Dependence of Convective Cloud Microphysical Properties on Environmental Conditions during the TRACER and ESCAPE Field Campaigns: A Synergistic Approach of Observations, Machine Learning, and Parcel Models

YONGJIE HUANG,[a] GREG M. MCFARQUHAR,[b,c] SAURABH U. PATIL,[b,c] LAN GAO,[c] MATEUSZ TASZAREK,[d] MING XUE,[a,c] ANDREW DZAMBO,[b] MENGISTU WOLDE,[e] LEONID NICHMAN,[e] CUONG NGUYEN,[e] KEYVAN RANJBAR,[e] NATALIA BLIANKINSHTEIN,[e] KENNY BALA,[e] PAVLOS KOLLIAS,[f,g] MICHAEL P. JENSEN,[g] QIXU MO,[h] ROELOF BRUINTJES,[i] CHONGAI KUANG,[g] AND TAMANNA SUBBA[g]

[a] *Center for Analysis and Prediction of Storms, University of Oklahoma, Norman, Oklahoma*
[b] *Cooperative Institute for Severe and High-Impact Weather Research and Operations, University of Oklahoma, Norman, Oklahoma*
[c] *School of Meteorology, University of Oklahoma, Norman, Oklahoma*
[d] *Adam Mickiewicz University, Poznań, Poland*
[e] *National Research Council Canada, Ottawa, Ontario, Canada*
[f] *School of Marine and Atmospheric Sciences, Stony Brook University, Stony Brook, New York*
[g] *Environmental Science and Technologies Department, Brookhaven National Laboratory, Upton, New York*
[h] *SPEC Inc., Boulder, Colorado*
[i] *NSF National Center for Atmospheric Research, Boulder, Colorado*

(Manuscript received 13 December 2024, in final form 5 August 2025, accepted 18 August 2025)

ABSTRACT: The sensitivity of convective clouds to aerosols and their interactions with environment, combined with limited observational constraints in parameterizations, introduces significant uncertainties in atmospheric models. This study investigates the dependence of convective cloud microphysical properties on environmental conditions using a synergistic approach that combines unique observations from the Tracking Aerosol Convection Interactions Experiment (TRACER) and Experiment of Sea Breeze Convection, Aerosols, Precipitation, and Environment (ESCAPE) field campaigns, machine learning techniques, and parcel model simulations with a superdroplet microphysics scheme. A random forest algorithm identifies in situ vertical velocity w, temperature T, and surface fine-mode aerosol mass concentration as the three most important environmental conditions influencing cloud properties including liquid water content (LWC), number concentration for particles with $D_{\max} < 50$ μm ($N_{c,<50}$), 50 μm $\leq D_{\max} \leq$ 3000 μm ($N_{c,50–3000}$), and droplet effective diameter D_e. The results show that LWC, $N_{c,<50}$, and $N_{c,50–3000}$ significantly increase with w in updrafts. Across w bins, as T decreases, LWC, D_e, and $N_{c,50–3000}$ increase, while $N_{c,<50}$ decreases, which are closely linked to the distance above cloud bases. Warmer cloud bases yield higher LWC, greater $N_{c,50–3000}$, and smaller $N_{c,<50}$, while polluted environments produce greater $N_{c,<50}$. Parcel model simulations successfully replicate these observed dependencies. The simulation results indicate that warmer cloud bases enhance condensation generating larger droplets, and differences in droplet sizes are then amplified through collision–coalescence, resulting in a greater $N_{c,50–3000}$. Polluted conditions result in a greater $N_{c,<50}$ primarily due to enhanced cloud condensation nuclei activation despite increased collision–coalescence rates compared to pristine conditions. This study provides *observed quantitative patterns* characterizing cloud microphysical properties as a function of key environmental parameters, offering valuable constraints for improving physics parameterizations and numerical models.

SIGNIFICANCE STATEMENT: This study explores how microscale characteristics of convective clouds change under varying meteorological and aerosol conditions. By analyzing data from two major field campaigns, combined with advanced machine learning and physics-based models, we obtained the quantitative patterns linking cloud microscale characteristics with key environmental factors. We also revealed the mechanisms by which warmer cloud bases produce larger droplets, while polluted environments lead to an increase in smaller droplets. These insights provide critical guidance for improving weather and climate models, helping to reduce forecast uncertainties. This study also highlights the value of integrating observations, machine learning, and numerical modeling to advance our understanding of cloud physics.

KEYWORDS: Convective clouds; Cloud microphysics; Aircraft observations; Idealized models; Aerosol indirect effect; Machine learning

1. Introduction

Convective clouds play important roles in Earth's weather and climate systems, as they transport aerosols, moisture, heat, and momentum through the troposphere, significantly influencing large-scale atmospheric circulations, weather patterns, radiative energy balance, and precipitation processes (e.g., Battan 1953; Robert 1982; Hartmann et al. 1992; Tung and Yanai 2002; Tromeur and Rossow 2010; Bony et al. 2015; Huang et al. 2019; Oertel and Schemm 2021).

The Intergovernmental Panel on Climate Change (IPCC) Sixth Assessment Report (AR6) highlights the challenge of

Corresponding author: Yongjie Huang, yongjie.huang@ou.edu; huangynj@gmail.com

DOI: 10.1175/JAS-D-24-0269.1

accurately representing convective clouds in global climate models (GCMs), noting low confidence in models' ability to capture convective clouds and their interactions with the surrounding environments (Masson-Delmotte et al. 2021). Due to the constraint of computing resources, current-generation GCMs are still run at coarse resolutions, in which small-scale cloud processes cannot be represented explicitly and shallow and deep convection are typically handled by subgrid-scale parameterizations (Masson-Delmotte et al. 2021). However, even current advanced cloud-resolving models using advanced microphysics schemes cannot realistically simulate convective clouds and especially their microphysical properties (e.g., Morrison and Milbrandt 2011; Fan et al. 2017; Xue et al. 2017; Falk et al. 2019; Huang et al. 2020, 2021; Li et al. 2023). Significant uncertainties and limited observational constraints persist in the parameterizations of cloud microphysics and their interactions with surrounding environments, in both numerical weather prediction and Earth system models (Khain et al. 2015; Morrison et al. 2020). Therefore, better understanding how environmental conditions, such as dynamic, thermodynamic, and aerosol conditions, influence cloud microphysical properties, along with improving the parameterization of these key processes using observational constraints, is essential for improving the representation of convective clouds in multiscale models (Guo et al. 2015; Hagos and Houze 2016; Rio et al. 2019; Masson-Delmotte et al. 2021).

Previous field experiments, such as the Cloud, Aerosol, and Monsoon Processes Philippines Experiment (CAMP^{2}Ex) (Reid et al. 2023); the Remote Sensing of Electrification, Lightning, and Mesoscale/Microscale Processes with Adaptive Ground Observations (RELAMPAGO) field campaign (Nesbitt et al. 2021); the Cloud, Aerosol, and Complex Terrain Interactions (CACTI) field campaign (Varble et al. 2021); the Convective Transport of Active Species in the Tropics (CONTRAST) experiment (Pan et al. 2017); the Studies of Emissions, Atmospheric Composition, Clouds and Climate Coupling by Regional Surveys (SEAC^{4}RS) field mission (Toon et al. 2016); the High Altitude Ice Crystals–High Ice Water Content (HAIC–HIWC) international field campaign (Strapp et al. 2016); the Midlatitude Continental Convective Clouds Experiment (MC3E) (Jensen et al. 2016); and the Deep Convective Clouds and Chemistry (DC3) field campaign (Barth et al. 2015), mostly observed convective clouds within mesoscale convective systems, although CAMP^{2}Ex and CACTI also sampled cumulus congestus clouds. However, the inherent complexity of these convective systems has posed challenges in fully understanding the intricate relationships between convective clouds and their surrounding environments. Although the Ice in Clouds Experiment–Tropical (ICE-T) project (Heymsfield and Willis 2014) sampled isolated convective clouds, it was conducted primarily over tropical maritime regions with a limited range of meteorological and aerosol conditions, and it focused mainly on upper-level ice formation processes. In contrast, more recent campaigns such as the Tracking Aerosol Convection Interactions Experiment (TRACER) (Jensen et al. 2023, 2025) and the Experiment of Sea Breeze Convection, Aerosols, Precipitation, and Environment (ESCAPE) (Kollias et al. 2025), both conducted in the vicinity of Houston, Texas, in 2022, used ground-based and airborne probes to observe numerous isolated convective cells under a wide range of meteorological and aerosol conditions. These efforts have provided unique datasets and unprecedented opportunities to investigate how environmental factors influence convective cloud properties in more controlled and isolated conditions.

These comprehensive observational datasets from the TRACER and ESCAPE field campaigns also open new opportunities for applying data-driven approaches to explore complex relationships in cloud systems. Although artificial intelligence (AI) and machine learning (ML) methods are not new and date back to the last century (Haupt et al. 2022), recent advances in computational power and data availability have driven an exponential growth in their applications for Earth sciences over the past few years. These techniques have become essential tools for both researchers and engineers (McGovern et al. 2024). Despite the promise of these technologies, significant challenges remain, particularly regarding their explainability, in which AI/ML methods often take inputs and produce outputs without providing users with physically interpretable insights (McGovern et al. 2019). Additionally, since AI/ML methods primarily identify statistical relationships in data, physical and dynamical models are still indispensable for uncovering the underlying physical mechanisms behind these relationships.

This study leverages the unique observations from the TRACER and ESCAPE field campaigns, combined with advanced machine learning and physical models, to explore the dependence of convective cloud microphysical properties on surrounding environmental conditions including meteorological and aerosol factors. It aims to address three key scientific questions:

- What are the key environmental parameters influencing isolated convective cloud microphysical properties?
- How do isolated convective cloud microphysical properties vary with changes in these key environmental parameters?
- What mechanisms underlie the observed dependence of isolated convective cloud microphysical properties on environmental parameters?

The machine learning method used in this study, a random forest algorithm, trained on data from the TRACER and ESCAPE field campaigns, identifies the key environmental parameters that influence cloud microphysical properties. To explore the underlying physical mechanisms, we employ a parcel model integrated with a superdroplet microphysics scheme. This study aims to provide observed quantitative patterns that characterize convective cloud microphysical properties as a function of key environmental parameters, while also offering a deeper understanding of the physical mechanisms that shape these patterns. It will be valuable for evaluating numerical models and constraining physics parameterization schemes in future studies.

The structure of the paper is as follows: Section 2 provides an overview of the data and methods used in this study. Section 3 presents a detailed analysis of cloud microphysical properties and their relationships with environmental parameters, integrating insights from observations, machine learning models, and parcel model simulations to explore the underlying physical mechanisms. Finally, conclusions are summarized in section 4.

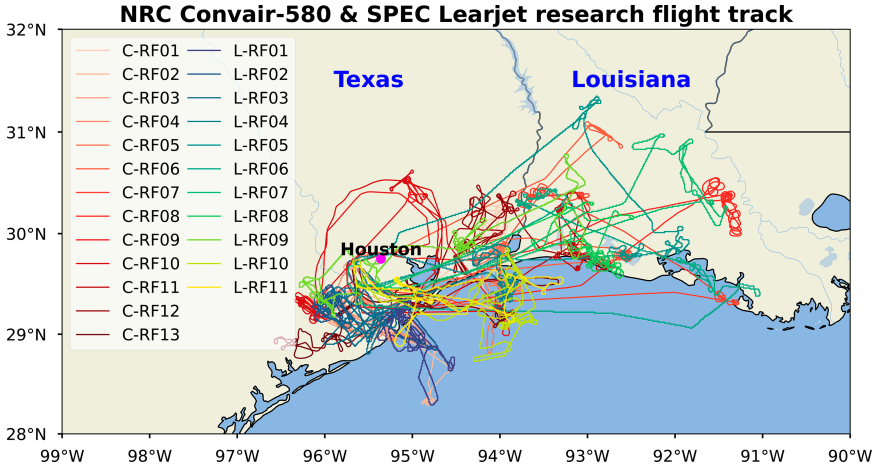

FIG. 1. Research flight tracks of NRC Convair-580 (C-RF01–C-RF13) and SPEC Learjet (L-RF01–L-RF11) during the ESCAPE field campaign.

2. Data and methods

a. Data

This study leverages unique observations collected during the TRACER and ESCAPE field campaigns conducted in the vicinity of Houston, Texas, in 2022, to examine the dependence of convective cloud microphysical properties on surrounding meteorological and aerosol conditions. TRACER was a long-term campaign that employed a comprehensive suite of ground-based instruments to continuously monitor convective cells, aerosols, and meteorological conditions over an extended period (Jensen et al. 2023). ESCAPE provided unique airborne in situ measurements of convective cloud microphysics from two research aircraft, supplemented by ground-based observations similar to those employed in TRACER (Kollias et al. 2025).

1) CLOUD MICROPHYSICAL PROPERTIES

In this study, we utilize in situ observations from cloud microphysics probes aboard the National Research Council of Canada (NRC) Convair-580 and Stratton Park Engineering Company (SPEC) Learjet aircraft to derive cloud microphysical properties from convection sampled during the ESCAPE field campaign. The NRC Convair-580 conducted 13 research flights, and the SPEC Learjet completed 11 flights (Fig. 1). These flights sampled regions encompassing areas west and east of Houston, Texas, the Gulf of America, and the coastal area of Louisiana (Fig. 1). This extensive geographic coverage indicates that the samples represent a diverse range of environmental conditions. The cloud probes installed on the aircraft include the scattering probes, cloud droplet probe, version 2 (CDP-2), and Fast CDP (FCDP) on the NRC Convair-580 and SPEC Learjet, respectively, and the optical array probe (OAP) 2-dimensional stereo (2D-S) imaging probe on both aircraft. Details regarding the parameters of these probes are shown in Table 1. We analyze data from both channels of the 2D-S probe, i.e., the horizontal H and vertical V channels, which are perpendicular to each other. The averaged size distributions and associated derived properties from these channels are utilized in the analysis. We also examined data from the scattering probes FCDP on the NRC Convair-580 and fast forward scattering spectrometer probe (FFSSP) on the SPEC Learjet but found that the liquid water content derived from CDP-2 on the NRC Convair-580 and FCDP on the SPEC Learjet were closer to the bulk measurements from the Nevzorov hotwire cloud water content probes on both aircraft, with ratios closer to 1 (not shown). Consequently, data from the CDP-2 on the NRC Convair-580 and FCDP on the SPEC Learjet are used to characterize small droplets with maximum dimension $D_{max} < 50$ μm. Through the examination of cloud samples collected during the ESCAPE field campaign, we found insufficient ice-phase samples for statistical analysis. Therefore, this study focuses exclusively on liquid-phase clouds, particularly during the initial stage of convection. We only include samples with number concentrations measured by the CDP-2 or FCDP instruments that are $\geq$10 cm^{-3}, as clouds with such high number concentrations have previously

TABLE 1. Cloud microphysics probes and their parameters.

Instrument	Manufacturer	Platform	Range (μm)	Resolution (μm)	Reference
CDP-2[a]	DMT	NRC Convair-580	2–50	2	Lance et al. (2010)
2D-S[b]	SPEC	NRC Convair-580	10–3000	10	Lawson et al. (2006)
FCDP[c]	SPEC	SPEC Learjet	2–50	2	Lawson et al. (2017)
2D-S	SPEC	SPEC Learjet	10–3000	10	Lawson et al. (2006)

[a] Cloud droplet probe, version 2.
[b] Two-dimensional stereo imaging probe.
[c] Fast cloud droplet probe.

been shown to be water clouds (Lance et al. 2010; Hu et al. 2021, 2022). We also manually examined the cloud particle imager (CPI) images for the selected cloud samples at temperatures < 0°C and found an almost complete absence of ice crystal images. Therefore, the influence of ice crystals on the analyses presented in this study can be considered negligible. Finally, a total of 16 853 cloud samples meet the criteria for this study. The particle size distributions derived from the CDP-2 or FCDP for particles with $D_{max} < 50$ μm and from the 2D-S for particles with 50 μm $\leq D_{max} \leq$ 3000 μm are used to estimate the number concentration for $D_{max} < 50$ μm ($N_{c,<50}$) and the number concentration for 50 μm $\leq D_{max} \leq$ 3000 μm ($N_{c,50–3000}$), respectively. The 2D-S data on the NRC Convair-580 and SPEC Learjet were processed as discussed by McFarquhar et al. (2023) and Lawson et al. (2022a), respectively, using different processing software. Statistical comparisons indicate substantial differences in the derived particle size distributions for small size bins between the NRC Convair-580 and SPEC Learjet 2D-S data, while size distributions for bins greater than 50 μm remain generally consistent (not shown). Therefore, only the derived particle size distributions for particles with 50 μm $\leq D_{max} \leq$ 3000 μm from 2D-S probes are used in this study. The liquid water content (LWC) for $D_{max} \leq$ 3000 μm is defined assuming spherical particles by summing the third moment of the size distribution for $D_{max} \leq$ 3000 μm (Table 2). The effective diameter D_e for $D_{max} \leq$ 3000 μm is defined as the ratio of the third moment to the second moment of the particle size distribution (Lacis and Hansen 1974; McFarquhar and Heymsfield 1998; Huang et al. 2021).

2) METEOROLOGICAL ENVIRONMENTAL PARAMETERS

The meteorological environmental conditions surrounding convective clouds, including dynamic and thermodynamic parameters, are described in Table 2. Different instruments and reanalysis products were used to derive these parameters. The air temperature T at the sample location was measured using the Rosemount model 102 probe, with an uncertainty of ±1°C on the NRC Convair-580 and ±0.5°C on the SPEC Learjet (Lawson and Cooper 1990) during the ESCAPE field campaign. Vertical wind speed w was measured using Aventech Aircraft Integrated Meteorological Measurement Systems (AIMMS)-20 Airdata probe, with an uncertainty of ±0.35 m s^{-1} on the NRC Convair-580 and ±0.5 m s^{-1} on the SPEC Learjet (Beswick et al. 2008) during the ESCAPE field campaign. An additional 54 dynamic and thermodynamic parameters, including various types of convective available potential energy (CAPE), convective inhibition (CIN), temperature lapse rates, wind shear, and others, were derived from atmospheric profile data using the thundeR package (Taszarek et al. 2023). These selected dynamic and thermodynamic parameters are well-known primary factors influencing the initiation, development, and decay of convection (Emanuel 1994; Stevens 2005; Giaiotti et al. 2007; Smith 2013). Since the TRACER field campaign provided longer-term continuous sounding observations compared to the ESCAPE campaign, the sounding data from TRACER are used in this study. However, due to the coarse spatiotemporal resolution of the soundings obtained during the TRACER intensive observation period (IOP) (1 June–30 September 2022), typically four times per day, hourly profile data from the fifth generation European Centre for Medium-Range Weather Forecasts (ECMWF) atmospheric reanalysis (ERA5), with a grid spacing of 0.25° (Hersbach et al. 2020), are used to estimate all meteorological environmental parameters in Table 2 except T and w. Taszarek et al. (2021) evaluated convective parameters including those listed in Table 2, derived from reanalysis datasets against rawinsonde data over Europe and North America, highlighting ERA5 as one of the most reliable reanalyses for investigating convective environments.

To assess the quality of ERA5 data specifically for the Houston region, the ERA5 profile data (air temperature, relative humidity, and winds) were compared against the independent observations from a total of 798 balloon-borne soundings at the ARM Mobile Facility 1 (AMF1) main site (29.670°N, 95.059°W) and the ancillary upwind site (29.328°N, 95.741°W) in Houston, Texas, during the TRACER IOP. The comparison presented in Fig. 2 shows that air temperature, relative humidity, eastward wind u, and northward wind v are generally consistent between the ERA5 and radiosonde profiles. The mean biases are about −0.10°C, 1.57%, 0.01 m s^{-1}, and 0.02 m s^{-1} in the respective fields, which are within the uncertainty range of the meteorological sensors used in radiosondes (Holdridge 2020). There is a negative temperature bias of up to −0.45°C within the boundary layer (Fig. 2a), which can influence atmospheric stability. Although our evaluation primarily focuses on vertical profiles of fundamental thermodynamic and kinematic variables, including air temperature, relative humidity, and wind components (u, v), from which the 54 meteorological parameters in this study are derived, we also compared key derived stability parameters such as CAPE and CIN from ERA5 against balloon-borne soundings collected during the TRACER IOP (not shown). The resulting biases align closely with those reported by Taszarek et al. (2021), who conducted a comprehensive evaluation of ERA5-derived sounding parameters, including those listed in Table 2, against radiosonde observations over North America. These findings support the reliability of ERA5 for investigating convective environments. There is a larger discrepancy in mean bias in relative humidity for pressures less than 300 hPa (nearly 10% at 200 hPa, Fig. 2b). However, the parameters derived from the profile data in Table 2 are predominantly influenced by variable values in the planetary boundary layer and free troposphere, and values in the upper troposphere (at pressures below 300 hPa) exert only a limited influence. Therefore, the ERA5 reanalysis dataset is sufficiently accurate for defining meteorological parameters.

The mean values (Mean), standard deviations (Std devs), and coefficients of variation (CV = Std dev/Mean, representing relative dispersion) for environmental parameters are also provided in Table 2. Considerable variability exists among these parameters. For instance, vertical velocity (w, CV = 3.92) and CIN (ML_CIN, CV = −1.02) exhibit notably high relative dispersion, whereas thermodynamic parameters such as lifting condensation level temperature (T_{LCL}, CV = 0.12), relative humidity at 0–2 km (RH_02km, CV = 0.10), and precipitable water (PRCP_WATER, CV = 0.12) show relatively lower variability. These statistical measures effectively summarize

TABLE 2. Cloud microphysical properties and environmental parameters, along with their respective Mean, Std dev, and CV. The variables indicated by asterisks are the final set of input variables used in the random forest models after removing multicollinear predictors.

Variable	Description	Units	Mean	Std dev	CV
$N_{c,<50}$	Cloud droplet number concentration for $D_{max} < 50$ μm	cm^{-3}	232.50	289.26	1.24
$N_{c,50–3000}$	Cloud droplet number concentration for 50 $\mu m \leq D_{max} \leq 3000$ μm	cm^{-3}	1.02	3.69	3.63
LWC	Cloud LWC for $D_{max} \leq 3000$ μm	$g\ m^{-3}$	0.65	1.05	1.62
D_e	Cloud droplet effective diameter for $D_{max} \leq 3000$ μm	μm	43.19	102.90	2.38
w^*	Vertical wind speed at sample locations	$m\ s^{-1}$	0.78	3.04	3.92
T^*	Air temperature at sample locations	°C	10.17	9.06	0.89
M_{aer}*	Surface mass concentration of fine-mode aerosols (sulfate, organic carbon, and black carbon)	$\mu g\ m^{-3}$	3.49	2.22	0.64
T_{LCL}	Air parcel temperature at the LCL, derived from the parcel at 1000 hPa	°C	19.77	2.29	0.12
T_{CCL}*	Air parcel temperature at the CCL, derived from the parcel at 1000 hPa	°C	18.81	2.31	0.12
MU_CAPE	CAPE, derived from the most unstable parcel	$J\ kg^{-1}$	2310.11	900.39	0.39
MU_CAPE_3km*	CAPE between surface and 3 km AGL, derived from the most-unstable parcel	$J\ kg^{-1}$	76.33	57.24	0.75
MUCIN	CIN, derived from the most-unstable parcel	$J\ kg^{-1}$	−18.38	23.20	−1.26
MU_WMAX	Estimated updraft speed (a square root of two times CAPE), derived from the most-unstable parcel	$m\ s^{-1}$	66.73	12.95	0.19
MU_MIXR	Mixing ratio at the height of the most-unstable parcel	$g\ kg^{-1}$	16.59	1.63	0.10
MU_cold_cloud	Depth of the most-unstable CAPE above freezing level	m	7481.33	915.78	0.12
MU_warm_cloud	Depth of the most-unstable CAPE below freezing level	m	5003.79	634.43	0.13
MU_ECAPE*	Entraining CAPE (ECAPE), derived from the most-unstable parcel	$J\ kg^{-1}$	622.16	463.76	0.75
MU_ECAPE_3km*	ECAPE between surface and 3 km AGL, derived from the most-unstable parcel	$J\ kg^{-1}$	18.47	20.01	1.08
MU_EWMAX	Estimated updraft speed (a square root of two times ECAPE), derived from the most-unstable parcel	$m\ s^{-1}$	32.50	13.71	0.42
MUML_CAPE	CAPE, derived from the most unstable 500-m mean layer parcel	$J\ kg^{-1}$	1949.37	777.14	0.40
MUML_CAPE_3km	CAPE between surface and 3 km AGL, derived from the most-unstable 500-m mean layer parcel	$J\ kg^{-1}$	54.14	45.67	0.84
MUML_CIN	CIN, derived from the most-unstable 500-m mean layer parcel	$J\ kg^{-1}$	−27.16	29.52	−1.09
MUML_WMAX	Estimated updraft speed (a square root of two times CAPE), derived from the most-unstable 500-m mean layer parcel	$m\ s^{-1}$	61.21	12.34	0.20
MUML_MIXR	Mixing ratio at the height of the most-unstable 500-m mean layer parcel	$g\ kg^{-1}$	16.09	1.45	0.09
MUML_cold_cloud*	Depth of the most-unstable 500-m mean layer CAPE above freezing level	m	7192.56	968.40	0.13
MUML_warm_cloud*	Depth of the most-unstable 500-m mean layer CAPE below freezing level	m	4752.94	648.77	0.14
MUML_ECAPE	ECAPE, derived from the most-unstable 500-m mean layer parcel	$J\ kg^{-1}$	551.17	411.39	0.75
MUML_ECAPE_3km	ECAPE between surface and 3 km AGL, derived from the most-unstable 500-m mean layer parcel	$J\ kg^{-1}$	13.48	15.65	1.16
MUML_EWMAX	Estimated updraft speed (a square root of two times ECAPE), derived from the most-unstable 500-m mean layer parcel	$m\ s^{-1}$	30.56	12.97	0.42
SB_CAPE	CAPE, derived from the surface-based parcel	$J\ kg^{-1}$	2288.09	871.77	0.38
SB_CAPE_3km	CAPE between surface and 3 km AGL, derived from the surface-based parcel	$J\ kg^{-1}$	74.81	56.00	0.75
SB_CIN	CIN, derived from the surface-based parcel	$J\ kg^{-1}$	−19.27	23.10	−1.20
SB_WMAX	Estimated updraft speed (a square root of two times CAPE), derived from the surface-based parcel	$m\ s^{-1}$	66.44	12.70	0.19
SB_MIXR	Mixing ratio at the height of the surface-based parcel	$g\ kg^{-1}$	16.61	1.66	0.10
SB_cold_cloud	Depth of the surface-based CAPE above freezing level	m	7477.12	916.42	0.12
SB_warm_cloud	Depth of the surface-based CAPE below freezing level	m	4999.16	640.87	0.13
SB_ECAPE	Entraining convective available potential energy, derived from the surface-based parcel	$J\ kg^{-1}$	583.84	419.25	0.72
SB_ECAPE_3km	ECAPE between surface and 3 km AGL, derived from the surface-based parcel	$J\ kg^{-1}$	16.86	18.22	1.08
SB_EWMAX	Estimated updraft speed (a square root of two times ECAPE), derived from the surface-based parcel	$m\ s^{-1}$	31.69	12.79	0.40
ML_CAPE	CAPE, derived from the mixed-layer parcel (defined by averaging potential temperature and water vapor mixing ratio over the 0–500 m AGL layer, and initialized from the surface)	$J\ kg^{-1}$	1912.78	729.20	0.38
ML_CAPE_3km	CAPE between surface and 3 km AGL, derived from the mixed-layer parcel	$J\ kg^{-1}$	51.86	44.00	0.85
ML_CIN*	CIN, derived from the mixed-layer parcel	$J\ kg^{-1}$	−29.29	29.74	−1.02
ML_WMAX	Estimated updraft speed (a square root of two times CAPE), derived from the mixed-layer parcel	$m\ s^{-1}$	60.70	11.88	0.20

TABLE 2. (*Continued*)

Variable	Description	Units	Mean	Std dev	CV
ML_MIXR	Mixing ratio at the height of the mixed-layer parcel	g kg^{-1}	16.10	1.43	0.09
ML_cold_cloud	Depth of the mixed-layer CAPE above freezing level	m	7180.99	963.56	0.13
ML_warm_cloud	Depth of the mixed-layer CAPE below freezing level	m	4732.72	642.55	0.14
ML_ECAPE	ECAPE, derived from the mixed-layer parcel	J kg^{-1}	513.37	368.28	0.72
ML_ECAPE_3km	ECAPE between surface and 3 km AGL, derived from the mixed-layer parcel	J kg^{-1}	12.00	14.05	1.17
ML_EWMAX	Estimated updraft speed (a square root of two times ECAPE), derived from the mixed-layer parcel	m s^{-1}	29.69	12.04	0.41
LR_01km*	Temperature lapse rate between surface and 1 km AGL	K km^{-1}	−8.49	1.60	−0.19
LR_03km	Temperature lapse rate between surface and 3 km AGL	K km^{-1}	−6.52	0.90	−0.14
LR_06km	Temperature lapse rate between surface and 6 km AGL	K km^{-1}	−6.24	0.37	−0.06
RH_01km	Mean relative humidity between surface and 1 km AGL layer	%	0.72	0.09	0.13
RH_02km	Mean relative humidity between surface and 2 km AGL layer	%	0.71	0.07	0.10
RH_14km	Mean relative humidity between 1 and 4 km AGL layer	%	0.62	0.13	0.20
RH_25km	Mean relative humidity between 2 and 5 km AGL layer	%	0.56	0.14	0.26
RH_36km	Mean relative humidity between 3 and 6 km AGL layer	%	0.51	0.15	0.30
BS_01km*	Bulk wind shear between surface and 1 km AGL	m s^{-1}	2.92	1.69	0.58
BS_03km	Bulk wind shear between surface and 3 km AGL	m s^{-1}	6.61	3.75	0.57
BS_06km*	Bulk wind shear between surface and 6 km AGL	m s^{-1}	7.38	3.62	0.49
PRCP_WATER*	Precipitable water (entire column)	mm	45.19	5.54	0.12

the range of environmental conditions associated with the cloud samples analyzed in this study.

3) AEROSOL ENVIRONMENTAL PARAMETER

Due to its superior temporal and spatial continuity relative to available observational aerosol datasets, the hourly aerosol reanalysis from the Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2) (Gelaro et al. 2017), at a grid spacing of 0.5° × 0.625°, is used to characterize the aerosol environment surrounding convective clouds. To evaluate the reliability of MERRA-2 aerosol products, we compared them with available observational datasets.

For evaluating temporal consistency, aerosol number concentrations measured by the ultrahigh sensitivity aerosol spectrometer (UHSAS; 100–1000 nm) during the TRACER IOP were compared with surface aerosol mass concentrations from the MERRA-2 3-hourly aerosol product, which provides mass concentrations in multiple size bins. A 4-month time series comparison (June–September 2022) reveals reasonable consistency between the two variables, with a correlation coefficient of 0.75 (not shown), suggesting that the surface mass concentration from MERRA-2 can serve as a useful proxy for fine-mode aerosol number concentration.

To further evaluate the spatial representativeness of the MERRA-2 aerosol data, the accuracy of the MERRA-2 aerosol optical depth (AOD) at 550 nm was evaluated using AOD data from both five Aerosol Robotic Network (AERONET) program stations in the Houston region (represented by green dots in Fig. 3a) and National Oceanic and Atmospheric Administration (NOAA) *Geostationary Operational Environmental Satellite 17* (*GOES-17*) during the TRACER IOP (Fig. 3a). The AERONET AOD data have an estimated uncertainty of ~0.02, determined as one standard deviation of the difference between the prefield calibration AOD and the interpolated calibration AOD (Giles et al. 2019). The *GOES-17* AOD product includes four quality levels: no retrieval, low-, medium-, and high-quality retrievals. High-quality AOD retrievals are recommended for quantitative applications due to their overall superior performance, although they have a higher uncertainty than the AERONET AOD, typically around 0.1 (https://www.star.nesdis.noaa.gov/atmospheric-composition-training/documents/G17_ABI_AOD_ReadMe.pdf). Therefore, only the high-quality AOD retrieval data from *GOES-17* are utilized and then regridded to match the MERRA-2 grid (represented by blue dots in Fig. 3a) for evaluation. Due to the sensitivity of satellite-based measurements to surface reflectance, high-quality *GOES-17* AOD retrievals are unavailable near the land–sea interface. Consequently, no comparison grid cells with MERRA-2 are shown in these areas (Fig. 3a). The root-mean-square error (RMSE) between MERRA-2 and AERONET AOD is 0.060 with a mean difference of 0.03 and a standard deviation of 0.05 (Fig. 3b). For the comparison between MERRA-2 and *GOES-17* AOD, the RMSE is 0.066 with a mean difference of 0.02 and a standard deviation of 0.06 (Fig. 3c). These differences exceed the uncertainty of AERONET AOD (~0.02) but remain below that of *GOES-17* AOD retrievals (~0.1). Despite some biases, the MERRA-2 AOD is generally consistent with both AERONET observations and *GOES-17* retrievals, aligning closely with the 1:1 lines (Figs. 3b,c), which effectively captures the spatial distribution of aerosols and represents a range of aerosol environments from clean to polluted conditions. In addition, aerosol number concentrations (100–1000 nm) measured by the UHSAS instrument aboard the NRC Convair-580 aircraft during the ESCAPE field campaign exhibit statistically significant positive correlations with collocated total aerosol mass concentrations from MERRA-2, integrated over a comparable size range (not shown). Their correlation coefficients are 0.77 ($p < 0.001$) for 2-min averaged UHSAS data and 0.58 ($p < 0.01$) for 10-min averages.

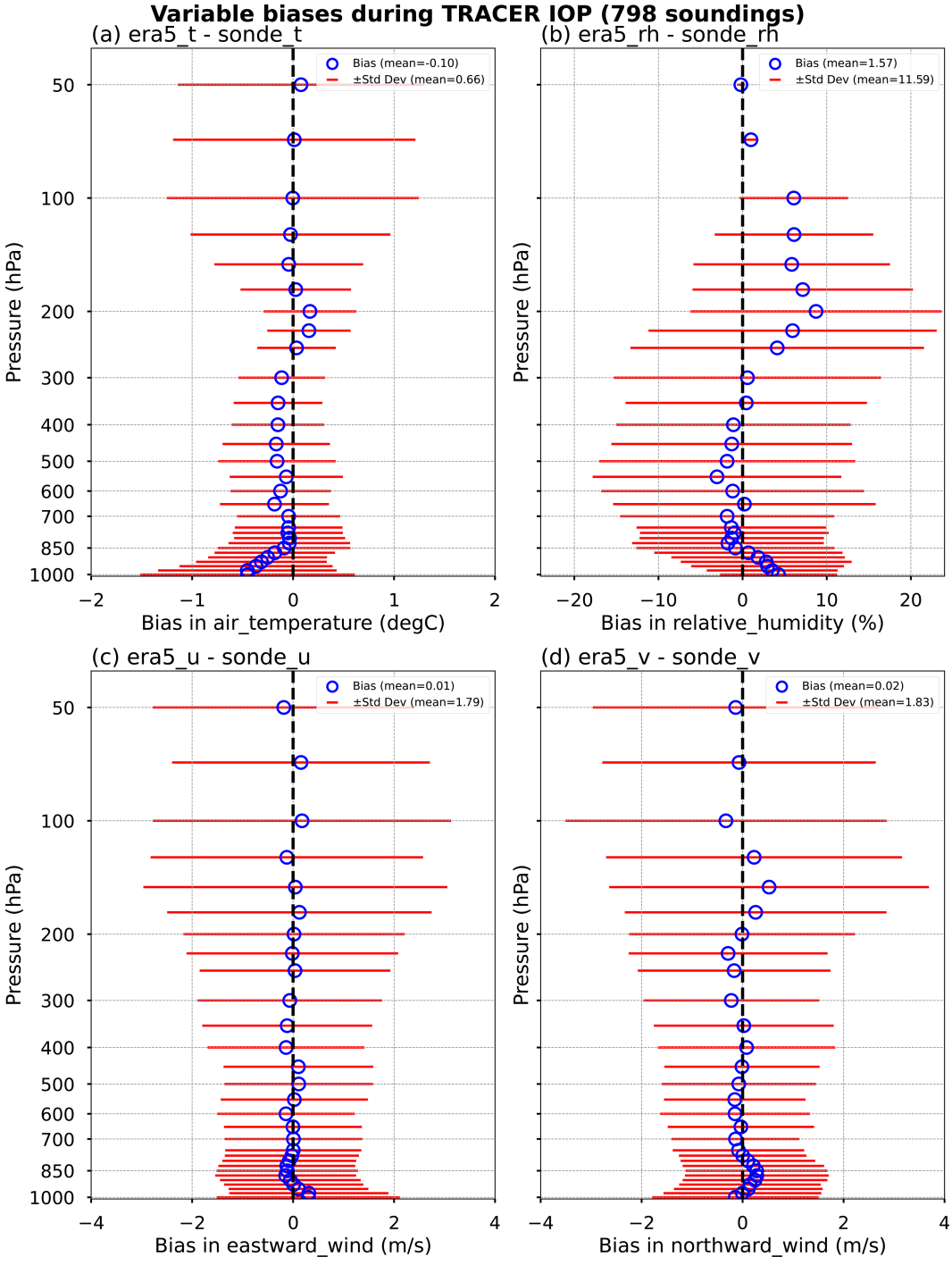

FIG. 2. Differences in (a) air temperature (°C), (b) relative humidity (%), (c) eastward wind (u; m s^{-1}), and (d) northward wind (v; m s^{-1}) between the ERA5 and balloon-borne soundings during the TRACER IOP. Blue circles indicate the mean bias at each level, and red bars represent the ±1 standard deviation.

Based on the comparisons with available observational datasets, the MERRA-2 aerosol products adequately represent both the spatial distribution and temporal evolution of aerosols in the study region. AOD is commonly used as a proxy for the concentration of cloud condensation nuclei (CCN) (Andreae 2009; Koren et al. 2010), but its reliability remains a topic of ongoing debate (Stier 2016; Veals et al. 2022; Varble et al. 2023). Following previous studies (e.g., Hasekamp et al. 2019; Pan et al. 2021; Block et al. 2024), we derive the surface mass concentration M_{aer} of fine-mode aerosols from MERRA-2

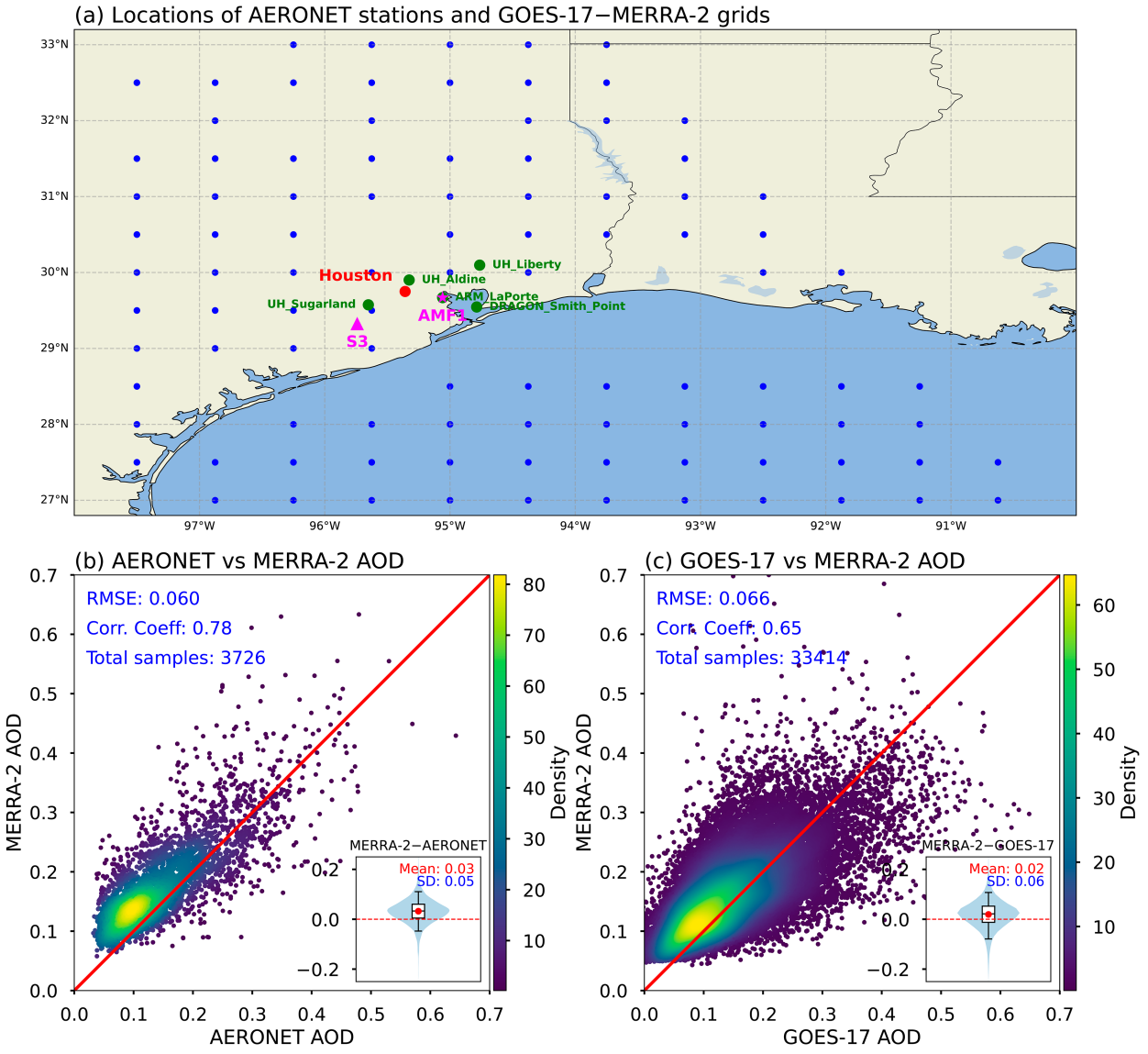

FIG. 3. Comparison of AOD at 550 nm between MERRA-2, AERONET, and *GOES-17* datasets during the TRACER IOP. (a) Locations of AERONET stations (green dots) and grid cells for the comparison between *GOES-17* and MERRA-2 (blue dots). Scatterplot of (b) AERONET AOD (*x* axis) vs MERRA-2 AOD (*y* axis) and (c) *GOES-17* AOD (*x* axis) vs MERRA-2 AOD (*y* axis) with density coloring. In (b) and (c), the color intensity of the scatter points indicates the data point density as estimated by a Gaussian kernel density estimate. The red line represents the 1:1 line. Key statistical metrics including RMSE, correlation coefficient, and total number of samples are annotated on each plot. The inset plot shows the distribution of difference [MERRA-2 minus AERONET in (a) and MERRA-2 minus *GOES-17* in (b)] with a violin plot and box plot. The box plot indicates the 5th, 25th, 50th (median), 75th, and 95th percentiles. The mean difference (red dot and annotation) and standard deviation (blue annotation) are also highlighted.

as an alternative CCN proxy. Specifically, we include sulfate, organic carbon, and black carbon components, while excluding dust aerosol due to its generally hydrophobic nature and limited effectiveness as CCN. Given the uncertainties in the MERRA-2 aerosol products, we intentionally avoid establishing quantitative relationships between M_{aer} and cloud microphysical properties. Instead, we categorize cloud samples into pristine and polluted regimes based on M_{aer}, thereby enhancing the robustness of our analysis and conclusions.

b. *Methods*

1) RANDOM FOREST

A random forest is an ensemble learning algorithm that builds multiple decision trees during training (Ho 1995; Breiman 2001). Each decision tree is trained on a random subset of both the training data and the predictors (or input variables), introducing randomness that reduces overfitting and enhances generalization. The random forest makes

predictions by aggregating the outputs of all individual decision trees, typically using majority voting for classification problems or averaging for regression tasks (Biau and Scornet 2016; Rigatti 2017). The random forest algorithm is particularly effective in handling noisy data, outliers, and large datasets and excels when the data contain complex interactions or nonlinear relationships (Biau and Scornet 2016; Genuer et al. 2017). A key feature of the random forest is its ability to provide predictor importance metrics, which are especially useful for identifying the environmental parameters that most strongly influence cloud properties in this study. The random forest has seen extensive application across atmospheric sciences, such as in atmospheric chemical composition forecasts (Brokamp et al. 2018; Grange et al. 2018), aviation turbulence diagnosis (Williams 2014), urban surface roughness prediction (Duan and Takemi 2021), ocean heat content anomaly mapping (Lyman and Johnson 2023), and weather forecasting (Gagne et al. 2017; Herman and Schumacher 2018; Burke et al. 2020; Hill et al. 2020; Loken et al. 2020; Mecikalski et al. 2021; Radford and Lackmann 2023).

In this study, we employ the random forest algorithm implemented in the open-source Python library, scikit-learn (Pedregosa et al. 2011), available online (https://scikit-learn.org). A random forest regression model is configured with 200 trees (n_estimators = 200) and a fixed random seed (random_state = 42), while all other parameters are kept at their default values. Obviously, covariability exists among the meteorological and aerosol environmental parameters listed in Table 2. To mitigate this redundancy, we applied hierarchical clustering based on Spearman rank-order correlation coefficients among the input variables. A correlation threshold of 0.6 was used to define clusters (other thresholds were tested and yielded consistent results). From each cluster, the variable with the highest variance (representing the greatest informational content) was retained. This reduced set of input variables, indicated by asterisks in Table 2, was then used to train random forest models to predict four cloud microphysical properties (i.e., LWC, D_e, $N_{c,<50}$, and $N_{c,50–3000}$).

2) PARCEL MODEL

The machine learning model primarily identifies statistical relationships between cloud microphysical properties and their corresponding environmental conditions. However, to uncover the physical mechanisms behind these relationships, physics-based models are necessary. In this study, the open-source Python package PySDM (Bartman et al. 2022), available online (https://github.com/open-atmos/PySDM), is used to set up an idealized parcel model to investigate the physical mechanisms driving the observed dependencies of cloud microphysical properties on environmental conditions. PySDM is designed to simulate the dynamics of particle populations, such as aerosol, cloud, and rain particles, immersed in moist air. It employs the Lagrangian particle–based approach known as the superdroplet method (SDM) to simulate the evolution of the particle size spectrum, focusing on processes such as condensational and collisional growth, fluid flow interaction, and chemical composition changes. The SDM Monte Carlo algorithm, introduced by Shima et al. (2009), is implemented to efficiently simulate collisional growth within the Lagrangian framework (Bartman et al. 2022). This Lagrangian particle-based method is advocated as a solution to the conceptual and practical challenges associated with traditional bulk and bin microphysics parameterization schemes (Morrison et al. 2020).

3. Results and discussion

a. Key environmental parameters

All environmental parameters indicated by asterisks in Table 2 are used as predictors to train random forest models to examine their relative importance in predicting cloud microphysical properties (i.e., LWC, D_e, $N_{c,<50}$, and $N_{c,50–3000}$) (Fig. 4). The dataset is randomly split, with 80% used for training and 20% reserved for testing. The importance here is based on the mean decrease in impurity, computed as the average reduction in node impurity (variance) attributed to each feature across all trees and then normalized so that the importance sum to one. In addition, permutation importance is also used to further evaluate key predictors, and the identified key environmental parameters remain consistent despite minor differences in the order of variable importance (not shown). Additionally, 100 iterations of Monte Carlo cross validation are performed to ensure the robustness of the results. The mean correlation coefficients (±standard deviation) from 100 forecasts are 0.72 (±0.05) for LWC, 0.76 (±0.03) for D_e, 0.74 (±0.02) for $N_{c,<50}$, and 0.71 (±0.04) for $N_{c,50–3000}$. The corresponding mean biases (±standard deviation) across the same 100 forecasts are 0.016 (±0.014) g m^{-3} for LWC, 0.24 (±1.12) μm for D_e, 5.23 (±3.34) cm^{-3} for $N_{c,<50}$, and 0.048 (±0.047) cm^{-3} for $N_{c,50–3000}$, respectively. These biases are relatively small compared to the mean values presented in Table 2, indicating that the models adequately fit the data and supporting the reliability of the subsequent variable importance analysis.

It should be noted that the primary goal of using Monte Carlo cross validation is to evaluate the variability in variable importance rankings and to ensure that the key parameters influencing cloud microphysical properties are consistently identified. While potential spatiotemporal correlations exist in the dataset, these are likely reduced due to the ESCAPE campaign's focus on isolated convective cells. To further mitigate such correlations, we also conducted 10-fold cross validation using a time-ordered dataset. Although this approach resulted in decreased forecast skill, the set of important variables remained consistent with those identified through Monte Carlo cross validation (not shown), reinforcing the robustness of the key environmental predictors. The decrease in forecast skill when training on one dataset and testing on a relatively independent one is expected. Forecast models generally perform best when trained on datasets that span a broad range of environmental conditions, particularly when the target variables exhibit strong nonlinear dependencies on the inputs. When the training data do not adequately represent the conditions present in the test set, achieving high forecast skill becomes inherently challenging.

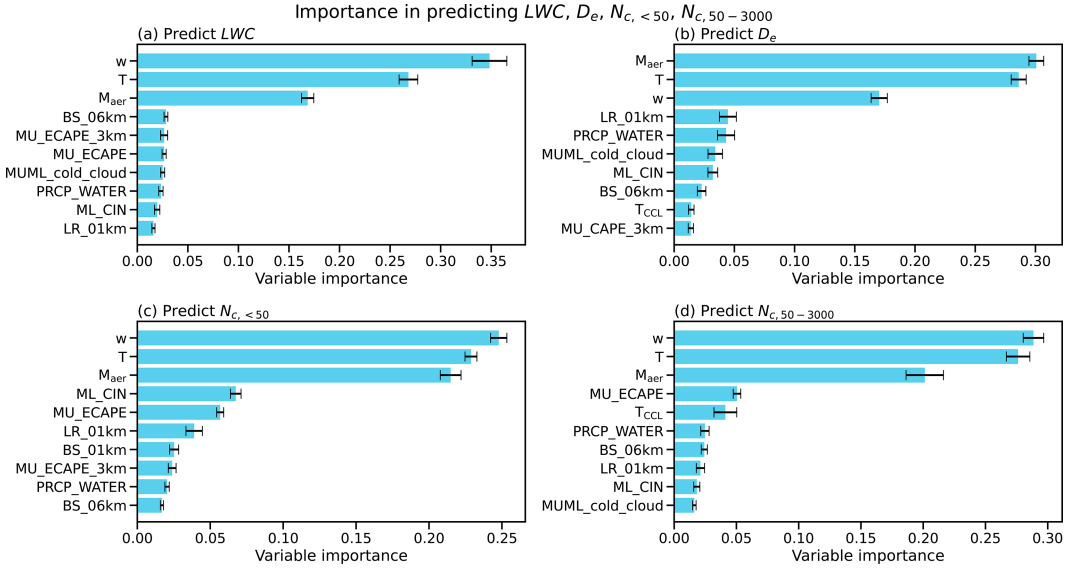

FIG. 4. Importance of top-10 environmental parameters in predicting (a) LWC, (b) D_e, (c) cloud droplet number concentration for $D_{max} < 50$ μm ($N_{c,<50}$), and (d) cloud droplet number concentration for 50 μm $\leq D_{max} \leq$ 3000 μm ($N_{c,50–3000}$) using a random forest model. The horizontal bars represent the mean relative importance of various environmental variables for each cloud property, with error bars showing the standard deviation of the importance scores over 100 iterations of Monte Carlo cross validation.

For predicting all cloud properties, the most important environmental parameters are consistently found to be vertical velocity w, air temperature T, and M_{aer}, with importance scores significantly higher than those of other environmental parameters (Fig. 4). It is worth noting that an additional random forest model was trained to simultaneously predict four cloud properties rather than each property individually. The combined model allows us to assess whether certain environmental parameters, such as w, T, and M_{aer}, are universally important across all cloud properties or only in specific cases. The results indicate that the three most important environmental parameters remained consistent with those identified when predicting cloud properties individually (not shown), suggesting that these parameters are robust predictors that influence multiple cloud properties consistently rather than just isolated cases. The importance of these parameters highlights the critical role of dynamic and thermodynamic conditions, as well as aerosol–cloud interactions, in cloud development. The result is robust, as evidenced by the error bars, which represent the standard deviation in variable importance across 100 Monte Carlo cross-validation runs. In predicting LWC, w plays a particularly influential role (Fig. 4a), underscoring the importance of dynamic conditions in modulating cloud water content. For predicting D_e, M_{aer} and T become more prominent (Fig. 4b), indicating the strong influence of aerosol and thermodynamic conditions on droplet growth. When predicting number concentrations ($N_{c,<50}$ and $N_{c,50–3000}$), the relative importance of the parameters is largely consistent, with w, T, and M_{aer} being the most influential. This highlights their combined impact, especially the dominant role of the dynamic condition w, in modulating cloud droplet concentrations.

Thus, the first scientific question is answered by identifying the three key environmental parameters influencing cloud properties: w, T, and M_{aer}. Since w and T are in situ measurements at the exact location the cloud microphysical properties are derived, they are expected to exhibit stronger correlations with in situ cloud properties. Aerosols M_{aer}, which serve as CCN, are critical for cloud particle formation. These three parameters directly govern the formation and growth of cloud particles and are therefore expected to be more strongly correlated with cloud microphysical properties. However, the ERA5-derived parameters primarily characterize the broader-scale thermodynamic and kinematic environment that influences convective evolution and intensity at a larger scale. As such, their influence on microphysical properties is more indirect, which may explain why they are generally less important for predicting microphysical properties compared to the more direct controls provided by w, T, and M_{aer}.

b. *Cloud properties as a function of environmental parameters*

This section addresses the second scientific question, namely, how cloud microphysical properties vary with changes in the key environmental parameters identified by the random forest models (i.e., w, T, and M_{aer}). Specifically, we examine the variation in cloud microphysical properties within the w–T parameter space. Given the greater uncertainty associated with M_{aer} data compared to the directly measured variables w and T, we primarily use it to separate the aerosol environment into polluted and pristine conditions in this study.

Figure 5 shows how cloud properties vary with changes in both w and T, where the mean cloud properties are illustrated for a two-dimensional grid of varying w and T bins. We focus on the mean values of cloud properties within the w–T parameter space for bins containing a sample percentage greater

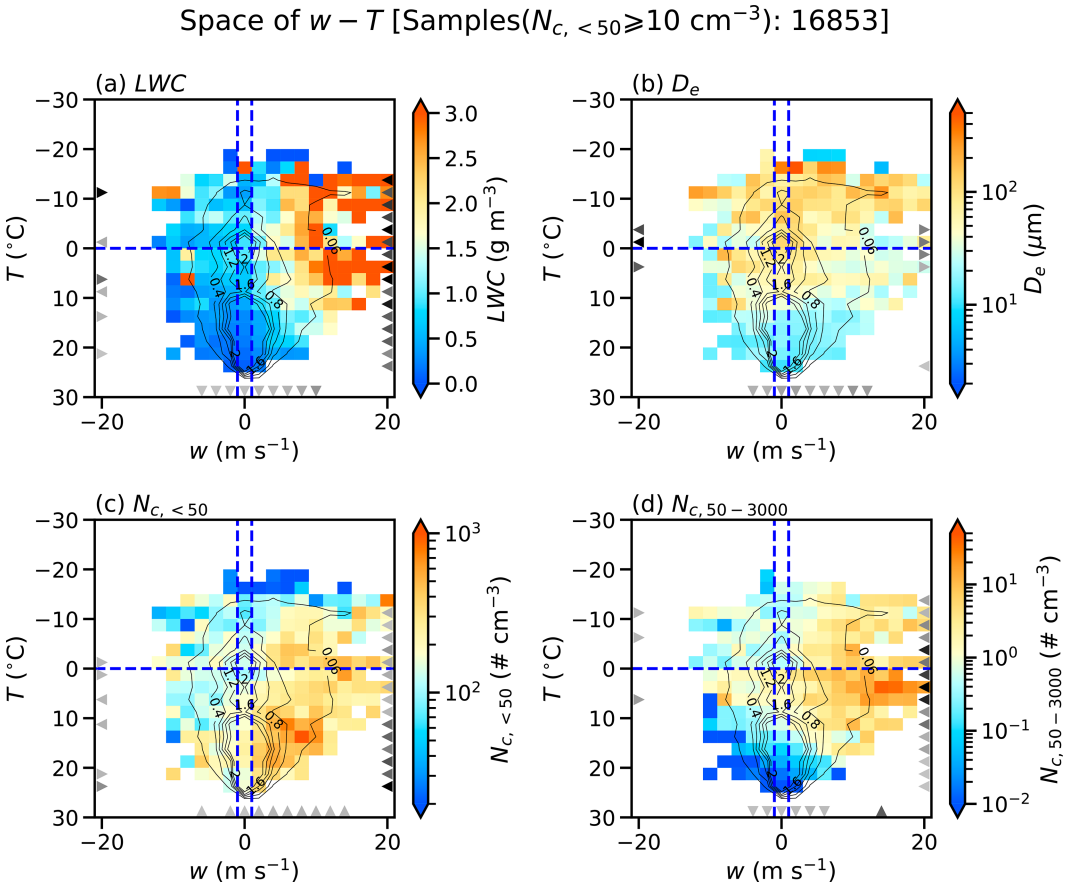

FIG. 5. Variation in cloud properties (a) LWC, (b) D_e, (c) $N_{c,<50}$, and (d) $N_{c,50–3000}$ in the w–T parameter space for 16 853 samples where $N_{c,<50} \geq 10$ cm^{-3}. The bin sizes for computing the variations for w and T are 2 m s^{-1} and 2.5°C, respectively. Colored shading areas represent the mean value of each cloud property within each bin. Contours indicate the sample percentage distribution at levels of 0.06%, 0.4%, 0.8%, 1.2%, 1.6%, and 2%. The triangles along the x axis represent the sign of regression coefficients (slopes) between cloud properties and T in each w bin: upward-pointing triangles (▲) indicate the positive slopes (cloud properties increase as T increases) and downward-pointing triangles (▼) indicate the negative slopes (cloud properties decrease as T increases). The triangles along the left y axis represent the sign of regression coefficients between cloud properties and negative w in each T bin, and the triangles along the right y axis correspond to the regression coefficients between cloud properties and positive w: rightward-pointing triangles (▶) indicate the negative slopes (cloud properties decrease as w increases) and leftward-pointing triangles (◀) indicate the positive slopes (cloud properties increase as w increases). Shading of every triangle indicates the magnitude of the slope (darker = larger |slope|). Only regression coefficients with sample sizes greater than 30 and a statistical significance level of 0.05 are shown. The horizontal blue dashed line indicates 0°C, and the vertical blue dashed lines denote −1 and 1 m s^{-1} in each panel.

than 0.06% (corresponding to at least 10 samples) (Fig. 5). The majority of samples are concentrated within the range of w between −1 and 1 m s^{-1} and T around 16° and −1°C. The number of samples decreases with decreasing temperature at $T < 0$°C, as ice-phase samples are excluded in this study and liquid-phase samples diminish further with lower temperatures below 0°C. The cloud properties exhibit a consistent and statistically significant variation with decreasing temperature. Specifically, LWC, D_e, and $N_{c,50–3000}$ increase, while $N_{c,<50}$ decreases, as T decreases, across different w bins. For LWC, values can be more than twice as high in low $|w|$ bins at $T < 0$°C compared to those at $T \geq 0$°C (Fig. 5a). Similarly, D_e is typically about twice as high and can exceed four times as high in some $|w|$ bins at $T < 0$°C compared to those at $T \geq 0$°C (Fig. 5b). In contrast, $N_{c,<50}$ is approximately twice as high, while $N_{c,50–3000}$ is about half as high, at $T \geq 0$°C compared to those around 0°C (Figs. 5c,d). This suggests that larger cloud particles are found at lower temperatures, while smaller particles are more common at higher temperatures, likely reflecting different growth times of cloud particles as expected from parcel theory. To test this hypothesis, we use T_{LCL} as an estimate for the cloud-base temperature and calculate the temperature difference between sample locations and the cloud base as $T - T_{\mathrm{LCL}}$. New random forest models are then trained, incorporating this additional

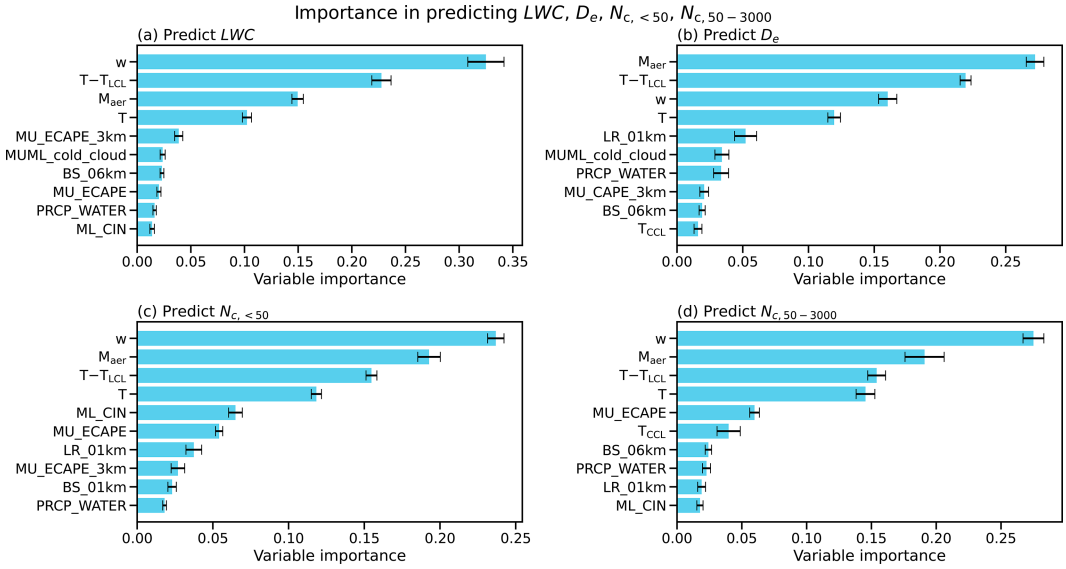

FIG. 6. As in Fig. 4, but for random forest models trained with an additional environmental parameter $T - T_{\mathrm{LCL}}$.

environmental parameter $T - T_{\mathrm{LCL}}$. The updated relative importance scores for the environmental parameters are shown in Fig. 6. Notably, $T - T_{\mathrm{LCL}}$ emerges as one of the three most important environmental factors influencing the four cloud microphysical properties, exceeding the importance of T in predicting each property (Fig. 6). This confirms that the observed variations in cloud microphysical properties with temperature, as shown in Fig. 5, are more closely linked to the distance above the cloud base, reflecting the growth time of cloud particles. For cloud properties as a function of vertical velocity w, LWC, $N_{c,<50}$, and $N_{c,50–3000}$ significantly increase with increasing w in updraft regions (Figs. 5a,c,d). This pattern corresponds to increased cloud particle formation in updraft regions. The variation in D_e with changing w (Fig. 5b) is relatively weaker compared to other cloud properties (Figs. 5a,c,d). This is consistent with the result shown in Fig. 4, which indicates that w is the third most important factor for D_e, while it is the primary factor influencing LWC, $N_{c,<50}$, and $N_{c,50–3000}$.

In general, the magnitude of the slope of cloud properties except LWC with respect to w in updraft regions exhibits greater variability across T bins in updraft regions (indicated by the grayscale-filled triangles along the right y axis in Fig. 5) compared to the variability of slopes with respect to T across different w bins (triangles along the x axis in Fig. 5). This difference is especially pronounced for $N_{c,<50}$ and $N_{c,50–3000}$ (Figs. 5c,d). For $N_{c,<50}$, the 95th–5th percentile ratio of slope magnitude with respect to w in updraft regions is 10.66, whereas it is only 6.20 with respect to T. Furthermore, the slope of $N_{c,<50}$ with respect to w in updraft regions increases significantly at higher temperatures between 10° and 25°C, compared to temperatures lower than 10°C (Fig. 5c). For $N_{c,50–3000}$, the contrast is even more pronounced, where the 95th–5th percentile ratio of slope magnitude with respect to w in updraft regions is 94.30, compared to just 2.51 for T. And $N_{c,50–3000}$ shows steeper gradients with w in updraft regions at lower temperatures near 0°C and smaller gradients at temperatures higher than 10°C (Fig. 5d). In contrast, the slope variability of LWC with respect to w in updraft regions across different T bins is relatively modest (Fig. 5a), indicating a more uniform sensitivity to w across temperature levels. This consistency likely results from the combined influences of both $N_{c,<50}$ and $N_{c,50–3000}$, whose sensitivities to w peak at different temperature levels. Taken together, these patterns suggest that cloud droplet activation (reflected in $N_{c,<50}$) is more sensitive to w at higher temperature (lower altitude) levels. While droplet growth via collision–coalescence (reflected in $N_{c,50–3000}$) responds more strongly to w at lower temperature (higher altitude) levels, where condensational growth has likely plateaued and coalescence processes become dominant. In addition, the signs of the slopes of cloud properties with respect to w in downdraft regions are inconsistent and exhibit more complex behavior. Disentangling this complexity requires a three-dimensional cloud-resolving model that explicitly represents the full range of convective cloud processes, which needs further investigation in future studies.

Furthermore, a significant characteristic is that D_e generally increases, while $N_{c,<50}$ decreases with increasing T_{LCL} (not shown), indicating a greater potential for the formation of larger droplets associated with higher cloud-base temperatures. To further elucidate this relationship and account for uncertainties in the use of T_{LCL} as a proxy for cloud-base temperature, the differences in mean cloud properties between the highest and lowest 30th percentiles of entire T_{LCL} samples across the w–T parameter space are displayed in Fig. 7. The use of other percentiles, such as the 20th and 40th, yields consistent results. The mean difference in T_{LCL} between the highest and lowest 30th percentile groups is 5.36°C, with standard deviations of 0.69° and 1.28°C, respectively. This suggests that the variations in cloud properties observed in these groups can effectively show differences between warmer and colder

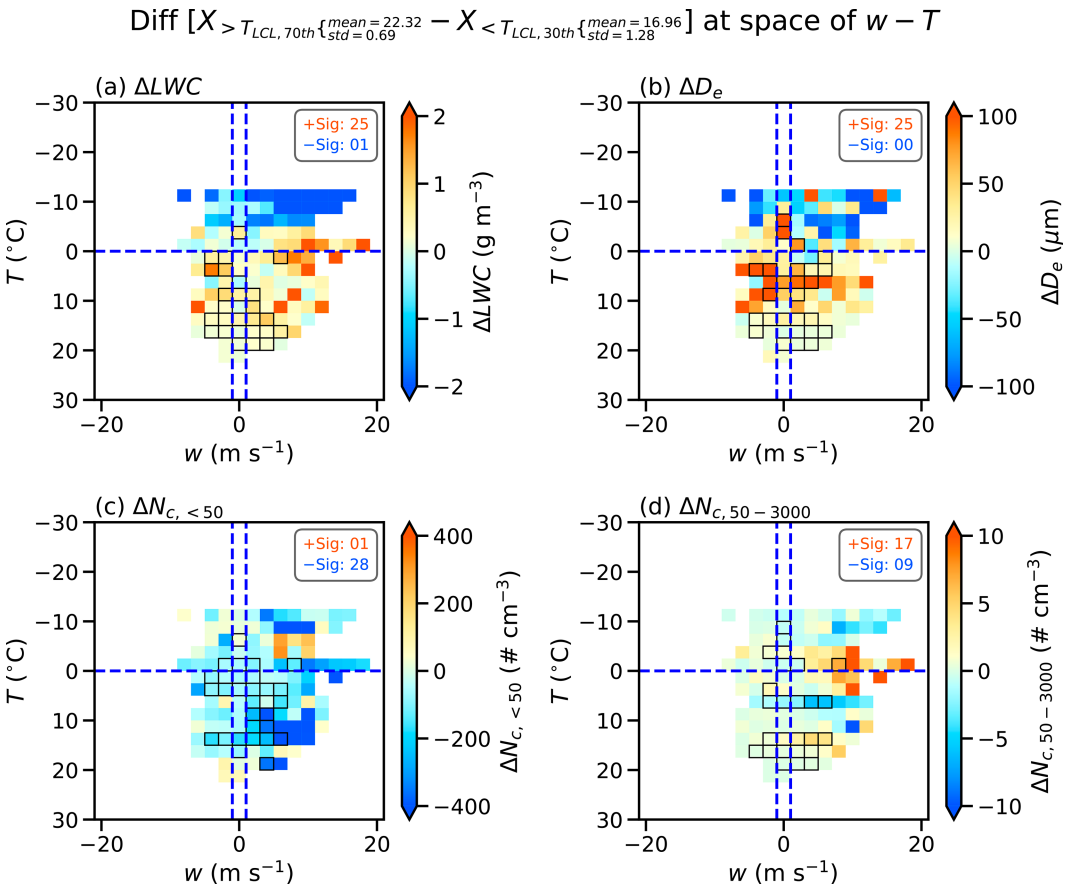

FIG. 7. Differences (shaded pixels) in mean cloud properties between the highest and lowest 30th percentiles of entire T_{LCL} samples across the w–T parameter space for (a) LWC, (b) D_e, (c) $N_{c,<50}$, and (d) $N_{c,50–3000}$. The highest 30th percentile group of T_{LCL} has a mean of 22.32°C with a standard deviation of 0.69°C, and the lowest 30th percentile group has a mean temperature of 16.96°C with a standard deviation of 1.28°C. Pixels outlined with black boxes represent the bins where the sample size is at least 10 in both groups and the mean differences are statistically significant at the 0.05 level based on a t test. The total number of significant positive (orange) and negative (blue) differences is displayed in the top-right corner of each panel. The bin sizes for w and T are 2 m s^{-1} and 2.5°C, respectively. The horizontal blue dashed line indicates 0°C, and the vertical blue dashed lines denote −1 and 1 m s^{-1} in each panel.

cloud bases. The majority of statistically significant bins in the w–T parameter space show predominantly positive differences for LWC (96.2%, Fig. 7a), D_e (100%, Fig. 7b), and $N_{c,50–3000}$ (65.4%, Fig. 7d), while predominantly negative differences are observed for $N_{c,<50}$ (96.6%, Fig. 7c). This indicates that higher water contents, more large droplets, and fewer small droplets are associated with warmer cloud-base conditions compared to colder ones. Moreover, these significant differences span a wide range of w and T bins, reinforcing that the variations in cloud properties between warmer and colder cloud bases are robust. Overall, the results suggest that warmer cloud bases are conducive to droplet growth.

To explore the dependence of cloud properties on aerosol conditions, we use various percentiles (i.e., 20th, 30th, and 40th) of entire M_{aer} samples to divide the observed cloud samples into polluted and pristine groups to analyze their differences in cloud properties. Similarly, the differences in mean cloud properties between the polluted and pristine conditions in the w–T parameter space are examined. The predominant statistically significant positive and negative differences in LWC and D_e depend on the choice of M_{aer} percentiles used to separate cloud samples into polluted and pristine categories (not shown). However, the results for droplet number concentrations $N_{c,<50}$ and $N_{c,50–3000}$ remain consistent across different percentiles. Given the uncertainty in the MERRA-2 M_{aer} product, we focus on the differences in mean cloud properties between the highest and lowest 20th percentiles of entire M_{aer} samples in the w–T parameter space (Fig. 8). The mean M_{aer} difference between the highest and lowest 20th percentile groups is 4.85 μg m^{-3}, with standard deviations of 3.10 and 0.23 μg m^{-3}, respectively. These values provide a high confidence in the separation of polluted and pristine conditions, supporting the robustness of the results discussed here. For increases in M_{aer}, the most statistically significant bins in the w–T parameter space show predominantly positive differences for $N_{c,<50}$ (100%, Fig. 8c) and predominantly

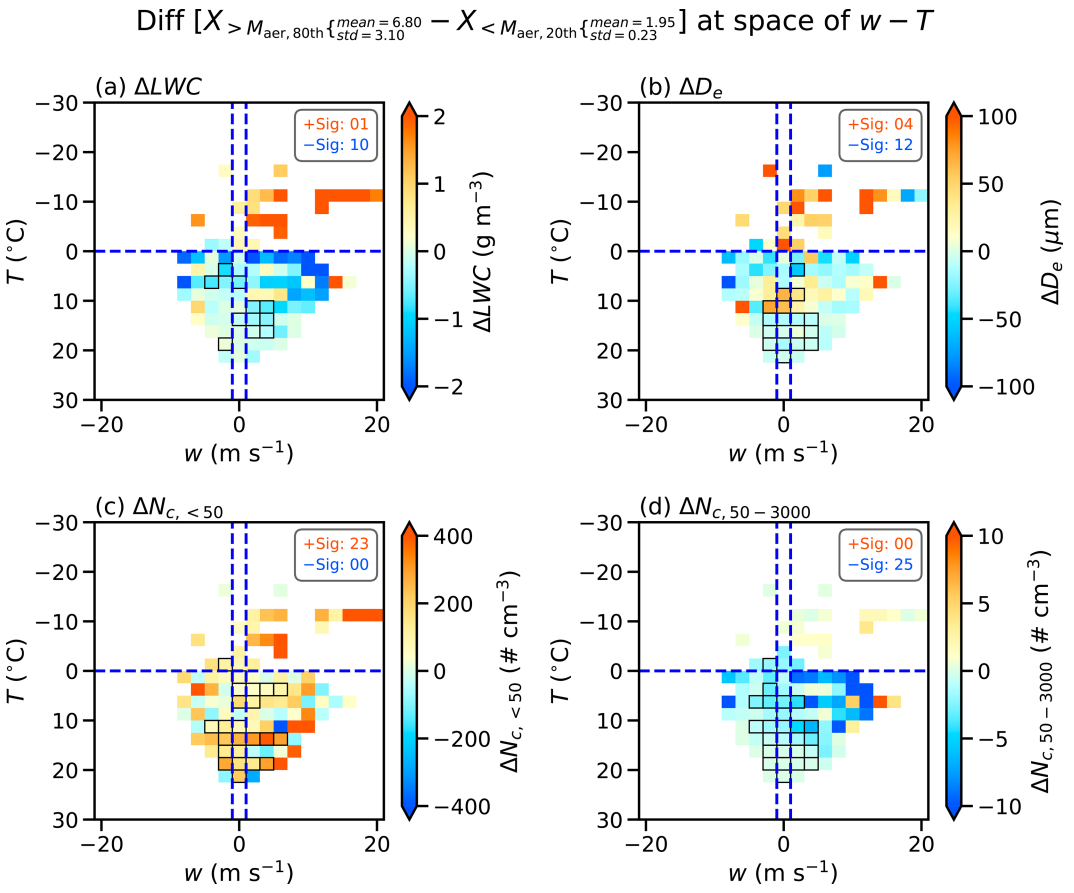

FIG. 8. Differences (shaded pixels) in mean cloud properties between the groups of the highest and lowest 20th percentile of entire M_{aer} samples in the w–T parameter space for (a) LWC, (b) D_e, (c) $N_{c,<50}$, and (d) $N_{c,50–3000}$. The highest 20th percentile group of M_{aer} has a mean of 6.80 μg m^{-3} with a standard deviation of 3.10 μg m^{-3}, and the lowest 20th percentile group has a mean of 1.95 μg m^{-3} with a standard deviation of 0.23 μg m^{-3}. Pixels outlined with black boxes represent the bins where the sample size is at least 10 in both groups, and the mean differences are statistically significant at the 0.05 level based on a t test. The total number of significant positive (orange) and negative (blue) differences is displayed in the top-right corner of each panel. The bin sizes for w and T are 2 m s^{-1} and 2.5°C, respectively. The horizontal blue dashed line indicates 0°C, and the vertical blue dashed lines denote −1 and 1 m s^{-1} in each panel.

negative differences for $N_{c,50–3000}$ (100%, Fig. 8d). This suggests that more small droplets form in polluted environments, while more large droplets are present under pristine conditions. In polluted environments, a higher aerosol concentration increases the availability of CCN, leading to a greater number of cloud droplets given the same water vapor and other conditions. However, this also reduces droplet size due to increased competition for available moisture during droplet growth if the LWC is constant (Twomey 1974, 1977). But, as previously mentioned, the predominance of statistically significant positive and negative differences in LWC and D_e between the polluted and pristine conditions varies when using different M_{aer} percentiles (from 20th to 40th, not shown). This inconsistency arises because LWC and D_e are influenced by different powers of droplet diameter: LWC is sensitive to the cubic power of diameter, while D_e is influenced by both the square and cubic powers. As a result, changes in the concentration of small and large droplets (e.g., a shift toward smaller droplets) can affect LWC and D_e in a complex, nonlinear way, where the net effect depends on the specific balance between the increase in small droplets and the decrease in large ones. Despite these inconsistent changes in LWC and D_e, more negative differences in D_e between high and low M_{aer} conditions are observed at lower levels, while more positive differences appear at higher levels, although the number of statistically significant samples is limited (Figs. 8a,b). This pattern suggests that droplet size differences between polluted and pristine conditions are primarily driven by enhanced CCN activation in polluted environments at lower levels, leading to more numerous but smaller droplets due to limited condensational growth. At higher levels, collision–coalescence processes become increasingly influential, resulting in the formation of larger droplets.

Overall, the observed variations in cloud microphysical properties as a function of key environmental parameters

(i.e., w, T, and M_{aer}) are clearly revealed in the w–T parameter space. Generally, cloud properties LWC, $N_{c,<50}$, and $N_{c,50–3000}$ increase with w in updraft regions. Across different w bins, LWC, D_e, and $N_{c,50–3000}$ tend to increase as T decreases, while $N_{c,<50}$ exhibits a decrease, which are closely linked to the distance above cloud bases. Higher water content, more large droplets, and fewer small droplets are associated with warmer cloud-base conditions compared to colder ones. Additionally, more small droplets form in polluted environments, while more large droplets are observed under pristine conditions. Thus, the results highlight the complex interactions between vertical velocity, temperature, and aerosols, which together play a crucial role in shaping cloud microphysics.

c. *Mechanisms driving observed dependencies*

The previous section presents the statistical relationships between cloud microphysical properties and their corresponding environmental parameters. In this section, the parcel model incorporating the superdroplet microphysics scheme (Shima et al. 2009; Bartman et al. 2022) described in section 2b is employed to investigate the physical mechanisms driving the observed dependence of cloud microphysical properties on environmental parameters.

A total of 442 parcel model simulations were conducted by systematically varying the initial conditions of cloud-base temperature, vertical velocity, and aerosol size distribution, based either directly on observations or within observed ranges (Fig. 5). The parcel initial temperatures, representing cloud-base temperatures T_{CB}, are set to 24°, 23°, 22°, 21°, 20°, 19°, 18°, 17°, 16°, 15°, 14°, 13°, and 12°C, respectively. The parcel, initialized at a pressure of 900 hPa and a relative humidity of 97%, rises adiabatically to 3-km height with a prescribed vertical velocity of 0.25, 0.5, 1, 1.5, 2, 2.5, 3, 4, 5, 6, 7, 8, 9, 10, 12, 15, and 20 m s^{-1}. A sensitivity test was conducted using varying initial conditions of pressure and relative humidity, revealing some impacts on specific values. However, the primary patterns and overall conclusions are consistent (not shown). Two aerosol regimes, representing pristine and polluted conditions, are considered. The corresponding aerosol size distributions are derived from scanning mobility particle sizer (SMPS) observations at the AMF1 site on 17 June 2022, during the TRACER field campaign (Fan et al. 2024; Kuang et al. 2024). The pristine aerosol size distribution is represented by three lognormal modes with total number concentrations of 359, 624, and 245 cm^{-3}; median radii of 9, 27, and 89.5 nm; and geometric standard deviations of 1.6, 1.6, and 1.4, respectively. The polluted aerosol size distribution consists of three lognormal modes with total concentrations of 10 331, 825, and 425 cm^{-3}; median radii of 10.5, 30, and 79.5 nm; and geometric standard deviations of 1.5, 1.4, and 1.4, respectively. The parcel model employs 200 superdroplets with a time step of 1 s, simulating only condensation and collision–coalescence processes, which are the dominant microphysical processes in the early stage convective clouds sampled during the ESCAPE campaign.

To enable a consistent comparison with observations, we also examine parcel model simulation samples with $N_{c,<50} \geq 10$ cm^{-3} at various heights above the cloud base in the $w - T$ parameter space. Evaluation shows that the parcel model generally reproduces the observed patterns of cloud properties in the w–T parameter space. Specifically, LWC and $N_{c,<50}$ increase with increasing w, and LWC, D_e, and $N_{c,50–3000}$ increase while $N_{c,<50}$ decreases as T decreases across different w bins (not shown). Therefore, the parcel model simulations are further used to investigate the mechanisms underlying the differences across various cloud-base temperatures and aerosol conditions (Figs. 9 and 10).

To investigate the differences between warmer and colder cloud-base conditions, simulation samples with 21°C $\leq T_{CB} \leq$ 23°C and 15°C $\leq T_{CB} \leq$ 17°C are selected based on the mean T_{LCL} of the two groups shown in Fig. 7. The predominance of statistically significant negative differences in $N_{c,<50}$ (100%, Fig. 9a) and positive differences in LWC (100%, not shown), D_e (100%, not shown), and $N_{c,50–3000}$ (96.8%, Fig. 9b) resembles the observational analysis (Fig. 7). Under similar relative humidity and aerosol conditions, air parcels originating from warmer cloud bases contain more water vapor, leading to greater condensation as air parcels ascend and cool to saturation at the same temperature levels, thereby producing higher LWC. Changes in D_e, however, reflect the net outcome of competing changes in small and large droplet populations. Consequently, our analysis below will focus on differences in $N_{c,<50}$ and $N_{c,50–3000}$, which more directly reveal shifts in droplet number concentrations. However, these differences within the same T and w bin are actually between droplets with different growth times. Within the same T and w bin, droplets originating from a warmer cloud base experience a longer growth time compared to those from a colder cloud base, which likely results in larger droplets in warmer cloud-base conditions, as expected from an adiabatic growth model. Consequently, this amplifies the differences in cloud properties between warmer and colder cloud-base conditions in the w–T parameter space. However, accurately determining observed cloud-base temperatures is challenging, whereas simulations easily provide them as prescribed initial conditions. Thus, the differences in $N_{c,<50}$ and $N_{c,50–3000}$ are examined in the w–($T - T_{CB}$) parameter space (Figs. 9c,d). Here, $T - T_{CB}$ represents the temperature difference between the cloud sample locations and their corresponding cloud-base temperatures. Within the same $T - T_{CB}$ and w bin, droplets have, to some extent, similar growth times since they are located at locations with the same temperature difference from their cloud bases. The differences in $N_{c,<50}$ and $N_{c,50–3000}$ between warmer and colder cloud bases decrease in the w–($T - T_{CB}$) parameter space (Figs. 9c,d) compared to those in the w–T parameter space (Figs. 9a,b). This suggests that the influence of cloud-base temperatures on $N_{c,<50}$ and $N_{c,50–3000}$ may appear overstated in the w–T parameter space compared to the w–($T - T_{CB}$) parameter space, as the former does not account for the vertical distance from cloud base, represented by $T - T_{CB}$, which reflects the history of droplet growth above cloud base according to parcel theory. However, statistically significant negative and positive differences continue to dominate $N_{c,<50}$ (100% of bins) and $N_{c,50–3000}$ (79.2% of bins) differences for warmer and colder cloud bases, respectively (Figs. 9c,d), indicating the robustness of these differences.

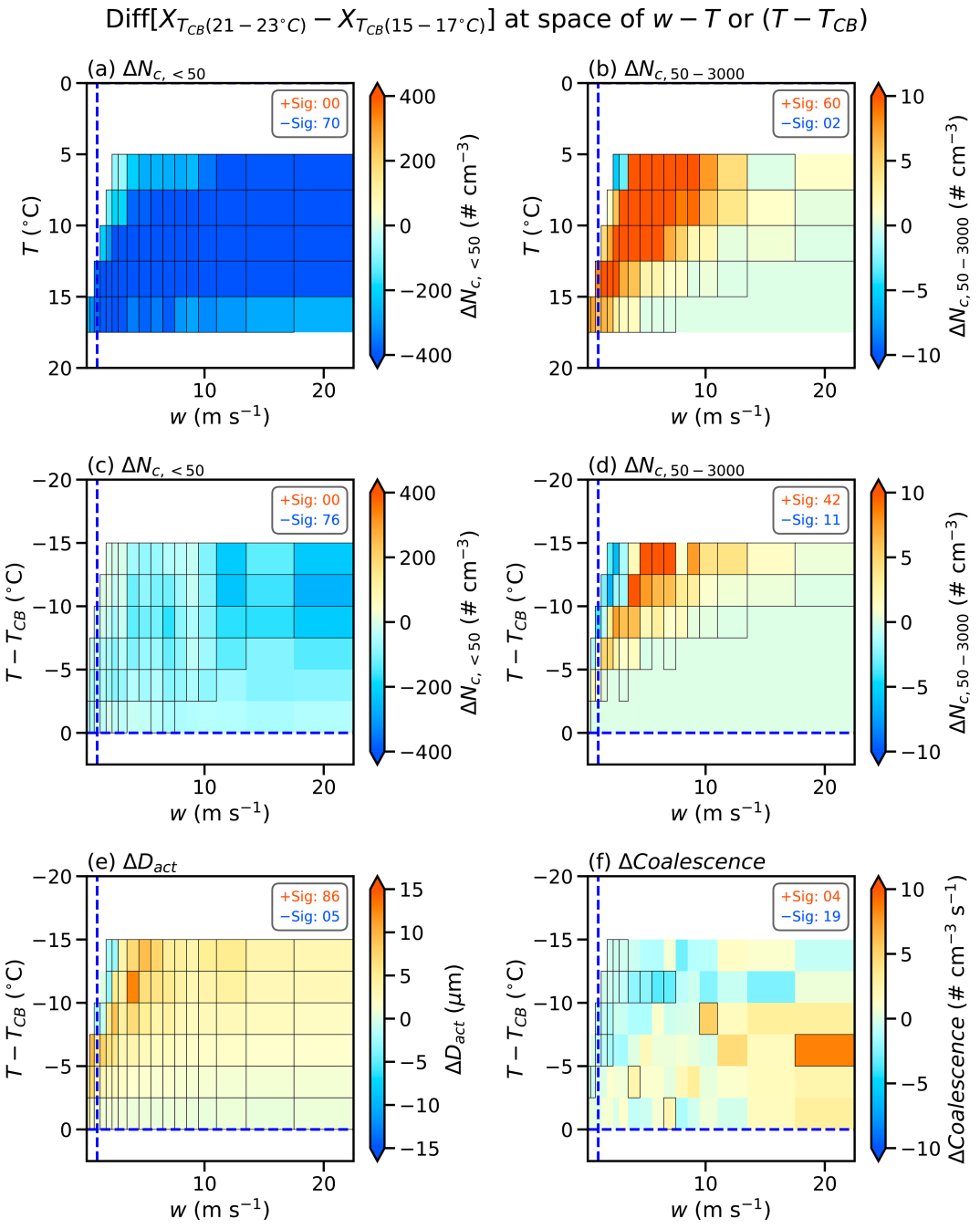

FIG. 9. Differences (shaded pixels) in mean cloud properties or process rates between warmer (21°C ≤ T_{CB} ≤ 23°C) and colder (15°C ≤ T_{CB} ≤ 17°C) cloud-base sample groups in the (a),(b) w–T and (c)–(f) w–($T - T_{CB}$) parameter spaces for (a),(c) $N_{c,<50}$, (b),(d) $N_{c,50–3000}$, (e) diameter of activated droplets D_{act}, and (f) droplet collision–coalescence rate. Pixels outlined with black boxes represent the bins where the sample size is at least 10 in both groups and the mean differences are statistically significant at the 0.05 level based on a t test. The total number of significant positive (orange) and negative (blue) differences is displayed in the top-right corner of each panel. The bin size for T or $T - T_{CB}$ is 2.5°C. The bin center for w corresponds to the prescribed vertical velocities in the parcel model. The horizontal blue dashed line indicates 0°C, and the vertical blue dashed line denotes 1 m s^{-1} in each panel.

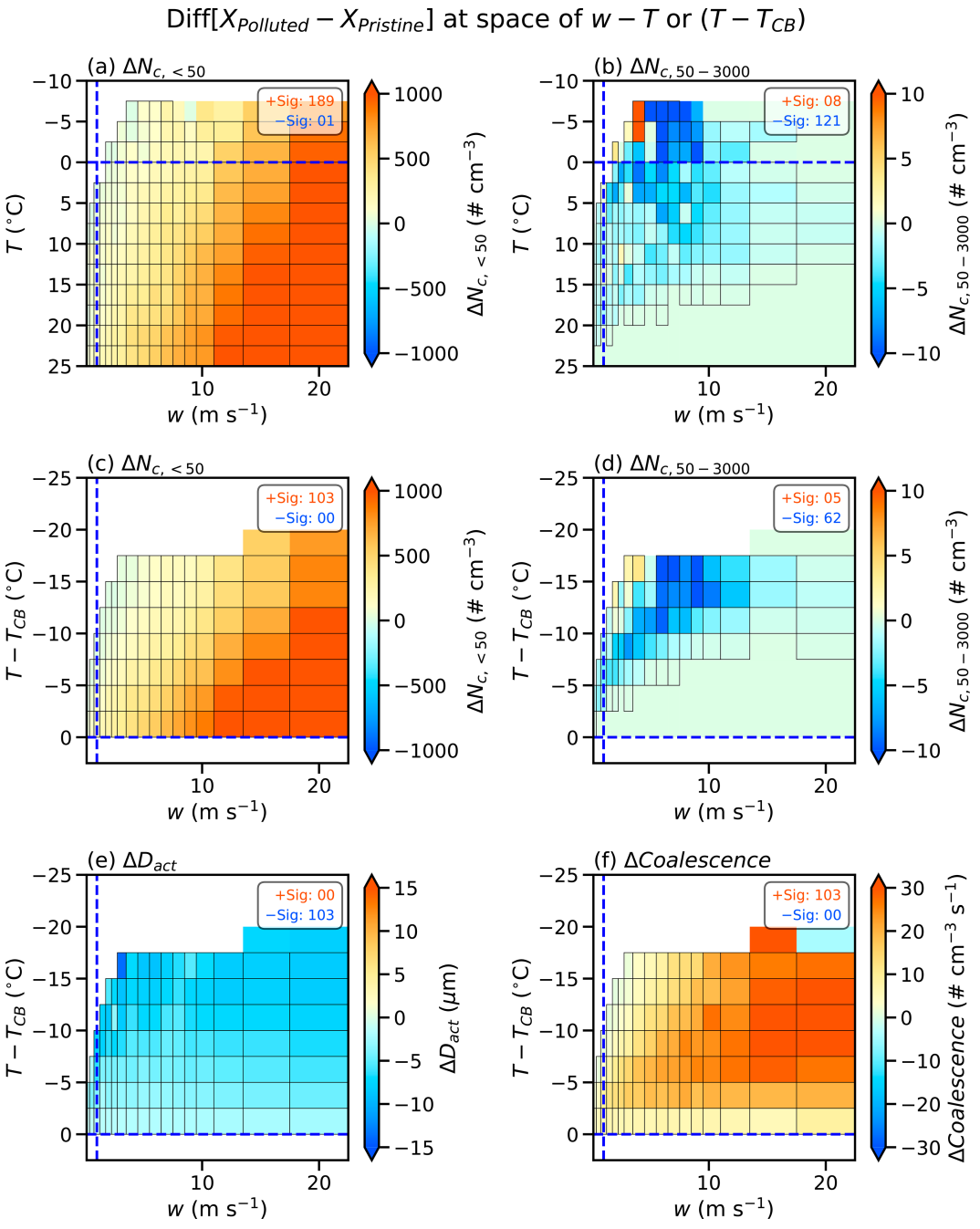

FIG. 10. As in Fig. 9, but for the differences between the polluted and pristine sample groups.

Examination of differences in the mean diameter of activated droplets D_{act} and collision–coalescence rates between warmer and colder cloud bases in the w–$(T - T_{CB})$ parameter space (Figs. 9e,f) reveals that collision–coalescence rates in four paired bins show statistically significant increases, primarily at lower levels in warmer cloud-base conditions, contributing to the increase in droplet sizes. However, the statistically significant positive difference in D_{act} plays a more dominant role (94.5% of bins, Fig. 9e). Another set of parcel model simulations was conducted by considering only the condensation process and turning off the collision–coalescence process. The results still consistently show higher $N_{c,50–3000}$ and lower $N_{c,<50}$ in warmer cloud-base conditions compared to colder ones, although the differences in $N_{c,<50}$ and $N_{c,50–3000}$ are reduced (not shown). This suggests that the greater amount of water vapor available for condensation in warmer cloud-base conditions results in higher condensation growth rates as air parcels rise and cool. The rapid condensation depletes water

vapor quickly, limiting the maximum supersaturation S_{max} that can be achieved (not shown). As a result, fewer CCN particles are activated, but those activated droplets grow larger due to reduced competition for water vapor. This is further supported by observational evidence showing a predominance of statistically significant positive differences in the effective diameter of cloud droplets with $D_{max} < 50$ μm (not shown) and negative differences in $N_{c,<50}$ (Fig. 7c) in warmer cloud-base conditions compared to colder ones, primarily associated with condensation growth. Subsequently, similar or slightly enhanced collision–coalescence rates further amplify the size difference between particles in warmer and colder cloud-base conditions. Ultimately, statistically significant negative $N_{c,<50}$ and positive differences $N_{c,50–3000}$ dominate (Figs. 9c,d), indicating a higher number of large droplets in warmer cloud-base conditions. These results highlight the crucial role of condensation during the initial stages of droplet growth, contrasting with previous studies that have predominantly emphasized the collision–coalescence process (e.g., Lawson et al. 2022b).

Similarly, to investigate differences between polluted and pristine conditions, differences in $N_{c,<50}$, $N_{c,50–3000}$, D_{act}, and the droplet collision–coalescence rate are examined in the w–T and w–$(T - T_{CB})$ parameter spaces as shown in Fig. 10. As discussed in the observational analysis of LWC and D_e (Figs. 8a,b), responses to aerosol perturbations are complex and nonlinear. Consequently, the predominance of statistically significant positive or negative differences between polluted and pristine conditions in both LWC and D_e varies depending on the chosen M_{aer} percentile. In the parcel model simulations, only 23.4% of bins in the w–T parameter space exhibit statistically significant differences in LWC between polluted and pristine conditions. This limited sensitivity may be a result of the use of an adiabatic parcel model, which neglects mixing and other nonadiabatic processes. For simulated D_e, statistically significant negative differences dominate (88.5%, not shown), consistent with the observed pattern (75%, Fig. 8b). These differences arise from changes in the droplet size distribution, which influence D_e in a complex, nonlinear way. The net effect depends on the relative increase in small droplets and decrease in larger ones. To better understand these underlying shifts, we focus below on differences in $N_{c,<50}$ and $N_{c,50–3000}$, which more directly reflect changes in droplet number concentrations. The predominance of statistically significant positive differences in modeled $N_{c,<50}$ (99.5%, Fig. 10a) and negative differences in modeled $N_{c,50–3000}$ (93.8%, Fig. 10b) is consistent with the observed patterns in the w–T parameter spaces (Figs. 8c,d). These differences remain robust even after removing the influence of droplets originating from varying cloud-base temperatures in the w–$(T - T_{CB})$ parameter space (100% for $N_{c,<50}$ and 92.5% for $N_{c,50–3000}$, Figs. 10c,d). In polluted environments, a higher concentration of aerosols provides more CCN for activation as air parcels rise and cool. This results in a higher concentration of small cloud droplets, as evidenced by the statistically significant positive differences in $N_{c,<50}$ (100%, Fig. 10c) and negative differences in D_{act} (100%, Fig. 10e). Although smaller droplets individually exhibit lower collision efficiencies due to their reduced sizes and similar terminal velocities, the overall collision–coalescence rate can be maintained or even enhanced due to the increased droplet number concentration. This is indicated by the statistically significant positive differences in collision–coalescence rates as a function of aerosol amount (100%, Fig. 10f). However, despite the increase in collision–coalescence rates, this process plays a secondary role in determining the difference in mean droplet sizes between polluted and pristine conditions with the statistically significant negative differences in D_e (not shown). The dominant factor influencing droplet sizes is the initial high concentration of small droplets resulting from enhanced CCN activation in polluted conditions. This is further supported by the predominance of statistically significant positive differences in $N_{c,<50}$ and negative differences in $N_{c,50–3000}$ (Figs. 10a–d), indicating a higher number of small droplets and a lower number of large droplets in polluted environments. This is also supported by observational evidence showing a predominance of statistically significant negative differences in the effective diameter of cloud droplets with $D_{max} < 50$ μm (not shown) and positive differences in $N_{c,<50}$ (Fig. 8c) in polluted environments compared to pristine ones, primarily associated with the condensation process. Therefore, the presence of pollution shifts the cloud droplet spectrum toward smaller sizes, with collision–coalescence processes playing a lesser role compared to the primary effect of increased CCN activation.

It should be noted that we primarily use the parcel model to examine the dominant microphysical processes, namely, condensation (including evaporation) and collision–coalescence in early stage convective clouds sampled during the ESCAPE campaign, to help explain some observed patterns, particularly those associated with warmer cloud-base conditions. However, the adiabatic parcel model does not capture the full complexity of convective cloud processes, as it omits several important mechanisms such as entrainment-mixing, rain sedimentation, and aerosol–cloud interactions. Accurately representing these processes would require a three-dimensional cloud-resolving model capable of explicitly simulating these convective clouds, which is beyond the scope of the present study.

4. Conclusions

To investigate the dependence of convective cloud microphysical properties on meteorological and aerosol conditions, a synergistic approach combining unique observations from the TRACER and ESCAPE field campaigns, machine learning techniques, and parcel model simulations is employed. The main findings are summarized as follows:

1) A random forest model identifies in situ vertical velocity, temperature, and surface fine-mode aerosol mass concentration as the three most important parameters affecting the observed cloud microphysical properties.
2) Cloud properties LWC, $N_{c,<50}$, and $N_{c,50–3000}$ significantly increase with increasing w in updraft regions. Across different w bins, LWC, D_e, and $N_{c,50–3000}$ tend to increase as T decreases, while $N_{c,<50}$ exhibits a decrease, which are closely linked to the distance above cloud bases, reflecting different growth times of cloud particles. Higher values of LWC, D_e, and $N_{c,50–3000}$, alongside lower values of $N_{c,<50}$,

are associated with warmer cloud-base conditions compared to colder ones, with the proportions of statistically significantly different T–w bins at 96.2%, 100%, 65.4%, and 96.6%, respectively. Additionally, higher values of $N_{c,<50}$ (100% of statistically significantly different T–w bins) and lower values of $N_{c,50–3000}$ (100% of statistically significantly different T–w bins) are observed in polluted environments compared to pristine conditions.

3) The parcel model, incorporating a superdroplet microphysics scheme, successfully reproduces the observed dependencies of cloud microphysical properties on environmental parameters. The simulations underscore the crucial role of cloud-base temperature and aerosol conditions in shaping cloud microphysical properties (i.e., LWC, D_e, $N_{c,<50}$, and $N_{c,50–3000}$). Warmer cloud bases enhance condensation rates, leading to fewer but larger droplets due to reduced CCN activation [100% of statistically significant bins with lower $N_{c,<50}$ and 79.2% with higher $N_{c,50–3000}$ in the w– ($T - T_{CB}$) parameter space], while collision–coalescence rates remain similar to those in colder cloud bases. This indicates that warmer cloud bases initially produce larger droplets through condensation, with the subsequent collision–coalescence further amplifying the size difference of particles between warmer and colder conditions. In contrast, polluted environments with higher aerosol concentrations produce a greater number of smaller droplets due to increased CCN activation, as indicated by 100% of statistically significant bins with lower $N_{c,<50}$ in the w–($T - T_{CB}$) parameter space. Although collision–coalescence rates also increase in polluted conditions, condensation (with 100% of statistically significant bins showing smaller activated droplet diameters) remains the primary mechanism driving the observed differences in cloud microphysical properties ($N_{c,<50}$ and $N_{c,50–3000}$) compared to pristine conditions.

Combining observations, machine learning, and numerical models provides a powerful approach for exploring and understanding the complex relationships between cloud properties and environmental parameters. This study presents observed quantitative patterns that characterize convective cloud microphysical properties as a function of key environmental parameters, specifically temperature and vertical velocity, offering valuable insights for evaluating numerical models and constraining physics parameterization schemes. The parcel model simulations also reveal mechanisms linking cloud-base temperature, aerosol conditions, and cloud microphysics, providing a deeper understanding of how environmental conditions shape cloud droplet properties and potentially influence cloud radiative and precipitation processes. Future research should focus on determining whether cloud-resolving models can *quantitatively* replicate these observed patterns and also tackle interactions between convective clouds and their surrounding environments, particularly aerosol–convection interactions, using data of isolated convective cells under varying environmental conditions observed during the TRACER field campaign. Given the limited aerosol observations available for this study, surface fine-mode aerosol mass concentration from the MERRA-2 reanalysis dataset is used as a proxy for CCN conditions. While this is reasonable, future studies would benefit from more comprehensive, collocated aerosol and cloud observations.

Acknowledgments. The authors appreciate Editor Dr. Adele L. Igel and three anonymous reviewers for their comments that contributed to improving the manuscript. This research was supported primarily by the U.S. Department of Energy's Atmospheric System Research, an Office of Science Biological and Environmental Research program, under Grant DE-SC0024317 and under Contract DE-SC0012704, and the National Science Foundation (Award AGS-2019968). The authors thank Paul Lawson for processing and sharing the ESCAPE SPEC Learjet data. The computing for this project was primarily performed at the University of Oklahoma (OU) Supercomputing Center for Education and Research (OSCER). The authors also acknowledge high-performance computing support from the Derecho system (https://doi.org/10.5065/qx9a-pg09) provided by the NSF National Center for Atmospheric Research (NCAR), sponsored by the National Science Foundation.

Data availability statement. ERA5 reanalysis dataset is available at https://doi.org/10.5065/BH6N-5N20. MERRA-2 dataset is available at https://doi.org/10.5067/KLICLTZ8EM9D (last access: 10 June 2024). NOAA *Geostationary Operational Environmental Satellite 17* (*GOES-17*) dataset is available at https://registry.opendata.aws/noaa-goes (last access: 10 June 2024). CDP-2 dataset on NRC Convair-580 is available at https://doi.org/10.26023/KKD0-PCW2-ZW0V (last access: 1 March 2024) (Ranjbar et al. 2023). 2D-S dataset on NRC Convair-580 is available at https://data.eol.ucar.edu/dataset/619.013 (last access: 1 March 2024) (McFarquhar et al. 2023). FCDP and 2D-S data on SPEC Learjet are available at https://doi.org/10.26023/P7WJ-1TFE-TZ01 (last access: 18 October 2022) (Lawson et al. 2022a). TRACER balloon-borne sounding dataset is available at https://doi.org/10.5439/1595321 (last access: 29 May 2024) (Keeler et al. 2024). AERONET AOD, version 3, dataset is available at https://aeronet.gsfc.nasa.gov/new_web/aerosols.html (last access: 29 March 2023) (Giles et al. 2019). The derived data can be downloaded at https://doi.org/10.5281/zenodo.17166818.

REFERENCES

Andreae, M. O., 2009: Correlation between cloud condensation nuclei concentration and aerosol optical thickness in remote and polluted regions. *Atmos. Chem. Phys.*, **9**, 543–556, https://doi.org/10.5194/acp-9-543-2009.

Barth, M. C., and Coauthors, 2015: The Deep Convective Clouds and Chemistry (DC3) field campaign. *Bull. Amer. Meteor. Soc.*, **96**, 1281–1309, https://doi.org/10.1175/BAMS-D-13-00290.1.

Bartman, P., and Coauthors, 2022: PySDM v1: Particle-based cloud modeling package for warm-rain microphysics and aqueous chemistry. *J. Open Source Software*, **7**, 3219, https://doi.org/10.21105/joss.03219.

Battan, L. J., 1953: Observations on the formation and spread of precipitation in convective clouds. *J. Atmos. Sci.*, **10**, 311–324,

https://doi.org/10.1175/1520-0469(1953)010<0311:OOTFAS>2.0.CO;2.

Beswick, K. M., M. W. Gallagher, A. R. Webb, E. G. Norton, and F. Perry, 2008: Application of the Aventech AIMMS20AQ airborne probe for turbulence measurements during the Convective Storm Initiation Project. *Atmos. Chem. Phys.*, **8**, 5449–5463, https://doi.org/10.5194/acp-8-5449-2008.

Biau, G., and E. Scornet, 2016: A random forest guided tour. *Test*, **25**, 197–227, https://doi.org/10.1007/s11749-016-0481-7.

Block, K., M. Haghighatnasab, D. G. Partridge, P. Stier, and J. Quaas, 2024: Cloud condensation nuclei concentrations derived from the CAMS reanalysis. *Earth Syst. Sci. Data*, **16**, 443–470, https://doi.org/10.5194/essd-16-443-2024.

Bony, S., and Coauthors, 2015: Clouds, circulation and climate sensitivity. *Nat. Geosci.*, **8**, 261–268, https://doi.org/10.1038/ngeo2398.

Breiman, L., 2001: Random forests. *Mach. Learn.*, **45**, 5–32, https://doi.org/10.1023/A:1010933404324.

Brokamp, C., R. Jandarov, M. Hossain, and P. Ryan, 2018: Predicting daily urban fine particulate matter concentrations using a random forest model. *Environ. Sci. Technol.*, **52**, 4173–4179, https://doi.org/10.1021/acs.est.7b05381.

Burke, A., N. Snook, D. J. Gagne II, S. McCorkle, and A. McGovern, 2020: Calibration of machine learning–based probabilistic hail predictions for operational forecasting. *Wea. Forecasting*, **35**, 149–168, https://doi.org/10.1175/WAF-D-19-0105.1.

Duan, G., and T. Takemi, 2021: Predicting urban surface roughness aerodynamic parameters using random forest. *J. Appl. Meteor. Climatol.*, **60**, 999–1018, https://doi.org/10.1175/JAMC-D-20-0266.1.

Emanuel, K. A., 1994: *Atmospheric Convection.* Oxford University Press, 580 pp., https://doi.org/10.1093/oso/9780195066302.001.0001.

Falk, N. M., A. L. Igel, and M. R. Igel, 2019: The relative impact of ice fall speeds and microphysics parameterization complexity on supercell evolution. *Mon. Wea. Rev.*, **147**, 2403–2415, https://doi.org/10.1175/MWR-D-18-0417.1.

Fan, J., and Coauthors, 2017: Cloud-resolving model intercomparison of an MC3E squall line case: Part I—Convective updrafts. *J. Geophys. Res. Atmos.*, **122**, 9351–9378, https://doi.org/10.1002/2017JD026622.

——, and Coauthors, 2024: Tracking Aerosol Convection Interaction Experiment (TRACER) Model Intercomparison Project (MIP) Roadmap. https://www.arm.gov/uploads/Roadmap-TRACER-MIP.pdf.

Gagne, D. J., II, A. McGovern, S. E. Haupt, R. A. Sobash, J. K. Williams, and M. Xue, 2017: Storm-based probabilistic hail forecasting with machine learning applied to convection-allowing ensembles. *Wea. Forecasting*, **32**, 1819–1840, https://doi.org/10.1175/WAF-D-17-0010.1.

Gelaro, R., and Coauthors, 2017: The Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2). *J. Climate*, **30**, 5419–5454, https://doi.org/10.1175/JCLI-D-16-0758.1.

Genuer, R., J.-M. Poggi, C. Tuleau-Malot, and N. Villa-Vialaneix, 2017: Random forests for big data. *Big Data Res.*, **9**, 28–46, https://doi.org/10.1016/j.bdr.2017.07.003.

Giaiotti, D. B., R. Steinacker, and F. Stel, 2007: *Atmospheric Convection: Research and Operational Forecasting Aspects.* CISM International Centre for Mechanical Sciences, Vol. 475, Springer, 227 pp., https://doi.org/10.1007/978-3-211-69291-2.

Giles, D. M., and Coauthors, 2019: Advancements in the Aerosol Robotic Network (AERONET) Version 3 database—Automated near-real-time quality control algorithm with improved cloud screening for sun photometer aerosol optical depth (AOD) measurements. *Atmos. Meas. Tech.*, **12**, 169–209, https://doi.org/10.5194/amt-12-169-2019.

Grange, S. K., D. C. Carslaw, A. C. Lewis, E. Boleti, and C. Hueglin, 2018: Random forest meteorological normalisation models for Swiss PM_{10} trend analysis. *Atmos. Chem. Phys.*, **18**, 6223–6239, https://doi.org/10.5194/acp-18-6223-2018.

Guo, H., J.-C. Golaz, L. J. Donner, B. Wyman, M. Zhao, and P. Ginoux, 2015: CLUBB as a unified cloud parameterization: Opportunities and challenges. *Geophys. Res. Lett.*, **42**, 4540–4547, https://doi.org/10.1002/2015GL063672.

Hagos, S., and R. Houze, 2016: Atmospheric system research treatment of convection in next-generation climate models: Challenges and opportunities workshop report. U.S. Department of Energy Tech. Rep. DOE/SC-ASR-16-002, 49 pp., https://www.arm.gov/publications/programdocs/doe-sc-asr-16-002.pdf.

Hartmann, D. L., M. E. Ockert-Bell, and M. L. Michelsen, 1992: The effect of cloud type on Earth's energy balance: Global analysis. *J. Climate*, **5**, 1281–1304, https://doi.org/10.1175/1520-0442(1992)005<1281:TEOCTO>2.0.CO;2.

Hasekamp, O. P., E. Gryspeerdt, and J. Quaas, 2019: Analysis of polarimetric satellite measurements suggests stronger cooling due to aerosol-cloud interactions. *Nat. Commun.*, **10**, 5405, https://doi.org/10.1038/s41467-019-13372-2.

Haupt, S. E., and Coauthors, 2022: The history and practice of AI in the environmental sciences. *Bull. Amer. Meteor. Soc.*, **103**, E1351–E1370, https://doi.org/10.1175/BAMS-D-20-0234.1.

Herman, G. R., and R. S. Schumacher, 2018: Money doesn't grow on trees, but forecasts do: Forecasting extreme precipitation with random forests. *Mon. Wea. Rev.*, **146**, 1571–1600, https://doi.org/10.1175/MWR-D-17-0250.1.

Hersbach, H., and Coauthors, 2020: The ERA5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, **146**, 1999–2049, https://doi.org/10.1002/qj.3803.

Heymsfield, A., and P. Willis, 2014: Cloud conditions favoring secondary ice particle production in tropical maritime convection. *J. Atmos. Sci.*, **71**, 4500–4526, https://doi.org/10.1175/JAS-D-14-0093.1.

Hill, A. J., G. R. Herman, and R. S. Schumacher, 2020: Forecasting severe weather with random forests. *Mon. Wea. Rev.*, **148**, 2135–2161, https://doi.org/10.1175/MWR-D-19-0344.1.

Ho, T. K., 1995: Random decision forests. *Proc. Third Int. Conf. on Document Analysis and Recognition*, Montreal, Quebec, Canada, Institute of Electrical and Electronics Engineers, 278–282, https://doi.org/10.1109/ICDAR.1995.598994.

Holdridge, D., 2020: Balloon-borne sounding system (SONDE) instrument handbook. DOE Office of Science Atmospheric Radiation Measurement (ARM) Program Tech. Rep. DOE/SC-ARM-TR-029, 38 pp., https://doi.org/10.2172/1020712.

Hu, Y., and Coauthors, 2021: Dependence of ice microphysical properties on environmental parameters: Results from HAIC-HIWC Cayenne field campaign. *J. Atmos. Sci.*, **78**, 2957–2981, https://doi.org/10.1175/JAS-D-21-0015.1.

——, and Coauthors, 2022: Dependence of ice crystal size distributions in high ice water content conditions on environmental conditions: Results from the HAIC-HIWC Cayenne campaign. *J. Atmos. Sci.*, **79**, 3103–3134, https://doi.org/10.1175/JAS-D-22-0008.1.

Huang, Y., Y. Liu, Y. Liu, and J. C. Knievel, 2019: Budget analyses of a record-breaking rainfall event in the coastal metropolitan city of Guangzhou, China. *J. Geophys. Res. Atmos.*, **124**, 9391–9406, https://doi.org/10.1029/2018JD030229.

——, Y. Wang, L. Xue, X. Wei, L. Zhang, and H. Li, 2020: Comparison of three microphysics parameterization schemes in the WRF model for an extreme rainfall event in the coastal metropolitan city of Guangzhou, China. *Atmos. Res.*, **240**, 104939, https://doi.org/10.1016/j.atmosres.2020.104939.

——, and Coauthors, 2021: Microphysical processes producing high ice water contents (HIWCs) in tropical convective clouds during the HAIC-HIWC field campaign: Evaluation of simulations using bulk microphysical schemes. *Atmos. Chem. Phys.*, **21**, 6919–6944, https://doi.org/10.5194/acp-21-6919-2021.

Jensen, M. P., and Coauthors, 2016: The Midlatitude Continental Convective Clouds Experiment (MC3E). *Bull. Amer. Meteor. Soc.*, **97**, 1667–1686, https://doi.org/10.1175/BAMS-D-14-00228.1.

——, and Coauthors, 2023: Tracking Aerosol Convection Interactions Experiment (TRACER) field campaign report. U.S. Department of Energy, Atmospheric Radiation Measurement user facility Tech. Rep. Doe/SC-Arm-23-038, 132 pp., https://doi.org/10.2172/2202672.

——, and Coauthors, 2025: Studying aerosol, clouds, and air quality in the coastal urban environment of southeastern Texas. *Bull. Amer. Meteor. Soc.*, https://doi.org/10.1175/BAMS-D-23-0331.1.

Keeler, E., K. Burk, and J. Kyrouac, 2024: Balloon-borne sounding system (SONDEWNPN). ARM Data Center, accessed 29 May 2024, https://doi.org/10.5439/1595321.

Khain, A. P., and Coauthors, 2015: Representation of microphysical processes in cloud-resolving models: Spectral (bin) microphysics versus bulk parameterization. *Rev. Geophys.*, **53**, 247–322, https://doi.org/10.1002/2014RG000468.

Kollias, P., and Coauthors, 2025: Experiment of Sea Breeze Convection, Aerosols, Precipitation, and Environment (ESCAPE). *Bull. Amer. Meteor. Soc.*, **106**, E310–E332, https://doi.org/10.1175/BAMS-D-23-0014.1.

Koren, I., G. Feingold, and L. A. Remer, 2010: The invigoration of deep convective clouds over the Atlantic: Aerosol effect, meteorology or retrieval artifact? *Atmos. Chem. Phys.*, **10**, 8855–8872, https://doi.org/10.5194/acp-10-8855-2010.

Kuang, C., A. Singh, J. Howie, C. Salwen, and C. Hayes, 2024: Scanning mobility particle sizer (AOSSMPS). ARM Data Center, accessed 27 September 2024, https://doi.org/10.5439/1476898.

Lacis, A. A., and J. Hansen, 1974: A parameterization for the absorption of solar radiation in the Earth's atmosphere. *J. Atmos. Sci.*, **31**, 118–133, https://doi.org/10.1175/1520-0469(1974)031<0118:APFTAO>2.0.CO;2.

Lance, S., C. A. Brock, D. Rogers, and J. A. Gordon, 2010: Water droplet calibration of the Cloud Droplet Probe (CDP) and in-flight performance in liquid, ice and mixed-phase clouds during ARCPAC. *Atmos. Meas. Tech.*, **3**, 1683–1706, https://doi.org/10.5194/amt-3-1683-2010.

Lawson, P., and W. A. Cooper, 1990: Performance of some airborne thermometers in clouds. *J. Atmos. Oceanic Technol.*, **7**, 480–494, https://doi.org/10.1175/1520-0426(1990)007<0480:POSATI>2.0.CO;2.

——, D. O'Connor, P. Zmarzly, K. Weaver, B. Baker, Q. Mo, and H. Jonsson, 2006: The 2D-S (stereo) probe: Design and preliminary tests of a new airborne, high-speed, high-resolution particle imaging probe. *J. Atmos. Oceanic Technol.*, **23**, 1462–1477, https://doi.org/10.1175/JTECH1927.1.

——, C. Gurganus, S. Woods, and R. Bruintjes, 2017: Aircraft observations of cumulus microphysics ranging from the tropics to midlatitudes: Implications for a "new" secondary ice process. *J. Atmos. Sci.*, **74**, 2899–2920, https://doi.org/10.1175/JAS-D-17-0033.1.

——, Q. Mo, and R. Bruintjes, 2022a: SPEC Learjet instrument data. UCAR/NCAR - Earth Observing Laboratory, accessed 18 October 2022, https://doi.org/10.26023/P7WJ-1TFE-TZ01.

——, R. Bruintjes, S. Woods, and C. Gurganus, 2022b: Coalescence and secondary ice development in cumulus congestus clouds. *J. Atmos. Sci.*, **79**, 953–972, https://doi.org/10.1175/JAS-D-21-0188.1.

Li, H., Y. Huang, Y. Luo, H. Xiao, M. Xue, X. Liu, and L. Feng, 2023: Does "right" simulated extreme rainfall result from the "right" representation of rain microphysics? *Quart. J. Roy. Meteor. Soc.*, **149**, 3220–3249, https://doi.org/10.1002/qj.4553.

Loken, E. D., A. J. Clark, and C. D. Karstens, 2020: Generating probabilistic next-day severe weather forecasts from convection-allowing ensembles using random forests. *Wea. Forecasting*, **35**, 1605–1631, https://doi.org/10.1175/WAF-D-19-0258.1.

Lyman, J. M., and G. C. Johnson, 2023: Global high-resolution random forest regression maps of ocean heat content anomalies using in situ and satellite data. *J. Atmos. Oceanic Technol.*, **40**, 575–586, https://doi.org/10.1175/JTECH-D-22-0058.1.

Masson-Delmotte, V., and Coauthors, 2021: *Climate Change 2021: The Physical Science Basis.* Cambridge University Press, 2392 pp., https://doi.org/10.1017/9781009157896.

McFarquhar, G. M., and A. J. Heymsfield, 1998: The definition and significance of an effective radius for ice clouds. *J. Atmos. Sci.*, **55**, 2039–2052, https://doi.org/10.1175/1520-0469(1998)055<2039:TDASOA>2.0.CO;2.

——, M. Wolde, L. Nichman, K. Bala, and S. Patil, 2023: ESCAPE: Optical Array Probes (OAP) microphysics data. UCAR/NCAR - Earth Observing Laboratory, accessed 1 March 2024, https://data.eol.ucar.edu/dataset/619.013.

McGovern, A., R. Lagerquist, D. J. Gagne II, G. E. Jergensen, K. L. Elmore, C. R. Homeyer, and T. Smith, 2019: Making the black box more transparent: Understanding the physical implications of machine learning. *Bull. Amer. Meteor. Soc.*, **100**, 2175–2199, https://doi.org/10.1175/BAMS-D-18-0195.1.

——, A. Bostrom, M. McGraw, R. J. Chase, D. J. Gagne II, I. Ebert-Uphoff, K. D. Musgrave, and A. Schumacher, 2024: Identifying and categorizing bias in AI/ML for Earth sciences. *Bull. Amer. Meteor. Soc.*, **105**, E567–E583, https://doi.org/10.1175/BAMS-D-23-0196.1.

Mecikalski, J. R., T. N. Sandmæl, E. M. Murillo, C. R. Homeyer, K. M. Bedka, J. M. Apke, and C. P. Jewett, 2021: A random-forest model to assess predictor importance and nowcast severe storms using high-resolution radar–GOES satellite–lightning observations. *Mon. Wea. Rev.*, **149**, 1725–1746, https://doi.org/10.1175/MWR-D-19-0274.1.

Morrison, H., and J. Milbrandt, 2011: Comparison of two-moment bulk microphysics schemes in idealized supercell thunderstorm simulations. *Mon. Wea. Rev.*, **139**, 1103–1130, https://doi.org/10.1175/2010MWR3433.1.

——, and Coauthors, 2020: Confronting the challenge of modeling cloud and precipitation microphysics. *J. Adv. Model. Earth Syst.*, **12**, e2019MS001689, https://doi.org/10.1029/2019MS001689.

Nesbitt, S. W., and Coauthors, 2021: A storm safari in subtropical South America: Proyecto RELAMPAGO. *Bull. Amer. Meteor. Soc.*, **102**, E1621–E1644, https://doi.org/10.1175/BAMS-D-20-0029.1.

Oertel, A., and S. Schemm, 2021: Quantifying the circulation induced by convective clouds in kilometer-scale simulations. *Quart. J. Roy. Meteor. Soc.*, **147**, 1752–1766, https://doi.org/10.1002/qj.3992.

Pan, L. L., and Coauthors, 2017: The Convective Transport of Active Species in the Tropics (CONTRAST) experiment. *Bull. Amer. Meteor. Soc.*, **98**, 106–128, https://doi.org/10.1175/BAMS-D-14-00272.1.

Pan, Z., D. Rosenfeld, Y. Zhu, F. Mao, W. Gong, L. Zang, and X. Lu, 2021: Observational quantification of aerosol invigoration for deep convective cloud lifecycle properties based on geostationary satellite. *J. Geophys. Res. Atmos.*, **126**, e2020JD034275, https://doi.org/10.1029/2020JD034275.

Pedregosa, F., and Coauthors, 2011: Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.*, **12**, 2825–2830.

Radford, J. T., and G. M. Lackmann, 2023: Improving High-Resolution Ensemble Forecast (HREF) system mesoscale snowband forecasts with random forests. *Wea. Forecasting*, **38**, 1695–1706, https://doi.org/10.1175/WAF-D-23-0005.1.

Ranjbar, K., K. Bala, and M. Woldo, 2023: ESCAPE: CDP and FCDP scattering probe data. UCAR/NCAR - Earth Observing Laboratory, accessed 1 March 2024, https://doi.org/10.26023/KKD0-PCW2-ZW0V.

Reid, J. S., and Coauthors, 2023: The coupling between tropical meteorology, aerosol lifecycle, convection, and radiation during the Cloud, Aerosol and Monsoon Processes Philippines Experiment (CAMP^{2}Ex). *Bull. Amer. Meteor. Soc.*, **104**, E1179–E1205, https://doi.org/10.1175/BAMS-D-21-0285.1.

Rigatti, S. J., 2017: Random forest. *J. Insur. Med.*, **47**, 31–39, https://doi.org/10.17849/insm-47-01-31-39.1.

Rio, C., A. D. Del Genio, and F. Hourdin, 2019: Ongoing breakthroughs in convective parameterization. *Curr. Climate Change Rep.*, **5**, 95–111, https://doi.org/10.1007/s40641-019-00127-w.

Robert, A., and Houze, Jr, 1982: Cloud clusters and large-scale vertical motions in the tropics. *J. Meteor. Soc. Japan*, **60**, 396–410, https://doi.org/10.2151/jmsj1965.60.1_396.

Shima, S., K. Kusano, A. Kawano, T. Sugiyama, and S. Kawahara, 2009: The super-droplet method for the numerical simulation of clouds and precipitation: A particle-based and probabilistic microphysics model coupled with a non-hydrostatic model. *Quart. J. Roy. Meteor. Soc.*, **135**, 1307–1320, https://doi.org/10.1002/qj.441.

Smith, R. K., 2013: *The Physics and Parameterization of Moist Atmospheric Convection*. Nato Science Series C, Vol. 505, Springer, 498 pp.

Stevens, B., 2005: Atmospheric moist convection. *Annu. Rev. Earth Planet. Sci.*, **33**, 605–643, https://doi.org/10.1146/annurev.earth.33.092203.122658.

Stier, P., 2016: Limitations of passive remote sensing to constrain global cloud condensation nuclei. *Atmos. Chem. Phys.*, **16**, 6595–6607, https://doi.org/10.5194/acp-16-6595-2016.

Strapp, J., and Coauthors, 2016: The High Ice Water Content (HIWC) study of deep convective clouds: Science and technical plan. FAA Rep. DOT/FAA/TC-14/31, 105 pp.

Taszarek, M., N. Pilguj, J. T. Allen, V. Gensini, H. E. Brooks, and P. Szuster, 2021: Comparison of convective parameters derived from ERA5 and MERRA-2 with rawinsonde data over Europe and North America. *J. Climate*, **34**, 3211–3237, https://doi.org/10.1175/JCLI-D-20-0484.1.

——, B. Czernecki, and P. Szuster, 2023: thundeR - A rawinsonde package for processing convective parameters and visualizing atmospheric profiles. *11th European Conf. on Severe Storms*, Bucharest, Romania, ESSL, ECSS2023-28, https://doi.org/10.5194/ecss2023-28.

Toon, O. B., and Coauthors, 2016: Planning, implementation, and scientific goals of the studies of emissions and atmospheric composition, clouds and climate coupling by regional surveys (SEAC^{4}RS) field mission. *J. Geophys. Res. Atmos.*, **121**, 4967–5009, https://doi.org/10.1002/2015JD024297.

Tromeur, E., and W. B. Rossow, 2010: Interaction of tropical deep convection with the large-scale circulation in the MJO. *J. Climate*, **23**, 1837–1853, https://doi.org/10.1175/2009JCLI3240.1.

Tung, W.-W., and M. Yanai, 2002: Convective momentum transport observed during the TOGA COARE IOP. Part II: Case studies. *J. Atmos. Sci.*, **59**, 2535–2549, https://doi.org/10.1175/1520-0469(2002)059<2535:CMTODT>2.0.CO;2.

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

——, 1977: The influence of pollution on the shortwave albedo of clouds. *J. Atmos. Sci.*, **34**, 1149–1152, https://doi.org/10.1175/1520-0469(1977)034<1149:TIOPOT>2.0.CO;2.

Varble, A. C., and Coauthors, 2021: Utilizing a storm-generating hotspot to study convective cloud transitions: The CACTI experiment. *Bull. Amer. Meteor. Soc.*, **102**, E1597–E1620, https://doi.org/10.1175/BAMS-D-20-0030.1.

——, A. L. Igel, H. Morrison, W. W. Grabowski, and Z. J. Lebo, 2023: Opinion: A critical evaluation of the evidence for aerosol invigoration of deep convection. *Atmos. Chem. Phys.*, **23**, 13 791–13 808, https://doi.org/10.5194/acp-23-13791-2023.

Veals, P. G., A. C. Varble, J. O. Russell, J. C. Hardin, and E. J. Zipser, 2022: Indications of a decrease in the depth of deep convective cores with increasing aerosol concentration during the CACTI campaign. *J. Atmos. Sci.*, **79**, 705–722, https://doi.org/10.1175/JAS-D-21-0119.1.

Williams, J. K., 2014: Using random forests to diagnose aviation turbulence. *Mach. Learn.*, **95**, 51–70, https://doi.org/10.1007/s10994-013-5346-7.

Xue, L., and Coauthors, 2017: Idealized simulations of a squall line from the MC3E field campaign applying three bin microphysics schemes: Dynamic and thermodynamic structure. *Mon. Wea. Rev.*, **145**, 4789–4812, https://doi.org/10.1175/MWR-D-16-0385.1.