Article

Not peer-reviewed version

Hyperspectral Retrieval of Vertical Cloud Droplet Profiles and Above-Cloud Integrated Water Vapor: Reducing Liquid Water Path Retrieval Bias with Application to EMIT

Andrew Buggee * and Peter Pilewskie

Posted Date: 31 March 2026

doi: 10.20944/preprints202603.2506.v1

Keywords: cloud remote sensing; retrievals; hyperspectral measurements; cloud optical properties

 doi:10.20944/preprints202603.2506.v1

Article

Hyperspectral Retrieval of Vertical Cloud Droplet Profiles and Above-Cloud Integrated Water Vapor: Reducing Liquid Water Path Retrieval Bias with Application to EMIT

Andrew John Buggee [1,2,*] and Peter Pilewskie [1,2]

[1] Laboratory for Atmospheric and Space Physics, University of Colorado Boulder, 80303, USA

[2] Department of Atmospheric and Oceanic Sciences, University of Colorado Boulder, 80303, USA

* Correspondence: andrew.buggee@lasp.colorado.edu

Highlights

What are the main findings?

- A Gauss-Newton optimal estimation retrieval simultaneously retrieves vertical profiles of cloud droplet effective radius and above-cloud integrated water vapor from hyperspectral solar backscatter measurements, reducing liquid water path bias from 45% (standard bispectral method) to 18% against in situ measurements.
- Application to 603 EMIT pixels over the southeast Pacific shows that MODIS liquid water path retrievals, which assume a vertically homogeneous cloud, exceed the hyperspectral estimate by 25.6% on average, consistent with the expected bias direction from prior comparisons with retrievals using microwave observations.

What are the implications of the main findings?

- Simultaneous retrieval of above-cloud water vapor is essential for accurate droplet profile retrievals: assuming a fixed water vapor column introduces biases of up to 0.4 μm and 1 μm in retrieved effective radius at cloud top and base, respectively.
- The upcoming CLARREO Pathfinder HySICS instrument, with radiometric uncertainty of 0.3% and 636 spectral channels, is expected to enable routine, accurate retrievals of vertical cloud droplet profiles, substantially advancing beyond current column-averaged multispectral approaches.

Abstract

Accurate liquid water path estimates derived from backscattered solar radiation require knowledge of the vertical structure of cloud droplet effective radius, yet standard bispectral retrievals assume a vertically homogeneous cloud and overestimate liquid water path by up to 45% compared with in situ measurements. We developed a Gauss-Newton optimal estimation retrieval that simultaneously estimates vertical profiles of cloud droplet effective radius and above-cloud integrated water vapor from hyperspectral solar backscatter measurements in the visible and shortwave infrared. The retrieval solves for effective radius at cloud top and base, cloud optical thickness, and above-cloud integrated water vapor in logarithmic space, using an a priori covariance matrix with off-diagonal elements derived from VOCALS-REx in-situ measurements. Tested on 69 simulated HySICS reflectance spectra constructed from in situ cloud microphysics, the hyperspectral retrieval reduces the average liquid water path error to 17.7%, compared to 45.2% for the standard bispectral method. Applied to 603 EMIT hyperspectral measurements over the southeast Pacific, MODIS-retrieved liquid water path exceeds the hyperspectral estimate by 25.6% on average. These results demonstrate that simultaneous retrieval of above-cloud water vapor is necessary for accurate droplet profile retrievals, and that the upcoming CLARREO Pathfinder instrument, with its 0.3% radiometric uncertainty, should enable routine vertical profiling of cloud droplet size.

 doi:10.20944/preprints202603.2506.v1

Keywords: cloud remote sensing; retrievals; hyperspectral measurements; cloud optical properties

1. Introduction

Marine stratus and stratocumulus clouds cover approximately 12% and 23% of the global ocean surface on the annual mean, respectively, and occur most frequently over cold eastern ocean basins [1]. These clouds typically form in the boundary layer and are strong reflectors of incident solar radiation, with an average top-of-atmosphere cooling effect of -1.4 W/m² and -8.3 W/m², respectively [2]. The abundance and cooling effect of marine stratus and stratocumulus make them vital to our understanding of Earth's radiative energy balance [3].

Theoretical models of warm, non-precipitating adiabatic clouds predict a vertical structure of droplet size that increases from cloud base to cloud top, with many in situ measurements verifying this prediction (King et al., 2013; Miles et al., 2000; Painemal and Zuidema, 2011; Yau and Rogers, 1996). Many field campaigns have verified this [4–6]. Vertical profiles of droplet size can be used to determine the droplet growth process and the mixing of cloud top with dry air above [7–9]. In addition, vertical droplet profiles are required for an accurate estimation of liquid water path, the vertically integrated water content, and cloud albedo [10–12].

For many decades, space-borne spectrally resolved measurements of reflected solar radiation over cloudy scenes have been used to remotely determine cloud droplet effective radius, defined as the ratio of the third moment to the second moment of the droplet size distribution, and cloud optical thickness, the number of photon mean free paths [13]. Twomey and Seton [14] outlined what is now considered the standard method for deriving cloud optical properties with spectral measurements in the visible (0.4 $\boldsymbol{\mu m}$ to 0.7 $\boldsymbol{\mu m}$) and shortwave infrared (1 $\boldsymbol{\mu m}$ to 2.5 $\boldsymbol{\mu m}$ for this study). This optical remote sensing method, often referred to as the two-band lookup table method (TBLUT), can be applied to measured reflectance in as few as two spectral bands: one at a wavelength where absorption by water is negligible and the other at a wavelength where water weakly absorbs [14–16]. While straightforward to implement, the TBLUT method derives a column weighted average of the actual vertical droplet size profile [17]. Retrieving a column weighted average, which neglects the vertical variation in droplet size, is one reason that the TBLUT method was found to overestimate liquid water path by as much as 25% [12,18].

A single shortwave infrared wavelength is used to estimate the effective radius in many operational retrieval algorithms. Due to the variability of absorption by liquid water in the shortwave infrared, this results in a wavelength-dependent effective radius. This was explained by Platnick [19], who showed that photons at different wavelengths penetrate to different depths within clouds due to the spectral dependence of droplet single scattering albedo. Thus, the retrieved droplet radius represents a weighted average over the vertical extent of the cloud, with the largest weighting occurring near cloud top [19]. This wavelength-dependent retrieval laid the foundation for using measurements at multiple shortwave infrared wavelengths to retrieve vertical profiles of droplet size that capture vertical variation [20–24].

Following Platnick [19], Chang and Li [21] proposed using MODIS (Moderate Resolution Imaging Spectroradiometer) measurements at three shortwave infrared spectral bands to retrieve the vertical dependence of effective droplet radius. Their method assumed a linear relationship between effective droplet radius and altitude within cloud. Kokhanovsky and Rozanov [22] outlined the mathematical framework for applying the Gauss-Newton optimal estimation technique to infer a vertical droplet profile using spectral measurements. They showed that four MODIS wavelengths could be used simultaneously with less computational cost than the lookup table method to solve for three variables, the effective radii at cloud top and cloud base, and cloud optical thickness, demonstrating their method with synthetic and real MODIS measurements. King and Vaughan [23] applied an optimal estimation technique to hundreds of synthetic spectral measurements throughout the visible and shortwave infrared to retrieve effective radii at cloud top and cloud base, cloud optical thickness and the integrated water vapor content above cloud. Buggee and Pilewskie [20] built on

 doi:10.20944/preprints202603.2506.v1

these results by developing an optimal estimation retrieval method to estimate vertical droplet profiles from MODIS measurements coincident with in situ data for validation. They determined that the uniqueness of the retrieved droplet profile is predominantly impacted by the measurement and forward model uncertainty, and to a lesser degree, the number of spectral channels used. See Buggee and Pilewskie [20] for a thorough description of these previous studies.

Hyperspectral sensors that sample the solar spectrum in many hundreds of contiguous spectral bands have been proposed and built, with some already launched and others soon to be [25–28]. Coddington et al. [29] and King and Vaughan [23] have shown that these measurements contain a wealth of information on cloud optical properties, which can be defined as "…both a consequence of and an expression for the solar radiative transfer characteristics of clouds [13]." The increase in spectral sampling and the reduction of radiometric uncertainty of upcoming space-borne hyperspectral imaging spectrometers can be leveraged to routinely retrieve vertical cloud droplet profiles, rather than a single column weighted average, enhancing our understanding of cloud droplet growth, cloud-aerosol interactions, and cloud albedo.

The Climate Absolute Radiance and Refractivity Earth Observatory (CLARREO) Pathfinder (CPF) is an upcoming space-borne hyperspectral imaging spectrometer that will soon deploy on the International Space Station (ISS), which occupies a near-circular orbit with an average altitude of 417 km and an inclination of 51.6° [28]. The full CPF swath width at ISS altitude (417 km) is 70 km, comprised of 480 cross-track pixels [28]. The CPF Hyperspectral Imager for Climate Science (HySICS) is the most radiometrically accurate space-borne hyperspectral imager ever built for Earth science, with a radiometric uncertainty of 0.3% [28]. HySICS will measure scattered radiation contiguously from 350 nm to 2300 nm with a spectral resolution of 6 nm and Nyquist sampling [30]. At the ISS altitude and with an instantaneous field of view of 0.02°, HySICS will have a spatial sampling of about 0.5 km at nadir after three-pixel binning. In this paper, we expanded on the vertical droplet profile retrieval method of Buggee and Pilewskie [20] to utilize the enhanced radiometric accuracy and spectral resolution of CPF and applied it to real hyperspectral measurements for the first time.

The research herein builds upon previous studies in several ways. First, we build on the results of King and Vaughan [23], who were the first to demonstrate the additional Shannon information content, the quantification of a measurement's ability to change an initial guess, throughout the visible and shortwave infrared that may improve the retrieval of a droplet profile. We show that retrieving the integrated water vapor content above cloud improves the accuracy of the retrieved droplet profile. Second, all previous studies attempting to retrieve a droplet profile using optimal estimation methods assumed the retrieved variables were uncorrelated [20,22,23]. This is not the best representation of the true physics of the problem. In this study, we used in situ measurements from the Variability of the American Monsoon Systems Ocean-Cloud-Atmosphere-Land Study Regional Experiment (VOCALS-REx) field campaign to compute the a priori covariance matrix, including the off-diagonal elements [31]. Third, previous studies showed the importance of transforming the optimal estimate framework into logarithmic space when the retrieved variables more closely follow lognormal distributions [32,33]. We show that the effective radii at cloud top and base, the cloud optical thickness and the integrated water vapor content above cloud are inherently non-negative and follow log-normal distributions more closely than normal distributions, warranting this transformation. Lastly, radiative transfer calculations using the HySICS slit functions were constructed using VOCALS-REx in situ data. By using measurements from the Cloud Droplet Probe (CDP) to define the cloud microphysics, and data from the European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis version 5 (ERA5) to define vertical profiles of temperature and water vapor concentration, quasi-realistic reflectances were computed [34,35]. This enabled a systematic study of forward model uncertainty by introducing components one at a time to understand their individual impacts on retrieval performance. The droplet profile retrieval is then tested on real hyperspectral measurements from the Earth Surface Mineral Dust Source Investigation (EMIT) imaging spectrometer and compared against retrievals using measurements from the

 doi:10.20944/preprints202603.2506.v1

instruments onboard the Aqua satellite [36]. With a similar spectral range, albeit coarser spectral resolution, the EMIT instrument acts as a surrogate for the upcoming CPF instrument.

Section 2 provides background information on previous studies and how they will be leveraged for the analysis within this paper. Section 3 describes the details the optimal estimation retrieval method and the construction of simulated HySICS reflectances. Section 3 also describes the selection of EMIT measurements with spatially and temporally coincident observations from instruments onboard the Aqua satellite, along with modifications to the retrieval method for application to real hyperspectral data. Section 4 presents results from applying our retrieval to simulated HySICS data, demonstrating the importance of simultaneously retrieving above-cloud integrated water vapor and quantifying retrieval performance using in situ-defined cloud profiles. Section 4 also presents results from applying our retrieval to EMIT measurements, with comparisons to coincident satellite observations. Section 5 provides a discussion of the results and concluding remarks on the utility of hyperspectral measurements for retrieving vertical cloud structure and above-cloud integrated water vapor.

2. Materials and Methods

2.1. Retrieval Methodology

We employ the Gauss-Newton iterative retrieval algorithm developed by Buggee and Pilewskie [20], an optimal estimation method grounded in Bayesian probability [37]. This approach combines prior information about retrieved variables with spectral measurements, each weighted by their respective uncertainties through covariance matrices. With every iteration, the algorithm updates the state vector by balancing the difference between estimated and true reflectances against deviations from the prior. For the complete mathematical formulation, see Buggee and Pilewskie [20].

At each iteration, a new state vector, $\boldsymbol{x_{i+1}} = (\boldsymbol{r_{top}}, \boldsymbol{r_{bot}}, \boldsymbol{\tau_c}, \boldsymbol{IWV_{ac}})$, is estimated:

$$\boldsymbol{x_{i+1}} = \boldsymbol{x_i} + \left(\mathbf{S_a^{-1}} + \mathbf{K_i^T}\ \mathbf{S_\epsilon^{-1}}\ \mathbf{K_i}\right)^{-1}\ [\mathbf{K_i^T}\ \mathbf{S_\epsilon^{-1}}\ (\boldsymbol{m} - \boldsymbol{R}(\boldsymbol{x_i})) + \mathbf{S_a}(\boldsymbol{x_i} - \boldsymbol{x_a})] \quad (1)$$

where matrices are indicated in capitalized boldface, and vectors are indicated in lowercase boldface. $\boldsymbol{x_i}$ is the state vector estimate of the $\boldsymbol{i^{th}}$ iteration, $\boldsymbol{x_a}$ is the a priori state vector, $\boldsymbol{S_a}$ is the a priori covariance matrix, $\boldsymbol{K_i}$ is the Jacobian matrix of $\boldsymbol{R}(\boldsymbol{x_i})$, and $\boldsymbol{S_\epsilon}$ is the measurement covariance matrix. The state vector, $\boldsymbol{x_{i+1}} = (\boldsymbol{r_{top}}, \boldsymbol{r_{bot}}, \boldsymbol{\tau_c}, \boldsymbol{IWV_{ac}})$, represents the effective radii at cloud top and base, the cloud optical thickness, and the integrated water vapor content above cloud, respectively. The measurement covariance matrix combines measurement and forward model uncertainties: $\boldsymbol{S_\epsilon} = \boldsymbol{S_m} + \boldsymbol{S_{fm}}$ [38]. $\boldsymbol{S_m}$ defines the measurement uncertainty at each wavelength and the correlation between measurements at different wavelengths. $\boldsymbol{S_{fm}}$defines the forward model uncertainty, which can be separated into two categories: sources proportional to the measured signal and uncertainties in surface reflectance [38]. The forward model, $\boldsymbol{R}$, computes reflectance at a set of wavelengths for a given cloud state, $\boldsymbol{x_i}$. The Jacobian represents the change in reflectance resulting from a perturbation in each state variable. Equation (1) balances several competing factors during each iteration: the difference between measured and computed reflectances $(\boldsymbol{m} - \boldsymbol{R}(\boldsymbol{x_i}))$, the difference between the current state estimate and the a priori $(\boldsymbol{x_i} - \boldsymbol{x_a})$, and the rate of change of the estimated measurements with respect to the current state variable $\mathbf{K_i} = \boldsymbol{\nabla R}(\boldsymbol{x_i})$.

We use the same convergence criteria as Buggee and Pilewskie [20]. The cost function, commonly denoted as $\boldsymbol{J}$, will refer to the $\boldsymbol{l^2}$-norm (the root-sum-square) of the difference between the forward model reflectances and the true measurements. Convergence is satisfied in two ways: (1) the percent difference of the cost function between two successive iterations falls below 3%, or (2) the cost function becomes less than or equal to the $\boldsymbol{l^2}$-norm of the total uncertainty (measurement and forward model), and is lower than the previous iteration [20,39]. Once convergence occurs, the posterior covariance matrix is computed. The uncertainties of the retrieved variables are the square root of the main diagonal [37]. It is important to note that this method does not find the global

 doi:10.20944/preprints202603.2506.v1

minimum, but rather a local minimum that depends strongly on the prior information and initial guess [33].

Several notable differences from the method of Buggee and Pilewskie [20] have been made for application with real hyperspectral data. (1) We updated the retrieval to estimate the integrated water vapor content above cloud in addition to the droplet sizes at cloud top and cloud base and cloud optical thickness. (2) Using in situ measurements from the VOCALS-REx field campaign, an a priori covariance matrix with off-diagonal elements was computed. (3) We transformed the Gauss-Newton problem into logarithmic space because the desired variables are inherently non-negative [40]. (4) A forward model covariance matrix was constructed and transformed into measurement space through a forward model Jacobian [33,37].

2.2. Construction of Simulated HySICS Reflectances and Forward Model Assumptions

Before applying our algorithm to real data, we tested our method on simulated HySICS reflectances to quantify retrieval accuracy when the forward model and radiometric uncertainty are precisely known. We first computed a set of reflectances in a manner similar to the synthetic data generated by King and Vaughan [23]. Cloud droplet profiles were modeled as smooth, monotonic, and adiabatic following Platnick [19], with 20 plane-parallel cloud layers, each with a gamma droplet distribution. The effective variance, which defines the width and asymmetry of the distribution, was defined as the column-average of the in situ measurements from the VOCALS-REx field campaign. We note that the assumption of a vertically constant effective variance is a simplification that is not valid for all clouds, and that in situ measurements have found that droplet distributions tend to narrow towards cloud top [5].

libRadtran was used to compute top-of-atmosphere reflected radiance spectra using 1-D DISORT [41,42]. Reflectance from 350 nm to 2300 nm at the HySICS sampling resolution of 3 nm was estimated by convolving the radiance spectrum with the HySICS slit functions and normalizing by the incident solar irradiance. The TSIS-1 hybrid solar reference spectrum was used [43]. We assumed a uniform radiometric uncertainty of 0.3% across all spectral channels [28]. This dataset was used to isolate the effect of incorrect above-cloud integrated water vapor assumptions on retrieval performance in the absence of other forward model uncertainties (Sect. 3.1).

We generated a second set of HySICS reflectances using VOCALS-REx in situ measurements to simulate more realistic clouds [31]. 69 in situ vertical profiles of non-precipitating clouds where a C-130 aircraft sampled the droplet distribution between cloud base and top were identified. This data set will be referred to as the in situ *measured profiles* for the remainder of the paper. Reflectances were computed from these profiles by defining each cloud using the in situ measured cloud boundaries (base and top) and droplet size distribution, and the in situ derived cloud optical thickness. We used definitions similar to [6], defining cloud top and base with a minimum liquid water content of 0.03 $\boldsymbol{g/m^3}$ and a minimum total droplet number concentration threshold of 25 $\boldsymbol{cm^{-3}}$. Therefore, cloud top and base were identified as the minimum and maximum altitudes where both criteria were satisfied.

For each in situ measured profile, the above cloud integrated water vapor content was calculated by integrating ERA5 water vapor concentration from the VOCALS-REx in situ measured cloud top altitude to the top of atmosphere. ERA5 combines historical observations from satellites, weather stations, and other sources with numerical weather prediction models through data assimilation to create a spatially and temporally complete, physically consistent reconstruction of the Earth's atmosphere from 1940 to present [35]. For each in situ measured profile, we found the ERA5 dataset closest in time and space. We note that the use of radiosonde data, which was extensive throughout the VOCALS-REx field campaign, was explored. However, the average spatial distance and time difference between the in situ measured profiles and the closest radiosonde measurement was 735 km and 63 minutes, respectively, whereas for ERA5 data these values were 9 km and 14 minutes, respectively. We chose to use ERA5 data over radiosonde measurements due to the higher temporal and spatial sampling. Often the same radiosonde file was found to be the closest in time and space

 doi:10.20944/preprints202603.2506.v1

for several in situ measured profiles, leading to redundant $\boldsymbol{IWV_{ac}}$ measurements. The choice between radiosonde and ERA5 data significantly impacted the a priori covariance matrix (see Sect. 2.5). Radiosonde measurements of above-cloud integrated water vapor averaged four times higher than corresponding ERA5 estimates for the same in situ profiles, which may be due to the large spatial and temporal offsets between the radiosonde and in situ measurements.

Our analysis was restricted to non-precipitating marine stratus and stratocumulus clouds over ocean with cloud optical thickness of at least three. The optical depth threshold was chosen to avoid a non-unique solution of effective radii for optically thin clouds [15]. We adopted some assumptions from the forward model used in the MODIS Collection 6 cloud optical properties retrieval algorithm. These included assuming maritime aerosols with an aerosol optical depth of 0.1 and using the Cox-Munk surface bidirectional reflectance model to account for the impact on reflectances over ocean with a particular wind speed and direction [17,44]. We used the US standard atmosphere to define vertical profiles of all atmospheric gases other than water vapor [45]. In addition, reflectance at one wavelength was assumed to be independent from reflectance at every other wavelength, resulting in a diagonal covariance matrix where each element along the main diagonal was defined as the square of the reflectance uncertainty at that wavelength. A diagonal covariance matrix does not reflect the true nature of the instrument, but it is a common simplifying assumption [22,23,38].

2.3. Retrieval of Above-Cloud Integrated Water Vapor Content

Historically, space-borne spectrometers used to infer cloud optical properties sampled regions of reflected solar radiation where water vapor weakly absorbs [46–48]. The advantage of avoiding spectral regions with appreciable absorption by water vapor is that accurate knowledge of the column water vapor amount is not crucial. However, absorption by water vapor is ubiquitous, with many weak to strong lines spanning the solar spectrum. King and Vaughan [23] showed there was additional information content in the wings of water vapor absorption bands that could be leveraged to improve the estimate of droplet size near cloud top. To leverage all CPF spectral measurements for the retrieval of a droplet profile, the above-cloud integrated water vapor content needs to be retrieved simultaneously.

Figure 1 shows a set of weighting functions derived from radiative transfer calculations using the HySICS slit functions. These weighting functions demonstrate the importance of including measurements in the wings of water vapor absorption regions for droplet profile retrievals. Weighting functions represent the conditional probability of a photon scattering in the upward direction at cloud top, given that it penetrated to some maximum optical depth, $\boldsymbol{\tau}$, within a cloud with optical thickness $\boldsymbol{\tau_c}$. A selection of wavelengths were chosen to emphasize a spread in the peak value. The set of weighting functions shown in Figure 1 demonstrate a larger degree of relative orthogonality than a set that doesn't sample the wings of water vapor absorption features (compare with Figure 1 from Buggee and Pilewskie [20]). Wavelengths close to a strong water vapor absorption feature shifts the proportion of reflected photons toward cloud top, whereas wavelengths with little water vapor and liquid water absorption penetrate further into the cloud.

This basis for retrieving the integrated water vapor content above cloud from backscattered light was outlined by Albert et al. [49], showing that marine stratus and stratocumulus have two opposing contributions to the total absorption by water vapor. The lowest layers of the atmosphere, where most of the water vapor resides, do not contribute to the reflected signal, thus lessening the total absorption (this effect is less prominent for optically thin clouds). On the other hand, multiple scattering increases the total radiation absorbed. Albert et al. [49] computed vertical weighting functions to determine the contribution of different cloud layers to the total absorption by water vapor, and found that the total absorption for optically thick clouds over ocean was primarily due to the above-cloud integrated water vapor content.

 doi:10.20944/preprints202603.2506.v1

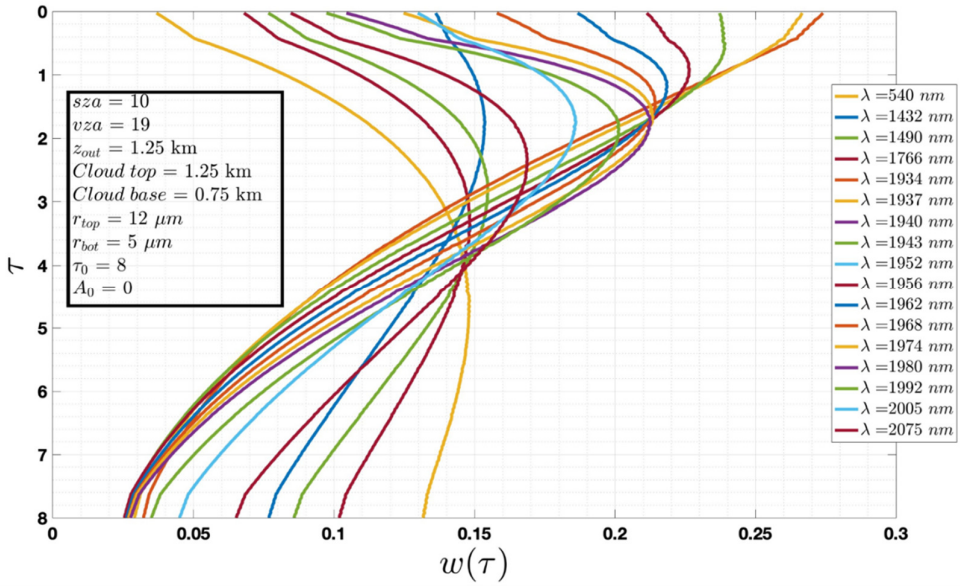


Figure 1. A set of weighting functions derived from simulated HySICS reflectances. The center wavelength of each spectral channel is listed in the legend. The text box on the left describes the solar and viewing zenith angles, the cloud top and base height, the effective radii and cloud top and base, the cloud optical thickness and the surface albedo.

The Gauss-Newton optimal estimation method requires an initial guess. Thus, we developed a multi-spectral method for estimating $\boldsymbol{IWV_{ac}}$ based on the differential absorption method by Albert et al. [49]. First, we ran our multi-spectral TBLUT retrieval to estimate the effective radius and cloud optical thickness, which is based on the algorithm described by Amarasinghe et al. [17]. Using these results, $\boldsymbol{IWV_{ac}}$ was estimated in a similar manner using a lookup table of computed reflectances at 900 $\boldsymbol{nm}$, 955 $\boldsymbol{nm}$, and 1127 $\boldsymbol{nm}$ for $\boldsymbol{IWV_{ac}}$ values between 0 and 40 $\boldsymbol{mm}$. These wavelengths provide an adequate spread in the bulk absorption coefficient of water vapor. The solution was determined by solving for the minimum least-squares fit between the look-up table and the simulated HySICS reflectances. This multi-spectral estimate was subsequently used as the initial guess in the hyperspectral retrieval of $\boldsymbol{IWV_{ac}}$, alongside $\boldsymbol{r_{top}}$, $\boldsymbol{r_{bot}}$, and $\boldsymbol{\tau_c}$.

2.4. Transforming the Gauss-Newton Formalism Into Log-Space

A core assumption of the Gauss-Newton iterative method is that the prior and measurement uncertainties follow a normal distribution [50]. Previous work has shown that the distribution of many positively-defined atmospheric variables more closely follow a log-normal distribution [32,33]. Maahn et al. [33] showed that transforming variables into log-space when they more closely follow a log-normal distribution resulted in an increased convergence rate from 86% to 100%.

To determine whether the retrieved variables better fit a normal or log-normal distribution, we created an a priori data set using VOCALS-REx in situ measurements to define $\boldsymbol{r_{top}}$, $\boldsymbol{r_{bot}}$, and $\boldsymbol{\tau_c}$, and ERA5 reanalysis data to define $\boldsymbol{IWV_{ac}}$ (see Sect. 2.1). $\boldsymbol{r_{top}}$ and $\boldsymbol{r_{bot}}$ were defined as the average effective radius over the top and bottom 25% of each situ measured vertical profile with respect to the normalized altitude within cloud, respectively. Averages were used because many of the in situ measurements showed rapid variation in droplet size near cloud top and base. By using an average to define the a priori values of $\boldsymbol{r_{top}}$ and $\boldsymbol{r_{bot}}$ rather than a single measurement, we found that the retrieved $\boldsymbol{LWP}$ was closer to the true in situ value. The cloud optical thickness was determined from the in situ measurement of the droplet size distribution: $\boldsymbol{\tau_c^{\lambda} = \int \int \sigma_e^{\lambda}(r)\, n(r,z)\, dr\, dz}$, where $\boldsymbol{\sigma_e^{\lambda}}$ is the spectral extinction cross section of liquid water, and $\boldsymbol{n(r,z)}$ is the droplet size distribution. The

 doi:10.20944/preprints202603.2506.v1

integrated water vapor content was derived from ERA5 data as the integral of water vapor concentration from the in-situ measured cloud top height to the top of the atmosphere.

Figure 2.a and 2.b show quantile-quantile plots for all four variables in linear and logarithmic space, respectively, with an R^2 value to indicate how well each variable fits a normal (panel a) and log-normal (panel b) distribution. All four variables have a higher R^2 value in panel (b), indicating that they more closely follow a log-normal distribution. Therefore, we make the following transformations to the measurement and state vectors:

$$\boldsymbol{m}' = \ln(\boldsymbol{m}), \qquad \boldsymbol{x}'_a = \ln(\boldsymbol{x}_a), \qquad \boldsymbol{x}'_i = \ln(\boldsymbol{x}_i) \tag{2}$$

The measurement and prior covariance matrices are transformed into log space by first taking the log of the prior data set and the measurements:

$$\boldsymbol{S}'_a = \boldsymbol{Cov}(\ln([\boldsymbol{r}_{top}, \boldsymbol{r}_{bot}, \boldsymbol{\tau}_c, \boldsymbol{IWV}_{ac}])) \tag{3}$$

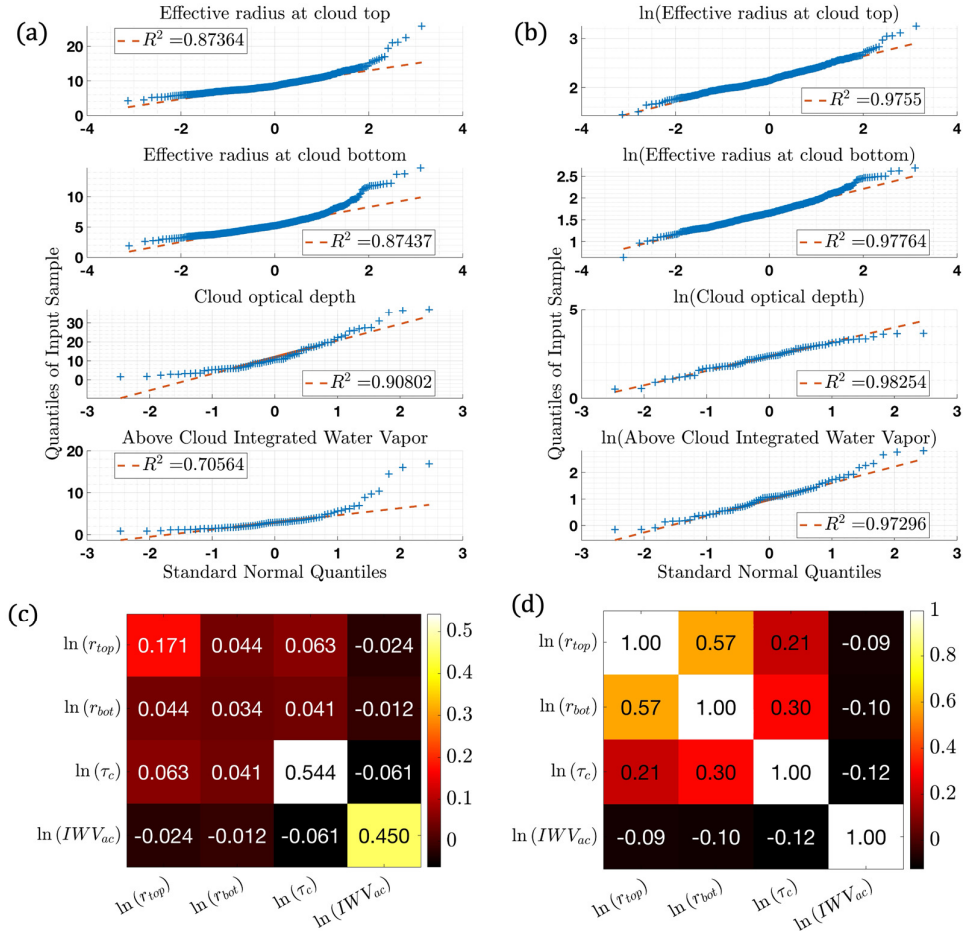
$$\boldsymbol{S}'_\epsilon = \boldsymbol{Cov}(\ln([\boldsymbol{m}_{\lambda_1}, \boldsymbol{m}_{\lambda_2}, \dots, \boldsymbol{m}_{\lambda_M}]))$$

We compute the Jacobian in log-space as follows [51]:

$$\boldsymbol{K}' = \frac{\partial \boldsymbol{m}'}{\partial \boldsymbol{x}'} = \frac{\partial \boldsymbol{m}}{\partial \boldsymbol{x}} \frac{\boldsymbol{x}}{\boldsymbol{m}} = \boldsymbol{K} \frac{\boldsymbol{x}}{\boldsymbol{m}} \tag{4}$$

Reformulating Equation (1) into log space, we solve for the logarithm of the state vector at each iteration:

$$\boldsymbol{x}'_{i+1} = \boldsymbol{x}'_i + \left(\mathbf{S}'_a{}^{-1} + \mathbf{K}'_i{}^{T}\, \mathbf{S}'_\epsilon{}^{-1}\, \mathbf{K}'_i\right)^{-1} \left[\mathbf{K}'_i{}^{T}\, \mathbf{S}'_\epsilon{}^{-1}\, (\boldsymbol{m}' - \boldsymbol{R}(\boldsymbol{x}'_i)) + \mathbf{S}'_a(\boldsymbol{x}'_i - \boldsymbol{x}'_a)\right] \tag{5}$$



 doi:10.20944/preprints202603.2506.v1

Figure 2. Panel (a) shows the Quantile-Quantile plot for all four retrieved variables. Panel (b) shows the same but for the natural logarithm of the in situ measurements. An $\boldsymbol{R^2}$ value is listed for each Quantile-Quantile plot, where a value of 1 indicates a perfect fit. Panel (c) shows the a priori covariance matrix in log space (Equation 3). Panel (d) shows the correlation coefficient.

2.5. Computation of the A Priori Covariance Matrix

In the Gauss-Newton formulation, the prior is defined by a Gaussian distribution with mean, $\boldsymbol{x_a}$, and covariance, $\boldsymbol{S_a}$ [37]. At each iteration, the difference between the current guess and the prior is computed. Solutions far from the prior are penalized. If the true value is not well represented by the prior, the result is likely to be inaccurate [33]. We used the prior dataset described in Sect. 2.3 to determine the a priori covariance matrix.

Figure 2.c shows the resulting matrix in log space. Values along the main diagonal show the variance of the effective radius at cloud top and cloud base, the cloud optical thickness, and the above-cloud precipitable water, respectively. The off-diagonal elements show the covariance between variables. To better interpret the results, Figure 2.d shows the correlation coefficient matrix, the covariance matrix normalized by the variance. Values in this matrix are often referred to as the Pearson correlation: $\boldsymbol{r} = {\boldsymbol{Cov}(\boldsymbol{X},\boldsymbol{Y})}/{\boldsymbol{\sigma_Y \sigma_X}}$, where $\boldsymbol{\sigma_X}$ is the variance in the random variable $\boldsymbol{X}$ [52]. Values range from [-1,1], implying a strong, positive linear and strong, negative linear relationship, respectively. The square of the Pearson correlation, sometimes called the coefficient of determination, defines the proportion of variance in one variable that can be explained by the linear relationship with the other [52].

The effective radii at cloud top and base have the strongest positive Pearson correlation of 0.57 and a coefficient of determination of 0.325. This suggests some vertical coherence between the droplet size at cloud top and base, such as a quasi-adiabatic structure. However, with only 32.5% of the variance in $\boldsymbol{r_{top}}$ explained by $\boldsymbol{r_{bot}}$ (and vice vera), decoupling processes, such as cloud top entrainment, likely breakdown this coherence. The Pearson correlations between cloud optical thickness and effective radii at cloud top and base are 0.21 and 0.3, with coefficients of determination of 0.044 and 0.09, respectively. Greater numbers of cloud condensation nuclei (CCN) often result in many small droplets, whereas small amounts of CCN often result in fewer large droplets. Thus, we expected a negative correlation between droplet size and cloud optical thickness, especially at cloud base where the droplet number concentration is determined by the available CCN. Interestingly, all Pearson correlations with integrated water vapor above cloud were of small magnitude and negative. This could be due to the mismatch of instruments between variables, where ERA5 reanalysis data was used to define $\boldsymbol{IWV_{ac}}$, and $\boldsymbol{r_{top}}$, $\boldsymbol{r_{bot}}$, and $\boldsymbol{\tau_c}$ were defined using measurements from a CDP instrument With only 69 in situ measured droplet profiles and spatial, temporal and physical differences between the CDP measurements and ERA5 reanalysis data, some of these correlations may be tenuous. The mechanisms behind the relationships described above require further investigation with larger in situ data sets before any confident conclusions can be reached.

2.6. Forward Model Uncertainty and Jacobian

Forward model uncertainty is often ignored despite having an impact on retrieval uncertainty [37]. Indeed, with HySICS radiometric uncertainty being only 0.3%, retrieval uncertainty is likely limited by forward model uncertainty. Several forward model assumptions that represent some of the largest sources of uncertainty according to previous studies were investigated: cloud top height uncertainty, deviations from the assumed smooth, monotonic adiabatic profile, and uncertainty in effective variance of the droplet size distribution [38,53]. We ignored surface reflectance uncertainty since we only considered clouds over ocean with an optical thickness of at least three. Additionally, forward model uncertainty due to assumed vertical profiles of temperature, pressure and atmospheric gasses deviating from the true environmental profiles was not incorporated.

 doi:10.20944/preprints202603.2506.v1

Forward model uncertainty in cloud top height was defined by assuming a value of 1.2 $\boldsymbol{km}$ for all retrievals using simulated HySICS reflectances, with a standard deviation of 0.218 $\boldsymbol{km}$. The mean and standard deviation were determined from VOCALS-REx in situ measured droplet profiles described in Sect. 2.1. From these measurements, it was determined that cloud top height followed a log-normal distribution. Forward model uncertainty due to the assumption of a smooth, monotonic adiabatic droplet profile was also determined through the use of VOCALS-REx in-situ measurements. Uncertainty exists in our reflectance calculations due to deviations between the true droplet profile and the retrieved smooth and monotonic profile (see Figure 4). The forward model used 20 plane-parallel layers to define clouds with vertical variation. For each layer, a mean and standard deviation effective radius was defined using a log-normal distribution, following the results from Figure 2. The mean was the value of the effective radius at that layer for the current iteration. The standard deviation was 2 $\boldsymbol{\mu m}$ for the two cloud layers at cloud top and bottom. Every other layer in between was defined with a standard deviation of 1 $\boldsymbol{\mu m}$. These values were determined from the VOCALS-REx in situ measurements used to create Figure 2.

To define the uncertainty of the assumed effective variance of the droplet size distribution, each in situ measured droplet profile was normalized along the vertical dimension and divided into 20 layers. With the ensemble of all measurements for each layer, a distribution shape and effective were solved for. Once again, we found a log-normal distribution best fit the effective variance in each layer. Similar to analysis reported for other field campaigns, we found that the effective variance decreases from cloud base to cloud top, demonstrating that the droplet size distribution narrows with respect to altitude [5]. We found the mean value of the effective variance at cloud top to be about 0.045, whereas the mean value at cloud base was about 0.1. While vertical variation in the effective variance can be used to define liquid water clouds in the libRadtran radiative transfer package, solving the radiative transfer equation with DISORT requires a single assumption of distribution shape and width. libRadtran uses a precomputed table to convert cloud properties of droplet size and liquid water content to Mie properties in order to solve the radiative transfer equation, and this table is constructed assuming a single effective variance. Therefore, the mean value of all vertical profiles of effective variance, which was 0.055, was assumed in all our radiative transfer calculations. The standard deviation of effective variance at each of the 20 cloud levels was used to construct the forward model covariance matrix. The values ranged from 0.05 at cloud bottom to 0.035 at cloud top.

Using the lognormal distributions that were found for the effective radius and effective variance at normalized cloud layers, and for the cloud top height, we computed the covariance matrix using Equation (3). To transform the forward model covariance matrix, $\boldsymbol{S_{fm}}$, into measurement space, such that the measurement and forward model covariance matrices can be summed, $\boldsymbol{S_\epsilon = S_m + S'_{fm}}$, a forward model Jacobian is required:

$$\boldsymbol{S'_{fm} = K'^{\,T}_{fm}\, S_{fm}\, K'_{fm}} \quad (6)$$

2.7. Applying Retrieval to Real EMIT Measurements

The retrieval method described in Sections 2.1 – 2.6 was carefully developed for application to real measurements. By introducing forward model uncertainties one at a time, we were able to study the impact of each assumption in the model. The culmination was the application of our retrieval method to quasi-realistic simulated HySICS reflectances. To test this development, we applied the retrieval to data from the EMIT hyperspectral imager flying on the ISS [36]. The only aspects of the retrieval method outlined in Sections 2.1 - 2.6 that were altered for application to EMIT measurements were the measurement covariance matrix and a few forward model assumptions. Only EMIT observations with spatially coincident observations from instruments onboard the Aqua satellite that were recorded within 10 minutes of one another were used in order to have a means of comparison.

 doi:10.20944/preprints202603.2506.v1

2.8. EMIT Measurements Overlapping with Aqua

We selected EMIT observations of marine stratus and stratocumulus over the southeast Pacific off the coast of Chile, the same region surveyed during the VOCALS-REx field campaign, in order to confidently use the same a priori covariance matrix developed in Sect. 2.5. We adapted the Langley Automated Sensor Inter-calibration System (LASICS) tool [54] to find EMIT observations that overlap with MODIS, AMSR-E (the Advanced Microwave Scanning Radiometer for EOS), and AIRS (the Atmospheric Infrared Sounder). We searched the EMIT data record from its start, on 1 January 2023, to 25 December 2025.

MODIS measurements of spectral reflectance in the visible and shortwave infrared are used to retrieve cloud effective radius, cloud optical thickness, and above-cloud precipitable water [53,55] at a spatial resolution of approximately 1 $\boldsymbol{km}$ [53]. The comparison between MODIS-retrieved effective radius and the retrieved droplet profile using the hyperspectral method with EMIT data is not straightforward because the two instruments have different sampling volumes. Table 1 summarizes the spatial resolution for the EMIT spectral measurements and retrieved products from MODIS, AMSR-E and AIRS. With a spatial resolution of 60 meters, EMIT is capable of detecting spatial heterogeneity within a MODIS pixel. Therefore, the comparison between the retrieved droplet profile from a single EMIT pixel with the retrieved effective radius and cloud optical thickness from the overlapping MODIS pixel can be misleading. To address this, we randomly selected 30 EMIT pixels within a single MODIS pixel and ran the hyperspectral retrieval. We also compared the MODIS-retrieved cloud optical thickness and liquid water path with our hyperspectral retrieval.

Table 1. Each instrument used for the analysis in Sect. 3, along with a description of what each instrument was used for and the corresponding spatial resolution.

Instrument	Usage	Spatial Resolution at Nadir
EMIT	Hyperspectral retrieval	60 m (measurement resolution)
MODIS	Comparison of τ_c, LWP, and pw_{ac}	~1 km (retrieval resolution)
AMSR-E	Comparison of LWP, and pw_{ac}	~10 km (retrieval resolution)
AIRS	Comparison of pw_{ac}	~50 km (retrieval resolution)

Microwave emission measurements from 12 channels between 6.9 and 89 GHz from the AMSR-E instrument are used to estimate cloud liquid water path [56,57]. Liquid water path is proportional to the absorption of microwave emission along the path, and these retrievals are less sensitive to assumptions about droplet size distribution and droplet vertical structure than retrievals using visible and short wave infrared radiation. This is one reason why retrievals of liquid water path using microwave emission are often used to verify cloud microphysics parameterizations within Earth System Models [58,59]. The spatial resolution of the AMSR-E/AMSR2 Unified L2B Global Swath Ocean Product is approximately 10 km along track × 5 km across track, resulting in a much larger sampling volume than that of EMIT [57]. Our retrieval of liquid water path was compared with the AMSR-E retrievals when the microwave retrieval quality was flagged as acceptable. However, we often found that the closest AMSR-E pixel was flagged as "Land/Bad Pixel". A previous study by Duncan et al. [60] found that land intrusions of 1-2% impact the retrieval, and then even side-lobe emission outside the nominal pixel field of view can be significant for observations near the coast, which is where all of the EMIT observations used in this analysis were located. The number of available comparisons was so low as to be statistically meaningless. However, we used the few examples we had as a guide while developing the retrieval. Future work will use a larger data set with more direct comparisons to microwave retrievals.

The AIRS instrument samples upwelling infrared radiation between 3.8 $\boldsymbol{\mu m}$ and 15 $\boldsymbol{\mu m}$ within 2378 spectral channels [61]. These measurements are used to retrieve vertical profiles of temperature and water vapor concentration, which are part of the AIRS version 7 Level 2 data products, have a spatial resolution of 50 $\boldsymbol{km}$ × 50 $\boldsymbol{km}$ at nadir [62]. We used AIRS retrievals to define the vertical profile of temperature and water vapor concentration within the forward model of our retrieval. We

 doi:10.20944/preprints202603.2506.v1

also used the AIRS water vapor concentration retrieval as another means of comparison with the hyperspectral retrieval of above-cloud precipitable water.

Using MODIS retrievals, we filtered for observations of liquid water clouds with a cloud optical thickness greater than three and a maximum horizontal inhomogeneity index of 2.5. Zhang and Platnick [63] showed the inhomogeneity index, the ratio of the standard deviation of the 16 sub-pixels of MODIS-measured reflectance at 250 m spatial resolution to the mean, was a strong indicator of whether horizontal inhomogeneity affected the retrieval of effective radius. The authors concluded that effective radius retrievals were biased when the cloud had an inhomogeneity index greater than 0.3 [63]. Homogeneity indices less than 0.1 are indicative of homogeneous scattering layers with negligible 3-D radiative biases [63]. We chose an inhomogeneity threshold of 2.5 because we found very few coincident observations between EMIT and MODIS that met the smaller thresholds defined by [63]. This likely introduced 3-D biases in the retrieval. Future work will explore mitigation strategies for these biases.

2.9. Covariance Matrix and Forward Model Assumptions Used in EMIT Retrieval

We used 257 of the 285 EMIT spectral channels for the retrieval, estimating the same variables defined in Sect. 2.1: the effective radii at cloud top and base, the cloud optical thickness, and the above-cloud integrated water vapor content. On-orbit EMIT measurements have been compared with spectral measurements from two RadCalNet calibration and validation reference sites, revealing discrepancy at wavelengths less than 500 $\boldsymbol{nm}$, potentially due to calibration error [25]. An additional discrepancy was found near 1290 $\boldsymbol{nm}$ at the seam of two order sorting filters [25]. For these reasons, EMIT measurements less than 500 $\boldsymbol{nm}$ and between 1245 and 1320 $\boldsymbol{nm}$ were omitted in the analysis. The EMIT spectral measurements were assumed to be independent from one another, resulting in a diagonal measurement covariance matrix where each element along the main diagonal was defined as the square of the reflectance uncertainty for that channel. Since the uncertainty of each EMIT spectral channel is not reported in the mission documentation, we assumed a radiometric uncertainty of 4% for all channels to account for instrument and forward model uncertainty, such as the uncertainty due to incorrect assumptions of the vertical profile of water vapor concentration, which was not captured in our forward model. We note that a diagonal covariance matrix does not reflect the true nature of partial correlation between radiances across the spectrum, but it is a common simplifying assumption [22,23,38]. The a priori covariance matrix defines the distribution of possible values for each variable, thus imparting vital information for the retrieval. We used the same a priori matrix for the retrieved variables described in Section 2.4 using a set of VOCALS-REx in situ observations.

For each coincident observation, we used the MODIS-retrieved cloud top height [53] to constrain our forward model. The forward model covariance matrix accounts for the uncertainty in this quantity by using the retrieval uncertainty and the same in situ-derived statistics reported in Sect. 2.6.

The AIRS Level 2 retrieval products include a vertical profile of estimated water vapor concentration and temperature [62]. For overlapping observations between EMIT and AIRS, these profiles are used to define the vertical profiles of water vapor and temperature within our forward model. The AIRS retrieval estimates temperatures between the surface and the stratosphere at values ranging from 1100 to 10 $\boldsymbol{mb}$, whereas the retrieval of water vapor concentration is estimated between the surface and the upper troposphere. For cloudy scenes, these retrievals are unable to estimate water vapor amount below the cloud due to the high absorption by clouds for many of the AIRS spectral channels. For altitudes outside of the AIRS retrieval range, or when an AIRS retrieval was not available, we used the US standard atmosphere to define water vapor concentration and temperature [45].

 doi:10.20944/preprints202603.2506.v1

3. Results

3.1. Retrievals Using Simulated HySICS Reflectances

First, we applied our retrieval to a set of simulated HySICS reflectances from clouds with smooth, monotonic droplet profiles, removing forward model uncertainty caused by deviations from these ideal profiles (see Sect. 2.4). We supplied the forward model with all the inputs used to create the simulated measurements, resulting in negligible forward model uncertainty. With these simulations, we retrieved r_{top}, r_{bot} and τ_c, and made an assumption about the total column precipitable water above cloud that is common in most retrievals of cloud droplet size [17,22,38,46]. Next, we retrieved IWV_{ac}, r_{top}, r_{bot} and τ_c simultaneously. This comparison highlights the impact of assuming the total column precipitable water amount on the retrieval of droplet size at cloud top and base. Lastly, we applied our retrieval to the set of simulated measurements that are defined using the VOCAL-REx in situ measurements, described in section 2.1. These measurements simulate more realistic clouds and the uncertainties that are encountered when using real data. We quantify the performance of the retrieval by computing the difference in LWP between our hyperspectral retrieval and the TBLUT estimate, the modified TBLUT estimate by Wood and Hartmann [64], and the true in situ value.

A proper comparison between the hyperspectral retrieval and the in situ measurements should include uncertainty in measured droplet size. However, the CDP droplet size uncertainty estimate is attributed to several factors that make it difficult to quantify [34]. Droplets that pass through the edges of the sampling area tend to have much higher uncertainty than droplets that pass through the center. Uncertainty due to coincidence, where multiple droplets pass through the sampling area within the sampling time of the detecting optics, is challenging to estimate because it depends on droplet size, particle concentration, and transit location within the sampling area. There are also limitations to the size resolution of the instrument due to the non-monotonic relationship between droplet size and the scattered laser light signal [34]. Lance et al. [34] used a water droplet generating system to determine the sizing accuracy of the CDP instrument. Using their results, we simplified the CDP measurement uncertainty for this analysis by defining an uncertainty of 20% for effective radii below 5 μm, 15% for radii greater than 5 μm and less than 10 μm, and 10% for radii greater than or equal to 10 μm.

3.1.1. How the Above-Cloud Integrated Water Vapor Impacts the Droplet Profile Retrieval

Figure 3 shows the impact of different assumptions of above-cloud integrated water vapor amounts on the droplet profile retrieval. The retrieval bias for the effective radius at cloud top, shown in panel (a), is at most 0.4 μm, less than the average MODIS retrieval uncertainty for marine stratocumulus analyzed in this study of about 0.9 μm. Despite the small magnitude, Figure 3.a reveals a consistent pattern in cloud top effective radius: When the assumed integrated water vapor content exceeds/is less than the true value, the retrieval compensates for the excess/insufficient absorption due to water vapor by reducing/increasing retrieved droplet sizes. Panel (b) shows a much larger spread in retrieval bias for the effective radius at cloud bottom, with a maximum value approaching 1 μm. In addition, the relationship established in panel (a) is not as clear in panel (b). The retrieval bias for droplet size at cloud top appears to have a linear relationship with the true integrated water vapor content, whereas the bias for droplet size at cloud bottom does not have the same type of linear relationship. In addition, the bias appears to plateau around a true IWV_{ac} value of 8 mm. Figure 3 highlights the additional uncertainty added to the retrieval of a cloud droplet profile and the importance of simultaneously retrieving above-cloud precipitable water.

 doi:10.20944/preprints202603.2506.v1

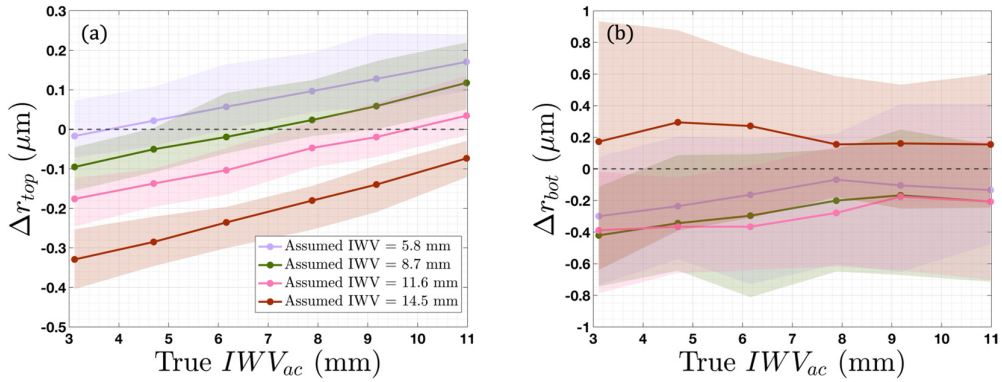


Figure 3. Retrieval bias, defined as the difference between the retrieved variable and the true value, for the effective radii at cloud top (a) and base (b). The true integrated water vapor content above cloud is the amount used to computed simulated HySICS reflectances. The four colored lines represent the four different assumptions of integrated water vapor content for the retrieval. The solid line represents the median and the shaded area represents the 25th and 75th percentile for about 800 retrievals (see Sect. 3.1).

3.1.2. Assessment of Retrieval Performance

The retrieval method was then applied to the set of simulated HySICS reflectances with cloud microphysics defined using VOCALS-REx in-situ measurements (see Sect. 2.2). Figure 4 shows three examples of retrieved profiles from the 69 simulated measurements, chosen to highlight distinct cases: (a) rapid change at cloud base, (b) minimal fluctuations throughout, and (c) rapid change at cloud top. Despite rapid fluctuations in droplet effective radius, the retrieved profiles closely follow the in situ measurements. The three cases shown in Figure 4 demonstrate close agreement between the in situ measured above-cloud integrated water vapor content and the retrieved value. The largest discrepancy shown is for the example with an optical thickness of 28.2, in panel (c). We know from Albert et al. [49] that our estimate will likely overestimate the true value for optically thick clouds due to the apparent elongation of the water vapor path due to multiple scattering.

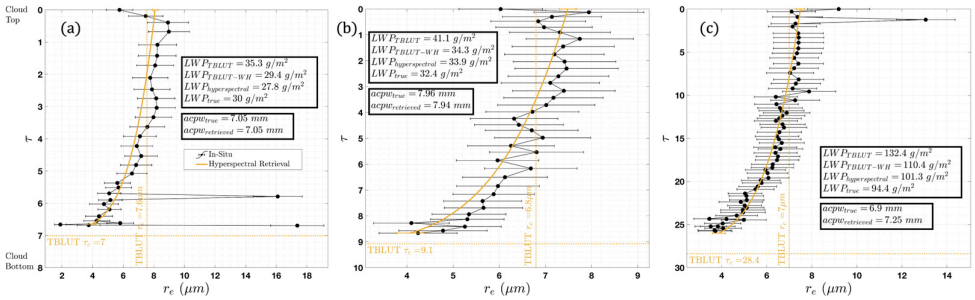


Figure 4. Comparison between in situ measured effective radius (black circles), TBLUT bispectral retrieval (dotted yellow lines), and retrieved vertical profile using all 636 HySICS spectral channels (solid yellow curve). Upper boxes show liquid water path from in situ measurements and three retrieval methods: TBLUT (Eq. 12), Wood-Hartmann adjustment (Eq. 13), and hyperspectral retrieval. Middle boxes show in situ and retrieved above-cloud integrated water vapor. Yellow horizontal bars indicate retrieval uncertainty for effective radius at cloud top and bottom. Black horizontal bars indicate the in situ measurement uncertainty.

The average percent difference between the in situ measured liquid water path and three different retrievals are listed in Table 2. Assuming vertically constant droplet number concentration and that the size parameter $x = 2\pi r/\lambda$ is much larger than unity, liquid water path can be estimated as the product of effective radius and cloud optical thickness:

 doi:10.20944/preprints202603.2506.v1

$$LWP = \frac{2}{3}\rho\,\tau_c\,r_e \quad (7)$$

where ρ is liquid water density. This expression forms the basis for liquid water path estimates in the MODIS Collection 6 cloud products [65], referred to hereafter as TBLUT LWP. Several studies have shown that retrieving liquid water path using Eq. (7) from visible/shortwave infrared retrievals of cloud optical properties results in an overestimation compared with retrievals of liquid water path using measurements of microwave emission[11,58,66].

An adjustment to the TBLUT LWP estimate by assuming vertical variation of droplet size results in values closer to microwave retrievals (Seethala and Horváth, 2010). Wood and Hartmann [64] derived a modified expression for adiabatic clouds where liquid water content varies linearly with height:

$$LWP = \frac{5}{9}\rho\,\tau_c\,r_e \quad (8)$$

This will be referred to as the TBLUT LWP_{WH} estimate for the remainder of the paper. To estimate LWP from the retrieved droplet profile, we computed liquid water content at each model cloud layer. Using the retrieved droplet profile and assumed effective variance at each layer, we computed the bulk extinction coefficient for liquid water using libRadtran and the MIEV0 Mie scattering code [41,67]. When assuming a size distribution, the libRadtran output represents the bulk extinction coefficient (km^{-1}) per liquid water content (g/m^3). Assuming linearly increasing liquid water content with height within cloud, we computed the liquid water content at each layer as:

$$LWC(z_n) = \frac{\tau_c}{\Delta z \sum_n \kappa'_n z_n} z_n \quad (9)$$

where z_n is the altitude of the n^{th} cloud layer, Δz is the thickness of the layer, and κ' is the bulk extinction coefficient per liquid water content. Liquid water path was then estimated by integrating the liquid water content over normalized height within the cloud.

Table 2. Average percent difference between three liquid water path retrievals and true in situ measured values ($N = 69$). The HySICS hyperspectral retrieval used the retrieved droplet profile to estimate LWP (Equation 9). The TBLUT retrieval uses bispectral effective radius and optical thickness to estimate LWP (Equation 7). The TBLUT retrieval with the Wood-Hartmann adjustment uses the same bispectral retrievals but applies a different coefficient to account for a linear liquid water content increase with height (Equation 8).

Liquid Water Path Retrieval	Avg. % difference (%) (N = 69)
HySICS Hyperspectral retrieval	17.7%
TBLUT retrieval	45.2%
TBLUT retrieval w/ Wood-Hartman adjustment	21.1%

Retrieved LWP was compared with true in situ measured LWP, along with TBLUT LWP estimates using both standard (Eq. 12) and Wood-Hartmann adjusted (Eq. 13) methods. The results in Table 2 demonstrate that the hyperspectral retrieval achieves significantly better agreement with in situ measurements (17.7% average difference) compared to the standard TBLUT method (45.2%). The Wood-Hartmann adjustment substantially improves the TBLUT estimate (21.1%), approaching the accuracy of the full hyperspectral retrieval. This suggests that accounting for vertical droplet size variation, whether through the simplified Wood-Hartmann correction or full profile retrieval, is essential for accurate LWP estimation.

Figure 5 shows comparisons between the hyperspectral retrievals of LWP, the TBLUT LWP retrievals, τ_c, and IWV_{ac} with true in-situ measured values used to create the measurements. Panel (a) shows that the hyperspectral retrieval of LWP exhibits a small, positive bias of 6.8%, with retrieval accuracy diminishing for in-situ LWP values exceeding 90 g/m^2. This degradation likely reflects increased forward model uncertainty for optically thicker clouds where photon penetration becomes

 doi:10.20944/preprints202603.2506.v1

more limited. Panel (b) shows the TLBUT LWP estimate exhibits a positive bias of 28%, about 4 times larger than the value for the hyperspectral retrieval. Panel (c) shows that the Wood-Hartmann adjustment leads to a more accurate LWP estimate, with a positive bias of 6.6%. These results are similar to those of Seethala and Horváth [11], who showed better agreement with microwave retrievals using the Wood-Hartmann adjustment. Panel (d) shows that the hyperspectral retrieval of cloud optical thickness is accurate across the range shown here, with a small bias of 0.2%. Panel (e) shows a fairly accurate retrieval of IWV_{ac}, with an underestimation of about 11%. Future work with a larger in situ dataset spanning a wider range of IWV_{ac} values will enable a more robust assessment of retrieval accuracy across diverse atmospheric conditions.

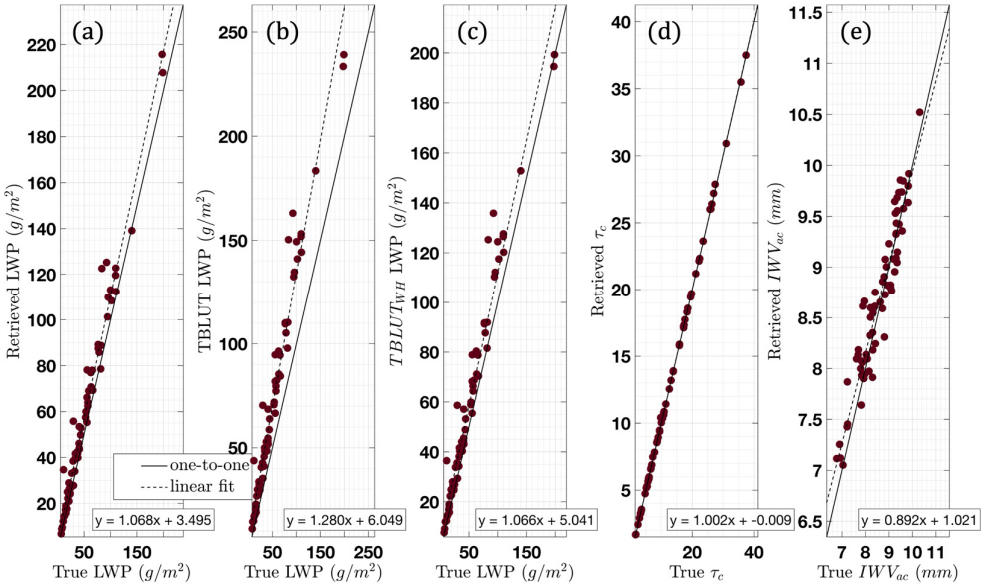


Figure 5. Performance of the (a) hyperspectral LWP retrieval, (b) the TBLUT LWP, (c) the TLBUT LWP_{WH} retrieval, and the hyperspectral retrieval of (d) cloud optical thickness and (e) above-cloud integrated water vapor. These retrievals used simulated HySICS reflectances. The true in situ measured value is shown on the x-axis. The solid line represents perfect agreement; a one-to-one line. The dashed line is the linear fit, mathematically described in lower righthand box.

3.2. Retrievals Using EMIT Measurements

Figure 6 shows an example scene from one of the 603 EMIT measurements used for this analysis. While we always found many EMIT pixels completely contained within a MODIS pixel, the MODIS pixels often straddled, or at times were outside of, the closest AMSR-E retrieval footprint. The AIRS Level 2 retrieval footprint is so much larger than the other instrument footprints that it often included non-cloudy regions or regions over land (see Table 1). This figure highlights the different ground sampled areas between EMIT, MODIS, AMSR-E and AIRS, demonstrating the difficulty with comparing retrievals from each.

 doi:10.20944/preprints202603.2506.v1

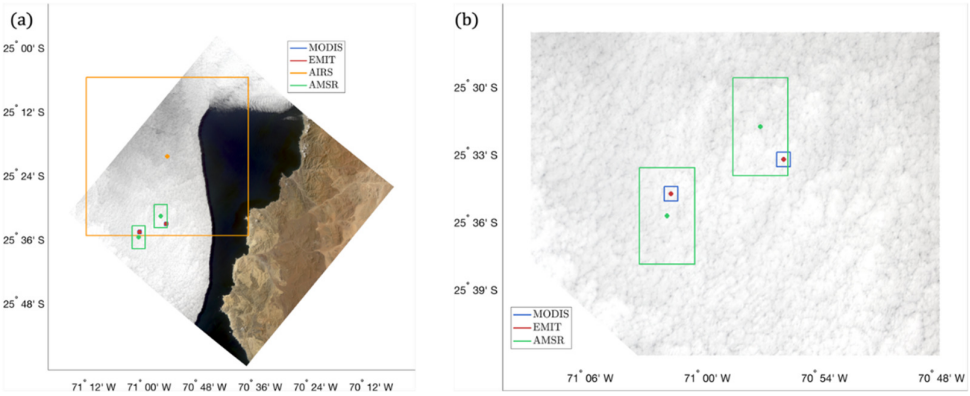


Figure 6. Panel (a) shows an EMIT swath recorded on 16 Sept. 2023 at 19:11:30 UTC, off the coast of Chile. The orange box shows an approximation of the closest AIRS Level 2 standard retrieval footprint in space and time to the EMIT measurements, shown in red. The green boxes show an approximation for the closest AMSR-E/AMSR2 Unified Level 2 Global Swath Ocean Product footprint to each EMIT measurement shown in red. Panel (b) is a cropped version of panel (a) to highlight the closest MODIS Level 2 retrieval, with the footprint approximated by the blue boxes.

Three examples of retrieved droplet profiles using EMIT measurements are shown in Figure 7. Without coincident in-situ measurements, we compared these retrievals with the liquid water path retrieved from MODIS, and AMSR-E measurements (when available), and the above cloud integrated water vapor retrieved from MODIS and AIRS measurements. Panel (a) is the only example shown with a usable overlapping AMSR-E retrieval of liquid water path. The average percent different between the MODIS retrieved effective radius and the hyperspectral retrieval of effective radius at cloud top was 9.2%, implying that the value derived from MODIS measurements typically exceeded the hyperspectral retrieval of effective radius at cloud top. We expected these values to be similar because of the MODIS retrieval bias towards droplet sizes at cloud top, as depicted in Figure 7. While the three examples shown in Figure 7 show close agreement between the hyperspectral retrieval of optical thickness and the MODIS retrieved value, we found the hyperspectral retrieval of optical thickness to vary spatially within all of the MODIS scenes used for this analysis, which is why we ran our retrieval on 30 different EMIT pixel measurements within each MODIS pixel used in this analysis. This was done in order to average out this spatial heterogeneity.

The average percent difference between the MODIS retrieval of liquid water path and the MODIS retrieval with the Wood-Hartmann adjustment with our hyperspectral retrieval using EMIT measurements is reported in Table 3. These results show that the MODIS retrieval, which assumes a vertically and horizontally homogeneous cloud, estimates a value of cloud liquid water that is 25.6% larger, on average, than our estimate using the retrieved vertical droplet profile. While we were not able to compare our retrieval of liquid water path with retrievals from AMSR-E often enough to report meaningful statistics, we note that previous work has reported a positive bias between MODIS and AMSR-E for overcast scenes, implying that the hyperspectral retrieval of liquid water path reported in Table 3 should be close to retrievals derived from microwave measurements [11]. The MODIS retrieval of liquid water path with the Wood-Hartmann adjustment was found to be 10.7% larger than our hyperspectral estimate, on average. This reinforces previous work that demonstrated this adjustment led to liquid water path retrievals in better agreement with retrievals derived from microwave measurements, which are thought to be a better representation of liquid water path [11].

 doi:10.20944/preprints202603.2506.v1

Table 3. The average percent difference between the MODIS-retrieved liquid water path and the hyperspectral retrieval using EMIT measurements is shown in the first row. The average percent difference between the MODIS-retrieved liquid water path using the Wood-Hartmann adjustment and the hyperspectral retrieval using EMIT measurements is shown in the second row. The EMIT hyperspectral retrieval used the retrieved droplet profile to estimate LWP (Equation 14).

Liquid Water Path Retrieval Comparison	Avg. % difference (%) (N = 603)
MODIS retrieval − EMIT Hyperspectral retrieval	25.6%
MODIS retrieval w/ Wood-Hartman adjustment − EMIT Hyperspectral retrieval	10.7%

Table 4 summarizes the comparison between the MODIS and AIRS retrievals for above cloud integrated water vapor content and the MODIS retrieval of cloud optical thickness with our hyperspectral retrieval. We found that our hyperspectral estimate of above cloud integrated water vapor content was consistently less than the retrievals of MODIS and AIRS by 26% and 28%, respectively. This bias in our retrieval may be due to the high uncertainty of the EMIT measurements. More than 90% of the retrievals converged after the first iteration, emphasizing the importance of an accurate initial guess. As demonstrated in Buggee and Pilewskie [20], radiometric and forward model uncertainty significantly increase the number of acceptable solutions. The MODIS retrieval of above cloud integrated water vapor content was found to be 9% smaller, on average, than the AIRS retrieval. Table 4 demonstrates a strong similarity between the MODIS retrieval of effective radius and our hyperspectral retrieval. The MODIS-derived values were found to be 3% smaller than our estimates, on average.

Table 4. The average percent difference between the MODIS-retrieved above cloud integrated water vapor content amount and the hyperspectral retrieval is shown in the first row. The average percent difference between the AIRS-retrieved above cloud integrated water vapor content amount and the hyperspectral retrieval using EMIT measurements is shown in the second row.

Above-cloud Integrated Water Vapor Retrieval Comparison	Avg. % difference (%) (N = 603)
MODIS retrieval − EMIT Hyperspectral retrieval	21.9%
AIRS retrieval − EMIT Hyperspectral retrieval	22%
MODIS retrieval − AIRS retrieval	-9.3%
Cloud Optical Thickness Retrieval Comparison	**Avg. % difference (%) (N = 603)**
MODIS retrieval − EMIT Hyperspectral retrieval	-4.3%

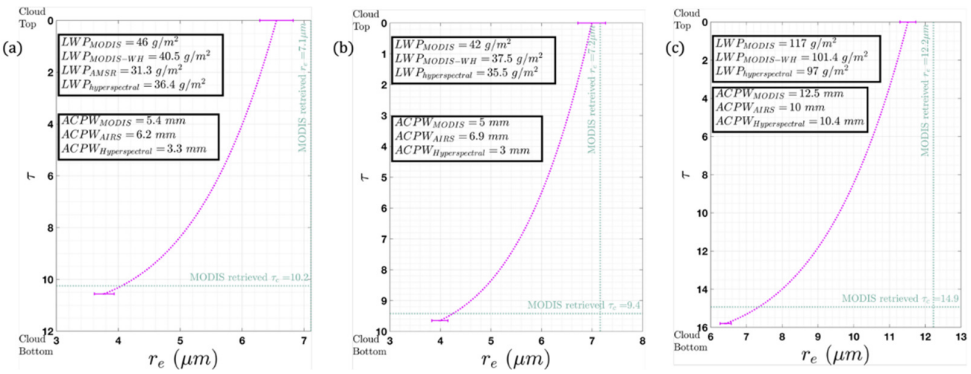


Figure 7. Comparison between the MODIS retrieval (dotted green lines), and the retrieved vertical profile using EMIT spectral channels (pink curve). Pink horizontal bars indicate retrieval uncertainty for effective radius at cloud top and bottom. Upper boxes compare the hyperspectral retrieval of LWP (Equation (14)) with the MODIS retrieval, the MODIS retrieval with the Wood-Hartmann adjustment, and the AMSR-E microwave retrieval (if

 doi:10.20944/preprints202603.2506.v1

available). Lower boxes compare the hyperspectral retrieval of above-cloud integrated water vapor with the MODIS, and AIRS estimates.

4. Summary and Discussion

We developed an optimal estimation retrieval method to simultaneously retrieve vertical profiles of cloud droplet effective radius and above-cloud integrated water vapor from hyperspectral measurements in the visible and shortwave infrared. This work builds upon the retrieval method of Buggee and Pilewskie [20] with several key enhancements: (1) simultaneous retrieval of above-cloud integrated water vapor in addition to droplet sizes at cloud top and base and cloud optical thickness, (2) transformation of the Gauss-Newton problem into logarithmic space to better represent the log-normal distributions of retrieved variables, (3) computation of the a priori covariance matrix including off-diagonal elements using in-situ measurements from the VOCALS-REx field campaign, and (4) incorporation of forward model uncertainty through a forward model Jacobian that accounts for deviations from assumed smooth adiabatic profiles, uncertainty in cloud top height, and uncertainty in the assumed effective variance of the droplet size distribution.

The physical basis for simultaneously retrieving above-cloud integrated water vapor stems from the work of Albert et al. [49], who showed that for optically thick marine clouds over ocean, the dominant contribution to water vapor absorption comes from the column above the cloud rather than below. We leveraged this by including spectral measurements in the wings of water vapor absorption features, which provide additional information content for constraining both the droplet profile [23,29]. Figure 1 demonstrated that weighting functions derived from HySICS spectral channels near water vapor absorption features exhibit greater relative orthogonality compared with channels outside these features, with wavelengths closer to strong absorption bands shifting the proportion of scattered photons toward cloud top. This enhanced orthogonality is fundamental to the ability to retrieve both the droplet profile and above-cloud integrated water vapor.

The impact of incorrectly above-cloud integrated water vapor assumptions on droplet profile retrievals using simulated HySICS reflectances with smooth, monotonic adiabatic profiles was quantified in Figure 3. Retrieval bias for radii at cloud top and base can reach values of approximately 0.4 $\boldsymbol{\mu m}$ and 1 $\boldsymbol{\mu m}$, respectively. When the assumed $\boldsymbol{IWV_{ac}}$ exceeded the true value, the retrieval compensated for the excess absorption by converging toward smaller droplet sizes. Conversely, when assumed $\boldsymbol{IWV_{ac}}$ was less than the true value, the algorithm converged toward larger droplet sizes. These results demonstrated that accurate droplet profile retrievals require simultaneous estimation of above-cloud integrated water vapor rather than relying on climatological assumptions.

We tested our retrieval method on 69 simulated HySICS reflectance spectra constructed using VOCALS-REx in situ measurements of cloud microphysics. This quasi-realistic dataset enabled systematic evaluation of retrieval performance against true values while maintaining precise control over forward model uncertainties. To quantify overall retrieval performance, liquid water path estimates from three retrievals were compared: the hyperspectral retrieval (Equation 9), the standard TBLUT method (Equation 7), and the Wood-Hartmann adjusted TBLUT method (Equation 8). Table 2 demonstrates that the hyperspectral retrieval achieved an average percent difference of 17.7% compared to in situ measured $\boldsymbol{LWP}$, substantially better than the standard TBLUT method (45.2%). The Wood-Hartmann adjustment, which accounts for a linear increase in liquid water content with height, improved the TBLUT estimate to 21.1%, approaching the accuracy of the full hyperspectral retrieval. This result is consistent with previous studies showing that accounting for vertical droplet size variation is essential for accurate $\boldsymbol{LWP}$ estimation [11,23]. Figure 5 revealed that while the hyperspectral $\boldsymbol{LWP}$ retrieval exhibited a small positive bias overall, retrieval accuracy diminished for large in situ $\boldsymbol{LWP}$ values exceeding approximately 100 $\boldsymbol{g/m^2}$. This degradation likely reflects increased forward model uncertainty for optically thicker clouds, where photon penetration into the lower cloud layers becomes more limited and the uncertainty for the retrieved droplet radius at cloud base increases. Cloud optical thickness retrievals (Figure 5.d) showed excellent agreement across all magnitudes, with retrieved values closely following the one-to-one line. Above-cloud integrated

 doi:10.20944/preprints202603.2506.v1

water vapor retrievals (Figure 5.e) demonstrated accurate performance, though future work with a larger in situ dataset spanning a wider range of $\boldsymbol{IWV_{ac}}$ values will enable more robust assessment across diverse atmospheric conditions.

The transformation of the retrieval problem into logarithmic space proved essential for improving convergence. All four retrieved variables—effective radii at cloud top and base, cloud optical thickness, and above-cloud integrated water vapor—more closely followed log-normal distributions than normal distributions, with $\boldsymbol{R^2}$ values higher in log space. This finding is consistent with previous studies showing that many atmospheric variables are inherently non-negative and better represented by log-normal distributions [32,33]. The a priori covariance matrix computed from VOCALS-REx in-situ measurements revealed physically meaningful correlations: effective radius at cloud top was found to have positive correlation coefficients with effective radius at cloud base (0.57) and cloud optical thickness (0.21). Interestingly, cloud optical thickness was found to have a higher correlation coefficient with the effective radius at cloud base (0.30) than at cloud top. Above cloud integrated water vapor content was found to have weak correlation coefficients with effective radii at cloud top (-0.09), cloud base (-0.1) and cloud optical thickness (-0.12). The specific mechanisms driving these correlation warrant further investigation with larger in situ data sets.

Application of our retrieval method to real hyperspectral measurements from the EMIT imaging spectrometer provided an initial demonstration of performance on real data, using coincident observations from MODIS and AIRS for comparison. EMIT, with a spectral range similar to HySICS but with coarser spectral resolution (285 vs. 636 channels) and substantially larger total uncertainty (4% assumed vs. 0.3% for HySICS), served as a surrogate for the upcoming CLARREO Pathfinder instrument. Table 3 showed that the MODIS retrieval, which assumes vertically and horizontally homogeneous clouds, estimated $\boldsymbol{LWP}$ to be 25.6% larger, on average, than our hyperspectral estimate using the retrieved droplet profile ($\boldsymbol{N = 603}$). The MODIS $\boldsymbol{LWP}$ with the Wood-Hartmann adjustment was 10.7% larger than our hyperspectral estimate, consistent with the direction and approximate magnitude of biases observed in previous studies [11].

The retrieval of above-cloud integrated water vapor from EMIT measurements was consistently lower than estimates from both MODIS and AIRS by 21.9% and 22%, respectively. This systematic underestimation could be due to the high radiometric uncertainty of the EMIT measurements: more than 90% of retrievals converged after the first iteration, suggesting that the measurement covariance matrix dominated the solution and drove the retrieval toward the a priori. We also note that the MODIS retrieval of above-cloud integrated water vapor was, on average, 9.3% smaller than the retrieval derived from AIRS measurements, indicating difference between the instruments that highlight the complication of direct validation in the absence of in-situ measurements. Future work should carefully characterize the sources of these biases, including the contribution of the a priori and the spatiotemporal representativeness of the comparison datasets.

Taken together, the comparison between simulated HySICS reflectances and real EMIT observations highlights how instrument characteristics directly affect retrieval performance. The larger radiometric uncertainty, coarser spectral resolution all contribute to the degraded retrieval performance of liquid water path and above cloud integrated water vapor content relative to simulated results. Nevertheless, the liquid water path and effective radius retrievals from EMIT remain physically consistent with coincident MODIS observations and reproduce the expected direction of the TBLUT bias, providing confidence that the retrieval methodology transfers successfully from simulated to real hyperspectral data. The HySICS instrument, with its order-of-magnitude improvement in radiometric accuracy over EMIT and nearly twice the number of spectral channels, is expected to improve overall performance relative to the EMIT results presented here.

Several limitations of this study warrant discussion. First, our forward model assumes 1-D plane-parallel radiative transfer, which does not accurately represent all cloud regimes. Horizontally inhomogeneous clouds can lead to 3-D radiative effects such as illumination and shadowing from net horizontal radiative transport. Previous studies have shown that horizontal inhomogeneity increases uncertainty in retrieved effective radius [63,68]. To limit these effects, EMIT observations with

 doi:10.20944/preprints202603.2506.v1

coincident MODIS-derived cloud inhomogeneity indices of less than 2.5 were used, though this threshold is substantially higher than the value of 0.3 recommended by Zhang and Platnick [63] for avoiding 3-D biases. A higher threshold was used to increase the number of observations that could be used in this analysis. Future work should explore more sophisticated approaches for identifying and correcting 3-D radiative effects in hyperspectral retrievals.

Second, we relied on a relatively small set of in situ observations ($\boldsymbol{N = 69}$) from a single field campaign conducted over a limited spatial region. The VOCALS-REx campaign surveyed marine stratus and stratocumulus clouds off the coast of Chile in the southeast Pacific over the course of two months, representing only one cloud regime under specific meteorological conditions. While this region is ideal for studying these cloud types due to their prevalence and relative homogeneity, the a priori covariance matrix derived from these measurements may not be representative of marine boundary layer clouds in other regions or under different synoptic conditions. For instance, clouds over the northeast Pacific, the Southern Ocean, or the Namibian coast may exhibit different relationships between droplet size, cloud optical thickness, and above-cloud integrated water vapor due to variations in aerosol sources, sea surface temperature, and large-scale subsidence patterns. The limited spatial extent also restricted our application to EMIT measurements, as we focused on the same geographic region to ensure the applicability of our a priori statistics. This constraint limited us to 603 coincident observations between EMIT and the instruments aboard Aqua that met our constraints for cloud optical thickness ($\boldsymbol{\tau_c \geq 3}$) and horizontal inhomogeneity ($\boldsymbol{H_\sigma \leq 2.5}$). Future work should incorporate in situ measurements from multiple field campaigns spanning diverse geographic regions and cloud regimes to develop more robust and generalizable a priori statistics, enabling a priori covariance matrices for different cloud types or meteorological conditions.

Our assumption of a smooth, monotonic adiabatic droplet profile does not capture the full complexity of real cloud microphysics. Figure 4 showed that in situ measurements exhibit rapid fluctuations in effective radius throughout the cloud, while our retrieved profiles necessarily represent smoothed approximations. However, Table 2 demonstrated that this simplification still enables substantially more accurate $\boldsymbol{LWP}$ estimation (17.7% error) compared to the vertically homogeneous assumption implicit in standard TBLUT retrievals (45.2% error). The success of the Wood-Hartmann adjustment (21.1% error) suggests that even simple parameterizations of vertical structure substantially improve retrieval accuracy. Future work should investigate the retrieval of multiple droplet sizes along the vertical dimension, rather than just values at cloud top and bottom, to improve the accuracy of the liquid water path retrieval. Additionally, the use of machine learning methods may be able to retrieve droplet profiles without strict assumptions that were made in the forward model developed for this analysis, such as an adiabatic droplet profile and the lack of drizzle-sized droplets.

The implications of this work extend beyond improved cloud property retrievals. Vertical profiles of droplet size provide insight into cloud droplet growth processes, the presence of drizzle, and the mixing of cloud top with dry air above [7–9]. More accurate liquid water path estimates are essential for constraining cloud albedo [10]. The ability to simultaneously retrieve above-cloud integrated water vapor may enable new investigations into the relationship between column integrated water vapor concentration and cloud droplet size. Future retrievals that estimate droplet size at more than just two heights within the cloud could be used to investigate the role of entrainment in modulating cloud albedo and lifetime. Looking toward future missions, the CLARREO Pathfinder HySICS instrument will provide unprecedented radiometric accuracy (0.3%) combined with contiguous spectral sampling from 350 to 2300 $\boldsymbol{nm}$. Our results using simulated HySICS reflectances demonstrate that this combination of high accuracy and spectral resolution should enable routine retrieval of vertical cloud droplet profiles.

Author Contributions: Conceptualization, A.J.B..; methodology, A.J.B..; software, A.J.B.; formal analysis, A.J.B.; investigation, A.J.B.; data curation, A.J.B.; writing—original draft preparation, A.J.B.; writing—review and

 doi:10.20944/preprints202603.2506.v1

editing, A.J.B. and P.P.; visualization, A.J.B.; supervision, P.P. All authors have read and agreed to the published version of the manuscript.

Funding: This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. 2040434.

Data Availability Statement: The retrieval algorithm developed for this paper is freely available on GitHub (https://github.com/andrewjbuggee/multispectral-retrieval-using-MODIS). The MODIS L1B and geolocation files (https://doi.org/10.5067/MODIS/MYD021KM.061 & https://doi.org/10.5067/MODIS/MOD021KM.061), and the L2 files (https://doi.org/10.5067/MODIS/MOD06_L2.061 & https://doi.org/10.5067/MODIS/MYD06_L2.061) used for retrieving droplet profiles are described within the previously mentioned GitHub repository and freely available at NASA's Level-1 and Atmosphere Archive & Distribution System Distributed Active Archive Center (LAADS-DAAC), hosted at NASA's Goddard Space Flight Center (GSFC): https://ladsweb.modaps.eosdis.nasa.gov/. The VOCALS-REx data used for comparison with the multispectral retrievals are similarly defined within the GitHub repository. These data are maintained by the National Center for Atmospheric Research Earth Observing Laboratory Field Data Archive (NCAR EOL) and are freely available at: https://doi.org/10.5065/D60863M8. The AMSR-E/AMSR2 Unified L2B Global Swath Ocean Products (https://doi.org/10.5067/9YQRFKKEPUP4) used for verifying the retrieved products are described in the GitHub repository and freely available from the NASA National Snow and Ice Data Center Distributed Active Archive Center.

Acknowledgments: The authors would like to thank Dr. Yolanda Shea for providing the LASICS code. We also want to thank Dr. Greg Kopp for discussions on the HySICS instrument and for providing the slit functions. This work utilized the Alpine high performance computing resource at the University of Colorado Boulder. Alpine is jointly funded by the University of Colorado Boulder, the University of Colorado Anschutz, and Colorado State University and with support from NSF grants OAC-2201538 and OAC-2322260.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

MODIS	Moderate Resolution Imaging Spectroradiometer
EMIT	Earth Surface Mineral Dust Source Investigation
AMSR-E	Advanced Microwave Scanning Radiometer for EOS
AIRS	Atmospheric Infrared Sounder
VOCALS-REx	Variability of the American Monsoon Systems Ocean-Cloud-Atmosphere-Land Study Regional Experiment

References

1. Wood, R. Stratus and Stratocumulus. *Encyclopedia of Atmospheric Sciences* 2015, *2*, 196–200.
2. L'Ecuyer, T.S.; Hang, Y.; Matus, A.V.; Wang, Z. Reassessing the Effect of Cloud Type on Earth's Energy Balance in the Age of Active Spaceborne Observations. Part I: Top of Atmosphere and Surface. *Journal of Climate* **2019**, *32*, 6197–6217, doi:10.1175/JCLI-D-18-0753.1.
3. Wood, R. Stratocumulus Clouds. *Monthly Weather Review* **2012**, *140*, 2373–2423, doi:10.1175/MWR-D-11-00121.1.
4. Miles, N.L.; Verlinde, J.; Clothiaux, E.E. Cloud Droplet Size Distributions in Low-Level Stratiform Clouds. *Journal of the Atmospheric Sciences* **2000**, *57*, 295–311, doi:10.1175/1520-0469(2000)057%3C0295:CDSDIL%3E2.0.CO;2.
5. Meyer, K.; Platnick, S.; Arnold, G.T.; Amarasinghe, N.; Miller, D.; Small-Griswold, J.; Witte, M.; Cairns, B.; Gupta, S.; McFarquhar, G.; et al. Evaluating Spectral Cloud Effective Radius Retrievals from the Enhanced

 doi:10.20944/preprints202603.2506.v1

MODIS Airborne Simulator (eMAS) during ORACLES. *Atmospheric Measurement Techniques* **2025**, *18*, 981–1011, doi:10.5194/amt-18-981-2025.

6. Painemal, D.; Zuidema, P. 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* **2011**, *116*, 1–16, doi:10.1029/2011JD016155.
7. Hill, A.A.; Feingold, G.; Jiang, H. The Influence of Entrainment and Mixing Assumption on Aerosol–Cloud Interactions in Marine Stratocumulus. *Journal of the Atmospheric Sciences* **2009**, *66*, 1450–1464, doi:10.1175/2008JAS2909.1.
8. Magaritz-Ronen, L.; Pinsky, M.; Khain, A. Drizzle Formation in Stratocumulus Clouds: Effects of Turbulent Mixing. *Atmospheric Chemistry and Physics* **2016**, *16*, 1849–1862, doi:10.5194/acp-16-1849-2016.
9. Martins, J.V.; Marshak, A.; Remer, L.A.; Rosenfeld, D.; Kaufman, Y.J.; Fernandez-Borda, R.; Koren, I.; Correia, A.L.; Zubko, V.; Artaxo, P. Remote Sensing the Vertical Profile of Cloud Droplet Effective Radius, Thermodynamic Phase, and Temperature. *Atmospheric Chemistry and Physics* **2011**, *11*, 9485–9501, doi:10.5194/acp-11-9485-2011.
10. Borg, L.A.; Bennartz, R. Vertical Structure of Stratiform Marine Boundary Layer Clouds and Its Impact on Cloud Albedo. *Geophysical Research Letters* **2007**, *34*, doi:10.1029/2006GL028713.
11. Seethala, C.; Horváth, Á. Global Assessment of AMSR-E and MODIS Cloud Liquid Water Path Retrievals in Warm Oceanic Clouds. *Journal of Geophysical Research: Atmospheres* **2010**, *115*, doi:10.1029/2009JD012662.
12. Miller, D.J.; Zhang, Z.; Ackerman, A.S.; Platnick, S.; Baum, B.A. The Impact of Cloud Vertical Profile on Liquid Water Path Retrieval Based on the Bispectral Method: A Theoretical Study Based on Large-eddy Simulations of Shallow Marine Boundary Layer Clouds. *Journal of Geophysical Research: Atmospheres* **2016**, *121*, 4122–4141, doi:10.1002/2015JD024322.
13. Stephens, G.L.; Christensen, M.; Andrews, T.; Haywood, J.; Malavelle, F.F.; Suzuki, K.; Jing, X.; Lebsock, M.; Li, J.L.F.; Takahashi, H.; et al. Cloud Physics from Space. *Quarterly Journal of the Royal Meteorological Society* **2019**, *145*, 2854–2875, doi:10.1002/qj.3589.
14. Twomey, S.; Seton, K.J. Inferences of Gross Microphysical Properties of Clouds from Spectral Reflectance Measurements. *Journal of Atmospheric Sciences* **1980**, *37*, 1065–1069.
15. Nakajima, T.; King, M.D. Determination of the Optical Thickness and Effective Particle Radius of Clouds from Reflected Solar Radiation Measurements. Part I: Theory. *Journal of Atmospheric Sciences* **1990**, *47*, 1878–1893.
16. Twomey, S.; Bohren, C.F. Simple Approximations for Calculations of Absorption in Clouds. *Journal of Atmospheric Sciences* **1980**, *37*, 2086–2094.
17. Amarasinghe, N.; Platnick, S.; Meyer, K. Overview of the MODIS Collection 6 Cloud Optical Property (MOD06) Retrieval Look-up Tables; 2017;
18. King, M.D.; Platnick, S.; Menzel, W.P.; Ackerman, S.A.; Hubanks, P.A. Spatial and Temporal Distribution of Clouds Observed by MODIS Onboard the Terra and Aqua Satellites. *IEEE Transactions on Geoscience and Remote Sensing* **2013**, *51*, 3826–3852, doi:10.1109/TGRS.2012.2227333.
19. Platnick, S. Vertical Photon Transport in Cloud Remote Sensing Problems. *Journal of Geophysical Research Atmospheres* **2000**, *105*, 22919–22935, doi:10.1029/2000JD900333.
20. Buggee, A.J.; Pilewskie, P. Retrieving Vertical Profiles of Cloud Droplet Effective Radius Using Multispectral Measurements from MODIS: Examples and Limitations. *Atmospheric Measurement Techniques* **2025**, *18*, 5299–5320, doi:10.5194/amt-18-5299-2025.
21. Chang, F.L.; Li, Z. Estimating the Vertical Variation of Cloud Droplet Effective Radius Using Multispectral Near-Infrared Satellite Measurements. *Journal of Geophysical Research Atmospheres* **2002**, *107*, 1–12.
22. Kokhanovsky, A.; Rozanov, V.V. Droplet Vertical Sizing in Warm Clouds Using Passive Optical Measurements from a Satellite. *Atmospheric Measurement Techniques* **2012**, *5*, 517–528, doi:10.5194/amt-5-517-2012.
23. King, N.J.; Vaughan, G. Using Passive Remote Sensing to Retrieve the Vertical Variation of Cloud Droplet Size in Marine Stratocumulus: An Assessment of Information Content and the Potential for Improved Retrievals from Hyperspectral Measurements. *Journal of Geophysical Research Atmospheres* **2012**, *117*, doi:10.1029/2012JD017896.

 doi:10.20944/preprints202603.2506.v1

24. Shang, H.; Hioki, S.; Penide, G.; Cornet, C.; Letu, H.; Riedi, J. Establishment of an Analytical Model for Remote Sensing of Typical Stratocumulus Cloud Profiles under Various Precipitation and Entrainment Conditions. *Atmospheric Chemistry and Physics* **2023**, *23*, 2729–2746, doi:10.5194/acp-23-2729-2023.
25. Thompson, D.R.; Green, R.O.; Bradley, C.; Brodrick, P.G.; Mahowald, N.; Dor, E.B.; Bennett, M.; Bernas, M.; Carmon, N.; Chadwick, K.D.; et al. On-Orbit Calibration and Performance of the EMIT Imaging Spectrometer. *Remote Sensing of Environment* **2024**, *303*, 113986, doi:10.1016/j.rse.2023.113986.
26. Mero, B.; Lake, K.; Upcott-Gill, A.; Edwards, A.; Herrington, M.; Coste, P.; Binter, F.; Lopez, S.; Fournier, S.; Ballester, M.; et al. Status of the Design, Accommodation, and Performance of the TRUTHS Hyperspectral Imager at the End of Phase B1. In Proceedings of the International Conference on Space Optics — ICSO 2022; SPIE, July 12 2023; Vol. 12777, pp. 258–266.
27. Nieke, J.; Despoisse, L.; Gabriele, A.; Weber, H.; Strese, H.; Ghasemi, N.; Gascon, F.; Alonso, K.; Boccia, V.; Tsonevska, B.; et al. The Copernicus Hyperspectral Imaging Mission for the Environment (CHIME): An Overview of Its Mission, System and Planning Status. In Proceedings of the Sensors, Systems, and Next-Generation Satellites XXVII; SPIE, October 19 2023; Vol. 12729, pp. 21–40.
28. Shea, Y.; Fleming, G.; Kopp, G.; Lukashin, C.; Pilewskie, P.; Smith, P.; Thome, K.; Wielicki, B.; Liu, X.; Wu, W. Clarreo Pathfinder: Mission Overview and Current Status. *International Geoscience and Remote Sensing Symposium (IGARSS)* **2020**, *4*, 3286–3289, doi:10.1109/IGARSS39084.2020.9323176.
29. Coddington, O.; Pilewskie, P.; Vukicevic, T. The Shannon Information Content of Hyperspectral Shortwave Cloud Albedo Measurements: Quantification and Practical Applications. *Journal of Geophysical Research Atmospheres* **2012**, *117*, 1–12, doi:10.1029/2011JD016771.
30. Kopp, G.; Smith, P.; Belting, C.; Castleman, Z.; Drake, G.; Espejo, J.; Heuerman, K.; Lanzi, J.; Stuchlik, D. Radiometric Flight Results from the HyperSpectral Imager for Climate Science (HySICS). *Geoscientific Instrumentation, Methods and Data Systems* **2017**, *6*, 169–191, doi:10.5194/gi-6-169-2017.
31. Wood, R.; Mechoso, C.R.; Bretherton, C.S.; Weller, R.A.; Huebert, B.; Straneo, F.; Albrecht, B.A.; Coe, H.; Allen, G.; Vaughan, G.; et al. The VAMOS Ocean-Cloud-Atmosphere-Land Study Regional Experiment (VOCALS-REx): Goals, Platforms, and Field Operations. *Atmospheric Chemistry and Physics* **2011**, *11*, 627–654, doi:10.5194/acp-11-627-2011.
32. Dubovik, O.; Fuertes, D.; Litvinov, P.; Lopatin, A.; Lapyonok, T.; Doubovik, I.; Xu, F.; Ducos, F.; Chen, C.; Torres, B.; et al. A Comprehensive Description of Multi-Term LSM for Applying Multiple a Priori Constraints in Problems of Atmospheric Remote Sensing: GRASP Algorithm, Concept, and Applications. *Front. Remote Sens.* **2021**, *2*, doi:10.3389/frsen.2021.706851.
33. Maahn, M.; Turner, D.D.; Löhnert, U.; Posselt, D.J.; Ebell, K.; Mace, G.G.; Comstock, J.M. Optimal Estimation Retrievals and Their Uncertainties: What Every Atmospheric Scientist Should Know. **2020**, doi:10.1175/BAMS-D-19-0027.1.
34. Lance, S.; Brock, C.A.; Rogers, D.; Gordon, J.A. 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* **2010**, *3*, 1683–1706, doi:10.5194/amt-3-1683-2010.
35. Copernicus Climate Change Service ERA5 Hourly Data on Pressure Levels from 1940 to Present 2018.
36. Green, R.O.; Mahowald, N.; Ung, C.; Thompson, D.R.; Bator, L.; Bennet, M.; Bernas, M.; Blackway, N.; Bradley, C.; Cha, J.; et al. The Earth Surface Mineral Dust Source Investigation: An Earth Science Imaging Spectroscopy Mission. In Proceedings of the 2020 IEEE Aerospace Conference; March 2020; pp. 1–15.
37. Rodgers, C.D. *Inverse Methods for Atmospheric Souding: Theory and Practice*; World Scientific, 2000; Vol. 2; ISBN 981-02-2740-X.
38. Poulsen, C.A.; Siddans, R.; Thomas, G.E.; Sayer, A.M.; Grainger, R.G.; Campmany, E.; Dean, S.M.; Arnold, C.; Watts, P.D. Cloud Retrievals from Satellite Data Using Optimal Estimation: Evaluation and Application to ATSR. *Atmospheric Measurement Techniques* **2012**, *5*, 1889–1910, doi:10.5194/amt-5-1889-2012.
39. Doicu, A.; Schreier, F.; Hess, M. Iteratively Regularized Gauss–Newton Method for Bound-Constraint Problems in Atmospheric Remote Sensing. *Computer Physics Communications* **2003**, *153*, 59–65, doi:10.1016/S0010-4655(03)00138-3.
40. Dubovik, O.V.; Lapyonok, T.V.; Oshchepkov, S.L. Improved Technique for Data Inversion: Optical Sizing of Multicomponent Aerosols. *Appl. Opt.* **1995**, *34*, 8422, doi:10.1364/AO.34.008422.

 doi:10.20944/preprints202603.2506.v1

41. Emde, C.; Buras-Schnell, R.; Kylling, A.; Mayer, B.; Gasteiger, J.; Hamann, U.; Kylling, J.; Richter, B.; Pause, C.; Dowling, T.; et al. The libRadtran Software Package for Radiative Transfer Calculations (Version 2.0.1). *Geoscientific Model Development* **2016**, *9*, 1647–1672, doi:10.5194/gmd-9-1647-2016.
42. Stamnes, K.; Tsay, S.-C.; Wiscombe, W.; Laszlo, I. DISORT, a General-Purpose Fortran Program for Discrete-Ordinate-Method Radiative Transfer in Scattering and Emitting Layered Media: Documentation of Methodology; 2000;
43. Coddington, O.M.; Richard, E.C.; Harber, D.; Pilewskie, P.; Woods, T.N.; Snow, M.; Chance, K.; Liu, X.; Sun, K. Version 2 of the TSIS-1 Hybrid Solar Reference Spectrum and Extension to the Full Spectrum. *Earth and Space Science* **2023**, *10*, e2022EA002637, doi:10.1029/2022EA002637.
44. Cox, C.; Munk, W. Measurement of the Roughness of the Sea Surface from Photographs of the Sun's Glitter. *J. Opt. Soc. Am., JOSA* **1954**, *44*, 838–850, doi:10.1364/JOSA.44.000838.
45. Anderson, G.P.; Chetwynd, J.H.; Clough, S.A.; She1tle, E.P.; Kneizys, F.X. *AFGL Atmospheric Constituent Profiles (0-120km)*; 1986;
46. Minnis, P.; Sun-Mack, S.; Young, D.F.; Heck, P.W.; Garber, D.P.; Chen, Y.; Spangenberg, D.A.; Arduini, R.F.; Trepte, Q.Z.; Smith, W.L.; et al. CERES Edition-2 Cloud Property Retrievals Using TRMM VIRS and Terra and Aqua MODIS Data—Part I: Algorithms. *IEEE Transactions on Geoscience and Remote Sensing* **2011**, *49*, 4374–4400, doi:10.1109/TGRS.2011.2144601.
47. Platnick, S.; King, M.D.; Ackerman, S.A.; Menzel, W.P.; Baum, B.A.; Riédi, J.C.; Frey, R.A. The MODIS Cloud Products: Algorithms and Examples from Terra. *IEEE Transactions on Geoscience and Remote Sensing* **2003**, *41*, 459–472, doi:10.1109/TGRS.2002.808301.
48. Sayer, A.M.; Poulsen, C.A.; Arnold, C.; Campmany, E.; Dean, S.; Ewen, G.B.L.; Grainger, R.G.; Lawrence, B.N.; Siddans, R.; Thomas, G.E.; et al. Global Retrieval of ATSR Cloud Parameters and Evaluation (GRAPE): Dataset Assessment. *Atmospheric Chemistry and Physics* **2011**, *11*, 3913–3936, doi:10.5194/acp-11-3913-2011.
49. Albert, P.; Bennartz, R.; Fischer, J. Remote Sensing of Atmospheric Water Vapor from Backscattered Sunlight in Cloudy Atmospheres. *Journal of Atmospheric and Oceanic Technology* **2001**, *18*, 865–874.
50. Rodgers, C.D. Information Content and Optimization of High-Spectral-Resolution Measurements. *Optical Spectroscopic Techniques and Instrumentation for Atmospheric and Space Research II* **1996**, *2830*, 136, doi:10.1117/12.256110.
51. Dubovik, O.; King, M.D. A Flexible Inversion Algorithm for Retrieval of Aerosol Optical Properties from Sun and Sky Radiance Measurements. *Journal of Geophysical Research: Atmospheres* **2000**, *105*, 20673–20696, doi:10.1029/2000JD900282.
52. Wilks, D.S. *Statistical Methods in the Atmospheric Sciences*; International geophysics series; 3rd ed.; Elsevier/Academic Press: Amsterdam ; Boston, 2011; ISBN 978-0-12-385022-5.
53. Platnick, S.; Meyer, K.G.; King, M.D.; Wind, G.; Amarasinghe, N.; Marchant, B.; Arnold, G.T.; Zhang, Z.; Hubanks, P.A.; Holz, R.E.; et al. The MODIS Cloud Optical and Microphysical Products: Collection 6 Updates and Examples from Terra and Aqua. *IEEE Transactions on Geoscience and Remote Sensing* **2017**, *55*, 502–525, doi:10.1109/TGRS.2016.2610522.
54. Gopalan, A.; Shea, Y.; Siman-tov, E.; Roithmayr, C.; Doelling, D.R.; Scarino, B.R. Langley Automated Sensor Inter-Calibration System (LASICS): Open Access Tools for Satellite Imager Inter- Calibration. Presented at the AMS 101st Annual Meeting, 2021.
55. Gao, B.-C.; Kaufman, Y.J. The MODIS Near-IR Water Vapor Algorithm;
56. Duncan, D.I.; Kummerow, C.D. A 1DVAR Retrieval Applied to GMI: Algorithm Description, Validation, and Sensitivities. *Journal of Geophysical Research: Atmospheres* **2016**, *121*, 7415–7429, doi:10.1002/2016JD024808.
57. Kummerow, C.; Ferraro, R.; Duncan, D. *AMSR-E/AMSR2 Unified L2B Global Swath Ocean Products, Version 1: User Guide*; NASA National Snow and Ice Data Center: Boulder, CO, USA, 2021;
58. Greenwald, T.J. A 2 Year Comparison of AMSR-E and MODIS Cloud Liquid Water Path Observations. *Geophysical Research Letters* **2009**, *36*, doi:10.1029/2009GL040394.
59. Liu, C.-C.; Zhu, Y.; Wang, L.; Bardeen, C.; Maloney, C.; Tilmes, S.; Vitt, F.; Kay, J.E.; Herrington, A.R.; Larson, V.E.; et al. Performance and Evaluation of a Sectional Cloud Microphysical Model (CARMA Cloud) in the Community Earth System Model (CESM2) 2026.

 doi:10.20944/preprints202603.2506.v1

60. Duncan, D.I.; Kummerow, C.D.; Meier, W.N. An Integrated Examination of AMSR2 Products Over Ocean. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2017**, *10*, 3963–3974, doi:10.1109/JSTARS.2017.2718535.
61. Strow, L.L.; Hannon, S.E.; Weiler, M.; Overoye, K.; Gaiser, S.L.; Aumann, H.H. Prelaunch Spectral Calibration of the Atmospheric Infrared Sounder (AIRS). *IEEE Transactions on Geoscience and Remote Sensing* **2003**, *41*, 274–286, doi:10.1109/TGRS.2002.808245.
62. Maning, E.; Kahn, B.; Fetzer, E.J.; Yue, Q.; Wong, S.; Kalmus, P.; Payne, V.; Wong, T.; Olsen, E.T.; Wilson, R.C.; et al. *AIRS/AMSU/HSB Version 7 Level 2 Product User Guide*; Jet Propulsion Laboratory: Pasadena, CA, USA, 2021; p. 197;.
63. Zhang, Z.; Platnick, S. An Assessment of Differences between Cloud Effective Particle Radius Retrievals for Marine Water Clouds from Three MODIS Spectral Bands. *Journal of Geophysical Research* **2011**, *116*, D20215, doi:10.1029/2011JD016216.
64. Wood, R.; Hartmann, D.L. Spatial Variability of Liquid Water Path in Marine Low Cloud: The Importance of Mesoscale Cellular Convection. *Journal of Climate* **2006**, *19*, 1748–1764, doi:10.1175/JCLI3702.1.
65. Platnick, S.; Meyer, K.G.; King, M.D.; Wind, G.; Amarasinghe, N.; Marchant, B.; Thomas Arnold, G.; Zhang, Z.; Hubanks, P.A.; Ridgway, B.; et al. *MODIS Cloud Optical Properties: User Guide for the Collection 6/6.1 Level-2 MOD06/MYD06 Product and Associated Level-3 Datasets*; 2018;
66. Chen, R.; Chang, F.-L.; Li, Z.; Ferraro, R.; Weng, F. Impact of the Vertical Variation of Cloud Droplet Size on the Estimation of Cloud Liquid Water Path and Rain Detection. *Journal of the Atmospheric Sciences* **2007**, *64*, 3843–3853, doi:10.1175/2007JAS2126.1.
67. Wiscombe, W.J. Improved Mie Scattering Algorithms. *Appl. Opt., AO* **1980**, *19*, 1505–1509, doi:10.1364/AO.19.001505.
68. Zhang, Z.; Ackerman, A.S.; Feingold, G.; Platnick, S.; Pincus, R.; Xue, H. 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* **2012**, *117*, 1–18, doi:10.1029/2012JD017655.