Image segmentation for dust detection using unsupervised machine learning

CyberTraining 2020: Big Data + High-Performance Computing + Atmospheric Sciences

Julie Bessac[1], Ling Xu[2], Manzhu Yu[3]

Faculty mentor: Aryya Gangopadhyay[4]; External mentor: Yingxi Shi[5]; Research Assistant: Pei Guo[4]

[1] Mathematics and Computer Science Division, Argonne National Laboratory; [2] Department of Mathematics, North Carolina A&T State University; [3] Department of Geography, Pennsylvania State University; [4] Department of Information Systems, University of Maryland Baltimore County; [5] NASA GSFC Climate and Radiation Laboratory

Technical Report HPCF–2020–17, hpcf.umbc.edu > Publications

Abstract

Dust and sandstorms originating from Earth's major arid and semi-arid desert areas can significantly affect the climate system and health. Many existing methods use heuristic rules to classify on a pixel-level regarding dust or dust-free. However, these heuristic rules are limited in applicability when the study area or the study period has changed. Based on a multi-sensor collocation dataset, we sought to utilize unsupervised machine learning techniques to detect and segment dust in multispectral satellite imagery. In this report, we describe the datasets used, discuss our methodology, and provide preliminary validation results.

1 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 plays a positive role in absorbing light radiation and the formation of clouds (Prospero, 1999). 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 (Middleton, 2019). Frequencies and intensities of local dust storms are observed to be increasing, bringing higher impacts on wildlife, human beings, and bio-community (Taylor et al., 2017).

A highly accurate and efficient method for dust detection is desired, which can predict the occurrence and intensity of dust events from high-quality dust observations in a timely fashion, and at the same time mitigate the adverse effects of dust storms. One such method of dust detection is to use various satellites and their observations. For example, the satellites that range from polar to geostationary orbiting, and the various spectral bands ranging from thermal infrared to LIDAR. It is noted that the polar-orbiting satellites generally have a high spatial resolution but limited temporal resolution. Examples include the Visible Infrared Imaging Radiometer Suite (VIIRS) on the Suomi National Polar-orbiting Partnership (Suomi NPP) satellite, the Ozone Mapping Profiling Suite (OMPS) on Suomi NPP, and Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) onboard the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO).

Existing methods for dust detection include using dust RGB product, heuristic thresholding, and calculating dust index. However, these methods suffer limitations. The dust signal interpretation is human-subjective and time-intensive, and therefore, lacking quantitative detection of the dust storm occurrences (Gonzalez and Briottet, 2017).

In this project, we aim to use the techniques of unsupervised machine learning to automate the process of dust detection. This approach addresses the limitations of the fixed thresholds and empirical parameters and ensures reproducibility. It is found that machine learning algorithms learn the complex relationships between dust occurrences and the spectral radiance from satellite images and make it possible to outperform the derived thresholds from statistical analysis (Kolios and Hatzianastassiou, 2019). To get started, we use the unsupervised machine learning method of Kmeans, the satellite data of VIIRS imagery, and the CALIPSO dust profile.

The paper is organized as follows. Section 2 introduces the datasets used in this study, which are VIIRS and CALIPSO. Section 3 describes the unsupervised machine learning methods that we have adopted for clustering dust extents, Kmeans. Section 4 shows the preliminary results, including image segmentation, method validation, and prediction. In the end, Section 5 discusses future directions.

2 Related works

Existing methods of detecting dust outbreaks from satellite remote sensing have been utilizing the brightness temperature difference (BTD) between Thermal Infrared (TIR) bands at around 11 μm and 12 μm wavelengths to detect dust clouds over land surfaces (Shenk and Curran, 1974). 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 (McClain, 1989; Prata, 1989). Later, a set of BTDs with corresponding heuristic thresholds was exploited to detect dust from meteorological clouds. Ackerman (1997) proposed using two BTDs, i.e., BT11–BT12 and BT8–BT11 (analyzing the signal at 8.5 μm and 11 μm wavelengths) to detect stratospheric volcanic aerosols over oceans. Similarly, Wald et al. (1998) used the same two BTDs to identify mineral dust over desert regions. Miller (2003) 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 (Sokolik et al., 1998; Pierangelo et al., 2004). Thus, these threshold-based methods are sensitive to different dust events, study areas, or different seasons (Darmenov and Sokolik, 2005; Baddock et al., 2009).

Recent dust detection methods also integrated 15-day rolling mean cloud screened BTD for each pixel (Ashpole and Washington, 2012). Moreover, these methods have employed more shortwave (UV, VIS, and NIR) bands to eliminate cloud effects (Miller et al., 2017), aiming to impose multiple fixed thresholds on calculated dust indices (Taylor et al., 2015; Marchese et al., 2017; She et al., 2018). However, these improvements of dust detection did not adequately address a few important issues, such as the sensitivity to airborne dust identification over bright surfaces (e.g., desert regions), the dependence of IR signals on dust plume features (e.g., plume height), the sensitivity of BTD to variability in surface emissivity, and the impact of cirrus clouds on the BTD signal (Ashpole and Washington, 2012; Chaboureau et al., 2007).

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. During the past decade, machine learning methods have been widely used in meteorology and atmospheric science to classify cloud type and estimate rainfall intensity (Lazri and

Ameur, 2018). More recently, a few researchers have started to investigate the performance of different machine (and deep) learning methods in dust detection. For example, Strandgren et al. (2017) developed an algorithm based on Artificial Neural Network (ANN) to study the characteristics of clouds and aerosols based on both SEVIRI and CALIOP. Kolios and Hatzianastassiou (2019) 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 author estimated AOD values during dust outbreaks in the Mediterranean region. However, the applicability of these methods to other study areas are still questionable, and the sensitivity of these methods to spectral bands used in the training is also not discussed.

3 Data preparation

3.1 VIIRS

The Visible Infrared Imaging Radiometer Suite (VIIRS) instrument observes and collects global satellite observations that span the visible and infrared wavelengths across land, ocean, and atmosphere (https://ncc.nesdis.noaa.gov/VIIRS/). It has 22 channels ranging from 0.41 µm to 12.01 µm. 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 750m 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.

3.2 CALIPSO

The Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) satellite has been providing 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) instrument 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) data includes 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 (Winker et al., 2009). In this study, we use the CALIPSO aerosol and cloud profiles prepared by Team 5 of CyberTraining 2019 (Cai et al., 2019). The profiles include aerosol subtypes and cloud profiles within 1km and 5km, 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.

3.3 VIIRS data download and data preprocess

Besides the collocated VIIRS and CALIOP data prepared by Team 5 of CyberTraining 2019, we also downloaded VIIRS granule (VNP02MOD and VNP03MOD products) using the API (https://sips.ssec.wisc.edu/#/products/api). In the collocated data, VIIRS has only one dimension on the CALIOP track. In order to obtain the spatial extent of dust, we downloaded the VIIRS granules corresponding to the collocated data. In this preliminary study, we selected three spatiotemporal ranges 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.

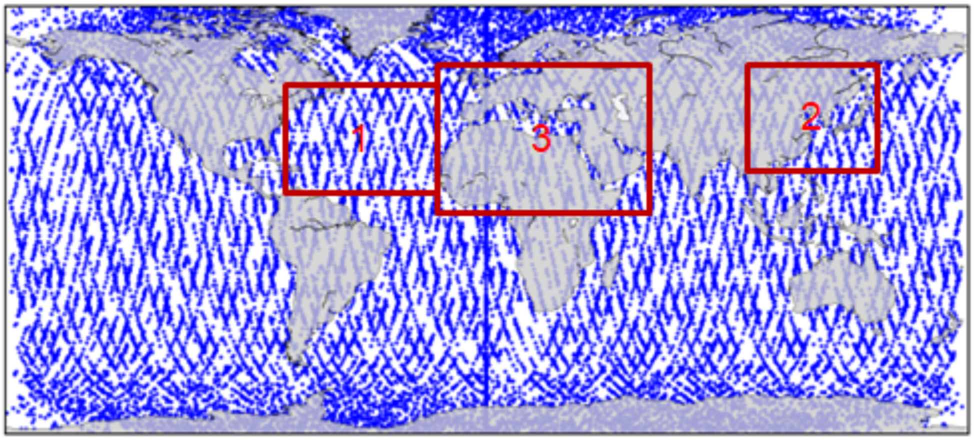


Figure 1. Center points of the entire 2014 collocated data (25139 granules) and the three study areas: (1) North Atlantic Ocean (74W-20W, 13N-43N), (2) Asian (110.9E-135.85E, 28.26N-44.38N), and (3) Northern Africa, Europe, and the Mediterranean (30W-60E, 0N-60N).

We then subset the VIIRS granule to the rectangular region based on the bounding boxes of the collocated data, and then crop the spectral bands into 256 by 256 pixels, keeping the spatial resolution of 750m. We used a moving 256 by 256 window and cropped the data so that the window is always intersecting the CALIPSO track, see Figure 2. To extract the information related to dust on CALIPSO track, we categorized the on-track pixels into five categories:

1) dust (dust only no cloud no other aerosols)
2) pure polluted dust
3) dust with polluted dust
4) dust or polluted dust with other aerosols
5) other aerosols only.

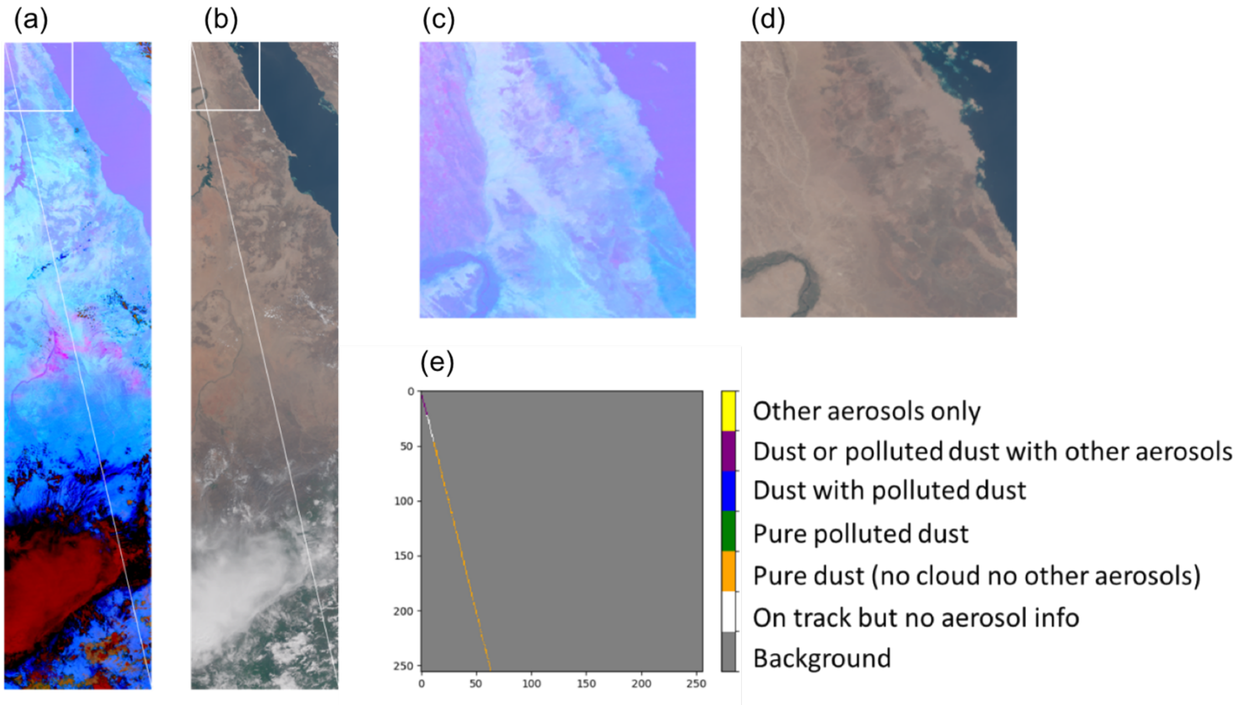


Figure 2. Illustration of data sets at a selected area in North Africa and Caribbean, (a) VIIRS dust composite, (b) VIIRS true color composite, (c) enlargement of the top left corner in (a), (d) enlargement of the top left corner in (b), (e) the dust category on CALIPSO track.

We performed some initial data exploration on the spectral bands. Figure 3 shows the boxplots which shows the probability distributions of the 16 considered VIIRS bands on the CALIPSO-track for three categories of dust and aerosol (0, 1, 4). It is observed that bands containing dust show less variance in distributions (smaller boxes) and lower intensity values. We will refer to this particular behavior as dust signature or dust profile for the distributions of bands according to the CALIOP dust and aerosol categories. One should note that bands M1 to M11 have most values of zero at night-time since they depend on visible wavelengths. The bands distributions will guide the interpretation of the Kmeans clusters in Sections 4 and 5.

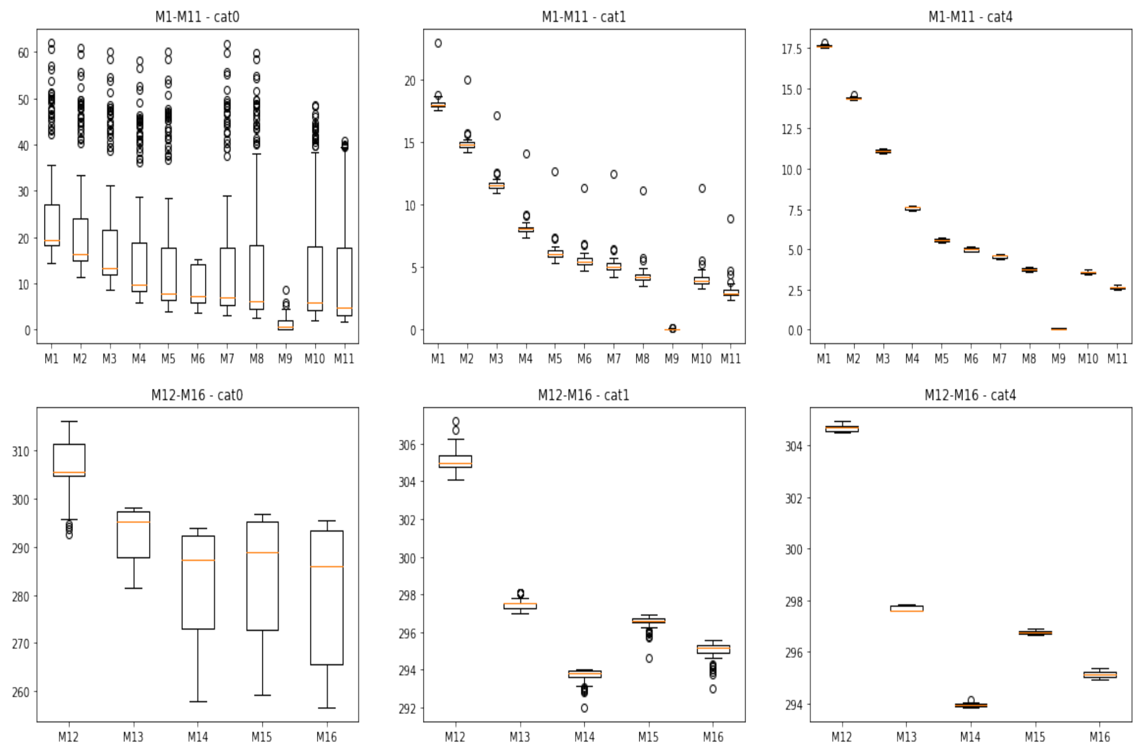


Figure 3: Boxplot of the 16 bands extracted on VIIRS that collocated with the CALIOP track, for different categories (0, 1 and 4) of dust and aerosols. Orange line: distribution median, box edges: 25% - and 75%-quantiles, dots are outliers.

4 Methods

Figure 4 illustrates the workflow of our methods. Details of the components are described in Sections 4.1 and 4.2. In Step 1, pixels on CALIPSO tracks are categorized into groups related to dust based on the VIIRS CALIPSO collocated data. Category 1 (pure dust) will be considered as the dust pixels, and the other categories are considered as dust-free in our first trials of experiments. In Step 2, each prepared VIIRS granule subset is clustered using K-means. The number of clusters (K) is determined using the L-curve method for optimization. In Step 3, the segmentation result is generated. Each cluster occupies a proportion of the VIIRS granule subset. In 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. In 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). In Step 6, the resulting dust extent is generated. In Step 7, pixels on track of CALIPSO are used to validate the resulting dust extent. The validation using existing aerosol products, such as the VIIRS Aerosol Environmental Data Record (EDR), is ongoing.

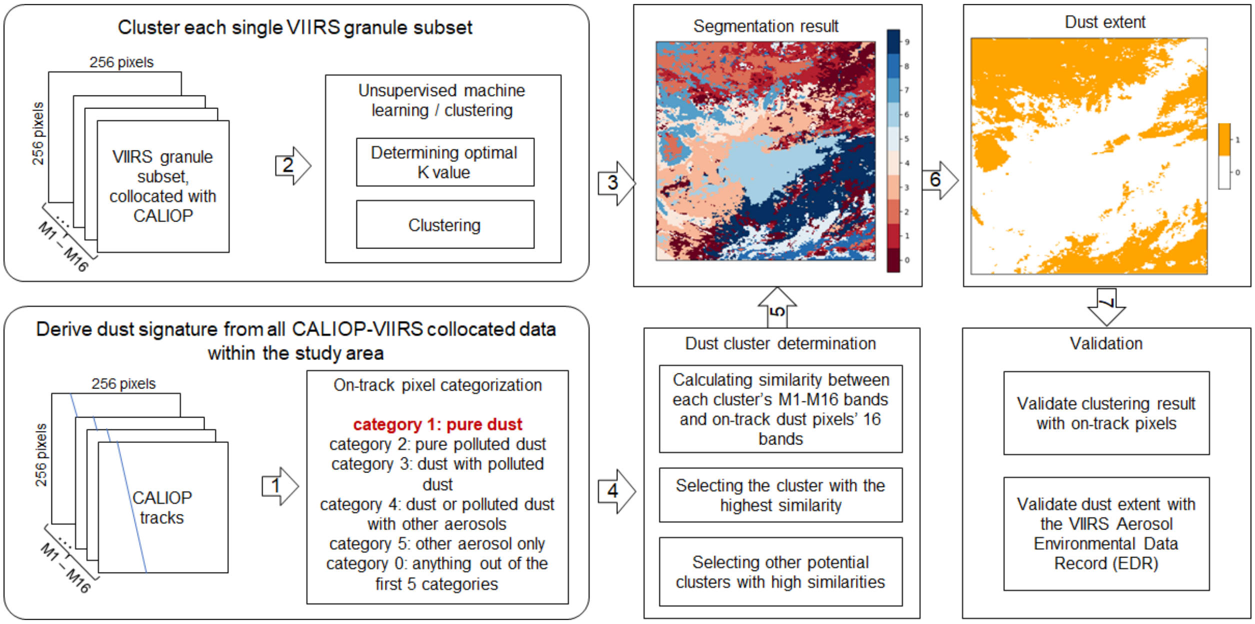


Figure 4. Workflow of the methods.

4.1 Unsupervised machine learning

We start the study with unsupervised learning techniques, specifically, the K-means clustering method (https://en.wikipedia.org/wiki/K-means_clustering). K-means clustering is a method that partitions a dataset of observations (x_1, …, x_n) into K (K≤n) sub-groups called clusters (C_1, …, 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, …, C_K) that minimizes the sum over the clusters of the within-cluster variance (within-cluster sum of squares WCSS also called inertia).

Clusters and their means are derived iteratively starting oftentimes from a random guess of the cluster means $m^{(1)}$, …, $m^{(K)}$, and by alternatively proceeding through the following two steps until reaching a stopping criterion. The first step is the **assignment** during which each observation is assigned to the cluster with a nearest mean $m^{(k)}$ in terms of Euclidean distance. The second step called **update** consists of re-calculating the mean $m^{(k)}$ of the observations assigned to each cluster during the previous assignment step. Several stopping criteria are used such as no change in the cluster means at a tolerance threshold, number of iterations, and no improvement in the cluster variance.

The optimal number of clusters K is determined empirically through the L-curve or elbow method, which consists of running K-means method several times with an increasing number of clusters and plotting the inertia or WCSS of the clustering as a function of the number of clusters. The optimal number of clusters is chosen as the number shows an inflection in the curve (an elbow shape). We illustrate in the following section the computation of the optimal number of clusters.

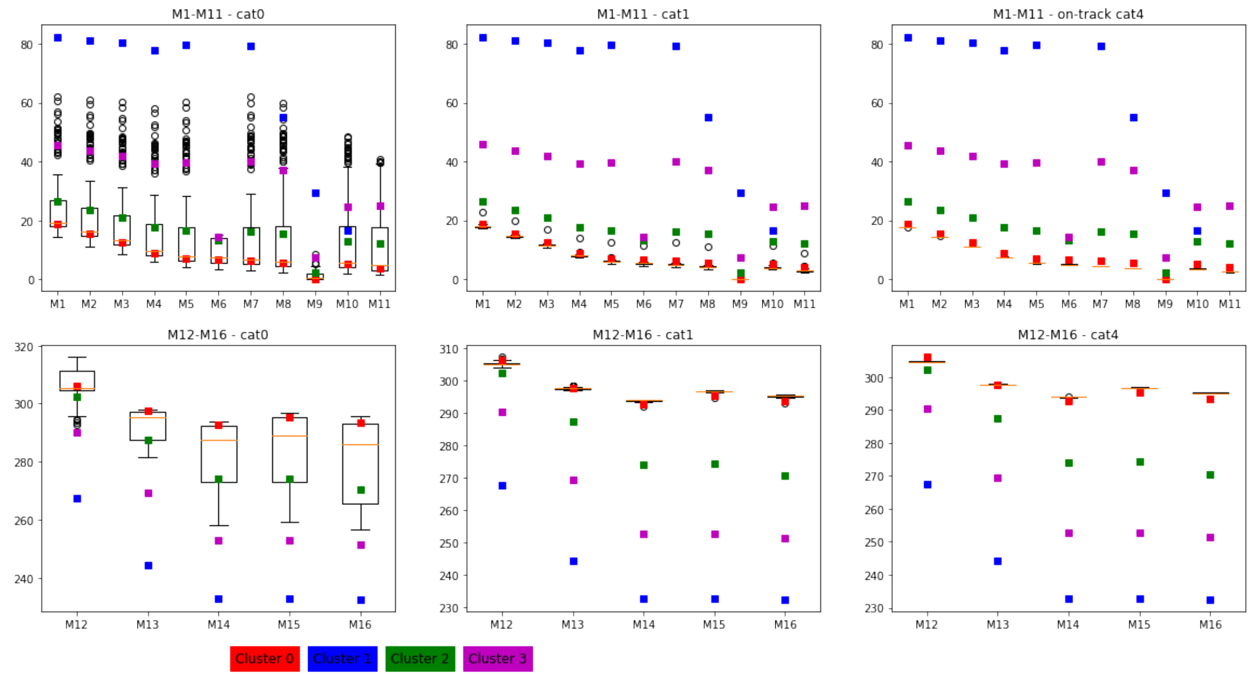
Many variations of the K-means clustering have been proposed in the literature: based on different initialization methods, distances, and different cluster representants such as the K-medoids where the median of each cluster is used instead of the mean. Indeed, in the K-means clustering centroids do not necessarily belong to the dataset since they are averages of dataset points. The use of the medians in the K-medoids is to ensure the physical meaning of the centroids.

In the following, we focus on the most commonly used techniques based on random initialization and Euclidean distance. We present results for the K-means and K-medoids.

4.2 Dust cluster determination

After obtaining the clusters based on the unsupervised machine learning algorithm, it is essential to determine which cluster (or potentially multiple clusters) represents dust. The cluster determination process relies on the collective dust signature within the 16 spectral bands in the CALIOP-VIIRS collocated data. Specifically, the CALIOP-VIIRS collocated data is firstly categorized based on aerosol subtypes and cloud information into six different groups: 1) **pure dust** (no cloud and no other aerosols), 2) pure polluted dust (no cloud and no other aerosols), 3) dust with polluted dust, 4) dust or polluted dust with other aerosols, 5) other aerosols only, and 6) other. In this project, we focus on segmenting dust pixels from the dust-free pixels on the VIIRS granule, therefore, only Category 1 (pure dust) is considered as dust pixels, and the other five categories are considered as dust-free. All Category 1 (pure dust) pixels within the study area are aggregated, and the values of VIIRS M1-M16 bands associated with each dust pixel are collected as the dust signature matrix (Figure 3). We use a set of metrics and statistics described in the following to select and interpret the K-means clusters.

First, the **similarities** between each cluster and the dust signature matrix are calculated and are displayed in Figure 5. The similarity function utilized here is the **Euclidean distance**. The cluster that has the highest similarity (smallest Euclidean distance) to the dust signature matrix is considered as the dust cluster. 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. Figure 5 also shows the similarities and dissimilarities between cluster centroids and bands extracted for each targeted dust and aerosol categories from CALIPSO-track data. To further quantify results in Figure 5, the Euclidean distance between the clustering centroids and the mean of all the bands in each CALIOP dust-aerosol categories is shown below. In this example case, Cluster 0 (red) has its centroid the closest to the distribution of the bands of pure dust (cat1).




```
Distance between centroids and bands mean per category
[[ 34.4   4.1   5.5]
 [187.5 221.4 222.6]
 [ 17.8  50.2  51.3]
 [ 83.8 120.4 121.6]]
Cluster minimizing distance centroids and bands mean per category
0
```

Figure 5: Boxplot of the 16 bands extracted on CALIPSO-track (same as Figure 3). Colored pixels represent the centroids of each cluster when a K-means clustering is performed with 4 clusters on the example dataset. In this example, cluster C_0 is visually the closest from the bands corresponding to pure dust (category 1 in central column). Euclidean distances computed confirms the closeness of cluster C_0 to the bands categorized as pure dust.

In addition to inspecting similarities between the centroids and spectral bands, a statistical exploration of each cluster is led. Figure 6 shows the statistical distribution of the example clusters on the CALIPSO track and on the entire studied image. We observe that the most prevalent cluster for the area is the most prevalent on the CALIPSO track as well. It appears that the most prevalent cluster is cluster C_0 that is being selected as minimizing the mean of the Euclidean distance for pure dust.

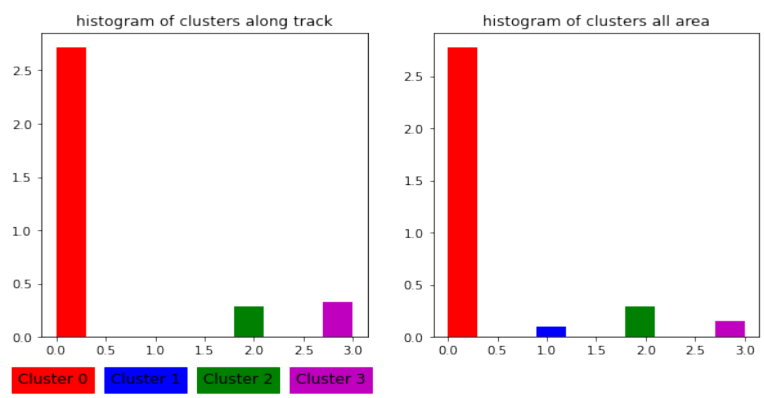


Figure 6: Repartition of clusters on the whole area and along the CALIOP track. Cluster C_0 minimizing the Euclidean distance between the centroids and the bands means in each dust-aerosol category is the most prevalent cluster.

Finally, in addition to showing on-track band distribution, we explore the distributions of the bands and the clusters for the entire study area. Figure 7 plots the boxplot of the band's distribution of the pixels in each cluster to further investigate the physical meaning. The probability distribution of each band for the entire study area is represented by a box, we show the distribution of the bands in each cluster from the K-means clustering and from the pure dust data of CALIPSO data. We can observe that several clusters distributions differ significantly from that of the pure dust bands on the left column. Additionally, colored pixels represent the centroids of each K-means cluster. We expect to use the different bands distributions for each cluster to guide the interpretation of the clusters. In this figure, we observe that Cluster 0 has boxplot distributions the most similar to the pure dust (most left boxplot) in terms of intensity and dispersion, and the centroid of Cluster 0 falls inside the boxplot showing a good agreement between the cluster centroid and the overall band distributions for this cluster. Additionally Clusters 1 and 3 show band distributions that differ significantly from the band distribution of the pure dust on CALIPSO-track (most left panel), we suspect these clusters contains little information about dust.

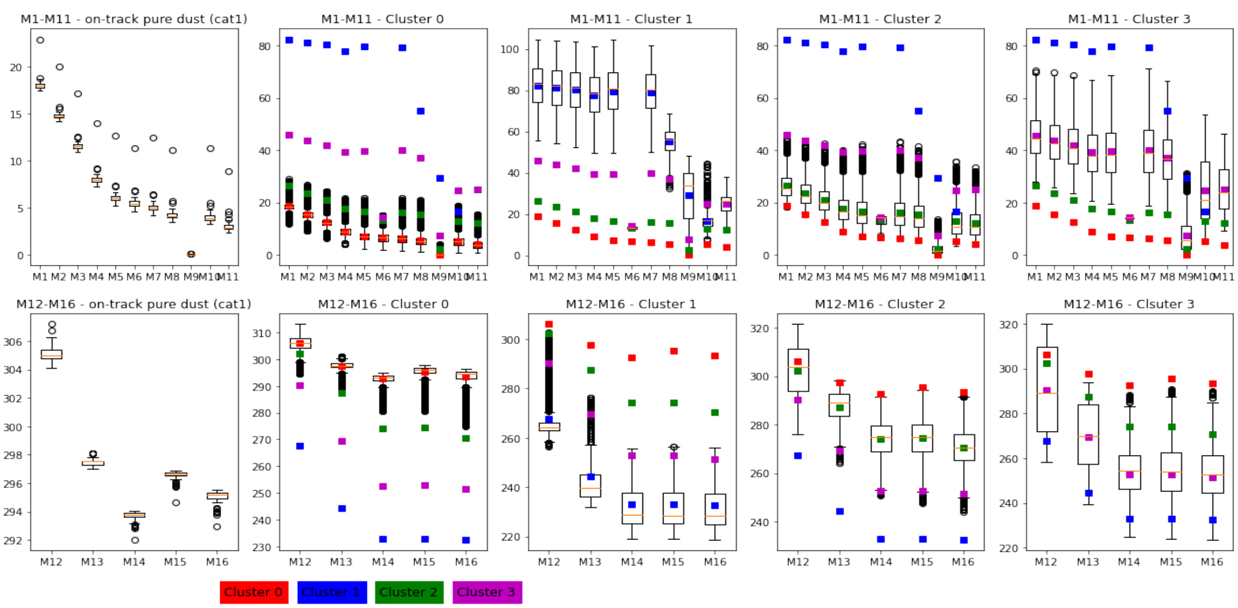


Figure 7: Boxplots of bands on CALIPSO track for pure dust category (category 1) (First column). Distribution of bands at each pixel of each cluster (second to fifth column). Colored squares are the centroids of each cluster.

We use this set of statistics and metrics to determine the candidate cluster containing the most dust information. This cluster is used in the following section to evaluate the performance of the K-means and K-medoids methods.

5 Experiments and results

In the following, we compare results of dust detection, using three K-means related methods (K-means, K-Medoids, Fuzzy C-means), at three different datasets (North Atlantic, Asian Spring, Northern Africa Summer), and at two different image sizes (small:256*256 pixels and large: entire VIIRS image snapshot).

Subsection 5.1 applies the K-means clustering on a single image using 16 VIIRS radiative bands. The size of the image is of 256*256 pixels. In order to improve these initial clustering results, we refine our experiments by exploring the accuracy of the K-means at given various land surface types. This is because each surface type has a distinctive radiative behavior influencing the radiative bands. Subsection 5.2 compares accuracy results with two variations of the K-means: K-Medoids and Fuzzy C-means. The K-medoids method is able to obtain a cluster within the dataset and thus improve the interpretability of each cluster. Subsection 5.3 applies the K-means on a single image using 3 selected VIIRS radiative bands. The goal is to examine the relative importance. Subsection 5.4 presents results of the clustering performed on larger images in order to explore a greater spatial extent of dust.

Several challenges are raised in this section. First, one needs to interpret the results of the K-means clusters and explain their physical meaning in terms of dust information. In order to tackle this challenge, the collocated CALIOP data are used to help characterize the clusters along the track, which contain the dust type present along the satellite track. Since the surface-type underneath the atmosphere influences the radiative transfers, a surface-type categorical variable is used to further interpret the K-means clusters. Additionally, several K-means clusters may contain dust information, especially since several types of dust (pure, polluted, etc.) are present. Following this direction, we will use the improved K-means described in Section 4.2. Finally, it is known that expert-human labeling is required in order to strengthen the interpretation of K-means clusters. Although it sounds daunting, if performed on a small number of images, this approach could guide the characterization and interpretation of the K-means clusters.

The second challenge is to determine the optimal number of clusters for all images. Nevertheless, we performed an L-curve method on a significant number of datasets, most of them showed an optimal number of clusters around **4,** as shown in Figure 8. In the following study, we will use 4 clusters. In future work, one can design a criterion to better determine an optimal number of clusters depending on the dataset features.

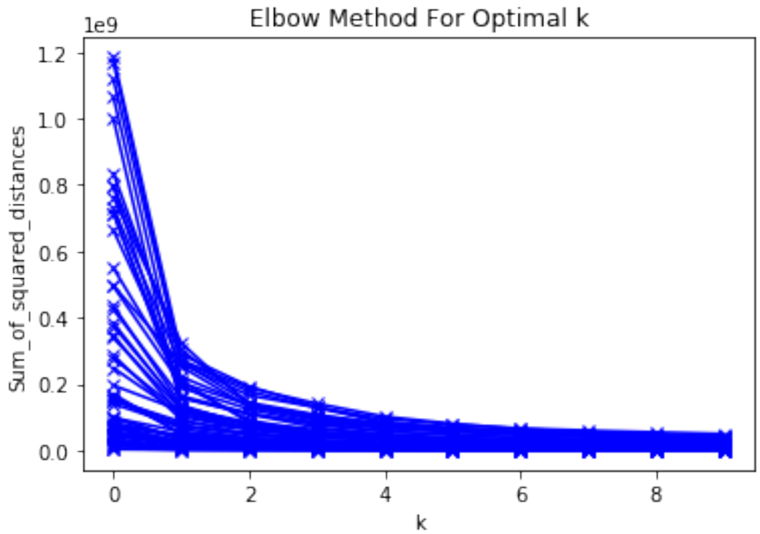
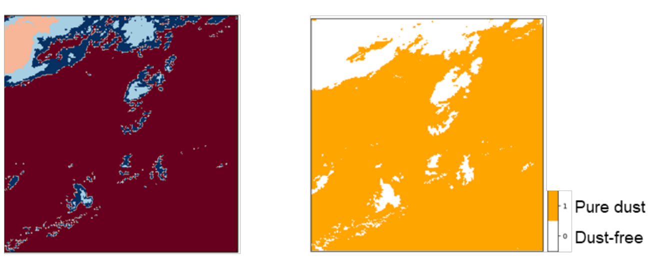


Figure 8: L-curve study of K-means clustering, the optimal number is around 4.

5.1 Dust extent extraction using K-means

As the first set of experiments, the K-means clustering is performed on 256*256 pixels images. Figure 9 show the (a) initial true colors images, the (b) RGB composite images, the (c) dust categories along CALIPSO-track, and (d) the segmentation results using K=4. The outcome predicted cluster using K-means is depicted in (e). It is seen that the dust extent in (e) coincides reasonably with the CALIOP pure dust in (c). To quantify such visual assessments, we summary the metrics in (f), which shows a reasonable accuracy (~70-80%) at determining pure dust (Category 1).

(a) True color composite (b) Dust composite (c) Dust categories on CALIOP track



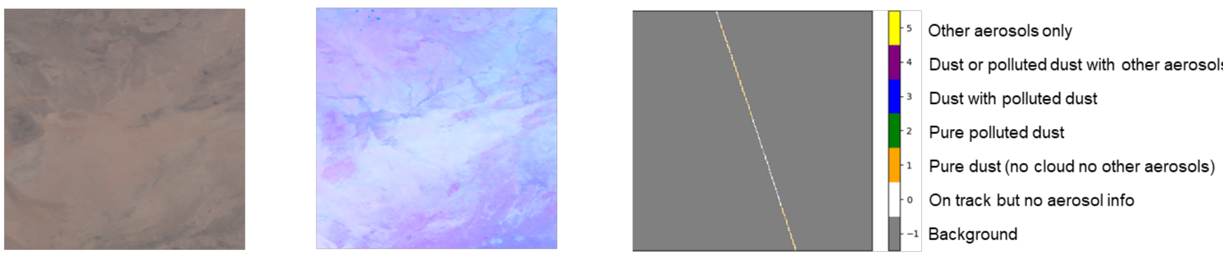
(e) Segmentation result (e) Resulting dust extent

	precision	recall	f1-score	support
Dust-free	1.00	0.41	0.58	137
Dust	0.59	1.00	0.74	118
Accuracy			0.68	255
Macro avg	0.80	0.70	0.66	255
Weighted avg	0.81	0.68	0.66	255

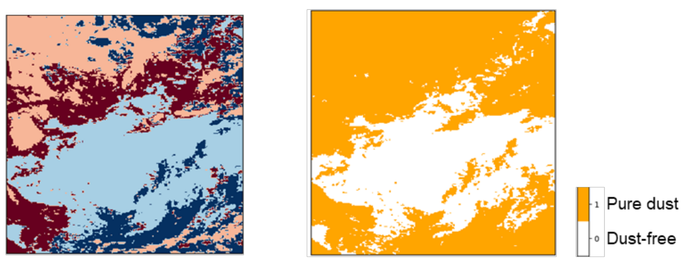
(f) On-track accuracy

Figure 9. (North Atlantic region) Composite images of VIIRS granule subset at 2014234t1724, dust categories on CALIPSO track, and resulting dust extents segmented from our methods.

Figures 10-11 show similar computations as Figure 9 but in two other regions. Together with Figure 9, these figures reveal a slightly different quality of clustering and accuracy, this might be because Figures 10 and 11 show an underlying land surface. Additionally, varying accuracy might be arising from the small spatial extent of the considered images. In the following sections, we propose different experiments to answer these questions.



(a) True color composite (b) Dust composite (c) Dust categories on CALIOP track

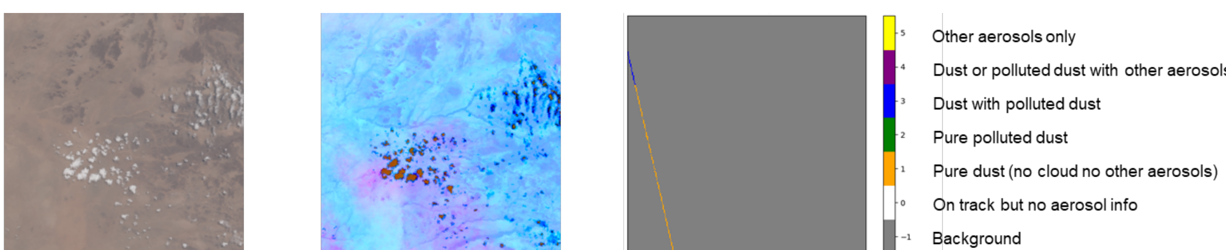


(d) Segmentation extent (e) Resulting dust extent

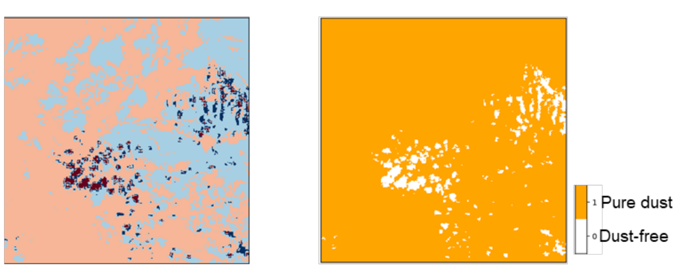
	precision	recall	f1-score	support
Dust-free	0.73	0.78	0.75	109
Dust	0.81	0.76	0.78	134
Accuracy			0.77	243
Macro avg	0.77	0.77	0.77	243
Weighted avg	0.77	0.77	0.77	243

(f) On-track accuracy

Figure 10. (Asian Spring region) Composite images of VIIRS granule subset at 2014147t0606, dust categories on CALIPSO track, and resulting dust extents segmented from our methods.



(a) True color composite (b) Dust composite (c) Dust categories on CALIOP track



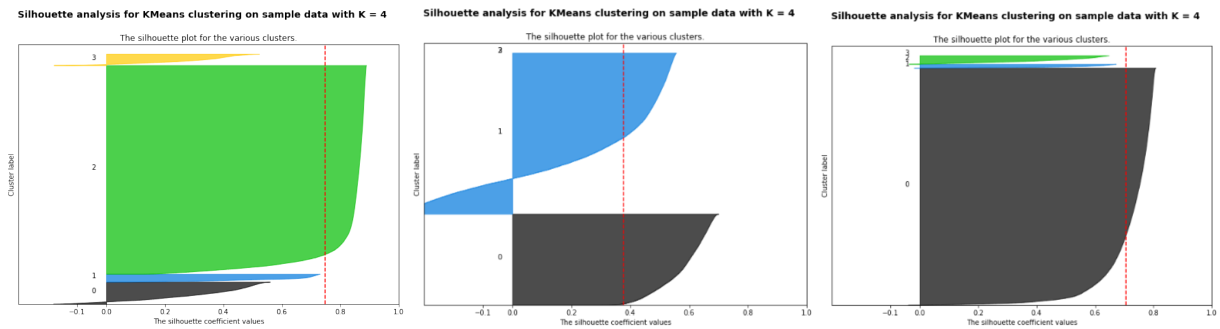
(d) Segmentation extent (e) Resulting dust extent

	precision	recall	f1-score	support
Dust-free	0.00	0.00	0.00	36
Dust	0.83	1.00	0.91	174
Accuracy			0.83	210
Macro avg	0.41	0.50	0.45	210
Weighted avg	0.69	0.83	0.75	210

(f) On-track accuracy

Figure 11. (Northern Africa Summer) Composite images of VIIRS granule subset at 2014147t0606, dust categories on CALIPSO track, and resulting dust extents segmented from our methods.

Additionally, we computed the silhouette coefficients to measure the quality of each cluster. The silhouette coefficient has 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 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. Figure 12 shows the average silhouette from all data-points is derived and the silhouette for each data-point is shown. The North Atlantic and North Africa regions examples show reasonably a good clustering property with an average silhouette significantly above 0.5. However, the Asian region example show a poor clustering property with many data-points wrongly assigned to clusters.



(a) North Atlantic region
Average silhouette: 0.749

(b) Asian Spring dust region
Average silhouette: 0.377

(c) North Africa region
Average silhouette: 0.707

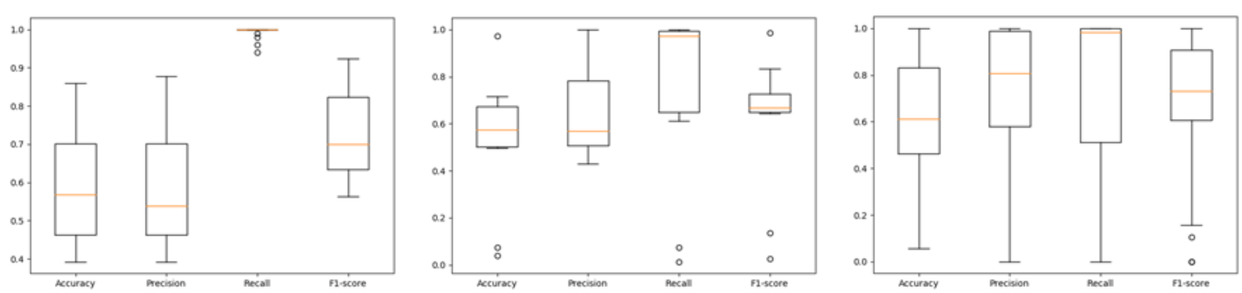
Figure 12. Silhouette coefficients the three examples in Figure 9-11

5.2 Average accuracy

In order to assess quantitatively the quality of the clustering, we calculated the mean values of accuracy metrics along the CALIPSO tracks regarding dust or dust-free. Three sets of experiments were conducted: Subsection 5.2.1) accuracy comparison among different study areas, Subsection 5.2.2) accuracy comparison over different surface types, and Subsection 5.2.3) accuracy comparison among K-means, K-medoids, and Fuzzy C-means.

5.2.1 Average accuracy using K-means within different study areas

Figure 13 displays the boxplots of the accuracy, precision, recall, and F1-score of all available images using the K-means over the datasets of three different study areas. It is observed that all three study regions have a median accuracy value around 0.6. Northern Africa summer study area shows a higher median precision (~0.8) over the other two study areas (~0.6). However, the Northern Africa summer study area generally has a wider range of accuracy values than the other two study areas.



(a) North Atlantic study area

(b) Asian dust study area

(c) Northern Africa summer study area

Figure 13. Box plots of accuracy, precision, recall, and F1-score for all the images in different study areas.

5.2.2 Average accuracy using K-means over different surface types

Figure 14 shows the box plots of the accuracy, precision, recall, and F1-score for all the images using K-means over different surface types. The proposed method performs better over barren with a precision of ~0.7, whereas the accuracy over water bodies and other surface types result in ~0.2.

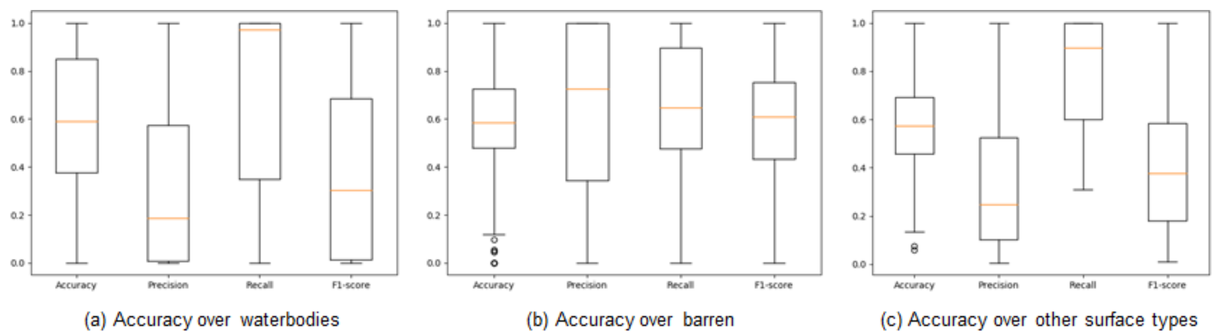


Figure 14. Box plots of accuracy, precision, recall, and F1-score for all the images over different surface types

5.2.3 Average accuracy using K-means, K-medoids, and Fuzzy C-means

Accuracy using different clustering methods, including K-means, K-medoids, and Fuzzy C-means did not show significant differences (Figure 15), therefore we continue our experiments using K-means.

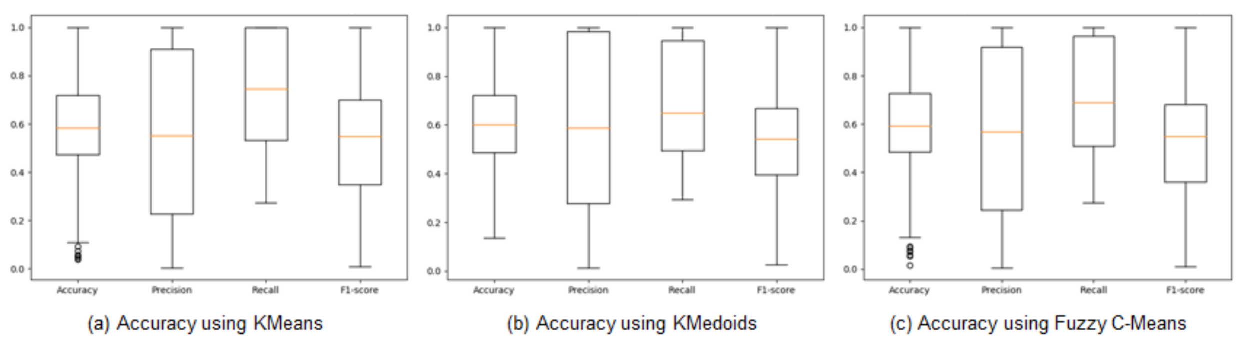


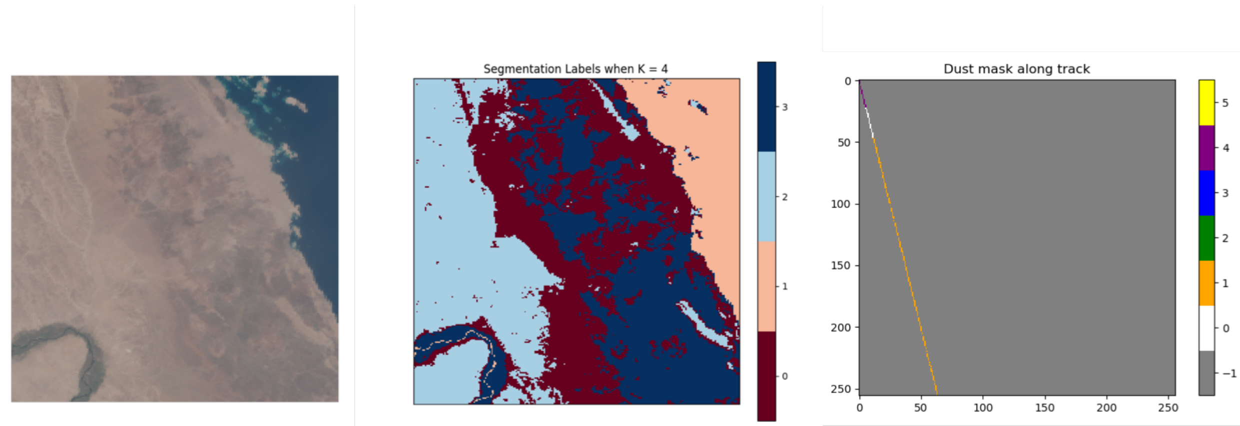
Figure 15. Box plots of accuracy, precision, recall, and F1-score for all the images using different clustering methods

5.3 K-means clustering on one single image using 3 VIIRS bands

We have been using all VIIRS radiative 16 bands so far in the above sessions. However, the importance of each band might not be the same. For example, we believe, by intuition, that the bands of M8, M9, M11 (cloud particle size information, cirrus cloud cover information, and clouds information) are more closely related to the dust presence.

In this subsection, we showed the results of using only three VIIRS bands, M8, M9, M11. Figure 16(a) shows the region of interest, Figure 15(b) shows the K-means segmentation using 4 clusters, Figure 16(c) displays the location of CALIPSO data which is used to validate the prediction. The corresponding accuracy table and confusion matrix are shown in Table 1. Figure 17 (a,b) compares the predicted dust regions using (a) 3 VIIRS bands, (b) 16 VIIRS bands. It is seen there is no significant difference between using 3 and 16 bands. When it comes to the (c) true dust composite image, the prediction seems to indicate the dust region, by eyeball.

The total number of possible combinations of these 16 bands is 16! ~= 10^13, which is inhibiting. A thorough investigation of the relationship between the dust presence and the VIIRS bands requires more knowledge on the bands and a smart design. We will save it for future work.



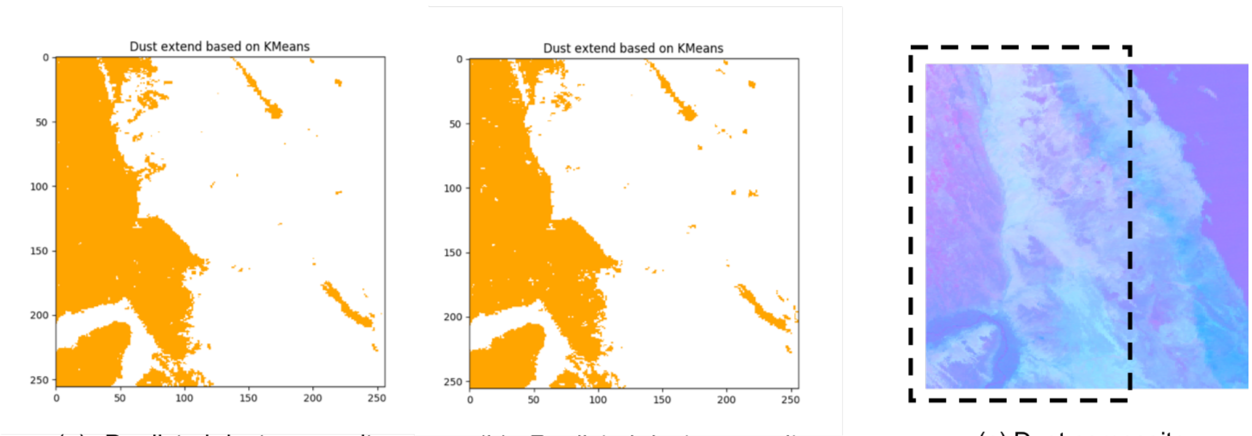
(a) VIIRS true color composite   (b) Kmeans segmentation, 3 bands   (c) CALIOP data along track

Figure 16. The test image, (a) VIIRS true color composite, (b) K-means segmentation (K=4), using 3 bands, (c) CALIOP data along the track.

K-Means	precision	recall	f1-score	support
0.0	0.00	0.00	0.00	47
1.0	**0.78**	0.81	0.79	199
macro avg	0.39	0.41	0.40	246
weighted avg	0.63	0.66	0.64	246

Confusion matrix	0	1
0	0	47
1	37	162

Table 1. Accuracy table and Confusion matrix, K-means, 3 bands.



(a) Predicted dust composite (using VIIRS 3 bands)

(b) Predicted dust composite (using VIIRS 16 bands)

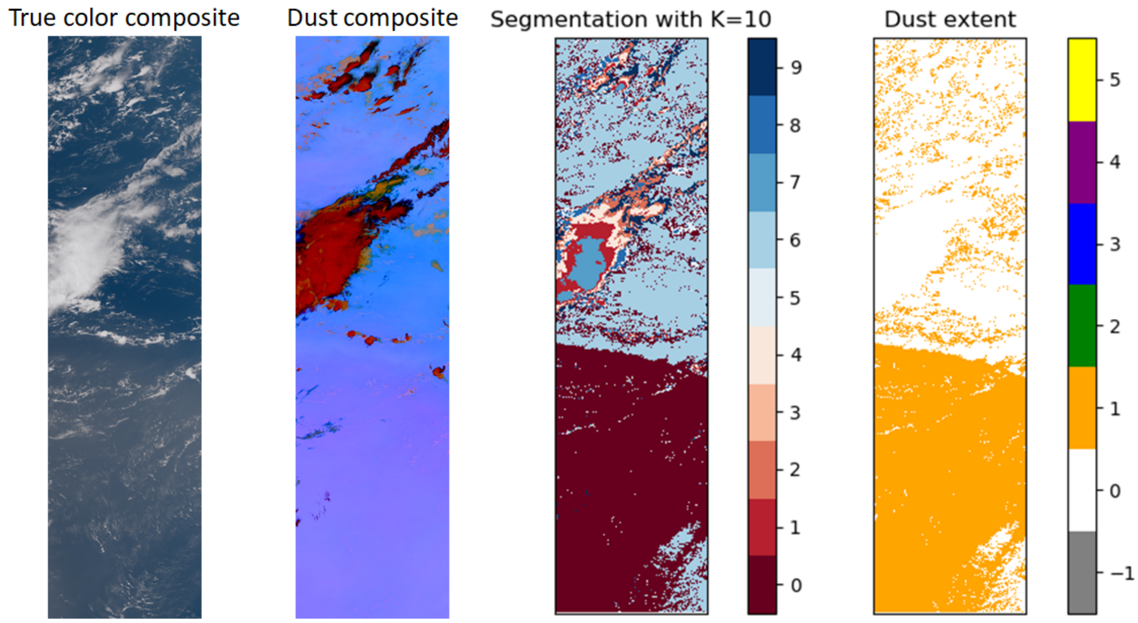
(c) Dust composite

Figure 17. The test image, comparing predicted results using (a) K-means, 3 bands, (b) K-means, 16 bands, c) dust composite image.

5.4 Experiment using larger VIIRS granule subset

We also tested the capability of our proposed method on a larger spatial scale. Instead of a 256*256-pixel matrix, we used the entire VIIRS subset collocated with the CALIPSO track and clustered the VIIRS subset using the proposed method. Generally, with a larger scale, the on-track accuracy improves. This accuracy improvement is expected because the sample size increases, and the dust is easier to detect as a mid-scale meteorological phenomenon.

The results of two experiments are shown below. By examining the composite images of VIIRS, the resulting dust extents are reasonable. Figure 17 shows the resulting dust extents are majorly in the lower half of the figure, which corresponds to the true color and dust composites. The accuracy scores at 0.71 and the mean silhouette scores at 0.3618. For the dust category, the precision of our method is 0.45, meaning the percentage of dust detected by our method that are truly dust; and the recall of our method is 0.69, meaning the percentage of observed dust that are correctly detected by our method.



Accuracy report

	precision	recall	F1-score	support
Dust-free	0.87	0.71	0.78	1952
Dust	0.45	0.69	0.55	668
accuracy			0.71	2620

Confusion matrix

Confusion matrix	Predicted - dust free	Predicted - dust
Actual - dust free	1386	566
Actual - dust	204	464

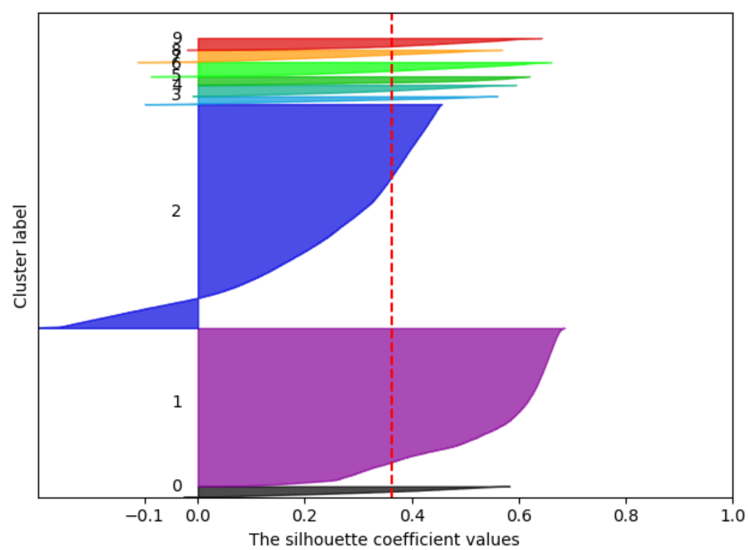
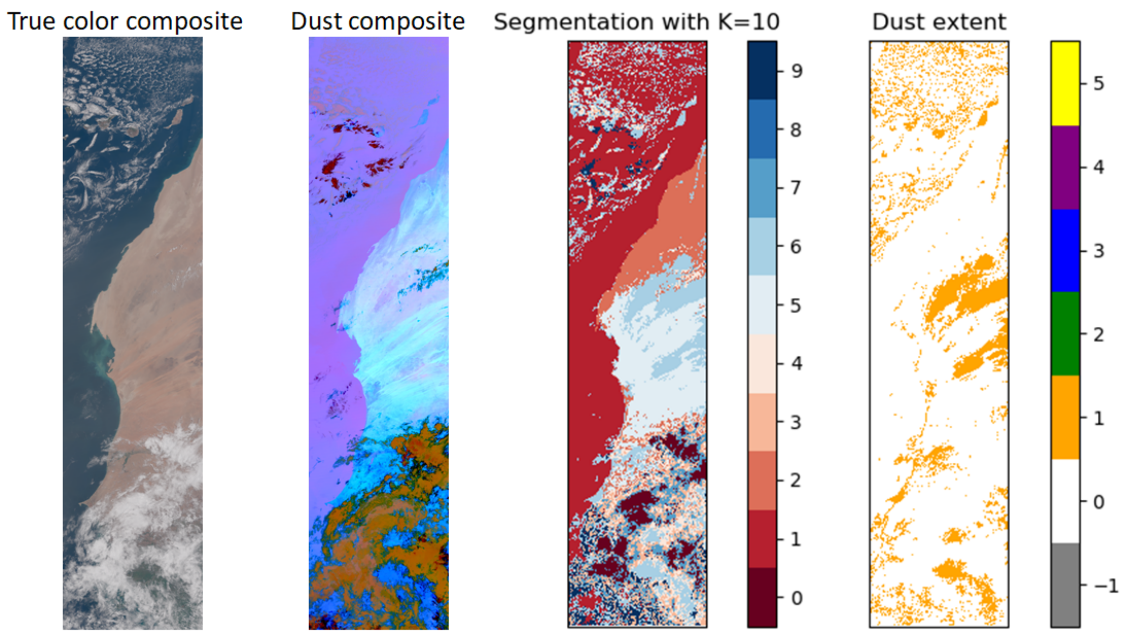


Figure 17. Composite images of VIIRS granule subset at 2014224t1712, resulting segmentation and dust extents segmented from our methods, and silhouette coefficient values for each cluster. Mean silhouette score: 0.3618

Figure 18 shows another dataset and the resulting dust extents are more scattered in different places. It is difficult to visually examine the dust extent from the true color and dust composites, but the on-track accuracy results in 0.73, and the mean silhouette scores at 0.4406. For the dust category, the precision of our method is 0.44, meaning the percentage of dust detected by our method that are truly dust; and the recall of our method is 0.19, meaning the percentage of observed dust that are correctly detected by our method. Among 678 observed dust pixels along CALIPSO track, only 127 of them were correctly clustered into the dust category. This is probably because this VIIRS imagery covers both ocean and land, and the different surface type impacts on the clustering result. We can observe the segmentation result having a clear distinction between ocean and land, indicating that our method is able to highlight the differences and similarities within the image. However, it lacks the capability of identifying dust clusters with surface type differences as the more significant patterns.



Accuracy report

	precision	recall	F1-score	support
Dust-free	0.77	0.92	0.84	1983
Dust	0.44	0.19	0.26	678
accuracy			0.73	2661

Confusion matrix

Confusion matrix	Predicted - dust free	Predicted - dust
Actual - dust free	1824	159
Actual - dust	551	127

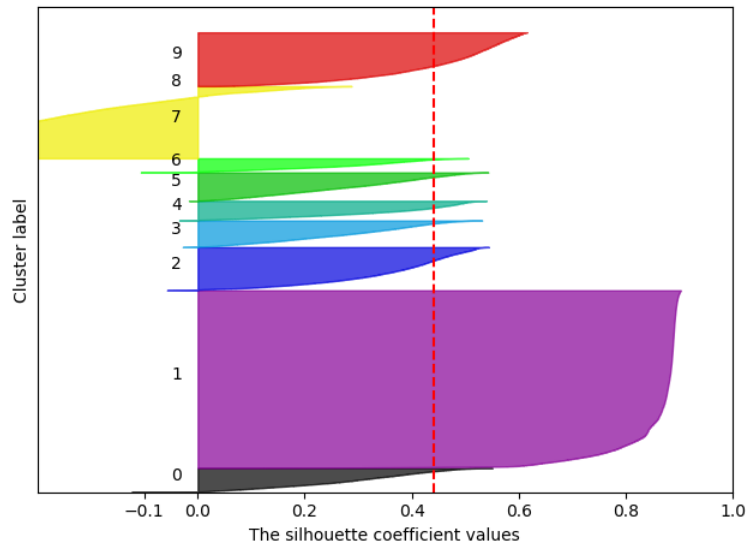


Figure 18. Composite images of VIIRS granule subset at 2014152t1436, resulting segmentation and dust extents segmented from our methods, and silhouette coefficient values for each cluster. Mean silhouette score: 0.4406.

6 Conclusions and future directions

In this project, we explored using unsupervised machine learning methods, especially K-means clustering, to identify dust extents from satellite imagery. We designed a workflow to extract existing dust profiles within spatiotemporal ranges based on CALIPSO dust profiles and used the dust profiles to select from the K-means clusters as the final dust extents. We examined the sensitivity of our method in different experiments, including average accuracy 1) in different study areas, 2) using different clustering methods, 3) using different combinations of VIIRS spectral bands. 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.

In future works, we will investigate semi-supervised techniques. Indeed, the access to the CALIOP data provides relevant information to guide the clustering; however, since these data are available along the CALIPSO track only, they cannot be used in a fully supervised setup for spatial clustering. In Wang et al. (2020), a semi-supervised clustering technique is proposed to segment land cover from remote sensing images. There are many additional variants of the setup of the proposed experiment that can be tested to improve the interpretation of the clusters and the accuracy of the classification. For instance, having a deeper knowledge of the band's characteristics for pure dust over large areas would help greatly.
We will also further validate the resulting dust extents by comparing it with other existing aerosol products. One of the products we tried in the experiments was VIIRS Aerosol EDR product (https://www.star.nesdis.noaa.gov/smcd/emb/viirs_aerosol/products_edr.php), but the product was not able to detect thin dust and failed to identify dust pixels in any of our validation time periods. Another possible product was the VIIRS Smoke/Dust Mask, but it does not produce days in 2014. We will continue to search for available products to further validate our results.

Acknowledgement

This work is supported by the grant "CyberTraining: DSE: Cross-Training of Researchers in Computing, Applied Mathematics and Atmospheric Sciences using Advanced Cyberinfrastructure Re- sources" 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.

References

Ackerman, S.A., 1997. Remote sensing aerosols using satellite infrared observations. *Journal of Geophysical Research: Atmospheres*, *102*(D14), pp.17069-17079.

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).

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.

Changjie Cai, Jangho Lee, Yingxi Rona Shi, Camille Zerfas, Pei Guo, and Zhibo Zhang. Dust Detection in Satellite Data using Convolutional Neural Networks. Technical Report HPCF-2019-15, UMBC High Performance Computing Facility, University of Maryland, Baltimore County, 2019.

Chaboureau, 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).

Darmenov, A. and Sokolik, I.N., 2005. Identifying the regional thermal-IR radiative signature of mineral dust with MODIS. *Geophysical Research Letters*, *32*(16).

Paul McClain, 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.

Pierangelo, C., Chédin, A., Heilliette, S., Jacquinet-Husson, N. and Armante, R., 2004. Dust altitude and infrared optical depth from AIRS.

Prata, A.J., 1989. Infrared radiative transfer calculations for volcanic ash clouds. *Geophysical research letters*, *16*(11), pp.1293-1296.

Prospero, J.M., 1999. Long-term measurements of the transport of African mineral dust to the southeastern United States: Implications for regional air quality. *Journal of Geophysical Research: Atmospheres*, *104*(D13), pp.15917-15927.

Lazri, M. and Ameur, S., 2018. Combination of support vector machine, artificial neural network and random forest for improving the classification of convective and stratiform rain using spectral features of SEVIRI data. *Atmospheric research*, *203*, pp.118-129.

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.

Middleton, N., 2019. Variability and trends in dust storm frequency on decadal timescales: Climatic drivers and human impacts. *Geosciences*, *9*(6), p.261.

Miller, S.D., 2003. A consolidated technique for enhancing desert dust storms with MODIS. *Geophysical Research Letters*, *30*(20).

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.

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.

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.

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), p.475.

Gonzalez, L. and Briottet, X., 2017. North Africa and Saudi Arabia day/night sandstorm survey (NASCube). *Remote Sensing*, *9*(9), p.896.

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.

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.

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.

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.

Wang, S., Chen, W., Xie, S. M., Azzari, G., & Lobell, D. B. (2020). Weakly Supervised Deep Learning for Segmentation of Remote Sensing Imagery. *Remote Sensing*, *12*(2), 207.