

# A global aerosol classification algorithm incorporating multiple satellite data sets of aerosol and trace gas abundances

M.J.M. Penning de Vries<sup>1</sup>, S. Beirle<sup>1</sup>, C. Hörmann<sup>1</sup>, J.W. Kaiser<sup>1</sup>, P. Stammes<sup>2</sup>,  
L.G. Tilstra<sup>2</sup>, O.N.E. Tuinder<sup>2</sup>, and T. Wagner<sup>1</sup>

<sup>1</sup>Max Planck Institute for Chemistry, Mainz, Germany

<sup>2</sup>Royal Netherlands Meteorological Institute (KNMI), De Bilt, Netherlands

*Correspondence to:* M. Penning de Vries  
(marloes.penningdevries@mpic.de)

**Abstract.** Detecting the optical properties of aerosols using passive satellite-borne measurements alone is a difficult task due to the broad-band effect of aerosols on the measured spectra and the influences of surface and cloud reflection. We present another approach to determine aerosol type, namely by studying the relationship of aerosol optical depth (AOD) with trace gas abundance, aerosol absorption, and mean aerosol size. Our new Global Aerosol Classification Algorithm, GACA, examines relationships between aerosol properties (AOD and extinction Ångström exponent from the Moderate Resolution Imaging Spectroradiometer (MODIS), UV Aerosol Index from the second Global Ozone Monitoring Experiment, GOME-2) and trace gas column densities (NO<sub>2</sub>, HCHO, SO<sub>2</sub> from GOME-2, and CO from MOPITT, the Measurements of Pollution in the Troposphere instrument) on a monthly mean basis. First, aerosol types are separated based on size (Ångström exponent) and absorption (UV Aerosol Index), then the dominating sources are identified based on mean trace gas columns and their correlation with AOD. In this way, global maps of dominant aerosol type and main source type are constructed for each season and compared with maps of aerosol composition from the global MACC (Monitoring Atmospheric Composition and Climate) model. Although GACA cannot correctly characterize transported or mixed aerosols, GACA and MACC show good agreement regarding the global seasonal cycle, particularly for urban/industrial aerosols. The seasonal cycles of both aerosol type and source are also studied in more detail for selected 5° × 5° regions. Again, good agreement between GACA and MACC is found for all regions, but some systematic differences become apparent: the variability of aerosol composition (yearly and/or seasonal) is often not well captured by MACC, the amount of mineral dust outside of the dust belt appears to be overestimated, and the abundance of secondary organic aerosols is underestimated in comparison with GACA. Whereas the presented study is of exploratory nature, we show that the developed algorithm is well suited to evaluate climate and atmospheric composition models by including aerosol type and source obtained from measurements into the comparison, instead of focusing on a single

25 parameter, e.g. AOD. The approach could be adapted to constrain the mix of aerosol types during  
the process of a combined data assimilation of aerosol and trace gas observations.

## 1 Introduction

Measurements of aerosol optical depth (AOD) – by ground-based, airborne, and satellite-borne instruments – have provided us with a good picture of the highly variable distribution of aerosols  
30 throughout the globe. The uncertainties in our knowledge of the global distribution of aerosol loading have become progressively smaller during the past decade owing to dedicated satellite-borne aerosol instruments like the Moderate Resolution Imaging Spectroradiometer and the Multi-angle Imaging Spectroradiometer (MODIS and MISR, see e.g., Remer et al., 2005; Kahn et al., 2005; Kokhanovsky and de Leeuw, 2009; Chin et al., 2014, and references therein). However, for many  
35 applications the aerosol amount tells only half of the story: to study the interaction between aerosols and clouds (Rosenfeld et al., 2014), to determine aerosol radiative effects, and for the development of mitigation strategies it is crucial to additionally know the aerosol type or source (e.g., IPCC, 2013). For remote sensing retrievals themselves, aerosol optical properties or some constraints on particle type are also needed to aid model selection in the inversion process.

40 The contribution of aerosols to the top-of-atmosphere radiance detected by satellite instruments is spectrally smooth, and due to the interfering signal from the surface, passive radiometers like MODIS cannot retrieve more than one or two pieces of information from their measurements: AOD and the Extinction Ångström Exponent, EAE. The EAE, as a proxy for the particle size distribution, turns out to be a very useful metric when characterizing aerosol types. Naturally emitted primary  
45 aerosols, such as mineral dust and sea salt, consist of relatively large particles with a size distribution centered at sizes  $> 1 \mu\text{m}$ . In contrast, secondary aerosols – those formed from components emitted in gaseous form – are generally (much) smaller than  $1 \mu\text{m}$  (i.e. the extinction is almost entirely due to small particles, Seinfeld and Pandis, 2006). The majority of such “fine” particles is often assumed to be of anthropogenic origin (Kaufman et al., 2002), although biomass burning aerosols, which consist  
50 mostly of fine particles, are not all human-induced. In addition, there are strong biogenic sources of small secondary organic aerosols. To further discriminate between aerosol types, differences in absorption can be exploited (as e.g., in Higurashi and Nakajima, 2002; Jeong and Li, 2005; Kim et al., 2007; Mielonen et al., 2009). This allows the distinction of desert dust (large particles that absorb in the UV range) from sea salt (large, but non-absorbing), and smoke (small, absorbing) from  
55 industrial pollution (small, weakly or non-absorbing), to name a few examples. In practice, such simple rules are often violated: aging of particles (hygroscopic growth, coating or other processes) or mixing of different aerosol types change the optical properties. To determine the (most probable) main aerosol source, more information is required. We use measurements of trace gas abundances as a source of this information.

60 Apart from naturally formed particles (desert dust and sea salt), aerosols are often accompanied by enhanced trace gas levels – because they were emitted by the same source, or were formed from those trace gases or from the same precursor. Hence, collocated measurements of trace gases can be used to determine the main source of aerosols. This has been exploited in a study by Veefkind et al. (2011), in which it was shown that the presence of significant correlation of AOD with trace gas concentrations, notably NO<sub>2</sub> and HCHO, is an indication of the main source of those aerosols. In a later publication, also involving data from the Ozone Monitoring Instrument (OMI), Torres et al. (2013) demonstrated that the use of CO data from the Atmospheric Infrared Sounder (AIRS) to identify smoke improves the aerosol retrieval by OMI. In the present study, we take these findings a step further and integrate them into an algorithm to determine the main aerosol type and its source on a global scale. We extend the analysis initiated by Veefkind et al. (2011) by adding CO abundance and aerosol optical properties. The resulting Global Aerosol Classification Algorithm, GACA, combines the EAE from MODIS and UV Aerosol Index (UVAI) from GOME-2 (Global Ozone Monitoring Experiment-2) to determine an aerosol type based on its size and absorption. Subsequently, trace gas vertical column densities (VCDs of NO<sub>2</sub>, HCHO, SO<sub>2</sub>, and CO) are used to infer the dominating source of the aerosols. The main results from this algorithm are seasonal maps that show the dominating aerosol type and source at 1° × 1° or 2° × 2° resolution, respectively.

GACA results are compared to aerosol composition from MACC (Monitoring Atmospheric Composition and Climate) reanalysis data on a global and regional scale. The MACC project provides data on atmospheric composition for the recent past and makes mid-term forecasts by combining state-of-the-art atmospheric modelling with satellite-based measurements (e.g., Inness et al., 2013). The model assimilates AOD from both MODIS instruments, using it to scale the total aerosol mixing ratio. The tropospheric aerosol types (or components) included in MACC are sea salt, desert dust, organic matter, black carbon, and sulfate. The comparison with model data highlights an important application of our algorithm: the improvement of emissions of both trace gases and aerosols in models (as suggested in e.g., Xu et al., 2013).

In this paper we present GACA and demonstrate its capabilities with seasonal global maps of aerosol type and main source, seasonal cycles of aerosol type and source in six selected regions, and several other applications. We find good agreement between results from GACA and MACC reanalysis in most cases; some important discrepancies between the data sets are discussed. The paper is structured as follows: first, we describe the instruments and data sets used in GACA. The algorithm is described in detail in Sect. 3. Global maps of aerosol type and aerosol source determined by GACA are presented and compared with maps of aerosol composition from the MACC reanalysis in Sect. 4, where the study of the seasonal cycle in six study regions is also shown. In Sect. 5 the sensitivity of GACA to various parameters is discussed, GACA results are compared to existing aerosol climatologies, and future improvements to the algorithm are suggested; the closing Sect. 6 contains our concluding remarks.

## 2 Instruments and data

### 2.1 Satellite instruments

There are two MODIS instruments in operation: one each on NASA’s Aqua and Terra satellites.  
100 Designed to detect aerosols, the MODIS instruments measure reflectances in 36 wavelength bands at high spatial resolution (of the order of  $1 \text{ km}^2$  or less) with a swath wide enough (2600 km) to provide daily global coverage (Justice and Townshend, 2002, and references therein). Aqua is part of the A-Train, crosses the equator at 13:30 LT and performs its daylight measurements on the ascending part of its orbit; Terra is in a daytime-descending orbit and has a local equator crossing time of about  
105 10:30 LT.

The MOPITT instrument (Pan et al., 1998) is also part of the payload of the Terra satellite. MOPITT pixels measure  $22 \text{ km} \times 22 \text{ km}$  and the swath of the instrument is 640 km, hence global coverage is reached approximately every 3 days.

GOME-2 on MetOp-A is a spectrometer that measures backscattered radiance in the UV-NIR  
110 range (240–790 nm) with a nominal spatial resolution of  $40 \text{ km} \times 80 \text{ km}$  (Callies et al., 2000). The swath width of the GOME-2 instrument is 1920 km, permitting global coverage in 1.5 days. MetOp-A was launched in 2006 into a daytime-descending orbit with a local equator crossing time of 09:30 LT.

### 2.2 Data sets

115 The data sets that are used as input to GACA are briefly introduced in this section; for details we refer to the literature and websites listed in Table 1.

#### 2.2.1 Aerosol Optical Depth and extinction Ångström exponent

Monthly mean values of AOD (or  $\tau$ ) from MODIS collection 5.1 were obtained at  $1^\circ \times 1^\circ$  resolution. The retrieval algorithms for aerosols over ocean and dark land are described in Remer et al. (2005)  
120 and Levy et al. (2007b), respectively. For bright surfaces where no AOD value is available from the dark target algorithm (mainly deserts), the Deep Blue product (Hsu et al., 2004) is used. In this study data from the MODIS instrument on Aqua are used – despite the better agreement of the overpass times of GOME-2 and Terra – because the Deep Blue data set of Terra reaches only up to 2007 due to a missing polarization correction. The Level-3 reprocessing of collection 6, in which the calibration  
125 of both MODIS instruments is improved and several other algorithm updates have been made (Levy et al., 2013; Lyapustin et al., 2014), is incomplete at the time of writing. MODIS AOD is given at 550 nm. Monthly mean EAE ( $\alpha$ ) is calculated according to Eq. (1) from the mean MODIS AOD:

$$\alpha = -\frac{\log(\tau_{\lambda_2}/\tau_{\lambda_1})}{\log(\lambda_2/\lambda_1)} \quad (1)$$

with  $\tau_\lambda$  the monthly mean AOD at the wavelengths  $\lambda_1 = 470$  nm and  $\lambda_2 = 660$  nm. Those are the  
130 only two channels for which AOD is determined for land, ocean, and bright surfaces. The EAE  
was chosen over the fine-mode fraction (FMF), because FMF is not part of the Deep Blue aerosol  
product, thus no aerosol size information would be available over deserts and other bright surfaces.  
A more detailed discussion of EAE and FMF appears in Sect. 5.2.

### 2.2.2 UV Aerosol Index

135 The UVAI is a semi-quantitative indicator of aerosols. Positive values of UVAI are generally re-  
ferred to as “Absorbing Aerosol Index (AAI)”, which is a measure of aerosols that absorb UV radi-  
ation (Torres et al., 1998; de Graaf et al., 2005). For UVAI  $< 0$ , which can be used for the detection  
of non-absorbing aerosols (Penning de Vries et al., 2009), the term “Scattering Index (SCI)” was  
suggested. The UVAI is a complex function of AOD, aerosol absorption, and layer altitude, and  
140 using it in a quantitative sense is not straightforward. However, in combination with auxiliary in-  
formation on aerosol abundance (i.e., AOD), information on aerosol absorption can be derived from  
UVAI. Although “AAI” is more often used in literature, we prefer to use the term “UVAI”, as we use  
both the positive and negative values of the Aerosol Index. Level-2 operational UVAI (determined  
using 340 and 380 nm GOME-2 reflectances) from the O3M SAF (Satellite Application Facility  
145 for Atmospheric Composition and UV Radiation, o3msaf.fmi.fi) were obtained from the Tropo-  
spheric Emission Monitoring Internet Service (TEMIS); a description of the algorithm can be found  
in de Graaf et al. (2005, 2014). The UVAI were corrected for the effects of instrument degradation  
using empirically derived in-flight reflection correction factors (Tilstra et al., 2012). The data were  
filtered for sunglint, single scattering angles smaller than  $90^\circ$ , and solar eclipses, as recommended  
150 in the “ATBD for the GOME-2 Aerosol products” by de Graaf et al. (2014). In addition, data with  
FRESCO effective cloud fractions (Wang et al., 2008) exceeding 0.2 or solar zenith angle (SZA)  
over  $80^\circ$  were discarded prior to gridding and averaging to comply with the data selection of the  
trace gases measured by GOME-2 (see next section).

### 2.2.3 Trace gases

155 Total column densities of  $\text{SO}_2$  and HCHO, and tropospheric column densities of  $\text{NO}_2$  are retrieved  
by DOAS analysis (Differential Optical Absorption Spectroscopy, see e.g., Platt and Stutz (2008);  
Richter and Wagner (2011)) of GOME-2 spectra in the UV-visible range.

For our study, TM4NO2A version 2.1 Level-2  $\text{NO}_2$  tropospheric VCDs were obtained from  
TEMIS. The retrieval of  $\text{NO}_2$  from GOME, similarly applied to GOME-2, is described in Boersma  
160 et al. (2004).

Version 12 GOME-2 HCHO VCDs were downloaded from h2co.aeronomie.be; the retrieval  
description can be found in De Smedt et al. (2012). Level-2 HCHO data are only available for  
SZA  $< 80^\circ$ .

Our retrieval of GOME-2 SO<sub>2</sub> data is described in detail in Hörmann et al. (2013). It takes into  
165 account non-linear effects that may occur for high SO<sub>2</sub> concentrations.

All GOME-2 trace gas data were filtered by FRESCO cloud fraction (CF < 0.2, unless stated  
otherwise) and SZA < 80°, subsequently gridded to 1° × 1° resolution and averaged for each month  
of the years 2007–2011.

Monthly mean, gridded version-6 MOPITT CO total VCDs were obtained from the Atmospheric  
170 Science Data Center (ASDC). We used results from the combined near- and thermal infrared (NIR-  
TIR) retrieval because combination of the two spectral regions greatly improves the sensitivity to  
the lower troposphere (Deeter et al., 2003, 2013). A recent validation of the NIR-TIR algorithm  
found relatively large random retrieval errors and bias drift (Deeter et al., 2013), but these are not  
expected to significantly influence our results for two reasons: first, we use monthly mean data on  
175 a coarse 1° × 1° grid which reduces random errors. And second, we use the excess CO (value minus  
background) instead of the absolute value, which should remove a time-dependent bias. The total  
excess CO column used here (denoted as ΔCO) is obtained by subtracting a background column  
that is the median of the data within each 5° latitude band. This procedure is needed due to the long  
life time of CO and allows the use of a single CO threshold value throughout the year and for the  
180 whole globe.

#### 2.2.4 MACC model data

The MACC reanalysis was developed and produced during the series of EU-funded GEMS (Global  
and regional Earth-system (Atmosphere) Monitoring using Satellite and in-situ data), MACC and  
MACC-II (MACC-Interim Implementation) projects. These projects developed the operational Coper-  
185 nicus Atmosphere Monitoring Services (CAMS), which was launched in November 2014. It de-  
livers global atmospheric composition analyses and forecasts and European air quality forecasts  
every day. While the main developments were aimed at real-time production, periodic reanalyses  
have been planned from the outset to provide consistent time series for various scientific applica-  
tions (Hollingsworth et al., 2008, [www.copernicus-atmosphere.eu](http://www.copernicus-atmosphere.eu)). The aerosol model is integrated  
190 into the European Centre for Medium-Range Weather Forecasts (ECMWF) Integrated Forecasting  
System (IFS) for numerical weather predictions and uses the total aerosol mixing ratio as a con-  
trol variable. Five types of tropospheric aerosols are included: sea salt, desert dust, organic matter,  
black carbon, and sulfate. Aerosols of natural origin (sea salt and desert dust) are related to model  
parameters (wind speed and soil moisture), whereas anthropogenic aerosol emissions come from  
195 inventories (Morcrette et al., 2009). In particular, biomass burning emissions are distributed with  
0.5° and 1 day resolution according to GFASv1.0 (Kaiser et al., 2012), with monthly budgets before  
2009 scaled to GFED3.0 (Van der Werf et al., 2010). The aerosol assimilation system uses AOD  
from both MODIS sensors at the time and location of overpass to scale the total aerosol abundance,

while retaining the fractional contribution of each aerosol component to the total mass (Benedetti et al., 2009).

### 3 Global Aerosol Classification Algorithm description

GACA is based on the outcome of several tests applied to the trace gas and aerosol data described in the previous section, and their correlation with AOD. The algorithm consists of two main parts: the first part, named GACA-type, assigns certain aerosol types to each data point within a grid box based on UVAI (a measure of aerosol absorption) and EAE (a measure of aerosol size). The second part, GACA-source, relates trace gas abundance to the different aerosol types and assigns the most probable aerosol source to each grid box. Both parts will be described in detail in Sects. 3.2 and 3.3, and are summarized in the decision tree in Fig. 2.

#### 3.1 Data selection

Prior to analysis, GACA performs a selection of data for each “grid box”. For a final map with a resolution of  $2^\circ \times 2^\circ$  (which was chosen as a compromise between spatial resolution and statistics), each grid box on the globe contains 4 data points per month, because the input monthly mean maps (of AOD, UVAI, EAE, and trace gas column densities) have a resolution of  $1^\circ \times 1^\circ$ . To improve statistics and stability of the algorithm, the data are grouped by season and five years of data (2007–2011) are combined, increasing the number of data points to 60. Grid boxes in which the monthly mean AOD never exceeds 0.05 are removed, as it is assumed that they cannot be reliably classified. The obtained data set is screened for missing values and outliers; the latter because the intention is to build a climatology of typical conditions, which should not be influenced by exceptional events. In addition, faulty retrievals (e.g. due to the South Atlantic Anomaly) are removed. Outliers are removed by repeated exclusion of data points exceeding the mean-plus-three-sigma criterion until all data fall within the three-sigma range. Whenever an AOD, EAE or UVAI outlier is encountered, all corresponding values (collocated AOD, UVAI, EAE, and trace gas columns) are removed from the data set. Trace gas outliers are also excluded, but in this case only the affected data point is removed. Hence, if an  $\text{NO}_2$  outlier is encountered, the  $\text{NO}_2$  value is removed, but  $\text{HCHO}$ ,  $\text{SO}_2$ , and  $\Delta\text{CO}$  columns and aerosol data are retained (i.e., in this case the mean  $\text{NO}_2$  VCD is calculated with one data point less than the means of the other trace gases and aerosol data; the same applies to the calculation of the correlation with AOD). If outliers are not removed from the data set, GACA results are not strongly affected, but the effects of local extreme events (fires, volcanic eruptions) become apparent. This is discussed in more detail in Sect. 5.1.

### 230 3.2 Aerosol type classification by GACA-type

Each point of the filtered data set is subsequently assigned one of nine aerosol types based on its UVAI and EAE values. In this study, aerosol types are defined by their size – small (S), medium (M), and large (L) – and the amount of aerosol absorption in the UV range – non-absorbing (NA), neutral (N), or absorbing (A) – as shown in the left panel of Fig. 1. The acronyms of aerosol types and sources are explained in Table 2.

The choice of UVAI and EAE thresholds is motivated by the right panel of Fig. 1, which displays monthly mean data (June–August 2007–2011) from regions which we assume to be dominated by one of four aerosol sources: mineral dust (14–26° N/16° W–8° E), smoke (4–16° S/14–30° E), biogenic secondary organic aerosols (30–36° N/80–90° W), and sea salt (0–10° S/120–140° W). The depicted aerosols are clearly separated by the EAE thresholds (sea salt from secondary organic aerosols; desert dust from smoke) and the UVAI thresholds (desert dust from sea salt; smoke from secondary organic aerosols). The choice of nine aerosol types instead of four (like in Higurashi and Nakajima, 2002) was motivated by the occurrence of situations where different particle types are mixed.

For each  $2^\circ \times 2^\circ$  grid box, the fraction of data points belonging to each aerosol type is computed and the most frequently observed type, weighted by AOD, is assumed to be the dominant type. Note that if the type classification is run on its own (i.e., not as input for the aerosol source assignment step), the statistics requirements are less strict and global maps can be produced on  $1^\circ \times 1^\circ$  resolution (e.g., Fig. 4).

### 250 3.3 Aerosol source assignment by GACA-source

The results from GACA-type are used as input for the second part of GACA: the determination of the dominant aerosol source. The main assumption underlying GACA-source is that enhancements in trace gas and aerosol abundance are caused by the same source. The algorithm computes means over all data points within a grid box (of AOD, UVAI, and trace gas VCDs) and correlations between AOD on the one hand, and UVAI and trace gas VCDs on the other. Together with the dominant aerosol type determined in the previous step, these data are used to assign a main aerosol source based on the outcome of two types of tests: (1) is the mean trace gas abundance or HCHO : NO<sub>2</sub> ratio above the threshold given in Table 3? (2) Is there a linear correlation (with  $R^2 > 0.25$ ) between AOD and UVAI or AOD and trace gas abundance? An overview of GACA-source can be found in the lower part of the decision tree in Fig 2.

Eight aerosol sources are discriminated in GACA-source: biomass burning smoke, desert dust, secondary biogenic, secondary urban/industrial, aged, volcanic sulfate, sea salt, and unknown sources. Each source and the selected classification criteria will be described in more detail in the following sections.



### 265 **3.3.1 Biomass burning smoke (BB)**

Fresh smoke from forest, agricultural, or grassland fires mainly consists of small particles (e.g., Dubovik et al., 2002; Eck et al., 2013) that absorb light in the UV and visible range. Co-emitted trace gases are NO<sub>2</sub>, HCHO, and CO; SO<sub>2</sub> only in very small amounts (Andreae and Merlet, 2001). In GACA-source, grid boxes are always designated BB when the main type is small absorbing. Biomass  
270 burning is also assigned if the absorbing aerosol criterion is fulfilled and either (1) mean CO or (2) correlation between  $\Delta$ CO and AOD or (3) mean HCHO and correlation between HCHO and AOD pass the threshold. The absorbing aerosol criterion requires that either (a) the dominant aerosol type is absorbing; or (b) the dominant type is neutral and a good correlation with a positive slope is found for UVAI and AOD, and mean AOD  $\geq$  0.15. This allows grid boxes with relatively small UVAI (e.g.,  
275 due to lower-lying aerosol layers or cloud contamination) to be designated as BB.

### **3.3.2 Desert dust (DD)**

Mineral dust consists of large, non-spherical particles that absorb UV radiation due mainly to their iron oxide content (Sokolik and Toon, 1999). The emission and transport of DD is linked to meteorology (i.e., wind fields) and land surface conditions and not to trace gas emissions. GACA-type assigns  
280 DD as a source to grid boxes that are dominated by large absorbing aerosols – unless they were already characterized as BB. To include aged DD plumes, medium-size and large neutral aerosol types can be attributed to DD if the absorbing aerosol criterion is fulfilled (see above), but additionally, the correlation of  $\Delta$ CO and AOD and means of the other trace gases (NO<sub>2</sub>, HCHO, and SO<sub>2</sub>) should be below their respective threshold values. The latter criterion serves to distinguish DD from BB and  
285 volcanic ash, but as a negative side-effect excludes polluted dust and cases of mixed desert dust and smoke.

### **3.3.3 Secondary aerosols biogenic origin (BIO)**

The small, non-absorbing aerosols that form by condensation of (semi-) volatile biogenic precursors are accompanied by enhanced levels of HCHO, as both are products of the oxidation of isoprene  
290 and other volatile organic compounds (Seinfeld and Pandis, 2006; Goldstein et al., 2009; Stavrou et al., 2009). To separate them from urban/industrial aerosols, the ratio of HCHO/NO<sub>2</sub> is required to be above a certain threshold value (given in Table 3).

### **3.3.4 Secondary aerosols of urban/industrial origin (URB)**

Due to the diversity of sources and chemical processing in industrialized environments, the URB  
295 source is very broadly defined in GACA-source. All grid boxes dominated by non-absorbing or neutral aerosol types that have enhanced NO<sub>2</sub> columns qualify. The only exception is grid boxes already characterized as BIO.

### 3.3.5 Aged/transported aerosols (AGED)

Air masses with enhanced  $\Delta\text{CO}$ , but low levels of  $\text{NO}_2$  are assumed to have been transported away  
300 from their sources. The AGED source is therefore assigned when CO, which has a long life time, is  
enhanced, but the shorter-lived  $\text{NO}_2$  is not. Aging may change average aerosol properties by dilution,  
mixing with other air masses, processing within clouds, or other mechanisms. Hence, all neutral and  
non-absorbing aerosol types qualify as AGED.

### 3.3.6 Volcanic sulfate (VOG)

305 Secondary aerosols formed by the reaction of volcanic  $\text{SO}_2$  with the atmosphere are named volcanic  
smog (VOG) here to distinguish them from anthropogenic sulfate. GACA-source can only detect  
VOG in remote locations, as one requirement for the assignment is the lack of enhancements in  $\text{NO}_2$   
and  $\Delta\text{CO}$ . In addition, the  $\text{SO}_2$  mean and correlation with AOD need to pass the thresholds. Freshly  
formed sulfate aerosols are small, but can grow rapidly due to their hygroscopicity; therefore small  
310 and medium-sized aerosol types can be assigned to VOG. Both non-absorbing and neutral aerosol  
types qualify because the sensitivity of UVAI to non-absorbing aerosols is not very high.

### 3.3.7 Sea salt (SS)

Breaking waves and bursting bubbles cause the release of sea salt particles. The particles are hygro-  
scopic and grow readily in the marine boundary layer, forming large, non-absorbing particles. The  
315 emission of SS depends mainly on wind speed and geography (e.g. coastlines), but is not associ-  
ated with the emission of trace gases. GACA-source attributes SS as a main source to grid boxes  
with mean AOD  $< 0.15$  and no trace gas enhancements; only non-absorbing and neutral, large and  
medium-size type aerosols are eligible candidates. GACA does not discriminate between grid boxes  
located over land and ocean, therefore the SS type is also regularly found over land and may be  
320 interpreted as a generic background type.

### 3.3.8 Unknown source (XX)

If all tests leading to the above-mentioned aerosol sources fail, but significant amounts of aerosols  
are detected (mean AOD  $> 0.05$ ) the aerosol source is set to “unknown”.

### 3.3.9 Source assignment

325 Means and correlation coefficients are calculated from all valid data points within a grid box if  
the fraction of valid points amounts to at least 25% of all points (down to an absolute minimum  
of 5). The tests performed by GACA-source are based on thresholds, given in Table 3, of which  
the values were chosen empirically. The source assignment criteria were chosen based on textbook  
knowledge (e.g., that biomass burning is associated with HCHO and CO emissions), as detailed for

330 each source type in Sects. 3.3.1-3.3.8, and were adjusted iteratively to obtain consistent results. The quantitative understanding of aerosol-trace gas relationships, however, is currently not sufficient to derive trace gas thresholds in a systematic way, hence the trace gas thresholds were determined in a more empirical fashion. The thresholds were empirically chosen high enough to exclude noise (or natural variability), but low enough that the associated sources are recognized. The  $\Delta\text{CO}$  threshold, 335 for example, was chosen low enough to include aged air masses. The  $\text{SO}_2$  threshold, on the other hand, had to be set sufficiently high to exclude noise. The thresholds were chosen independent of region and season to keep the algorithm globally consistent. A future development of GACA may be the adoption of threshold climatologies to better account for regional and seasonal variability of trace gas and aerosol emissions (see Sect. 5.4).

340 In Fig. 3 we demonstrate the algorithm for two  $2^\circ \times 2^\circ$  grid boxes: the first shows data from a region in central Africa ( $2-4^\circ \text{S}/18-20^\circ \text{E}$ ) during the biomass burning season, whereas the second is located west of Hawaii, in a region of volcanic outflow at  $16-18^\circ \text{N}$  and  $162-164^\circ \text{W}$ . The trace gas columns for June–August 2007–2011 are plotted together with their respective thresholds (colored lines), so that if data points lie above the respective threshold, the trace gas is assumed to be associated with the local aerosols. In the left panel (central Africa), HCHO (green) and  $\Delta\text{CO}$  (light blue) 345 are strongly enhanced. The level of  $\text{NO}_2$  (blue) clearly exceeds  $10^{-15}$ , the threshold for both  $\text{NO}_2$  and  $\text{SO}_2$ . This is in contrast to  $\text{SO}_2$ , which is close to or even below the detection limit, leading to scatter of data and negative values. The dominating source is biomass burning (BB), because (1) the dominating aerosol type is medium-size absorbing and (2) the correlation between  $\Delta\text{CO}$  and AOD 350 is high ( $R^2 = 0.71$ ).

Over the remote eastern Pacific Ocean (right panel) the trace gas means and correlations usually fall below the threshold values, but due to prodigious degassing of Mount Kilauea (especially in 2008) strongly enhanced  $\text{SO}_2$  columns can be observed in the selected grid box. In the atmosphere  $\text{SO}_2$  is converted to sulfate aerosols, resulting in a good correlation between AOD and  $\text{SO}_2$  of 355  $R^2 = 0.53$ . The dominating aerosol types are large neutral and large non-absorbing; the main source assigned to this grid box is volcanic sulfate (VOG).

## 4 Results

### 4.1 Aerosol type

We applied GACA-type to the data set from 2007–2011 to study the seasonal cycle of aerosol prop- 360 erties globally. Figure 4 shows maps of the dominating aerosol type on a  $1^\circ \times 1^\circ$  resolution for all four seasons. Focusing first on the summer (third panel), it can be seen that the dust belt, at around  $10-40^\circ \text{N}$ , is dominated by large particles (dark hues) with strong to moderate absorption (red and green tones). Smoke plumes from central Africa consist mostly of small to medium-size absorbing particles (orange and red), although there appears to be a significant contribution from large

365 absorbing (LA) particles, which is probably an artifact that will be discussed in more detail in the  
next section. North America, Europe and large parts of Asia are dominated by small, non-absorbing  
aerosols (light blue). Over ocean, particularly in the southern oceans, large particles (dark blue and  
green) dominate. Light gray areas denote regions where no AOD data were available (due to e.g.  
clouds, snow or ice cover, low sun) or where monthly mean AOD did not exceed 0.05 within the  
370 studied period.

In winter and spring (December–February; March–May) the contribution of mineral dust to the  
aerosol mix over China can be clearly seen: the aerosol type is dominated by larger, more strongly  
absorbing particles than in summer. The burning of cropland and agricultural waste in Southeast  
Asia stands out in spring, when aerosol types are predominantly absorbing (red and orange). The  
375 biomass burning season in South America, which starts in July–August and peaks in September–  
October, has a very different signature than that in Southern Africa: the particles are smaller and  
appear less absorbing. This may be a consequence of the difference in fuel type (e.g Eck et al.,  
2013), which leads to different trace gas and aerosol emission factors. But the main causes are  
probably the increased cloudiness, which leads to lower UVAI values and more data gaps in the  
380 trace gas products, and the large abundance of (non-absorbing) secondary organic aerosols. Despite  
the fact that wildfires occur frequently in summer in North America, BB is not selected as a major  
source there. This is because forest fires occur at irregular intervals, so that their signal is suppressed  
as a consequence of averaging data in time and space.

The frequency of occurrence of each aerosol type can be used to study changes in aerosol com-  
385 position as a function of time (or distance to the source). As an example, the westward transport of  
Saharan dust over the Atlantic Ocean is shown in Fig. 5. The upper panel displays the mean total  
AOD along a longitudinal transect from 10° E to 80° W, at 15–20° N (see yellow box in panel 3  
of Fig. 4). The lower panel presents the aerosol fraction, weighted by AOD, for the same transect.  
Only the three large aerosol types (LNA, LN, and LA) are shown; the other types never contribute  
390 more than 20 % to the total AOD. Close to the source, situated at roughly 10° E to 10° W, the aerosol  
load is almost completely made up of large absorbing particles (LA, brown triangles). West of about  
25° W, the fraction of large neutral aerosols (LN, green crosses) starts increasing until it becomes  
the dominating particle type at 50° W, where the total AOD has decreased to 0.3 (from a maximum  
of 0.75). This apparent change in absorption is mainly due to the fact that we use UVAI as a mea-  
395 sure for absorption: as UVAI increases with AOD and aerosol altitude, the gradual descent of the  
dust layer (Colarco et al., 2003), combined with the decreasing AOD causes UVAI to fall below the  
upper threshold value of 0.25. This indicates that GACA underestimates dust abundance far from its  
source.

## 4.2 Source type

400 The results from a run of GACA-source with data from 2007–2011 are shown in the form of seasonal global maps with  $2^\circ \times 2^\circ$  resolution in Fig. 6. The upper frame shows the main source type in winter. Most of the continental northern hemispheric aerosols are of urban/industrial origin (URB, dark blue), except where mineral dust (DD, red) predominates (in North Africa, the South-Arabian peninsula, and northwestern China). Biomass burning smoke (BB, dark red) can be found in sub-Saharan Africa in this season, as well as over parts of Southeast Asia. The forested part of South America is a large source of secondary organic particles (BIO, dark green). Aged aerosols (AGED, blue-gray) can be seen in the outflow from Asia (India, China) and are also found in the air masses transported from equatorial Africa over the Atlantic. Most of the aerosols over oceans are classified as sea salt (SS, light blue), although aerosols of undefined composition (XX, dark gray) are found in the Asian outflow over the Pacific and the African outflow over the Atlantic. The band of aerosols at  $40\text{--}60^\circ$  S (also seen in March–May) is caused by unrealistically high AOD mainly due to inaccurate wind speed assumptions and residual cloud contamination in the MODIS retrieval (Levy et al., 2013; Schutgens et al., 2013) and may be ignored. In spring and summer (second and third panel of Fig. 6) more dust is activated within the global dust belt. The amount of biomass burning smoke also increases as first the agricultural fires in Southeast Asia reach their springtime peak, and then the southern hemisphere fire season starts in summer. A conspicuous sulfate (VOG) plume is seen emerging from Hawaii and is mainly due to prodigious degassing in April–October 2008 by the Kilauea volcano ( $19.4^\circ$  N/ $155.3^\circ$  W), (see, e.g. Yuan et al., 2011; Beirle et al., 2014). The misclassification of SS aerosols over continents in the high latitudes is most apparent in fall (lower-most panel). These grid boxes show no enhanced trace gas concentrations and have mean AOD  $< 0.15$ , corresponding to the definition of SS in GACA. These aerosols may be regarded as background aerosols of which the source cannot reliably be determined by GACA.

405  
410  
415  
420

Whereas Fig. 6 depicts the main aerosol source, determined from all data points within a grid box, Fig. 7 shows the aerosol source determined for each of the nine aerosol types separately. The data are from June–August 2007–2011: the same data set as shown in the third panel of Fig. 6. The three absorbing aerosol types (small, medium-size and large) are shown in Fig. 7a–c. Medium-sized and large absorbing aerosols north of the equator are almost exclusively attributed to mineral dust; the apparent band of desert dust at  $60^\circ$  S is caused by a few data points with unrealistically high AOD, as mentioned above, in addition to erroneous (high) UVAI values that are probably caused by small scattering angles ( $90^\circ$  to  $100^\circ$ ) encountered in this region. The smoke plume off the southwestern coast of Africa in panel a is rather unusual, as biomass burning particles are usually small. This is caused by the use of EAE as a measure for aerosol size; although  $EAE < 0.75$ , the FMF in this region is on the order of 0.7, indicating a large fraction of small aerosols (not shown). It is unclear why EAE and FMF show opposing behavior in this region. We speculate that it has to do with the persistent low-cloud cover during the biomass burning season, which may cause enhanced cloud

430  
435

contamination and, possibly, a wrong choice of aerosol model. However, as it was pointed out in various studies, the size information retrieved by MODIS is not very reliable (e.g., Remer et al., 2005; Levy et al., 2010) and we do not pursue the issue further (but see Sect. 5.2 for a discussion on EAE and FMF). GACA assumes that the source of all small absorbing aerosols is biomass burning, therefore no other source type is seen to contribute in Fig. 7c.

Neutral aerosol types, shown in Fig. 7d–f, come from various sources: URB, SS, AGED, MIX, BB, and some DD. Large and medium-sized non-absorbing aerosols (Fig. 7g–h) are dominated by SS with contributions from URB and MIX. Small non-absorbing (Fig. 7i) is the dominating aerosol type throughout the eastern United States, most of Europe and eastern China (compare lower right panel of Fig. 4), where URB is the main source. Large parts of South America and southern Africa can be seen to emit BIO aerosols (which are assumed to be exclusively small non-absorbing particles), but the AOD-weighted main aerosol source in those regions is BB – in contrast to Southeast Asia, where BIO is the dominant aerosol source in this season (Fig. 6). Performing the analysis by GACA-source on each aerosol type separately thus allows insight into the aerosol mixture that cannot be seen when studying the main source map only.

Additional information on the sources can be gained by adding the trace gas information that was not directly used for source assignment. For example: the URB source is assigned based only on the presence of enhanced  $\text{NO}_2$  (after exclusion of BIO as source type; see Fig. 2), but the information on other trace gas means is retained. By adding binary coding, i.e. values of 1, 2, and 4 to grid boxes with enhanced mean values of HCHO,  $\text{SO}_2$ , and  $\Delta\text{CO}$ , respectively, we obtain the map presented in Fig. 8 for June–August 2007–2011. If only  $\text{NO}_2$  is enhanced, the gridbox has an index 0 and appears dark blue. If, in addition, HCHO is enhanced, the gridbox obtains an index  $0 + 1 = 1$  and appears in a lighter shade of blue. Gridboxes with enhanced  $\text{NO}_2$ , HCHO, and  $\Delta\text{CO}$  are indexed  $0 + 1 + 4 = 5$  and are shown in orange. Grid boxes with a main source other than URB are shown in light gray. This analysis reveals a great diversity in urban/industrial emissions. A clear separation can be seen in Europe, with enhancements of  $\text{NO}_2$  in the West, but additionally enhanced HCHO in the East, confirming the findings of (Veefkind et al., 2011). Further East, urban/industrial aerosols are again only associated with increased  $\text{NO}_2$  columns. Throughout most of the Indian subcontinent, both  $\text{NO}_2$  and HCHO are enhanced. The increased HCHO levels are mainly due to human activities: industrial and vehicle exhaust, biogenic emissions from agriculture (direct and indirect, by  $\text{CH}_4$  oxidation) and burning of biomass (e.g., household fires) (Stavrakou et al., 2009). A similar pattern can be seen over northern Thailand. Northeastern China emits large quantities of all trace gases investigated here, in addition to aerosols. It is one of only a few regions where anthropogenic  $\text{SO}_2$  can be detected from satellite – another being the Highveld in South Africa, which stands out in light blue (enhanced  $\text{NO}_2$  and  $\text{SO}_2$ ). Grid boxes colored yellow and orange (increased  $\Delta\text{CO}$ , or increased  $\Delta\text{CO}$  and HCHO levels) mostly appear at the edges of regions with intense biomass burning in this season (Central Africa, Central South America) and are influenced by fire emissions or are possibly

mis-classified. In southern South America, the South Atlantic Anomaly causes errors in trace gas retrievals that show up mainly in erroneously enhanced  $\text{SO}_2$  values. Although the HCHO retrieval  
475 is similarly affected, the threshold used in GACA-source is high enough to exclude those outliers. Urban aerosols in North America are accompanied by enhanced HCHO levels, which is mostly of biogenic origin (Goldstein et al., 2009; Stavrou et al., 2009). The filament with enhanced  $\Delta\text{CO}$  and HCHO seen on the Northeast coast of the United States is possibly due to transported wildfire smoke from the Northeast (Canada/Alaska). Similar patterns are found for the other seasons (see  
480 Fig. S1 in the Supplement).

### 4.3 Comparison with MACC

Main aerosol types from the MACC reanalysis for 2007–2011 were grouped by season and treated analogously to the measured data with respect to the minimum AOD threshold of 0.05 and the removal of outliers. The dominating aerosol component is shown in Fig. 9 in a similar fashion to  
485 Fig. 6, but there are two important differences.

1. The aerosol components are different (Benedetti et al., 2009): black carbon (BC, black) originates mostly from biomass burning, but also occurs in urban regions due to (e.g.) vehicle exhaust or household fires. In Fig. 9, AOD due to BC is additionally shown in contours (AOD = 0.02–0.1), to indicate the regions affected by biomass burning, as BC constitutes  
490 only a small fraction of aerosol emissions by fires. The contribution of organic matter (OM, green) to biomass burning smoke is much greater, but OM also has important biogenic and anthropogenic sources. The types desert dust (DD) and sea salt (SS) are equivalent to the source types of the same name in GACA and are therefore indicated with the same colors (red and light blue, respectively). Sulfate aerosols ( $\text{SO}_4$ ) are indicated in the same color, blue, as URB aerosols in GACA, because the sources are assumed to be similar. The aerosol type is set to MIX when none of the aerosol components contributes more than 50 % to the total AOD.  
495
2. MACC data are also inherently different from GACA data in that each data point contains contributions of each aerosol component (BC, OM, DD,  $\text{SO}_4$ , SS), whereas GACA determines only one dominating aerosol source per grid box, or, at most, one dominating aerosol source  
500 for each aerosol type found in a grid box.

At a first glance, the agreement between GACA-source and MACC is quite good (compare Figs. 6 and 9): the general spatial and seasonal patterns of DD and  $\text{SO}_4$  (or URB) agree well. The biomass burning regions roughly agree, although the model does not show BC in South America in summer, where GACA sees a lot of BB (mostly due to fires in August, as can be seen in MODIS fire count  
505 patterns). In addition, GACA selects BB as the main source of AOD in sub-Saharan Africa in the first half of the year, whereas in MACC DD dominates. On the other hand, the agricultural fires in Southeast Asia in spring are well captured by both GACA and MACC. The main sources of BIO

(or OM) agree in GACA and MACC, but the source in Southeast USA is missed by the model. The differences between MACC and GACA will be discussed in more detail in the following section, where regional seasonal cycles are investigated.

#### 4.4 Regional seasonal cycles

Six  $5^\circ \times 5^\circ$  regions were selected for the study of the seasonal cycle: Central South America (1), Southern Africa (2), Southeast USA (3), Northwest Europe (4), Thailand (5), and Northeast China (6); the regions are shown as enumerated yellow boxes in the third panel of Fig. 6. For each season of each year (2007–2011), the AOD of every aerosol type is shown in panels a1–a6 of Figs. 10–12. The dominant aerosol source was determined for each individual aerosol type separately and is shown in panels b1–b6. The AOD fractions are therefore equal in panels a and b of Figs. 10–12. For example: the bar representing the fall (September–November) of 2007 in Fig. 10a1 contains contributions from MA, SA, SN, and SNA types. The AOD fraction corresponding to SNA reappears in Fig. 10b1 in dark green (BIO), the dominant source of the SN fraction is URB (blue), and the summed AOD from SA and MA types is attributed to BB (brown). Panels c1–c6 of Figs. 10–12, finally, display the AOD corresponding to the MACC aerosol types for the same regions. All data presented in Figs. 10–12 can be found in Tables S1–S6 in the Supplement.

The first two regions, Central South America and Southern Africa (panels a1–c1 and a2–c2 of Fig. 10, respectively), are characterized by seasonal biomass burning. The fire season starts in late summer in South America; the highest number of fires is usually found in fall. The high year-to-year variability of biomass burning in this region is clearly reflected in all three panels. Both GACA and MACC ascribe the larger part of AOD in winter and spring to secondary organic aerosols (BIO and OM in GACA and MACC, respectively). Although the DD contribution in the model appears to be somewhat high (no DD is detected by GACA), the agreement between GACA and MACC is good for this example. Good agreement is also found for Southern Africa, where smoke forms the major part of the aerosol mixture during the fire season in summer, when the highest AOD are detected. All panels show that the year-to-year variation is much smaller than in South America. Urban/industrial aerosols appear to be overestimated by GACA, whereas MACC shows higher contributions of DD.

The regions Southeast USA and Northwest Europe are dominated by non-absorbing aerosols (Fig. 11a3 and a4). Throughout most of the year, aerosols over S.E. USA are of urban/industrial origin (URB and SO<sub>4</sub> for GACA and MACC, respectively). In summer this region is dominated by secondary organic aerosols (Goldstein et al., 2009), clearly seen by GACA (Fig. 11b3), which attributes nearly all AOD to BIO. MACC, on the other hand, only shows a slight increase in OM relative to the other seasons. The contributions of dust and sea salt to the aerosol mixture appear to be too large in the model in comparison to GACA results, which points to sources missing in the model: MACC scales the aerosol amount with MODIS AOD, but keeps the mass fractions of the different aerosol components constant (see Sect. 2.2). Hence if a source is missing, e.g. secondary



organic aerosols, the AOD due to those aerosols is spread over the remaining components. The small  
545 year-to-year variation observed in MACC aerosol composition is a result of this procedure.

There is no clear aerosol seasonal cycle recognizable in N.W. Europe (Fig. 11a4–c4): the AOD  
is rather constant throughout the year and the composition rarely deviates from the urban/industrial  
(URB and SO<sub>4</sub>) type. In winter there is a larger contribution of medium-size and large particles  
(Fig. 11a4), which GACA-source has trouble identifying but which MACC attributes to sea salt. As  
550 in all previous regions, the model sees significant amounts of dust that are not detected by GACA.  
This can partly be explained by too low deposition rates in the model, but may also be due to the fact  
that GACA does not select DD as a source if any trace gas means are enhanced (unless the aerosol  
type is large absorbing).

Figure 12 presents the seasonal cycle for two regions in Asia. In winter and particularly in spring,  
555 agricultural fires in Thailand release large quantities of smoke, as seen by both GACA and MACC  
(Fig. 12a5–c5). During the rainy season (June–October) secondary aerosols dominate, both from  
anthropogenic (URB and SO<sub>4</sub>) and biogenic sources (BIO and OM). MACC finds significant con-  
tributions of dust which are not seen by GACA.

In NE China, the seasonal mean AOD is greater than 0.5 throughout the year for each year from  
560 2007–2011 (Fig. 12a6–c6). Most of the AOD can be attributed to aerosols of anthropogenic origin  
(URB and SO<sub>4</sub>), but a large fraction is caused by mineral dust transported from deserts in Mongolia,  
northern China, and Kazakhstan, especially in winter and spring. In view of their sizes (medium-size  
to large), most of the aerosols characterized as BB by GACA are probably polluted dust or dust in  
the presence of pollution, i.e., NO<sub>2</sub>, HCHO, SO<sub>2</sub> or  $\Delta$ CO. The variability of the seasonal cycle of  
565 DD appears to be underestimated by MACC (compare Fig. 12a6 and c6). The amount of modeled  
BC in China is as high as for South America in the biomass burning season (see Fig. 10c1), which  
may reflected by the high levels of aerosol absorption found by GACA for northeastern China. The  
more probable source of absorbing aerosols is, however, desert dust.

## 5 Discussion

570 GACA is a threshold-based algorithm for the determination of dominant aerosol types and sources  
globally on a seasonal basis. In this section we investigate the robustness of the algorithm, motivate  
our choice of EAE (as opposed to FMF), and compare results from GACA with previously reported  
climatologies from measurements and models. Although the algorithm can be improved further by  
fine-tuning with regional settings and/or additional (satellite) data, the main objective of the current  
575 study is to explore what can be learned from the combination of different satellite data sets. We  
present some suggestions for future improvements to GACA in Sect. 5.4.

## 5.1 Sensitivity studies

It is clear that GACA results depend on the choice of thresholds and criteria for aerosol type and source determination. Most source assignments are rather robust and altering thresholds only causes  
580 small shifts of borders between different sources. Beyond being rooted in textbook knowledge, our criteria are justified by the consistency of the obtained results and the good general agreement with MACC model results. The basic assumption underlying GACA is that enhancements in trace gas and aerosol abundance are caused by the same source and wherever this is not the case, the algorithm fails. Correctly characterizing mixed air masses (e.g. dust with smoke or pollution) or transported  
585 aerosols (that may be present above or in addition to local pollution) thus is beyond the capabilities of GACA.

To investigate how robust GACA is with respect to effects of clouds, varying time ranges, and the treatment of outliers, we performed a series of tests. First, we applied different cloud filters to the GOME-2 data prior to gridding. Unfortunately, a similar test could not be performed on MO-  
590 PITT data, as we used gridded monthly means that had already been cloud-cleared. MODIS AOD is only retrieved under clear sky conditions, but because the field of view of the instrument is small, retrievals in between cloud patches are often possible in regions that would be considered cloudy by GOME-2. Setting the maximum effective cloud fraction (CF) to 0.05, 0.20, or 0.40 does not cause major changes in global maps of GACA-type and GACA-source (Figs. S2–S3 in the Supplement,  
595 respectively). Perhaps surprisingly, the results are still similar if only data with  $CF > 0.40$  are selected: the main difference is the disappearance of non-absorbing aerosol types due to the increase in data points with  $UVAI < 0$ . We conclude that measurements of  $NO_2$ , HCHO,  $SO_2$ , and UVAI in the presence of clouds contain enough information to be used for characterization of aerosol (or air mass) sources, at least on a monthly mean basis. Measurements of other trace gases, e.g. CO, are  
600 expected to be similarly useful (e.g., Liu et al., 2014).

The effects of varying the time range from the maximum of 15 months per season (5 years  $\times$  3 months) are rather trivial: the scatter increases with decreasing data amount, and so does the influence of one-time events, such as volcanic eruptions. We performed tests for the summer (June–August) and found that GACA-type and GACA-source results are very similar if data from 2007–  
605 2011 or 2008–2010 are used. Decreasing the time window further to July 2007–2011 (5 months) or to June–August 2009 causes noisy results with large data gaps (particularly over South America). For source determination of individual aerosol types (as in Fig. 8) the statistical requirements are even higher. Changing the resolution of GACA-source to  $1^\circ \times 1^\circ$  yields dominant source maps very similar to those in Fig. 6, but with several large data gaps, most notably over South America in  
610 summer.

In the standard GACA set-up, each data set is screened for outliers, which are then removed (see Sect. 3.1 for details). The reason for this procedure is that GACA is aimed at constructing a climatology in which exceptional events (large fires, volcanic eruptions, etc.) should not be rep-

resented. Another reason is the removal of artifacts, which are, however, only rarely encountered  
615 in the monthly averaged, gridded data sets used here – except in the region affected by the South  
Atlantic Anomaly. If GACA is run without removing outliers, the resulting source maps are very  
similar to those from the standard run (compare Fig. 6 with Fig. S4 in the Supplement); in fact, the  
map for winter does not change at all. The biggest change is found for the spring maps, where sev-  
620 eral volcanic sulfate (VOG) plumes appear, most prominently one from the Fernandina volcano on  
the Galapagos islands, which erupted in April 2009. VOG plumes from degassing (Kilauea, Hawaii,  
2008) and erupting (Nabro, Eritrea, 2011) volcanoes are also seen more clearly in the summer map  
when outliers are not removed. The largest change in summer is caused by the exceptional fire season  
that occurred in 2010 in Russia. Because GACA uses AOD weighting, the thick, persistent smoke  
625 plumes strongly influence the algorithm, despite the fact that the fires occurred in only 2 out of 15  
months considered. In South America more grid boxes are assigned to BB, replacing URB; the same  
is seen in fall, although there BB replaces several assignments of BIO if outliers are included in the  
analysis.

## 5.2 Extinction Ångström Exponent and Fine-Mode Fraction

Throughout this study, EAE is used as a measure of aerosol size, instead of the often-used FMF (also  
630 denoted as  $\eta$  in the MODIS literature). The main reason is consistency among the three MODIS  
aerosol algorithms: the Deep Blue algorithm does not output FMF, and although both dark target  
algorithms (land and ocean) provide values of FMF, the definitions are different. The MODIS over-  
ocean retrieval adjusts the abundance of two aerosol types – one fine-mode, one coarse-mode – to  
best fit the measured radiance at six wavelength bands. The two types are chosen from a total of  
635 nine aerosol types (four fine, five coarse), each represented by a single log-normal size distribution.  
The over-ocean FMF is the radiance fraction attributed to the fine-mode aerosol type (Remer et al.,  
2005). Over (dark) land, the FMF represents the weighting of fine-dominated and coarse-dominated  
*models*, which each consist of fine and coarse mode(s). In practice, FMF is essentially binary, rarely  
deviating from either 0 or 1 (Levy et al., 2010).

640 The definition of EAE, on the other hand, is unambiguous (Eq. 1). Throughout the course of the  
MODIS retrieval, AOD is determined at each of the wavelengths used in the retrieval, hence EAE  
can be computed for several wavelength combinations. Here, 470 and 660 nm were chosen, as these  
are the only two wavelengths used in all three MODIS retrievals. Despite the fact that like FMF, EAE  
is affected by a priori assumptions of aerosol optical properties and surface reflectance (over land),  
645 the monthly pattern of EAE corresponds to the global distribution of dust and non-dust (Remer et al.,  
2005) and this is sufficient for the application presented here. For spatially and temporally higher-  
resolved characterization studies, however, a different (or additional) metric may need to be used,  
e.g., size and/or shape from instruments like MISR (Kahn et al., 2005) or POLDER (Polarization  
and Directionality of the Earth's Reflectances, Tanré et al., 2011).

### 650 5.3 Comparison with other climatologies

Different aerosol climatologies of microphysical aerosol properties (or proxies) have been constructed using remotely sensed data in the past. The most established empirical climatologies are derived from AERONET data (Dubovik et al., 2002; Omar et al., 2005; Levy et al., 2007a; Lee et al., 2010). At a first glance, the agreement between GACA-source and AERONET-derived climatologies (e.g., Fig. 2 in Omar et al., 2005, Fig. 3 in Levy et al., 2007a, or Fig. 2 in Lee et al., 2010) is good. However, due to large differences in spatial sampling and the limited information available from AERONET, the informative value of such a comparison is limited. More recently, large-scale collaborations between various modeling groups have shown that a combination (or mean) of aerosol properties from different models perform better (i.e., display smaller differences with measurements) than the output of any single model (e.g., Kinne et al., 2013; Sessions et al., 2015). The resulting climatologies (Fig. 2 in Kinne et al., 2013 and Fig. 3 in Sessions et al., 2015) are in agreement with GACA regarding the dominating aerosol type. But again, the gain from such a comparison is limited, because there is no separation of aerosol types in the presented model climatologies apart from that between fine and coarse modes. It would be more interesting to compare the aerosol composition from the model climatologies with GACA-source, but this is beyond the scope of the current study. Recently published model data of global aerosol composition (Chin et al., 2014) allow a more detailed comparison with GACA-source results. The agreement between our Figs. 10–12 and Chin’s Fig. 6a (where regional annual average AOD composition from 1980–2009 are shown) is good; many of the discrepancies between GACA-source and GOCART (Goddard Chemistry Aerosol Radiation and Transport) model results may be attributed to the differences in geographical selection. There are, however, some important differences, two of which point to inaccuracies in the modeling of secondary organic aerosols. In the regions Southern USA and South America GOCART clearly underestimates the amount of organic matter contributing to aerosols. This is particularly evident in South America, where both GACA-source and MACC ascribe the major part of AOD to secondary organic aerosols throughout the year, whereas in GOCART sulfate aerosols contribute almost 50 % to the yearly mean AOD. Additionally, the amount of desert dust appears to be high compared to GACA. The general underestimation of secondary organic and biomass burning aerosols, as well as the overestimation of desert dust by the GOCART model is known (Chin et al., 2014) and might be remedied with the help of an algorithm like GACA.

### 680 5.4 Applications and improvements

The presented algorithm is an attempt at determining dominating aerosol types and sources on a global scale and mainly intends to show the potential of combined trace gas and aerosol data sets. The most important application of an algorithm like GACA is the improvement of model emissions of aerosols and trace gases, as suggested in the study by Xu et al. (2013). Not only models that

685 rely on data assimilation (like MACC, now succeeded by CAMS) may benefit from comparisons  
with GACA. The possibilities of selecting certain aerosol types (e.g., small non-absorbing aerosols)  
or sources (e.g., urban/industrial) for more detailed investigations of the relationships between AOD  
and trace gases are useful tools for the assessment of model performance regarding aerosols and may  
assist in finding strategies to improve aerosol parameterization. In addition, GACA is rather robust  
690 despite the flexibility with respect to temporal and spatial resolution and input data.

There is a multitude of possible adaptations for an algorithm like GACA, but here we focus on three.

1. Adaptation of GACA to shorter time periods and smaller spatial scales. The algorithm as  
such can be easily applied to daily Level-2 data (on a single-pixel scale), with the caveat  
that co-location of the measurements then becomes more important. This could be achieved  
695 using data from a single instrument (e.g. GOME-2 or OMI), from different instruments on the  
same platform (GOME-2 and Infrared Atmospheric Sounding Interferometer (IASI); OMI and  
Tropospheric Emission Spectrometer, TES), or from instruments closely following each other,  
as in the A-Train. Such an approach could be directly applied to atmospheric composition  
modeling through global data assimilation, e.g. in CAMS. Using the combined information  
700 from different satellite observations, the aerosol type could be updated in addition to the total  
AOD, yielding a more realistic mix of aerosol composition.

2. Application of GACA to cloudy data, i.e., aerosol and trace gas measurements of pixels with  
high cloud cover. As shown above, trace gas measurements of cloudy pixels contain enough  
information to be used for aerosol characterization. These would have to be combined with  
705 aerosol retrievals over clouds, e.g. from MODIS or OMI (Torres et al., 2012; Jethva et al.,  
2013, 2014).

3. Modification of GACA to ground-based data. For example, multi-axis-DOAS (MAX-DOAS)  
measurements of trace gases could be combined with aerosol data from a sun-photometer  
(e.g., Aerosol Robotic Network, AERONET) to assess local aerosol sources.

710 Possible future improvements include: (a) the use of more aerosol data, e.g. particle shape and  
aerosol layer height (e.g. from POLDER or MISR) or more trace gas data from GOME-2 (gly-  
oxal) or other instruments. (b) Make use of spatial and/or temporal patterns and correlations, e.g. by  
taking into account the results from neighboring grid boxes or by pattern recognition. (c) Replacing  
the fixed thresholds with a threshold climatology that depends on location and season.

## 715 **6 Conclusions**

Aerosols and trace gases are frequently co-located, and often even correlated, because they are (1)  
emitted by the same sources, e.g. in the case of biomass burning smoke; (2) formed from the same  
precursor, e.g. volatile organic compounds and secondary organic aerosols; or (3) formed from those

trace gases in the atmosphere, e.g. sulfate aerosols from  $\text{SO}_2$ . We exploit this fact for the assessment  
720 of the dominant aerosol source from satellite observations. In this paper, we introduce a strategy for  
the systematic classification of aerosols using the combination of aerosol optical depth and extinc-  
tion Ångström exponent from MODIS with UV Aerosol Index and trace gas columns ( $\text{NO}_2$ , HCHO,  
and  $\text{SO}_2$ ) from GOME-2, and CO columns from MOPITT. Our Global Aerosol Classification Al-  
725 gorithm, GACA, is separated into two main steps: first, an aerosol type is determined based on its  
optical properties; subsequently, trace gas information is added to appoint a dominant aerosol source.  
The obtained global yearly and seasonal maps are generally in good agreement with MACC model  
data, indicating that both are legitimate. However, systematic differences are also found: more desert  
dust and less secondary organic aerosols are indicated by MACC than by GACA. This demonstrates  
the potential of our method – combining aerosol and trace gas data – to evaluate and investigate  
730 aerosol treatment (parameterization, sources, transport, aging and removal processes) in air quality  
and climate models. One possible application of an algorithm like GACA is the updating of both  
aerosol and trace gas emissions, e.g. in CAMS (successor of MACC) or in GEOS-Chem, as sug-  
gested in the study by Xu et al. (2013). Since the mix of aerosol types is currently preserved in  
models, a combined data assimilation of aerosol and trace gas observations would lead to an overall  
735 more realistic representation of aerosols by models.

We find that the rather simple, threshold-based GACA suffices for very plausible results that are  
quite robust with respect to outliers, choice of time range and cloud fraction thresholds. We em-  
phasize, however, that the presented study is exploratory in nature. We provide several suggestions  
for improvement of the algorithm. With the coming new generation of space-based DOAS instru-  
740 ments with high spatial resolution, in particular TROPOMI (Tropospheric Monitoring Instrument  
on the polar-orbiting Sentinel 5p platform, Veefkind et al., 2012) and the geo-stationary Sentinel  
4 (Ingmann et al., 2012), more (cloud-free) data will be available. With such instruments, global  
aerosol type maps with even higher spatial and temporal resolution become feasible. These maps  
may find a wide range of applications: from modelers, who can use the information to verify emis-  
745 sions and aerosol processes, to scientists working to update aerosol climatologies used in the retrieval  
of aerosol optical depth (e.g., MODIS) or trace gas columns, and environmental policy makers, for  
the development of effective mitigation strategies.

*Acknowledgements.* We thank K. Mies (MPIC) for her assistance with  $\text{SO}_2$  retrievals and I. De Smedt (BIRA)  
for help with HCHO data. The O3M SAF is acknowledged for funding visiting scientist project O3-VS10-01  
750 (M. Penning de Vries) and for GOME-2 AAI data. We acknowledge the free use of tropospheric  $\text{NO}_2$  and  
HCHO column data from GOME-2 from [www.temis.nl](http://www.temis.nl). EUMETSAT is thanked for GOME-2 Level-1 data;  
NASA for MODIS AOD and MOPITT CO data, and ECMWF for MACC reanalysis data. J. W. Kaiser was  
supported by the FP7 EU project MACC-II (grant agreement no. 283576).

755 The article processing charges for this open-access publication were covered by the Max Planck Society.

## References

- Andreae, M. O., and Merlet, P.: Emission of trace gases and aerosols from biomass burning, *Global Biogeochem. Cy.*, 15, 955–966, doi:10.1029/2000GB001382, 2001.
- 760 Beirle, S., Hörmann, C., Penning de Vries, M., Dörner, S., Kern, C., and Wagner, T.: Estimating the volcanic emission rate and atmospheric lifetime of SO<sub>2</sub> from space: a case study for Kilauea volcano, Hawai'i, *Atmos. Chem. Phys.*, 14, 8309–8322, doi:10.5194/acp-14-8309-2014, 2014.
- Benedetti, A., Morcrette, J.-J., Boucher, O., Dethof, A., Engelen, R. J., Fisher, M., Flentje, H., Huneeus, N., Jones, L., Kaiser, J. W., Kinne, S., Mangold, A., Razinger, M., Simmons, A. J., and Suttie, M.: Aerosol analysis and forecast in the European Centre for Medium-Range Weather Forecasts Integrated Forecast System, 2. Data assimilation, *J. Geophys. Res.*, 114, D13205, doi:10.1029/2008JD011115, 2009.
- 765 Boersma, K. F., Eskes, H. J., and Brinksma, E. J.: Error analysis for tropospheric NO<sub>2</sub> retrieval from space, *J. Geophys. Res.*, 109, D04311, doi:10.1029/2003JD003962, 2004.
- Callies, J., Corpaccioli, E., Eisinger, M., Hahne, A., and Lefebvre, A.: GOME-2 – METOP's second-generation sensor for operational ozone monitoring, *ESA Bull.*, 102, 28–36, 2000.
- 770 Chin, M., Diehl, T., Tan, Q., Prospero, J. M., Kahn, R. A., Remer, L. A., Yu, H., Sayer, A. M., Bian, H., Geogdzhayev, I. V., Holben, B. N., Howell, S. G., Huebert, B. J., Hsu, N. C., Kim, D., Kucsera, T. L., Levy, R. C., Mishchenko, M. I., Pan, X., Quinn, P. K., Schuster, G. L., Streets, D. G., Strode, S. A., Torres, O., and Zhao, X.-P.: Multi-decadal aerosol variations from 1980 to 2009: a perspective from observations and a global model, *Atmos. Chem. Phys.*, 14, 3657–3690, doi:10.5194/acp-14-3657-2014, 2014.
- 775 Colarco, P. R., Toon, O. B., Reid, J. S., Livingston, J. M., Russell, P. B., Redemann, J., Schmid, B., Maring, H. B., Savoie, D., Welton, E. J., Campbell, J. R., Holben, B. N., and Levy, R.: Saharan dust transport to the Caribbean during PRIDE, 2. Transport, vertical profiles, and deposition in simulations of in situ and remote sensing observations, *J. Geophys. Res.*, 108, 8590, doi:10.1029/2002JD002659, 2003.
- 780 Deeter, M. N., Emmons, L. K., Francis, G. L., Edwards, D. P., Gille, J. C., Warner, J. X., Khattatov, B., Ziskin, D., Lamarque, J.-F., Ho, S.-P., Yudin, V., Attié, J.-L., Packman, D., Chen, J., Mao, D., and Drummond, J. R.: Operational carbon monoxide retrieval algorithm and selected results for the MOPITT instrument, *J. Geophys. Res.*, 108, 4399, doi:10.1029/2002JD003186, 2003.
- Deeter, M. N., Martínez-Alonso, S., Edwards, D. P., Emmons, L. K., Gille, J. C., Worden, H. M., Pittman, J. V., 785 Daube, B. C., and Wofsy, S. C.: Validation of MOPITT Version 5 thermal-infrared, near-infrared, and multispectral carbon monoxide profile retrievals for 2000–2011, *J. Geophys. Res. Atmos.*, 118, 6710–6725, doi:10.1002/jgrd.50272, 2013.
- de Graaf, M., Stammes, P., Torres, O., and Koelemeijer, R. B. A.: Absorbing Aerosol Index – Sensitivity analysis, application to GOME and comparison with TOMS, *J. Geophys. Res.*, 110, D01202, 790 doi:10.1029/2004JD005178, 2005.
- de Graaf, M., Tuinder, O., Tilstra, G., and Penning de Vries, M.: Algorithm Theoretical Basis Document: ATBD for the GOME-2 Aerosol Products, O3MSAF/KNMI/ATBD/002, KNMI, De Bilt, Netherlands, 1–35, 2014.
- De Smedt, I., Müller, J.-F., Stavrou, T., van der A, R., Eskes, H., and Van Roozendaal, M.: Twelve years of global observations of formaldehyde in the troposphere using GOME and SCIAMACHY sensors, *Atmos. Chem. Phys.*, 8, 4947–4963, doi:10.5194/acp-8-4947-2008, 2008.
- 795



- De Smedt, I., Van Roozendaal, M., Stavroukou, T., Müller, J.-F., Lerot, C., Theys, N., Valks, P., Hao, N., and van der A, R.: Improved retrieval of global tropospheric formaldehyde columns from GOME-2/MetOp-A addressing noise reduction and instrumental degradation issues, *Atmos. Meas. Tech.*, 5, 2933–2949, doi:10.5194/amt-5-2933-2012, 2012.
- 800 Dubovik, O., Holben, B., Eck, T. F., Smirnov, A., Kaufman, Y. J., King, M. D., Tanré, D., and Slutsker, I.: Variability of absorption and optical properties of key aerosol types observed in worldwide locations, *J. Atmos. Sci.*, 59, 590–608, doi:10.1175/1520-0469(2002)059<0590:VOAAOP>2.0.CO;2, 2002.
- Eck, T. F., Holben, B. N., Reid, J. S., Mukelabai, M. M., Piketh, S. J., Torres, O., Jethva, H. T., Hyer, E. J., Ward, D. E., Dubovik, O., Sinyuk, A., Schafer, J. S., Giles, D. M., Sorokin, M., Smirnov, A., and Slutsker, I.:  
805 A seasonal trend of single scattering albedo in southern African biomass-burning particles: implications for satellite products and estimates of emissions for the world’s largest biomass-burning source, *J. Geophys. Res. Atmos.*, 118, 6414–6432, doi:10.1002/jgrd.50500, 2013.
- Goldstein, A. H., Koven, C. D., Heald, C. L., and Fung, I. Y.: Biogenic carbon and anthropogenic pollutants combine to form a cooling haze over the southeastern United States, *PNAS*, 106, 8835–8840,  
810 doi:10.1073/pnas.0904128106, 2009.
- Herman, J. R., Bhartia, P. K., Torres, O., Hsu, C., Seftor, C., and Celarier, E.: Global distribution of UV-absorbing aerosols from Nimbus 7/TOMS data, *J. Geophys. Res.*, 102, 16911–16922, 1997.
- Higurashi, A. and Nakajima, T.: Detection of aerosol types over the East China Sea near Japan from four-channel satellite data, *Geophys. Res. Lett.*, 29, 1836, doi:10.1029/2002GL015357, 2002.
- 815 Hollingsworth, A., Engelen, R. J., Benedetti, A., Dethof, A., Flemming, J., Kaiser, J. W., Morcrette, J.-J., Simmons, A. J., Textor, C., Boucher, O., Chevallier, F., Rayner, P., Elbern, H., Eskes, H., Granier, C., Peuch, V.-H., Rouil, L., and Schultz, M. G.: Toward a monitoring and forecasting system for atmospheric composition: The GEMS Project, *B. Am. Meteorol. Soc.*, 89, 1147–1164, doi:10.1175/2008BAMS2355.1, 2008.
- Hörmann, C., Sihler, H., Bobrowski, N., Beirle, S., Penning de Vries, M., Platt, U., and Wagner, T.: Systematic  
820 investigation of bromine monoxide in volcanic plumes from space by using the GOME-2 instrument, *Atmos. Chem. Phys.*, 13, 4749–4781, doi:10.5194/acp-13-4749-2013, 2013.
- Hsu, N. C., Herman, J. R., Gleason, J. F., Torres, O., and Seftor, C. J.: Satellite detection of smoke aerosols over a snow/ice surface by TOMS, *Geophys. Res. Lett.*, 26, 1165–1168, 1999.
- Hsu, N. C., Tsay, S.-C., King, M. D., and Herman, J. R.: Aerosol properties over bright-reflecting source regions,  
825 *IEEE T. Geosci. Remote*, 42, 557–569, doi:10.1109/TGRS.2004.824067, 2004.
- Ingmann, P., Veihelmann, B., Langen, J., Lamarre, D., Stark, H., and Courrèges-Lacoste, G. B.: Requirements for the GMES Atmosphere Service and ESA’s implementation concept: Sentinels-4/-5 and -5p, *Remote Sens. Environ.*, 120, 58–69, doi:10.1016/j.rse.2012.01.023, 2012.
- Inness, A., Baier, F., Benedetti, A., Bouarar, I., Chabrilat, S., Clark, H., Clerbaux, C., Coheur, P., Engelen, R. J.,  
830 Errera, Q., Flemming, J., George, M., Granier, C., Hadji-Lazaro, J., Huijnen, V., Hurtmans, D., Jones, L., Kaiser, J. W., Kapsomenakis, J., Lefever, K., Leitão, J., Razinger, M., Richter, A., Schultz, M. G., Simmons, A. J., Suttie, M., Stein, O., Thépaut, J.-N., Thouret, V., Vrekoussis, M., Zerefos, C., and the MACC team: The MACC reanalysis: an 8 yr data set of atmospheric composition, *Atmos. Chem. Phys.*, 13, 4073–4109, doi:10.5194/acp-13-4073-2013, 2013.

- 835 IPCC: Climate Change 2013: The Physical Science Basis, Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P. M., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp., 2013.
- Jeong, M.-J. and Li, Z.: Quality, compatibility, and synergy analyses of global aerosol products derived from  
840 the advanced very high resolution radiometer and Total Ozone Mapping Spectrometer, *J. Geophys. Res.*, 110(D10), D10S08, doi:10.1029/2004JD004647, 2005.
- Jethva, H., Torres, O., Remer, L. A., and Bhartia, P. K.: A color ratio method for simultaneous retrieval of aerosol and cloud optical thickness of above-cloud absorbing aerosols from passive sensors: application to MODIS measurements, *IEEE T. Geosci. Remote*, 51, 3862–3870, doi:10.1109/TGRS.2012.2230008, 2013.
- 845 Jethva, H., Torres, O., Waquet, F., Chand, D., and Hu, Y.: How do A-train sensors intercompare in the retrieval of above-cloud aerosol optical depth? A case study-based assessment, *Geophys. Res. Lett.*, 41, 2013GL058405, doi:10.1002/2013GL058405, 2014.
- Justice, C. and Townshend, J.: Special issue on the moderate resolution imaging spectroradiometer (MODIS): a new generation of land surface monitoring, *Remote Sens. Environ.*, 83, 1–2, doi:10.1016/S0034-  
850 4257(02)00083-4, 2002.
- 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.
- 855 Kaiser, J. W., Heil, A., Andreae, M. O., Benedetti, A., Chubarova, N., Jones, L., Morcrette, J.-J., Razinger, M., Schultz, M. G., Suttie, M., and van der Werf, G. R.: Biomass burning emissions estimated with a global fire assimilation system based on observed fire radiative power, *Biogeosciences*, 9, 527–554, doi:10.5194/bg-9-527-2012, 2012.
- Kaufman, Y. J., Tanré, D., and Boucher, O.: A satellite view of aerosols in the climate system, *Nature*, 419,  
860 215–223, doi:10.1038/nature01091, 2002.
- Kim, J., Lee, J., Lee, H. C., Higurashi, A., Takemura, T. and Song, C. H.: Consistency of the aerosol type classification from satellite remote sensing during the Atmospheric Brown Cloud–East Asia Regional Experiment campaign, *J. Geophys. Res.*, 112(D22), D22S33, doi:10.1029/2006JD008201, 2007.
- Kinne, S., O’Donnell, D., Stier, P., Kloster, S., Zhang, K., Schmidt, H., Rast, S., Giorgetta, M., Eck, T. F., and  
865 Stevens, B.: MAC-v1: a new global aerosol climatology for climate studies, *J. Adv. Model. Earth Syst.*, 5, 704–740, doi:10.1002/jame.20035, 2013.
- Kokhanovsky, A. A. and de Leeuw, G. (eds.): *Aerosol Remote Sensing over Land*, Springer, Berlin, 2009.
- Lee, J., Kim, J., Song, C. H., Kim, S. B., Chun, Y., Sohn, B. J. and Holben, B. N.: Characteristics of aerosol types from AERONET sunphotometer measurements, *Atmospheric Environment*, 44(26), 3110–3117,  
870 doi:10.1016/j.atmosenv.2010.05.035, 2010.
- Levelt, P. F., van den Oord, G. H. J., Dobber, M. R., Mälkki, A., Visser, H., de Vries, J., Stammes, P., Lundell, J., and Saari, H.: The ozone monitoring instrument, *IEEE T. Geosci. Remote*, 44, 1093–1101, doi:10.1109/TGRS.2006.872333, 2006.

- 875 Levy, R. C., Remer, L. A., and Dubovik, O.: Global aerosol optical properties and application to Moderate Resolution Imaging Spectroradiometer aerosol retrieval over land, *J. Geophys. Res.*, 112, D13210, doi:10.1029/2006JD007815, 2007a.
- 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, 2007b.
- 880 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.
- 885 Liu, C., Beirle, S., Butler, T., Hoor, P., Frankenberg, C., Jöckel, P., Penning de Vries, M., Platt, U., Pozzer, A., Lawrence, M. G., Lelieveld, J., Tost, H., and Wagner, T.: Profile information on CO from SCIAMACHY observations using cloud slicing and comparison with model simulations, *Atmos. Chem. Phys.*, 14, 1717–1732, doi:10.5194/acp-14-1717-2014, 2014.
- 890 Lyapustin, A., Wang, Y., Xiong, X., Meister, G., Platnick, S., Levy, R., Franz, B., Korkin, S., Hilker, T., Tucker, J., Hall, F., Sellers, P., Wu, A., and Angal, A.: Science impact of MODIS C5 calibration degradation and C6+ improvements, *Atmos. Meas. Tech. Discuss.*, 7, 7281–7319, doi:10.5194/amtd-7-7281-2014, 2014.
- Mielonen, T., Arola, A., Komppula, M., Kukkonen, J., Koskinen, J., de Leeuw, G. and Lehtinen, K. E. J.: Comparison of CALIOP level 2 aerosol subtypes to aerosol types derived from AERONET inversion data, *Geophys. Res. Lett.*, 36(18), L18804, doi:10.1029/2009GL039609, 2009.
- 895 Morcrette, J.-J., Boucher, O., Jones, L., Salmond, D., Bechtold, P., Beljaars, A., Benedetti, A., Bonet, A., Kaiser, J. W., Razinger, M., Schulz, M., Serrar, S., Simmons, A. J., Sofiev, M., Suttie, M., Tompkins, A. M., and Untch, A.: Aerosol analysis and forecast in the European Centre for Medium-Range Weather Forecasts Integrated Forecast System: forward modeling, *J. Geophys. Res.*, 114, D06206, doi:10.1029/2008JD011235, 2009.
- 900 Omar, A. H., Won, J.-G., Winker, D. M., Yoon, S.-C., Dubovik, O., and McCormick, M. P.: Development of global aerosol models using cluster analysis of Aerosol Robotic Network (AERONET) measurements, *J. Geophys. Res.*, 110, D10S14, doi:10.1029/2004JD004874, 2005.
- Pan, L., Gille, J. C., Edwards, D. P., Bailey, P. L., and Rodgers, C. D.: Retrieval of tropospheric carbon monoxide for the MOPITT experiment, *J. Geophys. Res.*, 103, 32277–32290, doi:10.1029/98JD01828, 1998.
- 905 Penning de Vries, M. and Wagner, T.: Modelled and measured effects of clouds on UV Aerosol Indices on a local, regional, and global scale, *Atmos. Chem. Phys.*, 11, 12715–12735, doi:10.5194/acp-11-12715-2011, 2011.
- Penning de Vries, M. J. M., Beirle, S., and Wagner, T.: UV Aerosol Indices from SCIAMACHY: introducing the SCattering Index (SCI), *Atmos. Chem. Phys.*, 9, 9555–9567, doi:10.5194/acp-9-9555-2009, 2009.
- 910 Platt, U. and Stutz, J.: *Differential Optical Absorption Spectroscopy: Principles and applications*, Springer, Berlin, Heidelberg, 2008.

- Remer, L. A., Kaufman, Y. J., Tanré, D., Mattoo, S., Chu, D. A., Martins, J. V., Li, R.-R., Ichoku, C., Levy, R. C., Kleidman, R. G., Eck, T. F., Vermote, E., and Holben, B. N.: The MODIS Aerosol Algorithm, Products, and  
915 Validation, *J. Atmos. Sci.*, 62, 947–973, doi:10.1175/JAS3385.1, 2005.
- Richter, A. and Wagner, T.: The use of UV, visible, and near IR solar back scattered radiation to determine trace gases, in: *The remote sensing of tropospheric composition from space*, edited by: Burrows, J. P., Platt, U., and Borell, P., 67–121, Springer, Berlin, Heidelberg, 2011.
- Rosenfeld, D., Andreae, M. O., Asmi, A., Chin, M., de Leeuw, G., Donovan, D. P., Kahn, R., Kinne, S.,  
920 Kivekäs, N., Kulmala, M., Lau, W., Schmidt, K. S., Suni, T., Wagner, T., Wild, M., and Quaas, J.: Global observations of aerosol-cloud-precipitation-climate interactions, *Rev. Geophys.*, 52, 2013RG000441, doi:10.1002/2013RG000441, 2014.
- Schutgens, N. A. J., Nakata, M., and Nakajima, T.: Validation and empirical correction of MODIS AOT and AE over ocean, *Atmos. Meas. Tech.*, 6, 2455–2475, doi:10.5194/amt-6-2455-2013, 2013.
- 925 Seinfeld, J. H., and Pandis, S. N. (eds.): *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*, John Wiley & Sons, New Jersey, 2006.
- Sessions, W. R., Reid, J. S., Benedetti, A., Colarco, P. R., da Silva, A., Lu, S., Sekiyama, T., Tanaka, T. Y., Baladasano, J. M., Basart, S., Brooks, M. E., Eck, T. F., Iredell, M., Hansen, J. A., Jorba, O. C., Juang, H.-M. H., Lynch, P., Morcrette, J.-J., Moorthi, S., Mulcahy, J., Pradhan, Y., Razinger, M., Sampson, C. B., Wang, J.,  
930 and Westphal, D. L.: Development towards a global operational aerosol consensus: basic climatological characteristics of the International Cooperative for Aerosol Prediction Multi-Model Ensemble (ICAP-MME), *Atmos. Chem. Phys.*, 15, 335–362, doi:10.5194/acp-15-335-2015, 2015.
- Sokolik, I. N. and Toon, O. B.: Incorporation of mineralogical composition into models of the radiative properties of mineral aerosol from UV to IR wavelengths, *J. Geophys. Res.*, 104, 9423–9444,  
935 doi:10.1029/1998JD200048, 1999.
- Stavrakou, T., Müller, J.-F., De Smedt, I., Van Roozendaal, M., van der Werf, G. R., Giglio, L., and Guenther, A.: Evaluating the performance of pyrogenic and biogenic emission inventories against one decade of space-based formaldehyde columns, *Atmos. Chem. Phys.*, 9, 1037–1060, doi:10.5194/acp-9-1037-2009, 2009.
- Tanré, D., Bréon, F. M., Deuzé, J. L., Dubovik, O., Ducos, F., François, P., Goloub, P., Herman, M., Lifermann, A., and Waquet, F.: Remote sensing of aerosols by using polarized, directional and spectral measurements within the A-Train: the PARASOL mission, *Atmos. Meas. Tech.*, 4, 1383–1395, doi:10.5194/amt-4-1383-2011, 2011.
- Tilstra, L. G., de Graaf, M., Aben, I., and Stammes, P.: In-flight degradation correction of SCIAMACHY UV reflectances and Absorbing Aerosol Index, *J. Geophys. Res.*, 117, D06209, doi:10.1029/2011JD016957,  
945 2012.
- Torres, O., Bhartia, P. K., Herman, J. R., Ahmad, Z., and Gleason, J.: Derivation of aerosol properties from satellite measurements of backscattered ultraviolet radiation: theoretical basis, *J. Geophys. Res.*, 103, 17099–17110, 1998.
- Torres, O., Jethva, H., and Bhartia, P. K.: Retrieval of aerosol optical depth above clouds from OMI observations: sensitivity analysis and case studies, *J. Atmos. Sci.*, 69, 1037–1053, doi:10.1175/JAS-D-11-0130.1,  
950 2012.

- 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.
- 955 Van der Werf, G. R., Randerson, J. T., Giglio, L., Collatz, G. J., Mu, M., Kasibhatla, P. S., Morton, D. C., DeFries, R. S., Jin, Y., and van Leeuwen, T. T.: Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997–2009), *Atmos. Chem. Phys.*, 10, 11707–11735, doi:10.5194/acp-10-11707-2010, 2010.
- 960 Veeffkind, J. P., Boersma, K. F., Wang, J., Kurosu, T. P., Krotkov, N., Chance, K., and Levelt, P. F.: Global satellite analysis of the relation between aerosols and short-lived trace gases, *Atmos. Chem. Phys.*, 11, 1255–1267, doi:10.5194/acp-11-1255-2011, 2011.
- Veeffkind, J. P., Aben, I., McMullan, K., Förster, H., de Vries, J., Otter, G., Claas, J., Eskes, H. J., de Haan, J. F., Kleipool, Q., van Weele, M., Hasekamp, O., Hoogeveen, R., Landgraf, J., Snel, R., Tol, P., Ingmann, P., Voors, R., Kruizinga, B., Vink, R., Visser, H., and Levelt, P. F.: TROPOMI on the ESA Sentinel-5 Precursor: a GMES mission for global observations of the atmospheric composition for climate, air quality and ozone layer applications, *Remote Sens. Environ.*, 120, 70–83, doi:10.1016/j.rse.2011.09.027, 2012.
- 965 Wang, P., Stammes, P., van der A, R., Pinardi, G., and van Roozendaal, M.: FRESCO+: an improved O<sub>2</sub> A-band cloud retrieval algorithm for tropospheric trace gas retrievals, *Atmos. Chem. Phys.*, 8, 6565–6576, doi:10.5194/acp-8-6565-2008, 2008.
- 970 Xu, X., Wang, J., Henze, D. K., Qu, W., and Kopacz, M.: Constraints on aerosol sources using GEOS-Chem adjoint and MODIS radiances, and evaluation with multisensor (OMI, MISR) data, *J. Geophys. Res. Atmos.*, 118, 6396–6413, doi:10.1002/jgrd.50515, 2013.
- Yuan, T., Remer, L. A., and Yu, H.: Microphysical, macrophysical and radiative signatures of volcanic aerosols in trade wind cumulus observed by the A-Train, *Atmos. Chem. Phys.*, 11, 7119–7132, doi:10.5194/acp-11-7119-2011, 2011.

**Table 1.** Data sets that are used as input to GACA with appropriate literature references and websites.

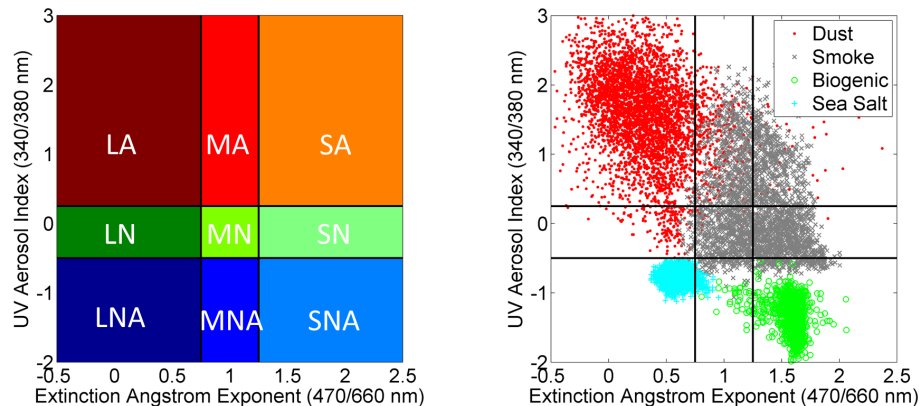
Data set	Instrument	Data version	Literature references	Source
AOD	MODIS	Coll. 5.1	Remer et al. (2005); Levy et al. (2007b); Hsu et al. (2004)	<a href="http://ladsweb.nascom.nasa.gov">http://ladsweb.nascom.nasa.gov</a>
UVAI	GOME-2	V. 4	de Graaf et al. (2005, 2014)	<a href="http://www.temis.nl/airpollution/absaai">www.temis.nl/airpollution/absaai</a>
NO <sub>2</sub> VCD	GOME-2	V. 2.1	Boersma et al. (2004)	<a href="http://www.temis.nl/airpollution/no2.html">www.temis.nl/airpollution/no2.html</a>
HCHO VCD	GOME-2	V. 12	De Smedt et al. (2012)	<a href="http://h2co.aeronomie.be">http://h2co.aeronomie.be</a>
SO <sub>2</sub> VCD	GOME-2		Hörmann et al. (2013)	own
CO VCD	MOPITT	V. 6	Deeter et al. (2003, 2013)	<a href="http://eosweb.larc.nasa.gov/project/mopitt/mopitt_table">http://eosweb.larc.nasa.gov/project/mopitt/mopitt_table</a>

**Table 2.** Abbreviations of aerosol types and sources used throughout this document.

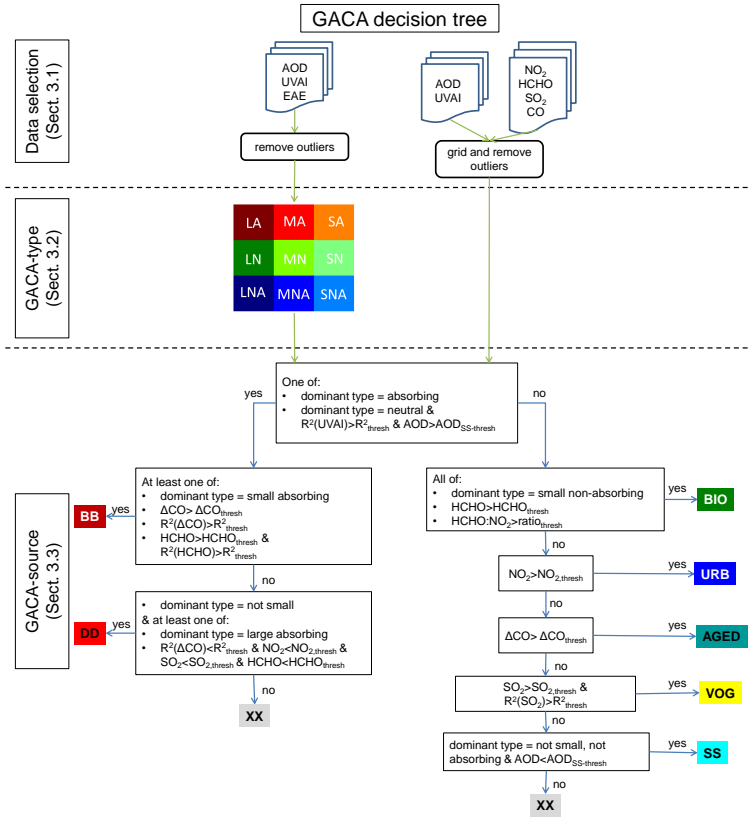
Acronym	Aerosol type/source/component	Occurrence
LA	Large absorbing	GACA-type
LN	Large neutral	GACA-type
LNA	Large non-absorbing	GACA-type
MA	Medium-size absorbing	GACA-type
MN	Medium-size neutral	GACA-type
MNA	Medium-size non-absorbing	GACA-type
SA	Small absorbing	GACA-type
SN	Small neutral	GACA-type
SNA	Small non-absorbing	GACA-type
BB	Biomass burning smoke	GACA-source
DD	Desert dust	GACA-source and MACC
BIO	Secondary aerosols of biogenic origin	GACA-source
URB	Secondary aerosols of urban/industrial origin	GACA-source
AGED	Aged aerosols	GACA-source
VOG	Volcanic sulfate	GACA-source
SS	Sea salt	GACA-source and MACC
XX	Unknown source	GACA-source
BC	Black carbon	MACC
OM	Organic matter	MACC
SO <sub>4</sub>	Sulfate	MACC
MIX	Mixture	MACC
na	not assessed	all

**Table 3.** Thresholds used in GACA. Variables are unitless except for the trace gas (excess) VCDs; these are given in molec cm<sup>-2</sup>.

Variable	Nominal range	Thresholds	GACA step
AOD	0–3	0.05	Filtering
AOD	0–3	0.15	GACA-source (Sea Salt)
EAE	0–2	0.75 and 1.25	GACA-type
UVAI	–2.5–+2.5	–0.5 and 0.25	GACA-type
NO <sub>2</sub> column	0–10 × 10 <sup>15</sup>	1 × 10 <sup>15</sup>	GACA-source
HCHO column	0–25 × 10 <sup>15</sup>	7 × 10 <sup>15</sup>	GACA-source
SO <sub>2</sub> column	0–20 × 10 <sup>15</sup>	1 × 10 <sup>15</sup>	GACA-source
ΔCO excess column	0–40 × 10 <sup>17</sup>	4 × 10 <sup>17</sup>	GACA-source
Ratio HCHO : NO <sub>2</sub>	0–100	4	GACA-source
Correlation coefficient, $R^2$	0–1	0.25	GACA-source

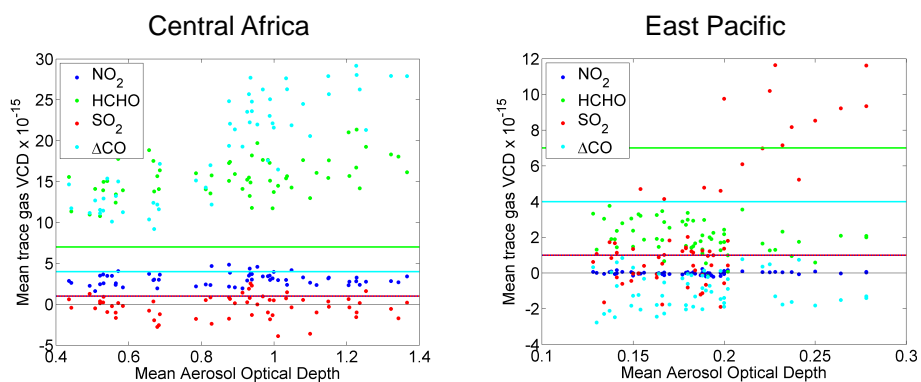


**Figure 1.** Speciation of aerosol types based on absorption (UVAI) and size (EAE). Left, aerosol types color-coded according to size (larger sizes have darker hues) and absorption (non-absorbing in blue, neutral in green, absorbing in red): LA, large absorbing; MA, medium-size absorbing; SA, small absorbing; LN, large, neutral; MN, medium-size, neutral; SN, small, neutral; LNA, large, non-absorbing; MNA, medium-size, non-absorbing; SNA, small, non-absorbing. Right, monthly mean UVAI and EAE within grid boxes in regions dominated by desert dust (red dots), biomass burning smoke (gray crosses), secondary biogenic aerosols (green circles), and sea salt (light blue pluses). Data are from June–August 2007–2011; see the text for the selected geographical regions.

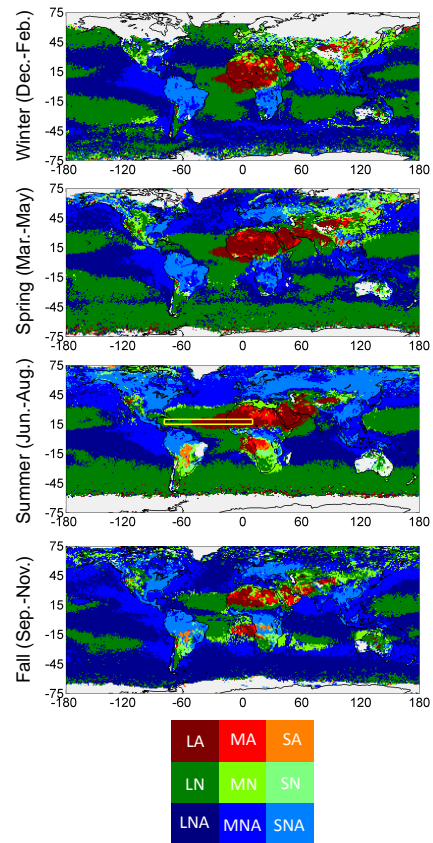


**Figure 2.** Schematic decision tree of GACA. The corresponding threshold values are given in Table 3. The mean value of a quantity, e.g.  $\Delta\text{CO}$ , is denoted “ $\Delta\text{CO}$ ”; the coefficient of correlation between AOD and a quantity, e.g. HCHO, is denoted “ $R^2(\text{HCHO})$ ”. Thresholds are denoted as (e.g.)  $\text{SO}_{2,\text{thresh}}$ ,  $R^2_{\text{thresh}}$ ,  $\text{ratio}_{\text{thresh}}$  (for the HCHO : NO<sub>2</sub> ratio threshold), or  $\text{AOD}_{\text{SS-thresh}}$  (for the maximum AOD allowed for SS classification). Other abbreviations are explained in Table 2.

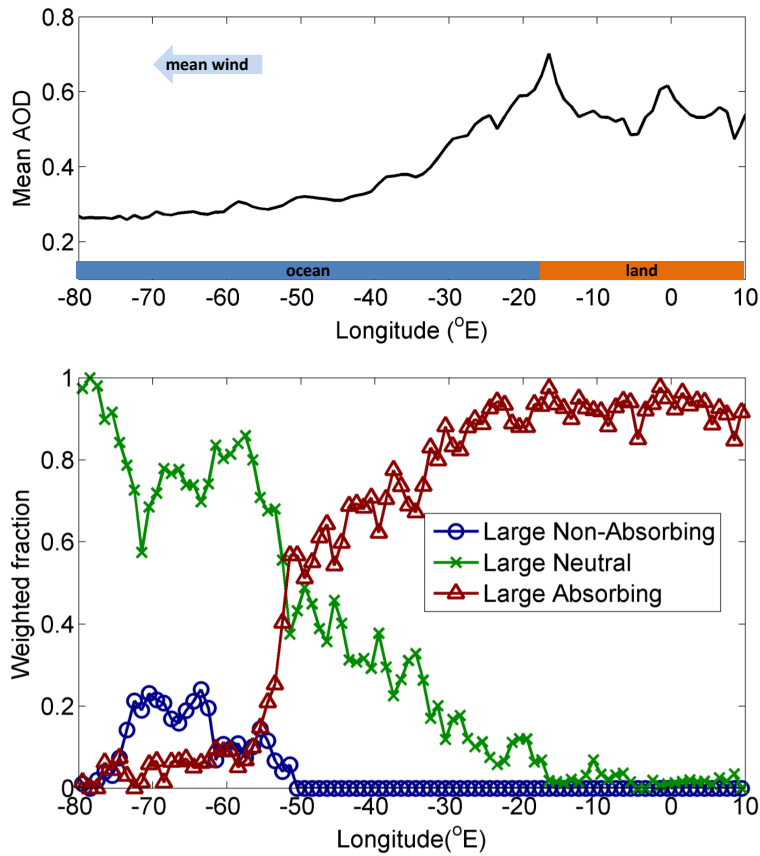




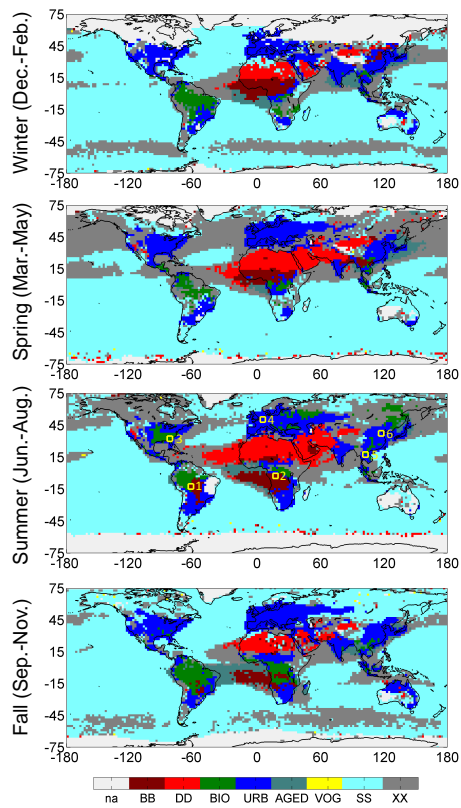
**Figure 3.** Relationship between  $1^\circ \times 1^\circ$  monthly mean values of AOD and trace gas columns (in molec cm<sup>-2</sup>) for a region in central Africa (2–4° S, 18–20° E; left panel) and in the East Pacific Ocean (16–18° N, 162–164° W; right panel) for July–August 2007–2011. Dots depict NO<sub>2</sub> (blue), HCHO (green), and SO<sub>2</sub> (red) VCDs and excess CO VCDs (light blue, scaled by a factor of 0.01) and their respective thresholds. The threshold values of NO<sub>2</sub> and SO<sub>2</sub> are identical (dotted blue and red lines). Note the differences in  $y$  axis scales.



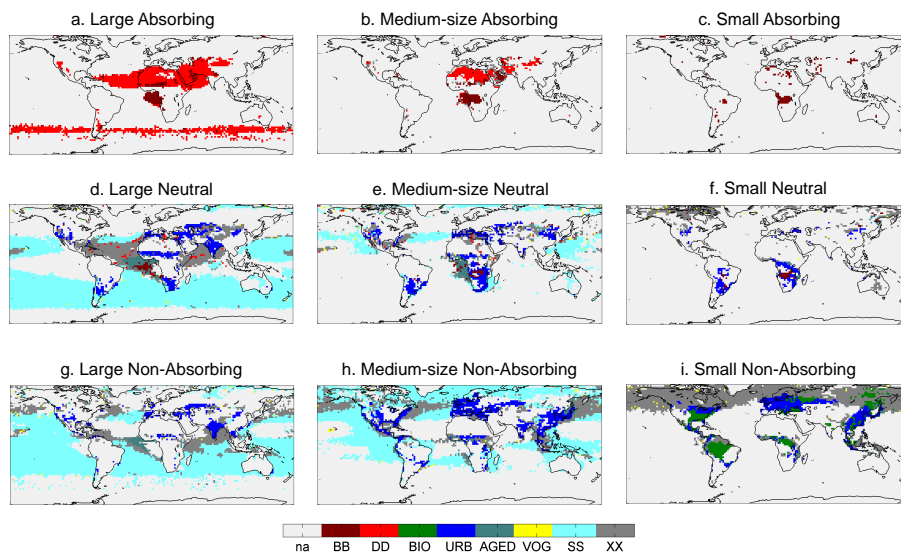
**Figure 4.** Seasonal cycle of global aerosol type distribution according to GACA. Data are from 2007–2011 and were divided into the four main seasons (from top to bottom): winter, spring, summer, and fall. The legend is given on the bottom; see Fig. 1 and Table 2 for aerosol type abbreviations. The yellow box indicates the region investigated in Fig. 5.



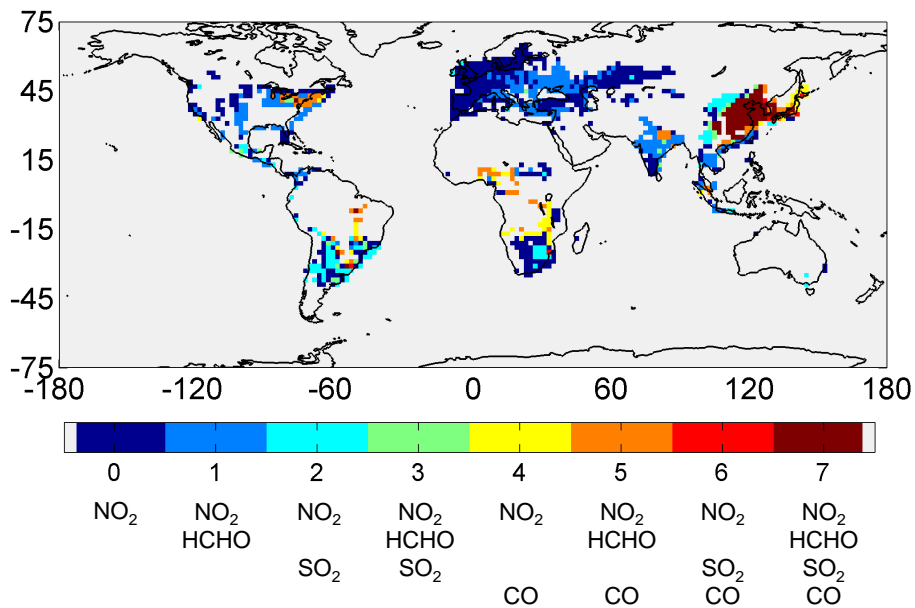
**Figure 5.** Transect showing transport of mineral dust plumes. Shown are summertime (June–August 2007–2011) data from 15–20° N, a region of Saharan dust outflow. Upper panel: mean AOD (total of all aerosol types); the mean wind direction is indicated by an arrow, and the surface type (land or ocean) is given at the bottom of the panel. Lower panel: AOD-weighted fraction of all aerosol types contributing > 20 % to AOD.



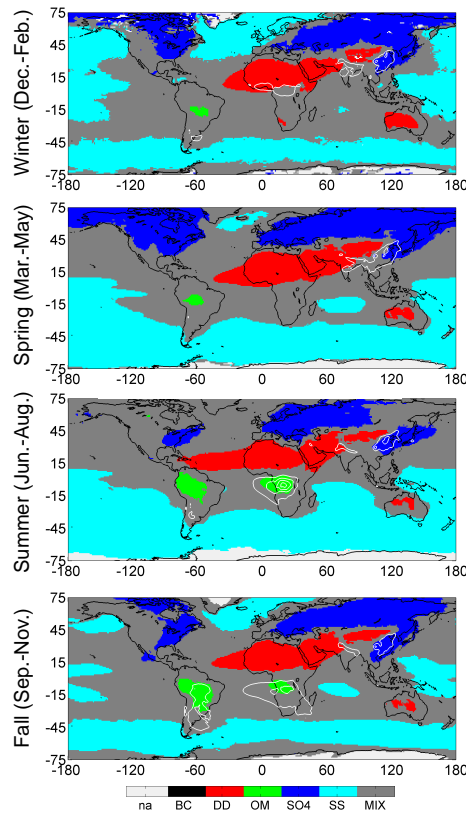
**Figure 6.** Seasonal cycle of global main aerosol source distribution according to GACA. Data are from 2007–2011 and were divided into the four main seasons (from top to bottom): winter, spring, summer, and fall. Aerosol source type abbreviations are given in Table 2; gray areas are not analyzed due to lack of data or too small mean AOD (see text for details). Enumerated yellow boxes in the third panel mark the regions investigated in Figs. 10–12, respectively.



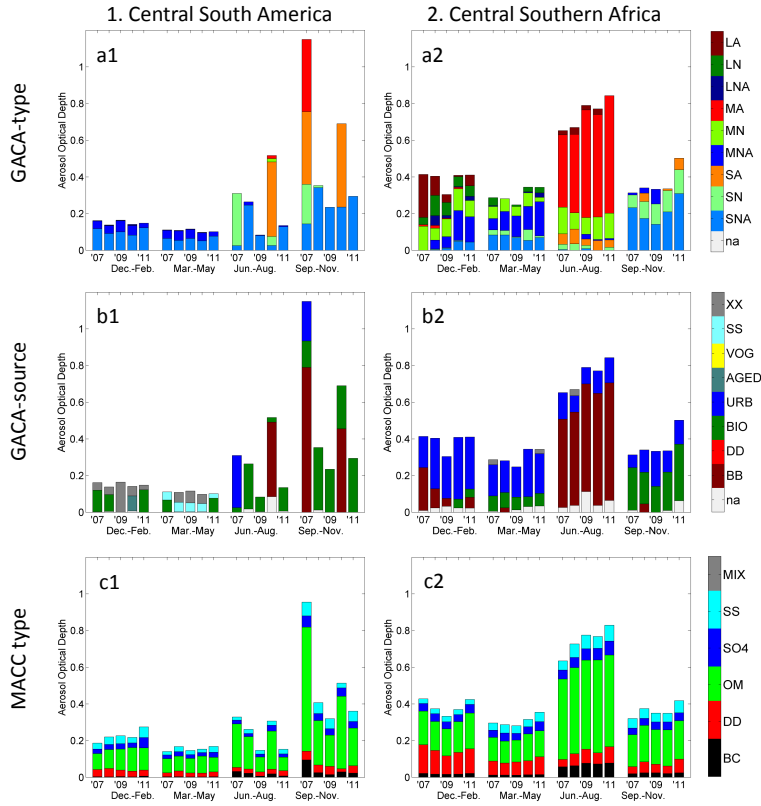
**Figure 7.** Global aerosol source for each aerosol type according to GACA for June–August 2007–2011. Aerosol source and type abbreviations are given in Table 2; gray areas do not contain more than 4 points belonging to the relevant aerosol type.



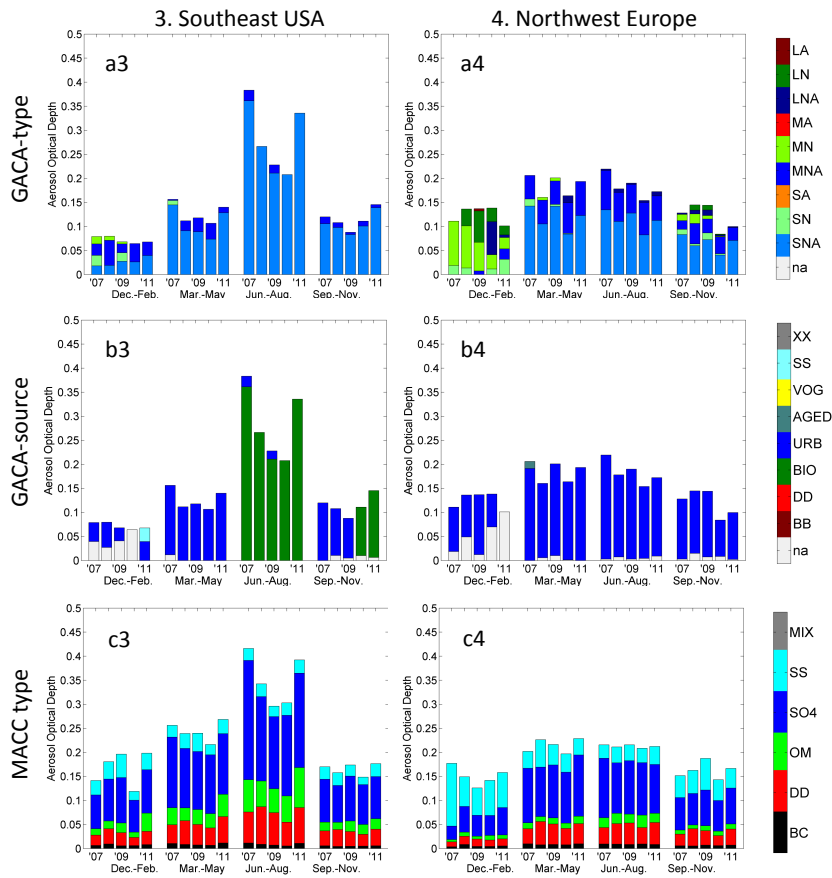
**Figure 8.** Trace gas composition for gridboxes with URB source for June–August 2007–2011. The presence of enhanced trace gas columns (in addition to  $\text{NO}_2$ ) is indicated by 1, 2, or 4 for HCHO,  $\text{SO}_2$ , and  $\Delta\text{CO}$ , respectively: 1 thus indicates enhanced  $\text{NO}_2$  and HCHO, 2 enhanced  $\text{NO}_2$  and  $\text{SO}_2$ , 3 enhanced  $\text{NO}_2$  and HCHO and  $\text{SO}_2$ , etc. Gray areas are not dominated by URB.



**Figure 9.** Seasonal cycle of global main aerosol type distribution according to MACC. Data are from 2007–2011 and were divided into the four main seasons (from top to bottom): winter, spring, summer, and fall. Aerosol types are black carbon (BC), mineral dust (DD), organic matter (OM), sulfate (SO<sub>4</sub>), sea salt (SS), and mixture (MIX). Light gray areas (na) are not analyzed due to too small mean AOD. As BC does not dominate anywhere, contours show mean BC amount (AOD 0.02–0.1) to indicate regions affected by smoke; see text for details.



**Figure 10.** Seasonal cycles of global aerosol type and source according to GACA and MACC for  $5^{\circ} \times 5^{\circ}$  regions in central South America ( $10\text{--}15^{\circ}\text{S}/60\text{--}65^{\circ}\text{W}$ ) and central southern Africa ( $0\text{--}5^{\circ}\text{S}/15\text{--}20^{\circ}\text{E}$ ). Data are grouped into four seasons and separated by year. Panels (a1) and (a2) mean AOD contribution of each aerosol type; (b1) and (b2) mean AOD contribution of aerosol source (determined from each aerosol type); (c1) and (c2) mean AOD contribution of aerosol types from MACC. Abbreviations are explained in Table 2.



**Figure 11.** Seasonal cycles of global aerosol type and source according to GACA and MACC for  $5^\circ \times 5^\circ$  regions in Southeast USA ( $30\text{--}35^\circ \text{N}/80\text{--}85^\circ \text{W}$ ) and Northwest Europe ( $48\text{--}53^\circ \text{N}/3\text{--}8^\circ \text{E}$ ). See Fig. 10 for details.



