

Interactive comment on “Assimilation of ground versus lidar observations for PM₁₀ forecasting” by Y. Wang et al.

Y. Wang et al.

wangy@cerea.enpc.fr

Received and published: 10 December 2012

We thank the reviewer for your useful comments on the following manuscript:

Journal: ACP

Title: Assimilation of ground versus lidar observations for PM10 forecasting

Author(s): Y. Wang et al.

MS No.: acp-2012-523

C10392

1 General comments

However, the paper is at times quite difficult to read and the main result of the paper - whether or not the Lidar stations are truly a meaningful improvement - gets lost in the details. Several statements are made redundantly, and details that aren't really that relevant to the analysis (e.g. a long-winded description of the choice of background errors) could be greatly shortened. At the moment it's rather easy to get lost in the details of this paper - I would like to see the authors make the experimental details simpler and more clear, so that the main point of the paper (i.e. that the assimilation of the LIDAR observations is more effective than that of the AirBase stations) is clear.

We have tried to make the paper easier to read, for example replacing the term “twin run” by “nature run” which is more explicit and changing some sentence positions (see answers to P299, L12-14:..., P300, L11-13:... and P304, L19-25:..., etc.). Some details were removed from the main text, e.g. in the section on the choice of DA method (see the answer to P296, Para1:...), description of the statistics, etc. However, the section on the choice of the horizontal and vertical correlation lengths has not been shortened, because it's crucial to understand and justify the parameters used for DA.

Very much space is dedicated to details of the assimilation experiments, while the space dedicated to actually discussing the results (section 7) is quite small. My overall recommendation is to simplify the discussion of experimental details (mostly by removing redundant statements, but also by carefully deciding which details are relevant to understanding the study and which are not), and devoting more time to explaining and contextualizing the results. By "contextualizing" I mean that the relevance of the results should be clear even to people who don't use this particular model or this particular assimilation approach.

Actually, the main results are shown in section 5, section 6 and section 7. In section 5, we studied the impact of correlation lengths to lidar DA and found the parameters which make lidar DA lead to better scores. We compared lidar network ($L_h = 200$ km

C10393

and $L_v = 0$) and AirBase network ($L_h = 200$ km and $L_v = 1500$ m) in section 6 to show the potential impact of lidar networks (12 stations) to improve PM₁₀ forecasts. We also studied the sensitivity to the number and to the lidars locations in section 7, which can help to design lidar network.

Some details were removed from the main text, e.g. in the section on the choice of DA method (see the answer to P296, Para1:...), description of the statistics, etc.

For clarity, the last paragraph of section 6 was replaced by “The results show that the impact on PM₁₀ forecast of assimilating data from a lidar network with 12 stations and data from a ground network AirBase with 488 stations are similar in terms of scores, although AirBase (resp. lidar) DA leads to slightly better scores for the first (resp. second) forecast day. We will study the sensitivity to the number and to the lidars locations in the next section.”.

Presently, section 7 really only tells us that, given the current experiment settings, the Lidar observations happen to give lower forecast errors on day 2 and higher forecast errors on day 1. Is that a good reason to expand and eventually assimilate the lidar network? You tell me!

We think your suggestion applies to section 6 rather than section 7, because we compared lidar network and AirBase network in section 6.

We took AirBase as a reference network in order to quantitatively show the potential impact of future ground-based lidar networks on analysis and short-term forecasts of PM₁₀ (section 6). Although the efficiency of assimilating the lidar network measurements is comparable or worse than that of assimilating concentration measurements from AirBase, we found advantages of lidar networks (only 12 lidar stations) against about 500 ground stations (section 6). Furthermore, we found that increasing the number of lidar (26, 76 or 488) improves the forecast scores in our paper.

For clarity, the following sentences have been added in conclusion section : “Because

C10394

AirBase network covers well western Europe and provides in situ surface measurements and AirBase measurements have been used for DA of PM₁₀, we took AirBase as an assimilation reference network.”, “Although AirBase DA can correct PM₁₀ concentrations at high levels because of the long vertical correlation length of the background errors, the lidar DA corrects PM₁₀ concentrations more accurately than the AirBase DA at high levels. The spatial and temporal influence of the assimilation of lidar observations is larger and longer.” and “Increasing the number of lidar improves the forecast scores. For example, the improvement of the RMSE becomes as high as 65% if 76 lidars are used, but a lidar network with many stations may be too expensive.”.

The writing could also be made more clear throughout by careful editing, and there are occasional grammatical and spelling errors throughout the document. I've made some technical suggestions below, but the manuscript should be carefully edited before put in final form.

We have edited carefully this new version. We used British English through out the manuscript. The English was improved. And we thank you for your suggestions of correction.

2 Specific comments

P293,L22: Is “applications of DA to PM10” equivalent to “applications of DA to air quality”? In that case, the beginning of this sentence is redundant. If there are other ways that DA has been done in the field of air quality, but with different aerosols, that could be mentioned (or at least this sentence made more clear.)

“applications of DA to PM10” is not equivalent to “applications of DA to air quality”. DA has been done for other species in the field of air quality, e.g. ozone and dust. For clarity, the following sentence is added “In air quality, Zhang et al. (2012) review chemical DA techniques developed to improve regional real-time air quality forecasting

C10395

model performance for ozone, PM₁₀, and dust. However, applications of DA to PM₁₀ forecast are still sparse.”

P294,L2: “was needed for the DA system” is pretty vague. It would make more sense to say something like, “was needed to yield useful air quality forecasts”, or whatever the case may be.

Yes, we agree. We changed “was needed for the DA system” to “was needed for the SDS forecasts” (SDS stands for sand and dust storm).

P295, L26: DA by definition always combines models and observations, not just in an OSSE.

We removed “In an OSSE,” on this line.

P296, Para1: The focus of this paragraph should be to justify why OI was used in this study over more complex algorithms like the EnKF and 4D-Var. However, a lot of detail are given here that make it easy for the reader to get lost. I think this sentence could be greatly simplified by mentioning that it has been shown that (a) 4D-Var apparently has certain weaknesses that make it suboptimal for PM10 assimilation, and (b) that the EnKF has both been shown to perform better and worse than OI/SI. If the statistical interpolation method used in the Denby et al (2008) reference is effectively similar to the OI method used here (I assume that they are, since the two terms are often used to describe similar algorithms in the literature), it might be easier to just refer to both algorithms as “OI”.

For clarity, we removed the following sentences : “In the EnKF used in air quality, the model uncertainties are approximated by the statistics of the ensemble generated by perturbing uncertain model parameters. It produces the best forecasts at the end of prediction periods.”, “The strongly constrained 4D-Var provides a moderate performance, because uncertainties are taken into account only at the initial date of the assimilation window.”. The following sentence is added to the end of this paragraph :

C10396

“Although aerosol assimilation could be performed with 4D-Var (Benedetti and Fisher, 2007), it may be limited to the use of a simplified aerosol model, as it is quite expensive for computation.”.

Yes, the SI method is similar to the OI method used. For clarity, we replaced “SI” by “OI”.

P296, L21-22: I would strongly suggest showing the OI analysis equations rather than referring the reader to the Tombette et al (2009) study - it seems to me that these are crucial to understanding what is done. Without seeing the basic DA update equation, a reader has only a weak handle on what the covariance matrix is there for.

The OI analysis equations have been included in this paragraph.

Section 4.1: This section introduced four somewhat-complicated measures of the performance of the assimilation. - Are any of these redundant, do they all contain unique information? This subsection should make clear why we need all four measures, i.e., what unique information each measure gives us. If there is a lot of redundancy, you could very much simplify this paper by getting rid of one or two measures.

In Boylan and Russel (2006), MFB and MFE are proposed to evaluate model performances against observations. RMSE and correlation are also often used in the aerosol modelling community. For clarity, the definition of the statistical indicators and their usefulness is added in appendix. The following sentences are added: “MFE and MFB bound the maximum error and bias and do not allow a few data points to dominate the statistics. They are often used to evaluate model performances against observations for aerosols (Boylan and Russel, 2006). The RMSE is a measure of the extent that the model deviates from the observations. Correlation is a measure of statistical relationships involving dependence between the observed and the modelled concentrations.”

P299, L12-14: At the end of this sentence, it would be good to append, “of the simulated PM10 concentrations with respect to the truth” - this will give the reader an idea

C10397

of why these measures are useful. Alternatively, put the sentence where the $\{o_i\}$ and $\{s_i\}$ are defined first in this paragraph.

We have appended “the results with respect to the truth” at the end of this sentence and put the definitions of o_i and s_i to the beginning of this paragraph.

P300, L4-5: There is no reason for the reader to be convinced here that the criterion specified by Boylan and Russel (2006) is applicable to this particular study. Is the Boylan and Russel study similar? Can you give a concise reason why we should accept these criteria?

The US EPA has issued minimal guidance on PM model performance evaluation metrics, goals, and criteria. Boylan and Russel (2006) recommended PM model performance goal and criterion that are based upon an analysis of numerous PM and visibility modelling studies. The PM model performance goal corresponds to the level of accuracy that is considered to be close to the best a model can be expected to achieve. The PM model performance criterion corresponds to the level of accuracy that is considered to be acceptable for modelling applications.

P300, L11-13: The sentence beginning with “Even though for an OSSE...” nicely explains why we are interested in the MFB and MFE measures. This sentence should go to the beginning of this paragraph! Also, it looks like the MFB and MFE measures are used only to meet the model performance criteria, while the RMSE and correlation are going to be used to evaluate the assimilation (analysis) relative to the truth. This could be made more clear.

You are right. The sentence beginning with “Even though for an OSSE...” has been moved to the beginning of this paragraph.

Yes, the MFB and MFE measures are used only to meet the model performance criteria, while the RMSE and correlation are used to evaluate DA runs relative to the truth. For clarity, the following sentence is added to the end of this paragraph: “for example

C10398

the calculation of RMSE and correlation”.

P300, L16: It is confusing to distinguish between “the truth” and “the true states (e.g. concentrations)” - for the latter, why not just say, “hypothetical PM10 observations”?

As said in P299 L5-6, the “truth” denotes a synthetic atmosphere state. The “true” states denote concentrations of this synthetic atmosphere in our paper. The “true” states are not “hypothetical PM10 observations”. Because the observations are made noisy, “PM10 observations” would correspond to the “true” states of PM10 perturbed with the observation errors (e.g. 35% or 43%).

P300, L23: It should be explained what is meant by “representativeness errors”. Where does the 35% come from? Why are different assumptions made about observation error in AirBase and in LIDAR?

The representativeness errors mathematically come from the transformation and projection operator which map the true state (e.g. continuous) to the model state (e.g. discrete). They measure the inabilities of the model to simulate subgrid scale processes. They depend on the resolution of the model state and the characteristics of the location. Here, 35% is an average value.

We considered the representativeness errors for AirBase, while we considered both the representativeness and instrumental errors for LIDAR, because the instrumental errors of AirBase is small.

P301: It is interesting that the observations are perturbed with a spatial covariance structure, as opposed to just adding random noise to each measurement. Can you explain why this more complex approach was chosen?

As we showed in Figure 4, those perturbations make the perturbed observations depend continuously on the vertical level and time. Just adding random noise to each measurement will produce discontinuous observations. In the reality, we do expect observational trends to be continuous.

C10399

P302, L28: I don't entirely understand "allows us to increase the duration of this impact". If I understood correctly, both the initial values of all (?) chemical constituents and aerosols are perturbed in the assimilation run relative to the "true" run, in order to make the error between them larger. If this was not done, in which case both runs would in some sense forget the initial conditions, what would happen? Would the two runs collapse back to the same state? If so, why are we interested in assimilating observations, if we can just recover the truth by letting the model run long enough?

P303, L12-14: Same comment as for P302, L28

Both initial gaseous and aerosols are perturbed in order to make the difference between the control run and the "truth" last as long as possible. For clarity, the sentence "Perturbing both initial gaseous and aerosol of species allow us to increase the duration of this impact." is replaced by "For this impact to last as long as possible, both gaseous and aerosol concentrations are perturbed."

In air quality models, the impact of initial conditions on PM₁₀ concentrations lasts for a few hours to a few days at most. It means that the two runs collapse back to the same state.

Although the impact of assimilating PM₁₀ observations to correct initial conditions lasts for a few days at most, we can use DA to correct our initial conditions and to improve forecasts up to 48 hours. Moreover, in reality, the asymptotic trajectory of the model would not match the real state of the atmosphere because of model errors, so that DA would help correct these.

P304, L19-25: The end of section 4 states that the specification of the background error covariance / correlation field is critical to the success of the assimilation. Then the beginning of section 5 states that "the definition of background correlations are relatively trivial". Which is it? It's also confusing that section 4 discusses using the Balgovid approach but then section the NMC method is mentioned; one has to read both parts over a few times to figure out what used for which part. I would suggest

C10400

integrating the discussion of background error statistics into section 5. Then make it clear that background error covariances are extremely important, but also explain what makes their specification easier in this (special) case where the model is a perfect representation of the truth. Some redundant statements can also be removed, e.g. "...are crucial for the success of the method" (P304, L7) and "...is crucial to the quality of the analysis" (P304, L17).)

At the end of section 4, we want to emphasise the specification of the background error correlations is very important, because they determine to what extent the background fields will be corrected.

At the beginning of section 5, we said "the definition of background correlations are relatively trivial", because it is trivial on mathematics, but the true atmospheric state is never known.

For clarity, as suggested by the reviewer, we have moved the first sentence of section 5 to the end of section 4, and the statement "and are crucial for the success of the method" (P304, L7) has been removed. But we kept the statement "...is crucial to the quality of the analysis" (P304, L17), because it is very useful to make the readers understand to what parameters the results depend on.

Figure 8: What do the vertical black lines denote? Are the RMSE and Correlation here defined between the assimilated analysis and truth? If so (or if otherwise), this should again be mentioned here. Also, the caption points top and bottom figures, but they are actually side by side. It would make the plots much more clear if they were clearly labeled "RMSE" and "Correlation" as headings. As in other figures, the axis labels could be much larger.

In Fig. 8, the vertical black lines denote the moments where forecasts of PM₁₀ begin. The assimilation period corresponds to the left of the black line, while the forecast period corresponds to the right of the black line. The following is added to the figure caption: "The vertical black lines denote the separation between the assimilation period

C10401

(to the left of the black lines) and the forecast (to the right of the black lines).”

Yes, the RMSE and correlation here are defined between the assimilated analysis and truth. The sentence “The scores are calculated over land grid points from the ground to the sixth level.” has been added in the caption.

“RMSE” and “Correlation” have been added as label and the character size enlarged.

General comment on section 5: It isn't entirely clear to me why the NMC method of estimating background error covariances is explained in such detail, only to manually test different decorrelation lengths anyway. Why not just show Figure 8 and then talk about the effects of the different decorrelation lengths? Also, it is very clear from Figure 8 that assimilating the LIDAR observations yields lower errors and a higher correlation to the truth than does the assimilation of AirBase stations only. It seems odd not mention this, either at the end of Section 5 or at the beginning of Section 6.

The NMC method is often used to estimate background error covariances, in meteorology and air quality modelling. Manually testing different correlation lengths is another method to validate conclusions of the NMC method. But the choice of the correlation lengths that were tested manually are obtained from the NMC method.

It is not clear from Figure 8 that assimilating the LIDAR observations always leads to lower errors and a higher correlation than assimilating AirBase observations only. That is why we compared them in detail in section 6.

P307, L1-3: “...because assimilation only influences initial conditions...” - this is only true if the assimilation is only applied at the initial time, but not if observations are assimilated at regular intervals.

Here, we wanted to say “...because assimilation only influences initial conditions of the forecast period ..”. We added “of the forecast period” in the paper.

P307, last para: Lines 18-19 state, “the AirBase DA leads to lower RMSE than column DA for most forecasts,” but then lines 28-29 say, “the column DA leads to lower or

C10402

similar RMSE as the AirBase DA for most forecasts.” I had to read the paragraph several times to understand how this is not completely contradictory. Please clarify the text by (clearly!!) stating what the difference is between what is shown in Fig. 9 and in Fig. 10. Since the experimental set-up is fairly complicated, it really needs to be more clear where we are looking at continuously-assimilated runs and where we are looking at pure forecasts.

Lines 18-19 state is the results of the first forecast day shown in Fig. 9. And lines 28-29 is the results of the second forecast day shown in Fig. 10. For clarity, we have split this paragraph into two. Furthermore, we added in the text: “For the second forecast day (Fig. 10), the relative impact of column DA and AirBase DA is different from the first forecast day (Fig. 10) : the column DA leads to lower or similar RMSE as the AirBase DA for most forecasts.”.

P309, L7-8: “The results shown in this paper suggest that the assimilation of lidar observation would improve PM10 forecast over Europe” - this statement should be made much more specific. Section 7 shows that the value of the lidar observations greatly increases on the second forecast day, relative to the first. Can the authors list other respects in which (according to this study) the lidar observations offer an advantage, and the extent of this advantage. Something along the lines of, “Assimilation of the lidar observations improves forecasts by x% relative to the standard AirBase measurements.”

The statement, “Although AirBase DA can correct PM₁₀ concentrations at high levels because of the long vertical correlation length of the background errors, the lidar DA corrects PM₁₀ concentrations more accurately than the AirBase DA at high levels. The spatial and temporal influence of the assimilation of lidar observations is larger and longer.”, has been added at the end of this paragraph.

We also added more details in the paragraph on the impact on the number of lidars : “As lidar stations are developing over Europe following volcanic eruptions in Iceland

C10403

(Chazette et al., 2012; Pappalardo et al., 2010), a sensitivity analysis has also been conducted on the number and locations of lidars.” and “For example, the improvement of the RMSE becomes as high as 65% if 76 lidars are used, but a lidar network with many stations may be too expensive.”.

3 Technical corrections

P293, L6-7: “Aerosols influence gaseous molecules photo-dissociation” – > “Aerosols influence the photo-dissociation of gaseous molecules”.

Corrected.

P294, L27: “analysis” – > “analyses”

Corrected.

P295, L4: No need to write “OSSE system” since OSSE already contains the word “Experiment” - Just writing “An OSSE is constituted...” is enough. Same goes for similar uses later on. This sentence also doesn’t make it clear what a twin run actually is, since really any model run can be considered “an approximate atmosphere”. The next sentence explains it, so the first sentence should be slightly restructured.

All “OSSE system” have been replaced by “OSSE” in this paper.

P296, L5: “the ensemble” – > “an ensemble”

This statement has been removed because of the comment P296, Para1.

P298, L13: “stations types” – > “station types”

We have changed “stations types” to “station types”.

P298, L23-24: Change the end of the sentence to “in order to better cover Western Europe”.

C10404

This paragraph is changed to “In this work, a network of 12 fictitious ground-based lidar stations covering western Europe is defined, as shown in Fig. 1, based on the lidar locations of existing observation stations, e.g. some stations from the European Aerosol Research Lidar Network (<http://www.earlinet.org/>).”.

P299, L3-4: This sentence is confusing and can be worded much more simply. I’d suggest something like, “Observation impact experiments for not-yet-existing observing systems require the simulation of an atmospheric state, from which the hypothetical observations can be generated.” Then all that is needed in the rest of the paragraph is to explain that we call this state the “truth”, and to describe the truth run used in this study.

As you suggest, this sentence P299, L3-4 is changed to “Observation impact experiments for not-yet-existing observing systems requires an atmospheric state, from which the hypothetical observations can be generated.”.

Section 4.3: Since this run is afterwards only referred to as the “truth”, perhaps it would be better to entitle this section “Truth simulation”, or something like that.

We think your suggestion applies to section 4.1 rather than section 4.3, because the twin run refers to the truth as specified P23299 L8 (1st version of the paper). For clarity, the title of this section “Twin run” is changed to “Nature run”.

P302, L8: Technically, the “twin run” wouldn’t be called that if this wasn’t a twin experiment. It’s probably better here to refer to the non-control run as the “assimilation run” - then it’s clear to the reader that this is the run where the observations are assimilated.

For clarity, the “twin run” was replaced by “nature run”.

P302, L15: There is a lot of redundant information in this paragraph. It’s better and more straightforward to state right away that this particular study performs identical twin experiments, so move the sentence starting with “We follow...” to the beginning of the paragraph. The part about fraternal-twin experiments isn’t really relevant to this

C10405

study and should probably be parenthetical. The most important point made in this paragraph is that the impact of the observations is usually overestimated in identical twin experiments - if the other details are trimmed, this will be more clear and then you won't have to repeat it at the beginning of the next paragraph.

The sentence starting with "We follow..." is moved to the beginning of the paragraph. The paragraph is re-organised and the most important point (that the impact of the observations is over-estimated) is stated at the end of the paragraph, so that we do not have to repeat it at the beginning of the next.

P303, L17-19: It is not clear why there should be a compromise between lower assimilation error and short assimilation time. - What is the benefit of keeping the assimilation time N low? Why would we not want to assimilate as many observations as possible?

In reality, when we use real observations, we assimilate as many observations as possible, if the computational cost of DA is not too high. However, since only the initial conditions are perturbed in our experiments (see Sect. 4.3), the difference between two forecasts initialised with different initial conditions only lasts for a few days. Therefore, we need to limit the assimilation period N, so that the differences in PM₁₀ concentrations between the runs with and without assimilation are still large at the beginning of the forecast period.

Figure 5: I would suggest a colormap that is white in the center (i.e. for small to no differences between truth and assimilation run), to make it easier to spot where the main differences are, but I'll leave this up to the authors. Also, the figure labels could be much larger. Also, it would be good to remind the reader where in the assimilation time 00:00 on 15 July falls (i.e. that it's the initial time).

The statement "which is the initial time of the first five-day experiment," is added to the end of "00:00 on 15 July". The figure labels are enlarged.

P303, L21-22: "The simulations use the same setup..." - this seems to have been made

C10406

abundantly clear in the description of the identical twin experiments.

Here, we want to explain that the simulations of "DA runs" have the same setup as the one of the control run. We prefer to keep it for clarity.

P306, L25: "of Fig. 8" seems misplaced here, since both tests are shown in Figure 8. (Or the sentence needs to be restructured.)

It is restructured to "In the following, we compare the DA test "AB 200km 1500m" of Fig. 8 for AirBase ($L_h = 200$ km and $L_v = 1500$ m) and the DA test "Col. 200km 0m" of Fig. 8 for the lidar network ($L_h = 200$ km and $L_v = 0$)."

References

- Benedetti, A. and Fisher, M. : Background error statistics for aerosols, Q. J. Roy. Meteor. Soc., 133, 391-405, 2007.
- Boylan, J. W., Russell, A. G. : PM and light extinction model performance metrics, goals, and criteria for three-dimensional air quality models. Atmos. Env., 40 (26), 4946-4959, 2006.
- Chazette, P., Bocquet, M., Royer, P., Winiarek, V., Raut, J.-C., Labazuy, P., Gouhier, M., Lardier, M., and Cariou J.-P. : Eyjafjallajökull ash concentrations derived from both Lidar and modeling, J. Geophys. Res. Atmospheres, 117, D00U14, doi:10.1029/2011JD015755, 2012.
- Pappalardo, G., Amodeo, A., Ansmann, A., Apituley, A., Arboledas, L.A., Balis, D., Böckmann, C., Chaikovsky, A., Comeron, A., D'Amico, G., De Tomasi, F., Freudenthaler, V., Giannakaki, E., Giunta, A., Grigorov, I., Gustafsson, O., Gross, S., Haeffelin, M., Iarlori, M., Kinne, S., Linné, H., Madonna, F., Mamouri, R., Mattis, I., McAuliffe, M., Molero, F., Mona, L., Müller, D., Mitev, V., Nicolae, D., Papayannis, A., Perrone, M.R., Pietruczuk, A., Pujadas, M., Putaud, J. P., Ravetta, F., Rizi, V., Serikov, I., Sicard, M., Simeonov, V., Spinelli, N., Stebel, K., Trickl, T., Wandinger, U., Wang, X., Wagner, F., and Wiegner, M. : EARLINET observations of the Eyjafjallajökull ash plume over Europe, SPIE Proceedings, 7832, doi:10.1117/12.869016, 2010.
- Zhang, Y., Bocquet, M., Mallet, V., Seigneur, C., and Baklanov, A. : Real-time air quality fore-

C10407

casting, part II: State of the science, current research needs, and future prospects, *Atmos. Env.*, 60, 656-676, 2012.

C10408