Evaluation of three Gap-Filling techniques for daily rainfall data sets: a case study in Portugal

Camilla Fagandini[1], Valeria Todaro[1], Maria Giovanna Tanda[1], Joao Lino Pereira[2], Leonardo Azevedo[2], and Andrea Zanini[1]

[1]Department of Engineering and Architecture, University of Parma, Parco Area delle Scienze 181/A, 43124, Parma, Italy (camilla.fagandini@unipr.it)

[2]CERENA/DECivil, Universidade de Lisboa, Instituto Superior Técnico, Av. Rovisco Pais ,628 1049-001 Lisbon, Portugal

In hydro meteorological temporal datasets, the lack of data is a common problem that can be caused by a variety of factors, including sensor malfunction, errors in measurement, and faults in data acquisition from the operators. Because complete time series are necessary for conducting trustworthy analysis, finding efficient solutions to this issue is crucial. In this work, a gap-filling approach using Kriging-based methods (Ordinary Kriging and Simple Cokriging) is presented and compared to a linear regression approach proposed by the Food and Agriculture Organization (FAO method). The proposed procedure consists of fitting semi-variogram models for each month using the available daily rainfall collected at all stations and averaged for the specific month in the reference period. The advantages are that only 12 monthly semi-variograms have to be built rather than one for each missing day of the dataset and that a greater amount of data at a time can be processed. Then, the Ordinary Kriging and Cokriging are used to estimate the daily precipitation where it is missed using the semi-variograms of the month of interest. The Cokriging method is applied considering the elevation data as the second variable. The FAO approach fills the gaps in rainfall time series by means of a linear relationship between the station that presents missing data and the best correlated station that has data gathered at the gap time. The approaches were compared using daily rainfall data from 60 rain gauges from the Portuguese case study of the InTheMED project for a 30-year reference period (1976-2005). To evaluate the effectiveness of the proposed approaches, one year of data (1985) was removed from some stations; missing precipitation data were estimated using data from the remaining precipitation stations by applying the three procedures. A cross-validation process and an analysis of the error statistics have been considered to determine the accuracy of the estimation for the three gap-filling methods. The outcomes pointed out that the geostatistical approaches outperformed the FAO method in daily estimation. The presented approach performed well in the study area, especially for the Ordinary Kriging, which well-estimated the daily missing data with a low computational effort. However, Cokriging did not significantly improve the estimates.