

Objective evaluation of mesoscale simulations of the Algiers 2001 flash flood by the model-to-satellite approach

N. Söhne, J.-P. Chaboureau, S. Argence, D. Lambert, and E. Richard

Laboratoire d'Aérodynamique UMR 5560 (UPS/CNRS), Toulouse, France

Received: 26 October 2005 – Revised: 16 December 2005 – Accepted: 19 December 2005 – Published: 14 March 2006

Abstract. An objective evaluation of mesoscale simulations by the model-to-satellite approach is performed. The model-to-satellite approach consists in calculating brightness temperatures (BT) from model variables with a radiative transfer code. It allows to compare directly and quantitatively simulations and observations by calculating statistical scores. This method is detailed and used herein to objectively evaluate an ensemble of Meso-NH simulations of the Algiers 2001 flash flood. In particular, the improvement due to the grid-nesting is shown.

1 Introduction

Current simulation ensemble experiments aim at quantifying the predictability of extreme weather events that are, until now, poorly forecasted. Accordingly, new tools for systematically evaluating mesoscale simulations are needed. Satellite observation can monitor a series of meteorological features with high space and time resolutions. The so-called model-to-satellite approach directly compares simulation and observation. It guarantees that errors come from the modeling side only (e.g. Chaboureau et al., 2002a). Furthermore, continuous and categorical statistical scores computed for a set of simulations allow to objectively classifying them. This new evaluation tool box is applied to the Algiers 2001 flash flood. This cyclone is resulting of the interaction between an upper-level trough over Spain and lower-level warm air moving north off the Sahara. Among 110 mm of rain was measured over Algiers in only three hours, between 06:00 and 09:00 UTC 10 November. An ensemble of simulations run with the French mesoscale model Meso-NH is studied. The simulated brightness temperatures (BT) calculated with a radiative transfer code are compared with METEOSAT and SSM/I observations.

Correspondence to: N. Söhne
(nathalie.sohne@aero.obs-mip.fr)

2 Data

2.1 Model

Meso-NH is a non-hydrostatic mesoscale model jointly developed by Météo-France and the Laboratoire d'Aérodynamique (Lafore et al., 1998). The microphysical scheme is a bulk mixed-phase cloud parameterization developed by Pinty and Jabouille (1998). An ensemble of 12 simulations has been done (Table 1).

These simulations vary by the initialization dates, the analyses, and by the eventual use of grid-nesting (Stein et al., 2000). Three initialization dates have been chosen: at 00:00 UTC and 12:00 UTC 9 November 2001 and at 00:00 UTC 10 November 2001. Three analyses, ARPEGE, ECMWF and modified ARPEGE, are used to initialize and force the simulation limit conditions. The modified ARPEGE analyses have been done by filling the trough over Spain at different places and for variable atmosphere thickness, thanks to a potential vorticity (PV) inversion technique, see Argence et al. (2005) for more details. For all the simulations, the (father) model domain covers Europe and North Africa (model 1) with a 50 km grid. Simulations with grid-nesting have two additional models with a 10 km grid centered on the Western Mediterranean (model 2) and a 2 km grid centered on Algiers. All results presented next are shown at 06:00 UTC 10 November for the domain of model 2.

2.2 Observations

Satellite observations offer a large spatial and temporal resolution. Each satellite channel gives an observation of a different meteorological feature. The MVIRI Infrared channel (IR) aboard METEOSAT senses the temperature emitted by the Earth and so, describes the cloud cover. The signature of high clouds, with cloud top higher than 6 km, is represented with $BT_{IR} \leq 250$ K (blue areas on Fig. 1, IR). In this case, this category includes all precipitating clouds. The MVIRI Water

Table 1. Configuration details of the simulation ensemble.

Name	Initialization	Analyses	Grid-nesting
ALG01	09 Nov 00:00 UTC	ARPEGE	no
ALG02	09 Nov 12:00 UTC	ARPEGE	no
ALG06	10 Nov 00:00 UTC	ARPEGE	no
ALG07	10 Nov 00:00 UTC	ECMWF	no
ALG11*	10 Nov 00:00 UTC	ARPEGE	2 ways
ALG13	10 Nov 00:00 UTC	ARPEGE	2 ways
DAL14	10 Nov 00:00 UTC	modif.	2 ways
ALG15	10 Nov 00:00 UTC	ARPEGE	1 way
DAL16	10 Nov 00:00 UTC	modif.	1 way
DAL02	10 Nov 00:00 UTC	modif.	no
DAL03	10 Nov 00:00 UTC	modif.	no
DAL04	10 Nov 00:00 UTC	modif.	no

* simulation using a different mixing length than other simulations

Vapor channel (WV), sensitive to the relative humidity, permits the study of the upper-level trough. The signature of the trough is defined with $BT_{WV} > 240$ K (yellow and red areas on Fig. 1, WV). The SSMI microwave frequencies used here are 37 GHz and 85.5 GHz in vertical polarization. The reference observation (07:30 UTC) is a combination from the 07:00 UTC and the 08:20 UTC images. Over ocean, the 37 GHz channel is sensitive to rain emission. Rain and cloud water have an emissivity ~ 1 and conduct to warmer BT than other free ocean which emissivity is ~ 0.5 . So, the signature of cloud water is defined with $BT_{37\text{ GHz V}} > 230$ K (along the Algerian coast on Fig. 1, 37 GHz). Over land, the emissivity is ~ 0.9 . The conditions on the 85 GHz channel are based on ice hydrometeor scattering, which makes the BT decrease strongly. So, the signature of graupels is defined with $BT_{85\text{ GHz V}} < 245$ K (blue and white areas on Fig. 1, 85 GHz).

2.3 Radiative code

The radiative transfer code RTTOV (Radiative Transfer for Tiros Operational Vertical Sounder) is maintained in the EU-METSAT NWP-SAF context. The model allows fast radiance and BT calculation for most of infrared and microwave radiometers. The latest version RTTOV-8 (Saunders and Brunel, 2004) is used here. In infrared, absorption cloud effect is taken into account with the grey body approximation. In microwaves, RTTOV takes now hydrometeor scattering into account, as well as the polarization due to the sea surface properties but scattering due to non-spherical particles is not modelised. On Fig. 1, simulated BTs are very realistic whatever the channel, in spite of a bias in the WV channel. So it is possible to objectively evaluate the simulations.

3 Scores for simulation evaluation

Category choice depends both of its physical and statistical significance. A minimal population density in each category is needed and each channel has a BT threshold that represents

a particular feature (defined Part 2.2). The bias between simulations and observations must be small in order to use the same threshold for simulations and the corresponding observations.

3.1 Continuous scores

Continuous scores measure the correspondence between the values of simulations and observations at grid-points. Correlation and the ratio of standard deviation simulated over observed, for the WV channel, are presented here. They are summarized in a Taylor diagram (Fig. 2).

Correlation for the WV channel is between 0.7 and 0.9 for all the simulations. The WV channel is well reproduced. The lowest correlation is obtained for the simulation initialized at 00:00 UTC (ALG01), then at 12:00 UTC (ALG02) 9 November, and then for the simulations initialized at 00:00 UTC 10 November. So, the sooner the initialization, the lower the correlation. The simulations initialized at 00:00 UTC 10 November present similar results, even if, correlation for the WV channel is a little larger for the one initialized with ECMWF analysis (ALG07) than the ones initialized with ARPEGE analyses (DALXX¹, ALG06, ALG1X). Note that simulations can only be distinguished for the infrared channels. In microwaves, the possible precipitating area is too small to influence these scores whatever the initialization date or analysis (figure not shown).

3.2 Categorical scores

Categorical scores measure the correspondence between simulated and observed occurrence of events at grid-points. They have been developed to focus on high precipitation rates and tornado detection (Ebert et al., 2004).

The meteorological situation is divided in, at least, two categories chosen by the user. Most often binary categories are used; the event happens or not. Comparison with observation, also divided in two categories, conducts to a 2×2 contingency table. This defines the number of hits (event is simulated and observed), false alarms (event is only simulated), misses (event is only observed), and the correct negatives (non-event in both simulation and observation). From this table many scores can be calculated. Only the Heidke Skill Score (HSS) is presented here. HSS measures the fraction of correct forecasts after eliminating those which would be correct due to chance. HSS range is $[-\infty; 1]$, and 1 is the perfect score. For the infrared channels and also the microwave channels, presented in Fig. 3, HSS is lower for the simulation initialized at 00:00 UTC 9 November (ALG01) than for the simulation initialized at 12 UTC 9 November (ALG02) and 00:00 UTC 10 November (ALG06). Again, the sooner the initialization, the lower the HSS. Moreover, as HSS presents a great variability between the simulations, it helps to evaluate more accurately the simulations initialized at the same time, 00:00 UTC 10 November, with different

¹X varies from 0 to 1 or 1 to 6. It is used to call all the simulations of a subensemble

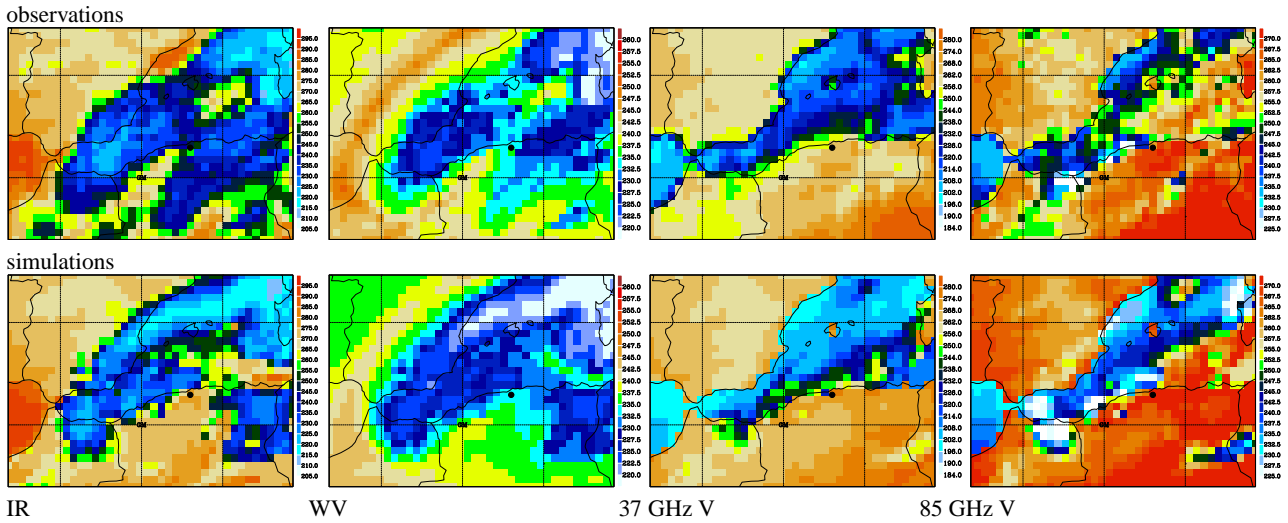


Fig. 1. From the left to the right, IR, WV, 37 GHz V, 85 GHz V channels BT (K) at 06:00 UTC 10 Nov, (top) observed by METEOSAT and SSMI, and (bottom) simulated (ALG06).

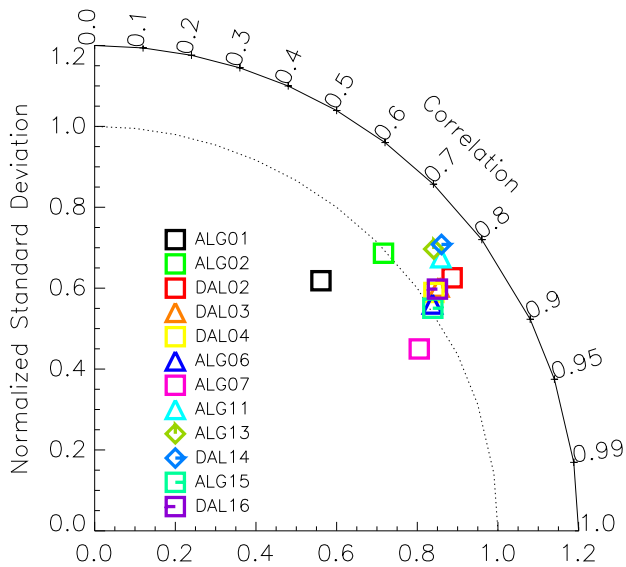


Fig. 2. Taylor diagram representing correlation and normalized standard deviation between simulation and observation at 06:00 UTC 10 Nov, on domain 2, for METEOSAT WV channel.

(ALG06, ALG07) or perturbed analyses (ALG06, DAL0X) for each feature. However, HSS variation is not the same for all the meteorological features. So, it is possible to determine what the best type of simulation for each meteorological feature is. Grid-nesting brings more variability, but continuous and categorical scores for simulations with grid-nesting (e.g. ALG11) are not always better (e.g. ALG06). This is due to the double penalty effect. This arises when at high resolution the event is more realistically simulated than at low resolution, but is misplaced. It induces that simulations at high resolution are penalized twice, first for missing an event, second for simulating it where it is not, and so producing a false alarm.

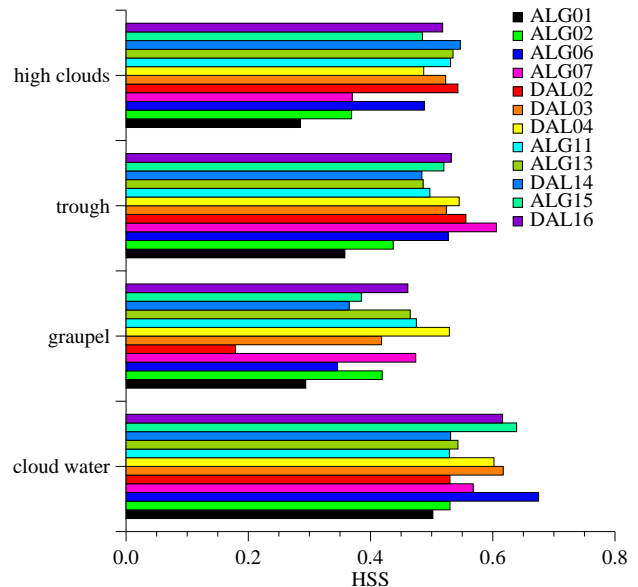


Fig. 3. HSS calculated on domain 2 with a 50 km grid at 06:00 UTC 10 Nov for the trough (WV), the high clouds (IR), the cloud water over sea (37 V) and the graupel other land (85 V) for all the simulations.

3.3 Fraction Skill Score

To compare different scale simulations the Fraction Skill Score (FSS) calculation (Roberts, 2005) is used. FSS calculation is based on fraction comparisons. For every grid square, the fraction of surrounding grid-squares within a given area that exceeds a particular threshold is calculated. FSS range is [0;1], and the perfect score is 1. When comparing simulations on their original grid, (Fig. 4a), FSS for simulations with grid-nesting (at 10 km, dotted lines) is lower than FSS for simulations without grid-nesting (50 km, solid lines).

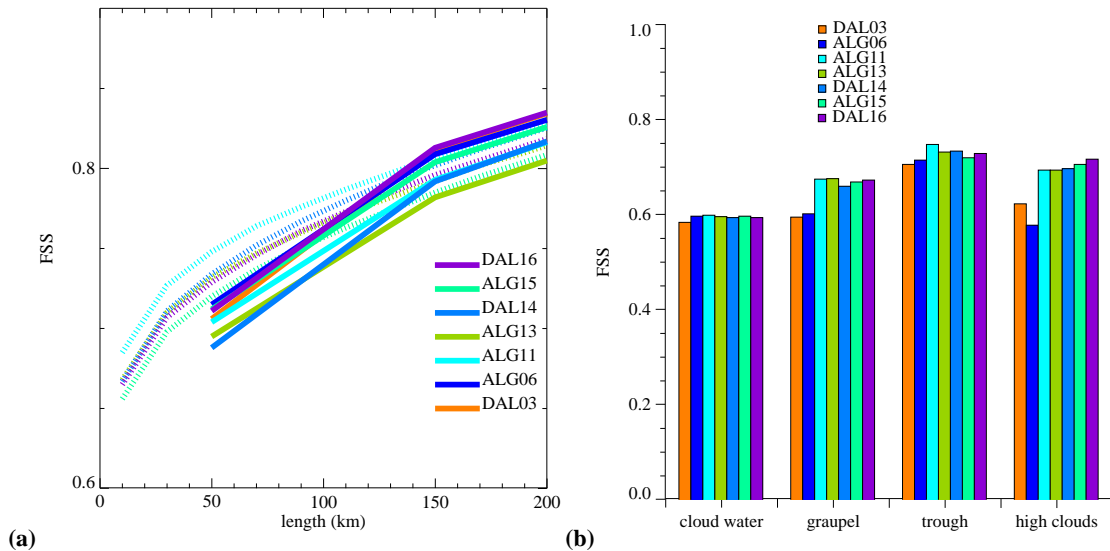


Fig. 4. FSS calculated on domain model 2 (a) with a 10 km grid (dotted lines) and a 50 km grid (solid lines) for simulations initialized at 00:00 UTC 10 November with the ARPEGE analysis for $BT_{IR} < 250$ and (b) at 50 km for all the frequencies studied.

This is an illustration of the double penalty. When comparing at a length of 50 km, simulations with grid-nesting have a better FSS than simulations without grid-nesting.

This is obtained for all the meteorological features, (Fig. 4b). Simulations with grid-nesting, the 5 bars to the right, have a better FSS than the 2 bars to the left (DAL03, ALG06) representing simulations without grid-nesting. FSS, by avoiding the double penalty, is able to show the improvement of the simulation when using high resolution.

4 Conclusions

A qualitative comparison between all the simulations and the observations shows that this event is well reproduced by the model. Then, scores with adapted thresholds to this situation and the available observations have been defined keeping in mind the physical meaning of the selection. Continuous scores allow to evaluate the whole simulation quality; simulation at different initialization date or very different analyses can be characterized. But, they are neither relevant to distinguish the simulations initialized at the same time nor to separate BT in the microwave channels whatever the date. On the contrary, categorical scores, here the HSS, focus on specific phenomena and so allow to classify specific meteorological features for all the simulations more accurately. Last, the improvement due to the grid-nesting can only be evaluated with the FSS.

Acknowledgements. The first author thanks the CNES and Météo-France to allow her making her Phd.

Edited by: V. Kotroni and K. Lagouvardos

Reviewed by: E. Heifetz

References

- Argence, S., Lambert, D., Richard, E., Söhne, N., Chaboureaud, J.-P., Crépin, F., and Arbogast, P.: High resolution sensitivity study of the Algiers 2001 flash flood to initial conditions by potential vorticity inversion, Proceedings of the 7th EGS Plinius conference on Mediterranean storms, 5–7 October 2005, Rethymon, Crete, Greece, 2005.
- Chaboureaud, 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(D17), 4310, doi:10.1029/2001JD000714, 2002.
- Ebert, E., Wilson, L. J., Brown, B. G., Nurmi, P., Brooks, H. E., Bally, J., and Jaenake, M.: Verification of Nowcasts from the WWRP Sydney 2000 Forecast Demonstration Project, *Wea. Forecast.*, 19, 73–96, 2004.
- Lafore, J. P., Stein, J., Asencio, N., et al.: The Meso-NH Atmospheric Simulation System. Part I: adiabatic formulation and control simulations, *Ann. Geophys.*, 16, 90–109, 1998.
- 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, Proceedings of the AMS conference on cloud physics, 17–21 August 1998, Everett, Wa, USA, 217–220, 1998.
- Roberts, N.: An investigation of the ability of a storm scale configuration of the Met Office NWP model to predict flood-producing rainfall, V 4.0, Forecasting Research Technical Report No. 455, Met Office, 2005.
- Saunders, R. and Brunel, P.: RTTOV-8-5 Users Guide, NWPSAF-MO-UD-008, V 1.7, 1.2, EUMETSAT, 2004.
- Stein, J., Richard, E., Lafore, J.-P., Pinty, J.-P., Asencio, N., and Cosma, S.: High-resolution non-hydrostatic simulations of flash-flood episodes with grid-nesting and ice phase parametrization, *Meteorol. Atmos. Phys.*, 72, 101–110, 2000.