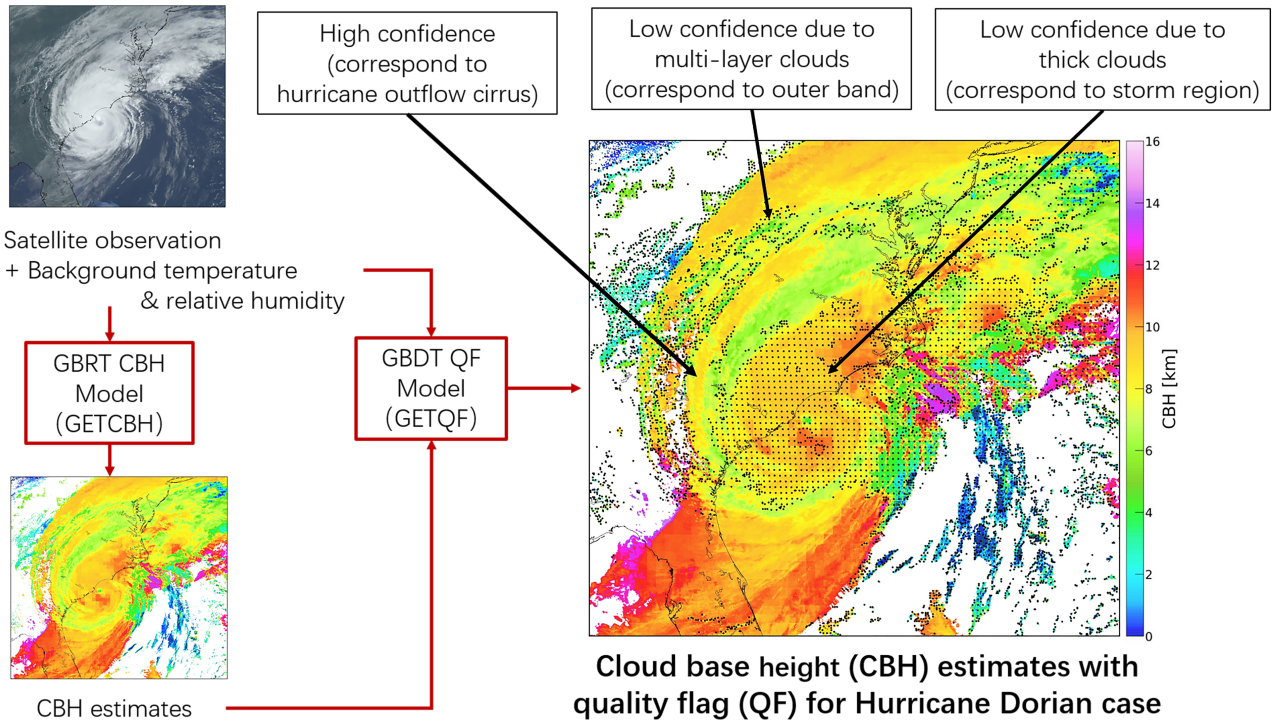
Graphical Abstract

Estimate of daytime single-layer cloud base height from Advanced Baseline Imager measurements

Han Lin, Zhenglong Li, Jun Li, Feng Zhang, Min Min, W. Paul Menzel



Highlights

Estimate of daytime single-layer cloud base height from Advanced Baseline Imager measurements

Han Lin, Zhenglong Li, Jun Li, Feng Zhang, Min Min, W. Paul Menzel

- Daytime cloud base height (CBH) is estimated from ABI level 1b data.
- Cloud phase information has high impact on CBH estimate.
- A quality flag model is built to classify CBH retrievals with high/low confidence.

Estimate of daytime single-layer cloud base height from Advanced Baseline Imager measurements

Han Lin[a,b,d], Zhenglong Li[b], Jun Li[c,1,*], Feng Zhang[d,e], Min Min[f], W. Paul Menzel[b]

[a]*Key Laboratory of Meteorological Disaster, Ministry of Education/ Collaborative Innovation Center on Forecast and Evaluation of Meteorological Disaster, Nanjing University of Information Science and Technology, Nanjing, 210044, China*
[b]*Cooperative Institute for Meteorological Satellite Studies, University of Wisconsin-Madison, Madison, 53705, WI, USA*
[c]*National Satellite Meteorological Center, China Meteorological Administration, Shanghai, 200232, China*
[d]*Shanghai Qi Zhi Institute, Shanghai, 200232, China*
[e]*Department of Atmospheric and Oceanic Sciences and Institute of Atmospheric Sciences, Fudan University, Beijing, 10081, China*
[f]*School of Atmospheric Sciences and Guangdong Province Key Laboratory for Climate Change and Natural Disaster Studies, Sun Yat-Sen University, Zhuhai, 519082, China*

Abstract

Cloud base height (CBH) is an important parameter to describe cloud state and is highly related to the vertical motions in the atmosphere. CBH information is critical for both aviation safety and synoptic analysis. In this study, daytime CBH is estimated directly from Geostationary Operational Environmental Satellite-R Series (GOES-16) Advanced Baseline Imager (ABI) level 1b data and the European Centre for Medium-Range Weather Forecasts' (ECMWF) fifth generation reanalysis (ERA5) data using the Gradient Boosted Regression Trees (GBRT) machine learning technique. The CBH estimate algorithm, which is named as GETCBH, covers the same areal extent as the full disk of the ABI/GOES-16 and only for single-layer clouds. The 2-years of CBH measurements from the Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) aboard Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) satellite is used as the label (which is the true value/class of the model output for regression/classification problem in machine learning terminology). A quality flag algorithm using another machine learning technique, the Gradient Boosted Decision Trees machine learning technique is developed to provide a confidence level for the CBH estimate. The evaluations show an overall root mean square error (RMSE) of 1.87 km

*corresponding author.
[1]Email address: jli_ssec@wisc.edu

and Pearson's correlation coefficient (Pearson's r) of 0.92 before any quality control. After excluding CBH estimates with low confidence (19.2% of all samples), the RMSE is reduced to 1.14 km, Pearson's r increases to 0.97, and 96% of the estimates are within 2 km of the CALIOP results. By analyzing model bias and feature importance, cloud phase information has the biggest impact on the CBH estimate, although all input features have positive impact on the estimate accuracy. Limited by the penetrability of CALIOP, GETCBH is valid for clouds with COD < 8.5. The CBH estimates have reduced accuracy (Pearson's r of 0.88) for optically thin clouds (clouds with cloud optical depth [COD] < 0.1) where little cloud information is contained in the ABI measurements, as well as for optically thick clouds (clouds with COD ≥ 3) where a larger proportion of opaque clouds is excluded. Furthermore, for the GBTCBH model using 9 months of CloudSat measurements as label, the CBH estimates are improved with an RMSE of 1.41 km and Pearson's r of 0.92. In a case study of Hurricane Dorian, CBHs for most of the single-layer clouds are successfully estimated with small errors and flagged with high confidence, for both high and low clouds. Deep convective clouds and multi-layer clouds, both of which are not included in the training, are reasonably flagged as low confidence with large CBH estimate errors. In this particular case, 65% of cloudy pixels have CBH estimate with high confidence in the scene. Daytime CBH with high spatial (2 km) and temporal (10 min) resolution can be derived from ABI measurements using this methodology.

Keywords: Cloud base height, GOES-16 ABI, CALIPSO CALIOP, Hurricane Dorian, Machine learning

1. Introduction

A commonly occurring atmospheric phenomenon, clouds often are described by their optical, microphysical, and macrophysical properties (e.g., cloud optical depth or COD, droplet number concentration, or cloud top/base height). Among them, cloud base height (CBH) is sparsely observed; yet it is an important parameter in determining the cloud radiative effect (Garrett et al., 2010; Yeo et al., 2018), the energy balance in climate simulations (Wetherald and Manabe, 1988), and the local thermal adjustment in numerical weather prediction (NWP) models (Ruppert et al., 2020). In addition, changing cloud vertical structures in-

fluence atmospheric circulation by directly altering the radiative heating/cooling rate and atmospheric static stability and by indirectly affecting the latent heating profile of deep convection (Wang and Rossow, 1998). Lastly, aviation guidance can also benefit from accurate CBH information (Mecikalski et al., 2007).

In routine observations, CBH retrieved from ground-based lidar ceilometer data offers the highest accuracy and is thus widely utilized in the field of aviation. However, the limited spatial coverage is a significant disadvantage for ground-based observations, which prevents broader utilization.

CBH estimates using space-borne observations have much improved spatial coverage over that from ground-based observations. But they are also more challenging, in part because of the larger distance from the cloud base than the cloud top to the satellite. Additional absorption and scattering by the clouds obscure some of the cloud base signal reaching space-borne sensors and make it more difficult to determine than the cloud top height (CTH). Despite the added complexity associated with estimating CBH, several estimation methods have been developed for different types of space-borne sensors.

For active sensors, accurate CBHs can be estimated from the Cloud Aerosol Lidar with Orthogonal Polarization (CALIOP) on the Cloud Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) satellite, including CBHs for optically thin clouds (Winker et al., 2010). The measurements from CALIPSO/CALIOP have been used to retrieve cloud profiles accurately, but the horizontal coverage is limited by its narrow scanning swath. Mülmenstädt et al. (2018) expanded the CBH coverage by extrapolating the CALIPSO/CALIOP CBH to nearby optically thin clouds. Another method extrapolating the CBH from active sensors to a broader spatial coverage was developed for CloudSat's Cloud Profiling Radar (CPR) measurements combined with the Moderate Resolution Imaging Spectroradiometer (MODIS) data on board the NASA Earth Observing System (EOS) Aqua platform (Miller et al., 2014; Sun et al., 2016).

Compared with active space-borne sensors, passive space-borne sensors, such as visible (VIS) /near infrared (NIR) /infrared (IR) imagers, have larger horizontal coverage, although the profile information from those passive sensor measurements is not nearly as extensive as that of the active sensors. Hence, the CBH estimates from passive airborne sensors often

depend on ancillary cloud products. For example, a method relating CBH to CTH and cloud water path (CWP) is operationally applied to data from the Visible Infrared Imaging Radiometer Suite (VIIRS) onboard the Suomi–National Polar-Orbiting Partnership (SNPP) platform (Seaman et al., 2017; Noh et al., 2017). Similarly, Böhm et al. (2019) derived CBH from CTH products using data from the Multi-angle Imaging SpectroRadiometer (MISR) onboard the EOS Terra platform when some preconditions, such as a cloud cluster with broken clouds, are met. This method assumes that the lowest CTH within a cloud cluster could correspond to the CBH seen from its side. Recently, Tan et al. (2019, 2021) took advantage of a random forest algorithm to estimate CBH from upstream cloud products of the Advanced Geostationary Radiation Imager (AGRI) onboard Fengyun-4A (FY-4A) and the Advanced Himawari Imager (AHI) on Himawari-8.

Even though the methods for CBH estimate from different sensors have been reported in recent decades, CBH is not one of the ABI baseline products (Schmit et al., 2005, 2009, 2017; Noh et al., 2017). ABI offers high spatio-temporal resolution data with a fixed nadir field of view, making it possible to continuously infer cloud base and provide input for aviation applications and weather analysis. Most existing CBH estimate methods using passive airborne sensors, that provide similar spatio-temporal resolutions, are based on level 2 cloud products. Our study examines whether a machine learning (ML) algorithm is capable of extracting CBH information directly from satellite level 1 measurements. This work introduces an ML based daytime CBH estimate algorithm that directly uses ABI level 1 measured radiances and reflectances, as well as European Centre for Medium-Range Weather Forecasts' (ECMWF) newly released fifth generation reanalysis (ERA5) ([dataset] European Centre for Medium-Range Weather Forecasts, 2017) temperature and relative humidity (RH) profiles, as input. The CBH estimate algorithm covers the same areal extent as the full disk of the ABI/GOES-16.

In recent years, ML in the field of remote sensing and meteorology has been proven to be effective for many non-linear problems such as estimating precipitation or CTH (Kühnlein et al., 2014; Håkansson et al., 2018; Min et al., 2017, 2020, and others).Unlike CTH, which has a rather clear physical relation with radiances or brightness temperatures (BTs) from IR bands, CBH is complicated by the clouds' strong absorption in the IR and scattering in the

VIS spectral bands. The ML technique could be useful for solving such a problem. In this study, the Gradient Boosted Regression Trees (GBRT) ML technique is used. The objective is to determine if an ML technique can be used to estimate CBH from ABI observations. The ERA5 dataset provides supplementary information regarding atmospheric profiles.

The data and ML-based CBH estimate models are described in Section 2. ABI, ERA5 and CALIOP data are briefly introduced in Subsection 2.1. Subsection 2.2 presents the data processing before model training. The settings for the models are noted in Subsection 2.3, and methods for model evaluation are detailed in Subsection 2.4. Results are analyzed in Section 3. Subsection 3.1 presents the general model errors (error in this study refers to the discrepancy between the estimated results and labels). Label is the true value/class of the model output for regression/classification problem in ML terminology. Feature importance and variables affecting CBH accuracy are analyzed in Subsection 3.2. A quality flag (QF) method is described in Subsection 3.3. Subsection 3.4 presents the results of an experiment using CloudSat CBH as label. The case studies comprising this work are discussed in Subsection 3.5. Finally, discussion and conclusions are found in Section 4.

2. Data and Methods

2.1. Data Source

ABI is a 16-band passive imaging radiometer designed to observe the Western Hemisphere and provide variable area imagery and radiometric information on the Earth's surface, atmosphere, and clouds. Located at 75.2°W, the ABI/GOES-16 full disk scanning has a temporal resolution of 5-15 minutes (depending on scan modes with a 10-minute flex mode becoming the default operating mode since April 2019), and a spatial resolution of 0.5 to 2 km at nadir. The 16 bands consist of 2 VIS, 4 NIR and 10 IR bands with central wavelengths ranging from 0.47 to 13.3 μm. More information can be found in Schmit et al. (2005, 2009, 2017). To maximize the use of ABI observations, Level 1b (L1b) data from all 16 bands at their finest resolution are used in the ML model (Table 1), including reflectance from VIS/NIR bands and radiances from IR bands.

To improve CBH estimates, the temperature and humidity profile information, especially in the troposphere, from the ERA5 reanalysis ([dataset] European Centre for Medium-Range

Weather Forecasts, 2017) are also used as input to the ML model. ERA5 hourly reanalysis data are $0.25^\circ \times 0.25^\circ$ gridded global data at 37 fixed vertical pressure levels from 1000 to 1 hPa. RH is chosen over the mixing ratio because it results to lower RMSE (about 0.07 km lower) of the CBH estimates in the tuning experiment. Only the lowest 27 fixed levels from 1000 to 100 hPa are used because most CBHs are found in this range. Since the pressures at the 27 levels are fixed, there is no need to use them as input. The surface pressure and surface skin temperature are also used (Table 1). ERA5 is chosen for its overall high accuracy and high spatio-temporal resolutions, which allows better understanding of the importance of temperature and RH profiles. These variables are chosen for their importance in forward radiative transfer calculations (Clough et al., 2005). Their importance in the CBH estimates will be quantitatively evaluated. It is important to point out that other NWP output (forecast or analysis) could be used instead of ERA5, i.e., GFS forecast data. However, one must make sure that the same feature data be used in both the training and the application.

CBH from CALIOP was obtained as the lowest height with a cloud fraction larger than 0, and provide the label for the ML model. CALIOP is a two-wavelength (532 and 1064 nm) polarization-sensitive lidar that provides high-resolution vertical profiles of aerosols and clouds (Winker et al., 2004, 2007). The latest Version 4 (V4) CALIOP cloud level 2 product (CAL_LID_L2_05kmCPro) has been generated based on its 532 nm laser. The information about cloud fraction at each level is provided at a horizontal resolution of 5 km and vertical resolution of 60 m over an altitude range from -0.5 to 30 km, where -0.5 km refers to the regions lower than sea level; also included are cloud column optical depth and quality control (QC) flags, e.g., extinction QC flags (Young et al., 2018) and cloud-aerosol discrimination score (Liu et al., 2019). This product is described online (`https://www-calipso.larc.nasa.gov/resources/calipso_users_guide/data_summaries/profile_data_v420.php#heading13`) and data can be downloaded through the Open-source Project for a Network Data Access Protocol (OPeNDAP, URL: `https://opendap.larc.nasa.gov/opendap/hyrax/CALIPSO/LID_L2_05kmCPro-Standard-V4-20/`)

CALIOP has a complete data record since the ABI/GOES-16 became operational on December 18, 2017. However, it is worth noting CALIOP's limited horizontal coverage and reduced signal-to-noise ratio (SNR) when the cloud optical depth (COD) is larger than 3

or less than 0.1 (Dupont et al., 2010). In comparison, the CloudSat 2B-GEOPROF-LIDAR product (Mace and Zhang, 2014) combining radar data from CPR onboard CloudSat and CALIOP lidar data is valid for a broader COD range. The spatial resolution of this product is about 2 km along track and 1 km across track. The vertical resolution is 240 m. However, this product has only 9 months overlap with ABI/GOES-16: October 10, 2018 through July 10, 2019. As a result, this study primarily uses CALIOP CBH products to develop the ML-based CBH estimate model. The potential advantage of CloudSat is also shown with the limited data. The CloudSat 2B-GEOPROF-LIDAR product can be accessed via direct FTP access (`ftp://ftp.cloudsat.cira.colostate.edu/2B-GEOPROF-LIDAR.P2_R05/`)

2.2. ML Models

Min et al. (2020) compared four decision-tree-based algorithms for CTH estimates for Himawari-8 AHI, and found the GBRT (Friedman, 2001) algorithm outperforms others with a minimum mean absolute error (MAE) down to 1.54 km with small fluctuations in the tuning experiments. Following their findings, this study uses GBRT for estimating CBH. Another ML technique, the deep neural network (DNN), was also tested, but the results were not as good. The RMSE and the absolute value of the mean bias error (MBE) are higher than those from this study (about 0.5 and 0.3 km respectively; further results and discussion not provided in this study).

GBRT is a non-parametric statistical ML regression algorithm comprised of a series of weak decision trees. GBRT builds an additive model in a forward stage-wise fashion. In each stage a regression tree is fit on the negative gradient of the given loss function (mean squared error is used). Each regression tree is a flowchart-like structure composed of nodes and the terminate nodes also known as leaves. A Python package, scikit-learn (Pedregosa et al., 2011), was utilized to establish the GBRT based CBH estimate models, named the GBRT model to EstimaTe Cloud Base Height (GETCBH). It uses ABI+ERA5 as features, and CALIOP as the label unless otherwise specified. If CloudSat is used as label, it is referred to as the CloudSat-based GETCBH. If pre-determined phase information is used to enhance GETCBH, it is referred to as the phase-enhanced GETCBH.

In addition, a Gradient Boosted Decision Trees (GBDT) based model to generate the QF to show the confidence level of the CBH is developed and named as the GBDT model to

EstimaTe Quality Flag for cloud base height (GETQF). It uses the same ABI+ERA5, along with the estimated CBH, as features and estimates confidence levels in two categories (low and high) for the CBH estimate.

GBRT has many applications, such as CTH (Min et al., 2020), solar power (Persson et al., 2017), and terrestrial latent heat flux (Wang et al., 2021), etc. That also means the optimization of parameters used by GBRT is necessary. Most tuning processes of the ML model rely on the minimization of both bias error and variance. The bias error refers to the difference between the estimated value and the expected value. Large bias means the model under-explains the relationship between features and label (underfitting). The variance refers to the error from the sensitivity to changes in the training dataset. A large variance means the model learned the random noise in the training data (overfitting). However, these two properties are often conflicted when trying to minimize them simultaneously. GBRT is a highly non-linear ML algorithm. Given enough input features and certain hyperparameters, it can fit the training dataset to the point where the bias error is close to 0. However, the model learns the noises at the same time, which increases variance and decreases generalization of the model. The bias-variance tradeoff, which can be accomplished by tuning the hyperparameters, is necessary in such cases.

As part of the tuning process for GETCBH, our goal is to minimize the discrepancy between the labels and CBH estimates of the validation dataset. By approaching that, the model can acquire a general best accuracy. The hyperparameters used in Min et al. (2020) are taken as an initial setting. Then each hyperparameter is tuned by cycle training the model with different values. The settings with the lowest RMSE for the validation dataset are chosen. For example, the number of regression trees to be performed is set to 500. The increase in this value tends to make the model have lower biases and higher variances. The fraction of samples (randomly selected for each regression tree) to be used for fitting the regression trees is 0.3. The decrease in this value leads to a reduction in both variance and the sensitivity of the model to different training datasets, yet with an increase in bias. The minimum number of samples required to split an internal node is 50. The model tends to have higher biases and lower variances when this value is increased. The minimum number of samples required to be at a leaf node is 25. It exhibits a behavior similar to the previous

parameter when this value is increased. The maximum depth of the individual regression trees is 9. The model tends to have lower biases and higher variances with deeper trees. The contribution weight of each tree (learning rate), which has a tradeoff with the number of boosting stages, is 0.02. Other hyperparameters are set to default (e.g., loss function is mean squared error).

The GBDT, with a core and a structure similar to GBRT but for classification problems, is also used in this study, but only for the QF of the CBH estimates. Again, the hyperparameter optimization was performed. And the hyperparameters are set as follows. The number of boosting stages to perform is set to 400; the fraction of samples to be used for fitting the individual base learners is 0.4; the minimum number of samples required to split an internal node is 275; the minimum number of samples required to be at a leaf node is 100; the maximum depth of the individual regression estimators is 14; and the contribution weight of each tree is 0.02. Other hyperparameters were set to default (e.g., loss function is the deviance for classification with probabilistic outputs).

It is worth noting that, for the consideration of reproducibility, the random states (seeds) of the models and other shuffling processes are set to a fixed number (42 in this investigation).

2.3. Data Processing

Two collocated datasets were generated for this study. The first dataset is for the model training and validation using CALIOP as the label for GETCBH. Two years of data (2018 and 2019) were used — 20% of which are randomly sampled for the validation dataset and the remaining 80% are used in the training dataset. The other dataset is for the model using CloudSat as the label for the CloudSat-based GETCBH. It contains 9 months of observations and the split between training and validation datasets is also 80% and 20%, respectively.

Accurate labels are essential for ML-based algorithms. Multiple measures were used to filter out failed/bad retrievals or suspicious samples. CALIOP samples were considered only if the solar zenith angle was smaller than 90° (daytime, for use of VIS and NIR bands); the extinction QC flag was equal to 0, 1 or 2 (unconstrained, constrained or initial lidar ratio reduced retrievals); the absolute value of cloud-aerosol discrimination score ranges from 50 to 100 (medium to high confidence); and the cloud column optical depth is successfully retrieved (cloud is detected and the retrieval of low-to-no confidence is excluded). Among them, solar

zenith angle is an important parameter in the radiative transfer equation. The cosine of it is used as an input to the model (Table 1). It is worth noting that, for the ABI cloud optical and microphysical properties product, daytime is defined as observations with a solar zenith angle of 65° or below (see `http://cimss.ssec.wisc.edu/clavr/dncomp/ATBD_DCOMP_v2.6.pdf`). However, in this study, only using samples with solar zenith angles less than 65° is not beneficial for the CBH estimates (the RMSE decrease is less than 0.1 km). For a more comprehensive temporal coverage, all samples with solar zenith angles smaller than 90° are included in this work.

After filtering, the remaining samples are classified as single-layer or multi-layer clouds. A multi-layer cloud is classified if any clear sky layer appears between cloudy layers. Note that only single-layer clouds are used for training and validation. Multi-layer clouds and opaque clouds (extinction QC flags equaling to 16, 17 or 18) are only used for demonstration for the case study in Section 3.5. After filtering and classifying, each sample was matched up with the nearest ABI pixel (2 km $\times$ 2 km at sub-satellite point) with a time difference less than 7.5 minutes and an ERA5 grid within $\pm$ 30 min period. After the collocation, there are 779,795 training and 194,949 validation samples.

For CloudSat samples, only those with good quality and average horizontal distances of less than 4 km between the lidar and CloudSat are used. After a similar process of collocation and screening for multi-layer clouds, there are 2,367,556 training and 591,889 validation samples from CloudSat. The sample size is much larger than that of CALIOP for two reasons: 1) the CloudSat product uses pulses averaged over about 1.1 km along-track distance versus 5 km for the CAL_LID_L2_05kmCPro product from CALIOP; 2) clouds that are opaque to CALIOP could be fully observed by CPR. To analyze the performance of the CloudSat-based GETCBH in different COD ranges, the CODs from CALIOP's CAL_LID_L2_05kmCPro product were collocated to the samples from CloudSat's 2B-GEOPROF-LIDAR product with the closest distance and time differences less than 2 minutes. The COD from CALIOP is chosen because CloudSat's COD product is not available for the corresponding period. The COD retrievals of low-to-no confidence are excluded. When comparing the GETCBH with the CloudSat-based GETCBH, using the COD from the same source can eliminate the influence from different COD retrieval methods.

Table 1: Label and Features for ML model. The parameters in parentheses following ABI features mean: (central wavelength, nadir spatial resolution, number of features within each IR pixel of 2 km resolution at nadir). The unit following ABI features is for radiance. And the parameters following the other features are: (number of features within each IR pixel of 2 km resolution at nadir).

Label	Cloud base height [unit: km]
ABI features	Reflectance (0.47 μm, 1 km, 4), Reflectance (0.64 μm, 0.5 km, 16), Reflectance (0.86 μm, 1 km, 4), Reflectance (1.37 μm, 2 km, 1), Reflectance (1.6 μm, 1 km, 4), Reflectance (2.2 μm, 1 km, 4), Radiance (3.9 μm, 2 km, 1), Radiance (6.2 μm, 2 km, 1), Radiance (6.9 μm, 2 km, 1), Radiance (7.3 μm, 2 km, 1), Radiance (8.4 μm, 2 km, 1), Radiance (9.6 μm, 2 km, 1), Radiance (10.3 μm, 2 km, 1), Radiance (11.2 μm, 2 km, 1), Radiance (12.3 μm, 2 km, 1), Radiance (13.3 μm, 2 km, 1) [unit: $\mathrm{mW/m^2sr^1cm^{-1}}$]
ERA5 features on pressure layers	Relative humidity (27) [unit: %], Temperature (27) [unit: K]
ERA5 surface features and others	Surface pressure (1) [unit: hPa], Surface skin temperature (1) [unit: K], Cosine of solar zenith angle (1), Cosine of GOES-16 zenith angle (1)

Table 1 shows the complete list of features used for GETCBH. They are chosen based on their potential importance in cloud, radiative transfer processes (reflectance/radiance, temperature, RH and angle information), and boundary conditions (surface information). Specifically, temperature contributes to IR radiance emission. RH contains the water vapor information relating to optical absorption. Compared to mixing ratio, RH is directly related to lifting condensation level (LCL) or convective condensation level (CCL), and thus is more connected to the cloud genesis (van Heerwaarden and de Arellano, 2008; Costa-Surós et al., 2016). Solar and satellite angles affect the optical path of radiation. The choice of reflectance instead of radiance for the VIS/NIR bands (0.47 – 2.2 μm) is because the influence of solar irradiance can be excluded in reflectance. Note that the various number of features for VIS and NIR bands arises from the different spatial resolutions. For example, bands 0.47 and 0.64 μm have 4 and 16 pixels respectively within 1 pixel of band 3.9 μm. Using all pixels for VIS/NIR bands within an IR pixel enables sub-pixel information (refer to the information with a scale finer than 2 km at nadir) in the estimation model. This leads to slightly but robustly more accurate results (about 0.01 km lower in both MBE and RMSE than using average values in contrast experiments). Last but not the least, surface pressure and surface skin temperature provide the information on the lowest boundary that is sometimes useful for IR bands measuring clouds which are not opaque enough to prevent ABI from seeing the surface.

2.4. Model Evaluation

A few intuitive error indicators, namely mean bias error (MBE), mean absolute error (MAE), and root mean square error (RMSE), were used to evaluate the CBH estimates. They are calculated as follows:

$$MBE = \frac{1}{n}\sum_{i=1}^{n}(y_{o,i} - y_{p,i}), \tag{1}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_{o,i} - y_{p,i}|, \tag{2}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_{o,i} - y_{p,i})^2}, \tag{3}$$

in which n is the sample size, and $(y_{o,i})$ is the i^{th} CALIOP CBH observation, and $(y_{p,i})$ is the i^{th} CBH estimate (prediction). A unitless index, Pearson correlation coefficient (Pearson's r) is also used. Note that all two-tailed p-values of Pearson's r calculated in this study are smaller than 0.0001 and not presented. The comparisons of Pearson's r between datasets are verified by Zou's 99% confidence interval (Diedenhofen and Musch, 2015)

To help understand the importance of features, the permutation importance (PI) is introduced. PI is measured by calculating how much RMSE accuracy deteriorates after permutating (randomly shuffling) a feature. Taking the surface skin temperature feature as an example, its PI is defined to be the decrease in the model score (negative RMSE, whose decrease means the increase in errors) when its positions are permutated between samples. The purpose of permutation is to disable the feature's contribution. To minimize the random uncertainties in calculating PI, 10 permutations are applied. PI can be negative if the permutation improves the accuracy. A negative PI means the feature's contribution is negative. Removing it from the feature list actually improves the model performance.

It is worth noting that, the PI calculation code in the scikit-learn package was partially revised to enable a more robust analysis. The original code was designed to calculate the PI of each feature, such as every single reflectance of the pixels with resolution finer than 2 km at nadir, independently. After revision, multiple features can be permutated simultaneously to achieve an overall PI of those features. This revision is needed because the PI of multiple features is not necessarily the sum of all PIs of individual features (Gregorutti et al., 2015).

3. Results

The analysis of the accuracy of GETCBH is presented in this Section. Feature importance is analyzed and they are ranked based on their PI values. In addition, GETQF is constructed to refine the accuracy and estimate quality. Finally, one case study is presented.

3.1. Overall Evaluation

GETCBH is trained using the GBRT algorithm. The hyperparameters are determined by tuning the model (see Section 2.2). Subsequently, the model is trained with the CALIOP training dataset, and evaluated with the CALIOP validation dataset.

Table 2 shows the evaluation results of the validation set. GETCBH has an MBE of -0.01 km and an RMSE of 1.87 km. These numbers are slightly larger than or comparable to the VIIRS CBH estimate (MBE of about 0.4 km and RMSE of 1.7 km) by Noh et al. (2017). Currently there is no operational CBH product generated from GEO satellites and therefore no product requirements are available. However, 84% of CBH estimates in the validation dataset using GETCBH have discrepancies within ± 2 km of CALIOP CBHs (the VIIRS CBH product requirement; Baker, 2011). Note that there is no procedure to quality control the CBH estimates. So, there could be CBH estimates with low confidence degrading MAE and RMSE.

Another way to understand the performance of GETCBH is through comparison with CTH estimate statistics. Min et al. (2020) used the same source of label (CALIOP) in their study for CTH estimates along with observations from AHI, which is similar to ABI. Their results show an MAE of 1.53 km and a standard deviation of 2.57 km for CTH. When the MAEs of CTH and CBH are scaled by their own standard deviations (5.61 km for CTH and 4.75 km for CBH), they are actually comparable to each other (0.27 for CTH and 0.25 for CBH).

To test the sensitivity of GETCBH on the training-validation sampling, the training-validation (with a split of 80% - 20%) dataset was resampled and re-trained with the same hyperparameters four additional times. Note that for more representative samplings, five validation datasets are set to be complementary subsets. Consistent results were obtained with the largest RMSE difference being 0.02 km and Pearson's r difference of 0.002. These small differences indicate the chosen hyperparameters are not sensitive to the sampling of the validation set.

Table 2 shows the estimated CBH has a small bias of 0.01 km. Detailed analysis shows the bias has vertical variation. Figure 1(B) shows the CALIOP measured and estimated CBHs both have histograms with two peaks, representing the low liquid water clouds and the high ice clouds, respectively. The CALIOP measured low clouds peak near the surface with a value about 1.7e4 samples, while the GETCBH estimates peak around 1.25 km with a value about 1.35e4 samples. For high clouds, the CALIOP measurements have a longer tail with CBH reaching 17.5 km; the GETCBH estimates in general are all less than 16.5 km.

Table 2: Errors (unit of km) in the CBH estimates using the training and validation set from GETCBH, the phase-enhanced GETCBH, and the CloudSat-based GETCBH. The phase-enhanced GETCBH is the same as GETCBH but with CALIOP phase information as an extra feature. Confidence level (± 2 km) is the percentage of CBH estimates with errors within ±2 km.

Experiment	MAE	MBE	RMSE	Pearson's r	Confidence Level (± 2 km)
Training set					
GETCBH	1.11	0.00	1.76	0.93	85%
Phase-enhanced GETCBH	0.67	0.00	1.01	0.98	95%
CloudSat-based GETCBH	0.71	0.00	1.27	0.94	92%
Validation set					
GETCBH	1.17	-0.01	1.87	0.92	84%
Phase-enhanced GETCBH	0.70	0.00	1.05	0.98	94%
CloudSat-based GETCBH	0.77	0.00	1.41	0.92	91%

These results indicate that the GETCBH estimates are underestimated for high clouds and overestimated for low clouds. The low clouds are defined with CBH lower than 7 km and high clouds higher than 7 km. The low clouds have a percentage of 72% with CBH lower than 4 km. That means occurrence of mid-level clouds in the training/validation data set is limited. For simplification, they are added in with the low clouds. The density plot in Figure 1(A) also shows a similar pattern. Around the high cloud peak, the CBHs from GETCBH are clearly underestimated; and around the low cloud peak, they are clearly overestimated. To further understand the GETCBH bias, Figure 1(C) presents the bias distribution vertically. It is clear that the bias of the GETCBH estimates has a vertical variation. The bias decreases from positive to negative with increased CBH. A similar bias pattern was observed in CTH estimates as well (Min et al., 2020).

Figure 1(A) also shows that most CBH errors are confined within a range of ± 2 km (about 84% of all samples). For CBH estimates with uncertainties larger than 2 km, 31.5% (or 5.0% of all samples) have uncertainties larger than 4 km. Removing those samples (whose uncertainties are larger than 4 km), would reduce the RMSE to 1.24 km. These are mostly high ice clouds badly estimated with low CBHs or low water clouds badly estimated with

high CBHs. The CBHs of these clouds could be improved if phase information is known.

It is not immediately clear what causes the vertical variation in the bias shown in Figure 1(C). One limitation in the above estimate is that the GETCBH does not discriminate ice clouds from liquid ones. The ice and liquid clouds are different in both their physical and optical properties, while ice clouds are typically located higher than liquid water clouds. Figure 1(C) shows that ice and water clouds exhibit different bias behavior. Not discriminating them in the GETCBH might lead to underestimates of CBH for ice clouds and overestimates for water clouds.

3.2. Feature Analysis

Figure 1 shows that the bias of GETCBH has height variation, where low clouds have a positive bias, and high clouds have a negative bias. In addition, the lower (higher) the low (high) CBHs are, the larger the absolute MBE values they have (Figure 1C). Cloud phase information could possibly help to reduce the CBH errors. To understand the potency of phase information in the CBH estimate, cloud phase index is used as a feature alongside those in Table 2 to establish the phase-enhanced GETCBH with the same settings as GETCBH (see Section 2.2). The cloud phase index is obtained from CALIOP by classifying cloud phase as 0 for unknown, 1 for liquid cloud and 2 for ice cloud. The water (ice) cloud phase is determined if the cloud phase of the column profile is all water (ice) clouds. Profiles with mixed or unknown phase are categorized as unknown; these are 6.6% of all samples. Table 2 shows that the addition of the cloud phase as a feature significantly improves the CBH estimates. The results of the phase-enhanced GETCBH have their RMSE significantly reduced from 1.87 km to 1.05 km, or a reduction of 44% (0.82 km reduction in RMSE). The MAE, the Pearson's r, and the confidence level are all significantly improved.

To better understand how features like phase information can affect the CBH estimate, Figure 2(A) presents the statistical PIs for all features listed in Table 1 for GETCBH and in Figure 2(B) for the phase-enhanced GETCBH. All features for each VIS/NIR band are permutated simultaneously to calculate an overall PI for that band. The 27 levels of RH or temperature are split into two groups: lower layer (1000 – 650 hPa, 13 levels) and upper layer (600 – 100 hPa, 14 levels). These two layers are dominated by liquid and ice clouds, and PI

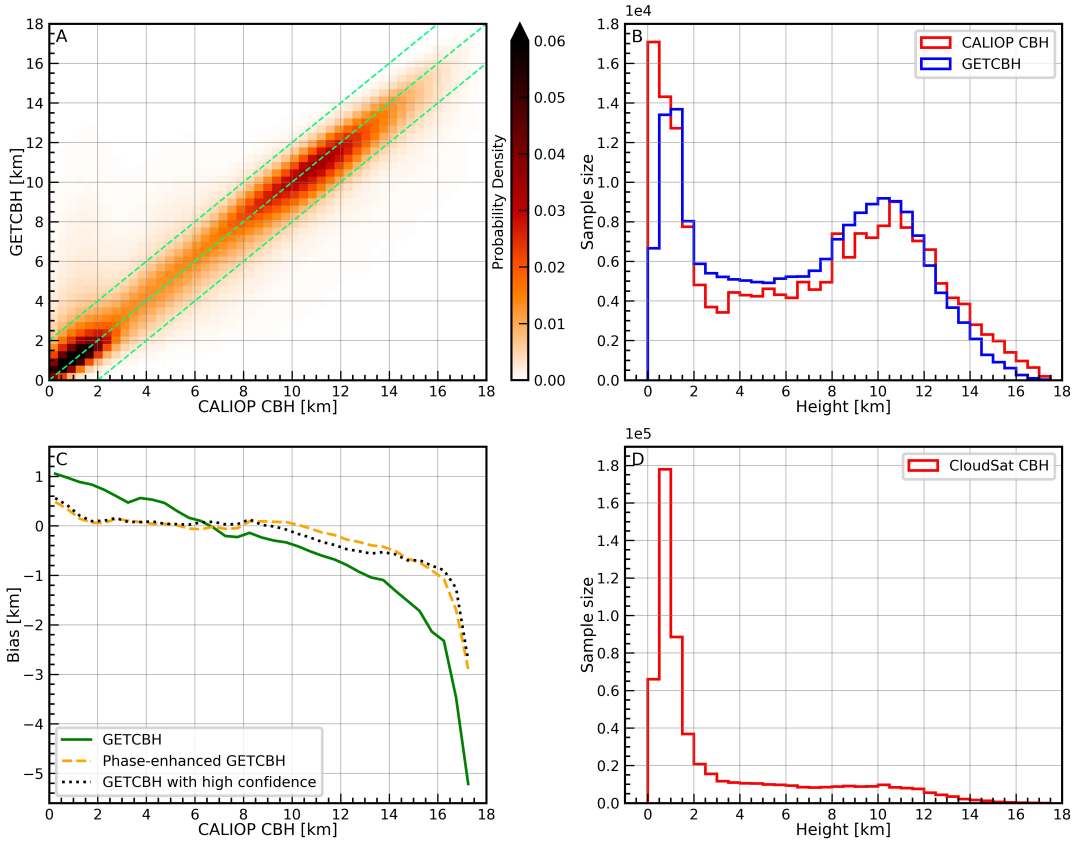


Figure 1: (A): Probability density scatter plot of CBH from CALIOP and GETCBH with the CALIOP CBH validation set. The color-bar shows the probability density, and three green dash lines illustrate an error of -2, 0 and 2 km. (B): The CBH histograms binned at 0.5 km from CALIOP (red) and the GETCBH (blue). (C): The MBE of the CBH estimate from GETCBH (green solid line), the phase-enhanced GETCBH (orange dashed line) and high confidence samples (black dotted line) of the GETCBH experiment as a function of CALIOP CBH (binned at 0.5 km interval). Only intervals with sample sizes larger than 50 are shown. (D): Similar to (B), but for the distribution of the CloudSat CBH validation set. Figures are plotted using the validation datasets.

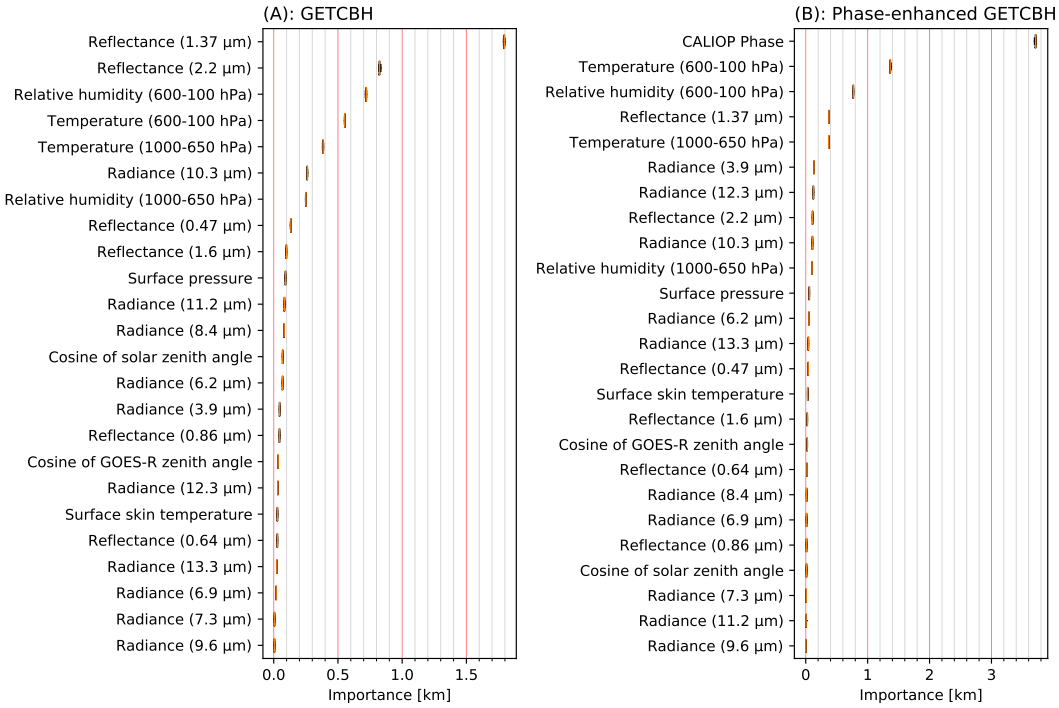


Figure 2: Permutation importance for (A) GETCBH and (B) the phase-enhanced GETCBH with the CALIOP CBH validation dataset. Whiskers represent the minimum and maximum of 10 times permutation, while the boxes (may be too narrow to be discerned) and orange line mean 25% and 75% quantiles and medium.

for each layer is calculated. And for both GETCBH and the phase-enhanced GETCBH, all PIs are positive, indicating every feature has a positive contribution to the CBH estimate.

In Figure 2(B), the cloud phase has an overwhelmingly large PI of 3.71 km, which explains the great improvement in the phase-enhanced GETCBH compared to GETCBH. The phase information greatly mitigates the estimate errors where high ice clouds are estimated with low CBHs and low water clouds are estimated with high CBHs. The percentage of samples with uncertainties larger than 4 km is greatly reduced to 0.7% in the phase-enhanced GETCBH from 5.0% in the GETCBH. In addition, the vertical variation of the bias is reduced as well. On average, the bias measured by the absolute MBE is reduced by 62%. It is calculated by the following two steps: 1) summing up the absolute value of MBEs in different height intervals for GETCBH and the phase-enhanced GETCBH, and 2) calculating the reduced percentage of the sum. It is interesting that the phase PI of 3.71 km is even larger than the GETCBH RMSE at 1.87 km. The permutation of the phase information degrades the accuracy of

the phase feature in the phase-enhanced GETCBH. This in turn leads to degraded CBH estimates, which are even worse than those from GETCBH where no direct phase information is used as a feature. This analysis indicates that a) accurate phase information is very useful in improving the CBH estimates, and b) incorrect or inaccurate phase information may degrade the CBH estimates. Therefore, if the phase information is not accurate enough, it should not be used as a feature.

The ABI reflectance at 1.37 μm has the largest PI for GETCBH at 1.79 km (Figure 2A), but has less significance in the phase-enhanced GETCBH with a PI of 0.37 km (Figure 2B). This likely indicates there exists information overlap between the 1.37 μm band and the CALIOP phase. This is consistent with the fact that the ABI 1.37 μm band is a strong water vapor absorption band widely used for cirrus detection. In other words, the contribution of the 1.37 μm band reflectance to the CBH estimates resides in its information about the cloud phase. The contribution is significantly reduced when better phase information from CALIOP is added. Figure 3 shows the distribution of the 1.37 versus 2.2 μm band reflectance for liquid and ice clouds. The 1.37 μm band reflectance of liquid clouds tends to be small, because of strong moisture absorption above the low cloud top. Conversely, the ice clouds tend to have a larger 1.37 μm band reflectance because of less moisture absorption above the high cloud top. The different responses of the 1.37 μm band reflectance to liquid and ice clouds provide useful phase information to the CBH estimate. Such information should be useful for CTH estimation as well. Similarly, another possible phase related feature is the 2.2 μm band reflectance used for cloud particle size estimates. The liquid and ice clouds show different peaks for their distributions of the 2.2 μm band reflectance (Figure 3). This feature has the second highest PI value in GETCBH. After the phase information from CALIOP is introduced, the 2.2 μm band reflectance PI decreases from 0.82 km to 0.14 km. This analysis identifies two phase-sensitive features and verifies the importance of phase information in the CBH estimate.

In Figure 2(B), apart from CALIOP Phase, the ERA5 upper layer temperature and RH information have the second and the third largest PIs, along with the lower layer temperature in fifth. In Figure 2(A), ERA5 has 5 features with PI values in the top 10. However, this does not mean that ERA5 fields contribute more CBH related information than ABI

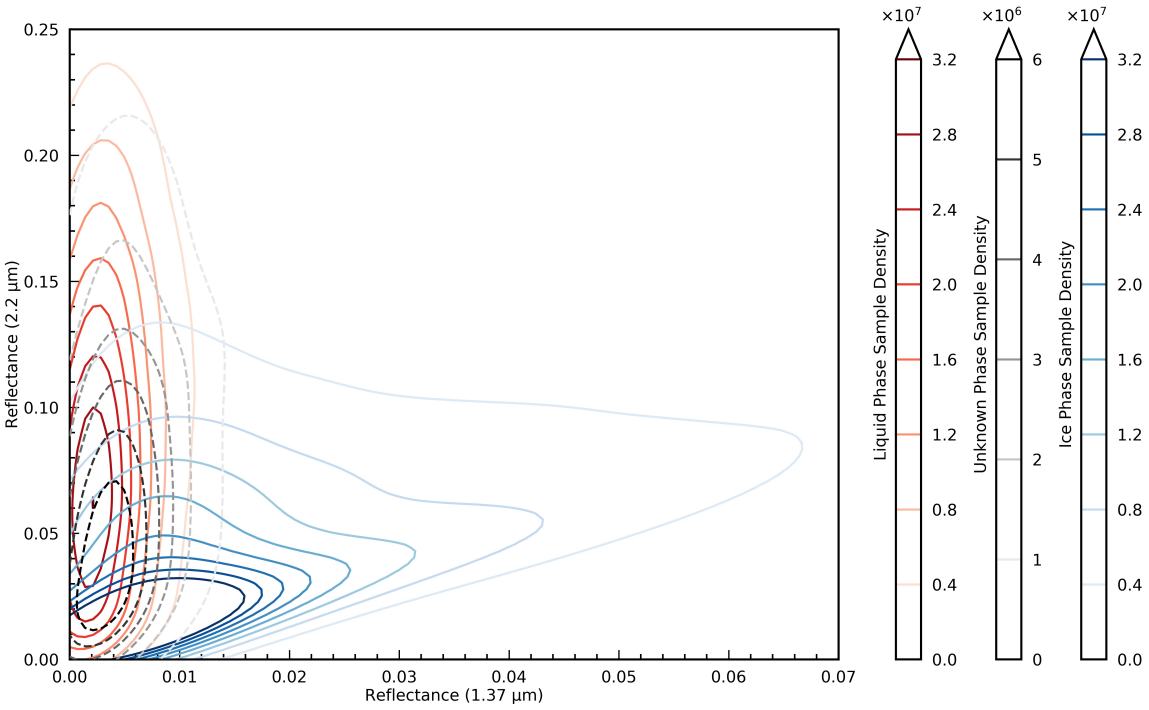


Figure 3: Sample density on reflectance of 1.37 and 2.2 μm bands. The red solid contour lines are for liquid phase samples and blue ones for ice phase. The black dashed contour lines stand for the samples with unknown phase.

measurements. Comparing to GETCBH in Figure 2(A), the main reason that the two largest PIs of ABI features are degraded is that the more accurate phase information from CALIOP is included in Phase-enhanced GETCBH. To better understand the values of the ERA5 fields and the ABI measurements in GETCBH, two experiments are conducted: 1) only ABI features are used as input and 2) only ERA5 features are used as input. The RMSE and Pearson's r of validation set are 2.14 km and 0.89, respectively, for the first model, while they are 2.60 km and 0.83, respectively, for the second model. These results indicate that the satellite measurements from ABI provide more useful information to CBH estimates than the ERA5 data.

When working with ABI observations, CALIPSO/CALIOP phase information is not always available. And given the importance of accurate phase information in improving the accuracy of the CBH estimate, two further tests are conducted. In the first test, the cloud top phase information from ABI L2 products was used as a feature to replace the CALIOP cloud phase information in the phase-enhanced GETCBH model. The validation shows reduced accuracy for the CBH estimate. The RMSE increased by about 0.03 km compared to GETCBH. This is probably due to two reasons: a) most of the information used for the ABI L2 phase retrieval is already included in the GETCBH features; and b) the ABI L2 cloud phase product is for the cloud top while phase information for the whole cloud layer is needed for a reliable CBH estimate. These results indicate that the ABI L2 cloud phase product as a proxy for all cloud layer phase information might not be accurate enough. Using it causes more degradation to the CBH estimate than not using it. This is consistent with previous findings that inaccurate phase information may degrade the quality of the CBH estimates.

Furthermore, Figure 3 shows there is substantial information about cloud phase in the ABI L1b measurements, especially the 1.37 and 2.2 μm reflectance. It is therefore possible to estimate cloud phase from ABI L1b measurements using a machine learning technique. In the second test, CALIOP phase index is used as a label, and cloud phase indices are estimated by the GBDT algorithm using the same input features and dataset as used by GETCBH (see Table 1 and Section 2.3). After a similar tuning process for GETCBH (see Section 2.2), the result of the phase estimate shows a total correct classification rate of 84.4%. Unfortunately, when adding this phase estimate as an additional feature to GETCBH, the improvements are

Table 3: CBH errors (in km for MAE, MBE and RMSE) using GETCBH with the validation set in different COD ranges.

COD range	Sample Size	MAE	MBE	RMSE	Pearson's r
[0.0, 0.1)	48978	1.62	0.29	2.57	0.88
[0.1, 0.5)	68736	1.17	0.05	1.80	0.92
[0.5, 1.0)	31022	0.96	-0.02	1.40	0.92
[1.0, 2.0)	24442	0.87	-0.31	1.29	0.94
[2.0, 3.0)	11900	0.80	-0.50	1.28	0.94
[3.0, 8.5)	9869	0.78	-0.55	1.30	0.88

subtle (RMSE decreases by less than 0.01 km in the best case). A similar test was carried out for CTH, and no obvious improvements were seen as well. The outcome indicates that GETCBH has already taken good advantage of the cloud phase and CTH information from the input data (e.g., ABI reflectance and radiance) and successfully used such information to improve the estimates. The CBH estimates can be further improved if additional cloud phase information with higher accuracy is available.

Apart from the phase information, background information also demonstrates substantial importance in Figure 2. The RH profile of the upper layer in GETCBH with a PI of 0.72 km exhibits the third highest PI, followed by the temperature of the same layer with a PI of 0.55 km. The temperature and RH profiles of the lower layer have the fifth and seventh highest PIs. The upper layer having higher PIs than the lower layer can be attributed to the larger sample size of ice clouds versus liquid clouds. The temperature profile is important because of its close connections to cloud radiative emission and cloud phase. The temperature profile is also proven to have great importance for CTH estimate by exhibiting the second highest importance in Min et al. (2020). In addition, RH is highly related to the formation of clouds (van Heerwaarden and de Arellano, 2008). These results indicate that both atmospheric temperature and RH background data are helpful for CBH estimates, in addition to ABI L1b measurements.

Beyond the impact of cloud phase, it is also important to understand how the GETCBH model performs under different cloud optical depths (COD). As mentioned in Dupont et al.

(2010), the CALIOP data for clouds with optical depths ranging from 0.1 to 3 have a relatively high signal-to-noise ratio (SNR). Based on CALIOP penetrating power, in this study, clouds with CALIOP measured COD smaller than 0.1 are defined as optically thin clouds, and clouds with COD larger than 3.0 are defined as optically thick clouds. It is important to point out that the opaque clouds in this study are not defined based on COD, but as those identified by CALIOP. A layer is considered opaque if 1) it is the lowest feature detected in a column, and 2) it is not classified as a surface return. In general, opaque clouds are optically thick, but not all thick clouds are opaque.

In Table 3, the statistics from the GETCBH are presented for different CODs. The CALIOP COD product is used as the classifier to categorize the samples into six groups. Comparing the results, samples with $1.0 \leq \text{COD} < 3.0$ have smaller RMSE and larger Pearson's r. Clouds optically thinner or thicker have larger RMSE and smaller Pearson's r. Specifically, CBH estimates for optically thin clouds have the largest RMSE of 2.57 km and a relatively low Pearson's r of 0.88. ABI measurements have reduced information for optically thin clouds. Estimating CBH in such cases is difficult, simply because not enough cloud information is in the input. For optically thick clouds, due to increased absorption by cloud particles, there is less information about the cloud base in ABI measurements, and the CALIOP CBH as the label has reduced accuracy as well. As a result, there is a slight degradation in the RMSE of the CBH estimate and a substantial reduction in Pearson's r comparing to clouds with $1.0 \leq \text{COD} < 3.0$. This analysis indicates that GETCBH can be applied to clouds with $0.0 \leq \text{COD} < 8.5$ and is more accurate for clouds with $0.1 \leq \text{COD} < 3.0$.

Note that the sample size for clouds with $3.0 \leq \text{COD} < 8.5$ is much smaller than clouds with $\text{COD} < 0.1$. That is because most optically thick clouds do not pass the CALIOP QC and are not included in the training and validation datasets. When applying GETCBH for those clouds that would have not passed the CALIOP QC, the CBH estimates may not be as accurate as shown in Table 3. Note that the upper limit of the COD of 8.5 is the maximum CALIOP COD in the CALIOP training and validation datasets.

From Table 3, it appears that COD can be used for generating the QF for the estimated CBH. For example, the COD can be estimated simultaneously along with the CBH. Then

the estimated COD can be used to remove optically thin clouds to improve the overall estimate quality. However, evaluations show that such COD estimates are not accurate enough (Pearson's r of 0.57) to warrant a meaningful QF for two reasons: 1) the COD mentioned above is referenced to the CALIOP 532 nm wavelength, which does not overlap with any ABI band and 2) there is not enough information about thin clouds with COD $<$ 0.1.

3.3. Quality Flag for CBH Estimate

Table 2 shows the overall quality of the CBH estimates. However, for users, a CBH QF before using the product is indispensable. Previous analysis has revealed that the CBH estimate quality is associated with several factors, such as the CBH itself, cloud phase, and COD. Directly combining the information from these sources to generate a QF is difficult. However, the GETQF model is built to estimate the quality of each CBH estimate using the GBDT algorithm.

The label of the GETQF is designed as a 2-class flag, which are classified based on the absolute error of the CBH estimates. A threshold of 1.6 km is explicitly set in our study. Low confidence is defined as having errors greater than the threshold, with high confidence defined as having errors smaller than the threshold. This threshold is configurable. The threshold determination for the GETQF model is based on the balance between the yield for the class of high confidence and its confidence level. A large threshold results in more CBH estimates with errors larger than 2 km mis-classified as high confidence. Therefore, the confidence level for the class of high confidence is small, although the yield is large. On the other hand, a small threshold may result in more CBH estimates with errors smaller than 2.0 km mis-classified as low confidence. Thus, the yield for the class of high confidence is small, although the confidence level is large. Besides, the threshold of 1.6 km is found effective in classifying CBH estimates for deep convective clouds as low confidence. In the training, weights of low confidence samples are set to increase linearly with the increase of absolute errors (absolute error / threshold for each sample) in sample, while high confidence samples are all weighted as 1. The samples with large errors are emphasized since the larger error for a given sample the more weight it is given. In addition to the features shown in Table 1, the CBH estimated by GETCBH is added as one more input feature due to its close

Table 4: CBH error (unit of km for MAE, MBE and RMSE) and confidence level of GETCBH with high/low confidence.

Sample Set	Sample Size	MAE	MBE	RMSE	Pearson's r	Confidence Level (± 2 km)
High Confidence	157482	0.74	0.01	1.14	0.97	96%
Low Confidence	37467	2.95	-0.10	3.58	0.69	36%

relationship to the estimate error and its ability to boost the classification accuracy (which is measured by the number of correctly classified samples / total samples, from 0.86 to 0.89) of the GETQF model. The setting of hyperparameters was introduced in Section 2.2. The CALIOP training and validation sets were used for GETQF.

The validation results are present in Table 4. The ± 2 km interval used to define the confidence level is chosen by referencing the VIIRS CBH product requirement (Baker, 2011). The two classes reasonably split the samples into two groups of different accuracies. The class of high confidence has a sample size about 81% of total samples. It has a confidence level of 96%, meaning 96% of errors within are ± 2 km, an MAE of 0.74 km, an RMSE of 1.14 km, and Pearson's r of 0.97. All these error indicators are comparable to those from the phase-enhanced GETCBH in Table 2. In addition, the bias is reduced (Figure 1C) by 60% on average compared to GETCBH results, which are also comparable to the phase-enhanced GETCBH results. The class of low confidence has a much lower estimate quality with an RMSE of 3.58 km, while the confidence level is only 36%. These results indicate that a) the estimates with low confidence are the main reason for the overall large RMSE of 1.87 km in Table 2, and b) the threshold of 1.6 km successfully separates good CBH estimates from bad ones.

To better illustrate the capability of GETQF, GETCBH is compared to the climatological guess for low/high clouds. The most frequent CBH values, which are about 1.00 km for low clouds and 10.75 km for high clouds, are taken as the climatological guess. For low clouds, the RMSE of GETCBH is 1.96 km (0.65 km or 25% smaller than the climatological guess). For high clouds, the RMSE of GETCBH is 1.79 km (0.5 km or 22% lower than the climatological guess). The quality control with GETQF further improves the CBH estimate accuracy.

For low clouds with high confidence, the RMSE of GETCBH is 1.18 km (1.43 km or 55%

smaller than the climatological guess). For high clouds with high confidence, the RMSE of GETCBH is 1.09 km (1.20 km or 52% smaller than the climatological guess). This analysis shows the improvement of the GETCBH over the simple climatological guess for both high and low clouds. The improvement becomes more significant after QC by GETQF.

Table 3 shows that optically thin clouds have the largest RMSE (2.57 km). It is larger than the general RMSE (1.87 km) of GETCBH. The large RMSE is caused by the lack of cloud information in the ABI L1b measurements. Fortunately, such samples can be identified by the GETQF. After filtering by GETQF, the optically thin clouds have 70.7% samples with high confidence. The high confidence samples have an RMSE of 1.44 km and Pearson's r of 0.96, respectively. These results indicate that, with the help of GETQF, GETCBH can be applied to clouds with COD < 8.5.

3.4. CBH Estimate with CloudSat Dataset

As analyzed in Section 3.3, no samples with COD larger than 8.5 are included in the CALIOP training and validation datasets for GETCBH. The CBH estimates for those clouds are expected not to be as accurate, and they are expected to be flagged with low confidence by the GETQF. The RMSE and Pearson's r of CBH estimates are also slightly degraded for clouds with $3.0 \leq$ COD < 8.5 compared with those for clouds with $1.0 \leq$ COD < 3.0. So, it appears that GETCBH is less reliable for optically thick clouds. Two factors may cause these results. First, the CALIOP CBH is not accurate enough due to strong cloud attenuation. And second, the ABI measurements have limited CBH information. In either case, the GETCBH model is not optimally trained, and the CBH estimate is going to have reduced accuracy.

To better understand these two reasons, the CloudSat-based GETCBH is built based on the 2B-GEOPROF-LIDAR product from CloudSat with higher penetrability (see Section 2.2). The CloudSat data are only available for a 9-month overlap with ABI/GOES-16 (October 10, 2018 through July 10, 2019). The majority of cloud samples in the northern hemisphere summer are not available. So, it is difficult to carry out a comprehensive analysis of the performance of the CloudSat-based GETCBH. However, it should provide some insights as to whether the GBRT algorithm can better estimate the CBH for optically thick clouds using information from the ABI and ERA5.

Table 5: CloudSat-based GETCBH errors (in km) using the CloudSat validation dataset in different COD ranges.

COD range	Sample Size	MAE	MBE	RMSE	Pearson's r
[0.0, 0.1)	34367	1.34	-0.19	2.35	0.83
[0.1, 0.5)	66963	1.37	0.50	2.22	0.89
[0.5, 1.0)	34760	1.14	0.44	1.84	0.90
[1.0, 2.0)	30425	0.84	0.11	1.38	0.94
[2.0, 3.0)	25986	0.75	-0.04	1.25	0.94
[3.0, 8.5]	251027	0.61	-0.04	1.05	0.93
[8.5, 45.6)	61537	0.33	-0.09	0.64	0.81

The CloudSat based CBHs have a different distribution to CALIOP based CBHs. As shown in Figure 1(D), the sample size of CloudSat based CBHs within 0.5-1.0 km is around 1.79e5, much larger than that of the high cloud peak (CBHs within 10-10.5 km) around 0.10e5. It is attributed to the higher penetrability of CPR to clouds. CPR could observe low CBHs of optically thick clouds. This will have substantial impact on the validation statistics. This will have substantial impact on the validation statistics.

Table 2 shows the overall accuracy of the CloudSat-based GETCBH. For Pearson's r, the CloudSat-based GETCBH is comparable to GETCBH, both of which are 0.92. The CloudSat-based GETCBH does show a reduced RMSE of 1.42 km, substantially smaller than 1.87 km of GETCBH. This RMSE reduction is mostly due to the different CBH distributions of these two datasets (Figure 1B, D). The low CBHs (CBH $<$ 7 km) account for 82% of the total samples in the CloudSat dataset and 49% in the CALIOP dataset. The two datasets are also different in many other aspects, such as the sample size, period, high/low cloud ratio, etc. Therefore, the comparison between them based on RMSE, MAE, and MBE is not objective. The Pearson's r will be used as the primary parameter for the comparison.

Table 5 presents the performance of the CloudSat-based GETCBH in different COD ranges. Similar to GETCBH in Table 3, the CloudSat-based GETCBH has the best performance for $1.0 \leq$ COD < 3.0 with a Pearson's r of 0.94, equal to that of GETCBH. For clouds with COD < 1.0, the CloudSat-based GETCBH has a lower Pearson's r than GETCBH. This

could be caused by the influence of the surface. For clouds with COD < 1.0, near-surface clouds are dominant in CloudSat-based samples, accounting for about 57% of samples versus 21% for CALIOP samples. For those clouds, the IR radiances are found to have relatively large PI scores for the CBH estimate, yet they have limited information about clouds. As a result, the CloudSat-based GETCBH algorithm could be confused by those radiances, leading to reduced performance. For clouds with $3.0 \leq$ COD < 8.5, the CloudSat-based GETCBH has a larger Pearson's r of 0.93. It is substantially larger than the 0.88 for GETCBH. These results indicate that the CloudSat-based GETCBH can handle optically thicker clouds better than GETCBH. It also suggests that ABI measurements may have more CBH information for optically thick clouds than those extracted by the GETCBH model. However, as pointed out earlier, the CloudSat only has 9 months of data. The limited sample does not warrant a mature CBH estimate model. Hence the case demonstration in the next section still uses the GETCBH model. Again, note that the much-reduced RMSE of 1.41 km from 1.87 km for GETCBH is mostly due to the dominance of low clouds.

For optically thick clouds with COD ≥ 8.5 (Pearson's r of 0.81), there is increased absorption by cloud particles. In addition, there is less information about the cloud base in the ABI measurements, making the CBH estimate difficult. This becomes more difficult for optically very thick clouds, such as deep convective clouds, which also have a large vertical extent. Nevertheless, for these clouds, the CloudSat-based GETCBH may learn that they are optically thick clouds. Even with little information about CBH in the ABI measurements, however, the model may assign the CBHs with information from the temperature and moisture profiles. The model may learn such information from either ERA5 profiles or the ABI measurements. Further analysis reveals that clouds with CBH < 2 km (sample size of 55,617) have an RMSE of 0.45 km, which is much smaller than the RMSE of 1.23 km for the clouds with CBH ≥ 2 km (sample size of 5,920). Overall, the relatively small RMSE of 0.64 km indicates that the CloudSat-based GETCBH may provide reasonable CBH estimates for some optically thick clouds, i.e. low clouds.

It is important to point out that the CODs shown in Table 5 are CALIOP measurements. They are not fully quality controlled, i.e., by extinction QC flags or cloud-aerosol discrimination score. Therefore, the accuracy of those CODs is not as reliable. The larger the CODs,

the less reliable they are. There are two major uncertainties: 1) for opaque clouds, the retrieved CODs are the CODs between CALIOP and the lowest feature it observed; 2) for water clouds, the multiply-scattered component of the signal may not be fully represented. Consequently, it is possible the actual COD is larger than those reported by CALIOP. However, the conclusion that the CloudSat-based GETCBH is more accurate in charactering CBH for optically thick clouds should stand.

In table 2, the Pearson's r of GETCBH, the phase-enhanced GETCBH, and the CloudSat-based GETCBH for the corresponding training dataset are 0.93, 0.98, and 0.94, respectively. All three of them are comparable to or slightly larger than the evaluation statistics (in Table 2) for the corresponding validation set. Other statistics from the training, such as MAE, MBE, and RMSE, are also comparable to or slightly smaller than the validation results. These slight degradations from the training to validation suggest that overfitting in the training is not severe.

3.5. Case Demonstration

GETCBH is applied to Hurricane Dorian on 5 September 2019. Specifically, in Figure 4, there are three regions that we will give special attention to: the high clouds over the Caribbean region (domain B), the hurricane (domain C), and the low clouds over the Pacific Ocean off the coast of South America (domain D). This case is chosen for its coverage of different types of clouds, such as high, low, single-layer, multi-layer, and deep convective clouds. This allows better understanding of the performance of GETCBH for different cloud types.

CBH estimates with low confidence (black shadow in Figure 4B, C and D) mainly occur in two scenarios: 1) the edges between low clouds and high clouds and 2) the clouds within and around Hurricane Dorian. For the boundaries between low and high clouds, such as the area around (90°W, 24°S) in Figure 4(D) and the area to the south of 13°N in Figure 4(B), the low confidence is due to the presence of multi-layer clouds. For clouds around the storm center (79°W, 33°N) in Figure 4(C), they are known as deep convective clouds and are optically very thick. There is very little information on the cloud base remaining after attenuation by the cloud layer. These two types of clouds are not included in the training. GETQF successfully identifies those CBH estimates as having low confidence.

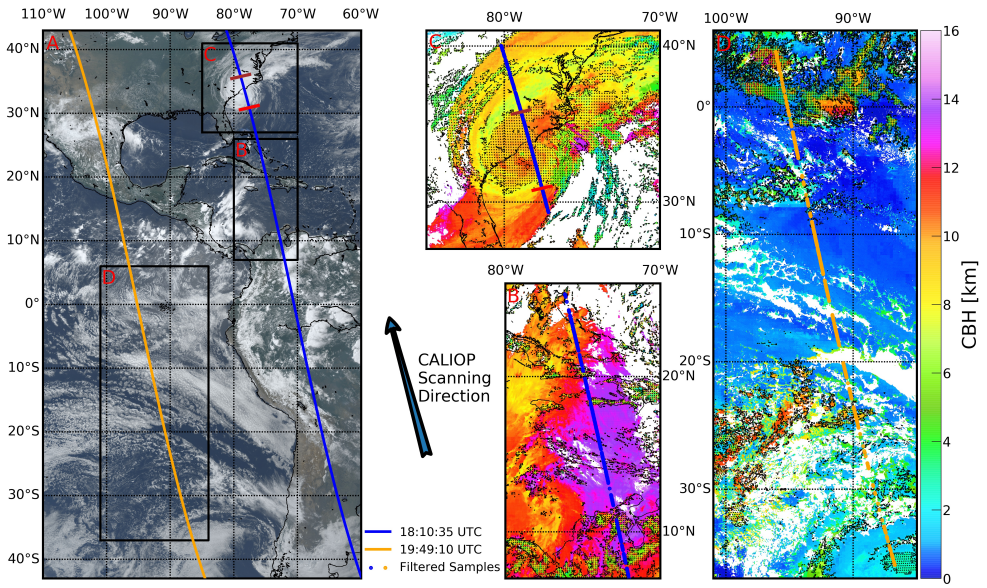


Figure 4: Hurricane Dorian and low/high cloud case study on Sep. 5th, 2019. A: ABI Natural True Color imagery at 18:40:12 UTC. The blue and orange lines show the CALIOP scanning tracks with the start time and direction in the right-side legends. Panels B, C and D show the CBH estimated by the GETCBH corresponding to regions 1, 2 and 3 in panel A. The ABI observation start-time of panels B and C is 18:40:12 UTC and that of panel D is 20:00:12 UTC. The clear sky regions are determined by the ABI L2 Clear Sky Mask product and shown in white color. The regions shadowed with black dots show CBH estimates with low confidence. The blue and orange dot lines show the CALIOP track. The red and brown lines indicate the extent of Hurricane Dorian in panel C as determined by the CALIOP CBH products and quality flags. It is worth noting that cloud patterns in panel D are slightly different from those in panel A due to the time difference.

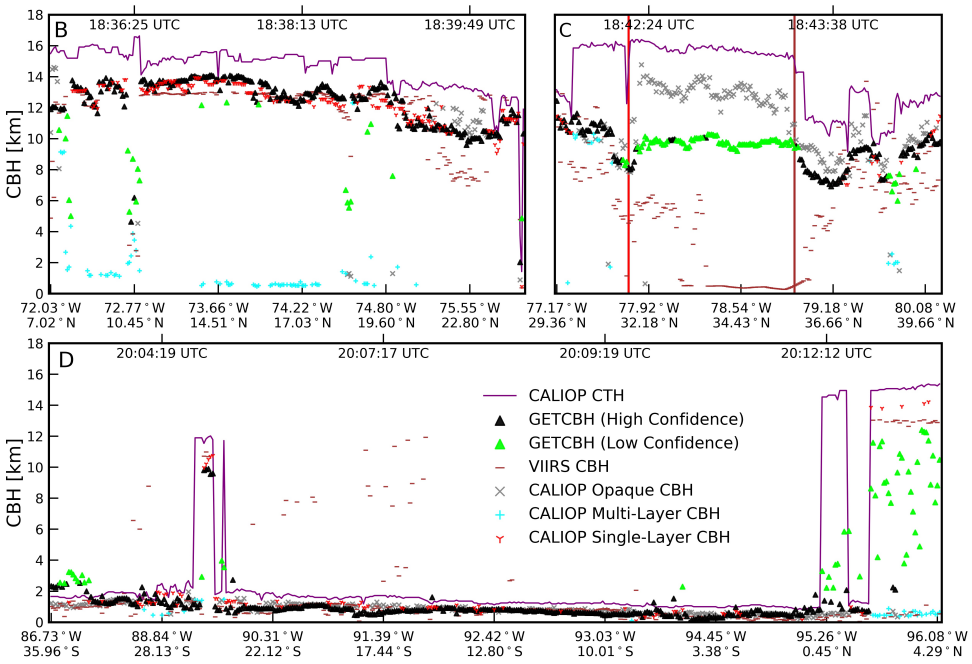


Figure 5: The cross-sections along CALIOP scanning tracks for Hurricane Dorian and low/high cloud case study. B, C and D: The cross-sections refer to the blue and orange dot lines in Figure 4(B), (C) and (D), respectively. The red and brown lines representing convective region in panel C are also marked in Figure 4(A) and (C). The purple line represents CTH estimates and the grey cross, cyan plus and red Y markers are opaque CBH (the lowest cloud feature detected from the cloud opaque to CALIOP), multi-layer CBH, and single-layer CBH from CALIOP observation respectively. Note that the opaque cloud can be either single- or multi-layer cloud. The black (lime green) triangle marker is the GETCBH estimated CBH with high confidence (low confidence). The brown bar marker is the valid CBH (invalid VIIRS CBHs are excluded) estimated from VIIRS observations.

Note that in ABI measurements, the samples with low confidence will have a larger percentage than those shown in Table 4 because some samples, i.e., thick or multi-layer clouds, are expected to have low confidence (as in Figure 4 and 5). However, those samples are excluded from the training and validation datasets (see Section 2.3). For the cases demonstrated here, the full disk has 35% of the cloudy pixels identified as low confidence compared to 19% in Table 4.

In the cross-section view shown in Figure 5, the CBH estimates (GETCBH) of single-layer clouds agree with the CALIOP observation (red Y) reasonably well for all cloud types (Figure 5B, C and D). Most of those CBH estimates have errors less than 1 km. In fact, those CBH

estimates are successfully characterized as high confidence (black triangles), for both high and low clouds. Most of the CBH estimates with low confidence (lime green triangles) occur in two types of clouds: opaque clouds (clouds which are opaque to CALIOP) and multi-layer clouds. These estimates have low confidence because the problem for multi-layer clouds is complex and GETCBH is not optimal for opaque clouds. The CBH estimate errors for those clouds are larger than 2 km with some as large as 12 km for multi-layer clouds.

For opaque clouds with low confidence, for example those over Hurricane Dorian (2019) in Figure 5(C), the ABI CBH estimates from GETCBH are clustered around 10 km. The CALIOP CBHs are around 13 km for opaque clouds; these incorrect results are caused by the opaqueness. These are within the hurricane in a region with strong convection and thus deep convective clouds. The CBHs are typically within 1 km of the surface (Hence and Houze, 2011). The ABI CBH estimates with large estimate errors are successfully characterized with low confidence. Outside Hurricane Dorian's convective zone, CALIOP detects many opaque high clouds as well. These are mostly outflow cirrus with much higher CBHs than within the convective zone. The CBH estimates for these high opaque clouds appear to be reasonable for two reasons: 1) the estimates are comparable to the CALIOP CBHs for some of the multi-layer clouds around 31°N in Figure 5(C) (to the left of the red vertical line), and 2) the estimates are lower than the CALIOP opaque CBHs. To the right of the brown vertical line in Figure 5(C), around 38.6°N, there are some opaque/multi-layer clouds estimated with low confidence. These correspond to the clouds in the outer band of the hurricane to the west and north of the storm. In Figure 4(C), to the west of 80°W, between the black dotted regions of the storm and the clouds in the outer band to west of the storm, there is a half-ring-shaped region, where the estimated CBH exhibits high confidence, which corresponds to the hurricane outflow cirrus.

For opaque low clouds with low confidence at 35°S in Figure 5(D), GETCBH overestimates the CBH by 1-2 km when compared with the CALIOP measurements. However, for the majority of the CALIOP opaque low clouds in Figure 5(D), the estimates are characterized with high confidence, and the estimated CBHs are close to the CALIOP CBHs. Of those CBH estimates, 71.4% have errors smaller than 0.3 km. This is better seen from the zoom-in in Figure 6(D). These results show that the CBHs from GETCBH are highly accurate for

low clouds over oceans with quality flags of high confidence.

In addition, most of the samples marked as CALIOP multi-layer CBH (cyan plus) exhibit significant overestimates (Figure 5B, C and D). This result is expected because 1) there are multiple cloud bases and 2) the extinction of signal over a large range. In one of the tests during the development of the GETCBH model, the RMSE of the CBH model is increased by 61.0% (from 1.87 km to 3.01 km) after including multi-layer cloud samples (20.6% of total samples) in both the training and validation datasets. This indicates that the ABI measurements have limited resolving power of the cloud base information for multi-layer clouds. Based on these results, multi-layer clouds were excluded in the collocation datasets.

The above analysis indicates that the GETCBH estimated CBHs are reliable with high confidence for most of the single-layer clouds. Since the training dataset contains only single-layer clouds with COD < 8.5 that passed the CALIOP QC tests, the estimate becomes less reliable when applied to single-layer clouds that are optically thick or to multi-layer clouds.

To further study the quality of CBH estimates from ABI observations, the CBH product from VIIRS (Seaman et al., 2017; Noh et al., 2017) temporally and spatially collocated with CALIOP samples is included in Figure 5 and 6. Note that both the Joint Polar Satellite System (JPSS)-1 spacecraft and the Suomi National Polar-orbiting Partnership (SNPP) spacecraft have a VIIRS instrument on board. Only the observations from JPSS-1 are used in this study because they are closer in time to the CALIOP observations (within 24 minutes compared to 33 minutes for SNPP).

In Figure 5 and 6, the CBH estimates from ABI and VIIRS both appear to have good accuracy for the high and low non-deep convective single-layer clouds. There are 185 non-opaque single-layer cloud samples in Figure 5. The RMSE of ABI CBH estimates for those single-layer clouds is 0.77 km, while the same statistic for VIIRS CBH estimates is 1.82 km. ABI CBH estimates from GETCBH show higher accuracy for non-opaque single-layer clouds in this case. This is not surprising since GETCBH is specially trained for such clouds. For clouds in the deep convective zone in Figure 5(C), VIIRS CBH estimates are significantly lower than ABI CBHs. Some are close to the ocean surface, which is reasonable for hurricanes. VIIRS CBH estimates take direct advantage of the CCL, which is estimated from a NWP analysis data in regions of deep convection, circumventing their lack of cloud base information

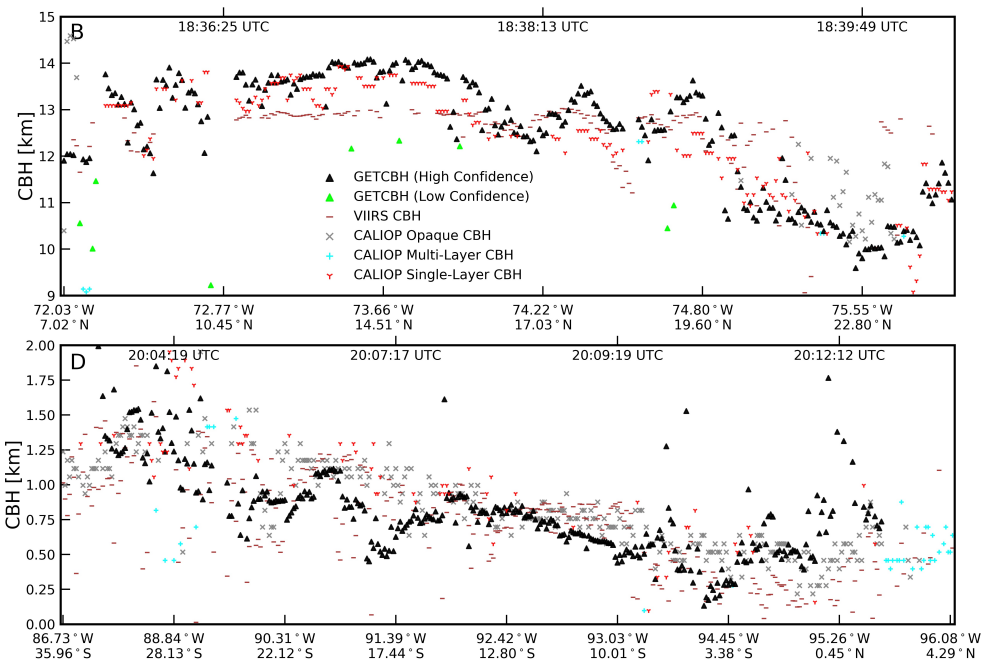


Figure 6: The zoom-in of the cross-sections along CALIPSO scanning tracks for the low/high cloud case study. B: High clouds with CBH between 9 and 15 km in Figure 5(B). D: Low clouds with CBH between 0 and 2 km in Figure 5(D).

in deep convective clouds. For multi-layer clouds, both ABI and VIIRS CBH estimates tend to have overestimates, as in the clouds to the north of 2°N in Figure 5(D). However, the ABI estimates successfully characterize those with low confidence.

It is interesting that samples around 35°S in Figure 5(D) have discrepancies between GETCBH and CALIOP CBHs around 1.5 km. It is likely due to the extremely dry air above those clouds which is verified by the total column water vapor from the ERA5 dataset (not shown). The 1.37 μm band is a water vapor absorption band. The reflectance of this band provides critical cloud phase information. It has the highest PI in GETCBH. The dry air above those clouds leads to high reflectance of this band, confusing the GETCBH to treat them as ice clouds and introducing large uncertainty in the CBH estimate. Fortunately, most of these samples are successfully marked as low confidence by GETQF.

In general, the ABI CBH estimates from GETCBH have high accuracy for single-layer non-opaque clouds in this case study, whether they are low or high clouds. While opaque and multi-layer clouds are challenging, GETQF successfully characterizes those CBH estimates

as low confidence.

4. Conclusions and Discussions

This study introduces ML algorithms to estimate daytime CBH from GOES-16 ABI Level 1b data, with help from the ERA5 reanalysis of temperature and moisture profiles. The CALIOP measured CBH is used as the label in the training process. A matchup file containing ABI reflectance and radiances, ERA5 profiles of temperature and relative humidity, as well as CALIOP cloud base height and cloud phase, is generated, covering 2 years (2018 and 2019). The split between training and validation datasets is 80% and 20% respectively.

The results show that GETCBH has an RMSE of 1.87 km for the CALIOP validation dataset. Analyzing the distribution of CBH bias reveals a vertical variation, i.e., underestimate for high ice clouds and overestimate for low water clouds. To reduce the vertical variation of the bias, the CALIOP cloud phase is added as one additional feature in GETCBH. The RMSE is reduced from 1.87 km to 1.05 km, a significant reduction of 43.9%. The Pearson's r increases from 0.92 to 0.98, while the vertical variation for the bias is reduced by 62%. These results indicate that the accurate cloud phase information is critically important to enhance the CBH estimates and significantly mitigates the vertically varied bias.

Feature impacts are analyzed using the permutation importance technique. The results confirm that accurate cloud phase information has the highest importance for CBH estimates. In addition, inaccurate or incorrect phase information may significantly degrade the CBH estimate as well. Such that if the phase information is not accurate enough, it should not be used as a feature. When there is no accurate cloud phase information, as in GETCBH, the reflectance of the ABI 1.37 μm band radiance becomes the most important feature, because this band contains significant information about cloud phase. When accurate cloud phase information is used as a feature, the importance of the 1.37 μm band is significantly reduced. This is also true for the 2.2 μm band. The background information (RH and temperature profiles) also plays an important role for CBH estimate.

Both ERA5 and CALIOP phase information are shown to be useful for improving the CBH estimate. It appears to be reasonable to include ABI L2 products as features to improve the CBH estimate. Experiments have been carried out to test several ABI L2 products as

additional features, such as cloud phase, COD, and CTH. And none of the three products improves the CBH estimate significantly. The RMSE is only reduced by less than 0.01 km in the best case. This is likely due to the fact that most of the information used for the ABI L2 phase retrieval is already included in the GETCBH features, especially the ABI L1b measurements. Adding them as additional features does not add additional useful information about CBH.

The CBH estimates using GETCBH are evaluated for different CODs. For clouds with COD from 0.1 to 8.5, the RMSE ranges from 1.28 to 1.80 km, lower than the overall RMSE of 1.87 km. The CBH estimates of clouds with COD from 1.0 to 3.0 have the smallest RMSE, likely due to the best balance between cloud signal from the base in ABI measurements and cloud attenuation in CALIOP CBH. Thinner clouds with reduced cloud signals, or thicker clouds with strong attenuation, have reduced CBH accuracy.

While the CBH bias/errors are found to be related to COD, cloud phase, and the CBH, it is difficult to use such information for quality control. GETQF is developed to provide the confidence level of the CBH estimates for each sample. For the validation dataset, the high confidence samples (81% of total samples) have a small RMSE of 1.14 km, and CBH estimates are 96% likely to have errors within $\pm$ 2 km. The samples with low confidence (19% of total sample) have an RMSE of 3.58 km, and only 36% of the CBH errors are within $\pm$ 2 km. These results indicate that estimates with low confidence are the main reason for the overall large RMSE of 1.87 km. For ABI measurements, about 65% of cloudy pixels can be characterized as high confidence CBH estimates.

The CloudSat-based GETCBH is experimentally built using the 9-month CloudSat 2B-GEOPROF-LIDAR product as the label. Compared with GETCBH, the CloudSat-based GETCBH can be better applied to optically thick clouds with $3.0 \leq \text{COD} < 8.5$, and has larger Pearson's r than for GETCBH. Like GETCBH, the CBHs of the CloudSat-based GETCBH have reduced accuracy for both optically thin clouds with COD <0.1 (Pearson's r of 0.83) and optically thick clouds with COD $\geq$ 8.5 (Pearson's r of 0.81). When there are enough CloudSat data, the CloudSat-based GETCBH should be used instead of the GETCBH to provide more accurate CBH estimates for optically thick clouds.

The GETCBH algorithm is applied to Hurricane Dorian (2019). The CBH estimates

are in line with CALIOP CBHs for most single-layer clouds with discrepancies less than 1 km. Those estimates are successfully characterized as high confidence by GETQF. Even though some clouds are opaque to CALIOP, the CBH estimates appear to be reasonable when compared with nearby non-opaque clouds or the surface. The CBH estimates for deep convective clouds and multi-layer clouds are characterized as low confidence, and are therefore not included in the training; in addition, the CBH estimates show much larger estimate errors. In the comparisons between ABI and VIIRS CBHs, the ABI CBH estimates show a better overall agreement with CALIOP for non-opaque single-layer clouds with an RMSE of 0.77 km versus 1.82 km for VIIRS.

There are three limitations of the GETCBH algorithms. First, the low confidence CBH estimates for deep convective clouds and multi-layer clouds have errors that can be more than 10 km and should be considered failed CBH estimations. For a multi-layer cloud, the problem is complex with multiple cloud bases and the extinction of signal over a large range. For deep convective clouds, there is simply little or no information from the cloud base. Both multi-layer and deep convective clouds are not included in the training dataset. Second, given the high importance of the 1.37 μm band reflectance, the CBH estimate can be problematic in some unusual scenarios. The reflectance of the 1.37 μm band is a water vapor absorption band, which makes it useful for phase determination of ice clouds. The water vapor content above low water cloud tops, if too little, may cause GETCBH to treat them incorrectly as ice clouds and introduce large uncertainties to the CBH estimate. Fortunately, GETQF appears to be able to recognize the confusion and assign low confidence to the CBH estimation in the Hurricane Dorian case demonstration. Third, the ERA5 data is useful in improving the CBH estimate accuracy, but they are not available in real-time or near real-time (NRT). If the GBRT based CBH estimation algorithm is to be run in NRT, the ERA5 data must be replaced with other NWP short-range forecast data that is available in real-time. A new NRT GBRT CBH model needs to be trained. That will likely result in a decrease in the performance of the CBH estimate because of reduced accuracy of NWP data.

In summary, the methodology developed in this study can be applied to geostationary advanced imager measurements for deriving a daytime CBH product with high spatial and temporal resolution. It is limited to single-layer clouds with optical thickness not too thick.

For Climate Data Record (CDR) generation, ERA5 and imager radiances can be used together as input; while for NRT CBH generation, the NWP short-range forecast data and the imager radiances can be used together as features to build an NRT GETCBH. More representative training sets are beneficial to estimate CBH product in NRT for various applications.

Acknowledgements

This work is partly funded by the National Natural Science Foundation 41975020 (NSF; for Min Min), the China Scholarship Council (CSC; for Han Lin), and the GOES-R Science Program. We would like to acknowledge all open access data and python packages used in this study, including but not limited to ERA5 reanalysis, CALIPSO/CALIOP, CloudSat, ABI/GOES-16 and scikit-learn packages.

References

Baker, N., 2011. Joint Polar Satellite System (JPSS) VIIRS cloud base height algorithm theoretical basis document (ATBD). JPSS Ground Project Code 474-00045, NASA GSFC. URL: http://www.star.nesdis.noaa.gov/jpss/documents/ATBD/D0001-M01-S01-015_JPSS_ATBD_VIIRS-Cloud-Base-Height.pdf.

Böhm, C., Sourdeval, O., Mülmenstädt, J., Quaas, J., Crewell, S., 2019. Cloud base height retrieval from multi-angle satellite data. Atmos. Meas. Tech. 12, 1841–1860. URL: https://www.atmos-meas-tech.net/12/1841/2019/, doi:10.5194/amt-12-1841-2019.

Clough, S., Shephard, M., Mlawer, E., Delamere, J., Iacono, M., Cady-Pereira, K., Boukabara, S., Brown, P., 2005. Atmospheric radiative transfer modeling: a summary of the aer codes. J. Quant. Spectrosc. Radiat. Transfer 91, 233–244. doi:10.1016/j.jqsrt.2004.05.058.

Costa-Surós, M., Stachlewska, I.S., Markowicz, K., 2016. Comparing water vapor mixing ratio profiles and cloud vertical structure from multiwavelength raman lidar retrievals and radiosounding measurements, in: EPJ Web of Conferences, EDP Sciences. p. 24005. doi:10.1051/epjconf/201611924005.

[dataset] European Centre for Medium-Range Weather Forecasts, 2017. Ecmwf reanalysis v5 (era5). URL: https://doi.org/10.5065/D6X34W69.

Diedenhofen, B., Musch, J., 2015. cocor: A comprehensive solution for the statistical comparison of correlations. PloS one 10, e0121945. doi:10.1371/journal.pone.0121945.

Dupont, J.C., Haeffelin, M., Morille, Y., Noël, V., Keckhut, P., Winker, D., Comstock, J., Chervet, P., Roblin, A., 2010. Macrophysical and optical properties of midlatitude cirrus clouds from four ground-based lidars and collocated CALIOP observations. J. Geophys. Res. Atmos. 115. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2009JD011943, doi:10.1029/2009JD011943.

Friedman, J.H., 2001. Greedy function approximation: A gradient boosting machine. Ann. Statist. 29, 1189–1232. URL: https://doi.org/10.1214/aos/1013203451, doi:10.1214/aos/1013203451.

Garrett, T.J., Schmidt, C.T., Kihlgren, S., Cornet, C., 2010. Mammatus clouds as a response to cloud-base radiative heating. J. Atmos. Sci. 67, 3891–3903. URL: https://journals.ametsoc.org/view/journals/atsc/67/12/2010jas3513.1.xml, doi:10.1175/2010JAS3513.1.

Gregorutti, B., Michel, B., Saint-Pierre, P., 2015. Grouped variable importance with random forests and application to multiple functional data analysis. Comput. Stat. Data Anal. 90, 15–35. doi:10.1016/j.csda.2015.04.002.

Håkansson, N., Adok, C., Thoss, A., Scheirer, R., Hörnquist, S., 2018. Neural network cloud top pressure and height for MODIS. Atmos. Meas. Tech. 11, 3177–3196. URL: https://www.atmos-meas-tech.net/11/3177/2018/, doi:10.5194/amt-11-3177-2018.

van Heerwaarden, C.C., de Arellano, J.V.G., 2008. Relative humidity as an indicator for cloud formation over heterogeneous land surfaces. J. Atmos. Sci. 65, 3263–3277. doi:10.1175/2008JAS2591.1.

Hence, D.A., Houze, R.A., 2011. Vertical structure of hurricane eyewalls as seen by the trmm

precipitation radar. J. Atmos. Sci. 68, 1637–1652. URL: https://journals.ametsoc.org/view/journals/atsc/68/8/2011jas3578.1.xml, doi:10.1175/2011JAS3578.1.

Kühnlein, M., Appelhans, T., Thies, B., Nauß, T., 2014. Precipitation estimates from msg seviri daytime, nighttime, and twilight data with random forests. J. Appl. Meteorol. Clim. 53, 2457–2480. doi:10.1175/JAMC-D-14-0082.1.

Liu, Z., Kar, J., Zeng, S., Tackett, J., Vaughan, M., Avery, M., Pelon, J., Getzewich, B., Lee, K.P., Magill, B., Omar, A., Lucker, P., Trepte, C., Winker, D., 2019. Discriminating between clouds and aerosols in the CALIOP version 4.1 data products. Atmos. Meas. Tech. 12, 703–734. URL: https://amt.copernicus.org/articles/12/703/2019/, doi:10.5194/amt-12-703-2019.

Mace, G.G., Zhang, Q., 2014. The cloudsat radar-lidar geometrical profile product (rl-geoprof): Updates, improvements, and selected results. J. Geophys. Res. Atmos. 119, 9441–9462. doi:10.1002/2013JD021374.

Mecikalski, J.R., Feltz, W.F., Murray, J.J., Johnson, D.B., Bedka, K.M., Bedka, S.T., Wimmers, A.J., Pavolonis, M., Berendes, T.A., Haggerty, J., Minnis, P., Bernstein, B., Williams, E., 2007. Aviation applications for satellite-based observations of cloud properties, convection initiation, in-flight icing, turbulence, and volcanic ash. B. Am. Meteorol. Soc. 88, 1589–1607. URL: https://journals.ametsoc.org/view/journals/bams/88/10/bams-88-10-1589.xml, doi:10.1175/BAMS-88-10-1589.

Miller, S.D., Forsythe, J.M., Partain, P.T., Haynes, J.M., Bankert, R.L., Sengupta, M., Mitrescu, C., Hawkins, J.D., Vonder Haar, T.H., 2014. Estimating Three-Dimensional Cloud Structure via Statistically Blended Satellite Observations. J. Appl. Meteorol. Clim. 53, 437–455. URL: https://doi.org/10.1175/JAMC-D-13-070.1, doi:10.1175/JAMC-D-13-070.1.

Min, M., Li, J., Wang, F., Liu, Z., Menzel, W.P., 2020. Retrieval of cloud top properties from advanced geostationary satellite imager measurements based on machine learning algorithms. Remote Sens. Environ. 239, 111616. URL: http://www.sciencedirect.com/science/article/pii/S0034425719306364, doi:10.1016/j.rse.2019.111616.

Min, M., Wu, C., Li, C., Liu, H., Xu, N., Wu, X., Chen, L., Wang, F., Sun, F., Qin, D., et al., 2017. Developing the science product algorithm testbed for Chinese next-generation geostationary meteorological satellites: Fengyun-4 series. J. Atmos. Sci. 31, 708–719. doi:10.1007/s13351-017-6161-z.

Mülmenstädt, J., Sourdeval, O., Henderson, D.S., L'Ecuyer, T.S., Unglaub, C., Jungandreas, L., Böhm, C., Russell, L.M., Quaas, J., 2018. Using CALIOP to estimate cloud-field base height and its uncertainty: the Cloud Base Altitude Spatial Extrapolator (CBASE) algorithm and dataset. Earth Syst. Sci. Data 10, 2279–2293. URL: https://essd.copernicus.org/articles/10/2279/2018/, doi:10.5194/essd-10-2279-2018.

Noh, Y.J., Forsythe, J.M., Miller, S.D., Seaman, C.J., Li, Y., Heidinger, A.K., Lindsey, D.T., Rogers, M.A., Partain, P.T., 2017. Cloud-Base Height Estimation from VIIRS. Part II: A Statistical Algorithm Based on A-Train Satellite Data. J. Atmos. Ocean. Tech. 34, 585–598. URL: https://doi.org/10.1175/JTECH-D-16-0110.1, doi:10.1175/JTECH-D-16-0110.1.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al., 2011. Scikit-learn: Machine learning in python. J. Mach. Learn. Res. 12, 2825–2830. doi:10.5555/1953048.2078195.

Persson, C., Bacher, P., Shiga, T., Madsen, H., 2017. Multi-site solar power forecasting using gradient boosted regression trees. Sol. energy 150, 423–436. doi:10.1016/j.solener.2017.04.066.

Ruppert, J.H., Wing, A.A., Tang, X., Duran, E.L., 2020. The critical role of cloud–infrared radiation feedback in tropical cyclone development. Proc. Natl. Acad. Sci. USA 117, 27884–27892. URL: https://www.pnas.org/content/117/45/27884, doi:10.1073/pnas.2013584117.

Schmit, T.J., Griffith, P., Gunshor, M.M., Daniels, J.M., Goodman, S.J., Lebair, W.J., 2017. A Closer Look at the ABI on the GOES-R Series. B. Am. Meteorol. Soc. 98, 681–698. URL: https://doi.org/10.1175/BAMS-D-15-00230.1, doi:10.1175/BAMS-D-15-00230.1.

939 Schmit, T.J., Gunshor, M.M., Menzel, W.P., Gurka, J.J., Li, J., Bachmeier, A.S., 2005.
940 INTRODUCING THE NEXT-GENERATION ADVANCED BASELINE IMAGER ON
941 GOES-R. B. Am. Meteorol. Soc. 86, 1079–1096. URL: 10.1175/BAMS-86-8-1079, doi:10.
942 1175/BAMS-86-8-1079.

943 Schmit, T.J., Li, J., Ackerman, S.A., Gurka, J.J., 2009. High-Spectral- and High-Temporal-
944 Resolution Infrared Measurements from Geostationary Orbit. J. Atmos. Ocean. Tech. 26,
945 2273–2292. URL: 10.1175/2009JTECHA1248.1, doi:10.1175/2009JTECHA1248.1.

946 Seaman, C.J., Noh, Y.J., Miller, S.D., Heidinger, A.K., Lindsey, D.T., 2017. Cloud-
947 Base Height Estimation from VIIRS. Part I: Operational Algorithm Validation against
948 CloudSat. J. Atmos. Ocean. Tech. 34, 567–583. URL: https://doi.org/10.1175/
949 JTECH-D-16-0109.1, doi:10.1175/JTECH-D-16-0109.1.

950 Sun, X.J., Li, H.R., Barker, H.W., Zhang, R.W., Zhou, Y.B., Liu, L., 2016.
951 Satellite-based estimation of cloud-base heights using constrained spectral radi-
952 ance matching. Q. J. Roy. Meteor. Soc. 142, 224–232. URL: https://rmets.
953 onlinelibrary.wiley.com/doi/abs/10.1002/qj.2647, doi:10.1002/qj.2647,
954 arXiv:https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.2647.

955 Tan, Z.H., Huo, J., Ma, S., Han, D., Wang, X., Hu, S., Yan, W., 2021. Estimating cloud
956 base height from Himawari-8 based on a random forest algorithm. Int. J. Remote Sens.
957 42, 2485–2501. URL: https://doi.org/10.1080/01431161.2020.1854891, doi:10.1080/
958 01431161.2020.1854891.

959 Tan, Z.H., Ma, S., Han, D., Gao, D., Yan, W., 2019. Estimation of cloud base height
960 for FY-4A satellite based on random forest algorithm. J. Infrared Millim. Waves 38,
961 381. URL: http://www.clp.ac.cn/EN/Article/OJbddf2ede2d22edd7, doi:10.11972/j.
962 issn.1001-9014.2019.03.020.

963 Wang, J., Rossow, W.B., 1998. Effects of Cloud Vertical Structure on Atmospheric Cir-
964 culation in the GISS GCM. J. Climate 11, 3010–3029. doi:10.1175/1520-0442(1998)
965 011<3010:EOCVSO>2.0.CO;2.

Wang, L., Zhang, Y., Yao, Y., Xiao, Z., Shang, K., Guo, X., Yang, J., Xue, S., Wang, J., 2021. Gbrt-based estimation of terrestrial latent heat flux in the haihe river basin from satellite and reanalysis datasets. Remote Sens. 13, 1054. doi:10.3390/rs13061054.

Wetherald, R.T., Manabe, S., 1988. Cloud feedback processes in a general circulation model. J. Atmos. Sci. 45, 1397 – 1416. URL: https://journals.ametsoc.org/view/journals/atsc/45/8/1520-0469_1988_045_1397_cfpiag_2_0_co_2.xml, doi:10.1175/1520-0469(1988)045<1397:CFPIAG>2.0.CO;2.

Winker, D.M., Hunt, W.H., Hostetler, C.A., 2004. Status and performance of the caliop lidar, in: Laser Radar Techniques for Atmospheric Sensing, International Society for Optics and Photonics. pp. 8–15. doi:10.1117/12.571955.

Winker, D.M., Hunt, W.H., McGill, M.J., 2007. Initial performance assessment of caliop. Geophys. Res. Lett. 34. doi:10.1029/2007GL030135.

Winker, D.M., Pelon, J., Coakley, J. A., J., Ackerman, S.A., Charlson, R.J., Colarco, P.R., Flamant, P., Fu, Q., Hoff, R.M., Kittaka, C., Kubar, T.L., Le Treut, H., Mccormick, M.P., Mégie, G., Poole, L., Powell, K., Trepte, C., Vaughan, M.A., Wielicki, B.A., 2010. The CALIPSO Mission: A Global 3D View of Aerosols and Clouds. B. Am. Meteorol. Soc. 91, 1211–1230. URL: https://doi.org/10.1175/2010BAMS3009.1, doi:10.1175/2010BAMS3009.1.

Yeo, H., Park, S.J., Kim, B.M., Shiobara, M., Kim, S.W., Kwon, H., Kim, J.H., Jeong, J.H., Park, S.S., Choi, T., 2018. The observed relationship of cloud to surface longwave radiation and air temperature at Ny-Ålesund, Svalbard. Tellus B 70, 1–10. URL: https://doi.org/10.1080/16000889.2018.1450589, doi:10.1080/16000889.2018.1450589.

Young, S.A., Vaughan, M.A., Garnier, A., Tackett, J.L., Lambeth, J.D., Powell, K.A., 2018. Extinction and optical depth retrievals for CALIPSO's Version 4 data release. Atmos. Meas. Tech. 11, 5701–5727. URL: https://www.atmos-meas-tech.net/11/5701/2018/, doi:10.5194/amt-11-5701-2018.

List of Figure Captions:

Figure 1. (A): Probability density scatter plot of CBH from CALIOP and GETCBH with the CALIOP CBH validation set. The color-bar shows the probability density, and three green dash lines illustrate an error of -2, 0 and 2 km. (B): The CBH histograms binned at 0.5 km from CALIOP (red) and GETCBH (blue). (C): The MBE of the CBH estimate from GETCBH (green solid line), the phase-enhanced GETCBH (orange dashed line) and high confidence samples (black dotted line) of the GETCBH experiment as a function of CALIOP CBH (binned at 0.5 km interval). Only intervals with sample sizes larger than 50 are shown. (D): Similar to (B), but for the distribution of the CloudSat CBH validation set. Figures are plotted using the validation datasets.

Figure 2. Permutation importance for (A) GETCBH and (B) the phase-enhanced GETCBH with the CALIOP CBH validation dataset. Whiskers represent the minimum and maximum of 10 times permutation, while the boxes (may be too narrow to be discerned) and orange line mean 25% and 75% quantiles and medium.

Figure 3. Sample density on reflectance of 1.37 and 2.2 μm bands. The red solid contour lines are for liquid phase samples and blue ones for ice phase. The black dashed contour lines stand for the samples with unknown phase.

Figure 4. Hurricane Dorian and low/high cloud case study on Sep. 5th, 2019. A: ABI Natural True Color imagery at 18:40:12 UTC. The blue and orange lines show the CALIOP scanning tracks with the start time and direction in the right-side legends. Panels B, C and D show the CBH estimated by the GETCBH corresponding to regions 1, 2 and 3 in panel A. The ABI observation start-time of panels B and C is 18:40:12 UTC and that of panel D is 20:00:12 UTC. The clear sky regions are determined by the ABI L2 Clear Sky Mask product and shown in white color. The regions shadowed with black dots show CBH estimates with low confidence. The blue and orange dot lines show the CALIOP track. The red and brown lines indicate the extent of Hurricane Dorian in panel C as determined by the CALIOP CBH products and quality flags. It is worth noting that cloud patterns in panel D are slightly different from those in panel A due to the time difference.

Figure 5. The cross-sections along CALIOP scanning tracks for Hurricane Dorian and low/high cloud case study. B, C and D: The cross-sections refer to the blue and orange dot lines in Figure 4(B), (C) and (D), respectively. The red and brown lines representing

convective region in panel C are also marked in Figure 4(A) and (C). The purple line represents CTH estimates and the grey cross, cyan plus and red Y markers are opaque CBH (the lowest cloud feature detected from the cloud opaque to CALIOP), multi-layer CBH, and single-layer CBH from CALIOP observations, respectively. Note that the opaque cloud can be either single- or multi-layer cloud. The black (lime green) triangle marker is the GETCBH estimated CBH with high confidence (low confidence). The brown bar marker is the valid CBH (invalid VIIRS CBHs are excluded) estimated from VIIRS observations.

Figure 6. The zoom-in of the cross-sections along CALIPSO scanning tracks for the low/high cloud case study. B: High clouds with CBH between 9 and 15 km in Figure 5(B). D: Low clouds with CBH between 0 and 2 km in Figure 5(D).