

Stochastic versus Dynamical Downscaling of Ensemble Precipitation Forecasts

ELISA BRUSSOLO

CIMA Research Foundation, Savona Italy

JOST VON HARDENBERG

ISAC-CNR, Torino, Italy

NICOLA REBORA

CIMA Research Foundation, Savona, Italy

(Manuscript received 30 September 2008, in final form 30 March 2009)

ABSTRACT

The assessment of hydrometeorological risk in small basins requires the availability of skillful, high-resolution quantitative precipitation forecasts to predict the probability of occurrence of severe, localized precipitation events. Large-scale ensemble prediction systems (EPS) currently provide forecast scenarios down to a resolution of about 50 km. High-resolution, nonhydrostatic, limited-area ensemble prediction systems provide dynamically based forecasts by extending these scenarios to smaller scales, typically on the order of 10 km. This work explores an alternative approach to the use of limited-area ensemble prediction systems, by directly applying a stochastic downscaling technique to large-scale ensemble forecasts. The performances of these two different approaches for three well-predicted precipitation events in northwestern Italy during 2006 are compared. Ensemble forecasts provided by the ECMWF EPS, downscaled using the Rainfall Filtered Autoregressive Model (RainFARM) stochastic technique, and ensemble forecasts obtained from the Consortium for Small-Scale Modeling Limited-Area Ensemble Prediction System (COSMO-LEPS) are considered. A dense network of rain gauges is used for verification. It is found that the probabilistic forecast skill of stochastically downscaled ensembles may be comparable with that of dynamically downscaled ensembles, using a range of standard forecast skill measures. Stochastic downscaling is suggested as a tool for benchmarking the performance of dynamical ensemble downscaling systems.

1. Introduction

Precipitation intensity represents the crucial atmospheric variable for the operational assessment of hydrometeorological risks; unfortunately, it is also one of the most difficult to forecast reliably by current weather models. Uncertainty in quantitative precipitation forecasts (QPFs) arises as a result of measurement errors, incomplete observations, insufficiently resolved initial conditions, an incomplete representation of the physics of the problem, finite numerical resolution, and sensitivity to initial conditions of atmospheric models. At large scales the different sources of uncertainty in me-

teorological forecasts can be captured by dynamical ensemble forecasting techniques (Epstein 1969; Leith 1974; Toth and Kalnay 1993; Palmer 1993; Molteni et al. 1996; Toth and Kalnay 1997), in which the probability of occurrence of different meteorological scenarios is estimated from the relative frequency of different forecast ensemble members. Existing operational global ensemble forecast systems, such as the Ensemble Prediction System (EPS) of the European Centre for Medium-Range Weather Forecasts (ECMWF) and the global ensemble system of the National Centers for Environmental Prediction (NCEP), reach nominal spatial scales of down to 50 km and typically consist of tens of ensemble members. The operational ECMWF ensemble (Molteni et al. 1996), with its 51 members, is one of the largest and better-resolved ensembles.

Civil protection applications, such as flood warnings in small Alpine basins, require forecasts and an estimate

Corresponding author address: Jost von Hardenberg, Institute of Atmospheric Sciences and Climate (ISAC-CNR), Corso Fiume 4, 10133, Torino, Italy.
E-mail: j.vonhardenberg@isac.cnr.it

of uncertainty at spatial scales of the order of tens of kilometers or less, significantly smaller than those currently provided by these systems. To address this issue, different operational systems have been implemented, in which large-scale ensemble forecasts are extended to smaller scales by nesting high-resolution, nonhydrostatic, limited-area atmospheric models in the general circulation scenarios. Examples of such local-area ensemble forecast systems are the Consortium for Small-scale Modeling's Limited-Area Ensemble Prediction System (COSMO-LEPS; Molteni et al. 2001; Marsigli et al. 2001; Marsigli et al. 2005), the High-Resolution Limited-Area Model/Limited-Area Model Ensemble Prediction System (HIRLAM/LAMEPS; Frogner and Iversen 2002; Frogner et al. 2006; Sattler and Feddersen 2005), and the NCEP's Short-Range Ensemble Forecasting (SREF) system (Hamill and Colucci 1997, 1998; Du et al. 1997, 2006). This dynamical downscaling approach allows modelers to take into account small-scale phenomena that are relevant for the formation of intense localized precipitation, such as orographic and nonhydrostatic effects. Additional small-scale uncertainty is captured by varying model physics or initial conditions. Because of computational limits, local area ensemble forecasting systems typically comprise a small number of ensemble members, on the order of 10–20 members. The number of members used has been found to influence the performance statistics in these systems (Marsigli et al. 2005). Because of the incomplete representation of small-scale processes, the limited resolution of the observational network used for initialization and data assimilation, and the numerical details, these models often present limited forecast skill (de Elià et al. 2002) and underestimate variability (Harris et al. 2001) at the smallest scales.

In addition to these dynamical, physically based, downscaling approaches, several stochastic downscaling methods have been developed in the hydrometeorological literature. These techniques exploit the observed spatiotemporal structure of precipitation fields to construct ensembles of stochastic precipitation fields that conserve the large-scale features of the flow and present realistic statistical properties at small scales [see Ferraris et al. (2003) for a comparison of different approaches]. They permit the rapid creation of large ensembles of fields, also at very high spatial and temporal resolutions, allowing for a smooth evaluation of the probabilities of occurrence of intense precipitation events over small basins. The large number of ensemble members made possible by these methods allows for limiting residual errors in the forecast due to undersampling of the forecast probability distribution function (Buizza and Palmer 1998). Their computational burden is consider-

ably less than physically based downscaling. Although these techniques have been originally developed to downscale single large-scale deterministic forecasts, they represent a possible alternative to limited-area models to extend large-scale ensemble forecasts to small scales.

This work compares the performance of these two very different ensemble forecast downscaling approaches. To this purpose we consider ensemble forecasts provided by the ECMWF EPS, downscaled in space using the Rainfall Filtered Autoregressive Model (RainFARM) stochastic technique (Rebora et al. 2006a), and ensembles of forecasts obtained from the COSMO-LEPS limited-area prediction system (which also uses ECMWF EPS ensemble members as boundary conditions), for three intense precipitation events over northern Italy in 2006. The statistical properties of the fields produced with these two techniques are compared, and the skill of the resulting ensembles is verified against direct precipitation measurements from a dense network of rain gauges.

2. Ensemble forecasts and verification data

a. Ensemble forecasts

Since 1992, an EPS has been running operationally at the ECMWF. Different possible future circulation scenarios are generated through the perturbation of initial conditions obtained from the assimilation of a large dataset of observations from the entire globe (Buizza et al. 1999; Molteni et al. 1996; Toth et al. 2001). In its configuration at the time of the events considered in this study (2006), the EPS system evolved 50 initial perturbations—generated using the singular-vector technique—plus an unperturbed control run (Buizza and Palmer 1995). The model grid has 62 vertical levels and a horizontal T_L399 resolution, corresponding to a grid spacing of around 50 km. In this work we use EPS precipitation forecasts accumulated over six hours, available on a regular grid with 0.5° spacing.

COSMO-LEPS (Molteni et al. 2001; Marsigli et al. 2001, 2005) is operationally run by the Italian Regional Hydro-Meteorological Service of Emilia-Romagna [Agenzia Regionale Prevenzione e Ambiente Dell'Emilia-Romagna-Servizio IdroMeteoClima (ARPA-SIMC)]. A clustering technique is used to select a few representative members from the EPS ensemble, which provide boundary and initial conditions for 16 ensemble members (in the year 2006), advanced in time using the nonhydrostatic COSMO Lokal Modell (Doms and Schättler 1998), on a spatial grid with a resolution of about 10 km (obtained from a rotated regular latitude–longitude grid). The system provides operational forecasts

of precipitation accumulated over six hours, on the same grid.

b. Stochastic downscaling

The ECMWF EPS fields are downscaled using the stochastic model RainFARM (Rebora et al. 2006a,b). This downscaling model is based on a nonlinear transformation of a linearly correlated stochastic field, obtained by extrapolation to small scales of the large-scale power spectrum of the forecast. A power-law functional form is assumed for the extrapolation, and random Fourier phases are used at the unresolved scales. The downscaling procedure maintains exactly the large-scale features of the original fields down to given reliability scales L_o and T_o and generates stochastic variability only at smaller scales. The reliability scales 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. As shown by Rebora et al. (2006a), the precipitation fields produced by RainFARM correctly reproduce the small-scale statistics of precipitation—such as the scaling properties of the main statistical moments, the spatiotemporal correlation structure of the fields, and the spectrum of generalized fractal dimensions.

Using this procedure, for each member of the EPS ensemble we generate one stochastic realization with a temporal resolution of six hours and a spatial resolution of 0.1° , close to the 10-km resolution of the LEPS forecasts. This leads to an ensemble of 51 spatiotemporal fields downscaled stochastically in space. Please note that because the EPS and LEPS models work on very different grids, the grid of the downscaled EPS fields does not coincide with the LEPS grid. Nonetheless, the areas of grid elements are similar (about 90 and 100 km² for the downscaled EPS and LEPS grids, respectively); for this reason, we prefer to use them directly rather than interpolating data from one grid to the other. Numerical weather prediction models should not be considered reliable at their nominal resolution (Patterson and Orszag 1971; Harris et al. 2001), so we consider the EPS ensemble forecast fields reliable at the larger scales of $L_o = 1^\circ$ in space and $T_o = 6$ h in time. The downscaling area is the dashed boundary in Fig. 1.

c. Study cases

Three intense precipitation events in 2006 over northwestern Italy are used as study cases: 15 August, 13 September, and 5 December. These events were selected among the five events in 2006 that led to serious civil protection alarms over northwestern Italy, maximizing event intensity and a good agreement between

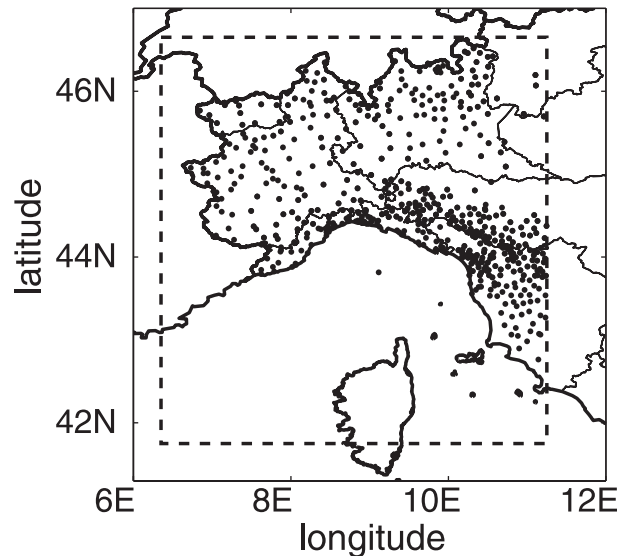


FIG. 1. The study area and the network of rain gauges used for the verification of QPFs. Each point indicates the location of a rain gauge. The dashed line indicates the area over which EPS forecasts were downscaled using the RainFARM technique.

ensemble forecasts and rain gauge observations. These events were characterized by the following synoptic scenarios:

From 15 to 18 August: A disturbance affected northwestern Italian regions for two days, leading to high instability, severe thunderstorms, and heavy and widespread precipitation.

From 13 to 16 September: A deep low-pressure structure extended from the British Islands to the Sahara, leading to an intense warm moist flow over all of Italy and causing unstable conditions with heavy precipitation, particularly over northern Italy.

From 5 to 8 December: A low-pressure structure centered over Sardinia led to convective instability with intense bursts of precipitation over northwestern Italy.

Because our goal is to compare the ability of both the stochastic downscaling and the LEPS system in extending the large-scale predictions to small scales, we limit our analysis only to precipitation events that were reasonably well predicted by the EPS system: Large-scale errors in the EPS forecasts would simply be propagated to smaller scales both by a stochastic downscaling method and by the LEPS system, making a meaningful comparison very difficult. For the same reason, because the skill of quantitative precipitation forecasts drops rapidly with lead time (Sanders 1986), we limit our study to the first 72 h of the LEPS and EPS forecasts, starting at 1200 UTC.

d. Verification dataset

The observational dataset consists of measurements from a dense regional network of rain gauges, provided by the Italian Civil Protection Department, over a study area of $5^\circ \times 5^\circ$ covering northwestern Italy, described in Brussolo et al. (2008) and shown in Fig. 1. A large number of rain gauges (about 600) are available for the periods of interest. For each case study, we consider only rain gauges for which a complete, uninterrupted 72-h time series is available and we accumulate rainfall over 6-h periods. The rain gauge observations are area averaged on the $0.1^\circ \times 0.1^\circ$ grid of the downscaled EPS forecasts and on the $10 \text{ km} \times 10 \text{ km}$ grid of the LEPS forecast fields, using the Thiessen technique. Throughout this work, all of the statistics on the EPS and LEPS forecasts (apart from power spectra) are computed only over grid boxes containing at least one rain gauge.¹

3. Statistical agreement

As a first step in comparing dynamical LEPS ensemble forecasts with stochastically downscaled EPS fields, we simply compare their main statistical properties and verify their agreement with those of observed precipitation over the study area. We will proceed with a point-by-point verification of the forecasts provided by the two downscaling approaches against observations in a following section. The variable considered in the analyses is always 6-h accumulated precipitation.

a. Area-averaged precipitation

Because downscaling cannot correct large-scale forecast errors, as discussed earlier, we chose three events in 2006 that were well predicted at large scales, at least in terms of precipitation volumes. Figure 2 reports the evolution in time of 6-h accumulated precipitation averaged over the study area, comparing the EPS and LEPS ensemble members with observed precipitation. These plots confirm that in all three events considered, both the EPS and the LEPS ensembles captured correctly the time evolution of total volume of 6-h accumulated precipitation, being able to forecast the time of peak activity and bracketing the observation at almost all times.

¹ It is well known that upscaling small numbers of rain gauges by area averaging may lead to limited representativeness and large uncertainties (Bras and Rodriguez-Iturbe 1976; Ciach and Krajewski 1999); on the other hand, because of the density of the available observational network, in our data few small-scale grid boxes contain more than 2–3 rain gauges. Possible techniques to assess this representativeness error have been discussed in Tustison et al. (2001) and Brussolo et al. (2008) but are not considered here for simplicity.

b. Spatial correlation structure

The stochastic downscaling method RainFARM has been shown (Rebora et al. 2006a) to be able to reproduce correctly the small-scale correlation structure of observed precipitation, by extending the large-scale power spectrum to small scales. We verify to what extent this procedure leads to a spatial correlation structure comparable to that of LEPS forecasts, by comparing in Fig. 3 the spatial power spectra of the LEPS forecasts and of the downscaled EPS ensemble. At large scales (below wavenumber 10, corresponding to a spatial scale of 50 km), LEPS ensemble members present the same (on 13 September 2006) or higher variance than the EPS ensemble members. This can be attributed to the possibility of the nonhydrostatic limited-area model to generate additional precipitation compared with the parameterizations used in the general circulation model used in the EPS system (Tibaldi et al. 2006). We also see that while the downscaled EPS spectra follow a power law at small scales by construction, the LEPS spectra drop more rapidly, until at small scales, above wavenumber $k = 10$, they present slightly less variability than the downscaled EPS ensemble. It is well known that numerical weather prediction models present a pronounced drop in variability at their smallest resolved scales (Harris et al. 2001). The RainFARM stochastic downscaling technique avoids this problem by simply extrapolating the large-scale spectrum to these scales.

c. Amplitude distributions

We proceed with comparing the amplitude distributions of the forecast fields at different spatial scales by reporting in Fig. 4 the first four statistical moments for the downscaled EPS and the LEPS ensembles. Specifically, the 6-h accumulated precipitation fields are spatially averaged over boxes of different sizes. From the resulting spatiotemporal fields, we compute the global mean, variance, skewness, and kurtosis for each ensemble member. The same statistics are computed also for the observations and reported in the figure.² The figure shows that all moments are very well captured by the LEPS ensemble at all aggregation scales. In fact, each statistic is always contained between the ensemble extremes at most aggregation scales and, apart from a few exceptions for the highest moments, typically falls well within 88% of the ensemble members. The downscaled EPS members underestimate mean and variance

² Because the EPS and the LEPS grids are different, the observation statistics computed on these grids are slightly different.

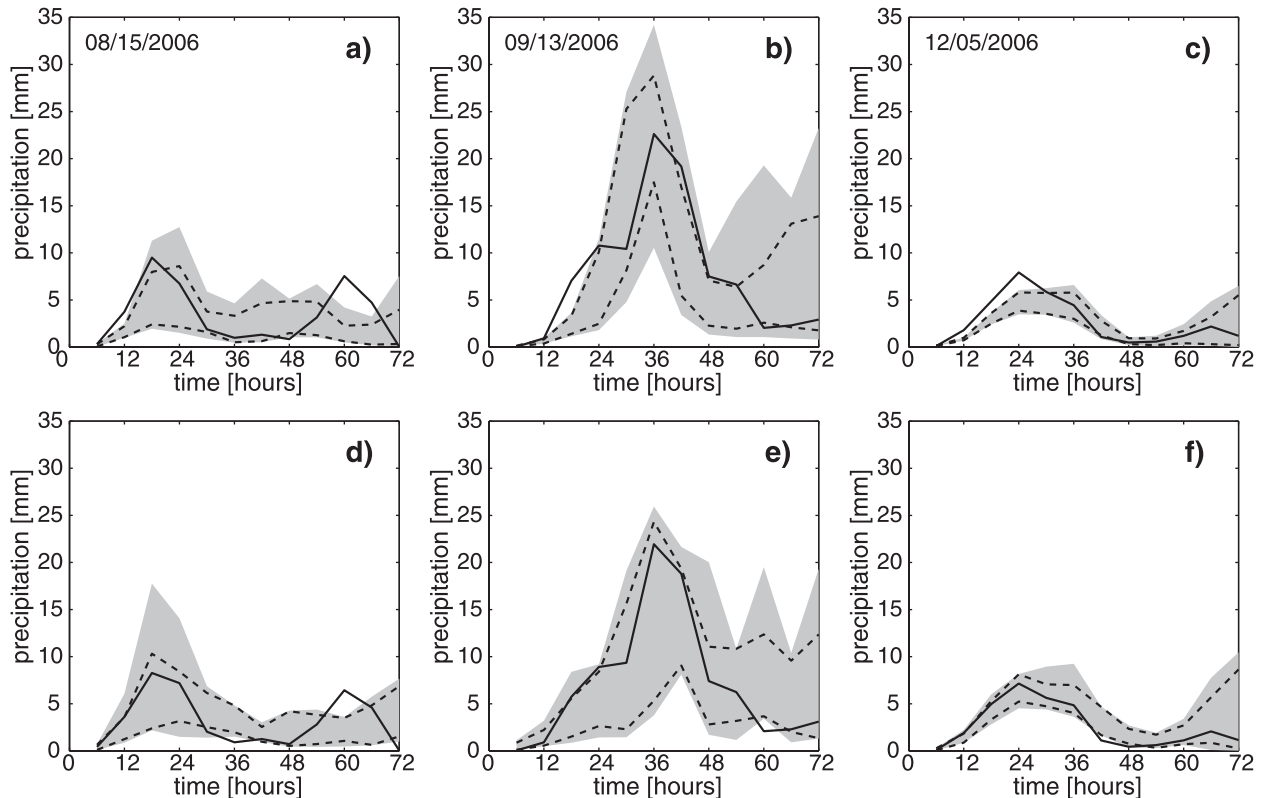


FIG. 2. Plots, as a function of time, of 6-h accumulated precipitation averaged over the study area. To consider homogeneous areas and only the large-scale features of the forecasts, EPS and LEPS fields were averaged over grid boxes of $0.5^{\circ} \times 0.5^{\circ}$ and $50 \text{ km} \times 50 \text{ km}$, respectively, containing at least 20 active rain gauges. Observations were averaged over the same domains using the Thiessen technique. The gray bands represent the ensemble extremes. The dashed lines bracket 88% of the ensemble members. Average rain gauge observations are drawn as black lines. Events shown are (a),(d) 15 Aug; (b),(e) 13 Sep; (c),(f) 5 Dec. Comparisons between the (a)–(c) EPS ensemble and rain gauge observations, and (d)–(f) LEPS ensemble and rain gauge observations.

of the observations, but these are still contained between the ensemble extremes at the smallest scale of 0.1° . The results of the previous section on power spectra suggest that variance is underestimated, particularly at large scales, a fact that is confirmed in Figs. 4d–f by a severe underestimation at the largest aggregation scales for the events on 15 August and 5 December 2006. The agreement in terms of higher-order moments is similar for both downscaled EPS and LEPS. This suggests that the ability of both downscaling methods in reproducing extreme precipitation peaks is similar.

4. Forecast verification

a. Precipitation fields

We start with a qualitative assessment of the ability of both downscaling methods in forecasting point-by-point precipitation, by reporting in Fig. 5 a comparison between observations and the best members of each forecast ensemble. For each study event, we choose the “best member” in each EPS and LEPS ensemble

that minimizes the root-mean-square (RMS) difference, with the observations at the rain gauge positions over the entire period of 72 h considered. The figure reports the corresponding forecast fields at the moment of peak intensity of the event (see Fig. 2) and, for illustration, the observations, linearly interpolated on the LEPS grid. The figure shows that on the one hand, the LEPS ensemble contains members that are capable of capturing intense precipitation in areas that are characterized by significant orography, such as the alpine areas in northwestern Italy and the coastal region of Liguria. It has to be noted, though, that the detailed positioning of precipitation structures, with respect to observations, agrees only if averaged over length scales significantly larger than the smallest resolved scales and that their intensity is not always captured accurately. Stochastic downscaling, on the other hand, because it does not use orographic information, distributes precipitation at random inside each box of side $L_o = 1^{\circ}$. Nonetheless, at larger scales the EPS ensemble has members that possess some skill in predicting and placing correctly large-scale

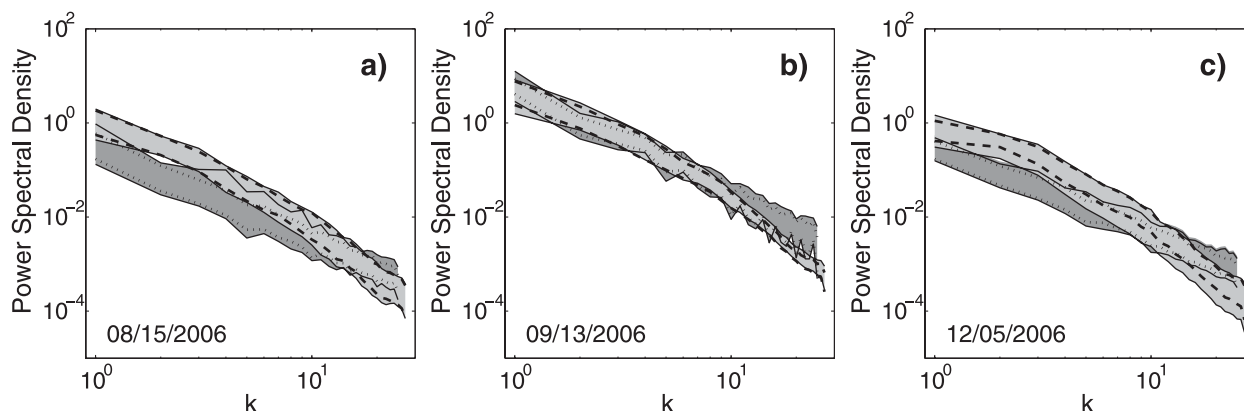


FIG. 3. Spatial power spectra for the LEPS ensemble (light gray) and downscaled EPS ensemble (dark gray) for the events of (a) 15 Aug; (b) 13 Sep; (c) 5 Dec. The dashed lines bracket 88% of the ensemble members. The gray bands represent the ensemble extremes. Note that the particular choice of 88% confidence bands was made because this represents 14 of the 16 members for the LEPS ensemble and 45 of the 51 members for the EPS ensemble.

average precipitation, as can be seen for the intense precipitation predicted over the Alps in Figs. 5g,h and in Liguria in Figs. 5g,i. From these figures it is clear that the LEPS system is capable of producing forecasts which, at a given instant in time, can agree better with observations compared with forecasts produced by stochastic downscaling. In the context of a probabilistic ensemble forecasting system though, it is not possible, a priori, to know which ensemble member will be the best. The forecast is expressed using the ensemble to derive the probability of occurrence of precipitation at a given instant in time and position in space, and all ensemble members contribute equally to this measure. In the following sections, we will analyze if the advantage of dynamical downscaling is maintained when applied in the context of probabilistic ensemble forecasting.

b. RMSE

We proceed with computing the root-mean-square error (RMSE) between each ensemble forecast of 6-h accumulated precipitation and the rain gauge observations, reported in Fig. 6. Also in this case, different spatial scales are considered by spatially aggregating the forecast fields and observations onto coarser grids. For each ensemble member separately, we average the squared forecast errors computed at every available grid point and at every time step. The figure reports the range of RMSE values achieved by the different ensemble members. The agreement between downscaled EPS and LEPS on 13 September and 5 December 2006 is striking because both forecast models span approximately the same RMSE ranges at all scales. For the event starting on 15 August 2006, the downscaled EPS

ensemble presents at small scales an even lower RMSE compared to the LEPS forecasts. Note that in all cases, these RMSE values are comparable to the standard deviations of the observations (cf. Fig. 4), a clear indication that neither the stochastic downscaling procedure (as could be expected) nor the dynamical LEPS downscaling procedure produce ensemble members that match the observations at all points. Nonetheless, both forecast systems produce significantly lower RMSE values with respect to forecasts that have absolutely no skill in the spatial localization of precipitation. We reproduce such reference no-skill forecasts by shuffling in space the positions of the rain gauges before comparing forecast and observations. This way the temporal evolution of total precipitation averaged over the study area (which, as we have shown in Fig. 2, is well represented by the forecasting ensembles) is preserved, but spatial correlations and pattern positioning are destroyed. The continuous lines in the figure represent, for the EPS and LEPS forecast, the limit above which 95% of the RMSE values of 200 such no-skill forecasts fall and lie significantly above the largest RMSE values of both downscaled EPS and LEPS forecasts. In addition, as we will see in the following section, both ensemble forecasting systems show skill in a probabilistic sense.

c. Brier skill scores

We provide a probabilistic evaluation of the skill of the two downscaling approaches by computing Brier scores (Wilks 1995) for the exceedence of a fixed threshold, chosen as the 90th percentile (over all available grid points and times) of observed 6-h precipitation; the threshold was computed separately for each event and for different aggregation scales. These Brier scores are

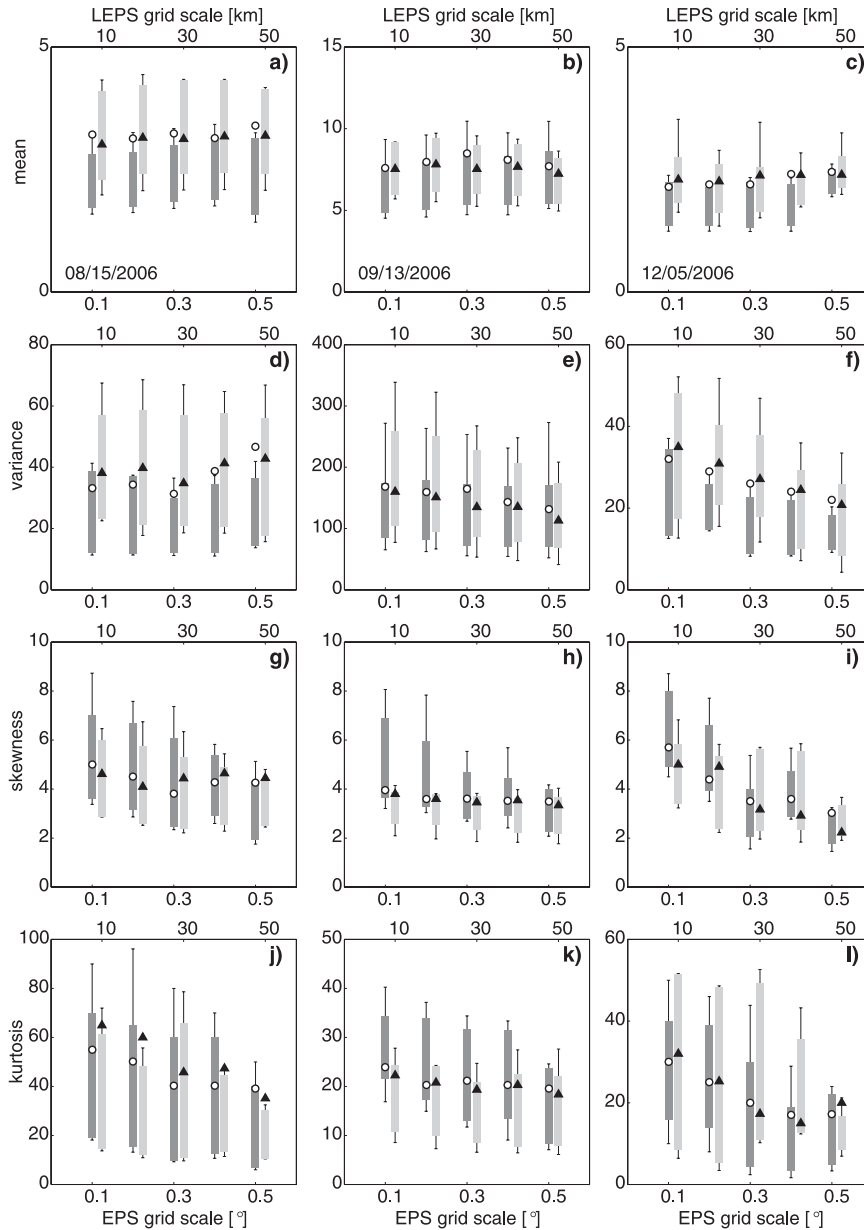


FIG. 4. Comparison of the statistical moments of the distribution for the events of (left) 15 Aug; (middle) 13 Sep; (right) 5 Dec. For each aggregation spatial scale, the light gray (dark gray) bars represent the LEPS (downscaled EPS) ensembles. The variable considered is 6-h accumulated precipitation. The thick bars bracket 88% of the ensemble members, and the fine bars indicate ensembles extremes. The observations are indicated as white circles (black triangles) for EPS (LEPS). (a)–(c) Mean, (d)–(f) variance, (g)–(i) skewness, and (j)–(l) kurtosis.

compared with a null hypothesis of no-forecast skill. As was done in the previous section, we obtain ideal no-skill forecasts by shuffling in space the positions of the rain gauges before comparing forecast and observations. Figure 7 reports Brier skill scores (BSS) with respect to a large number (200) of such no-skill forecasts. The performance of the LEPS and the downscaled EPS ensemble

forecasts is extremely similar at all scales. In fact, although the EPS ensemble event on 15 August 2006 achieves on average a slightly larger BSS, the opposite is true on 5 December 2006. On 13 September 2006, both systems present very low BSS that, at all aggregation scales larger than the smallest (10 km or 0.1°), are not significantly different from zero, corresponding to no-skill forecasts.

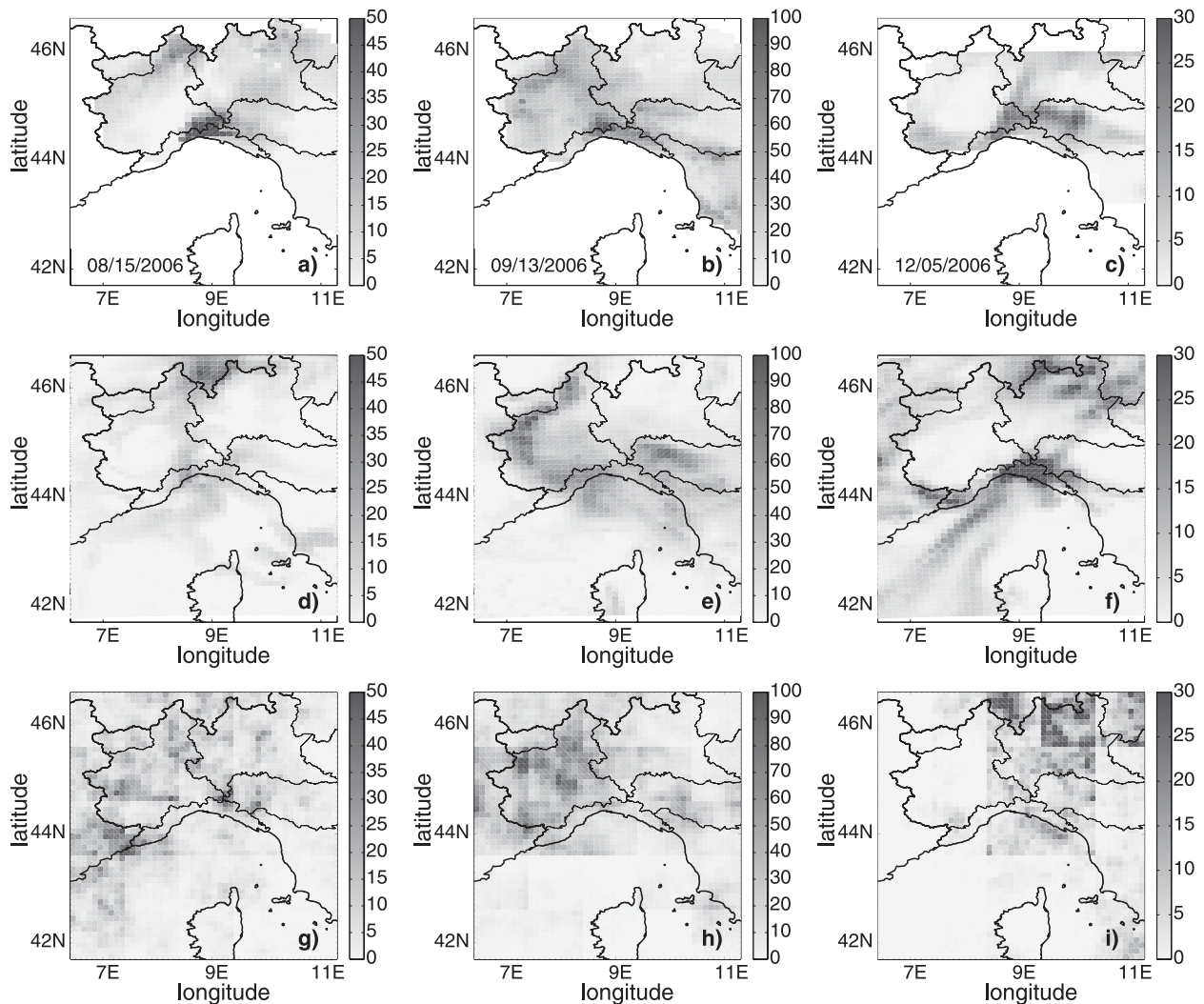


FIG. 5. Comparison of the best member EPS and LEPS forecasts with observations at the moment of peak event intensity for the events of (left) 15 Aug, at $t = 18$ h; (middle) 13 Sep, at $t = 36$ h; (right) 5 Dec, at $t = 24$ h. Refer to text for details on the choice of the best members. The variable considered is 6-h accumulated precipitation. (top row) Observations, linearly interpolated from the rain gauge positions for visualization purposes; (middle row) LEPS forecasts; and (bottom row) downscaled EPS forecasts.

d. Relative operating characteristic

Finally, we compare the discrimination of the probabilistic forecasts (i.e., their ability to distinguish different outcomes) by computing their relative operating characteristic (ROC) curves (Mason 1982; Harvey et al. 1992). For a probabilistic system, the ROC curve illustrates the varying quality of the forecast system at different levels of forecast probability (the probability threshold used to raise a warning).

After again defining as a threshold the 90th percentile of the observed precipitation of each event, we compute hit rates and false-alarm rates provided by the ensemble forecast at different forecast probability levels and compare them in Fig. 8 for the three events at the finest

spatial resolution. The performance of the LEPS and the downscaled EPS ensemble forecasts is extremely similar at this scale, with good skill. In fact, both ensemble systems present ROC curves that fall in the upper-left half of the plot, far from the diagonal no-skill line, and although for the event on 15 August 2006 the LEPS ensemble presents a slightly better ROC curve, the opposite is true on 5 December 2006. On 13 September 2006, the ROC curves of the two systems are almost indistinguishable.

5. Discussion and conclusions

Dynamical downscaling with nested limited-area models has the enormous potential of being able to reproduce processes important at small scales, such as

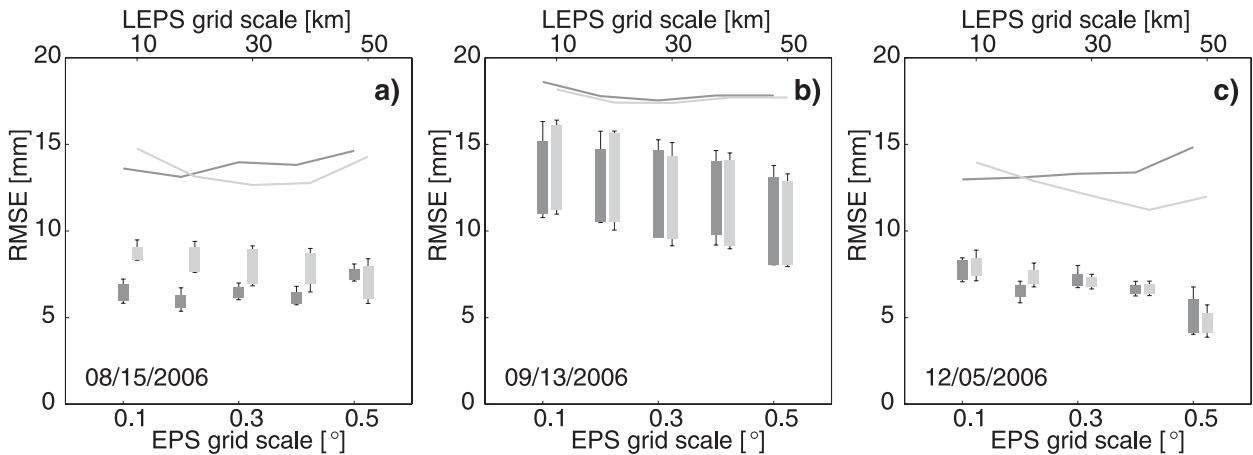


FIG. 6. Comparison of RMSE for the events of (left) 15 Aug; (middle) 13 Sep; (right) 5 Dec. For each aggregation spatial scale, the light gray (dark gray) bars represent the LEPS (downscaled EPS) ensembles. The thick bars bracket 88% of the ensemble members, and the fine bars indicate ensemble extremes. The results are compared with limits provided by 200 reference, no-skill forecasts; 95% of their RMSE values fall above the continuous lines (same coloring as for bars).

nonhydrostatic and orographic effects, and thus to forecast small-scale precipitation features and extremes. On the other hand, it has been shown that, although such downscaling systems can correctly reproduce small-scale properties of the flow in a statistical sense, their forecasting skill is very sensitive to the resolution of the boundary conditions and the initial conditions provided (de Elià et al. 2002; Laprise et al. 2000). This severely limits the skill that can be achieved when they are applied as “intelligent interpolators” in forecasting chains in which no such high-resolution boundary and initial conditions are provided, as happens in most current implementations. When applied to downscale large-

scale ensemble forecasts, current operational implementations of nested limited-area models sacrifice on the number of ensemble members, by downscaling only a part of the available large-scale ensemble and thus abandoning potentially useful probabilistic information it may provide.

Our results presented earlier show that a stochastic downscaling method that is capable of correctly reproducing the small-scale statistical features of observed precipitation may lead to downscaled ensemble forecasts with an essentially equivalent skill, both in terms of RMSE and in its ability to estimate the probabilities of occurrence of intense precipitation. The similar

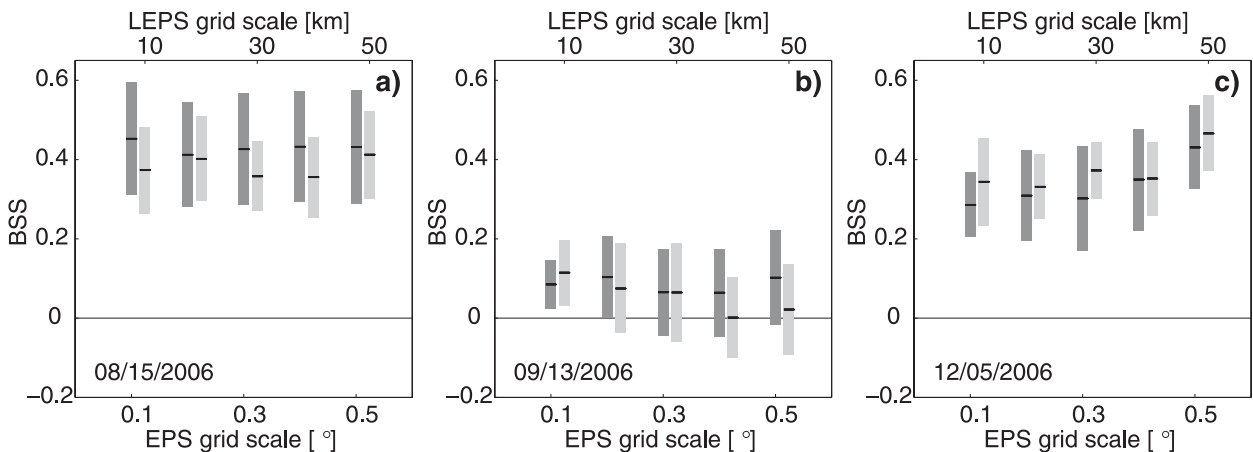


FIG. 7. Comparison of Brier skill scores for the events of (a) 15 Aug; (b) 13 Sep; (c) 5 Dec, obtained comparing forecast Brier scores with reference scores obtained from no-skill forecasts produced shuffling in space the position of rain gauge data. A BSS of 0 corresponds to forecasts no more skillful than the reference, whereas a BSS of 1 corresponds to a perfect probabilistic forecast. The bars indicate the range covered by 95% of the BSS obtained by comparing with 200 independent, reference, no-skill forecasts. The light gray (dark gray) bars show results for the LEPS (downscaled EPS) ensembles. A thin horizontal line marks the average BSS.

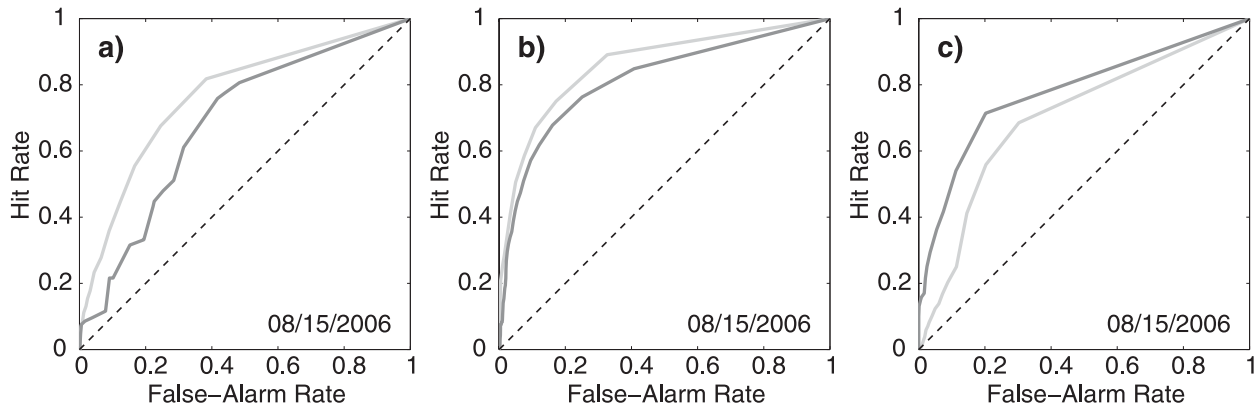


FIG. 8. Comparison of ROC curves (for the exceedence of the 90th percentile of the observations) for the events of (a) 15 Aug; (b) 13 Sep; (c) 5 Dec. The light gray (dark gray) curve represents the LEPS (downscaled EPS) ensembles.

performance of the two downscaling systems considered in this work can be interpreted as a combination of two effects: on one hand, the limited-area model used provides only a marginal advantage over stochastic downscaling in overall forecasting skill at small scales (as shown by the RMSE measures); on the other hand, the downscaled EPS ensemble preserves the number of members of the large-scale ensemble, thus exploring a broader range of scenarios, with a possible advantage in terms of probabilistic skill (we verified by randomly subsampling the downscaled EPS ensemble that if the number of members is reduced to 16, that Brier scores get worse by about 20%, on average).

These results suggest that the skill of current limited-area ensemble systems in forecasting precipitation at small scales could be improved in the future, for example, by providing high-resolution boundary and initial conditions by assimilating small-scale local measurements and by extending the number of ensemble members. At the moment, stochastic downscaling represents a computationally fast alternative to dynamical downscaling for civil protection applications in areas not covered by current limited-area models, and it could be used as a tool to benchmark current nested limited-area models. In such contexts the choice of the scales below which stochastic downscaling intervenes becomes crucial. A forecaster should select the space–time scales at which a forecast model’s prediction skill is the highest and then use stochastic downscaling to extend the forecasts to finer resolutions at which the model is less skillful or unresolved.

Three sample events were considered in this work. The conclusions of this study should be expanded by the exploration of a larger database of events, considering different models and forecast periods. In addition, in the present study we explored extending each large-scale ensemble member with only one stochastic downscaling

realization. Larger ensembles may help to limit residual errors in the forecast due to undersampling of the forecast probability distribution function by dressing the ensemble members with additional variability (Roulston and Smith 2003; von Hardenberg et al. 2007) and thus may lead to improved probabilistic skill of stochastic downscaling ensembles.

Acknowledgments. We thank Luca Ferraris and Franco Siccardi for their useful comments and discussions. EB was supported by the CIMA Research Foundation. This work was partially supported by the project Proscenio, funded by the Italian Civil Protection Department.

REFERENCES

- 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.
- Brussolo, E., J. von Hardenberg, L. Ferraris, N. Reborja, and A. Provenzale, 2008: Verification of quantitative precipitation forecasts via stochastic downscaling. *J. Hydrometeorol.*, **9**, 1084–1094.
- Buizza, R., and T. N. Palmer, 1995: The singular-vector structure of the atmospheric general circulation. *J. Atmos. Sci.*, **52**, 1434–1456.
- , and —, 1998: Impact of ensemble size on ensemble prediction. *Mon. Wea. Rev.*, **126**, 2503–2518.
- , M. Miller, and T. N. Palmer, 1999: Stochastic simulation of model uncertainty in the ECMWF ensemble prediction system. *Quart. J. Roy. Meteor. Soc.*, **125**, 2887–2908.
- Ciach, G. J., and W. F. Krajewski, 1999: On the estimation of radar rainfall error variance. *Adv. Water Resour.*, **22**, 585–595.
- de Elià, R., R. Laprise, and B. Denis, 2002: Forecasting skill limits of nested, limited-area models: A perfect-model approach. *Mon. Wea. Rev.*, **130**, 2006–2022.
- Doms, G., and U. Schättler, 1998: The nonhydrostatic limited-area model LM (Lokal Modell) of DWD. Part I: Scientific documentation. Deutscher Wetterdienst Rep. LM F90 1.35, 172 pp.

- Du, J., S. L. Mullen, and F. Sanders, 1997: Short-range ensemble forecasting of quantitative precipitation. *Mon. Wea. Rev.*, **125**, 2427–2459.
- , J. McQueen, G. DiMego, Z. Toth, D. Jovic, B. Zhou, and H.-Y. Chuang, 2006: New dimension of NCEP short-range ensemble forecasting (SREF) system: Inclusion of WRF members. Preprints, *WMO Expert Team on Ensemble Prediction Systems*, Exeter, United Kingdom, WMO, 5 pp. [Available online at http://www.emc.ncep.noaa.gov/mmb/SREF/WMO06_full.pdf.]
- Epstein, E. N., 1969: Stochastic dynamic prediction. *Tellus*, **21**, 739–759.
- Ferraris, L., S. Gabellani, N. Rebora, and A. Provenzale, 2003: A comparison of stochastic models for spatial rainfall downscaling. *Water Resour. Res.*, **39**, 1368–1384.
- Frogner, I. L., and T. Iversen, 2002: High-resolution limited-area ensemble predictions based on low-resolution targeted singular vectors. *Quart. J. Roy. Meteor. Soc.*, **125**, 1321–1341.
- , H. Haakenstad, and T. Iversen, 2006: Limited-area ensemble predictions at the Norwegian Meteorological Institute. *Quart. J. Roy. Meteor. Soc.*, **132**, 2785–2808.
- Hamill, T. M., and S. J. Colucci, 1997: Verification of Eta-RSM short-range ensemble forecasts. *Mon. Wea. Rev.*, **125**, 1312–1327.
- , and —, 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.
- Harvey, L. O., K. R. Hammond, C. M. Lusk, and E. F. Mross, 1992: The application of signal detection theory to weather forecasting behavior. *Mon. Wea. Rev.*, **120**, 863–883.
- Laprise, R., M. Ravi Varma, B. Denis, D. Caya, and I. Zawadzki, 2000: Predictability of a nested limited-area model. *Mon. Wea. Rev.*, **128**, 4149–4154.
- Leith, C. E., 1974: Theoretical skill of Monte Carlo forecasts. *Mon. Wea. Rev.*, **102**, 409–418.
- Marsigli, C., A. Montani, F. Nerozzi, T. Paccagnella, S. Tibaldi, F. Monteni, and R. Buizza, 2001: A strategy for high-resolution ensemble prediction. II: Limited-area experiments in four Alpine flood events. *Quart. J. Roy. Meteor. Soc.*, **127**, 2095–2115.
- , F. Boccanera, A. Montani, and T. Paccagnella, 2005: The COSMO-LEPS mesoscale ensemble system: Validation of the methodology and verification. *Nonlinear Processes Geophys.*, **12**, 527–536.
- Mason, I., 1982: A model for assessment of weather forecasts. *Aust. Meteor. Mag.*, **30**, 291–303.
- Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliagis, 1996: The ECMWF ensemble prediction system: Methodology and validation. *Quart. J. Roy. Meteor. Soc.*, **122**, 73–119.
- , C. Marsigli, A. Montani, F. Nerozzi, and T. Paccagnella, 2001: A strategy for high-resolution ensemble prediction. Part I: Definition of representative members and global model experiments. *Quart. J. Roy. Meteor. Soc.*, **127**, 2069–2094.
- Palmer, T. N., 1993: Extended-range atmospheric prediction and the Lorenz model. *Bull. Amer. Meteor. Soc.*, **74**, 49–65.
- Patterson, G., and S. Orszag, 1971: Spectral calculations of isotropic turbulence: Efficient removal of aliasing interaction. *Phys. Fluids*, **14**, 2538–2541.
- Rebora, N., L. Ferraris, J. von Hardenberg, and A. Provenzale, 2006a: RainFARM: Rainfall downscaling by a filtered autoregressive model. *J. Hydrometeor.*, **7**, 724–738.
- , —, —, and —, 2006b: Rainfall downscaling and flood forecasting: A case study in the Mediterranean area. *Nat. Hazards Earth Syst. Sci.*, **6**, 611–619.
- Roulston, M. S., and L. A. Smith, 2003: Combining dynamical and statistical ensembles. *Tellus*, **55A**, 16–30.
- Sanders, F., 1986: Trends in skill of Boston forecasts at MIT, 1966–84. *Bull. Amer. Meteor. Soc.*, **67**, 170–176.
- Sattler, K., and H. Feddersen, 2005: Limited-area short-range ensemble predictions targeted for heavy rain in Europe. *Hydrol. Earth Syst. Sci.*, **9**, 300–312.
- Tibaldi, S., T. Paccagnella, C. Marsigli, A. Montani, and F. Nerozzi, 2006: Limited-area ensemble forecasting: The COSMO-LEPS system. *Predictability of Weather and Climate*, T. Palmer and R. Hagedorn, Eds., Cambridge University Press, 489–513.
- Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.*, **74**, 2317–2330.
- , and —, 1997: Ensemble forecasting at NCEP and the breeding method. *Mon. Wea. Rev.*, **125**, 3297–3319.
- , Y. Zhu, and T. Marchock, 2001: The use of ensembles to distinguish between forecasts with small and large uncertainty. *Wea. Forecasting*, **16**, 463–477.
- Tustison, B., D. Harris, and E. Foufoula-Georgiou, 2001: Scale issues in verification of precipitation forecasts. *J. Geophys. Res.*, **106**, 11 775–11 784.
- von Hardenberg, J., L. Ferraris, N. Rebora, and A. Provenzale, 2007: Meteorological uncertainty and rainfall downscaling. *Nonlinear Processes Geophys.*, **14**, 193–199.
- Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences: An Introduction*. International Geophysical Series, Vol. 59, Academic Press, 464 pp.