A geostatistical approach to multisensor rain field reconstruction and downscaling

P. Fiorucci[1], P. La Barbera[2], L.G. Lanza[2] and R. Minciardi[1]

[1]University of Genova, Dept. of Communication, Computer and System Science, Opera Pia 13, 16145 Genova, Italy
[2]University of Genova, Dept. of Environmental Engineering, 1Montallegro, 16145 Genova, Italy
Email for corresponding author: luca@diam.unige.it

Abstract

A rain field reconstruction and downscaling methodology is presented, which allows suitable integration of large scale rainfall information and rain-gauge measurements at the ground. The former data set is assumed to provide probabilistic indicators that are used to infer the parameters of the probability density function of the stochastic rain process at each pixel site. Rain-gauge measurements are assumed as the ground truth and used to constrain the reconstructed rain field to the associated point values. Downscaling is performed by assuming the *a posteriori* estimates of the rain figures at each grid cell as the *a priori* large-scale conditioning values for reconstruction of the rain field at finer scale. The case study of an intense rain event recently observed in northern Italy is presented and results are discussed with reference to the modelling capabilities of the proposed methodology.

Keywords: Reconstruction, downscaling, remote sensing, geostatistics, Meteosat

Introduction

Direct measurements of space-time rainfall, desirable in many hydrological studies, are subject to operational limitations that reduce monitoring capabilities to the frustrating case of negligible spatial information when compared to the size of the catchment areas on the ground. Rain-gauge measurements are therefore usually referred to as being at the point scale in rainfall monitoring.

Since the first geostationary platform for meteorological use was made operational in the late '60s, the chance of obtaining quantitative estimation of the rain field over extended areas by means of suitable retrieval from newly available indirect measurements has faced hydrologists and remote sensing scientists. Several empirical, semi-empirical and physically-based algorithms are now available in the literature for the retrieval of space-time precipitation, based on the remote sensing of more or less related variables such as radiance temperatures, brightness temperatures, etc. (see e.g. Barrett and Martin (1981), Barrett and Beaumont (1994) for a review of operational methodologies).

Though significant contributions to the understanding and modelling of meteorological processes at the global, synoptic and meso-scales based on remote sensing can be found in the literature, the very indirect nature of data coming from remote sensors makes quantitative estimation of the so-called 'instantaneous' rain field hardly reliable in a deterministic approach. More encouraging results have been obtained at climatological scales, where the total rainfall accumulation over sufficiently large intervals in time (seasons or years) is obtained as the sum of numerous estimates of the rain field, which ensures reduced variances and reliable estimation of the mean field.

It is quite reasonable to assume that remotely sensed estimates of instantaneous rain fields can be viewed as probabilistic indicators of the actual rain rate figures and used to infer at least one parameter of the probability density function for rain intensity at each pixel site. This represents the assumption underlying the methodology proposed in this paper, aimed at suitable integration of rainfall estimates from remote sensing with the observations from a number of rain-gauges at the ground. Rain-gauge measurements are assumed as the ground truth even though clear evidence of non-negligible measurement errors affecting both traditional and modern instruments is increasing (Sevruk and Lapin, 1993).

In addition, estimates of space-time rainfall are provided operationally by numerical weather prediction models both

at the global (General Circulation Models, GCM) and synoptic to meso-scale (Limited Area Models, LAM). Again, the information is reliable at large scales while the uncertainty associated with rainfall estimates increases with increased resolution in both space and time. In this case, the uncertainties involved derive from the simplifying hypotheses that are needed to run operational physically-based models of the atmosphere (hydrostatic approach, rough description of the orography, use of climatological data for model initialisation, etc.), with significant implications on the actual predictability of the precipitation field.

Reliable monitoring and prediction of space-time rainfall is therefore affordable at resolution scales that are in the order of 5 to 10 km in space and three to six hours in time. At finer scales, though radar rain maps are becoming increasingly reliable, estimation of the rain field is most commonly performed by interpolation based on point measurements such as geostatistics or even simpler algorithms. Simple Kriging is the best known geostatistical technique; it is based on the assumption that the expected value of the regionalised variable is a known constant, although this is unrealistic for many processes. In the case of universal Kriging, the assumption of a known drift is simply unreasonable because statistical parameters are rarely known *a priori* for natural systems (Papamichail and Metaxa, 1996). Geostatistical techniques have been developed that do not require that the basic statistical characteristics of the rain process (usually the mean and covariance structure) are known, as these are derived from the same data to be interpolated (Bastin and Gevers, 1985; Bacchi and Borga, 1993). Exploitation of data sets coming from different sensors is possible through the so called co-Kriging, which has been used e.g. for integration of rain gauge and radar data (Krajewski, 1987; Delrieu *et al.*, 1988; Seo *et al.*, 1990a,b; Seo 1998a,b). Geostatistical techniques have also been developed for use in case of fractional coverage situations (Barancourt *et al*, 1992; Seo, 1998a).

In this work, the basic statistics (second order description) are derived from interpretation of the information content of the large scale sensor, while the rain-gauge data — which are viewed as completely reliable measurements — are used as quantitative constraints for the generation of the rain field. The geostatistical approach presented in this paper allows reconstruction of the deterministic component of the rain field based on both large scale and local data, together with the associated random component, which is expressed in terms of its statistical parameters. In addition, estimation of the actual reliability of the reconstructed rain field is made possible at the pixel scale, so that uncertainty can be handled and quantified for eventual uses of the derived information.

The same approach has been used to determine the sub-grid distribution of rainfall down to the scales that are required by the hydrological modelling of small to medium size catchments in mountainous terrain. The proposed downscaling algorithm is again constrained to the rain-gauge data whenever these may be available at a finer resolution in time.

Problem statement

Consider the target region as a spatial domain that is discretised, at each time step ΔT, by a regular grid with $M \cdot N$ cells. Each grid cell has total area $\Delta S = \Delta x \cdot \Delta y$. A random variable R_k ($k = 1,...., M \cdot N$) is associated with each cell of the grid to represent the stochastic rainfall process, **R**, at each location. Within the target region, a certain number, *G*, of rain gauge measurements may be available, which provide the constraining rainfall values $\mathbf{r}_l = \left(r_j \quad j = 1,...,G \right)$. Also, from a large scale sensor χ, the conditioning indicators at each grid cell x_k ($k = 1,...., M \cdot N$) can be derived after appropriate interpretation of the original data. The term "indicator" denotes that these are not actual rainfall measurements and their probabilistic interpretation is explained below.

The following assumptions are made:

1. the joint probability distribution of variables $Z_k = \ln(R_k)$ ($k = 1,...., M \cdot N$) is Gaussian — then, the marginal distribution of each variable $Z_k = \ln(R_k)$ is $N\left(\mu_{\ln(R_k)}, \sigma^2_{\ln(R_k)} \right)$; hence the rainfall variable R_k has a log-normal probability distribution $LN(\mu_{R_k}, \sigma^2_{R_k})$;

2. some known relationships hold between the expected value and variance of the random variable R_k and the conditioning indicators x_k in the form:

$$\mu_{R_k} = E[R_k] = f(x_k) \tag{1}$$

$$\sigma^2_{R_k} = E\left[\left(R_k - \mu_{R_k} \right)^2 \right] = g(x_k) \tag{2}$$

In addition, it is assumed that the relationship (1) does not allow that μ_{R_k} takes exactly null values and therefore no cells with $\mu_{R_k} = 0$ can be obtained. Fractional coverage in the final rain field can be obtained by the application of an appropriate threshold level.

3. the covariance matrix of $\mathbf{Z} = \left(Z_1, Z_2,...., Z_{M \cdot N} \right)$, the stochastic log-rainfall process, can be estimated on the basis of the available information; using such statistical structure together with Eqns. (1) and (2), the stochastic

log-rainfall process $\mathbf{Z}$ is completely determined and can be assumed as the *a priori* formulation of the log-rainfall field; the best *a priori* estimate of that field is

$$\hat{\mathbf{z}} = E[\mathbf{Z}] = \left(\mu_{\ln(R_1)}, \mu_{\ln(R_2)}, \ldots, \mu_{\ln(R_{M\cdot N})} \right) \qquad (3)$$

4. the rain-gauge measurements r_j, obtained as area average values after application of suitable variance reduction coefficients to the fluctuations of the actual rain-gauge figures around the expected values μ_{R_j}, are considered as realisations of R_j (and therefore $z_j = \ln(r_j)$ as realisations of Z_j) in all grid cells where at least one rain-gauge is located.

Let:

$$m_{\hat{\mathbf{z}}} = \frac{1}{M \cdot N} \sum_{k=1}^{M \cdot N} \mu_{\ln(R_k)} \qquad (4)$$

and

$$s_{\hat{\mathbf{z}}}^2 = \frac{1}{M \cdot N} \sum_{k=1}^{M \cdot N} \left(\mu_{\ln(R_k)} - m_{\hat{\mathbf{z}}} \right)^2 \qquad (5)$$

be the sample mean and variance of the log-rainfall field $\hat{\mathbf{z}}$ that are derived by the large scale sensor χ. If the expected value $\mu_{\ln(R_k)}$ is constant over the target region, the field is said to be homogeneous and its sample variance obviously is equal to zero.

Let vector $\mathbf{Z}$ be partitioned into two sub-vectors, the first collecting the random variables corresponding to cells where a conditioning realisation is known, and the second containing all other variables, in the form:

$$\mathbf{Z} = \left(\mathbf{Z}_1, \mathbf{Z}_2 \right) = \left(Z_1, \ldots, Z_G, Z_{G+1}, \ldots, Z_{M\cdot N} \right) \qquad (6)$$

where, in general, $G \ll M \cdot N$, as the rain-gauge measurements cover only a small portion of the target region. The expected value and covariance matrix of the stochastic process $\mathbf{Z}$ are given by:

$$E[\mathbf{Z}] = \mu_{\mathbf{Z}} = \left(\mu_{\mathbf{Z}_1}, \mu_{\mathbf{Z}_2} \right) \qquad (7)$$

where the generic element of the covariance matrix is defined as:

$$\sigma_{\ln(R_i),\ln(R_j)}^2 = E\left[\left(\ln(R_i) - \mu_{\ln(R_i)} \right) \cdot \left(\ln(R_j) - \mu_{\ln(R_j)} \right) \right] \qquad (9)$$

and Σ is assumed to be positive definite and symmetric, which implies $\Sigma_{12} = \Sigma_{21}$.

For evaluation of the off-diagonal terms in (8) the correlation coefficient, or normalised covariance, is defined as:

$$\rho_{ij} = \frac{\sigma_{\ln(R_i),\ln(R_j)}^2}{\sigma_{\ln(R_i)} \cdot \sigma_{\ln(R_j)}} \qquad (10)$$

which, by assumption, can be estimated from the available information.

The joint probability density function of $\mathbf{Z}_2$ conditional on the available realisations $\mathbf{z}_1 = \left(z_1, \ldots, z_G \right)$ at the rain-gauge locations can be obtained using standard techniques (Box *et al.*, 1994, p.281) as:

$$p\left(\mathbf{Z}_2 | \mathbf{z}_1 \right) = (2\pi)^{-(M\cdot N-G)} \cdot \left| \Sigma_{22.11} \right|^{-\frac{1}{2}} \cdot \qquad (11)$$

$$\exp\left[-\frac{\left(\mathbf{Z}_2 - \mu_{2.1} \right)^T \cdot \Sigma_{22.11}^{-1} \cdot \left(\mathbf{Z}_2 - \mu_{2.1} \right)}{2} \right]$$

with

$$\mu_{2.1} = \mu_{\mathbf{Z}_2} + \beta_{2.1} \cdot \left(\mathbf{z}_1 - \mu_{\mathbf{Z}_1} \right) = E\left(\mathbf{Z}_2 | \mathbf{z}_1 \right) \qquad (12)$$

$$\Sigma_{22.11} = \Sigma_{22} - \beta_{2.1} \cdot \Sigma_{12} \qquad (13)$$

$$\beta_{2.1} = \Sigma_{21} \cdot \Sigma_{11}^{-1} \qquad (14)$$

where $\Sigma_{22.11}$ is positive definite, and $\beta_{2.1}$ is also known as the regression matrix.

$$\Sigma = \begin{bmatrix} \sigma_{\ln(R_1)}^2 & \cdot & \sigma_{\ln(R_1),\ln(R_G)}^2 & \sigma_{\ln(R_1),\ln(R_{G+1})}^2 & \cdot & \sigma_{\ln(R_1),\ln(R_{M\cdot N})}^2 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \sigma_{\ln(R_G),\ln(R_1)}^2 & \cdot & \sigma_{\ln(R_G)}^2 & \sigma_{\ln(R_G),\ln(R_{G+1})}^2 & \cdot & \sigma_{\ln(R_G),\ln(R_{M\cdot N})}^2 \\ \sigma_{\ln(R_{G+1}),\ln(R_1)}^2 & \cdot & \sigma_{\ln(R_{G+1}),\ln(R_G)}^2 & \sigma_{\ln(R_{G+1})}^2 & \cdot & \sigma_{\ln(R_{G+1}),\ln(R_{M\cdot N})}^2 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \sigma_{\ln(R_{M\cdot N}),\ln(R_1)}^2 & \cdot & \sigma_{\ln(R_{M\cdot N}),\ln(R_G)}^2 & \sigma_{\ln(R_{M\cdot N}),\ln(R_{G+1})}^2 & \cdot & \sigma_{\ln(R_{M\cdot N})}^2 \end{bmatrix} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \qquad (8)$$

Rain field reconstruction and downscaling

The probability density function defined in Eqn. (11) can be used for conditional generation of stochastic rainfall fields based on the available multi-sensor observation.

Besides, as for $E[Z_2 \mid z_1] = \mu_{2.1}$ the multivariate distribution $p\left(\mathbf{Z}_2 \big| \mathbf{z}_1\right)$ has its maximum density, it can be viewed as natural to select $\mathbf{\mu}_{2.1}$ as an estimate of the log-rainfall field in all cells where no rain-gauge measurements are available. Equation (12) can thus be interpreted as a procedure that, on the basis of the available ground based data, allows the *a priori* estimate of the rainfall field in un-gauged cells to be corrected, thus providing the *a posteriori* conditional estimate of the rain field .

The generic element $z_k^* = E\left[Z_k \big| \mathbf{z}_1\right]$ $\left(k = G+1,....,M \cdot N\right)$ can be determined using Eqn. (12), once it is rewritten as

$$z_k^* = \mu_{\ln(R_k)} + \sum_{j=1}^{G} \beta_{kj}\left(z_j - \mu_{\ln(R_j)}\right)$$

$$k = G+1,....,M \cdot N \tag{15}$$

β_{kj} being the generic element of matrix $\beta_{2.1}$.

In this view, z_k^* is the log-rainfall estimate in any generic un-gauged cell and

$$\mathbf{z} = \left(\mathbf{z_1}, \mathbf{z}^*\right) \tag{16}$$

where

$$\mathbf{z}^* = \left[z_k^* \quad k = G+1,.....,M \cdot N\right] \tag{17}$$

is the best *a posteriori* estimate of the log-rainfall field conditional on the available rain-gauge data.

The error variance of such an estimate,

$$\sigma_{e_{z_k}}^2 = E\left[\left(z_k - z_k^*\right)^2\right]$$

$$k = G+1,....,M \cdot N \tag{18}$$

corresponds to the kth diagonal element of $\Sigma_{22.11}$. Obviously, the error variance of the components of $\mathbf{z}_1$ is equal to zero by assumption.

The *explained variance* can be defined as

$$S_{z_k}^2 = \sigma_{Z_k}^2 - \sigma_{e_{z_k}}^2 + s_{\hat{z}}^2$$

$$k = G+1,....,M \cdot N \tag{19}$$

and the *normalised explained variance* as

$$\tilde{S}_{z_k}^2 = \left(1 - \frac{\sigma_{e_{z_k}}^2}{\sigma_{Z_k}^2 + s_{\hat{z}}^2}\right) \tag{20}$$

This coefficient gives a measure of the deterministic part of the precipitation process identified from the proposed procedure, i.e. the signal part that is described by the available rain gauges and the rainfall field estimated by sensor χ .

The rainfall estimate r_k^* $\left(k = G+1,..,M \cdot N\right)$ can be obtained through

$$r_k^* = \exp\left[z_k^* + \frac{1}{2}\sigma_{e_{z_k}}^2\right] \tag{21}$$

$$\sigma_{e_{r_k}}^2 = \exp\left\{2 \cdot \left[z_k^* + \sigma_{e_{z_k}}^2\right]\right\} - \exp\left\{2 \cdot \left[z_k^* + \frac{1}{2}\sigma_{e_{z_k}}^2\right]\right\} \tag{22}$$

and $\mathbf{r} = \left(\mathbf{r_1}, \mathbf{r}^*\right)$ can be defined as the best *a posteriori* estimate of the rainfall field.

To discuss the conceptual characteristics of the estimator in Eqn. (15), let us consider a stochastic log-rainfall field where only one rain gauge measurement $z_p = \ln\left(r_p\right)$ is available for conditioning. In this case matrix $\mathbf{\beta}_{2.1}$, defined by (14), is simply a column vector of the form:

$$\mathbf{\beta}_{2.1} = \left[\frac{\sigma_{\ln(R_k)}}{\sigma_{\ln(R_p)}} \cdot \rho_{kp} \quad k = 2,....,M \cdot N\right] \tag{23}$$

so that

$$z_k^* = E\left[Z_k \big| z_p\right] = \mu_{\ln(R_k)} + \frac{\sigma_{\ln(R_k)}}{\sigma_{\ln(R_p)}} \cdot \rho_{kp} \cdot \left(z_p - \mu_{\ln(R_p)}\right) \tag{24}$$

The variance of the estimator error can be obtained by (13) as:

$$\sigma_{e_{z_k}}^2 = \sigma_{\ln(R_k)}^2 \cdot \left(1 - \rho_{kp}^2\right) \quad \left(k = 2,..,M \cdot N\right) \tag{25}$$

which provide

$$\tilde{S}_{z_k}^2 = \left(1 - \frac{\sigma_{\ln(R_k)}^2 \cdot \left(1 - \rho_{kp}^2\right)}{\sigma_{\ln(R_k)}^2 + s_{\hat{z}}^2}\right) \tag{26}$$

The following observations can be noted:
(a) when the *a priori* variance $\sigma_{\ln(R_k)}^2$ is independent of k, the correction term to be added to the *a priori* estimate

$\mu_{\ln(R_k)}$ (in Eqn. 24) is a function of the correlation coefficient only; thus, assuming ρ_{kp} as a suitable decreasing function of distance, the shorter the distance between cells k and p, the larger such a correction term will be. Obviously, there will be no correction if the two cells are uncorrelated; similar considerations hold for $\sigma_{e_{z_k}}^2$.

(b) when the *a priori* variance $\sigma_{\ln(R_k)}^2$ varies with k, Eqn. (24) implies larger corrections for cells having larger *a priori* uncertainty, i.e. larger $\sigma_{\ln(R_k)}^2$ (at the same distance from p); note that, when $\sigma_{\ln(R_k)}/\sigma_{\ln(R_p)}$ is greater than 1, corrections larger than $\left(z_p - \mu_{\ln(R_p)}\right)$ are possible.

As regards the normalised explained variance, note that:

(a) in the case where the random field is assumed to be homogeneous, i.e. when $s_{\hat{z}}^2 = 0$, it turns out $\tilde{S}_{e_{z_k}}^2 = \rho_{kp}^2$; thus the normalised explained variance

reaches 100% in the constrained cell (where information is totally reliable) and decreases monotonically with ρ_{kp} (i.e. as the distance increases, following the above assumption on the function ρ_{kp}).

(b) when the random field is not homogeneous, the corrections to the estimates $\left(z_k^* - \mu_{\ln(R_k)}\right)$ for unconstrained cells are increasingly reliable as $s_{\hat{z}}^2$ increases.

In Fig. 1 the pattern of the *a posteriori* estimate z_k^*, the variance of the estimator error $\sigma_{e_{z_k}}^2$, and the normalised explained variance $\tilde{S}_{e_{z_k}}^2$ are depicted column-wise for a sample field with dimensions $M=20$ and $N=50$ in the three different cases where the information from the large scale sensor χ only (first row), the large scale sensor and one conditioning gauge (second row), and the conditioning gauge with a homogeneous base field (third row) are used to obtain the conditional rainfall field.
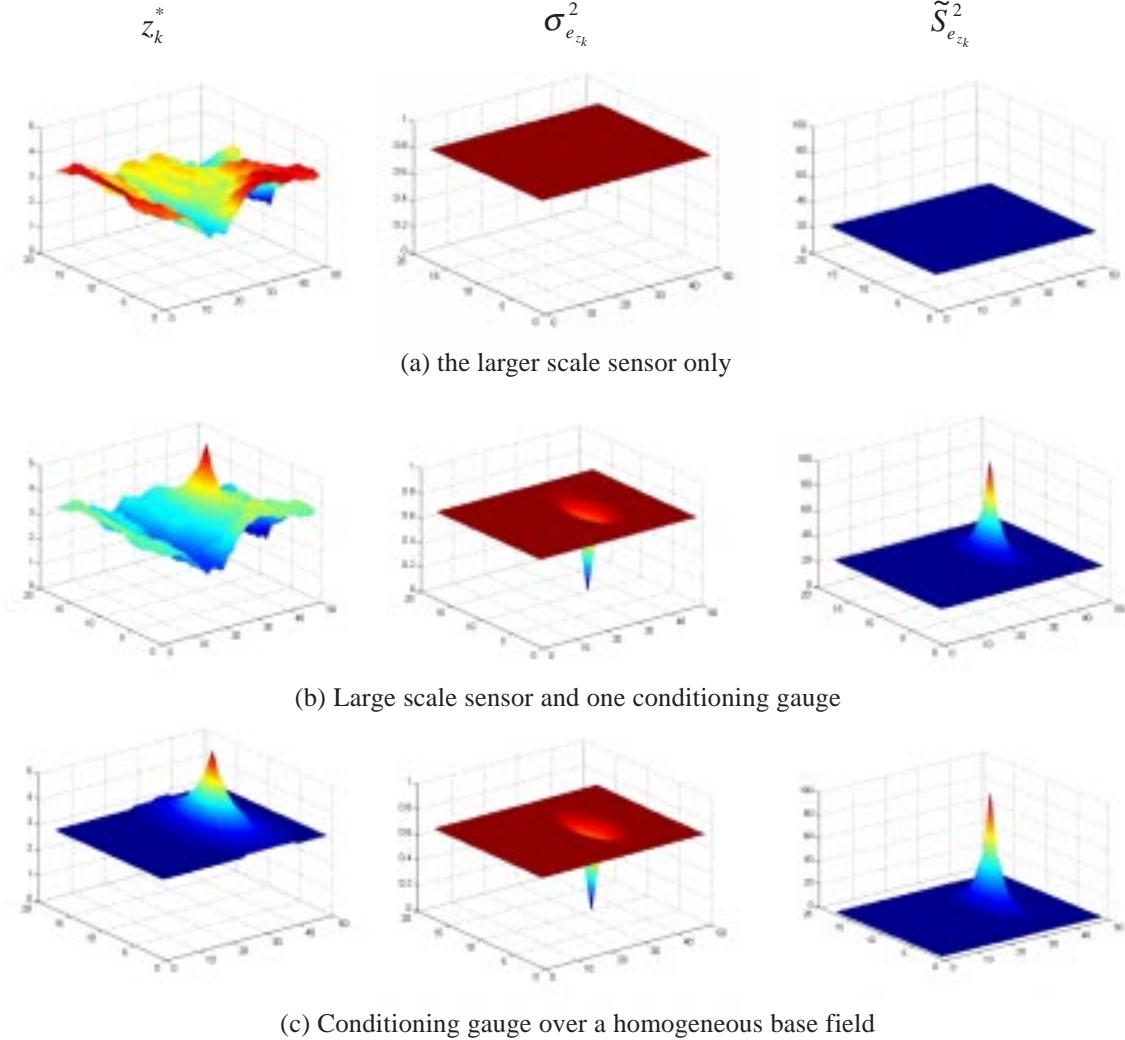
$$z_k^* \qquad\qquad \sigma_{e_{z_k}}^2 \qquad\qquad \tilde{S}_{e_{z_k}}^2$$

(a) the larger scale sensor only

(b) Large scale sensor and one conditioning gauge

(c) Conditioning gauge over a homogeneous base field

Fig. 1. *Comparison of the spatial variability of* z_k^*, $\sigma_{e_{z_k}}^2$ *and* $\tilde{S}_{e_{z_k}}^2$ *in a sample case where a) the larger scale sensor only, b) the larger scale sensor and one conditioning gauge, and c) the conditioning gauge over a homogeneous base field, are used for reconstruction field.*

Consider now a given sub-grid partition of the target region both in time, using $\Delta T^{'} = \Delta T / \tau$, and space, using $\Delta S^{'} = \Delta x \cdot \Delta y / \delta^2$, which leads to the new grid dimensions $m = \delta \cdot M$ and $n = \delta \cdot N$ with $\delta > 1$ and $\tau > 1$ (see Fig. 2). Again, a random variable R_h $(h = 1,...., m \cdot n)$ can be associated with each sub-grid cell to represent the stochastic rainfall process, **R,** at each location. The rain-gauge measurements are now attributed to the corresponding sub-grid cells using the proper area reduction factor, which is a function of the cell size.

The expected value and variance of the random variable R_h are now derived from :

$$\mu_{R_h} = E[R_h] = \frac{r_k}{\tau} \quad \forall h \in P(k) \tag{27}$$

$$\sigma^2_{R_h} = CV^2 \cdot \mu^2_{R_h} \quad \forall h \in P(k) \tag{28}$$

where r_k is the kth element of $\mathbf{r} = (\mathbf{r_1}, \mathbf{r^*})$, $P(k)$ = {sub-grid cells belonging to the kth cell at larger scale} and CV is the coefficient of variation.

The covariance matrix of $\mathbf{Z} = (Z_1, Z_2,...., Z_{m \cdot n})$— the new log-rainfall process at sub-grid scale — can also be evaluated, so that the *a posteriori* estimate of the rain field at the larger scales acts here as the *a priori* estimate of the sub-grid scale process in all ungauged locations. It is also possible to define the stochastic log-rainfall process

$$\mathbf{Z} = (\mathbf{Z_1}, \mathbf{Z_2}) = (Z_1,...., Z_G, Z_{G+1},....., Z_{mn}) \tag{29}$$

and, therefore, obtain the joint probability distribution of $(\mathbf{Z_2} | \mathbf{z_1})$, conditional on the G rain-gauge measurements at sub-grid scale.

To calibrate the mappings defined by Eqns. (1) and (2) and validate the proposed procedure, it would be necessary to run several experiments with varying model parameters and finally compare the estimated conditional rainfall fields with that observed. Unfortunately, such rigorous validation is never possible, due to the common lack of any direct measurements of spatially distributed rain fields. Indirect methods are, therefore, usually addressed to perform both calibration and validation exercises.

One possible technique for the evaluation of random field estimates, known as the *jacknife* method, is based on alternately suppressing one out of the available conditioning sensors at the ground and evaluation, for each generated conditional rain field, of the marginal probability distribution of the random variable R_j at the corresponding grid cell. This is compared with the measured value at that cell and the procedure is repeated recursively for all the G nodes where a rain gauge is located (see Papamichail and Metaxa, 1996).

The procedure for conditional generation is therefore evaluated on the basis of the following performance index:

$$MSE = \frac{1}{G} \sum_{j=1}^{G} e_j \tag{30}$$

where

- $e_j = (r_j - r_j^*)^2$ is the error reduction at each cell j where a sensor is located;
- r_j is the "ground truth" measured at cell j;
- r_j^* is the rainfall estimate at location j when measurement r_j is neglected;

The calibration parameters are contained in the functions $f(x_k)$ and $g(x_k)$, used to transform the measurements provided by the large scale sensor χ into the *a priori* mean and variance of the random field, and in the selected model for the correlation function ρ_{ij}. Application of the validation procedure to the reconstruction of a real rain field is described later with reference to the case study analysed.
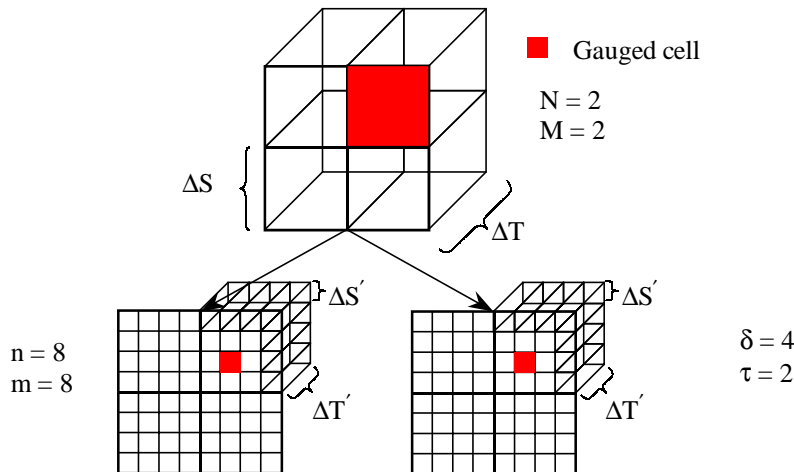

Fig. 2. *Conceptual scheme of the downscaling procedure*

A case study

For the storm event on May 28th 1998 in northern Italy, the meteorological situation at synoptic scale is well described by the Meteosat imagery in the thermal infrared (IR) band. A sequence of such images is provided in Fig. 3 corresponding to the core of the event over the Liguria region of Italy (see zoomed boxes).

The event analysed is embedded in a large-scale synoptic disturbance associated with a pressure low located over north-west Europe and producing a flow of humid air rising
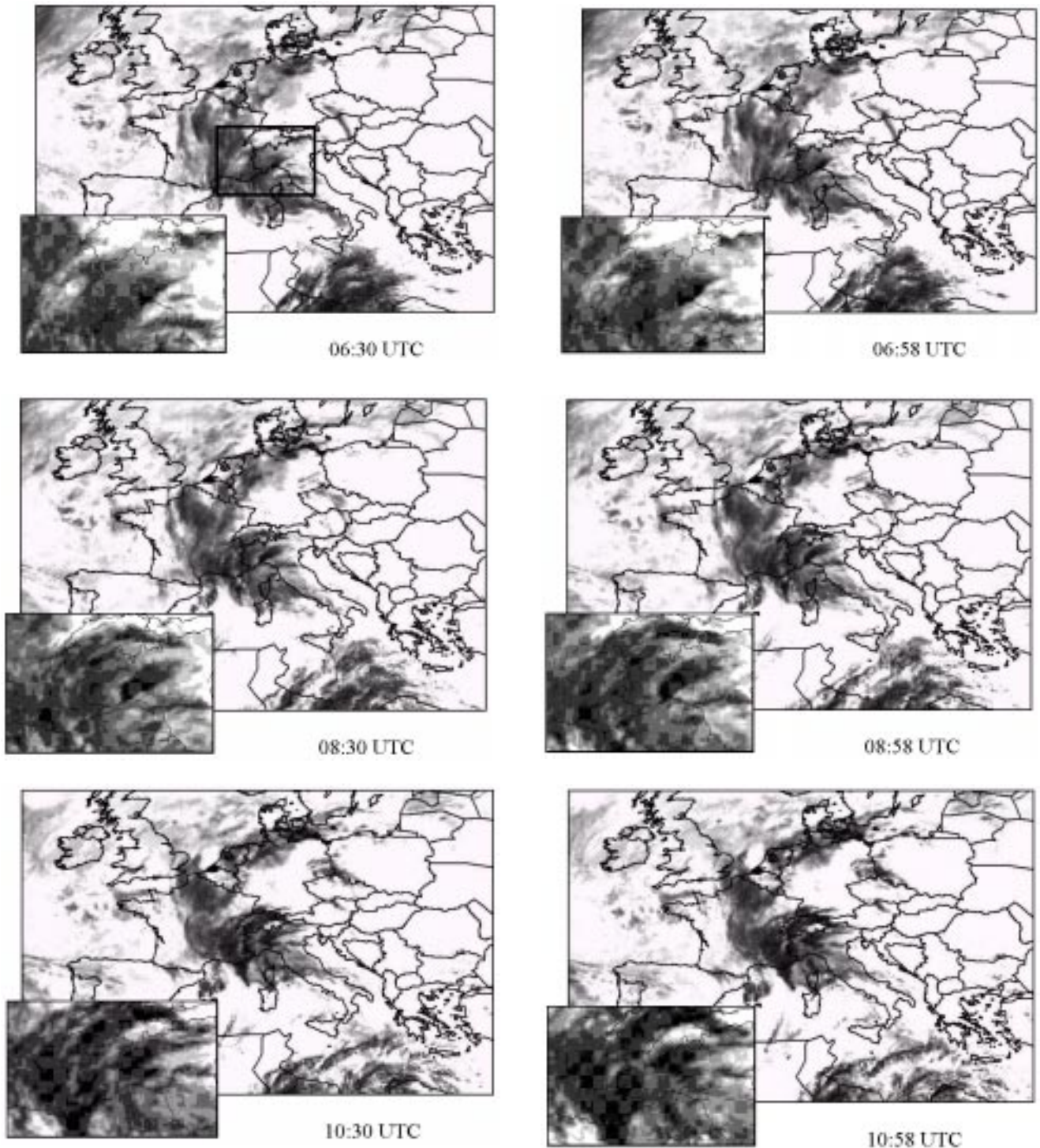

Fig. 3. *Sequence of IR Meteosat images from 06.30 to 10.58 UTC for the event of May 28th 1998 over Europe and zoomed over northern Italy where the study area is located*

from the south of the Mediterranean Sea towards the coastal European regions. Such meteorological conditions persist for a few days due to the blocking mechanism caused by a wide high located over Eastern Europe.

During the evolution of the event, humid airflows impinge on the west coast of Liguria causing strong low level convergence at local scales and intense precipitation over the study area in the time window 05:00 to 16:00 UTC, .

Low level convergence is associated with areas where the highest precipitation is observed, with values up to 100 mm in six hours, with the maximum rain rates at the ground being recorded at Mele, reaching 47 mm h^{-1}. The location of the study area and the distribution of the available rain gauges are shown in Fig. 4 . The space-time distribution of rainfall has been determined over the study area (about 25 000 km^2) by using 44 active rain-gauges from the local Hydrographic Service, most of them providing cumulated rainfall in hourly intervals.

The rain field has been discretised using a regular mesh of 50×18 pixels, each a square of side 5 km. To be coherent with the temporal resolution of the rain-gauge information, hourly data have been used for reconstruction of the rain field at the original space scale (therefore after aggregation of the half-hourly original Meteosat images). For the sake of brevity, only results referring to a single time slot (09:00 UTC) are shown here.

Similarly, rain-gauge rainfall fluctuations have been made coherent with the spatial resolution of the Meteosat images by application of appropriate area reduction factors to comply with the adopted pixel size.

To apply the proposed procedure for reconstruction of the precipitation field the following parameters must be specified:

1. second-order description (μ_{R_k}, $\sigma^2_{R_k}$) of the marginal p.d.f. $LN(\mu_{R_k}, \sigma^2_{R_k})$ for $k = 1,....,900$;
2. correlation coefficients ρ_{ij} $(i, j = 1,......,900)$ for the precipitation field.

Derivation of the parameters of the marginal p.d.f. from the rain-gauge data alone is scarcely reliable as the available stations are quite sparse over the region and not uniformly distributed. It seems more reasonable to estimate such parameters from the large-scale sensor information, which is represented in this case study by the Meteosat imagery. It is well known that IR Meteosat data represent the top cloud temperature and a suitable function is needed for mapping the radiance temperature data (grey levels) into the corresponding precipitation figures (e.g. Adler and Negri, 1988).
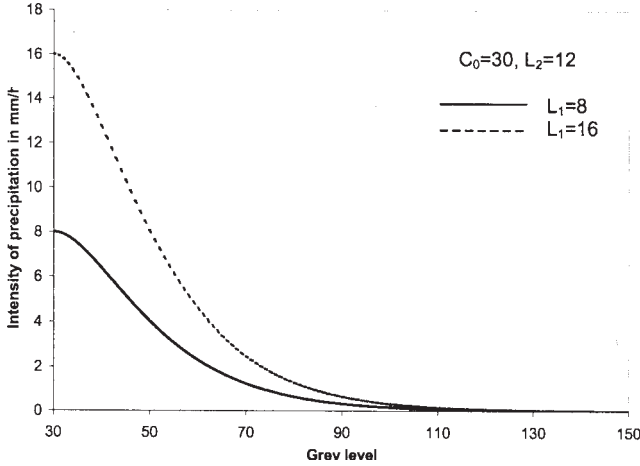

Fig. 4. *Location of the study area and distribution of the available rain gauges. The dotted square denotes the area where downscaling is applied.*

Fig. 5. *Mapping from Meteosat data (grey levels) to rain mean value using Eqn. 31 with sample parameters.*

The mapping of the Meteosat measurements (for each single pixel) to the mean values $\mu_{R_k} = E[R_k]$ has been obtained by using the following relationship:

$$f(x_k) = L_1 \cdot \left(\frac{x_k - C_0}{L_2} + 1 \right) \cdot \exp\left(-\frac{x_k - C_0}{L_2} \right) \quad (31)$$

where C_0 is the grey level threshold, L_2 is the shape parameter and L_1 is a suitable coefficient (see Fig. 5).

The above relationship may be considered as a suitable continuous refinement of the empirical rule proposed by Adler and Negri (1988). A constant *CV* (equal to 1) has been assumed, so that

$$g(x_k) = (CV \cdot f(x_k))^2 \quad (32)$$

As for the dependence of the spatial correlation coefficient on the distance, the following decay model has been used:

$$\rho_{ij} = \exp\left(-\frac{d_{ij}}{\gamma} \right) \quad (33)$$

with d_{ij} the Euclidean distance between pixels *i* and *j*.

Calibration of the coefficients in functions (31) and (33) has been performed by estimating the MSE for a series of sets of parameter values. In Table 1 a synthesis of such results obtained is reported, as regards the calibration of parameter γ.

Table 1. Results for the case study analysed with $C_0 = 30$, $L_1 = 8$ and $L_2 = 12$

	$\gamma = 0$	$\gamma = 1$	$\gamma = 2$	$\gamma = 3$	$\gamma = 4$	$\gamma = 8$
MSE	3032	2405	**1935.1**	2134.3	3056	3537.2

From repeated experiments the set of parameter values which allows the most reliable reconstruction of the rainfall field corresponds to $C_0 = 30$, $L_1 = 8$, $L_2 = 12$ and $\gamma = 2$. The reconstructed rain field **r** obtained by integration of the data from the two sensors (with parameters $C_0 = 30$, $L_1 = 8$, $L_2 = 12$ and $\gamma = 2$) and the map of standard deviations $\sigma_{e_{r_k}}^2 \ (k = 1,....,M \cdot N)$ of the random variables for the reconstructed field are shown in Figs. 6 and 7. In Fig. 8 the reliability map is presented for $\widetilde{S}_{z_k}^2 \ (k = 1,....,M \cdot N)$ the derived rain field.

The observed ground truth and the estimate of the rain figures are compared in Figs. 9 to 12. In Fig. 9 rain estimates from Meteosat data, using (31) with parameters $C_0 = 30$, $L_1 = 8$, $L_2 = 12$, and rain-gauge data are compared. In Fig. 10 the rain figures obtained with the proposed procedure using a base homogeneous field (the mean value of rain-gauge measurements) are compared with the rain-gauge observations.

In Fig. 11 the rain figures obtained with the proposed procedure using Meteosat data with the set of parameters selected ($C_0 = 30$, $L_1 = 8$, $L_2 = 12$ and $\gamma = 2$), are compared with the available ground truth, while in Fig. 12 the same graph is presented where a single outlier — which can be due to some measurement error — is eliminated.

The proposed procedure seems to provide the best performances when integration of Meteosat data and rain-gauge measurements is addressed. This reflects the obvious observation that the additional information contained in Meteosat data does contribute to a better rain field reconstruction than the use of rain-gauge measurements alone (homogeneous field). The scatterplot in Fig. 12 is obtained by eliminating the contribution of the station of Mele (see Fig. 2); note that the correlation obtained is about 0.8. The measurement in Mele can be considered an outlier if compared to the other instruments. Besides, this station is out of the influence zone of the other instruments and, when its contribution is neglected, the rain field information in that region only comes from the Meteosat images. It seems, therefore, reasonable to eliminate this information in the validation phase.

Downscaling of the reconstructed rain field has been possible thanks to the availability of high-resolution data from a sub-set of the available rain-gauges. Data with resolution in time of 15 minutes have been used and the sub-grid scale distribution of rainfall has been determined down to a spatial scale of about 3 km² (1/9 of the original scale). The sub-region already identified in Fig. 6 where the highest rain rates were observed in the case study analysed, has been resized with a finer grid and the mean values at the former scale have been used for determination of the parameters of the new joint probability density
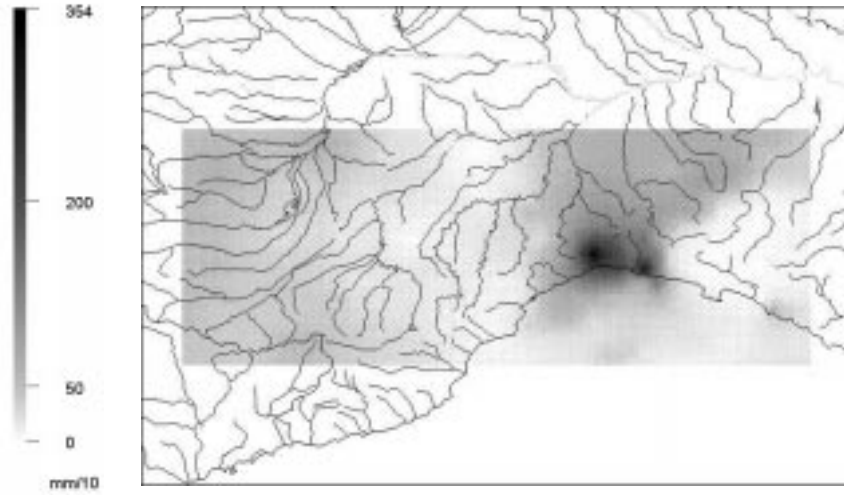
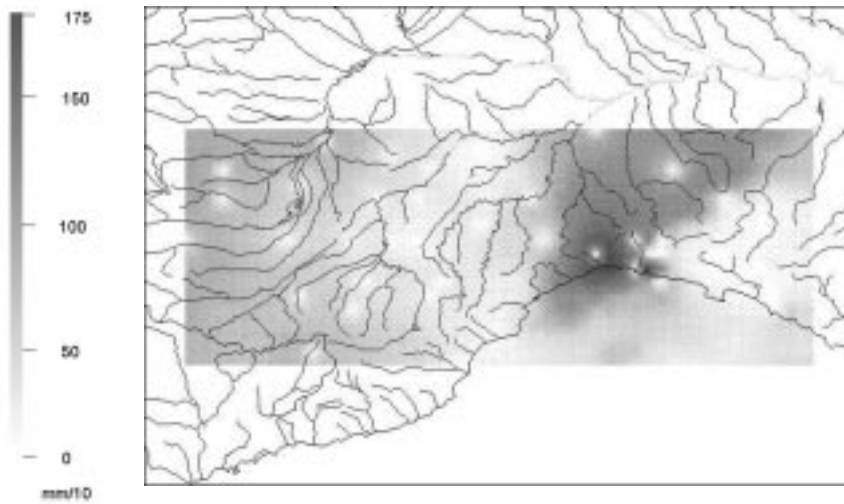Fig. 6. *Rainfall field **r** from Meteosat data constrained to rain-gauge*

Fig. 7. *Standard deviation* $\sigma^2_{e_{r_k}}$ $\left(k=1,.....,M\cdot N\right)$ *of rain fluctuations with respect to the Meteosat field constrained to rain-gauge*
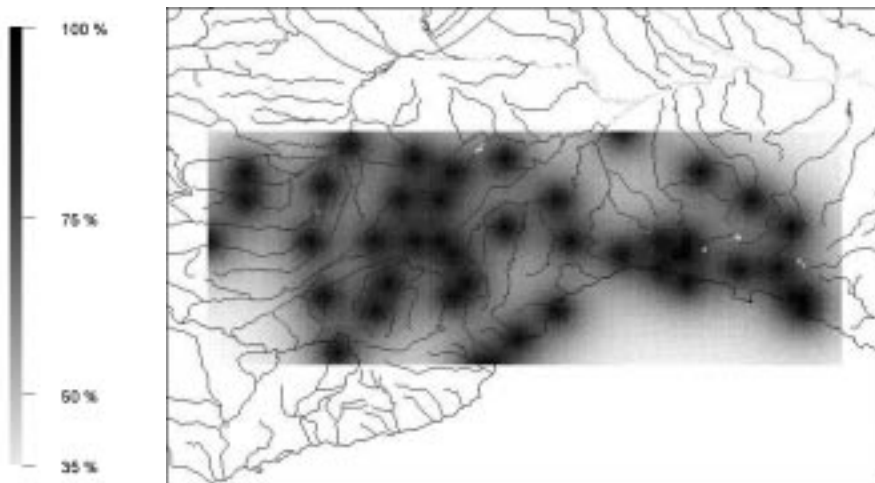

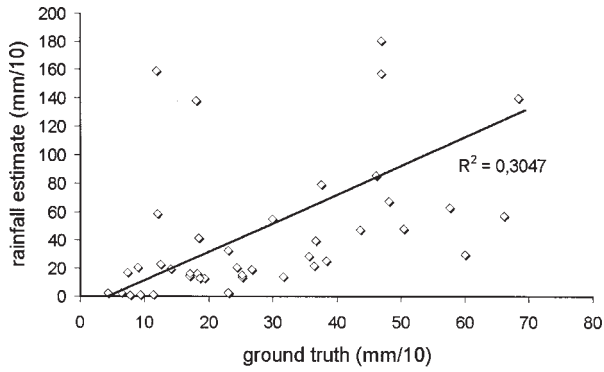Fig. 8. *Reliability map* $\tilde{S}^2_{z_k}$ $\left(k=1,....,M\cdot N\right)$ *of the rain field*

Fig. 9. *Meteosat data and ground truth*
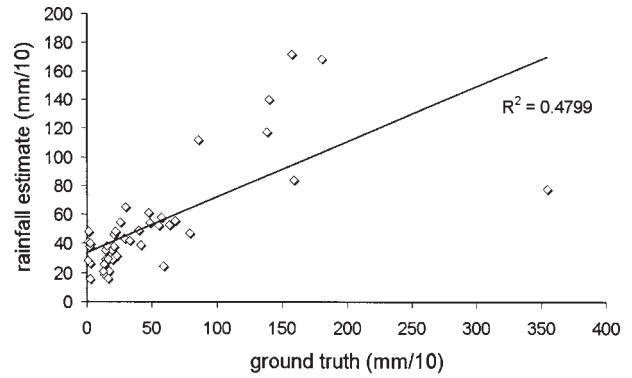
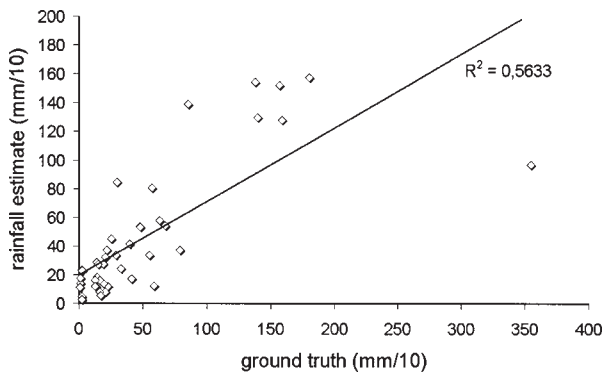Fig. 10. *Estimate values using homogeneous field and ground truth*

Fig. 11. *Estimate values using Meteosat and ground truth*
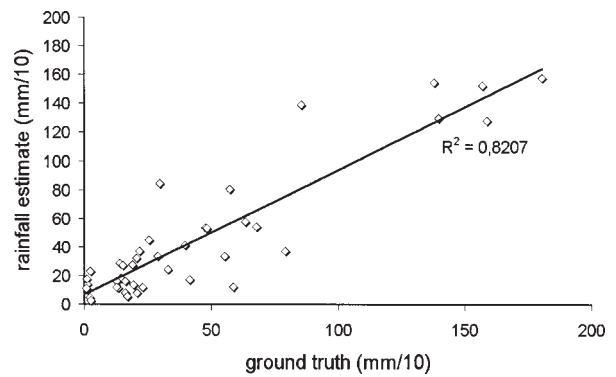

Fig. 12. *Same as in Fig. 11 with a single outlier eliminated*
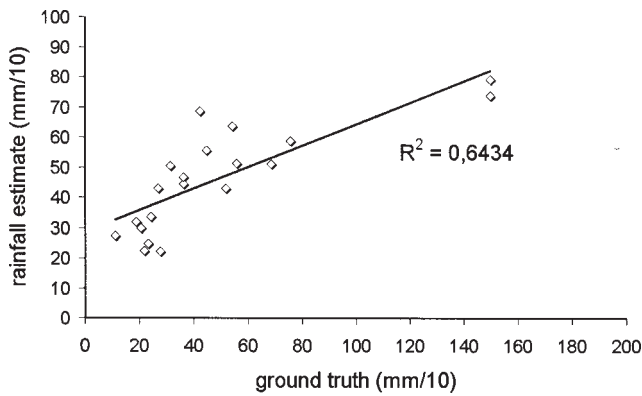

Fig. 13. *Downscaling results and ground truth*

function at the finer scale. The same procedure is then used for determination of the rainfall field at a finer scale, constrained at point values (five rain-gauges available) at the corresponding new time scale.

The scatterplot in Fig. 13 is analogous to those in Figs. 9–12, though the scarce number of rain-gauges available reduces the significance of the validation exercise in this context.

Obviously, downscaling increases the explained variance

of the process with respect to that obtained at the original scale. However, the variance of the rain field at finer scale is greater than the variance at the original scale. Therefore, with respect to the total variance, the explained variance decreases provided any additional information at finer scales (e.g. radar rain fields) is included.

In the first column of Fig. 14 the rainfall fields at 15-minute intervals obtained after downscaling the 1-hour field are shown while in the second column, maps of the standard deviation for the random variables of the reconstructed fields are reported.

Conclusions

Rain field reconstruction and downscaling procedures have been proposed for integration of rainfall data coming from different sources. Each of them provides different information in terms of reliability and scale. The procedures presented lead to the following results:

(1) mapping of the expected values of the reconstructed rainfall field at the same resolution of the available large scale information, using point data as a constraint;
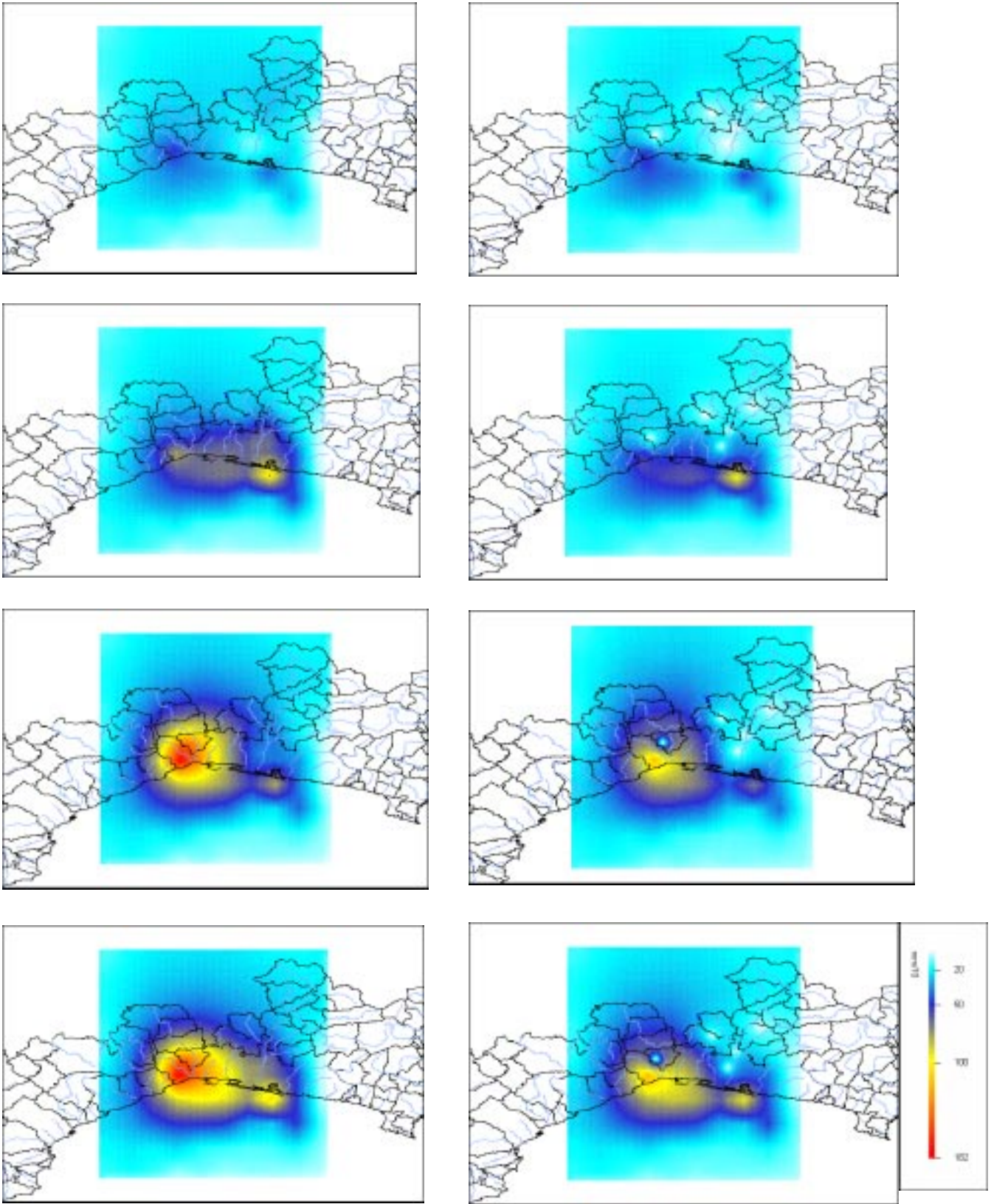
(2) mapping of the residual variance of the precipitation

Fig. 14. *Downscaling of the reconstructed rain field (on the left) and associated standard deviation maps (on the right) for the case study analysed*

process, i.e. a measure of the information content of the reconstructed rain field;

(3) quantitative evaluation of the reliability of the rainfall field at each cell of the grid.

The probabilistic information obtained about the rainfall field can be used for stochastic generation of different scenarios, with the purpose of determining the risk associated with specific space-time rainfall extremes in the region of concern, conditional on the available information on the specific rain event.

The procedure developed is based on the definition of various functions, which are used for conversion of the original information (as provided by the available sensors) into quantitative information about rainfall intensity, or better into suitable parameters characterising the probabilistic distribution of the rainfall field. Obviously, validation of such functions and their parameters is needed by means of an extensive comparison with actual rainfall measurements in a large number of case studies under various climatological conditions.

Throughout the paper we have assumed a constant CV over the whole rain field. The method can be modified easily to account for variable CV, provided some physical meaning is associated with the variability of these parameters. One possibility is to investigate the role of local precipitation-enhancing mechanisms in the study area and to model them e.g. orographic forcing through suitable patterns (maps) of CVs. This could be valuable in the downscaling frameworks as it allows one to include small-scale sources of variability in the process which are not yet captured by the large-scale observations or the individual point measurements of rainfall.

Further important features of space-time rainfall are not fully resolved by the proposed rain field reconstruction model, such as the bursting nature of rainfall with fractional coverage and alternate rain / no rain periods. In the approach presented, the modelling of such features is limited and reduced to the pattern that can be observed by the available rain-gauges. In the case of a sparse network, the fractional pattern between contiguous rain-gauges cannot be accounted for in this model. Future developments, possibly in synergy with intermittent models of space-time rainfall (e.g. Lanza, 2000) may allow integration of such modelling capabilities. This should proceed in parallel with the development of enhanced observing systems, such as radar networks, to obtain a measure of the space-time characteristics of fractional rain fields.

References

Adler, R.F. and Negri, A.J., 1988. A satellite infrared technique to estimate tropical convective and stratiform rainfall, *J. Clim. Appl. Meteorol.,* **27**, 30–51.

Bacchi, B. and Borga, M., 1993. Spatial correlation patterns and rainfall fields analysis, *Excerpta,* **7**, 7–40.

Bastin, G. and Gevers, M., 1985. Identification and optimal estimation of random fields from scattered point-wise data, *Automatica*, **21**, 139–155.

Barancourt, C., Creutin, J.D. and Rivoirard, J., 1992. A method for delineating and estimating rainfall fields, *Water Resour. Res*., **28**, 1133–1144.

Barrett, E.C. and Beaumont, M.J., 1994. Satellite rainfall monitoring: an overview, *Remote Sensing Reviews,* **11**, 23–48.

Barrett, E.C. and Martin, D.W., 1981. *The use of satellite data in rainfall monitoring*, Academic Press, London.

Box, G.E.P., Jenkins, G.M. and Reinsel, G.C., 1994. *Time Series Analysis (Forecasting and Control)*, Prentice-Hall, New Jersey.

Delrieu, G., Bellon, A. and Creutin, J.D., 1988. Estimation de lames d'eau spatiales à l'aide de données de pluviomètres et de radar météorologique - Application au pas de temps journalier dans la région de Montréal, *J. Hydrol*., **98**, 315–344.

Krajewski, W.F., 1987. Cokriging radar-rainfall and rain gage data, *J. Geophys. Res*., **92**, 9571–9580.

Lanza, L.G., 2000. A conditional simulation model of intermittent rain fields, *Hydrol. Earth Sys. Sci.*, **4**, 173–183.

Papamichail, D.M. and Metaxa, I.G., 1996. Geostatistical analysis of spatial variability of rainfall and optimal design of a rain gauge network, *Water Resour. Manag*., **10**, 107–127.

Seo, D.J., Krajewski, W.F. and Bowles, D.S. 1990a. Stochastic interpolation of rainfall data from rain gages and radar using cokriging – 1 – Design of experiments, *Water Resour. Res*., **26**, 469–477.

Seo, D.J., Krajewski, W.F., Azimi-Zonooz, A. and Bowles, D.S., 1990b. Stochastic interpolation of rainfall data from rain gages and radar using cokriging – 2 – Results, *Water Resour. Res.*, **26**, 915–924.

Seo, D.J., 1998a. Real-time estimation of rainfall fields using rain gage data under fractional coverage conditions, *J. Hydrol*., **208**, 25–36.

Seo, D.J., 1998b. Real-time estimation of rainfall fields using radar rainfall and rain gage data, *J. Hydrol*., **208**, 37–52.

Sevruk, B. and Lapin, M., 1993. *Precipitation Measurement and Quality Control*. Proc. Symp. on Precipitation and Evaporation, Slovak Hydromet. Institute, Bratislava and Swiss Federal Inst. of Technology, Dep. of Geography, Zurich.