



Automatic removal of outliers in hydrologic time series and quality control of rainfall data: processing a real-time database of the Local System for Flood Monitoring in Kłodzko County, Poland

Bartłomiej Mizinski (1), Tomasz Niedzielski (2,3), Maciej Kryza (2), and Mariusz Szymanowski (2)

(1) University of Wrocław, Mathematical Institute, Wrocław, Poland, (2) University of Wrocław, Institute of Geography and Regional Development, Wrocław, Poland (tomasz.niedzielski@uni.wroc.pl), (3) Space Research Centre, Polish Academy of Sciences, Warsaw, Poland

Real-time hydrological forecasting requires the highest quality of both hydrologic and meteorological data collected in a given river basin. Large outliers may lead to inaccurate predictions, with substantial departures between observations and prognoses considered even in short term. Although we need the correctness of both riverflow and rainfall data, they cannot be processed in the same way to produce a filtered output. Indeed, hydrologic time series at a given gauge can be interpolated in time domain after having detected suspicious values, however if no outlier has been detected at the upstream sites. In the case of rainfall data, interpolation is not suitable as we cannot verify the potential outliers at a given site against data from other sites especially in the complex terrain. This is due to the fact that very local convective events may occur, leading to large rainfall peaks at a limited space. Hence, instead of interpolating data, we rather perform a flagging procedure that only ranks outliers according to the likelihood of occurrence. Following the aforementioned assumptions, we have developed a few modules that serve a purpose of a fully automated correction of a database that is updated in real-time every 15 minutes, and the main objective of the work was to produce a high-quality database for a purpose of hydrologic rainfall-runoff modeling and ensemble prediction. The database in question is available courtesy of the County Office in Kłodzko (SW Poland), the institution which owns and maintains the Local System for Flood Monitoring in Kłodzko County. The dedicated prediction system, known as HydroProg, is now being built at the University of Wrocław (Poland). As the entire prediction system, the correction modules work automatically in real time and are developed in R language. They are plugged in to a larger IT infrastructure. Hydrologic time series, which are water levels recorded every 15 minutes at 22 gauges located in Kłodzko County, are tested for presence of outliers using the Rosner's test performed. Although the assumption of data normality may not always be justified, we use Rosner's test to just flag extreme data which then is validated with interrelated gauges in order to seek true outliers. Rainfall data are also observed with the same sampling interval at 18 sites and quality control criteria are applied. Two tests are used to verify the data quality: range test, based on regional extreme values, and persistence test. Rainfall data are never altered, but one of three flags are assigned: 0 – good measurement, 1 – suspicious (persistence test failed), 2 – failure (range test failed). The overall performance of the database correction approach is acceptable, and its most crucial feature is automation – a key element in real-time rapid prediction solutions.