中国科学院大气物理研究所2023年度学术年会

——innovation、association、progress——

A Machine Learning-based Cloud Detection Algorithm for FY-4A AGRI

Meijia Xiao　Disong Fu　Hongrong Shi　Guocheng Wang　Xiangao Xia

Institute of Atmospheric Physics， Chinese Academy of Sciences, Beijing,100020

Abstract： **Cloud Masking is one of the most essential products for satellite remote sensing and downstream applications. This study develops machine learning-based (ML-based) cloud detection algorithms using spectral observations for the Advanced Geosynchronous Radiation Imager onboard the FY 4A geostationary satellite.Collocated active observations from Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) are used to provide reference labels for model development and validation. In this study, 26 models were built based on whether the solar altitude angle was greater than 90°, different types of subsurface(water,vegetation,bare,ice), and different seasons (spring,summer,autumn,winter), with eligible snow and ice samples counted throughout the year because of the small number of samples. AutoML automatically searches for the optimal parameters for the 26 models, which greatly saves the time for parameterization. We also compared the performance of this paper's algorithmic model with that of CLM, the official cloud detection product of Feng Yun, for each of the five evaluation metrics, and the results showed that our algorithms are superior to CLM.**

Data and Method

■ Data

- **FY-4A_L1_04km_AGRI**
- **Caliop_L2_01km_Clay**
- **MCD12C1**

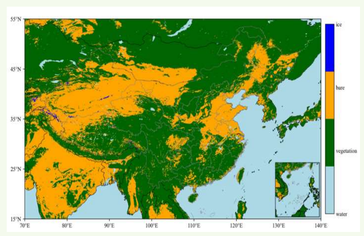


Fig.1 Distribution of the four types of underlayment

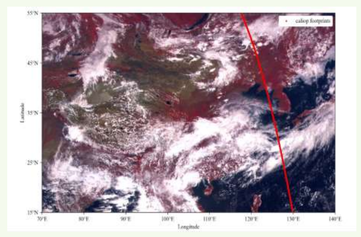


Fig.2 Schematic of FY4A and CALIOP matching

■ Method:matching rule

➢ Nearest match method

■ ML model : RF + AutoML

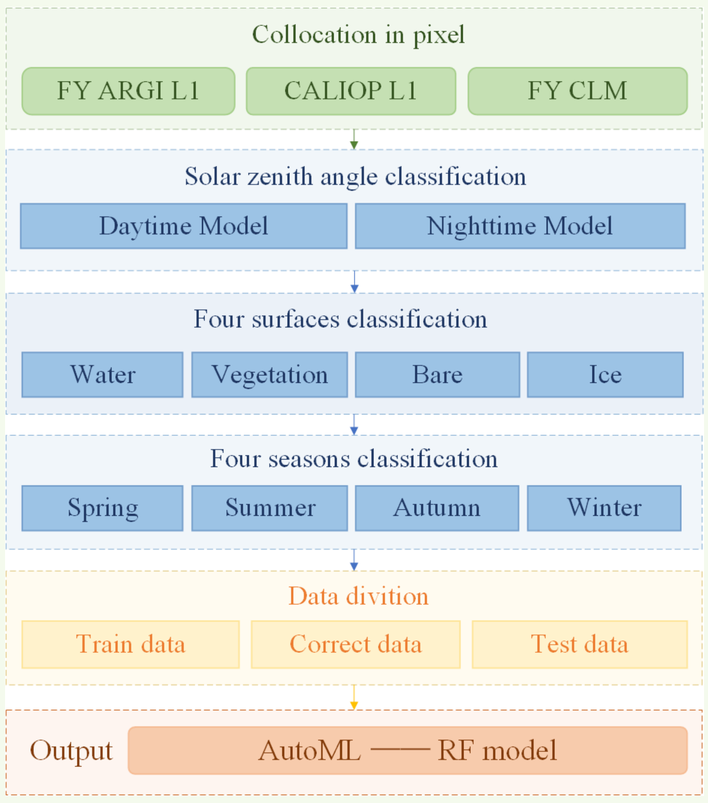


Fig.3 Flowchart of RF-based cloud detection algorithm

Results and Discussion

■ RF model:

Fig.4 Percentage of importance scores for 26 model features (top 5)

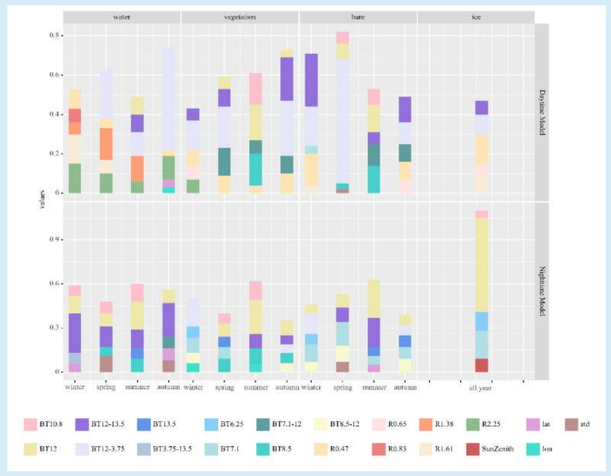
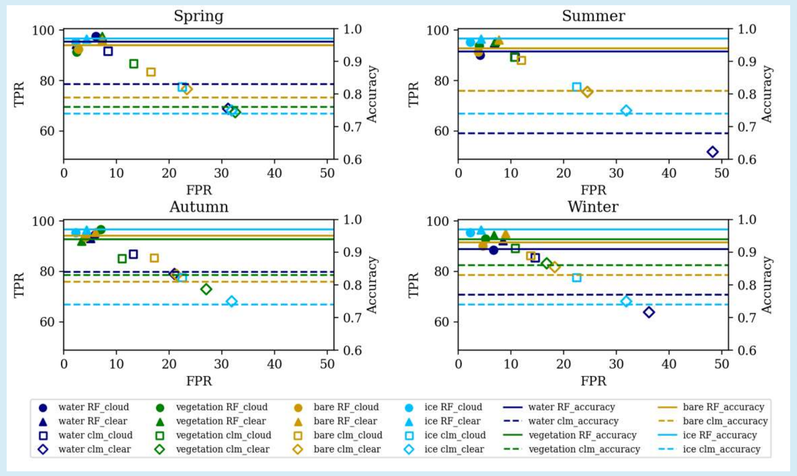
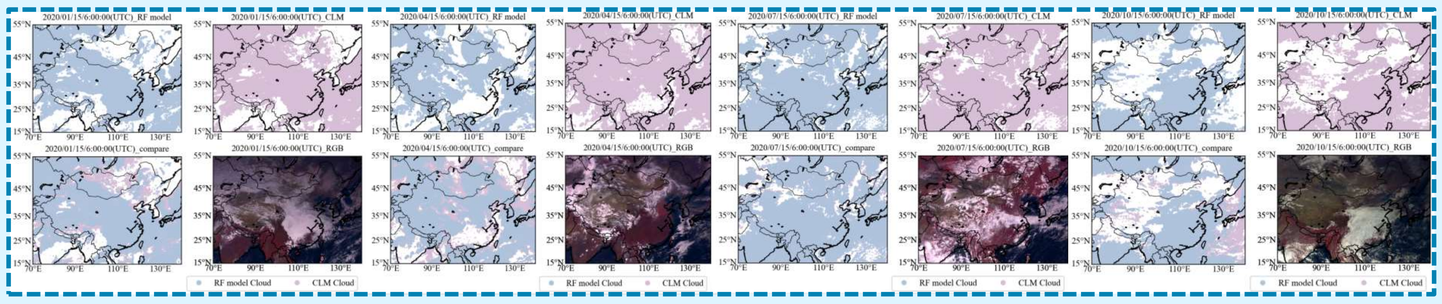


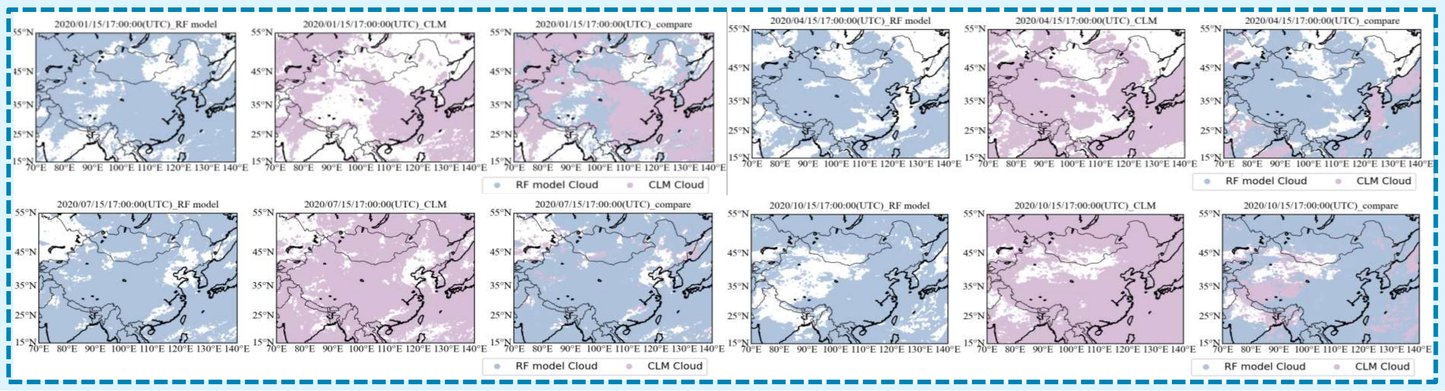
Fig.5 Comparison of TPR, FPR and accuracy based on daytime algorithm



■ **Daytime model**: Individual examples of daytime algorithm models based on four subsurfaces and four seasons



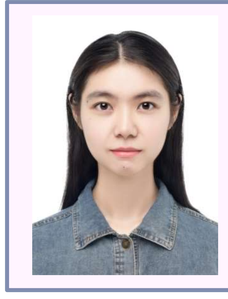
■ **Nighttime model**: Individual examples of nighttime algorithm models based on four subsurfaces and four seasons



Conclusion

In this study, daytime and nighttime algorithms were developed based on the random forest algorithm for four seasons (spring, summer, autumn, winter) and four types of subsurface (water, vegetation, bare , ice). The training, testing, and validation set truth values required for algorithm development were obtained from CALIOP radar data. Five evaluation metrics are proposed to compare 26 model results with CLM, the official cloud detection product for storm clouds. On average compared to CLM, the daytime algorithm model accuracy is improved by about 15% and the nighttime algorithm model accuracy is improved by about 10%; the daytime algorithm model FPR is reduced by about 11% and the nighttime algorithm model FPR is reduced by about 8%. From the analysis of individual cases, the algorithm model inverts cloud edges better than CLM but overestimates clear sky pixels.

About the Author



MEIJIA XIAO was born in 1997. She received the B.S. degree from Chengdu University of Information Engineering, China, in 2020. She is currently pursuing the Ph.D. degree at the Institute of Atmospheric Physics, Chinese Academy of Sciences. Her research interests mainly focus on satellite remote sensing inversion of cloud\fog, machine learning, etc.

E-mail address:3538764076@qq.com

References:

[1] Wang C , Platnick S , Meyer K ,et al.A Machine Learning-Based Cloud Detection and Thermodynamic Phase Classification Algorithm using Passive Spectral Observations[J]. 2019.DOI:10.5194/amt-2019-409.

[2] Liu C , Yang S , Di D ,et al.A Machine Learning-based Cloud Detection Algorithm for the Himawari-8 Spectral Image[J].Advances in Atmospheric Sciences, 2021.DOI:10.1007/s00376-021-0366-x.

[3] Yang, Yikun ,et al.. Machine learning-based retrieval of day and night cloud macrophysical parameters over East Asia using Himawari-8 data. Remote Sensing of Environment. 273. 112971. 10.1016/j.rse.2022.112971.

[4] Shi et al., First estimation of high-resolution solar photovoltaic resource maps over China with Fengyun-4A satellite and machine learning. RSER,2023,18,113549.