Multiple imputation of rainfall missing data in the Iberian Mediterranean context

Juan Javier Miró, Vicente Caselles, María José Estrela

PII: S0169-8095(17)30125-4
DOI: doi: 10.1016/j.atmosres.2017.07.016
Reference: ATMOS 4011

To appear in: *Atmospheric Research*

Received date: 1 February 2017
Revised date: 7 July 2017
Accepted date: 17 July 2017

Please cite this article as: Juan Javier Miró, Vicente Caselles, María José Estrela , Multiple imputation of rainfall missing data in the Iberian Mediterranean context, *Atmospheric Research* (2016), doi: 10.1016/j.atmosres.2017.07.016

MULTIPLE IMPUTATION OF RAINFALL MISSING DATA IN THE IBERIAN MEDITERRANEAN CONTEXT

Juan Javier Miró[1], Vicente Caselles[1], María José Estrela[2]

1 Departament de Física de la Terra i Termodinàmica, Facultat de Física, Universitat de València, Doctor Moliner 50, Burjassot, Valencia 46100, Spain

2 Departament de Geografia, Facultat de Geografia i Història, Universitat de València, Avda. Blasco Ibáñez 28, Valencia 46010, Spain

ABSTRACT

Given the increasing need for complete rainfall data networks, in recent years have been proposed diverse methods for filling gaps in observed precipitation series, progressively more advanced that traditional approaches to overcome the problem. The present study has consisted in validate 10 methods (6 linear, 2 non-linear and 2 hybrid) that allow multiple imputation, i.e., fill at the same time missing data of multiple incomplete series in a dense network of neighboring stations. These were applied for daily and monthly rainfall in two sectors in the *Júcar* River Basin Authority (east Iberian Peninsula), which is characterized by a high spatial irregularity and difficulty of rainfall estimation. A classification of precipitation according to their genetic origin was applied as pre-processing, and a Quantile-Mapping adjusting as post-processing technique. The results showed in general a better performance for the non-linear and hybrid methods, highlighting that the Non-linear PCA (NLPCA) method outperforms considerably the Self Organizing Maps (SOM) method within non-linear approaches. On linear methods, the Regularized Expectation Maximization method (RegEM) was the best, but far from NLPCA. Applying EOF filtering as post-processing of NLPCA (hybrid approach) yielded the best results.

Key words: Rainfall, missing data, gap filling, daily data, monthly data, imputation method comparison, dense network.

1. INTRODUCTION

The presence of gaps in the observed series of precipitation has been an old and well-known problem, but difficult to solve (Eischeid *et al.*, 2000; Auer *et* al, 2005; Ramos-Calzado *et al.*, 2008). This is particularly true under climates which, like the Mediterranean, have a high spatial irregularity and a marked concentration in daily rainfall extremes (Romero *et al*., 1998; Martín-Vide, 2004; Monjo and Martín Vide, 2016), which leads to greater uncertainties in any estimated data (Monjo *et al*., 2015). Therefore, in such cases the missing data imputation by the more traditional methods, tending to use simple weighting or naïve methods from nearby available stations, can lead to large errors (Ramos-Calzado *et al*., 2008; Kalteh and Hjorth, 2009). This is especially true if the availability of nearby stations is low in some time frame.

Likewise, the very heterogeneous temporal coverage of the observed series makes it difficult to create complete datasets that fulfill an adequate spatial detail and temporal coverage. These datasets are necessary to characterize in detail the climate trends and also are increasingly demanded by the society. In fact, the presence of multiple gaps and different coverage periods in the observed series generates uncertainty in the actual analysis of climate trends (Duffy *et al.*, 2001). This makes it necessary to apply a suitable fill of gaps before any trend analysis.