AN AEROSOL TYPE CLASSIFICATION METHOD BASED ON REMOTE SENSING DATA IN GUANGDONG, CHINA

Yangcheng Zheng [1, 2], Lili Li [3], Yunpeng Wang [1, *]

[1] Guangzhou Institute of Geochemistry, Chinese Academy of Sciences, Guangzhou, China – wangyp@gig.ac.cn
[2] University of Chinese Academy of Sciences, Beijing, China - zhengyangcheng17@mails.ucas.ac.cn
[3] College of Environmental Science and Engineering, Zhongkai University of Agricultural and Engineering, Guangzhou, China – lilili@gig.ac.cn

Commission III, WG III/8

ABSTRACT:

This paper provides an aerosol classification method based on remote sensing data in Guangdong, China in year 2010 and 2011. Aerosol Optical Depth, Angstrom Exponent and Ultraviolet Aerosol Index, as important properties of aerosols, are introduced into classification. Data of these three aerosol properties are integrated to establish a 3-dimension dataset, and k-means clustering algorithm with Mahalanobis distance is used to find out four clusters of the dataset, which respectively represents four aerosol types of urban-industrial, dust, biomass burning and mixed type. Prior knowledge about the understanding of each aerosol type is involved to associate each cluster with aerosol type. Temporal variation of the aerosol properties shows similarities between these two years. The proportion of aerosol types in different cities of Guangdong Province is also calculated, and result shows that in most cities urban-industrial aerosols takes the largest proportion while the mixed type aerosols takes the second place. Classification results prove that k-means cluster algorithm with Mahalanobis distance is a brief and efficient method for aerosol classification.

1. INTRODUCTION

Atmospheric aerosol particles can influence the global climate change not only by direct scattering and absorbing the solar radiation, but also by indirect ways like acting as cloud condensation nuclei (Ealo et al., 2018). Aerosols' diverse optical properties, component and size distribution can lead to different effects on climate (Chen et al., 2019). For example, dust aerosols have large size distribution and a scattering tendency, which can result in the reduction of visibility (Renard et al., 2018). Biomass burning aerosols have small size particles and an absorbing tendency because of their black carbon component (Hamill et al., 2016). The rapid increasing emissions of secondary aerosols caused by industrialization and urbanization can lead to an increase in air pollution and even extremely severe haze events, which have harmful impacts on environment, economic and health (Huang et al., 2014). Large quantities of aerosol particles produced by biomass burning can influence cloud microphysical or even disturb biogeochemical cycles (Crutzen et al., 1990). Aerosol type classification will help refine aerosol optical models and reduce uncertainties in satellite observations of the global aerosol and in modeling aerosol impacts on climate (Dubovik et al., 2002). Thus, making sense of the aerosol types is great of significance and further effort should be put on it.

Traditional methods that monitor and analyze the aerosol types mainly depend on ground-based observation sites, which are limited to space. And due to the long-distance transport of aerosols, sites' observation doesn't always demonstrate the real environment. Using remote sensing data to classify the aerosol types has the advantages of higher spatial and temporal resolution, and a consecutive series of historical data can be obtained. Kalapureddy et al. (2009) analyzed the aerosol types over the

Arabian Sea in the pre-monsoon season and found that a single aerosol type in each location can only exist under specific conditions while the presence of mixed type aerosols is more usual. Sreekanth (2013) analyzed an 8-year data set of aerosol properties in Bangalore, India and regarded that aerosol transformation processes were expected responsible for the mixed type aerosols. In addition, ground AERONET-based classification was also useful in interpreting aerosol acquisitions from remote sensing data (Hamill et al., 2016).

Studies about aerosol type classification are mainly based on aerosol properties which can be extracted from remote sensing data product. Aerosol Optical Depth (AOD) is an aerosol optical property which describes the aerosol's ability of extinction, while Angstrom Exponent (AE) is used to measure the size distribution of aerosol particles. If aerosol has a large size distribution, then its AE value is small, and a large value for small size distribution. Ultraviolet Aerosol Index (UVAI) is an optical property which describes the aerosol's absorbing ability. Absorbing aerosols yield positive UVAI values, while non-absorbing or weak-absorbing aerosols yield small negative UVAI values. Different types of aerosol's properties are distinguished with each other, making them possible to be characterized by remote sensing data. Patrick uses aerosol parameters of Angstrom Exponent, Single Scattering Albedo (SSA) and Refraction Index (RI) to classify the aerosol types. Kumar et al. (2018) established AOD-AE threshold model to classify the key aerosol type and calculate the frequency distribution in Yangtze River Delta, China. Chen et al. (2016) combines Aerosol optical Depth with Aerosol Relative Optical Depth (AROD) to classify the aerosol types. Traditional classification methods focused on giving a threshold value for each aerosol property, and could separate the dataset by linear plane or hyperplane. However, for those datasets with intensive

---

* Corresponding author.

inner structure, threshold values will not help. Clustering algorithm will be more suitable to discover the inner structure of the multi-dimension dataset so it is applied in this paper.

The research region in this study is Guangdong, a coastal province with the most rapid development of industrialization and urbanization in China. Guangdong has diverse types of aerosol and according to the emissions sources, aerosol types are divided as urban-industrial (UI), biomass burning (BB), mixed type and dust. These four aerosol types mainly include all the emission sources in Guangdong. A further understanding for aerosol types and their properties will help to make control of aerosols in Guangdong. Tan et al. (2016) studied the aerosol optical properties and the mixing state of black carbon in the Pearl River Delta of Guangdong and gave a description on black carbon's absorption ability. Yin et al. (2017) used numerical models to identify the source contributions to aerosols pollutants and industry sources were the major contributors in Guangdong. However, previous researches about aerosol types in Guangdong mainly concentrated on a particular type of aerosol or lacked for spatial and temporal resolution. Thus, aerosol classification with remote sensing data in Guangdong Province will make sense of the aerosol sources and the pollutants raised by aerosols.

This paper aims at giving a method for aerosol classification. Guangdong Province and time from 2010 to 2011 are chosen as research region and research period. Aerosol properties of Aerosol Optical Depth, Angstrom Exponent and Ultraviolet Aerosol Index from different data sources are integrated to establish a 3-dimension dataset for aerosol classification. K-means clustering algorithm with Mahalanobis distance is involved to find out four clusters of the dataset, which respectively represents four aerosol types of urban-industrial, dust, biomass burning and mixed type.

2. DATA AND METHOD

2.1 Data Acquisition

Aerosol properties of AOD, AE and UVAI are integrated to establish a 3-dimension dataset. AOD and AE data are obtained from the product MOD04_L2 of Moderate-Resolution Imaging Spectroradiometer (MODIS) with a spatial resolution of 10 km x 10 km. MOD04_L2 is a daily level-2 product which monitors the ambient aerosol properties. AOD data is from the subset 'Optical_Depth_Land_And_Ocean' at 0.55 micron while AE is from the subset 'Deep_Blue_Angstrom_Exponent_Land'. AE is calculated between 0.412/0.47 micron for bright surfaces and 0.47/0.65 micron for vegetated surfaces. UVAI comes from Ozone Monitoring Instrument (OMI) on the NASA's Aura spacecraft with a spatial resolution of 13 km x 24 km and temporal resolution of 98.8 minutes. It is a dataset that describes a multi-decadal Fundamental Climate Data Record (FCDR) of aerosol properties from a 32-year record of satellite near-UV observations. UVAI is a measure of the departure of the observed spectral dependence of the near-UV upwelling radiation at the top of the actual Earth surface-atmosphere system from that calculated for a hypothetical pure molecular atmosphere bounded at the bottom, which is quantitatively defined as:

$$UVAI = -100 log \frac{I_{obs}}{I_{cal}} \qquad (1)$$

where $\qquad I_{obs}$ = observed radiance
$\qquad\qquad\qquad I_{cal}$ = calculated radiance

Time series from January 2010 to December 2011 is chosen. And geolocation from 20°N to 26°N and 109°E to 118°E are restrained in order to cover the whole Guangdong Province. Each product is incomplete and has invalid data in different dates or regions, so matching up the pixels with valid AOD, AE and UVAI values in the same day is necessary to satisfy the training process. After the extraction, totally 3,064 pixels for year 2010 and 6263 pixels for year 2011 are integrated as the training data set. Training pixels are arranged in the format of vectors, and the ith pixel x_i in the dataset can be represented as:

$$x_i = [AOD_i, AE_i, UVAI_i] \qquad (2)$$

2.2 K-means Clustering Algorithm

The aerosol classification process is an unsupervised method because there is no label data for error calculation and correction. Thus, clustering algorithm is involved to find the implicit structure of intensive clusters in the multi-dimension data set without label data. Among several kinds of clustering methods, k-means clustering algorithm is the most common used for its simplicity and efficiency. The first process of training is to randomly generate the vectors **c** of 4 clusters, which correspond with 4 aerosol types to be classified in this study. For each training point $\mathbf{x}_i$, distance from $\mathbf{x}_i$ to each cluster $\mathbf{c}_j$ is calculated by turns and the cluster with minimum distance is chosen as the training point's classification result. An iteration is finished after classifying all the training points by once, and the new center vector $\mathbf{c}_j$ of each cluster are recalculated using the formula (3). The new j^{th} center **c** after an iteration is defined as:

$$c_j = [E(AOD), E(AE), E(UVAI)] \quad (j = 1, 2, 3, 4) \qquad (3)$$

where E means the average exception of each property.

In a new turn of iteration, a training point will be allocated to a new cluster if distance from this point to the new cluster's center is smaller than the former one. In iteration, the centers of each cluster and the number of points allocated to each cluster are dynamically changing. Do the iteration until the numbers of training points that change classification result is smaller than a certain number or times of iteration reach a limited value. After clustering, inner structure within the 3-dimension dataset will be found and training points will be allocated to each cluster.

2.3 Mahalanobis Distance

Common k-means clustering algorithm uses Euclidean distance to measure the distance between training point $\mathbf{x}_i$ and the cluster center $\mathbf{c}_j$, but Euclidean distance has the limit of solving the scale variation and correlation problems within different dimensions. AOD, AE and AI data are not in the same scale and the correlations among these parameters are not clear. For example, AOD represents the ability of extinction while UVAI represents the ability of absorption, so UVAI may be correlated with AOD. As a 0result, Mahalanobis distance is applied in this study instead of Euclidean distance. Mahalanobis distance D_M of two vectors can be calculated as followed:

$$D_M = (\boldsymbol{x} - \boldsymbol{c})^T \boldsymbol{S}^{-1} (\boldsymbol{x} - \boldsymbol{c})^{1/2} \qquad (4)$$

where $\qquad$ **S** = covariance matrix of AOD, AE and AI

The covariance matrix can account for the diffusiveness of a cluster and cross-correlations among different dimensions of a cluster. Compared with Euclidean distance, Mahalanobis

distance is scale-invariant and can efficiently eliminate the correlation of each dimension.

3. RESULTS AND DISCUSSIONS

3.1 Temporal Variation of Aerosol Properties

Average monthly values of AOD, AE and AI are calculated respectively to show the temporal variation in 2010 and 2011 (Figure 1). Results show that AOD data have the similar temporal characteristics between 2010 and 2011, with the highest value in spring (March in 2010 and April in 2011) (Figure 1a). Higher values of AE in autumn and winter indicate a smaller size distribution of aerosol (Figure 1b). UVAI values also reach largest in spring (March in 2010 and February in 2011), which means that aerosols consist of more absorbing component in these months.
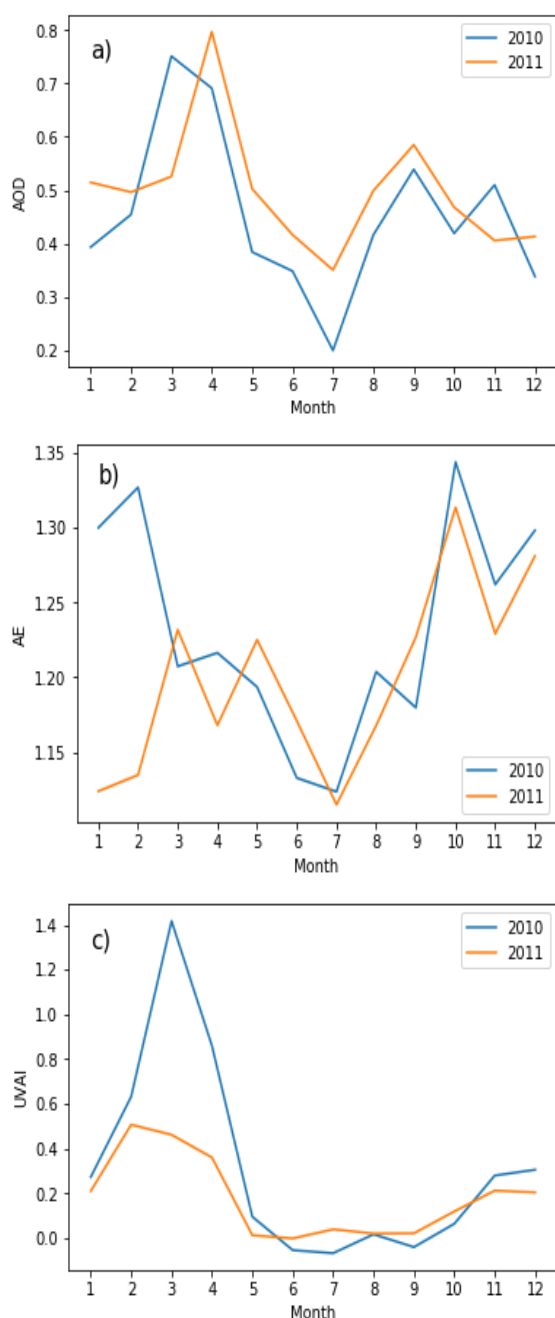

Figure 1. Temporal variation of 2010 and 2011, (a) AOD, (b) AE, (c) UVAI

Though the variation tendency between 2010 and 2011 is the same, there remains differences in absolute values. The reason may be that the aerosol properties' values shown in Figure 1 don't include pixels without complete value of AOD, AE or UVAI, so pixels from certain regions or dates are excluded. In addition, the choice of data sources and the algorithms used to inverse the data may contribute to the differences.

3.2 Results of K-means Clustering Algorithm

After the process the k-means clustering algorithm, the center of each cluster has been found out (Table 1). For the dimension of AOD, cluster 4 has the largest value approximately equal to 1 and the rest clusters have of values near to 0.5. For the dimension of AE, cluster 3 has the smallest value, which means to the aerosol type with the largest size distribution. Cluster 1's UVAI value is negative, while others' is positive, so aerosols with different ability of absorption are separated.

	AOD	AE	UVAI
2010			
Cluster 1	0.555	1.380	-0.301
Cluster 2	0.447	1.445	0.585
Cluster 3	0.510	0.484	0.332
Cluster 4	0.995	1.357	0.862
2011			
Cluster 1	0.445	1.459	-0.367
Cluster 2	0.526	1.372	0.616
Cluster 3	0.571	0.525	0.102
Cluster 4	1.075	1.540	0.765

Table 1. The center of each cluster in 2010 and 2011

Similar results are obtained between year 2010 and 2011. Figure 1 shows that the centers of clusters between 2010 and 2011 are located in the similar coordinates in the 3-dimension scatter plot, indicating that the similar clusters are achieved and the same classification logic can be used between 2010 and 2011.
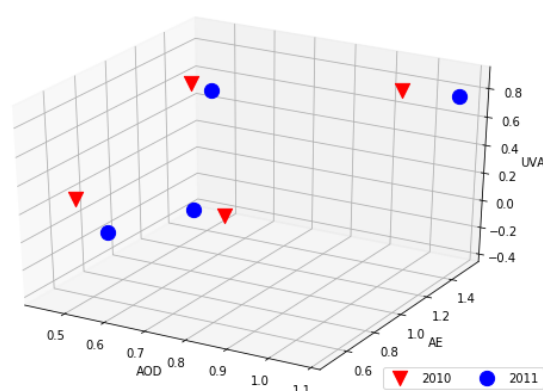

Figure 2. 3-dimension scatter plot of the centres of each cluster in year 2010 and 2011

Take the classification results of 2010 for a further analyze. Projecting the classification results of different clusters into 2-dimension scatter plot will give a more direct understanding (Figure 3). Figure 3a shows that cluster 3 can be easily distinguished with other aerosol types for its large size

distribution. Aerosols of cluster 2 and 4 can't be separated directly in Figure 3a because they have the similar value range in both AE and AOD dimensions. However, in the scatter plot of AOD and UVAI they can be easily characterized (Figure 3b). Figure 3c shows that cluster 3 and 4 can not only separated by AE values, but also by AOD values.
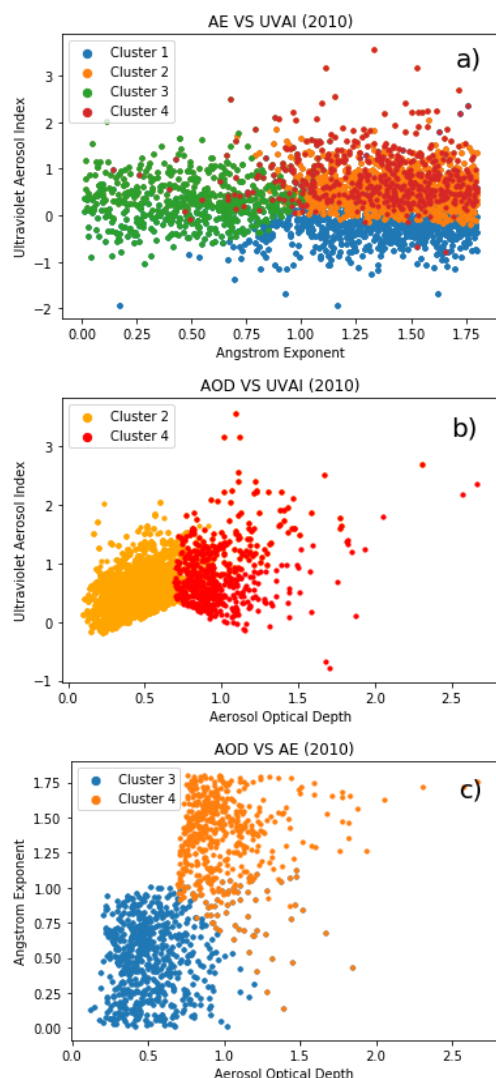

Figure 3. Scatter plots of different dimension combination in 2010, (a) AE VS UVAI, (b) AOD VS UVAI, (c) AOD VS AE

3.3 Aerosol Types Characterization

By far, aerosol clusters with different size distribution, ability of extinction and absorption are distinguished. But they should be associated with aerosol types and prior knowledge about aerosol type characterization should be involved. Previous studies have researched how to characterize the cluster's aerosol type. Hamill et al. (2016) used aerosols from particulate regions to define the reference aerosol types. Aerosols from desert are regarded as dust, with characteristic of large size distribution, while aerosols from Amazon forest are regarded as biomass burning's product, with small size distribution and high absorbing tendency, which means high angstrom exponent value and high aerosol index value. Chen et al. (2019) established an AOD-AE models and classify the aerosol types by their threshold. Pixels with 0.2<AOD<1.2 and 0<AE<0.5 is regarded as dust, while

0.5<AOD<1.2 and 1.5<AE<2.5 is regarded as urban-industrial or biomass burning.

Dust aerosols have a large size distribution, so cluster 3 is regarded as dust type with the smallest AE values. Biomass burning and urban industrial have similar AOD and AE, but biomass burning aerosol consists of more black carbon and has higher AI, so cluster 4 is regarded as biomass burning, and cluster 3 is regarded as urban-industrial. The last cluster is regarded as the mixed aerosol type, which means that the characteristic of this type doesn't match any type of aerosol in prior knowledge. After the clustering and characterization, four clusters with their corresponding aerosol types are achieved. When a new pixel with valid AOD, AE and UVAI values needs to be classified, calculate its Mahalanobis distance with each cluster center and the aerosol type with minimum distance will be allocated to this pixel.

3.4 Aerosol Partition in Guangdong Province

The partition of each aerosol type in cities of Guangdong Province is shown in the Table 2. Due to the lack of data, some important cities, such as Shenzhen, Dongguan, which have less than 100 training points, are not shown in this table for the result is not accurate.

Cities	Mixed	UI	Dust	BB
Shaoguan	27.5	55.3	10.7	6.5
Meizhou	25.6	62.7	7.5	4.2
Qingyuan	32.7	51.8	11.2	4.3
Heyuan	27.0	63.1	7.1	2.8
Chaozhou	20.9	66.5	4.7	7.9
Zhaoqing	28.5	55.4	5.0	11.1
Jieyang	34.6	56.3	3.5	5.6
Huizhou	33.1	57.8	3.6	5.5
Guangzhou	26.9	61.1	5.1	6.9
Shanwei	31.8	59.2	3.6	5.4
Yunfu	29.1	51.9	4.1	14.9
Jiangmen	37.7	44.3	3.7	14.3
Yangjiang	28.2	56.4	5.7	9.7
Maoming	25.6	55.5	7.3	11.6
Zhanjiang	30.3	43.8	6.7	19.2
Average	29.3	56.1	5.9	8.7

Table 2. The proportion of aerosol types in different cities of Guangdong Province (%)

Among the four aerosol types, urban-industrial takes an average of 56.1%, mixed type 30.3%, dust 6.7% and biomass burning 16.7%. Guangdong is a province with highly urbanization and industrialization, so urban-industrial aerosol taking the most part consists with the facts. Mixed type aerosols may consist of marine aerosols, biogenic aerosols from nature sources and others, which results in mixed aerosol properties and difficulties in recognition. Biomass burning aerosols, the only cluster with largest size distribution and strong absorption tendency, is expected to be responsible for the increase of Angstrom Exponent in winter season and increase of Ultraviolet Aerosol Index in spring season. The requirement for fuel combustion in winter and straw burning on cropland in spring may contribute to the increase of biomass burning.

3.5 Discussion

In the process of training, the initial centres of the four clusters are given by random, but different random initial centres can reach to the same classification results, which means that the

training process of k-means clustering algorithm is convergent in this study and this cluster algorithm is suitable for aerosol classification. For further study, other aerosol optical properties such as Single Scattering Albedo and Refraction Index can be brought in to establish a data set with more dimensions. Other classification methods like Support Vector Machine (SVM) also can be used to make a cross validation with k-means clustering algorithm.

Marine aerosols are not taken into consideration in this paper. Previous observation gave marine aerosols a strong absorption ability (UVAI>1), but the cluster with largest UVAI values obviously should be biomass burning type for its small size distribution. Time variation of UVAI also corresponds with biomass burning events in Guangdong province. In addition, aerosols with large UVAI values are distributed not only in coastal cities like Zhanjiang, Jiangmen, but also in inland cities like Zhaoqing. Confusion about the distribution of marine aerosols makes it difficult to be recognized. Another confusion is that mixed type aerosols' UVAI value is negative, which means weak absorption or non-absorption aerosols. This may be accounted for the change of aerosol' properties after mixing, or mixed type aerosols consist of another aerosol type not mentioned in this paper, which both need further research.

4. CONCLUSIONS

This paper provides a method for aerosol classification based on remote sensing data in Guangdong, China. Aerosol properties of Aerosol Optical Depth, Angstrom Exponent and Ultraviolet Aerosol Index are used for classification and with k-means clustering algorithm, four clusters are found out. Combined with prior knowledge, the clusters are associated with aerosol type respectively. After training, when classifying a new pixel with valid AOD, AE and UVAI values, allocate it to the aerosol type with minimum Mahalanobis distance. The similar classification results between year 2010 and 2011 prove that k-means cluster algorithm with Mahalanobis distance is a brief and efficient method for aerosol classification. Proportion of each aerosol type in Guangdong is also calculated and result shows that in most cities urban-industrial aerosols takes the largest proportion while the mixed type aerosols takes the second place, which consists with the facts.

REFERENCES

Chen, Q.X., Shen, W.X., Yuan, Y., Tan, H.P., 2019. Verification of aerosol classification methods through satellite and ground-based measurements over Harbin, Northeast China. *Atmospheric Research*, 216, 167-175.

Chen, Q.X., Yuan, Y., Shuai, Y., Tan H.P., 2016. Graphical aerosol classification method using aerosol relative optical depth. *Atmospheric Environment*, 135, 84-91.

Crutzen, P.J., Andreae, M.O., 1990. Biomass Burning in the Tropics: Impact on Atmospheric Chemistry and Biogeochemical Cycles. *Science*, 250, 1669-1678.

Dubovik, O., Holben, B., Eck, T.F., Smirnov, A., Kaufman, Y.J., King, M.D., Tanre, D., Slutsker, I., 2001. Variability of Absorption and Optical Properties of Key Aerosol Types Observed in Worldwide Locations. *Journal of the Atmospheric Sciences*, 59, 590-608.

Ealo, M., Alastuey, A., Perez, N., Ripoll, A., Querol, X., Pandolfi, M., 2018. Impact of aerosol particle sources on optical properties in urban, regional and remote areas in the north-western Mediterranean. *Atmospheric Chemistry and Physics*, 18, 1149-1169.

Hamill, P., Giordano, M., Ward, C., Giles, D., Holben, B., 2016. An AERONET-based aerosol classification using the Mahalanobis distance. *Atmospheric Environment*, 140, 213-233.

Huang, R.J., Zhang, Y.L., Bozzetti, C., Ho, K.F., Cao, J.J., Han, Y.M., Daellenbach, K.R., Slowik, J.G., Platt, S.M., Canonaco, F., 2014. High secondary aerosol contribution to particulate pollution during haze events in China. *Nature*, 514, 218-222.

Kalapureddy, M.C.R., Kaskaoutis, D.G., Raj, P.E., Devara, P.C.S., Kambezidis, H.D., Kosmopoulos, P.G., Nastos, P.T., 2009. Identification of aerosol type over the Arabian Sea in the pre-monsoon season during the Integrated Campaign for Aerosols, Gases and Radiation Budget (ICARB). *Journal of Geophysical Research*, 114, D17203.

Kumar, K.R., Kang, N., Yin, Y., 2018. Classification of key aerosol types and their frequency distributions based on satellite remote sensing data at an industrially polluted city in the Yangtze River Delta, China. *International Journal of Climatology*, 38, 320-336.

Renard, J.B., Dulac, F., Durand, P., Bourgeois, Q., Denjean, C., Vignelles, D., Coute, B., Jeannot, M., Verdier, N., Mallet, M., 2018. In situ measurements of desert dust particles above the western Mediterranean Sea with the balloon-borne Light Optical Aerosol Counter/sizer (LOAC) during the ChArMEx campaign of summer 2013. *Atmospheric Chemistry and Physics*, 18, 3677-3699.

Sreekanth, V., 2014. On the classification and sub-classification of aerosol key types over south central peninsular India: MODIS–OMI algorithm. *Science of the Total Environment*, 468-469, 1086-1092.

Tan, H.B., Liu, L., Fan, S.J., Li, F., Yin, Y., Cai, M.F., Chan, P.W., 2016. Aerosol optical properties and mixing state of black carbon in the Pearl River Delta, China. *Atmospheric Environment*, 131, 196-208.

Yin, X.H., Huang, Z.J., Zheng, J.Y., Yuan, Z.B., Zhu, W.B., Huang, X.B., Chen, D.H., 2017. Source contributions to PM2.5 in Guangdong province, China by numerical modelling: Results and implications. *Atmospheric Research*, 186, 63-71.