

Image segmentation for dust detection using semi-supervised machine learning

Manzhu Yu

*Department of Geography
Pennsylvania State University
University Park, PA, USA
mqy5198@psu.edu*

Julie Bessac

*Mathematics and Computer Science Division
Argonne National Laboratory
Lemont, IL, USA
jbessac@anl.gov*

Ling Xu

*Department of Mathematics
North Carolina A&T State University
Greensboro, NC, USA
lxu@ncat.edu*

Aryya Gangopadhyay

*Department of Information Systems
University of Maryland,
Baltimore County
Baltimore, MD, USA
gangopad@umbc.edu*

Yingxi Shi

*Joint Center for Earth Systems Technology
University of Maryland,
Baltimore County
Baltimore, MD, USA
yshi2@umbc.edu*

Jianwu Wang

*Department of Information Systems
University of Maryland,
Baltimore County
Baltimore, MD, USA
jianwu@umbc.edu*

Abstract—Dust plumes originating from the Earth’s major arid and semi-arid areas can significantly affect the climate system and human health. Many existing methods have been developed to identify dust from non-dust pixels from a remote sensing point of view. However, these methods use empirical rules and therefore have difficulty detecting dust above or below the detectable thresholds. Supervised machine learning methods have also been applied to detect dust from satellite imagery, but these methods are limited especially when applying to areas outside the training data due to the inadequate amount of ground truth data. In this work, we proposed an automatic dust segmentation framework using semi-supervised machine learning, based on a collocated dataset using Visible Infrared Imaging Radiometer Suite (VIIRS) and Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO). The proposed method utilizes unsupervised machine learning for segmentation of VIIRS imagery, and leverages the guidance from the dust labels using the dust profile product of CALIPSO to determine the dust clusters as the final product. The dust clusters are determined based on the similarity of spectral signature from dust pixels along the CALIPSO tracks. Experiment results show that the accuracy of the proposed framework outperforms the traditional physical infrared method along CALIPSO tracks. In addition, the proposed method performs consistently over three different study areas, the North Atlantic Ocean, East Asia, and Northern Africa.

Index Terms—dust detection, semi-supervised machine learning, multi-sensor remote sensing, image segmentation

I. INTRODUCTION

Dust events are common meteorological phenomena in arid and semi-arid regions, often arising when strong winds uplift fine-grained dust particles from the surface of the Earth. Atmospheric dust has various impacts on air quality, weather formation, radiative balance, biogeochemical cycles, and climate. On the other hand, dust storms are usually damaging. Due to climate change, the dynamics of dust storms at a local scale have changed drastically along with climate and

weather variables, such as total precipitation and average wind speed [1]. Frequencies and intensities of local dust storms are observed to be increasing, bringing higher impacts on wildlife, human society, and bio-community [2].

Satellite remote sensing can collect spatiotemporally continuous information of atmospheric components, but it is difficult to differentiate dust from other atmospheric components especially over bright surfaces. Depending on the spatial, temporal, and spectral characteristics, different satellite sensors have varying capabilities in detecting dust. For example, the Visible Infrared Imaging Radiometer Suite (VIIRS) on the Suomi National Polar-orbiting Partnership (Suomi NPP) satellite has a relatively broad spatial coverage (3040 km in the cross-track direction), making it possible to capture more dust storm. Meanwhile, the depolarization observation from Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) on-board the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) enhances the sensor’s ability of accurately detect non-spherical particles (*e.g.* dust). However, the lidar beam only covers a very narrow stripe (333m in the cross-track direction). To detect dust events and extract their spatial extent, it is beneficial to combine the above two types of satellite observations in a systematic approach.

Traditional methods of detecting dust from remote sensing imagery utilize the dust spectrum signatures, such as dust visualization (*e.g.* dust false color imaging), empirical thresholds, dust index [3]–[5], and the Deep Blue algorithm [6], [7]. These methods mainly utilize measurements at wavelengths of 8.5, 11.0, and 12.0 μm and calculate the brightness temperature differences (BTDs) to differentiate dust from clouds, surfaces, and other atmospheric particles. These methods are generally heuristic, by utilizing predefined thresholds that applied to general conditions, thus they may fail at extreme circumstances such as very optically thin or thick dust plumes or observing conditions that depart largely from the norm.

To address this problem, researchers have been using supervised machine learning methods to learn the relationship between the dust occurrences and the spectral radiance from satellite images in a pixel-wise way [8]–[10]. The trained models are then used to determine whether an observed pixel on the imagery is dust. However, supervised machine learning methods suffer from several drawbacks. Firstly, they require a set of training examples where dust pixels are labeled, but labeling the imagery pixel-wise is labor intensive and time-consuming. As an alternative to human labeling, CALIOP vertical feature mask and aerosol profile products are commonly used to categorize aerosols, but the CALIOP labels are limited spatially and may not cover the entire observing geometries [11]. Existing supervised machine learning methods for dust detection [8], [9] utilized CALIOP dust labels as ground truth, but the accuracy using the trained model on-track of CALIPSO to predict off-track is not guaranteed. Even if the on-track test accuracy is high, the off-track prediction accuracy can be low, because the model training process is limited by the amount of on-track dust labels. Secondly, dust over land is difficult to extract from satellite imagery because separating dust contribution from complex terrains and surface contributions can be challenging. Therefore, labeling dust from imagery by human beings is also prone to error. Thirdly, existing supervised machine learning methods were trained using data on the track of CALIPSO and do not take the spatial variation of dust plume into consideration. It is known that dust storms appear in a spatial cluster - rather than a pixel - and evolve over time; thus it is more suitable to take the spatial context of dust storms into account.

Unsupervised machine learning, on the other hand, can be used to extract patterns without pre-existing labels, making it suitable for dust detection from satellite imagery where labels are rare. However, unsupervised machine learning will not be able to utilize the dust labels from CALIPSO pixels, thus the results from unsupervised machine learning methods are difficult to relate to specific physical meanings.

This research aims to develop an automatic framework using semi-supervised machine learning to detect dust extent from VIIRS satellite imagery by integrating unsupervised machine learning and CALIPSO dust labels. The K-means clustering is firstly applied to segment VIIRS imagery into clusters, and then one or multiple clusters are selected as dust clusters based on their similarities to the dust spectral signature summarized from a collocated CALIPSO-VIIRS dataset. The dust spectral signature demonstrates the densely distributed spectral values of dust observations in different spectral bands. This approach addresses the limitations of the fixed thresholds and empirical parameters and ensures reproductivity.

The novelty of this research comes from three aspects:

- 1) **Methodology.** As discussed above, both supervised and unsupervised machine learning methods have limitations for the task of dust detection. This research proposes a novel semi-supervised machine learning method that will address these limitations. The method firstly clusters the satellite imagery with an unsupervised machine

learning method; and then it determines one or multiple dust clusters from knowledge distilled in step one. The dust cluster determination is based on the similarity between each cluster and the summarized CALIOP dust labels from the same spatiotemporal region.

- 2) **Application.** To the authors' knowledge, this is the first time that the semi-supervised machine learning method is used for dust detection from satellite imagery. The semi-supervised machine learning method is suitable for this specific task, because it will leverage the limited dust labels from CALIOP and produce segmentation results with a certain level of physical meanings. However, this semi-supervised machine learning is novel and different from existing methods for image segmentation (discussed in Section II.C.) due to the limited number of CALIOP dust labels.
- 3) **Experiment.** The validation of dust detection results used in the proposed method is a combination of accuracy metrics from both supervised and unsupervised machine learning methods. Traditional accuracy metrics from only one type of machine learning methods are not adequate to evaluate the quality of results. Therefore, this research utilizes classification metrics – precision, recall, and F_1 -score – to evaluate the results where the pixels are on-track of CALIPSO; and then utilizes silhouette scores to evaluate the quality of clustering. Therefore, the results are evaluated both on-track and off-track.

The paper is organized as follows. Section II reviews the literature on dust detection from remote sensing imagery. Section III states details of the datasets and the preprocessing steps. Section IV describes the proposed framework of dust extent extraction. Section V shows the results, including image segmentation, average accuracy, and method comparison. In the end, Section VI discusses future directions.

II. RELATED WORKS

A. Traditional dust detection methods

Existing methods of detecting dust plumes from passive satellite remote sensing have been utilizing the brightness temperature difference (BTD) between Thermal Infrared (TIR) bands at around $11\ \mu\text{m}$ and $12\ \mu\text{m}$ wavelengths to detect dust clouds over land surfaces [12]. This method assigns pixels as dust pixels with BTD values lower than zero based on the understanding that the desert dust exists when the BTD values generally decrease up to below zero [13], [14]. Later, a set of BTDs with corresponding heuristic thresholds was exploited to detect dust from meteorological clouds. Ackerman [3] proposed using two BTDs, *i.e.*, BT11–BT12 and BT8–BT11 (analyzing the signal at $8.5\ \mu\text{m}$ and $11\ \mu\text{m}$ wavelengths) to detect stratospheric volcanic aerosols over oceans. Similarly, Wald *et al.* [15] used the same two BTDs to identify mineral dust over desert regions. Miller [16] enhanced the investigation of daytime airborne dust over water and land. However, BTD has strong correlations with various land and dust properties,

such as the particle size distribution, chemical composition, and dust layer height [17], [19]. Thus, these threshold-based methods are sensitive to different dust events, study areas, or different seasons [19], [20]. Deep Blue algorithm was developed to calculate aerosol optical depth over land using the 412 nm band, where the contrast between aerosols and the bright surfaces such as deserts is larger, making it easy to differentiate aloft dust from dust on the surface [6], [7]. However, the Deep Blue algorithm can not be used over water.

Recent dust detection methods also integrated 15-day rolling mean cloud screened BTDR for each pixel [4]. Moreover, these methods have employed more shortwave bands to eliminate cloud effects [21], aiming to impose multiple fixed thresholds on calculated dust indices [5], [22], [23]]. However, these improvements of dust detection did not adequately address some important issues, such as the sensitivity to airborne dust identification over bright surfaces (*e.g.*, desert regions), the dependence of infrared signals on dust plume features (*e.g.*, plume height), the sensitivity of BTDR to variability in surface emissivity, and the impact of cirrus clouds on the BTDR signal [4], [24].

B. Supervised machine learning for dust detection

To address the limitations of fixed thresholds or empirical parameters, machine learning, including deep learning algorithms, can be utilized to learn the complex relationships between dust occurrences and the spectral radiance from satellite imagery, thus making it possible to outperform the derived thresholds from statistical analysis. A few researchers have started to investigate the performance of different machine (and deep) learning methods in dust detection. Existing studies have explored the common classification methods, including Support Vector Machines, Random Forests, Deep Neural Networks, and Convolutional Neural Networks [8], [9]. It is found that machine learning algorithms outperform the derived thresholds from statistical analysis [10]. For example, Strandgren *et al.* [25] developed an algorithm based on Artificial Neural Network (ANN) to study the characteristics of clouds and aerosols based on both SEVIRI and CALIOP. Shi *et al.* [8] compared different supervised machine learning methods, including SVM, Random Forest, Logistic Regression, on the capability of classifying dust or dust-free pixels along CALIPSO tracks using the spectral information from Moderate Resolution Imaging Spectroradiometer (MODIS) satellite. They also developed a hybrid classification model combining the physical infrared method with machine learning methods. Cai *et al.* [9] compared Logistic Regression, K-Nearest-Neighbor, Random Forest, Feed Forward Neural Network, and Convolutional Neural Network on the capability of classifying dust or dust-free pixels along CALIPSO tracks using the spectral information from VIIRS satellite. Kolios and Hatzianastassiou [10] utilized an ANN model to learn the relationship between the aerosol optical depth (AOD) values, obtained at the stations of AERONET, and the combinations of brightness temperatures of SEVIRI. The authors estimated AOD values during dust outbreaks in the Mediterranean re-

gion. These existing methods are using supervised machine learning, which is prone to the drawbacks mentioned in the Introduction section. At this moment, unsupervised and semi-supervised machine learning has not yet been investigated in dust detection from remote sensing imagery.

C. Semi-supervised machine learning for image segmentation

While semi-supervised machine learning methods have not yet been explored for dust detection, there are existing research using semi-supervised machine learning for various kinds of semantic image segmentation tasks.

Semantic image segmentation is commonly conducted using supervised machine learning. One of the best performing methods is Deep Convolutional Neural Network (DCNN), which can perform semantic segmentation using pixel-level labels [26]. Acquiring such pixel-level annotation is time-consuming, thus researchers have been exploring ways to train models using weak annotations, such as bounding box annotations or image-level labels. Papandreou *et al.* [27] explored using either bounding boxes and image-level labels or a mixture of few strongly labeled and many weakly labeled images to conduct semantic image segmentation using Expectation-Maximization (EM) methods. For dust detection, the CALIOP dust labels are pixel-based, but the pixels are limited on CALIPSO tracks. These labels are neither sufficient to form a bounding box, nor sparse enough to be one label per image.

Another direction is to use adversarial learning for semi-supervised semantic segmentation [28]. This type of semantic segmentation utilizes fully labeled images and unlabeled images to perform segmentation. For example, Hung *et al.* [29] proposed an adversarial learning method for semi-supervised semantic segmentation, and leverage unlabeled images to enhance the segmentation results. A fully-convolution discriminator network is trained between the labeled map and the input image to optimize the segmentation loss, the adversarial loss, and a semi-supervised loss based on the confidence map. For dust detection, the CALIOP dust labels on-track are not able to form fully labeled images, thus the adversarial learning is not suitable. Further semi-supervised machine learning methods for the task of dust detection are yet to be explored.

III. STUDY AREA AND DATA PREPARATION

A. Study areas

In this study, we selected three study areas covering different spatiotemporal ranges, see Fig. 1, (1) North Atlantic Ocean (74W-20W, 13N-43N) for the whole year of 2014, (2) Asian (110.9E-135.85E, 28.26N-44.38N) in Spring season (March, April, and May) in 2014, and (3) Northern Africa, Europe, and the Mediterranean (30W-60E, 0N-60N) in the Summer season (June, July, and August) in 2014. These three study areas have different challenges regarding dust detection from satellite imagery. The North Atlantic study area receives aloft dust from Sahara desert in Africa's west coast and the major task of dust detection in this area is to differentiate dust pixels from the ocean [30]. The Asian spring study area receives dust from the Gobi desert and contains polluted dust from industrial

sources, and the major challenge in this area is the difficulty of extracting dust from complex surface types [31]. The Northern Africa study area has the most dust events during the summer season, usually from aloft dust from the Sahara desert, and the challenge is to differentiate atmospheric dust from desert dust on the ground [32]. These three study areas are chosen to explore the capability and robustness of the proposed method.

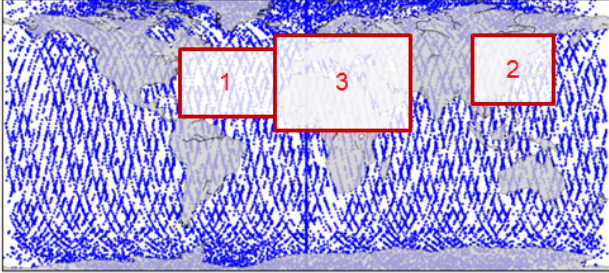


Fig. 1. Center points of the entire 2014 collocated data (25139 granules) and the three study areas: (1) North Atlantic Ocean (74W-20W, 13N-43N), (2) Asian (110.9E-135.85E, 28.26N-44.38N), and (3) Northern Africa, Europe, and the Mediterranean (30W-60E, 0N-60N).

B. VIIRS

The Visible Infrared Imaging Radiometer Suite (VIIRS) instrument observes and collects global satellite observations that span the visible and infrared (IR) wavelengths across land, ocean, and atmosphere (<https://ncc.nesdis.noaa.gov/VIIRS/>). VIIRS uses passive radiometer sensors. It has 22 channels ranging from $0.41 \mu\text{m}$ to $12.01 \mu\text{m}$. Five of these channels are high-resolution image bands or I-bands, and sixteen serve as moderate-resolution bands or M-bands. In this study, we use the 16 M-bands with 750 m spatial resolution across visible/reflective, near IR, shortwave IR, medium-wave IR, and longwave IR. Within these M-bands, M1-M5 and M7 primarily provide ocean color aerosol information, M6 provides atmospheric correction information, M8 provides cloud particle size information, M9 provides cirrus cloud cover information, M10 provides snow fraction information, M11 provides clouds information, M12-M13 and M15-M16 provide sea surface temperature and fires, and M14 provides cloud top properties.

C. CALIPSO

The Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) provide range-resolved information on the vertical distribution of aerosols and clouds (<https://www-calipso.larc.nasa.gov/>). The Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) the instrument on-board CALIPSO uses a two-wavelength elastic backscatter laser that transmits linear polarized light at 532 nm and 1064 nm , coupled with a receiver telescope of 1 m diameter that measures the perpendicular and parallel components of the attenuated backscatter at 532 nm and the total attenuated backscatter at 1064 nm . The CALIOP Level 2 (L2) aerosol and cloud profile products include information on the aerosol and cloud backscatter coefficient at 532 nm and 1064 nm , and

the particle depolarization ratio at 532 nm . CALIOP emits 20 laser pulses per second and measures curtains of attenuated backscatter profiles along the satellite track with a vertical resolution of up to 30 m [33]. CALIOP also provides a vertical feature mask product profiling aerosol subtypes and cloud profiles within 1 km and 5 km , along with other information in the data. The aerosol subtypes are marine, dust, polluted continental/smoke, clean continental, polluted dust, elevated smoke, and dusty marine.

D. VIIRS data download and data preprocessing

This research builds on a collocated VIIRS-CALIPSO dataset from Cai *et al.* [9]. The collocated data contains VIIRS observations only on the CALIPSO tracks to reduce the necessary file size. However, the CALIPSO tracks are too narrow (333 m in the cross-track direction) for dust spatial extent detection. To obtain the spatial extent of dust, we downloaded the VIIRS granules (each covering a swath of 3040 km in the cross-track direction) corresponding to the collocated data. Two VIIRS products (VNP02MOD and VNP03MOD products) are downloaded from the Atmosphere SIPS product search website operated by the University of Wisconsin Madison. VNP02MOD is the product for VIIRS/NPP Moderate Resolution 6-Min L1B Swath 750 m , and VNP03MOD is the VIIRS/NPP Moderate Resolution Terrain-Corrected Geolocation 6-Min L1 Swath 750 m .

Then we subset the VIIRS granule to the rectangular region based on the bounding boxes of the collocated data (Fig. 2). To extract the information related to dust on the CALIPSO track, we categorized the on-track pixels into two categories based on aerosol sub-types provided by the CALIOP Level 2 product: 1) dust: pixels containing dust or polluted dust, 2) non-dust: pixels not containing dust or polluted dust. Both categories are cloud screened.

IV. METHODS

The automatic framework of detecting dust extent is illustrated in the workflow, see Fig.3. Step 1, pixels on CALIPSO tracks are categorized into dust and non-dust. Step 2, each prepared VIIRS granule subset is clustered using K-means. Step 3, the segmentation result is generated. Each cluster occupies a proportion of the VIIRS granule subset. Step 4, the dust signature of the study area is generated based on all dust pixels on CALIPSO tracks, and the dust signature is essentially a matrix with each dust pixel stored in a row, and the corresponding VIIRS spectral band values stored in a column. Step 5, similarities of the VIIRS spectral band values between each cluster in the segmentation result and the dust signature are examined to determine if the resulting cluster is more likely to be dust. Cluster(s) with high similarity values will be considered as dust cluster(s). Step 6, the resulting dust extent is generated. Step 7, pixels on track of CALIPSO are used to validate the resulting dust extent. Silhouette scores are also calculated to evaluate the quality of clustering.

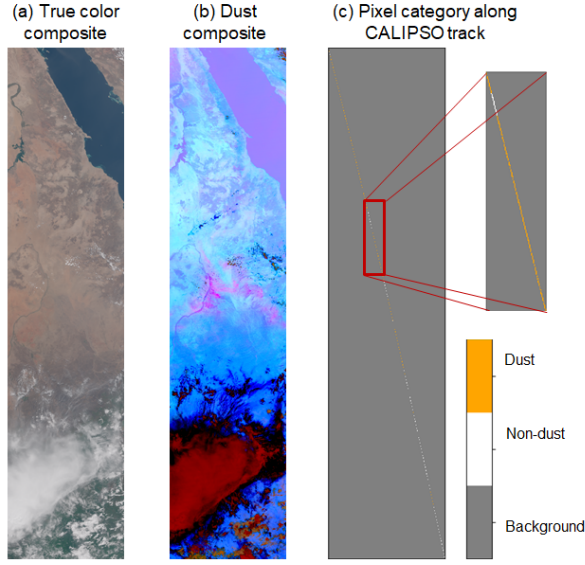


Fig. 2. Illustration of data sets at a selected area in North Africa and the Mediterranean, (a) VIIRS true-color composite, (b) VIIRS dust composite, (c) the pixel category on CALIPSO track.

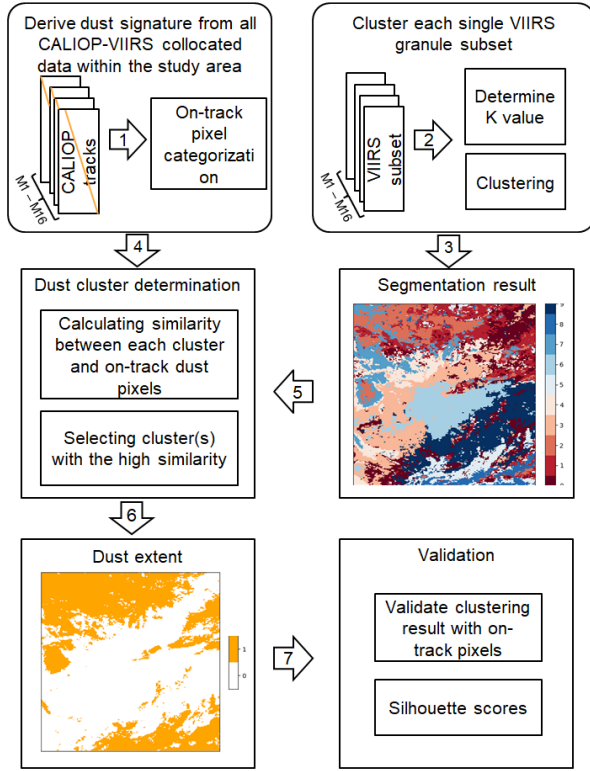


Fig. 3. Workflow of the methods.

A. K-means clustering

K-means clustering is used to segment the VIIRS imagery into different clusters initially. K-means clustering [34] is a method that partitions a dataset of observations $(x_1, \dots, x_n)$

into K ($K \leq n$) sub-groups called clusters $(C_1, \dots, C_K)$. Each cluster C_K is identified by its mean m_k value and generally an arbitrary label k . Observations from the dataset are assigned to the cluster with the nearest mean m_k , the Euclidean distance is generally used to measure the closeness to the cluster center. The underlying principle of the K-means clustering is to create a partition $(C_1, \dots, C_K)$ that minimizes the sum over the clusters of the within-cluster variance. This within-cluster sum of squares (WCSS) is also called inertia.

Clusters and their means are derived iteratively starting from a random guess of the cluster means $m_1, \dots, m_K$, and by alternatively proceeding through the following two steps until reaching a stopping criterion. The first step is called assignment. During which each observation is assigned to the cluster with the nearest mean m_k in terms of Euclidean distance. The second step is called an update which recalculates the mean m_k of the observations assigned to each cluster during the previous step. Several stopping criteria are used, such as no change in the cluster means at a tolerance threshold, the number of iterations, and no improvement in the cluster variance.

Many variations of the K-means clustering have been proposed in the literature based on different initialization methods, distances, or different cluster representatives. For example, the K-medoids use the median of each cluster instead of the mean. Nevertheless, these methods provide similar results as the current K-means, and therefore their results are not included in this report.

B. Dust cluster determination

After obtaining the clusters based on the K-means, it is essential to determine which cluster (or potentially multiple clusters) represents dust. We performed an initial data exploration of the spectral bands. The probabilistic distribution of each band on the CALIPSO track given their dust or non-dust category is shown in Fig. 4, using the North Atlantic region. In the figure, the x-axis labels the band number, and the y-axis shows the value of the spectral bands. It is observed that for bands M1-M11 (top two subplots), dust and non-dust bands display similar mean values (orange line); however, the non-dust category spreads out wider, indicating a larger variance. Similar features are seen in bands M12-M16 (bottom two subplots). The other two study areas show different probability distributions, but the difference between dust and non-dust categories remains. The significant spectral difference between dust and non-dust categories inspired us to find the most similar clusters to the dust category in spectral behavior and consider them as dust clusters.

Then, the dust clusters from the initial K-means are determined by incorporating information on dust pixels along the CALIPSO track. First, all dust pixels from CALIPSO within a selected study area are aggregated, and the values of VIIRS M1-M16 bands associated with each dust pixel are collected as the dust signature (right column of Fig. 4). The **dust signature** is essentially a matrix $D = [N, M]$, where N is the number of dust pixels, and M is the number of spectral

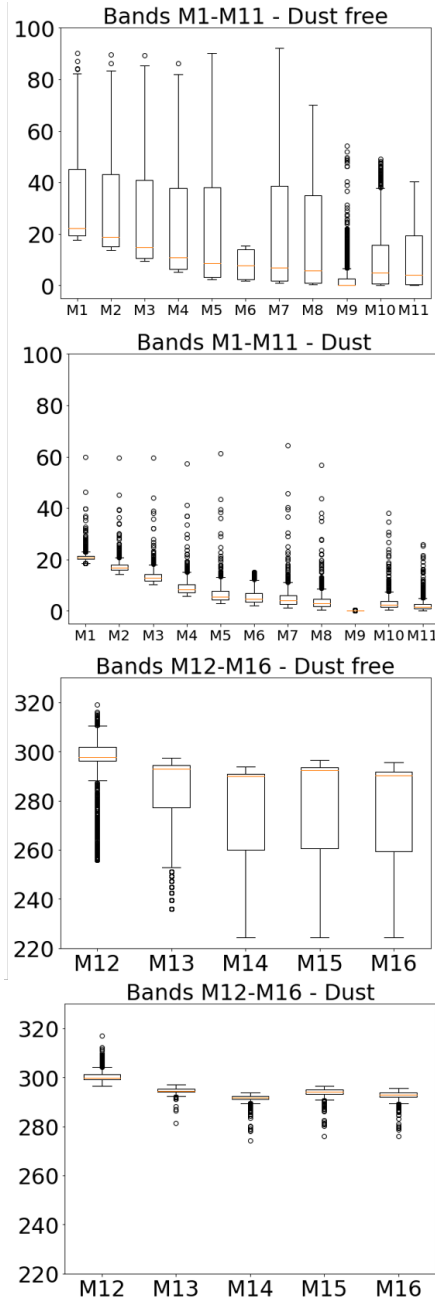


Fig. 4. (North Atlantic region) Boxplot shows the probability distributions of all the bands on the CALIPSO track in the dust and non-dust categories. Right column will be referred to as part of the dust signature.

bands (M1-M16). For each study area, a specific dust signature matrix is composed.

Second, for a specific VIIRS imagery, after K -means clustering, the **similarities** between each cluster and the dust signature matrix is calculated in a pairwise way. For each cluster, the number of pairwise similarity calculation is $N \times n$, where n is the number of pixels within the cluster. The dust signature matrix can be extensive regarding the N number. To reduce the calculation, we select only 10,000 samples randomly from the dust signature matrix.

The similarity function utilized here is the **Euclidean distance**. The cluster that has the highest similarity to the dust signature matrix is considered as the dust cluster. Additionally, several K-means clusters may contain dust information, especially since several types of dust (pure, polluted, etc.) are present. If the similarity values of other clusters to the dust signature matrix are within a valid range, *i.e.*, the similarity values are also high enough, then these clusters are considered as potential dust clusters. Potential dust clusters can complement the small dust region effect when the number of clusters (K) is large.

V. EXPERIMENTS AND RESULTS

In this section, we demonstrate the dust extent resulted from the proposed method, along with the accuracy metrics comparing to dust profiles on the collocated CALIPSO tracks and the silhouette scores that evaluate the quality of the clustering result. We compared the performances of the proposed method in different study areas. We also computed the accuracy values resulted from the proposed method and the physical infrared method. In the experiment, the number of clusters is set to be $K = 10$ by observing the clustering quality using a significant number of datasets.

Experiments were conducted on the *taki* server of UMBC High Performance Computing Facility (HPCF; <https://hpcf.umbc.edu/>). Our proposed method was implemented in Python with the support of libraries, such as Satpy (for satellite image processing), scikit-learn (for K-means clustering), and matplotlib (for graphics generation).

A. Dust extent extraction using the proposed method

Results from the three different study areas are demonstrated in the following. To quantify the accuracy of dust detection, the pixels along the CALIPSO track are used to verify the result. **This can be considered as verifying a classification of dust vs. non-dust along the CALIPSO track.** If a detected dust pixel is truly in the dust category on the CALIPSO track, then the detection is successful for the particular pixel. Similarly, if a detected non-dust pixel is truly not in the dust category on the CALIPSO track, then the detection is also successful for this particular pixel. Other cases are considered as unsuccessful detection.

For the North Africa region in the summer season, Fig. 5 shows the (a) VIIRS true colors composite image, the (b) VIIRS dust composite image, the (c) the segmentation results using $K = 10$, and (d) the extracted dust extent of the VIIRS imagery. The overall accuracy is 0.78, with the dust category having a precision of 0.69 and a recall of 0.33, and the non-dust category having a precision of 0.79 and a recall of 0.95. For the dust category, the percentage of dust detected by the proposed method that is CALIPSO observed dust is 0.69; and the recall is 0.33, meaning the percentage of CALIPSO observed dust that is correctly detected by the proposed method.

For the North Atlantic study area, Fig. 6 shows the composite images and the dust extent result generated from the

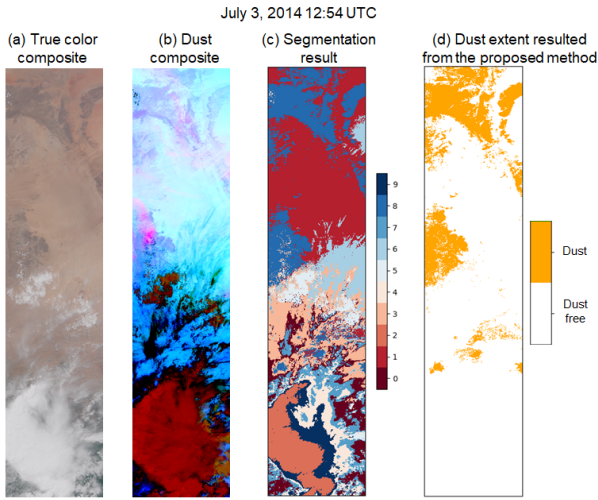


Fig. 5. (North Africa/Summer) Composite images of the VIIRS granule subset at 2014184t1254 and dust extent result.

proposed method. The overall accuracy is 0.72, with the dust category having a precision of 0.66 and a recall of 0.99, and the non-dust category having a precision of 0.91 and a recall of 0.44.

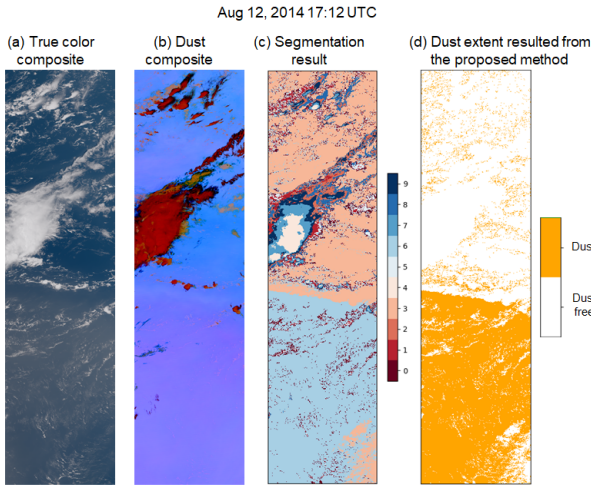


Fig. 6. (North Atlantic) Composite images of the VIIRS granule subset at 2014224t1712 and dust extent result.

For the Asian dust study area in the Spring season, Fig. 7 shows the composite images and the dust extent result generated from the proposed method. The overall accuracy is 0.79, with the dust category having a precision of 0.61 and a recall of 0.77, and the non-dust category having a precision of 0.9 and a recall of 0.8.

Additionally, we computed the silhouette coefficients to measure the clustering quality. The silhouette coefficient takes values between -1 and 1 , and it measures how well each data-point belongs to each cluster. The silhouette coefficient is interpreted as follows: the closer to 1 , the better it is; a

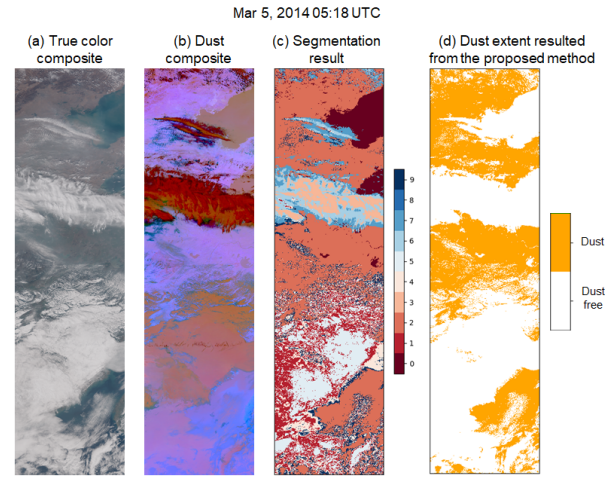


Fig. 7. (Asia/Spring) Composite images of the VIIRS granule subset at 2014064t0518 and dust extent result.

value around and below 0.5 indicates that data-point should belong to neighboring clusters; negative values indicate that data-points poorly belong to the assigned cluster.

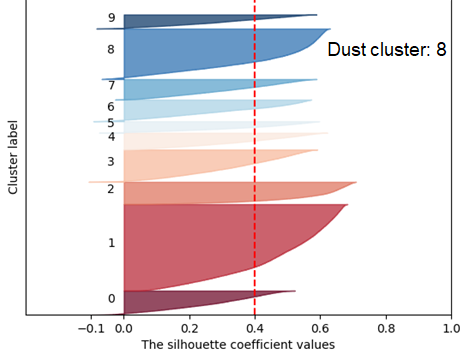
Fig. 8 shows the average silhouette derived from all data-points, and the silhouette for each data-point is shown. The three examples from different study areas show the clustering quality at around 0.4 , which indicates that neighboring clusters can be further merged. The low silhouette scores are due to the high number of clusters. We found in the experiment that decreasing the number of clusters can improve the silhouette score. However, since the goal of the clustering task is to extract dust extent, we intentionally set the number of clusters high to segment the image as detailed as possible, and then use the similarity calculation to merge potential dust clusters. As can be seen in Fig. 8c, two clusters (2 and 8) are selected are dust clusters and merged to be the final dust extent. Therefore, the proposed method complements to relatively low silhouette values.

B. Comparison of average accuracy with the physical infrared method

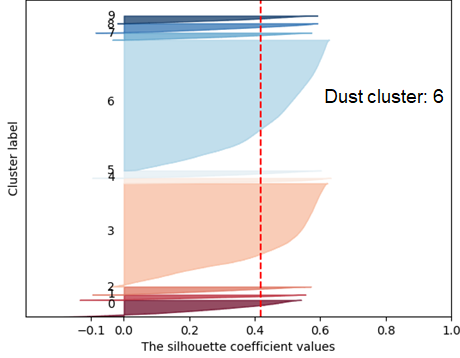
Table I illustrates the mean accuracy value and the standard deviation generated from all sample imagery using the proposed method, i.e., the mean overall accuracy at 0.64 and the standard deviation at 0.11 . The mean values and standard deviations of precision and recall for both the dust and the non-dust categories are also listed. Overall, the proposed method achieves a high mean precision for the non-dust category (0.82) and a high recall for the dust category (0.68).

There are potential ways to improve on-track accuracy of the proposed method. First, the dust signature summarized for each spatiotemporal region is not yet refined. A random sample of $10,000$ CALIOP dust labels from each specific spatiotemporal region is selected to form the dust profile. However, the selected samples may not represent every possibility of dust pixels. Second, the dust signature from CALIOP labels is

(a) Northern Africa summer example: July 3, 2014 12:54 UTC



(b) North Atlantic region example: Aug 12, 2014 17:12 UTC



(c) Asian spring dust example: Mar 5, 2014 05:18 UTC

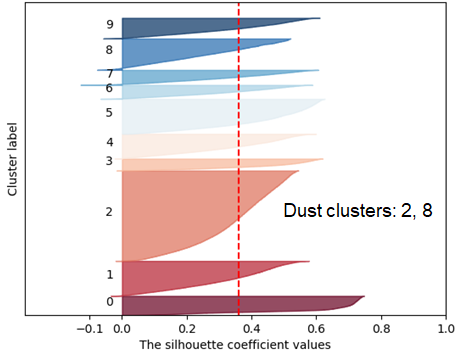


Fig. 8. Silhouette coefficients of the three examples in Figs.5-7. (a) North Africa region, average silhouette: 0.40. (b) North Atlantic region, average silhouette: 0.42, (c) Asian Spring dust region, average silhouette: 0.36.

TABLE I
COMPARING THE **ON-TRACK** ACCURACY BETWEEN THE PROPOSED METHOD AND THE PHYSICAL INFRARED METHOD.

	Accuracy Proposed method		Accuracy Ackerman (1997) [3]	
	Mean	STD	Mean	STD
Overall accuracy	0.64	0.11	0.52	0.13
Precision (dust)	0.52	0.16	0.44	0.13
Recall (dust)	0.68	0.32	0.99	0.03
Precision (non-dust)	0.82	0.15	0.88	0.27
Recall (non-dust)	0.60	0.24	0.23	0.16

limited by CALIPSO's narrow coverage. Integrating multiple years' CALIOP data may improve the representativeness of dust pixels. Third, the proposed method is sensitive to the number of clusters (K). Each separate imagery might require a different initial K number, whereas we specify the number of clusters as $K = 10$ in the experiments.

Comparing with the physical infrared method [3], the proposed method achieves a higher mean overall accuracy (0.64 vs. 0.52). For the dust category, the proposed method results in a higher mean precision (0.52 vs. 0.44), but a lower recall (0.68 vs. 0.99). This indicates that the proposed method has a higher percentage of detected dust that is CALIPSO observed dust, but a lower percentage of CALIPSO observed dust that is truly detected. For the non-dust category, the proposed method results in a lower mean precision (0.82 vs. 0.88), but a higher recall (0.60 vs. 0.23). This indicates that the proposed method has a lower percentage of detected non-dust that are CALIPSO observed non-dust, but a higher percentage of CALIPSO observed non-dust that are truly detected.

Fig. 9 shows the dust extent extracted from the physical infrared method, where the dust extents are more significant than the ones generated from the proposed method. Compared to the true-color and dust composites, it can be observed that the physical infrared method generates several mistakenly extracted dust regions. For the case of July 3, 2014, the physical infrared method wrongly treats the arid area in the Sahara region as dust. For the case of Mar 5, 2014, the physical infrared method also treats the entire southern part of the imagery as dust, where the true-color imagery clearly shows a thick cloud. It seems that the physical infrared method differentiates more likely the cloud (white regions in Fig. 9) from other pixels (yellow regions in Fig. 9), but sometimes it also fails to detect the cloud as well.

C. Comparison of average accuracy using the proposed method among three different study areas

Table II illustrates the mean value and standard deviation of accuracy metrics using the proposed method over the three different study areas. Overall, the proposed method performs best detecting the Asian Spring dust with an accuracy of 0.71 but performs worst (overall accuracy: 0.60) detecting dust extents in the North Atlantic region.

For the dust category, the mean precision ranges 0.45 – 0.55 for the three study areas, indicating 50% of detected dust is CALIPSO observed dust; but the mean recall shows significant differences among the three areas. The mean recall for the North Atlantic region is the highest (0.99), indicating that almost all CALIPSO observed dust pixels are successfully detected. The mean recall for North African Summer dust is the lowest (0.42), indicating that less than half of the CALIPSO observed dust pixels are detected. One of the assumptions for the high recall values in the North Atlantic region is that the surface type in this region is relatively simple: ocean or cloud. It is relatively easier than detecting dust in the other two study areas. The North African study area contains arid areas or deserts, which makes it difficult to differentiate

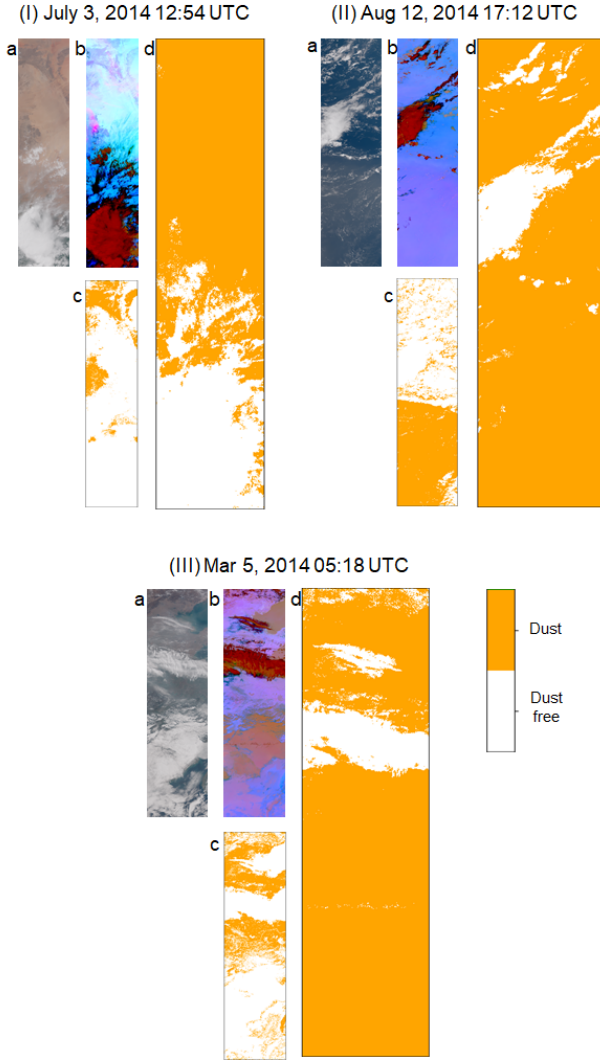


Fig. 9. Dust extent resulted from the physical infrared method for the three example imagery of Figs.5-7. Sub-images a (true color composites), b (dust composites), and c (results generated from the proposed method) are replicates of images in Figs.5-7. Subimages d are results generated by Ackerman [3].

TABLE II
COMPARING THE **ON-TRACK** ACCURACY USING THE PROPOSED METHOD
AMONG THREE DIFFERENT STUDY AREAS.

	North Africa Summer		North Atlantic		Asia Spring	
Overall accuracy	Mean 0.64	STD 0.10	Mean 0.60	STD 0.12	Mean 0.71	STD 0.07
Precision (dust)	0.56	0.15	0.51	0.13	0.47	0.19
Recall (dust)	0.42	0.19	0.99	0.01	0.67	0.33
Precision (non-dust)	0.68	0.11	0.96	0.03	0.87	0.09
Recall (non-dust)	0.78	0.12	0.33	0.13	0.70	0.15

from aloft dust particles. The Asian study area contains a more

complex mixture of surface types, including shrubs, forests, and deserts.

For the non-dust category, the mean precision ranges 0.68 – 0.96, with the highest precision over the North Atlantic area, and the lowest precision over the North Africa area. This indicates that in the North Atlantic area, almost all of the detected non-dust are CALIPSO observed non-dust, but in the North Africa area, only less than 70% of the detected non-dust are CALIPSO observed non-dust. The mean recall values for the North Africa and Asia study areas are relatively high (0.78 and 0.70), whereas the mean recall for the North Atlantic study area is relatively low (0.33). This indicates that in the North Atlantic area, only 33% of the CALIPSO observed non-dust pixels are detected. This is probably due to the fact that the CALIPSO observed non-dust pixels include cloud pixels and pixels of other aerosol types, such as marine aerosols, continental aerosols, or smoke. The North Atlantic study area contains significantly more cloud pixels than the other study areas, so the mean recall value is the lowest.

VI. CONCLUSIONS AND FUTURE DIRECTIONS

In this paper, we are using a semi-supervised machine learning method to identify dust extents from satellite imagery. We designed a workflow to extract dust signature within spatiotemporal ranges based on CALIOP identified dust pixels and used these data to select from the K-means clusters as the final dust extents. We examined the sensitivity of our method in different experiments, including 1) testing the capability in detecting dust in different study areas, 2) comparing the average accuracy to a physical infrared method, 3) comparing the average accuracy among different study areas. We also validated our results, 1) using the common classification accuracy matrix for all the pixels along the CALIPSO tracks, and 2) using silhouette coefficient scores to evaluate the clustering performances.

For the future work, we will improve this study in the following directions. First, we will further validate the resulting dust extents by comparing it with other existing aerosol and dust products. One of the products we tried in the experiments was the Visible Infrared Imaging Radiometer Suite Environmental Data Record (VIIRS Aerosol EDR) product [35], but the product was not able to detect thin dust and failed to identify dust pixels in any of our validation time periods. Second, we will continue search for available products to validate our results further. In addition, this research utilizes the CALIOP data to guide the clustering, but the CALIOP data is only available along the track; thus, it cannot be used in a fully supervised setup for spatial clustering. Third, we will explore semi-supervised approaches as an alternative to segment dust from imagery using a small number of CALIOP-derived dust labels.

The semi-supervised machine learning method for dust detection can be easily adopted for other spectral and thermal satellite products, such as the Level 1 products from MODIS, Geostationary Satellite Server (GOES), and Spinning Enhanced Visible and Infrared Imager (SEVIRI). In addition,

the procedure of composing a signature matrix using a limited amount of available ground-truth labels can be applied to the detection of various other types of extreme events, such as wildfires, tropical cyclones, dust storms.

ACKNOWLEDGMENT

This work is supported by the grant “CyberTraining: DSE: Cross-Training of Researchers in Computing, Applied Mathematics and Atmospheric Sciences using Advanced Cyberinfrastructure Resources” from the National Science Foundation (grant no. OAC-1730250). The hardware used in the computational studies is part of the UMBC High Performance Computing Facility (HPCF). The facility is supported by the U.S. National Science Foundation through the MRI program (grant nos. CNS-0821258, CNS-1228778, and OAC-1726023) and the SCREMS program (grant no. DMS-0821311), with additional substantial support from the University of Maryland, Baltimore County (UMBC). See hpcf.umbc.edu for more information on HPCF and the projects using its resources. Julie Bessac was partially supported by the U.S. Department of Energy, Office of Science, Advanced Scientific Computing Research Program under contract DE-AC02-06CH11357.

REFERENCES

- [1] Middleton, N., 2019. Variability and trends in dust storm frequency on decadal timescales: Climatic drivers and human impacts. *Geosciences*, 9(6), p.261.
- [2] Taylor, C.M., Belušić, D., Guichard, F., Parker, D.J., Vischel, T., Bock, O., Harris, P.P., Janicot, S., Klein, C. and Panthou, G., 2017. Frequency of extreme Sahelian storms tripled since 1982 in satellite observations. *Nature*, 544(7651), pp.475-478.
- [3] Ackerman, S.A., 1997. Remote sensing aerosols using satellite infrared observations. *Journal of Geophysical Research: Atmospheres*, 102(D14), pp.17069-17079.
- [4] Ashpole, I. and Washington, R., 2012. An automated dust detection using SEVIRI: A multiyear climatology of summertime dustiness in the central and western Sahara. *Journal of Geophysical Research: Atmospheres*, 117(D8).
- [5] Taylor, I., Mackie, S. and Watson, M., 2015. Investigating the use of the Saharan dust index as a tool for the detection of volcanic ash in SEVIRI imagery. *Journal of Volcanology and Geothermal Research*, 304, pp.126-141.
- [6] Hsu, N.C., Tsay, S.C., King, M.D. and Herman, J.R., 2006. Deep blue retrievals of Asian aerosol properties during ACE-Asia. *IEEE transactions on geoscience and remote sensing*, 44(11), pp.3180-3195.
- [7] Hsu, N.C., Jeong, M.J., Bettenhausen, C., Sayer, A.M., Hansell, R., Seftor, C.S., Huang, J. and Tsay, S.C., 2013. Enhanced Deep Blue aerosol retrieval algorithm: The second generation. *Journal of Geophysical Research: Atmospheres*, 118(16), pp.9296-9315.
- [8] Shi, P., Song, Q., Patwardhan, J., Zhang, Z., Wang, J. and Gangopadhyay, A., 2019, September. A Hybrid Algorithm for Mineral Dust Detection Using Satellite Data. In 2019 15th International Conference on eScience (eScience) (pp. 39-46). IEEE.
- [9] Cai, C., Lee, J., Shi, Y., Zerfas, C., Guo, P., and Zhang, Z., 2019. Dust Detection in Satellite Data using Convolutional Neural Networks. Technical Report HPCF-2019-15, UMBC High Performance Computing Facility, University of Maryland, Baltimore County
- [10] Kolios, S. and Hatzianastassiou, N., 2019. Quantitative Aerosol Optical Depth Detection during Dust Outbreaks from Meteosat Imagery Using an Artificial Neural Network Model. *Remote Sensing*, 11(9), p.1022.
- [11] Huang, J., Minnis, P., Yi, Y., Tang, Q., Wang, X., Hu, Y., Liu, Z., Ayers, K., Trepte, C. and Winker, D., 2007. Summer dust aerosols detected from CALIPSO over the Tibetan Plateau. *Geophysical Research Letters*, 34(18).
- [12] Shenk, W.E. and Curran, R.J., 1974. The detection of dust storms over land and water with satellite visible and infrared measurements. *Monthly Weather Review*, 102(12), pp.830-837.
- [13] McClain, P. E., 1989. Global sea surface temperatures and cloud clearing for aerosol optical depth estimates. *International Journal of Remote Sensing*, 10(4-5), pp.763-769.
- [14] Prata, A.J., 1989. Infrared radiative transfer calculations for volcanic ash clouds. *Geophysical research letters*, 16(11), pp.1293-1296.
- [15] Wald, A.E., Kaufman, Y.J., Tanré, D. and Gao, B.C., 1998. Daytime and nighttime detection of mineral dust over desert using infrared spectral contrast. *Journal of Geophysical Research: Atmospheres*, 103(D24), pp.32307-32313.
- [16] Miller, S.D., 2003. A consolidated technique for enhancing desert dust storms with MODIS. *Geophysical Research Letters*, 30(20).
- [17] Sokolik, I.N., Toon, O.B. and Bergstrom, R.W., 1998. Modeling the radiative characteristics of airborne mineral aerosols at infrared wavelengths. *Journal of Geophysical Research: Atmospheres*, 103(D8), pp.8813-8826.
- [18] Pierangelo, C., Chédin, A., Heilliette, S., Jacquinet-Husson, N. and Armante, R., 2004. Dust altitude and infrared optical depth from AIRS.
- [19] Darmenov, A. and Sokolik, I.N., 2005. Identifying the regional thermal-IR radiative signature of mineral dust with MODIS. *Geophysical Research Letters*, 32(16).
- [20] Baddock, M.C., Bullard, J.E. and Bryant, R.G., 2009. Dust source identification using MODIS: a comparison of techniques applied to the Lake Eyre Basin, Australia. *Remote Sensing of Environment*, 113(7), pp.1511-1528.
- [21] Miller, S.D., Bankert, R.L., Solbrig, J.E., Forsythe, J.M., Noh, Y.J. and Grasso, L.D., 2017. A Dynamic Enhancement With Background Reduction Algorithm: Overview and Application to Satellite-Based Dust Storm Detection. *Journal of Geophysical Research: Atmospheres*, 122(23), pp.12-938.
- [22] Marchese, F., Sannazzaro, F., Falconieri, A., Filizzola, C., Pergola, N. and Tramutoli, V., 2017. An enhanced satellite-based algorithm for detecting and tracking dust outbreaks by means of SEVIRI data. *Remote Sensing*, 9(6), p.537.
- [23] She, L., Xue, Y., Yang, X., Guang, J., Li, Y., Che, Y., Fan, C. and Xie, Y., 2018. Dust detection and intensity estimation using Himawari-8/AHI observation. *Remote Sensing*, 10(4), p.490.
- [24] Chaboureaud, J.P., Tulet, P. and Mari, C., 2007. Diurnal cycle of dust and cirrus over West Africa as seen from Meteosat Second Generation satellite and a regional forecast model. *Geophysical research letters*, 34(2).
- [25] Strandgren, J., Bugliaro Goggia, L., Sehnke, F. and Schröder, L., 2017. Cirrus cloud retrieval with MSG/SEVIRI using artificial neural networks. *Atmospheric Measurement Techniques (AMT)*, 10(9), pp.3547-3573.
- [26] Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K. and Yuille, A.L., 2014. Semantic image segmentation with deep convolutional nets and fully connected crfs. *arXiv preprint arXiv:1412.7062*.
- [27] Papandreou, G., Chen, L.C., Murphy, K.P. and Yuille, A.L., 2015. Weakly- and semi-supervised learning of a deep convolutional network for semantic image segmentation. In *Proceedings of the IEEE international conference on computer vision* (pp. 1742-1750).
- [28] Mondal, A.K., Dolz, J. and Desrosiers, C., 2018. Few-shot 3d multi-modal medical image segmentation using generative adversarial learning. *arXiv preprint arXiv:1810.12241*.
- [29] Hung, W.C., Tsai, Y.H., Liou, Y.T., Lin, Y.Y. and Yang, M.H., 2018. Adversarial learning for semi-supervised semantic segmentation. *arXiv preprint arXiv:1802.07934*.
- [30] Van Der Does, M., Korte, L.F., Munday, C.I., Brummer, G.J.A. and Stuut, J.B.W., 2016. Particle size traces modern Saharan dust transport and deposition across the equatorial North Atlantic. *Atmospheric Chemistry & Physics*, 16(21).
- [31] Li, G., Chen, J., Ji, J., Yang, J. and Conway, T.M., 2009. Natural and anthropogenic sources of East Asian dust. *Geology*, 37(8), pp.727-730.
- [32] Engelstaedt, S. and Washington, R., 2007. Atmospheric controls on the annual cycle of North African dust. *Journal of Geophysical Research: Atmospheres*, 112(D3).
- [33] Winker, D.M., Vaughan, M.A., Omar, A., Hu, Y., Powell, K.A., Liu, Z., Hunt, W.H. and Young, S.A., 2009. Overview of the CALIPSO mission and CALIOP data processing algorithms. *Journal of Atmospheric and Oceanic Technology*, 26(11), pp.2310-2323.
- [34] Lloyd, S., 1982. Least squares quantization in PCM. *IEEE transactions on information theory*, 28(2), pp.129-137.
- [35] JPSS Visible Infrared Imaging Radiometer Suite Environmental Data Record (VIIRS_EDR), URL: https://www.star.nesdis.noaa.gov/smcd/emb/viirs_aerosol/products_edr.php