Referee #1

In this paper several experiments are conducted to explore the sensitivity of a convective scale ensemble-based data assimilation system to the: length of the assimilation window, inclusion of additive inflation, variance of the reflectivity error. The experiments are conducted using real radar observations and combining latent heat nudging with the assimilation of conventional observations and radar reflectivity.
Although the topic is relevant and the experiments presented by the authors are interesting. I think that some aspects of the manuscript needs improvement. I give details of these in the following comments.

Major comments:
Some aspects of the presentation needs to be improved. For example, some paragraphs needs to be reorganized to improve the clarity of the manuscript. Also some figures can be merged in order to reduce the total number of figures in the paper. I provide examples of these changes as minor comments.
A reply to each comment is provided in the "Minor comments" section, indicating also the changes made to the document.

In this paper Latent Heat Nudging is combined with "direct" assimilation of radar reflectivity using EnKF. The introduction suggests that what is examined in this paper is the assimilation of radar reflectivity, however all the experiments use Latent Heat Nudging of precipitation estimated from radar reflectivity. I feel that the objective of the work should be reformulated since what is being explored is the added value of the LHN with "direct" assimilation of reflectivity. In this context an experiment which do not used LHN will also provide interesting results for comparison and discussion and will strength the conclusions. Also, if this is the focus of the paper references of previous work discussing these issues should be included in the introduction. I believe that LHN and EnKF has been combined in the development of the Rapid Update Cycle developed for the US.
As suggested by both the reviewers, an experiment (rad60_nolhn) in which conventional data and reflectivity volumes are assimilated without LHN is added to the manuscript. Details of the set-up employed are described in section 2.4 (as for all the other experiments) while results and the comparison with rad60 (in which LHN is applied together to the assimilation of reflectivity volumes) are provided in section 3.1.
No significant changes can be noticed in the results when comparing rad60 and rad60_nolhn. Therefore, it is decided to not switch off the LHN for the other experiments. In fact, this choice does not affect the results of the sensitivity tests that are presented in this work and, at the same time, the LHN allows to use radar derived information on the state of the atmosphere in the whole Italian country, despite reflectivity volumes can be assimilated, at present, only over Northern Italy.
Some modifications and additions are made in the abstract, in the introduction and in the conclusions to better empathize the use of LHN in combination to the assimilation of reflectivity volumes. Furthermore, it is explicitied that a test is performed to evaluate the impact of switching off LHN when reflectivity volumes are assimilated

Minor comments
P1L6 -We evaluated the impact or In this work the impact of … is evaluated
Done.

P1L8 – SAL is not described.
A brief explanation of the SAL technique was already provided but, to improve the clarity, the sentence is modified as follows:
A 4 days test case on February 2017 is considered and the verification of QPFs is performed using the SAL technique, an object-based method which allows to decompose the error in precipitation fields in terms of structure (S), amplitude (A) and location (L).

P1L9 - Missing stop before Results
Done.

P1L12 of additive inflation
Done.

P1L19 from the issue of … to decision making in …
Done.

P1L22 - convection allowing models are a significant improvement in this direction. I think this should be mentioned explicitly because it allows the use of reflectivity data to improve the initial conditions.
The sentence is modified as follows:
In recent years, the increase of available computing resources has allowed to increment NWP spatial resolution and to improve the accuracy of parametrization schemes, enabling to develop convection-permitting models (Clark et al., 2016).

P2L9 can you provide a reference for this?
Done: Schraff et al., 2016.

P2P14 - Recently particle filter has been successfully applied to convective allowing data assimilation see for example: Poterjoy, J.,2016: A Localized Particle Filter for High-Dimensional Nonlinear Systems. Mon. Wea. Rev., 144, 59–76, https://doi.org/10.1175/MWR-D-15-0163.1
This sentence is added and a previous sentence (in the new version of the manuscript is P2L10) is changed according to this addition:
Another option may be to employ particle filters but, despite the efforts to overcome the dimensionality challenges of these assimilation techniques (e.g. Poterjoy, 2016), it is still not feasible for operational applications.

P2L25 There are some previous work as well that deal with the issue of assimilation reflectivity in an EnKF starting (as far as I know) from the following paper: Snyder, C. and F. Zhang, 2003: Assimilation of Simulated Doppler Radar Observations with an Ensemble Kalman Filter. Mon. Wea. Rev., 131, 1663–1677, https://doi.org/10.1175//2555.1
The sentence and the subsequent one are modified as follows:
Conversely, only few tries have been made to directly assimilate reflectivity volumes in a convection permitting model employing EnKF techniques (e.g. Snyder and Zhang, 2003), especially in an operational framework (Bick et al., 2016). Despite some promising results, many issues affect the assimilation of reflectivity volumes at high spatial resolution and several aspects need to be further investigated.

P2L29 I think this sentence may need more clarification. There is no mention to non-linear effects. I believe that one of the main reasons why a short window is desirable is because non linear effects will become important for longer windows.
The sentence is modified as follows:
In EnKF methods, a short window would be desirable to avoid that dynamical features leave the area where computed localized increments are significant (Buehner et al., 2010a) and to better preserve the gaussianity of the ensemble which can be compromised by non-linearities (Ferting et al., 2007).

P2L32: Why when reflectivity volumes are assimilated the window length becomes more crucial?
Due to their high resolution, reflectivity observations can better define the small-scale features of weather phenomena. To improve the understanding, the sentence is modified as follow:
When reflectivity volumes are assimilated, the window length becomes even more crucial since these observations allow to catch small scale features of the atmosphere (Houtekamer and Zhang, 2016).

P3L1 by the use of short localization scales.
Done.

P3L2 replace instability by imbalance (this also applies to other parts of the manuscript).
Done.

P3L6 are known
Done.

P3L4 In this paragraph the issue of observation error correlation should be mentioned as an additional challenge when dealing with radar data assimilation.
This sentence is added:
Finally, a further challenge is the estimation of the observational error correlation especially when dealing with radar data assimilation, due to the high density of this type of observations.

Figure 2, 3 and 4 can be merged into one single figure.
We merged figures 2 and 3. We have decided to keep Figure 4 separate as it relates purely to the verification of results. In addition, we decided to plot, over the verification domain, the rain-gauges stations used in the verification. As a result, some small modifications are made in sections 2.3 and 2.5.

P4L14 Is this scheme an online estimation scheme? What is the horizontal localization scale used in the experiments?
RTPP is an online scheme.
For horizontal localization, an 80km length scale is used for conventional data (as Bick, 2016) while 16km for radar volumes (as Bick, 2016). The following sentences are added at the first paragraph of section 2.2 to include these information:
To avoid spurious long-distance correlations in the background error covariance matrix, analyses are performed independently for each model grid point taking into account only nearby observations (observation localization). Observations are weighted according to their distance from the grid point considered using the Gaspari-Cohn correlation function (Gaspari and Cohn, 1999). In the present work, two different values of the Gaspari-Cohn localization length-scale are employed for conventional and radar observations: 80 km for the former, 16 km for the latter (as done by Bick et al., 2016).

P5L4 A discussion of possible implications of using a B matrix designed for low resolution models should be presented here. It would also be good to discuss previous work that shows positive impact associated with the inclusion of additive noise for convective scale data assimilation: Dowell, D.C. and L.J. Wicker, 2009: Additive Noise for Storm-Scale Ensemble Data Assimilation. J. Atmos. Oceanic Technol., 26, 911–927,https://doi.org/10.1175/2008JTECHA1156.1
Some additions and modifications have been made in the last part of the paragraph:
Since Q is not known, it is assumed to be proportional (by a factor smaller than 1) to a static background error covariance B (Mitchell and Houtekamer, 2000). This technique has already been employed with a positive impact in convective scale data assimilation (e.g. Dowell and Wicker, 2009). In the present work, additive inflation is used together with multiplicative inflation and to RTPP only in one experiment, employing a climatological B-matrix from the 3D-VAR of the Icosahedral Nonhydrostatic (ICON) global model (Zängl et al., 2015). Although the use of a lower resolution B-matrix may not allow to properly characterize the model error at the smallest scales, the same configuration has been gainfully employed at MeteoSwiss (Leuenberger and Merker, 2018).

P5L13, the paragraph starting here should be merged with the previous paragraph.
Done. Note that the whole paragraph regarding LHN has been moved at the end of the section. In this way, first all observations (conventional and non conventional) assimilated by KENDA are described, then those employed for the LHN are described.

P5L15 remove the :). Also it would be good to provide a reference for the quality control that is applied to radar data in general.

The reference for the quality control has been added and the statement has been changed to better explain how this control is carried out, as follows:

Data coming from each station undergo a quality control that removes those with low quality. The quality depends on different factors such as ground clutter, beam blocking, range distance, vertical variability and attenuation as described in Rinollo et al. (2013).

How is the issues associated with complex terrain handled in this case (e.g. beam blocking)

For the SRI product the effect of the beam blocking is combined with the other parameters that enter into the generation of quality. The use of different radars to generate the composite, taking for different radars the points with the highest quality, fulfill the domain and the direct beam blocking effect is lost.

P6L1 This sentence is not clear please revise it.

The sentence and the previous one are modified as follows:

The scheme, which is applied continuously during the integration of the model, acts in rescaling temperature profiles with an adjustment of the humidity field according to the ratio between observed and modelled rain rates. LHN has been gainfully employed in different frameworks, including forecasts over complex terrain (Leuenberger and Rossa, 2004; Leuenberger and Rossa, 2007).

Figure 3 should include the effect of beam blocking to have a better idea of the area actually covered by the radar

The figure 3 has been merged to figure 2.

As specified in two previous answers, due to the fact that it arises from a composite, the SRI product fulfill the domain presented in figure. The four radars highlighted in red are assimilated in the KENDA system taking all data volumes, for this reason it has less sense to indicate where radar beams are blocked, because this affects only first elevations and elevations higher than firsts have the same domain. The caption of the figure is misleading. For this reason it has been changed.

P6L15 Here it would be nice to add a reference for the data quality control.

Reflectivity volumes used in the KENDA systems come directly from two different Regional Meteorological Services. For this reason they are subject to different cleaning procedures described, if present, by internal documentation in italian, not suitable for being used as a reference.

P7L5 Is the superobbing approach used also in the vertical? Is the superobbing considered in the observation operator as well?

Superobbing is applied only in the horizontal to both the observed and simulated field, as well as the threshold of 5 dBZ.. To avoid misunderstandings about this, the subsequent sentence is modified as follows:

Furthermore, before performing superobbing on the observed and simulated fields, a threshold of 5 dBZ is applied to both fields in order to avoid that large innovations associated to non-precipitating signals would lead to large analysis increments without physical relevance.

P7L10 What do the authors mean by “on average along the vertical”?

It meant that the whole volume was taken into account in its entirety, calculating errors spanning all the vertical extension of data. The sentence is modified as follows:

Employing all radar data available during the test case, a reflectivity observational error equal to 5 dBZ is estimated, as found also by Tong and Xue (2005).

P7L13 In this section a description of the experimental setting is presented. The clarity of the first two paragraphs needs to be improved. For example some operational systems are mentioned that are not used in the rest of the paper. It would be good to have a comparison between the operational systems and the experimental system, but in this section only the information regarding the experiments should be included. Table I caption, replace trial by experiment

The operational system is mentioned because several aspects of it (like boundary conditions) are replicated in the experimental set-up. Furthermore, the whole operational set-up is substantially used in the conv60 experiment, the "control" experiment against which we evaluate if the assimilations of reflectivity volumes is advantageous. Anyway to improve the understanding, some small modifications have been made to section 2.4 and, in particular, P7L29 (which in the new version of the manuscript is P8L21) is modified as follows:

In the control experiment, called conv60, the set-up of the operational chain, described in the previous two paragraphs, is replicated. In particular, this means that in conv60 experiment only conventional data are assimilated using KENDA through cycles of 60 minutes and the LHN is performed during the forecast step of each assimilation cycle.

Which is the output frequency of the model for the data assimilation cycle?

The output frequency is equal to the time step of model, that is equal to 18 seconds.

P8L13 from February 3rd to February 7th

Done.

P8L16 new precipitation systems

Done.

Verification: It would be nice to show some examples of how the analysis look like and how the forecast look like in comparison with the observations. This will help to have a general idea on how well the system is working and how accurate the forecasts are.

Even if we understand the point of the reviewer, we do not think that showing a case in which the assimilation of reflectivity volumes improves the accuracy of forecast can really add some useful information to the reader. In fact, as shown by verification scores, case in which a positive impact can be found alternates to some in which the impact is negative. At the same time, we agree with the reviewer who said that the number of figures was already quite large.

The figure in which the areal averaged precipitation is shown is based on the use of dependent data for the verification of the assimilation system. The authors said that since all the experiment are verified in this way this should not be a problem. However for me this is not convincing, since validating with dependent data might not detect issues like overfitting. An analysis closes to the observation is not necessarily the best analysis.

Verification based on areal averaged precipitation employs independent data, since the only observations considered are rain-gauges which are not assimilated. To improve the understanding, some modifications have been done in section 2.5.

P9L15 Describe the SAL acronym

SAL acronym (which derives from the name of components) is now described in the abstract of the paper. In the verification section was already present.

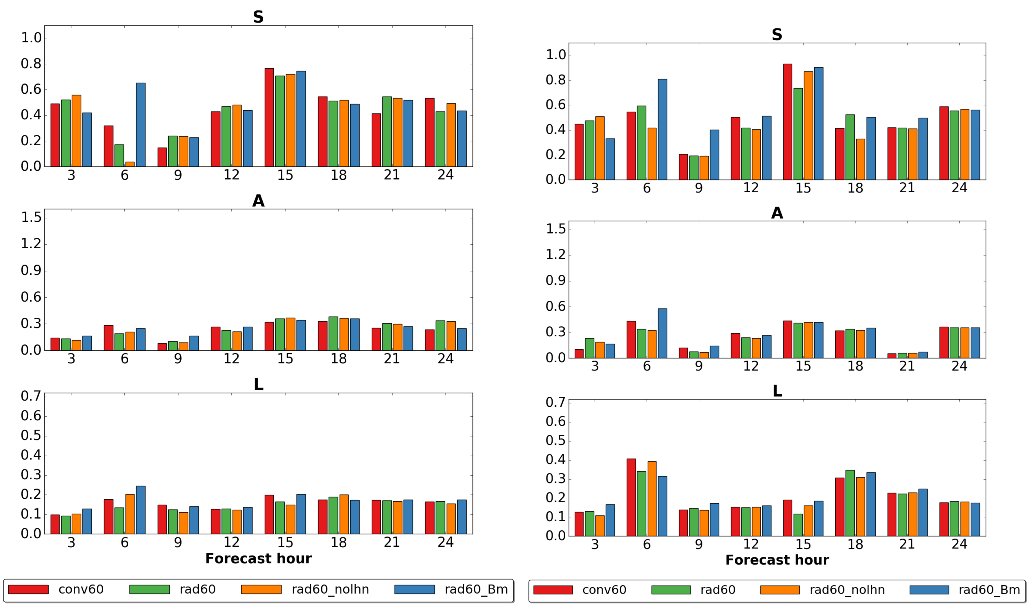
P9L17 Here the authors said that SRI observations are not independent, however the SAL approach is applied to the precipitation forecasts and not to the analysis. In this sense the observations are independent because these observations has not been assimilated yet.

P9L17 and P9L18 are removed

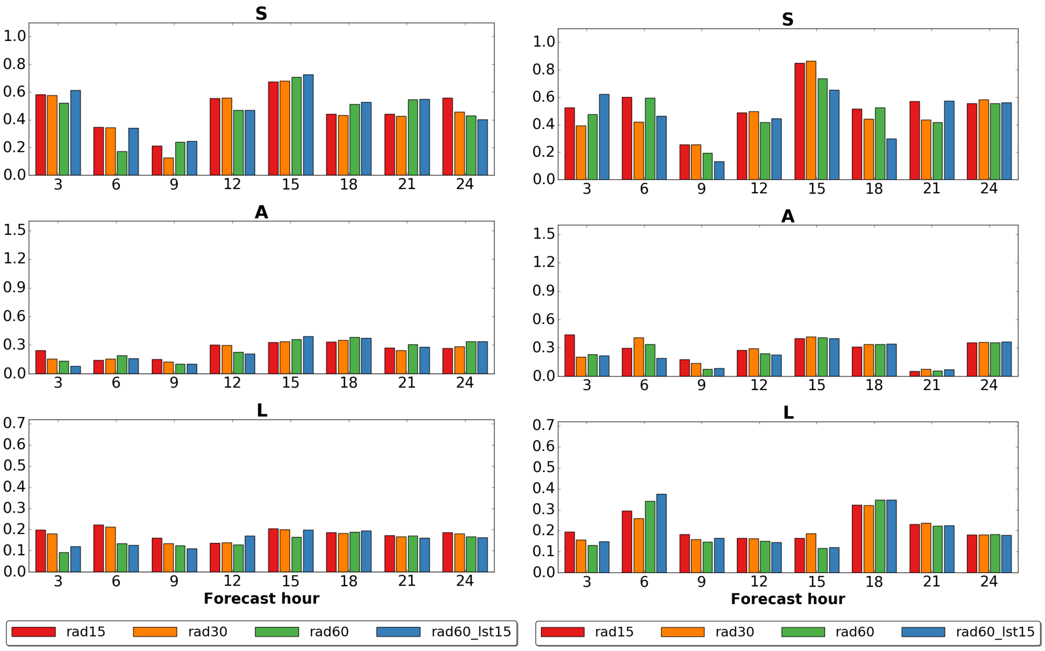
P9L30 Please provide more detailed explanation on this limitation of SAL. My understanding from this paragraph is that SAL can only work with one precipitation system at a time. But this is difficult to guarantee even if the domain is very small as proposed by the authors.
The problem of using SAL in a large domain is that, if precipitating system are strongly different, results will not be representative for the weakest system. To better explain the problem this sentence is added:
In fact, if the domain contains strongly differing meteorological systems, then results obtained using the SAL technique may not be representative for the weakest one.

P10L16 Only results concerning 1 mm are shown. Are the results sensitive to the threshold used in the SAL method? Can the authors comment on the results obtained with other thresholds as well?
Verification with SAL has been done also using a threshold of 3 mm, but results are not included in the manuscript because they are not significantly different from those obtained with a 1 mm threshold. Anyway, we add a sentence in the paper about this (P12L9 of the new version of the manuscript):
Verification using a 3 mm threshold was also performed but, since results do not differ significantly from those obtained with a 1 mm threshold, they are not shown here.

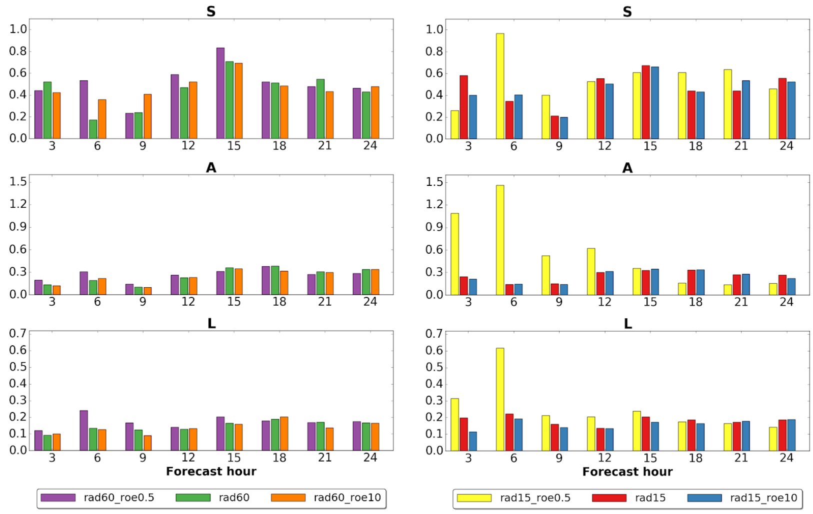
We report here the plots: on the left the ones for a 1 mm threshold (shown in the paper), on the right those with a 3mm threshold.
When comparing conv60 to rad60, rad60_nolhn and rad60_Bm no significant difference can be noticed and the conclusions written in the manuscript are unaffected: the assimilation of reflectivity volumes (applying or not at the same time LHN) does not improve consistently the accuracy of forecast precipitation; instead, using additive inflation slightly worsen results.

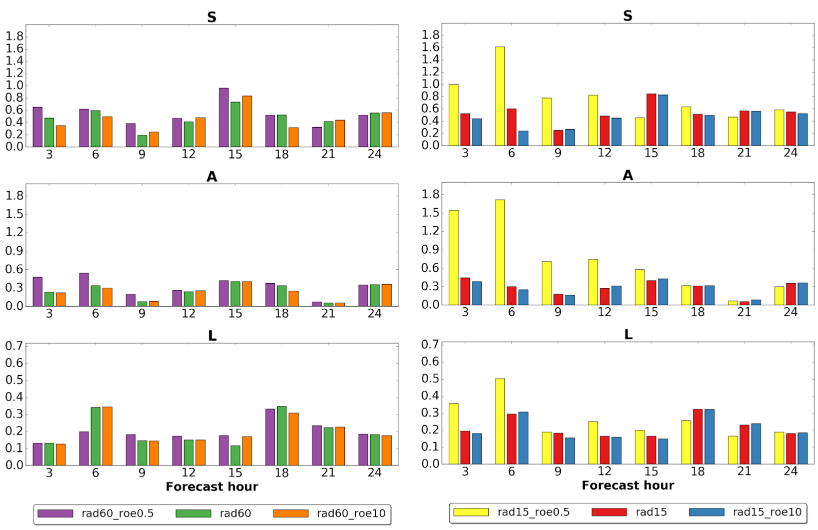


When considering the length of assimilation cycles, small differences can be noticed when using a threshold of 3 mm instead of 1 mm. At forecast time +3h, rad15 is the worst while rad30 and rad60 are similar. At +6h, rad30 is slightly better than rad60. Anyway, as stated in the manuscript, shortening the length of assimilation cycles does not affect significantly the accuracy of forecast precipitation.



Finally, regarding the reflectivity observational error (roe), verification with a 3mm threshold (below) leads to same conclusions obtained when considering a 1 mm threshold (above). When a very small value of roe (0.5dBZ) is employed to assimilate reflectivities throughout cycles of 15 min, accuracy of precipitation forecasts is strongly worsened. On the contrary, using a roe=10dBZ does not have a relevant impact.





P12L8 It is not clear how the kinetic energy spectra can be used to identify the effect of the imbalance. May be the evolution of the spectra with the forecast lead time would be a better tool to detect the presence of small scale noise that arises as a consequence of the assimilation of observations (in a similar way as it is done with pressure tendencies).

What we want to assess here is if shortening cycles from 60 minutes to 15 minutes introduces small scale noise in the analyses. Following Skamarock (2004), the fact that KE spectra are identical for the two experiments (especially at the highest wavenumbers) allows us to state that no imbalances are introduced when employing 15 minutes cycles instead of 60 minutes cycles. To improve the understanding of this point, the sentence P12L11 in the original version of the manuscript (P14L7 in the new version of the manuscript) is modified as follows:

Kinetic energy spectra of rad15 (red) and rad60 (green) are almost overlapping, even at very small wavelength, indicating that shortening the length of cycles from 60 to 15 minutes does not introduce imbalances in the analyses(Skamarock, 2004).

Referee #2

General comments:
I do recognize the authors' efforts on assimilating the radar data in their regional data assimilation system, KENDA. However, I felt that the setups of the experiments cannot clearly illustrate the impact of radar reflectivity on precipitation prediction, given that radar information has been injected into the model state through latent heat nudging. Also, with a high-resolution setup, it is somewhat surprising that there is no benefit from more rapid updates. I am also concerned a potential systematic underestimation of precipitation (and humidity) when a strong dependence on radar data is tested. These seem to lead to issues of radar data quality or how the authors handle the raw radar data. Based on these concerns, I will recommend major revision for this manuscript and hope the authors can address the following comments in their revised manuscript.

Major comments:
1. I understand that the assimilation configuration used in this study attempts to be close to the operation settings as much as they could. However. a big question is whether the justification of the impact from radar data on precipitation is fair, given that the precipitation nudging is always applied and the observations for verification contains both information of radar and surface rain gauges. Is it possible to conduct experiments without LHN for clean comparison? E.g. an experiment assimilates conventional data only and compares with the experiment that assimilates conventional and radar data. And, compare the results with the rain gauges data?

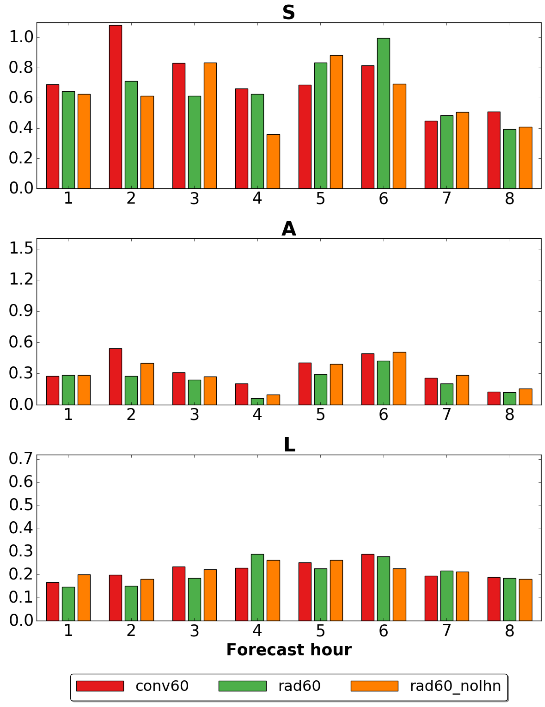
- Does the result imply that LHN is more effective than radar data assimilation?

As suggested by both the reviewers, an experiment (rad60_nolhn) in which conventional data and reflectivity volumes are assimilated without LHN is added to the manuscript. Details of the set-up employed are described in section 2.4 (as for all the other experiments) while results and the comparison with rad60 (in which LHN is applied together to the assimilation of reflectivity volumes) are provided in section 3.1.
No significant changes can be noticed in the results when comparing rad60 and rad60_nolhn. This means that LHN is no more effective than radar data assimilation. Therefore, it is decided to not switch off the LHN for the other experiments. In fact, this choice does not affect the results of the sensitivity tests that are presented in this work and, at the same time, the LHN allows to use radar derived information on the state of the atmosphere in the whole Italian country, despite reflectivity volumes can be assimilated, at present, only over Northern Italy.
Some modifications and additions are made in the abstract, in the introduction and in the conclusions to better empathize the use of LHN in combination to the assimilation of reflectivity volumes. Furthermore, it is explicitied that a test is performed to evaluate the impact of switching off LHN when reflectivity volumes are assimilated.
Regarding verification, we stress that areal average precipitation (Figure 4 in the new version of the manuscript) is computed during the assimilation cycles and using rain-gauges as observations (independent data, since they are not assimilated). Regarding SAL verification, it is applied to the forecasts. In this sense the observations (radar estimation corrected by rain-gauges) are independent because these observations has not been assimilated yet.

2. Intuitively, the assimilation of radar data is expected to improve the precipitation. It doesn't seem to be a reasonable choice to me that only examining the absolute value of the components of SAL, without trying to distinguish the possibility of overforecasting or underestimation of the precipitation.
Although in the manuscript only the absolute value of the components of SAL is shown, we have looked carefully at SAL values for each experiment and forecast. Only when significant results are found (like the underestimation of average precipitation in rad15_roe0.5 experiment) they are reported in the manuscript. Therefore we don't think that there is a way to show results (synthetically) better than the one we provided.

- In most of the literate using radar data, the impact is mostly seen in the first 6-h forecast and some even only last for 3 hours. Do the authors see a clear impact for the 1-h or 2-h lead time?
  When considering precipitation accumulated hourly and applying a 1 mm threshold for the

verification (see figure below), a clear positive impact when assimilating reflectivity volumes can be noticed at forecast lead time +2h, while a small positive impact can be noticed at +1h and +3h (in this case, especially for rad60 that is when LHN is applied together to the assimilation of reflectivity volumes). In this article, we focused on longer forecast lead times because they are much more interesting from an operational point of view.



3. It is not too surprising to me that rad60_BM has a worse performance since the perturbations used to augment the ensemble-based background error covariance may be in larger scale. I will suggest either remove this experiment or illustrate the reasons that degrades the performance.

Actually, using the same set-up (perturbations from the ICON model into regional data assimilation with COSMO and KENDA) has been gainfully employed at MeteoSwiss (Leuenberger, D. and Merker, C.: "Additive Covariance Inflation in an operational, convective-scale NWP Ensemble Kalman Filter Assimilation System", Poster contribution at the International Symposium on Data Assimilation (ISDA) 2018, https://isda2018.wavestoweather.de/5 program/poster_presentations/p6_1_leuenberger.pdf.).

Anyway, to improve the understanding, some modifications/additions are made from P5L21 in the new version of the manuscript:

Since Q is not known, it is assumed to be proportional (by a factor smaller than 1) to a static background error covariance B (Mitchell and Houtekamer, 2000). This technique has already been employed with a positive impact in convective scale data assimilation (e.g. Dowell and Wicker, 2009). In the present work, additive inflation is used together with multiplicative inflation and to RTPP only in one experiment, employing a climatological B-matrix from the 3D-VAR of the Icosahedral Nonhydrostatic (ICON) global model (Zängl et al., 2015). Although the use of a lower resolution B-matrix may not allow to properly characterize the model error at the smallest scales, the same configuration has been gainfully employed at MeteoSwiss (Leuenberger and Merker, 2018).

4. It is unclear to attribute the degradation of using a sub-hourly assimilation window to location of rainfall nuclei (Page 13, line, 4). Can the authors explain why a more rapid update (15 or 30-min window) lead to a worse performance than the one using a 60-min window since using a short assimilation window does not introduce the imbalance issue?

When we noticed the small degradation of performance associated to shorter cycle, we thought that the cause might be the application of some balance constraints (hydrostatic and saturation adjustments) when computing analysis. With shorter cycles they are applied more frequently. So, we decided to remove them but results were almost unaffected. Therefore, it is not yet clear to us why results degradates but we want to stress that the degradation is really small and when considering different threshold (like 3mm) for the verification the degradation becomes even smaller.

5. The authors explain that a larger A component in SAL with the use of small observation error (roe0.5) is due to a systematic underestimation of the average precipitation over the domain or as the example showing a result of decreased humidity. With a strong confidence in observations, such results will be highly dominated by the characteristics of the radar reflectivity data. Do the authors observe that the rain estimated by radar data is underestimated as compared with the rain gauge data? Is there a calibration issue such as the attenuating effect in radar data or the QC procedure (O-B > 5dBz) before the superobservations were constructed?
Raw radar volumes are pre-processed before their use in the KENDA system. Volumes come from different Regional Meteorological Services and they are subject to different cleaning/calibration procedures. Corrections take into account clutter, beam blocking and attenuation. Unfortunately, as explained to the other reviewer, these procedures are described, if present, by internal documentation in italian, not suitable for being used as a reference.
The "procedure" O-B>5dBZ is not a quality check, it is explained more in detail in the next replies section regarding set-ups about radar data assimilation.
However the text was not clear and the sentence P7L4 in the new version of the manuscript has been modified:
Before assimilation raw reflectivity are pre-processed taking into account non meteorological echoes, beam blocking and attenuation to improve the quality of data. In particular, it is important to eliminate the clutter signal...

- In the experiments of rad60_roe10 and rad60_roe0.5, is the QC during assimilation the same?
  Yes, the QC is the same.
- I don't quite follow with the rationale in the last paragraph on Page 15. With rad15_roe0.5, It should be the assimilation introduces the small-scale features, instead of losing the ability to "correct" the small-scale errors. To verify this, I suggest that the authors can compare the KE spectrum before (background) and after (analysis) assimilation.
  In the paragraph the expression "small scale structures" was used wrongly to indicate "small scale noise". Now it is corrected. We stress that, analyzing KE spectra, we only want to assess if some configurations employed to assimilate reflectivity volumes provide more balanced analyses than other configurations. According to Skamarock (2004), higher values of KE at the smallest wavelengths of experiment rad15_roe0.5 compared to those of rad60 indicate that the former experiment is more unbalanced than the latter.

6. Information and setups about Radar data assimilation are not clear.

- Although Bick et al. (2016) had described the operator characteristics, and other radar data management. It will still be essential for this paper to briefly provide the important information related to data assimilation. For example, the volume used to construct the superobservation (degree, gate, etc..?). Are all the radar data from different observation intervals with different radars used for constructing the superobservations?
  Informations regarding volume data in input are given in the description in the section "assimilated data": volumes have a range resolution of 1km, while the azimuthal resolution is 1 degree for Bric della Croce and Settepani and 0.9 degree for San Pietro Capofiume and Gattatico. Superobservations are made individually for each acquisition.
- Page 7, line 6: Is there a precondition to reject (O-B > 5dBz) to avoid large innovations associated to non-precipitating signals? If (O-B >5dBz), doesn't it mean that observation tend to have more rain water? Are the assimilation/forecast results sensitive to such choice?
  To avoid misunderstandings, we stress that what the reviewer calls "O-B > 5dBz" means that reflectivity values which are smaller than 5 dBZ are set equal to 5 dBZ. This correction is made for

both the observation and background fields. This is done to avoid that large innovations associated to non-precipitating signals would lead to large analysis increments without physical relevance. This choice does not imply that observation field tend to have more water, since the "correction" is applied to both observation and background fields. We have not tested different values of threshold, since the value of 5 dBZ has already been employed in other studies (Bick et al., 2016).

- If the horizontal grid-spacing of the analysis domain is 2.2km, isn't it too coarse to have superobservations with resolution of 10km?
  Actually, analysis weights are computed on a coarser grid (6.6km). Anyway, even if the analysis grid had been 2.2 km, a higher resolution for superobbing would have violate the assumption that observations are independent.
  Some explanations about the analysis coarse grid are added at the end of section 2.2:
  The KENDA suite also allows to compute the analysis weights on a coarsened grid (Yang et al., 2009). Weights computed on this coarsened grid are then interpolated to the model grid and afterwards used to calculate analysis increments. In this way, the computational cost is decreased without affecting negatively the accuracy of analysis (Yang et al., 2009). In the present study, a coarsening factor equal to 3 is employed
- Since Bick et al. (2016) used an observation error of 10dBz, is there a particular reason why this study reduces the observation error to 5dBz?
  As stated in the last paragraph of section 2.3, the value of 5 dBZ has been estimated by applying Desroziers' statistics to our case study. To stress this concept, it is stated also at the beginning of section 3.3, modifying the sentence as follows:
  In addition to the value of 5 dBZ employed so far, which was estimated applying the diagnostic described in Desroziers et al. (2005) to this case study, two other values of roe are tested: 10 dBZ and 0.5 dBZ.
- Page 7, line 28: Isn't the radial velocity also assimilated? It's not clear to me why the authors only emphasize on the contribution from reflectivity.
  In this work only reflectivity has been assimilated. The radar operator gives the possibility to assimilate also radial winds, but at the moment we haven't tested them yet. As it does not seem too clear, we have modified the paragraph on P6L11:
  Although the operator gives the possibility to assimilate both radial winds and reflectivities, in the present work only reflectivity volumes are assimilated.

Minor comments

Please provide the following Information

- Page 3, line 28: what is the model top of the model?
  This sentence is added:
  The model top is at 22 km.
- Page 5, line 4: please spell out the full name of the ICON model.
  Done.
- Page 5, line11, 14: It's not clear the composite map is composed by what data? Radar only? Or weighted average with the surface rain rate? Is this the same as the observations employed to perform SAL? (Page 9, lines 16-17)
  SRI from the radar composite is composed only by radar data. To perform SAL, SRI (used for LHN during assimilation) is corrected using rain-gauges. Please note that SAL is employed only for the verification of forecasts. Some small modifications are made in the paragraph to improve its clarity.
- Page, 7, line19: I would suggest to cite the original reference for the LETKF scheme: Hunt et al. 2007.
  In Bonavita et al. (2010) the implementation of the LETKF in COMet is described in detail, so we think it is a better reference than Hunt et al. (2007).
- Should I assume that the horizontal grid-spacing of the domain for assimilation is 2.2km?
  No, as reported above, a coarse analysis grid is employed. Details about that are added at the end of section 2.2.

- Page 10, line 20: “observed rainfall field consists of at least 1000 grid points”=> It would be better to change the observed target based on the definition of area (e.g. 50km x 50km?).
  The sentence is modified as follows:
  The average is computed considering only cases in which the observed rainfall field consists of at least 1000 grid points (3 events at lead time +6h, 4 otherwise), which is approximately equal to an area of 50x50 km^2.
- Page 16, line 3-4: Actually, a lot of efforts have been devoted to assimilation of radar reflectivity data already. I am not sure why the authors have such statement
  We refer to the assimilation of reflectivity volumes in an operational data assimilation system. The sentence is modified as follows:
  Assimilation of radar data is a challenging issue and most of the previous studies is devoted to the assimilation of rainfall estimation, while few to the direct employment of reflectivity observations in an operational data assimilation system.

Data assimilation of radar reflectivity volumes in a LETKF scheme

Thomas Gastaldo[1,2], Virginia Poli[1], Chiara Marsigli[1], Pier Paolo Alberoni[1], and Tiziana Paccagnella[1]

[1]Arpae Emilia-Romagna Hydro-Meteo-Climate Service - Bologna, Italy
[2]University of Bologna - Bologna, Italy

Correspondence to: Thomas Gastaldo (thomas.gastaldo2@unibo.it)

Abstract. Quantitative precipitation forecast (QPF) is still a challenge for numerical weather prediction (NWP), despite the continuous improvement of models and data assimilation systems. In this regard, the assimilation of radar reflectivity volumes should be beneficial, since the accuracy of analysis is the element that most affects short-term QPFs. Up to now, ~~very~~ few attempts have been made to assimilate these observations in an operational set-up, due to the large amount of computational resources needed and to several open issues, like the arise of imbalances in the analyses and the estimation of the observational error. In this work, ~~it is evaluated~~ we evaluate the impact of the assimilation of radar reflectivity volumes employing a Local Ensemble Transform Kalman Filter (LETKF), implemented for the convection permitting model of the COnsortium for Small-scale Modelling (COSMO). A 4 days test case on February 2017 is considered and ~~QPF is evaluated in terms of~~ the verification of QPFs is performed using the SAL technique, an object-based method which allows to ~~evaluate structure , amplitude and location of precipitation fields~~ decompose the error in precipitation fields in terms of structure (*S*), amplitude (*A*) and location (*L*). Results obtained assimilating radar reflectivity volumes are compared to those of the operational system of the Hydro-Meteo-Climate Service of the Emilia-Romagna region (Arpae-SIMC), in which only conventional data are employed ~~.~~ and latent heat nudging (LHN) is applied using surface rainfall intensity (SRI) estimated from the Italian radar network data. The impact of assimilating reflectivity volumes using LETKF in combination or not to LHN is assessed. Furthermore, some sensitivity tests are performed to evaluate the ~~impact of the~~ effects of additive inflation, of the ~~lenght~~ length of assimilation windows and of the reflectivity observational error. Finally, balance issues are assessed in terms of kinetic energy spectra and providing some examples of how these affect prognostic fields.

1 Introduction

Numerical weather prediction (NWP) models are widely used in meteorological centres to produce forecasts of the state of the atmosphere. In particular, they play a key role in the forecast of precipitation (Cuo et al., 2011), which arouses a great interest due to the many applications in which it is involved, ~~like~~ from the issue of severe weather warnings ~~or for~~ to decision making in several branches of agriculture, industry and transportation. Therefore, an accurate quantitative precipitation forecast (QPF) is of great value for society and economic activities.

In recent years, the increase of available computing resources has allowed to increment NWP ~~model~~ spatial resolution and to improve the accuracy of parametrization schemes, enabling to develop convection-permitting models (Clark et al., 2016). Despite that, QPF is still a challenge since it is affected by uncertainties in timing, location and intensity (Cuo et al., 2011; Röpnack et al., 2013). These errors arise partly from the chaotic behaviour of the atmosphere and from shortcomings in the model physics (Berner et al., 2015), but the main factor which affects the quality of QPF, especially in the short range (3-12 hours), is the accuracy of initial conditions (Dixon et al., 2009; Clark et al., 2016).

The initial condition (analysis) is generally produced by a data assimilation procedure which combines model state (background or first guess) and observations to provide the best estimate of the actual state of the atmosphere at a given time. In the last decades, different assimilation schemes have been proposed and implemented operationally in meteorological centres around the world. They can be divided in ~~two~~ different families: those based on a variational approach, like three-dimensional variational data assimilation (3D-Var: Courtier et al., 1998) and four-dimensional variational data assimilation (4D-Var: Buehner et al., 2010b)~~and~~ , those based on the ensemble Kalman filter (EnKF: Evensen, 1994; Houtekamer and Mitchell, 1998) and those based on the particle filter (PF; see van Leeuwen, 2009 for a review). At the convective scale, EnKF methods seem to be preferable to variational schemes (Schraff et al., 2016). In fact, they determine explicitly the background error covariance, which is highly flow-dependent at the convective scale. Furthermore, in a variational scheme it is not straightforward to update any variable of a NWP model since an explicit linear and adjoint relation to the control vector of prognostic variables is needed. These problems are partly addressed by employing hybrid EnKF-Variational techniques (like Wang et al., 2008; Gustafsson and Bojarova, 2014) but these approaches have mostly been applied to larger scale NWP. Another option may be to employ particle filters but, despite the efforts to overcome the dimensionality challenges of these assimilation techniques (e.g. Poterjoy, 2016), it is still not feasible for operational applications. Several variants of EnKF have been suggested (for a survey refer to Meng and Zhang, 2011) and one of the most popular is the local ensemble transform Kalman filter (LETKF), proposed by Hunt et al. (2007). It is used operationally in several meteorological centres like at COMET (Bonavita et al., 2010), at MeteoSwiss employing the version of the scheme developed for the COSMO consortium (Schraff et al., 2016) and for research purposes at the Japan Meteorological Agency (JMA; Miyoshi et al., 2010) and at the European Centre of Medium-Range Weather Forecasts (ECMWF; Hamrud et al., 2015)

The quality of the analysis is not determined only by the data assimilation scheme employed, but also by the quality and amount of observations that can be assimilated. With this aim, the assimilation of radar observations can be very beneficial, since they are highly dense in space (both horizontally and vertically) and in time. Up to now, several attempts have been made to improve the quality of analyses and subsequently the accuracy of QPFs by assimilating rainfall data estimated from radar reflectivity observations (Jones and Macpherson, 2006; Leuenberger and Rossa, 2007; Sokol, 2009; Davolio et al., 2017). Conversely, only ~~recently EnKF techniques have been employed~~ few tries have been made to directly assimilate reflectivity volumes in a convection permitting model ~~(Bick et al., 2016) with~~ employing EnKF techniques (e.g. Snyder and Zhang, 2003), especially in an operational framework (Bick et al., 2016). Despite some promising results~~. However~~, many issues affect the assimilation of reflectivity volumes at high spatial resolution and several aspects need to be further investigated.

First of all, the length of the assimilation window, which is one of the key aspects of any data assimilation system, has to be examined. In EnKF methods, a short window would be desirable to avoid that dynamical features leave the area where computed localized increments are significant (Buehner et al., 2010a) and to better preserve the gaussianity of the ensemble which can be compromised by non-linearities (Ferting et al., 2007). On the other hand, a too short window would lead to an increase of imbalances in the analysis, since the model has no the time to filter spurious gravity waves, introduced at each initialization, throughout the forecast step of the assimilation cycle. When reflectivity volumes are assimilated, the window length becomes even more crucial since these observations allow to catch small scale features of the atmosphere (Houtekamer and Zhang, 2016). In order to exploit the high temporal frequency of these data, which is essential to properly characterize fast developing and moving precipitation systems, it seems reasonable to employ short windows to assimilate, in each cycle, only observations collected very close to the analysis time. Furthermore, the choice of a short window is encouraged by the use of ~~a severe spatial localization~~ short localization scales, which has to be employed since small scales features are observed~~(Houtekamer and Zhang, 2016)~~. Conversely, the big amount of radar observations enhances the ~~instability~~ imbalance issue and, therefore, the ~~instability~~ imbalances generated in the model by each initialisation should be checked and kept under control.

Another important issue is how to determine the observational error for radar reflectivities. As for any other observation, this is influenced by three different sources: instrumental errors, representativity errors and observation operator errors. Since none of these ~~is~~ are known, the choice of its value is not straightforward and can be estimated only in a statistical sense. Considering the amount of radar data, a correct estimation of the observational error is crucial, since even a small departure from the correct value can have a large impact on the quality of the analyses. ~~Finally~~Moreover, it should be taken into account that the use of the radar data is highly dependent on the observation operator adopted and its biases should also be studied and ideally removed. Finally, a further challenge is the estimation of the observational error correlation especially when dealing with radar data assimilation, due to the high density of this type of observations..

At Arpae-SIMC, the Hydro-Meteo-Climate Service of the Emilia-Romagna region, in Italy, a LETKF scheme is used to provide the initial conditions to the convection-permitting components of the operational modeling chain, consisting of one deterministic run and of one ensemble system both at 2.2 km of horizontal resolution. Currently, only conventional data are assimilated throughout the LETKF scheme and latent heat nudging ((LHN; Stephan et al., 2008) is performed using rainfall intensity estimated from the Italian radar network data. The purpose of this paper is to present the first results obtained ~~by assimilating in this scheme the reflectivity volumes of the Italian radar network~~when also reflectivity volumes are assimilated using the LETKF scheme. In particular, the impact of assimilating reflectivity volumes in combination or not with LHN is evaluated. Furthermore, it is studied the sensitivity of the obtained analysis to two important characteristics of the assimilation cycle: the length of each cycle and the observational error attributed to the radar reflectivities.

This paper is organised as follows. In section 2, the model and the data assimilation system employed are described, as well as the observations employed and the set-up of the experiments performed. Furthermore, the verification method is explained. In section 3 results are shown and discussed. In section 4 some conclusions are drawn.

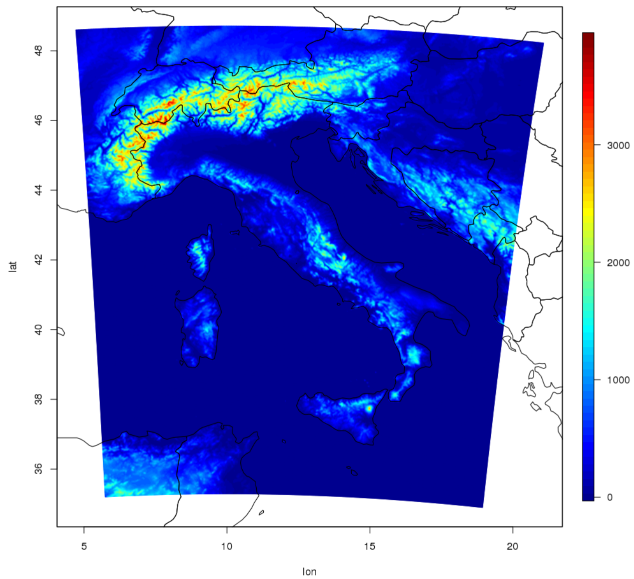


Figure 1. Integration domain and corresponding orography of the COSMO model employed in this study.

2 Data, model and methodology

2.1 The COSMO model

The COSMO model (Baldauf et al., 2011) is a non-hydrostatic limited-area model developed by the multi-national COnsortium for Small-scale Modelling (COSMO) and it is designed for both operational NWP and several research applications. It is based on the primitive equations describing compressible flows in a moist atmosphere and the continuity equation is replaced by a prognostic equation for the pressure perturbation (deviation from a reference state). The prognostic variables involved in these equations are the three dimensional wind vector, temperature, pressure perturbation, turbulent kinetic energy (TKE) and specific amount of water vapour, cloud water, cloud ice, rain, snow and graupel.

In the present study, the COSMO model is run at 2.2 km horizontal resolution over a domain covering Italy and part of the neighbouring countries (Figure 1) and employing 65 terrain-following hybrid layers. The model top is at 22 km.

Regarding set-up and parametrizations, deep convection is resolved explicitly while the shallow convection is parametrized following the non-precipitating part of Tiedtke scheme (Tiedtke, 1989). Cloud formation and decay is controlled by a Lin-type one moment bulk microphysics scheme which includes all the prognostic microphysical species (Lin et al., 1983; Seifert and Beheng, 2001). The turbulent parametrization is based on a TKE equation with a closure at level 2.5, according to Raschendorfer (2001). Radiative effects are described by the δ-two-stream radiation scheme of Ritter and Geleyn (1992) for short-wave and long-wave fluxes. Finally, the lower boundary conditions at the ground are provided by the multi-layer soil model TERRA (Doms et al., 2011).

2.2 The KENDA system

The KENDA system (Schraff et al., 2016) implements for the COSMO model the LETKF scheme described by Hunt et al. (2007). In this implementation, the method is fully four dimensional, that is all observations collected during the assimilation window contribute to determine the analysis and the related model equivalents are computed using the prognostic variables at the proper observation time. To avoid spurious long-distance correlations in the background error covariance matrix, analyses are performed independently for each model grid point taking into account only nearby observations (observation localization). Observations are weighted according to their distance from the grid point considered using the Gaspari-Cohn correlation function (Gaspari and Cohn, 1999). In the present work, two different values of the Gaspari-Cohn localization length-scale are employed for conventional and radar observations: 80 km for the former, 16 km for the latter (as done by Bick et al., 2016).

The limited size of the ensemble, combined to the assumption of a perfect model made in the LETKF scheme, leads to an underestimation of the background and analysis variances (e.g. Anderson, 2009) and, as a consequence, the quality of analyses is negatively affected. To address this issue, KENDA provides some techniques to enlarge the spread of the ensemble (for a complete description of each of them refer to Schraff et al., 2016). Here, multiplicative covariance inflation (Anderson and Anderson, 1999) and the relaxation to prior perturbation (RTPP; Zhang et al., 2004) are employed. The former consists in inflating the analysis error covariance by a factor ρ greater than one which is estimated following Houtekamer et al. (2005). The latter lies on the relaxation of the analysis ensemble perturbations towards the background ensemble perturbations by replacing at each grid point the analysis perturbation matrix in ensemble space $\mathbf{W}^a$ by

$$(1-\alpha_p)\mathbf{W}^a + \alpha_p\mathbf{I} \tag{1}$$

where $\mathbf{I}$ is the identity matrix and $\alpha_p = 0.75$ (see also Harnisch and Keil, 2015). Another approach provided by KENDA to account for model error is the additive inflation. The basic idea is to add random noise with mean $\mathbf{0}$ and covariance $\mathbf{Q}$ to the analysis ensemble members, where $\mathbf{Q}$ is the model error covariance matrix (Houtekamer and Mitchell, 2005). Since $\mathbf{Q}$ is not known, it is assumed to be ~~proportion~~ proportional (by a factor smaller than 1) to a static background error covariance $\mathbf{B}$ (Mitchell and Houtekamer, 2000). ~~Additive~~ This technique has already been employed with a positive impact in convective scale data assimilation (e.g. Dowell and Wicker, 2009). In the present work, additive inflation is used together with multiplicative inflation and to RTPP only in one experiment, ~~using~~ employing a climatological $\mathbf{B}$-matrix from the 3D-VAR of the ~~ICON~~ Icosahedral Nonhydrostatic (ICON) global model (Zängl et al., 2015). Although the use of a lower resolution $\mathbf{B}$-matrix may not allow to properly characterize the model error at the smallest scales, the same configuration has been gainfully employed at MeteoSwiss (Leuenberger and Merker, 2018).

The KENDA suite also allows to compute the analysis weights on a coarsened grid (Yang et al., 2009). Weights computed on this coarsened grid are then interpolated to the model grid and afterwards used to calculate analysis increments. In this way, the computational cost is decreased without affecting negatively the accuracy of analysis (Yang et al., 2009). In the present study, a coarsening factor equal to 3 is employed.

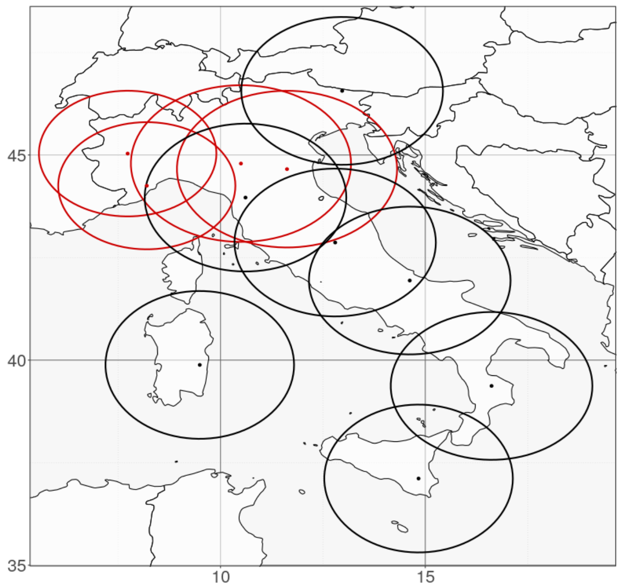


Figure 2. ~~Domain covered by the composite of the~~ The Italian radar network. For each radar the approximate coverage area is shown. All radars contribute to the composite generation employed in LHN. Radars highlighted in red are used to directly assimilate reflectivity volumes throughout KENDA.

2.3 Assimilated data

KENDA allows the assimilation of both conventional and non conventional observations.

Conventional observations assimilated in this work include aircraft measurements (AMDAR) of temperature and horizontal wind, surface station measurements (SYNOP) of 10 m horizontal wind, 2 m temperature, 2 m relative humidity and surface pressure, radiosonde data (TEMP) of temperature, horizontal wind and humidity.

~~Fields of surface rainfall rate (SRI) are also assimilated in each member of the assimilation ensemble using a Latent Heat Nudging Scheme (LHN; Stephan et al., 2008). SRI data come from the composite of Italian radar network and are distributed by the National Department of Civil Protection and cover the area of Figure 2.~~

~~These data have a temporal resolution of 10 minutes and a spatial resolution of 1 km, but before the assimilation they are interpolated at the model resolution. The composite is obtained as a weighted average of surface rain rates from single stations, where weights are represented by quality. Data coming from each radar station undergo:) a quality control that removes data with low quality depending in particular on the distance from the radar station itself. These fields are assimilated through the Latent Heat Nudging scheme, based on the fact that the latent heat, integrated along the vertical column, is approximately proportional to the precipitation observed. It acts in rescaling temperature profiles with an adjustment of the humidity field according to the ratio between observed and modelled rain rates. LHN is applied continuously during the integration of the~~

~~model and it is able to capture the dynamical structure and the right rainfall amount of the storm in a perfect environment and it can be gainfully used also in complex terrain (Leuenberger and Rossa, 2004; Leuenberger and Rossa, 2007). In the KENDA framework this allows to have the model first guess closer to the observed atmospheric state, improving the analysis quality.~~

With regards to non conventional observation, KENDA allows also the assimilation of radar reflectivity volumes and radial winds. Radar data are assimilated through the Efficient Modular VOlume RADar Operator (EMVORADO) expressly designed for the COSMO model. It simulates the radar reflectivity factor and radial velocities processing the COSMO model fields one radar system at a time. Operator characteristics, resolution and the management of no-precipitation information are described in (Bick et al., 2016).

~~In the present work, radar reflectivities~~ Although the operator gives the possibility to assimilate both radial winds and reflectivities, in the present work only reflectivity volumes are assimilated~~, but only on Northern Italy. Radar reflectivity~~. Reflectivity volumes come from four different radar stations over Northern Italy (red circles in Figure 2): Bric Della Croce (Piedmont Region), Settepani (Liguria Region), Gattatico and San Pietro Capofiume (Emilia-Romagna Region). Due to the complex orography of the considered area, radar are placed at very different altitudes and have different acquisition strategies. Observations are acquired every 10 minutes for Bric Della Croce radar, every 5 minutes for Settepani radar, every 15 minutes for San Pietro Capofiume radar and every 15 minutes starting from minutes ~~05~~ 5 and 10 of each hour for Gattatico radar. ~~At lowest beam elevation angle the covered domain is presented in Figure ??.~~

~~Domain covered by radar systems used in this study at their lowest elevation beam angle~~

~~Raw reflectivities are pre-processed to eliminate non meteorological echoes, in particular, to eliminate the clutter signal that would affect the analysis retrieval introducing spurious observations.~~ Data have a range resolution of 1 km, while the azimuthal resolution is 1 degree for Bric Della Croce and Settepani and 0.9 degree for San Pietro Capofiume and Gattatico. ~~Due to the~~ Before assimilation raw reflectivity are pre-processed taking into account non meteorological echoes, beam blocking and attenuation to improve the quality of data. In particular, it is important to eliminate the clutter signal that would affect the analysis retrieval introducing spurious observations. However, due to the fact that volumes from single radars undergo different pre-processing, it is not possible to define a homogeneous quality criterion. For this reason, all data in the volume ~~,~~ that are not rejected from pre-processing step ~~,~~ are supposed to have the same quality and are used into the assimilation cycle.

The high temporal and spatial density of observations is precious to estimate the initial state of numerical weather forecast. This allows to gather a lot of information on the real state of the atmosphere, but it determines an increase in analysis computational cost, in data transfer time and in memory disk occupation. Moreover, a spatial and/or temporal high density violates the assumption made in the most part of assimilation schemes: the non-correlation of observational errors. To reduce the total amount of data and to extract essential content of information, the superobbing technique is chosen. In this way, reflectivities over a defined area are combined through a weighted mean into one single observation representative of the desired greater spatial scale. As in Bick et al. (2016), the horizontal resolution chosen in this work for the superobbing is equal to 10 km. Furthermore, before performing superobbing on the observed and simulated fields, a threshold of 5 dBZ is applied to ~~observed and simulated~~ both fields in order to avoid that large innovations associated to non-precipitating signals would lead to large analysis increments without physical relevance.

To evaluate the observational error associated to reflectivity volumes, a diagnostic based on statistical averages of observations-minus-background and observations-minus-analysis residuals, as described in Desroziers et al. (2005), is used. ~~The estimated value is, on average along the vertical,~~ Employing all radar data available during the test case, a reflectivity observational error (*roe*) equal to 5 dBZ is estimated, as found also by Tong and Xue (2005).

Finally, fields of surface rainfall intensity (SRI) are also assimilated in each member of the assimilation ensemble using a latent heat nudging scheme. SRI data come from the composite of the Italian radar network (all circles in Figure 2) and are distributed by the National Department of Civil Protection. These data have a temporal resolution of 10 minutes and a spatial resolution of 1 km, but before the assimilation they are interpolated at the model resolution. Data coming from each station undergo a quality control that removes those with low quality. The quality depends on different factors such as ground clutter, beam blocking, range distance, vertical variability and attenuation as described in Rinollo et al. (2013). The composite is then obtained as a weighted average of surface rain rates from single radar stations, where weights are represented by quality. These fields are assimilated through the LHN scheme, based on the assumption that the latent heat, integrated along the vertical column, is approximately proportional to the precipitation observed. The scheme, which is applied continuously during the integration of the model, acts in rescaling temperature profiles with an adjustment of the humidity field according to the ratio between observed and modelled rain rates. LHN has been gainfully employed in different frameworks, including forecasts over complex terrain (Leuenberger and Rossa, 2004; Leuenberger and Rossa, 2007). Our hypothesis is that, in the KENDA framework, LHN allows to have the model first guess closer to the observed atmospheric state, improving the analysis quality. For this reason, in all experiments (except one) presented here, LHN is applied together to the direct assimilation of reflectivity volumes through KENDA.

2.4 Experimental set-up

The KENDA system is implemented operationally at Arpae using an ensemble of 20 members plus a deterministic run, which is obtained by applying the Kalman gain matrix for the ensemble mean to the innovations of the deterministic run itself. In principle, ensemble mean analyses can be deployed to initialize the deterministic forecasts, but this would lead to some inaccuracies since the mean of a non-Gaussian ensemble is generally not in balance (Schraff et al., 2016). For this reason the deterministic branch is added to the system, which differs from the ensemble ones only due to boundary conditions. The ensemble members use lateral boundary conditions provided each 3 hours at a 10 km horizontal resolution by the ensemble of the data assimilation system of the Centro Operativo per la Meteorologia (COMet), based on a LETKF scheme (Bonavita et al., 2010). The deterministic run employs hourly boundary conditions provided by a 5 km version of COSMO run at Arpae (COSMO-5M) which domain covers a large part of the Mediterranean basin ans surrounding countries.

In the operational set-up employed at Arpae, the COSMO model configuration described in Section 2.1 is adopted for all the 21 members. At present, in the operational chain only conventional observations are assimilated and LHN is performed on each member of the ensemble. The KENDA analyses are used operationally to provide initial conditions to COSMO-2I, the 2.2 km determinstic run initialized twice a day at 00 UTC and 12 UTC and to COSMO-2I EPS, an ensemble which is run every day at 00 UTC for a 48 hours forecast range. In this work, deterministic forecasts starting from the KENDA deterministic analysis are also performed, in order to evaluate the quality of the analysis also from its impact when used to initialise a forecast.

To evaluate the impact of the assimilation of reflectivity radar volumes, several experiments are performed employing different configurations. The complete list is provided in Table 1. In the control experiment, called *conv60*, the set-up of the operational chain~~is replicated:~~ , described in the previous two paragraphs, is replicated. In particular, this means that in *conv60* experiment only conventional data are assimilated ~~in KENDA in~~ using KENDA through cycles of 60 minutes and the LHN is performed during the forecast step of each assimilation cycle. ~~All the other experiments involve the assimilation of both conventional data and reflectivity volumes, in addition to LHN.~~ In the *rad60* experiment, radar measurements are assimilated using a reflectivity observation error (*roe*) of 5 dBZ and a 60 minutes assimilation window is employed. A comparison with *conv60*, from which *rad60* differs only due to the inclusion of radar data in the KENDA system, allows an assessment of whether, under the same conditions, the assimilation of reflectivity observations improves the quality of analyses. ~~The~~ In *rad60_nolhn* and *rad60_Bm* experiments the same set-up of *rad60* is employed~~in the experiment *rad60_Bm*, but~~ , but in the former the LHN procedure is switched off in order to assess the impact of the assimilation of radar data only by means of KENDA, while in the latter additive inflation is applied to increase the spread of the ensemble.

Trial	Window length [min]	Assimilated obs.	roe [dBZ]	Note
conv60	60	conv.	-	-
rad60	60	conv. + radar	5	-
rad60_nolhn	60	conv. + radar	5	No LHN
rad60_Bm	60	conv. + radar	5	Additive inflation
rad30	30	conv. + radar	5	-
rad15	15	conv. + radar	5	-
rad60_lst15	60	conv. + radar	5	Use obs. in the last 15 min. of the window
rad60_roe10	60	conv. + radar	10	-
rad60_roe0.5	60	conv. + radar	0.5	-
rad15_roe10	15	conv. + radar	10	-
rad15_roe0.5	15	conv. + radar	0.5	-

Table 1. Experimental set-up of each ~~trial~~ experiment including the length of the assimilation cycles, the type of observations assimilated, the reflectivity observation error (*roe)* associated to radar data and any additional feature.

All the other experiments involve the assimilation of both conventional data and reflectivity volumes, in addition to LHN. In order to test the impact of assimilating only observations which are not too far from the analysis time, experiments on the duration of the assimilation windows are performed. This is tested by comparing *rad60* to experiments *rad30* and *rad15* which differ from *rad60* only for the length of the assimilation window, equal to 30 and 15 minutes respectively. An alternative way to assimilate only the most relevant observations is to select in each cycle a subset of data including the closest to the analysis time. In the experiment *rad60_lst15* an assimilation window of 60 minutes is employed but only the observations (both conventional and radar reflectivities) collected in the last 15 minutes of the cycle are taken into account.

Since the estimation of observation error is not straightforward and different techniques can be applied, it is worth to evaluate the sensitivity of the assimilation system to this parameter. In addition to the value of 5 dBZ employed in the previous experiments, two other values are selected: 10 dBZ or 0.5 dBZ. Both of them are tested employing a 60 minutes assimilation window (*rad60_roe10* and *rad60_roe0.5*) and using 15 minutes cycles (*rad15_roe10* and *rad15_roe0.5*).

The experiments described above are carried out over a period of almost 4 days from ~~3 February 2017~~ February 3rd at 06 UTC to ~~7 February 2017~~ February 7th at 00 UTC ~~.~~ in 2017. During 3 and 4 February, middle tropospheric circulation over Northern and Central Italy was dominated by southwesterly divergent flows associated with the passage of some precipitating systems. In 5 February a trough moved from France to Italy and this caused the formation of new precipitations in Northern Italy. During 6 February the trough moved slowly from Central Italy to the southern part of the country and ~~precipitations~~ precipitation systems weaken gradually. For each experiment, a set of 5 forecasts up to 48 hours is initialized using the analyses generated during the assimilation procedure. Initialization times employed are 04 February at 00 and 12 UTC, 5 February at 00 and 12 UTC and 6 February at 00.

2.5 Verification

The performance of each experiment described in the previous section is assessed in terms of QPF employing two methods. The first, employed for the verification during assimilation cycles, consists of comparing the ~~area-average~~ areal average values of 3-hourly precipitation: average precipitation forecasted by the model over an area is compared against the average precipitation observed by raingauges over the same area. The raingauge stations used in this work (nearly 1500~~stations) over the same area~~) are shown in Figure 3; note that they are approximately in the region where reflectivity volumes are assimilated. In order to have comparable samples, ~~precipitation forecasted by the model~~ model precipitation is first interpolated on station location ~~(~~by selecting the value at the nearest grid point~~).~~ .

The second method, applied for the verification of forecasts, employs the SAL metrics (Wernli et al., 2008), an object based verification score which allows to overcome the limitations of traditional scores for convection-permitting models, like the double-penalty problem (Rossa et al., 2008). The detection of individual objects in the accumulated precipitation fields is achieved by considering continuous areas of grid points exceeding a selected threshold. Comparing objects from observed and forecast fields, SAL provides information about the structure S, the amplitude A and the location L errors of QPF. A perfect match between forecast and observations would lead to $S = A = L = 0$; the more values differ from 0, the greater the disagreement between model and observations. More in detail, a too sharp/flat (broad/small) structure of forecast precipitation compared to observations is associated to positive (negative) values of S; an overestimation (underestimation) of average rainfall over the domain is associated to positive (negative) values of A; a misplacement of precipitation nuclei leads to positive values of L. Note that L can range between 0 and 2, while S and A between -2 and 2.

Observations employed to perform SAL consist in 3-hourly accumulated precipitation estimated from the Italian radar network (SRI) and corrected using rain-gauges data. The ~~radar estimates are derived from the SRI product which is assimilated~~

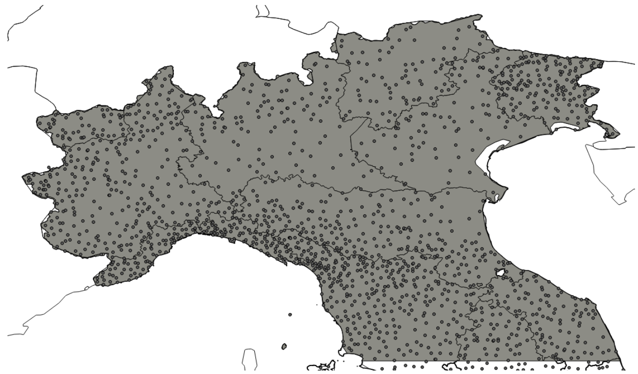


Figure 3. Raingauges (dots) used for the verification of area-average precipitation during the assimilation procedure. In gray the domain employed for SAL verification is depicted.

~~in each run with the LHN method, therefore they are not independent observations. Nevertheless, the assimilation takes place in all experiments, making possible the use of this product for the evaluation of their relative performances. The~~ radar-raingauges adjustment, adapted for a radar composite, derives from the method described in Koinstinen and Puhakka (1981). The original method comprises two terms: a range dependency adjustment and a spatial varying adjustment. In our case, only the second term is taken into account due to the fact that, in overlapping areas of the composite, rainfall estimation is obtained combining data from different radars and, therefore, the original information on the range distance from the radar is lost. The correction is based on a weighted mean of the ratio between rain gauges and estimated radar rainfall amount calculated over the station locations. Weights are a function of the distance of the grid point from the station and of a filtering parameter calculated as the mean spacing between 5 observations. Then a smoothing factor is applied to the correction.

The verification area is shown in gray in Figure 3.

~~Verification domain over Northern Italy is highlighted in dark grey~~

This choice is made to assess the impact of the assimilation of radar reflectivity volumes in the region where these data are actually observed. Furthermore, regarding SAL, in Wernli et al. (2009) it is recommended to use a domain not larger than 500×500 km^2 since, otherwise, the domain may include different meteorological systems making the interpretation of results problematic. In fact, if the domain contains strongly differing meteorological systems, then results obtained using the SAL technique may not be representative for the weakest one.

3 Results

3.1 Impact of assimilating the radar reflectivities

A preliminary assessment of the impact of assimilating radar reflectivity volumes with the KENDA system is provided by comparing *conv60*, in which only conventional observations are employed, and *rad60*, in which radar reflectivity data are

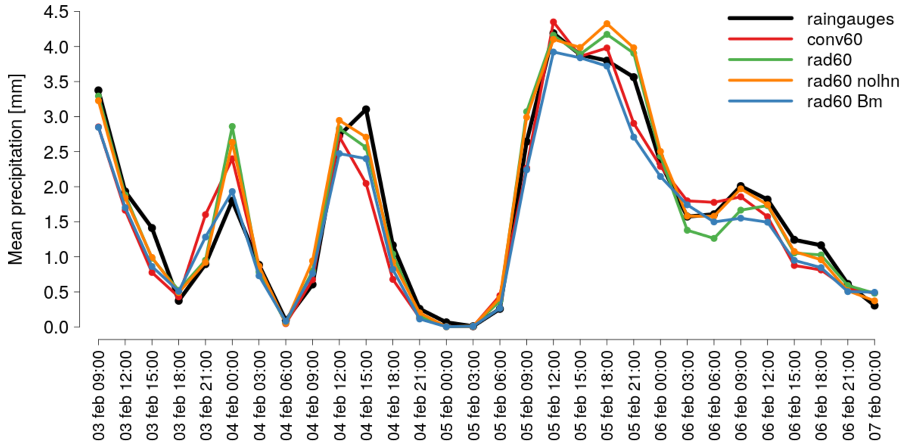


Figure 4. ~~Area-average~~ Areal average 3h precipitation for rain-gauges (black) in the verification area shown in Figure 3 and at the corresponding model forecast relative to experiments *conv60* (red), *rad60* (green), *rad60_nolhn* (orange) and *rad60_Bm* (blue), during the assimilation procedure.

added. It is reminded that LHN ~~of radar precipitation estimates~~ using SRI data is also applied in both experiments. ~~Area-average~~ Areal average 3-hourly precipitation forecasted during the assimilation procedure is displayed in Figure 4 for ~~three~~ these experiments, employing precipitation recorded by rain-gauges (black line) as independent observation. Since the duration of each assimilation cycle is 1 hour, the precipitation forecasted by the model during each hour is accumulated in order to obtain the 3-hourly precipitation. Overall, the correspondence of *rad60* (green, which will be used from here onwards to identify uniquely this experiment) to observations is equal or better than that of *conv60* (red). In some cases the improvement is particularly relevant, like at 21 UTC on 3 February and at 15 UTC on 4 February. Only at 00 UTC on 4 February the performance of *conv60* is clearly better than that of *rad60*.

Since in *rad60* experiment both reflectivity volumes and surface rainfall intensity are assimilated (the former using KENDA, the latter by LHN), the impact of assimilating reflectivity volumes may be hidden. To avoid this and to assess the influence of reflectivity observations in the assimilation procedure, LHN is switched off in *rad60_nolhn* experiment. Results displayed in Figure 4 show that average precipitation during assimilation cycles of *rad60_nolhn* (orange line) is similar to that of *rad60* with an improvement on February 6th between 03 UTC to 09 UTC.

The same set-up of *rad60* is used in the experiment *rad60_Bm*, but with the addition of additive inflation to enlarge the ensemble spread. Mean precipitation during the assimilation cycles (blue line in Figure 4) differs from that of *rad60* but, since slight improvements at some instants are compensated by slight deteriorations at others, the overall impact of the use of additive inflation ~~seems to be neutral~~cannot be judged.

For the five forecasts initialized from the analyses of each experiments, the precipitation is verified using SAL and employing a 1 mm threshold to identify rainfall objects. Verification using a 3 mm threshold was also performed but, since results do not differ significantly from those obtained with a 1 mm threshold, they are not shown here. In Figure 5 the average of the absolute

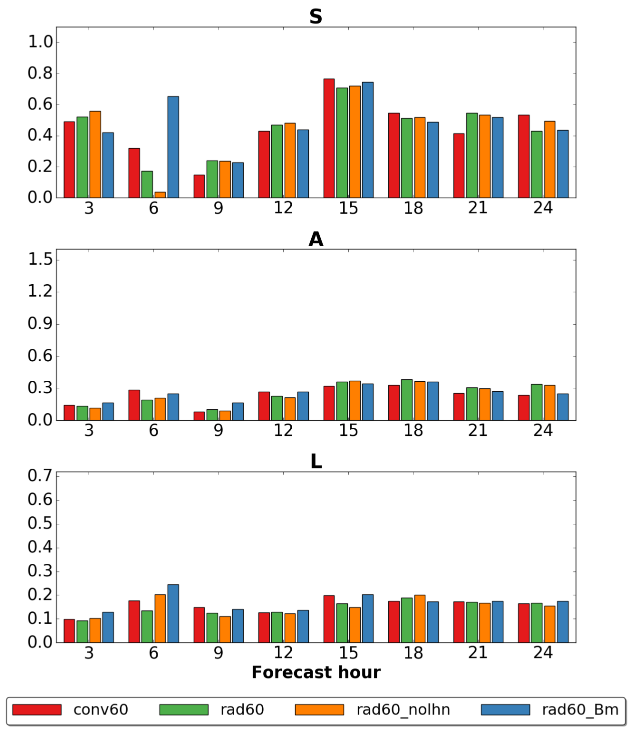


Figure 5. Average of the absolute value of each component of SAL over the 5 forecasts initialized from *conv60* (red), *rad60* (green), *rad60_nolhn* (orange) and *rad60_Bm* (blue) analyses. Three-hourly precipitation is considered employing a threshold of 1mm. Cases in which the observed precipitation field consist of less than 1000 points are not taken into account in the average.

value of each component of SAL is plotted as a function of lead time. Although forecasts are up to 48 hours, the verification is shown only for the first 24 hours, since after this lead time scores of the different experiments become very close. The average is computed considering only cases in which the observed rainfall field consists of at least 1000 grid points (3 events at lead time +6h, 4 otherwise), which is approximately equal to an area of $50 \times 50\mathrm{km}^2$. Using the absolute value of the components of SAL, only the magnitude of the error is considered, loosing the information on the type of error (e.g., for *A*, an overestimation of forecast precipitation cannot be distinguished from an underestimation). This choice slightly limits the potential of SAL but provides an intuitive picture of the overall performance of each experiments (a similar approach is employed by Davolio et al., 2017). Differences between forecasts initialized from *conv60* and from *rad60* analyses are very small. Overall, the location of precipitations (*L* component) of *rad60* forecasts is only slightly improved compared to *conv60* forecasts. The amplitude error *A* is generally smaller in the first 12 hours, but from 15h onward the *conv60* forecasts outperform the *rad60* forecasts. Regarding the structure component *S*, smaller errors for *rad60* forecasts at some lead times are counterbalanced by smaller errors for *conv60* at other lead times, in a non coherent way. Therefore, even if the use of analyses obtained by assimilating reflectivity volumes affects the structure of forecast precipitation, it is not possible to state if it is improved or deteriorated.

~~The~~ When LHN is not performed and radar data are assimilated only using KENDA, results (orange bars in Figure 5) are not significantly affected. Only a meaningful improvement can be noticed in the structure component at lead time +6h, but, at the same time, the error in the location component is increased. Since the combined assimilation of reflectivity volumes with KENDA and SRI by LHN does not have a negative impact on the quality of precipitation forecasts compared to the assimilation of only reflectivity observations, it is decided to not switch off the LHN for the other experiments. In fact, this choice does not affect the results of the sensitivity tests that are presented in this work and, at the same time, the LHN allows to use radar derived information on the state of the atmosphere in the whole Italian country, despite reflectivity volumes can be assimilated, at present, only over Northern Italy.

Finally, the addition of the additive inflation (*rad60_Bm*, blue bars in Figure 5) does not show a positive impact. At lead time +6h a clear worsening of each of the 3 components of SAL can be noticed.

3.2 Impact of the length of the assimilation cycles

To obtain some insights about this topic, assimilation cycles of 15 and 30 minutes (respectively *rad15* and *rad30*) are tested and the results are compared with those obtained with the 60 minutes window (*rad60*). In the same way described in the previous subsection, SAL verification is computed and averaged over the 5 forecasts initialized from the analyses of each trial. Results are shown in Figure 6 where the green bars are associated to *rad60* while red and orange to *rad15* and *rad30* respectively. Regarding the location error, during the first 6 hours of forecast the performance of *rad15* is similar to that of *rad30* and both are worse than *rad60*. Afterwards, the 3 experiments provides a very similar performance. In terms of structure and amplitude components, the differences among the 3 different window lengths are generally small and non coherent over the 24 hours of forecast.

In a further test, conventional and radar observations are assimilated only if collected during the last 15 minutes of each assimilation cycle of 60 minutes (*rad60_lst15*). In this way, the total amount of assimilated data is reduced and the increments computed by the LETKF scheme should be more appropriate for computing the analysis, since the observations time is always very close to the analysis time. Actually, verification shown in Figure 6 (blue bars) does not point out an improvement of *rad60_lst15* if compared to *rad60*. In fact, except for a worsening in the *S* component at forecast time +3h and +6h, performance of both experiments is very similar. Therefore, the assimilation of data closer to analysis time does not improve the forecast quality.

In order to evaluate the ~~instability~~ imbalance issue, the kinetic energy (KE) spectra of the experiments is computed following the method described in Errico (1985). Curves displayed in Figure 7 are obtained as an average over the whole assimilation period (from 3 February at 06 UTC to 7 February at 00 UTC) of KE spectra computed each hour using analysis values of u, v and w over the whole domain. Kinetic energy spectra of *rad15* (red) and *rad60* (green) are almost overlapping, even at very small wavelength, indicating that shortening the length of cycles from 60 to 15 minutes does not introduce imbalances in the analyses (Skamarock, 2004). Furthermore, both spectra have a $-5/3$ dependence on the wavenumber beyond a wavelength of 15-20 km, in agreement with observed spectra at the mesoscale (Nastrom and Gage, 1985). Same considerations apply also to KE spectra of *rad30* and to *rad60_lst15*, which are not shown.

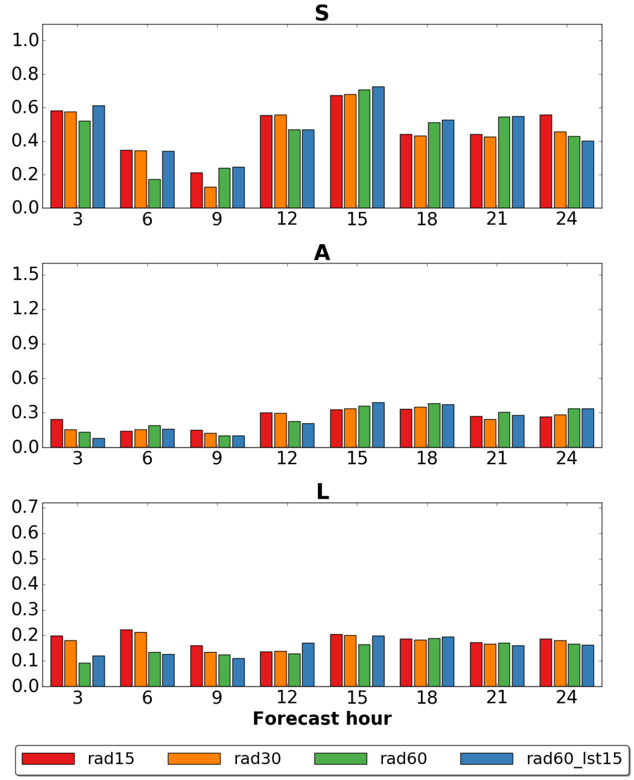


Figure 6. As in Figure 5 but considering experiments *rad15* (red), *rad30* (orange), *rad60* (green) and *rad60_lst15* ~~)~~(blue).

As a conclusion, with the current set-up, the use of a sub-hourly window length does not introduce imbalances in the analysis. It is anyway observed a slightly worsening of the precipitation forecast, especially in terms of location of rainfall nuclei.

3.3 Impact of changing the reflectivity observational error

A set of experiments is performed to investigate the impact of the reflectivity observation error in the assimilation scheme. In addition to the value of 5 dBZ employed so far, which was estimated applying the diagnostic described in Desroziers et al. (2005) to this case study, two other values of *roe* are tested: 10 dBZ and 0.5 dBZ. The former is employed by Bick et al. (2016) for the assimilation of reflectivity volumes from the German radar network using KENDA and COSMO and, therefore, should be reasonable also for the present study. The latter is a deliberately extreme value that may be chosen in the case of a great confidence in the quality of radar observations. These two different values of *roe* are used in assimilation cycles of 60 minutes (*rad60_roe0.5* and *rad60_roe10*) and 15 minutes (*rad15_roe0.5* and *rad15_roe10*). Therefore, they can be compared with the experiments with our standard value of roe $=5$dBZ, respectively *rad60* and *rad15*.

Verification of forecasts initialized from the analyses of these experiments is reported in Figure 8. Regarding the experiments with a 60 minutes assimilation cycle (left panel), the performance of *rad60_roe0.5* is clearly worse than that of *rad60*. In fact,

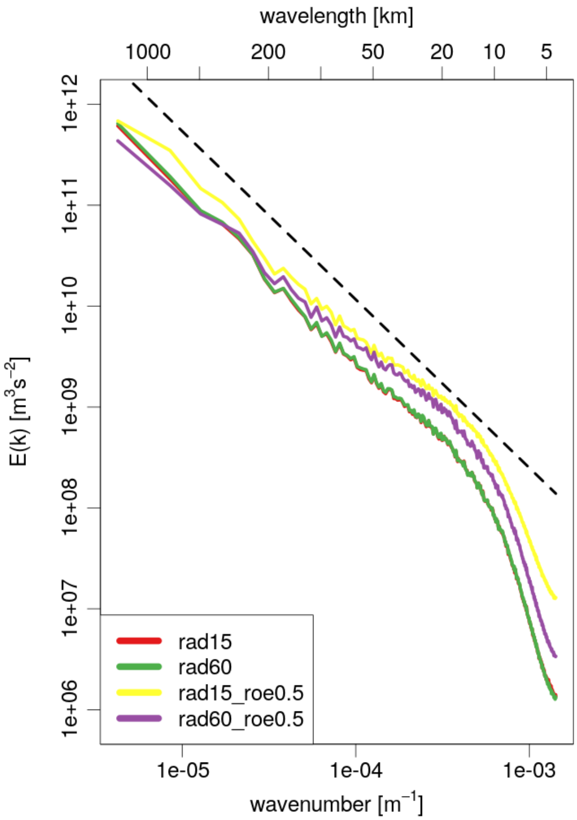


Figure 7. Kinetic energy (KE) spectra computed following the method described by Errico (1985). Each curve is obtained averaging KE spectra computed with a frequency of one hour during the assimilation procedure and employing analysis values of u, v and w over the whole model domain. The spectra are displayed for experiments *rad15* (red), *rad60* (green), *rad15_roe0.5* (yellow) and *rad60_roe0.5* (violet). The dashed black line represents a function with a dependence to the wavenumber equal to $-5/3$.

up to the lead time of +15h, each component of SAL for the former (violet) is almost equal or greater than that of the latter (green) with the only exception for S at +3h. From +18h onwards, differences between the two become very small. On the other hand, the performance of *rad60_roe10* forecasts (orange) is very similar to that of *rad60* at any lead time, apart for the error in the structure of precipitation at +6h and +9h which is significantly greater for roe $= 10$dBZ. When considering assimilation cycles of 15 minutes (right panel in Figure 8), the worsening of forecast precipitation employing a roe $= 0.5$dBZ (yellow) compared to roe $= 5$dBZ (red) is further enhanced. In particular, precipitation in the first 12 hours is largely mis-placed and its total amount is widely different from observations. In this regard, the verification of individual forecasts (not shown here) reveals that the large error in A component is due to a systematic underestimation of the average precipitation over the domain. Regarding *rad15_roe10* (blue), similarly to what observed for 60 minutes cycles, the use of roe $= 10$dBZ instead of 5 dBZ

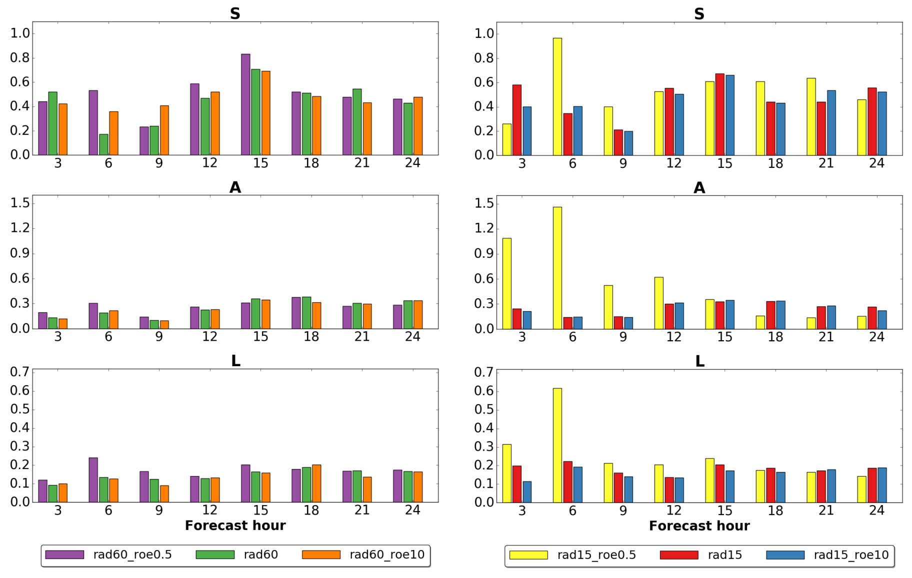


Figure 8. As in Figure 5 but considering, in the left panel, experiments *rad60_roe0.5* (violet), *rad60* (green) and *rad60_roe10* (orange) while, in the right panel, experiments *rad15_roe0.5* (yellow) *rad15* (red) and *rad15_roe0.5* (blue).

does not affect radically the quality of forecasts, even if a slight improvement in each component of SAL can be observed at +3h.

The overall poor quality of *rad15_roe0.5* forecasts is the direct consequence of the poor quality of the analyses from which they are initialized. As an example, in Figure 9 it is shown the mean sea level pressure (MSLP) and specific humidity at 850 hPa of *rad15_roe0.5* (right column) analysis on February 5 at 12 UTC and it is compared with the same quantities for the analysis of *rad60* (central column) and of the Integrated Forecasting System (IFS) of ECMWF (left column). Slight variations can be observed between IFS and *rad60* analyses and it seems reasonable that they may simply arise from differences between models and assimilation systems. Conversely, *rad15_roe0.5* analysis exhibits a noticeable increase in MSLP and a decrease in specific humidity over Northern Italy. This is in agreement with the decrease in forecast precipitation previously described.

In the same way as described in Section 3.2, KE spectra are computed for *rad15_roe0.5* and *rad60_roe0.5* and displayed in Figure 7. In both cases, at the smallest wavelength the KE is significantly grater that that of *rad15* or *rad60* and this is particularly evident for *rad15_roe0.5*. This behaviour is indicative of the presence of some undesired noise at small scales (Skamarock, 2004). Therefore, employing a value of *roe* equal to 0.5 dBZ, the assimilation system is not able to correctly remove small scale ~~structures~~noise, especially when really short cycles are employed. Furthermore, the excess of energy associated to the highest wavenumber modes propagates to the larger scales and the slope of the curves at wavelengths greater than 15 km differs from -5/3.

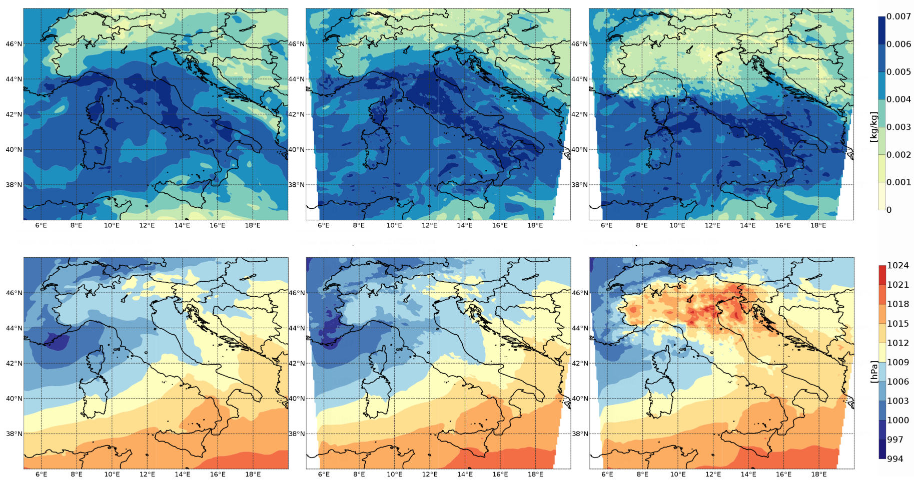


Figure 9. Mean sea level pressure (top) and specific humidity at 850 hPa analysis on February 5 at 12 UTC for IFS (left) *rad60* (middle) and *rad15_roe0.5* (right).

4 Conclusions

In the present work, the assimilation of reflectivity volumes in a high resolution model employing a LETKF scheme is evaluated. Assimilation of radar data is a challenging issue and most of the previous studies is devoted to the assimilation of rainfall estimation, while ~~very~~ few to the direct employment of reflectivity observations in an operational data assimilation system. Here, results in terms of QPF obtained assimilating reflectivity volumes from 4 radars of the Italian network are shown and compared to those produced with the current operational assimilation system of Arpae, in which only conventional data are employed. Furthermore, some sensitivity tests are performed to investigate the impact of some parameters which can substantially affect the quality of analyses.

The assimilation of radar reflectivity volumes with our selected set-up (*rad60*) only slightly improves QPF both during the assimilation procedure and for the subsequent forecasts, compared to the assimilation of only conventional data. ~~This~~ At first glance, this result could be partly ascribed by the fact that radar data are already assimilated in the *conv60* experiment, in form of LHN of radar precipitation estimate. ~~Even~~ In fact, even if the precipitation estimate is only a product derived by applying a complex algorithm to the volume of reflectivities, nevertheless its ingestion could influence the analysis so much that it is difficult to add benefit with the assimilation of reflectivities themselves. However, when LHN is not applied and radar data are assimilated only through KENDA (*rad60_nolhn*), QPF accuracy is not improved. As a consequence, at this stage, the direct assimilation of reflectivity volumes does not outperform significantly LHN and the two techniques may be applied together without loss in QPF accuracy. It has also been shown that, even if the spread of the ensemble is very small (not shown), its enlargement by employing additive inflation does not improve the performance, instead, it leads to a modest worsening of the results. One possible reason for this behaviour can be the use of a climatological $\mathbf{B}$ matrix generated from the ICON model run at a very different resolution. This test should be repeated when a $\mathbf{B}$ matrix provided by the COSMO model over the Italian domain will be available.

For the case study considered in this work, the assimilation of data close to analysis time (at most collected 15 minutes before) does not improve the quality of forecast obtained when all observations collected in the whole assimilation window are employed. Nevertheless, this results suggests that this configuration can be employed without evident downsides to reduce the computational cost of the KENDA system. Further tests would be necessary to evaluate if the same conclusion arises when only observations at the analysis time are assimilated. Investigation of the instabilities generated with short assimilation cycles shows that the use of a sub-hourly window length does not introduce imbalances in the analysis, but it slightly worsens the forecast of precipitation, especially in terms of location of rainfall nuclei.

With regards to the observational error, it is found that a value of *roe* equal to 0.5 dBZ negatively affects the quality of the analyses and of the subsequent forecasts, because the model is not able to remove noise at the smallest scales. This leads to large errors in all prognostic fields in the area where radar data are assimilated and, as a consequence, to a very poor quality of the forecasts. This is particularly significant when 15 minutes assimilation cycles are employed, in which case forecast precipitation is strongly underestimated and mis-located. Conversely, a value of 10 dBZ does not degrade results both in 60 minutes and 15 minutes cycles and further tests are necessary to find out if a value grater than 5 dBZ can provide better results. Another improvement of results may be obtained when *roe* is dependent on the range, elevation and radar station, but a better comprehension and estimation of this value is mandatory before testing this configuration.

Competing interests. No competing interests.

Acknowledgements. TEXT

References

Anderson, J. L.: Spatially and temporally varying adaptive covariance inflation for ensemble filters, Tellus A, 61, 72–83, https://doi.org/10.1111/j.1600-0870.2008.00361.x, http://dx.doi.org/10.1111/j.1600-0870.2008.00361.x, 2009.

Anderson, J. L. and Anderson, S. L.: A Monte Carlo Implementation of the Nonlinear Filtering Problem to Produce Ensemble Assimilations and Forecasts, Monthly Weather Review, 127, 2741–2758, https://doi.org/10.1175/1520-0493(1999)127<2741:AMCIOT>2.0.CO;2, https://doi.org/10.1175/1520-0493(1999)127<2741:AMCIOT>2.0.CO;2, 1999.

Baldauf, M., Seifert, A., Förstner, J., Majewski, D., Raschendorfer, M., and Reinhardt, T.: Operational Convective-Scale Numerical Weather Prediction with the COSMO Model: Description and Sensitivities, Monthly Weather Review, 139, 3887–3905, https://doi.org/10.1175/MWR-D-10-05013.1, https://doi.org/10.1175/MWR-D-10-05013.1, 2011.

Berner, J., Fossell, K. R., Ha, S.-Y., Hacker, J. P., and Snyder, C.: Increasing the Skill of Probabilistic Forecasts: Understanding Performance Improvements from Model-Error Representations, Monthly Weather Review, 143, 1295–1320, https://doi.org/10.1175/MWR-D-14-00091.1, https://doi.org/10.1175/MWR-D-14-00091.1, 2015.

Bick, T., Simmer, C., Trömel, S., Wapler, K., Hendricks Franssen, H.-J., Stephan, K., Blahak, U., Schraff, C., Reich, H., Zeng, Y., and Potthast, R.: Assimilation of 3D radar reflectivities with an ensemble Kalman filter on the convective scale, Quarterly Journal of the Royal Meteorological Society, 142, 1490–1504, https://doi.org/10.1002/qj.2751, http://dx.doi.org/10.1002/qj.2751, 2016.

Bonavita, M., Torrisi, L., and Marcucci, F.: Ensemble data assimilation with the CNMCA regional forecasting system, Quarterly Journal of the Royal Meteorological Society, 136, 132–145, https://doi.org/10.1002/qj.553, http://dx.doi.org/10.1002/qj.553, 2010.

Buehner, M., Houtekamer, P. L., Charette, C., Mitchell, H. L., and He, B.: Intercomparison of Variational Data Assimilation and the Ensemble Kalman Filter for Global Deterministic NWP. Part I: Description and Single-Observation Experiments, Monthly Weather Review, 138, 1550–1566, https://doi.org/10.1175/2009MWR3157.1, https://doi.org/10.1175/2009MWR3157.1, 2010a.

Buehner, M., Houtekamer, P. L., Charette, C., Mitchell, H. L., and He, B.: Intercomparison of Variational Data Assimilation and the Ensemble Kalman Filter for Global Deterministic NWP. Part II: One-Month Experiments with Real Observations, Monthly Weather Review, 138, 1567–1586, https://doi.org/10.1175/2009MWR3158.1, https://doi.org/10.1175/2009MWR3158.1, 2010b.

Clark, P., Roberts, N., Lean, H., Ballard, S. P., and Charlton-Perez, C.: Convection-permitting models: a step-change in rainfall forecasting, Meteorological Applications, 23, 165–181, https://doi.org/10.1002/met.1538, http://dx.doi.org/10.1002/met.1538, mET-14-0154.R1, 2016.

Courtier, P., Andersson, F., Hackley, W., Vasiljevic, D., Hamrud, M., Hollingsworth, A., Rabier, F., Fisher, M., and Pailleux, J.: The ECMWF implementation of three dimensional variational assimilation (3D-Var). I: Formulation, Quarterly Journal of the Royal Meteorological Society, 124, 1783–1807, https://doi.org/10.1002/qj.49712455002, 1998.

Cuo, L., Pagano, T. C., and Wang, Q. J.: A Review of Quantitative Precipitation Forecasts and Their Use in Short- to Medium-Range Streamflow Forecasting, Journal of Hydrometeorology, 12, 713–728, http://www.jstor.org/stable/24912965, 2011.

Davolio, S., Silvestro, F., and Gastaldo, T.: Impact of Rainfall Assimilation on High-Resolution Hydrometeorological Forecasts over Liguria, Italy, Journal of Hydrometeorology, 18, 2659–2680, https://doi.org/10.1175/JHM-D-17-0073.1, https://doi.org/10.1175/JHM-D-17-0073.1, 2017.

Desroziers, G., Berre, L., Chapnik, B., and Poli, P.: Diagnosis of observation, background and analysis-error statistics in observation space, Quarterly Journal of the Royal Meteorological Society, 131, 3385–3396, https://doi.org/10.1256/qj.05.108, http://dx.doi.org/10.1256/qj.05.108, 2005.

Dixon, M., Li, Z., Lean, H., Roberts, N., and Ballard, S.: Impact of Data Assimilation on Forecasting Convection over the United Kingdom Using a High-Resolution Version of the Met Office Unified Model, Monthly Weather Review, 137, 1562–1584, https://doi.org/10.1175/2008MWR2561.1, https://doi.org/10.1175/2008MWR2561.1, 2009.

Doms, G., Förstner, J., Heise, E., Herzog, H.-J., Mironov, D., Raschendorfer, M., Reinhardt, T., Ritter, B., Schrodin, R., Schulz, J.-P., and Vogel, G.: A Description of the Nonhydrostatic Regional COSMO Model, Part II: Physical Parameterization, Tech. rep., http://www.cosmo-model.org/content/model/documentation/core/default.htm, 2011.

Dowell, D. C. and Wicker, L. J.: Additive Noise for Storm-Scale Ensemble Data Assimilation, Journal of Atmospheric and Oceanic Technology, 26, 911–927, https://doi.org/10.1175/2008JTECHA1156.1, https://doi.org/10.1175/2008JTECHA1156.1, 2009.

Errico, R. M.: Spectra Computed from a Limited Area Grid, Monthly Weather Review, 113, 1554–1562, https://doi.org/10.1175/1520-0493(1985)113<1554:SCFALA>2.0.CO;2, https://doi.org/10.1175/1520-0493(1985)113<1554:SCFALA>2.0.CO;2, 1985.

Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, J. Geophys. Res, 99, 10 143–10 162, 1994.

Ferting, E. J., Harlim, J., and Hunt, B. R.: A comparative study of 4D-VAR and a 4D Ensemble Kalman Filter: perfect model simulations with Lorenz-96, Tellus A, 59, 96–100, https://doi.org/10.1111/j.1600-0870.2006.00205.x, https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1600-0870.2006.00205.x, 2007.

Gaspari, G. and Cohn, S. E.: Construction of correlation functions in two and three dimensions, Quarterly Journal of the Royal Meteorological Society, 125, 723–757, https://doi.org/10.1002/qj.49712555417, https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.49712555417, 1999.

Gustafsson, N. and Bojarova, J.: Four-dimensional ensemble variational (4D-En-Var) data assimilation for the HIgh Resolution Limited Area Model (HIRLAM), Nonlinear Processes in Geophysics, 21, 745–762, https://doi.org/10.5194/npg-21-745-2014, https://www.nonlin-processes-geophys.net/21/745/2014/, 2014.

Hamrud, M., Bonavita, M., and Isaksen, L.: EnKF and Hybrid Gain Ensemble Data Assimilation. Part I: EnKF Implementation, Monthly Weather Review, 143, 4847–4864, https://doi.org/10.1175/MWR-D-14-00333.1, https://doi.org/10.1175/MWR-D-14-00333.1, 2015.

Harnisch, F. and Keil, C.: Initial Conditions for Convective-Scale Ensemble Forecasting Provided by Ensemble Data Assimilation, Monthly Weather Review, 143, 1583–1600, https://doi.org/10.1175/MWR-D-14-00209.1, https://doi.org/10.1175/MWR-D-14-00209.1, 2015.

Houtekamer, P. L. and Mitchell, H. L.: Data Assimilation Using an Ensemble Kalman Filter Technique, Monthly Weather Review, 126, 796–811, https://doi.org/10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2, https://doi.org/10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2, 1998.

Houtekamer, P. L. and Mitchell, H. L.: Ensemble Kalman filtering, Quarterly Journal of the Royal Meteorological Society, 131, 3269–3289, https://doi.org/10.1256/qj.05.135, http://dx.doi.org/10.1256/qj.05.135, 2005.

Houtekamer, P. L. and Zhang, F.: Review of the Ensemble Kalman Filter for Atmospheric Data Assimilation, Monthly Weather Review, 144, 4489–4532, https://doi.org/10.1175/MWR-D-15-0440.1, https://doi.org/10.1175/MWR-D-15-0440.1, 2016.

Houtekamer, P. L., Mitchell, H. L., Pellerin, G., Buehner, M., Charron, M., Spacek, L., and Hansen, B.: Atmospheric Data Assimilation with an Ensemble Kalman Filter: Results with Real Observations, Monthly Weather Review, 133, 604–620, https://doi.org/10.1175/MWR-2864.1, https://doi.org/10.1175/MWR-2864.1, 2005.

Hunt, B. R., Kostelich, E. J., and Szunyogh, I.: Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter, Physica D: Nonlinear Phenomena, 230, 112 – 126, https://doi.org/https://doi.org/10.1016/j.physd.2006.11.008, http://www.sciencedirect.com/science/article/pii/S0167278906004647, data Assimilation, 2007.

Jones, C. D. and Macpherson, B.: A latent heat nudging scheme for the assimilation of precipitation data into an operational mesoscale model, Meteorological Applications, 4, 269–277, https://doi.org/10.1017/S1350482797000522, https://rmets.onlinelibrary.wiley.com/doi/abs/10.1017/S1350482797000522, 2006.

Koinstinen, J. and Puhakka, T.: An improved spatial gauge-radar ad-justment technique, in: Proc. 20^{th} Conference on Radar Meteorology, pp. 179–186, AMS, 1981.

Leuenberger, D. and Merker, C.: Additive Covariance Inflation in an operational, convective-scale NWP Ensemble Kalman Filter Assimilation System, Poster contribution at the International Symposium on Data Assimilation (ISDA), https://isda2018.wavestoweather.de/program/poster_presentations/p6_1_leuenberger.pdf, 2018.

Leuenberger, D. and Rossa, A.: Revisiting the latent heat nudging scheme for the rainfall assimilation in convective system, in: Proc. of 3^{th} European Radar Conference on Radar Meteorology and Hydrology (ERAD), pp. 162–167, Visby, Gotland, Sweden, 6–10 September 2004, 2004.

Leuenberger, D. and Rossa, A.: Revisiting the latent heat nudging scheme for the rainfall assimilation of a simulated convective storm, Meteorol. Atmos. Phys., 98, 195–215, 2007.

Lin, Y.-L., Farley, R. D., and Orville, H. D.: Bulk Parameterization of the Snow Field in a Cloud Model, Journal of Climate and Applied Meteorology, 22, 1065–1092, https://doi.org/10.1175/1520-0450(1983)022<1065:BPOTSF>2.0.CO;2, https://doi.org/10.1175/1520-0450(1983)022<1065:BPOTSF>2.0.CO;2, 1983.

Meng, Z. and Zhang, F.: Limited-Area Ensemble-Based Data Assimilation, Monthly Weather Review, 139, 2025–2045, https://doi.org/10.1175/2011MWR3418.1, https://doi.org/10.1175/2011MWR3418.1, 2011.

Mitchell, H. L. and Houtekamer, P. L.: An Adaptive Ensemble Kalman Filter, Monthly Weather Review, 128, 416–433, https://doi.org/10.1175/1520-0493(2000)128<0416:AAEKF>2.0.CO;2, https://journals.ametsoc.org/doi/abs/10.1175/1520-0493%282000%29128%3C0416%3AAAEKF%3E2.0.CO%3B2, 2000.

Miyoshi, T., Sato, Y., and Kadowaki, T.: Ensemble Kalman Filter and 4D-Var Intercomparison with the Japanese Operational Global Analysis and Prediction System, Monthly Weather Review, 138, 2846–2866, https://doi.org/10.1175/2010MWR3209.1, https://doi.org/10.1175/2010MWR3209.1, 2010.

Nastrom, G. D. and Gage, K. S.: A Climatology of Atmospheric Wavenumber Spectra of Wind and Temperature Observed by Commercial Aircraft, Journal of the Atmospheric Sciences, 42, 950–960, https://doi.org/10.1175/1520-0469(1985)042<0950:ACOAWS>2.0.CO;2, https://doi.org/10.1175/1520-0469(1985)042<0950:ACOAWS>2.0.CO;2, 1985.

Poterjoy, J.: A Localized Particle Filter for High-Dimensional Nonlinear Systems, Monthly Weather Review, 144, 59–76, https://doi.org/10.1175/MWR-D-15-0163.1, https://doi.org/10.1175/MWR-D-15-0163.1, 2016.

Raschendorfer, M.: The new turbulence parametrization of LM, COSMO newsletter, 1, 89–97, 2001.

Rinollo, A., Vulpiani, G., Puca, S., Pagliara, P., Kaňák, J., Lábó, E., Okon, L., Roulin, E., Baguis, P., Cattani, E., Laviola, S., and Levizzani, V.: Definition and impact of a quality index for radar-based reference measurements in the H-SAF precipitation product validation, Natural Hazards and Earth System Sciences, 13, 2695–2705, https://doi.org/10.5194/nhess-13-2695-2013, https://www.nat-hazards-earth-syst-sci.net/13/2695/2013/, 2013.

Ritter, B. and Geleyn, J.-F.: A Comprehensive Radiation Scheme for Numerical Weather Prediction Models with Potential Applications in Climate Simulations, Monthly Weather Review, 120, 303–325, https://doi.org/10.1175/1520-0493(1992)120<0303:ACRSFN>2.0.CO;2, https://doi.org/10.1175/1520-0493(1992)120<0303:ACRSFN>2.0.CO;2, 1992.

Röpnack, A., Hense, A., Gebhardt, C., and Majewski, D.: Bayesian Model Verification of NWP Ensemble Forecasts, Monthly Weather Review, 141, 375–387, https://doi.org/10.1175/MWR-D-11-00350.1, https://doi.org/10.1175/MWR-D-11-00350.1, 2013.

Rossa, A., Nurmi, P., and Ebert, E.: Overview of methods for the verification of quantitative precipitation forecasts, pp. 419–452, Springer Berlin Heidelberg, Berlin, Heidelberg, https://doi.org/10.1007/978-3-540-77655-0_16, https://doi.org/10.1007/978-3-540-77655-0_16, 2008.

Schraff, C., Reich, H., Rhodin, A., Schomburg, A., Stephan, K., Periáñez, A., and Potthast, R.: Kilometre-scale ensemble data assimilation for the COSMO model (KENDA), Quarterly Journal of the Royal Meteorological Society, 142, 1453–1472, https://doi.org/10.1002/qj.2748, http://dx.doi.org/10.1002/qj.2748, 2016.

Seifert, A. and Beheng, K. D.: A double-moment parameterization for simulating autoconversion, accretion and selfcollection, Atmospheric Research, 59-60, 265 – 281, https://doi.org/https://doi.org/10.1016/S0169-8095(01)00126-0, http://www.sciencedirect.com/science/article/pii/S0169809501001260, 13th International Conference on Clouds and Precipitation, 2001.

Skamarock, W. C.: Evaluating Mesoscale NWP Models Using Kinetic Energy Spectra, Monthly Weather Review, 132, 3019–3032, https://doi.org/10.1175/MWR2830.1, https://doi.org/10.1175/MWR2830.1, 2004.

Snyder, C. and Zhang, F.: Assimilation of Simulated Doppler Radar Observations with an Ensemble Kalman Filter, Monthly Weather Review, 131, 1663–1677, https://doi.org/10.1175//2555.1, https://doi.org/10.1175//2555.1, 2003.

Sokol, Z.: Effects of an assimilation of radar and satellite data on a very-short range forecast of heavy convective rainfalls, Atmospheric Research, 93, 188 – 206, https://doi.org/https://doi.org/10.1016/j.atmosres.2008.11.001, http://www.sciencedirect.com/science/article/pii/S0169809508003402, 4th European Conference on Severe Storms, 2009.

Stephan, K., Klink, S., and Schraff, C.: Assimilation of radar-derived rain rates into the convective-scale model COSMO-DE at DWD, Quarterly Journal of the Royal Meteorological Society, 134, 1315–1326, https://doi.org/10.1002/qj.269, http://dx.doi.org/10.1002/qj.269, 2008.

Tiedtke, M.: A Comprehensive Mass Flux Scheme for Cumulus Parameterization in Large-Scale Models, Monthly Weather Review, 117, 1779–1800, https://doi.org/10.1175/1520-0493(1989)117<1779:ACMFSF>2.0.CO;2, https://doi.org/10.1175/1520-0493(1989)117<1779:ACMFSF>2.0.CO;2, 1989.

Tong, M. and Xue, M.: Ensemble Kalman Filter Assimilation of Doppler Radar Data with a Compressible Nonhydrostatic Model: OSS Experiments, Monthly Weather Review, 133, 1789–1807, https://doi.org/10.1175/MWR2898.1, https://doi.org/10.1175/MWR2898.1, 2005.

van Leeuwen, P. J.: Particle Filtering in Geophysical Systems, Monthly Weather Review, 137, 4089–4114, https://doi.org/10.1175/2009MWR2835.1, https://doi.org/10.1175/2009MWR2835.1, 2009.

Wang, X., Barker, D. M., Snyder, C., and Hamill, T. M.: A Hybrid ETKF–3DVAR Data Assimilation Scheme for the WRF Model. Part I: Observing System Simulation Experiment, Monthly Weather Review, 136, 5116–5131, https://doi.org/10.1175/2008MWR2444.1, https://doi.org/10.1175/2008MWR2444.1, 2008.

Wernli, H., Paulat, M., Hagen, M., and Frei, C.: SAL—A Novel Quality Measure for the Verification of Quantitative Precipitation Forecasts, Monthly Weather Review, 136, 4470–4487, https://doi.org/10.1175/2008MWR2415.1, https://doi.org/10.1175/2008MWR2415.1, 2008.

Wernli, H., Hofmann, C., and Zimmer, M.: Spatial Forecast Verification Methods Intercomparison Project: Application of the SAL Technique, Weather and Forecasting, 24, 1472–1484, https://doi.org/10.1175/2009WAF2222271.1, https://doi.org/10.1175/2009WAF2222271.1, 2009.

Yang, S.-C., Kalnay, E., Hunt, B., and Neill, E. B.: Weight interpolation for efficient data assimilation with the Local Ensemble Transform Kalman Filter, Quarterly Journal of the Royal Meteorological Society, 135, 251–262, https://doi.org/10.1002/qj.353, https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.353, 2009.

Zängl, G., Reinert, D., Rípodas, P., and Baldauf, M.: The ICON (ICOsahedral Non-hydrostatic) modelling framework of DWD and MPI-M: Description of the non-hydrostatic dynamical core, Quarterly Journal of the Royal Meteorological Society, 141, 563–579, https://doi.org/10.1002/qj.2378, http://dx.doi.org/10.1002/qj.2378, 2015.

Zhang, F., Snyder, C., and Sun, J.: Impacts of Initial Estimate and Observation Availability on Convective-Scale Data Assimilation with an Ensemble Kalman Filter, Monthly Weather Review, 132, 1238–1253, https://doi.org/10.1175/1520-0493(2004)132<1238:IOIEAO>2.0.CO;2, https://doi.org/10.1175/1520-0493(2004)132<1238:IOIEAO>2.0.CO;2, 2004.