Evaluation and utilization of MODIS and CALIPSO aerosol retrievals over a complex terrain in Himalaya

Ashish Kumar[1, 2], Narendra Singh[1], Anshumali[2], Raman Solanki[3]

[1]Aryabhatta Research Institute of Observational Sciences (ARIES), Manora peak, Nainital, India

[2]Indian Institute of Technology (Indian School of Mines), Dhanbad, India

[3]National Astronomical Research Institute of Thailand (NARIT), Chiangmai, Thailand

Abstract

The study elucidate upon the evaluation of satellite retrievals with ground based aerosol optical depth (AOD) measurements, their utilization in LiDAR ratio (LR) estimation, boundary layer (BL) height determination and the case studies on aerosol transport over Himalayan region. The AOD retrievals from the latest level-2 data collections (C5.1 and C6.0) of MODerate resolution Imaging Spectroradiometer (MODIS) onboard Aqua and Terra satellites and Cloud-Aerosol LiDAR and Infrared Pathfinder Satellite Observations (CALIPSO) versions (4.10 and 3) are subjected for quantitative analysis to assess the level of agreement with the quality assured level-2 ground based AErosol RObotic NETwork (AERONET) measurements over Manora peak (29.36° N, 79.46°E), a high altitude site in the Himalayas. Analysis revealed that the AOD from the latest MODIS Terra C6.0 deep blue (DB) 30 km × 30 km and CALIPSO ver. 4.10 (overpass within ~100 km distance) are in a very good agreement ($R \geq 0.9$) with that from coincident AERONET measurements averaged over the span of ±30 minutes. About 77 % of the AOD retrieved using MODIS and ~ 87 % from CALIPSO were found to be within the expected error (EE) limits. The AOD comparison between MODIS Terra C6.0 DB and CALIPSO ver. 4.10, suggested their synergic use for aerosol characterization over Himalayas. In comparison to the ver. 3, CALIPSO ver. 4.10 is found to have undergone substantial changes, and their long term inter-comparison in the grid 28.86°-29.86° N and 78.96°-79.96° E revealed that their vertical feature and aerosol sub-types are in agreement of ~ 94.6 % and ~ 68.6 %, respectively. Utilizing the AOD retrievals from AERONET and MODIS collections, the iteratively computed LR for three LiDAR systems was found to be lower (< 16) during winter and higher (> 43) during summer. Study on the BL height estimations suggested that the wavelet covariance transform (WCT) method for CALIPSO could be the best choice as compared to the threshold method, and complements well with the specific humidity gradient method used with the radiosonde observation. Case studies on the continental transport of smoke plumes emanating from crop-residue burning in post-monsoon, and long range transport of aerosols and dust over the region in summer are also discussed using the collocated measurements from ground-based AERONET and LiDAR, in conjunction with MODIS, CALIPSO, reanalysis data and trajectory modeling.

Keywords: Aerosols, CALIPSO, AOD, AERONET, MODIS, LiDAR, LR, radiosonde, dust, smoke

1. Introduction

The phenomena such as fossil fuel combustion and biomass burning are directly linked to the anthropogenic activities across the globe, affecting the weather and climate at various spatial and temporal scales. These anthropogenic sources as well as the natural sources like air-borne dust, storms etc. can alter the concentration, chemical composition, size distribution and shapes of the atmospheric aerosols (Boucher, 2015). Any such alterations in aerosol distribution can affect the climate on regional as well as on global scales (IPCC, 2014; Hansen and Sato, 2016). The understanding of the atmospheric aerosol sources and their variations over a region in conjunction with the prevailing meteorological conditions may improve the knowledge of atmospheric processes such as the radiation balance, cloud formation, precipitation and chemical processes aloft. Rising concerns on climate change demand better insight of the physical and optical properties of the aerosols by means of ground and satellite based measurements such as AERONET, LiDAR, CALIPSO and MODIS. The correlations and improved understanding on the relationship between ground based and space borne observations are also essential in formulating the reliable current and future predictions (Ramachandran and Kedia, 2013). Moreover, it is important that any artefacts or inconsistencies associated with theoretical or operational exactitudes in the aerosol measurements are to be checked and understood.

Past studies have emphasized that satellites are the best tool for broader understanding of aerosol parameters on a global scale, however, satellite measurements possess some uncertainties, especially, at the local scale which can be quantified through their assessment with the ground based measurements (Kokhanovsky et al., 2007; Hersey et al., 2015). In this context, it is important that the satellite based latest release of aerosol products are to be examined, from time to time and corrected with the ground truth on a regional scale at finer spatial resolutions.

Nevertheless, while dealing with the aerosol optical product retrieval algorithms, it is quite common to make some priori assumptions in the retrieval processes that sometime may lead to the erroneous results and incorrect conclusions. One such assumption is the unknown aerosol LR value of any Mie LiDAR system whose wrong selection may produce uncertainty in the calculation of aerosol extinction coefficients and AOD values. Likewise, the aerosol retrieval algorithms based on satellite data demand such assumptions regarding aerosol optical properties e. g. single scattering albedo (SSA) and refractive index (Kokhanovsky et al., 2007; Wang et al., 2011). Hence, the rigorous assessment of these products is essential for studies on aerosol distribution.

Furthermore, the regional climate, particularly, along the slopes of the mountain regions, is being greatly affected due to the deleterious anthropogenic interventions. The preliminary assessment of climate change with impact studies on temperature and rainfall, snow cover and glaciers, biodiversity, streams and rivers, agriculture and other sectors conducted by state of Uttarakhand have been reported (Mishra, 2014; UCOST and USERC, 2012). However, there are a very limited studies focusing on the long term impacts of aerosols on Himalayan ecosystem, due to lack of high resolution ground based measurements (Mal et al., 2016). Studies focused on Himalayan region are of paramount importance, as the occurrences of cloudbursts, flash floods, landslides etc. have increased over the region due to the human's overexploitation of natural resources by rapid urbanization, industrialization, deforestation, emissions from forest-fires, transportation etc. (Valdiya, 2008; Tiwari and Joshi, 2016).

Considering the aforementioned facts, an attempt is made to evaluate the latest versions of satellite aerosol products mainly AOD, at regional scale and compared/validated with the ground truth as previously done by the researchers (e.g. Choudhry et al., 2012; Solanki and

Singh, 2014) on the earlier versions. This would enable subsequent usage of these products for understanding the aerosol characteristics and their impact over the region. Focusing the AERONET observations over the Himalayan region, LR for different LiDAR systems is estimated and discussed using collocated measurements along with the MODIS satellite retrievals. The latest CALIPSO aerosol products are also quantitatively evaluated with its earlier versions and utilized in BL height determination over the complex terrain, as BL evolution is a key parameter to understand the vertical transport of pollutants. Hence, making use of the evaluated data sets, the transport mechanism of the aerosols from distant regions (continental and long range) is studied and explained with the trajectory model, reanalysis data, and satellite products. The subsequent sections describe about site, instrumentation and data, methodologies, results and discussion, which is followed by the conclusion at the end.

2. Site, instrumentation and data

2.1. Site Description

Manora peak (29.36° N, 79.46°E, 1939 m amsl) is a high altitude regional representative site in the central Himalayas located near the city of Nainital in the state of Uttarakhand (Solanki and Singh, 2014; Solanki et al., 2016). The study using ground and satellite based measurements over the site amidst undulating topography in the free tropospheric conditions can be of great relevance. This pristine site is surrounded by the Himalayan mountain ranges and towards its South is the Indo-Gangetic plains (known as Tarai). During the past two decades, industrialization has grown up rapidly in these Tarai portions (Kazuo, 2014) and the pollutants are being transported to the site quite often (Ojha et al., 2012; Sarangi et al., 2014). Therefore, the site has a great advantage to study the continental as well as long range transport of the pollutants, and additionally it provides the background values of the aerosol parameters. Further details of the site, variations in meteorology and synoptic-wind patterns can be found elsewhere (Ojha et al., 2014; Singh et al., 2016). The data sources used in the present study are described in the subsequent sub-sections.

2.2. LiDAR observations

During the period from 2006 - 2014, three Mie LiDAR systems at Manora peak were utilized for the vertical profiling of atmospheric aerosols in the free troposphere. The first system was operated during 2006-2008 (Hegde et al., 2009), and the second system between 2010 and the mid of 2011 (Bangia et al., 2011). The third system named as LiDAR for Atmospheric Measurement and Probing (LAMP) is an upgraded version of the first one and was made operational since October 2011 (Solanki et al., 2013; Solanki and Singh, 2014). LAMP is much more compact monostatic version of the first one and is equipped with RS-232 and Ethernet interfaces, built-in acousto-optic modulator for Q-switching and high quality optical assemblies. **Table 1** summarizes the major differences among all the three versions of LiDAR systems. All LiDAR systems at the study site were operated in late-evening hours under cloud free conditions on the days considered in this study, and the data acquired is presented collectively.

Table 1.

Technical specifications of the Mie LiDAR systems operated at the site.

Parameters	**LiDAR-I**	**LiDAR-II**	**LiDAR-III**
Wavelength	532 nm	532 nm	532 nm
Telescope	Cassegrain, 150 mm dia, ~ 1 mrad Focal ratio – f/9	Cassegrain, 380 mm dia, ~ 6 mrad Focal ratio – f/15	Cassegrain, 150 mm dia, ~ 400 μrad Focal ratio – f/9

Laser Type	Q-switched, Nd:YAG	Q-switched, Nd:YAG	Acousto-optic, Q-switched, Nd:YAG
Beam expander	8X	10X	8X
Resolution	30 m	300 m	15 m
Complete Overlap	150 m	300 m	90 m

2.3. AERONET measurements

The AERONET program is an inclusive federation of ground-based remote sensing aerosol networks established by National Aeronautics and Space Administration (NASA) and PHOtométrie pour le Traitement Opérationnel de Normalisation Satellitaire (PHOTONS) and greatly expanded by networks and collaborators from national agencies, institutes, and other partners (Holben et al., 1998). The program provides a long-term database of globally distributed observations of aerosol optical, microphysical and radiative properties. AERONET measurements are considered to be the ground truth due to its worldwide use and acceptability in the validation and bias corrections of the satellite retrievals (Bréon et al., 2011; Bibi et al., 2015; Bilal et al., 2016).

In the present study, the quality assured and well calibrated, level-2 AERONET data sets are used that include automatic cloud screening and utilize the tools such as 1-min stability, diurnal stability, smoothness tests etc. The day-time measurement of columnar aerosol parameters at wavelengths between 440 – 870 nm using the collocated AERONET sun photometer system, are utilized with the coincident night-time LiDAR observations. The AERONET provides the high quality data on a wider scale across the globe, so the methodologies adopted in the present work can be utilized by the larger science community.

2.4. MODIS products

MODIS is a key Earth observing instrument launched aboard NASA's Terra (MOD) and Aqua (MYD) satellites on 18 December 1999 and 4 May 2002 respectively (Savtchenko et al., 2004). Terra's orbit around the Earth is so timed that it passes from North to South across the Equator (descending node) in the morning, while Aqua passes South to North over the Equator (ascending node) in the afternoon. MODIS satellite passes over the study region twice a day and specifically, Terra crosses between 10:00 – 11:00 hours local time (LT), while Aqua between 13:00 – 14:00 hours LT. MODIS Terra and Aqua satellites view the entire Earth's surface in every 1 to 2 days, acquiring data since March 2000 for Terra, and July 2002 for Aqua in 36 spectral bands between 0.4 and 14.4 µm. The acquired MODIS data are available in the hierarchy of levels (level-1 to 4) and grouped in four broad disciplines – land, atmosphere, ocean and cryosphere. The collections are also defined in MODIS data that represent the versions of MODIS data production algorithm (Savtchenko et al., 2004; Remer et al., 2005). In the present work, level-2 MODIS aerosol collections (C5.1 and C6.0) and active fire location product (C6.0) available under atmosphere and land disciplines respectively are used.

The latest level-2 MODIS aerosol product collections C5.1 and C6.0 over land and ocean are based on two algorithms, namely the deep blue (DB) and dark target (DT) (Remer et al., 2005; Levy et al., 2013; Bilal et al., 2016). DT has separate algorithms for land and ocean, whereas DB is for the land retrieval only. Both C5.1 and C6.0 contains the standard 10 km spatial resolution MODIS Terra (MOD04_L2) and Aqua (MYD04_L2) retrievals. To cater the need of resolving the local aerosol gradients and regional features in a much precise manner, the MODIS C6.0 production includes the DT aerosol product with 3 km spatial resolution under both Terra (MOD04_3K) and Aqua (MYD04_3K) platforms. Recent studies revealed that the

MODIS 3 km land product is less reliable and requires continued evaluation in contrast to the standard 10 km product (Remer et al., 2005; Levy et al., 2013; Remer et al., 2013; Nichol and Bilal, 2016; He et al., 2017). Studies were carried out on the validation of MODIS 10 km aerosol retrievals over land with the ground based measurements, and over ocean with the shipborne measurements (Remer et al, 2002; Remer et al., 2005; Wang et al., 2011; Sayer et al., 2013). Majority of the cited studies have found reliable and good agreements of the 10 km retrievals with ground based measurements. Therefore, here 10 km MODIS level-2 latest C6.0 (DT and DB) and C5.1 (DT) with quality flag 3 were chosen for assessment and comparison with the ground truth over the region.

For one of the case studies presented in section 4, the MODIS C6.0 standard active fire location product MCD14ML is extracted from NASA Fire Information for Resource Management System (FIRMS) database which is produced using the most up-to-date algorithms in the form of monthly files containing the geographic location, date, brightness temperature, updated fire radiative power (FRP), fire type and the confidence levels for each fire pixel detected by the Terra and Aqua MODIS sensors. The confidence estimate is expressed in percentage and is classified as 0% - 29 % for low, 30% - 79% for nominal, and 80% - 100% for high fire-events (Giglio et al., 2003; Giglio, 2005).

2.5. CALIPSO products

CALIPSO was launched in April 2006 under a joint mission of NASA and the French space agency, Centre National d’Etudes Spatiales (CNES). It is equipped with a dual wavelength (550 and 1064 nm) polarization LiDAR system referred as Cloud and Aerosol LiDAR with Orthogonal Polarization (CALIOP) for providing the long term database of global aerosol vertical profiles (Winker et al., 2009 and 2010). The CALIOP laser transmitter is a diodepumped Nd:YAG laser that emits simultaneous co-aligned pulses at 532 and 1064 nm. The laser generates optical pulses of ~20 ns long with 110 mJ of energy at both the wavelengths. The receiver sub-systems measures the backscattered signal intensity at 1064 nm and the two backscattered orthogonal polarization components at 532 nm (Winker et al., 2009 and 2010; Hunt et al., 2009).

At present, the researchers worldwide, are utilizing the CALIPSO products to a great extent in order to understand the impact of aerosol and cloud on the Earth's radiation budget. The CALIPSO/CALIOP (*ver. 3 and 4.10*) aerosol products used in this study are:

- *Level-1B* products (temporal resolution: 0.05 sec, vertical and spatial resolution: 30 m (0-8.2 km) and 333 m)
- *Level-2* products:
  - Aerosol profile (temporal resolution: 5.92 sec, vertical and spatial resolution: 60 m × 5 km)
  - Aerosol layer (temporal resolution: 0.74 sec, spatial resolution: 5 km)
  - Vertical feature mask (VFM) product (temporal resolution: 0.74 sec, vertical and spatial resolution: 30 m (up to 8.2 km) and 333 m)

2.6. Reanalysis products

The reanalysis products are produced from the available atmospheric observations and dynamic models. There are a number of reanalysis products available on the global scale such as NCEP-NCAR reanalysis (NNR), ERA-40, ERA-Interim, Modern-Era Retrospective analysis for Research and Applications (MERRA) etc. (Decker et al., 2012). In the present study, to ascertain the sources of dust transport, data obtained from MERRA-2 (ver. 5.12.4) is utilized. It is the latest available reanalysis product released by NASA Global Modeling and Assimilation Office (GMAO), and is based on the Earth observing system (EOS) satellite observations (Bosilovich et al., 2016). The 6-hourly ERA-Interim wind products is also used to understand the prevailing wind pattern over the site during the period of study.

2.7. Air mass trajectory model

To trace the sources of air masses on the days showing high AOD variabilities over the site, the backward trajectory analysis is carried out using National Oceanic and Atmospheric Administration (NOAA) Hybrid Single Particle Lagrangian Integrated Trajectory (HYSPLIT) model (Draxler and Hess, 1997, 1998). The model utilizes several meteorological parameters like rainfall, humidity, temperature and solar radiation flux for computing the air mass trajectories at different height levels. The trajectory analysis basically characterizes the air masses and the origin, in order to understand the impact on meteorological conditions and the aerosol transport (Draxler and Hess, 1997, 1998; Stein et al., 2015).

3. Methodology

3.1. Selection criteria for evaluation of satellite products

Along with the intermittent observations made using LiDAR systems, the collocated AERONET measurements were also available for the period of 2008- 2012. The AERONET observation were made from the site in two phases – each during April 2008-February 2011 (Nainital Station) and August 2011-March 2012 (ARM_Nainital Station). Out of the datasets collected during above period, common days of reliable measurements with best temporal match were selected. Based on the LiDAR profiles the selected data were further screened for clear sky conditions. In this process the usable datasets turned out to be 37, and considered for the analysis.

In order to match with the above identified 37 days, the selection of MODIS data sets is done on the basis of closest overpass to the site and the availability of AERONET data during the overpass. The details on the number of spatially and temporally coincident data sets (*N*) obtained as a result are given in **Table 2**. To achieve a valid comparison between AOD values measured from MODIS and AERONET instruments, a well-known spatio-temporal averaging technique is adopted (Ichoku et al., 2002) and multiple metrics were utilized to quantify the results. Taking into account the mean MODIS AOD values within 20 km × 20 km and 30 km × 30 km grid and AERONET AOD averaged over ±30 minutes and ±15 minutes, various cases on spatio-temporal combinations were examined and the levels of agreement were established.

MODIS being the passive remote sensor provides a single columnar value of aerosol parameter, which lacks the information of aerosol vertical distribution, a rather important parameter to quantify aerosol effects in the atmosphere. In this context, the CALIPSO satellite products were also evaluated and used in the BL height estimation and case studies on the evolution and transport of aerosols. In order to understand the associated changes (vertical features and aerosol sub-types) in the two versions (ver. 3 and 4.10) of CALIPSO data over complex Himalayan terrain, level-2 VFM profiles available within the grid of ± 0.5° ranging from 28.86°-29.86° N and 78.96°-79.96° E for the period August 2006- April 2017 were analysed. The results in the form of confusion matrix are presented and discussed in relevance to the changes in the feature types and aerosol sub-types between ver. 3 and ver. 4.10 data. Further, to evaluate the AOD values from two versions of CALIPSO with AERONET, a total of 23 data sets were identified based on the criteria of high cloud aerosol discrimination (CAD) score (between -35 and -100), the presence of 5 or more valid vertical profiles up to 4.5 km altitudes within the horizontal distance of ~ 100 km from the site, and the availability of coincident AERONET measurements within ± 30 minutes of the closest CALIPSO ground track.

For the period October 2006 – December 2014, 54 good cases of AOD measurements both from the MODIS Terra DB C6.0 and CALIPSO (within 100 km) were identified, and the same were utilized in their inter-comparison. The selection procedure is based on the screening and coincidence constraints, and the sequence is as follows:

- To account for the best temporal match between the CALIPSO and MODIS satellite overpasses, only the day-time CALIPSO profiles were considered where the time difference between two observations is within 3 hours.
- For MODIS, the averaged AOD values (550 nm) with quality flag 3, measured within 30 km × 30 km from the site were considered.
- For CALIPSO, the average of the column AOD values (532 nm) reported in CALIPSO level-2 aerosol layer product is used. The selection criteria set for the AOD is: CAD score (between -35 to -100), Extinction QC 532 flag (0 or 1), column optical depth uncertainty (between 0 and 0.5 × AOD), CALIOP initial LR = final LR, surface elevation > 1200 m, and the horizontal averaging ≤ 80 km (Young and Vaughan, 2009; Vaughan et al., 2016).

3.2. LR estimations

The LR for any single wavelength ground-based LiDAR is a key parameter that needs to be known for the retrieval of aerosol vertical profiles. Basically, it is the ratio of aerosol extinction coefficient (α_{aer}) and the aerosol backscatter coefficient (β_{aer}) that is linked to the regional aerosol characteristics like shape, size and composition. A-priori hypotheses for LR in the range between 20 to 100 sr is quite common, but LR computed by constraining the AOD from LiDAR through AERONET or MODIS measurements can be the better choice than former (He et al., 2006). In the later approach, initially the LiDAR range-corrected signal (RCS) is processed for AOD computation with a fixed LR, which then undergoes several iterations to produce an adjusted LR at a point where the difference between LiDAR derived AOD (τ_{LiDAR}) and the AOD retrieved from AERONET ($\tau_{AERONET}$) or MODIS (τ_{MODIS}) measurements is minimal. Similar approach is adopted here to find out the best LR values for the three LiDAR systems operated in night-time under clear-sky conditions during different seasons. The adjusted range-independent LR values, where the LiDAR AOD showed the best match within the tolerance of ±0.5 %, are considered to be the final LR. To account for any discrepancies between day and night-time AOD measurements from AERONET and the ground based LiDAR, about 95% of the data sets were so chosen that the diurnal variations in AOD and AE (440-870 nm) fall in the limits of ±0.05 and ±0.2, respectively (Amiridis et al., 2011). Such a bound is employed to ensure that the intrusion of aerosol from other locations is almost insignificant and the aerosol loading remains nearly the same during day and night over the site.

The LiDAR AOD in LR retrieval process has been computed using the relation:

$$\tau_{LiDAR,532} = \int_{z_0}^{z_1} \alpha_{aer}(z)dz + \int_{z_1}^{z_2} \alpha_{aer}(z)dz \quad (1)$$

where, z_0 = height at which the LiDAR system is installed, z_1 = height at which complete overlap occurs (150 m, 300 m and 90 m considered for the three LiDAR systems, respectively), and z_2 = upper height limit considered for the columnar AOD retrievals (assumed as 4.5 km above ground level). Considering the uniform distribution of aerosols between z_0 to z_1, it is assumed that for the three LiDAR systems, the maximum of 7.5 %, 15 % and 5 % of the AOD values, respectively are confined within the respective overlap regions.

3.3. BL height estimation

The BL height is an important meteorological parameter that determines the extent to which the dispersion of pollutants, heat and moisture take place, and is very useful parameter for weather, climate and pollution studies (Monks et al., 2009). In this context, an accurate determination of BL height, using different data sources over the complex high altitude site, where the upslope and downslope airflows vary with time, is of great interest. From Manora peak site, the radiosonde launches were conducted four times a day during 2011-2012 (Singh et al., 2016), so taking this an advantage, and considering the fact that BL depth can be derived from the radiosonde (Seibert et al., 2000; Singh et al., 2016) and CALIPSO data (Jordan et al., 2010; McGrath Spangler and Denning, 2012), an attempt is made to estimate and compare the BL height computed from the in-situ radiosonde observations and near coterminous CALIPSO level-1B data (< 100 km overpass distance; ver. 4.10). To ascertain cloud-free cases, the parameters (signal intensity at the surface, depolarization ratio, color ratio and vertical features) from CALIPSO level-1B and 2 data products are examined for the period June 2011- March 2012. After discarding the cloud contaminated profiles, a total of 10 day-time cloud-free CALIPSO profiles, in the temporal match (< 2 hours) with the radiosonde observations were identified and selected for BL height estimation. The day-time cases were selected to avoid the influence of the residual layer and heavy surface inversion (Su et al., 2017).

To estimate the BL heights from radiosonde, the vertical gradient method is used for the potential temperature (PT) and specific humidity (SH), that is expressed as:

$$\frac{dX(y_i)}{dy} = \frac{X(y_{i+1}) - X(y_i)}{y_{i+1} - y_i} \quad (2)$$

Here, $X(y_k)$ is used to represent the PT or SH values at altitude y_k, where k represents the height intervals i up to 3.2 km amsl in vertical, that is selected on the basis of the characteristics studied over Manora peak (Singh et al., 2016). The BL height is identified as the location of the maximum vertical gradient for PT and SH changes (Seibert et al., 2000; Seidel et al., 2010).

To retrieve the BL height from CALIPSO, two methods, namely the threshold (Melfi et al., 1985; Johnson et al., 2010) and WCT (Brooks, 2003; Compton et al., 2013), are used. With the threshold method, the BL height is determined by finding the steepest gradient in total aerosol backscatter coefficient profiles (CALIPSO level-1B). In WCT method, the Haar wavelet function is applied to the total aerosol backscatter coefficient profile, and its first maxima where the sharpest decrease in the total aerosol backscatter coefficient occurs is taken up as the BL height (Baars et al., 2008). The implementation of WCT method is described using two equations (Gamage and Hagelberg, 1993; Brooks, 2003):

$$W_f\ (a,b) = \frac{1}{a}\int_{z_b}^{z_t} \beta'_{total,532}(z)\psi\left(\frac{z-b}{a}\right)dz \quad (3)$$

$$\text{and, } \psi\left(\frac{z-b}{a}\right) = \begin{cases} +1\,; & b-\frac{a}{2} \le z < b \\ -1\,; & b \le z \le b+\frac{a}{2} \\ 0\,; & elsewhere \end{cases} \quad (4)$$

where, a = dilation parameter (scale); b = vertical translation i.e. altitude at which the wavelet function is centered; $\beta'_{total,532}(z)$ = CALIPSO level-1B total aerosol backscatter coefficients as the function of altitude; $W_f(a,b)$ = wavelet covariance transform as a function of scale and translation; $\psi\left(\frac{z-b}{a}\right)$ is the Haar wavelet function, described as a symmetrical square wave with positive and negative going amplitudes; z_t and z_b are the top and bottom altitudes of $\beta'_{total,532}(z)$.

For retrieval of BL height with high degree of accuracy using the WCT method, it is essential that the dilation, 'a' should be carefully chosen (Brooks, 2003). At small dilation value, due to the spurious gradients and noisy $W_f\ (a,b)$ profile, it becomes very difficult to estimate the correct BL height, and at extremely high dilation value, the BL height becomes too high or sometime may even get missed. Therefore, the BL height estimated from the mean profile of wavelet covariance transform, $< W_f\ (b) >$, generated across the mid-range of dilation values, is considered as the optimum BL height in the present work, which is expressed as:

$$< W_f\ (b) > = \frac{1}{n} \sum_{i=1}^{n} W_f\ (a_i, b) \qquad (5)$$

and the final BL height $= max < W_f\ (b) >$, for z_b < b < z_t. This approach of selecting and averaging multiple wavelet dilation values will reduce the bias in the final BL height estimation. An example demonstrating the sensitivity analysis and the selection of appropriate dilation range for a typical CALIPSO level-1B (ver. 4.10) total aerosol backscatter coefficients profile of 16 June 2011 is available in the Supplementary data (**Figure S1 (a-d)**).

3.4. Retrieval error and wavelength conversions

The EEs associated with AERONET measured AOD (τ) and the corresponding MODIS retrievals are $EE_{AERONET}$ = ± 0.01 to ±0.02 and EE_{MODIS} = ± (0.05+0.15τ), respectively (Holben et al., 1998; Eck et al., 1999; Remer et al., 2005). Similarly, the EE associated with CALIPSO measurement is $EE_{CALIPSO}$ = ± (0.05+0.4τ) (Winker et al., 2009).The uncertainties in the AERONET measurements are wavelength (λ) dependent and are generally higher in the ultra violet spectral ranges.

MODIS and AERONET measure AOD at two different wavelengths 550 and 500 nm respectively, and to make a valid comparison between the two, AOD at 500 nm is converted to AOD at 550 nm by taking into account the Angstrom exponent (AE) provided by AERONET in the wavelength range of 440 - 870 nm, using the relation as follows (Eck et al., 1999):

$$\tau_{required} = \tau_{measured} \left[\frac{\lambda_{required}}{\lambda_{measured}}\right]^{-AE} \qquad (6)$$

where, $AE = -\frac{log\left[\frac{\tau_{\lambda_1}}{\tau_{\lambda_2}}\right]}{log\left[\frac{\lambda_1}{\lambda_2}\right]}$, τ_{λ_1} and τ_{λ_2} are the AOD values at wavelengths λ_1 and λ_2.

Similar analogy is used to make the comparison of CALIPSO (532 nm) with AERONET. In all the wavelength conversions, it is assumed that the errors introduced were negligible.

3.5. Performance parameters

To evaluate the performance of the aerosol retrievals from MODIS and CALIPSO/CALIOP satellite products, the following statistical parameters were computed on *N* coincident data sets:

(i) Mean bias error (MBE) = $\frac{\sum(\tau_{Satellite}-\tau_{AERONET})}{N}$, is the measure of overall bias error and the values > 0 indicate overestimation, whereas the values < 0 represent underestimation of the satellite retrieved AODs with the ground truth.

(ii) Average error ratio (AER) = $\frac{\sum(\tau_{MODIS}-\tau_{AERONET})}{N} \mathrm{X} \frac{1}{EE}$, is the measure of the comparison between the actual error and the EE. |AER| ≤ 1 is the good match, and |AER| > 1 represents the poor match. In case of satellite data, AER< 0 represents underestimation of the measurement, and AER > 0 reveals overestimation.

(iii) Root mean square error (RootMSE) – It is the root mean square of the error in the regression, computed as the square root of the reduced Chi-square i.e. $\sqrt{Reduced\ \chi^2}$. It provides the variability/standard deviations of the data from the regression line. Lower the value of RootMSE, better will be the agreement between the regression and the data.

(iv) Percentage mean relative deviation (MRD) = $\frac{100}{N}\sum_{i=1}^{N}\frac{|M_{version2}-M_{version1}|}{M_{version2}}$, is the measure of mean divergence of the data version 2 ($M_{version2}$) from the version 1 ($M_{version1}$).

(v) Percentage EE (%EE) = Percentage of AOD values falling within EE limits that is defined in section 3.4. If out of N AOD values, M values falls within the EE limits, then, $\%\, EE = \frac{M}{N} \times 100$.

(vi) Standard deviation (SD) – It is a measure of how spread out a data set is, and is equal to the square root of the variance. Mathematically, for N samples (X_1, X_2,… X_N), $SD = \sqrt{\frac{\sum_{i=1}^{N}(X_i - \bar{X})^2}{N-1}}$, where, $\bar{X}$ is the mean value of N samples.

(vii) Standard error of mean (SEM) = $\frac{SD}{\sqrt{N}}$, and is the measure of the variability associated with estimating a mean.

4. Results and discussion

4.1. Evaluation of Satellite aerosol retrievals – MODIS and CALIPSO

4.1.1. Assessment of MODIS collections with coincident AERONET measurements

In order to make an assessment of MODIS collections (C5.1 and C6.0), four cases on spatial and temporal comparisons were considered. In the first two cases, the average of MODIS AOD within 20 km × 20 km from the study site are compared with the average AOD from AERONET within (a) ±30 minutes and (b) ±15 minutes, and in the latter two cases, the average of MODIS AOD within 30 km × 30 km are compared with the average AOD from AERONET within (c) ±30 minutes and (d) ±15 minutes of the closest Terra/Aqua overpass occurrences. From the statistics on the four spatio-temporal combinations, as given in **Table 2**, case (c) is found to perform better with almost all the metrics in good agreement as compared to other three cases. Since, MODIS Terra shows high correlation (R ~0.90) and high %EE (~62) in comparison to Aqua (R ~0.75, %EE ~54) with ground based AERONET measurements. Therefore, one has to essentially decide upon the most suitable collection of Terra retrievals out of C5.1 DT, C6.0 DT and C6.0 DB.

Table 2.

Statistical summary on AOD (550 nm) comparison for the four spatio-temporal cases.

		Case-(a)	**Case-(b)**	**Case-(c)**	**Case-(d)**
MODIS Terra DT: C5.1, C6.0 DB: C6.0	N	24	24	27	27
	Pearson Correlation (mean±SD±SEM)	0.90±0.05±0.03	0.89±0.05±0.03	0.90±0.04±0.02	0.90±0.03±0.02
	RootMSE (mean±SD)	0.068±0.02	0.069±0.02	0.067±0.01	0.068±0.02
	%EE (mean±SD)	52.8±17.4	48.6±17.3	61.7±15.5	56.8±17.1
	MBE (mean±SD)	0.052±0.04	0.052±0.04	0.054±0.04	0.054±0.04
	AER (mean±SD)	0.13±0.14	0.14±0.16	0.105±0.10	0.12±0.13
MODIS Aqua DT: C5.1, C6.0 DB: C6.0	N	22	21	22	22
	Pearson Correlation (mean±SD±SEM)	0.72±0.06±0.03	0.66±0.02±0.01	0.75±0.06±0.04	0.69±0.04±0.02
	RootMSE (mean±SD)	0.081±0.003	0.089±0.003	0.084±0.006	0.093±0.004
	%EE (mean±SD)	57.5±2.6	53.9±14.6	54.5±9.1	51.5±13.8
	MBE (mean±SD)	0.057±0.03	0.056±0.03	0.059±0.03	0.055±0.03
	AER (mean±SD)	0.102±0.06	0.12±0.10	0.12±0.08	0.13±0.11

In order to achieve the reasonable assessment with the ground truth, the study in case (c) is extended for 174 coincident AOD measurements available during 2008-2010. Each of the three collections is subjected to one-one line comparison with AERONET AOD values as shown in **Fig. 1**. From the figure, it is evident that MODIS Terra C5.1 DT showed high correlation (R ~0.92) and low RootMSE (~ 0.087), thereby reflecting lowest variability, whereas, MODIS Terra C6.0 DB demonstrated the highest percentage of MODIS AOD values falling within the defined EE ($\pm 0.05 \pm 0.15\tau$) boundary (%EE ~77.01). However, MODIS Terra C6.0 DT is showing good correlation (R ~0.90), but the least %EE (~45.98) among all.

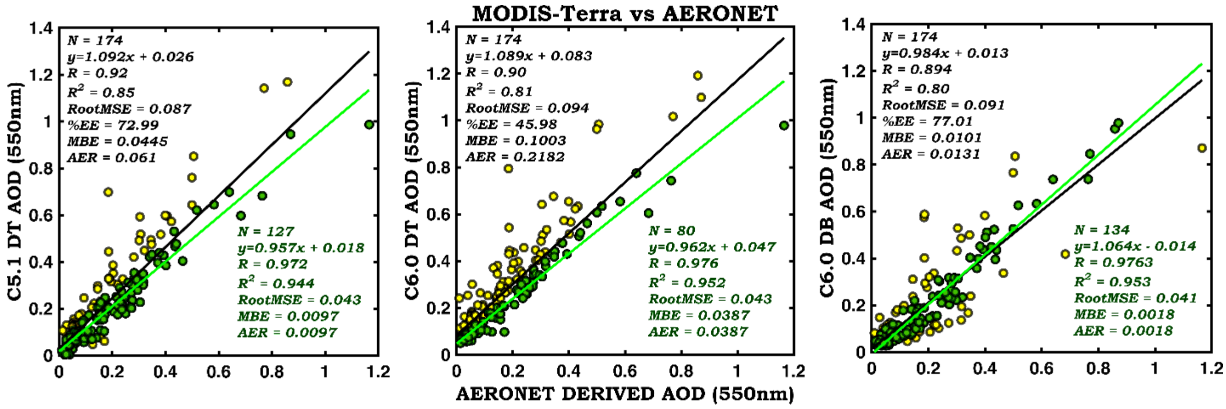

Fig.1. Scatter plot with one-one line comparison for MODIS Terra and AERONET AOD measurements; green + yellow = complete, green = for values within %EE.

Based on the %EE values, a total of 127, 80 and 134 data points respectively for C5.1 DT, C6.0 DT and C6.0 DB are found to be within EE limits, which are further subjected to statistical analysis. The statistics confirms that the MODIS Terra C6.0 DB is the best choice among others.

MBE and AER values for the three collections are found to be positive, indicating the overestimation of MODIS AOD as compared to the ground based AERONET measurement. The overestimation may be attributed to the huge spatial differences in measurements, as the ground-based AOD measurement through AERONET is a point observation, whereas the MODIS retrievals of AODs are over 10 km x 10 km at each instance. Owing to the large spatial coverage, the MODIS retrieved AOD may get influenced due to the presence of small clouds, geographical locations etc.

4.1.2. Inter-comparison of CALIPSO versions

CALIPSO mission announced the release of ver. 4.10 data product on 8 November 2016 and, in comparison to the earlier ver. 3, the quality in ver. 4.10 release is enhanced with the inclusion of the updated digital elevation map (DEM) from CloudSat and high-quality MERRA-2 product (Vaughan et al., 2016).

In going from CALIPSO ver. 3 to ver. 4.10, major code and algorithm modifications were implemented, e.g. improved data filtering strategies, changes in the calibration algorithms for both 532 nm and 1064 nm, and the revised probability density functions (PDFs) in CAD algorithm. To investigate the changes in the vertical features provided by two CALIPSO versions, an analysis is performed by extracting the feature classification flag (FCF) of each detected layer from the VFM files for the CALIPSO transacts within the defined geographical region of 28.86°-29.86° N and 78.96°-79.96° E for the period of August 2006- April 2017. The vertical feature type, with a confidence level of at least 'medium' i.e. $50 \leq |\text{CAD score}| < 70$ for aerosol and cloud layers (confirmed using FCF bits 4 and 5), is obtained by decoding the FCF bits 1-3 in decimal form (Vaughan et al., 2016). The changes w.r.t. ver. 3, as observed in ver. 4.10 are explained by constructing a confusion matrix as given in the **Table 3.** The overall agreement between ver. 3 and ver. 4.10 in this case is computed by summing the samples which remained unchanged (e.g. clear air – clear air, cloud - cloud) divided by the total number of samples expressed in percentage, and is found to be 94.64 % when a total of 437 day and night-time profiles were taken into account. The level of disagreement between ver. 3 and ver. 4.10 is ~1 % higher in the night-time profiles as compared to the day-time profiles (refer Supplementary data **Table S1 and S2**).