Experiencing 1D-Var+nudging technique in the COSMO model

Virginia Poli, Tiziana Paccagnella, Pier Paolo Alberoni, Davide Cesari and Paolo Patruno

ARPA Emilia-Romagna, Bologna, Italy

1 Introduction

At very high resolution the use of dense and highly frequent observations become crucial in assimilation systems. In the proposed variational approach the retrieval of temperature and humidity profiles from radar derived surface rain rate is performed firstly, employing two linearized parameterizations of large-scale condensation and convection originally developed for the ECMWF model. The obtained profiles are then used as “pseudo” observations into the nudging scheme of the high resolution COSMO model.
The aim of this work is to test if the developed framework outperforms the currently running latent heat nudging (LHN hereafter) scheme which was specifically designed for the assimilation of radar rain rates.
One of the main reason for investing in this type of scheme is the rapidly decrease of the positive impact of radar data when using the LHN as progressing into the forecast as documented in Stephan et al. [8]. As suggested in this paper, a possible cause for this lack of persistence is the weak coupling between the LHN temperature adjustments and the model dynamic. The LHN effectively acts only rescaling the temperature profiles with an adjustment in the humidity field which is not consistent with the cloud scheme prediction. 1D-Var algorithm is built instead on a physically based operator which reproduces all of the processes that take place in the cloud and then should be able to vertically re-distribute in a coherent way the heat released by the rain formation process.
To assess the quality of this approach full model integrations with and without the assimilation of the retrieved profiles are finally used to quantify the impact of rain rate assimilation in improving the forecasted precipitation events.

2 1D-Var theory

In variational data assimilation, the goal is to find the optimal model state, the analysis, $\mathbf{x}_a$, that simultaneously minimizes the distance to the observations, $\mathbf{y}_0$, and a background model state, $\mathbf{x}_b$, usually coming from a previous short-range forecast. When the background and observation errors are uncorrelated and have a Gaussian distribution, then the maximum likelihood estimator of the state vector, $\mathbf{x}$, is the minimum of the following cost function

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x}-\mathbf{x}_b)^T\mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b) + \frac{1}{2}(H(\mathbf{x})-\mathbf{y})^T\mathbf{R}^{-1}(H(\mathbf{x})-\mathbf{y}) \tag{1}$$

where H is the operator simulating the observed data from the model variable $\mathbf{x}$, $\mathbf{R}$ is the observation error matrix which includes measurement errors and representativeness errors, including errors in H, and $\mathbf{B}$ is the background error covariance matrix of the state $\mathbf{x}_b$. The superscripts -1 and T denote inverse and transpose matrices, respectively.

Under the hypothesis of linearity for the observation operator (i.e. $H(\mathbf{x}) = H(\mathbf{x}_b) + \mathbf{H}(\mathbf{x}-\mathbf{x}_b)$) the optimal analysed state can then be found by solving $\nabla J(\mathbf{x}) = 0$ where

$$\nabla J(\mathbf{x}) = \mathbf{B}^{-1}(\mathbf{x}-\mathbf{x}_b) + \mathbf{H}^T\mathbf{R}^{-1}(H(\mathbf{x})-\mathbf{y}) \tag{2}$$

which leads to the expression for the analysis

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}(\mathbf{y}-\mathbf{y}_b) \tag{3}$$

with $\mathbf{K} = \mathbf{A}\mathbf{H}^T\mathbf{R}^{-1}$ and $\mathbf{A} = (\mathbf{B}^{-1} + \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}$; $\mathbf{K}$ is the Kalman gain matrix, and $\mathbf{A}$ is the analysis error covariance matrix.

If the analysis is performed independently for each atmospheric column at the location of the observed quantities, the variational technique is said to be one dimensional (1D-Var) and the dimension of $\mathbf{x}$ reduces to the number of model levels times the number of control variables, thereby simplifying the minimization process. In this study the model state $\mathbf{x}$ contains vertical profiles of temperature and specific humidity and surface

pressure (i.e. $\mathbf{x}_b = (T, q, P_s)$) derived from the regional non-hydrostatic forecast model COSMO. The observation vector $\mathbf{y}$ contains the surface rain rate estimation from the radar network and is therefore a scalar. H is the diagnostic moist process model which converts temperature and humidity increments into rain rate increments. Tangent-linear and adjoint versions of H are available in order to avoid the excessive computational cost of a minimization based on finite-difference Jacobians.
The moist physics parameterizations used in this work are an adapted version of codes initially developed at ECMWF for the assimilation of global-scale satellite rainy microwave radiances [1] and of radar rain rates over the U.S.A. [5, 6]. They consist of two linearized parameterization of large-scale condensation [10] and convection [4] whose sensitivity to input perturbations is more linear than that of [9] parameterization. This ensures a smoother behaviour of the minimization and avoids excessive increments that may cause convergence problems [1]. The simulated surface rain rate therefore comprises both convective (RR_{conv}) and large scale (RR_{strat}) contributions.

Large-scale precipitation, RR_{strat}, is diagnosed from the grid-mean amount of cloud condensate, q_c, as:

$$RR_{strat} = q_c c_0 \left[1 - \exp\left\{ - \left(\frac{q_c}{q_c^{crit}} \right)^2 \right\} \right] \qquad (4)$$

where q_c^{crit} (set to 0.5 kg kg^{-1}) is the critical value of the in-cloud water content at which precipitation generation starts and c_0 is the conversion factor (equal to 4.167×10^{-4} s^{-1}).

Finally, the 1D-Var minimization core finds the solution $\mathbf{x}_a$ through the minimization of $J(\mathbf{x})$.

3 Experiments setup

To formulate the procedure in a computationally light way and to be sure that some hypotheses underlaying the variational assimilation are verified, some topics are examined in details.

Firstly the 1D-Var algorithm is not enclosed in the COSMO code. This means that to make use of it the assimilation cycle needs to be doubled. Fields from a first COSMO nudging cycle are extracted every 15 minutes to feed the 1D-Var scheme with vertical profiles to start minimization. Then retrieved profiles in output are nudged by repeating the first assimilation cycle. The major problem associated to this off-line application is that the retrieval of the analysed profiles of humidity and temperature are not updated during the 12 hours of the assimilation cycle. To overcome this difficulty and, hence, to mitigate the effect due to the use of old profiles, the assimilation cycle is divided into 4 interval defining a framework similar to the one of the Rapid Update Cycle (Figure 1).

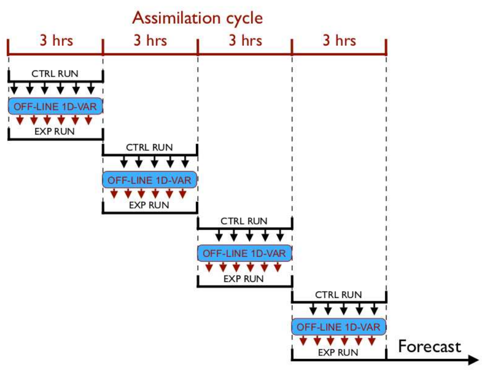


Figure 1: 1D-Var+nudging framework.

The second point of interest regards the data thinning.
The use of data with very high spatial and temporal resolution should guarantee improvements in the initial condition knowledge. Nevertheless, Liu and Rabier [3] showed how high density observations with correlated errors can produce a degradation of the analysis because of the potential spreading of error in correlated neighboring pixels. The most intuitive and commonly used thinning method is to reduce the amount of selected observations in predefined areas or at specified intervals [2]. Moreover, in our specific case the amount of data, coming from the italian network managed by the National Civil Protection Department, over the selected

domain is very large (57491 points every 15 minutes). The use of all the available observations generates AOFs so big to cause the killing of forecast runs by the system because of memory problems. As a first attempt, one observation every 5 grid points in both directions was taken, but, due to the poor results in forecasted precipitation fields the thinning procedure proposed by Lopez [7] has been chosen. As a result, only those points for which first guess and observed rain rates are greater than zero are used in the 1D-Var scheme.

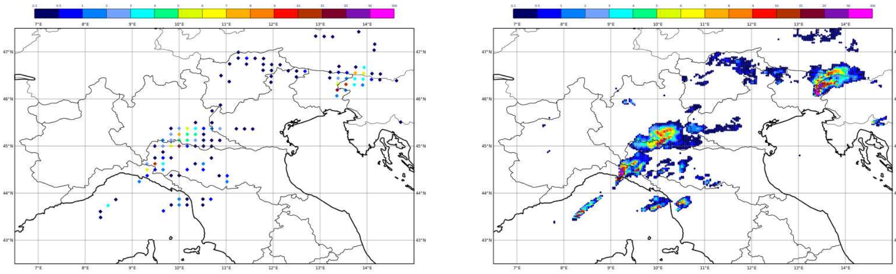


Figure 2: Regular data thinning (left) and suggested Lopez thinning (right) of precipitation field.

It is evident from Figure 2 how this two thinning methods are different in terms of number of points and structures in input to the 1D-Var minimization scheme.

The third topic is the bias correction.
The variational approach works in a statistically optimal way if observations and model errors are unbiased. Physics implemented in the forward operator, which is a simplified version of the cloud scheme implemented in the ECMWF forecast model, is different from the actual one implemented into the COSMO model. This means that, given a set of temperature and humidity profiles, precipitation fields generated by the cloud model diverge from those produced by the COSMO model. The differences between the linearized cloud model and COSMO have been compared by means of their diagnosed surface rain rate. In particular strong rain rates are not produced and the mean rainfall field is weak and diffuse (Figure 3) even taking note of the fact precipitation is not determined only by the “physical” balance of the total water contained in a 1D column but it also depends on dynamical driven processes.

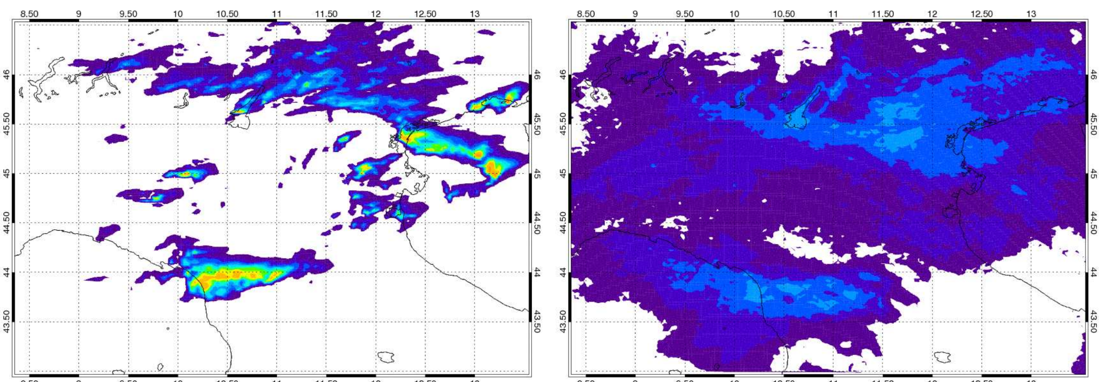


Figure 3: Mean surface rain rate for COSMO (left) and cloud (right) model calculated during the 12 hours assimilation cycle starting at 00 UTC of the 29^{th} of July 2010.

At first the bias correction was determined from the distribution of observed and simulated mean rainfall fields and it was applied to those observed precipitation rates for which there was an overestimation/underestimation compared to cloud model values. In this way the correction factor was evaluated a posteriori only for case studies. But, due to the difficulty in deducing this correction factor in a straightforward way not only for case studies, the idea was to change some parameters (such as convective cloud cover, autoconversion timescale of large cloud condensate to precipitation and autoconversion rate of convective cloud water to convective precipitation) trying to diminishing the spread effect observed in Figure 3 and to generate stronger rain rates. Mean rainfall fields following from different parameterizations are shown in Figure 4.

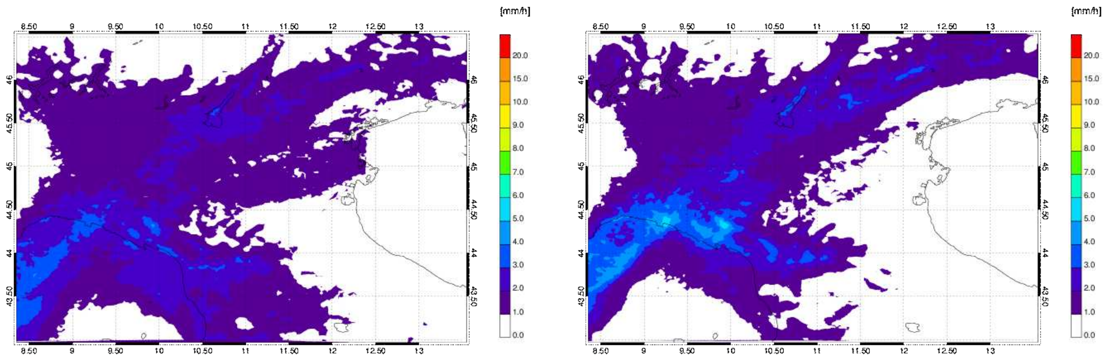


Figure 4: Mean rainfall fields obtained with the standard (left) and convective (right) parameterization. Data are from the 12 hours assimilation cycle starting at 00 UTC of the 26^{th} of September 2012

3 Case studies and results verification

To test the proposed methodology some case studies were chosen with two prerequisites to be fulfilled:

- presence of convective structures (short-lived and small-scale structures) in order to take advantage from 1 km resolution observations;
- high resolution COSMO model failure in forecasted precipitation in the operational configuration.

These requirements are demanding and limit the research of cases mainly in the summer season with a sharp restriction of the whole possible dataset.
For all of the three case studies presented, simulation runs consist of an assimilation and a forecast cycle both 12 hours long. A first verification is made by comparing qualitatively accumulated precipitation fields from the assimilation and the forecast cycle. The impact is assessed against the operational run, used as a control run, in which only conventional observations are assimilated through nudging scheme, and a run with LHN.

Then outcomes are verified quantitatively by means of the areal mean of accumulated precipitation over a selected domain. The area of verification, shaded in blue in Figure 5, is centered over the Northern Italy. Considered values are:

- 12 h accumulated precipitation in the assimilation cycle;
- hourly accumulated precipitation in the forecast cycle up to 12 hours.

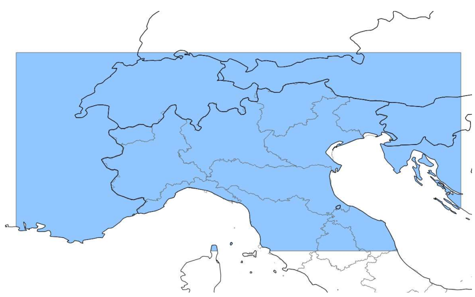


Figure 5: Selected domain used for the verification of results.

The first case study occurred during the Hymex Special Observation Period (SOP). The goal of the experiment is to resolve the underestimation of forecasted precipitation over Liguria and on the Apennines between Tuscany and Emilia-Romagna region. For this instance, the two different configurations of the 1D-Var are tested. Accumulated rainfall fields at the end of the assimilation cycle (Figure 6) display small modifications due to the change in parameterization parameters. Over the Alps the precipitation is a bit more widespread with an intensification of the convective core when radar observed profiles are assimilated. Over the Liguria region,

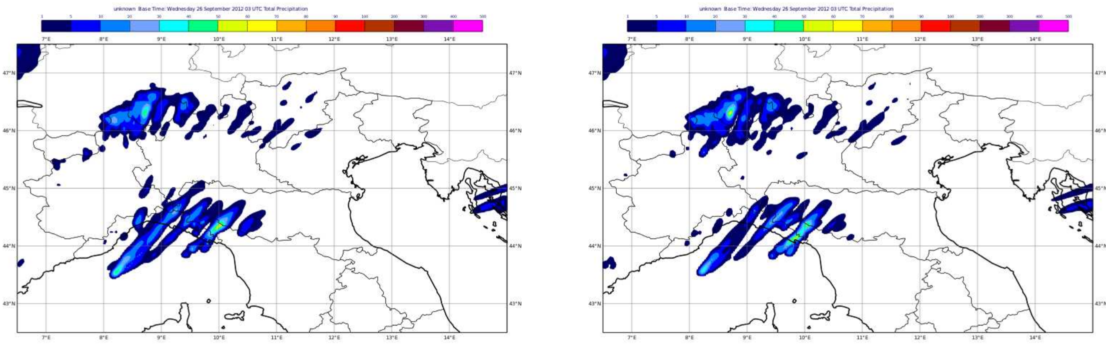


Figure 6: Accumulated rainfall fields over the assimilation cycle starting at 00 UTC of the 26^{th} of September 2012 obtained with the standard (left) and convective (right) 1D-Var configuration

an attenuation of precipitation can be recognized over the west structures, while there is an intensification in the east direction with the splitting in two parts of the convective nucleus.

Hence all of the runs are compared by means of accumulated precipitation fields.

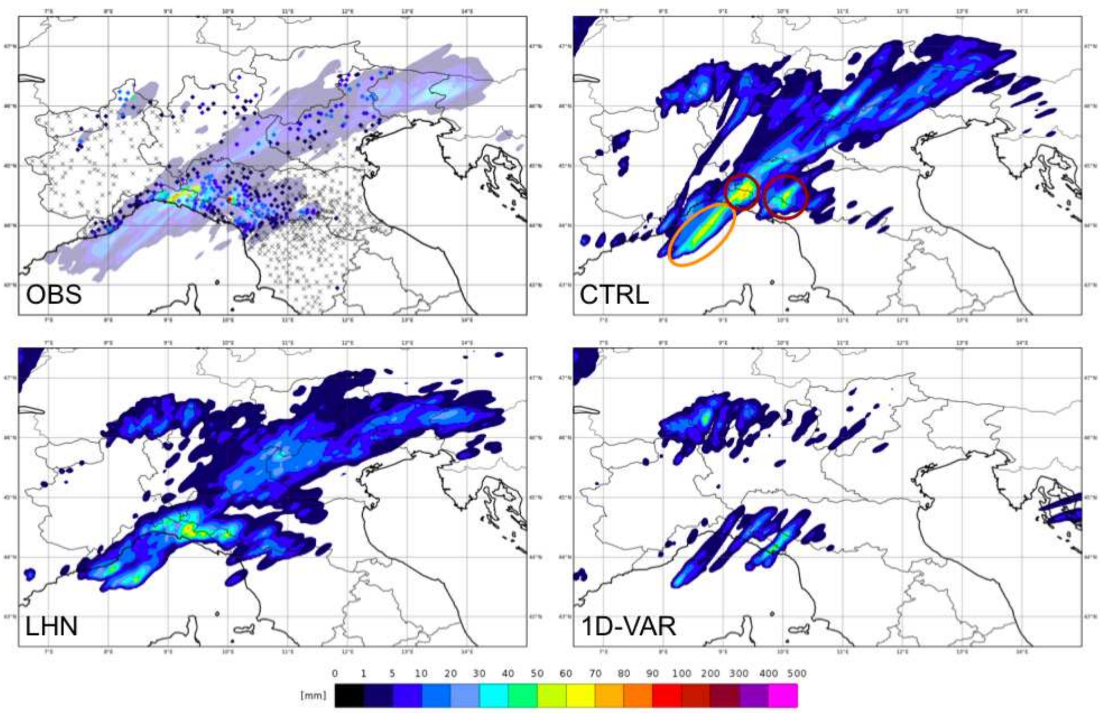


Figure 7: Accumulated rainfall fields at the end of the assimilation cycle for rain gauges and radar (top left), for control run (top right), for LHN run (bottom left) and for 1D-Var+nudging run (bottom right).

In the upper left panel of Figure 7 accumulated precipitation measured by rain gauges (diamonds) is displayed over the same field retrieved by the radar network (shaded area). These are observations that are used for the qualitative verification of output fields. Control run (upper right side) correctly predicts the pattern of precipitation field even if there is an overestimation over the Tyrrhenian sea (orange circle), and an underestimation over Apennines (red circles). LHN run (lower left panel) improves results decreasing precipitation over the sea, but does not reproduce highest rain rates over the Apennines. 1D-Var+nudging run (lower right panel), instead, predicts in a wrong way the rainfall fields with a general underestimation. The overestimation over the Alps cannot be verified due to the lack of rain gauges and the probable blindness of radar in that region.

The comparison of accumulated rainfall over the forecast cycle does not show great differences among the 3 runs (Figure 8). All of the runs miss the precipitation over Liguria and over Apennines between Tuscany and Emilia-Romagna where an improvement of the forecasted fields are expected (pink circles). Instead, runs with the assimilation of radar observations better predict precipitation over Northeastern Italy where there was an overestimation in the control run.

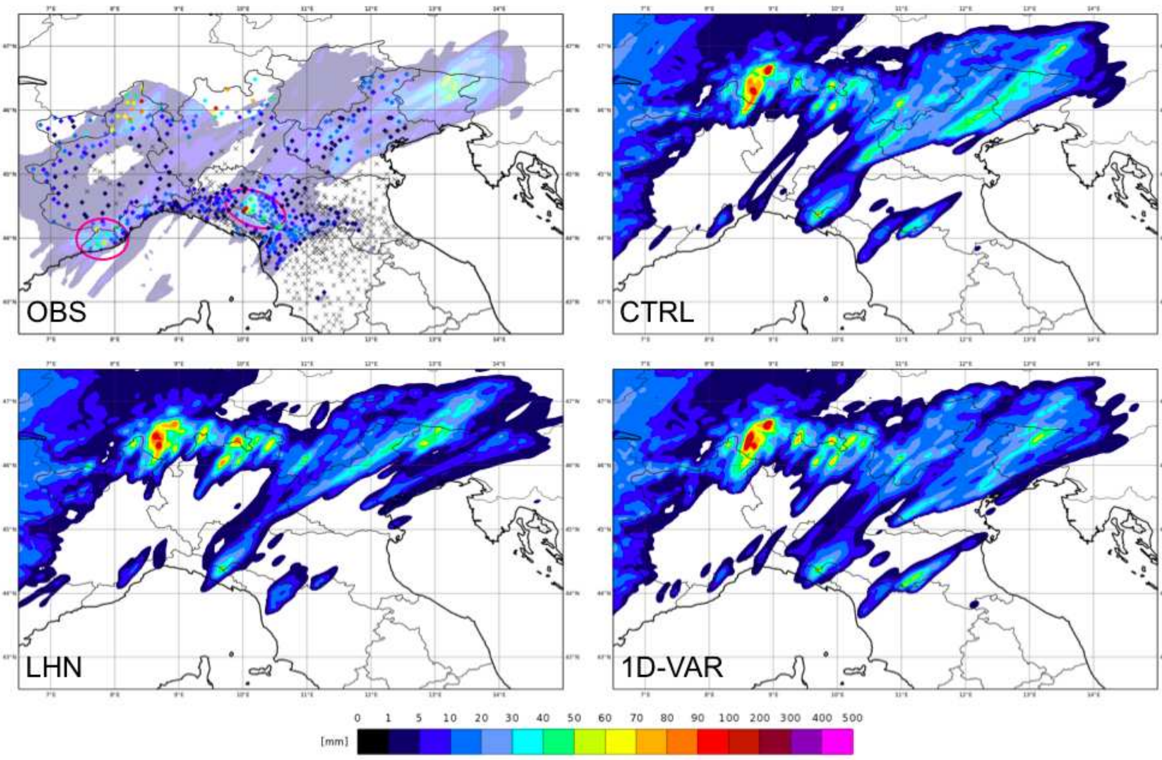


Figure 8: Accumulated rainfall fields at the end of forecast cycle for rain gauges and radar (top left), for control run (top right), for LHN run (bottom left) and for 1D-Var+nudging run (bottom right).

In the areal mean precipitation graph (Figure 9) all of the model runs start with a very low mean precipitation with respect to the observed one (blue line). The two different 1D-Var parameterizations (standard configuration in yellow, convective one in red) give quite the same results and have a positive trend towards observations. In the assimilation cycle (on the left) LHN (pink line) improves with respect to the control run. In the forecast cycle, control and LHN runs are slightly different and in the first three hours they are both better than 1D-Var+nudging run. 1D-var seems to go better between the 4^{th} and the 5^{th} hour. Behind this time all runs are slightly different. Observation information is substantially lost. At the end of forecast period, precipitation is clearly underestimated.

The second case study starts at the 00 UTC of 21^{th} of July 2012. Rainfall patterns both during assimilation and forecast are quite the same. The greatest difference is in the amount of precipitation. During the assimilation cycle (Figure 10) the control run overestimates all of the field, the overestimate of LHN is localized only over Northeastern Italy while, as settled before, 1D-Var presents a general underestimation. In the forecast (Figure 11) all of the runs miss the precipitation in the areas evidenced by red circles. Control and LHN runs display a very strong convective core not observed (pink circle), while over the same area 1D-Var+nudging run is completely dry. Surprisingly 1D-Var+nudging run gives more intense precipitation with respect to the other integrations.

The chart of Figure 12 shows very low precipitation values. In this case the best estimates are from 1D-Var+nudging run, while for the other runs there is a common overestimation. In the first hours the forecast runs are too wet with respect to the observations and they have a similar trend which does not fit the observed one.

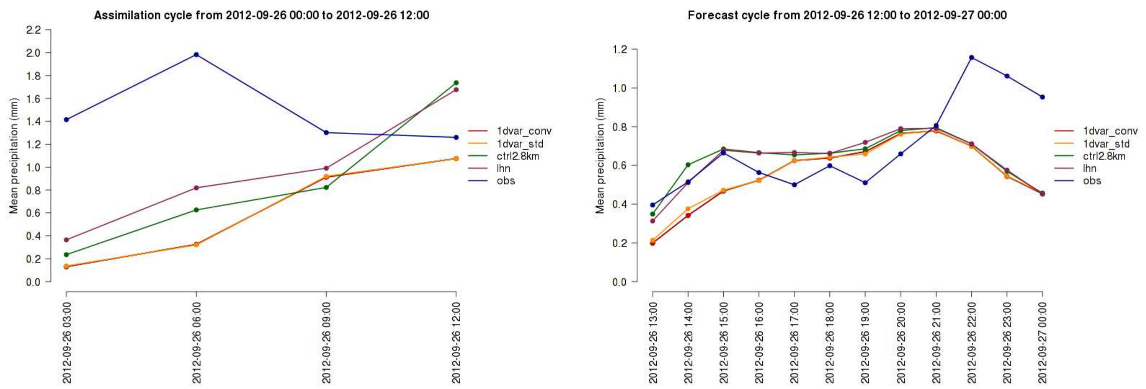


Figure 9: Areal mean precipitation in function of time for the assimilation (left) and forecast (right) cycle for the different runs against observation (blue line) for the 26^{th} of September 2012 case study.

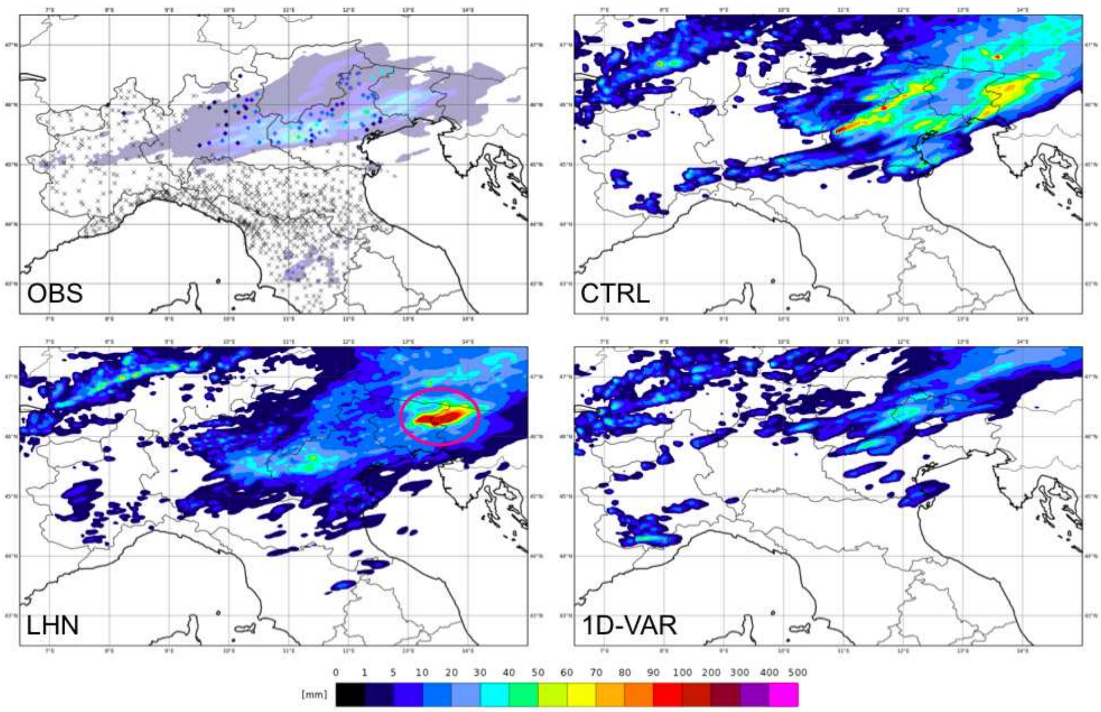


Figure 10: Accumulated rainfall fields at the end of the assimilation cycle for rain gauges and radar (top left), for control run (top right), for LHN run (bottom left) and for 1D-Var+nudging run (bottom right).

For the last case study only quantitative results are presented due to the small differences in forecasted patterns (Figure 13). From the quantitative point of view, in the assimilation cycle what can be recognized is that LHN and control runs have the same trend of observations even if the overestimation of the LHN is greater than the one of the control run. The 1D-Var+nudging run performs better in the first 9 hours and then maintains its tendency by underestimating precipitation over the last hours. In the forecast cycle, as expected, the run starting from the 1D-Var+nudging analysis starts dryer than the others, which are too wet. All of the three forecasts lose the peak in the observed precipitation at 18 UTC. However, as seen before and as a common result, can be observed that in the forecast cycle the influence of assimilated observations is completely loss after few hours. Moreover, tendencies of forecasted precipitation are not able to follow the great changes in the observations.

From these case studies, it was expected that assimilation of 1D-Var derived profiles should trigger some instability and should produce greater amount of precipitation mainly because only points where first guess and observations are greater than zero. Due to the small changes in the forecasted rainfall fields, and due to a

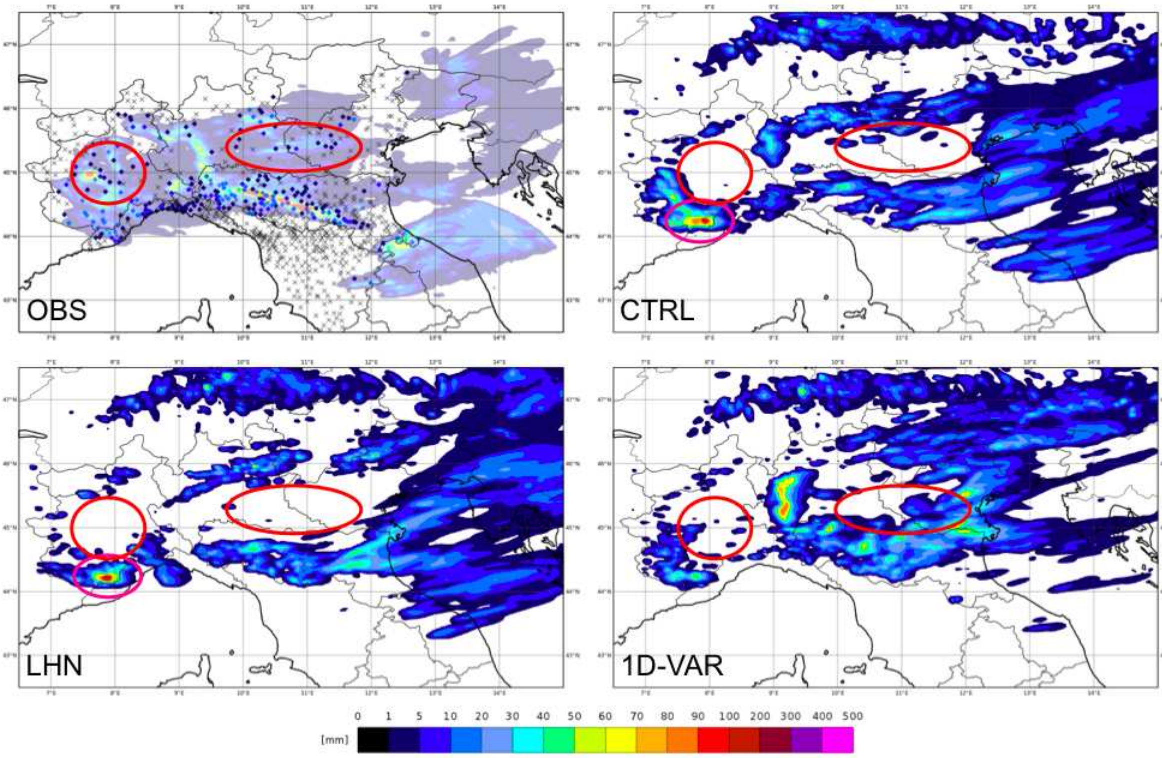


Figure 11: Accumulated rainfall fields at the end of forecast cycle for rain gauges and radar (top left), for control run (top right), for LHN run (bottom left) and for 1D-Var+nudging run (bottom right).

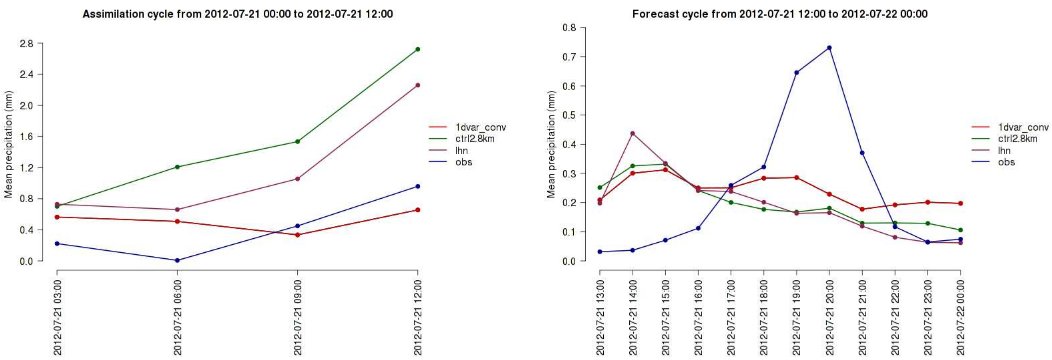


Figure 12: Areal mean precipitation in function of time for the assimilation (left) and forecast (right) cycle for the different runs against observation (blue line) for the 21^{th} of July 2012 case study.

general drying effect associated to the assimilation of 1D-Var retrieved profiles, a backward analysis starting from the 1D-Var algorithm was performed.
First investigation is made examining all of the profiles that come out from the 1D-Var scheme. Statistically the 70% of inputs converges providing temperature and humidity profiles to be nudged in COSMO. For the rest of the profiles the minimization of the cost function fails and points are discarded. A more detailed analysis is made over this sample. A direct comparison between observed radar rain rates and 1D-Var derived rain rates (Figure 14) shows how the minimization fails for the most part of points for which precipitation is moderate/heavy. Hence the information coming from points which should mainly contribute in producing rainfall is completely lost. The change of 1D-Var configuration is capable to increase the number of points associated to higher precipitation, but the strongest convective core structures are to a large extent lacking.

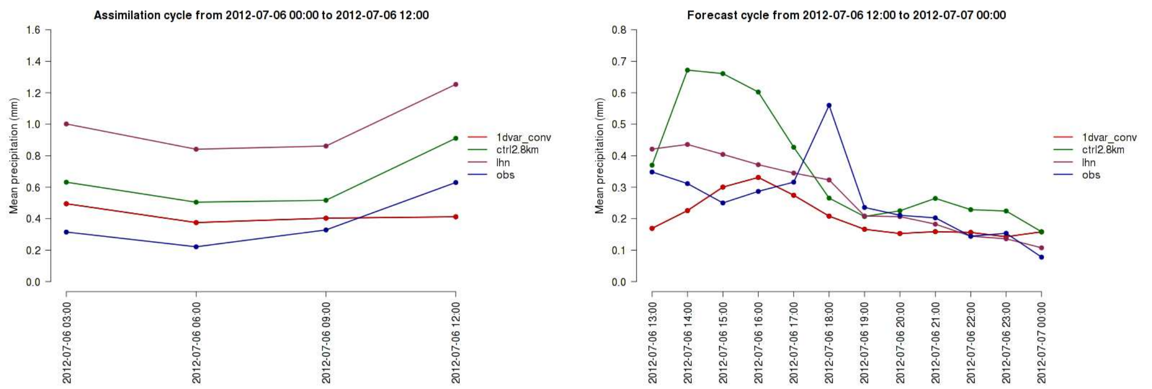
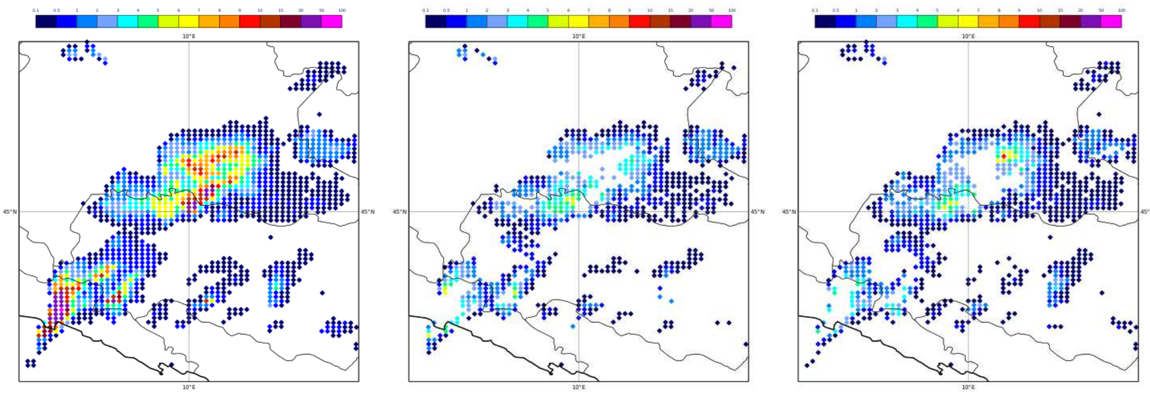


Figure 13: Areal mean precipitation in function of time for the assimilation (left) and forecast (right) cycle for the different runs against observation (blue line) for the 6^{th} of July 2012 case study.



(a) (b) (c)

Figure 14: (a) Observed radar rain rate for the 26^{th} of September 2012; (b) Output from 1D-Var scheme for the standard configuraztion; (c) Output from 1D-Var scheme for the convective configuration.

4 Summary and Outlook

In latest years different attempts were made in order to understand how and how much the assimilation of radar data through the 1D-Var+nudging technique affects the precipitation forecast.

Products were verified subjectively and objectively examining 12 hours accumulated precipitation. Despite changes, results show that LHN scheme outperforms the proposed methodology. These unsatisfactory results are mainly due to two different reasons:

- the moist physics implemented in the 1D-Var differs from the one of the COSMO model;
- the use of a linearized moist physics that has been designed at coarse resolutions is not appropriate to represent intense precipitation events by very high resolution models.

These conclusions imply that this methodology is not suitable for the assimilation with the COSMO model of high density rain rate estimates based on radar data.

References

[1] P. Bauer, P. Lopez, A. Benedetti, D. Salmond and E. Moreau *Implementation of 1D+4D-Var assimilation of precipitation-affected microwave radiances at ECMWF. I: 1D-Var*, Q. J. R. Meteorol. Soc., **132**, 2006,2277–2306.

[2] V. Bondarenko, T. Ochotta, D. Saupe and W. Wergen *The interaction between model resolution, ob-*

servation resolution and observation density in data assimilation: a two-dimensional study, Proceedings 11^{th} Symposium on Integrated Observing and Assimilation Systems for the Atmosphere, Oceans, and Land Surface, San Antonio, Texas, 2007.

[3] Z.-Q. Liu and F. Rabier *The interaction between model resolution, observation resolution and observation density in data assimilation: a one-dimensional study*, Q. J. R. Meteorol. Soc., **28**, 2002, 1367–1386.

[4] P. Lopez and E. Moreau *A convection scheme for data assimilation: description and initial test*, Q. J. R. Meteorol. Soc., **131**, 2005, 409–436.

[5] P. Lopez and P. Bauer *"1D+4D-Var assimilation of NCEP Stage-IV radar and gauge hourly precipitation data at ECMWF*, Mon. Weather Rev., **135**, 2007, 2506–2524.

[6] P. Lopez and P. Bauer *Experimental assimilation of ground-based radar data in ECMWF 4D-Var*, Proc. of 5^{th} European Radar Conference on Radar Meteorology and Hydrology (ERAD), Helsinki, Finland, 30 June–4 July 2008.

[7] P. Lopez 2010 *Direct 4D-Var assimilation of NCEP Stage IV radar and gauge precipitation data at ECMWF*, Tech. Memo., **627**, 2010.

[8] K. Stephan, S. Klink and C. Schraff *Assimilation of radar-derived rain rates into the convective-scale model COSMO-DE at DWD*, Q. J. R. Meteorol. Soc., **134**, 2008, 1315–1326.

[9] M. Tiedtke *A comprehensive mass flux scheme for cumulus parameterization in large-scales models*, Mon. Weather Rev., **117**, 1989, 1779–1800.

[10] A. M. Tompkins and M. Janisková *A convection scheme for data assimilation: description and initial test*, Q. J. R. Meteorol. Soc., **130**, 2004, 2495–2517.