Downscaling of ECMWF Ensemble Forecasts for Cases of Severe Weather: Ensemble Statistics and Cluster Analysis

ČEDO BRANKOVIĆ, BLAŽENKA MATJAČIĆ, AND STJEPAN IVATEK-ŠAHDAN

Croatian Meteorological and Hydrological Service, Zagreb, Croatia

ROBERTO BUIZZA

European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, United Kingdom

(Manuscript received 3 August 2007, in final form 15 January 2008)

ABSTRACT

Dynamical downscaling has been applied to global ensemble forecasts to assess its impact for four cases of severe weather (precipitation and wind) over various parts of Croatia. It was performed with the Croatian 12.2-km version of the Aire Limitée Adaptation Dynamique Développement International (ALADIN) limited-area model, nested in the ECMWF T_L255 (approximately 80 km) global ensemble prediction system (EPS). The 3-hourly EPS output was used to force the ALADIN model over the central European/northern Mediterranean domain.

Results indicate that the *identical* clustering algorithm may yield differing results when applied to either global or to downscaled ensembles. It is argued that this is linked to the fact that a downscaled, higher-resolution ensemble resolves more explicitly small-scale features, in particular those strongly influenced by orographic forcing. This result has important implications in limited-area ensemble prediction, since it implies that downscaling may affect the interpretation or relevance of the global ensemble forecasts; that is, it may not always be feasible to make a selection (or a subset) of global lower-resolution ensemble members that might be representative of all possible higher-resolution evolution scenarios.

1. Introduction

Synoptic case studies help to identify the capabilities of numerical models to simulate physical and dynamical processes related to severe weather like, for example, strong precipitation, flash floods, gale force winds, and hail storms. Although these studies are normally based on a limited amount of readily available observational data, they provide insights into the strengths and weaknesses of the models under assessment.

The focus of a case study analysis is usually closely linked with the characteristics of the models used in the investigation—whether they are relatively low-resolution atmospheric global circulation models (AGCMs) or relatively high-resolution regional models (RMs). When investigating the properties of an AGCM, the main concern is to establish whether it is able to repro-

duce large-scale circulation that preceded and ultimately led to a given severe weather event. For example, using the European Centre for Medium-Range Weather Forecasts (ECMWF) deterministic AGCM, Jung et al. (2004) studied several European storms: the Dutch storm of 1953, the Hamburg storm of 1962, and the October 1987 storm that severely affected the northern France and the United Kingdom. They concluded that, despite some underestimation of the severity of the storms, reliable predictions of severe weather by ECMWF global model were possible several days in advance.

By contrast, when investigating the properties of a regional model, attention is focused on smaller-scale phenomena, since RMs enable a better spatial and temporal resolution, normally required to accurately analyze synoptic- and mesoscale atmospheric features related to severe weather. For example, Shutts (1990) claims that the Met Office fine-mesh model successfully predicted the details of the October 1987 storm. Several synoptic-scale studies have been made in the Croatian Meteorological and Hydrological Service (CMHS)

Corresponding author address: Čedo Branković, Croatian Meteorological and Hydrological Service, Grič 3, 10000 Zagreb, Croatia.

E-mail: cedo.brankovic@cirus.dhz.hr

with a limited-area model, normally employed for short-range operational forecasting. Various local phenomena have been studied, many of them related to severe weather (e.g., Tudor and Ivatek-Šahdan 2002; Ivančan-Picek et al. 2003; Strelec-Mahović and Drvar 2005; Ivatek-Šahdan and Ivančan-Picek 2006). Although in each of the above CMHS studies a single RM run was made, the results enabled a detailed insight into various aspects of processes defined on relatively small scales.

In the past decade, ensembles of forecasts brought new improvements, both qualitative and quantitative, to the process of operational forecasts and have become an indivisible part of operational forecasting practice in many weather services. Thus, the 1990s pioneering work at ECMWF (e.g., Molteni et al. 1996; Buizza 1997; Palmer et al. 1997), the National Centers for Environmental Prediction (NCEP; Toth and Kalnay 1993), and the Meteorological Service of Canada (Houtekamer et al. 1996) has led to a widespread application of the atmospheric predictability theory in the medium range (e.g., Lorenz 1982). Global ensemble forecasts have been used to revisit interesting cases: for example, Buizza and Hollingsworth (2002) used the ECMWF ensemble prediction system (EPS) to study the December 1999 European storms. They found that the EPS was a valuable tool for assessing the risk of severe weather and issuing early warnings. Similarly, Buizza and Chessa (2002) studied the impact of a stochastic parameterization (Buizza et al. 1999) on the EPS forecasts of a severe storm that affected the United States in January 2000, and concluded that such a parameterization improved the capability of the EPS to simulate the storm's development. Jung et al. (2005) showed that the ensemble forecasts could have provided extremely valuable supplementary information to that generated by a single, high-resolution (deterministic) forecast in the case of the October 1987 storm. They concluded that the EPS was capable of predicting large uncertainties associated with the timing of the storm.

Because of their relatively coarse resolutions, AGCMs cannot resolve small-scale features, and their usefulness to study phenomena at subsynoptic scales is limited. Severe weather events are often associated with a relatively large variability of atmospheric small-scale parameters, which in turn might be related to variations in local topography. Thus, in order to better evaluate and understand local processes, and to be able to verify them against observations, there is a need to carry out dynamical downscaling of global ensembles. In general, dynamical downscaling is a method to obtain high-resolution weather parameters by integrating

an RM that is forced with initial and boundary conditions defined from a coarse-resolution AGCM. The downscaled ensembles might be useful for further applications in e.g., hydrological and crop-yield models. Forecasting flash floods, for example, may be possible by exploiting the uncertainty information from a downscaled ensemble, in contrast to relatively smooth ensemble averages that could mask such an event.

The purpose of this study is to identify the differences in statistical properties between forecast ensembles made by an AGCM and the downscaled forecast ensembles made by an RM. Such an analysis is extended to clusters that were derived in an identical way from both global and regional ensembles. The detailed synoptic analysis of global and regional model ensembles is shown and discussed in the companion paper by B. Matjačić et al. (2008, unpublished manuscript).

There is no unique approach to dynamical downscaling of ensemble forecasts (e.g., Stensrud et al. 1999; Montani et al. 2001; Marsigli et al. 2001; Frogner and Iversen 2002). Some authors, for example, applied the so-called representative members approach (e.g., Molteni et al. 2001; Montani et al. 2001). This implies various intermediate steps in order to define a reduced number of members for dynamical downscaling that characterize all possible evolution scenarios of the global model EPS. It is important to emphasize that in the current study dynamical downscaling is applied to *all* global model ensemble members. While some pieces of information are inevitably lost in the selection of representative members, no such a loss in our study is possible. Thus, dynamical downscaling of the whole global model ensemble represents a more complete way of applying ensembles to smaller spatial scales. Because of a relatively large computational demand, such an approach might not, at present, be viable in an operational practice. However, within the project Grand Limited Area Model Ensemble Prediction System (GLAMEPS; http://www.ecmwf.int/about/special_projects/iversen_GLAMEPS/index.html), currently running at ECMWF, a distributed multimodel production system is planned which would enable individual countries to produce ensemble members and exchange results in real time.

In the next section experiments and methodology are described. Section 3 deals with a general (statistical) comparison between the results from the global model and regional model ensembles. In section 4, an assessment of the global and regional models' clusters, together with the clustering statistics, is presented. Summary and conclusions are given in section 5.

2. Experimental design

a. Models and dynamical downscaling

The global ensembles were generated using the ECMWF model cycle that was in the operational suite from 18 October 2004 until 5 April 2005 (cycle 28R4). The model was integrated at the T_L255 spectral resolution (approximate horizontal resolution 80 km) with 40 levels in the vertical and a 45-min time step. ECMWF EPS (hereafter referred to as ECEPS) initial conditions (ICs) were made as in ECMWF operations. The global model ensembles contained 51 members each.

To create lateral boundary conditions (LBCs) for dynamical downscaling by the hydrostatic version of the Aire Limitée Adaptation Dynamique Développement International (ALADIN) RM, the 3-hourly output from ECEPS was used. LBCs were defined separately from ECMWF model levels and surface data. The upper-air LBCs were interpolated from ECMWF model levels to ALADIN model levels, and also converted to the appropriate format. For surface fields, a selected output from the ECMWF 4-layer land surface model (Van den Hurk et al. 2000) was scaled to accommodate the ALADIN 2-layer surface scheme (Giard and Bazile 2000). The ECMWF surface parameters of soil temperature, soil water, and snow depth were converted to ALADIN's soil temperature, soil water and ice, and snow depth. The ALADIN ICs were handled in the same way as LBCs. No other modification or manipulation with ECEPS output data has been carried out.

The full set of such defined ICs and LBCs from ECEPS was then applied to ALADIN; that is, ALADIN ensembles (hereafter referred to as ALEPS) also contained 51 members. The ALADIN RM was run at the regular 12.2-km grid, at the Lambert conformal projection, with 37 levels in the vertical and with the time step of 514 s (8.6 min). The central point of the integration domain was positioned at 46.2°N, 17°E with 229 grid points in the x direction and 205 grid points in the y direction, thus covering central Europe and the northern Mediterranean. The ALADIN dynamical core is similar to that of ECEPS [the integrated forecasting system (IFS)]; however, the physical parameterizations in the two models are different. The references that describe the ALADIN RM in more details are given in, for example, Ivatek-Šahdan and Tudor (2004).

In addition to physical parameterizations, one of the most important differences between ECEPS and ALEPS is the definition of orography field. Figures 1a,b shows orography in the ALADIN integration domain for both models interpolated, for the comparison

purpose, to the regular $0.5 \times 0.5^\circ$ latitude–longitude grid. The 200-m contour encircles approximately the same area in both models. However, while the maximum height of the Alps in ECEPS is below 1800 m, in the ALADIN model it exceeds 3400 m. The maximum orographic difference between ECEPS and ALEPS models, located in the western Alps at about 46°N, 8°E, exceeds 1800 m (Fig. 1c). When compared to ECEPS, the orography in Croatia has been “raised” in ALEPS between 300 and 600 m in the eastern Adriatic coastal region and its hinterland. It is expected that such differences in orography would contribute to the differences in orography-related fields, like, for example, precipitation or surface temperature.

Because of the large difference between the two model orographies, variations in meteorological parameters appear already at the initial time $T = 0$. For example, vertical cross sections indicate the initial temperature difference in ensemble means of up to 2° – 3° C (not shown). These variations are confined to the lower troposphere and are zero or negligible above 700 hPa. By $T + 6$ h they are reduced over the lowland parts of the integration domain; however, they remain over high mountains.

b. Data

ECMWF ensembles were run with a 3-hourly output for dynamical downscaling with the ALADIN model for two summer (SU) and two autumn (AU) cases (see Table 1 for the list of cases). Whereas ensemble statistics are based on all four cases, the clustering statistics are studied for only three cases. They are the cases of severe weather that occurred over various parts of Croatia in the summer of 2003 and autumn 2004. Based on their intensity and inflicted damage, these three cases by no means may be considered as extreme weather events when compared with some other Mediterranean storms discussed by, for example, Montani et al. (2001, 2003) or by Tripoli et al. (2005). However, they are typical of severe weather that occasionally hits Croatia and therefore are appropriate to be studied by an ensemble prediction system. The fourth synoptic case did not, in terms of severe weather, affect Croatia. For this case, the ECEPS indicated a possible increase in precipitation and wind in the westernmost part of Croatia, and therefore it could be characterized as a sort of a “mild” false alarm.

3. Comparison of the ensembles' performances

A range of performance scores have been computed for both ensembles using the ECMWF operational analyses as verification, which appears to be the best

FIG. 1. Model orography at $0.5^\circ \times 0.5^\circ$ resolution in (a) ALADIN, (b) ECEPS, and (c) the difference: ALADIN minus ECEPS. Contours in (a) and (b) are 400 m starting from 200 m. Contours in (c) are 300 m with positive contours in red and negative contours in blue. The boundaries of Croatia are indicated in (c).

TABLE 1. The four ensembles and three synoptic cases described and discussed in the main body of the paper.

Case	Initial time and date	Target forecast period for case study
SU1	0000 UTC 2 Jul 2003	$T + 54$ to $T + 66$
SU2	0000 UTC 26 Jul 2003	$T + 66$ to $T + 78$
AU1	0000 UTC 12 Nov 2004	$T + 36$ to $T + 72$
AU2	0000 UTC 17 Nov 2004	

common verification field on the lowest-resolution grid (a kind of lowest-common-denominator verification field). However, it might be assumed that, because of the difference in spatial scales between the two ensembles, the choice of ECMWF analysis as the reference for the comparison would favor ECEPS. As it is demonstrated below, for most parameters considered this does not prove to be the case. The exception is vertical velocity, which is most susceptible to the influence of model orography, in particular in the lower troposphere (see the discussion below). In the following, all the fields from both ensembles and from ECMWF analysis were interpolated to the regular $0.5^\circ \times 0.5^\circ$ latitude–longitude grid (approximately 110×75 km at 45°N). For ALADIN, the 12-point quadratic interpolation was used, which introduces a negligible smoothing effect into the high-resolution fields. According to Tustison et al. (2001), such a spatial interpolation inevitably introduces a nonzero scale-dependent error that is independent of model performance. The horizontal resolution of the then ECMWF analysis was T511, approximately 0.35° .

In Table 2, the following three simple and relatively crude accuracy measures are shown: the ensemble mean absolute difference with respect to ECMWF operational analysis δ_e , the ensemble root-mean-square error σ_e (RMSE; also measured with respect to ECMWF operational analysis), and the ensemble spread S_e measured with respect to ensemble mean. For an ensemble of M individual forecast F_i , ensemble mean absolute difference δ_e with respect to verifying analysis A is given by

$$\delta_e = \frac{1}{M} \sum_i^M |F_i - A|, \quad (1)$$

and likewise, ensemble RMSE σ_e could be defined as the following:

$$\sigma_e = \left[\frac{1}{M} \sum_i^M (F_i - A)^2 \right]^{1/2}. \quad (2)$$

Here δ_e and σ_e measure overall departures of the two sets of ensembles from ECMWF verifying analysis and

represent some sort of aggregated forecast error. Ensemble spread S_e is computed as the distance between each individual ensemble member F_i and the mean of the ensemble $\bar{F}_e$, and averaged over all ensemble members. The expression is essentially similar to that for σ_e where the verifying analysis A is replaced by the ensemble mean $\bar{F}_e$ with one degree of freedom less than in σ_e

$$S_e = \left[\frac{1}{M-1} \sum_i^M (F_i - \bar{F}_e)^2 \right]^{1/2}. \quad (3)$$

In terms of probability distribution, the spread of an ensemble measures the dispersion of forecast states. In the presence of small model errors, small spread usually indicates that the ensemble mean is expected to be relatively skilful. All statistical quantities discussed here, δ_e , σ_e , and S_e , are computed for various upper-air parameters at the $T + 48$ h forecast step over the domain 36° – 56°N , 2° – 32°E and for all synoptic cases considered. It is assumed that the selection of meteorological parameters in Table 2 would facilitate the evaluation on what upper-air parameter may serve as a proxy for clustering of the near-surface parameters, such as for example precipitation (see section 4c).

Overall, Table 2 shows that the mean absolute difference δ_e for ALEPS does not differ substantially from that for ECEPS. In 13 out of the total 32 synoptic case–parameter combinations (8 parameters times 4 synoptic cases), the ALEPS δ_e was larger than the ECEPS δ_e (indicated by the bold typeset in Table 2). In 17 combinations, the opposite is seen, that is, the ECEPS δ_e was larger than the ALEPS δ_e ; and in 2 combinations the δ_e values were identical in both models. When the values for vertical velocity ω are excluded from the consideration the total number of the ALEPS combinations being larger than ECEPS drops dramatically from 13 to 5. Such a result broadly implies that, in our four synoptic cases, the dynamical downscaling reduces the forecast error for most parameters considered except for ω . In other words, a better horizontal resolution may introduce errors in parameters closely dependent on orography if they are “verified” against a relatively coarse-resolution reference field. However, if a (hypothetical) high-resolution analysis of the ω field were available, it may show larger errors for ALEPS as the phase errors become more visible, and a higher-resolution forecast may perform worse on point-to-point measures (see, e.g., Baldwin et al. 2001).

When σ_e values for ω parameters are excluded from Table 2, a fairly regular pattern of higher σ_e values for ECEPS than for ALEPS could be clearly seen (18 out of 24 values). This may seem somewhat surprising bear-

TABLE 2. The $T + 48$ h absolute mean difference with respect to ECMWF operational analysis (two left columns), RMSE (two middle columns), and ensemble spread for ECEPS and ALEPS ensembles computed over the domain 36° – 56° N, 2° – 32° E. Boldface indicates the larger value of the two ensembles.

Parameter	Synoptic case	δ_e		σ_e		S_e	
		ECEPS	ALEPS	ECEPS	ALEPS	ECEPS	ALEPS
Z500 (dam)	SU1	1.26	1.20	1.64	1.56	1.41	1.36
	SU2	1.07	1.07	1.45	1.46	1.24	1.29
	AU1	3.73	3.67	4.91	4.83	4.33	4.22
	AU2	2.87	2.78	4.10	3.96	3.49	3.40
Z700 (dam)	SU1	1.14	1.03	1.50	1.36	1.15	1.08
	SU2	1.09	1.05	1.39	1.35	1.08	1.04
	AU1	2.65	2.61	3.40	3.34	2.95	2.85
	AU2	1.98	1.91	2.64	2.59	2.26	2.23
T850 (K)	SU1	1.10	1.03	1.44	1.35	1.17	0.92
	SU2	1.30	1.40	1.70	1.79	1.18	1.11
	AU1	1.74	1.67	2.43	2.35	1.93	1.84
	AU2	1.74	1.52	2.39	2.04	1.66	1.43
Wind 850 (m s^{-1})	SU1	2.38	2.43	3.10	3.18	2.02	2.12
	SU2	2.05	2.05	2.83	2.85	2.03	2.09
	AU1	3.57	3.61	4.87	4.97	3.48	3.61
	AU2	3.81	3.89	5.47	5.45	3.54	3.63
TH 500/1000 (dam)	SU1	1.38	1.33	1.76	1.73	1.37	1.28
	SU2	1.38	1.34	2.00	1.91	1.50	1.54
	AU1	2.94	2.83	4.16	4.07	3.65	3.55
	AU2	3.08	2.76	4.92	4.48	3.62	3.42
TH 700/1000 (dam)	SU1	0.92	0.88	1.17	1.14	0.95	0.81
	SU2	0.98	1.16	1.39	1.49	1.08	1.04
	AU1	1.57	1.49	2.23	2.16	1.86	1.78
	AU2	1.58	1.33	2.32	2.00	1.63	1.46
ω 500 (Pa s^{-1})	SU1	0.19	0.24	0.30	0.39	0.13	0.27
	SU2	0.15	0.18	0.26	0.33	0.19	0.26
	AU1	0.34	0.41	0.52	0.67	0.24	0.48
	AU2	0.22	0.27	0.35	0.44	0.16	0.30
ω 700 (Pa s^{-1})	SU1	0.20	0.26	0.30	0.42	0.15	0.30
	SU2	0.17	0.20	0.28	0.36	0.19	0.29
	AU1	0.36	0.43	0.58	0.71	0.28	0.47
	AU2	0.23	0.31	0.35	0.48	0.20	0.34

ing in mind a potentially higher variability in ALEPS that would have stemmed from better-resolved smaller spatial scales. A partial explanation for such results is that most values for the nonomega parameters in Table 2 are related to upper-air fields that usually have relatively smooth features above model orography.

The higher ALEPS than ECEPS σ_e values for wind magnitude at 850 hPa is indicative of an increased influence of a high-resolution orography on the low-level circulation variability. Clearly, the higher ALEPS orography “interferes” with the 850-hPa atmospheric flow in more grid points than in ECEPS. Consistently larger δ_e and σ_e values for the autumn than for summer cases, irrespective of the model, is associated with an increased natural atmospheric variability due to the seasonal cycle.

For temperature at 850 hPa (T850), both models largely underestimate analyzed temperature over much of western and central Europe (not shown), that is, in

the regions with a relatively inconspicuous orography (Fig. 1). These results indicate that the largest differences between the two models are not always associated with high orography.

Ensemble spread S_e (the last two columns in Table 2) is constantly higher in ALEPS than in ECEPS for wind at 850 hPa (wind 850) and both ω fields. For ω , the spread at the initial time ($T = 0$) has already a larger amplitude in ALEPS than in ECEPS. Therefore, at small spatial scales in ALADIN there is more potential for the upscale error growth, in particular for those fields that are in general more influenced by orography. This statistic is consistent with ensemble RMSE with respect to operational analysis σ_e .

To quantify the statistical significance of the differences between ALEPS and ECEPS distributions of scores shown in Table 2, the Wilcoxon–Mann–Whitney rank-sum test (hereafter referred as to WMW; see, e.g., Wilks 1995) was applied. This is a nonparametric test

where data distribution does not assume any specific form, and it is “resistant,” that is, it is not affected by a few outliers. Because of a small sample, for all variables in Table 2 (except for ω fields), one value of the WMW rank-sum test was computed following the rescaling of the scores.

Consider, for example, the four mean absolute difference values $\delta_e(j)$ for geopotential height at the 500-hPa level (Z500) of ECEPS (first four values in the first column of Table 2). Each rescaled value is defined as $sc(j) = [\delta_e(j) - m]/s$, where m and s are the mean and the standard deviation of the four values, respectively. Once all 48 rescaled scores for the first 6 parameters in Table 2 have been defined, the WMW rank-sum test was computed. The same procedure was applied to the RMSE σ_e and ensemble spread S_e . The WMW rank-sum test has a value of 23% for the mean absolute difference, 24% for the RMSE, and 26% for the ensemble spread, indicating probabilities that the two sets of values for mean absolute differences (RMSE, ensemble spread) belong to the same distribution. Thus, the null hypothesis that “the two distributions of scores come from the same underline distribution” can be rejected at the 77% (76%, 74%) level. To further assess the significance of the difference between the ECEPS and ALEPS scores, a bootstrapping method (e.g., Press et al. 1995) has been applied to compute percentiles and confidence intervals of the rescaled scores: results indicate that for all scores in Table 2 ALEPS has, on average, lower values than ECEPS.

The WMW rank-sum test has not been applied to ω fields because, according to Table 2, their values clearly fall into a distribution that cannot be ranked together with the scores from the other parameters. In addition, at very small scales, orographically induced “irregularities” affect the overall distribution of ω to be very different from that at the coarse-resolution global model (not shown).

Finally, the spectral distributions of the 850-hPa wind magnitude variability in the two models have been compared. Figures 2a,b show power spectra for the AU1 case as a function of longitude (x distance for ALEPS) and oscillation period, calculated from model data at respective original resolutions for the domain nearest to 43°–47°N, 13°–19°E. The term “nearest” is used here because the original model grids do not correspond exactly to the above geographical domain (because of different resolutions and projections). A small error is introduced in this case, but, for the relatively small domain considered, it does not have a critical overall impact on the results shown. The above domain was chosen because all major orographic differences over Croatia and its hinterland are found there (Fig.

1c). The time series from $T + 0$ to $T + 120$ h were sampled every 6 h at each grid point, averaged over the y direction (latitude) and over all ensemble members. The vertical axis in Figs. 2a,b denotes high frequency (top) and low frequency (bottom) for a given forecast extent, and the horizontal axis indicates the spatial variation in the west–east direction.

Results indicate that the ECEPS has more power than ALEPS in low frequencies. In both models, the low-frequency maxima are extending beyond day 5, indicating that some atmospheric phenomena have a major impact on time scales beyond the limit of our integrations. In ALEPS, a somewhat “noisier” spectrum is apparent, caused by a finer horizontal resolution and a stronger orographic impact. At higher frequencies, there is more power in ALEPS than in ECEPS. The maximum is found at about 1350 km from the origin of the ALEPS domain, approximately corresponding to 16.5°E. This is the longitude where an intensive synoptic activity in the AU1 case occurred [cf. Fig. 8 and B. Matjačić et al. (2008, unpublished manuscript)].

Figure 2c shows that, in shorter waves, there is more kinetic energy in ALEPS than in ECEPS as the forecast time evolves (note that only the first 45 waves are shown, because, at the model level nearest to 850 hPa, energy is rather small for the higher wavenumbers). At initial time $T + 0$, both models have almost identical power up to the wavenumber 15. For the higher wavenumbers (>15), ALEPS displays higher energy levels than ECEPS. This essentially implies that the interpolation to the ALEPS grid adds energy to high wavenumbers; this is presumably associated with the areas where there is more variability in the orography. At $T + 48$ h, energy in ALEPS has increased relative to that at $T + 0$, whereas it drops in ECEPS. At this forecast time step, the largest difference between the models is found for wavenumbers larger than 15, signifying a dominant impact of small spatial scales. At $T + 96$ h, the graphs are a little lower in both ensembles than in the corresponding ones at $T + 48$ h, indicating that energy has leveled off. The overall difference between the two ensembles at this forecast time step remains very similar to that at $T + 48$ h.

4. Clustering

Clustering is the method by which individual forecasts from an ensemble that are close to each other are grouped together, thus condensing the large volumes of information in ensembles. The “closeness” of individual forecasts is based upon some objective criterion, for example, the RMS difference among individual members. Clustering enables the identification of some

FIG. 2. Spectra of the 850-hPa wind magnitude as a function of longitude for ECEPS (x direction with respect to the origin of the integration domain for ALEPS) and 6-hourly periods in AU1 case for (a) ECEPS and (b) ALEPS. Contour interval (CI) $\equiv 4 \text{ m}^2 \text{ s}^{-2}$. In (c) the ensemble mean kinetic energy as a function of wavenumber for SU2 case at the ALADIN and ECMWF model level nearest to 850 hPa.

atmospheric features (flows, synoptic regimes) that might attain an increased probability for the occurrence. For example, in a forecasting system with relatively small systematic errors the most populated cluster would be normally accepted as the one having the highest probability (chance) of realization.

Once the clustering criterion has been defined (e.g., RMS difference), the choice of clustering method is usually arbitrary, possibly resulting with different outcomes; that is, the clustering results are by no means unambiguous. For example, even an enlargement (a reduction) of the clustering geographical domain may

have an effect on clustering results since parts of various synoptic systems (lows, fronts, etc.) may then be included (missed out) in the clustering procedure. However, our main concern is to evaluate the differences between the two ensembles that have been clustered using the *identical* clustering method over the same domain.

The most common algorithm for clustering is the so-called Ward hierarchical clustering algorithm (Anderberg 1973). The algorithm merges pair of clusters (initially pair of individual members) that will result in the minimum sum of squared distances between the points and the centroids of their respective groups, summed over all resulting groups. The process is repeated until a predefined number of clusters (or other criterion) is reached (see, e.g., Wilks 1995). It is at the core of the standard ECMWF clustering algorithm (Atger 1999).

We are not suggesting that one clustering method is better than the other. Our analysis and conclusions emphasize the comparison between two identically manipulated ensembles rather than focusing on the choice of the clustering method. After clustering both ensembles, the comparison of the results would enable a better insight and a more detailed assessment of potential benefits of dynamical downscaling. The smaller spatial scales contained in ALEPS may affect some ensemble properties, and eventually clustering results for ALEPS could be different from those for ECEPS.

In the following, we show and discuss the results of the clustering method that is based upon the choice of predefined cluster number (PCN). It has been decided beforehand that no more than three clusters would be generated; that is, $PCN = 3$. In addition, we discuss some results of the clustering method that is based upon the proportion of ensemble variance explained (PVE). PVE is defined by the following expression: $PVE = (V_c/V_e) \cdot 100$, where V_e represents total ensemble variance and V_c stands for intercluster variance. At the starting point of clustering procedure $V_c = V_e$, that is, every single member of an ensemble is assumed to be a separate “cluster”, thus yielding $PVE = 100\%$. In the consecutive steps, as the number of clusters is being reduced (because ensemble members are grouped together), V_c gets smaller. The clustering will automatically stop when the PVE falls below a given threshold. In our case, the PVE was set to 50%, which represents a compromise threshold—a sufficiently large variance explained without too many clusters generated. Again, the PVE clustering algorithm is applied in identical way to both sets of ensembles. As before, both ECEPS and ALEPS data were interpolated to the regular $0.5^\circ \times 0.5^\circ$ latitude–longitude grid. The clustering was performed over the domain 36° – 56° N, 2° – 32° E for

TABLE 3. Cluster size of each cluster (1, 2, 3) for ECEPS and ALEPS. Column C is the number of common members in both ECEPS and ALEPS from the most populated or the second most populated (boldface) clusters. The base clustering time is $T + 48$ h.

Parameter	Synoptic case	ECEPS clusters				ALEPS clusters		
		1	2	3	C	1	2	3
Z500	SU1	22	23	6	<i>18</i>	33	12	6
	SU2	30	12	9	<i>21</i>	11	21	19
	AU1	32	13	6	<i>18</i>	20	24	7
	AU2	16	23	12	<i>20</i>	32	14	5
Z700	SU1	31	14	6	<i>23</i>	24	21	6
	SU2	19	20	12	<i>17</i>	17	18	16
	AU1	26	11	14	<i>24</i>	24	12	15
	AU2	17	19	15	<i>19</i>	17	19	15
T850	SU1	19	18	14	<i>16</i>	30	10	11
	SU2	14	18	19	<i>19</i>	24	18	9
	AU1	34	10	7	<i>33</i>	34	8	9
	AU2	19	24	8	<i>23</i>	26	16	9
Wind 850	SU1	22	22	7	<i>21</i>	27	10	14
	SU2	23	12	16	<i>15</i>	24	23	4
	AU1	25	14	12	<i>24</i>	26	14	11
	AU2	21	18	12	<i>21</i>	26	18	7
TH 500/1000	SU1	28	11	12	<i>21</i>	32	13	6
	SU2	21	19	11	<i>19</i>	20	20	11
	AU1	32	9	10	<i>17</i>	27	21	3
	AU2	29	16	6	<i>29</i>	29	16	6
TH 700/1000	SU1	27	11	13	<i>14</i>	28	13	10
	SU2	31	14	6	<i>18</i>	15	18	18
	AU1	33	9	9	<i>30</i>	30	9	12
	AU2	24	19	8	<i>20</i>	20	24	7
ω 500	SU1	24	14	13	<i>22</i>	27	16	8
	SU2	11	28	12	<i>9</i>	16	21	14
	AU1	10	30	11	<i>12</i>	18	16	17
	AU2	18	20	13	<i>14</i>	24	17	10
ω 700	SU1	22	10	19	<i>18</i>	25	15	11
	SU2	34	10	7	<i>19</i>	20	21	10
	AU1	32	8	11	<i>23</i>	12	24	15
	AU2	19	19	13	<i>10</i>	24	15	12

all parameters from Table 2 at the $T + 48$ h forecast step.

a. Cluster size and common members

First we discuss some basic properties of clustering for the two sets of ensembles and the differences between them. For the PCN clustering method, Table 3 shows the cluster size (number of individual members) for all three clusters in both ECEPS and ALEPS. The ordering of clusters in Table 3 was defined by the clustering algorithm; that is, they have not been sorted out according to their size. Overall, between ECEPS and ALEPS there are more differences than similarities. Only in 2 out of the total 32 parameter–synoptic case combinations the clustering algorithm yields identical

result: for geopotential height at the 700-hPa level (Z700) and thickness between the 500- and 1000-hPa levels (TH 500/1000), both in the AU2 synoptic case. In both cases, the ECEPS and ALEPS clusters have not only the same number of members but also the same members. However, the resulting mean fields for corresponding clusters may not be necessarily identical, although in our case they are very similar (not shown). It should be pointed out that the cluster means could also be similar even with some differences in the cluster membership. A few outliers may not alter the mean pattern of a large number of highly similar members.

For most combinations in Table 3 the differences in size between ECEPS and ALEPS clusters are relatively large. Figure 3 illustrates these differences for Z500 in the AU1 synoptic case—the colored areas are the differences between cluster means and ECMWF analysis (cluster “errors”). It may be argued that there is some similarity between the most populated clusters: 32 members in ECEPS (Fig. 3b) and 24 members in ALEPS (Fig. 3c). However, the error pattern and amplitude for the second most populated clusters (13 members in ECEPS, Fig. 3d; 20 members in ALEPS, Fig. 3a) look very different. The third ALEPS most populated cluster (7 members, Fig. 3e) shows some similarity with the ECEPS cluster number 2 (Fig. 3d), as does the second most populated ALEPS cluster with the third ECEPS clusters (6 members). Clearly, from Fig. 3, similar error patterns between the ECEPS and ALEPS clusters arise from including identical individual ensemble members in both clusters. For example, all 7 members from the ALEPS third cluster (Fig. 3e) are included in the second ECEPS cluster (with 13 members; Fig. 3d).

When the PVE clustering method is applied, the number of clusters generally increases by one or two, but for some parameters (wind 850 and ω fields) such an increase is even larger. However, in the autumn season, the PVE clustering method produces the same answer as the PCN clustering for geopotential height and TH 500/1000. Thus, in certain situations (parameter, season), for a given clustering criterion (in our case it was the RMS difference), different clustering methods (PCN, PVE) can produce identical results.

The number of common members, that is, ensemble members residing in both ECEPS and ALEPS most populated or second most populated clusters, is shown in the middle column in Table 3 (denoted C). Even in the case when ECEPS and ALEPS clusters have a high number of common members, errors in cluster means may not look identical or even similar (not shown). Many common members do not necessarily guarantee similarity among clusters from the two different popu-

lations. Since an identical clustering algorithm is applied to both sets of ensembles at the same grid, the different results indicate the effect of dynamical downscaling.

b. Clustering statistics

To quantify in more details the differences between the ECEPS clusters and the ALEPS clusters the following additional calculations were performed: (i) the mean distance d_i of the i th ALEPS cluster members from the j th ECEPS cluster mean (centroid), and vice versa; (ii) the distance D_i between the i th ALEPS and the j th ECEPS centroids; and (iii) the so-called representative members of ALEPS and ECEPS clusters. Consider first the mean distance d_i as defined in (i). If dynamical downscaling does not affect clustering properties and the *same* clustering algorithm is used, then the mean distance of the i th ALEPS cluster with respect to the j th ECEPS cluster centroid should be smallest for $i = j$. The same consideration is applicable to the distance D_i from (ii) where only cluster centroids in both ALEPS and ECEPS are taken into account. The definitions of d_i and D_i differ in the sense that for d_i the cluster’s internal spread is implicitly taken into consideration. Finally, for each cluster C_i , the representative member is defined as the ensemble member with smallest ratio between the average distance from ensemble members belonging to cluster C_i and the distance from ensemble members *not* belonging to cluster C_i (see Molteni et al. 2001 for details).

For the i th ALEPS cluster, the mean absolute distance d_i (averaged over all cluster members) from the j th ECEPS cluster centroid C_j^G is computed as

$$d_i = \frac{1}{M_i} \sum_k |F_k^R - C_j^G|. \quad (4)$$

The summation index k runs over all F_k^R members of the i th ALEPS cluster (in total M_i members of the i th cluster). Similarly, the mean distance d_j of the j th ECEPS cluster from the i th ALEPS cluster centroid C_i^R can also be calculated.

For our 32 synoptic case–parameter combinations and PCN = 3, there are in total 96 pairs of ALEPS and ECEPS clusters that fulfill the criterion $i = j$. In 51 out of these 96 pairs the mean distance d_i between the ALEPS clusters and ECEPS centroids C_j^G is the smallest. In other words, there are almost as many pairs (45) for which dynamical downscaling causes the distance d_i to be the smallest when $i \neq j$. The best results are found for Z700, where in all 12 pairs for which $i = j$ the distance d_i is the smallest. Similar statistics emerge

ALEPS

ECEPS

FIG. 3. The 500-hPa geopotential height clusters for the AU1 synoptic case at $T + 48$ h for (left) ALEPS and (right) ECEPS. CI = 4 dam for cluster means (solid), and 3 dam for errors with respect to ECMWF operational analysis (colored).

when the distances $D_i = |C_i^R - C_j^G|$ between ALEPS and ECEPS cluster centroids are computed. In this case, in 51 out of 96 pairs the distance D_i is the smallest when $i = j$. For most pairs when $i = j$ and D_i is not the smallest (in total 45), the distances between the ECEPS

and ALEPS cluster centroids are, on average, relatively large with respect to the minima found. Thus, a relatively large fraction of the ALEPS centroids exhibits a major shift from their ECEPS match. To summarize, this kind of statistics indicates that dynamical downscal-

FIG. 4. Scatter diagrams for pairs of normalized mean distances d_i and d_j for (a) Z500, (b) Z700, and (c) $\omega 700$.

ing induces nonnegligible differences between clusters from global and regional ensembles even if the clustering algorithm was identical.

To further support the findings above, Fig. 4 shows examples of scatter diagrams when all pairs of distances d_i and d_j are plotted against each other. For each parameter and season, the distances are normalized by

ensemble spread in order to obtain comparable values. For a given parameter and $PCN = 3$, there are 9 possible pairs (combinations) of i and j indices; hence 9 symbols for each season. If there were no impact of dynamical downscaling on clustering properties, the symbols would be positioned along the diagonal. Thus, it could be assumed that the departures from the di-

agonal measure the impact of dynamical downscaling on regional clusters. Three “types” of scatter diagrams can be seen in Fig. 4. For Z500 (Fig. 4a; a similar scatter diagram can be obtained for TH 500/1000 and TH 700/1000), the dispersion is relatively large, whereas for Z700 (Fig. 4b; and similarly for T850 and wind 850) it is less so. For the latter, there is a tendency of the clusters to group closer to the diagonal, in particular in autumn, possibly indicating some significance of seasonal cycle on clustering. For both ω fields ($\omega 700$ in Fig. 4c), the symbols are arranged closely along the line that is rotated relative to the diagonal. This “tightness” in ω fields is explained, at least partly, by a relatively larger ensemble spread for vertical velocity found in ALEPS than in ECEPS (see Table 2) that has an effect on the normalized distance d_i .

We also compare the (interensemble) distances D_i with the (intraensemble) distances among the ECEPS centroids $\Delta_i = |C_i^G - C_j^G|$. The distances Δ_i may serve as a reference against which to judge the significance of the distances D_i . The results broadly confirm the above findings. When the ω fields are excluded from consideration, the number of $D_i > \Delta_i$ is about equal to the number of $D_i < \Delta_i$ (with negligible number of $D_i = \Delta_i$); when ω fields are included, the number of $D_i > \Delta_i$ increases considerably.

Finally, we briefly compare and discuss cluster representative members for both ALEPS and ECEPS defined by following the procedure proposed by Molteni et al. (2001). For the same cluster index (that is, when $i = j$) only in 27 out of total 96 ALEPS–ECEPS pairs (less than 30%) the representative members in both populations are found to be identical. The distance $r_i = |R_i^R - R_j^G|$ between the ALEPS representative member R_i^R and the ECEPS representative member R_j^G for identical clusters (that is, when $i = j$) essentially yields the results similar to those for d_i and D_i discussed above: r_i was found to be the smallest in 46 out of 96 combinations. For the PVE clustering method, it was found that most representative members are identical to those from the PCN method above.

These results imply that similarities found among the members of a given global model cluster might not necessarily be seen or carried over to the subsequent downscaled forecasts.

c. Clustering proxy for precipitation

When upper-air fields are considered as the basis for the clustering of precipitation, the question of which parameter could be used as the best-defined proxy arises. Since no unambiguous relationship between precipitation and upper-air fields exists, various testing has

been performed. Clustering time for precipitation proxies is based on the hindsight knowledge of each synoptic case; that is, we already knew the time intervals when severe weather events (in this case heavy precipitation) occurred. This, of course, would not be possible in an operational forecasting environment; however, our aim is to find out the best possible estimates for the clustering of parameters that constitute severe weather.

1) RAIN GAUGE DATA

Before discussing the clustering results, Croatia’s rain gauge data for the first three synoptic cases are described [the reader is referred to B. Matjačić et al. (2008, unpublished manuscript) for a detailed synoptic analysis of the four case studies from Table 1].

In Fig. 5a, the 24-h rain gauge accumulated precipitation for the period between 0600 UTC 4 July and 0600 UTC 5 July indicates the rain totals between 50 and 56 mm. On 4 July between 0800 and 1000 LT, the towns of Šibenik and Split (marked by open black circle and open square, respectively) were hit by heavy rain. In Split more than 30 mm fell in 40 min. The then ECMWF operational deterministic T_L511 model (with horizontal resolution of approximately 40 km) predicted between 1 and 5 mm in 12 h. Operational ALADIN predicted between 1 and 10 mm; however, it was shifted farther inland from the place where maximum precipitation actually occurred.

During the night of 28/29 July between 2000 and 0400 UTC many stations in northwest Croatia reported heavy precipitation with thunder and strong winds (Fig. 5b). The rain gauge at Kapela (shown as open square) measured 60.9 mm in 24 h (Zagreb is marked by black diamond.). The ECMWF operational deterministic model predicted between 5 and 10 mm of rain. Operational ALADIN predicted more than 10 mm; however, it was slightly displaced to the south of the region of interest.

Figure 5c shows that heavy precipitation affected the southern Adriatic coastal region between 0600 UTC 13 November and 0600 UTC 14 November. In Split (marked by open square), 31.7 mm of rain fell in 24 h. In Ploče (triangle) and Dubrovnik, 45.5 and 41.1 mm of rain, respectively, fell in 6 h, between 0700 and 1300 UTC on 14 November. In 24 h it ultimately amounted to 94.2 and 76.9 mm, respectively. The maximum precipitation of 123.4 mm in 24 h was measured in Gornje Sitno (diamond), and 87.8 mm was recorded in Blato (hexagon). For this synoptic case the ECMWF T_L511 model predicted precipitation amounts higher than 50 mm(12 h)^{−1} only farther south, and it predicted no precipitation in the northern Adriatic. Similar results were obtained by the operational ALADIN model.

FIG. 5. 24-h accumulated precipitation for cases (a) SU1, (b) SU2, and (c) AU1. Only rain gauges with more than 20 mm (24 h)⁻¹ are shown.

2) MODEL RESULTS

For the SU1 case, Fig. 6 shows the precipitation rate based on the clustering of TH 700/1000 in all ECEPS (left) and ALEPS (right) clusters. Precipitation for this parameter in the ALEPS cluster 1 (most populated cluster; Fig. 6, top right), is closest to observations from Fig. 5a. The 30-mm contour is only slightly shifted eastward relative to observations. The ALEPS cluster 1 contains all 9 members of the ECEPS cluster 3 (Fig. 6, bottom left), the ECEPS cluster that produced the highest precipitation rate in the area of interest (the rectangle in Fig. 6). Based on this, it may be assumed that there is a good correspondence between the global and regional models' members that would generate higher rainfall totals, even if the clustering algorithm distributes those members into the different clusters. However, an additional 14 members are also contained in the ALEPS cluster 1, indicating that the added value to the ALEPS cluster 1 comes partly from these additional members.

More importantly, it could be argued that the increased precipitation in the southern Dalmatian hinterland in ALEPS, compared to ECEPS, is due to the impact of a better-resolved ALADIN orography. From Fig. 1c, it is clear that, in the area considered, the orography difference between the ALADIN and the ECEPS models is more than 600 m in places. For this synoptic case, the orientation of the mountains in the eastern Adriatic hinterland must have also played a crucial role to increasing precipitation [this is discussed in detail in B. Matjačić et al. (2008, unpublished manuscript)]. A similar increase in the ALEPS precipitation is seen over central Italy where the Apennines represented in ALADIN are higher by more than 1000 m than in ECEPS.

It is important to emphasize that the best ALEPS result shown in Fig. 6 is related to the most populated cluster. However, for some other parameters this is not the case, and other, less populated clusters are closest to observations. Of course, such an analysis might be relevant only in hindsight; in practice, for real-time forecasting, less populated clusters would only serve as an indication of a possible alternative development; that is, they would quantify forecast uncertainty.

When compared with the SU2 verification data (Fig. 5b), precipitation based on the clustering of ALEPS is no better than that based on the clustering of ECEPS (Fig. 7). The additional clustering for other base times, forecast time intervals, and parameters brought no improvement to either ensemble. The detailed analysis reveals that only a negligible fraction of ensemble members predicted a precipitation rate close to ob-

served (not shown). However, their impact is masked by forcing each of these individual members into any of the three clusters.

For the AU1 case (Fig. 5c), ECEPS was relatively successful irrespective of the clustering parameter and a high degree of consensus among individual members was attained. For the $\omega 500$ clustering base, Fig. 8 (left) shows that all ECEPS clusters generate more than 20 mm of precipitation $(12 \text{ h})^{-1}$ in the southeastern Adriatic and more than 30 mm in or near the Dubrovnik area. Moreover, a gradual reduction in precipitation amounts toward the northern Adriatic was also correctly predicted: in the northern Adriatic town of Senj (marked as SE) the 12-h accumulation centered at 0000 UTC 14 November was only 7 mm.

The ALEPS clusters (Fig. 8, right) indicate correctly more than 30 mm/12hr precipitation in the Ploče (PL) area and more than 50 mm $(12 \text{ h})^{-1}$ in the Dubrovnik (DU) area, respectively. The largest precipitation amounts are found further south down the eastern Adriatic coast. Clearly, the ALEPS clusters improved the detailed distribution of high precipitation amounts in the southern Adriatic relative to ECEPS, although the performance of the latter was very good indeed. When compared to observations, both ECEPS and ALEPS performed better than ECMWF and ALADIN deterministic models.

For this synoptic case, there is only a little difference among various parameters used as the clustering base for precipitation. Such consistency could be useful in an operational forecasting practice, since it would eliminate redundant results.

5. Summary and conclusions

The impact of dynamical downscaling on ECMWF global model ensemble forecasts for four synoptic cases over Croatia was investigated. The 5-day global forecasts were generated using the ECMWF 51-member EPS (ECEPS). The 3-hourly ECEPS outputs were used to force the ALADIN limited-area model over the central Europe and northern Mediterranean domain, thus creating downscaled ensembles (ALEPS) with 51 members each. The ECEPS was run at the T_{L255} spectral resolution (roughly corresponding to 80 km) with 40 levels in the vertical, and ALEPS was run at the 12.2-km regular grid with 37 vertical levels. The outputs from both the ECEPS and ALEPS ensembles were manipulated in the same way in order to determine the impact of dynamical downscaling and possible gains that could be attained for studying synoptic cases of severe weather.

Two measures of forecast error calculated with re-

FIG. 6. The 12-h accumulated precipitation between $T + 54$ and $T + 66$ in (top) cluster 1, (middle) cluster 2, and (bottom) cluster 3 for (left) ECEPS and (right) ALEPS in the SU1 synoptic case. Clustering is based on TH 700/1000. The rectangle denotes the verification area shown in Fig. 5a. Contours are 1, 5, 10, 20, ... mm (12 h)⁻¹

FIG. 7. As in Fig. 6 but for the 12-h accumulated precipitation between $T + 66$ and $T + 78$ in (top) cluster 1, (middle) cluster 2, and (bottom) cluster 3 in the SU2 synoptic case. Clustering is based on TH 500/1000. The rectangle denotes the verification area shown in Fig. 5b.

ECEPS

ALEPS

ECMWF_255 00UTC, 12h precipitation 20041112 +54-42 h cl_1 - 10 members [56/ 2/ 36/ 32]
base_w500+48: 2 3 8 10 15 17 29 34 44 47

ALADIN 00UTC 12h prec: 20041112 +54-42 h cl_1 - 18 members [56/ 2/ 36/ 32]
base_w500+48: 2 3 8 12 15 17 19 21 23 29 32 33 34 39 41 44 47 50

ECMWF_255 00UTC, 12h precipitation 20041112 +54-42 h cl_2 - 30 members [56/ 2/ 36/ 32]
base_w500+48: 0 1 4 6 9 11 13 14 16 18 20 22 24 25 26 27 28 30 31 33 35 37 38 40 41 45 46 48 49 50

ALADIN 00UTC 12h prec: 20041112 +54-42 h cl_2 - 16 members [56/ 2/ 36/ 32]
base_w500+48: 4 6 10 11 14 16 18 20 22 24 27 30 31 35 40 48

ECMWF_255 00UTC, 12h precipitation 20041112 +54-42 h cl_3 - 11 members [56/ 2/ 36/ 32]
base_w500+48: 5 7 12 19 21 23 32 36 39 42 43

ALADIN 00UTC 12h prec: 20041112 +54-42 h cl_3 - 17 members [56/ 2/ 36/ 32]
base_w500+48: 0 1 5 7 9 13 25 26 28 36 37 38 42 43 45 46 49

FIG. 8. As in Fig. 6 but for the 12-h accumulated precipitation between $T + 42$ and $T + 54$ in (top) cluster 1, (middle) cluster 2, and (bottom) cluster 3 in the AU1 synoptic case. Clustering is based on ω_{500} . The rectangle denotes the verification area shown in Fig. 5c.

spect to ECMWF operational analysis (mean absolute difference and RMSE) over the downscaling domain indicate that the errors in ECEPS and ALEPS upper-air fields are comparable. However, when the ω field is excluded from the consideration, the overall picture changes in favor of ALEPS, indicating that dynamical downscaling can lead to a forecast error reduction in the free atmosphere. The spatial distribution of model errors indicates that the largest differences between the two models are not necessarily associated with relatively large orographic differences.

The impact of dynamical downscaling on clustering results has been investigated in some details. When the same clustering algorithm is applied to both sets of ensembles, the resulting difference in the size and membership of clusters appears to be large. Results indicate that the ALEPS smaller spatial scales affect ensemble properties and may cause different clustering outcomes, even if the ECEPS and the ALEPS clusters include the same members. In approximately one-half of all (parameter–season) combinations considered, dynamical downscaling has caused nonnegligible differences between global and regional clusters even when identical clustering algorithm was applied. This implies that the properties identified via representative members in global clusters may not always be extended to regional clusters. This result has a clear implication for the design of limited-area ensemble prediction systems: it suggests that care must be taken when choosing the global representative members for dynamical downscaling, in particular if they serve as proxies for fields that are highly dependent on small-scale orographic features (like, for example, precipitation discussed in Molteni et al. 2001 and Marsigli et al. 2001).

The precipitation rates in three out of the four synoptic cases representing typical severe weather (or storms) that occasionally occur over the maritime and continental parts of Croatia have been discussed in some detail. Although the results from this limited sample do not detect an unequivocal advantage of the higher-resolution ensemble, the analysis indicates that the improvement of precipitation rate and pattern in downscaled ensembles could be linked to a more accurate simulation of spatial and temporal scales of synoptic events.

Acknowledgments. We are grateful to Jim Hansen from the U.S. Naval Research Laboratory and to the two other anonymous reviewers for their constructive criticism, comments, and suggestions that greatly improved the original manuscript. We are also indebted to Laura Ferranti from ECMWF for help with power spectra calculation, to Josip Juras and Zoran Pasarić

from Geophysical Institute, University of Zagreb, and to Antonio Stanešić from CMHS for useful discussions and collaboration.

REFERENCES

- Anderberg, M. R., 1973: *Cluster Analysis for Applications*. Academic Press, 359 pp.
- Atger, F., 1999: Tubing: An alternative to clustering for the classification of ensemble forecasts. *Wea. Forecasting*, **14**, 741–757.
- Baldwin, M. E., S. Lakshmiarahan, and J. S. Kain, 2001: Verification of mesoscale features in NWP models. Preprints, *Ninth Conf. on Mesoscale Processes*, Fort Lauderdale, FL, Amer. Meteor. Soc., 8.3.
- Buizza, R., 1997: Potential forecast skill of ensemble prediction and spread and skill distributions of the ECMWF ensemble prediction system. *Mon. Wea. Rev.*, **125**, 99–119.
- , and P. Chessa, 2002: Prediction of the U.S. storm of 24–26 January 2000 with the ECMWF Ensemble Prediction System. *Mon. Wea. Rev.*, **130**, 1531–1551.
- , and A. Hollingsworth, 2002: Storm prediction over Europe using the ECMWF Ensemble Prediction System. *Meteor. Appl.*, **9**, 289–305.
- , M. Miller, and T. N. Palmer, 1999: Stochastic representation of model uncertainties in the ECMWF Ensemble Prediction System. *Quart. J. Roy. Meteor. Soc.*, **125**, 2887–2908.
- 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.*, **128**, 1321–1341.
- Giard, D., and E. Bazile, 2000: Implementation of a new assimilation scheme for soil and surface variables in a global NWP model. *Mon. Wea. Rev.*, **128**, 997–1015.
- Houtekamer, P. L., L. Lefèvre, J. Derome, H. Ritchie, and H. L. Mitchell, 1996: A system simulation approach to ensemble prediction. *Mon. Wea. Rev.*, **124**, 1225–1242.
- Ivančan-Picek, B., D. Glasnović, and V. Jurčec, 2003: Analysis and Aladin prediction of heavy precipitation event on the eastern side of the Alps during Map IOP 5. *Meteor. Z.*, **12**, 103–112.
- Ivatek-Šahdan, S., and M. Tudor, 2004: Use of high-resolution dynamical adaptation in operational suite and research impact studies. *Meteor. Z.*, **13**, 99–108.
- , and B. Ivančan-Picek, 2006: Effects of different initial and boundary conditions in Aladin/HR simulations during MAP IOPs. *Meteor. Z.*, **15**, 187–197.
- Jung, T., E. Klinker, and S. Uppala, 2004: Reanalysis and reforecast of three major European storms of the twentieth century using the ECMWF forecasting system. Part I: Analyses and deterministic forecasts. *Meteor. Appl.*, **11**, 343–361.
- , —, and —, 2005: Reanalysis and reforecast of three major European storms of the twentieth century using the ECMWF forecasting system. Part II: Ensemble forecasts. *Meteor. Appl.*, **12**, 111–122.
- Lorenz, E. N., 1982: Atmospheric predictability experiments with a large numerical model. *Tellus*, **34**, 505–513.
- Marsigli, C., A. Montani, F. Nerozzi, T. Paccagnella, S. Tibaldi, F. Molteni, 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.
- Molteni, F., R. Buizza, T. N. Palmer, and T. Petroliaigis, 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.
- Montani, A., C. Marsigli, F. Nerozzi, T. Paccagnella, and R. Buizza, 2001: Performance of the ARPA-SMR limited-area ensemble prediction system: Two flood cases. *Nonlinear Processes Geophys.*, **8**, 387–399.
- , —, —, —, 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.
- Palmer, T. N., J. Barkmeijer, R. Buizza, and T. Petroliaigis, 1997: The ECMWF Ensemble Prediction System. *Meteor. Appl.*, **4**, 301–304.
- Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, 1995: *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, 994 pp.
- Shutts, G. J., 1990: Dynamical aspects of the October 1987 storm: A study of a successful fine-mesh simulation. *Quart. J. Roy. Meteor. Soc.*, **116**, 1315–1347.
- Stensrud, D. J., H. E. Brooks, J. Du, S. Tracton, and E. Rogers, 1999: Using ensembles for short-range forecasting. *Mon. Wea. Rev.*, **127**, 433–446.
- Strelec-Mahović, N., and D. Drvar, 2005: Hailstorm on 04 July 2003—A case study. *Croat. Meteor. J.*, **40**, 381–384.
- Toth, Z., and E. Kalnay, 1993: Ensemble forecasting at NMC: The generation of perturbations. *Bull. Amer. Meteor. Soc.*, **74**, 2317–2330.
- Tripoli, G. J., C. M. Medaglia, S. Dietrich, A. Mugnai, G. Panegrossi, S. Pinori, and E. A. Smith, 2005: The 9–10 November 2001 Algerian flood: A numerical study. *Bull. Amer. Meteor. Soc.*, **86**, 1229–1235.
- Tudor, M., and S. Ivatek-Šahdan, 2002: MAP IOP 15 case study. *Croat. Meteor. J.*, **37**, 1–14.
- Tustison, B., D. Harris, and E. Foufoula-Georgiou, 2001: Scale issues in verification of precipitation forecasts. *J. Geophys. Res.*, **106**, 11 775–11 784.
- Van den Hurk, B. J. J. M., P. Viterbo, A. C. M. Beljaars, and A. K. Betts, 2000: Offline validation of the ERA40 surface scheme. ECMWF Tech. Memo. 295, 42 pp.
- Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences: An Introduction*. Academic Press, 467 pp.