Review of “Monitoring aerosols over Europe: an assessment of the potential benefit of assimilating the VIS04 measurements from the future MTG/FCI geostationary imager” by Descheemaecker et al, 2018

General comments

This paper presents the results from an OSSE study to assess the impact of the FCI AOD observations from a geostationary instrument on continental scale PM10 analyses. The study is significant as there is a lot of attention for new instruments on geostationary instruments and their usefulness in air quality applications. OSSEs are useful tools to evaluate the value of the new instruments prior to their operation or even prior to their production. In the study the authors have put large effort in the realism of the OSSE and demonstrating this, which is positive and provides trust in the results.

The paper is well structured and leads you through the different steps of the experiment.

The main comments I have are related to the choices made in the study and the amount of discussion.

- To identify the value of the new instrument it should ideally be compared to the current situation. Currently most operational models assimilate surface observations of PM10 and PM2.5 and some also AOD observations from e.g. MODIS. It would be beneficial to show the added value of the FCI observations on top of these common observations. How beneficial are these observations from the geostationary instrument as compared to observations from an orbiting satellite with lower temporal resolution. Perhaps an experiment could be done with assimilation of only one observation per day, as compared to the hourly observations, do you see an added value? Can the satellite observations replace or add value to the surface observations (which are much cheaper). This is something is really miss in the discussion to support the value of this instrument for aerosol modelling.

- I would suggest more room for discussion of the results. There is a large focus on the areas/cases where it goes well, but some discussion is required on the situations where it does not work so well. The assimilation distributes the increments based on the fractions, so it can not correct for errors in the size distributions. The same holds for the vertical distributions. The AOD does not provide any vertical information, so what happens when the vertical profile is wrong, such as for example in figure 16 bottom right, you can see that the system does not work in these cases. So when the AOD in underestimated but the surface PM is overestimated (or the other way around), the AOD observations will increase the PM at all levels also at the surface, leading to worse results. This discussion on AOD-PM relationship and the importance of having correct vertical distributions or vertical information is missing and should be added. Especially since you are showing that your NR shows overestimation of AOD and underestimation of PM, when you would use that version of the model for assimilation I think you would get problems.

Specific comments

- The model runs are performed at a resolution of approximately 20km which is soon not really representative anymore for regional air quality forecasts/analysis. Many models already run at a 0.1 deg resolution. Especially since the observations are available at a 2km resolution I wonder why it is not chosen to run at a higher resolution. Somewhere it is mentioned that the goal of the study is to evaluate the impact for continental modelling, Please elaborate a bit more on this choice and for the

discussion part I would add the potential to look at the impact at urban scales, maybe in a follow-up study.

- For the vertical distribution of the aerosol emissions, a fixed profile is used as I understand correctly, is this not depending on the source type of the emissions, e.g. car emissions at the surface and emissions from industry higher up?

- The inclusion or exclusion of SOA is not clear throughout the paper. Somewhere it is mentioned that SOA is added to the NR and CR1 by using a percentage of the primary carbon species. But further along it looks as if the SOA is not included in the computation of AOD (p9, line 16-17), is that correct? Your nature run is the ‘‘real world’ so then also the synthetic observations should include the SOA. What is the impact of excluding SOA in the AOD synthetic observations?

- Section 3 is very hard for me to follow, it is not within my field of expertise but I get the impression that the AOD is only computed from the model concentrations without taking into account any radiate transfer modelling? Is this a correct assumption from my side? because then the sensitivity of the instrument to different altitudes in the atmosphere is not taken into account which can lead to overoptimistic results. On the other hand the errors are computed in a very accurate way. A lot of attention goes to the simulation of these errors, which are very important for the realism of the OSSE. However I think the amount of figures/tables and text dedicated to this part of the study is out of proportion and needs condensation and rewriting.

- p 11, Filtering: A lot of observations are removed due to filtering. This is an important comment. The added value of geostationary satellites lies in their temporal resolution. If you only retain 1 to 4 observations per day, is there still a large added value, is this representative of the future real situation? Please add some discussion

- Location of observations, the observations seem to be concentrated over central Europe, how representative are the results at the AQeR for other regions? You can see that the plots for the entire domain provide different conclusions than the plots for the AQeR stations.

- Validity of CR-NR: The statistical metrics have been compared to metrics from literature. Two different papers have been used, but if I am correct both evaluating the ensemble of models which is always better than the individual models. Have you also compared to individual model results. The CR-NR seem to be smaller than the NR – real observations difference.

- Only one observation per hour is used because of the system. As is suggested in the conclusion I would make a super observation. It is mentioned that avoiding overoptimistic results is one of the reasons for this choice but I do not agree as this will probably be done once the observations become available in real life, so I do not see why this would lead to overoptimistic results

- Spatially 3 out of 4 observations are removed trough filtering for convergence of cost function, while already a lot of observations are removed due to clouds etc. Is a larger impact of the observations foreseen without this spatial filtering?

Technical corrections

- p1., line 17 Abstract, change 4-month**s** to 4-month

- p2. Line 6. which lead → which lead**s**

- p2, line 13, Only the WHO limit values are mentioned, but it would also be good to include the official EU limit values.

- p2, line 20, there are more appropriate references for the CAMS services, please add the website and the paper from Marécal which is use further along in the paper.

- p2 line 30, MODIS is now also available in a 1x1 km product (MAIAC)

- p4, line 10-11 this sentence is unclear to me, do you mean by combining AOD and error characteristics?

- p4, line 12, CR, which should represent……(something like the current situation, the situation without use of the observations)

- p4, line 25-26. Also when you are using two different models, you should evaluate this.

- p4, line 27, → as MOCAGE is used for **both**….

- p7, line 20, FGE is **also** used

- Figure 2, I find it very hard to see the NR background in central Europe with all the overlying circles and the small plots. It is mentioned that the variability and maxima are well represented, but I cannot evaluate this when I do not see the background.

- p8, lines 20-24. The underestimation is indeed common, I do think there are many more possible reasons for this, such as underestimation of emissions in cold winter periods, and perhaps the modelling of stable winter conditions with shallow surface layers.

- P8, line 16: maxima, I would change this word, as I relate maxima to the absolute maximum values, while I think you mean the location of the mamixum values.

- p12, line 5 slowest → slower

- p13, line 2, especially the CR4: but the bias for CR4 is quite small….

- p13, line 24, here the purpose of the paper is mentioned, but this should be stated more clearly in the introduction, especially the focus on the continental scale.

- p14, line 25-28 Figure 14 versus figure 12, I found it hard to see the improvement, while a large improvement is mentioned, maybe it would be helpful to direct the reader to some specific areas where it is visible. Tables 9 and 10 are clear but only cover central Europe.

- p15, lines 5-10 please add here the discussion of AOD-PM relation as suggested in the general comments

- p15, line 29-20, what is meant with high spatial and temporal episode?

- p16, summary, please also add the case where it does not work (simulation 4 at the surface, averaged over whole domain).