Relationships between convective structure and transport of aerosols to the upper troposphere deduced from satellite observations

Sudip Chakraborty,¹ Rong Fu,¹ Jonathon S. Wright,² and Steven T. Massie³

Corresponding author: S. Chakraborty, Jackson School of Geosciences, University of Texas at Austin, JGB 5.336, 2305 Speedway, Austin, TX 78703, USA. (sudipm@utexas.edu)

¹Jackson School of Geosciences, University of Texas at Austin, Austin, Texas, USA.
²Center for Earth System Science, Tsinghua University, Beijing, China.
³Laboratory for Atmospheric and Space Physics (LASP), University of Colorado, Boulder, 3665 Discovery Drive, Boulder, CO

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/2015JD023528

©2015 American Geophysical Union. All Rights Reserved.
Abstract. We estimate the extent of upper tropospheric aerosol layers (UT ALs) surrounding mesoscale convective systems (MCSs) and explore the relationships between UT AL extent and the morphology, location, and developmental stage of collocated MCSs in the tropics. Our analysis is based on satellite data collected over equatorial Africa, South Asia, and the Amazon basin between June 2006 and June 2008. We identify substantial variations in the relationships between convective properties and aerosol transport by region and stage of convective development. The most extensive UT ALs over equatorial Africa are associated with mature MCSs, while the most extensive UT ALs over South Asia and the Amazon are associated with growing MCSs. Convective aerosol transport over the Amazon is weaker than that observed over the other two regions despite similar transport frequencies, likely due to the smaller sizes and shorter mean lifetimes of MCSs over the Amazon. Variations in UT ALs in the vicinity of tropical MCSs are primarily explained by variations in the horizontal sizes of the associated MCSs, and are largely unrelated to aerosol loading in the lower troposphere. We also identify potentially important relationships with the number of convective cores, vertical wind shear, and convective fraction during the growing and mature stages of MCS development. Relationships between convective properties and aerosol transport are relatively weak during the decaying stage of convective development. Our results provide an interpretive framework for devising and evaluating numerical model experiments that examine relationships between convective properties and ALs in the upper troposphere.
1. Introduction

Atmospheric aerosols exert significant influences on the Earth’s radiation balance and surface temperature [Twomey, 1977a, b; Menon et al., 2002]. These influences depend not only on the radiative properties of aerosols, but also on their vertical distribution [Keil and Haywood, 2003; McComiskey and Feingold, 2008; Satheesh and Ramanathan, 2000]. For example, aerosols in the upper troposphere (UT) may increase planetary albedo under clear sky conditions, but reduce planetary albedo when located above clouds. Aerosols in the UT typically have longer lifetime and more persistent and far-reaching radiative effects than aerosols at lower levels [Lacis et al., 1992]. Increases in aerosols in the UT can influence the formation of in situ cirrus clouds [Khain et al., 2008; Froyd et al., 2009], enhance the lifetime of convective anvil clouds [Bister and Kulmala, 2011], and increase water vapor transport into the stratosphere by distributing the water content of ice clouds among a larger number of smaller crystals [Sherwood, 2002; Bister and Kulmala, 2011; Su et al., 2011].

Most previous studies of convective influences on aerosol transport to the UT have focused on the convective transport of insoluble gas-phase aerosol precursors. Convective transport of existing aerosols to the UT has been suggested by cloud resolving models [Ekman et al., 2006; Fan et al., 2009] and measurements made during field campaigns [Andreae et al., 2004; Heald et al., 2011; Heese and Wiegner, 2008; Nakata et al., 2013]; however, observational evidence of large-scale convective aerosol transport was unclear until recent satellite measurements revealed large areas of persistent aerosol layers in the tropical tropopause layer (TTL) over the Asian monsoon region [Vernier et al., 2011]. Moreover,
Solomon et al. [2011] showed that background stratospheric aerosol loading has increased since 2000, reducing the global radiative forcing by about 0.1 W m$^{-2}$. These studies highlight the potential importance of convective aerosol transport on climatic scales and raise a number of important questions. How do aerosols that reach the upper troposphere survive wet scavenging during convective transport? Is convective transport of aerosols limited to certain geographic regions or aerosol types, or does it occur on global tropical and climatological scales? What conditions favor convective transport of aerosols?

Large samples that cover a variety of different climate regimes are needed to develop a general statistical characterization of the factors that influence convective transport of aerosols. Satellite measurements can provide these samples. For example, as shown by Vernier et al. [2011], the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) instrument suite can detect aerosol layers in the UT and TTL (between 10 and 18 km altitude). Observations from CALIPSO have also been combined with observations from the CloudSat cloud radar to infer the properties of cloud and aerosol layers [Kato et al., 2011] and the distributions of stratiform and upper-level clouds [Zhang et al., 2010]. Several studies have used measurements of carbon monoxide (CO) made by the Aura Microwave Limb Sounder (MLS) as a proxy for convective transport of aerosols generated during biomass burning [Jiang et al., 2008, 2009; Di Pierro et al., 2011; Knippertz et al., 2011]. CALIPSO, CloudSat, and Aura MLS are all part of the A-Train satellite constellation [L’Ecreuyer and Jiang, 2010].

The role of wet scavenging on aerosol concentrations in convective environments has been extensively studied using both field observations [Martin and Pruppacher, 1978; Okita et al., 1996; Pratt et al., 2010] and numerical model simulations [Ervens et al.,
By contrast, relatively few studies have focused on the influence of convective dynamic structure on aerosol transport to the UT. The dynamic structure of convection varies substantially on diurnal timescales and cannot be adequately described by polar-orbiting satellites such as those in the A-Train. However, geostationary satellites, such as those used in the International Satellite Cloud Climatology Projects (ISCCP) Cloud Tracking data, provide observations of a number of important parameters that either directly or indirectly describe key aspects of convective dynamic structure throughout the life cycles of convective systems [Machado et al., 1998].

Here, we collocate instantaneous 3-hourly ISCCP Cloud Tracking data with CloudSat, CALIPSO, Aura MLS, and other related A-Train satellite measurements to examine variations in the influence of convective dynamic structure on aerosol transport by both region and stage of convective development.

We focus on MCSs over three tropical regions (Fig. 1): equatorial Africa (10°S–10°N; 10°W–40°E), South Asia (0°–40°N, 70°–100°E), and the Amazon basin (15°S–5°N; 40°W–80°W). This approach allows us to examine upper tropospheric aerosol layers over three important tropical sources of aerosol pollution with substantially different convective regimes [Petersen and Rutledge, 2001]. Aerosols from biomass and fossil fuel burning, dusts, and different types of sulfate, nitrate, and ammonium aerosols are predominant over these regions [Chung and Ramanathan, 2004; Fan et al., 2004; Huang et al., 2013]. Mesoscale convective systems (MCSs) over equatorial Africa are prototypical of tropical continental convective systems, and contain some of the deepest and most intense convection in the world. MCSs over South Asia and the Amazon basin are generally shallower and less intense than MCSs over equatorial Africa, and share many characteristics in com-
mon with convection over tropical oceans. We restrict our analysis of MCSs over South Asia to the peak monsoon rainy season (June–August); these rainy season MCSs typically have longer lifetimes and greater horizontal extents than MCSs over the Amazon basin.

2. Data and Methodology

We use a suite of several satellite data sets to describe the properties of MCSs and detect the existence and extent of aerosol and pollution layers in the nearby atmosphere. Table 1 lists key properties of the data sets and information regarding data access. This section presents an overview of our approach (Section 2.1), the sources, limitations, and uncertainties of the data sets (Section 2.2), our approach to collocating and analyzing the data (2.3), and the methodology used to construct statistical models of the data (2.4).

2.1. Identification and description of clouds and aerosol layers

Although satellites cannot directly measure the dynamic properties of MCSs (such as vertical velocity, mass flux, and vorticity), they are able to measure many physical properties of clouds that are related to these dynamic properties. The International Satellite Cloud Climatology Project (ISCCP) has produced data that tracks a suite of convective morphological properties at three-hourly resolution, including the number of convective cores (NCC), the convective fraction (CF), and the radius associated with observed convective systems [Machado et al., 1998]. A typical MCS contains a stratiform part and a convective core part; the ratio of the latter to the total cloud area is called the convective fraction (CF). NCCs within the convective part act as a pathway for the updraft that contains water vapor, aerosols, and many other gases to the top of the convection. NCCs are identified by their high reflectivities caused by heavy rainfall. Observations of ice
water content (IWC) data from CloudSat are used to infer the height of the detrainment layer (HDL) [Mullendore et al., 2009]. CloudSat also provides an estimate of cloud top height (CTH); however, CloudSat is primarily sensitive to larger hydrometeors and cannot detect the relatively small ice particles in cirrus anvils near the top of tropical convective storms. We therefore estimate CTH using measurements from the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO) satellite [Winker et al., 2009], which, like CloudSat, is part of the A-Train satellite constellation [L’Ecuyer and Jiang, 2010]. The Aura Microwave Limb Sounder (MLS), which is also part of the A-Train, provides measurements of IWC in upper troposphere that complement CALIPSO and CloudSat observations of convective anvil clouds [Wu et al., 2008]. Precipitation rates are a fundamental measure of convective intensity and can influence the wet scavenging of aerosols; we use gridded estimates of precipitation rate from the Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis (TMPA) [Huffman et al., 2007], which are provided at the same temporal resolution (and approximately the same spatial resolution) as ISCCP data. In addition to cloud physical properties, environmental properties such as vertical wind shear (VWS) can play an important role in cloud formation, storm development, and convective aerosol transport [Houze, 2004; Thorpe et al., 1982; Kingsmill and Houze, 1999; Weisman and Rotunno, 2004; Moncrieff, 1978]. We derive VWS using data from the Modern-Era Retrospective-analysis for Research and Applications (MERRA) [Rienecker et al., 2011], provided MERRA successfully detects the observed convective event.

It is difficult to detect aerosol layers in the vicinity of convection. We therefore use three complementary sets of aerosol measurements made by instruments in the A-Train satellite
constellation. Our analysis of aerosol transport is based primarily on CALIPSO observations, validated and supplemented by observations from the Ozone Monitoring Instrument (OMI) onboard the Aura satellite and the Moderate Resolution Imaging Spectroradiometer (MODIS) onboard the Aqua satellite. CALIPSO provides high-vertical resolution profiles of aerosols along a narrow swath, and is sensitive to thin layers of aerosols above clouds and above the boundary layer. By contrast, MODIS provides observations of column-integrated aerosol optical depth (AOD) over a wide swath. These measurements are dominated by aerosol loading in the boundary layer. OMI uses observations of ultraviolet radiation to measure aerosol index (AI) and reflectivity. Because its instrumental wavelength is much smaller than cloud droplets or ice particles, OMI can be used to observe aerosol layers above clouds.

2.2. Data

The core of our analysis is the ISCCP DX data set. These data include cloud properties that have been derived from ISCCP B3 infrared and visible radiances, and are provided every 3 hours between July 1983 and December 2009 at 30 km spatial resolution. Measurement uncertainties are approximately ±2% for infrared wavelengths and ±5% for visible wavelengths. Cloud identification in the ISCCP DX dataset is based on brightness temperature and visible reflectance [Machado et al., 1998]. Once identified, MCSs are tracked by matching 28 different parameters at three-hourly intervals within 5° × 5° domains [see Machado et al., 1998, for details]. Only MCSs with radii greater than 100 km and lifetimes longer than 6 h are considered. We use ISCCP DX data to identify the time of origin and track the evolution of MCSs over three regions in the tropics. We focus on variations in location, radius, NCC, CF, and stage of storm development. MCSs

©2015 American Geophysical Union. All Rights Reserved.
can undergo several cycles of growth and decay as they propagate through different locations, with related changes in convective dynamic structure. The developmental stage of an MCS is calculated from the continuity equation by estimating the areal time rate of expansion ratio

$$\frac{1}{A} \frac{\partial A}{\partial t} = \nabla \cdot V,$$

where $A$ is the area of the convective system and $V$ is the horizontal wind vector \cite{Machado et al., 1998}. Large positive values of AE ($AE > 0.1$) are associated with the growing stage of the convective life cycle, whereas large negative values ($AE < -0.1$) are associated with the decaying stage. Values of AE associated with mature MCSs are close to 0 ($-0.1 \leq AE \leq 0.1$).

The primary CloudSat instrument is a 94 GHz nadir-pointing cloud profiling radar (CPR), which measures radar backscatter from clouds as a function of distance from the instrument \cite{Stephens et al., 2002}. Where possible, we use Level 2 (along-track) Cloud Water Content-Radar Only (2B-CWC-RO) profiles of cloud water content (CWC) to determine the HDL of MCSs observed by ISCCP. Convective detrainment occurs when the convective air mass loses buoyancy. This loss of buoyancy leads to vertical convergence of IWC, which is balanced by horizontal divergence. The convective detrainment layer is therefore linked to a local increase in IWC \cite{Mullendore et al., 2009}. We estimate the rate of change of CloudSat IWC with respect to height ($\partial IWC/\partial z$), and define the base of the detrainment layer as the peak in positive $\partial IWC/\partial z$ and the top of the detrainment layer as the peak in negative $\partial IWC/\partial z$. The HDL is then defined as the center of this layer.

©2015 American Geophysical Union. All Rights Reserved.
CloudSat profiles of CWC are provided in 125 vertical bins with a vertical resolution of 240 m. The horizontal footprint is 1.4 km × 1.7 km (across track × along-track), with samples collected at 1.1 km intervals along the orbit. Liquid and ice water content are retrieved using an algorithm that combines active remote sensing data with a priori data via both forward and backward models. 2B-CWC-RO retrievals may fail in regions of high reflectivity (indicating strong precipitation); however, our focus is on IWC near the top of the cloud, where precipitation rates are generally low. The retrieval algorithm assumes a constant ice crystal density; this assumption allows for the introduction of a correction factor based on the complex refraction index of ice particles. We limit our analysis to IWC data for which the quality flag is set to zero, indicating measurements that are of good quality and suitable for scientific research.

The primary instrument in the CALIPSO suite is the Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) [Winker et al., 2009]. We use Level 2 Version 3-01 Vertical Feature Mask (VFM) data, which includes both cloud and aerosol products at 0.33–5 km horizontal resolution and 30–180 m vertical resolution. The CALIOP transmits laser signals simultaneously at two wavelengths (532 nm and 1064 nm) with a pulse repetition rate of 20.16 Hz. The receiver measures backscatter intensity at 1064 nm and two orthogonally polarized components at 532 nm. Misclassification of aerosols and clouds can occur due to a variety of reasons, including differences in aerosol type (such as dense smoke or dust), proximity of the aerosols to the edge of the cloud, the presence of optically thin clouds, and many more scenarios [Liu et al., 2009]. We limit ambiguity between clouds and aerosols by considering only those measurements with high cloud–aerosol discrimination (CAD) scores (absolute value > 70) [Liu et al., 2004, 2009; Omar et al., 2009]. CALIPSO CAD scores
indicate the confidence in cloud–aerosol discrimination, and are based on five parameters: layer-mean attenuated backscatter at 532 nm, layer-mean attenuated backscatter color ratio, layer-mean volume depolarization ratio, altitude, and latitude [Liu et al., 2009]. Negative values (−100 to 0) are associated with aerosols and positive values (0 to 100) are associated with clouds. CALIPSO Version 3 has been validated by several previous studies [Kacenelenbogen et al., 2011; Koffi et al., 2012; Redemann et al., 2012], and the use of CAD score has been shown to provide a reliable discrimination between clouds and aerosols with a classification error of only 2.01% [Liu et al., 2014]. CALIPSO is unable to detect aerosols when the aerosol backscatter signal is less than the instrument sensitivity of $2 - 4 \times 10^{-4}$ km$^{-1}$ sr$^{-1}$.

We use precipitation rates from Version 7 of the TMPA daily gridded precipitation product [Huffman et al., 2007]. These data are based on observations from several satellites and rain gauge networks, and are provided every three hours at $0.25^\circ \times 0.25^\circ$ spatial resolution between 50$^\circ$S and 50$^\circ$N.

We derive VWS from MERRA data as the difference between the mean horizontal wind speed ($|v|$) in the 925–850 hPa layer and the mean wind speed in the 250–200 hPa layer, divided by the difference in mean geopotential height (z) between the two layers [Petersen et al., 2006]:

$$VWS \equiv \frac{\partial |v|}{\partial z} \approx \frac{\langle |v|_{LT} \rangle - \langle |v|_{UT} \rangle}{\langle z_{LT} \rangle - \langle z_{UT} \rangle}$$

(2)

where $\langle x \rangle$ indicates the layer mean of quantity $x$. We use six-hourly reanalysis winds at $0.67^\circ \times 0.5^\circ$ horizontal resolution for ISCCP observations at 00, 06, 12, and 18 UTC, supplemented by three-hourly forecast winds at $1.25^\circ \times 1.25^\circ$ horizontal resolution for ISCCP.
observations at 03, 09, 15, and 21 UTC. The reanalysis winds include data assimilation; the forecast winds are products of the forecast model alone. Both products are reported on 42 vertical pressure levels. We use MERRA analyses of vertical pressure velocity ($\omega$) and IWC (in-between 500-100 hPa) to verify that MERRA successfully represents the occurrence of the collocated convective event.

OMI measures backscatter radiance between 270 nm and 500 nm, and provides daytime observations of clouds, aerosol layers, and surface UV irradiance. The OMI footprint is $13 \times 24$ km. We use OMI aerosol index (AI) to validate upper tropospheric aerosol layers detected by CALIPSO in the vicinity of MCSs during the daytime. AI is based upon differences in observed and model radiance ratios and detects the presence of UV-absorbing aerosols in regions where ozone absorption is very small [Torres et al., 2012].

MODIS observes AOD in clear sky conditions across a swath approximately 2300 km wide, which allows us to assess the environmental aerosol loading in the vicinity of an MCS. We use AOD at 550 nm from Aqua MODIS (as opposed to Terra MODIS) because the Aqua satellite is part of the A-Train, so that Aqua MODIS observes approximately the same locations as CALIPSO, CloudSat, and MLS at approximately the same time. Like OMI, MODIS AOD is only available for the daytime. We use two MODIS products: the Level 2 MYD04-L2 data and the Level 3 MYD08-D3 daily gridded product. The former allows us to characterize environmental aerosol loading during the convective time frame for daytime events only, while the latter provides a measure of environmental AOD in the region where the convection took place. MYD04-L2 aggregates observed aerosol concentrations and optical properties into $10 \times 10$ km pixels. MYD08-D3 further aggregates these observations into a global $1^\circ \times 1^\circ$ grid on a daily basis. AOD in both datasets
is based on observations using MODIS channels 1 through 7 and 20 (of 36 visible and infrared channels).

Several previous studies have indicated that MODIS AOD observations are reliable over land regions [Levy et al., 2010] and have evaluated the associated uncertainties [Chu et al., 2003, 2002; Remer et al., 2005]. We only consider MCSs with polluted pixels located nearby, where polluted pixels are defined as pixels with AOD greater than 0.3 [Lee et al., 2012; Livingston et al., 2014]. AOD measurements near clouds can be highly uncertain [Tackett and Di Girolamo, 2009; Varnai and Marshak, 2009], both because of signal degradation due to clouds and because AOD varies due to spatial heterogeneity in relative humidity in the vicinity of clouds. To limit these uncertainties, we evaluate the presence or absence of polluted air within a radial distance of 2° from the cloud boundary (where the latter is based on ISCCP data).

We supplement the above data with Version 3.3 MLS retrievals of IWC and CO. This version of the MLS data accounts for contamination of CO by Milky Way core galaxy radiation, flagging contaminated retrievals as bad data [Pumphrey et al., 2009]. We analyze MLS retrievals of CO and IWC between 216 and 83 hPa in the vicinity of the detected MCSs. This pressure range is within the valid range for both CO and IWC. IWC retrievals are filtered to remove profiles for which the status flag of the MLS temperature retrieval is odd. We then apply the ‘2σ − 3σ’ screening process suggested by Livesey et al. [2013] to limit the effects of spectroscopic and calibration uncertainties. This process has the effect of removing IWC retrievals that are less than three standard deviations above the mean in 10° latitude bands. Other potential uncertainties in IWC include the effects of gravity waves originating in the wintertime stratosphere; however, these uncertainties are most

©2015 American Geophysical Union. All Rights Reserved.
pronounced over mid-to-high latitudes during wintertime, whereas our analysis focuses on the tropics. MLS CO retrievals are filtered to remove profiles for which the status flag is even, the convergence flag is greater than 1.4, the quality flag is less than 1.1 (0.2 at 83 hPa), or the retrieved IWC at 147 hPa is greater than or equal to 8 mg m$^{-3}$. We also eliminate CO retrievals for which either the precision or the retrieved value is less than zero.

2.3. Collocation criteria

ISCCP, TRMM, and MERRA data are provided every three hours, while CloudSat, CALIPSO, Aqua MODIS, OMI, and Aura MLS are provided along the orbits of their respective satellites. A-Train satellites orbit the earth sixteen times per day, with equator-crossing times at approximately 01:30 and 13:30 local time. The regions we consider are located in the tropics, so that A-Train observation times approximately correspond to A-Train equator-crossing times.

We initially collocate the ISCCP convective tracking data with CloudSat 2B-CWC-RO profiles based on time (within ±1 h) and location (within the radius of the storm). We start by checking whether the CloudSat orbit passed through an MCS detected by ISCCP. We then check to see whether MODIS observed aerosols within a 2° radial distance from the outer boundary of the MCS, since we intend to examine the factors that control convective aerosol transport from the lower troposphere to the UT within MCSs. During nighttime, we use CALIPSO data set to ensure the presence of aerosols below 4 km altitude within a range of 2° radial distance from the outer boundary of the MCS. If the MCS occurred within a polluted environment, we use CALIPSO to characterize the aerosol profiles surrounding the MCS. We identify possible convective transport of aerosols
by selecting mesoscale convective systems (detected by ISCCP and CloudSat) that occurred in a polluted environment (detected by MODIS and CALIPSO) with aerosol layers in both the lower troposphere and the upper troposphere (detected by CALIPSO). We then examine the evolution of the MCS before and after collocation to determine whether it was observed during a growing, mature, or decaying stage. We use gridded data from TMPA and MERRA to determine the precipitation rate and VWS associated with each MCS. As both MERRA and TMPA are available at 3-hourly intervals, we calculate precipitation rate and VWS by averaging all data within the estimated (circular) area of the cloud indicated by ISCCP (defined by the center and the radius). We use Aura MLS IWC to supplement observations of deep convection and Aura MLS CO to supplement observations of aerosol transport [Jiang et al., 2008].

We validate the upper tropospheric aerosol layers detected by CALIPSO using OMI (see below), and the lower tropospheric aerosol layers using MODIS. This validation is limited to daytime MCSs, because MODIS and OMI are only able to observe aerosols during the day. This limitation restricts the total number of collocated samples to 963 (353 growing MCSs, 400 mature MCSs, and 210 decaying MCSs). OMI is able to distinguish between absorbing and non-absorbing aerosols: positive values of AI indicate the presence of absorbing aerosols (such as dust or smoke), while near-zero or negative values indicate the presence of non-absorbing aerosols or cloud particles. An OMI measurement with AI close to 1 and a reflectivity greater than about 0.15 likely indicates a cloud–aerosol mixture, while a measurement with AI larger than 1 and reflectivity larger than 0.25 likely indicates aerosols above clouds (Dr. Omar Torres, personal communication, 3 October 2015).
79% of collocated OMI and CALIPSO measurements in the vicinity of convective anvils agree regarding the presence (or absence) of aerosols at the cloud top.

Torres et al. [2012] used simultaneous measurements from A-Train satellites (specifically CALIPSO, MODIS, and OMI) to detect the properties of aerosol and cloud layers over the southern Atlantic Ocean. They used large values of AI measured by OMI to identify the presence of aerosol layers above clouds, and CALIPSO measurements at 532 nm to deduce information about the vertical profile of aerosols and clouds. CALIPSO consistently confirmed the presence of enhanced aerosol layers above clouds observed by OMI. They further used MODIS true color images to determine the horizontal extent of the observed clouds. The success of their study and others [Yu et al., 2012] shows that approximately simultaneous measurements made by satellites (such as the A-Train satellites, which follow each other at a maximum interval of ~15 minutes) can be used to describe the collective properties of cloud and aerosol layers. Moreover, measurements from additional satellite instruments provide independent support and validation of the core data sets we use in this study.

Figure 2 shows an example of A-Train satellite measurements collocated with an MCS identified and tracked using ISCCP data. The ISCCP data provide the location of the center and the radius of the system every three hours (Fig. 2a). This MCS was first observed over the southern Amazon (16.7° S, 57.3° W) at 21 UTC on 20 January 2007. The MCS moved west as it developed, where it intersected with the A-Train tracks at approximately 06 UTC on 21 January 2007, when the MCS was in a growing stage (AE = 0.36). The radius of the system grew from 158 km at 03 UTC to 228 km at 06 UTC. According to collocated TMPA data, the mean precipitation rate at this time was 0.69 mm h⁻¹.

©2015 American Geophysical Union. All Rights Reserved.
CALIPSO and CloudSat passed over the eastern side of the MCS as it moved westward at 05:27 UTC, and Aura MLS passed over just west of the center of the system at 05:35 UTC. The convection began to decay after 09 UTC, eventually disappearing after 15 UTC on 21 January 2007. It was only observed by the A-Train satellites once.

Figure 2b shows profiles of IWC and increases in CO relative to background concentrations (calculated as the mean clear-sky CO observed within 1000 km of the system on 21 January 2007) retrieved by Aura MLS during the overpass of this system. Large values of IWC were observed both within the system and immediately to its north, with a maximum value of 43 mg m\(^{-3}\). Large positive CO anomalies above the storm indicate that polluted air was transported to the upper troposphere, particularly near the geographic center of the system. Figure 2c shows observations of cloud and aerosol layers made by CloudSat and CALIPSO as they passed over the system. The CALIPSO measurements show both extensive anvil layers and aerosol layers in the upper troposphere, approximately 10–12 km above the surface on either sides of the MCS in-between 10\(^\circ\)S and 22\(^\circ\)S. The vertical cloud layer at 10\(^\circ\)S is detected with low or no confidence by the CAD algorithm and is termed as false determination. However, the columnar layer of aerosols detected at 9\(^\circ\)S is not a false determination and indicates a real aerosol feature. CloudSat profiled the vertical structure of the central part of the cloud system, which CALIPSO was partially unable to detect due to attenuation in the lidar signal. Large values of CWC (indicative of NCCs) were observed within the system, particularly near its northern edge.

CloudSat and CALIPSO observations collected during overpasses that passed close to the center of an MCS can be used to verify the size and location of the same MCS reported by ISCCP. Cloud sizes based on CALIPSO observations of anvil clouds are slightly
greater than those reported by ISCCP, while cloud sizes based on CloudSat observations are slightly less than those reported by ISCCP. These differences can be explained by differences in the sensitivities of these instruments to small ice particles and optically thin clouds. The central location of MCSs based on CloudSat is within $\pm 1^\circ$ of the center reported by ISCCP. Differences in the central location could be attributable to horizontal propagation of the storm during the time gap between when it is observed by ISCCP and when it is observed by the A-Train satellites.

The occurrence of convective aerosol transport is inferred from changes in the vertical distribution of aerosol pixels relative to expected background values. The background profile of aerosol pixels is derived by averaging all clear-sky CALIPSO data over one month in a $1^\circ \times 1^\circ$ grid box with a vertical resolution of 1 km between the surface and 20 km altitude. Aerosol pixel counts near convection are defined as the number of aerosol pixels along the satellite track within $\pm 2^\circ$ (222 km) of the boundary of the MCS along the satellite track, but not within the system itself. Aerosol pixel counts are thus calculated along the same horizontal length on both sides of every system, regardless of its size. The background profile from the grid cell containing the MCS is subtracted from the number of aerosol pixels observed in the vicinity of the convection at each altitude. The differences are then binned in 2 km vertical increments. For example, the change in the number of aerosol pixels at 9 km altitude is defined as the sum of differences between the observed and background profiles in the 8–9 km and 9–10 km layers. This procedure mitigates uncertainties associated with horizontal advection of aerosols into the vicinity of the MCS. The vertical distributions of convective detrainment height based on CloudSat IWC (Fig. 3) and MERRA vertical velocity (not shown) also support the hypothesis.
that the detected aerosol layers are primarily transported to the UT by the collocated convective systems. The number of background aerosol pixels is very low above 4 km altitude, consistent with previous results based on CALIPSO [Huang et al., 2013] and Department of Energy Atmospheric Radiation Measurement (ARM) observations [Turner et al., 2001]. The detected aerosols are unlikely to be freshly nucleated, as previous studies indicate that only a small number of nucleation mode aerosols reach the convective anvil [Ekman et al., 2006]. Hence, we use the vertically integrated extent of the convective aerosol layers (ALs), defined as the total number of aerosol pixels between 4 and 20 km within a range of 2° radial distance from the outer boundary of the MCS, to explore the influence of convective dynamic properties on aerosol transport.

Due to laser attenuation, CALIPSO can only observe aerosols near the periphery of the cloud systems; it cannot detect aerosols inside clouds with substantial water content (see, e.g., Fig. 2c). Our analysis is therefore based on aerosol profiles detected on the periphery of MCSs, either above or beneath the stratiform portions of the system. Aqua MODIS and OMI are completely unable to detect aerosol layers underneath clouds. These instruments are not suitable for detecting aerosols when the CWC is large.

We use MERRA data to provide a meteorological context for each MCS, including the grid-scale vertical pressure velocity ($\omega$) and the vertical wind shear (VWS) in the area surrounding the convection. The occurrence of convection in MERRA matches well with the occurrence of convection observed by the A-Train satellites. MERRA simulates strong negative values of $\omega$ (indicative of upward motion) and detects IWC in-between 500-100 hPa (indicative of presence of a system extending from mid altitudes to anvil level) in 98% of the collocated cases. We eliminate the other 2% of cases from the analysis.

©2015 American Geophysical Union. All Rights Reserved.
Our study includes a total of 963 MCSs (systems with lifetimes more than 6 hours) over equatorial Africa, South Asia, and the Amazon basin between June 2006 and June 2008. We restrict our analysis of MCSs over South Asia to systems that occurred during the peak monsoon months (June–August 2006, June–August 2007, and June 2008). The 963 identified systems include 353 growing MCSs, 400 mature MCSs, and 210 decaying MCSs. We also briefly examine 111 short-lived convective systems (systems with lifetimes less than 6 hours) over the Amazon basin.

Our collocation methodology (Section 2.3) contains several potential sources of random error, such as measurement uncertainties and differences in the part of the MCS observed by the A-Train satellite instruments (e.g., edge versus center, convective versus stratiform parts of the MCS, etc.). To mitigate the effects of these random errors, we aggregate each sample into 50 bins of approximately equal size according to the observed aerosol layer extent and perform the regression using the mean values in each bin. For example, the full sample of 963 MCSs is separated into 50 bins, each containing between 19 and 20 MCSs. Missing data are masked and excluded from the mean value in each bin.

2.4. Statistical model construction

The statistical models we consider here are based on multiple linear regression, in which the dependent variable (or output) is modeled as a linear combination of the independent variables (or predictors):

\[
y_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij},
\]  

(3)
where $\beta_0$ is the intercept and $\beta_j, j \in [1,p]$ is the coefficient associated with the $j$th of $p$ predictors. In our analysis, the dependent variable $y_i$ is the aerosol layer extent, defined as the aerosol pixel count relative to the local background aerosol pixel count based on CALIPSO observations. This variable can be either negative (indicating dilution of the upper tropospheric background aerosol layer) or positive (indicating aerosol transport). The independent variables $x_{ij}$ include the radius, number of convective cores (NCC), and convective fraction (CF) observed from ISCCP data, the cloud top height (CTH) from CALIPSO, the vertical wind shear (VWS) from MERRA, the precipitation rate from TRMM, and the aerosol optical depth (AOD) from MODIS (i.e., $p = 7$). We standardize all predictors so that they have a mean of zero and a variance of one.

Traditional linear regression selects the coefficients $\hat{\beta}$ that minimize the residual sum of squares

$$
\sum_{i=1}^{N} \left( y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2.
$$

(4)

This approach is unsuited to coefficient determination and variable selection for modeling physical systems like deep convection, for which two or more of the predictors may be tightly correlated. In particular, the coefficients for tightly correlated predictors may be large, with opposite signs that cause compensation in the model. To mitigate this effect, we apply the elastic net method for coefficient shrinkage [Zou and Hastie, 2005]. This method applies a penalty on the size of the coefficients. The coefficients $\hat{\beta}$ are then determined by minimizing the penalized residual sum of squares

©2015 American Geophysical Union. All Rights Reserved.
\[
\sum_{i=1}^{N} \left( y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij} \right)^2 + 0.5\alpha \sum_{j=1}^{p} |\beta_j| + 0.25\alpha \sum_{j=1}^{p} \beta_j^2. \tag{5}
\]

The penalty \(\sum_{j=1}^{p} |\beta_j|\) is often referred to as the \(L_1\) prior (used in Lasso regression) and the penalty \(\sum_{j=1}^{p} \beta_j^2\) is referred to as the \(L_2\) prior (used in ridge regression). The elastic net approach strikes a balance between the Lasso approach (which will typically choose one of the correlated predictors and discard the others) and the ridge regression approach (which will typically shrink the coefficients of correlated predictors toward each other). The parameter \(\alpha\) in Eq. 5 controls the amount of coefficient shrinkage (larger values correspond to greater shrinkage), and is defined to be 0.1 in our models. We have determined this value of \(\alpha\) to be optimal using both the grid search and randomized search methods for parameter estimation. In Grid search method, we try different parameter values on a regular grid and choose the one that provides the best representation of the data, whereas in randomized search method, we randomly sample the parameter space and choose the best value based on how well the model fits the data.

Our goal is to identify the subset of predictors that best explain variability in the dependent variable. We therefore use recursive feature elimination to identify subsets of independent variables that maximize the explained variance in the sample. It eliminates the redundant or irrelevant independent variables that make no useful contribution in predicting the dependent variable. We are then left with seven subsets with sizes of one to seven, where the subset of size one includes the single predictor that maximizes the explained variance, the subset of size two includes the pair of predictors that maximizes the explained variance, and so on. The subset of size seven includes all seven predictors.

©2015 American Geophysical Union. All Rights Reserved.
We apply a random five-fold cross-validation to train and test the models using optimal subsets of various size. Specifically, we randomly divide the sample into five approximately equal bins. We then use 80% of the data to construct a statistical model based on each optimal subset of size one to seven, and then test the model using the remaining 20% of the data. This approach is applied recursively, so that each 20% is used as test data. We repeat this random five-fold cross-validation process 100 times, so that 500 models are trained and tested. We then calculate the coefficients $\hat{\beta}$ as the average coefficients over those 500 models. Repeating this process generally results in identical coefficients to three significant digits.

The optimal model is chosen by identifying the subset of independent variables that minimizes the mean squared error in the test data. Choosing a model that is too simple (i.e., including too few predictors) will underfit the training data, with errors in the test data dominated by large biases. Choosing a model that is too complex (i.e., including too many predictors) will overfit the training data, with errors in the test data dominated by large variance. Using this procedure, we are able to construct statistical learning models that strike a balance between maximizing explained variance (via recursive feature elimination) and minimizing mean squared error (via cross-validation).

3. Results

In principle, aerosols can be transported to the UT by either long-range transport or by local convective transport. The slopes of isentropic surfaces, sedimentation, and vertical motion driven by diabatic heating can all influence the height of an aerosol layer, so that the vertical distribution of aerosol layers is unlikely to remain constant during long-range transport. By contrast, the vertical locations of aerosol layers recently transported to

©2015 American Geophysical Union. All Rights Reserved.
the UT by local convective transport are likely to match the vertical locations of the associated convective detrainment layers. Figure 3 shows the distributions of aerosol layer height (HAL), cloud detrainment height (HDL), and cloud top height (CTH) for all MCSs detected by ISCCP with aerosol layers in the upper troposphere. The vertical distribution of aerosol layers observed in the vicinity of convection is very similar to the vertical distribution of convective detrainment heights. This similarity supports the hypothesis that local convective transport dominates the formation of UT aerosol layers in the vicinity of convective systems, and suggests that joint application of CALIPSO and CloudSat is a reasonable approach to detecting convectively-generated aerosol layers.

The cloud top heights detected by CALIPSO are located several kilometers above the convective detrainment layers, likely due to the sensitivity of the CALIPSO lidar to smaller ice particles and relatively thin cirrus clouds. The sensitivity of the lidar enables us in detecting CTHs more accurately than CloudSat. These clouds may be associated with overshooting convection or local in situ condensation due to convectively-generated gravity waves.

Figure 4 shows variations in IWC and CO in the upper troposphere during different stages of convective development. Both IWC and CO tend to be larger during the growing stage than during the mature or decaying stages. This result indicates that convection is typically deepest and most intense during the growing stage, with stronger transport of near-surface air to the UT.

This hypothesis is further supported by Fig. 5, which shows variations in aerosol transport and convective properties during different stages of the convective life cycle. The vertically-integrated extent of the convective aerosol layer (i.e., the total number of...
aerosol pixels surrounding the MCSs between 4 and 20 km relative to background values) is significantly larger during the growing and mature stages of convection than during the decaying stage. Likewise, the mean height of the aerosol layer is significantly higher during the growing stage (8.8±0.5 km) than during the mature stage (6.8±0.6 km), and significantly higher during the mature stage than during the decaying stage (4.7±0.7 km). These variations in the height and extent of aerosol layers may be attributed to differences in the intensity and extent of convection. The intensity of convection is highest during the growing stage, as indicated by larger mean values of CTH, CF, and rain rate. Mature convective events are nearly as intense (differences in the mean values of CF and rain rate are not significant between growing and mature events), and are also more extensive (as indicated by larger mean values of radius and NCC). The mean values of all variables are significantly lower for decaying MCSs than for growing or mature MCSs.

Figure 6 shows profiles of the magnitude and frequency of convective aerosol transport over the equatorial Africa, South Asia, and Amazon regions. The magnitude of convective aerosol transport is inferred by subtracting a local clear-sky background profile from profiles of aerosol pixels observed within ±222 km of the boundary of the MCS along the satellite track (but not within the boundary of the MCS; see Section 2.3 for details). This metric can be thought of as the extent of the convective aerosol layer in a given altitude range. The magnitude of aerosol transport is bimodal, particularly over equatorial Africa and South Asia, with peaks in the middle troposphere (approximately 5 km altitude) and upper troposphere (approximately 9–11 km altitude). This distribution of aerosol changes is consistent with the vertical profile of convective detrainment, and indicates that convective influences on aerosol concentrations are strongest at the levels where mid-
level congestus and deep convective clouds detrain. The extent and frequency of convective aerosol layers are largest during the mature stage over equatorial Africa and during the growing stage over South Asia. Over the Amazon, the frequency and mean extent of convective aerosol layers are substantially larger during the growing stage than during the mature or decaying stages. The frequency of convective aerosol layers over the Amazon during the growing stage is comparable to the frequencies of convective aerosol layers over equatorial Africa and South Asia, but the extent of these aerosol layers is much smaller (∼50%) over the Amazon.

Differences in convective aerosol transport among the three regions may also be explained by differences in convective properties. For example, mean radius and NCC peak during the mature stage over equatorial Africa and peak during the growing stage over the Amazon (not shown). This difference may explain why aerosol layers are most extensive during the mature stage over equatorial Africa but most extensive during the growing stage over the Amazon. Mean radius and NCC also peak during the mature stage over South Asia; however, differences between the growing and mature stage are much smaller over South Asia than over equatorial Africa. Moreover, unlike over equatorial Africa, these differences are not statistically significant. Growing systems are almost as large as mature systems over South Asia (mean radius of 438.2±55.6 km relative to 499.3±57.7 km) and have nearly as many convective cores (18.6±4.8 relative to 20.5±4.6). When coupled with the greater intensity of growing MCSs relative to mature MCSs and the greater potential for mature MCSs to occur in environments with aerosol concentrations already diluted by wet scavenging, these differences provide a plausible explanation for why convective aerosol transport is strongest during the growing stage over South Asia.
The smaller mean extent of convective aerosol layers over the Amazon relative to convective aerosol layers over equatorial Africa and South Asia has several possible explanations. MCSs over the Amazon have a shorter mean lifetime (\(\sim 19 \text{ h}\)), a much smaller mean radius (\(\sim 120 \text{ km}\)), and higher mean rain rates (\(\sim 1.5 \text{ mm h}^{-1}\)) than MCSs over equatorial Africa or South Asia. The differences in lifetime and radius are likely to reduce the area influenced by the MCSs over the Amazon relative to the larger, longer-lived systems over the other regions, while the difference in rain rate may enhance wet scavenging. Short-lived convective systems are also more common over the Amazon than over the other regions; the frequent precipitation associated with these systems may act more effectively to flush aerosols out of the atmosphere.

Our results indicate that convective aerosol transport is stronger, deeper, and more frequent during the growing and mature stages of the convective life cycle than during the decaying stage over all three analyzed regions. Although this conclusion is unsurprising given the well-known differences in convective intensity between growing systems and decaying systems [Machado et al., 1998], it provides a strong indication that our approach is able to capture variations in both convective structure and aerosol transport. We therefore use these data to construct a series of statistical models. This approach allows us to establish firmer connections between measures of convective intensity and aerosol transport, identify the most influential variables, and evaluate the potential for using these variables to predict the occurrence and extent of aerosol layers generated by convective transport in MCSs.

4. Statistical model results and discussion
Figure 7 shows the fraction of variance of aerosol layer extent in the full data set explained by the predictors in the optimal statistical models for MCSs in different stages of the convective life cycle (see Section 2.4 for details). Table 2 lists the associated coefficients and evaluation metrics for each model based on test data alone. Approximately 80% of the variance in UT aerosol layers (ALs) associated with all MCSs can be explained by three predictors: radius, NCC, and VWS. Likewise, approximately 58% of the variance in ALs associated with growing MCSs can be explained by two predictors (radius and VWS), approximately 75% of the variance in ALs associated with mature MCSs can be explained by three predictors (radius, NCC, and CF) and approximately 52% of the variance in ALs associated with decaying MCSs can be explained by two predictors (radius and AOD).

The most influential predictor in each case is the size (radius) of the MCS.

Figure 8 shows the covariation of the dependent and independent variables in the optimal statistical model for ALs associated with all MCSs. Radius and NCC are both significantly positively correlated with AL extent ($R = 0.92$ and $R = 0.66$, respectively), as well as with each other ($R = 0.92$). As a result, the strong positive coefficient for radius ($\beta_{RAD} = 6.66$) is compensated by a strong negative coefficient for NCC ($\beta_{NCC} = -54.1$) despite the positive single-variable correlation between NCC and AL extent. The negative dependence of AL extent on NCC is at least partially a surrogate for regional differences in the relationship between radius and AL extent. For example, mean AL extent is larger over equatorial Africa (1904 ± 169) than over South Asia (1432 ± 423) (see also intercept values for these two regions in Table 3), but mean radius is slightly larger over South Asia (431 ± 36 km) than over equatorial Africa (364 ± 17 km). By contrast, the mean value of NCC is almost twice as large over South Asia (17.5 ± 2.9) as over equatorial Africa.
NCC therefore becomes a useful surrogate variable for representing the smaller mean extent of UT aerosol layers over South Asia, and its coefficient is negative despite the positive linear relationship between NCC and AL extent.

The role of VWS is more difficult to parse. VWS accounts for 3% of the variance in AL extent for all MCSs and 5% of the variance in AL extent for growing MCSs (Fig. 7). It is tempting to ascribe the negative coefficients for VWS in these models to the adverse impacts of strong VWS on convective development. However, the overall linear correlation between VWS and UT aerosol layer extent is weak, and Fig. 8 and Fig. 9 suggest a more complicated relationship. Local regression indicates a positive relationship between VWS and aerosol layer extent at smaller values of VWS ($R \approx 0.24$ for VWS $\leq 10 \text{s}^{-1}$) and a negative relationship at larger values of VWS ($R \approx -0.38$ for VWS $> 10 \text{s}^{-1}$), although neither relationship is significant at the 95% confidence level ($p = 0.1 \sim 0.2$). Moreover, VWS is most influential over South Asia (Table 3), where it is positively correlated with AL extent ($R = 0.3$, $p = 0.03$).

VWS can be beneficial to the growth of an MCS and the transport of water vapor and aerosols from the boundary layer to the UT. The presence of VWS between low levels and upper levels produces steady storms [Thorpe et al., 1982] and provides a slant-wise path for air to ascend through the system [Houze, 2004]. Wind shear allows updrafts to rise over cold pools formed by downdrafts at an angle between $34^\circ$ and $76^\circ$ [Kingsmill and Houze, 1999], so that large values of VWS may be associated with convective systems that contain distinct regions of updrafts and downdrafts [Weisman and Rotunno, 2004]. Slantwise ascent and separation of updraft and downdraft regions both favor aerosol transport through the convective cores, because they allow the aerosols to bypass scavenging.
in downdrafts. Moreover, cold pool formation and layer overturning are associated with enhanced storm propagation, which is a function of stability and shear of the environment [Moncrieff, 1978]. Storms subject to stronger VWS may therefore be more likely to sample locations in which lower tropospheric aerosols have not yet been scavenged by earlier precipitation.

We are unable to fully clarify the relationship between VWS and AL extent based on these data alone, but our results do permit some insight that could be used to develop model-based studies to further explore this relationship. First, the relationship between VWS and AL extent is strongest during the growing stage. This result is consistent with expectations: growing MCSs are strongly rooted in the boundary layer, while mature MCSs are maintained primarily by latent heating and gravity wave dynamics in the free troposphere [Houze, 2004] and decaying MCSs have weaker updraft speeds and fewer convective cores. Second, the relationship between VWS and UT aerosol layer extent is strongest over South Asia. Mean VWS in convective environments is approximately twice as large over South Asia (15.4 ± 1.1 × 10^{-4} s^{-1}) as over equatorial Africa (7.5 ± 0.5 × 10^{-4} s^{-1}) or the Amazon basin (8.1 ± 1.4 × 10^{-4} s^{-1}), suggesting that the relationship between VWS and UT aerosol layer extent may be more pronounced in optimal-shear environments. Finally, like NCC, the dependence of UT aerosol layer extent on VWS may be confounded by regional differences in VWS. The stronger wind shear over South Asia than over equatorial Africa, combined with the weaker dependence of AL extent on radius over South Asia relative to over equatorial Africa, may partially explain the negative values of $\hat{\beta}_{VWS}$ shown in Table 2. It is also possible that the strong relationship between VWS and AL extent over South Asia represents sub-regional variability with the

©2015 American Geophysical Union. All Rights Reserved.
larger South Asia domain; however, while VWS depends strongly on latitude ($R = -0.57$), most of the MCSs are clustered within the same latitude band ($10^\circ$N–$18^\circ$N) and there is no significant relationship between latitude and AL extent.

Figure 10 shows the covariation of the dependent and independent variables in the optimal statistical model for ALs associated with mature MCSs. The primary difference between the model for mature MCSs and the model for all MCSs is that CF is flagged as important for AL extent associated with mature MCSs, particularly over equatorial Africa (Table 3). The linear relationship between CF and AL extent for mature MCSs over all regions is weak and statistically insignificant ($R = 0.06, p = 0.68$), but the linear relationship between CF and AL extent for mature MCSs over equatorial Africa is significant ($R = 0.33, p = 0.02$). The former is again likely due to regional differences (the largest values of CF are observed over the Amazon, and are associated with smaller systems and less extensive ALs), while the latter likely reflects the actual relationship between CF and AL extent during the mature stage of MCS development. This conclusion is supported by the positive value of $\hat{\beta}_{CF}$ in both cases, which indicates that, given the same radius, larger values of CF are associated with more extensive aerosol layers.

Figure 11 shows the covariation of the dependent and independent variables in the optimal statistical model for ALs associated with decaying MCSs. In addition to the dependence on radius, a small fraction of the variance ($\sim 2\%$) is explained by AOD. Surprisingly, the coefficient for AOD is negative. This may again be partially attributable to regional variability (mean AOD is significantly larger for decaying MCSs over South Asia but mean AL extent is largest for decaying MCSs over equatorial Africa); however, this negative relationship is also identified for AL extent over equatorial Africa alone (Ta-
ble 3). One possibility is that larger values of AOD indicate environmental air that has experienced less wet scavenging, and are therefore evidence of either (1) greater horizontal displacement of the MCS between its growing and mature stages (when ALs were more likely to develop in the UT) and its decaying stage, or (2) lower peak convective intensity, which could potentially reduce both convective transport and washout. Along with the weakness of relationships between AOD and AL extent in other stages of convective development, the negative relationship between AOD and AL extent during the decaying stage underscores the fundamental independence of UT aerosol layer extent relative to aerosol loading in the lower troposphere.

Figure 12 illustrates the expected variations of AL extent according to the regression models summarized in Table 2. Although the models for all MCSs and mature MCSs contain three independent variables, we have used the strong correlation between NCC and radius (NCC = 0.0483 × RAD − 6.561 for all MCSs; NCC = 0.0482 × RAD − 6.871 for mature MCSs) to illustrate them in two dimensions. Figure 12a–b shows that AL extent is expected to increase with increasing radius and decreasing VWS. Likewise, AL extent increases with increasing radius and CF during the mature stage (Fig. 12c), and increases with increasing radius and decreasing AOD during the decaying stage (Fig. 12d). The models explain much of the qualitative variance in the data, but consistently overestimate the lowest values of AL extent and consistently underestimate the highest values of AL extent. These errors may reflect nonlinear relationships that are unaccounted for in the model construction.
5. Summary

The morphology and dynamic structure of MCSs is potentially critical for convective aerosol transport throughout their life cycle, yet relatively few studies have evaluated the relationships between the morphology of MCSs and aerosol transport at the scales of tropical continents. Our analysis uses multiple years of satellite data to evaluate relationships between convective properties and aerosol transport to the UT over three different tropical regions and during three different stages of MCS development.

Many previous studies of aerosol transport have focused on variations in UT CO, which has been used as a surrogate for biomass burning aerosols [Jiang et al., 2008]. Here, we have introduced an approach that combines ISCCP and A-Train satellite observations to explore the relationships between convective characteristics and direct observations of aerosol layers in the nearby UT. A-Train observations allow us to examine the detailed vertical structure of clouds and aerosol layers in the tropical UT, while ISCCP data provides valuable context with respect to both the large-scale environment and the stage of convective development. We focus on three stages of convective development, which have been defined according to changes in the area of the underlying MCS: the growing stage, the mature stage, and the decaying stage. Our approach yields a sample of 963 MCSs covering three tropical regions: equatorial Africa, South Asia, and the Amazon basin. This sample size is considerably larger than previous attempts to infer relationships between convective properties and aerosol transport, and allows us to develop more robust statistical models. We find that the dynamic structure of convective systems and the influence of this structure on aerosol transport vary significantly during the convective life cycle.
Although it is generally expected that the magnitude and frequency of convective aerosol transport are fundamentally tied to the strength of the convection, our analysis represents the first attempt to characterize the evolution of these relationships throughout the MCS life cycle. Aerosol transport is strongest during the growing stage of MCS development over South Asia, but strongest during the mature stage of MCS development over equatorial Africa. These differences are attributable to a combination of (1) larger differences in convective intensity between the growing and mature stages over equatorial Africa than over South Asia and (2) the greater likelihood of diluted aerosol concentrations during the mature stage due to scavenging by earlier precipitation. Growing MCSs are almost equally as strong as mature MCSs over South Asia, but growing MCSs are considerably weaker than mature MCSs over equatorial Africa. The variation of convective aerosol transport frequency by developmental stage is qualitatively similar over all three basins, but convectively-generated ALs are considerably less extensive over the Amazon basin than over the other two regions. Rain rates are typically higher in convective systems over the Amazon than in convective systems over South Asia or equatorial Africa, which may enhance wet scavenging over the Amazon. Moreover, convective size and lifetime are considerably shorter over the Amazon than over South Asia or equatorial Africa, which may limit aerosol transport.

Our results show that much of the variability in convectively-generated UT aerosol layers (ALs) can be explained by variations in the size (radius) of the MCS. This dependence does not arise from the definition of ALs, as AL extents are calculated over the same horizontal distance along the CALIPSO track regardless of MCS size. The number of convective cores (NCC) is also identified as influential in several cases, although its coefficient in the
multiple regression models is opposite in sign to its coefficient in single-variable regression. NCC is strongly correlated with radius \((R = 0.92)\), and careful analysis suggests that the primary function of NCC in the model is to account for regional variability in the relationship between radius and AL extent. Stronger vertical convective flux is associated with a larger number of convective cores embedded in a MCS, and the larger vertical flux leads to more aerosol being transported to the 10-12 km altitude levels.

Optimal statistical models relating convective properties to AL extent vary considerably by both region and stage of the convective life cycle. In particular, vertical wind shear (VWS) is related to AL variability during the growing stage, convective fraction (CF) is related to AL variability during the mature stage, and aerosol optical depth is related to AL variability during the decaying stage. Regional variability in several convective properties confounds physical interpretation of the statistical relationships. This confusion may be particularly important for VWS, which receives a negative coefficient when all regions are considered, but a positive coefficient when only South Asia is considered.

Our analysis indicates that VWS may play an important role in aerosol transport during the growing stage. The VWS coefficient appears to act as a surrogate for regional variability when all regions are considered together, but AL extent over South Asia increases significantly with increasing VWS \((R = 0.30, p = 0.03)\). A positive relationship between VWS and aerosol transport is physically plausible, given the close relationship between convective and boundary layer properties during the growing stage and the potential for VWS to enhance convective aerosol transport by establishing separate regions for updrafts and downdrafts. Our results suggest that this physical interpretation is most likely, but we are unable to conclusively rule out alternative explanations (such as sub-regional
co-variability in VWS and AL extent). A full evaluation of this hypothesis will require detailed model simulations.

Our statistical model results further indicate that mature systems with similar sizes will typically produce more extensive ALs when the CF is larger. This positive dependence on CF is only observed for mature systems. By contrast, the extent of ALs associated with decaying systems is negatively correlated with aerosol optical depth (AOD). This result is intuitively surprising, and may reflect a positive relationship between convective aerosol transport to the UT and wet scavenging in the lower troposphere over the preceding hours. We find no significant correlations between AOD and UT AL extent, although we note that our analysis is restricted to MCSs that occurred in environments that were at least somewhat polluted (i.e., at least some nearby pixels with AOD ≥ 0.3).

Our results provide valuable context regarding the relative importance of different convective properties over different regions and at different stages of the convective life cycle. In particular, we have identified a potentially important relationship between VWS and aerosol transport during the growing stage of convective development, and we have shown that variations in AL extent in the upper troposphere are largely independent of variations in lower-tropospheric AOD above the threshold value. These results demonstrate the potential value of ISCCP observations for developing deeper physical interpretations of the detailed profiles of cloud and aerosol properties provided by the A-Train satellites, and provide an interpretive framework for constructing and evaluating numerical model simulations of convective aerosol transport.

Acknowledgments. SC and RF were supported by NASA Aura Science Team Grant (No. NNX11AE72G). We thank Dr. Laura Pan and Dr. Omar Torres for helpful com-
ments and suggestions. The supercomputer at the University of Texas has been used to store and analyze the data. We acknowledge the providers of the ISCCP, CALIPSO, CloudSat, Aura MLS, OMI, Aqua MODIS, and MERRA data (see Table 1).

References


Heese, B., and M. Wiegner (2008), Vertical aerosol profiles from Raman polarization lidar observations during the dry season AMMA field campaign, *J. Geophys. Res. Atmos.*, ©2015 American Geophysical Union. All Rights Reserved.


Kacenelenbogen, M., et al. (2011), An accuracy assessment of the CALIOP/CALIPSO version 2/version 3 daytime aerosol extinction product based on a detailed multi-sensor,


Livingston, J. M., et al. (2014), Comparison of MODIS 3 km and 10 km resolution aerosol optical depth retrievals over land with airborne sunphotometer measurements during ©2015 American Geophysical Union. All Rights Reserved.


©2015 American Geophysical Union. All Rights Reserved.
Okita, T., H. Hara, and N. Fukuzaki (1996), Measurements of atmospheric SO2 and
SO42-, and determination of the wet scavenging coefficient of sulfate aerosols for the
winter monsoon season over the Sea of Japan, *Atmos. Environ.*, 30(22), 3733–3739,

Omar, A. H., et al. (2009), The CALIPSO automated aerosol classification and li-
10.1175/2009jtecha1231.1.

Petersen, W. A., and S. A. Rutledge (2001), Regional variability in tropical convection:

of convection and lightning activity in the southern Amazon as a function of cross-

Pratt, K. A., et al. (2010), In situ chemical characterization of aged biomass-
burning aerosols impacting cold wave clouds, *J. Atmos. Sci.*, 67(8), 2451–2468, doi:
10.1175/2010jas3330.1.

Pumphrey, H. C., R. E. Cofield, M. J. Filipiak, and N. J. Livesey (2009), An all-

Redemann, J., M. A. Vaughan, Q. Zhang, Y. Shinozuka, P. B. Russell, J. M. Livingston,
M. Kacenelenbogen, and L. A. Remer (2012), The comparison of MODIS-Aqua (C5)
and CALIOP (V2 & V3) aerosol optical depth, *Atmos. Chem. Phys.*, 12(6), 3025–3043,
doi:10.5194/acp-12-3025-2012.

©2015 American Geophysical Union. All Rights Reserved.


©2015 American Geophysical Union. All Rights Reserved.


Figure 1. Definition of the analysis domains (red boxes). Shading shows mean precipitation rates (in mm d$^{-1}$) during the analysis period (June 2006–June 2008) based on the Tropical Rainfall Measuring Missing (TRMM) Multisatellite Precipitation Analysis (TMPA).
Figure 2. An example of a collocated MCS, including (a) the origin and development of the MCS, (b) the vertical distribution of IWC and fractional CO anomalies in the upper troposphere along the MLS track, and (c) the vertical distribution of aerosol and cloud layers along the CALIPSO and CloudSat track. This MCS was first detected by ISCCP at 21 UTC 20 January 2007 at 16.7°S and 57.3°W. The system then moved west, where the A-Train satellites observed it at approximately 06 UTC 21 January. The dark and light blue circles in (a) show the approximate central position and radius of the MCS at the A-Train overpass time. The cloud was growing when CALIPSO and CloudSat (red line) observed its trailing edge at 05:27 UTC and Aura MLS (green line) observed the system near its center at 05:35 UTC. The central positions of each satellite footprint are shown as white circles along the satellite track.

©2015 American Geophysical Union. All Rights Reserved.
Figure 3. Box and whisker plots showing the distributions of the height of the aerosol layer (HAL), the height of maximum convective detrainment (HDL), and the cloud top height (CTH) for collocated MCSs. The bottom and top of each box represent the upper and lower quartiles, respectively, and the horizontal line inside the box represents the median. The ends of the whiskers are the minimum and maximum of the data. MCSs with HAL < 6 km (the minimum observed detrainment layer) have been excluded.
Figure 4. Mean profiles of (a) IWC and (b) CO in the upper troposphere based on Aura MLS profiles collocated with growing (red), mature (purple), and decaying (blue) MCSs. Shading represents intervals of one standard error around the mean.
Figure 5. Variations in mean values of (a) vertically-integrated anomaly in number of aerosol pixels, (b) aerosol layer height (HAL), (c) storm radius, (d) cloud top height (CTH), (e) number of convective cores (NCC), (f) convective fraction (CF), and (g) precipitation rate during different stages of the convective life cycle. Error bars and shading represent intervals of two standard errors around the mean.
Figure 6. Vertical profiles of aerosol transport associated with growing (red), mature (purple), decaying (blue) and short-lived convective events (gray) over (a) equatorial Africa, (b) South Asia, and (c) the Amazon basin. The left side of each panel shows the vertical distribution of the fraction of convective events for which the anomaly in aerosol pixels relative to the background profile is positive. The right panel shows the mean anomaly in the number of aerosol pixels at that height. The vertical resolution is 2 km. Error bars indicate intervals of one standard error around the mean.
Figure 7. Fraction of variance of aerosol layer extent in the binned data explained by the independent variables in the optimal statistical models for aerosol transport associated with (a) all MCSs, (b) growing MCSs, (c) mature MCSs, and (d) decaying MCSs. Gray area shows fraction of variance not explained.
Figure 8. Scatter matrix showing covariability among the independent variables (ordinates in top three rows; abscissas in left three columns) and the dependent variable (ordinate in bottom row; abscissa in right column) in the optimal statistical model for aerosol transport associated with all MCSs. The model independent variables (in order of importance) are radius, number of convective cores (NCC), and vertical wind shear (VWS). The dependent variable is the vertically-integrated anomaly in aerosol pixels (see text for details). Red lines in scatter plots show linear relationships; blue lines show local linear regressions using a lowess filter. Blue curves along the diagonal show the density distributions of each variable (scale as shown in the top left ordinate).
Figure 9. As in Fig. 8, but for growing MCSs. The independent variables are radius and vertical wind shear (VWS).
Figure 10. As in Fig. 8, but for mature MCSs. The independent variables are radius, number of convective cores (NCC), and convective fraction (CF).
Figure 11. As in Fig. 8, but for decaying MCSs. The independent variables are radius and aerosol optical depth (AOD).
Figure 12. Graphical representations of the statistical models for (a) all MCSs, (b) growing MCSs, (c) mature MCSs, and (d) decaying MCSs. The models for all MCSs and mature MCSs have been collapsed into two dimensions by using the strong linear correlation between radius and NCC (see text and Figs. 8 and 10). A scatter plot of the binned data for each case is shown for reference, and the fraction of variance explained by the linear model is shown at the top right of each panel.
Table 1. Specifications for data sets used in this paper, including observation or model platform, product identifier, version, resolution (H: horizontal, V: vertical, T: time), and data source. All but MERRA are based on satellite observations.

<table>
<thead>
<tr>
<th>Platform</th>
<th>Identifier</th>
<th>Version</th>
<th>Resolution</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aura MLS</td>
<td>IWC,CO</td>
<td>v3.3 L2</td>
<td>(H) ~300 km × 7 km (V) ~3 km</td>
<td>GES DISC&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>CALIPSO</td>
<td>CAL_LID_L2_VFM</td>
<td>v3.01 L2</td>
<td>(H) ~0.33–5 km (V) ~60–180 m</td>
<td>LARC ADSC&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>CloudSat</td>
<td>2B-CWC-RO</td>
<td>R04 L2</td>
<td>(H) 2.5 km (V) 250 m</td>
<td>CS DPC&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>ISCCP</td>
<td>Convection Tracking</td>
<td>DX L3</td>
<td>(H) 30 km × 30 km (T) 3-hourly</td>
<td>GISS&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>MERRA</td>
<td>MAI6NPANA</td>
<td>v1, stream 3</td>
<td>(H) 0.5° × 0.67° (V) 42 pressure levels (T) 6-hourly</td>
<td>GES DISC</td>
</tr>
<tr>
<td></td>
<td>MAI3CPASM</td>
<td>v1, stream 3</td>
<td>(H) 1.25° × 1.25° (V) 42 pressure levels (T) 3-hourly</td>
<td></td>
</tr>
<tr>
<td>MODIS</td>
<td>MYD04-L2</td>
<td>v5 L2</td>
<td>(H) 10 km × 10 km (T) daily</td>
<td>LAADS&lt;sup&gt;e&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>MYD08-D3</td>
<td>v5 L3</td>
<td>(H) 1° × 1°</td>
<td></td>
</tr>
<tr>
<td>OMI</td>
<td>OMAERUV</td>
<td>v3 L2</td>
<td>(H) 13 km × 24 km (T) 3-hourly</td>
<td>GES DISC</td>
</tr>
<tr>
<td>TRMM</td>
<td>3B42</td>
<td>v7 L3</td>
<td>(H) 0.25° × 0.25°</td>
<td>GES DISC</td>
</tr>
</tbody>
</table>

<sup>a</sup> Goddard Earth Sciences Data and Information Services Center (http://disc.sci.gsfc.nasa.gov)

<sup>b</sup> Langley Atmospheric Science Data Center (https://www-calipso.larc.nasa.gov)

<sup>c</sup> CloudSat Data Processing Center (http://www.cloudsat.cira.colostate.edu)

<sup>d</sup> Goddard Institute for Space Studies (http://isccp.giss.nasa.gov)

<sup>e</sup> Level 1 and Atmosphere Archive and Distribution System (http://ladsweb.nascom.nasa.gov)

©2015 American Geophysical Union. All Rights Reserved.
Table 2. Coefficients for background-relative aerosol pixel counts between 4 and 20 km in optimal statistical models by stage of convective development, along with the correlation coefficient between the expected and predicted values in the test data ($R^2$) and the explained variance in the test data (EV).

<table>
<thead>
<tr>
<th>Stage</th>
<th>β₀</th>
<th>β_{RAD}</th>
<th>β_{NCC}</th>
<th>β_{VWS}</th>
<th>β_{CF}</th>
<th>β_{AOD}</th>
<th>$R^2$</th>
<th>EV</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>1690</td>
<td>6.66</td>
<td>-54.1</td>
<td>-44.6</td>
<td></td>
<td></td>
<td>0.64 ± 0.02</td>
<td>69 ± 5%</td>
</tr>
<tr>
<td>Growing</td>
<td>1950</td>
<td>5.19</td>
<td>-52.6</td>
<td></td>
<td></td>
<td></td>
<td>0.22 ± 0.04</td>
<td>32 ± 3%</td>
</tr>
<tr>
<td>Mature</td>
<td>1860</td>
<td>6.15</td>
<td>-67.4</td>
<td>9.80</td>
<td></td>
<td></td>
<td>0.62 ± 0.02</td>
<td>66 ± 2%</td>
</tr>
<tr>
<td>Decaying</td>
<td>1120</td>
<td>4.99</td>
<td></td>
<td>-745</td>
<td>-44.6</td>
<td>-52.6</td>
<td>0.09 ± 0.05</td>
<td>24 ± 3%</td>
</tr>
</tbody>
</table>

©2015 American Geophysical Union. All Rights Reserved.
Table 3. Coefficients for background-relative aerosol pixel counts between 4 and 20 km in optimal statistical models by region. Data for equatorial Africa is separated into different stages of the convective life cycle (G: growing; M: mature; D: decaying). Sample sizes for South Asia and the Amazon basin are too small to separate by stage of convective life cycle.

<table>
<thead>
<tr>
<th></th>
<th>$\beta_0$</th>
<th>$\beta_{RAD}$</th>
<th>$\beta_{NCC}$</th>
<th>$\beta_{VWS}$</th>
<th>$\beta_{CF}$</th>
<th>$\beta_{AOD}$</th>
<th>$\beta_{CTH}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Africa</td>
<td>1960</td>
<td>5.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Africa (G)</td>
<td>2100</td>
<td>4.35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Africa (M)</td>
<td>2350</td>
<td>5.39</td>
<td>$-66.1$</td>
<td>6.62</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Africa (D)</td>
<td>1300</td>
<td>3.96</td>
<td></td>
<td>$-776$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>South Asia</td>
<td>1632</td>
<td>3.65</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>57.5</td>
</tr>
<tr>
<td>Amazon</td>
<td>475</td>
<td>2.63</td>
<td>$-92.2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

©2015 American Geophysical Union. All Rights Reserved.