**Atmospheric
Chemistry
and Physics
Discussions**

Interactive comment on "Assimilation of ground versus lidar observations for PM$_{10}$ forecasting" *by* Y. Wang et al.

Y. Wang et al.

wangy@cerea.enpc.fr

We thank the referee 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

1   General comments

In general, I found the paper interesting but difficult to read.

We tried to make the paper easier to read, for example by replacing the term "twin run" by "nature run" which is more explicit. Some details were removed from the main text, e.g. description of the statistics, etc.

It also lacks the strong message, which was evidently expected by the authors when the work was started.

Because a lidar network with continuous measurements does not yet exist, we expected this work to prove the impact of a ground-based lidar network on short-term forecasts of PM$_{10}$ and to help future lidar network projects to design the lidar network, e. g. number and locations of lidar stations. For clarity, the following sentence is added in introduction section : "This work aims to investigate the usefulness of future ground-based lidar network on analyses and short-term forecasts of PM$_{10}$ and to help future lidar network projects to design lidar networks, e. g. number and locations of lidar stations. ".

Therefore, in our work, we started with defining 12 fictitious lidar stations which are based upon the lidar locations of existing observation stations and we increased the number of fictitious lidar stations to 26, 76 or 488 in the sensitivity study section (section 7). For clarity, the description of a lidar network with 12 lidar stations in this paper 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/).".

The efficiency of extended lidar network appeared to be comparable or worse than that of already existing AirBase. The difference of 2% in the RMS reduction at the second day of the forecast is evidently insignificant.

Yes, we agree with the reviewer : assimilating *only* 12 lidar stations measurements do not lead to better scores (lower RMSE and higher correlations) than assimilating 488 AirBase stations measurements, even on the second forecast day, although the scores

are sensibly better.

Intuitively, the opposite should be true: the possibility to assimilate 3D fields should increase the model scores quite a lot. It did not realize in the current experiment, probably because of very few lidar sites. However, extension of this network is costly.

The improvement are low with only 12 lidar stations in DA. However, as shown in the sensitivity study section (section 7), increasing the number of lidars (to 26, 76 or 488) improves the forecast scores. For example, assimilating 76 lidar stations measurements lead to much better scores than AirBase. For clarity, the following sentence is added in the conclusion section : "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.".

Therefore, demonstration of sufficiency of AirBase network for data assimilation tasks is probably the most interesting conclusion of the paper, which is bound to raise further discussion. Therefore, I favour the publication providing the below comments are taken into account.

AirBase network covers well western Europe and provides in situ surface measurements. AirBase measurements have been used for DA of aerosols (e.g. PM10) in Denby et al. (2008); Tombette et al. (2009). We took AirBase as a assimilation reference network in order to quantitatively show the potential impact of future ground-based lidar networks on analysis and short-term forecasts of PM_{10}. Therefore, the similar efficiency of assimilating 12 lidar and AirBase measurements is also, if not the main, conclusion of this paper. For clarity, the following sentence has been added in conclusion section : "Because AirBase network covers well western Europe and provides in situ surface measurements and AirBase measurements have been used for DA of PM_{10}, we took AirBase as an assimilation reference network.".

2   Specific comments

Introduction. L.36-40. This sounds as a biased statement: data assimilation is quite efficiently used in the AQ world by many groups, both in research and, since recent, in operational forecasting. PM10 is indeed not heavily addressed yet â for many reasons not discussed in the paper. However, limitation to PM10-only narrows the review and excludes the main groups routinely working with data assimilation in Europe and US. I suggest to give a broader outlook here.

A reference to Zhang et al. (2012) is added to this statement, as it is a recent review paper of DA used in air quality.

L.60-65 and elsewhere. Why lidars for PM10? One would argue that lidars are more sensitive to PM2.5 since the surface-to-mass ratio for fine particles is higher. Also, actual aerosol size spectrum seems to favour finer particles, which provide the main surface-wise contribution. No such discussion is given but a coupe of sentences would be useful.

Yes, you are right. The following sentences were added in the text. "Because the surface-to-mass ratio for fine particles ($PM_{2.5}$, particulate matter with a diameter smaller than $2.5\ \mu m$) is high, they largely contribute to the measured lidar signal. However, the contribution of coarse particles may not be negligeable as shown by Randriamiarisoa et al. (2006) who estimated it to be about 19%. The relative contribution of $PM_{2.5}$ may increase with altitude (Chazette et al., 2005), but it is difficult to quantify."

L.167-175. These equations are trivial and unnecessary in the paper body. They should either be moved to annex or eliminated.

These equations are moved to an appendix section.

L176-178. I did not understand. Two sets of criteria are given with the same statement that if they are satisfied, the model performance goal is met. So, which of the sets is

the right one?

Actually, those two sets of criteria are given for two different model performances : a goal and a criterion.

If the model performance goal is met, the level of accuracy that is considered to be close to the best a model can be expected to achieve. If the model performance criteria is met, the level of accuracy that is considered to be acceptable for modelling applications.

L.178-182. And again something strange: the statement construction suggests that the PM2.5 and PM10 scores are different but the actual wording says that in both cases the criterion is met. That should be rewritten. And why PM2.5? It is not the goal of the paper, neither it is mentioned elsewhere.

The statement "whereas for PM$_{2.5}$ both the model performance goal and criterion are met for the two networks," is changed to "whereas for PM$_{2.5}$ the model performance goal is met for both networks,".

Yes, PM$_{2.5}$ is not the goal of this paper. But PM$_{2.5}$ is an other important species to check if the simulation of aerosol compares well to observations for particle mass.

L.184. The opposite is true: accuracy of the twin run is of high importance because its synthetic measurements must reproduce the real pattern and real situation. Otherwise conclusions on the network design and quality will not be applicable for real life.

The sentences "Even though, for an OSSE study, the accuracy of the nature run compared with real observations is usually not a major concern, the twin run should produce typical features of the phenomena of interest." is moved to the beginning of this paragraph and changed to "For an OSSE study, the accuracy of the nature run compared with real observations is important, and the nature run should produce typical features of the phenomena of interest.".

L.203. "as now explained" should be removed.

It is done.

L.211. Please explain how the error covariance matrix "sigma-capital" is obtained from the function $f(d_v, d_t)$. Only after that one can proceed into Cholesky decomposition of this matrix.

We added the sentence "Each component of the covariance matrix Σ may be written as $\Sigma_{ij} = f(d_v(x_i, x_j), d_t(x_i, x_j))$." in this statement in order to explain how the error covariance matrix "sigma-capital" is obtained from the function $f(d_v, d_t)$.

Eq.7. Please state the features of gamma clearly: normal distribution requires mean and variance to be fully defined.

It is changed to "γ is a random vector whose components are a standard normal distribution (of mean 0 and variance 1).".

L.236. It is a hand-waving justifying an easy-to-do but not rigorous approach. I agree that over-estimation of the DA efficiency will be for both in-situ and lidar networks - but will it be overestimated to the same extent? The features of the datasets are much too different to claim that. For instance, initial conditions are "forgotten" faster near the ground than aloft - and (in)correct emission/deposition at/near the surface will overwhelm the impact of transport or transformation representation, which are the main players higher up. In that sense, "ideal" model choice will probably favour lidars more than ground stations. Some discussion of this kind is needed here.

In the AirBase DA run, we introduced a vertical correlation length ($L_v = 1500$m. Because of the large value of L_v, the assimilation of AirBase ground measurements affect high vertical levels. Actually, the correction of high vertical levels for AirBase DA is very important for the DA to be efficient. Therefore, the over-estimation of the DA efficiency is probably of the same extent for lidar and AirBase DA. As you can see in Fig. 8 and 12 that scores of lidar and AirBase DA converge similarly. The following sentence is added to the paper : "(the assimilation of both ground and lidar observations lead to

corrections at high vertical levels, as discussed in section 5)"

L.258-259. Did not understand the sentence. Shouldn't it be just removed?

We wanted to say that two types of DA runs are performed in our OSSE. One type assimilates AirBase observations; Another assimilates lidar observations. For clarity, "Two runs" is changed "Two different types of DA runs".

L.261-262. Why the first model level?? It is wrong in general and directly contradicts to the section 5, which suggests much more rigorous approach. The whole paragraph seems to be forgotten from some historical text edition.

It is a mistake. The sentence has been changed to "from the first level (20m above the ground) to the sixth level (1950m above 25 the ground) of the model".

L.270-277. Standard deviation instead of variance would be better here: more intuitively understandable numbers.

In order to make this statement in line with Section 4.2, we suggest to keep "variance". Thanks for your comprehension.

L.285-286. An empty sentence, which seems to have no connection to the section. Please remove.

We agree with the reviewer that the statement does not make sense at the beginning of section 5. We moved it to the end of section 4 where it belongs.

L.306. This is strange. Why would the horizontal correlation length decrease in the free troposphere in comparison with ABL? A common knowledge is that the fields are much smoother in FT, so should be the errors, shouldn't they? A short discussion is needed here.

Yes, the fields are smoother in the free troposphere than in the low troposphere. This model behaviour in the free troposphere where the horizontal correlation length decrease is a consequence of prescribed aerosol boundary conditions and the numerical

C10401

algorithm. Because the background error is estimated by the differences between a simulation with 24 forecast and a simulation with assimilating ground measurements in the NMC method (the only error sources is the ground measurements) and the same boundary conditions are used for both simulations, too small background error appears at higher levels. By contrast, the numerical noise becomes important and leads to short length correlations at high levels. A similar behaviour also is shown in Benedetti and Fisher (2007); Pagowski et al. (2010).

For clarity, a following discussion is added in the paper : "This is a consequence of the prescribed aerosol boundary conditions and the numerical algorithm. Because the background error is estimated by the differences between a simulation with 24 forecast and a simulation with assimilating ground measurements in the NMC method (the error sources are the ground measurements) and the same boundary conditions are used for both simulations, too small background error appears at high levels. By contrast, the numerical noise sometimes becomes important and leads to short length correlations at high levels. A similar behaviour is shown in Benedetti and Fisher (2007); Pagowski et al. (2010).".

Section 6. Here the problem shows up, owing to the above-criticised choice of the ideal-model twin run setup. The selection of exactly the same model for twin, control, and DA runs, combined with perturbation of only the initial conditions, lead to reduction of the model error with time in all runs. Indeed, with influence of initial conditions fading out with the forecast length, all runs fall to the same main path, i.e. become non-distinguishable from each other. This is why RMSE reduces and the correlation coefficient grows in all runs all the time. The role of DA is only to make it faster. Such discussion is entirely missing. A reader without deep experience with OSSE is left guessing why RMSE continues to decrease also outside assimilation window, whereas in all real applications it will immediately start growing. Explanation should be added and also related to Figure 8.

The statement "Because only the initial conditions (pollutant concentrations) are differ-

C10402

ent between the nature run and the control runs (see section 4.3), and because the influence of initial conditions fade out with the forecast time, all control runs converge (RMSEs decrease to 0 and correlations increase to 1 in Fig. 8). The role of DA is to accelerate this convergence, to make RMSEs decrease and correlations increase faster." is added and related to Fig. 8 to explain the fact that RMSEs decrease and correlations increase with time.

L.364-365. I see no "powerful impact" whatsoever - the impact of those networks is practically identical. Those few % of difference to either direction are well inside the method uncertainty and simplifications of the setup. The strongest statement possible here is that those networks lead to a similar effect, despite 40-fold smaller number of lidar sites. However, the costs can be comparable - or higher in case of the lidar network.

Yes, we can see that they lead to a similar effect. The sentence has been replaced by "The results show that the impact on PM_{10} 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.".

Note that this result is independent from the cost issue which can later be discussed.

Section 7. I am strongly lacking the sensitivity run with 9 existing lidar stations. Comparison of this, already existing lidar network, with already existing AirBase would be very useful for practical applications. I would advise to do this exercise.

Although some stations of our 12-station network exist and do make continuous measurements, there is not yet any operational lidar network. We based our work on what could be a future lidar network, but it is not operational yet and the design (the number and locations of lidar stations) has not yet been decided. In particular, our 12-station network is inspired from existing observation stations, but it is not a proper planned extension of a given lidar network.

L.381 and elsewhere. The word "global" should not be used in this context: improving European network does not lead to better global forecasts. I guess, the authors meant "in-general" or "overall" in all such occasions.

The word "globally" is changed to "overall" in all such occasions.

L.396. I would remove the word "strongly". With all the simplifications in the setup and very similar scores against AirBase there is no reason to make so bold statement. Actually, the boldness of the message is lowered already by the next sentence, so let's be consistent.

You are right. The word "strongly" has been removed.

L.408-415. This paragraph is neither understandable nor adds anything to the conclusions. Please remove.

We removed the sentence "Because we did not have enough available lidar observations in western Europe, we did not perform DA with a combination of real lidar and AirBase observations.", which does not add information to this paper. However, we kept relating mass concentration to optical properties, as this is crucial for future work based on this paper.

General remark. Please do not provide too many digits. For instance, out of three digits in 56.1% improvement, maximum 2 are really known, i.e. 56% should be shown.

Yes, we agree. We now provide at most 2 digits for every number, e.g. in sections abstract, conclusions and Table 1.

3 Figures

Figure 4 is unreadable. Figure 5 is unreadable. Figure 8. Instead of empty names "text X" in the legend, a meaningful abbreviation would be very helpful. For instance,

instead of "test 1", one can use "AB 50km 1500m". That would dramatically simplify the analysis of the very busy pictures. Also, the range for correlation coefficient should start from 0.5 or 0.6 to make the lines distinguishable. Figure 12 and 13. Empty names "Network X" should be replaced with meaningful abbreviations, e.g. Net.YY, where YY is the number of sites.

In Fig. 8, legends are changed to the form like "AB 50km 1500m". In Fig. 12 and 13, the legend "Network X" is changed to "Net.YY", where YY is the number of lidar stations. The labels character size is made larger in all figures.

References

Benedetti, A. and Fisher, M. : Background error statistics for aerosols, Q. J. Roy. Meteor. Soc., 133, 391-405, 2007.

Chazette, P., Randriamiarisoa, H., Sanak, J., and Couvert P. : Optical properties of urban aerosol from airborne and ground-based in situ measurements performed during the ESQUIF program, J. Geophys. Res, 110, D0220610.1029/2004JD004810, 2005.

Denby, B., Schaap, M., Segers, A., Builtjes, P., and Horálek, J. : Comparison of two data assimilation methods for assessing PM_{10} exceedances on the European scale, Atmos. Env., 42, 7122-7134, 2008.

Pagowski, M., Grell, G. A., McKeen, S. A., Peckham, S. E., and Devenyi, D. : Three-dimensional variational data assimilation of ozone and fine particulate matter observations: some results using the Weather Research and Forecasting - Chemistry model and Grid-point Statistical Interpolation, Quaterly Journal of the Royal Meteorological Society, 136, 2013-2024, DOI: 10.1002/qj.700, 2010.

Randriamiarisoa, H., Chazette, P., Couvert, P., Sanak, J., and Mégie, G. : Relative humidity impact on aerosol parameters in a Paris suburban area, Atmos. Chem. Phys., 6, 1389-1407, 2006.

Tombette, M., Mallet, V., and Sportisse, B. : PM_{10} data assimilation over Europe with the optimal interpolation method, Atmos. Chem. Phys., 9, 57 -70, 2009.

Zhang, Y., Bocquet, M., Mallet, V., Seigneur, C., and Baklanov, A. : Real-time air quality fore-

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