Introduction and objectives

Hydrometeorological data acquired by a monitoring network are potentially affected by errors, originating from various causes, that can compromise their applicability in describing hydrological phenomena (Madsen, 1989). A traditional control of the data, based on a “manual” inspection, is not suitable for real-time applications, where the information on the observed phenomenon has to be promptly available.

Therefore, it appears necessary to develop methods for automatic quality control of the data that make it possible to perform a preliminary analysis of the acquired information aimed at identifying potential anomalies in the observations (Abbott, 1986; Reek et al., 1992).

In this paper, an automatic quality control procedure of daily rainfall data is presented (Sciuto et al., 2008), which extends a method previously developed for monthly rainfall data (Campisano et al., 2002). In particular, two types of controls are proposed. The first control is oriented to verify the null/not null nature of observed precipitation, without regard to its amount and therefore aims to detect mainly errors due to the total blockage of the rain gauge. Once a not null observation is labeled as correct, the second control aims at verifying its value through a comparison with confidence intervals of fixed probability. Both controls are based on neural networks, appropriately trained by making use of historical observations already subject to manual inspection for correctness.

Quality control procedure

Due to the intermittent nature and high variability of daily precipitation, confidence intervals cannot be estimated from historical data of the considered station (Sciuto et al., 2008). The proposed procedure consists of two different neural networks: the first neural network, defined binary, is used to estimate the presence or not of rainfall value in the series of the target station, on the basis of contemporary observations in reference stations. The second neural network is adopted to detect errors in the amount of rainfall, on the basis of confidence intervals, only when the first neural network has confirmed that for the examined day a not-null rainfall value is expected.

Selection of reference stations

Reference stations have been first selected among the neighboring ones within a maximum distance Δr=30 km and altitude difference ΔH=200 m. Then, a criteria based on the frequency of contemporary presence/absence of rain has been adopted, by computing the following conditional probability:

$$P_{ij} = \frac{N_{ij}}{N_i}$$

where $N_i$ and $N_{ij}$ are the number of days respectively with or without rainfall contemporary observed in two stations and Nt is the total number of observations.

The proposed methodology has been applied to daily rainfall observed from 1950 to 2004 in selected Sicilian stations, which belong to the real-time monitoring network of the Water Observatory of the Regional Agency for Waste and Water in Sicily (formerly, the Sicilian Regional Hydrographic Office).

Applications

In Figure 4 not null values observed at the station of Mirabella Iimbacciari, that have already passed the first quality control through the binary network, and relative confidence intervals (P=95%) based on the reference stations’ data computed by neural network, are shown.

In Table I, the percentage of the validated null and not null data are reported, with reference to a few considered stations, together with some statistics (R2 and RMSE) describing the goodness of the model estimates. From the table, it is possible to observe that such percentage is generally about 85%.

Conclusions

On the basis of the obtained results it can be concluded that, when the percentage of errors in the series increases, the probability that not validated data are not correct will increase too, while the probability that validated data are correct will decrease.

References

