First of all we would like to thank the reviewer for his/her in-depth review of the paper and the useful comments. All additional reference suggested by the review were added to the manuscript. 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.

General comments:

1. It would be useful to introduce mathematically aerosol prediction as an initial and boundary conditions problem as opposed to aerosol projection, which is essentially a boundary condition problem. This will help understand that data assimilation is particularly important for aerosol prediction, while for future projection emissions scenarios are the key factor.

This has been done in the section "Aerosol Prediction Models".

"Moreover, although some of the data requirements presented here are shared with aerosol models for climate applications, here we focus on numerical aerosol prediction at the short and medium-range (up to 10 days). In this context, we are essentially dealing with an initial and boundary condition problem for which the requirements for assimilation have high priority. For sub-seasonal to seasonal aerosol prediction, which is not dealt with here specifically, requirements on ocean state and variability are also important as well requirements for the development of prognostic emission models. In the wider context of aerosol projections for climate prediction, the emphasis is much more on emission scenarios and the requirements will consequently be different."

2. At the core of any transport model, there is an advection solver. Models use different solvers, with some creating spurious waves. These numerical oscillations are generally smooth out with a diffusive scheme, creating numerical (unphysical) diffusion. These drawbacks are too important to be ignored, and I would recommend addressing them. An example of discrepancy generated by advection schemes has been discussed by Ginoux (2003). He showed that poor representation of dust size distribution in models was primarily due to the numerical solver of sedimentation.

In the section "Aerosol Prediction Models" we added the following sentence:

"Each [model] relies on different dynamical cores, advection solvers, and aerosol microphysics schemes that necessarily generate a large degree of diversity among the various models (see for example Ginoux, (2003)). The range of horizontal and vertical resolutions across the models is also very diverse, as is inline versus offline architecture."

3. A source of error considered in data assimilation is the inconsistency between simulated and observed variables. This is discussed in the manuscript but what is missing is the description of the dependent variables of the prognostic equations in these models. You should mention that dependent variables of these equations are mass/number concentrations, as it will clarify the discussion, while observations are mostly optical properties. Passing from one to the other necessitates assumptions and consequently error.

This sentence has been added in the section "Data assimilation for aerosol prediction".

"As discussed previously, most aerosol assimilation systems rely at the moment on products such as AOD, rather than raw measurements such as satellite radiances. However, the tendency in the future will likely be towards the use of satellite radiances, either raw or aggregated and possibly cloud-cleared, for consistency with the current approach in NWP. This represents a challenge for both the model developers and the data providers and might also involve joint development of observation operators. The last point is particularly true considering that there is a fundamental

inconsistency between simulated and observed variables. The prognostic variables in the model are the mass/number concentrations of the individual species whereas the observed variables are mostly optical properties. Converting from one to the other necessitates assumptions and consequently is a source of error which has to be mitigated."

4. Is ocean data assimilation not important to be mentioned for seasonal to sub-seasonal aerosol prediction? How could you make any correct aerosol prediction without representing the right phase of large-scale oscillation such as ENSO or NAO? Maybe you should add a sentence about this without developing as it is beyond the scope of the paper.

This has been included in the reply to comment 1, and reported above. We should stress that at this point most aerosol prediction systems are focussing on the short to medium-range, with most systems predicting up to day 5. Sub-seasonal and seasonal aerosol prediction is very much at its beginning. Most of the requirements will be similar to the medium-range prediction, but of course the complexity of the system will increase and the role of other drives such as ocean initialization will become more important.

5. Emission of several aerosols depends strongly on vegetation. For example, biomass burning will obviously depend on the amount of biomass, dust emission is drastically reduced in presence of any vegetation cover, and the emission of biogenic organic precursors depends on vegetation cover. It may be valuable in this paper to include data requirements for vegetation cover, as new model developments often increase the level of interactions between vegetation and aerosols. Evans et al. (2016) showed that dust variability in Australia is amplified by dynamic vegetation in agreement with satellite observations. Also, are there any recommendations to validate land model results used for aerosol prediction?

While this is a very valid point, we feel that to add requirements on vegetation cover would go beyond the scope of the paper. In the case of the dust forecast, one option is to include NRT data (with one day lag) from MODIS vegetation index in the source map instead a vegetation climatology, that it is most commonly used. Another thing is to consider this vegetation changes in the meteorological solver.

6. An additional application of aerosol forecasting model is to provide support during field campaign. The model provides direct information on aerosol optical thickness and concentrations for effective flight planning, while feedbacks from measurements constantly evaluate the model for successful model improvements (Chin et al., 2003).

This has been added to the section "Aerosol Prediction Models".

"These systems are used for various applications, including, but not exclusively, global air quality forecasts (dust and biomass burning), operation impacts, boundary conditions for regional systems and flight campaign planning (Chin et al., 2003).

SPECIFIC COMMENTS:

Section 1.2. I would suggest adding some sentences related to above comments 1 to 3.

This has been addressed.

Section 1.3. You may want to mention the use of forecasting model to support field campaigns (see above comment 6).

See above.

Section 3.2.2. Last paragraph. Increasing resolution does not necessarily mean better model skills. It may request new tuning of parameters of subscale processes (e.g. orographic gravity wave drag), as well as larger ensemble runs due to large variability. I wish I could propose a reference related to aerosol, but Kapnick et al. (2018) discusses such issues for the prediction of snow over the western US.

We agree with the reviewer. This sentence has been added in the section "Aerosol Prediction Models":

"In general, increasing resolution does not necessarily mean better model skills as it may request new tuning of parameters of subscale processes (e.g. orographic gravity wave drag), as well as larger ensemble runs due to high variability."

We did not include the reference as not directly relevant.

Section 3.4. This section on dust and the following on sea-salt are much shorter than the previous section related to biomass burning. Is there a justification for it? Section 3.4. No discussion on dust sources, which is the base of any dust study and modeling. All dust models use a form or another of preferential dust sources defined by Prospero et al. (2002) and adapted for global models (Ginoux et al., 2001; Tegen et al., 2002; Zender et al., 2003; Ginoux et al., 2012). These source functions were necessary because soil properties from global inventories (e.g. FAO) were and still are unrealistic in arid and hyper-arid regions. Although, model representing the physical processes of dust emission have been around for a long time (e.g. Marticorena and Bergametti, 1995; Shao, 2001), they have to be adapted to accommodate major discrepancy in soil texture datasets, the driver of dust emission. There is the interesting work of objectively compare different dust source inventories (Cakmur et al., 2006). It may be adequate to perform similar exercise with more recent inventories.

This section is now 3.2
We agree that these aspects were under-emphasised in the submitted paper and have added paragraphs that address these points. Moreover the other sections were also shortened and re-organized to make the whole paper more homogeneous and readable.

"For a reliable prediction of mineral dust aerosol, sufficiently accurate knowledge of both the emitting soil and the deflating winds is needed. Both aspects suffer from insufficient observational constraints, creating a large challenge for quantitative emission predictions. Important source regions globally include the Sahara/Sahel, Southwest Asia/Middle East, Taklimakan/Gobi deserts of China, Australia and the Southwest United States/adjacent Mexico (Prospero et al 2002). However, larger source regions show substantial fine structure and throughout the world there are also many individual sources such as in Patagonia, the Arctic plains, and countless dry or drying lake beds. Estimating dust emission sources can also be performed from satellite data (for examples see Huneeus et al (2012), Schutgens et al (2012), Yumimoto and Takemura (2013), Escribano et al (2016), Escribano et al (2017), Di Tomaso et al (2017)).

Dust models typically employ maps of dust source functions (e.g. Zender et al (2003), Ginoux et al (2012)), because soil properties in arid and hyper-arid regions from global inventories are insufficient to provide consistent soil texture information. This includes aspects such as soil particle size distribution and binding energies but also the existence of roughness elements and soil moisture content that impact on mobilization thresholds. See Darmenova (2009) for a comprehensive review. This severely limits the level of complexity that can be put into models representing the physical processes of dust emission (e.g. Marticorena and Bergametti (1995), Shao et al (2001), Kok et al

(2014)). In order to get a better understanding of the involved uncertainties, an update to the objective comparison of different dust source inventories by Kakmur et al (2006) would be desirable and could be extended to take into account uncertainties in the dust emission parameterization itself."

Section 3.4. Not one word on soil texture, soil moisture, vegetation cover, and mineralogy, although these are key elements to simulate dust emission, distribution and optical properties. I would recommend including them in a paragraph with references.

We have added substantial text on this,
"In addition to that, dust emission is further complicated by suppressing influences of soil moisture (Fecan et al, 1998) and vegetation cover, including brown vegetation from a previous rainy period (Kergoat et al, 2017), which can vary on relatively small time and spatial scales. This is particularly acute in the semi-arid Sahel with its seasonal vegetation, also creating large variations in surface roughness (Cowie et al, 2013). There is currently a debate to what extent the mineralogy of emitted dust particles should be taken into account, as this would alter both its interactions with radiation (Journet etal 2014) and cloud microphysics. While certainly an interesting field of research, the former aspect is probably more relevant on longer timescales, while the latter is not even considered in most current dust prediction models."
"Finally, the dust-focused satellite data should be complemented by improved space-born assessments of soil moisture and vegetation cover (green and brown) to better characterize varying conditions in source regions (Kergoat et al 2017)."

On the other hand, there is a discussion on the difficulty to represent sub-scale dry and wet convection. These are important processes for dust emission, but it may be better to discuss boundary layer parametrization in a "meteorological" section. Why are you mentioning 3 field campaigns? And these ones in particular, are the others less important?

"The wind requirements for dust modelling are quite distinct from those of other components of the model, since the uplift occurs as a result of very rare high wind-speed events, the "tail" of the very wind distribution. The importance of processes such as the day-time breakdown of the LLJ and haboobs are specific to dust modelling as such we have found that discussion of these issues is best placed in the dust section, rather than a general meteorological section.
In the original draft we refer to the AMMA, Fennec, BoDeX and JADE field campaigns (as well as the CV-DUST project and Cape Verde Observatory) as these have deployed networks of stations, in some cases long-lived, in remote areas similar to those that we believe are required to improve dust in NWP. The authors have been involved in other less campaigns, such as GERBILS and SAMUM, which we have not cited as they are less relevant here, but would welcome other specific suggestions for campaigns we have missed.

Section 3.4. Satellite data. You mention IASI but there are more than 2 groups working on retrieving dust from the data. Geostationary satellites have their own quality for aerosol prediction, and SEVIRI has been quite useful to retrieve dust sources (Schepanski et al., 2007), or detect haboobs (Ashpole and Washington, 2013). Also, I would mention the promising results from GRASP algorithm (Chen et al., 2018).

We have added a reference to Klüser et al., 2012 for IASI.
We already discuss use of geostationary infrared data, but have added a note on its use for source detection, "Infrared products are being developed but still have biases related to atmospheric

moisture (Banks et al., 2013). These would need to be further improved and provided in near-real time for data assimilation, but have been useful for source detection (Schepanski et al., 2007)."
NWP needs information on the land surface, dust and dust-generating winds, independent of the uplift mechanism, so we have not discussed haboob detection.
We have added a reference to GRASP,
"or those produced with the GRASP algorithm (Chen et al., 2018) are promising but have more limited space-time coverage."

Section 3.5. There is no mention of the temperature dependency of sea salt emission. Most models are now considering it, specifically the parameterization of Jaegle et al. (2011)

This sentence has been added to the section on sea salt emissions (**now section 3.3**).

"Jaegle et al (2011) found discrepancies between modelled and observed marine aerosol concentrations correlated with sea surface temperature; significant improvement in agreement was found when the model sea spray source function was modified to include a temperature dependence. This result is consistent with a number of laboratory studies…"

Section 3.6. This section is detailing removal processes of one model (NAAP), but they are generally treated quite differently in other models. It reads as a technical report of the NAAP model. Also, it seems that important processes are missing, such as in cloud scavenging, Bergeron process, etc. It would be more useful to learn about the method to parameterize the different physical processes rather than learning what is useful or not to run NAAP.

NAAP is not a specific model, but stands for Numerical Atmospheric Aerosol Prediction.

Section 3.6. Line 703-705. I would mention the work by Yu et al. (2017), which allows evaluating dust deposition by combining MODIS and CALIOP data.

The following sentence has been added:

"Recently, Yu et al (2015) tried to infer dust deposition by combining MODIS and Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) data. Their CALIOP-based multi-year mean estimate of dust deposition matches better with estimates from in situ measurements and model simulations than previous satellite-based estimates."

Section 4.2.2 Line 790. Reference(s) would be useful.

References have been added.

Section 4.2.5. This section is again focusing on model (CAMS) to discuss its problems. Why should I care about this model if I am not using it?

Specific mentions to the CAMS model have been removed.

Line 874: There is method to derive aerosol speciation from AERONET (Schuster et al.,2005), and more recently there are promising possibilities with GRASP algorithm (Torres et al., 2017)

References have been added. The sentence now reads:

"Wherever direct speciation measurements are possible, those would be best suited to be used to correct model prediction of a given aerosol species. These could be measurements derived from a (relatively dense) network of ground-based instruments and/or from satellites. Some promising

results to derive aerosol speciation from AERONET observations have been obtained by Schuster et al. (2005) and more recently by Torres et al. (2017) using the Generalized Retrieval of Aerosol and Surface Properties (GRASP) algorithm.