Evaluation of Liquid Cloud Albedo Susceptibility in E3SM Using Coupled Eastern North Atlantic Surface and Satellite Retrievals

Adam C. Varble[1], Po-Lun Ma[1], Matt W. Christensen[1], Johannes Mülmenstädt[1], Shuaiqi Tang[1], and Jerome Fast[1]

5   [1]Atmospheric and Global Change Division, Pacific Northwest National Laboratory, Richland, Washington, 99354, USA

Correspondence to: Adam C. Varble (adam.varble@pnnl.gov)

Abstract. The impact of aerosol number concentration on cloud albedo is a persistent source of spread in global climate predictions due to multi-scale, interactive atmospheric processes that remain difficult to quantify. We use 5 years of geostationary satellite and surface retrievals at the U.S. Department of Energy (DOE) Atmospheric Radiation Measurement (ARM) Eastern North Atlantic

10   (ENA) site in the Azores to evaluate the representation of liquid cloud albedo susceptibility for overcast cloud scenes in DOE Energy Exascale Earth System Model version 1 (E3SMv1) and provide possible reasons for model-observation discrepancies.

The overall distribution of surface 0.2% CCN concentration values is reasonably simulated but simulated liquid water path (LWP) is lower than observed and layer-mean droplet concentration (N_d) comparisons are highly variable depending on the N_d retrieval technique. E3SMv1's cloud albedo is greater than observed for given LWP and N_d values due to a lesser cloud effective

15   radius than observed. However, the simulated albedo response to N_d is suppressed due to a solar zenith angle (SZA)-N_d correlation created by the seasonal cycle that is not observed. Controlling for this effect by examining the cloud optical depth (COD) shows that E3SMv1's COD response to CCN concentration is greater than observed. For surface-based retrievals, this is only true after controlling for cloud adiabaticity because E3SMv1's adiabaticities are much lower than observed. Assuming a constant adiabaticity in surface retrievals as done in TOA retrievals narrows the retrieved $\ln N_d$ distribution, which increases the cloud albedo sensitivity

20   to $\ln N_d$ to match the TOA sensitivity.

The greater sensitivity of COD to CCN is caused by a greater Twomey effect in which the sensitivity of N_d to CCN is greater than observed for TOA-retrieved N_d, and once model-observation adiabaticity differences are removed, this is also true for surface-retrieved N_d. The LWP response to N_d in E3SMv1 is overall negative as observed. Despite reproducing the observed LWP-N_d relationship, observed clouds become much more adiabatic as N_d increases while E3SMv1 clouds do not, associated with more

25   heavily precipitating clouds that are partially but not completely caused by deeper clouds and weaker inversions in E3SMv1. These cloud property differences indicate that the negative LWP-N_d relationship is likely not caused by the same mechanisms in E3SMv1 and observations. The negative simulated LWP response also fails to mute the excessively strong Twomey effect, highlighting potentially important confounding factor effects that likely render the LWP-N_d relationship non-causal. N_d retrieval scales and assumptions, particularly related to cloud adiabaticity, contribute to substantial spreads in the model-observation comparisons,

30   though enough consistency exists to suggest that aerosol activation and convective drizzle processes are critical areas to focus E3SMv1 development for improving the fidelity of aerosol-cloud interactions in E3SM.

1 Introduction

Aerosol effects on liquid clouds are a longstanding leading source of uncertainty in climate projections (Boucher et al. 2013, Bellouin et al. 2020, Smith et al. 2020). As aerosols acting as cloud condensation nuclei (CCN) increase, the number of cloud

35   droplets (N_d) increases while droplet size decreases if holding other cloud properties constant. This increases the cloud albedo and is known as the first indirect, cloud albedo, or Twomey effect (Twomey 1974, 1977, Coakley et al. 1987, Radke et al. 1989).

However, this near-instantaneous aerosol-cloud interaction (ACI) can be buffered or amplified by slower cloud adjustments to the new cloud state including changes in cloud fraction (CF) and/or liquid water path (LWP) (Gryspeerdt et al. 2016, 2019, Christensen et al. 2020, and references therein). Increasing N_d can increase cloud lifetime and thus CF and/or LWP by suppressing precipitation

40   that removes liquid from the cloud (Albrecht 1989, Pincus and Baker 1994), which is known as the second indirect, cloud lifetime, or Albrecht effect. However, it can also potentially decrease CF and/or LWP in non-precipitating clouds when dry air entrainment driven evaporation increases. This can occur via decreased cloud droplet sedimentation with increased cloud top radiative and evaporational cooling in stratocumulus clouds (Ackerman et al. 2004, Bretherton et al. 2007). Increasing N_d may also decrease cloud updraft equilibrium supersaturation, thus increasing condensate (Kogan and Martin 2004, Koren et al. 2014), but smaller

45   droplets may also increase cloud edge evaporation, thus decreasing condensate (Jiang et al. 2006, Xue and Feingold 2006, Small et al. 2009). There can also be several confounding factors affecting correlations between LWP and N_d including meteorological correlations (Zhang et al. 2022). These individual factors influencing effective radiative forcing caused by ACI (ERFaci) are controlled by complex, cross-scale interactions between evolving clouds and their surrounding environment. Better isolating and quantifying these interactions and the processes that control them is required for reducing uncertainty and improving model

50   parameterizations.

The change in cloud albedo for a given change in N_d or aerosols is called the cloud albedo susceptibility (Platnick and Twomey 1994). This metric is commonly decomposed into contributions from the first and second indirect effect where the second indirect effect can be further separated into changes in CF versus changes in LWP (e.g., Quaas et al. 2008, Mülmenstädt et al. 2019, Bellouin et al. 2020). These terms can then be further decomposed to isolate the response of N_d to a change in aerosols, which we

55   refer to as the aerosol activation term while acknowledging that N_d sink processes such as precipitation and evaporation that also influence the N_d vs. CCN relationship. Because cloud base CCN number concentration is rarely directly observed and very difficult to accurately retrieve from satellite measurements (Shinozuka et al. 2015), aerosol optical depth (AOD) or aerosol index (AI) are often used instead with the caveat that they can only be retrieved for clear sky and are thus offset in space from clouds (Stier 2016). These terms have been quantified in many observational and large eddy simulation (LES) studies that have elucidated complex

60   dependencies on cloud properties and atmospheric state that can yield susceptibilities of varying sign and magnitude (Zhang et al. 2022). The spread in estimated global Twomey effects remains substantial due to the need for imperfect aerosol proxies, N_d retrievals, and statistical techniques used for observational quantification (Quaas et al. 2020), imperfect model parameterizations (Gryspeerdt et al. 2017), and an unknown change in aerosols between preindustrial and present day (Carslaw et al. 2013, Ghan et al. 2016). CF and LWP adjustments are even more uncertain given their operation over longer timescales and larger spatial scales.

65   The LWP response to N_d in liquid clouds has been shown in observations and LES to be positive or negative depending on precipitation rate and evaporation via sedimentation-radiation-entrainment feedbacks (e.g., Lebsock et al. 2008, Chen et al. 2014, Glassmeier et al. 2019, Hoffman et al. 2020). The most recent net global $\frac{dlnLWP}{dlnN_d}$ estimates are negative (-0.3 to 0) with a positive value in precipitating clouds with $N_d < \sim 30$ cm^{-3} and a value as low as -0.4 in non-precipitating clouds with higher N_d values (Bellouin et al. 2019). However, estimates vary even more substantially and go strongly positive on smaller scales in which strong

70   aerosol perturbations exist (Christensen et al. 2022). These values remain highly uncertain due to observational uncertainty and substantial modulation by atmospheric conditions such as inversion strength, relative humidity, and boundary layer depth (e.g., Gryspeerdt et al. 2019, Possner et al. 2020).

Present-day aerosol and cloud statistics are commonly used to quantify cloud albedo susceptibility or ERFaci, which are then used to evaluate ACI in climate models. Several previous studies have compared cloud albedo susceptibility, ERFaci, or similar

75   metrics in climate models with observations, often finding that models have a greater susceptibility than observed (Boucher et al. 2013, Ghan et al. 2016). More weight has been traditionally given to observational estimates, but recent studies have shown that

they can have significant biases, particularly depending on the aerosol variable used, with AOD being problematic (Penner et al. 2011), particularly near cloud edges (Christensen et al. 2017). Gryspeerdt et al. (2020) show that this can be largely attributed to the forcing being computed differently in observational and model datasets, and that climate models and observations have surprisingly good agreement if consistent methods are applied to each, though still with significant spread and disagreement on the sign and magnitude of cloud adjustments. In contrast to recent observational and LES studies showing a net decrease in global LWP in response to a N_d increase, climate models most commonly produce a net LWP increase (Quaas et al. 2009, Gryspeerdt et al. 2020). This may be due to model representation of precipitation suppression with increasing N_d but not buffering effects (e.g., Stevens and Feingold 2009) associated with condensate sedimentation, radiative cooling, entrainment, and evaporation (Ghan et al. 2016, Michibata et al. 2016, Toll et al. 2017). Indeed, there are suggestions that global storm-resolving models may better represent LWP responses to N_d perturbations (Sato et al. 2018, Terai et al. 2020).

A difficulty in evaluating process parameterizations with observations is that such processes cannot be directly observed and need to be inferred from properties. A commonly used approach for analyzing processes given only statistics of select properties is plotting joint distributions and heatmaps of variables known to control and/or respond to processes (e.g., Suzuki et al. 2010, 2013, 2015, Gryspeerdt et al. 2016, 2017, 2019, Jing and Suzuki 2018). Others have examined cloud susceptibilities to aerosols in the context of cloud and atmospheric state properties that modulate their magnitude (e.g., Douglas and L'Ecuyer 2019, 2020, Zhang et al. 2022). These approaches coupled with linear regressions to quantify relationships following numerous past studies are used in this study to assess how well the U.S. Department of Energy's Energy Exascale Earth System Model version 1 (E3SMv1; Golaz et al. 2019, Rasch et al. 2019) with ~1° grid spacing reproduces observationally estimated cloud albedo susceptibility controls across temporal (5- to 60-min) and spatial (4-km to 1°) scales and to identify possible reasons for discrepancies with observations.

Observational and modeling datasets as well as the study methodology are described in section 2. Comparisons of observed and simulated Twomey, LWP susceptibility, and aerosol activation contributions to cloud albedo susceptibility over the Eastern North Atlantic are discussed in sections 3-4 along with possible reasons for E3SM-observation differences. Finally, conclusions are presented in section 5.

2 Datasets and Methodology

2.1 Approach

Several methodological decisions are made to improve the interpretability of model-observation comparisons in this study. Both surface and satellite observational retrievals of cloud properties are used in comparisons with model output because of uncertainties associated with each, for N_d in particular (Grosvenor et al. 2018). Although in situ cloud droplet measurements are more accurate than remote sensing retrievals, they are much rarer, which results in sampling biases. Surface and satellite retrievals are also applied to model output to yield multiple model datasets. Observational retrievals are analyzed at their highest resolution as well as coarse resolutions consistent with model output. This approach yields 4 observational and 3 E3SMv1 datasets where the output spread among them can be viewed as a rough measure of uncertainty stemming from retrieval assumptions and resolution. Lastly, analyses are confined to situations with single layer liquid clouds and greater than 95% cloud fraction to improve the accuracy of remote sensing retrievals (e.g., see Grosvenor et al. 2018).

Long-term surface retrievals require a fixed site that experiences frequent marine liquid clouds with aerosol, cloud, and atmospheric state measurements, and have several advantages over satellite retrievals. CCN observations eliminate the need to use column-integrated optical properties such as AOD or AI, and they can be co-located with clouds rather than offset in space. In addition, no adiabaticity assumption is required for retrieving N_d, adiabaticity can be estimated, and variables such as cloud depth

115 and rain rate can be well quantified. This approach lends itself to Eulerian, climatological statistics as opposed to Lagrangian trajectory analyses (e.g., Pincus et al. 1997, Johnson et al. 2000, Eastman et al. 2016, Goren et al. 2019, Mohrmann et al. 2019, Christensen et al. 2020, 2023) or comparisons of situational aerosol plume effects on clouds from ship tracks, volcanoes, or other significant emission sources (see Christensen et al. (2022) and references therein), though these strategies provide valuable, complementary perspectives.

120 **2.2 Observations**

Surface datasets from 2016-2020 are collected from the Atmospheric Radiation Measurement (ARM) Eastern North Atlantic (ENA) site located on Graciosa Island in the Azores at about 39°N and 28°W (Mather and Voyles 2013). The site experiences a range of marine, mid-latitude meteorological and cloud conditions (Rémillard and Tselioudis 2015, Mechem et al. 2018) with frequent drizzling stratocumulus (Rémillard et al. 2012, Giangrande et al. 2019, Jensen et al. 2021) that make it ideal for this study of aerosol

125 effects on single layer liquid clouds (Wood et al. 2015). Surface CCN concentration at 0.2% supersaturation is estimated by a CCN counter (ARM user facility 2016a) that varies supersaturation setpoints between 0 and 1% over the course of an hour from which a polynomial is fit to the data to provide the CCN spectra as a function of supersaturation (ARM user facility 2016b). Although aerosol measurements at ENA are occasionally impacted by emissions from the airport where the site is located (Gallo et al. 2020), this had a minimal impact on hourly 0.2% CCN statistics and thus we do not control for it. However, there are missing CCN

130 measurements before 23 June 2016, between 31 July and 4 December 2018, after 28 October 2020. Thus, analyses involving CCN cover a shorter period than the full 5 years. Cloud LWP is retrieved from a 3-channel microwave radiometer (Turner et al. 2007, ARM user facility 2014c) with an estimated uncertainty of ~20 g m^{-2} for low LWPs increasing to ~10% for large LWPs (Caddedu et al. 2013). A multifrequency shadowband radiometer is used to retrieve cloud optical depth (COD) and layer-mean effective radius (R_{eff}) (Min and Harrison 1996, ARM user facility 2014a). Cloud base and top heights are retrieved from the Active Remote

135 Sensing of Clouds product (Clothiaux et al. 2000, ARM user facility 2015) that combines vertically pointing ceilometer and Ka-band radar measurements. Cloud base and top temperatures are then estimated by matching temperature profiles retrieved from interpolated radiosonde measurements scaled by microwave radiometer precipitable water retrievals (ARM user facility 2013a) to cloud base and top heights. Surface rain rates are retrieved using a Parsivel-2 disdrometer (ARM user facility 2014b) and optical rain gauge (ARM user facility 2013d). All-sky and clear sky downwelling broadband shortwave irradiances are obtained from a

140 pyranometer via the RADFLUX product (Long and Ackerman 2000, ARM user facility 2013c), from why cloud effective albedo is computed. These variables are then either averaged or interpolated to 5-min and 60-min intervals depending on whether the temporal frequency of the variable is less than or greater than that interval. N_d is then derived using Equation 1 following McComiskey et al. (2009):

$$N_d = \frac{2^{-\frac{5}{2}}}{kH}\left(\frac{4\pi\rho_{liq}}{3LWP}\right)^2\left(\frac{5COD}{3\pi Q}\right)^3 ,$$ (1)

145 where H is the cloud depth, k is the ratio of the drop volume mean radius to R_{eff}, Q is the droplet scattering efficiency, and ρ_{liq} is the liquid water density. LWP, COD, and cloud depth inputs from surface retrievals are averaged to the 5- or 60-min intervals. While it is more physically realistic to compute N_d at the highest resolution possible and average it to coarser scales, that would be inconsistent with the E3SM-simulated surface retrievals that use inputs from a 1° grid. We assume $Q = 2$ following Bennartz (2007) and $k = 0.74$ following Brenguier et al. (2011). k commonly varies between 0.5 and 0.9 depending on the drop size distribution.

150 Adiabatic LWP is estimated from retrieved cloud base and top temperature and pressure, from which cloud adiabaticity is computed. Lastly, CF is estimated from downwelling shortwave fluxes (Long et al. 2006, ARM user facility 2013c).

Hourly satellite-based cloud retrievals are obtained from the NASA Visible Infrared Solar-infrared Split-Window Technique (VISST) dataset (Minnis 2008, 2011; ARM user facility 2013b, 2014c, 2018a-b). These retrievals use MeteoSat 10 and 11 geostationary satellite measurements to estimate top-of-atmosphere (TOA) radiative fluxes including all-sky albedo, cloud LWP and IWP, COD, and cloud top R_{eff}, height, and temperature for 4-km wide pixels and 0.5-deg regions, with the 0.5-deg retrievals including CF. To match the E3SMv1 1° grid, the 0.5° retrievals are averaged to 1° grids with each 0.5° value weighted by CF for CF-dependent variables to yield in-cloud rather than all-sky values. From these, cloud layer mean N_d is retrieved using Equation 2 following Bennartz (2007):

$$N_d = \frac{\sqrt{\alpha \Gamma_{ad}}}{k} \left(\frac{4\pi \rho_{liq}}{3}\right)^2 \left(\frac{5COD}{3\pi Q}\right)^3 (2LWP)^{-\frac{5}{2}},$$ (2)

where α is cloud adiabaticity and Γ_{ad} is the adiabatic liquid water content lapse rate. Equation 2 assumes adiabatically stratified clouds and uses inputs of COD, LWP, and cloud top temperature (for calculation of Γ_{ad}) that are obtained from the VISST retrievals. Bennartz (2007) showed that the uncertainty in this retrieval is less than 80% for LWP exceeding 30 g m^{-2} and CFs exceeding 0.8. Cloud adiabaticity is highly variable (e.g., Merk et al. 2016), which along with the assumed k value of 0.74, cloud inhomogeneity, LWP uncertainty, and satellite viewing angle, contributes to N_d retrieval uncertainty. Grosvenor et al. (2018) reported N_d retrieval uncertainties of 54-78% depending on the averaging area for ideal conditions, while Gryspeerdt et al. (2022) reports lesser uncertainties of 30-50% for overcast stratocumulus with parameter thresholds to remove likely biased samples, most of which we employ in this study (see Section 2.4). Surface-measured CCN is interpolated to satellite retrieval times for analyses relating satellite-retrieved cloud properties with CCN concentration. An example of a closed cell stratocumulus case is shown to highlight some of the key retrieved variables from the surface and satellite data (Fig. 1).

2.3 Model Output

The E3SMv1 (Golaz et al. 2019) Atmosphere Model (Rasch et al. 2019) is run with ne30np4 horizontal resolution (approximately 1° grid spacing) for 2016-2020 with hourly output. Large-scale meteorological conditions are constrained via nudging the horizontal winds toward the Modern-Era Retrospective Analysis for Research and Applications version 2 (MERRA-2) (Gelaro et al. 2017) with a relaxation timescale of 6 hours. Several of the variables used in analyses are directly outputted including grid-scale LWP, CF by height, surface and TOA radiative fluxes for computing cloud-effective and TOA albedo, temperature and moisture by height, surface rain rate, and surface CCN concentration at 0.2% supersaturation.

Other variables require derivation. These include cloud base height, which is derived by summing height levels at which CF is greater than 0 weighted by that level's CF contribution minus the integrated CF below that level to the total 2D low level CF moving upward from the surface until the 2D low level cloud fraction is reached. This assumes maximum cloud overlap. Mathematically, this is written as:

$$Cloud\ Base\ Height = \sum_{k=1}^{nlev} \left\{ Z_k max \left[\frac{CF_k - max(CF_1,...,CF_{k-1})}{max(CF_1,..., CF_{nlev})}, 0 \right] \right\},$$ (3)

where k is the height level with 1 being the lowest level and $nlev$ being the highest level, Z_k is the height at level k, and CF_k is the cloud fraction at level k. As an example, if the 2D low level cloud fraction is 80%, and the lowest level clouds are located at 1000 m with a CF of 60%, and then the next height level at 1200 m has a CF of 80% that is equal to the low cloud fraction, the cloud base level would be (60/80)*1000 m + ((80 − 60)/80)*1200 m = 1050 m. However, a complication arises due to limited model vertical resolution. The existence of cloud at a point means that the cloud layer thickness is equal to the distance between the model half levels bounding the point, but if the cloud was fully resolved, it could be anything greater than 0 up to this value. To account for this effect in comparisons to observations, cloud bases are computed separately using both the half levels directly below the

cloud mass levels and using the cloud mass levels, and then these values are averaged as a best estimate. The same method is
applied for estimating cloud top heights but integrating normalized CF-weighted height levels downward from above cloud top
until the 2D low level cloud fraction is consumed. Cloud base and top temperatures are computed in the same way. The cloud
layer-mean N_d is computed by averaging N_d at each level in the cloud weighted by the CF at that level. Note that this is the grid-
scale "stratiform" N_d whereas LWP and CF include convective contributions to be consistent with observations.

The satellite-based N_d retrieval leverages simulated MODerate resolution Imaging Spectroradiometer (MODIS) retrievals
(Pincus et al. 2012) using the Cloud Feedback Model Intercomparison Project Observation Simulator Package (COSP; Bodas-
Salcedo et al. 2011) version 2 (Swales et al. 2018). Simulated MODIS-retrieved LWP and COD with cloud top temperature are
used to derive N_d following the same method described in section 2.1 for VISST satellite observations. Cloud top R_{eff} is also
retrieved from COSP. The simulated surface-based N_d retrieval uses vertically integrated cloud plus rain water content (since
surface-based observations are also impacted by rain) with COSP-simulated COD and cloud depth following the method described
in section 2.1 for ARM surface-based measurements. Cloud adiabaticity is also computed following the same method used for
observations.

2.4 Comparison Methods

All datasets and variables used in comparisons are listed in Table 1 including 5- and 60-minute averaged surface observations (Obs
Sfc 5min, Obs Sfc 60min), as well as 4-km and 1° satellite observations (Obs TOA 4km, Obs TOA 1deg). Three E3SMv1 datasets
are used. One uses E3SMv1 direct output (E3SM), while the others use surface-estimated N_d (E3SM Sfc) and TOA-estimated N_d
(E3SM TOA). LWP estimates also slightly differ between these 3 E3SMv1 datasets. To be consistent with surface measurements,
E3SM surface retrievals use total LWP inclusive of rain and convective liquid from the Zhang-McFarlane (Z-M) scheme (Zhang
and McFarlane 1995) that accounts for 15% of the total LWP in single layer liquid cloud situations at ENA. The direct output
E3SM dataset uses summed column-integrated grid scale and Z-M cloud water (not including rain water path), and TOA retrievals
use COSP simulated LWP. Times with drizzle can influence the accuracy of retrievals (Cadeddu et al. 2020) but are not removed
except for surface-based observations when drizzle at the surface is sufficient to obscure MWR LWP retrievals. For some variables
in Table 1 such as albedo, CCN concentration, COD, and R_{eff}, the value from a single dataset is used for the others that do not have
uniquely derived values. For these variables, only differing sampling constraints produce slight differences in values between
datasets. Since only overcast conditions are analyzed, all-sky albedo varies as cloud albedo varies, and thus we use all-sky values
in analyses to avoid estimating clear sky contributions.

Comparisons between E3SMv1 and observational datasets are confined to specific situations to limit sources of retrieval
uncertainty and possible contributors to dataset differences. All comparisons are limited to the column over the ENA site. In
E3SMv1 datasets, single layer liquid cloud situations are isolated by removing times with COSP-simulated MODIS IWP > 0 or
cloud top temperature < 0°C. E3SMv1-predicted IWP is not used since it is commonly slightly greater than 0. This same IWP and
cloud top temperature filtering is applied to VISST datasets in which it is also possible for multiple liquid cloud layers to exist due
to a lack of sufficient TOA data to remove such situations, though surface vertically pointing radar measurements indicate that
such situations are not common for the overcast, liquid cloud situations considered. For surface-based observations, only single
layer cloud situations are considered, and situations with cloud top temperature < 0°C are removed using the radar-detected cloud
boundaries described in section 2.2. Only situations with CF > 95%, solar zenith angle (SZA) < 65°, LWP > 20 g m^{-2}, and COD >
4 are included following some recommendations in Grosvenor et al. (2018) and Gryspeerdt et al. (2022). E3SMv1 and TOA
measurements of these variables are spatial averages, whereas surface measurements of these variables are averaged over 5- and
60-minute periods in which cloud fraction is estimated using a ceilometer. For analyses relating cloud properties to surface CCN

concentrations, only times with cloud base potential temperature < 2°C warmer than the near surface potential temperature are included, situations that are most likely to have surface-coupled cloud bases. This will not remove all uncoupled clouds, but it

230 allows retaining of greater sample sizes. Other cloud-surface coupling indices produced similar results as the potential temperature metric (Fig. S1).

3 Cloud, Aerosol, and Radiation Properties

The sample size of overcast single layer liquid clouds at ENA depends on the resolution and sensitivity of each dataset. The average warm, liquid CF and percentage of times with CF > 95% for times without supercooled and ice clouds are greater for the 5-minute

235 surface and 4-km TOA retrievals than the 60-minute surface and 1° TOA retrievals (Table 2). This is likely the result of increasing probability of encountering overlying cloud layers as scale increases. The average warm, liquid CF in E3SMv1 for times without overlying clouds is slightly lower than observed. However, the percentage of time that CF exceeds 95% is similar for satellite-observed and COSP-simulated TOA estimates (20 vs. 23%) with the 60-min surface estimate a bit lower (17%) and direct model output having a greater occurrence (31%). SZA, LWP, and COD constraints further reduce observational hourly sampling to

240 between 649 and 1,381 with between 1,697 and 1,941 hourly E3SM samples depending on the dataset. Analyses involving surface CCN concentration have even fewer samples. This is partly because of the need for clouds to be more likely coupled with the surface, though observed samples drop further than for E3SM due to missing or bad CCN data, resulting in 197-328 hourly observational samples but 1,303-1,459 E3SM samples.

For times with overcast, warm liquid clouds and sufficient SZA, distributions of TOA albedo, cloud effective albedo at the

245 surface, and cloud optical depth are shown in Figure 2. E3SMv1 has slightly greater median TOA albedos than observed (Fig. 2a and 3), though median surface-estimated cloud effective albedos are slightly lower (Fig. 2b and 3). This TOA difference will be shown later to be the result of differing SZA-N_d correlations between E3SMv1 and observations. Indeed, median COD is similar between E3SMv1 and TOA observations (Fig. 3). Surface-estimated COD values are greater than simulated, consistent with greater cloud effective albedos caused by greater LWPs being sampled with similar N_d values (Fig. 3 and 4a-b). Median LWP is lesser in

250 E3SMv1 as compared to observations (Fig. 3 and 4a) by ~30% (62-68 vs. 78-92 g m^{-2}). Unlike LWP, N_d is notably greater in E3SMv1 TOA than TOA observations (95 vs. 56-61 cm^{-3} medians; Fig. 3and 4b) and droplet R_{eff} is notably smaller (9.6 vs. 13 μm median values; Fig. 3 and 4c) though E3SMv1 direct outputted N_d is lesser with a median value of 70 cm^{-3}. While surface-retrieved R_{eff} observations are also greater than simulated, 60-min surface-retrieved N_d observations are similar to E3SMv1 surface-retrieved N_d, both having median values near 110 cm^{-3} (Fig. 3 and 4b-c), and 0.2% surface CCN concentration distributions are also similar

255 (Fig. 3 and 4d). Model-observation N_d comparisons clearly depend tremendously on how N_d is retrieved with typical values changing by up to a factor of 2 based on retrieval method. This emphasizes the importance of using multiple different retrievals to assess and better interpret model-observation differences.

4 Cloud Albedo Susceptibility

Differences in observed and simulated cloud properties can result from differences in atmospheric states and/or errors from subgrid

260 scale parameterizations. Such errors do not necessarily imply that the responses of clouds and radiation to aerosol perturbations are incorrect, and it is these responses that are the primary focus here. In particular, the response of cloud albedo to CCN is evaluated in this section. The cloud albedo (*A*) susceptibility to changes in CCN number concentration is evaluated by separating the Twomey effect and LWP response components with Equation 4 (Quaas et al 2008):

$$\frac{dA}{dlnCCN} = \left(\frac{\partial A}{\partial lnN_d} + \frac{\partial A}{\partial lnLWP}\frac{dlnLWP}{dlnN_d}\right)\frac{dlnN_d}{dlnCCN},$$ (4)

265   Given the overcast cloud condition requirement, the CF response is neglected, and we analyze all sky albedo for TOA relationships. The following sections quantify the individual terms of Equation 4 within each of the E3SMv1 and observation datasets within the context of retrieval uncertainties. Analyses are also performed to assess possible causes for model-observation discrepancies.

4.1 Twomey Effect Comparison

We first evaluate the Twomey effect signified by $\frac{\partial A}{\partial lnN_d}\frac{dlnN_d}{dlnCCN}$, which describes the response of albedo to a change in N_d via a
270   change in CCN.

4.1.1 Albedo Dependence on Droplet Concentration

Isolating the Twomey effect requires accounting for the effect of changes in LWP on A. The effect of both LWP and N_d on A is visualized in Figures 5a-d and 6a-d, which show heatmaps of median A within bins of N_d and LWP for the various datasets analyzed. All datasets show similar patterns with A increasing foremost with increasing LWP and secondarily with increasing N_d, though the
275   A sensitivity to N_d is muted in E3SMv1, particularly for TOA estimates. Indeed, absolute differences between E3SMv1 and observations in Figure 5e-f and 6e-f show that E3SMv1 A is much greater than observed for relatively low N_d values, a difference that decreases as N_d increases. $\frac{\partial A}{\partial lnN_d}$ is further quantified by multiple linear regression. Regression coefficients confirm that $\frac{\partial A}{\partial lnLWP}$ is greater than $\frac{\partial A}{\partial lnN_d}$ by a factor of 2-8 depending on the dataset considered (Fig. 7). All dataset fits have Pearson correlation coefficients (r) of 0.81-0.88 (Fig. 7), showing that LWP and N_d alone predict much of the A variability. Consistent with Figures 5-
280   6, E3SMv1 coefficients have a similar response of A to LWP as observations but the response of A to N_d is about half that observed. The estimated response of A to N_d also depends on how N_d is retrieved. Whereas a surface-retrieved N_d value decreases $\frac{\partial A}{\partial lnN_d}$ relative to the actual E3SMv1-predicted N_d value (green vs. blue open diamonds), a TOA-retrieved N_d does the opposite (orange vs. blue open circles) (Fig. 7). This difference is explained by differences in cloud adiabaticity assumptions made by TOA and surface retrievals. TOA retrievals assume an adiabaticity of 80% while surface retrievals allow adiabaticity to vary by leveraging
285   cloud depth measurements. If cloud depths are scaled to values that produce adiabaticity of 80% with N_d recomputed from these scaled cloud depths, then the sensitivity of albedo to N_d increases (light green relative to dark green symbols in Fig. 7). The increase is especially dramatic for E3SMv1, where surface-retrieved $\frac{\partial A}{\partial lnN_d}$ assuming 80% adiabaticity now exceeds the value derived from direct model output in agreement with the higher TOA retrieval value relative to direct model output. Clearly cloud adiabaticity can affect $\frac{\partial A}{\partial lnN_d}$, a topic discussed further in the next section. While the resolution of the N_d retrieval input data does not affect
290   TOA $\frac{\partial A}{\partial lnN_d}$, the surface $\frac{\partial A}{\partial lnN_d}$ decreases by 15% when switching from 5-min to 60-min inputs, a difference that is less than the model-observation differences (Fig. 7). Thus, there is agreement between datasets that E3SMv1 significantly underestimates the N_d effect on A despite reasonable sensitivity of A to LWP.

4.1.2 Factors Affecting Model-Observation Differences

We first investigate how SZA impacts differences in observed and simulated $\frac{\partial A}{\partial lnN_d}$. SZA and N_d are not correlated in observations,
295   but E3SMv1 SZA decreases as N_d increases (Fig. S2). This difference between E3SMv1 and observations is evident in Figure 8 where E3SMv1 has greater SZA values for relatively low N_d and lesser SZA values for relatively high N_d. We first check whether

this correlation is caused by the diurnal cycle. There is an early afternoon N_d minimum in E3SMv1 datasets (dashed lines in Fig. S3e) that correlates with a SZA minimum (Fig. S3b) whereas observed variables lack such a correlation. However, this is the opposite SZA-N_d correlation in Figures S2 and 8, so this cannot be the cause of the SZA-N_d correlation in those figures. Surface

300 CCN concentration reaches a minimum in early afternoon (Fig. S4a), which is less apparent in observations than E3SMv1 and could be the cause for the simulated early afternoon N_d minimum. We next investigate the seasonal cycle. Simulated N_d strongly peaks in July with median values that are more than twice the wintertime minimum (Fig. 9e), a seasonal cycle that is anti-correlated with the SZA and A seasonal cycles (Fig. 9b-c). R_{eff} also exhibits a notable seasonal cycle, but LWP does not. Thus, seasonality is likely the cause for the E3SMv1 SZA-N_d correlation. Observations also exhibit N_d seasonality consistent with past studies (e.g.,

305 Wood et al. 2015) but with a peak in May that does not correlate with the SZA seasonal cycle. Both E3SMv1 and observations also exhibit strong seasonal cycles in surface CCN concentration (Fig. S4b), and despite being constrained to specific cloud situations, the wintertime minimum and summer maximum in CCN concentration are consistent with the results of Zheng et al. (2018). The observed maximum is 2 months earlier than simulated, possibly contributing to the ~2-month earlier peak in observed maximum N_d relative to E3SMv1 that leads to the differing relationship of N_d with SZA between E3SMv1 and observations.

310 To remove the impact of SZA, we analyze the COD response to N_d and LWP. Observed and simulated $\frac{\partial lnCOD}{\partial lnLWP}$ values are similar, which is consistent with $\frac{\partial A}{\partial lnLWP}$ comparisons (Fig. 10). Whereas observed and simulated $\frac{\partial A}{\partial lnN_d}$ values significantly differed, observed and simulated $\frac{\partial COD}{\partial lnN_d}$ values are much more similar (Fig. 10). The difference that remains is E3SM and E3SM Sfc values being less than surface-based observations. However, controlling for cloud adiabaticity differences between E3SM and observations removes this difference (light green symbols in Fig. 10). As described in section 4.1.1, this is done by recomputing

315 N_d in surface retrievals with scaled cloud depths to values that would be associated with 80% adiabatic LWP. This increases $\frac{\partial lnCOD}{\partial lnLWP}$ by over 10% for observations and almost 50% for E3SMv1, leading to values that are almost identical to TOA retrieved values that assume 80% adiabaticity. To visualize this, Figure 11 shows absolute differences of lnCOD between E3SM and observational datasets as a function of LWP and N_d (lnCOD values for each dataset are shown in Fig. S5). Note that when the exact same retrieval with a constant adiabaticity is applied to both E3SM and observations (Fig. 11a, c), there is virtually no difference in lnCOD for a

320 given LWP and N_d because N_d is computed from COD, LWP, and adiabaticity or cloud depth. When adiabaticity is not held constant, lnCOD values for surface-based retrievals in E3SMv1 and observations diverge (Fig. 11b). Why is this? As shown in Figure 12, cloud adiabaticity is frequently lower than 80% and decreases as N_d decreases and LWP increases. E3SMv1 also has much more subadiabatic clouds than observed. When N_d is recomputed using a constant adiabaticity, it causes a shift in the N_d distribution. For the case of constant 80% adiabaticity, low N_d values with adiabaticity much less than 80% increase more than higher N_d values that have higher adiabaticity values. This causes a narrowing of the lnN_d distribution, which increases $\frac{\partial lnCOD}{\partial lnN_d}$ (Fig.

325 13). Because adiabaticity also decreases with increasing LWP, the shift in the lnN_d distribution varies by LWP (Fig. 13), causing a slight decrease in $\frac{\partial lnCOD}{\partial lnLWP}$ (Fig. 10) that can also affect $\frac{\partial lnCOD}{\partial lnN_d}$. Thus, the grossly subadiabatic clouds in E3SMv1 suppress the change of COD with N_d relative to more adiabatic observed clouds and retrievals with constant 80% adiabaticity.

Though simulated and observed values of $\frac{\partial lnCOD}{\partial lnN_d}$ agree once cloud adiabaticity effects on N_d retrievals are removed, lnCOD

330 is always greater than observed for the E3SM direct model output (Fig. 11d-e). This is potentially the result of lesser R_{eff} values in E3SMv1 than observed for given LWP and N_d values (Fig. 14c-d; R_{eff} values for each dataset are shown in Fig. S6). While R_{eff} is also systematically lower for given LWP and N_d values in TOA retrievals (Fig. 14a), this does not lead to COD differences as a function of LWP and N_d in the TOA datasets because the exact same N_d calculation is used for Obs TOA and E3SM TOA with

sole dependence on LWP, COD, and the constant k parameter that relates droplet volume mean radius to R_{eff}. The same applies for surface retrievals (Fig. 14b), but non-systematic differences are created by differing adiabaticity values and surface retrieved R_{eff} that is sensitive to the entire cloud layer whereas E3SM Sfc R_{eff} reflects near cloud top values obtained from COSP. R_{eff} is dependent on the parameterized cloud droplet size distribution in the Morrison-Gettelman (MG2) microphysics scheme (Morrison and Gettelman 2008, Gettelman and Morrison 2015, Gettelman et al. 2015), and that it is lower in direct model output than retrievals means that the parameterized size distribution differs from that assumed in observations. Although it is possible that remote sensing retrieved R_{eff} is high biased, recent studies show limited satellite retrieval biases when compared with in situ measurements (Witte et al. 2018, Kang et al. 2021). In addition, aircraft in situ measurements from the ACE-ENA campaign (Wang et al. 2022) support remotely sensed R_{eff} values being greater in observations than E3SMv1 with corresponding lower N_d values (Wu et al. 2020). These results provide further support that the simulated SZA-N_d correlation caused by seasonality plus simulated cloud adiabaticity that differs from observations mute the effect of N_d on A in E3SMv1, while smaller R_{eff} in E3SMv1 than observed amplify A for a given N_d.

4.1.3 Aerosol Activation into Cloud Droplets

Twomey effect differences between E3SMv1 and observations also depend on the response of N_d to CCN ($\frac{dlnN_d}{dlnCCN}$). This term is expected to strongly depend on aerosol activation, though with modulation by N_d sinks such as evaporation and precipitation. It is only evaluated for situations in which clouds are more likely to be coupled with the surface where CCN measurements are made. Although Jones et al. (2011) use a threshold of 0.5°C for the difference in cloud base and near surface potential temperature, we increase this to 2°C to retain more samples and to account for uncertainty in potential temperature measurements obtained from interpolated soundings that are often separated by 12 hours.

Figure 15 shows N_d as a function of surface 0.2% CCN for all datasets, and correlation coefficients from Theil-Sen robust linear regression are provided in Figure 16. The sensitivity of N_d to CCN is greater for E3SM direct output than any other dataset. The sensitivity reduces for E3SM TOA but remains greater than Obs TOA datasets, and it reduces further for E3SM Sfc, which aligns well with Obs Sfc datasets. The fit is weaker in observations than E3SMv1 as shown by the spread of N_d values for a given CCN value (Fig. 15) and the correlation coefficients in Figure 16 (0.29-0.54 for observations vs. 0.39-0.85 for E3SMv1 datasets). A substantial portion of this spread is a result of errors in TOA- and surface-retrieved N_d values. This is shown by the greater spread in the COSP- and surface-simulated E3SMv1 N_d relationships as compared to the direct N_d relationship, quantified by differences in E3SMv1 correlation coefficients (0.85 (direct) vs. 0.65 (satellite) and 0.39 (surface)) as compared to observational coefficients (0.46-0.54 (TOA) and 0.29-0.30 (surface)). Why do surface retrieved $\frac{dlnN_d}{dlnCCN}$ values agree between E3SMv1 and observations while TOA values do not? As for $\frac{\partial lnCOD}{\partial lnN_d}$ discrepancies, this contrast is caused by variable cloud adiabaticity in surface retrievals and constant 80% adiabaticity in TOA retrievals. When surface retrieved N_d is recomputed assuming 80% adiabaticity, E3SMv1's N_d sensitivity to CCN substantially increases, but the observed sensitivity does not (dark green to light green symbols in Fig. 16). Thus, once the effect of cloud adiabaticity differences between E3SMv1 and observations on N_d are removed, it becomes clear that the sensitivity of N_d to CCN is too high in E3SMv1. Aircraft observed N_d vs. CCN concentration during ACE-ENA also supports a conclusion of greater sensitivity in E3SMv1 (Tang et al. 2023). This explains why surface-retrieved N_d values agree between E3SM Sfc and Obs Sfc but are greater for E3SM TOA than Obs TOA. Lower adiabaticities in E3SMv1 than observed lower N_d further from their adiabatic value, but when that effect is removed, it becomes clear that E3SMv1 has higher N_d values than observed despite reasonable CCN values because of a greater sensitivity of N_d to CCN.

These comparisons suggest that aerosol activation in E3SMv1 is potentially too high. This is consistent with the findings of Ghan et al. (2011) for the Abdul-Razzak Ghan (ARG) scheme (Abdul-Razzak and Ghan 2000) used in E3SM. Gong et al. (2023) also find that the ARG scheme coupled with the Cloud Layers Unified by Binormals (Golaz et al. 2002, Larson and Golaz 2005) and the four-mode Modal Aerosol Module (Liu et al. 2016) in the Community Earth System Model version 2.1 with the Community Atmosphere Model version 6 (Danabasoglu et al. 2020) produces greater cloud supersaturations than retrieved from observations at ENA. With a similar setup in E3SMv1, this could also be influencing the $\frac{dlnN_d}{dlnCCN}$ differences shown here. However, it is also possible that unrealistic N_d sinks including precipitation and evaporation contribute to $\frac{dlnN_d}{dlnCCN}$ being greater in E3SMv1 than observed, and more investigation into these processes is required.

E3SMv1-observation $\frac{dlnN_d}{dlnCCN}$ differences offset $\frac{\partial A}{\partial lnN_d}$ differences for surface retrievals, producing Twomey effect and albedo susceptibility estimates that are similar (Fig. 17). For TOA estimates, $\frac{\partial A}{\partial lnN_d}$ differences are too great to overcome, leading to a weaker than observed Twomey effect and albedo susceptibility (Fig. 17). Removing SZA-N_d correlation effects by examining COD susceptibility shows in fact that the Twomey effect is greater in E3SMv1 than observed, though this is only true for surface retrievals once model-observation cloud adiabaticity differences are removed. Removing adiabaticity effects gives Twomey effects that are about 30-40% greater in E3SMv1 than observed, while lnCOD susceptibility can be up to a factor of 2 greater depending on dataset (Fig. 18).

4.2 LWP Susceptibility Comparison

Since only overcast cloud conditions are considered, the second aerosol indirect effect is confined to the LWP susceptibility to N_d ($\frac{dlnLWP}{dlnN_d}$), which can accentuate or mute the Twomey effect. Figure 19 shows N_d-bin normalized joint distributions N_d and LWP for the various datasets analyzed. Like the Twomey effect, the quantified LWP responses depend on the datasets considered since LWP and N_d values shift depending on how they are retrieved. Somewhat poor linear fits are readily apparent, as highlighted by low correlation coefficients (0.21-0.58) in Figure 20. E3SM susceptibilities range from -0.3 to -0.4, which is similar to those in surface observations, while TOA observations have weaker susceptibilities around -0.2. Removing cloud adiabaticity effects in surface retrievals results in a weaker LWP susceptibility in E3SM than observations, while the opposite is true for TOA retrievals. Thus, there is no consensus among comparisons, and all that can be said is that E3SMv1 has similar values to observed, which is a success in that most GCMs fail to produce a negative LWP susceptibility (Quaas et al. 2009, Gryspeerdt et al. 2020).

Previous satellite retrieval and LES studies discussed in the introduction have highlighted an inverted V response of LWP to N_d which has been hypothesized to be caused by drizzle suppression at relatively low N_d values increasing LWP but entrainment-driven evaporation mechanisms reducing LWP for relatively high N_d values in which non-drizzling clouds. To assess whether LWP responses to N_d change as N_d increases, linear fits are applied for 2 different ranges of N_d shown as dashed black lines in Figure 19a-g and quantified in Figure 20. Observational datasets become more negative as N_d increases in agreement with past satellite retrieval studies. However, the opposite trend exists in E3SMv1 direct output and TOA datasets where LWP responses become less negative as N_d increases, with the caveat that uncertainty is high for $N_d < 50$ cm^{-3} due to limited sampling that is possibly associated with constraining analyses to overcast conditions. For surface retrievals, E3SM's LWP susceptibility becomes more negative with increasing N_d like observed, but once adiabaticity is held constant, the opposite occurs in disagreement with observations. Thus, the change in E3SM's LWP susceptibility with increasing N_d is generally opposite to the expectation from proposed physical mechanisms in past observational and LES studies, indicating possibly different mechanisms that cause its overall negative sign than in observations.

4.2.1 Potential Physical Mechanisms Affecting Model-Observation Comparisons

Are negative LWP susceptibilities caused by physical processes, and if so, are they the same in observations and E3SMv1? If
410   clouds are responding to the entrainment-evaporation mechanism, one might expect the LWP susceptibility to change with the
inversion strength and above inversion RH. To assess this, the median LWP susceptibility is plotted in joint EIS-700-hPa RH bins
in Figure 21. No clear dependencies on EIS or RH arise in surface observations (Fig. 21a), while the LWP response becomes less
negative as RH increases in TOA observations (Fig. 21c), consistent with more widespread satellite measurements of marine warm
clouds in Chen et al. (2014). However, the simulated LWP response to N_d becomes more negative as RH increases for a given EIS
415   (Fig. 21b, d, e), which indicates that entrainment-driven evaporation is not a major control on the negative LWP response in
E3SMv1.

However, it is apparent from the sample size contours in Figure 21 that E3SM and observations cover different EIS and 700-
hPa RH conditions. Indeed, for overcast cloud conditions at ENA, E3SMv1 has weaker than observed inversion strengths and
higher than observed above inversion relative humidity (Fig. 22a-b). Even though the simulation winds are nudged to MERRA2,
420   thermodynamic state is not and can have errors develop. These differences are associated with clouds in E3SMv1 that have higher
and colder cloud tops than observed (Fig. S7c-d) with greater cloud depths (Fig. 21c). TOA estimates of cloud depth are shallower
than surface observations and E3SMv1 due to assuming that clouds are 80% adiabatic whereas most clouds have lesser adiabaticity
(Fig. 21d). For a given LWP and N_d value, E3SMv1 can have substantially deeper clouds than observed, often by a couple hundred
meters (Fig. S8), and weaker inversions in E3SMv1 are present for all LWP-N_d combinations (Figs. S9-10). While E3SM direct
425   model output and surface retrievals have inversions that strengthen as N_d increases with clouds that become shallower, the EIS
gradient is notably absent in TOA datasets and the cloud depth dependence on N_d changes sign. This is likely associated with the
lack of variable cloud adiabaticity in TOA datasets and serves as caution for using such datasets alone to infer physical processes
and evaluate model output. Past studies have shown that weaker stability with boundary layer and cloud deepening causes the LWP
susceptibility to become more negative (e.g., Possner et al. 2020, Zhang et al. 2022). Thus, one might expect a more negative LWP
430   susceptibility in E3SMv1 than observed if it were properly representing processes controlling the LWP response to N_d, but this is
not the case, again suggesting that there are other causes for the negative simulated sign.

Another factor affecting LWP susceptibility is cloud adiabaticity (dark vs. light green symbols in Fig. 20). Most of the sampled
observed clouds are subadiabatic, consistent with the results of Wu et al. (2020) and with a mean of 71% that is only slightly higher
than the 63% found in Merk et al. (2016) and not far below the 80% assumed in TOA retrievals. However, E3SMv1's mean
435   adiabaticity is 27%. 68% of simulated clouds have an adiabaticity < 30%, whereas this is only 16% in observations (Fig. 21d). On
the other hand, 64% of observed clouds are more than 60% adiabatic but only 12% of clouds sampled in E3SMv1 reach this
threshold. Thus, an assumption of 80% adiabaticity for E3SM clouds is way off. In observations, clouds become more adiabatic
as N_d increases, potentially due to drizzle suppression as N_d increases (Fig. 12). However, clouds only become slightly more
adiabatic as N_d increases in E3SMv1 (Fig. 12). Median hourly surface rain rates as a function of LWP and N_d are shown in Figure
440   23 with rain rates less than 0.001 mm h^{-1} removed due to limited observational sensitivity. Observed clouds with less than 100 g
m^{-2} LWP or greater than 50 cm^{-3} N_d usually do not have rain reaching the surface, but this is common in E3SMv1 where sensitivity
of surface rain rate to N_d and LWP are muted. The hourly surface rain rate exceeds 0.001 mm h^{-1} 62% of the time in E3SMv1 as
compared to 15-18% in disdrometer and optical rain gauge observations, while rates exceed 0.01 mm h^{-1} nearly half the time in
E3SMv1 as compared to ~10% in observations (Table S1). Total surface rainfall accumulations are also nearly 3 times greater in
445   E3SMv1 as evidenced by average rain rates inclusive of all times in Table S1. Too frequent and light precipitation is a well-known
problem in GCMs (Stephens et al. 2010, Song et al. 2018) including E3SM (Ma et al. 2022). 84% of this surface rainfall is produced
in the Z-M convection parameterization (Table S1) while the frequency of stratiform drizzle is closer to observed. Figure S11

shows that surface hourly rain rates compare much more favorably when only stratiform rain rates for E3SM and E3SM Sfc are considered, indicating the Z-M parameterization is a primary driver of the drizzle frequency and amount biases. This is partly caused by deeper clouds in E3SM that more heavily precipitate (Fig. S12), likely a result of weaker inversions in E3SM. However, even for given cloud depth and N_d values, clouds precipitate more heavily (Fig. S12) and are less adiabatic (Fig. S13) than observed. It is probable that the excessive convective drizzle in E3SMv1 contributes to much less adiabatic clouds than observed, a difference that increases as N_d increases when observed drizzle ceases but E3SMv1 clouds continue to drizzle with only slightly lesser rates (Figs. 12, 23, and S12-13). It's possible that this muted sensitivity of rain rate and adiabaticity to N_d contributes to the negative relationship of LWP with N_d, but more research is needed to better understand how rain rate, LWP, and N_d interact to affect albedo susceptibility.

A final important consideration for interpreting LWP susceptibility quantification is that the LWP-N_d joint distribution is likely not entirely due to N_d effects on LWP. Confounding factors such as meteorology or mesoscale cloud structures may combine with microphysical processes to create overall negative LWP-N_d relationships. Indeed, present day minus pre-industrial E3SMv1 simulations produce a positive rather than negative LWP susceptibility (Christensen et al. 2023). Albedo and COD susceptibilities are also generally slightly less than Twomey effects despite LWP susceptibility estimates that are large enough to completely cancel the Twomey effects (Figs. 17-18), suggesting that spread in the $\frac{dlnLWP}{dlnN_d}$ and $\frac{dlnN_d}{dlnCCN}$ relationships and unaccounted for covariations between lnLWP and lnCCN partly counteract the estimated LWP response. Thus, further research is needed to develop methods for sufficiently evaluating cloud adjustments in GCMs using present day observations.

5 Conclusions

Surface-based and satellite observations collected at the ARM ENA site in the Azores over 5 years were used to evaluate factors controlling single layer liquid cloud albedo susceptibility for overcast cloud conditions in E3SMv1. While a single geographical location such as the ENA site is not representative of all environments and cloud types, its relatively comprehensive measurements and retrievals including CCN number concentration, atmospheric stability and humidity, radiative fluxes, and cloud LWP, N_d, rain rate, and adiabaticity in a marine environment with frequent single layer liquid clouds susceptible to aerosol influences allows for detailed model evaluation targeting specific processes. This methodology provides valuable context for more traditional model evaluation via longer-term, global metrics. Surface retrievals with 5- and 60-min resolutions, geostationary satellite retrievals with 4-km and 1° resolutions, and 1° resolution E3SMv1 datasets using different N_d and LWP retrievals from raw model output, COSP satellite simulator output, and surface variables, are all analyzed to assess the robustness of model-observation comparisons.

Simulated cloud albedo values are greater for given LWP and N_d values due to R_{eff} being smaller than retrieved in observations. However, cloud albedo sensitivities to N_d and LWP are well simulated by E3SMv1 after accounting for a simulated SZA-N_d correlation caused by seasonality that is not observed by analyzing COD susceptibility and after accounting for differences in observed and simulated cloud adiabaticity. Both effects mute the simulated sensitivity of cloud albedo to N_d. E3SMv1 adiabaticities are much lower than observed and the common 80% assumed in satellite retrievals, and adiabaticities affect the sensitivity of cloud albedo to lnN_d by altering the width of the lnN_d distribution. LWP is also less than observed while N_d is greater than observed if cloud adiabaticity differences between E3SMv1 and observations are removed, despite a similar overall distribution of surface CCN concentration to observed. Because of logarithmic sensitivities, this means that linear perturbations in N_d will result in a lesser Twomey effect in E3SMv1 compared to observations at ENA, though perturbations in LWP could potentially result in a greater effect.

Greater N_d values in E3SMv1 may result from excessive activation of CCN. This drives a Twomey effect that is too large by 30-40% when using TOA and surface retrieved N_d values. However, this difference again only emerges after controlling for SZA and cloud adiabaticity. LWP decreases as N_d increases like observed, something that is not produced by most GCMs. However, the LWP susceptibility becomes more negative in E3SMv1 as above inversion RH increases, a trend that is not observed and not

490 consistent with drier conditions facilitating more evaporation. In addition, the simulated LWP susceptibility does not become more negative as N_d increases like observed. Thus, it is unlikely that an entrainment-driven evaporation mechanism is operating in E3SMv1. E3SMv1 also has weaker inversions and deeper clouds than observed, which partly contributes to excessive convectively driven drizzle that is a common bias in GCMs (Stephens et al. 2010), though a substantial bias remains after controlling for cloud depth. There is also little decrease in drizzle rate and little increase in adiabaticity as N_d increases for a given LWP or cloud depth

495 in E3SMv1, which is in stark contrast to observations. Because of these major differences in observed and simulated cloud properties, the causes for the similar negative LWP susceptibility in E3SMv1 and observations likely differ. The negative LWP responses also only slightly decrease the overall cloud albedo susceptibilities, especially for E3SMv1, indicating that there are confounding effects that render a statistical correlation between LWP and N_d insufficient for assessing N_d impacts on LWP.

Because E3SMv1 has a greater Twomey effect than observed and because adjustments lower the COD susceptibility more in

500 observations than E3SMv1, simulated COD susceptibility values are about double those observed after controlling for cloud adiabaticity differences, which is consistent with overly negative ERFaci assessed in previous studies (Golaz et al. 2019, Rasch et al. 2019, Wang et al. 2020, Ma et al. 2022, Zhang et al. 2022). This also indicates that greater $\frac{dlnN_d}{dlnCCN}$ in E3SMv1, as modulated by aerosol activation, N_d evaporation, and precipitation, along with lesser muting by adjustments, are the primary drivers of greater radiative susceptibilities to aerosols in E3SMv1.

505 Differences between E3SMv1 and observations often exceed the observational spreads after critically controlling for cloud adiabaticity model-observation differences, but the spreads are still substantial. Both retrieval scale and simplifying assumptions contribute similar magnitudes of uncertainty in quantifying cloud albedo susceptibility and the Twomey effect, while retrieval assumptions contribute the most uncertainty in quantifying the LWP susceptibility. A major contributor to these discrepancies is the retrieval of N_d. For the most accurate N_d retrieval, it is best to compute N_d at the highest resolution possible and average it to

510 coarser scales (McComiskey and Feingold 2012, Feingold et al. 2022). However, because E3SM surface and TOA N_d retrievals were performed using E3SMv1 1° output, 1° and 60-min observational N_d retrievals were instead computed using inputs averaged to those scales so that comparisons used consistent methods. This creates differences as a function of data resolution. However, the spreads in Twomey effect and LWP susceptibility estimates from E3SMv1 are of similar magnitude to observational spreads despite being computed at a constant resolution. This occurs because the retrieval assumptions create shifts in the N_d distribution

515 relative to its true shape in direct model output. Retrieval biases are to be expected given assumptions of constant N_d in the cloud layer, drizzle contamination, and scaled adiabatic droplet growth between cloud base and cloud top, but this clearly demonstrates how significant the impacts can be from such assumptions in quantifying aerosol-cloud interactions.

Multiple model-observation comparison approaches like those used in this study coupled with in situ data comparisons provide critical context in the absence of well-quantified sampling and retrieval biases. Comparative analyses used in this study are being

520 implemented into the open source ESMAC-Diags software package (Tang et al. 2022, 2023) and expanded to additional locations with differing aerosol, cloud, and meteorological conditions. However, despite designing this study to have interpretable model-observation comparisons, it does not overcome several key issues that have led to persistent uncertainty of ERFaci through successive generations of climate models. These include the small magnitude of the forcing relative to the shortwave cloud radiative effect, its dependency on massive cross-scale atmospheric interactions, an inability to isolate specific model components that

525 contribute to errors, and an inability to estimate such effects with sufficient accuracy in observations (Mülmenstädt and Feingold

2018). Furthermore, it does not overcome the problem of present-day statistically derived cloud susceptibilities being insufficient to describe susceptibilities based on present-day minus pre-industrial conditions (e.g., Ghan et al. 2016, Christensen et al. 2022).

There are several implications of this study's results. First, the diurnal and seasonal cycles of LWP and N_d can modulate cloud albedo susceptibility due to correlations with the SZA, which impacts interpretations of model-observation differences. Second, large uncertainties exist due to N_d retrieval assumptions that limit our ability to accurately quantify factors affecting cloud albedo susceptibility. Continued and expanded integration of complementary in situ, remote sensing, and high-resolution model datasets will aid this effort. Critical to properly predicting albedo susceptibility are improved understanding and quantification of LWP and CF adjustments that respond to poorly simulated small-scale entrainment and precipitation processes operating over days and 100s of kilometers, processes that are not well represented in GCMs. Of equal importance is improving parameterized activation of aerosol particles into cloud droplets, which is known to be problematic (e.g., Ghan et al. 2011) and can turn realistic simulated radiation responses to N_d perturbations into greatly amplified (as in this study) or muted responses to aerosol perturbations. Recent work shows promise in improving aerosol activation parameterization (e.g., Silva et al. 2021). Further studies are required to continue improvement and to critically improve simulated convective drizzle processes that influence quantification of both the Twomey effect and cloud adjustments via impacts on cloud adiabaticity. Lastly, more studies are needed to assess how confounding factors such as erroneous responses to meteorology (e.g., Eastman et al. 2021) contribute to susceptibilities and interpretations of model-observation comparisons.

Code and Data Availability

Files containing all the observations, E3SM output variables, and necessary E3SM setup files used in this study can be downloaded from https://portal.nersc.gov/project/m3525/avarble/varble_et_al_2023_acp/. Python notebooks used to make these files including retrievals of several variables can also be downloaded from the same location. Original DOE ARM datasets at the ENA site used in this research can be downloaded from https://adc.arm.gov/discovery/#/results/site_code::ena.

Supplement

Table S1 and Figure S1-S13 are available in the Supplement.

Author Contributions

ACV wrote the manuscript, performed analyses, and created visualizations. PLM performed E3SMv1 simulations. Research conceptualization, and reviewing and editing of the manuscript, was performed by ACV, PLM, MWC, JM, ST, and JF.

Competing Interests

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

Acknowledgements

This research was supported by the "Enabling Aerosol cloud interactions at Global convection-permitting scalES (EAGLES)" project (74358), funded by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research,

Earth System Model Development program area. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility located at Lawrence Berkeley National Laboratory, operated under contract no. DE-AC02-05CH11231 using NERSC awards ALCCERCAP0016315, BER-

560   ERCAP0015329, BER-ERCAP0018473, and BER-ERCAP0020990. We thank the numerous AMR instrument and data mentors for providing data. The Pacific Northwest National Laboratory is operated for the U.S. Department of Energy by Battelle Memorial Institute under contract no. DE-AC05-76RL01830.

References

Abdul-Razzak, H., and Ghan, S. J.: A parameterization of aerosol activation: 2. Multiple aerosol types, *Journal of Geophysical*
565   *Research*, 105, 6837-6844, doi:10.1029/1999JD901161, 2000.

Ackerman, A. S., Kirkpatrick, M. P., Stevens, D. E., and Toon, O. B.: The impact of humidity above stratiform clouds on indirect aerosol climate forcing, *Nature*, 432, 1014–1017, doi:10.1038/nature03174, 2004.

Albrecht, B. A.: Aerosols, Cloud Microphysics, and Fractional Cloudiness, *Science*, 245, 1227–1230, doi:10.1126/science.245.4923.1227, 1989.

570   Atmospheric Radiation Measurement (ARM) user facility: Interpolated Sonde (INTERPOLATEDSONDE), 2016-01-01 to 2020-12-31, Eastern North Atlantic (ENA) Graciosa Island, Azores, Portugal (C1), compiled by M. Jensen, S. Giangrande, T. Fairless and A. Zhou, ARM Data Center, dataset accessed 2021-08-12 at doi:10.5439/1095316, 2013a.

Atmospheric Radiation Measurement (ARM) user facility: Minnis Cloud Products Using Visst Algorithm (VISSTGRIDM10MINNIS), 2016-01-01 to 2018-02-20, Eastern North Atlantic (ENA) External Data (satellites and others)
575   (X1), ARM Data Center, dataset accessed 2021-08-12, 2013b.

Atmospheric Radiation Measurement (ARM) user facility: Surface Meteorological Instrumentation (MET), 2016-01-01 to 2020-12-31, Eastern North Atlantic (ENA) Graciosa Island, Azores, Portugal (C1), compiled by J. Kyrouac and Y. Shi, ARM Data Center, dataset accessed 2021-08-12 at doi:10.5439/1786358, 2013c.

Atmospheric Radiation Measurement (ARM) user facility: Cloud Optical Properties from MFRSR Using Min Algorithm
580   (MFRSRCLDOD1MIN), 2016-01-01 to 2020-12-31, Eastern North Atlantic (ENA) Graciosa Island, Azores, Portugal (C1), compiled by D. Zhang, ARM Data Center, dataset accessed 2021-12-30 at doi:10.5439/1027296, 2014a.

Atmospheric Radiation Measurement (ARM) user facility: Laser Disdrometer Quantities (LDQUANTS), 2016-01-01 to 2020-12-31, Eastern North Atlantic (ENA) Graciosa Island, Azores, Portugal (C1), compiled by J. Hardin, S. Giangrande and A. Zhou. ARM Data Center, dataset accessed 2021-08-12 at doi:10.5439/1432694, 2014b.

585   Atmospheric Radiation Measurement (ARM) user facility: Minnis Cloud Products Using Visst Algorithm (VISSTPXM10MINNIS), 2016-01-01 to 2018-02-20, Eastern North Atlantic (EN) External Data (satellites and others) (X1), ARM Data Center, dataset accessed 2021-08-12, 2014c.

Atmospheric Radiation Measurement (ARM) user facility: MWR Retrievals with MWRRET Version 2 (MWRRET2TURN), 2016-01-01 to 2020-12-31, Eastern North Atlantic (ENA) Graciosa Island, Azores, Portugal (C1), compiled by K. Gaustad,
590   ARM Data Center, dataset accessed 2021-12-30 at doi:10.5439/1566156, 2014d.

Atmospheric Radiation Measurement (ARM) user facility: Active Remote Sensing of CLouds (ARSCL) product using Ka-band ARM Zenith Radars (ARSCLKAZRBND1KOLLIAS), 2016-01-01 to 2020-12-31, Eastern North Atlantic (ENA) Graciosa Island, Azores, Portugal (C1), compiled by K. Johnson, S. Giangrande and T. Toto, ARM Data Center, dataset accessed 2021-08-12 at doi:10.5439/1393438, 2015.

595   Atmospheric Radiation Measurement (ARM) user facility: Cloud Condensation Nuclei Particle Counter (AOSCCN1COLAVG),
        2016-06-22 to 2020-10-28, Eastern North Atlantic (ENA) Graciosa Island, Azores, Portugal (C1), compiled by A. Koontz and
        C. Flynn, ARM Data Center, dataset accessed 2021-08-12 at doi:10.5439/1255094, 2016a.

      Atmospheric Radiation Measurement (ARM) user facility: Cloud Condensation Nuclei Particle Counter
        (AOSCCN1COLSPECTRA), 2016-06-22 to 2020-10-28, Eastern North Atlantic (ENA) Graciosa Island, Azores, Portugal
600   (C1), compiled by J. Uin, ARM Data Center, dataset accessed 2021-08-12 at doi:10.5439/1342134, 2016b.

      Atmospheric Radiation Measurement (ARM) user facility: Minnis Cloud Products Using Visst Algorithm
        (VISSTGRIDM11MINNIS), 2018-02-20 to 2020-12-31, Eastern North Atlantic (EN) External Data (satellites and others)
        (X1), ARM Data Center, dataset accessed 2021-08-12, 2018a.

      Atmospheric Radiation Measurement (ARM) user facility: Minnis Cloud Products Using Visst Algorithm
605   (VISSTPX2DM11MINNIS), 2018-02-20 to 2020-12-31, Eastern North Atlantic (EN) External Data (satellites and others)
        (X1), ARM Data Center, dataset accessed 2021-08-12, 2018b.

      Bellouin, N., Quaas, J., Gryspeerdt, E., Kinne, D., Stier, P., Watson-Parris, D., Boucher, O., Carslaw, K. S., Christensen, M.,
        Daniau, A.-L., Dufresne, J.-L., Feingold, G., Fiedler, S., Forster, P., Gettelman, A., Haywood, J. M., Lohmann, U., Malavelle,
        F., Mauritsen, T., McCoy, D. T., Myhre, G., Mülmenstädt, J., Neubauer, D., Possner, A., Rugenstein, M., Sato, Y., Schulz,
610   M., Schwartz, S. E., Sourdeval, O., Storelvmo, T., Toll, V., Winker, D., and Stevens, B.: Bounding global aerosol radiative
        forcing of climate change, *Reviews of Geophysics*, 58, e2019RG000660, doi:10.1029/2019RG000660, 2020.

      Bennartz, R.: Global assessment of marine boundary layer cloud droplet number concentration from satellite, *Journal of
        Geophysical Research Atmospheres*, 112, D02201, doi:10.1029/2006JD007547, 2007.

      Boucher, O., Randall, D., Artaxo, P., Bretherton, C., Feingold, G., Forster, P., Kerminen, V.-M., Kondo, Y., Liao, H., Lohmann,
615   U., Rasch, P., Satheesh, S., Sherwood, S., Stevens, B., and Zhang, X.: *Clouds and aerosols, in Climate change 2013: The
        physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on
        Climate Change*, edited by T. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex,
        and P. Midgley, chap. 7, pp. 571–658, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA,
        doi:10.1017/CBO9781107415324.016, 2013.

620   Brenguier, J.-L., Burnet, F., and Geoffroy, O.: Cloud optical thickness and liquid water path - does the k coefficient vary with
        droplet concentration?, *Atmospheric Chemistry and Physics*, 11, 9771-9786, doi:10.5194/acp-11-9771-2011, 2011.

      Bretherton, C. S., Blossey, P. N., and Uchida, J.: Cloud droplet sedimentation, entrainment efficiency, and subtropical
        stratocumulus albedo, *Geophysical Research Letters*, 34, L03813, doi:10.1029/2006GL027648, 2007.

      Cadeddu, M. P., Ghate, V. P., and Mech, M.: Ground-based observations of cloud and drizzle liquid water path in stratocumulus
625   clouds, *Atmospheric Measurement Techniques*, 13, 1485–1499, doi:10.5194/amt-13-1485-2020, 2020.

      Carslaw, K., Lee, L., Reddington, C., Pringle, K. J., Rap, A., Forster, P. M., Mann, G. W., Spracklen, D. V., Woodhouse, M. T.,
        Regayre, L. A., and Pierce, J. R.: Large contribution of natural aerosols to uncertainty in indirect forcing, *Nature*, 503, 67–71,
        doi:10.1038/nature12674, 2013.

      Chen, YC., Christensen, M., Stephens, G., and Seinfeld, J. H.: Satellite-based estimate of global aerosol–cloud radiative forcing
630   by marine warm clouds. *Nature Geosci*ence, 7, 643–646, doi:10.1038/ngeo22142, 2014.

      Christensen, M. W., Neubauer, D., Poulsen, C. A., Thomas, G. E., McGarragh, G. R., Povey, A. C., Proud, S. R., and Grainger, R.
        G.: Unveiling aerosol–cloud interactions – Part 1: Cloud contamination in satellite products enhances the aerosol indirect
        forcing estimate, *Atmospheric Chemistry and Physics*, 17, 13,151–13,164, doi:10.5194/acp-17-13151-2017, 2017.

Christensen, M. W., Jones, W. K., and Stier, P.: Aerosols Enhance Cloud Lifetime and Brightness along the Stratus-to-Cumulus Transition, *Proceedings of the National Academy of Sciences of the United States of America*, 117, 17591-17598, doi:10.1073/pnas.1921231117, 2020.

Christensen, M. W., Gettelman, A., Cermak, J., Dagan, G., Diamond, M., Douglas, A., Feingold, G., Glassmeier, F., Goren, T., Grosvenor, D. P., Gryspeerdt, E., Kahn, R., Li, Z., Ma, P.-L., Malavelle, F., McCoy, I. L., McCoy, D. T., McFarquhar, G., Mülmenstädt, J., Pal, S., Possner, A., Povey, A., Quaas, J., Rosenfeld, D., Schmidt, A., Schrödner, R., Sorooshian, A., Stier, P., Toll, V., Watson-Parris, D., Wood, R., Yang, M., and Yuan, T.: Opportunistic experiments to constrain aerosol effective radiative forcing, *Atmospheric Chemistry and Physics*, 22, 641-674, doi:10.5194/acp-22-641-2022, 2022.

Christensen, M. W., Ma, P.-L., Wu, P., Varble, A. C., Mülmenstädt, J., and Fast, J.: Evaluation of aerosol-cloud interactions in E3SSM using a Lagrangian framework, *Atmospheric Chemistry and Physics*, 23, 2789-2812, doi:10.5194/acp-23-2789-2023, 2023.

Clothiaux, E. E., Ackerman, T. P., Mace, G. G., Moran, K. P., Marchand, R. T., Miller, M. A., and Martner, B. E.: Objective determination of cloud heights and radar reflectivities using a combination of active remote sensors at the ARM CART sites, *Journal of Applied Meteorology*, 39, 645-665. doi:10.1175/1520-0450(2000)039<0645:ODOCHA>2.0.CO;2, 2000.

Coakley, J. A., Bernstein, R. L., and Durkee, P. A.: Effect of Ship-Stack Effluents on Cloud Reflectivity, *Science*, 237, 1020–1022, doi:10.1126/science.237.4818.1020, 1987.

Danabasoglu, G., Lamarque, J.-F., Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., Emmons, L. K., Fasullo, J., Garcia, R., Gettelman, A., Hannay, C., Holland, M. M., Large, W. G., Lauritzen, P. H., Lawrence, D. M., Lenaerts, J. T. M., Lindsay, K., Lipscomb, W. H., Mills, M. J., Neale, R., Oleson, K. W., Otto-Bliesner, B., Phillips, A. S., Sacks, W., Tilmes, S., van Kampenhout, L., Vertenstein, M., Bertini, A., Dennis, J., Deser, C., Fischer, C., Fox-Kemper, B., Kay, J. E., Kinnison, D., Kushner, P. J., Larson, V. E., Long, M. C., Mickelson, S., Moore, J. K., Nienhouse, E., Polvani, L., Rasch, P. J., and Strand, W. G.: The Community Earth System Model Version 2 (CESM2), *Journal of Advances in Modeling Earth Systems*, *12*, e2019MS001916, doi:10.1029/2019MS001916, 2020.

Douglas, A. and L'Ecuyer, T.: Quantifying variations in shortwave aerosol–cloud–radiation interactions using local meteorology and cloud state constraints, *Atmospheric Chemistry and Physics*, 19, 6251–6268, doi:10.5194/acp-19-6251-2019, 2019.

Douglas, A. and L'Ecuyer, T.: Quantifying cloud adjustments and the radiative forcing due to aerosol–cloud interactions in satellite observations of warm marine clouds, *Atmospheric Chemistry and Physics*, 20, 6225–6241, doi:10.5194/acp-20-6225-2020, 2020.

Eastman, R., Wood, R., and Bretherton, C. S.: Time Scales of Clouds and Cloud-Controlling Variables in Subtropical Stratocumulus from a Lagrangian Perspective, *Journal of the Atmospheric Sciences*, 73, 3079-3091, doi:10.1175/JAS-D-16-0050.1, 2016.

Eastman, R., Terai, C. R., Grosvenor, D. P., and Wood, R.: Evaluating the Lagrangian Evolution of Subtropical Low Clouds in GCMs Using Observations: Mean Evolution, Time Scales, and Responses to Predictors, *Journal of the Atmospheric Sciences*, 78, 353-372, doi:10.1175/JAS-D-20-0178.1, 2021.

Feingold, G., Goren, T., and Yamaguchi, T.: Quantifying albedo susceptibility biases in shallow clouds, *Atmospheric Chemistry and Physics*, 22, 3303-3319, doi:10.5194/acp-22-3303-2022, 2022.

Gallo, F., Uin, J., Springston, S., Wang, J., Zheng, G., Kuang, C., Wood, R., Azevedo, E. B., McComiskey, A., Mei, F., Theisen, A., Kyrouac, J., and Aiken, A. C.: Identifying a regional aerosol baseline in the eastern North Atlantic using collocated measurements and a mathematical algorithm to mask high-submicron-number-concentration aerosol events, *Atmospheric Chemistry and Physics*, 20, 7553-7573, doi:10.5194/acp-20-7553-2020, 2020.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G.,
Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., Silva, A. M. da, Gu, W., Kim,
G.-K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert,
S. D., Sienkiewicz, M., and Zhao, B.: The Modern-Era Retrospective Analysis for Research and Applications, Version 2
(MERRA-2), *Journal of Climate*, 30, 5419–5454, doi:10.1175/JCLI-D-16-0758.1, 2017.

Gettelman, A., and Morrison, H.: Advanced two-moment bulk microphysics for global models. Part I: Off-line tests and
comparison with other schemes, *Journal of Climate*, 28, 1268–1287, doi:10.1175/JCLI-D-14-00102.1, 2014.

Gettelman, A., Morrison, H., Santos, S., Bogenschutz, P., and Caldwell, P. M.: Advanced two-moment bulk microphysics for
global models. Part II: Global model solutions and aerosol-cloud interactions, *Journal of Climate*, 28, 1288–1307,
doi:10.1175/JCLI-D-14-00103.1, 2015.

Ghan, S. J., Abdul-Razzak, H., Nenes, A., Ming, Y., Liu, X., Ovchinnikov, M., Shipway, B., Meskhidze, N., Xu, J., and Shi, X.:
Droplet nucleation: Physically-based parameterizations and comparative evaluation, *Journal of Advances in Modeling Earth
Systems*, 3, M10001, doi:10.1029/2011MS000074, 2011.

Ghan S., Wang M., Zhang S., Ferrachat S., Gettelman A., Griesfeller J., Kipling Z., Lohmann U., Morrison H., Neubauer D.,
Partridge D. G., Stier P., Takemura T., Wang H., and Zhang K.: Challenges in constraining anthropogenic aerosol effects on
cloud radiative forcing using present-day spatiotemporal variability, *Proceedings of the National Academies of Sciences of the
United States of America*, 113, 5804–5811, doi:10.1073/pnas.1514036113, 2016.

Giangrande, S. E., Wang, D., Bartholomew, M. J., Jensen, M. P., Mechem, D. B., Hardin, J. C., and Wood, R., 2019: Midlatitude
oceanic cloud and precipitation properties as sampled by the ARM Eastern North Atlantic Observatory, *Journal of
Geophysical Research: Atmospheres*, 124, 4741– 4760, doi:10.1029/2018JD029667, 2019.

Glassmeier, F., Hoffmann, F., Johnson, J. S., Yamaguchi, T., Carslaw, K. S., and Feingold, G.: An emulator approach to
stratocumulus susceptibility, *Atmospheric Chemistry and Physics*, 19, 10191–10203. doi:10.5194/acp-19-10191-2019, 2019.

Golaz, J.-C., Larson, V. E., and Cotton, W. R.: A PDF-Based Model for Boundary Layer Clouds. Part I: Method and Model
Description, *Journal of the Atmospheric Sciences*, 59, 3540–3551, doi:10.1175/1520-555
0469(2002)059<3540:APBMFB>2.0.CO;2, 2002.

Golaz, J. C., Caldwell, P. M., Van Roekel, L. P., Petersen, M. 665 R., Tang, Q., Wolfe, J. D., Abeshu, G., Anantharaj, V., Asay-
Davis, X. S., Bader, D. C., Baldwin, S. A., Bisht, G., Bogenschutz, P. A., Branstetter, M., Brunke, M. A., Brus, S. R., Burrows,
S. M., Cameron-Smith, P. J., Donahue, A. S., Deakin, M., Easter, R. C., Evans, K. J., Feng, Y., Flanner, M., Foucar, J. G.,
Fyke, J. G., Griffin, B. M., Hannay, C., Harrop, B. E., Hoffman, M. J., Hunke, E. C., Jacob, R. L., Jacobsen, D.W., Jeffery,
N., Jones, P.W., Keen, N. D., Klein, S. A., Larson, V. E., Leung, L. R., Li, H. Y., Lin, W. Y., Lipscomb, W. H., Ma, P. L.,
Mahajan, S., Maltrud, M. E., Mametjanov, A., McClean, J. L., McCoy, R. B., Neale, R. B., Price, S. F., Qian, Y., Rasch, P. J.,
Eyre, J. E. J. R., Riley, W. J., Ringler, T. D., Roberts, A. F., Roesler, E. L., Salinger, A. G., Shaheen, Z., Shi, X. Y., Singh, B.,
Tang, J. Y., Taylor, M. A., Thornton, P. E., Turner, A. K., Veneziani, M., Wan, H., Wang, H. L., Wang, S. L., Williams, D.
N., Wolfram, P. J., Worley, P. H., Xie, S. C., Yang, Y., Yoon, J. H., Zelinka, M. D., Zender, C. S., Zeng, X. B., Zhang, C. Z.,
Zhang, K., Zhang, Y., Zheng, X., Zhou, T., and Zhu, Q.: The DOE E3SM Coupled Model Version 1: Overview and Evaluation
at Standard Resolution, *Journal of Advances in Modeling Earth Systems*, 11, 2089–2129, doi:10.1029/2018ms001603, 2019.

Gong, X., Wang, Y., Xie, H., Zhang, J., Lu, Z., Stratmann, F., Wex, H., Liu, X., and Wang, J.: Maximum supersaturation in the
marine boundary layer clouds over the North Atlantic, *AGU Advances*, 2023, in review.

Goren, T., Kazil, J., Hoffmann, F., Yamaguchi, T., and Feingold, G.: Anthropogenic air pollution delays marine stratocumulus
break-up to open-cells, *Geophysical Research Letters*, 46, 14135– 14144, doi:10.1029/2019GL085412, 2019.

715   Grosvenor, D. P., Sourdeval, O., Zuidema, P., Ackerman, A., Alexandrov, M. D., Bennartz, R., Boers, R., Cairns, B., Chiu, J. C., Christensen, M., Deneke, H., Diamond, M., Feingold, G., Fridlind, A., Hunerbein, A., Knist, C., Kollias, P., Marshak, A., McCoy, D., Merk, D., Painemal, D., Rausch, J., Rosenfeld, D., Russchenberg, H., Seifert, P., Sinclair, K., Stier, P., van Diedenhoven, B., Wendisch, M., Werner, F., Wood, R., Zhang, Z., and Quaas, J.: Remote sensing of droplet number concentration in warm clouds: A review of the current state of knowledge and perspectives, *Reviews of Geophysics*, 56, 409–453, doi:10.1029/2017RG000593, 2018.

720   Gryspeerdt, E., Quaas, J., and Bellouin, N.: Constraining the aerosol influence on cloud fraction, *Journal of Geophysical Research Atmospheres*, 121, 3566-3583, doi:10.1002/2015JD023744, 2016.

Gryspeerdt, E., Quaas, J., Ferrachat, S., Gettelman, A., Ghan, S., Lohmann, U., Morrison, H., Neubauer, D., Partridge, D. G., Stier, P., Takemura, T., Wang, H., Wang, M., and Zhang, K.: Constraining the instantaneous aerosol influence on cloud albedo, *Proceedings of the National Academy of Sciences of the United States of America*, 114, 4899–4904, 725   doi:10.1073/pnas.1617765114, 2017.

Gryspeerdt, E., Goren, T., Sourdeval, O., Quaas, J., Mülmenstädt, J., Dipu, S., Unglaub, C., Gettelman, A., and Christensen, M.: Constraining the aerosol influence on cloud liquid water path, *Atmospheric Chemistry and Physics*, 19, 5331-5347, doi:10.5194/acp19-5331-2019, 2019.

Gryspeerdt, E., Mülmenstädt, J., Gettelman, A., Malavelle, F. F., Morrison, H., Neubauer, D., Partridge, D. G., Stier, P., Takemura, 730   T., Wang, H., Wang, M., and Zhang, K.: Surprising similarities in model and observational aerosol radiative forcing estimates, *Atmospheric Chemistry and Physics*, 20, 613–623, doi:10.5194/acp-20-613-2020, 2020.

Gryspeerdt, E., McCoy, D. T., Crosbie, E., Moore, R. H., Nott, G. J., Painemal, D., Small-Griswold, J., Sorooshian, A., and Ziemba, L.: The impact of sampling strategy on the cloud droplet number concentration estimated from satellite data, *Atmospheric Measurement Techniques*, 15, 3875–3892, doi:10.5194/amt-15-3875-2022, 2022.

735   Hoffmann, F., Glassmeier, F., Yamaguchi, T., and Feingold, G.: Liquid Water Path Steady States in Stratocumulus: Insights from Process-Level Emulation and Mixed-Layer Theory, *Journal of the Atmospheric Sciences*, 77, 2203-2215, doi:10.1175/JAS-D-19-0241.1, 2020.

Jensen, M. P., V. P. Ghate, D. Wang, D. K. Apoznanski, M. J. Bartholomew, S. E. Giangrande, K. L. Johnson, and Thieman, M. M.: Contrasting characteristics of open- and closed-cellular stratocumulus cloud in the eastern North Atlantic, *Atmospheric* 740   *Chemistry and Physics*, 21, 557–14 571, doi:10.5194/acp-21-14557-2021, 2021.

Jiang, H., Xue, H., Teller, A., Feingold, G., and Levin, Z.: Aerosol effects on the lifetime of shallow cumulus, *Geophysical Research Letters*, 33, L14806. doi:10.1029/2006GL026024, 2006.

Jing, X., and Suzuki, K.: The impact of process-based warm rain constraints on the aerosol indirect effect, *Geophysical Research Letters*, 45, 10,729–10,737, doi:10.1029/2018GL079956, 2018.

745   Johnson, D. W., Osborne, S., Wood, R., Suhre, K., Johnson, R., Businger, S., Quinn, P. K., Wiedensohler, A., Durkee, P. A., Russell, L. M., Andreae, M. O., O'Dowd, C., Noone, K. J., Bandy, B., Rudolph, J., and Rapsomanikis, S.: An overview of the Lagrangian experiments undertaken during the North Atlantic regional Aerosol Characterisation Experiment (ACE-2), *Tellus B: Chemical and Physical Meteorology*, 52, 290–320, doi:10.3402/tellusb.v52i2.16110, 2000.

Jones, C. R., Bretherton, C. S., and Leon, D.: Coupled vs. decoupled boundary layers in VOCALS-REx, *Atmospheric Chemistry* 750   *and Physics*, 11, 7143–7153, doi:10.5194/acp-11-7143-2011, 2011.

Kang, L., Marchand, R. T., and Smith, W. L.: Evaluation of MODIS and Himawari-8 low clouds retrievals over the Southern Ocean with in situ measurements from the SOCRATES campaign, *Earth and Space Science*, 8, e2020EA001397, doi:10.1029/2020EA001397, 2021.

Kogan, Y. L., and Martin, W. J.: Parameterization of Bulk Condensation in Numerical Cloud Models, *Journal of Atmospheric Sciences*, 51, 1728-1739, doi:10.1175/1520-0469(1994)051<1728:POBCIN>2.0.CO;2, 1994.

Koren, I., Dagan, G., and Altaratz, O.: From aerosol-limited to invigoration of warm convective clouds, *Science*, 344, 1143–1146, doi:10.1126/science.1252595, 2004.

Larson, V. E., and Golaz, J.-C.: Using Probability Density Functions to Derive Consistent Closure Relationships among Higher-Order Moments, *Monthly Weather Review*, 133, 1023–1042, doi:10.1175/MWR2902.1, 2005.

Lebsock, M. D., Stephens, G. L., and Kummerow, C.: Multisensor satellite observations of aerosol effects on warm clouds, *Journal of Geophysical Research*, 113, D15205, doi:10.1029/2008JD009876, 2008.

Liu, X., Ma, P.-L., Wang, H., Tilmes, S., Singh, B., Easter, R. C., Ghan, S. J., and Rasch, P. J.: Description and evaluation of a new four-mode version of the Modal Aerosol Module (MAM4) within version 5.3 of the Community Atmosphere Model, *Geoscientific Model Development*, 9, 505–522, doi:10.5194/gmd-9-505-2016, 2016.

Long, C. N., and Ackerman, T. P.: Identification of clear skies from broadband pyranometer measurements and calculation of downwelling shortwave cloud effects, *Journal of Geophysical Research – Atmospheres*, 105, 15609–15626.,doi:10.1029/2000JD900077, 2000.

Long, C. N., Ackerman, T. P., Gaustad, K. L., and Cole, J. N. S.: Estimation of fractional sky cover from broadband shortwave radiometer measurements. *Journal of Geophysical Research – Atmospheres*, 111, D11204. doi:10.1029/2005JD006475, 2006.

Ma, P.-L., Harrop, B. E., Larson, V. E., Neale, R. B., Gettelman, A., Morrison, H., Wang, H., Zhang, K., Klein, S. A., Zelinka, M. D., Zhang, Y., Qian, Y., Yoon, J.-H., Jones, C. R., Huang, M., Tai, S.-L., Singh, B., Bogenschutz, P. A., Zheng, X., Lin, W., Quaas, J., Chepfer, H., Brunke, M. A., Zeng, X., Mülmenstädt, J., Hagos, S., Zhang, Z., Song, H., Liu, X., Pritchard, M. S., Wan, H., Wang, J., Tang, Q., Caldwell, P. M., Fan, J., Berg, L. K., Fast, J. D., Taylor, M. A., Golaz, J.-C., Xie, S., Rasch, P. J., and Leung, L. R.: Better calibration of cloud parameterizations and subgrid effects increases the fidelity of the E3SM Atmosphere Model version 1, *Geoscientific Model Development*, 15, 2881–2916, doi:10.5194/gmd-15-2881-2022, 2022.

Mather, J. H. and Voyles, J.W.: The ARM Climate Research Facility: A Review of Structure and Capabilities, *Bulletin of the American Meteorological Society*, 94, 377–392, doi:10.1175/bams-d-11-00218.1, 2013.

McComiskey, A., Feingold, G., Frisch, A. S., Turner, D. D., Miller, M. A., Chiu, J. C., Min, Q., and Ogren, J. A.: An assessment of aerosol-cloud interactions in marine stratus clouds based on surface remote sensing, *Journal of Geophysical Research – Atmospheres*, 114, D09203, doi:10.1029/2008JD011006, 2009.

McComiskey, A. and Feingold, G.: The Scale Problem in Quantifying Aerosol Indirect Effects, *Atmospheric Chemistry and Physics*, 12, 1031–1049, doi:10.5194/acp-12-1031-2012, 2012.

Mechem, D. B., Wittman, C. S., Miller, M. A., Yuter, S. E., and de Szoeke, S. P.: Joint synoptic and cloud variability over the Northeast Atlantic near the Azores, *Journal of Applied Meteorology and Climatology*, 57, 1273-1290, doi:10.1175/JAMC-D-17-0211.1, 2018.

Merk, D., Deneke, H., Pospichal, B., and Seifert, P.: Investigation of the adiabatic assumption for estimating cloud micro- and macrophysical properties from satellite and ground observations, *Atmospheric Chemistry and Physics*, 16, 933-952, doi:10.5194/acp-16-933-2016, 2016.

Michibata, T., Suzuki, K., Sato, Y., and Takemura, T.: The source of discrepancies in aerosol–cloud–precipitation interactions between GCM and A-Train retrievals, *Atmospheric Chemistry and Physics*, 16, 15413–15424, doi:10.5194/acp-16-15413-2016, 2016.

Min, Q, and Harrison, L. C.: Cloud properties derived from surface MFRSR measurements and comparison with GOES results at the ARM SGP site, *Geophysical Research Letters*, 23, 1641−1644, doi:10.1029/96GL01488, 1996.

795   Minnis, P., Nguyen, L., Palikonda, R., Heck, P. W., Spangenberg, D. A., Doelling, D. R., Ayers, J. K., Smith, J. W. L., Khaiyer, M. M., Trepte, Q. Z., Avey, L. A., Chang, F.-L., Yost, C. R., Chee, T. L., and Szedung, S.-M.: Near-real time cloud retrievals from operational and research meteorological satellites, *Proceedings Volume 7107, Remote Sensing of Clouds and the Atmosphere XIII*, doi:10.1117/12.800344, 2008.

Minnis, P., Sun-Mack, S., Young, D. F., Heck, P. W., Garber, D. P., Chen, Y., Spangenberg, D. A., Arduini, R. F., Trepte, Q. Z., Smith, W. L., Ayers, J. K., Gibson, S. C., Miller, W. F., Hong, G., Chakrapani, V., Takano, Y., Liou, K. N., Xie, Y., and Yang,
800   P.: CERES Edition-2 Cloud Property Retrievals Using TRMM VIRS and Terra and Aqua MODIS Data—Part I: Algorithms, *IEEE Transactions on Geoscience and Remote Sensing*, 49, 4374-4400, doi:10.1109/tgrs.2011.2144601, 2011.

Mohrmann, J., Bretherton, C. S., McCoy, I. L., McGibbon, J., Wood, R., Ghate, V., Albrecht, B., Sarkar, M., Zuidema, P., and Palikonda, R.: Lagrangian Evolution of the Northeast Pacific Marine Boundary Layer Structure and Cloud during CSET, *Monthly Weather Review*, 147, 4681-4700, doi:10.1175/MWR-D-19-0053.1, 2019.

805   Morrison, H., and Gettelman, A.: A new two-moment bulk stratiform cloud microphysics scheme in the Community Atmosphere Model, version 3 (CAM3). Part I: Description and numerical tests, *Journal of Climate*, 21, 3642–3659, doi:10.1175/2008JCLI2105.1, 2008.

Mülmenstädt, J. and Feingold, G.: The radiative forcing of aerosol–cloud interactions in liquid clouds: Wrestling and embracing uncertainty, *Current Climate Change Reports*, 4, 23–40, doi:10.1007/s40641-018-0089-y, 2018.

810   Mülmenstädt, J., Gryspeerdt, E., Salzmann, M., Ma, P.-L., Dipu, S., and Quaas, J.: Separating radiative forcing by aerosol–cloud interactions and rapid cloud adjustments in the ECHAM–HAMMOZ aerosol–climate model using the method of partial radiative perturbations, *Atmospheric Chemistry and Physics*, 19, 15415–15429, doi:10.5194/acp-19-15415-2019, 2019.

Penner, J. E., Xu, L., and Wang, M.: Satellite methods underestimate indirect climate forcing by aerosols, *Proceedings of the National Academy of Sciences of the United States of America*, 108, 13,404-13,408, doi:10.1073/pnas.1018526108, 2011.

815   Pincus, R. and Baker, M.: Effects of precipitation on the albedo susceptibility of clouds in the marine boundary layer, *Nature*, 372, 250–252, doi:10.1038/372250a0, 1994.

Pincus, R., Baker, M. B., and Bretherton, C. S.: What Controls Stratocumulus Radiative Properties? Lagrangian Observations of Cloud Evolution, *Journal of the Atmospheric Sciences*, 54, 2215–2236, doi:10.1175/1520-0469(1997)054<2215:WCSRPL>2.0.CO;2, 1997.

820   Pincus, R., Platnick, S., Ackerman, S. A., Hemler, R. S., and Hofmann, R. J. P.: Reconciling simulated and observed views of clouds: MODIS, ISCCP, and the limits of instrument simulators, *Journal of Climate*, 25, 4699-4720, doi:10.1175/JCLI-D-11-00267.1, 2012.

Platnick, S., and Twomey, S.: Determining the Susceptibility of Cloud Albedo to Changes in Droplet Concentration with the Advanced Very High Resolution Radiometer, *Journal of Applied Meteorology and Climatology*, 33, 334-347,
825   doi:10.1175/1520-0450(1994)033<0334:DTSOCA>2.0.CO;2, 1994.

Possner, A., Eastman, R., Bender, F., and Glassmeier, F.: Deconvolution of boundary layer depth and aerosol constraints on cloud water path in subtropical stratocumulus decks, *Atmospheric Chemistry and Physics*, 20, 3609–3621, doi:10.5194/acp-20-3609-2020, 2020.

Quaas, J., Boucher, O., Bellouin, N., and Kinne, S.: Satellite-based estimate of the direct and indirect aerosol climate forcing,
830   *Journal of Geophysical Research*, 113, D05204. doi:10.1029/2007JD008962, 2008.

Quaas, J., Ming, Y., Menon, S., Takemura, T., Wang, M., Penner, J., Gettelman, A., Lohmann, U., Bellouin, N., Boucher, O., Sayer, A., Thomas, G., McComiskey, A., Feingold, G., Hoose, C., Kristjánsson, J., Liu, X., Balkanski, Y., Donner, L., Ginoux, P., Stier, P., Grandey, B., Feichter, J., Sednev, I., Bauer, S., Koch, D., Grainger, R., Kirkevåg, A., Iversen, T., Seland, Ø.,

Easter, R., Ghan, S., Rasch, P., Morrison, H., Lamarque, J.-F., Iacono, M., Kinne, S., and Schulz, M.: Aerosol indirect effects—General circulation model intercomparison and evaluation with satellite data, *Atmospheric Chemistry and Physics*, 9, 8697–8717, doi:10.5194/acp-9-8697-2009, 2009.

Quaas, J., Arola, A., Cairns, B., Christensen, M., Deneke, H., Ekman, A. M. L., Feingold, G., Fridlind, A., Gryspeerdt, E., Hasekamp, O., Li, Z., Lipponen, A., Ma, P.-L., Mülmenstädt, J., Nenes, A., Penner, J. E., Rosenfeld, D., Schrödner, R., Sinclair, K., Sourdeval, O., Stier, P., Tesche, M., van Diedenhoven, B., and Wendisch, M.: Constraining the Twomey effect from satellite observations: issues and perspectives, *Atmospheric Chemistry and Physics*, 20, 15079–15099, doi:10.5194/acp-20-15079-2020, 2020.

Radke, L. F., Coakley, J. A., and King, M. D.: Direct and Remote Sensing Observations of the Effects of Ships on Clouds, *Science*, 246, 1146–1149, doi:10.1126/science.246.4934.1146, 1989.

Rasch, P. J., Xie, S., Ma, P.-L., Lin, W., Wang, H., Tang, Q., Burrows, S. M., Caldwell, P., Zhang, K., Easter, R. C., Cameron-Smith, P., Singh, B., Wan, H., Golaz, J.-C., Harrop, B. E., Roesler, E., Bacmeister, J., Larson, V. E., Evans, K. J., Qian, Y., Taylor, M., Leung, L. R., Zhang, Y., Brent, L., Branstetter, M., Hannay, C., Mahajan, S., Mametjanov, A., Neale, R., Richter, J. H., Yoon, J.-H., Zender, C. S., Bader, D., Flanner, M., Foucar, J. G., Jacob, R., Keen, N., Klein, S. A., Liu, X., Salinger, A., Shrivastava, M., and Yang, Y.: An Overview of the Atmospheric Component of the Energy Exascale Earth System Model, *Journal of Advances in Modeling Earth Systems*, 11, 2377–2411, doi:10.1029/2019MS001629, 2019.

Rémillard, J., Kollias, P., Luke, E., and Wood, R.: Marine Boundary Layer Cloud Observations in the Azores. *Journal of Climate*, 25, 21, 7381-7398, doi:10.1175/JCLI-D-11-00610.1, 2012.

Rémillard, J., and Tselioudis, G.: Cloud regime variability over the Azores and its application to climate model evaluation, *Journal of Climate*, 28, 9707–9720, doi:10.1175/JCLI-D-15-0066.1, 2015.

Saponaro, G., Sporre, M. K., Neubauer, D., Kokkola, H., Kolmonen, P., Sogacheva, L., Arola, A., de Leeuw, G., Karset, I. H. H., Laaksonen, A., and Lohmann, U.: Evaluation of aerosol and cloud properties in three climate models using MODIS observations and its corresponding COSP simulator, as well as their application in aerosol–cloud interactions, *Atmospheric Chemistry and Physics*, 20, 1607–1626, doi:10.5194/acp-20-1607-2020, 2020.

Sato, Y., Goto, D., Michibata, T., Suzuki, K., Takemura, T., Tomita, H., and Nakajima, T.: Aerosol effects on cloud water amounts were successfully simulated by a global cloud-system resolving model, *Nature Communications,* 9, 985, doi:10.1038/s41467-018-03379-6, 2018.

Shinozuka, Y., Clarke, A. D., Nenes, A., Jefferson, A., Wood, R., McNaughton, C. S., Strom, J., Tunved, P., Redemann, J., Thornhill, K. L., Moore, R. H., Lathem, T. L., Lin, J. J., and Yoon, Y. J.: The relationship between cloud condensation nuclei (CCN) concentration and light extinction of dried particles: indications of underlying aerosol processes and implications for satellite-based CCN estimates, *Atmospheric Chemistry and Physics*, 15, 7585-7604, doi:10.5194/acp-15-7585-2015, 2015.

Silva, S. J., Ma, P.-L., Hardin, J. C., and Rothenberg, D.: Physically regularized machine learning emulators of aerosol activation, *Geoscientific Model Development*, 14, 3067–3077, doi:10.5194/gmd-14-3067-2021, 2021.

Small, J. D., Chuang, P. Y., Feingold, G., and Jiang, H.: Can aerosol decrease cloud lifetime?, *Geophysical Research Letters*, 36, L16806, doi:10.1029/2009GL038888, 2009.

Smith, C. J., Kramer, R. J., Myhre, G., Alterskjær, K., Collins, W., Sima, A., Boucher, O., Dufresne, J.-L., Nabat, P., Michou, M., Yukimoto, S., Cole, J., Paynter, D., Shiogama, H., O'Connor, F. M., Robertson, E., Wiltshire, A., Andrews, T., Hannay, C., Miller, R., Nazarenko, L., Kirkevåg, A., Olivié, D., Fiedler, S., Lewinschal, A., Mackallah, C., Dix, M., Pincus, R., and Forster, P. M.: Effective radiative forcing and adjustments in CMIP6 models, *Atmospheric Chemistry and Physics*, 20, 9591–9618, doi:10.5194/acp-20-9591-2020, 2020.

Song, H., Zhang, Z., Ma, P., Ghan, S. J., and Wang, M.: An Evaluation of Marine Boundary Layer Cloud Property Simulations in the Community Atmosphere Model Using Satellite Observations: Conventional Subgrid Parameterization versus CLUBB, *Journal of Climate*, 31, 2299-2320, doi:10.1175/JCLI-D-17-0277.1, 2018.

Stephens, G. L., L'Ecuyer, T., Forbes, R., Gettelmen, A., Golaz, J.-C., Bodas-Salcedo, A., Suzuki, K., Gabriel, P., and Haynes, J.: Dreary state of precipitation in global models, *Journal of Geophysical Research*, 115, D24211, doi:10.1029/2010JD014532, 2010.

Stevens, B., and G. Feingold: Untangling aerosol effects on clouds and precipitation in a buffered system, *Nature,* 461, 607–613, doi:10.1038/nature08281, 2009.

Stier, P.: Limitations of passive remote sensing to constrain global cloud condensation nuclei, *Atmospheric Chemistry and Physics*, 16, 6595–6607, doi:10.5194/acp-16-6595-2016, 2016.

Suzuki, K., Nakajima, T. Y., and Stephens, G. L.: Particle growth and drop collection efficiency of warm clouds as inferred from joint CloudSat and MODIS observations, *Journal of the Atmospheric Sciences*, 67, 3019–3032, doi:10.1175/2010JAS3463.1, 2010.

Suzuki, K., Stephens, G. L., and Lebsock, M. D.: Aerosol effect on the warm rain formation process: Satellite observations and modeling, *Journal of Geophysical Research: Atmospheres*, 118, 170–184. doi:10.1002/jgrd.50043, 2013.

Suzuki, K., Stephens, G. L., Bodas-Salcedo, A., Wang, M., Golaz, J.-C., Yokohata, T., and Tsuyoshi, K.: Evaluation of the warm rain formation process in global models with satellite observations, *Journal of the Atmospheric Sciences*, 72, 3996–4014, doi:10.1175/JAS-D-14-0265.1, 2015.

Swales, D. J., Pincus, R., and Bodas-Salcedo, A.: The Cloud Feedback Model Intercomparison Project Observational Simulator Package: Version 2, *Geoscientific Model Development*, 11, 77–81, doi:10.5194/gmd-11-77-2018, 2018.

Tang, S., Fast, J. D., Zhang, K., Hardin, J. C., Varble, A. C., Shilling, J. E., Mei, F., Zawadowicz, M. A., and Ma, P.-L.: Earth System Model Aerosol-Cloud Diagnostics (ESMAC Diags) package version 1: assessing E3SM aerosol predictions using aircraft, ship, and surface measurements. *Geoscientific Model Development*, 15, 4055-4076. doi:10.5194/gmd-15-4055-2022, 2022.

Tang, S., Varble, A. C., Fast, J. D., Zhang, K., Wu, P., Dong, X., Mei, F., Pekour, M., Hardin, J. C., and Ma, P.-L.: Earth System Model Aerosol-Cloud Diagnostics (ESMAC Diags) package version 2: assessments of aerosols, clouds, and aerosol-cloud interactions through field campaign and long-term observations, *Geoscientific Model Development*, 2023, in review.

Terai, C. R., Pritchard, M. S., Blossey, P., and Bretherton, C. S.: The impact of resolving subkilometer processes on aerosol-cloud interactions of low-level clouds in global model simulations, *Journal of Advances in Modeling Earth Systems*, 12, e2020MS002274, doi:10.1029/2020MS002274, 2020.

Toll, V., Christensen, M., Gassó, S., and Bellouin, N.: Volcano and ship tracks indicate excessive aerosol-induced cloud water increases in a climate model, *Geophysical Research Letters*, 44, 12,492– 12,500, doi:10.1002/2017GL075280, 2017.

Turner, D. D., Clough, S. A., Liljegren, J. C., Clothiaux, E. E., Cady-Pereira, K. E., and Gaustad, K. L.: Retrieving liquid water path and precipitable water vapor from Atmospheric Radiation Measurement (ARM) microwave radiometers, *IEEE Transactions on Geoscience and Remote Sensing*, 45, 3680−3690, doi:10.1109/TGRS.2007.903703, 2007.

Twomey, S.: Pollution and the planetary albedo, *Atmospheric Environment*, 8, 1251–1256, doi:10.1016/0004-6981(74)90004-3, 1974.

Twomey, S.: The influence of pollution on the shortwave albedo of clouds, *Journal of the Atmospheric Sci*ences, 34, 1149–1152, doi:10.1175/1520-0469(1977)034<1149:TIOPOT>2.0.CO;2, 1977.

Xue, H., and Feingold, G.: Large-Eddy Simulations of Trade Wind Cumuli: Investigation of Aerosol Indirect Effects, *Journal of the Atmospheric Sciences*, 63, 1605-1622, doi:10.1175/JAS3706.1, 2006.

915 Wang, H., Easter, R. C., Zhang, R., Ma, P.-L., Singh, B., Zhang, K., Ganguly, D., Rasch, P. J., Burrows, S. M., Ghan, S. J., Lou, S., Qian, Y., Yang, Y., Feng, Y., Flanner, M., Leung, L. R., Liu, X., Shrivastava, M., Sun, J., Tang, Q., Xie, S., and Yoon, J.-H.: Aerosols in the E3SM Version 1: New developments and their impacts on radiative forcing, *Journal of Advances in Modeling Earth Systems*, 12, e2019MS001851, doi:10.1029/2019MS001851, 2020.

Wang, J., Wood, R., Jensen, M. P., Chiu, J. C., Liu, Y., Lamer, K., Desai, N., Giangrande, S. E., Knopf, D. A., Kollias, P., Laskin, A., Liu, X., Lu, C., Mechem, D., Mei, F., Starzec, M., Tomlinson, J., Wang, Y., Yum, S. S., Zheng, G., Aiken, A. C., Azevedo, E. B., Blanchard, Y., China, S., Dong, X., Gallo, F., Gao, S., Ghate, V. P., Glienke, S., Goldberger, L., Hardin, J. C., Kuang, C., Luke, E. P., Matthews, A. A., Miller, M. A., Moffet, R., Pekour, M., Schmid, B., Sedlacek, A. J., Shaw, R. A., Shilling, J. E., Sullivan, A., Suski, K., Veghte, D. P., Weber, R., Wyant, M., Yeom, J., Zawadowicz, M., and Zhang, Z.: Aerosol and Cloud Experiments in the Eastern North Atlantic (ACE-ENA), *Bulletin of the American Meteorological Society*, 103, E619-E641, doi:10.1175/BAMS-D-19-0220.1, 2022.

Witte, M. K., Yuan, T., Chuang, P. Y., Platnick, S., Meyer, K. G., Wind, G., and Jonsson, H. H.: MODIS retrievals of cloud effective radius in marine stratocumulus exhibit no significant bias, *Geophysical Research Letters*, 45, 10,656-10,664, doi:10.1029/2018GL079325, 2018.

Wood, R., Wyant, M., Bretherton, C. S., Rémillard, J., Kollias, P., Fletcher, J., Stemmler, J., Szoeke, S. D., Yuter, S., Miller, M., Mechem, D., Tselioudis, G., Chiu, J. C., Mann, J. A. L., O'Connor, E. J., Hogan, R. J., Dong, X., Miller, M., Ghate, V., Jefferson, A., Min, Q., Minnis, P., Palikonda, R., Albrecht, B., Luke, E., Hannay, C., and Lin, Y.: Clouds, aerosol, and precipitation in the marine boundary layer: An ARM mobile facility deployment, *Bulletin of the American Meteorological Society*, 96, 419–440, doi:10.1175/BAMS-D-13-00180.1, 2015.

Wu, P., Dong, X., Xi, B., Tian, J., and Ward, D. M.: Profiles of MBL cloud and drizzle microphysical properties retrieved from ground-based observations and validated by aircraft in situ measurements over the Azores, *Journal of Geophysical Research: Atmospheres*, 125, e2019JD032205, doi:10.1029/2019JD032205, 2020.

Wu, P., Dong, X., and Xi, B.: A Climatology of Marine Boundary Layer Cloud and Drizzle Properties Derived from Ground-Based Observations over the Azores, *Journal of Climate*, 33, 10133-10148, doi:10.1175/JCLI-D-20-0272.1, 2020.

Zhang, G. J., and McFarlane, N. A.: Sensitivity of climate simulations to the parameterization of cumulus convection in the Canadian climate center general-circulation model, *Atmosphere-Ocean*, 33, 407–446, doi:10.1080/07055900.1995.9649539, 1995.

Zhang, J., Zhou, X., Goren, T., and Feingold, G.: Albedo susceptibility of northeastern Pacific stratocumulus: the role of covarying meteorological conditions. *Atmospheric Chemistry and Physics*, 22, 861-880, doi:10.5194/acp-22-861-2022, 2022.

Zhang, K., Zhang, W., Wan, H., Rasch, P. J., Ghan, S. J., Easter, R. C., Shi, X., Wang, Y., Wang, H., Ma, P.-L., Zhang, S., Sun, J., Burrows, S. M., Shrivastava, M., Singh, B., Qian, Y., Liu, X., Golaz, J.-C., Tang, Q., Zheng, X., Xie, S., Lin, W., Feng, Y., Wang, M., Yoon, J.-H., and Leung, L. R.: Effective radiative forcing of anthropogenic aerosols in E3SM version 1: historical changes, causality, decomposition, and parameterization sensitivities, *Atmospheric Chemistry and Physics*, 22, 9129–9160, doi:10.5194/acp-22-9129-2022, 2022.

Zheng, G., Y. Wang, A. C. Aiken, F. Gallo, M. P. Jensen, P. Kollias, C. Kuang, E. Luke, S. Springston, J. Uin, R. Wood, and Wang, J.: Marine boundary layer aerosol in the eastern North Atlantic: seasonal variations and key controlling processes, *Atmospheric Chemistry and Physic*s, 18, 17615-17635, doi:10.5194/acp-18.17615-2018, 2018.

Dataset	Short Name	Color	Variables
5-min surface retrievals	Obs Sfc 5min	Black Line	Cloud fraction (CF), Cloud effective albedo, Cloud liquid water path (LWP), Cloud layer mean drop number concentration (N_d), Cloud optical depth (COD), Cloud layer mean drop effective radius (R_{eff}), Cloud base and top heights (CBH, CTH), Cloud base and top temperatures (CBT, CTT), Surface cloud condensation nuclei (CCN), Estimated inversion strength (EIS), 700-hPa relative humidity (RH), Cloud adiabaticity, Surface rain rate
60-min surface retrievals	Obs Sfc 60min	Green Line	
4-km satellite retrievals	Obs TOA 4km	Red Line	CF (1-deg only), TOA albedo, Cloud LWP, Cloud layer mean N_d, COD, Near cloud top drop R_{eff}, CTT, solar zenith angle (SZA), ice water path (IWP)
1-deg satellite retrievals	Obs TOA 1deg	Orange Line	
1-deg E3SM output	E3SM	Blue Fill	CF, Cloud effective albedo, TOA albedo, Cloud LWP, Cloud layer mean N_d, CBH, CTH, CBT, CTT, Surface CCN, EIS, 700-hPa RH, Cloud adiabaticity, Surface rain rate, SZA
1-deg E3SM COSP retrievals	E3SM TOA	Green Fill	CF, Cloud LWP, Cloud layer mean N_d, COD, Near cloud top drop R_{eff}
1-deg E3SM surface retrievals	E3SM Sfc	Orange Fill	Cloud layer mean N_d

Table 1: Each dataset analyzed including its short name and color used in figures along with variables provided by each dataset that are used in comparisons.

Dataset	Warm, Liquid Cloud Samples	Warm Liquid CF (%)	% Time (CF > 95%)	Samples (CF > 95%)	Samples (CF > 95%, SZA ≤ 65°)	Samples (CF > 95%, SZA ≤ 65°, COD > 4, LWP > 20 g m⁻²)	Samples (CF > 95%, SZA ≤ 65°, COD > 4, LWP > 20 g m⁻², CCN_{fit} > 0, θ_{diff} < 2°C)
Obs Sfc 5min	190,565	70	41	79,609	22,247	12,643	3,907
Obs Sfc 60min	19,594	55	17	3,243	852	649	197
Obs TOA 4km	17,310	75	36	6,301	1,716	1,381	328
Obs TOA 1deg	19,891	59	20	3,960	1,017	990	224
E3SM	28,224	54	31	8,710	2,403	1,939	1,459
E3SM Sfc	28,224	54	31	8,710	2,403	1,941	1,303
E3SM TOA	21,688	48	23	6,557	1,871	1,697	1,459

955 **Table 2:** Warm (T ≥ 0°C) liquid cloud fraction and sampling as filters are applied for each dataset for only situations with no sub-freezing clouds detected using both IWP and cloud top temperature constraints. For E3SM, IWP is derived from the COSP simulator due to an abundance of very low ice concentrations in the upper troposphere that would remove too many warm, liquid cloud samples. Sensitivity tests indicate that these low ice concentrations that COSP does not detect have little impact on albedo. For non-TOA datasets, multi-layer warm cloud situations have additionally been removed. The Obs TOA 4-km CF is derived from measurements over a 0.5° region.

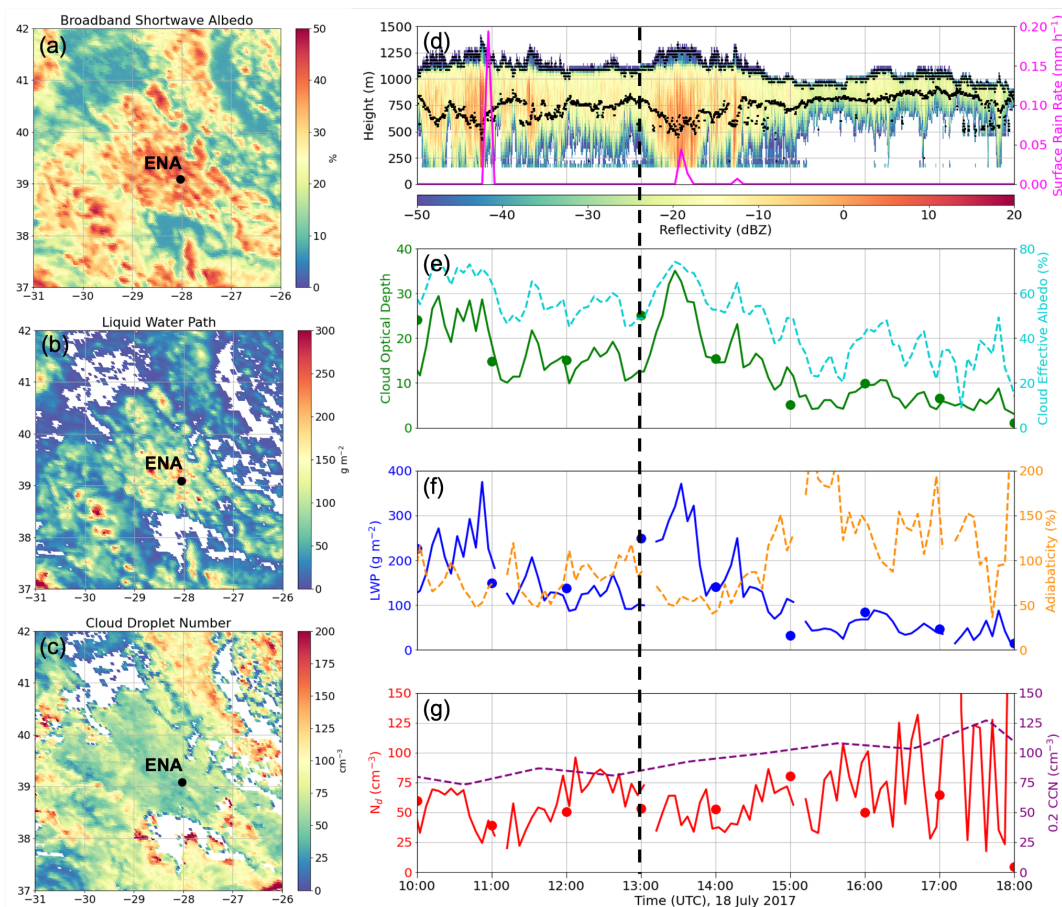



Figure 1: An example single layer liquid cloud case at the ARM ENA site showing snapshots of 4-km satellite-retrieved (a) TOA albedo, (b) LWP, and (c) N_d at 1300 UTC 18 July 2017. An 8-hour period of the same event is shown with surface observations of (d) Ka-band reflectivity and lidar-retrieved cloud base with 5-min optical rain gauge surface rain rate, (e) 5-minute surface-retrieved COD (solid) and cloud effective broadband shortwave albedo (dashed), (f) 5-min LWP (solid) and cloud adiabaticity (dashed), and (g) 5-minute N_d (solid) and hourly 0.2% CCN concentration (dashed). Satellite-retrieved COD, LWP, and N_d values are plotted on the time series with circles. The dashed vertical black line indicates the time of the satellite snapshots. The surface site is noted in the satellite images.

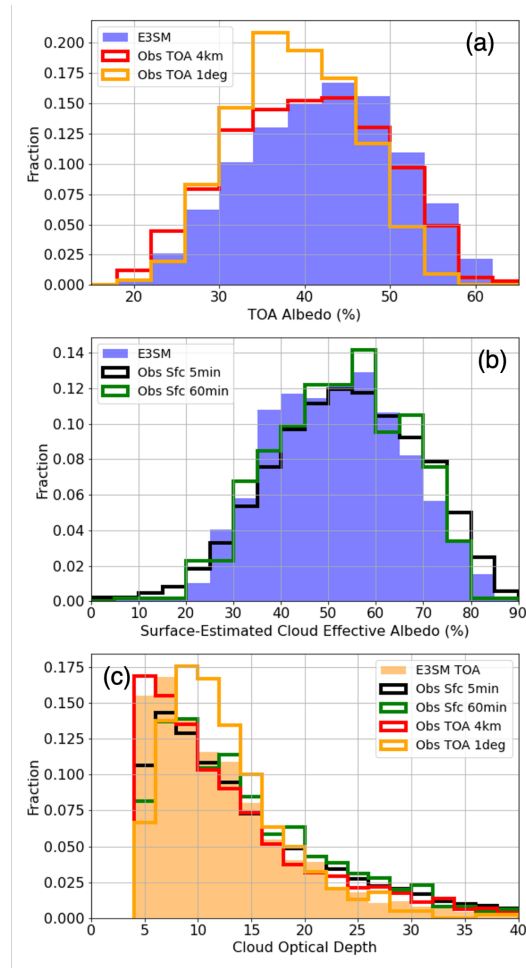



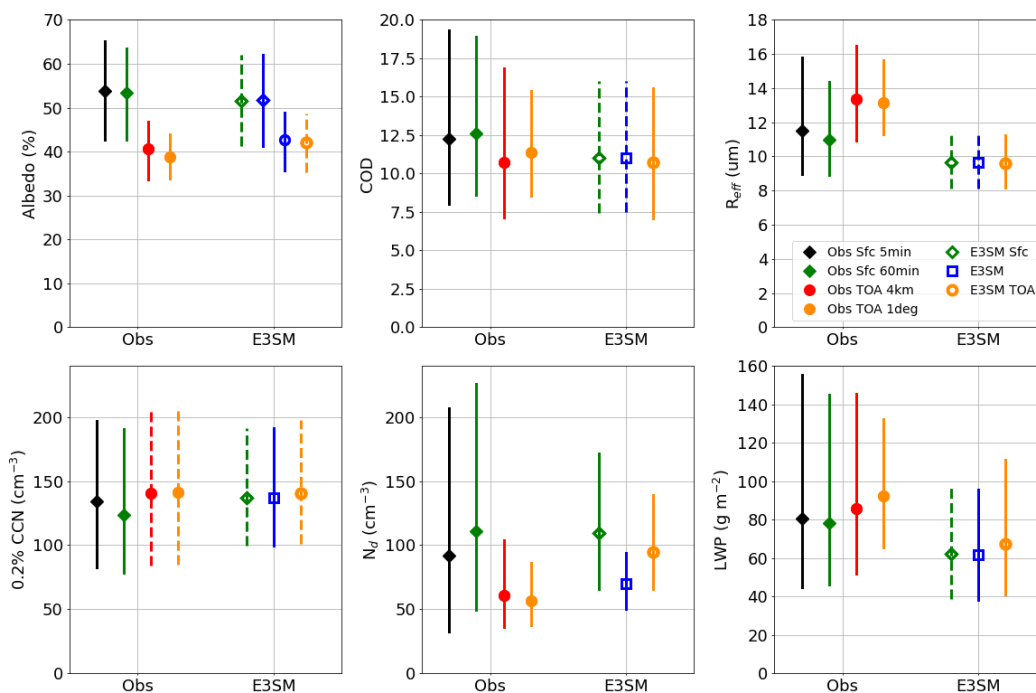
Figure 2: Probability distributions of observed and simulated (a) TOA albedo, (b) surface-estimated cloud effective albedo, and (c) cloud optical depth. Datasets are excluded when they are similar to another already shown due to being derived from the shown dataset.

970

Figure 3: Median (symbol) and interquartile range (vertical bar) values of key variables. Dashed vertical bars indicate datasets sampled from solid vertical bar datasets including Obs TOA from Obs Sfc CCN values, E3SM and E3SM Sfc COD and R_{eff} from E3SM TOA, E3SM Sfc and TOA albedos and CCN from E3SM, and E3SM Sfc LWP from E3SM.

975

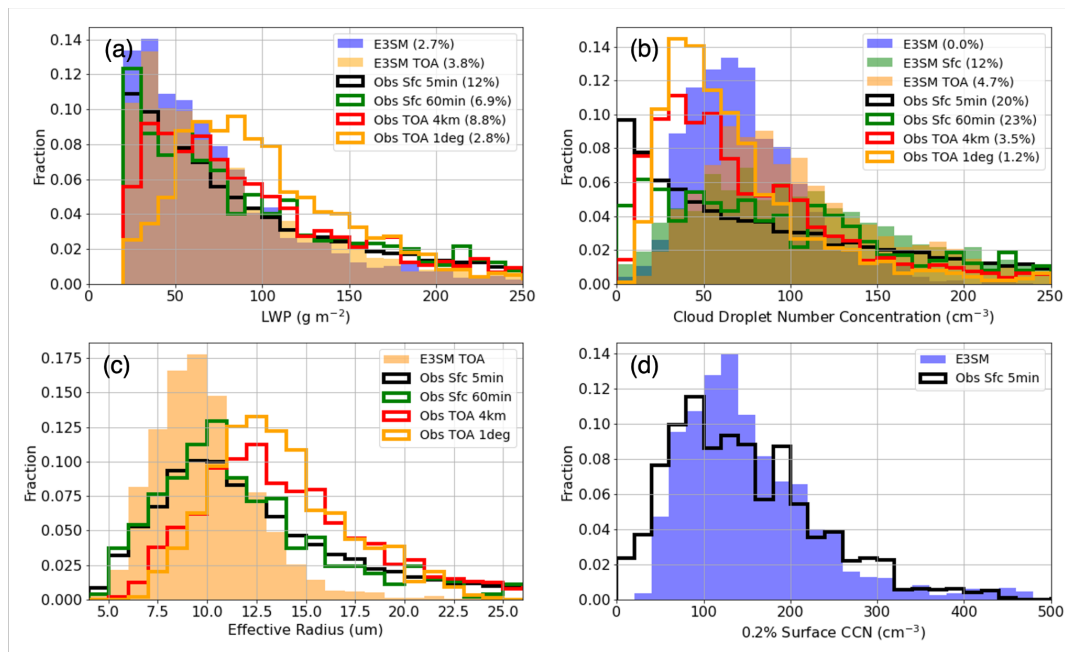



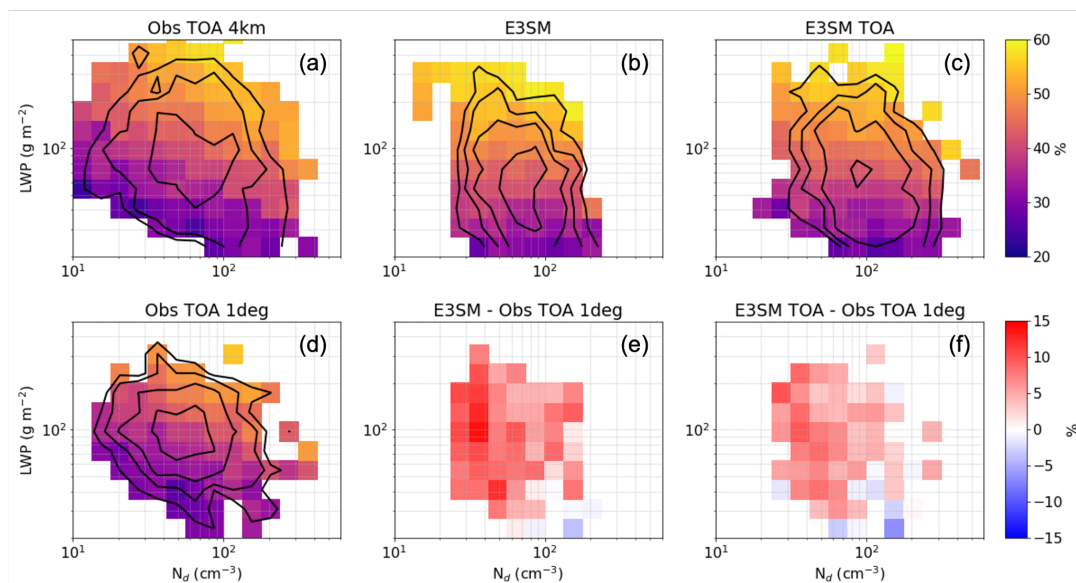
Figure 4: Probability distributions of observed and simulated (a) cloud LWP, (b) cloud layer mean N_d, (c) cloud droplet R_{eff} and (d) 0.2% surface CCN. Percentages in the legends indicate how many samples exceed the range of the x-axis. Datasets are excluded when they are similar to another already shown due to being derived from the shown dataset.

980

Figure 5: Median TOA albedo vs. N_d and LWP for (a) Obs TOA 4km, (b) E3SM, (c) E3SM TOA, and (d) Obs TOA 1deg. Absolute differences are also shown between (e) E3SM and Obs TOA 1deg, and (f) E3SM TOA and Obs TOA 1deg. Black contours indicate sample size thresholds of 0.4, 0.8, 1.6, and 3.2%, respectively.

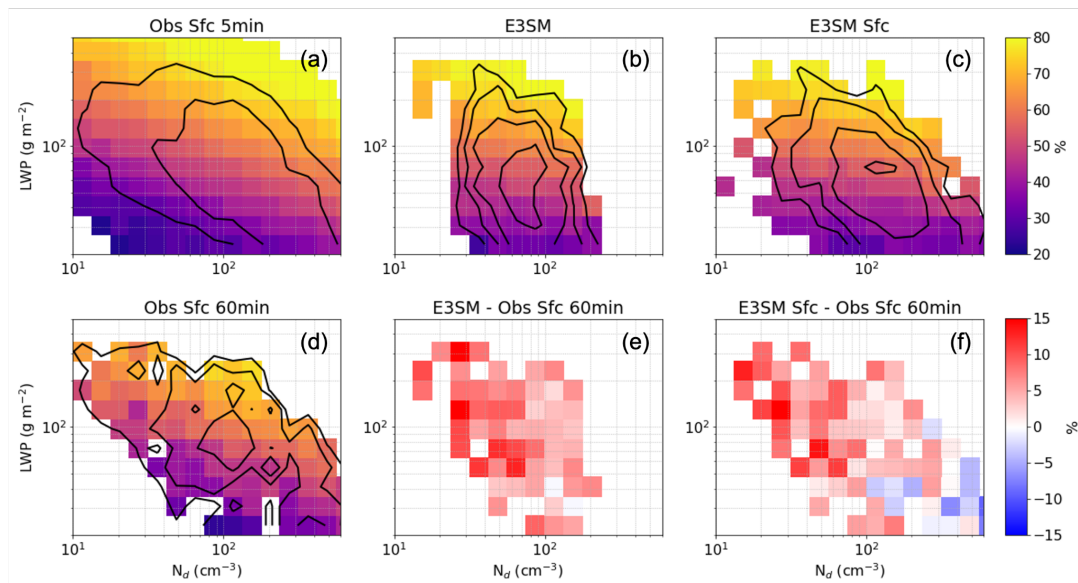



Figure 6: Median cloud effective albedo vs. N_d and LWP for (a) Obs Sfc 5min, (b) E3SM, (c) E3SM Sfc, and (d) Obs Sfc 60min. Absolute differences are also shown between (e) E3SM and Obs Sfc 60min, and (f) E3SM Sfc and Obs Sfc 60min. Black contours indicate sample size thresholds of 0.4, 0.8, 1.6, and 3.2%, respectively.

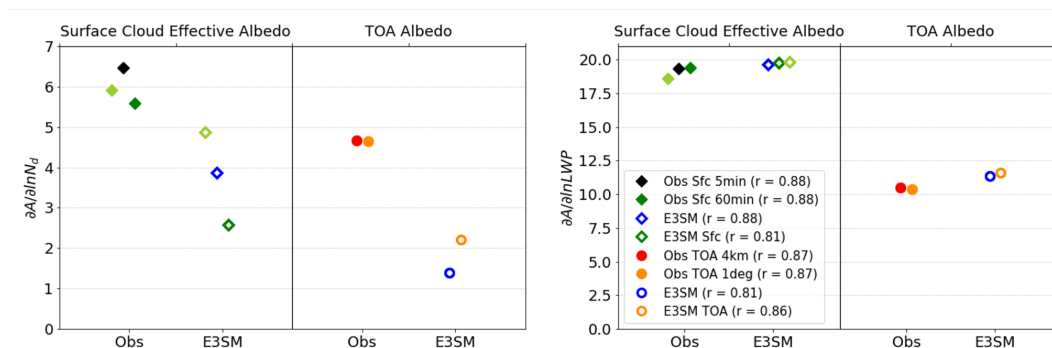



Figure 7: Surface cloud effective albedo and TOA albedo as functions of (left) $\ln N_d$ and (right) lnLWP for observational and E3SM datasets. Estimates are obtained from ordinary least squares multiple linear regression with Pearson correlation coefficients shown in the legend. Light green symbols represent Obs Sfc 60min and E3SM Sfc datasets with a recomputed N_d that assumes 80% adiabaticity like the TOA retrievals.

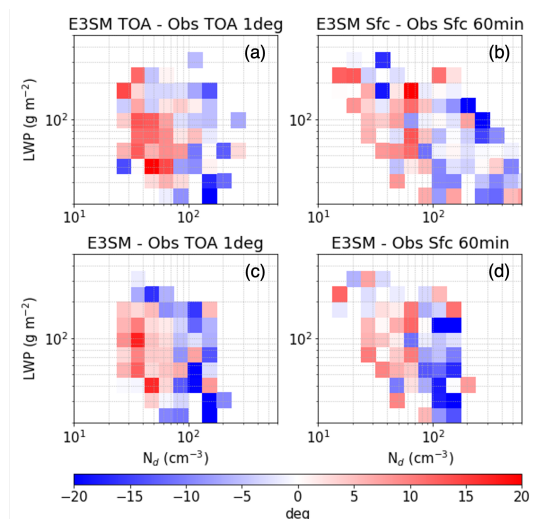



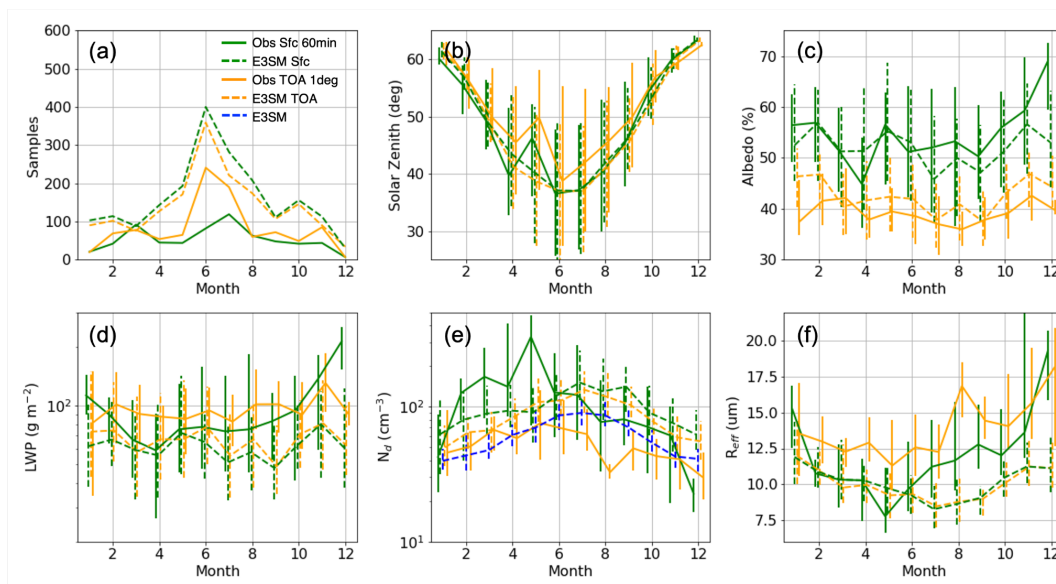
Figure 8: Median absolute SZA differences between (a) E3SM TOA and Obs TOA 1deg, (b) E3SM Sfc and Obs Sfc 60min, (c) E3SM and Obs TOA 1deg, and (d) E3SM and Obs Sfc 60min as functions of N_d and LWP.

995

Figure 9: Seasonal cycles of (a) number of samples and interquartile ranges (vertical bars) connected by medians for (b) SZA < 65°, (c) TOA or cloud effective albedo, (d) cloud LWP, (e) layer-mean N_d, and (f) R_{eff}.

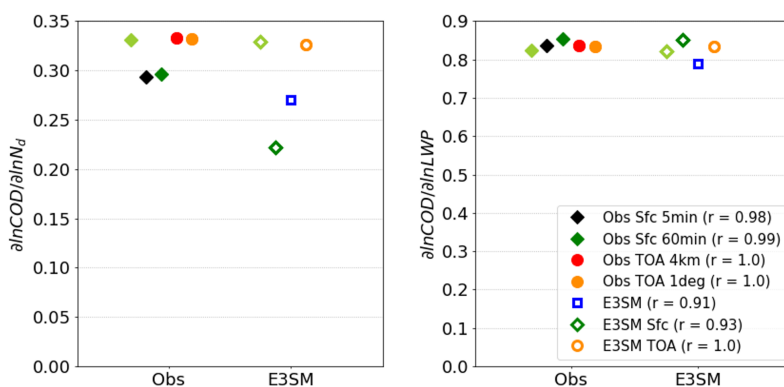



Figure 10: lnCOD as a function of (left) lnN_d and (right) lnLWP for observational and E3SM datasets. Estimates are obtained from ordinary least squares multiple linear regression with Pearson correlation coefficients shown in the legend. Light green symbols represent Obs Sfc 60min and E3SM Sfc datasets with a recomputed N_d that assumes 80% adiabaticity like the TOA retrievals.

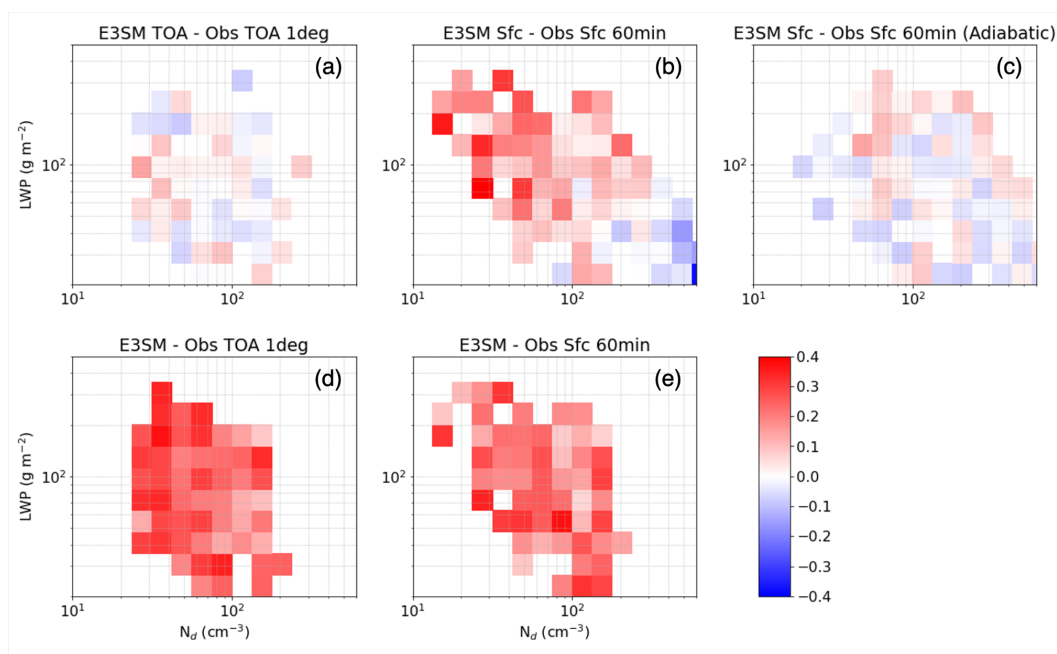



Figure 11: Median absolute lnCOD differences between (a) E3SM TOA and Obs TOA 1deg, (b) E3SM Sfc and Obs Sfc 60min, (c) E3SM Sfc and Obs Sfc 60min using N_d retrievals that assume 80% adiabaticity, (d) E3SM and Obs Sfc 60min, and (e) E3SM and Obs TOA 1deg as functions of N_d and LWP.

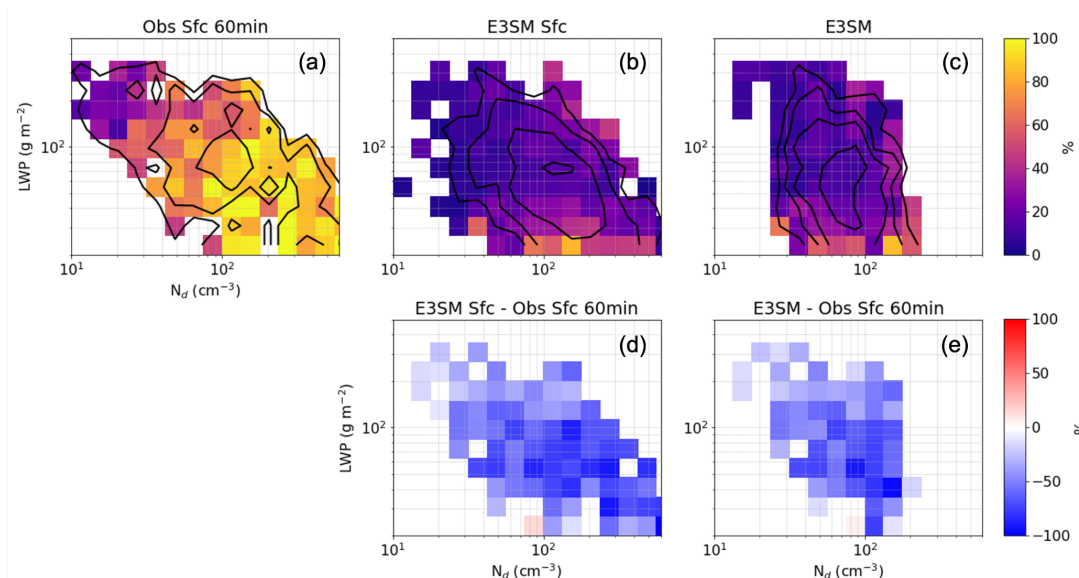



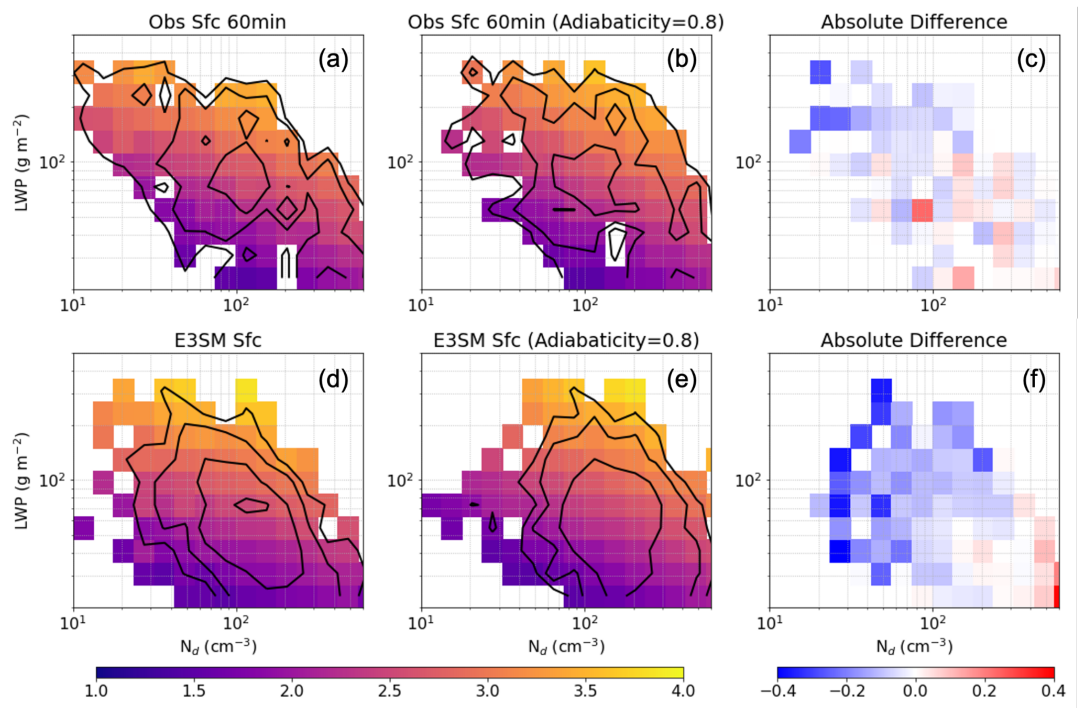
Figure 12: Median cloud adiabaticity as a function of LWP and N_d for (a) Obs Sfc 60min, (b) E3SM Sfc, and (c) E3SM. (d) Absolute differences between E3SM Sfc and Obs Sfc 60min. (e) Absolute differences between E3SM and Obs Sfc 60min. Black contours indicate sample size thresholds of 0.4, 0.8, 1.6, and 3.2%, respectively.

1010

Figure 13: Median lnCOD as a function of LWP and N_d for (a) Obs Sfc 60min, (b) Obs Sfc 60min using Nd derived assuming 80% adiabaticity, (c) Obs Sfc 60min (Adiabaticity=0.8) minus Obs Sfc 60min, (d) E3SM Sfc, (e) E3SM Sfc assuming 80% adiabaticity, and (f) E3SM Sfc (Adiabaticity=0.8) minus E3SM Sfc. Black contours indicate sample size thresholds of 0.4, 0.8, 1.6, and 3.2%, respectively.

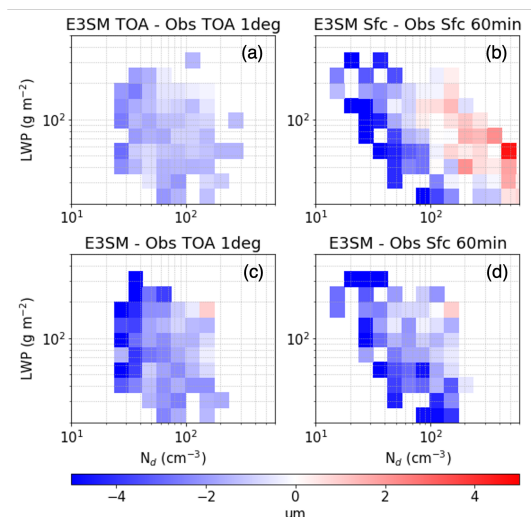



Figure 14: Median absolute R_{eff} differences between (a) E3SM TOA and Obs TOA 1deg, (b) E3SM Sfc and Obs Sfc 60min, (c) E3SM and Obs TOA 1deg, and (d) E3SM and Obs Sfc 60min as functions of N_d and LWP.

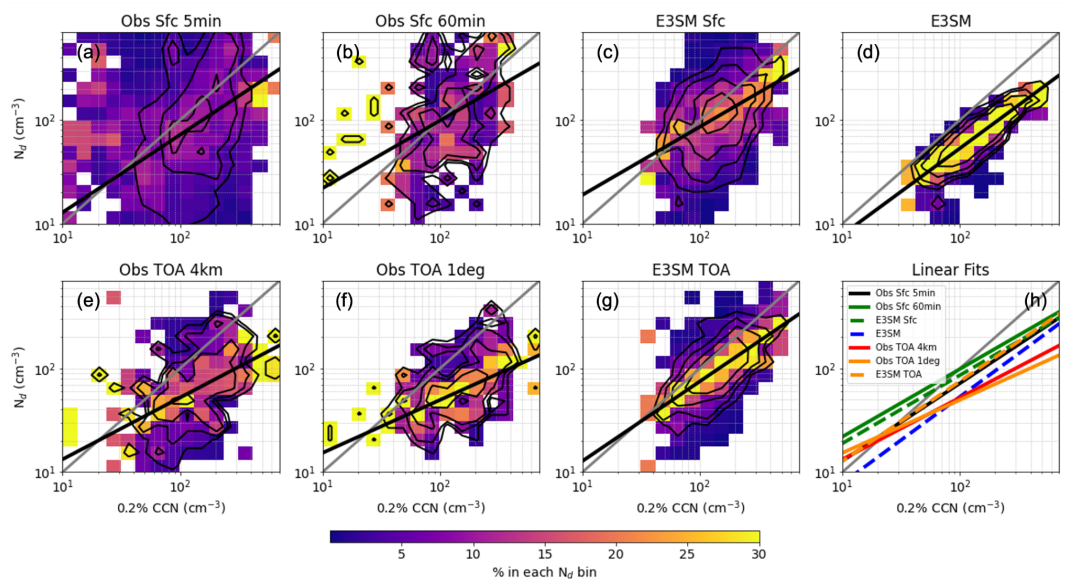



Figure 15: Joint distributions of N_d and 0.2% CCN number concentration normalized by CCN bin for (a) Obs Sfc 5min, (b) Obs Sfc 60min, (c) E3SM Sfc, (d) E3SM, (e) Obs TOA 4km, (f) Obs TOA 1deg, and (g) E3SM TOA. Thin black contours indicate sample size thresholds of 0.4, 0.8, 1.6, and 3.2%, respectively. Theil-Sen linear fits are overplotted in thick solid black with the 1:1 line in gray. Linear fits for each dataset are overplotted on one another in (h).

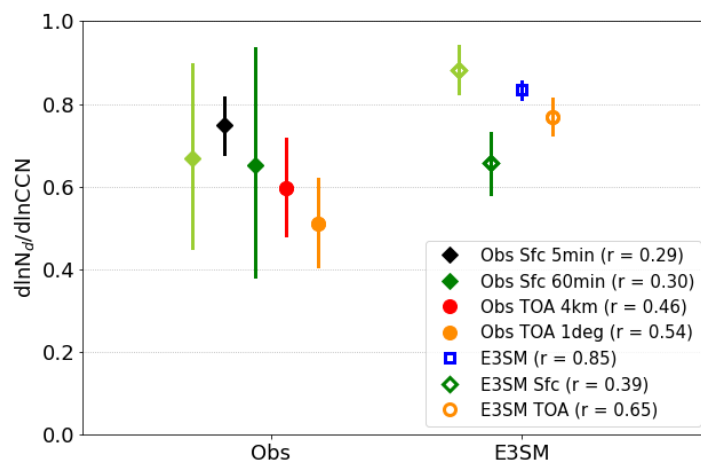



Figure 16: lnN$_d$ as a function of lnCCN for observational and E3SM datasets. Estimates are obtained from Theil-Sen robust linear regression with 95% confidence intervals. Pearson correlation coefficients are shown in the legend. Light green symbols represent Obs Sfc 60min and E3SM Sfc datasets with a recomputed N$_d$ that assumes 80% adiabaticity like the TOA retrievals.

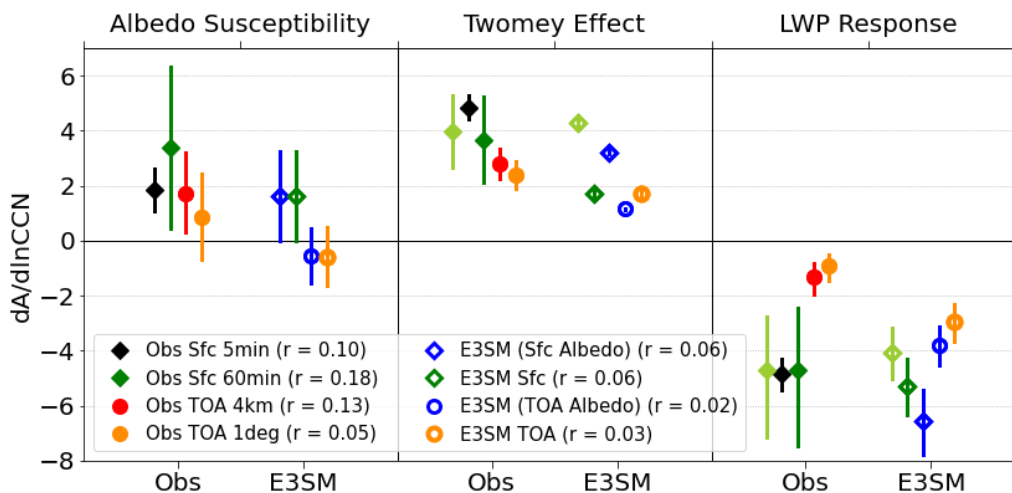



Figure 17: (left) Cloud albedo susceptibility to CCN concentration with (middle) Twomey effect and (right) LWP response terms for observational and E3SM datasets with 95% confidence intervals. Pearson correlation coefficients for albedo regressed on lnCCN are shown in the legend. Light green symbols represent Obs Sfc 60min and E3SM Sfc datasets with a recomputed N$_d$ that assumes 80% adiabaticity like the TOA retrievals.

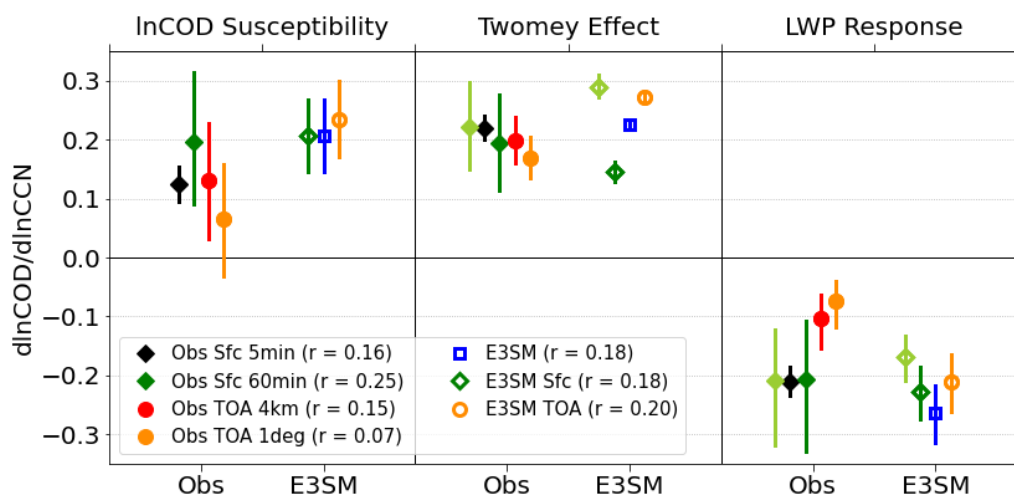



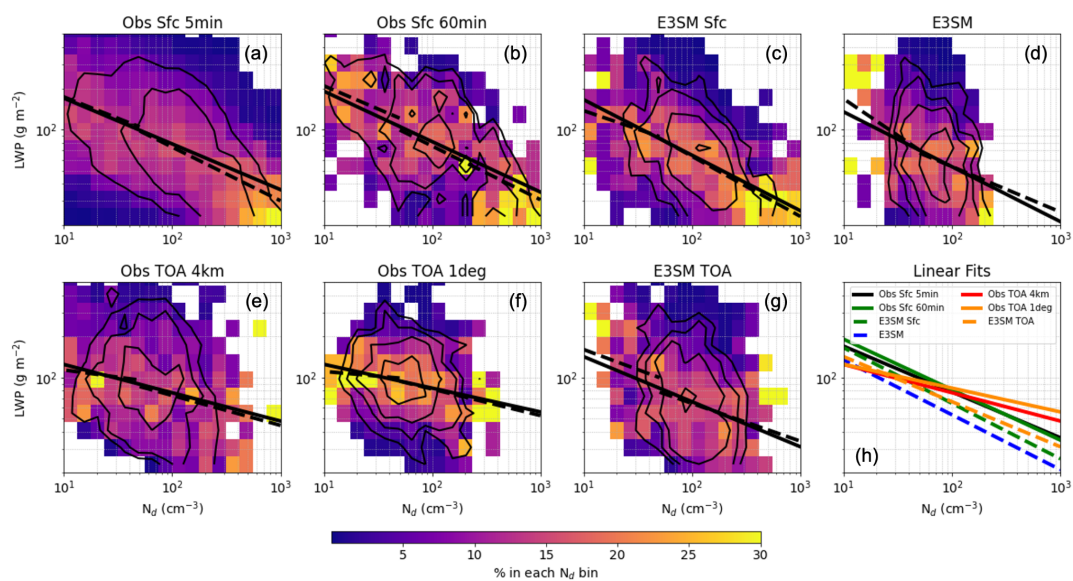
Figure 18: (left) lnCOD susceptibility to CCN concentration with (middle) Twomey effect and (right) LWP response terms for observational and E3SM datasets with 95% confidence intervals. Pearson correlation coefficients for lnCOD regressed on lnCCN are shown in the legend. Light green symbols represent Obs Sfc 60min and E3SM Sfc datasets with a recomputed N_d that assumes 80% adiabaticity like the TOA retrievals.

1035

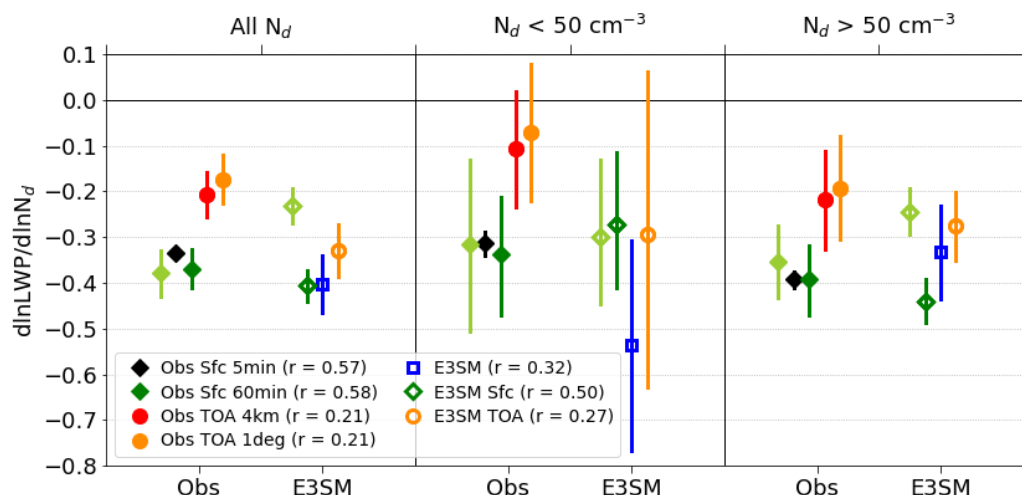
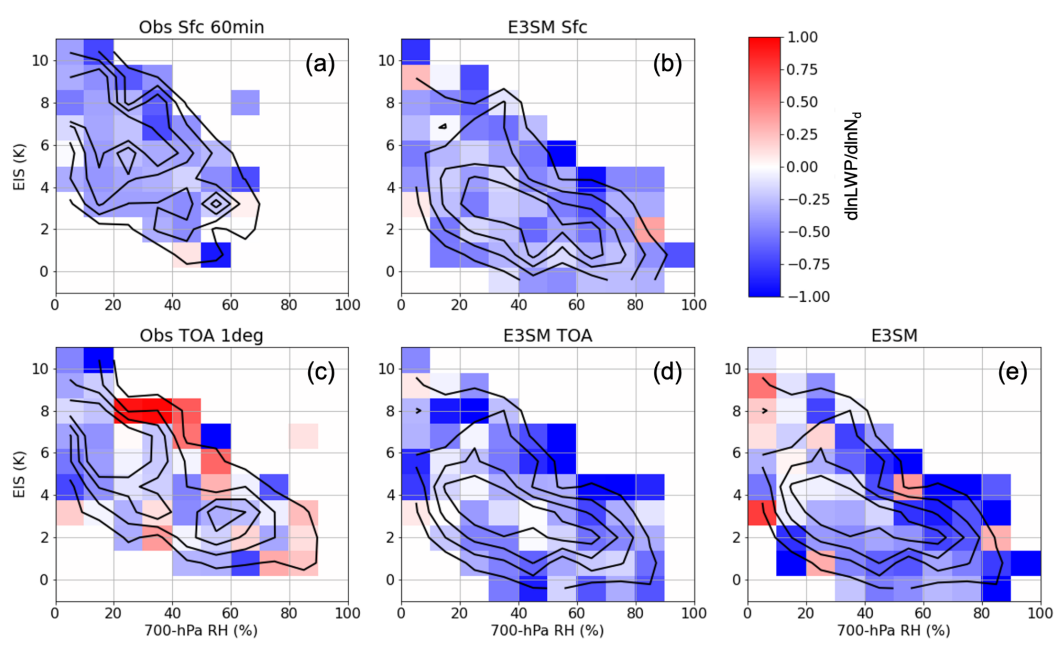
Figure 19: Joint distributions of LWP and N_d normalized by N_d bin for (a) Obs Sfc 5min, (b) Obs Sfc 60min, (c) E3SM Sfc, (d) E3SM, (e) Obs TOA 4km, (f) Obs TOA 1deg, and (g) E3SM TOA. Thin black contours indicate sample size thresholds of 0.4, 0.8, 1.6, and 3.2%, respectively. Single Theil-Sen linear fits are overplotted in thick solid black, and piecewise fits are overplotted in thick dashed black. Linear fits for each dataset are overplotted on one another in (h).

1040

Figure 20: lnLWP as a function of lnN$_d$ for all N$_d$ values, N$_d$ < 50 cm^{-3}, and N$_d$ > 50 cm^{-3} for observational and E3SM datasets. Estimates and 95% confidence intervals are obtained from Theil-Sen robust linear regressions with Pearson correlation coefficients shown in the legend. Light green symbols represent Obs Sfc 60min and E3SM Sfc datasets with a recomputed N$_d$ that assumes 80% adiabaticity like the TOA retrievals.

1045   **Figure 21:** Median LWP response ($\frac{dlnLWP}{dlnN_d}$) as a function of EIS and 700-hPa RH for (a) Obs Sfc 60min, (b) E3SM Sfc, (c) Obs TOA 1deg, (d) E3SM TOA, and (e) E3SM. Black contours indicate sample size thresholds of 0.4, 0.8, 1.6, and 3.2%, respectively.

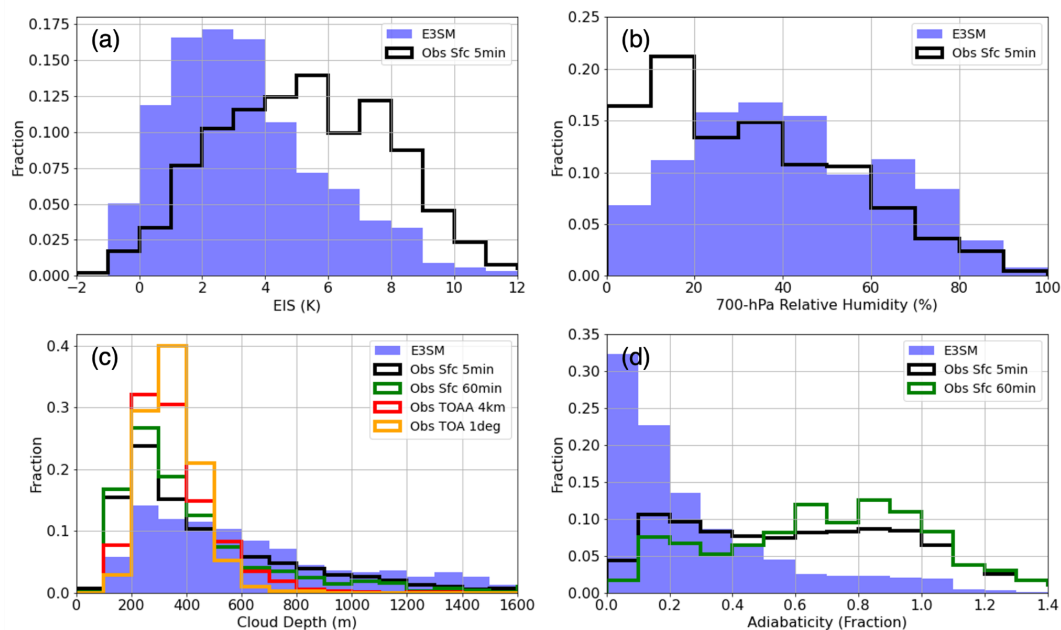



Figure 22: (a) Estimated inversion strength (EIS), (b) 700-hPa RH, (c) cloud depth, and (d) cloud adiabaticity PDFs for E3SM and observations. Observations are derived from interpolated radiosondes.

1050

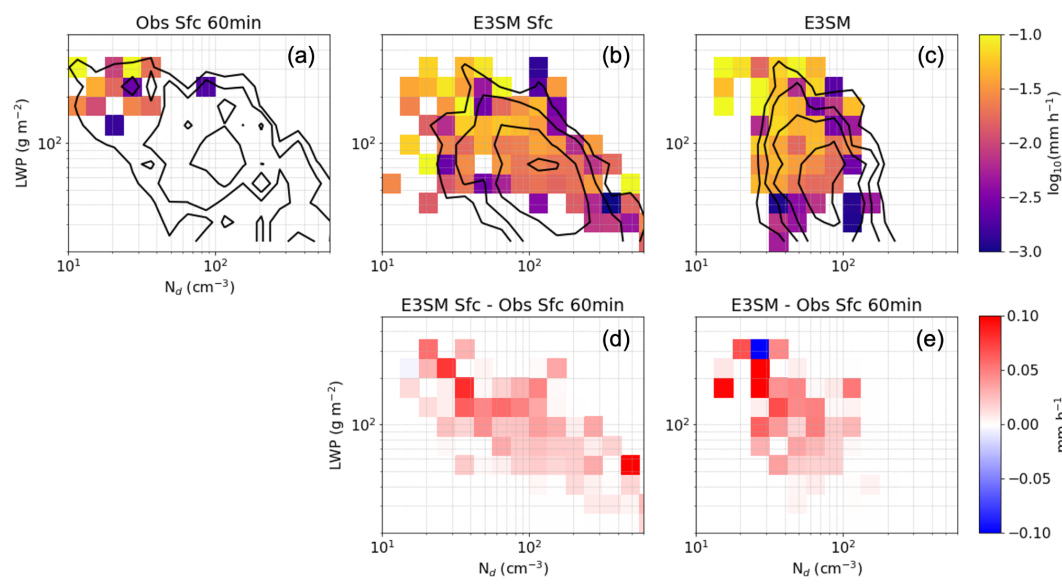



Figure 23: Median surface hourly rain rate as a function of LWP and N_d for (a) Obs Sfc 60min disdrometer retrievals, (b) E3SM Sfc, and (c) E3SM. (d) Absolute differences between E3SM Sfc and Obs Sfc 60min. (e) Absolute differences between E3SM and Obs Sfc 60min. Only rain rates exceeding 0.001 mm h[-1] are included in statistics due to observational sensitivity limitations. Black contours indicate sample size thresholds of 0.4, 0.8, 1.6, and 3.2%, respectively.