Statistical quality control of daily precipitation data

G. Sciuto (1), B. Bonaccorso (2), A. Cancelliere (3), and G. Rossi (4)

(1) Geographical Institute, University of Bonn, Bonn, Germany (guidosciuto@geographie.uni-bonn.de), (2) Department of Civil and Environmental Engineering, University of Catania, Catania, Italy (bbonacco@dica.unict.it), (3) Department of Civil and Environmental Engineering, University of Catania, Catania, Italy (acance@dica.unict.it), (4) Department of Civil and Environmental Engineering, University of Catania, Catania, Italy (grossi@dica.unict.it)

Hydrometeorological data acquired by a monitoring network are potentially affected by different kind of errors that can compromise their applicability in describing hydrological phenomena. Clearly, a traditional control of the data, based on a manual inspection, is not feasible when the information about the observed phenomenon has to be promptly available, as is the case of real-time applications.

Therefore, automatic data quality control procedures are necessary in order to perform a fast preliminary screening of the acquired data and to identify possible anomalies which can be further analyzed by an operator, so that the time necessary for an in-depth inspection of the data is considerably reduced.

To this end, an automatic quality control procedure of daily precipitation, designed to automatically detect erroneous data to be submitted for further manual controls, is herein presented. Quality control of daily precipitation is based on confidence intervals derived by means of neural networks on the basis of contemporaneous data observed at reference stations, since the presence of zero values in the series and the strong variability of precipitation at daily time scale do not allow reliable confidence intervals to be estimated from historical data from the same station.

The proposed procedure has been applied to daily precipitation data observed from 1950 to 2004 in automatic stations in Sicily (Italy), belonging to the real time monitoring network of the Sicilian Regional Hydrographic Office. Such application enables validation of more than 80% of the data. Also the accuracy of the procedure in detecting erroneous data is verified by introducing known errors into the available datasets, supposed as correct, and by computing the probabilities of correctly classifying data as validated or not validated. Results indicate that the procedure enables to reduce substantially the number of data to be processed manually, while maintaining a fairly good accuracy in detecting correct and non-correct data.