Evaluation of MODIS and Himawari-8 Low Clouds Retrievals over the Southern Ocean with In Situ Measurements from the SOCRATES Campaign

Litai Kang[1], Roger T Marchand[2], and William Smith, Jr.[3]

[1]Department of Atmospheric Sciences, University of Washington
[2]University of Washington
[3]NASA Langley Research Center

November 24, 2022

Abstract

Aircraft observations collected during the Southern Ocean Cloud Radiation Aerosol Transport Experimental Study (SOCRATES) in January-February of 2018 are used to evaluate cloud properties from three satellite-imager datasets: (i) the Moderate Resolution Imaging Spectroradiometer (MODIS) level 2 (collection 6.1) cloud product, (ii) the CERES-MODIS Edition 4 cloud product, and (iii) the NASA SatCORPS Himawari-8 cloud product.

Overall the satellite retrievals compare well with the in situ observations, with little bias and modest to good correlation coefficients when considering all aircraft profiles for which there are coincident MODIS observations. The Himawari-8 product does, however, show a statistically significant mean bias of about 1.2 μm for effective radius (r_e) and 2.6 for optical depth (τ) when applied to a larger set of profiles with coincident Himawari-8 observations.

The low overall mean-bias in the r_e retrievals is due in part to compensating errors between cases that are non- or lightly-precipitating, with cases that have heavier precipitation. r_e is slightly biased high (by about 0.5 to 1.0 μm) for non- and lightly-precipitating cases and biased low by about 3 to 4 μm for heavily-precipitating cases when precipitation exits near cloud top. The bias in non- and lightly-precipitating conditions is due to (at least in part) having assumed a drop size distribution in the retrieval that is too broad. These biases in the r_e ultimately propagate into the retrieved liquid water path and number concentration.

Evaluation of MODIS and Himawari-8 Low Clouds Retrievals over the Southern Ocean with In Situ Measurements from the SOCRATES Campaign

L. Kang[1], R. T. Marchand[1], and W. L. Smith[2]

[1] Department of Atmospheric Sciences, University of Washington, Seattle, Washington, USA

[2] NASA Langley Research Center, Hampton, Virginia, USA

Corresponding author: Litai Kang (kanglt@uw.edu)

Key Points:

- Overall imager-based bi-spectral retrievals are found to work reasonably well for Southern Ocean overcast (closed-cell) stratocumulus.
- Effective radius is biased high by 0.5 to 1.0 µm for non or lightly-precipitating cases and biased low by about 3 to 4 µm otherwise.
- These biases are due in part to (1) the assumed drop size distribution being too broad and (2) precipitation being present near cloud top.

Abstract

Aircraft observations collected during the Southern Ocean Cloud Radiation Aerosol Transport Experimental Study (SOCRATES) in January-February of 2018 are used to evaluate cloud properties from three satellite-imager datasets: (i) the Moderate Resolution Imaging Spectroradiometer (MODIS) level 2 (collection 6.1) cloud product, (ii) the CERES-MODIS Edition 4 cloud product, and (iii) the NASA SatCORPS Himawari-8 cloud product.

Overall the satellite retrievals compare well with the in situ observations, with little bias and modest to good correlation coefficients when considering all aircraft profiles for which there are coincident MODIS observations. The Himawari-8 product does, however, show a statistically significant mean bias of about 1.2 μm for effective radius (r_e) and 2.6 for optical depth (τ) when applied to a larger set of profiles with coincident Himawari-8 observations.

The low overall mean-bias in the r_e retrievals is due in part to compensating errors between cases that are non- or lightly-precipitating, with cases that have heavier precipitation. r_e is slightly biased high (by about 0.5 to 1.0 μm) for non- and lightly-precipitating cases and biased low by about 3 to 4 μm for heavily-precipitating cases when precipitation exits near cloud top. The bias in non- and lightly-precipitating conditions is due to (at least in part) having assumed a drop size distribution in the retrieval that is too broad. These biases in the r_e ultimately propagate into the retrieved liquid water path and number concentration.

Plain Language Summary

Clouds play a crucial role in the weather and climate system. Satellite data can provide useful information on cloud properties (such as the size of the cloud droplets, the amount of the liquid water, and the number of droplets in a given volume of the clouds) over large areas and at high spatial and temporal resolutions. However, satellite cloud properties are determined or retrieved from satellite measurements by employing a variety of simplifying assumptions that can lead to large uncertainties in some conditions. In-situ measurements of clouds from aircraft provide more direct observations and can be used as ground truth to evaluate and improve the performance of the satellite retrievals. This study focuses on clouds over the Southern Ocean and uses aircraft measurements from Southern Ocean Cloud Radiation Aerosol Transport Experimental Study (SOCRATES) in January-February of 2018 to evaluate cloud properties from three satellite observations. It is found that the satellite observations generally compare well with aircraft measurements with little bias. However, satellite observations tend to overestimate the size of the cloud droplets, when clouds are not precipitating or are lightly precipitating, while for clouds with heavier precipitation, the satellite observations tend to underestimate the size of the cloud droplets.

1 Introduction

The Southern Ocean (SO) is the one of the cloudiest regions in the world, in large part because of extensive stratiform marine boundary layer (MBL) cloud (Mace et al., 2009, 2010). Compared against satellite datasets, climate models and present-day reanalysis products predict too little MBL cloud over the SO, especially in the cold sectors of SO cyclones (Williams et al., 2013; Bodas-Salcedo et al., 2014; Naud et al., 2014). The insufficient cloud cover causes significant biases in shortwave radiative fluxes over the SO (Trenberth and Fasullo, 2010;

Schneider and Reusch, 2016) and contributes to biases in the simulated surface air and sea surface temperatures (Sallée et al., 2013; Bodas-Salcedo et al., 2016). In turn, these model biases have impacts on regional and global circulations, including influencing the position and strength of the Southern Hemisphere midlatitude jet, the Inter-Tropical Convergence Zone (ITCZ) and cross-hemispheric energy transports (Ceppi et al., 2012, 2013; Hwang and Frierson, 2013; Kay et al., 2016).

Cloud properties, such as clouds effective radius (r_e), optical depth (τ), liquid water path (LWP) and cloud droplet number concentration (N_d) are central in understanding the physics of MBL clouds and their radiative effect. Geostationary and polar orbiting satellite visible and infrared observations have long been used to retrieve MBL cloud microphysical characteristics and have been widely used for the study of SO clouds, cloud-aerosol interactions and for the evaluation of global models (e.g. Meskhidze and Nenes, 2006; Haynes et al., 2011; McCoy et al., 2015; Bodas-Salcedo et al., 2016; Vergara-Temprado et al., 2018). However, the accuracy of satellite retrievals over the SO is questionable, as satellite retrievals have been infrequently evaluated against in situ measurements in this region, due in part to the remoteness of the region and a paucity of in situ measurements. The validation, empirical relationships, and apriori data used in satellite retrieval algorithms are mostly based on data collected in the Northern Hemisphere and might not be applicable over the SO. In general, low-level SO clouds are thought to be more frequently multilayered, mixed phase, and contain more supercooled liquid water than in the northern hemisphere, conditions which poses significant challenges for satellite retrievals (Morrison et al., 2011; Huang et al., 2014).

A direct evaluation of satellite cloud retrievals can be made using in situ measurements from aircraft, and many such studies have been done over the years, including in recent years for the Southeast Pacific (Painemal and Zuidema, 2011; Min et al., 2012; King et al., 2013) and Northeastern Pacific (Noble and Hudson, 2015). There have been a few cases where in situ measurements have been collected from aircraft over the SO. Four transects over the SO were made during the HIAPER Pole to Pole Observations (HIPPO) experiment (Wofsy, 2011). HIPPO confirmed the existence of extensive supercooled liquid water in the region but collected insufficient data to directly evaluate coincident satellite microphysical retrievals. More recently, in situ measurements from 20 flights were made over the SO to the west and south of Tasmania (43-45°S, 145–148°E) during the austral winter during 2013-2015 by Ahn et al. (2017). These flights focused on the microphysical properties of low-level clouds, which were found to be commonly precipitating, patchy and mixed phase. Ahn et al. (2018, hereafter A18) compared in situ observations from 11 of these flights to cloud products from Moderate Resolution Imaging Spectroradiometer (MODIS) and found an overestimation of MODIS cloud droplet effective radius (r_e) in comparison with in situ measurements.

More recently, Southern Ocean Cloud Radiation Aerosol Transport Experimental Study (SOCRATES) collected airborne in situ measurements over the SO. During SOCRATES, NSF deployed the Gulfstream-V (GV) research aircraft to Hobart, Tasmania from January to February of 2018. From Hobart, the GV flew a total of 15 research flights over the SO as far as 62°S, sampling aerosol, cloud and precipitation properties in situ, as well as remotely with cloud (W-band) radar and high spectral resolution lidar. SOCRATES provides an opportunity to evaluate satellite cloud products and retrieval assumptions during the austral summer over the SO. In this study, we evaluate low altitude cloud microphysical properties retrieved from satellites using airborne in situ measurements collected during SOCRATES. After describing the datasets and methods used in section 2, in section 3 we compare the satellite retrievals of effective radius (r_e), optical depth (τ), liquid water path (LWP) and cloud droplet number concentration (N_d) from three datasets that are based on observations from MODIS (Platnick et al., 2003) and Himawari-8 (geostationary weather satellite; Bessho et al., 2016) with values obtained from the GV in situ measurements. This is followed in section 4, by a more detailed examination of retrieval assumptions and other factors that are responsible for differences between the satellite retrievals and in situ data, with conclusions and final remarks given in section 5.

SOCRATES specifically targeted stratocumulus primarily overcast or closed cell stratocumulus, that reside in the cold sectors of low-pressure system, and the SOCRATES data do not represent a meteorologically unbiased set of conditions. Nonetheless stratocumulus clouds are a significant fraction of all SO low clouds (Wood, 2012), and our focus on these clouds during SOCRATES results from a recognition that these clouds lay at the heart of difficulties that many models are having in simulating the climate of the SO. Based on results from other regions, one expects that satellite retrievals for these relatively spatially homogenous clouds should work well (e.g., Painemal and Zuidema, 2011; hereafter PZ11). In section 5, we discuss conditional sampling issues and how results obtained here related to previous evaluation studies over the Pacific and over the SO (e.g., A18; Zhao et al., 2020) in more detail.

2 Data and Methods

2.1 SOCRATES Flights and In situ measurements

During SOCRATES, the GV was equipped a suite of instruments measuring aerosol, cloud and thermodynamic variables. A total of 15 research flights were flown over the Southern Ocean, which are marked by the black lines in Figure 1. Typically, the GV sampled clouds with several flight modules in each flight, which consists of a combination of ramp ascents and descents, as well as level (fixed altitude) legs above, below, and in cloud. Here we focus on vertical profiles for cloud microphysical properties constructed from flight segments where the aircraft completely ascended or descended through all low-altitude clouds, and (as detailed below) multiple low-level cloud layers were occasionally present.

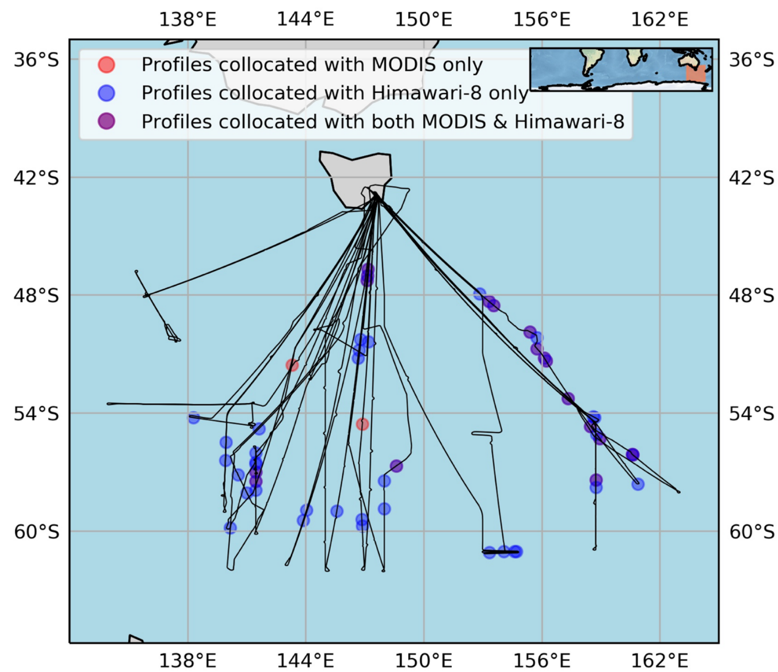

Figure 1. GV aircraft trajectory (black lines) during the SOCRATES. Clouds vertical profiles collocated with MODIS retrievals are marked by red dots and Himawai-8 retrievals by blue dots, and both satellites by purple circles.

This study focuses on the cloud microphysical properties measured by the particle-sizing-instruments, as listed in Table 1. SOCRATES GV data are available via the Earth Observing Laboratory data archive (https://data.eol.ucar.edu/project/552), with links to specific datasets given in the Table 1. Here we reply primarily on (i) the Cloud Droplet Probe (CDP) - an optical instrument that measures the concentration of cloud droplets in 30 size bins ranging from 2-50 µm, by measuring the light forward scattered by individual cloud droplets as they pass through a laser beam oriented across the aircraft flight direction, (ii) the Two-Dimensional Stereo (2DS) probe - an optical array probe that records the images of hydrometeors using two orthogonal laser beams that cross in the middle of the sample volume and measures particle size based on the shadow (blockage of the lidar beam) for particles ranging from about 10 to 1280 µm with a 10 µm bin-width as they cross the optical array; (iii) the Two-Dimensional Cloud (2DC) probe – an optical array probe that measures particle ranging from 37.5 to 1612.5 µm with 25 µm bin-width. We combined droplet size spectra from the CDP with that from the 2DS (or 2DC) to calculate several microphysical properties. Details and uncertainties associated with instruments and the spectra-merging processes are discussed in Section 4.4.

Table 1 In Situ Instruments

Instruments	Method	Measurements	References
Cloud Droplet Probe (CDP)	Forward scattered light	Droplet diameter within 2-50 μm, 30 bins (1 μm bin-width for sizes < 14 μm; 2 μm bin-width for sizes >=16 μm)	Lance et al. (2010) https://data.eol.ucar.edu/dataset/552.002
Two-Dimensional Stereo probe (2DS)	2-D image	Droplet diameter within 10-1280 μm(10 μm bin-width)	Wu and McFarquhar (2019) https://data.eol.ucar.edu/dataset/552.047
Two-Dimensional Cloud optical array probe (2DC)	2-D image	Droplet diameter within 37.5- 1612.5 μm, 64 bins (25 μm bin-width)	Wu and McFarquhar (2019) https://data.eol.ucar.edu/dataset/552.046

Note: For spherical particles droplet size is diameter. All the data is available at 1Hz temporal resolution. CDP data is included in the SOCRATES Navigation, State Parameter, and Microphysics Flight-Level Data, and this study uses version 1.3 of this dataset. The version number of 2DC and 2DS is 1.1. 2DC data is not available for research flight RF02.

Microphysical properties are calculated from the combined CDP and 2DS (or 2DC) data. Specifically, effective radius (r_e) is computed as the ratio of the third to the second moment of a droplet size distribution

$$r_e = \frac{\sum_{i=1}^{N} {r_i}^3 \cdot n_i}{\sum_{i=1}^{N} {r_i}^2 \cdot n_i} \tag{1}$$

where r_e is the effective radius at a given time, r_i the droplet radius of each bin, n_i the droplet concentration(#/cm^3) per bin, and N the total number of the bins. Cloud droplet number concentration (N_d) is computed as

$$N_d = \sum_{i=1}^{N} n_i \tag{2}$$

Cloud optical depth (τ) is calculated by vertically integrating the volume extinction coefficient (β)

$$\beta = \sum_{i=1}^{N} \pi Q_e {r_i}^2 n_i \tag{3}$$

where the extinction efficiency Q_e is approximately 2.

Liquid water content (LWC) is calculated as

$$\mathrm{LWC} = \frac{4}{3}\pi\rho_w \sum_{i=1}^{N} {r_i}^3 \cdot n_i \tag{4}$$

where the ρ_w is the density of liquid water. We calculate LWP by integrating LWC from cloud base to the cloud top. Following Wood et al. (2011) and PZ11, cloud top and cloud base are defined as the highest and lowest altitude with LWC greater than 0.03 gm^{-3}. The calculated value of LWP is not sensitive to this threshold.

In order to identify the cases (vertical profiles) where the precipitation is present, we calculate LWP for the droplets with diameters larger than 50 μm from 2DS probe (i.e. the vertically integrated precipitation liquid water path) which we will denote as PWP. Following the definition of King et al. (2013), clouds profiles are categorized into three groups: non-precipitating (PWP ≤ 2 g m^{-2}), lightly-precipitating (2 g m^{-2} < PWP ≤10 g m^{-2}) or heavily-precipitating (PWP > 10 g m^{-2}). The phase of the clouds is discussed using the ice phase fraction μ_{ice} (Korolev et al., 2017), defined as μ_{ice} = IWP/(IWP+LWP), where ice water path (IWP) is the vertical integral of ice water content (IWC) obtained from the 2DS and only includes ice particles >= 200 um, as derived by Wu and McFarquhar (2019). Later in the article, we discuss the implications of this restriction.

Figure 2 shows the verticals profiles of in situ r_e, LWC and N_d as a function of the normalized height (that is the position within cloud normalized such that 1 is cloud top and 0 is cloud base). Here only profiles of single layered clouds are shown (meaning profiles with multiple low-level clouds layers are not included). The thin lines are from individual aircraft penetrations (dots shown in Figure 1), while the thick line and purple shading shows the average profile and standard error, respectively. The standard error is the standard deviation divided by the square root of the number of profiles and is provided to give a sense for the one-sigma (66%) uncertainty in the average.

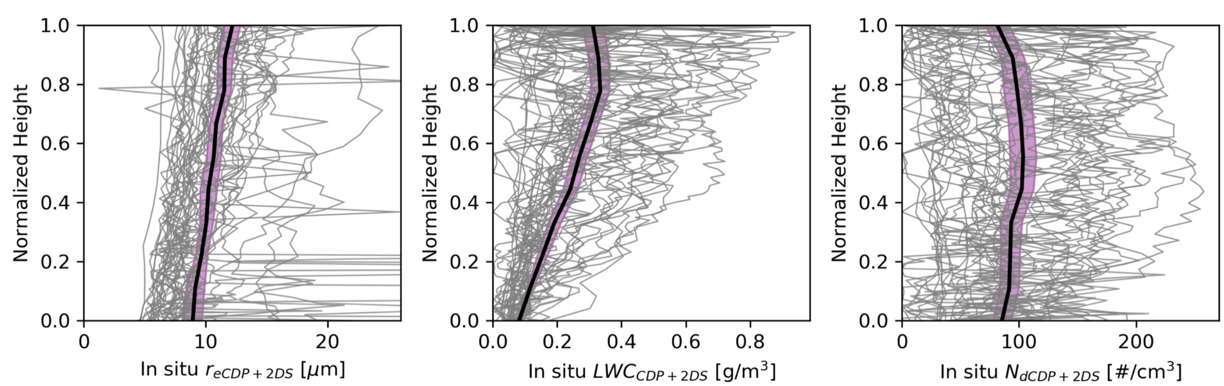

Figure 2. In situ profiles (single layered) of r_e, LWC and N_d normalized with respect to position within the cloud such that 1 indicates cloud top and 0 indicates cloud base. The thick black solid line is the mean of all the profiles. Purple shadings indicate the standard error. Spikes in the effective radius plot, in particular, occur when the CDP records only a few small (~ 10 μm) cloud-size droplets, but some precipitation size particles are present.

On average, both r_e and LWC increase roughly linearly with height, while N_d remains relatively constant with height, which is what one expects for cloud under an adiabatic assumption. We note that while r_e increases linearly with height, the value of r_e is not especially small at cloud base, and the total change in r_e (on average) is only a few microns. LWC near the cloud top deviates from a linear increasing LWC, which may be due to entrainment, but could also be due to the aircraft passing through clouds with a horizontally varying cloud top height or cases where a thin-cloud-layer exists above a thicker-layer that is not resolved by the aircraft sampling. The data used here are sampled at 1 Hz, which is roughly equivalent to a horizontal sampling distance of 137 m, while ascents and descents rates where typical about 5-7 m s^{-1} yielding a vertical resolution of about 6 m. The thickness of cloud layers varied from 88 to 2421 m, and with multilayer clouds often featuring multiple thin layers. The light lines in Figure 2 shows that individual aircraft penetrations do not always show an adiabatic-like profile, but of course the aircraft profiles we are creating are not necessarily sampling individual updrafts or downdrafts within a cloud, and any individual profile does not represent the actual profile of cloud properties at any specific location. Nonetheless the horizontal distance sampled by the aircraft is roughly consistent with the 1-to-few km pixel size being used in the satellite retrievals.

In Section 3, we compare satellite retrieved r_e, τ, LWP and N_d with in situ values from individual profiles. In that analysis, the in situ cloud spectra (drop size distribution) are first aggregated from the top of the cloud to the point where the optical depth reaches one (unless noted otherwise), and average spectra are used to compute r_e. Since both τ and LWP are integrated quantity, they are computed over the entire cloud layer by vertically integrating LWC and β, respectively. The mean value of N_d for each profile is used (rather than the value near cloud top) to reduce sampling uncertainty. In the plots in section 3, the variability in the in situ r_e is show by the standard deviations of the values over the top 1 OD of the cloud, and the variability in the in situ N_d is shown by the standard deviations of the values taken over the cloud profile. In order to estimate the uncertainty associated with the LWP and τ, we fit a set of lines to individual profiles that bound the vertical variations in LWC and β. Details are given in the supporting information.

2.2 Satellite Products and Collocation

MODIS level 2 Collection 6.1 cloud products from Aqua platform (MYD06) are used in this study. Detailed descriptions for the MODIS cloud product can be found in Platnick et al. (2017). Data is available at https://ladsweb.modaps.eosdis.nasa.gov/. The product includes cloud microphysical information with 1 km resolution (at nadir) based on a bi-spectral method using a non-absorbing visible-wavelength channel and one absorbing shortwave infrared channel, following the approach developed by Nakajima and King (1990). MODIS provides three set of retrievals, based on three different absorbing channels at 1.6, 2.1, and 3.7 µm. During the evaluation, we mainly focus on the retrievals using the 3.7 µm channel since r_e retrievals from this band are expected to be less influenced by 3D scattering effects (more on this later in the document) and often show the best agreement with the in situ measurements (e.g. King et al., 2013). Comparison between 1.6 and 2.1 µm are also provided and discussed in section 3.2. In general, the comparison of satellite retrieved r_e and in situ measured r_e requires consideration of the vertical penetration of the photons into the cloud. As reported in the previous studies (Platnick, 2000; King et al., 2013; Zhang and Platnick, 2011; Nakajima et al., 2010), one expect that r_e retrieval at 3.7 µm is more sensitive to the cloud droplets near cloud top due to the stronger absorption (smaller penetration depth), while r_e retrievals at 1.6 and 2.1 µm are more representative for the droplets deeper into the clouds due to the relatively weaker absorption (larger penetration depth). Thus (as described in section 2.1), we calculated in situ cloud top r_e from using cloud droplet spectrum averaged over 1 optical depth (OD) at the top of the cloud. We also have examined the impact of using a threshold of two and three optical depths but found this had little effect on the results. In addition to MYD06 product, we also used MODIS level 3 MYD03 product for the geolocation fields, and MODIS Level-1B data set MYD02QKM for the calibrated radiances to calculate the homogeneity index (section 4.1).

In addition to the operational MODIS retrievals, we also evaluated the CERES-MODIS cloud product retrieval Edition 4 (Trepte et al., 2019; Minnis et al., 2020). This retrieval product is produced by the CERES team at NASA Langley and is used in generating CERES radiative flux products (Kato et al., 2013). Although CERES-MODIS pixel level data are not publicly available (publically available data are limited to gridded level 3 products), we include in the supplementary material (Table S2) the mean of the CERES-MODIS retrievals collocated with the aircraft vertical profiles, which is used in all of the analysis presented here. While the microphysical properties of low clouds (which we examine in this article) are also based on the bi-spectral technique, the underlying codes were developed independently and apply different techniques to account for absorption due to above-cloud-water vapor and different criteria to identify low clouds and when to apply the bi-spectral retrieval. CERES-MODIS algorithm processes MODIS radiance data with every other scanline and every 4th pixel from the original MODIS 1 km resolution (i.e. 339 pixels per scanline, instead of 1354 pixels).

This study also evaluates cloud retrievals produced by the NASA SatCORPS group based on Himawari-8 observations (Trepte et al., 2019; Minnis et al., 2020). Data is available at https://data.eol.ucar.edu/dataset/552.027. Himawari-8 is a Japanese geostationary meteorological satellite launched in Oct. 2014. The SATCORPS Himawari-8 retrievals have 2-km resolution at nadir (at the equator) and are available every 10-minutes during GV aircraft flight dates. Details of the cloud property retrieval methodology are given in Minnis et al. (2011, 2008), and largely follow the approach used for CERES-MODIS and use near-infrared imagery at 3.9 μm.

All three satellite products provide retrievals for r_e and τ, based on 1D radiative transfer calculations and the satellite retrieved r_e and τ can be used to derive LWP. The formulation varies depending on the assumed vertical structure (profile shape) of the cloud LWC and r_e. For a vertically homogeneous cloud having a constant LWC and r_e with altitude, one obtains (Borg and Bennartz, 2007),

$$\mathrm{LWP} = \frac{4\rho_w}{3Q_e}\tau \cdot r_e \tag{5}$$

while for an adiabatically stratified cloud having a linearly increasing LWC and r_e with altitude (Wood and Hartmann, 2006) one obtains,

$$\mathrm{LWP} = \frac{10\rho_w}{9Q_e}\tau \cdot r_e \tag{6}$$

where ρ_w is the density of water, and again the extinction efficiency Q_e is about 2. These two expressions differ by a constant factor, with the vertically homogenous assumption giving a 20% larger LWP. Both MODIS and CERES-MODIS operational algorithm calculate and provide LWP based on equation (5), while SatCORPS Himawari-8 dataset include LWP calculated with the equation (6). Previous studies (e.g. King et al., 2011) have found that LWP computed using the vertically homogeneous formulation is usually positively biased for marine stratocumulus, which is not surprising given the overall adiabatic-like profiles of oceanic boundary layer clouds (Seethala and Horváth, 2010). We likewise find this assumption provides a better match with the observations, and in the later discussion use equation (6) assuming adiabatically stratified clouds except where specifically stated otherwise.

Although N_d is not provided in any of the satellite products, it can likewise be derived from passive satellite observations using r_e and τ assuming a one-dimensional cloud and following the assumption that clouds have an adiabatic-like profile, in which LWC increase linearly from cloud base to cloud top, given by (Bennartz et al., 2007; Grosvenor et al., 2018):

$$N_d = \frac{\sqrt{5}}{2\pi k}\left(\frac{f_{ad}c_w}{Q_e\rho_w}\right)^{1/2}\frac{\tau^{1/2}}{{r_e}^{5/2}} = C\cdot\frac{\tau^{1/2}}{{r_e}^{5/2}} \quad (7)$$

which is basically the product of the ratio of $\tau^{1/2}$ and ${r_e}^{5/2}$ and a constant *C*. The constant C is determined by several parameters, with $Q_e \approx 2$, and k, c_w and f_{ad} given by:

$$\mathrm{k} = \left(\frac{r_v}{r_e}\right)^3 \quad (8)$$

$$c_w = \frac{\rho c_p}{L_v}(\Gamma_m(T,P) - \Gamma_d) \quad (9)$$

$$f_{ad} = \frac{LWP}{LWP_{ad}} \quad (10)$$

The k parameter is a measure of the droplet spectrum width and is given by the third power of the ratio between volume radius (r_v) to r_e, where $r_v = \left(\frac{3LWC}{4\pi\rho_w N_d}\right)^{1/3}$. k is often assumed to be a constant with a value of 0.8 in retrievals. Later in this study, we calculate k values using the in situ r_v and r_e values for the cloud profiles. c_w is the rate of increase of LWC with height (i.e. the condensation rate) and is a weak function of temperature and pressure and is often assumed to be a constant range from 1 to 2.5 g m^{-3} km^{-1} (Albrecht et al., 1990; Min et al., 2012). Again, later in this study, we examine the mean and variability of c_w. c_p =1004 J K^{-1} kg^{-1} is the specific heat of dry air at constant pressure, L_v= 2.5 ×10^6 [J kg-1] is the latent heat of vaporization, Γ_d and Γ_m are the dry and moist adiabatic lapse rate, respectively. f_{ad} is called the adiabacity factor and describes how close the observed cloud is to a true adiabatic cloud layer (while still assuming the liquid water content increases linearly with height). Typically, this factor is assumed to be a constant 0.8 (e.g., Bennartz et al., 2007). Again, later in this study, we calculate f_{ad} values for the cloud profiles.

A key step in comparing satellite retrievals with in situ measurements is collocation. Here we averaged the satellite pixels around the location of the in situ profile within a 5 pixels × 5 pixels box for MODIS, 3 pixels ×3 pixels box for CERES-MODIS, and 3 pixels × 3 pixels box for Himawari-8. Changing the size of the averaging box by a factor of 2 has a negligible impact on the results. While in most cases, all of the satellite retrievals within these boxes correctly identified the cloud as low-level (<3 km) liquid clouds, in a few cases there were some scattered high clouds in the vicinity. In our box averages, we include only those satellite pixels which are identified as low-level liquid clouds by the retrieval algorithms. We did reject a few cases (match up points) because the apparent cloud-top-height (CTH) did not match the in situ aircraft measurement with in 1 km (satellite reported CTH > 3 km). While there are not a sufficient number of the poor CTH cases to quantify errors for these cases, we note that all of the satellite imager retrievals assume single layer clouds. Situations in which an optically thin high-altitude (ice) clouds overlays an optically thicker low-altitude (liquid) clouds is a long-standing problem for imager-based retrievals but filtering for CTH < 3 km appears to be satisfactory for the present analysis.

The in situ aircraft measurements and satellite retrievals do not necessarily occur simultaneously. In our analysis we account for the time offset by adjusting the box location for cloud advection, and we set a maximum time offset between the in situ and satellite data to be 1 hour for MODIS (and CERES-MODIS) and 10 minutes for Himawari-8. Specifically, we account for the distance clouds traveled using the in situ wind speed averaged near the cloud top, an approach which is similar to that employed by PZ11.

After the above filtering and processing, there remained 20 in situ cloud profiles (from 8 flights) closely aligned with Aqua MODIS overpasses, and 51 profiles (from 14 flights) closely aligned with Himawari-8 products. In total 53 in situ cloud profiles are used in this study and statistics are provided in Table.2. The circles marked on Figure 1 show the location of these profiles and Table S1 in the supporting information lists in situ properties for each profile.

Table 2. Summary of the Statistics for In Situ Measurements Used in the Satellite Evaluation

Variable	N	Mean	Standard Error
τ	53(21,18,14)	12.6(6.7,19.1,13.2)	1.3(1.5,2.4,2.1)
LWP [$g \cdot m^{-2}$]	53(21,18,14)	84.8(34.7,120.6,114.0)	9.1(7.0,13.3,18.4)
N_d [$\#/cm^{-3}$]	53(21,18,14)	86.8(91.4,109.7,50.3)	7.5(12.2,12.8,7.9)
r_e[μm]	53(21,18,14)	11.4(8.7,11.0,16.0)	0.6(0.4,0.4,1.3)

Note. N is the number of the data points. For each cell, value in front of parentheses is the statistics for all the collocated profiles, while the 1st, 2nd and 3rd value is that for non-precipitating, lightly-precipitating, and heavily-precipitating cases.

Table 3. Summary of the Comparison Statistics Between MODIS Retrievals and In Situ Measurements

Variable	N	R	Mean Bias	Mean	Standard Error
τ	20(6,7,7)	0.91	0.1(1.0,-0.6,0.1)	13.8(5.3,20.7,14.2)	1.0(1.6,1.5,1.8)
LWP [$g \cdot m^{-2}$]	20(6,7,7)	0.83	1.6(6.1,0.2,-0.8)	96.1(30.3,138.7,109.8)	8.5(9.1,9.4,20.9)
LWP_{vh} [$g \cdot m^{-2}$]	20(6,7,7)	0.82	16.1(10.4,20.9,16.1)	110.5(34.6,159.4,126.7)	8.8(9.4,10.1,21.3)
N_d [$\#/cm^{-3}$]	20(6,7,7)	0.68	-9.1(7.2,-32.8,0.6)	76.9(62.9,100.8,65.0)	8.3(8.1,15.4,12.0)
N_d_obs-mean-k_fad [$\#/cm^{-3}$]	20(6,7,7)	0.68	-8.1(8.0,-31.5,1.4)	77.8(63.7,102.0,65.8)	8.3(8.2,15.4,12.2)
N_d_obs_cbc_k_fad [$\#/cm^{-3}$]	20(6,7,7)	0.78	-7.2(0.8,-23.3,2.2)	78.8(56.5,110.2,66.5)	7.2(10.1,10.3,13.5)
$r_{e3.7}$ [μm]	20(6,7,7)	0.9	0.0(1.0,0.7,-1.6)	12.5(10.4,11.9,15.0)	0.5(0.3,0.3,1.0)
$r_{e2.1}$ [μm]	20(6,7,7)	0.83	0.7(2.4,1.0,-1.1)	13.2(11.9,12.2,15.4)	0.6(1.1,0.5,0.7)
$r_{e1.6}$ [μm]	20(6,7,7)	0.84	-0.1(0.8,0.6,-1.6)	12.5(10.3,11.8,15.0)	0.6(1.1,0.4,1.0)

Note. N is the number of the data points. R is the correlation coefficient. For each cell, value in front of parentheses is the statistics for all the collocated profiles, while the 1st, 2nd and 3rd value is that for non-precipitating, lightly-precipitating, and heavily-precipitating cases. LWP is calculated assuming adiabatically stratified cloud with equation $\mathrm{LWP} = \frac{10\rho_w}{9Q_e}\tau \cdot r_e$ for the satellite retrievals. $\mathrm{LWP_{vh}}$ is calculated assuming vertically homogeneous cloud with equation $\mathrm{LWP} = \frac{4\rho_w}{3Q_e}\tau \cdot r_e$. Satellite N_d is retrieved with typically assumed constants (k=0.8,f_{ad}=0.8), Nd_obs-mean-k_fad is retrieved by setting k and f_{ad} to the mean of the in situ values, and Nd_obs_cbc_k_fad by using case-by-case in situ value of k and f_{ad}. During the N_d, retrieval, condensation rate(c_w) is calculated using the satellite-retrieved cloud top temperature and pressure.

Table 4. Summary of the Comparison Statistics Between CERES-MODIS Retrievals and In Situ Measurements

Variable	N	R	Mean Bias	Mean	Standard Error
τ	20(6,7,7)	0.91	1.5(1.8,1.8,0.9)	15.2(6.1,23.2,15.1)	0.9(0.9,1.7,1.8)
LWP [$g \cdot m^{-2}$]	20(6,7,7)	0.79	12.2(12.2,25.3,-0.7)	106.7(36.4,163.8,109.8)	9.7(5.4,13.2,22.9)
$\mathrm{LWP_{vh}}$ [$g \cdot m^{-2}$]	20(6,7,7)	0.79	33.6(19.4,58.0,21.3)	128.0(43.6,196.5,131.8)	10.9(5.2,14.7,24.7)
N_d [#/cm^{-3}]	20(6,7,7)	0.49	-7.9(3.7,-41.5,15.7)	77.9(59.4,92.0,79.7)	10.7(3.8,17.3,19.0)
N_d_obs-mean-k_fad [#/cm^{-3}]	20(6,7,7)	0.49	-7.0(4.4,-40.4,16.7)	78.9(60.1,93.1,80.7)	10.7(3.8,17.3,19.2)
N_d_obs_cbc_k_fad [#/cm^{-3}]	20(6,7,7)	0.59	-5.9(-3.1,-33.4,19.1)	79.9(52.6,100.2,83.1)	10.1(6.0,12.4,21.2)
r_e[μm]	20(6,7,7)	0.78	0.2(1.5,1.5,-2.2)	12.7(10.9,12.6,14.3)	0.6(0.3,0.2,1.3)

Table 5. Summary of the Comparison Statistics Between SatCORPS Himawari-8 Retrievals and In Situ Measurements

Variable	N	R	Mean Bias	Mean	Standard Error
τ	51(19,18,14)	0.79	2.6(4.0,2.1,1.4)	15.7(11.1,21.2,14.7)	1.0(1.8,1.6,1.9)
LWP [$g \cdot m^{-2}$]	51(19,18,14)	0.64	16.1(21.8,21.1,2.0)	103.7(58.7,141.7,116.1)	8.7(9.4,11.4,24.5)
$\mathrm{LWP_{vh}}$ [$g \cdot m^{-2}$]	51(19,18,14)	0.64	36.9(33.6,49.4,25.2)	124.5(70.5,170.0,139.3)	10.1(11.9,13.3,28.0)
N_d [#/cm^{-3}]	51(19,18,14)	0.77	-1.6(6.3,-17.7,8.5)	87.0(102.9,92.1,58.8)	5.1(9.3,7.0,8.1)
N_d_obs-mean-k_fad [#/cm^{-3}]	51(19,18,14)	0.77	-0.5(7.5,-16.5,9.2)	88.0(104.2,93.2,59.5)	5.1(9.4,6.9,8.2)
N_d_obs_cbc_k_fad [#/cm^{-3}]	51(19,18,14)	0.77	0.1(4.4,-13.0,11.2)	88.7(101.1,96.7,61.4)	5.3(10.8,5.3,8.9)
r_e[μm]	51(19,18,14)	0.84	1.2(1.4,1.7,0.3)	12.3(9.9,12.4,15.6)	0.3(0.4,0.3,0.8)

3 In Situ and satellite retrievals comparisons

In this section, satellite retrievals from MODIS, CERES-MODIS and Himawari-8 are compared with the in situ measurements of τ, r_e, LWP and N_d. Statistics summarizing the comparison between the in situ and three satellites products provided in Table 3 to 5, respectively. We begin the analysis with τ and r_e, after which we focus on LWP and N_d, which are derived from τ and r_e.

3.1 Cloud optical depth

Figure 3 compares the in situ derived τ with satellite-retrieved τ. The vertical bars show the standard deviation of τ for the pixels within the collocated satellite match-up box that is used for averaging (see section 2). In many cases the vertical bars exceed 5, showing that there is typically a large horizontal variability in τ on the satellite pixel-scale. MODIS $\tau_{3.7}$ correlates well, R=0.91, with in situ values, having a mean bias of only 0.1. CERES-MODIS τ also correlates wells, R = 0.91, and mean bias of 1.5, which is not statistically different from zero at a 95% confidence level (as the one-sigma uncertainty in the mean, that is the standard error, is about 1). Himawari-8 τ is not as well correlated with R = 0.79 and a mean bias of 2.6, which is nominally significant at 95% confidence. However, the Himawari-8 data yield about the same as the mean bias as CERES-MODIS when restricted the 18 cases common to all three datasets (Figure 3d), with R = 0.86 and a mean bias = 1.88. Thus, the overall lower performance suggested by the full set of Himawari-8 match-ups is due to having more, and more difficult cases. In particular, there are more cases with multiple low-level cloud layers (that is, multiple layers below 3 km; grey filled dots) in the Himawari-8 set, and in general, cases which are more spatially variable (notice the larger vertical uncertainty bars in panel c). As will be discussed further in Section 5, a mean bias near 2.5 with a one-sigma certainty of near 1 is reasonably good performance and is consistent with expectations based on previous studies. Table 3 to 5 also lists the statistics for cases in different precipitation regime (non-precipitating, lightly-precipitating, and heavily-precipitating), and we will discuss these results in more detail in the context of the LWP retrieval in section 3.3, after examining the effective radius.

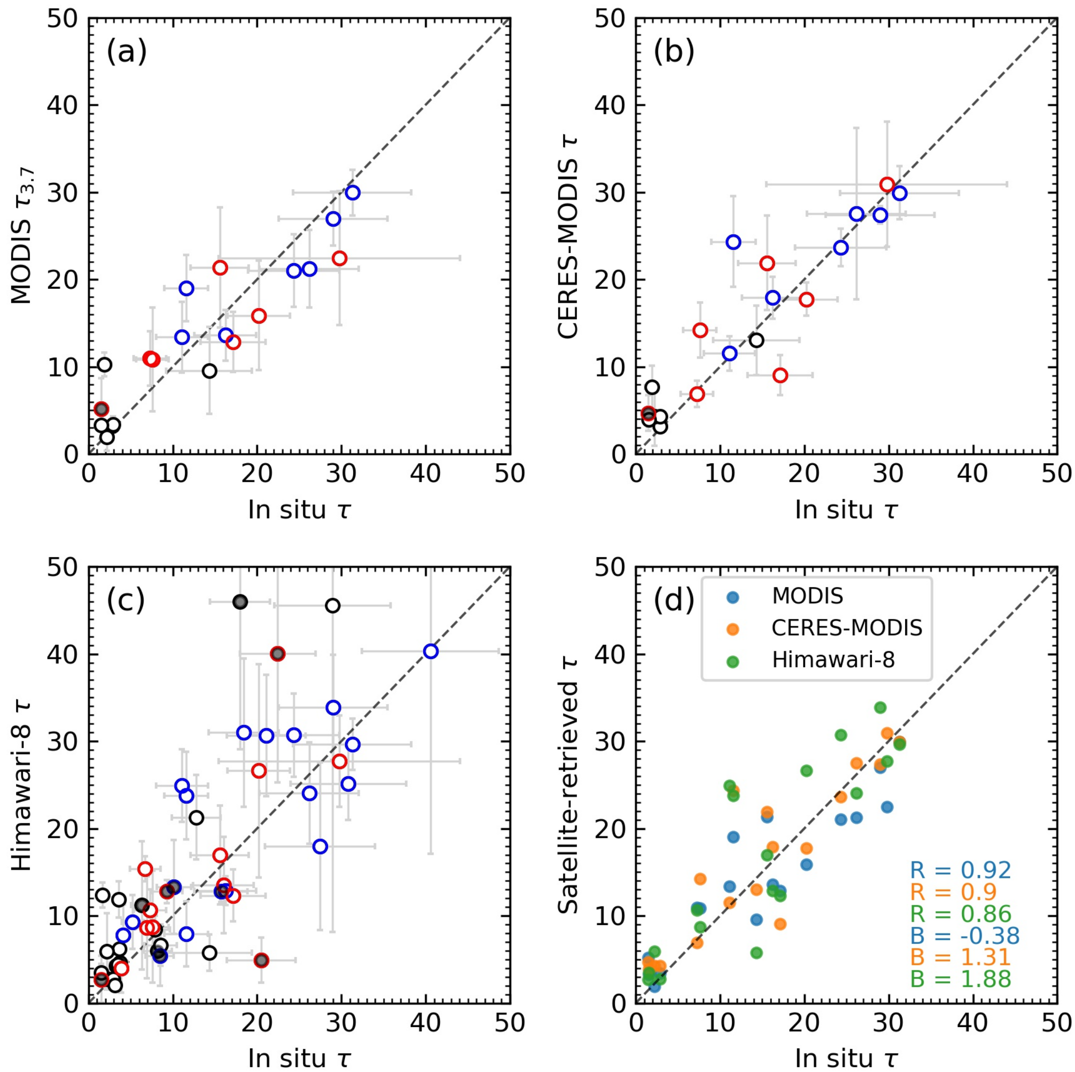

Figure 3. Comparison of τ from in situ measurements (CDP+2DS) and satellite retrievals for each case (aircraft vertical profile) based on (a) MODIS (MYD06 3.7 µm channel), (b) CERES-MODIS, (c) Himawari-8, and (d) for all three retrievals limited to the cases common to all three datasets. The vertical uncertainty bars indicate the standard deviation of τ within a box centered on the aircraft after correcting for advection (see text section 2).The horizontal uncertainty bars are estimated by fit a set of lines to individual profiles that bound the vertical variations in β. Black, blue, red open circles indicate cases that are non-precipitating (PWP < 2 gm^{-2}), lightly-precipitating (2 g m^{-2} < PWP <10 gm^{-2}) or heavily-precipitating (PWP > 10 gm^{-2}), respectively. Grey filled dots indicate those in situ profiles when there are multiple low-level cloud layers (cloud top of all layers is less than 3km). For text in panel (d), R indicates the correlation coefficient and B indicates the mean bias (satellite – in situ) for each dataset (of the specified color).

3.2 Effective Radius

The comparison between satellite derived r_e and in situ r_e is shown in Figure 4. Here the in situ r_e is derived from the merged spectrum of CDP and 2DS. MODIS $r_{e3.7}$ correlates well with in situ r_e (R=0.9) and has a mean bias of 0.0 μm. In spite of being for the same set of cases, perhaps surprisingly the correlation between CERES-MODIS and in situ r_e is not quite as good at, R=0.78. Nonetheless, the mean bias of CERES r_e is small at 0.2 μm and not significant different from zero at the 95% level of confidence. As for Himawari-8, the overall results are similarly good with the correlation between retrieved r_e and in situ r_e being 0.84, though the retrieved r_e are generally larger than in situ r_e, with a mean bias of 1.2 μm (which is significantly different from zero at 95% level of confidence). However, as was the case for optical depth, the difference in the Himawari-8 bias is due to additional cases analyzed, and the bias reduces to -0.29 μm when restricted to the set of cases common to all three retrievals (see Figure 4, panel d).

As shown by the red symbols in Figure 4, larger negative errors are associated with some heavily-precipitating cases (PWP > 10 gm^{-2}), while most non- and lightly-precipitating cases have a small positive bias. To demonstrate further how the error in r_e retrieval is related to the presences of precipitation, in Figure 5 the r_e retrieval error is plotted as a function of PWP. For MODIS r_e retrievals (panel a), there is a large negative bias associated with four cases, all of which have a PWP greater than 12 g m^{-2}. The same four cases are also negatively biased in CERES-MODIS and Himawari-8 retrievals. When consider all cases, there is more variability in the Himawari-8 retrieval error when PWP is greater than about 10 g m^{-2} than when PWP is less than about 10 g m^{-2}. The mean bias of Himawari-8 r_e for these heavily-precipitating cases is small, -0.3 μm while the bias for non- and lightly precipitation is 1.4 μm. When restricted to the 18 cases common to all three datasets, the mean bias for the non- and lightly precipitating cases is similar and statistically significant in all three datasets, with values of 0.78 μm, 1.52 μm, and 0.62 μm for MODIS, CERES-MODIS and Himawari-8 r_e retrievals, respectively. The presence of precipitation is clearly an important factor, and this will be explored in greater depth in section 4.

As mentioned in the section 2.2, MODIS r_e retrievals are also available based on observations at 1.6 μm and 2.1 μm in addition to 3.7 μm. The difference in the three MODIS r_e retrievals is influenced by the different absorption in different bands, with the photon penetration depth being largest at 1.6 μm and smallest at 3.7 μm. Figure 6(a) shows a comparison between all three MODIS r_e retrievals with in situ r_e. Both $r_{e2.1}$ and $r_{e1.6}$ correlate well with in situ r_e, with R=0.83 and R = 0.84, respectively, being slightly smaller than that of $r_{e3.7}$ (R=0.91). As is also shown in Figure 6, there is one case (marked by cross), with an unusually large difference among the three channels. This difference is likely due to the inhomogeneity of cloud scene, as will be discussed in Section 4.1. When this case is excluded, the mean bias in $r_{e2.1}$ and $r_{e1.6}$ (taken across all cases) is only 0.30 μm and 0.17 μm, respectively. However, as is the case for $r_{e3.7}$, there is marked variation with amount of precipitation. Similar to Figure 5, circles in Figure 6 (c) and (d) shows retrieval error of $r_{e2.1}$ and $r_{e1.6}$ as a function of PWP. Overall, a positive bias still exists for non- and lightly-precipitating cases, and the four cases associated with large negative bias in $r_{e3.7}$ (Figure 5a) continue to show a negative bias in $re_{2.1}$, though to a smaller extent.

To compare the MODIS r_e retrievals from different wavelength, Figure. 6(b) shows the difference between $r_{e3.7}$ and $r_{e2.1}$ (or $r_{e1.6}$) as a function of PWP. In general, $r_{e2.1}$ is larger than $r_{e3.7}$ with most points (orange circles) having positive difference (located above the zero line), and this positive difference become more obvious for the heavily-precipitating clouds. A similar positive difference is found for $r_{e1.6} - r_{e3.7}$, with $r_{e2.1}$ typically being closer to $r_{e3.7}$ than $r_{e1.6}$. As the amount of precipitation tends to increase with depth into the cloud, the increase in particle size for the precipitating cases is consistent with the expectation since photons at 2.1 μm can penetrate deeper into the cloud than at 3.7 μm. This does not explain, however, why $r_{e2.1}$ or $r_{e1.6}$ is larger for the non-precipitating cases (where one might expect the opposite behavior) suggesting that factors other than vertical variation, penetration depth and precipitation are important in the difference. This result is consistent with previous studies that shows $r_{e2.1}$ (or $r_{e1.6}$) tend to be larger than $r_{e3.7}$(e.g. PZ11; King et al., 2013).

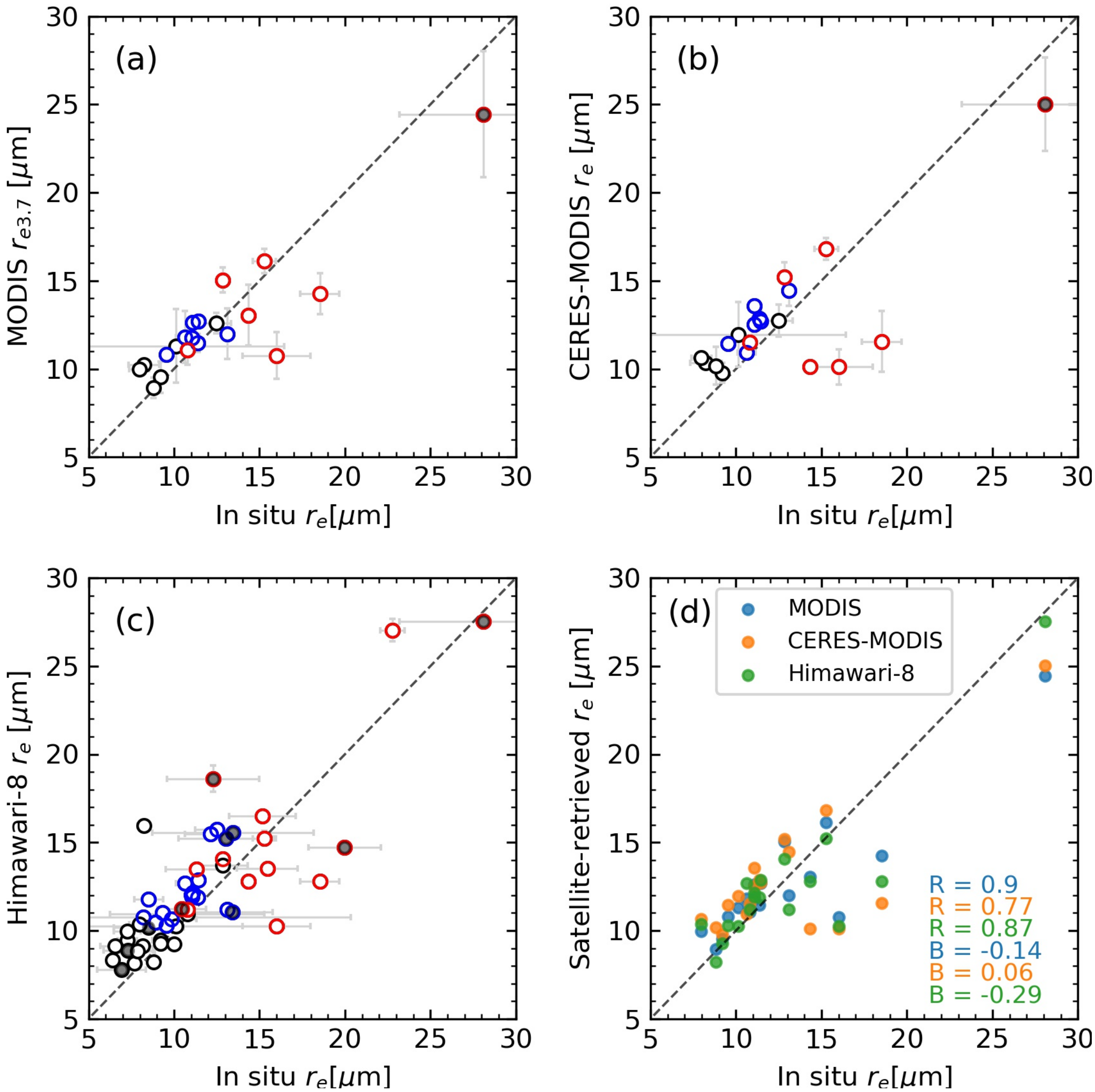

Figure 4. Comparison of r_e from in situ measurements (CDP+2DS) and r_e retrieved by (a) MODIS 3.7 μm channel, (b) CERES-MODIS, (c) Himawari-8, and (d) limited to the cases common to all three datasets. Symbols, vertical-uncertainty-bars and text-in-panel-(d) are the same as Figure 3. The horizontal-uncertainty-bars are the standard deviation near cloud top (see section 2).

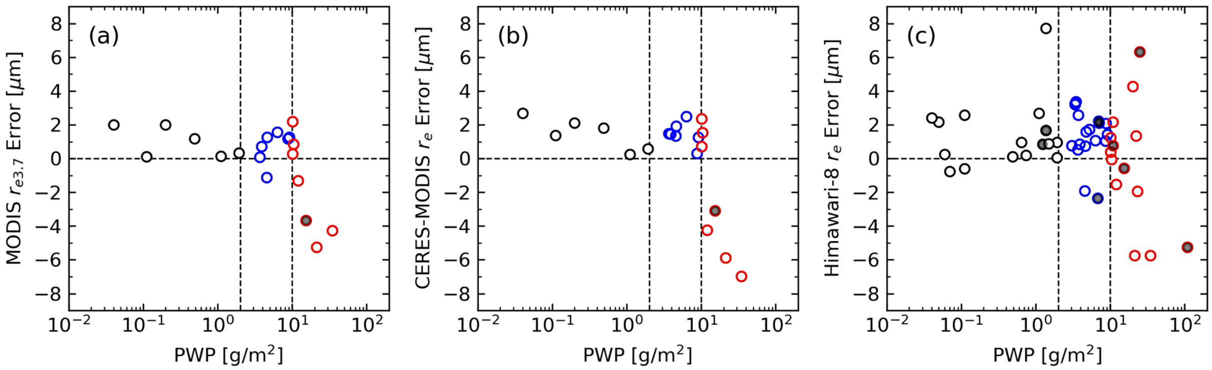

Figure 5. Satellite r_e retrieval errors as a function of vertically integrated precipitation water path (PWP). Symbols are same as that in Figure 3, with two vertical dashed line indicating the thresholds of 2 g m^{-2} and 10 g m^{-2} used to define non- and lightly-precipitating categories.

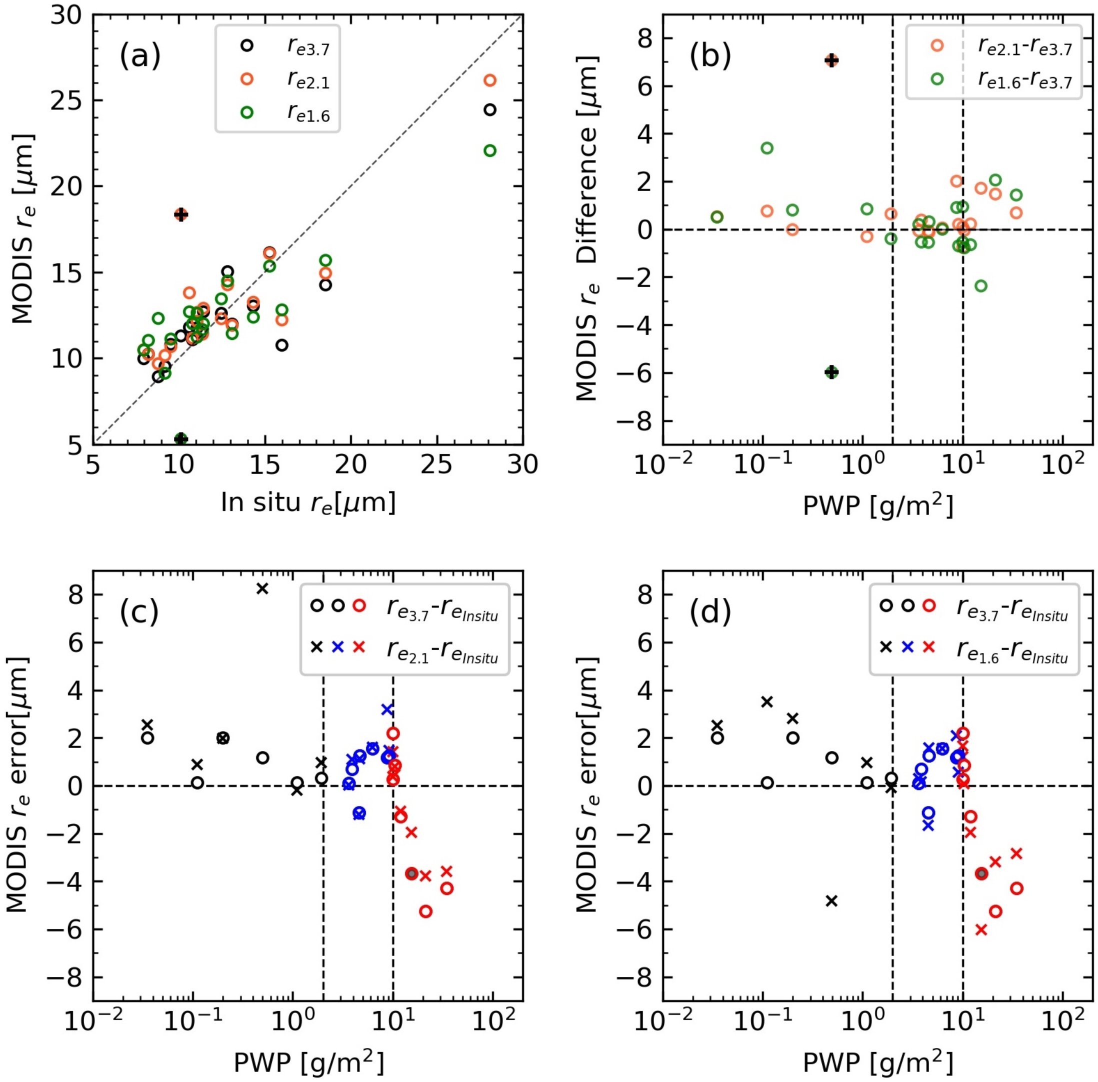

Figure 6. (a) MODIS r_e retrievals as three wavelengths versus the in situ r_e (CDP+2DS). Cross symbol in panel (a) denotes point with unusually large difference that is likely due to spatial heterogeneity (see text). (b) Difference between MODIS $r_{e3.7}$ and $r_{e2.1}$ (or $r_{e1.6}$) as a function of PWP. (c) MODIS $r_{e3.7}$ error (circles) and $r_{e2.1}$ error (x's) as a function of PWP. (d) same as panel c, except for x's are for $r_{e1.6}$ error. In (c) and (d), the color code is the same that in Figure 5 and earlier figures. The two vertical dashed lines in panels (b) and (c) denote the thresholds of 2 gm^{-2} and 10 gm^{-2} used to define non- and lightly precipitating categories.

3.3 Cloud Liquid Water Path

Figure 7 shows a comparison between in situ LWP and satellite derived LWP, calculated using equation (6), which assumes clouds are adiabatic. MODIS LWP correlates well with in situ LWP (R=0.83) and has a mean bias of only 1.6 g m^{-2}, while for CERES-MODIS, R=0.79 and the mean bias is 12.2 g m^{-2} (which is not significantly different from zero at the 95% level). For Himawari-8 using all 51 cases, R = 0.64 and mean bias is 16.1 g m^{-2} (not significant at 95%), with better performance for single layered cases in Himawari-8 retrievals (R = 0.8 and mean bias = 15.8

g m^{-2}), and with similar performance to MODIS when restricting to the set of cases common to all three satellite retrievals (panel d).

In the literature, satellite derived values for LWP are sometimes obtained by assuming a vertically homogeneous cloud (equation 5) rather than an adiabatic cloud (equation 6). Tables 3 to 5 provide mean bias and other error statistics using this alternative formulation. As one might expect given that the in situ profiles of LWC (see section 2) do show an adiabatic-like profile, the adiabatic formulation for LWP produces better overall results, whereas the vertically homogeneous assumption results in a statistically significant overestimation of LWP.

One expects a positive error in r_e or τ (meaning the retrieved value is too large) will result in a positive error in LWP (regardless of which of the two LWP formulations is used), and indeed we find this to be true, as shown in Figure 8. For all three satellite products, the bias in LWP is positively correlated with bias in r_e, with the R of 0.52, 0.69, and 0.59, respectively, and positively correlated with bias in τ, with the R of 0.92, 0.88, and 0.91. Note there are more black and blue points (associated with non- and lightly-precipitating profiles) in the upper right quadrant in Figure 8 in panels (b) and (c). In section 3.2 it was noted that non- and lightly-precipitating cases have a small positive (satellite > in situ) mean bias in r_e in all three retrieval datasets. Likewise, the optical depth for the non- and lightly-precipitating cases is also slightly biased in the CERES-MODIS and Himawari-8 datasets, as is evident in Fig. 8e and 8f which show fewer points in lower left quadrant than upper right quadrant (see also Tables 4 and 5). The positive bias in r_e and τ combine to create a small (but statistically significant) bias of 19.22 and 21.58 g m^{-2} in the LWP. In the operation MODIS MYD06 product, on the other hand, there are does not appear to be an LWP bias associated with non- and lightly-precipitating cases; and these points have a mean bias of only 2.93 g m^{-2}. This is because the bias in effective radius is countered by a small compensating error in τ of about -0.6 for MODIS for lightly-precipitating cases (note the points in lower left of Fig. 8d) The small bias of -0.6 is not itself statistically significant, as so it is ambiguous as to whether this compensation is coincidental. If coincidental, one expects that MODIS LWP would also have a small bias in LWP for non- and lightly-precipitating clouds given that it appears to have a similar bias in r_e, but all we can conclude is based on the data we have is that no bias in LWP.

While there is no statistically significant bias associated with the heavily-precipitating cases (red circles), there is considerable variability with these cases having largest positive and negative errors in r_e, τ, and LWP. The standard error (uncertainty in the mean) is greater than 20 g m^{-2} for the heavily-precipitating cases in all three datasets. In particular, the handful of cases identified as having large negative error in r_e (retrieved r_e is too small) have the largest underestimate in LWP.

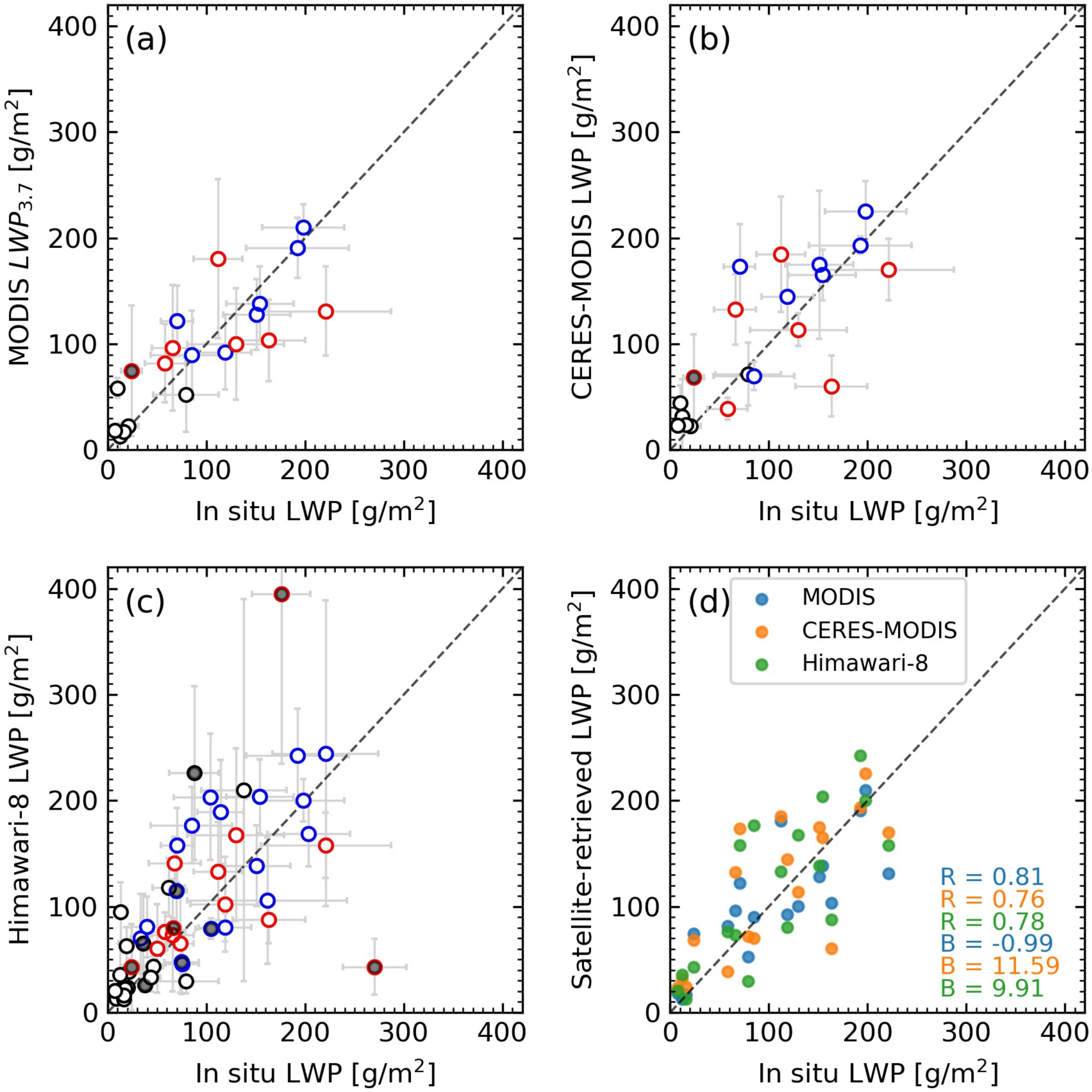

Figure 7. Comparison of LWP from in situ measurements (CDP+2DS) and retrieved by (a) MODIS 3.7 μm channel, (b) Himawari-8, (c) CERES-MODIS, and (d) limited to the cases common to all three datasets. LWP are retrieved from satellite assuming adiabatically stratified cloud. Symbols, uncertainty bars, and text in panel (d) are the same as that in Figure 4.

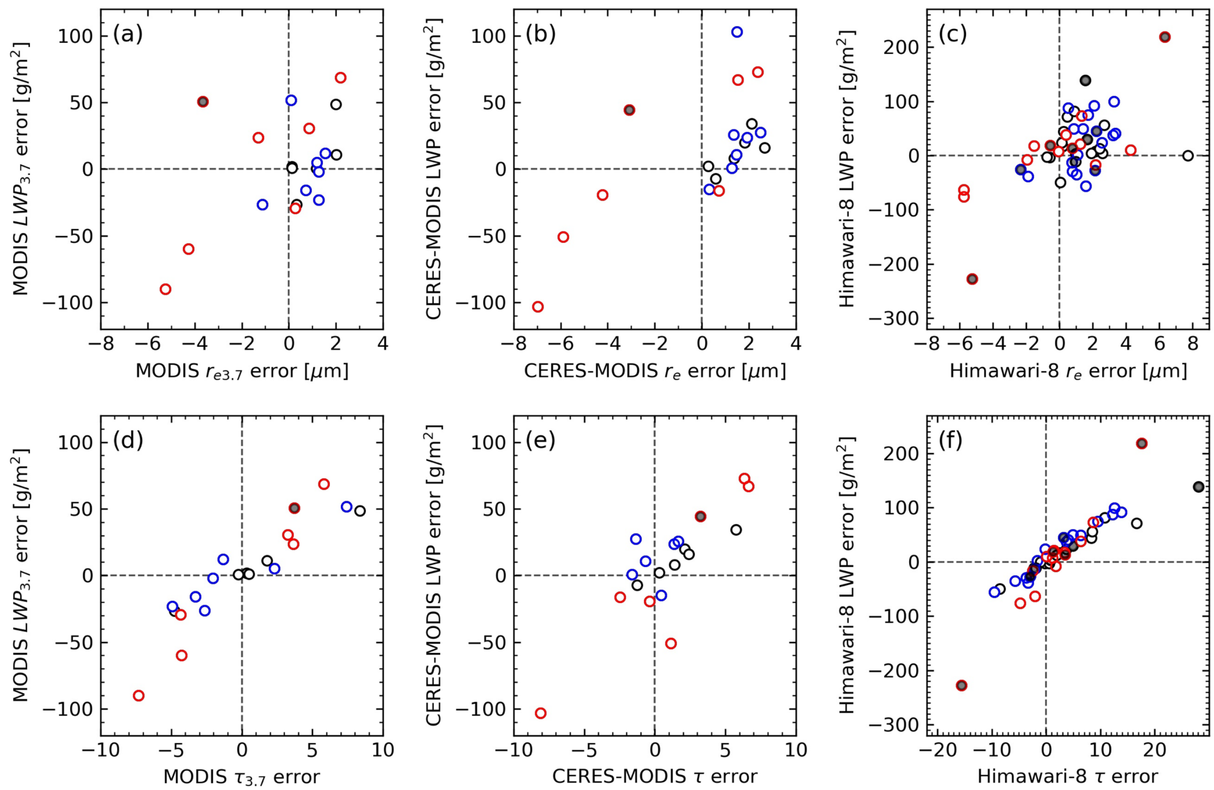

Figure 8. Difference between satellite derived LWP and in situ LWP as a function of retrieval error in r_e and τ. The 1st column is for MODIS, the 2nd column is for CERES-MODIS, and the 3rd column is for Himawari-8. Symbols are the same as Figure 3.

3.4 Cloud Droplet Number Concentration

Figure 9 compares the satellite derived N_d with the in situ values. When considering all comparison points (regardless of whether or not precipitation is present), the MODIS, CERES-MODIS and Himawari-8 N_d retrievals are biased by only -9.1, -7.9 and -1.6 #/cm^{-3}, respectively. These biases are not significantly different from zero at the 95% level of confidence and are small or modest relative to the overall mean of 86.8 #/cm^{-3} (Table 2). As was the situation for LWP (discussed above in section 3.3), the impact of precipitation on the bias in N_d retrievals is complicated by the correlation between errors in r_e and τ and is somewhat different in each of the three datasets and also depends to amount of precipitation present. In all three satellite datasets, the errors in r_e and τ tend to cancel out producing relatively little bias in N_d. The only statistically significant bias we find are for the lightly-precipitating category, where MODIS and CERES-MODIS retrievals have underestimated the N_d by about 30 to 40 #/cm^{-3} ,and Himawari-8 retrievals have underestimated the N_d by -17.7 #/cm^{-3} (from an overall mean of about 100 #/cm^{-3}). We note that the correlation between the retrieved and in situ values is poorer for N_d (ranging from 0.49 to 0.77) than for r_e, τ, and LWP. At the end of this section, we examine in more details the effect of random errors (variability from profile-to-profile) in the retrieved N_d.

Perhaps equally importantly, we find a large bias error in cases with multiple low-level clouds for Himawari-8, with a mean bias of 23.4 #/cm^{-3}. There are only 10 cases where multiple low-level clouds are present, but the difference is significant because these cases have smaller droplet concentration (mean value about 52.4 #/cm^{-3} with a mean-absolute deviation of 27 #/cm^{-3}). The MODIS and CERES-MODIS retrievals include only one such multilayer case, and we can't directly assess if the results would be the similar for multilayer cases for these two datasets but given the similar physical basis of the retrievals it seems likely that the MODIS-based retrievals would have similar difficulty. Unfortunately, it is difficult to identify when multiple low-level cloud layers are present from satellite VIS-IR imagery alone, however other measurements such as CALIPSO lidar backscatter might be used to detect the presence of such layers in combined retrievals algorithms. When multilayer clouds are removed from the set of cases examined the three datasets have similar mean biases of -9.7, -8.6, -7.7 #/cm^{-3} for MODIS, CERES-MODIS, and Himawari-8, respectively.

The above assessment for N_d is based on an assumed value for k of 0.8, f_{ad} of 0.8, and using c_w value calculated using equation (9) with satellite retrieved cloud top temperature and pressure. Using in situ measurements, f_{ad} can be calculated using equation (10). Doing so, we find a mean value of 0.74 for the 43 single layered profiles. Likewise, the k factor can be calculated using the SOCRATES data based on equation (8). The value for k is not generally constant over the depth of the clouds, but typically is larger toward cloud top because the droplet size distribution is narrower. In Figure 10, we plot histograms of the calculated k factors near cloud top (integrated extinction from cloud top less than 1) and for all vertical levels. The averaged k for cloud top is 0.76±0.08, which is slightly larger than averaged k for the whole cloud layer 0.73±0.09. Using both the mean cloud-top k of 0.76 and mean value f_{ad} of 0.74 has little net effect on the retrieval, with the resulting mean bias for N_d from MODIS, CERES-MODIS and Himawari-8 becoming -8.1, -7.0 and -0.5 #/cm^{-3}. We also find that, if one uses values of k and f_{ad} obtained from the in situ data on case-by-case basis for the N_d retrieval, there is likewise little change in the mean bias (-7.2, -5.9 and 0.1 #/cm^{-3}). The small net change in the bias occurs because that the impact of decreasing f_{ad} opposes (or compensates) for the effect of increasing k in equation (7). That is, what is important for the retrieval is the ratio sqrt(f_{ad})/k, which remains nearly constant and produces no net bias (systematic error) in the retrieval.

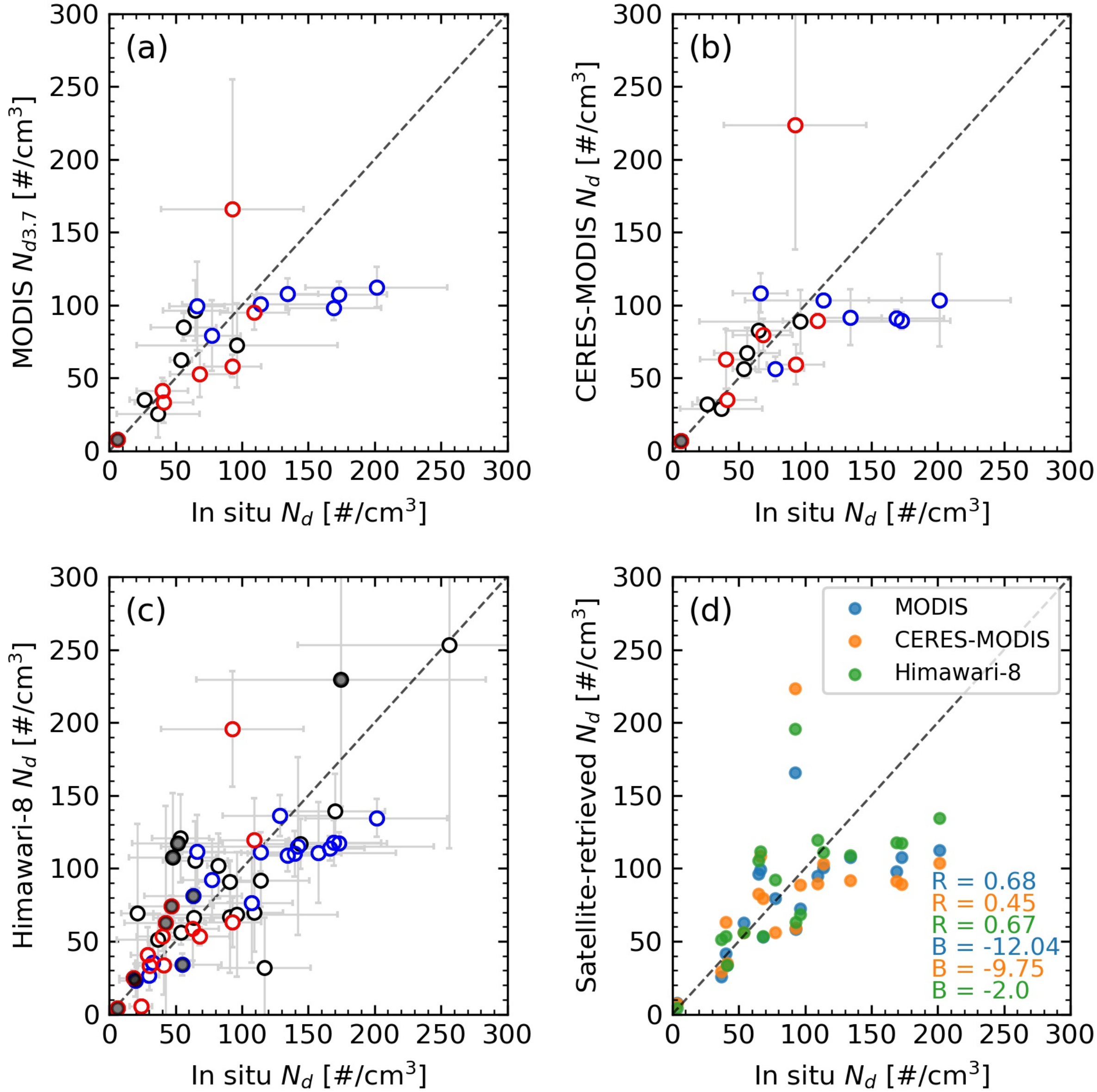

Figure 9. Comparison of N_d from in situ measurements (CDP+2DS) and retrieved by (a) MODIS 3.7 µm channel, (b) CERES-MODIS, (c) Himawari-8, and (d) limited to the cases common to all three datasets. Symbols, uncertainty bars, and text in panel (d) are the same as that in Figure 4.

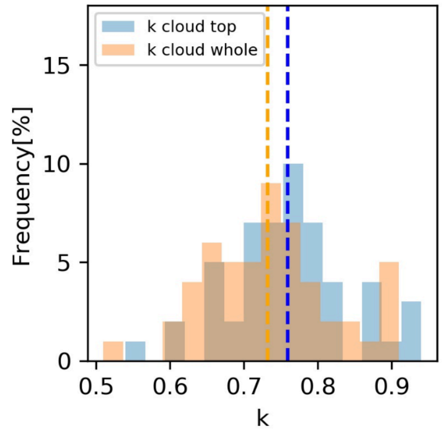

Figure 10. Histogram of k factor at cloud top (averaged over 1 optical depth) and averaged over whole cloud layer.

3.4.1 Uncertainty Analysis for N_d

Following Grosvenor et al. (2018) and Bennartz (2007), one can estimate the contribution of random errors (or uncertainty) in input variables in equation (7) to the random error in N_d, using a Gaussian error propagation formulation as shown in equation (11). The derivation assumes the input errors are normally distributed and uncorrelated with each other.

$$\left|\frac{\partial N_d}{N_d}\right|^2 = \left|\frac{1}{2}\frac{\partial c_w}{c_w}\right|^2 + \left|\frac{1}{2}\frac{\partial f_{ad}}{f_{ad}}\right|^2 + \left|\frac{\partial k}{k}\right|^2 + \left|\frac{1}{2}\frac{\partial \tau}{\tau}\right|^2 + \left|\frac{5}{2}\frac{\partial r_e}{r_e}\right|^2 + \text{other} \qquad (11)$$

Here, the term "other" represents the contribution of additional error sources other than input variables, which we neglect here, see Grosvenor et al. (2018) for additional discussion. In short, the expected fractional error in N_d would be given by square root of the sum of the squares of the fractional errors in the input terms on the right-hand-side of equation (11). For each input variable, we have calculated the fractional error for the inputs using the case-by-case (profile-by-profile) SOCRATES single-layered collocated profiles. For example, for Himawari-8 we approximate ∂r_e as the standard deviation of (retrieved r_e – in situ r_e) which equals 1.93 µm and r_e as the mean in situ value of 11.73 µm, and so $\left|\frac{5}{2}\frac{\partial r_e}{r_e}\right|$ = 41.14%.

Table 6 lists the percentage fractional error for each term (not squared) in the equation (11). Note that the column $\left|\frac{\partial N_d}{N_d}\right|$ given here is calculated from the data (same as the other columns) not calculated based on equation (11), while $\left|\frac{\partial N_d}{N_d}\right|_{calc}$ is calculated based on equation (11) with terms on the right-hand-side of equation (11) as input values. As one might intuitively expect from equation (7) and (11), errors in N_d are sensitive to changes in r_e, since r_e is raised to the power of 5/2. Our estimates show that error in r_e is the largest source for N_d error, with highest relative error contribution, followed by error in τ. As for assumed constants, variability in c_w, k and f_{ad} can also contribute to N_d error but based on variability observed during SOCRATES the impact is smaller than that of r_e, though we note the SOCRATES samples data are limited to summertime stratocumulus. One might notice that the sum of the expected percent fractional error doesn't "add up" to the $\left|\frac{\partial N_d}{N_d}\right|$ calculated on a case-by-case basis. This is because there are correlations between error terms that are not considered in equation (11). Nonetheless, it seems safe to conclude that error in r_e have a relatively large impact on the uncertainty in the N_d retrieval as compared with other source, with a total (case-to-case) uncertainty between about 40 and 55%.

Table 6. Expected Percent Fractional Error (uncertainty) in N_d due to Contributions from Different Sources

	$\left\lvert\frac{\partial N_d}{N_d}\right\rvert$	$\left\lvert\frac{\partial N_d}{N_d}\right\rvert_{calc}$	$\left\lvert\frac{1}{2}\frac{\partial c_w}{c_w}\right\rvert$	$\left\lvert\frac{\partial k}{k}\right\rvert$	$\left\lvert\frac{1}{2}\frac{\partial f_{ad}}{f_{ad}}\right\rvert$	$\left\lvert\frac{1}{2}\frac{\partial \tau}{\tau}\right\rvert$	$\left\lvert\frac{5}{2}\frac{\partial r_e}{r_e}\right\rvert$
MODIS	41.93%	48.17%	2.83%	9.68%	17.35%	14.99%	41.14%
CERES-MODIS	54.18%	63.4%	2.41%	9.68%	17.35%	14.71%	58.33%
Himawari-8	36.6%	55.9%	1.59%	10.72%	16.11%	22.05%	47.56%

4 Error Analysis

Satellite imager retrievals examined in this article invoke several assumptions about cloud structure and microphysics, and errors are likely to arise when these assumptions are violated in the real world. In this section, we focus on errors in the effective radius retrieval, which are arguably the most statistically robust errors identified in section 3, to assumptions in the bi-spectral retrieval, as well as examine some uncertainties in our analysis approach. Specifically, in section 4.1 we examine errors related to the to the assumption of horizontally homogeneous (i.e. plane-parallel or 1D) clouds. The bi-spectral retrieval also assumes the shape of the cloud droplet size distribution (DSD) can be represented by a simple function with a single mode. In the case of the MODIS, CERES-MODIS and Himawari-8 bi-spectral retrievals examined in this article, a modified gamma distribution with a fixed effective variance is assumed. Larger liquid droplets absorb more SWIR radiation than smaller droplets, and at its core, the bi-spectral technique is using the difference in absorbed radiation (between the visible and SWIR) to determine particle size. In simple terms, the larger droplets are (on average), the larger the absorption is, and the smaller the ratio of SWIR reflectance to visible wavelength becomes. The retrieval therefore also has some sensitivity to the width of the DSD. In sections 4.2, we show that when there is large contribution from larger precipitating droplets near cloud top, these cases are associated with significant underestimate in the effective radius, and in section 4.3 we examine errors associated with the assumed width for the size distributions for the non- and lightly precipitating cases. Last in section 4.4, we discuss uncertainties related to the in situ probes and analysis technique.

4.1 Horizontal inhomogeneity

The bi-spectral retrieval technique (at least as originally developed and applied here to MODIS and Himawari-8 observations) assumes clouds are horizontally homogeneous (i.e. plane-parallel or 1D). Of course, in reality the cloud fields often exhibit significant horizontal variability, and the breakdown of the 1D assumption can lead to systematic errors during the retrieval (e.g. Marshak et al., 2006; Zhang et al., 2012). To assess the impact of horizontal inhomogeneity on the retrieval error, we examine the relationship between heterogeneity in the satellite visible imagery and errors in effective radius using the H_σ index, defined as (Liang et al., 2009)

$$H_{\sigma,\lambda} = \frac{stdev[R(\lambda)]}{mean[R(\lambda)]} \tag{12}$$

which is the ratio of the standard deviation to the mean of the reflectance within the domain. For MODIS, we calculated H_σ using the MODIS (MYD03 product) radiance at 0.86 µm for the same 5 × 5 pixel analysis box used in the comparisons in section 3. Similarly, we calculate H_σ for Himawari-8 reflectance at 0.8 µm for using the same 3 × 3 pixel analysis box. The MODIS radiance is observed at 250m (nadir) resolution at 0.86 µm, which is finer than the 1 km grid used for the MODIS cloud property retrievals. The results shown here are based on the 250 m data, but we find our results do not differ appreciably if the radiance data is first reduced to 1 km resolution. Our adoption of this metric stems from previous research suggesting that clouds with H_σ <0.3 are sufficiently homogeneous that errors due to 1D assumption are likely small is this situation, while larger values associated with more heterogenous cloud fields have significant retrieval biases (Zhang and Platnick, 2011; Zhang et al., 2012). Figure 11 shows the error in the retrieved r_e from MODIS and Himawari-8 as a function of H_σ. Overall most points have a value for H_σ smaller than 0.3, and there is no clear dependence in the biases for these points. However, there are a few points with H_σ>0.3. For the one case with H_σ~0.7, there is a large difference in the three r_e retrievals from MODIS (based on different SWIR bands) which motivated us to remove this point from the analysis in Section 3.2. For this heterogenous point, the MODIS 3.7 µm band retrieval has the least error, which is consistent with the analysis in Zhang and Platnick (2011) and other studies that have suggested that this band is less susceptible to 3D scattering effects. For Himawari-8, there are two cases with H_σ > 0.5 that show relatively large error in r_e. Overall, most of the cases we evaluated are relatively homogenous with no dependence on H_σ, which suggests that horizontal heterogeneity is not a dominant source of r_e error for our evaluation result. We also examined whether errors in retrieved τ show any dependence on H_σ, since previous studies suggested that the retrieved τ can be smaller than the actual τ due to heterogeneity (Grosvenor et al., 2018). We found that retrieved τ error likewise shows no clear dependence on H_σ for our cases (figure not shown).

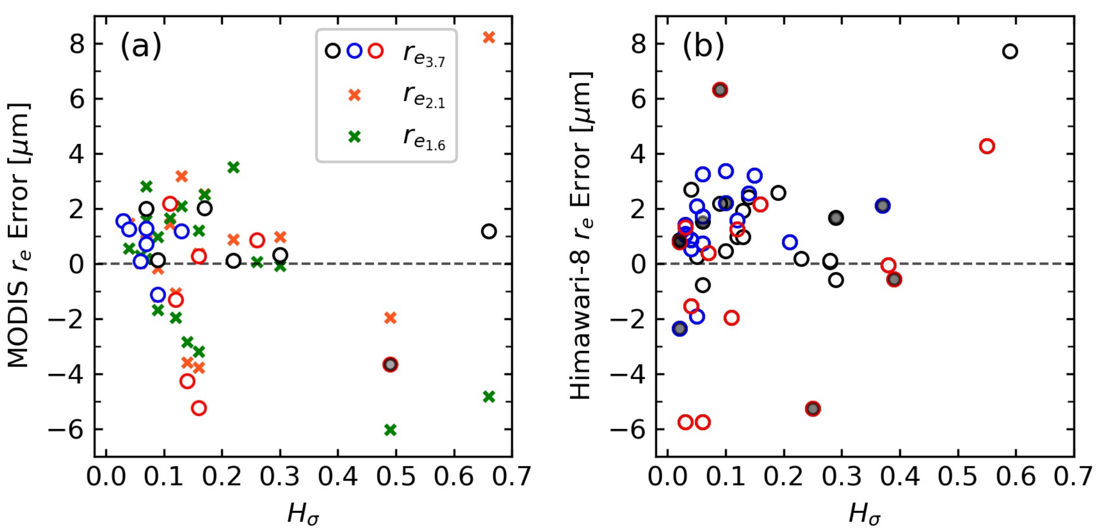

Figure 11. Satellite retrieval r_e error as a function of horizontal inhomogeneity index for (a) MODIS and (b) Himawari-8. Symbols are the same as Figure 3. The additional x-symbols in (a) represent the error of MODIS $r_{e2.1}$ and $r_{e1.6}$.

4.2 The Presence of Drizzle at Cloud Top

The presence of drizzle can significantly impact the r_e retrieval. Minnis et al. (2004) and Zhang (2013) show that the presence of drizzle drops can result in underestimation in retrieved r_e. In our study, section 3.2 we find that r_e is underestimated for some (but not all) heavily-precipitating cases. To further assess the contribution of the droplets larger than 50 µm, we calculated the ratio of mean LWC over the top 1 OD of the cloud for droplets with diameters > 50 µm (i.e. precipitation water content, PWC) and droplets with diameters < 50 µm (i.e. cloud water content, CWC). Figure.12 shows difference between satellite retrieved r_e and in situ r_e as a function of this ratio PWC / CWC.

For the simplicity, only relatively homogeneous cases with H_σ <0.3 are considered here. Most of the cases have a ratio < 0.1, which means the contribution from larger drizzle mode is small. Underestimation of r_e was found for three heavily cases with large contribution from drizzle particles (ratio > 0.2). This demonstrates that it is not the presence of drizzle in the cloud (characterized by PWP), but the presence of drizzle near cloud top that is important. We note that if we ignore particle larger than 50 µm, and calculate in situ r_e only from the CDP, the difference between the satellite retrieved r_e and in situ r_e (showing as crosses in Figure 12) are smaller for these three heavily-precipitating cases, but the satellite retrieved r_e still shows underestimation, especially for two of the cases, demonstrating that this effect is not an artifact resulting from the merging of the CDP and 2DS (more on this in section 4.4).

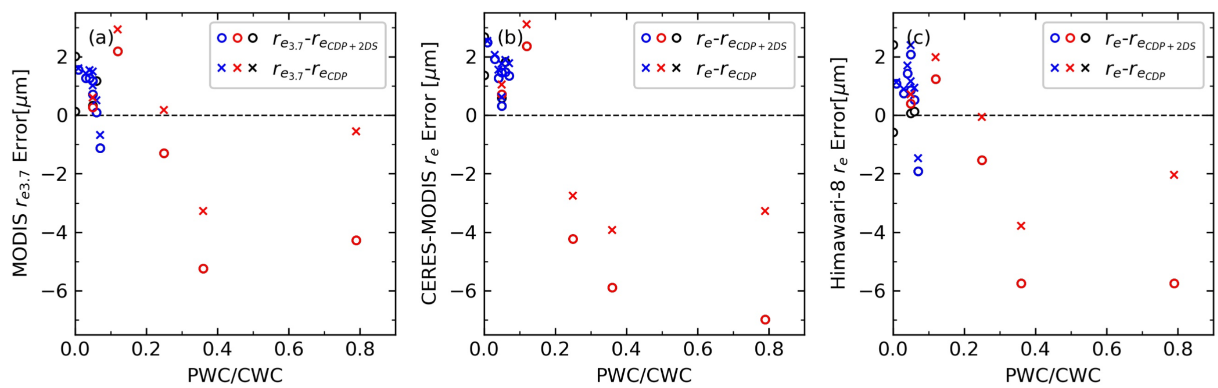

Figure 12. (a) MODIS $r_{e3.7}$ error as a function the ratio of mean LWC over the top 1 OD of the cloud for droplets with diameters > 50 µm (i.e. precipitation water content, PWC) and droplets with diameters < 50 µm (i.e. cloud water content, CWC). (b) and (c) are the same as (a) except for CERES-MODIS and Himawari-8 r_e. Only cases with H_σ <0.3 are considered here. Colors and symbols are same that in Figure 5, with open circles representing the difference between satellite retrieved r_e and in situ value calculated using merged DSD from CDP and 2DS, while cross represent the difference between retrieved r_e and in situ r_e calculated using the CDP only.

4.3 Droplet size distribution width (for non- and lightly-precipitating clouds)

Satellite retrievals algorithms typically make assumptions regarding the shape of cloud droplet size distribution (DSD). The MODIS, CERES-MODIS, and Himawari-8 retrievals examined here assume a modified gamma distribution which can be written as (Hansen, 1971)

$$n(r) = N_0 r^{(1-3v_e)/v_e} e^{-r/(r_e v_e)} \quad (13)$$

where r is the droplet radius, N_0 is a constant, and v_e is effective variance given by (Hansen, 1971)

$$v_e = \frac{\int_0^\infty (r - r_e)^2 \pi r^2 n(r) dr}{{r_e}^2 \int_0^\infty \pi r^2 n(r) dr} \quad (14)$$

For the gamma distribution one can show that k = $(1-v_e)(1-2v_e)$. Thus, the width of the DSD can be assessed using v_e or k factor. In the retrievals, MODIS assumes a modified gamma distribution with a fixed variance v_e of 0.1 (Platnick et al., 2016), as do CERES-MODIS and Himawari-8 (W. L. Smith, personal communication, 2020). $v_e = 0.1$ corresponds to k =0.72. Of course, the actual DSD may not be well approximated by a gamma distribution with v_e of 0.1 and this will impact the retrieved r_e (Arduini et al., 2005).

In order to explore the width of the cloud DSD with respect to precipitation amount, we plot the in situ estimated k factor as a function PWP in Figure 13(a). Here k is calculated using equation (8) and no assumption regarding the shape of the DSD is made. Consistent with our earlier analysis and focus on values need for retrieval, here the k factor is determined average taking over the region at the top of the cloud corresponding to an optical depth of 1, and for simplicity, we only consider single layered clouds. The k factor tends to decrease (the distribution becomes broader) with increasing PWP. The mean k factor for non-precipitating, lightly-precipitating and heavily-precipitating cases is 0.80, 0.77 and 0.70. In particular, the observed DSD width of the non-precipitating and lightly-precipitating cases is narrower than the assumed value (that is, k is greater than 0.72). We likewise calculated v_e for those non-precipitating and lightly-precipitating cases using equation (14) with the cloud DSD from CDP probe averaged over 1 optical depth. The mean value of v_e is for these non-precipitating and lightly-precipitating cases is about 0.068, which is narrower than the assumed v_e =0.1 in the retrieval.

We discussed the impact of bias and uncertainty in the k factor on N_d in section 3.4. A quantitative assessment of the impacts of uncertainty (or bias) in k (or v_e) on the r_e retrieval is more difficult and arguably requires detailed radiative transfers calculations using a variety of values for v_e. However, we can gauge the impact of the droplet width on r_e retrieval based on result published by PZ11, who examined the impact of the distribution width on the r_e retrieval using a log-normal σ (σ_{log}). We estimated σ_{log} of the in situ measured DSD using a least squares minimization. We opted to use a minimization approach in order to obtain a best fit for a log-normal distribution to the bulk of the observed cloud particles, and to minimize the impact of unusually small or large particles (outliers in the data), which we found to significantly broaden the estimated σ_{log}. Details are given in the supplementary material.

As shown in Figure 11b, σ_{log} is negatively correlated with k, because broader DSD means smaller k and larger σ_{log}, while a narrower DSD means larger k and smaller σ_{log}. The σ_{log} for non-precipitating and lightly-precipitating of single layered cases averaged over the top 1 OD is 0.16. PZ11 undertook radiative transfer simulations to understand how the retrieved r_e is impacted when the true value $\sigma_{\log}$ is smaller than the value assumed in the radiative transfer calculation. They found that when actual $\sigma_{\log}$ is smaller than the assumed value of 0.35 (equivalent to $v_e = 0.1$), the retrieved r_e is also larger, and retrieved r_e would be overestimated (biased high). Specifically, PZ11 compared the retrieved r_e assuming $\sigma_{\log}$ =0.35 and 0.2, and found retrieved was r_e overestimated by as much as 0.58 μm. This result is broadly similar to result published by Chang and Li (2001) who found that a change of ±0.15 in $\sigma_{\log}$ resulted change of about ±1 μm in the mean of the r_e retrievals (starting from a nominal value of 0.35 for $\sigma_{\log}$ with $r_e = 10$ μm).

We concluded, therefore, that much of the positive-bias in effective radius for the non- and lightly-precipitating cases (shown in section 3.2 to range from 0.5 to about 1.0 μm) is likely due to having an assumed effective variance that is a bit too large, or stated more generally, an assumed DSD in the retrieval which is too wide for the SO clouds observed during SOCRATES. As a caveat, we note that the solar and view geometries in the present case are not identical to those in previous studies that examine the width of the DSD and its impacts on the retrieval. We do not expect this is a significant factor for the solar and view geometry during SOCRATES, as the experiment took place during the Southern Hemisphere summer primarily in the afternoon when the sun is reasonably high with a solar zenith angle less than 60°. Nonetheless, the above conclusion should perhaps be quantified using full radiative transfer calculations for the precise conditions observed during SOCRATES, and more generally evaluated over the range of solar and view geometries encountered over the Southern Ocean to more fully assess the impact, especially as regards possible seasonal impacts. Such is beyond the scope of the present study, and is left as a topic for future work.

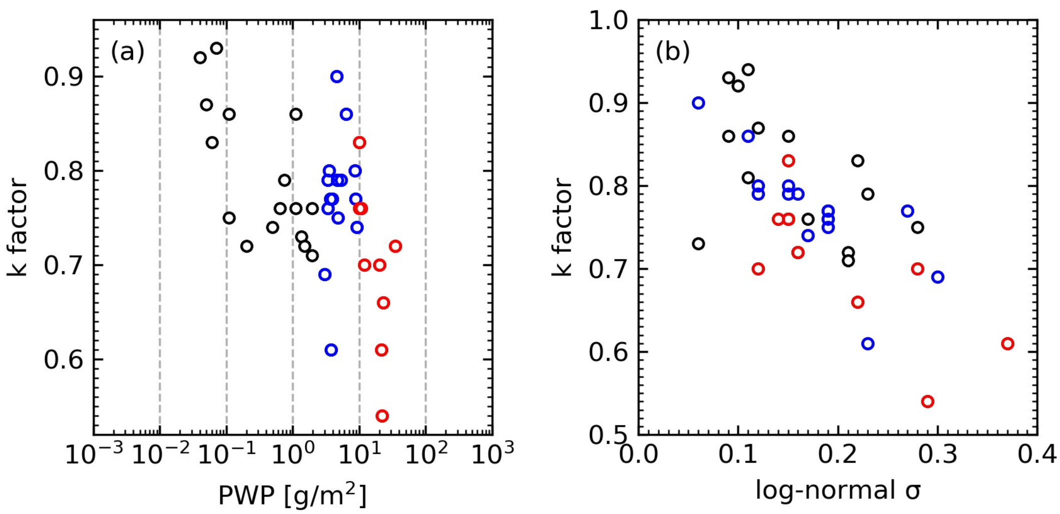

Figure 13. (a) k factor as a function of precipitation water path (PWP); (b) Scatter plot between log-normal σ and k. Only single layered clouds are shown here.

4.4 Uncertainty due to instrumentation

In most of the preceding analysis we calculated in situ r_e from the DSD obtained by merging measurements from the CDP and 2DS. Specifically, we used all the CDP bins (which includes particles up to 50 μm) and combine it with the DSD from the 2DS for bins larger than 50 μm, the same approach as used by King et al. (2013). We have also explored merging the CDP and 2DC (as was used by PZ11), as well as a second alternative (ALT) approach for merging the CDP and 2DS, in which we use the DSD from CDP for bins smaller than 25 µm, the DSD from the 2DS for bins larger than 50 µm, and use the larger values between the two probes for the intermediate bin (25 to 50µm). Figure S3 in the supplementary material shows an example of the CDP, 2DS and 2DC spectra and the result merged DSD.

Table 6 along with Figures 14 and 15, summarize the impact of using different probes or the merge approach has on the in situ r_e and estimated error in the satellite retrieved r_e. Since 2DC probe is not available for research flight RF02, only data from other flights are considered. For the 19 profiles available for MODIS, mean in situ r_e calculated using different probes or merging methods varies from 11.54 µm with the CDP only to 13.2 µm with CDP+2DS (ALT). To visualize the difference of in situ $r_{e,}$ Figure 14 shows box plots of in situ r_e for cases collocated with different sensors. Overall, CDP+2DS (ALT) gives largest in situ r_e in all precipitation regime. In situ r_e from CDP+2DS (King) is smaller than CDP+2DS (ALT), because counts in the intermediate bin (25 to 50 µm) from the CDP are typically smaller than that from 2DS. In situ r_e from CDP+2DC tend to be smaller than that from CDP+2DS, and close to that from CDP only, as counts from 2DC bins are usually smaller than that from 2DS.

Naturally, the impact of using different probes or merge approach is much more important for the heavily precipitating cases than for the non- or lightly-precipitating cases. Nonetheless, even for the non- or lightly-precipitating cases, using the CDP+2DS (ALT) merging increases the r_e and can (at least partially) offset the estimated error (see Figure 15). Taking the MODIS $r_{e3.7}$ as an example, for light-precipitating cases the mean error in MODIS $r_{e3.7}$ is about 0.98, 0.98, 0.71, and 0.07 µm when compared with in situ r_e from CDP, CDP+2DC, CDP+2DS (King), CDP+2DS (ALT), respectively. Using CDP+2DS (ALT) effectively appears to eliminate the bias for the lightly precipitating cases, and the bias for non-precipitating cases, while not eliminated is reduced going from 1.27 µm from CDP+2DC only to 0.66 µm from CDP+2DS (ALT). However, the bias for heavily-precipitating cases gets worse, going from 0.34 µm estimated using CDP+2DC to -2.41 µm using CDP+2DS (ALT).

Thus, regardless of how we merged the CDP and 2DS data, there is a fundamental difference in the bias for the different precipitating categories. If one calculates the bias across all precipitating categories the CDP+2DS (ALT) formulation produces the smallest error but does so only by balancing the errors across the different categories. This same pattern is weaker in the CERES-MODIS and Himawari-8, but is qualitatively similar.

Past studies (e.g. King et al., 2013) suggest that the counts in the CDP below 50 µm are more reliable and we have therefore focused on using CDP+2DS (King) formulation in our analysis. But we note there is a measurement issue here that needs to be addresses for future field campaigns, specifically that efforts are needed to reduce the uncertainty in measured number concentration for particles between about 20 and 100 µm.

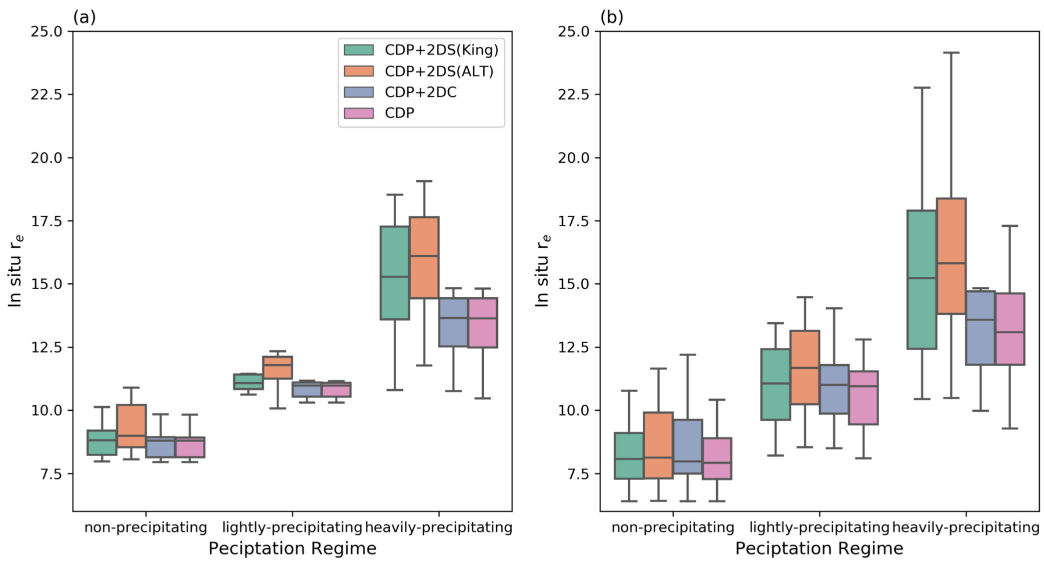

Figure 14. Box plots in situ r_e for cases collocated with (a) MODIS and CERES-MODIS (19 profiles), (b) Himawari-8. There are four kinds of in situ r_e obtained with different instruments and merging methods: CDP+2DS (King approach), CDP+2DS (Alternative approach), CDP+2DC, and CDP only. Since the 2DC probe is not available for research flight RF02, only data from other flights with collocated profiles are considered.

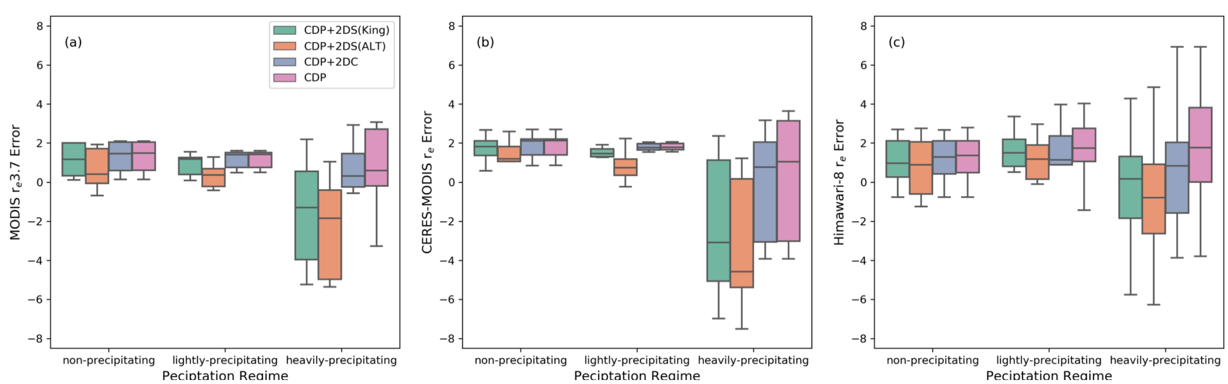

Figure 15. Box plots of error in satellite derived r_e from (a) MODIS (b) CERES-MODIS (c) Himawari-8 when compared with different in situ r_e. There are four kinds of in situ r_e obtained with different instruments and merging methods: CDP+2DS (King approach), CDP+2DS (Alternative approach), CDP+2DC, and CDP only. Since 2DC probe is not available for research flight RF02, only data from other flights with collocated profiles are considered.

Table 6. Statistics for r_e Using Different Probes or Merging Methods and Corresponding Estimates of Error in Satellite Retrieved r_e

	Different probes/methods			
	CDP+2DS (King)	CDP+2DS (ALT)	CDP+2DC	CDP
Mean r_e that collocated with MODIS[μm]	12.55 (8.87,11.17,16.55)	13.21 (9.33,11.81,17.37)	11.7 (8.73,10.9,14.62)	11.54 (8.72,10.9,14.18)
Mean r_e that collocated with Himawari-8 [μm]	11.48 (8.44,11.03,15.95)	12.01 (8.75,11.57,16.77)	11.58 (8.59,11.45,15.6)	10.61 (8.28,10.62,13.59)
Mean error of MODIS $r_{e3.7}$	-0.03 (1.13,0.71,-1.59)	-0.69 (0.66,0.07,-2.41)	0.82 (1.27,0.98,0.34)	0.98 (1.27,0.98,0.78)
Mean error of MODIS $r_{e2.1}$	0.74 (2.92,1.04,-1.11)	0.08 (2.45,0.4,-1.93)	1.59 (3.06,1.31,0.82)	1.75 (3.06,1.32,1.26)
Mean error of MODIS $r_{e1.6}$	-0.14 (0.79,0.65,-1.58)	-0.8 (0.32,0.01,-2.4)	0.71 (0.93,0.92,0.35)	0.88 (0.93,0.93,0.79)
Mean error of CERES-MODIS r_e	0.17 (1.71,1.47,-2.22)	-0.49 (1.25,0.83,-3.04)	1.02 (1.85,1.74,-0.29)	1.02 (1.85,1.74,-0.29)
Mean error of Himawari-8 r_e	0.91 (1.43,1.35,-0.31)	0.38 (1.13,0.81,-1.13)	0.81 (1.28,0.93,0.04)	0.81 (1.28,0.93,0.04)

Note: Since 2DC probe is not available for research flight RF02, only data from other flights with collocated profiles are considered here for comparison. For each cell, value in front of parentheses is the statistics for all the collocated profiles, while the 1st, 2nd and 3rd value is that for non-precipitating, lightly-precipitating, and heavily-precipitating cases.

5 Summary, Discussion and Conclusions

5.1 Summary

Satellite retrievals of cloud properties have been widely used to study clouds over the Southern Ocean, but our confidence in these retrievals has been limited by a lack of verification studies due in no small part to a lack of in situ observations. In this study, cloud properties observed from aircraft during the Southern Ocean Cloud Radiation Aerosol Transport Experimental Study (SOCRATES) in January and February of 2018 are used to evaluate retrievals of cloud microphysical properties for Southern Ocean stratocumulus based on the widely used visible-shortwave-infrared bi-spectral technique. In particular, three datasets are examined: (i) the Moderate Resolution Imaging Spectroradiometer (MODIS) level 2 (collection 6.1) cloud product, (ii) the CERES-MODIS edition 4 product, and (iii) the NASA SatCORPS Himawari-8 product. The analysis focused on the evaluation of retrieved cloud optical depth (τ) and effective radius (r_e), as well as liquid water path (LWP) and cloud droplet number concentration (N_d) which are derived from τ and r_e under the assumption that the cloud has a linear adiabatic-like profile of liquid water content.

Our analysis focused on the use of vertical profiles of cloud properties constructed from individual aircraft penetrations through the stratocumulus. Analysis of the cloud vertical structure shows that SO stratocumulus do have an adiabatic-like structure on average, with both r_e and LWC increasing roughly linearly with height, while N_d remains relatively constant with height. The stratocumulus examined were largely closed-cell (or at least overcast). For most of the aircraft profiles, collocated satellite imagery had a homogeneity index (equivalent to the fractional standard deviation of the cloud reflectance) of less than 0.3 and the retrievals show little evidence of error related to horizontal inhomogeneity

When evaluated across all aircraft profiles, there was no statistically significant bias (at the 95% level of confidence) between the retrieved and aircraft-based estimates for LWP or N_d, and reasonable or good correlations were found. Only the SatCORPS Himawari-8 product showed a statistically significant mean bias for τ and r_e (2.6 and 1.2 µm, respectively). However, this bias was only clear when applied to a larger set of cases than available to the MODIS overpasses, and when restricted to only those cases collocated with all three retrievals, Himawari-8 likewise show no significant bias in τ or r_e. Nonetheless, given that all three retrievals are based on the same bi-spectral technique, it seems likely that were MODIS and CERES-MODIS retrievals available for all collocation points they too would have a small bias in τ and r_e. A close examination shows that the low overall mean-bias in the retrievals is due in part to compensating errors between cases (vertical profiles) which were non- or lightly-precipitating (PWP < 10 gm^{-2}) with heavily-precipitating cases (PWP > 10 gm^{-2}). Below we summarize the key results for cloud optical depth (τ) and effective radius (r_e), liquid water path (LWP) and droplet number concentration (N_d), especially as relates to the presence of drizzle.

Effective Radius (r_e)

- We find a small positive bias in r_e for non- or lightly-precipitating cases of about 0.5 to 1 µm in all three datasets (satellite retrievals are slightly too large).
- This small positive r_e is due (at least in part) to the assumed Drop Size Distribution (DSD) width being too wide in the retrievals. In the retrievals, the DSD is assumed to be a modified-gamma distribution with an effective variance (v_e) of 0.1, which is larger than the value calculated from in situ measurements for non- or lightly-precipitating cases of 0.068. Previous studies of polarimetric data has also suggested that v_e for the marine clouds is likely to be narrower than is assumed in the satellite retrievals (e.g. Benas et al., 2019; Di Noia et al., 2019).
- We also find that the width of DSD increases (the k parameter decreases) as the PWP increases, suggesting it might be possible to parameterize this relationship as part of a combined imager-radar retrieval, in which the radar would constrain the PWP.
- Collectively, cases with relatively heavy precipitation (PWP > 10 gm^{-2}) have a negative bias (opposite in sign to the non- and lightly-precipitating cases). Not all heavily precipitating cases are negatively biased, but rather large biases occurred when significant precipitation was found near cloud top (PWC/CWC > 0.2). In these few cloud-top-precipitation cases, biases in r_e ranged from about -2 to -6 µm (satellite retrieved values are too small - see Figure 12). This result is not qualitatively dependent on inclusion of 2DS data in the calculation of r_e, but quantitatively the size of the bias

does depend on if (and how) 2DS data are merged with the CDP data to obtain the full DSD.

Optical Depth (τ) AND Liquid Water Path (LWP)

- CERES-MODIS and Himawari-8 are found to have a small positive bias in τ of about 2 to 3 (satellite retrievals are too large) for non- and lightly-precipitating cases. This bias is close but is not significant at the 95% level of confidence. MODIS (MYD06), on the other hand, do not appear to be biased for these cases (and instead was found to have a small negative bias for lightly precipitating clouds).
- LWP is derived based on τ and r_e in combination with an assumption about the cloud vertical structure. LWP retrievals based on the assumption of an adiabatic cloud structure compare well with the in situ observations (and are unbiased when averaged over all cases), while the assumption of a constant profile in LWC and r_e results in a significant overestimate in the LWP (~ +20%).
- For non- and lightly-precipitating cases, the small positive bias in r_e and τ for CERES-MODIS and Himawari-8 combine to produce a statistically significant bias of about +20 g m^{-2} in the LWP for these cases. MODIS (MYD06), on the other hand, was not biased by its small positive bias in r_e because of the small compensating bias in optical depth (about of -0.6) for the same lightly-precipitating cases.
- Heavily-precipitating case do not show a significant bias in τ or LWP for any of the three datasets. However, in all three, there is larger variability associated with the heavily-precipitating cases, with these cases having both the largest positive and largest negative errors in r_e, τ, and LWP. In particular, the handful of cases identified as having large negative errors in r_e (due to significant precipitation near cloud top) had the largest underestimate in LWP.

Cloud Droplet Number Concentration (N_d)

- N_d is also derived using τ and r_e. The formulation assumes the cloud is sub-adiabatic, meaning the total LWP is smaller than that for a true adiabatic cloud (of the same thickness, temperature and pressure) by a factor f_{ad}, but the LWC still increases linearly with altitude about cloud base (while N_d is constant). The formulation also depends on the DSD width (expressed via the parameter, k) and a condensation rate (that depends on pressure and temperature).
- While there is considerable variability from profile to profile, the SOCRATES data show that on average the SO stratocumulus LWC does increase linearly with height above cloud base and N_d is nearly constant. However, the relationship between r_e and LWC is not fixed, and the k-parameter generally varies with altitude.
- Overall, the N_d retrieval works reasonably well for our SO cases, as long as one uses the condensation rate that is appropriate for the SO (and this can be estimated reasonably well from the cloud top temperature).

- Errors in r_e and τ tend to cancel out producing relatively little bias in N_d. With respect to our precipitation classification, the only statistically significant bias in N_d that we find is in the lightly-precipitating category, where MODIS and CERES-MODIS retrievals have underestimated the N_d by about 30 to 40 #/cm^{-3}, and Himawari-8 retrievals have underestimated the N_d by -17.7 #/cm^{-3} (from an overall mean of about 100 #/cm^{-3}).
- Perhaps more problematically, we also found a bias of about 23.4 #/cm^{-}in N_d retrieval for cases featuring multiple low–level (< 3 km) clouds in the SatCORPS Himawari-8 dataset. The presence of optically thin layers with low droplet concentrations was found in 10 of the 53 profiles with collocated Himawari-8 data. Only one such case occurred in the set of cases with collocated MODIS data.
- Using assumed values of 0.8 for both f_{ad} and the k-parameter causes little bias in the retrieval because there is a cancellation of error between f_{ad} (observed mean = 0.74) and k-parameter (observed mean at cloud top = 0.76). However, using k and fad on a case-by-case basis does improve the correlation between the retrieved and in situ N_d.
- The profile-to-profile uncertainty (the percent mean fractional error) based on ~ 5 km × 5 km spatial averages of the N_d retrieval was found to be 40 to 55%, driven primarily by errors in effective radius (see Table 6).

5.2 Comparison with Previous Studies

Overall, our results broadly agree with the past evaluations studies of the MODIS bi-spectral retrievals technique for overcast stratocumulus. For instance, PZ11 reported that the MODIS retrieved $r_{e2.1}$ was overestimated by 15%-20% (mean bias of 2.08 μm) in comparison with cloud top r_e using 20 profiles (from mostly non- and lightly-precipitating subtropical stratocumulus) over the southeast Pacific (to the west of South America) during The VAMOS Ocean-Cloud-Atmosphere-Land Study (VOCALS), while Min et al. (2012) reported a mean bias of 1.75 μm using 17 non-precipitating cases from VOCALS. While we focused on the $r_{e3.7}$, we likewise find the MODIS $r_{e2.1}$ is overestimated, though by a slightly smaller amount of ~10% (mean bias of 1.12 μm) for 12 homogeneous non- and lightly-precipitating cases.

Closer to our region of study, A18 evaluated MODIS retrievals in wintertime stratocumulus over the Southern Ocean near Tasmania. Like us, A18 find that MODIS underestimates the effective radius of heavily-precipitating clouds and overestimates the effective radius of non-precipitating clouds, and like us A18 identify the width of the drop size distribution as a significant factor in the MODIS overestimate for non-precipitating clouds. However A18 found an overestimation of $r_{e2.1}$ by ~13 μm on average for non-precipitating clouds. While a variety of factors contribute to this rather large effective radius bias (see discussion A18), the broken and patchy nature of the clouds they observed, which were primarily open cell or disorganized stratocumulus, is a major factor. The MODIS and Himawari-8 bi-spectral retrievals are based on an assumption of 1D radiative transfer and are known to work poorly for broken and spatially heterogonous clouds, and to substantially overestimate r_e on average for broken clouds (e.g. Marshak et al., 2006). A18 did include two flights with overcast (closed cell) stratocumulus. According to their Table 1, the average in situ effective radius for these two cases were 8.6 and 7.5 μm (which is consistent with the smaller values we observed during SOCRATES for non-precipitating clouds) while the MODIS retrieved values of $r_{e3.7}$ are near 12.6 μm on both flights (which is within the range we found for non-precipitating clouds but toward the high side), resulting in a bias of 4 to 5 μm (which is several μm bigger than our bias for this cloud type). Our SOCRATES cases included only one non-precipitating case with a bias larger than 4 μm, and this case was one of our cases with a relatively large cloud heterogeneity index. Thus, we speculate that the somewhat larger bias found by A18 for their overcast cases might be a consequence of cloud heterogeneity. (We note that A18 do a report a heterogeneity index for their cases, but the index they use is the standard MODIS product index which looks at variability of 250m pixel radiances within each 1 km pixel used in the optical depth retrieval, and does not characterize the variability of the larger scene or collocation box used in the analysis). We also note that the observations A18 use in their analysis are not restricted to the region near cloud top. One expects the effective radius (in non-precipitating clouds) will be smaller below cloud top and this might well have reduced the magnitude of the in situ estimate (and increase the apparent bias) by a few microns.

Very recently, Zhao et al. (2020), hereafter Z20, evaluated MODIS and Himawari-8 r_e using SOCRATES measurements for a subset of the flights that we have analyzed. Their results differ from ours in several key respects. Their analysis was based on two approaches: (1) measurements taken when the aircraft was flying horizontally (level legs) within about 200 m of cloud top and (2) vertical profiles created from aircraft ramps through the cloud (which is similar to our study). Based on the horizontal flight data, Z20 report a mean bias with Himawari-8 of 4.39 μm for liquid phase clouds and 2.24 μm for mixed phase clouds (see their Figure 4), while for MODIS $r_{e3.7}$ they report a bias in of about 2 μm for both liquid and mixed-phase clouds (see their Figure 7). It is not clear from their manuscript whether the comparison for Himawari-8 is based on only CDP or the combination of CDP + 2DS (while their MODIS comparison is clearly based on the combination) which might explain some of the difference between their Himawari-8 and MODIS results, but more importantly, in both comparisons the collocated in situ data with Himawari-8 never has an effective radius value greater than about 11.5 μm and the in situ data collocated with MODIS never has a value for r_e larger than about 9.4 μm. This fundamentally differs from what we find. We frequently find in situ values for r_e are larger than 12 μm for profiles that contain precipitation (which is common place) and this seems consistent with previous studies. We note that in their analysis of aircraft profile data, Z20 find their profiles (1) have vertical mean values for r_e that is larger than the average for their horizontal flight legs and (2) the profile values near cloud top suggest a bias for Himawari-8 that is near (or below) 1 μm (see their Figures 6). As such, their vertical profile data is consistent with our results and inconsistent with the horizontal flight data they present. We speculate that when creating their 10s horizontal leg data averages that periods with low or no condensate (with small values of effective radius) or perhaps drop-outs in the data have somehow biased the 10s averages. In general, we suggest that averages of effective radius should either (1) be weighted by liquid-water-content or total number concentration, or (2) better yet, a single DSD should be summed (generated) from the measured counts for the full averaging period and effective radius (and other parameters that characterize the distribution) calculated from this single DSD.

As noted above, Z20 subdivide their results between liquid and mixed-phase clouds. They identify mixed phase as those where the ratio of liquid water content from the CDP (where presumably all CDP observed particles are assumed to be liquid) divided by the total condensed water (estimated from measurements by a Closed-Path Hygrometer, CLH-2) is less than 0.85. We suggest that the approach used by Z20 is problematic because it relies on measurements from two different instruments, where each measurement has a nominal uncertainty of 10 to 15%, and the instruments can (and do) have different response times and sensitivities to icing in supercooled environments. This means that the measurement uncertainty alone can easily cause the ratio of liquid-to-total condensate to be less than 0.85. In fact, we have been unable to reproduce Z20's results in this regard and find that in many of our aircraft profiles LWC for the CDP is greater than TWC from the CLH-2 such that the ratio has unphysical values greater than one. Consequently, we have examined the ratio of ice-to-total condensate for precipitation based on the 2DS only, whose imagery has been processes following Wu and Mcfarquhar (2019) to identify ice particles >= ~ 200 um. Whereas Z20 find that the majority of the cloud is mixed phase, we find that only 4 out of 53 of our profiles contain even 10% ice from the perspective of the 2DS (Figure S4). Of course, it could well be the case that numerous small-ice particles are present and the 2DS-only estimate that we use is substantially underestimating the contribution of ice. But one expects that small ice particles will very rapidly grow in size via the Wegener–Bergeron–Findeisen process, such that (while our 2DS-only) estimate might underestimate the mass of ice, we would detect its presence. Overall, we find no distinction between cases that contained large-ice from those without large-ice, in any significant way, for any choice of the ice-mass-fraction. Ultimately Z20 conclude that phase does not matter (bias is about the same for liquid and mixed-phase), and in this sense we agree. Nonetheless, we do not believe the majority of the cloud should be considered mixed-phase. At present, evaluation of cloud phase (across the full range of SOCRATES instruments) remains on ongoing area of research by SOCRATES instrument teams, and more work is needed to understand the performance of instruments under the challenging conditions encountered.

5.3 Conclusions

Regardless of whether our speculations regarding A18 and Z20 are correct, we conclude there is a consistent pattern between studies in which show there are statically significant biases associated with the MODIS and Himawari-8 bi-spectra retrievals of r_e for overcast SO stratocumulus as compared with in situ aircraft measurements, even when comparisons are appropriately restricted to near cloud-top observations. At least here and in A18, the bias depends significantly on precipitation within the cloudy column, and we conclude that the presence of precipitation near cloud top (not just within cloud) is of particular importance. We find the bias for non- or lightly-precipitating stratocumulus to be consistent with (if a bit smaller) than those identified during VOCALS for subtropical stratocumulus, and find (as other studies have) that this bias is due (at least in part) to the width (shape) of the assumed drop size distribution, which is too broad in the bi-spectral retrieval for non-precipitating marine stratocumulus. In general, precipitation is associated with wider distributions, and the observed DSDs is not always well characterized using a monomodal log-normal or gamma size distribution (see supplementary material). The biases in optical depth are less robust and typically not statistically significant at the 95% level of confidence, depending on dataset and precipitation category. Errors in effective radius and optical depth propagate into the retrieved LWP and N_d in somewhat complex ways, as errors in the effective radius and optical depth are correlated (again depending on the presence of precipitation). A summary is given in section 5.1 with more detailed discussions given in sections 3.3 and 3.4. In general, we find the bias and case-to-case uncertainty in the satellite N_d retrievals is smaller than one might expect simply from the bias and random errors in effective radius because of this correlation.

We stress the SOCRATES measurement were collected in the afternoon and during the SH summer where solar zenith angles are less than 60^o (conditions under which theoretical studies suggest the bi-spectral retrieval should work well for homogeneous clouds). So we are not surprised to find the bi-spectral retrieval works similarly well during SOCRATES as has been found with subtropical stratocumulus. We suggest that additional research should be undertaken using detailed radiative transfer simulations to model and better understand how variations in the DSD and its vertical and horizontal structure are likely to effect retrievals at larger solar zenith angles typical of the SO in other seasons and other times of day; and more generally suggest that additional measurements should be collected during the SO winter.

In section 4.4, we examined briefly the impact combining data from the CDP, 2DS and 2DC probes has on our analysis. The agreement between satellite retrievals and in situ r_e shows some dependence on the choices of in situ probes and merging methods. Several evaluation studies (e.g. King et al., 2013; Platnick & Valero, 1995; Witte et al., 2018) have considered the uncertainty of in situ measurement of r_e. For instance, Witte et al. (2018) compared the r_e measured by phase Doppler interferometer (PDI) with MODIS $r_{e2.1}$ and revisited the evaluation studies over the Pacific (e.g. Noble and Hudson, 2015; PZ11) using different instruments during the same three campaigns. Witte et al. (2018) found no apparent systematic bias (mean bias of -0.22 µm) in retrieved $r_{e2.1}$. Indeed, as we show in section 4.4 we can merge the CDP and 2DS data in such a way that there is little overall bias in the r_e, but this result is obtain by balancing the errors between non-, lightly- and heavily precipitating case and there is a fundamental difference in the bias for the different precipitating categories. In short, as these studies highlight, there are significant uncertainties associated with in situ measurements, and a continued need for improved in situ measurements.

Acknowledgments, Samples, and Data

This work was supported by the U.S. National Science Foundation on grant 1660609 and would not have been possible except for the work of a large and dedicated group of scientist, engineers and planners, too numerous to list here. In particular, we want to call out the contribution from the many individuals associated with the NCAR Earth Observing Laboratory, SOCRATES Science and Aircraft Instrument Teams. The authors wish to acknowledge the SOCRATES Project for providing in situ measurements via the SOCRATES Data Archive Center (SDAC) at NCAR's Earth Observing Laboratory. In particular, low rate navigation, state parameter, and microphysics flight-level data https://doi.org/10.5065/D6M32TM9; 2DS data https://doi.org/10.26023/8HMG-WQP3-XA0X; and 2DC data https://doi.org/10.26023/E95A-FKYF-7P0R. We also thank the MODIS, CERES-MODIS and SatCORPS teams, and acknowledge the use of data products produced by them. MODIS products were obtained from https://ladsweb.modaps.eosdis.nasa.gov/ including: (1) MODIS level 2 Collection 6.1 cloud products from Aqua platform (MYD06)(http://dx.doi.org/10.5067/MODIS/MYD06_L2.061), (2) MODIS level 3 MYD03 geolocation product(http://dx.doi.org/10.5067/MODIS/MYD03.006), and (3) MODIS Level-1B MYD02QKM Calibrated Radiances product(http://dx.doi.org/10.5067/MODIS/MYD02QKM.061). NASA SatCORPS Himawari-8 data during the SOCRATES were obtained from SDAC at https://doi.org/10.5065/D6CC0ZFJ. Table S1 in the supporting information includes all the case-by-case values needed to generate the figures containted in this article.

References

Ahn, E., Huang, Y., Chubb, T. H., Baumgardner, D., Isaac, P., de Hoog, M., et al. (2017). In situ observations of wintertime low-altitude clouds over the Southern Ocean. *Quarterly Journal of the Royal Meteorological Society*, *143*(704), 1381–1394. https://doi.org/10.1002/qj.3011

Ahn, E., Huang, Y., Siems, S. T., & Manton, M. J. (2018). A Comparison of Cloud Microphysical Properties Derived From MODIS and CALIPSO With In Situ Measurements Over the Wintertime Southern Ocean. *Journal of Geophysical Research: Atmospheres*, *123*(19), 11,120-11,140. https://doi.org/10.1029/2018JD028535

Albrecht, B. A., Fairall, C. W., Thomson, D. W., White, A. B., Snider, J. B., & Schubert, W. H. (1990). Surface-based remote sensing of the observed and the Adiabatic liquid water content of stratocumulus clouds. *Geophysical Research Letters*, *17*(1), 89–92. https://doi.org/10.1029/GL017i001p00089

Arduini, R. F., Minnis, P., Smith Jr, W. L., Ayers, J. K., Khaiyer, M. M., & Heck, P. (2005). *Sensitivity of satellite-retrieved cloud properties to the effective variance of cloud droplet size distribution*. Science Applications International Corporation, Hampton, Virginia; NASA.

Benas, N., Meirink, J. F., Stengel, M., & Stammes, P. (2019). Sensitivity of liquid cloud optical thickness and effective radius retrievals to cloud bow and glory conditions using two SEVIRI imagers. *Atmospheric Measurement Techniques*, *12*(5), 2863–2879. https://doi.org/10.5194/amt-12-2863-2019

Bennartz, R. (2007). Global assessment of marine boundary layer cloud droplet number concentration from satellite. *Journal of Geophysical Research Atmospheres*, *112*(2), 1–16. https://doi.org/10.1029/2006JD007547

Bessho, K., Date, K., Hayashi, M., Ikeda, A., Imai, T., Inoue, H., et al. (2016). An introduction to Himawari-8/9 — Japan's new-generation geostationary meteorological satellites. *Journal of the Meteorological Society of Japan*, *94*(2), 151–183. https://doi.org/10.2151/jmsj.2016-009

Bodas-Salcedo, A., Hill, P. G., Furtado, K., Williams, K. D., Field, P. R., Manners, J. C., et al. (2016). Large contribution of supercooled liquid clouds to the solar radiation budget of the Southern Ocean. *Journal of Climate*, *29*(11), 4213–4228. https://doi.org/10.1175/JCLI-D-15-0564.1

Bodas-Salcedo, A., Williams, K. D., Ringer, M. A., Beau, I., Cole, J. N. S., Dufresne, J. L., et al. (2014). Origins of the solar radiation biases over the Southern Ocean in CFMIP2 models. *Journal of Climate*, *27*(1), 41–56. https://doi.org/10.1175/JCLI-D-13-00169.1

Borg, L. A., & Bennartz, R. (2007). Vertical structure of stratiform marine boundary layer clouds and its impact on cloud albedo. *Geophysical Research Letters*, *34*(5), 1–5. https://doi.org/10.1029/2006GL028713

Ceppi, P., Hwang, Y. T., Frierson, D. M. W., & Hartmann, D. L. (2012). Southern Hemisphere jet latitude biases in CMIP5 models linked to shortwave cloud forcing. *Geophysical Research Letters*, *39*(19), 1–5. https://doi.org/10.1029/2012GL053115

Ceppi, P., Hwang, Y. T., Liu, X., Frierson, D. M. W., & Hartmann, D. L. (2013). The relationship between the ITCZ and the Southern Hemispheric eddy-driven jet. *Journal of Geophysical Research Atmospheres*, *118*(11), 5136–5146. https://doi.org/10.1002/jgrd.50461

Di Noia, A., Hasekamp, O. P., van Diedenhoven, B., & Zhang, Z. (2019). Retrieval of liquid water cloud properties from POLDER-3 measurements using a neural network ensemble approach. *Atmospheric Measurement Techniques*, *12*(3), 1697–1716. https://doi.org/10.5194/amt-12-1697-2019

Grosvenor, D. P., Sourdeval, O., Zuidema, P., Ackerman, A., Alexandrov, M. D., Bennartz, R., et al. (2018). Remote Sensing of Droplet Number Concentration in Warm Clouds: A Review of the Current State of Knowledge and Perspectives. Reviews of Geophysics, 56(2), 409–453. https://doi.org/10.1029/2017RG000593

Hansen, J. E. (1971). Multiple Scattering of Polarized Light in Planetary Atmospheres Part II. Sunlight Reflected by Terrestrial Water Clouds. *Journal of the Atmospheric Sciences*, *28*(8), 1400–1426. https://doi.org/10.1175/1520-0469(1971)028<1400:MSOPLI>2.0.CO;2

Haynes, J. M., Jakob, C., Rossow, W. B., Tselioudis, G., & Brown, J. B. (2011). Major characteristics of Southern Ocean cloud regimes and their effects on the energy budget. *Journal of Climate*, *24*(19), 5061–5080. https://doi.org/10.1175/2011JCLI4052.1

Huang, Y., Siems, S. T., Manton, M. J., & Thompson, G. (2014). An evaluation of WRF simulations of clouds over the southern ocean with a-train observations. *Monthly Weather Review*, *142*(2), 647–667. https://doi.org/10.1175/MWR-D-13-00128.1

Hwang, Y.-T., & Frierson, D. M. W. (2013). Link between the double-Intertropical Convergence Zone problem and cloud biases over the Southern Ocean. *Proceedings of the National Academy of Sciences*, *110*(13), 4935 LP – 4940. https://doi.org/10.1073/pnas.1213302110

Kato, S., Loeb, N. G., Rose, F. G., Doelling, D. R., Rutan, D. A., Caldwell, T. E., et al. (2013). Surface irradiances consistent with CERES-derived top-of-atmosphere shortwave and longwave irradiances. *Journal of Climate*, *26*(9), 2719–2740. https://doi.org/10.1175/JCLI-D-12-00436.1

Kay, J. E., Wall, C., Yettella, V., Medeiros, B., Hannay, C., Caldwell, P., & Bitz, C. (2016). No access global climate impacts of fixing the Southern Ocean shortwave radiation bias in the Community Earth System Model (CESM). *Journal of Climate*, *29*(12), 4617–4636. https://doi.org/10.1175/JCLI-D-15-0358.1

King, N. J., Bower, K. N., Crosier, J., & Crawford, I. (2013). Evaluating modis cloud retrievals with in situ observations from VOCALS-REx. *Atmospheric Chemistry and Physics*, *13*(1), 191–209. https://doi.org/10.5194/acp-13-191-2013

Korolev, A., McFarquhar, G., Field, P. R., Franklin, C., Lawson, P., Wang, Z., et al. (2017). Mixed-Phase Clouds: Progress and Challenges. *Meteorological Monographs*, *58*, 5.1-5.50. https://doi.org/10.1175/AMSMONOGRAPHS-D-17-0001.1

Lance, S., Brock, C. A., Rogers, D., & Gordon, J. A. (2010). Water droplet calibration of the Cloud Droplet Probe (CDP) and in-flight performance in liquid, ice and mixed-phase clouds during ARCPAC. *Atmospheric Measurement Techniques*, *3*(6), 1683–1706. https://doi.org/10.5194/amt-3-1683-2010

Liang, L., Di Girolamo, L., & Platnick, S. (2009). View-angle consistency in reflectance, optical thickness and spherical albedo of marine water-clouds over the northeastern Pacific through MISR-MODIS fusion. *Geophysical Research Letters*, *36*(9), 1–5. https://doi.org/10.1029/2008GL037124

Mace, G. G. (2010). Cloud properties and radiative forcing over the maritime storm tracks of the Southern Ocean and North Atlantic derived from A-Train. *Journal of Geophysical Research Atmospheres*, *115*(10), 1–26. https://doi.org/10.1029/2009JD012517

Mace, G. G., Zhang, Q., Vaughan, M., Marchand, R., Stephens, G., Trepte, C., & Winker, D. (2009). A description of hydrometeor layer occurrence statistics derived from the first year of merged Cloudsat and CALIPSO data. *Journal of Geophysical Research Atmospheres*, *114*(8), 1–17. https://doi.org/10.1029/2007JD009755

Marshak, A., Platnick, S., Várnai, T., Wen, G., & Cahalan, R. F. (2006). Impact of three-dimensional radiative effects on satellite retrievals of cloud droplet sizes. *Journal of Geophysical Research Atmospheres*, *111*(9), 1–12. https://doi.org/10.1029/2005JD006686

McCoy, D. T., Burrows, S. M., Wood, R., Grosvenor, D. P., Elliott, S. M., Ma, P. L., et al. (2015). Natural aerosols explain seasonal and spatial patterns of Southern Ocean cloud albedo. *Science Advances*, *1*(6). https://doi.org/10.1126/sciadv.1500157

Meskhidze, N., & Nenes, A. (2006). Phytoplankton and Cloudiness in the Southern Ocean. *Science*, *314*(5804), 1419 LP – 1423. https://doi.org/10.1126/science.1131779

Min, Q., Joseph, E., Lin, Y., Min, L., Yin, B., Daum, P. H., et al. (2012). Comparison of MODIS cloud microphysical properties with in-situ measurements over the Southeast Pacific. *Atmospheric Chemistry and Physics*, *12*(23), 11261–11273. https://doi.org/10.5194/acp-12-11261-2012

Minnis, P., Arduini, R. F., Young, D. F., Ayers, J. K., Albrecht, B. A., Sharon, T., & Stevens, B. (2004). An examination of the impact of drizzle drops on satellite retrieved effective particle sizes. In *Proceedings of the 14th International Conference on Clouds and Precipitation*. Bologna, Italy.

Minnis, P., Nguyen, L., Palikonda, R., Heck, P. W., Spangenberg, D. A., Doelling, D. R., et al. (2008). Near-real time cloud retrievals from operational and research meteorological satellites. *Remote Sensing of Clouds and the Atmosphere XIII*, *7107*, 710703. https://doi.org/10.1117/12.800344

Minnis, P., Sun-Mack, S., Chen, Y., Chang, F.-L., Yost, C. R., Smith Jr., W. L., et al. (2020). *CERES MODIS cloud product retrievals for Edition 4, Part I: Algorithm changes*. *IEEE Trans. Geosci. Remote Sens.*

Minnis, P., Sun-Mack, S., Chen, Y., Khaiyer, M. M., Yi, Y., Ayers, J. K., et al. (2011). CERES Edition-2 Cloud Property Retrievals Using TRMM VIRS and Terra and Aqua MODIS Data—Part II: Examples of Average Results and Comparisons With Other Data. *IEEE Transactions on Geoscience and Remote Sensing*, *49*(11), 4401–4430. https://doi.org/10.1109/TGRS.2011.2144602

Morrison, A. E., Siems, S. T., & Manton, M. J. (2011). A three-year climatology of cloud-top phase over the Southern Ocean and North Pacific. *Journal of Climate*, *24*(9), 2405–2418. https://doi.org/10.1175/2010JCLI3842.1

Nakajima, T. Y., Suzuki, K., & Stephens, G. L. (2010). 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*(6), 1884–1896. https://doi.org/10.1175/2009JAS3280.1

Nakajima, T., & King, M. D. (1990). Determination of the optical thickness and effective particle radius of clouds from reflected solar radiation measurements. Part I: theory. *Journal of the Atmospheric Sciences*, *47*(15), 1878–1893. https://doi.org/10.1175/1520-0469(1990)047<1878:DOTOTA>2.0.CO;2

Naud, C. M., Booth, J. F., & Del Genio, A. D. (2014). Evaluation of ERA-Interim and MERRA cloudiness in the southern ocean. *Journal of Climate*, *27*(5), 2109–2124. https://doi.org/10.1175/JCLI-D-13-00432.1

Noble, S. R., & Hudson, J. G. (2015). MODIS comparisons with northeastern Pacific in situ stratocumulus microphysics. *Journal of Geophysical Research: Atmospheres*, *120*(16), 8332–8344. https://doi.org/10.1002/2014JD022785

Painemal, D., & Zuidema, P. (2011). Assessment of MODIS cloud effective radius and optical thickness retrievals over the Southeast Pacific with VOCALS-REx in situ measurements. *Journal of Geophysical Research: Atmospheres*, *116*(D24). https://doi.org/10.1029/2011JD016155

Platnick, S. (2000). Vertical photon transport in cloud remote sensing problems. *Journal of Geophysical Research Atmospheres*, *105*(D18), 22919–22935. https://doi.org/10.1029/2000JD900333

Platnick, S., King, M. D., Ackerman, S. A., Menzel, W. P., Baum, B. A., Riedi, J. C., & Frey, R. A. (2003). The MODIS cloud products: algorithms and examples from Terra. *IEEE*

Transactions on Geoscience and Remote Sensing, *41*(2), 459–473. https://doi.org/10.1109/TGRS.2002.808301

Platnick, S., Meyer, K. G., King, M. D., Wind, G., Amarasinghe, N., Marchant, B., et al. (2017). The MODIS Cloud Optical and Microphysical Products: Collection 6 Updates and Examples from Terra and Aqua. *IEEE Transactions on Geoscience and Remote Sensing*, *55*(1), 502–525. https://doi.org/10.1109/TGRS.2016.2610522

Sallée, J. B., Shuckburgh, E., Bruneau, N., Meijers, A. J. S., Bracegirdle, T. J., Wang, Z., & Roy, T. (2013). Assessment of Southern Ocean water mass circulation and characteristics in CMIP5 models: Historical bias and forcing response. *Journal of Geophysical Research: Oceans*, *118*(4), 1830–1844. https://doi.org/10.1002/jgrc.20135

Schneider, D. P., & Reusch, D. B. (2016). Antarctic and Southern Ocean surface temperatures in CMIP5 models in the context of the surface energy budget. *Journal of Climate*, *29*(5), 1689–1716. https://doi.org/10.1175/JCLI-D-15-0429.1

Seethala, C., & Horváth, Á. (2010). Global assessment of AMSR-E and MODIS cloud liquid water path retrievals in warm oceanic clouds. *Journal of Geophysical Research Atmospheres*, *115*(13), 1–19. https://doi.org/10.1029/2009JD012662

Smith, W., & Minnis, P. (2018). NASA SatCORPS Himawari Cloud Retrieval Data. Version 2.1 (Version 2.1) [Data set]. UCAR/NCAR - Earth Observing Laboratory. https://doi.org/10.5065/D6CC0ZFJ Accessed 10 Jun 2020.

Trenberth, K. E., & Fasullo, J. T. (2010). Simulation of present-day and twenty-first-century energy budgets of the southern oceans. *Journal of Climate*, *23*(2), 440–454. https://doi.org/10.1175/2009JCLI3152.1

Trepte, Q. Z., Bedka, K. M., Chee, T. L., Minnis, P., Sun-Mack, S., Yost, C. R., et al. (2019). Global Cloud Detection for CERES Edition 4 Using Terra and Aqua MODIS Data. *IEEE Transactions on Geoscience and Remote Sensing*, *57*(11), 9410–9449. https://doi.org/10.1109/TGRS.2019.2926620

UCAR/NCAR-Earth Observing Laboratory. (2019). Low Rate (LRT - 1 sps) Navigation, State Parameter, and Microphysics Flight-Level Data. Version 1.3 (Version 1.3) [Data set]. UCAR/NCAR - Earth Observing Laboratory. https://doi.org/10.5065/D6M32TM9 Accessed 10 Jun 2020.

Vergara-Temprado, J., Miltenberger, A. K., Furtado, K., Grosvenor, D. P., Shipway, B. J., Hill, A. A., et al. (2018). Strong control of Southern Ocean cloud reflectivity by ice-nucleating particles. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(11), 2687–2692. https://doi.org/10.1073/pnas.1721627115

Williams, K. D., Bodas-Salcedo, A., Déqué, M., Fermepin, S., Medeiros, B., Watanabe, M., et al. (2013). The transpose-AMIP II experiment and its application to the understanding of southern ocean cloud biases in climate models. *Journal of Climate*, *26*(10), 3258–3274. https://doi.org/10.1175/JCLI-D-12-00429.1

Witte, M. K., Yuan, T., Chuang, P. Y., Platnick, S., Meyer, K. G., Wind, G., & Jonsson, H. H. (2018). MODIS Retrievals of Cloud Effective Radius in Marine Stratocumulus Exhibit No Significant Bias. *Geophysical Research Letters*, *45*(19), 10,656-10,664. https://doi.org/10.1029/2018GL079325

Wofsy, S. C. (2011). HIAPER Pole-to-Pole Observations (HIPPO): Fine-grained, global-scale measurements of climatically important atmospheric gases and aerosols. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *369*(1943), 2073–2086. https://doi.org/10.1098/rsta.2010.0313

Wood, R. (2012). Stratocumulus clouds. *Monthly Weather Review*, *140*(8), 2373–2423. https://doi.org/10.1175/MWR-D-11-00121.1

Wood, R., & Hartmann, D. L. (2006). Spatial variability of liquid water path in marine low cloud: The importance of mesoscale cellular convection. *Journal of Climate*, *19*(9), 1748–1764. https://doi.org/10.1175/JCLI3702.1

Wood, R., Bretherton, C. S., Leon, D., Clarke, A. D., Zuidema, P., Allen, G., & Coe, H. (2011). An aircraft case study of the spatial transition from closed to open mesoscale cellular convection over the Southeast Pacific. *Atmospheric Chemistry and Physics*, *11*(5), 2341–2370. https://doi.org/10.5194/acp-11-2341-2011

Wu, W., & McFarquhar, G. (2019). NSF/NCAR GV HIAPER 2D-S Particle Size Distribution (PSD) Product Data. Version 1.1 (Version 1.1) [Data set]. UCAR/NCAR - Earth Observing Laboratory. https://doi.org/10.26023/8HMG-WQP3-XA0X Accessed 10 Jun 2020.

Wu, W., & McFarquhar, G. (2019). NSF/NCAR GV HIAPER Fast 2DC Particle Size Distribution (PSD) Product Data. Version 1.1 (Version 1.1) [Data set]. UCAR/NCAR - Earth Observing Laboratory. https://doi.org/10.26023/E95A-FKYF-7P0R Accessed 10 Jun 2020.

Zhang, Z. (2013). On the sensitivity of cloud effective radius retrieval based on spectral method to bi-modal droplet size distribution: A semi-analytical model. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *129*, 79–88. https://doi.org/10.1016/j.jqsrt.2013.05.033

Zhang, Z., & Platnick, S. (2011). An assessment of differences between cloud effective particle radius retrievals for marine water clouds from three MODIS spectral bands. *Journal of Geophysical Research Atmospheres*, *116*(20), 1–19. https://doi.org/10.1029/2011JD016216

Zhang, Z., Ackerman, A. S., Feingold, G., Platnick, S., Pincus, R., & Xue, H. (2012). Effects of cloud horizontal inhomogeneity and drizzle on remote sensing of cloud droplet effective radius: Case studies based on large-eddy simulations. *Journal of Geophysical Research Atmospheres*, *117*(19), 1–18. https://doi.org/10.1029/2012JD017655

Zhao, L., Zhao, C., Wang, Y., Wang, Y., & Yang, Y. (2020). Evaluation of Cloud Microphysical Properties Derived From MODIS and Himawari-8 Using In Situ Aircraft Measurements Over the Southern Ocean. *Earth and Space Science*, *7*(5), e2020EA001137. https://doi.org/10.1029/2020EA001137