AGU ADVANCING EARTH AND SPACE SCIENCE

Earth and Space Science

RESEARCH ARTICLE
10.1029/2020EA001290

How Long Is Too Long? Variogram Analysis of AERONET Data to Aid Aerosol Validation and Intercomparison Studies

Andrew M. Sayer[1,2]

[1]Universities Space Research Association, Greenbelt, MD, USA, [2]Ocean Ecology Laboratory, NASA Goddard Space Flight Center, Greenbelt, MD, USA

Key Points:
- Variograms can characterize size and time scales of aerosol variation
- These can inform matchup thresholds in data validation and intercomparisons
- Biomass burning regions tend to show fastest aerosol variation

Correspondence to:
A. M. Sayer,
andrew.sayer@nasa.gov

Citation:
Sayer, A. M. (2020). How long is too long? Variogram analysis of AERONET data to aid aerosol validation and intercomparison studies. *Earth and Space Science*, *7*, e2020EA001290. https://doi.org/10.1029/2020EA001290

Received 5 JUN 2020
Accepted 31 AUG 2020
Accepted article online 14 SEP 2020

Abstract Geophysical data sets derived from satellite sensors, ground/airborne instrumentation, and computational models are often compared against each other. A common example is the validation of satellite aerosol optical depth (AOD) retrievals against measurements from Aerosol Robotic Network (AERONET) Sun photometers. Spatiotemporal mismatch between data set sampling means that uncaptured variation in the underlying geophysical field introduces apparent disagreement into such comparisons, known as representation or collocation matchup uncertainty. This study uses variogram analysis of AERONET data to estimate temporal mismatch uncertainties and decorrelation time scales for the global AERONET record. As well as total AOD, the fine- and coarse-mode AODs, Ångström Exponent (AE), and fine-mode fraction (FMF) of AOD are analyzed. Globally, a time difference of 30 min typically induces from 0.011–0.035 variation in AOD. For total, fine, and coarse AODs the typical time to decorrelation is around 2–10 days. For AE and FMF it is 3–33 days; that is, aerosol systems often persist significantly longer than individual events in them. Biomass burning regions tend to show the largest and fastest subdaily AOD variability and also longest times to decorrelation. Some sites show significant season-to-season variations in behavior. These results can be used to inform site-specific time collocation thresholds for aerosol validation analyses and account for temporal variation when estimating data set uncertainty. They also have implications for comparisons between different satellite products or models, data aggregation, and time series analyses. Results are provided on a site-by-site basis to facilitate use by other researchers.

Plain Language Summary It is common to compare different data sets that are measuring the same thing against each other, to check how consistent they are. Satellite data sets measuring aerosols—small particles in the atmosphere such as dust, smoke, smog, or volcanic ash—are often compared with ground-based data sets, which are very accurate but sparse. When comparing data sets like this, often the ground measurements are not made at exactly the same time as the satellite flies over. This time mismatch leads to a source of uncertainty in such comparisons, because a difference in the data could be from both measurement error and real variability. This study presents a method to quantify the typical level of variability of aerosols for different time mismatches, to allow for a better understanding of differences in such comparisons between different data sets.

1. Introduction

The Earth system is constantly changing, and our sampling of geophysical quantities, whether remotely sensed or modeled, is incomplete. As articulated well by Davis et al. (1994) and Anderson et al. (2003), these facts mean that studies of quantities such as aerosol optical properties implicitly or explicitly make decisions about spatial and temporal sampling, aggregation, relevant scales, and methods for analysis. For example, satellite aerosol data sets typically provide both so-called Level 2 (L2) products—near-instantaneous snapshots of a particular part of the world as seen along the platform's orbit—and Level 3 (L3) products representing spatiotemporal aggregates or averages of L2 data. L3 grids are often daily or monthly, with monthly aggregates in particular being used widely for climatological, trend, and model intercomparison studies (e.g., Hsu et al., 2012; Kinne et al., 2006; Sogacheva et al., 2020; Wei et al., 2019). Models and reanalyses typically provide output at regular (hourly to quarter-daily) time steps although monthly aggregates are arguably the most commonly used in research (e.g., Kim et al., 2014; Sand et al., 2017). Ground-based aerosol remote sensing such as from the Aerosol Robotic Network (AERONET, Holben et al., 1998) often provides

data on a regular instrument sampling cadence (order of several minutes), except where confounded by sampling gaps due to clouds, night (for solar sensors), or system maintenance, as well as daily, monthly, or climatological aggregates.

Validation and L2 intercomparison exercises impose collocation criteria on spatial and temporal location difference between data sets in order to determine whether or not a particular pair of points from data sets are considered "matched up" and therefore comparable (e.g., Ichoku et al., 2002, for the methodology used widely in satellite aerosol validation analyses). L3 and model intercomparisons typically aggregate data to a chosen time step before doing the comparisons, such as the monthly analyses mentioned previously. If daily rather than monthly comparisons are performed, even sensors on satellite platforms with similar nominal local solar crossing times may on a given day view a location at quite different times if their orbits are not the same (e.g., Sayer et al., 2017). Data assimilation systems typically ingest satellite and ground-based aerosol data as 3- to 6-hr composites (Benedetti et al., 2018; Xian et al., 2019), meaning (especially for polar-orbiting sensors) that the observations are assigned to model times that can be several hours separated from the time they were taken. Often, these decisions were adopted as convention several decades ago, based on computational and ease-of-use (rather than geophysical) considerations and have broadly become accepted as standard practice. In most cases they are applied globally; that is, the same matchup (or other) criteria are applied to all parts of the data set, regardless of the fact that the spatial and temporal scales of geophysical variability may differ significantly from one location to the other.

In recent years, however, increasing attention has been given to aggregation and matchup-related decisions in aerosol optical depth (AOD) studies, with a goal of quantifying and decreasing representativity and aggregation-related sources of uncertainty. These are sometimes referred to as "representation error" or "collocation mismatch uncertainty" (Schutgens et al., 2017; Virtanen et al., 2018). Note that spatial and temporal variations are coupled—no location exists in isolation from all others—though the main focus of the present study is the time domain.

Several of these studies have examined temporal aspects. Anderson et al. (2003) used assorted ground-based, airborne, and space-based aerosol optical properties (AOD and aerosol scattering coefficient) and found temporal scales of variation on the order of 2 hr to 2 days (and length scales of 40–400 km). Autocorrelation coefficients on spatial (±25 km) and temporal (±0.5–1 hr) scales used for satellite analysis (e.g., Ichoku et al., 2002; Kahn et al., 2010) in these data were generally in excess of 0.9. For the data sets examined, correlations approached 0 after around a week (*e*-folding times of ~2–3 days) but then increased again after a lag of around 1 year, due to the repeatability of aerosol annual cycles. Alexandrov et al. (2004) used a ground-based shadowband radiometer network in Oklahoma, USA (with fairly low AOD), and found that structure functions of temporal variation could be described by power laws, with scale breaks around ~1,000 and ~10,000 s (corresponding to ~15 min and ~3 days, respectively). Kaku et al. (2018) found *e*-folding times for surface aerosol mass in the southern United States around 3 days, but of column AOD in the same region under 1 day. A relevant (and application-dependent) question not addressed directly on a large scale by these studies, however, is on what scale does the magnitude of variation exceed some scientifically relevant threshold?

Virtanen et al. (2018) quantified this variation using AERONET data from a dense field campaign deployment in the mid-Atlantic United States. They validated satellite AOD using different spatial and temporal matchup thresholds and examined how the statistics of the comparison varied. They found a weak response (change around 0.02) of correlation coefficient to temporal thresholds between ±0.1 and ±2 hr and spatial thresholds between ±0.1 and ±1°. Their Figure S2 showed, for the sites in this region, the typical standard deviation of AERONET 550 nm AOD for a ±1-hr time period centered on satellite overpasses ranged from 0.01–0.03, which is larger than the roughly 0.01 uncertainty on the AERONET AOD. As the AERONET AOD uncertainty is largely driven by calibration it shows systematic behavior dependent on air mass and is not noise, so this variation is likely real rather than an instrumental artifact (Eck et al., 1999; Holben et al., 1998). This implies that the choice of a ±1-hr window might introduce an additional uncertainty of up to 0.01–0.03 (for this particular region, during this summer campaign period, and separate from other components of representation uncertainty).

In a series of papers using a variety of high-resolution model simulations and data sets Schutgens, Gryspeerdt, et al. (2016), Schutgens, Partridge, et al. (2016), Schutgens et al. (2017), and Schutgens (2020)

have investigated multiple aspects of representation uncertainty. These studies have numerous implications for the way comparisons are done and representation uncertainty is assessed, particularly for L3-type studies such as model evaluation. From a temporal point of view, a general message is that matchups should be done on as fine a time scale as possible. For a monthly or annual comparison between two data sets it is better to collocate the data on a finer (e.g., daily) basis and then aggregate only the matched points to a month or year than to compare the monthly or yearly products directly. Due to their main focus (model or satellite L3 evaluation) and the limitations of the data, however, these analyses have mostly considered hourly, daily, or longer composites and not the subhourly to several-hourly windows often used in satellite validation or intercomparison analyses.

This study aims to complement and extend this growing body of work by providing (1) estimates of the characteristic time scales of variation of aerosol optical properties at AERONET sites across the world and (2) the typical magnitude of variation for a given time mismatch. These properties include total column AOD, Ångström exponent (AE), fine-mode fraction (FMF) of AOD, and fine- and coarse-mode AODs. The data sets generated in this work are freely available to download from Sayer (2020), to encourage their use by the community. This will enable researchers to select location and seasonally appropriate time thresholds for validation and intercomparison analyses, to provide a balance between data volume and representation error.

Section 2 introduces the AERONET products and variogram analysis method used. Section 3 presents some results from a large-scale perspective, looking at the global picture of time scales of aerosol variation and also provides an example of the effect of changing temporal matchup thresholds on validation of Moderate Resolution Imaging Spectroradiometer (MODIS), Multiangle Imaging SpectroRadiometer (MISR), and POLarization and Directionality of the Earth's Reflectances (POLDER) aerosol products at the Kanpur, India, AERONET site. Section 4 provides a summary of the main findings and next steps. As mentioned previously, results for each AERONET site analyzed have been archived in the Zenodo repository at Sayer (2020).

2. AERONET Data and Variogram Analysis

2.1. AERONET Data

The main data set used here is the AERONET Version 3, Level 2 (cloud-screened, post-deployment calibrated, and quality-assured) direct Sun AOD data base (Giles et al., 2019). Data from all 1,527 sites (at the time of writing) from the start of the network in 1993 to the end of 2019 are considered. AOD is analyzed at a wavelength of 550 nm as this is a common reference wavelength in satellite retrievals and model simulations. AERONET does not measure AOD at 550 nm so here it is obtained by a second-order polynomial fit of AOD to wavelength (both quantities in log space), following the recommendations of Eck et al. (1999) and Schuster et al. (2006) and has an uncertainty around 0.01.

The standard AERONET AE product defined over the wavelength range 440 to 870 nm is also used; AE is the negative of the slope of log AOD versus log wavelength. Hereafter, references to AOD and AE implicitly refer to these wavelengths, unless stated otherwise. Uncertainty on AERONET AE is difficult to generalize because it is a derivative measure, so sensitive to both the magnitude of the AOD as well as any spectral correlations in the uncertainty of the AOD (Wagner & Silva, 2008) but is considerate most reliable in moderate or high aerosol loading (AOD ~0.2 or more).

The second data set is the AERONET spectral deconvolution algorithm (SDA) record. This provides estimates of FMF, fine-mode AOD, and coarse-mode AOD based on observed spectral characteristics of the (Version 3 Level 2) direct-Sun AOD data (O'Neill et al., 2003, 2008). Coverage is lower than the direct-Sun data due to additional requirements (e.g., minimum number of channels) and quality checks in the SDA algorithm. Uncertainty on FMF decreases with increasing AOD and depends on aerosol type and is 0.1–0.2 for AOD around 0.1 (Kleidman et al., 2005; O'Neill et al., 2008). The SDA data are provided at 500 nm but here converted to 550 nm (using total and fine-mode AE provided within the SDA data set) for consistency with the total column AOD analysis. This has a negligible effect on the results shown here.

2.2. Variogram Analysis

2.2.1. Definitions

Semivariograms provide a measure of the magnitude and scales of variation within a field. They have been used in remote sensing for decades (Curran, 1988, for an early overview), most commonly in the case of data fusion or gap filling by kriging (Chatterjee et al., 2010; Nguyen et al., 2012; Shi & Cressie, 2007; Yu et al., 2011, for some aerosol examples). Other aerosol uses include Hewson et al. (2013) to examine relationships between aerosol and rainfall, Snik et al. (2014) to estimate the uncertainty and best spatial resolution of AOD measured using an attachment for cellphone cameras, Sullivan et al. (2015, 2017) to quantify scales of aerosol variation over North America, and Alexandrov et al. (2016) who developed a representation of daily MODIS over-ocean AOD on regional scales in terms of spatial variograms.

Definitions and some properties are given in Chapter 4 of Cressie and Wikle (2011). In brief, a semivariogram γ is half the variance of the difference between values of a field Y at two locations x, $x+h$ separated by distance h:

$$\text{var}\,(Y(x)-Y(x+h)) = 2\gamma(h). \tag{1}$$

In the present case, the field Y corresponds to an aerosol optical property (e.g., AOD) at locations (times) x, $x+h$. If the field is intrinsically stationary (i.e., the variance of the difference is the same between any two points of a given separation) and isotropic (i.e., variations back and forward in time are statistically equivalent), then γ is dependent only on the distance h between points, often known as the lag. Note that x and h are sometimes represented in bold type as vector (i.e., multidimensional) quantities; the present analysis is concerned with one-dimensional (time series) data so scalar italic notation is used.

The quantity 2γ is known as the variogram; as mentioned by Davis et al. (1994), it is also referred to as the second-order structure function outside the geostatistics literature. Bachmaier and Backes (2011) note that the terms semivariogram and variogram are often used interchangeably and that there is confusion about the prefix "semi". They show that while γ is half the variance of the difference between field values at lag h (Equation 1), it is the full variance of field values at lag h, which is the quantity required for kriging applications. As the present analysis is concerned with uncertainty introduced in comparisons by temporal mismatch between observations, however, the most relevant quantity here is the square root of the variogram $\sigma = \sqrt{2\gamma}$, that is, transferring from variance to standard deviation space. This root variogram σ can then be used as a proxy for the representation uncertainty induced by temporal mismatch.

2.2.2. Empirical Estimates and Fitting

The true semivariogram is unknown so is generally estimated empirically and fit to a parametric model. Under the above assumptions the empirical semivariogram $\hat{\gamma}$ can be estimated as follows:

$$\hat{\gamma}(h \pm \delta) = \frac{1}{2N(h \pm \delta)} \sum_{i,j}^{N(h \pm \delta)} \left(y_i - y_j\right)^2. \tag{2}$$

Here, $\hat{\gamma}$ is computed for 54 logarithmically spaced time bins between 0.1 hr (6 min) and ~20,000 hr (~833 days). The quantity δ represents a tolerance factor accounting for the fact that points may not be spaced exactly according to the bin lag. Here, δ is taken as 5% of the bin central time, with a floor of 1.5 min and a ceiling of 1 day. As δ tends toward 0, $\hat{\gamma}$ becomes an unbiased estimator of γ (Cressie & Wikle, 2011). N is the number of data pairs within each bin with lag $h \pm \delta$, and y_i, y_j are the paired points within each bin. Note that as data coverage is limited to daytime, lags near 12 hr correspond to pairs of points near dawn and dusk, while lags of only a few hours or somewhat above 12 hr can be found throughout the day. It is clear from Equation 2 that this represents half the mean squared difference between pairs of points separated by time h (and that as such the quantity σ is the root-mean-square difference).

Empirical semivariograms are often fit to parametric models to provide a continuous function. Here, the commonly used powered exponential covariance function (PECF) (Cressie & Wikle, 2011) is used:

$$\gamma(h) = a_0 + a_1\left(1 - \exp\left[-\left(\frac{h}{a_2}\right)^{a_3}\right]\right) \tag{3}$$

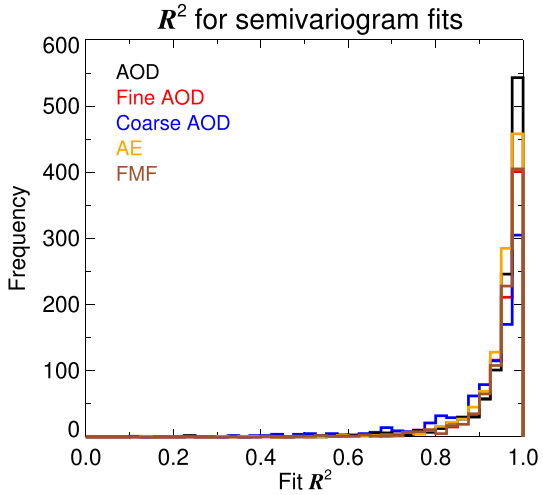

Figure 1. Histograms of coefficients of determination of R^2 between logarithms of observed empirical semivariograms and piecewise exponential covariance function fit values for all sites. Results for total aerosol optical depth (AOD) are shown in black, fine-mode AOD in red, coarse-mode AOD in blue, Ångström exponent (AE) in orange, and fine-mode fraction (FMF) of AOD in brown.

for $h > 0$, where a_0, a_1, a_2, and a_3 are coefficients determined by least squares fitting using the MPFIT software package (Markwardt, 2009). The fit is done in $\log(\hat{\gamma}(h))$ space (base 10) as the data typically cover several orders of magnitude, and the behavior is of interest at both extremes. First guesses at parameters a_0 to a_3 are taken as 10^{-4}, 0.1, 1 hr, and 1, respectively.

To increase robustness only bins with at least 50 data pairs are used in the fit, and only empirical semivariograms with at least 27 of the bins (i.e., half, sampling a substantial portion of the lag space) sufficiently populated are considered. This leaves 1076 sites for direct-Sun calculations and 992 for SDA. The median coefficient of determination R^2 between observed $\log(\hat{\gamma}(h))$ and fit $\log(\gamma(h))$ is above 0.95 for all quantities, indicating that the PECF model captures the variation of the data well in most cases (Figure 1). Note that due to the properties of logarithms, these R^2 are also equivalent for derived σ. Unless otherwise stated analyses below exclude those fits for which $R^2 < 0.6$ (corresponding to <2% of the data), as the PECF model is a poorer representation in those cases. The choice of R^2 threshold within reasonable variations does not significantly alter the population of sites retained.

2.2.3. Nugget, Sill, Range, and Autocorrelation

Regardless of the form assumed for the semivariogram, several quantities (the nugget, sill, and range) are commonly derived from it and interpreted. These are illustrated conceptually in Figure 2. By definition for the true semivariogram $\gamma(h = 0) = 0$, as the variance of the difference with zero lag must be 0. The resulting discontinuity at the origin for the empirical semivariogram (due to measurement and/or model specification error) is known as the nugget n and for a model of the form in Equation 3, $n = a_0$.

As h approaches infinity, γ tends toward the sill (the variance of the underlying distribution); here, $s = a_0 + a_1$. In their analysis of MODIS over-ocean AOD data, Alexandrov et al. (2016) found that some grid cells analyzed showed a partial saturation feature whereby the sill appeared to be reached but then, at a larger lag, the variogram began increasing again. They interpreted this as due to local nonstationarity, that is, that in such cases the AOD field is better represented by a combination of a 2-D linear spatial gradient plus the stationary field. This behavior was rarely seen in the data here, likely due in part due to the largest lag considered being ~2 years, over which time trends in aerosol optical properties are likely to be small, especially in comparison to s. Some sites did show oscillation of $\hat{\gamma}$ near 1-year lag; this was also seen in the temporal autocorrelation analysis of Anderson et al. (2003) and arises as many aerosol sources show interannual-repeating patterns. However, cases of severe long-term nonstationarity are likely to be caught by the $R^2 < 0.6$ threshold on the PECF fit discussed previously. Seasonal nonstationarity is examined later.

The range r is the time taken for γ to reach the sill. As Equation 3 is asymptotic the true range is undefined, and the effective range is often taken to be the point where the semivariogram reaches 95% of the sill; given $e^{-3} \approx 0.05$, if the nugget is negligible, then

$$\left(\frac{r}{a_2}\right)^{a_3} \approx 3, \tag{4}$$

and so by rearrangement,

$$r \approx a_2(3^{1/a_3}). \tag{5}$$

Note that from Equation 3, $a_3 = 1$ corresponds to an exponential decay of γ with increasing lag. Figure 3 shows that $a_3 \approx 1$ for total and fine AOD (central 68% of points, i.e., one standard deviation, between ~0.83 and 1.15 for both), albeit slightly skewed to lower values. There is a more pronounced shift to values below 1 for the other quantities. This implies slightly slower than exponential decay, on average, and supports the use of the PECF over a pure exponential model.

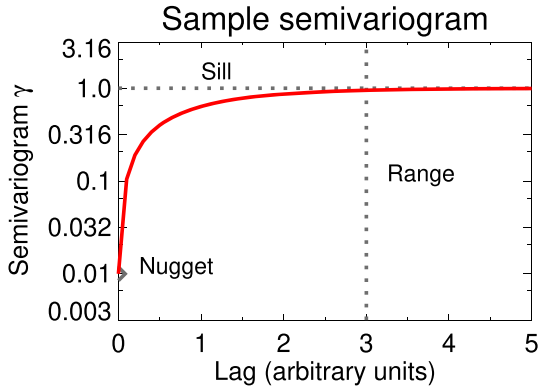

Figure 2. Example semivariogram (red) generated using the powered exponential covariance function model (Equation 3) with $a_0 = 0.01$, $a_1 = 1$, $a_2 = 1$, and $a_3 = 1$. The nugget, sill, and range are indicated by the diamond, horizontal, and vertical dashed lines, respectively.

These quantities can also be used to estimate the lag-h autocorrelation of the field $\rho(h)$. Haslett (1997) discussed the application of empirical semivariograms to time series analyses and noted they have some advantages in strictness of assumptions over other methods to estimate autocorrelation. The relation between these two is as follows (cf. Equation 2 of Bachmaier & Backes, 2011):

$$\gamma(h) = s(1 - \rho(h)). \tag{6}$$

From Equations 3 and 6 it can be seen that, when the nugget is small compared to the sill, the e-folding time for ρ is approximately the fit coefficient a_2. Finally, by rearrangement of Equation 3, the lag for at which a given value of γ is reached is

$$h = a_2 \left[-\ln\left(1 - \frac{(\gamma - a_0)}{a_1}\right)\right]^{1/a_3} \tag{7}$$

Fit coefficients for Equation 3, R^2, and empirical variogram values and data counts are freely available to download from Sayer (2020) for each AERONET site and each optical property. For AOD, two other quantities are also provided. First is the time where the fit root variogram $\sigma = 0.01$ (i.e., the standard deviation of the difference reaches the nominal AERONET direct-Sun uncertainty), using Equation 7 and bearing in mind $\sigma = \sqrt{2\gamma}$. Second is σ for various lags of interest. All calculations involving time units are in hours.

3. The Global Picture From AERONET

3.1. Time Scales and Magnitudes of Variation

3.1.1. Kanpur as an Example

An example set of root variograms (recall $\sigma = \sqrt{2\gamma}$) is shown for the AERONET site Kanpur, India, in Figure 4. This site was chosen as an example as it lies in the Indo-Gangetic Plain (IGP), a densely populated area with a complex aerosol burden, hence of high interest for air quality and climate studies (Dey & Di Girolamo, 2011; Giles et al., 2011; Ramanathan et al., 2007; van Donkelaar et al., 2016). Aerosol properties at Kanpur exhibit distinct seasonality: peak AOD (and lowest AE/FMF) from April–June when the dust contribution to AOD is highest, monsoon from June–September leading to increased aerosol washout (and potential for hygroscopic growth of remaining aerosols), and an increased relative contribution of fine-mode aerosol from biomass burning and urban/industrial activities in other months with a secondary peak (higher AOD and FMF/AE) from November–January (Eck et al., 2010; Giles et al., 2011, 2012; Singh et al., 2004). Complexity in aerosol and surface reflectance characteristics also mean it can be a challenging site for satellite aerosol optical property retrievals (Jethva et al., 2007, 2010).

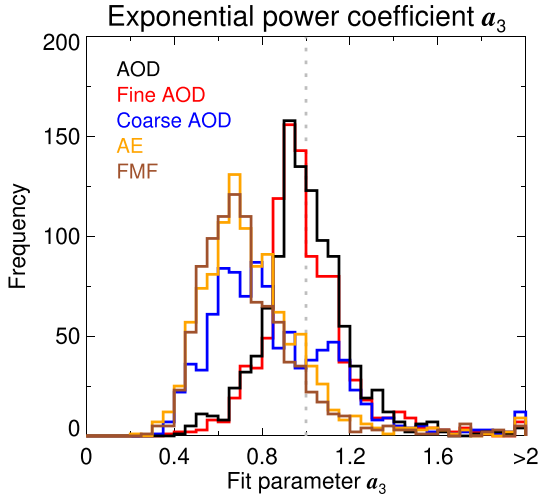

Figure 3. Histograms of fit coefficients a_3 (Equation 3) for all sites. Results for total aerosol optical depth (AOD) are shown in black, fine-mode AOD in red, coarse-mode AOD in blue, Ångström exponent (AE) in orange, and fine-mode fraction (FMF) of AOD in brown.

Figure 4 shows that σ for total AOD is dominated by variations in fine-mode AOD, with that of coarse-mode AOD around a factor of 3 smaller. At a 30-min lag, σ for total AOD is around 0.055, significantly larger than the (mostly systematic) AERONET AOD uncertainty of ~0.01 (Eck et al., 1999). Note that this does not imply that a validation exercise using a ±30-min AERONET averaging window incurs a temporal mismatch uncertainty of 0.055; rather, this is an upper bound on likely mismatch uncertainty for validation exercises using such a window. In many cases there will be multiple AERONET observations within a given window, in which case averaging (which is standard practice) or picking the closest AERONET observation in time would mitigate this somewhat. Either way, consistent with results from Virtanen et al. (2018) for North America, this implies that time mismatch may be a more significant contribution than AERONET measurement uncertainty to the total uncertainty in the comparison.

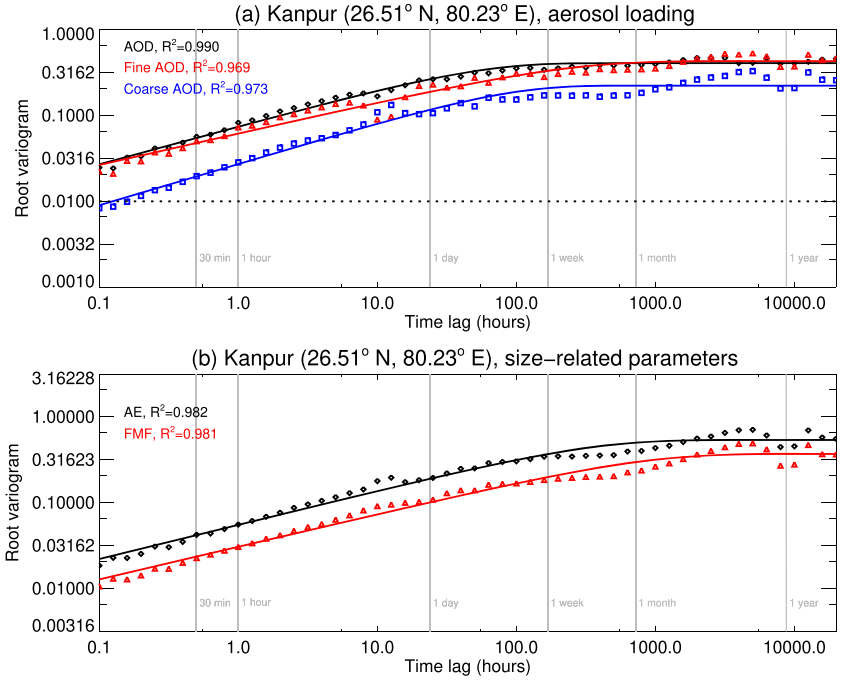

Figure 4. Root variograms $\sigma = \sqrt{2\gamma}$ of aerosol optical properties for Kanpur, India. Panel (a) shows results for total aerosol optical depth (AOD) in black, together with fine AOD in red and coarse AOD in blue. Panel (b) shows results for Ångström exponent (AE) in black and fine-mode fraction (FMF) of AOD in red. Diamonds show points, and lines show fits (Equation 3); the fit coefficient of determination in log space R^2 is also provided in the legend. The horizontal dashed line in (a) shows $\sigma = 0.01$. Vertical lines indicate various time lags of interest.

Time profiles of σ for AE and FMF show a very similar shape to each other, suggesting that these quantities capture similar underlying variations, with the former being roughly a factor of 2 larger. This magnitude difference is not unexpected given that FMF varies from 0 to 1 by definition (O'Neill et al., 2003), and AE varies roughly in the range 0 to 2 going from coarse-dominated to fine-dominated aerosol columns (Eck et al., 1999). Range r (time to decorrelation of the field) is lower for AOD (~7 days) than AE or FMF (~49 and 122 days, respectively), which is consistent with seasonal-scale aerosol source and transport mechanisms being significantly longer than individual aerosol events within them. The potential partial saturation feature from around 1 week (i.e., approximately the time of the AOD range) to 1 month in AE and FMF may reflect the time scale of individual aerosol events within a season. The difference between AE and FMF range may indicate differences in the information captured in these quantities, as well as uncertainties of model parametrisation. Note σ for all quantities shows a brief decrease around 1-year lag; as mentioned previously, this reflects the year-to-year broad-scale consistency in aerosol features.

For all optical properties in Figure 4, there is a departure from the fit around 10-hr lag. As AERONET data are only collected during sunlit hours, these points generally correspond to a midmorning observation paired with an early evening one. This could be indicative of diurnal cycles with shapes such that AOD tends to be more closely related between these times than would be expected (as the total and fine AOD points lie below the fit line), that is, some nonstationarity. Several studies have examined diurnal cycles of aerosol optical properties at Kanpur. Smirnov et al. (2002) looked at AOD on an annual basis and Singh et al. (2004) at AOD and AE on a seasonal basis. While Smirnov et al. (2002) found (their Figure 2a) a fairly flat average AOD throughout the day, declining during the evening, Singh et al. (2004) found (their Figure 7) that in all seasons except premonsoon the AOD at midmorning and early evening were more similar to each other than to AOD through the afternoon. This is consistent with the dip around 10 hr in Figure 4. The pattern is less clear for AE. It is also worth noting that the present study benefits from a significantly longer data record than Smirnov et al. (2002) or Singh et al. (2004).

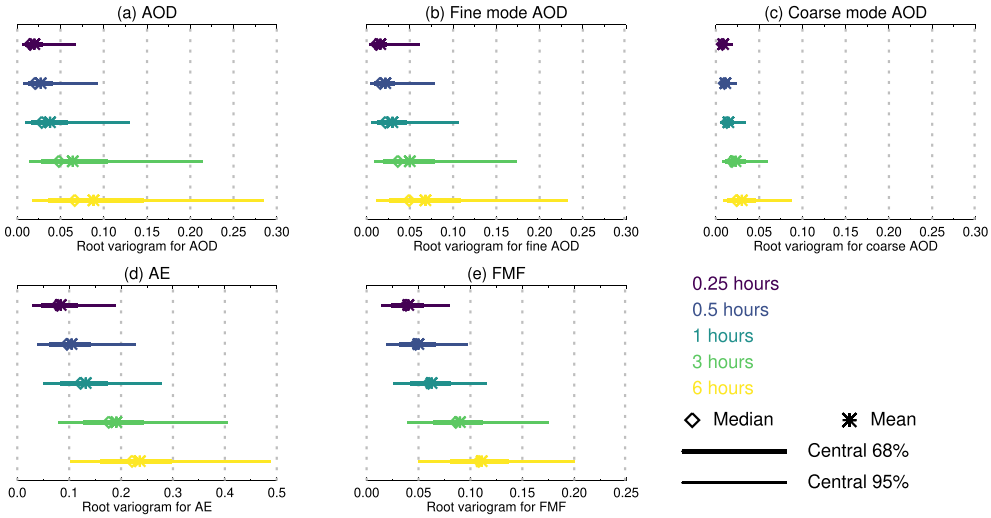

Figure 5. Box-whisker plots showing $\sigma = \sqrt{2\gamma}$ at time lags between 0.25 and 6 hr for aerosol optical properties across all sites. Panels show (a) total aerosol optical depth (AOD), (b) fine-mode AOD, (c) coarse-mode AOD, (d) Ångström exponent (AE), and (e) fine-mode fraction (FMF) of AOD. The legend is given in the bottom right.

Bigger picture, this acts as a reminder that diurnal cycles can influence the empirical semivariograms. In a practical sense, this is not expected to be too important for the present application (provided the magnitudes and time scales of short-term variation do not show too strong a diurnal cycle) as it is most relevant on time scales of several hours, while most validation/intercomparison analyses use shorter (i.e., subhour) colocation criteria. One option could be to construct semivariograms using only starting points from a given local time range (similar to the seasonal analyses performed later), although in that case there would be a trade-off between level of detail and available data volume.

3.1.2. Global Variations

Showing results aggregated for all sites, Figure 5 indicates how potential mismatch uncertainties change as a function of lag, for lags between 0.25 and 6 hr. The former is a shorter time than that typically used (0.5 or 1 hr) for validation of satellite aerosol products (Ichoku et al., 2002) so represents an attempt to decrease mismatch uncertainties while not decreasing total data volume too much. For validation studies given the AERONET sampling cadence is typically 5–15 min in most cases the actual time lag (and so mismatch uncertainty) would therefore be closer to the 0.25-hr results, even if a longer averaging window is permitted. This is examined later in section 3.2.2 using various satellite products.

The final two lags considered (3 and 6 hr) correspond to typical model assimilation time steps (i.e., data aggregated sequentially in chunks of these lengths and then assimilated on these time steps, Benedetti et al., 2018; Xian et al., 2019). In this sense it is representative of a typical upper bound of mismatch uncertainty for a perfect polar-orbiting satellite retrieval, which just misses the cutoff for inclusion in one assimilation cycle and so is assimilated just under 3 or 6 hr late. The actual mismatch uncertainty would again depend on the actual lag between observation and assimilation time.

For all quantities and times, distributions of σ are positively skewed, with a long tail and mean values larger than median values. This indicates that a subset of sites have much larger temporal variability than the others. The main implication of this is that, for satellite validation analyses, it may be beneficial to consider different temporal matchup windows for different sites; where this is shorter than normal this represents an effort to control mismatch uncertainty, and where longer than normal an effort to increase sampling without introducing significant additional uncertainty. For AOD (the most commonly validated quantity) the median (and central 68% of values) for σ is 0.019 (0.011–0.035) for a 30-min lag; at 15 min, this decreases to 0.014 (0.009–0.025), and for 1 hr this increases to 0.025 (0.014–0.049), that is, roughly a 25% decrease or increase by halving or doubling the window, respectively. Note this indicates that mismatch uncertainties may exceed notional AERONET radiometric uncertainty even for this shortest time lag tested. These results also suggest that the 5-min revisit time of the new generation of geostationary imagers (e.g., Schmit et al., 2017) should be

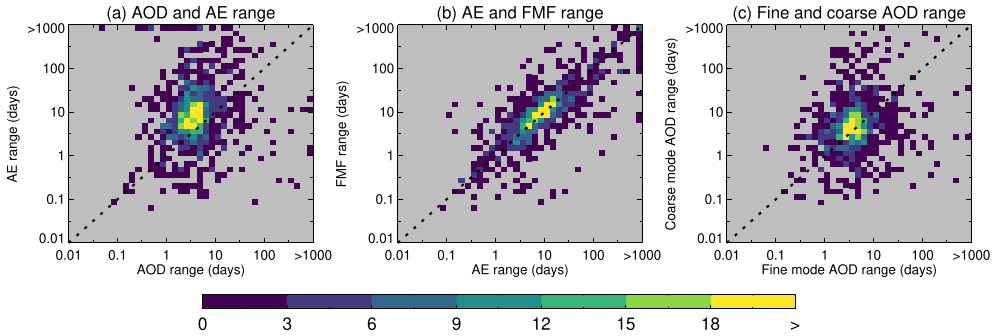

Figure 6. Scatter density plots of semivariogram range compared between various aerosol optical properties. Panel (a) shows total aerosol optical depth (AOD) and Ångström exponent (AE), (b) shows AE and fine-mode fraction (FMF) of AOD, and (c) shows fine and coarse AODs.

sufficient to resolve aerosol temporal variations of scientific interest (assuming retrieval algorithms are sufficiently capable).

Figure 6 shows the relationship between semivariogram range *r* for different optical properties. The reader is reminded that "range" refers to time to decorrelation of the field, rather than the variation in magnitude of the optical property at a given site: a measure of length rather than magnitude. The ranges for all quantities are most commonly between roughly 3 and 10 days. There is a modest relationship between ranges for different quantities; that is, sites with a longer AOD range also have a longer AE range. This is likely related to variations of persistence of aerosol sources and aerosol lifetimes. Ranges for AE and FMF are very strongly correlated (as expected, because they carry similar information), while those between fine- and coarse-mode AODs less so. This is also consistent with different source and sink characteristics for different particle sizes. As at Kanpur (Figure 4), ranges tend to be slightly longer for AE and FMF than AOD, indicating aerosol seasonality is longer than individual aerosol features. Coarse-mode range also tends to be slightly higher than fine-mode range.

Specifically, the median (and central 68% of values) for range are 3.9 (1.9–8.2) days for AOD, 3.8 (1.8–9.3) days for fine AOD, 4.5 (1.7–13) days for coarse AOD, 8.5 (2.7–33) days for AE, and 10.3 (3.0–33) days for FMF. This implies that monthly composites of satellite data or model simulations typically sample 3–10 distinct aerosol events from one larger-scale aerosol system. It also implies that satellite instrument with revisit times on these scales or longer, such as the Along-Track Scanning Radiometer (ATSR) series, MISR, and Cloud-Aerosol LIdar with Orthogonal Polarization (CALIOP), can often be considered independent observations when aggregating from daily to monthly or longer time scales. The ATSRs and MISR have revisit frequencies of order 5 and 9 days at the equator, respectively, while CALIOP's curtain sampling means only the subsatellite track is sampled and that with a 16-day revisit period (for each of daytime and nighttime). Practically, revisit time is further decreased by cloud cover. In contrast, broader-swath sensors on polar-orbiting platforms (e.g., MODIS and several others) sample (suitable conditions permitting) roughly once a day and geostationary potentially several times an hour, meaning these data are highly autocorrelated. While uncertainties in satellite L3 aerosol products have not yet been broadly developed, these autocorrelation structures should be accounted for. The same implies to temporal composites of AERONET data as well. Ranges of order weeks (for AOD) to a month (for AE or FMF) do not necessarily mean that there is no autocorrelation between successive monthly composites, as aerosol systems do not start and end in perfect synchronicity with calendar months. Indeed, monthly aerosol trend analyses often estimate and account for 1-month lag autocorrelation when calculating trend uncertainty estimates (Hsu et al., 2012; Zhang & Reid, 2010).

A small number (approximately 4%) of sites have unusually short (<0.1 day) or long >300 days) range for AE or FMF. Manual inspection reveals (not shown) that these are typically short-term sites operated for up to several months. As such (for the cases of short range) the results may be sensitive to the weather system during the period of operation, and (for very long range) values near the sill were not reached and so the relevant fit parameters were not well constrained by the data.

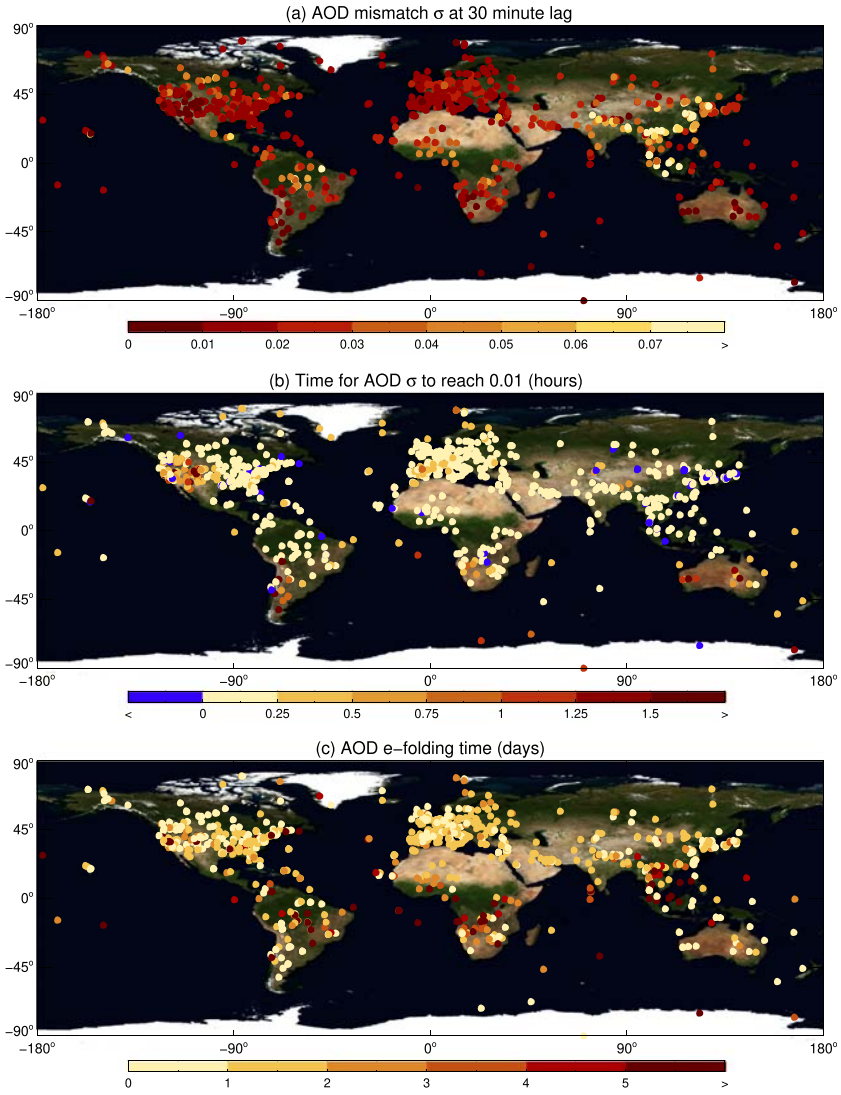

Figure 7. Global maps of aerosol optical depth (AOD) mismatch uncertainty results. Panel (a) shows $\sigma = \sqrt{2\gamma}$ evaluated at a 30-min lag. Panel (b) shows the time at which σ reaches 0.01; values in blue here indicate $\sigma > 0.01$ for zero lag. Panel (c) indicates the e-folding time (coefficient a_2 in Equation 3) for AOD.

As AERONET sites are distributed unevenly across the world, it is instructive to examine maps of the data as well as aggregated distributions. Some results for AOD are shown in Figure 7. In Figure 7a, higher σ at 30-min lag is most commonly seen in regions associated with high aerosol loading, especially biomass burning aerosols. The most striking feature stretches across the IGP, southeastern Asia, and eastern Asia, but clusters of $\sigma > 0.03$ are also seen across parts of Africa, South America, and boreal regions associated with summertime biomass burning. This is somewhat unfortunate in that such places also tend to be comparatively challenging for satellite aerosol retrievals and model simulations. Additionally, in these areas σ may approach or exceed the Global Climate Observing System (GCOS) goal for error on an AOD climate data record of the greater of 0.03 or 10% (GCOS, 2011). Thus, the temporal mismatch uncertainty may be nonnegligible compared to the goal error to be assessed, clouding judgments of how successful a given data set is.

Figure 7b shows that achieving $\sigma \leq 0.01$ across most of the world requires a time mismatch smaller than 15 min. In this respect it is fortunate that the sampling cadence at most AERONET sites is between 2 and 15 min, with 5 min being common, meaning that in many cases the actual lag in validation exercises should

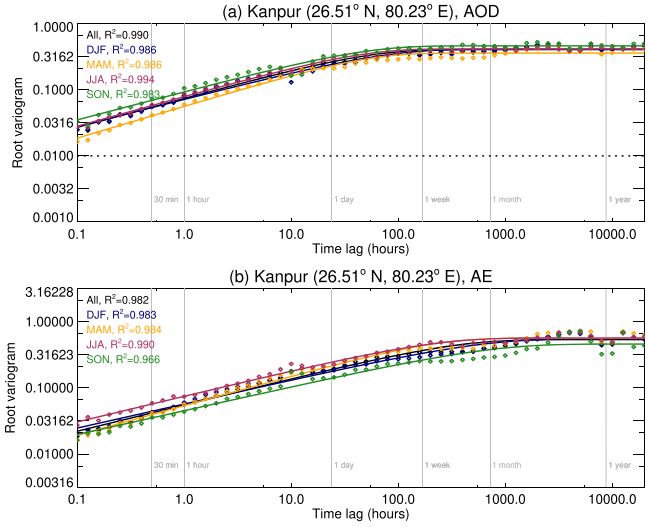

Figure 8. Seasonal root variograms $\sigma = \sqrt{2\gamma}$ for Kanpur, India (cf. Figure 4). Panel (a) shows results for total aerosol optical depth (AOD), and panel (b) shows results for Ångström exponent (AE). In both, fits using all points are in black, those using data from December–February (DJF) in navy, those from March–May (MAM) in orange, from June–August (JJA) in maroon, and from September–November (SON) in green. Diamonds show points, and lines show fits (Equation 3); the fit coefficient of determination in log space R^2 is also provided in the legend. The horizontal dashed line in (a) shows $\sigma = 0.01$. Vertical lines indicate various time lags of interest.

be shorter than this. In some cases $\sigma \geq 0.01$ at zero lag; as mentioned earlier this reflects measurement uncertainty or the PECF model breaking down at small lag but does imply that mismatch uncertainties are likely to be nonnegligible for small lags. Most of the world lies below the 30-min lag typically adopted for satellite validation exercises. This indicates than even with this window nonnegligible mismatch error can be introduced if there are a few missing AERONET observations around the time of the satellite overpass. It also implies that comparing satellite instruments on different orbits may often require strict collocation (although algorithm errors may still swamp mismatch uncertainties in some cases). Some mountainous or elevated areas (e.g., the Rocky Mountains in North America, Andes in South America, and central plateau of southern Africa), as well as much of Australia and some ocean sites, indicate that longer time windows may be tolerable (increasing data volume for comparison) without introducing significant additional uncertainty.

Finally, Figure 7c shows the *e*-folding time for autocorrelation (one third of the range when $a_3 \approx 1$; Equation 5). Longer time periods tend again to be most strongly associated with areas influenced by biomass burning aerosol transport. This is not quite a mirror of Figure 7a, though, as areas of heavy dust or urban/industrial loading do not stand out as prominently. Combined, these two panels show that biomass burning regions show both the largest short-term temporal variability in AOD and the longest time to decorrelation, that is, the longest aerosol "events".

3.1.3. Seasonal Variogram Variations

The site-to-site variations discussed in section 3.1.2 imply that temporal variation is strongly dependent on the geophysical processes driving the aerosol system. To that end, it is plausible that the semivariograms are not truly stationary, that is, that semivariograms constructed from observations beginning in different seasons may exhibit different behavior. Some differences would also arise as subsamples by definition have a smaller data volume and are noisier, although for a well-sampled site it is expected that they would remain sufficiently robust.

This is briefly investigated here by repeating the construction of $\hat{\gamma}$ and fitting of γ for subsets of data. These subsets correspond to the data pairs where the earlier observation fell in a given season, for each of the canonical four seasons, that is, December–February (DJF), March–May (MAM), June–August (JJA), and September–November (SON). The seasonal variation Δ_σ for a given lag can then be defined as

$$\Delta_\sigma(h) = \max\left(\sigma(h)_{seas}\right) - \min\left(\sigma(h)_{seas}\right), \tag{8}$$

where σ_{seas} denotes the seasonal root variograms. The relative seasonal variation is then $\Delta_\sigma(h)/\sigma(h)$. For example, suppose for the all-data case $\sigma = 0.02$ at some lag, and the individual seasons' results provided $\sigma_{seas} = \{0.015, 0.018, 0.022, 0.025\}$. Then $\Delta_\sigma = 0.01$ and the relative seasonal variation is $0.01/0.02 = 0.5$. Note that these canonical four seasons may not be the most appropriate representation of aerosol seasonality at all sites. However, they represent a balance between providing some seasonal specificity to the results while maintaining sufficient data volume at a high proportion of sites for analysis.

An example is shown for AOD and AE for Kanpur in Figure 8. The fit R^2 is similar for the full data set and the subsets. The data for the four seasons show very similar behavior but different magnitudes; overall, for AOD σ is largest in SON when fine-mode aerosol dominates and smallest in MAM as the dust contribution increases (Eck et al., 2010; Giles et al., 2012). For AE σ is highest in JJA (as the dust peak ends and monsoon begins), and the lowest in SON. At 30 min, for AOD, $\Delta_\sigma = 0.029$ and the relative seasonal variation $0.029/0.055 = 0.53$. This indicates nonstationarity at a seasonal level at this site, implying that temporal mismatch uncertainties may be better modeled at a seasonal level. Note that the departure around 10-hr lag shown in Figure 4 is also present here, consistent with the discussion above and seasonality of diurnal cycles seen by Singh et al. (2004).

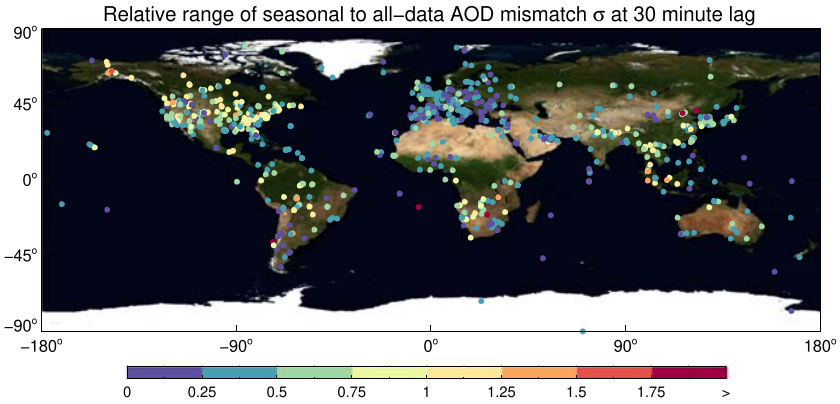

Figure 9. Relative seasonal-to-annual variation (see text) in root variograms $\sigma = \sqrt{2\gamma}$ for aerosol optical depth (AOD) evaluated at a lag of 30 min.

Extending this, Figure 9 shows the relative seasonality in σ for AOD at a 30-min lag for all (851) AERONET sites providing sufficient data and fit quality for at least two seasons. Overall, the median (and central 68% of values) for σ for these sites are 0.019 (0.011–0.035); for Δ_σ this is 0.008 (0.003–0.024), and for the relative seasonal variation 0.42 (0.19–0.80). This means for the median site the relative seasonal variation in σ is 42%, or around 0.008, which might not be important for some purposes. However, larger values are found in many of the aforementioned regions showing significant seasonality in fine-mode aerosol loading. This may be because in many biomass burning regions the smoke-free season is fairly pristine compared to the smoky season (e.g., Martin et al., 2010; Queface et al., 2011).

3.2. Influence on Validation Analysis at Kanpur, India

3.2.1. Satellite Data and Matchup Method

Here the influence of time thresholds on validation of satellite AOD retrievals at Kanpur is examined. This site is on the upper end of the typical range for AOD σ in Figure 5. Three different satellite AOD products are used:

(1) The Collection 6.1 MODIS Aqua data set (mid-2002 to present), corresponding to retrievals from the combination of Deep Blue (DB) and Dark Target (DT) algorithms (Gupta et al., 2016; Hsu et al., 2019) at 10-km nominal horizontal pixel size. MODIS is a broad-swath (2,330 km) single-view passive imager. This merged data set is described in Sayer et al. (2014) and is intended to represent a 'best of' data set with improved coverage and at least as high quality as the individual algorithms (on global average). Over land, the merging process includes a map of climatological surface category (arid, vegetated, or transitional) for each calendar month (and most pixels around Kanpur fall into the latter two categories in most months). Dependent on this and retrieval quality assurance flags, either one algorithm's results are used or the two are averaged. As a result DT retrievals are used in DJF and SON; in MAM DT accounts for about two thirds of retrievals, while the others are DB or average; and in JJA almost 90% of retrievals are from DB.

(2) The Version 23 MISR data set (2000 to present), described by Garay et al. (2020). MISR is a narrow-swath (380 km) multiangle passive imager. The main updates affecting retrievals over land compared to earlier MISR retrievals are calibration updates and a finer L2 spatial resolution (4.4-km horizontal pixel size compared to 17.6 km in previous versions). MISR AOD is provided at 558 nm and converted to 550 nm using the AE provided in the data (increasing AOD by up to 3%).

(3) The Version 2.1 "models" processing of the Generalized Retrieval of Atmospheric and Surface Properties (GRASP) algorithm for POLDER (2005–2013). POLDER was a broad-swath (1,600 km) multiangle passive polarimeter. GRASP is described by Dubovik et al. (2011, 2014); a main conceptual difference from the MODIS and MISR algorithms used here is that it takes advantage of polarimetry and temporal compositing of data for a retrieval with fewer assumptions about aerosol and surface properties. Data are provided on a 10-km sinusoidal grid; the AOD is provided at 565 nm, and converted to 550 nm using the AE provided in the data (increasing AOD by up to 8%).

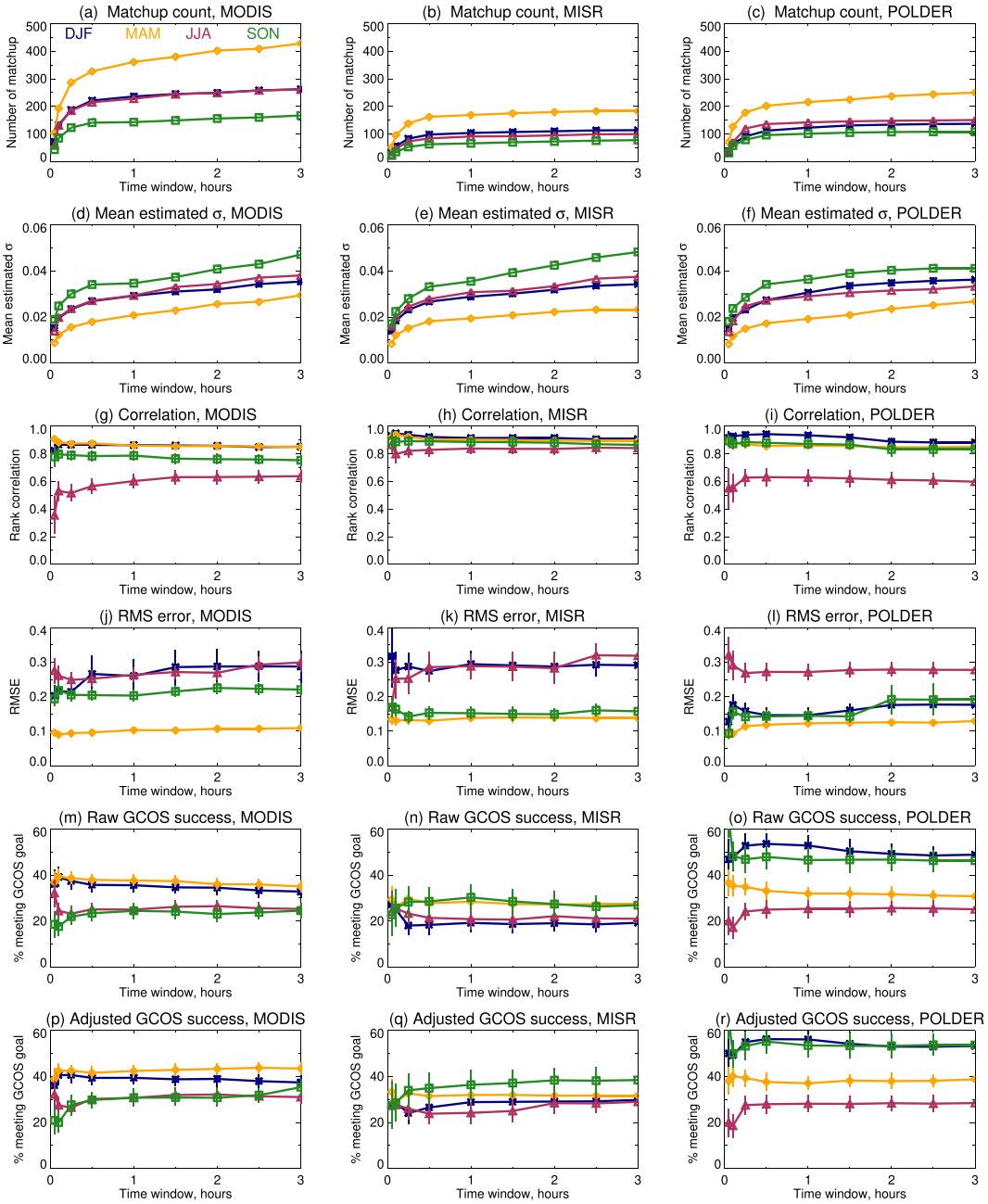

Figure 10. Selected statistics for validation of (left) Moderate Resolution Imaging Spectroradiometer (MODIS), (middle) Multiangle Imaging Spectoradiometer (MISR), and (right) POLarization and Directionality of Earth's Reflectances (POLDER) aerosol optical depth (AOD) against Aerosol Robotic Network (AERONET) at Kanpur, India. Panels (a)–(c) show the number of matchups, (d–f) the mean σ estimated for the matchups, (g–i) Spearman's rank correlation coefficient, and (j–l) the root-mean-square error (RMSE). Panels (m)–(o) and (p)–(r) show the fraction of retrievals agreeing with AERONET within the Global Climate Observing System (GCOS) goal AOD uncertainty of Max(0.03,10%), without and with adjustment for matchup uncertainty, respectively. In all panels, matchups from December–February (DJF) are in navy, those from March–May (MAM) in orange, from June–August (JJA) in maroon, and from September–November (SON) in green. Vertical bars indicate uncertainties on the metrics, as described in the text.

The AERONET AOD at 550 nm is calculated as described in section 2.1. A variety of time thresholds are used to consider whether the satellite retrieval and AERONET observation are matched up; in each case, only the nearest AERONET point in time (rather than the average within the windows) is used. This is because if all AERONET measurements within a window lie only before or only after the satellite time (rather than being

spread around it), it is likely that the nearest point in time may be more representative of the aerosol field at overpass time than the average. This differs from prior validation analyses and is suggested that it is explored for future similar work, because of the high temporal autocorrelation on short scales found here and in previous studies (e.g., Alexandrov et al., 2004; Anderson et al., 2003).

Spatially, the median of all satellite retrievals within ±25 km of the AERONET is taken to represent the satellite AOD (and note the merged MODIS product is as provided is already filtered to remove poor-quality retrievals). The use of median instead of mean was suggested and applied in Sayer et al. (2018, 2019) to increase the robustness of the analysis to outliers in the satellite data, such as single-pixel retrieval artifacts or true heterogeneity within the averaging area which can skew the mean. Thresholds around ±25 km are used widely in validation analyses. This spatial threshold is kept constant here to isolate the contribution from changing time thresholds, although future work beyond the scope of this study could develop more locally specific spatial aggregation protocols accounting for the unique characteristics of individual AERONET sites.

Note that these spatial and temporal thresholds were chosen initially by Ichoku et al. (2002) based on visual inspection of the movement of aerosol features across the Atlantic Ocean in satellite images in summer 1988. They observed motion about 50 km over 1 hr (~6,000 km over 5–6 days) and picked this over shorter windows such that such a box should yield sufficient (up to 25) samples for the nominal 10-km MODIS retrieval to be statistically representative. These thresholds have since been adopted widely for regional and global validation studies from many satellite data sets (and intuitively seem reasonable), although to this author's knowledge there has been no systematic global investigation of aerosol motion scales for validation purposes (aside from the AERONET-centric perspective within this work).

Several metrics are presented with uncertainty estimates: These are the standard deviation in the metric calculated using the bootstrap technique (sampling the matchups with replacement 100 times; Efron, 1979). These are important because uncertainties in summary metrics are seldom reported in aerosol validation exercises, yet obtaining precise metrics often takes a much larger sample size than intuitively expected (Schonbrödt & Perugini, 2013). Note however that (particularly for smaller sample sizes) the bootstrap method tends to underestimate the uncertainty of a metric as only the sample (rather than population) variance is sampled.

3.2.2. Validation Results

Figure 10 shows various validation metrics as a function of the time window considered when collocating the data. Results are presented on a seasonal basis as there is clear seasonal variation in the validation results; this is expected given the variation in aerosol and surface characteristics throughout the year (Eck et al., 2010; Jethva et al., 2007, 2010). Further, it implies that at sites with such seasonal variation that presenting only annual summary statistics may be insufficient, particularly if these statistics are to be compared against those of different algorithms, which may have different seasonality in data coverage or error characteristics.

Overall, a plurality of matchups are obtained in MAM (Figures 10a–10c), at which time the aerosol column tends to be dominated by dust (Eck et al., 2010). Fewest are obtained during SON (from the late monsoon through to increasing presence of fine-mode aerosols). This pattern is consistent between sensors, although magnitudes differ due to the narrower swath of MISR versus MODIS and shorter mission length of POLDER compared to the others. As expected, increasing time window increases the number of matchups obtained. This response begins to plateau from 30 min onward, which makes sense because if it is sufficiently cloudy to preclude AERONET observation within these time windows, it is likely that there will be few or no satellite retrievals within the ±25 km spatial window.

Figures 10d–10f show the mean value of σ for matchups obtained within the given time window. Note that this is different from σ at the mean time lag, due to the nonlinearity of Equation 3. This is highest in SON and lowest in MAM (cf. Figures 8) and exceeds 0.01 in all seasons even for a 15-min matchup window. Figures 10g–10i show Spearman's rank correlation between AERONET and satellite AOD. Rank correlation is preferable to the (more commonly used) Pearson's linear correlation because the latter is more sensitive to outliers in the data (such as from the skewed distributions common for AOD, e.g., O'Neill et al., 2000; Sayer & Knobelspiesse, 2019) or in the relationship (due, e.g., to cloud-contaminated retrievals) and assumes linearity (rather than just monotonicity) in the relationship. For most seasons and sensors, there is a decrease in

correlation around 0.05 with increasing temporal window. This is consistent with σ in Figures 10d–10f introducing additional variation in the comparison, although as the decline is modest suggests that in this case it is a fairly small component of the total variation. This is a larger sensitivity than found by Virtanen et al. (2018) for the eastern United States, consistent with the spatial patterns shown in Figure 7.

This is further supported by Figures 10g–10i, which shows the root-mean-square error (RMSE) of the retrieval. This is estimating by subtracting in quadrature the AERONET direct-Sun AOD uncertainty (about 0.01, Eck et al., 1999) and mean σ (Figures 10d–10f) from the observed root-mean-square difference. Subtracting these components provides a better estimate of the actual retrieval error, as recommended by Virtanen et al. (2018) and Sayer et al. (2020). This shows roughly a factor of 3 seasonal variation but is fairly insensitive to time window (within uncertainties); in part this is because σ is accounted for in the calculation, and in part this is because retrieval errors for this site are fairly large (especially in DJF for MODIS/MISR and JJA for all sensors). JJA coincides with the Indian monsoon so it is plausible that cloud masking difficulties and/or 3-D radiative transfer effects (unaccounted for in retrieval algorithms) lead to increased retrieval error.

Figures 10m–10r show the percentage of matchups achieving the GCOS goal AOD error of the greater of 0.03 or 10% (GCOS, 2011). The former panels (m–o) are for this requirement directly, while the latter (p–r) account for AERONET and temporal representation uncertainty as in the RMSE panels. While Figures 10m–10o shows a 5–10% decline with increasing time window, there is fairly large uncertainty at the shortest windows due to the small sample sizes, especially for MISR and POLDER). In Figures 10p–10r this is somewhat more stable with time, as expected when comparison uncertainty is accounted for.

From this perspective, the commonly used 30-min window seems reasonable for the MODIS algorithm at this site. It strikes a balance between data volume (being around the knee of the matchups-vs.-time curves in Figures 10a–10c) and temporal mismatch uncertainty. If retrieval errors were significantly smaller or temporal variation somewhat larger, then a shorter threshold might be needed to avoid apparent degradation in retrieval performance at the cost of lower data volume. The choice is more difficult for sensors for which the number of matchups is more limited (e.g., MISR and POLDER, especially at shorter windows). Longer windows make estimation of quantities such as RMSE and GCOS success rate more precise, but at the cost of degrading the apparent correlation against AERONET.

4. Conclusions

It is difficult to robustly assess the uncertainty in a data set through a comparison exercise without a thorough understanding of the uncertainty introduced by the comparison technique itself. This analysis builds on earlier work by Ichoku et al. (2002), Schutgens et al. (2017), Virtanen et al. (2018), Sayer et al. (2020), and others in developing and refining methods to evaluate satellite data by providing a first attempt to quantify the temporal component of such uncertainties, for the global AERONET network, on time scales relevant for validation and similar analyses. It cannot provide the unknown true temporal mismatch for every given case. Rather, it provides a typical value that can be used in a statistical sense to estimate mismatch uncertainties, and to guide decisions made when using or comparing these data sets (whether satellite vs. ground-based validation exercises, aggregation for data assimilation, or comparisons between satellite and/or model data fields).

These results are provided for each AERONET site freely downloadable from Sayer (2020) to encourage their dissemination and use. One can (and should) account for this mismatch in validation exercises. This can be done, for example, when estimating RMSE of a data set from observed route mean squared deviation (which is rarely done) or when assessing compliance of a data set with either its own uncertainty estimates (Sayer et al., 2020) or GCOS goal uncertainties (GCOS, 2011). The importance of this is dependent on the characteristics of the AERONET site (with largest and most rapid mismatch uncertainties in high-AOD conditions, and especially during biomass burning periods) as well as the error characteristics of the satellite retrieval (or model simulation) at the relevant locations and times. For the median site and a 30-min time mismatch, this is larger than the nominal AERONET AOD uncertainty of around 0.01. For mismatches around 15 min more commonly encountered in validation exercises (due to the AERONET sampling cadence), typical values range from 0.009 to 0.025 across the bulk of sites.

There is an inherent tension between increasing spatial/temporal windows to increase data volume (providing more precise estimates of validation metrics), at the cost of increased representation uncertainty (which at the same time degrades the quality of some of those estimates). The example here concerned a long-term AERONET site and mature missions (roughly 20-year records from MODIS and MISR and 9 years from POLDER)—in some senses a best case for validation data volumes at a given site. Even at these lengths the uncertainty on some validation metrics was significant, especially for MISR. This implies that this trade-off may be more difficult to balance for a shorter-term AERONET site, or for a new satellite mission (where it is often more critical to assess instrument and algorithm performance).

While the focus is on total AOD, results are also provided for AE, FMF, and fine- and coarse-mode AODs. It is worth mentioning that uncertainty estimates for these quantities in general are less widely available, so the present analysis presents a general quantification of the contribution from noise and temporal mismatch uncertainty.

The results also have implications for longer-term (e.g., monthly) composites of aerosol properties, which are commonly used for climate applications. The time to decorrelation for AOD is typically 2–8 days, while for AE and FMF is longer and more variable (3–33 days). This means that data sets with daily or near-daily repeat coverage (such as MODIS) are expected to be highly autocorrelated from day to day, while those from narrower-swath (e.g., the ATSRs, MISR) or curtain sampling (e.g., CALIOP) instruments may be considered independent observations for monthly aggregates. This makes a contribution toward the development of uncertainty estimates for L3 satellite data products.

Several avenues for future research follow from this work. One might be to extend work such as Schutgens, Gryspeerdt, et al. (2016) and build a model to assess the spatial component of mismatch uncertainty in validation exercises and provide a similar quantification. Another would be to attempt to extend it (spatially and temporally, and potentially also as a function of time of day) to full global coverage, which would enable better quantification of mismatch uncertainties for purposes such as data aggregation to time steps for model data assimilation or daily L3 satellite product comparison. This might be accomplished with a high-fidelity high-resolution Earth system model simulation, such as the Goddard Earth Observing System Model 5 Nature Run (G5NR, Putman et al., 2014). The G5NR has already been used to provide realistic aerosol fields for purposes such as observing system simulation experiments (Castellanos et al., 2018), assessing the shapes of AOD distributions for data aggregation (Sayer & Knobelspiesse, 2019), and assessing representation errors on larger scales (Schutgens, 2020). Finally, although the present study is focused on aerosol optical properties, the technique can in principle be applied to other data sets with available validation time series such as remote sensing reflectance (Zibordi et al., 2009).

Data Availability Statement

AERONET data are freely available from https://aeronet.gsfc.nasa.gov (Version 3, Level 2.0 direct Sun and SDA products), and MISR and MODIS satellite data are freely available following registration from https://earthdata.nasa.gov (MIL2ASAEv003 and MYD04_L2 Collection 6.1, respectively). POLDER data are also freely available following registration from https:/www.grasp-open.com (Version 2.1 "models" L2 processing). In each case the available data records from the start of data sets to end of 2019 were obtained. The semivariogram fits and derived parameters generated in this work are freely available to download from Sayer (2020).

Acknowledgments

This research was funded by the NASA PACE project. The AERONET team and site PIs are thanked for the creation and stewardship of the Sun photometer data and satellite retrieval teams and data centers for the satellite products used. This manuscript was prepared using the Overleaf online Editor (https://overleaf.com). The author is grateful to Prof. Noel Cressie (University of Wollongong, Australia) for useful discussions and suggested reading and Dr. P. Jeremy Werdell (NASA GSFC) for comments and questions on a draft of this work. The author thanks one anonymous referee for his/her helpful review of this manuscript. Zenodo is thanked for hosting the data.

References

Alexandrov, M. D., Geogdzhayev, I. V., Tsigaridis, K., Marshak, A., & Levy, R. (2016). New statistical model for variability of aerosol optical thickness: Theory and application to MODIS data over ocean. *Journal of the Atmospheric Sciences*, *73*, 821–837. https://doi.org/10.1175/JAS-D-15-0130.1

Alexandrov, M. D., Marshak, A., Cairns, B., Lacis, A. A., & Carlson, B. E. (2004). Scaling properties of aerosol optical thickness retrieved from Ground-Based measurements. *Journal of the Atmospheric Sciences*, *61*, 1024–1039. https://doi.org/10.1175/1520-0469(2004)061<1024:SPOAOT>2.0.CO;2

Anderson, T. L., Charlson, R. J., Winker, D. M., Ogren, J. A., & Holmén, K. (2003). Mesoscale variations of tropospheric aerosols. *Journal of the Atmospheric Sciences*, *60*(1), 119–136. https://doi.org/10.1175/1520-0469(2003)060<0119:MVOTA>2.0.CO;2

Bachmaier, M., & Backes, M. (2011). Variogram or semivariogram? Variance or semivariance? Allan variance or introducing a new term? *Mathematical Geosciences*, *43*, 735–740. https://doi.org/10.1007/s11004-011-9348-3

Benedetti, A., Reid, J. S., Knippertz, P., Marsham, J. H., Di Giuseppe, F., Rémy, S., et al. (2018). Status and future of numerical atmospheric aerosol prediction with a focus on data requirements. *Atmospheric Chemistry and Physics*, *18*, 10,615–10,643. https://doi.org/10.5194/acp-18-10615-2018

Castellanos, P., Da Silva, A. M., Darmenov, A. S., Buchard, V., Govindaraju, R. C., Ciren, P., & Kondraguntha, S. (2018). A geostationary instrument simulator for aerosol observing system simulation experiments. *Atmosphere*, *10*(1), 2. https://doi.org/10.3390/atmos10010002

Chatterjee, A., Michalak, A. M., Kahn, R. A., Paradise, S. R., Braverman, A. J., & Miller, C. E. (2010). A geostatistical data fusion technique for merging remote sensing and ground-based observations of aerosol optical thickness. *Journal of Geophysical Research*, *115*, D20207. https://doi.org/10.1029/2009JD013765

Cressie, N., & Wikle, C. K. (2011). *Statistics for spatio-temporal data*. Hoboken, NJ, USA: Wiley.

Curran, P. J. (1988). The semivariogram in remote sensing: An introduction. *Remote Sensing of Environment*, *24*(3), 493–507. https://doi.org/10.1016/0034-4257(88)90021-1

Davis, A., Marshak, A., Wiscombe, W., & Cahalan, R. (1994). Multifractal characterizations of nonstationarity and intermittency in geophysical fields: Observed, retrieved, or simulated. *Journal of Geophysical Research*, *99*(D3), 8055–8072. https://doi.org/10.1029/94JD00219

Dey, S., & Di Girolamo, L. (2011). A decade of change in aerosol properties over the Indian subcontinent. *Geophysical Research Letters*, *38*, L14811. https://doi.org/10.1029/2011GL048153

Dubovik, O., Herman, M., Holdak, A., Lapyonok, T., Tanré, D., Deuzé, J. L., et al. (2011). Stastitically optimized inversion algorithm for enhanced retrieval of aerosol properties from spectral multi-angle polarimetric satellite observations. *Atmospheric Measurement Techniques*, *4*, 975–1018. https://doi.org/10.5194/amt-4-975-2011

Dubovik, O., Lapyonok, T., Litvinov, P., Herman, M., Fuertes, D., Ducos, F., et al. (2014). GRASP: A versatile algorithm for characterizing the atmosphere. *SPIE: Newsroom*. https://doi.org/10.1117/2.1201408.005558

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

Eck, T. F., Holben, B. N., Sinyuk, A., Pinker, R. T., Goloub, P., Chen, H., et al. (2010). Climatological aspects of the optical properties of fine/coarse mode aerosol mixtures. *Journal of Geophysical Research*, *115*, D19205. https://doi.org/10.1029/2010JD014002

Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, *7*(1), 1–26. https://doi.org/10.1214/aos/1176344552

GCOS (2011). Systematic observation requirements for satellite-based data products for climate, 2011 update: World Meteorological Organization. https://library.wmo.int/doc_num.php?explnum_id=3710 Climate Observing System report GCOS-154.

Garay, M. J., Witek, M. L., Kahn, R. A., Seidel, F. C., Limbacher, J. A., Bull, M. A., et al. (2020). Introducing the 4.4 km spatial resolution Multi-Angle Imaging SpectroRadiometer (MISR) aerosol product. *Atmospheric Measurement Techniques*, *13*, 593–628. https://doi.org/10.5194/amt-13-593-2020

Giles, D. M., Holben, B. N., Eck, T. F., Sinyuk, A., Smirnov, A., Slutsker, I., et al. (2012). An analysis of AERONET aerosol absorption properties and classifications representative of aerosol source regions. *Journal of Geophysical Research*, *117*, D17203. https://doi.org/10.1029/2012JD018127

Giles, D. M., Holben, B. N., Tripathi, S. N., Eck, T. F., Newcomb, W., Slutsker, I., et al. (2011). Aerosol properties over the IndoGangetic plain: A mesoscale perspective from the TIGERZ experiment. *Journal of Geophysical Research*, *116*, D18203. https://doi.org/10.1029/2011JD015809

Giles, D. M., Sinyuk, A., Sorokin, M. G., Schafer, J. S., Smirnov, A., Slutsker, I., et al. (2019). Advancements in the Aerosol Robotic Network (AERONET) version 3 database automated near-real-time quality control algorithm with improved cloud screening for Sun photometer aerosol optical depth (AOD) measurements. *Atmospheric Measurement Techniques*, *12*, 169–209. https://doi.org/10.5194/amt-12-169-2019

Gupta, P., Levy, R. C., Mattoo, S., Remer, L. A., & Munchak, L. A. (2016). A surface reflectance scheme for retrieving aerosol optical depth over urban surfaces in MODIS Dark Target retrieval algorithm. *Atmospheric Measurement Techniques*, *9*, 3293–3308. https://doi.org/10.5194/amt-9-3293-2016

Haslett, J. (1997). On the sample variogram and the sample autocovariance for non-stationary time series. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *46*, 475–484. https://doi.org/10.1111/1467-9884.00101

Hewson, M., McGowan, H., Phinn, S., Peckham, S., & Grell, G. (2013). Exploring aerosol effects on rainfall for Brisbane, Australia. *Climate*, *1*, 120–147. https://doi.org/10.3390/cli1030120

Holben, B. N., Eck, T. F., Slutsker, I., Tanré, D., Buis, J. P., Setzer, A., et al. (1998). AERONET: A federated instrument network and data archive for aerosol characterization. *Remote Sensing of Environment*, *66*, 1–16. https://doi.org/10.1016/S0034-4257(98)00031-5

Hsu, N. C., Gautam, R., Sayer, A. M., Bettenhausen, C., Li, C., Jeong, M. J., et al. (2012). Global and regional trends of aerosol optical depth over land and ocean using SeaWiFS measurements from 1997 to 2010. *Atmospheric Chemistry and Physics*, *12*, 8037–8053. https://doi.org/10.5194/acp-12-8037-2012

Hsu, N. C., Lee, J., Sayer, A. M., Kim, W., Bettenhausen, C., & Tsay, S.-C. (2019). VIIRS Deep Blue aerosol products over land: Extending the EOS long term aerosol data record. *Journal of Geophysical Research: Atmospheres*, *124*, 4026–4053. https://doi.org/10.1029/2018JD029688

Ichoku, C., Chu, D. A., Mattoo, S., Kaufman, Y. J., Remer, L. A., Tanré, D., et al. (2002). A spatio-temporal approach for global validation and analysis of MODIS aerosol products. *Geophysical Research Letters*, *29*(12), 1616. https://doi.org/10.1029/2001GL013206

Jethva, H., Satheesh, S. K., & Srinivasan, J. (2007). Assessment of second generation MODIS aerosol retrieval (Collection 005) at Kanpur, India. *Geophysical Research Letters*, *34*, L19802. https://doi.org/10.1029/2007GL029647

Jethva, H., Satheesh, S. K., Srinivasan, J., & Levy, R. C. (2010). Improved retrieval of aerosol size-resolved properties from Moderate Resolution Imaging Spectroradiometer over India: Role of aerosol model and surface reflectance. *Journal of Geophysical Research*, *115*, D18213. https://doi.org/10.1029/2009JD013218

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

Kaku, K. C., Reid, J. S., Hand, J. L., Edgerton, E. S., Holben, B. N., Zhang, J., & Holz, R. E. (2018). Assessing the challenges of surface level aerosol mass estimates from remote sensing during the SEAC4RS and SEARCH campaigns: Baseline surface observations and remote sensing in the southeastern United States. *Journal of Geophysical Research: Atmospheres*, *123*, 7530–7562. https://doi.org/10.1029/2017JD028074

Kim, D., Chin, M., Yu, H., Diehl, T., Tan, Q., Kahn, R. A., et al. (2014). Sources, sinks, and transatlantic transport of North African dust aerosol: A multimodel analysis and comparison with remote sensing data. *Journal of Geophysical Research: Atmospheres*, *119*, 6259–6277. https://doi.org/10.1002/2013JD021099

Kinne, S., Schulz, M., Textor, C., Balkanski, Y., Bauer, S., Berntsen, T., et al. (2006). An AeroCom initial assessment—Optical properties in aerosol component modules of global models. *Atmospheric Chemistry and Physics*, *6*, 1815–1834. https://doi.org/10.5194/acp-6-1815-2006

Kleidman, R. G., O'Neill, N. T., Remer, L. A., Kaufman, Y. J., Eck, T. F., Tanré, D., et al. (2005). Comparison of Moderate Resolution Imaging Spectroradiometer (MODIS) and Aerosol Robotic Network (AERONET) remote-sensing retrievals of aerosol fine mode fraction over ocean. *Journal of Geophysical Research*, *110*, D22205. https://doi.org/10.1029/2005JD005760

Markwardt, C. B. (2009). Non-Linear least squares fitting in IDL with MPFIT. In D. Bohlender, P. Dowler, & D. Durand (Eds.), *Astronomical Society of the Pacific* (Vol. 411, pp. 251–254). San Francisco. https://pages.physics.wisc.edu/~craigm/idl [Accessed April 2 2020].

Martin, S., Andreae, M. O., Artaxo, P., Baumgardner, D., Chen, Q., Goldstein, A. H., et al. (2010). Sources and properties of Amazonian aerosol particles. *Reviews of Geophysics*, *48*, RG2002. https://doi.org/10.1029/2008RG000280

Nguyen, H., Cressie, N., & Braverman, A. (2012). Spatial statistical data fusion for remote-sensing applications. *Journal of the American Statistical Association*, *107*, 1004–1018. https://doi.org/10.1080/01621459.2012.694717

O'Neill, N. T., Eck, T. F., Smirnov, A., Holben, B. N., & Thulasiraman, S. (2003). Spectral discrimination of coarse and fine mode optical depth. *Journal of Geophysical Research*, *108*(D17), 4559. https://doi.org/10.1029/2002JD002975

O'Neill, N. T., Eck, T., Smirnov, A., Holben, B., & Thulasiraman, S. (2008). Spectral deconvolution algorithm technical memo: NASA Goddard Space Flight Center. Revision April 9, 2008, version 4.1, available online from https://aeronet.gsfc.nasa.gov/new_web/PDF/tauf_tauc_technical_memo1.pdf [Accessed April 2020].

O'Neill, N. T., Ignatov, A., Holben, B. N., & Eck, T. F. (2000). The lognormal distribution as a reference for reporting aerosol optical depth statistics; Empirical tests using multi-year, multi-site AERONET Sunphotometer data. *Geophysical Research Letters*, *27*(20), 3333–3336. https://doi.org/10.1029/2000GL011581

Putman, W., da Silva, A., Ott, L. E., & Darmenov, A. (2014). Model configuration for the 7-km GEOS-5 Nature Run, Ganymed release: Goddard Space Flight Center, National Aeronautics and Space Administration. Technical Report GMAO Office Note No 5 (Version 1.0).

Queface, A. J., Piketh, S. J., Eck, T. F., Tsay, S.-C., & Mavume, A. F. (2011). Climatology of aerosol optical properties in southern Africa. *Atmospheric Environment*, *45*(17), 2910–2921. https://doi.org/10.1016/j.atmosenv.2011.01.056

Ramanathan, V., Ramana, M. V., Roberts, G., Kim, D., Corrigan, C. E., Chung, C. E., & Winker, D. (2007). Warming trends in Asia amplified by brown cloud solar absorption. *Nature*, *448*, 575–578. https://doi.org/10.1038/nature06019

Sand, M., Samset, B. H., Balkanski, Y., Bauer, S., Bellouin, N., Berntsen, T. K., et al. (2017). Aerosols at the poles: An AeroCom Phase II multi-model evaluation. *Atmospheric Chemistry and Physics*, *17*, 12,197–12,218. https://doi.org/10.5194/acp-17-12197-2017

Sayer, A. M. (2020). Supporting information for "How long is too long? Variogram analysis of AERONET data to aid aerosol validation and intercomparison studies". https://doi.org/10.5281/zenodo.3988487

Sayer, A. M., Govaerts, Y., Kolmonen, P., Lipponen, A., Luffarelli, M., Mielonen, T., et al. (2020). A review and framework for the evaluation of pixel-level uncertainty estimates in satellite aerosol remote sensing. *Atmospheric Measurement Techniques*, *13*, 373–404. https://doi.org/10.5194/amt-13-373-2020

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

Sayer, A. M., Hsu, N. C., Lee, J., Kim, W. V., Dubovik, O., Dutcher, S. T., et al. (2018). Validation of SOAR VIIRS over water aerosol retrievals and context within the global satellite aerosol data record. *Journal of Geophysical Research: Atmospheres*, *123*, 13,496–13,526. https://doi.org/10.1029/2018JD029465

Sayer, A. M., Hsu, N. C., Lee, J., Kim, W., & Dutcher, S. (2019). Validation, stability, and consistency of MODIS Collection 6.1 and VIIRS Version 1 Deep Blue aerosol data over land. *Journal of Geophysical Research: Atmospheres*, *124*, 4658–4688. https://doi.org/10.1029/2018JD029598

Sayer, A. M., & Knobelspiesse, K. D. (2019). How should we aggregate data? methods accounting for the numerical distributions, with an assessment of aerosol optical depth. *Atmospheric Chemistry and Physics*, *19*, 15,023–15,048. https://doi.org/10.5194/acp-19-15023-2019

Sayer, A. M., Munchak, L. A., Hsu, N. C., Levy, R. C., Bettenhausen, C., & Jeong, M.-J. (2014). MODIS Collection 6 aerosol products: Comparison between Aqua's e-Deep Blue, Dark Target, and merged data sets, and usage recommendations. *Journal of Geophysical Research: Atmospheres*, *119*, 13,965–13,989. https://doi.org/10.1002/2014JD022453

Schmit, T. J., Griffith, P., Gunshor, M. M., Daniels, J. M., Goodman, S. J., & Lebair, W. J. (2017). A closer look at the ABI on the GOES-R series. *Bulletin of the American Meteorological Society*, *98*, 681–698. https://doi.org/10.1175/BAMS-D-15-00230.1

Schonbrödt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality*, *47*(5), 609–612. https://doi.org/10.1016/j.jrp.2013.05.009

Schuster, G. L., Dubovik, O., & Holben, B. N. (2006). Angstrom exponent and bimodal aerosol size distributions. *Journal of Geophysical Research*, *111*, D07207. https://doi.org/10.1029/2005JD006328

Schutgens, N. A. J. (2020). Site representativity of AERONET and GAW remotely sensed AOT and AAOT observations. *Atmospheric Chemistry and Physics*, *20*, 7473–7488. https://doi.org/10.5194/acp-20-7473-2020

Schutgens, N. A. J., Gryspeerdt, E., Weigum, N., Tsyro, S., Goto, D., Schulz, M., & Stier, P. (2016). Will a perfect model agree with perfect observations? The impact of spatial sampling. *Atmospheric Chemistry and Physics*, *16*, 6335–6353. https://doi.org/10.5194/acp-16-6335-2016

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

Schutgens, N. A. J., Tsyro, S., Gryspeerdt, E., Goto, D., Weigum, N., Schulz, M., & Stier, P. (2017). On the spatio-temporal representativeness of observations. *Atmospheric Chemistry and Physics*, *17*, 9761–9780. https://doi.org/10.5194/acp-17-9761-2017

Shi, T., & Cressie, N. (2007). Global statistical analysis of MISR aerosol data: A massive data product from NASA's Terra satellite. *Environmetrics*, *18*, 665–680. https://doi.org/10.1002/env.864

Singh, R. P., Dey, S., Tripathi, S. N., Tare, V., & Holben, B. (2004). Variability of aerosol parameters over Kanpur, northern India. *Journal of Geophysical Research*, *109*, D23206. https://doi.org/10.1029/2004JD004966

Smirnov, A., Holben, B. N., Eck, T. F., Slutsker, I., Chatenet, B., & Pinker, R. T. (2002). Diurnal variability of aerosol optical depth observed at AERONET (Aerosol Robotic Network) sites. *Geophysical Research Letters*, *29*(23), 2115. https://doi.org/10.1029/2002GL016305

Snik, F., Rietjens, J. H. H., Apituley, A., Volten, H., Mijling, B., Di Noia, A., et al. (2014). Mapping atmospheric aerosols with a citizen science network of smartphone spectropolarimeters. *Geophysical Research Letters*, *41*, 7351–7358. https://doi.org/10.1002/2014GL061462

Sogacheva, L., Popp, T., Sayer, A. M., Dubovik, O., Garay, M. J., Heckel, A., et al. (2020). Merging regional and global aerosol optical depth records from major available satellite products. *Atmospheric Chemistry and Physics*, *20*, 2031–2056. https://doi.org/10.5194/acp-20-2031-2020

Sullivan, R. C., Levy, R. C., da Silva, A. M., & Pryor, S. C. (2017). Developing and diagnosing climate change indicators of regional aerosol optical properties. *Scientific Reports*, *7*, 18903. https://doi.org/10.1038/s41598-017-18402-x

Sullivan, R. C., Levy, R. C., & Pryor, S. C. (2015). Spatiotemporal coherence of mean and extreme aerosol particle events over eastern North America as observed from satellite. *Atmospheric Environment*, *112*, 126–135. https://doi.org/10.1016/j.atmosenv.2015.04.026

van Donkelaar, A., Martin, R. V., Brauer, M., Hsu, N. C., Kahn, R. A., Lecy, R. C., et al. (2016). Global estimates of fine particulate matter using a combined Geophysical-Statistical method with information from satellites, models, and monitors. *Environmental Science & Technology*, *50*(7), 3762–3772. https://doi.org/10.1021/acs.est.5b05833

Virtanen, T. H., Kolmonen, P., Sogacheva, L., Rodriguez, E., Saponaro, G., & de Leeuw, G. (2018). Collocation mismatch uncertainties in satellite aerosol retrieval validation. *Atmospheric Measurement Techniques*, *11*, 925–938. https://doi.org/10.5194/amt-11-925-2018

Wagner, F., & Silva, A. M. (2008). Some considerations about Ångström exponent distributions. *Atmospheric Chemistry and Physics*, *8*, 481–489. https://doi.org/10.5194/acp-8-481-2008

Wei, J., Peng, Y., Mahmood, R., Sun, L., & Guo, J. (2019). Intercomparison in spatial distributions and temporal trends derived from multi-source satellite aerosol products. *Atmospheric Chemistry and Physics*, *19*, 7183–7207. https://doi.org/10.5194/acp-19-7183-2019

Xian, P., Reid, J. S., Hyer, E. J., Sampson, C. R., Rubin, J. I., Ades, A., et al. (2019). Current state of the global operational aerosol multimodel ensemble: An update from the International Cooperative for Aerosol Prediction (ICAP). *Quarterly Journal of the Royal Meteorological Society*, *145*, 176–209. https://doi.org/10.1002/qj.3497

Yu, C., Chen, L., Su, L., Fan, M., & Li, S. (2011). Kriging interpolation method and its application in retrieval of MODIS aerosol optical depth. In *2011 19th International Conference on Geoinformatics*. https://doi.org/10.1109/GeoInformatics.2011.5981052

Zhang, J., & Reid, J. S. (2010). A decadal regional and global trend analysis of the aerosol optical depth using a data-assimilation grade over-water MODIS and Level 2 MISR aerosol products. *Atmospheric Chemistry and Physics*, *10*, 10,949–10,963. https://doi.org/10.5194/acp-10-10949-2010

Zibordi, G., Mélin, F., Berthon, J., Holben, B., Slutsker, I., Giles, D., et al. (2009). AERONET-OC: A network for the validation of ocean color primary products. *Journal of Atmospheric and Oceanic Technology*, *26*, 1634–1651. https://doi.org/10.1175/2009JTECHO654.1