

# IMPROVING RETRIEVAL ACCURACY FOR AEROSOL OPTICAL DEPTH BY FUSION OF MODIS AND CALIOP DATA

*Bo Han, Hongpeng Ding, Yingying Ma, Wei Gong*

Original scientific paper

MODIS and CALIOP are two independent observation instruments in the A-train satellite constellation. They both provide aerosol optical depth (AOD) retrievals and scan the same points on the Earth's surface within a two-minute interval. With different design principles, MODIS and CALIOP instruments obtain varying AOD retrieval accuracies under different conditions. In this paper, we propose a two-stage fusion approach, including an analysis stage and an integration stage, to improve AOD retrieval accuracy. In the analysis stage, we systematically analyse conditions where MODIS retrieves well while CALIOP does not, and vice-versa. In the integration stage, we combine AOD retrievals from both instruments together by drawing on the other's strong points to make up one's weak points. We test the fusion approach on the two-year collocated data from MODIS, CALIOP and AERONET. The fusion result is significantly more accurate than AOD retrievals from any single observation facility.

**Keywords:** MODIS; CALIOP; AERONET; Aerosol Optical Depth; Fusion

## Poboljšanje točnosti pronalaženja podataka za optičku dubinu aerosola spajanjem podataka MODIS i CALIOP

Izvorni znanstveni članak

MODIS i CALIOP su dva neovisna promatračka instrumenta u konstelaciji A-train satelita. Obadva daju podatke za optičku dubinu aerosola (AOD - aerosol optical depth) i skeniraju iste točke na površini zemlje unutar intervala od dvije minute. Različitih principa dizajna, MODIS i CALIOP instrumenti daju različite točnosti podataka AOD pod različitim uvjetima. U ovom radu predlažemo pristup dvostepene fuzije, uključujući stadij analize i stadij integracije, kako bi se poboljšala točnost pronalaženja AOD podataka. U stadiju analize sistematski analiziramo uvjete pod kojima se MODIS-om podaci dobro pronalaze dok CALIOP-om ne i obrnuto. U stadiju integracije, kombiniramo AOD podatke obadva instrumenta zajedno koristeći prednosti jednoga da bi se nadomjestili nedostaci drugoga. Ispitujemo pristup fuzije na podacima skupljenim dvije godine uz MODIS, CALIOP i AERONET. Taj rezultat fuzije je znatno točniji od AOD prikupljenih podataka s bilo kojeg pojedinačnog instrumenta za promatranje.

**Cljučne riječi:** MODIS; CALIOP; AERONET; optička dubina aerosola; fuzije

### 1 Introduction

Aerosol distribution has a profound impact on global climate system. Currently there are many environmental observation facilities on satellites to derive aerosol optical depth (AOD) globally for aerosol monitoring and research. Some of these are on the same satellite constellation, scanning the same points on the Earth's surface in the same orbit and within a very small time interval. It makes possible to obtain multiple aerosol observations for a specific location from different satellites at almost the same time. The design principles of each observation facility are different. Consequently, previous validation results based on AERONET observations showed the AOD retrieval accuracies derived from different facilities vary under different conditions [1-10]. The results present us with two questions: whether there is a systematic and data-driven approach that can relate observing conditions to AOD retrieval accuracy for an instrument, and whether it is possible to combine AOD retrievals from two observation instruments together, that take advantage of relative advantages of each sensor to make more accurate retrievals?

We have explored the answers to these questions in the context of AOD measurements retrieved from the Moderate Resolution Imaging Spectroradiometer (MODIS) and the Cloud-Aerosol Lidar with Orthogonal Polarization (CALIP) aboard the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) satellite. MODIS and CALIOP are two independent observation instruments on Aqua and CALIPSO satellites respectively, in the A-train satellite constellation. MODIS

observes Earth from polar orbit in 36 wavelength bands ranging from 0,415  $\mu\text{m}$  to 14,5  $\mu\text{m}$  at different spatial resolution. It derives AOD by matching the TOA (top of the atmosphere) reflectance to the simulated values stored in lookup tables by using domain knowledge about aerosol models and forward simulation algorithm [1]. Its AOD retrieval algorithm has been validated and is constantly improving. The C005 AOD products developed by second-generation retrieval algorithms significantly improve the retrieval accuracy compared to the previous C004 products. Its expected error for most land AODs are within  $\pm (0.05+0.15 \times \text{AOD})$  and errors for ocean AODs are even lower [11-14]. Validation results based on AERONET sites show that under certain conditions C005 AODs have achieved accurate retrievals. Under other conditions, however, they show systematic deviation. For example, under the condition of low aerosol loading, the retrieve of AOD is more accurate in the eastern United States, Western Europe and Southern Africa sites (because their spectrum is closer to the middle "dark" and moderate "green" regimes); whereas the AOD retrieval is often overestimated by MODIS in western United States and Central Asia's sites - because of its relatively bright surface and less greenness, MODIS overestimates the AOD [15]. In another example, Jethva et al. found that the errors in the fine-dominated AOD over the Indian region were large for most AOD retrievals in the C005 data [16].

The CALIPSO satellite was designed by carefully controlling its orbit to ensure the acquisition of collocated near-simultaneous measurements with other A-train satellites. The primary instrument on the CALIPSO satellite for global profiling of aerosols and clouds is

CALIOP. This instrument estimates the aerosol layer optical depths and total column optical depth by backscattering lidar signals [17–20]. Validation studies show that CALIOP achieved an average AOD retrieval bias of -13% relative to AERONET over 147 global sites from June 2006 to May 2009 [9]. The results show that CALIOP is biased 20% and 12% below the AERONET AODs in the Northern African sites and Middle Eastern region, respectively. When these global AOD comparisons are analyzed according to aerosol types, scientists found that CALIOP presents the most significant biases at a very high confidence level for dust and marine aerosol. In addition, A.H. Omar found that generally CALIOP AODs are lower than AERONET AODs especially at small optical depths, possibly due to undetectable tenuous aerosol layers caused by low signal-to-noise ratio [8]. These comparisons illustrate that MODIS and CALIOP retrieve AOD with different accuracy under different conditions. However, in those studies, the conditions that determine the retrieval accuracy of the instruments are not fully understood, especially from the aspect of aerosol modelling and spatial temporal analysis.

In this paper, we propose a two-stage fusion approach, including an analysis stage and an integration stage, to improve AOD retrieval accuracy. The analysis stage is based on a data-driven decision-tree algorithm. It systematically analyses the relationship between different observing conditions and accuracy of AOD retrievals for each instrument. The integration stage combines AOD retrievals with different accuracy (categorized as accurate and inaccurate retrievals in the paper) from both instruments to determine a fusion plan. The fusion approach is tested by the collocated data among MODIS, CALIOP and AERONET from April 2, 2009 to April 1, 2011.

## 2 Data sets

To identify the conditions determining the AOD retrieval accuracy for MODIS and CALIOP, and to further evaluate the performance of our fusion approach, we analyze the collocated data among MODIS, CALIOP and AEROSOL ROBOTICS NETWORK (AERONET) from April 2, 2009 to April 1, 2011. Similar to previous validation research and our previous study, the AOD retrievals from MODIS, CALIOP and AERONET are compared in visible wavelength 550 nm [9, 15, 21–23]. Data products and spatial and temporal collocation requirements are outlined in the following paragraphs.

### 2.1 AERONET data

The AERONET measurements of AOD are relatively accurate (on the order of 0.02) when compared with satellite retrievals. As a result, they are widely used as "ground truth" for validation of satellite AOD retrievals [11, 24]. In this paper, Level 2.0 cloud-screened and quality-assured AERONET data is obtained for 199 sites globally between April 2, 2009 and April 1, 2011. These sites cover land, coast, desert and marine surface types. AERONET provides AOD retrievals in seven spectral bands (340, 380, 440, 500, 670, 870, and 1020 nm). As

AOD retrieval in 550 nm is not directly provided, it is interpolated by log-linearity from AOD values at 440 and 870 nm, as shown in our previous research [21–23].

### 2.2 MODIS data

MODIS/Aqua Collection 005 product suites are used in this study since Aqua is in A-Train constellation. The suites include a level 2 aerosol product MYD04\_L2 at a 10-km resolution, a level-1B sub-sampled calibrated radiance data MYD02SSH at a 5-km resolution and a level 2 cloud mask product MYD35 at a 1-km resolution. The three data sets with different spatial resolutions are synchronized in the spatial coincidence square region with an area of 40×40 km<sup>2</sup> surrounding an AERONET site [11, 13]. The region is required to contain at least one non-cloud pixel from MYD35 and at least one MODIS land AOD retrieval with quality flag QA ≥ 3 or at least one MODIS ocean AOD retrieval with QA ≥ 1 from MYD04\_L2.

### 2.3 CALIOP data

CALIOP level 2 version 3 cloud-free aerosol layer products are used in this study at a 5-km resolution around AERONET sites, with the following two additional settings: CALIOP cloud-aerosol discrimination (CAD) score < -20 and extinction quality control flag QC=0 for all layers [9, 19]. The two settings require the absence of clouds in the CALIOP data columns and consider only quality-controlled aerosol data. For keeping the same wavelength of AOD with MODIS and AERONET, we interpolated the CALIOP AOD retrievals at 550 nm by log-linearity from AOD values at 532 and 1064 nm.

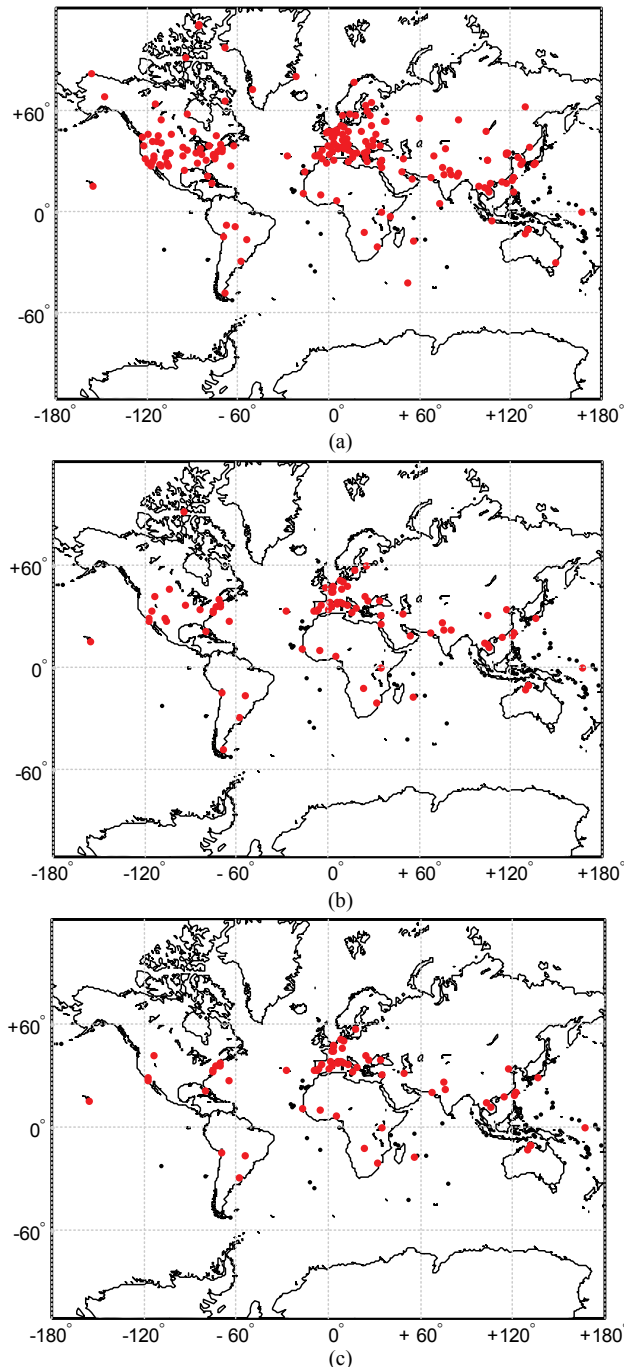
### 2.4 Collocated data

We collocated MODIS and AERONET data from April 2, 2009 to April 1, 2011, using the similar spatial-temporal coincidence criteria as Ichoku et al. [11]. The spatial coincidence box is approximately 40 km by 40 km, with the AERONET station at the centre covering a grid of 4 by 4 MODIS aerosol retrieval pixels. Spatial mean values for MODIS attributes are calculated and synchronized with the temporal mean values of the AERONET observations taken within ±30 minutes of MODIS overpass. By the coincidence criteria, the collocated data include 6351 overpasses covering 197 out of 199 AERONET sites globally. The spatial distribution of MODIS-AERONET collocated data is illustrated as red circles in Fig. 1a.

Using a very similar spatial-temporal coincidence criteria (a 40km by 40km square box and within ±30 minutes of CALIOP overpass), we collocated CALIOP and AERONET data from April 2, 2009 to April 1, 2011. A total number of 486 collocated data points are found covering 82 AERONET sites globally. Their spatial distribution is shown in Fig. 1b.

The MODIS-AERONET collocated data are matched with the CALIOP-AERONET collocated data employing the above spatial-temporal coincidence criteria. We obtained 322 MODIS-CALIOP-AERONET collocated

data covering 65 AERONET sites globally. Their spatial distribution is presented in Fig. 1c.



**Figure 1** (a) Spatial distribution of MODIS-AERONET collocated data, (b) Spatial distribution of CALIOP-AERONET collocated data, (c) Spatial distribution of MODIS-CALIOP-AERONET collocated data

### 3 Methodology

To integrate AOD retrievals from MODIS and CALIOP, we propose a two-stage fusion approach, including an analysis stage and an integration stage.

The goal of the analysis stage is to develop a systematic and data-driven approach to analyze the correlation between various conditions and AOD retrieval accuracy of MODIS and CALIOP instruments. In this study, the accuracy of the retrievals is simply categorized as two class: accurate and inaccurate AOD retrievals,

which are checked against an expected error (EE) envelope as,

$$EE = \pm(0.05 + 0.15 \times AERONET\_AOD) \quad (1)$$

It is a common standard used in AOD validation as shown in many studies [11, 13, 15]. Furthermore, an accurate retrieval is defined by an instrument AOD retrieval that falls within the EE envelope,

$$|AERONET\_AOD - Instrument\_AOD| \leq |EE| \quad (2)$$

Vice versa, AOD retrievals that fall outside the error box are defined as inaccurate retrievals. In this way, each collocated instrument AOD retrieval record will be categorized as two types of labels, either in EE envelope (class EE1) or outside of EE envelope (class EE0). Thereby, the data analysis can be considered as a classification problem: classifying conditions that determines accurate or inaccurate AOD retrievals from an instrument, given collocated AOD retrievals and their category labels.

Classification techniques can solve this problem. The output of a classification algorithm is the class label defined by Eq. (2). Its inputs are the collected information from a collocated data set.

For MODIS-AERONET collocated data, by our previous research, the following five types of attributes are potentially correlated with AOD retrievals and they are used as input-attribute candidates: 1) mean value and standard deviation of  $\rho$  MODIS AOD, its detected altitude and angstrom exponent (AE) in a spatial box; 2) mean value and standard deviation of surface reflectance at wavelength of 470 nm, 550 nm, 660 nm, 860 nm, 1240 nm, 1640 nm, 2120 nm, and their derived NDVI\_swir. The NDVI\_swir is defined as,

$$NDVI\_swir = \frac{\rho_{1240} - \rho_{2120}}{\rho_{1240} + \rho_{2120}} \quad (3)$$

3) observation geometry information including solar zenith, solar azimuth angle, sensor zenith, sensor azimuth angle, scattering angle; 4) percentage of cloud-free pixels in spatial box, percentage of pixels on a specific surface (water, coastal, desert, land) in spatial box; 5) observing temporal information (day, hour, minute) and spatial information (latitude, longitude). In addition, previous studies showed MODIS AOD retrieval accuracy varies in different areas [1]. With latitude and longitude information, we clustered 6351 MODIS-AERONET collocated points into 17 centres by using expectation maximization algorithm. The distance from each point to these 17 cluster centres is also measured as spatial attributes.

For CALIOP-AERONET collocated data, similar as the above collocated data, the following five types of attributes are used as input-attribute candidates: 1) mean value and standard deviation of CALIOP-AOD, its detected altitude and angstrom exponent (AE) in spatial box; 2) majority aerosol type across layers in scanning columns; 3) altitude, pressure, humidity for the base and top of the layer with the thickest and thinnest AOD; 4)

majority aerosol type (marine, desert dust, polluted dust, clean continental, polluted continental, biomass burning), cloud optical depth; 5) observing temporal information (day, hour, minute) and spatial information (latitude, longitude).

Both MODIS-AERONET and CALIOP-AERONET collocated data sets provide a large number of attributes. With limited training data size, it will result in over-fitting problems in decision tree classification. Therefore, we applied information gain measure to select the most informative attributes. The information gain for an attribute  $A$  in a data set  $S$  is defined as,

$$Infor - Gain(A) = Infor(A) - Infor_A(S) \tag{4}$$

Where,

$$Infor(A) = - \sum_{i=0}^1 p_i \log_2(p_i) \tag{5}$$

$$Infor_A(S) = \sum_{j=1}^m \frac{|S_j|}{|S|} \times Infor(S_j) \tag{6}$$

Here,  $p_i$  is the probability that a data record in  $S$  labeled as class  $EE_i$  ( $i = 0$  or  $1$ ) and is computed as  $|EE_i S|/|S|$ . For an element in  $A$ ,  $Infor(A)$  denotes the average amount of information required to identify its class label.  $m$  is the number of discrete values the attribute  $A$  has.  $S_j$  is the subset in  $S$  where attribute  $A$  is valued at  $a_j$ .

In the data mining domain, there are many classification techniques, such as support vector machine, neural networks, logistic regression, etc. Compared with these classifiers, a decision tree has the distinct advantage of constructing classification rules and these rules are easy for interpretation. Also a decision tree is a competitive classifier by achieving high accuracy in many applications. Therefore, we select decision tree techniques as classifiers to discover conditions where MODIS or CALIOP can retrieve AODs well.

There are multiple decision tree algorithms available. Specifically, we apply C4.5 algorithm, a widely-used decision tree method, in this research. It can deal with both nominal and numerical types of attributes.

In the research, by using five-folds cross validation, we iteratively explore combination of attributes and apply C4.5 decision tree algorithms to classify data. The selected attributes achieve the largest classification accuracy are retained.

The resulted C4.5 decision tree holds a flowchart-like structure (illustrated in Fig. 2), where each internal node (illustrated by rectangles) denotes a test on an attribute, each branch represents pass ways splitting by the test, and each leaf node (illustrated by ovals) holds a class category label. Each path from the root to a leaf node consists of a decision rule. A rule can be written as an "IF" part and a "Then" part. "IF" part consists of a combined condition, where each splitting criterion (such as specifying a data attribute in a value range) along a given path from the tree is logically "AND" connected. The "Then" part declares the class label for the records satisfying the combined condition in "IF" part. In the context of this study, the

class label is either ACCURATE or INACCURATE, indicating accurate AOD retrievals within  $EE$  envelope (class  $EE_i$ ) or inaccurate AOD retrievals outside of  $EE$  envelope (class  $EE_0$ ) respectively.

Ideally, each rule for a given class label covers many data records in a single class, rather than two classes. In another word, it is preferable to have a rule that produces an exact classification of records, i.e., to be pure. However, it is practically difficult to obtain pure rules in data with complex nature. Nevertheless, it is possible to obtain some rules containing a collection of records mostly from one class  $EE_i$  and just few from the other class  $EE_j$  ( $i, j = 0$  or  $1, i \neq j$ ). We use the following measure "confidence" to evaluate the accuracy of a rule,

$$confidence = \frac{|EE_i|}{|EE_i| + |EE_j|} \tag{7}$$

where,  $|EE_i|$  ( $i = 0$  or  $1$ ) denotes the number of records satisfying the combined condition in "IF" part and its corresponding class label is  $EE_i$  in a rule.

In addition, we compute measure "support" of a rule to calculate what is the percentage of records in all data set  $D$  that satisfy the rule,

$$support = \frac{|EE_i| + |EE_j|}{|D|} \tag{8}$$

Rules that satisfy both a minimum confidence threshold and a minimum support threshold are defined as strong rules.

In this way, we can obtain four sets of strong condition rules: 1) rule sets R1 where MODIS AOD retrievals are in class  $EE_1$ ; 2) rule sets R2 where MODIS AOD retrievals are in class  $EE_0$ ; 3) rule sets R3 where CALIOP AOD retrievals are in class  $EE_1$ ; 4) rule sets R4 where CALIOP AOD retrievals are in class  $EE_0$ .

Based on R1 and R2, MODIS AOD retrievals are categorized into three subsets: MA, MI and MU. MA includes accurate retrievals supported by strong rules R1. MI includes inaccurate retrievals supported by strong rules R2. The remaining retrievals as unknown instances in MODIS AOD (no strong rule supports) form the subset MU. In similar way, CALIOP AOD retrievals are categorized into subsets CA, CI and CU. CA covers accurate retrievals, CI covers inaccurate retrievals and CU covers remaining unknown retrievals.

In the fusion stage, a fusion plan is established according to definition of MA, MU, MI and CA, CU, CI, as shown in Tab. 1.

Table 1 Fusion plan

CA	CU	CI	CA
MA	Apply average of MA and CA	Apply MODIS retrieval MA	Apply MODIS retrieval MA
MU	Apply CALIOP retrieval CA	Apply average of MU and CU	Apply MODIS retrieval MU
MI	Apply CALIOP retrieval CA	Apply CALIOP retrieval CU	Apply average of MI and CI

In the MODIS-CALIOP-AERONET collocated data set, each record has both MODIS AOD retrieval and



CALIOP AOD retrieval. Each retrieval falls into one of three categories: accurate, inaccurate or unknown.

The MODIS AOD retrieval and CALIOP AOD retrieval are regarded as in the same level if they are in the same category – because they are measured by the same expected error envelope and supported by the same category of rules. Therefore, averaging is used as the fusion method for balancing for a robust predictor. For example, in the case of AOD retrievals in MI and CI, since the two instruments are in totally different design principles, they show different errors in different conditions, especially for outliers. Thereby, we choose their average as the final fusion results by smoothing the outlier error effects.

If MODIS AOD retrievals are accurate (MA) while CALIOP AOD retrievals are inaccurate (CI) or unknown (CU), we use MA as the fusion results. Similarly, if CALIOP AOD retrievals are inaccurate (CI) and MODIS AOD retrievals are unknown (MU), we would like to use MU as the fusion results by exploring the possibility that MODIS AOD retrievals might be accurate.

In the same way, the fusion operations can be conducted for situations where CALIOP AOD retrievals are accurate while MODIS AOD retrievals are not accurate, or situations where CALIOP AOD retrievals are unknown while MODIS AOD retrievals are inaccurate.

Based on the above two-stage approach, we obtain the final fusion results Fusion\_AOD. The results are evaluated on four measurements: mean absolute error (MAE), relative absolute bias (RAB), root mean square error (RMSE) and R-square (R<sup>2</sup>). They are defined as below,

$$MAE = \frac{\sum_i |Fusion\_AOD_i - AERONET\_AOD_i|}{N} \quad (9)$$

$$RAB = \frac{1}{N} \cdot \sum_i \frac{|Fusion\_AOD_i - AERONET\_AOD_i|}{AERONET\_AOD_i} \quad (10)$$

$$RMSE = \sqrt{\frac{\sum_i (Fusion\_AOD_i - AERONET\_AOD_i)^2}{N}} \quad (11)$$

$$R^2 = 1 - \frac{\sum_i (Fusion\_AOD_i - f(AERONET\_AOD_i))^2}{\sum_i (Fusion\_AOD_i - mean(Fusion\_AOD))^2} \quad (12)$$

#### 4 Experimental results and discussion

In the analysis stage of our experiments, we use the J48 classifier, which is a Java implementation of C4.5 decision tree algorithm in an open-source software Weka [25]. The confidence threshold is set to 80% and the support threshold is set to 0.5%. The parameter minNumObj in a J48 classifier controls the minimum number of instances per leaf in the decision tree. It is set to 50.

For MODIS-AERONET collocated data set, 14 attributes are finally selected in C4.5 decision tree classification. They are mean value of MODIS AOD retrievals (M\_AOD), standard deviation of M\_AOD (STD\_M\_AOD), NDVIs<sub>swir</sub>, surface reflectance at

wavelength 2.1 (Ref\_7), Scattering Angle (SA), the percentage of cloud free pixels (cloud\_free%), the percentage of coastal pixels (coastal%), the percentage of desert pixel (desert%), observation altitude (M\_Alt), standard deviation of altitude (STD\_M\_Alt), longitude (Lon), latitude (Lat), distance from a Northern Europe point with Lat=57.70 and Lon=21.47 (DNEP), distance to a Western Europe point with Lat=43.74 and Lon=6.54 (DWEP). MAE of all MODIS-AERONET collocated records is 0.066.

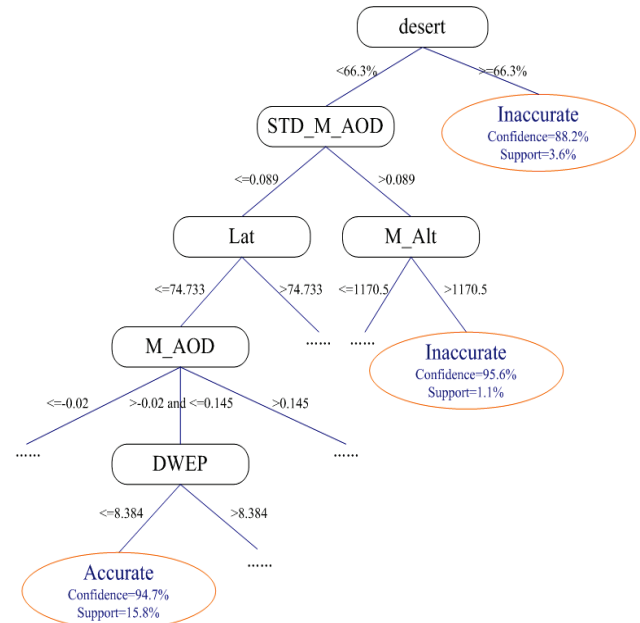


Figure 2 A portion of decision tree flowchart by analysis of MODIS-AERONET collocated data

Fig. 2 shows a portion of decision tree flowchart resulted from the MODIS-AERONET collocated data. It illustrates one accurate retrieval condition and two inaccurate retrieval conditions. The full decision tree flowchart is complex and cannot visualize clearly in the paper. In the full decision tree, the extracted five strong condition rules deriving accurate MODIS AOD retrievals are listed in Tab. 2.

The analysis of each rule is described as below,

MA1: If DWEP<=8.384 and desert%<=66.3% and M\_AOD<=0.145 and M\_AOD>-0.020 and STD\_M\_AOD<=0.089 and Lat<=74.733

Then ACCURATE (confidence=94.7%, support=15.8%, MAE=0.027)

The numbers in the bracket show the effectiveness of rule MA1 according to three measures. Support value shows that 15.8% of records in the MODIS\_AERONET collocated dataset satisfy the requirements of rule MA1. Confidence value shows that 94.7% MODIS AOD retrievals supported by rule MA1 are accurate retrievals. MAE presents the mean absolute error of MODIS AOD for all records supported by rule MA1.

The core combined condition of MA1 is DWEP<=8.384 and M\_AOD<=0.145 and M\_AOD>-0.020, because they cover almost the same set of records as in rule MA1.

Since DWEP represents the distance to a Western Europe point with Lat=43.74 and Lon=6.54, the rule MA1 suggests that in the Western Europe (with the clustering

center at DWEP), MODIS has accurate AOD retrievals on light aerosol loadings. It coincides with MODIS C005 validation results by previous studies [15].

MA2: If  $Ref\_7 \leq 0.0792$  and  $DWEP > 8.384$  and  $desert\% \leq 66.3\%$  and  $M\_AOD \leq 0.145$  and  $M\_AOD > -0.020$  and  $STD\_M\_AOD \leq 0.089$  and  $STD\_M\_Alt \leq 449.13$  and  $Lat \leq 74.733$

Then ACCURATE (confidence=93.1%, support=11.0%, MAE=0.029)

Condition  $Ref\_7 \leq 0.0792$  shows the surface is unusually dark. The rule MA2 suggests that in the case of light aerosol loading ( $M\_AOD \leq 0.145$ ), MODIS has accurate AOD retrievals in dark surface outside of desert area ( $desert\% \leq 66.3\%$ ). It proves the effectiveness of MODIS C005 algorithm on dark-surface land targets. In previous validation results [15], Levy et al. concluded that MODIS underestimates AOD (by 0.02 or more) where the surface is unusually dark ( $Ref\_7 < 0.05$ ). From this rule, we see that the underestimate shift has an upper bound and has high confidence within EE envelope.

**Table 2** Condition Rules Deriving accurate MODIS AOD Retrievals

ID	Conditions	Confidence	Support	MAE
MA1	DWEP ≤ 8.384 and desert% ≤ 66.3% and M_AOD ≤ 0.145 and M_AOD > -0.020 and STD_M_AOD ≤ 0.089 and Lat ≤ 74.733	94.7%	15.8%	0.027
MA2	Ref_7 ≤ 0.0792 and DWEP > 8.384 and desert% ≤ 66.3% and M_AOD ≤ 0.145 and M_AOD > -0.020 and STD_M_AOD ≤ 0.089 and STD_M_Alt ≤ 449.13 and Lat ≤ 74.733	93.1%	11%	0.029
MA3	Ref_7 > 0.0792 and SA ≤ 140.93 and DWEP > 8.384 and desert% ≤ 66.3% and M_AOD ≤ 0.145 and M_AOD > -0.020 and STD_M_AOD ≤ 0.089 and STD_M_Alt ≤ 449.13 and Lat ≤ 74.733	85.8%	12.0%	0.036
MA4	cloud_free% > 98.5% and NDVIswir > 0.323 and M_AOD > 0.145 and STD_M_AOD ≤ 0.089 and desert% ≤ 66.3% and Lon > -104.7 and Lat ≤ 61.846	91.8%	6.7%	0.040
MA5	cloud_free% ≤ 98.5% and Ref_7 < 0.008 and M_AOD > 0.145 and STD_M_AOD ≤ 0.089 and desert% ≤ 66.3% and Lon > -104.7 and Lat ≤ 61.846	84.0%	2.0%	0.046

MA3: If  $Ref\_7 > 0.0792$  and  $SA \leq 140.93$  and  $DWEP > 8.384$  and  $desert\% \leq 66.3\%$  and  $M\_AOD \leq 0.145$  and  $M\_AOD > -0.020$  and  $STD\_M\_AOD \leq 0.089$  and  $STD\_M\_Alt \leq 449.13$  and  $Lat \leq 74.733$

Then ACCURATE (confidence=85.8%, support=12.0%, MAE=0.036)

The rule MA3 suggests that if surfaces are not unusually dark ( $Ref\_7 > 0.0792$ ), scattering angle plays an important role on MODIS retrievals for areas outside of desert area ( $desert\% \leq 66.3\%$ ). In the combined

condition, small or medium scattering angles ( $SA \leq 140.93$ ) correspond to light aerosol loading points ( $M\_AOD \leq 0.145$ ). In the validation of C005 algorithm [15], Levy et al. show that larger AOD is associated with larger scattering angle and it verifies the rule MA3.

MA4: If  $cloud\_free\% > 98.5\%$  and  $NDVIswir > 0.323$  and  $M\_AOD > 0.145$  and  $STD\_M\_AOD \leq 0.089$  and  $desert\% \leq 66.3\%$  and  $Lon > -104.7$  and  $Lat \leq 61.846$

Then ACCURATE (confidence=91.8%, support=6.7%, MAE=0.040)

AERONET cloud screen algorithm has different design strategy from MODIS cloud screen algorithm. Thereby, the clear sky identified by AERONET site may contain cloud pixels detected by MODIS within a spatial range [15]. The condition  $cloud\_free\% > 98.5\%$  indicates a clear sky detected by both instruments MODIS and AERONET.

The rule MA4 suggests that in the cases of clear sky ( $cloud\_free\% > 98.5\%$ ), moderate/heavy aerosol loadings ( $M\_AOD > 0.145$ ), and green surface ( $NDVIswir > 0.323$ ), MODIS retrievals are accurate within homogeneous area ( $STD\_M\_AOD \leq 0.089$ ) where Longitude > -104.7 and Latitude ≤ 61.846. It coincides with Levy's validation results that the difference between MODIS and AERONET AODs is small when NDVIswir is around 0.4 [15]. MA4 further lists the detailed combined condition for accurate retrievals in green surface.

**Table 3** Condition Rules Deriving Inaccurate MODIS AOD Retrievals

ID	Conditions	Confidence	Support	MAE
MI1	desert% > 66.3%	88.2%	3.6%	0.221
MI2	desert% ≤ 66.3% and M_Alt > 1170.5 and STD_M_AOD > 0.089	95.6%	1.1%	0.203
MI3	desert% ≤ 66.3% and M_Alt ≤ 1170.5 and STD_M_AOD > 0.191	82.4%	1.2%	0.258
MI4	M_AOD ≤ -0.020 and desert% ≤ 66.3% and STD_M_AOD ≤ 0.089 and Lat ≤ 74.733	92.2%	1.6%	0.087
MI5	Lat > 74.733 and desert% ≤ 66.3% and STD_M_AOD ≤ 0.089	100%	0.6%	0.180

MA5: If  $cloud\_free\% \leq 98.5\%$  and  $Ref\_7 \leq 0.008$  and  $M\_AOD > 0.145$  and  $STD\_M\_AOD \leq 0.089$  and  $desert\% \leq 66.3\%$  and  $Lon > -104.7$  and  $Lat \leq 61.846$

Then ACCURATE (confidence=84.0%, support=2.0%, MAE=0.046)

The condition  $cloud\_free\% \leq 98.5\%$  shows retrieval records with a bit of cloudy detected by MODIS. This rule suggests that in the cases of cloudy, extremely dark surface ( $Ref\_7 \leq 0.008$ ) and moderate/heavy aerosol loadings ( $M\_AOD > 0.145$ ), MODIS AOD retrievals are accurate. It verifies the effectiveness of MODIS second-generation AOD retrieval algorithm about its assumptions of the surface optical characteristics.

In the decision tree, the extracted five strong condition rules deriving inaccurate MODIS AOD retrievals are listed in Tab. 3.

The detailed analysis of each rule is listed as below,

MI1: If  $desert\% > 66.3\%$

Then INACCURATE (confidence=88.2%, support=3.6%, MAE=0.221)

This rule MI1 suggests MODIS has inaccurate AOD retrievals on desert ( $\text{desert}\% > 66.3\%$ ). Desert has a relatively brighter surface than dark-targets where aerosol signals are comparatively weak for MODIS retrieval. We investigated the details and found under the condition, 90.8% of supported records overestimate AOD. It agrees with Levy et al's validation results [15] that MODIS poorly overestimates AOD retrievals on bright surface.

MI2: If  $\text{desert}\% \leq 66.3\%$  and  $M\_Alt > 1170.5$  and  $STD\_M\_AOD > 0.089$

Then INACCURATE (confidence=95.6%, support=1.1%, MAE=0.203)

This rule shows MODIS has inaccurate AOD retrievals on heterogeneous areas ( $STD\_M\_AOD > 0.089$ ) which are in the high altitudes ( $M\_Alt > 1170.5$ ) and are outside of deserts ( $\text{desert}\% \leq 66.3\%$ ). In Levy's validation results [15], they showed MODIS compare poorly with AERONET on elevated targets, such as on plateau, which has a relative brighter surface.

MI3: If  $\text{desert}\% \leq 66.3\%$  and  $M\_Alt \leq 1170.5$  and  $STD\_M\_AOD > 0.191$

Then INACCURATE (confidence=82.4%, support=1.2%, MAE=0.258)

This rule MI3 shows MODIS has inaccurate AOD retrievals on extreme heterogeneous areas ( $STD\_M\_AOD > 0.191$ ) which are in the low altitude ( $M\_Alt \leq 1170.5$ ) and are outside of deserts ( $\text{desert}\% \leq 66.3\%$ ).

By combining MI2 and MI3, we can derive a new rule CMI straightforward:

CMI: If  $\text{desert}\% \leq 66.3\%$  and  $STD\_M\_AOD > 0.191$

Then INACCURATE (confidence=84.4%, support=1.4%, MAE=0.264)

This rule shows that in heterogeneous areas with extreme complex nature ( $STD\_M\_AOD > 0.191$ ), either in low altitudes or high altitudes, AODs varies greatly in a spatial box and they lead to large standard deviation of MODIS AOD. In this case, it cannot match well with AERONET AOD.

MI4: If  $M\_AOD \leq -0.020$  and  $\text{desert}\% \leq 66.3\%$  and  $STD\_M\_AOD \leq 0.089$  and  $Lat \leq 74.733$

Then INACCURATE (confidence=92.2%, support=1.6%, MAE=0.087)

Under the combined condition, the average AERONET AOD of all covered records is 0.052, whereas MODIS AOD retrievals of all these covered records are negative ( $M\_AOD \leq -0.020$ ). This rule shows MODIS underestimates AODs in light aerosol loadings and permits negative values. It coincides with Levy's analysis [15].

MI5: If  $Lat > 74.733$  and  $\text{desert}\% \leq 66.3\%$  and  $STD\_M\_AOD \leq 0.089$

Then INACCURATE (confidence=100%, support=0.6%, MAE=0.180)

This rule shows MODIS has inaccurate retrievals on areas in high latitude ( $Lat > 74.733$ ). These areas are near to poles, mostly covered by ice and leading to a brighter surface. A bright surface will make the aerosol signals are comparatively weak for MODIS retrieval.

For CALIOP-AERONET collocated data set, 7 attributes are selected in C4.5 decision tree classification. They are cloud flag (C\_Cloud), mean value of CALIOP AOD retrievals (C\_AOD), standard deviation of C\_AOD

(STD\_C\_AOD), the thickest aerosol layer type in a scanning column (MAX\_AOD\_TYPE), the thinnest aerosol layer type in a scanning column (MIN\_AOD\_TYPE), the altitude of the base for the thickest aerosol layer (MAX\_AOD\_LBA), the temperature in the middle level of the thickest aerosol layer (MAX\_AOD\_LMT). The parameter minNumObj in J48 is set to 15. MAE of all CALIOP-AERONET collocated records is 0.106.

In the decision tree, the extracted two strong condition rules deriving accurate CALIOP AOD retrievals are listed in Table IV.

The detailed explanation and analysis for each rule is listed as below,

CA1: If C\_Cloud=NO and MAX\_AOD\_LBA  $\leq 2.119$  and MAX\_AOD\_TYPE=POLLUTED\_CONTINENTAL and STD\_C\_AOD  $\leq 0.210$

Then ACCURATE (confidence=80.7%, support=11.7%, MAE=0.043)

The condition MAX\_AOD\_LBA  $\leq 2.119$  indicates the detected thickest layer throughout a column is not high, or say the thickest layer is in the reasonable height of aerosol loadings. The rule suggests that in the case of clear sky (C\_Cloud=No), the aerosol type for the thickest layer is polluted continental (MAX\_AOD\_TYPE=POLLUTED\_CONTINENTAL), reasonable height of aerosol loadings (MAX\_AOD\_LBA  $\leq 2.119$ ) and homogeneous area (STD\_C\_AOD  $\leq 0.210$ ), CALIOP has accurate retrievals. It coincides with the validation results from Schuster [9].

**Table 4** Condition Rules Deriving Accurate CALIOP AOD Retrievals

ID	Conditions	Confidence	Support	MAE
CA1	C_Cloud=NO and MAX_AOD_LBA $\leq 2.119$ and MAX_AOD_TYPE=POLLUTED_CONTINENTAL and STD_C_AOD $\leq 0.210$	80.7%	11.7%	0.043
CA2	C_Cloud=NO and MAX_AOD_LBA $\leq 2.119$ and MAX_AOD_TYPE=MARINE and STD_C_AOD $\leq 0.210$	84.4%	9.3%	0.042

CA2: If C\_Cloud=NO and MAX\_AOD\_LBA  $\leq 2.119$  and MAX\_AOD\_TYPE=MARINE and STD\_C\_AOD  $\leq 0.210$

Then ACCURATE (confidence=84.4%, support=9.3%, MAE=0.042)

The rule suggests that in the case of clear sky (C\_Cloud=No), the aerosol type for the thickest layer is marine (MAX\_AOD\_TYPE=MARINE), reasonable height of aerosol loadings (MAX\_AOD\_LBA  $\leq 2.119$ ) and homogeneous area (STD\_C\_AOD  $\leq 0.210$ ), CALIOP has accurate retrievals. Under the combined condition, 84.4% of satisfied records are accurate retrievals and MAE of these accurate retrievals are only 0.025. We check in the cloudy and marine aerosol situation (C\_Cloud=YES and MAX\_AOD\_TYPE=MARINE), MAE of the corresponding CALIOP retrievals is 0.171, far larger than the results in clear sky. It illustrates that for the same type of marine aerosols, cloud condition makes a big difference on CALIOP retrievals.

The three strong condition rules deriving inaccurate CALIOP AOD retrievals are listed in Tab. 5.

The detailed explanation and analysis for each rule is listed as below,

C11: If  $MAX\_AOD\_LBA > 2.119$

Then INACCURATE (confidence=87.5%, support=9.9%, MAE =0.220)

**Table 5** Condition Rules Deriving Inaccurate CALIOP AOD Retrievals

ID	Conditions	Confidence	Support	MAE
C11	$MAX\_AOD\_LBA > 2.119$	87.5%	9.9%	0.220
C12	$MAX\_AOD\_LBA \leq 2.119$ and $STD\_C\_AOD > 0.210$ and $C\_Cloud = YES$ and $MIN\_AOD\_TYPE = DESERT$	81.8%	4.5%	0.266
C13	and $MAX\_AOD\_LBA \leq 2.119$ and $STD\_C\_AOD > 0.024$ and $STD\_C\_AOD \leq 0.210$	82.4%	3.5%	0.280

The rule suggests that if the detected thickest layer throughout a column is not in the reasonable height for aerosol loadings, CALIOP has inaccurate retrievals. In the condition, 87.5% of covered retrievals are inaccurate and the mean error of (AERONET AOD - CALIOP AOD) equals 0.188, suggesting that most CALIOP retrievals miss some aerosol layers throughout a column.

C12: If  $MAX\_AOD\_LBA \leq 2.119$  and  $STD\_C\_AOD > 0.210$

Then INACCURATE (confidence=81.8%, support=4.5%, MAE=0.266)

The rule suggests that if detected aerosols are in the reasonable height, but in the heterogeneous areas ( $STD\_C\_AOD > 0.210$ ), CALIOP has inaccurate retrievals.

C13: If  $C\_Cloud = YES$  and  $MIN\_AOD\_TYPE = DESERT$  and  $MAX\_AOD\_LBA \leq 2.119$  and  $STD\_C\_AOD > 0.024$  and  $STD\_C\_AOD \leq 0.210$

Then INACCURATE (confidence=82.4%, support=3.5%, MAE=0.280)

The rule suggests in the cases of cloudy sky ( $C\_Cloud = YES$ ), the aerosol type for the thinnest layer is DESERT ( $MIN\_AOD\_TYPE = DESERT$ ), reasonable height of aerosol loadings ( $MAX\_AOD\_LBA \leq 2.119$ ), and moderately heterogeneous areas ( $0.024 < STD\_C\_AOD \leq 0.210$ ), CALIOP has inaccurate retrievals.

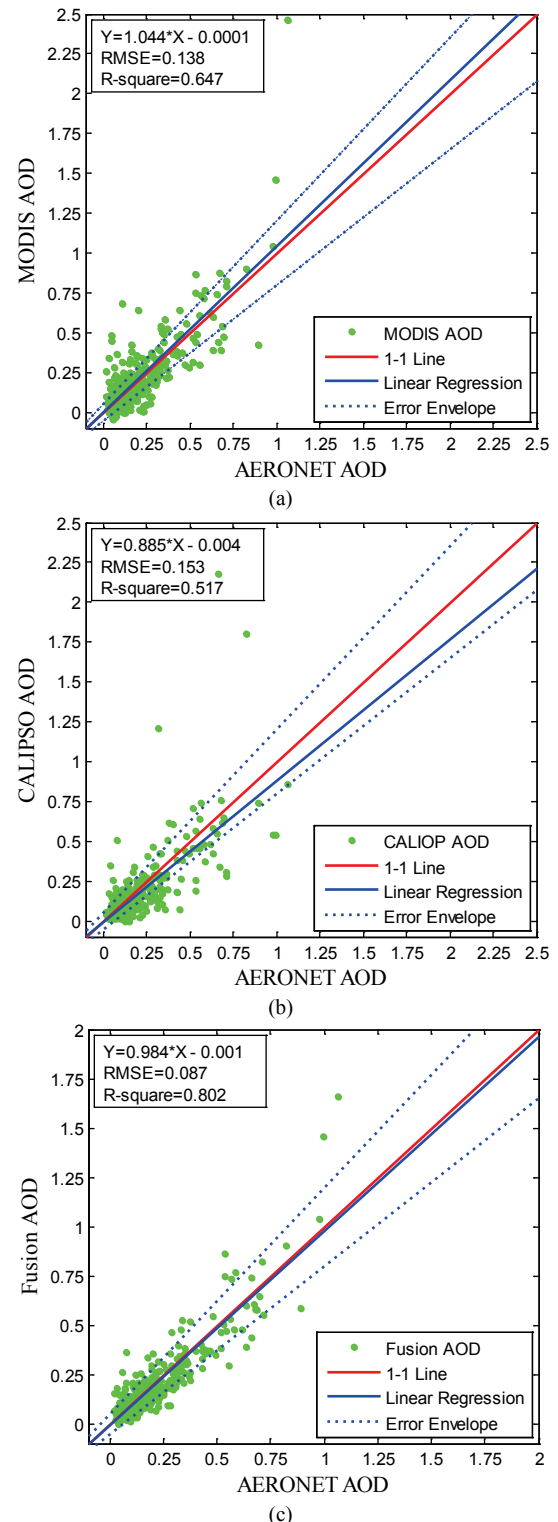
Based on the above rules, we categorize the MODIS retrievals from MODIS-CALIOP-AERONET collocated data into three subsets: MA, MU and MI; and categorize the CALIOP retrievals from MODIS-CALIOP-AERONET collocated data into three subsets as well: CA, CU and CI.

Next, we apply the fusion plan presented in Table 1 on MODIS-CALIOP-AERONET collocated data. The final fusion results are presented in Tab. 6.

In Tab. 6, we see that by using measures MAE, RAB, RMSE,  $R^2$ , the fusion results all are significantly better than either MODIS AOD retrievals or CALIOP AOD retrievals. For example, the root mean square error (RMSE) for MODIS and CALIOP are 0.138 and 0.153 respectively. The fusion result has been significantly decreased to 0.087. For  $R^2$  measure, fusion result has achieved to 0.802 from 0.647 for MODIS retrievals and 0.517 for CALIOP retrievals.

**Table 6** Compare Fusion Results with MODIS/CALIOP AOD Retrievals

	MAE	RAB	RMSE	$R^2$
MODIS AOD	0.081	39.25%	0.138	0.647
CALIOP AOD	0.086	41.60%	0.153	0.517
Fusion	0.059	28.23%	0.087	0.802



**Figure 3** (a) Comparison of MODIS AOD retrievals with AERONET AOD retrievals ( $R^2=0.647$ ) (b) Comparison of CALIOP AOD retrievals with AERONET AOD retrievals ( $R^2=0.517$ ) (c) Comparison of fusion results with AERONET AOD retrievals ( $R^2=0.802$ )

To intuitively compare MODIS AOD retrievals, CALIOP AOD retrievals and our fusion results, we visualize the collocated data points in Fig. 3a ÷ 3c.



From Fig. 3a ÷ 3c, we see that compared with the points from MODIS or CALIOP AOD retrievals, the points from fusion results are more closely clustered near the 1-1 line and the linear regression line has the smallest deviation from the 1-1 line. It illustrates the fusion results are more accurate than MODIS or CALIOP retrievals by comparing with AERONET AODs. This proves the effectiveness of our fusion approach.

We also compared MODIS AOD retrievals and CALIOP AOD retrievals in the experiment with results from others' previous study. Levy et al. compared MODIS C005 AODs over dark land with AERONET collocations at 550 nm wavelength, the result shows RMSE=0.116,  $R^2=0.778$  (corresponding to correlation coefficient  $R=0.882$ ) [15]. But their collocation points only cover dark targets. They noted that MODIS is comparatively weak for brighter surface. Our experiments include collocation points on bright deserts. It explains the reason why our result achieves RMSE = 0.138 and  $R^2 = 0.647$ , relatively worse than theirs. Omar et al. performed an analysis of the correlations between cloud-cleared CALIOP 532 nm and AERONET 500 nm AOD over 92 land sites and 57 coastal sites [8]. The 613 land collocations achieves  $R^2 = 0.348$  (corresponding to correlation coefficient  $R = 0.59$ ) and the 468 coastal collocations achieves  $R^2 = 0.292$  (corresponding to correlation coefficient  $R = 0.54$ ). They compared CALIOP and AERONET AOD retrievals in different wavelength and also their selected collocated sites are different from ours. These reasons may explain the difference. Especially, their result shows that the differences between the CALIOP AOD retrievals and the AERONET AOD measurements are independent of the surface type. It is helpful to explain in some surface type condition (such as desert), CALIOP retrieves AODs well but MODIS does not.

## 5 Conclusion

MODIS and CALIOP are two independent observation instruments in the A-train satellite constellation. They both provide aerosol retrievals at nearly the same locations on the Earth's surface with a two-minute interval. Due to the different design principles of the two observation instrument, their accuracy of AOD retrievals varies under different conditions. In this work, a two-stage fusion approach is proposed, including an analysis stage and an integration stage, to combine AOD retrievals with different accuracy from each instrument together in order to improve AOD retrieval accuracy. In the analysis stage, a data-driven decision-tree algorithm is used to systematically analyze the relationship between the observation conditions and accuracy of AOD retrievals for instrument MODIS and CALIOP respectively. In the integration stage, based on the discovered condition rules, we combine AOD retrievals from both instruments together to form a fusion plan. We test the fusion approach on the collocated data among MODIS, CALIOP and AERONET from April 2, 2009 to April 1, 2011. From decision-tree analysis, we obtain five accurate retrieval conditions and five inaccurate retrieval conditions for MODIS, and two accurate retrieval conditions and three inaccurate retrieval conditions for

CALIOP. The final fusion result achieves AOD retrieval accuracy with  $R^2 = 0.802$  and RMSE = 0.087, which is significantly more accurate than AOD retrievals from any single observation facility (For MODIS retrievals,  $R^2 = 0.647$ , RMSE = 0.138, and for CALIOP retrievals,  $R^2 = 0.517$ , RMSE = 0.133).

In our future research, we will investigate more interesting topics by fusion of MODIS and CALIOP data. MODIS and CALIOP are designed with different principles and their observed information can complement with one another. We will test if the complement information will bring new insights for AOD retrievals for some specific areas, such as urban areas, high altitude areas, ice surface areas or in cloud fields. In addition, it will be interesting to explore the fusion information from MODIS and CALIOP will be helpful for air quality forecast applications.

Meanwhile, in the next step, we would like to apply our fusion approach to other remote sensor combination. Except that other sensors collect different observation attributes from MODIS and CALIOP and they require different data pre-processing, we can apply the same decision tree techniques to discover the accurate retrieval conditions for each sensor in the analysis stage and similarly form an integration plan in the fusion stage.

## Acknowledgements

This work was supported by the National Natural Science Foundation of China under Grants U1531122 and 61272272, and by the Natural Science Foundation of Hubei province under Grant 2015CFA058. We highly appreciate the support from NASA Goddard Space Flight Centre for the MODIS data, NASA Langley Research Centre for CALIOP data, and AERONET science team for AERONET data.

## 6 References

- [1] Levy, R. C.; Remer, L. A.; Mattoo, S.; Vermote, E.; Kaufman, Y. J. Second generation algorithm for retrieving aerosol properties over land from MODIS spectral reflectance. // *J. Geophys. Res.* Vol. 112, No. D13211, Jul. 2007. DOI: 10.1029/2006JD007811
- [2] Ichoku, C.; Chu, D. A.; Mattoo, S.; Kaufman, Y. J.; Remer, L. A.; Tanré, D.; Slutsker, I.; Holben, B. N. A spatio-temporal approach for global validation and analysis of MODIS aerosol products. // *Geophys. Res. Lett.*, vol. 29, no. 12, 8006, Jun. 2002. DOI: 10.1029/2001GL013206
- [3] Levy, R. C.; Remer, L. A.; Martins, J. V.; Kaufman, Y. J. Evaluation of the MODIS aerosol retrievals over ocean and land during CLAMS. // *J. Atmos. Sci.* 62, 4(2005), pp. 974-992. DOI: 10.1175/JAS3391.1
- [4] Ichoku, C.; Remer, L. A.; Kaufman, Y. J.; Levy, R.; Chu, D. A.; Tanre, D.; Holben, B. N. MODIS observation of aerosols and estimation of aerosol radiative forcing over southern Africa during SAFARI 2000. // *J. Geophys. Res.* Vol. 108, No. D13, 8499, July. 2003. DOI: 10.1029/2002JD002366
- [5] Abdou, W. A.; Diner, D. J.; Martonchik, J. V.; Bruegge, C. J.; Kahn, R. A.; Gaitley, B. J.; Crean, K. A.; Remer, L. A.; Holben, B. Comparison of coincident MISR and MODIS aerosol optical depths over land and oceans scenes containing AERONET sites. // *J. Geophys. Res.* Vol. 110, No. D10 D10S07, May. 2005. DOI: 10.1029/2004JD004693

- [6] Tripathi, S. N.; Dey, S.; Chandel, A.; Srivastava, S.; Singh, R. P.; Holben, B. N. Comparison of MODIS and AERONET derived aerosol optical depth over the Ganga Basin, India. // *Ann. Geophys.* 23, 4(2005), pp. 1093-1101. DOI: 10.5194/angeo-23-1093-2005
- [7] Chylek, P.; Henderson, B. G.; Lesins, G. Aerosol Optical Depth Retrieval over the NASA Stennis Space Center: MTI, MODIS, and AERONET. // *IEEE Trans. on Geos. and Remote Sens.* 43, 9(2005), pp. 1978-1983. DOI: 10.1109/TGRS.2005.851773
- [8] Omar, A. H.; Winker, D. M.; Tackett, J. L.; Giles, D. M.; Kar, J.; Liu, Z.; Vaughan, M. A.; Powell, K. A.; Trepte, C. R. CALIOP and AERONET aerosol optical depth comparisons: One size fits none. // *J. Geophys. Res. Atmos.* 118, 10(2013), pp. 4748-4766. DOI: 10.1002/jgrd.50330
- [9] Schuster, G. L.; Vaughan, M.; MacDonnell, D. et al. Comparison of CALIPSO aerosol optical depth retrievals to AERONET measurements, and a climatology for the LIDAR ratio of dust. // *J. Atmos. Chem. Phys.* 12, 16(2012), pp. 7431-7452. DOI: 10.5194/acp-12-7431-2012
- [10] Kacenelenbogen, M.; Vaughan, M. A.; Redemann, J.; Hoff, R. M.; Rogers, R. R.; Ferrare, R. A.; Russell, P. B.; Hostetler, C. A.; Hair, J. W.; Holben, B. N. An accuracy assessment of the CALIOP/CALIPSO version 2/version 3 daytime aerosol extinction product based on a detailed multi-sensor, multi-platform case study. // *J. Geophys. Res. Atmos.* 11, 8(2011), pp. 3981-4000. DOI: 10.5194/acp-11-3981-2011
- [11] Chu, D. A.; Kaufman, Y. J.; Ichoku, C.; Remer, L. A.; Tanré, D.; Holben, B. N. Validation of MODIS aerosol optical depth retrieval over land. // *Geophys. Res. Lett.* 29, 12(2002), pp. 1617-1621. DOI: 10.1029/2001GL013205
- [12] Chu, D. A.; Holben, B. N.; Dubovik, O.; Smirnov, A.; Martins, J. V.; Li, R. R.; Ahmad, Z. Validation of MODIS aerosol retrieval over ocean. // *Geophys. Res. Lett.* 29, 12(2002). DOI: 10.1029/2001GL013205
- [13] Remer, L. A.; Kaufman, Y. J.; Tanre, D.; Mattoo, S.; Chu, D. A.; Martins, J. V.; Li, R.-R.; Ichoku, C.; Levy, R. C.; Kleidman, R. G.; Eck, T. F.; Vermote, E.; Holben, B. N. The MODIS aerosol algorithm, products and validation. // *J. Atmos. Sci.* 62, 4(2005), pp. 947-973. DOI: 10.1175/JAS3385.1
- [14] Levy, R. C.; Remer, L. A.; Dubovik, O. Global aerosol optical properties and application to Moderate Resolution Imaging Spectroradiometer aerosol retrieval over land. // *J. Geophys. Res. Vol.* 112, No. D13, D13210, Jul. 2007. DOI: 10.1029/2006JD007815
- [15] Levy, R. C.; Remer, L. A.; Kleidman, R. G.; Mattoo, S.; Ichoku, C.; Kahn, R.; Eck, T. F. Global evaluation of the Collection 5 MODIS dark-target aerosol products over land. // *J. Atmos. Chem. Phys.* 10, 6(2010), pp. 14815-14873. DOI: 10.5194/acp-10-10399-2010
- [16] Jethva, H.; Satheesh, S. K.; Srinivasan, J. Assessment of second generation MODIS aerosol retrieval (Collection 005) at Kanpur, India. // *J. Geophys. Res. Lett.* Vol. 34, No. 19, L19802, Oct. 2007. DOI: 10.1029/2007GL029647
- [17] Winker, D. M.; Pelon, J.; Coakley, J. A.; Ackerman, S. A. et al. The CALIPSO Mission: A Global 3D View of Aerosols and Clouds. // *Bull. Amer. Meteor. Soc.* 91, 9(2010), pp. 1211-1229. DOI: 10.1175/2010BAMS3009.1
- [18] Winker, D. M.; Vaughan, M. A.; Omar, A.; Hu, Y.; Powell, K. A.; Liu, Z.; Hunt, W. H.; Young, S. A. Overview of the CALIPSO Mission and CALIOP Data Processing Algorithms. // *J. Atmos. Oceanic Technol.* 26, 11(2009), pp. 2310-2323. DOI: 10.1175/2009JTECHA1281.1
- [19] Young, S. A.; Vaughan, M. A. The Retrieval of Profiles of Particulate Extinction from Cloud-Aerosol Lidar Infrared Pathfinder Satellite Observations (CALIPSO) Data: Algorithm Description. // *J. Atmos. Oceanic Technol.* 26, 6(2009), pp. 1105-1119. DOI: 10.1175/2008JTECHA1221.1
- [20] Liu, Z.; Vaughan, M.; Winker, D.; Kittaka, C.; Getzewich, B.; Kuehn, R.; Omar, A.; Powell, K.; Trepte, C.; Hosterler, C. The CALIPSO Lidar Cloud and Aerosol Discrimination: Version 2 Algorithm and Initial Assessment of Performance. // *J. Atmos. Oceanic Technol.* 26, 7(2009), pp. 1198-1213, DOI: 10.1175/2009JTECHA1229.1
- [21] Vucetic, S.; Han, B.; Mi, W.; Li, Z.; Obradovic, Z. A Data-Mining Approach for the Validation of Aerosol Retrievals. // *IEEE Geosci. Remote Sens. Lett.* 5, 1(2008), pp. 113-117. DOI: 10.1109/LGRS.2007.912725
- [22] Han, B.; Vucetic, S.; Braverman, A.; Obradovic, Z. A Statistical Complement to Deterministic Algorithms for the Retrieval of Aerosol Optical Thickness from Radiance Data. // *Engineering Applications of Artificial Intelligence.* 19, 7(2006), pp. 787-795. DOI: 10.1016/j.engappai.2006.05.009
- [23] Xu, Q.; Han, B.; Li, Y.; Braverman, A.; Obradovic, Z.; Vucetic, S. Improving Aerosol Retrieval Performance by Integrating AERONET, MISR, and MODIS Data Products. // 8th Int'l Conf. on Information Fusion, Philadelphia, PA, 2005.
- [24] Holben, B. N.; Eck, T. F.; Slutsker, I.; Tanre, T.; Buis, J. P.; Setzer, A.; Vermote, E.; Reagan, J. A.; Kaufman, Y. J.; Nakajima, T.; Lavenu, F.; Jankowiak, I.; Smirnov, A. AERONET: A federated instrument network and data archive for aerosol characterization. // *Remote Sens. Environ.* 37, (1998), pp. 2403-2412. DOI: 10.1016/s0034-4257(98)00031-5
- [25] Hall, M.; Frank, E.; Holmes, G.; Pfahringer, B.; Reutemann, P.; Witten, I. The WEKA data mining software: an update. // *SIGKDD Explorations.* 11, 1(2009). DOI: 10.1145/1656274.1656278

#### Authors' addresses

**Bo Han, Associate Professor**

**Hongpeng Ding, Graduate Student**

International School of Software, Wuhan University,

37 Luoyu Road, Wuhan, China

E-mail: bhhan@whu.edu.cn

E-mail: 2010282160014@whu.edu.cn

**Yingying Ma, Associate Professor**

**Wei Gong, Professor**

The State Key Laboratory of Information Engineering in

Surveying, Mapping, and Remote Sensing, Wuhan University,

37 Luoyu Road, Wuhan, China

E-mail: yym863@whu.edu.cn

E-mail: weigong@lmars.whu.edu.cn