Reviewer 1 (Andrew Sayer)

I'd like to thank Andy for his dedication to this paper, providing many useful comments and suggesting several improvements.

I am posting this review under my own name (Andrew Sayer) due to ongoing collaborations with the author. I feel I am able to provide an unbiased review. I reviewed the previous version of this manuscript, and recommended major revisions (mostly organizational/clarity rather than technical points). This revised manuscript has a more readable structure and better balance of figures. My technical comments raised in the previous round of review have also been addressed. Similar suggestions were made by the other referees.

For me, a stand-out conclusion is that hourly collocation results in a more reasonable comparison between AERONET-like and model aerosol fields than daily or yearly. I suspect the same may be true for satellite AOD fields as well (although that is not a topic of this analysis). This suggests to me that it would be worthwhile generating hourly composites of AERONET data and satellite level 3 products, for a more direct comparison with model fields. This is important because a significant component of the representation error here is systematic due to clear sky bias (i.e. the average representation error tends to be negative, and going to hourly makes this less negative, as well as reducing the variation of the representation error). So I see this as a major action item on the data producer community.

A second conclusion is that going from 1 degree grid boxes to 0.5 degree grid boxes has a much smaller influence (for AOT). This suggests that the 1 degrees used by many data sets now may be a sufficient balance of detail and storage overhead. However, for AAOT it seems the difference becomes more important (mean error and mean signless error roughly halve going from 1 to 0.5 degrees). Therefore, as AAOT comparisons become more prominent, perhaps it is worth the community (modelers plus satellite data sets) making the shift from 1 to 0.5 degrees as well.

These two conclusions are highlighted by Figure 4 (and for North America, 5) and Table 6 for AOT, and Figure 12 for AAOT. Personally, I would suggest emphasizing this by adding to the title of the paper, "implies that hourly colocations will result in more meaningful, representative data comparisons" or something similar. This way the conclusion is even clear from the title. I leave this up to the author, however, as it's a question of style preference.

I do have a few further comments:

Figure 10: My understanding is that the left part of this Figure should match the middle part of Figure 4. Both are the representation error in yearly AOT, at a 1 degree grid, based on level 2 direct Sun AOT sampling. However, if one compares the numbers above the figure, they are not quite the same. Is this an error, or am I misunderstanding something? This should be checked and either corrected or clarified.

Well spotted. The difference is due to the selection of stations. In Fig. 10 I included only stations that observe 12 months out of the year since I am comparing yearly to monthly errors. In Fig. 4 I included all stations, so also those that might not have any observations during some months of the year. This different selection entirely explains the small differences between Fig. 4 and 10. I have added text to the caption of Fig. 10 to explain this. Note that in both cases, only stations below 60 degree N/S latitude were used.

Conclusions: In my previous review I had raised the question of "is a take-away that AERONET and satellites should provide hourly aggregates" and the author's response was "But don't these data already come at hourly or daily resolution?" They come at daily but not hourly, and this study suggests there can be quite some benefit from going to hourly. The Conclusions currently states that hourly colocation is beneficial, but I do think it is worthwhile to explicitly state "it would be good for data providers to provide hourly aggregates of certain parameters (e.g. AOT) to facilitate these comparisons". I doubt that many data users will be motivated to go back and create their own hourly aggregates from available highest-resolution data so if we want people to take this recommendation, I think it should be framed as a direct request to them (AERONET plus satellite). I know this study was not directly about satellites but again, if satellites are to be used as an evaluation resource for the models, the same points apply.

I believe I was thinking of satellite data; at least for polar orbiters daily and hourly data will not be different (except at high latitudes). But the benefit of hourly (instead of daily/monthly) data is a point worth making, I will add this to the Conclusions.

I also think it would be worth adding a couple of sentences about the possible merits of 0.5 degree grid boxes (for satellites, I think MISR and SeaWiFS are the only publicly-available data sets at this resolution?) compared to the 1 degree that is much more standard. Again, unless it is framed as an ask and data providers act on it, data users are unlikely to create them. Of course for e.g. AeroCom this is also dependent on model resolutions so perhaps that is a topic for discussion at the next AeroCom meeting. Still I would mention this in the Conclusions so it sticks in the mind of the reader.

I am not sure my study allows conclusions to be drawn about optimal *satellite* aggregation levels. I'm assuming the truth in this case would be the 1 or 0.5 degree averaged AOD. In the case of clear-sky AOD, a perfect satellite retrieval would not show a representation error! In the case of all-sky AOD, I think it would depend on the length-scales involved: one for the obscuring cloud field; one for the observed aerosol field; which aggregation level performs better. In my 2017 ACP paper, I considered all-sky representation errors given a 1 degree satellite aggregate and truth grid box of 0.5 to 3 degrees (Fig. 15), showing that a mismatch between these sizes can easily double the representation error. But that is an answer to an altogether different question and I consider this issue of optimal satellite aggregation level unresolved.

Other than that, I recommend that the paper be accepted for publication. I am happy to read any further revision if that would be helpful.

Reviewer 5 (Tero Mielonen)

I'd like to thank Tero for his work on this paper. His thorough reading found several typing errors and inconsistencies and helped to improve this paper.

As I'm collaborating with the author on another topic, I'm writing this review with my own name. Despite this collaboration, I believe that I'm able to provide an objective review of this manuscript.

In the manuscript Schutgens uses high resolution simulations (GEOS5) to evaluate the spatio-temporal representativity of AOT and AAOT observations done at AERONET and GAW sites. The topic is scientifically very interesting and the analysis is well executed. The author has satisfactorily addressed the concerns of the reviewers on the first round. I recommend that the manuscript is accepted for publication in ACP after minor revision.

General and specific comments:

Page 1, line 19: "correlate strongly throughout the year", I'm not sure if I understood this correctly. Do you mean that the monthly representation errors correlate with each other or something else?

They correlate well with each other, i.e. January errors correlate strongly with February errors but less strongly with June errors (although the correlation is far from zero). I have modified the text.

Introduction: Several abbreviations (e.g. AERONET, GAW, AAOT, AEROCOM) are mentioned in the text but not defined. Abbreviations should be defined in the abstract and then again at the first instance in the rest of the text.

Done.

Page 3, line 21: Was there a minimum number of observations required for the calculation of an hourly average? Or do you assume that even a single observation is representative enough?

Here a single value is assumed sufficient.

Page 3, line 37: "although for dust and biomass burning aerosol higher AOT at 440 nm ≥ 0.5 were needed", this was hard to follow. Do you mean that for dust and biomass burning aerosols the SSA errors are in the 0.03 range only for AOTs larger than 0.5?

Yes, text has been clarified.

Page 3, line 56: Please clarify here what are the GAW-ABS measurements and how do you calculate the AAOT from them. "Surface properties" are mentioned which makes me think about aerosol surface properties but I'm guessing you mean ground-based in-situ observations of light absorption coefficients at some wavelenght(s)? If I guessed right, then how do you calculate the AAOT from the absorption coefficients? Do you assume some kind of a vertical profile and integrate that?

I do not use GAW observations, only their geolocations. I then assume that a future GAW-ABS network will include AERONET like stations.

Page 4, line 47: Figure S4 doesn't seem to include any sites above 60 degrees latitude.

Correct. The Figure stops at 60^o N. I have corrected the text.

Page 5, line 11: "the simulation captures spatial variation rather well", this seems to be true on yearly time scale but do you know if it holds for shorter time scales as well?

Fig 1 of course suggests that temporal variation at the AERONET sites is captured reasonably well but I do not know if that also means that, say, spatial variation in monthly averages is captured well. I redid the analysis in Fig 1 for the months of January and June with the following results for mean AOT:

For a minimum of 100 observations per site:

	Pcorr	*Nr of sites*
January	*0.89*	*47*
June	*0.89*	*96*

For a minimum of 10 observations per site:

	Pcorr	*Nr of sites*

January	*0.76*	*126*
June	*0.83*	*145*

Compare this to Pcorr=0.75 and Nr=216 for a minimum of 100 observations per site in Fig. 1

I would conclude that this also holds on monthly time-scales.

Page 5, line 16: "overestimation of dust", could you clarify here that do you mean the dust load or dust AOT? Overestimation of dust load could be explained by differences in meteorology and consequent changes in dust emissions but overestimation of dust AOT could also be influenced by the optical properties of dust used in the simulation. Did you check how the comparison looks if you separate Africa into northern and southern part? As the northern part is dominated by dust and the southern part by biomass burning aerosols, the analysis could help disentangle the contributions from dust and carbonaceous aerosols.

This text refers to results by Gelaro et al. They studied AOT (not loads) and their analysis was based on global averages.

Page 5, line 39: I'm not sure if you are aware, but AERONET Inversion L1.5 data has a handy flag called If_Retrieval_is_L2(without_L2_0.4_AOD_440_threshold). You could use that to relax the AOT limit but not the other requirements for L2.0.

Good to know, I have found it in the V3 data. The flag is not present in V2 data which were used in this paper.

Page 5, line 77: "Inversion data is generally closer to the equator", not sure what you mean with this. Do you mean that inversion data has larger SZAs even though the sites that produce the most Inversion data are close to the Equator? This could be related to the differences in the measurement principles (direct vs. almucantar).

Yes, likely. It's a pattern in the data but I do not think it's important to the present study.

Page 5, line 78: I didn't understand how the overestimation of dust AOT is related to the observational coverage. Could you please clarify?

Notice that here I am talking about L2.0 Inversion data. One important constraint is that AOD@440nm > 0.4. If G5NR overestimates AOT at dust sites (which are over-represented in real Inversion data), this will lead to larger temporal coverage in the OSSE than in the real observations.

Page 5, line 85: Instrument malfunction and maintenance will likely affect all observations, not just inversion products, so they are not likely to explain the difference between comparisons with direct and inversion data.

True. I have adapted the text to reflect this.

Page 6, line 25: You mentioned in the replies to the first round of reviewers that likely reason for the regional gradients is cloudiness. I believe it would be good to mention that in the text as well. This is an interesting detail because the MODIS AOT also has/had this kind of east-west trend over US. To my understanding, that was caused by land surface properties/orography.

I think this is explained in the following paragraph, p. 6 l. 28-40

Page 6, line 31: wet growth → hygroscopic growth

Done.

Page 6, line 39: Doesn't the usage of clear-sky data make sense also for the AERONET observations, especially inversion products, as they are based on observations from cloudless parts of the atmosphere?

Yes, and this is accounted for. No simulated AERONET site can observe under cloudy conditions. But the AERONET site "sees" only a small part of the atmosphere in a 1 or 2 degree grid-box. So part of that grid-box may be cloudy and yet the site may still be able to make observations.

Page 6, line 101: "substantial reduction in representation error can be seen for for r > 1 sites", this is true if you compare r = 0 and r > 1 sites but there doesn't seem to be such a big difference between r =1 and r > 1 sites, at leas based on the mean errors.

> should have been >= . This has now been changed.

Table 6: Please, clarify in the heading what does the "90 %" stand for.

90% quantile. This has now been clarified.

Section 5.2: There's a large gap between the heading and the text.

Very rarely Latex will do this to make other bits and pieces fit better. I will keep an eye out for it in final production.

Page 7, line 50: Anthropogenic emissions didn't have a diurnal cycle but biomass burning did. Did you look at the results from South America and southern Africa in this perspective? These kind of regional analysis could strengthen the conclusions here.

An interesting idea but it would probably require substantial extra work. First there is the question of realism of the biomass burning diurnal profiles. G5NR's documentation (Putman et al. 2014) is scant on detail but I gather a very simple diurnal profile has been imposed on-line (i.e. it is not part of the QFED emission dataset). In comparison, the diurnal profiles used in Schutgens et a. 2016, 2017 were based on statistics of traffic, industry etc. and integral to the emission dataset. In these two papers, substantial differences between daily and hourly collocation were found. Second, sites affected by biomass burning are also affected by other sources and there is no easy way to separate the two contributions. At the very least one would have to identify sites affected by biomass burning. A very preliminary and coarse attempt at such an analysis yielded nothing.

That said, Fig. 2 shows that AAOT diurnal variation is under-estimated in G5NR, we know that the anthropogenic sources for absorbing aerosol have no diurnal cycle, and in Schutgens et al. 2016 (which does include realistic diurnal profiles) daily representation errors are significantly larger than hourly errors.

Page 7, line 60: "Sect. 6, f", there's an extra "f" at the end of the line

Corrected.

Page 8, line 42: The shift from spatio-temporal to spatial representation errors comes rather suddenly. It would help the reader if there would be a short description (and a reference) how the spatial representation errors were calculated, either in this section or in the methods section.

Agreed. I have included a brief explanation (at the start of the section).

Page 8, line 69: Thank you for sharing this ranking data with easy access! Would it be possible to include also AERONET sites above 60 degrees latitude? You mentioned in the text that near the poles the simulated pixels become too small but is it an issue already at 70 or 80 degrees latitude?

It is currently not possible to quantitatively assess when this becomes an issue but a G5NR grid-box at 60º latitude is only 50% (3.5 km) the area of a grid-box near the equator (6.9 km). A grid-box at 70º is only 34% (2.4 km) and a grid-box at 80º only 17% (1.2 km). At the same time, the real field-of-view of an AERONET site does not change and will remain in the order of 5-10 km. The current code is built on the assumption that a single G5NR grid-box represents this FOV.

Section 6: If I understood this correctly, Kinne's ranking is based on site centered analysis whereas in this study the grid was fixed so the sites may not be in the center of the grid boxes. For finer grids this probably doesn't matter much but it might affect the results at 4 degree grid. What is your view on this?

Correct, Kinne's original ranking is based on grid-boxes centered around the site, in the current paper, however, the grid itself is fixed and sites may be near the edges of a grid-box. Note that in the case of model or satellite L3 evaluation, people will be forced to interpret Kinne's ranking in the latter way.

The analysis presented in the current paper (Fig 9) shows the impact of r on representation errors. Unless r and the location of sites within a grid-box are in some weird and unexpected way correlated, I do not expect an impact on the conclusions that I draw.

Page 8, line 3: You mention several examples where the behaviour of site specific representation errors differ from the "rules" defined on the basis of all sites. It would be an interesting and valuable addition if you could give some explanation why things do not go as expected. Does it depend on local meteorology, aerosol types or something else?

Yes, yes and something else as well ☺. I do not know if G5NR provides sufficient data to disentangle these aspects but it would be a lot of work. In addition, the deeper you delve into a model, the more apparent will become its deficiencies.

While I understand the interest in the causes, I'm not sure what could be learned from it at this stage. The most important 'lesson' is that one cannot take general 'rules' derived from the entire dataset and expect them to apply to each individual site. This is not surprising: local meteorology can be expected to cause a lot of variation in representation errors even when sources remain fixed and constant.

Page 9, line 26: "ground-based remote sensing observations", can you say it like this? If I understood correctly, the GAW absorption observations are in-situ observations.

Yes and yes. I have assumed sun-photometers at the GAW-ABS that in reality do not exist. See also my earlier comment.

Page 10, line 24: There's something wrong in the author list: "K??rcher"

Corrected.

Figures 4, 6, 7, 9, 10, 12, S2, S3, S5-S8: What are the black circle and bar? I'm guessing mean and median, but which is which?

Bar is median, circle is mean. This is now explained at the end of Sect. 3.

Figure 8: Did you check how the representation errors behave as a function of AOT? I think GAW stations are often designed to observe the background concentrations, meaning lower AOTs, so I'm just wondering if the difference in the altitude dependence between AERONET and GAW is caused solely by topography or do aerosol concentrations also play a role. Of course they are linked so it is hard separate their effects.

I would expect higher GAW stations to see less AOT (shorter air column, farther from sources) and this is indeed the case. However, the correlation of representation error with AOT is lower (-0.3) than with altitude (-0.5) suggesting maybe that height is the dominant factor.

Figure 10: The mean statistics for yearly errors in this figure are not exactly the same as in Figure 4. Shouldn't they be the same? Then another question about the monthly representation errors. How can you calculate monthly representation errors from yearly averages (the brown bar)?

There is a small difference in selection of sites. In Fig 10 only sites that provided observations 12 months out of the year were used. This has now been explained in the caption. As to your second questions, I can't! That brown bar represents a monthly collocation protocol. Caption has been changed to reflect this.

Figure 11: How does the number and spatial distribution of the sites change during the year? I would guess that not all sites provide data constantly throughout the year due the seasonal changes and maintenance. Would the graph look the same if all the months had the same sites?

Exactly for that reason this analysis was done using only sites that provide observations 12 months out of the year. Caption has been updated to reflect this.

Site representativity of AERONET and GAW remotely sensed AOT and AAOT observations

Nick A.J. Schutgens[1]

[1]Department of Earth Science, Vrije Universiteit Amsterdam, 1081 HV Amsterdam, the Netherlands

Correspondence: Nick Schutgens (n.a.j.schutgens@vu.nl)

Abstract. Remote sensing observations from the AERONET (AErosol RObotic NETwork) and GAW (Global Atmosphere Watch) networks are intermittent in time and have a limited field-of-view. A global high-resolution simulation (GEOS5 Nature Run) is used to conduct an Observing System Simulation Experiment (OSSE) for AERONET and GAW observations of AOT (Aerosol Optical Thickness) and AAOT (Absorbing Aerosol Optical Thickness) and estimate the spatio-temporal representativity of individual sites for larger areas (from 0.5^o to 4^o in size).

GEOS5 NR and the OSSE are evaluated and shown to have sufficient skill, although daily AAOT variability is significantly underestimated while the frequency of AAOT observations is over-estimated (both resulting in an under-estimation of temporal representativity errors in AAOT).

Yearly representation errors are provided for a host of scenarios: varying grid-box size, temporal collocation protocols, and site altitudes are explored. Monthly representation errors ~~are shown to correlate strongly throughout the year~~show correlations from month to month, with a pronounced annual cycle that suggests temporal averaging may not be very successfull in reducing multi-year representation errors. The collocation protocol for AEROCOM (AEROsol Comparisons between Observations and Models) model evaluation (using daily data) is shown to be sub-optimal and the use of hourly data is advocated instead. A previous subjective ranking of site *spatial* representativity (Kinne et al., 2013) is analysed and a new objective ranking proposed. Several sites are shown to have yearly representation errors in excess of 40%.

Lastly, a recent suggestion (Wang et al., 2018) that AERONET observations of AAOT suffer a positive representation bias of 30% globally is analysed and evidence is provided that this bias is likely an overestimate (the current paper finds 4%) due methodological choices.

1 Introduction

As the temporal sampling of observations is often intermittent and their field-of-view limited, the ability of observations to represent the weather or climate system is negatively affected (Nappo et al., 1982). This adverse effect can be described through a representation error, which allows comparison to e.g. observational errors or model errors.

Representation errors have been receiving more attention recently, in a variety of fields: solar surface radiation (Hakuba et al., 2014b, a; Schwarz et al., 2017, 2018), sea surface temperatures (Bulgin et al., 2016), trace gases (Sofieva et al., 2014; Coldewey-Egbers et al., 2015; Lin et al., 2015; Boersma et al., 2016), water vapour (Diedrich et al., 2016), cloud susceptibility (Ma et al., 2018) and even climate data (Cavanaugh and Shen, 2015; Director and Bornn, 2015). In the field of aerosol, most work has been on the representativity of satellite measurements (Kaufman et al., 2000; Smirnov, 2002; Remer et al., 2006; Levy et al., 2009; Colarco et al., 2010; Sayer et al., 2010; Colarco et al., 2014; Geogdzhayev et al., 2014), either using satellite data or model data. A new development is the use of local spatially relatively dense measurement networks (Shi et al., 2018; Virtanen et al., 2018).

As aerosols are known to vary over short time and spatial scales (Anderson et al., 2003; Kovacs, 2006; Santese et al., 2007; Shinozuka and Redemann, 2011; Weigum et al., 2012; Schutgens et al., 2013), aerosol studies are likely to experience large representation errors. Indeed, Schutgens et al. (2016b) (S16b hereafter) showed that representation errors due to temporal sampling in both satellite and AERONET (AErosol RObotic NETwork) observations were of similar magnitude as actual model errors and often larger than observational errors. Similarly, Schutgens et al. (2016a) (S16a hereafter) showed that the narrow field-of-view of in-situ measurements could lead to large differences from area averages (monthly RMS differences of $10-80\%$ for ~~201×210~~ 210×210 km^2, depending on the type of measurement and the location of the site). Recently, Schutgens et al. (2017) (hereafter S17) considered the combined impact of spatio-temporal sampling on the representativeness of remote sensing data (both satellite and ground-based). They provide representation uncertainty estimates and optimal strategies when dealing with different observing systems (ground networks, polar orbiting satellites with varying revisit times, or geostationary satellites).

In this paper, a global one-year high-resolution simulation of the atmosphere (GEOS5 Nature Run) is used to conduct an Observing System Simulation Experiment to estimate representation errors for remote sensing measurements of aerosol optical thickness (and its absorptive counterpart) as observed by the global networks AERONET and GAW (Global Atmosphere Watch). In S16a and S17, regional high-resolution simulations covering a month were used to study representation errors. This prevented an analysis of such errors world-wide and on longer time-scales. In addition, the limited spatio-temporal domains made evaluation of the high-resolution simulation difficult. These issues are addressed in the current study. Note that the current paper does not replace previous work (which also considers satellite, in-situ and flight measurements) but extends it. In addition, the current study allows us to evaluate a recent suggestion by Wang et al. (2018) that representation errors in AERONET AAOT observations are positively biased (by $\sim 30\%$) which would help to explain the observed underestimation of AAOT in global models (Bond et al., 2013).

Representation errors are not only determined by observational sampling but also by how these observations are put to use. If observations are used to evaluate models, different protocols (or strategies) exist to temporally collocate model data and observations. For instance, within AEROCOM (AEROsol Comparisons between Observations and Models), an oft-used strategy is daily collocation: daily averages of observations are collocated with daily model data. The different sampling of model and observations throughout the day are ignored (e.g. most remote sensing observations only observe a small part of the diurnal cycle). In contrast, hourly collocation uses hourly model data that is collocated with hourly averages of observations. S17 showed that in the case of remote sensing observations daily collocation allows significantly larger representation errors than hourly collocation. A third protocol would be yearly collocation which is seldom used these days in model evaluation as it yields large representation errors (S16b). However, if remote sensing observations are used to construct a yearly climatology, effectively a yearly collocation protocol is used.

In data assimilation the representation error is often (but not always) thought to include effects from incorrectly modelled sub-grid processes. In this paper, the representation error is purely thought of as resulting from the different sampling by observations and models.

Section 2 describes the high-resolution simulation data and AERONET observations used in this study. The OSSE for estimating representation errors is briefly explained in Sect. 3 but more details can be found in S17. An evaluation of the high-resolution simulation with a particular focus on its use in an OSSE is given in Sect. 4. While the simulation shows deviations from AERONET observations, the agreement is deemed sufficient to study representation errors. Representation errors in AERONET AOT & AAOT are studied in Sect. 5. A ranking of AERONET sites in terms of their representativity is given in Sect. 6. As may be expected, the paper finishes with a summary of the conclusions (Sect. 7).

2 Data

2.1 GEOS-5 Nature Run

The GEOS-5 Nature Run (G5NR here-after) is a 2-year global, non-hydrostatic simulation from June 2005 to May 2007 at a 0.0625^o grid-resolution (~ 7 km near the equator). Not just a simulation of standard meteorological parameters (wind, temperature, moisture, surface pressure), G5NR includes tracers for common aerosol species (dust, seasalt, sulfate, black and organic carbon) and several trace gases: O_3, CO and CO_2. The simulation is driven by prescribed sea-surface temperature and sea-ice, daily volcanic and biomass burning emissions, as well as monthly high-resolution inventories of anthropogenic sources (Putman et al., 2014). As it is a nature run (i.e. no meteorological nudging), the meteorology in G5NR can deviate substantially from the actual weather in 2006.

Aerosol in GEOS-5 are calculated using the Goddard Chemistry, Aerosol, Radiation, and Transport (GOCART) module (Chin et al., 2002) that uses 15 tracers to describe externally mixed species of organic carbon, black carbon, sulphate, sea-salt and dust. Biomass burning emissions are obtained from QFED (Quick Fire Emissions Dataset) (Suarez et al., 2013) with a diurnal cycle imposed online. Anthropogenic emissions of organic and black carbon use EDGAR-HTAP (Emissions Database for Global Atmospheric Research-Hemispheric Transport of Air Pollution) emissions (Janssens-maenhout et al., 2012) which were rescaled to match AEROCOM Phase II emissions. Non-shipping anthropogenic SO_2 emissions come from EDGAR v4.1.

Evaluation of G5NR (Gelaro et al., 2015) against NASA/GMAO MERRA (Modern-Era Retrospective analysis for Research and Applications) Aerosol Reanalysis (da Silva et al., 2012) suggest that global organic carbon, black carbon and sulphate AOT are underestimated by $30-40\%$ while dust AOT is overestimated by $\sim 50\%$. Global sea-salt AOT is similar to MERRA within 10%. Hence, Castellanos et al. (2019) derived global rescaling factors for aerosol speciated AOT in G5NR but these are not used in the current study (true scaling factors are unlikely to be global, it is unclear what to do about AAOT and the focus here is on relative errors anyway). Comparison with AEROCOM models shows that G5NR sulphate life-times are quite low (at 2.7 days) while the other species fairly agree with the AEROCOM multi-model mean. G5NR shows reasonable cloud fractions compared to CERES-SSF (Clouds and the Earth's Radiant Energy System-Single Scanner Footprints), although in the equatorial/sub-tropical region (30S-30N), G5NR has a deficit of partially cloudy scenes. In addition there are too few clouds off western continental coasts and the southern branch of the ITCZ is too strong. CALIOP (Cloud-Aerosol Lidar with Orthogonal Polarization) data suggests G5NR cloud fraction are too low, especially over equatorial/sub-tropical lands in the Northern Hemisphere, and too high in the northern polar region.

For this study, the following hourly G5NR data for 2006 were obtained: see Table 1.

Table 1. G5NR data used in this study

short name	description
totexttau	aerosol total column extinction at 550 nm
totscatau	aerosol total column scattering at 550 nm
swtdn	TOA* downward short-wave radiation
cldtot	total cloud area fraction
phis	surface geopotential height
bceman	monthly anthropogenic burning BC emissions
bcembb	monthly biomass burning BC emissions

*: Top Of Atmosphere

2.2 AERONET observations & geolocations

AERONET (Holben et al., 1998) data were obtained from `https://aeronet.gsfc.nasa.gov`. For 2006, AOT from Direct Sun Version 3 L2.0 (Giles et al., 2019; Smirnov et al., 2000) and AOT & AAOT from Inversion Version 2 L1.5 and L2.0 (Holben et al., 2006) were logarithmically interpolated to values at 550 nm and averaged over an hour. For all years starting in 1992, geolocation data were obtained for all sites (1144 in total).

The DirectSun dataset contains only AOT (at multiple wavelengths). These observations are based on direct transmission measurements of solar light and have high accuracy of ± 0.01 (Eck et al., 1999; Schmid et al., 1999). The Inversion dataset contains both AOT and AAOT (at multiple wavelengths) and these observations are based on measurements of scattered solar light from multiple directions. This inversion uses radiative transfer calculations (Dubovik and King, 2000) and yields larger errors than the DirectSun measurements. In particular, Dubovik et al. (2000) showed that Single Scattering Albedo (SSA) errors decrease with increasing AOT and estimated SSA errors of ± 0.03 for water-soluble aerosol at AOT at 440 nm ≥ 0.2 ~~although for dust and~~ and for dust or biomass burning aerosol ~~higher AOT~~ at 440 nm ≥ 0.5~~were needed~~. Consequently, one important distinction between Inversion L1.5 and L2.0 data is a minimum threshold of AOT at 440 nm ≥ 0.4 used in the latter (improved cloud screening is another distinction). Inversion L2.0 is a subset of the L1.5 dataset. For an intercomparison of AERONET SSA with flight campaign data, see Schafer et al. (2014).

In the current study, only AOT at 550 nm is used and the Inversion L2.0 AOT at 440 nm criterion is adapted to AOT at 550 nm ≥ 0.25. This is the minimum value of AOT at 550 nm present in actual Inversion L2.0 data, but also corresponds to AOT at 440 nm = 0.4 for small particles (Ångström exponent = 2.1). As a result, the OSSE in this paper is rather lenient when it comes to selecting valid observations similar to Inversion L2.0.

2.3 GAW geolocations

GAW geolocation data were obtained from NILU (Norwegian Institute for Air Research). Two networks were used: the GAW-AOT network which comprises 29 sun-tracking sun photometers that measure AOT; and the GAW-ABS network which comprises 81 filter instruments that measure surface properties. The real GAW-ABS network is not capable of measuring a columnar (A)AOT but here we will assume it does, similar to AERONET, and consider its representation errors.

3 Method: analysis of representation errors

The representation error is defined as the difference between a perfect observation (i.e. no observational error) and a truth value (area average), see also S16a and S17. Here, a self-consistent high resolution simulation will be used to generate both observation and truth in a so-called Observing Systems Simulation Experiment. The representation error may refer to instantaneous values or time averages. This work concerns itself mostly with yearly averages (and some monthly averages). For instantaneous and daily error values, see S16a and S17. The mapping from G5NR data to the data used in this study is given in Table 2.

Perfect observations are generated from the high-resolution simulation by choosing the data at the location of an AERONET or GAW site and sub-sampling those data in time according to certain conditions for solar zenith angles (SZA), cloud-fraction and AOT. Table 3 lists the threshold conditions for which observations will be possible. Values for SZA and AOT are inferred from real AERONET data

Table 2. Mapping from G5NR data to data used in this study

G5NR	this study	units
totexttau	AOT	
totextau-totscatau	AAOT	
$\frac{180}{\pi}\,\mathrm{arccos}\,(\mathrm{swtdn}/1367)$	SZA	degrees
cldtot	cloud fraction	
phis/9.81	geopotential altitude	m
bceman+bcembb	BC emissions	kg/m^2 s

files. The maximum cloud-fraction was tuned to obtain similar temporal coverage of observations as real AERONET data (see Sect. 4 and Fig. 3 but the impact of tuning is small).

Table 3. Conditions for valid AERONET observations as simulated in this study

source	maximum SZA	maximum cloud-fraction	minimum AOT
DirectSun L2.0	80^o	0.01	0.0
Inversion L1.5	80^o	0.01	0.03
Inversion L2.0	80^o	0.01	0.25

The truth is generated from the high-resolution simulation by averaging AOT and AAOT over a large area (0.5^o to 4^o grid-boxes) and further averaging in time. Here we should distinguish three different protocols depending on how one intends to use the observations, see Table 4. In the case of a gridded climatology derived from observations, the truth should be an average over a continuous long-term time range (say a year). In the case of model evaluation, it is possible to resample model data to the times of the observations. E.g. within the AEROCOM community, a daily collocation protocol is often used, where daily model data is used for days with observations only (irrespective of the temporal sampling of those observations throughout the day). To assess representation errors in this case, the truth needs to be sampled accordingly to days with observations before yearly averages are determined. The same protocols were also explored in S17.

The current methodology differs slightly from S17 in that:

1. a different model is used to construct the OSSE,

2. previously, SZA was assumed to be sufficiently high for a fixed fraction of the day (10 hours). In the current work, SZA is calculated from downward-welling TOA SW radiation and will vary with geo-location and time-of-day,

Table 4. Collocation protocols

collocation protocol	purpose
yearly	gridded climatology
daily	model evaluation (AEROCOM)
hourly	model evaluation

3. previously, the truth was generated for grid-boxes centered on the observations. In the current work, those grid-boxes are assumed regularly spaced from 0^o to 360^o longitude and -90^o to 90^o latitude. The AERONET and GAW sites can be located anywhere within those grid-boxes (at their real geo-location),

4. previously, the high-resolution simulation had a constant grid-size of (about) 10 km. In the current work, the grid-size varies but has a constant angular size of 0.0625^o ($\sim$ 7 km at the equator).

The last point implies that the simulation grid-box used for the observation decreases towards zero as we approach the poles. Since this is clearly undesirable (field-of-view will remain on the order of several kilometers), we will limit our analysis of *representation errors* to latitudes below 60^o. The ~~exceptions are the Figures 5and S4~~exception is Fig. 5.

Our methodology allows separation of the factors that determine the representation error: spatial extent of the grid-box, and observational intermittency due to low SZA, high cloud-fraction or low AOT. We will not present such *causal analysis* in this paper (see S17 instead) but will refer to it to explain results.

To show the distributions of representation errors, box-whisker plots using the 2, 9, 25, 75, 91 and 98% quantiles will be used in this paper (in addition, the median is shown as a bar and the mean as a circle). For a normal distribution, these quantiles will be equally spaced. Any skewness or extended wings in a distribution will be readily visible. In addition to quantiles, the values of mean error and the mean sign-less error will be provided. The mean sign-less error is deemed more relevant than the standard deviation as 1) it includes biases; 2) the errors are seldom normally distributed, and a standard deviation is very sensitive to larger errors ("out-liers"). For a normal distribution with a mean of zero and a standard deviation of one, the mean sign-less error is ~ 0.8. The correlation used in this paper is the Pearson correlation coefficient that assesses linear relationships. Regression slopes were calculated with a robust Ordinary Least Squares regressor (OLS bisector from the IDL `sixlin` function, Isobe et al. (1990)). This regressor is recommended when there is no proper understanding of the errors in the independent variable, see also Pitkänen et al. (2016).

4 Evaluation of G5NR and OSSE

In this section, G5NR is evaluated with real AERONET observations of AOT and AAOT, with special focus on its usefulness in an OSSE. As G5NR generates its own meteorology that deviates from 2006, one might expect differences between simulation and observations. Simulated data were nevertheless collocated to the time of the observations (within the hour) to ensure the same temporal sampling throughout the days, the months and the year.

The mean and standard deviation of AOT and AAOT per site are shown in Fig. 1, top row. In general, simulated site-mean AOT shows good agreement with the observations with correlations around 0.75 and slopes around 0.84. Simulated site-mean AAOT does not agree as nicely with the observations but there is still correlation (0.48) (the evaluation of AAOT will be affected by large measurement errors). The agreement in standard deviation suggests that simulated and observed AOT ~~and AAOT~~ (and AAOT) show similar temporal variation. But the global agreement also suggests that the simulation captures spatial variation rather well. This is also true on shorter length scales, as an analysis by region shows in Table 5. Europe appears to be the exception but this is mostly due to a few southern sites. As the table shows: without those sites, correlation increases significantly. This may be related to the overestimation of dust and underestimation of carbonaceous & sulphate aerosol in G5NR (Gelaro et al., 2015), which will affect north-south gradients in AOT in Europe. DRAGON (Distributed Regional Aerosol Gridded Observation Networks, see Holben et al. (2018)) campaigns might allow evaluation of the spatial distribution of simulated AOT at even smaller length-scales (10's of kilometers) but are not available for 2006.

Table 5. Correlation in modelled and observed yearly site-mean AOT

region	nr	correlation
World	216	0.75
Europe	55	0.26
Europe*	26	0.68
Africa	32	0.86
Asia	34	0.82
N. America	49	0.81
S. America	13	0.91

*: southern AERONET sites removed from analysis

The top row of Fig. 1 was created using only sites that provide a minimum of 100 real observation throughout 2006. The lower row shows how this criterion affects results. As the minimum number of observations per site increases, so do the correlations, probably due to a reduction in statistical noise (partly due to different simulated and actual meteorologies). But the overall bias also increases. This criterion selects for sites with lower cloudiness (higher number of observations) until predominantly northern African and Saudi Arabian sites are left for a minimum of 500 observations per site. The increase in bias is thus likely due to the overestimation of dust AOT that was mentioned earlier.

Note that AAOT is here evaluated with L1.5 data. The L2.0 data have a minimum AOT threshold which results in fewer observations and fewer available sites overall. Although L1.5 is considered a less reliable product, the evaluation with L2.0 (which now uses a minimum of 30 observations per site) yields a similar but slightly poorer result for G5NR, see Fig. S1, and over a shorter range of values.

Figure 2 shows mean values per site for the daily difference in maximum and minimum AOT. Again, good agreement for simulated AOT is seen but AAOT compares rather poorly. However, its correlation is still above 0.6 and it is clear that the simulation *underestimates* daily AAOT variation. The impact of AAOT measurement error on daily variation is likely reduced as the variation is a difference between two measurements (pers. comm. with T. Eck and O. Dubovik).

Figure 3 shows the temporal coverage (or frequency of observation) per site as a function of latitude. G5NR's simulated coverage is calculated using the conditions described in Table 3 (and explained in Sec. 3). This coverage would be 100% if observations are available 24 hours a day, 365 days a year. In practice it cannot be higher than 50% due to the day-night cycle, and will be less due to cloudiness or low AOT.

The bimodal structure that is visible in both the simulation and observations is due to SZA variation (which reduces coverage towards the poles) and cloudiness (which reduces coverage near the equator). Simulated and real coverage per site are not expected to agree well due to meteorological differences and down times from site maintenance. Still, the results suggests that the OSSE predicts similar frequency of Direct Sun observations as actually observed.

However, the OSSE also simulates more Inversion observations in the Northern hemisphere than actually occur. This suggests there are additional factors in observational coverage that are not accounted for in Table 3. One factor is that real Inversion measurements are attempted less frequently (several times per day) than Direct Sun measurements (several times per hour). Other factors may include inversion failure at low SZA (real observations show that Inversion data generally have larger SZA than DirectSun data even though Inversion data is generally closer to the equator) and overestimation of dust AOT in G5NR (largest over-estimates of coverage occur for Sahara and Saudi Arabia sites). Yet another issue is that successful inversion requires a high degree of azimuthal symmetry in the measurements. In essence, this is a built-in check on the magnitude of *spatial* representation errors which will lower temporal coverage of the observations but is not considered in the OSSE due to lack of information.

Finally, instrument malfunction & maintenance are not taken in to account, although that would affect AOT coverage as well.

In all, it seems that G5NR can realistically simulate spatial and temporal variation in AOT and AAOT, at least on the scales accessible by the available observations. There is some underestimation of daily AOT variation and significant underestimation of daily AAOT variation. G5NR can also be used to fairly realistically simulate frequency of observation (temporal coverage), although it will over-estimate this for the Inversion products in the Northern Hemisphere.

5 Results

5.1 Representation errors in yearly AOT

Figure 4 shows yearly representation errors for AERONET DirectSun L2.0 AOT observations as a function of model grid-box size, for the three collocation protocols (see Table 4). As grid-box size changes from 4^o to 0.5^o, errors for hourly collocation are more than halved from 13% to 5% while those for daily collocation change only from 17% to 12%. In contrast, errors for yearly collocation ($\sim 22\%$) are dominated by temporal sampling and do not depend much on grid-box size. Smaller representation errors for hourly collocation can also be seen in a regional analysis, see Fig. S2. The hourly collocation is especially beneficial when using the Inversion L2.0 AOT product, which allows large representation errors due to the condition of a minimum AOT at 440 nm ($>= 0.4$) for valid observations, see Fig. S3, even though it results in a global 9% bias.

The impact of collocation protocol can also be shown through the total number of sites that yield errors larger than, say, 10%: 821 (yearly), 653 (daily), 235 (hourly) out of 1108 AERONET stations in total, for a grid-box of $1^o \times 1^o$.

In addition to larger representation errors in general, the yearly and daily collocation protocols also allow significant biases across the AERONET network. Regionally, spatial patterns with east-west or north-south gradients in the representation errors exist, see Fig. 5 and Fig. S4. Such patterns are absent or at least much reduced for hourly collocation.

The biases in regional and global distributions of representation errors for yearly and daily collocations are strongly affected by cloudiness. Higher humidity in the cloudy part of a grid-box increases AOT through ~~wet~~ hygroscopic growth The area averages used to calculate representation errors have been derived for the *entire* grid-box (all-sky), both clear and cloudy parts. Representation errors for clear-sky parts of grid-boxes are lower for the yearly and daily collocation protocols, see Fig. 6. In certain situations, it seems more realistic to use only the clear part of the grid-box in calculating representation errors: e.g. when the grid-box average stands in for an aggregated satellite product. In this paper, focus will be on the all-sky representation error.

Table 6 shows absolute values of the yearly representation errors for different collocation protocols (yearly and hourly) and grid-box sizes. The statistical metrics provided are the mean of the sign-less representation error over all AERONET sites, and the 90% quantile of the sign-less representation error (an indication of the large representation errors possible for some sites). Using absolute values allows a comparison with the AERONET AOT measurement error of 0.01 (Eck et al., 1999; Schmid et al., 1999). This is the error for individual measurements, and not that of a yearly average which is likely to be much smaller. Clearly, representation errors are larger than measurement errors.

Results so far suggest that the daily collocation is a significant improvement over the yearly collocation. This is in contrast to S17 (Fig. 7) where the representation errors for daily and *monthly* collocation were found to be similar. The absence of diurnal (anthropogenic) emission profiles in G5NR may cause underestimation of representation errors for the daily collocation in the current study.

It is interesting to compare the representation errors of two different networks, AERONET and GAW. AERONET was not designed with representativity in mind but the GAW network was. Nevertheless, Fig. 7 suggests that GAW sites exhibit slightly larger representation errors than AERONET. In particular, GAW error statistics are strongly skewed to negative values. In the G5NR OSSE, GAW sites are located at higher altitudes and more often on isolated mountains than AERONET sites (G5NR site altitudes correlate very well with real altitudes, $R = 0.98$, but tend to underestimate by 28 m on average, with a standard deviation of 171 m). A look at yearly representation errors for the hourly collocation reveals a systematic altitude dependence, see Fig. 8. A high altitude site on an isolated mountain will observe a shorter atmospheric column than the surrounding grid-box (most of which is at lower altitudes) which will cause a negative representation error. Note that AERONET sites do not show this dependence on altitude for 1^o grid-boxes, probably because they are located more often on mountains surrounded by similar mountaineous terrain.

Finally, a comparison is made with a previous study into AERONET representation errors (Kinne et al., 2013). Using a range score r, see Table 7, they ranked sites according to their representativity for larger domains. This ranking is subjective in that it is non-quantitative, based on personal knowledge of the sites and only defines representativity in broad terms. The range scores are only available for sites that had at least 5 months of data before 2008. Using the methodology of this paper, representation errors were calculated for all sites of a certain range score, see Fig. 9. For large grid boxes of 4^o (~ 450 km near equator), the impact of the range score on representation error is quite small. While there is a visually arresting change in the error distribution for $r > 1$ (wide flanks are changed into a broader center), the mean sign-less error barely changes. This rather weak dependence on range score suggests that Kinne et al. (2013) overestimated the size

Table 6. Absolute representation errors for AERONET sites

metric	protocol	4^o	2^o	1^o	0.5^o	0.0625^o
mean	yearly	0.043	0.042	0.042	0.042	0.044
	hourly	0.021	0.015	0.011	0.008	0.000
90 % quantile	yearly	0.086	0.079	0.082	0.083	0.086
	hourly	0.052	0.033	0.029	0.017	0.000

of the domains ($>= 500$ km for $r > 1$) for which their sites were representative. On the other hand, for a grid-box of 1^o a substantial reduction in representation error can be seen for ~~$r > 1$~~ $r \geq 1$ sites. However, this only occurs for the hourly collocation: Kinne et al. (2013) did not consider the temporal sampling of the observations which causes large representation errors. An alternative ranking of representativity will be introduced in Sect. 6.

5.2 Representation errors in monthly AOT

Surprisingly, monthly representation errors are not that much larger than yearly errors, see Fig.10. If monthly errors for the same site were independent and random, one would expect them to be $\sim \sqrt{12} \approx 3.5$ larger than yearly errors but that is not the case. As a matter of fact, monthly errors are strongly correlated from month to month, throughout the year, see Fig. 11. The increase in correlation with January after September, is probably due to yearly cycles in meteorology and emissions and very likely to be a realistic aspect of representation errors. The implication of this is that multi-year averages may not reduce representation errors as strongly as one would hope.

This analysis also provokes the question whether representation errors (per site) should be seen as mostly biases or random errors (see also Schwarz et al. (2018)). Preliminary analysis suggests that both cases can occur. I.e. some sites show large variations (including sign changes) in representation error from month to month, and as a consequence a strongly reduced yearly representation error. Here monthly representation errors may be interpreted as mostly random. Other sites show monthly representation errors with not much variation and as a consequence yearly representation errors are similar to the monthly errors. There the representation error is better characterised as a bias. A proper analysis of this would require significantly longer time-series of data than are currently available. Further discusion of this can be found in Sect. 6.

5.3 Representation errors in AAOT

The discussion of representation errors for Inversion L1.5 AAOT will be shorter than that for DirectSun L2.0 AOT, as the main conclusion is identical: the hourly collocation yields smaller representation errors than the other protocols, see Fig. 12. Note also that representation errors in AAOT are of a similar magnitude as for AOT. One obvious difference is that AAOT representation errors tend to be positively biased while the AOT errors were negatively biased. While the latter was due to cloudiness as discussed before, the positive bias for AAOT is more difficult to explain. It appears that a combination of conditions (location of the sites, necessity of day-light, clear skies and a minimum AOT of 0.03) together conspire to create these positive biases. Only over the Amazon can a simple explanation be found: the clear sky condition prevents many observations outside the biomass burning season, explaining large positive biases for yearly collocation (see also Fig. S5, discussed later).

Even more than for AOT, representation errors for AAOT are very similar for the daily and hourly collocations. As discussed before, this is likely due to the absence of diurnal (anthropogenic) emissions profiles. The daily variation of AAOT is strongly underestimated by G5NR (see Sect 4 and Fig. 2).

For completeness' sake, an analysis of AAOT representation errors for different regions (Fig. S5), different products (Fig. S6), different networks (Fig. S7) and different range scores by Kinne et al. (Fig. S8) are given in the supplement. Overall the conclusions are very similar to those for AOT.

The similarity in general behaviour of representation errors for AOT and AAOT should not be taken to mean that these errors are identical per site. As discussed in Sect. 6, ~~f~~ representation errors for AOT and AAOT ~~or~~ at individual sites can be very different. Ultimately this is due to the different sources of AOT and AAOT which leads to different spatio-temporal distributions in the atmosphere.

5.4 Comparison to recent results from Wang et al. '18

Recently Wang et al. (2018) suggested that the observed underestimation of AAOT by AEROCOM models (Bond et al., 2013) may be due to *spatial* representation errors. Spatial representation errors are entirely due to the narrow field-of-view of AERONET observations (i.e. the intermittent temporal sampling of these observations is ignored). Their analysis found that AERONET Inversion L1.5 AAOT representation errors exhibit a global bias of

Table 7. Range scores for AERONET sites in (Kinne et al., 2013)

range score	spatial domain	number of sites	comments
0	100 km	120	includes mountainous sites
1	300 km	106	
2	500 km	28	
3	900 km	6	

30% for $2^o \times 2^o$ model grid-boxes, which would help explain the aforementioned underestimation by the global models. As AERONET sites need to be serviceable, they are often found near roads and urban build-up, i.e. near sources of absorbing aerosol. Compared to the larger area of global model grid-boxes, these sites would quite naturally observe larger AAOT. Thus, Wang et al. (2018) concluded that at least part of the underestimation of modelled AAOT is an artefact, created by the location of the AERONET sites.

Wang et al.'s idea is quite persuasive and indeed one can see evidence of such positive representation errors in Fig. 13 where sites in major cities like London, Paris, Madrid and Barcelona clearly exhibit positive representation errors. (For another example, see Fig. 3b in S17 concerning surface black carbon concentrations). But Wang's study found such biases for the majority of AERONET sites, not just a few located in big cities. As a matter of fact, the current study shows no evidence of this global bias of 30%. Instead it finds a global bias of only 9%, dominated by a few sites with large positive representation errors (median bias over all sites: 4%).

Wang et al. (2018) performed an analysis very much like the one in this study with one crucial difference. As they did not have a global simulation at high resolution like G5NR, they downscaled results from a standard global simulation at $2.5^o \times 1.27^o$ resolution. The downscaling was accomplished with the help of a high-resolution ($0.1^o \times 0.1^o$) black carbon emission map (Wang et al., 2016). It is possible to simulate this procedure using the high-resolution G5NR black carbon emission maps and AAOT simulations (the AAOT simulation was first coarsened over $2^o \times 2^o$) and explain the different results in Wang et al. (2018) and the current study.

Figure 14 shows AAOT *spatial* representation errors as estimated by the current study and by Wang's methodology as simulated with G5NR data. A global bias of 25%, not very different from the original 30% mentioned in Wang et al. (2018), is found for the Wang analysis whose representation errors yield a strongly skewed distribution over all sites. In contrast, the present study yields a more symmetric distribution with a much smaller bias. Unlike in the Wang analysis this bias is dominated by just a few sites with large positive representation errors.

The analysis above is a self-consistent evaluation of Wang's methodology. Using high-resolution black carbon emission data to downscale coarse model AAOT fields ignores redistribution of absorbing aerosol due to small scale (at and below the coarse model's grid-box) advective and turbulent transport as well as removal by local precipitation (Wang et al. were aware of this limitation but could not assess its impact). It also ignores local orography and the contribution of absorbing dust to AAOT. The result is that there is very little correlation between representation errors as estimated by the two methods, see Fig. 15. As a matter of fact, representation errors from the current study do not show a systematic dependence on emission distributions, unlike the representation errors from Wang's methodology.

6 A ranking of representativity for the AERONET sites

A ranking of AERONET and GAW sites in terms of their *spatial* representativity for AOT and AAOT can be found at Schutgens (2019). Only sites below 60^o latitude are considered, and temporal sampling of observations is ignored. The latter was done for two reasons: 1) as discussed in Sect 2 and 4, temporal sampling of observations is considered less accurately modelled by the OSSE than spatial variability; 2) both S17 and the current study show that once hourly collocation is used, the remaining representation error is similar although slightly larger than the spatial representation error.

Relative representation errors are classed according to bins: 0-5% (rank 1), 5-10% (rank 2), 10-20% (rank 3), 20-40% (rank 4), 40% and higher (rank 5). The accuracy of this ranking depends of course on the skill of G5NR and the OSSE, but also on statistical noise due to the use of a single year of data. The latter source of uncertainty was assessed using a block bootstrap method (Efron, 1979) on the time-series per site. Typically more than 85% of all resampled time-series yield a representation error in the same class as the original time-series. For large grid-boxes (4^o) and small errors ($< 10\%$), this may drop down to 66% of the resampled time-series. For those resampled time-series that yielded a different ranking, this ranking was only off by 1. It then seems that statistical noise does not prevent a robust classification of yearly relative spatial representation errors. The impact of G5NR and OSSE skill on the classification can currently not be assessed.

Compared to the subjective ranking by Kinne et al. (2013), the new ranking is objective because the rank is related to a well-defined representation error that is quantified bottom-up from known emission sources and calculated meteorology. That in itself is of course no guarantee for accuracy.

Inspection of the rankings turns up several interesting points. Analysis in the previous sections determined a few "rules" for the behaviour of representation errors (e.g. errors decrease when the grid-box size decreases) but these can easily be "broken" for specific sites: a smaller grid-box may actually lead to larger representation errors (e.g. AOE_Baotou, Ascension_Island, Aras_de_los_Olmos), monthly errors may be substantially larger than yearly errors (e.g. ARM-Darwin, BORDEAUX). Also, representation errors for AOT and AAOT may be very different: Bayfordbury shows small yearly representation errors for AOT but large errors for AAOT, while Mace_Head shows the opposite.

7 Conclusions

Remote sensing observations from the AERONET and GAW networks are intermittent in time and have a limited field-of-view. Consequently such observations have limited ability to represent (Absorbing) Aerosol Optical Thickness, or (A)AOT, over larger areas. The resulting spatio-temporal *representation* error is here analysed using a high-resolution simulation of global aerosol (GEOS5 Nature Run, ~ 7 km resolution near equator). Using G5NR, an Observing System Simulation Experiment (OSSE) was constructed that simulates the frequency of AERONET observations taking Solar Zenith Angle, cloud fraction and AOT values into account.

This work extends previous work on temporal representation with global low-resolution models (Schutgens et al., 2016b) to spatio-temporal representation. It also extends previous work on spatio-temporal representation with regional high-resolution simulations (Schutgens et al., 2016a, 2017) to the global domain. The current work is more limited in scope than the previous studies and only considers ground-based remote sensing observations. For satellite remote sensing, see Schutgens et al. (2016b) and Schutgens et al. (2017). For in-situ measurements, see Schutgens et al. (2016a) and Schutgens et al. (2017).

G5NR and the OSSE are evaluated and found to show significant skill in AOT and reasonable skill in AAOT. AERONET mean AOT per site, as well as yearly and daily variability were estimated quite correctly, usually within a factor less than $2\times$. Considering that G5NR generates its own meteorology, G5NR AOT correlated very well ($R \approx 0.75$) with the observations. Similarly, the OSSE was surprisingly good at simulating the overall pattern of observational coverage (frequency of AOT observation). Results were not as good for AAOT but still acceptable. Yearly AAOT variability was slightly underestimated while daily AAOT variability was severely underestimated. The latter is possibly related to the absence of diurnal anthropogenic emission profiles in G5NR. For representativity studies that take diurnal variations into account, see Schutgens et al. (2016a, 2017). In addition, the OSSE tended to overestimate the frequency of AAOT observations per site (although this was shown to have no impact on representation errors).

Both yearly and monthly representation errors are provided for observations from ground sites that attempt to represent larger areas (from 0.5^o to 4^o in size). The monthly representation errors are shown to be strongly correlated throughout the year. For some sites this is an expression of a bias but that is not universally the case. In any case, monthly representation errors can not be treated as independent and this has (negative) consequences for the reduction of representation errors in multi-year averages. Other conclusions are: 1) AERONET derived climatologies allow for substantial representation errors (yearly collocation allows errors of typically 20%, see Fig. 4); 2) AEROCOM evaluation protocol is sub-optimal (daily collocation can show errors of 25% in coherent regional patterns). Instead hourly collocation is advocated~~. Also, ~~, 3) the representativity of AERONET and GAW sites was shown to be not very different, although AERONET sites seem to be more affected by nearby sources while GAW sites seem more affected by their altitude. Finally, a subjective ranking (Kinne et al., 2013) of the *spatial* representativity of sites was analysed and shown to broadly agree with the current study, although it appears to overestimate represented spatial domain sizes and judges several sites as less representative than the current analysis. A new objective ranking is also presented.

While the current study's focus is on strategies for model evaluation with original ('All Points') AERONET data, it does allow recommendations to be made for the optimal aggregation level of observational data. Hourly products are preferred to daily or monthly products as they allow users to perform hourly collocation which in turn yields significantly smaller representation errors. This should hold for both satellite and AERONET data.

Spatial representation errors have been used to reconcile observations and global simulations of AAOT. Bond et al. (2013) showed that global models tend to significantly underestimate AAOT but Wang et al. (2018) suggested that AERONET AAOT observations may suffer from a global 30% representation bias. In contrast, the current analysis finds a much smaller bias of 9% which is more-over strongly influenced by a few sites with large positive representation errors due to their proximity to black carbon sources. Judiciously excluding those sites significantly reduces the bias even further (4%). The large positive representation errors found by Wang et al. are shown to be due to methodological choices that limit the realism of their OSSE.

Several questions remain and seem interesting for follow-up studies: 1) how can we evaluate the representativity rankings~~?~~; 2) how do OSSE errors affect estimated representation errors~~?~~; 3) how will diurnal emission profiles impact results~~?~~;

4) can representation errors at any site be decomposed in a bias and random error (possibly with temporal correlations over several months)?; 5) what are representation errors like in multi-year averages?

Code and data availability. G5NR data can be obtained from `https://gmao.gsfc.nasa.gov/global_mesoscale/7km-G5NR/data_access`, AERONET data can be obtained from `https://aeronet.gsfc.nasa.gov`. Analysis code was written in IDL and is available from the author upon request.

Author contributions. NS designed the experiments, carried them out and prepared the manuscript.

Competing interests. No competing interests are present

Acknowledgements. NS thanks the NASA Global Modelling and Assimilation Office team that conducted the GEOS-5 Nature Run simulation, in particular Arlindo da Silva and Ravi Govindaraju for help in obtaining the G5NR data. NS thanks the PI(s) and Co-I(s) and their staff for establishing and maintaining the many AERONET sites used in this investigation. NS thanks Ann Mari Fjaeraa (NILU) for providing GAW-AOT and GAW-ABS geolocation data. NS is also grateful to Rong Wang, Björn H. Samset and Gunnar Myhre for valuable discussions. Three anonymous reviewers, Andrew Sayer and Tero Mielonen provided very useful commentary and NS gratefully acknowledges their contributions. The figures in this paper were prepared using David W. Fanning's Coyote Library for IDL. This work is part of the Vici research programme with project number 016.160.324, which is (partly) financed by the Dutch Research Council (NWO).

References

Anderson, T. E., Charlson, R. J., Winker, D. M., Ogren, J. A., and Holmen, K.: Mesoscale Variations of Tropospheric Aerosols, J. Atmospheric Sciences, 60, 119–136, 2003.

Boersma, K. F., Vinken, G. C. M., and Eskes, H. J.: Representativeness errors in comparing chemistry transport and chemistry climate models with satellite UV-Vis tropospheric column retrievals, Geoscientific Model Development, 9, 875–898, https://doi.org/10.5194/gmd-9-875-2016, 2016.

Bond, T. C., Doherty, S. J., Fahey, D. W., Forster, P. M., Berntsen, T., Deangelo, B. J., Flanner, M. G., Ghan, S., Kärcher, B., Koch, D., Kinne, S., Kondo, Y., Quinn, P. K., Sarofim, M. C., Schultz, M. G., Schulz, M., Venkataraman, C., Zhang, H., Zhang, S., Bellouin, N., Guttikunda, S. K., Hopke, P. K., Jacobson, M. Z., Kaiser, J. W., Klimont, Z., Lohmann, U., Schwarz, J. P., Shindell, D., Storelvmo, T., Warren, S. G., and Zender, C. S.: Bounding the role of black carbon in the climate system: A scientific assessment, Journal of Geophysical Research Atmospheres, 118, 5380–5552, https://doi.org/10.1002/jgrd.50171, 2013.

Bulgin, C. E., Embury, O., and Merchant, C. J.: Sampling uncertainty in gridded sea surface temperature products and Advanced Very High Resolution Radiometer (AVHRR) Global Area Coverage (GAC) data, Remote Sensing of Environment, 177, 287–294, https://doi.org/10.1016/j.rse.2016.02.021, http://dx.doi.org/10.1016/j.rse.2016.02.021, 2016.

Castellanos, P., da Silva, A. M., Darmenov, A. S., Buchard, V., Govindaraju, R. C., Ciren, P., and Kondragunta, S.: A Geostationary Instrument Simulator for Aerosol Observing System Simulation Experiments, Atmosphere, 10, https://doi.org/10.3390/atmos10010002, 2019.

Cavanaugh, N. R. and Shen, S. S. P.: The effects of gridding algorithms on the statistical moments and their trends of daily surface air temperature, Journal of Climate, 28, 9188–9205, https://doi.org/10.1175/JCLI-D-14-00668.1, 2015.

Chin, M., GInoux, P., KInne, S., Torres, O., Holben, B. N., Duncan, B. N., Martin, R. V., Logan, J. A., Higurashi, A., and Nakajima, T.: Tropospheric Aerosol Optical Thickness from the GOCART Model and Comparisons with Satellite and Sun Photometer Measurements, Journal of the Atmospheric Sciences, 59, 461–483, 2002.

Colarco, P., Silva, A., Chin, M., and Diehl, T.: Online simulations of global aerosol distributions in the NASA GEOS - 4 model and comparisons to satellite and ground - based aerosol optical depth, Journal of Geophysical Research : Atmospheres, 115, https://doi.org/10.1029/2009JD012820, 2010.

Colarco, P. R., Kahn, R. A., Remer, L. A., and Levy, R. C.: Impact of satellite viewing-swath width on global and regional aerosol optical thickness statistics and trends, Atmospheric Measurement Techniques, 7, 2313–2335, https://doi.org/10.5194/amt-7-2313-2014, 2014.

Coldewey-Egbers, M., Loyola, D. G., Koukouli, M., Balis, D., Lambert, J. C., Verhoelst, T., Granville, J., Van Roozendael, M., Lerot, C., Spurr, R., Frith, S. M., and Zehner, C.: The GOME-type Total Ozone Essential Climate Variable (GTO-ECV) data record from the ESA Climate Change Initiative, Atmospheric Measurement Techniques, 8, 3923–3940, https://doi.org/10.5194/amt-8-3923-2015, 2015.

da Silva, A., Colarco, P. Darmenov, A., Buchard-Marchant, V., Randles, C., and Govinaradju, R.: Overview of the MERRA Aerosol Reanalysis: Toward an Integrated Earth System Analysis, in: 4th WCRP International Conference on Reanalyses, 2012.

Diedrich, H., Wittchen, F., Preusker, R., and Fischer, J.: Representativeness of total column water vapour retrievals from instruments on polar orbiting satellites, Atmospheric Chemistry and Physics, 16, 8331–8339, https://doi.org/10.5194/acp-16-8331-2016, 2016.

Director, H. and Bornn, L.: Connecting point-level and gridded moments in the analysis of climate data, Journal of Climate, 28, 3496–3510, https://doi.org/10.1175/JCLI-D-14-00571.1, 2015.

Dubovik, O. and King, M. D.: A flexible inversion algorithm for retrieval of aerosol optical properties from Sun and sky radiance measurements, Journal of Geophysical Research, 105, 20 673, https://doi.org/10.1029/2000JD900282, http://doi.wiley.com/10.1029/2000JD900282, 2000.

Dubovik, O., Smirnov, A., Holben, B. N., King, M. D., Kaufman, Y. J., Eck, T. F., and Slutsker, I.: Accuracy assessments of aerosol optical properties retrieved from Aerosol Robotic Network (AERONET) Sun and sky radiance measurements, Journal of Geophysical Research, 105, 9791–9806, https://doi.org/10.1029/2000JD900040, http://doi.wiley.com/10.1029/2000JD900040, 2000.

Eck, T. F., Holben, B. N., Reid, J. S., Smirnov, A., O'Neill, N. T., Slutsker, I., and Kinne, S.: Wavelength dependence of the optical depth of biomass burning, urban, and desert dust aerosols, J. Geophysical Research, 104, 31 333–31 349, 1999.

Efron, B.: Bootstrap methods: another look at the jackknife, The annals of Statistics, 7, 1—-26, 1979.

Gelaro, R., Putman, W. M., Pawson, S., Draper, C., Molod, A., Norris, P. M., Ott, L., Privé, N., Reale, O., Achuthavarier, D., Bosilovich, M., Buchard, V., Chao, W., Coy, L., Cullather, R., Silva, A., Darmenov, A., and Errico, R. M.: Evaluation of the 7-km GEOS-5 Nature Run, Tech. Rep. NASA / TM – 2014-104606, NASA, 2015.

Geogdzhayev, I., Cairns, B., Mishchenko, M. I., Tsigaridis, K., and van Noije, T.: Model-based estimation of sampling-caused uncertainty in aerosol remote sensing for climate research applications, Quarterly Journal of the Royal Meteorological Society, 140, 2353–2363, https://doi.org/10.1002/qj.2305, http://doi.wiley.com/10.1002/qj.2305, 2014.

Giles, D. M., Sinyuk, A., Sorokin, M. G., Schafer, J. S., Smirnov, A., Slutsker, I., Eck, T. F., Holben, B. N., Lewis, J. R., Campbell, J. R., Welton, E. J., Korkin, S. V., and Lyapustin, A. I.: Advancements in the Aerosol Robotic Network (AERONET) Version 3 database – automated near-real-time quality control algorithm with improved cloud screening for Sun photometer aerosol optical depth (AOD) measurements, Atmospheric Measurement Techniques, 12, 169–209, 2019.

Hakuba, M. Z., Folini, D., Sanchez-lorenzo, A., and Wild, M.: Spatial representativeness of ground-based solar radiation measurements - extension to the full Meteosat disk, Journal of Geophysical Research: Atmospheres, 16, 50 673, https://doi.org/10.1002/jgrd.50673., 2014a.

Hakuba, M. Z., Folini, D., and Wild, M.: Solar absorption over Europe from collocated surface and satellite observations, Journal of Geophysical Research: Atmospheres, pp. 3420–3437, https://doi.org/10.1002/2013JD021421, 2014b.

Holben, B. N., Eck, T. F., Slutsker, I., Tanre, D., Buis, J. P., Setzer, A., Vermote, E., Reagan, J. A., Kaufman, Y. J., Nakajima, T., Lavenu, F., Jankowiak, I., and Smirnov, A.: AERONET-A Federated Instrument Network and Data Archive for Aerosol Characterization, Remote Sensing of Environment, 66, 1–16, 1998.

Holben, B. N., Eck, T. F., Slutsker, I., Smirnov, A., Sinyuk, A., Schafer, J., Giles, D., and Dubovik, O.: Aeronet's Version 2.0 quality assurance criteria, in: Remote Sensing of the Atmosphere and Clouds, edited by Tsay, S.-C., Nakajima, T., Singh, R. P., and Sridharan, R., vol. 6408, pp. 134 – 147, International Society for Optics and Photonics, SPIE, https://doi.org/10.1117/12.706524, https://doi.org/10.1117/12.706524, 2006.

Holben, B. N., Kim, J., Sano, I., Mukai, S., Eck, T. F., Giles, D. M., Schafer, J. S., Sinyuk, A., Slutsker, I., Smirnov, A., Sorokin, M., Anderson, B. E., Che, H., Choi, M., Crawford, J. H., Ferrare, R. A., Garay, M. J., Jeong, U., Kim, M., Kim, W., Knox, N., Li, Z., Lim, H. S., Liu, Y., Maring, H., Nakata, M., Pickering, K. E., Piketh, S., Redemann, J., Reid, J. S., Salinas, S., Seo, S., Tan, F., Tripathi, S. N., Toon, O. B., and Xiao, Q.: An overview of mesoscale aerosol processes , comparisons , and validation studies from DRAGON networks, Atmospheric Chemistry and Physics, 18, 18 655– 18 671, 2018.

Isobe, T., Feigelson, E. D., Akritas, M. G., and Babu, G. J.: Linear regression in Astronomy I, The Astrophysical Journal, 364, 104–113, 1990.

Janssens-maenhout, G., Dentener, F., Aardenne, J. V., Monni, S., Pagliari, V., Orlandini, L., Klimont, Z., Kurokawa, J.-i., Akimoto, H., Ohara, T., Wankmüller, R., Battye, B., Grano, D., Zuber, A., and Keating, T.: EDGAR-HTAP : a harmonized gridded air pollution emission dataset based on national inventories, Tech. rep., JRC, Institute for Environment and Sustainability, https://doi.org/10.2788/14102, 2012.

Kaufman, Y. J., Holben, B. N., Tanre, D., Slutsker, I., Smimov, A., and Eck, T. F.: Will aerosol measurements from Terra and Aqua polar orbiting satellites represent the daily aerosol abundance and properties?, Geophysical Research Letters, 27, 3861–3864, 2000.

Kinne, S., O'Donnel, D., Stier, P., Kloster, S., Zhang, K., Schmidt, H., Rast, S., Giorgetta, M., Eck, T. F., and Stevens, B.: MAC-v1: A new global aerosol climatology for climate studies, Journal of Advances in Modeling Earth Systems, 5, 704–740, https://doi.org/10.1002/jame.20035, http://doi.wiley.com/10.1002/jame.20035, 2013.

Kovacs, T.: Comparing MODIS and AERONET aerosol optical depth at varying separation distances to assess ground-based validation strategies for spaceborne lidar, Journal of Geophysical Research, 111, D24 203, https://doi.org/10.1029/2006JD007349, http://www.agu.org/pubs/crossref/2006/2006JD007349.shtml, 2006.

Levy, R. C., Leptoukh, G. G., Kahn, R., Zubko, V., Gopalan, A., and Remer, L. a.: A critical look at deriving monthly aerosol optical depth from satellite data, IEEE Transactions on Geoscience and Remote Sensing, 47, 2942–2956, https://doi.org/10.1109/TGRS.2009.2013842, 2009.

Lin, M., Horowitz, W., Cooper, O. R., Tarasick, D., Conley, S., Iraci, L. T., Johnson, B., Leblanc, T., Petropavlovskikh, I., and Yates, E. L.: Revisiting the evidence of increasing springtime ozone mixing ratios in the free troposphere over Western North America, Geophysical Research Letters, 42, https://doi.org/10.1002/2015GL065311.Received, 2015.

Ma, P.-l., Rasch, P. J., Chepfer, H., Winker, D. M., and Ghan, S. J.: Observational constraint on cloud susceptibility weakened by aerosol retrieval limitations, Nature Communications, 9, https://doi.org/10.1038/s41467-018-05028-4, 2018.

Nappo, C., Caneill, J., Furman, R., Gifford, F., Kaimal, J., Kramer, M., Lockhart, T., Pendergast, M., RA, P., Randerson, D., Shreffler, J., and Wyngaard, J.: The workshop on the representativeness of meteolorogical observations, June 1981, Boulder, Colo, Bulletin of the American Meteorological Society, 63, 761–764, 1982.

Pitkänen, M. R. A., Mikkonen, S., Lehtinen, K. E. J., Lipponen, A., and Arola, A.: Artificial bias typically neglected in comparisons of uncertain atmospheric data, Geophysical Research Letters, 43, 10,003–10,011, https://doi.org/10.1002/2016GL070852, http://doi.wiley.com/10.1002/2016GL070852, 2016.

Putman, W., da Silva, A., Ott, L., and Darmenov, A.: Global Modeling and Assimilation Office Model Configuration for the 7-km GEOS-5 Nature Run , Ganymed Release (non-hydrostatic 7km Global Mesoscale Simulation), Tech. Rep. GMAO Office Note No 5, NASA, http://gmao.gsfc.nasa.gov/pubs/office{_}notes, 2014.

Remer, L., Kaufman, Y., and Kleidman, R.: Comparison of Three Years of Terra and Aqua MODIS Aerosol Optical Thickness Over the Global Oceans, IEEE Geoscience and Remote Sensing Letters, 3, 537–540, https://doi.org/10.1109/LGRS.2006.879562, http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1715312, 2006.

Santese, M., De Tomasi, F., and Perrone, M. R.: AERONET versus MODIS aerosol parameters at different spatial resolutions over southeast Italy, Journal of Geophysical Research, 112, D10 214, https://doi.org/10.1029/2006JD007742, http://www.agu.org/pubs/crossref/2007/2006JD007742.shtml, 2007.

Sayer, A. M., Thomas, G. E., Palmer, P. I., and Grainger, R. G.: Some implications of sampling choices on comparisons between satellite and model aerosol optical depth fields, Atmospheric Chemistry and Physics, 10, 10 705–10 716, https://doi.org/10.5194/acp-10-10705-2010, http://www.atmos-chem-phys.net/10/10705/2010/, 2010.

Schafer, J. S., Eck, T. F., Holben, B. N., Thornhill, K. L., Anderson, B. E., Sinyuk, A., Giles, D. M., Winstead, E. L., Ziemba, L. D., Beyersdorf, A. J., Kenny, P. R., Smirnov, A., and Slutsker, I.: Intercomparison of aerosol single-scattering albedo derived from AERONET surface radiometers and LARGE in situ aircraft profiles during the 2011 DRAGON-MD and DISCOVER-AQ experiments, Journal of Geophysical Research: Atmospheres, 119, 7439–7452, https://doi.org/10.1002/2013JD021166.Received, 2014.

Schmid, B., Michalsky, J., Halthore, R., Beauharnois, M., Harnson, L., Livingston, J., Russell, P., Holben, B., Eck, T., and Smirnov, A.: Comparison of Aerosol Optical Depth from Four Solar Radiometers During the Fall 1997 ARM Intensive Observation Period, Geophysical Research Letters, 26, 2725–2728, 1999.

Schutgens, N.: Representativeness of AERONET and GAW aerosol observation sites, https://doi.org/10411/XDZD4A, https://hdl.handle.net/10411/XDZD4A, 2019.

Schutgens, N., Gryspeerdt, E., Weigum, N., Tsyro, S., Goto, D., Schulz, M., and Stier, P.: Will a perfect model agree with perfect observations? The impact of spatial sampling, Atmospheric Chemistry and Physics Discussions, 16, https://doi.org/10.5194/acp-2015-973, http://www.atmos-chem-phys-discuss.net/acp-2015-973/, 2016a.

Schutgens, N., Partridge, D. G., and Stier, P.: The importance of temporal collocation for the evaluation of aerosol models with observations, Atmospheric Chemistry and Physics, 16, 1065–1079, https://doi.org/10.5194/acp-16-1065-2016, 2016b.

Schutgens, N., Tsyro, S., Gryspeerdt, E., Goto, D., Weigum, N., Schulz, M., and Stier, P.: On the spatio-temporal representativeness of observations, Atmospheric Chemistry and Physics, 17, 9761–9780, https://doi.org/10.5194/acp-2017-149, https://www.atmos-chem-phys-discuss.net/acp-2017-149/, 2017.

Schutgens, N. J., Nakata, M., and Nakajima, T.: Validation and empirical correction of MODIS AOT and AE over ocean, Atmospheric Measurement Techniques, 6, 2455–2475, https://doi.org/10.5194/amt-6-2455-2013, http://www.atmos-meas-tech.net/6/2455/2013/, 2013.

Schwarz, M., Folini, D., Hakuba, M. Z., and Wild, M.: Spatial representativeness of surface-measured variations of downward solar radiation, Journal of Geophysical Research: Atmospheres, 122, 13 319–13 337, https://doi.org/10.1002/2017JD027261, http://doi.wiley.com/10.1002/2017JD027261, 2017.

Schwarz, M., Follini, D., Hakuba, M., and Wild, M.: From Point to Area : Worldwide Assessment of the Representativeness of Monthly Surface Solar Radiation Records, Journal of Geophysical Research : Atmospheres, 123, 13 857–13 874, https://doi.org/10.1029/2018JD029169, 2018.

Shi, X., Zhao, C., Jiang, J. H., Wang, C., Yang, X., and Yung, Y. L.: Spatial Representativeness of PM 2.5 Concentrations Obtained Using Reduced Number of Network Stations, Journal of Geophysical Research: Atmospheres, 123, 3145–3158, https://doi.org/10.1002/2017JD027913, http://doi.wiley.com/10.1002/2017JD027913, 2018.

Shinozuka, Y. and Redemann, J.: Horizontal variability of aerosol optical depth observed during the ARCTAS airborne experiment, Atmospheric Chemistry and Physics, 11, 8489–8495, https://doi.org/10.5194/acp-11-8489-2011, 2011.

Smirnov, A.: Diurnal variability of aerosol optical depth observed at AERONET (Aerosol Robotic Network) sites, Geophysical Research Letters, 29, 2115, https://doi.org/10.1029/2002GL016305, http://doi.wiley.com/10.1029/2002GL016305, 2002.

Smirnov, A., Holben, B. N., Eck, T. F., Dubovik, O., and Slutsker, I.: Cloud-Screening and Quality Control Algorithms for the AERONET Database, Remote Sensing of Environment, 73, 337–349, 2000.

Sofieva, V. F., Kalakoski, N., Päivärinta, S. M., Tamminen, J., Kyrölä, E., Laine, M., and Froidevaux, L.: On sampling uncertainty of satellite ozone profile measurements, Atmospheric Measurement Techniques, 7, 1891–1900, https://doi.org/10.5194/amt-7-1891-2014, 2014.

Suarez, M. J., Darmenov, A. S., and da Silva, A.: The Quick Fire Emissions Dataset (QFED) - Documentation of versions 2 . 1 , 2 . 2 and 2 . 4, Tech. rep., NASA, 2013.

Virtanen, T. H., Kolmonen, P., Sogacheva, L., Rodríguez, E., Saponaro, G., and Leeuw, G. D.: Collocation mismatch uncertainties in satellite aerosol retrieval validation, Atmospheric Measurement Techniques, 11, 925–938, 2018.

Wang, R., Balkanski, Y., Boucher, O., Ciais, P., Schuster, G. L., Chevallier, F., Samset, B. H., Liu, J., Piao, S., Valari, M., and Tao, S.: Estimation of global black carbon direct radiative forcing and its uncertainty constrained by observations, Journal of Geophysical Research, 121, 5948–5971, https://doi.org/10.1002/2015JD024326, 2016.

Wang, R., Andrews, E., Balkanski, Y., Boucher, O., Myhre, G., Samset, B., Schulz, M., Schuster, G. L., Valari, M., and Tao, S.: Geophysical Research Letters, Geophysical Research Letters, 45, 2106–2114, https://doi.org/10.1002/2017GL076817, 2018.

Weigum, N. M., Stier, P., Schwarz, J. P., Fahey, D. W., and Spackman, J. R.: Scales of variability of black carbon plumes over the Pacific Ocean, Geophysical Research Letters, 39, https://doi.org/10.1029/2012GL052127, http://doi.wiley.com/10.1029/2012GL052127, 2012.

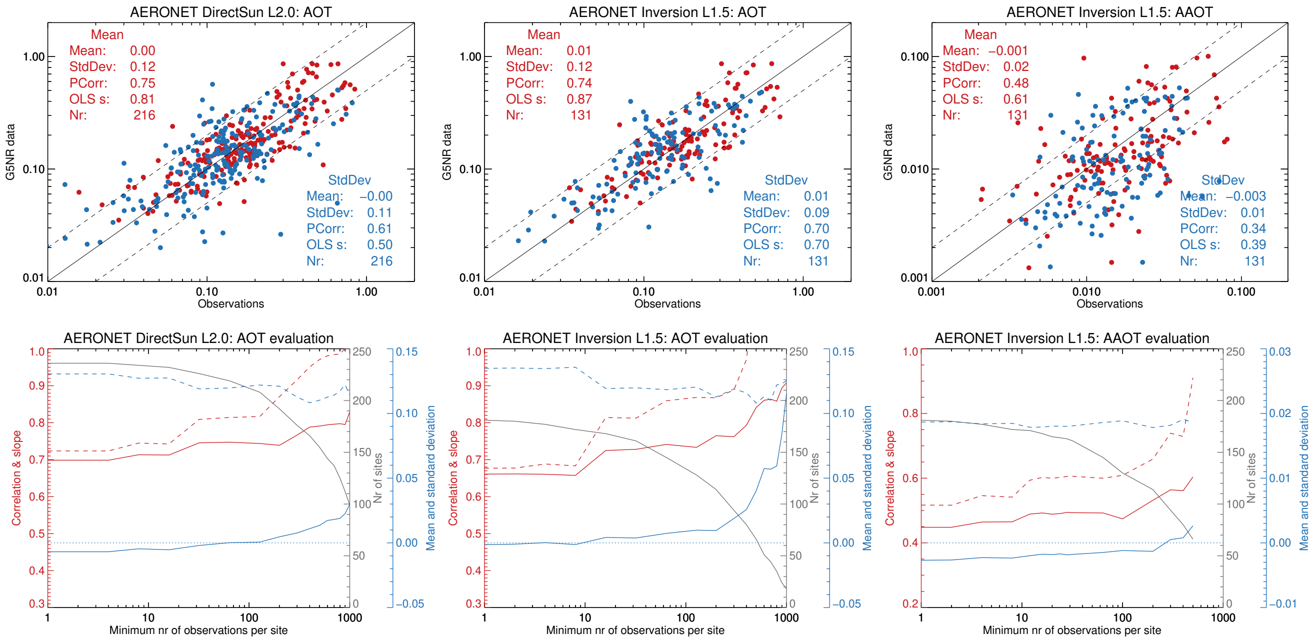

Figure 1. Evaluation of the G5NR simulation of AOT and AAOT with AERONET data. The top row shows evaluation against three different datasets. Each dot represents the yearly mean or standard deviation for a single AERONET site (with at least 100 observations in 2006); the mean value is shown in red and the standard deviation in blue. The coloured text summarizes the statistics over all data points in the figure. In the bottom row, the impact of the minimum required number of observations per site on those summary statistics (for means) is shown. Colours relate lines to axes and have different meaning than in the top row. Red solid is correlation, red dashed is slope, blue solid is mean, and blue dashed is standard deviation. In all figures, hourly G5NR model data was collocated in time & space with AERONET observations before calculating site statistics.

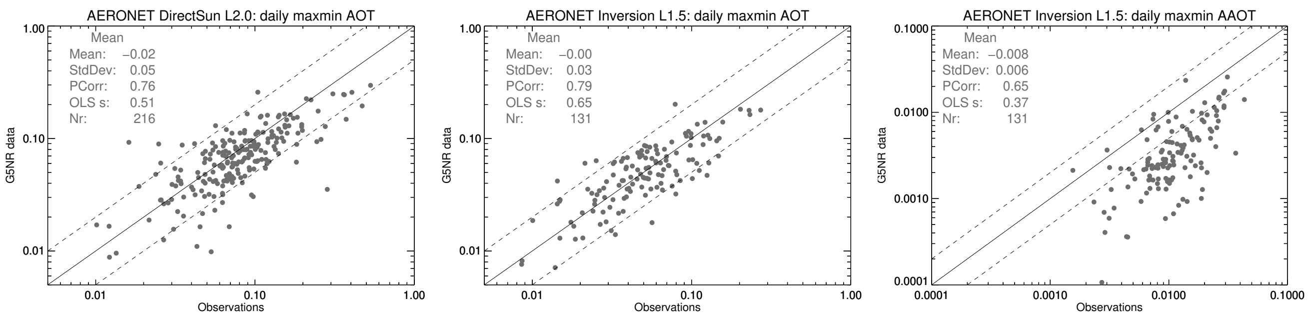

Figure 2. Evaluation of the G5NR simulation of AOT and AAOT with AERONET data. Each dot represents the yearly average of daily variation (maximum minus minimum value) for a single AERONET site (with at least 100 observations in 2006). The grey text summarizes the statistics over all data points in the figure. In all figures, hourly G5NR model data was collocated in time & space with AERONET observations before calculating site statistics.

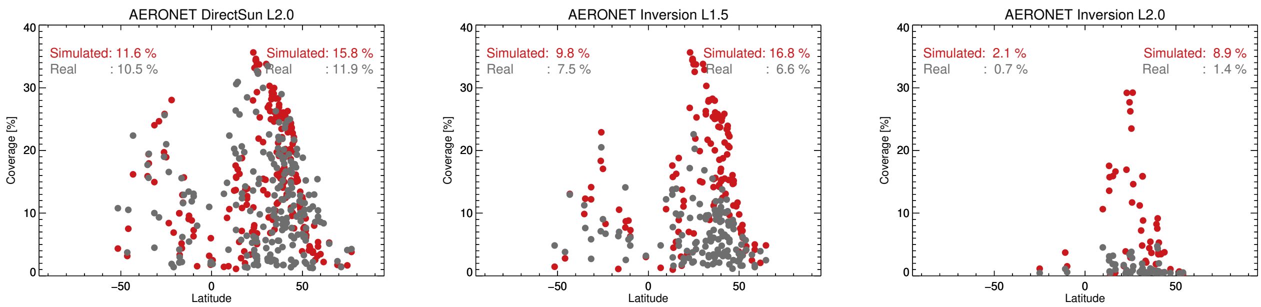

Figure 3. Evaluation of the temporal coverage predicted by the OSSE with AERONET observations. Each dot represents temporal coverage (or frequency of observation) for a single AERONET site (with at least 100 observations in 2006, at least 30 observations for Inversion L2.0). The grey dots are real AERONET data, the red dots are simulated by the methodology described in Sec. 3. The numbers in the graph are temporal coverages estimated by hemisphere.

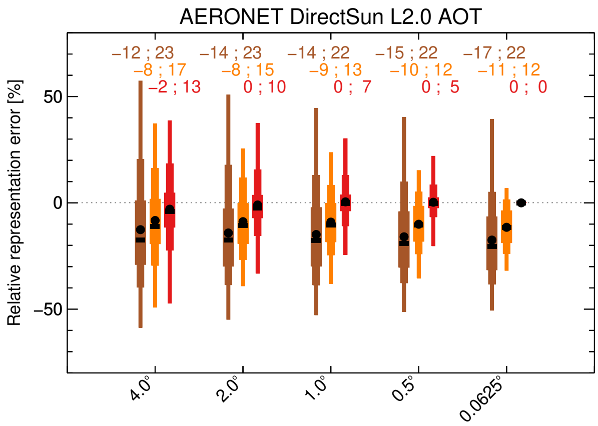

Figure 4. Yearly representation errors for AOT from DirectSun L2.0 AERONET for different model grid-box sizes. The colours indicate different collocation protocols: yearly (brown), daily (orange) and hourly (red). Numbers on top are mean of the errors and mean of the sign-less errors.

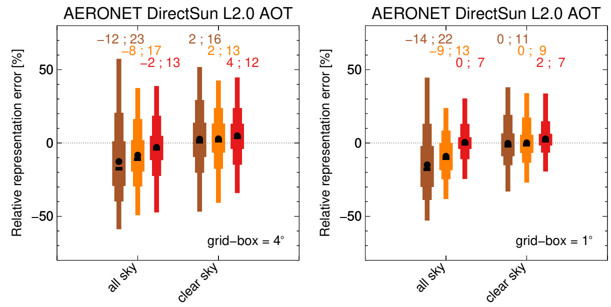

Figure 6. Yearly representation errors for AOT from from DirectSun L2.0 AERONET using all-sky or clear sky conditions and model grid-box size of 4^o (left) or 1^o (right). The colours indicate different collocation protocols: yearly (brown), daily (orange) and hourly (red). Numbers on top are mean of the errors and mean of the sign-less errors.

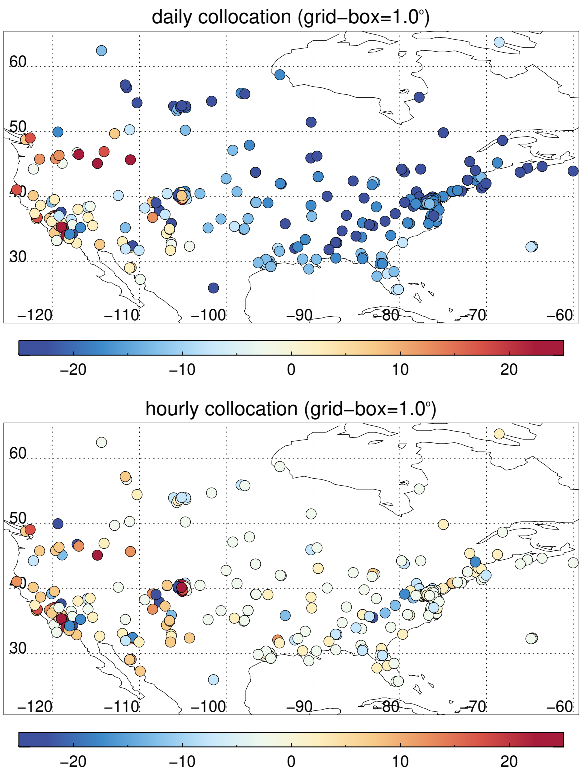

Figure 5. Yearly representation errors [%] for AOT from DirectSun L2.0 AERONET in Northern America, for two different collocation protocols (top: daily; bottom: hourly) and a model grid-box size of 1^o.

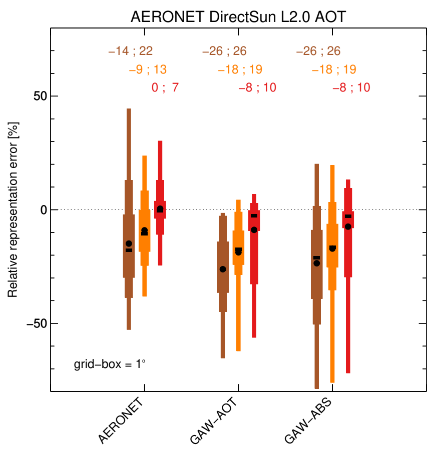

Figure 7. Yearly representation errors for AOT from Direct Sun L2.0 AERONET and GAW and a model grid-box size of 1^o. The colours indicate different collocation protocols: yearly (brown), daily (orange) and hourly (red). Numbers on top are mean of the errors and mean of the sign-less errors.

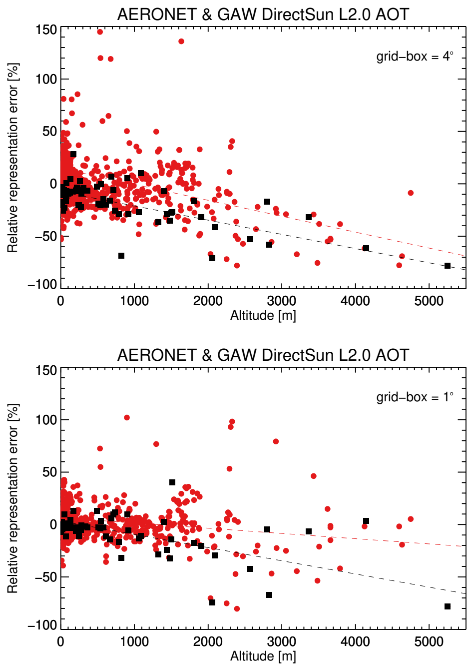

Figure 8. Yearly representation errors for AOT from Direct Sun L2.0 AERONET (red circles) and GAW (black squares) as a function of site altitude, for a model grid-box size of either 4^o or 1^o; using hourly collocation.

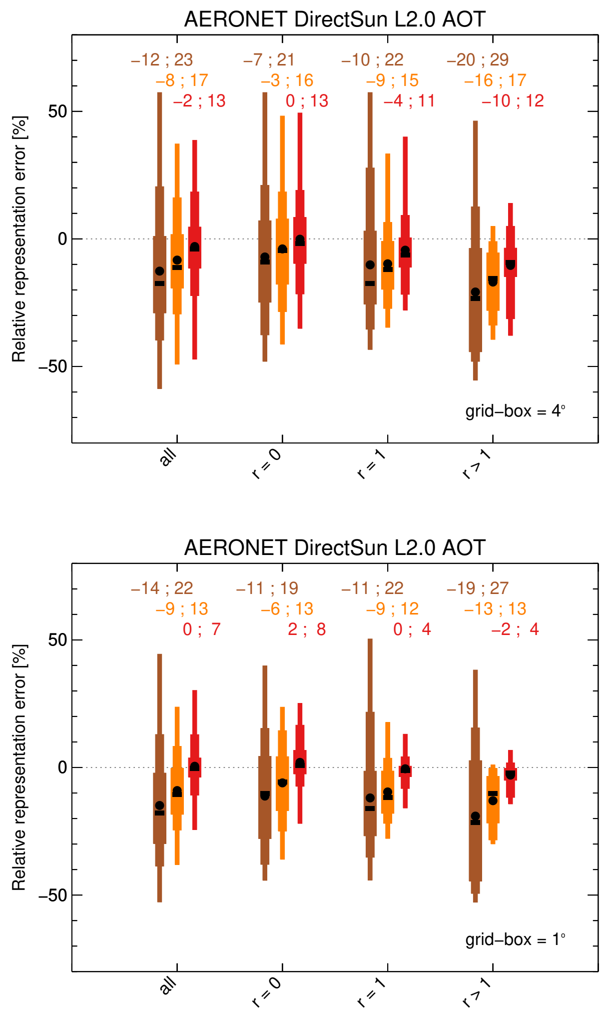

Figure 9. Yearly representation errors for AOT from DirectSun L2.0 AERONET for different range scores r by Kinne et al. (2013), for a model grid-box size of either 4^o or 1^o. The colours indicate different collocation protocols: yearly (brown), daily (orange) and hourly (red). Numbers on top are mean of the errors and mean of the sign-less errors.

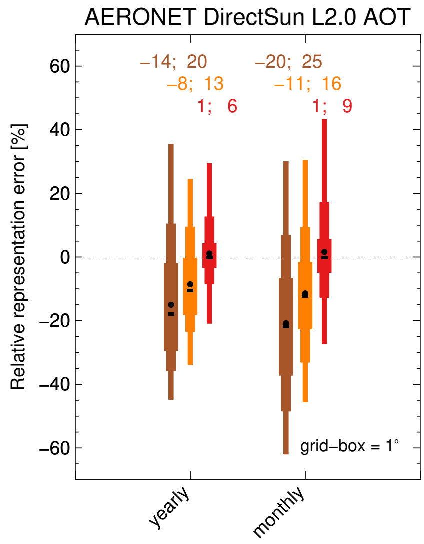

Figure 10. Yearly and monthly representation errors for AOT DirectSun L2.0 AERONET, for a model grid-box size of 1^o. In contrast to Fig. 4, only sites that provide observations 12 months out of the year are used in this analysis. The colours indicate different collocation protocols: yearly (brown), daily (orange) and hourly (red). Obviously, for the brown bar on the right a monthly protocol was used. Numbers on top are mean of the errors and mean of the sign-less errors.

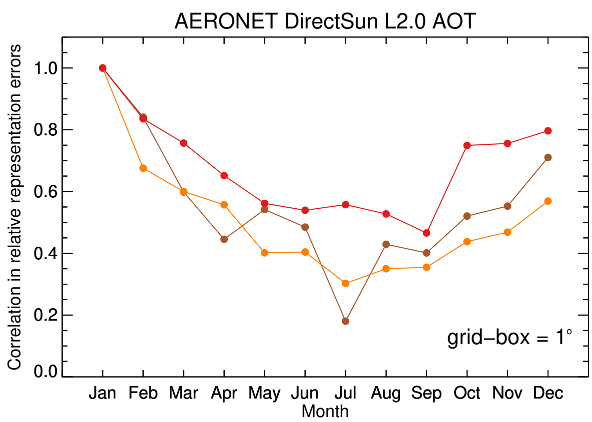

Figure 11. Correlation in monthly representation errors with errors for January, for AOT DirectSun L2.0 AERONET, for a model grid-box size of 1^o. Only sites that provide observations 12 months out of the year are used in this analysis. The colours indicate different collocation protocols: ~~yearly~~ monthly (brown), daily (orange) and hourly (red).

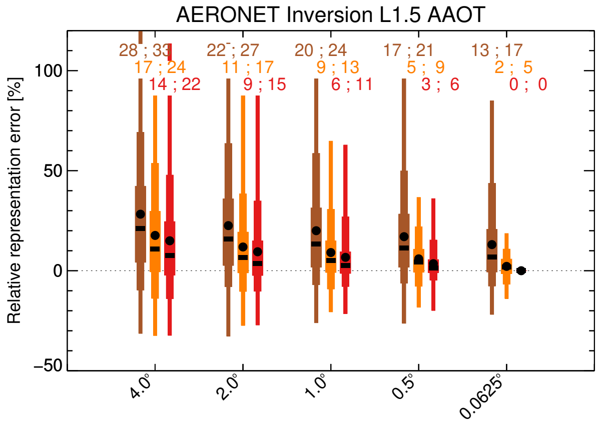

Figure 12. Yearly representation errors for AAOT from Inversion L1.5 AERONET for different model grid-box sizes. The colours indicate different collocation protocols: yearly (brown), daily (orange) and hourly (red). Numbers on top are mean of the errors and mean of the sign-less errors.

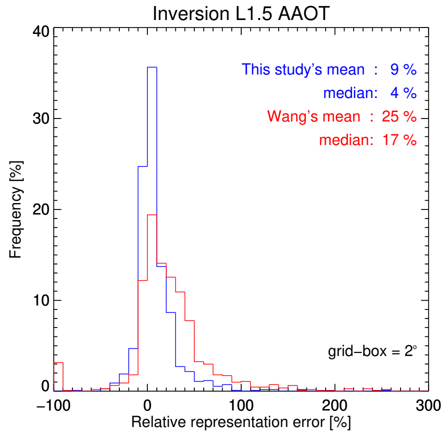

Figure 14. Yearly representation errors for AAOT from Inversion L1.5 AERONET as estimated in this paper or using the methodology from Wang et al. (2018) and a model grid-box size of 2^o. The representation error shown is the spatial representation error (Schutgens et al., 2017), i.e. temporal sampling of observations is ignored.

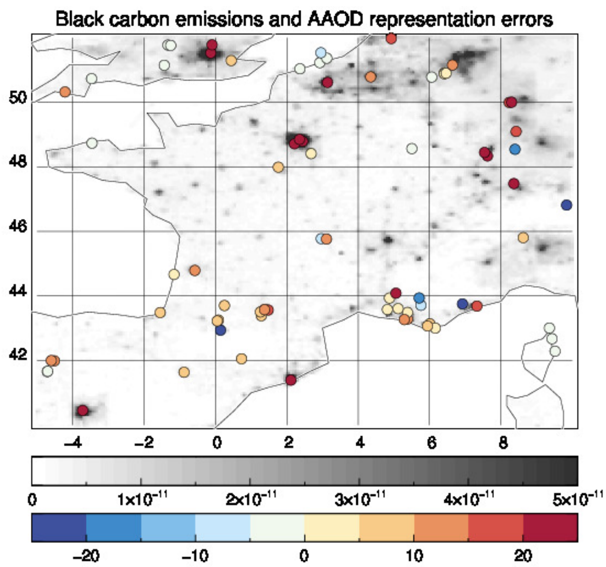

Figure 13. Black carbon emissions over France, Europe, with the representation errors in AAOT from Inversion L1.5 AERONET super-imposed. The top colourbar (white-black) represents emissions ([kg/m$^2 s$]), and the bottom colourbar (blue-red) represents relative representation errors ([%]). Only *spatial* representation errors are shown, i.e. the temporal sampling of observations is ignored.

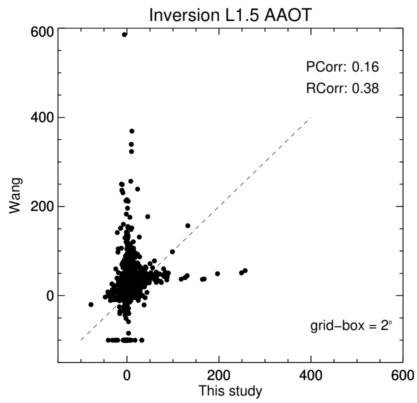

Figure 15. Comparison of yearly representation errors for AAOT from Inversion L1.5 AERONET as estimated in this paper or using the methodology from Wang et al. (2018) and a model grid-box size of 2^o. The representation error shown is the spatial representation error (Schutgens et al., 2017), i.e. temporal sampling of observations is ignored. Also shown are the Pearson linear correlation (PCorr) and rank correlation (RCorr) between the data. The dashed line shows $y = x$.