Verification of Precipitation Forecasts from Two Limited-Area Models over Italy and Comparison with ECMWF Forecasts Using a Resampling Technique

Christophe Accadia, Stefano Mariani, and Marco Casaioli

Agenzia per la Protezione dell'Ambiente e per i Servizi Tecnici, Rome, Italy

Alfredo Lavagnini

Istituto di Scienze dell'Atmosfera e del Clima, Consiglio Nazionale delle Ricerche, Rome, Italy

Antonio Speranza

Dipartimento di Matematica e Informatica, Università di Camerino, Camerino, Italy

ABSTRACT

This paper presents the first systematic limited area model (LAM) precipitation verification work over Italy. A resampling technique was used to provide skill score results along with confidence intervals. Two years of data were used, starting in October 2000. Two operational LAMs have been considered, the Limited Area Model Bologna (LAMBO) operating at the Agenzia Regionale Prevenzione e Ambiente-Servizio Meteorologico Regionale (ARPA-SMR) of the Emilia–Romagna region, and the QUADRICS Bologna Limited Area Model (QBOLAM) running at the Agenzia per la Protezione dell'Ambiente e per i Servizi Tecnici (APAT). A 24-h forecast skill score comparison was first performed on the native 0.1° high-resolution grids, using a Barnes scheme to produce the observed 24-h accumulated rainfall analysis. Two nonparametric skill scores were used: the equitable threat score (ETS) and the Hanssen and Kuipers score (HK). Frequency biases (BIA) were also calculated. LAM forecasts were also remapped on a lower-resolution grid (0.5°), using a nearest-neighbor average method; this remapping allowed for comparison with ECMWF model forecasts, and for LAM intercomparisons at lower resolution, with the advantage of reducing the skill score sensitivity to small displacements errors. LAM skill scores depend on the resolution of the verification grid, with an increase when they are verified on a lower-resolution grid. The selected LAMs have a higher BIA compared to ECMWF, showing a tendency to overforecast precipitation, especially along mountain ranges, possibly due to undesired effects from the large-scale and/or convective precipitation parameterizations. Lower ECMWF BIA accounts for skill score differences. LAMBO precipitation forecasts during winter (adjusted for BIA differences) have less misses than ECMWF over the islands of Sardinia and Sicily. Higher-resolution orography definitely adds value to LAM forecasts.

1. Introduction

Weather forecasting, and rainfall prediction in particular, can be a very difficult task in the Mediterranean area. This is due to many factors, such as strong evaporation from the Mediterranean Sea followed by advection and orographic lifting of moist air. Moreover, synoptic system interaction with complex orography may force mesoscale processes like secondary cyclogenesis. Additional complications might be the influence from the subtropical jet, and the paucity of valuable observations on the southern boundary, just to name a few.

This basin is surrounded by largely populated areas, so the impact of a good (or bad) precipitation forecast is vastly amplified, especially for civil defense and flooding management. A typical example is the 13–16 October 2000 Piedmont flood (Gabella and Mantovani 2001), which was well handled by the Italian civil defense, due to a good warning issued few days in advance. It is then of great interest for the national and regional meteorological services to understand how their operational NWP limited area models (LAMs) forecast rainfall in such a complicated area. LAM precipitation forecast verification is then an important part of the forecasting process, since it allows for the identification of the pros and cons of a particular model. Quantitative precipitation forecast (QPF) skill is viewed in some operational centers as an indicator of the general capability of a NWP model to produce a

Corresponding author address: Christophe Accadia, Agenzia per la Protezione dell'Ambiente e per i Servizi Tecnici (APAT), Via Curtatone 3, 00185 Rome, Italy.
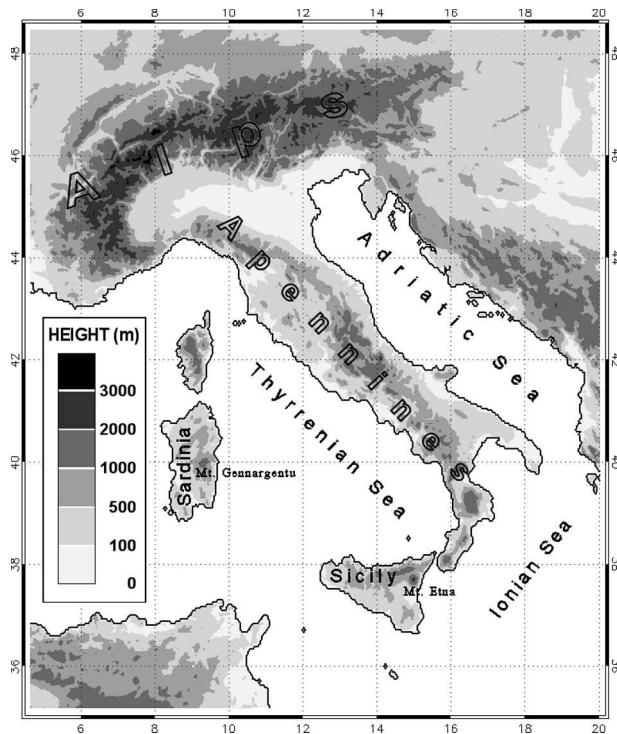E-mail: christophe.accadia@apat.it

Fig. 1. Geomorphological map of Italy and surrounding areas.

Fig. 2. Extension of the 10-km LAM inner domains: solid line, QBOLAM; dashed line, LAMBO. The verification area is the cross hatched zone.

good forecast (Mesinger 1996). Moreover an examination of QPF forecast skill provides valuable feedback to the modeling community, giving insights into how parameterization schemes model the physical processes linked to precipitation.

In the framework of the European Project INTERREG II C (*Gestione del territorio e prevenzione dalle inondazioni* or land management and floods prevention), Accadia et al. (2003a) have carried out a quantitative precipitation forecast assessment over the Piedmont and Liguria regions of Italy for three operational LAMs using categorical statistics to quantify the skill of predicting the occurrence of rain. The considered time range was 8 months. A resampling hypothesis testing technique, proposed by Hamill (1999), was used to check the statistical significance of the categorical skill score differences. In that study the considered LAMs have shown statistically equivalent skill scores over the Piedmont and Liguria regions (northwestern Italy).

Mass et al. (2002) point out that only a few works have presented long-period objective precipitation verification studies. In this paper the aforementioned methodology will be applied to study the 24-h QPF produced by two operational LAMs over the entire territory of Italy including Sicily and Sardinia (see Fig. 1 for geographical reference), for a longer time period from 1 October 2000 to 31 October 2002 (761 days). This long time period ensures the statistical robustness of the results, especially for high values of 24-h accumulated
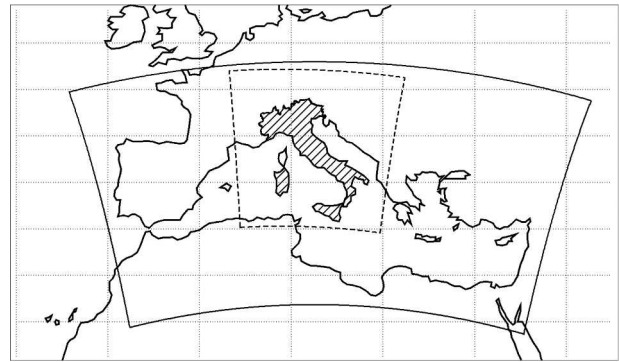
rainfall. Such a study has never been performed systematically over Italy. Further, this statistical assessment may indicate more generally how LAMs predict precipitation in a complex area such as the western Mediterranean.

The QUADRICS Bologna Limited Area Model (QBOLAM) and the Limited Area Model Bologna (LAMBO) were used in this study. QBOLAM is running operationally at the Agenzia per la Protezione dell'Ambiente e per i Servizi Tecnici [APAT, formerly the Department of National Technical Services of the Italian Cabinet Presidency (DSTN-PMC)]. QBOLAM is a parallelized version of the Bologna Limited Area Model (BOLAM; Buzzi et al. 1994), which has shown very good forecast capabilities in the Comparison of Mesoscale Prediction and Research Experiments (COMPARE; Georgelin et al. 2000; Nagata et al. 2001) project.

LAMBO is an operational version of the Eta Model (Lazic and Telenta 1990) running at Agenzia Regionale Prevenzione e Ambiente-Servizio Meteorologico Regionale (ARPA-SMR) of Emilia–Romagna (Paccagnella et al. 1992).

Each LAM was verified on its own original high-resolution grid with grid spacing of about 10 km. Such domains are shown in Fig. 2. To produce an observed rainfall analysis less sensitive to such grid-box size, a two-pass Barnes objective analysis scheme was used (Barnes 1964, 1973; Koch et al. 1983). Another issue explored in this work is the unambiguous comparison of LAM precipitation forecasts with those produced by a global model. With this aim in mind, LAMBO and QBOLAM 24-h accumulated precipitation forecasts were compared with equivalent forecasts produced by the European Centre for Medium-Range Weather Forecasts (ECMWF).

Since these compared models have significantly different grid-box sizes, it was necessary to verify precipitation forecasts on a common grid. QBOLAM and LAMBO precipitation forecasts were remapped onto a regular 0.5°-spaced ECMWF grid using a remapping
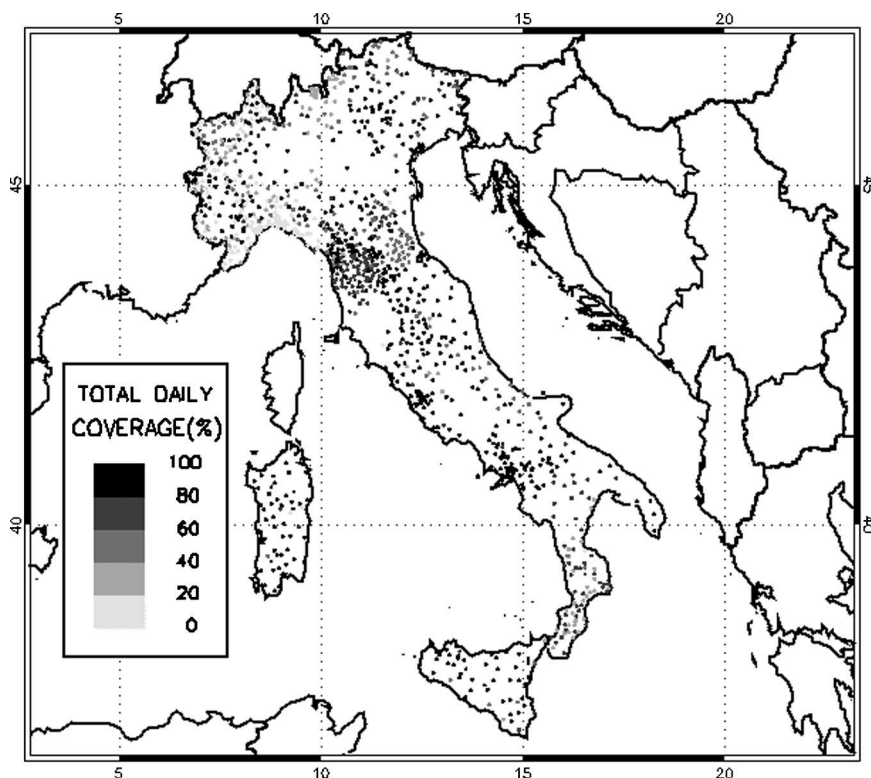
FIG. 3. Geographical distribution of rain gauges used to verify LAMs and ECMWF models. The shading indicates the total daily coverage of the observing stations during the considered 761-day period.

technique that conserves, to a desired degree of accuracy, the total forecast precipitation of the native grid (Baldwin 2000; Accadia et al. 2003b; Mesinger 1996). In this case, rain gauge observations were simply box averaged, but grid boxes with less than 3 rain gauges were neglected, as suggested by Cherubini et al. (2002).

The statistics used to evaluate model precipitation forecasts and to perform the intercomparison are the bias score (BIA; Wilks 1995), the equitable threat score (ETS; Schaefer 1990), and the Hanssen–Kuipers score (HK; Hanssen and Kuipers 1965; McBride and Ebert 2000).

The paper is organized as follows. Section 2a describes the precipitation rain gauge data used, with a very short description of the Barnes objective analysis scheme, while in sections 2b and 2c the QBOLAM and LAMBO models, respectively, are described. Section 2d describes the ECMWF model. Section 3a gives an overview of the nonprobabilistic skill scores used. Section 3b describes the resampling technique. Results of high-resolution verification are presented in section 4a, and low-resolution results are presented in section 4b, whereas section 4c describes the seasonal and regional skill score variations. A discussion of the results is presented in section 5. Conclusions and final remarks are given in section 6.

2. Verification and forecast data

a. Precipitation data

The rain gauge data used in this study covered the land area of Italy, starting from 1 October 2000 to 31 October 2002, a total of 761 days. Rain gauge observations were collected from several datasets, whose spatial coverage is shown in Fig. 3.

The largest part of the datasets was provided by the former Servizio Idrografico e Mareografico Nazionale (SIMN; Italian National Hydrographic and Marigraphic Service) for a total of 923 rain gauges spread over continental and peninsular Italy. These rain gauges were located primarily in the major Italian hydrographic basins; for this reason, some data-void areas, including Sardinia and Sicily, were present. To obtain a more representative precipitation distribution across Italy, rain-gauge data were also collected from different regional services.

A set of 68 rain gauges over the Emilia–Romagna region, pertaining to local networks, has been provided by ARPA-SMR for the same time period. A previously used 8-month (from October 2000 to May 2001) dataset of 389 gauges over the Liguria and Piedmont regions has also been utilized (Accadia et al. 2003a). Despite the limited duration of these datasets, they have been

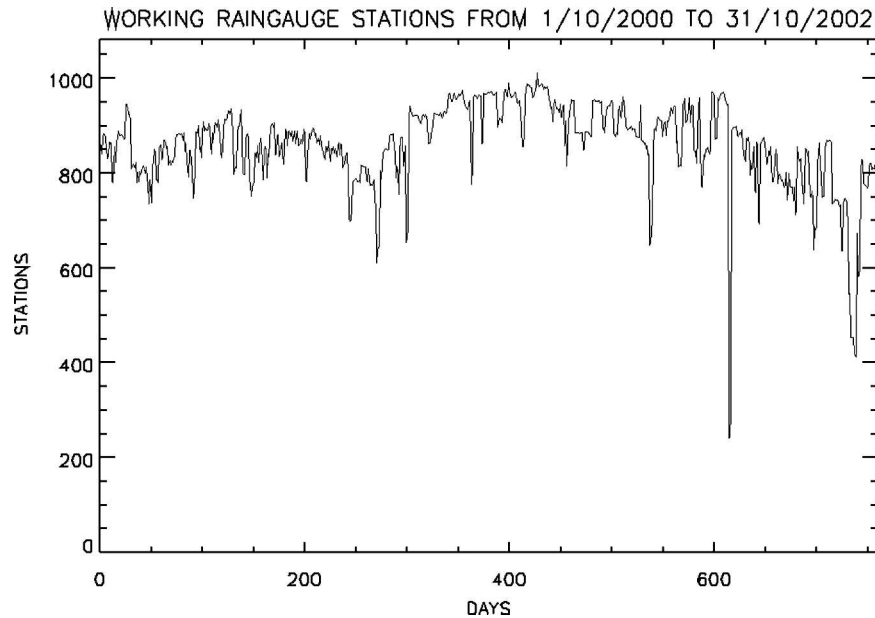WORKING RAINGAUGE STATIONS FROM 1/10/2000 TO 31/10/2002

FIG. 4. Time series of the number of working rain gauge stations during the considered time period.

deemed useful to include, since the SIMN network does not thoroughly cover the Piedmont and Liguria regions, nor does it have stations above 1500 m. Finally, precipitation observations from the regional rain gauge networks of Marche, Sardinia, Sicily, and Valle d'Aosta from October 2000 to October 2002 have been obtained and included in the rainfall analysis.

Observations were accumulated on a daily basis, starting from 0000 UTC, with rain gauges sampling intervals ranging from 5 min to 24 h. A series of quality control procedures were applied. First of all, time series were automatically inspected for internal consistency, in order to spot possible outliers, and corrupted or missing data. A second pass allowed assigning, where possible, a sensible value in place of the corrupted or missing records. For example, rain gauges measuring cumulative precipitation (updating a counter) allowed spotting possible outliers using a simple buddy check in time. If the rain gauge counter was constant before and after the anomalous value, the outlier was corrected assigning the same value as the previous record. On the other hand, if the counter varied, the record was considered to be "no data." Finally, suspicious measurements from available already accumulated 24-h observations were manually verified against nearby stations, taking into account the different meteorological situations using available Meteosat imagery.

Figure 3 shows the percentage of time coverage of each rain gauge station, whereas Fig. 4 shows the daily number of active rain gauges over Italy. As can be seen, not all the stations were active for the whole period, varying from at least 20% to 100% of the time coverage.

A Barnes (1973) Gaussian weighted-averaging scheme was applied to produce a precipitation analyses at 0.1°. This technique assigns a weight to an observation as a function of distance between the observation and grid-box center. The two-pass implementation described by Koch et al. (1983) was applied. A first pass was performed to produce a first-guess precipitation analysis, followed by a second pass, which increases the amount of detail from the prior pass. The convergence parameter was set to 0.3 for both passes, while the average data spacing has been set to 0.2°. This setting is consistent with the constraint that the ratio between grid size and average data spacing lies between 0.3 and 0.5 (Barnes 1964, 1973). Grid points that do not have a rain gauge within a radius of 0.15° were neglected to avoid the excessive rainfall spreading introduced by the analysis scheme on grid points far from the rain gauges' actual locations. For a thorough discussion on the application of the method, the reader is referred to Accadia et al. (2003b).

Observation thresholds used in this study of the nonparametric scores are 0.5, 5.0, 10.0, 20.0, 30.0, and 40.0 mm $(24 \text{ h})^{-1}$.

As mentioned in the introduction, it is important to consider long time periods in order to represent the precipitation statistics reasonably well. This is particularly important in the Mediterranean area, since many (and most) intense rainfall events occur during the fall and early winter, with relatively few events during summer.

Table 1 quantifies the occurrence of rainy days over the considered 2-yr period (in terms of intervals of observed 24-h accumulated precipitation) as a function of

TABLE 1. Number of days for the selected threshold intervals of 24-h observed rainfall, as a function of the percentage coverage of the verification domain (peninsular Italy, Sardinia, and Sicily). Percentage of coverage is given by the ratio between the number of QBOLAM grid points (see text) affected by rainy events (as a function of observation threshold ranges) and the daily total number of grid points affected by all active rain gauges. Observation threshold ranges are in [mm (24 h)$^{-1}$].

Coverage (%)	0.0–0.5	0.5–5.0	5.0–10.0	10.0–20.0	20.0–30.0	30.0–40.0	>40.0
0–1	7	63	230	324	497	622	637
1–5	1	113	229	194	172	110	86
5–10	3	125	139	111	69	24	26
10–25	40	282	160	123	23	5	10
25–50	137	178	3	9	0	0	2
50–75	176	0	0	0	0	0	0
75–100	397	0	0	0	0	0	0

percentage spatial coverage over the verification domain. Percentage spatial coverage is given by the ratio between the number of QBOLAM grid points (0.1° grid) affected by rain events each day (i.e., precipitation within a specified observation threshold range) and the daily total number of grid points affected by all active rain gauges. The lowest threshold interval [0.0–0.05 mm (24 h)$^{-1}$ threshold range] actually gives information on the occurrence of nonrainy days. More than one-half of the selected days are not affected by precipitation events [0.0–0.5 mm (24 h)$^{-1}$ threshold range] over more than 75% of Italy. As expected, as thresholds increase, the spatial percentage coverage progressively decreases.

Verification data for the comparison with ECMWF forecasts on the 0.5° grid were simply box averaged (i.e., the Barnes scheme was not used), neglecting grid boxes affected by less than three rain gauges, in order to ensure a stable (and representative) 24-h rainfall analysis. This procedure is equivalent to the upscaled observation (USO) verification method proposed by Cherubini et al. (2002).

b. Description of model QBOLAM

QBOLAM is a parallel version of the BOLAM model described by Buzzi et al. (1994), developed at the Bologna branch of the Istituto di Scienze dell'Atmosfera e del Clima-Consiglio Nazionale delle Ricerche [ISAC-CNR, formerly the Institute for Physics and Chemistry of the Atmosphere (FISBAT)]. It is a finite-difference, primitive equation, hydrostatic, limited area model running operationally on a QUADRICS APE-100 parallel computer at APAT as an element of the Poseidon sea wave and tidal forecasting system (Speranza et al. 2004), which also includes a wave model (WAM), and the Princeton Ocean Model (POM) nested in a finite-element model of the Venice Lagoon (VL-FEM). Analysis and forecast lateral boundary conditions are provided by ECMWF. A 60-h forecast with 0.3° horizontal grid spacing starts daily at 1200 UTC. The first 12-h forecasts are neglected (spinup time). The low-resolution forecast provides boundary conditions to the high-resolution 48-h forecast run. Only the first 24-h forecasts are considered here. Out-

puts are available every 3 h, and for verification purposes precipitation is accumulated to 24 h. The horizontal inner grid domain covers the entire Mediterranean Sea (see Fig. 2), with 40 levels in the vertical (sigma coordinates). The inner grid horizontal grid box spacing is 0.1° for both latitude and longitude on the original computational Arakawa C grid. Additionally, QBOLAM is used to provide a surface wind forecast to the WAM model within the Poseidon sea wave and tidal forecasting system. Since rainfall prediction was not a primary model task, simplified parameterization schemes were used to avoid parallelization problems. Boundary layer fluxes are represented by analytic formulas (Louis et al. 1982). A simplified radiation scheme (Page 1986; Ruti et al. 1997) is used. The convection parameterization is based on the Kuo (1974) scheme. The model configuration was unaltered during the considered time period.

c. Description of the LAMBO model

LAMBO is a hydrostatic primitive equation model running operationally at ARPA-SMR, derived from the Eta Model (Mesinger et al. 1988; Janjić 1994; Black 1994), and developed by the University of Belgrade and National Centers for Environmental Prediction (NCEP). Prognostic variables are staggered on an Arakawa E grid. The radiation parameterization of Ritter and Geleyn (1992) is used; boundary layer fluxes are computed using a second-order closure scheme. Convective processes are parameterized with the Betts–Miller scheme (Betts 1986; Betts and Miller 1986). LAMBO runs on two one-way nested grids, whose horizontal grid spacings are about 20 and 10 km, respectively. Initial and boundary conditions on the outer domain are provided by ECMWF analysis and forecasts. Precipitation forecast outputs were accumulated for 6 h. As for QBOLAM, only the first 24 h of a forecast are considered here, starting from 0000 UTC. The LAMBO configuration was unchanged from 2000 to 2002.

d. Description of the ECMWF model

The ECMWF model is a global spectral model, described in detail by Simmons et al. (1989). Stratiform

and convective precipitation are parameterized using the Tiedke (1993) prognostic cloud scheme. However in October 1999 a new cloud parameterization scheme was implemented (Jakob and Klein 2000), taking into account the vertical distribution of cloud layers. Teixeira (1999) thoroughly described the ECMWF model's parameterizations. The model was run with 60 levels starting from October 1999, using a 12-h four-dimensional variational data assimilation scheme (4DVAR; Courtier et al. 1994; Rabier et al. 1998, 2000) and spectral horizontal resolution of T_L 319 until 21 November 2000, when the deterministic mode was upgraded to T_L 511.

The forecasts starting at 1200 UTC were considered, and the verification focused on precipitation accumulated between 12- and 36-h lead time. ECMWF precipitation forecasts were retrieved from its Meteorological Archival and Retrieval System (MARS), with a horizontal grid spacing of 0.5° in latitude and longitude.

3. Statistical methodology

a. Nonprobabilistic scores

Nonprobabilistic statistics were used to quantify the skill of the model precipitation forecasts. These verification measures are based on a categorical dichotomous statement (i.e., a yes–no statement). It is then possible, with a given set of matched rain forecasts and observations, to build a 2×2 contingency table. An event is identified when a forecast or the observed precipitation is below or above a threshold. The combination of different possibilities between observations and forecast defines the contingency table. For each precipitation threshold four categories of hits, false alarms, misses, and correct no-rain forecasts (a, b, c, and d as shown in Table 2) are defined. The scores used in this study are the above-mentioned ETS and HK scores (see Schaefer 1990; Hanssen and Kuipers 1965) and the BIA score (Wilks 1995).

The BIA is the ratio between the predicted rain frequency and the observed rain frequency. It is defined by

$$\text{BIA} = \frac{a + b}{a + c}. \tag{1}$$

This score measures the relative frequency of precipitation forecasts and observations; hence, it is a measure of the model's tendency to overforecast (BIA > 1.0) or underforecast (BIA < 1.0) rain occurrence. It is not a measure of forecast accuracy (for a review, see Wilks 1995 or Ebert et al. 2003b).

The ETS score, proposed by Schaefer (1990), is calculated as

$$\text{ETS} = \frac{a - a_r}{a + b + c - a_r}, \tag{2}$$

TABLE 2. Contingency table of possible events for a selected threshold.

		Rain observed	
		Yes	No
Rain forecast	Yes	a	b
	No	c	d

where a_r is a correction factor of "model hits expected from a random forecast":

$$a_r = \frac{(a + b)(a + c)}{a + b + c + d}. \tag{3}$$

The ETS ranges from –1/3 to 1. An ETS equal to one indicates a perfect forecast, while an ETS close to zero or negative indicates poor rain forecasting skill. It is now used for rainfall verification in many operational centers (Ebert et al. 2003a; Mesinger et al. 1997; Mesinger 1996).

The HK score is a measure of the accuracy both for events and nonevents (McBride and Ebert 2000):

$$\text{HK} = \frac{(ad - bc)}{(a + c)(b + d)}. \tag{4}$$

This score is independent of the event and nonevent marginal distributions; that is, HK does not depend on the number of rain and nonrain events existing in the sample set, as other scores do. The HK method rewards more a correct "yes" forecast if the event is less likely (Wilks 1995). A perfect forecast has an HK score equal to 1.0, while a score of –1.0 means that the precipitation forecast capability is inferior to a random forecast. A random or constant forecast receives a score of 0.0.

b. The resampling method

When performing model comparison, a measure of uncertainty on score differences should be given. Performing a hypothesis test provides a confidence interval for the score difference that exists between two competing models. The hypothesis testing method applied, which was originally proposed by Hamill (1999), is based on a resampling technique called the bootstrap method (Diaconis and Efron 1983). This is a computer-based method that builds a PDF consistent with the selected null hypothesis. A short description of the methodology is presented here; for more details the reader is referred to Hamill (1999) and Accadia et al. (2003b).

The null hypotheses used in this study are that the differences in skill score (either ETS or HK) and BIA are zero between the two competing model forecasts M1 and M2; for

H_0,

$$\text{ETS}_{M1} - \text{ETS}_{M2} = 0.0.$$

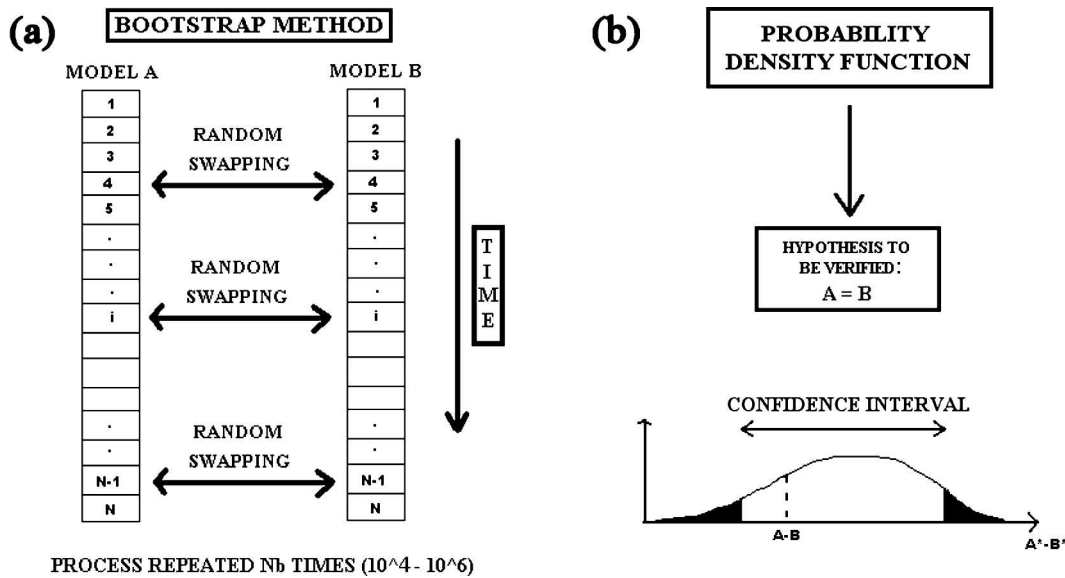$$\text{BIA}_{M1} - \text{BIA}_{M2} = 0.0. \tag{5}$$

FIG. 5. Schematic description of the bootstrap technique. (a) The random swapping of paired data (daily contingency tables) is necessary to build resampled data. (b) Given a confidence level, a confidence interval is identified from the resampled score difference distribution.

The alternative hypotheses are

H_A:

$$ETS_{M1} - ETS_{M2} \neq 0.0.$$

$$BIA_{M1} - BIA_{M2} \neq 0.0. \tag{6}$$

The test is also applied to HK, which is used in place of ETS. The scores are computed for each threshold from a sum of n daily contingency tables obtained from 24-h accumulated observations and forecast comparisons.

Random sampling of daily contingency tables is performed from the available paired data (Fig. 5a), followed by a significance assessment of the test based on a comparison of the observed statistics with the numerically built statistics (Fig. 5b). The general method does not need any assumptions concerning the probability density function. The random sampling is performed 10 000 times for each comparison at each threshold. A 95% significant level is assumed for all tests performed in this study. The null hypothesis H_0 is tested by a two-tailed test using the percentile method (Wilks 1995).

Mesinger (1996) has noticed that skill scores like the ETS can be affected by model BIA. In fact, a random forecast has little skill for intense precipitation; hence, at higher thresholds, a model with a greater BIA should normally exhibit a greater ETS. To perform a fair comparison, Hamill (1999) suggested a BIA adjustment procedure. This procedure allows for the assessment of the effect of BIA score differences on the skill scores.

In fact, skill scores like ETS reward hits more than they penalize false alarms, so a model with BIA greater than one looks artificially skillful when compared with a model with a lower BIA.

Since the analysis is done by comparing two models at a time, one model is defined as the reference (the one that has the BIA closer to 1 for almost all thresholds), while the other is dubbed the competitor. The BIA adjustment procedure is performed for each threshold independently. It consists of changing the forecast threshold when calculating the competitor BIA score in a way that allows for the possibility that the two BIA scores could coincide. This procedure will be applied in this work. For a detailed discussion about the effects of the BIA adjustment procedure on the skill scores, see Accadia et al. (2003b).

The application of any hypothesis test is rather incomplete without a discussion of the correlation of time series and possible space correlations in the dataset. To avoid spatial correlations contaminating the hypothesis test results, the contingency table elements were tallied over the whole set of grid points available in Italy each day (see Hamill 1999). Moreover Table 3 (for high-resolution comparison) and Table 4 (for low-resolution comparison) show that the score time series were negligibly autocorrelated when observations and forecasts are accumulated for 24 h. For Tables 3 and 4, to test the significance of the Pearson lag-1 autocorrelation (r), Fisher's r to Z transformation (Fisher 1925) was applied. This transformation changed the correlation variable r into a normally distributed variable Z; in this way it was possible to assign a correlation confidence interval by means of the standard error.

TABLE 3. Pearson lag-1 autocorrelation of daily scores over Italy, which have been calculated using precipitation forecasts from high-resolution LAMs (LAMBO and QBOLAM) for the entire time interval. For each correlation coefficient, a Fisher's confidence value is also shown.

Threshold [mm (24 h)$^{-1}$]	Model	BIA		ETS		HK	
		Lag-1 autocorrelation	Fisher's confidence value	Lag-1 autocorrelation	Fisher's confidence value	Lag-1 autocorrelation	Fisher's confidence value
0.5	LAMBO	0.2751	±0.0592	0.3689	±0.0553	0.2584	±0.0597
	QBOLAM	0.2821	±0.0556	0.2049	±0.0579	0.1219	±0.0595
5.0	LAMBO	0.0950	±0.0668	0.3496	±0.0597	0.2632	±0.0628
	QBOLAM	0.1835	±0.0613	0.2950	±0.0580	0.1763	±0.0615
10.0	LAMBO	0.0920	±0.0704	0.3497	±0.0632	0.2328	±0.0672
	QBOLAM	0.1918	±0.0646	0.2788	±0.0624	0.1768	±0.0650
20.0	LAMBO	0.0161	±0.0805	0.3483	±0.0725	0.2669	±0.0749
	QBOLAM	0.1571	±0.0742	0.2512	±0.0720	0.1232	±0.0749
30.0	LAMBO	0.1659	±0.0884	0.2937	±0.0863	0.2598	±0.0848
	QBOLAM	0.1933	±0.0835	0.2083	±0.0843	0.0855	±0.0861
40.0	LAMBO	0.0676	±0.1047	0.2973	±0.1019	0.2689	±0.0976
	QBOLAM	0.1917	±0.0969	0.1074	±0.1015	0.0874	±0.0998

4. Verification results

One of the declared aims of this work is to see how LAMs score when verified on their own (~0.1°) high-resolution grids. It is clear that a traditional verification method (i.e., grid-box based) is very demanding at high resolutions. Small displacement errors in weather forecasts are likely, and these may produce the so-called double penalty effect on the scores (e.g., a one grid-box shift results in both a miss and a false alarm). The verification exercise on the 0.5° grid in latitude and longitude allows comparing skill score results with the high-resolution one; furthermore, it permits an assessment of how these two LAMs perform in comparison with the

TABLE 4. As in Table 3 except that daily scores have been calculated using precipitation forecast from low-resolution remapped LAMs (LAMBO and QBOLAM) and the ECMWF global model.

| Threshold [mm (24 h)$^{-1}$]	Model	BIA		ETS		HK	
		Lag-1 autocorrelation	Fisher's confidence value	Lag-1 autocorrelation	Fisher's confidence value	Lag-1 autocorrelation	Fisher's confidence value
0.5	LAMBO	0.1766	±0.0638	0.3360	±0.0585	0.2675	±0.0611
	QBOLAM	0.2900	±0.0554	0.2602	±0.0565	0.1759	±0.0587
	ECMWF	0.1624	±0.0589	0.2692	±0.0570	0.2242	±0.0574
5.0	LAMBO	0.0773	±0.0702	0.2917	±0.0662	0.2539	±0.0661
	QBOLAM	0.1319	±0.0640	0.2973	±0.0603	0.1754	±0.0631
	ECMWF	−0.0029	±0.0650	0.2196	±0.0672	0.1909	±0.0627
10.0	LAMBO	0.1108	±0.0767	0.2716	±0.0761	0.2356	±0.0734
	QBOLAM	0.2817	±0.0655	0.2858	±0.0668	0.1918	±0.0685
	ECMWF	0.0233	±0.0710	0.2557	±0.0765	0.2079	±0.0680
20.0	LAMBO	0.1766	±0.0871	0.2574	±0.0950	0.2140	±0.0858
	QBOLAM	0.2786	±0.0766	0.3053	±0.0785	0.1765	±0.0804
	ECMWF	0.0570	±0.0826	0.2311	±0.1019	0.2247	±0.0787
30.0	LAMBO	0.1128	±0.1069	0.2325	±0.1221	0.2093	±0.1036
	QBOLAM	0.1619	±0.0988	0.2323	±0.1043	0.1183	±0.1001
	ECMWF	0.1431	±0.0992	0.1797	±0.1387	0.2213	±0.0964
40.0	LAMBO	0.0401	±0.1264	0.2268	±0.1458	0.1763	±0.1227
	QBOLAM	0.2679	±0.1109	0.2192	±0.1276	0.1742	±0.1158
	ECMWF	0.1639	±0.1159	0.2929	±0.1768	0.2709	±0.1105
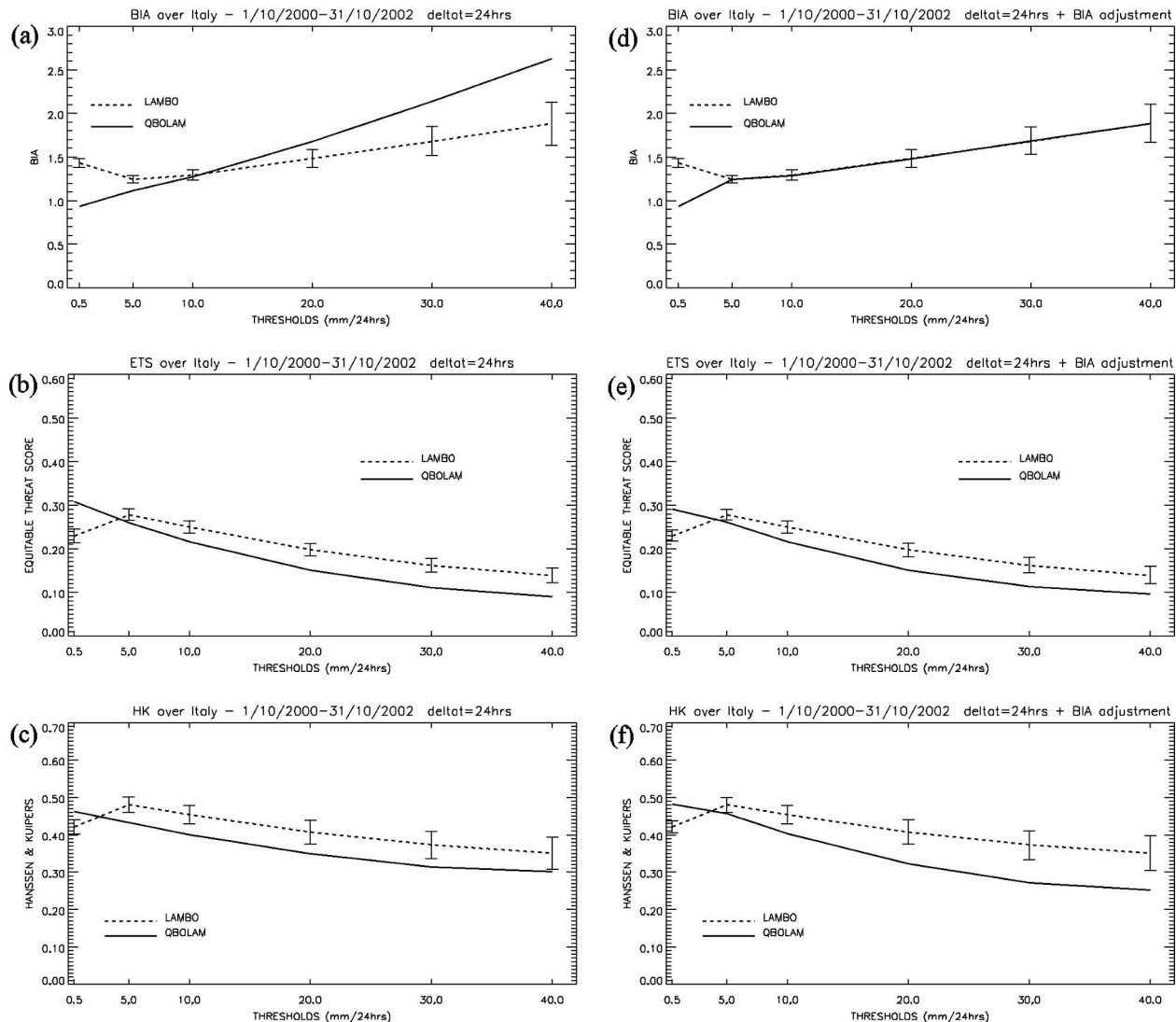
FIG. 6. The 24-h accumulation scores calculated for a set of 761 days over Italy for QBOLAM and LAMBO on their original high-resolution grids. (a) BIA, (b) ETS, and (c) HK are calculated without BIA adjustment; the same scores are shown in (d), (e), and (f), respectively, after BIA adjustment. The confidence intervals are referenced to LAMBO scores, indicating the 2.5th and the 97.5th percentiles of the resampled distributions.

ECMWF global model. All results are presented with and without the BIA adjustment.

a. High-resolution verification

The LAMBO and QBOLAM models show a BIA score that is above 1.0 for almost all thresholds, with a single exception at the lowest threshold for QBOLAM (Fig. 6a). The BIA scores increase with threshold value, with QBOLAM having a higher BIA for thresholds above 10.0 mm $(24 \text{ h})^{-1}$. This result indicates that both models have an overforecasting tendency for medium to high precipitation thresholds, with an increase in false alarms. This tendency is accentuated for QBOLAM.

ETS scores ranged from 0.3 to 0.1 at higher thresholds (Fig. 6b). Bootstrap results show that the differences existing between LAMBO and QBOLAM are statistically significant, with QBOLAM scoring slightly less than LAMBO for thresholds above 0.5 mm $(24 \text{ h})^{-1}$. This ETS difference is accentuated after performing the same comparison with the BIA adjustment (Fig. 6e). The HK scores give similar results (Figs. 6c,f). There is actually a slight decrease in the relative QBOLAM HK after the BIA adjustment for thresholds above 20.0 mm $(24 \text{ h})^{-1}$. This happens because both hits and false alarms are actually reduced by the application of the procedure, since the QBOLAM BIA is shifted toward LAMBO BIA. QBOLAM captures well the 0.5 mm $(24 \text{ h})^{-1}$ rain–no-rain situations, but has
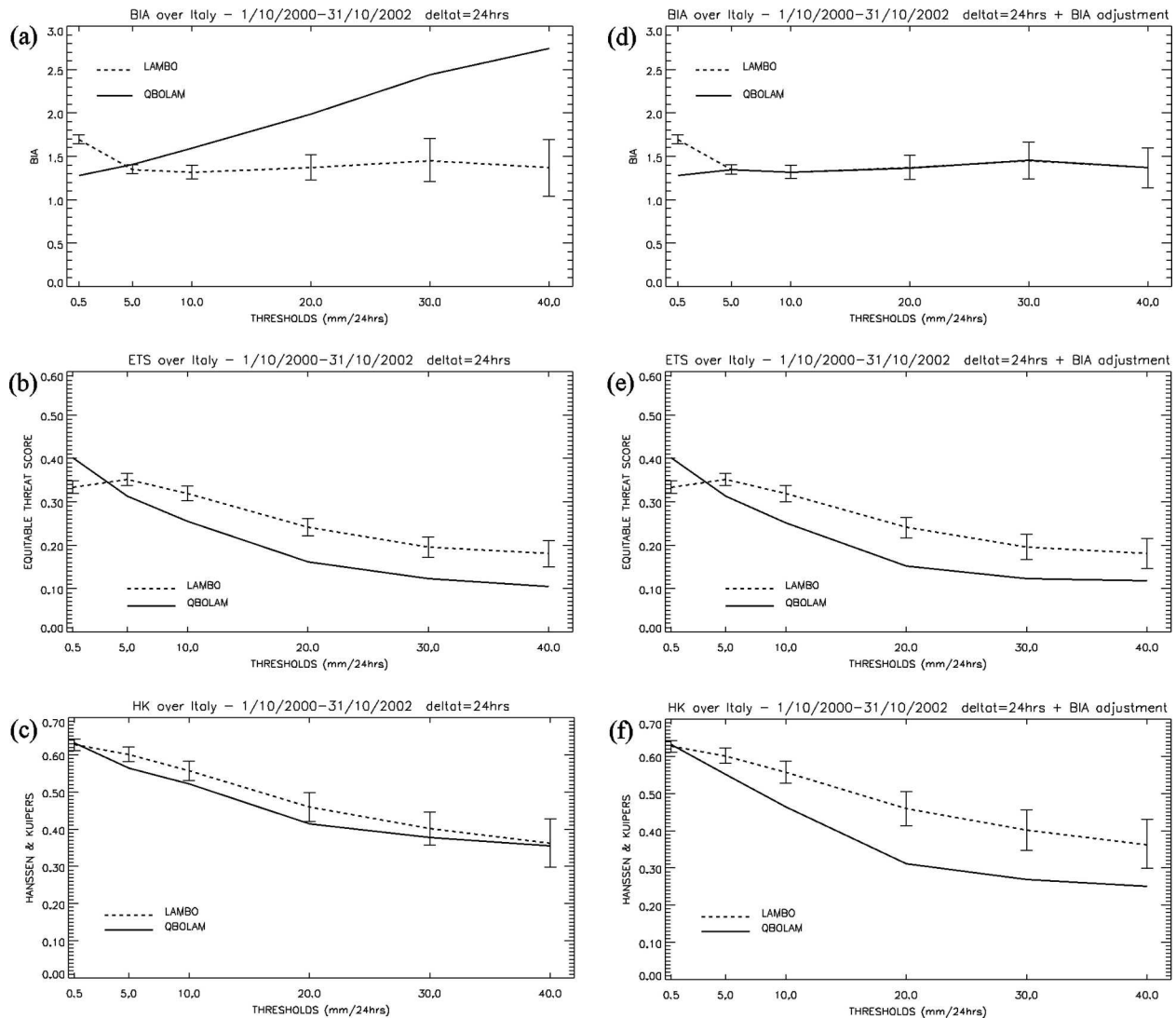
FIG. 7. As in Fig. 6, but for the 24-h accumulation scores calculated for a set of 761 days over Italy for LAMBO and QBOLAM remapped on the 0.5° grid.

lower scores compared to LAMBO for all the other thresholds.

b. Low-resolution verification

The same verification exercise was performed on the low-resolution grid after remapping precipitation forecasts (Fig. 7).

First, note that the BIA scores change when compared with the high-resolution results (Fig. 7a). The remapping produces a general increase of QBOLAM BIA at all thresholds, while LAMBO BIA is stabilized around a value of 1.4, with an actual BIA decrease for thresholds above 20.0 mm $(24 \text{ h})^{-1}$. The LAMBO BIA increases for the two lowest thresholds, an effect similar to the QBOLAM change. BIA score differences between LAMBO and QBOLAM are statistically signifi-

cant, with the possible exception of the score at 5.0 mm $(24 \text{ h})^{-1}$. The ETS scores (Fig. 7b) are slightly higher than those at high resolution, with LAMBO markedly improving, while QBOLAM has a less noticeable improvement for thresholds above 0.5 mm $(24 \text{ h})^{-1}$. These findings are consistent with results presented by Gallus (2002). Also the HK comparison (Fig. 7c) shows the same qualitative behavior, although for the lowest and the two highest thresholds the results are statistically similar. The application of BIA adjustment (Figs. 7e,f) shows that LAMBO is unambiguously more skillful than QBOLAM for thresholds above 0.5 mm $(24 \text{ h})^{-1}$.

As mentioned before, the upscaling method applied to LAM precipitation forecasts allows the intercomparison with 24-h accumulated ECMWF forecasts. The results are shown in Figs. 8 and 9 for LAMBO and QBOLAM, respectively.
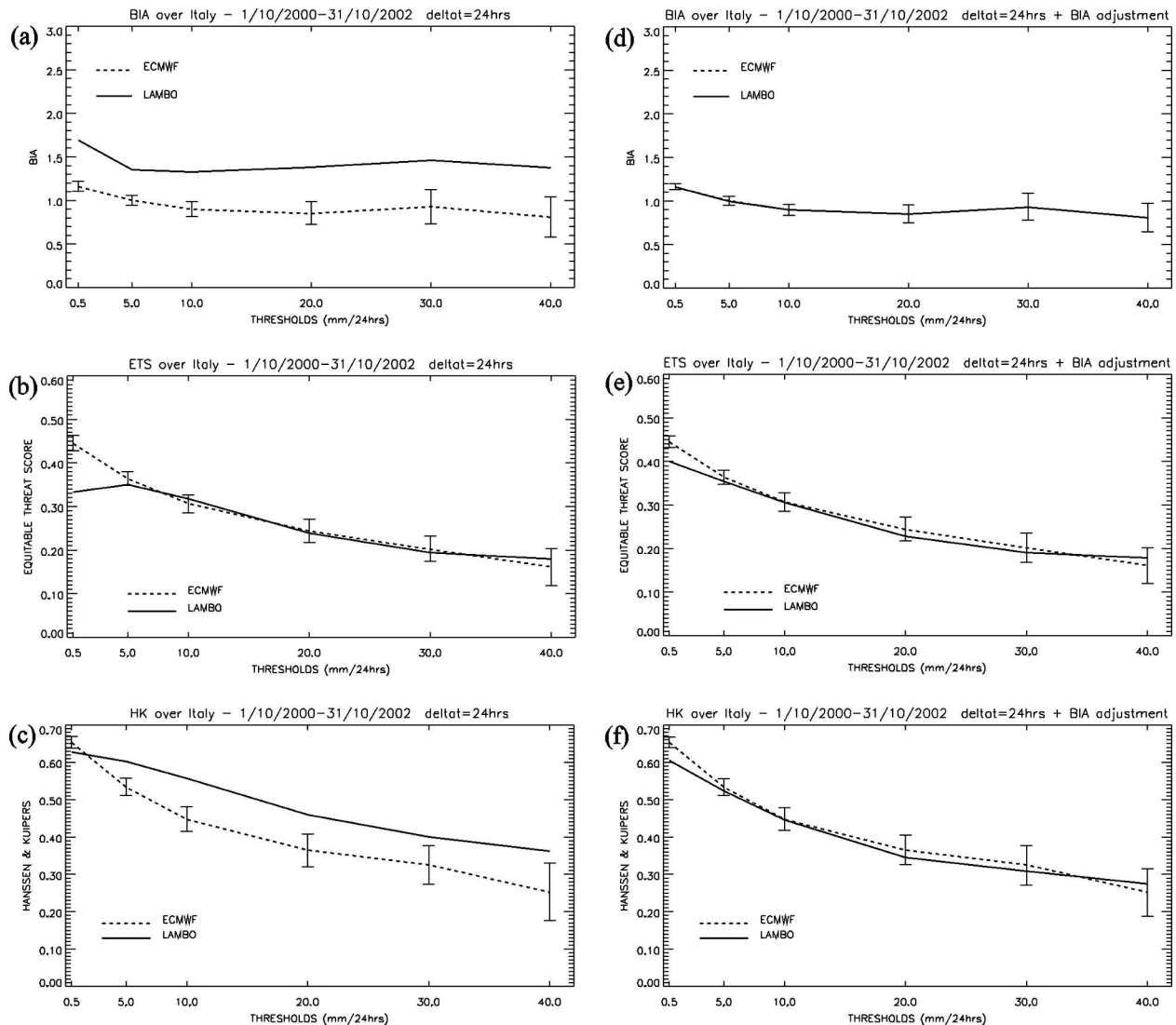
FIG. 8. As in Fig. 6, but for the 24-h accumulation scores calculated for a set of 761 days over Italy for LAMBO remapped on the 0.5° grid and ECMWF model.

The ECMWF model has a BIA score around 1 for almost all thresholds, while LAMBO BIAs are systematically higher, as described before (Fig. 8a). LAMBO ETS scores are statistically similar to those of ECMWF, apart from the lowest threshold (Fig. 8b). The outcomes of the HK comparison between the LAMBO and ECMWF precipitation forecasts are particularly interesting (Fig. 8c). LAMBO shows HK scores sensibly higher than those of ECMWF for all thresholds above 5.0 mm $(24 \text{ h})^{-1}$. The same comparison after the BIA adjustment interestingly suggests that this difference can be explained by the higher BIA of LAMBO. In fact the results can be interpreted, remembering that HK can be written as a sum of the accuracy for events (probability of detection, POD $= a/(a + c)$) and the accuracy of nonevents (probability of a null event, PON $= d/(b + d)$ (McBride and Ebert 2000; Wilks 1995), as

$$ \text{HK} = \frac{a}{a + c} + \frac{d}{b + d} - 1, \tag{7} $$

with the last term simply normalizing between -1 and 1. Then, as said before, if the competitor BIA is reduced toward the reference BIA, the hits and false alarms are also reduced. The false alarm reduction increases the nonevents accuracy, while the reduction of hits decreases the rain event accuracy. Furthermore, the number of correct no-rain forecasts outnumbers the other members of the contingency table, so the nonevent accuracy is only slightly increased (Accadia et al. 2003b). The ETS scores, except for the lower threshold (Figs. 8b,e), are statistically similar, and almost insensitive to the BIA adjustment, indicating that the reduction of hits and false alarms is about the same.

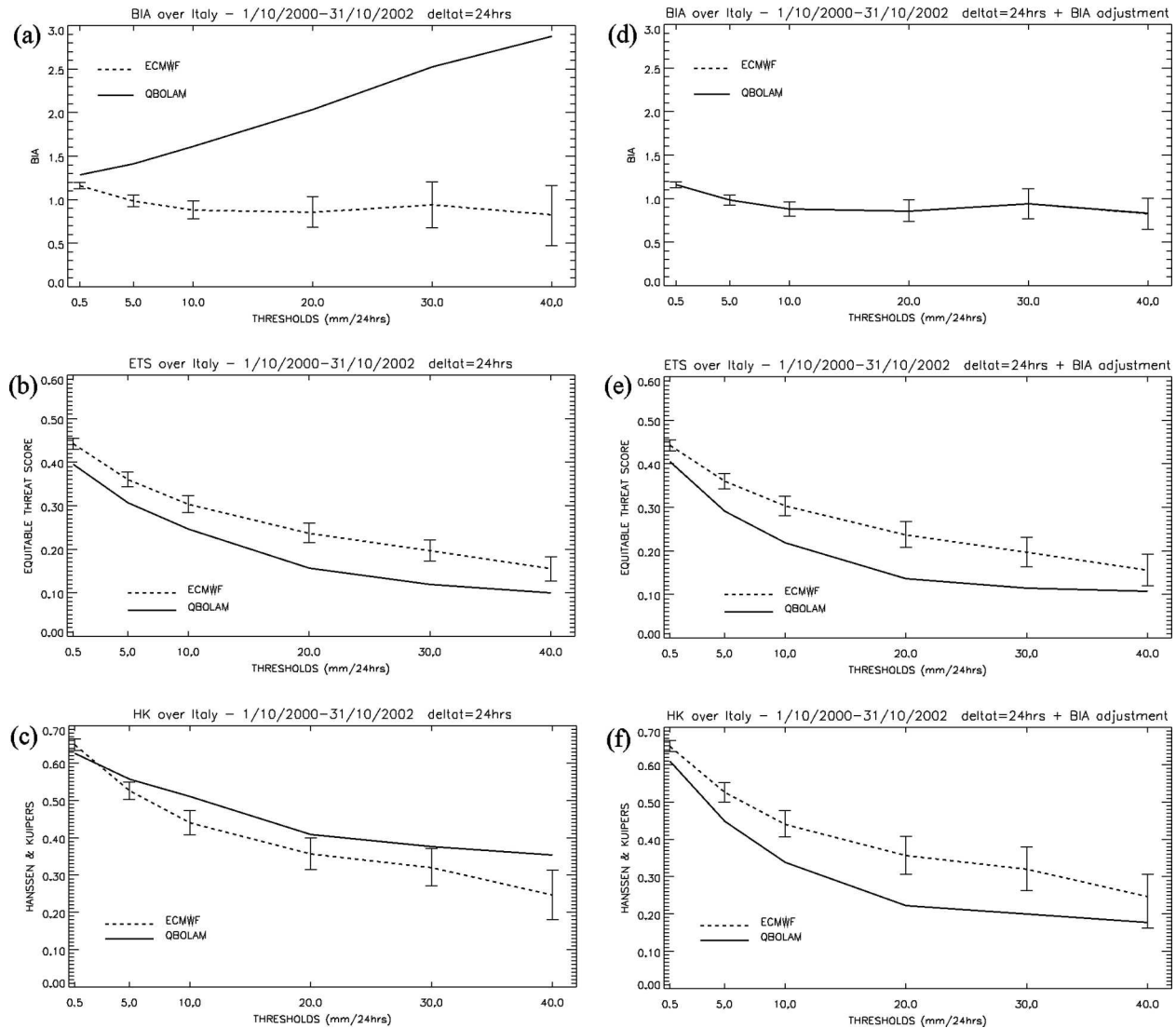Synthesizing this, LAMBO produces more false

FIG. 9. As in Fig. 6, but for the 24-h accumulation scores calculated for a set of 761 days over Italy for QBOLAM remapped on the 0.5° grid and ECMWF model.

alarms, but has a good hit rate, and so it has about the same ETS scores as does the ECMWF model; meanwhile LAMBO is also rewarded by the HK score, since it is less sensitive to false alarms. If calibrated for BIA differences, the forecasts are not significantly different in skill.

The same qualitative behavior is shown when QBOLAM is compared with the ECMWF model (Fig. 9), but in this case the results indicate, in a statistically significant way, that this model does not have the same rainfall forecasting capabilities as the ECMWF model.

c. Seasonal and regional skill variation

A long verification period produces results that are statistically robust, but additional information on the temporal variation of model performance is desirable. To study the seasonal behavior of the models, the scores have been calculated for each month for the 5.0 mm $(24 \text{ h})^{-1}$ threshold on the 0.5° verification grid. The comparison is again presented with confidence intervals obtained with the bootstrap technique. The 5.0 mm $(24 \text{ h})^{-1}$ threshold has been selected because it is representative enough to distinguish between precipitation events and nonevents and still has robust statistics on a monthly basis.

First, LAMBO and QBOLAM are compared and results are presented in Fig. 10. BIA scores (Fig. 10a) show an increase from May to September 2001, and in March and June 2002. QBOLAM is reaching higher values, especially during summer months, with some statistically significant differences. QBOLAM ETS and
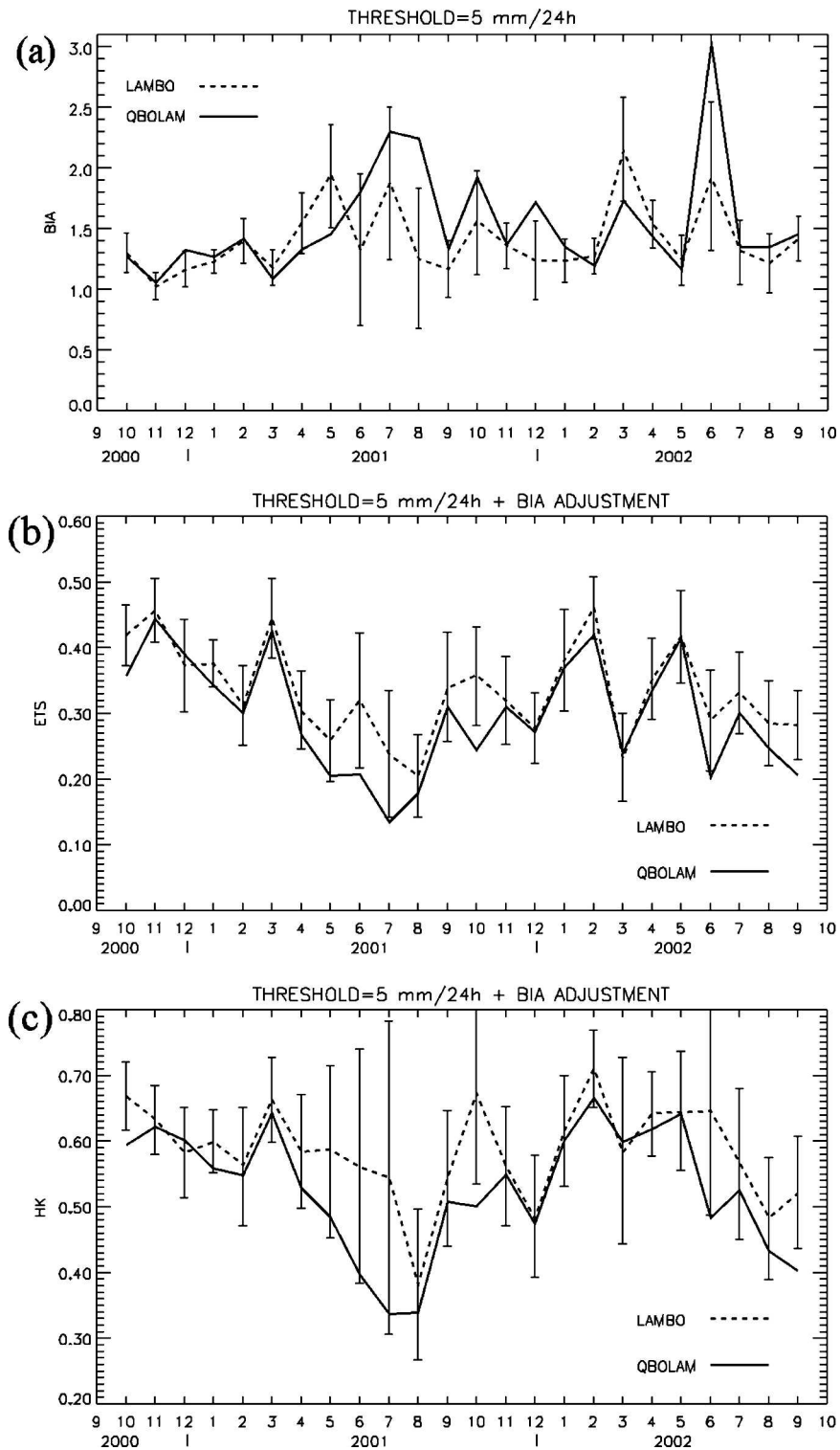
FIG. 10. The 24-h accumulation (a) BIA, (b) ETS, and (c) HK scores for the 5 mm $(24 \text{ h})^{-1}$ threshold computed on a monthly basis over Italy, over 2 yr, for QBOLAM and LAMBO remapped on the 0.5° grid. The confidence intervals indicate the 2.5th and the 97.5th percentiles of the resampled distributions. QBOLAM skill scores are computed after BIA adjustment against LAMBO.

HK skill scores are computed after applying the BIA adjustment. The BIA comparison, together with the observed ETS and HK drops (Figs. 10b,c), are consistent with the assumption that models have difficulties in forecasting convective events, more likely during spring and summer. The HK value of around 0.5 and an ETS around 0.3 indicate that both models captured the rainfall events above 5.0 mm $(24 \text{ h})^{-1}$ with fair skill. QBOLAM BIA sometimes goes above 2.0 during the summer months, indicating that it is overforecasting precipitation. The shown BIA scores computed at high and low resolutions indicate that the crude implementation of the Kuo (1974) convection scheme is not able to produce reliable precipitation fields. Actually, during the daily forecast check activities, "gridpoint storms" were observed in some cases, especially over North Africa during the warm season.

The monthly comparison of the LAMBO and ECMWF scores, computed for the same threshold, is presented in Fig. 11. The HK and ETS comparisons were computed after the LAMBO BIA was adjusted against the ECMWF BIA. This is actually needed to interpret the skill score comparisons, since LAMBO shows a BIA systematically higher than ECMWF (with some statistically significant differences). ECMWF shows two small peaks: one during July 2001 and the other during June 2002. The ETS and HK scores are statistically similar, with the only exception being August 2001. The HK comparison without BIA adjustment indicates that LAMBO's HK score is often higher than the corresponding ECMWF score, although many differences were not statistically significant (not shown). This indicates that BIA differences account for HK differences.

A similar monthly comparison (again, applying the BIA adjustment) between QBOLAM and ECMWF forecasts (Fig. 12) shows that QBOLAM has a significantly lower ETS during the summer months (Fig. 12b). This can be interpreted by looking to the corresponding BIA scores (Fig. 12a). The high number of false alarms produced by QBOLAM reflects on the low summer ETS scores.

Are there differences in model skill that depend on location? A possible way to answer this question is to analyze for a selected threshold how hits, false alarms, and misses are located on the low-resolution grid, for a particular time range. This choice also permits us to see whether there are changes due to a particular season. Correct no-rain forecasts are not shown, because they are distributed almost uniformly over the territory.

For brevity, only two seasonal time ranges are shown: one from 1 October 2000 to 31 March 2001, and the other from 1 April to 30 September 2001. The 10.0 mm $(24 \text{ h})^{-1}$ threshold has been selected to focus on medium to high precipitation events.

Hits, false alarms, and misses for LAMBO and QBOLAM during the first "winter" period are shown in Fig. 13, while the same elements of the contingency table for the same models during the selected "summer" period are shown in Fig. 14. ECMWF results for the two time ranges are shown in Fig. 15. Finally, LAM results are presented in Figs. 16 and 17 selecting the actual forecast thresholds (shown in Table 5) that adjust LAM BIA to ECMWF BIA for the selected time period.

Let us focus on the cold season first. The LAMBO, QBOLAM, and ECMWF models show a large number of hits over the Alps, and in the northern part of the Apennines, but there are some interesting differences. Hits from the LAMs (Fig. 13, without BIA adjustment) show a tendency to lie on the western coast of the Italian peninsula, along the Apennines, while ECMWF does not markedly show this behavior. Moreover ECMWF has very few hits over Sicily and Sardinia. LAMs hits, though, are mainly explained by the larger BIA (as evidenced by their higher false alarm rates in Figs. 13b and 13e), since after the BIA adjustment many hits over the western coast disappear (Fig. 16).

All models produce a number of false alarms along the Alps and over the central Apennines, with QBOLAM also producing the most over the islands. This is still true for both LAMs after the BIA adjustment (Figs. 16b,e). The ECMWF model has many misses over Sardinia and Sicily, as well as over the northern Apennines. LAMBO has fewer misses than QBOLAM over the same area. Apparently, as can be seen from a comparison of Figs. 16c and 16f (after the application of the BIA adjustment procedure) with Figs. 15c and 15f, LAMs have fewer misses than ECMWF over the two main Italian islands during the "winter" period.

The most striking feature for the warm-season behavior of the considered models is the high number of false alarms produced by QBOLAM over the Alps and over the central Apennines, especially over higher peaks. This result is confirmed after the BIA adjustment (Fig. 17f). Hits tend to be concentrated along the Apennines and the Alps for all models. Actually, few hits are scored over major islands by LAMs, while ECMWF scores almost none. The price to pay is that both LAMs forecast some false alarms. In fact, an inspection of Fig. 17 (LAMs after BIA adjustment against ECMWF) shows that LAMBO still has comparable hits, slightly more false alarms, and fewer misses than ECMWF. On the other hand, QBOLAM has reduced skill over peninsular Italy. In general, the quality of all of the considered models is quite low over Sardinia and Sicily during the warm season.

5. Discussion

The comparison of the low-resolution verification results with the corresponding high-resolution ones for LAMBO and QBOLAM gives not only information about how LAM forecasts verify on different grids, but also on the forecast precipitation structure. Verification of upscaled LAM precipitation forecasts may depend
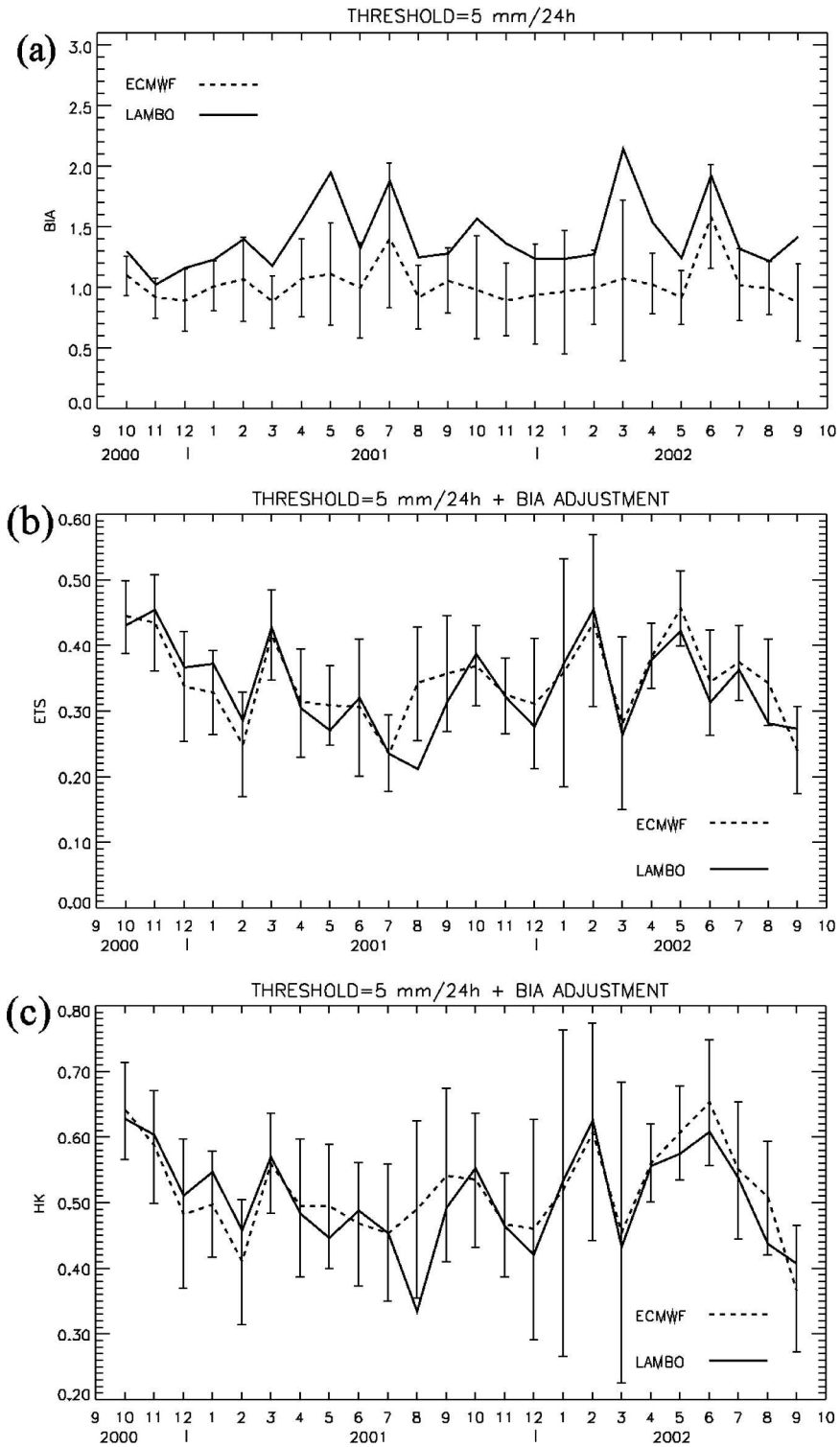
FIG. 11. As in Fig. 10, but for ECMWF and LAMBO remapped on the 0.5° grid. The confidence intervals indicate the 2.5th and the 97.5th percentiles of the resampled distributions. LAMBO skill scores are computed after BIA adjustment against ECMWF.
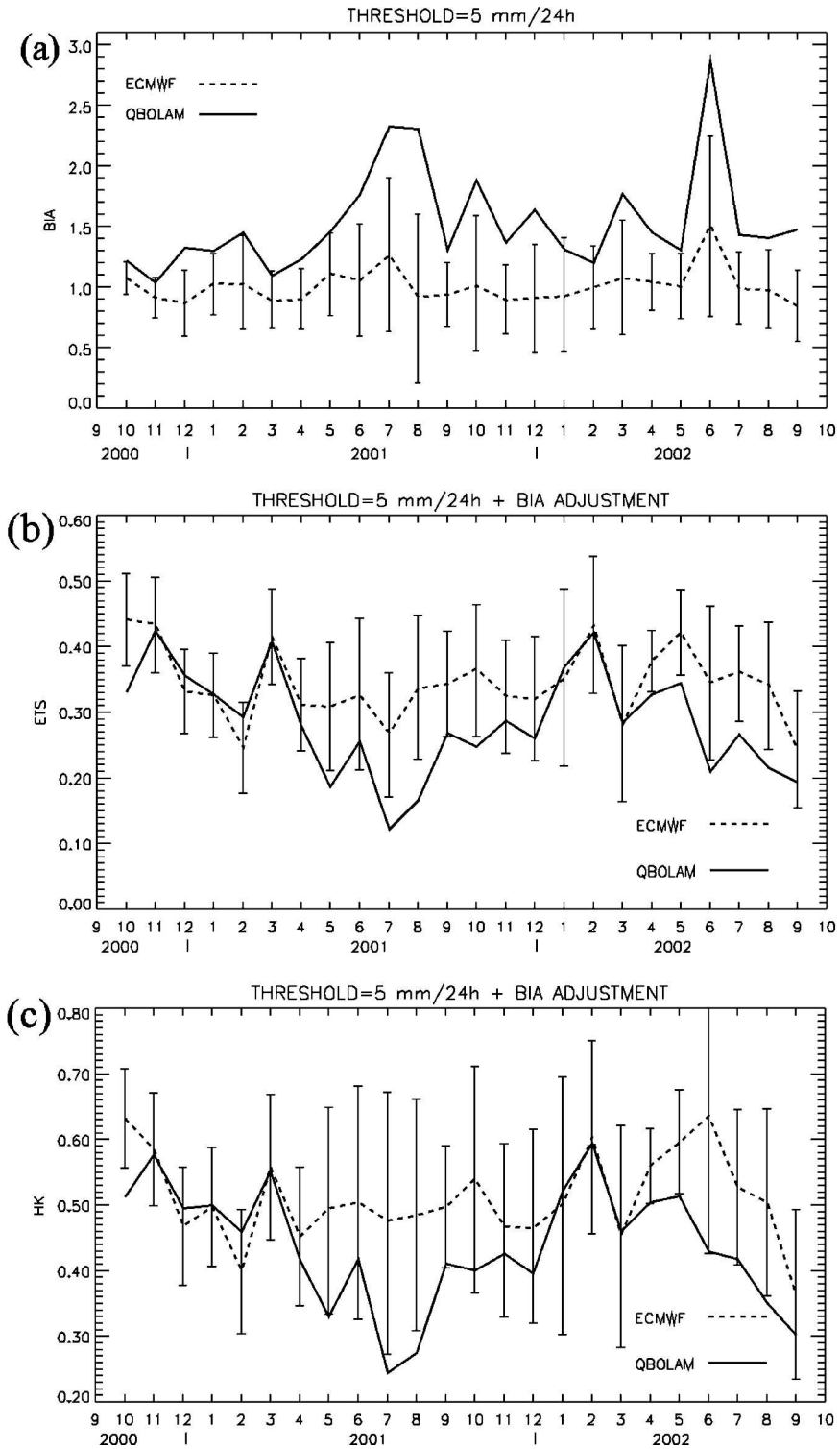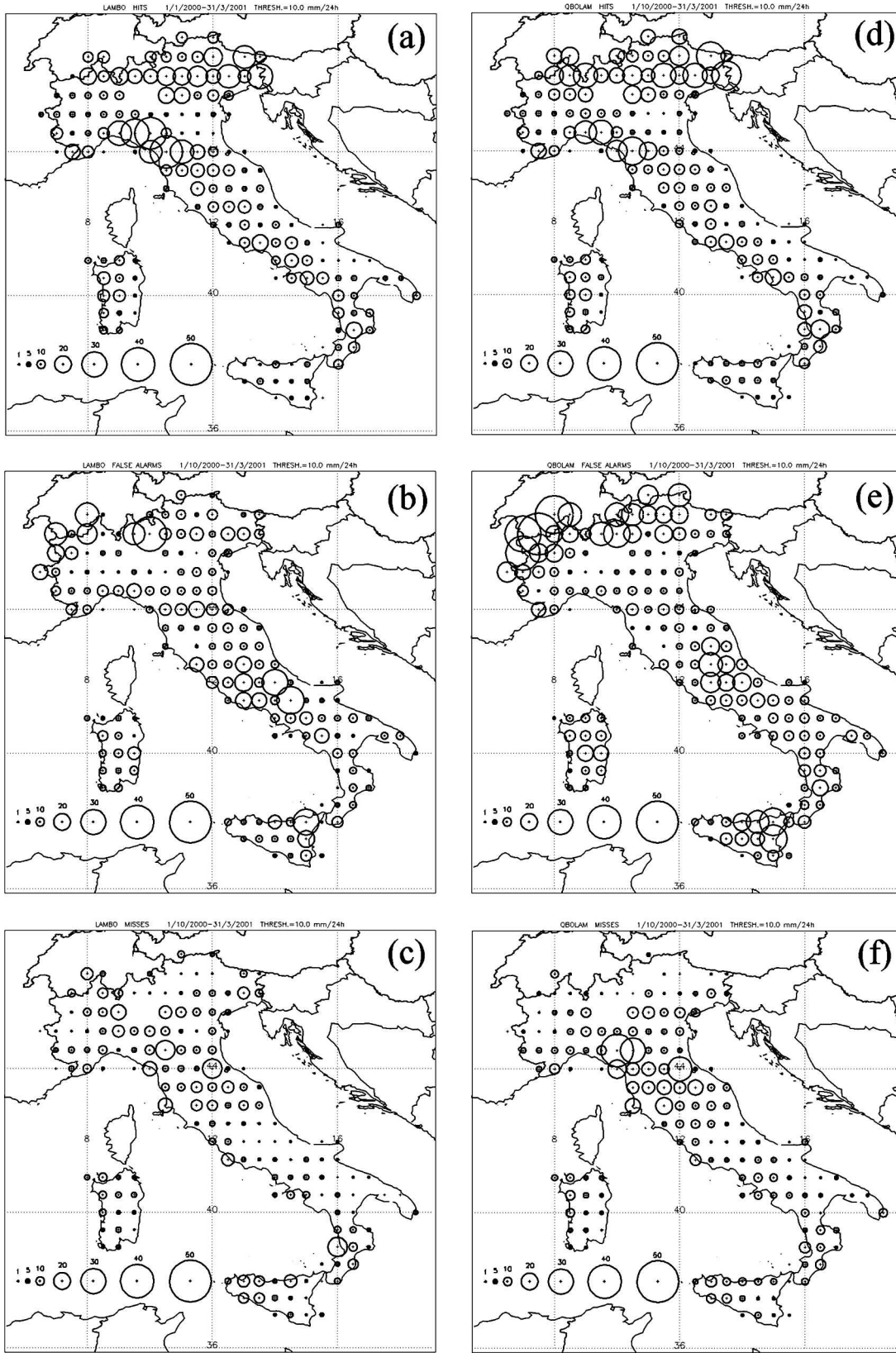
FIG. 12. As in Fig. 10, but for ECMWF and QBOLAM remapped on the 0.5° grid. The confidence intervals indicate the 2.5th and the 97.5th percentiles of the resampled distributions. QBOLAM skill scores are computed after BIA adjustment against ECMWF.

FIG. 13. Hits, false alarms, and misses, for the 10 mm $(24 \text{ h})^{-1}$ threshold, for (a)–(c) LAMBO and (d)–(f) QBOLAM for the "winter" period from 1 Oct 2000 to 31 Mar 2001, respectively.
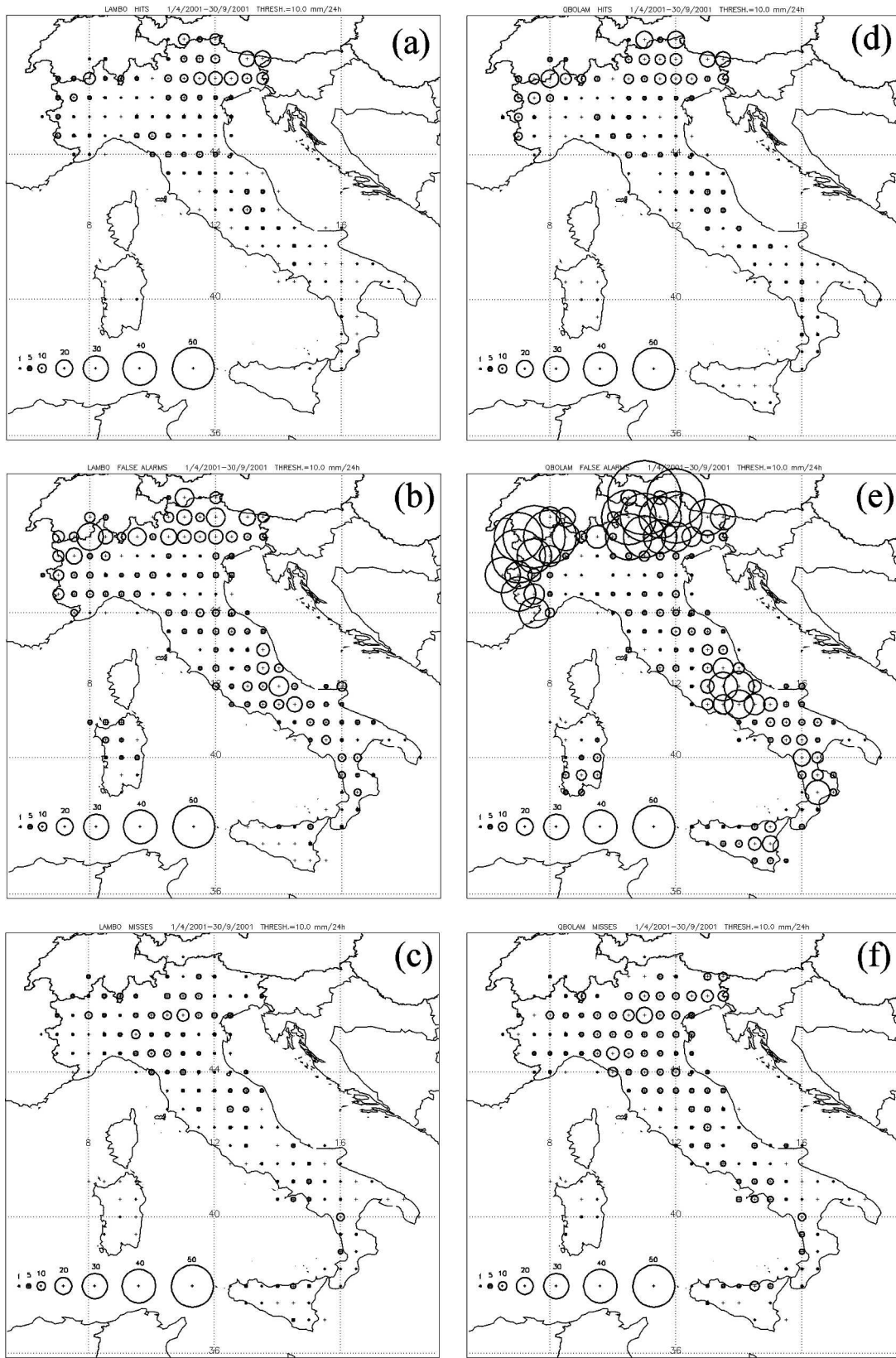
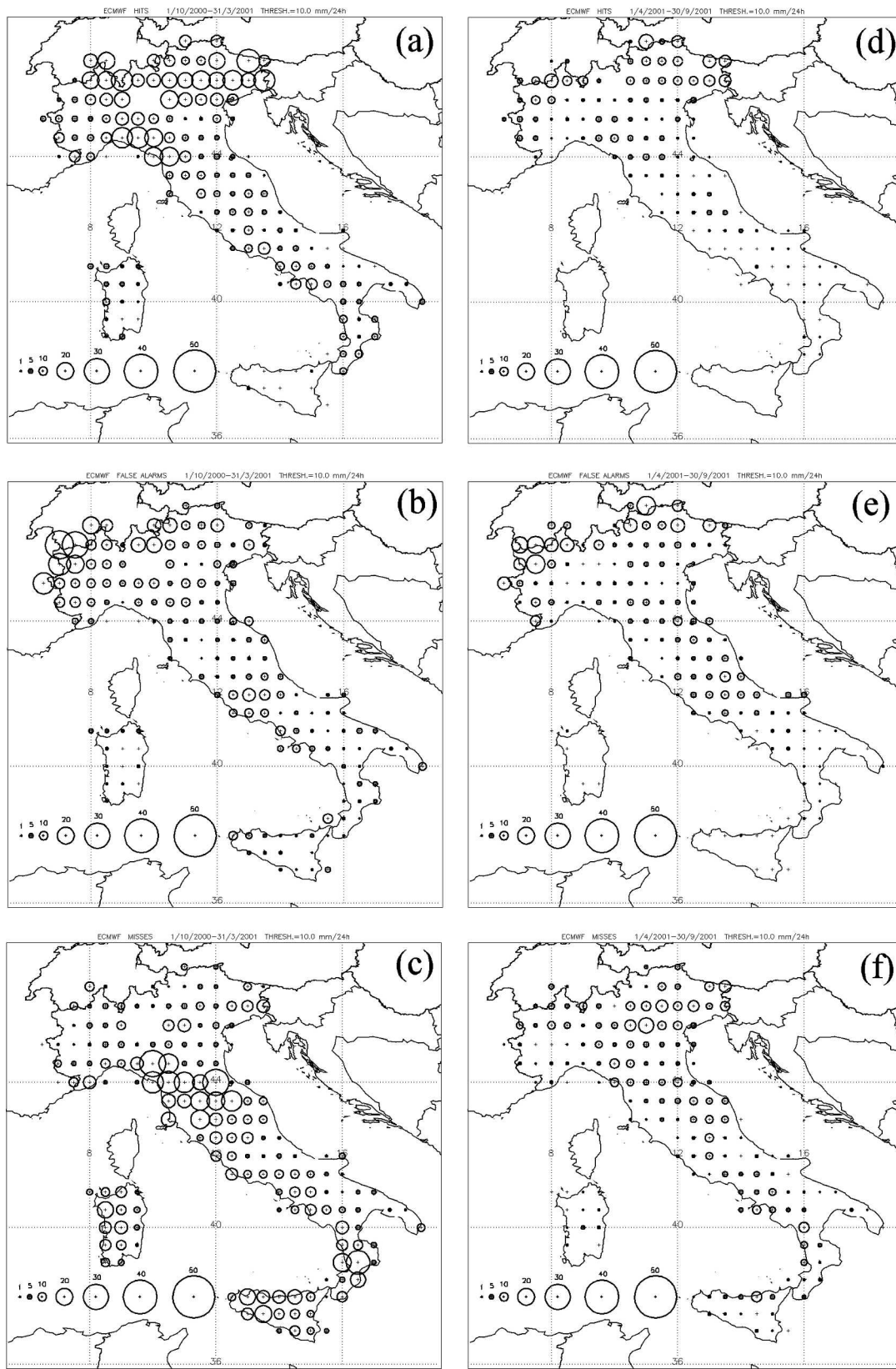FIG. 14. As in Fig. 13 but from the "summer" period: 1 Apr–30 Sep 2001.

FIG. 15. Hits, false alarms, and misses for ECMWF, for the 10 mm $(24 \text{ h})^{-1}$ threshold, for (a)–(c) the winter period, 1 Oct 2000–31 Mar 2001, and (d)–(f) the summer period, 1 Apr–30 Sep 2001, respectively.
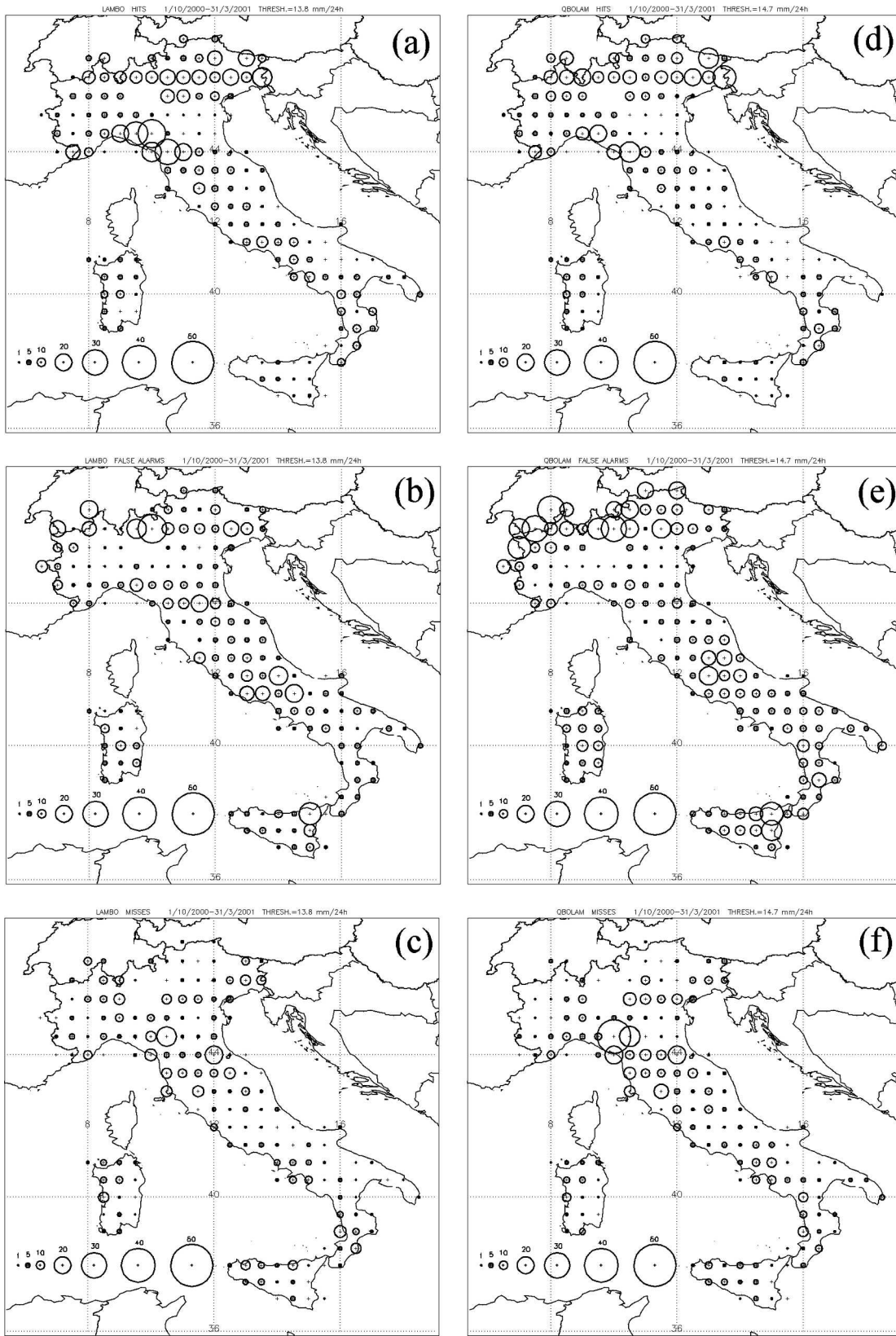
FIG. 16. As in Fig. 13 but after recalculation of the forecast thresholds according to the BIA adjustment against ECMWF BIA.
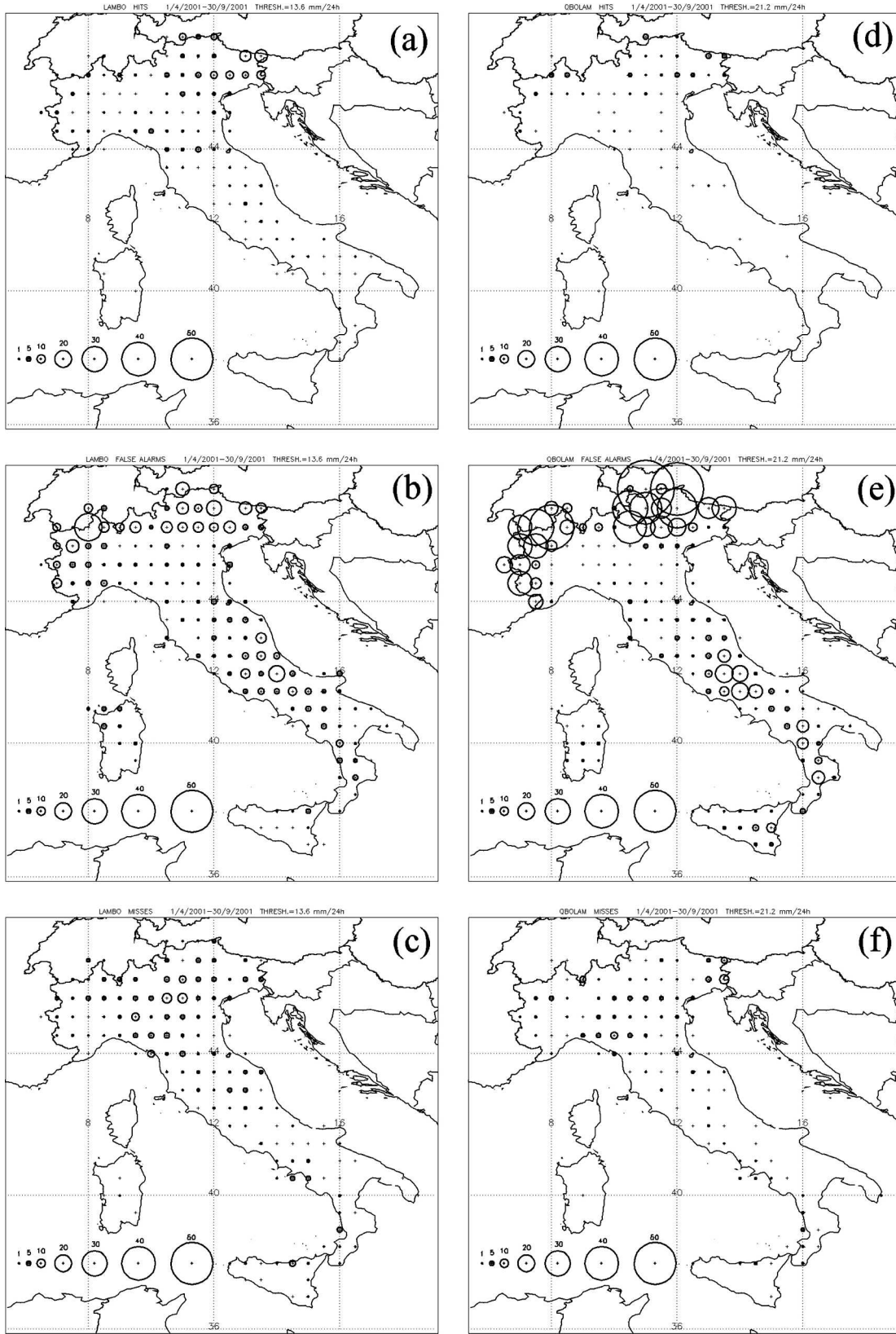
FIG. 17. As in Fig. 14 but after recalculation of the forecast thresholds according to the BIA adjustment against ECMWF BIA.

TABLE 5. LAM forecast thresholds [mm (24 h)$^{-1}$] after BIA adjustment against ECMWF, for the two selected time periods. The actual observation threshold is 10 mm (24 h)$^{-1}$.

LAM	Time period	
	1 Oct 2000–31 Mar 2001	1 Apr–30 Sep 2001
LAMBO	13.80	13.56
QBOLAM	14.70	21.20

TABLE 6. Percentage of distributions of the observed 24-h rainfall when analyzed on the LAM high-resolution grids and on the low-resolution grid. Only the results for observations on the QBOLAM 10-km native grid are presented, since they are very similar to those relative to the LAMBO grid.

Rainfall categories [mm (24 h)$^{-1}$]	High-resolution grid frequency	Low-resolution grid frequency
0.0–0.5	71.31	68.59
0.5–5.0	15.52	17.82
5.0–10.0	5.44	5.92
10.0–20.0	4.53	4.69
20.0–30.0	1.67	1.64
30.0–40.0	0.71	0.63
>40.0	0.82	0.72

on how small-scale details are reproduced, and it is likely that, although upscaled, LAM forecasts may still have a spatial structure that is different from forecasts originally made on a larger grid size. For example, Weygandt et al. (2004) observed that smoothness of the forecast precipitation field generally rewards more skill scores like ETS. One possible reason (apart from grid-size difference) is that any procedure, applied to reduce numerical instabilities, smoothes the fields (typically temperature, specific humidity) from which the forecast precipitation is derived (Chéruy et al. 2004; Harris et al. 2001).

As mentioned before, there is a prevalent increase in BIA scores for both models when upscaling from the 0.1° to the 0.5° grid. Indeed, an increase is noticeable for all thresholds below 10 mm (24 h)$^{-1}$ for LAMBO, while QBOLAM has a systematic increase for all thresholds. This indicates that false alarms are generally increased, due to the remapping on the 0.5° grid. This can be explained by two possible effects induced by remapping. First, local precipitation forecast maxima can be reduced by the remapping procedure, overspreading light precipitation over larger areas (Accadia et al. 2003b). Second, it should be noted that in the low-resolution verification, many of the 0.1° grid points, not considered in the high-resolution verification, are now included. LAM-predicted rainfall areas can be larger than one single 0.1° grid box, especially for low precipitation values. Then the remapped forecast could include some "extra" precipitation, resulting on average in an increase in forecast 24-h rainfall.

On the other hand, verification results could be affected by observation upscaling, since the simple average of rain gauge observations may decrease the observed maxima and increase the minima. This last hypothesis was checked by computing the observed rainfall frequencies at the different thresholds intervals for the high- and low-resolution grids (Table 6), in order to assess whether distribution differences are statistically significant. The simple box averaging at lower scale decreases the frequency of observed events below 0.5 mm (24 h)$^{-1}$ and above 20.0 mm (24 h)$^{-1}$. There is a small increase for the remaining thresholds. A one-tailed Chi-square test, with a 95% confidence level, was performed in order to check the goodness of fit between low-resolution and high-resolution frequencies. The results, with a threshold value of $\chi_T^2 = 0.246$ and the probability of 0.999 to obtain a chi square greater than χ_T^2, indicate that the two distributions are not sig-

nificantly different. These results indicate that rainfall analyses are not crucial in explaining the observed BIA changes; thus, LAM forecast upscaling has a major impact on scores. Moreover, the observed BIA increase suggests that LAM 24-h accumulated precipitation fields on the original high-resolution grid are generally overspread, in the sense that forecast precipitation affects many contiguous grid boxes on the original higher-resolution grids. This is obviously true for low precipitation values [<10 mm (24 h)$^{-1}$], which are related to stratiform phenomenology.

All models show a net reduction in skill when examining progressively larger thresholds. This indicates that a general difficulty in predicting correctly the magnitude and the location of rainfall events. This is true when verifying at high or low resolution. LAM low-resolution verification reduces the sensitivity to small displacement errors, concomitantly increasing skill scores. Apparently, LAMs often correctly predict the general weather situation (e.g., convective storms over a particular region) but misforecast the precise position of the rain events. The results suggest a general difficulty in predicting specific timing and location of convective events, which are frequently observed in the Mediterranean, both in the winter and summer seasons. Similar findings were presented by Islam et al. (1993) for the Tropics.

Results presented in section 4c help to better define this picture. The monthly score intercomparison (Figs. 10–12) mainly evidences the strong "wet" BIA of the LAMs during the warm season, and the correspondent skill score drop, especially for QBOLAM. On the other hand, such overforcasting can also produce an artificial increase in the skill scores (Hamill 1999), which is removed by the BIA adjustment technique, as discussed during the LAMBO–ECMWF comparison presented in section 4c. These results, which are consistent with the ones shown in Fig. 8, can be interpreted using the HK score (7). The LAMBO probability of detection [POD; first term on rhs in Eq. (7)] is higher compared to that of ECMWF, due to its higher rain event frequency. On the other hand, for the same reason, the probability of detecting correct no-rain forecasts [PON;

second term on rhs in Eq. (7)] is slightly reduced. The net result is a HK score that is higher for LAMBO than ECMWF. The application of the BIA adjustment decreases the POD (due to the reduction of hits), while the PON is increased, due to the decrease of false alarms (not shown). So LAMBO produced slightly higher HK scores compared to ECMWF, but this is due to its general tendency to overforecast precipitation, which artificially inflates the HK score [for a more detailed discussion on the effects of the BIA adjustment on skill scores the reader is referred to Accadia et al. (2003b)].

The spatial distribution of the contingency table elements (Figs. 13–17) provides some evidence of a geographic characterization of the different models' forecast error. The role of orography is apparent, since results show that a big part of the signal is due to precipitation over mountain ranges. Moreover, the increased resolution of LAMs shows a beneficial impact on precipitation forecasting, especially over the main Italian islands during the winter season, which have important mountain ranges, like Mount Etna in Sicily and Mount Gennargentu in Sardinia (see Fig. 1).

The overforecasting of precipitation along mountain ranges might be due to a modeling mismatch of the interaction of large-scale and convective parameterizations with orography. A known effect is the specific humidity diffusion along sigma coordinates over a mountain ridge, which contributes to the triggering of exceedingly high precipitation over mountain crests (A. Buzzi 2004, personal communication).

6. Conclusions

This paper documents what is, to our knowledge, the first systematic precipitation verification study over Italy. The 24-h precipitation forecasts from the operational limited area models LAMBO and QBOLAM were verified, using a 2-yr rain gauge observations dataset. A resampling technique has been used in order to provide nonprobabilistic categorical skill scores with confidence intervals. Moreover, a comparative study was also performed, checking ECMWF global model precipitation forecasts against the LAMs upscaled forecasts.

High-resolution verification requires that precipitation location be forecast very precisely. When this is not possible, skill scores at high resolution will tend to be lower than when verified on a reduced-resolution grid. In fact, both the QBOLAM and LAMBO models have better scores when verified on the 0.5° grid.

Ebert and McBride (2000) have proposed an alternative verification approach that seems to be promising. In addition to the simple grid-box verification they apply the contiguous rain area (CRA) approach. The contiguous rain area of a particular weather system is defined by choosing a specified precipitation isohyet,

which bounds the union of the observed and forecast precipitation areas (rain entities). The displacement error is assessed by appropriately shifting the forecast precipitation field so that the total squared difference against the observed rain field is minimized. In this way it is possible to decompose the total error into components due to location, rain volume, and pattern. Thus, this method seems to be more appropriate for verifying LAMs forecasts on their original grids. Mass et al. (2002) indicates that high horizontal resolution LAMs may provide realistic results, but this does not necessarily mean that the objectively scored grid-box-measured accuracy increases. The application of the CRA method in mesoscale verification will be an object of our future studies.

All of the considered models, including ECMWF, show less skill forecasting intense precipitation, due to the inherent difficulty in correctly locating (or representing) convective cells, squall lines, mesoscale convective systems, or mesoscale cyclones. Models may capture the meteorological situation correctly, but they may also be prone to some errors in timing, location, and intensity of rainfall. A net advantage offered by LAMs is the capability to better reproduce rainfall events during the winter season over relatively small islands, like Sardinia and Sicily. This is particularly true for LAMBO in our specific case, since this result is also confirmed after the application of the BIA adjustment against ECMWF.

LAMs, and to a lesser extent the ECMWF global model, tend to overforecast precipitation events across mountain ranges, resulting in a number of false alarms. This may indicate that precipitation parameterizations interact with orography producing excessive rainfall. A better understanding of the interactions of large-scale and convective precipitation schemes with orography could result in a better representation of mesoscale and convective phenomena, improving LAMs quantitative precipitation forecasts.

Finally, LAMBO and QBOLAM can be considered "frozen" versions, while the ECMWF global model is continuously evolving. This should be taken into account when performing a comparison between such a global model and mesoscale models. The sole additional information of a more detailed description of the orography has a slight beneficial impact on the miss rate obtained by LAMBO on the Italian main islands during winter, while QBOLAM pays a noticeable toll due to the simplified parameterizations needing to be used in order to implement the model on the QUADRICS computer. An updated version will run on a different parallel computer within 2005.

labria, Liguria, Marche, Piedmont, Sicily, and Valle d'Aosta, and to Servizio Agrometeorologico Regionale (SAR) of Sardinia, which provided the rainfall dataset from their rain gauge networks. By the way, we thank also Eng. M. Bussettini (Department of Internal and Marine Waters Protection, APAT, Rome, Italy) for his help in collecting observed precipitation data from Regional Services. We thank Mr. C. Transerici (ISAC–CNR, Rome, Italy) for his help in writing and testing the first version of the remapping code and also in collecting ECMWF global model forecasts. Finally, we thank three anonymous reviewers for their helpful suggestions that improved the quality of the paper.

REFERENCES

Accadia, C., and Coauthors, 2003a: Application of a statistical methodology for limited area model intercomparison using a bootstrap technique. *Nuovo Cimento, 26C,* 61–77.

——, S. Mariani, M. Casaioli, A. Lavagnini, and A. Speranza, 2003b: Sensitivity of precipitation forecast skill scores to bilinear interpolation and a simple nearest-neighbor average method on high-resolution verification grids. *Wea. Forecasting, 18,* 918–932.

Baldwin, M. E., cited 2000: QPF verification system documentation. [Available online at http://wwwt.emc.ncep.noaa.gov/mmb/ylin/pcpverif/scores/docs/mbdoc/pptmethod.html.]

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

——, 1973: Mesoscale objective analysis using weighted time-series observations. NOAA Tech. Memo. ERL NSSL-62, National Severe Storms Laboratory, Norman, OK, 60 pp. [NTIS COM-73-10781.]

Betts, A. K., 1986: A new convective adjustment scheme. Part I: Observational and theoretical basis. *Quart. J. Roy. Meteor. Soc., 112,* 677–691.

——, and M. J. Miller, 1986: A new convective adjustment scheme. Part II: Single column tests using GATE wave, BOMEX, ATEX and arctic air-mass data sets. *Quart. J. Roy. Meteor. Soc., 112,* 693–709.

Black, T., 1994: The new NMC mesoscale Eta Model: Description and forecast examples. *Wea. Forecasting, 9,* 265–278.

Buzzi, A., M. Fantini, P. Malguzzi, and F. Nerozzi, 1994: Validation of a limited area model in cases of Mediterranean cyclogenesis: Surface fields and precipitation scores. *Meteor. Atmos. Phys., 53,* 53–67.

Cherubini, T., A. Ghelli, and F. Lalaurette, 2002: Verification of precipitation forecasts over the Alpine region using a high-density observing network. *Wea. Forecasting, 17,* 238–249.

Chéruy, F., A. Speranza, A. Sutera, and N. Tartaglione, 2004: Surface winds in the Euro-Mediterranean area: The real resolution of numerical grids. *Ann. Geophys., 22,* 4043–4048.

Courtier, P., J. N. Thépaut, and A. Hollingsworth, 1994: A strategy for operational implementation of 4D-Var, using an incremental approach. *Quart. J. Roy. Meteor. Soc., 120,* 1367–1387.

Diaconis, P., and B. Efron, 1983: Computer-intensive methods in statistics. *Sci. Amer., 248,* 116–130.

Ebert, E. E., and J. L. McBride, 2000: Verification of precipitation in weather systems: Determination of systematic errors. *J. Hydrol., 239,* 179–202.

——, U. Damrath, W. Wergen, and M. E. Baldwin, 2003a: The WGNE assessment of short-term quantitative precipitation forecasts. *Bull. Amer. Meteor. Soc., 84,* 481–491.

——, ——, ——, and ——, 2003b: Supplement to the WGNE assessment of short-term quantitative precipitation forecasts. *Bull. Amer. Meteor. Soc., 84,* ES10–ES11.

Fisher, R. A., 1925: *Statistical Methods for Research Workers.* Oliver and Boyd, 239 pp.

Gabella, M., and R. Mantovani, 2001: The floods of 13–16 October 2000 in Piedmont (Italy): Quantitative precipitation estimates using radar and a network of gages. *Weather, 56,* 337–351.

Gallus, W. A., 2002: Impact of verification grid-box size on warm-season QPF skill measures. *Wea. Forecasting, 17,* 1296–1302.

Georgelin, M., and Coauthors, 2000: The second COMPARE exercise: A model intercomparison using a case of a typical mesoscale orographic flow, the PYREX IOP3. *Quart. J. Roy. Meteor. Soc., 126,* 991–1030.

Hamill, T. M., 1999: Hypothesis tests for evaluating numerical precipitation forecasts. *Wea. Forecasting, 14,* 155–167.

Hanssen, A. W., and W. J. A. Kuipers, 1965: On the relationship between the frequency of rain and various meteorological parameters. *Meded. Verh., 81,* 2–15.

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.

Islam, S. B., L. Rafael, and K. A. Emanuel, 1993: Predictability of mesoscale rainfall in the tropics. *J. Appl. Meteor., 32,* 297–310.

Jakob, C., and S. A. Klein, 2000: A parametrization of the effects of cloud and precipitation overlap for use in general circulation models. *Quart. J. Roy. Meteor. Soc., 126,* 2525–2544.

Janjić, Z. I., 1994: The step-mountain eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. *Mon. Wea. Rev., 122,* 927–945.

Koch, S. E., M. desJardins, and P. J. Kocin, 1983: An interactive Barnes objective map analysis scheme for use with satellite and conventional data. *J. Climate Appl. Meteor., 22,* 1487–1503.

Kuo, H. L., 1974: Further studies of the parameterization of the influence of cumulus convection on large scale flow. *J. Atmos. Sci., 31,* 1232–1240.

Lazic, L., and B. Telenta, 1990: *Documentation of the UB/NMC (University of Belgrade and National Meteorological Center, Washington) Eta Model.* WMO Tropical Meteorology Research Programme (TMRP) Rep. 40, WMO/TD-366, 308 pp.

Louis, J. F., M. Tiedke, and J. F. Geleyn, 1982: A short history of the PBL parameterization at ECMWF. *Proc. Workshop on Planetary Boundary Layer Parameterization*, Shinfield Park, Reading, United Kingdom, ECMWF, 59–79.

Mass, C. F., D. Ovens, K. Westrick, and B. A. Colle, 2002: Does increasing horizontal resolution produce more skillful forecasts? *Bull. Amer. Meteor. Soc., 83,* 407–430.

McBride, J. L., and E. E. Ebert, 2000: Verification of quantitative precipitation forecasts from operational numerical weather prediction models over Australia. *Wea. Forecasting, 15,* 103–121.

Mesinger, F., 1996: Improvements in quantitative precipitation forecasting with the Eta regional model at the National Centers for Environmental Prediction: The 48-km upgrade. *Bull. Amer. Meteor. Soc., 77,* 2637–2649.

——, Z. I. Janjić, S. Nickovic, D. Gavrilov, and D. G. Deaven, 1988: The step-mountain coordinate: Model description and performance for cases of Alpine lee cyclogenesis and for a case of an Appalachian redevelopment. *Mon. Wea. Rev., 116,* 1493–1518.

——, T. L. Black, and M. E. Baldwin, 1997: Impact of resolution and of the eta coordinate on skill of the Eta Model precipitation forecasts. *Numerical Methods in Atmospheric and Oceanic Modeling.* C. Lin, R. Laprise, and H. Ritchie, Eds., Canadian Meteorological and Oceanographic Society/NRC Research Press, 399–423.

Nagata, M., and Coauthors, 2001: Third COMPARE workshop: A model intercomparison experiment of tropical cyclone inten-

sity and track prediction. *Bull. Amer. Meteor. Soc.,* **82,** 2007–2020.

Paccagnella, T., S. Tibaldi, R. Buizza, and S. Scoccianti, 1992: High resolution numerical modelling of convective precipitation over northern Italy. *Meteor. Atmos. Phys.,* **50,** 143–163.

Page, J. K., 1986: *Prediction of Solar Radiation on Inclined Surfaces.* Vol. 3, Series F, Solar Energy R&D in the European Community, Dordrecht Reidel, 459 pp.

Rabier, F., J. N. Thépaut, and P. Courtier, 1998: Extended assimilation and forecast experiments with a four-dimensional variational assimilation system. *Quart. J. Roy. Meteor. Soc.,* **124,** 1861–1887.

——, H. Järvinen, E. Klinker, J.-F. Mahfouf, and A. Simmons, 2000: The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics. *Quart. J. Roy. Meteor. Soc.,* **126,** 1143–1170.

Ritter, B., and J. F. Geleyn, 1992: A comprehensive radiation scheme for numerical weather prediction models with potential applications in climate simulations. *Mon. Wea. Rev.,* **120,** 303–325.

Ruti, P. M., C. Cacciamani, T. Paccagnella, A. Bargagli, and C. Cassaro, 1997: Intercomparison between BATS and LSPM surface schemes, using point micrometeorological data set. *Beitr. Phys. Atmos.,* **70,** 201–220.

Schaefer, J. T., 1990: The critical success index as an indicator of warning skill. *Wea. Forecasting,* **5,** 570–575.

Simmons, A. J., D. M. Burridge, M. Jarraud, C. Girard, and W. Wergen, 1989: The ECMWF medium-range prediction models: Development of the numerical formulations and the impact of increased resolution. *Meteor. Atmos. Phys.,* **40,** 28–60.

Speranza, A., and Coauthors, 2004: POSEIDON: An integrated system for analysis and forecast of hydrological, meteorological and surface marine fields in the Mediterranean area. *Nuovo Cimento,* **27C,** 329–345.

Teixeira, J., 1999: The impact of increased boundary layer vertical resolution on the ECMWF forecast system. ECMWF Tech. Memo. 268, 55 pp.

Tiedke, M., 1993: Representation of clouds in large-scale models. *Mon. Wea. Rev.,* **121,** 3040–3061.

Weygandt, S. S., A. Loughe, S. Benjamin, and J. Mahoney, cited 2004: Scale sensitivities in model precipitation skill scores during IHOP. Preprints, *22d Conf. on Severe Local Storms,* Hyannis, MA, Amer. Meteor. Soc., CD-ROM, P4.7.

Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences.* Academic Press, 467 pp.