Regional-scale analysis of extreme precipitation from short and fragmented records

Regional-scale analysis of extreme precipitation from short and fragmented records

Andrea Libertino[a,*], Paola Allamano[b], Francesco Laio[a], Pierluigi Claps[a]

[a] *Dipartimento di Ingegneria dell'Ambiente, del Territorio e delle Infrastrutture, Politecnico di Torino, Corso Duca degli Abruzzi, 24, 10129, Torino, Italy*
[b] *Waterview SRL, Corso Castelfidardo, 30/A, 10129, Torino, Italy*

Abstract

Rain gauge is the oldest and most accurate instrument for rainfall measurement, able to provide long series of reliable data. However, rain gauge records are often plagued by gaps, spatio-temporal discontinuities and inhomogeneities that could affect their suitability for a statistical assessment of the characteristics of extreme rainfall. Furthermore, the need to discard the shorter series for obtaining robust estimates leads to ignore a significant amount of information which can be essential, especially when large return periods estimates are sought. This work describes a robust statistical framework for dealing with uneven and fragmented rainfall records on a regional spatial domain. The proposed technique, named "patched kriging" allows one to exploit all the information available from the recorded series, independently of their length, to provide extreme rainfall estimates in ungauged areas. The methodology involves the sequential application of the ordinary kriging equations, producing a homogeneous dataset of synthetic series with uniform lengths. In this way, the errors inherent to any regional statistical estimation can be easily represented in the spatial domain and, possibly, corrected. Furthermore, the homogeneity of the obtained series, provides robustness toward local artefacts during the parameter-estimation phase. The application to a case study in the north-western Italy demonstrates the potential of the methodology and provides a significant base for discussing its advantages over previous techniques.

*Corresponding author at: DIATI, Politecnico di Torino, Corso Duca degli Abruzzi, 24, Torino, Italy. Tel.:+390110905655

Email address: andrea.libertino@polito.it (Andrea Libertino)

Keywords: regional analysis, annual maxima, fragmented records, rainfall extremes, L-moments
PACS: 92.40.Ea
2010 MSC: 86A05

1. Introduction

Probabilistic modelling of extreme rainfall has a crucial role in flood risk estimation and consequently in the design and management of flood protection projects [1]. The first attempts to establish a mathematical relation between intensity and frequency of rainfall goes back to as early as 1932 [2]. Since then, many studies (e.g., [3]) have been carried out, aimed at providing the rainfall depths for different return periods and durations. Complete overviews on the different approaches adopted from several countries around the globe can be found, e.g, in [4, 5].

Intensity-Duration-Frequency (IDF) and Depth-Duration-Frequency (DDF) curves are commonly adopted in water resources engineering for both planning, designing and operating of water resource projects and for land and people protection purposes [6]. These curves are usually developed considering the historical records for different durations and adopting the index-rainfall method, in which the quantile of the extreme rainfall comes as the product of an "index value" (i.e., usually the mean) and a growth curve (i.e., the non-dimensional inverse of the frequency distribution $F(x)$).

Two approaches are commonly adopted for fitting a probability distribution to the series of maxima: (i) the "block" method, that consists in selecting the maximum rainfall occurring over a fixed period (usually 1 year) and (ii) the "peak-over-threshold" method, in which all the rainfall data exceeding some pre-specified threshold are considered [7]. The method (i) is widely adopted in Italy for design rainfall estimation, and a large dataset of annual maxima for duration 1-3-6-12-24 hours is available, which dates back to the early twentieth century.

Due to the significant developments of the theory of extreme value in the last two decades [7, 8] the methodologies for rainfall frequency analysis are nowadays quite established and robust, both at the single-station and at the regional scale. However, the correct reproduction of complex hydrometeorological processes requires not only long, but also serially complete and reliable observations [9, 10] from a dense and spatially uniform monitoring network. A non-uniform and non-continuous dataset can prevent a reliable

application of the aforementioned methodologies at the regional scale leading to inconsistencies.

It is thus evident that, despite the existence of established rainfall frequency analysis techniques, operational and methodological problems concerning their applications still arise.

Rainfall time series are often plagued with missing values creating sporadic and/or continuous gaps in their records. The fragmented behaviour traces back to the activation and dismissal of rain gauges, attributable to station relocation, service interruptions, replacement/renewal of the sensor, changes in the ownership of the station, etc. The characteristics of the stations (location and elevation, type of sensor, etc.) may also change before and after the interruptions, with consequent problems in attributing the data to a unique homogeneous sample. Despite these problems are quite common, even in developed countries, many practical applications and statistical methodologies have little or no tolerance to missing values [9, 11]

The treatment of gaps in the records or relocation of rain gauges, especially when dealing with large databases, requires the set-up of specifically-conceived methodologies aimed at bypassing or reconciling the inconsistencies [12]. Two approaches can be adopted for dealing with non-uniform sets of records: (i) a precautionary approach, that consist in assuming a minimum acceptable threshold of record length and discard the series shorter than the threshold and (ii) a preservative approach, focused on the identification of methodologies aimed at extracting all the available information even from the shorter records. While, on the one hand the approach (i) can discard important information hidden in the shorter records, affecting the results of the regional rainfall frequency analysis, the approach (ii) turns out to be complex, computationally demanding, and can lead to errors when based on non-robust assumptions [11].

A number of procedures for recovering information from short records can be found in the literature. Various authors propose the adoption of interpolation techniques along the time-axis, to estimate the missing data of environmental series (linear or logistic regression, polynomial or spline interpolation, inverse distance weighting, ordinary kriging, etc. - see, e.g., [13, 14]). The statistical techniques available include also artificial neural networks and nearest neighbours [15, 16], approaches based on Kalman filters [17], non-linear mathematical programming [18] and normal-ratio and inverse distance weighting methods [19].

In [9] it is argued that the complexity and the computational burden

associated with these techniques often make them unsuitable for an application over large scales. This usually leads to the adoption of conceptually over-simplified approaches (e.g., filling the gaps with fixed values, often corresponding to the sample average of the series) not adequate to represent the complexity of the phenomena. The authors propose a simple method based on the analysis of the auto-correlation structure of the series, amenable for a quick filling of sporadic gaps. However, the technique is viable if the percentage of missing values in the time series is limited. When the gaps are frequent and systematic (e.g., in developing countries [20]) and when data show low auto-correlation in time, this approach is not effective.

Even when long uninterrupted rainfall records are available, an IDF relation is basically valid only at the point where it is estimated. Rain gauges are generally not evenly distributed in space, and they allow only for a point estimation of the parameters of the rainfall distribution. To extend estimates to ungauged locations, rainfall data are usually interpolated, either by considering the distribution parameters estimated at the station location (e.g., [21, 22]), or by estimating the IDFs after pooling the available data within homogeneous areas defined by geographical boundaries, or centred around a location of interest (see, e.g., [23]). In the presence of data scarcity, some recent studies also propose to include external sources of data (e.g., remote sensing data [24]) in the procedure. Regional techniques for rainfall frequency analysis actually build representative growth curves from larger samples resulting from pooling. On the other hand, the use of a regional frequency curve is suitable only when the spatial dependence is weak enough to enable transferring information to a site of interest from the surrounding gauged sites [25]. When spatial dependence is significant, as in the presence of high discontinuity in the rainfall distribution, or due to different climatic and orographic conditions, different approaches should be preferred. For instance, [26] propose a statistical approach that involves the adoption of a bootstrap algorithm aimed at providing complete annual maxima series at each location, taking into account all data observed at surrounding stations with decreasing importance when distance increases. This kind of approach allows one to overcome the problem of data filling, but the bootstrap procedure produces results that deviate significantly from the sample spatial distribution, ignoring the existence of long and reliable records at some locations.

In this work, a simple approach able to provide a set of complete series of rainfall data for each location of the domain under analysis is proposed.

The methodology, described in section 3.1, is summarized in figure 1. It is based on the sequential application of the ordinary kriging equation to the values recorded annually in the region of interest. The so-called "patched kriging" procedure preserves the spatio-temporal information of the annual maxima recorded by the monitoring network, "patching" them together, i.e., considering each record just like a point in the (x,y,t) space (where x and y are the planimetric coordinates and t is the time).

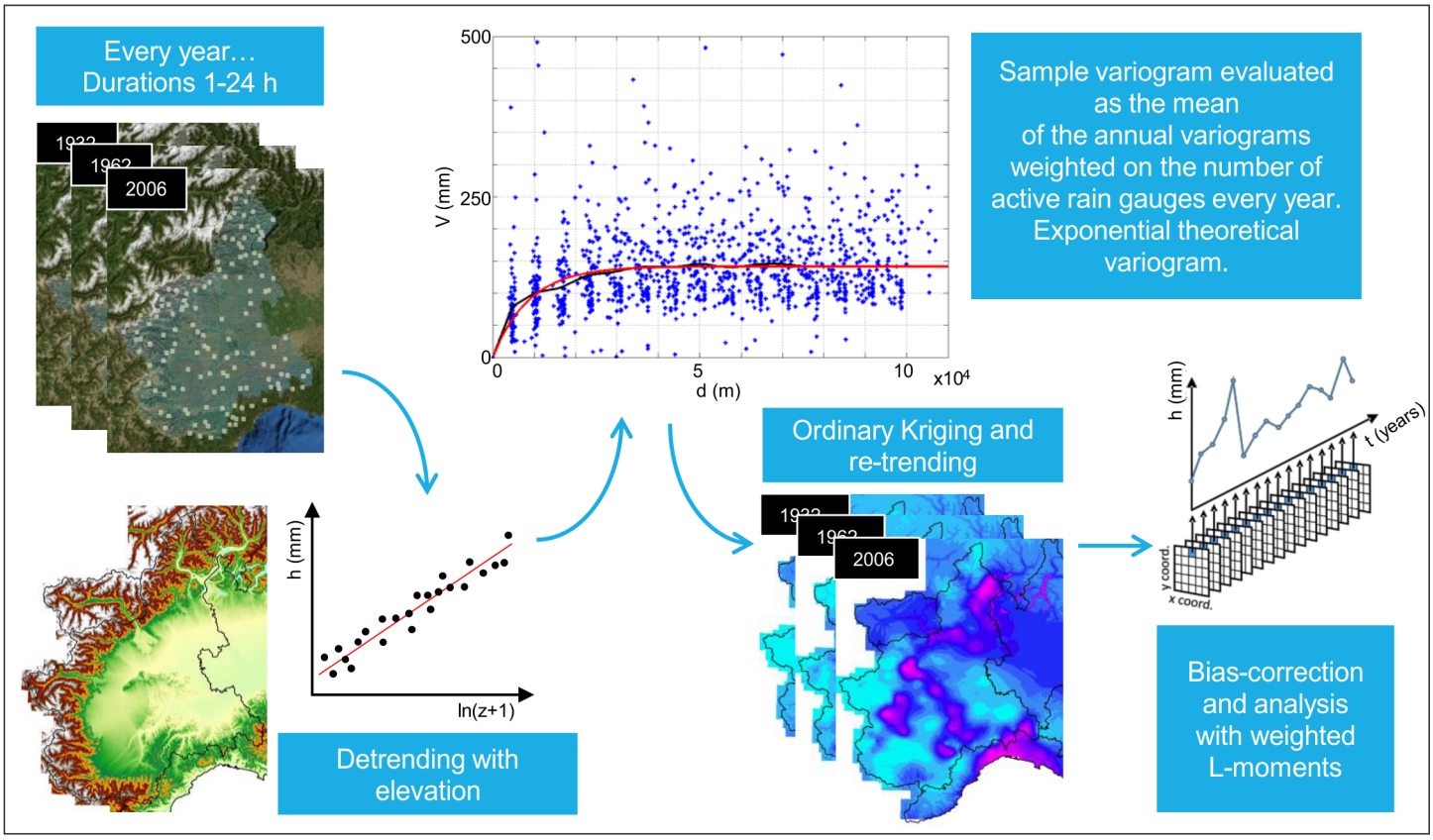

Figure 1: Flow chart of the "patched kriging" methodology.

From an operational point of view, this methodology has a low computational cost and does not require to work with stationary or significantly autocorrelated data, as it does not involve any interpolation along the time-axis. This feature proves to be particularly effective when dealing with frequent rain gauge relocations, allowing on the one hand to maximize the usable information at gauged sites, and on the other to extend the analysis to the

ungauged ones.

2. Data and case study

The region considered for the demonstration of the proposed methodology refers to the Piemonte region, an area of about 30000 km^2 in the North-Western part of Italy, shown in figure 2a. The area is characterized by a very heterogeneous orography, flat or hilly in the centre, surrounded by the Alps in the North-West and by the Ligurian Apennines in the South, with the minimum elevations of the order of a few tens of meters a.s.l. and the maximum ones exceeding 4000 m a.s.l. Several regional-scale hydrological analyses have been performed with a focus on this area (e.g. [27, 28, 29]); in all cases, the availability of accurate extreme-rainfall statistics is an essential prerequisite for obtaining consistent results.

A dataset of annual maximum rainfall depths over duration intervals of 1, 3, 6, 12 and 24 hours from 1928 to 2010 has been assembled for this analysis. The data before the '90s were collected from the publications of the National Bureau for Hydro-Meteorological Monitoring ($SIMN$). After 1987 the network was gradually taken over by the Regional Environmental Agency ($ARPA$ Piemonte) that removed, substituted or relocated some of the stations. Gauge data from neighbouring regions has also been considered to limit the edge effects. Overall, nearly 500 gauging stations have worked for at least one year in the considered period.

Annual maximum values have been extracted from the original rainfall series by the competent authorities using sliding time windows [30, 31]. The original series have a resolution in time varying from 1 hour for the oldest stations to 5 minutes for the most recent ones.

Figure 2b illustrates the data availability over time. It shows how irregular the available database is. This is a rather typical situation in Italy: only very few of the stations have a complete uninterrupted record, while the large majority has experienced interruptions, relocation or replacement/renewal of the instruments. As a consequence, more than 50% of the considered rain gauges have series shorter than 20 years, as shown in figure 2c.

In this context it is clear that, despite the large and dense rain gauge network available, a regional frequency analysis in the study area would require a preliminary work aimed at tracking the modifications in the network and harmonizing the whole database.

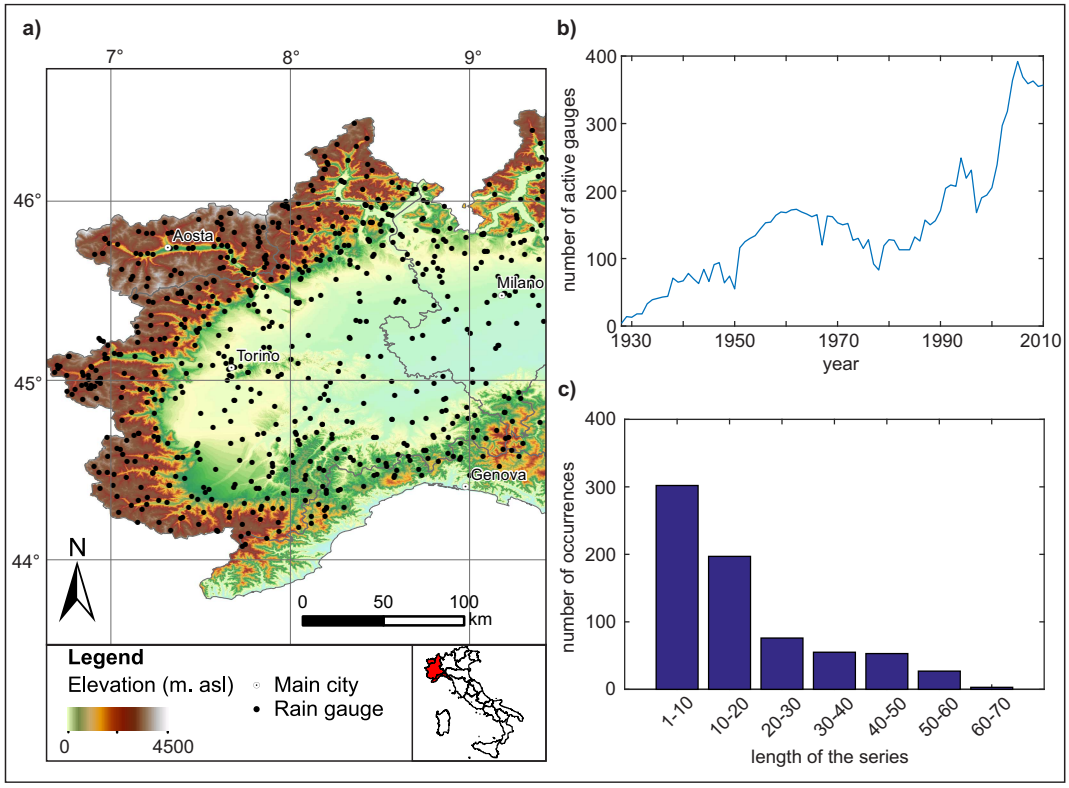

Figure 2: (a) The study area and the location of the available stations. (b) Number of active station per year and (c) number of available series per class of record length in the study area.

Numerous examples of gap-filling techniques for time series (see e.g., a review in [32]) and of space-time interpolation of rainfall data over relatively coarse grids (e.g., [33, 34]) are available in the literature. Less attention has been paid to the treatment of discontinuous records coming from a network of rain gauges with spatially varying positions. In these cases the usual approach is to exclude the series shorter than a given threshold, setting a minimum length suitable for the statistical analysis. However, this leads to exclude a large potential of information, affecting the robustness of the results. Consider, e.g., a station where less than 20 years of data have been recorded before being relocated few kilometres apart and that, after the relocation, has recorded an additional series shorter than 20 years. Setting a minimum length of the series equal to 20 would lead to lose almost 40 years of data.

The information content of the short series can be significant, especially in the presence of intense and localized rainfall events. Figure 3a shows the available series of 24-hour annual maxima for the "Caselle" rain gauge (45.19°N, 7.65°E, WGS84). During year 2008 a severe localized thunderstorm occurred in the area, with the rain depth approaching 300 mm in 24 hours. In that year, only the "Caselle" rain gauge recorded such a large rainfall amount, as shown in figure 3b. All the information related to this severe rainfall event is contained in a 7-years long time series, that would be ignored in many of the traditional frequency analysis techniques. In the following sections we describe how the proposed methodology allows at preserving this kind of information while maintaining a set of robust statistical procedures for the estimation of the design rainfall at a generic location.

3. Methods

3.1. The patched kriging technique

The proposed approach, called "patched kriging", allows one to produce regular spatial datasets by analysing the available rainfall data year-by-year, assuming that spatial gradients can act as a proxy for temporal gradients [35]. In this procedure, each measurement is considered a point in the three dimensional (x,y,t) space.

The "patched" procedure is amenable for application with any spatial interpolation method (e.g., Inverse Weighted Distance, etc.). In this work, we propose the use of the kriging interpolation method [36], because it can provide useful information on the estimation uncertainty at each location. Vari-

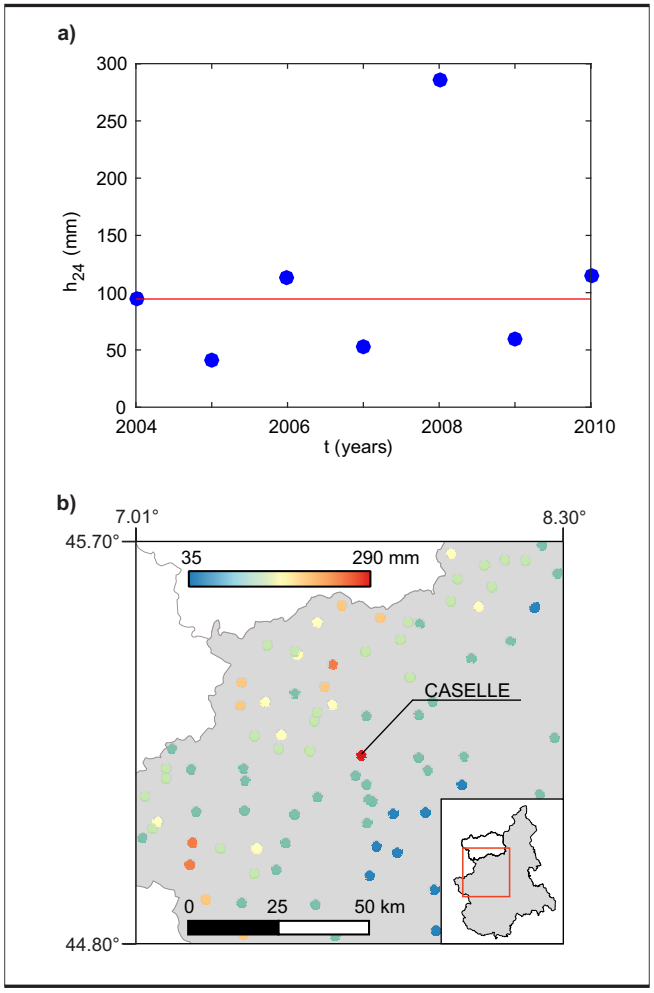

Figure 3: (a) Annual maxima for the 24-hours duration series of the "Caselle" rain gauge (45.19°N, 7.65°E, WGS84). The red line shows the median of the series. (b) Annual maxima for the 24-hours duration recorded during year 2008 at a sub-sample of the database.

ous kriging methods (i.e., simple kriging, ordinary kriging, universal kriging, etc.) have been developed based on assumptions about the model. At this stage, we have not found any significant advantage in choosing a particular kriging method. Therefore, for the sake of simplicity, ordinary kriging is considered. Detailed description of the ordinary kriging algorithms is available in the geostatistical literature (e.g., [37]).

Ordinary kriging assumes that the spatial variation of data is stationary and ergodic across the domain [38]. Kriging relies on the assumption that all the random errors are second-order stationary. This means that the covariance between any two random errors depends only on the distance and, possibly, on the direction that separates them, not on their exact locations. This leads to the need to analyse and remove the possible correlation between rainfall and elevation, especially in areas characterized by a complex orography [39]. The analysis of the correlation with topography also allows one to compensate for the lack of information at the small scale, improving the global performance of the method [40]. Various approaches have been adopted in the literature for dealing with this problem: among the others [41, 42] propose to perform linear regression on precipitation vs elevation, subtract the regressed elevation effect and perform the kriging on the elevation-adjusted data. The same approach has been adopted with positive outcomes in [43] for the Alpine area. Similarly, in this work the relation between h_d (mm), i.e. the annual maximum precipitation with duration d (h), and elevation z (m) is assumed to follow the equation:

$$h_d = m \cdot \ln(z + 1) + m_0 + \varepsilon_d \tag{1}$$

where m is the slope of the regression line, m_0 (mm) is the intercept and ε_d (mm) the residual. The logarithm of elevation is adopted as an independent variable, in order to limit the weight that linear interpolation would attribute to the stations placed at low altitudes. The regression procedure takes into account the values recorded at all the stations in all the years simultaneously. This stems from the assumption that the relationships between precipitation and elevation is invariant over time.

Once assessed the regression significance, de-trended at-station precipitation values $h_{d,0}$ (mm) are computed for all the durations by removing the elevation effects from the observed value h_d.

The degree of spatial dependence in the kriging approach is expressed

using a sample variogram given by:

$$V(L) = \frac{1}{n(L)} \sum_{L_{ij}} (\alpha_i - \alpha_j)^2 \tag{2}$$

where $V(L)$ is the variance, which is defined over observations α_i and α_j lagged successively by lag-distance L, with $n(L)$ representing the number of pairs of the sample separated by lag L [32].

De-trended values are therefore used to define the annual sample variograms. For each year Y in the observation interval, the annual sample variogram $V_Y(L)$ is computed according to equation 2. A global sample variogram is obtained averaging the annual sample variograms, each weighted by the number of active stations in the considered year. The sample variogram is then converted to an analytical function, i.e., the theoretical variogram, $\gamma(L)$. Generally, several variogram models are tested before selecting a particular one. In this study the four most widely used variogram models (i.e., spherical, exponential, Gaussian and circular) have been considered [32]. After a visual analysis of the empirical variograms for the considered durations and some preliminary first-attempt fits of the models to the data, the exponential form is adopted:

$$\gamma(L) = c_3 + c_1(1 - \mathrm{e}^{-L/c_2}) \tag{3}$$

where L (m) is the lag-distance, c_1 (mm^2), c_2 (m), c_3 (m) are the sill, the range parameters and the nugget of the variogram, respectively [44]. The nugget effect is neglected by setting c_3=0, considering the rain gauge records not affected from measurement errors. This is a strong assumption, but as the work deals with annual maximum values, the impact of the instrumental error can be considered not significant for the aim of the analysis at this stage. The other variogram parameters are fitted to the data by minimizing the root mean square error.

We work on a gridded 250 m x 250 m domain that is set equal to the resolution of the Digital Terrain Model used, after considering a reasonable balance between the topographic detail and the station spatial density. If more than one rain gauge falls in the same cell, the largest measured value is considered.

Ordinary kriging equations are applied independently in each year. For each location, the values recorded at the nearest gauged cells are weighted according to the variogram and used to estimate the local value. Since we have

neglected the nugget effect, measurements in gauged cells are automatically preserved.

According to the literature, the number of nearest gauged cells to be considered is arbitrary and depends on the sampling pattern and on the covariance matrix structure [45]. While, on the one hand, using the whole sample for applying the kriging equations could grant shorter computational cost, as the estimation domain is the same for all the cells of the grid, on the other hand, smaller neighbourhoods are preferred when there is the need to represent small-scale variability. Moreover, some authors [46] underline that the use of large neighbourhoods does not lead to a significant increase in the robustness of the estimation, as the weight associated to a distant observation quickly tends to zero [45]. Therefore, usually, only the stations in a neighbourhood of the estimation point are considered. Some authors suggest to consider a number of stations around 10-20 [47], even though the size of the neighbourhood should be selected according to the c_2 parameters of the variogram. In this work, the significant variations of both the number and the spatial distribution of the stations along the time axis leads to the need of summarizing the spatial information in a weighted mean variogram. Considering the value of the range of this variogram for assuming the width of the estimation domain could affect the results, specially in years and in areas with a low density of information, leading to consider an insufficient number of rain gauges. After a preliminary sensitivity analysis, aimed at preventing the flattening of the estimated values on a global regional mean, the estimation domain is therefore limited to the nearest 10 rain gauges, for all the cells, for all the years.

Sequential kriging application leads to the development of a set of grids (as many as the considered years), containing the estimated values of precipitation maxima for each location of the study area, configuring a "cube" of rainfall data in the (x,y,t) space (figure 4a), which will be referred to as the "rainfall cube". The ordinary kriging equation provides also a "variance cube", containing the kriging variance for each cell in each year. The kriging variance is a measure of the uncertainty of the estimation for the values predicted by kriging.

"Coring" the "cube" along the t-axis (i.e., extracting a complete series, once fixed a pair of x and y coordinates, by varying t) one can obtain complete "cored series" (i.e., complete series extracted from the cube) for each x-y pair (figure 4b). Each uninterrupted annual maxima series, related to a generic cell in the considered domain, is associated to a series of kriging variances,

informing about the uncertainty of each data. The length of all the series equals the length of the considered time period.

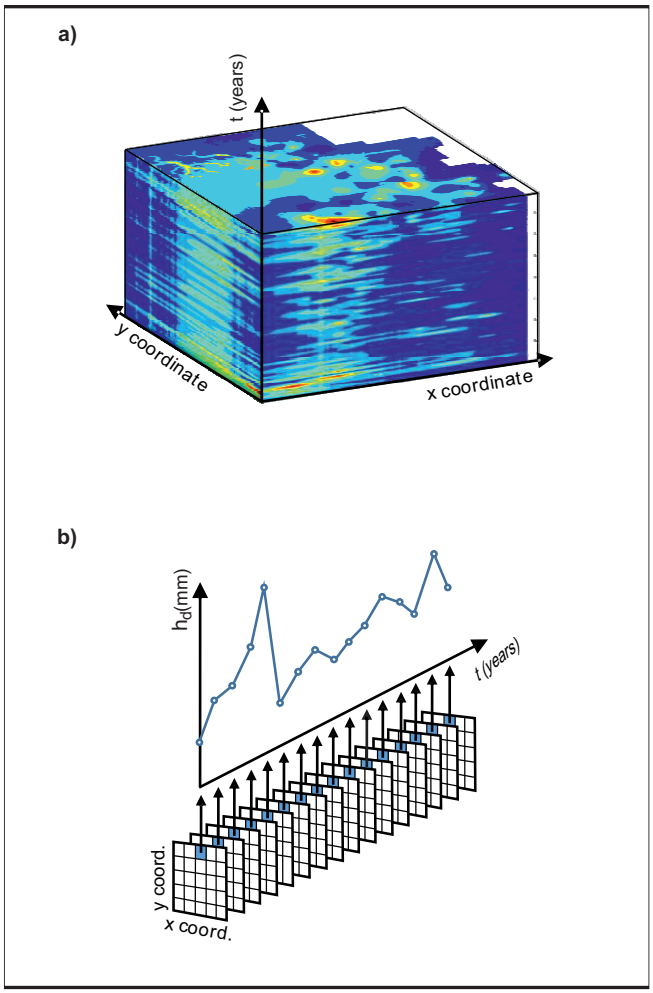

Figure 4: (a) The "rainfall cube" obtained with the "patched kriging". (b) Example of the extraction of a "cored series" from the "rainfall cube".

3.2. Application

The "patched kriging" technique is applied to the study area in Piemonte. Annual extremes for each duration d are considered as separate series, so as to obtain 5 different series per rain gauge, leading to 5 rainfall and variance cubes.

Table 1: Parameters of the regression equation 1 for precipitation versus elevation, for different durations (* indicates a significant trend at a 5% level).

d (h)	m	m_0 (mm)
1	-3.56*	48.73
3	-2.57*	55.72
6	-0.32	54.59
12	3.52*	49.08
24	8.34*	44.54

Regression of rainfall depths with the logarithm of elevation has been carried out, considering equation 1 for the 5 durations. Results are reported in table 1. The trends significance is evaluated with a Student's t test with an acceptance level α=0.05.

Referring to the coefficients in table 1, the maximum annual precipitations of duration 1 and 3 hours show a declining trend with elevation, which loses significance for the duration 6 hours and becomes a positive trend for the durations of 12 and 24 hours. This justifies the absence of the expected increasing trend of the intercept of the regression lines with the duration, and is consistent with the findings of [43] that relate the different behaviour with the nature of the events typical of the different durations (mostly convective for shorter durations, stratiform for longer ones).

Using the coefficients reported in table 1, the de-trended precipitation ($h_{d,0}$) is obtained. For d=6, we set $h_{6,0}$=h_6, due to the lack of significance of the hypothesis $m \neq 0$.

We then proceed with the definition of the sample and theoretical variograms according to equation 3. Annual sample variograms (not shown) are characterized by a large annual variability, partially ascribable to the low data density in the first analysed years, that leads to sample variograms with large variance. To avoid loss of robustness, as previously noted, the annual variograms are weighted according to the number of annual active stations. Table 2 reports the coefficients of the obtained theoretical variograms for the different durations.

With the application of the ordinary kriging equations, as described in section 3.1 a set of 5 "rainfall cubes" (one per duration) with the related "variance cubes" is obtained.

Table 2: Estimated parameters for the theoretical exponential variograms

d (h)	c_1 (mm^2)	c_2 (m)
1	142	6709
3	334.7	8798
6	574.2	10240
12	1051	11520
24	2028	13650

3.3. Weighting the L-moments

In order to guarantee a robust data-based approach, the proposed methodology aims at preserving as much as possible the statistics of the original series in the cored ones. This operation should be treated with caution, considering the different length of the original series [11] (e.g., extracting the characteristics of a 80 years long series from a subset of 10-20 data can lead to large bias, as the characteristics of the sample can be not consistent with the characteristics of the corresponding complete series). The "patched kriging" technique helps to increase the robustness of the operation. It allows one to preserve the recorded data, filling in the gaps with spatially estimated values.

In order to take into account the different nature of the data (i.e., part of the core is measured and part is estimated by kriging) differential weight is given to each value in the evaluation of the characteristics of the cored series (i.e., more weight is given to the measured values and to the values estimated in years with more observations). The kriging variance is then considered to weight the contribution of each value to the estimation of the sample L-moments of the series. The kriging variance is a measure of the uncertainty of the estimation: it is larger in cells far from gauged locations and, for a fixed cell, it increases/decreases when the number of annual available stations in its proximity decreases/increases. For instance, in figure 5a the fast increase of the kriging variance when getting far from the stations is shown. Moreover, considering the northern part of the study area, for year 1987 (5a above), when it totally lacks active stations, the variance reaches very large values while it shows generally lower values (around 1700 mm^2) for year 2010, when a dense network is available.

In the detail, for evaluating the sample L-moments of a cored series, a weight $w_i = \sigma^2_{max}/\sigma^2_i$ is assigned to the i-th value of the series, characterized

by the σ_i^2 kriging variance (with $\sigma_{max}^2 = \max(\sigma_i^2)$ for the considered series). Further details are available in the Appendix A.

As an example, the "Caselle" station, mentioned in section 2, was installed in 2004. The cored series of the annual maxima for 24 hour duration of the cell related to its location is reported in figure 5b. The mean of the cored series (i.e., the red dashed line) turns out to be significantly lower than the one of the original series of 7 data (i.e.,the yellow line). Analysing the series of the kriging variance of the "Caselle" location (figure 5c) one can note the sensitivity of this parameter to the number of globally available stations: as previously mentioned, the kriging variance increases/decreases with the decrease/increase of the number of active gauges. It drastically decreases in year 2004, when the station has been activated.

From figure 5b we can also observe that the weighted mean, evaluated with the weights reported in figure 5d (left axis), is almost equal to the mean related to the period 2004-2010. When a station is located in a previously ungauged cell, the kriging variance decreases drastically and this leads to give virtually zero weight to all the previously kriged values. Considering the lack of reliability of L-moments estimated on short series, this phenomenon should be avoided, as this would undermine the benefit of the "patched kriging" methodology. A maximum threshold w_{max} is therefore set. For $w_i > w_{max}$, w_i=w_{max} is considered. After some sensitivity analysis, aimed at giving large enough weight to the measured values without denying the contribution of the reconstructed ones, we set w_{max}=10. The final weights adopted for the "Caselle" cell are reported in figure 5d (right axis), and the resulting mean values is shown in figure 5a with a black line.

4. Analysis and validation of the patched series

4.1. Series validation

At first, in points where sample data are available, the cored series are validated by comparing their L-moments with those of the measured series. L-moments have been considered for evaluating the quality of the results, as they provide information on the underlying probability distributions.

Given the lack of significance of the shorter series from a statistical point of view (i.e., as previously mentioned, the L-moments estimated from short fragmented series can be not-consistent with the real characteristics of the related uninterrupted series) the validation is restricted to the series with more than 20 years of data. Figure 6 reports the comparison between τ, τ_3

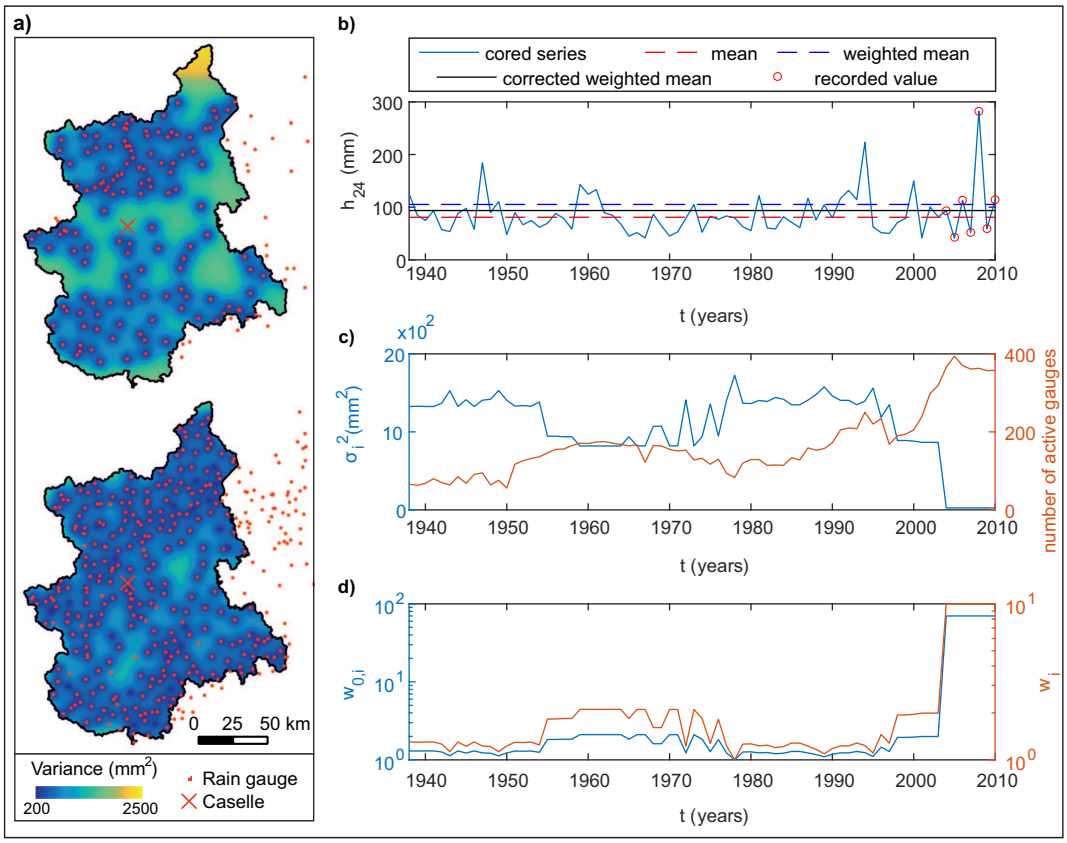

Figure 5: (a) Map of the kriging variance for year 1987 (above) and year 2010 (below). The red cross shows the location of the "Caselle" rain gauge (45.19°N, 7.65°E, WGS84), installed in 2004. (b) Cored series of the "Caselle" rain gauge for 24 hours duration. The red circles mark the recorded values. All the other values are estimated with the "patched kriging" technique. The mean of the series, the weighted mean and the weighted mean with w_{max} threshold are also shown. (b) Kriging variance series for the "Caselle" location (left axis) related to the number of active gauge per year (right axis). (c) Series of the weights related to the "Caselle" series (left axis). The right axis refers to the same series, after correcting it, by setting w_{max}=10.

and τ_4 (i.e., the coefficient of L-variation, L-skewness and L-kurtosis respectively [7]) of the measured versus the estimated series for the five durations mentioned above.

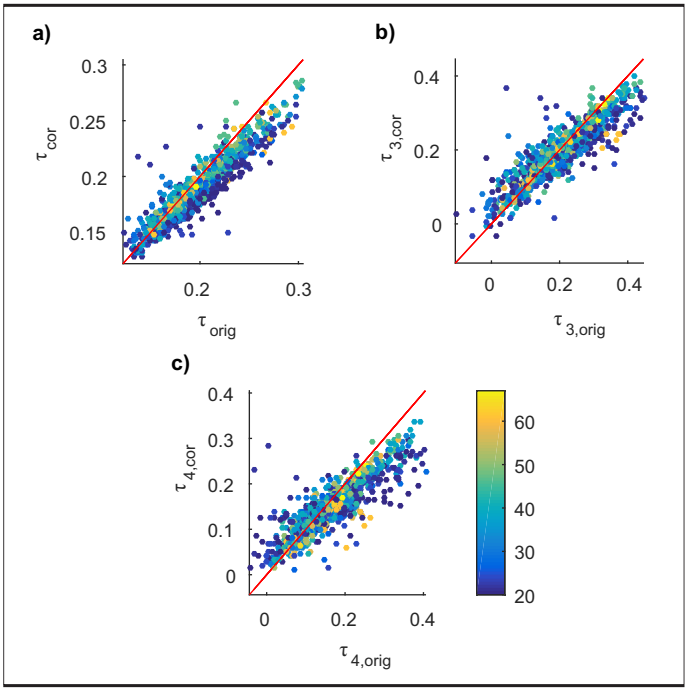

Figure 6: (a) τ, (b) τ_3 and (c) τ_4 of the measured (i.e., orig) versus the cored (i.e., cor) series for all the durations. The chromatic scale refers to the length of the series.

The comparison demonstrates the ability of the methodology to preserve the L-moments, except for a slight underestimation of the τ of the cored series, as seen in panel (a) in figure 6 that compares the measured with the cored series for all the durations.

To assess the performance of the methodology even in cells without sample data, or with a number of data that does not allow for a robust estimation of the sample L-moments, the clouds of the sample L-moments of the cored series in the L-moments ratio diagrams [23] are compared with those of the original series with more than 20 years of data, considering all the durations together (figure 7a-b).

A significant underestimation of the second order L-moment (τ) is evident from the analysis of panel (b), while a slight underestimation of the τ_3 and τ_4 values appears from panel (a); this implies that the cored series

denote smaller variability along the time axis than the original ones. This is an expected drawback when applying a spatial interpolation technique, and is consistent with what emerges from the analysis of the gauged cells in figure 6a. As the underestimation of τ leads to underestimation of the design rainfall, a correction procedure has been developed, as described in the following section.

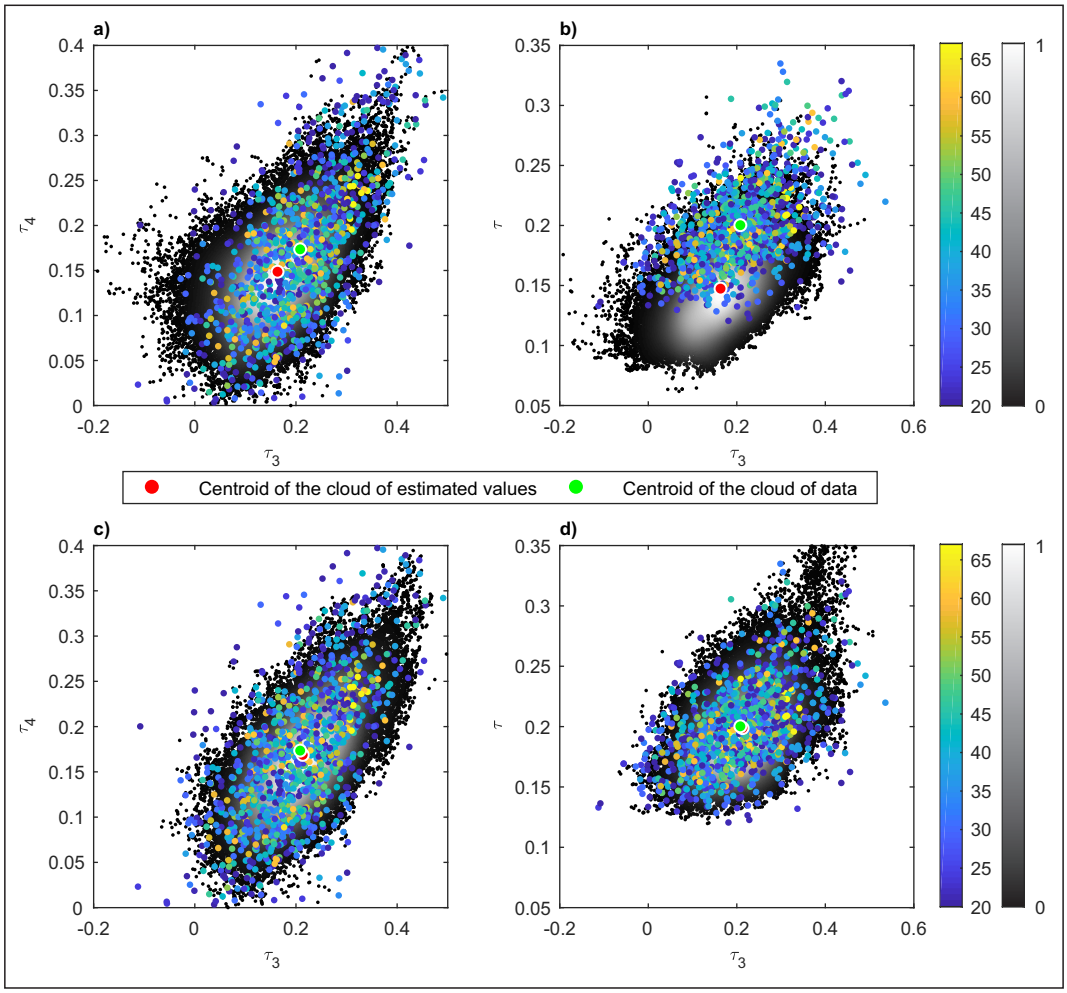

Figure 7: L-moments ratio diagrams (a),(b) before and (c),(d) after the correction considering all the durations. The greyscaled cloud of points represent the cored series. The greyscale is proportional to the density of points. The coloured dots represent the original series. The colourscale is related to the length of the series.

4.2. Correction of the bias in the variance

By construction, a weighted average of identically distributed random variables has a different statistical distribution than the variables themselves. The "patched kriging" technique involves merging the locally observed values and the interpolated ones, which can therefore have different statistical distributions. This operation can potentially introduce bias and, in particular, lead to reduce the coefficient of variation of the estimates. To correct this behaviour a bias-correction procedure is proposed, conceived at increasing the variance of the cored series.

Consider a situation when a series $x_i(t)$ is obtained from the "patched kriging" methodology. The temporal average is $\bar{x}_i$ and, as shown in figure 7b, the $x_i(t)$ values are underdispersed around $\bar{x}_i$. A natural way to avoid the underdispersion would be to inflate the distance from the mean through multiplication by a factor K_0:

$$\hat{x}_i(t) - \bar{x}_i = K_0 \cdot (x_i(t) - \bar{x}_i) \tag{4}$$

with $K_0 > 1$. However, equation 4 can lead to negative rainfall values, that are obviously not acceptable. Equation 4 is thus applied to the logarithms of the variables, leading to:

$$K = \frac{\ln(\hat{x}_i(t)) - \ln(\bar{x}_i)}{\ln(x_i(t)) - \ln(\bar{x}_i)} \tag{5}$$

the correction equation then reads:

$$\hat{x}_i(t) = \bar{x}_i \left(\frac{x_i(t)}{\bar{x}_i} \right)^K. \tag{6}$$

For calibrating the K coefficient we start from the heuristic observation that the distance of the analysed location from the closer gauged cells is one of the main determinants of the bias. Cells far from gauging stations are expected to be more affected by the smoothing effect of the interpolation and thus to show less variability around the average. Hence:

- If the target point is close to a gauging station, the distribution of the cored series will likely be very similar to the one of the original series, and then correction should be very limited.

- When the target point moves further away from the gauging stations, the smoothing effect becomes very relevant and the correction becomes essential.

We therefore expect the correction factor K to be an increasing function of the distance from the rain gauges, i.e.

$$K = f(D_s) \tag{7}$$

where D_s is the distance, f is an increasing function and $f(D_s = 0) = 1$.

The distance from the rain gauges is computed as follows. For each year we assign to each cell a D_s (km) value, representing the inverse average distance of the cell from the nearest 10 gauged cells (the ones considered when the kriging equations are applied), evaluated as:

$$D_s = \frac{1}{\frac{1}{10}\sum_{j=1}^{10}(\frac{1}{\delta_j})} \tag{8}$$

with δ_j being the distance of the cell from the j-th closest gauged one. We consider the inverse average distance (and not the standard average distance) in order to assign a D_s value approaching zero when the cell coincides with a gauged cell.

In order to estimate the dependence of K on D_s, we take the average of equation 5, conditioned on D_s. We note that, on the right-hand side of equation 5, we have at the numerator a variable which is independent of D_s, by definition (otherwise the correction would not be effective). The average thus reads:

$$\frac{\Delta(D_s)}{a_0} = \mathrm{E}\left[\frac{1}{|\ln(x_i(t)) - \ln(\bar{x}_i)|}\middle|_{D_s}\right] = f(D_s) \tag{9}$$

where $\mathrm{E}\left[\frac{1}{|\ln(x_i(t))-\ln(\bar{x}_i)|}\middle|_{D_s}\right]$ is the average, conditioned on a specific D_s value, and $a_0 = \mathrm{E}\left[\frac{1}{|\ln(x_i(t))-\ln(\bar{x}_i)|}\middle|_{D_s=0}\right]$.

In practice, the $\Delta(D_s)$ value is estimated separately for each duration. For each year, we build equally consistent D_s classes to compute $\Delta(D_s)$, considering all the cells belonging to each class. The ($\Delta(D_s)$,D_s) pairs belonging to all years are then pooled together and the median value for each D_s class is considered. They are represented as dots in figure 8.

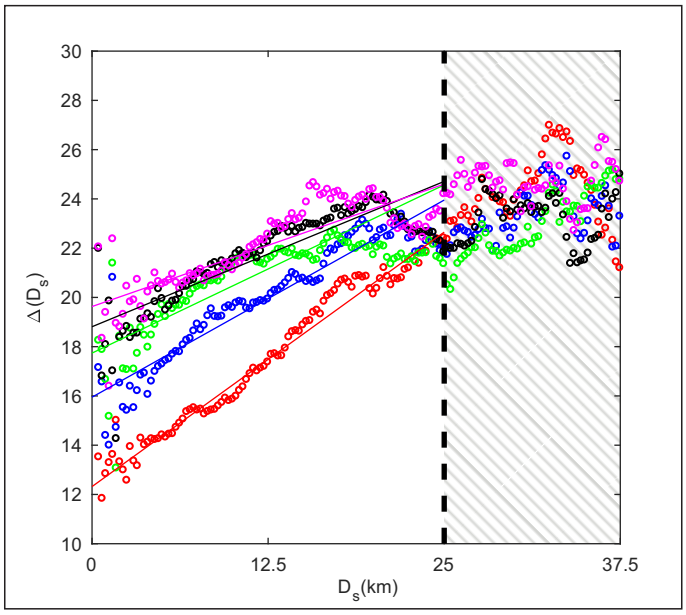

Figure 8: Median value of $\Delta(D_s)$ for equally spaced D_s classes. Regression lines refers to the [0, 25] km range.

The increase of $\Delta(D_s)$ with increasing distance is clear for all durations, for D_s values up to 25 Km, which confirms our hypotheses on the influence of the distance on the distribution of the bias. Furthermore, it emerges from figure 8 that in this range the relation between D_s and Δ can be approximated with a linear equation:

$$\Delta(D_s) = a_0 + a_1 \cdot D_s \tag{10}$$

with a_1 (km^{-1}) representing the slope of the regression line. Combining equation 10 with equations 9 and 7 we obtain:

$$K(D_s) = 1 + \beta \cdot D_s \tag{11}$$

with $\beta = \frac{a_1}{a_0}$ (km^{-1}). For $D_s > 25$ km the behaviour becomes less consistent, probably due to the small number of stations with large D_s available: due to the difficulty of calibrating a proper relationship, K is kept constant in this range. Considering that, as previously mentioned, the slope of the regression line changes for the different durations, the final correction factor reads:

$$K(D_s, d) = \begin{cases} 1 + \beta(d) \cdot D_s & D_s \leq 25, \\ 1 + \beta(d) \cdot 25 & D_s > 25. \end{cases} \tag{12}$$

Table 3: Coefficients β of the correction function $K(D_s, d)$ for the different durations d.

d (h)	β (km^{-1})
1	0.034
3	0.020
6	0.015
12	0.013
24	0.010

$\beta(d)$ values for the different durations are reported in table 3.

Once assigned to each cell of each year a suitable correction factor (equation 12), all the "rainfall cubes" are corrected according to equation 6 and the L-moments ratio diagrams are re-computed. Results are reported in figure 7. Comparing the diagrams of the corrected values (panels (c) and (d)) with those of the original cored series (panels (a) and (b)), it is evident that the correction procedure works correctly, making the τ and τ_3 values of the cored series consistent with the L-coefficients of the observed ones. Considering the position of the centroid of the cloud of the cored series and comparing it with the one of the data, it is indeed clear that, after the correction, the methodology is able to provide unbiased results.

For further assessing the quality of the obtained results, for each duration, the 26 series with more than 50 years of data are considered for carrying out a leave-one-out cross-validation procedure. The limit of 50 years of data has been selected for limiting the computational burden of the operation, that would be extremely large if considering the whole dataset. Leave-one-out cross-validation is a special case of cross-validation where the number of folds equals the number of instances in the dataset [48]. The whole "patched kriging" technique is then performed leaving one of the series out at-a-time, obtaining interpolated values year by year, and correcting those values with equation 6. Figure 9 compares each recorded annual maximum value with the corresponding cored one, obtained from the cross-validation procedure. The shape of the scatter suggests that the "patched kriging" technique is able to provide not only patched series with L-moments consistent with those of the original ones, but also to reconstruct reliable annual maxima at ungauged areas.

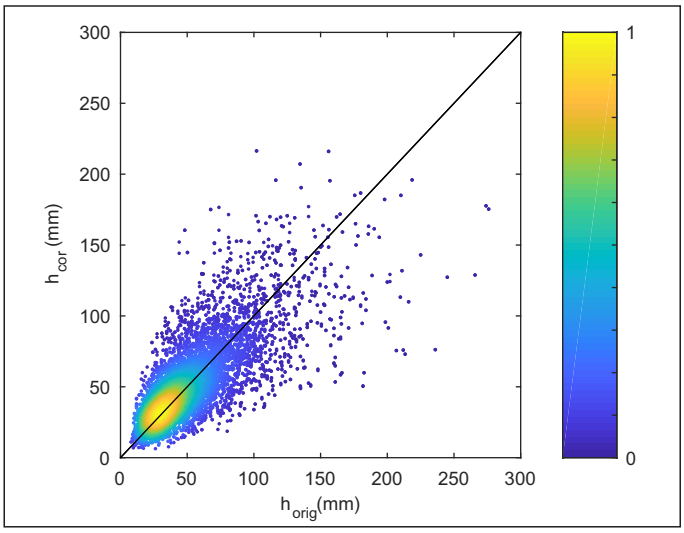

Figure 9: Comparison between the annual maxima for the different durations of the series with more than 50 years of data and the corresponding cored values, estimated with the "patched kriging" technique in a cross-validation environment. The colourscale is proportional to the density of the points.

4.3. IDF curves

By considering the cored series, the coefficients a and n of the average IDF in the commonly adopted form $\bar{h} = ad^n$ are estimated for each cell in the study area.

Figure 10a-b shows the parameters distribution over the study area. To assess the validity of the results, the relative differences between the values of the parameters evaluated with the original series and the ones estimated with the cored ones is considered. The maps of the the spatial distribution of the differences (omitted) shows that no particular spatial clustering can be observed. Significant differences between the two sets of parameters are mainly related to the length of the original series, as shown in figure 10c-d. Comparing the differences with the length of the series, a decreasing trend with the length of the series is obtained, as explainable from the sampling variance theory. The "patched kriging" allows for a robust data-based spatial estimation of the IDF curves by increasing the robustness of the estimation at gauged sites, by filling the gaps in the series with data spatially consistent with the surrounding stations, and by allowing for the spatialization of the parameters to ungauged areas.

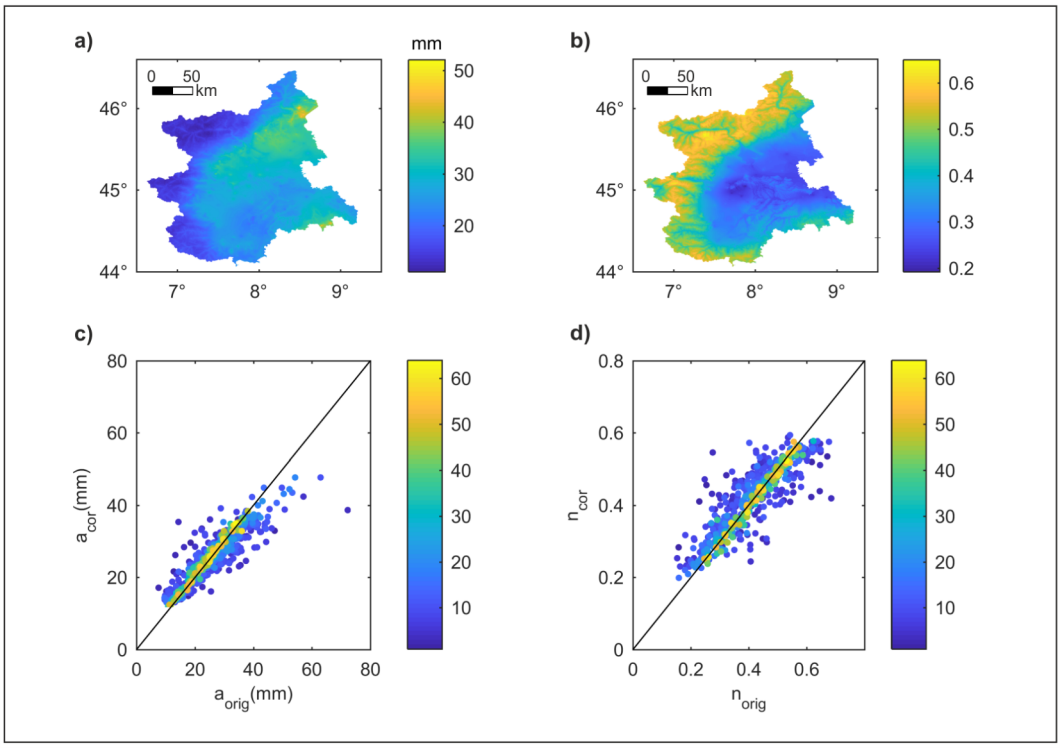

Figure 10: (a) a and (b) n parameters of the mean IDF curve. (c) a and (d) n parameters of the mean IDF curve estimated with the original and cored series. The colourscale refers to the length of the original series.

5. Frequency analysis at ungauged sites

In order to estimate the design rainfall for a given return period for a generic point in the domain under analysis, it is necessary to identify a probability distribution representing the annual maxima. It would be then possible to estimate the rainfall depth $\hat{h}_{d,T}$ related to a duration d and a return period T, using the average IDF curve previously identified and the growth factor K_T:

$$\hat{h}_{d,T} = ad^n K_T \tag{13}$$

Different probability distributions have been used in the literature to statistically represent the growth factor. Even if the identification of the best probability distribution lies beyond the scope of this work, in this section we illustrate a preliminary analysis of the distribution of the considered dataset, aimed at showing the potentiality of the "patched kriging" in providing a spatially consistent frequency analysis.

Figure 11a shows the points and curves representing different distributions commonly used in the analysis of extreme values. Plotting the L-moments of the cored series allows one to visually evaluate the global behaviour of the samples.

The diagram confirms that the Gumbel distribution is a good candidate to represent extreme precipitations at the regional scale, despite the centroid of the cloud of points is slightly shifted towards larger τ_3 values. To identify the amount of variability due to the sample size with a Monte Carlo procedure, 25000 series with a length of 72 year have been randomly extracted from a Gumbel distribution with scale and position parameters set to 1. This allows one to build a region in the (τ_3,τ_4) space occupied by parameters resampled from the original Gumbel function. In this region it is easy to delimit the 90% and 95% acceptance areas, that have been overlapped to the points estimated from the actual samples (see figure 11a). Most of the actual points fall into the domain of the Gumbel distribution. For the series characterized by larger skewness and kurtosis values the GEV distribution can be a viable alternative, despite the use of distributions with three parameters increases the uncertainty associated to the estimates. This uncertainty depends on the inherent difficulty in estimating the shape parameter of the distribution, especially in the presence of short and unevenly distributed records.

Figure 11b shows the spatial distribution of the cells whose L-moments fall inside the theoretical acceptance area of the Gumbel distribution. As expected, a regular pattern of Gumbel and non-Gumbel cores can be hardly

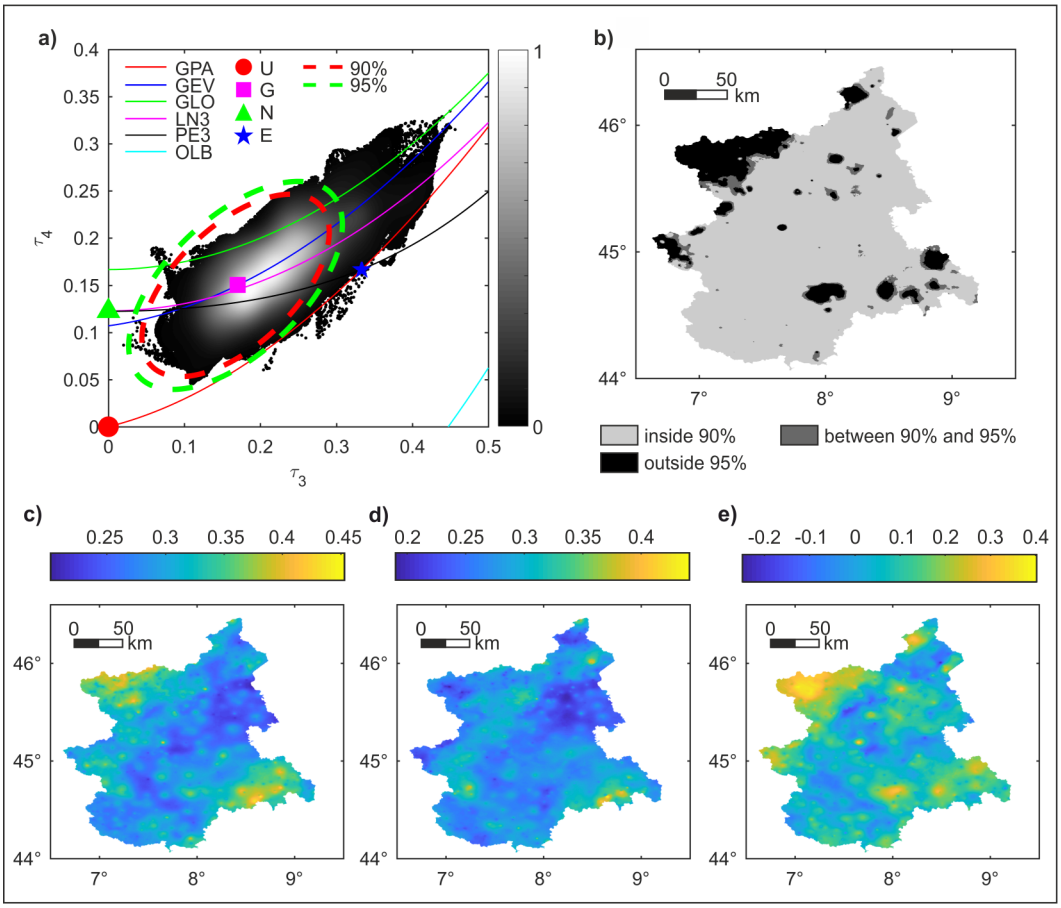

Figure 11: (a) L-moments ratio diagram related to the cored series. The ellipses represent respectively the 95% (green) and the 90% (red) acceptance area, defined by bootstrapping from a Gumbel distribution. The colour scale is proportional to the density of the points. Key to distributions: E - Exponential, G - Gumbel, N - Normal, U - Uniform, GPA - Generalized Pareto, GEV - Generalized Extreme Value, GLO - Generalized Logistic, $LN3$ - Lognormal, $PE3$ - Pearson type III. OLB is the overall lower bound of L-kurtosis as function of τ_3. (b) Spatial distribution of the cells falling inside the 90% and 95% acceptance area of the Gumbel distribution. Dimensionless scale (θ_2^*) parameter of the (c) Gumbel and (d) GEV distribution, normalized on the mean rainfall depth. (e) Shape parameter (θ_3) of the GEV distribution.

defined, due to the complex topography and to the different characteristics of the events generating annual maxima for different durations at the regional scale [5]. The mixture of scales involved in data generation (local and synoptic scale events) and the effect of orography on storm generation (particularly significant in the north-western Alpine and south-eastern Appenninic areas) does not allow the identification of a unique regional probability distribution. In addition, boundaries effects may occur at the edge of the analysed domain.

The growth factor of the GEV distribution can be expressed for a given return period T by the equation [49]:

$$K_T = 1 - \frac{\theta_2^*}{\theta_3}\Big[\Gamma(1-\theta_3) - \Big(-\ln\Big(\frac{T-1}{T}\Big)\Big)^{-\theta_3}\Big] \tag{14}$$

with $\theta_2^* = \frac{\theta_2}{\mu}$, where μ is the mean, $\theta_2 > 0$ as the scale parameter and θ_3 as the non-dimensional shape parameter. When θ_3=0, the GEV reduces to the Gumbel distribution [1]:

$$K_T = 1 + \theta_2^*\Big[-\gamma_E - \ln\Big(-\ln\Big(\frac{T-1}{T}\Big)\Big)\Big] \tag{15}$$

with γ_E as the Euler-Mascheroni constant.

We estimate the parameters of the distributions for each cell of the grid, both with the constraint θ_3=0 (forcing the use of a Gumbel distribution) and letting the shape parameter be freely estimated.

For the parameter estimation of both the distributions we adopt the L-moments methodology [50]. In detail, we use the average weighted L-moments among the different durations for estimating the dimensionless parameters of the Gumbel and GEV distribution. Maps of the estimated parameters are reported in figure 11, panels (c),(d) and (e).

The "patched kriging" allows not only for a consistent spatialization of the local information to ungauged areas but, as it emerges from the maps, to pursue a more robust estimation of the distribution parameters. For instance, the shape parameter of the GEV distribution, estimated with the original series, takes on values between -1 and 6. However, the shape parameter usually assumes values in a much narrower range, smaller or larger values being ascribable to an excessive sampling variability in small samples (see figure 11e). Moreover, negative shape parameters of the GEV distribution may be just an artefact of the data, attributable to bias in the estimation of the sample L-moments [51, 52]. In this study we obtain values in the [-0.2,0.4] range, with the large majority of data cores providing a $\theta_3 > 0$ value.

6. Conclusions

We propose a methodology for estimating rainfall extremes at ungauged sites in the presence of short and fragmented records, providing the basis for a spatially homogeneous and reliable frequency analysis of rainfall extremes on wide areas.

Treating each recorded annual maximum like a point in the (x,y,t) space, the "patched kriging" technique allows one to overcome the problems concerning the filling and merging of fragmented records, exploiting in the same time all the available information from the measurements, providing series consistent with the available measurements. Once a suitable correction factor for increasing the variability of the obtained series is applied, the "patched kriging" technique is able to reconstruct reliable annual maximum values also in ungauged areas.

The "rainfall cube" produced by the "patched kriging" technique provides greater robustness during the distribution estimation phase than other available procedures. The information concerning the estimation uncertainty is carried out thanks to a "variance cube" assembled with the estimation variance, per location, per year.

The "best" probability distribution can be therefore estimated at each location in the gridded domain. Despite a complete frequency analysis is beyond the aims of this paper, a exploratory methodology aimed at defining the global behaviour of different distributions at the regional scale is also proposed. Referring to the Piemonte region case study, the methodology confirmed good performances of the Gumbel distribution at a regional scale. As the procedure provides specific patterns of the areas of acceptability of the different distributions, application results allow for more in-depth meteorological and morphological analyses aimed at explaining the spatial variability of extreme rainfall.

From this perspective, the proposed methodology offers a powerful and expeditious procedure, suitable to grant an at-site evaluation of the best distribution and of the related quantiles, in the framework of a regional frequency analysis always consistent with the available data.

7. Acknowledgements

This work was funded by the ERC Project "CWASI" [grant number 647473] and by the FESR INTERREG IVA Italia - Svizzera 2007-2013,

STRADA 2.0 Project. The authors wish to thank Secondo Barbero and ARPA Piemonte for providing the data and for the useful discussion on the development of the methodology, Mattia Iavarone for cleaning and merging the database and Simon Michael Papalexiou and another anonymous reviewer for their valuable comments and suggestions to improve the quality of the paper.

Due to some legal restrictions, the full dataset can not be made freely available. Sources of the dataset are available at [53, 54, 55].

Appendix A. Weighted L-moments

Given a sample with size n sorted in ascending order: $x_{1:n} \leq x_{2:n} \leq ... \leq x_{n:n}$, and considering:

$$b_r = n^{-1} \sum_{i=r+1}^{n} \frac{(i-1)(i-2)...(i-r)}{(n-1)(n-2)...(n-r)} x_{i:n} \tag{A.1}$$

the sample L-moments can be written as:

$l_1 = b_0$

$l_2 = 2b_1 - b_0$

$l_3 = 6b_2 - 6b_1 + b_0$

and, in general:

$$l_{r+1} = \sum_{k=0}^{r} p^*_{r,k} b_k \tag{A.2}$$

with r=0,1,...,n-1 and $p^*_{r,k} = \frac{(-1)^{r-k}(r+k)!}{(k!)^2(r-k)!}$.

In order to take into account the different nature of the data each value is weighted according to the estimation variance associated with it. In the detail, to the i-th value of the considered cored series, characterized by σ_i^2 estimation variance, is assigned a weight $w_i = \sigma^2_{max}/\sigma_i^2$, with $\sigma^2_{max} = \max(\sigma_i^2)$ for the considered series. Once defined $W_i = \sum_{k=1}^{i} w_i$, each cored series (all characterized by the same length n) acquires an effective length $m = W_n$. Concretely, the weighting procedure inserts a number of virtual ties, aimed at giving more weight to some values than to others, so the effective length of a cored series equals the sum of its weights. Considering the $y_{j:m}$ elements of the series series including the virtual ties, sorted in ascending order, the equivalent of equation A.1 for the weighted series can be written as:

$$b_r = m^{-1} \sum_{j=r+1}^{m} \frac{(j-1)(j-2)...(j-r)}{(m-1)(m-2)...(m-r)} y_{j:m} \tag{A.3}$$

with $y_{(j)}$=$x_{(i)}$ for $1 + W_{i-1} \leq j \leq W_i$.

Evaluating equation A.3, the L-moments weighted on the estimation variances can be obtained from A.2. For simplicity we report in the following the explicit form of A.3 for $r = 1, 2, 3, 4$, used in this study.

$$b_0 = \frac{1}{m} \sum_{i=1}^{n} w_i x_{(i)} \tag{A.4}$$

$$b_1 = \frac{1}{m(m-1)} \sum_{i=1}^{n} w_i x_{(i)} \Big(W_{i-1} + \frac{1}{2}(w_i - 1) \Big) \tag{A.5}$$

$$b_2 = \frac{1}{m(m-1)(m-2)} \sum_{i=1}^{n} w_i x_{(i)} \Big(\frac{1}{3} w_i^2 + w_i(W_{i-1} - 1) + \frac{2}{3} - 2W_{i-1} + W_{i-1}^2 \Big) \tag{A.6}$$

$$\begin{aligned} b_3 = \tfrac{1}{m(m-1)(m-2)(m-3)} \textstyle\sum_{i=1}^{n} \tfrac{1}{4} w_i x_{(i)} \Big(w_i^3 + w_i^2(4W_{i-1} - 6) + \\ + w_i(6W_{i-1}^2 - 18W_{i-1} + 11) + 4W_{i-1}^3 - 18W_{i-1}^2 + 22W_{i-1} - 6) \Big) \end{aligned} \tag{A.7}$$

References

[1] D. Koutsoyiannis, Advances in Urban Flood Management, Taylor and Francis, London, 2007, Ch. A critical review of probability of extreme rainfall: principles and models.

[2] M. Bernard, Formulas for rainfall intensities of long durations, ASCE 96 (1932) 592–624.

[3] C. Svensson, D. A. Jones, Review of rainfall frequency estimation methods, Journal of Flood Risk Management 3 (4) (2010) 296–313. doi:10.1111/j.1753-318X.2010.01079.x.

[4] A. Castellarin, S. Kohnová, L. Gaál, A. Fleig, J. Salinas, A. Toumazis, T. Kjeldsen, N. Macdonald, Review of applied-statistical methods for flood-frequency analysis in Europe, NERC/Centre for Ecology & Hydrology, 2012.

[5] J. Szolgay, J. Parajka, S. Kohnová, K. Hlavčová, Comparison of mapping approaches of design annual maximum daily precipitation, Atmospheric Research 92 (3) (2009) 289–307. doi:10.1016/j.atmosres.2009.01.009.

[6] D. Koutsoyiannis, D. Kozonis, A. Manetas, A mathematical framework for studying rainfall intensity-duration-frequency relationships, Journal of Hydrology 206 (12) (1998) 118 – 135. doi:10.1016/S0022-1694(98)00097-3.

[7] S. Coles, An introduction to statistical modeling of extreme values, Springer Series in Statistics, Springer-Verlag, London, 2001.

[8] R.-D. Reiss, M. s. Thomas, Statistical analysis of extreme values, from insurance, finance, hydrology and other field, Birkhaser Verlag, Basel, Boston, Berlin, 2001.

[9] C. Pappas, S. M. Papalexiou, D. Koutsoyiannis, A quick gap filling of missing hydrometeorological data, Journal of Geophysical Research: Atmospheres 119 (15) (2014) 9290–9300, 2014JD021633. doi:10.1002/2014JD021633.

[10] D. Koutsoyiannis, Statistics of extremes and estimation of extreme rainfall: I. theoretical investigation / statistiques de valeurs extrmes et estimation de prcipitations extrmes: I. recherche thorique, Hydrological Sciences Journal 49 (4) (2004) 591–610. doi:10.1623/hysj.49.4.575.54430.

[11] R. S. Teegavarapu, A. Nayak, Evaluation of long-term trends in extreme precipitation: Implications of in-filled historical data use for analysis, Journal of Hydrology 550 (2017) 616 – 634.

[12] F. Acquaotta, S. Fratianni, C. Cassardo, R. Cremonini, On the continuity and climatic variability of the meteorological stations in torino, asti, vercelli and oropa, Meteorology and atmospheric physics 103 (1-4) (2009) 279–287. doi:10.1007/s00703-008-0333-4.

[13] D. R. Maidment, et al., Handbook of hydrology., McGraw-Hill Inc., 1992.

[14] D. Koutsoyiannis, A. Langousis, 2.02 - precipitation, in: P. Wilderer (Ed.), Treatise on Water Science, Elsevier, Oxford, 2011, pp. 27 – 77. doi:10.1016/B978-0-444-53199-5.00027-0.

[15] A. A. Elshorbagy, U. Panu, S. Simonovic, Group-based estimation of missing hydrological data: I. approach and general methodology, Hydrological Sciences Journal 45 (6) (2000) 849–866.

[16] A. Elshorbagy, S. Simonovic, U. Panu, Estimation of missing streamflow data using principles of chaos theory, Journal of Hydrology 255 (1) (2002) 123–133. doi:10.1016/S0022-1694(01)00513-3.

[17] N. Alavi, J. S. Warland, A. A. Berg, Filling gaps in evapotranspiration measurements for water budget studies: evaluation of a kalman filtering approach, Agricultural and Forest Meteorology 141 (1) (2006) 57–66. doi:10.1016/j.agrformet.2006.09.011.

[18] R. S. Teegavarapu, Spatial interpolation using nonlinear mathematical programming models for estimation of missing precipitation records, Hydrological Sciences Journal 57 (3) (2012) 383–406. doi:10.1080/02626667.2012.665994.

[19] R. S. Teegavarapu, V. Chandramouli, Improved weighting methods, deterministic and stochastic data-driven models for estimation of missing precipitation records, Journal of Hydrology 312 (1) (2005) 191–206. doi:10.1016/j.jhydrol.2005.02.015.

[20] R. T. Clarke, R. D. de Paiva, C. B. Uvo, Comparison of methods for analysis of extremes when records are fragmented: A case study using amazon basin rainfall data, Journal of Hydrology 368 (14) (2009) 26 – 29. doi:10.1016/j.jhydrol.2009.01.025.

[21] M. Ashraf, J. C. Loftis, K. Hubbard, Application of geostatistics to evaluate partial weather station networks, Agricultural and forest meteorology 84 (3-4) (1997) 255–271. doi:10.1016/S0168-1923(96)02358-1.

[22] D. E. Myers, Spatial interpolation: an overview, Geoderma 62 (1) (1994) 17–28. doi:10.1.1.539.5564.

[23] J. R. M. Hosking, J. R. Wallis, Regional frequency analysis: an approach based on L-moments, Cambridge University Press, 1997.

[24] M. U. Qamar, M. Azmat, M. A. Shahid, D. Ganora, S. Ahmad, M. J. M. Cheema, M. A. Faiz, A. Sarwar, M. Shafeeque, M. I. Khan, Rainfall extremes: a novel modeling approach for regionalization, Water Resources Management 31 (6) (2017) 1975–1994.

[25] T. Buishand, Extreme rainfall estimation by combining data from several sites, Hydrological Sciences Journal 36 (4) (1991) 345–365.

[26] F. Uboldi, A. Sulis, C. Lussana, M. Cislaghi, M. Russo, A spatial bootstrap technique for parameter estimation of rainfall annual maxima distribution, Hydrology and Earth System Sciences 18 (3) (2014) 981–995. doi:10.5194/hess-18-981-2014.

[27] F. Laio, D. Ganora, P. Claps, G. Galeati, Spatially smooth regional estimation of the flood frequency curve (with uncertainty), Journal of hydrology 408 (1) (2011) 67–77.

[28] D. Ganora, F. Laio, P. Claps, An approach to propagate streamflow statistics along the river network, Hydrological sciences journal 58 (1) (2013) 41–53.

[29] M. U. Qamar, D. Ganora, P. Claps, Monthly runoff regime regionalization through dissimilarity-based methods, Water resources management 29 (13) (2015) 4735–4751.

[30] M. A. van Montfort, Sliding maxima, Journal of Hydrology 118 (1-4) (1990) 77–85.

[31] S. M. Papalexiou, Y. G. Dialynas, S. Grimaldi, Hershfield factor revisited: Correcting annual maximum precipitation, Journal of Hydrology 542 (2016) 884–895.

[32] R. S. Teegavarapu, Floods in a changing climate: extreme precipitation, Cambridge University Press, 2012.

[33] M. Haylock, N. Hofstra, A. Klein Tank, E. Klok, P. Jones, M. New, A european daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006, Journal of Geophysical Research: Atmospheres 113 (D20). doi:10.1029/2008JD010201.

[34] F. A. Isotta, C. Frei, V. Weilguni, M. Perčec Tadić, P. Lassegues, B. Rudolf, V. Pavan, C. Cacciamani, G. Antolini, S. M. Ratto, et al., The climate of daily precipitation in the alps: development and analysis of a high-resolution grid dataset from pan-alpine rain-gauge data, International Journal of Climatology 34 (5) (2014) 1657–1675. doi:10.1002/joc.3794.

[35] R. Singh, T. Wagener, K. van Werkhoven, M. E. Mann, R. Crane, A trading-space-for-time approach to probabilistic continuous streamflow

predictions in a changing climate accounting for changing watershed behavior, Hydrology and Earth System Sciences 15 (11) (2011) 3591–3603. doi:10.5194/hess-15-3591-2011.

[36] A. G. Journel, C. J. Huijbregts, Mining geostatistics, Academic press, 1978.

[37] E. H. Isaaks, M. R. Srivastava, An Introduction to Applied Geostatistics, Oxford University Press, USA, 1990.

[38] M. A. Oliver, R. Webster, Kriging: a method of interpolation for geographical information systems, International Journal of Geographical Information System 4 (3) (1990) 313–332. doi:10.1080/02693799008941549.

[39] D. L. Phillips, J. Dolph, D. Marks, A comparison of geostatistical procedures for spatial analysis of precipitation in mountainous terrain, Agricultural and Forest Meteorology 58 (1) (1992) 119–141. doi:10.1016/0168-1923(92)90114-J.

[40] C. Prudhomme, D. W. Reed, Mapping extreme rainfall in a mountainous region using geostatistical techniques: a case study in scotland, International Journal of Climatology 19 (12) (1999) 1337–1356. doi:10.1002/(SICI)1097-0088(199910)19:12¡1337::AID-JOC421¿3.0.CO;2-G.

[41] S.-H. Chua, R. L. Bras, Optimal estimators of mean areal precipitation in regions of orographic influence, Journal of Hydrology 57 (1-2) (1982) 23–48.

[42] S. L. Dingman, D. M. Seely-Reynolds, R. C. Reynolds, Application of kriging to estimating mean annual precipitation in a region of orographic influence, JAWRA Journal of the American Water Resources Association 24 (2) (1988) 329–339.

[43] P. Allamano, P. Claps, F. Laio, C. Thea, A data-based assessment of the dependence of short-duration precipitation on elevation, Physics and Chemistry of the Earth, Parts A/B/C 34 (10) (2009) 635–641. doi:10.1016/j.pce.2009.01.001.

[44] A. E. Gelfand, P. Diggle, P. Guttorp, M. Fuentes, Handbook of spatial statistics, CRC press, 2010.

[45] R. A. Olea, Geostatistics for engineers and earth scientists, Technometrics 42 (4) (2000) 444–445.

[46] U. Heinrich, Zur Methodik der räumlichen Interpolation mit geostatistischen Verfahren, Springer, 1992.

[47] M. Kolov, P. Hamouz, Approaches to the analysis of weed distribution, John Wiley and Sons, Ltd, 2016, pp. 438–453.

[48] C. Sammut, G. I. Webb, Encyclopedia of machine learning, Springer Science & Business Media, 2011.

[49] A. F. Jenkinson, The frequency distribution of the annual maximum (or minimum) values of meteorological elements, Quarterly Journal of the Royal Meteorological Society 81 (348) (1955) 158–171. doi:10.1002/qj.49708134804.

[50] J. R. M. Hosking, J. F. Wallis, Parameter and quantile estimation for the generalized pareto distribution, Technometrics 29 (3) (1987) 339–349. doi:10.2307/1269343.

[51] J. Hosking, J. Wallis, A comparison of unbiased and plotting-position estimators of l moments, Water Resources Research 31 (8) (1995) 2019–2025.

[52] S. M. Papalexiou, D. Koutsoyiannis, Battle of extreme value distributions: A global survey on extreme daily rainfall, Water Resources Research 49 (1) (2013) 187–201.

[53] The CUBIST Team, Cubist, accessed: 2016-08-01.
URL `http://www.cubist.polito.it`

[54] ARPA Piemonte, Banca dati meteorologica, accessed: 2016-08-01.
URL `http://www.regione.piemonte.it/ambiente/aria/rilev/ariaday/annali/meteoro`

[55] ARPA Lombardia, Progetto STRADA, accessed: 2016-08-01.
URL `http://idro.arpalombardia.it/pmapper-4.0/map.phtml`