Aerosol data assimilation in the chemical-transport model MOCAGE during the TRAQA/ChArMEx campaign: Aerosol optical depth

Bojan Sic, Laaziz El Amraoui, Andrea Piacentini, Virginie Marécal, Emanuele Emili, Daniel Cariolle, Michael Prather, and Jean-Luc Attié

This paper describes a data assimilation system for aerosol applied to a regional model. While the DA system itself is not novel, the interest of the paper lies in the verification of DA results with observations very different from those used in the data assimilation. Assimilated observations are remotely sensed AOT from MODIS but part of the verification is done with flight campaign measurements of particle number densities. Sic et al show that although they assimilate only column-integrated optical properties (AOT), even particle number density profiles can be improved. This is a very interesting result.

Sic et al. go on to explain this in terms of how the modelled profile shape changes in subsequent assimilation cycles. Although this is not really surprising to specialists in aerosol data assimilation, it can be hoped that this paper will help affect a better understanding in a more general audience who often have trouble accepting that data assimilation of 2D properties can improve 3D distributions.

This point might be brought across even stronger with some additional analyses. Some suggestions are

- Trajectory analyses of selected air parcels identified during previous assimilation cycles,
- A data denial experiment where the DeepBlue observations are removed (and so observations close to dust sources are missing),
- Evaluation against LIDAR data, e.g. Earlinet.

That said, the paper is acceptable for publication as it is (after minor changes suggested below).

Specific comments:

Abstract:

L 14: "reduced the bias in the AOD (from 0.050 to 0.006) and increased the correlation (from 0.74 to 0.88)": please state for which dataset (MODIS, SEVIRI, AERONET). Also, include a standard deviation or RMS error. If any spatio-temporal averaging was done beforehand, specify so now.

Introduction:

P 2, L 3: Neither Textor nor Lee consider efforts to improve models. Textor evaluates an ensemble of models, while Lee developed a methodology to better understand a single model. True, all of this with a goal to finally improving the models but neither of these papers actually gets there. There is a vast literature out there about model evaluation and development, and more (and more relevant) references should be included.

P2, L 6: "diversity of parameterizations": there are many causes (e.g. dynamical cores, emissions, parameter uncertainty) for model diversity, parametrisation of processes is only one of them.

P2, L19: Please include at least the following: Collins et al JGR 2001 (OI), Weaver et al JAS 2005, Generoso et al JGR 2007 (3D-VAR), Zangh et al JGR 2008 (3D-VAR), Yumimoto et al. ACP 2008 (4D-VAR), Sekiyama et al. ACP 2010 (EnKF), Schwartz et al. 2012 JGR (3D-VAR), Saide et al ACP 2013 (3D-VAR), Rubin et al ACP 2016 (EnKF)

P3, L26: "The particle size distribution for each type is divided into 6 size bins, characterized by the particle average diameter and mass. Each aerosol bin is then treated as a passive tracer: aerosols are emitted, transported and removed from the atmosphere, however there are no transformations or chemical reactions between aerosol types, between size bins or with gases." To put it differently: a 2-moment scheme (mass & numbers) is used to describe aerosols in each bin but no coagulation or condensation processes are modeled? If these latter processes are not considered, wouldn't a simple mass scheme do just as well?

P4, L7: "We use daily BB emissions for better synoptic forecasts, which is not possible with the monthly mean emissions of Lamarque et al. (2010)." But this begs the question how a diurnal cycle in emissions would affect your standard model run (and your DA results for that matter). Emission datasets with (imposed) diurnal cycles for anthropogenic emissions exist for the European domain, e.g. TNO-INERIS. Why were such datasets not used?

P5, L2: "The model is then run over a cycle length (1 hour) to obtain the analysed trajectory." I could not find any clear statement over which domain observations are assimilated. Sect. 4 seems to suggest only over MEDIO2. But why then a cycle of 1 hour? New MODIS observations will happen only twice every day (at best). Since the cycle is only 1hr, isn't 3D-FGAT very similar to standard 3D-VAR?

P5, L13: Reformat "Courtier et al. (1994)"

P6, L3: "nor does it need to contain the inter-bin covariances"

P6, L6: "could be weighted by 5 different quantities, like number or mass concentration, or extinction coefficient". I disagree. If the innovation is zero (i.e. forecast AOT agrees with observations), it makes sense that the relative distribution of aerosol species does not change. This limits how you can redistribute aerosol. I also don't understand how number and masses would make a difference in your system. From an earlier description, I gather that masses and numbers in each bin are closely related (since there are no inter-bin processes or gaseous condensations).

P6, L20: "also by taking into account the hygroscopicity of sea salt aerosols" Surely the hygroscopicity of sulfate and organic carbons is considered as well? This is part of the OPAC (by the way, I believe the authors may be mistaken the climatology GADS for the aerosol optical properties database OPAC)

P6, L27 and further: I am not sure what this adds to the paper. Consider removing it or clarifying its importance.

P7, L16: "defines the observation and representativeness errors. These errors are considered to be non-correlated, which means that all non-diagonal members (covariances) in the matrix R are zero". This is an assumption that is unlikely to be true, especially given that your super-observations are aggregates over 0.2 by 0.2 degrees (a very short distance). At the very least you need to state this, better yet if you can provide justification for your assumption.

P8, L1: "As the optimal parameters, we estimated that the percentage for the errors of the model should be twice as large as for the observations (24% and 12% respectively)." I don't understand what the 24 and 12% refer to. R should be based on realistic assessment of the observational error (including representativeness error), independent of the data assimilation. So is 12% a typical observational error? But it seems estimated from Eq 13, so it is not independent from the data assimilation system.

P8, L10: "The implemented horizontal correlation length is 0.4." This is based on other studies or just an educated guess?

P8, Section 4: It appears that outside MEDIO2 no observations are assimilated? Why not? Wouldn't you want to constrain inflows across the MEDIO2 boundaries?

P8, L20: MODIS C5 is known to have systematic biases, e.g. due to windspeed and cloud cover over ocean but also over land. This has greatly improved in C6 so why did you not use this? Also, these error estimates are very optimistic (see Fig 20 in Schutgens et al AMT 2013). Even C6 has higher error estimates.

P8, L25: Earlier the authors mention representativeness errors. Do you assume these are essentially zero because of the creation of super-observations? But you will seldom have grid-filling MODIS observations so representativeness may still be an issue. Also, over distances of less than 50km aerosol is usually strongly correlated. As a result, retrieval errors and representativeness errors may also be strongly correlated. Yet you assume they are not. Please discuss this.

P9, L5: What are typical SEVIRI AOD errors over ocean? Did you consider assimilating these data (which have a much higher time resolution than MODIS)?

P9, L8: 3m and 5m should likely be 3km and 5km.

P9, L12: Why did the authors use AERONET L1.5 data? Since their experiment is from 2012, L2 data are available and generally far superior in cloud screening. It seems they do not consider AE as a observation for evaluation. Why not? Have they considered using AERONET inversion data for evaluation? SSA measurements would allow for an interesting evaluation of aerosol speciation.

P9 & 10 PCASP and LOAC: is anything known of measurement errors for these instruments (after the time-averaging)?

P10, L12: Is there a separate spin-up of the assimilation as well? Or do the authors start evaluating results after the very first assimilation cycle?

P10, L13: "The assimilation cycles in the experiment have a length of one hour." I guess most observations are made near the same time so why use a one hour cycle?

P10, L12: "(directly forecasted" Is this the direct experiment or the forecast that is part of the assimilation experiment?
P10, L22: This suggests that biases are small (RMS error ~ standard deviation) and that suggests you are comparing forecast and analysis after some spin-up of the assimilation system?

P12, L3: "MODIS overpasses each point twice" Undoubtably the authors are aware there are two MODIS instruments, each passes over a point once a day.

P 13, Fig 4: How do the authors interpret the difference in direct model run for MODIS (Fig 1) and SEVIRI (Fig 4). The first one seems unbiased but in the second one the model is clearly biased low. Is this because Fig 1 really compares the forecast and not the direct model run? If so, how does the direct model compare against MODIS and does it look similar as Fig 4?

P 18 Fig 8: I find this very interesting and a major result of this paper. Maybe panel d) can be changed to include assimilated observations during the day of the flight? It would help understand what information was available and what not.

P20, L3: "Sometimes, the generalized multiplicative change of the aerosol profile in the assimilation can produce unsatisfactory effects in some layers. For the second flight, although the concentrations in the plume are hugely improved, near the surface the increase of aerosols lead to a larger overestimation in the model." To me it seems these two sentences make more sense if their order is reversed.

P20, L14: "This demonstrates that the continuous assimilation of good quality AOD observations can correct a shape of the aerosol vertical profile, although a single AOD assimilation cycle can only expand or shrink the profile shape (as the AOD observations do not contain the information on the vertical). For the profile in Fig. 9c, by comparing the forecast and the analysis of the same assimilation window, we see that the single AOD assimilation cycle expands the profile but does not change its shape, what multiple cycles does." I agree with this assessment and think it shows nicely the power of data assimilation. However, the authors may want to reconsider earlier statements like P20, L6 “the aerosol concentration in the assimilation run at a certain height can be correct *only* if the profile shape is well simulated in the direct model run.” which is far more restrictive.

P20, L18: “multiple cycles” it may be useful to point out to readers that it is not just the multiple assimilation cycles that improve the profile but also assimilation throughout a spatial domain. So dust originating from different regions and transported at different levels is adjusted in previous cycles before it ends up in the main plume. One question is then how important the DeepBlue observations are to your assimilation. I expect quite important because they contain information closest to individual dust source regions.

P20, L23: “Therefore, we cannot follow the dust plume by following an air parcel.” I think I understand this sentence but the dust is still advected by the winds: following air parcels at different heights should give you the dust plume or all Lagrangian analyses are useless (HYSPLIT or FLEXPART).

P21, Fig 10 (panel d): is it possible to somehow plot Lagrangian trajectories as well to show how different air parcels move before contributing to the plume? This would be a very strong visual support for the nice analysis in 6.6

P22, L9: “In doing so, our assimilation system showed to be more efficient in lowering overestimated AOD values in the model than increasing underestimated values” This has been shown in several other assimilation studies as well. It’s due to model errors being defined relative to the model forecast which is unavoidable if you assume the model to be unbiased. I expect your solution (in L14) would not work: if the model substantially underestimates actual AOD (i.e. outside the model’s error envelope), adjusting observational error will not help (not to mention that it is close to fudging).

P23, L12: Also, Benedetti uses a 4D-VAR method not 3D-FGAT

P23, L19-28: This is the first time that mention is made of the very different size ranges of model and observations. The authors should state the model range (I could not find it in their paper) and how they dealt with the discrepancy (I assume they only compared number concentrations in overlapping bins). Ideally, this should be described in resp. the model and observation section but not here in the discussion. Instead of this long paragraph a single sentence that there was insufficient information to compare size distributions should be sufficient.

P24,L6: Instead of ‘influence’ I suggest ‘improve’.