Int J Appl Earth Obs Geoinformation 73 (2018) 721–735

Contents lists available at ScienceDirect

Int J Appl Earth Obs Geoinformation

journal homepage: www.elsevier.com/locate/jag

Atmospheric scene classification using CALIPSO spaceborne lidar measurements in the Middle East and North Africa (MENA), and India

Foad Brakhasi[a,*], Aliakbar Matkan[a], Mohammad Hajeb[a], Kourosh Khoshelham[b]

[a] Remote Sensing and GIS Center, University of Shahid Beheshti, Tehran, Iran
[b] Department of Infrastructure Engineering, The University of Melbourne, Parkville, 3010, Australia

ARTICLE INFO

Keywords:
Aerosol
CALIOP
SVM
MENA
Classification
VFM

ABSTRACT

This paper presents a new algorithm based on the support vector machine (SVM) for classifying the Cloud-Aerosol Lidar Infrared Pathfinder Satellite Observations (CALIPSO) data into classes of clean air, cloud, thin aerosol, dense aerosol, surface, subsurface and totally attenuated. The procedure is as follows: At first, the considered features based on CALIPSO data are prepared. Brightness Temperature Differences between 10 and 12 μm ($BTD_{11\text{-}12}$) is then used to better discriminate dense aerosols from clouds. The particle density feature proposed in this research is another feature participating in the classification. Training samples are automatically extracted by applying strict thresholds on the features. A wrapper feature selection is performed to rank the features based on their performance. Four post-processing steps are implemented to correct some misclassified cells e.g. edges of clouds and high-level clouds. The proposed algorithm was implemented on 4 datasets in the Middle East and North Africa (MENA), and India with various types and densities of aerosol. An accuracy assessment based on the comparison between the obtained results and ground truth samples indicated 0.94, 0.96 4, 0.92 and 0.89 kappa coefficients for the datasets. A statistical hypothesis test demonstrated that our SVM classification overcame CALIPSO vertical feature mask (VFM) product. The experimental result indicates the high accuracy of the proposed algorithm for the atmosphere scene classification using CALIPSO data.

1. Introduction

Aerosols are solid particles or suspended liquid in the atmosphere (Wong et al., 2013), which are produced as a result of natural activities such as volcanoes, storms as well as human activities including burning fossil fuels and traffic (Wiltshire, 2011; Gong and Ma, 2012; Kokkalis et al., 2012; Zhu et al., 2016). These particles are pumped into the troposphere by convection systems, and last from several days to several weeks (Haywood and Boucher, 2000; Yu et al., 2012). Aerosols have had significant impacts on ecosystem health and humans, land and sea, biogeochemical cycles, characteristics of clouds, as well as weather systems and climate, especially in recent decades (Feng and Christopher, 2014; Basha et al., 2015; Yu et al., 2015; Lee et al., 2016). Awareness of the uncertainties stemmed from aerosols as the most important sources in the studies of climate systems and climate change is significant (Banks and Brindley, 2013; Tomasi et al., 2015). Aerosols and suspended particles in the atmosphere resulted in the reduction of accuracy in remote sensing and its difference with ground-based measurements, therefore in the process of atmospheric correction, it is necessary to consider the concentration of aerosol in the region. Studies have shown that there are significant differences in the spatial and temporal distribution of aerosols (Tomasi et al., 2015), so monitoring and estimating the temporal and spatial variations of aerosols is very important (Ma and Gong, 2012). The location of Middle East in the desert zones of the earth which is the main sources of aerosols caused a large number of dust storms, and it has changed the atmospheric composition. (Abdi Vishkaee et al., 2011, 2012; Shahsavani et al., 2012; Rezazadeh et al., 2013; Alam et al., 2014; Basha et al., 2015; Parajuli et al., 2016; Hermida et al., 2018).

There are several methods for studying aerosols, including ground-based measurements (Welton et al., 2000; Holben et al., 2001; Murayama et al., 2001; Chung et al., 2005; He and Yi, 2015; Han et al., 2015; Ningombam et al., 2014; Vijayakumar et al., 2016; Campbell et al., 2016), aerosols models (McMurry, 2000; Nowottnick et al., 2015; Geng et al., 2015; Vijayakumar et al., 2016; Prijith et al., 2016) and remote sensing (Winker et al., 2007; Bréon et al., 2011; Knippertz and Todd, 2012; Kokhanovsky, 2013; Feng and Christopher, 2014). Ground-Based Measurement is a common method for aerosols monitoring which is done discretely and pointwise on ground stations. This method is not able to show the temporal and spatial variations of aerosols due to the

* Corresponding author.
E-mail address: brakhasi.f@gmail.com (F. Brakhasi).

https://doi.org/10.1016/j.jag.2018.07.017
Received 11 March 2018; Received in revised form 14 July 2018; Accepted 15 July 2018

point of being, especially in areas with low station numbers. AERONET (Aerosol Robotic Network) is a global network of devices for measuring aerosols which uses the Sun photometer passive instrument which is widely used to evaluate satellite observations and simulation models (Lin et al., 2014; Geng et al., 2015). Despite the proper accuracy of AERONET data, a major problem is low-density stations (Ma et al., 2011; Bibi et al., 2015). Another method is aerosol models (Nowottnick et al., 2015; Prijith et al., 2016). These models often have a lot of inputs and the results of these models are also associated with uncertainty (Yu et al., 2015). These models show only specific layers in vertical resolution and do not take into account phenomenon such as sea breeze and convection which are of a moderate scale. In addition, the performance of meteorological models is low when the area is small (McMurry, 2000).Remote sensing technology is known as one of the most effective techniques for the simultaneous study of climate and atmospheric phenomena, and it is possible to study the track of transport of aerosols on the global scale. Passive sensors such as MODIS (Moderate Resolution Imaging Spectroradiometer) are widely used in aerosol studies, but such sensors are limited to daily measurements for measuring in the visible wavelengths (Ma et al., 2011). It is also not capable of separating the path radiance from top-of-atmosphere radiance (Omar et al., 2009). Products of aerosols of MODIS sensor is heavily influenced by surface conditions in the pixel condition, and the field reflection affects their data (Ma et al., 2011). Most importantly, the vertical distribution of aerosols is not possible by such sensors (Kar et al., 2015). Ground-based passive remote sensing can provide information on columnar features of aerosols; however, vertical profiles of aerosols can be measured only by active remote sensing. Micro Pulse lidar is a known device for its portability over the years. MPLNET (Micro-Pulse Lidar Network) network has been established at AERONET stations (Welton et al., 2000). Micro Pulse lidar measures the backscatter coefficient of an aerosol profile at a wavelength of 532 nm in high vertical resolution (30–75 m)

The CALIPSO satellite is a joint mission of CNES and NASA, launched in 2006. This satellite carries an active Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) sensor that is capable of measuring the microphysical properties of atmospheric particles and providing vertical profiles of them with high spatial resolution and also measuring their depolarization (Winker et al., 2009; Ma et al., 2011; Kar et al., 2015). This sensor operates throughout the day without affecting the surface conditions of the earth. These unique characteristics make the CLIPSO data an appropriate source to the atmospheric science.

2. Related works

A large number of studies have used CALIPSO data for a wide range of atmospheric studies. CALIPSO profiles were utilized for various fields including; testing dust transfer (Liu et al., 2008; Abdi Vishkaee et al., 2012; Cabello et al., 2012; Yu et al., 2015; Francis et al., 2017), aerosol distribution (Chen et al., 2012; Mishra and Shibata, 2012), aerosol radiation (Feng and Christopher, 2014; Basha et al., 2015) and validation of climate models (Koffi et al., 2012). NASA presents an online bibliography related to atmospheric researches using CALIPSO data (https://www-calipso.larc.nasa.gov/resources/bibliographies.php).

Atmospheric scene classification is another challenging issue in which CALIOP lidar data have been widely utilized (Liu et al., 2009; Omar et al., 2009; Vaughan et al., 2009; Chen et al., 2010; Lu et al., 2011; Ma et al., 2011; Gong and Ma, 2012; Naeger et al., 2013; Liu et al., 2014, 2015). The discrimination of clouds and atmospheric aerosols is the first and the most important step in this filed which was emphasized in several works (Vaughan et al., 2005; Liu et al., 2005, 2009; Ma et al., 2011; Burton et al., 2013; Liu et al., 2014). Existing methods can be divided into two general categories; methods based on probability density functions (PDFs-based) and machine learning. In PDFs-based methods, a probability function is first produced defining the probability that attribute X is associated with each of the classes (aerosol or cloud). In the other word, using these PDFs, separation rules of classes are specified using a set of thresholds on input features.

Liu et al., 2004, before the launch of the CALIPSO satellite, used a two-dimensional (2D) and three-dimensional (3D) probability density function (so-called version 1 PDF) to classify the data of two-wavelength lidar. The features used include the layer mean attenuated backscatter at 532 nm, layer-integrated 1064-nm to 532-nm volume color ratio and mid-layer altitude. The results indicate a significant superiority of the 3D density function in comparison the 2D probability density function which mistakenly classifies the low ratio of thin clouds as an aerosol. CAD (Cloud-Aerosol Discrimination) is another algorithm to separate cloud from aerosol which uses a multi-dimensional probability density function (Liu et al., 2009). In 2009, Liu and his colleagues proposed the CAD-V2 algorithm based on the 3D probability density function for the CALIPSO satellite version 2 lidar data. They acknowledged that this algorithm has some misclassification in the classification of thick dust, smoke near the emission regions, and thin-light clouds in polar regions. In 2010, Liu et al. (2010) developed the CAD-V3 algorithm based on a five-dimensional (5D) probability density function for the CALIPSO satellite version 3 data. They concluded that the discrimination cloud from aerosol in the 5D space is better than the 3D space that was previously used for version 2 data. Studies have shown that even with using a 5D density function the overlap between the aerosol-density function with the cloud-density function is inevitable. An effective way to reduce this disadvantage is to use thermal bands. The CLIM (Combined Lidar and IR Measurement) algorithm was provided by Chen et al., (2010) which uses the combination of CALIOP lidar data and brightness temperature difference (BTD) of IIR sensor, both from the CALIPSO satellite, to detect the thick dust from ice clouds. This algorithm depends on infrared measurements and is practical for both day and night conditions. The results indicated that misclassification dust by the CLIM algorithm is much lower than the CAD-V2 method. In 2014, Liu and his colleagues used the CLIM algorithm to distinguish and validate the layers of dust aerosols and clouds in Sahara deserts and compared their results with CAD-V2 and CAD-V3 algorithms. The results have shown that CLIM is significantly better than CAD-V3 and CAD-V2 due to using thermal data. The CLIM method reduced the number of misclassified layers of dense dust layers. Misclassifications caused by CLIM in this area are mostly related to the mixture of dust and cloud layers. In other research's, the thermal data of other sensors was also considered for this purpose. The BTD CAD algorithm that uses infrared satellite measurements was used by Naeger and colleagues in 2013 to reduce the frequency of misclassified caused from CAD algorithm of CALIPSO satellite. This algorithm uses 11 and 12 μm infrared of MODIS sensors on the Aqua satellite or SEVIRI (Spinning Enhanced Visible and Infrared Imager) mounted on the MSG (Meteosat Second Generation) satellite which is very sensitive to dust concentrations. The results showed that the use of MODIS thermal data has been led up to better results.

A PDF-based method for cloud-aerosol classification and also subclass aerosol classification is provided by ASDC (Atmospheric Science Data Center) in the VFM product of level 2 data. This product is generated by 3 primary algorithms including SYBAL (Selective, Iterated Boundary Location), SCA (scene classification algorithms) and HERA (hybrid extinction retrieval algorithm). SIBYL is an algorithm proposed by Vaughan and his colleagues in 2009 to detect clouds and aerosols in signals that are mixed with the noise of the CALIOP sensor backscatter. SCA was composed of three different algorithms of which CAD is the main algorithm to discriminate clouds from aerosols (refer to Liu et al., 2005 and 2009). Burton et al., 2013 showed that the classification conformity of the CALIOP VFM product with HSRL-1(High Spectral Resolution Lidar) classification for marine pollution, industrial pollution, thick layers of dust, smoke and polluted dust is 62 and 54, 80, 13, and 35 percent respectively. According to the findings, the internal boundaries between different types of aerosol in the VFM product do not correctly represent the accurate boundary between the aerosol classes.