Characterization of liquid cloud profiles using global collocated active radar and passive polarimetric cloud measurements

Yutong Wang[1,2], Huazhe Shang[1], Chenqian Tang[3], Jian Xu[4], Tianyang Ji[5], Wenwu Wang[1,2], Lesi Wei[1], Yonghui Lei[1], Jiancheng Shi[4], Husi Letu[1]

[1]State Key Laboratory of Remote Sensing and Digital Earth, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100101, China
[2]University of Chinese Academy of Sciences, Beijing 100049, China
[3]Department of Geosciences, University of Oslo, Oslo 0371, Norway
[4]National Space Science Center, Chinese Academy of Sciences, Beijing 100190, China
[5]College of Computer Science, Inner Mongolia University, Hohhot 010021, China

Correspondence to: Huazhe Shang (shanghz@radi.ac.cn), Husi Letu (husiletuw@hotmail.com)

Abstract. Stratiform liquid cloud profiles are key to deciphering cloud life cycles, microphysical processes, and climate change impacts. Nevertheless, remote sensing of cloud vertical structure remains largely unresolved. CloudSat active measurements provide cloud microphysical profile products but are restricted to narrow orbital tracks. Multiangle passive imagers, such as Polarization and Directionality of Earth's Reflectance (POLDER), are capable of generating a variety of cloud properties with broad area coverage; however, they lack key prior knowledge and effective methods for obtaining cloud vertical information. Focusing on single-layer cloud profile retrieval, we first reveal the structural characteristics of stratiform cloud effective radius (CER) profiles based on global CloudSat data and find that the dominant structures include triangle-shaped and monotonically decreasing profiles, which account for approximately 88.5% of global liquid CER profiles. Furthermore, we propose a novel approach to estimate the structural characteristics of triangle-shaped profiles from POLDER observations like the properties of the profile turning point (TP). This approach integrates vertical structure morphology recognition with a combination of fitting methods and machine learning models. The cloud profiles are then accurately reconstructed using physical parameterization models. Our retrieval results exhibit good consistency with active observations, with an RMSE of 1.1μm for TP_CER and 0.1 for the normalized cloud optical thickness at the TP. This research advances the parameterization of liquid cloud profiles and enables profile structural characteristic retrieval based on a multiangle passive imager. Our findings provide valuable insights into improving the understanding and modelling of cloud processes in weather and climate systems.

1 Introduction

Low stratiform liquid clouds, which cover nearly 30% of the Earth's surface (Warren et al., 1986; Warren et al., 1988; Warren et al., 2007; Wood, 2012; Wood, 2015), play a crucial role in the climate system due to their extensive coverage and significant radiative effects, including reflecting shortwave solar radiation and absorbing longwave radiation from the Earth

(Slingo, 1990; Chen et al., 2000; Greenwald et al., 1995). These clouds are a key component of climate and must be accurately represented in general circulation models (GCMs) (Dong and Minnis, 2023; Turner et al., 2007). Compared with ice clouds, which have complex vertical structures and high uncertainty, stratiform liquid clouds are relatively homogeneous in the horizontal direction and have a certain thickness (Zhang et al., 2010; Mace et al., 2009). Therefore, these clouds are relatively ideal for studying cloud properties and growth processes using satellite observations and models.

The cloud microphysical profile is an indispensable parameter for describing cloud vertical structures, such as the profile of cloud effective radius (CER) and liquid water content (LWC) (Chen et al., 2008). Studies have demonstrated that cloud vertical structures are closely linked to the cloud life cycle and precipitation characteristics, atmospheric circulation, cloud–precipitation microphysical processes, and conditions for artificial rainfall (Nakajima et al., 2010a, b; Zhao et al., 2024; Breon and Doutriaux-Boucher, 2005; Sinclair et al., 2021). Accurately and comprehensively detecting and quantifying the vertical structural characteristics and geographical distribution of clouds is highly important for reducing uncertainties in the impact of clouds on climate change and exploring the role of clouds and related feedbacks in complex processes (Rosenfeld and Lensky, 1998; Kessler, 1969; Liu et al., 2006).

Active remote sensing systems, such as ground-based, satellite-borne, and airborne radars, provide precise vertical microphysical cloud data but are constrained by narrow observational swaths (Battaglia et al., 2020; Protat et al., 2009; Fox and Illingworth, 1997). In contrast, passive remote sensing enables large-scale cloud monitoring through multispectral measurements of reflected solar radiation and emitted thermal radiation, offering broad coverage with high spatiotemporal resolution (Nakajima et al., 2010a; Letu et al., 2020; Tana et al., 2023; Shi et al., 2025; Letu et al., 2023; Tang et al., 2025). However, conventional plane-parallel cloud assumptions in passive satellite cloud retrievals contradict natural three-dimensional cloud structures (Platnick, 2001; Horváth and Davies, 2004). Integrating passive and active satellite measurements can effectively combine their complementary advantages, thereby significantly enhancing the ability of passive sensors to retrieve cloud vertical microphysical properties.

Polarimetric multiangle imagers are widely regarded as pivotal instruments for acquiring multidimensional information in global and regional cloud property retrievals (Wang et al., 2022; Bréon and Goloub, 1998; Shang et al., 2019). Unlike conventional passive optical satellite payloads, these advanced sensors synergistically combine multiangle, multipolarization, and multispectral characterization capabilities, thereby maximizing observational information for individual pixels and specific targets (Dubovik et al., 2019). Moreover, the CER can be retrieved more reliably and robustly through polarized multiangle observation data, which also provides effective variance information, than through use of the MODIS imager (Breon and Doutriaux-Boucher, 2005; Bréon and Goloub, 1998; Shang et al., 2019). These factors significantly increase the potential for retrieving cloud vertical properties from passive satellite observations. The POLDER series represents the most mature polarimetric multiangle payloads internationally. Although POLDER-3, the final payload launched in 2004, was

decommissioned in 2013, the upcoming 3MI payload onboard the EUMETSAT Polar System Second Generation (EPS-SG) program will effectively inherit and improve POLDER-3's observational capabilities (Fougnie et al., 2018).

While significant research efforts have focused on cloud vertical structure retrieval (Rosenfeld and Lensky, 1998; Chen et al., 2020; Alexandrov et al., 2020; Barker et al., 2011; Leinonen et al., 2019; Shang et al., 2023), substantial challenges persist in large-scale vertical microphysical characterization, including insufficient prior knowledge from active sensors to guide passive retrieval algorithms and poorly understood correlations between key profile features and other cloud parameters. Additionally, comprehensive statistical analyses examining global-scale cloud profiles from structural and morphological perspectives are lacking. These gaps highlight the need for enhanced integration of active and passive systems with advanced and novel approaches in cloud profile retrieval. To address the above challenges, this paper focuses on quantifying and retrieving the vertical microphysical characteristics of single-layer stratiform liquid clouds. Using observation data from the CloudSat cloud profile radar (CPR), we extract global stratiform liquid cloud profile data over nearly three years. Through a novel perspective of analyzing profile shapes and structural characteristics, we aim to compensate for the lack of a priori knowledge of cloud profile retrieval by passive observation, to understand the significance of profile shape in the cloud life cycle, and to explore the correlation between structural features of the cloud profile and other cloud parameters. Moreover, this study reports the first retrieval of key cloud profile features from POLDER/Parasol satellite observations and the reconstruction of complete cloud profiles.

2 Datasets

2.1 CloudSat data

The main mission of the CloudSat satellite is to detect cloud vertical structures and improve the understanding of cloud abundance, distribution, structure and radiation characteristics. To date, the CloudSat official website has released observations from 2006 to 2020. The instrument carried by CloudSat is the Cloud Profile Radar (CPR), which is a 94-GHz millimeter-wave radar with a sensitivity 1,000 times that of a standard weather radar. The CPR transmits energy to the Earth and calculates the energy returned by the cloud as a function of distance. Global CER profiles of liquid stratiform clouds are derived from the latest version (R05) of CloudSat's 2B-CWC-RO product. This dataset provides vertical measurements of hydrometeor water content, number concentration, and effective radius, featuring 125 layers at 240-m vertical resolution across a 30-km detection range (Austin, 2007). Moreover, we employ the 2B_CLDCLASS_radar product (R05) of CloudSat for cloud layer information, including cloud layer type, cloud layer base height, cloud layer top height, land/sea flag, etc. The precipitation flags are obtained from the 2B-CLDCLASS product (R05) of CloudSat. Users can download all standard CloudSat products from the official CloudSat Data Processing Center at https://www.cloudsat.cira.colostate.edu.

As part of the A-train satellite constellation, CloudSat can also offer synergistic observation capabilities with other passive sensors, such as MODIS and POLDER. Although CloudSat/CPR is sensitive to hydrometers, uncertainties inevitably exist in these products: (1) Near the cloud base, the presence of drizzle or raindrops—due to predefined threshold settings—may hinder the ability to reliably distinguish them from cloud droplets, (2) contamination by drizzle and raindrops can cause deviations in the observed cloud droplet size distribution from the theoretical distribution, thereby introducing errors in the retrieval process (Austin, 2007), and (3) due to the limitations of 240-m resolution, it may not be possible to identify ultra-thin layer structures below 240m.

2.2 Parasol data

The POLDER-3/Parasol payload is a multiangle, multipolarized, and multispectral instrument designed for atmospheric aerosol, cloud, water vapor, and radiation budget studies. Operating from September 2004 to October 2013, POLDER-3 features three polarized (490, 670, and 865 nm) and six nonpolarized (443, 565, 763, 765, 910, and 1020 nm) observation channels, providing atmospheric data from up to 16 angles with a nadir resolution of 5.3×6.2 km (Deschamps et al., 1994). Parasol joined the A-Train constellation in 2005, but regrettably, it drifted away from the formation in 2009. Nevertheless, the combined POLDER and CloudSat data still hold significant research value. This study employs their observations for the remote sensing of profiles, offering insights for future sensor combinations such as 3MI and EarthCARE.

Our study uses POLDER-3 Level-1 (L1_B) products and cloud optical thickness parameters from Level-2 (RB2) products as input data for estimating the structural characteristics of cloud profiles. Moreover, we employ the CER retrieved by an improved primary cloudbow retrieval (PCR) algorithm (Shang et al., 2019) and the cloud-base height and cloud-top height retrieved based on POLDER data. The PCR algorithm permits an extended range of CER (3–25μm) and EV (0.01–0.29) estimates and a higher resolution (40–60 km) in the retrieval by using POLDER polarized measurements from both primary and supernumerary cloudbow regions. The retrieval algorithm for cloud base height will be thoroughly described in a forthcoming article and is therefore not discussed here

3 Methodology

The study workflow (Fig. 1) is divided into three main parts: first, classifying the shapes of CER profiles from CloudSat data and examining their structural features; second, conducting correlation analyses between the structural features of these profiles and other relevant cloud parameters; and third, parameterizing cloud profiles, retrieving their key characteristics and reconstructing the complete profile. The methods employed in this study are introduced below.

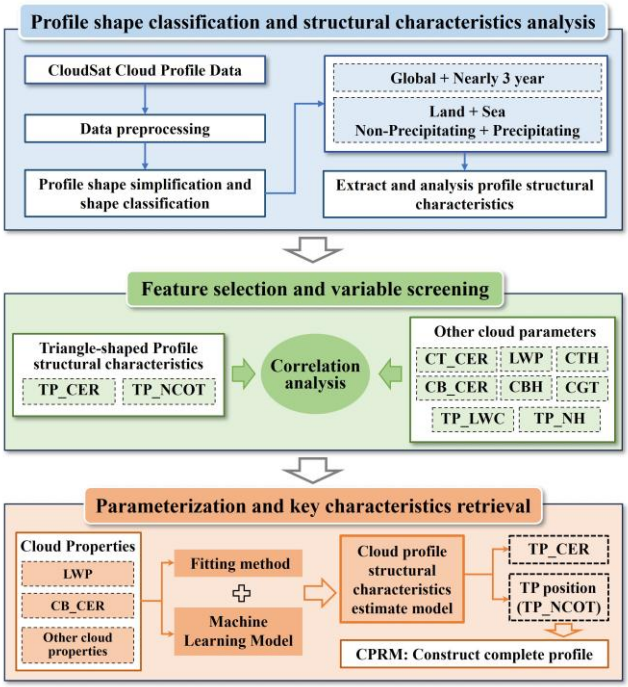

Figure 1: Research Framework Flowchart (TP_CER represents the cloud effective radius at the profile turning point; TP_NCOT represents the normalized cloud optical thickness at the profile turning point; CT_CER represents the cloud effective radius at the cloud top; CB_CER represents the cloud effective radius at the cloud base; TP_LWC represents the liquid water content at the profile turning point; TP_NH represents the normalized height at the profile turning point; LWC represents the liquid water content; CTH represents the cloud-top height; CBH represents the cloud-base height; CGT represents the cloud geometric thickness.)

3.1 Cloud profile data preprocessing and shape simplification

To ensure the accuracy and consistency of the research data and to provide a reliable basis for subsequent analysis, we preprocessed the CloudSat profile data. Preprocessing involves mainly data screening and matching, as well as standardizing the data to better reveal the structural characteristics of the profiles. Due to the complexity of multilayered clouds and the dominance of single-layer cases, we limit our investigation to single-layer cloud profiles. Based on global CloudSat data product over nearly three years (2013, 2019, and the first eight months of 2020), we extracted approximately 12.47 million

CER profiles of single-layer stratocumulus and stratus clouds, with each profile linked to multiple auxiliary cloud parameters. The auxiliary information includes profile observation time, geographic coordinates, cloud-top height, cloud-base height, geometric thickness, liquid water path, corresponding liquid water content (LWC) profile for each CER profile, cloud type, land/sea flag, and precipitation flag. Moreover, the normalized optical thickness and normalized height at each layer of the profile are calculated for subsequent comparative analysis. Notably, the normalized optical thickness is 0 at the cloud top and 1 at the cloud base, whereas the normalized height is 1 at the cloud top and 0 at the cloud base. The selection of 2013, 2019, and 2020 is not arbitrary but strategically chosen to align with existing polarized multi-angle payload observations. Furthermore, since our study focuses on relatively homogeneous and stable single-layer stratiform liquid clouds, localized atmospheric anomalies do not impact the statistical results presented herein.

Here, we propose a shape simplification scheme primarily based on the Visvalingam-Whyatt line simplification (VM) algorithm (Visvalingam, 2016) to extract the essential shape characteristics of profiles. This approach effectively simplifies complex profile geometries while preserving their fundamental structure, thereby eliminating interference with the extraction of key profile features. The VM algorithm, which was originally designed for vector data processing, preserves geometric characteristics while reducing the number of data points. The simplification steps include (1) setting a distance threshold (decimation factor) as the simplification criterion; (2) calculating the area of triangles formed by consecutive points and identifying the smallest area (Amin); and (3) deleting the middle vertex if Amin < A and repeating the process until all areas exceed the threshold. This approach minimizes angular changes and maintains the profile's true form, enabling faster and more accurate identification of turning points in subsequent analyses.

3.2 Feature selection and variable screening

Feature selection and variable screening were conducted for the subsequent retrieval of key profile characteristics through correlation analysis. The turning-point cloud effective radius (TP_CER), normalized optical thickness (TP_NCOT), and related cloud parameters are extracted from CloudSat-derived CER profiles exhibiting an "increasing-then-decreasing" shape. The related cloud parameters include cloud-top CER (CT_CER), cloud-base CER (CB_CER), the liquid water path (LWP), cloud-top height (CTH), cloud-base height (CBH), cloud geometric thickness (CGT), normalized height at the turning point (TP_NH), and liquid water content at the turning point (TP_LWC). The correlation coefficients between these cloud parameters and the key CER profile structural characteristics (TP_CER and TP_NCOT) are calculated. The parameter combinations that exhibit a high correlation with the key structural characteristics of CER profiles are selected as input variables for the subsequent model.

3.3 Retrieval of profile key characteristics and profile reconstruction

For the retrieval of stratiform liquid cloud profiles, the estimation of key profile characteristics is performed first, followed by inputting these key characteristics into a physical parameterization model to reconstruct the complete cloud profile. The

estimation methods for the key profile structural features (TP_CER and TP_NCOT) are carefully selected based on the results of the correlation analysis. When the correlation analysis shows that certain parameters have strong correlations with the key profile structural features, a multiple regression model is employed to estimate these features. Conversely, when all the parameters exhibit weak correlations with the key profile structural features, we compare the performance of the machine learning model with that of multiple regression and select the approach that demonstrates superior accuracy. Additionally, different combinations of input parameters are compared to achieve the highest estimation accuracy and computational efficiency.

The optimal multiple linear regression model is selected based on the number of highly correlated variables in the correlation analysis: the selected cloud parameters with higher correlations serve as independent variables ($X_1, X_2, \ldots, X_k$), whereas TP_CER is treated as the dependent variable Y, establishing a multiple linear regression model. The model takes the following form:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k \tag{1}$$

where β_0 is the constant term and $\beta_1, \beta_2, \ldots, \beta_k$ denote the regression coefficients. Approximately 50% of the CloudSat dataset is used as the training set to estimate the regression coefficients through least squares methods and obtain optimal model parameters, and the remaining 50% is used for validation.

Given its demonstrated high stability in cloud property retrievals and previous testing, the random forest is chosen as the primary machine learning model for estimating key features. Compared with multiple decision trees, the random forest model trains each tree through random sampling and feature selection. The ratio of the training set to the test set remains approximately 1:1, which is consistent with the regression models. Ultimately, the most accurate method and optimal input parameter combination are applied to POLDER data to retrieve the key structural characteristics of CER profiles. The accuracy metrics, including the root mean square error (RMSE), Pearson correlation coefficient (R), coefficient of determination (R^2) and relative root mean square error (rRMSE), are used to evaluate the accuracy of the parameter estimation.

To extend profile feature estimation to the scope of passive observations, we constructed a cost function F based on L2 product parameters from POLDER-3 observations covering all pixels within the observational range, as well as passive observation L2 product parameters of triangle-shaped profiles in CloudSat observations. The L2 product parameters include COT, CGT, the temperature profile (10 layers), and the water vapor profile (9 layers). The cost function is defined as follows:

$$F(i,j,m) = \sum_{v=1}^{V} \left[\frac{P_v(i,j) - P_v(m)}{P_v(m)} \right]^2 : m \in [0, M] \tag{2}$$

where i and j are the row and column numbers of the pixel; m is the index of the active-passive synergistic pixel with an Inc_Dec profile; and P_v represents the value of the L2 product parameters. In this study, the threshold for the cost function is set to 0.5. If the cost function of a pixel is less than 0.5, the atmospheric environment forming the cloud and the profile shape of the cloud within that pixel are similar to the triangle-shaped profiles along the CloudSat observation trajectory.

The estimated profile structural characteristics are input into the cloud profile reconstruction model (CPRM) to reconstruct complete profiles. The CPM, proposed by Shang et al. (Shang et al., 2023), is a parameterization model derived from extensive cloud profile data simulated by the Colorado State University (CSU) Regional Atmospheric Modeling System (RAMS). The CPM assumes that the particle radius profile and liquid water content profile are either linear or triangular. It identifies eight main parameters necessary to fully describe a cloud profile: cloud geometric thickness (z_c), cloud optical thickness (τ), turning-point normalized optical thickness (t_m) measured from the cloud top, effective radius at the cloud base (r_b), effective radius at the cloud top (r_t), effective radius at the turning point (r_m), effective variance of the gamma particle distribution (v_e), and slope (k) of the cloud droplet number concentration profile (N). The CPRM is an extension of the CPM, designed to reconstruct detailed cloud profiles using limited measurement data from active detection (Shang et al., 2025), which can be iteratively computed based on the retrieved key profile structural features (TP_CER and TP_NCOT) until it converges to a complete profile structure with continuous layer information.

3.4 Match-up between POLDER and CloudSat data

To validate the profile structural characteristics retrieved by passive satellite observations, a match-up process between POLDER and CloudSat observations is conducted. We focus on March 2007 and identified coincident orbits that contained a high number of stratiform cloud profiles exhibiting a triangle-shaped vertical structure in CloudSat data. A specific dataset from March 2, 2007 (POLDER observation time between 06:41:09 and 07:24:06 UTC) is selected for detailed analysis in Section 4.4. The POLDER Level 2 (RB2) product served as the primary dataset for matching with CloudSat observations. With a spatial resolution of approximately 16 km, this product is notably coarser than CloudSat's resolution of less than 2 km. To establish correspondence between the datasets, the Euclidean distance between each POLDER RB2 pixel center and all CloudSat data points within the POLDER RB2 pixel is computed. Owing to the resolution discrepancy, a single POLDER RB2 pixel often contains multiple CloudSat data points. In such cases, only the CloudSat data point closest to the center of the POLDER_RB2 pixel is retained.

Through the matching process, cloud optical thickness (COT), latitude, longitude, and other relevant data are extracted from the POLDER RB2 product. These coordinates are then used to extract cloud base height (CBH), cloud top height (CTH), and cloud-top effective radius (CT_CER) obtained through the retrieval algorithm. CBH and CTH are retrieved from the POLDER L1 product, which has a native resolution of 6 km, matching the resolution of the source data. CT_CER is retrieved from the POLDER L1 product at a 50 km resolution.

3.5 Cloud base height retrieval algorithm

带格式的: 标题 2

The cloud base height (CBH) retrieval algorithm(Ji et al., 2025) employed in this study is based on a deep neural network trained on Parasol L1 measurements collocated active sensor observations. The training dataset comprises Parasol L1 data—including intensity from 14 viewing angles in the oxygen A-band (763 nm and 765 nm channels), longitude, latitude, elevation, and cloud indicator—along with the corresponding CBH values and cloud detection information obtained from the CloudSat-CALIPSO L2 product 2B-CLDCLASS-LIDAR. Data from March, June, September, and December 2007 are primarily used, with the last seven days of each month reserved for testing and the remaining data used for training. To ensure high-quality training data, only cases where Parasol confidently detected cloudy scenes and CloudSat identified single-layer clouds are retained. Spatial collocation accuracy is constrained to within 0.01°, while temporal discrepancies are negligible due to the near-simultaneous observations from the A-Train satellites. The model utilizes geographic coordinates (longitude, latitude, elevation) and multi-angle oxygen A-band information from Parasol as inputs to predict CBH, with CloudSat-derived heights serving as ground truth. This method enables the CBH retrieval using only passive observations as input. The validation results indicate that the retrieval achieves a mean absolute error (MAE) of 0.78 km, a bias of 0.22 km, and a correlation coefficient (R) of 0.82.

It is important to note that the development, comprehensive validation, and detailed methodological discussion of this algorithm are beyond the scope of this study. A full description of the algorithm has been submitted to a separate journal and is currently under review. To ensure transparency and reproducibility, we provide a complete documentation of the algorithm in the Supplementary Material of this article. Furthermore, the code and pre-trained model have been made publicly available (see the Data and Code Availability section).

4 Results

4.1 Typical shape and structural characteristic analysis of CER profiles

We conduct categorical statistics based on three criteria, namely, cloud type (stratocumulus/stratus), underlying surface type (sea/land), and precipitation occurrence (nonprecipitating/precipitating), to investigate the characteristic effective particle profiles across different cloud categories. Fig. 2 presents the results of the quantitative analysis of the spatial distribution and categorical proportions of the profile data within our dataset. Spatially, single-layer stratiform liquid cloud profiles occur more frequently over mid- to low-latitude oceans in the Southern Hemisphere and high-latitude oceans in the Northern Hemisphere, whereas generally lower occurrence rates occur over land areas. The striated features apparent in Fig. 2(a) result from CloudSat's narrow-swath observations. Fig. 2(b) more visually illustrates the significant land-ocean disparity: for stratiform cloud profiles, the surface type ratio is 88.3% ocean to 11.7% land, whereas for stratocumulus profiles, the ratio is 84.4% ocean to 15.6% land. For both cloud types, the nonprecipitating-to-precipitating ratio is approximately 2:3, and

precipitating clouds are slightly more frequent. It should be noted that the CloudSat CWC-RO product is known to miss a portion of single-layer liquid clouds, either due to masking by surface clutter or because their signal falls below the radar's detection threshold (Lamer et al., 2020; Schulte et al., 2023). As a result, the true ratio of nonprecipitating to precipitating clouds is likely higher than reported in this study. The vertical distribution patterns are presented in Fig. 2(c). Notably, 99.46% of sea and 99.62% of land liquid stratiform cloud profiles are concentrated within layers 3-11. Based on the above analysis, given CloudSat's vertical resolution of 240 m per bin and the fact that approximately 99.5% of the profiles are concentrated within 11 bins, this study focuses on single-layer stratiform liquid clouds with a geometric thickness of less than 2.64 km.

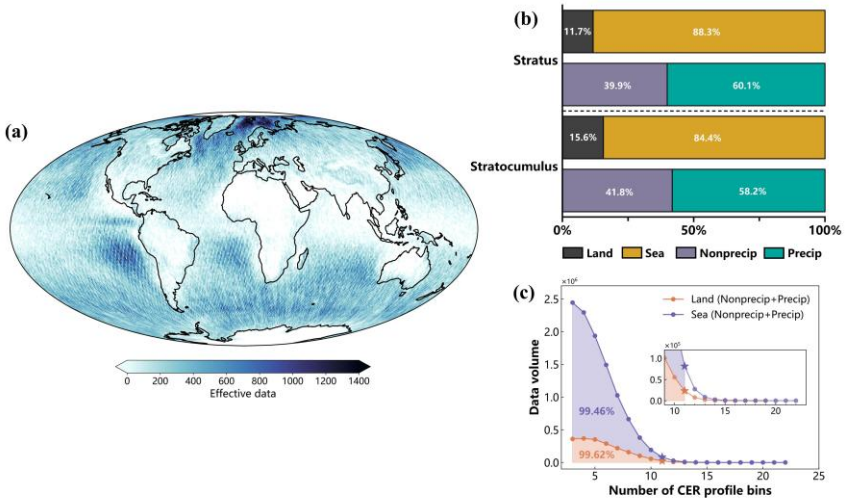

Figure 2: Distribution and composition of effective data on the investigated liquid stratiform cloud profiles. (a) Geographical distribution of effective data; (b) statistical chart of land/sea and precipitation/non-precipitation composition of effective data; (c) distribution and number of profile valid data bins.

Through an extensive literature review and visual analysis of CloudSat single-layer liquid cloud profiles, the vertical variation of cloud effective radius (CER) can be classified into four distinct shapes based on the monotonicity between adjacent layers: (1) triangle shaped (Inc_Dec), increasing then decreasing; (2) monotonically decreasing (Mono_Dec); (3) monotonically increasing (Mono_Inc); and (4) decreasing then increasing (Dec_Inc). These shapes depict the vertical variation in the CER from the cloud base to the top. The four shapes can be simply expressed by the following formulas:

$$\text{Inc_Dec: } CER_1 < CER_2 < \cdots < CER_k > CER_{k+1} > \cdots > CER_N, 1 < k < N \quad (3)$$

带格式的: 段落间距段前: 0.5 行

$$\text{Mono_Dec: } CER_1 > CER_2 > \cdots > CER_N \quad (4)$$

$$\text{Mono_Inc: } CER_1 < CER_2 < \cdots < CER_N \tag{5}$$

$$\text{Dec_Inc: } CER_1 > CER_2 > \cdots > CER_k < CER_{k+1} < \cdots < CER_N, 1 < k < N \tag{6}$$

带格式的: 段落间距段后: 0.5 行

Where CER_i denote the CER at the *i*-th vertical level (bin), *i*=1 corresponds to the cloud base and *i*=N to the cloud top. Systematic classification and statistical analysis confirm these patterns (Fig. 3(a)). Collectively, these four shapes account for 90.1% of the observed CER profiles, with Shapes 2 (Mono_Dec: 48.8%) and 1 (Inc_Dec: 39.7%) being the most prevalent, highlighting their dominance in the liquid stratiform cloud life cycle. The remaining 9.1% represent complex-shaped profiles that do not conform to these four categories, with further analysis of these cases presented in the Appendix A (Table A4). Approximately 60% of these profiles contain only single segments inconsistent with Shapes 1 and 2. Although our profile simplification program can reduce complex shapes to simpler forms, there is controversy regarding the specific categories to which these shapes belong. Taking Fig. A1 as an example, Complex Shape 1 and Complex Shape 2 can be simplified into different primary shapes. Therefore, we believe that these 9.1% of complex profile shapes can be further analyzed in subsequent studies, but it is unnecessary to include them in the follow-up parts presented in this study, as their inclusion would introduce unnecessary errors into our retrieval prior knowledge.

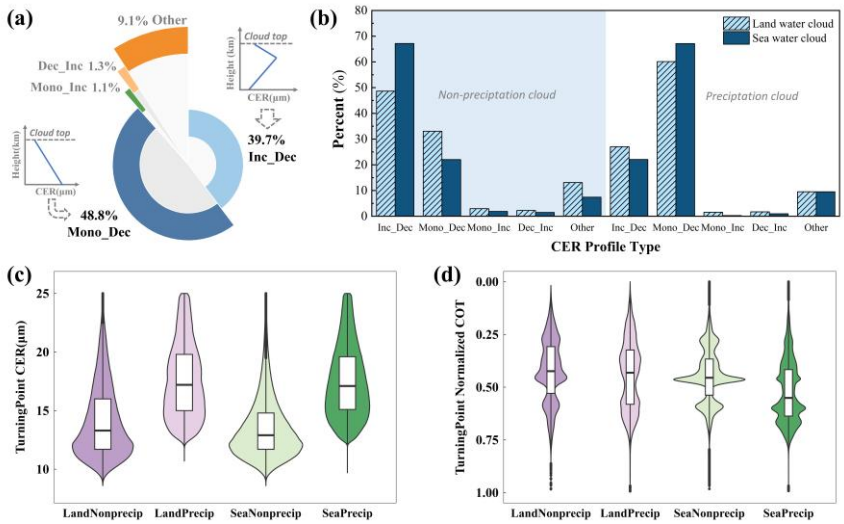

Figure 3: Categorical statistics of stratiform cloud profile shape and the distribution of CER and normalized optical thickness at the TP of Inc_Dec (Shape 1) profiles. (a) Shape of all profile data; (b) shape of land/sea profile data; (c) TP_CER of stratiform cloud profiles; (d) TP_NCOT of stratiform cloud profiles.

These shapes correspond to distinct development stages of the cloud life cycle (Fig. 4). Initial updrafts drive adiabatic growth, leading to a reduction in droplet size with increasing optical thickness. Mature clouds exhibit enhanced evaporation at the cloud top due to dry air entrainment, further reducing droplet size (Shape 1, Fig. 4(a)). In contrast, collision-coalescence near the cloud base promotes droplet growth and spectral broadening (Shape 2, Fig. 4(d)). The distinction between Fig. 4(c) and (d) lies in the presence of precipitation below the cloud base, which results in larger base particle sizes. If precipitation-induced droplet accumulation in the lower cloud layer reduces droplet size (based on Fig. 4(d)), the resulting pattern resembles that in Fig. 4(b).

Since the Inc_Dec shape is more complex than the Mono_Dec shape is and more structural characteristics require investigation (e.g., turning points of profiles), this study focuses specifically on analyzing the structural characteristics of Inc_Dec profiles. The analysis establishes a foundation for subsequent feature parameterization and estimation. Unlike other shapes, the key parameters essential for describing Inc_Dec profiles are the turning-point CER (TP_CER) and the turning-point normalized optical thickness (TP_NCOT). Fig. 3(c) shows the TP_CER distribution for Inc_Dec profiles. Nonprecipitating stratiform clouds exhibit a single-peak TP_CER distribution between 8 and 25μm, peaking near 12μm. Precipitating stratiform clouds exhibit a single-peak TP_CER distribution between 10 and 25μm, peaking near 17μm. The analysis reveals several key results: (1) precipitating cloud profiles present larger (by 3–5μm) TP_CERs than nonprecipitating clouds do, and (2) oceanic stratiform cloud profiles present a narrower TP_CER range than their land counterparts do.

The distributions of TP_NCOT for stratocumulus and stratus clouds are shown in Fig. 3(d). TP_NCOT indicates the position of the TP within the cloud profile. The key observations include the following: (1) the TP_NCOT of liquid stratiform clouds exhibits a multipeak distribution; (2) nonprecipitating cloud TPs are concentrated in the upper optical thickness region (NCOT: 0–0.5), while land-precipitating cloud TPs show a symmetric distribution around NCOT=0.5, and oceanic-precipitating cloud TPs are distributed predominantly between 0.5 and 1. This suggests that precipitating cloud TPs occur closer to the cloud base than nonprecipitating clouds do, which is consistent with the cloud life cycle illustrated in Fig. 3. (3) The peak of the oceanic precipitating cloud distribution clusters around NCOT=0.5.

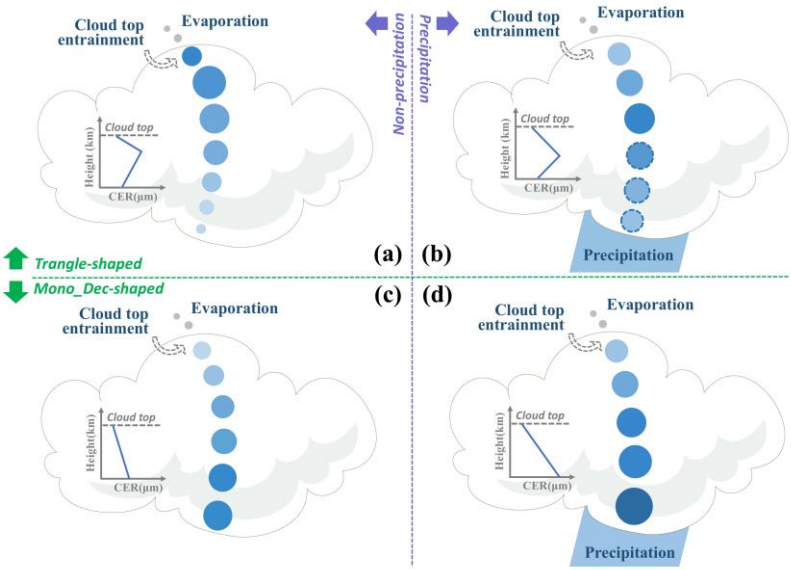

Figure 4: Schematic diagram of the cloud life cycle stages corresponding to different profile shapes. (a) and (b) represent cloud droplet development corresponding to Inc_Dec (triangle-shaped) profiles; (c) and (d) represent cloud droplet development corresponding to Mono_Dec profiles.

4.2 Correlation analysis of profile structural features

During the profile retrieval process, to determine the main structure of the cloud profile, it is necessary to obtain the relevant information at the profile TP, as well as at the bottom of the cloud. Many previous studies have focused on the retrieval of cloud bottom parameters; however, few studies have explored cloud profile TP information, so we would like to establish a parameterized scheme and estimation method for cloud profile TP-related information. To further investigate the correlation between the structural features of the profile and other cloud parameters, we initially examined the relationships of both TP_CER and TP_NCOT with other cloud parameters for Inc_Dec profiles. The study randomly selected 4800 data points for each of the eight stratiform clouds to be analyzed, and the correlations between the TP_CERs of the eight stratiform clouds and the nine cloud parameters are presented in Fig. 5(a). For different types of clouds, TP_CER has a strong correlation with cloud-base CER and the liquid water path (LWP), which generally range from 0.75 to 0.92, among which the cloud-base CER has the highest correlation with the TP_CER of the stratiform cloud profiles. Moreover, TP_CER also has a strong correlation with the liquid water content at the TP (TP_LWC), and the correlation between stratiform cloud TP_CER and

TP_LWC fluctuates over a wider range, between 0.462 and 0.784. The cloud-base height, TP_NH, and TP_NCOT have weak correlations with TP_CER, with correlations in the range of -0.26 to 0.13. The cloud-top CER and cloud-top height have different degrees of correlation for the TP_CERs of different types of clouds: the correlation of the cloud-top CER is greater for the TP_CER of nonprecipitating stratiform clouds (0.42–0.51) than that of precipitating stratiform clouds (0.23–0.27), and the correlation of cloud-top height with the TP_CER of the sea stratocumulus and stratus profiles is greater (0.41–0.49) than that with the land stratiform clouds (0.18–0.29).

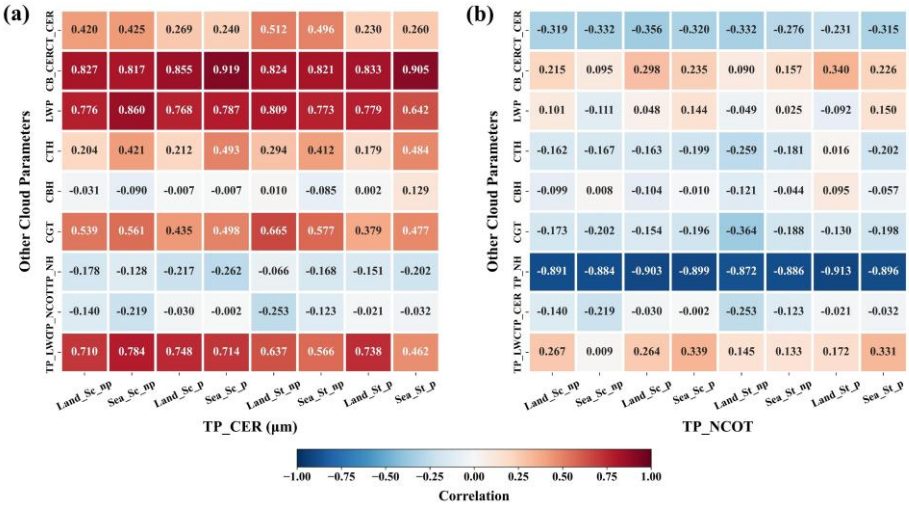

Figure 5: Correlation of TP parameters with other cloud parameters. (a) correlation of turning point CER with other cloud parameters; (b) correlation of the normalized optical thickness at turning point with other cloud parameters. The nine cloud parameters corresponding to the vertical axis are, from top to bottom: cloud top CER (CT_CER), cloud bottom CER (CB_CER), liquid water path (LWP), cloud top height (CTH), cloud bottom height (CBH), cloud geometric thickness (CGT), normalized height at the turning point (TP_NH), normalized optical thickness at the turning point (TP_NCOT), and liquid water content at the turning point (TP_LWC). The nine types of clouds corresponding to the horizontal axis are, from left to right: land non-precipitation stratocumulus (Land_Sc_np), sea non-precipitation stratocumulus (Sea_Sc_np), land precipitation stratocumulus (Land_Sc_p), sea precipitation stratocumulus (Sea_Sc_p), land non-precipitation stratus (Land_St_np), sea non-precipitation stratus (Sea_St_np), land precipitation stratus (Land_St_p), sea precipitation stratus (Sea_St_p).

Fig. 5(b) shows the correlations between the TP_NCOT of stratiform cloud profiles and other cloud parameters. In marked contrast to TP_CER, the correlations of almost all cloud parameters with TP_NCOT, except for TP_NH, are relatively weak, and almost all of them are in the range of ±0.3. Since both TP_NH and TP_NCOT indicate where the TP occurs in the cloud,

it is understandable why they are highly correlated. The weak correlation for TP_NCOT stems from the fact that the TP position is largely independent of common cloud parameters such as droplet size, cloud water content, and cloud thickness. Instead, it is primarily influenced by microphysical processes like cloud-top entrainment and precipitation formation, leading to a relatively random distribution of TP_NCOT within the cloud layer. This inherent randomness makes it inherently difficult to estimate TP_NCOT using conventional correlation-based method.

To further investigate several parameters that are highly correlated with TP_CER and their relative distributions with respect to TP_CER, we randomly selected 4800 samples with four cloud characteristics (Land Nonprecip, Sea Nonprecip, Land Precip, and Sea Precip) to generate scatter density plots. The relative distributions of parameters highly correlated with the TP_CER of stratiform cloud profiles are shown in Fig. 6(a)-6(p). As illustrated in Fig. 6(a)-(d), the scatter points are located mostly below and close to the 1:1 line, indicating a strong linear correlation between the CB_CER and TP_CER of the profile. For sea cloud profiles, TP_CER also shows a strong linear correlation with the LWP. In contrast, Fig. 6(e) and (g) show that, for land clouds, while TP_CER still strongly correlates with the LWP, the scatter density plot exhibits a vertical distribution along TP_CER, with points diverging from the density center toward both sides. Sea cloud scatter points, however, show a radial distribution pattern, dispersing symmetrically from the center. In Fig. 6(i)–(l), the scatter points form nearly horizontal stripes, with most points concentrated around several CGT values. This is due to CloudSat's measurement method, which results in a discontinuous discrete distribution of CGT, thereby lowering the linear correlation between CGT and TP_CER. TP_LWC shows a high linear correlation with TP_CER, with multiple density centers visible in the scatter density plots. As shown in Fig. 6(m), (n), and (p), nearly all plots exhibit two density centers. The two density centers observed in the relationship between the TP_CER and TP_LWC reflect two distinct cloud microphysical regimes. One is primarily driven by condensational growth, which tends to occur under low aerosol and stable conditions, resulting in higher LWC for a given droplet size. The other is dominated by collision-coalescence, typical in relative high aerosol and dynamically active environments, leading to lower LWC for the same droplet size.

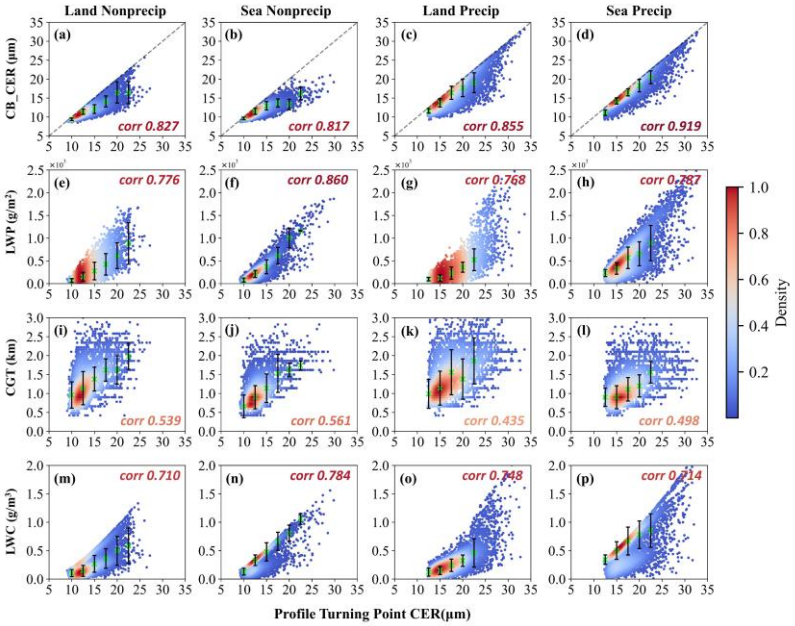

Figure 6: Relative distribution of cloud parameters (high correlation with TP_CER) and profile TP_CER. The green × represents binned median of the parameter represented by the vertical axis; the black line segment represents binned median of the parameter represented by the vertical axis ±σ.

4.3 Estimation of key structural features of CER profiles

Based on the previous analysis, CB_CER and the LWP clearly have good linear correlations with TP_CER. Therefore, we employ multiple linear regression to estimate TP_CER for cloud profiles with four different characteristics using CB_CER and the LWP as parameters. The goal is to derive empirical fitting formulas for TP_CER based on these two parameters. Three combinations of dependent variables were selected: (1) CB_CER and the LWP, (2) CB_CER alone, and (3) the LWP alone. For all four cloud types, using the combination of CB_CER and the LWP for multiple linear regression clearly yields the highest accuracy for estimating TP_CER. The validation results of TP_CER estimation, shown in Fig. 7, indicate that the RMSEs are 1.19 for Sea Nonprecip, 1.30 for Sea Precip, 1.75 for Land Nonprecip, and 1.96 for Land Precip. The estimation accuracy for sea cloud profiles is greater than that for land cloud profiles. Additionally, for sea cloud profiles, the use of both parameters for estimation significantly improves accuracy compared with the use of only one parameter. However, for land cloud profiles, the accuracy of TP_CER estimation using only CB_CER is comparable to the accuracy when both CB_CER

and the LWP are used, suggesting that CB_CER alone is sufficient for estimating TP_CER for land cloud profiles. The final empirical fitting coefficient, predictive performance and more specific verification results are shown in the Appendix A.

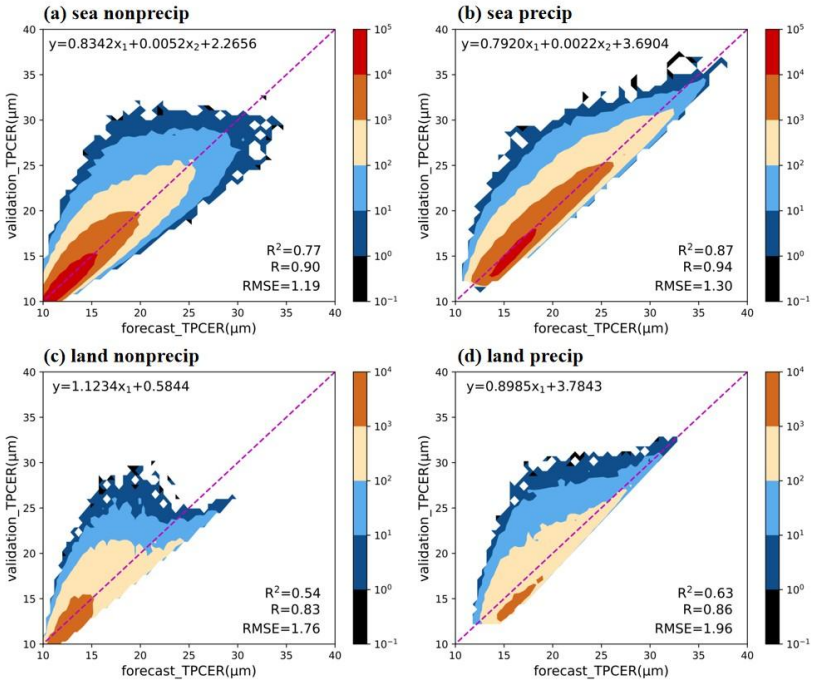

Figure 7: Accuracy verification contour plots of the optimal TP_CER estimation method for four different types of clouds (sea nonprecipitating, sea precipitating, land nonprecipitating, and land precipitating).

In addition to TP_CER, we use random forest and multiple linear regression methods, considering various parameter combinations to estimate TP_NCOT, according to the relatively weak correlation between other cloud parameters and TP_NCOT. The results indicate that the random forest method performs well for estimating TP_NCOT. The highest accuracy for estimating TP_NCOT is achieved by combining four parameters—CB_CER, CT_CER, CGT, and the LWP—with RMSEs ranging from 0.1 to 0.12 and an R value of approximately 0.6, as shown in Fig. 8. Except for CTH, which can somewhat replace CGT for estimating TP_NCOT, other parameter combinations significantly reduce estimation accuracy.

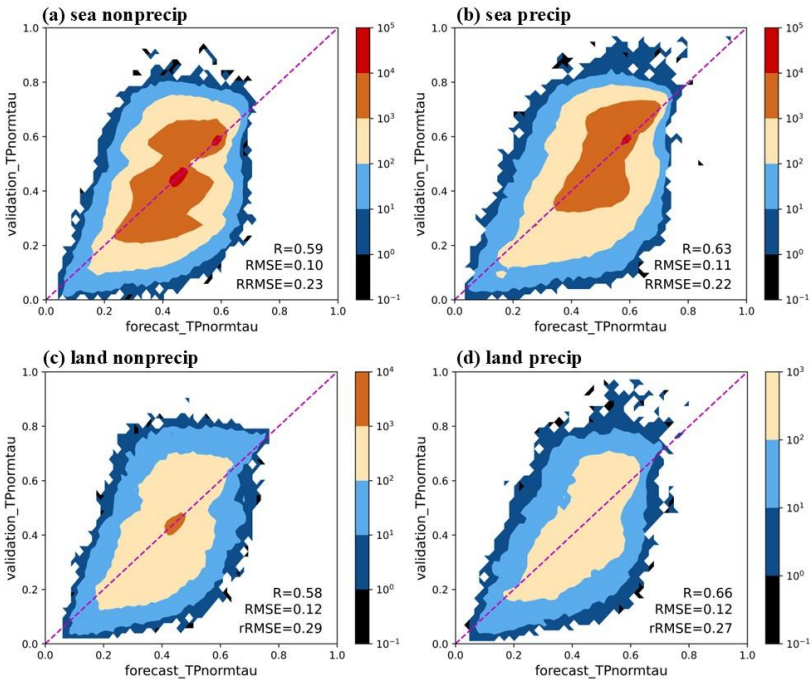

Figure 8: Accuracy verification contour plots of the optimal TP_NCOT estimation method for four different types of clouds (sea nonprecipitating, sea precipitating, land nonprecipitating, and land precipitating).

This study utilizes POLDER-3 Levels 1 and 2 (RB2) data, with observations commencing on March 2, 2007, at 06:41:09, in conjunction with CloudSat products, to perform active-passive satellite data matching. The primary objective is to investigate the applicability of the proposed profile structure characterization method for retrieving cloud vertical structures from passive satellite data. This study focuses on a typical stratocumulus cloud region over the Indian Ocean, located west of Oceania, within the geographical range of 40°S–65°S and 100°E–125°E, as shown in Fig. 9(a). The cloud-base CER is difficult to obtain directly through satellite observations. Here, we first analyzed the statistical distribution of CloudSat-observed CB_CER and found its probability density function to be highly regular. Based on this observation, we developed a multivariate regression model using known parameters (CTH, the LWP, CT_CER, CBH) to estimate CB_CER. The method achieved excellent results, with the highest retrieval accuracy (for sea nonprecipitating clouds) and an RMSE of 1.13μm. Based on the aforementioned estimation method, we estimated the key structural features of the cloud field profile in Fig.

9(a)—TP_CER and TP_NCOT—with the results shown in Fig. 9(f). Here, we estimate only the TP information for profiles classified as Inc_Dec.

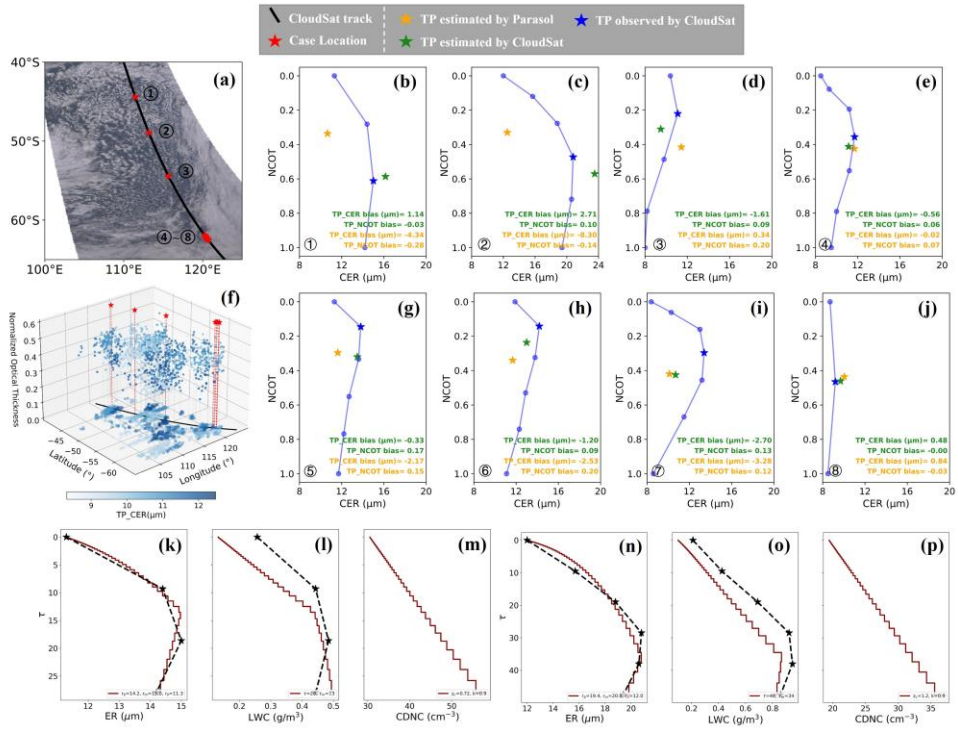

Figure 9: Comparative results of estimating the TP using passive data (Parasol) and active data (CloudSat). (a) True-color image drawn by Parasol observation on March 2, 2007; (b)–(e) and (g)–(j) represent the 8 profile cases indicated in (a) and present a comparison between the TP estimated by Parasol data and those estimated by CloudSat data, as well as the profiles and TP observed by CloudSat; (f) shows the TP_CER and TP_NCOT results estimated by POLDER observation. The last row of panels in the figure represents the profiles reconstructed by CPRM; (k)–(m) and (n)–(p) correspond to cases 1 and 2, respectively.

The estimation accuracy is validated by comparing the estimated TP parameters derived from the POLDER-3/Parasol and CloudSat input parameters against the actual CloudSat measurements, as illustrated in Fig. 9(b)-(e) and (g)-(j). Eight representative profiles with TPs are selected based on CloudSat data. The comparative analysis reveals that the estimation accuracy when CloudSat parameters are used slightly surpasses that when POLDER-3 data are used. In particular, cases 3, 4,

5, 6, and 8 demonstrate satisfactory estimation performance, whereas cases 1, 2, and 7 exhibit relatively lower accuracy. This discrepancy can be attributed to three primary factors: (1) The coarse resolution of POLDER products cannot capture inherent subpixel heterogeneity. This study utilizes the POLDER Level 2 RB2 product (16 km resolution) along with cloud top height (CTH), cloud base height (CBH) (both at 6 km), and cloud top effective radius (CER) data (50 km resolution), all of which are derived from POLDER Level 1 products. Compared to CloudSat data, the coarser resolution of POLDER may cause biases resulting from subpixel heterogeneity. We conduct a further analysis of the eight cases in Fig. A2 by averaging the CloudSat CER profiles within each corresponding POLDER pixel along the altitude dimension. This process effectively aggregates the high-resolution CloudSat profiles to the spatial scale of a POLDER pixel, simulating what POLDER would likely "see". The resulting averaged profiles are then compared against our validation data—the CloudSat profile closest to the center of the POLDER pixel. Although this study specifically targets horizontally relatively homogeneous single-layer stratiform water clouds, subpixel heterogeneity—resulting from POLDER's coarse resolution—remains one of the main sources of error in estimating the structural parameters of cloud profiles. (2) Inherent physical limitation: Vertically integrated signal. The retrieval of CER from POLDER observations differs from traditional dual-channel methods, relying instead on the directional characteristics of polarized reflectance within the cloud bow scattering angle range. It should be emphasized that all such passive retrieval techniques essentially provide a vertically integrated measurement—a weighted average signal sensitive to microphysical properties from the cloud top downward through a depth determined by cloud optical thickness. This fundamental characteristic inherently increases the uncertainty in retrieving vertical structural features. (3) Propagation of input uncertainties. Errors in the upstream inputs (COT, CT_CER from POLDER, and CTH/CBH from the combination of POLDER and ancillary data) inevitably propagate into the LWP and CB_CER. This accumulated uncertainty then propagates into errors in the final estimated profile parameters, including the TP_CER and its location. Cases with higher sub-pixel heterogeneity or where the cloud-top CER is less representative of the layer-average are particularly susceptible to this propagation effect.

5 Discussion and conclusion

The primary goal of this study was to analyze the structural features and shapes of single-layer stratiform liquid cloud CER profiles using global CloudSat data, with a focus on understanding how these profiles represent different stages of the cloud life cycle. We also aimed to retrieve key profile characteristics from multiangle passive imager observations and reconstruct complete cloud profiles using physical parameterization models.

Profile analysis of global single-layer stratiform liquid clouds reveals two dominant profile shapes: Inc_Dec (triangular shape, 39.7%) and the Mono_Dec (48.8%), which represent nearly 90% of cases. These shapes occur in both precipitating and nonprecipitating clouds, reflecting different lifecycle stages. For Inc_Dec profiles, the turning-point (TP) CER (TP_CER) and its position are structurally significant, showing strong correlations with cloud-base CER (CB_CER), the liquid water

path (LWP), cloud geometric thickness (CGT), and liquid water content at the TP (TP_LWC). In contrast, the normalized optical thickness at the TP (TP_NCOT) depends primarily on its normalized height (TP_NH) and weakly on other parameters.

Multilinear regression is applied using POLDER-3 data to estimate TP_CER. For maritime clouds, combining CB_CER and the LWP achieves high accuracy (RMSE: 1.19–1.30μm), whereas continental clouds require only CB_CER (RMSE: 1.75–1.96μm). For TP_NCOT, random forest outperforms linear regression, with optimal results (RMSE≈0.1) using CB_CER, cloud-top CER (CT_CER), CGT, and the LWP. Cloud-top height (CTH) could partially substitute for CGT.

The primary challenges in retrieving profile structural features originate from the following aspects: (1) The coarse resolution of POLDER products restricts the ability to capture sub-pixel cloud heterogeneity; however, by concentrating on relatively uniform single-layer stratiform liquid clouds, this study partially mitigates the resulting retrieval uncertainties. It should be noted that sub-pixel heterogeneity can inevitably introduce certain errors, particularly at cloud boundaries. Nevertheless, Shang et al. (2015) pointed out that the error caused by sub-pixel heterogeneity in cloud effective radius (CER) retrieval does not exceed 10%, which remains within an acceptable range. (2) The estimation of CB_CER remains subject to certain uncertainties due to the inherent challenges in retrieving cloud base microphysical properties from passive observations; (3) The 240-m vertical resolution of CloudSat is insufficient to resolve ultra-thin cloud layers or capture fine-scale in-cloud structures, such as sharp inversion layers or thin drizzling layers near cloud base.

To address the issues mentioned above, the improvement strategies below can be implemented: (1) Internationally, there are currently polarimetric multi-angle payloads with higher spatial resolution and greater observation angles that have been launched or are planned for launch. For instance, China's DPC/GF-5 achieves a spatial resolution of nadir 3.3 km; the 3MI/Metop-SG developed by the European Space Agency offers a spatial resolution of nadir 4 km, supports up to 21 observation angles, and incorporates near-infrared bands. These capabilities collectively enable higher-resolution CER retrieval. (2) In terms of estimating the CB_CER: a) Introduce meteorological factors—such as ERA5 reanalysis data (e.g., wind speed, temperature, humidity, pressure, vertical velocity)—to assist in estimating the CB_CER, thereby enhancing the physical characterization of the cloud-base environment; b) Optimize existing methods for directly retrieving cloud-bottom particle size using passive observations (Level 1 or Level 2 products), improve the robustness of the retrieval model, and clarify the applicable boundaries of the method—specifically, determining the optical thickness threshold beyond which passive observations can no longer capture cloud-bottom information; c) Incorporate an uncertainty weighting framework to dynamically adjust the contribution weights of different input parameters based on their reliability, thereby refining the retrieval accuracy of CB_CER and reducing dependence on CloudSat-derived empirical relationships. (3) For the issue that CloudSat is insufficient to distinguish thin clouds or fine cloud structures smaller than 240m, in the future, the EarthCARE/CPR observation data with higher vertical resolution can be adopted to alleviate this problem. EarthCARE/CPR

demonstrates notable advancements over CloudSat, most significantly through its finer vertical resolution of 100 m compared to CloudSat's 240 m. Additional enhancements include higher detection sensitivity, Doppler-based vertical wind measurements, and synchronized multi-sensor observational capabilities. These improvements are expected to deliver enhanced observational capabilities for characterizing finer-scale cloud microphysical processes and their interactions with atmospheric dynamics.

Meanwhile, the validation results indicate that the RMSE of stratiform cloud profile structural characteristics over land is significantly higher than that over sea. This discrepancy is considered to be mainly attributable to the following factors: a) Sub-pixel Surface Heterogeneity: Variations in surface reflectance among different land cover types (e.g., vegetation, bare soil, urban areas) lead to mixed-pixel effects, complicating the decoupling of cloud optical properties. b) Aerosol Interference: Higher and spatiotemporally variable aerosol loadings over land can perturb cloud signals either indirectly by altering cloud microphysics (e.g., through cloud condensation nuclei effects) or directly via scattering. c) Surface Heating Effects: The lower thermal inertia of land surfaces results in more complex boundary-layer dynamics, increasing spatiotemporal variability in cloud base height and cloud layer thickness, which in turn elevates retrieval uncertainty. d) Interference from complex terrain and high-albedo surfaces: Complex terrain (e.g., mountains) and high-albedo surfaces (e.g., snow cover) are prone to causing false positives in cloud detection or overestimation of optical thickness. It is suggested that the following strategies could be adopted in the future to improve the estimation accuracy of stratiform cloud profile structural characteristics over land：a) Integration of land cover classification data (e.g., MODIS Land Cover product); b) Integration of aerosol ancillary data: Multi-source aerosol observations (e.g., MERRA-2 reanalysis data, AERONET ground-based measurements) could be incorporated to better constrain retrieval parameters in regions affected by aerosol-cloud interactions; c) Development of advanced retrieval algorithms: More sophisticated methods, such as machine learning or deep learning approaches, could be employed to better represent the complex relationships between land surface, atmosphere, and clouds.

Future work should focus on higher-resolution observations and improved retrieval methods to refine cloud structural analysis. In summary, this study advances methods for estimating TP characteristics in liquid clouds but underscores the need for enhanced observational capabilities and hybrid active-passive approaches to fully resolve profile uncertainties. Additionally, our work on the parameterization and retrieval of liquid cloud profiles through multiangle passive imagers provides valuable insights that can further improve the understanding and modeling of cloud processes in weather and climate systems.

Appendix A

Table A1. Table of acronyms.

Abbreviation	Full term
CALIPSO	Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations
CB_CER	Effective Radius of Cloud Base
CBH	Cloud Base Height
CER	Cloud Effective Radius
CGT	Cloud Geometric Thickness
CNES	Centre National d'Études Spatiales
CPM	Cloud Profile Model
CPR	Cloud Profile Radar
CPRM	Cloud Profile Reconstruction Model
CSU	The Colorado State University
CT_CER	Effective Radius of Cloud Top
CTH	Cloud Top Height
Dec_Inc	Decreasing then increasing
ECMWF	European Centre for Medium-Range Weather Forecasts
GCM	General Circulation Model
Inc_Dec	Increasing then Decreasing
LWP	Liquid Water Path
MLR	Multiple Linear Regression
Mono_Dec	Monotonically Decreasing
Mono_Inc	Monotonically Increasing
POLDER	Polarization and Directionality of Earth's Reflectance
RAMS	Regional Atmospheric Modeling System
RB2	Level 2 products of POLDER-3
RF	Random Forest
RMSE	Root Mean Square Error
TP	Turning Point
TP_CER	Cloud Effective Radius at the Turning Point
TP_LWC	Liquid Water Content at the Turning Point
TP_NCOT	Normalized Cloud Optical Thickness at the Turning Point
TP_NH	Normalized Height at the Turning Point

Table A2. Empirical coefficient and predictive performance of estimating TP_CER via the multiple linear regression method.

Cloud profile type	Empirical coefficient			Predictive performance		
	β_0	β_1(CB_CER)	β_2(LWP)	R^2	R	RMSE
Sea Nonprecip	2.2656	0.8342	0.0052	0.77	0.90	1.19
Sea Precip	3.6904	0.7920	0.0022	0.87	0.94	1.30
Land Nonprecip	0.5844	1.1234	—	0.54	0.83	1.76
Land Precip	3.7843	0.8985	—	0.63	0.86	1.96

Table A3. Statistics on the number of stratiform clouds categorized by land and sea, precipitating and nonprecipitating.

	Sc	St	Sc+ St
Land Nonprecip	1034462	15823	1050285
Sea Nonprecip	4016879	135926	4152805
Land Precip	853197	28468	881665
Sea Precip	6188612	199916	6388528
Sum	12093150	380133	12473283

Table A4. Further statistics on CER profiles of complex shapes ("Other": category V in profile shape).

	Sc+St		Sc+St	
	Land Nonprecip	Sea Nonprecip	Land Precip	Sea Precip
Sum total	137699	308344	84026	608258
Percentage of situation 1[a]	33.57%	33.57%	47.88%	49.68%
Percentage of situation 2[b]	31.41%	36.72%	19.07%	18.93%
Intersection of 1 and 2[c]	8.40%	11.72%	6.31%	6.21%
Percentage of situation 1+2	56.58%	58.57%	60.63%	62.40%

[a] situation1 refers to a situation where only one segment of the profile does not correspond to the increasing and then decreasing shape profile of shape1; [b] situation2 refers to a situation where only one segment of the profile does not correspond to the monotonically decreasing shape profile of shape2. [c] There is an intersection of situation1 and situation2, i.e., a profile that matches both situation1 and situation2 (Intersection of 1+2), which needs to be subtracted out when calculating the sum of the two in order to avoid double counting.

Table A5. Validation accuracy of TP_CER linear regression when different parameters are used.

Independent variables	Dependent variables	R^2	R	RMSE	Regression data volume	Validation data Volume
TP_CER (Sea Nonprecip)	CB_CER; LWP	**0.77**	**0.90**	**1.19**	1400000	1389529
	CB_CER	0.52	0.82	1.59		
	LWP	0.61	0.85	1.47		
TP_CER (Sea Precip)	CB_CER; LWP	**0.87**	**0.94**	**1.30**	700000	710750
	CB_CER	0.82	0.92	1.46		
	LWP	0.35	0.78	2.37		
TP_CER (Land Nonprecip)	CB_CER; LWP	**0.54**	**0.83**	**1.75**	250000	260999
	CB_CER	0.54	0.83	1.76		
	LWP	0.28	0.76	2.43		
TP_CER (Land Precip)	CB_CER; LWP	**0.63**	**0.86**	**1.96**	120000	118706

Table A6. Validation accuracy of TP_NCOT predictions when different parameters and methods are used.

Independent variables	Dependent variables	R	RMSE	rRMSE	Method
TP_NormCOT (Sea Nonprecip)	CB_CER, CT_CER, CGT, LWP	**0.59**	**0.10**	**0.23**	RF
	CB_CER, CT_CER, CTH, LWP	0.56	0.10	0.23	
	CT_CER, CGT, LWP	0.39	0.12	0.27	
	CB_CER, CT_CER, LWP	0.49	0.11	0.25	
	CB_CER, CT_CER, CGT, LWP	0.51	0.11	0.24	MLR
TP_NormCOT (Sea Precip)	CB_CER, CT_CER, CGT, LWP	0.63	0.11	0.22	RF
	CB_CER, CT_CER, CTH, LWP	**0.64**	**0.11**	**0.21**	
	CT_CER, CTH, LWP	0.40	0.14	0.26	
	CB_CER, CT_CER, CGT, LWP	0.52	0.12	0.24	MLR
TP_NormCOT (Land Nonprecip)	CB_CER, CT_CER, CGT, LWP	**0.58**	**0.12**	**0.29**	RF
	CB_CER, CT_CER, CTH, LWP	0.53	0.13	0.30	
	CT_CER, CTH, LWP	0.37	0.14	0.33	
	CB_CER, CT_CER, CGT, LWP	0.54	0.12	0.29	MLR
TP_NormCOT (Land Precip)	CB_CER, CT_CER, CGT, LWP	**0.66**	**0.12**	**0.27**	RF
	CB_CER, CT_CER, CTH, LWP	0.65	0.12	0.28	
	CT_CER, CGT, LWP	0.33	0.16	0.35	
	CB_CER, CT_CER, CGT, LWP	0.60	0.13	0.29	MLR

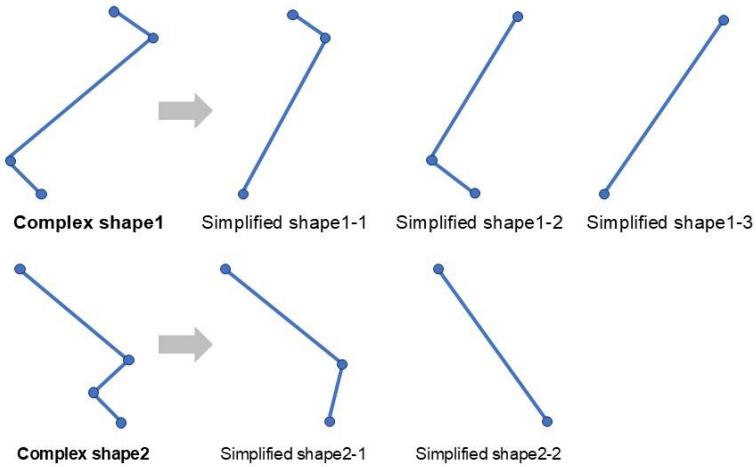

Figure A1. Complex CER profile shapes and their possible corresponding simplified shapes (examples).

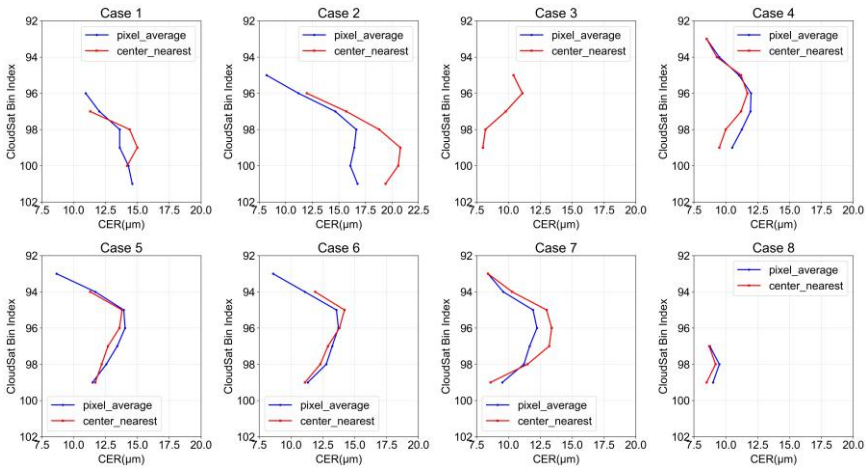

Figure A2. Comparison of the average CER profile (averaged by height) within the same POLDER pixel versus the CER profile closest to the center of the POLDER pixel.

Data and Code Availability. All datasets used in this work are open-source. The CloudSat datasets are available from the CloudSat Data Processing Center of the Cooperative Institute for Research in the Atmosphere (http://www.cloudsat.cira.colostate.edu/). The Parasol products are available from ICARE Data and Services Center (https://www.icare.univ-lille.fr/). The code and pre-trained model for the cloud base height retrieval algorithm, along with a minimal dataset required to reproduce the key results, have been deposited on Zenodo under a permanent DOI: [10.5281/zenodo.17082185].

Author contribution. HS and HL outlined the project, HS and YW conceived the methodology and performed the experiments, YW composed the manuscript, and all authors revised the manuscript. HS and HL funded, supervised, and encouraged the research.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Acknowledgements. This work was supported by the National Key Research and Development Program of China (Grant No. 2023YFB3905900) and National Natural Science Foundation of China (B) (Grant No.42522507). We are grateful to the Data Processing Center (DPC) managed by the Cooperative Institute for Research in the Atmosphere (CIRA) for providing CloudSat data, which has greatly facilitated our research. The authors would also like to thank ICARE Data and Services Center, Villeneuve d'Ascq Cedex, France, for providing the POLDER-3/Parasol product (https://www.icare.univ-lille.fr/).

References

Alexandrov, M. D., Miller, D. J., Rajapakshe, C., Fridlind, A., van Diedenhoven, B., Cairns, B., Ackerman, A. S., and Zhang, Z.: Vertical profiles of droplet size distributions derived from cloud-side observations by the research scanning polarimeter: Tests on simulated data, Atmospheric Research, 239, 104924, https://doi.org/10.1016/j.atmosres.2020.104924, 2020.

Austin, R.: Level 2B radar-only cloud water content (2B-CWC-RO) process description document, Data Processing Center, 24, 2007.

Barker, H. W., Jerg, M. P., Wehr, T., Kato, S., Donovan, D. P., and Hogan, R. J.: A 3D cloud-construction algorithm for the EarthCARE satellite mission, Quarterly Journal of the Royal Meteorological Society, 137, 1042-1058, https://doi.org/10.1002/qj.824, 2011.

Battaglia, A., Kollias, P., Dhillon, R., Roy, R., Tanelli, S., Lamer, K., Grecu, M., Lebsock, M., Watters, D., Mroz, K., Heymsfield, G., Li, L., and Furukawa, K.: Spaceborne Cloud and Precipitation Radars: Status, Challenges, and Ways Forward, Reviews of Geophysics, 58, e2019RG000686, https://doi.org/10.1029/2019RG000686, 2020.

Bréon, F.-M. and Goloub, P.: Cloud droplet effective radius from spaceborne polarization measurements, Geophysical Research Letters, 25, 1879-1882, https://doi.org/10.1029/98GL01221, 1998.

Breon, F. M. and Doutriaux-Boucher, M.: A comparison of cloud droplet radii measured from space, IEEE Transactions on Geoscience and Remote Sensing, 43, 1796-1805, 10.1109/TGRS.2005.852838, 2005.

Chen, R., Wood, R., Li, Z., Ferraro, R., and Chang, F.-L.: Studying the vertical variation of cloud droplet effective radius using ship and space-borne remote sensing data, Journal of Geophysical Research: Atmospheres, 113, https://doi.org/10.1029/2007JD009596, 2008.

Chen, T., Rossow, W. B., and Zhang, Y.: Radiative Effects of Cloud-Type Variations, Journal of Climate, 13, 264-286, https://doi.org/10.1175/1520-0442(2000)013<0264:REOCTV>2.0.CO;2, 2000.

Chen, Y., Chen, G., Cui, C., Zhang, A., Wan, R., Zhou, S., Wang, D., and Fu, Y.: Retrieval of the vertical evolution of the cloud effective radius from the Chinese FY-4 (Feng Yun 4) next-generation geostationary satellites, Atmos. Chem. Phys., 20, 1131-1145, 10.5194/acp-20-1131-2020, 2020.

Deschamps, P. Y., Breon, F. M., Leroy, M., Podaire, A., Bricaud, A., Buriez, J. C., and Seze, G.: The POLDER mission: instrument characteristics and scientific objectives, IEEE Transactions on Geoscience and Remote Sensing, 32, 598-615, 10.1109/36.297978, 1994.

Dong, X. and Minnis, P.: Stratus, Stratocumulus, and Remote Sensing, Fast Processes in Large-Scale Atmospheric Models: Progress, Challenges, and Opportunities, 141-199, 2023.

Dubovik, O., Li, Z., Mishchenko, M. I., Tanré, D., Karol, Y., Bojkov, B., Cairns, B., Diner, D. J., Espinosa, W. R., Goloub, P., Gu, X., Hasekamp, O., Hong, J., Hou, W., Knobelspiesse, K. D., Landgraf, J., Li, L., Litvinov, P., Liu, Y., Lopatin, A., Marbach, T., Maring, H., Martins, V., Meijer, Y., Milinevsky, G., Mukai, S., Parol, F., Qiao, Y., Remer, L., Rietjens, J., Sano, I., Stammes, P., Stamnes, S., Sun, X., Tabary, P., Travis, L. D., Waquet, F., Xu, F., Yan, C., and Yin, D.: Polarimetric remote sensing of atmospheric aerosols: Instruments, methodologies, results, and perspectives, Journal of Quantitative Spectroscopy and Radiative Transfer, 224, 474-511, https://doi.org/10.1016/j.jqsrt.2018.11.024, 2019.

Fougnie, B., Marbach, T., Lacan, A., Lang, R., Schlüssel, P., Poli, G., Munro, R., and Couto, A. B.: The multi-viewing multi-channel multi-polarisation imager – Overview of the 3MI polarimetric mission for aerosol and cloud characterization, Journal of Quantitative Spectroscopy and Radiative Transfer, 219, 23-32, https://doi.org/10.1016/j.jqsrt.2018.07.008, 2018.

Fox, N. I. and Illingworth, A. J.: The Retrieval of Stratocumulus Cloud Properties by Ground-Based Cloud Radar, Journal of Applied Meteorology, 36, 485-492, https://doi.org/10.1175/1520-0450(1997)036<0485:TROSCP>2.0.CO;2, 1997.

Greenwald, T. J., Stephens, G. L., Christopher, S. A., and Vonder Haar, T. H.: Observations of the global characteristics and regional radiative effects of marine cloud liquid water, Journal of climate, 8, 2928-2946, 1995.

Horváth, Á. and Davies, R.: Anisotropy of water cloud reflectance: A comparison of measurements and 1D theory, Geophysical Research Letters, 31, https://doi.org/10.1029/2003GL018386, 2004.

Ji, T., Shang, H., Bao, F., Liu, Z., Wang, Y., Bao, S., Yin, S., Shi, C., Wang, H., Liu, Z., and Letu, H.: Retrieval of the base heights and cloud geometric thicknesses of clouds based on the PARASOL measurement, SSRN, https://ssrn.com/abstract=5515438 or http://dx.doi.org/10.2139/ssrn.5515438, 2025.

Kessler, E.: On the Distribution and Continuity of Water Substance in Atmospheric Circulations, in: On the Distribution and Continuity of Water Substance in Atmospheric Circulations, edited by: Kessler, E., American Meteorological Society, Boston, MA, 1-84, 10.1007/978-1-935704-36-2_1, 1969.

Lamer, K., Kollias, P., Battaglia, A., and Preval, S.: Mind the gap – Part 1: Accurately locating warm marine boundary layer clouds and precipitation using spaceborne radars, Atmos. Meas. Tech., 13, 2363-2379, 10.5194/amt-13-2363-2020, 2020.

Leinonen, J., Guillaume, A., and Yuan, T.: Reconstruction of Cloud Vertical Structure With a Generative Adversarial Network, Geophysical Research Letters, 46, 7035-7044, https://doi.org/10.1029/2019GL082532, 2019.

Letu, H., Yang, K., Nakajima, T. Y., Ishimoto, H., Nagao, T. M., Riedi, J., Baran, A. J., Ma, R., Wang, T., Shang, H., Khatri, P., Chen, L., Shi, C., and Shi, J.: High-resolution retrieval of cloud microphysical properties and surface solar radiation using Himawari-8/AHI next-generation geostationary satellite, Remote Sensing of Environment, 239, 111583, https://doi.org/10.1016/j.rse.2019.111583, 2020.

Letu, H., Ma, R., Nakajima, T. Y., Shi, C., Hashimoto, M., Nagao, T. M., Baran, A. J., Nakajima, T., Xu, J., Wang, T., Tana, G., Bilige, S., Shang, H., Chen, L., Ji, D., Lei, Y., Wei, L., Zhang, P., Li, J., Li, L., Zheng, Y., Khatri, P., and Shi, J.: Surface Solar Radiation Compositions Observed from Himawari-8/9 and Fengyun-4 Series, Bulletin of the American Meteorological Society, 104, E1772-E1789, https://doi.org/10.1175/BAMS-D-22-0154.1, 2023.

Liu, Y., Daum, P. H., McGraw, R., and Wood, R.: Parameterization of the autoconversion process. Part II: Generalization of Sundqvist-type parameterizations, Journal of the atmospheric sciences, 63, 1103-1109, 2006.

Mace, G. G., Zhang, Q., Vaughan, M., Marchand, R., Stephens, G., Trepte, C., and Winker, D.: A description of hydrometeor layer occurrence statistics derived from the first year of merged Cloudsat and CALIPSO data, Journal of Geophysical Research: Atmospheres, 114, https://doi.org/10.1029/2007JD009755, 2009.

Nakajima, T. Y., Suzuki, K., and Stephens, G. L.: Droplet Growth in Warm Water Clouds Observed by the A-Train. Part I: Sensitivity Analysis of the MODIS-Derived Cloud Droplet Sizes, Journal of the Atmospheric Sciences, 67, 1884-1896, https://doi.org/10.1175/2009JAS3280.1, 2010a.

Nakajima, T. Y., Suzuki, K., and Stephens, G. L.: Droplet Growth in Warm Water Clouds Observed by the A-Train. Part II: A Multisensor View, Journal of the Atmospheric Sciences, 67, 1897-1907, https://doi.org/10.1175/2010JAS3276.1, 2010b.

Platnick, S.: Approximations for horizontal photon transport in cloud remote sensing problems, Journal of Quantitative Spectroscopy and Radiative Transfer, 68, 75-99, https://doi.org/10.1016/S0022-4073(00)00016-9, 2001.

Protat, A., Bouniol, D., Delanoë, J., O'Connor, E., May, P. T., Plana-Fattori, A., Hasson, A., Görsdorf, U., and Heymsfield, A. J.: Assessment of Cloudsat Reflectivity Measurements and Ice Cloud Properties Using Ground-Based and Airborne Cloud Radar Observations, Journal of Atmospheric and Oceanic Technology, 26, 1717-1741, https://doi.org/10.1175/2009JTECHA1246.1, 2009.

Rosenfeld, D. and Lensky, I. M.: Satellite-Based Insights into Precipitation Formation Processes in Continental and Maritime Convective Clouds, Bulletin of the American Meteorological Society, 79, 2457-2476, https://doi.org/10.1175/1520-0477(1998)079<2457:SBIIPF>2.0.CO;2, 1998.

Schulte, R. M., Lebsock, M. D., and Haynes, J. M.: What CloudSat cannot see: liquid water content profiles inferred from MODIS and CALIOP observations, Atmos. Meas. Tech., 16, 3531-3546, 10.5194/amt-16-3531-2023, 2023.

Shang, H., Hioki, S., Penide, G., Cornet, C., Letu, H., and Riedi, J.: Establishment of an analytical model for remote sensing of typical stratocumulus cloud profiles under various precipitation and entrainment conditions, Atmospheric Chemistry and Physics, 23, 2729-2746, 2023.

Shang, H., Chen, L., Bréon, F. M., Letu, H., Li, S., Wang, Z., and Su, L.: Impact of cloud horizontal inhomogeneity and directional sampling on the retrieval of cloud droplet size by the POLDER instrument, Atmos. Meas. Tech., 8, 4931-4945, 10.5194/amt-8-4931-2015, 2015.

Shang, H., Letu, H., Wei, L., Ma, R., Wang, Y., Cai, Z., Yin, S., and Shi, C.: Remote sensing of liquid cloud profiles based on an analytical cloud profiling model, Science China Earth Sciences, 68, 998-1012, 10.1007/s11430-024-1532-2, 2025.

Shang, H., Letu, H., Bréon, F.-M., Riedi, J., Ma, R., Wang, Z., Nakajima, T. Y., Wang, Z., and Chen, L.: An improved algorithm of cloud droplet size distribution from POLDER polarized measurements, Remote Sensing of Environment, 228, 61-74, https://doi.org/10.1016/j.rse.2019.04.013, 2019.

Shi, C., Letu, H., Nakajima, T. Y., Nakajima, T., Wei, L., Xu, R., Lu, F., Riedi, J., Ichii, K., Zeng, J., Shang, H., Ma, R., Yin, S., Shi, J., Baran, A. J., Xu, J., Li, A., Tana, G., Wang, W., Na, Q., Sun, Q., Yang, W., Chen, L., and Shi, G.: Near-global monitoring of surface solar radiation through the construction of a geostationary satellite network observation system, The Innovation, 6, 100876, https://doi.org/10.1016/j.xinn.2025.100876, 2025.

Sinclair, K., van Diedenhoven, B., Cairns, B., Alexandrov, M., Dzambo, A. M., and L'Ecuyer, T.: Inference of Precipitation in Warm Stratiform Clouds Using Remotely Sensed Observations of the Cloud Top Droplet Size Distribution, Geophysical Research Letters, 48, e2021GL092547, https://doi.org/10.1029/2021GL092547, 2021.

Slingo, A.: Sensitivity of the Earth's radiation budget to changes in low clouds, Nature, 343, 49-51, 10.1038/343049a0, 1990.

Tana, G., Ri, X., Shi, C., Ma, R., Letu, H., Xu, J., and Shi, J.: Retrieval of cloud microphysical properties from Himawari-8/AHI infrared channels and its application in surface shortwave downward radiation estimation in the sun glint region, Remote Sensing of Environment, 290, 113548, https://doi.org/10.1016/j.rse.2023.113548, 2023.

Tang, C., Shi, C., Letu, H., Yin, S., Nakajima, T., Sekiguchi, M., Xu, J., Zhao, M., Ma, R., and Wang, W.: Development of a hybrid algorithm for the simultaneous retrieval of aerosol optical thickness and fine-mode fraction from multispectral satellite observation combining radiative transfer and transfer learning approaches, Remote Sensing of Environment, 319, 114619, https://doi.org/10.1016/j.rse.2025.114619, 2025.

Turner, D. D., Vogelmann, A. M., Austin, R. T., Barnard, J. C., Cady-Pereira, K., Chiu, J. C., Clough, S. A., Flynn, C., Khaiyer, M. M., Liljegren, J., Johnson, K., Lin, B., Long, C., Marshak, A., Matrosov, S. Y., McFarlane, S. A., Miller, M., Min, Q., Minimis, P., O'Hirok, W., Wang, Z., and Wiscombe, W.: Thin Liquid Water Clouds: Their Importance and Our Challenge, Bulletin of the American Meteorological Society, 88, 177-190, https://doi.org/10.1175/BAMS-88-2-177, 2007.

Visvalingam, M.: The Visvalingam Algorithm: Metrics, Measures and Heuristics, The Cartographic Journal, 53, 242-252, 10.1080/00087041.2016.1151097, 2016.
WANG, Y., MA, J., LI, J., HONG, J., and LI, Z.: Review of cloud polarimetric remote sensing, National Remote Sensing Bulletin, 26, 852-872, 10.11834/jrs.20221404, 2022.
Warren, S. G., Eastman, R. M., and Hahn, C. J.: A survey of changes in cloud cover and cloud types over land from surface observations, 1971–96, Journal of climate, 20, 717-738, 2007.
Warren, S. G., Hahn, C. J., London, J., Chervin, R. M., and Jenne, R. L.: Global distribution of total cloud cover and cloud type amounts over land, United States, Other Information: Paper copy only, copy does not permit microfiche production. Original copy available until stock is exhausted, 1986.
Warren, S. G., Hahn, C. J., London, J., Chervin, R. M., Jenne, R. L., Colorado Univ, B. C. O. C. I. f. R. i. E. S., Colorado Univ, B. C. O. D. o. A. P., Atmospheric, S., and National Center for Atmospheric Research, B. C. O.: Global distribution of total cloud cover and cloud type amounts over the ocean, United States, 10.2172/5415329
Other Information: This report contains 37 microfiche supplements, 1988.
Wood, R.: Stratocumulus Clouds, Monthly Weather Review, 140, 2373-2423, https://doi.org/10.1175/MWR-D-11-00121.1, 2012.
Wood, R.: CLOUDS AND FOG | Stratus and Stratocumulus, in: Encyclopedia of Atmospheric Sciences (Second Edition), edited by: North, G. R., Pyle, J., and Zhang, F., Academic Press, Oxford, 196-200, https://doi.org/10.1016/B978-0-12-382225-3.00396-0, 2015.
Zhang, Z., Platnick, S., Yang, P., Heidinger, A. K., and Comstock, J. M.: Effects of ice particle size vertical inhomogeneity on the passive remote sensing of ice clouds, Journal of Geophysical Research: Atmospheres, 115, https://doi.org/10.1029/2010JD013835, 2010.
Zhao, Y., Li, J., Wang, Y., Zhang, W., and Wen, D.: Warming Climate-Induced Changes in Cloud Vertical Distribution Possibly Exacerbate Intra-Atmospheric Heating Over the Tibetan Plateau, Geophysical Research Letters, 51, e2023GL107713, https://doi.org/10.1029/2023GL107713, 2024.