Low Cloud Detection for the GOES ABI using a Random Forest Classifier

John M. Haynes¹, Yoo-Jeong Noh¹, Steven D. Miller¹, and Andrew K. Heidinger²

(1) Cooperative Institute for Research in the Atmosphere, Fort Collins, CO (2) NOAA/NESDIS, Madison, WI. Email: john.haynes@colostate.edu

Introduction

- Passive remote sensing instruments like the Advanced Baseline Imager (ABI) on GOES-16/17, are particularly effective at revealing attributes of the topmost layer of clouds (their height, particle size, water content, etc.). However in multiday situations, information on those lower layers is limited.
- NOAA customers, however (particularly in the aviation community), may be particularly interested in the presence of low clouds. This is especially relevant for general aviation in terms of appropriability of VFR vs. IFR conditions, the presence of aircraft icing, mountain obscuration, etc.
- The ABI satellite image at left shows a cold air outbreak spreading southward across the Great Plains. Low clouds are occurring behind the front in the cold air.
- The white/grey areas are high clouds identified by their cold IR brightness temperatures; the blue areas are low clouds detected using a 10.3 vs. 3.9 μm threshold.
- There are low clouds "hiding" under the cirrus at the red "X"? Our meteorological experience says "yes", but this isn’t at all evident from a satellite algorithm alone.
- The purpose of this study is to explore use of Machine Learning to improve our ability to detect low clouds from satellites, especially to augment existing satellite products.

Original (non-ML) Cloud Product

- The baseline cloud height identification product that we wish to improve is called Cloud Cover Layers (CCL). In its earliest form, this algorithm identified ABI pixels as containing exactly one of (Low), (Mid), or (High) cloud, based mostly on the associated cloud top height determination.
- In the past three years, we have improved this product by using a priori statistical relationships derived from the spaceborne CloudSat radar and CALIPSO lidar (Noh et al. 2017) to derive the geometric thickness of the topmost cloud layer (Δz) for any pixel observed by the ABI. When combined with the cloud top height (CTH), this new information allows clouds to occupy more than one vertical level. The following diagram demonstrates how a cloud that used to be simply classified as H (high) can now be classified as H+M (high plus mid):
- We can evaluate this algorithm using overlaps of the CloudSat radar and CALIPSO lidar through the ABI full disk sector. This is an example vertical cross section through some clouds that ABI-only algorithms identified as containing only cirrus:

Method

- A Random Forest (RF) classifier (Breiman 2001) is used for this work. For each of the matched CloudSat/CALIPSO/ABI pixels in the training set, the binary question is asked “Is there a low cloud (≤ 642 hPa) present?”
- Data used for training has been determined experimentally. Initially we included all ABI channels in the training, but a selection of physically relevant channels, ratios, and differences has led to better results.
- Currently we use the following for training during daytime:
  - An ABI difference channel (Δνs)
  - A split window channel
- At night, training data does not use channels (≥ 3.9 μm).

Results

- After training, the daytime results are evaluated on an independent test data set, consisting of 2.1 million CloudSat/CALIPSO/ABI matchups (Oct 2017, Jan 2019)
- TPR (Probability of Detection) of low cloud

Data / Preprocessing

- We want to improve the ABI Cloud Cover Layers product by "filling in" the missing low clouds (see bottom left panel of this poster) using machine learning.
- CloudSat radar and CALIPSO lidar will provide the "truth"; as active sensors, they detect most clouds in the atmosphere.
- The parallax-corrected GOES-16 ABI data are matched to the nearest pixel along CloudSat / CALIPSO ground track, in all 16 bands.
- Currently, surrounding pixels are not utilized (but we expect improvements by doing so).
- National Snow and Ice Data Center (NSIDC) daily microwave SSM/I-based land/sea surface characterization, the GTOPO30 elevation map, and atmospheric relative humidity profiles from ECMWF model are also matched to the same location.

Figure 1 shows that the new algorithm (dark bars) has increased POD compared to the original algorithm (light grey bars) for all cloud types except for fog. Improvement is most significant for the Cirrus and Enhancing cloud categories, which contain significant fraction of multi-cloud layers.

Note that the new algorithm has an increased rate of false positives compared to the original algorithm; there are now a few cases that slightly increase POD but there is no significant increase in missed detections. Improvements in these cases are mostly due to improved detection of cloud bases, particularly for cirrus, for which the original algorithm is unable to identify the heights of other non-cirrus layers.

Figure 2 shows the POD and FAR for low cloud detection for the new algorithm (dark bars) and original algorithm (light bars) for all ABI cloud types.

Note that the new algorithm has increased rate of false positives compared to the original algorithm; there are now a few cases that slightly increase POD but there is no significant increase in missed detections. Improvements in these cases are mostly due to improved detection of cloud bases, particularly for cirrus, for which the original algorithm is unable to identify the heights of other non-cirrus layers.

Figure 3 shows a 4×4 grid of SSM/I images of cloud detection for the new algorithm (dark green) and original algorithm (light green) for all ABI cloud types.

Note that the new algorithm has increased rate of false positives compared to the original algorithm; there are now a few cases that slightly increase POD but there is no significant increase in missed detections. Improvements in these cases are mostly due to improved detection of cloud bases, particularly for cirrus, for which the original algorithm is unable to identify the heights of other non-cirrus layers.

Figure 4 shows a 4×4 grid of SSM/I images of cloud detection for the new algorithm (dark green) and original algorithm (light green) for all ABI cloud types.

Note that the new algorithm has increased rate of false positives compared to the original algorithm; there are now a few cases that slightly increase POD but there is no significant increase in missed detections. Improvements in these cases are mostly due to improved detection of cloud bases, particularly for cirrus, for which the original algorithm is unable to identify the heights of other non-cirrus layers.

Figure 5 shows a 4×4 grid of SSM/I images of cloud detection for the new algorithm (dark green) and original algorithm (light green) for all ABI cloud types.

Note that the new algorithm has increased rate of false positives compared to the original algorithm; there are now a few cases that slightly increase POD but there is no significant increase in missed detections. Improvements in these cases are mostly due to improved detection of cloud bases, particularly for cirrus, for which the original algorithm is unable to identify the heights of other non-cirrus layers.

Figure 6 shows a 4×4 grid of SSM/I images of cloud detection for the new algorithm (dark green) and original algorithm (light green) for all ABI cloud types.

Note that the new algorithm has increased rate of false positives compared to the original algorithm; there are now a few cases that slightly increase POD but there is no significant increase in missed detections. Improvements in these cases are mostly due to improved detection of cloud bases, particularly for cirrus, for which the original algorithm is unable to identify the heights of other non-cirrus layers.

Figure 7 shows a 4×4 grid of SSM/I images of cloud detection for the new algorithm (dark green) and original algorithm (light green) for all ABI cloud types.

Note that the new algorithm has increased rate of false positives compared to the original algorithm; there are now a few cases that slightly increase POD but there is no significant increase in missed detections. Improvements in these cases are mostly due to improved detection of cloud bases, particularly for cirrus, for which the original algorithm is unable to identify the heights of other non-cirrus layers.