A COMPARISON OF METHODS FOR THE APPROXIMATION AND ANALYSIS OF RAINFALL FIELDS IN ENVIRONMENTAL APPLICATIONS

G. Patané[a],[*] A. Cerri[a], V. Skytt[b], S. Pittaluga[a], S. Biasotti[a], D. Sobrero[a], T. Dokken[b], M. Spagnuolo[a]

[a] CNR-IMATI, Genova, Italy - (patane,cerri,pittaluga,biasotti,sobrero,spagnuolo)@ge.imati.cnr.it
[b] SINTEF, Oslo, Norway - (vibeke.skytt,tor.dokken)@sintef.no

Commission III, WG III/5

KEY WORDS: Rainfall fields, change detection, environmental analysis

ABSTRACT:

Digital environmental data are becoming commonplace and the amount of information they provide is huge, yet complex to process, due to the size, variety, and dynamic nature of the data captured by the available sensing devices. Making use of the data largely relies on the availability of efficient methods to extract meaningful information, and requires to process the environmental events at the speed data are acquired. This paper focuses on the evaluation of methods to approximate observed rain data, in real conditions of sparsity of the observations. The novelty stands in the selection of a particularly complex area, Liguria region, located in the north-west of Italy, where the orography and the closeness to the sea causes complex hydro-meteorological events. Approximation results are compared on a fine granularity in terms of cumulated rain interval used, gathered from two different rain gauge networks, with different characteristics and spatial distribution. Moreover, beside traditional cross-validation comparison, we provide a qualitative comparison based on the analysis of the number and location of maxima of the approximation. Rain maxima are indeed crucial features of rain fields needed for storm tracking, to support effective monitoring of meteorological events.

1. INTRODUCTION

The large amount of digital data provide an extremely rich, yet difficult to process, amount of information about our environment, geographic and meteorological phenomena. The geographical area selected for presenting our results, the Liguria region in Italy, is an exemplary case study: the territory is characterized by an articulated orography close to the sea, with many small catchment basins that are highly influenced by local maxima of precipitation. Moreover, the proximity to the sea causes additional problems during storms, concurring to the creation of secondary low pressure areas, also known as the Genova Low, which increases the amount of precipitation and increases the risk of critical flash floods. The continuous observation of rain data during critical events, as well as the analysis of historical time series of precipitation, are definitely crucial to support a better understanding and monitoring of hydro-geological risks, such as floods and landslides (Keefer et al., 1987, Hong et al., 2007, Wake, 2013, Hou et al., 2014).

In this context, the paper discusses the evaluation of three different approximation techniques in relation to their suitability to capture the behavior of precipitation events: LR (Locally Refinable) B-Splines and meshless approximation with kriging, and Radial Basis Functions (RBFs). The comparison of methods for rainfalls approximation has been addressed in the literature both at the theoretical level (Scheuerer et al., 2013) and for domain-specific analysis (Skok and Vrhovec, 2006). Our study contributes to this topic extending the analysis to other approximation techniques, LR B-Splines in particular, and using a new setting for the comparison, inspired by the theory of *topological persistence* (Edelsbrunner et al., 2002). The basic idea is that in order to characterize precipitation events, it is important to focus on the main features of the rainfall fields and their configuration, discarding irrelevant details that do not contribute to understanding the overall event structure. With this motivation in mind, the *prominence*

---
[*]Corresponding author

of precipitation maxima is measured through the notion of *persistence*, which allows for hierarchically organize maxima by importance, and possibly filter out irrelevant (i.e., non-prominent) ones. Based on this, we developed a criteria to compare different approximation methods, by analyzing the number and location of the most prominent maxima they produce. Finally, we want to remark that the interest here is to evaluate the performance of the approximation methods in real conditions of sparsity: the number of the measuring gauges is quite low with respect to the area covered and their distribution is quite uneven. This fact makes the experimental evaluation more interesting.

The comparative study was conducted selecting Liguria as area of interest, and the precipitation event recorded on September 29, 2013: the precipitation was characterized by light rain with 2 different thunderstorms, which caused local flooding and landslides. The observed rain data are heterogeneous both in terms of spatial distribution of the rain gauges and acquisition frequency, therefore adding a further variability that deserves analysis. Moreover, results are shown for the integration of another source of rain data, namely, extracted by radar data acquired the day of the selected event.

For the approximation of the sparse rainfall data, we considered LR B-Spline and two meshless approximation, kriging and radial basis functions. LR B-Splines are particularly useful as a compact representation of functions over large domains: they use a (locally) regular domain parameterization and can be locally refined according to the required approximation error. Ordinary kriging and RBFs use a variogram or a kernel, which are adapted to the spatial distribution of the data (e.g., through the selection of the kernel width). The three approximation methods define slightly different functions, whose behavior is studied both at the numerical level (accuracy, sensitivity to sparsness, computational issues) and at a qualitative level, by measuring the difference among the configuration of precipitation maxima induced by the three approximation techniques.

To contextualize better the comparison discussed, we start with a short overview of related work on rain observation methods, approximation and comparison techniques (Sect. 2.). We present the setting adopted for the evaluation with details on the rain event and metrics used for the comparison (Sect. 3.). Then, we give the formal definition of the three approximation methods discussed (Sect. 4.) and discuss the performances of the approximations schemes with respect to approximation, sparsity, and computational aspects (Sect. 5.). Then, the approximation schemes are compared by analyzing the difference in the configuration and prominence of the detected maxima (Sect. 6.). Finally (Sect. 7.), we summarize our study.

2. RELATED WORK

We briefly review previous work on measuring, approximating, and analyzing rainfall data and precipitation fields.

Measuring rainfall data Rainfall intensities are traditionally derived by measuring the rain rate through rain gauges, weather radar, or by measuring the variations in soil moisture with microwave satellite sensors (Brocca et al., 2014). Even though satellite precipitation analysis allows the estimation of rainfall data at a global scale and in areas where ground measures are sparse, the evaluation of light rainfalls is generally difficult, thus generating an underestimation of the cumulated rainfalls (Kucera et al., 2013). To bypass this issue, in (Brocca et al., 2014) the soil water balance equation is applied to extrapolate the daily rainfall from soil moisture data. The integration of rainfall data at regional and local levels is also intended to provide a more precise approximation of the underlying phenomenon on urban areas, which are sensitive to spatial variations in rainfalls (Segond, 2007). Furthermore, the spatial and temporal variations (e.g., speed, direction) of rainfalls are important to characterize their variability and peaks, together with their effects on catchments.

Approximating rainfall data Different approaches have been developed for the approximation of rainfall data. In (Thiessen, 1911), rainfalls recorded in the closest gauge are associated with un-sampled locations, by identifying a Voronoi diagram around each weather station and assigning the measured rainfall to the respective cell. In 1972, the U.S. National Weather Service proposed to estimate the unknown rainfall values as a weighted average of the neighboring values; the weights are the inverse of the squares of the distances between the un-sampled locations and each rainfall sample. The underlying assumption is that the samples are autocorrelated and their estimates depend on the neighboring values. This method has been extended in (Teegavarapu and Chandramouli, 2005) through the modified inverse distance and the correlation weighting method, the inverse exponential and nearest neighbor distance weighting method, and the artificial neural network estimation. In (McRobie et al., 2013), storms are modeled as clusters of Gaussian rainfall cells, where each cell is represented as an ellipse whose axis is in the direction of the movement and the rainfall intensity is a Gaussian function along each axis (Willems, 2001).

McCuen (McCuen, 1989) proposed the *isoyetal method* that allows the hydrologists to take into account the effects of different factors (e.g., elevation) on the rainfall field by drawing lines of equal rainfall depths among the rain-gauges and taking into account the main factors that influence the distribution of the rain field. Then, the rainfalls at new locations are approximated by interpolation starting from the isohyets. Geo-statistical approaches allow us to take into account the spatial correlation between neighboring samples and to predict the values at new locations (Journel and Huijbregts, 1978, Goovaerts, 1997, Goovaerts,

2000). Furthermore, the geo-statistic estimator includes additional information, such as weather-radar data (Creutin et al., 1988, Azimi-Zonooz et al., 1989) or elevation from a digital model (Goovaerts, 2000, Di Piazza et al., 2011).

Comparing rainfall data approximations For the comparison of the precipitation fields originated from different approximation schemes, we have adopted a number of standard metrics to assess the differences and performance of the schemes. Moreover, we extend the evaluation approach by comparing the differences in the configurations of meaningful features, namely prominent maxima, of the approximated fields. The motivation for this evaluation is that precipitation maxima convey important information for *storm tracking*, a crucial analysis of dynamic measures of rain data.

In storm tracking, different sets of meaningful features associated with distinct time frames, are matched to track their evolution along time. Previous work in this area focuses on the identification of regions of interest on radar images, usually characterized by high reflectivity and sufficiently large area. Various characteristics of these regions, such as centroids, area, major/minor radii, and orientation, are computed. Finally, regions are matched across the two consecutive time frames,according to the idea that the best candidate for matching minimizes some distance between the considered characteristic (Lakshmanan and Smith, 2009). For example, the TITAN algorithm (Dixon and Wiener, 1993) combines both centre of mass and area of regions for final decision of tracking. The SCIT algorithm (Johnson et al., 1998, Han et al., 2009) forecasts the centroid locations of cells at a given time: regions at the next time step are then assigned to the closest centroid location within a certain radius.

In this paper, we take inspiration from a storm tracking strategy recently proposed in (Biasotti et al., 2015), and apply it to the comparison of the different fields. The approach is based on a topological analysis of rainfall data, which focuses on the most prominent precipitation maxima instead of regions. Indeed, the granularity of the analysis is more appropriate for the characteristics of the geographic area selected; at the same time, the introduction of an ad-hoc bottleneck distance allows for matching prominent maxima of two consecutive time frames, and hence tracking their evolution along time. The same strategy for matching maxima is used to compare the configuration of maxima of different approximation results, treating them as if they were snapshots at different times.

3. CASE STUDY AND EVALUATION METRICS

The area selected for the evaluation is the Liguria region, in the north-west of Italy. Liguria can be described as a long and narrow strip of land, squeezed between the sea, the Alps and the Apennines mountains, with the watershed line running at an average altitude of about 1000 m. The orography and the closeness to the sea make this area particularly interesting for hydrometeorological events, frequently characterized by heavy rain due to Atlantic low pressure area, augmented by a secondary low pressure area creating from the Ligurian sea (Genoa Low). Moreover the several and small catchments are causing fast flooding events, and even small rivers exhibit high hydraulic energy due to the quick variation of altitude. This is the main motivation behind our analysis, which targets the understanding of the best approximation method to capture important and potentially dangerous precipitation events.

In Liguria, observed rainfall data are captured by two different rain gauges networks. The first rain gauge network is owned by

the ARPAL team of Regione Liguria, and consists of 143 professional measure stations distributed over the whole region; the measures are acquired every 5-20 minutes, and the stations are connected by GPRS and radio link connection, producing about 2 MB data per day. The second rain gauge network is owned by the Genova municipality and consists of 25 semi-professional measuring stations spread within the city boundary; the acquisitions are done every 3 minutes, and the stations are linked by GPRS or LAN connections, with an average production of 1Mb data per day. The configuration of the rain gauge networks is shown in Fig. 1.

The two rain gauge networks act as sampling devices of the true precipitation field, working at two different scales, that is, at two different spatial distributions. Since the temporal interval is different for each network, we have cumulated the station rainfalls to a step of 30 minutes. This selection is also motivated by the desire to produce a fine-grained evaluation of the approximation methods in the perspective of a real-time precipitation monitoring. Note that the cumulated interval is a much smaller than the one used in (Skok and Vrhovec, 2006), where an interval of 24 hours was used. Concerning the precipitation events, we selected a rainy day, September 29, 2013, which was characterized by light rain over Liguria with 2 different thunderstorms that caused local flooding and landslides, without damages. The maximum rain-rate over all time step is $60mm/30'$ and the average rain-rate is $1.12mm/30'$.

To establish a formal evaluation setting, let us formulate the problem of rainfall approximation as follows. Given a set of points $\mathcal{P} := \{\mathbf{p}_i\}_{i=1}^n$, which represent the position of the measurement instruments, let $f : \mathcal{P} \to \mathbb{R}$ be the precipitation field, measured at the n sample points. An approximation of f is defined as $F : \mathbb{R}^2 \to \mathbb{R}$ such that $d(F(\mathbf{p}) - f(\mathbf{p})) \leq \epsilon$ for some required distance $d(\cdot, \cdot)$ and threshold ϵ. When $d(F(\mathbf{p}) - f(\mathbf{p})) = 0$ the approximation is an interpolation of f. The map F can be used to evaluate the value of the precipitation at any point other then those in $\mathcal{P}$, with results differing according to the approach used to define F. In our case, we will consider three different F approximation functions. The color coding used for the illustration of the computed approximation ranges from blue (i.e., smallest rain rate) to brown (i.e., highest rain rate), passing through green, yellow, and orange (i.e., intermediate values).

To compare the approximations, we adopt a cross-validation strategy, implemented in two ways. First, every rainfall station at $\mathbf{p}_i$ is iteratively turned off, that is, it is not used in the computation of F; the approximation function F obtained is sampled at that position $\mathbf{p}_i$ and compared with the rain value measured at $\mathbf{p}_i$, which acts as a ground truth (leave-one-out strategy). Second, the rainfall data measured by the municipality stations are used as ground-truth to validate the values approximated from the ARPAL data set: in this setting, the cross-validation aims at evaluating the capability of the different methods to estimate the local features of rain fields interpolated over a sparse data set, with different spatial distribution.

The comparative study also includes the analysis of the spatial configuration of local maxima extracted from the rainfall fields produced by each approximation scheme. In this case, local maxima are endowed with a notion of prominence borrowed from topological persistence, which is used to quantify the importance that a maximum has in characterizing the associated rainfall field. This comparative analysis is motivated by the fact that, in order to understand the evolution of precipitations and tracking their changing along time, local maxima and their configuration provide a useful description for capturing the important elements of
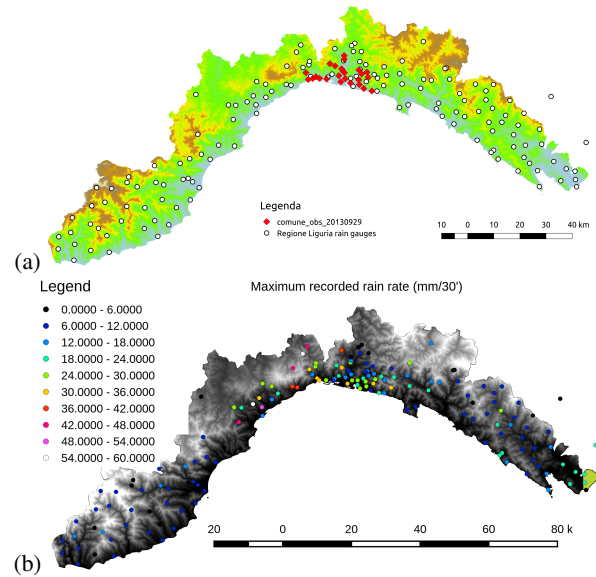

Figure 1: (a) Input rainfall measures at 143 stations (regional level, white points) and 25 stations (municipality level, red circles). (b) Map of the rain rate maximum recorded at each weather station, which highlights that only the central west of the region has been involved by heavy rain and the remaining part were interested by drizzle.

the underlying rainfall field. Indeed, they have a relevant semantic content and, at the same time, are formally well-defined. For this set of experiments, the approximated rainfalls were sampled at the vertices of a digital terrain model, producing a discretization of the precipitation field whose maxima were compared. The DTM used is coming from the SRTM (Shuttle Radar Topography Mission (Farr et al., 2007)), available in public domain at the URL `http://www2.jpl.nasa.gov/srtm/`, and with a spatial resolution of 100 mt.

4. THEORETICAL BACKGROUND

In the following, we give an overview of the three approximation methods compared and of the persistence analysis framework used to analyze the evolution of precipitations (Fig. 2).

4.1 Approximation schemes

LR B-Splines The rainfall values are parameterized on the xy-values of the corresponding geographic location and the rainfall is approximated by a 2.5D LR B-spline surface (Dokken et al., 2013). The approximation of the rainfall data is performed by an iterative procedure starting from a lean tensor-product B-spline surface being constantly equal to zero. For each iteration the distance between the current surface and the rainfall data is computed, the surface is refined locally where a given tolerance is not met, and the surface coefficients are updated using *Multilevel B-spline approximation (MBA)* (Lee et al., 1997) adapted for LR B-splines. The MBA method is a local and explicit approximation method, where the surface coefficients are updated based on the data points situated in the support of the corresponding B-spline. The performance depends on three components, which are done at each iteration step: the refinement of the LR B-spline, distance computations, and update of the surface coefficients. The latter two elements are the most time consuming. For each iteration, the coefficients are updated twice and one additional distance computation is performed. Let the number of data points be N. The number of non-zero B-splines for each data point varies, but will
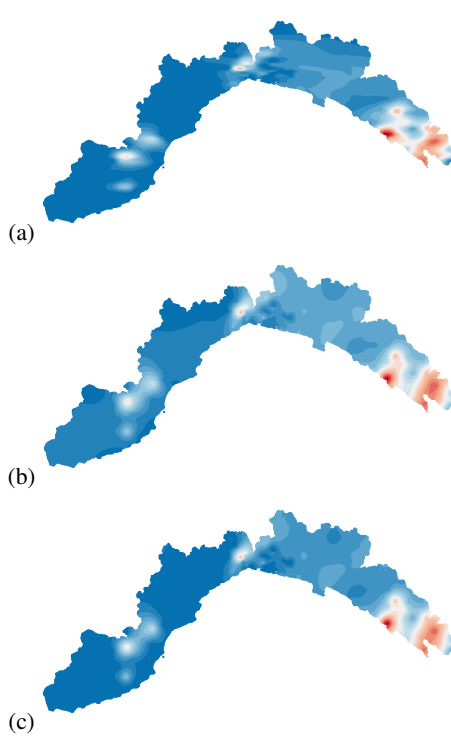
(a)

(b)

(c)

Figure 2: Rainfall fields computed for the event of September, 23: (a) LR B-Spline, (b) RBF, (c) Kriging.

Figure 3: A function $F : \mathcal{M} \to \mathbb{R}$, color-coded from blue (low) to red (high) values, and the associated local maxima having persistence greater than $\alpha(\max F - \min F)$, with $\alpha = 0.05$, 0.15 (middle) and 0.25.

be in the magnitude of $(d_1 + 1) \times (d_2 + 1)$ where d_1 and d_2 are the polynomial degrees in the two parameter directions of the surface. The surface is bi-quadratic so $d_1 = d_2 = 2$. In our tests, the algorithm is run with 20 iterations giving a total of $3 \times 20 \times N \times 9$ bi-variate B-spline evaluations.

Implicit approximation with radial basis functions Implicit approximation computes the map $F(\mathbf{p}) := \sum_{i=1}^{n} \alpha_i \varphi_i(\mathbf{p})$ as a linear combination of the basis $\mathcal{B} := \{\varphi_i(\mathbf{p}) := \varphi(\|\mathbf{p} - \mathbf{p}_i\|_2)\}_{i=1}^{n}$, where φ is the kernel function (Aronszajn, 1950, Dyn et al., 1986, Micchelli, 1986, Patanè et al., 2009). Depending on the properties of φ, we distinguish globally- (Carr et al., 2001, Turk and O'Brien, 2002) and compactly-supported (Wendland, 1995, Morse et al., 2001) supported radial basis functions. Then, the coefficients $(\alpha_i)_{i=1}^{n}$ solve a $n \times n$ linear system, which is achieve by imposing the interpolating constraints $F(\mathbf{p}_i) = f(\mathbf{p}_i)$, $i = 1, \ldots, n$. Since a $n \times n$ linear system is solved once, the computational cost of the approximation with globally- and locally-supported RBFs is $O(n^3)$ and $O(n \log n)$, respectively. In our experiments, we have chosen the Gaussian kernel $\varphi(st) := \exp(-st)$, which has a global support; in fact, its fast decay makes it suitable to approximate rainfalls with a sparse spatial distribution that change quickly in time. To this end, the width of each basis function is automatically adapted to the local sampling density by selecting its width according to the local spatial distribution of the rainfall stations (Dey and Sun, 2005, Mitra and Nguyen, 2003).

Kriging The previous two approximation methods do not take into account in an explicit manner; correlation among observations may have unwanted effects especially in the case of unevenly distributed observations. Furthermore, there is no natural mechanism for propagating the individual quality of the observations into a quality description of the estimation. A class of methods that takes care of these issues is kriging, (Wackernagel, 2003), which is a common technique in environmental sciences and a special case of the maximum likelihood estimation. The underlying assumptions are that the quality of the observations
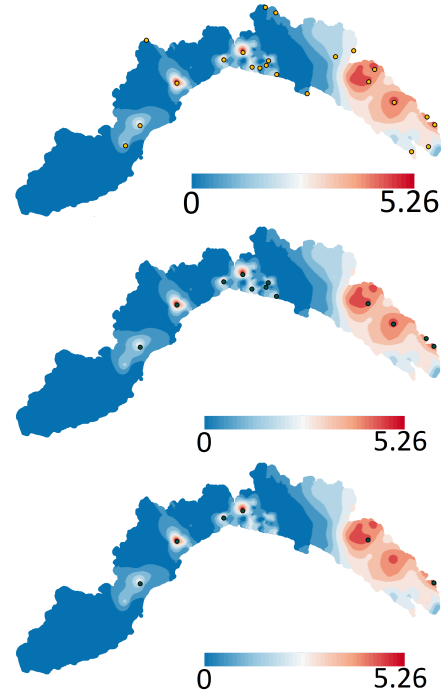
is given as variance values, and that the covariance between observations only depends on their mutual spatial or temporal distance, and not on their location. Formally, kriging is expressed as $F(\mathbf{p}) := \sum_{i=1}^{n} \omega_i f(\mathbf{p}_i)$, where the weights ω_i are defined as $\Omega = \mathbf{C}^{-1} \times \mathbf{D}$ with $\mathbf{C}$ the covariance matrix of the measured values and $\mathbf{D}$ is the matrix defined by the covariance between the known values and the points to be estimated. The covariance is expressed by the variogram model, which reflects the priors on the spatial variability of the values. The main problem with kriging is the low computational efficiency, as the solution of the linear systems scales quadratically with the number of observations. In the implementation used, the problem is addressed by combining kriging with deterministic spatial division techniques, which efficiently restrict the number of observations to the closest ones. More specifically, the Kd-tree is used to select only the 20 closest neighbors for the matrix inversion.

4.2 Prominent rainfall maxima *via* persistence analysis

The importance of precipitation maxima is evaluated by means of the *persistence analysis*. Given a scalar field $F : \mathcal{M} \to \mathbb{R}$ (e.g., the interpolated rainfall field) and sweeping t in $\mathbb{R}$, new connected components of the superlevel sets $\mathcal{M}^t = \{\mathbf{p} \in \mathcal{M} : F(\mathbf{p}) \geq t\}$ are either born, or previously existing ones are merged together. A connected component is associated with a local maximum $\mathbf{p}$ of F, where the component is first born. When two components corresponding to local maxima $\mathbf{p}_1, \mathbf{p}_2, F(\mathbf{p}_1) < F(\mathbf{p}_2)$, merge together the component corresponding to $\mathbf{p}_1$ dies. In this case, the component associated with the smaller local maximum is merged into that associated with the larger one. Each local maximum $\mathbf{p}$ of F is associated with its a *persistence value* $\text{pers}_F(\mathbf{p})$, which is defined as the difference between the birth and the death level of the corresponding connected component. Maxima associated with a higher persistence value identify relevant features and structures of the underlying phenomena.

To compute the local maxima and the associated persistence val-

Method	Max	Mean	Median	Std. dev.	MSE
Method	[mm]	[mm]	[mm]	[mm]	[mm^2]
Ord. krig.	32.44 (54.1%)	0.02	5.85E-05	2.38	5.64
RBFs	37.80 (63.0%)	0.97	0.34	2.12	5.44
LR B-Splines	27.2 (45.3%)	-0.04	1.20E-5	2.73	7.05

Table 1: Summary statistics for the error distribution of the cross validation.
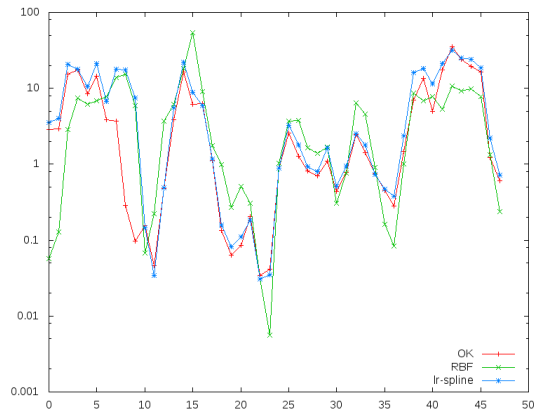

Figure 4: Leave-one-out cross validation MSE: y-axis reports the MSE [mm^2] for each time step x-axis.

ues, F is interpolated on the vertices of a triangle mesh $\mathcal{M}$. The points of $\mathcal{M}$ are first sorted in decreasing values, from $\max F$ to $\min F$; then, the classical 0th-persistence algorithm (Edelsbrunner et al., 2002, Edelsbrunner and Harer, 2010) is used. The cost of sorting the n points of $\mathcal{M}$ is $O(n \log n)$; by using a union-find data structure, the persistence algorithm requires linear storage and running time at most proportional to $O(m\alpha(m))$, where m is the number of edges in the mesh and $\alpha(\cdot)$ is the inverse of the Ackermann function. An example for the extraction of local maxima at three different persistence levels is given in Fig. 3.

5. APPROXIMATION BEHAVIOR

The first set of results we discuss is related to the comparison of the behavior of the three methods according to approximation performance and computational complexity. Concerning the leave-one-out cross-validation strategy, we have checked the results by computing the three approximation fields turning off, iteratively, each rainfall station at $\mathbf{p}_i$, for each cumulated interval. The value of the approximation function F obtained was then compared at $\mathbf{p}_i$ with the rain value measured by the corresponding rain gauge at $\mathbf{p}_i$, acting as a ground truth. The statistics of the evaluation are shown in Table 1; in this case, ordinary kriging and LR B-Splines have the smaller maximum error, but the RBFs have a smaller mean-squares error and standard devitation. In Fig. 4, the plot of the MSE distribution for the three methods is shown, per each time interval.

The second set of results concerns the cross-validation done using the rainfall data measured by the municipality stations as ground-truth to validate the values that approximate only the ARPAL data set. This validation aims at gathering indicators on the behavior, in terms of accuracy, on different spatial distribution of the sample points. This approach is meaningful as the two observation networks cover an overlapping region of the study area. The network from Genova municipality is located within the boundary of the city and is denser than the ARPAL one, which covers the whole study area, and some of the ARPAL stations are located in the Genova municipality. Comparing the approximation results at these two scales, we have evaluated the sensitivity of the approximation to local distributions of the samples and the capability to estimate the local features of rain fields interpolated over
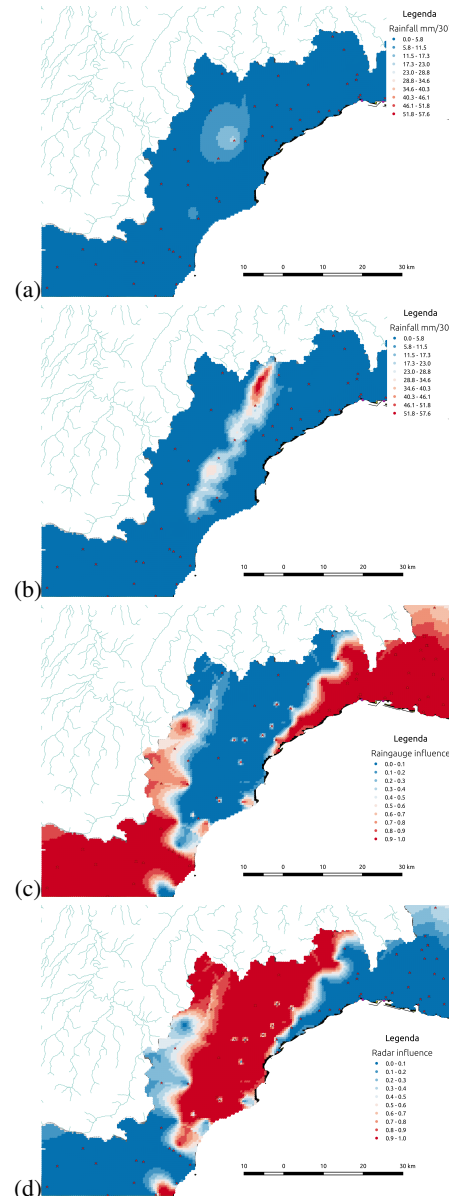

(a)
(b)
(c)
(d)

Figure 5: Ordinary kriging approximation of rainfalls computed with (a) rain gauges and (b) integrated with radar measurements. (c) Rain gauges weights and (d) radar data set mapped in (b).

a sparser data set. The results are shown in Table 2: in this case, ordinary kriging and LR B-Splines have the smaller maximum error, but the RBFs have a smaller mean-squares error.

Finally, concerning computational complexity, the different algorithms have been tested over a 64 bits workstation 8 cores at 1.6GHZ and RAM of 16 GB. The system runs an Ubuntu 14.04LTS with 3.13.0 kernel. The run of LR B-Splines takes 19.33 seconds to compute the approximation over the whole region (20K points) for the 48 time intervals. For the same task, the ordinary kriging takes 1.746 seconds and RBFs approximation takes 6.23 seconds. One important point to make here is that, for all the methods, the computational complexity and the timing collected are well below the time interval analyzed (30min). This important characteristic tells us that we could use any of them for real-time monitoring of the rain events. The analysis carried until now does not tell us much about the scalability of the methods for a larger set of observation points, where the computational complexity could become an issue. Preliminary results of this situation are presented later on, in Fig. 5, where results of kriging

Method	Max [mm]	Mean [mm]	Median [mm]	Std. dev. [mm]	MSE [mm^2]
Ord. krig.	28.62 (47.7%)	0.59	3.26E-3	4.45	20.21
RBFs	36.77 (61.2%)	1.41	0.44	3.25	12.58
LR B-Splines	30.39 (50.6%)	0.59	3.71E-3	4.45	20.19

Table 2: Summary statistics for the error distribution of the accuracy evaluation at different scales.

Table 3: Statistics for the average number of extracted persistent maxima.

Method	$\tau = 0.05$	$\tau = 0.15$	$\tau = 0.25$	$\tau = 0.35$
Ord. krig.	7.4375	3.4792	2.3334	1.7292
RBFs	6.8750	3.4	2.4250	1.7250
LR B-Splines	7.6667	3.7917	2.5208	1.8333

obtained integrating radar rain data are presented.

6. PERSISTENT RAINFALL MAXIMA

Tables 3 and 4 report the comparative results about the extraction of persistent maxima when considering the rainfall fields produced by the three approximation schema. For our tests, rainfall fields have been computed based on the data set described in Sect. 3.. Hence, for each approximation scheme, we considered the 48 approximated fields, one for each cumulative step. For each field F, the associated persistence maxima have been extracted according to four different values for the persistence threshold ε, namely $\varepsilon = \tau(\max F - \min F)$ with $\tau = 0.05$, 0.15, 0.25, 0.35. In practice, a maximum is preserved only if its persistence is larger than ε, while the others are filtered away. Table 3 reports the total number of extracted persistent maxima, averaged by the amount of considered cumulative steps; Table 4 shows the maximum number of local maxima that have been extracted, method by method, from the 48 rainfall fields. Despite some slight differences in the results, the general trend is to have a decreasing number of persistent maxima as the threshold τ increases. This is actually not surprising, since a higher persistence threshold implies that a larger portion of local maxima are pruned out. Also, for low values of the persistence threshold, we can relate the number of detected maxima to the smoothness of the considered approximation: in this view, the RBF schema appears to have a higher smoothing effect, as indicated by the smaller number of maxima characterized by a low persistence value.

6.1 Comparing sets of persistent maxima

In order to refine the above comparative analysis, we make use of the tracking procedure introduced in (Biasotti et al., 2015) to quantitatively assess a (dis)similarity measure between two sets of local maxima, originated from the three approximation schema when considering the same cumulative step. Before presenting results, we briefly recall the main ideas of (Biasotti et al., 2015).

For two sets $\mathcal{F}, \mathcal{G}$ of local maxima of two rainfall fields $F, G : \mathcal{M} \to \mathbb{R}$, it is possible to compare them by measuring the cost of moving the points associated with one function to those of the other one, with the requirement that the longest of the transportations should be as short as possible. Interpreting the local maxima in $\mathcal{F}$ and $\mathcal{G}$ as points in $\mathbb{R}^3$ (i.e., geographical position and persistence value), the collections of local maxima are compared through the *bottleneck distance* between $\mathcal{F}$ and $\mathcal{G}$, which is defined as

$$d_B(\mathcal{F}, \mathcal{G}) = \inf_{\gamma} \sup_{\mathbf{p}} d(\mathbf{p}, \gamma(\mathbf{p})),$$

where $\mathbf{p} \in \mathcal{F}$, γ ranges over all the bijections between $\mathcal{F}$ and $\mathcal{G}$, $d(\cdot, \cdot)$ is the *pseudo-distance*

$$d(\mathbf{p}, \mathbf{q}) := \min\{\|\mathbf{p} - \mathbf{q}\|, \max\{\text{pers}_F(\mathbf{p}), \text{pers}_G(\mathbf{q})\}\},$$

which measures the cost of moving $\mathbf{p}$ to $\mathbf{q}$, and $\|\cdot\|$ is a weighted modification of the Euclidean distance. In practice, the cost of

Table 4: Statistics for the maximum number of extracted persistent maxima.

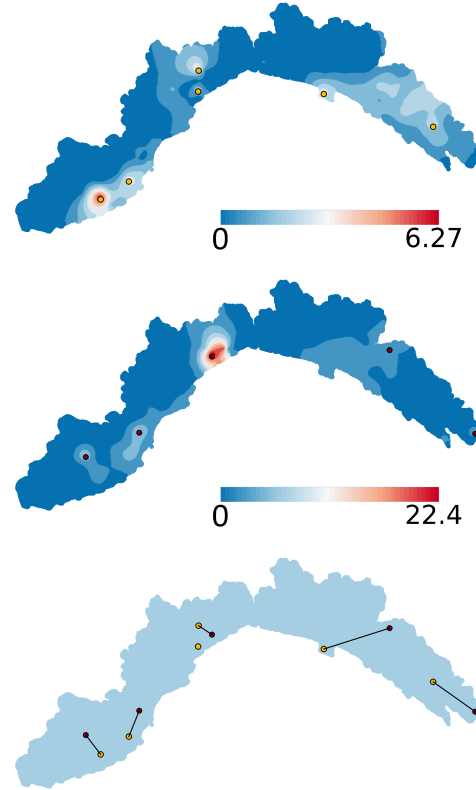| Method	$\tau = 0.05$	$\tau = 0.15$	$\tau = 0.25$	$\tau = 0.35$
Ord. krig.	20	12	8	5
RBFs	18	10	8	5
LR B-Splines	21	10	8	5

Figure 6: Two fields $F, G : \mathcal{M} \to \mathbb{R}$, color-coded from blue (low) to red (high) values, and the associated local maxima. On the right, bottleneck matching between local maxima.

taking $\mathbf{p}$ to $\mathbf{q}$ is measured as the minimum between the cost of moving one point onto the other and the cost of moving both points onto the plane $xy : z = 0$. Matching a point $\mathbf{p}$ with a point of xy, which can be interpreted as the annihilation of $\mathbf{p}$, is allowed by the fact that the number of points for $\mathcal{F}$ and $\mathcal{G}$ is usually different. The matching γ between the points of $\mathcal{F}$ and those of $\mathcal{G}$, for which d_B is actually occurred, is referred to as a *bottleneck matching* (Fig. 6). Through the bottleneck matching and the bottleneck distance, it is then possible to derive quantitative information about the differences in the spatial arrangement and the rain measurements for the points in $\mathcal{F}$ and $\mathcal{G}$.

The bottleneck distance can be evaluated by applying a pure graph-theoretic approach or by taking into account geometric information that characterize the assignment problem. We opt for a graph-theoretic approach, which is independent of any geometric constraint and our implementation is based on the push-relabel maximum flow algorithm (Cherkassky and Goldberg, 1997). For each iteration, the algorithm runs in $O(k^{2.5})$, where k is the number of local maxima involved in the comparison. We note that the computational complexity is not an issue, because the number of points to be considered is very limited in general. For example, in tracking applications the number of persistent maxima to be monitored is usually no more than a dozen for each time sample.

Experimental results For each cumulative step, we consider the rainfall fields interpolated by the three methods, and extract

Table 5: Table of results for the averaged geographical distance (measured in km) between sets of local maxima (Liguria area size: 5.410 km^2).

Method1/Method2	$\tau = 0.05$	$\tau = 0.15$	$\tau = 0.25$	$\tau = 0.35$
Ord. krig / RBFs.	15.514 km	2.798 km	2.768 km	2.685 km
RBFs / LR B Spline	35.08 km	10.30 km	5.71 km	4.054 km
krig / LR B-Splines	35.36 km	6.528 km	3.388 km	2.326 km

Table 6: Table of results for the averaged rainfall distance (measured in mm) between sets of local maxima.

Method1/Method2	$\tau = 0.05$	$\tau = 0.15$	$\tau = 0.25$	$\tau = 0.35$
Ord. krig/RBFs.	4.127 mm	2.438 mm	1.900 mm	1.884 mm
RBFs/LR B Sp.	9.029 mm	5.111 mm	3.930 mm	3.960 mm
Krig./LR B-Sp.	6.707 mm	3.345 mm	2.105 mm	1.991 mm

the sets of local maxima according to the four persistence thresholds discussed above. For each threshold, the three collections of persistent maxima are pairwise compared as follows. Since geographic coordinates and rainfall measurements come with different reference frames and at different scales, local maxima to be matched are first normalized so that their coordinates range in [0,1]; then, they are processed by computing the associated bottleneck matching and the bottleneck distance, and afterwards projected back in the original reference frames. Finally, a measure of their distance in terms of both geographical coordinates and rainfall values is derived by combining the information contained in the bottleneck matching and the associated numerical (dis)similarity score. Precisely, we consider the *geographical* and *rainfall distances*, which are defined as the largest difference in geographical position and rainfall value, respectively, for two persistent maxima that have been paired by the bottleneck matching.

Tables 5 and 6 report the obtained results, in terms of geographical and rainfall distances, respectively, averaged by the total number of considered cumulative steps. To have a clearer picture of the comparative evaluation in terms of the two distances, the results included in Tables 5 and 6 should be jointly interpreted for each persistence threshold. For instance, when $\tau = 0.05$ we have (relatively) high values for the geographical distance together with quite low rainfall distance values: this can be interpreted as slight numerical variations for the three approximations, possibly appearing spatially far one from each other. Note, however, that in this view that the RBF and the kriging techniques appear to have a more similar behavior, both producing higher values for the geographical and rainfall distances when compared with LR-B Splines. On the other hand, moving to higher persistence thresholds, the values of geographical distance decrease, as an effect of filtering out non-relevant maxima. As a consequence, the corresponding rainfall distance values reveal now the differences occurring at prominent maxima, which appear to be quite small.

We conclude by proposing in Table 7 a similar analysis to compare the results that are obtained when rainfall fields are interpolated by considering either observed rainfall measurements or an integration of these data with radar acquisitions (Sect. 5.). Indeed, integrated data can reveal useful, e.g., for tracking applications: although rainfall measurements are more reliable, integrating them with radar data makes it possible to extend the rainfall field interpolation in larger areas and to have a clearer picture about the temporal evolution of the associated precipitation event. As can be seen from the results in Table 7, characterized by high values in both the geographic and the rainfall distance, considering radar data can sensibly change the spatial location and the rainfall value of persistent maxima. This can be interpreted as the introduction of complementary information with respect to rainfall measurements, which hopefully can help in having a clearer understanding of precipitation events.

Table 7: Averaged geographical (measured in km) and rainfall distance (measured in mm) between sets of local maxima.

Krig/(Radar + Krig)	$\tau = 0.05$	$\tau = 0.15$	$\tau = 0.25$	$\tau = 0.35$
Geogr. dist.	45.453 km	23.387 km	18.568 km	13.28 km
Rainfall dist.	23.97 mm	21.227 mm	18.533 mm	16.396 mm

7. CONCLUSIONS AND FUTURE WORK

The aim of this study was the comparison of different spatial approximation methods finalized to compute the amount of rainfalls for hydro-metereological analysis and civil protection. As a final remark, we point out that all these approaches easily support the integration of further sources of rain measures, for instance those captured by radar. Fig. 5 shows the results of estimation achieved by kriging (a) with the rain gauges only and (b) with the integration of this data with radar information. In (c,d), we show the contribution of each data set in the estimated map plotted in (b). The color scale varies from blue to red corresponding respectively to a contribution varying from null to full. Since we have selected only two different data sets, the two maps in (c,d) are complementary.

Finally, we plan to proceed further with the presented comparison framework, including several more aspects and extending the evaluation to more elaborate correlation analysis, taking into account other relevant data, such as terrain morphology, satellite imagery, and meteorological situation. We will further investigate this possibility and especially the effect on approximation results on storm tracking.

REFERENCES

Aronszajn, N., 1950. Theory of reproducing kernels. Trans. of the American Mathematical Society 68, pp. 337–404.

Azimi-Zonooz, A., Krajewski, W., Bowles, D. and Seo, D., 1989. Spatial rainfall estimation by linear and non-linear co-kriging of radar-rainfall and raingage data. Stochastic Hydrology and Hydraulics 3(1), pp. 51–67.

Biasotti, S., Cerri, A., Pittaluga, S., Sobrero, D. and Spagnuolo, M., 2015. Persistence-based tracking of rainfall field maxima. Topology-Based Methods in Visualization 2015, In press.

Brocca, L., Ciabatta, L., Massari, C., Moramarco, T., Hahn, S., Hasenauer, S., Kidd, R., Dorigo, W., Wagner, W. and Levizzani, V., 2014. Soil as a natural rain gauge: Estimating global rainfall from satellite soil moisture data. Journal of Geophysical Research: Atmospheres 119(9), pp. 5128–5141.

Carr, J. C., Beatson, R. K., Cherrie, J. B., Mitchell, T. J., Fright, W. R., McCallum, B. C. and Evans, T. R., 2001. Reconstruction and representation of 3D objects with radial basis functions. In: ACM Siggraph, pp. 67–76.

Cherkassky, B. V. and Goldberg, A. V., 1997. On Implementing the Push-Relabel Method for the Maximum Flow Problem. Algorithmica 19(4), pp. 390–410.

Creutin, J., Delrieu, G. and Lebel, T., 1988. Rain measurement by raingage-radar combination: a geostatistical approach. Journal of Atmospheric and Oceanic Technology 5(1), pp. 102–115.

Dey, T. K. and Sun, J., 2005. An adaptive MLS surface for reconstruction with guarantees. In: ACM Symposium on Geometry Processing, pp. 43–52.

Di Piazza, A., Conti, F., Noto, L., Viola, F. and La Loggia, G., 2011. Comparative analysis of different techniques for spatial interpolation of rainfall data to create a serially complete monthly time series of precipitation for sicily, italy. International Journal of Applied Earth Observation and Geoinformation 13(3), pp. 396–408.

Dixon, M. and Wiener, G., 1993. Titan: Thunderstorm identification, tracking, analysis, and nowcasting–a radar-based methodology. Journal of Atmospheric and Oceanic Technology 10(6), pp. 785–797.

Dokken, T., Lyche, T. and Pettersen, K. F., 2013. Polynomial splines over locally refined box-partitions. Computer Aided Geometric Design 30(3), pp. 331–356.

Dyn, N., Levin, D. and Rippa, S., 1986. Numerical procedures for surface fitting of scattered data by radial functions. SIAM Journal on Scientific and Statistical Computing 7(2), pp. 639–659.

Edelsbrunner, H. and Harer, J., 2010. Computational Topology: An Introduction. American Mathematical Society.

Edelsbrunner, H., Letscher, D. and Zomorodian, A., 2002. Topological persistence and simplification. Discrete Computational Geometry 28, pp. 511–533.

Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E. and Roth, L., 2007. The shuttle radar topography mission. Reviews of Geophysics 45(2), pp. 361–363.

Goovaerts, P., 1997. Geostatistics for natural resources evaluation. Oxford University Press.

Goovaerts, P., 2000. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. Journal of Hydrology 228(1), pp. 113–129.

Han, L., Fu, S., Zhao, L., Zheng, Y., Wang, H. and Lin, Y., 2009. 3D Convective Storm Identification, Tracking, and Forecasting–An Enhanced TITAN Algorithm. Journal of Atmospheric and Oceanic Technology 26(4), pp. 719–732.

Hong, Y., Adler, R. and Huffman, G., 2007. An experimental global prediction system for rainfall-triggered landslides using satellite remote sensing and geospatial datasets. IEEE Trans. on Geoscience and Remote Sensing 45(6), pp. 1671–1680.

Hou, A. Y., Ramesh, K. K., Neeck, S., Azarbarzin, A. A., Kummerow, C. D., Kojima, M., Oki, R., Nakamura, K. and Iguchi, T., 2014. The global precipitation measurement mission. Bulletin of the American Meteorological Society.

Johnson, J. T., MacKeen, P. L., Witt, A., Mitchell, Stumpf, G. J., Eilts, M. D. and Thomas, K. W., 1998. The Storm Cell Identification and Tracking Algorithm: An Enhanced WSR-88D Algorithm. Wea. Forecasting 13(2), pp. 263–276.

Journel, A. G. and Huijbregts, C. J., 1978. Mining geostatistics. Academic press.

Keefer, D., Wilson, R., Mark, R., Brabb, E., Brown, W., Ellen, S., Harp, E., Wieczorek, G., Alger, C. and Zatkin, R., 1987. Real-time landslide warning during heavy rainfall. Science 238, pp. 921–925.

Kucera, P. A., Ebert, E. E., Turk, F. J., Levizzani, V., Kirschbaum, D., Tapiador, F. J., Loew, A. and Borsche, M., 2013. Precipitation from space: advancing earth system science. Bulletin of the American Meteorological Society 94, pp. 365–375.

Lakshmanan, V. and Smith, T., 2009. An Objective Method of Evaluating and Devising Storm-Tracking Algorithms. Wea. Forecasting 25(2), pp. 701–709.

Lee, S., Wolberg, G. and Shin, S. Y., 1997. Scattered data interpolation with multilevel b-splines. IEEE Trans. on Visualization and Computer Graphics 3(3), pp. 228–244.

McCuen, R. H., 1989. Hydrologic analysis and design. Prentice-Hall Englewood Cliffs, NJ.

McRobie, F. H., Wang, L.-P., Onof, C. and Kenney, S., 2013. A spatial-temporal rainfall generator for urban drainage design. Water Science and Technologies 68(1), pp. 240–249.

Micchelli, C. A., 1986. Interpolation of scattered data: Distance matrices and conditionally positive definite functions. Constructive Approximation 2, pp. 11–22.

Mitra, N. J. and Nguyen, A., 2003. Estimating surface normals in noisy point cloud data. In: Proc. of Computational Geometry, ACM Press, pp. 322–328.

Morse, B. S., Yoo, T. S., Chen, D. T., Rheingans, P. and Subramanian, K. R., 2001. Interpolating implicit surfaces from scattered surface data using compactly supported radial basis functions. In: IEEE Shape Modeling and Applications, pp. 89–98.

Patanè, G., Spagnuolo, M. and Falcidieno, B., 2009. Topology- and error-driven extension of scalar functions from surfaces to volumes. ACM Trans. on Graphics 29(1), pp. 1–20.

Scheuerer, M., Schaback, R. and Schlather, M., 2013. Interpolation of spatial data - a stochastic or a deterministic approach? European Journal of Applied Mathematics 24, pp. 601–629.

Segond, M.-L., 2007. The significance of spatial rainfall representation for flood runoff estimation: A numerical evaluation based on the Lee catchment. Journal of Hydrology 347, pp. 116 – 131.

Skok, G. and Vrhovec, T., 2006. Considerations for interpolating rain gauge precipitation onto a regular grid. Meteorologische Zeitschrift 15(5), pp. 545–550.

Teegavarapu, R. S. V. and Chandramouli, V., 2005. Improved weighting methods, deterministic and stochastic data-driven models for estimation of missing precipitation records. Journal of Hydrology 312(1), pp. 191–206.

Thiessen, A. H., 1911. Precipitation averages for large areas. Monthly weather review 39(7), pp. 1082–1089.

Turk, G. and O'Brien, J. F., 2002. Modelling with implicit surfaces that interpolate. ACM Siggraph 21(4), pp. 855–873.

Wackernagel, H., 2003. Multivariate geostatistics: an introduction with applications. 2nd edn, Springer-Verlag.

Wake, B., 2013. Flooding costs. Nature Climate Change 3(9), pp. 1671–1680.

Wendland, H., 1995. Real piecewise polynomial, positive definite and compactly supported radial functions of minimal degree. Advances in Computational Mathematics 4(4), pp. 389–396.

Willems, P., 2001. A spatial rainfall generator for small spatial scales. Journal of Hydrology 252, pp. 126 – 144.