods are also widely used in hydrology. Under the stacking ensemble learning framework, Gu et al. [5] used four ML models, namely k-nearest neighbors (KNN), extreme gradient boosting (XGB), support vector regression (SVR), and artificial neural networks (ANN), for monthly rainfall prediction. Their results showed that the stacking ensemble learning method can synthesize the advantages of each single ML model to produce a good simulation accuracy in their study area (Taihu Basin, China).

Uncertainty in hydrological forecasting has always existed. Statistical methods are an effective means to quantitatively deal with this uncertainty. Romero-Cuellar et al. [6] developed an extension of the Model Conditional Processor (MCP), which merged clusters with Gaussian mixture models to offer an alternative solution to manage heteroscedastic errors. Case studies indicated that this new post-processer had significant potential in generating more reliable, sharper, and more accurate monthly streamflow predictions than the MCP and MCP using a truncated normal distribution, especially in dry catchments.

Statistical methods are highly important in engineering hydrological design, such as sample processing, model fitting, and statistical parameter estimation in frequency analysis. The Special Issue published five papers on this topic [7–11]. Among them, the first two papers [7,8] are mainly about model selection and parameter estimation in frequency analysis under stationary conditions, and the last three papers [9–11] mainly focus on frequency analysis under nonstationary conditions caused by changing environment. Shao et al. [7] revised the method for the regional frequency analysis of extreme precipitation using regional L-moments methods. A Monte Carlo (MC) simulation was conducted to determine the appropriate probability distribution. A case study in Jiangsu province, China shows that the frequency estimations based on this revised regional frequency analysis are in good agreement with observations. This study provides a new perspective in regional frequency analysis. Song et al. [8] used the maximum likelihood estimation (MLE) method to estimate the parameters and confidence intervals of quantiles of the four-parameter exponential gamma (FPEG) distribution. The FPEG distribution was then applied to precipitation data of the Weihe watershed in China. The results showed that FPEG distribution is a good candidate for modeling annual precipitation data. Considering that the use of the FPEG distribution has received only limited attention from the hydrological community, this finding may provide guidance for estimating design values of random variables in other parts of the world.

Due to the changing environment, hydrological design such as frequency analysis faces nonstationary difficulties. Zeng et al. [9] used Bayesian nonstationary time-varying moment models to investigate the annual maximum flood peak (AMFP) risk and return period. Two climate covariates were defined to exhibit a significant positive correlation with AMFP. The results indicated that the climate-informed model demonstrated the best performance on extreme flood frequency analysis as well as sufficiently explaining the variability of extreme flood risk. This study provides a good alternative to extreme flood frequency analysis under the context of climate change. Li et al. [10] found that the 1-day annual maximum flood volume (AMFV) exhibits a significant correlation with AMFP in a case study at Longmen Reservoir (China). Moreover, they developed a copula-based bivariate nonstationary flood frequency analysis to investigate environmental effects on the dependence of flood peak and volume. The results showed that the design floods estimated by bivariate nonstationary joint distribution would increase largely compared with the ones estimated in a univariate nonstationary context. On this subject, Li and Qin [11] introduced the mechanism-based reconstruction (Me-RS) method into this topic and demonstrated a case study on the calculation of design annual runoff under nonstationary conditions in the Jialu River Basin, China. In the Me-RS framework, the nonstationary hydrological series was transformed into stationary series for forthcoming hydrological design calculation. The results showed that the Me-RS method not only had theoretical support, but also its obtained design values were consistent with the actual condition and had much smaller uncertainty. Thus, the method provides an effective tool for annual runoff frequency analysis under nonstationary conditions. The findings are very useful because the statistical characteristics of many rivers around the world exhibit complex nonstationary changes.

3. Conclusions

The above 11 papers contribute to the increasing interest in the studies of hydrometeorological changes, hydrological prediction and uncertainty analysis, and engineering hydrological design under stationary/nonstationary conditions. The Guest Editors hope that readers will be inspired by this Special Issue and further innovate in the field of statistical hydrology. In particular, the era of "big data" has arrived, which will bring new development opportunities to statistical hydrology.

Funding: This research was funded by the National Natural Science Foundation of China, grant number 41877147, and the Open Research Fund of State Key Laboratory of Simulation and Regulation of Water Cycle in River Basin (China Institute of Water Resources and Hydropower Research), grant number IWHR-SKL-202013.

Acknowledgments: The authors of this paper, who served as the Guest Editors of this Special Issue, wish to thank the journal editors, all authors submitting papers, and the referees who contributed to revise and improve the 11 published papers. In addition, we would like to thank Xia Wu of Hohai University and Shanika Zhu, section managing editor of this journal for their contributions to communication and coordination in the organization of this Special Issue.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ahemaitihali, A.; Dong, Z. Spatiotemporal Characteristics Analysis and Driving Forces Assessment of Flash Floods in Altay. *Water* **2022**, *14*, 331. [CrossRef]
2. Dong, L.; Guo, Y.; Tang, W.; Xu, W.; Fan, Z. Statistical Evaluation of the Influences of Precipitation and River Level Fluctuations on Groundwater in Yoshino River Basin, Japan. *Water* **2022**, *14*, 625. [CrossRef]
3. Lang, Y.; Jiang, Z.; Wu, X. Investigating the Linkage between Extreme Rainstorms and Concurrent Synoptic Features: A Case Study in Henan, Central China. *Water* **2022**, *14*, 1065. [CrossRef]
4. He, R.; Chen, Y.; Huang, Q.; Wang, W.; Li, G. Forecasting Summer Rainfall and Streamflow over the Yangtze River Valley Using Western Pacific Subtropical High Feature. *Water* **2021**, *13*, 2580. [CrossRef]
5. Gu, J.; Liu, S.; Zhou, Z.; Chalov, S.R.; Zhuang, Q. A Stacking Ensemble Learning Model for Monthly Rainfall Prediction in the Taihu Basin, China. *Water* **2022**, *14*, 492. [CrossRef]

6. Romero-Cuellar, J.; Gastulo-Tapia, C.J.; Hernández-López, M.R.; Prieto Sierra, C.; Francés, F. Towards an Extension of the Model Conditional Processor: Predictive Uncertainty Quantification of Monthly Streamflow via Gaussian Mixture Models and Clusters. *Water* **2022**, *14*, 1261. [CrossRef]
7. Shao, Y.; Zhao, J.; Xu, J.; Fu, A.; Wu, J. Revision of Frequency Estimates of Extreme Precipitation Based on the Annual Maximum Series in the Jiangsu Province in China. *Water* **2021**, *13*, 1832. [CrossRef]
8. Song, S.; Kang, Y.; Song, X.; Singh, V.P. MLE-Based Parameter Estimation for Four-Parameter Exponential Gamma Distribution and Asymptotic Variance of Its Quantiles. *Water* **2021**, *13*, 2092. [CrossRef]
9. Zeng, H.; Huang, J.; Li, Z.; Yu, W.; Zhou, H. Nonstationary Bayesian Modeling of Extreme Flood Risk and Return Period Affected by Climate Variables for Xiangjiang River Basin, in South-Central China. *Water* **2022**, *14*, 66. [CrossRef]
10. Li, Q.; Zeng, H.; Liu, P.; Li, Z.; Yu, W.; Zhou, H. Bivariate Nonstationary Extreme Flood Risk Estimation Using Mixture Distribution and Copula Function for the Longmen Reservoir, North China. *Water* **2022**, *14*, 604. [CrossRef]
11. Li, S.; Qin, Y. Frequency Analysis of the Nonstationary Annual Runoff Series Using the Mechanism-Based Reconstruction Method. *Water* **2022**, *14*, 76. [CrossRef]