JOURNAL OF GEOPHYSICAL RESEARCH, VOL. ???, XXXX, DOI:10.1029/,

Evaluation of NASA Deep Blue/SOAR aerosol retrieval algorithms applied to AVHRR measurements

A. M. Sayer[1,2], N. C. Hsu[2], J. Lee[2,3], N. Carletta[2,4], S.-H. Chen[2,4], and A. Smirnov[2,4]

N. Carletta, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA.

S.-H. Chen, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA.

N. C. Hsu, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA.

J. Lee, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA.

A. M. Sayer, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA. (andrew.sayer@nasa.gov)

A. Smirnov, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA.

[1]Goddard Earth Sciences Technology and

Abstract. The Deep Blue (DB) and Satellite Ocean Aerosol Retrieval (SOAR) algorithms have previously been applied to observations from sensors like the Moderate Resolution Imaging Spectroradiometers (MODIS) and Sea-viewing Wide Field-of-view Sensor (SeaWiFS) to provide records of mid-visible aerosol optical depth (AOD) and related quantities over land and ocean surfaces respectively. Recently, DB and SOAR have also been applied to Advanced Very High Resolution Radiometer (AVHRR) observations from several platforms (NOAA11, NOAA14, and NOAA18), to demonstrate the potential for extending the DB and SOAR AOD records. This study provides an evaluation of the initial version (V001) of the resulting AVHRR-based AOD data set, including validation against Aerosol Robotic Network (AERONET) and ship-borne observations, and comparison against both other AVHRR AOD

Research (GESTAR), Universities Space Research Association.

[2]NASA Goddard Space Flight Center, Greenbelt, MD, USA.

[3]Earth Systems Science Interdisciplinary Center (ESSIC), University of Maryland, College Park, MD, USA.

[4]Science Systems and Applications, Inc., Lanham, MD, USA.

records and MODIS/SeaWiFS products at select long-term AERONET sites. Although it is difficult to distil error characteristics into a simple expression, the results suggest that one standard deviation confidence intervals on retrieved AOD of $\pm(0.03+15\,\%)$ over water and $\pm(0.05+25\,\%)$ over land represent the typical level of uncertainty, with a tendency towards negative biases in high-AOD conditions, caused by a combination of algorithmic assumptions and sensor calibration issues. Most of the available validation data are for NOAA18 AVHRR, although performance appears to be similar for the NOAA11 and NOAA14 sensors as well.

1. Introduction

Remote sensing of aerosol optical depth (AOD) from space has been performed using a wide variety of sensor types. Passive polar-orbiting single-view imaging radiometers such as the Advanced Very High Resolution Radiometer (AVHRR), Sea-viewing Wide Field-of-view Sensor (SeaWiFS), Moderate Resolution Imaging Spectroradiometer (MODIS), Medium Resolution Imaging Spectroradiometer (MERIS), and Visible Infrared Imaging Radiometer Suite (VIIRS) include several important features suited for this task. Specifically, they typically have moderate spatial pixel sizes (sub-km to several km), broad swaths (providing views of a given location on the Earth approximately daily), and make measurements in bands at solar (and often thermal) wavelengths sensitive to the atmospheric aerosol loading. Thus they have been widely used via various techniques for such applications over both land (e.g. *Hsu et al.*, 2004, *Levy et al.*, 2007, *von Hoyningen-Huene et al.*, 2011, *Lyapustin et al.*, 2011) and water (e.g. *Stowe et al.*, 1997, *Tanré et al.*, 1997, *Mishchenko et al.*, 1999, *Ahmad et al.*, 2010, *Sayer et al.*, 2012a, 2017) surfaces.

Similarities in observation characteristics between sensors of this type facilitate the application of similar retrieval techniques, moving towards the goal of a long-term consistent AOD record. AVHRR is particularly advantageous for long-term monitoring as the first was launched in 1978 and AVHRRs are still flying at the present time. Even with a common algorithm, however, this goal is beset by numerous challenges as no two sensors are identical; issues such as precise measurement capabilities, cloud screening, and calibration, among others, can influence sensor-to-sensor data consistency (e.g. *Jeong and Li*, 2005, *Li et al.*, 2009, *Kahn et al.*, 2011, *Mishchenko et al.*, 2012). Other instrument types offer important capabilities of their own relevant to aerosol retrieval (e.g. multi-angle, polarimetry, UV wavelengths, hyperspectral, lidar). These features provide additional or alternative information content (e.g. *Hasekamp and Landgraf*, 2007), although such sensors often lack comparatively in some combination of pixel size, swath width, or data record length.

The long time series of the AVHRRs motivated recent efforts to apply versions of the over-land Deep Blue (DB, *Hsu et al.*, 2004, 2006, 2013, *Sayer et al.*, 2012b) and over-water Satellite Ocean Aerosol Retrieval (SOAR, *Sayer et al.*, 2012a, 2017) algorithms, which have previously been applied to AOD retrieval from SeaWiFS, MODIS, and VIIRS, to the AVHRRs. An initial version (V001) of an AVHRR Deep Blue data set, combining DB and SOAR retrievals, has been created for a subset of the AVHRR sensor records (those flying on NOAA11 from 1989-1990; NOAA14 from 1995-1999; NOAA18 from 2006-2011). Although the individual instruments were operational for longer, the satellites were launched with nominal daytime Equatorial local solar crossing times around 1:30 pm and drifted later while in orbit, which has consequences for sampling and time series analysis. Thus the time periods processed to date were chosen to cover the periods where these sensors had Equatorial crossing times between 1:30 pm and 3 pm, most comparable with other early-afternoon platforms (e.g. the A-Train).

The new data set is freely available to download, along with a user guide, from https://portal.nccs.nasa.gov/datashare/AVHRRDeepBlue. Additional information and documentation is provided at https://deepblue.gsfc.nasa.gov. The specific implementation of these algorithms to the AVHRRs is described in a companion paper, *Hsu et al.*

[2017]. Note the data set is referred to as 'AVHRR Deep Blue' although it is composed of both the distinct DB and SOAR algorithms.

The goal of this study is to evaluate these new data products, thereby providing guidance for data users and suggest directions for refinement for a future processing of the whole multi-satellite AVHRR data record. This is accomplished through several sets of comparisons. First, ground-truth reference data from the Aerosol Robotic Network (AERONET, *Holben et al.*, 1998), Maritime Aerosol Network (MAN, *Smirnov et al.*, 2009), and earlier ship-based AOD observations provide a validation. Second, comparing to existing related satellite-based AOD records provides broader-scale context. Section 2 describes the data products used, and the general validation methodology. The following Sections 3 and 4 provide a validation of the SOAR over-water and DB over-land AOD retrievals respectively, while Section 5 is a comparison of the new AVHRR Deep Blue data set against other satellite products. Section 6 provides a brief summary.

2. Data set descriptions

2.1. AVHRR Deep Blue/SOAR AOD retrievals

The adaptation of the DB and SOAR algorithms to the AVHRRs is described by *Hsu et al.* [2017]. The physical principles behind the AVHRR application of the algorithms are the same as those behind the SeaWiFS, MODIS, and VIIRS versions. However, as only two solar bands are available for most of the AVHRRs and there is no on-board solar band calibration, various algorithmic constraints and cloud screening tests must be tightened to avoid unstable or unphysical results. Brief descriptions of some key features of the AVHRR implementation follow.

The primary data products are the AOD at wavelengths of 550 nm (due to its common use as a reference wavelength in the scientific community), AOD at AVHRR band 1, and (over water only) AOD at band 2. In general, mentions of AOD without a specific wavelength should be taken to refer to 550 nm. The exact central wavelengths of these bands vary slightly between the different AVHRRs, and are referred to herein at 630 nm and 830 nm respectively in the general discussion for simplicity. All calculations, however, use exact sensor-specific wavelengths. Specifically, central wavelengths are 636, 636, and 633 nm for band 1, and 810, 820, and 848 nm for band 2, for NOAA11, 14, and 18 respectively. Multiple AVHRR solar band calibrations have been derived; this initial version of the data set uses that of *Vermote and Kaufman* [1995], which is also used for NASA's long-term normalized difference vegetation index (NDVI) data sets, although the use of other calibrations will be investigated for future versions.

Over land, DB has two methods of estimating surface reflectance for a given pixel, depending on whether the location has a bright (e.g. barren ground, urban areas) or vegetated surface. For bright surfaces, a global seasonally-varying data base of surface reflectance is constructed using a similar method to the minimum reflectance technique, applied to the whole sensor record. For the other applications of Deep Blue (cf. *Hsu et al.*, 2013) the primary wavelength is 412 nm, at which the surface reflectance is fairly dark, even for deserts. AVHRR lacks this channel so band 1 near 630 nm is used instead. As the surface is typically somewhat brighter at 630 nm than 412 nm, however, the aerosol signal is somewhat reduced, and the resulting AOD uncertainty is larger. Over the brightest surfaces (e.g. snow, salt pans, some deserts) the surface is too bright and no retrieval is performed to due a lack of sensitivity to AOD variations. Over vegetated surfaces, reflectance is estimated dynamically, as it often varies more rapidly in time than over arid surfaces. As AVHRR lacks shortwave infrared (SWIR) bands which are useful to track these changes, the surface reflectance is modelled as an empirical function of NDVI. A similar approach was previously developed for SeaWiFS DB (*Hsu et al.*, 2013), as SeaWiFS also lacks SWIR channels, and was found to perform well (*Sayer et al.*, 2012b). Full details of both approaches are provided by *Hsu et al.* [2017]. Note that separate surface data bases and NDVI relationships are constructed for each sensor, as they each have slightly different spectral response functions.

For both land surface types, the aerosol optical model is assumed on a regional and seasonal basis, due to the aforementioned limited information content of AVHRR. These models are drawn from the same sets of models used for other DB applications, adapted to AVHRR wavelengths. Once the surface reflectance has been obtained, band 1 AOD is retrieved directly from the AVHRR measurement. The AOD at 550 nm is extrapolated from this using an assumed (regionally and seasonally-dependent) Ångström exponent (AE) based on AERONET climatologies (*Hsu et al.*, 2017). Thus, both 550 nm and band 1 AOD are provided within the data set, even though AVHRR has no band near 550 nm.

Over water, SOAR uses both bands 1 and 2 in a simultaneous inversion to determine AOD and the best-fitting aerosol optical model from a choice of dust, fine-mode dominated, and maritime optical models. Surface reflectance includes contributions from wind-speed-dependent foam and Sun glint, as well as 'underlight' from suspended pigments, although this latter term is small for AVHRR bands. This is essentially the same basic approach as in the SeaWiFS retrieval (*Sayer et al.*, 2012a), although the AVHRR algorithm makes use of improvements to the surface reflectance model and aerosol optical models (e.g. nonspherical dust) which were developed during the VIIRS implementation of SOAR (*Sayer et al.*, 2017). Unlike these other applications, for AVHRR the fine mode fractional contribution to AOD is fixed (one different value for each aerosol type), rather than retrieved directly, again due to the limited spectral information provided by AVHRR. The AOD at 550 nm is then obtained in a self-consistent approach using the retrieved aerosol loading and best-fitting aerosol optical model. This model and its associated AE are also reported in the data set. As over land, separate lookup tables are created for each AVHRR sensor.

Each retrieval also has an associated quality assurance (QA) flag between 1 and 3. QA=1 ('poor') indicates internal tests (*Hsu et al.*, 2017) suggest some potential problem, such as cloud-contamination or an improper surface model, so the retrieval is likely to be quantitatively less reliable. These retrievals should not be used for most applications. QA=3 ('good') pass all checks, and are therefore least likely to suffer from these issues. QA=2 ('moderate') retrievals are an intermediate category. Most retrievals are assigned either QA=1 or QA=3. In this analysis, only retrievals with QA=2 or 3 are used, which is the general recommendation for almost all data users.

The resulting data Level 2 (L2) products are provided at approximately $8.8 \times 8.8\,\mathrm{km}^2$ horizontal pixel size at the sub-satellite point (2×2 Global Area Coverage AVHRR pixels) for daytime (solar zenith angle $<84°$) land and ocean pixels free from cloud, snow/ice, or Sun glint. Level 3 (L3) daily/monthly composites are also available, created from QA≥ 2 retrievals gridded to 1° resolution. Consecutive orbits from AVHRR overlap, particularly at high latitudes, and so some L3 daily grid cells contain contributions from multiple orbits, spaced approximately 90 minutes apart. Note that the AVHRR daily L3 data product requires at least 5 retrievals for a grid cell to be valid, and the monthly mean at least 3 days with sufficient data within a month.

As with other Deep Blue data products, and indeed many other satellite data sets, the uncertainty on retrieved AOD is a function of the true AOD. This is somewhat unavoidable given the nature of the measurements and required retrieval assumptions. An expected error (EE) envelope is defined, intended to represent a one-standard-deviation confidence envelope around the retrieved AOD (e.g. *Sayer et al.*, 2013), such that one standard deviation of retrievals (i.e. about 68 %) match the ground truth AOD to within this level (and, following Gaussian statistics, approximately 95 % within twice the EE envelope, etc.). For the initial AVHRR Deep Blue data set, the EE is taken (*Hsu et al.*, 2017) as ±(0.03+15 %) over water, and ±(0.05+25 %) over land (with AOD defined relative to the Sun photometer values, i.e. a diagnostic rather than prognostic measure), for all wavelengths considered. These may be refined further in the future.

The validation analysis includes discussions of the fraction f of points where the AVHRR-AERONET difference is smaller than the EE. By the definition of the EE, the target in the ideal case for a useful uncertainty metric is thus $f \approx 0.68$, with substantially lower values indicating retrieval errors are on average larger than this envelope, and substantially higher values indicating that performance under this circumstance appears better than anticipated.

2.2. Other satellite AOD products used

2.2.1. AVHRR

Two other over-ocean AVHRR AOD retrieval algorithms are also examined in this work. The first is from the Global Aerosol Climatology Project (GACP), described most recently by *Geogdzhayez et al.* [2015], which provides monthly AOD at 550 nm and AE over ocean on a 1° grid. The second is the NOAA aerosol climate data record (CDR) version 3 (*Zhao*, 2016, *Zhao et al.*, 2016), which provides AOD at band 1 (but not 550 nm); daily and monthly data (level 3) are provided on a 0.1° grid. Various approaches to retrieve AOD over land from the AVHRRs have been proposed and demonstrated on local or regional scales (e.g. *Knapp and Stowe*, 2002, *Riffler et al.*, 2010, *Mei et al.*, 2014, *Gao et al.*, 2016). However these have not been applied to the full AVHRR records to produce global publicly-available data products. The one exception is the Polar Multi-sensor Aerosol product (PMAp, *EUMETSAT*, 2016), although this has only been applied to the AVHRR sensors on the MetOp platforms in forward-processing, and does not overlap with the current DB/SOAR AVHRR record. Therefore the comparison with other AVHRR products is restricted to the aforementioned over-ocean data records only.

2.2.2. SeaWiFS

The SeaWiFS mission covered the time period September 1997-December 2010, with a small number of temporary outages, and the SeaWiFS Deep Blue data set includes DB retrievals over land (*Sayer et al.*, 2012b, *Hsu et al.*, 2013) and the initial application of SOAR over water (*Sayer et al.*, 2012a). The current version 4 is used; specifically, the monthly mean 550 nm AOD product at 1° spatial resolution. As many 1° grid cells contain both land and water pixels, the product provides both a 'combined' land/ocean data set as well as results from the DB/SOAR algorithms individually. The latter are used here. A level 3 monthly grid cell from this product is only filled if it contains data from at least 3 different days within a given month, and a grid cell is valid on a particular day if it contains at least 3 retrievals passing QA checks.

2.2.3. MODIS

This study makes use of MODIS AOD from the Aqua platform (launched in May 2002 and still operational), as it has a similar local solar crossing time (1:30 pm at the Equator for the daytime nodes) to the nominal orbit times of the platforms hosting the specific AVHRR sensors considered in this study. Over land, data from the DB algorithm are used (*Hsu et al.*, 2013, *Sayer et al.*, 2013). Over water, since there has not yet been an application of SOAR to MODIS, the standard MODIS ocean retrieval product (*Tanré et al.*, 1997, *Levy et al.*, 2013) is used as a point of reference. This shares similar physical principles to SOAR (multispectral inversion based on matching observed reflectances to results from radiative transfer models) but numerous algorithmic specifics are different. Both DB and the MODIS ocean algorithm provide 550 nm AOD, used herein.

The results in this work are taken from the current Collection 6 level 3 monthly product (identifier MYD08_M3). Note that this product as standard does not have any thresholds applied to determine whether a grid cell is sufficiently well-sampled or not to be meaningful (i.e. one retrieval passing QA checks in a whole month results in a populated monthly mean AOD). In practice most populated grid cells contain several hundred retrievals from multiple days, but a small fraction contain only a dozen or so. There is no way within the MYD08_M3 data product to identify how many individual days contributed to a specific cell within a given month. As a result, to mitigate the influence of a small number of sparsely-populated cells, an additional filtering step is applied herein to remove grid cells with fewer than 30 retrievals within a month. The specific threshold chosen does not strongly affect the results presented herein.

2.3. AERONET

Starting from a few sites in the early 1990s, AERONET has expanded to provide several hundred sites with multi-year (in some cases decadal or longer) aerosol observations (*Holben et al.*, 1998, 2001), as well as dedicated deployments during intensive field campaigns (*Holben et al.*, 2017). The Cimel Sun photometers used by AERONET provide observations of columnar spectral AOD and water vapor from direct-Sun observations with a temporal frequency of approximately 3-15 minutes (dependent on site, and limited to daytime cloud-free periods), as well as a range of products from the spectral deconvolution of the AOD (*O'Neill et al.*, 2003), and from inversions of almucantur scans (*Dubovik and King*, 2000). The direct-Sun products have become a standard for satellite/model AOD validation, due to the low level of uncertainty ($\sim$0.01 at midvisible and longer wavelengths; *Eck et al.*, 1999) and consistency in instrument calibration and data processing between sites and in time.

This study uses the AERONET direct-Sun version 2 level 2 (cloud-screened and quality-assured; *Smirnov et al.*, 2000a) data products. All instruments provide a standard set of wavelengths (440, 675, 870, and 1020 nm for AOD), and some include additional wavelengths. In this analysis, AERONET AOD are interpolated spectrally to 550 nm as well as band 1 and 2 central wavelengths for the individual AVHRR sensor in question. This interpolation is performed using the closest available AERONET wavelength and the AE, and adds negligible additional uncertainty.

AVHRR and AERONET data are compared by averaging satellite data within 25 km of the AERONET site and AERONET data within $\pm$30 minutes of the satellite overpass, which has been the standard approach (cf. previously-cited satellite AOD papers). This approach is designed to mitigate the influence of spatiotemporal variability on the comparison, although it cannot eliminate sampling differences entirely (see discussion by *Hyer et al.*, 2011 and *Kahn et al.*, 2011). When comparing DB land retrievals only AVHRR land pixels are considered, and when comparing SOAR ocean retrievals only AVHRR water pixels are considered; as noted previously, the AVHRR data are QA-filtered prior to this averaging process. A matchup is valid if there is at least one AVHRR retrieval in the spatial window and at least one AERONET observation in the temporal window.

Note that no AERONET matchups are available for the NOAA11 part of the analysis, because no sites were active during the time period available in the V001 AVHRR Deep Blue data set (1989-1990).

2.4. Ship-borne AOD measurements

The Maritime Aerosol Network (MAN; *Smirnov et al.*, 2009, 2011) is a complement to AERONET, consisting of ship-based AOD measurements made from hand-held Microtops II Sun photometers. These can be used to determine spectral AOD with an approximate uncertainty of 0.02 (*Knobelspiesse et al.*, 2004), i.e. slightly greater than that of the stationary Cimel instruments used in AERONET but still sufficient for a validation of satellite/model data sets. Measurements are made on cruises where equipment and personnel are available, thus enabling Sun photometer-based AOD validation in open ocean regions. With the exception of two pilot cruises in 2004 and 2005, the MAN data base includes cruises from 2006 onwards (and are most frequent in more recent years). Thus, of the satellites considered in this study, MAN data are only available for NOAA18. Here, the ‘series average’ (data acquired with a gap of <2 minutes between observations) level 2 MAN product is used, with the same matchup methodology as applied over land (Section 2.3).

Ship-based measurements of this type were also made prior to the formal establishment of MAN. *Smirnov et al.* [2002] provide a discussion of some. As they were collected by a variety of researchers and not formalised into a consistent data base, the available information (both in terms of AOD wavelengths and precision of spatial/temporal location data) for these earlier cruises is more variable. These measurements are used to provide validation for the NOAA11 and NOAA14 data. To increase the available NOAA11 data record, data from the year 1991 were also included rather than just the 1989-1990 period provided in the initial AVHRR Deep Blue data release. It is possible, however, that 1991 results will be subject to different error characteristics as the June 1991 eruption of Mt. Pinatubo injected a sizeable amount of aerosol into the atmosphere, which spread to cover much of the globe, persisted for several years, with both different optical properties and vertical distribution from tropospheric aerosols found in periods free from strong eruptions (*Lambert et al.*, 1992, *Mishchenko and Geogdzhayev*, 2007).

The same spectral interpolation technique is applied throughout. For NOAA14 matchups, data content and format were similar to those of MAN and so the same matchup criteria are used. NOAA11 matchups typically provide latitude/longitude information to the nearest degree, and data reported as 'morning' or 'afternoon' averages (note NOAA11, 14, and 18 had early-afternoon overpass times), which is somewhat less precise than the MAN data. Where this is the case, all available AVHRR retrievals over a 200 km radius from this point on a given day are averaged and compared with the reported ship-based temporal average. This inevitably increases sampling-related uncertainty, which should be borne in mind in the interpretation, although given the limitations of the available data, it is the best that can be done and the only option to provide a validation for these earlier satellite missions.

3. AOD validation over ocean

3.1. AERONET island/coastal sites

AERONET has expanded significantly through the past few decades. As a result, matchups over water were obtained at 40 island/coastal sites during the NOAA18 period, but only 20 provided data for NOAA14, and none for NOAA11. Summary statistics for the comparisons for these two satellites are provided in Table 1. The focus of this discussion is on NOAA18 data, since it has the greatest available comparison volume, for brevity; conclusions concerning error characteristics over ocean, where not discussed explicitly, are qualitatively and quantitatively similar for NOAA14.

The overall tendencies of AOD retrieval error over ocean for NOAA18 are shown in Figure 1, which splits the data according to AERONET AOD and AE, thus providing a simple categorisation into background (low-AOD), elevated-AOD coarse-mode dominated (i.e. dust, typically), and elevated-AOD fine-mode dominated (i.e. smoke/continental) conditions. The 550 nm and band 1 results show similar behavior; for these bands there is a tendency for a slight positive AOD bias of order 0.02 in the cleanest scenes, gradually decreasing as AOD increases, with a negative bias of approximately 10 % in high-AOD conditions. The neutral point of AOD bias around 0 is about 0.15-0.2. On the other end, the ability to examine the statistics of extreme conditions is limited, since the 95th percentile of AERONET AOD is only 0.48 for these matchups. Note that in a statistical sense a tendency for positive offset is expected for the cleanest conditions, due to the simple fact that AOD is positive definite, so in comparison to a 'truth' reference even with a random distribution of errors the aggregate is likely to be biased positive (i.e. negative AOD is unphysical so an underestimate of AOD in conditions close to zero AOD is impossible). For band 2 (near 848 nm for NOAA18), the bias is more small and positive throughout, and only becomes negative, and to a lesser extent, for dust-like conditions.

The median and central 68 % of retrieval errors fall within or are similar to the EE lines in Figure 1 in most conditions, suggesting that this metric provides a reasonable approximation of retrieval uncertainty on aggregate. It would be desirable in future versions to address biases so that binned statistics of this type fall closer to the zero line. These bias tendencies are indicative of a bias in some combination of sensor calibration or radiative transfer assumptions (most likely aerosol optical model or trace gas absorption, which is not negligible, especially for AVHRR band 2; *Tanré et al.*, 1992).

Figure 2 is analogous to Figure 1, except for NOAA14; the general tendencies between the two are similar, except that (particularly for band 2) the biases are more positive, by around 0.015-0.02 dependent on wavelength (Table 1). Since the two apply the same algorithm, it is likely that calibration differences are the major reason for the discrepancy here. It should also be noted that the data volume is smaller for NOAA14 by about an order of magnitude (in terms of number of matchups) and a factor of two (in terms of sites), since AERONET was less widespread during this period.

Returning to NOAA18, Figure 3 shows site-by-site statistics at 550 nm for the over-ocean comparison. Spatial patterns are similar for NOAA14 data, as well as for data at other wavelengths (not shown). Correlation coefficients tend to be high (0.8-1) for sites with a large dynamic range of AOD (largely continental outflow regions), and smaller for low-AOD regions, where the range of AOD becomes more comparable to the retrieval EE.

Biases tend to be small (magnitude <0.015 at most sites), with the sign dependent on whether it is a predominantly clean or high-AOD region, consistent with Figure 1. Note a few areas with high positive AOD bias are sites in complex coastal areas, particularly ICIPE Mbita (on the shores of Lake Victoria), Hong Kong, Taihu (a large lake near Shanghai), and Darwin (northern Australia). In these areas it is possible that either the turbid water mask is not working effectively, or some pixels identified as ocean are in fact mixed land and ocean, thereby providing a brighter signal than would be expected for an open-ocean scene. These sites are also the ones at which the fraction of points matching AERONET within the EE are significantly lower than the target of 68 %.

The data were also examined for possible biases with respect to changing near-surface wind speeds or total column water vapor amount (omitted for brevity), although these were small (less than 0.02 change in median bias across the range of the variables). Overall, this analysis suggests that caution should be taken in analysis of retrievals in complex coastal environments, and particularly lakeshores, but otherwise typical AOD retrieval biases are close to zero and uncertainty is of order $\pm(0.03+15\,\%)$. Over ocean, the root mean square error (RMSE) at individual sites is typically in the range 0.05-0.075 (Figure 3). Since the biases are in most cases significantly smaller than this, it is unlikely that decreasing the bias, whether through improvements to calibration or baseline aerosol optical models, will significantly decrease the RMSE over ocean or shrink the EE envelope on a global basis. This is a fundamental consequence of AVHRR's limited spectral information and band digitisation. Consequently, improving the correlation over low-AOD ocean sites may be difficult. This suggests that the best path forward for improvement to the ocean retrieval may be to focus on improvement to QA tests in turbid or coastal waters, as these aforementioned sites are those with highest RMSE and lowest compliance with the EE metric.

AOD bias characteristics over ocean are similar to the SeaWiFS application of SOAR reported by *Sayer et al.* [2012a], i.e. a small positive bias in low-AOD conditions but ∼10 % low bias in high-AOD conditions. The optical models in both cases are based on AERONET version 2 inversions, and are common to both sensors (except for the case of dust, where AVHRR adopts a nonspherical model which had not been developed at the time the SeaWiFS data set was created). The similar bias characteristics may plausibly indicate systematic biases in the aerosol optical models (e.g. insufficient absorption), although other causes such as sensor calibration cannot be discounted at the present time. A version 3 AERONET inversion product is expected to become available within the next year or so, at which point the two versions will be compared to see if there is any systematic shift in retrieved size distribution or absorption. If so, updated optical models can be derived and implemented in future SeaWiFS/AVHRR reprocessings.

Finally, although aerosol type should not be considered a primary retrieval data product here, Figure 4 shows histograms of the AERONET AE split according to whether SOAR identified each matchup as predominantly dust-dominated, fine-mode dominated, or clean marine (in terms of retrieved best-fit aerosol optical model). As a reminder (Section 2.1), SOAR sequentially performs the retrieval for each aerosol optical model and reports the best-fitting. So, over ocean, it is instructive to see to what extent SOAR's judgement of likely aerosol optical model compares so the AE (which is related to aerosol fine/coarse-mode optical dominance) derived from AERONET. Over land, the AVHRR application of DB uses a fixed aerosol optical model dependent on location and season and so such a comparison is not possible.

For both the 'all points' and 'AERONET AOD ≥0.2' cases, the general picture is reasonable in that the most common AE when SOAR picks the dust model are low (0-0.5), while the most common AE when SOAR picks the fine-dominated model are higher (1.3-1.8). The distributions do however have fairly long tails, indicating cases where the inferred likely aerosol type from AVHRR is probably incorrect. Therefore, while they may often be reasonable, the best-fitting optical model should not be taken alone as a definitive indicator of likely type or origin of the observed aerosols in the column.

The 'marine' histograms are both broader, reflecting both the more potentially more mixed nature of clean scenes, but also the fact that AERONET AE is somewhat uncertain in low-AOD conditions (e.g. *Wagner and Silva*, 2008). Note also that the marine AE histograms skew to more positive values than expected for typical remote ocean conditions, as reported by *Smirnov et al.* [2011] based on extensive ship-borne observations, which is probably related to the fact that the available AERONET sites are, by their nature, situated in island/coastal areas (i.e. additional continental influence) which may be expected to have a different fine/coarse aerosol partition from the open ocean. This points to the need for validation in both coastal and remote regions.

3.2. Ship-based observations

Table 2 presents statistics of the comparison between AVHRR and ship-based AOD measurements. The results are in general agreement with those obtained in Section 3.1 for coastal/island AERONET sites. Figure 5 shows the locations of each matchup for each sensor, colored to show the aerosol optical model chosen by SOAR in each case.

Although the data volume is small and this is a series of instantaneous snapshots rather than a climatology, it does match intuitive expectations (i.e. open-ocean conditions tend to be have chosen the optical model for clean marine aerosols, and dust/fine-dominated aerosols are chosen largely downwind of expected source locations typical for these aerosol types). This is broadly in agreement with the histograms shown previously in Figure 4.

The 80 matchups with NOAA11 come from two distinct sources. The first is measurements made by Y. Villevalde in the Pacific and North Atlantic oceans, reported in *Villevalde et al.* [1994] and *Smirnov et al.* [1995a]. These both predominantly sampled low-AOD conditions represented of the clean marine atmosphere; for the cruises as a whole, *Villevalde et al.*, 1994 report mean 551 nm AOD of 0.13 and 0.11, and AE of 0.56 and 0.99, for the Pacific and Atlantic legs respectively. The NOAA11 data are in good agreement with these cases, and indeed SOAR chose the 'clean marine' optical model (*Sayer et al.*, 2012c) in almost all these cases. The second set of measurements were led by O. Yershov and took place on several cruises in the North Atlantic, Mediterranean, and Black Sea, and are described by *Smirnov et al.* [1995b]. These sampled both open-ocean and continentally-influenced air masses. One outlying case from a Mediterranean leg of these cruises is responsible for the lower correlation and higher RMSE of these data compared to the NOAA14/NOAA18 observations in Table 2. Manual examination of this case reveals a dust plume near the reported ship location; since the geolocation information of these early ship-borne data were less precise than in later records (Section 2.4), this is likely attributable to sampling differences rather than true retrieval error. AVHRR retrieved in the dust plume with AOD around 0.65, but the ship, potentially up to 50 km in space and several hours distant in time, may have sampled outside the plume, reporting AOD around 0.15. Overall, however, the matchups with NOAA11 are consistent with comparable performance to the later AVHRR sensors, and as noted, the available validation data for this time period are very limited.

All of the 20 NOAA14 matchups come from measurements between the US East Coast and Bermuda during the summer 1996 TARFOX campaign, described in *Smirnov et al.* [2000b]. This cruise sampled a mixture of clean marine and continentally-influenced air masses; the matchups with NOAA14 were all low to moderate AOD (0.1-0.35). All are in excellent agreement with the ship-based data (correlation 0.98 or higher, and RMSE 0.03 or lower, depending on wavelength, and 100 % matching within the EE). While encouraging, it is important to emphasise that this is a small number of measurements from a small region and a limited time period, so should not be taken to imply that the performance of the NOAA14 data set is superior to the others.

The NOAA18 matchups are from a broader set of cruises (see *Smirnov et al.*, 2009, 2011) and cover many different regions (Figure 5). Comparison statistics are broadly similar to those for the earlier AVHRR sensors (Table 2) and the AERONET island/coastal sites (Table 1). In particular, the AOD bias tends to become more positive (or less negative), and RMSE to decrease, as wavelength increases. The increased uncertainty at 550 nm is expected since this AOD represents a slight extrapolation beyond the wavelength range of AVHRR measurements, so is not as well-constrained (i.e. it is quite sensitive to the AE, which is assumed rather than retrieved). Nevertheless, all data in Table 2 have $f>0.68$, suggesting the EE may be smaller than the assumed $\pm(0.03+15\,\%)$ over ocean, consistent with results in Table 1 for island/coastal AERONET locations.

4. AOD validation over land

A total of 427 AERONET sites, shown in Figure 6, provided matchups with NOAA18 over land. Due to the larger variety of aerosol sources and sinks over land compared to ocean, as well as the increased heterogeneity of terrain, unevenness of distribution of AERONET sites, and regional rather than global nature of many analyses, regional as well as global statistics are provided in Table 3. The boundaries of these regions are also shown in Figure 6. Their definition is a balance between trying to keep areas with similar aerosol/surface conditions together, and regions frequently used in analyses, balanced by the distribution of the AERONET sites. As such it is inherently somewhat subjective but provides a balance between level of detail, conciseness, and data volume. Figure 7 shows (for the 304 sites providing at least 25 matchups) site-by-site correlation, bias, and fraction matching within the over-land EE of $\pm(0.05+25\,\%)$, and Figure 8 an examination of retrieval error characteristics as a function of AOD and AE.

From Table 3, globally, 69 % of matchups agree with AERONET within the EE at 550 nm, and 74 % for band 1 (633 nm for NOAA18). Globally and regionally, the RMSE tends to be 10-20 % larger at 550 nm compared to band 1, and the AOD bias is less positive (or more negative) at 550 nm than band 1. The AOD bias at both wavelengths (Figure 8) also tends to be small and positive in low-AOD conditions, but more negative (relative bias around -20 %) at high AODs, meaning it is on the lower end of the EE envelope; on average it is small and negative at most sites (between 0 and -0.05; Figure 7). These bias characteristics share similarities with those found over ocean (Section 3.1). Further, the DB algorithm has two methods for modelling land surface reflectance (*Hsu et al.*, 2013, 2017): a method based on NDVI used over vegetated regions, and a surface data base for brighter surfaces (deserts, mountains, urban areas), and similar bias characteristics are found in both (Figure 8). This makes it likely that sensor calibration is a contributing factor, since similar biases are found in both land and ocean algorithms, and for the two different over-land surface reflectance determination methods.

Table 3 and Figure 7 also indicate that there is regional variability in performance. Around half of global matchups are in North America or Europe, due to the density of the AERONET network in these areas, and about 8 % are in the 'boreal' region (mostly tundra or forested regions at high Northern latitudes). At sites in these regions, the DB algorithm tends to perform well, with biases often smaller than 0.025 and more than 68 % of retrievals matching within the EE. These are regions where the NDVI-based surface reflectance determination method predominates. A fairly high quality of performance is also seen in the South America, South Africa, and Oceania regions, although there is some tendency to underestimate AOD in high-AOD conditions. Lower correlations at sites in some of these regions (particularly Oceania) again reflect that the dynamic range of AOD is fairly small compared to the magnitude of retrieval uncertainty.

Performance at tropical sites, particularly in the Sahel, Arabian Peninsula, Indian subcontinent, and eastern Asia is poorer. This is likely due to a combination of the brighter surface (less sensitivity to the aerosol signal, and potentially being near the critical albedo where the TOA signal is invariant with AOD, e.g. *Seidel and Popp*, 2012), high variability in aerosol composition (i.e. single aerosol models and the assumed AE for conversion of band 1 AOD to 550 nm is less appropriate), and higher frequency of cirrus clouds (which are harder to detect in AVHRR than sensors which have bands around 1.37 μm like MODIS/VIIRS). Tropical cirrus cloud contamination is particularly problematic in south-eastern Asia, and can affect Sun photometer data as well as satellite retrievals. *Chew et al.* [2011] examined collocated Sun photometer and lidar data at Singapore and found residual cirrus contamination present in around a third of the Sun photometer; the resulting AOD bias for these cases was around 0.034, which is somewhat larger than the instruments' nominal uncertainty. They also found that the bias induced in the Sun photometer data was larger than the typical bias introduced into satellite AOD from cirrus contamination, so it is possible that the negative biases are in part due to this effect. These regions often perform more poorly than others in over-land AOD retrieval algorithms, so the difficulty is not limited to AVHRR or DB (*Levy et al.*, 2010, *Kahn et al.*, 2010, *Sayer et al.*, 2012b, 2013, 2014, *Reid et al.*, 2013, *Popp et al.*, 2016). This is also reflected in Figure 8, in that uncertainties tend to be slightly larger for bright regions where the data base method was used to estimate surface reflectance.

The available data volume for NOAA14 is an order of magnitude smaller (6,668 matches from 123 sites, 58 of which provided at least 25 matches). As over ocean, this is due to the more limited extent of the AERONET network during the 1995-1999 period. No AERONET sites were active over the OCE region at this time. Figures 9 and 10 characterize the AOD- and site- dependence of validation statistics for NOAA14 respectively, and show the same tendencies as were observed for NOAA18 data in Figures 7 and 8. Table 4 summarises statistics globally and regionally. The regional dependence of these statistics is in general similar to that of NOAA18 (cf. Table 3), although the very limited data volume in some regions makes it more difficult to assess how representative some of these statistics are, particularly in Asia. Figure 10 suggests similar AOD- and type-dependence of retrieval errors, thus it appears as though the over-land data from the two sensors share similar error characteristics on the whole.

For both AVHRR sensors, a summary is that the retrieval tends to perform well in areas with darker (more vegetated) surfaces, and where the aerosol type is not too variable in time. In these cases the biases are small and the retrieval uncertainty is probably better than $\pm(0.05+25\,\%)$, tracking the temporal variability of AOD well but with a tendency to underestimate the AOD of high-AOD events. In more complicated tropical environments, the data should be used with more caution, as there is a greater tendency to underestimate AOD. However the correlation often remains high, suggesting the ability to identify high-AOD events, despite this underestimation. Development of future versions of the AVHRR DB products will therefore focus on better QA-filtering of data in these regions, whether more appropriate aerosol optical models can be found, and development of separate error models for the two different surface reflectance determination methods (NDVI vs. data base) and/or geographic regions. The AOD biases relative to RMSE are larger over land (e.g. Figure 7) than ocean, suggesting that decreasing the bias (which over land could be achieved with improved radiometric calibration and/or surface reflectance determination) could lead to non-negligible decreases in RMSE and shrinking the EE envelope.

The error characteristics for the DB AVHRR data over land also share some common features with validation results from DB applied to SeaWiFS (*Sayer et al.*, 2012b) and MODIS (*Sayer et al.*, 2013, 2014). All show better performance over vegetated than bright land surfaces. This is a consequence of the fact that the aerosol signal is in general comparatively stronger over a vegetated (darker) surface, and the dynamic surface reflectance model employed by DB over such surfaces helps in tracking temporal/directional variations. The similarity in bias characteristics between instruments, however, is harder to explain. As AVHRR lacks bands in the blue spectral region which are key for the SeaWiFS/MODIS applications of DB, errors caused by aerosol optical model assumptions in SeaWiFS/MODIS would not necessarily be expected to be the same. This similarity may therefore be in part coincidental. The fact that AVHRR DB (land) and SOAR (water) AOD biases show similar behavior, despite being independent algorithms, suggests that sensor calibration plays some role in AVHRR's biases.

5. Comparison with other satellite products

5.1. AVHRR over ocean

As discussed in Section 2.2.1, the main other AVHRR data sets available at present are the over-ocean NASA GACP and NOAA CDR products. It is difficult to do a direct three-way comparison between these and SOAR, as there are differences in the available wavelengths (i.e. GACP provides only 550 nm, CDR only band 1, SOAR both) and aggregation levels in both time (GACP provides only monthly, CDR daily and monthly, neither orbit-level) and space (CDR is on a 0.1° grid while GACP and SOAR are at 1°) between the data sets. As a balance, this analysis provides a comparison of seasonal composites for the year 2006 from NOAA18. This year was notable for aerosol events including a strong dust storm in March, intense fires in north-eastern Russia and China in May, and a strong El Niño leading to an intense biomass burning season in Indonesia, peaking in September/October (*Carboni et al.*, 2012, *Marlier et al.*, 2013, *Field et al.*, 2016). A seasonal comparison means that the effects of calibration, sampling, and retrieval algorithm cannot be directly separated, but it allows for a big-picture comparison which is more akin to the way many data users approach these products (i.e. monthly or longer composites), and, as noted, the types of comparison possible are constrained by the available data products.

The CDR product is aggregated to 1° to match the others (cf. Section 2.2.1), and differences in monthly means are calculated before averaging to provide seasonal means and differences. Figure 11 shows the resulting seasonal 550 nm AOD maps from SOAR, as well as difference maps between SOAR and GACP/CDR (comparing SOAR and GACP 550 nm AOD, and then SOAR and CDR band 1 AOD, i.e. comparing common wavelengths in both cases).

It is immediately apparent that the differences between SOAR and GACP, and SOAR and CDR, show contrasting behaviour in several regions (e.g. SOAR is somewhat lower than GACP in the Southern Ocean but somewhat higher than CDR in this region). Part of the difference between the SOAR/GACP and SOAR/CDR comparisons is due to the different wavelengths between the two comparisons (the former pair is 550 nm and the latter 630 nm), although this should be a small effect (<0.02) in most cases since this wavelength difference is not that large. Thus, where the SOAR/GACP and SOAR/CDR comparisons show offsets of opposing signs, it is likely that the difference is dominated by some combination of calibration, algorithm, and sampling, rather than this wavelength difference.

The SOAR AOD is higher than both GACP and CDR in many high-AOD continental outflow regions (e.g. the Saharan dust belt, central African biomass burning, north-eastern Asia). Differences in such regions are expected to be particularly large, because of the limited information content of the sensor and so need to make considerable simplifying assumptions about aerosol optical model (size distribution and refractive index). SOAR picks from one of several bimodal optical models, while both GACP and CDR assume the same aerosol properties for fine and coarse aerosol modes globally. This will lead to larger (systematic) errors in high-AOD conditions (as scattering/absorption properties of marine and dust, smoke, continental, or other aerosol types differ). As GACP and CDR assume a spherical coarse mode (SOAR includes nonspherical dust), further, errors will exhibit a larger angular-dependence in the case of nonspherical dust, and such errors will not necessarily cancel out through averaging to a longer time scale (e.g. *Zhao et al.*, 2004, *Lee et al.*, 2017). Some further analysis of the implications of the aerosol optical model assumptions, as it pertains to the GACP product and long-term trends in particular, is provided by *Mishchenko et al.* [2012]. Such differences in these areas are therefore expected, although it is interesting that SOAR AOD is higher than the others in these cases given that Figure 1 indicates a tendency to underestimate the AOD in high-AOD conditions. Therefore it is possible that GACP/CDR are biased more negatively.

Validation of the GACP product was performed by *Liu et al.* [2004], although this was on a monthly 1° basis as opposed to an instantaneous basis (as performed for SOAR herein), and predated NOAA18's launch. *Geogdzhayez et al.* [2015] did not present additional validation for NOAA18, although noted that there did not appear to be sensor-to-sensor discontinuities between the GACP record from different sensors, by using years where data from multiple overlapping sensors were available. Hence, it is plausible that the bias tendencies of NOAA18 are similar to those found for the earlier sensors by *Liu et al.* [2004], which were ship-based measurements indicating a random error of 0.04 and positive bias around 11 %. In this sense the fact that SOAR AOD (which appears to have a small bias with respect to AERONET/MAN in clean conditions) is higher than GACP is also unusual if the NOAA18 GACP record really does have a positive bias. The version 3 CDR product has not been validated extensively, particularly for NOAA18, although available analyses (*Zhao et al.*, 2004, *Zhao*, 2016) suggest a systematic error at 630 nm in open-ocean condition of order 0.03, and random errors of order 0.11. CDR also allows retrieval of negative AOD (down to -0.2), although unphysical, in an attempt to stop retrieval errors in low-AOD conditions being positively skewed (*Zhao*, 2016). It is not clear how error characteristics are likely to change in areas of high aerosol loading. Due to the small data volume, and monthly rather than instantaneous comparison, it is difficult to disentangle how algorithm and sampling may be combining to cause the observed offsets in low- and high-AOD regions.

Smaller differences in open-ocean conditions may arise from factors such as the relative aggressiveness of cloud screening, both in terms of the risk of cloud contamination, which typically causes high-AOD artifacts, and relative sampling of near-cloud vs. far-from-cloud pixels, the former of which may have real higher AODs due to e.g. aerosol humidification (*Twohy et al.*, 2009, *Várnai et al.*, 2013). *Zhao et al.* [2013] found differences in zonal or monthly mean AVHRR-derived AOD at 630 nm of up to 0.04 dependent upon strictness of cloud masking. Detection of optically-thin cirrus clouds is particularly difficult for the AVHRRs compared to e.g. MODIS as they lack a band near 1.37 μm, which is sensitive to high clouds. Additional regional offsets can be explained by the fact that the GACP algorithm assumes a globally-constant near-surface wind speed of 7 ms^{-1} (*Mishchenko et al.*, 1999) while SOAR uses ancillary meteorological information to calculate the influence of wind speed on surface reflectance for each retrieval. This constant-wind assumption is known to lead to regional offsets in AOD of either sign of order 0.01-0.02, dependent on typical local wind speeds, and also mean that Sun glint can be under- or overscreened (*Zhang and Reid*, 2006, *Sayer et al.*, 2010). The CDR product uses a constant Lambertian albedo (*Zhao*, 2016), which is more-or-less equivalent to a constant wind speed, although does include a Sun-glint contribution as well.

Another notable offset is that SOAR-GACP is quite negative in the Southern Ocean while SOAR-CDR is positive (and somewhat smaller). The phenomenon of high Southern Ocean AOD is found in several satellite data sets (including GACP but not CDR), but not seen in AERONET or MAN and so thought to be partially an artifact. The feature is also seen in Northern storm tracks, but is less prominent due to cloud and land cover. The causes were investigated by *Toth et al.* [2013], with a main focus on MODIS data, who concluded that cloud contamination was responsible for up to 30-40 % but other assumptions (such as a fixed assumed surface wind speed) were responsible for the rest. It therefore seems likely that this conclusion is applicable to the AVHRR products as well. The fact that this is not seen in CDR may suggest that cloud contamination is the larger factor relevant for the AVHRRs, and its absence in CDR an indication of more aggressive cloud masking; following *Zhao et al.* [2013], a fairly strict cloud mask was adopted in the version 3 CDR product.

5.2. MODIS over land and ocean

As many research applications take monthly AOD products as a basis, rather than L2 data, it is instructive to see how similar such composites are between the new AVHRR data set and other commonly-used products such as MODIS. Figure 12 provides such a comparison between NOAA18 AVHRR and MODIS Aqua monthly data (Section 2.2.3), constructed from the overlapping time period of the two sensors (2006-2011). To increase the robustness of statistics only grid cells containing data from at least 24 months are considered, which removes points in areas of high cloud cover (e.g. tropical rainforests) and high latitudes where clouds and polar night strongly limit coverage in some months. Figure 12(a) shows the main global features of AOD are represented in the AVHRR data. Note that as this is a multiannual mean composite, the strength of seasonal features can be attenuated. The other panels provide important contextual information.

Over the open ocean, AVHRR AOD is often lower than MODIS by 0-0.03. This is consistent with AVHRR having a near-zero AOD bias in such conditions (Section 3), and MODIS having a positive bias of order 0.015 on average (*Sayer et al.*, 2012d, *Levy et al.*, 2013). For some grid cells near the Equator a positive offset is seen instead, which may be due to the aforementioned greater difficulty of thin cirrus cloud detection in AVHRR than MODIS. In general over the remote ocean the correlation coefficient varies from 0-0.8, dependent on the precise region. This is because the seasonal variation in AOD is small relative to the retrieval uncertainties (which tend to have a non-negligible systematic component), such that a large correlation is only found in areas with seasonal or periodic continental aerosol transport. AVHRR has a slightly more negative offset at high latitudes, which is consistent with *Toth et al.* [2013] who identified cloud contamination as a probable contributing cause to a high band of AOD in MODIS. The RMS difference is small (0-0.03 over the cleanest ocean regions, 0.03-0.06 over other open oceans), but higher in these storm tracks, likely due again to cloud contamination in MODIS.

AOD is also lower in AVHRR than MODIS over dust aerosol outflow regions of Africa and Asia, consistent both with a slight low bias in AVHRR, and a positive (on average) bias in MODIS due to its lack of nonspherical dust aerosol models (*Levy et al.*, 2003, *Zhang* *and Reid*, 2006, *Banks et al.*, 2017). The correlation in these outflow regions remains high (0.8-1), indicating that both track the same seasonal and interannual variability in dust transport. Over smoke outflow regions East of southern Africa the correlation is similarly high and offset/RMS difference small. In contrast, the RMS difference over the smoke outflow region from southern Africa into the southern Atlantic is larger and correlation lower. Closer examination reveals that this is due to some of this smoke being masked as cloud in the AVHRR data, resulting in it being underrepresented in the monthly composite. This doesn't show in Figure 12(c) because this smoke transport only occurs in a few months of the year.

As over ocean, the regions over land with low correlation between AVHRR and MODIS monthly composites tend to be those with fairly persistent low AODs such as large parts of Australia and mountainous areas of North and South America. In most of these areas the offset and RMS difference between the two sensors tends to be 0-0.03, confirming that the two are consistent in this lack of temporal variation. In contrast the two are highly correlated, with fairly low bias, in smoke source regions in South America and Africa. Intermediate regions (i.e. fairly low AOD but moderate seasonality), such as much of the Americas and Europe, have intermediate correlation and small biases (0-0.03, of either sign). Improving correlation or decreasing RMS in these areas may be difficult as both sensors show fairly small biases with respect to AERONET in these regions, although those of AVHRR are slightly larger (Section 4 and *Sayer et al.*, 2013). Larger offsets and/or RMS differences are found in three land regions:

1. Near the limits of AVHRR spatial coverage around bright deserts. These differences are likely dominated by a combination of AVHRR retrieval error, and differences in spatial coverage.

2. Over high-AOD regions of China. These often have limited sampling due to high cloud cover; available validation suggests MODIS DB has less bias than AVHRR DB. Refinement of seasonal aerosol optical model assumptions may help, although this region has very high spatiotemporal variation in aerosol sources.

3. In central Asia, most notably around Iran and surrounding countries. This difference has been traced to a limitation of the MODIS C6 DB product in this area, which has been fixed for the upcoming Collection 6.1 reprocessing. Future data versions should show a higher level of consistency.

5.3. Time series comparison at AERONET sites

A goal of the Deep Blue aerosol project is to move towards consistency in AOD derived from multiple satellite sensors using similar measurement types and retrieval techniques. As such, this Section examines the AOD time series obtained at selected long-term AERONET sites covering the era to which DB/SOAR have been applied. Although there are several hundred AERONET sites in operation, very few have operated continuously or with few gaps since the mid-1990s, which limits the extent of the comparison. A total of five sites are considered in this analysis. Over ocean these are Capo Verde (Atlantic dust outflow) and Wallops (US East coast continental outflow); there is unfortunately no well-sampled long-term 'clean marine' site covering both the NOAA14 and NOAA18 eras. Over land they are Alta Floresta (Brazilian rainforest with seasonal biomass burning), NASA Goddard Space Flight Center (GSFC, suburban Eastern US) and Mongu (situated in a semi-arid part of Zambia with seasonal biomass burning). Four of these five AERONET sites were identified by *Li et al.* [2016] as providing a moderate or high level of representivity of their surrounding regions on these 1° spatial scales; the other (Wallops) was not evaluated by *Li et al.* [2016]. Thus, although the choice of AERONET sites is strongly constrained by the limited number which have been in operation for much of the period from the mid-1990s until 2011, it is fortunate that these sites appear to sample air masses representative of the spatial scales of satellite level 3 products.

In addition to AERONET and AVHRR, the time series analysis uses the monthly mean MODIS and SeaWiFS data sets described in Section 2.2. The AERONET daily mean product is used to calculate both the monthly mean AOD (for those months with at least 3 days with observations) as well as the central one standard deviation (68 %) range of daily mean AOD, to provide an indication of day-to-day AOD variability within the month and thus indicate those periods where sampling issues are most likely to be important. For all the satellite products, the grid cell in which the AERONET site lies was used to extract the time series.

The resulting mean AOD time series, with the AERONET variability providing a shaded background, are shown in Figures 13 and 14 for the ocean and land sites respectively. The correlation and median bias between AVHRR and the other monthly mean AOD data sets are given in Table 5. The comparison against AERONET here provides an additional look at the validation. Even though SeaWiFS and MODIS are retrievals and not a ground truth like AERONET, the rationale for providing statistics comparing AVHRR to each of these is to assess the level of consistency between the satellite products, which is subtly different than assessing the error in the AVHRR data. Thus these analyses provide different but complementary information. Note that as the NOAA14 time series processed ends in 1999, prior to the launch of the Aqua platform in 2002, there is no comparison between this pair. Both NOAA14 and NOAA18 AVHRR are represented by black lines, but as there is no temporal overlap between the two the provenance of each part of the time series is unambiguous.

The time series all provide similar AOD magnitude and seasonality, and monthly mean values typically lie within the central 68 % range of daily means observed by AERONET for the month in question, which is encouraging. Correlation coefficients range between 0.72 and 0.99 (Table 5), confirming that the seasonal and interannual variability are broadly consistent between AVHRR and the other data sets. Biases are often of similar magnitude between NOAA14 and NOAA18 AVHRR, and in terms of correlation coefficient, there does not appear to be a systematic pattern whereby NOAA14 or NOAA18 is systematically more strongly-correlated with AERONET or the other data sets. However it is important to note that (particularly for NOAA14) the number of overlapping months between data sets is small. Hence, there is inherently likely to be larger uncertainty on these statistics compared with, for example, the instantaneous matchup statistics obtained in validation with AERONET direct-Sun data (see e.g. *Schonbrödt and Perugini*, 2013 for discussions of uncertainties in the estimation of correlation coefficients). Despite the small data volume, the data do suggest the future potential for combining multi-sensor data sets like DB and SOAR to produce a consistent long-term data record, possibly after further bias-correction steps such as have been developed for data assimilation applications (e.g. *Zhang and Reid*, 2006, *Hyer et al.*, 2011, *Schutgens et al.*, 2013).

6. Conclusions

A primary goal of the Deep Blue aerosol project is to be able to create a long-term aerosol data record with broadly consistent error characteristics that is based on the use of satellite sensors with similar measurement capabilities. The approach is to apply similar retrieval algorithms that account for the particular characteristics of each sensor. The feasibility of using the AVHRRs for AOD retrieval over ocean has been established for decades, but existing over-land AOD retrievals proposed for AVHRR have been limited in scope.

This study has established that the DB and SOAR algorithms can be adapted for use with the AVHRR sensors to retrieve AOD over land (aside from snow-covered or very bright desert) and ocean surfaces. As well as providing an over-ocean record with comparable heritage to the other SOAR algorithms, this opens up (for the first time for AVHRR) near-global over-land AOD products on both an instantaneous (i.e. Level 2 orbit-level) and aggregated (Level 3 daily/monthly) basis. The bulk of the available validation data is for NOAA18, although the results indicate a similar quality of performance, to the extent that can be diagnosed, from the earlier NOAA11 and NOAA14 AVHRR instruments as well. This is encouraging in terms of being able to extend these data records back in time, particularly for the new over-land capability. The sparsity of available validation data prior to the mid-1990s will, however, present a challenge for evaluation when the algorithms are applied to the earlier AVHRRs.

The typical level of uncertainty on instantaneous AOD retrieved, which appears to be around $\pm(0.05+25\,\%)$ over land and $\pm(0.03+15\,\%)$ over water, is a little higher than the application of DB/SOAR or similar algorithms to more advanced similar sensors such as

SeaWiFS, MODIS, and VIIRS. This is due to the well-known more limited capabilities of the AVHRR sensors (only two broad reflective solar bands, without on-board calibration). However this should still be sufficient for many quantitative scientific applications, and may be able to be reduced further by refinement of retrieval algorithm and sensor calibration. In particular, AOD time series at long-term AERONET sites examined are well-correlated and typically exhibit small biases with respect to both AERONET and other satellite products. Differences between AVHRR and MODIS AOD data are generally consistent with their known error characteristics, and can hopefully be decreased in future versions. This suggests that the future goal of creating a harmonized data set from multiple sensors, which would be a great advantage for the study of multi-decadal variations in aerosol loading, is achievable. To assess and improve upon the sensor calibration used in the creation of the data set, to further aerosol optical models refine, and to extend DB/SOAR processing to the whole AVHRR record, making use of available validation data and periods of overlap from multiple sensors, are therefore the next steps toward this goal.

Acknowledgements

Further information about Deep Blue, including file formats, documentation, and data download locations for the various data sets, is available at https://deepblue.gsfc.nasa.gov. Data hosting resources were provided by the NASA High-End Computing (HEC) Program through the NASA Center for Climate Simulation (NCCS) at Goddard Space Flight Center; the AVHRR Deep Blue data products are freely available from https://portal.nccs.nasa.gov/datashare/AVHRRDeepBlue. The MERRA data used have been provided by the Global Modeling and Assimilation Office (GMAO) at NASA Goddard Space Flight Center (https://gmao.gsfc.nasa.gov). AERONET and MAN data are available from https://aeronet.gsfc.nasa.gov. The responsible investigators for AERONET sites and MAN cruises are thanked for the creation and stewardship of the Sun photometer data records. GACP data were obtained from https://gacp.giss.nasa.gov/data/time_ser. NOAA AVHRR ocean AOD data were obtained from https://www.ncei.noaa.gov/data/avhrr-aerosol-optical-thickness, and X. Zhao (NOAA) is thanked for discussions about the CDR product. Data processing was facilitated by use of the GNU Parallel utility by *Tange* [2011]. The Editor and three reviewers are thanked for numerous useful comments, which improved the content and clarity of the manuscript.

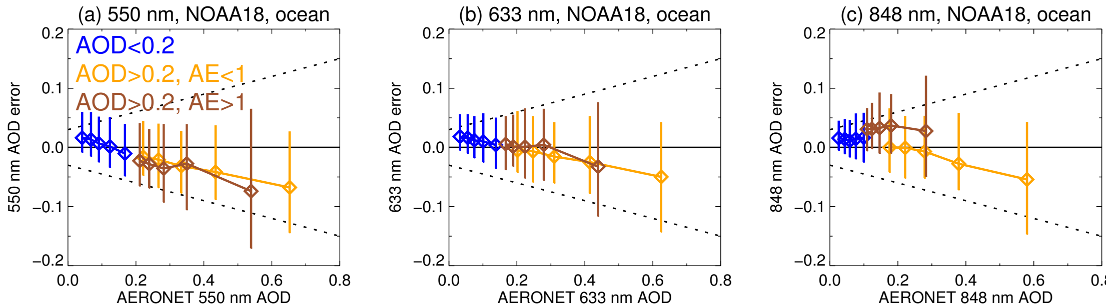

Figure 1. Binned median (points) and central 68 % (lines) over-ocean AOD retrieval error (AVHRR-AERONET) for NOAA18, for (a) 550 nm, (b) band 1, and (c) band 2. Data are split into (blue) AERONET AOD at 550 nm <0.2, (orange) AERONET AOD at 550 nm ≥0.2 and AE<1, and (brown) AERONET AOD at 550 nm ≥0.2 and AE≥1. Matchups within each category are divided into five equally-populated bins. Dashed black lines indicate the EE, ±(0.03+15 %).

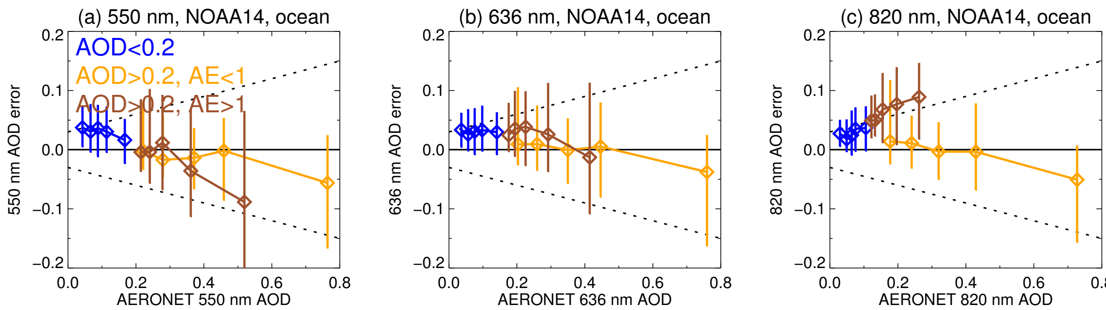

Figure 2. As Figure 1, except for NOAA14.

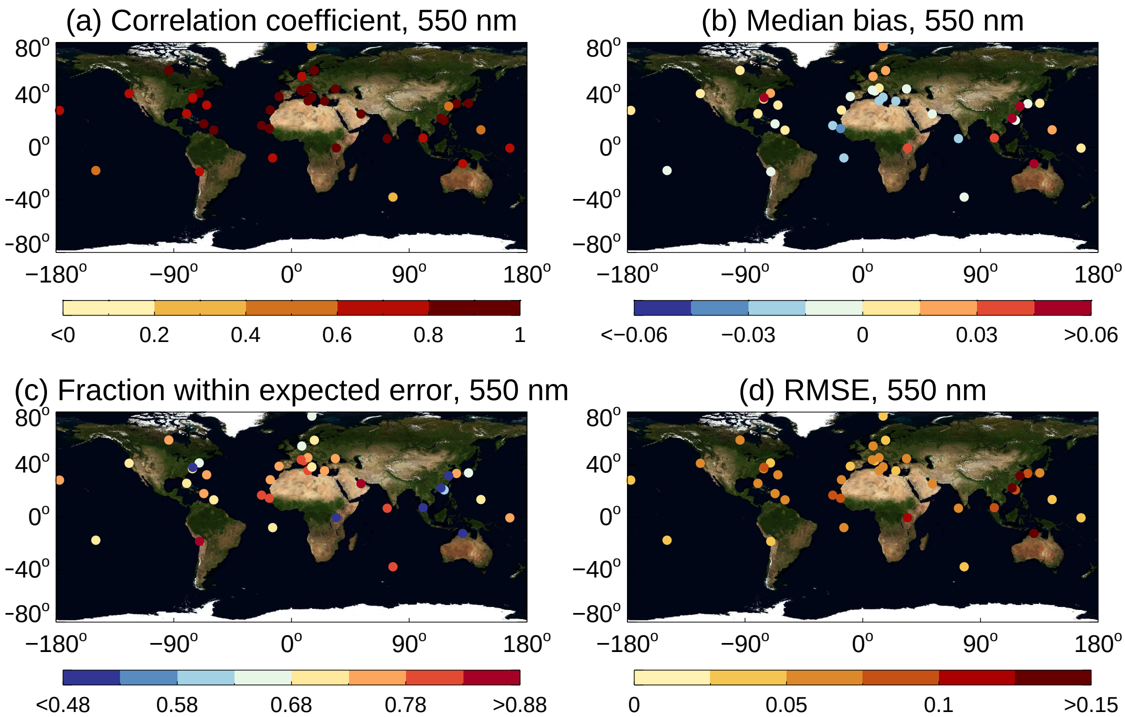

Figure 3. Site-by-site (a) correlation coefficient, (b) median bias, (c) fraction agreeing within the EE, and (d) root mean square error for over-ocean NOAA18 and AERONET matchups at 550 nm.

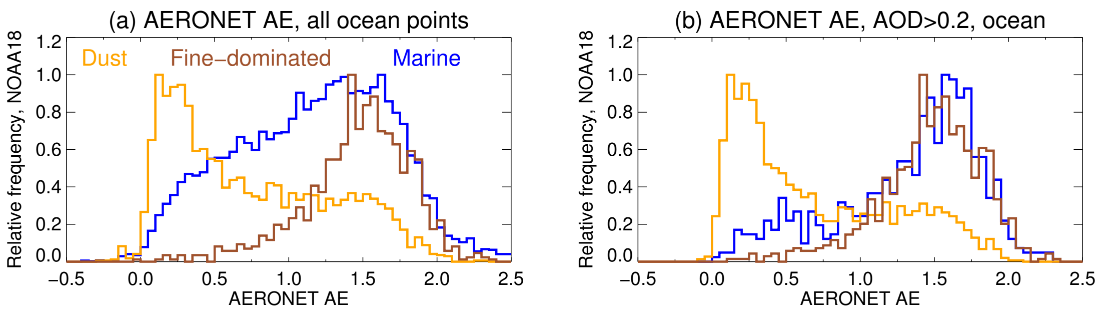

Figure 4. Histograms of AERONET AE, for (a) all ocean matchups with NOAA18, and (b) only NOAA18 matchups where AERONET AOD at 550 nm ≥0.2. Points split to show cases where SOAR chose (blue) maritime, (orange) dust, and (brown) fine-dominated aerosol optical models.

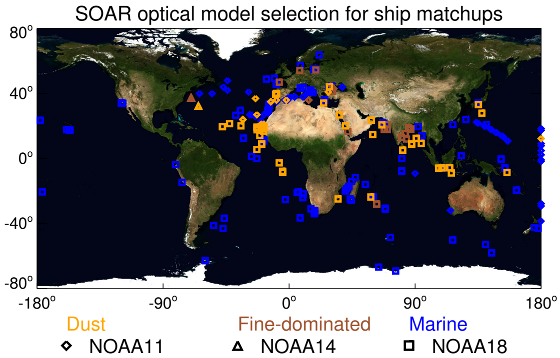

Figure 5. Optical models chosen by SOAR for the AVHRR/ship matchups. Orange indicates matchups where the dust model was chosen, brown the fine-dominated model, and blue the maritime model. Diamonds, triangles, and squares indicate NOAA11, NOAA14, and NOAA18 respectively.

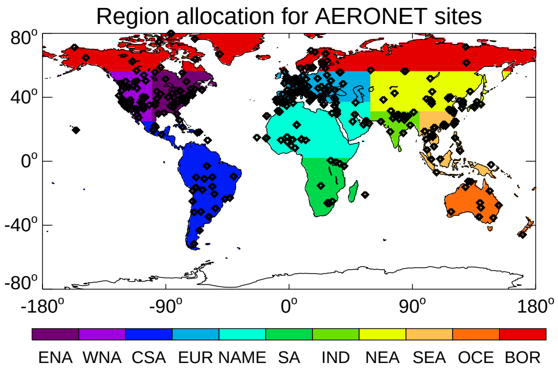

Figure 6. Site locations (black diamonds) and region assignment for over-land NOAA18 DB and AERONET matchups. Regions are boreal (BOR), Eastern North America (ENA), Western North America (WNA), Central/South America (CSA), Europe (EUR), North Africa/Middle East (NAME), Southern Africa (SA), Indian subcontinent (IND), North-Eastern Asia (NEA), South-Eastern Asia (SEA), and Oceania (OCE).

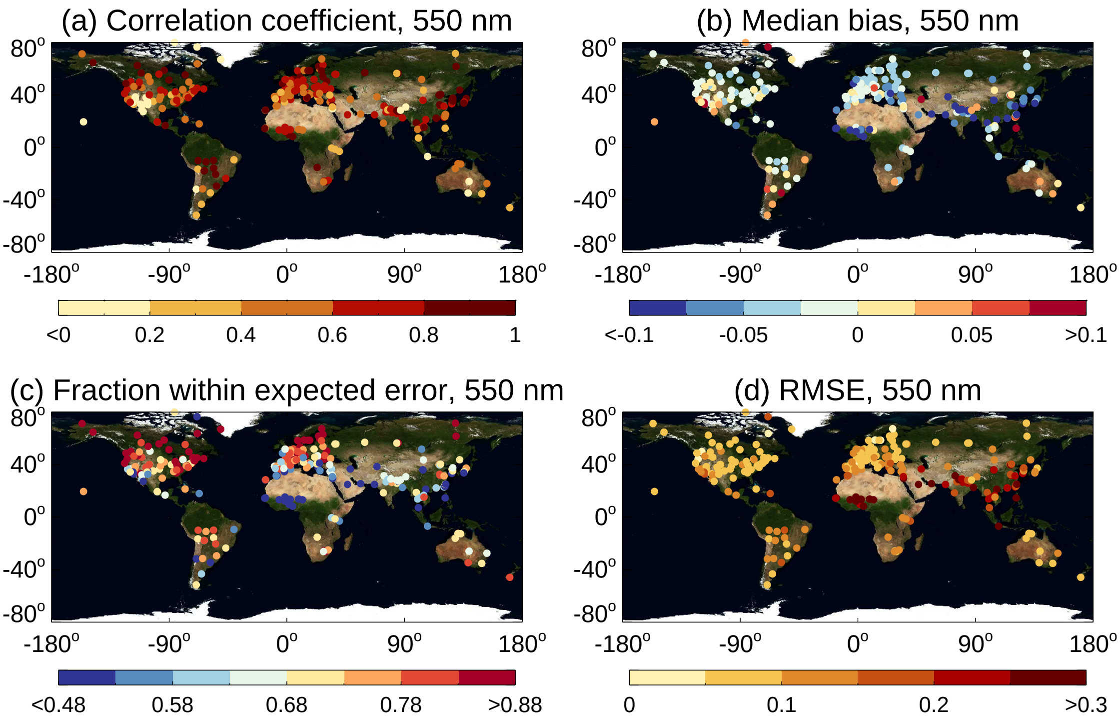

Figure 7. As Figure 3, except for the comparison between NOAA18 and AERONET sites over land, and note different color scale range in panels (b) and (d). Data shown only for sites with at least 25 matchups.

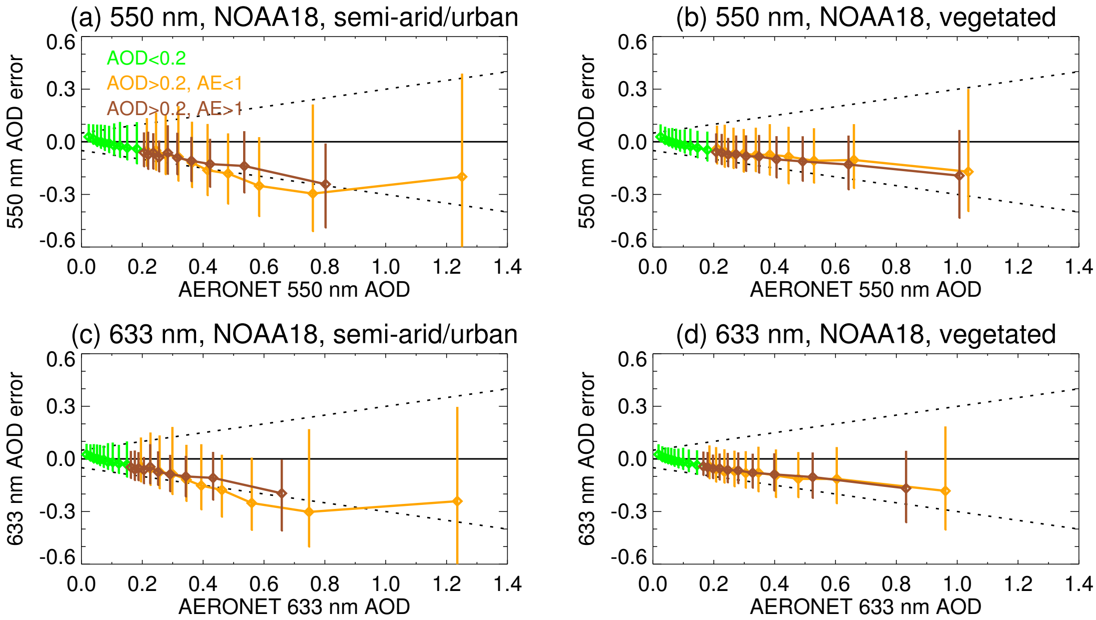

Figure 8. As Figure 1, except for NOAA18 DB matchups over land at 550 nm and band 1, and matchups within each category are divided into 10 equally-populated bins. Note that axis ranges are also different. Data shown separately for (a,c) matchups from semi-arid/urban pixels where the surface reflectance data base method was used and (b,d) vegetated pixels where the NDVI-based surface reflectance model was used. The low-AOD 'background' set are also indicated in green, rather than blue. Dashed black lines indicate the over-land EE, ±(0.05+25 %).

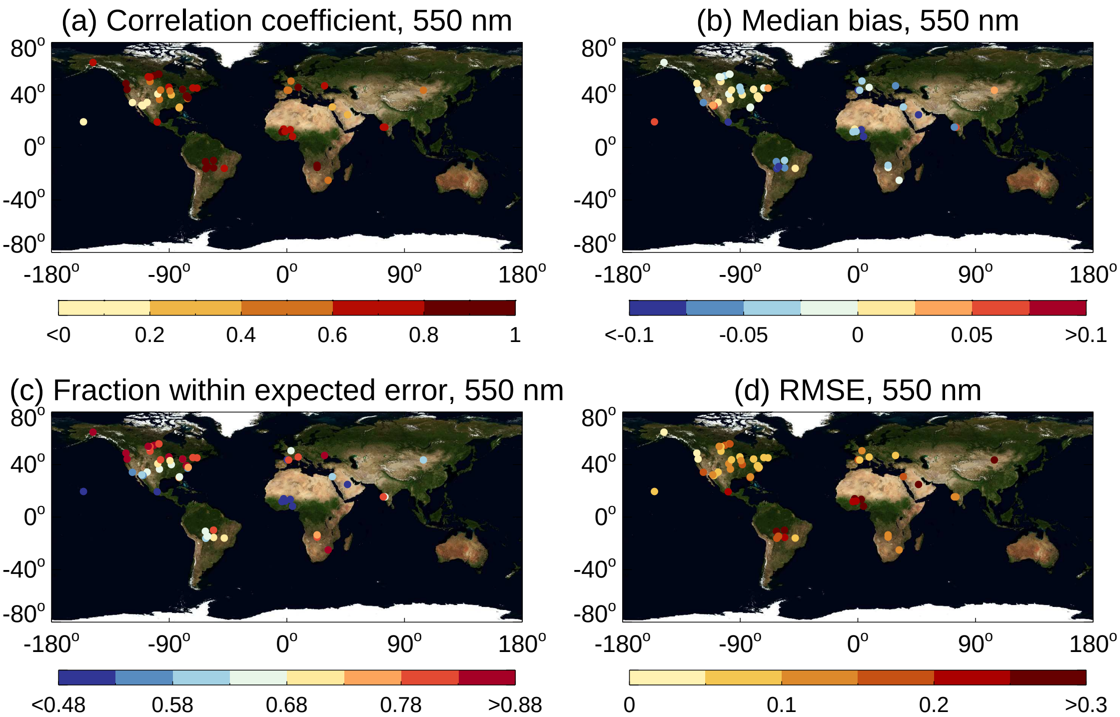

Figure 9. As Figure 7, except for the comparison between NOAA14 and AERONET sites over land.

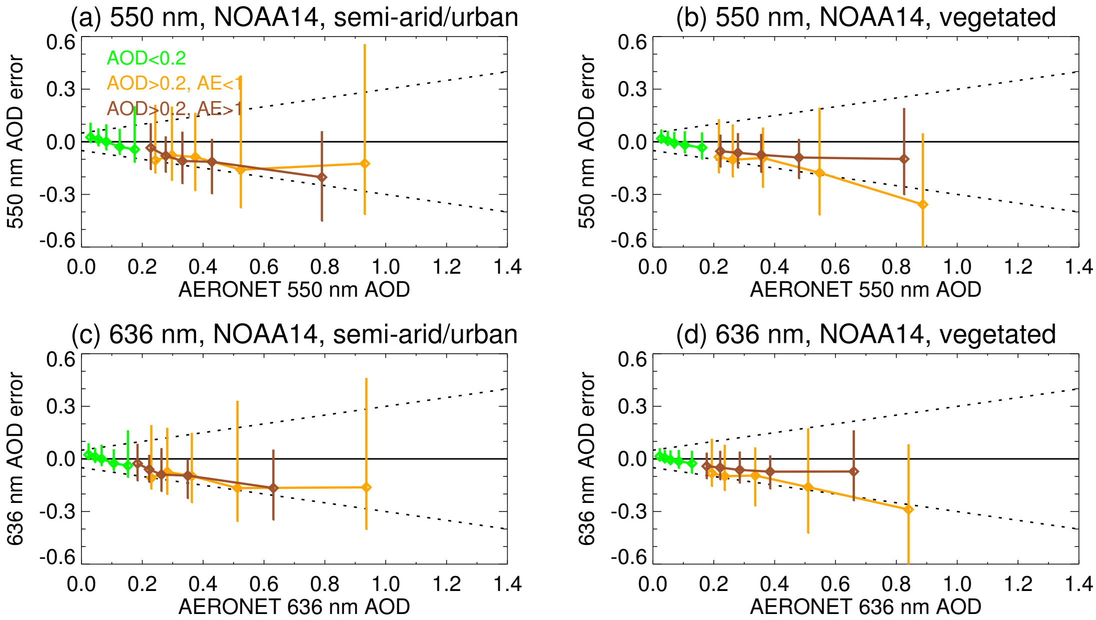

Figure 10. As Figure 8, except for NOAA14 matchups over land, and with half the number of bins in each category due to the smaller data volume.

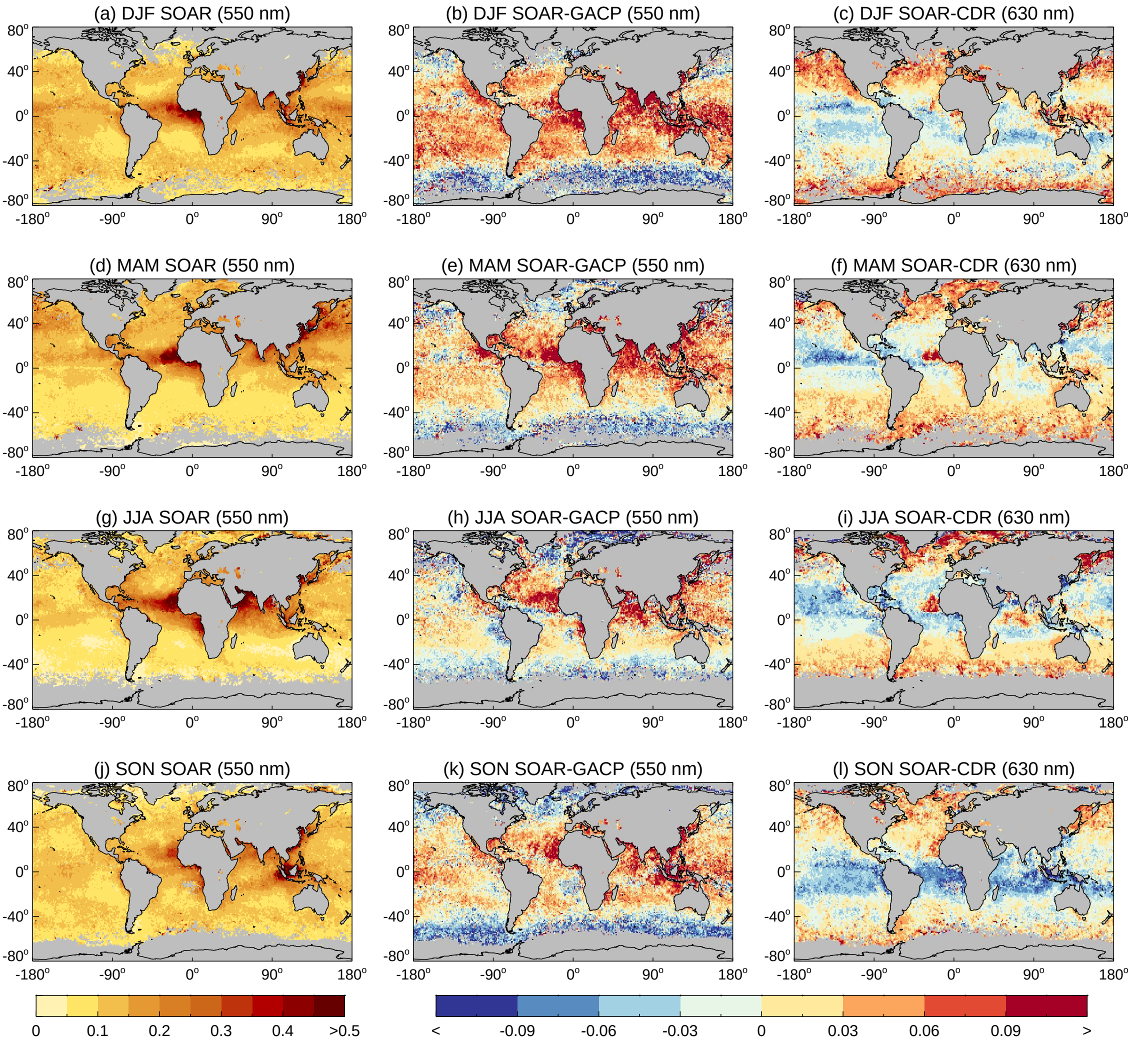

Figure 11. Seasonal composites from the year 2006, of NOAA18 (left column) SOAR 550 nm AOD, (center) SOAR-GACP 550 nm AOD, and (right) SOAR-CDR band 1 AOD. Grid cells without valid data are shaded in grey.

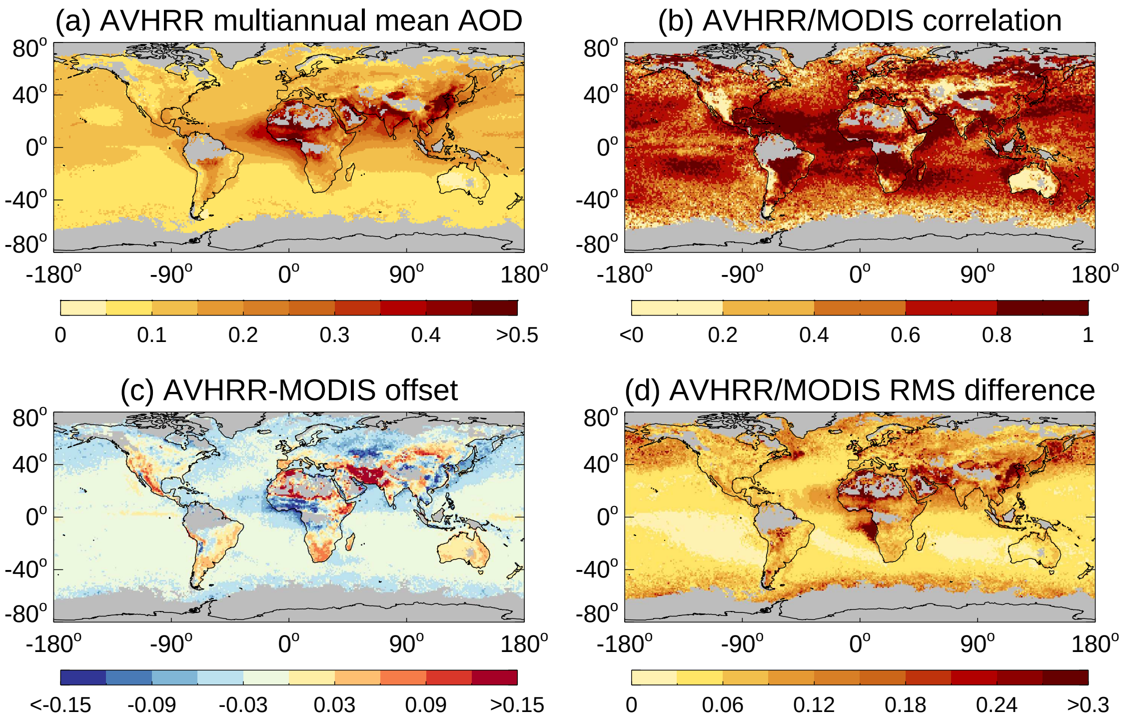

Figure 12. Global statistics of comparison between NOAA18 AVHRR and MODIS Aqua AOD at 500 nm. Panels show (a) multiannual mean AVHRR AOD from matched monthly points, (b) Pearson's correlation coefficient, (c) the median AVHRR-MODIS offset, and (d) the RMS difference between the two. Grid cells without sufficient valid data are shaded in grey.

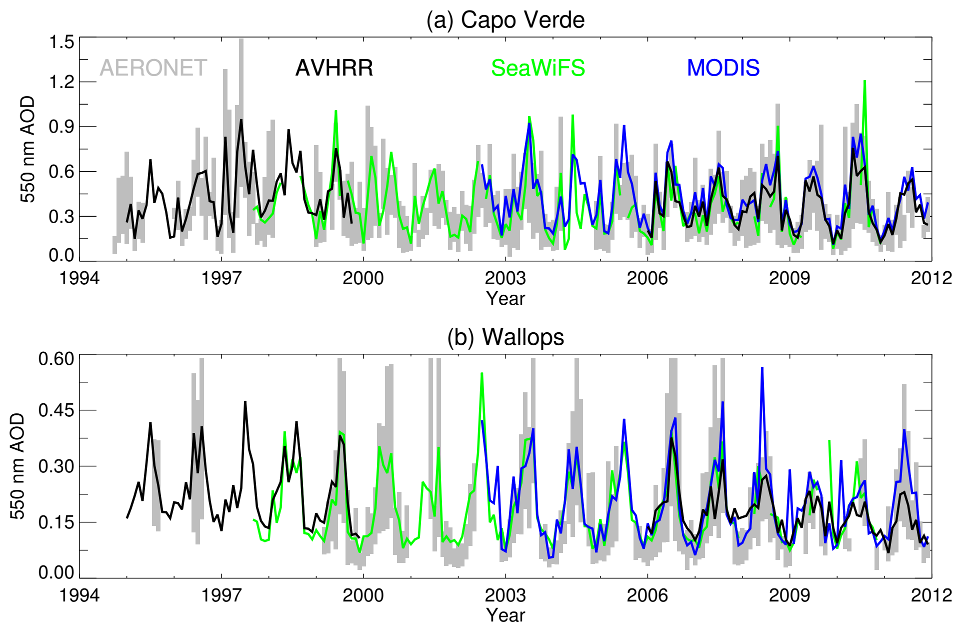

Figure 13. Time series of 550 nm AOD at two long-term coastal/island AERONET sites. The shaded grey area indicates the central 68 % range of AERONET daily mean AOD within a month. Black, green, and blue lines indicate AVHRR (SOAR), SeaWiFS (SOAR), and MODIS Aqua (ocean) retrieved monthly mean AOD respectively.

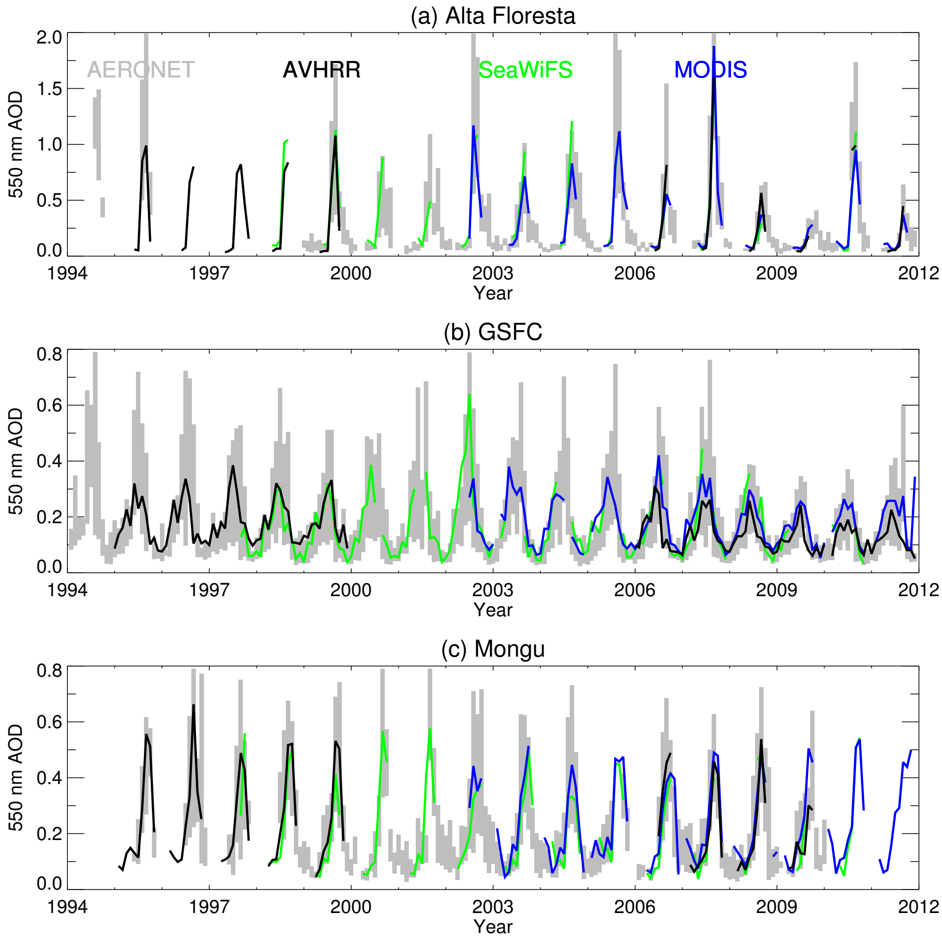

Figure 14. Time series of 550 nm AOD at three long-term land AERONET sites. The shaded grey area indicates the central 68 % range of AERONET daily mean AOD within a month. Black, green, and blue lines indicate AVHRR, SeaWiFS, and MODIS Aqua retrieved monthly mean AOD respectively, in all cases from the DB algorithm.

References

Ahmad, Z., B. A. Franz, C. R. McClain, E. J. Kwiatowska, J. Werdell, E. P. Shettle, and B. N. Holben (2010), New aerosol models for the retrieval of aerosol optical thickness and normalized water-leaving radiances from the SeaWiFS and MODIS sensors over coastal regions and open oceans, *Appl. Opt.*, *49*(29), 5545–5560, doi:10.1364/AO.49.005545.

Banks, J. R., H. E. Brindley, G. Stenchikov, and K. Schepanski (2017), Satellite retrievals of dust aerosol over the Red Sea and the Persian Gulf (2005-2015), *Atmos. Chem. Phys.*, *17*, 3987–4003, doi:10.5194/acp-17-3987-2017.

Carboni, E., G. E. Thomas, A. M. Sayer, R. Siddans, C. A. Poulsen, R. G. Grainger, C. Ahn, D. Antoine, S. Bevan, R. Braak, H. Brindley, S. DeSouza-Machado, J. L. Deuzé, D. Diner, F. Ducos, W. Grey, C. Hsu, O. V. Kalashnikova, R. Kahn, P. R. J. North, C. Salustro, A. Smith, D. Tanré, O. Torres, and B. Veihelmann (2012), Intercomparison of desert dust optical depth from satellite measurements, *Atmos. Meas. Tech.*, *5*, 1973–2002, doi:10.5194/amt-5-1973-2012.

Chew, B. N., J. R. Campbell, J. S. Reid, D. M. Giles, E. J. Welton, S. V. Salinas, and S. C. Liew (2011), Tropical cirrus cloud contamination in sun photometer data, *Atm. Env.*, *45*(37), 6724–6731, doi:10.1016/j.atmosenv.2011.08.017.

Dubovik, O., and M. D. King (2000), A flexible inversion algorithm for retrieval of aerosol optical properties from Sun and sky radiance measurements, *J. Geophys. Res.*, *105.*

Eck, T. F., B. N. Holben, J. S. Reid, O. Dubovik, A. Smirnov, N. T. O'Neill, I. Slutsker, and S. Kinne (1999), Wavelength dependence of the optical depth of biomass burning, urban, and desert dust aerosols, *J. Geophys. Res.*, *104*(D24), 31,333–31,349.

EUMETSAT (2016), Polar Multi-Sensor Aerosol Product: User guide, V2, *Tech. rep.*, EUMETSAT, Darmstadt, Germany, available online at www.eumetsat.int under Data > Technical Documents > GDS Metop > PMAp [Accessed 27 Feb 2017].

Field, R. D., G. van der Werf, T. Fanin, E. J. Fetzer, R. Ruller, H. Jethva, R. Levy, N. J. Livesey, M. Luo, O. Torres, and H. M. Wordern (2016), Indonesian fire activity and smoke pollution in 2015 show persistent nonlinear sensitivity to El Niño-induced drought, *Proc. Natl. Acad. Sci.*, *113*(33), 9204–9209, doi:10.1073/pnas.1524888113.

Gao, L., J. Li, L. Chen, L. Zhang, and A. K. Heidinger (2016), Retrieval and validation of atmospheric aerosol optical depth from AVHRR over China, *IEEE Trans. Geosci. Remote Sens.*, *54*(1), 6280–6291, doi:10.1109/TGRS.2016.2574756.

Geogdzhayez, I. V., M. I. Mishchenko, J. Li, W. B. Rossow, L. Liu, and B. Cairns (2015), Extension and statistical analysis of the GACP aerosol optical thickness record, *Atmos. Res.*, *164-165*, 268–277, doi:10.1016/j.atmosres.2015.05.013.

Hasekamp, O. P., and J. Landgraf (2007), Retrieval of aerosol properties over land surfaces: capabilities of multi-viewing-angle intensity and polarization measurements, *Appl. Opt.*, *46*(16), 3332–3344, doi:10.1364/AO.46.003332.

Holben, B. N., T. F. Eck, I. Slutsker, D. Tanré, J. P. Buis, A. Setzer, E. Vermote, J. A. Reagan, Y. J. Kaufman, T. Nakajima, F. Lavenu, I. Jankowiak, and A. Smirnov (1998), AERONET: A federated instrument network and data archive for aerosol characterization, *Remote Sens. Environ.*, *66*, 1–16, doi:10.1016/S0034-4257(98)00031-5.

Holben, B. N., D. Tanré, A. Smirnov, T. F. Eck, I. Slutsker, N. Abuhassan, W. W. Newcomb, J. S. Schafer, B. Chatenet, F. Lavenu, Y. J. Kaufman, J. Vande Castle, A. Setzer, B. Markham, D. Clark, R. Frouin, R. Halthore, A. Karneli, N. T. O'Neill, C. Pietras, R. T. Pinker, K. Voss, and G. Zibordi (2001), An emerging ground-based aerosol climatology: Aerosol optical depth from AERONET, *J. Geophys. Res.*, *106*(D11), 12,067–12,097, doi:10.1029/2001JD900014.

Holben, B. N., et al. (2017), An overview of meso-scale aerosol processes, comparison and validation studies from DRAGON networks, *Atmos. Chem. Phys. Discuss.*, doi:10.5194/acp-2016-1182.

Hsu, N. C., S.-C. Tsay, M. D. King, and J. R. Herman (2004), Aerosol properties over bright-reflecting source regions, *IEEE Trans. Geosci. Remote Sens.*, *42*(3), 557–569, doi: 10.1109/TGRS.2004.824067.

Hsu, N. C., S.-C. Tsay, M. D. King, and J. R. Herman (2006), Deep Blue retrievals of Asian aerosol properties during ACE-Asia, *IEEE Trans. Geosci. Remote Sens.*, *44*(11), 3180 – 3195, doi:10.1109/TGRS.2006.879540.

Hsu, N. C., M.-J. Jeong, C. Bettenhausen, A. M. Sayer, R. Hansell, C. S. Seftor, J. Huang, and S.-C. Tsay (2013), Enhanced Deep Blue aerosol retrieval algorithm: the second generation, *J. Geophys. Res.*, *118*, 9296–9315, doi:10.1002/jgrd.50712.

Hsu, N. C., J. Lee, A. M. Sayer, N. Carletta, S.-H. Chen, C. J. Tucker, and S.-C. Tsay (2017), Retrieving near-global aerosol loading over land and ocean from AVHRR, *J. Geophys. Res.*, submitted.

Hyer, E. H., J. S. Reid, and J. Zhang (2011), An over-land aerosol optical depth data set for data assimilation by filtering, correction, and aggregation of MODIS Collection 5 optical depth retrievals, *Atmos. Meas. Tech.*, *4*, 379–408, doi:10.5194/amt-4-379-2011.

Jeong, M. -J. and Z. Li (2005), Quality, compatibility, and synergy analyses of global aerosol products derived from the advanced very high resolution radiometer and Total Ozone Mapping Spectrometer, *J. Geophys. Res.*, *110*, D10S08, doi:10.1029/2004JD004647.

Kahn, R. A., B. J. Gaitley, M. J. Garay, D. J. Diner, T. F. Eck, A. Smirnov, and B. N. Holben (2010), Multiangle Imaging SpectroRadiometer global aerosol product assessment by comparison with the Aerosol Robotic Network, *J. Geophys. Res.*, *115*(D23209), doi: 10.1029/2010JD014601.

Kahn, R. A., M. J. Garay, D. L. Nelson, R. C. Levy, M. A. Bull, D. J. Diner, J. V. Martonchik, E. G. Hansen, L. A. Remer, and D. Tanré (2011), Response to 'Toward unified satellite climatology of aerosol properties: 3. MODIS versus MISR versus AERONET', *J. Quant. Spectrosc. Radiative Trans.*, *112*(5), 901–909, doi:10.1016/j.jqsrt.2010.11.001.

Knapp, K. R., and L. L. Stowe (2002), Evaluating the potential for retrieving aerosol optical depth over land from AVHRR Pathfinder atmosphere data, *J. Atmos. Sci.*, *59*, 279–293, doi:10.1175/1520-0469(2002)059<0279:ETPFRA>2.0.CO;2.

Knobelspiesse, K. D., C. Pietras, G. S. Fargion, M. Wang, R. Frouin, M. A. Miller, A. Subramaniam, and W. M. Balch (2004), Maritime aerosol optical thickness measured by handheld Sun photometers, *Remote Sens. Environ.*, *93*(1–2), 87–106, doi:10.1016/j.rse.2004.06.018.

Lambert, A., R. G. Grainger, J. J. Remedio, C. D. Rodgers, M. Corney, and F. W. Taylor (1992), Measurements of the evolution of the Mt. Pinatubo aerosol cloud by ISAMS, *Geophys. Res. Lett.*, *20*(12), 1287–1290, doi:10.1029/93GL00827.

Lee, J., N. C. Hsu, A. M. Sayer, C. Bettenhausen, and P. Yang (2017), AERONET-based non-spherical dust optical models and effects on the VIIRS Deep Blue/SOAR over-water aerosol product, *J. Geophys. Res.*, submitted.

Levy, R. C., L. A. Remer, D. Tanré, Y. J. Kaufman, C. Ichoku, B. N. Holben, J. M. Livingston, P. B. Russell, and H. Maring (2003), Evaluation of the Moderate-Resolution Imaging Spectroradiometer (MODIS) retrievals of dust aerosol over the ocean during PRIDE, *J. Geophys. Res.*, *108*(D19), doi:10.1029/2002JD002460.

Levy, R. C., L. A. Remer, S. Mattoo, E. F. Vermote, and Y. J. Kaufman (2007), Second-generation operational algorithm: Retrieval of aerosol properties over land from inversion of Moderate Resolution Imaging Spectroradiometer spectral reflectance, *J. Geophys. Res.*, *112* (D13211), doi:10.1029/2006JD007811.

Levy, R. C., L. A. Remer, R. G. Kleidman, S. Mattoo, C. Ichoku, R. Kahn, and T. F. Eck (2010), Global evaluation of the Collection 5 MODIS dark-target aerosol products over land, *Atmos. Chem. Phys.*, *10*, 103,999–10,420, doi:10.5194/acp-10-10399-2010.

Levy, R. C., S. Mattoo, L. A. Munchak, L. A. Remer, A. M. Sayer, F. Patadia, and N. C. Hsu (2013), The Collection 6 MODIS aerosol products over land and ocean, *Atmos. Meas. Tech.*, *6*, 2989–3034, doi:10.5194/amt-6-2989-2013.

Li, J., X. Zhao, R. Kahn, M. Mishchenko, L. Remer, K. -H. Lee, M. Wang, I. Laszlo, T. Nakajima, and H. Maring (2009), Uncertainties in satellite remote sensing of aerosols and impact on monitoring its long-term trend: a review and perspective, *Ann. Geophys.*, *27*, 2755–2770, doi: 10.5194/angeo-27-2755-2009.

Li, J., X. Li, B. E. Carlson, R. A. Kahn, A. A. Lacis, O. Dubovik, and T. Nakajima (2016), Reducing multi-sensor satellite monthly mean aerosol optical depth uncertainty part I: Objective assessment of current AERONET locations, *J. Geophys. Res.*, *121*, doi:10.1002/2016JD025469.

Liu, L., M. I. Mishchenko, I. G. Geogdhzhayev, A. Smirnov, S. M. Sakerin, D. M. Kabanov, and O. A. Ershov (2004), Global validation of two-channel AVHRR aerosol optical thickness retrievals over the oceans, *J. Quant. Spect. Rad. Trans.*, *88* (1-3), 97–109, doi: 10.1016/j.jqsrt.2004.03.031.

Lyapustin, A., Y. Wang, I. Laszlo, R. Kahn, S. Korkin, L. Remer, R. Levy, and J. S. Reid (2011), Multiangle implementation of atmospheric correction (MAIAC): 2. Aerosol algorithm, *J. Geophys. Res.*, *116* (D03211), doi:10.1029/2010JD014986.

Marlier, M. E., R. S. DeFries, A. Voulgarakis, P. L. Kinney, J. T. Randerson, D. T. Shindell, Y. Chen, and G. Faluvegi (2013), El Niño and health risks from landscape fire emissions in southeast Asia, *Nature Clim. Change*, *3*, 131–136, doi:10.1038/nclimate1658.

Mei, L. L., Y. Xue, A. A. Kokhanovsky, W. von Hoyningen-Huene, G. de Leeuw, and J. P. Burrows (2014), Retrieval of aerosol optical depth over land surfaces from AVHRR data, *Atmos. Meas. Tech.*, *7*, 2411–2420, doi:10.5194/amt-7-2411-2014.

Mishchenko, M. I., and I. V. Geogdzhayev (2007), Satellite remote sensing reveals regional tropospheric aerosol trends, *Opt. Express*, *15*(12), 7423–7438, doi:10.1364/OE.15.007423.

Mishchenko, M. I., I. V. Geogdzhayez, B. Cairns, W. B. Rossow, and A. A. Lacis (1999), Aerosol retrievals over the ocean by use of channels 1 and 2 AVHRR data: sensitivity analysis and preliminary results, *Appl. Opt.*, *38*(36), 7325–7341, doi:10.1364/AO.38.007325.

Mishchenko, M. I., L. Liu, I. V. Geogdzhayev, J. Li, B. E. Carlson, A. A. Lacis, B. Cairns, and L. D. Travis (2012), Aerosol retrievals from channel-1 and -2 AVHRR radiances: Long-term trends updated and revisited, *J. Quant. Spect. Rad. Trans.*, *113*(15), 1974–1980, doi:10.1016/j.jqsrt.2012.05.006.

O'Neill, N. T., T. F. Eck, A. Smirnov, B. N. Holben, and S. Thulasiraman (2003), Spectral discrimination of coarse and fine mode optical depth, *J. Geophys. Res.*, *108*(D17), 4559–4573, doi:10.1029/2002JD002975.

Popp, T., et al. (2016), Development, production and evaluation of aerosol climate data records from European satellite observations (Aerosol_cci), *Remote Sens.*, *8*(5), doi:10.3390/rs8050421.

Reid, J. S., E. J. Hyer, R. S. Johnson, B. N. Holben, R. J. Yokelson, J. Zhang, J. R. Campbell, S. A. Christopher, D. G. L., L. Giglio, R. E. Holz, C. Kearney, J. Miettinen, E. A. Reid, F. J. Turk, J. Wang, P. Xian, G. Zhao, R. Balasubramanian, B. N. Chew, S. Janjai, N. Lagrosas, P. Lestari, N. H. Lin, M. Mahmud, A. Nguyen, B. Norris, N. T. K. Oanh, M. Oo, S. V. Salinas, E. J. Welton, and S. C. Liew (2013), Observing and understanding the Southeast Asian aerosols system by remote sensing: an initial review and analysis for the Seven Southeast Asian Studies (7 SEAS) program, *Atmos. Res.*, *122*, 303–468, doi:10.1016/j.atmosres.2012.06.005.

Riffler, M., C. Popp, A. Hauser, F. Fontana, and S. Wunderle (2010), Validation of a modified AVHRR aerosol optical depth retrieval algorithm over Central Europe, *Atmos. Meas. Tech.*, *3*, 1255–1270, doi:10.5194/amt-3-1255-2010.

Sayer, A. M., G. E. Thomas, and R. G. Grainger (2010), A sea surface reflectance model for (A)ATSR, and application to aerosol retrievals, *Atmos. Meas. Tech.*, *3*, 813–838, doi: 10.5194/amt-3-813-2010.

Sayer, A. M., N. C. Hsu, C. Bettenhausen, Z. Ahmad, B. N. Holben, A. Smirnov, G. E. Thomas, and J. Zhang (2012a), SeaWiFS Ocean Aerosol Retrieval (SOAR): Algorithm, validation, and comparison with other data sets, *J. Geophys. Res.*, *117*(D03206), doi:10.1029/2011JD016599.

Sayer, A. M., N. C. Hsu, C. Bettenhausen, M.-J. Jeong, B. N. Holben, and J. Zhang (2012b), Global and regional evaluation of over-land spectral aerosol optical depth retrievals from Sea-WiFS, *Atmos. Meas. Tech.*, *5*, 1761–1778, doi:10.5194/amt-5-1761-2012.

Sayer, A. M., A. Smirnov, N. C. Hsu, and B. N. Holben (2012c), A pure marine aerosol model, for use in remote sensing applications, *J. Geophs. Res.*, *117*(D05213), doi:10.1029/2011JD016689.

Sayer, A. M., A. Smirnov, N. C. Hsu, L. A. Munchak, and B. N. Holben (2012d),Estimating marine aerosol particle volume and number from Maritime Aerosol Network data, *Atmos. Chem. Phys.*, *12*, 8889–8909, doi:10.5194/acp-12-8889-2012.

Sayer, A. M., N. C. Hsu, C. Bettenhausen, and M.-J. Jeong (2013), Validation and uncertainty estimates for MODIS Collection 6 "Deep Blue" aerosol data, *J. Geophys. Res.*, *118*, 7864–7872, doi:10.1002/jgrd.50600.

Sayer, A. M., L. A. Munchak, N. C. Hsu, R. C. Levy, C. Bettenhausen, and M.-J. Jeong (2014), MODIS Collection 6 aerosol products: Comparison between Aqua's e-Deep Blue, Dark Target, and merged data sets, and usage recommendations, *J. Geophys. Res.*, *119*(1396513989), doi: 10.1002/2014JD022453.

Sayer, A. M., N. C. Hsu, C. Bettenhausen, R. E. Holz, J. Lee, G. Quinn, and P. Veglio (2017), Cross-calibration of S-NPP VIIRS moderate-resolution reflective solar bands against MODIS Aqua over dark water scenes, *Atmos. Meas. Tech.*, *10*, 1425–1444, doi:10.5194/amt-2016-238.

Schonbrödt, F. D., and M. Perugini (2013), At what sample size do correlations stabilize?, *J. Res. Person.*, *47*(5), 609–612, doi:10.1016/j.jrp.2013.05.009.

Schutgens, N. A. J., M. Nakata, and T. Nakajima (2013), Validation and empirical correction of MODIS AOT and AE over ocean, *Atmos. Meas. Tech.*, *6*, 2455–2475, doi:10.5194/amt-6-2455-2013.

Seidel, F. C., and C. Popp (2012), Critical surface albedo and its implications to aerosol remote sensing, *Atmos. Meas. Tech.*, *5*, 1653–1665, doi:10.5194/amt-5-1653-2012.

Smirnov, A., Y. Villevalde, N. T. O'Neill, A. Royer, and A. Tarussov (1995a), Aerosol optical depth over the oceans: analysis in terms of synoptic air mass types, *J. Geophys. Res.*, *100*(D8), 16,639–16,650.

Smirnov, A., O. Yershov, and Y. Villevalde (1995b), Measurement of aerosol optical depth in the Atlantic Ocean and Mediterranean Sea, in *Proceedings of SPIE, 2582*, pp. 203–214, doi: 10.1117/12.228530.

Smirnov, A., B. N. Holben, T. F. Eck, O. Dubovik, and I. Slutsker (2000a), Cloud-screening and quality control algorithms for the AERONET database, *Remote Sens. Environ.*, *73*(3), 337–349.

Smirnov, A., B. N. Holben, O. Dubovik, N. T. O'Neill, L. A. Remer, T. F. Eck, I. Slutsker, and D. Savoie (2000b), Measurement of atmospheric optical properties on U.S. Atlantic coast sites, ships, and Bermuda during TARFOX, *J. Geophys. Res.*, *105*(D8), 9887–9901, doi: 10.1029/1999JD901067.

Smirnov, A., B. N. Holben, Y. J. Kaufman, O. Dubovik, T. F. Eck, I. Slutsker, C. Pietras, and R. H. Halthore (2002), Optical properties of atmospheric aerosol in maritime environments, *J. Atmos. Sci.*, *59*, 501–523.

Smirnov, A., B. N. Holben, I. Slutsker, D. M. Giles, C. R. McCLain, T. F. Eck, S. M. Sakerin, A. Macke, P. Croot, G. Zibordi, P. K. Quinn, J. Sciare, S. Kinne, M. Harvey, T. J. Smyth, S. Piketh, T. Zielinski, A. Proshuninsky, J. I. Goes, N. B. Nelson, P. Larouche, V. F. Radionov, P. Goloub, K. K. Moorthy, R. Matarresse, E. J. Robertson, and F. Jourdin (2009), Maritime aerosol network as a component of aerosol robotic network, *J. Geophys. Res.*, *112*(D06204), doi:10.1029/2008JD011257.

Smirnov, A., B. N. Holben, D. M. Giles, I. Slutsker, N. T. O'Neill, T. F. Eck, A. Macke, P. Croot, Y. Courcoux, S. M. Sakerin, T. J. Smyth, T. Zielinski, G. Zibordi, J. I. Goes, M. J. Harvey, P. K. Quinn, N. B. Nelson, V. F. Radionov, C. M. Duarte, R. Losno, J. Sciare, K. J. Voss, S. Kinne, N. R. Nalli, E. Joseph, K. Krishna Moorthy, D. S. Covert, S. K. Gulev, G. Milinevsky, P. Larouche, S. Belanger, E. Horne, M. Chin, L. A. Remer, R. A. Kahn, J. S. Reid, M. Schulz, C. L. Heald, J. Zhang, K. Lapina, R. G. Kleidman, J. Griesfeller, B. J. Gaitley, Q. Tan, and T. L. Diehl (2011), Maritime Aerosol Network as a component of AERONET-first results and comparison with global aerosol models and satellite retrievals, *Atmos. Meas. Tech.*, *4*, 583–597, doi:10.5194/amt-4-583-2011.

Stowe, L., A. Ignatov, and R. Singh (1997), Development, validation, and potential enhancements to the second-generation operational aerosol product at NOAA/NESDIS, *J. Geophys. Res.*, *102*(D14), 16,923–16,934.

Tange, O. (2011), GNU Parallel - the command-line power tool, *;login: The USENIX Magazine*, pp. 42–47.

Tanré, D., B. N. Holben, and Y. J. Kaufman (1992), Atmospheric correction algorithm for NOAA-AVHRR products: theory and application, *IEEE Trans. Geosci. Remote Sens.*, *30*(2), 231–248, doi:10.1109/36.134074.

Tanré, D., Y. J. Kaufman, M. Herman, and S. Mattoo (1997), Remote sensing of aerosol properties over oceans using the MODIS/EOS spectral radiances, *J. Geophys. Res.*, *102*(D14), 16,971–16,988, doi:10.1029/96JD03437.

Toth, T. D., J. Zhang, J. R. Campbell, J. S. Reid, Y. Shi, R. S. Johnson, A. Smirnov, M. A. Vaughan, and D. M. Winker (2013), Investigating enhanced Aqua MODIS aerosol optical depth retrievals over the mid-to-high latitude Southern Oceans through intercomparison with co-located CALIOP, MAN, and AERONET data sets, *J. Geophys. Res. Atmos.*, *118*, 4700–4714, doi:10.1002/jgrd.50311.

Twohy, C. H., J. A. Coakley Jr., and W. R. Tahnk (2009), Effect of changes in relative humidity on aerosol scattering near clouds, *J. Geophys. Res.*, *114*(D0505), doi:10.1029/2008JD010991.

Várnai, T., A. Marshak, and W. Yang (2013), Multi-satellite aerosol observations in the vicinity of clouds, *Atmos. Chem. Phys.*, *13*, 3899–3908, doi:10.5194/acp-13-3899-2013.

Vermote, E. and Y. J. Kaufman (1995), Absolute calibration of AVHRR visible and near-807 infrared channels using ocean and cloud views, *Int. J. Remote Sens.*, *16*(13), 2317–2340, doi: 10.1080/01431169508954561.

Villevalde, Y. V., A. Smirnov, N. T. O'Neill, S. P. Smyshlyaev, and V. V. Yakovlev (1994), Measurements of aerosol optical depth in the Pacific Ocean and the North Atlantic, *J. Geophys. Res.*, *99*(D10), 20,983–20,988, doi:10.1029/94JD01618.

von Hoyningen-Huene, W., J. Yoon, M. Vountas, L. G. Istomina, G. Rohen, T. Dinter, A. A. Kokhanovsky, and J. P. Burrows (2011), Retrieval of spectral aerosol optical thickness over land using ocean color sensors MERIS and SeaWiFS, *Atmos. Meas. Tech.*, *4*, 151–171, doi:10.5194/amt-4-151-2011.

Wagner, F., and A. M. Silva (2008), Some considerations about Ångström exponent distributions, *Atmos. Chem Phys.*, *8*, 481–489, doi:10.5194/acp-8-481-2008.

Zhang, J., and J. S. Reid (2006), MODIS aerosol product analysis for data assimilation: Assessment of over-ocean level 2 aerosol optical thickness retrievals, *J. Geophys. Res.*, *111*(D22207), doi:10.1029/2005JD006898.

Zhao, T. X.-P., O. Dubovik, A. Smirnov, B. N. Holben, J. Sapper, C. Pietras, K. J. Voss, and R. Frouin (2004), Regional evaluation of an advanced very high resolution radiometer (AVHRR) two-channel aerosol retrieval algorithm, *J. Geophys. Res.*, *109*(D02204), doi:10.1029/2003JD003817.

Zhao, T. X.-P., P. K. Chan, and A. K. Heidinger (2013), A global survey of the effect of cloud contamination on the aerosol optical thickness and its long-term trend derived from operational AVHRR satellite observations, *J. Geophys. Res. Atmos.*, *118*, 2849–2857, doi:10.1002/jgrd.50278.

Zhao, X. (2016), Climate Data Record (CDR) Program Climate Algorithm Theoretical Basis Document (C-ATBD) AVHRR Aerosol Optical Thickness (AOT), *Tech. rep.*, NOAA, report number CDRP-ATBD-0096 revision 3, available online from https://www.ncei.noaa.gov/data/avhrr-aerosol-optical-thickness/access/doc/ [Accessed March 7 2017].

Zhao, X., A. K. Heidinger, and A. Walther (2016), Climatology analysis of aerosol effect on marine water cloud from long-term satellite climate data records, *Remote Sens.*, *8*(4), 300, doi:10.3390/rs8040300.

Table 1. Statistics of validation between AVHRR and AERONET AOD measurements for SOAR over-water retrievals; n denotes the number of points, R Pearson's correlation coefficient, f the fraction matching within the EE, and RMSE the root mean square error. The bias is the median AVHRR-AERONET bias. Statistics are given separately for 550 nm and AVHRR bands 1 and 2 (columns labelled 550, 630, 830 respectively).

Satellite	n	R			Bias			f			RMSE		
		550	630	830	550	630	830	550	630	830	550	630	830
NOAA14	1,227	0.92	0.94	0.94	0.022	0.027	0.030	0.64	0.64	0.61	0.071	0.065	0.064
NOAA18	13,412	0.86	0.88	0.90	0.0002	0.009	0.014	0.73	0.74	0.72	0.088	0.076	0.061

Table 2. As Table 1, except for the comparison between SOAR AVHRR retrievals and ship-based AOD measurements.

Satellite	n	R			Bias			f			RMSE		
		550	630	830	550	630	830	550	630	830	550	630	830
NOAA11	80	0.70	0.67	0.61	0.011	0.012	0.016	0.76	0.76	0.79	0.081	0.080	0.077
NOAA14	20	0.98	0.99	0.99	-0.017	-0.004	0.003	1.0	1.0	1.0	0.029	0.021	0.019
NOAA18	252	0.92	0.93	0.92	-0.019	-0.014	-0.007	0.79	0.82	0.83	0.071	0.056	0.047

Table 3. Statistics of validation between NOAA18 AVHRR and AERONET AOD measurements for DB over-land retrievals, globally and by region (as indicated in Figure 6). Statistics are defined as in Table 1, given separately for 550 nm and AVHRR band 1 (columns labelled 550 and 630 respectively).

Region	n	R		Bias		f		RMSE	
		550	630	550	630	550	630	550	630
Global	89,104	0.80	0.81	-0.014	-0.010	0.69	0.74	0.15	0.13
BOR	7,155	0.86	0.85	-0.016	-0.010	0.81	0.87	0.073	0.062
ENA	11,582	0.65	0.64	-0.010	-0.006	0.79	0.84	0.087	0.070
WNA	11,080	0.48	0.47	0.008	0.009	0.66	0.70	0.11	0.094
CSA	5,745	0.90	0.89	0.015	0.014	0.66	0.70	0.12	0.10
EUR	26,319	0.63	0.63	-0.018	-0.013	0.74	0.78	0.099	0.082
NAME	10,451	0.73	0.74	-0.052	-0.045	0.47	0.50	0.28	0.27
SA	2,277	0.70	0.67	-0.021	-0.015	0.72	0.77	0.12	0.098
IND	3,346	0.79	0.79	-0.058	-0.050	0.68	0.71	0.19	0.17
NEA	6,483	0.86	0.86	-0.039	-0.031	0.65	0.68	0.21	0.18
SEA	2,402	0.70	0.69	-0.046	-0.035	0.62	0.65	0.22	0.18
OCE	2,264	0.37	0.36	-0.002	-0.001	0.74	0.78	0.089	0.074

Table 4. As Table 3, except for NOAA14 AVHRR over land, and the OCE row is omitted due to a lack of sites in this region during the 1995-1999 time period.

Region	n	R		Bias		f		RMSE	
		550	630	550	630	550	630	550	630
Global	6,668	0.84	0.82	-0.010	-0.009	0.71	0.74	0.17	0.16
BOR	284	0.66	0.64	-0.004	-0.002	0.92	0.94	0.051	0.042
ENA	2,153	0.81	0.77	0.0003	-0.001	0.78	0.82	0.12	0.11
WNA	1,132	0.77	0.73	-0.002	-0.002	0.77	0.80	0.094	0.082
CSA	628	0.91	0.89	-0.041	-0.031	0.62	0.66	0.24	0.23
EUR	583	0.78	0.77	-0.033	-0.031	0.72	0.76	0.11	0.095
NAME	945	0.73	0.74	-0.067	-0.062	0.42	0.43	0.30	0.28
SA	622	0.86	0.83	-0.029	-0.026	0.78	0.82	0.13	0.11
IND	99	0.76	0.76	0.017	0.006	0.72	0.76	0.12	0.098
NEA	209	0.54	0.55	0.017	0.018	0.62	0.64	0.24	0.21
SEA	13	0.51	0.45	0.013	0.009	0.69	0.85	0.15	0.15

Table 5. Statistics of multi-sensor time series comparison for locations shown in Figures 13 and 14. Columns show the correlation coefficient R and median (AVHRR-other) bias at each location, separately for NOAA14 and NOAA18 AVHRR, between monthly mean 550 nm AOD. Note the NOAA14 and MODIS Aqua time series do not overlap.

Statistic	R						Bias					
Comparison	AERONET		SeaWiFS		MODIS		AERONET		SeaWiFS		MODIS	
NOAA platform	14	18	14	18	14	18	14	18	14	18	14	18
Ocean sites												
Capo Verde	0.85	0.91	0.72	0.84	-	0.94	0.032	0.026	0.070	0.036	-	-0.036
Wallops	0.95	0.84	0.89	0.76	-	0.75	0.040	0.020	0.039	0.012	-	-0.006
Land sites												
Alta Floresta	0.91	0.99	0.98	0.95	-	0.96	-0.063	-0.029	-0.057	0.005	-	-0.016
GSFC	0.81	0.77	0.91	0.75	-	0.74	0.014	-0.009	0.033	-0.004	-	-0.028
Mongu	0.92	0.93	0.88	0.87	-	0.90	-0.006	-0.024	0.046	0.030	-	-0.032