Research and Application of Hydrological Time series Outlier Detection

Y.F Yu (1), Y.L Zhu (1), D.S Wan (2), and S.J Li (1)

(1) 1) College of Computer & Information, Hohai University, Nanjing 210098, China, (2) National Engineering Research Center of Water Resources Efficient Utilization Engineering Safety, Nanjing 210098, China

Abstract:Outliers are observations that appear to be inconsistent with the remainder of the data set and outliers refer to extreme high levels in stage analyses. The existence of outliers in hydrologic data sets affects the decision-making process related to design, operation, and management of water resources. Insufficient information on outliers limits our understanding and predictive ability of such extreme hydrologic phenomena. The identification of suspicious or outliers is to avoid data contamination, recognize mixed groups, and avoid distortion of the underlying probability distribution(s). An outlier can generally be categorized either as an error or a true observation. This research focuses on the problem of detecting extreme values or outliers from hydrological time series data. The main consideration is the fact that data closer in time are more correlated to each other than those farther apart. Combined with characteristics of hydrologic data, two statistic terms - mean change and mean rate of change - as well as the outliers in stage time series based on these terms are defined. And then, two variations of a method that uses 1) two-sided mean and 2) one-sided accumulative change from the neighborhoods of data point and a threshold value to compare the difference between the mean and the observed data value are proposed to detecting the outliers in stage time series. Experiments with the data collected from different hydrographic stations were performed, whose analysis results indicate the proposed methods are fast and can be used for stage time series analysis.

Keyword:Hydrological Time series; Outliers detection; Two-sided mean; One-sided accumulative change