Geophysical Research Abstracts
Vol. 21, EGU2019-6226, 2019
EGU General Assembly 2019


# A Random Forest Approach for Cloud Base Analysis

Chao Li
CMA Public Meteorological Service Centre, Beijing, China (lichao@cma.gov.cn)

The temporal and spatial characteristics of cloud base height have an important effect on the evolution of cloud system, its information was obtained by fusion analysis of multi-source observation (i.e. remote sensing data, field cloud base and meteorological observation via various instruments) and numerical model prediction. The core idea of current broadly adopted step-by-step fusion approach is to seek the greatest consistency of multi-source information which arises two problems, first, the consistency rules are difficult to obtain,second, the setting of threshold contains uncertainty and situation depends.

Machine learning uses statistical techniques to give computer systems the ability to "learn" from data. Random forests (RF), one of the state of art machine learning algorithms, is an ensemble learning method for classification. RF operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes or mean prediction of the individual trees. Fernández-Delgado, M., et al. (2014) evaluated 179 classifiers arising from 17 families via 121 data sets, concluded that “The classifiers most likely to be the bests are the random forest (RF)”.

In this study,RF was used to integrate various information to establish a data-driven analysis approach (RF-CB) with consideration of physical mechanism.Using the best evaluated data-set from Atmospheric Radiation Measurement (ARM) in the Southern Great Plains (SGP), we establish the RF-CB combing ground observation, radar,total sky image, satellite radiation data and NCEP prediction.Also, the resulting human readable learning rules and the variable importance measures are studied to deeper understand relationship among various sources of information.

If tested successfully, the RF-CB scheme can potentially (a) speed up the performance of cloud analysis model, (b) be helpful in leading to a deeper understanding of physical mechanism.