Summer 2023

Cloud Phase Distribution at a Global-Scale and the Governing Factors Using Satellite and In-Situ Airborne Observations

Dao Wang
San Jose State University

CLOUD PHASE DISTRIBUTIONS AT A GLOBAL-SCALE AND THE GOVERNING FACTORS USING SATELLITE AND IN-SITU AIRBORNE OBSERVATIONS

A Thesis

Presented to

The Faculty of the Department of Meteorology and Climate Science

San José State University

In Partial Fulfillment

of the Requirements for the Degree

Master of Science

by

Dao Wang

August 2023

The Designated Thesis Committee Approves the Thesis Titled

CLOUD PHASE DISTRIBUTIONS AT A GLOBAL-SCALE AND THE GOVERNING FACTORS USING SATELLITE AND IN-SITU AIRBORNE OBSERVATIONS

by

Dao Wang

APPROVED FOR THE DEPARTMENT OF METEOROLOGY AND CLIMATE SCIENCE

SAN JOSÉ STATE UNIVERSITY

August 2023

Minghui Diao, Ph.D.	Department of Meteorology and Climate Science
Qian Tan, Ph.D.	Department of Meteorology and Climate Science
Jie Gong, Ph.D.	NASA Goddard Space Flight Center

ABSTRACT

CLOUD PHASE DISTRIBUTIONS AT A GLOBAL-SCALE AND THE GOVERNING FACTORS USING SATELLITE AND IN-SITU AIRBORNE OBSERVATIONS

by Dao Wang

Cloud thermodynamic phase distributions play a crucial role in accurately representing cloud radiative effects and feedback in a changing climate. The partitioning of cloud thermodynamic phases (ice, liquid, and mixed phase) significantly influences Earth's surface temperature and its ability to mitigate the impact of global warming. Satellite-based cloud phase data are frequently used for the evaluation of global climate models, yet validation of them against in-situ observations is still lacking. This study examines global cloud phase distributions and their determinant factors by validating three satellite-based cloud phase products against an extensive in-situ airborne dataset. CALIPSO exhibits the most similar ice phase profiles, CloudSat tends to overestimate mixed phase frequency, and DARDAR overestimates ice phase frequency. The comparison results reveal variations in ice phase frequency across latitudes and seasons. Spatiotemporal mismatches have minimal impacts on the main findings, emphasizing statistical robustness. Machine learning techniques are employed to explore key determinant factors, such as temperature, relative humidity, vertical velocity, and aerosol effects. Temperature is the most influential factor in cloud phase distribution, while relative humidity determines in-cloud and clear-sky conditions. This study provides guidelines for globally validating satellite-based cloud products and enhances our understanding of factors influencing cloud phase distributions.

ACKNOWLEDGEMENTS

D. Wang would like to acknowledge the funding from the National Science Foundation Office of Polar Programs (OPP) grant #1744965 and SJSU Division of Research and Innovation award number 22-LUG-08-006. M. Diao would like to acknowledge funding from the NASA ACCDAM grant ROSES-2020 80NSSC21K1457. D. Wang also received the Walker Fellowship from San Jose State University. We would like to thank the science team that support each of the eleven NSF field campaigns, the scientists and staff from the UCAR Earth Observing Laboratory / Research Aviation Facility, and the science team of CALIPSO, CloudSat, and DARDAR.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

2DS – Two-Dimensional Stereo Probe
AIRS – Atmospheric Infrared Sounder
AMSU-A – the Advanced Microwave Sounding Unit-A
ANN – Artificial Neural Network
CALIOP – Cloud-Aerosol Lidar with Orthogonal Polarization
CALIPSO – Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation
CALIPSO-GOCCP – CALIPSO-GCM Oriented Cloud Calipso Product
CAM6 – Community Atmosphere Model version 6
CERES – Clouds and the Earth's Radiant Energy System
CDP – Cloud Droplet Probe
cm^{-3} – per cubic centimeter
CMIP3 – Coupled Model Intercomparison Projects 3
CMIP5 – Coupled Model Intercomparison Projects 5
CONTRAST - CONvective TRansport of Active Species in the Tropics
CPR – Cloud Profiling Radar
CRE – Cloud Radiative Effects
CSET – Cloud Systems Evolution in the Trades
CTH – Cloud Top Height
DARDAR – raDAR/liDAR project
DC3 – Deep Convective Clouds and Chemistry Project
dDist –differences in Distance
DJF – December, January, and February
dTime – difference in Time
EP – East Pacific (120˚W–180˚W)
Fast-2DC – the Fast-2 Dimensional Cloud probe
GBRTs – Gradient Boosting Regression Trees
GCM – Global Climate Model
GHz – Gigahertz
GV – Gulfstream-V
HIPPO – HIAPER Pole-to-pole Observations
Hz – Hertz
IceF – Ice phase occurrence Frequency
$IceF_{10}$ – Ice phase occurrence Frequency exceeding 10%
$IceF_{50}$ – Ice phase occurrence Frequency exceeding 50%
$IceF_{90}$ – Ice phase occurrence Frequency exceeding 90%
In-cloudF – In-cloud occurrence Frequency
INPs – Ice Nucleating Particles
IWC – Ice Water Contents
JJA – June, July, and August
km – kilometer
L^{-1} – per Liter
LWC – Liquid Water Contents

LWP – Liquid Water Path
ms^{-1} – meter per second
MAM – March, April, and May
M_{CDP} – Mass concentration of the CDP probe
ML – Machine Learning
MOSAIKS – Multi-task Observation using Satellite Imagery and Kitchen Sinks
Na – aerosol number concentration
Na_{100} – Number concentrations of Aerosols larger than 100 nm
Na_{500} – Number concentrations of Aerosols larger than 500 nm
NASA – National Aeronautics and Space Administration
NCAR – the National Center for Atmospheric Research
N_{CDP} – particle Number concentration of the CDP probe
Nd:YAG – Neodymium-doped Yttrium Aluminum Garnet
NH – Northern Hemisphere
nm – nanometer
NSF – the U.S. National Science Foundation
ORCAS – The O2/N2 Ratio and CO2 Airborne Southern Ocean Study
OTREC – Organization of Tropical East Pacific Convection
PDF – probability density function
PREDICT – PRE-Depression Investigation of Cloud Systems in the Tropics
r_e – effective radius
RF – research flight
RH_{ice} – relative humidity with respect to ice
RH_{liq} – relative humidity with respect to liquid
RICE – Rosemount Icing Detector
SEVIRI – The Spinning Enhanced Visible and InfraRed Imager
SH – Southern Hemisphere
SO – Southern Ocean
SOCRATES – Southern Ocean Clouds, Radiation, Aerosol Transport Experimental Study
SON – September, October, and November
START08 – Stratosphere-Troposphere Analyses of Regional Transport
TORERO – Tropical Ocean tRoposphere Exchange of Reactive halogen species and Oxygenated voc
UHSAS – Ultra-High Sensitivity Aerosol Spectrometer
VFM – Vertical Feature Mask
VCSEL – Vertical Cavity Surface Emitting Laser
WBF – Wegener-Bergeron-Findeisen process
WINTER – Wintertime Investigation of Transport, Emissions, and Reactivity
WMO – World Meteorological Organization
WP – West Pacific (120˚E–180˚E)
μm – micrometer

1. Validation of Satellite-based Cloud Phase Distribution Using Global-Scale In-Situ Airborne Observations

1.1 Introduction

Three cloud thermodynamic phases – liquid, mixed, and ice phase, have different radiative effects on the Earth's incoming and outgoing radiation. A previous study using satellite observations showed that the liquid, mixed and ice phases have negative, positive, and negative net cloud radiative effects (CRE) at the top of the atmosphere (Matus & L'Ecuyer, 2017). These differences in CRE can be explained by the vertical level, microphysical and macrophysical properties of different types of clouds. For example, liquid clouds are generally located in the lower troposphere, reflect more incoming shortwave radiation, and have relatively longer lifetime. On the other hand, ice clouds, such as cirrus clouds, tend to be higher in altitude and more transparent for shortwave radiation. They also trap outgoing longwave radiation and reheat the atmosphere. In fact, the higher and colder clouds have a greater warming effect on the Earth's system due to the larger temperature difference between the cloud layers and surface (Liou, 1992; Lohmann & Gasparini, 2017).

Simulations of Earth's climate are largely affected by the representation of cloud phases in global climate models (GCMs). Previously, the GCMs in the Coupled Model Intercomparison Projects 3 (CMIP3) and 5 (CMIP5) were found to have a large radiative bias over the Southern Ocean (e.g., Bodas-Salcedo et al., 2012, 2014, 2016; Frey & Kay, 2018; Kay et al., 2016; McCoy et al., 2014a, 2014b), mainly due to the misrepresentation of liquid phase below 0°C (i.e., supercooled liquid water) in the model parameterizations. These studies show radiative biases in the GCMs can be up to 10–20 W m-2 in the southern high latitudes due to the underestimation of reflected shortwave radiation by the supercooled liquid water. In addition,

other studies found that cloud feedback to a changing climate is also highly sensitive to the current state of the cloud phase (e.g., Flynn & Mauritsen, 2020; Tan et al., 2016; Terai et al., 2016; Zelinka et al., 2020). This is because the currently existing ice crystals could turn into liquid droplets as temperature increases and such a melting process would absorb energy from the surrounding air, acting as a buffer to global warming. This process is known as the cloud phase feedback in the climate system (Mitchell et al., 1989).

It is crucial to observational data to improve cloud parameterizations in GCMs, especially for accurately representing cloud phases. Satellite data have been frequently used to compare the global distributions of cloud phases and the associated radiative effects with various GCMs (e.g., Ahn et al., 2018; Barker et al., 2008; Ceccaldi et al., 2013; Cesana & Storelvmo, 2017; Hu et al., 2021). CALIPSO and CloudSat are two of the satellites from the original Afternoon Constellation (A-Train), equipped with LiDAR and RADAR active sensors, respectively. These two sensors have the advantage of providing near global, daily coverage cloud properties for over a decade. These sensors measure real-time observations that have the potential to vary significantly through time, whereas passive sensors with a perspective from space usually measure column-integrated quantities (Yost et al., 2010). Several previous studies used different types of products derived from CALIPSO and CloudSat to validate GCMs. Tan et al. (2016) used CALIPSO data to improve the representation of supercooled liquid water in the NCAR CAM6 model. Bodas-Salcedo et al. 2014 used the CALIPSO-GOCCP data to validate several climate models. Kay et al. 2016 use the COSP simulator, which is a satellite remote sensing simulator equipped with the NCAR CESM model.

Data qualities of these satellite products are therefore important for increasing the fidelity of climate predictions. They are not only dictated by the nature of the sensors but also by the algorithms used to process the retrievals. The impacts of using different algorithms were evaluated by Cesana et al. (2016). They compared three CALIPSO LiDAR-only cloud phase products to identify the differences in cloud phase identification caused by algorithms and instrumental designs. Three satellite-derived cloud phase products are frequently used for analyzing cloud phase distributions on a global scale, including LiDAR-only CALIPSO Vertical Feature Mask (VFM) (Y. Hu et al., 2009; Vaughan et al., 2019), RADAR-LiDAR combined CloudSat (Wang, 2019) and DARDAR (Delanoë and Hogan, 2008a) products. However, no previous study has validated all three products against in-situ observations. While the CALIPSO cloud phase product is solely based on CALIOP lidar measurements, the data generation processes of CloudSat and DARDAR cloud phase products integrate the combined lidar and radar measurements from CALIPSO and CloudSat. It is unclear if using the combined lidar and radar measurements would reduce the uncertainty in cloud phase identification compared with solely using the lidar measurements. In addition, since the CALIPSO lidar sensor has limited detection of optically thick clouds or multi-layer clouds, the CALIPSO observations potentially have a biased sampling of cloud top properties in these scenarios (Cesana et al., 2016). Thus, it is unclear if the frequency distributions of in-cloud conditions versus clear-sky conditions would differ between lidar and the combined lidar + radar datasets.

Several main challenges exist for validating satellite observations against other types of measurements. For example, ground-based observations have fixed locations at a limited number of ground stations, which cannot be used to validate the global-scale distributions of

cloud phases for satellite products. In-situ airborne observations use a moving platform and therefore have relatively larger geographical coverage than ground-based observations, but it requires multiple aircraft flight campaigns to achieve a global view. Previously, Diao et al. (2013) also quantified in impacts of arbitrary selection of comparison windows for selecting satellite and aircraft observations collocated in space and time based on validation of the NASA AIRS-AMSU-A satellite for its retrieval of temperature and water vapor.

Expanding on to the few previous studies that used a limited number of research flights for validating satellite observations of cloud phases at a specific geographical location, this study broadens the scope by providing a global-scale perspective of cloud phase distributions using a composite in-situ dataset and three frequently used satellite cloud phase products (i.e., CALIPSO, CloudSat, and DARDAR). A compositive in-situ observation dataset is developed using 11 NSF campaigns with a spatial coverage nearly from Pole to Pole (Longitude: 128°E ~ 180°E and 180°W ~ 37°W; Latitude: 84°N ~ 75°S). Additionally, several key factors affecting the cloud phase distributions and the satellite comparison results are investigated, including various latitudinal and longitudinal locations seasonal variability, and spatiotemporal variabilities (ranging from tens to hundreds of kilometers, as well as 3–12 hours). This work is designed as follows. Detailed descriptions of in-situ and satellite cloud phase datasets are given in Section 1.2. In Section 1.3, two different sampling methods to select satellite data for comparisons with in-situ aircraft observations are described. The results are shown in Section 1.4, where we quantify, visualize, and investigate the individual factors affecting the occurrences of clouds and cloud phases. Lastly, Section 1.5 summarizes the key conclusions and provides implications for future satellite retrieval development.

1.2. Data and Methods

Several observational platforms and datasets are described below, including in-situ observations from eleven flight campaigns funded by the US National Science Foundation (NSF) and three satellite-based cloud phase products. Detailed information of in-situ observed and remotely sensed datasets are described in Tables 1 and 2, respectively.

Table 1. Summary of eleven NSF campaigns, including their name, the number of research flights, time, location, and flight hours.

Campaign Name and abbreviation	# of Flights	Time of the Campaign	Longitude and latitude coverages	Flight Hours	All- Sky Hours	Clear-Sky Hours	In-Cloud hours
Stratosphere-Troposphere Analyses of Regional Transport (START08)	18	April – June 2008	26°N – 62°N, 118°W – 86°W	121	28	23	5
HIAPER Pole-to-pole Observations - deployments 2 – 5 (HIPPO 2-5)	46	Oct–Nov 2009; Mar–Apr 2010; Jun–July 2011; Aug–Sept 2011	84°N – 67°S, 128°E – 105°W	333	140	117	23
PRE-Depression Investigation of Cloud Systems in the Tropics (PREDICT)	26	Aug – Sept 2010	10°N – 28°N, 87°W – 37°W	175	11	10	1
Deep Convective Clouds and Chemistry Project (DC3)	22	May – June 2012	26°N – 42°N, 107°W – 79°W	136	37	31	6
Tropical Ocean tRoposphere Exchange of Reactive halogen species and Oxygenated voc (TORERO)	17	Jan – Feb 2012	42°S – 13°N, 106°W – 70°W	134	30	29	1
CONvective TRansport of Active Species in the Tropics (CONTRAST)	17	Jan – Feb 2014	21°S – 41°N, 133°E – 105°W	128	22	20	2
Wintertime Investigation of Transport, Emissions, and Reactivity (WINTER)	13	Feb – March 2015	29°N – 42°N, 86°W – 67°W	95	57	54	3

Cloud Systems Evolution in the Trades (CSET)	16	July – August 2015	20°N – 43°N, 118°W – 157°W	114	43	42	1
The O_2/N_2 Ratio and CO_2 Airborne Southern Ocean Study (ORCAS)	18	Jan – Mar 2016	75°S – 19°S, 92°W – 50°W	95	40	33	7
Southern Ocean Clouds, Radiation, Aerosol Transport Experimental Study (SOCRATES)	15	Jan – Feb 2018	62°N – 43°N, 133°W – 164°W	112	87	74	13
Organization of Tropical East Pacific Convection (OTREC)	22	Aug – Sept 2019	3°N – 13°N, 95°W – 77°W	123	9	7	2
All campaign combined	**230**			**1566**	**504**	**440**	**64**

Table 2. Descriptions of three satellite products and their cloud phase variables

Satellite and Instrument	Data Availability	Dataset	Variables	Phase Categorization	Resolution
CloudSat – CPR (Cloud Profiling Radar)	2006/06/12 d.163 ~ current (Missing: 2017/12/06 ~ 2018/08/10)	2B-CLDCLASS-LIDAR	CloudPhase	1 = Ice Cloud 2 = Mix Cloud 3 = Liquid Cloud -9 = Missing Value	Vertical: 240 m Horizontal: 1.4km x 1.8km
CALIPSO - CALIOP (Cloud-Aerosol Lidar with Orthogonal Polarization)	(V4.20) 2006/06/12 ~ 2020/06/30 (V4.21) 2020/07/01 ~ current	LID_L2_VFM-Standard-V4-20	Feature_Classification_Flags	0 = Unknown/Not Determined 1 = Ice 2 = Water 3 = Horizontal Oriented Ice	Vertical/Horizontal resolution: 30m/333m (−0.5 km to 8.2 km) 60m/1000m (8.2 to 20.2 km) 180m/1667m (20.3 to 30.1 km)

DARDAR (raDAR/liDAR)	(v2.1.1) 2006/06/15 ~ 2017/10/25	DARDAR-CLOUD_v2.1.1	DARMASK_Simplified_Categorization	-9 = ground -1 = do not know 0 = clear 1 = ice 2 = ice + supercooled 3 = liquid warm 4 = supercooled 5 = rain 6 = aerosol 7 = maybe insects 8 = stratospheric feature	Vertical: 60 m Horizontal: 1.1 km

1.2.1. In Situ Airborne Observations

A total of 11 NSF campaigns are used in this study, including the Stratosphere-Troposphere Analyses of Regional Transport (START08) (Pan et al., 2010), the HIAPER Pole-to-Pole Observations (HIPPO) (Wofsy, 2011), the PRE-Depression Investigation of Cloud Systems in the Tropics (PREDICT) campaign (Montgomery et al., 2012), the Deep Convective Clouds and Chemistry Project (DC3) (Barth et al., 2015), the Tropical Ocean tRoposphere Exchange of Reactive halogen species and Oxygenated voc (TORERO) campaign (Volkamer et al., 2015), CONvective TRansport of Active Species in the Tropics (CONTRAST) campaign (Pan et al., 2017), the Wintertime INvestigation of Transport, Emissions, and Reactivity (WINTER) campaign (Thronton et al., 2014), the Cloud Systems Evolution in the Trades (CSET) campaign (Albrecht et al., 2019), the O2/N2 Ratio and CO2 Airborne Southern Ocean Study (ORCAS) (Stephens et al., 2018), the Southern Ocean Clouds, Radiation, Aerosol Transport Experimental Study (SOCRATES) (McFarquhar et al., 2021), and the Organization of Tropical

East Pacific Convection (OTREC) campaign (Fuchs-Stone et al., 2020). The comprehensive dataset covers a longitude range of 120°E – 40°W and a latitude range of 75°S – 90°N, providing us with a unique analysis of global cloud phase distribution.

NSF Gulfstream-V research aircraft was the main platform for 10 out of 11 NSF campaigns, except for the WINTER campaign, which use the NSF C-130 research aircraft. Various sizes of cloud hydrometeors are measured by several in-situ cloud probes, including the Cloud Droplet Probe (CDP), the Fast-2 Dimensional Cloud (Fast-2DC) probe, and the Two Dimension Stereo (2DS) cloud probe. The measurement ranges of these cloud probes are as follows: 2 – 50 μm for CDP, 62.5 – 3200 μm for 2DC, and 40 – 5000 μm for 2DS. For almost all the 11 NSF campaigns, the CDP and 2DC probes are used to derive cloud microphysical properties. Only for the NSF SOCRATES campaign, CDP and 2DS probes were used due to quality issues with the 2DC probe in several research flights in that campaign.

For each in-situ cloud probe, the number concentrations of particles are reported at 1-Hz resolution. The 1-Hz measurements of each cloud probe are further categorized into large aerosols, liquid droplets (including supercooled liquid droplets), and ice crystals. This phase identification method was previously described in D'Alessandro et al. (2019) and Yang et al. (2021) with more details. Fundamentally, this method uses the relationship between particle number concentration (N_{CDP}) and mass concentration (M_{CDP}) of the CDP probe to define whether the 1-Hz observations are (i) large aerosols (satisfying either $N_{CDP} \leq 10{-}1.5$ cm−3 or $M_{CDP} \leq 10{-}3.4$ g m−3); (ii) liquid droplets (satisfying both $N_{CDP} \geq 10{-}0.5$ cm−3 and $M_{CDP} > 10{-}3.4$ g m−3); or (iii) ice crystals (satisfying both $10{-}1.5 < N_{CDP} < 10{-}0.5$ cm−3 and $M_{CDP} > 10{-}3.4$ g m−3). The phase identification for the 2DC and 2DS probes are similar to each

other, which uses parameters such as ambient temperature, particle number concentrations, maximum particle diameter, and the standard deviation of particle size distribution to define either liquid or ice dominant 1-Hz measurements. Furthermore, visual inspections as well as intercomparisons with other in-situ probes such as Rosemount ICEing (RICE) probe and King probe are also part of the quality control routine to verify the automatically designated cloud phase.

After categorizing the 1-Hz observations of individual cloud probes, liquid water content (LWC) and IWC are calculated for each second of measurement by each probe. For the 1-Hz observations defined as liquid droplets, spherical shape, and liquid water density are used for calculating LWC. To derive IWC for observations defined as ice crystals, small ice particles (i.e., diameter $\leq$ 75 um) derive IWC using the mass-Dimension (m-D) relationship for small ice based on Brown and Francis (1995), while the m-D relationship of large ice in that paper was used to derive IWC for large ice crystals (diameter >75 μm). Lastly, the total IWC and total LWC are derived by combining the IWC or LWC from individual cloud probes, respectively. Ice mass fraction is calculated using the total IWC divided by the sum of total IWC and total LWC. Such ice mass fraction at three ranges (<0.1, 0.1 – 0.9, and >0.9) each represent liquid phase, mixed phase, and ice phase based on integrated in-situ observations at 1-Hz resolution. These three cloud phases are further used for validating the three satellite products described below.

1.2.2. NASA CALIPSO (LID_L2_VFM_Standard-V4-20)

The NASA CALIPSO satellite carries the Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP), which is a spaceborne dual-wavelength polarization lidar using an

Nd:YAG laser that operates at 1064 nm and 532 nm (Y. Hu et al., 2009; Winker et al., 2007). CALIOP, which is designed with a broad dynamic range, has the unique advantage of measuring a variety of aerosols and cloud hydrometeors. High-resolution vertical profiles of clouds and aerosols, including their microphysical and optical properties, have been provided near globally since the CALISPO satellite was launched in April 2006.

The primary data product from CALIPSO used in this paper is the Vertical Feature Mask (VFM), which is a Level 2 product. The VFM data categorizes the spatial distributions of various features, such as cloud type, aerosol type (e.g., dust, polluted smoke, etc.), and cloud thermodynamic phase (Vaughan et al., 2004). Quantities derived from CALIOP retrievals, such as backscatter and linear depolarization ratio, can be used to identify particle sizes and orientations, which can be further used to separate ice and liquid phases (Sassen, 1991).

Cloud phase product is provided by “Feature_Classification_Flags” inside the files named “LID_L2_VFM-Standard-V4-20” (Table 2). The cloud phase product has different vertical and horizontal resolutions at three different altitudinal regions, as explained in Hunt et al. (2009) in their Table 4. The low-altitude region is located between -0.5 km and 8.2 km, with a vertical (horizontal) resolution of 30 m (333 m). The mid-altitude region is located from 8.2 km to 20.2 km, with a vertical (horizontal) resolution of 60 m (1 km). The high-altitude region is from 20.2 km to 30.1 km, with a vertical (horizontal) resolution of 180 m (1.667 km). The cloud phase product includes three categories – liquid, ice, and horizontal-oriented ice (HOI). In this study, HOI was grouped into the ice phase.

1.2.3. NASA CloudSat (2B-CLDCLASS-LIDAR)

The CloudSat cloud phase product is derived from the combination of CALIPSO CALIOP lidar and CloudSat Cloud Profiling Radar (CPR) observations. The “Cloud Phase” variable from the 2B-CLDCLASS-LIDAR product is used in this work (Table 2). This cloud phase dataset has a vertical resolution of 240 m and a horizontal resolution of 1.4 km by 1.8 km (Deng et al., 2010).

The CloudSat CPR radar uses 94 GHz sensors. Their signals travel through the atmospheric medium with bidirectional attenuation due to absorption by liquid water droplets, precipitation-sized particles, and gas (mainly water vapor). Since the lidar and radar have different sensitivities to different sizes and orientations of particles, the combined radar and lidar data could potentially complement each other when identifying cloud layers and cloud phases. Several fundamental differences in ice and liquid phase are used to develop the algorithm for identifying cloud phases (Wang, 2019; Wang & Sassen, 2001). Ice particles in clouds are significantly larger than water droplets, leading to distinct vertical distributions. Large ice particles have a terminal falling velocity of approximately 1 m/s, while small water droplets have falling speed that are negligible. This disparity in falling speeds affects the vertical distribution of clouds. In ice clouds, large particles tend to accumulate near the lower region, whereas in water clouds, larger water droplets are typically found closer to the top. Stratiform mixed phase clouds exhibit rapid ice particle growth within the mixed-phase layer, resulting in the formation of ice virga or precipitation beneath this region. Furthermore, there are orders of magnitude differences in the number concentrations of cloud droplets and ice crystals. Ice crystal number concentrations generally increase with decreasing temperature and can range

from less than 1/L to as high as 100/L in the temperature range where homogeneous ice nucleation occurs. In contrast, water droplet concentrations range from as low as 10 cm^{-3} in clear marine boundary layers to as high as 500 cm^{-3} in polluted continental boundary layers (Wang, 2019).

1.2.4. DARDAR (raDAR/liDAR) project (DARDAR_CLOUD.v2.1.1)

The DARDAR (raDAR/liDAR) project uses the integrated the CloudSat radar and the CALIPSO lidar observations and retrieves cloud properties. This project is supported by Laboratoire Atmosphères, Milieux, Observations Spatiales, and the University of Reading. The cloud phase product used in this study is part of the DARDAR-CLOUD dataset. It is similar to the CloudSat 2B-CLDCLASS-LIDAR product as they both use CALIOP and CPR measurements, but DARDAR uses a different algorithm to define cloud phases. The algorithm "Varcloud" is developed based on Delanoë and Hogan (2008b), providing several cloud properties, including ice water content (IWC), visible extinction coefficients, ice cloud effective radius (r_e), etc.

The cloud phase variable is stored as "DARMASK_Simplified_Categorization", which has a vertical resolution of 60 m and a horizontal resolution of 1.4 km (Delanoë and Hogan, 2008a). DARDAR cloud mask includes ice, ice + supercooled, liquid warm, supercooled, rain, and other aerosol indicators (Table 2). In this study, liquid warm, supercooled, and rain were classified as liquid phase. The category called "ice + supercooled" is categorized as the mixed phase.

1.3. Comparison Methods

This study developed two methods to compare satellite observations with in-situ airborne observations (Fig. 1). The main differences between the two methods are the spatial and temporal windows used to select satellite observations surrounding in-situ observations. Ideally, selecting a satellite observation sample that is perfectly collocated at the same time and location as an in-situ observation sample would minimize the impacts of spatial and temporal mismatches on the comparison results. Yet this type of perfect collocation is very rare in reality since both satellites and aircraft are moving objects in space and time with different trajectories and velocities. For example, CALIPSO and CloudSat are both sun-synchronous and orbit the Earth at an altitude of 705 km with a speed of approximately 7 km/s (Winker et al., 2003; Stephens et al., 2002). Conversely, the NSF Gulfstream-V research aircraft usually operates between the surface and 14 km in altitude with a true airspeed of around 200–250 m/s. Furthermore, aircraft campaigns often have limited geographical coverage and durations compared with the near global coverage of satellite observations. In addition, very few campaigns were specifically designed to follow the satellite trajectories in order to validate satellite observations (Barker et al., 2008). To overcome these challenges, this study developed two methods for sampling satellite data at various ranges of spatiotemporal windows, described as follows.

1.3.1. Method 1: selecting satellite samples in proximity to 1-Hz in-situ observations

Method 1 first selects satellite samples that are within a time window, which is from 12 hours before the first second of each research flight (RF) to 12 hours after the last second of that flight. These samples are further reduced to a spatial window bounded by the maximum

and minimum values of the latitude and longitude of that specific flight. For each pair of satellite and 1-Hz in-situ observation samples being compared, their differences in time (dTime) and in horizontal space (dDist) are calculated. Specifically, dDist represents the arc distance between two samples using their latitudinal and longitudinal positions. An additional restriction is applied to each satellite sample, which is dDist has to be within ± 200 km surrounding each second of in-situ observations. This method of selecting a range of dTime and dDist for comparing satellite and aircraft observations is similar to that used in Diao et al. (2013) as illustrated in their Figure 4.

After applying these restrictions, vertical columns of satellite samples are selected surrounding each second of in-situ observations. To further restrict the vertical levels for comparisons, each second of aircraft is matched with only one altitudinal bin inside a vertical column of a certain satellite product. This altitudinal bin is the layer which the 1-D flight track is transected through. The satellite samples selected via Method 1 for three datasets – CALIPSO, CloudSat, and DARDAR are illustrated in Fig. 1a-1c, respectively. The CALIPSO and CloudSat longitudinal and latitudinal positions were based on their respective orbit as part of the A-Train constellation, while the location of the DARDAR product is based on the collocated footprints of CALIPSO-CloudSat at the ground (Delanoë and Hogan, 2008b).

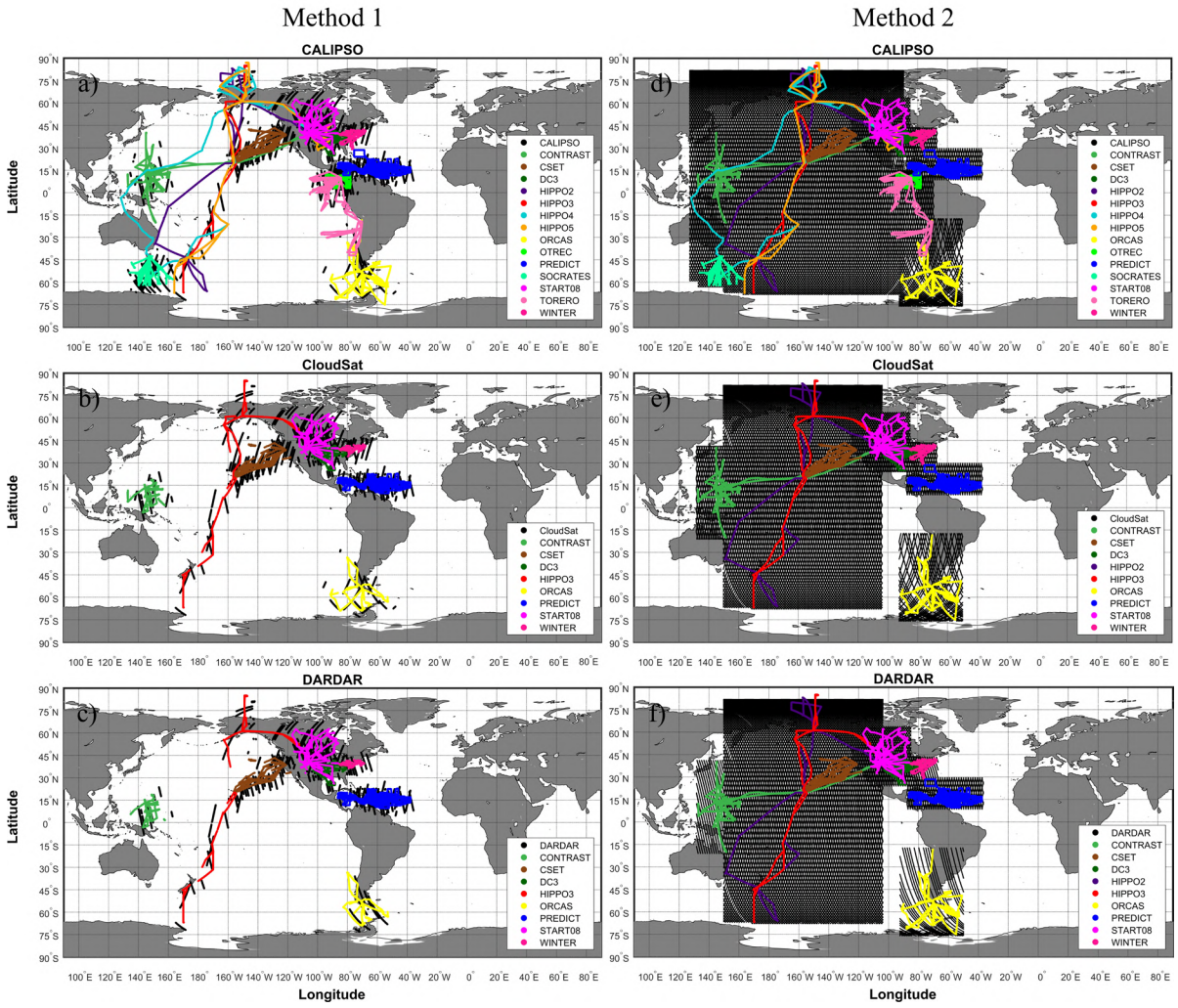

Figure 1. A global map for flight tracks of 11 NSF campaigns and the selected three satellite cloud phase products for comparisons. CALISPO, CloudSat and DARDAR are selected via (a-c) Method 1 and (d-f) Method 2, respectively. Aircraft flight tracks in (d-f) represent the entire campaigns, while (a-c) only show flight tracks collocated with satellite samples.

1.3.2. Method 2: selecting satellite samples for each flight campaign

Compared with Method 1, Method 2 uses a wider spatiotemporal window to select satellite data for individual campaigns, which allows more satellite samples to be evaluated in the comparisons. The temporal window of Method 2 starts 1 day before the first day of each campaign and ends 1 day after the last day of each campaign. A 3-D campaign domain is

defined by the minimum and maximum values of latitude and longitude for each campaign. Method 2 uses a spatial domain that extends 1 degree from each side of such campaign domain to select satellite observations surrounding each campaign. Another difference between the two methods is that Method 2 uses the entire vertical columns of satellite samples to conduct a statistical comparison with in-situ observations. Spatial distributions of satellite data sampled via Method 2 are illustrated in the global map in Fig. 1e-1f.

Overall, Method 1 serves the purpose of choosing satellite samples that are relatively closer to the in-situ observations to reduce the impacts of spatial and temporal variabilities in cloud phase distributions on the comparison results. Method 2, on the other hand, serves the purpose of providing a larger sample size of satellite observations to be evaluated against in-situ observations, which can be further used to examine variabilities in different seasons and latitudinal and longitudinal bands.

1.4 Results

1.4.1 A Case Study Comparison Using One Research Flight

A case study is conducted to illustrate the comparisons of cloud phase detected by in-situ aircraft observations and three satellite products (Fig. 2). This case is the Research Flight (RF) 08 from the NSF START08 flight campaign on May 6, 2008. The START08 campaign targeted the physical and chemical processes in the extratropical upper troposphere (UT) and lower stratosphere (LS). The outbound part of the flight was sampled across a weakening squall in eastern Oklahoma and northern Texas based on radar observations from the National Weather Service. The inbound part of the flight followed an air mass with low stability. The final descent back to Broomfield, Colorado sampled a weak convection with virga.

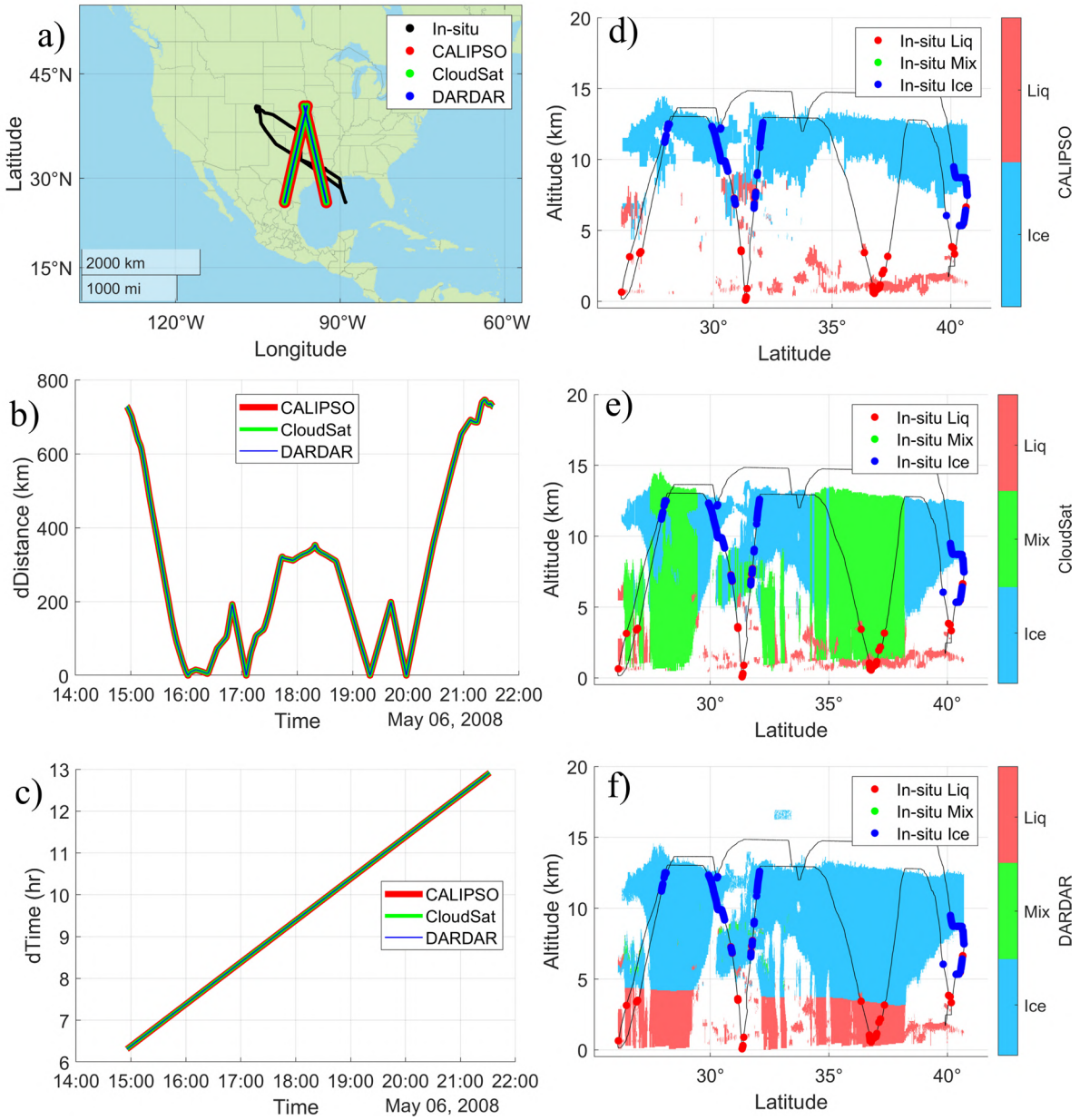

Figure 2. A case study of NSF START08 research flight 8 that compares in-situ observations with three satellite products. (a) a map of the locations of in-situ and satellite observations. (b) dDist, arc distance between research aircraft and satellites. (c) Time differences, dTime. (d-f) Curtain plots of satellite-based cloud phases (in shades) and in-situ observations of liquid, mixed and ice phase (in markers).

The satellite data shown in this case study are selected in a similar way as Method 1 but do not apply any restrictions on dDist or altitudinal levels for illustrative purposes. Figure 2a illustrates the relative positions between in-situ and satellite observations in a 2-D bird view, while Fig. 2b and 2c illustrate the constantly changing dDist and dTime between observation platforms. The three satellite products show almost identical locations and times relative to the aircraft observations, indicating minimal impacts from spatial and temporal variations among these three satellite products on their evaluation results.

This specific case study was selected based on the presence of a well-defined and stationary deep convection system that developed in the northwestern regions of Texas and Oklahoma. The deep convection structure exhibited minimal movement throughout the day, with a minimum dTime of approximately 6 hours. This extended period of minimal movement provided an ideal opportunity to thoroughly examine and analyze the consistency and agreement between in-situ and satellite observations. By focusing on a case with limited movement, valuable insights were gained into the intricacies of cloud dynamics and the performance and accuracy of satellite observations in capturing the complex behavior of the deep convection system. The vertical structure of cloud layers are investigated through curtain plots (Figure 2d-2f). The satellite products show an extensive, thick layer of ice clouds in the middle to upper troposphere, which may originate from the anvil outflow cirrus from the squall line. All three satellite products detected cloud layers most of the time when cloud hydrometeors were detected in aircraft observations. For the flight track segments closer to Broomfield, CO (i.e., latitudes beyond 40°N), the satellites were further away from the aircraft track as illustrated by larger dDist values in Fig. 2b. This may be the main reason that ice

crystals sampled around this area in the flight were not detected by the satellites. Comparing the in-cloud occurrences among three satellite products, CALIPSO shows the lowest cloud coverage, with mostly clear- sky conditions between 2–8 km. This could be due to the lidar signal attenuation through thicker ice cloud layers at higher altitudes. On the contrary, both CloudSat and DARDAR show extensive cloud coverage from the surface to 12 km in this case. The in-cloud occurrences in the CloudSat cloud phase product are also very similar to DARDAR, showing similar cloud fraction and location.

When further examining the separation between ice and liquid phase, the ice and liquid hydrometeors observed by the research aircraft cloud probes are mostly captured correctly by three satellite products, with ice hydrometeors being detected as either ice or mixed phase, and liquid hydrometeors being detected as liquid or mixed phase by satellites. To further contrast the three satellites on their phase partitioning, DARDAR shows a distinct separation between ice and liquid phases around 3–4 km, which indicates a temperature-dependent threshold used in their cloud phase partitioning algorithm. CALIPSO and CloudSat have almost identical occurrences of the liquid phase, except for a few occasions when CALIPSO allows the liquid phase to occur up to 8 km (e.g., near 32° latitude), while CloudSat categorizes these points as a mixed phase. A unique feature of CloudSat product is the thick layers of mixed phase stretching from 2 km to 12 km. In comparison, the original CALIPSO data do not include mixed phase. The DARDAR cloud phase includes mixed phase but has very few occurrences.

Both CloudSat and DARDAR detected deep convection and anvil structures in both outbound and inbound parts of the flight, but their identified cloud phases differ from each other. CloudSat identified two convective pillar clouds with anvil clouds on the side, likely

altostratus and/or cirrus clouds, as well as rainbands on the south of deep convection (26°N ~ 27°N) categorized as liquid and mixed-phase clouds in an alternating pattern. A similar feature of alternating phases surrounding deep convection was reported previously (Hu et al., 2021). The two convective pillars are mostly categorized as mixed phase in CloudSat but are categorized as partly ice and partly liquid by DARDAR. For the location around 35°N - 37°N, DARDAR shows better agreement with the in-situ observations as they both capture the liquid phase, while CloudSat shows mixed phase.

Overall, this case study illustrates that all three satellite products are able to capture the hydrometeors at similar locations as the in-situ observations, demonstrating the capability of using in-situ observations with relatively small variations in space and time for satellite validation. This case study also illustrates the key differences among the three satellite products in terms of in-cloud occurrences and cloud phase partition. The distinct feature of lower in-cloud occurrences seen in CALIPSO is consistent with the physical limitation in the CALIOP LiDAR system, which is primarily used for detecting optically thin clouds (Winker et al., 2007; Zhang et al., 2010). On the other hand, since CPR RADAR is able to penetrate optically thick clouds (Im et al., 2005), CloudSat and DARDAR cloud phase products show higher in-cloud occurrences but with different phase partitioning.

1.4.2. Vertical Distributions of Cloud Phase Frequencies using eleven NSF campaigns

The cloud phase occurrence frequencies at various altitudes are examined in Fig. 3. In-cloud frequency is calculated in each 500-m bin as the number of in-cloud samples (i.e., the sum of all cloud phase samples) divided by the total number of samples (i.e., clear-sky plus in-cloud conditions) in that bin. In addition, cloud phase frequency is calculated for each vertical

bin as the number of samples of a specific cloud phase is divided by the number of in-cloud samples. That is, cloud phase frequency represents the frequency of a cloud phase among all cloud phases, while the in-cloud frequency represents the frequency of in-cloud conditions among all-sky conditions.

Method 2 (Fig. 2e, 2f, and 2g), which allows more satellite samples in wider spatiotemporal windows for comparisons, are compared with respective satellite samples selected by Method 1 for their in-cloud frequencies. The fluctuations of in-cloud frequencies shown in Method 1 are smoothed out when using larger samples in Method 2. In addition, Method 2 shows similar in-cloud frequencies as Method 1 below 10 km for CALIPSO, as well as similar profiles between 4–10 km for CloudSat and DARDAR. However, Method 2 shows lower in-cloud frequencies at 10–15 km for all three satellite products. This may reflect the flight designs of certain campaigns, such as the DC3 campaign targeting anvil outflow cirrus, and the PREDICT campaign targeting the upper levels of tropical cyclones, which leads to a slightly higher in-cloud frequency in the UT/LS region around the flight track compared with other regions in the entire flight domain. This result indicates that in-cloud frequency comparisons are highly sensitive to the spatial and temporal proximity between different observation platforms. As a result, smaller spatiotemporal comparison windows are recommended for in-cloud frequency comparison.

Cloud phase frequency distributions of all observations show increasing frequencies of the ice phase and decreasing frequencies of the liquid phase with increasing altitudes, consistent with the fundamental thermodynamic process of liquid hydrometeors converting to ice

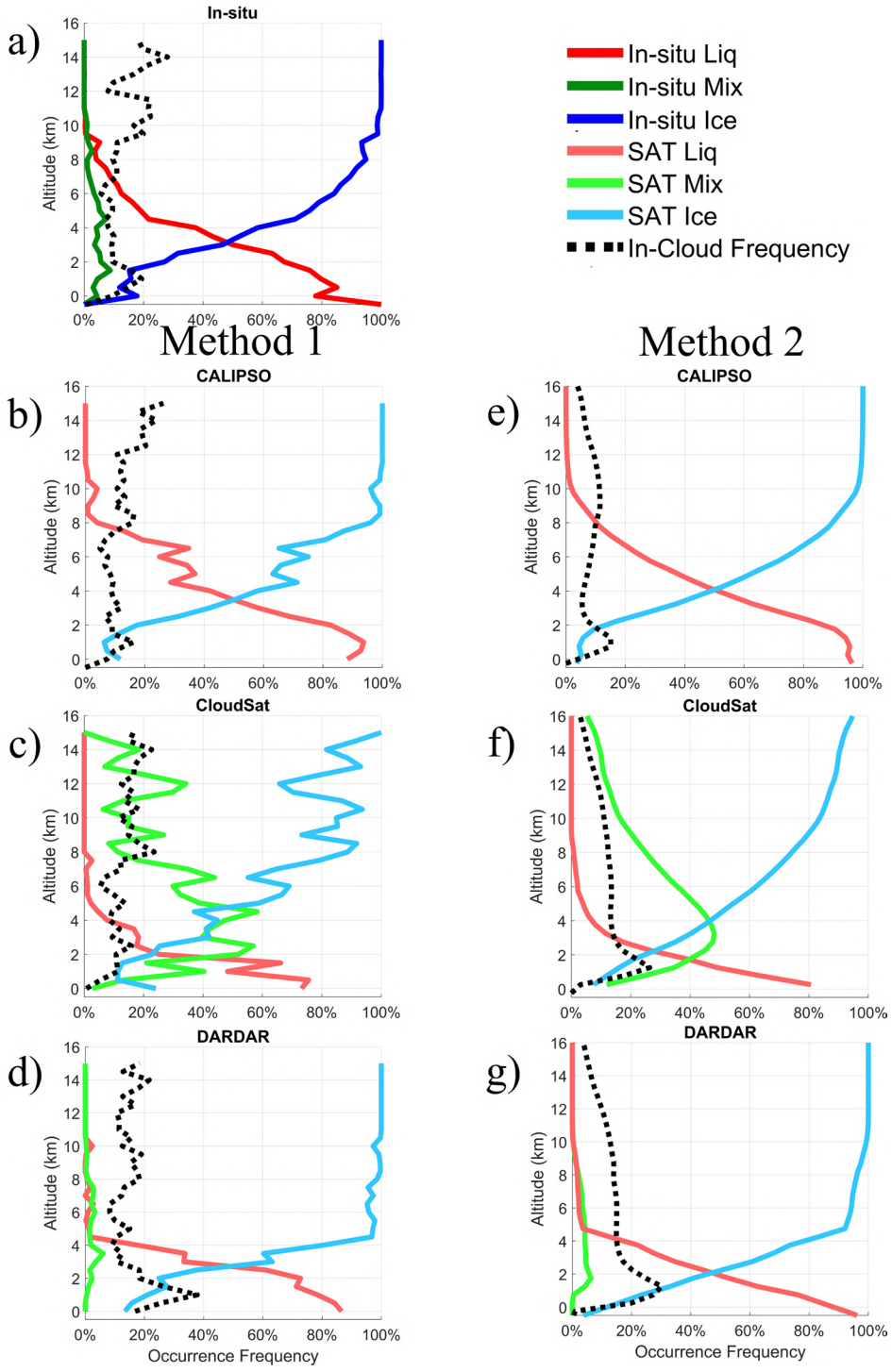

Figure 3. Vertical profiles of occurrence frequencies of three cloud phases and in-cloud conditions. (a) In-situ observations from 11 flight campaigns. (b-d) Satellite-based cloud phase vertical distributions using Method 1. (e-g) Comparisons using Method 2.

hydrometeors with decreasing temperatures. Unlike the comparisons of in-cloud frequencies that show different results using Methods 1 and 2, both comparison methods show similar features of cloud phase frequency distributions, including the trend relative to altitude and the frequency range of each cloud phase.

The altitudes where ice and liquid-containing phase (i.e., liquid plus mixed phase) have the same occurrence frequencies are at 3 km, 3.5 km, 4.5 km, and 3 km using method 1 sampling for in-situ observations, CALIPSO, CloudSat, and DARDAR, respectively. These altitudes also represent the locations where the dominant phase transitions from liquid dominant to ice dominant as altitudes increase. This vertical level can also be defined as where ice phase frequency equals 50%. The reason that CALIPSO and DARDAR show slightly lower altitudes of this transition level than in-situ observations is because these two satellite products have slightly higher frequencies of ice phase from the surface to 10 km than in-situ observations. The locations above which ice phase frequencies exceed 95% are 9.5 km, 8.5 km, 15 km, and 4.5 km for the in-situ and three satellite datasets, respectively. In addition, the locations below which liquid phase frequencies exceed 95% are 1.5 km, 1 km, and 0.5 km, and not applicable to these datasets, respectively. The identification of near-surface cloud phases by CloudSat is influenced by surface cluttering, specifically the echoes produced by the CPR within the range of the Earth's surface to 1 km. This interference has an impact on the observation of low-level clouds and precipitation. This comparison shows that CALIPSO has the most similar distributions of ice and liquid phase compared with observations, even though CALIPSO does not include mixed phase in its cloud phase product. In comparison, DARDAR allows too many

ice phase clouds to occur at lower altitudes and therefore never allows liquid phase to exceed 95% frequency.

The frequencies of mixed phase from in-situ observations are relatively low (i.e., less than 5%). DARDAR shows similar low frequencies of mixed phase less than 5%. CloudSat significantly overestimates mixed phase frequencies in all altitudes with frequency values ranging from 10% to 60%, consistent with the feature shown in the case study (Fig. 2e). this spurious mixed phase in CloudSat likely should have been categorized as all ice phase above 8 km and as all liquid phase below 2 km, since the in-situ observations show >90% of ice and liquid phase at these two altitude ranges, respectively. The peak position of the in-situ observed mixed phase is around 5 km, which is close to the transition level from ice to liquid- containing phase. Interestingly, for CloudSat and DARDAR, the peak positions of mixed phase frequencies are also concurrent with their respective transition levels. This feature indicates that the locations where both ice and liquid phases occur frequently individually as separate cloud segments (near 50% frequency) also have the highest probability of generating and maintaining mixed phase cloud segments.

1.4.3. Latitudinal Distributions of Ice Phase Frequency in a Global View

The latitudinal distribution of ice phase frequency helps identify discrepancies between satellite and in-situ cloud phase observations. It can provide insights into the performance of satellites in different latitude ranges. The zonal average of ice phase occurrence frequency is compared between in-situ and satellite observations using comparison Method 1 (Fig. 4) and Method 2 (Fig. 5). The ice phase frequency is first calculated as the number of ice phase samples divided by the total number of in-cloud samples in a given bin, which has a vertical

resolution of 500 m and a horizontal resolution of 2.5° latitude. In addition, a 3 × 3 moving average is applied to each latitudinal by altitudinal bin to reduce the impacts of missing data in certain bins. Ice frequency lines (IceF) are defined as the lowest altitudes where ice phase frequency reaches a certain percentage within each 2.5° latitudinal bin, including frequencies of 10% ($IceF_{10}$), 50% ($IceF_{50}$), and 90% ($IceF_{90}$). A moving average is also applied to three IceF lines after they are calculated from high-frequency data. IceF lines illustrate the transition between liquid- containing phase and the ice phase at various altitudes and latitudes. We define the altitudinal range between $IceF_{10}$ and $IceF_{90}$ as the mixing region.

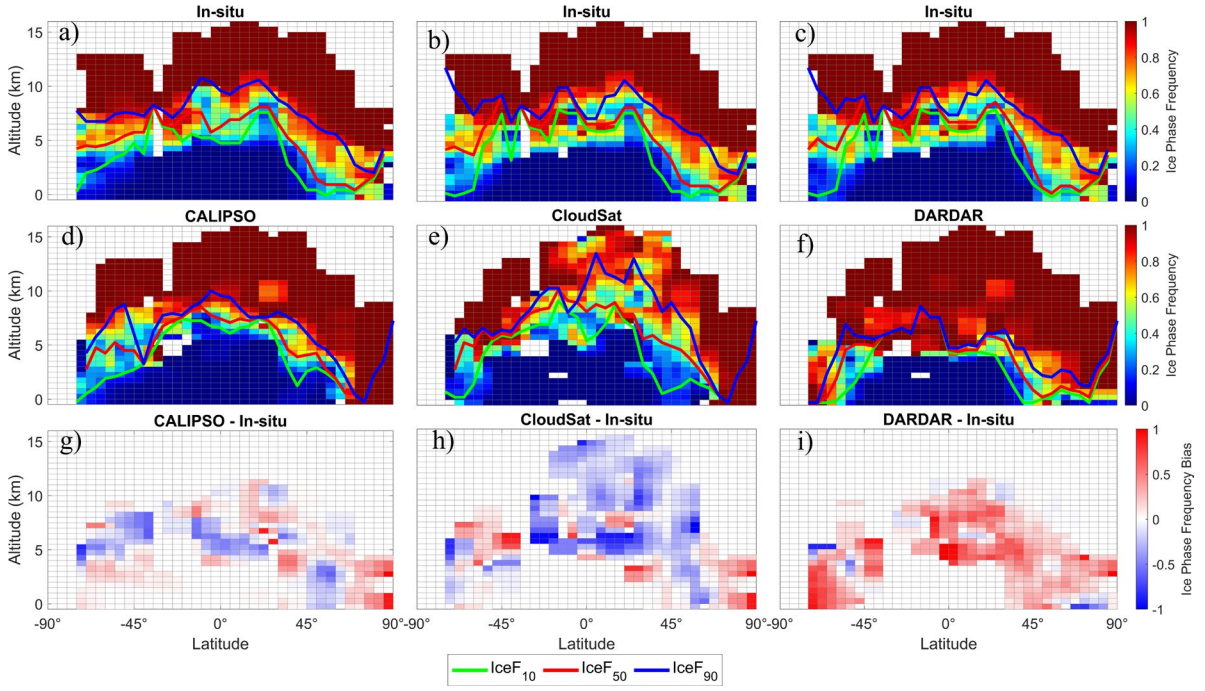

Figure 4. Latitudinal – altitudinal distributions of ice phase occurrence frequency using Method 1. Columns 1–3 represent comparisons for three respective satellite products – CALIPSO, CloudSat, and DARDAR, respectively. Rows 1, 2, and 3 show in-situ data, satellite data, and their differences (satellite minus in-situ).

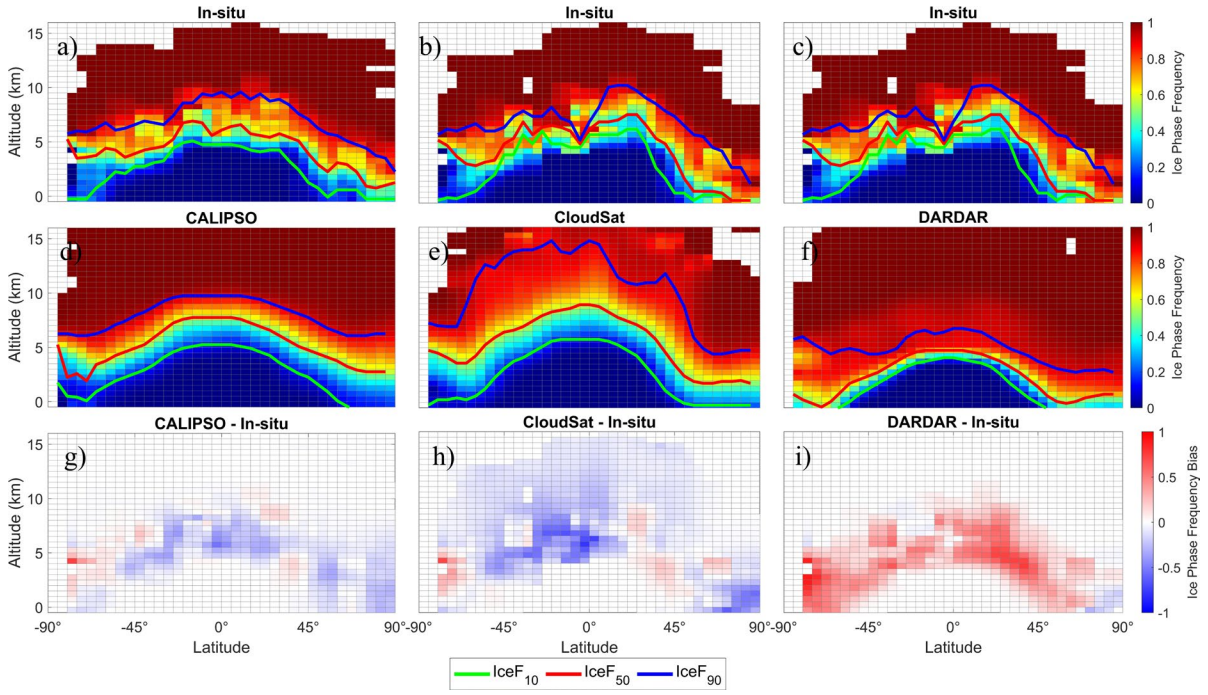

Figure 5. Similar to Figure 4, except for using Method 2. Unlike Figure 4, in-situ observations in Figure 5 represent the entire flight campaigns

Three different sub-sets of in-situ observations are shown in Fig. 4a, 4b, and 4c, which represent the in-situ observations that satisfy the spatiotemporal collocation criteria of Method 1 and can be further compared with CALIPSO, CloudSat, and DARDAR, respectively. Similarly, the top row of Fig. 5 represents the sub-sets of in-situ observations that satisfy the spatiotemporal selection criteria of Method 2 for the respective satellite dataset. Note that even though Method 2 allows satellite products within the entire domain of the flight campaigns to be used for comparison, time gaps still exist in different satellite data and therefore not all aircraft observations are matched with all three satellite products, leading to the variations in the sub-sets of in-situ observations in the top row.

Latitudinal dependence of ice phase frequency distribution is shown in the in-situ observations (Fig. 4a, 4b, and 4c). In the tropics (30°N–30°S), the mixing region is located

between 5 and 10 km, which is higher in altitude compared with the high-latitudinal mixing region. In the midlatitudes in the Northern Hemisphere (NH), the mixing region shows a sharp decrease in altitude by 5 km as latitude increases from 30°N to 60°N. Comparatively, the Southern Hemisphere (SH) midlatitudes show a smaller decline (by 2.5 km) in the mixing region height. The NH polar region (60°N–90°N) shows a relatively stable mixing region height at 0.5–5 km, while the SH polar region (60°S–90°S) shows a lower mixing region height around 2.5–7.5 km. Overall, significant hemispheric differences are shown by the in-situ observations, with the NH showing higher ice phase frequency than the SH at the same altitudes. One possible explanation for such hemispheric differences of ice phase frequencies is that the majority of flight campaigns in the SH took place in the Austral summer (e.g., SOCRATES and ORCAS were in Austral summer, with only a few flights from HIPPO in the Austral fall), while the flight campaigns in the NH occur in all seasons. The impacts of seasonal variability will be further discussed in Section 1.4.5.

These hemispheric differences and latitudinal dependence are consistently shown in all three sub-sets of the in-situ observations, regardless of the small variations in their samples. One main difference among the three sub-sets of in-situ observations occurs around the equator, i.e., Figures 4b and 4c show a dip in the mixing region at the equator compared with Figure 4a. Another main difference is at the 60°S–75°S, where a wider mixing region is shown in Figures 4b and 4c compared with Figure 4a.

Among all three satellite cloud phase products, CALIPSO (Fig. 4d) shows the most similar location and thickness of the mixing region at various latitudes compared with in-situ observations. At 45°S to 60°N, the thickness of the mixing region in CALIPSO is 2–3 km,

which is slightly thinner than that of in-situ observations (3 – 5 km). In the SH high latitudes (45°S to 90°S), the CALIPSO and in-situ observations have a similar thickness of mixing region, but CALIPSO shows a lower height of mixing region than in-situ observations, indicating that CALIPSO may have misrepresented supercooled liquid water as ice phase in this region. In the NH polar region (60°N to 90°N), both CALIPSO and CloudSat data have ice phase frequency at or closer to 1, while DARDAR allows a small amount of liquid phase below 1 km. This indicates that all three satellite products overestimate ice phase frequencies in the NH polar region by up to 50% from the surface to 5 km. Interestingly, the in-situ observations show a small increase in mixing region height at the NH and SH polar regions with increasing latitudes. Such an increase is captured by satellite data for the NH polar region but they overestimate this increasing trend. Such an increase could be caused by the reduction of available ice nucleating particles (INPs) in the polar regions, which may reduce the likelihood of ice formation compared with the midlatitudes.

When comparing CloudSat and DARDAR data against in-situ observations, CloudSat shows a thicker mixing region (i.e., the thickness of 5–7 km), while DARDAR shows a thinner mixing region (thickness of 2–3 km) for most latitudes. The height of the mixing region is also higher in CloudSat (i.e., 5–12.5 km in altitude) and lower in DARDAR (4–7.5 km in altitude) compared with in-situ observations. CloudSat (Fig. 4e) allows liquid-containing phase, mostly likely mixed phase, to occur up to 12.5 km, much higher than the in-situ observed value. DARDAR (Fig. 4h) shows a sharp transition between the liquid-containing and ice phase (similar to the case study results). It only allows the liquid phase to frequently occur in a narrow

latitudinal range (45°N to 45°S), while in-situ observations frequently show liquid-containing phase in the SH polar region as well.

The last row of Fig. 4 (Fig. 4e, 4h, and 4i) shows the satellite biases of ice phase frequency, calculated as frequencies of a certain satellite product minus the in-situ observations for the same latitudinal and altitudinal bin. CALIPSO shows a mixture of positive and negative biases in ice phase frequency, which underestimates ice phase frequency in the mid-troposphere around the mixing region and overestimates it above and below the mixing region. CloudSat shows the most negative biases except for the polar regions. DARDAR shows almost exclusively positive biases except near the surface at the NH polar region and near 5 km at the SH polar region.

Figure 5 shows similar main features compared with Fig. 4, including similar location and thickness of mixing region for in-situ observations, a mixture of positive and negative biases in CALIPSO, significant negative biases in CloudSat, and positive biases in DARDAR. Similar to the smoothing effect of Method 2 shown in Fig. 3, Fig. 5 also shows a reduction of fluctuations in three IceF lines for each dataset when a larger sample size is analyzed. In addition, since Method 2 uses entire columns of samples instead of a specific altitude for comparison with aircraft data, Fig. 5 also shows more samples in the high altitudes in satellite data. In-situ observations (Fig. 4a, 4b, 4c) show a similar thickness from $IceF_{10}$ to $IceF_{50}$, compared with $IceF_{50}$ to $IceF_{90}$. This feature is also well represented by CALIPSO. CloudSat shows a similar range from $IceF_{10}$ to $IceF_{50}$, but a much thicker range from $IceF_{50}$ to $IceF_{90}$. This feature is most likely caused by an overestimation of mixed phase up to the high altitudes. DARDAR shows a similar $IceF_{50}$ to $IceF_{90}$, but a much thicker range from $IceF_{10}$ to $IceF_{50}$,

which indicates that the transition from liquid-dominant to a mixture of liquid and ice occurs very fast in DARDAR, similar to the sharp transition seen in Fig. 3.

Overall, the similarity between Methods 1 and 2 for the main features seen in the in-situ observations and the main validation results for each satellite product demonstrates the statistical robustness of using these eleven NSF flights for the comparisons.

1.4.4. Regional Variations of Ice Phase and In-Cloud Frequencies

Average vertical profiles of ice phase frequencies (Fig. 6) and in-cloud frequencies (Fig. 7) are examined for three regions: Southern Ocean (35°S – 70°S), tropics (5°S–30°N), and NH extratropics (30°N–90°N). In addition, the effects of selecting collocated satellite samples in various temporal windows are examined for the ice phase and in-cloud frequencies. Specifically, three ranges of dTime are analyzed using Method 1, including dTime within 3 hrs, 6 hrs, and 12 hrs. Such analysis of the temporal window is similar to the method used by Diao et al. (2013). Similar to the frequency definition used in Fig. 3–5, ice phase frequency is the number of ice phase over the number of in-cloud samples, while in-cloud frequency is the number of in-cloud samples over the total number of all-sky conditions. Only the in-situ observations that have collocated satellite samples are shown in each panel, which leads to slight variation in the observed profiles depending on which satellite product is being evaluated.

Based on in-situ observations, the tropics have lower ice phase frequency at each altitudinal bin compared with the Southern Ocean and NH extratropics. This is mainly due to the fact that temperatures at the tropics are higher than other latitudes for the same altitudes, while ice phase frequency is temperature dependent (i.e., lower ice phase frequency with higher temperature).

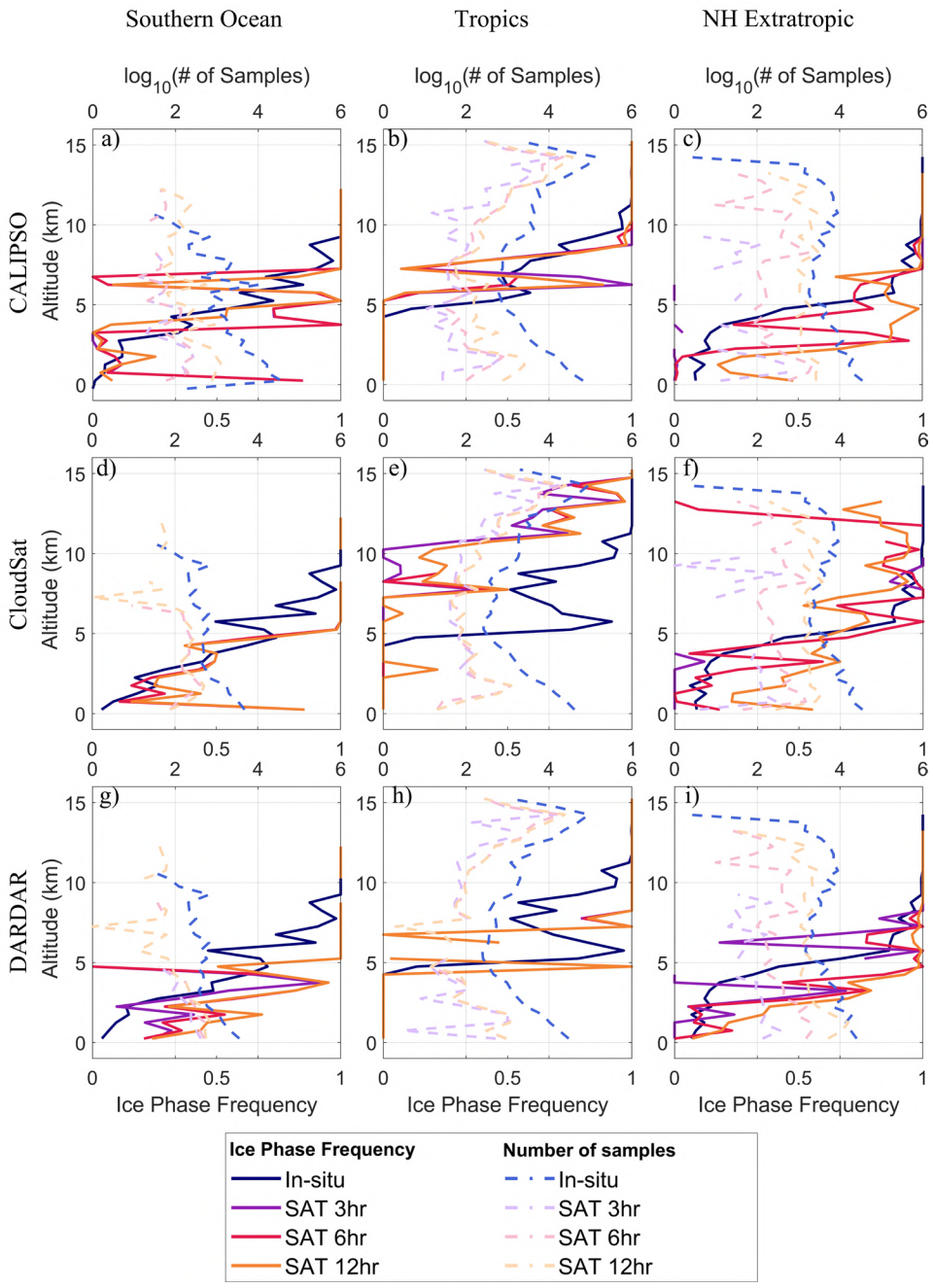

Figure 6. Regional variations of ice phase frequency vertical profiles. Columns 1 – 3 represent three latitudinal bands – the Southern Ocean, tropics, and NH extratropics. Solid lines denote ice phase frequency with different time window restrictions. Dotted lines denote the log10 number of samples of in-cloud conditions within different time windows restrictions.

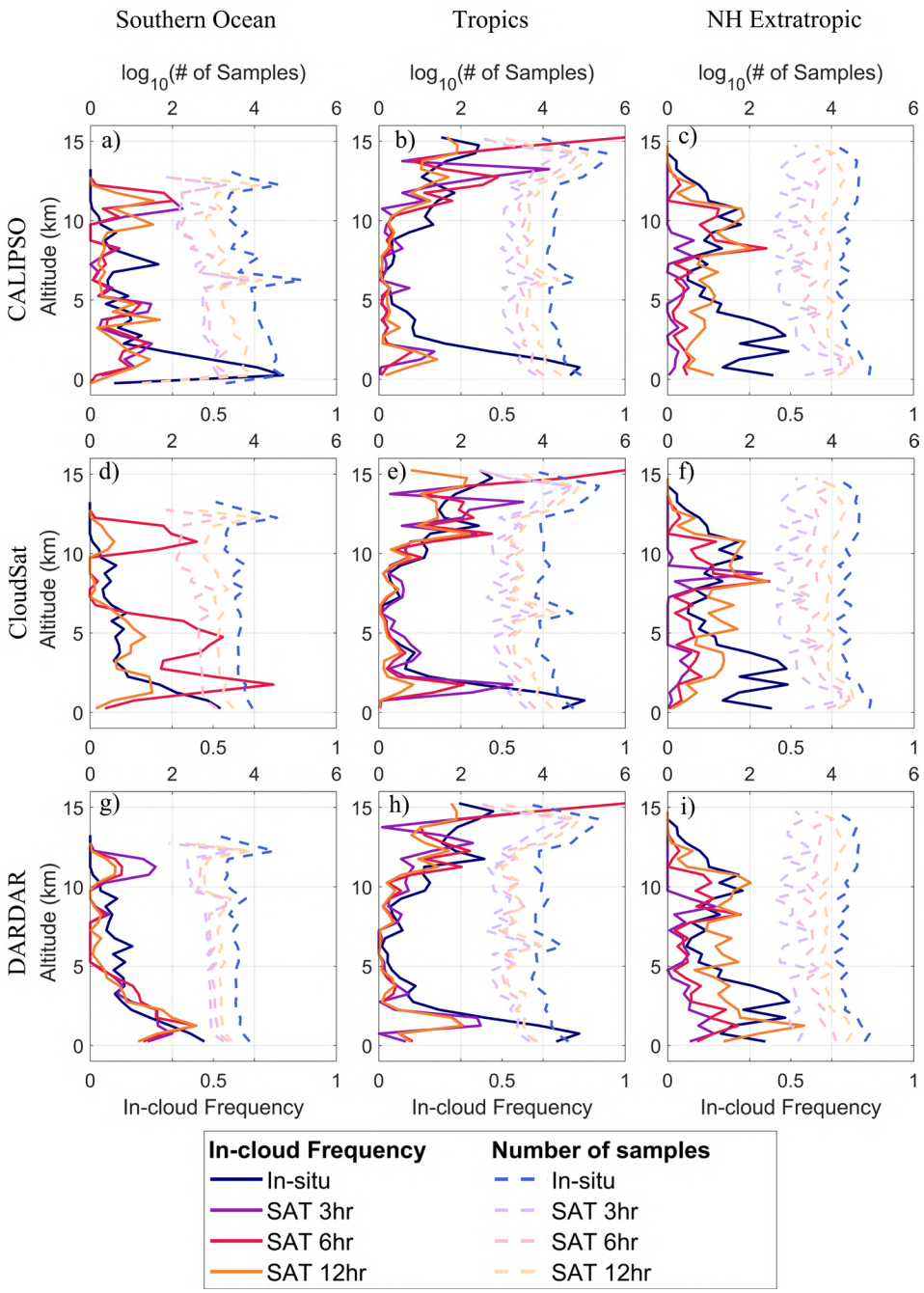

Figure 7. Similar to Figure 6, but for in-cloud frequency. Dotted lines denote the logarithmic-scale number of samples of all-sky conditions within different time windows restrictions.

The in-situ observations show a linear increase of ice phase frequency with increasing altitudes over the Southern Ocean (Fig. 6a). That is, ice phase frequency increases from 0 to 0.5 when altitudes increase from surface to 5 km, then from 0.5 to 1 when altitudes increase from 5 to 8 km. In the tropics (Fig. 6b), such an increasing trend is faster when ice phase frequency increases from 0 to 0.5 within the range of 4 – 5.5 km, while the latter part of the increasing frequency from 0.5 to 1 occurs at a similar range (i.e., 5 – 11 km) to the in-situ observations. In the NH extratropics (Fig. 6c), the ice phase frequency only changes by 0.1 below 4 km or above 5.5 km but changes drastically from 0.25 to 0.75 from 4 to 5.5 km. These results indicate that the Southern Ocean has the thickest mixing region ($IceF_{10}$ to $IceF_{90}$ is between 1–7 km), indicating that supercooled liquid water does not rapidly turn into ice phase as altitude increases. This result is consistent with numerous observational studies that reported extensive distributions of supercooled liquid water over the Southern Ocean (e.g., Desai et al., 2023; Yang et al., 2021). The NH extratropics has the thinnest mixing region between 4 and 6 km, indicating that once ice crystal forms in the NH extratropics, supercooled liquid water turns into ice phase such as via Wegener-Bergeron-Findeisen process or riming.

Three satellite data are able to capture the increasing ice phase frequencies with altitudes, as well as the lower ice phase frequencies in the tropics for the same altitudes compared with the other two regions. In the Southern Ocean, CALIPSO shows the most similar $IceF_{50}$ at 4.5 km as shown in the observations, while DARDAR shows much lower $IceF_{50}$ at 2.5 km, consistent with the overestimation of ice phase frequency in DARDAR shown in Fig. 4i and 5i. However, it is important to note that the limited availability of data from CloudSat and DARDAR during key NSF campaigns (HIPPO-3 and ORCAS) hinders their suitability for

comparison in the Southern Ocean. For NH extratropics, the CALIPSO data within 6 hrs of the aircraft observations show similar $IceF_{50}$ near 4.5 km as the in-situ observations, while CloudSat and DARDAR both show $IceF_{50}$ near 2.5 km. On the other hand, CALIPSO data within 12 hrs of the aircraft observations show $IceF_{50}$ near 2.5 km as well, which indicates that the temporal window used to restrict satellite validation has a significant impact on the comparison result in the NH extratropics. Such impact of temporal variability is smaller in the Southern Ocean and tropics. Such larger temporal variabilities of ice phase frequencies in the NH extratropics may be caused by larger diurnal variations in temperature profiles over land than over the ocean. Since most of the research flights took place during the daytime, within 6 hrs of the flights are mostly still in the daytime while within 12 hrs of the flights could be both daytime and nighttime. CloudSat also shows increases in ice phase frequencies for each altitudinal bin when the temporal window increases from within 6 hr to within 12 hr. DARDAR does not show such a strong effect from the temporal window, possibly because DARDAR already shows strong positive biases of ice phase and thus the diurnal variation is not as significant.

In-cloud frequencies are compared among three regions in Fig. 7. A high-altitudinal cirrus cloud layer is seen in the in-situ observations, which is located at higher altitudes in the tropics due to the higher tropopause height. This is consistent with previous studies that reported high occurrence frequencies of cirrus clouds slightly below the thermal tropopause (e.g., Diao et al., 2015). All three satellite data also capture the existence of such a cirrus layer. For the mid and lower troposphere, the in-cloud frequencies are around 25% to 40% in the in-situ observations, which is relatively similar to all three satellite data. Using different temporal windows

generally can change the in-cloud frequencies of satellite data by 0.1 to 0.2, which is not as significant as the impacts on ice phase frequencies as seen in the NH extratropics (Fig. 6c and 6f). The most significant impacts of temporal windows occur at the cirrus level over the Southern Ocean (Fig. 7a, 7d, and 7g), with lower in-cloud frequency at the wider window (within 12 hrs), indicating higher cirrus occurrences at daytime than nighttime over this region.

1.4.5. Seasonal and Longitudinal Variability of Ice Phase Frequency based on Satellite Data

The previous analysis in Section 1.4.4 indicates variations in ice phase frequency with different geographical locations and temporal windows. To further examine the impacts of spatial and temporal variabilities of ice phase frequency, we compare satellite observations in four longitudinal ranges (Fig. 8) and four seasons (Fig. 9). Larger sample sizes are needed in order to examine these spatial and temporal variabilities; therefore we examine one-year samples from the three satellite products without applying the selection criteria. In-situ observations are not involved in this analysis since the average of all in-situ observations is already shown in Fig. 5.

The four longitudinal ranges include global samples and three longitudinal bands: West Pacific (WP) (120˚E–180˚E), East Pacific (EP) (120˚W–180˚W), and America (20˚W–120˚W). The global distributions of ice phase frequencies of all three satellites (Fig. 8a–8c) are similar to those shown in Fig. 4 and 5, even though the global distributions in Fig. 8 allows more regions to be examined compared with those selected by Method 2 (as illustrated in the map of Fig. 1b). Among three satellite data, CALIPSO (Fig. 8a) still shows the most similar mixing region compared with in-situ observations (Fig. 5a) In addition, the main biases in the satellite observations are similar to those shown in Fig. 4 and 5, including the lack of

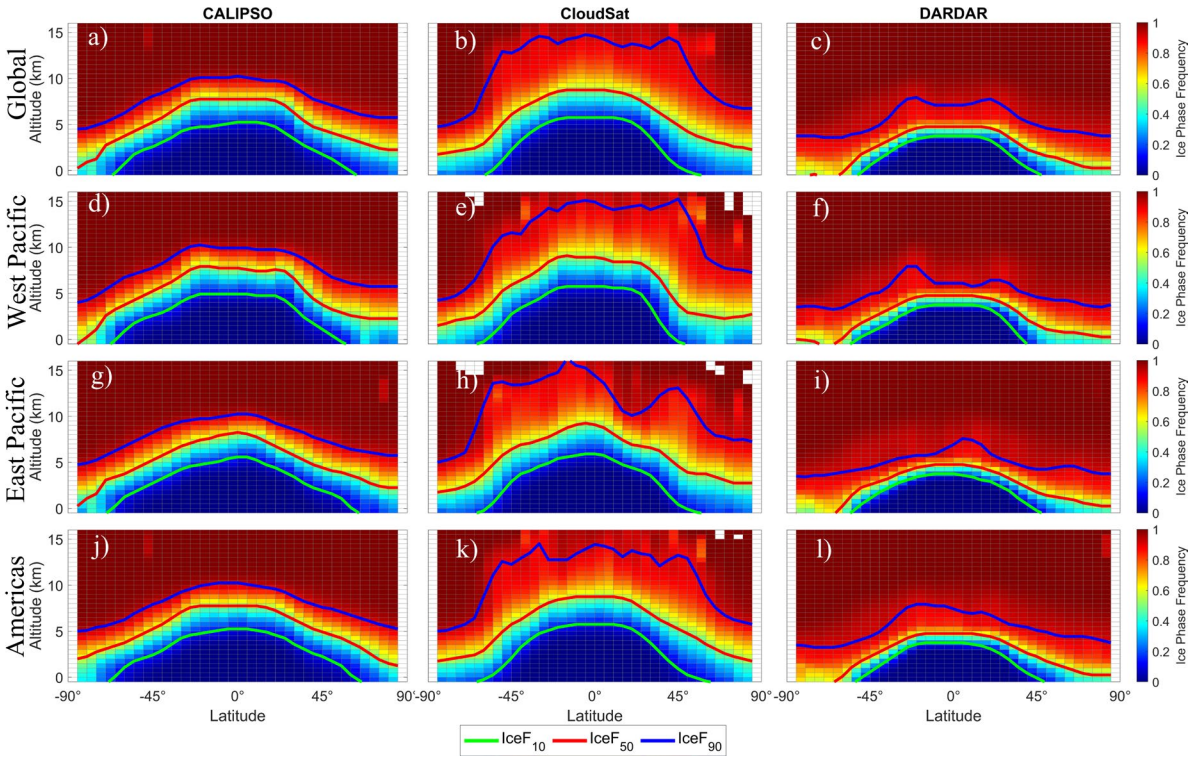

Figure 8. Latitudinal-altitudinal view of ice phase frequency for three longitudinal bands: West Pacific, East Pacific, and Americas. Satellite data represent the entire year of 2010. Colored solid lines denote 10% (green), 50% (red), and 90% (blue) ice phase frequency.

hemispheric differences, the overestimation (underestimation) of ice phase frequencies in DARDAR (CloudSat), as well as the overestimation (underestimation) of the mixing region thickness in CloudSat (DARDAR). This indicates that the previous validation results against the in-situ observations are statistically robust when using either Method 1 or Method 2 to restrict satellite data surrounding 1-Hz in-situ observations or flight campaigns. This also indicates that the differences between in-situ observations and satellite data cannot solely be attributed to their spatial mismatches.

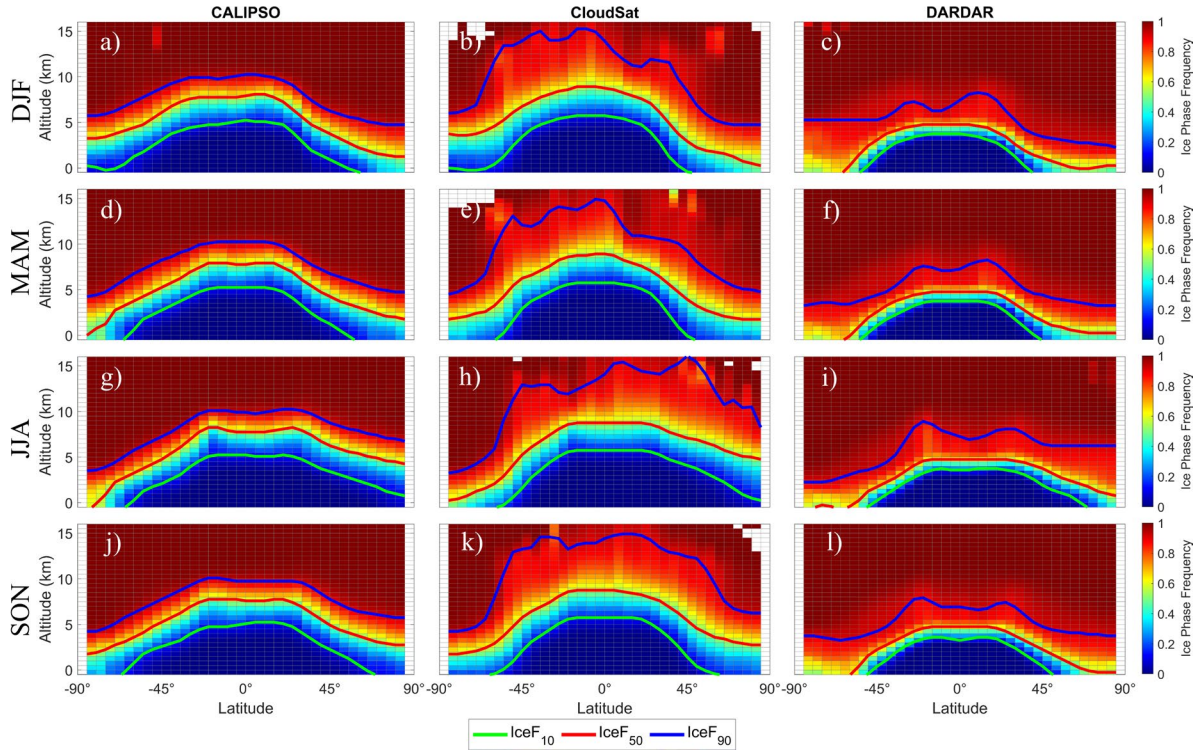

Figure 9. Similar to Figure 8 but separated by four seasons using three satellite products in the entire year of 2010.

In terms of seasonal variability (Fig. 9), CALIPSO and CloudSat show similar variabilities in ice phase frequencies in four seasons: December, January, and February (DJF); March, April, and May (MAM); June, July, and August (JJA); and September, October, and November (SON). That is, when NH is in boreal winter (DJF) or when SH is in austral winter (JJA), the winter hemisphere shows higher ice phase frequencies and lower IceF lines in the extratropics compared with other seasons for the same latitudinal/altitudinal bin. Such seasonal variability is shown most clearly in the extratropics CALIPSO and CloudSat, while the tropics have relatively constant ice phase frequencies among four seasons. DARDAR shows much smaller variations among the four seasons, with only slightly lower ice phase frequencies in the NH polar region in JJA compared with other

seasons. This lack of seasonal variability in DARDAR may be caused by its overestimation of the ice phase in general, which is consistent with the lack of sensitivity to temporal windows shown in Fig. 6i. Overall, the results from Fig. 8 and 9 indicate that the spatial variability in different longitudinal bands with different surface conditions (over land or ocean) have lower impacts than the seasonal variability.

1.5. Discussion and Conclusion

This study aims to investigate the accuracy of cloud phase identification and categorization by satellite in comparison to in-situ direct detection. To achieve this, a merged in-situ dataset was used to visualize and quantify the differences and similarities in satellite products. As shown in Fig. 3, a case study was conducted to simulate a temporal and spatial averaging comparison of cloud phase distribution. The results indicate that CALIPSO effectively captures the dependence of ice and liquid phase distribution with height. However, DARDAR tends to identify mixed phase distribution accurately but detects ice and liquid phase cross at a lower altitude due to its dependence on temperature-related variables. Fig. 2e visualizes how CloudSat over-predicts mixed phase clouds throughout the entire column. Furthermore, the thickness of the CloudSat mixing layer can be seen in Fig. 5 c, Fig. 8, and Fig. 9. Despite this, the ice phase distributions from CloudSat compare well against in-situ observations, except for mixed phase clouds, which is similar to CALIPSO. These findings emphasize the importance of combining satellite data with in-situ observations to improve the accuracy of cloud phase identification and categorization. Moreover, this study contributes to a better understanding of the strengths and limitations of different satellite algorithms for cloud phase detection.

The study investigated the ability of satellite cloud phase identification and categorization in comparison with in-situ direct detection. The results of the investigation highlighted some limitations and biases in the current satellite detection algorithms. For instance, CALIPSO single instruments (LiDAR-only) showed the inability to penetrate optically thick clouds (depicted in Fig. 2d), showed the inability to identify the cloud phase from deep convection vertical structure (depicted in Fig. 2d). Also, the retrieval algorithm used by CloudSat (Fig. 2e) showed a bias towards mixed phases, leading to over-predictions of mixed phase clouds throughout the entire column. In comparing DARDAR (Fig. 2f) with in-situ, the results did not support this product well, and more investigation is needed to adjust satellite instruments and algorithms, leading to better satellite detection. The temporal and spatial averaging comparison method has more advantages than in-situ and satellite collocational comparison, but more in-situ campaigns are required to verify how well satellites can detect cloud phases globally. The organized satellite-aircraft collocation verification can significantly improve the satellite algorithm, which can then be used on a global scale to improve the global circulation models.

The methodologies used in this study were effective. Biases between different satellite instruments and algorithms must be considered when comparing minor details. In conclusion, satellite cloud phase detection, identification, and algorithm improvement are still needed, although satellites have the benefit of spatial coverage and the ability to identify large-scale features. The combination of LiDAR and RADAR on satellites has proved the ability to acquire atmospheric characteristics at multiple altitudes, providing valuable information for weather and climate research. Future improvements in satellite technology are anticipated and will be immensely helpful for weather satellite technology, climatology research, and weather model

development. In the future, two satellite missions, NASA EarthCARE and NASA AoS, will utilize multi-sensor setups to enhance cloud phase detection. EarthCARE includes instruments such as the Atmospheric Lidar for aerosol and thin cloud profiles, the Cloud Profiling Radar for vertical cloud profiles and Doppler measurements, the Multi-Spectral Imager for across-track cloud and aerosol information, and the Broad-Band Radiometer for top-of-the-atmosphere radiance and flux measurements. The NASA AoS combines multiple satellites with instruments including a single-frequency Doppler Radar for cloud and light precipitation motion, a microwave Radiometer for various atmospheric measurements, and a Backscatter Lidar for aerosol and cloud backscatter. These missions will benefit from the findings in this chapter to optimize their cloud phase detection capabilities. It is essential to continue research into optimizing satellite algorithms and improving detection capabilities to enhance the accuracy of satellite cloud phase products on a global scale.

2. A machine learning approach to examine key factors controlling cloud phase distributions based on in-situ and satellite observations

2.1. Introduction

Cloud phase detection is an important aspect of atmospheric research, as it provides valuable information about the composition and behavior of clouds. In-situ aircraft observations and satellite remote detection are two common methods for gathering this data. However, there can be discrepancies between the results obtained from these two methods. Machine learning models provide a powerful tool for evaluating the similarity between these two sources of data. The use of machine learning models is crucial in evaluating the similarity between in-situ aircraft observations and satellite remote detection of cloud phases. This paper utilizes data from NASA CALIPSO, NASA CloudSat, and DARDAR satellite products, as well as 11 NSF EOL aircraft observation campaigns. By leveraging the power of machine learning, a deeper understanding of the relationship between these two methods of cloud phase detection can be obtained.

Recent studies using machine learning have significantly contributed to atmospheric research and the improvement of satellite observations. Various studies have demonstrated the effectiveness and necessity of incorporating machine learning techniques for satellite validation and enhancement. Kuma et al. (2023) utilized satellite observations from CERES along with ground-based data from WMO stations to train an artificial neural network (ANN) for estimating the cloud-type (i.e., cirrus, cirrostratus, cirrocumulus, altocumulus, stratocumulus, etc.) occurrence probability. The ANN showed promising performance in reproducing observed cloud types, highlighting its potential in accurately characterizing cloud phenomena. Kim et al. (2020) developed a machine learning model that used spectral

information from Meteosat-9 SEVIRI data and satellite viewing geometry as predictors, with CloudNet LWP data as the target. By employing Gradient Boosting Regression Trees (GBRTs), the model achieved a better agreement with independent CloudNet LWP observations compared to traditional physics-based approaches. This result underscores the potential of machine learning methods in advancing retrieval techniques for cloud properties. Rolf et al. (2021) employed daytime satellite imagery to train MOSAIKS, a machine-learning model that demonstrated competitive performance in various tasks. The model's ability to generalize out-of-sample across space and outperform ground truth interpolation in multiple tasks highlights its potential in capturing complex atmospheric patterns and features. While these studies provide valuable insights into the effectiveness of machine learning in satellite validation and improvement, it is worth noting that most of them primarily focus on 2-D satellite imagery to evaluate horizontal atmospheric features. In our current research, we aim to expand on these findings by exploring the advantages and limitations of utilizing vertical down-looking scanning satellites. To achieve this, we will compare and analyze the data obtained from these satellites with in-situ airborne observations. By leveraging different approaches and datasets, we hope to gain a comprehensive understanding of the capabilities and limitations of satellite-based observations for studying the atmosphere and improve our ability to accurately characterize and predict cloud phenomena.

The existing research on satellite cloud phase detection validation often overlooks the important correlation and biases caused by thermodynamic, dynamic conditions, and aerosol effects when compared with in-situ airborne observations. To address this gap, future studies should focus on analyzing these relationships more comprehensively. Noh et al. (2011)

highlight the challenge of assigning the proper microphysical phase in CloudSat retrievals. They propose potential solutions, such as utilizing CALIPSO data to identify supercooled liquid layers at cloud tops or estimating liquid water content based on temperature profiles or adiabatic liquid water content profiles. These approaches could help improve the accuracy of cloud phase determination. Coopman et al. (2020) emphasize the strong relationship between ice crystal size, cloud droplet size, and cloud top height (CTH). They also emphasize the impact of aerosols on glaciation temperature (T_g). They suggest that leveraging machine learning techniques could enhance global models by considering these aerosol effects. Additionally, incorporating reanalysis data on aerosol information through collocation would be interesting for future research. In our study, we aim to address these challenges by integrating machine learning algorithms with in-situ observed thermodynamics, dynamic conditions, and aerosol data. This integration will enable us to assess the correlation between in-situ cloud phase observations and satellite detections more comprehensively. By incorporating these additional factors, we can improve our understanding of cloud phase identification and enhance the accuracy of satellite-based cloud observations.

In this paper, a detailed analysis of the cloud phase detection similarity between in-situ aircraft observations and satellite remote detection using fine-tree machine learning models is presented. The results demonstrate the effectiveness of using machine learning to evaluate the similarity between these two methods of cloud phase detection. One key finding of the analysis is that clouds behave very differently at different temperatures and RHi ranges. In particular, the temperature range between -40°C to 0°C is the most interesting to evaluate the difference in satellite and in-situ cloud phase detection. The machine learning model provides an aspect

that describes how temperature range, relative humidity with respect to ice, and vertical velocity affect the performance of cloud phase detection.

2.2. Data

2.2.1 In-situ Observation

A global in-situ dataset is created by combining 11 individual NSF campaigns, including START08 (Pan et al., 2010), HIPPO (Wofsy, 2011), PREDICT (Montgomery et al., 2012), DC3 (Barth et al., 2015), TORERO (Volkamer et al., 2015), CONTRAST (Pan et al., 2017), WINTER (Thornton et al., 2014), CSET (Albrecht et al., 2019), ORCAS (Stephens et al., 2018), SOCRATES (McFarquhar et al., 2021), and OTREC (Fuchs-Stone et al., 2020). This extensive dataset covers a vast longitudinal range, spanning from 120°E to 40°W, and a wide latitudinal range from 75°S to 90°N. This expansive coverage provides us with a unique opportunity to analyze and understand the global distribution of cloud phases across various regions and climatic zones. By encompassing diverse geographical locations and weather patterns, this dataset offers valuable insights into the complex dynamics of cloud formation and behavior on a global scale.

The cloud phase data utilized in this study were obtained from the works of D'Alessandro et al. (2019) and Yang et al. (2021). These studies employed a cloud phase identification method that processes high-resolution, 1-Hz in-situ data captured by a range of cloud probes, including the Cloud Droplet Probe (CDP), the Fast-2 Dimensional Cloud (Fast-2DC) Probe, and the Two Dimension Stereo (2DS) Probe. Each of these probes is designed to detect cloud hydrometeors within specific size ranges. The CDP captures particles ranging from 2 to 50 μm in size, the Fast-2DC Probe covers the range of 62.5 to 3200 μm, and the 2DS Probe detects

particles within the size range of 40 to 5000 μm. By utilizing these advanced cloud probes, the cloud phase identification method employed in this study ensures accurate and comprehensive characterization of cloud phase properties at different size scales.

Onboard the in-situ airborne observations, various measurements are conducted to gather important data. The temperature data is obtained using the Rosemount temperature probe. The RHi is calculated using the Vertical Cavity Surface Emitting Laser (VCSEL) hygrometer, and it is derived using Equation 11 from Murphy and Koop (2005). The vertical wind speed is directly measured as an in-situ observation. The concentrations of aerosol particles between 55 to 1000 nm, known as Na100 and Na500, are detected using the in-situ Ultra-High Sensitivity Aerosol Spectrometer (UHSAS). These in-situ measurements provide valuable information for further analysis and comparison.

2.2.2 Satellite Observations

For the collocation between in-situ observations and satellite data, three specific satellite cloud phase products have been selected: NASA CALIPSO Vertical Feature Mask (VFM), NASA CloudSat 2B-CLDCLASS-LIDAR, and Laboratoire Atmosphères, Milieux, Observations Spatiales DARDAR-CLOUD. These products are part of the Afternoon Constellation (A-Train) constellation, with CALIPSO and CloudSat later transitioning to the C-Train.

The CALIPSO VFM utilizes LiDAR as its primary instrument for categorizing the data based on the LiDAR backscatter and linear depolarization ratio. The VFM provides information on various features, including aerosol types (e.g., smoke, marine, volcanic ash), cloud types (e.g., cirrus, deep convection), and cloud thermodynamic phase (Vaughan et al.,

2004). Both CloudSat 2B-CLDCLASS-LIDAR and DARDAR-CLOUD combine data from the radar onboard NASA CloudSat and the LiDAR from CALIPSO. This combination leverages the strengths of both sensor types, as LiDAR is better at detecting optically thin clouds while radar excels at observing optically thick clouds. The main difference between CloudSat and DARDAR data lies in their algorithm development. The CloudSat algorithm is based on Wang and Sassen (2001), while the DARDAR algorithm, known as "Varcloud," is based on Delanoë and Hogan (2008b).

By utilizing these three satellite products, it becomes possible to identify and quantify the performance of different algorithm and instrumentation combinations, providing a comprehensive and quantitative representation of cloud phase identification from satellite observations.

2.3. Machine Learning Setup

In this study, a fit ensemble of learners for classification was employed, utilizing decision tree learners and discriminant analysis classifiers. The goal was to create a predictive classification ensemble by incorporating all available predictor variables in the dataset. Additionally, a reduced set of predictors was used to train another ensemble, enabling the assessment of variable selection's impact on predictive accuracies. The process flow chart is shown in Fig. 10. The decision to use decision tree machine learning over other methods like neural networks, K-nearest neighbors (KNN), and others was based on key factors. Decision trees provide interpretability, handle different data types efficiently, and require minimal preprocessing. In contrast, neural networks lack interpretability and can be computationally expensive, while KNN may struggle with high-dimensional data and require additional steps.

Other methods may have limitations in handling nonlinear relationships, outliers, or missing values. Considering these factors, decision trees were favored for their interpretability, efficiency, and versatility in handling various data complexities, making them a valuable choice for accurate cloud phase detection.

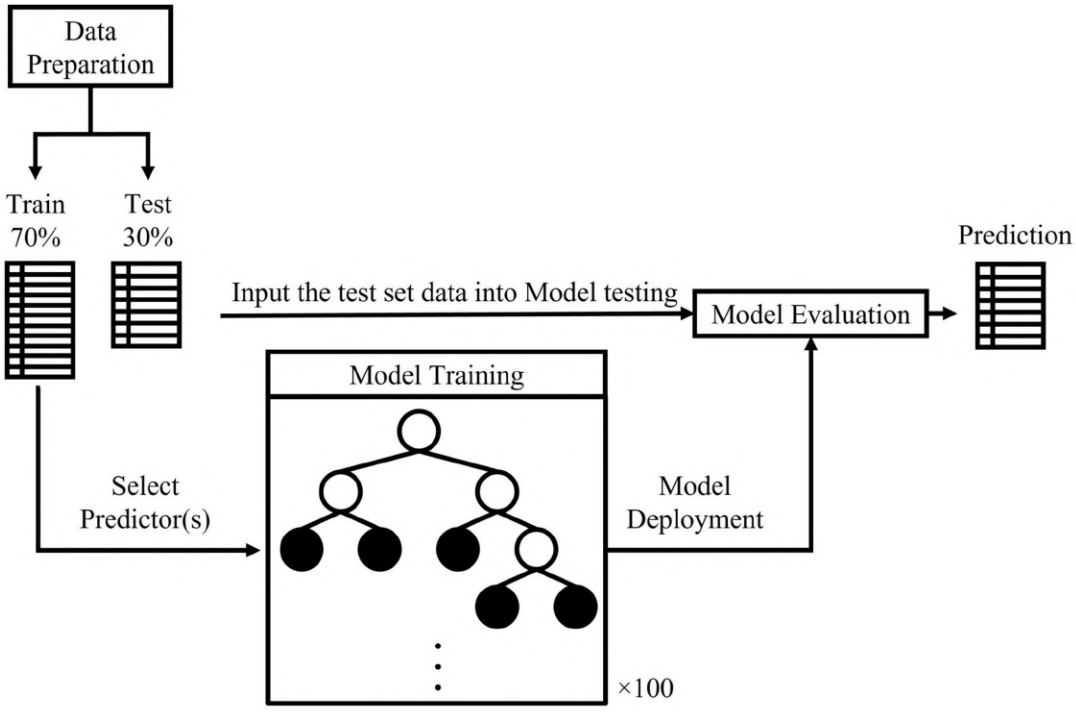

Figure 10. Flow chart of the Machine Learning process. Data preparation split the full dataset into a 30% test set and a 70% train set. Before the fine tree model ensemble training predictor(s) is selected. After model training, the model is evaluated using the test set and outputs a prediction dataset.

For the construction of decision tree ensembles, a uniformly specified algorithm was utilized to determine optimal splits on categorical predictors, select the split criterion, and determine the number of predictors used for each split (Gordon et al., 1984; Loh & Shih, 1997). Random Forest, a popular ensemble model, was employed where all individual decision tree models were generated simultaneously during training. During prediction, each tree

independently produced its own prediction, and the final prediction was determined through a majority conclusion among the trees. This approach leveraged the ensemble agreement of the trees to enhance accuracy, reduce overfitting, and provide more robust predictions. The implementation of "surrogate" decision splits was another aspect of the study. Surrogate splits allowed decision trees to incorporate up to 10 surrogate splits at each branch node, acting as backups for cases where original splits couldn't be effectively used, such as with missing values or weak features. Surrogate splits aimed to capture similar patterns or relationships as the original splits, improving model robustness and accuracy in datasets with missing values.

The decision tree ensembles were fine tree models with a specified "MaxNumSplits" parameter set to 100, determining the maximum number of decision splits or branch nodes per tree. This parameter controlled the depth of the trees, with lower values leading to shallower and simpler models and higher values allowing for more complex patterns to be captured. To address the data sample imbalance between clear sky and in-cloud conditions, the "RUSBoost" random undersampling boosting method was employed (Seiffert et al., 2008). RUSBoost combined random undersampling of the majority class with boosting techniques to create a strong predictive model. RUSBoost helped handle class imbalance and improved the model's ability to accurately predict both majority and minority classes. The decision tree template was chosen as the base learner for the ensemble, and the "NumLearningCycles" parameter determined the number of iterations of training. Increasing the value of "NumLearningCycles" added more weak learners to the ensemble, refining the model's predictions. The value of 100 was selected for "NumLearningCycles" to strike a balance between predictive power and computational resources. This approach of utilizing ensembles, surrogate splits, and RUSBoost

demonstrated the aim of improving predictive performance and addressing challenges such as missing values and class imbalance in the dataset.

In the validation process, the test set includes both the predictors (features) and the true answers (cloud phase labels). The machine learning model uses the predictors from the test set to make predictions on the cloud phase. After the predictions are made, the true answers from the test set are compared with the predicted values to assess the accuracy and performance of the model. This comparison allows for evaluating how well the machine learning model aligns with the actual cloud phase observations in the test set. By comparing the predicted values with the true answers, the model's performance can be assessed and validated against the known cloud phase data in the test set.

2.4. Results

2.4.1 Thermodynamic and Dynamic Distribution of Ice Phase Occurrence and In-Cloud Frequency

2.4.1.1 Temperature and RHi Distribution

The evaluation of in-cloudF provides crucial insights into the distribution and occurrence of clouds at different temperatures and RHi ranges. This metric is derived by dividing the cloud occurrence (including liquid, mixed, and ice phase clouds) by the total sky conditions detected through in-situ observations. Similarly, the iceF offers a measure of the relationship between temperature, RHi, and satellite observations. The iceF is calculated by dividing the occurrence of ice phase clouds by the total number of in-cloud detections.

Analyzing the in-cloudF from CALIPSO (Fig. 11d) reveals a striking similarity to the in-situ observations (Fig. 11a), indicating the accuracy of CALIPSO in capturing cloud occurrences. However, CALIPSO exhibits slightly higher cloud occurrences below -20°C and

above 100% RHi (Fig. 11g), suggesting possible biases in its cloud detection algorithm. Both CloudSat and DARDAR (Fig. 11g and 11j) exhibit comparable distributions of in-cloudF, implying consistent patterns of cloud occurrence. CloudSat demonstrates a modest dependence on temperature and RHi, with higher cloud occurrences observed at colder temperatures and higher RHi values. DARDAR shares similar characteristics with CloudSat but displays cloud occurrences at multiple locations above -20°C, without any apparent relationship with temperature and RHi.

The number of samples shown in Fig 11 (c, f, I, and l) represents the available data for comparison. However, the number of samples from the satellite data is limited due to sampling restrictions. The in-situ composite dataset provides comprehensive data below 100% RHi across all temperature. However, there is a decrease in data between -20°C to -40°C and 50% to 100% RHi. Additionally, there are fewer data points above 100% RHi. In all three satellite products, there is a higher availability of data between 0°C to -20°C and -40°C to -70°C for this evaluation.

Examining the iceF, CALIPSO's observations (Fig. 11e) exhibit a comparable distribution to the in-situ data (Fig. 11b), confirming its ability to detect the presence of the ice phase. The temperature dependence of the ice phase is significant, with no ice phase detection above 0°C and a complete ice phase below -40°C. In the temperature range between 0°C and -40°C, CALIPSO's iceF demonstrates a strong correlation with RHi. Similarly, DARDAR (Fig. 11k) exhibits temperature dependence, although it shows ice occurrence even above 0°C. The iceF of CloudSat (Fig. 11h) demonstrates a pronounced reliance on the temperature within the range of 0°C to -40°C. However, below -40°C, the iceF does not reach 100% for CloudSat,

suggesting potential issues with its cloud phase identification algorithm, leading to an over-prediction of mixed-phase clouds.

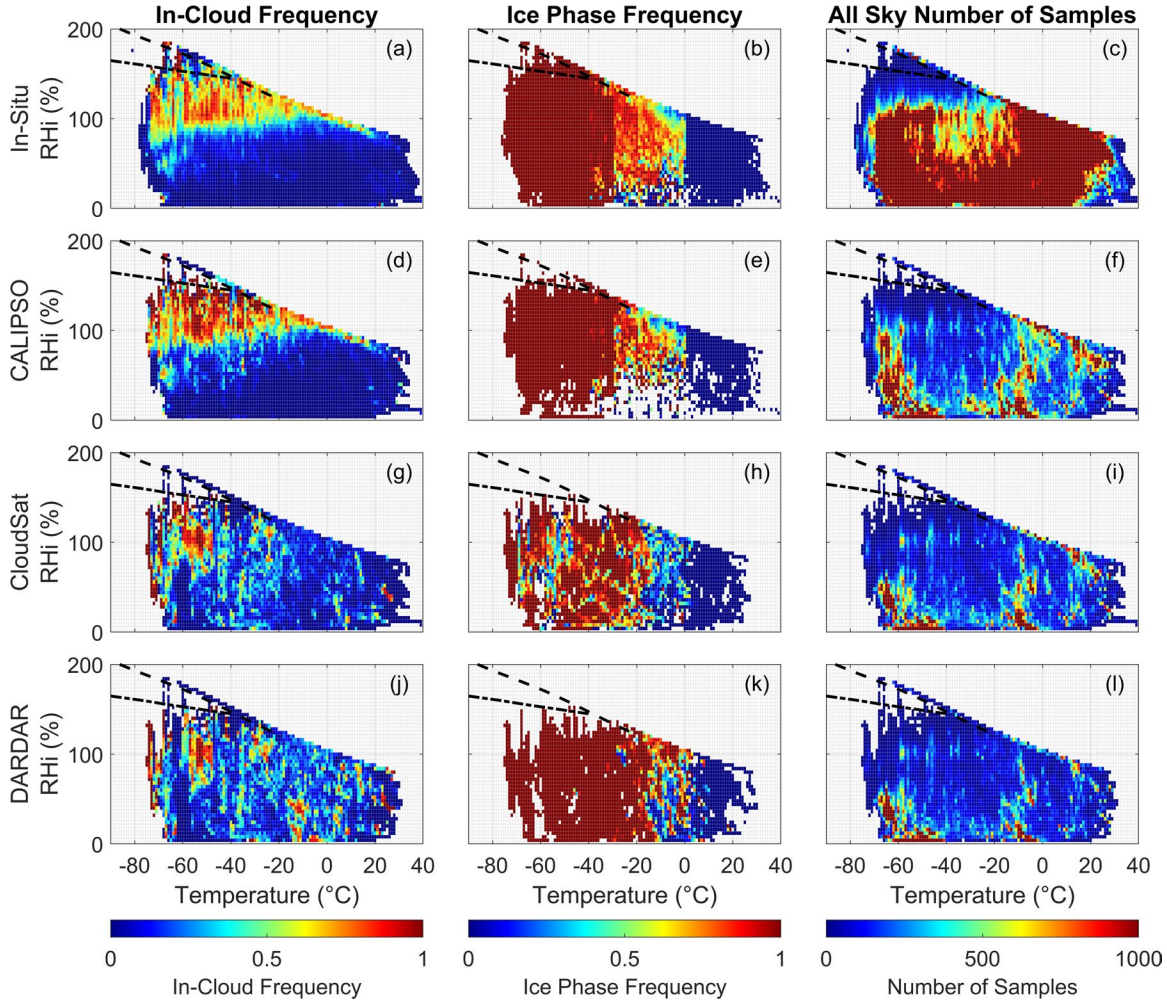

Figure 11. Distribution of RHi at different temperatures from (a, d, g, and j) in-cloudF, (b, e, h, and k) iceF, and (c, f, i, and l) the number of samples. The dashed-dotted line indicates the homogenous freezing threshold for 0.5 µm aerosols based on Koop et al. (2000). The dotted line shows the liquid saturation line calculated based on saturation vapor pressure with respect to liquid (Murphy and Koop, 2005).

2.4.1.2 Temperature and Vertical Velocity Distribution

Similarly to the previous section, this section examines the distribution of temperature and w and its impact on cloud occurrence and ice phase frequency.

The in-cloudF observed in-situ (Fig. 12a) reveals a higher cloud occurrence at lower temperatures and higher vertical velocities, both upward and downward. This pattern is also evident in the three satellite cloud phase products (Fig. 12d, 12g, and 12j). Notably, DARDAR (Fig. 12j) detects more clouds above -20°C compared to the other observations.

The iceF exhibited by the in-situ measurements, CALIPSO, and DARDAR (Fig. 12b, 12e, and 12k) shows a dependence on temperature rather than vertical velocity. However, CloudSat (Fig. 12h) displays a lower ice phase frequency below -40°C, consistent with the distribution observed in the previous figure. Additionally, an intriguing pattern emerges between -20°C and -40°C for CloudSat, where the iceF decreases to 50% at a vertical velocity of ±1 m/s.

These findings shed light on the relationship between temperature, vertical velocity, and cloud properties. The observed higher cloud occurrence at lower temperatures and increased vertical velocities suggest a possible influence of atmospheric dynamics on cloud formation and behavior. The temperature-dependent ice phase frequency highlights the sensitivity of ice formation to thermal conditions. The unique patterns exhibited by CloudSat, such as the lower iceF below -40°C and the notable decrease in iceF at specific vertical velocities, warrant further investigation into the underlying mechanisms and potential biases in cloud phase identification algorithms. The number of samples depicted in Fig 12 (c, f, I, and l) reflects the available data

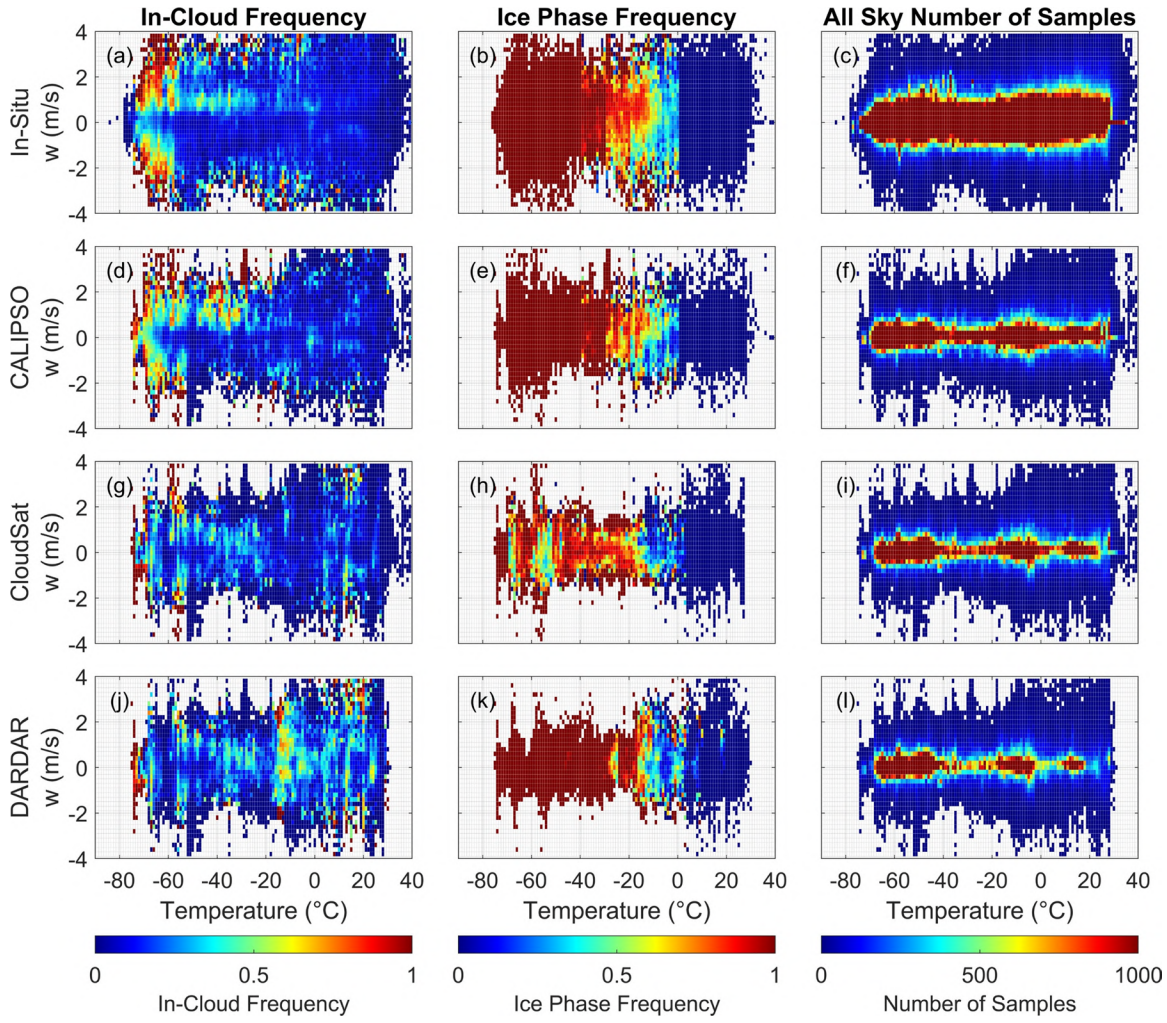

Figure 12. Similar to Figure 11, but for the Distribution of w at different temperatures from (a, d, g, and j) in-cloudF, (b, e, h, and k) iceF, and (c, f, i, and l) the number of samples.

for comparison, similar to Figure 11. In-situ and satellite observations all reveal a substantial amount of data sampled where vertical velocity is between -1 m/s and 1 m/s., as well as increased sampling between 0°C to -20°C and -40°C to -70°C temperature ranges.

The evaluation of in-cloudF and iceF provides valuable insights into the distribution and occurrence of clouds and their relation to temperature, RHi, and w. The comparison of satellite observations, such as CALIPSO, CloudSat, and DARDAR, with in-situ measurements reveals

similarities and differences in cloud occurrence and ice phase detection. CALIPSO exhibits accurate cloud occurrence and ice phase detection, although it shows slight biases at specific temperatures and RHi ranges. CloudSat and DARDAR demonstrate consistent patterns of cloud occurrence, with CloudSat displaying temperature and RHi dependence. However, potential issues with cloud phase identification are observed, particularly in over-predicting mixed-phase clouds. The iceF analysis confirms temperature dependence, with CALIPSO showing significant sensitivity to temperature and RHi, while DARDAR and CloudSat exhibit similar temperature patterns but with some deviations. The examination of vertical velocity reveals higher cloud occurrence at lower temperatures and increased vertical velocities. These findings emphasize the importance of evaluating satellite algorithms and contribute to improving satellite-based cloud observations and enhancing our understanding of cloud processes in the atmosphere.

2.4.2 Cloud phase probability density function of in-situ temperature and RHi

The cloud phases probability density function (PDF) (Fig. 13) represents the normalized distribution of the total number of samples in different temperature and RHi ranges, based on in-situ observations of clear sky, liquid phase, mixed phase, and ice phase. The PDF for the in-situ clear sky reveals that all three satellite products exhibit consistency with the in-situ observations in terms of temperature distribution. However, when considering the PDF-RHi distribution, all satellites display higher PDF values than the in-situ clear sky (Fig. 13a), indicating a higher frequency of cloudy conditions as perceived by the satellites.

For the liquid phase PDF (Fig.13b and 13f), CALIPSO shows the most similar trend to the in-situ observations, while CloudSat exhibits a sharp decrease outside the 0°C to 10°C range,

and DARDAR shows an over-identification between 0°C and -20°C. In terms of the PDF-RHi distribution, all three satellite products display a peak at around 80% RHi, indicating a consistent identification of in-cloud conditions. Moreover, the higher PDF values below 100% RHi suggest that the satellites project a more cloud-prone atmosphere compared to the in-situ observations. In the mixed phase distribution, both CloudSat and DARDAR show an over-detection of the mixed phase when compared to the in-situ observations in the temperature distribution. The PDF-RHi distributions reveal that both satellite products tend to over-identify the mixed phase below 100% RHi by orders of magnitude.

The PDF for the ice phase exhibits nearly perfect alignment with the in-situ observations. CALIPSO and CloudSat detect fewer ice phase occurrences than the in-situ observations at temperatures above -20°C, indicating a perception that the ice phase is less frequent at warmer temperatures. However, the DARDAR algorithm projects a higher ice phase PDF than the in-situ observations at temperatures above -20°C. The satellite products consistently display PDF-RHi distributions with lower PDF values than the in-situ observations above 70% RHi, and higher PDF values below 70% RHi.

The distribution of the number of samples for PDF for each cloud phase based on in-situ and satellite observations (Fig. 14). In clear sky conditions, the in-situ data exhibit the highest number of samples across all temperature ranges and RHi ranges, followed by CALIPSO, CloudSat, and DARDAR. However, for temperatures below -40°C and RHi above 125%, all three satellite datasets provide similar amounts of data. Regarding the liquid phase, in-situ observations have the most data above 0°C and RHi greater than 50%, while CALIPSO and CloudSat have the most data at RHi below 50%. CloudSat dominates the mixed phase

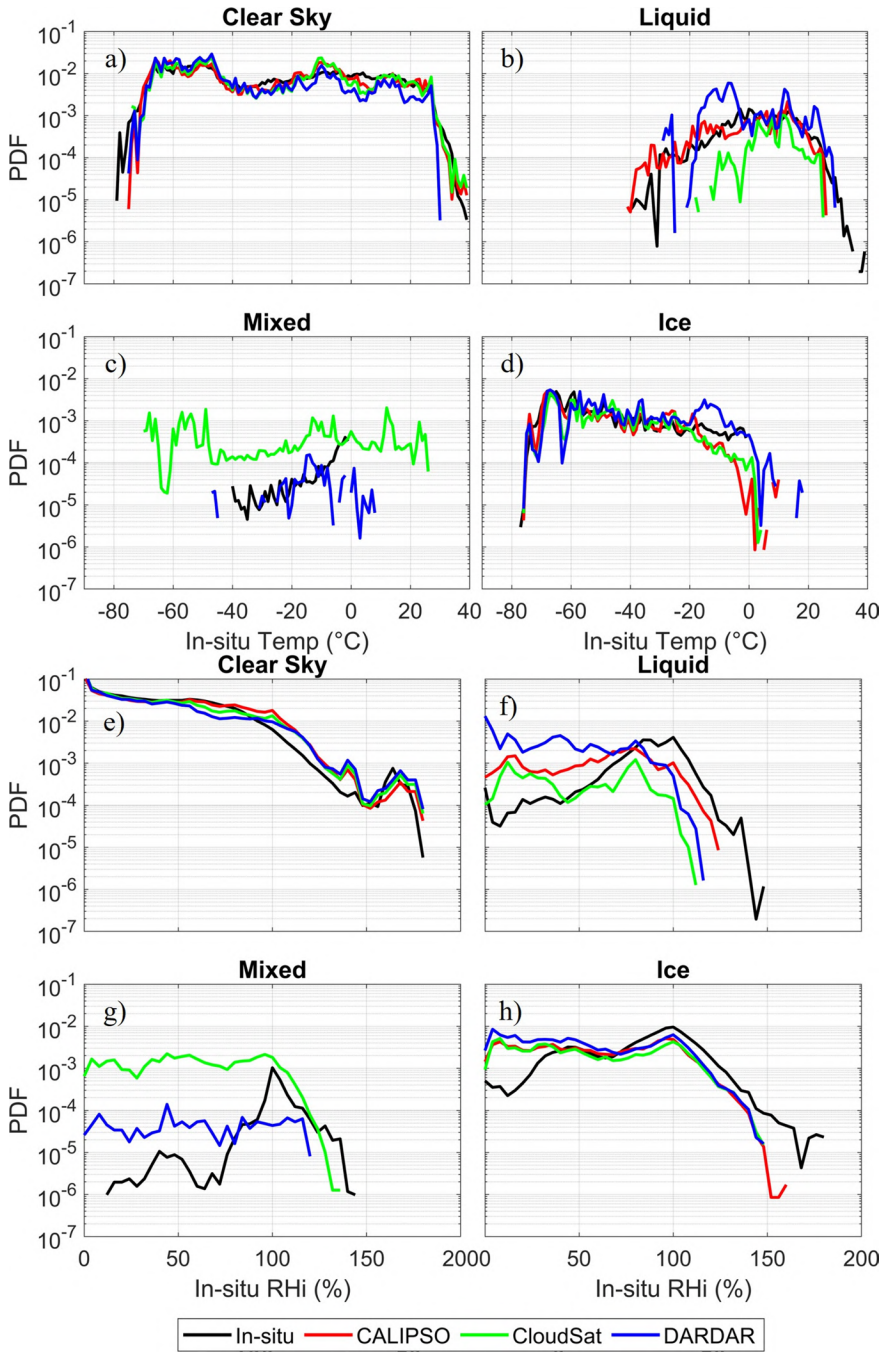

Figure 13. cloud phase probability density function with respect to (a - d) temperature and (e - h) RHi.

category. Finally, the ice phase is predominantly represented by the in-situ observations in all temperature ranges and for RHi values above 25%.

Overall, the PDF analysis provides a clearer understanding of how each satellite platform performs in distinguishing clear sky from the three cloud phases. In conclusion, the satellites tend to identify more clouds in the RHi range below 70% and less cloud detection in the RHi range above 70%, as compared to the PDF derived from the in-situ observations.

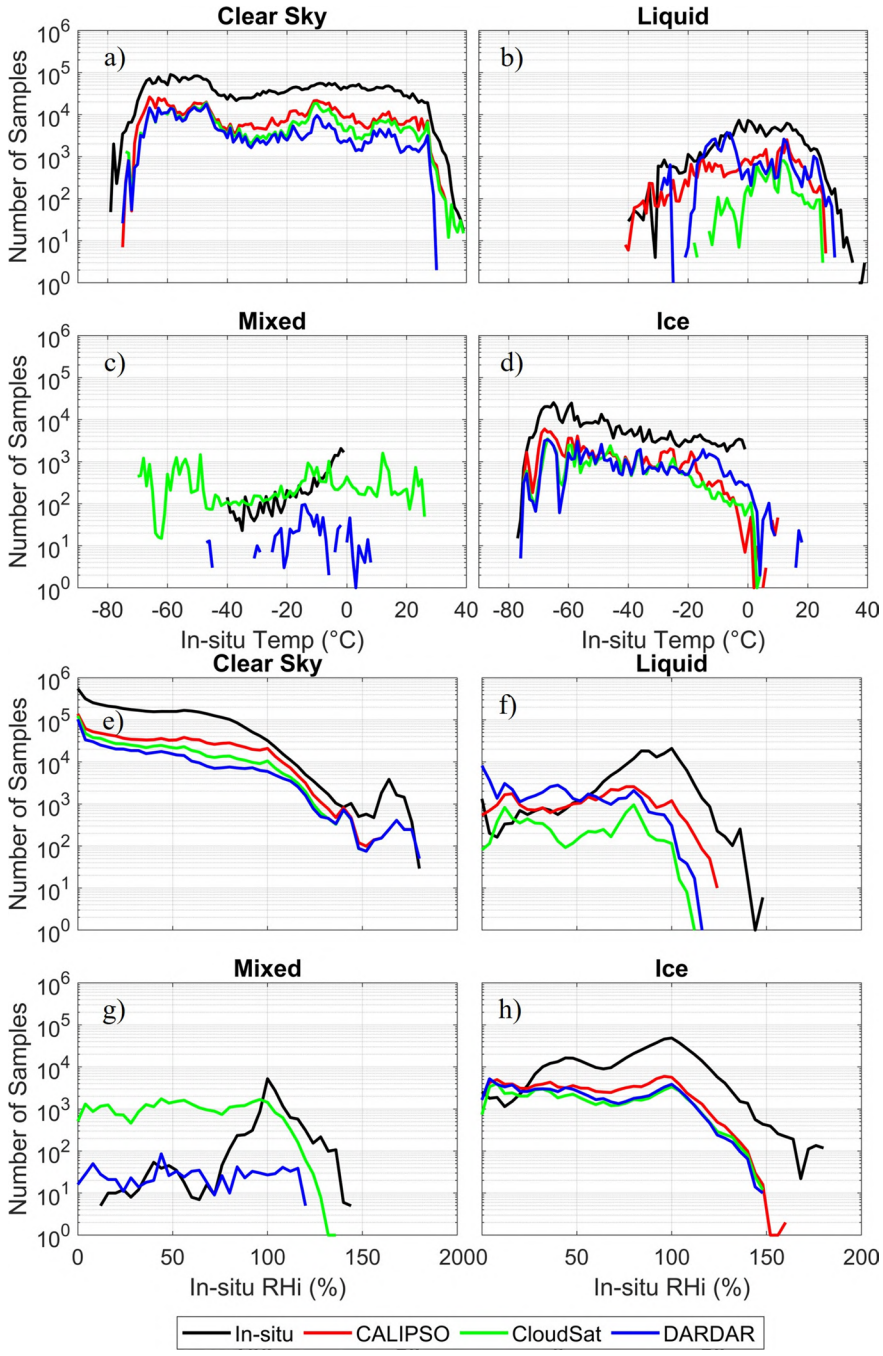

Figure 14. cloud phase probability density function number of samples with respect to (a - d) temperature and (e - h) RHi.

2.4.3 Linear regression fit of ice phase and in-cloud frequency at in-situ temperature and RHi

2.4.3.1 Ice Phase Occurrence Frequency

Ice phase frequencies per temperature bin provide valuable insights into the distribution and comparison of ice phase between satellite and in-situ aircraft measurements. By dividing the dataset into distinct temperature intervals, we can analyze the occurrence of supercooled liquid and ice phases within the mixed phase temperature range. The evaluation involves comparing machine learning models (T model, T+RHi model, and T+RHi+w model) with in-situ and satellite observations.

Figure 15 presents the results for each temperature range. Observations indicate a decreasing trend in IceF with increasing temperature, with CloudSat encountering challenges below -40°C. The machine learning models capture this trend to varying degrees. The T+RHi model shows an increase in IceF for in-situ observations between -40°C to -30°C, while the satellites consistently exhibit a decrease in IceF with higher temperatures. The T+RHi+w model follows a similar trend to the T+RHi model.

Below -40°C, only the observation from CloudSat shows ice phase frequencies below 100%, which is well captured by the machine learning models. CALIPSO and DARDAR exhibit a decrease in frequencies closer to -40°C. Between -40°C and -30°C, observations maintain IceF above 80% for all platforms, whereas the models struggle to predict high ice phase frequencies. The best agreement occurs in the temperature range of -20°C to -10°C, where the models reasonably represent the observations.

In summary, machine learning models effectively identify broader trends in ice phase

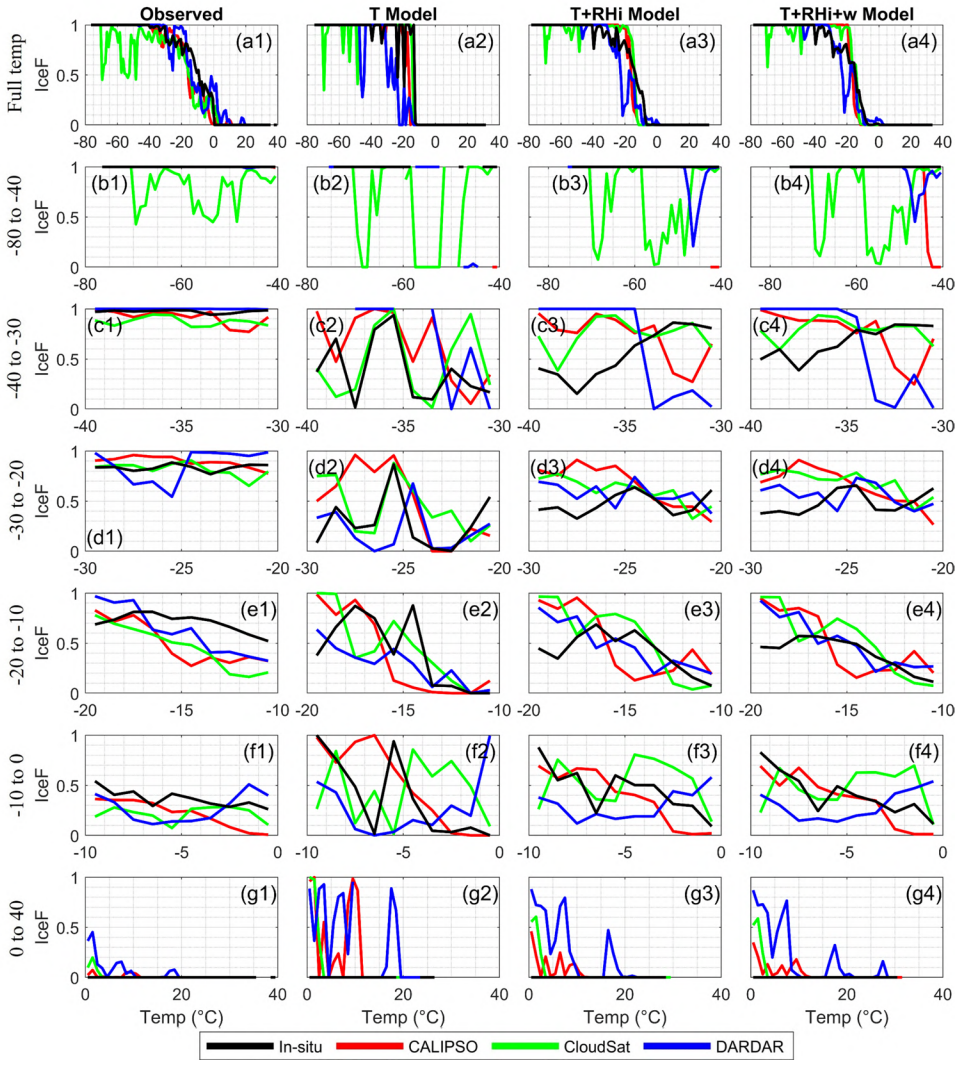

Figure 15. The iceF with respect to temperature. Column 1 shows the observation distribution, column 2 is the T model, column 3 is the T+RHi model, and column 4 shows the T+RHi+w model. Row a shows the IceF distribution in full temperature range, row b shows temperature range between -80°C to -40°C, row c is -40°C to -30°C, row d is -30°C to -20°C, row e is -20°C to -10°C, row f is -10°C to 0°C, and row g shows temperature range above 0°C.

frequencies across a wide temperature range. However, capturing the similarity of ice phase frequencies between specific temperature ranges presents a greater challenge. The T+RHi model performs well in predicting ice phase frequencies, while the T+RHi+w model shows

limited improvement, and the T model's predictions are less organized. Figure 16 reveals that the T model performs well in relation to observed RHi, while the T+RHi and T+RHi+w models struggle to capture the relationship between ice phase distribution and RHi.

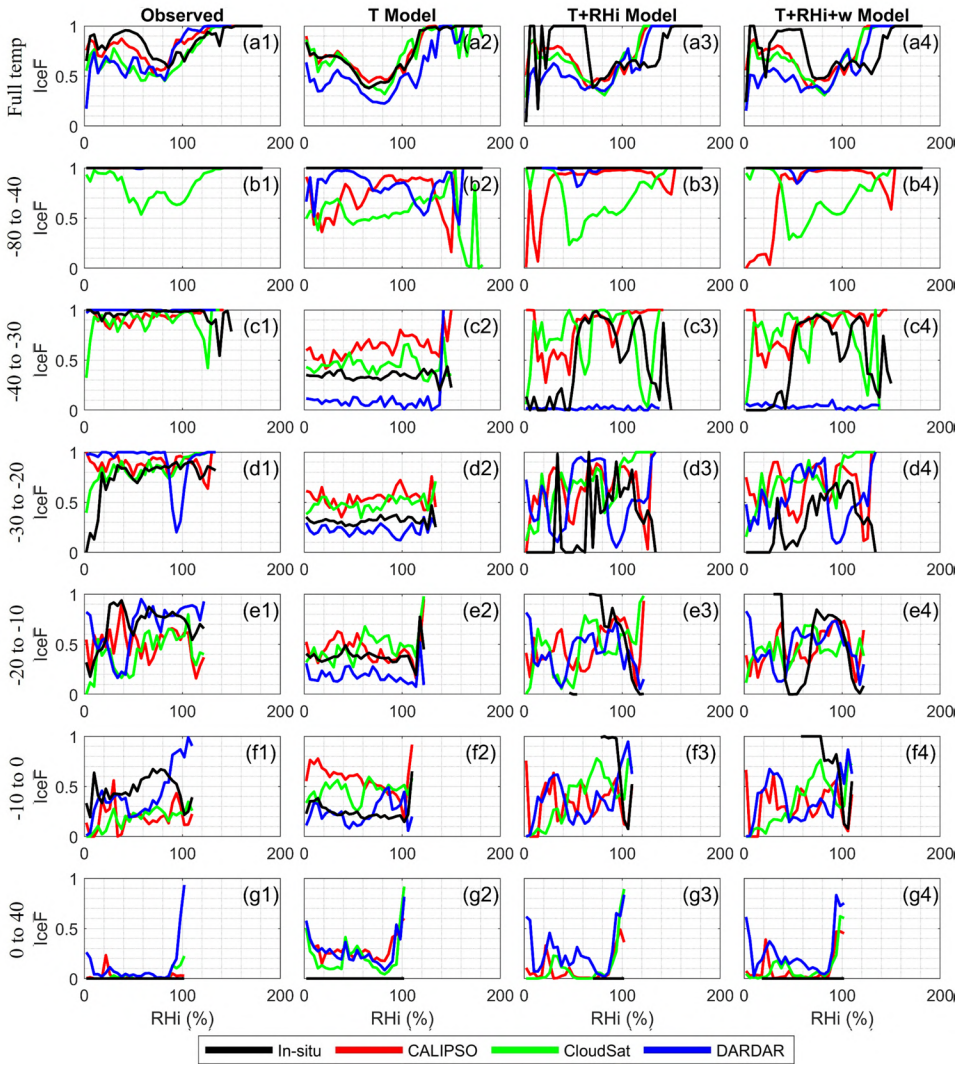

Figure 16. Similar to Figure 15 but The iceF with respect to RHi.

2.4.3.2 In-Cloud Occurrence Frequency

The comparison between in-cloudF observations and ML predictions in Fig. 17 and Fig. 18 focuses on evaluating the performance of ML models across different temperature ranges. Overall, there are good agreements between satellite platforms and in-situ observations, indicating the effectiveness of ML models in capturing cloud occurrence patterns. However, DARDAR shows higher in-cloudF than the in-situ observations in the temperature range of -20°C to 20°C, suggesting the cloud identification algorithm that require further investigation.

The limited performance of the T model across all temperature ranges can be attributed to its sole reliance on temperature as a predictor. Cloud formation involves multiple factors beyond temperature, such as relative humidity with respect to ice and atmospheric vertical velocity. The omission of these factors restricts the T model's ability to accurately represent cloud occurrence. The T+RHi model improves the prediction by incorporating relative humidity but still overpredicts cloud occurrence in all temperature ranges. The T+RHi+w model does not show significant improvement compared to the T+RHi model.

Figure 18 illustrates that the T model is not aware of the relationship between cloud occurrence and relative humidity, and overpredicts cloud occurrence to almost 100% across all temperature ranges. Both the T+RHi model and T+RHi+w model struggle to understand that low relative humidity should not result in high in-cloudF, leading to increased randomness in their predictions. In contrast, the T+RHi model and T+RHi+w model consistently perform well and mirror each other in most cases. By incorporating relative humidity alongside temperature, these models enhance their predictive capabilities. Considering multiple variables provides a

more comprehensive understanding of cloud formation conditions and improves the accuracy of cloud occurrence predictions.

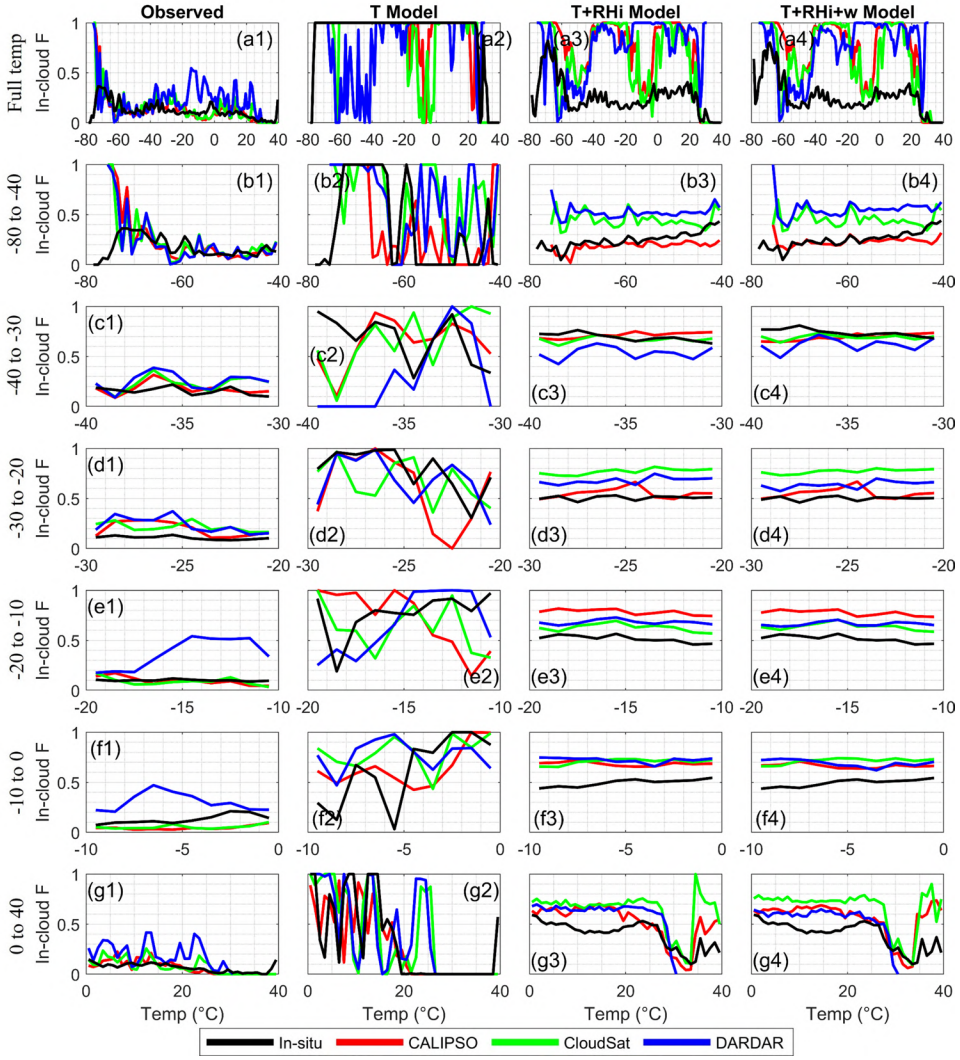

Figure 17. Similar to Figure 15, but for the in-cloudF with respect to temperature.

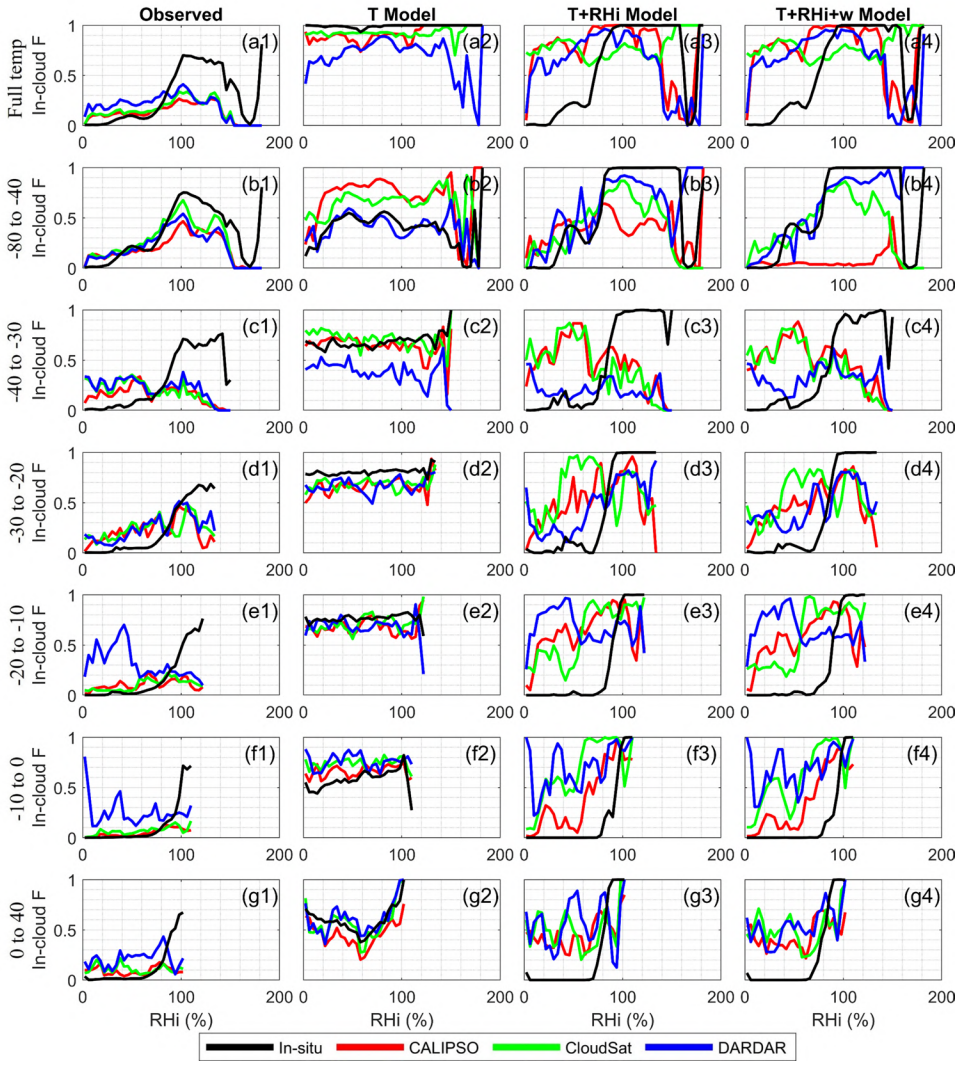

Figure 18. Similar to Figure 15, but for the in-cloudF with respect to RHi.

This comparison emphasizes the importance of incorporating multiple variables when developing ML models for cloud prediction. The temperature should be complemented by other relevant parameters like relative humidity to capture the complex nature of cloud formation. These findings highlight the need for comprehensive approaches to modeling cloud

occurrence, leading to more accurate predictions and a deeper understanding of the interplay between temperature, humidity, and cloud dynamics.

2.4.4 Using ML to evaluate each variable importance to the cloud condition

2.4.4.1 In-cloud and clear sky condition evaluation

This section focuses on comparing the reliance of different observation platforms by separating cloud phase data into the clear sky and in-cloud conditions (Fig. 19). By utilizing single predictor models, the dependence of each platform on specific variables can be evaluated, and by combining T, RHi, and w into a single model, the ML model can leverage the strengths of each predictor. The last two rows (Fig. 19f and 19g) use a Na_{500} subset that only includes when the cloud phase detection has a correlated Na_{500} observation. The inclusion of total aerosol number concentration (Na) is important since aerosols are present in both clear sky and in-cloud conditions, allowing for the examination of differences caused by Na.

The distribution of clear sky and in-cloud conditions is shown in the top row (Fig. 19 row a), with all platforms exhibiting a similar ratio of cloud occurrence. In the single predictor ML setup, the RHi model appears to be the best predictor for in-cloud or clear sky conditions for in-situ observations. For CALIPSO and CloudSat, RHi is better at predicting clear sky conditions, while the T model performs better at predicting in-cloud conditions. DARDAR shows that the T model and w model are similar in predicting clear sky conditions, and the T model and RHi model show similar predictability for in-cloud conditions.

The T+RHi+w model combines the advantages of each predictor. Clear sky predictions from in-situ observations are mainly contributed by the RHi model, while the improvement in

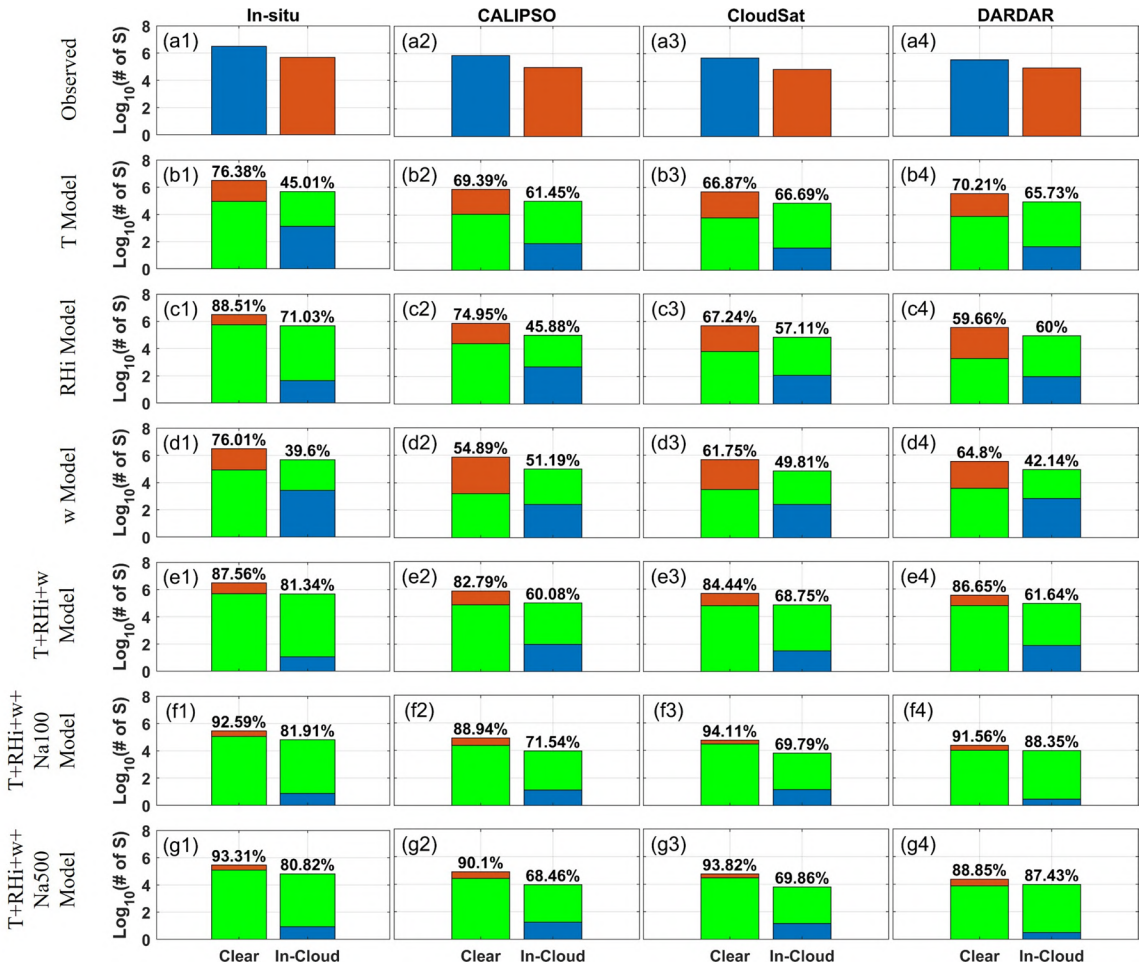

Figure 19. The observed in-cloud and clear sky occurrence in comparison with machine learning predictions. Column 1 shows in-situ, column 2 shows CALIPSO, column 3 shows CloudSat, and column 4 shows DARDAR. Model predictors are shown on the left side of each row. In row a, blue bar and orange bar indicated the clear sky or in-cloud conditions, accordingly. In ML, the bright green indicates correct prediction, the orange indicates the prediction of in-cloud, and blue indicates the prediction of clear sky.

in-cloud predictions is likely due to the combination of T and RHi. All satellite predictions show improvement, with clear sky predictions increasing from around 70% to above 80% using the combined predictors. Similarly, for in-cloud conditions, improvements are observed for both in-situ and all three satellite products compared to the single predictor models.

The inclusion of Na_{500} data in the T+RHi+w+Na100 model leads to further improvements in both clear sky and in-cloud conditions for in-situ and all satellites. DARDAR benefits the most from incorporating Na_{500} data, along with the inclusion of Na_{100} as a predictor. Interestingly, all the in-cloud predictions using the T+RHi+w+Na500 model are worse than those of the T+RHi+w+Na100 model, while clear sky predictions remain the same.

Overall, this analysis demonstrates the varying reliance of different observation platforms on specific predictors and highlights the improvements achieved by combining multiple variables in ML models. The inclusion of aerosol concentration as a predictor shows the potential for enhancing predictions in certain conditions.

2.4.4.2 Evaluating four sky conditions (Clear sky, Liquid, Mixed, and Ice phase)

This evaluation expands on the previous section by analyzing the distribution of cloud phases in addition to the overall appearance of clouds. The statistics of cloud phase distribution are shown in the top row (Fig. 20 row a), with the clear sky being the dominant phase across all platforms, followed by ice phase clouds. Liquid phase clouds are the third most common phase in in-situ observations and DARDAR, while mixed phase clouds detected by CloudSat surpass the frequency of liquid phase clouds.

The predictions from the T model highlight the high-temperature dependence of ice phases in in-situ observations, CALIPSO, and CloudSat, with relatively accurate predictions. However, DARDAR exhibits lower accuracy, achieving only around 50% accuracy in predicting ice phase clouds. Liquid phase clouds are predicted well by CALIPSO, CloudSat, and DARDAR. Clear sky predictions perform poorly across all platforms. The clear sky predictions from in-situ observations heavily rely on the RHi predictor. CALIPSO and

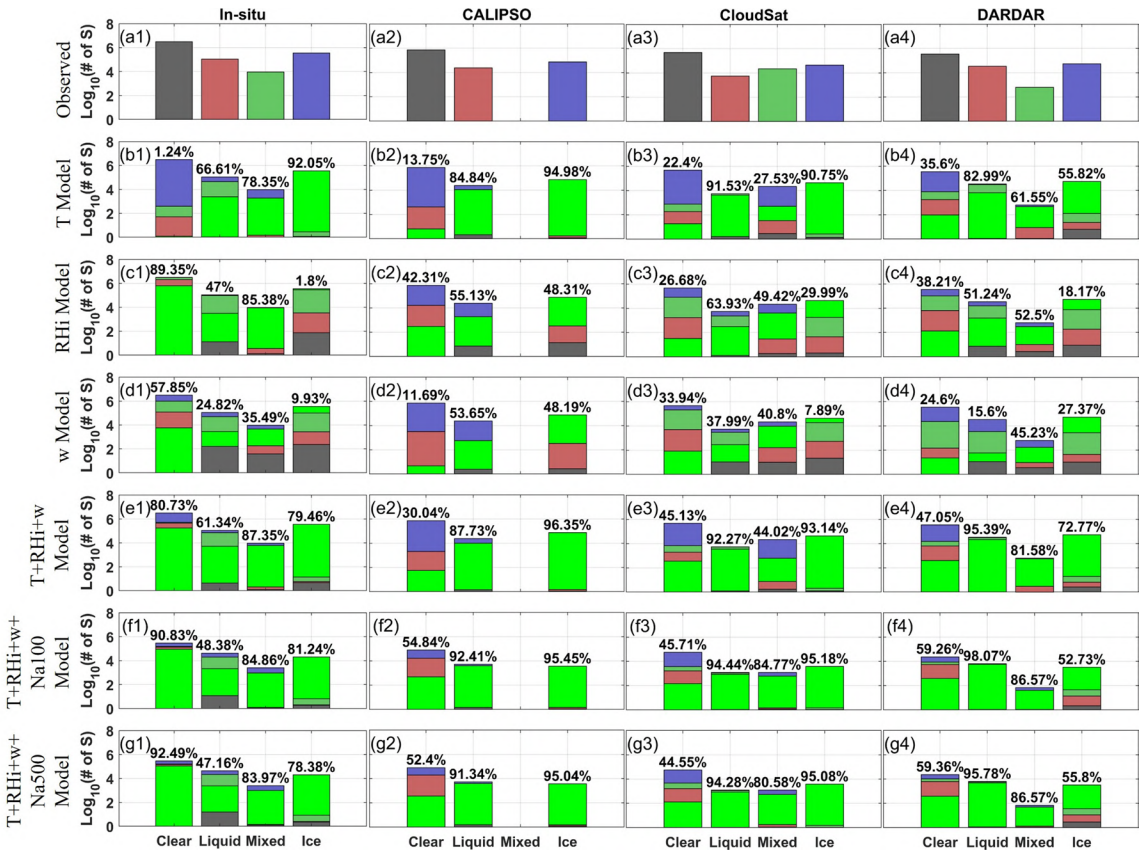

Figure 20. Similar to Fig. 19, but for clear sky, liquid, mixed, and ice phase occurrence in comparison with machine learning predictions.

CloudSat show slightly better clear sky predictions using RHi compared to T and w predictors. The w model demonstrates an inability to accurately identify cloud phases.

The T+RHi+w model shows significant improvements compared to the single predictor models for in-situ observations, CloudSat, and DARDAR across all cloud conditions. However, CALIPSO's clear sky prediction accuracy decreases compared to the RHi model.

When incorporating the Na subset in the ML model for in-situ observations, the accuracy of liquid phase predictions decreases compared to the T+RHi+w model. Similarly, the accuracy of mixed phase and ice phase predictions for DARDAR also decreases compared to the T+RHi+w model. However, in general, there are small improvements across all platforms

compared to the T+RHi+w model. CALIPSO shows the most improvement in clear sky prediction, CloudSat improves in mixed phase prediction, and DARDAR shows improvement in clear sky prediction.

Overall, this evaluation provides insights into the performance of ML models in predicting specific cloud phases. The T+RHi+w model demonstrates improvements in predicting cloud phases compared to single predictor models, but the inclusion of Na as a predictor has mixed effects on prediction accuracy, depending on the platform and cloud condition.

2.4.4.3 In-cloud examination

The cloud phase combination comparisons (Fig. 21) are used to demonstrate the correspondence between cloud phases identified through in-situ observations and satellite data. The all-phase combination provides a distribution of how each in-situ cloud phase aligns with its satellite detection counterpart, with clear sky represented as 0, liquid phase as 1, mixed phase as 2, and ice phase as 3. The combination is structured with the in-situ observation as the first digit and satellite observations as the second digit. Each bar represents a unique combination of cloud phases, and the bright green color indicates the correct predicted amount of the cloud phase combination. Each column represents in-situ, CALIPSO, CloudSat, and DARDAR. In contrast to the previous analysis containing both clear sky and three cloud phases, the in-cloud cloud phase combination only focuses on liquid, mixed, and ice phase combinations by in-situ and satellites.

In the in-situ analysis (Fig. 21 a – e), the best-performing models for predicting the liquid phase are the T model, followed by the RHi model, and finally the w model. When it comes to

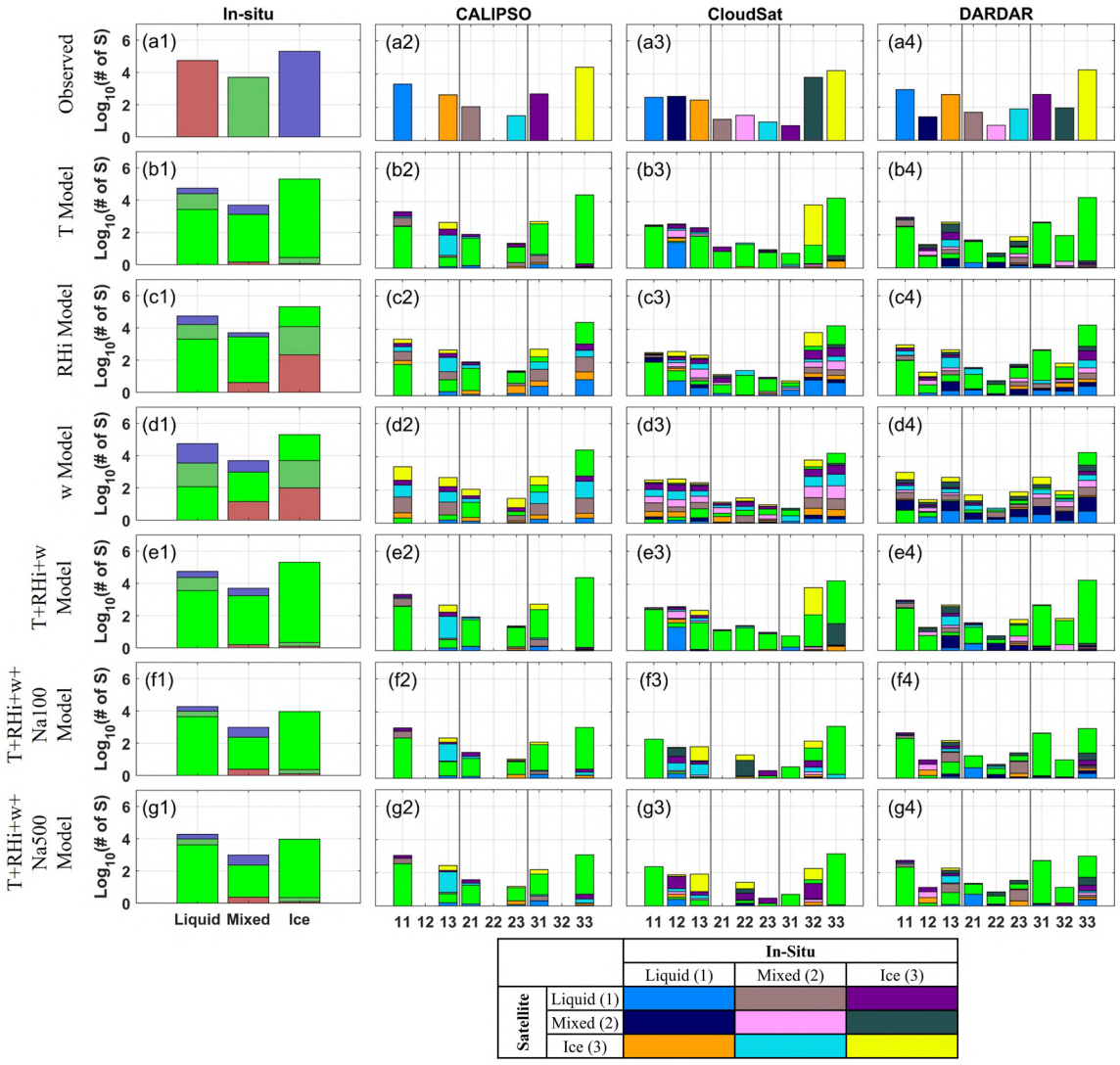

Figure 21. in-cloud cloud phase combination comparison by combining in-situ and satellite observations of cloud phase to demonstrate each model predictors setups variability.

predicting the mixed phase, the RHi model and T model exhibit the best performance, while the w model performs similarly to them. In terms of ice phase prediction, the T model demonstrates exceptionally accurate predictions, while the RHi model and w model yield similar results. Overall, the T model and RHi model outperform the w model in various aspects, with the T+RHi+w model combining the strengths of both the T model and RHi model, where

the T model has the most significant influence. This analysis highlights the strengths and weaknesses of each model and provides valuable insights into their performance for different cloud phases.

The comparison between in-situ and satellite cloud phase combination ML performance reveals a similarity to the in-situ only model prediction. The T model stands out as the best performer for the liquid and ice phases. However, for the CloudSat platform, the "33" combination is often misidentified as "32", and the combination "12" does not exhibit strong predictive capabilities. In the case of DARDAR, the mixed phase section shows that combination "22" is the best predicted, while combinations "21" and "23" have average predictions. The RHi model demonstrates a good representation of the mixed phase combinations, aligning with the in-situ only ML performance. The w model, as in the previous section, exhibits poor and random predictions. The T+RHi+w model's performance is also similar to the previous section, with the CALIPSO "13", CloudSat "12", and DARDAR "13", "21", and "23" combinations being the least accurately predicted among all.

2.5. Discussion and Conclusion

The results of the study provide insights into the thermodynamic and dynamic distribution of ice phase occurrence, in-cloud frequency, and cloud phase identification accuracy. The analysis compared satellite observations from CALIPSO, CloudSat, and DARDAR with in-situ measurements to evaluate their performance.

In terms of in-cloudF, CALIPSO demonstrated accurate cloud occurrence detection, although slight biases were observed at specific temperature and RHi ranges. CloudSat and DARDAR exhibited consistent patterns of cloud occurrence, with CloudSat showing a modest

dependence on temperature and RHi. However, potential issues with cloud phase identification were noted, particularly in over-predicting mixed-phase clouds. Regarding the iceF, CALIPSO showed a strong correlation with temperature and RHi, with no ice phase detection above 0°C and a complete ice phase below -40°C. DARDAR also exhibited temperature dependence but detected ice occurrence even above 0°C. CloudSat demonstrated pronounced reliance on temperature within the range of 0°C to -40°C. However, it did not reach 100% iceF below -40°C, indicating potential issues with its cloud phase identification algorithm and over-prediction of mixed-phase clouds.

The analysis of temperature and vertical velocity distribution revealed that lower temperatures and higher vertical velocities were associated with higher cloud occurrence. The temperature-dependent ice phase frequency highlighted the sensitivity of ice formation to thermal conditions. Unique patterns exhibited by CloudSat, such as lower iceF below -40°C and a decrease in iceF at specific vertical velocities, require further investigation into underlying mechanisms and potential biases in cloud phase identification algorithms. The cloud phase PDF analysis compared the distribution of cloud phases between satellite products and in-situ observations. CALIPSO generally exhibited good agreement with in-situ observations, while CloudSat and DARDAR showed some deviations. CloudSat and DARDAR tended to over-identify mixed-phase clouds, particularly at lower RHi values. The ice phase PDF showed nearly perfect alignment with in-situ observations, but some differences were observed at higher temperatures. Linear regression analysis of ice phase and in-cloud frequency using machine learning models showed that temperature played a significant role in ice phase occurrence, with decreasing iceF at higher temperatures. Machine learning models

captured broader trends in ice phase frequencies but faced challenges in accurately predicting frequencies within specific temperature ranges. In terms of in-cloud occurrence frequency, the incorporation of multiple variables such as temperature, RHi, and vertical velocity improved the models' predictive capabilities.

The evaluation of each variable's importance in cloud conditions revealed varying reliance on specific predictors across different observation platforms. The combination of temperature, RHi, and vertical velocity improved predictions for both clear sky and in-cloud conditions. The inclusion of aerosol concentration as a predictor further enhanced predictions, particularly for DARDAR. Overall, these findings emphasize the importance of evaluating satellite algorithms and incorporating multiple variables in machine learning models for cloud prediction. The accurate identification and understanding of cloud phases are crucial for improving satellite-based cloud observations and advancing our knowledge of cloud processes in the atmosphere.

References

Ahn, E., Huang, Y., Siems, S. T., & Manton, M. J. (2018). A comparison of cloud microphysical properties derived from MODIS and CALIPSO with in situ measurements over the wintertime southern ocean. *Journal of Geophysical Research: Atmospheres*, *123*(19), 11,120–11,140. https://doi.org/10.1029/2018JD028535

Albrecht, B., Ghate, V., Mohrmann, J., Wood, R., Zuidema, P., Bretherton, C., et al. (2019). Cloud system evolution in the trades (CSET): Following the evolution of boundary layer cloud systems with the NSF–NCAR GV. *Bulletin of the American Meteorological Society*, *100*(1), 93–121. https://doi.org/10.1175/BAMS-D-17-0180.1

Barker, H. W., Korolev, A. V., Hudak, D. R., Strapp, J. W., Strawbridge, K. B., & Wolde, M. (2008). A comparison between CloudSat and aircraft data for a multilayer, mixed phase cloud system during the Canadian CloudSat-CALIPSO validation project. *Journal of Geophysical Research: Atmospheres*, *113*(D8). https://doi.org/10.1029/2008JD009971

Barth, M. C., Cantrell, C. A., Brune, W. H., Rutledge, S. A., Crawford, J. H., Huntrieser, H., et al. (2015). The deep convective clouds and chemistry (DC3) field campaign. *Bulletin of the American Meteorological Society*, *96*(8), 1281–1309. https://doi.org/10.1175/BAMS-D-13-00290.1

Bodas-Salcedo, A., Andrews, T., Karmalkar, A. V., & Ringer, M. A. (2016). Cloud liquid water path and radiative feedbacks over the southern ocean. *Geophysical Research Letters*, *43*(20), 10,938-10,946. https://doi.org/10.1002/2016GL070770

Bodas-Salcedo, A., Williams, K. D., Field, P. R., & Lock, A. P. (2012). The Surface Downwelling solar radiation surplus over the southern ocean in the met office model: The role of midlatitude cyclone clouds. *Journal of Climate*, *25*(21), 7467–7486. https://doi.org/10.1175/JCLI-D-11-00702.1

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

Brown, P. R. A., & Francis, P. N. (1995). Improved measurements of the ice water content in cirrus using a total-water probe. *Journal of Atmospheric and Oceanic Technology*, *12*(2). https://doi.org/10.1175/1520-0426(1995)012<0410:imotiw>2.0.co;2

Ceccaldi, M., Delanoë, J., Hogan, R. J., Pounder, N. L., Protat, A., & Pelon, J. (2013). From CloudSat-CALIPSO to EarthCare: Evolution of the DARDAR cloud classification and its comparison to airborne radar-lidar observations. *Journal of Geophysical Research: Atmospheres*, *118*(14), 7962–7981. https://doi.org/10.1002/JGRD.50579

Cesana, G., Chepfer, H., Winker, D., Getzewich, B., Cai, X., Jourdan, O., et al. (2016). Using in situ airborne measurements to evaluate three cloud phase products derived from

CALIPSO. *Journal of Geophysical Research, 121*(10), 5788–5808. https://doi.org/10.1002/2015JD024334

Cesana, G., & Storelvmo, T. (2017). Improving climate projections by understanding how cloud phase affects radiation. *Journal of Geophysical Research, 122*(8). https://doi.org/10.1002/2017JD026927

Coopman, Q., Hoose, C., & Stengel, M. (2020). Analysis of the thermodynamic phase transition of tracked convective clouds based on geostationary satellite observations. *Journal of Geophysical Research: Atmospheres, 125*(11). https://doi.org/10.1029/2019JD032146

D'Alessandro, J. J., Diao, M., Wu, C., Liu, X., Jensen, J. B., & Stephens, B. B. (2019). Cloud phase and relative humidity distributions over the southern ocean in austral summer based on in situ observations and CAM5 simulations. *Journal of Climate, 32*(10), 2781–2805. https://doi.org/10.1175/JCLI-D-18-0232.1

Delanoë, J., & Hogan, R. J. (2008a). DARDAR-CLOUD document. *Université de Lille*.

Delanoë, J., & Hogan, R. J. (2008b). A variational scheme for retrieving ice cloud properties from combined radar, lidar, and infrared radiometer. *Journal of Geophysical Research, 113*(D7), D07204. https://doi.org/10.1029/2007JD009000

Deng, M., Mace, G. G., Wang, Z., & Okamoto, H. (2010). Tropical composition, cloud and climate coupling experiment validation for cirrus cloud profiling retrieval using cloudsat radar and CALIPSO lidar. *Journal of Geophysical Research Atmospheres, 115*(17). https://doi.org/10.1029/2009JD013104

Desai, N., Diao, M., Shi, Y., Liu, X., & Silber, I. (2023). Ship-based observations and climate model simulations of cloud phase over the southern ocean. *Journal of Geophysical Research: Atmospheres, 128*(11), e2023JD038581. https://doi.org/10.1029/2023JD038581

Diao, M., Jensen, J. B., Pan, L. L., Homeyer, C. R., Honomichl, S., Bresch, J. F., & Bansemer, A. (2015). Distributions of ice supersaturation and ice crystals from airborne observations in relation to upper tropospheric dynamical boundaries. *Journal of Geophysical Research, 120*(10). https://doi.org/10.1002/2015JD023139

Diao, M., Jumbam, L., Sheffield, J., Wood, E. F., & Zondlo, M. A. (2013). Validation of AIRS/AMSU-A water vapor and temperature data with in situ aircraft observations from the surface to UT/LS from 87°N-67°S. *Journal of Geophysical Research Atmospheres, 118*(12). https://doi.org/10.1002/jgrd.50483

Flynn, C. M., & Mauritsen, T. (2020). On the climate sensitivity and historical warming evolution in recent coupled model ensembles. *Atmospheric Chemistry and Physics, 20*(13), 7829–7842. https://doi.org/10.5194/ACP-20-7829-2020

Frey, W. R., & Kay, J. E. (2018). The influence of extratropical cloud phase and amount feedbacks on climate sensitivity. *Climate Dynamics*, *50*(7–8), 3097–3116. https://doi.org/10.1007/S00382-017-3796-5/FIGURES/13

Fuchs-Stone, Raymond, D. J., & Sentić, S. (2020). OTREC2019: Convection over the east pacific and southwest caribbean. *Geophysical Research Letters*, *47*(11), e2020GL087564. https://doi.org/10.1029/2020GL087564

Gordon, A. D., Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). Classification and regression trees. *Biometrics*, *40*(3), 874. https://doi.org/10.2307/2530946

Hu, X., Ge, J., Li, W., Du, J., Li, Q., & Mu, Q. (2021). Vertical structure of tropical deep convective systems at different life stages from CloudSat observations. *Journal of Geophysical Research: Atmospheres*, *126*(21). https://doi.org/10.1029/2021JD035115

Hu, Y., Winker, D., Vaughan, M., Lin, B., Omar, A., Trepte, C., et al. (2009). CALIPSO/CALIOP cloud phase discrimination algorithm. *Journal of Atmospheric and Oceanic Technology*, *26*(11). https://doi.org/10.1175/2009JTECHA1280.1

Hunt, W. H., Vaughan, M. A., Powell, K. A., & Weimer, C. (2009). CALIPSO lidar description and performance assessment. *Journal of Atmospheric and Oceanic Technology*, *26*(7). https://doi.org/10.1175/2009JTECHA1223.1

Im, E., Wu, C., & Durden, S. L. (2005). Cloud profiling radar for the cloudSat mission. *IEEE Aerospace and Electronic Systems Magazine*, *20*(10). https://doi.org/10.1109/MAES.2005.1581095

Kay, J. E., Bourdages, L., Miller, N. B., Morrison, A., Yettella, V., Chepfer, H., & Eaton, B. (2016). Evaluating and improving cloud phase in the community atmosphere model version 5 using spaceborne lidar observations. *Journal of Geophysical Research: Atmospheres*, *121*(8), 4162–4176. https://doi.org/10.1002/2015JD024699

Kim, M., Cermak, J., Andersen, H., Fuchs, J., & Stirnberg, R. (2020). A new satellite-based retrieval of low-cloud liquid-water path using machine learning and meteosat SEVIRI data. *Remote Sensing*, *12*(21), 3475. https://doi.org/10.3390/RS12213475

Koop, T., Luo, B., Tsias, A., & Peter, T. (2000). Water activity as the determinant for homogeneous ice nucleation in aqueous solutions. *Nature*, *406*(6796), 611–614. https://doi.org/10.1038/35020537

Kuma, P., A-M Bender, F., Schuddeboom, A., McDonald, A. J., & Seland, Ø. (2023). Machine learning of cloud types in satellite observations and climate models. *Atmospheric Chemistry and Physics*, *23*, 523–549. https://doi.org/10.5194/acp-23-523-2023

Liou, K. N. (1992). Radiation and cloud processes in the atmosphere. *Oxford Monographs on Geology and Geophysics*, 255–339. Retrieved from https://global.oup.com/academic/product/radiation-and-cloud-processes-in-the-atmosphere-9780195049107

Loh, W.-Y., & Shih, Y.-S. (1997). Split selection methods for classification trees. *Statistica Sinica, 7*, 815–840.

Lohmann, U., & Gasparini, B. (2017). A cirrus cloud climate dial? *Science, 357*, 248-249. https://doi.org/10.1126/science.aan3325

Matus, A. V., & L'Ecuyer, T. S. (2017). The role of cloud phase in Earth's radiation budget. *Journal of Geophysical Research, 122*(5), 2559–2578. https://doi.org/10.1002/2016JD025951

McCoy, D. T., Hartmann, D. L., & Grosvenor, D. P. (2014a). Observed southern ocean cloud properties and shortwave reflection. Part I: Calculation of SW flux from observed cloud properties. *Journal of Climate, 27*(23), 8836–8857. https://doi.org/10.1175/JCLI-D-14-00287.1

McCoy, D. T., Hartmann, D. L., & Grosvenor, D. P. (2014b). Observed southern ocean cloud properties and shortwave reflection. Part II: Phase changes and low cloud feedback. *Journal of Climate*, *27*(23), 8858–8868. https://doi.org/10.1175/JCLI-D-14-00288.1

McFarquhar, G. M., Bretherton, C. S., Marchand, R., Protat, A., DeMott, P. J., Alexander, S. P., et al. (2021). Observations of clouds, aerosols, precipitation, and surface radiation over the southern ocean: An overview of CAPRICORN, MARCUS, MICRE, and SOCRATES. *Bulletin of the American Meteorological Society*, *102*(4), E894–E928. https://doi.org/10.1175/BAMS-D-20-0132.1

Mitchell, J. F. B., Senior, C. A., & Ingram, W. J. (1989). CO2 and climate: A missing feedback? *Nature, 341*(6238), 132–134. https://doi.org/10.1038/341132a0

Montgomery, M. T., Davis, C., Dunkerton, T., Wang, Z., Velden, C., Torn, R., et al. (2012). The pre-depression investigation of cloud-systems in the tropics (PREDICT) experiment: Scientific basis, new analysis tools, and some first results. *Bulletin of the American Meteorological Society*, *93*(2), 153–172. https://doi.org/10.1175/BAMS-D-11-00046.1

Murphy, D. M., & Koop, T. (2005). Review of the vapour pressures of ice and supercooled water for atmospheric applications. *Quarterly Journal of the Royal Meteorological Society, 131*(608), 1539–1565. https://doi.org/10.1256/QJ.04.94

Noh, Y. J., Seaman, C. J., Haar, T. H. V., Hudak, D. R., & Rodriguez, P. (2011). Comparisons and analyses of aircraft and satellite observations for wintertime mixed-phase clouds. *Journal of Geophysical Research Atmospheres*, *116*(18). https://doi.org/10.1029/2010JD015420

Pan, L. L., Atlas, E. L., Salawitch, R. J., Honomichl, S. B., Bresch, J. F., Randel, W. J., et al. (2017). The convective transport of active species in the tropics (CONTRAST) experiment. *Bulletin of the American Meteorological Society*, *98*(1), 106–128. https://doi.org/10.1175/BAMS-D-14-00272.1

Pan, L. L., Bowman, K. P., Atlas, E. L., Wofsy, S. C., Zhang, F., Bresch, J. F., et al. (2010). The stratosphere–troposphere analyses of regional transport 2008 experiment. *Bulletin of the American Meteorological Society*, *91*(3), 327–342. https://doi.org/10.1175/2009BAMS2865.1

Rolf, E., Proctor, J., Carleton, T., Bolliger, I., Shankar, V., Ishihara, M., et al. (2021). A generalizable and accessible approach to machine learning with global satellite imagery. *Nature Communications*, *12*(1), 1–11. https://doi.org/10.1038/s41467-021-24638-z

Sassen, K. (1991). The polarization lidar technique for cloud research: A review and current assessment. *Bulletin - American Meteorological Society*, *72*(12). https://doi.org/10.1175/1520-0477(1991)072<1848:TPLTFC>2.0.CO;2

Seiffert, C., Khoshgoftaar, T. M., Van Hulse, J., & Napolitano, A. (2008). RUSBoost: Improving classification performance when training data is skewed. *Proceedings - International Conference on Pattern Recognition.* https://doi.org/10.1109/ICPR.2008.4761297

Stephens, B. B., Long, M. C., Keeling, R. F., Kort, E. A., Sweeney, C., Apel, E. C., et al. (2018). The O2/N2 ratio and CO2 airborne southern ocean study. *Bulletin of the American Meteorological Society*, *99*(2), 381–402. https://doi.org/10.1175/BAMS-D-16-0206.1

Stephens, G. L., Vane, D. G., Boain, R. J., Mace, G. G., Sassen, K., Wang, Z., et al. (2002). The cloudsat mission and the A-Train: A new dimension of space-based observations of clouds and precipitation. *Bulletin of the American Meteorological Society*, *83*(12). https://doi.org/10.1175/bams-83-12-1771

Tan, I., Storelvmo, T., & Zelinka, M. D. (2016). Observational constraints on mixed-phase clouds imply higher climate sensitivity. *Science*, *352*(6282), 224–227. https://doi.org/10.1126/SCIENCE.AAD5300/SUPPL_FILE/AAD5300-TAN-SM.PDF

Terai, C. R., Klein, S. A., & Zelinka, M. D. (2016). Constraining the low-cloud optical depth feedback at middle and high latitudes using satellite observations. *Journal of Geophysical Research: Atmospheres*, *121*(16), 9696–9716. https://doi.org/10.1002/2016JD025233

Thornton, J. A., Brown, S. S., Jaeglè, L., Cohen, R., Jimenez, J.-L., Weber, R., & Roberts, J. M. (2014). Winter investigation of transport, emissions, and reactivity: An NSF funded aircraft and ground-based investigation. *National Oceanic and Atmospheric Administration Chemical Sciences Laboratory.*

Vaughan, M. A., Young, S. A., Winker, D. M., Powell, K. A., Omar, A. H., Liu, Z., et al. (2004). Fully automated analysis of space-based lidar data: An overview of the CALIPSO retrieval algorithms and data products. *Laser Radar Techniques for Atmospheric Sensing*, *5575*. https://doi.org/10.1117/12.572024

Vaughan, M., Pitts, M., Trepte, C., Winker, D., Detweiler, P., Garnier, A., et al. (2019). Cloud-aerosol LIDAR infrared pathfinder satellite observations data management system data products catalog release 4.50 document No: PC-SCI-503. *National Aeronautics and Space Administration*.

Volkamer, R., Baidar, S., Campos, T. L., Coburn, S., DiGangi, J. P., Dix, B., et al. (2015). Aircraft measurements of BrO, IO, glyoxal, NO2, H2O, O2-O2 and aerosol extinction profiles in the tropics: Comparison with aircraft-/ship-based in situ and lidar measurements. *Atmospheric Measurement Techniques*, *8*(5), 2121–2148. https://doi.org/10.5194/AMT-8-2121-2015

Wang, Z., & Sassen, K. (2001). Cloud type and macrophysical property retrieval using multiple remote sensors. *Journal of Applied Meteorology*, *40*(10). https://doi.org/10.1175/1520-0450(2001)040<1665:CTAMPR>2.0.CO;2

Wang, Z. (2019). CloudSat 2B-CLDCLASS-LIDAR product process description and interface control document. *Colorado State University*.

Winker, D. M., Hunt, W. H., & McGill, M. J. (2007). Initial performance assessment of CALIOP. *Geophysical Research Letters*, *34*(19). https://doi.org/10.1029/2007GL030135

Winker, D. M., Pelon, J. R., & McCormick, M. P. (2003). The CALIPSO mission: Spaceborne lidar for observation of aerosols and clouds. In *Lidar Remote Sensing for Industry and Environment Monitoring III*, *4893*. https://doi.org/10.1117/12.466539

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

Yang, C. A., Diao, M., Gettelman, A., Zhang, K., Sun, J., McFarquhar, G., & Wu, W. (2021). Ice and supercooled liquid water distributions over the southern ocean based on in situ observations and climate model simulations. *Journal of Geophysical Research: Atmospheres*, *126*(24). https://doi.org/10.1029/2021JD036045

Yost, C. R., Minnis, P., Ayers, J. K., Spangenberg, D. A., Heymsfield, A. J., Bansemer, A., et al. (2010). Comparison of GOES-retrieved and in situ measurements of deep convective anvil cloud microphysical properties during the tropical composition, cloud and climate coupling experiment (TC 4). *Journal of Geophysical Research*, *115*. https://doi.org/10.1029/2009jd013313

Zelinka, M. D., Myers, T. A., McCoy, D. T., Po-Chedley, S., Caldwell, P. M., Ceppi, P., et al. (2020). Causes of higher climate sensitivity in CMIP6 models. *Geophysical Research Letters*, *47*(1), e2019GL085782. https://doi.org/10.1029/2019GL085782

Zhang, Y., Klein, S. A., Boyle, J., & Mace, G. G. (2010). Evaluation of tropical cloud and precipitation statistics of community atmosphere model version 3 using CloudSat and CALIPSO data. *Journal of Geophysical Research Atmospheres*, *115*(12). https://doi.org/10.1029/2009JD012006