



Verification of ensemble forecasts of Mediterranean high-impact weather events against satellite observations

J.-P. Chaboureau¹, O. Nuisser², and C. Claud³

¹Laboratoire d'Aérodynamique, Université de Toulouse, CNRS, UMR5560, Toulouse, France

²GAME/CNRM, CNRS/INSU Groupe d'étude de l'atmosphère météorologique (GAME), URA1357, Toulouse, France

³Laboratoire de Météorologie Dynamique/IPSL, CNRS, UMR8539, Ecole Polytechnique, Palaiseau, France

Correspondence to: J.-P. Chaboureau (jean-pierre.chaboureau@aero.obs-mip.fr)

Received: 21 March 2012 – Revised: 18 June 2012 – Accepted: 11 July 2012 – Published: 3 August 2012

Abstract. Ensemble forecasts at kilometre scale of two severe storms over the Mediterranean region are verified against satellite observations. In complement to assessing the forecasts against ground-based measurements, brightness temperature (BT) images are computed from forecast fields and directly compared to BTs observed from satellite. The so-called model-to-satellite approach is very effective in identifying systematic errors in the prediction of cloud cover for BTs in the infrared window and in verifying the forecasted convective activity with BTs in the microwave range. This approach is combined with the calculation of meteorological scores for an objective evaluation of ensemble forecasts. The application of the approach is shown in the context of two Mediterranean case studies, a tropical-like storm and a heavy precipitating event. Assessment of cloud cover and convective activity using satellite observations in the infrared (10.8 μm) and microwave regions (183–191 GHz) provides results consistent with other traditional methods using rainfall measurements. In addition, for the tropical-like storm, differences among forecasts occur much earlier in terms of cloud cover and deep convective activity than they do in terms of deepening and track. Further, the underdispersion of the ensemble forecasts of the two high-impact weather events is easily identified with satellite diagnostics. This suggests that such an approach could be a useful method for verifying ensemble forecasts, particularly in data-sparse regions.

1 Introduction

In the last few decades, improvements in the data assimilation, modelling and observing systems have resulted in good progress in predicting weather at the synoptic scale. Simmons and Hollingsworth (2002) examined the forecast errors for the 500-hPa height and mean-sea-level pressure produced by the European Centre for Medium-Range Weather Forecasts (ECMWF). They reported that the improvement between 1990 and 2001 for the Northern Hemisphere corresponded to a 1-day extension of the forecast range at which a given level error was reached. Nowadays, the root-mean square error of 1-day 500 hPa height forecasts has fallen below the 10 m level typical of radiosonde measurement errors. However, accurate forecasts of high-impact weather events are still challenging because of our inaccurate knowledge of the state of the atmosphere and model errors. This has led to the development of global operational ensemble prediction systems (EPSs) to sample all the uncertainty sources for the initial state of the atmosphere (e.g. Molteni et al., 1996). When running ensemble forecasts with limited-area models (LAMs), the lateral boundary conditions are an additional source of uncertainties to be considered. Furthermore, the rapid growth of convective-scale perturbations has led to the development of specific methods for LAMs, such as the shifting initialization technique, the use of multi models or physical parameterizations in a model, the selection of members from large-scale forecasts, and the addition of perturbations to initial and boundary conditions (see Argence et al., 2008; Davolio et al., 2009; Vich et al., 2011; Vié et al., 2011; Tapiador et al., 2012, among others).

Work dedicated to verifying forecasts has accompanied this modelling effort. In particular, the need for verification of forecasts is strong for sensible weather such as cloud and rain fields. First, cloud cover and rainfall are meteorological variables that are of importance to the general public. Second, it can be crucial to accurately predict diabatic processes, e.g. in cases of rapid cyclogenesis or flash floods. Third, as small differences in large-scale forcing, such as 500-hPa height, can result in large errors in the cloud and rain fields, assessment of the latter is of great interest as a critical measure of the model performance. Ebert et al. (2003) verified short-range quantitative precipitation forecasts from 11 operational numerical weather prediction models against rain gauge observations. They concluded that the skill for forecasts of rain greater than 20 mm per day was generally quite low, reflecting the difficulty in predicting heavy rain accurately in time and space.

As a complement to conventional rainfall measurements with rain gauge networks, which are unevenly distributed, satellite observations provide useful information on both cloud cover and rainfall over the whole globe, including data-sparse areas such as oceans. A way of using satellite observation to verify model outputs consists in the so-called model-to-satellite approach. In this approach, radiative quantities such as brightness temperature (BT) are calculated from the forecast fields and directly compared with satellite observations. In conducting the forecast verification in the observation space, errors from the observation are reduced to the instrumental uncertainties. In particular, this approach avoids the drawback of potential systematic differences between satellite retrievals and forecasts that may appear because of different assumptions in the retrieval algorithm and the meteorological model. It offers also the advantage to use satellite observations for verification purposes in near real time. The approach has already been previously applied to radiometer channels, mostly in the thermal infrared window, to estimate model cloudiness and identify drawbacks in cloud parameterizations (Morcrette, 1991; Chaboureau et al., 2000, 2002; Chevallier et al., 2001; Argence et al., 2008; Chaboureau et al., 2008; Grasso et al., 2008; Otkin and Greenwald, 2008; Otkin et al., 2009, among others).

Few studies have shown the advantage of using satellite observations combined with the model-to-satellite approach to evaluate the skill of forecast systems in predicting cloud cover. Söhne et al. (2008) performed verifications of cloud cover forecasts with satellite observations over West Africa, a data-sparse region. They showed a dependency of the forecast skill on the intensity of the synoptic forcing. However, the forecasts performed at low resolution (32 km horizontal grid spacing) mostly showed shortcomings in the representation of convection and clouds. Clark and Chaboureau (2010) demonstrated the benefits of using satellite observations to identify sources of uncertainties in a kilometre-scale forecast of heavy precipitation over southern France. They related the performance of the precipitation forecast to the prediction

of the intensity of the humidity flux from the sea during the stratiform regime and to the timely triggering of convection over the sea during the convective episode.

The purpose of the present study is to show the advantage of using satellite observations and the model-to-satellite approach to verify ensemble forecasts. This study is part of the “Forecast and projection in climate scenario of Mediterranean intense events: uncertainties and propagation on environment”, the MEDUP project. This project aims to characterize the propagation of sources of uncertainties with the forecast and climate projection for Mediterranean high-impact weather events. It lies within the framework of HyMeX in that it develops modelling and forecasting tools that could be deployed during this 10-yr program. Two Mediterranean cases were investigated in terms of accuracy and skill for cloud cover and rain. They were taken from two previous MEDUP case studies (Vié et al., 2011; Chaboureau et al., 2012). For both cases, a forecast ensemble was built using a convection-permitting model. The first one concerned intense cyclogenesis leading to the formation of a tropical-like storm or medicane (Mediterranean “hurricane”) over southeastern Italy (Moscatello et al., 2008; Davolio et al., 2009; Claud et al., 2010; Laviola et al., 2011; Pantillon et al., 2012, among others). For that case, Chaboureau et al. (2012) built different atmospheric states by simply shifting the initialization time. The so-called time-lagged initialization method generated a set of perturbed initial conditions that allowed them to study the effect of initial-condition uncertainties on the evolution of the medicane. The second case was a heavy precipitation event over southern France. Vié et al. (2011) built an ensemble forecast at kilometre scale from a global ensemble forecast and examined the impact of lateral boundary conditions on the prediction of the rain event.

The paper is organized as follows. Section 2 presents the forecasts and the verification approach. Section 3 gives the results obtained from the ensemble forecasts of the medicane. Section 4 describes the cloud verification of the ensemble forecasts of the heavy precipitation event. Section 5 concludes the paper.

2 Forecasts and verification approach

2.1 Meso-NH forecasts

The forecasts of the medicane were made with the non-hydrostatic model Meso-NH (Lafore et al., 1998) version 4.7 using the two-way interactive grid-nesting method with triply nested grids. The model domains (Fig. 1a) had horizontal grid spacings of 32, 8, and 2 km. For the inner grid, deep convection was explicitly resolved. The model included a turbulence parameterization (Cuxart et al., 2000), a microphysical scheme that predicts the evolution of the mixing ratios of six water species (water vapor, cloud droplet,

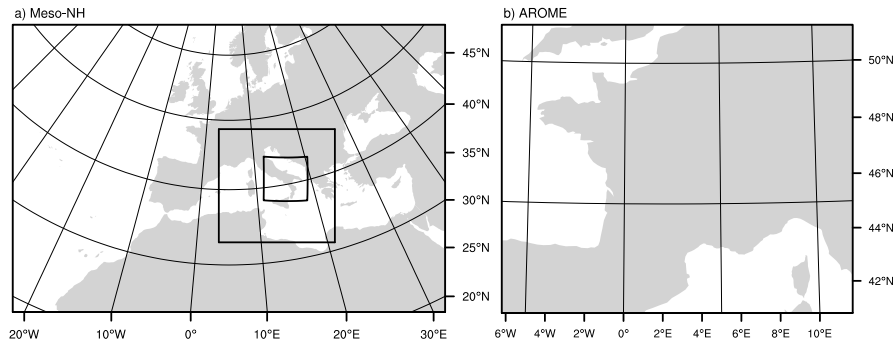


Fig. 1. Domains of (a) the outer Meso-NH model and (b) AROME. In (a) the boxes show the location of the two inner models.

raindrop, ice crystal, snow and graupel, Pinty and Jabouille, 1998) and a subgrid cloud cover and condensate content schemes (Chaboureau and Bechtold, 2002, 2005). The forecasts were designed to contrast the impact of the initial conditions on the development of the medicane (Chaboureau et al., 2012). Two sets of three lagged forecasts each were run using initial and boundary conditions provided by either Action de Recherche Petite Echelle Grande Echelle (ARPEGE) or ECMWF analyses (referred to as ARP and ECM experiments, respectively). This set of six members was built by shifting the initialization time of the runs. The first members started on 25 September at 00:00 UTC (ARPC and ECMC), the second at 06:00 UTC (ARPB and ECMB) and the last at 12:00 UTC (ARPA and ECMA). The experiments were then integrated during the rapid development of the medicane until 26 September at 18:00 UTC. For further details on the simulation setup, the reader is referred to Chaboureau et al. (2012).

2.2 AROME forecasts

The forecasts of the heavy precipitating event were performed with the operational convective-permitting Application of Research to Operations at Mesoscale (AROME) model from Météo-France (Seity et al., 2011), at a horizontal grid spacing of about 2.5 km (see the model domain in Fig. 1b). AROME is based on the nonhydrostatic version of the adiabatic equations of the limited area model *Aire Limitée Adaptation Dynamique développement International* (ALADIN), using physical parameterizations from the research model Meso-NH, which includes the microphysical scheme of Pinty and Jabouille (1998), the turbulence parameterization of Cuxart et al. (2000) and the shallow convection of Pergaud et al. (2009). The ensemble forecast was designed to assess the impact of uncertainty on large-scale lateral boundary conditions (LBCs) by providing the AROME simulations with LBCs from the members of the global ensemble prediction system *Prévision d'Ensemble ARPEGE* (PEARP). The PEARP ensemble has 11 members (hereafter P0, P1, etc.) obtained by adding perturbations, which blends

a breeding technique and calculation of singular vectors. For each PEARP member run for 24 h, a 24-h AROME forecast was issued at 12:00 UTC over one full month from 6 October to 5 November 2008. Here we used the AROME ensemble forecast starting on 1 November. Further details can be found in Vié et al. (2011).

2.3 Satellite observations and simulated brightness temperatures

Two types of observations were used for the verification of forecasts and interpolated onto the model grid. The Meteosat Second Generation (MSG) observations obtained from the Spinning Enhanced Visible and Infra-Red Imager (SEVIRI) have a temporal resolution of 15 min and a spatial sampling of 3 km at sub-satellite point. Here, we used 3-hourly measurements of BT in the thermal infrared window (10.8 μm), which is mainly sensitive to the temperature of opaque clouds at their top. Clouds are much more transparent for microwave radiation, which can give some information on cloud and rain content, depending on the frequency. The Advanced Microwave Sounding Unit (AMSU-B) on the polar-orbiting National Oceanic and Atmospheric Administration (NOAA)-15 to -17 platforms replaced by Microwave Humidity Sounder (MHS) on the NOAA-18 and MetOp-2 platforms allowed us to sense the rainfall intensity up to every 3 h and with a field of view of 16 km at nadir. These two sounders share similar moisture channels, for which their slightly different radiometric characteristics did not affect the rainfall detection used here (Claud et al., 2012). Tables 1 and 2 give the time of overpasses for the medicane and the heavy precipitation event, respectively. Following Funatsu et al. (2007), we used the observations from AMSU-B/MHS moisture channels (183–191 GHz). In the absence of any hydrometeors, channel 3, which senses humidity in the upper troposphere, measures lower BTs than channel 4, which senses the middle troposphere. The latter in turn shows lower BTs than channel 5, which peaks in the lower troposphere. In the presence of icy hydrometeors, large amounts of which are preferentially found at low levels, radiation can be efficiently

Table 1. Overpasses in September 2006 for the medicane.

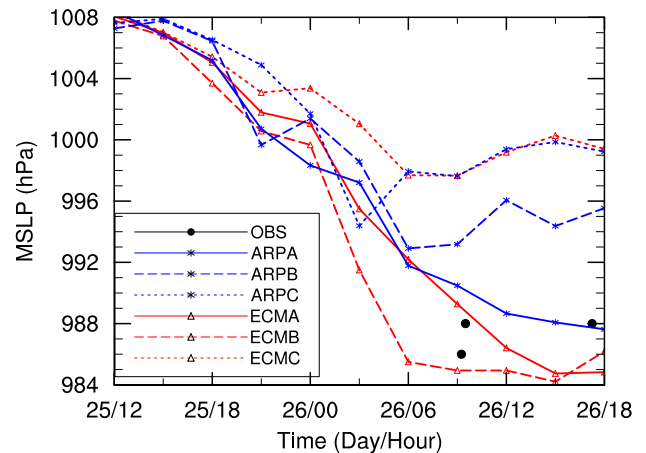
Satellite	Day	Time (UTC)	Time selected for forecast
NOAA-16	25	14:33	15:00
NOAA-15	25	16:16	18:00
NOAA-17	25	21:10	21:00
NOAA-18	26	01:43	00:00
NOAA-15	26	04:33	06:00
NOAA-18	26	11:28	12:00
NOAA-16	26	14:24	15:00
NOAA-15	26	16:54	18:00

Table 2. Overpasses in November 2008 for the heavy precipitating event.

Satellite	Day	Time (UTC)	Time selected for forecast
NOAA-16	01	16:58	18:00
MetOp-2	01	20:01	21:00
NOAA-18	02	01:16	00:00
NOAA-16	02	05:18	06:00
MetOp-2	02	09:57	09:00
NOAA-18	02	12:45	12:00

scattered so the weighting function for channel 5 peaks at a higher level in the atmosphere, thereby depressing BT to values close to that of channel 3. Over the Mediterranean, Funatsu et al. (2007) found that a difference between channels 3 and 5 (hereafter, B3m5) of > -8 K corresponded statistically to moderate rainfall (about 10 mm in 3 h) when compared with Tropical Rainfall Measuring Mission (TRMM) retrievals. Based on the above principle, but with a higher threshold (zero) applied to the all possible combinations of moisture channels (i.e., channels 3 minus 5, channels 4 minus 5, and 3 minus 4 simultaneously > 0 K), a deep convection threshold (DCT) was used to detect deep convection over the Mediterranean (Funatsu et al., 2007, 2009). The statistical analysis of Funatsu et al. (2007) reveals that DCT generally corresponds to heavy rainfall (about 20 mm in 3 h).

These two types of observations in the infrared and microwave regions were simulated from the model fields of temperature, water vapour and hydrometeors using the radiative transfer code RTTOV (Radiative Transfer for Tiros Operational Vertical Sounder) version 8.7 (Saunders et al., 2005). In the thermal infrared window, surface emissivity was given by the Ecoclimap database (Masson et al., 2003), SEVIRI viewing angles were computed for each model grid point, and the grey body approximation was considered for clouds (Chevallier et al., 2001). Radiative properties for water and ice clouds were taken from Hu and Stamnes (1993) and Baran and Francis (2004), respectively. In the microwave region, surface emissivity was calculated over sea using the

**Fig. 2.** Time evolution of MSLP minimum from 15:00 UTC, 25 September to 18:00 UTC, 26 September 2006; the black dots are values recorded at stations over land (Moscatello et al., 2008).

FASTEM code and set to the typical value for bare soil elsewhere. Absorption and scattering effects by hydrometeors were taken into account using precomputed Mie tables for liquid water, cloud ice, rain, and precipitating ice (Bauer, 2001). As the AMSU-B/MHS viewing angles and observation times vary with each orbit, the synthetic AMSU-B BT were calculated at fixed times (every 3 h) and angle (nadir view) and for the NOAA-16 platform only. The time delay and the approximations in the calculation could result in systematic errors. They are, however, of second order with respect to the spread of the ensemble forecasts as shown below.

2.4 Rain retrievals

The verification was completed by a comparison of the precipitation amount. For the heavy precipitating case, we used 3-hourly measurements by rain gauges over France. Each rain gauge was compared with the nearest model grid point and precipitation was averaged over the rain gauges. The use of rain gauges limited the verification to land areas only, and was not suitable for oceanic cases such as the medicane. For that case, we therefore used rain retrievals of Laviola and Levizzani (2011) based on a linear combination of the AMSU-B moisture channels. The retrieval used the spectral difference in the depression of the microwave radiation due to scattering by icy hydrometeors and absorption by raindrops. First results obtained from a comparison with rain retrievals from TRMM products were good and a validation study is underway (Laviola and Levizzani, 2011).

2.5 Verification approach

In the following, the verification of cloud and rain forecasts uses two measures, one of accuracy and the other of skill. As a measure of accuracy, the bias between forecasted

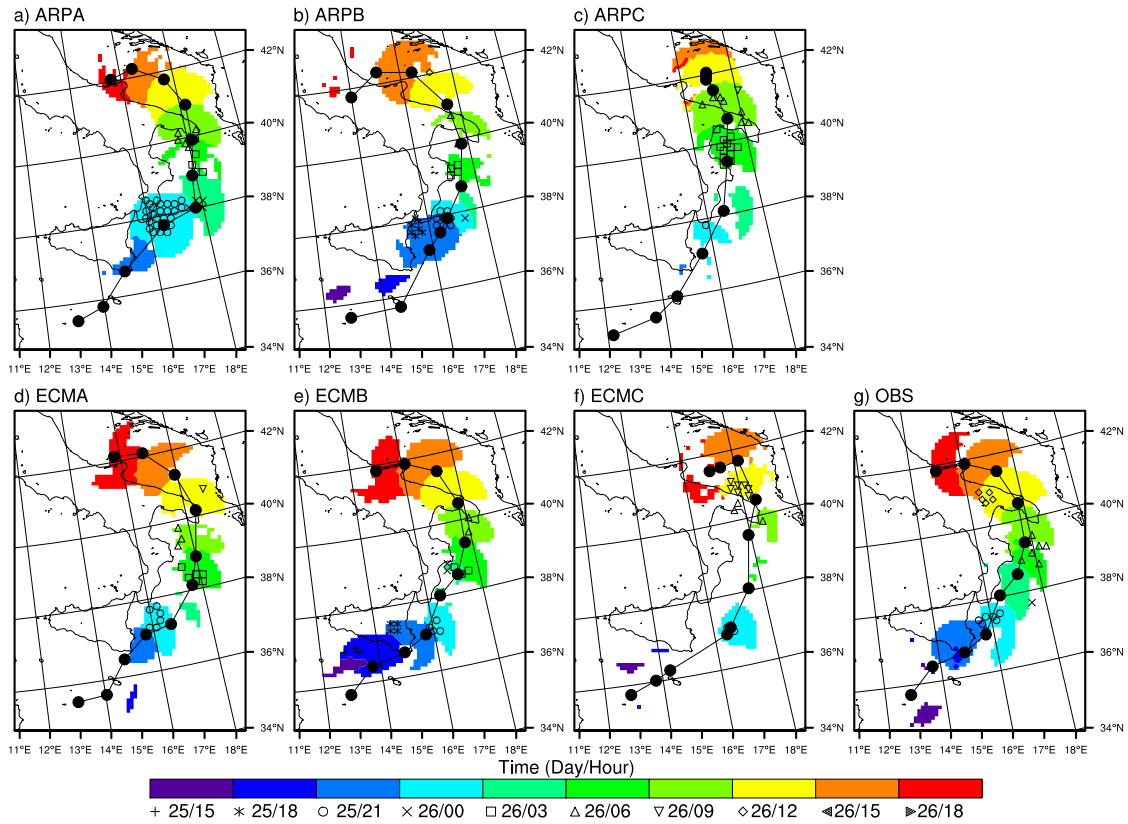


Fig. 3. Time (day/hour) associated with SEVIRI 10.8 μm BT less than 250 K (shading) and AMSU-B DCT (symbols) from 15:00 UTC, 25 September to 18:00 UTC, 26 September 2006 within an 80-km circle centred on the medicane for (a–f) Meso-NH forecasts and (g) MSG and AMSU-B observations. The positions and track of the medicane are shown at 3 h intervals with filled dots and lines, respectively.

and observed fields of cloud and rain was calculated in order to identify any systematic model error. Skill was estimated using a categorical score that quantify the matching between observed and simulated forecasts at gridpoints. For a given threshold, a contingency table was formed by classifying events as either non-high-cloud (non-rain) or high-cloud (rain) in the observation and the forecast. Among the large number of categorical scores in the literature, the categorical Symmetric Extreme Dependency Score (SEDS) proposed by Hogan et al. (2009) was used because of its very attractive properties. First, SEDS is equitable for large samples, meaning that a random forecast yields an expected score of zero. Second, SEDS is difficult to hedge because of its transpose symmetry (swapping the observations and the forecast does not change the score). Last, SEDS is independent of the frequency of occurrence of the quantity being verified, which is important for the verification of rare event forecasts, as those of heavy rainfalls. SEDS is defined as

$$SEDS = \frac{\ln\{(a + b)/n\} + \ln\{(a + c)/n\}}{\ln(a/n)} - 1 \quad (1)$$

where n is the total number of elements and a , b and c represent the number of hits, false alarms and misses, respectively.

SEDS typically lies between 0 and 1, the values for random and perfect forecasts, respectively.

3 The medicane case

The medicane was born during the night of 24 September 2006 in the lee of the Atlas Mountains. It subsequently moved eastward over the Strait of Sicily on the morning of 25 September. It deepened strongly on the morning of 26 September, while transiting over the Ionian Sea, and became a medicane at 09:00 UTC with a full tropospheric warm core over the Adriatic Sea.

The cyclone development is first examined in terms of its deepening by looking at the time evolution of the mean-sea-level pressure (MSLP) minimum (Fig. 2). Because the medicane crossed the southeastern tip of Italy and landed over eastern Italy, the MSLP minimum was recorded twice, with the lowest value reaching 986 hPa at 09:15 UTC on 26 September (Moscatello et al., 2008). Every forecast showed an important deepening in the first 12 h, from about 1008 to 1000 hPa. In the following 12 h or so, only three forecasts (ARPA, ECMA and ECMB) attained a minimum

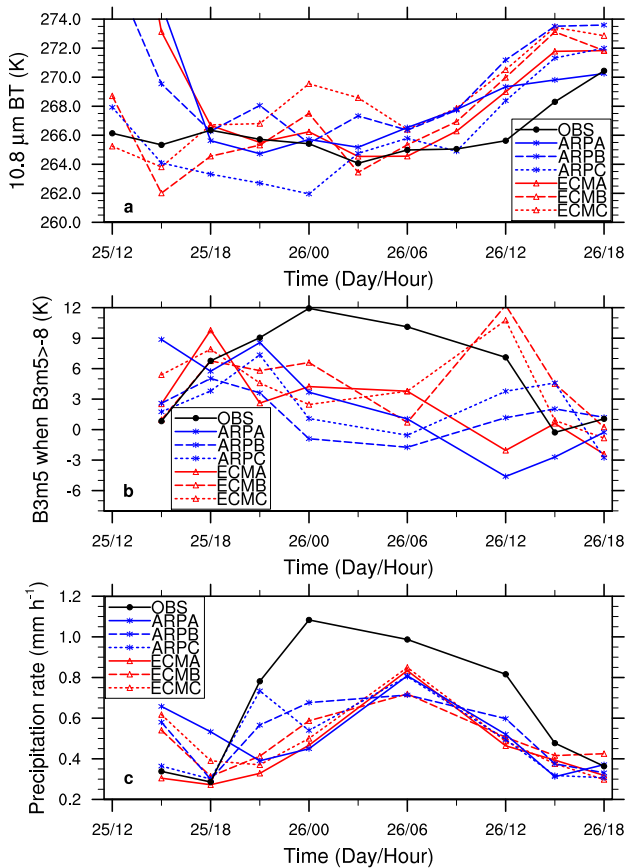


Fig. 4. Time series of mean (a) $10.8\mu\text{m}$ BT (K), (b) B3m5 (K) when B3m5 is larger than -8 K, and (c) instantaneous precipitation (mm h^{-1}). Fields were averaged over the domain shown in Fig. 3.

value less than 988 hPa. The other forecasts did not reach the recorded MSLP minimum value and failed to predict a tropical-like storm, as they did not develop a warm core (Chaboureau et al., 2012).

The track is the other important characteristic in a cyclone forecast. It is shown with the cloud cover and DCT in a radius of 80 km centred on the medicane, every 3 h from 15:00 UTC 25 September to 18:00 UTC, 26 September (Fig. 3). Here, the threshold of BT less than 250 K was chosen to diagnose deep, high clouds from the MSG infrared window channel (as in Argence et al., 2009). Deep, high clouds and DCT were first observed at 21:00 UTC, 25 September (Claud et al., 2010) as the medicane moved close to the southeastern tip of Sicily, where the orography acted as a trigger for deep convection. From 21:00 UTC onwards, deep, high clouds were observed in the vicinity of the medicane until its final landing over Italy at 18:00 UTC, 26 September, while DCT was diagnosed until 12:00 UTC, 26 September. All the forecasts showed a track close to the one observed on 25 September but they diverged substantially as the medicane moved over the Adriatic Sea. Interestingly, differences among forecasts occurred much earlier in terms of cloud cover and deep

convective activity than they did in terms of deepening and track. For example, ARPC and ECMC showed much less deep convective activity than the other forecasts during the night of 25 September. Consistently with a cyclone that did not deepen enough after 06:00 UTC, 26 September, ARPB produced less deep, high clouds than the three successful forecasts in the final 12 h. Because deep convection is the main mechanism in the deepening of a medicane, the larger the activity of clouds and convection, the more realistic the deepening and track. This result suggests that the forecast of such an intense mesocyclone could be verified in near real time just by examining its cloud cover and deep convective activity. In that particular case, the forecasts of the medicane from ARPC and ECMC could have been discarded from 25 September afternoon, as could the one from ARPB after 06:00 UTC, 26 September only. Note also that the too-large convective activity in ARPA at 21:00 UTC, 25 September would lead one to consider the forecast carefully.

A quantitative assessment is provided with the time evolution of $10.8\mu\text{m}$ BT, B3m5 and precipitation in Fig. 4, these cloud- and rain-related fields being averaged over the domain shown in Fig. 3. The observed $10.8\mu\text{m}$ BT equals 266 K most of the time and increases in the last 6 h up to 270 K. The forecasted $10.8\mu\text{m}$ BTs agree rather well with the observation, with less than 4 K of difference. An exception was the first few hours of the forecasts that started the latest (ECMA, ARPA, and ARPB), which showed BTs larger than 274 K due to the cloud spinup. As observed with B3m5, the largest convective activity occurred between 18:00 UTC, 25 September and 12:00 UTC, 26 September. Comparison with B3m5 suggests that none of the forecasts produced enough moderate to large convective rain on the morning of 26 September. Rain retrievals from AMSU observation also indicate that the model produced too little rain, but with relatively good timing, all the forecasts producing a maximum of rain on 06:00 UTC on 26 September, i.e., with only a 6-h delay relative to the observed peak. At 12:00 UTC on 26 September, ECMA and ECMC overestimated the B3m5 signal while they underestimated rainfall. Such an apparent contradiction could be partly attributed to a misrepresentation of the radiative properties of snow. Previous studies have shown that Meso-NH forecasts for convective situations are able to simulate the correct microwave BT signal in presence of a large graupel content (Wiedner et al., 2004). However, a relative disagreement was found at frequencies higher than 90 GHz for cases in which the depressed signal was mainly due to large amounts of snow (Meirolid-Mautner et al., 2007).

Overall, the comparison with averaged fields showed neither systematic bias nor particular outliers in the forecast ensemble. Almost the same average amount of rain and $10.8\mu\text{m}$ BT were forecasted. A larger spread was found for B3m5 as gridpoints with B3m5 > -8 K were less numerous than for the other variables. Note that the forecasts with the mean B3m5 closest to the observations (e.g. ECMC at 12:00 UTC, 26 September) were not necessarily the most successful ones

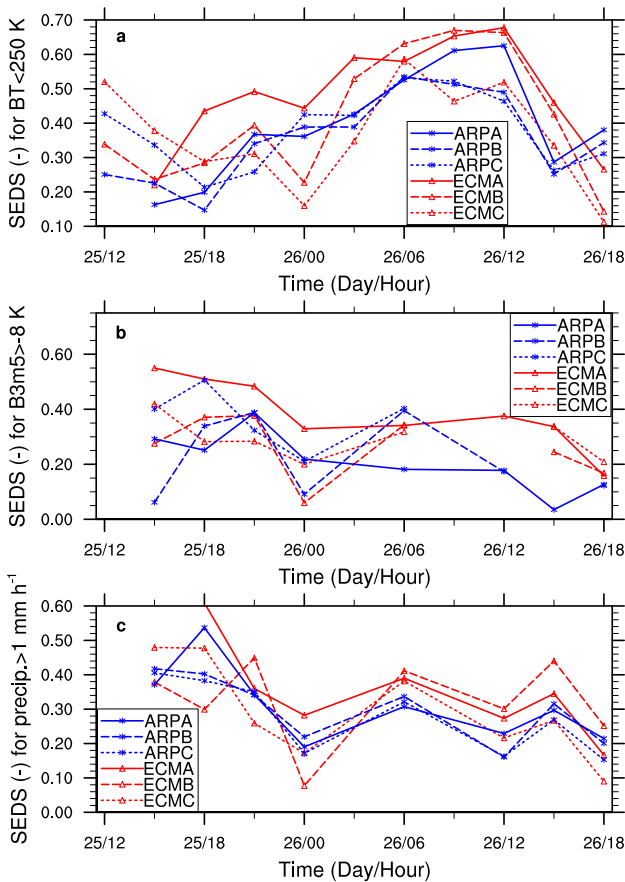


Fig. 5. Time series of SEDS for (a) 10.8 μm BT less than 250 K, B3m5 larger than -8 K and (c) instantaneous precipitation greater than 1 mm h^{-1} . SEDS scores were calculated over the domain shown in Fig. 3.

in terms of deepening and track. This is partly because the averaging was done over the domain shown in Fig. 3 that encompasses the medicane and the thunderstorms ahead the upper-level trough. As a consequence, this result contrasts with the ability to clearly distinguish the most successful forecasts from the others by looking at the cloud cover and deep convective activity in the vicinity of the cyclone (as seen with Fig. 3). It is therefore important to use a measure of skill to fully characterize the performance of forecasts.

To complement the examination of the averaged fields, a categorical score quantified the ability of the model to forecast a meteorological event at the right location. Here, we applied the SEDS score to 10.8 μm BT, B3m5 and precipitation with thresholds of 250 K, -8 K and 1 mm h^{-1} , respectively (Fig. 5). Larger SEDS for deep, high cloud were generally obtained for the three forecasts that were most successful in terms of deepening and track. This was particularly true for the morning of the 26 September, when the cyclone rapidly developed into a medicane. Comparisons against AMSU observations, based on either the B3m5 diagnostic or the rain

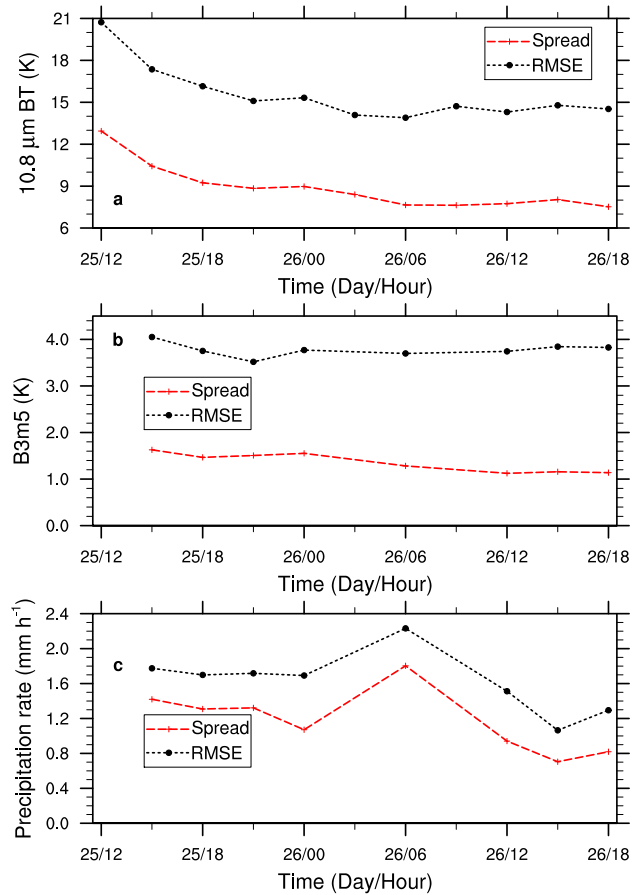


Fig. 6. Time series of spread and RMSE for (a) 10.8 μm BT, (b) B3m5 and (c) instantaneous precipitation. Fields were calculated over the domain shown in Fig. 3.

retrieval, also showed higher scores for the most successful forecast than for the others, but with smaller differences in score between forecasts. The time lag and missing data in AMSU observations made the comparison a little uncertain (the time window used here was 3 h). Note that some values of SEDS were undefined as no hit was forecasted (e.g. on 26 September, 12:00 UTC for ARPC and ECMC). In conclusion, these results obtained for cloud and rain are consistent with those on track and deepening. The best forecasts are those starting on 25 September, 12:00 UTC. A better prediction of cloud cover and rainfall is generally associated with a better prediction of deepening and track. Consequently, in the absence of any ground based data, a verification of forecasts of cyclone prediction could be based on satellite diagnostics only.

The time-lagged initialization technique shows a high sensitivity of the medicane forecast to initial conditions. Similar results in terms of deepening and track were obtained using different models and analyses by Davolio et al. (2009): forecasts starting at 12:00 UTC on 25 September were more successful than the forecasts starting at 00:00 UTC. The use

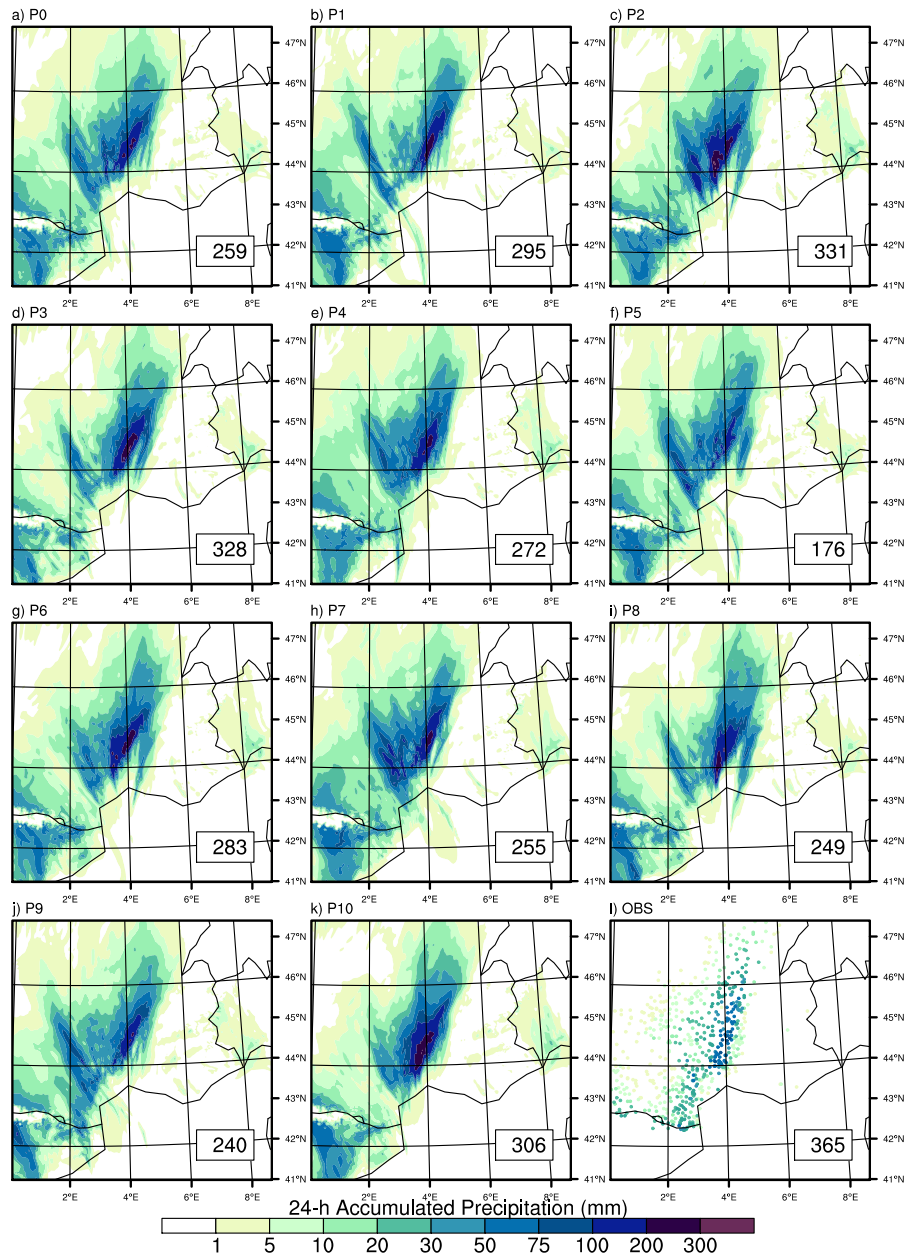


Fig. 7. Accumulated precipitation (mm) between 12:00 UTC, 1 November and 12:00 UTC, 2 November 2008 from (a–k) AROME simulations and (i) rain gauges (over France only). The number in the white box gives the maximum precipitation in (mm).

of ensemble scores allows us to look at another aspect of the multi-analysis ensemble: its spread. The averaged root-mean square error (RMSE) between forecasts and observation and the spread among the forecasts are shown for $10.8\ \mu\text{m}$ BT, B3m5 and precipitation rate in Fig. 6. For the three quantities, spread is always lower than RMSE. In particular, the ratio between spread and RMSE is about 0.5 for $10.8\ \mu\text{m}$ BT and B3m5. This is much less than the ratio for precipitating rate (about 0.8), which accounts for numerous zero values corresponding to non-raining grid points. In other words,

the satellite diagnostics point out the underdispersion of the multi-analysis ensemble more strongly than the rain rate.

4 The heavy precipitating event

The case of 1–2 November 2008 was a convective system developing in a quasi-stationary frontal system associated with a trough over western France. Such an event is typical of the Mediterranean area because moderate rain is often linked to upper-level systems, as evidenced by satellite observations

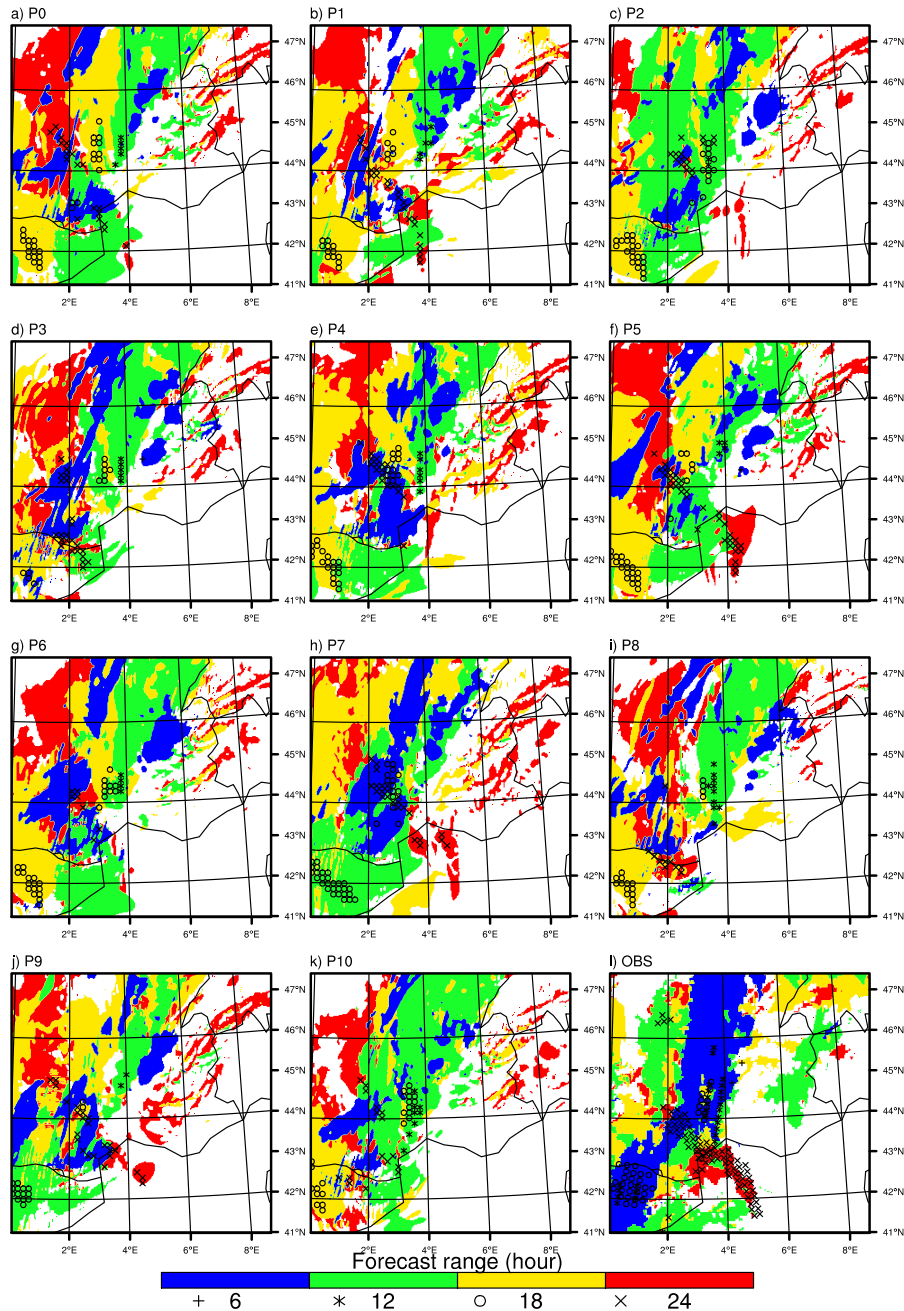


Fig. 8. Forecast range (hour) associated with SEVIRI 10.8 μm BT less than 225 K (shading) and AMSU-B DCT (symbols) from 18:00 UTC, 1 November to 12:00 UTC, 2 November 2008 for (a–k) AROME simulations and (i) MSG and AMSU-B observations.

(Chaboureau and Claud, 2006; Funatsu et al., 2008). It was characterized by a strong upper-level synoptic scale circulation, an intense low-level jet bringing moist, unstable air to the Massif Central foothills and large uncertainties on the global forecast (Vié et al., 2011). Rainfall accumulated in 24 h mostly along a southwest–northeast line (Fig. 7). Larger amounts were recorded over the southern side of the Massif Central, up to 365 mm. All forecasts showed significant rain

amounts around the Pyrenees and the foothills of the Massif Central along the same line as observed. There were obvious differences in precipitation between the ensemble members, in location as well as in intensity. The forecasts differed on the maximum amount of rain, ranging from 176 mm for P5 to 331 mm for P2. Because of the formation of some thunderstorms over the Gulf of Lions, some forecasts (P1, P5, P7, P9) produced rain over the sea while others did not. The skill

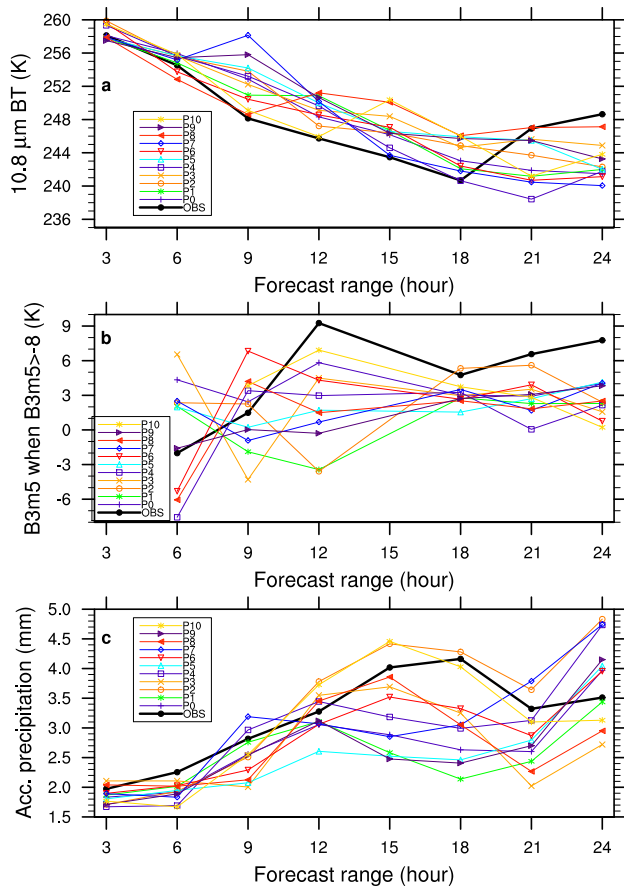


Fig. 9. Time series of mean (a) $10.8\ \mu\text{m}$ BT (K), (b) B3m5 (K) when B3m5 is larger than -8 K, and (c) 3-h accumulated precipitation (mm). Fields were averaged over the domain shown in Fig. 8.

of the forecasts over sea cannot however be verified with rain gauge measurements.

The history of the rain event is summarized by the 6-hourly time associated with deep convective clouds and DCT based on satellite diagnostics (Fig. 8). Deep convective clouds were observed on the upper-left part of the domain in the first 12-h range, then around the Gulf of Lions with an enhanced convective intensity, as diagnosed by DCT. All the ensemble members forecasted deep convective activity mostly organized into a southwest–northeast line in the first 12-h range. They differed mainly in the next 12-h range. For example, only P1, P5, P7 and P9 produced convective rain over the sea in agreement with the observation. Some forecasts, like P5 and P9, showed little convective activity over the Cevennes (around 44° N, 4° E), which resulted in an underestimation of the 24-h accumulated rainfall maximum.

The time evolution of observed $10.8\ \mu\text{m}$ BT, B3m5, and 3-h accumulated precipitation showed an increase in cloud and rain activity with the forecast range (Fig. 9). The minimum in the $10.8\ \mu\text{m}$ BT and the maximum in accumulated precipitation were both achieved at 18-h forecast range, while

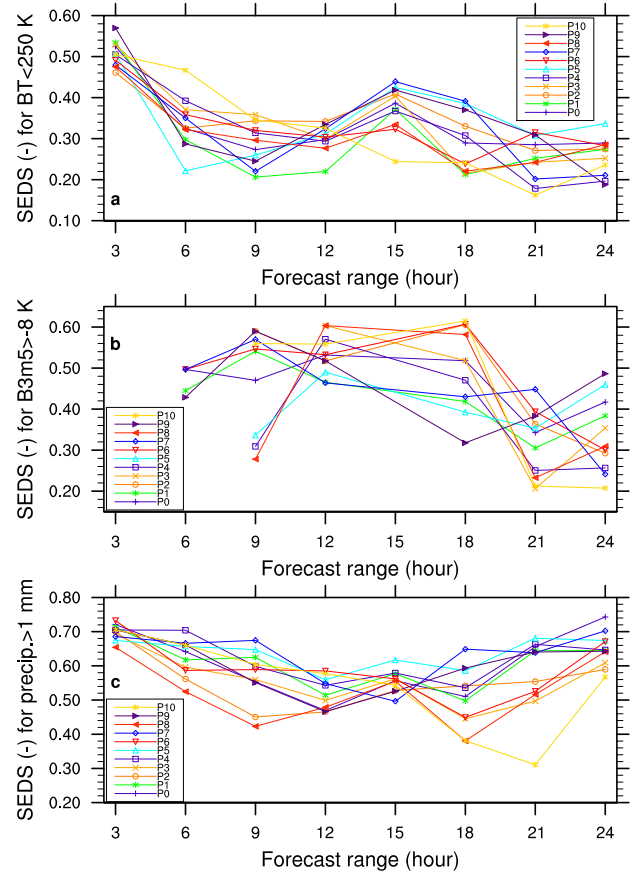


Fig. 10. Time series of SEDS for (a) $10.8\ \mu\text{m}$ BT less than 250 K, (b) B3m5 larger than -8 K and (c) 3-h accumulated precipitation greater than 1 mm. SEDS scores were calculated over the domain shown in Fig. 8.

B3m5 attained its maximum at 12-h forecast range. The ensemble members captured the overall evolution of the three variables in time and intensity. All the forecasts tended to underestimate the cloud cover until the 18-h range, however. This is consistent with the underestimation of both B3m5 and the accumulated precipitation. In the last 6-h period, they all overestimated the cloud cover and some members overestimated the 3-h accumulated precipitation. While the forecasts agreed with each other well in the first 12-h range, they were much more dispersive afterwards. A spread such as seen here for all three variables was interpreted as the impact of the PEARP lateral boundary conditions on the AROME ensemble forecasts (Vié et al., 2011).

The relatively large dispersion in the averaged cloud- and rain-related quantities between the AROME ensemble forecasts was also found on the associated SEDS scores (Fig. 10). The thresholds of 250 K, -8 K and 1 mm were applied to $10.8\ \mu\text{m}$ BT, B3m5 and accumulated precipitation, respectively. Some members had higher scores like P5, P7 and P9 on $10.8\ \mu\text{m}$ BT and precipitation at the 15-h and 18-h

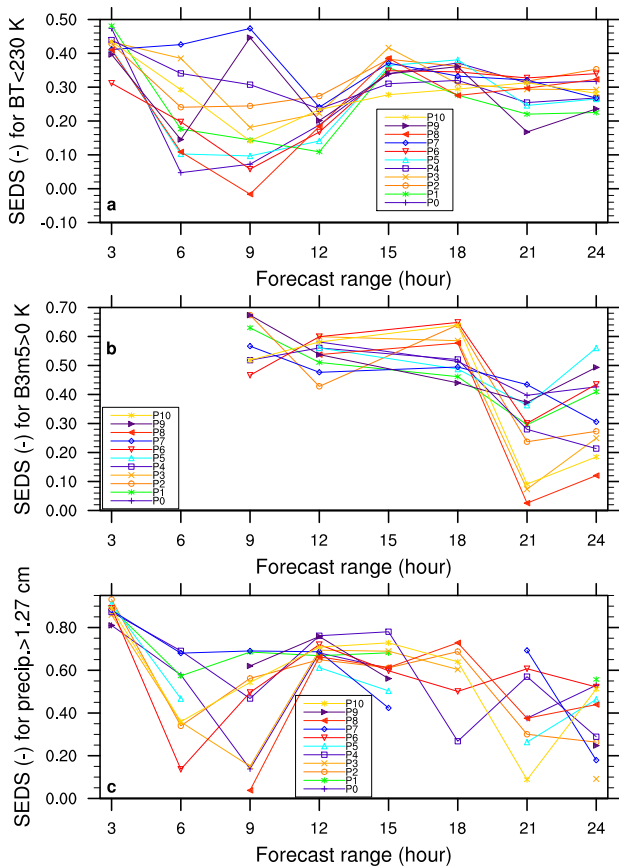


Fig. 11. Time series of SEDS for (a) 10.8 μm BT less than 230 K, (b) B3m5 larger than 0 K and (c) 3-h accumulated precipitation greater than 1.27 cm. SEDS scores were calculated over the domain shown in Fig. 8.

forecast ranges. These members were, however, characterized by smaller SEDS for B3m5 at the 18-h forecast range. This lower skill in predicting moderate to convective rain explained the smallest maximum in 24-h accumulated precipitation previously noted. In contrast, P2 and P10 fitted the time evolution of accumulated precipitation well but showed the lowest SEDS for precipitation from the 15-h forecast range onward (and, in the case of P10, for the 10.8 μm BT).

SEDS was also examined for the same quantities, but using thresholds corresponding to deeper clouds and heavier precipitation. The thresholds of 230 K, 0 K and 1.27 cm (0.5 inches) were applied to 10.8 μm BT, B3m5 and accumulated precipitation, respectively (Fig. 11). In the first 9-h range, the small-scale, deep convective activity yielded a large variation in SEDS between the forecasts for cloud cover and rain. In the absence of any hit for B3m5, SEDS is undefined for all the forecasts in the first 6-h range. From 12- to 15-h ranges, all the members performed rather similarly, but this relative agreement broke down afterward. Note that the skill of P10 was poor for B3m5 and precipitation, but at the same level as the other forecasts for the 10.8 μm BT. This un-

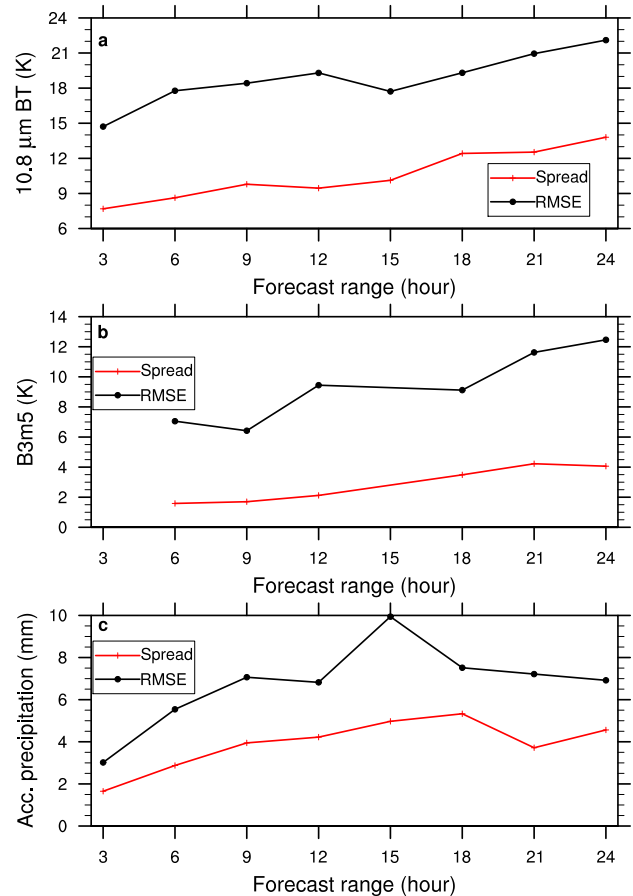


Fig. 12. Time series of spread and RMSE for (a) 10.8 μm BT, (b) B3m5 and (c) 3-h accumulated precipitation. Fields were calculated over the domain shown in Fig. 8.

derlines the advantage of using satellite diagnostics sensing the content of clouds rather than the temperature at their tops.

In summary, consistent results were obtained for the cloud and rain variables: none of the forecasts performed very well in predicting the cloud cover and rain at the right location and time. This suggests that the ensemble did not sample all the sources of uncertainty. As for the medicane, the underdispersion of the ensemble forecasts is illustrated with the averaged RMSE between forecasts and observation and the spread among the forecasts (Fig. 12). Whatever the cloud- and rain-related quantities, spread is always lower than RMSE. This confirms the consistency in the results obtained for the cloud cover and rain. The ratio between spread and RMSE varies between 0.3 and 0.7 and, as for the medicane, it is for 10.8 μm BT and B3m5 that they are the lowest, suggesting that the satellite diagnostics is more sensitive to the underdispersion of the ensemble.

The underdispersion of the ensemble forecasts can be further illustrated using a rank histogram applied to satellite observations. The rank histogram (Talagrand et al., 1997) is a

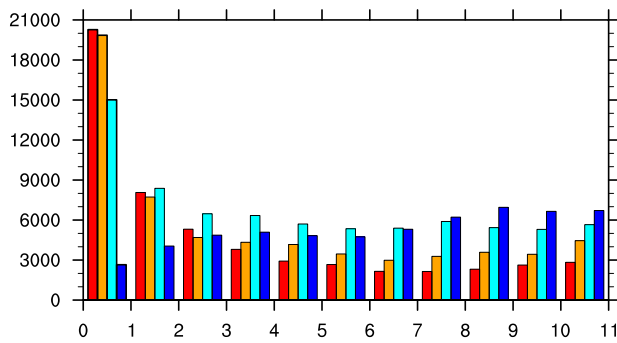


Fig. 13. Rank histograms for the $10.8\ \mu\text{m}$ BT at forecast ranges of 6 h (red boxes), 12 h (orange), 18 h (cyan) and 24 h (blue).

commonly used diagnostic for ensemble forecasts of a scalar variable x . At every time and location, the N -member forecast ensemble gives N forecast values of x , which define $N + 1$ intervals. The verifying observation is then ranked in ascending order within each group of $N + 1$ intervals, and a histogram is constructed from these ranks. The ideal result would be that the rank histogram was flat, or uniform, which would be suggestive of the ensemble prediction system performing well with respect to representing the forecast uncertainty. In contrast to precipitation for which values are often null, the $10.8\ \mu\text{m}$ BT generally shows a distribution in a U-shape (see Söhne et al., 2008, for examples of BT distribution in convective areas). This renders the building of the rank histogram for the $10.8\ \mu\text{m}$ BT easier, as shown in Fig. 13, for the 6-hourly forecasts. For forecast ranges of 6 and 12 h, the verifying observations fall in the lowest interval. This indicates that forecasts are biased toward values larger than observed, in agreement with the underestimation of the cloud cover previously noted. At 18 h the bias reduced, and at 24 h range the histogram is rather flat, indicating a certain degree of the reliability of the ensemble. Although calculated for one forecast day only, this result is consistent with the improvement in the histogram shape with the forecast range found by Vié et al. (2011) for the 925-hPa wind speed over a full month of forecasts.

5 Conclusions

The verification of two ensemble forecasts done by convective-permitted models at kilometre scale has been considered: a tropical-like storm over southern Italy and a heavy precipitation event over southern France. The ensemble forecasts were verified following two paths. First, forecasts were evaluated against ground based measurements, i.e. track and deepening for the medicane and rainfall for the heavy precipitation event. Second, the verification was achieved for the first time using the model-to-satellite approach and satellite diagnostics dedicated to cloud and rain fields. These variables are among the most difficult weather

variables to predict and to verify. It is therefore important to develop methodology to monitor the performance of numerical weather prediction systems in a systematic way.

For both cases, overall consistency was found between the traditional and model-to-satellite approaches. Thus, in the absence of any ground based data, forecasts of cloudy weather events can be verified using satellite observations only. This shows that the model-to-satellite approach is a useful tool for the verification of forecasts over sea or other data-sparse areas. Moreover, results from the tropical-like storm show some member forecasts that failed to predict a fully developed medicane, where overpredicted MSLP values showed a lack of deep convective activity. This suggests that forecasts can be verified in near real time just by comparing predicted and observed BT fields using pertinent satellite diagnostics. Such an ability to quickly evaluate the quality of the cloud forecasts produced by numerical weather prediction models can be essential for short-range forecasting. A further application dedicated to the uncertainty sampling of ensemble forecasts was shown for the medicane and the heavy precipitation event. For both cases, a too less dispersive forecast ensemble was found. This was partly because the Meso-NH and AROME ensemble forecasts used here were designed to investigate the uncertainties on initial and lateral boundary conditions only, respectively. The development of an ensemble approach that would enlarge the ensemble spread by combining uncertainties from initial and lateral-boundary conditions and model errors is currently under investigation.

Acknowledgements. We thank the editor and the anonymous referees for their constructive comments. This work was carried out in the framework of the MEDUP project (grant ANR-07-VULN-06-001), funded by the “Vulnérabilité Milieux et Climat (VMC)” programme of ANR. Computer resources were allocated by GENCI (project 0569). Satellite observations were provided by SATMOS, ICARE and the NOAA’s Comprehensive Large Array-data Stewardship System.

Edited by: M.-C. Llasat

Reviewed by: two anonymous referees



The publication of this article is financed by CNRS-INSU.

References

- Argence, S., Lambert, D., Richard, E., Chaboureau, J.-P., and Söhne, N.: Impact of initial conditions uncertainties on the predictability of heavy rainfall in the Mediterranean: A case study, *Quart. J. Roy. Meteor. Soc.*, 134, 1775–1788, 2008.
- Argence, S., Lambert, D., Richard, E., Chaboureau, J.-P., Arbogast, P., and Maynard, K.: Improving the numerical prediction of a cyclone in the Mediterranean by local potential vorticity modifications, *Quart. J. Roy. Meteor. Soc.*, 135, 865–879, 2009.
- Baran, A. J. and Francis, P. N.: On the radiative properties of cirrus cloud at solar and thermal wavelengths: a test of model consistency using high-resolution airborne radiance measurements, *Quart. J. Roy. Meteor. Soc.*, 130, 763–778, 2004.
- Bauer, P.: Including a melting layer in microwave radiative transfer simulations for clouds, *Atmos. Res.*, 57, 1619–1632, 2001.
- Chaboureau, J.-P. and Bechtold, P.: A simple cloud parameterization derived from cloud resolving model data: diagnostic and prognostic applications, *J. Atmos. Sci.*, 59, 2362–2372, 2002.
- Chaboureau, J.-P. and Bechtold, P.: Statistical representation of clouds in a regional model and the impact on the diurnal cycle of convection during Tropical Convection, Cirrus and Nitrogen Oxides (TROCCINOX), *J. Geophys. Res.*, 110, D17103, doi:10.1029/2004JD005645, 2005.
- Chaboureau, J.-P. and Claud, C.: Satellite-based climatology of Mediterranean cloud systems and their association with large-scale circulation, *J. Geophys. Res.*, 111, D01102, doi:10.1029/2005JD006460, 2006.
- Chaboureau, J.-P., Cammas, J.-P., Mascart, P., Pinty, J.-P., Claud, C., Roca, R., and Morcrette, J.-J.: Evaluation of a cloud system life-cycle simulated by Meso-NH during FASTEX using METEOSAT radiances and TOVS-3I cloud retrievals, *Quart. J. Roy. Meteor. Soc.*, 126, 1735–1750, 2000.
- Chaboureau, J.-P., Cammas, J.-P., Mascart, P., Pinty, J.-P., and Lafore, J.-P.: Mesoscale model cloud scheme assessment using satellite observations, *J. Geophys. Res.*, 107, 4301, doi:10.1029/2001JD000714, 2002.
- Chaboureau, J.-P., Söhne, N., Pinty, J.-P., Meirold-Mautner, I., Defer, E., Prigent, C., Pardo, J.-R., Mech, M., and Crewell, S.: A midlatitude cloud database validated with satellite observations, *J. Appl. Meteor. Climatol.*, 47, 1337–1353, 2008.
- Chaboureau, J.-P., Pantillon, F., Lambert, D., Richard, E., and Claud, C.: Tropical transition of a Mediterranean storm by jet crossing, *Quart. J. Roy. Meteor. Soc.*, 137, 596–611, <http://dx.doi.org/10.1002/qj.960>, 2012.
- Chevallier, F., Bauer, P., Kelly, G., Jakob, C., and McNally, T.: Model clouds over oceans as seen from space: comparison with HIRS/2 and MSU radiances, *J. Climate*, 14, 4216–4229, 2001.
- Clark, H. and Chaboureau, J.-P.: Uncertainties in short-term forecasts of a Mediterranean heavy precipitation event: Assessment with satellite observations, *J. Geophys. Res.*, 115, D22213, doi:10.1029/2010JD014388, 2010.
- Claud, C., Alhammoud, B., Funatsu, B. M., and Chaboureau, J.-P.: Mediterranean hurricanes: large-scale environment and convective and precipitating areas from satellite microwave observations, *Nat. Hazards Earth Syst. Sci.*, 10, 2199–2213, doi:10.5194/nhess-10-2199-2010, 2010.
- Claud, C., Alhammoud, B., Funatsu, B. M., Lebeaupin Brossier, C., Chaboureau, J.-P., Béranger, K., and Drobinski, P.: A high resolution climatology of precipitation and deep convection over the Mediterranean region from operational satellite microwave data: development and application to the evaluation of model uncertainties, *Nat. Hazards Earth Syst. Sci.*, 12, 785–798, doi:10.5194/nhess-12-785-2012, 2012.
- Cuxart, J., Bougeault, P., and Redelsperger, J.-L.: A turbulence scheme allowing for mesoscale and large-eddy simulations, *Quart. J. Roy. Meteor. Soc.*, 126, 1–30, 2000.
- Davolio, S., Miglietta, M. M., Moscatello, A., Pacifico, F., Buzzi, A., and Rotunno, R.: Numerical forecast and analysis of a tropical-like cyclone in the Ionian Sea, *Nat. Hazards Earth Syst. Sci.*, 9, 551–562, doi:10.5194/nhess-9-551-2009, 2009.
- Ebert, E. E., Damrath, U., Wergen, W., and Baldwin, M. E.: The WGNE assessment of short-term quantitative precipitation forecasts, *Bull. Amer. Meteor. Soc.*, 84, 481–492, 2003.
- Funatsu, B. M., Claud, C., and Chaboureau, J.-P.: Potential of Advanced Microwave Sounding Unit to identify precipitating systems and associated upper-level features in the Mediterranean region: Case studies, *J. Geophys. Res.*, 112, D17113, doi:10.1029/2006JD008297, 2007.
- Funatsu, B. M., Claud, C., and Chaboureau, J.-P.: A 6-year AMSU-based climatology of stratospheric intrusions and associated precipitation distribution in the Mediterranean region, *J. Geophys. Res.*, 113, D15120, doi:10.1029/2008JD009918, 2008.
- Funatsu, B. M., Claud, C., and Chaboureau, J.-P.: Comparison between the large-scale environment of moderate and intense precipitating systems in the Mediterranean region, *Mon. Weather Rev.*, 137, 3933–3959, 2009.
- Grasso, L. D., Sengupta, M., Dostalek, J. F., Brummer, R., and Demaria, M.: Synthetic satellite imagery for current and future environmental satellites, *Int. J. Remote Sens.*, 29, 4373–4384, 2008.
- Hogan, R. J., O'Connor, E. J., and Illingworth, A. J.: Verification of cloud-fraction forecasts, *Quart. J. Roy. Meteor. Soc.*, 135, 1494–1511, 2009.
- Hu, Y. X. and Stamnes, K.: An accurate parameterization of the radiative properties of water clouds suitable for use in climate models, *J. Climate*, 6, 728–742, 1993.
- Lafore, J.-P., Stein, J., Asencio, N., Bougeault, P., Ducrocq, V., Duron, J., Fischer, C., Hérelil, P., Mascart, P., Masson, V., Pinty, J.-P., Redelsperger, J.-L., Richard, E., and Vilà-Guerau de Arellano, J.: The Meso-NH Atmospheric Simulation System, Part I: adiabatic formulation and control simulations. Scientific objectives and experimental design, *Ann. Geophys.*, 16, 90–109, 1998, <http://www.ann-geophys.net/16/90/1998/>.
- Laviola, S. and Levizzani, V.: The 183-WSL fast rain rate retrieval algorithm: Part I: Retrieval design, *Atmos. Res.*, 99, 443–461, 2011.
- Laviola, S., Miglietta, M. M., and Levizzani, V.: Satellite and numerical model investigation of two mesoscale convective systems over central Mediterranean, *J. Hydrometeorol.*, 12, 634–649, 2011.
- Masson, V., Champeaux, J.-L., Chauvin, C., Meriguet, C., and Lacaze, R.: A global database of land surface parameters at 1 km resolution for use in meteorological and climate models, *J. Climate*, 16, 1261–1282, 2003.
- Meirold-Mautner, I., Prigent, C., Defer, E., Pardo, J. R., Chaboureau, J.-P., Pinty, J.-P., Mech, M., and Crewell, S.: Radiative transfer simulations using mesoscale cloud model outputs: comparisons with passive microwave and infrared satellite observations for mid-latitudes, *J. Atmos. Sci.*, 64, 1550–1568, 2007.

- Molteni, F., Buizza, R., Palmer, T. N., and Petroliagis, T.: The ECMWF ensemble prediction system: Methodology and validation, *Quart. J. Roy. Meteor. Soc.*, 122, 73–119, 1996.
- Morcrette, J.-J.: Evaluation of model-generated cloudiness: satellite-observed and model-generated diurnal variability of brightness temperature, *Mon. Weather Rev.*, 119, 1205–1224, 1991.
- Moscattello, A., Miglietta, M. M., and Rotunno, R.: Observational analysis of a Mediterranean “hurricane” over southeastern Italy, *Weather*, 63, 306–311, 2008.
- Otkin, J. A. and Greenwald, T. J.: Comparison of WRF model-simulated and MODIS-derived cloud data, *Mon. Weather Rev.*, 136, 1957–1970, 2008.
- Otkin, J. A., Greenwald, T. J., Sieglaff, J., and Huang, H. L.: Validation of a large-scale simulated brightness temperature dataset using SEVIRI satellite observations, *J. Appl. Meteor. Climatol.*, 48, 1613–1626, 2009.
- Pantillon, F., Chaboureau, J.-P., Lac, C., and Mascart, P.: On the role of a Rossby wave train during the extratropical transition of hurricane Helene (2006), *Quart. J. Roy. Meteor. Soc.*, in press, doi:10.1002/qj.1974, 2012.
- Pergaud, J., Masson, V., Malardel, S., and Couvreux, F.: A parameterization of dry thermals and shallow cumuli for mesoscale numerical weather prediction, *Bound. Lay. Meteorol.*, 132, 83–106, 2009.
- Pinty, J.-P. and Jabouille, P.: A mixed-phase cloud parameterization for use in a mesoscale non-hydrostatic model: simulations of a squall line and of orographic precipitations, in: *Conf. on cloud physics*, Everett, WA, Amer. Meteor. Soc., 217–220, 1998.
- Saunders, R., Matricardi, M., Brunel, P., English, S., Bauer, P., O’Keeffe, U., Francis, P., and Rayer, P.: RTTOV-8 Science and validation report, Tech. rep., NWP SAF Rep., 41 pp., 2005.
- Seity, Y., Brousseau, P., Malardel, S., Hello, G., Bénard, P., Bouttier, F., Lac, C., and Masson, V.: The AROME-France convective-scale operational model, *Mon. Weather Rev.*, 139, 976–991, 2011.
- Simmons, A. J. and Hollingsworth, A.: Some aspects of the improvement in skill of numerical weather prediction, *Quart. J. Roy. Meteor. Soc.*, 128, 647–677, 2002.
- Söhne, N., Chaboureau, J.-P., and Guichard, F.: Verification of cloud cover forecast with satellite observation over West Africa, *Mon. Weather Rev.*, 136, 4421–4434, 2008.
- Talagrand, O., Vautard, R., and Strauss, B.: Evaluation of probabilistic prediction systems, in: *Proc. ECMWF Workshop on Predictability*, Reading, United Kingdom, ECMWF, 1–25, 1997.
- Tapiador, F., Tao, W., Shi, J., Angelis, C., Martinez, M., Marcos, C., Rodriguez, A., and Hou, A.: A comparison of perturbed initial conditions and multiphysics ensembles in a severe weather episode in Spain, *J. Appl. Meteor. Climatol.*, 51, 489–504, 2012.
- Vich, M., Romero, R., and Brooks, H. E.: Ensemble prediction of Mediterranean high-impact events using potential vorticity perturbations, Part I: Comparison against the multiphysics approach, *Atmos. Res.*, 102, 227–241, 2011.
- Vié, B., Nuissier, O., and Ducrocq, V.: Cloud-resolving ensemble simulations of Mediterranean heavy precipitating events: uncertainty on initial conditions and lateral boundary conditions, *Mon. Weather Rev.*, 139, 403–423, 2011.
- Wiedner, M., Prigent, C., Pardo, J. R., Nuissier, O., Chaboureau, J.-P., Pinty, J.-P., and Mascart, P.: Modeling of passive microwave responses in convective situations using outputs from mesoscale models: comparison with TRMM/TMI satellite observations, *J. Geophys. Res.*, 109, D06214, doi:10.1029/2003JD004280, 2004.