A rainfall spatial interpolation algorithm based on inhomogeneous kernels

Lorenzo Campo, Elisabetta Fiori, Luca Molini

CIMA Research Foundation, Savona, Italy (lorenzo.campo@cimafoundation.org)

FONDAZIONE CIMA
CIMA RESEARCH FOUNDATION
CENTRO INTERNAZIONALE IN MONITORAGGIO AMBIENTALE
INTERNATIONAL CENTRE ON ENVIRONMENTAL MONITORING

cc BY

Introduction

Rainfall fields constitute the main input of hydrological distributed models, both for long period water balance and for short period flood forecast and monitoring. The importance of an accurate reconstruction of the spatial pattern of rainfall is, thus, well recognized in several fields of application: agricultural planning, water balance at watershed scale, water management, flood monitoring. The latter case is particularly critical, due to the strong effect of the combination of the soil moisture pattern and of the rainfall pattern on the intensity peak of the flood. Despite the importance of the spatial characterization of the rainfall height, this variable still presents several difficulties when an interpolation is required. Rainfall fields present spatial and temporal alternance of large zero-values areas (no-rainfall) and complex pattern of non zero heights (rainfall events). Furthermore, the spatial patterns strongly depend on the type and the origin of rain event (convective, stratiform, orographic) and on the spatial scale. Different kind of rainfall measures and estimates (rainfall gauges, satellite estimates, meteo radar) are available, as well as large amount of literature for the spatial interpolation: from Thiessen polygons to Inverse Distance Weight (IDW) to different variants of kriging, neural network and other deterministic or geostatistic methods. In this work a kernel-based method for interpolation of point measures (raingauges) is proposed, in which spatially inhomogeneous kernel are used. For each gauge a particular kernel is fitted with the objective of minimize the error of interpolation measured on each available raingauge. In this way the local features of the field are taken in consideration. The kernel are assumed to be Gaussian with diagonal covariance matrices, and the parameter to be fitted is a multiplier of the matrix itself (assumed to be the identity in the general case). The method was applied on a set of 8 years of measurements (2006-2013) of raingauges in Northern Italy.

1. Interpolation method

The selected interpolation method makes use of gaussian kernels: the value of the rainfall interpolated in a point is given by the weighted average of the values observed at the same instant in the other "neighbour" raingauges, within a certain radius. The weight of each raingauge decreases with the distance from the contributing raingauge as a gaussian bivariate function with given covariance matrix. The matrix is assumed to be diagonal and equal to the identity multiplied by a scalar (no privileged directions).

$$P(x)=\frac{\sum_{i=1}^{N} K(d_i)P_i}{\sum_{i=1}^{N} K(d_i)}$$

Unique kernel for all raingauges

$$P(x)=\frac{\sum_{i=1}^{N} K_i(d_i)P_i}{\sum_{i=1}^{N} K(d_i)}$$

Different kernel for all raingauges

where:

$P(x)$ = rainfall in the point x
$K(d_i)$ = value of the kernel function computed at distance d_i (unique kernel for all raingauges)
d_i = distance of the point x from the raingauge i

P_i = rainfall observed in the raingauge i
$K_i(d_i)$ = value of the kernel function computed at distance d_i (each raingauge has a different kernel, i.e. a gaussian function with a different covariance matrix)

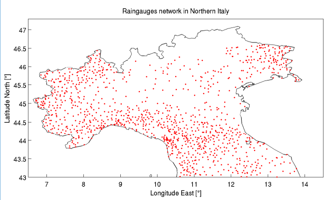

Figure 1. Raingauge network in North Italy.

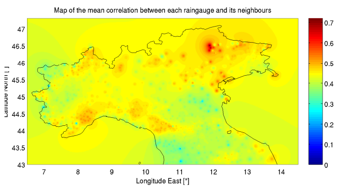

Figure 2. Map of the average correlation between each rainfall time series and those of the neighbours (other raingauges within 50 km of distance).

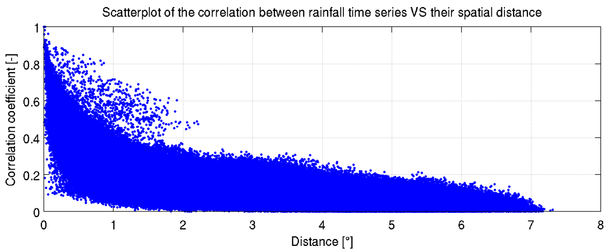

Figure 3. Scatterplot of the correlation between the rainfall time series observed VS their distance. It was assumed to consider as "neighbours" of each raingauges, the other sensors within a range of 50 km, corresponding to an average correlation of 0.6

2. Methodology

In order to fit the multiplier of the covariance matrix to assign to the kernel of each of the N raingauges, the following algorithm was used:

- For each raingauge i:
  - The neighbors (other raingauges within a given range) are selected
  - A minimization problem is solved by trying to reconstruct the values observed at the raingauge i interpolating only the values of the neighbors with the kernel method. It is assumed that the kernels are all equal, the cost function to be minimized is the RMSE between the interpolated values and the observed ones:

$$RMSE=\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(P_{INT}-P_{OBS}\right)^2}$$

In this way, an optimal values of the multiplier m of the covariance matrix is found

- At the end, for each raingauge there are several optimal values of the multiplier m_i (each raingauge is the neighbor of different raingauges and, in general, will have different values of the multiplier, see Figure **4**)
- The final value of the multiplier assigned to a given raingauge is obtained in two possible ways:
  - Case Single Kernel (SK) As the mean of all values (unique kernel for all the raingauges)
  - Case Multiple Kernels (MK) by a weighted average of the multipliers m_i, in which the weights are given by the percentage of improvement of the RMSE or reconstruction with respect to the case of unique kernel:

$$w_i=\frac{RMSE_{MK}-RMSE_{SK}}{RMSE_{MK}}$$

where:
$RMSE_{MK}$ = RMSE obtained with the fitted kernel
$RMSE_{SK}$ = RMSE obtained with the single kernel (unique for all raingauges)

In the Figures of Section 3, the two Cases are compared.

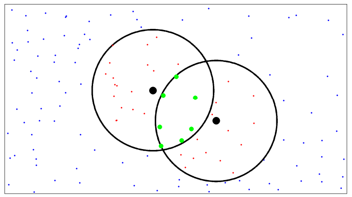

Figure 4. Given two raingauges (black large dots), the "neighborhood" of both is showed (black circles). The red points are raingauges the are assumed to be neighbours of the given sensors, the green points are neighbours of both.

3. Results

The fit of the kernel was carried on considering observed time series in Northern Italy (see Figure 1) data from 2008 to 2010, while the computation of the scores was done using time series from 2011 to 2012 as validation period.
Comparing Single Kernel and Multiple Kernel, a small improvement was achieved with the second method. The RMSE decreased of about 13% and the Bias was almost completely eliminated, while other scores like False Alarm (FA) and Probability of Detection (POD) resulted only slightly improved. In all cases, the variance of the scores computed on all the raingauges decreased. The following Figures show the comparisons.

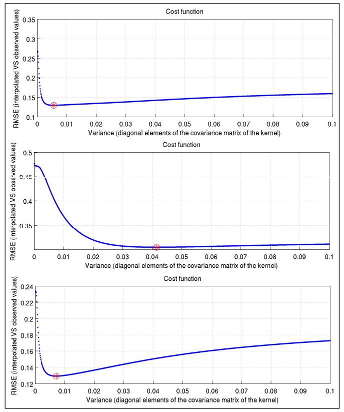

Figure 5. Examples of cost function: RMSE to be minimized with respect to the Variance (diagonal elements of the covariance matrix of the kernel).

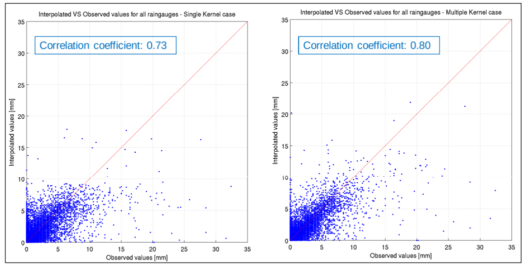

Figure 6. Scatterplot between observed and interpolated values for Single (left) and Multiple (right) Kernels.

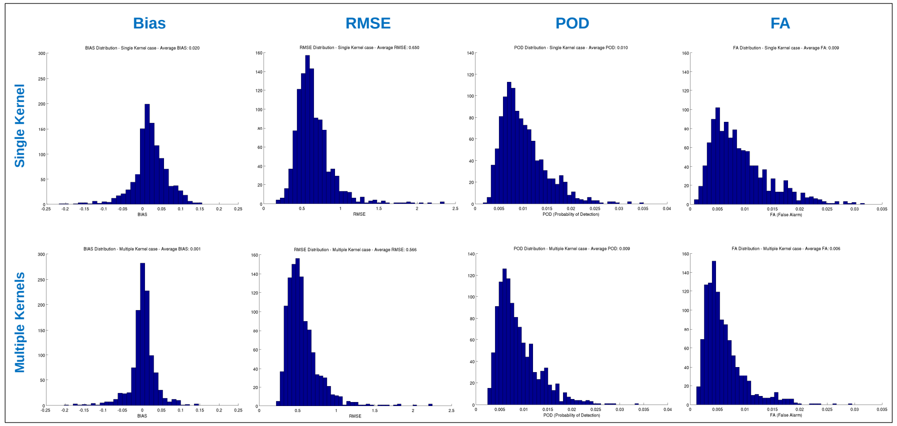

Figure 6. Distributions of different scores between the Single and Multiple Kernels cases. The compared scores are BIAS, RMSE (Root Mean Squared Error), POD (Probability of detection) and FA (False Alarm). All the scores are computed considering the ability of the interpolation to reconstruct the values observed in a raingauge excluded from the analysis. This procedure is repeated for all available sensors.