**Validation and Error Analysis for A Deep Learning Aerosol Retrieval Method**

Xingfeng Chen (1), Fengjie Zheng(2), Gerrit de Leeuw (1)(3), Kaitao Li (1), Zhengqiang Li (1)

1 Aerospace Information Research Institute, Chinese Academy of Sciences, 20 Datun road, Chaoyang District, Beijing, 100101, China

2 School of Space Information, Space Engineering University, Beijing, 101416，China

3 Royal Netherlands Meteorological Institute (KNMI), R&D Satellite Observations, 3731GA De Bilt, The Netherlands

Email: chenxf@aircas.ac.cn; zhengfengjie84@163.com

**KEY WORDS:** Deep Learning; Aerosol Retrieval; Neural Network; Validation; Fine Mode Fraction

**ABSTRACT:** The retrieval of multiple aerosol parameters from satellite remote sensing data is a difficult task. A deep learning method named NNAero (Neural Network based AEROsol retrieval) trained by MODIS (MODerate resolution Imaging Spectroradiometer) and AERONET (Aerosol Robotic Network) data was used to produce 10 years (2010-2019) Aerosol Optical Depth (AOD) and Fine Mode Fraction (FMF) data over eastern China. The NNAero AOD and FMF productions were validated using AERONET (Aerosol RObotic NETwork) and SONET (Sun–Sky Radiometer Observation Network) data, showing the good performance of the NNAero FMF retrieval (68% within Expected Error of ± (0.03 + 20%)) compared to previous studies. The evaluation of the NNAero AOD using MODIS Deep Blue AOD shows good spatial consistency. Validation results were analyzed to determine retrieval error sources, showing that aerosol absorption and scattering and surface brightness jointly contribute to the FMF retrieval error.

# INTRODUCTION

A number of satellites has been launched for atmospheric aerosol monitoring due to the important impact of aerosols on climate change, environment and human health (Kaufman et al., 2002; IPCC,

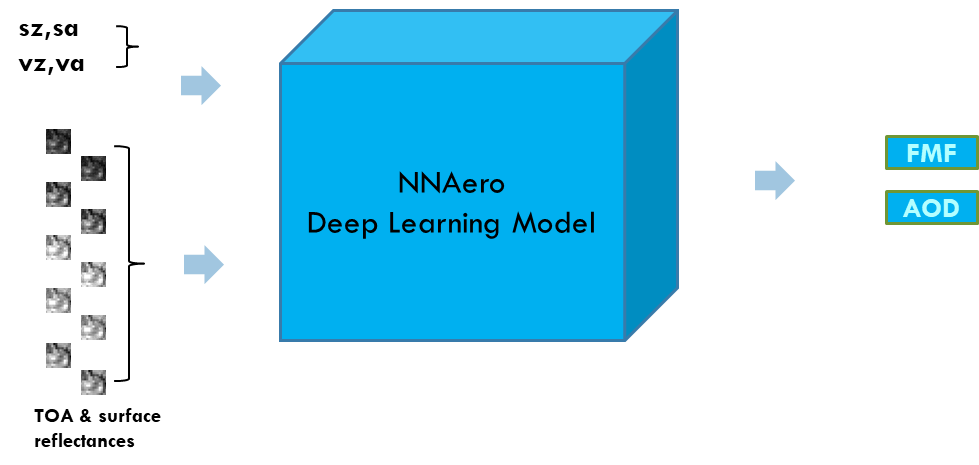
2013). Methods such as Dark Target (DT, Levy et al., 2013), Deep Blue (DB, Hsu et al., 2013) and Grasp (Dubovik et al., 2011) have been developed to retrieve global aerosol remote sensing products, which contribute to the global radiation budget calculation and air pollution monitoring.

The Aerosol Optical Depth (AOD) has been retrieved using a variety of satellite-based sensors. Aerosol Fine Mode Fraction (FMF) is defined as the fraction of fine mode particles in the total aerosol particles contributing to the AOD. But FMF is hard to retrieve over land with as high an accuracy as AOD using multi-spectral measurements only. Some efforts for FMF retrieval have been made using radiative method (Yan et al., 2017; Levy et al., 2013), but the accuracy is still not good. In order to figure out whether and how the satellite multi-spectral sensors, in particular the Moderate Resolution Imaging Spectroradiometer (MODIS), can retrieve FMF, an artificial Neural Network based AEROsol retrieval (NNAero) method to jointly retrieve FMF and AOD using MODIS data was developed by Chen et al. (2020). The NNAero-retrieved FMF accuracy was been improved significantly compared to previous studies.

For long-term and comprehensive validation for NNAero, FMF and AOD were retrieved for a period of 10 years (2010-2019) over north and east China, encompassing a wide variety of aerosol and surface conditions. AERONET (Holben et al, 2001) and SONET (Li et al., 2018) ground-based observations (some sites are shared) were used to validate the NNAero satellite products. In this paper, we present the validation and error analysis results.

# NNAero METHOD

The NNAero method was described in detail in Chen et al., (2020). In brief, the deep learning model NNAero was trained using TOA (Top Of the Atmosphere) and surface spectral reflectances observed by MODIS as input data, the FMF and AOD measured by AERONET/SONET were used as output data. Also, the solar and viewing angles were included in the input data. After the model was trained, FMF and AOD can be retrieved (or predicted) using the NNAero neural network model, which only uses MODIS data as input. The strategy is shown in Fig. 1.



**Figure 1.** The multi-input neural network architecture of MODIS FMF and AOD prediction. The inputs include TOA and surface reflectance images (input1) and four geometric angles (input2: sz for solar zenith, sa for solar azimuth, vz for viewing zenith and va for viewing azimuth). Method details can be found in Chen et al., (2020).

Five spectral bands of MODIS data were selected as shown in Table 1. The spatial resolution is 0.5 km as the reflectance is extracted from “MOD02HKM” (TOA radiance, is converted to TOA reflectance) and “MOD09” (surface reflectance) datasets. Cloud mask is acquired from MOD35 data.

**Table 1.** Spectral bands selected from MODIS data

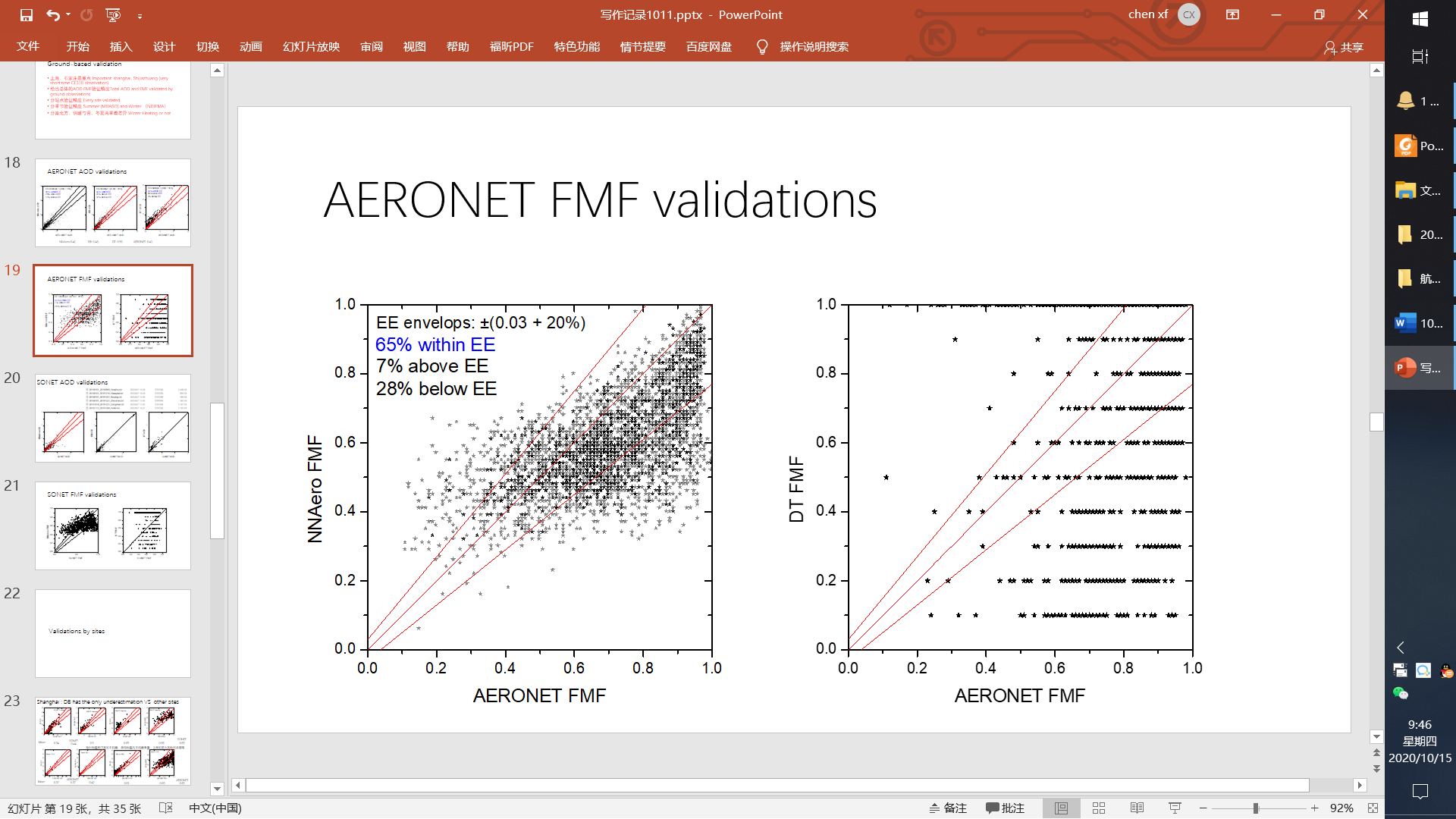
|  |  |
| --- | --- |
| Serial Number of band | Wavelength (nm) |
| 1 | 620—670 |
| 2 | 841—876 |
| 3 | 459—479 |
| 4 | 545—565 |
| 7 | 2105—2135 |

# AERONET VALIDATION

Chen et al., (2020) used several years of data from nine AERONET sites over a rather small area for independent validation, showing that 68% of the FMF values are within the Expected Error of ± (0.03 + 20%), and the AOD values are within 68% of the EE ± (0.05 + 15%). Here we validate the NNAero FMF and AOD using long-term data during 2010-2019 (10 years).

# NNAero FMF Vallidation

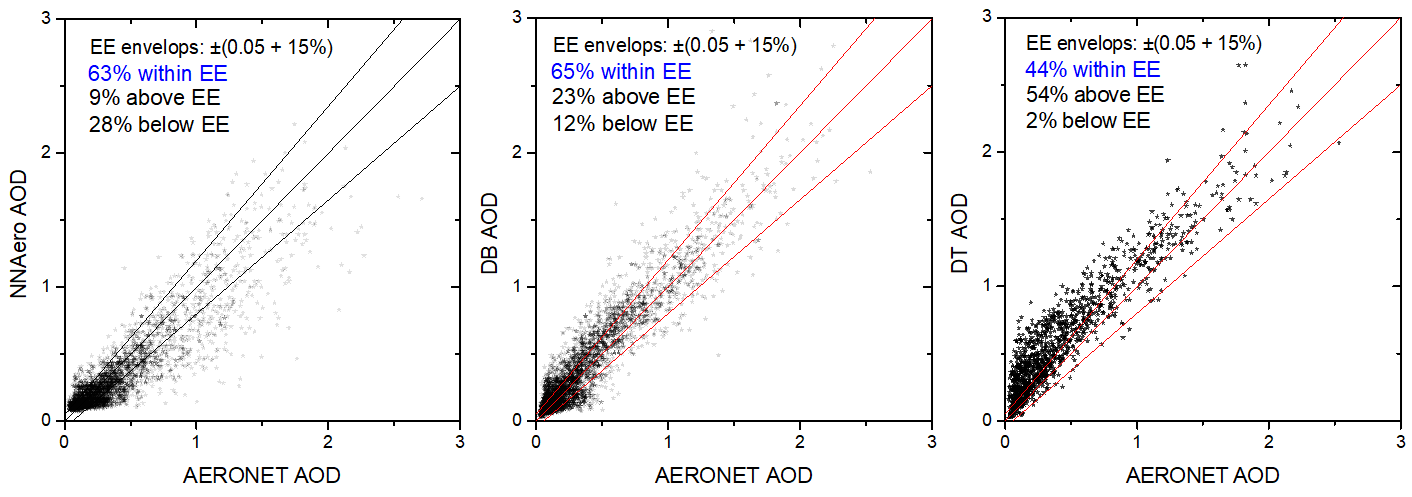
Fig.2 shows scatterplots of the NNAero and DT FMF versus AERONET and SONET FMF, for all 10 years of data. The left figure shows that 65% of the NNAero FMF data is within the EE ± (0.03 + 20%), which is slightly less good than the primary validation which was 68%, but over a small area. The right part of Fig.2 shows that DT FMF is discrete and the values are almost all too small as compared to the ground-based values.



**Figure 2**. MODIS NNAero FMF and DT FMF validations against AERONET and SONET FMF using ten years of data (2010-2019). The AERONET data include all available measurements during 2010-2019 from sites: Beijing\_RADI, Xianghe, SONET\_Xingtai, Xinglong, Xuzhou, Shijiazhuang\_SZF and Shijiazhang\_CHEY, Taihu. The data level and processing method is described in Chen et al., 2020.

# NNAero AOD Validation

Fig.3 shows that for NNAero AOD 63% of the values are within the EE of ± (0.05 + 15%). The DB AOD has the best accuracy with 65% within EE, while the DT AOD has the lowest accuracy with only 44% within EE. The difference between the primary validation in Chen et al., 2020 may be caused by using different AERONET dataset with ten years’ time. Overall, the NNAero AOD is underestimated.



**Figure 3**. Scatterplots of MODIS NNAero AOD, DB AOD and DT AOD against AERONET AOD using ten years data (2010-2019) of data.

# Averaged Fine Mode AOD Validation

The AODf (defined by Eq.1) of both satellite and ground observations were averaged over all 10 years (2010-2019) and the results are compared in Fig. 4. The averaged AODf data from 4 SONET sites, Shanghai, Nanjing, Hefei and Songshan, for the years 2016-2019 were also added. Observations during a very short time (such as a week) in total at one site were not used to calculate the mean AODf.

(1)



**Figure 4**. Comparison of averaged AODf at different AERONET/SONET sites (red dot curve) with 10-year averaged satellite observations (black squares) using MODIS NNAero AODf.

As shown in Fig. 4, the variation of the MODIS NNAero AODf over the different sites is similar to that of the AERONET ground-based AODf. The AODf is related to the fine mode particles originating from anthropogenic sources and precursor gases, and has more information on PM2.5 than AOD.

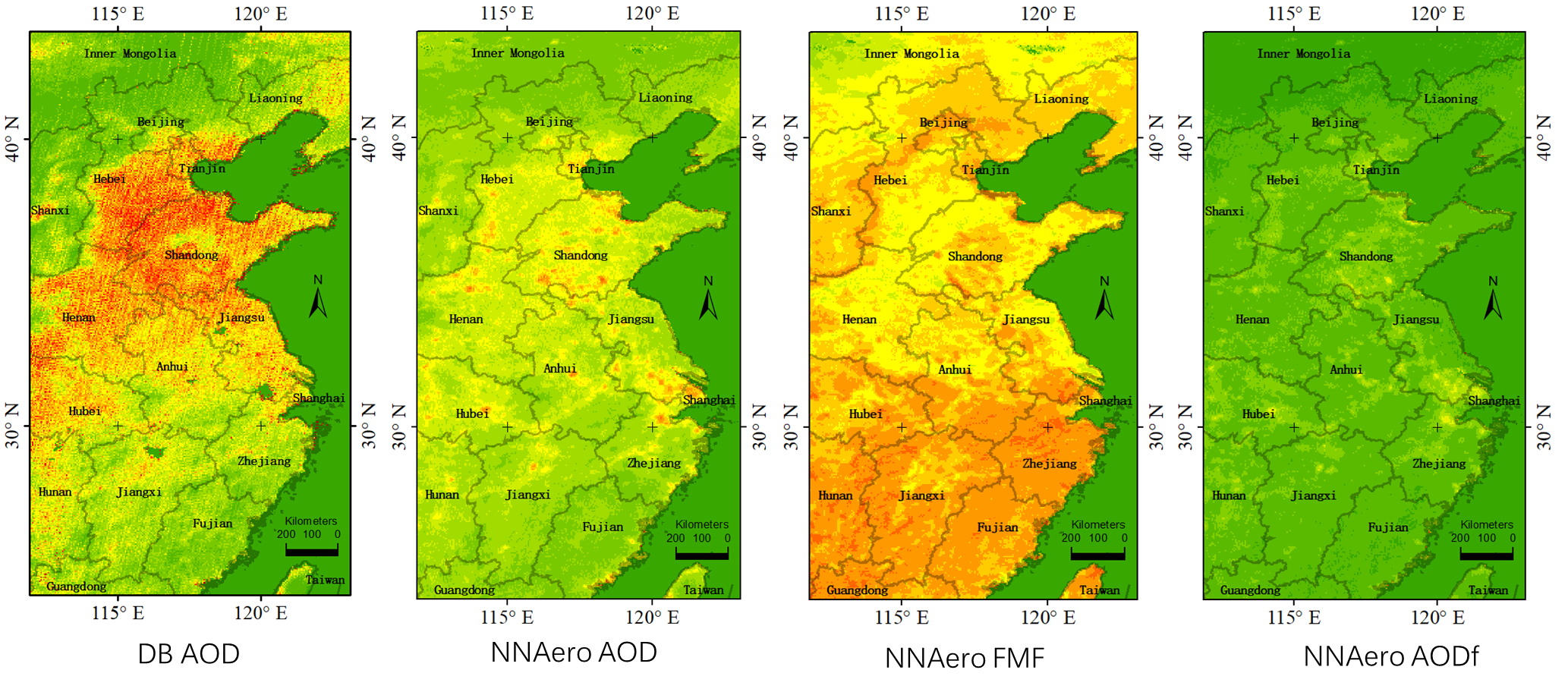
# Error Analysis

The primary error analysis has been introduced in Chen et al., 2020. The contributions of aerosol amount (AOD value), aerosol absorption and different satellites to errors in NNAero FMF and AOD were analyzed. The coupling effect of aerosol absorption and FMF has been found to be an important error source.

Here we extracted the TOA reflectance at the 2130 nm wavelength band, and set 0.05, 0.1 and 0.15 thresholds for validation. The NNAero FMF accuracy decreases with increasing reflectance.

# COMPARISON OF SATELLITE PRODUCTS

The NNAero AOD, FMF and AODf provide much information about aerosol pollution. MODIS data were processed using NNAero over the eastern and northern part of China. The 2010’s annual mean distributions of DB AOD, NNAero AOD, NNAero FMF, NNAero AODf are presented in Fig.5.

****

**Figure 5**. 2010’s annual mean distributions of DB AOD, NNAero AOD, NNAero FMF, NNAero AODf over the eastern and northern parts of China.

AOD and AODf are correlated with aerosol pollution. The data in Fig. 5 show that NNAero AOD is underestimated compared with DB AOD, but the spatial distributions are consistent. The northern China plain and Yangtze River delta areas are heavily polluted. South China and inner Mongolia have better air quality. The NNAero AODf shows some difference between rural and urban, more studies jointly using population and industry data may reveal more information.

# DISCUSSION AND CONCLUSION

After the validation, we found some problems requiring further research:

1. The coupling effect between multiple aerosol parameters introduces error. In this case, the decoupling strategy should be studied. Considering decoupling physics’ mechanism, multiplexing neural network combination or decision tree should be tested.
2. The neural network structure may introduce error. Two kinds of sites were compared: one set which was included in the training dataset and the other one not. The sites included in training have better accuracy. This may be caused by the Convolutional Neural Network (CNN) structure can remember some shape and texture information. The Full Connected Neural Network and other structure should be tested.
3. The samples should be reprocessed to get a better distribution. In the real atmosphere, low FMF cases (<0.2) are rare compared with high FMF, so the “Machine Learning” learned more from samples of high FMF value. The sample augmentation for a uniform distribution may decrease the approximation error.

Totally, the NNAero FMF and AOD were validated, and get high accuracies. The NNAero AOD may have some underestimation, so we do not recommend to use it for quantitative applications and analysis. The NNAero FMF is an obvious improvement compared with previous methods.

The improvement of the NNAero method is further pursued and expected to provide better accuracy, wider coverage and more aerosol product types.

# References

Chen, X., Leeuw, G. D., Arola, A., Liu, S., Li, Z., Zhang, K., 2020. Joint retrieval of the aerosol fine mode fraction and optical depth using MODIS spectral reflectance over northern and eastern china: artificial neural network method. Remote Sensing of Environment, 249, pp.112006.

Dubovik, O., Herman, M., Holdak, A., Lapyonok, T., Tanré, D., Deuzé, J. L., Ducos, F., Sinyuk, A., and Lopatin, A.: Statistically optimized inversion algorithm for enhanced retrieval of aerosol properties from spectral multi-angle polarimetric satellite observations, Atmos. Meas. Tech., 4, 975-1018, 2011.

Holben, B.N., Tanré, D., Smirnov, A., Eck, T.F., Slutsker, I., Abuhassan, N., Newcomb, W.W., Schafer, J.S., Chatenet, B., Lavenu, F., Kaufman, Y.J., Castle, J.V., Setzer, A., Markham, B., Clark, D., Frouin, R., Halthore, R., Karneli, A., O'Neill, N.T., Pietras, C., Pinker, R.T., Voss, K., Zibordi, G., 2001. An emerging ground-based aerosol climatology: Aerosol optical depth from AERONET. J. Geophys. Res. Atmos. 106 (D11), pp.12067–12097.

Hsu, N.C., Jeong, M.J., Bettenhausen, C., Sayer, A.M., Hansell, R., Seftor, C.S., Huang, J., Tsay, S.C., 2013. Enhanced deep blue aerosol retrieval algorithm: the second generation. J. Geophys. Res. Atmos. 118 (16), pp.9296–9315.

IPCC, 2013. Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY. From <https://doi.org/10.1017/CBO9781107415324>.

Kaufman, Y.J., Tanre, D., Boucher, O., 2002. A satellite view of aerosols in the climate system. Nature 419, pp.215–223.

Levy, R.C., Mattoo, S., Munchak, L.A., Remer, L.A., Sayer, A.M., Patadia, F., Hsu, N.C., 2013. The collection 6 MODIS aerosol products over land and ocean. At. Meas. Tech. 6, pp.2989–3034

Li, Z.Q., Xu, H., Li, K.T., Li, D.H., Xie, Y.S., Li, L., Zhang, Y., Gu, X.F., Zhao, W., Tian, Q.J., Deng, R.R., Su, X.L., Huang, B., Qiao, Y.L., Cui, W.Y., Hu, Y., Gong, C.L., Wang, Y.Q., Wang, X.F., Wang, J.P., Du, W.B., Pan, Z.Q., Li, Z.Z., Bu, D., 2018. Comprehensive study of optical, physical, chemical, and radiative properties of total columnar atmospheric aerosols over China: an overview of sun–sky radiometer observation network (SONET) measurements. Bull. Am. Meteorol. Soc. 99 (4), pp.739–755.

Yan, X., Li, Z., Shi, W., Luo, N., Wu, T., Zhao, W., 2017. An improved algorithm for retrieving the fine-mode fraction of aerosol optical thickness, part 1: algorithm development. Remote Sens. Environ. 192, pp.87–97