Statistical identification of orographic effects in the regional analysis of extreme rainfall

Pierluigi Furcolo, Anna Pelosi, Fabio Rossi

Dipartimento di Ingegneria Civile

Università degli Studi di Salerno

Via Giovanni Paolo II

84084 Fisciano (SA)

Italy

Submitted to: Hydrological Processes - July 2015

Revised: October 2015

Corresponding author:

Pierluigi Furcolo

Email: p.furcolo@unisa.it

This is the peer reviewed version of the following article Statistical identification of orographic effects in the regional analysis of extreme rainfall Hydrol. Process. 30, 1342–1353 (2016), ISSN:0885-6087***, which has been published in final form at http://dx.doi.org/10.1002/hyp.10719. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Use of Self-Archived Versions***

Abstract

Regional models of extreme rainfall must address the spatial variability induced by orographic obstacles. However, the proper detection of orographic effects often depends on the availability of a well-designed rain gauge network. The aim of this study is to investigate a new method for identifying and characterizing the effects of orography on the spatial structure of extreme rainfall at the regional scale, including where rainfall data are lacking or fail to describe rainfall features thoroughly.

We analyze the annual maxima of daily rainfall data in the Campania region, an orographically complex region in Southern Italy, and introduce a statistical procedure to identify spatial outliers in a low order statistic (namely the mean). The locations of these outliers are then compared to a pattern of orographic objects that has been *a priori* identified through the application of an automatic geomorphological procedure. The results show a direct and clear link between a particular set of orographic objects and a local increase in the spatial variability of extreme rainfall. This analysis allowed us to objectively identify areas where orography produces enhanced variability in extreme rainfall. It has direct implications for rain gauge network design criteria and has led to promising developments in the regional analysis of extreme rainfall.

Keywords: regional analysis; extreme rainfall; Kriging with uncertain data; orographic barriers.

1. Introduction

The regional analysis of extreme rainfall has to deal with the spatial variability of the rainfall process, which is driven by many physical factors, particularly those related to regional topography and its complex interactions with atmospheric circulation.

In the field of Earth and atmospheric sciences, physical modeling of atmospheric processes driven by orography at the event scale has long been a challenging problem (Smith, 1979; Rotunno and Ferretti, 2001; Roe, 2005 among the others). In recent years, major advances in meteorological modeling have been achieved, supported by the increasing availability of high-resolution data from remote sensors. These improvements have led to better real-time forecasting systems, especially through the introduction of very sophisticated, three-dimensional, non-hydrostatic models of atmospheric dynamics (Grell et al., 1995; Rotunno and Houze, 2007; Malguzzi et al., 2006).

Long-term probabilistic predictions of extreme rainfall at regional scale present a different problem. These predictions are primarily based on statistical analyses of extreme rainfall time series recorded by a network of rain gauges at the regional scale. In this case, one of the main issues is to identify and model the effects of orography on particular statistical parameters identified for rainfall extremes. Physically based models of single storms are not sufficient for this task: Statistical parameters associated with annual rainfall maxima are affected by a multitude of events that have different characteristics. As a result, the final impact on these parameters is difficult to predict. To better describe how orography influences these parameters, orographic data must be embedded into a regional model of extreme rainfall.

The typical regionalization approach, which is based on the application of regression relationships within homogeneous regions (Matalas and Gilroy, 1968), usually makes use of local descriptors of topography and tends to over-simplify the complex effects of orography on extreme rainfall (among others, Poreh and Mechrez, 1984; Corradini, 1985; Givone and Meignen, 1990; Weisse and Bois, 2001; Roe, 2005). As a result, regression errors are often spatially correlated, suggesting that a regression model alone cannot provide a satisfactory interpretation of the spatial variability of the field. A common solution to this problem is to split the region into smaller, more homogeneous sub-regions, or to couple a regression model with a spatial, non-parametric model of the residuals. The spatial model is usually based on the assumption of an intrinsically stationary random field: linear geostatistics provides BLUE (Best Linear Unbiased Estimator) estimators for this type of field. In addition, a geostatistical approach provides tools for evaluating the interpolation uncertainty related to network density and allows researchers to define the optimal criteria for network design (among others, Rodriguez-Iturbe and Meja, 1974; Bras and Rodriguez-Iturbe, 1976; Bacchi and Kottegoda, 1995; Bacchi, 1996; Pardo-Igúzquiza, 1998).

The approach of coupling a regression model with a spatial, non-parametric model of the residuals aims at describing the general dependence of rainfall on site-based topographic indices and capturing the complex interactions of rainfall with orography using spatial interpolation techniques. It can indeed give adequate results for technical purposes, but only if the spatial field is stationary and if a sufficiently dense rain gauge network is available. To improve the predictive ability of the regional models, increasingly sophisticated regression and spatial interpolation models have been introduced. The primary goal of these efforts is to add topographic covariates into the models, through the definition of features at the site level that are synthetic descriptors of topography at a broader scale (Slimani and Obled, 1986; Prudhomme and Reed, 1999; Weisse and Bois, 2001; Allamano et al., 2009). Thus, significant advances in regional modeling accuracy have been reached with a relatively moderate increase in the number of model parameters. Another way to improve regional models of extreme rainfall is to approach the description of a site's orography in new ways. For example, identification and characterization of the most relevant orographic features in a region can be used instead of relying on local or locally averaged descriptors of the topography obtained from the DEM (Digital Elevation Model). A recent paper from Cuomo et al. (2011) introduced a GIS-based, objective geomorphological procedure to identify and characterize orographic objects from a DEM of topography. The results of this geomorphological procedure can provide new and potentially valuable information relevant to several traditional problems of applied hydrology. This study will investigate their potential applications in the regional analysis of extreme rainfall.

In this paper, we consider a relatively wide and orographically complex region in Southern Italy, the Campania region. Then, in Section 3, we introduce a statistical procedure to identify "anomalous" spatial variability in the mean annual maxima of daily rainfall (here, "anomalous" refers to the null hypothesis of the weak stationary nature of the spatial field at the regional scale). The procedure addresses the non-Gaussian distribution of the original data and the presence of heteroscedastic site-based sampling errors (generated by the variable length of the time series). Anomalous values are strongly suspected to be caused by amplifications in intense rainfall driven by orography. By comparing the locations of anomalous gauges, as identified using this statistical procedure, with independently identified orographic objects, we observed a relationship between a particular set of orographic objects and the increased spatial variability of the rainfall field at small scale. This suggests that the selected topographic elements might effectively act as orographic barriers during rainfall events and may provide a key for interpreting observed amplifications in the spatial variability of extreme rainfall. This preliminary analysis provides some basic results that can be useful for improving the methods of rain gauge network analysis and design and building more accurate regional models of extreme rainfall.

In the next section, we present the study region, the hydrological dataset used for this study and a summary of the geomorphological identification of the orographic objects in the region. Section 3 contains a summary of the mathematical methods used in this paper, with a particular focus on the geostatistical techniques that are commonly used within regionalization studies. These methods provide the foundation for the statistical procedure presented in Section 4, along with the procedure's application to the case study, in which we identify anomalous spatial variability of the mean annual maxima of daily rainfall in the study region. In Section 5, we summarize and discuss the results by comparing the positions of the spatial outliers with the positions of the potential orographic barriers. Future work and potential applications of the results are discussed in the final section.

2. Study area, rainfall data set and orographic descriptors

The area selected for this study roughly corresponds to the Campania region in Southern Italy and includes some surrounding areas (Figure 1). Its complex orography interacts with wet air masses that come predominantly (but not exclusively) from the Tyrrhenian Sea and affects the spatial variability of many rainfall events. The current regional report on rainfall extremes in this area was prepared within the CNR's (Consiglio Nazionale delle Ricerche, National Research Council) VAPI (VAlutazione delle PIene, Flood

Estimation) project (Versace et al., 1989; Rossi and Villani, 1995; 1998), a national project designed to analyze the frequency of extreme rainfall and river floods at a regional scale. The regionalization of the annual daily rainfall maxima is based on a TCEV (Two Component Extreme Value) parent distribution (Rossi et al., 1984). For dimensionless high order parameters (coefficients of variation, skewness and kurtosis), the hypothesis that the entire region is statistically homogeneous is accepted. The average of the annual daily rainfall maxima was regionalized by identifying 6 sub-regions based on an estimated linear regression associated with elevation. The large number of sub-regions suggests that the spatial variability of the lowest order statistics may be strongly influenced by an orographic pattern.

This work is based on annual daily rainfall maxima because they represent the most reliable and largely available data in the region. A summary of the characteristics of the time series for the dataset is presented in Table 1. Given the results of previous regional studies, we focused on the mean of the annual maxima, which shows significant spatial variability when compared with its statistical variability. A map of Central-Southern Italy is presented in Figure 1, along with the locations of the 605 rain gauges used for the analysis: 245 of them are located inside the actual region of interest, as shown in the same figure. We included a comfortable buffer on all edges of the region of interest to reduce accidental border effects. We used a rainfall database collected within the VAPI project (http://www.gndci.cnr.it/it/vapi/welcome_it.htm). The database was updated after the publication of the regional report up to the year 2000.

The topography of the study area was modeled using a 20 m resolution DEM. Figure 1 illustrates the presence of a relevant orographic barrier, the Apennines along the inner section of the peninsula, together with many other mountains that run between the coastline and the Apennines. Cuomo et al. (2011) identify and classify mountains through an automatic and multi-scalar analysis of geomorphometric characteristics of the topography. The method is similar to well-known drainage network ordering systems and is based on the topographic concepts of key contour, key saddle, summit point and prominence (see Figure 2 for a summary of the method). The procedure is implemented in a GIS environment and is able to identify, delimit and order mountains and hills automatically. Simultaneously, the procedure finds the parent relationship between orographic entities and organizes the ordered mountains into an orographic hierarchy.

In this work, the union of orographic ranges (4^{th} order) and orographic groups (3^{rd} order) is chosen as the set of orographic objects that affects the spatial variability of the annual daily rainfall maxima (Figure 3). Here, this choice is driven by a comparison with the results of the presented statistical analysis of the spatial variability of annual daily rainfall maxima. However, this observed result is worth deeper investigation, since it is surely related to the interaction between the typical scale of the orographic objects in space and elevation and the scales of the meteorological events that most frequently produce the annual maxima.

3. Mathematical methods

In this section, we describe well-known mathematical methods relevant to this study and to the procedure proposed in Section 4.

3.1. Geostatistical methods for data affected by noise

An intrinsically stationary, isotropic random field $Z(\mathbf{x})$ has a unique spatial structure that can be expressed through the semi-variogram of the field (Cressie, 1993; Journel and Huijbregts, 1978):

$$\gamma(d)=\frac{1}{2}E\left[Z(\mathbf{x}+\mathbf{d})-Z(\mathbf{x})\right]^2;\quad \left(d=\|\mathbf{d}\|\right) \tag{1}$$

Given a set of measurements of the field $\left\{Z_i^*; i=1,..,n\right\}$ available at each site $\left\{\mathbf{x}_i; i=1,..,n\right\}$, the semi-variogram $\gamma(d)$ is estimated by fitting suitable parametric analytic functions to the experimental semi-variogram (Kitadinis, 1983; Todini and Pellegrini, 1999).

If the observed data Z_i^* are noise-free observations of the real values Z_i, the Ordinary Kriging Predictor (OKP) is the optimal linear estimator, which is the best possible estimator if the field is also Gaussian (Journel and Huijbregts, 1978). The Ordinary Kriging system provides a set of weights $\left\{w_i; i=1,\dots,n\right\}$ that define the best (minimum variance) unbiased linear estimator:

$$\hat{Z}_K(\mathbf{x})=\sum_{i=1}^{n} w_i \cdot Z_i^* \tag{2}$$

The variance of the estimator is also known:

$$\sigma_K^2(\mathbf{x})=2\cdot\sum_{j=1}^{n} w_j\cdot\gamma\left(\left\|\mathbf{x}-\mathbf{x}_j\right\|\right)-\sum_{k=1}^{n}\sum_{j=1}^{n} w_k w_j\cdot\gamma\left(\left\|\mathbf{x}_k-\mathbf{x}_j\right\|\right) \tag{3}$$

This feature makes the kriging estimator an appealing tool for network design (Rodriguez-Iturbe and Meja, 1974; Bras and Rodriguez-Iturbe, 1976; Bacchi, 1996). In fact, the variance in Equation (3) is solely a function of the geometry of the network and the semi-variogram of the field.

The presence of additional noise in the observed field (measurement errors, sampling errors, etc.) affects the optimal nature of the Ordinary Kriging Predictor, although its impacts on the predictor's unconditional unbiasedness are minimal (Borgault, 1994; Furcolo and Villani, 1998). When spatial interpolation of the field is the only desired result, Ordinary Kriging often represents a quick and reasonably accurate predictor. However, when the distribution and variance of predictor residuals is of interest, the noise in the data can dramatically affect the results and, consequently, the geostatistical technique must be modified.

In the fairly general case of a field contaminated with heteroscedastic uncorrelated noise, an optimal variant of the OKP, the Kriging with Uncertain Data (de Marsily, 1986; Slimani and Obled, 1986; Mazzetti and Todini, 2009), can be defined. Here, we use the KUD system as derived by de Marsily (1986). The application of this estimator requires knowledge of the semi-variogram of the true (noiseless) field and the noise variance at each sampling site. These requirements, which appear fairly strong, are not unrealistic in many applications, particularly where the noise variance is known *a priori* and the 'true' semi-variogram can, as a result, be estimated from the experimental variogram.

Onsite noise is defined as:

$$\delta_i = \delta(\mathbf{x}_i) = Z_i^* - Z_i \tag{4}$$

where $\{\delta_i; i = 1,..,n\}$ is a set of uncorrelated Gaussian noise with zero mean and variance $\sigma_i^2 = Var\left[\delta(\mathbf{x}_i)\right]$. Given the semi-variogram of the noiseless field $\gamma(d)$, the Kriging with Uncertain Data system varies slightly from the Ordinary Kriging Predictor and is expressed as follows (de Marsily, 1986, p.300):

$$\begin{cases} \sum_{j=1}^{n} w_j \gamma\left(\left\|\mathbf{x}_i - \mathbf{x}_j\right\|\right) - w_i \sigma_i^2 + m = \gamma\left(\left\|\mathbf{x}_i - \mathbf{x}\right\|\right) & i = 1,..,n \\ \sum_{j=1}^{n} w_j = 1 \end{cases} \tag{5}$$

Here, the unknowns are the n weights $\{w_i; i = 1,\ldots,n\}$ and m, which is the Lagrange multiplier for the unbiasedness constraint. The variance of the prediction error relative to the noiseless field $Z(\mathbf{x}) - \hat{Z}_K(\mathbf{x})$ becomes:

$$\sigma_K^2(\mathbf{x}) = 2 \cdot \sum_{j=1}^{n} w_j \cdot \gamma\left(\left\|\mathbf{x} - \mathbf{x}_i\right\|\right) - \sum_{k=1}^{n} \sum_{j=1}^{n} w_k w_j \cdot \gamma\left(\left\|\mathbf{x}_i - \mathbf{x}_j\right\|\right) + \sum_{j=1}^{n} w_j{}^2 \cdot \sigma_i^2 \tag{6}$$

Differently from the OKP, the exactness property does not hold for the KUD: the predictor acts as a filter and does not produce the exact measured values at gauged sites. As a result, the predictor variance at gauged sites is greater than zero. It is worth remarking that the prediction error is related to the noiseless field. This implies that the residuals at gauged sites $Z_i^* - \hat{Z}_K(\mathbf{x}_i)$ have a larger variance (σ_K^{2*} in the next equation) than the prediction error itself:

$$\sigma_K^{2*}(\mathbf{x}_i) = Var\left(Z_i^* - \hat{Z}_K(\mathbf{x}_i)\right) = \sigma_K^2(\mathbf{x}_i) + \sigma_i^2(1 - 2w_i) \tag{7}$$

When the spatial interpolation addresses a directly measurable physical quantity, onsite noise is representative of the measurement errors and it is usually neglected without relevant consequences for the result. However, if the field of a statistical parameter is analyzed, the onsite noise is produced by sampling errors. The variance of sampling errors depends on the length of the series, which in turn varies from site to site, producing heteroscedasticity. Inter-station correlations, caused by rainfall events that produce an annual maximum simultaneously at different gauges, can further complicate the picture, because it may introduce a correlation structure in the noise field as well.

In this study, the presence of sampling errors in the data is explicitly considered using the KUD estimator. We also analyzed all the couples of annual maxima time series and calculated the amount of correlation that is due to the same rainfall events (by checking the dates of the annual maxima). In this way we were able to assess the effect of inter-station correlation, and it was found to be negligible. Therefore, it is assumed that the sampling errors are heteroscedastic and uncorrelated.

3.2. Cross-validation and distribution of standardized residuals

Cross-validation is a method generally used to assess the performance of kriging estimators. It is based on a jackknife kriging estimation of values at sampling sites using only the values available from the other (n-1) sites:

$$\hat{Z}_{JN}(\mathbf{x}_i) = \hat{Z}_K(\mathbf{x}_i) \Big| \left\{ Z^*_1, \ldots, Z^*_{i-1}, Z^*_{i+1}, \ldots, Z^*_n \right\} \qquad i = 1, \ldots, n \tag{8}$$

For the OKP, the differences between the sampled (noise-free) and estimated values represent the estimation errors. These random errors are expected to have a Gaussian distribution, with zero mean and a variance that corresponds to the estimated predictor variance σ^2_{JN}. They can be then standardized as follows:

$$e_i = \frac{Z^*_i - \hat{Z}_{JN}(\mathbf{x}_i)}{\sigma_{JN}(\mathbf{x}_i)} = \frac{Z(\mathbf{x}_i) - \hat{Z}_{JN}(\mathbf{x}_i)}{\sigma_{JN}(\mathbf{x}_i)} \tag{9}$$

Based on the hypotheses of the kriging predictor, the standardized errors should follow a standard Gaussian distribution. They may still be affected by a small amount of spatial correlation; however, this can be eliminated by using ortho-normal residuals (Kitadinis, 1993), though these are less objectively defined. It is common for standardized errors to show little or no correlation structure, as is the case in this study. Given this, the present study uses simple and objectively defined standardized errors instead of ortho-normal errors.

When onsite sampling variance is introduced and kriging is used as a filter (KUD), the jackknife residuals are the sum of the kriging errors and the data errors. Yet, if the data errors are as defined in Equation (4) and the basic kriging hypotheses still hold, these residuals can be standardized using the variance in Equation (7), with $w_i = 0$ (because it is a jackknife estimation):

$$e_i^* = \frac{Z^*_i - \hat{Z}_{JN}(\mathbf{x}_i)}{\sigma_{JN}^*(\mathbf{x}_i)} \tag{10}$$

The standardized residuals $\{e_i^*\}$ are expected to be distributed according to a standardized normal distribution as well. The observed frequency distribution of the standardized residuals is estimated through the Weibull plotting position $P_k = k/(n+1)$ and compared with the expected normal distribution. Here, k is the position of a datum in the ordered sample $\{\varepsilon_k = e^*_{\pi_n(k)}; \varepsilon_k \geq \varepsilon_{k-1}\}$ as defined through a permutation $\pi_n(k)$ of the indices in the interval $\{1,\ldots,n\}$. For a given datum in the sample, the Weibull plotting position actually represents the expected value of the non-exceedance probability, whose distribution is in turn known to be a beta distribution with parameters $a = k$ and $b = n - k + 1$:

$$f_{P_k}(p) = \frac{p^a \cdot (1-p)^b}{\mathrm{B}(a,b)} = \frac{\Gamma(a+b) \cdot p^a \cdot (1-p)^b}{\Gamma(a) \cdot \Gamma(b)} \tag{11}$$

When plotted on a Normal Probability Chart, the empirical points represented by the pairs (ε_k, P_k) are supposed to organize with a good approximation along a straight line corresponding to a Gaussian CDF (Cumulative Distribution Function). It is possible to test the actual agreement between the observations and the theoretical model by defining confidence intervals around the plotting positions, according to the beta distribution in Equation (11).

This approach provides the basis for the identification of spatially anomalous values that will be carried out in Section 4.

3.3. Preliminary analysis and transformation of the field

To apply linear geostatistics under the most favorable hypotheses, it is convenient to obtain a marginal distribution of the field that is as Gaussian as possible. This property can be achieved through a suitable transformation of the original variable. Typically, for unimodal, positively skewed random variables the Box and Cox (1964) transformation represents a satisfactory solution. By means of a shape parameter λ the original field $Y(\mathbf{x})$ is transformed into a new, approximately Gaussian field:

$$Z(\mathbf{x}) = \begin{cases} \dfrac{Y(\mathbf{x})^{\lambda} - 1}{\lambda} & 0 < \lambda < 1 \\ \\ \ln Y(\mathbf{x}) & \lambda = 0 \end{cases} \tag{12}$$

For a given set of data, an estimation of λ can be obtained by finding the value that makes the sample skewness the closest possible to zero.

In addition to its (approximate) Gaussian property, the field $Z(\mathbf{x})$ obtained through the Box-Cox transformation also shows a more stable variance. If $Y(\mathbf{x})$ is the field of a statistical parameter, it is then possible to attribute most of the variability of the variance of the sampling errors affecting Z_i^*, or the onsite estimations of $Z(\mathbf{x}_i)$, to the variable size of the recorded time series.

4. Identification of spatial outliers in extreme rainfall data and their interpretation related to orographic effects

Using the rainfall data for the study area described in Section 2, we now present an iterative statistical procedure to identify anomalous variability in the spatial field of the mean annual daily rainfall maxima. This procedure is based on the use of linear geostatistics. As such, it is quite general and can be easily exported to other regions. The comparison of statistically identified spatial outliers to objectively defined orographic barriers identified *a priori* of this analysis will provide new information and perhaps better insight into the variability of spatial characteristics associated with extreme rainfall.

4.1. Preliminary data analysis and transformation

A preliminary screening of the data highlights the presence of a positive skewness in the original annual maxima series. For short samples, this skewness would affect the Gaussian properties of the sample average. The Box-Cox transformation (Equation (12)) was applied to address this issue. Looking at the graph in Figure 4, the natural logarithmic transformation is chosen ($\lambda = 0$). A simple first order approximation of the mean of the transformed variable is the logarithm of the mean. This transformation was then used in the next geostatistical analysis.

At a regional scale, the transformed data show a negligible dependence on elevation (on average, 2% increment every 1000 m of elevation, Figure 5). This may not be true at a local scale; however, this preliminary step is only intended to assess if any global trends can be distinguished in order to de-trend the data and work on the residuals. In this case, there was no need for any further pre-treatment of the data.

The experimental semi-variogram of the log-transformed data shows a well-defined structure, to which an exponential shape can be fitted (Figure 6a):

$$\gamma_Z\left(d\right) = n + s \cdot \left[1 - \exp\left(-\frac{3d}{r}\right)\right] \tag{13}$$

where parameters have the following values:

- nugget: n = 0.0160;
- partial sill: s = 0.0539;
- range: r = 122 km.

The presence of a nugget in the experimental semi-variogram likely relates to a micro-scale structure in the field or to the presence of noise in the data. In this case study, data are known to be affected by a statistical error related to the estimation of the mean from a finite sample. This error has a variance that depends on the length of each time series and is therefore variable among sites. A first order approximated estimation (small $\overline{c}_v$) of the sample variance $\sigma_Z^2\left(x_i\right)$ of Z(x) at site *i* can be found using the following equation:

$$\hat{\sigma}_Z^2(x_i) = \frac{\overline{c}_v^{\,2}}{N_i} \quad i = 1,\ldots,n \tag{14}$$

where $\overline{c}_v$ is the regional average of the coefficient of variation of the original series of the annual maxima and N_i is the length of the series at site *i*.Given the relatively large number of gauged sites, each point of the experimental variogram is the average of a very large number of pairs from the raw variogram. The effect of the noise on the experimental semi-variogram, then, is a general increment as large as the average variance of the errors. The quantitative assessment of the variance of the sampling errors allows us to separate the nugget component produced by errors in the data from the remaining part caused by an unknown micro-scale effect (see Table 2). A possible explanation of this ‘apparent’ micro-scale structure will be discussed later. The semi-variogram of the noiseless field is then obtained by subtracting the mean error variance from the entire experimental variogram.

4.2. Iterative statistical procedure for the identification of spatial outliers

The presence of spatial outliers produced by the existence of an enhanced, small scale variability in the field actually contaminates the data-set and affects the estimation of the variogram. The identification of these anomalous values is based on the analysis of the residuals of a cross-validation procedure (Section 3.2). The KUD estimator was run for the jackknife estimation of the field at all sampling points and the standardized cross-validation residuals e_i^* defined in Equation (10) were calculated. These residuals do not show any correlation structure, although the observed variance is slightly smaller than 1 (Figure 6b).

The introduction of sampling variance, which differs significantly from site to site, is mandatory in an application that tries to identify spatial outliers from standardized residuals. In this application, time series have a length that varies from 5 to 77 years; by using the Ordinary Kriging Predictor, the shorter series would automatically be more likely to produce anomalous errors.

The Normal Probability Chart in Figure 6c shows the standardized residuals obtained in the study region, along with their 95% confidence intervals and the line corresponding to the normal CDF, fitted through the 20th and the 80th percentiles. The overall shape of the empirical distribution shows little agreement with the normal straight line, especially at the upper tail, where an evident departure from the expected theoretical distribution is observed. The actual shape of the upper tail shows larger under-estimation errors than expected; those errors are then related to exceptionally high values of the data. These data are considered spatial outliers when the expected probability of the standardized residuals, based on the normal distribution model, exceeds the limits of the confidence intervals in the plot. A possible explanation for the presence of spatial outliers in the data is the effect of rainfall amplification caused by orography.

As discussed previously, the presence of outliers in the data can affect the experimental semi-variogram and, consequently, the jackknife estimates. For this reason, the procedure must be iterated according to the following scheme:

1) a new sample without the anomalous values is obtained;
2) the new sample is used to estimate the experimental semi-variogram and to fit a theoretical model;
3) a new set of jackknife estimates is calculated by applying KUD to the whole dataset (previous outliers included);
4) a new set of anomalous values is identified;
5) if the outliers are the same as in the previous iteration, the procedure ends; otherwise, a new iteration is performed.

In this case study, three iterations were needed before the procedure was terminated. The final semi-variograms of the data (without outliers) and the standardized residuals, along with the Normal Probability Chart showing the set of spatial outliers and the final standardized residuals, are presented in Figure 7.

It is interesting to note how the nugget in the experimental semi-variogram in the last iteration is almost completely attributable to sampling errors. After subtracting the mean error variance, the nugget of the noiseless field is much lower than in the first iteration and very close to zero. This suggests that the hypothetical micro-scale structure was indeed an apparent one; it was likely one of the effects of the anomalous data contained in the dataset on the semi-variogram. Once those data were removed, the experimental semi-variogram corresponded to an almost perfectly continuous noiseless field.

4.3. Validation of the iterative procedure

To validate the accurate performance of the proposed iterative procedure, a reduced dataset without spatial outliers was tested using the same procedure. This verified whether the Gaussian properties of the standardized residuals remained. Both the variogram and the jackknife procedure were run on a set of data from which the spatial outliers were removed to check if the standardized residuals at the remaining sites were close to a standard normal distribution. Despite the appearance, the results of this validation are not obvious at all. In the previous section, it was implicitly assumed that the departure from the theoretical normal distribution of the standardized residuals was only caused by the presence of a few anomalous values in the data. Here, the aim is to check:

- if this assumption can be considered true; or if, aside from the spatial outliers, there is a more general departure from the hypotheses that support the application of linear geostatistics in the region; and
- if the iterative procedure is indeed able to find all spatial outliers in the data.

The experimental and fitted semi-variograms are the same as in the last iteration of the identification procedure. The residuals of the jackknife procedure were then calculated using non-outlying data only. The Normal Probability Chart shows a very good alignment of the standardized residuals along a straight line (Figure 8). In general, if after a few iterations validation has not occurred, a more comprehensive preliminary analysis of the homogeneity of the study region would be required.

On the one hand, this simple validation confirms the stationary and linear nature of the field when anomalous gauges are removed. On the other hand, it underlines the importance of a properly designed monitoring network; if outliers had not been observed it would not have been possible to detect the phenomenon.

5. Summary of the results and discussion

The iterative procedure identified 19 anomalous gauges within the inner study region (the region of interest, where border effects should have less of an effect). These gauges are a small fraction of the total number of gauges (19 against 245); however, their effect on the geostatistical characterization of the field is anything but negligible (for example, the nugget in the experimental semi-variogram).

The basic statistics (Table 3) reveal that the average value of the field is much higher in the 19 anomalous gauges than in the other gauges of the region.

The presence of these spatial outliers should be investigated to better understand the cause of such high variability in the field recorded around those sites. More specifically, it would be interesting to assess whether the mean is affected by the presence of a few very intense events in the series or by a more general amplification of the events producing the annual maximum. The presence of outliers in the time series is well detected through the calculation of higher uneven order statistics (the skewness coefficient is used here). Skewness coefficients are reported in Table 3 for both anomalous and non-anomalous data. It is possible to see how similar the values are. The relatively higher value of the coefficient of variation for the anomalous stations is of particular interest. This value may just be the result of statistical errors in the estimation (for the anomalous subset, the coefficient is averaged over a small number of stations); nevertheless, if this is true, it may also result from an amplification effect that varies from event to event. This would increase the dispersion of the distribution of the annual maxima. Regardless, the increased dispersion affects both the lower and the upper tail of the distribution almost equally, as the skewness coefficient suggests.

Another, perhaps more important question is whether it is possible to identify the physical factors that influence the rainfall field in such a way as to produce spatial outliers that are clearly evident in the mean annual maxima. In a climatologically homogeneous region like the one analyzed, orography is the geographic characteristic that affects the spatial variability of rainfall the most. This is confirmed by a preliminary examination of the results, as shown in Figure 9. When the anomalous gauges within the boundary of the inner study region (the region of interest) are represented on the orographic map, those stations are located very near orographic ridges. The problem is how this information can be formalized in an objective, reproducible and effective (for modeling) way. The results from Cuomo et al. (2011) were used to describe the orography of the region analyzed in this study.

When the spatial outliers are compared with the orographic objects of the 3rd and 4th orographic orders (ranges and groups), we find that most spatial outliers lie inside the perimeter of these objects (Figures 3 and 9).

A summary of rain gauge network characteristics is given in Table 4. Together, this table and Figure 9 reveal a lower network density exactly where it is more necessary: in areas with orographic features. This is a consequence of the historical inaccessibility of these places due to morphology, vegetation and a lack of infrastructure. The result is that the enhanced spatial variability of rainfall caused by orography not only is not properly observed, but it is even less observed than what a network with homogeneous average density would allow. Nevertheless, a consistent number of anomalous gauges have been found and the phenomenon has at least been detected.

The relative frequency of spatial anomalies within orographic boundaries is compared to the frequency of spatial anomalies outside orographic boundaries in Table 5. A direct, significant relationship between the geomorphologically identified boundaries and the presence of high spatial variability in the rainfall field is indicated. All of the statistics confirm this observation. Nearly 70% of the anomalous gauges are within the orographic boundaries, whereas only 20% of all the network gauges are in those areas. The frequency of occurrence of the outliers within the orographic boundaries is nearly 30%, while it is as little as 3% beyond those boundaries.

6. Conclusions

We analyzed the spatial variability of a low order statistic for series of annual daily rainfall maxima through an iterative statistical procedure that detected the presence of enhanced small scale variability in some areas of the region. By comparing the results of the statistical procedure to orographic objects independently identified using the geomorphological approach developed by Cuomo et al. (2011), we identified a convergence of orographic barriers that affect the spatial variability of extreme rainfall. These results open a new avenue of research that includes orographic effects in regional models of rainfall extremes, especially in areas where the morphology of orographic barriers is particularly complex. For example, in the case study addressed here, the characteristics of the objects represented by orographic ranges (4^{th} order) and groups (3^{rd} order) provide a foundation for the statistical modeling of the drift produced by orographic barriers and could lead to more accurate regional models for extreme rainfall. Moreover, the results of this work underline the need for a rain gauge network that can capture the higher spatial variability of rainfall induced by orographic obstacles. Our preliminary results can provide the basis for defining better criteria for network designs and updates. A proper model of orographic effects on the spatial structure of a rainfall field would also better define the rules governing network updates; at the same time, a more accurate spatial sample of the rainfall field would allow improved detection of orographic effects and improve their modeling.

Both of these applications are under development for the region presented in this work and show promising preliminary results.

Acknowledgements

This work was developed within a research project funded by FARB of the University of Salerno and the PhD program in Civil and Environmental Engineering at the University of Salerno.

References

Allamano, P., Claps, P., Laio, F. and Thea, C., 2009. A data-based assessment of the dependence of short duration precipitation on elevation, Phys. Chem. Earth, 34, 635-641, doi:10.1016/j.pce.2009.01.001.

Alpert, P. and Shafir, H., 1989. A physical model to complement rainfall normals over complex terrain, J. Hydrol., 110, 51-62.

Bacchi B. and Kottegoda, N.T., 1995. Identification and calibration of spatial correlation patterns of rainfall, J. of Hydrol., 165, 311-348.

Bacchi B. (Ed), 1996. Metodi di ristrutturazione e progetto delle reti pluviometriche, CNR-GNDCI pub. n. 1620 (U.O. 1.33), Salerno, Italia, 98 pp.

Borgault, G., 1994. Robustness of noise filtering by kriging analysis, Math. Geol., 26, 733-752.

Box, G. E. P. and Cox, D. R., 1964. An analysis of transformations, J. R. Stat. Soc. Ser. B, 26, 211-252.

Bras, R. L., and Rodríguez-Iturbe, I., 1976. Network design for the estimation of areal mean of rainfall events, Water Resour. Res., 12, 1185–1195, doi:10.1029/WR012i006p01185.

Corradini, C., 1985. Analysis of the effects of orography on surface rainfall by a parameterized numerical model, J. Hydrol., 77, 19-30.

Cressie, N.A., 1993. Statistics for spatial data, John Wiley and Sons, NY, 928 pp.

Cuomo, A., Guida, D. and Palmieri, V., 2011. Digital orographic map of peninsular and insular Italy, J. Maps, 2011, 447-463. 10.4113/jom.2011.1209.

de Marsily, G., 1986. Quantitative hydrogeology: groundwater hydrology for engineers, Academic Press, Orlando, FL, 440 pp.

Furcolo, P. and Villani, P., 1998. Spatial analysis of uncertain data: an application to extreme rainfall, Int. J. Systems Res. Info. Sci., Special Issue on A. I. Techniques, 8, 151-181.

Givone, C. and Meignen, X., 1990. Influence of topography on spatial distribution of rain, Hydrology of Mountainous Areas, IAHS Publ. 190, 57-65.

Grell, G.A., Dudhia, J. and, Stauffer, D.R., 1995. A description of the fifth-generation Penn State NCAR Mesoscale Model (MM5), NCAR Tech. Note TN-398 + STR. 122 pp.

Journel, A.G. and Huijbregts, C.J., 1978. Mining Geostatistics, Academic Press, London, 600 pp.

Kitadinis, P.K., 1993. Geostatistics, Maidment, D.R. (Eds.), Handbook of hydrology.

Kitadinis, P.K., 1983. Statistical estimation of polynomial generalised covariance functions and hydrologic applications, Water Resour. Res., 19, 909-921.

Malguzzi, P., Grossi, G., Buzzi, A., Ranzi, R. and Buizza, R., 2006. The 1966 'century' flood in Italy: a meteorological and hydrological revisitation, J. Geophys. Res., 111, D24106, doi:10.1029/2006JD007111.

Matalas, N.C. and Gilroy, E.J., 1968. Some comments on regionalization in hydrologic studies, Water Resour. Res., 4 (6), 1361-1369.

Mazzetti C. and E. Todini, 2009. Combining Weather Radar and Raingauge Data for Hydrologic Applications, in: Flood Risk Management: Research and Practice – Samuels et al. (eds), Taylor & Francis Group, London, ISBN 978-0-415-48507-4

Pardo-Igúzquiza, E., 1998, Optimal selection of number and location of rainfall gauges for areal rainfall estimation using geostatistics and simulated annealing, J. Hydrol., 210, 206-220.

Poreh, M. and Mechrez, E., 1984. The combined effect of wind and topography on rainfall distribution, J. Hydrol., 72, 1-23.

Prudhomme, C. and Reed, D., 1999. Mapping extreme rainfall in a mountainous region using geostatistica ltechniques: a case study in Scotland, Int. J. Climatol., 19, 1337-1356.

Rodríguez-Iturbe, I., and Mejía, J. M., 1974. The design of rainfall networks in time and space, Water Resour. Res., 10, 713–728, doi:10.1029/WR010i004p00713.

Roe, G.H., 2005. Orographic precipitation, Annu. Rev. Earth Planet. Sci., 33:645–71, doi: 10.1146/annurev.earth.33.092203.122541.

Rossi, F., Fiorentino, M., and Versace, P., 1984. Two Component Extreme Value Distribution for Flood Frequency Analysis, Water Resour. Res., 20, 847-856.

Rossi, F. and Villani, P., 1995. Valutazione delle piene in Campania, CNR-GNDCI Report, 310 pp.

Rossi, F. and Villani, P., 1998. A project for regional analysis of floods in Italy, in Andah, K. (ed), Managing hydro-geological disasters in a vulnerable environment, CNR pub. No.1900, 141-168.

Rotunno, R. and Ferretti, R., 2001. Mechanisms of intense Alpine rainfall, J. Atmosph. Sc., 58, 1732-1749.

Rotunno, R. and Houze, R.A., 2007. Lessons on orographic precipitation from the Mesoscale Alpine Programme, Q. J. R. Meteorol. Soc., 133, 811-830.

Slimani, M. and Obled, Ch., 1986, Regionalization of extreme rainfall parameters through Kriging and correlation with topography descriptors, Int. Symp. on Flood Frequency and Risk, Baton Rouge (Louisiana).

Smith, R.B., 1979. The influence of mountains on the atmosphere, Adv. Geoph., 21, 87-230.

Todini, E. and Pellegrini, F., 1999. A maximum likelihood estimator for semi-variogram parameters in Kriging, in: GeoENV II – Geostatistics for environmental applications, Gòmez-Hernàndez, J., Soares, A. and R. Froidevaux (Eds.), Kluwer Academic Publisher, The Netherlands, 187-198.

Versace, P., Ferrari, E., Gabriele, S. and Rossi, F., 1989. Valutazione delle piene in Calabria, CNR-IRPI Report, Geodata, Cosenza.

Weisse, A.K. and Bois, P., 2001.Topographic effects on statistical characteristics of heavy rainfall and mapping in the French Alps, J. Appl. Meteor., 40, 720–740.