Comparing Methods for the Approximation of Rainfall Fields in Environmental Applications

G. Patané[a1], 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*

Abstract

Digital environmental data are becoming commonplace and the amount of information they provide is complex to process, due to the size, variety, and dynamic nature of the data captured by sensing devices. The paper discusses an evaluation framework for comparing methods to approximate observed rain data, in real conditions of sparsity of the observations. The novelty brought by this experimental study stands in the geographical area and heterogeneity of the data used for evaluation, aspects which challenge all approximation methods. The Liguria region, located in the north-west of Italy, is a complex area for the orography and the closeness to the sea, which cause complex hydro-meteorological events. The observed rain data are highly heterogeneous: two data sets come from measured rain gathered from two different rain gauge networks, with different characteristics and spatial distributions over the Liguria region; the third data set come from weather radar, with a more regular coverage of the same region but a different veracity. Finally, another novelty of the paper is brought by the proposal of an application-oriented perspective on the comparison. The approximation models the rain field, whose maxima and their evolution is essential for an effective monitoring of meteorological events. Therefore, we adapt a storm tracking technique to the analysis of the displacement of maxima computed by the different methods, used as a dissimilarity measure among the approximation methods analyzed.

1. Introduction

The large amount of digital data provides 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 articulated orography is characterized by 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 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). A robust approximation method, resilient to errors, is therefore highly desirable.

In this context, the paper presents the results of the evaluation of six approximation techniques, which give insights into their suitability to capture the behavior of precipitation events: the nearest neighbor method, the piecewise linear approximation with barycentric coordinates, the inverse distance weighting, kriging, the Locally Refinable (LR) B-Splines, and the Radial Basis Functions (RBFs). The comparison of methods for rainfall 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 more approximation techniques, such as the LR B-Splines, 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. With this motivation in mind, the *prominence* of precipitation maxima is measured through the notion of *persistence*, which allows for hierarchically organize maxima by importance, and possibly filter out irrelevant ones. Based on this, we developed an approach to compare different approximation methods based on the analysis of the number and location of the most prominent maxima they produce.

Our focus is on the evaluation of approximation performance 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. The evaluation results give also insights on the influence of integrating radar data in the approximation: rain data extracted from radar measures provide a complementary information with respect to rain gauges, less accurate but with a wider and more stable coverage. The integration of measured rain and radar data gives insights on the reliance on radar-driven approximations in case of failures of some rainfall stations during heavy storms.

For this study, we considered some of the best known methods in this field (nearest neighbor, piecewise linear, inverse distance weighting, kriging) and some other methods that have been applied mostly in the field of computer graphics, but have interesting properties for this application (LR B-Splines and RBFs).

The nearest neighbor method provides a rough approximation by defining the rainfall approximation at a point as equal to the rainfall value measured by its closest rainfall station. Barycentric coordinates are typically applied to compute piecewise linear approximations on triangle, or more generally, polygonal meshes, and they can be computed very efficiently, but generally provide a lower accuracy when we consider sparse data. 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. The inverse distance weighting and ordinary kriging are very well-known approximation methods in this field. Ordinary kriging uses a variogram to capture the spatial distribution of the input data; similarly, RBFs use a kernel, which can be also adapted to the spatial distribution of the data, through the selection of the kernel width. Among other techniques, we mention Poisson based methods. The solution to the Poisson equation, which has been applied to surface reconstruction from point sets (Kazhdan et al., 2006), can be written in terms of the harmonic kernel. Therefore, Poisson methods provide results analogous to the approximation with RBFs induced by the harmonic kernel. Since the harmonic kernel tends to over-smooth the solution, in our experiments we will focus on the approximation with RBFs induced by the Gaussian kernel.

The aforementioned approximation methods define different functions, whose behavior is studied both at the numerical level (accuracy, sensitivity to sparseness, computational issues) and at a qualitative level by measuring the differences among the configuration of precipitation maxima induced by the six techniques. The comparative study was conducted selecting Liguria as area of interest, and two precipitation events recorded on September 29, 2013 and January 17, 2014, characterized by different meteorological situation and events. For the latter event, we also used rain data extracted from weather radar acquisition.

To contextualize better the comparison, 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). We give the formal definition of the six approximation methods discussed (Sect. 4) and discuss their performances with respect to accuracy, behavior with respect to 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 micro-wave 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). The combined use of rain height measured at rain gauges and radar-derived ones provides locally accurate but spatially anisotropic measures (around gauges) with globally distributed detailed data. Furthermore, we mention that 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 used 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 Voronoi cell. Back to the 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 iso-hyets. 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, 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 their differences. Moreover, we have extended the evaluation approach by comparing the differences in the configurations of meaningful features of the precipitation fields, namely prominent maxima. The motivation for this evaluation is that precipitation maxima convey important information for *storm tracking*, a crucial analysis of dynamic measures of rain data, where meaningful features associated with distinct time frames, are matched to track their evolution along time.

There is a rich literature on storm tracking, mostly using a region-based approach, where regions in radar images are characterized by high reflectivity and sufficiently large area. Various characteristics of these regions, such as centroids, area, major/minor radii, and orientation, are computed, see for instance (Lakshmanan and Smith, 2009; Dixon and Wiener, 1993; Han et al., 2009). However, we underline that the focus of the paper is not storm tracking. Indeed, we use the storm tracking measure recently proposed in Biasotti et al. (2015) for 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 distance, combining geographical distance and the measured rainfall difference, allows for matching and tracking prominent maxima along time. The same strategy for matching maxima is used to evaluate the displacement of the maxima of the different approximated fields, treating them as if they were snapshots at different times.

3. Case studies 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, located 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 hydro-meteorological events (Sect. 3.1), frequently characterized by heavy rain due to Atlantic low pressure area, augmented by a secondary low pressure area created by the Ligurian sea (Genova Low). Moreover, the several and small catchments typically cause 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 (Sect. 3.2), which targets the understanding of the best approximation method to capture important and potentially dangerous precipitation events.

3.1. Rainfall stations and radar data

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 resolution of the rain gauges is $0,2mm$ while their accuracy is in the range of 2% error threshold. 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 and temporal 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 two different rainy days, September 29, 2013 and January 17, 2014. The first event was characterized by light rain over the whole Liguria and 2 different rainstorms that caused local flooding and landslides, without damages. The second event was characterized by the transit of different fronts with well distributed rain, and was part of a rainy period that caused several deaths and a train derail. The maximum rain-rate over all time step is 60 mm/30$'$ and the average rain-rate is $1.12mm/30'$. For the second event, we also used the rainfall measured every 10 minutes provided by the polarimetric weather radar of Liguria, deployed by ARPAL. The radar scans cover an area of about 134 Km, and the rainfall measures extracted from the scan are sampled on a grid with 1 Km of resolution. The calibration of the model used to extract rain data from the weather radar has been done using the same rain gauges network.

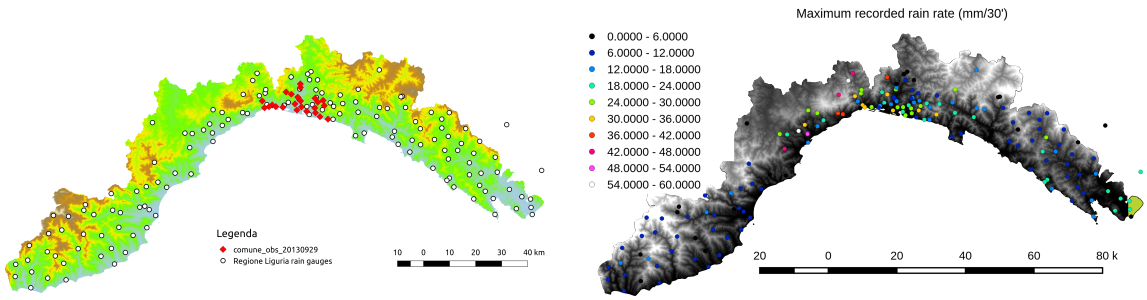
In addition to real data, we adopt also a synthetic rain field as an additional ground-truth defined using a module of the GRASS-GIS software, which produces a fractal field based on spectral synthesis methods (Saupe, 1988) (Fig. 2). Generated values have been scaled to be in the range of the rainfall values. To simulate a set of rain gauges, we sample the synthetic rainfall field with 200 points, randomly placed in a grid that contains Liguria. The goal here is to evaluate how the fields obtained with the six methods are far from the synthetic one in the whole grid.

3.2. Evaluation settings

To establish a formal evaluation setting, let us formulate the problem of rainfall approximation as follows. Given a set of points $\mathcal{P} := \{\mathbf{p}_i\}_{i=1}^{n}$, let us call $f : \mathcal{P} \rightarrow \mathbb{R}$ the precipitation field, known only at the n sample points in $\mathcal{P}$, which represent the positions of the measurement instruments and/or the nodes of the regular grid associated with the radar image. An approximation of f is defined as $F : \mathbb{R}^2 \rightarrow \mathbb{R}$ 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 than those in $\mathcal{P}$, with results differing according to the approach used to define F. In our case, we will consider six different F approximation functions.

To compare the approximations, we adopt a cross-validation strategy, exploiting the sets of data we have at regional and municipality level. Every rainfall station at $\mathbf{p}_i$ is iteratively turned off, that is, it is not used in the computation of F; the resulting approximation function F 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). Then, 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 distributions.

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



(a) (b)

Figure 1: (a) Input rainfall measures at 143 stations (regional level, white points) and 25 stations (municipality level, red circles). (b) Map of the maximum rain rate 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. Radar data used for our experiments cover the whole region and the scale of the color coding is *mm*.

maximum has in characterizing the associated rainfall field.

For this set of experiments, the approximated precipitation fields were sampled at the vertices of a digital terrain model, extracted from the SRTM (Shuttle Radar Topography Mission (Farr et al., 2007)), available in the public domain at the URL `http://www2.jpl.nasa.gov/srtm/`, and with a mesh size of 100 m.

4. Theoretical background

We give an overview of the six approximation methods (Sect. 4.1) and of the persistence analysis framework used to analyze the evolution of precipitations (Sect. 4.2).

4.1. Approximation schemes

We briefly review the following approximation schemes: nearest neighbor method, piecewise linear approximation with barycentric coordinates, LR B-splines, implicit approximation with radial basis functions, and kriging.

Nearest neighbor approximation. The value of the rainfall approximation at a sample $\mathbf{p}$ is equal to the rainfall value measured by the rainfall station closest to $\mathbf{p}$.

Piecewise linear approximation. Given an input triangulation $\mathcal{T}$ (e.g., the Delaunay triangulation of the rainfall stations) and assuming that a set of values $(f(\mathbf{p}_i))_{i=1}^{n}$ (e.g., the rainfall values) is associated with the vertices $\mathcal{P} := \{\mathbf{p}_i\}_{i=1}^{n}$ of $\mathcal{T}$, the piecewise linear approximation $f(\mathbf{p})$ at a sample $\mathbf{p}$ in $\mathcal{T}$ is computed by identifying

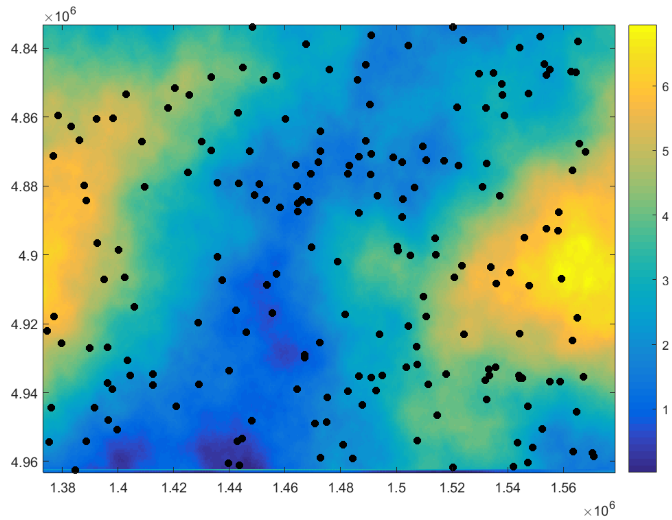
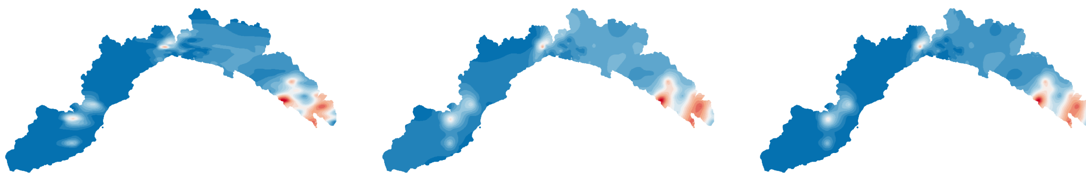


Figure 2: The synthetic field and the position of the 200 random rain gauges (black dots) used as ground truth for the evaluation of the approximation schemes. The values of the field varies from $0mm$ (blue) to $6.94mm$ (yellow).

the triangle $t \in \mathcal{T}$ of vertices $(\mathbf{p}_i, \mathbf{p}_j, \mathbf{p}_k)$ that contains $\mathbf{p}$ and expressing $f(\mathbf{p}) := \alpha_1 f(\mathbf{p}_i) + \alpha_2 f(\mathbf{p}_j) + \alpha_3 f(\mathbf{p}_k)$ as a linear combination of the f-values at the vertices of t. In this case, the coefficients $(\alpha_i)_{i=1}^3$, $\alpha_i \geq 0$, $i = 1, 2, 3$, $\alpha_1 + \alpha_2 + \alpha_3 = 1$, are the barycentric coordinates of $\mathbf{p}$ with respect to the vertices of t.

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). Similar to tensor product B-spline surfaces, the LR B-spline surfaces are defined from basis functions (B-splines) that have a local support. 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: refinement of the LR B-spline, distance computations, and update of the surface coefficients. The latter two elements are the most time



(a) (b) (c)

Figure 3: Rainfall fields computed for the event of September, 23: (a) LR B-Spline, (b) RBF, (c) Kriging. Colors represent the field values from low (blue) to high (red).

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 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 (Fig. 3a).

Inverse distance weighting. Using the inverse weighted distance (IWD) schema, the unknown value z, at location $\mathbf{p}_0$ can be estimated as a linear combination of nearby measured values, with the weights being inversely proportional to some power of the distance between observations and $\mathbf{p}_0$; i.e.,

$$z(\mathbf{p}_0) = \sum_{i=1}^{N} \omega_i z(\mathbf{p}_i), \quad \omega_i = \frac{d_{i0}^p}{\sum_{i=1}^{N} d_{i0}^p},$$

where the weights ω are expressed as functions of distances d. The basic idea of the IDW method is that observations that are close to each others on the ground tend to be more similar than those ones that are further apart (Tobler, 1970); hence, observations closer to $\mathbf{p}_0$ receive a larger weight. This interpolation method requires the choice of the exponent p (e.g., $p := 2$) and of a search radius R or, alternatively, the minimum number N of points required for the interpolation. Greater values of the exponent p assign greater influence to values closest to the interpolated point, with the result turning into a mosaic of tiles (like a Voronoi diagram) with nearly constant interpolated value for large values of p.

Implicit approximation with Radial Basis Functions. The 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

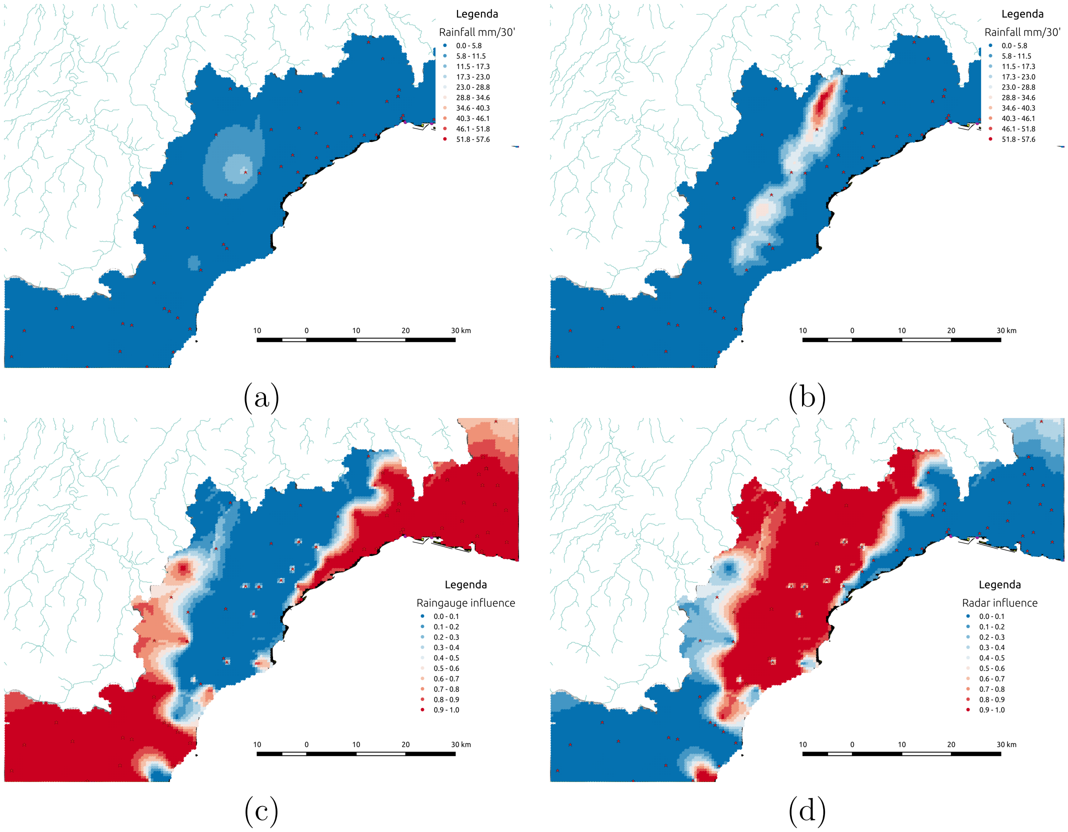


Figure 4: Ordinary kriging approximation of rainfalls computed with (a) rain gauges and (b) integrated with radar measurements. (c) Rain gauges weights and (d) radar data set mapped in (b). Colors represent the field values from low (blue) to high (red).

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- (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 achieved by imposing the interpolating constraints $F(\mathbf{p}_i) = f(\mathbf{p}_i)$, $i = 1, \dots, n$. 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 $\varphi(st) := \exp(-st^{1/2})$, which has a global support; in fact, its fast decay makes it suitable to approximate rainfalls with a sparse spatial distribution and that change quickly in time (Fig. 3b). 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

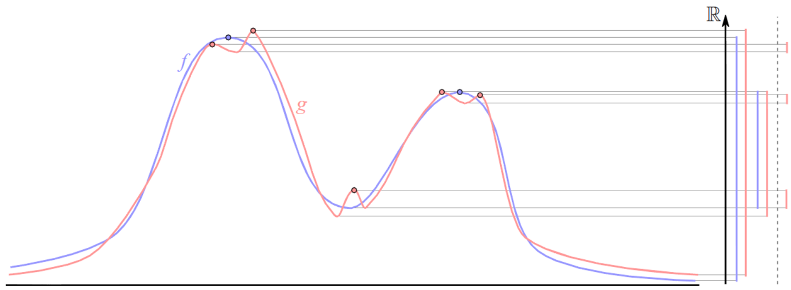


Figure 5: Two scalar fields f, g and their local maxima. On the right, pictorial representation for the persistence of each local maxima. Segments on the right of the dotted line stand for the persistence of topological noise.

Sun, 2005; Mitra and Nguyen, 2003).

Kriging. The previous approximation methods do not take into account in an explicit manner the correlation among observations, which 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 (Fig. 3c).

Formally, kriging is expressed as $F(\mathbf{p}) := \sum_{i=1}^{n} \omega_i f(\mathbf{p}_i)$, where the weights $\omega := (\omega_i)_{i=1}^{n}$ are the solution to the linear system $\mathbf{C}\omega = \mathbf{d}$, where $\mathbf{C}$ is the covariance matrix of the of the input points, $\mathbf{d}$ is the array of the covariance between the positions of the rainfall stations and the points that belong to a neighborhood of the sample point. 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, and in our tests we have used a constant variogram, whose nugget is set equal to 10% and the range is $30Km$. Fig. 4 shows the results obtained by kriging when radar rain data are integrated.

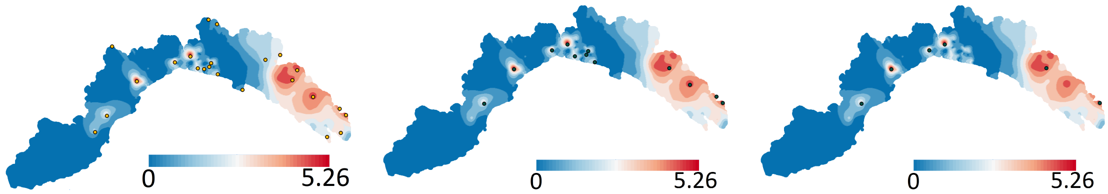


Figure 6: A function $F : \mathcal{M} \rightarrow \mathbb{R}$, color-coded from blue (low) to red (high) values, and the associated local maxima having persistence greater than $\alpha(\max F - \min F)$, with $\alpha = 0.05$, 0.15 (middle) and 0.25.

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 $F : \mathcal{M} \rightarrow \mathbb{R}$ (e.g., the interpolated rainfall field), persistence analysis is used to study the evolution of the connectivity in the superlevel sets $\mathcal{M}^t = \{\mathbf{p} \in \mathcal{M} : F(\mathbf{p}) \geq t\}$, for $t \in (-\infty, +\infty)$. Sweeping t from $+\infty$ to $-\infty$, new connected components of $\mathcal{M}^t$ are either born, or previously existing ones are merged together. A connected component C is associated with a local maximum $\mathbf{p}$ of F, where the component is first born. The value $F(\mathbf{p})$ is referred to as the birth time of C. When two components corresponding to local maxima $\mathbf{p}_1$, $\mathbf{p}_2$, with $F(\mathbf{p}_1) < F(\mathbf{p}_2)$, merge together, we say that 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 *persistence value* $\mathrm{pers}_F(\mathbf{p})$, 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, while maxima having a low persistence value are interpreted as local information or noise (Fig. 5).

To compute the local maxima and the associated persistence values, 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 $O(n \log n)$; after sorting, 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. 6.

Method	Max [mm]	Mean [mm]	Median [mm]	Std. dev. [mm]	MSE [mm^2]
Syntethic					
Ord. krig.	0.93 (14.1%)	0.15	0.10	0.22	0.04
RBFs	0.55 (8.3%)	0.14	0.10	0.18	0.03
LR B-Splines	0.48 (7.1%)	0.16	0.11	0.22	0.05
IDW	1.16 (17.4%)	0.22	0.15	0.31	0.097
NN	1.46 (21.7%)	0.29	0.20	0.41	0.17
BC	0.79 (11.9%)	0.34	0.11	0.88	0.83
Day 1					
Ord. krig.	32.44 (54.1%)	0.02	0	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	0	2.73	7.05
IDW	27.2 (45.32%)	0.78	0.055	2.58	6.66
NN	35.2 (58.67%)	0.83	0.0	2.94	8.63
BC	27.19 (45.33%)	0.71	0.02	2.40	5.79
Day 2					
Ord. krig.	16.6 (88.3%)	1.95	1.18	2.88	8.61
RBFs	16.59 (88.3%)	1.28	0.80	1.97	3.88
LR B-Splines	16.6 (88.3%)	1.27	0.79	1.98	3.95
IDW	7.6 (40.5%)	1.19	0.79	1.82	3.34
NN	16.6 (88.3%)	1.49	0.80	2.35	5.52
BC	16.6 (88.3%)	1.35	0.80	2.08	4.36

Table 1: Statistics for the error distribution of the cross validation.

5. Approximation behavior

In the following, we discuss the comparison of the behavior with respect to the approximation performance (Sect. 5.1), the local analysis of the field differences (Sect. 5.2), and the computational complexity (Sect. 5.3).

5.1. Approximation accuracy

For the leave-one-out cross-validation strategy, we have checked the results by computing the six 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; the approximation methods behave in a slightly different way depending on the three scenarios.

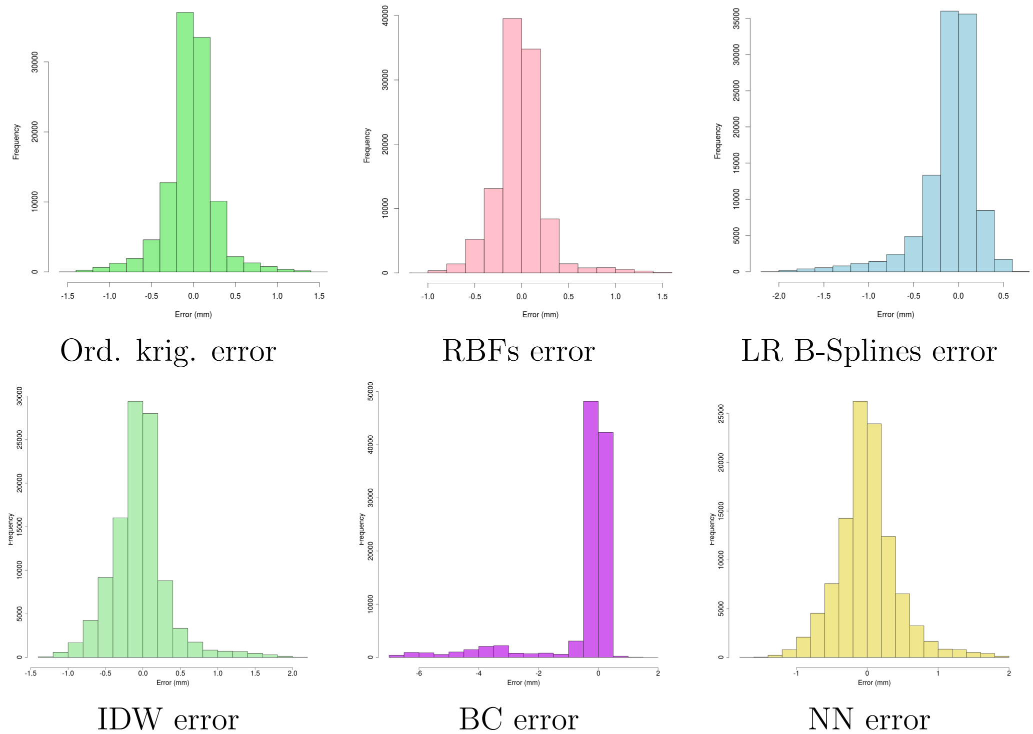


Figure 7: Histograms of the differences among the three approximated rainfall fields and a synthetic ground truth.

The nearest neighbor perform worst in most of the scenarios while ordinary kriging, LR B-spline and RBFs have overall the better performances. In the synthetic and the day-2 case studies, the best performances are achieved by IDW (day-2) followed by RBFs and LR B-splines, while ordinary kriging has a larger maximum error in the synthetic case and larger mean absolute error for day-2. In the day-1 case study, ordinary kriging and LR B-Splines have a smaller maximum error, but the RBFs have a lower mean-squares error and standard deviation.

The histogram of the error in the Fig. 7 shows clearly that the distribution of ordinary kriging and nearest neighbor are normally distributed while RBF and IDW have positive skewness and LR B-spline and BC have a negative skewness. The global behavior of the approximation techniques is also well shown in the map depicted in Fig. 8, the distance between the approximated fields and synthetic ground truth are plotted: ordinary kriging produces more spots that are characterized by an error higher than RBF and LR B-spline.

The second set of results concerns the cross-validation with the rainfall data

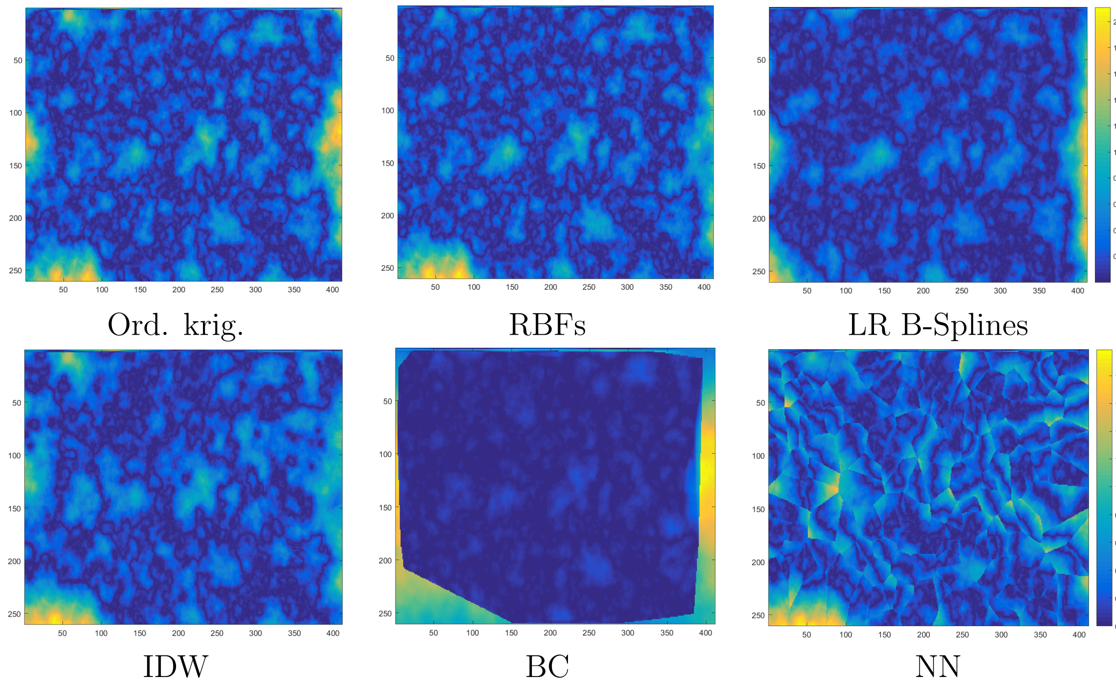


Figure 8: Error between the six approximated fields and the synthetic ground truth.

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 distributions 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 sparser data set. According to the results in Table 2, IDW and BC have the smaller maximum error, but the RBFs have a smaller mean-squares error. The worst performance is achieved by the nearest neighbor method.

5.2. Local analysis of the field differences

To measure the smoothness of the approximated rainfall fields, we compare the corresponding normalized gradients (Fig. 9). More precisely, given the approximated rainfall fields F_1, F_2 and the gradients ∇F_1 and ∇F_2, their point-wise difference at

Table 2: Statistics for the error distribution of the accuracy evaluation at different scales for day 1.

Method	**Max** $[mm]$	**Mean** $[mm]$	**Median** $[mm]$	**Std. dev.** $[mm]$	**MSE** $[mm^2]$
Ord. krig.	28.62 (47.7%)	0.59	0.01	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	0.01	4.45	20.19
IDW	27.2 (45.3%)	1.80	0.3	4.27	18.55
BC	27.2 (45.3%)	1.84	0.3	4.36	19.32
NN	35.2 (58.6%)	1.90	0.2	4.86	24.01

the node (i, j) of a uniform grid contained in the bounding box of Liguria is measured as Biasotti et al. (2007)

$$d(\nabla F_1, \nabla F_2)(i, j) := 1 - |\langle \nabla F_1(i, j), \nabla F_2(i, j)\rangle_2| \,.$$

As expected, the behavior of the gradients and their dot product reflects the punctual difference of the rainfall fields. We also compare the rainfall approximations looking at the differences of the rain values assumed on the DTM and the local smoothness of the three fields. First, we show the point-wise difference of the rainfall fields (Figs. 11, 12 at page 30). As expected, the difference of the fields is zero at the rain stations and, for the radar data also in the nodes of the regular grid. Since for the kriging approximation we adopted a local support, it gives a slightly perturbed approximation of the field far from the rain gauges and the radar nodes. Furthermore, the approximations with LR B-Splines and RBFs have a smoother behavior and a lower approximation error. Finally, Fig. 13 (page 32) represents the difference of the gradients over the selected grid: it can be seen that kriging has noisy values far from the sampling points, as a matter of the local behavior of the algorithm; RBF and ordinary kriging behave in a similar way near the samples while LR B-Spline show difference in the gradient with respect to both other methods.

5.3. Computational complexity

The computational complexity of the different algorithms has 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 computational time is measured on the rainfall data from the first day and with only rain gauges (no radar).

The run of BC takes 61.54 seconds to compute the approximation over the whole region (20K points) for the one time interval. For the same task, the ordinary kriging takes 1.746 seconds and RBFs approximation takes 6.23 seconds. The IDW takes the

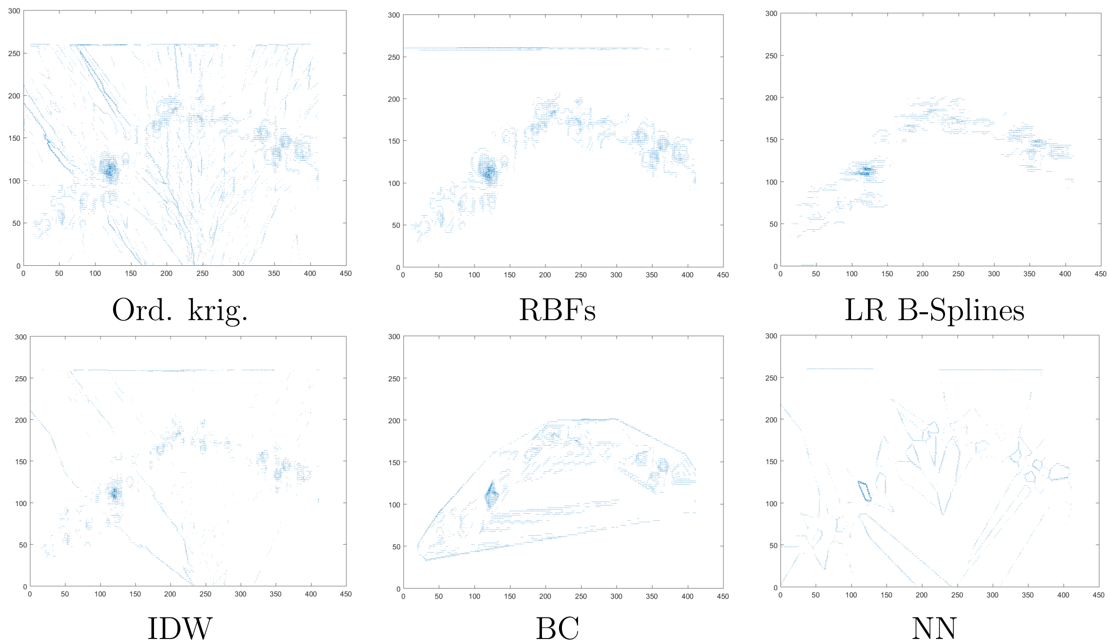


Figure 9: Gradient field of the six methods; in these images, the fields are approximated only with the rain stations.

shortest times needing only 1.61 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 on 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.

6. Analysis of rainfall field

We discuss the identification (Sect. 6.1), comparison (Sect. 6.2), and tracking (Sect. 6.3) of persistent maxima.

6.1. Identification of rainfall maxima

Tables 3-6 report the comparative results about the extraction of persistent maxima when considering the rainfall fields produced by the approximation schema using the ARPAL rainfall stations and when these stations are integrated with the radar data. For these tests, we used the rain data of the first precipitation event and

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

Method	$\tau = \mathbf{0.05}$	$\tau = \mathbf{0.15}$	$\tau = \mathbf{0.25}$	$\tau = \mathbf{0.35}$
Ord. krig.	28.31	11.27	6.70	4.37
IDW	27.15	13.55	7.69	4.57
RBFs	18.54	10.31	6.12	4.08
LR B-Splines	20.54	12.50	7.41	4.67
BC	24.54	12.92	7.98	4.88
NN	18.52	12.33	7.94	5.13

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

Method	$\tau = \mathbf{0.05}$	$\tau = \mathbf{0.15}$	$\tau = \mathbf{0.25}$	$\tau = \mathbf{0.35}$
Ord. krig.	48	19	13	9
IDW	40	22	17	10
RBFs	26	17	11	9
LR B-Splines	28	18	13	10
BC	70	55	42	12
NN	24	18	14	10

radar data (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 a 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 on the rainfall fields approximated from the rainfall stations only. Table 4 shows the maximum number of local maxima that have been extracted, method by method, from the 48 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 situation 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.

Similarly, Tables 5 and 6 report the same data when the approximation schema integrate also the radar data. The trend to have a decreasing number of persistent maxima as the threshold τ increases is confirmed and much more evident. Indeed,

Table 5: Statistics for the average number of extracted persistent maxima with radar data.

Method	$\tau = \mathbf{0.05}$	$\tau = \mathbf{0.15}$	$\tau = \mathbf{0.25}$	$\tau = \mathbf{0.35}$
Ord. krig.	99.22	26.46	10.91	5.15
IDW	114.05	26.54	10.67	5.11
RBFs	93.35	24.85	9.85	4.68
LR B-Splines	113.46	31.58	13.05	6.33
BC	105.48	29.5	12.29	5.69
NN	124.33	37.98	15.65	6.94

Table 6: Statistics for the maximum number of extracted persistent maxima with radar data.

Method	$\tau = \mathbf{0.05}$	$\tau = \mathbf{0.15}$	$\tau = \mathbf{0.25}$	$\tau = \mathbf{0.35}$
Ord. krig.	175	62	27	13
IDW	191	61	24	11
RBFs	165	60	24	11
LR B-Splines	195	76	35	19
BC	181	70	32	12
NN	222	97	48	21

the approximation of the fields with such a higher number of constraints introduces a quite large number of local maxima that are not really relevant and that are discarded when the persistence threshold increases. Our tests further confirm that the RBF schemes generally have a higher smoothness, as indicated by the slightly smaller number of maxima.

6.2. Comparing sets of persistent maxima

In order to refine the above comparative analysis, we use 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. Data are considered the same cumulative step.

According to (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-*

Table 7: Average geographical distance (Km) between sets of local maxima (Liguria area size: $5.410 Km^2$).

Method1/Method2	$\tau = \mathbf{0.05}$	$\tau = \mathbf{0.15}$	$\tau = \mathbf{0.25}$	$\tau = \mathbf{0.35}$
Ord. krig / RBFs.	71.19 Km	13.76 Km	4.67 Km	2.46 Km
Ord. krig / LR B-Splines	104.59 Km	52.42 Km	29.66 Km	25.45 Km
Ord. Krig / IDW	142.32 Km	31.86 Km	17.26 Km	5.71 Km
Ord. Krig / BC	92.02 Km	43.03 Km	25.61 Km	5.92 Km
Ord. Krig / NN	110.69 Km	96.32 Km	61.58 Km	30.49 Km
RBFs / LR B Spline	81.85 Km	54.47 Km	28.79 Km	14.46 Km
RBFs / IDW	61.66 Km	39.48 Km	21.16 Km	5.82 Km
RBFs / BC	85.33 Km	45.47 Km	25.03 Km	9.68 Km
RBFs / NN	117.52 Km	94.96 Km	59.08 Km	35.91 Km
LR B Spline / IDW	75.48 Km	65.23 Km	36.67 Km	20.98 Km
LR B Spline / BC	87.26 Km	71.34 Km	35.85 Km	26.43 Km
LR B Spline / NN	121.40 Km	106.46 Km	74.18 Km	50.19 Km
IDW / BC	88.79 Km	39.71 Km	33.44 Km	18.14 Km
IDW / NN	469.05 Km	396.51 Km	274.33 Km	169.76 Km
BC / NN	121.09 Km	97.78 Km	60.34 Km	41.06 Km

distance

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

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 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. 10). 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 characterizes 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

Table 8: Average geographical distance (Km) between sets of local maxima (Liguria area size: $5.410 Km^2$).

Method1/Method2	$\tau = \mathbf{0.05}$	$\tau = \mathbf{0.15}$	$\tau = \mathbf{0.25}$	$\tau = \mathbf{0.35}$
Ord. krig + radar/ RBFs. + radar	186.01 Km	135.04 Km	7.27 Km	4.94 Km
Ord. krig + radar/ LR B-Splines + radar	121.49 Km	60.99 Km	30.10 Km	11.79 Km
Ord. Krig + radar/ IDW + radar	136.48 Km	70.59 Km	36.863 Km	7.99 Km
Ord. Krig + radar / BC + radar	125.16 Km	94.78 Km	36.85 Km	20.44 Km
Ord. Krig + radar / NN + radar	147.49 Km	75.01 Km	64.52 Km	15.73 Km
RBFs + radar / LR B Spline + radar	223.89 Km	150.99 Km	23.20 Km	6.21 Km
RBFs + radar / IDW + radar	493.61 Km	304.45 Km	107.18 Km	30.79 Km
RBFs + radar / BC + radar	159.47 Km	75.10 Km	64.51 Km	15.73 Km
RBFs + radar / NN + radar	234.11 Km	168.15 Km	50.72 Km	14.10 Km
LR B Spline + radar/ IDW + radar	130.95 Km	71.61 Km	44.30 Km	16.26 Km
LR B Spline + radar / BC + radar	125.16 Km	94.79 Km	36.86 Km	20.45 Km
LR B Spline + radar / NN + radar	142.97 Km	104.58 Km	63.38 Km	30.22 Km
IDW + radar / BC + radar	147.49 Km	75.01 Km	64.52 Km	15.73 Km
IDW + radar / NN + radar	140.82 Km	80.73 Km	54.68 Km	18.89 Km
BC + radar / NN + radar	147.49 Km	75.00 Km	64.51 Km	15.73 Km

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.

6.3. Tracking rainfall maxima

For each cumulative step, we consider the interpolated rainfall fields, and extract 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 for two persistent maxima paired by the bottleneck matching.

Tables 7, 8, 9 and 10 report the obtained results, in terms of geographical and rainfall distances, respectively, averaged by the total number of considered cumula-

Table 9: Average rainfall distance (mm) between sets of local maxima.

Method1/Method2	$\tau = \mathbf{0.05}$	$\tau = \mathbf{0.15}$	$\tau = \mathbf{0.25}$	$\tau = \mathbf{0.35}$
Ord. krig / RBFs.	3.83 mm	1.87 mm	1.15 mm	0.28 mm
Ord. krig. / LR B-Sp.	3.63 mm	3.32 mm	2.63 mm	1.74 mm
Ord. Krig / IDW	3.17 mm	3.24 mm	2.56 mm	1.33 mm
Ord. Krig / BC	4.32 mm	2.64 mm	1.91 mm	1.32 mm
Ord. Krig / NN	3.81 mm	3.03 mm	2.45 mm	2.03 mm
RBFs / LR B Sp.	3.21 mm	3.13 mm	3.10 mm	2.13 mm
RBFs / IDW	4.09 mm	3.76 mm	2.58 mm	1.68 mm
RBFs / BC	3.33 mm	2.59 mm	2.10 mm	1.54 mm
RBFs / NN	3.03 mm	3.08 mm	2.68 mm	2.57 mm
LR B Sp. / IDW	3.88 mm	3.42 mm	3.02 mm	1.48 mm
LR B Spline / BC	4.01 mm	3.51 mm	2.74 mm	2.01 mm
LR B Spline / NN	3.33 mm	3.28 mm	2.47 mm	2.17 mm
IDW / BC	4.53 mm	3.32 mm	2.65 mm	1.60 mm
IDW / NN	15.35 mm	13.12 mm	10.18 mm	8.24 mm
BC / NN	3.11 mm	2.45 mm	2.31 mm	1.65 mm

tive steps. To have a clearer picture of the comparative evaluation in terms of the two distances, these results 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. From this perspective, approximations with RBFs and kriging have an analogous behavior, both producing higher values for the geographical and rainfall distances when compared with LR-B Splines. Moving to higher persistence thresholds, the values of the geographical distance decrease, as an effect of filtering out non-relevant maxima, and the corresponding rainfall distance values reveal now the differences occurring at prominent maxima, which appear to be quite small.

In Table 11, we show a similar comparison of the results obtained when rainfall fields are interpolated by considering either observed rainfall measurements or an integration of these data with radar acquisitions (Sect. 5). Integrated data can reveal useful information for rainfall tracking over time, as a matter of the higher spatial and temporal resolution of radar data with respect to point-wise rainfall fields measured by instruments at the ground level. 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. According to the results

Table 10: Average rainfall distance (mm) between sets of local maxima.

Method1/Method2	$\tau = \mathbf{0.05}$	$\tau = \mathbf{0.15}$	$\tau = \mathbf{0.25}$	$\tau = \mathbf{0.35}$
Ord. krig + radar / RBFs. + radar	10.22 mm	7.39 mm	4.22 mm	1.47 mm
Ord. krig. + radar / LR B-Sp. + radar	11.35 mm	9.80 mm	8.52 mm	5.18 mm
Ord. Krig + radar / IDW + radar	14.00 mm	10.63 mm	6.33 mm	2.76 mm
Ord. Krig + radar / BC + radar	11.55 mm	9.44 mm	7.89 mm	4.51 mm
Ord. Krig + radar / NN + radar	11.17 mm	11.10 mm	9.75 mm	8.39 mm
RBFs + radar / LR B Sp. + radar	11.38 mm	10.22 mm	8.86 mm	5.79 mm
RBFs + radar / IDW + radar	42.34 mm	32.27 mm	21.96 mm	12.43 mm
RBFs + radar / BC + radar	11.18 mm	11.11 mm	9.75 mm	8.38 mm
RBFs + radar / NN + radar	11.52 mm	11.70 mm	10.91 mm	7.68 mm
LR B Sp. + radar / IDW + radar	13.99 mm	12.25 mm	8.99 mm	6.19 mm
LR B Spline + radar / BC + radar	11.55 mm	9.44 mm	7.89 mm	4.52 mm
LR B Spline + radar / NN + radar	11.30 mm	10.38 mm	9.16 mm	5.30 mm
IDW + radar / BC + radar	11.55 mm	11.11 mm	9.75 mm	8.39 mm
IDW + radar / NN + radar	13.79 mm	12.93 mm	10.87 mm	8.21 mm
BC + radar / NN + radar	11.17 mm	11.10 mm	9.76 mm	8.39 mm

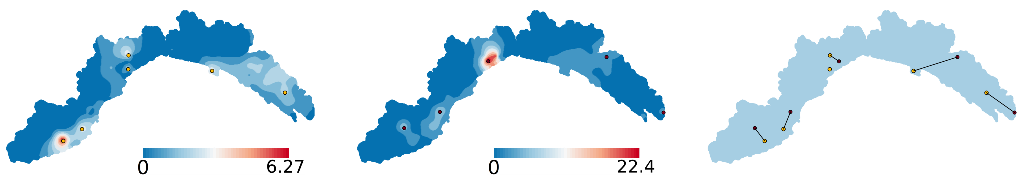


Figure 10: Two fields $F, G : \mathcal{M} \rightarrow \mathbb{R}$, color-coded from blue (low) to red (high) values, and the associated local maxima. On the right, bottleneck matching between local maxima.

in Table 11, which are characterized by high values in both the geographic and the rainfall distance, radar data can sensibly change the spatial location and the rainfall value of persistent maxima. This result can be interpreted as the introduction of complementary information with respect to rainfall measurements, which hopefully support a clearer understanding of precipitation events.

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. For the approximation of rainfall data all the approaches differ provide satisfactory results except for nearest neighbor that lower performance in most of our test, with a slightly better behavior for LR Splines and RBFs, and

Table 11: Average geographical (Km) and rainfall distance (mm) between sets of local maxima.

Krig/ (Radar + Krig)	$\tau =$ **0.05**	$\tau =$ **0.15**	$\tau =$ **0.25**	$\tau =$ **0.35**
Geogr. dist. Rainfall dist.	146.08 Km 17.52 mm	100.68 Km 17.23 mm	104.25 Km 16.90 mm	83.48 Km 16.04 mm
RBF/ (Radar + RBF)	$\tau =$ **0.05**	$\tau =$ **0.15**	$\tau =$ **0.25**	$\tau =$ **0.35**
Geogr. dist. Rainfall dist.	95.38 Km 17.71 mm	96.13 Km 17.25 mm	93.29 Km 16.56 mm	88.88 Km 16.21 mm
LR B-spline/ ((Radar +LR B-Spline)	$\tau =$ **0.05**	$\tau =$ **0.15**	$\tau =$ **0.25**	$\tau =$ **0.35**
Geogr. dist. Rainfall dist.	93.53 Km 18.77 mm	94.72 Km 18.26 mm	86.56 Km 17.953 mm	77.09 Km 16.74 mm
IDW / ((Radar + IDW)	$\tau =$ **0.05**	$\tau =$ **0.15**	$\tau =$ **0.25**	$\tau =$ **0.35**
Geogr. dist. Rainfall dist.	106.46 Km 18.52 mm	116.31 Km 17.00 mm	107.07 Km 16.32 mm	90.63 Km 15.79 mm
BC / ((Radar + BC)	$\tau =$ **0.05**	$\tau =$ **0.15**	$\tau =$ **0.25**	$\tau =$ **0.35**
Geogr. dist. Rainfall dist.	91.58 Km 17.66 mm	94.49 Km 17.25 mm	99.88 Km 16.81 mm	79.65 Km 16.35 mm
NN / ((Radar + NN)	$\tau =$ **0.05**	$\tau =$ **0.15**	$\tau =$ **0.25**	$\tau =$ **0.35**
Geogr. dist. Rainfall dist.	101.07 Km 18.93 mm	96.58 Km 18.68 mm	104.61 Km 18.71 mm	87.66 Km 17.83 mm

differences that are in any case significant with respect to the resolution and error of the measuring instruments. Those methods, moreover, easily support the integration of further sources of rain measures, for instance those captured by radar. The tests were run on events that were particularly targeting the comparison under conditions of sparsity and heterogeneity of accuracy, but there is no reason to think that the trends delineated would not consistently extend to other conditions of measurement locations and orographic context.

At the theoretical level, 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. At the application level, there are several venues for exploiting the comparison framework we have built:

for instance, the identification of the areas where a significant difference among approximations persist, could indicate where new rain gauges could improve best the monitoring of the precipitation events. Most importantly, we believe that the proposed comparison framework will support more robust storm tracking methods via a better understanding of the nature of the underlying precipitation event.

Acknowledgments We acknowledge the precious comments of the anonymous Reviewers, which helped us to improve the paper content and presentation. This work has been supported by the FP7 Integrated Project IQmulus, FP7-ICT-2011-318787. The rainfall data sets are courtesy of ARPAL, Regione Liguria, and of the Genova Municipality.

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

Azimi-Zonooz, A., Krajewski, W., Bowles, D., 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), 51–67.

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

Biasotti, S., Patanè, G., Spagnuolo, M., Falcidieno, B., 2007. Analysis and comparison of real functions on triangulated surfaces. Curve and Surface Fitting Modern Methods in Mathematics, 41–50.

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

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

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

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

Dey, T. K., 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., 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), 396–408.

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

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

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

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

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

Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., 2007. The shuttle radar topography mission. Reviews of Geophysics 45 (2), 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), 113–129.

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

Hong, Y., Adler, R., 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), 1671–1680.

Hou, A. Y., Ramesh, K. K., Neeck, S., Azarbarzin, A. A., Kummerow, C. D., Kojima, M., Oki, R., Nakamura, K., Iguchi, T., 2014. The global precipitation measurement mission. Bulletin of the American Meteorological Society 95.

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

Kazhdan, M., Bolitho, M., Hoppe, H., 2006. Poisson surface reconstruction. In: Computer Graphics Forum. pp. 61–70.

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

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

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

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

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

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

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

Mitra, N. J., 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., 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., Falcidieno, B., 2009. Topology- and error-driven extension of scalar functions from surfaces to volumes. ACM Trans. on Graphics 29 (1), 1–20.

Saupe, D., 1988. Algorithms for random fractals. In: The science of fractal images. Springer, pp. 71–136.

Scheuerer, M., Schaback, R., Schlather, M., 8 2013. Interpolation of spatial data - a stochastic or a deterministic approach? European Journal of Applied Mathematics 24, 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, 116 – 131.

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

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

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

Tobler, W. R., 1970. A computer movie simulating urban growth in the detroit region. Economic geography 46 (sup1), 234–240.

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

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

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

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

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

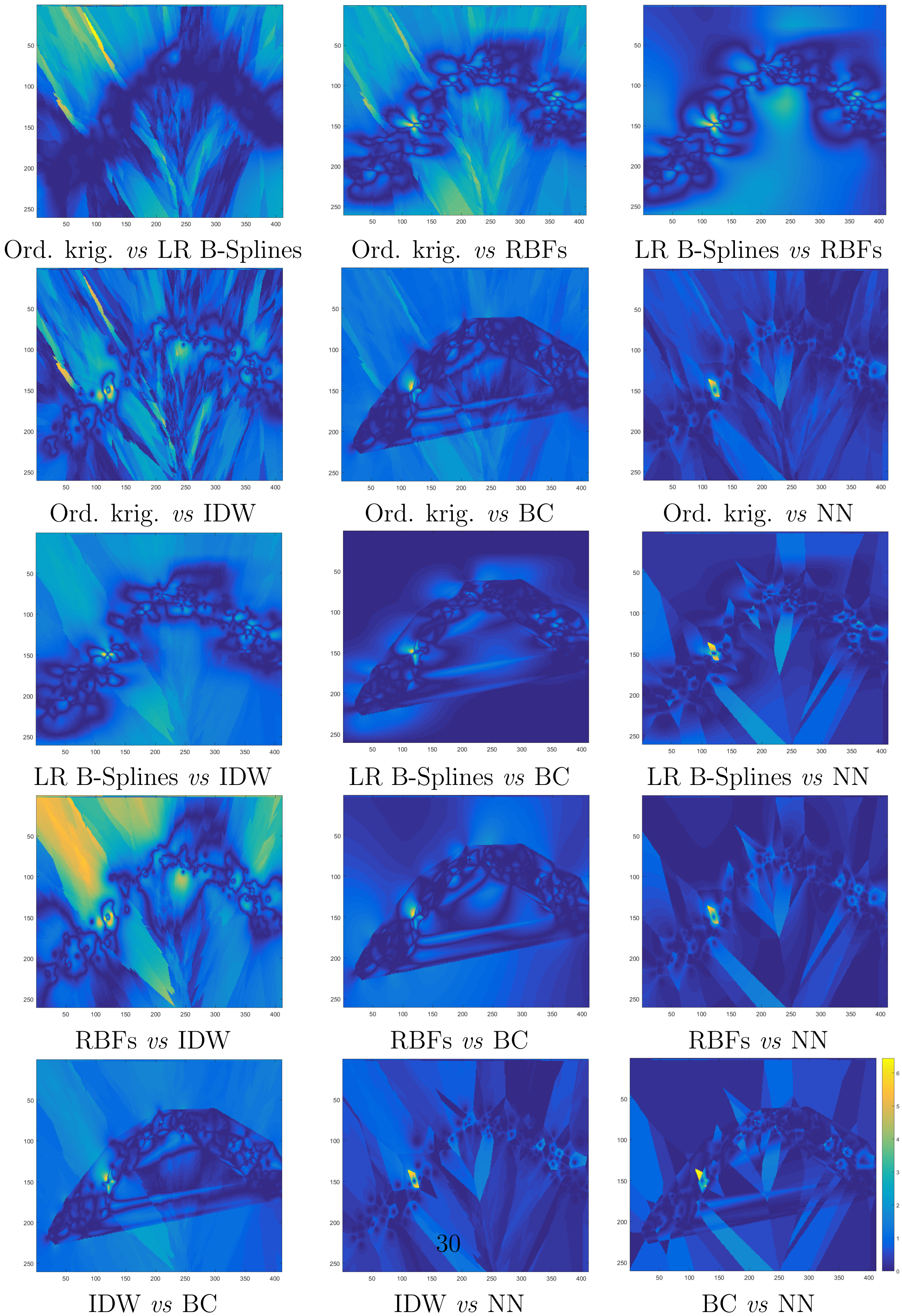
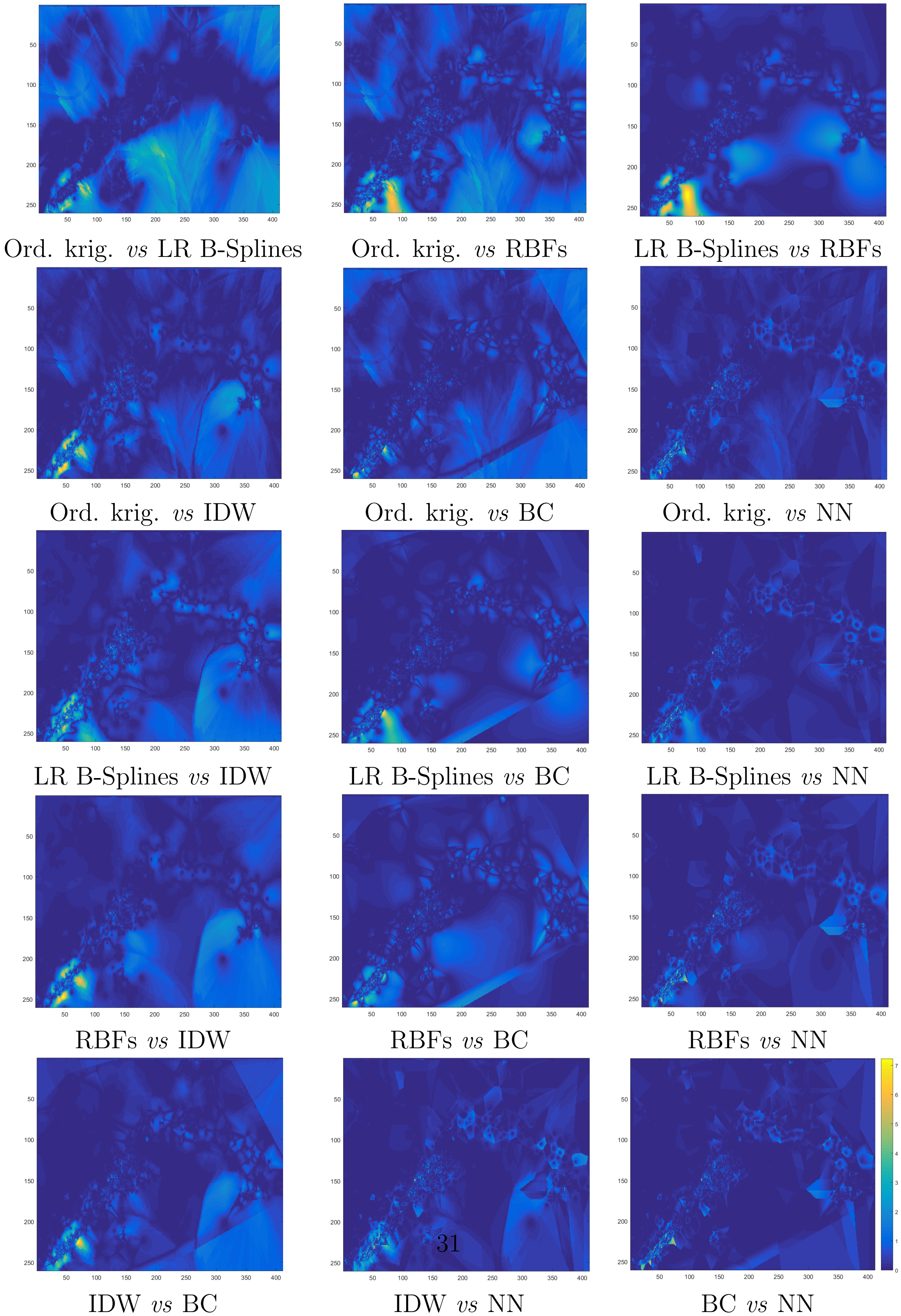


Figure 11: Point-wise difference of the rainfall fields evaluated on the rain stations. Colors represent the difference from low (blue) to high (yellow) values.



Ord. krig. *vs* LR B-Splines | Ord. krig. *vs* RBFs | LR B-Splines *vs* RBFs

Ord. krig. *vs* IDW | Ord. krig. *vs* BC | Ord. krig. *vs* NN

LR B-Splines *vs* IDW | LR B-Splines *vs* BC | LR B-Splines *vs* NN

RBFs *vs* IDW | RBFs *vs* BC | RBFs *vs* NN

IDW *vs* BC | IDW *vs* NN | BC *vs* NN

Figure 12: Point-wise difference of the rainfall fields integrated with the radar data. Colors represent the difference from low (blue) to high (yellow) values.

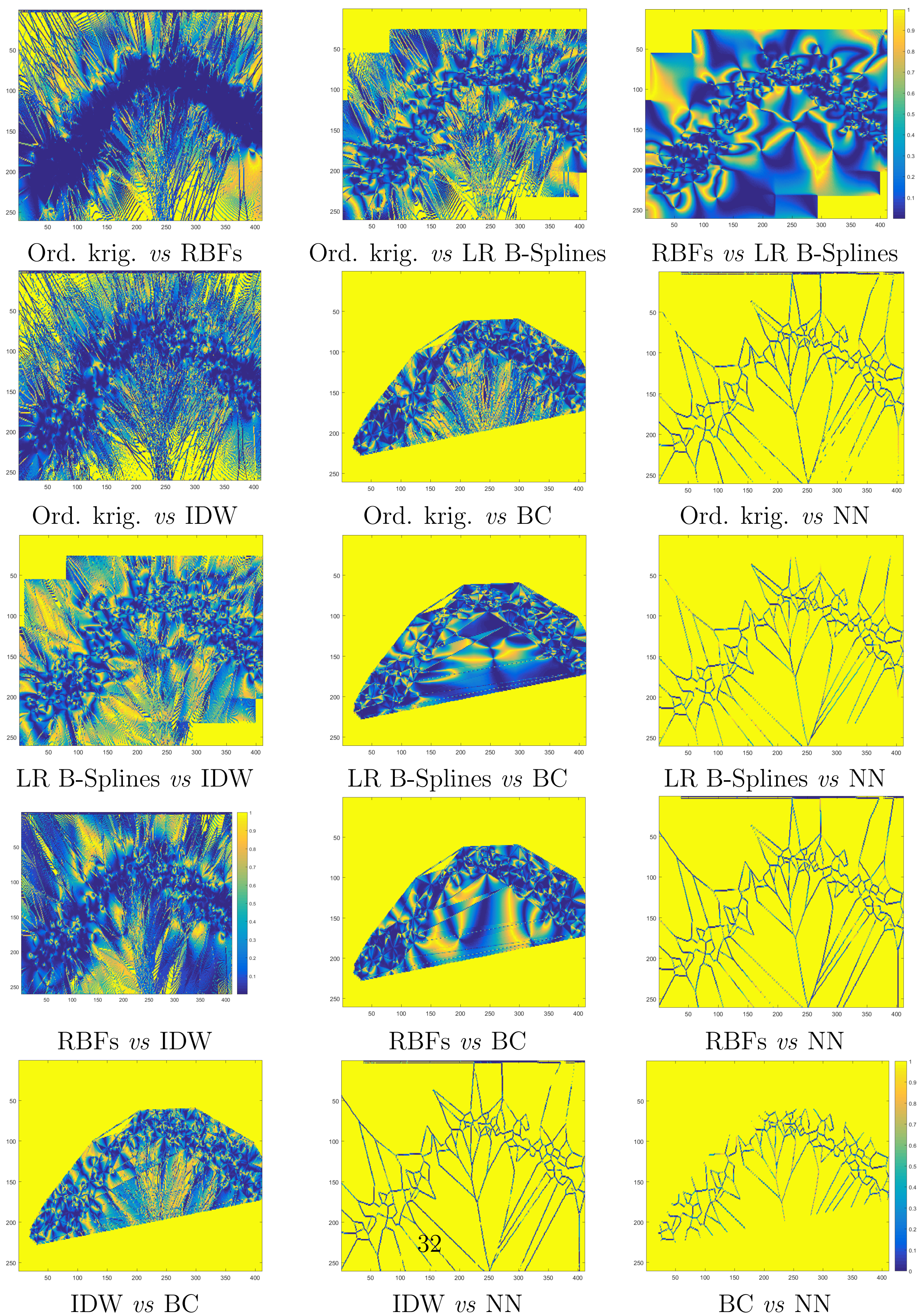


Figure 13: Local difference of the gradients of the six fields; colors represent the value of the distance d over the model grid from 0 (blue) to 1 (yellow).