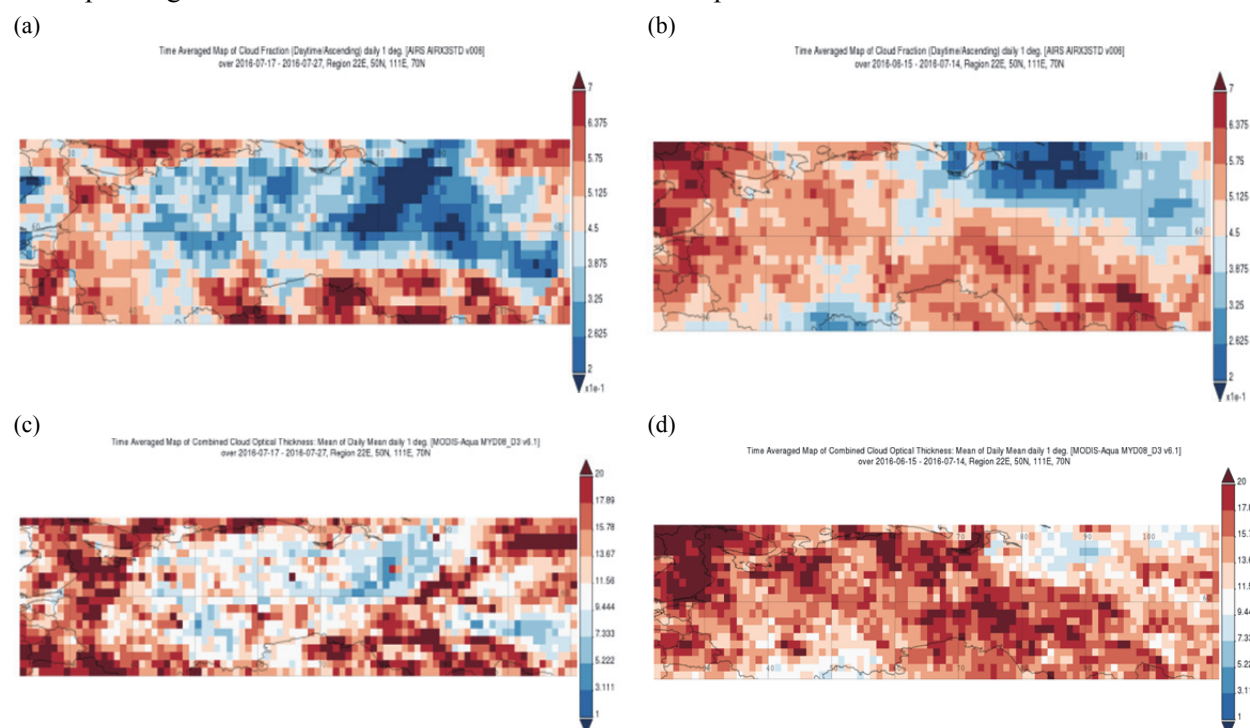


Referee's comment: *At P17L9, the approach to harmonize the model and satellite data, in terms of temporal and spatial sampling is described in detail. One important aspect, which I did not see described: as part of the satellite/model harmonization, how was the model data filtered for retrieval quality, by which, masking out of retrievals under cloudy conditions is meant? At P5L23, the predominantly cloudless conditions during this period are mentioned, but a quick inspection of the MODIS true color imagery and retrieved AOD show a mix of cloudy and clear sky conditions during the second half of July 2016. Under cloudy conditions, much of the AOD is masked out, as is the case for the OMI AOD and AAOD retrievals (for L3, at least, and presumably for many of the L2 pixels and individual CO retrievals). Is CTM filtered accordingly to not introduce a discrepancy due to the inclusion of simulated data under cloudy conditions where the retrievals fail? This could even be as basic as an ad-hoc threshold of cloud fraction or cloud optical depth, depending on what is available, so that the CTM data are 'biased' toward clear sky conditions in the same way as the satellite data. Or are the CTM cloud fields in sufficient enough agreement with those seen by the satellites that the retrieval co-location sampling handles this?*

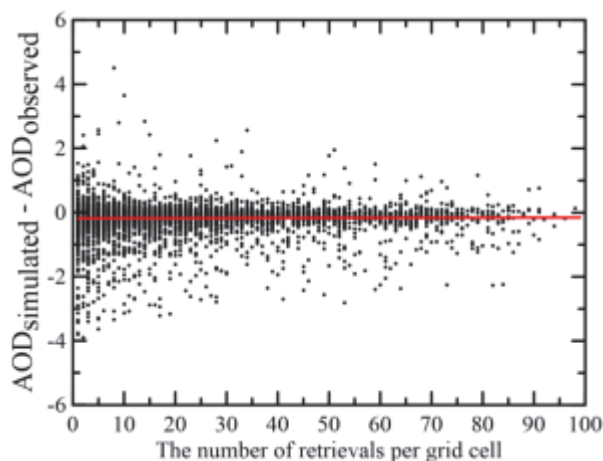
First of all, we would like to clarify our remark concerning “predominantly cloudless conditions” during the BB aerosol outflow event. According to the Level-3 AIRS retrievals (see Fig. C1 below), the cloud fraction was indeed typically less than 0.5 over the region (40–80°E, 55–65°N) and period (17–27 July) most affected by the outflow of BB plumes from Siberian fires. Furthermore, the cloud fraction was smaller and clouds were less dense than typical for a summer period in the same region (cf. panels (a) and (c) with panels (b) and (d) in Fig. C1). That said, we recognize that our remark could be confusing, particularly because the cloud fractions inferred from different observations can differ significantly. The corresponding sentence is corrected in the revised manuscript.



**Figure C1.** Time-averaged maps of (a, b) the cloud fraction and (c, d) the cloud optical depth according to the data retrieved from the AIRS and MODIS observations, respectively, for the periods (a, c) 17 – 27 July 2016 and (b, d) 15 June – 14 July 2016. The data are shown only for the study region. Source: Giovanni, <https://giovanni.gsfc.nasa.gov/giovanni/>, last access: 1 November 2020.

Our satellite/model harmonization procedure did not involve any filtering of the model data with respect to the cloud coverage. However, since each available retrieval falling into a given grid cell was matched with the corresponding simulations in time and since grid cells/hours corresponding to the continuous cloud coverage are barely represented in the satellite observations, both the satellite data and simulations are “automatically” weighed towards the clear sky conditions. Furthermore, as part of a preliminary analysis, we examined whether or not the lack of satellite retrievals for cloudy scenes could introduce significant systematic discrepancies between the simulated and modeled data fields by plotting the relationship between the difference of the observed and retrieved values of AOD in a given grid cell in

the study region and period and the number of retrievals per grid cell. This analysis is illustrated in Fig. C2 provided below. If there were strong systematic discrepancies between the observations and simulations due to the lack of observations for cloudy scenes, then the difference of the observed and retrieved values would probably depend on the number of retrievals per grid cell (since this number is inversely proportional to the fraction of cloudy scenes). However, we did not observe such a dependence and therefore concluded that biases introduced into our analysis by cloudy scenes in the satellite observations are not significant. We included a concise discussion of this point in Sect. 2.5 of the revised manuscript.



**Figure C2.** The differences between the simulated AOD values (for the ‘bb\_vbs’ scenario) and their counterparts retrieved from the MODIS observations as a function of the number of AOD retrievals per grid cell. Both the simulated and retrieved data were gridded on a  $1^\circ \times 1^\circ$  model grid with the hourly temporal resolution and were not subject to any pre-selection procedure. The red line shows the best linear fit to the data. Only each 10<sup>th</sup> data point is shown in the plot to preserve its readability. Note that the simulations shown here do not include the background part, which is the reason for a small negative bias in the simulated values.

Referee's comment: P1L12: *In the abstract, consider a basic description of importance of how the VBS scheme improved agreement between the model and satellite data (i.e. in Fig 4 b,d). This point is worth mentioning.*

We thank the referee for this useful suggestion. The corresponding remark is included in the abstract of the revised manuscript.

Referee's comment: P21L6: *In Figure 2b and 2d, I would suggest that the AOD scale range from, say, 0-2, to get a better sense of the AOD enhancement over the receptor region centered on Moscow.*

We have re-drawn Fig. 2 using the scale suggested by the referee. The AOD enhancement associated with anthropogenic emissions from Moscow (if this is what the referee meant) is now clearly seen in the simulations (Fig. 2d), but the corresponding enhancement in the observations is more smeared (Fig. 2c). These spatial differences between the observed and simulated AOD values are typically less than 0.1 and are likely due to either a minor positive bias in the MODIS retrievals or underestimation of the AOD associated with biogenic aerosol by the model. These minor differences could not significantly affect the results of our analysis. In addition, we would like to note that the receptor region is somewhat shifted from the center of the plume northward in order to exclude the impact of minor local fires (in the south-western part of the domain, see Fig. 1) on the analysis results.

Referee's comment: P21L9: *by correlation, do you mean spatial pattern correlation or something else?*

We meant the correlation between the daily values of the spatially averaged retrievals and simulations. This is clarified in the revised manuscript.

Referee's comment: *Minor editorial point throughout: use the word ‘rather’ fewer times.*

We thank the referee for this useful suggestion. We tried to avoid using the word ‘rather’ in the revised manuscript unless it is necessary.