Verification of Quantitative Precipitation Forecasts via Stochastic Downscaling

Verification of Quantitative Precipitation Forecasts via Stochastic Downscaling

ELISA BRUSSOLO
CIMA Research Foundation, Savona, Italy

JOST VON HARDENBERG
ISAC-CNR, Lecce, Italy

LUCA FERRARIS
CIMA Research Foundation, Savona, and DIST, University of Genova, Genoa, Italy

NICOLA REBORA
CIMA Research Foundation, Savona, Italy

ANTONELLO PROVENZALE
ISAC-CNR, Torino, Italy

(Manuscript received 16 November 2007, in final form 11 February 2008)

ABSTRACT

The use of dense networks of rain gauges to verify the skill of quantitative numerical precipitation forecasts requires bridging the scale gap between the finite resolution of the forecast fields and the point measurements provided by each gauge. This is usually achieved either by interpolating the numerical forecasts to the rain gauge positions, or by upscaling the rain gauge measurements by averaging techniques. Both approaches are affected by uncertainties and sampling errors due to the limited density of most rain gauge networks and to the high spatiotemporal variability of precipitation. For this reason, an estimate of the sampling errors is crucial for obtaining a meaningful comparison. This work presents the application of a stochastic rainfall downscaling technique that allows a quantitative comparison between numerical forecasts and rain gauge measurements, in both downscaling and upscaling approaches, and allows a quantitative assessment of the significance of the results of the verification procedure.

1. Introduction

The assessment of hydrogeological risk in small catchments requires the availability of skillful, high-resolution quantitative precipitation forecasts (QPFs), with a lead time of at least 24–48 h (Ferraris et al. 2002; Siccardi et al. 2005), which is usually provided by the output of a limited area circulation model (LAM; Bacchi et al. 2003). Within this framework, the verification of the model prediction skill represents an essential step for the development of efficient operational forecasting chains.

Different observational sources are available for this verification, such as radar and satellite observations and rain gauge data. Dense networks of rain gauges, in particular, provide direct observations of precipitation with high resolution in time at specific locations in space (see, e.g., Colle et al. 1999). There is, however, a scale gap, or inconsistency, between the point measurements provided by rain gauges and the spatial scales resolved by LAMs, which are typically of the order of 5–10 km. In addition, meteorological forecasts cannot be considered reliable at their nominal resolution (Patterson and Orszag 1971; Harris et al. 2001), so that a comparison with forecast fields smoothed on even larger scales is required.

The two main methods used to bridge the scale gap are smoothing (averaging) the rain gauge data at the scale of the forecast (point-to-area upscaling) or, vice versa, interpolating the forecast to the gauge positions (area-to-point downscaling).[1] There are two main dif-

Corresponding author address: Jost von Hardenberg, ISAC-CNR, Strada Provinciale Lecce-Monteroni Km. 1.2, 73100 Lecce, Italy.
E-mail: j.vonhardenberg@isac.cnr.it

[1] Alternative, multiscale, methods have also been developed, such as scale recursive estimation (Tustison et al. 2003).

DOI: 10.1175/2008JHM994.1

ficulties with these comparisons: On one hand, averaging small numbers of rain gauges at the scale of the forecast leads to limited representativeness and large uncertainties (Bras and Rodriguez-Iturbe 1976; Ciach and Krajewski 1999). Conversely, smooth interpolation of the forecast grid points at the position of the rain gauges underestimates the variability of the local precipitation at these points. Both problems are enhanced by the intermittent nature of precipitation (Hendrick and Comer 1970; Zawadski 1973).

To cope with these "representativeness errors," Tustison et al. (2001) suggested the use of stochastic models capable of representing the statistical properties of precipitation at multiple scales. In particular, rainfall downscaling techniques (Droegemeier et al. 2000; Ferraris et al. 2003) allow us to derive, from a single precipitation forecast with limited spatial and temporal resolution, higher-resolution stochastic ensembles of precipitation fields, which reproduce the statistical properties of observed precipitation at small scales, while conserving the large-scale features of the forecast. As such, rainfall downscaling provides a stochastic interpolation technique that is able to correctly reproduce, in a statistical sense, the small-scale structure of precipitation and can thus represent a fundamental aid for QPF verification.

As a specific example of this application, we use stochastic downscaling to verify precipitation forecasts issued by the Consortium for Small-scale Modeling's limited-area nonhydrostatic model (COSMO-LAMI; Marsigli et al. 2001; Montani et al. 2003) over an area covering northwestern Italy and compare them with direct measurements from a dense network of rain gauges, for three different events that took place in 2005 and 2006. To this end, we follow both a downscaling and an upscaling approach, based on the Rainfall Filtered Autoregressive Model (RainFARM) stochastic downscaling method introduced by Rebora et al. (2006b).

The rest of this paper is organized as follows. In section 2 we describe the forecast model, the observational data, and the RainFARM downscaling technique. In section 3 we bridge the scale gap by directly interpolating the forecasted precipitation fields to the positions of the rain gauges by means of RainFARM. In this way, a stochastic ensemble of forecasts is created at each rain gauge position and its statistical agreement with the observations is explored. In section 4, we take the upscaling approach and aggregate the observational data and LAM forecasts at large scales. In this case, rainfall downscaling is used to generate stochastic ensembles of high-resolution forecasts that allow us to estimate the sampling error and the significance of the results. Section 5 gives our conclusions and perspectives.

2. Data and models

a. The precipitation forecasts

In this paper we consider QPF fields issued by the COSMO-LAMI model (Marsigli et al. 2001; Montani et al. 2003), a regional version of the Lokal Model (Doms and Schattler 1998), used for operational and research forecasting in Italy. Three precipitation events over northwestern Italy, which started, respectively, at 1200 UTC on 10 April 2005, 1200 UTC on 7 September 2005, and 0200 UTC on 14 September 2006, and whose occurrence was predicted by the numerical model, are used as sample cases. These events were selected among eight events in the period 2005–06, which led to serious civil protection alarms in the area of interest, choosing those with the largest intensity and available number of rain gauge measurements. For each event, the maximum forecast lead time provided by COSMO-LAMI was 72 h, starting at 0000 UTC on, respectively, 10 April 2005, 7 September 2005, and 13 September 2006. To mimic the operational conditions, we focus on the forecasts that were issued between 12 and 24 h before the start of the intense precipitation event. Each forecast field consists of precipitation accumulated over 3 h, with a spatial resolution of 7 km.

These three case studies presented the following synoptic situations:

- *10–12 April 2005:* A low pressure center over the northern part of the Mediterranean Sea was maintained for 2 days by an Arctic airflow. The cyclone led to convective instability with intense bursts of precipitation over the northern regions of Italy.
- *7–9 September 2005*: Strong convective activity over southeastern France and northwestern Italy was due to a low pressure center positioned between the Iberian Peninsula and the western Mediterranean Sea. The situation over the northern Italian regions was characterized by heavy precipitation due to organized storm structures.
- *13–15 September 2006*: A deep low pressure structure extended from the British Islands to the Sahara with an intense warm moist flow over all of Italy. This scenario brought unstable conditions with heavy precipitation, especially over the northern regions of Italy.

b. The observational network

The verification dataset consists of measurements from a dense regional network of rain gauges provided by the Italian national Civil Protection Department.

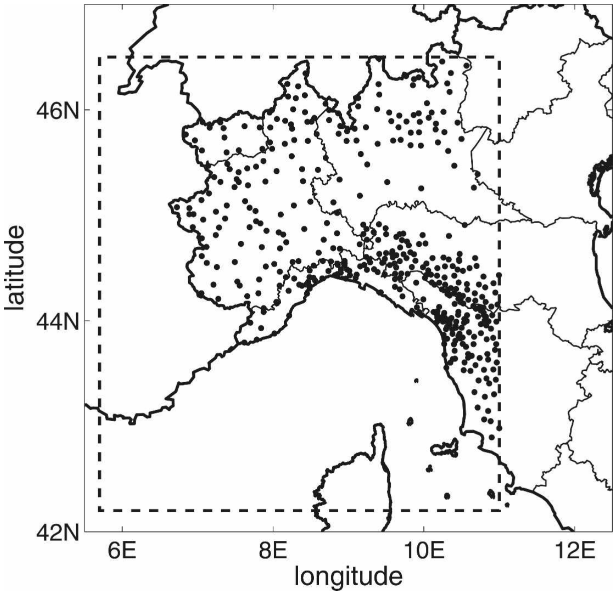

FIG. 1. The study area and the network of the rain gauges used for the verification of the QPFs. Each point indicates the location of a rain gauge.

This is an extremely dense operational network, never used before in its entirety in a verification study. The study area is a window of 448 km × 448 km covering northwestern Italy, in which a large number of rain gauges with instrumental resolution of 0.2 mm is available for the periods of interest. Figure 1 illustrates the position of the rain gauges and the study area. For each case study, we consider only rain gauges for which a complete, uninterrupted, 72-h-long time series is available (10–12 April 2005, 284 rain gauges; 7–9 September 2005, 200 rain gauges; 13–15 September 2006, 545 rain gauges) and focus on the total rainfall volume and on the maximum hourly precipitation intensity over the full duration of the time series (72 h).

c. *The stochastic downscaling model*

Stochastic downscaling techniques are used to generate, from forecast fields with coarser resolution, a high-resolution stochastic ensemble of precipitation fields that reproduce the statistical properties of the observed precipitation while conserving the features of the large-scale forecast (Rebora et al. 2006b). A wide range of stochastic downscaling techniques is reported upon in the literature, ranging from point-process models (Waymire et al. 1984) to fractal cascades (Lovejoy and Mandelbrot 1985; Perica and Foufoula-Georgiou 1996; Menabde et al. 1997) to autoregressive models (e.g., Bell 1987). The skill of such models was compared in Ferraris et al. (2003), finding similar ability (or inability) among the different methods to reproduce one-point probability distribution functions, two-point correlations in space and time, and the spectrum of generalized dimensions of observed precipitation.

In the following, we use the recently developed Rainfall Filtered Autoregressive Model (RainFARM; Rebora et al. 2006a,b) to downscale precipitation forecasts. This model is based on the nonlinear transformation of a linearly correlated Gaussian stochastic field, obtained by extrapolating the large-scale spatiotemporal power spectrum of the forecast to the small, unresolved scales. In this approach, we define the spatial and temporal scales, L_0 and T_0, above which the QPF is assumed to be reliable and is kept as it is. Below these scales, stochastic realizations of the precipitation field are obtained by (i) inverting a power spectrum with random Fourier phases, which produces a Gaussian field, and (ii) taking the exponential of the field generated in step (i), possibly with the inclusion of a small precipitation threshold below which the value of precipitation is set to zero. The small-scale form of the power spectrum is obtained by small-scale extrapolation of the QPF spectrum measured on scales larger than L_0 and T_0, assuming a power-law spectral shape. The "reliability scales" L_0 and T_0 represent the only free parameters of the model. All other parameters, such as the spectral slopes used to extrapolate the spectrum, are directly estimated from the large-scale properties of the fields.

With this approach, we are able to generate an ensemble of high-resolution stochastic precipitation fields (each member of the ensemble corresponding to a different choice of the random Fourier phases at small scales), which all coincide with the original QPF when aggregated on scales larger than L_0 and T_0. As shown by Rebora et al. (2006b) (to whom we refer for details), the precipitation fields produced by RainFARM correctly reproduce the small-scale statistics of the precipitation, such as the scaling properties of the main statistical moments, the spatiotemporal correlation structure of the fields, and the spectrum of the generalized fractal dimensions, and capture the temporal persistence of the observed precipitation also at scales smaller than the reliability scales.

3. Area-to-point downscaling

Area-to-point interpolation of forecast fields by classical methods, such as simple nearest neighbors, linear interpolation, or distance-weighted interpolation (Barnes 1964; Cressman 1959), underestimates the variability of the precipitation observed by rain gauges. Conditional stochastic interpolation (such as kriging-based methods) and stochastic downscaling represent alternative interpolation techniques that can provide a

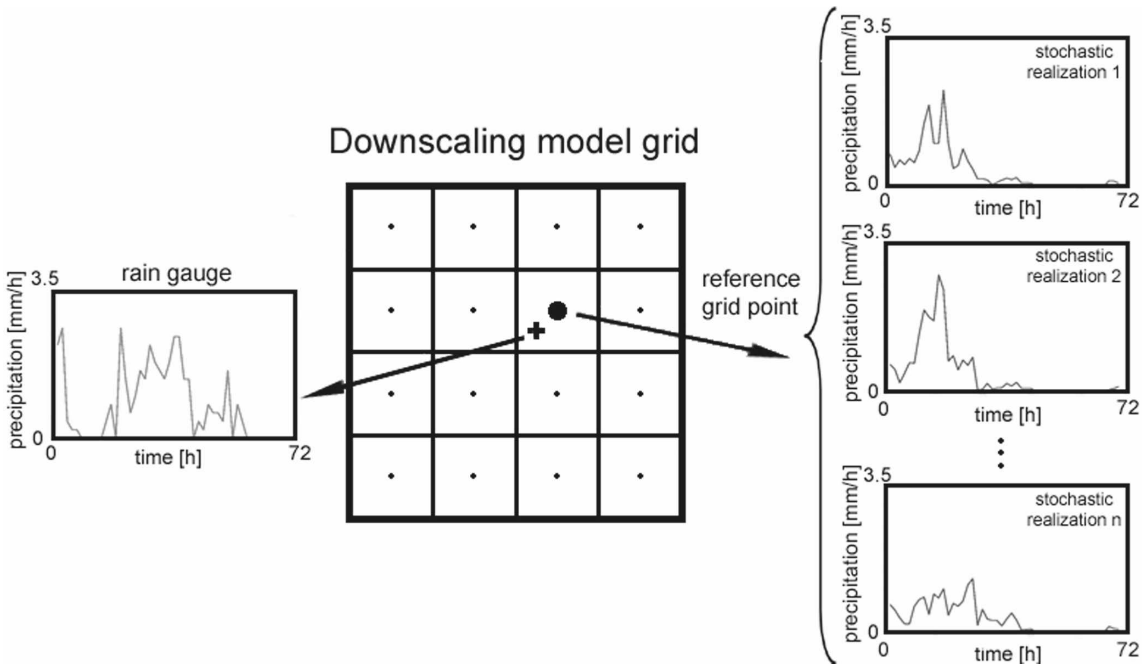

Fig. 2. An illustration of the forecast verification procedure based on stochastic rainfall downscaling.

solution to this problem. Clearly, the fields generated by these methods cannot be used as single deterministic forecasts, but ensembles of stochastic fields can be used to gauge the range of variability expected at each rain gauge position. It is then possible to verify whether the observed precipitation falls within this range.

In the following, we use the RainFARM downscaling technique to generate, for each numerical precipitation forecast, an ensemble of 50 high-resolution stochastic precipitation fields with a temporal resolution of 1 h and a spatial resolution of 875 m (1/8 of the COSMO-LAMI spatial resolution). The time series of precipitation measured at each rain gauge, accumulated over 1-h periods, are then compared with the time series generated by RainFARM at the nearest downscaling grid point. Figure 2 illustrates the procedure.

Two sample precipitation time series are reported in Fig. 3, where they are compared with the ensemble of time series obtained by applying the downscaling procedure described above to the QPF. The gray bands bracket the interval containing 95% of the members of the downscaling ensemble. The example in Fig. 3a shows good agreement between the downscaled forecast and the measurement, as the observed precipitation time series falls within the 95% confidence bands produced by downscaling the QPF. By contrast, the rain gauge measurements in Fig. 3b provide an example of an error of the meteorological model, as the observed precipitation is very different from any time series obtained by downscaling the numerical QPF.

The agreement between the observations and the downscaled QPF time series for all available rain gauges can be quantified with the help of rank histograms, a verification technique commonly used for evaluating ensemble forecasts (Hamill and Colucci 1998). Rank histograms check "where" an observation falls with respect to an ensemble of forecast data. In this approach, each verification data point is compared to the corresponding values of a forecast ensemble and the verification rank (defined as the number of ensemble values exceeding it) is recorded. Repeating this procedure for all verification instances allows the construction of a frequency histogram of the verification ranks.

In an ideal and well-balanced ensemble prediction system, a verification data point is equally likely to lie between any two ordered adjacent ensemble members (or to fall outside the ensemble range on either side of the distribution). For this reason, the rank histogram of a successful ensemble prediction is expected to present a uniform, flat distribution (Hamill 2001). Deviations from a flat distribution are expected in the case of errors from the meteorological model.

We apply the same technique to the situation considered here and use as verification data the accumulated precipitation over 72 h and the maximum hourly precipitation intensity observed at each rain gauge during the event. These two measures are compared with the equivalent quantities obtained for the ensemble of precipitation fields generated by applying RainFARM to

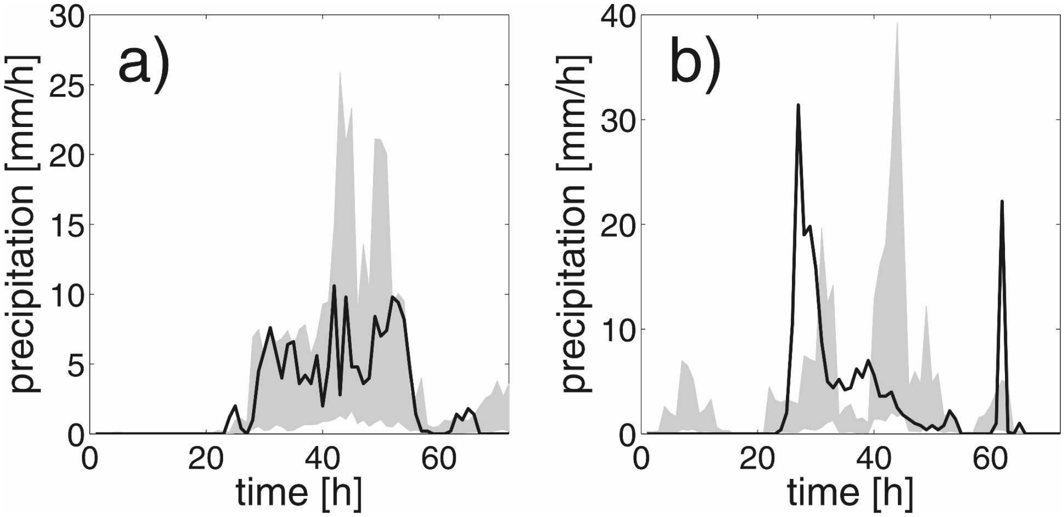

FIG. 3. Examples of single-site rain gauge time series. The gray bands include 95% of the individual members of the ensemble obtained by stochastically downscaling the numerical forecast. (a) Rain gauge at Front Malone, Turin, Italy. (b) Rain gauge at Colle San Bernardo, Cuneo, Italy. Both cases refer to the event of 13–15 Sep 2006. In the downscaling procedure, the forecast was considered reliable down to scales of $L_0 = 28$ km and $T_0 = 3$ h.

the COSMO-LAMI forecasts, and the corresponding rank histograms are computed. Before discussing the results provided by rank histograms, however, we need to address two basic issues. First, we need to define what a "flat" rank histogram is, in a statistical sense. Second, we need to verify that the RainFARM downscaling technique is able to correctly represent the variability of the precipitation at the unresolved scales.

One way to answer the first question and quantitatively determine flatness bounds is based on the use of surrogate data. To this end, we compare two datasets that are statistically identical and thus obtain a bound on the sampling errors. We construct a new ensemble of downscaled stochastic fields (the "surrogate data") and derive from this new set the corresponding precipitation time series at each rain gauge position. Comparing each of these surrogate time series with the original ensemble of downscaled time series and determining their rank histograms, we obtain a measure of the expected variability in each class of the rank histogram due to statistical variability and sampling errors. Fixing a confidence level of 95%, this procedure allows for defining error bars such that if less than 5% of the values of the rank histograms lie outside the bars, the histogram can safely be considered to be flat; that is, the downscaled fields are compatible with the observations.

To answer the second question, that is, whether RainFARM correctly represents the small-scale statistics of precipitation, we apply the procedure first for an ideal, control case. To this end, we construct a perfect forecast by aggregating the rain gauge data of the event of 13–15 September 2006 on a scale of 56 km and 6 h. This "perfect forecast" is then downscaled with RainFARM, generating an ensemble of 50 downscaled, stochastic forecast fields. Assigning to each rain gauge the time series at the closest grid point of these downscaled fields, we obtain a corresponding ensemble of time series at each rain gauge position. The rank histograms comparing the original observations with the stochastic forecasts are reported in Fig. 4. For each rain gauge, the number of forecasts exceeding the observation is counted and the corresponding rank histogram is reported.[2] The 95% confidence bands obtained in the way described above are also reported upon in the figure. The values of the rank histogram for this control case fall outside the 95% confidence bands only for about 5% of the histogram classes, confirming that RainFARM is able to reconstruct the observed variability of the precipitation at small scales and that, as expected, a perfect forecast leads to a flat distribution of the rank frequencies.

Using the same approach, we can now analyze the three events described above. In the following, we use different choices of the reliability scales below which the numerical QPFs are stochastically downscaled. We vary their values from the nominal scales of the forecast

[2] One important issue in this type of comparison is related to the presence of a large number of zeroes, and to the treatment they receive. In the construction of the rank histograms, we did not count rain gauges where both the observation and the forecasts were below the rain gauge resolution (0.2 mm).

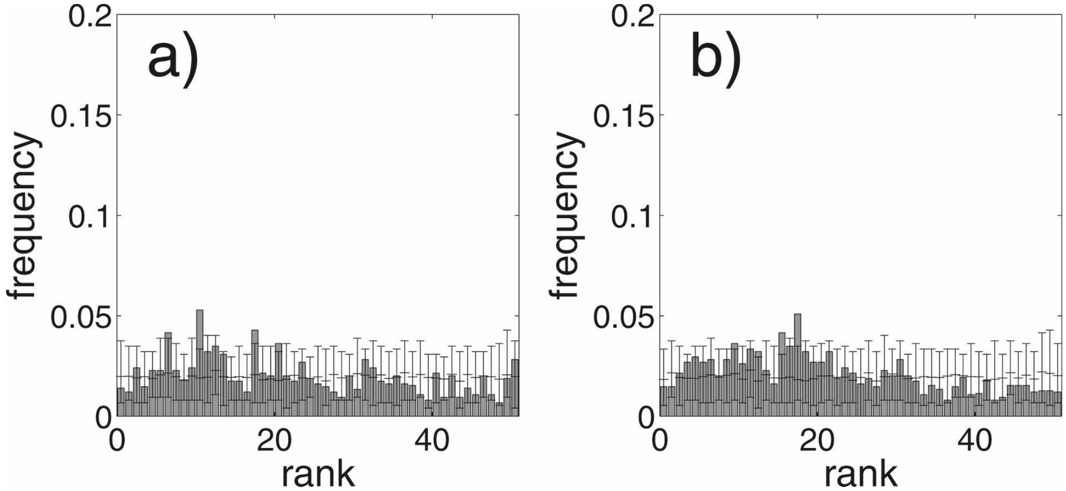

FIG. 4. Rank histograms for the perfect forecast case. (a) Precipitation accumulated over 72 h, and (b) hourly precipitation maxima within the 72 h. In both cases, the reliability scales used in the downscaling procedure are L_0 = 28 km and T_0 = 3 h. Error bars indicate the 95% confidence bands obtained as described in the text.

fields (7 km and 3 h) up to values of 112 km and 24 h. Figure 5 reports, as a function of the reliability scale, the percentage of rank histogram classes that fall within the 95% confidence bands (determined as discussed for the control). The results reported in Fig. 5 indicate that the events on 10–12 April 2005 and on 13–15 September 2006 display significant forecast errors, both for the accumulated and for the maximum precipitation. For

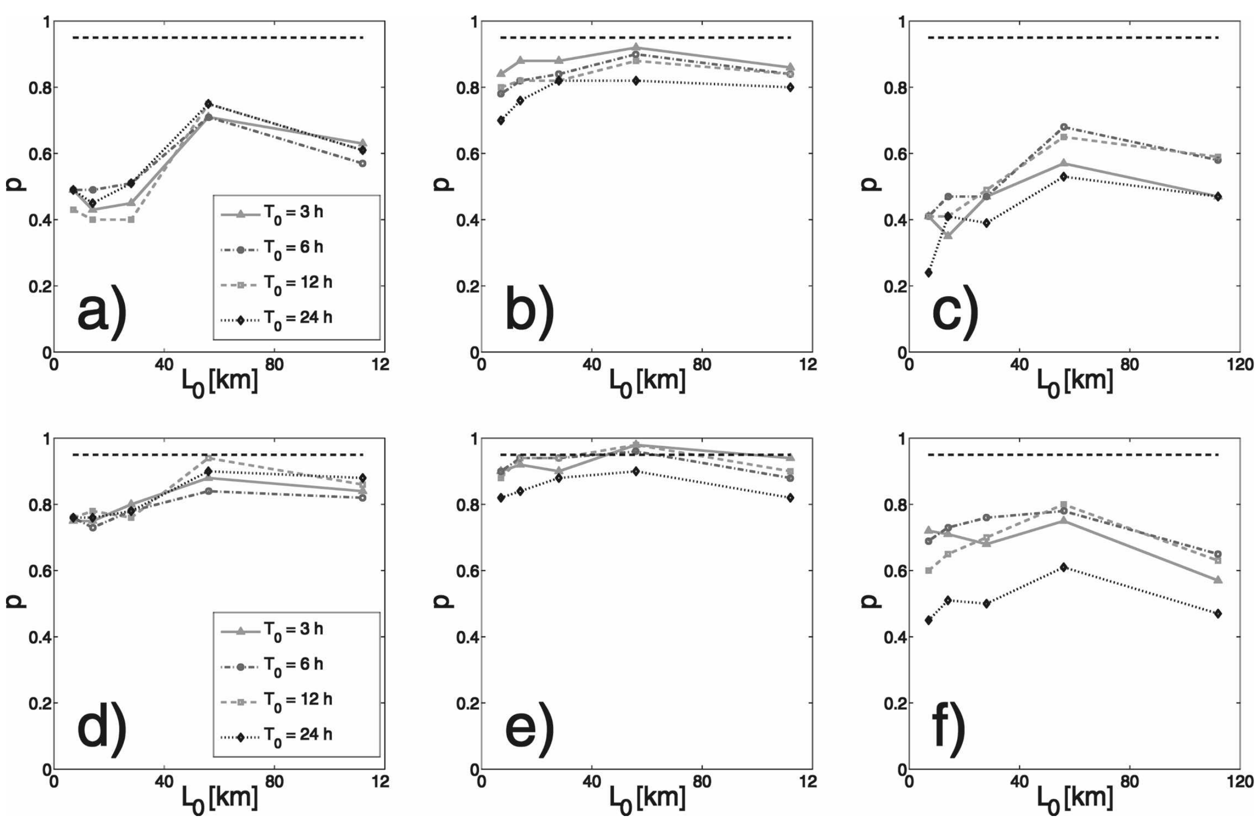

FIG. 5. Percentage of the classes of the rank histogram that fall within the 95% confidence bands, as a function of the values of the spatial and temporal reliability scales. (a), (d) Event of 10–12 Apr 2005; (b), (e) event of 7–9 Sep 2005; and (c), (f) event of 13–15 Sep 2006. (a)–(c) Precipitation accumulated over 72 h; (d)–(f) maximum hourly precipitation intensity within the 72 h.

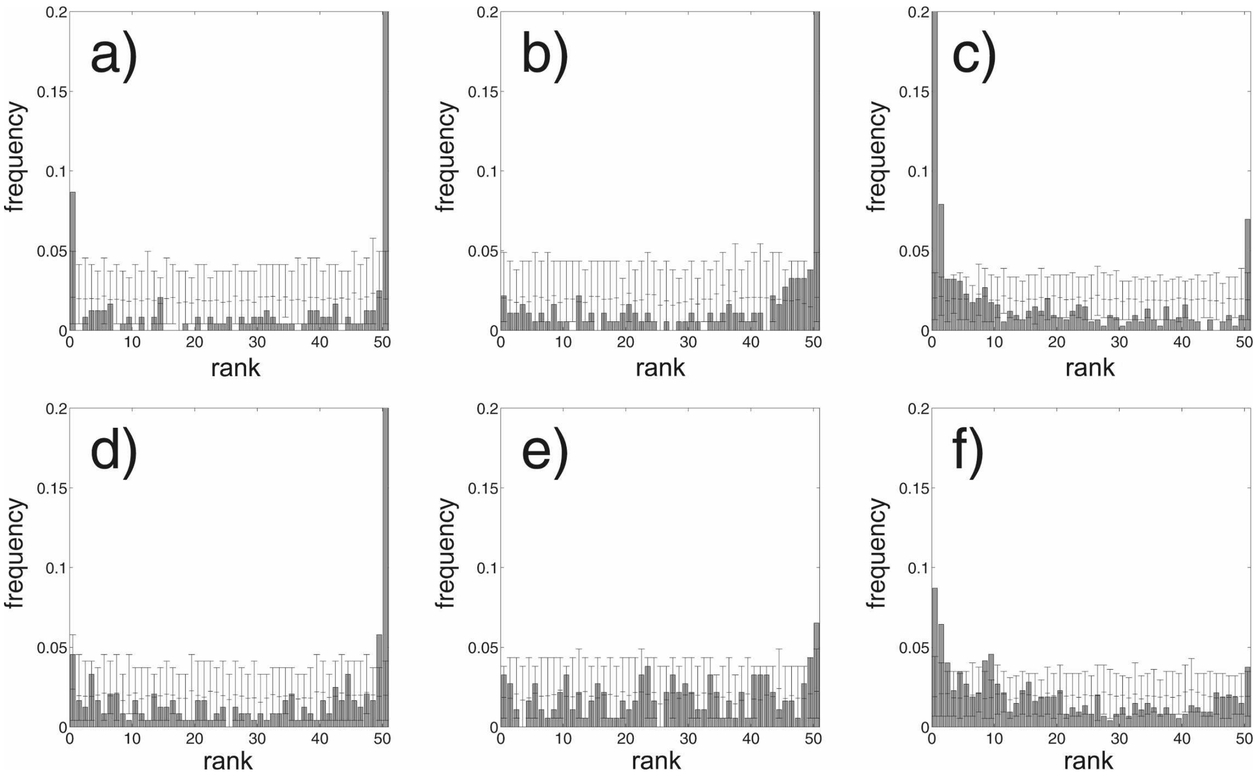

FIG. 6. Rank histograms corresponding to the reliability scales showing the least forecast error in Fig. 5. The ranks are determined by counting the number of stochastic downscaled time series exceeding the observation at each rain gauge. (a), (d) Event of 10–12 Apr 2005; (b), (e) event of 7–9 Sep 2005; and (c), (f) event of 13–15 Sep 2006. (a)–(c) Precipitation accumulated over 72 h; (d)–(f) maximum hourly precipitation intensity within the 72 h. Error bars indicate the 95% confidence bands, obtained as described in the text.

no value of the reliability scales does the percentage of rank histogram classes falling within the confidence bands reach 95%, an indication of significant deviation from the flat distribution expected for no model error. Only the forecast of the maximum hourly precipitation forecast for the event of 7–9 September 2005 shows good skill by reaching a maximum above the 95% threshold, at the values of reliability scales $L_0 = 56$ km and $T_0 = 3$ h. The corresponding plots for the accumulated precipitation in the same event do not reach the 95% threshold but they get very close to it for the same values of the reliability scales.

To get more insight into the nature of the meteorological forecast errors, in Fig. 6 we show the rank histograms corresponding to the reliability scales achieving the best score in Fig. 5. The precipitation maxima forecasted for the event of 7–9 September 2005 display an almost flat rank distribution, with only a small indication of the overestimation due to the high histogram value in the largest rank class. The corresponding accumulated precipitation forecast, on the other hand, clearly reveals an overestimate of the meteorological forecast, as indicated by the exceedingly large histogram value in the largest rank class. The event on 10–12 April 2005 mainly shows an overestimate in the meteorological forecast, while on 13–15 September 2006 the observed precipitation was underestimated for a large number of rain gauges, both in terms of the accumulated and the maximum precipitation. On both dates there is also an indication of an overestimate of the spread of the meteorological forecast compared to the observations, since both ends of the rank histogram present values above the confidence bands.

A point of concern is that, in principle, the resolution of the downscaled fields (875 m) is still much larger than the very small, pointlike, area spanned by a rain gauge. We verified some of the results in this section using a finer, 437.5-m, resolution for the 13–15 September 2006 event. The analysis shown in Fig. 5 also provided the same results at this smaller resolution, indicating that the chosen resolution is fine enough to represent rain gauge measurements.

4. Point-to-area upscaling

Upscaling procedures combine rain gauge measurements to provide an estimate of the average precipita-

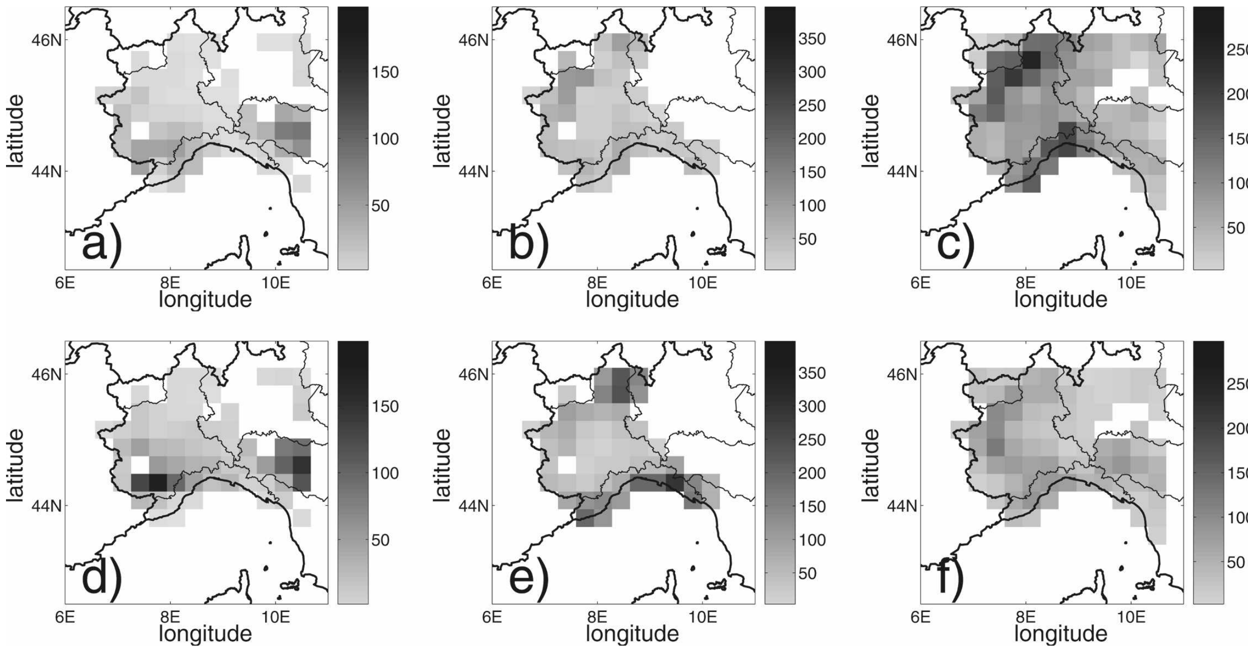

FIG. 7. Maps of precipitation volumes (mm), accumulated over the entire period of 72 h and aggregated on a spatial scale of 28 km, for the three events. (a), (d) Event of 10–12 Apr 2005; (b), (e) event of 7–9 Sep 2005; and (c), (f) event of 13–15 Sep 2006. (a)–(c) Observed precipitation from rain gauges; (d)–(f) COSMO-LAMI forecast.

tion over a large area. Averaging a limited number of rain gauges, however, is inevitably accompanied by potentially severe sampling errors. This problem has been discussed in detail in the past (Rodriguez-Iturbe and Mejia 1974; Zawadski 1973), and different methods, mainly based on estimating the small-scale correlation structure of precipitation, have been developed to identify the contribution of sampling errors to the total error (see, e.g., Anagnostou et al. 1999; Ciach and Krajewski 1999).

An alternative approach, which we follow here, is again based on the use of stochastic rainfall downscaling. We aim to determine the sampling error distribution that would be expected if the forecast were perfect. Since downscaling models are constructed to reproduce the spatiotemporal variability of precipitation at the small unresolved scales, they can be used to estimate such intervals by means of a Monte Carlo approach. After generating a large ensemble of stochastic forecasts by downscaling a QPF field, the corresponding stochastic time series at each rain gauge position can be obtained. The spatial average of these time series, derived in the same way as for the observations, leads to an ensemble of large-scale average precipitation estimates whose distribution allows us to estimate the sampling errors expected for a perfect meteorological forecast.

As an example, we consider the three events already discussed above and use the COSMO-LAMI rain forecasts aggregated on a scale of $L = 28$ km in space and $T = 3$ h in time. The aggregated forecast is compared with the observed precipitation, obtained by averaging the measurements of all rain gauges located in each grid box of size $L \times L$. In the following, as a simple example, we use plain arithmetic averages to aggregate rain gauge measurements, but for operational verification purposes, methods with lower RMSE and which better account for the spatial dependence of the observations, such as block kriging, could be used. The resolution chosen in this example allows for an adequate number of rain gauges in each box (on average, we have five rain gauges in each box), while keeping a sufficiently large number of grid boxes over the study area. The maxima and accumulated precipitation over 72 h are computed for both the aggregated forecast and the averaged observations. The resulting maps of the averaged observations and forecasts are shown in Figs. 7 and 8, for the accumulated precipitation and the precipitation maxima, respectively. Visual inspection of these maps suggests that in all three cases the meteorological forecast was affected by strong errors at the scale considered here. Only the precipitation maxima for the event on 7–9 September 2005 show a reasonable correspondence between the meteorological forecast and the observations, as was already indicated by the behavior of the rank histograms discussed in the previous section.

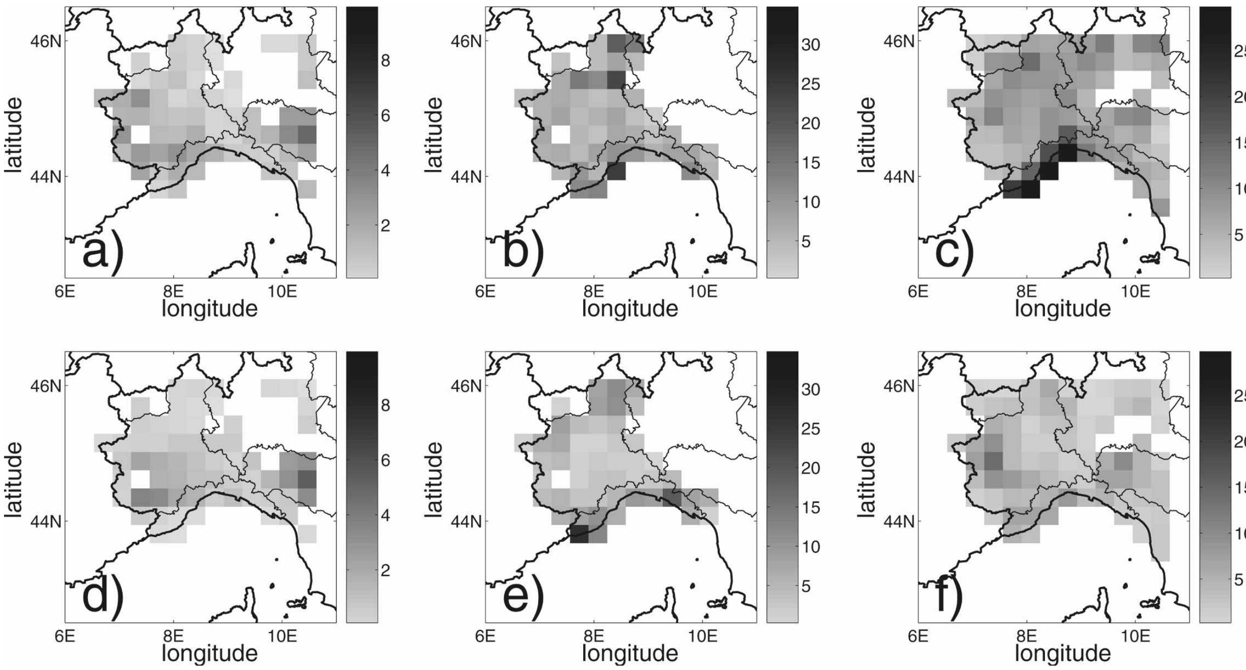

FIG. 8. Maps of the maximum hourly precipitation intensity (mm h^{-1}) in the entire period of 72 h, aggregated on a spatial scale of 28 km, for the three events. (a), (d) Event of 10–12 Apr 2005; (b), (e) event of 7–9 Sep 2005; and (c), (f) event of 13–15 Sep 2006. (a)–(c) Observed precipitation from rain gauges; (d)–(f) COSMO-LAMI forecast.

To verify the meteorological forecasts, in Fig. 9 we show the scatterplots, over the whole set of grid boxes containing at least four rain gauges, of the forecasts versus the observations. A perfect forecast would produce scatterplots where the points are aligned, with small spread, along the 45° line. In general, the numerical meteorological forecasts display a large scatter with respect to the upscaled observations and deviate from the 45° line, suggesting the presence of potentially severe forecast errors and bias.

To provide a quantitative basis to the above statement, we must assess whether the observed deviations from the 45° line are significant. To this end, we consider the ensemble of the time series already discussed in the previous section, obtained by stochastically downscaling the QPF and taking, for each rain gauge, the closest grid point in the downscaled field. In this way, for each member of the ensemble we get a set of numerical time series that have the same spatial distribution as the set of rain gauges. However, each numerical time series is obtained exactly from the numerical forecast that we want to verify. By aggregating these time series on the same spatiotemporal scales as was done for the true rain gauge data, we get a precipitation field that differs from the original numerical QPF only because of sampling errors and statistical variability (i.e., its expected value is equal to the original QPF, and if we had a very large number of time series, then we should exactly recover the original QPF). For each member of the ensemble, we can then obtain a scatterplot of the original field versus the sampled one; the bands that contain, say, 95% of the estimates provide a measure of the confidence limits on the sampling errors, and are shown in gray in Fig. 9. The behavior of the scatterplots confirms the results of the downscaling analysis reported in the previous section: In the event of 7–9 September 2005, the QPF displays good forecast skill for hourly precipitation maxima, as shown by the large number of rain gauges falling within the gray bands, but it provides an overestimate of the accumulated precipitation. For the events of 10–12 April 2005 and 13–15 September 2006, the forecasts of the precipitation maxima are less skillful than for the September 2005 event; in particular, the April 2005 accumulated precipitation is significantly overestimated while the September 2006 accumulated precipitation is significantly underestimated.

5. Summary and conclusions

In this work we have discussed how stochastic downscaling can be effectively used as a tool to verify quantitative precipitation forecasts obtained from numerical models against point observations provided by rain

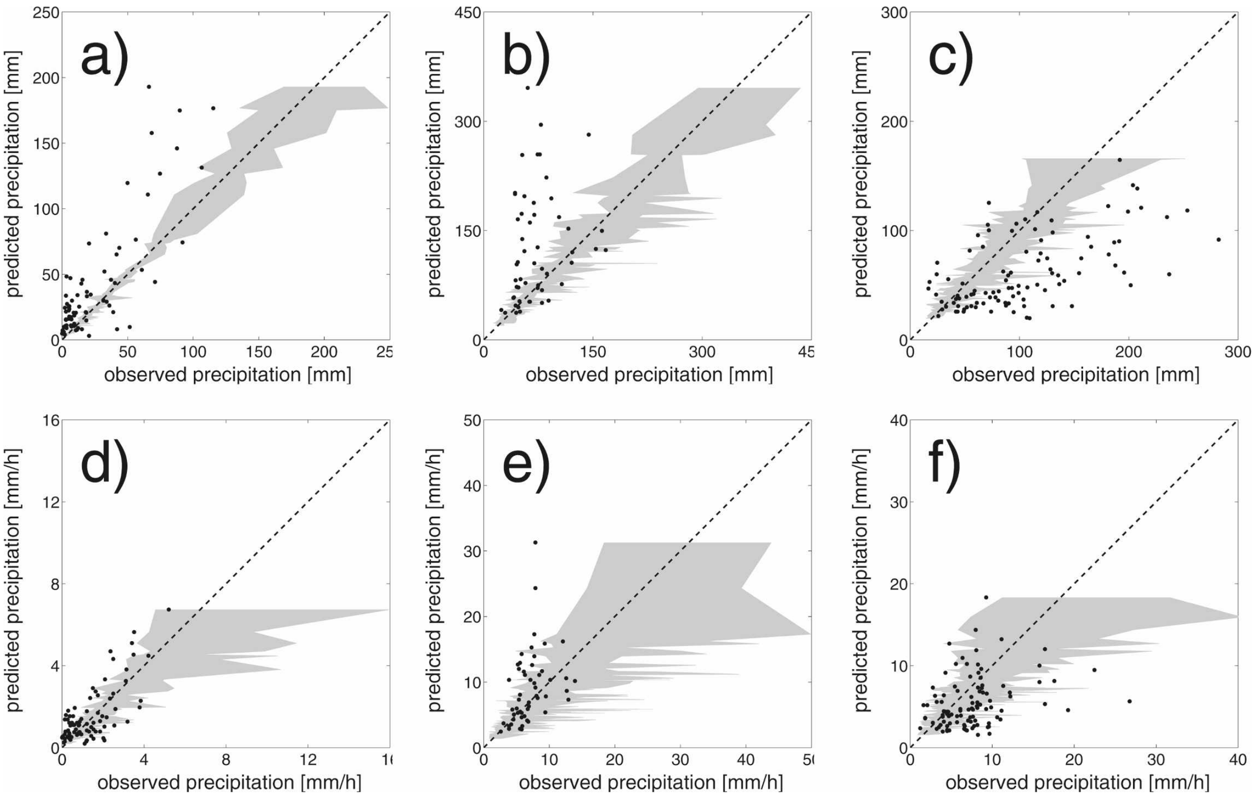

FIG. 9. Scatterplots of observed vs forecasted precipitation, aggregated on the scale $L = 28$ km and $T = 3$ h. (a), (d) Event of 10–12 Apr 2005; (b), (e) event of 7–9 Sep 2005; and (c), (f) event of 13–15 Sep 2006. (a)–(c) Precipitation accumulated over 72 h; (d)–(f) maximum hourly precipitation intensity within the 72 h. The gray bands represent the 95% confidence limits for sampling errors, obtained as described in the text.

gauges. Downscaling techniques can be used as an empirical way to generate the small-scale variability and correlation structure of precipitation, starting from larger-scale QPF fields, and to characterize representativeness errors using ensembles of downscaled fields. When the verification is performed by area-to-point interpolation, the downscaling approach allows the characterization of the expected variability at rain gauge positions and to check whether the QPF is compatible with the observed precipitation data. Investigating different aggregation scales also makes it possible to determine the scales above which a meteorological model can be considered reliable. When the opposite (point to area) averaging approach is used for verification, Monte Carlo downscaling ensembles provide a measure of the expected range of the variability of areal averages based on a limited number of rain gauges.

A crucial point for the feasibility of the approach described in this paper is the availability of downscaling techniques that are able to capture the variability and the correlation structure of the observed precipitation at small spatial and temporal scales. The RainFARM technique used here meets this condition, and so do downscaling methods such as that described by Venugopal et al. (1999).

On a side note, the present analysis has shown that, at least in the three events considered here, the numerical precipitation forecasts were affected by a significant error. Only in one case (the event of 7–9 September 2005) was the QPF able to correctly predict the observed hourly precipitation maxima (see Fig. 5). Even in this case, however, the total precipitation volumes were still affected by significant error. The QPFs for the other two events were affected by significant over- or underestimates, suggesting that, at least at the scales considered here (of the order of 30–50 km in space and 3 h in time), the meteorological forecast cannot be considered quantitatively reliable. We have repeated the analyses considering only the first, second, or third intervals of 24 h in the 72-h-long forecast, and found analogous results, with an expected degradation of the forecast skill in the last 24-h segment.

Clearly, this conclusion cannot be taken as a general statement and a more complete verification study over a larger database of events and with other numerical precipitation forecast models is required, possibly also exploring verification metrics that differ from those considered here. In our opinion, a careful verification of numerical QPFs is an essential, albeit sometimes discarded, component in the construction of effective operational chains for hydrometeorological risk assessment. The preliminary results reported here raise some concerns, and a large-scale verification study is called for. The downscaling approach discussed in this work provides a possible methodology for this type of verification studies.

Acknowledgments. We thank Franco Siccardi, Carlo Cacciamani, Stefano Tibaldi, Marina Morando, and Luca Molini for insightful comments and useful discussions on these results. We are grateful to Efi Foufoula-Georgiou and two anonymous reviewers, whose comments greatly helped us to improve the manuscript. This work was supported by the project "Proscenio," funded by the Italian national Civil Protection Department.

REFERENCES

Anagnostou, E. N. B., W. F. Krajewski, and J. Smith, 1999: Uncertainty quantification of mean-areal radar-rainfall estimates. *J. Atmos. Oceanic Technol.,* **16,** 206–215.

Bacchi, B., R. Ranzi, and E. Richard, 2003: Hydrometeorological processes and floods in the Alps: An overview. *Hydrol. Earth Syst. Sci.,* **7,** 785–798.

Barnes, S. L., 1964: A technique for maximizing details in numerical weather map analysis. *J. Appl. Meteor.,* **3,** 396–409.

Bell, T. L., 1987: A space–time stochastic model of rainfall for satellite remote sensing studies. *J. Geophys. Res.,* **92** (D8), 9631–9643.

Bras, R. L., and I. Rodriguez-Iturbe, 1976: Evaluation of mean square error involved in approximating the areal average of a rainfall event by discrete summation. *Water Resour. Res.,* **12,** 1185–1195.

Ciach, G. J., and W. F. Krajewski, 1999: On the estimation of radar rainfall error variance. *Adv. Water Resour.,* **22,** 585–595.

Colle, B. A., K. J. Westrick, and C. F. Mass, 1999: Evaluation of MM5 and Eta-10 precipitation forecasts over the Pacific Northwest during the cool season. *Wea. Forecasting,* **14,** 137–154.

Cressman, G. P., 1959: An operational objective analysis system. *Mon. Wea. Rev.,* **87,** 367–374.

Doms, G., and U. Schattler, 1998: The non-hydrostatic limited area model LM (Lokal Modell) of DWD. Part I: Scientific Documentation. German Weather Service, Rep. LM F90 1.35, 172 pp.

Droegemeier, K. K., and Coauthors, 2000: Hydrological aspects of weather prediction and flood warnings: Report of the Ninth Prospectus Development Team of the U.S. Weather Research Program. *Bull. Amer. Meteor. Soc.,* **81,** 2665–2680.

Ferraris, L., R. Rudari, and F. Siccardi, 2002: The uncertainty in the prediction of flash floods in the northern mediterranean environment. *J. Hydrometeor.,* **3,** 714–727.

——, S. Gabellani, N. Rebora, and A. Provenzale, 2003: A comparison of stochastic models for spatial rainfall downscaling. *Water Resour. Res.,* **39,** 1368–1384.

Hamill, T. M., 2001: Interpretation of rank histograms for verifying ensemble forecasts. *Mon. Wea. Rev.,* **129,** 550–560.

——, and S. J. Colucci, 1998: Evaluation of Eta–RSM short-range ensemble probabilistic precipitation forecasts. *Mon. Wea. Rev.,* **126,** 711–724.

Harris, D., E. Foufoula-Georgiou, K. K. Droegemeier, and J. J. Levit, 2001: Multiscale statistical properties of a high resolution precipitation forecast. *J. Hydrometeor.,* **2,** 406–418.

Hendrick, R. L., and G. H. Comer, 1970: Space variations of precipitation and its implications for rain gauge network design. *J. Hydrol.,* **10,** 151–163.

Lovejoy, S., and B. Mandelbrot, 1985: Fractal properties of rain and a fractal model. *Tellus,* **37A,** 209–232.

Marsigli, C., A. Montani, F. Nerozzi, T. Paccagnella, S. Tibaldi, F. Monteni, and R. Buizza, 2001: A strategy for high-resolution ensemble prediction. Part II: Limited-area experiments in four Alpine flood events. *Quart. J. Roy. Meteor. Soc.,* **127,** 2095–2115.

Menabde, M., A. Seed, D. Harris, and G. Austin, 1997: Self-similar random fields and rainfall simulations. *J. Geophys. Res.,* **102D,** 13 509–13 515.

Montani, A., C. Marsigli, F. Nerozzi, T. Paccagnella, S. Tibaldi, and R. Buizza, 2003: The Soverato flood in southern Italy: Performance of global and limited-area ensemble forecasts. *Nonlinear Processes Geophys.,* **10,** 261–274.

Patterson, G., and S. Orszag, 1971: Spectral calculations of isotropic turbulence: Efficient removal of aliasing interaction. *Phys. Fluids,* **14,** 2538–2541.

Perica, S., and E. Foufoula-Georgiou, 1996: Model for multiscale disaggregation of spatial rainfall based on coupling meteorological and scaling description. *J. Geophys. Res.,* **101** (D21), 26 347–26 361.

Rebora, N., L. Ferraris, J. von Hardenberg, and A. Provenzale, 2006a: Rainfall downscaling and flood forecasting: A case study in the Mediterranean area. *Nat. Hazards Earth Syst.,* **6,** 611–619.

——, ——, ——, and ——, 2006b: The RainFARM: Rainfall downscaling by a filtered autoregressive model. *J. Hydrometeor.,* **7,** 724–738.

Rodriguez-Iturbe, I., and J. M. Mejia, 1974: On the transformation of point rainfall to areal rainfall. *Water Resour. Res.,* **10,** 729–736.

Siccardi, F., G. Boni, L. Ferraris, and R. Rudari, 2005: A hydrometeorological approach for probabilistic flood forecast. *J. Geophys. Res.,* **110,** D05101, doi:10.1029/2004JD005314.

Tustison, B., D. Harris, and E. Foufoula-Georgiou, 2001: Scale issues in verification of precipitation forecasts. *J. Geophys. Res.,* **106,** 11 775–11 784.

——, E. Foufoula-Georgiou, and D. Harris, 2003: Scale-recursive estimation for multisensor quantitative precipitation forecast verification: A preliminary assessment. *J. Geophys. Res.,* **108,** 8377, doi:10.1029/2001JD001073.

Venugopal, V., E. Foufoula-Georgiou, and V. Sapozhnikov, 1999: Evidence of dynamic scaling in space–time rainfall. *J. Geophys. Res.,* **104** (D24), 31 599–31 610.

Waymire, E. C., V. K. Gupta, and I. Rodriguez-Iturbe, 1984: A spectral theory of rainfall intensity at the meso-beta scale. *Water Resour. Res.,* **20,** 1453–1465.

Zawadski, I., 1973: Errors and fluctuations of raingauge estimates of areal rainfall. *J. Hydrol.,* **18,** 243–255.