Atmospheric
Measurement
Techniques

Application of spectral analysis techniques to the intercomparison of aerosol data – Part 4: Synthesized analysis of multisensor satellite and ground-based AOD measurements using combined maximum covariance analysis

J. Li[1,2], B. E. Carlson[1], and A. A. Lacis[1]

[1]NASA Goddard Institute for Space Studies, New York, NY, USA
[2]Department of Applied Physics and Applied Math, Columbia University, New York, NY, USA

Correspondence to: J. Li (jl2862@columbia.edu)

Abstract. In this paper, we introduce the usage of a newly developed spectral decomposition technique – combined maximum covariance analysis (CMCA) – in the spatiotemporal comparison of four satellite data sets and ground-based observations of aerosol optical depth (AOD). This technique is based on commonly used principal component analysis (PCA) and maximum covariance analysis (MCA). By decomposing the cross-covariance matrix between the joint satellite data field and Aerosol Robotic Network (AERONET) station data, both parallel comparison across different satellite data sets and the evaluation of the satellite data against the AERONET measurements are simultaneously realized. We show that this new method not only confirms the seasonal and interannual variability of aerosol optical depth, aerosol-source regions and events represented by different satellite data sets, but also identifies the strengths and weaknesses of each data set in capturing the variability associated with sources, events or aerosol types. Furthermore, by examining the spread of the spatial modes of different satellite fields, regions with the largest uncertainties in aerosol observation are identified. We also present two regional case studies that respectively demonstrate the capability of the CMCA technique in assessing the representation of an extreme event in different data sets, and in evaluating the performance of different data sets on seasonal and interannual timescales. Global results indicate that different data sets agree qualitatively for major aerosol-source regions. Discrepancies are mostly found over the Sahel, India, eastern and southeastern Asia. Results for eastern Europe suggest

that the intense wildfire event in Russia during summer 2010 was less well-represented by SeaWiFS (Sea-viewing Wide Field-of-view Sensor) and OMI (Ozone Monitoring Instrument), which might be due to misclassification of smoke plumes as clouds. Analysis for the Indian subcontinent shows that here SeaWiFS agrees best with AERONET in terms of seasonality for both the Gangetic Basin and southern India, while on interannual timescales it has the poorest agreement.

1 Introduction

Global aerosol properties are highly variable in space and time. Aerosols from different regions generally have different chemical compositions, emission sources, and are subject to different meteorological conditions. Understanding the spatial and temporal variability of aerosols is critical in quantifying their direct and indirect climate effects. Satellite observations have become and will be an indispensable source of information about aerosol characteristics for use in various assessments of climate change (King et al., 1999). In the past decade, many satellite sensors have been developed to monitor global aerosol properties and have greatly advanced our knowledge of aerosols and their variability. These aerosol products have been validated against ground-based measurements from the Aerosol Robotic Network (AERONET, Holben et al., 1998; Dubovik et al., 2002), and their data accuracy and reliability are confirmed (e.g.,

Levy et al., 2010; Kahn et al., 2005; Sayer et al., 2012; Torres et al., 2007). As a result, they have been extensively used in various aerosol and climate-related studies. For example, Kalashnikova and Kahn (2005) used Multiangle Imaging Spectroradiometer (MISR) and Moderate Resolution Imaging Spectroradiometer (MODIS) aerosol products to study mineral dust plume evolution over the Atlantic, Torres et al. (2010) studied the anomalous biomass burning in the Southern Hemisphere using aerosol retrievals from the Ozone Monitoring Instrument (OMI) and MODIS, and Hsu et al. (2012) investigated global and regional trends in aerosol optical depth using Sea-viewing Wide Field-of-view Sensor (SeaWiFS) measurements. In these studies, usually only one or two data sets were used to study the physical problem. With multiple data sets available, it is desirable to take advantage of all available pieces of information in one analysis in order to yield more reliable results. Several authors have used aerosol retrievals from multiple sensors in their study. Nabat et al. (2013) created a 4-D climatology of monthly mean aerosol optical depth over the Mediterranean using nine satellite-derived AOD (aerosol optical depth) products. Carboni et al. (2012) evaluated desert dust optical depth retrievals from eight different satellite instruments. Another application of multisensor aerosol data is to validate and constrain aerosol parameterizations in climate models. Kinne et al. (2003, 2006) compared global monthly mean aerosol properties between AeroCom aerosol modules and several satellite data sets. Liu et al. (2006) assessed the GISS (Goddard Institute for Space Studies) ModelE aerosol climatology against multiple satellite retrieval products. In these multisensor applications, although different data sets achieved an overall global agreement, considerable regional differences were revealed that were associated with different aerosol sources or transport regimes. Regional differences between satellite-retrieved aerosol properties were also reported for India (Prasad and Singh, 2007), for South America (Ahn et al., 2008), and for southeastern Asia (Xiao et al., 2009). Therefore, effective and efficient use of multisensor data sets requires an understanding of the strengths and weaknesses of each data set in representing different aerosol types and variability in different regions of the world.

Previously, we have demonstrated that spectral decomposition techniques such as principal component analysis (PCA) can be effectively used to examine the spatial and temporal variability in multidimensional aerosol observations (Li et al., 2009, 2011, 2013). Many global and regional aerosol-source regions and their seasonal and interannual variability are successfully captured by the dominant orthogonal modes. We furthermore introduced the maximum covariance analysis (MCA) method that allows verifying the variability revealed by a particular satellite data set through the comparison with ground-based measurements from AERONET (Li et al., 2014a). And in Li et al. (2014b), we applied combined principal component analysis (CPCA) to achieve a parallel examination and comparison of the

spatial and temporal variability in aerosol optical depth as measured by multiple satellite data sets. The CPCA method is powerful in both confirming the agreement and finding locations and times of disagreement between the satellite data sets. However, a major drawback is that the CPCA methodology by itself does not accommodate the inclusion of scattered ground observations, as combining different fields assumes equal weight and is thus only suitable for gridded data with the same spatial mapping. The MCA does incorporate ground-based data; however, its results alone are not sufficient to select which data set best characterizes aerosol variability for a particular region, in that the method only evaluates one satellite data set against AERONET. For multisensor data analysis, it is necessary to simultaneously examine the capability of each data set in representing aerosol variability for particular regions, in order to determine which data set or data sets provide the best constraints on the aerosol property for the regions of interest. Such information is critical in many aspects of satellite data application, such as developing aerosol parameterization schemes and extending station measurements to a broader spatial context. In this study, we develop a new technique – the combined maximum covariance analysis (CMCA) – to bridge the gap between MCA and CPCA by examining and comparing spatial and temporal variability retrieved by multiple satellite sensors as well as incorporating more randomly distributed ground-based station data such as AERONET. Compared with previous techniques, the advantages of the CMCA include (1) integrating all available information from both satellite and surface measurements resulting in a more complete view of the picture; (2) the common modes of variability revealed through CPCA can be further confirmed, and the problems in each satellite data set can be identified through the comparison with ground-truth measurements; and (3) the examination and comparison is associated with specific aerosol sources, types or events, which are essential for both understanding the physics of the problem and improving satellite retrievals.

The goals of this paper are to introduce and highlight the utility of the CMCA technique and thereby promote its usage by the aerosol data community. We describe data selection, preprocessing and the detailed analysis procedure in Sects. 2 and 3. In Sect. 4, we present the results of our global analysis and two representative regional case studies that demonstrate the usefulness of this technique, while readers are welcome to use the method to explore additional regions based on their specific interest. Finally, a summary and discussion of potential extended usage of the CMCA technique is given in Sect. 5.

2   Data sets

We use monthly mean, gridded AOD products from four satellite sensors: MODIS, MISR, OMI and SeaWiFS. These four data sets have all been validated against ground

observations and have reasonably good global coverage. Only over-land data are used primarily because the majority of AERONET stations are located over land. The ground-based observations are from 58 selected AERONET stations (considered a benchmark for satellite data). The period of study is chosen to be from January 2005 to December 2010, which corresponds to the period of the longest overlap for the four satellite data records. Finally, because OMI AOD is reported at 500 nm while MODIS, MISR and SeaWiFS report AOD at multiple wavelengths, to facilitate parallel comparison, we interpolate MODIS, MISR and SeaWiFS AOD to 500 nm according to the Ångström power law, which is a linear interpolation using AOD measured at different wavelengths on logarithm scale. The wavelengths used for each sensor are detailed below.

2.1 MODIS

The MODIS instrument is a multispectral radiometer, which has the capability of retrieving aerosol microphysical and optical properties over land and ocean (Tanré et al., 1997; Levy et al., 2007). The 2330 km swath width of the MODIS instrument produces a global coverage in 1 or 2 days and captures most of aerosol variability due to this high sampling frequency. The MODIS on the Aqua platform is used here, as Terra MODIS AOD is not as complete as Aqua over desert regions. The official Level 3 monthly mean AOD product at $1° \times 1°$ resolution is used for this study (MYD08_M3, collection 5.1, available from ftp://ladsweb.nasacom.nasa.gov/allData/51/MYD08_M3). We use QA (quality assurance) weighted averages ("*QA_Mean_Mean" variables; Hubanks et al., 2008) for both dark target (DT; Levy et al., 2010) and deep blue (DB) AOD retrievals (Hsu et al., 2004, 2006). The deep-blue algorithm covers most of the dust regions and is thus important for the analysis. Note that in the MODIS collection 6 data, merged DT and DB will be provided as a standard product (Levy et al., 2013). However, since the collection 6 data are not yet available for Level 3, we merge these two products by ourselves here. The DT and DB products are combined following the procedure described by Levy et al. (2013) for collection 6, which determines the selection of DT or DB product according to the MODIS NDVI (normalized difference vegetation index) climatology. Specifically, for NDVI > 0.3, DT data are selected, for NDVI < 0.2, DB data are selected and for $0.2 \leq NDVI \leq 0.3$, an average of DT and DB AOD is used. Nonetheless, this merging of DT and DB products will result in a seasonally varying product type for some regions with seasonally varying NDVI, especially for semiarid areas such as the Sahel, western US, northern India and eastern China. To examine the consistency of these two products, in Fig. 1 we plot the merged DT and DB AOD time series and the NDVI time series at four AERONET stations with changing vegetation type: Banizoumbou, Beijing, Bratt's Lake and Kanpur. Note the NDVI time series only shows the climatology without interannual variability, which
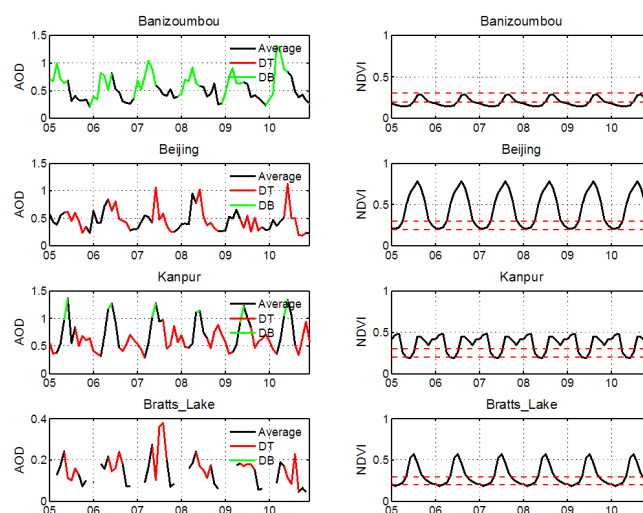

Figure 1. Merged MODIS DT and DB AOD products at four AERONET stations with seasonally varying vegetation index (NDVI). The merging of the data appears consistent.

is the same as Levy et al. (2013). We find that the time series appear rather smooth, indicating that the merging of DT and DB products has negligible influence on the overall data consistency. The MODIS AOD is interpolated to 500 nm using measurements at 470 and 660 nm.

2.2 MISR

The MISR is a multiangle sensor with nine push-broom cameras on the EOS Terra platform. The zonal overlap of the common swath of all nine cameras is at least 360 km in order to provide multiangle coverage in 9 days at the Equator, and 2 days at the poles (Diner et al., 1998). Compared to MODIS, the multiangle view of MISR performs better over bright surfaces (Kahn et al., 2005, 2010). However, MISR has a narrower swath width leading to a longer revisit time, and thus may not fully resolve short timescale variability. In this study, we use version 31 Level 3 gridded monthly products, available at http://eosweb.larc.nasa.gov. The original $0.5° \times 0.5°$ data resolution has been rescaled to $1° \times 1°$. The rescaling is performed by assigning equal weights to each subgrid, and the final $1° \times 1°$ grid is considered valid only when more than half of the subgrids have valid data. The data are also interpolated to 500 nm using measurements at the four MISR wavelengths of 446, 555, 672 and 865 nm.

2.3 SeaWiFS

The SeaWiFS instrument was launched on the SeaStar spacecraft in 1997. It is also a wide view imager with a swath width of 1502 km and covered the globe in approximately 2 days. The SeaWiFS over-land aerosol retrieval uses the deep blue algorithm developed by Hsu et al. (2004, 2006). The AOD data over land have been validated using AERONET

measurements (Sayer et al., 2012). Here we use the standard Level 3 monthly mean AOD product version 004, available at http://mirador.gsfc.nasa.gov/. The data are converted to 500 nm using the reported AOD values at 412, 490 and 670 nm.

2.4 OMI

The OMI sensor (Levelt et al., 2006) on the EOS Aura satellite has been providing global aerosol measurements since October 2005. The OMI instrument also has a wide swath of 2600 km and produces daily global coverage. The AOD data used here are derived from the UV (ultraviolet) algorithm (OMAERUV; Torres et al., 2007, 2013). OMAERUV makes use of the instrument's two near-UV channels to retrieve AOD and single scattering albedo at 388 nm, and the 500 nm AOD reported in the standard product is converted according to the spectral dependence of the assumed aerosol model (Torres et al., 2007, 2013; Ahn et al., 2008). While the reliability of the 500 nm AOD is affected by aerosol model assumptions, comparisons with AERONET, MODIS and MISR showed reasonable agreement (Torres et al., 2007; Ahn et al., 2008). Moreover, the upgraded OMI algorithm by Torres et al. (2013), which made use of aerosol layer information derived from CALIPSO (Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation) and AIRS (Atmospheric Infrared Sounder), produced noticeable improvements on the retrieval of dust and smoke aerosols. Additionally, the evaluation work by Ahn et al. (2014) on the upgraded algorithm indicated improved agreements with ground-based observations and comparable accuracy with MODIS deep blue algorithm and MISR retrievals over arid and semiarid areas. Here we use collection 003 data from the upgraded algorithm at $1° \times 1°$ spatial resolution, available from the Goddard Earth Sciences Data and Information Services Center (http://mirador.gsfc.nasa.gov/). Note that as the current OMAERUV algorithm does not explicitly account for ocean color effects, retrievals over ocean are limited to absorbing aerosols as identified by the aerosol index. Therefore, it is only used over land and in regional analyses. The wide swath of OMI provides daily global coverage. However, its relatively large footprint ($13 \times 24\,km^2$ at nadir) makes cloud contamination a more serious issue in OMI retrievals (Torres et al., 2007).

2.5 AERONET

AERONET (Holben et al., 1998) is a ground-based sun-photometer network with over 400 stations globally. The AERONET AOD is derived from direct beam solar measurements (Holben et al., 2001) at two UV channels at 340 and 380 nm, and five visible channels at 440, 500, 675, 870 and 1020 nm. The measurements from AERONET are usually regarded as ground truth when assessing satellite retrievals of aerosol properties. In this study, we also consider the AOD

variability represented by AERONET data as the benchmark against which we evaluate the different satellite data sets. The data used are the version 2 Level 2 quality-assured and cloud-screened (Smirnov et al., 2000; Holben et al., 2006) monthly mean AOD product. As the CMCA technique requires the construction of the temporal cross-covariance matrix, the completeness of the AERONET AOD time series is critical to the success of the analysis. Therefore, we select stations primarily based on the availability of a continuous data record for the study period of 2005 to 2010. Three steps are involved in the selection and quality control of AERONET data: (1) data from all stations are automatically screened by a threshold of at least eight monthly mean data points each year from 2005 to 2010; (2) the selected stations are further manually screened by removing stations with relatively large gaps (≥ 3 months) in the time series. This is because we need to interpolate to fill the gaps and generally interpolation with gaps greater than three data points will result in large uncertainty. (3) A few stations that do not strictly meet the above criteria are added to account for regions with representative aerosol variability. These stations are primarily based in Asia, including Pune and Gandhi College in India, Mukdahan in Thailand and Singapore. A total of 58 stations are selected globally. Figure 2 shows the distribution and associated aerosol types of the selected stations, and Table 1 lists the station name, location, aerosol type and the number of available monthly mean data points. The aerosol-type information for the AERONET stations is mostly obtained from existing references including Kinne et al. (2003), Kahn et al. (2010), and Garcia et al. (2012). For several stations, not found in the literature, the aerosol type is inferred from the station description available on the AERONET website (http://aeronet.gsfc.nasa.gov/). The AERONET AOD is converted to 500 nm using measurements from 380 to 870 nm by applying a second-order polynomial fitting of ln (AOD) vs. ln (wavelength), as recommended by Eck et al. (1999).

3 Methodology

3.1 Treatment of missing data

As mentioned in Sect. 2.5, the completeness of time series is critical to the construction of the temporal covariance matrix. For AERONET data, we apply a linear interpolation to the time series to fill the gaps. The interpolation is performed on the deseasonalized data constructed by removing the multiyear average seasonal cycle, so that the influence of interpolation on the seasonal variability will be minimized. The full data series is then reconstructed by adding the seasonal cycle back. To test the impact of interpolation on the variability of the original time series, we perform cross-validation using the time series from eight stations (Banizoumbou, Bratt's Lake, CARTEL, Capo Verde, Carpentras, Dakar, GSFC and IER Cinzana) which have less than two missing data points in
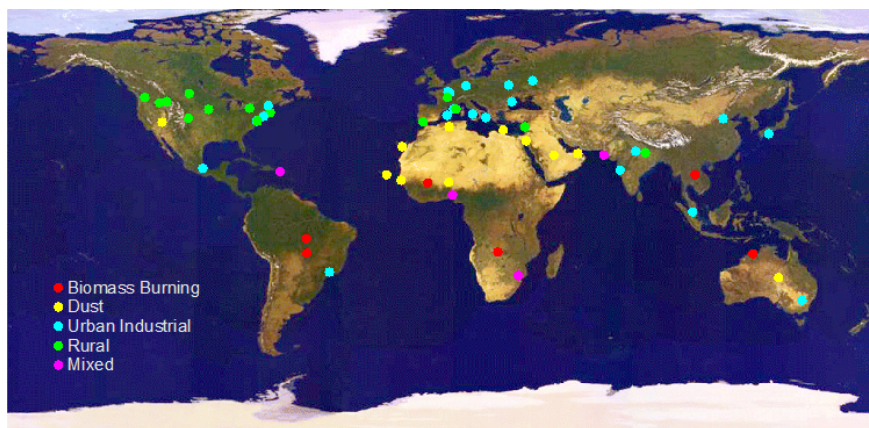
Figure 2. Locations and aerosol types of the 58 selected stations used in this analysis.

the entire study period. Specifically, tests are conducted for gaps with one to three missing data conditions. For the cases with one missing data point, each time one data point is removed from the original time series and treated as missing data. Interpolation is performed on the rest of the time series, and the mean square error (MSE) is calculated between the interpolated and original time series. The procedure is repeated until every data point has been removed once. The process is the same for two or three missing data point cases except each time two or three consecutive data points are treated as missing. The average of the MSE in then compared with the variance (representing variability) of the original time series. We find that for almost all stations, interpolation for the one or two missing data point cases results in a less than 3 % perturbation on the variability, while with three data point gaps, the error can increase to 8 %, although the magnitude of the influence depends on the variability of the original time series. In Fig. 3, we show an example for the IER Cinzana station. Three random gaps with one, two and three missing data points, respectively, are imposed on the original time series, and the comparisons between the interpolated and raw time series are shown. It is seen that the interpolation generally captures the variability of the raw time series for the one and two data point gap cases. While with three data point gaps, interpolation clearly misses the peak in March 2009. Note that the variability at IER Cinzana is relatively high compared to that at many of the other stations. Thus the interpolation performs better for stations with less variability. We therefore consider that interpolating over one or two missing data point intervals should not seriously impact the results of this analysis.

For the satellite data, we focus on the 60° S–60° N domain where the monthly mean products have nearly full coverage. Nonetheless, we do find that there a few regions with continually missing data. These regions include the Tibet Plateau for SeaWiFS and OMI and the intertropical convergence zone for SeaWiFS. For these regions, we apply a data
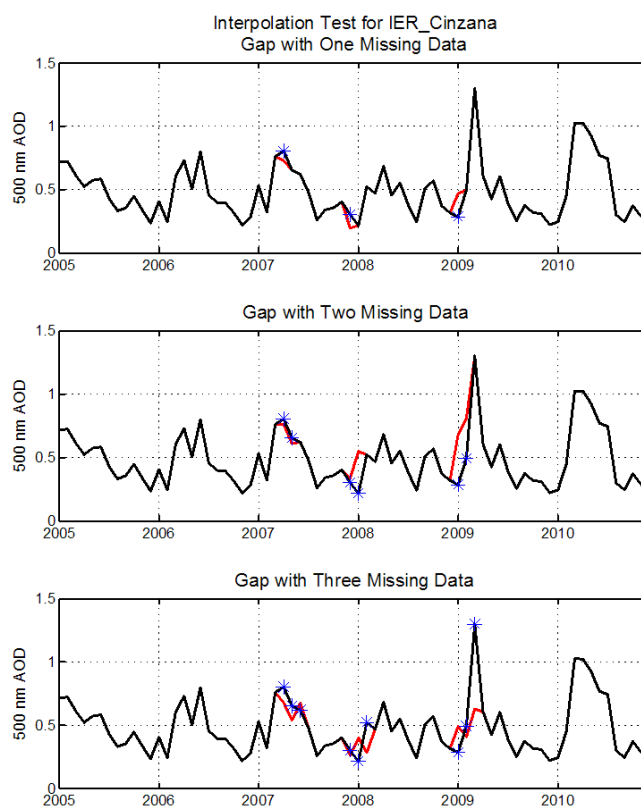

Figure 3. Interpolation test using data for the IER Cinzana station. Three gaps with one, two and three data points are removed and interpolated back. The black curve shows the original time series and the red parts are interpolated. The blue stars indicate missing data for each experiment.

availability mask to each monthly mean map to exclude them from the analysis. Figure 3 shows the mask of the four data sets. The hole of central Australia on the MODIS data mask is not due to missing data but because of negative values in the DT retrievals. The MODIS DT algorithm allows small

Table 1. Location, aerosol type, and number of monthly means used in this analysis (N) for the selected AERONET stations.

Station name	Longitude	Latitude	Type	N
North America				
Billerica	−71.269	42.528	Rural	62
CARTEL	−71.931	45.379	Urban	70
CCNY	−73.949	40.821	Urban	60
Egbert	−79.75	44.226	Rural	65
GSFC	−76.84	38.992	Urban	72
MD Science Center	−76.617	39.283	Urban	67
SERC	−76.500	38.883	Rural	58
Bratt's Lake	−104.7	50.28	Rural	70
BSRN BAO Boulder	−105.006	40.045	Rural	68
Railroad Valley	−115.962	38.504	Dust	63
Rimrock	−116.992	46.487	Rural	65
Saturn Island	−123.133	48.783	Rural	64
Sioux Falls	−96.626	43.736	Rural	58
South America				
Alta Floresta	−56.104	−9.871	Biomass	62
CUIABA-MIRANDA	−56.021	−15.729	Biomass	56
Sao Paulo	−46.735	−23.561	Mixed[a]	58
Europe				
Avignon	4.878	43.933	Urban	62
Barcelona	2.117	41.386	Urban	67
Carpentras	5.058	44.083	Urban	70
Dunkerque	2.368	51.035	Urban	67
Forth Crete	25.282	35.333	Dust	69
Hamburg	9.973	53.568	Urban	59
IMS-METU-ERDEMLI	34.255	36.565	Rural	60
Lecce University	18.111	40.335	Urban	61
Lille	3.142	50.612	Urban	69
Minsk	27.601	53.92	Urban	64
Missoula	−114.083	46.917	Rural	68
Moldova	28.816	47.000	Urban	62
Moscow MSU MO	37.510	55.700	Urban	64
OHP OBSERVATOIRE	5.71	43.935	Rural	62
Palaiseau	2.208	48.700	Rural	67
Rome Tor Vergata	12.647	41.84	Urban	64
Africa				
Banizoumbou	2.665	13.541	Dust	71
Blida	2.881	36.508	Dust	68
Capo Verde	−22.935	16.733	Dust	71
Dakar	−16.969	14.394	Dust	70
Evora	−7.912	38.568	Rural	61
IER Cinzana	−5.934	13.278	Mixed[b]	72
Ilorin	4.34	8.32	Mixed[b]	62
Izana	−16.499	28.309	Dust	72
Mongu	23.151	−15.254	Biomass	59
Skukuza	31.587	−24.992	Mixed[c]	63

[a] mixture of urban industrial and biomass burning, [b] mixture of dust and biomass burning,
[c] mixture of urban industrial and biomass burning, [d] mixture of urban industrial and dust,
[e] mixture of urban industrial and oceanic.

Table 1. Continued.

Station name	Longitude	Latitude	Type	N
Asia				
Kanpur	80.232	26.513	Urban	63
Gandhi College	84.128	25.871	Rural	40
Pune	73.805	18.537	Urban	45
Karachi	67.030	24.870	Mixed[d]	56
Beijing	116.381	39.997	Urban	66
Shirahama	135.357	33.693	Urban	68
Singapore	103.78	1.298	Urban	56
Mukdahan	104.676	16.607	Biomass	58
SEDE BOKER	34.782	30.855	Dust	69
Dhadnah	56.325	25.513	Dust	59
Solar Village	46.397	24.907	Dust	66
Australia				
Birdsville	139.346	−25.899	Dust	59
Canberra	149.111	−35.271	Urban	58
Lake Argyle	128.749	−16.108	Biomass	59
Central America				
La Parguera	−67.045	17.97	Mixed[e]	64
Mexico City	−99.182	19.334	Urban	58

sets and because converting these AODs to 500 nm on a logarithmic scale leads to imaginary numbers, we removed the central Australia data from the MODIS data. Overall, the removed data only account for a small portion of the global map and do not affect the results in other regions.

3.2 Combined maximum covariance analysis

The CMCA technique can be viewed as a combination of MCA and CPCA analyses. A detailed mathematical description of PCA and MCA (a.k.a. SVD – singular value decomposition) analyses can be found in Björnsson and Venegas (1997). Bretherton et al. (1992) provided a comprehensive intercomparison of these methods in the analysis of climate data. MCAs have been widely used to find coupled modes between observations of different variables, such as sea surface temperature and wind stress (Chang et al., 1997) and sea surface temperature and precipitation (Knight et al., 2006), or between observations from different regions to establish remote links (e.g., X. Li et al., 2014). CPCA is less widely used mainly due to the "equal weight" assumption which will be discussed in the following text. In Li et al. (2014a, b), the two techniques were first applied to the intercomparison of different measurements of the same variable. Also see Li et al. (2014a, b) for the detailed description of the MCA and CPCA methods. In CMCA, a SVD is performed between the joint satellite data matrix and AERONET data matrix to extract the modes of variability that maximize the covariance between these two fields. In this way, the modes retain the orthogonality feature, and the leading modes will both have the highest correlation between the two data fields and explain the most variance of each individual field. Specifically, we arrange each satellite data field

negative values to avoid statistical bias (Levy et al., 2009). For central Australia, the DT and DB products are averaged during the austral winter (June, July, August) when NDVI is between 0.2 and 0.3. However, the DT AOD is consistently negative while the DB AOD is usually quite small during this season. As a result, the averaged values often become negative. Because negative AODs are not present in the other data

and AERONET data by space and time dimension as

$$\mathbf{X} = \begin{bmatrix} x_{1,1} & \cdots & x_{n,1} \\ \vdots & \ddots & \vdots \\ x_{m,1} & \cdots & x_{m,n} \end{bmatrix},$$ (1)

where m is the number of spatial locations (number of grid boxes for satellite data and number of stations for AERONET) and n is the number of measurements at each location (length of the data time series). For the satellite data sets m equals $360 \times 180 = 64\,800$ (the number of grid cells), and for AERONET m equals 58 (the number of stations). n equals 72, which is the number of months in the 6-year period. The data are centered by removing the temporal mean from each row of $\mathbf{X}$. In addition, we also create an anomaly data matrix by removing the multiyear-averaged seasonal cycle from each row, in order to examine the interannual variability.

After organizing the data sets in this manner, the data matrix of the satellites are combined into one large $4\,m \times n$ matrix as

$$\mathbf{X}_{\mathrm{sat}} = \begin{bmatrix} \mathbf{X}_{\mathrm{MODIS}} \\ \mathbf{X}_{\mathrm{MISR}} \\ \mathbf{X}_{\mathrm{SeaWiFS}} \\ \mathbf{X}_{\mathrm{OMI}} \end{bmatrix}.$$ (2)

It is important to note that combining the data matrices assumes equal weight, which requires that the fields being combined have the same measure. For the question here, as the four fields are the measurement of the same physical quantity (AOD) and mapped to the same spatial resolution ($1° \times 1°$), this prerequisite is satisfied. Nonetheless, due to sampling and retrieval uncertainties, the satellite fields are not measuring the "true AOD state" and differences will appear in the analysis results. These differences will be the focus of examination and comparison, which offer insights into possible issues of the data sets.

Next, we construct the cross-covariance matrix $\mathbf{C}$ between the joint satellite field X_{sat} and the AERONET data matrix $\mathbf{X}_{\mathrm{AERONET}}$ as

$$\mathbf{C} = \frac{1}{m-1} \mathbf{X}_{\mathrm{AERONET}} \mathbf{X}_{\mathrm{sat}}^{\mathrm{T}},$$ (3)

where $X_{\mathrm{AERONET}}^{\mathrm{T}}$ denotes the transpose of X_{AERONET}. The dimensions of $\mathbf{C}$, X_{AERONET} and X_{sat} are 72×72, 58×72 and $64\,800 \times 72$, respectively. The orthogonal modes that maximize the covariance between X_{sat} and X_{AERONET} are then found by a SVD of $\mathbf{C}$:

$$\mathbf{C} = \mathbf{U}\mathbf{S}\mathbf{V}^{\mathrm{T}},$$ (4)

where $\mathbf{U}$ and $\mathbf{V}$ are orthogonal matrices of dimensions 58×72 and $64\,800 \times 72$, respectively, and $\mathbf{S}$ is a 72×72 matrix containing the singular values. The columns of $\mathbf{U}$ and $\mathbf{V}$ are singular vectors for X_{AERONET} and X_{sat}, respectively, and

each pair of singular vectors represent covarying modes between the two data fields. In the SVD, the singular values in $\mathbf{S}$ which is the covariance between each pair of singular vectors are organized in descending order of magnitude, so that the first mode represents the most covariance between the two fields. As the covariance can be expressed as

$$\mathrm{cov}(\mathbf{X}, \mathbf{Y}) = r_{X,Y} \sqrt{S_X^2 S_Y^2},$$ (5)

where $\mathrm{cov}(\mathbf{X},\mathbf{Y})$ denotes the covariance between $\mathbf{X}$ and $\mathbf{Y}$, $r_{X,Y}$ denotes the correlation between $\mathbf{X}$ and $\mathbf{Y}$, and S_X^2 and S_Y^2 are the variances of $\mathbf{X}$ and $\mathbf{Y}$ respectively, maximizing the covariance implies the maximization of both the correlation and the variances. Therefore, the leading modes will represent the correlated variability in the two data sets and account for most of the variance.

The singular values in $\mathbf{U}$ and $\mathbf{V}$ are the spatial patterns of AERONET data and the combined satellite field, respectively. To find the spatial pattern of each individual satellite field, we divide $\mathbf{V}$ back into four segments as

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_1 \\ \mathbf{V}_2 \\ \mathbf{V}_3 \\ \mathbf{V}_4 \end{bmatrix}.$$ (6)

Each segment will have dimension $m \times n$, whose columns are the spatial patterns of each individual satellite data set. The time series $\mathbf{A}$ and $\mathbf{B}$ describing how each mode oscillates in time are then found by projecting $\mathbf{U}$ back to X_{AERONET} and projecting $\mathbf{V}$ back to $\mathbf{X}_{\mathrm{sat}}$:

$$\mathbf{A} = \mathbf{X}_{\mathrm{AERONET}} \mathbf{U},$$ (7)
$$\mathbf{B} = \mathbf{X}_{\mathrm{sat}} \mathbf{V}.$$ (8)

By letting σ_i denote the ith element of $\mathbf{S}$, the fraction of squared covariance (SCF) explained by the ith mode is then given as

$$\mathrm{SCF} = \frac{\sigma_i^2}{\sum_{j=1}^{N} \sigma_j^2}.$$ (9)

The major advantage of CMCA over MCA and CPCA is that CMCA effectively incorporates all available information. We will be able to examine the coherency as well as discrepancies across satellite data sets in parallel, and to further identify the strengths and weaknesses of each data set by evaluating their individual spatial modes against the AERONET results.

Meanwhile, some limitation and caveats of CMCA or spectral decomposition techniques in general should be kept in mind before applying them. First, caution must be taken when combining different data fields. As previously noted, the data fields being combined are assumed to have equal
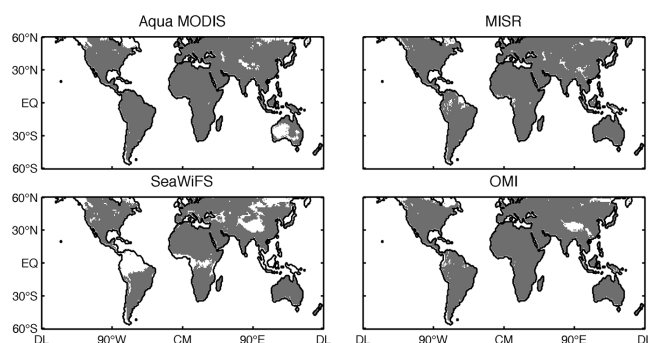
Figure 4. Data mask for the four satellite data sets. The white areas over land show the grid boxes with persistently missing data (more than half of the entire time series) that are removed from this analysis.

Figure 5. Variances explained by the first 20 modes of the CMCA analysis of global data from satellites and AERONET.

weight. Therefore, it is usually difficult to combine measurements of different quantities, as their degree of variability can be difficult to measure and compare. Second, the spatial representativeness of the ground-based measurements may affect the interpretation of the results. The disagreements between satellite and AERONET could also be attributed to the low spatial representativeness of the AERONET station rather than data accuracy. See Li et al. (2014a) for a discussion on this issue. Another possible issue is "mode leakage" which may also affect the physical interpretation. As these techniques are mathematical in nature, it is possible that some unrelated variability may appear in the same mode or that a single variability may be split into different modes. Li et al. (2014a) discussed an example for South America, where the signal is separated into both seasonal and semiseasonal modes due to the phase difference. This issue is not significant in comparison studies like the current one, as we are mainly concerned with the consistency in dominant variability. However, special attention must be paid to these artifacts when using the techniques to study physical problems. Finally, the major advantages of spectral techniques are dimensional reduction and efficiency. By simultaneously comparing the space–time variability, it is possible to quickly identify possible problems with (or differences between) the individual data sets, while it is not possible to find the cause of the problem or sources of uncertainty. Therefore, further examination of the regions identified by the spectral analysis is still needed to uncover the causes or mechanisms responsible for the disagreement.

4   Results

We demonstrate the usage of the CMCA technique by presenting the results of the global analysis for both the full and anomaly data, followed by two typical regional examples.
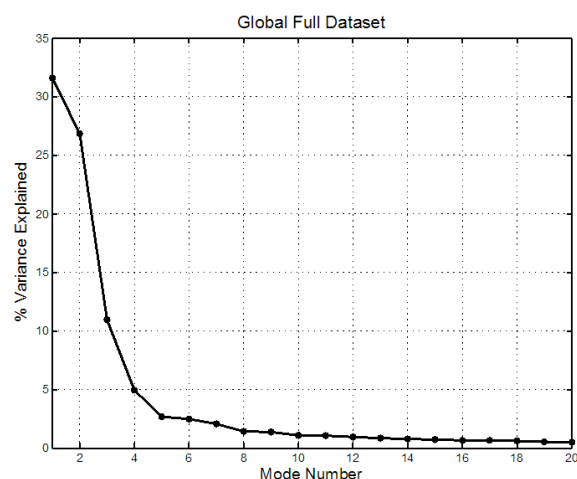
4.1   Global analysis – full data set

An analysis is first performed on the full data sets with the seasonal cycle left in. Figure 5 shows the variance explained by the first 20 modes. Each of the first three modes explains greater than 10 % of the variance, and these three modes together explain ∼ 70 % of the total variance. Moreover, there is a sharp drop in variance from mode 3 to mode 4. Based on these facts, we determine the first three modes to be dominant. The spatial patterns and PC time series are displayed in Fig. 6. The results are very similar to the those of PCA and MCA as presented by our previous studies (Li et al., 2014a, b), and thus we only briefly summarize them here. The first two modes represent the seasonal behavior of aerosol loading for the Northern Hemisphere and Southern Hemisphere, respectively. The two modes should be primarily controlled by meteorological conditions, in particular, the seasonal changes of temperature, wind and precipitation which affect the emission, transport and removal of aerosols. The third mode captures regions with semiannual variability including the Sahel, northern India and southeastern Asia. South America also shows a weak positive signal in this mode because the September peak in PC 3 overlaps the peak biomass-burning season over South America (see Li et al., 2014a). Special attention should be paid, however, to the agreements and disagreements between the signals at the AERONET stations and those of the underlying maps, as these provide information on the capability of each data set to represent the associated aerosol variability. For example, in mode 1, all four data sets and AERONET exhibit positive signals with similar strength over dust dominated regions of northwestern Africa and the Arabian Peninsula. This is an indication that dust variability over these regions is well represented by all of the satellite data sets. The same applies to the biomass-burning aerosol-source regions of South
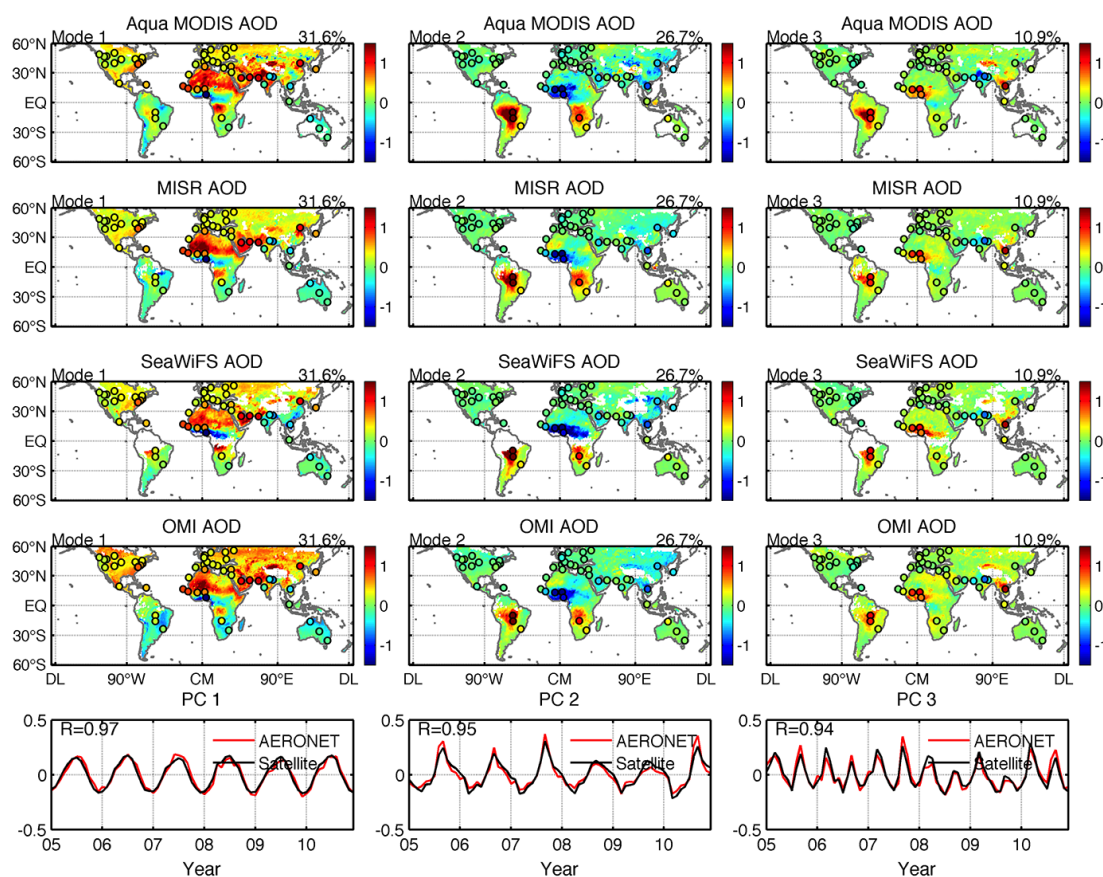
Figure 6. The first three modes of CMCA analysis of global satellite and AERONET data. The number in the upper right corner of each spatial map indicates the variance explained by this mode. The R value on the PC panels indicates the correlation coefficient between the time series for AERONET and that for the combined satellite field. The color of the dots indicates the strength of the AERONET signal and shares the same color scale as satellite data. Overall, the spatial patterns of all four data sets agree with AERONET for northern Africa, the Arabian Peninsula (mode 1), South America, southern Africa and the Sahel (mode 2).

America, the Sahel and southern Africa shown in mode 2. Mode 3 also reveals reasonable agreement in the semiannual variability of aerosol optical depth. Note the correlations between the PC time series are also quite high (above 0.9) for these three modes. However, notable differences can also be identified across the data sets. An obvious example is the Indian subcontinent. In mode 1, MODIS, MISR and OMI all have positive signals over this region, while SeaWiFS has weak negative signals. Turning to AERONET, we find that the three stations over this region also have negative signals, consistent with SeaWiFS but different from MODIS, MISR and OMI. It is thus highly possible that SeaWiFS captures well the seasonality of aerosol variability over the Indian subcontinent while the other three data sets may have lower skills over this region. As a result, this region will be examined in greater detail in the next section. In fact, with spatial modes from multiple satellites, regions with the highest uncertainty can be highlighted by examining the spread (standard deviation) of the four spatial patterns. Figure 7 shows the standard deviation fields of the four spatial maps for each

mode. Regions with largest spread are marked by red rectangles. For mode 1, in addition to India, eastern Asia also appears to have larger disagreement. This region has been an emerging global-aerosol-source region over the past decade. The aerosol composition is complicated with heavy pollution from industrialized urban areas, especially in eastern China, and also seasonal dust pollution from central northern Asia, which contributes to the difficulty in satellite retrievals. However, as most AERONET stations in eastern China were established in recent years, we found almost no qualified stations for the purpose of this study. The large disagreement across the satellite measurements over this region therefore suggests the necessity for continuous monitoring of aerosol properties of the surface in this region. For mode 2, South America, the Sahel, central Asia and Borneo appear to have the largest discrepancy. Looking back to Fig. 6, it can be seen that for South America and the Sahel, MODIS and SeaWiFS both have strong positive and negative signals, respectively, and in good agreement with AERONET, while the signals for MISR and OMI are generally weaker,
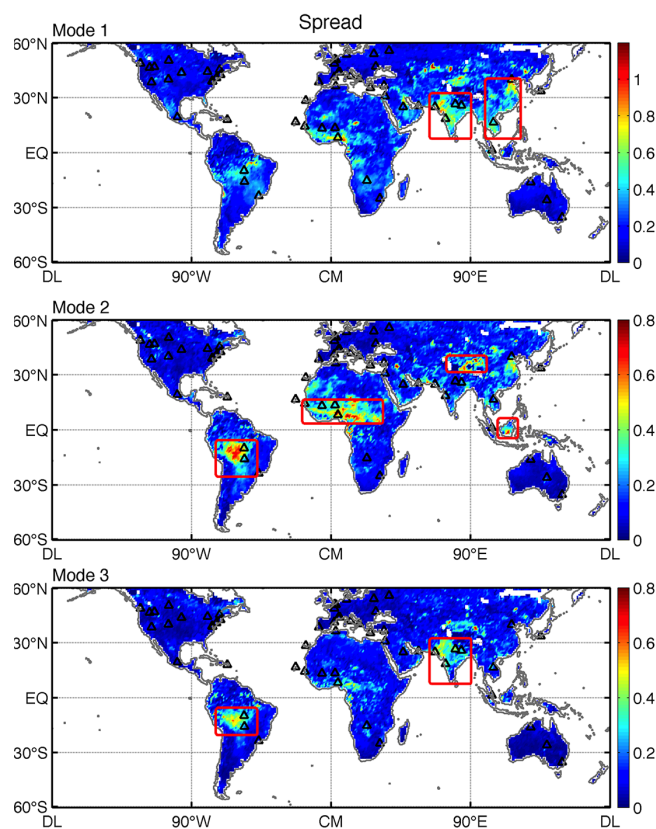
Figure 7. Satellite data standard deviation (spread) maps for the three modes shown in Fig. 6. Regions with largest spread, and thus highest uncertainty, are indicated by the red boxes.

Figure 8. Variances explained by the first 20 modes of the CMCA analysis of global satellite and AERONET anomaly data.

especially for MISR over the Sahel and OMI over South America. Li et al. (2014b) have discussed the problems in these two satellite data sets for these two regions and found underestimation in MISR and OMI during the peak biomass-burning season over South America, while MODIS overestimates the biomass-burning peak (see also Levy et al., 2010). The weaker spring/fall seasonality for MISR over the Sahel is due to its underestimation of AOD during the (boreal) fall and overestimation of AOD during the (boreal) spring (Li et al., 2014b). The CMCA successfully confirms these conclusions with the help of AERONET. Li et al. (2014b) also investigated the problem for central Asia around the Takla Makan Desert and indicated that the low sampling frequency of MISR may miss dust emission events and thus lead to an underestimation of the variability. Unfortunately, there is no AERONET station in this area to confirm this hypothesis. The disagreement over Borneo in mode 2 comes from the positive signals seen on MODIS and MISR maps, but no signal in OMI. SeaWiFS has consistently missing data over this region due to its difficulty with cloud screening. With the lack of IR (infrared) channels, cloud detection over land depends on spatial variability, while some aerosol plumes may appear similar to clouds. Satellite retrievals tend to have
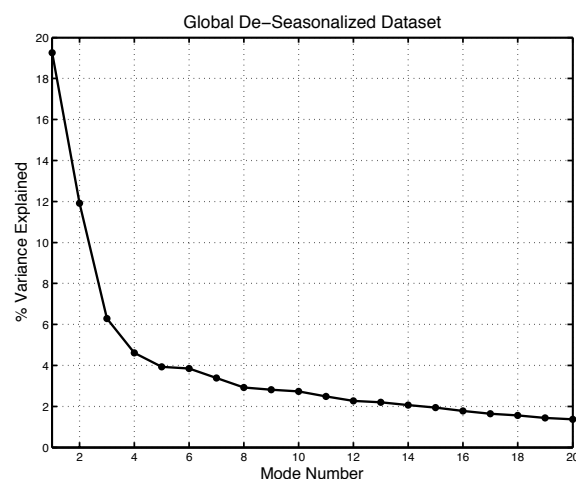
conservative cloud screening in order to avoid misclassification of clouds as aerosols. Therefore, in regions such as Borneo where small clouds and aerosol plumes often exist, it is likely that some smoke plumes are screened out as clouds (A. Sayer, personal communication, August 2012). Again no qualified AERONET station is available here. As this region is a major biomass-burning source region (Generoso et al., 2003; van der Werf et al., 2006), it should also be the focus of future AERONET instrumentation deployment. A few AERONET stations have already been established and have begun acquiring data in recent years, and a more complete study of the seasonal and interannual variability of this region will become possible in the future. The differences in mode 3 are similar to those in modes 1 and 2 and we therefore omit the discussion here. It should be kept in mind that the disagreements observed in Fig. 7 could be attributed to both data accuracy and observability of each sensor as well as the representativeness of AERONET station data. Low (high) bias in the data at high AOD conditions will lead to weaker (stronger) seasonality, such as the high bias in MODIS seasonality over South America. The observability defined here refers to the capability of the sensor to fully represent aerosol properties both spatially and temporally. It reflects the combined effect of satellite sampling associated with swath width and aerosol obscuration by clouds at the time of satellite overpass. In the analysis here using Level 3 data, the observability issues could be more dominant and account for most discrepancies, such as those for the Takla Makan Desert and Borneo. Moreover, while AERONET retrievals are highly accurate, they may also have this "observability" issue due to the sampling frequency and the presence of clouds. The spatial representativeness of each AERONET station may also affect the interpretation of the results, which was discussed by Li et al. (2014a).
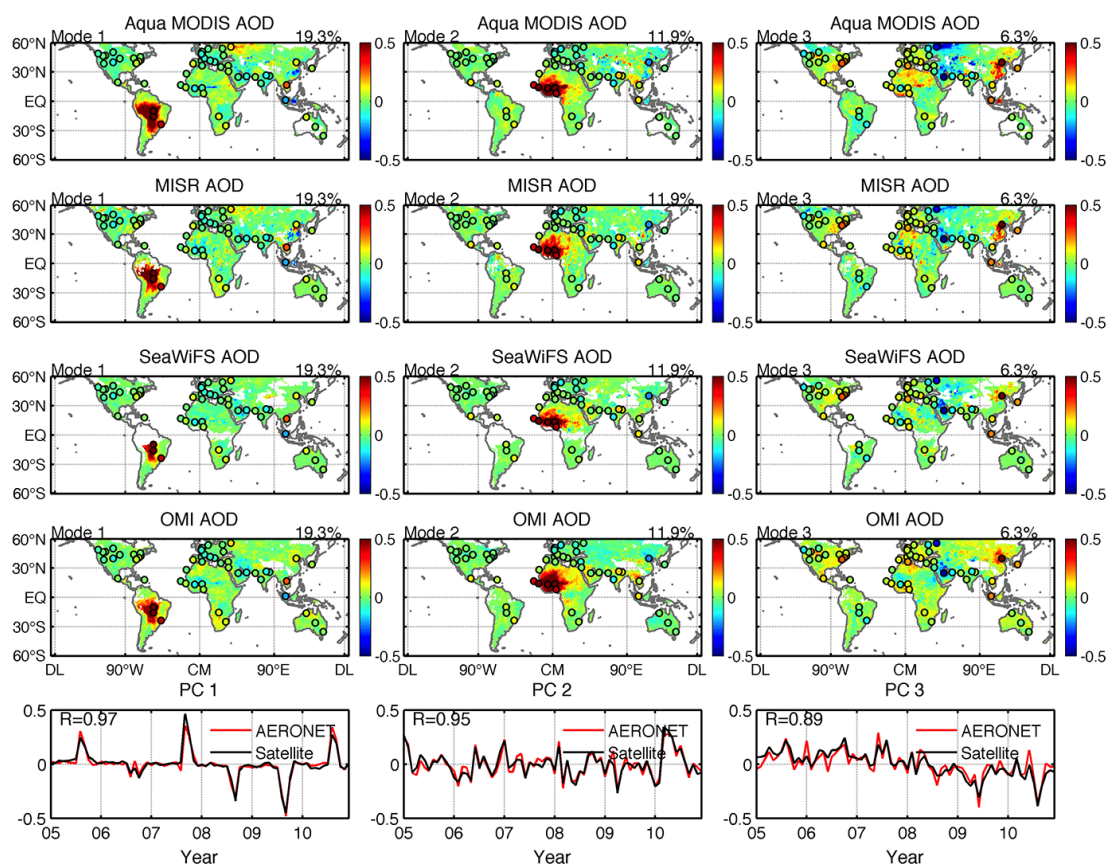
Figure 9. The first three modes of the CMCA analysis of anomaly data, representing interannual variability for South America, northwestern Africa and eastern Asia, respectively. The four satellite data sets also agree well with AERONET.

4.2 Global analysis – anomaly data set

With respect to the results of the analysis of the anomaly data set, we again chose to present the first three modes based on the behavior of variance explained with the curve shown in Fig. 8. These three modes, as shown in Fig. 9, are also consistent with Li et al. (2013, 2014a, b) and reveal aerosol-source regions and their interannual variability. It is encouraging that all four satellite data sets agree well with AERONET qualitatively. Quantitative examination of the standard deviation maps (Fig. 10) reveals discrepancies in the signal strength over South America and the Sahel, which are similar to those in Fig. 7 and previously discussed by Li et al. (2013). In mode 3, eastern Europe is highlighted with larger uncertainty. This is related to an extreme event and will be further investigated in the next section. Eastern and southeastern Asia also appear in the spread map of mode 3 which again suggests that additional observations are needed in these areas. While the global results mainly confirm our previous findings, the advantage of using CMCA is clearly seen: comparing multiple satellite data sets in parallel and simultaneously validating the variability associated with specific aerosol types and/or source regions against AERONET in

one spatial map. MCA only allows making comparisons between one data set and AERONET, while cross-comparisons between satellite data sets are not possible, and CPCA can only be used to compare different satellite data sets but it is not able to evaluate the strengths and weaknesses of the individual data sets because it cannot accommodate the ground-truth AERONET data. CMCA successfully solves these two problems and makes full use of all of the available information. It is also an efficient way to provide insights into possible problems and highlight the regions with the largest uncertainty in AOD.

4.3 Regional analysis

In this section, we present the results of two regional case studies. These studies focus on the added information content of the temporal variability, and demonstrate the advantage of the CMCA technique in identifying problems associated with extreme events, interannual variability and seasonal variability. In the global analysis of the anomaly data (focusing on interannual variability), we identified a "hot spot" in eastern Europe, i.e., a region of large disagreement between the four data sets (mode 3 in Fig. 9). Here we further examine this

disagreement using CMCA by isolating this region. CMCA is performed over Europe within the spatial domain of 6° W–56° E, 40–60° N. The first mode of the anomaly data, shown in Fig. 11, clearly highlights the eastern European region. This mode accounts for 42.3 % of the variance. On the spatial maps, both MODIS and MISR exhibit strong positive signals while SeaWiFS and OMI only have weak signal. The two AERONET stations located in this region also have positive signals in accordance with MODIS and MISR but disagree with SeaWiFS and OMI. The PC time series of this mode exhibits a high peak in August 2010. Therefore, this mode is most likely associated with the documented intense Russian wildfire in the summer of 2010 (Witte et al., 2011; Konovalov et al., 2011; Chubarova et al., 2012). And the patterns of the spatial maps of the four satellites indicate that MODIS and MISR capture this event while it is less well-represented in the SeaWiFS and OMI data sets. To confirm this conclusion, we compare the time series between the AERONET data and the satellite data at the Moscow MSU MO station, located at the center of the positive anomaly with the strongest signal. The results are presented in Fig. 12 and it is clearly seen that AERONET data at this station are mostly temporally flat except for an extremely strong peak in 2010. MODIS and MISR agree well with AERONET with peaks of similar strength. SeaWiFS also has a peak in 2010, but much weaker compared to AERONET, while the OMI data do not show any outstanding peaks in this year. Various reasons may account for the problems in SeaWiFS and OMI. For example, overconservative cloud screening may mistake smoke pixels for clouds, and the row anomaly developed in the OMI instrument since 2008 (http://www.knmi.nl/omi/research/product/rowanomaly-background.php) may lead to OMI missing this event due to reduced sampling. The results from our CMCA analysis suggest that the retrieval of AOD by SeaWiFS and OMI may need to be improved for this region to sufficiently represent this type of extreme event. Our next example focuses on the analysis of annual variability over the Indian subcontinent, which is another major source of discrepancy revealed through the global analysis (see Fig. 7). A major difficulty encountered for India is that few AERONET stations over this area have qualified data records for the construction of the temporal covariance matrix. Therefore, we only have four stations available for this analysis.

Nonetheless, the distribution of these stations does cover the typical aerosol-source regions of the Gangetic Plain, Thar Desert and southern India. Figure 13 shows the first two modes over India, which account for ∼ 98 % of the variance. The first mode mainly represents the variability of dust aerosols around the Thar Desert. The PC has a regular summer/winter (boreal) seasonal cycle. The second mode highlights the Gangetic Plain in northern India, and its PC time series displays a semiannual variability with two peaks in the late (boreal) spring–summer and the fall seasons, respectively. The Gangetic Plain has highly variable
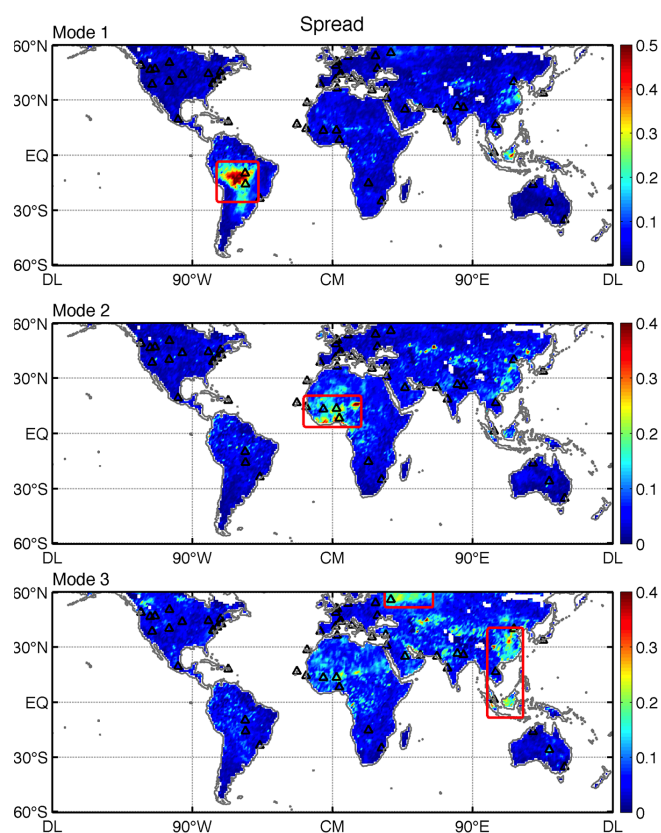

Figure 10. Satellite data standard deviation (spread) maps of the three CMCA modes shown in Fig. 9. Regions with the largest spread, and thus highest uncertainty, are indicated by red boxes.

aerosol types in different seasons. During the pre-monsoon (March–May) and monsoon seasons (June–August), this region is primarily influenced by dust aerosols, while during the post-monsoon (September–November) and winter seasons (December–January), anthropogenic aerosols compose a larger fraction of the total aerosol loading (Singh et al., 2004; Dey and Di Girolamo, 2010). The four data sets all agree with AERONET over the Thar Desert in mode 1. However, with respect to the Gangetic Plain, more differences appear. In mode 1, only SeaWiFS shows a coherent signal with AERONET over these regions with negative signals around the two AERONET sites in the Gangetic Plain (Kanpur and Gandhi College). The other three satellite data sets, especially MODIS, have positive signals and are opposite to those of AERONET. The seasonality of the Gangetic Plain is further isolated in mode 2. In this mode, SeaWiFS and MODIS capture the semiannual variability relatively well, while the signals for MISR and OMI are comparatively weaker than those observed by AERONET, SeaWiFS and MODIS. This result implies that disagreements in the seasonality of AOD for the Gangetic Plain may exist in the MODIS, MISR and OMI data sets and should therefore be examined further. We also examine the interannual variability of the Gangetic
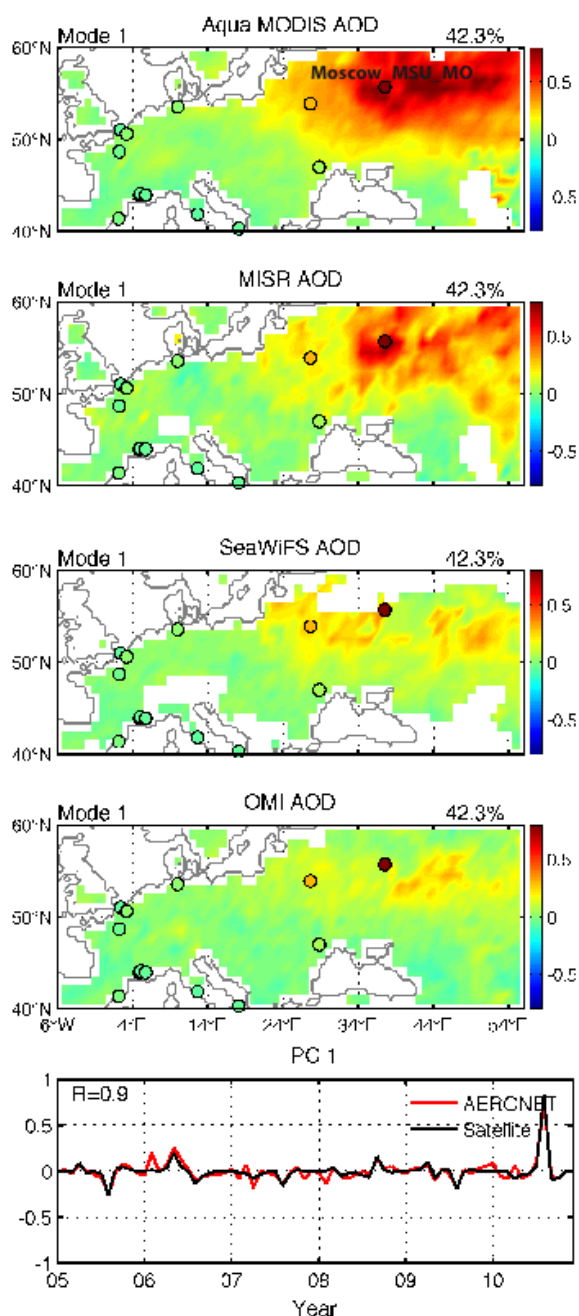
Figure 11. The first CMCA mode over Europe showing the intense wildfire in Russia. Both MODIS and MISR exhibit strong positive anomalies and agree with AERONET. The distribution of positive signals in SeaWiFS and OMI are less extensive.
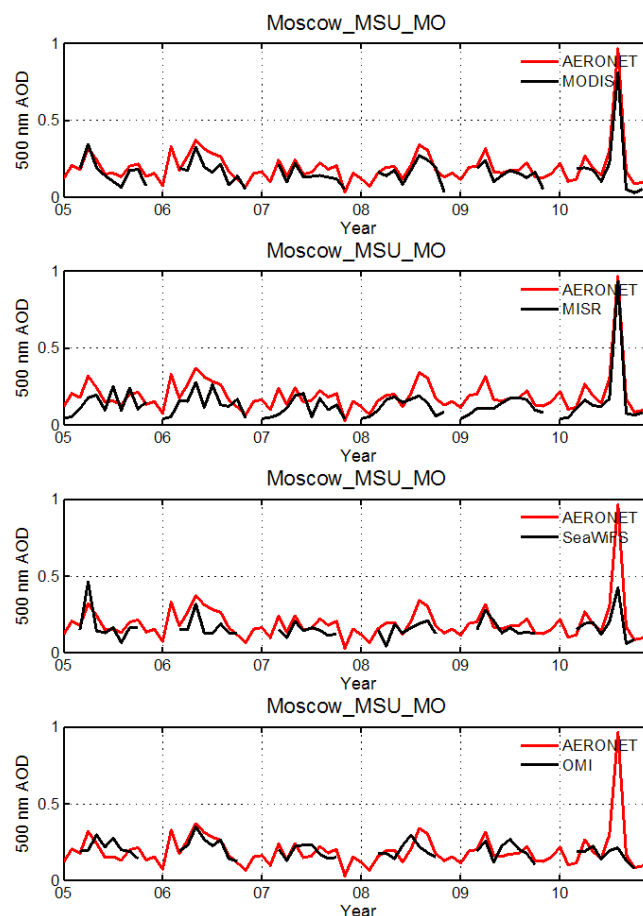

Figure 12. Comparison between the AOD time series for the satellite and AERONET data at the Moscow MSU MO station (the location is indicated in the top panel of Fig. 11).

Plain region using the anomaly data. This region appears in the dominant mode, which is shown in Fig. 14. Interestingly, while the SeaWiFS data sets best represent the seasonal variability of AOD over the Gangetic Plain, Fig. 14 indicates that on an interannual timescale, this data set differs the most from AERONET compared to the other three data sets. The positive anomalies on the SeaWiFS spatial map are

both narrower and weaker. To explain this paradox in the Sea-WiFS data, as well as the problems in the MODIS, MISR and OMI data sets, we compare the time series from the AERONET measurements and satellite data for two stations: Kanpur and Gandhi College. The raw time series, multiyear-averaged seasonal cycle, and the anomaly time series for each of the satellite data plotted against AERONET at Kanpur and Gandhi College are shown in Figs. 15 and 16, respectively. The correlation coefficient between the satellite and AERONET time series on each panel is indicated in the upper-left corner in black. For the Kanpur station, we can see that, overall, SeaWiFS data have the highest correlation with AERONET for the raw time series and seasonal cycle. Especially for the latter, the correlation is above 0.9. Compared with the AERONET time series, MISR and OMI both have an overall low bias, which is larger during the winter months. For MODIS, however, there is an overall high bias during the summer months but an underestimation during the winter. These differences lead to a stronger summer peak and weaker winter peak in MODIS, MISR and OMI
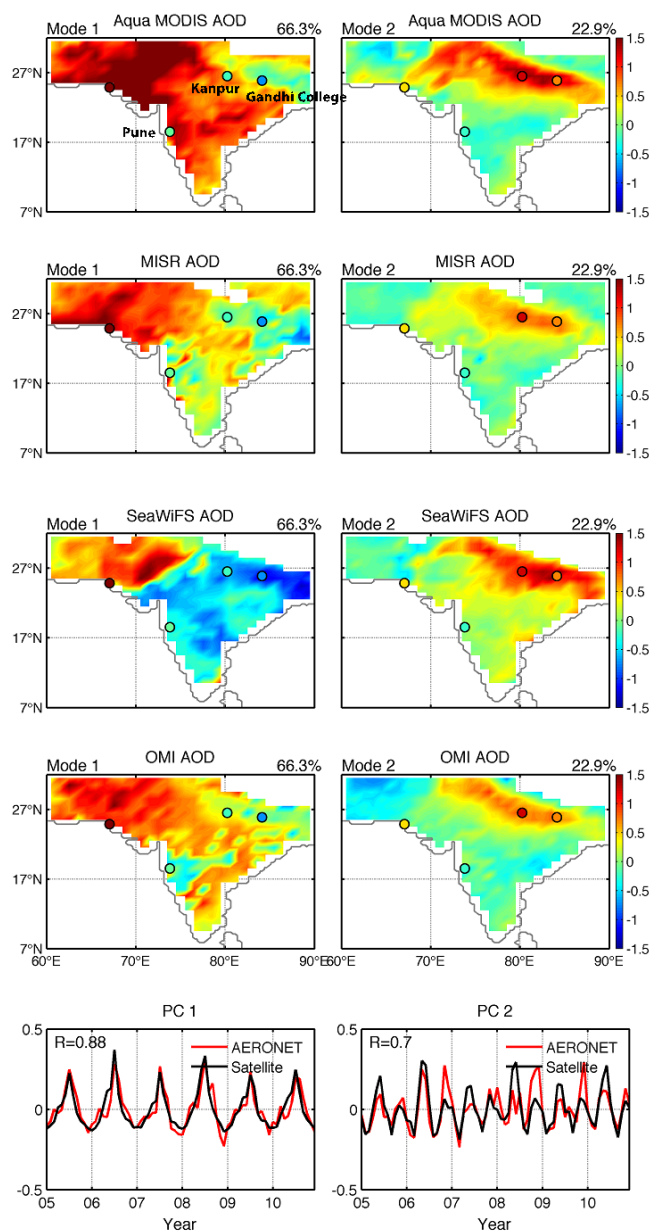
Figure 13. The first two CMCA modes over India, representing aerosol seasonal variability for the Thar Desert and Gangetic Basin, respectively. In mode 1, the magnitude of the signal of Kanpur is −0.18, and for Gandhi College it is −0.70. In mode 2, the signal for Kanpur is 1.26 and for Gandhi College it is 0.73. Compared with MODIS, MISR and OMI, the SeaWiFS spatial maps agree well with AERONET for both modes 1 and 2 and best capture the seasonality over the Gangetic Plain region.

data, which are responsible for the positive projection of the winter–summer seasonality (PC 1 of Fig. 13) on these three data sets. However, for AERONET and SeaWiFS, the intensity of the winter peak is comparable to or even stronger than the summer peak. As a result, the variability of these two data sets is captured by PC 2, which has an associated semiannual
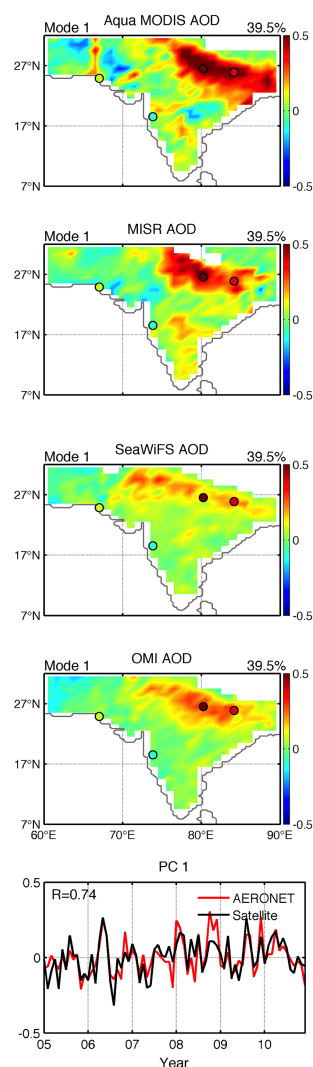

Figure 14. The first mode of the anomaly data over the Indian subcontinent. Unlike Fig. 13, on interannual timescales, MODIS and MISR best represent the AOD variability over the Gangetic Plain, while the SeaWiFS and OMI patterns have less coherency with AERONET.

timescale. The comparison between the interannual variability using AOD anomalies (right column in Fig. 15), however, displays a completely different picture. Unlike the raw time series and seasonal cycle, SeaWiFS now has the lowest correlation with AERONET on an interannual timescale. Not only does it fail to capture several strong anomalies in 2005, 2008 and 2009, but also the variance of the time series is considerably lower than that of AERONET. The variance for the SeaWiFS anomaly time series is 0.0041, while that for AERONET is 0.0136, and those for MODIS, MISR and OMI are 0.0113, 0.0083 and 0.0051, respectively. Accordingly, the weaker signal in the SeaWiFS spatial mode in Fig. 14 is attributed to both this low correlation and low variance. The results for Gandhi College convey the same
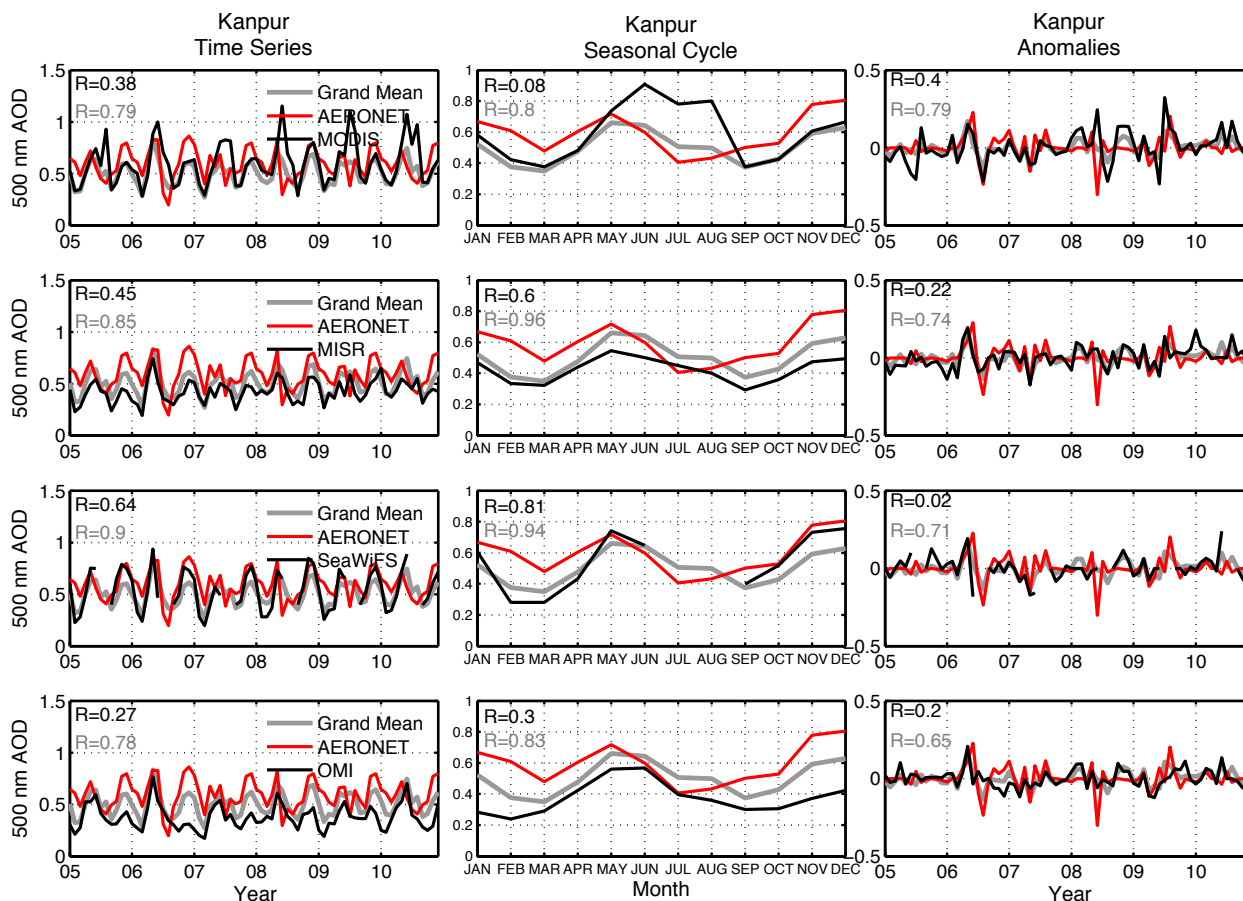
Figure 15. Comparison between the raw AOD time series (left column), multiyear-averaged seasonal cycle (middle column) and anomaly time series (right column) for Kanpur (location marked in the top-left panel of Fig. 13). The gray curves show the grand mean of all five data sets. The R values in each panel indicate the correlation coefficient between satellite and AERONET (black) and between satellite and grand mean (gray). SeaWiFS data have the highest correlation with AERONET and the grand mean in terms of seasonality; however, its agreement with AERONET in terms of interannual variability is not as good as the other three data sets.

information, indicating that SeaWiFS has the highest correlation with AERONET for season variability but the lowest for interannual variability. Moreover, because each data set has certain limitations, such as missing data for SeaWiFS in July and August, low sampling frequency for MISR, row anomalies for OMI, and AERONET stations possibly not fully representing the entire satellite grid box, we also compared each data set to the grand mean (shown in gray curves) which may be a better representation of the truth. The correlation between the time series of each satellite and the grand mean is indicated in the upper-left corner of each panel in gray numbers. As expected, the agreement between the satellite data and the AERONET grand mean is better. The results are consistent with the AERONET comparison i.e., SeaWiFS best represents seasonal variability but has the poorest agreement for interannual variability.

From the global analysis and regional studies, we can clearly see that the CMCA technique is both an efficient and effective method for the analysis and comparison of multisensor data. On a first order, spectral decomposition reduces data dimensionality and limits the comparison to only the first few leading modes that explain the bulk of the variance in the data. Moreover, by integrating all available information, many variations, source regions and events can be further confirmed. Most importantly, the analysis helps us to identify the strengths and weaknesses of each data set in representing aerosol variability for specific regions and on different timescales, which is essential for understanding the capability of the data and making the best use of it.

5   Summary

In this paper, we introduce an improved spectral decomposition technique for use in multiple data comparison and evaluation, based on the principal component analysis and maximum covariance analysis. By extracting the modes of variability that maximize the covariance between the combined satellite field and ground-based AERONET observations, the
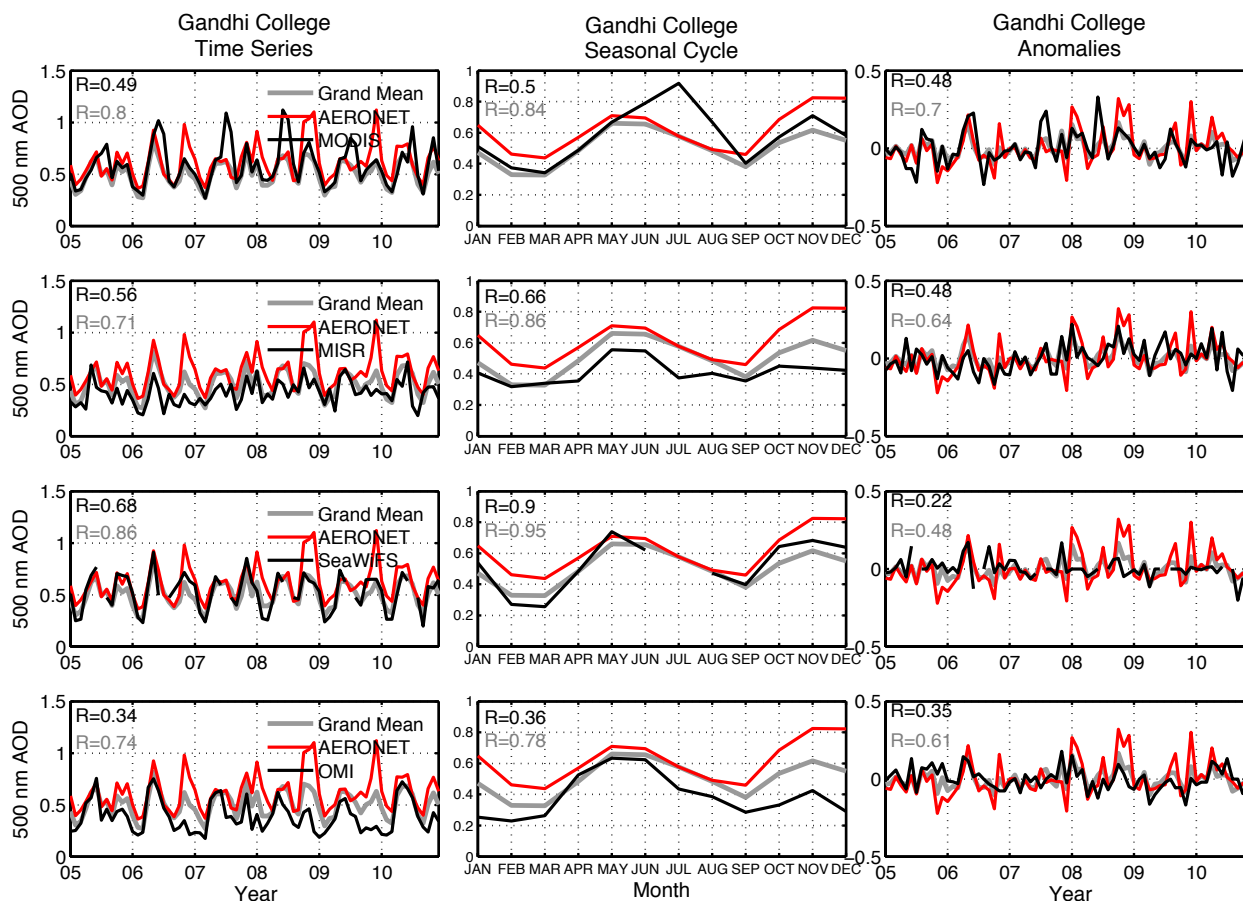
Figure 16. As Fig. 15 but for Gandhi College, the other station in the Gangetic Plain. The results are basically consistent with those for Kanpur.

CMCA has the advantage of evaluating each individual data set using AERONET simultaneously. In addition, most results are clearly associated with specific aerosol-source regions, events or temporal scales represented by each orthogonal mode, which associates the comparison with real physical phenomena.

Examples of global and two representative regional analyses are presented and discussed to show the usage of the CMCA method. Globally, the results indicate that all four data sets reasonably agree with AERONET for major aerosol-source regions, including dust over northern Africa and the Arabian Peninsula, biomass-burning over South America and southern Africa and mixed aerosol types over the Sahel. The interannual variability of the source regions also agrees well. These results suggest that these patterns are the most believable and that we should be confident in using all or any of the four satellite data sets in the study of aerosol properties and their temporal variability over these regions.

The purpose of the regional case studies is to illustrate the ability of the CMCA method to identify potential problems in certain the data sets. The strengths and weaknesses of each data set are identified through direct comparison between the

positive/negative signals in the spatial patterns of the satellite and AERONET data maps. The nature of the problem can then be further examined by comparing the raw time series. Moreover, the capability of each data set in capturing the variability on seasonal and interannual timescales can be separately assessed. The results from our regional analysis indicate that SeaWiFS and OMI did not fully represent the intense Russian wildfire in August 2010. Their signals are weaker compared to MODIS and MISR for this event. The AOD seasonality over the Indian Gangetic Plain and southern India needs to be improved for MODIS, MISR and OMI. SeaWiFS has the best agreement with AERONET on the seasonal variability over this region; however, on interannual timescales its agreement is poorer than that for MODIS, MISR and OMI.

Because the main purpose of this paper is to present the CMCA technique, we did not analyze all interesting regions. However, readers are encouraged to use this technique for comprehensive analyses covering more regions and events, or in studying specific regions of their interest. Although this technique has been applied between satellite and AERONET data, there is no doubt that it can be adapted for model–data

comparison and validation as well as for use with other ground-based network measurements (e.g., MPLnet). Model validation is an important potential application of the CMCA method. On the one hand, with multiple observational data sets available, it is desirable to incorporate all pieces of information to yield a more robust validation. One the other hand, as chemical transport models are usually constrained using satellite observations, large uncertainty in observations will also result in poorly constrained model fields. Therefore, places where retrieval skills are low often correspond to those where model fields are inaccurate. For example, Trivitayanurak et al. (2012) found poor agreement between the GEOS (Goddard Earth Observing System)-Chem simulated AOD and MODIS AOD for the southeastern Asia region due to the uncertainties in satellite retrieval. The CMCA technique will identify these regions, and thus provide insights into the problems in either the satellite, the model or both. When using model data, the model-data field should be treated in same manner as AERONET is used here, i.e., when the model resolution is coarser than satellite data. However, for models with a comparable spatial resolution to the satellite data, the model field can be treated as one of the satellite fields and directly compared with the satellite data sets and AERONET. Traditional model validation usually compares averaged time series between model and data for the globe and several representative regions, while the CMCA offers a new approach with a simultaneous spatial and temporal view. It also provides an effective and efficient way to identify problems that are not easily detected by traditional methods. With the continuous development of remote sensing data sets as well as climate models, we believe this technique will become a useful tool for data retrieval, data analysis and the modeling community.

References

Ahn, C., Torres, O., and Bhartia, P. K.: Comparison of ozone monitoring instrument UV aerosol products with Aqua/Moderate Resolution Imaging Spectroradiometer and Multiangle Imaging Spectroradiometer observations in 2006, J. Geophys. Res., 113, D16S27, doi:10.1029/2007JD008832, 2008.

Ahn, C., Torres, O., and Jethva, H.: Assessment of OMI near-UV aerosol optical depth over land, J. Geophys. Res. Atmos., 119, 2457–2473, doi:10.1002/2013JD020188, 2014.

Björnsson, H. and Venegas, S. A.: A manual for EOF and SVD analysis of climate data, CCGCR Rep. 97-1, McGill University, Montréal, QC, Canada, 52 pp., 1997.

Bretherton, C. S., Smith, C., and Wallace, J. M.: An intercomparison of methods for finding coupled patterns in climate data, J. Climate, 5, 541–560, 1992.

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

Chang, P., Ji, L., and Li, H.: A decadal climate variation in the tropical Atlantic Ocean from thermodynamic air-sea interactions, Nature, 385, 516–518, 1997.

Chubarova, N., Nezval', Ye., Sviridenkov, I., Smirnov, A., and Slutsker, I.: Smoke aerosol and its radiative effects during extreme fire event over Central Russia in summer 2010, Atmos. Meas. Tech., 5, 557–568, doi:10.5194/amt-5-557-2012, 2012.

Dey, S. and Di Girolamo, L.: A climatology of aerosol optical and microphysical properties over the Indian subcontinent from 9 years (2000–2008) of Multiangle Imaging Spectroradiometer (MISR) data, J. Geophys. Res., 115, D15204, doi:10.1029/2009JD013395, 2010.

Diner, D. J., Beckert, J. C., Reilly, T. H., Bruegge, C. J., Conel, J. E., Kahn, R., Martonchik, J. V., Ackerman, T. P., Davies, R., Gerstl, S. A. W., Gordon, H. R., Muller, J.-P., Myneni, R., Sellers, R. J., Pinty, B., and Verstraete, M. M.: Multiangle Imaging Spectro-Radiometer (MISR) escription and experiment overview, IEEE Trans. Geosci. Remt. Sens., 36, 1072–1087, 1998.

Dubovik, O., Holben, B. N., Eck, T. F., Smirnov, A., Kaufman, Y. J., King, M. D., Tanre, D., and Slutsker, I.: Variability of absorption and optical properties of key aerosol types observed in worldwide locations, J. Atmos. Sci., 59, 590–608, 2002.

Eck, T. F., Holben, B. N., Reid, J. S., Dubovik, O., Smirnov, A., O'Neill, N. T., Slutsker, I., and Kinne, S.: Wavelength dependence of the optical depth of biomass burning, urban and desert dust aerosols, J. Geophys. Res., 104, 31333–31350, 1999.

García, O. E., Díaz, J. P., Expósito, F. J., Díaz, A. M., Dubovik, O., Derimian, Y., Dubuisson, P., and Roger, J.-C.: Shortwave radiative forcing and efficiency of key aerosol types using AERONET data, Atmos. Chem. Phys., 12, 5129–5145, doi:10.5194/acp-12-5129-2012, 2012.

Generoso, S., Bréon, F.-M., Balkanski, Y., Boucher, O., and Schulz, M.: Improving the seasonal cycle and interannual variations of biomass burning aerosol sources, Atmos. Chem. Phys., 3, 1211–1222, doi:10.5194/acp-3-1211-2003, 2003.

Holben, B. N., Eck, T. F., Slutsker, I., Tanre, D., Buis, J. P., Setzer, A., Vermote, E., Reagan, J. A., Kaufman, Y., Nakajima, T., Lavenu, F., Jankowiak, I., and Smirnov, A.: AERONET – A federated instrument network and data archive for aerosol characterization, Remote Sens. Environ., 66, 1–16, 1998.

Holben, B. N., Tanre, D., Smirnov, A., Eck, T. F., Slutsker, I., Abuhassan, N., Newcomb, W. W., Schafer, J., Chatenet, B., Lavenue, F., Kaufman, Y. J., Vande Castle, J., Setzer, A., Markham, B., Clark, D., Frouin, R., Halthore, R., Karnieli, A., O'Neill, N. T., Pietras, C., Pinker, R. T., Voss, K., and Zibordi, G.: An emerging ground-based aerosol climatology: Aerosol Optical Depth from AERONET, J. Geophys. Res., 106, 12067–12097, 2001.

Holben, B. N., Eck, T. F., Slutsker, I., Smirnov, A., Sinyuk, A., Schafer, J. S., Giles, D. M., and Dubovik, O.: AERONET's version 2.0 quality assurance criteria, Proc. SPIE, 6408, doi:10.1117/12.706524, 2006.

Hsu, N. C., Tsay, S. C., King, M. D., and Herman, J. R.: Aerosol properties over bright reflecting source regions, IEEE Trans. Geosci. Remote Sens., 42, 557–569, 2004.

Hsu, N. C., Tsay, S. C., King, M. D., and Herman, J. R.: Deep blue retrievals of Asian aerosol properties during ACE-Asia, IEEE Trans. Geosci. Remote Sens., 44, 3180–3195, 2006.

Hsu, N. C., Gautam, R., Sayer, A. M., Bettenhausen, C., Li, C., Jeong, M. J., Tsay, S.-C., and Holben, B. N.: Global and regional trends of aerosol optical depth over land and ocean using SeaWiFS measurements from 1997 to 2010, Atmos. Chem. Phys., 12, 8037–8053, doi:10.5194/acp-12-8037-2012, 2012.

Hubanks, P., King, M., Platnick, S., and Pincus, R.: MODIS atmosphere L3 gridded product algorithm theoretical basis document Collection 005 Version 1.1, Tech. Rep. ATBD-MOD-30, NASA, 2008.

Kahn, R. A., Gaitley, B. J., Martonchik, J. V., Diner, D. J., Crean, K. A., and Holben, B.: Multiangle Imaging Spectroradiometer (MISR) global aerosol optical depth validation based on 2 years of coincident Aerosol Robotic Network (AERONET) observations, J. Geophys. Res., 110, D10S04, doi:10.1029/2004JD004706, 2005.

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

Kalashnikova, O. V. and Kahn, R. A.: Mineral dust plume evolution over the Atlantic from MISR and MODIS aerosol retrievals, J. Geophys. Res., 113, D24204, doi:10.1029/2008JD010083, 2008.

King, M. D., Kaufman, Y. J., Tanré, D., and Nakajima, T.: Remote sensing of tropospheric aerosols from space: Past, present, and future, 1999.

Kinne, S., Lohmann, U., Feichter, J., Schulz, M., Timmreck, C., Ghan, S., and Kaufman, Y.: Monthly averages of aerosol properties: A global comparison among models, satellite data, and AERONET ground data, J. Geophys. Res., 108, 4634, doi:10.1029/2001JD001253, 2003.

Kinne, S., Schulz, M., Textor, C., Guibert, S., Balkanski, Y., Bauer, S. E., Berntsen, T., Berglen, T. F., Boucher, O., Chin, M., Collins, W., Dentener, F., Diehl, T., Easter, R., Feichter, J., Fillmore, D., Ghan, S., Ginoux, P., Gong, S., Grini, A., Hendricks, J., Herzog, M., Horowitz, L., Isaksen, I., Iversen, T., Kirkevåg, A., Kloster, S., Koch, D., Kristjansson, J. E., Krol, M., Lauer, A., Lamarque, J. F., Lesins, G., Liu, X., Lohmann, U., Montanaro, V., Myhre, G., Penner, J., Pitari, G., Reddy, S., Seland, O., Stier, P., Takemura, T., and Tie, X.: An AeroCom initial assessment –optical properties in aerosol component modules of global models, Atmos. Chem. Phys., 6, 1815–1834, doi:10.5194/acp-6-1815-2006, 2006.

Knight, J. R., Folland, C. K., and Scaife, A. A.: Climate impacts of the Atlantic Multidecadal Oscillation, Geophys. Res. Lett., 33, L17706, doi:10.1029/2006GL026242, 2006.

Konovalov, I. B., Beekmann, M., Kuznetsova, I. N., Yurova, A., and Zvyagintsev, A. M.: Atmospheric impacts of the 2010 Russian wildfires: integrating modelling and measurements of an extreme air pollution episode in the Moscow region, Atmos. Chem. Phys., 11, 10031–10056, doi:10.5194/acp-11-10031-2011, 2011.

Levelt, P. F., van den Oord, G. H. J., Dobber, M. R., Malkki, A., Huib Visser, de Vries, J., Stammes, P., Lundell, J. O. V., and Saari, H.: The ozone monitoring instrument, Geosci. Remote Sens., 44, 1093–1101, 2006.

Levy, R. C., Remer, L. A., Mattoo, S., Vermote, E. F., and Kaufman, Y. J.: 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, 2007.

Levy, R. C., Remer, L. A., Tanré, D., Mattoo, S., and Kaufman, Y. J.: Algorithm for remote sensing of tropospheric aerosol over dark targets form MODIS: Collections 005 and 051, revision 2, available at: http://modis-atmos.gsfc.nasa.gov/_docs/ATBD_MOD04_C005_rev2.pdf (last access: July 2013), 2009.

Levy, R. C., Remer, L. A., Kleidman, R. G., Mattoo, S., Ichoku, C., Kahn, R., and Eck, T. F.: Global evaluation of the Collection 5 MODIS dark-target aerosol products over land, Atmos. Chem. Phys., 10, 10399–10420, doi:10.5194/acp-10-10399-2010, 2010.

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

Li, J., Carlson, B. E., and Lacis, A. A.: A study on the temporal and spatial variability of absorbing aerosols using Total Ozone Mapping Spectrometer and Ozone Monitoring Instrument Aerosol Index data, J. Geophys. Res., 114, D09213, doi:10.1029/2008JD011278, 2009.

Li, J., Carlson, B. E., and Lacis, A. A.: El Niño-Southern Oscillation correlated aerosol Ångström Exponent anomaly over the tropical Pacific discovered in satellite measurements, J. Geophys. Res., 116, D20204, doi:10.1029/2011JD015733, 2011.

Li, J., Carlson, B. E., and Lacis, A. A.: Application of spectral analysis techniques in the intercomparison of aerosol data: 1. An EOF approach to analyze the spatial-temporal variability of aerosol optical depth using multiple remote sensing data sets, J. Geophys. Res. Atmos., 118, 8640–8648, doi:10.1002/jgrd.50686, 2013.

Li, J., Carlson, B. E., and Lacis, A. A.: Application of spectral analysis techniques in the inter-comparison of aerosol data, Part II: Using maximum covariance analysis to effectively compare spatio-temporal variability of satellite and AERONET measured aerosol optical depth, J. Geophys. Res., 119, 153–166, doi:10.1002/2013JD020537, 2014a.

Li, J., Carlson, B. E., and Lacis, A. A.: Application of spectral analysis techniques in the inter-comparison of aerosol data, Part III:

Using combined PCA to compare spatio-temporal variability of MODIS, MISR and OMI aerosol optical depth, J. Geophys. Res., 119, 4017–4042, doi:10.1002/2013JD020538, 2014b.

Li, X., Holland, D. M., Gerber, E. P., and Yoo, C.: Impacts of the north and tropical Atlantic Ocean on the Antarctic Peninsula and sea ice, Nature, 505, 538–542, 2014.

Liu, L., Lacis, A. A., Carlson, B. E., Mishchenko, M. I., and Cairns, B.: Assessing Goddard Institute for Space Studies ModelE aerosol climatology using satellite and ground-based measurements: A comparison study, J. Geophys. Res., 111, D20212, doi:10.1029/2006JD007334, 2006.

Nabat, P., Somot, S., Mallet, M., Chiapello, I., Morcrette, J. J., Solmon, F., Szopa, S., Dulac, F., Collins, W., Ghan, S., Horowitz, L. W., Lamarque, J. F., Lee, Y. H., Naik, V., Nagashima, T., Shindell, D., and Skeie, R.: A 4-D climatology (1979–2009) of the monthly tropospheric aerosol optical depth distribution over the Mediterranean region from a comparative evaluation and blending of remote sensing and model products, Atmos. Meas. Tech., 6, 1287–1314, doi:10.5194/amt-6-1287-2013, 2013.

Prasad, A. K. and Singh, R. P.: Comparison of MISR-MODIS aerosol optical depth over the Indo-Gangetic basin during the winter and summer seasons (2000–2005), Remote Sens. Environ., 107, 109–119, 2007.

Sayer, A. M., Hsu, N. C., Bettenhausen, C., Jeong, M.-J., Holben, B. N., and Zhang, J.: Global and regional evaluation of overland spectral aerosol optical depth retrievals from SeaWiFS, Atmos. Meas. Tech., 5, 1761–1778, doi:10.5194/amt-5-1761-2012, 2012.

Singh, R. P., Dey, S., Tripathi, S. N., Tare, V., and Holben, B.: Variability of aerosol parameters over Kanpur, northern India, J. Geophys. Res., 109, D23206, doi:10.1029/2004JD004966, 2004.

Smirnov, A., Holben, B. N., Eck, T. F., Dubovik, O., and Slutsker, I.: Cloud screening and quality control algorithms for the AERONET database, Remote Sens. Env., 73, 337–349, 2000.

Tanré, D., Kaufman, Y. J., Herman, M., and Mattoo, S.: Remote sensing of aerosol properties over oceans using the MODIS/EOS spectral radiances, J. Geophys. Res., 102, 16971–16988, doi:10.1029/96JD03437, 1997.

Torres, O., Tanskanen, A., Veihelmann, B., Ahn, C., Braak, R., Bhartia, P. K., Veefkind, P., and Levelt, P.: Aerosols and surface UV products from Ozone Monitoring Instrument observations: An overview, J. Geophys. Res., 112, D24S47, doi:10.1029/2007JD008809, 2007.

Torres, O., Chen, Z., Jethva, H., Ahn, C., Freitas, S. R., and Bhartia, P. K.: OMI and MODIS observations of the anomalous 2008–2009 Southern Hemisphere biomass burning seasons, Atmos. Chem. Phys., 10, 3505–3513, doi:10.5194/acp-10-3505-2010, 2010.

Torres, O., Ahn, C., and Chen, Z.: Improvements to the OMI near-UV aerosol algorithm using A-train CALIOP and AIRS observations, Atmos. Meas. Tech., 6, 3257–3270, doi:10.5194/amt-6-3257-2013, 2013.

Trivitayanurak, W., Palmer, P. I., Barkley, M. P., Robinson, N. H., Coe, H., and Oram, D. E.: The composition and variability of atmospheric aerosol over Southeast Asia during 2008, Atmos. Chem. Phys., 12, 1083–1100, doi:10.5194/acp-12-1083-2012, 2012.

van der Werf, G. R., Randerson, J. T., Giglio, L., Collatz, G. J., Kasibhatla, P. S., and Arellano Jr., A. F.: Interannual variability in global biomass burning emissions from 1997 to 2004, Atmos. Chem. Phys., 6, 3423–3441, doi:10.5194/acp-6-3423-2006, 2006.

Witte, J. C., Douglass, A. R., da Silva, A., Torres, O., Levy, R., and Duncan, B. N.: NASA A-Train and Terra observations of the 2010 Russian wildfires, Atmos. Chem. Phys., 11, 9287–9301, doi:10.5194/acp-11-9287-2011, 2011.

Xiao, N., Shi, T., Calder, C. A., Munroe, D. K., Berrett, C., Wolfinbarger, S., and Li, D.: Spatial characteristics of the difference between MISR and MODIS aerosol optical depth retrievals over mainland Southeast Asia, Remote Sens. Environ., 113, 1–9, doi:10.1016/j.rse.2008.07.011, 2009.