



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

Nick A.J. Schutgens<sup>1</sup>

<sup>1</sup>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  $0.5^\circ$  to  $4^\circ$  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, with a pronounced annual cycle. 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 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).

5 As aerosol 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. (2016a) (S16a hereafter) showed that representation errors due to temporal sampling in both satellite and AERONET observations were of similar magnitude as actual model errors and often larger than observational errors. Similarly, Schutgens et al. (2016b) (S16b hereafter) showed that the narrow field-of-view of in-  
10 situ measurements could lead to large differences from area averages (monthly RMS differences of 10 – 80% for  $201 \times 210$  km<sup>2</sup>, 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 error estimates and optimal strategies when dealing with different observing systems (ground networks, polar orbiting satellites with varying return times, or geo-stationary satellites). S16b In this paper,  
15 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. In S16b 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  
20 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  
25 use. If observations are used to evaluate models, different protocols (or strategies) exist to temporally collocate model data and observations. For instance, within AEROCOM, 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  
30 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 (S16a). 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  
35 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  $0.0625^\circ$  grid-resolution ( $\sim 7$  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).

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 (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%. (Note that Castellanos et al. (2019) derived global rescaling factors for aerosol speciated AOD in G5NR. How such scaling factors will affect AAOD is unknown. True scaling factors are unlikely to be global, and representation errors in this paper are relative anyway. In this paper the original, i.e. not rescaled, model data will be used). 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. Clouds in G5NR show 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 data were obtained from <https://aeronet.gsfc.nasa.gov>. For 2006, AOT from Direct Sun Version 3 L2.0 and AOT & AAOT from Inversion Version 2 L1.5 and L2.0 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).

## 5 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 spectrometers that measure AOT; and the GAW-ABS network which comprises 81 surface-based filter instruments. While GAW-ABS is not capable to provide (A)AOT measurements, here we will assume a potential remotely sensed columnar product, similar to AERONET (A)AOT, and consider its representation errors.

## 10 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). Here, a self-consistent high resolution simulation will be used to generate both observation and truth (a so-called OSSE), as was first described in S16b and extended in S17. 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 S16b 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-sample those data in time according to certain conditions for solar zenith angles (SZA), cloud-fraction and AOT. Table 3 lists the conditions for which observations will be possible. Values for SZA and AOT are inferred from real



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

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

AERONET data files. The maximum cloud-fraction was slightly tuned to obtain similar temporal coverage of observations as real AERONET data (see Sect. 4 and Fig. 4 but the impact 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°	0.01	0.0
Inversion L1.5	80°	0.01	0.03
Inversion L2.0	80°	0.01	0.25

The truth is generated from the high-resolution simulation by averaging AOT and AAOT over a large area (0.5° to 4° 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  $0^\circ$  to  $360^\circ$  longitude and  $-90^\circ$  to  $90^\circ$  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  
5 varies but has a constant angular size of  $0.0625^\circ$  ( $\sim 7$  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 to latitudes below  $60^\circ$ .

Our methodology allows separation of the factors that determine the representation error: spatial extent of the grid-box, and  
10 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.

### 3.1 Statistical parameters

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. For a normal distribution, these quantiles will be equally spaced. Any skewness or extended wings in a  
15 distribution will be readily visible. In addition to quantiles, the mean error and the mean sign-less error will be provided. The mean sign-less error (or mean absolute 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  $\sim 0.8$ .

## 4 Evaluation of G5NR and OSSE

20 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 in AOT and AAOT per site are shown in Fig. 1, top row. In general, simulated AOT shows good agreement with the observations with correlations around 0.75 and slopes around 0.84. Simulated AAOT does not agree as nicely with the observations but there is still correlation ( $\sim 0.45$ ) (the evaluation of AAOT will of course be affected by large measurement errors). The agreement in standard deviation suggests that simulated and observed AOT 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. Removing them from the analysis, significantly increases correlation. 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) campaigns might allow evaluation of the spatial distribution of simulated AOT at even smaller length-scales (10's of kilometers) but they did not start until 2010.

**Table 5.** Correlation in modelled and observed yearly site 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 criterium 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 criterium 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 of  $\sim 0.25$  which results in less observations and less 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. 2, and over a shorter range of values.

Figure 3 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. AAOT measurement errors are not expected to have a big impact on daily variation (which is the difference between two measurements).

Figure 4 evaluates the OSSE and shows temporal coverage (or frequency of observation) per site as a function of latitude. G5NR's simulated coverage is calculated using the limitations described in Table 3 (and explained later 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. 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 limiting factors in observational coverage that are not accounted for in Table 3. One factor is that real Inversion measurements are simply 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). Finally, instrument malfunction & maintenance are not taken in to account, which will explain some of the discrepancy.

In all, it seems that G5NR can realistically simulate spatial and temporal variation in AOT and AAOT, although 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 for the Inversion products in the Northern Hemisphere. Further evidence for G5NR's applicability in an OSSE is given in Fig. 12 where it is shown that the present study agrees with an earlier analysis by Kinne et al. (2013) on the most representative AERONET sites.

## 5 Results

### 5.1 Representation errors in AOT

Figure 5 shows yearly representation errors for AERONET DirectSun L2.0 AOT observations, for three different collocation protocols (see Table 4), as a function of model grid-box size. Hourly collocation yields the smallest representation errors, and this is more pronounced for smaller grid-box sizes. As grid-box size changes from  $4^\circ$  to  $0.5^\circ$ , hourly collocation errors are more than halved from 13% to 5% while those for daily collocation change only from 17% to 12%. By construction, hourly



collocation errors become zero for a grid-box size equal to  $0.0625^\circ$  (the resolution of G5NR). In contrast, yearly collocation errors are dominated by temporal sampling and do not depend much on grid-box size.

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  $1^\circ \times 1^\circ$ . Also, the yearly and daily collocation protocols yield significant bias across the AERONET network. Regionally that bias translates into spatial patterns with east-west or north-south gradients in the representation errors, see Fig. 6 and 7. Such patterns are absent or at least much reduced for hourly collocation. Representation errors for different regions are shown in Fig. 8.

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

Representation errors for AOT do not differ much for the Direct Sun L2.0 and Inversion L1.5 products, see Fig. 9. However, the condition of a minimal AOT ( $\geq 0.25$ ) for valid observations causes large but unsurprising errors for the Inversion L2.0 product. This issue with the Inversion L2.0 data is well-known but the current analysis may be the first realistic estimate of incurred errors. Figure 9 also shows results for two sensitivity studies where observational coverage in the Northern Hemisphere was artificially lowered (see discussion in last paragraph of Sect. 4) but this has no clear impact as temporal coverage is quite low anyway.

AERONET was not designed with representativity in mind but the GAW network was. Nevertheless, Fig. 10 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. A high altitude site on an isolated mountain will observe a shorter atmospheric column than the surrounding grid-box which will cause a negative representation error, see Fig. 11. Actual site altitudes vary from -410 to 5320 m. G5NR site altitudes correlate very well ( $r = 0.98$ ) but tend to underestimate by 28 m on average, with a random error of 171 m.

Previous work by Kinne et al. (2013) ranked AERONET sites according to their representativity, see Table 6. 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 ranking is 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 ranking, see Fig. 12. For large grid boxes of  $4^\circ$  ( $\sim 450$  km near equator), the impact of ranking 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 suggests that Kinne et al. (2013) overestimated the size of the domains ( $\geq 500$  km for  $r > 1$ ) for which their sites were representative. On the other hand, for a grid-box of  $1^\circ$  a substantial reduction in representation error can be seen for  $r > 1$  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. A new ranking of representativity will be introduced in Sect. 6.

Sofar the area averages used to calculate representation errors have been derived for the *entire* grid-box, both the clear and cloudy parts. Under certain circumstances, it may be more realistic to use only the clear part. Examples are the evaluation



of aggregated satellite products with AERONET (like AERONET, satellites can not observed aerosol when there are clouds), or the evaluation of certain models that explicitly calculate clear-sky AOT (usually by estimating clear-sky humidity from grid-box averaged humidity). Representation errors for clear-sky parts of grid-boxes are improved for the yearly and daily collocation protocols, see Fig. 13.

**Table 6.** 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	

## 5 5.2 Representation errors at monthly time-scales

Surprisingly, monthly representation errors are not that much larger than yearly errors, see Fig.14. 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 clearly not the case. As a matter of fact, monthly errors are strongly correlated from month to month, throughout the year, see Fig. 15. 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 with strong correlations (see also Schwarz et al. (2018)). Our preliminary analysis suggests that at the monthly scale, both cases can occur. Figure 16 shows both maximum and minimum monthly errors by site as a function of yearly error. Many sites show large variations in monthly representation errors, but significantly reduced yearly errors, suggesting that the errors are essentially random. However, some sites also show very similar monthly maxima and minima, and yearly errors, suggesting that these errors are better interpreted as biases. Further discussion of this can be found in Sect. 6.

## 5.3 Representation errors in AAOT

Representation errors for Inversion L1.5 AAOT product are shown in Fig. 17. As for AOT, representation errors decrease with decreasing grid-box sizes, although the decrease is small for for yearly collocation. Significant positive biases can be seen for all protocols and large grid-box sizes. These biases are partly due to the AOT > 0.25 criterium for valid observations, which translates into an AAOT = (1 - SSA)AOT > 0.025 criterium for SSA = 0.9. However, other reasons for the positive bias are the proximity of AERONET sites to sources of absorbing aerosol and the impact of orography (e.g. see Schutgens et al. (2017))



and Sect. 6. Unsurprisingly the hourly collocation protocol shows the smallest positive bias and reduces it faster for decreasing grid-box size.

Strikingly, daily collocation yields very similar errors as hourly collocation. This is very likely due to a limitation in the OSSE. The daily variation of AAOT is strongly underestimated by G5NR (see Sect 4 and Fig. 3), possibly due to an absence of diurnal anthropogenic emission profiles.

Regionally, there is some variation in representation errors but not a lot, see Fig. ???. The exception is for the yearly collocation protocol which allows significant biases for sites in South America and Africa. This is related to the AOT criterium for valid observations and the dominant influence of episodic biomass burning for these two continents: outside the burning season much less observations are made. Consequently the observations will favour the absorbing biomass burning aerosol.

A comparison between AERONET and GAW, Fig. 19, shows error distributions that are positively skewed for AERONET and negatively skewed for GAW-AOT. The smaller bias for GAW than AERONET is due to a balancing of the positive bias due to the AOT criterium for valid observations and the negative bias due to site altitude (see also Fig 10 and its discussion).

An analysis of the impact of the site rankings by Kinne et al. (2013), shows similar results for AAOT as for AOT, see Fig 20.

#### 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. Their analysis found that AERONET Inversion L1.5 AAOT representation errors exhibit a global bias of 30% for  $2^\circ \times 2^\circ$  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. 21 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  $2.5^\circ \times 1.27^\circ$  resolution. The downscaling was accomplished with the help of a high-resolution ( $0.1^\circ \times 0.1^\circ$ ) 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 and explain the different results in Wang et al. (2018) and the current study.

Figure 22 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. 23. As a matter of fact, representation errors from the current study do not show a systematic dependence on emission distributions.

## 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° 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 bin boundaries 0%, 5%, 10%, 20%, 40% and up. Using a block bootstrap method on time-series per site, the uncertainty in yearly representation errors was assessed. 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°) and small errors (< 10%), this may drop down to 66% of the resampled time-series. In any case, yearly relative spatial representation errors can be classed robustly. Of course, G5NR and the OSSE are not perfect, which will introduce an uncertainty into the ranking that can not currently 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 as does grid-box size) 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 AOT or AAOT 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 OSSE was constructed that simulates the frequency of AERONET observations taking SZA, 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. AERONET mean AOT per site, as well as yearly and daily variability were estimated correctly within a factor less than  $2\times$ . Considering that G5NR generates its own meteorology, G5NR 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 impressive. 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  $0.5^\circ$  to  $4^\circ$  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% globally); 2) AEROCOM evaluation protocol is sub-optimal (daily collocation allows errors of typically 25% in coherent regional patterns). Instead hourly collocation was advocated. Also, 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.

*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 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. 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?;
- 5 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](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
- 10 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
- 15 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. 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

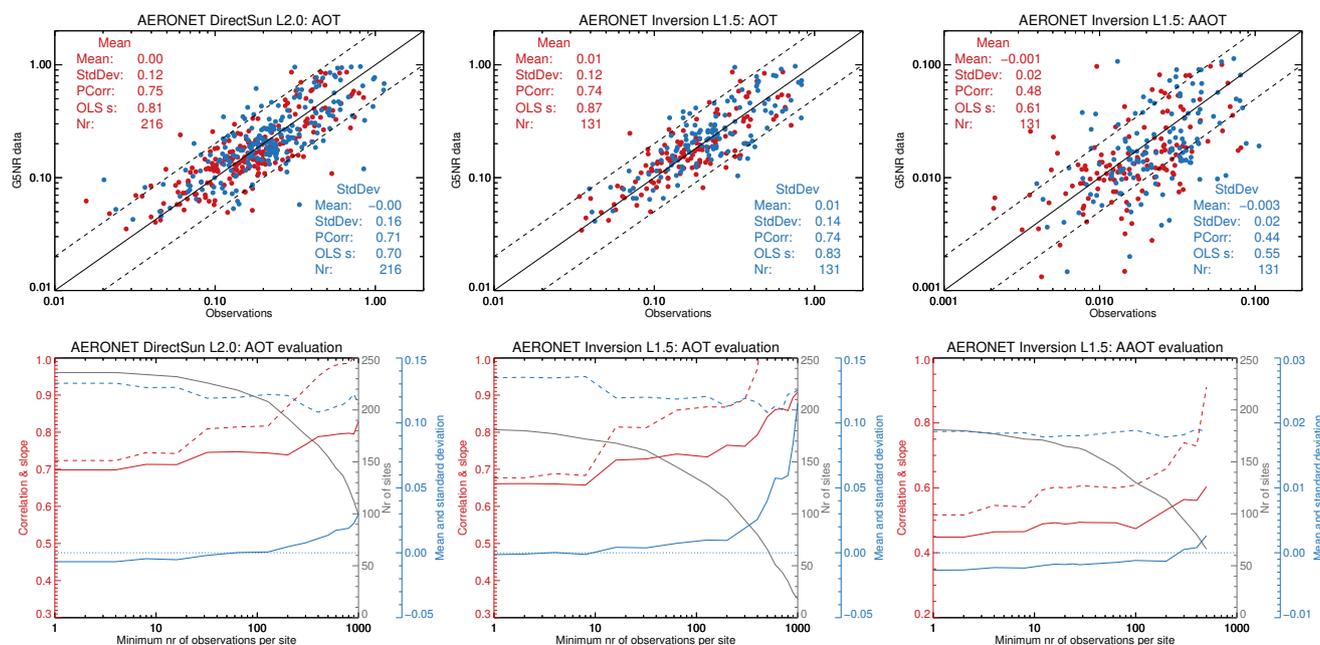
- Anderson, T. E., Charlson, R. J., Winker, D. M., Ogren, J. A., and Holmen, K.: Mesoscale Variations of Tropospheric Aerosols, *J. Atmospheric Sciences*, 60, 119–136, 2003.
- Boersma, K. F., Vinken, G. C. M., and Eskes, H. J.: Representativeness errors in comparing chemistry transport and chemistry 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., Kyrcher, 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 Roozendaal, 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., Wittechen, 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.



- Gelaro, R., Putman, W. M., Pawson, S., Draper, C., Molod, A., Norris, P. M., Ott, L., Privé, N., Reale, O., Achuthavariar, 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.
- 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, 50673, <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.
- 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, D24203, <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 meteorological observations, June 1981, Boulder, Colo, *Bulletin of the American Meteorological Society*, 63, 761–764, 1982.
- 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.
- 10 Schutgens, N.: Representativeness of AERONET and GAW aerosol observation sites, <https://doi.org/10411/XDZD4A>, <https://hdl.handle.net/10411/XDZD4A>, 2019.
- Schutgens, N., 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.
- 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.
- 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.
- 25 Schwarz, M., Follini, D., Hakuba, M., and Wild, M.: From Point to Area : Worldwide Assessment of the Representativeness of Monthly Surface Solar, *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.
- 30 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.
- 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.
- 35 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.



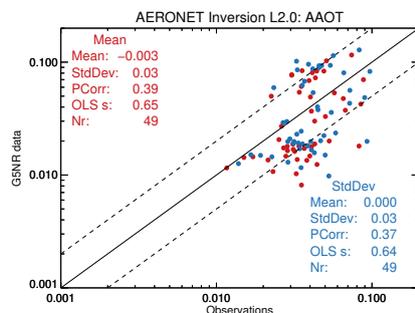
**Figure 1.** Evaluation of the G5NR simulation of AOT and AAOT with AERONET data. The top row shows evaluation against three different datasets. Each dot represents statistics for a single AERONET site (with at least 100 observations in 2006); the mean value is shown in red and the standard deviation 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 is shown.

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.

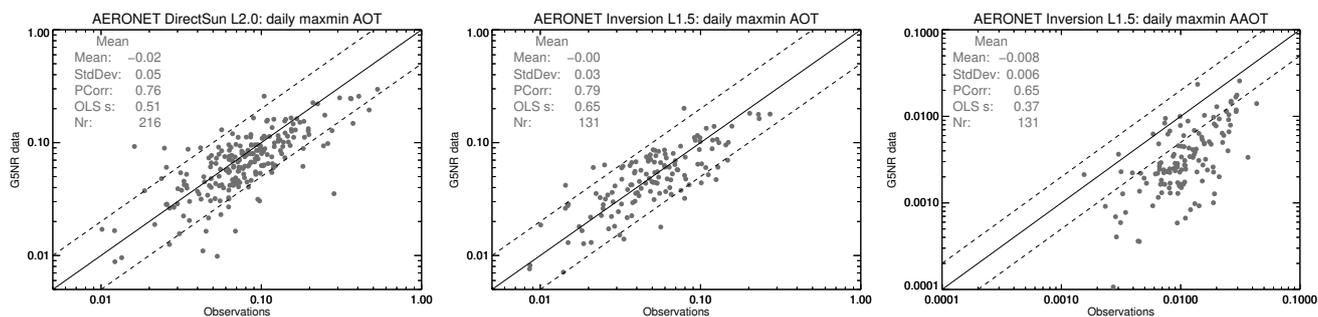
Wang, R., Balkanski, Y., Boucher, O., Ciais, P., Schuster, G. L., Chevallier, F., Samset, B. H., Liu, J., Piao, S., Valari, M., and Tao, S.: Estimation of global black carbon direct radiative forcing and its uncertainty constrained by observations, *Journal of Geophysical Research*, 121, 5948–5971, 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

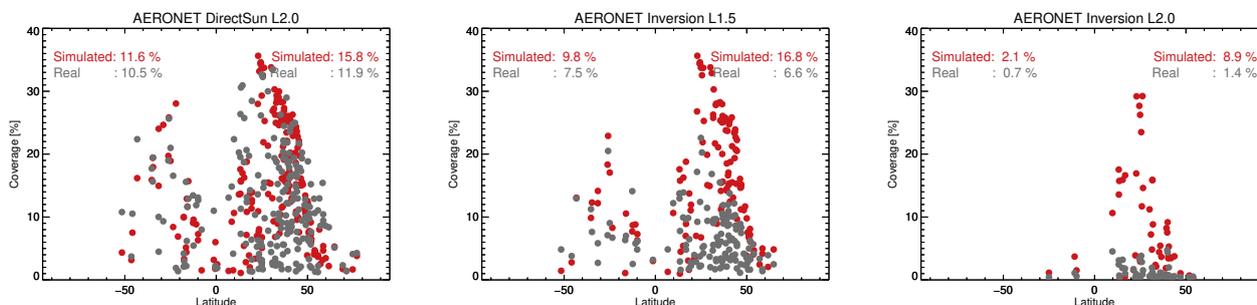
Weigum, N. M., Stier, P., Schwarz, J. P., Fahey, D. W., and Spackman, J. R.: Scales of variability of black carbon plumes over the Pacific Ocean, *Geophysical Research Letters*, 39, <https://doi.org/10.1029/2012GL052127>, <http://doi.wiley.com/10.1029/2012GL052127>, 2012.



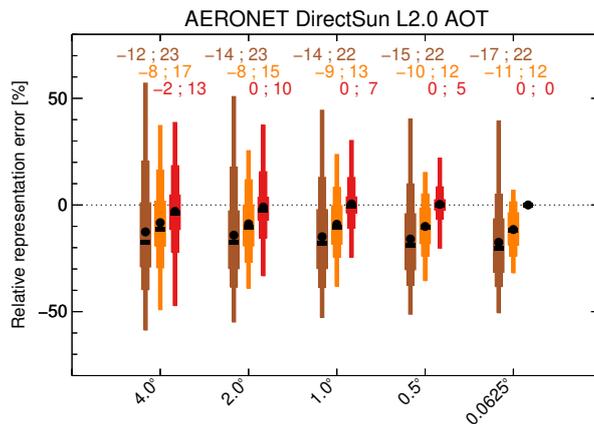
**Figure 2.** Evaluation of the G5NR simulation of AAOT with L2.0 AERONET data. Each dot represents statistics for a single AERONET site (with at least 30 observations in 2006); the mean value is shown in red and the standard deviation in blue. The coloured text summarizes the statistics over all data points in the figure.



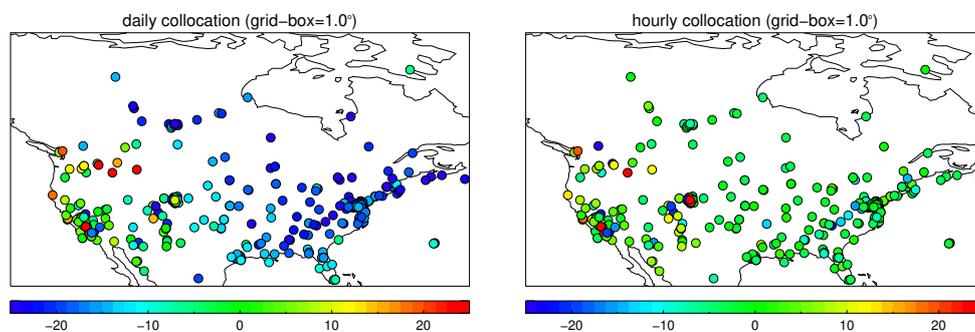
**Figure 3.** Evaluation of the G5NR simulation of daily variation in AOT and AAOT with AERONET data. Each dot represents statistics for a single AERONET site (with at least 100 observations in 2006). The grey text summarizes the statistics over all data points in the figure.



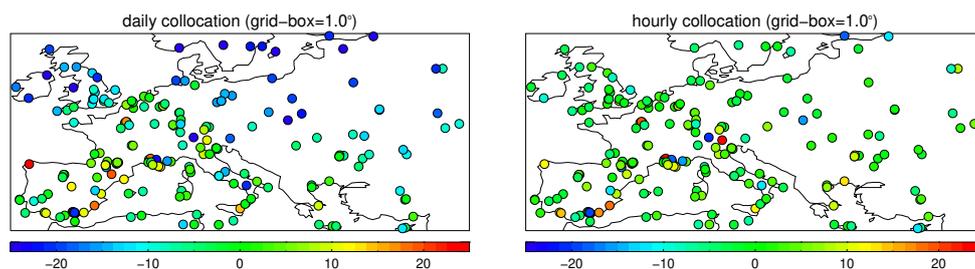
**Figure 4.** Evaluation of the observational coverage predicted by the OSSE with AERONET observations. Each dot represents statistics for a single AERONET site (with at least 100 observations in 2006, at least 30 observations for Inversion L2.0). The grey dots are real AERONET data, the red dots are simulated by the methodology described in Sec. 3. The numbers in the graph are temporal coverages estimated by hemisphere.



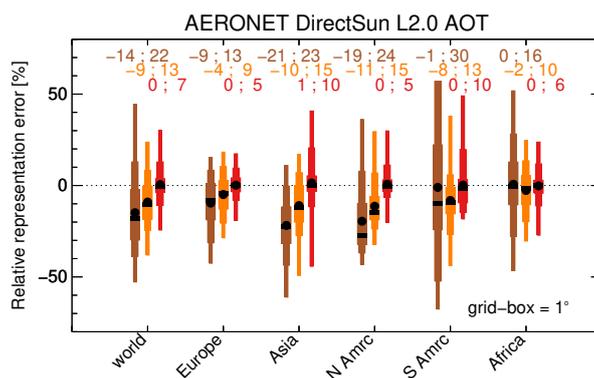
**Figure 5.** Yearly representation errors for AOT from DirectSun L2.0 AERONET for different model grid-box sizes. The colours indicate different collocation protocols: yearly (brown), daily (orange) and hourly (red). Numbers on top are mean of the errors and mean of the sign-less errors.



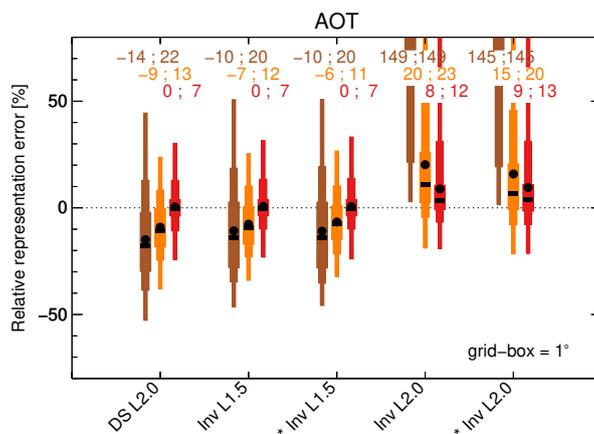
**Figure 6.** Yearly representation errors for AOT from DirectSun L2.0 AERONET in Northern America, for two different collocation protocols (top: daily; bottom: hourly) and a model grid-box size of  $1^\circ$ .



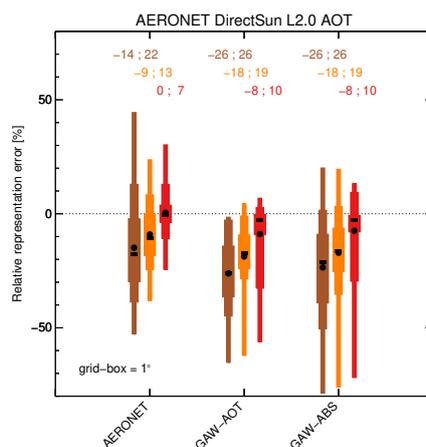
**Figure 7.** Yearly representation errors for AOT from DirectSun L2.0 AERONET in Europe, for two different collocation protocols (top: daily; bottom: hourly) and a model grid-box size of  $1^\circ$ .



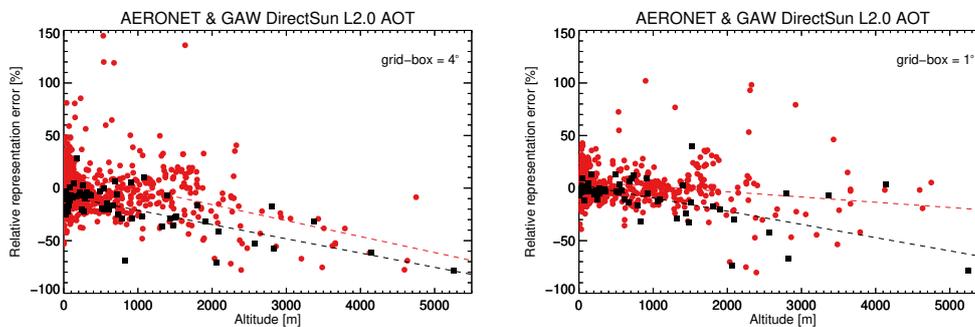
**Figure 8.** Yearly representation errors for AOT from DirectSun L2.0 AERONET in different regions and a model grid-box size of  $1^\circ$ .



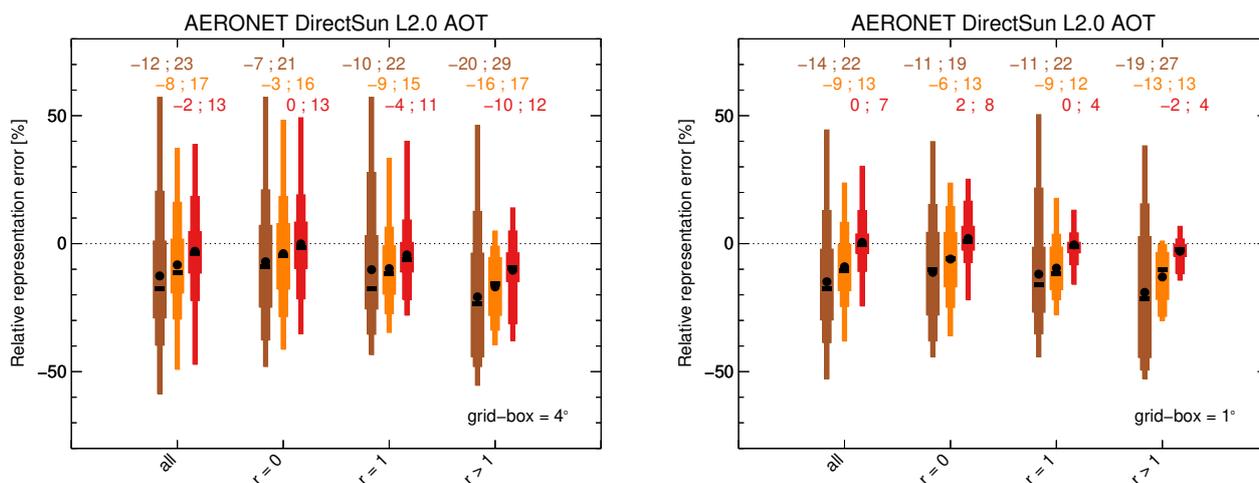
**Figure 9.** Yearly representation errors for AOT from AERONET for different products and a model grid-box size of  $1^\circ$ . The \* refers to Inversion products with artificially lowered temporal coverage in the Northern Hemisphere (see Sect. 4).



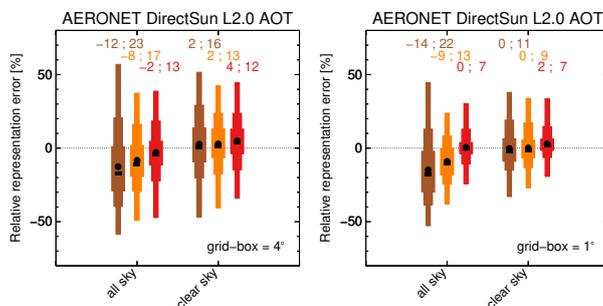
**Figure 10.** Yearly representation errors for AOT from Direct Sun L2.0 AERONET and GAW and a model grid-box size of  $1^\circ$ .



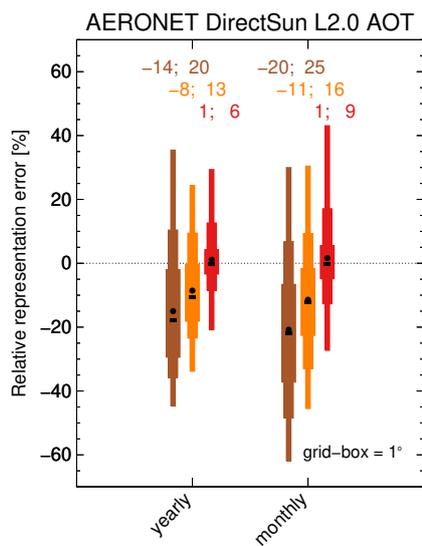
**Figure 11.** Yearly representation errors for AOT from Direct Sun L2.0 AERONET (red) and GAW (black) as a function of site altitude, for a model grid-box size of either  $4^\circ$  or  $1^\circ$ ; using hourly collocation.



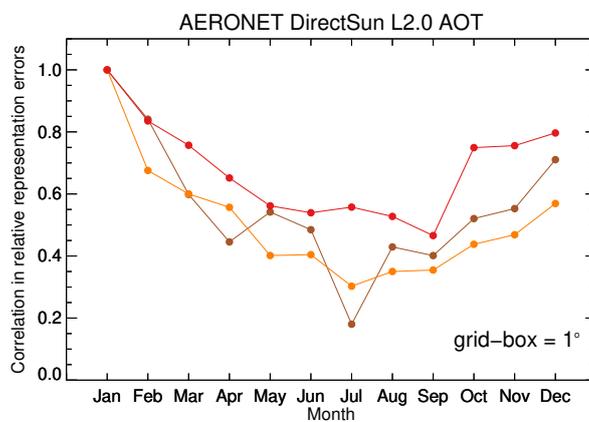
**Figure 12.** Yearly representation errors for AOT from DirectSun L2.0 AERONET for different representation rankings by Kinne et al. (2013), for a model grid-box size of either 4° or 1°.



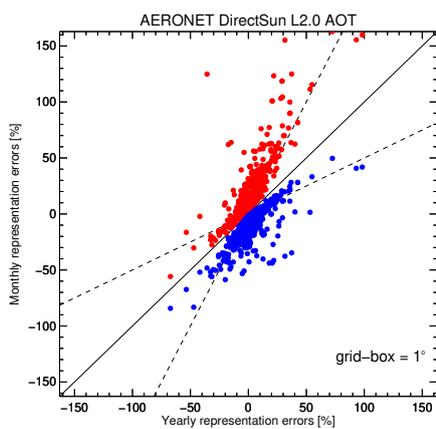
**Figure 13.** Yearly representation errors for AOT from from DirectSun L2.0 AERONET using all-sky or clear sky conditions and model grid-box size of 4° (left) or 1° (right).



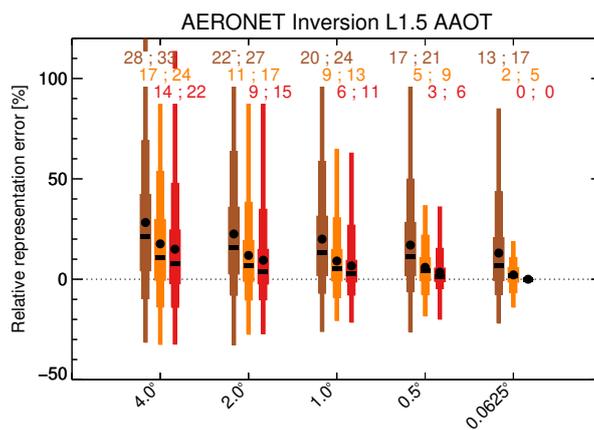
**Figure 14.** Yearly and monthly representation errors for AOT DirectSun L2.0 AERONET, for a model grid-box size of  $1^\circ$ .



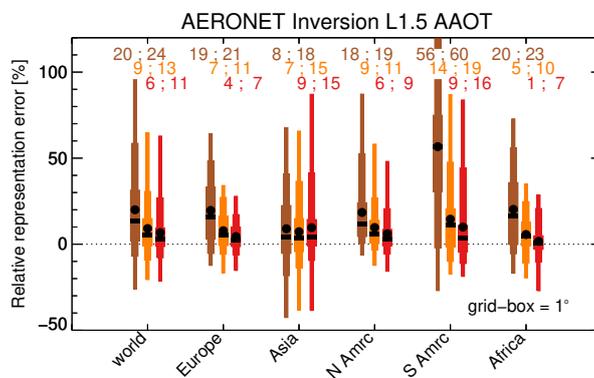
**Figure 15.** Correlation in monthly representation errors with errors for January, for AOT DirectSun L2.0 AERONET, for a model grid-box size of  $1^\circ$ .



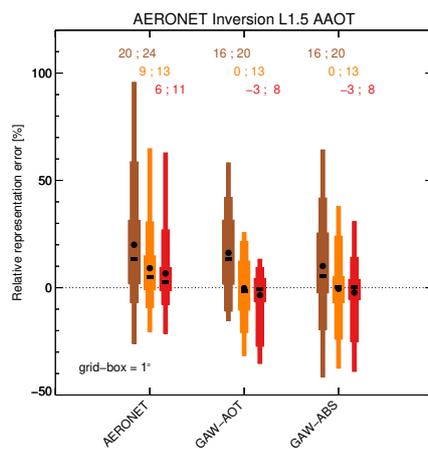
**Figure 16.** Maximum (red) & minimum (blue) monthly representation errors versus yearly representation errors, for AOT DirectSun L2.0 AERONET, for a model grid-box size of  $1^\circ$  and hourly collocation



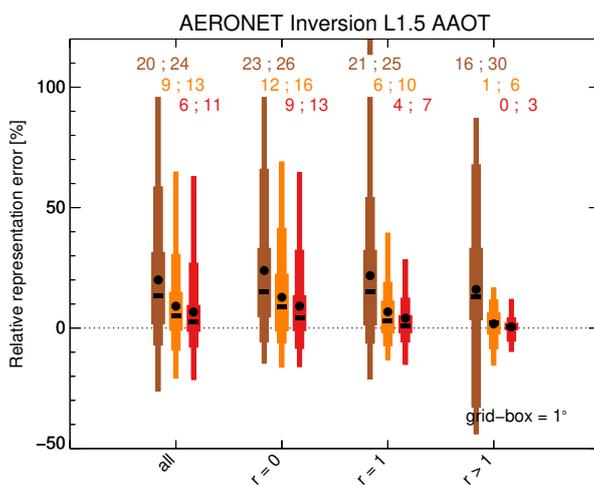
**Figure 17.** Yearly representation errors for AAOT from Inversion L1.5 AERONET for different model grid-box sizes. The colours indicate different protocols: no collocation (brown), daily collocation (orange) and hourly collocation (red). Numbers on top are mean of the errors and mean of the sign-less errors.



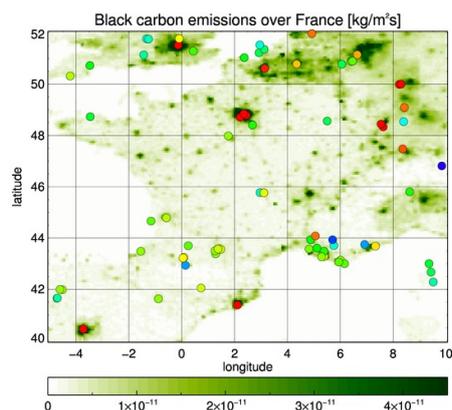
**Figure 18.** Yearly representation errors for AAOT from Inversion L1.5 AERONET in different regions and a model grid-box size of 1°.



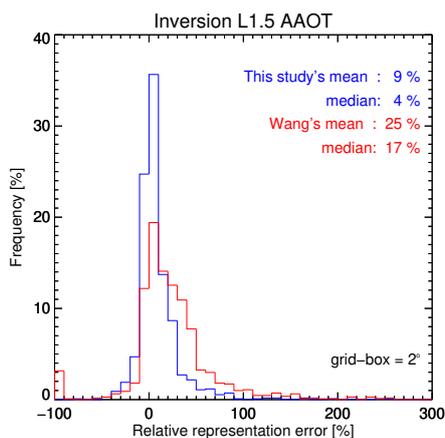
**Figure 19.** Yearly representation errors for AAOT from Inversion L1.5 AERONET and GAW and a model grid-box size of  $1^\circ$ .



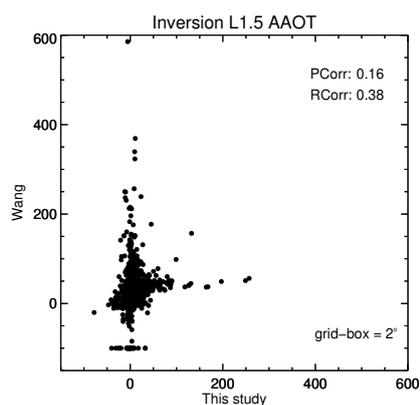
**Figure 20.** Yearly representation errors for AAOT from Inversion L1.5 AERONET for different representation rankings by Kinne et al. (2013), for a model grid-box size of  $1^\circ$ .



**Figure 21.** Black carbon emissions over France, Europe, with the representation errors in AAOT from Inversion L1.5 AERONET superimposed. The representation errors use the same colour bar as in Fig. 7 and runs from  $-25\%$  (blue) to  $+25\%$  (red).



**Figure 22.** 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  $2^\circ$ . The representation error shown is the spatial representation error (Schutgens et al., 2017), i.e. temporal sampling of observations is ignored.



**Figure 23.** 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  $2^\circ$ . 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.