

First of all, we would like to thank the reviewer for his/her in-depth review of the paper and the useful comments. The paper has been extensively re-written aiming at giving the full picture with the utmost clarity. Please see below for detailed answers to the reviewer's suggestions.

- 1. Section 1.1, Point 1. You might add something along the lines of “reasonably well documented” to the data distribution requirements.**
- 2. Section 1.1, Point 3, line 102. “... for verification and model refinement purposes, ...”**
- 3. Section 1.2, lines 118-119. Perhaps: “... data from an observational platform and simulations from a model.” Just to distinguish measurements from model results.**
- 4. Section 1.3, line 223. Perhaps: “... aerosol particle emission, secondary production, and removal.”**

The suggested sentences have been added/corrected.

- 5. Section 2.1. Could you say more about why “ensembles” in general seem to work so much better than individual models? For example, wouldn't the selection of ensemble members matter?**

For aerosols, Sessions et al (2015) and more recently Xian et al (2018, submitted to QJRMS) have shown that the ensemble is the top performer. Selection of members does not matter as long as individual members are truly independent and of roughly equal skill.

- 6. Section 3.1, line 278. Might be: “... whereas in remote regions, transports and aerosol processes control the uncertainty.”**

Added.

- 7. Section 3.1, line 288. By “DA” I assume you mean data assimilation, but this is not defined previously, as best I can tell.**

This has been defined.

- 8. Section 3.1, lines 295-296. Might be: “... temporal resolution, speciation, aerosol size distribution, and hygroscopicity.” Wouldn't hygroscopicity matter for aerosol-cloud interaction and particle removal efficiency in the models? Perhaps aerosol light-absorption properties is a bit more removed from the considerations for forecasting, but I thought I'd just mention that as a factor that might also be worthy of mention here.**

Hygroscopicity is indeed very important as it determines the aerosol optical properties and affects assimilation of, for example, AOD and in turn, the aerosol forecasts. It is now mentioned explicitly in the text.

- 9. Section 3.2, around line 305. I'm a bit confused by this discussion. “Biogenic” aerosol often refers to the aerosol produced by secondary processes from gases emitted by natural vegetation. You seem to mention all these elements in this section, but as written, it is difficult to discern how they relate to each other.**

We tried to reflect the fact that there is no agreed nomenclature that encompasses all “anthropogenic” aerosols. “Biogenic” can also refer to natural occurring primary, secondary or to aerosol emitted from anthropogenic sources which have certain organic materials in them. We agree that it is confusing. From the point of view of the paper, we just treat this category as distinct from “natural” aerosols such as sea salt and dust. **Please note that now this is section 3.4.**

**10. Section 3.2.1. Regarding the accuracy of inventories, another issue is “small” sources. For satellite-based inventories, sources too small to be detected by satellite fall into this category. To take one example, this can be a big issue for smoke inventories of agricultural-burning regions.**

The following sentence has been added to acknowledge the problem: “Moreover satellite-based inventories may miss “small” sources as it is the case for smoke inventories in agricultural burning regions.”

**11. Section 3.2.4. I’m wondering whether the requirements for particle vertical distribution, light-absorbing properties, and/or Mass Extinction Efficiency should be covered in this or one of the other subsections here.**

We agree with the reviewer that section 3.2.4 or other subsections of 3.2 could have been used to cover particle vertical distribution, light absorbing properties, and/or Mass Extinction Efficiency. However, we consider it to be more appropriate to cover these in the following sections since section 3.2 describes gridded emission inventories which estimate emission by combining activity data with emission factors, and thus the mentioned parameters are not relevant to estimate the emissions.

The afore mentioned parameters are addressed in the manuscript as follows:

Vertical distribution: in lines 497-508 of section 3.3.3 to highlight the importance of injection height in Fire emissions and then 509 and 510 to mention to importance of the injection profile and the need for observations constraining this source of uncertainty. Lines 845-851, in section 4.2.3, also stress the need for information of the vertical particle structure through vertically resolved observations. Finally, lines 1099-1103, in section 5.2, emphasise on the need to combine vertically integrated observations with vertically resolved and surface observations.

Extinction: lines 870-872 in section 4.2.5 to underline the importance of the extinction of single species in order to estimate their radiative impact.

Absorption properties: lines 884-886 of section 4.2.5 to highlight its use to constrain absorbing aerosols in the model. Lines 1129 to 1133 of section 5.3 reveals its need (together with other aerosol optical properties) to evaluate the direct and semi-direct effect on aerosol absorption properties.

**12. Section 3.2.5, lines 377-378. I think there is a typo regarding the subscripts.**

Typo corrected.

**13. Section 3.3, Line 400 ff. In small-fire regions, the required factors can be much larger. See, e.g., Petrenko et al., JGR 2017, doi:10.1002/2017JD026693.**

Reference to Petrenko et al. (2017) has been added. The sentence now reads:  
“In small-fire regions, the required factors can be much larger (Petrenko et al, 2017).  
**Please note that this is now section 3.5.**

**14. Section 3.3.1, lines 463-464. Overlying smoke opacity and fire emissivity are two additional factors that might be mentioned here (I know they are mentioned elsewhere in this section). A similar point applies to “small” fires (see point 10 above). More generally, the fire-emissions subsections seem a bit longer than necessary – a little reorganization could help consolidate and remove a bit of “rambling.” This paper is quite long, which is fine, but it would be a service to readers to consolidate as much as possible.**

This comment has been addressed with a radical consolidation of the section on biomass burning (now section 3.5).

**15. Section 3.3.1. Another, empirical approach relating FRP directly with smoke emission should be mentioned here: Ichoku, C., and L. Ellison, 2014. Global top-down smoke-aerosol emissions estimation using satellite fire radiative power measurements. *Atmosph. Chem. Phys.*, 14, doi:10.5194/acp-14-6643-2014.**

The reference has been added, thank you.

**16. Section 3.3.1, Line 501-506. Especially as this is a review paper, some references regarding inverse modeling would be appropriate here.**

Reference to Huneeus et al 2012 and Escribano et al 2017 have been added.

**17. Section 3.3.3, line 564. Another appropriate reference here, for completeness, would be Val Martin et al. 2012, Space-based observations constraints for 1-D plume-rise models. *J. Geophys. Res.* 117, D22204, doi:10.1029/2012JD018370.**

The reference has been added, thanks.

**18. Section 3.4, lines 596-597. It would be helpful here to include one more sentence, saying what they accomplished. [i.e. “A first step in creating such a network was undertaken during the recent Fenec project (Hobby et al., 2013; Roberts et al., 2017).”]**

**Please note that now this is section 3.2.**

We have added text to address this point:

“A first step in creating such a network was undertaken during the recent Fenec project, which deployed stations in 2011 (Hobby et al 2013): the deployed stations could not be maintained beyond 2013 (Roberts et al 2017), so do not provide continuous monitoring or a long climatology, but have demonstrated that (i) reporting the sub 3-minute variance in winds is generally unimportant, but resolving the diurnal cycle is critical, (ii) there are substantial biases even in analysed winds, which miss the summer-time wind maximum in the central Sahara, and (iii) that it is important to evaluate

dust uplift together with model winds, and that observational records of this relationship are invaluable (Roberts et al, 2018). “

**19. Section 3.4, lines 603-604. Would pressure or wind sensors be most efficacious in this case?** [ i.e. A much denser network of high-quality pressure observations is needed to better constrain models in this regard]

We have added text to address this question,  
“A much denser network of high-quality pressure and wind observations is needed to better constrain models in this regard. Pressure measurements have the advantage of being less affected by local (and often sub-grid) conditions (e.g. topographic circulations, inhomogeneities in roughness) than wind measurements, and have, through data assimilation, a far greater impact on the analysed heat low which, in turn, controls the model winds.

**20. Section 3.4, lines 629-633. Perhaps this reference would be useful here: J. Li et al., 2016. Reducing Multi-sensor Monthly Mean Aerosol Optical Depth Uncertainty Part I: Objective Assessment of Current AERONET Locations. J. Geophys. Res. 122, doi:10.1002/2016JD026308.**

Thanks. We have added this reference.

**21. Section 3.4 overall. I note that the desert dust section goes into much less detail about processes (e.g., surface roughness length, mobilization thresholds, etc.) than the corresponding discussion in the biomass burning section.**

We have added significant new text to address these points, covering vegetation and roughness and dust sources.

**22. Section 3.6, line 685. It might be worth mentioning here that dry removal also depends on particle size, shape, density, and hygroscopicity.**

Hygroscopicity has been added in the list of variables.

**23. Section 4.1, first paragraph. To some extent, by assimilating radiances rather than retrieved quantities, all the assumptions and issues treated in the satellite retrievals get pushed onto the model. The assumptions involved are likely to be more consistent, as you note, but not necessarily better, given the attention the modelers must pay to all the other components of the model.**

We fully agree with the reviewer on this point. This is how the sentence reads:

“The optimality of assimilating retrieved aerosol products versus radiances and the choice of a suitable algorithm or method for fast radiative transfer in the shortwave are still being debated. On the one hand direct radiance assimilation avoids the problem in the diversity between the model and the retrieval assumptions (aerosol type, refractive index, meteorological parameters, etc.), on the other hand the complexity of the observations might complicate or even prevent the implementation of radiance assimilation, especially for advanced sensors such as multi-angle instruments or polarimeters. In the end, the most pragmatic approach prevails in an operational

context, hence the assimilation currently depends heavily on the availability of good quality retrieval products with reliable uncertainty estimates.”

**24. Section 4.1, lines 750-757. Are there any appropriate references for this material?**

Several references on the various assimilation approaches have been added.

**25. Section 4.2, lines 771-777. This largely duplicates the discussion in Section 4.1.**

This has been reworded to avoid duplication.

**26. Section 4.2.2, lines 797-804. One appropriate reference here might be: Zhang, J., and J.S. Reid, 2006. MODIS aerosol product analysis for data assimilation: Assessment of over-ocean level 2 aerosol optical thickness retrievals, J. Geophys. Res. 111, doi:10.1029/2005JD006898.**

The reference has been added.

**27. Section 4.2.2, Points 1 and 2. AeroCom and AeroSat are spearheading considerable work in the area of pixel-level uncertainties for satellite aerosol retrievals. One example publication: Witek, M. et al., 2018. New approach to the retrieval of AOD and its uncertainty from MISR observations over dark water. Atmosph. Meas. Tech. doi.org/10.5194/amt-11-429-2018.**

The reference has been added, thanks for the suggestion.

**28. Section 4.2.4. In practical terms, is there specific temporal sampling that would address specific data-assimilation needs?**

This is a difficult question to answer. We tried to be constructive without being too prescriptive. This is how the section now reads:

“The issue of temporal resolution is similar to that of spatial resolution. In principle high-temporally resolved data are beneficial to the analysis, particularly because they provide information on the diurnal aerosol variability. However, issues connected to large data volume may arise. This is particularly true for datasets coming from geostationary satellites which are now providing data at with temporal resolution of 10-15 minutes. In some cases, such data have to be heavily thinned or averaged (Saide et al 2014). This is obviously only a technical limitation which might not be applicable across the range of assimilation systems. For example, the new generation of Japanese geostationary satellites, Himawari 8-9 (Bessho et al 2016) provides excellent data that have been demonstrated to be of use for data assimilation (Yumimoto et al 2016). For ground-based instruments, similar considerations can be made, although data volume might not be as high.”

**29. Section 4.2.5, lines 842-848. Shouldn't this have been covered, actually in more detail, in the earlier, measurement sections? Also, obtaining pixel-level uncertainties on any retrieval-produced speciation is another issue about which something might be said.**

All sections have been heavily rewritten for clarity and readability. This has been addressed in the most recent version of the manuscript.

**30. Section 4.2.5, lines 865-866. *The idea that dust and sea salt are “coarse mode,” whereas pollution and smoke are “fine mode,” is a gross oversimplification. Both dust and sea salt have size-distribution tails that extend into the fine mode, and often dominate the fine mode. If speciation really matters for the modeling applications under consideration, this needs to be clarified.***

The oversimplification is a by-product of modelling a complex natural phenomena at the global scale with the use of parameterizations. Of course, nature is a continuum and the distinction between coarse and fine particles is purely academic.

**31. Section 4.2.5. *How good would the particle size and AAOD information need to be to make a significant contribution to NWP?***

From the paper: “The accuracy of AAOD would need to be comparable to that of total AOD for the product to have an impact in the analysis.” Said this, any inclusion of AAOD would be beneficial.

**32. Section 4.2.5, lines 873-874. *Surface measurement will not get the transported aerosol, and where there are local aerosol sources, interpreting the results can be complex.***

We mention in the paper that the ground-based network should be relatively dense. Of course, this will still be limited to over-land. However, we feel that the surface measurements have an important role to play.

**33. Section 4.2.6. *MODIS aerosol observations are effectively continued by VIIRS. However, the data records for other instruments, such as CALIPSO and MISR, are at greater risk.***

This is unfortunate. We comment on this and refer to follow-on mission lidar such as Aeolus and EarthCARE that can help alleviate the problem.

**34. Section 4.2.7, lines 901-904. *However, errors in a slope derived from two or more spectral AOD measurements can be large. And if there are several modes in the column, interpretation of AE is not straightforward.***

This has been acknowledged. The paper now reads: “On the other hand, errors in a slope derived from two or more spectral AOD measurements can be large. Moreover interpretation of AE is not straightforward in a column where several aerosol modes are present. The usefulness of AE over AOD (or fine/coarse mode AOD) is still a matter of debate in the retrieval and assimilation communities.”

**35. Section 4. *Aside from data assimilation, are there other aerosol forecast applications, and if so, could their requirements be summarized or at least mentioned?***

The focus of the paper is on user requirements for operational aerosol prediction, and data assimilation is one of the tools to improve this prediction. We do not consider data assimilation as an end in its own but rather as a mean.

**36. Section 5.1, line 951. The residence time of aerosols is “short” compared to... This matters for the discussion here. Transported aerosol can stay aloft for days, even exceeding a week, in the troposphere. Please clarify what is meant here.**

The sentence has been expanded:

“Since the atmospheric residence time of aerosol particles in the troposphere is relatively short (from hours to ~1 week, depending on species-specific physical processes and meteorological conditions) and the footprint area of a single station may be limited, there is a need for ground-based observation networks with sufficient density and representativeness of stations.”

**37. Section 5.1, lines 957-959. However, spatial and usually temporal sampling by commercial aircraft tends to be highly skewed. You might elaborate on how the limitations affect application to aerosol forecasting.**

This is acknowledged in the text: “Data collected from commercial aircraft can provide invaluable observations for model evaluation (e.g., In-service Aircraft for a Global Observing System, IAGOS; <http://www.iagos.org/>). At the moment, however, this is not established for operational aerosol applications. Moreover due to the spatial and temporal skewness of the distribution of data collected from aircraft (often more dense close to airports), some care needs to be put into assimilating them into operational systems.”

**38. Section 5.1. This is a good summary of network capabilities. It would be helpful for the purposes of the current paper to summarize the strengths and limitations as they relate to aerosol forecasting in particular, e.g., desired site locations, coincident meteorological observations, etc. (Many networks also have a strong climate or air quality focus.)**

The following sentences have been added to the section 5.1:

“Referring to the requirements of observations as outlined in the introduction (i.e., ease of access/ consistency, uncertainty, and speed of delivery), globally consistent and available datasets such as for AOD from AERONET or NASA satellite by default currently drive the evaluation process and consequently model development. AERONET’s ability to provide high accuracy of fine and coarse mode AOD data over the globe with typical preliminary data availability within 6-24 hours makes it a favored metric variable (Sessions et al., 2015). Likewise, the maturity, coverage, speed and ease of access of MODIS aerosol retrievals makes MODIS AOD retrievals the dominate satellite verification product (as discussed in section 4 favored for data assimilation as well). This dominance of AOD to some degree is to the exclusion of perhaps more applicable baseline variables not meeting the noted observational requirement, such as PM<sub>2.5</sub> /PM<sub>10</sub> or aerosol vertical distribution. As discussed in section 3, additional evaluation variables related to model microphysics (chemical composition, absorption, size, full solar and IR radiative properties, etc.) are only sporadically available, and rarely collected simultaneously.”

And in this section is also mention that “it is mandatory to provide additional information on the observation site with a correct classification based on its spatial representation (regional or global) and its localization (environment types and emission types).”

Additionally in section 5.3 it is mentioned:

“To fully understand processes, more sites with co-located observations of visibility, cloud, radiation, vertical profiles of temperature, relative humidity as well as winds and aerosol properties would be highly desirable. Precipitation and deposition observations are also extremely relevant for benchmarking. Innovative designs for global measurement systems (existing technological platforms such as commercial aircraft, cell phones, cars, etc.) should be further exploited.”

**39. Section 5.2, line 1012. Perhaps, in one place in the paper, you could clarify what is meant by “high temporal resolution” for the applications under consideration, and elsewhere refer to that section for clarification. I get the impression that 3-hourly temporal resolution is desired for most of the applications considered here, but maybe not all.**

This is considered in the most recent version of the manuscript.

**40. Section 5.3, lines 1044-1048. This seems like a fairly comprehensive list. Are some species higher priority than others, perhaps at different locations? Generally for this section, are some quantities higher priority than others?**

It is clear that aerosol dominant species depends on the region but considering the focus of the manuscript is global aerosol prediction models all the aerosol chemical species must be in the list. We can rephrase the sentence but when you have chemical samples, you can consider all the species. By priorities we would like to say that all the observations are important but would include first size distribution, then, number concentration and finally, chemical speciation.