Supplement of Atmos. Chem. Phys., 26, 59–76, 2026
https://doi.org/10.5194/acp-26-59-2026-supplement

Supplement of

Machine learning reveals strong grid-scale dependence in the satellite N_{d}–LWP relationship

Matthew W. Christensen et al.

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

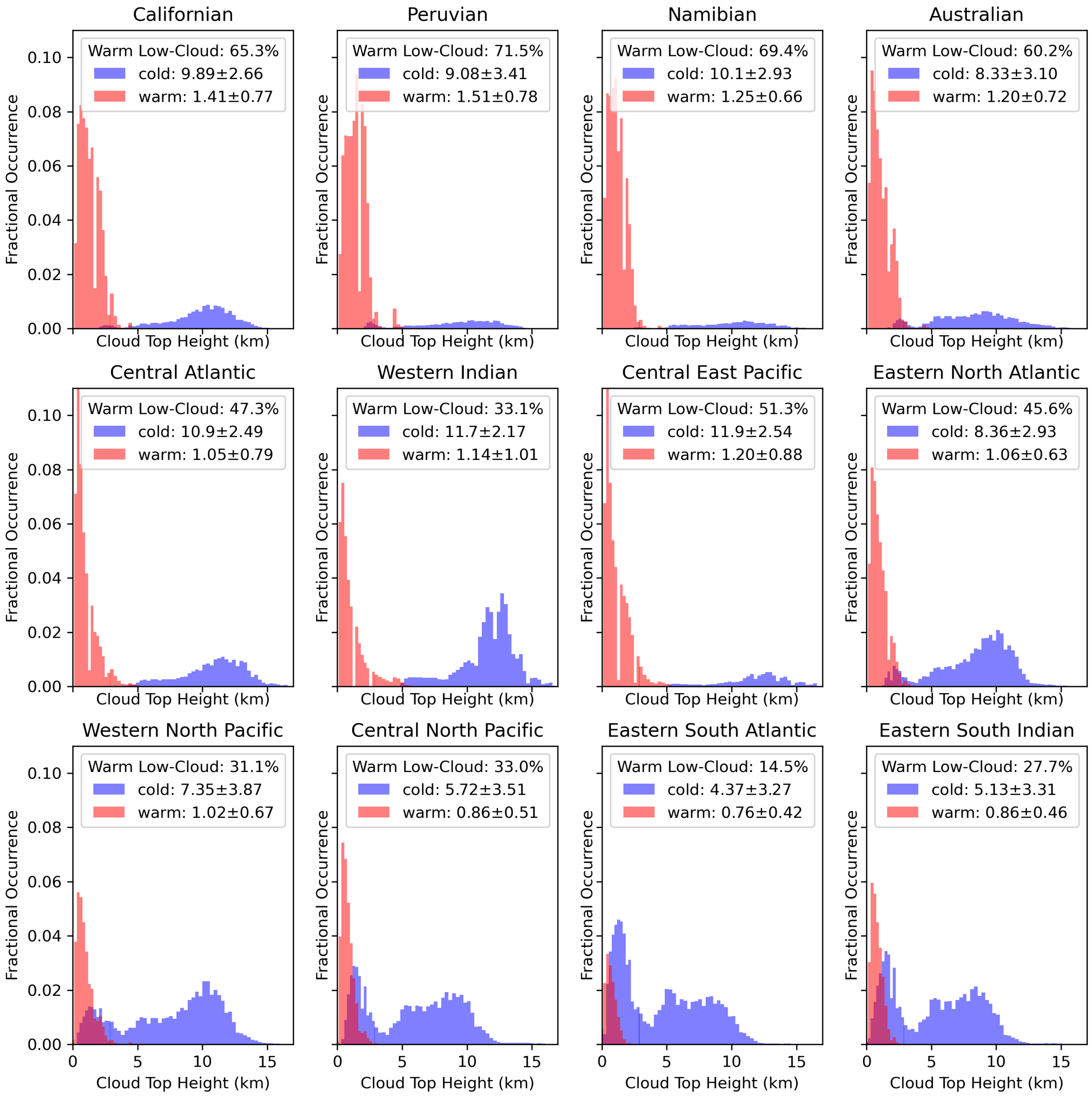

Figure S1. Histogram showing the fractional occurrence of the number of MODIS-retrieved cloud top heights grouped into 75 bins, divided by the total number of possible L2 pixels, sorted by warm (CTT > 273 K) and cold (CTT < 273 K) cloud top temperatures for each region in this study. The percentage of warm low-level clouds below 3 km, along with the mean and standard deviation of the cloud top heights for each distribution, are provided.

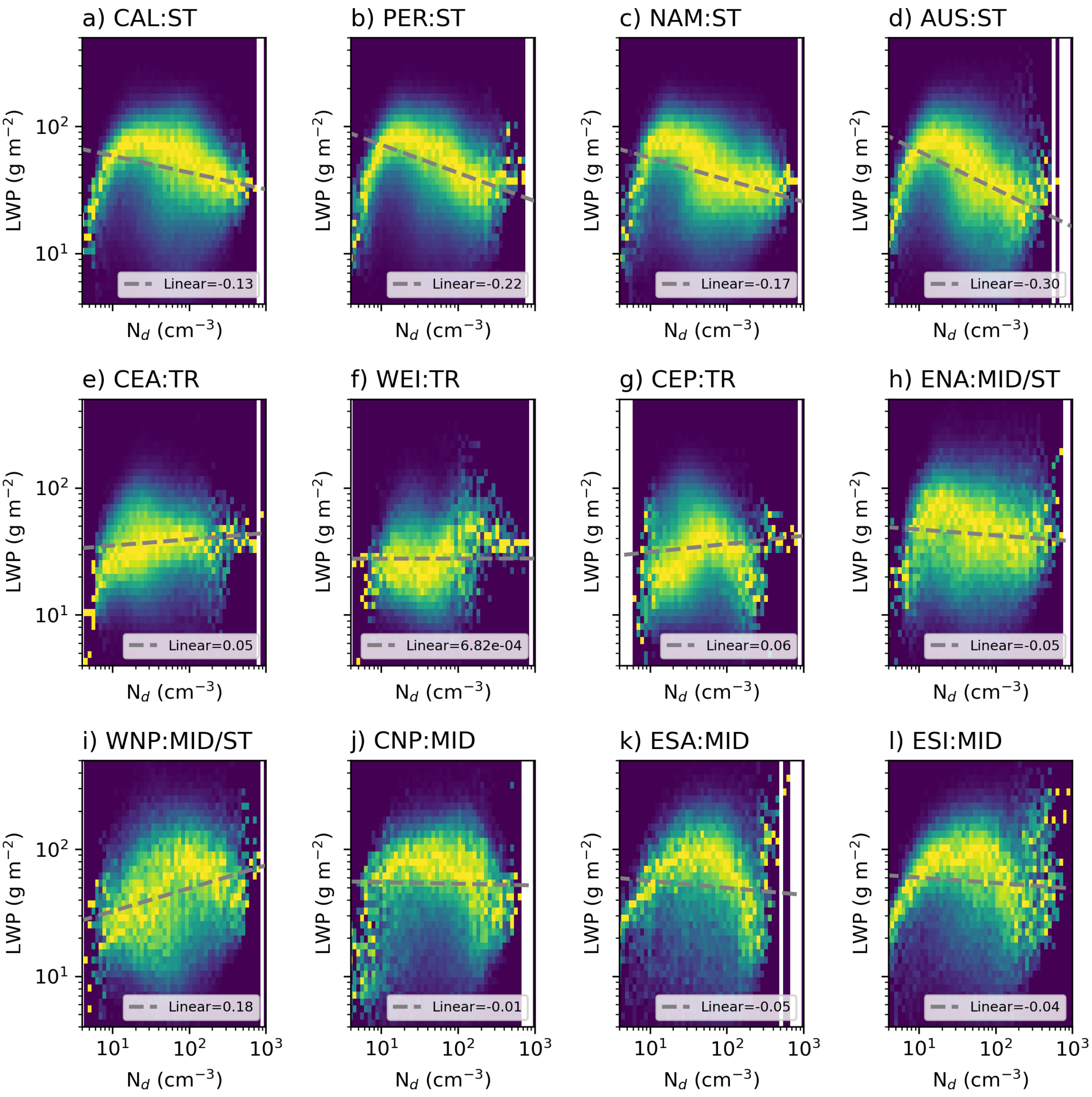

Figure S2. The N_d–LWP relationship expressed using a 2D histogram of the joint frequencies of the LWP and N_d normalized by the bin number of N_d using 5 years of MODIS cloud retrievals aggregated at 1° spatial scale resolution over 20°x 20° domain for each region described in Figure 1 of this study. Subtropical (ST), Tropical (TR), Midlatitude (MID), and mixed (MID/ST) regions are denoted followed by the prefix name of the region. Linear least squares fit (gray dashed line) and associated slope values are provided.

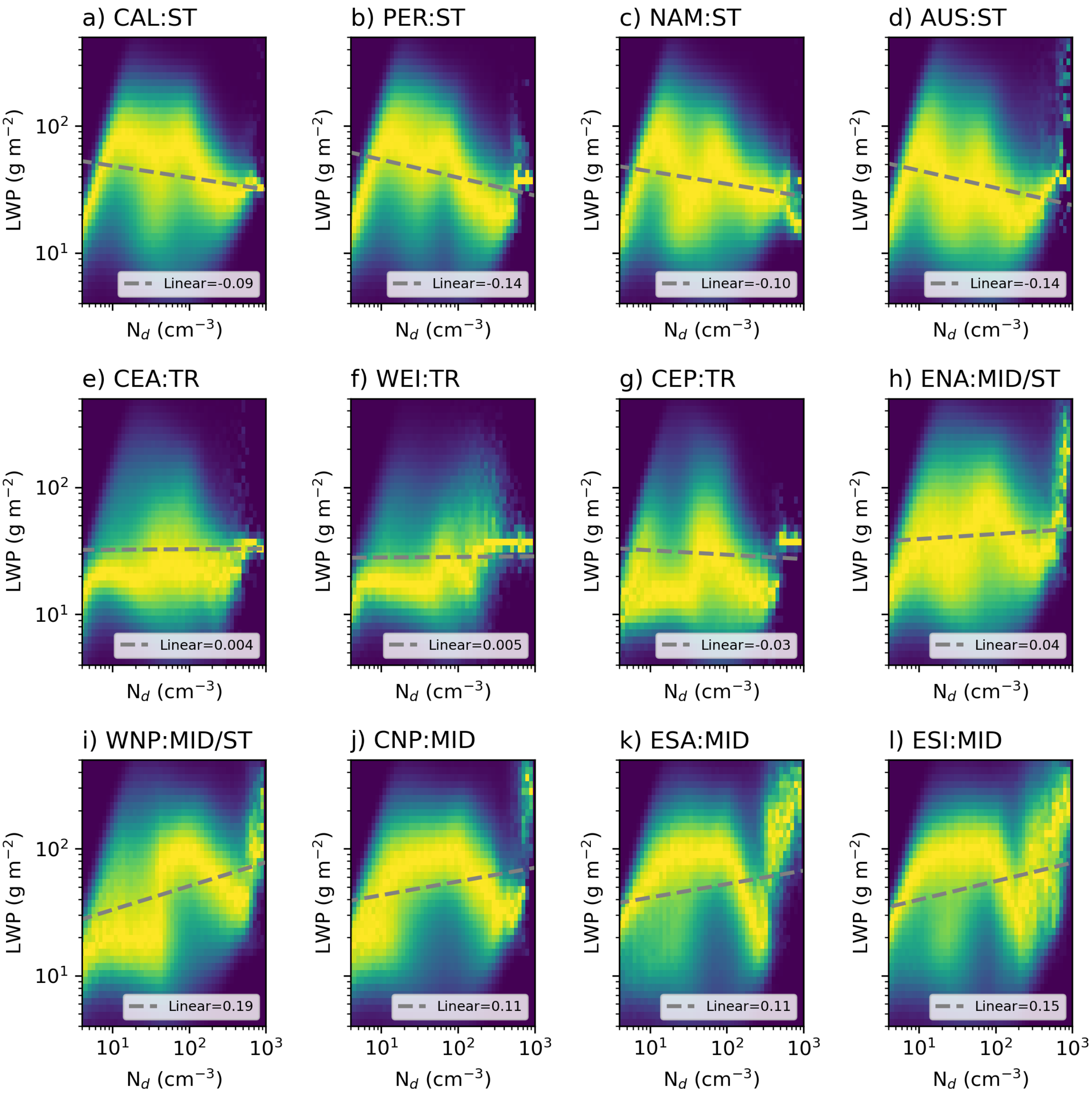

Figure S3. Same as Figure SFigE3 except using 0.1° spatial resolution data.

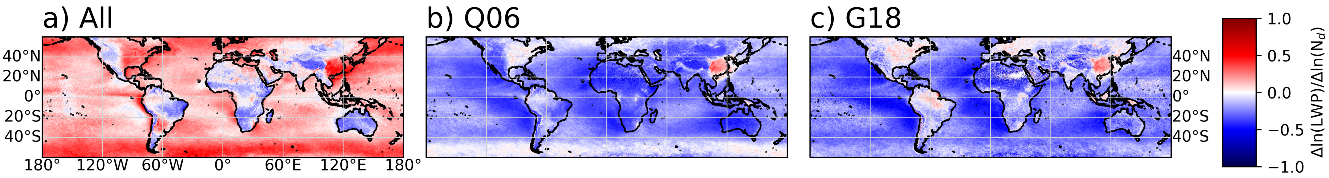

Figure S4. Linear least squares fit between the log of LWP and log of N_d using each composite: all a), **Q06** b), and **G18** c) for each 0.1° region of the globe for the 5 year period.

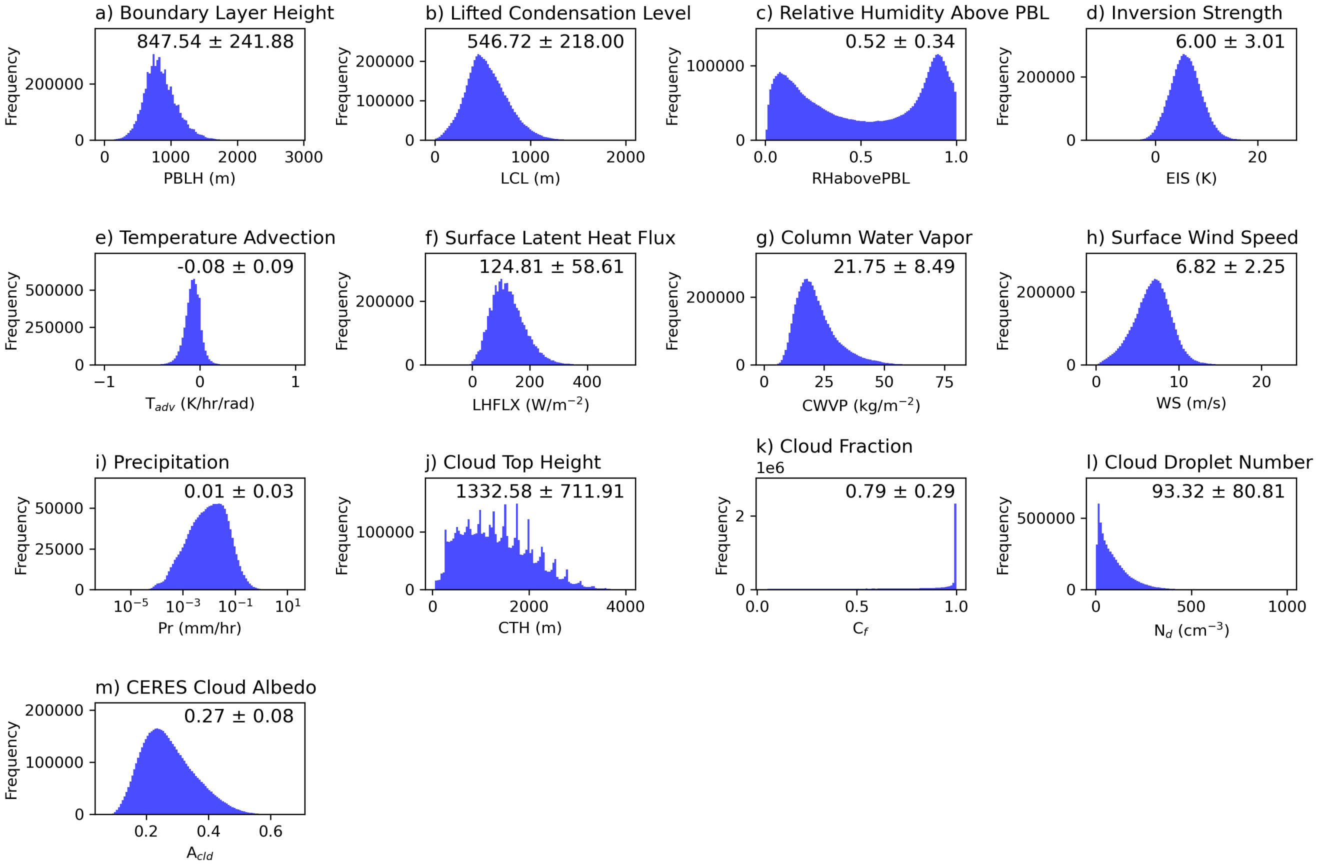

Figure S5. Histograms of the input predictor variables (a – o) into the random forest model shown for 5 years of 1degree data over the California region. Means and standard deviations are provided on each sub-plot.

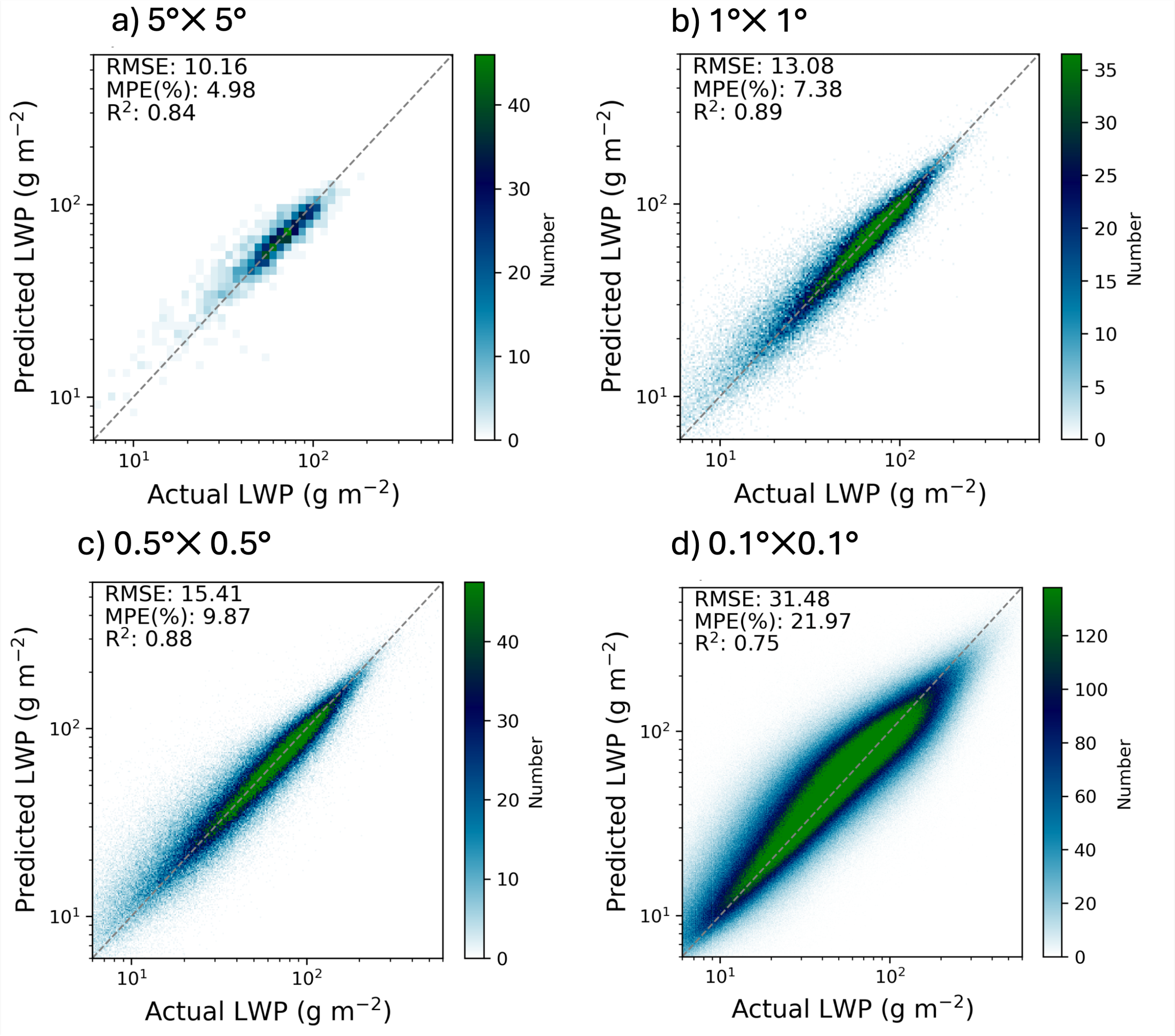

Figure S6. Random forest model predictions of LWP as a function of N_d for 5° (a), 1° (b), 0.5° (c), and 0.1° (d) spatial resolutions over the California region.

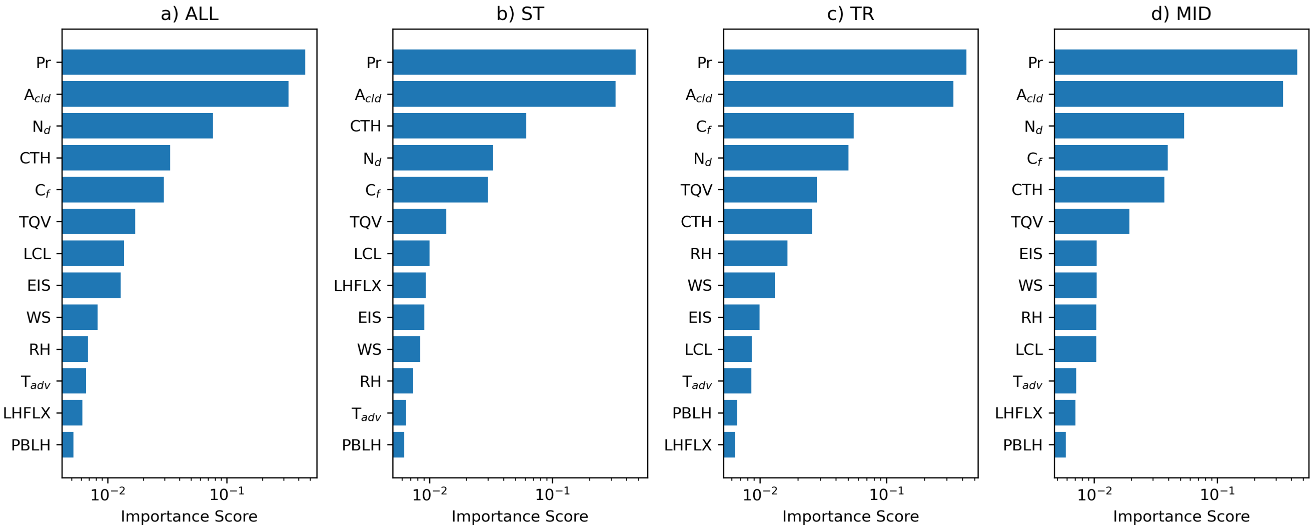

Figure S7. Random forest feature importance score in the N_d-LWP relationship shown for the combined subtropical (ST), tropical (TR), and midlatitude (MID) locations in Figure 1. Values are normalized and summed for each region.

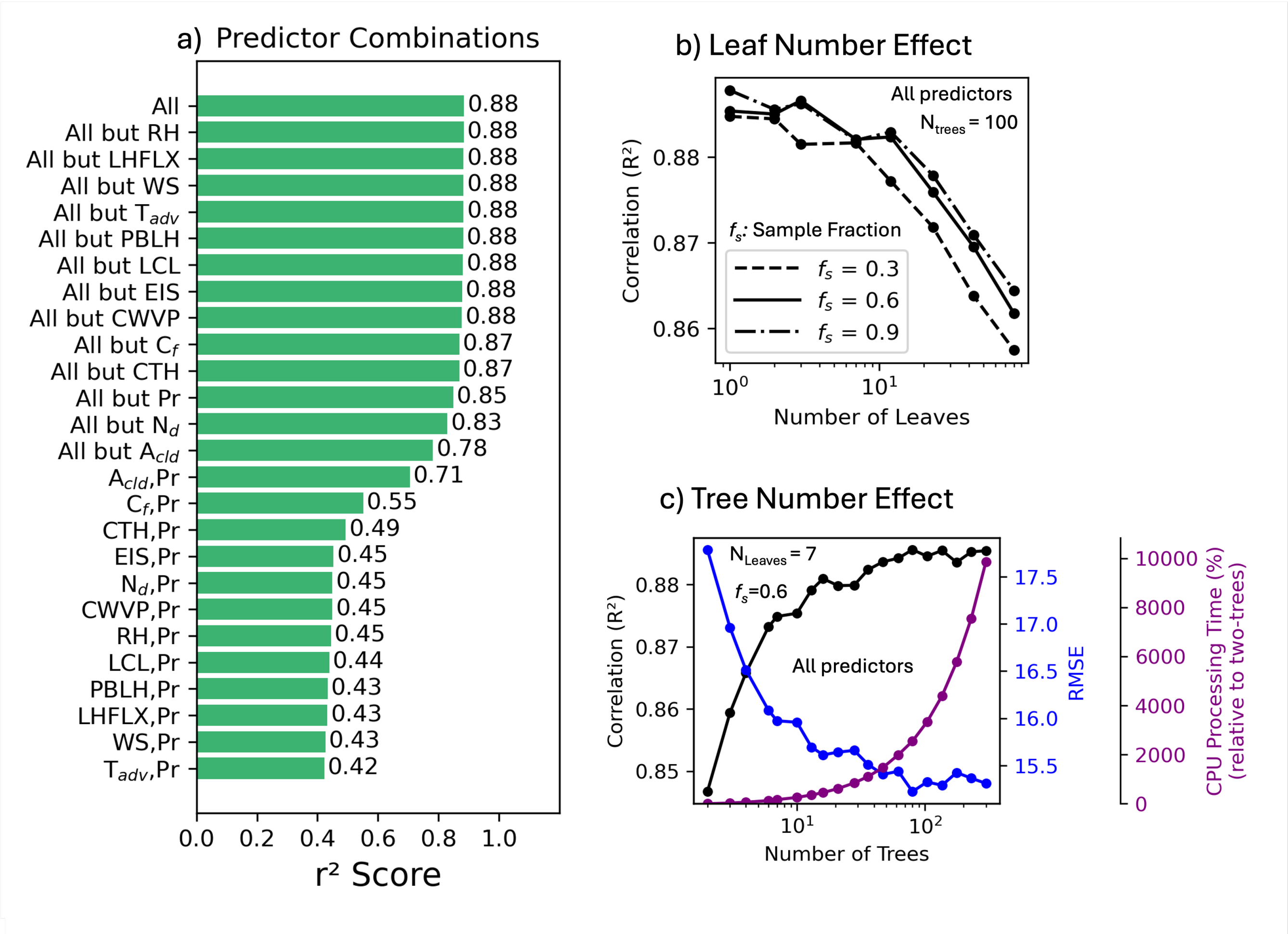

Figure S8. Sensitivity of model performance to hyperparameter and predictor selection at 1° resolution. Coefficient of determination (r^2) for models trained using different predictor combinations, where All refers to all 13 cloud controlling variables (a). Impact of the minimum number of samples per leaf node on r^2 for sample fractions of 0.3 (dashed), 0.6 (solid), and 0.9 (dashed-dotted). Effect of the number of trees on RMSE, r^2, and CPU processing time normalized by training the model with 2-trees (c) over the California region.

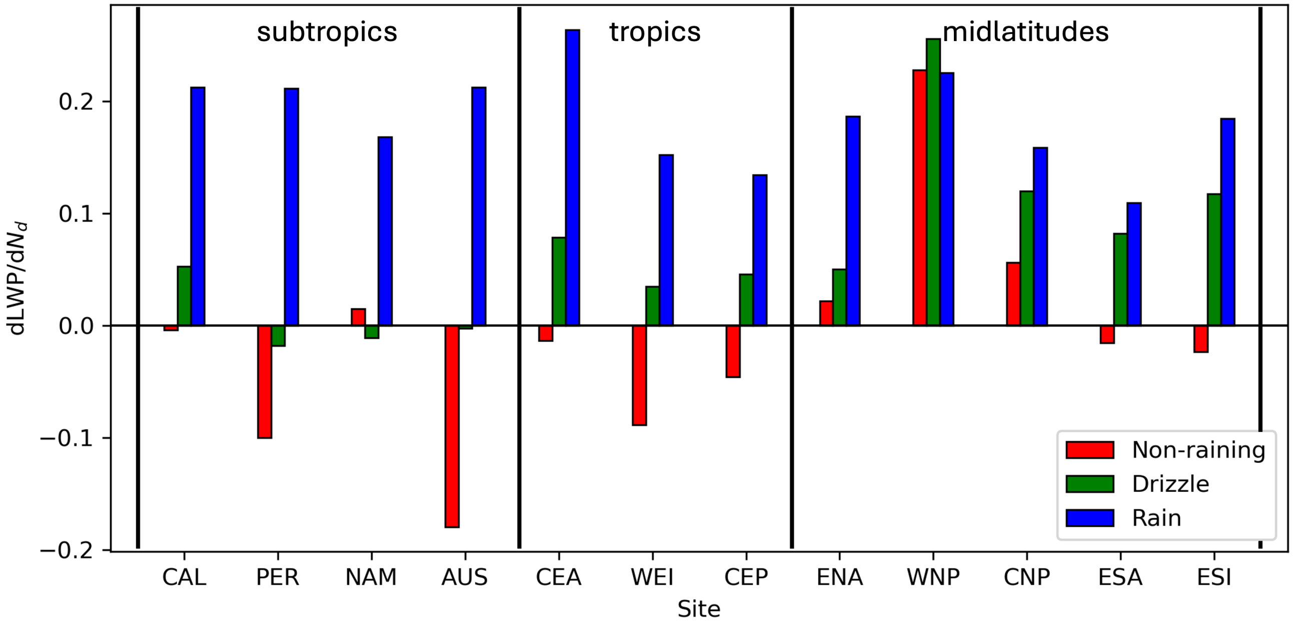

Figure S9. Linear least squares fit of the N_d-LWP MODIS relationship using the 0.1° product displayed for each region where the data was considered non-raining (Pr = 0 mm/hr; red), drizzling (0<Pr<0.05 mm/hr; green), and raining (0.05< Pr < 2.0 mm/hr; blue) from AMSRE retrievals.

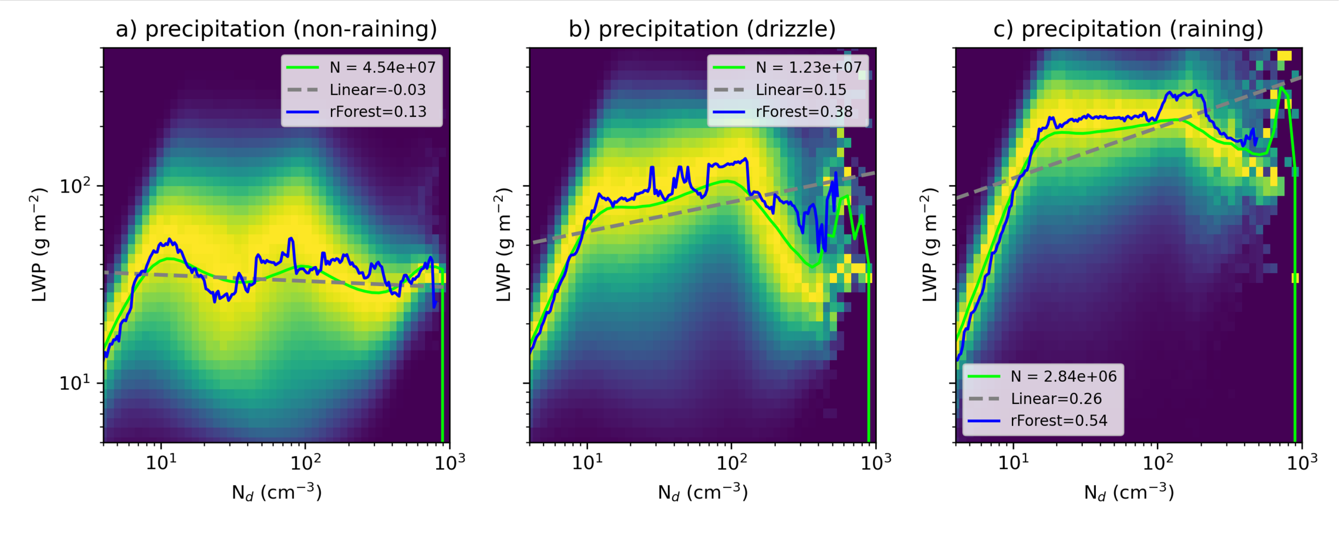

Figure S10. The N_d–LWP relationship over the California region using the 0.05° gridded data, using the same criteria as Figure 8, but composited by surface precipitation rate: (a) 0 mm/hr (non-raining clouds), (b) 0 < Pr < 0.05 mm/hr (drizzle), and (c) 0.05 < Pr < 0.2 mm/hr (rain). An ordinary least squares (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 shown for constant surface precipitation values: 0 mm/hr (non-raining), 0.01 mm/hr (drizzle), and 0.1 mm/hr (rain). Note, the remaining cloud controlling variables are allowed to change as a function of N_d.

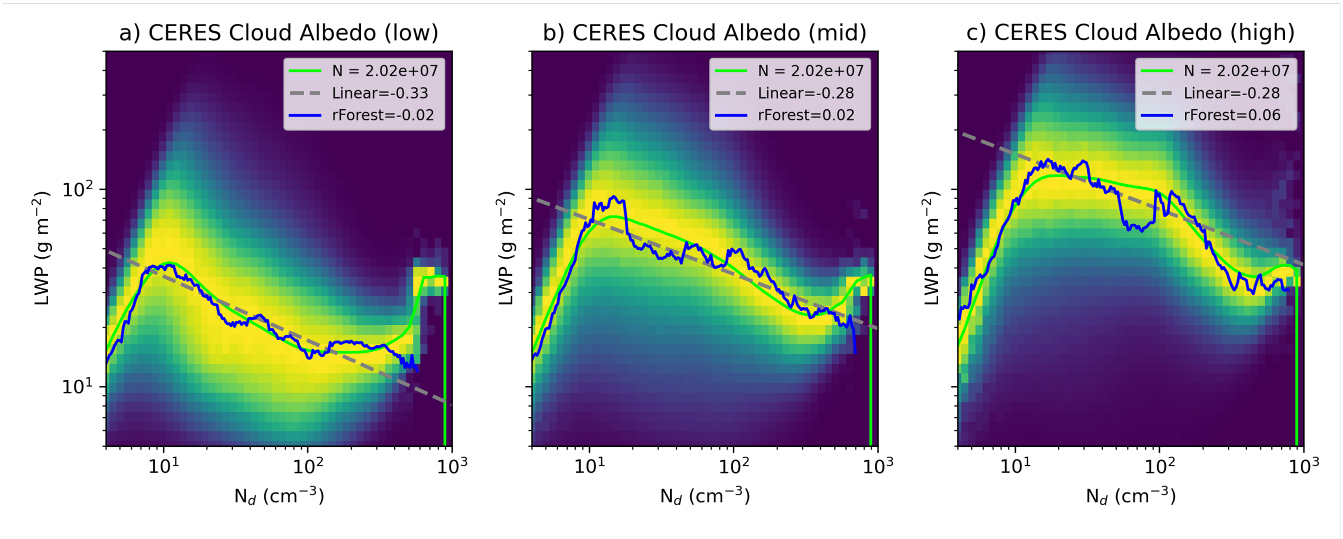

Figure S11. The N_d-LWP relationship over the California region using the 0.05° gridded data and the same criteria as Figure 8 but composited by CERES cloud albedo, low cloud albedo (0<A$_{cld}$<0.25) (a), mid cloud albedo (0.25<A$_{cld}$<0.4) (b), and high cloud albedo (0.4<A$_{cld}$<1) (c). An OLS 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 shown for constant cloud albedo values: 0.2 for low, 0.3 for mid, and 0.325 for high. Note, the remaining cloud controlling variables are allowed to change as a function of N_d.

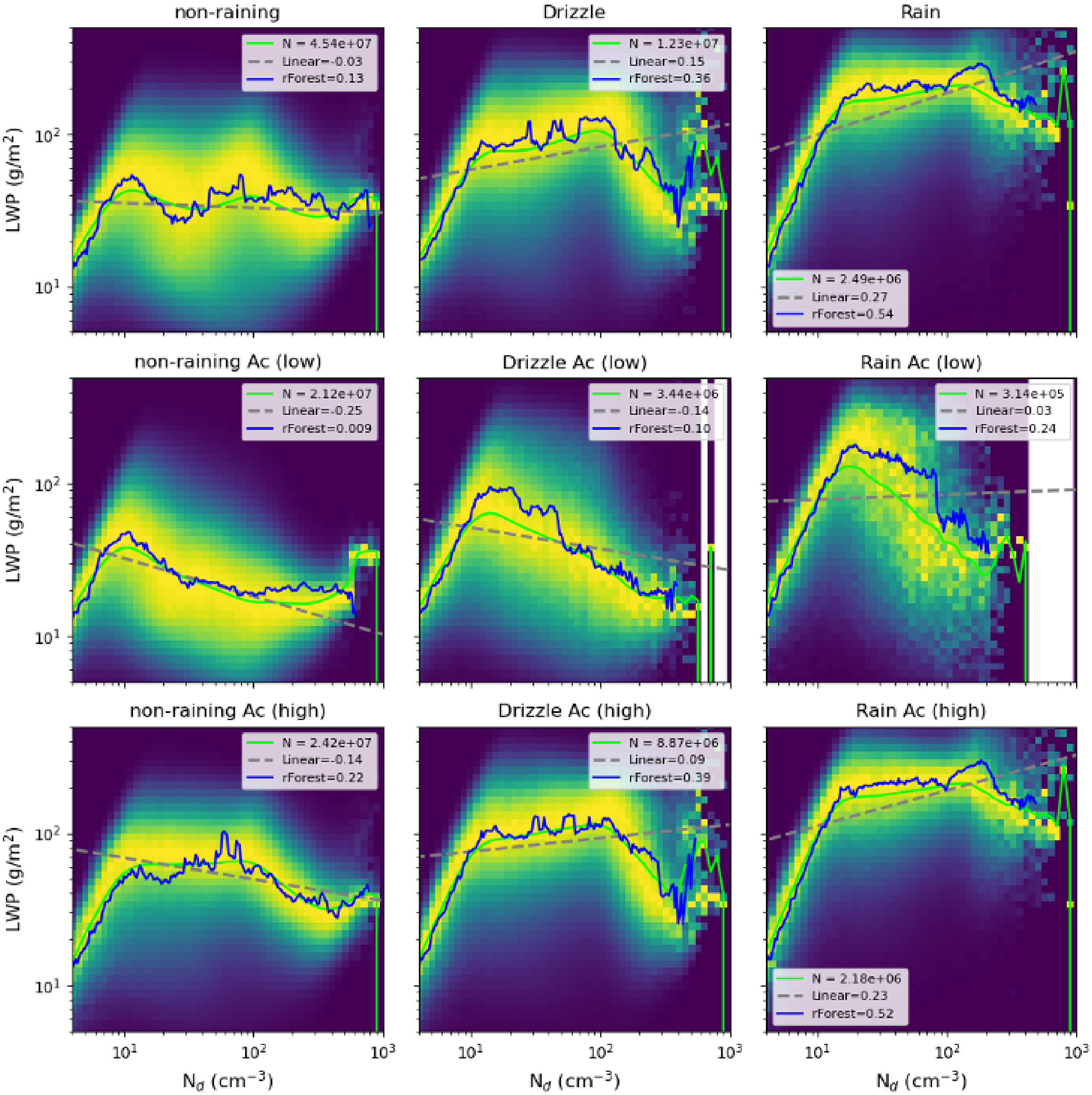

Figure S12. The N_d-LWP relationship over the California region using the 0.05° gridded data and the same criteria as Figure 8 but composited by precipitation rate, 0 mm/hr (non-raining clouds) (a), 0 < Pr < 0.05 mm/hr (drizzle) (b), and 0.05 < Pr < 0.2 (rain) (c) and with each regime further composited by low cloud albedo for non-raining (d), drizzling (e), and raining(f), and high cloud albedo for non-raining (g), drizzling (h), and raining (i). 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, is displayed.

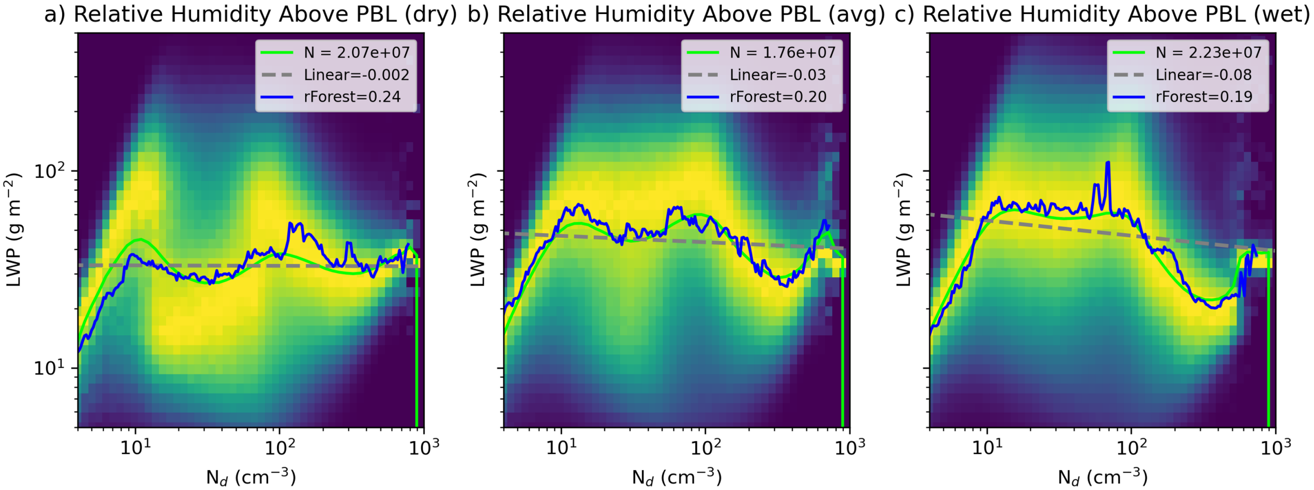

Figure S13. The N_d-LWP relationship over the California region using the 0.05° gridded data and the same criteria as Figure 8, but composited by relative humidity above the boundary layer into low ($0 < \mathrm{RH} < 0.33$) (a), mid ($0.33 < \mathrm{RH} < 0.67$) (b), and high ($0.67 < \mathrm{RH} < 1.0$) (c) levels. 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 shown for constant relative humidity values: 0.25 for dry, 0.5 for avg, and 0.9 for wet. Note, the remaining cloud controlling variables are allowed to change as a function of N_d.

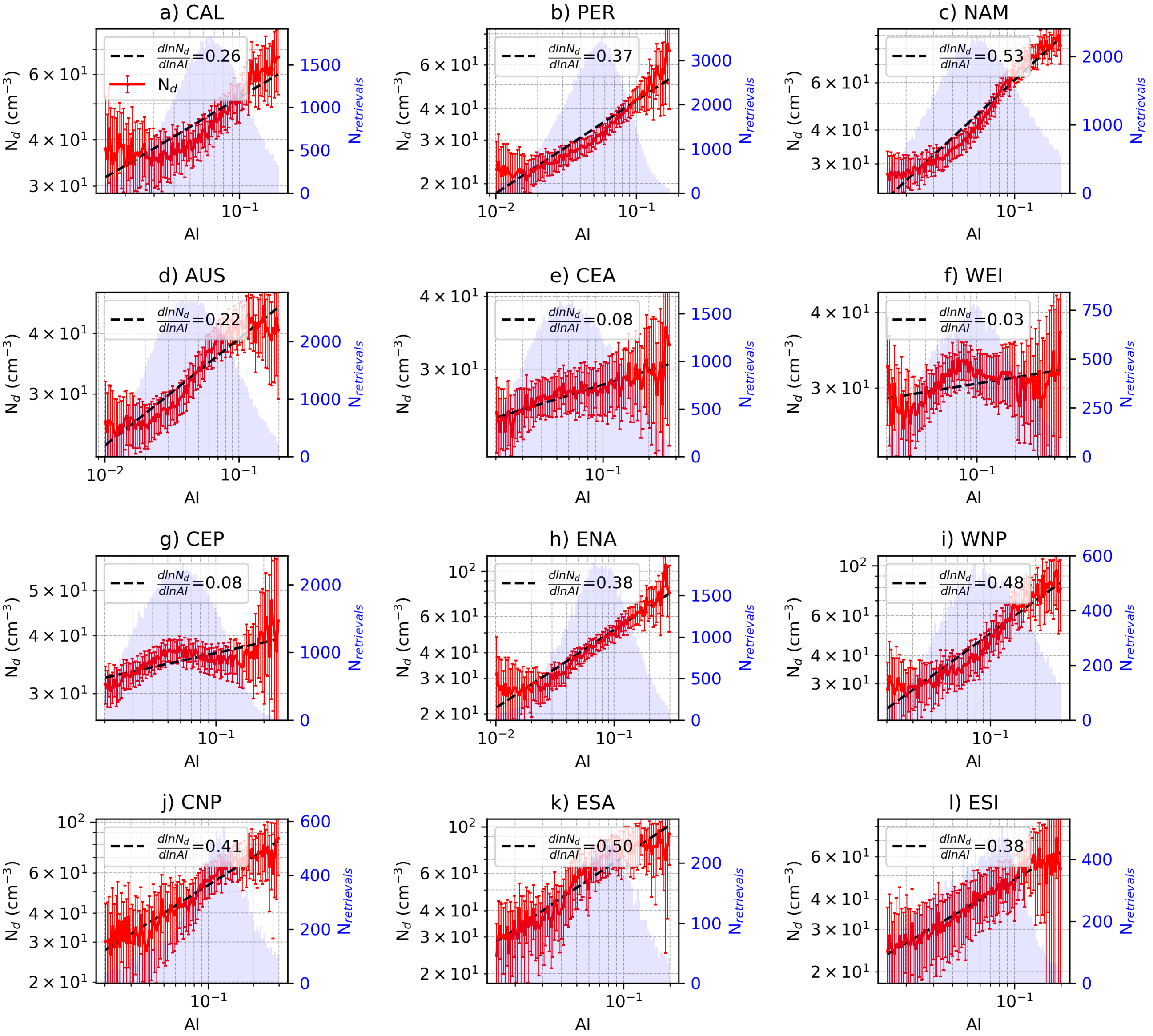

Figure S14. The $AI - N_d$ relationship in each region (a-l) computed from MODIS retrievals using five years of the 1° gridded product. Linear least squares fit (black dashed line) and slopes ($\frac{d \ln N_d}{d \ln AI}$) are provided.

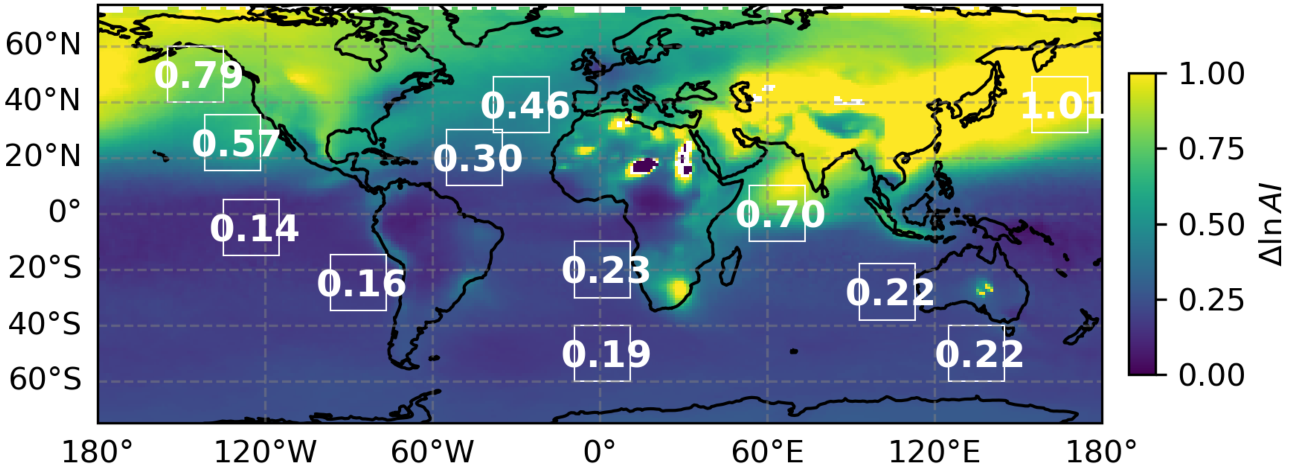

Figure S15. Change in aerosol index between present day and pre-industrial day conditions. Average values for each region are displayed in white.

Table S1. Overview of predictor variables and their respective data sources.

Predictor Variable	Source / Dataset
Planetary Boundary Layer Height (PBLH)	Reanalysis data
Lifted Condensation Level (LCL)	Calculated from reanalysis data
Relative Humidity above PBL Height (rhAbovePBL)	Calculated from reanalysis data
Estimated Inversion Strength (EIS)	Calculated from reanalysis data
Surface Temperature Advection (Tadv)	Calculated from reanalysis data
Surface Latent Heat Flux (LH)	Reanalysis data
Total Column Water Vapor (tqv)	Reanalysis data
10-m Surface Wind Speed (ws10)	Reanalysis data
Surface Precipitation	AMSR-E satellite retrieval
Cloud Top Height (CTH)	MODIS satellite retrieval
Cloud Fraction (CF)	MODIS satellite retrieval
Cloud Albedo (A_{cld})	CERES satellite retrieval
Cloud Droplet Number Concentration (N_d)	Calculated from MODIS satellite retrieval

Table S2. Overview of random forest hyperparameters and associated values used in this study.

Hyperparameter	Value	Description
Number of trees	100	Total number of decision trees in the random forest ensemble. A larger number improves stability and accuracy but increases computational cost.
Minimum leaf size	7	Minimum number of samples required to form a terminal leaf node. Controls model complexity; smaller values allow deeper trees that may capture finer variability but risk overfitting.
Sample fraction	0.6	Fraction of the training data randomly sampled (with replacement) to train each tree, defining the bootstrap sample size and influencing model diversity.

Table S3. Random forest predictions of LWP for each region using the 0.1°-resolution dataset. Shown are the number of samples (N$samples$), the linear least-squares fit of dLWP/dNd for non-raining and raining conditions, the Pearson correlation coefficient (R^2), the mean percentage error (MPE), and the top three predictor variables ranked by importance (from highest to lowest).

Region	$N_{samples}$	dLWP/dlnN$_d$	dLWP/dlnN$_d$	R^2	MPE (%)	Importance Order
		(non-raining)	(raining)			
CAL	2.26e+07	-0.003	0.24	0.75	22.0	Pr,A$_{cld}$,N$_d$
PER	2.32e+07	-0.06	0.33	0.71	26.0	Pr,A$_{cld}$,CTH
NAM	2.34e+07	0.04	0.24	0.74	20.7	Pr,A$_{cld}$,CTH
AUS	2.03e+07	-0.08	0.29	0.74	27.7	Pr,CTH,N$_d$
CEA	1.04e+07	-0.02	0.29	0.71	31.2	Pr,CTH,N$_d$
WEI	5.19e+06	-0.06	0.15	0.68	32.5	Pr,CTH,N$_d$
CEP	1.32e+07	-0.03	0.16	0.75	29.5	Pr,CTH,N$_d$
ENA	1.42e+07	0.08	0.31	0.73	32.7	Pr,A$_{cld}$,CTH
WNP	9.64e+06	0.24	0.33	0.78	26.0	Pr,A$_{cld}$,C$_f$
CNP	1.16e+07	0.12	0.30	0.77	25.2	Pr,A$_{cld}$,TQV
ESA	6.77e+06	0.07	0.29	0.80	23.3	Pr,A$_{cld}$,C$_f$
ESI	1.02e+07	0.10	0.34	0.75	29.5	Pr,A$_{cld}$,CTH

Table S4. Performance of the random forest model applied to the California region for predicting cloud albedo, 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	1.02	CF, LWP, N$_d$, CTH, TQV, RH
1°	0.88	1.50	CF, LWP, N$_d$, TQV, CTH, INV
0.5°	0.87	1.94	CF, LWP, N$_d$, TQV, CTH, INV
0.1°	0.75	2.76	CF, LWP, N$_d$, CTH, TQV, PR
0.05°	0.71	3.07	LWP, N$_d$, CTH, TQV, LCL, PR

Table S5. Performance of the random forest model applied to the California region for predicting cloud fraction, 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°	0.7	39.58	A$_{cld}$, N$_d$, LWP, RH, T$_{adv}$, LCL
1°	0.73	46.88	A$_{cld}$, N$_d$, LWP, CTH, LCL, RH
0.5°	0.72	38.42	A$_{cld}$, N$_d$, LWP, CTH, LCL, RH
0.1°	0.69	18.62	CTH, N$_d$, LCL, LWP, A$_{cld}$, RH
0.05°	0.61	6.57	CTH, N$_d$, LCL, LWP, A$_{cld}$, HFLX

Table S6. List of cloud and radiative effects from aerosol perturbations at increasing grid-resolution for midlatitude regions (CNP, ESA, and ESI) clouds. Radiative scaling is defined as $(-1.) * cf * F^{\downarrow} * \phi_{atm} * \frac{d\ln Nd}{d\ln AI} * d\ln AI$.

	Grid Resolution			
	5°	1°	0.5°	0.1°
Twomey [W/m^2]	-0.83±1.04	-1.16±0.35	-1.40±0.27	-1.21±0.18
LWP Adjustment [W/m^2]	0.05±0.12	-0.10±0.07	-0.21±0.12	-0.20±0.07
CF Adjustment [W/m^2]	-0.38±0.18	-0.42±0.12	-0.27±0.06	-0.10±0.02
RF Forcing [W/m^2]	-1.15±1.31	-1.68±0.44	-1.88±0.21	-1.51±0.13
Cloud Fraction	0.73±0.03	0.71±0.04	0.74±0.04	0.81±0.03
Radiative Scaling [W/m^2]	-22.41±0.94	-21.84±1.08	-22.83±1.14	-24.93±0.96
$dA_{cld}/dLWP$ [m^2/g]	0.001±2.19e-04	0.001±9.24e-05	6.75e-04±7.04e-05	1.94e-04±3.92e-05
$dLWP/d\ln N_d$ [g/m^2]	-1.32±4.49	5.14±4.04	14.2±7.82	40.5±5.54
d_{Acld}/dCF	0.25±0.09	0.12±0.01	0.07±0.01	0.03±0.007
$dCF/d\ln N_d$	0.08±0.05	0.16±0.03	0.16±0.006	0.13±0.008

Table S7. List of cloud and radiative effects from aerosol perturbations at increasing grid-resolution for tropical regions (CEA, WEI, and CEP). Radiative scaling is defined as $(-1.) * cf * F^{\downarrow} * \phi_{atm} * \frac{d\ln Nd}{d\ln AI} * d\ln AI$.

	Grid Resolution			
	5°	1°	0.5°	0.1°
Twomey [W/m^2]	-0.25±0.12	-0.46±0.08	-0.60±0.07	-0.50±0.05
LWP Adjustment [W/m^2]	-0.07±0.006	-0.02±0.04	-0.02±0.04	-0.01±0.005
CF Adjustment [W/m^2]	-0.07±0.05	-0.12±0.02	-0.16±0.04	-0.06±0.03
RF Forcing [W/m^2]	-0.39±0.15	-0.60±0.06	-0.77±0.04	-0.57±0.07
Cloud Fraction	0.30±0.05	0.31±0.04	0.36±0.04	0.54±0.05
Radiative Scaling [W/m^2]	-9.21±1.69	-9.39±1.14	-11.11±1.15	-16.76±1.46
$dA_{cld}/dLWP$ [m^2/g]	0.001±2.04e-04	0.001±1.47e-04	8.47e-04±1.04e-04	2.16e-04±4.88e-05
$dLWP/d\ln N_d$ [g/m^2]	7.03±1.38	2.26±4.14	2.27±4.48	3.84±1.50
d_{Acld}/dCF	0.07±0.03	0.13±0.02	0.11±0.02	0.03±0.02
$dCF/d\ln N_d$	0.08±0.03	0.10±0.03	0.13±0.04	0.11±0.02