



Semi-automatic error detection in hourly precipitation measurements using a Naive Bayes Classifier

G. Barentsen, D. Dehem and C. Tricot

Royal Meteorological Institute of Belgium (geert.barentsen@oma.be, didier.dehem@oma.be, christian.tricot@oma.be)

Hourly precipitation measurements from a network of 94 automatic rain gauges in the southern part of Belgium, operated by the Hydrological Service (SETHY) of the regional Ministry of Equipment and Transport (MET), are validated on a daily basis by the Royal Meteorological Institute of Belgium (RMI). The validation procedure corrects for random errors caused by rain gauge failures or obstructions. The errors are detected by comparing the rain gauge measurements with reference estimates resulting from spatial and temporal interpolation and from radar precipitation estimates. The detection is performed in a semi-automatic way using traditional visualisations and confidence interval tests.

Recently, a Naive Bayes Classifier (NBC) was introduced to assist in the detection of errors. A Naive Bayes Classifier is a simple, efficient and self-learning probability classification method, known for its surprising performance in other domains such as pattern recognition and e-mail spam filtering. The NBC estimates the likelihood of a measurement to be incorrect given the reference estimates, based on the history of manual corrections applied in the past. We find that the NBC detects 82% of the measurements with at least 1 mm error and 98% of the measurements with at least 2 mm error, while only 0.7% of the measurements that are believed to be correct are classified as error. The approach is found to be more flexible and powerful in detecting errors than manual rules or decision trees. In addition, the strategy may easily be applied to other meteorological parameters.