Machine Learning Reveals Strong Grid-Scale Dependence in the Satellite N_d–LWP Relationship

Matthew W. Christensen[1], Andrew Geiss[1], Adam C. Varble[1], and Po-Lun Ma[1]

[1]Atmospheric, Climate, and Earth Sciences Division, Pacific Northwest National Laboratory, Richland, WA 99354, Washington, USA

Correspondence: Matthew Christensen (matt.christensen@pnnl.gov)

Abstract. The relationship between cloud droplet number concentration (N_d) and liquid water path (LWP) is highly uncertain yet crucial for determining the impact of aerosol-cloud interactions (ACI) on Earth's radiation budget. The N_d-LWP relationship is examined using a machine learning (ML) random forest model applied to five years of satellite data at grid resolutions ranging from 10° to 0.05° in 12 distinct regions. In the subtropics, the shape of the N_d-LWP relationship switches from an *inverted-V* at 1° grid-resolution to an "M" shape at 0.1° resolution with decreased $\frac{d\ln LWP}{d\ln N_d}$ sensitivity. Tropical and midlatitude regions generally show a more positive sensitivity. Cloud sampling and filtering also influence this slope, wherein the exclusion of thin clouds, as commonly performed to reduce retrieval uncertainty, leads to strongly negative sensitivity across all regions. Precipitation is primarily responsible for driving the strength of the sensitivity, with strong positive slopes in raining clouds and negative and/or neutral responses found in non-raining clouds. A new method to compute radiative forcing from the ML model shows a robust Twomey radiative forcing across all regions and grid resolutions. However, LWP and cloud fraction ~~rapid adjustments~~ adjustments to the radiative forcing, which are $\sim 50\%$ or smaller than the Twomey effect, decrease to negligible values with higher spatial resolution data. As Earth system models move toward higher spatial resolutions in the future, evaluating the LWP and CF adjustment contributions to the radiative forcing budget at these finer resolutions will be essential for evaluation and model development.

1 Introduction

Clouds are highly reflective and significantly cool the Earth, helping to keep the planet habitable. The more abundant they are, and the more water they contain, the greater the amount of solar radiation they reflect (Stephens, 1978). Increased aerosol concentrations can elevate planetary albedo by enhancing cloud reflectance through an increase in cloud droplet number concentration (N_d), cloud liquid water path (LWP), and cloud fraction (CF). However, if LWP and CF decrease as aerosol loading increases, more sunlight will be absorbed by the Earth, leading to a warming radiative effect. While the effective radiative forcing from aerosol-cloud interactions (ACI) (ERF_{aci}; Gryspeerdt et al., 2019) has a net cooling effect globally, estimates remain highly uncertain (Bellouin et al., 2020) due to limited understanding, large retrieval uncertainties, and challenges quantifying and attributing causality from the non-linear relationship between N_d and LWP.

The N_d–LWP relationship is non-linear and typically appears as an inverted-V shape when displayed as a column-normalized 2D histogram representing the conditional distribution $P(\text{LWP} \mid N_d)$ ~~Gryspeerdt et al. (2019)~~(Gryspeerdt et al., 2019). This shape manifests from a rise in LWP as N_d increases below a threshold of approximately 30 cm^{-3}, followed by a decrease for higher N_d. Examples of this relationship can be seen in satellite observations (Gryspeerdt et al., 2019), large eddy simulations (LES; Hoffmann et al., 2020), and in the Energy Exascale Earth System Model (E3SM; Christensen et al., 2023). The prevailing hypothesis governing this relationship provides separate physical explanations for the behavior in both the high and low N_d regimes. The ascending branch (at low N_d) results from the suppression of precipitation caused by more, but smaller, cloud droplets allowing LWP to increase (Albrecht, 1989). In the descending branch, entrainment of dry air into non-raining clouds tends to decrease LWP (Ackerman et al., 2004) by increasing N_d further. Dry air entrainment affects both branches, but when the clouds are non-precipitating, the LWP is subject to larger decreases (Chen et al., 2014).

Although this relationship has been used to infer causality in ACI, several factors complicate a direct causal interpretation. Aerosol effects can make the link between aerosol concentration and N_d nonlinear due to processes such as precipitation suppression and evaporation-entrainment feedbacks (Albrecht, 1989; Ackerman et al., 2004). Satellite retrieval errors and sampling limitations—such as the exclusion of thin clouds due to uncertainties or biases introduced by retrieval inaccuracies—further affect analyses (Grosvenor et al., 2018). Feedbacks, including wet scavenging and entrainment, impact N_d retrievals, while meteorological confounders, such as dry air intrusion coinciding with elevated aerosol levels, can reduce cloud water paths; each of these factors are detailed further in (Gryspeerdt et al., 2019). Recent assessments show that the propagation of spatial variability and errors in satellite retrievals can lead to the misinterpretation of positive LWP adjustments as negative in subtropical clouds, thereby leading to an underestimate of ERF_{aci} (Arola et al., 2022). The extent to which this relationship is controlled by these or a combination of these factors and whether similar biases or drivers are applicable outside of the subtropics remains largely unknown.

Current understanding of the N_d–LWP relationship derived from satellite observations at global scales is largely based on coarse spatial resolution data (e.g., $4° \times 4°$ or $1° \times 1°$ grids). While Arola et al. (2022) examined finer spatial resolutions ($0.25° \times 0.25°$) of the N_d–LWP relationship it was over a limited area of the North Pacific, leaving an open question as to whether the results hold outside of the subtropics and over larger spatial extents. General circulation models (GCMs) with regionally refined meshes, such as the Simple Cloud-Resolving E3SM Atmosphere Model (SCREAM), have started generating ACI statistics at increasingly finer resolutions, down to 3 km (Caldwell et al., 2021). This raises pertinent questions about how the statistics of ACI change with varying spatial-scale resolutions, and whether meteorological regimes influence the 2D histograms of the N_d-LWP relationship. We ~~have~~ generated a series of collocated global datasets at ~~a series of spatial resolutions from~~ progressively higher grid-scale spatial resolutions (10°~~× 10~~, 5°~~down to 0.05~~, 1°~~×~~, 0.5°, 0.1°, and 0.05°) and applied a ~~Machine Learning~~ machine learning (ML) model to extract non-linear behavior in the N_d–LWP relationship. From these new datasets and tools, we are able to answer the following questions:

- How does the structure of the N_d–LWP relationship change as the spatial grid resolution increases to finer scales?
- How do subtropical, tropical, and midlatitude regions differ in their N_d–LWP relationships?

- How does satellite filtering and sampling clouds with different characteristics influence the N_d–LWP relationship?
- What are the primary meteorological drivers shaping the N_d–LWP relationship?
- What is the impact of changing spatial resolution on the estimated radiative effects of ACI?

In this study, we aim to address these questions by examining how grid-scale resolution, regional differences, and meteorological factors influence the N_d–LWP relationship, and by applying a random forest ML model to our data set to enhance our understanding and prediction of these complex interactions and changes in radiative forcing. The data sets are described in section 2, methods involving statistical sampling and ML methods are described in section 3, results are described in section 4, and conclusions in section 5.

2 Data

2.1 Satellite Observations

The MODIS Collection 6.1 cloud product is derived from the observations acquired from the Aqua satellite, which follows a polar orbit crossing the equator at approximately 1:30 pm local time. This product includes retrievals of cloud optical properties such as effective droplet radius (R_e) and optical thickness (τ_c) at multiple wavelengths (1.6, 2.1, and 3.7 μm) and cloud thermodynamic phase, as well as infrared retrievals of the cloud top temperature (CTT), pressure (CTP), and height (CTH) (Nakajima and King, 1990; Platnick et al., 2017). These products are retrieved at a nominal spatial resolution of 1 km at the surface at nadir. Due to oblique viewing angles and projection effects onto Earth's curved surface, MODIS pixel size increases from 1 km at nadir to nearly 4 km at the swath edges. MODIS data are provided as 1354 × 2030-pixel granules. This dataset includes three filters applied to the MODIS ~~data~~ retrievals for liquid warm clouds:

1. **All**: Includes retrieved cloud properties where phase=2 (liquid), R_e > 2.1 μm, and CTT > 268 K.
2. **Q06**: Includes all filters from the **All** composite plus τ_c > 4 and R_e > 4 μm. This filter is called "**Q06**" because it uses the same set of constraints as those used in Quaas et al. (2006).
3. **G18**: Includes all properties from the **Q06** composite plus 5 km CF > 0.9, solar zenith angle (θ_{solar}) < 65°, satellite zenith angle ($\theta_{satellite}$) < 55°, and sunglint pixel index (SPI) < 30°. This filter is called "**G18** because it uses the same set of constraints as those used in Grosvenor et al. (2018).

Following the same approach (and terminology for the filter names) as Gryspeerdt et al. (2022), N_d is computed for each composite using the equation $N_d = \gamma f(\text{CTT}) \tau_c^{1/2} R_e^{-5/2}$, where γ is the adiabatic condensation growth rate taking a value of 1.37×10^{-5} m$^{-0.5}$, and $f(\text{CTT}) = 0.0192T - 4.293$ is the temperature-dependent condensation rate determined from the CTT retrieval.

AMSR-E, onboard NASA's Aqua satellite, operates at multiple microwave frequencies, allowing it to retrieve cloud water path and surface precipitation rate. It has a swath width of about 1445 km with a footprint of approximately 13 km^2 at the

surface. Version 2 of the AMSR-E AE_Ocean(Rain) product provides retrievals of columnar cloud and rain water path as well as surface precipitation over each footprint using the 36.5 GHz channel (Wentz and Meissner, 2004).

The CERES instrument measures top of atmosphere radiances in the shortwave (0.3 – 5 μm), window (8 –12 μm), and total (0.3 to 200 μm) spectral channels, with a spatial resolution of approximately 20 km at nadir. Operating in cross-track, along-track, and rotating azimuth plane modes aboard the Aqua satellite, CERES scans from limb-to-limb to achieve daily global coverage. It provides retrievals of instantaneous shortwave (SW) top-of-atmosphere (TOA) radiative fluxes by incorporating MODIS cloud properties, aerosol retrievals, and meteorological parameters from the Global Modeling and Assimilation Office (GMAO) in Angular Distribution Models (ADMs) to retrieve all-sky ocean TOA fluxes to an accuracy of 6% (Loeb et al., 2009) in the Single Scanner Footprint TOA/Surface Fluxes and Clouds (SSF).

2.2 Meteorological Quantities

MERRA-2 (Modern-Era Retrospective analysis for Research and Applications, Version 2) is a reanalysis product developed by GMAO. It provides meteorological data spanning from 1980 to the present, assimilating observations from various satellites and ground-based stations. MERRA-2 offers a spatial resolution of approximately 0.625° (about 50 km) and includes 72 vertical levels spanning the atmosphere. The temporal resolution is 1 hour for all surface meteorological variables and 3 hours for 3D fields. We utilize vertical profiles of temperature and specific humidity to compute estimated inversion strength (Wood and Bretherton, 2006) and humidity above the boundary layer, near-surface meteorological variables to compute the near surface temperature advection and various other cloud controlling factors based on the 3D winds. MERRA-2 computes planetary boundary layer height (PBLH) as the the lowest level at which the heat diffusivity drops below a threshold value (for more details see McGrath-Spangler et al., 2015). These meteorological quantities have been shown to strongly influence ACI relationships (Wall et al., 2023). The MERRA-2 products are temporally interpolated to match the instantaneous time of the MODIS satellite overpass and spatially resampled using a KDTree approach (Bentley, 1975) and bilinear interpolation to match the MERRA-2 products to the pixel-scale resolution of the MODIS instrument for each individual L2 granule.

European Centre for Medium-Range Weather Forecasting (ECMWF) Reanalysis v5 (ERA5) is a reanalysis dataset that provides global meteorological data from 1950 to the present with 31 km spatial and hourly temporal resolution (Hersbach et al., 2020). It assimilates satellite and ground-based observations, including atmospheric motion vectors from cloud tops, offering strong wind constraints. We also match this reanalysis product to the instantaneous footprint from MODIS and use it to train our ML model.

3 Methods

3.1 Data Product and Aggregation

MODIS cloud retrievals at 1 km spatial resolution are gridded globally into daily files from 2007 – 2011 by 10°, 5°, 1°, 0.5°, 0.1°, and 0.05° regions. MODIS data is aggregated within each grid-box (at these different resolutions yielding, on average,

1E6, 2.5E5, 1E4, 2.5E3, 1E3, and 25 number of daily samples over the grid-box for each grid-resolution, respectively). All products are aggregated in the grid resolutions down to 0.5°. At 0.1° and 0.05° resolutions the data products at coarser spatial resolutions than the grid (i.e. AMSR-E of 13 km, CERES of 25 km) are bilinearly interpolated in space between the locations of the level-2 satellite footprints to the grid box of the gridded product in question. All of the cloud product averages are composed of warm liquid clouds only and if any ice clouds are detected within a grid-cell it is not used in the analysis.

3.2 Random Forest Model

A random forest model is an ensemble learning technique that combines multiple decision trees to improve predictive accuracy and robustness. It works by aggregating predictions from each tree, trained on random subsets of the data with replacement, and is effective in handling complex relationships and high-dimensional datasets (Breiman, 2001). Chen et al. (2022) employs a random forest algorithm to investigate the impacts of volcanic aerosols and meteorology on clouds surrounding Iceland. This approach enables the comparison of cloud responses of volcanic aerosol perturbations, isolating ACI signals. We adopt a similar approach by constructing a regression forest consisting of 100 trees independently trained to predict LWP, N_d, and CTH based on the following 13 predictor variables: PBLH, isentropic lifted condensation level based on the surface air temperature, pressure, and dew point (LCL), relative humidity above the height level of the PBLH (rhAbovePBL), estimated inversion strength (EIS), horizontal temperature advection at the surface (Tadv), surface latent heat flux (LH), total column water vapor (tqv), 10-m surface wind speed (ws10), surface precipitation (AMSRE-E), CTH (MODIS), CF (MODIS), cloud albedo (A_{cld}; CERES), and N_d (MODIS). A comprehensive list of predictors and their respective data sources is provided in Table S1.

The random forest is constructed from 100 trees with a minimal leaf size of 7 and no leaf merging. ~~Each tree~~ The full dataset was randomly partitioned into three independent subsets: 65% for training, 25% for testing, and 10% for validation. Randomized sampling partitions, as opposed to sequential (e.g., yearly) splits, did not have a significant impact on the model's outcomes. Within the training set, each decision tree in the random forest is trained on approximately 60% of the training ~~dataset with replacement, utilizing~~ data (sampled with replacement), while the remaining 40% serves as out-of-bag ~~observations to test tree performance. Our cloud~~ data for internal performance evaluation. The validation set was used for tuning model hyperparameters described in Table S2. After hyperparameter tuning, the model was retrained using both the training and validation data (75% total) to optimize performance, ensuring that the test set remained unseen and provided an unbiased assessment of model accuracy. Cloud controlling variables were primarily chosen from the selection used in the studies of Andersen et al. (2023) and Wall et al. (2023). Model performance is further evaluated using different hyperparameter values, such as the number of trees and the minimum number of samples per leaf, which are discussed below. Choosing predictors that have significant influence on cloud properties helps in effectively capturing their variability to best ensure that the model can discern meaningful patterns and relationships crucial for understanding complex atmospheric processes.

4 Results

4.1 Regions

Twelve oceanic regions are used in this study (Fig. 1a), encompassing the subtropics (California; CAL, Peruvian; PER, Namibian; NAM, and Australian; AUS, similar to Klein and Hartmann, 1993), tropics (Central East Pacific; CEP, Central Atlantic; CEA, and Western Indian; WEI), mid-latitudes (Central North Pacific; CNP, Eastern South Atlantic; ESA, and Eastern South Indian; ESI), and mixed regions (Eastern North Atlantic; ENA and Western North Pacific; WNP), with tropical and mid-latitude regions selected to ensure representation in both hemispheres and alignment within similar latitude belts. The subtropical locations contain a large abundance (greater than 60%) of warm low-level (below 3 km) marine stratocumulus cloud (Fig. S1). Tropical and Mid-latitude locations also contain substantial amounts of warm boundary layer clouds, but these regions have significant differences in sea surface temperature (Fig. 1b) with much higher amounts of mid and high-level clouds (Fig. 1c). On average, planetary boundary layer depths are significantly lower in the mid-latitude regions but have the heaviest precipitation rates (Fig. 1e) where a long tail (positive skewness) is apparent indicating a higher propensity for heavier rainfall by frequent large-scale storm-track systems. Precipitation rates are also large in the tropics (compared to the subtropics). The diversity in meteorological conditions across regions provides essential test-beds to isolate and study the impact of different cloud states and environmental controls on ACI.

4.2 Impact of grid-spacing

Six spatial resolutions with grid spacing from 10° to 0.05° are used to examine how the shape of the N_d–LWP relationship changes within each region. The total number of grid-cells over the 5 year period in the region increase roughly an order of magnitude for each grid resolution (numbers used in each grid-resolution 7.3E3, 2.9E4, 6.9E5, 2.6E6, 5.2E7, 1.9E8, respectively). Using the **All** warm-cloud filter at 1° grid resolution reveals a pronounced *inverted-V* distribution with distinct ascending and descending branches for the California region (Fig. 2c). This result agrees with Gryspeerdt et al. (2019). As the spatial resolution increases and becomes finer, the *inverted-V* shape morphs into multiple modes more closely resembling an "M" shape (Fig. 2e and f). The *inverted-V* shape of the N_d–LWP relationship can be modeled using a piecewise linear function in log–log space, following a similar approach to Gryspeerdt et al. (2019). The function is defined with a single turning point corresponding to the median in the probability distribution function (PDF) of LWP as a function of N_d:

$$\ln L = \begin{cases} \ln L_p + m_1(\ln N_d - \ln N_{d,p}), & \text{if } N_d < N_{d,p} \\ \ln L_p + m_2(\ln N_d - \ln N_{d,p}), & \text{if } N_{d,p} \geq N_{d,p} \end{cases} \tag{1}$$

Here, $N_{d,p}$ denotes the turning point—defined as the mode in LWP for a given N_d—with slope coefficients m_1 and m_2 representing the log–log gradients to the left and right of $N_{d,p}$, respectively, and L_p the corresponding LWP value at the turning point. At 1° resolution, the N_d–LWP distributions exhibit a well-defined *inverted-V* structure. However, as spatial resolution increases (e.g., 0.1° and 0.05°), the median of the PDF deviates from this structure and reveals a more prominent

"M" shape relative to the piecewise linear fit, reflecting increased subgrid variability. We focus on results at 0.1° resolution in the main analysis to balance computational cost with spatial fidelity, unless otherwise noted.

To explain why the "M" shape emerges at the 0.1° and 0.05° resolutions, a detailed analysis of τ_c and R_e is conducted. Since LWP is proportional to the product of τ_c and R_e (i.e., $LWP = 2/3\rho_w\tau_c R_e$, where ρ_w is the density of water), the localized decrease in LWP in the bottom of the "M" trough (Fig. 2e and f), which does not appear in the classic *inverted-V* shape (Fig. 2c), occurs due to a higher occurrence of clouds with smaller τ_c (Fig. ~~S2b~~3b). This change in LWP is primarily driven by variations in τ_c, as the changes in R_e between grid resolutions are similar (Fig. ~~S2c~~3c). The statistical differences between these datasets indicate that finer resolutions capture more variability in cloud optical properties, leading to a wider spread in LWP values. This results in a lower median LWP at finer scales due to the increased detection of smaller optical depth clouds.

4.3 Regional Differences

The N_d–LWP relationship varies significantly across subtropical, tropical, and midlatitude regions (Fig. 4). In subtropical regions, a distinct *inverted-V* distribution with clear ascending and descending branches is evident at a 1° grid resolution (Fig. ~~S3~~S2). At higher resolutions, this pattern transitions to an "M" shape, that is robust across all dominant subtropical stratocumulus locations (Fig. 4d ~~, Fig. S4~~and Fig. S3). Consistent with Gryspeerdt et al. (2019), the linear-fit slope of N_d-LWP (over the whole range in N_d) is negative in the subtropics at 1° grid resolution. It is also negative at 0.1° grid resolution albeit the slope is significantly less negative at higher spatial resolution. In the tropics, the shapes of these distributions are less pronounced, and LWP exhibits a more neutral/positive trend with N_d. In midlatitudes, the relationship sometimes resembles an *inverted-V*, but with a significant increase in LWP at high N_d. Linear-fit slopes in midlatitudes tend to be more positive than in the subtropics and tropical regions especially at higher spatial resolution. While the relationships remain broadly consistent within subtropical, tropical, and midlatitude regions, variations in dominant cloud types (which respond differently to aerosols, e.g. see, Christensen et al., 2016), as well as differences in large-scale meteorology and zonal gradients across these domains, can also play a role in shaping the N_d–LWP relationship, a topic we will further explore using ML.

4.4 Impact of sampling and filtering

The accuracy of τ_c and R_e retrievals, which are used to compute LWP and N_d, is generally improved by removing thin and/or broken cloud fields. Thicker overcast cloud fields have less noise from shortwave radiation scattering off a homogenous cloud-scene and more closely adhere to the scattering assumptions inherent in the plane-parallel approximation (McGarragh et al., 2018). A common approach to reduce these errors is to require τ_c to be greater than 4. This more stringent filtering is applied in the **Q06** and **G18** composites. However, thinner clouds, which are removed in these composites, are generally more sensitive to aerosol perturbations (Platnick and Twomey, 1994).

Figure 5 illustrates the impact of removing thin cloud retrievals from the calculation of the slope in $\Delta \ln LWP / \Delta \ln N_d$ across the globe using the 1° gridded product. When **All** clouds are included (Figure 5a), the slope is predominantly positive, except in the subtropics. However, using 0.1° data results in less negative slopes in these regions (Fig. ~~S5~~S4), consistent with the observed differences between 1° and 0.1° resolutions in Figure 4. When thin clouds are removed using the **Q06**

and **G18** composites (Figure 5b-c), the global distribution of slopes becomes predominantly negative, except in the southern hemisphere mid-latitude storm track, Polynesia, and various continental regions. Negative slopes are reported in the literature when applying similar τ_c thresholds (e.g., Figure 2 in Gryspeerdt et al., 2019, which used the thicker cloud composite of **G18**). Removing thin clouds reduces the number of cloud retrieval samples by 54% (Fig. 5d), highlighting the trade-off between reducing retrieval uncertainties and altering the sign of the LWP adjustment.

Figure 6 illustrates the effect of increasing the τ_c threshold on the slope of the N_d-LWP relationship. Beyond an optical thickness of approximately 1.25, the inclusion of thicker clouds with lower susceptibility results in a progressively more negative slope, converging with other composites around an optical thickness of 10. The reliability of cloud property retrievals for thin clouds depends on several factors, including sea state roughness, CF, satellite viewing and zenith angles, and particle size. McGarragh et al. (2018) highlight the challenges of retrieving accurate optical thickness values below 0.1, as subvisible cirrus (optical depth > 0.03) can impact the retrieval when they exist (as determined by spaceborne lidar; Reverdy et al., 2012). However, they demonstrate that retrievals are generally reliable for optical depths greater than 1.0, particularly over the ocean, where surface reflectance is well constrained. Given the strong influence of cloud filtering on slope estimates and the significant uncertainties in LWP adjustments, improved constraints on thin clouds are essential for refining radiative forcing estimates. The **Q06** and **G18** datasets may thus be too conservative and result in overly strong negative slopes. Therefore, to maximize sample size in N_d-LWP relationship, while recognizing the retrieval uncertainties that are intrinsic to thin and broken clouds, we will use the **All** dataset to assess meteorological impacts in our ML analysis.

5 Machine Learning the N_d-LWP Relationship

The random forest model is particularly useful for assessing feature importance because it accounts for non-linear interactions between variables, a critical capability given the inherently non-linear nature of the ACI problem. To isolate the impact of meteorology on ACI, a random forest model is used to predict LWP based on 13 cloud-controlling factors (Table S1) in all twelve regions of our study. Fig. ~~S6~~ S5 shows histograms of these factors for the California region, where over 60 million samples from the 0.05° gridded product were used to train the model. It's noteworthy that initially, ∼180 million warm cloud retrievals were available, but requiring joint AMSR-E and CERES observations reduced this to ∼100 million, with additional filtering for ice-free grid boxes further narrowing the sample to 60 million. Figure 7a shows that the model performs very well in predicting LWP when compared to the test dataset ~~. All sample indices were randomly shuffled using a uniform permutation, with 75% assigned to training and the remaining 25% withheld for testing, yielding a~~ achieving a Pearson correlation coefficient squared (r^2) of over 0.72. The relative uncertainty in the predicted LWP is ∼25% compared to the test data. Figure 7b highlights the predictors with the highest correlation values, indicating their importance. In the random forest model, importance is determined by evaluating each feature's contribution to the reduction in impurity (e.g., Gini impurity or entropy) across all trees in the forest, with higher importance scores indicating greater influence on the model's predictions (Breiman, 2001).

Random forest models with identical hyperparameter settings were trained separately for each grid resolution and each region in this study. Fig. ~~S7 and Table ??~~ S6 and Table 1 shows the accuracy tends to improve with coarser spatial resolution data,

however, these coarser grid-resolution have greatly reduced numbers of samples and lack "M" shaped LWP-N_d relationships. The importance of each factor predicted by the ML model are also consistent across regions (Fig. ~~S8).~~ S7). The model also shows robust and stable performance in terms of r^2, mean percentage error , and the ranking of variable importance across all 12 regions in our study (Table S3). In all spatial resolution datasets (excluding the coarsest), precipitation ranks as the most significant cloud controlling factor for predicting LWP (Table 1). Furthermore, A_{cld}, N_d, CTH and LCL rank towards the top of the list of "important" variables predicting LWP. ~~Precipitation (probability and intensity) and N_d are closely associated with LWP, typically increasing as LWP rises in warm clouds, though precipitation can also decrease LWP, and the bidirectional nature of these relationships makes it difficult to fully isolate cause and effect (e.g., in CloudSat observations of L'Ecuyer et al., 2009; Lebsock et al.~~

As noted earlier, our selection of hyperparameters and cloud-controlling factors was guided by Chen et al. (2022), Andersen et al. (2023), and Wall et al. (2023). To assess the impact of these choices, we trained the model 100 different times at 0.5° resolution (instead of a higher resolution as this would have been too computationally expensive), varying hyperparameters and predictor combinations and evaluating the model against our validation dataset. Using all predictor variables simultaneously in the training yields the highest r^2 values (Fig. ~~S9a~~S8a), while removing single individual predictors has modest effects unless key variables like precipitation, N_d, or A_{cld} are excluded. Including cloud albedo alongside precipitation significantly improves r^2 when using only two predictors which is not surprising given the high "importance" of these variables. A minimum of seven samples per leaf node was selected based on how r^2 responded to increasing this value (Fig. ~~S9b~~S8b); larger values force the tree to group more data in each decision, resulting in shallower trees with fewer splits, which helps prevent overfitting but may reduce predictive accuracy. Sample fraction, which controls the proportion of training data used to grow each tree, showed little effect on r^2, so we adopted the default value of 0.6, which introduces some randomness and improves generalization. Increasing the number of trees tends to decrease RMSE and increase r^2; we selected 100 trees, where RMSE plateaued and r^2 approached its maximum value (Fig. ~~S9c~~S8c), while balancing computational cost, for example, training with 100 trees requires ∼300% more CPU time than with two trees. Finally, the risk of overfitting, fitting the model too closely to noise or idiosyncrasies in the training data rather than learning generalizable patterns, is likely minimal in our case. The random forest models are small relative to the size and diversity of the dataset, and model skill does not improve at coarser spatial resolutions where overfitting would be most likely (i.e. see Table ~~??~~1). In fact, performance slightly degrades at the coarsest resolution (5°), supporting the conclusion that the models are not simply memorizing the data.

The random forest model successfully predicts the shape of the N_d-LWP relationship (~~Fig. S10~~Figure 8) thereby enabling detailed examination of the non-linear impact of cloud-controlling factors on LWP. LWP is predicted by the ML model while holding each cloud-controlling variable at a fixed value. When all predictor variables except N_d are held constant (set to either their median values, shown by the green line, or their average values, shown by the blue line in ~~Fig. S10~~Figure 8), the resulting LWP distribution exhibits an *inverted-V* shape as a function of N_d but not an "M" shape. However, allowing the cloud controlling variables to vary with N_d, by using the median value of each cloud-controlling variable for each binned N_d value from 4 to 1000 cm^{-3} reveals the "M" shape (black line in ~~Fig. S10~~Figure 8). This suggests that the co-variability of N_d with other cloud-controlling factors plays a crucial role in shaping the LWP response. While Goren et al. (2024) recently

highlighted the importance of LWP covariability with other cloud controlling factors, they identified PBLH as a primary driver shaping LWP, whereas our results suggest it plays a lesser role in this relationship with precipitation and cloud albedo being much stronger contributing factors.

5.1 Impact of Precipitation

Precipitation is identified by the ML model as the most influential cloud-controlling factor in predicting LWP. It significantly impacts the N_d-LWP relationship, generally leading to increasingly positive slopes of $\frac{d\ln LWP}{d\ln N_d}$ as precipitation rates increase. The average slope is estimated using both an ordinary least squares (OLS) fit in log–log space and a numerical differentiation approach based on finite differences that computes the mean of local slopes between adjacent points along the random forest predictions as a function of N_d. While the OLS fit captures the overall trend, the finite-difference approach reflects the average instantaneous rate of change, which can differ in shape-sensitive cases such as *inverted-V* functions. The OLS fitted slope is consistent across all regions (Fig. ~~S11~~S9). Furthermore, the random forest ML model accurately captures the shape of LWP as a function of N_d with ~~incresaing~~ increasing slopes using finite-differences as precipitation increases across composites (Fig. ~~S12~~S10).

~~On average, the ML model shows minimal~~ Precipitation (probability and intensity) and N_d are closely associated with LWP, typically increasing as LWP increases in warm clouds. While LWP and precipitation generally increase together as clouds deepen, in more developed or heavily drizzling systems, efficient rainout processes can deplete cloud liquid water, leading to a reduction in LWP and a bidirectional response in $\frac{d\ln LWP}{d\ln N_d}$ (e.g., in CloudSat observations of L'Ecuyer et al., 2009; Lebsock et al., 2008; Ch . The ML model supports this effect, showing little variation in LWP with respect to N_d ~~in~~ for non-precipitating clouds~~. This is~~ , likely because increased aerosol ~~(and thus, N_d) concentrations cannot~~ concentrations cannot further suppress drizzle in ~~non-raining clouds , a process that would otherwise enhance the cloud water path in precipitating clouds. Since this mechanism is absent, the LWP response remains flat or negative, a result that also agrees with A-train observatings including CloudSat in Chen et al. (2014). However, in~~ clouds that are already non-raining—yielding a flat or slightly negative response consistent with A-Train observations (Chen et al., 2014). In precipitating clouds, by contrast, higher N_d tends to suppress precipitation by reducing droplet size and ~~inhibiting collision-coalescence~~limiting collision–coalescence. As shown in Fig. ~~S12c, LWP increases sharply as a function of~~ S10c, LWP rises sharply with N_d before increasing more ~~slowly at approximately~~ gradually beyond about 20 cm^{-3}~~, as N_d continues to rise~~. The random forest model performs consistently well across all 12 regions, exhibiting a similar pattern of small slightly positive or negative $\frac{d\ln LWP}{d\ln N_d}$ sensitivities for non-precipitating clouds and larger positive sensitivities for raining clouds (Table S3).

5.2 Impact of Cloud Albedo

Cloud albedo plays the next most significant role in modulating the N_d-LWP relationship. Under all-sky conditions, we observe an "M" pattern (~~Fig. S10~~Figure 8), but when stratifying the data into low cloud albedo ($0<A_{cld}<0.25$), average cloud albedo ($0.25<A_{cld}<0.4$), and high cloud albedo ($0.4<A_{cld}<1$), the relationship shifts to solely an *inverted-V* shape (Fig. ~~S13~~S11), even at the highest 0.05° grid-resolution. For dimmer clouds, the peak of the *inverted-V* occurs at a relatively lower N_d (around 10

cm^{-3}), while for brighter clouds, the peak is broader and spans a wider range of concentrations (20-80 cm^{-3}), with the LWP shifted to larger values. The linear least squares fit is negative in each composite, while the finite-difference method applied to the random forest predictions yields a weaker average slope due to the pronounced positive increase at low N_d, which is not well captured by the OLS fit. Nevertheless, the consistency of slopes across composites suggests that the influence of precipitation—which tends to steepen the slope—is similar across A_{cld} groupings.

To test whether the positive LWP response to precipitation is still observed even after compositing the data by A_{cld}, which drives an overall negative OLS linear regression fit of the N_d-LWP relationship, the data is further composited by precipitation. Fig. ~~S14~~ S12 shows negative slopes of the relationship for the non-raining composites. The slope is especially negative for low-albedo clouds. As precipitation becomes heavier in the drizzle and raining regimes the slopes become more positive and larger in both the OLS and finite differences slopes. This suggests that while the observed negative slope may be shaped by cloud albedo binning, the emergence of a more positive slope is more directly tied to the presence of precipitation. However, it is important to note that cloud albedo itself is not an independent driver of cloud microphysics but rather an outcome of variables such as LWP and N_d, among others. Therefore, interpreting slope changes as being "controlled" by cloud albedo may misrepresent causal relationships and itself serves as a means for binning clouds of varying microphysical quantities.

5.3 Impact of Free Troposphere Humidity

Free tropospheric humidity has often been proposed as a key factor driving ~~a negative LWP adjustment~~ negative LWP adjustments (Ackerman et al., 2004; Chen et al., 2014; Gryspeerdt et al., 2019). However, Fig. ~~S15~~ S13 does not support this expectation, that higher relative humidity leads to a more positive LWP slope as droplet concentrations increase. Based on this first-level binned analysis, LWP appears largely unaffected by relative humidity.

To verify this, we further decompose the response using the most influential variables identified by the random forest method: first by precipitation, then by cloud albedo, and finally by relative humidity. Figure 9 shows that the N_d-LWP OLS slope remains negative for non-precipitating clouds under all cloud controlling factor groupings. For dimmer clouds, A_c(low), higher relative humidity (moist) is actually associated with stronger LWP decreases (not increases) with N_d. Only in drizzling, high-albedo clouds does the LWP OLS slope become positive under an increased RH. If anything, higher free-tropospheric relative humidity decreases LWP. Despite a seeming consensus in the literature (Ackerman et al., 2004; Chen et al., 2014; Gryspeerdt et al., 2019), our findings suggest that relative humidity above the PBL plays a relatively minor role in ACI compared to other factors like precipitation state and cloud macroscopic properties such as albedo in subtropical stratocumulus clouds. This analysis is included to explicitly demonstrate that, contrary to prior expectations, free-tropospheric humidity exerts only a weak influence on LWP. Detailed measurements and/or modeling of cloud top entrainment or divergence may be needed to close this research gap instead of relying on inferred relationships to above PBL relative humidity.

5.4 Radiative Effect

We introduce a new method using ML random forest model predictions for computing aerosol indirect radiative forcing. This approach captures nonlinear relationships between variables, simplifies the computation of partial derivatives while holding

other variables constant (e.g. LWP), and eliminates the need for data stratification methods such as binning. The ML model ~~by itself can thus provides direct predictions of the Twomeyand rapid adjustment radiative effects(assuming~~ can therefore directly predict the Twomey, LWP, and cloud fraction radiative effects, provided it has been trained with the necessary variables~~)~~. The shortwave radiative forcing due to a change in anthropogenic aerosols can be expressed as

$$\Delta F = -\Delta\alpha \overline{F^{\downarrow}} \phi_{atm} \tag{2}$$

where the negative sign denotes that an increase in planetary albedo reduces the net downward (absorbed) shortwave flux, consistent with the convention that positive ΔF denotes a warming (increase in absorbed energy). Here, $\overline{F^{\downarrow}}$ is the annual mean incoming solar radiative flux at the top of the atmosphere ~~, with a globally averaged~~ (global mean value of 340.2 W m^{-2}~~so the~~), enabling comparison of relative changes in the outgoing flux between regions~~can be compared~~ , and ϕ_{atm} is ~~the transfer function that accounts~~ a transfer function accounting for the average ~~albedo of the air~~ atmospheric attenuation above the surface and clouds, ~~taking an average value of~~ typically taken as 0.7. The change in planetary albedo due to observed variations in N_d and aerosol index (AI) is given by:

$$\Delta\alpha = \frac{d\alpha}{d\ln N_d}\frac{d\ln N_d}{d\ln AI}\overline{\Delta \ln AI}, \tag{3}$$

where $\frac{d\ln N_d}{d\ln AI}$ represents the aerosol-cloud sensitivity, which relates changes in cloud droplet concentration to changes in AI. AI is the product of aerosol optical depth (AOD retrieved at 550 nm) and the Ångström exponent computed using the AOD at 550 nm and 865 nm wavelength pairs derived from MODIS, which is found to be a better proxy for column cloud condensation nuclei than AOD alone (Quaas et al., 2006). The relationship is positive across our regions with an average value of approximately 0.31 ± 0.16 (Fig. ~~S16~~S14). Stronger slopes are typically retrieved in stratocumulus-dominated regions, while weaker slopes tend to occur under more unstable atmospheric conditions with lower cloud fraction. Weaker slopes under these conditions may be partially caused by larger error contributions stemming from satellite retrieval artifacts related to aerosol humidification and 3D scattering effects between clouds with lower cloud fractions (Gryspeerdt and Stier, 2012; Christensen et al., 2017). Higher precipitation rates in tropical regions is also associated with smaller N_d, complicating the causal direction of aerosols and their role on clouds or potentially in this case, the precipitation impact on aerosols as shown in Christensen et al. (2023).

The $\overline{\Delta \ln AI} = \overline{\ln\left(\frac{AI^{PD}}{AI^{PI}}\right)}$ term represents the average log-change in aerosol optical depth due to the influence of anthropogenic aerosols (i.e., based on present-day, AI^{PD}, and pre-industrial levels, AI^{PI}). An Earth system model is needed to estimate this quantity. The average value of $\overline{\Delta \ln AI}$ is estimated from 1° spatial resolution E3SM simulations of 5-year average present-day and pre-industrial aerosol emissions, the global results of which are displayed in Fig. ~~S17~~S15, and over our regions, $\overline{\Delta \ln AI} = 0.41 \pm 0.27$.

The derivative of planetary albedo, defined as $\alpha = (1 - CF)\alpha_{clr} + CF\alpha_{cld}$, where α_{clr} is clear-sky albedo, with respect to N_d is carried out using the chain rule expansion described in Bellouin et al. (2020):

$$\frac{d\alpha}{d\ln N_d} = CF\left(\left.\frac{\partial \alpha_{cld}}{\partial \ln N_d}\right|_{\text{LWP, CF}} + \frac{\partial \alpha_{cld}}{\partial LWP}\frac{dLWP}{d\ln N_d} + \frac{\partial \alpha_{cld}}{\partial CF}\frac{dCF}{d\ln N_d}\right). \tag{4}$$

where the clear-sky contribution to the planetary albedo is not included because it is part of the direct radiative effect (i.e. $(\alpha_{cld}-\alpha_{clr})\frac{dCF}{d\ln N_d}$ term). All three terms in equation 4 represent radiative sensitivities, with the first and second terms restricted to cloudy regions. When combined with equations 2 and 3, these terms correspond to the Twomey radiative effect (where LWP and CF are held constant), LWP adjustment, and CF adjustments. For completeness, the full expression can be written as

$$\Delta F = -\overline{F^{\downarrow}}\phi_{atm} CF\left[\left.\frac{\partial \alpha_{cld}}{\partial \ln N_d}\right|_{\text{LWP, CF}} + \frac{\partial \alpha_{cld}}{\partial LWP}\frac{dLWP}{d\ln N_d} + \frac{\partial \alpha_{cld}}{\partial CF}\frac{dCF}{d\ln N_d}\right]\frac{d\ln N_d}{d\ln AI}\overline{\Delta \ln AI}. \tag{5}$$

To implement this framework, three separate random forest models are trained to predict A_{cld}, LWP, and CF based on the same set of cloud-controlling variables. However, when training the model to predict A_{cld}, we exclude A_{cld} as a predictor and replace it with LWP, applying a similar replacement strategy when predicting CF. This approach ensures consistency in the input variables across all models. The performance of the ML models using A_{cld} and CF predictors (instead of LWP) is summarized in Tables ~~S1 and S2~~S4 and S5. These models achieve r^2 values that are comparable to those obtained when predicting LWP. Across most grid resolutions, CF, LWP, and N_d are identified as the most "important" terms influencing A_{cld}. Similarly, cloud top height, N_d , and the LCL play significant roles in predicting CF.

The ~~first term~~three terms in parentheses correspond to the **Twomey**, **liquid water path**, and **cloud fraction** radiative effects, respectively. These are multiplied by a radiative scaling factor defined as $(-1.) * CF * \overline{F^{\downarrow}} * \frac{d\ln N_d}{d\ln AI}\overline{\Delta \ln AI}$ where the negative sign indicates that an increase in albedo (from higher N_d) reduces the net downward shortwave flux. The first term (Twomey effect) inside the parenthesis, $\left.\frac{\partial \alpha_{cld}}{\partial \ln N_d}\right|_{\text{LWP, CF}}$, is evaluated using a radiative kernel approach, similar to that described in Wall et al. (2023). First, the frequency of occurrence from cloud retrievals within logarithmic bins of LWP and linear bins of CF is estimated from the full dataset (Figure 10a). Second, the sensitivity of A_{cld} to N_d is computed within each joint LWP–CF bin, where the random forest model uses fixed values corresponding to the midpoints of the bin. Predictions of A_{cld} are generated across the full range of N_d within each bin and are fit using a linear least squares method (Figure 10b). To compute uncertainty, the slope within each bin is computed three times using the median values of other cloud-controlling variables and twice more using the median $\pm$ 1 standard deviation. Finally, the resulting Twomey radiative kernel (Figure 10c) is calculated from the product of the normalized PDF with the sensitivity summed over all LWP and CF bins (i.e., $\left.\frac{\partial \alpha_{\text{cld}}}{\partial \ln N_d}\right|_{\text{LWP, CF}} = \sum_{i=1}^{N_{\text{LWPbins}}}\sum_{j=1}^{N_{\text{CFbins}}} N_{\text{pdf},i,j}\left(\frac{\partial A_{\text{cld}}}{\partial N_d}\right)_{i,j}$). Uncertainties are propagated using the same approach. For the California region, this results in a Twomey radiative effect slope of $\left.\frac{\partial \alpha_{\text{cld}}}{\partial \ln N_d}\right|_{\text{LWP, CF}} = 0.067 \pm 0.0085$.

The next two terms, the LWP and CF adjustment terms, are computed without the need for a radiative kernel since LWP and CF vary with N_d. Figure 11a shows a logarithmic increase in A_{cld} as LWP increases. (Note, the small scale variability

is an artifact of random forest regressors, where averaging outputs from multiple decision trees creates a piecewise constant function that appears smooth, but specific trees cause changes in output nodes as input variables vary, and predictions were made using one RF output per N_d bin with median values of the other inputs for that bin.) This relationship is expected given the analytic two-stream approximation $A_{cld} \simeq \frac{\tau_c}{\gamma+\tau_c}$, where γ depends on the degree of forward scattering but for water clouds is approximately 13.33, with $\tau_c = \left(\frac{3}{2}\right) \frac{LWP}{R_e}$ (Stephens, 1978). Figure 11b shows the familiar *inverted-V* relationship (at 1° spatial resolution) with a strong increase at low N_d followed by a peak and decline to larger LWP. The product of $\frac{\partial A_{\text{cld}}}{\partial \text{LWP}} \frac{\text{dLWP}}{d \ln N_d}$ represents the sensitivity for the LWP adjustment which for California is -0.016 ± 0.007–negative and about 5 times smaller than the Twomey effect. Thus, clouds lose LWP as N_d increases causing A_{cld} to decrease. Negative LWP adjustments have been identified in natural experiments like ship and volcano tracks (Christensen et al., 2022) as well as in general for warm boundary layer stratocumulus clouds (Gryspeerdt et al., 2019) generally assumed to be caused by enhanced cloud top entrainment and dessication drying of the clouds as they become more polluted (Ackerman et al., 2004).

The CF adjustment is generally associated with much more uncertainty. Observed positive correlations are typically found between cloud cover fraction and aerosol optical thickness. However, these correlations are influenced by physical processes as well as uncertainties caused by artifacts, such as cloud contamination of satellite-retrieved N_d under low cloud fraction conditions, co-variation of cloud fraction with relative humidity/wind speed, and cloud processing of aerosols, which may bias the magnitude of the CF adjustment (Quaas et al., 2010; Grandey et al., 2013). Figure 11c shows that A_{cld} increases as a function of CF until approximately 0.8, where a sharp increase suddenly occurs as the cloud scene becomes fully overcast. A similar conclusion, albeit replacing A_{cld} with τ_c (which, using the 2-stream approximation, is a good proxy for A_{cld}), was also found in Coakley et al. (2005). The rapid rise towards overcast conditions may be a retrieval artifact caused by three-dimensional radiative transfer effects. As the separation between the clouds becomes comparable to the cloud thickness, radiation escaping through the sides of clouds has a high probability of being scattered upward by nearby clouds, thereby contributing significantly to the reflected radiances (Welch and Wielicki, 1985). Figure 11d shows that CF rapidly rises with N_d to about 85 cm^{-3} then flattens out for larger values of N_d. The mechanism behind the increase in CF with N_d has been hypothesized to be due to the reduction in precipitation efficiency, leading to more persistent clouds (Albrecht, 1989) and possibly due to concurrent increases in evaporation and cloud breakup. Gryspeerdt et al. (2016) suggests that the observed correlation may be due to meteorological covariations and artifacts in cloud properties. The product of $\frac{\partial A_{cld}}{\partial CF}$ and $\frac{dCF}{dlnNd}$ = 0.011 ± 0.01 for the California region. It is positive, suggesting that increased aerosol levels enhance cloud fraction and cloud albedo causing further radiative cooling.

The radiative adjustments are affected by spatial resolution. Figure 11e - h shows that the strength of all the adjustment slopes tend to decrease when using the 0.1° gridded product compared to 1°. The shift from *inverted-V* to "M" shapes (comparing 11b with 11f) is associated with weaker slopes (more positive values). Additionally, the cloud albedo sensitivity to cloud fraction and liquid water path decreases at higher spatial resolutions. These results suggest that data aggregation has a profound influence on the strength of the estimated ACI relationship. Similar conclusions about the grid-scale dependence of ACI have been noted in previous studies (McComiskey and Feingold, 2012; Feingold et al., 2016), but those did not explicitly examine

the N_d–LWP relationship, apply machine learning, decompose the radiative forcing into components, or analyze a broad range of satellite grid resolutions at the global scale—further supporting the significance of our findings.

All three terms are now combined using Equation 5 to estimate the total radiative forcing. Table 2 lists radiative forcing estimates for the combined subtropical regions. Aside from the coarsest resolution, the Twomey effect remains relatively unchanged as a function of grid resolution, with values of approximately -1 ± 0.2 W m^{-2}. The LWP adjustment contributes very little to the net radiative forcing. At 1° resolution, the CF adjustment makes up over a half of the response, but at finer resolutions, the adjustment terms become negligible. While CF increases on average as a function of grid-resolution (from 0.6 to 0.75), this contribution leads to a slightly larger radiative scaling (since the forcing is proportional to the cloud fraction). However, the primary reason for the reduction in the adjustment terms is the weakened LWP and CF sensitivities to N_d at higher spatial resolutions. While the sensitivities are slightly different between regions (radiative forcing being largest in the mid-latitudes and weakest in the tropics; Tables ~~S3 and S4~~S6 and S7), these ACI relationships to spatial resolution are similar across regions despite having different meteorological and cloud regimes.

6 Conclusions

We have developed a comprehensive satellite and reanalysis dataset gridded from coarse (10°) to fine (0.05°) spatial resolutions using 3 different commonly used warm-cloud filters (Gryspeerdt et al., 2022) over a 5 year time period. Using this data, a series of outstanding questions that give rise to significant uncertainty in the quantification of ERF$_{aci}$ and the N_d-LWP relationship have been addressed with the use of ML.

How does the structure of the N_d–LWP relationship change as the spatial grid resolution increases to finer scales? The structure of the N_d–LWP relationship changes significantly with increasing grid resolution. At coarser resolutions, the relationships found within the subtropics exhibits a classic *inverted-V* shape. As the resolution becomes finer, the relationship becomes more detailed, revealing multiple modes, including an "M" shape at scales approaching 0.1°. This increase in resolution captures more variability and impact of optically thinner clouds, resulting in a wider spread of LWP values when stratifying by primary cloud controlling variables (like precipitation and cloud albedo) identified using a random forest ML model.

How do subtropical, tropical, and midlatitude regions differ in their N_d–LWP relationships? The N_d–LWP relationship varies significantly between tropical, subtropical, and midlatitude regions. In subtropical regions, the relationship typically exhibits an *inverted-V* or "M" shape, with negative linear slopes, which become less negative at increasing spatial resolution. In tropical regions, the relationship does not show these distinct shapes and the linear sensitivity is roughly flat. In midlatitude regions, the relationship is generally positive with less coherent*inverted-V* or "M" structures, indicating that regional meteorological conditions significantly influence the N_d–LWP relationship.

How does satellite filtering and sampling clouds with different characteristics influence the N_d–LWP relationship? Satellite filtering and sampling of clouds with different properties have a substantial impact on the N_d–LWP relationship. Including **All** retrieved warm clouds, which importantly include thin clouds, results in a predominantly positive slope, while

more stringent filtering (e.g., removing thinner clouds) leads to mostly negative slopes. Thicker, opaque clouds improve the accuracy of cloud property retrievals but are less sensitive to aerosol perturbations. This indicates that the choice of cloud sampling and filtering criteria can significantly alter the interpretation of the N_d–LWP relationship and its associated radiative effects.

What are the primary meteorological drivers shaping the N_d–LWP relationship? A random forest model identified precipitation and cloud albedo as critical factors shaping the relationship. Precipitation strongly influences the N_d-LWP relationship, with steeper positive slopes as precipitation rate increases. In non-precipitating clouds, LWP remains flat due to the absence of drizzle suppression, while in precipitating clouds, increasing N_d suppresses rainfall, which increases LWP before leveling off. Binning the relationship by cloud albedo leads to negative slopes of the *inverted-V* distribution, while precipitation controls the positive slopes. Relative humidity above the PBL has a minor influence on ACI compared to precipitation state and cloud macroscopic properties like albedo.

What is the impact of changing spatial resolution on the radiative effects of ACI? The Twomey radiative effect is the dominant term, with LWP adjustments being small by comparison; this is consistent with a breadth of observed natural laboratory results by Toll et al. (2019). The radiative forcing is robust across all grid resolutions. On the other hand, the ~~rapid~~ LWP and CF radiative adjustments are strongly affected by spatial grid resolution, wherein the CF adjustment makes up a 1/2 of the response at 1° resolution. Part of the sensitivity decrease is due to shifting from an *inverted-V* to "M" in subtropical regions, but a larger part of the response is due to a weaker A_{cld} relationship with LWP and CF as spatial resolution increases.

Estimates of ACI relationships are sensitive to spatial resolution, regional variations, and satellite sampling methods. Associated radiative forcing is impacted by these factors, which underscores the need to evaluate high-resolution, region-specific Earth system model representations of these processes. ML models can enhance our understanding by capturing non-linear effects, identifying key predictors like precipitation and cloud albedo, and offering new capabilities for quantifying aerosol radiative forcing. As Earth system models increase spatial resolution, these data and analyses will be useful for evaluating ACI in warm clouds, which is necessary for identifying model deficiencies and will be addressed in a follow on companion paper.

Data availability. All of the six globally gridded resolution datasets used in this study have been archived and provided through DataHub https://doi.org/10.25584/3000993 (last access: 10 November 2025). The MODIS collection 6 products are available at https://earthdata.nasa.gov (last access: 17 August 2024). The CERES SYN Ed4a 4 product is available at https://ceres.larc.nasa.gov (last access: 17 August 2024). MERRA-2 data were obtained from https://goldsmr4.gesdisc.eosdis.nasa.gov/data/MERRA2/ (last access: 17 August 2024; NASA, 2023). ECMWF ERA5 data were obtained from https://www.ecmwf.int/en/forecasts/dataset/ecmwf-reanalysis-v5 (last access: 17 August 2024). AMSR-E precipitation data were obtained from https://n5eil01u.ecs.nsidc.org/DP1/AMSA/AU_Rain.001/ (last access: 17 August 2024).

Appendix A

A1

Author contributions. MWC carried out the analysis and wrote the paper with contributions from all co-authors. AG assisted in the design of the ML random forest model. PLM provided E3SM simulations. Research and development ideas, as well as writing and editing, were contributed to by AG, ACV, JM, and PLM.

Competing interests. MWC is an Editor for ACP.

Disclaimer. TEXT

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~~(ESMD) and Regional and Global Model Analysis (RGMA) program areas. 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-ERCAP0015329, BER-ERCAP0018473, and BER-ERCAP0020990. 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

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, https://doi.org/10.1038/nature03174, 2004.

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

Andersen, H., Cermak, J., Douglas, A., Myers, T. A., Nowack, P., Stier, P., Wall, C. J., and Wilson Kemsley, S.: Sensitivities of cloud radiative effects to large-scale meteorology and aerosols from global observations, Atmospheric Chemistry and Physics, 23, https://doi.org/10.5194/acp-23-10775-2023, 2023.

Arola, A., Lipponen, A., Kolmonen, P., Virtanen, T. H., Bellouin, N., Grosvenor, D. P., Gryspeerdt, E., Quaas, J., and Kokkola, H.: Aerosol effects on clouds are concealed by natural cloud heterogeneity and satellite retrieval errors, Nat Commun, 13, 7357, https://doi.org/10.1038/s41467-022-34948-5, arola, Antti Lipponen, Antti Kolmonen, Pekka Virtanen, Timo H Bellouin, Nicolas Grosvenor, Daniel P Gryspeerdt, Edward Quaas, Johannes Kokkola, Harri eng 821205/EC | Horizon 2020 Framework Programme (EU Framework Programme for Research and Innovation H2020)/ 337549/Academy of Finland (Suomen Akatemia)/ 446279238 / GZ QU 311/27-1/Deutsche Forschungsgemeinschaft (German Research Foundation)/ England 2022/11/30 Nat Commun. 2022 Nov 30;13(1):7357. doi: 10.1038/s41467-022-34948-5., 2022.

Bellouin, N., Quaas, J., Gryspeerdt, E., Kinne, S., 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., Mulmenstadt, J., Neubauer, D., Possner, A., Rugenstein, M., Sato, Y., Schulz, 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, e2019RG000 660, https://doi.org/10.1029/2019RG000660, e2019RG000660 10.1029/2019RG000660, 2020.

Bentley, J. L.: Multidimensional binary search trees used for associative searching, Commun. ACM, 18, 509–517, https://doi.org/10.1145/361002.361007, 1975.

Breiman, L.: Random Forests, Machine Learning, 45, https://doi.org/10.1023/A:1010933404324, 2001.

Caldwell, P. M., Terai, C. R., Hillman, B., Keen, N. D., Bogenschutz, P., Lin, W., Beydoun, H., Taylor, M., Bertagna, L., Bradley, A. M., Clevenger, T. C., Donahue, A. S., Eldred, C., Foucar, J., Golaz, J.-C., Guba, O., Jacob, R., Johnson, J., Krishna, J., Liu, W., Pressel, K., Salinger, A. G., Singh, B., Steyer, A., Ullrich, P., Wu, D., Yuan, X., Shpund, J., Ma, H.-Y., and Zender, C. S.: Convection-Permitting Simulations With the E3SM Global Atmosphere Model, Journal of Advances in Modeling Earth Systems, 13, e2021MS002 544, https://doi.org/https://doi.org/10.1029/2021MS002544, e2021MS002544 2021MS002544, 2021.

Chen, Y., Haywood, J., Wang, Y., et al.: Machine learning reveals climate forcing from aerosols is dominated by increased cloud cover, Nature Geoscience, 15, https://doi.org/10.1038/s41561-022-00991-6, 2022.

Chen, Y. C., Christensen, M., Stephens, G., et al.: Satellite-based estimate of global aerosol–cloud radiative forcing by marine warm clouds, Nature Geoscience, 7, https://doi.org/10.1038/ngeo2214, 2014.

Christensen, M. W., Chen, Y.-C., and Stephens, G. L.: Aerosol indirect effect dictated by liquid clouds, Journal of Geophysical Research: Atmospheres, 121, 14,636–14,650, https://doi.org/https://doi.org/10.1002/2016JD025245, 2016.

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, https://doi.org/10.5194/acp-17-13151-2017, 2017.

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, https://doi.org/10.5194/acp-22-641-2022, 2022.

Christensen, M. W., Ma, P.-L., Wu, P., Varble, A. C., Mülmenstädt, J., and Fast, J. D.: Evaluation of aerosol–cloud interactions in E3SM using a Lagrangian framework, Atmospheric Chemistry and Physics, 23, 2789–2812, https://doi.org/10.5194/acp-23-2789-2023, 2023.

Coakley, J. A., Friedman, M. A., and Tahnk, W. R.: Retrieval of Cloud Properties for Partly Cloudy Imager Pixels, Journal of Atmospheric and Oceanic Technology, 22, 3 – 17, https://doi.org/10.1175/JTECH-1681.1, 2005.

Feingold, G., McComiskey, A., Yamaguchi, T., Johnson, J. S., Carslaw, K. S., and Schmidt, K. S.: New approaches to quantifying aerosol influence on the cloud radiative effect, Proceedings of the National Academy of Sciences, 113, 5812–5819, https://doi.org/10.1073/pnas.1514035112, 2016.

Goren, T., Chourdhury, G., Kretzschmar, J., and McCoy, I.: Co-variability drives the inverted-V sensitivity between liquid water path and droplet concentrations, EGUsphere, 2024, 1–18, https://doi.org/10.5194/egusphere-2024-2245, 2024.

Grandey, B. S., Stier, P., and Wagner, T. M.: Investigating relationships between aerosol optical depth and cloud fraction using satellite, aerosol reanalysis and general circulation model data, Atmospheric Chemistry and Physics, 13, 3177–3184, 2013.

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., Hünerbein, 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. B., 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, https://doi.org/10.1029/2017rg000593, go3fd Times Cited:197 Cited References Count:234, 2018.

Gryspeerdt, E. and Stier, P.: Regime-based analysis of aerosol-cloud interactions, Geophysical Research Letters, 39, https://doi.org/https://doi.org/10.1029/2012GL053221, 2012.

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

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, https://doi.org/10.5194/acp-19-5331-2019, 2019.

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, https://doi.org/10.5194/amt-15-3875-2022, 2022.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., De Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., de Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.-N.: The ERA5 global reanalysis, Quarterly Journal of the Royal Meteorological Society, 146, 1999–2049, https://doi.org/https://doi.org/10.1002/qj.3803, 2020.

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, https://doi.org/10.1175/JAS-D-19-0241.1, 2020.

Klein, S. A. and Hartmann, D. L.: The Seasonal Cycle of Low Stratiform Clouds, Journal of Climate, 6, https://doi.org/10.1175/1520-0442(1993)006<1587:TSCOLS>2.0.CO;2, 1993.

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

L'Ecuyer, T. S., Berg, W., Haynes, J., Lebsock, M., and Takemura, T.: Global observations of aerosol impacts on precipitation occurrence in warm maritime clouds, Journal of Geophysical Research: Atmospheres, 114, https://doi.org/https://doi.org/10.1029/2008JD011273, 2009.

Loeb, N., Wielicki, B., Doelling, D., Smith, G., Keyes, D., Kato, S., Manalo-Smith, N., and Wong, T.: Toward Optimal Closure of the Earth's Top-of-Atmosphere Radiation Budget, J. Climate, 22, 748–766, https://doi.org/10.1175/2008JCLI2637.1, 2009.

McComiskey, A. and Feingold, G.: The scale problem in quantifying aerosol indirect effects, Atmospheric Chemistry and Physics, 12, 1031–1049, https://doi.org/10.5194/acp-12-1031-2012, 2012.

McGarragh, G. R., Poulsen, C. A., Thomas, G. E., Povey, A. C., Sus, O., Stapelberg, S., Schlundt, C., Proud, S., Christensen, M. W., Stengel, M., Hollmann, R., and Grainger, R. G.: The Community Cloud retrieval for CLimate (CC4CL) – Part 2: The optimal estimation approach, Atmospheric Measurement Techniques, 11, https://doi.org/10.5194/amt-11-3397-2018, 2018.

McGrath-Spangler, E. L., Molod, A., Ott, L. E., and Pawson, S.: Impact of planetary boundary layer turbulence on model climate and tracer transport, Atmospheric Chemistry and Physics, 15, 7269–7286, https://doi.org/10.5194/acp-15-7269-2015, 2015.

Nakajima, T. and King, M. D.: Determination of the Optical Thickness and Effective Particle Radius of Clouds from Reflected Solar Radiation Measurements. Part I: Theory, Journal of Atmospheric Sciences, 47, 1878 – 1893, https://doi.org/10.1175/1520-0469(1990)047<1878:DOTOTA>2.0.CO;2, 1990.

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, 33, 334–347, https://doi.org/10.1175/1520-0450(1994)033<0334:DTSOCA>2.0.CO;2, 1994.

Platnick, S., Meyer, K. G., King, M. D., Wind, G., Amarasinghe, N., Marchant, B., Arnold, G. T., Zhang, Z., Hubanks, P. A., Holz, R. E., Yang, P., Ridgway, W. L., and Riedi, J.: The MODIS Cloud Optical and Microphysical Products: Collection 6 Updates and Examples From Terra and Aqua, IEEE Transactions on Geoscience and Remote Sensing, 55, 502–525, https://doi.org/10.1109/TGRS.2016.2610522, 2017.

Quaas, J., Boucher, O., and Lohmann, U.: Constraining the total aerosol indirect effect in the LMDZ and ECHAM4 GCMs using MODIS satellite data, Atmospheric Chemistry and Physics, 6, 947–955, https://doi.org/10.5194/acp-6-947-2006, 2006.

Quaas, J., Stevens, B., Stier, P., and Lohmann, U.: Interpreting the cloud cover-aerosol optical depth relationship found in satellite data using a general circulation model, Atmospheric Chemistry and Physics, 10, 6129–6135, 2010.

Reverdy, M., Noel, V., Chepfer, H., and Legras, B.: On the Origin of Subvisible Cirrus Clouds in the Tropical Upper Troposphere, Atmospheric Chemistry and Physics, 12, 2012.

Stephens, G. L.: Radiation profiles in extended water clouds. II: Parameterization schemes, Journal of the Atmospheric Sciences, 35, 2123–2132, 1978.

Toll, V., Christensen, M., Quaas, J., and et al.: Weak average liquid-cloud-water response to anthropogenic aerosols, Nature, 572, 51–55, https://doi.org/10.1038/s41586-019-1423-9, 2019.

Wall, C. J., Storelvmo, T., and Possner, A.: Global observations of aerosol indirect effects from marine liquid clouds, Atmospheric Chemistry and Physics, 23, https://doi.org/10.5194/acp-23-13125-2023, 2023.

Welch, R. M. and Wielicki, B. A.: A radiative parameterization of stratocumulus cloud fields, Journal of the Atmospheric Sciences, 42, 2888–2897, 1985.

Wentz, F. J. and Meissner, T.: AMSR-E/Aqua L2B Global Swath Ocean Products derived from Wentz Algorithm, Version 2, Tech. rep., NASA National Snow and Ice Data Center Distributed Active Archive Center, Boulder, Colorado, USA, https://doi.org/10.5067/AMSRE/AE_OCEAN.002, [Indicate subset used]. [Date Accessed]., 2004.

Wood, R. and Bretherton, C. S.: On the Relationship between Stratiform Low Cloud Cover and Lower-Tropospheric Stability, Journal of Climate, 19, 6425 – 6432, https://doi.org/10.1175/JCLI3988.1, 2006.

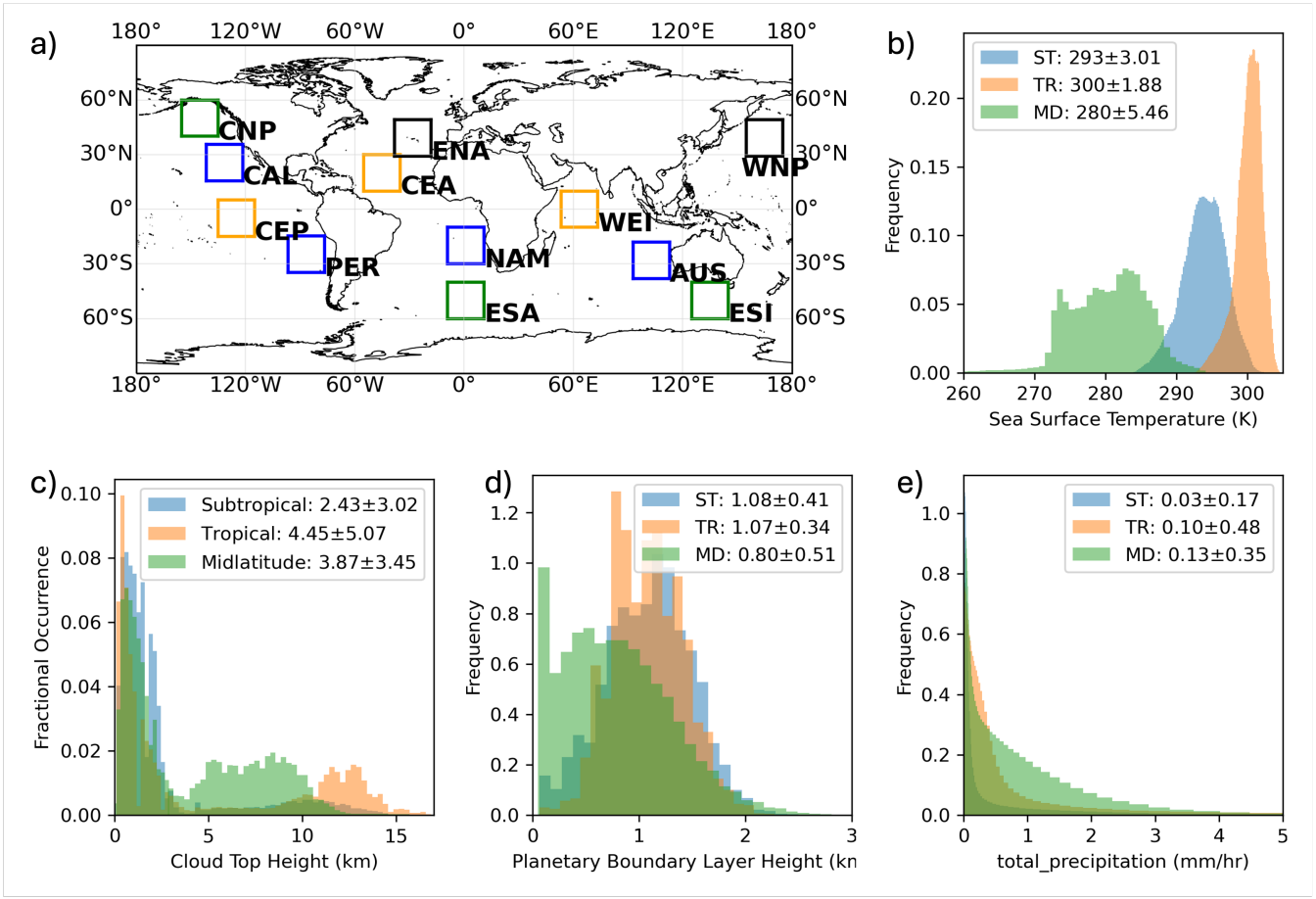

Figure 1. Oceanic regions 20° × 20° in domain size (a) are displayed. The combined subtropical (ST) regions include California (CAL), Peruvian (PER), Namibian (NAM), and Australian (AUS); the tropical (TR) regions include Central East Pacific (CEP), Central Atlantic (CEA), and Western Indian (WEI); and the mid-latitude (MD) regions include Central North Pacific (CNP), Eastern South Atlantic (ESA), and Eastern South Indian (ESI). Mixed regions—Eastern North Atlantic (ENA) and Western North Pacific (WNP)—are not included in the histograms. MERRA-2 Sea Surface Temperature (b), MODIS retrieved Cloud Top Height (c), MERRA-2 Planetary Boundary Layer Depth (d), and MERRA-2 Precipitation Rate (e) are determined from 1-year (2010) data. Means and standard deviations of each distribution are provided.

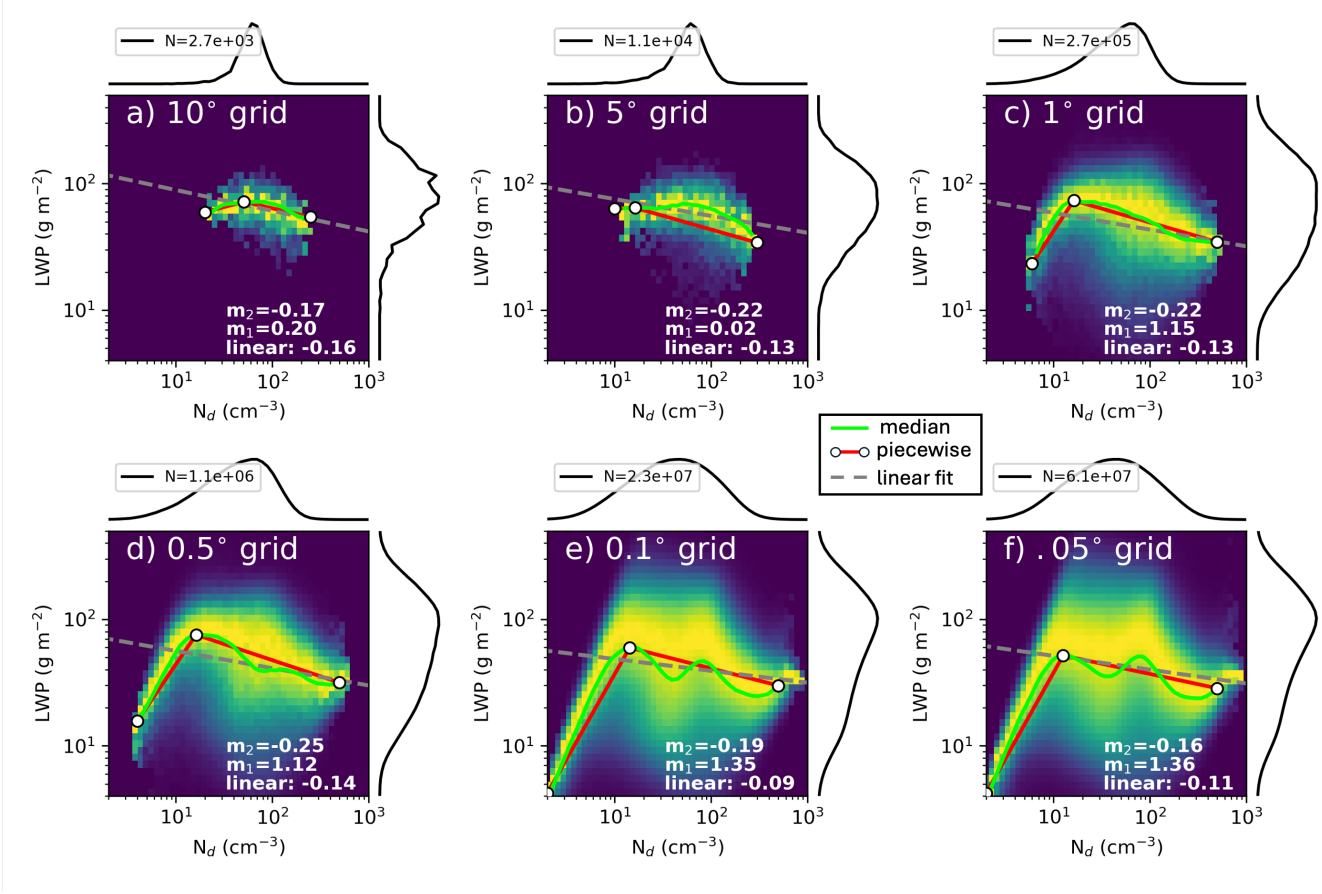

Figure 2. The N_d–LWP relationship is shown as a 2D histogram of joint frequencies, normalized by the total number of binned N_d retrievals, using five years of MODIS data at spatial resolutions of 10° (a), 5° (b), 1° (c), 0.5°, 0.1°, and 0.05° over a 20° × 20° domain in California (Figure 1). Black lines indicate the distributions of N_d (top) and LWP (right). Red lines represent piecewise fits to the ascending and descending branches, with slope values (m_n) provided. The gray dashed line shows the linear least-squares fit of LWP against N_d using all retrieved values.

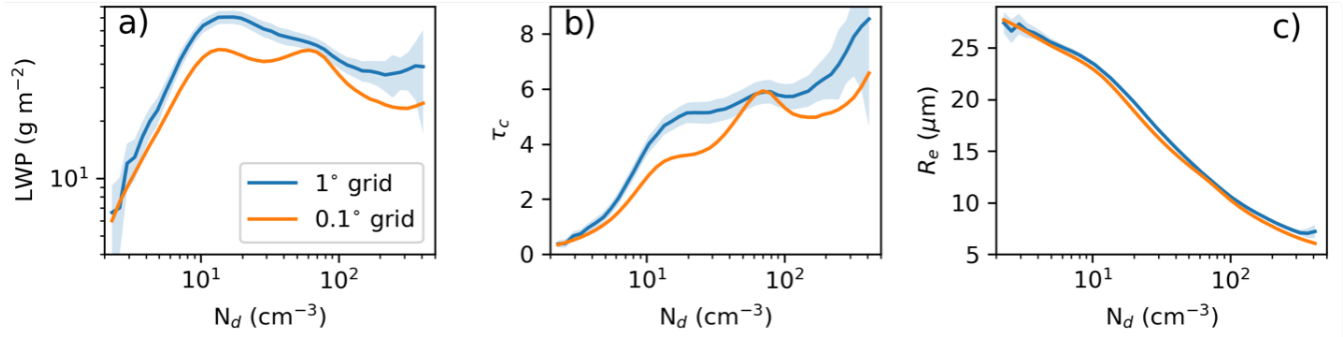

Figure 3. Median LWP (a), cloud optical thickness (τ_c) (b), and droplet effective radius (R_e) from 50 bins increasing by the log in N_d for 5 years of data across the Peruvian region averaged into 1° (blue) and 0.1° (orange) grids.

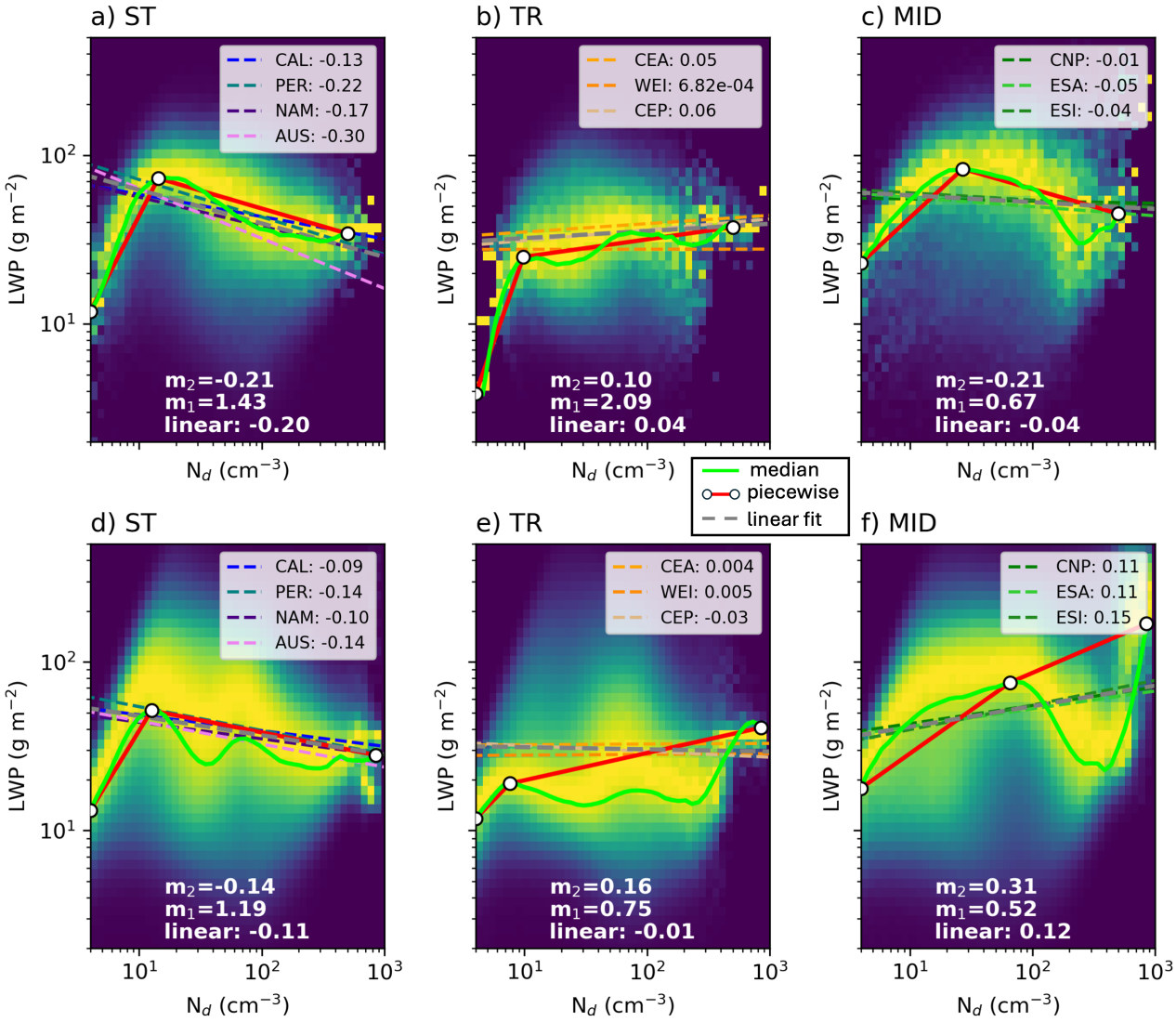

Figure 4. The N_d–LWP relationship expressed using a 2D histogram of the frequency of the LWP and N_d using 5 years of MODIS cloud retrievals aggregated at 1° (a,b,c) and 0.1° resolutions for combined subtropical (ST), tropical (TR), and midlatitude (MID) locations in Figure 1. Red lines indicate the piecewise slopes in the ascending and descending branches of the distribution, dashed lines represent the linear least squares fit line for each region.

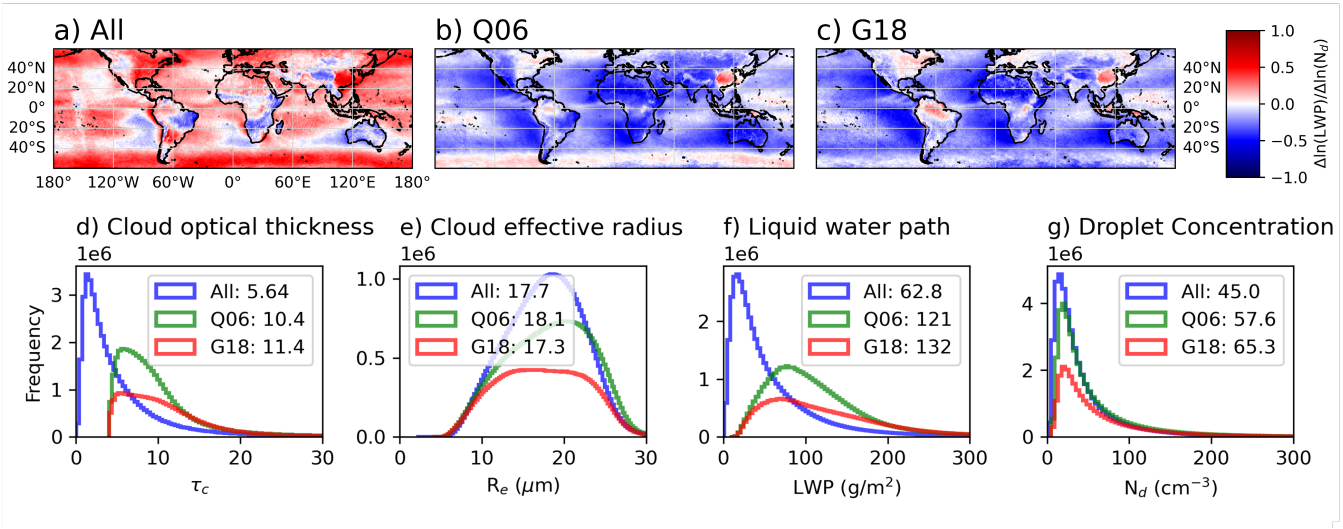

Figure 5. Linear least squares fit between the log of LWP and log of N_d using each composite: **All** (a), **Q06** (b), and **G18** (c) for each 1° region of the globe for the 5 year period. Histograms showing the impact of sampling on each composite of τ_c (d), R_e (e), LWP (f), and N_d (g) are displayed.

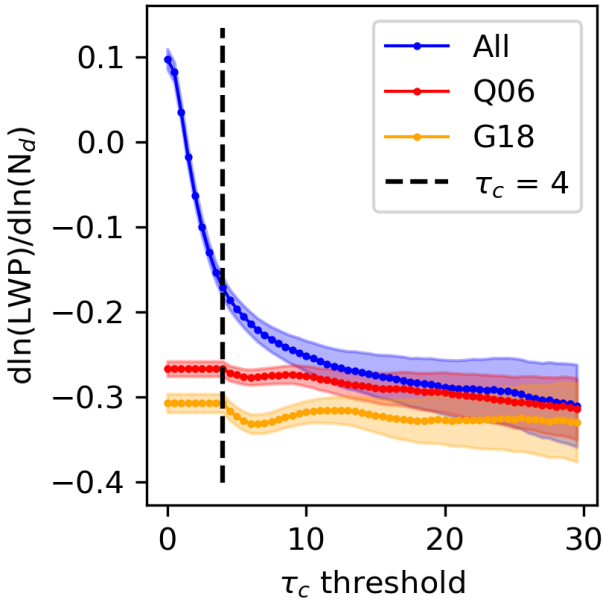

Figure 6. Linear least squares fit between the log of LWP and log of N_d using 5 years of 1° gridded data at the ENA site as a function of increasing the threshold of τ_c. Slope means (circle) for the All (blue), **Q06** (red), and **G18** (orange) composites are displayed along with the τ_c threshold of 4 used ubiquitously in **Q06** and **G18**.

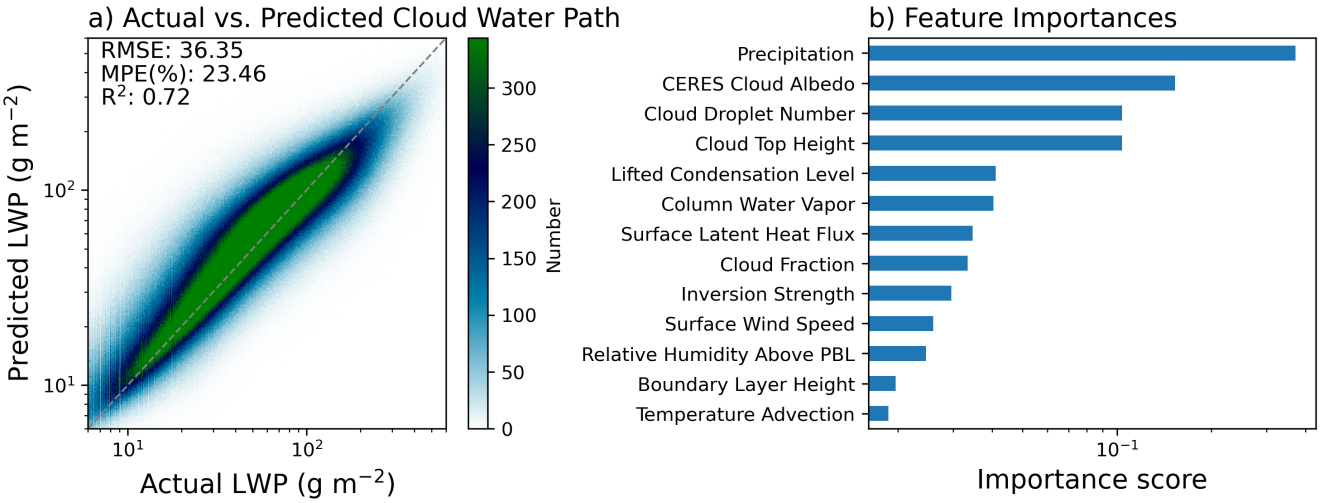

Figure 7. Random forest model predictions of LWP compared to observed LWP using a 25% holdout testing dataset for the California region at 0.05° grid resolution, with listed values of the Root Mean Square Error (RMSE), Mean Percentage Error (MPE), and Pearson's R-squared correlation coefficient (R^2) (a). The relative importance described by Breiman (2001) for each cloud-controlling factor is displayed in (b).

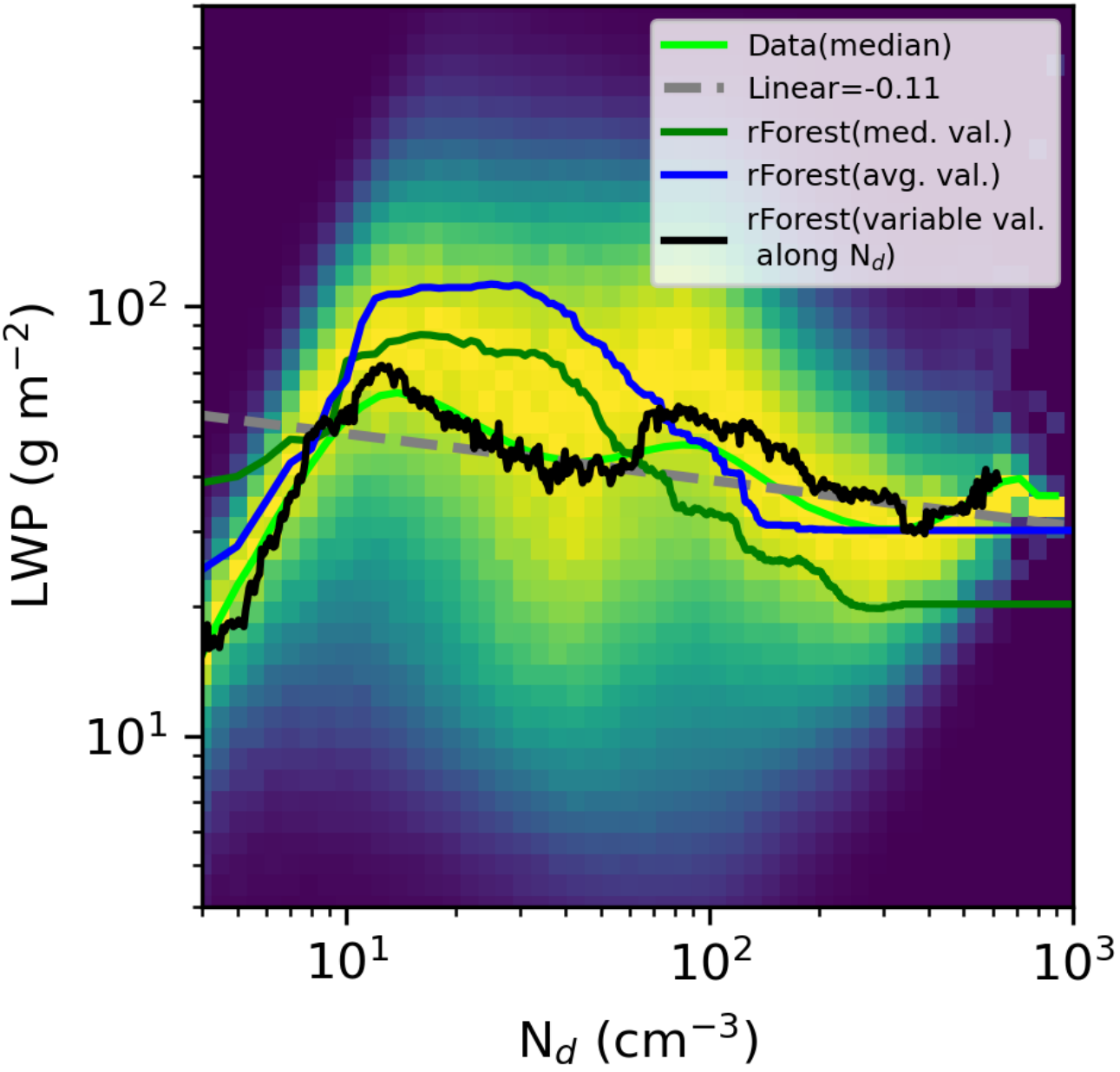

Figure 8. The N_d-LWP relationship ~~is composited by precipitation rate into non-raining clouds ($0 < Pr < 0.05$ mm/hr;~~ displayed as a 2D histogram normalized by the maximum number of retrievals in each N_d bin, ~~b~~using 5 years of MODIS cloud retrievals aggregated at 0.05° resolution for California. The light green line represents the median of the actual data distribution, ~~e~~while the dark green, ~~d)~~blue, ~~drizzle (e, f, g, h),~~ and ~~raining clouds~~ black lines correspond to random forest (~~$0.05 < Pr < 0.2$ mm/hr; i, j, k, l~~rForest) ~~. Within these composites~~model predictions of LWP based on single median values, ~~the data is further divided by low cloud albedo ($0 < A_c < 0.25$) and higher cloud albedo ($0.25 < A_c < 1.0$)~~single mean values, and ~~finally by low relative humidity above~~ varying the ~~boundary layer (dry~~median of all cloud-controlling factors within each N_d bin, ~~red labels) and high relative humidity (moist, blue labels)~~respectively . ~~An OLS~~ Linear least squares fit ~~to the observational data~~ (~~dashed~~ gray dashed line) and ~~to the random forest prediction (solid blue line), along with the average~~ associated slope ~~estimated by numerical differentiation of the prediction using finite differences, are displayed~~value is provided.

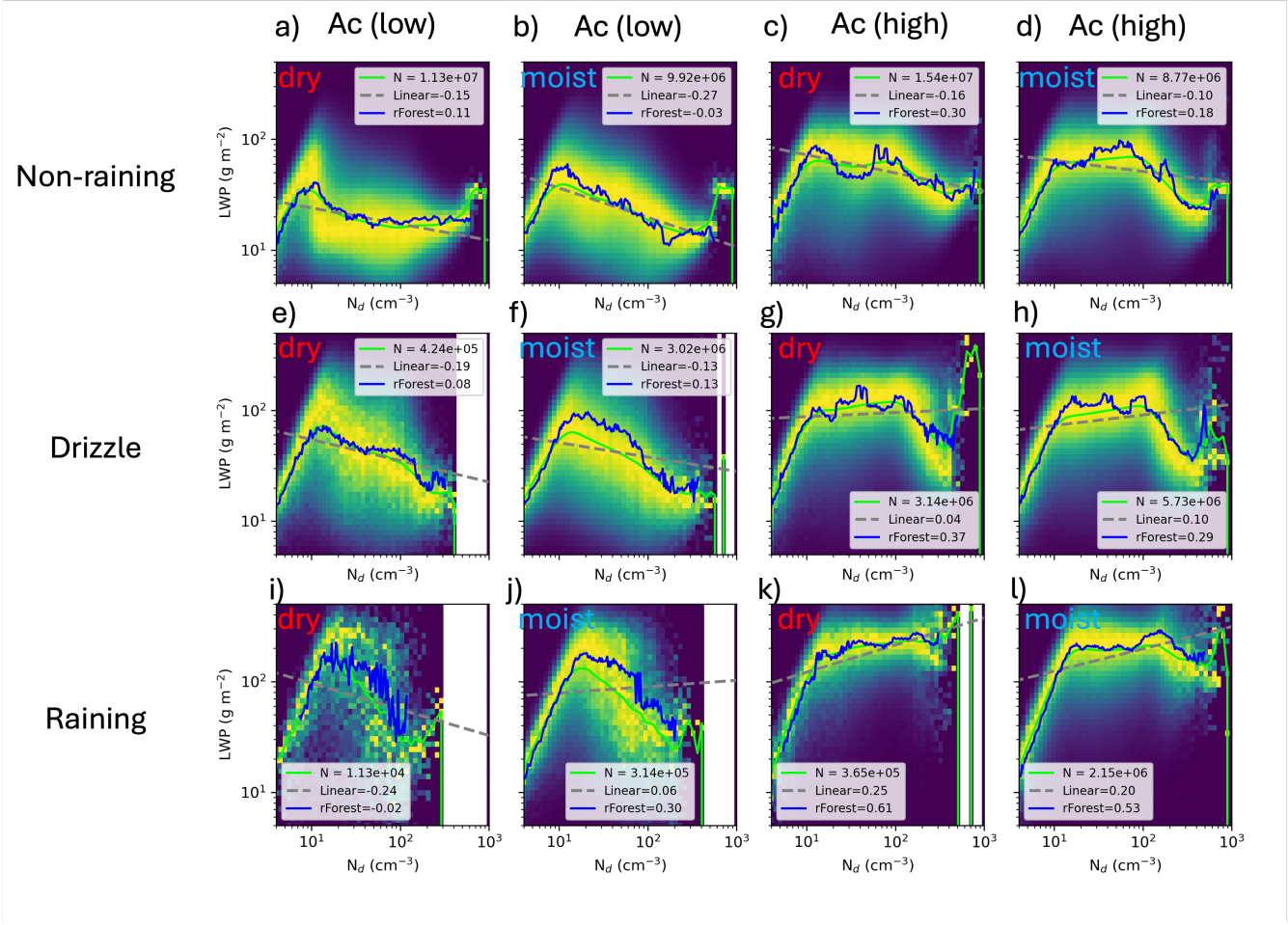

Figure 9. The N_d-LWP relationship is composited by precipitation rate into non-raining clouds ($0 < \mathrm{Pr} < 0.05$ mm/hr; a, b, c, d), drizzle (e, f, g, h), and raining clouds ($0.05 < \mathrm{Pr} < 0.2$ mm/hr; i, j, k, l) for the California region. Within these composites, the data is further divided by low cloud albedo ($0 < A_c < 0.25$) and higher cloud albedo ($0.25 < A_c < 1.0$), and finally by low relative humidity above the boundary layer (dry, red labels) and high relative humidity (moist, blue labels). An OLS fit to the observational data (dashed gray line) and to the random forest prediction (solid blue line), along with the average slope estimated by numerical differentiation of the prediction using finite differences, are displayed.

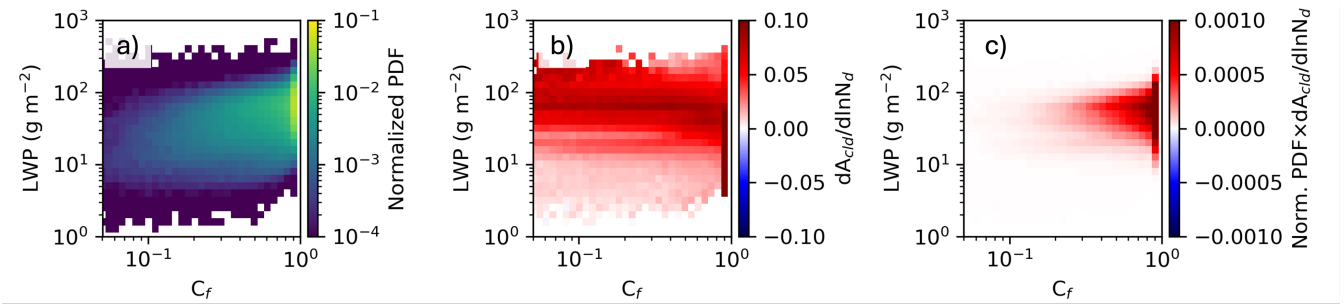

Figure 10. Radiative kernel for computing the Twomey effect using predictions of A_{cld} from the random forest model for the California region using 1° grid-spacing data. Number of retrievals falling into log-bins of LWP and CF are normalized by the total (a), sensitivity of changes in A_{cld} for each bin determined using finite differences of the random forest predictions (b), and resulting Twomey radiative kernel (c) given by the product of (a) and (b).

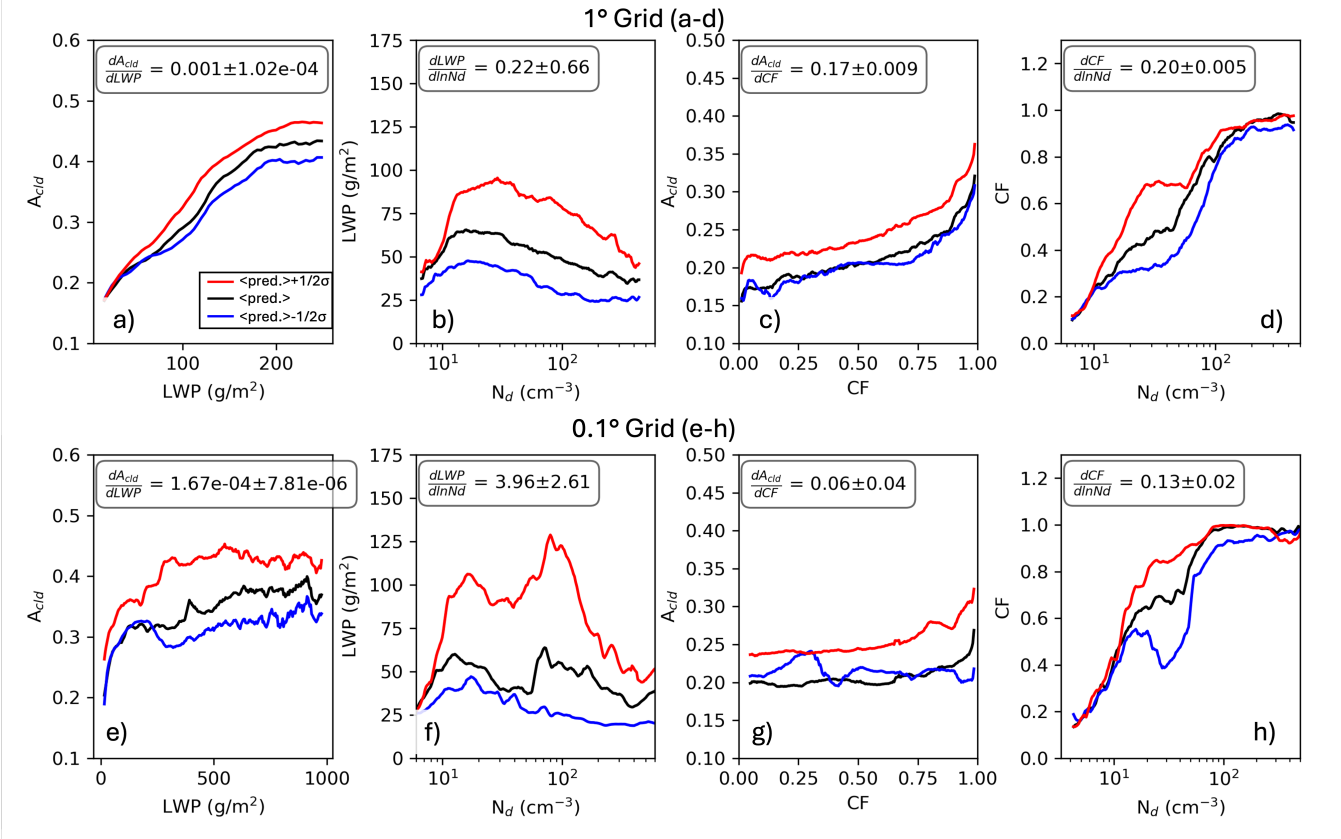

Figure 11. ACI relationships over the California region used to compute liquid water path and cloud fraction adjustments using 1° (a-d) and 0.1° (e-h) grid-resolutions for the relationship between LWP-A_{cld}, N_d-LWP, CF-A_{cld}, and N_d-CF predicted using the random forest model. Three different curves represent the relationship using predictions from the median of the cloud controlling factors (black), median $+0.5$ standard deviation (red), and median -0.5 standard deviation (blue) of the cloud controlling variables. Average and standard deviation of the three slopes estimated by numerical differentiation of the prediction using finite differences are provided.

Table 1. Performance of the random forest model for predicting LWP over the California region, evaluated using the Pearson's coefficient (R^2), mean percentage error (MPE), and the top six variables ranked by importance from highest to lowest.

Resolution	R^2	MPE (%)	Importance Order
5°	0.84	4.9	A_{cld}, N_d, Pr, RH, LCL, TQV
1°	0.89	7.4	Pr, A_{cld}, N_d, TQV, CTH, C_f
0.5°	0.88	9.9	Pr, A_{cld}, N_d, CTH, TQV, C_f
0.1°	0.75	21.9	Pr, A_{cld}, N_d, CTH, C_f, LCL
0.05°	0.66	27.5	Pr, A_{cld}, N_d, CTH, LCL, TQV

~~Performance of the random forest model for predicting LWP, evaluated using the Pearson's coefficient (R^2), mean percentage error (MPE), and the top six variables ranked by importance from highest to lowest.~~

Table 2. List of cloud and radiative effects from aerosol perturbations at increasing grid-resolution for subtropical ~~clouds~~regions (California, Peruvian, Namibian, and Australian). Radiative scaling is defined as ~~$(-1.) * CF * F^{\downarrow} * \phi_{atm} * \frac{d\ln Nd}{d\ln AI} * d\ln AI$~~ $(-1.) * CF * \overline{F^{\downarrow}} * \frac{d\ln N_d}{d\ln AI} \overline{\Delta \ln AI}$.

	Grid Resolution			
	5°	1°	0.5°	0.1°
Twomey [W/m^2]	-0.56±0.24	-0.96±0.28	-1.09±0.20	-0.90±0.16
LWP Adjustment [W/m^2]	0.14±0.09	-0.01±0.01	-0.05±0.02	-0.03±0.02
CF Adjustment [W/m^2]	-0.51±0.13	-0.49±0.09	-0.36±0.05	-0.11±0.05
RF Forcing [W/m^2]	-0.93±0.43	-1.46±0.35	-1.50±0.23	-1.04±0.19
Cloud Fraction	0.59±0.04	0.58±0.04	0.62±0.04	0.76±0.03
Radiative Scaling [W/m^2]	-18.29±1.10	-17.81±1.32	-19.17±1.31	-23.43±1.07
$dA_{cld}/dLWP$ [m^2/g]	0.001±2.14e-04	9.48e-04±1.15e-04	6.64e-04±4.66e-05	1.55e-04±1.67e-05
$dLWP/d\ln N_d$ [g/m^2]	-5.83±3.34	0.89±1.00	3.79±1.47	7.43±6.42
d_{Acld}/dCF	0.17±0.03	0.15±0.02	0.10±0.007	0.03±0.01
$dCF/d\ln N_d$	0.17±0.03	0.19±0.02	0.19±0.003	0.14±0.009