



1 **Earth System Model Aerosol-Cloud Diagnostics Package**
2 **(ESMAC Diags) Version 2: Assessments of Aerosols, Clouds and**
3 **Aerosol-Cloud Interactions Through Field Campaign and Long-**
4 **Term Observations**

5 Shuaiqi Tang¹, Adam C. Varble¹, Jerome D. Fast¹, Kai Zhang¹, Peng Wu¹, Xiquan Dong², Fan
6 Mei¹, Mikhail Pekour¹, Joseph C. Hardin³, and Po-Lun Ma¹

7 ¹Pacific Northwest National Laboratory, Richland, WA, USA

8 ²University of Arizona, Tucson, AZ, USA

9 ³Unaffiliated scientist

10 *Correspondence to:* Shuaiqi Tang (shuaiqi.tang@pnnl.gov)

11

12 **Abstract.**

13 Poor representations of aerosols, clouds and aerosol-cloud interactions (ACI) in Earth System Models
14 (ESMs) have long been the largest uncertainties in predicting global climate change. Huge efforts have
15 been made to improve the representation of these processes in ESMs, and key to these efforts is
16 evaluation of ESM simulations with observations. Most well-established ESM diagnostics packages focus
17 on the climatological features; however, they are lack of the process-level understanding and
18 representations of aerosols, clouds, and ACI. In this study, we developed an ESM aerosol-cloud
19 diagnostics package (ESMAC Diags) to facilitate routine evaluation of aerosols, clouds and aerosol-cloud
20 interactions simulated by the Department of Energy's (DOE) Energy Exascale Earth System Model
21 (E3SM). This paper documents its version 2 functionality (ESMAC Diags v2), which has substantial
22 updates from its version 1 (Tang et al., 2022a). The simulated aerosol and cloud properties have been
23 extensively compared with in-situ and remote-sensing measurements from aircraft, ship, surface and
24 satellite platforms in ESMAC Diags v2. It currently includes six field campaigns and two permanent sites
25 covering four geographical regions: Eastern North Atlantic, Central U.S., Northeastern Pacific and
26 Southern Ocean, where frequent liquid or mixed-phase clouds are present and extensive measurements
27 are available from the DOE Atmospheric Radiation Measurement user facility and other agencies.
28 ESMAC Diags v2 generates various types of single-variable and multi-variable diagnostics, including
29 percentiles, histograms, joint histograms and heatmaps, to evaluate model representation of aerosols,
30 clouds, and aerosol-cloud interactions. Select examples highlighting ESMAC Diags capabilities are
31 shown using E3SM version 2 (E3SMv2). E3SMv2 in general can reasonably reproduces many observed
32 aerosol and cloud properties, with biases in some variables such as aerosol particle and cloud droplet sizes
33 and number concentrations. The coupling of aerosol and cloud number concentrations may be too strong
34 in E3SMv2, possibly indicating a bias in processes that control aerosol activation. Furthermore, the liquid
35 water path adjustment to perturbed cloud droplet number concentration behaves differently in E3SMv2
36 and observations, which warrants a further study to improve the cloud microphysics parameterizations in
37 E3SMv2.

38



39 1. Introduction

40 Poor representations of aerosols, clouds and aerosol-cloud interactions (ACI) in Earth System Models
41 (ESMs) have long been the largest uncertainties in predicting global climate change (IPCC, 2021).
42 Challenges come from several aspects: first, there are many aerosol properties (e.g., number, size, phase,
43 shape, composition) and cloud micro- and macro-physical properties (e.g., fraction, water content,
44 number and size of liquid and ice hydrometeors) that affect Earth's climate. Coincident measurements of
45 these properties remain largely under-sampled due to substantial spatiotemporal variability and logistical
46 difficulties for making such measurements. Second, there are complex interactive processes between
47 aerosols, clouds, and ambient meteorological conditions, many of which are not fully understood, but are
48 critical to properly interpreting relationships between observable properties. Third, many ACI processes
49 are nonlinear, multi-scale processes that involve feedbacks depending on cloud types and meteorological
50 regimes, which also shift in space and time, presenting challenges for assessing causal effect and
51 representing such processes in ESMs.

52 Huge efforts have been made to improve the representation of aerosols, clouds and ACI in ESMs. Key to
53 these efforts is evaluation of ESM simulations with observations. Many modeling centers have developed
54 standardized diagnostics packages to document ESM performance. For aerosol and cloud properties, most
55 diagnostic packages rely heavily on satellite measurements as evaluation data (e.g., AMWG, 2021;
56 E3SM, 2021; Eyring et al., 2016; Gleckler et al., 2016; Maloney et al., 2019; Myhre et al., 2013; Schulz
57 et al., 2006). Satellite remote sensing measurements have global or near global coverage but limited
58 spatial and temporal resolution. They are also unable to retrieve some variables, especially for aerosol
59 properties such as cloud condensation nuclei (CCN) number concentration, while many cloud
60 microphysical retrievals such as droplet number concentration have large uncertainties (e.g., Grosvenor et
61 al., 2018). This limits their application to robustly quantify aerosols, clouds and ACI processes. In-situ
62 measurements from ground, aircraft or ship platforms from field campaigns are also used in a few
63 projects to evaluate ESMs (e.g., Reddington et al., 2017; Watson-Parris et al., 2019; Tang et al., 2022a;
64 Zhang et al., 2020). Some of these field campaigns were conducted over remote or poorly sampled
65 locations, which are highly valuable for model evaluation despite limited spatial coverage and time
66 periods. Moreover, the DOE Atmospheric Radiation Measurement (ARM) user facility has conducted
67 continuous field measurements at a few sites for multiple years. These long-term high-resolution field
68 measurements have also been demonstrated to be valuable for evaluating ESMs (e.g., Zhang et al., 2020).

69 In response to the need for more ESM diagnostics for evaluating ACI processes, Tang et al. (2022a)
70 developed an ESM aerosol-cloud diagnostics package (ESMAC Diags) to facilitate the routine evaluation
71 of aerosols, clouds and ACI simulated by the Department of Energy's (DOE) Energy Exascale Earth
72 System Model (E3SM, Golaz et al., 2019). It includes diagnostics that leverage in-situ measurements
73 from multiple platforms during six field campaigns since 2013, which are not included in previous
74 diagnostics tools (e.g., Reddington et al., 2017). Version 1 of ESMAC Diags (ESMAC Diags v1, Tang et
75 al., 2022a) mainly focuses on aerosol properties. We present here version 2 of ESMAC Diags (ESMAC
76 Diags v2) that is a direct extension of ESMAC Diags v1 with two major additions:

77 1. measurements from satellite and long-term diagnostics at the ARM Southern Great Plains
78 (SGP) and Eastern North Atlantic (ENA) sites.

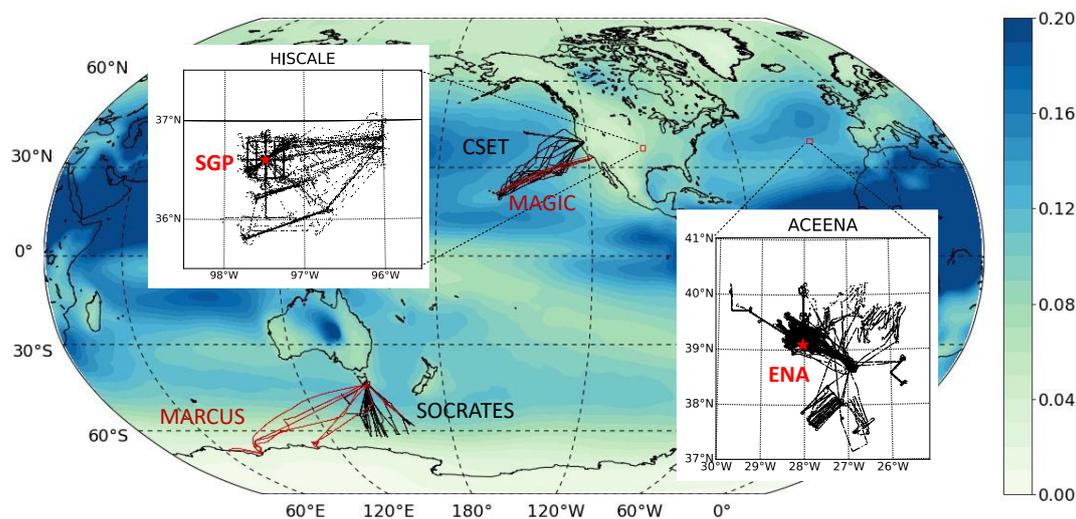


79 2. diagnostics for cloud properties and aerosol-cloud interactions.

80 The new measurements, as well as major data quality controls are introduced in Section 2. Additional
 81 discussions on retrieval uncertainties of cloud microphysical properties are performed in Section 3.
 82 Details of the code structure of ESMAC Diags v2, which is substantially changed since version 1, are
 83 described in Section 4. Section 5 provides selected examples of single-variable and multi-variable
 84 diagnostics using ESMAC Diags v2 to highlight its capabilities. Lastly, Section 6 provides a summary.

85 **2. Aerosol and cloud measurements from ground, aircraft, ship and satellite platforms**

86 Following the initial development in version 1, ESMAC Diags v2 continues to focus on six field
 87 campaigns conducted in four geographical regions: the Central U.S. (CUS, where the ARM Southern
 88 Great Plains (SGP) site is located), Eastern North Atlantic (ENA), Northeastern Pacific (NEP), and
 89 Southern Ocean (SO). Information on the six field campaigns is shown in Table 1 and their locations are
 90 shown in Figure 1, each reproduced from Table 1 and Figure 3 in Tang et al. (2022a).



91
 92 Figure 1. Aircraft (black) and ship (red) tracks for the six field campaigns. Red stars in the
 93 enlarged map indicate two ARM fixed sites: SGP and ENA, that have long-term
 94 measurements available for model diagnostics. Overlaid is aerosol optical depth at
 95 550nm averaged from 2014 to 2018 simulated in E3SMv1. (Reproduced from Figure 3 in
 96 Tang et al., 2022a)

97 Table 1. Descriptions of the field campaigns used in this study. (Reproduced from Table 1
 98 in Tang et al., 2022a)

Campaign*	Period	Platform	Typical Conditions	Reference
HI-SCALE	IOP1: 24 Apr – 21 May 2016 IOP2: 28 Aug – 24 Sep 2016	Ground, aircraft (IOP1: 17 flights, IOP2: 21 flights)	Continental cumulus with high aerosol loading	(Fast et al., 2019)



ACE-ENA	IOP1: 21 Jun – 20 Jul 2017 IOP2: 15 Jan – 18 Feb 2018	Ground, aircraft (IOP1: 20 flights, IOP2: 19 flights)	Marine stratocumulus with low aerosol loading	(Wang et al., 2021)
MAGIC	Oct 2012 – Sep 2013	Ship (18 legs)	Marine stratocumulus to cumulus transition with low aerosol loading	(Lewis and Teixeira, 2015; Zhou et al., 2015)
CSET	1 Jul – 15 Aug 2015	Aircraft (16 flights)	Same as above	(Albrecht et al., 2019)
MARCUS	Oct 2017 – Apr 2018	Ship (4 legs)	Marine liquid and mixed phase clouds with low aerosol loading	(McFarquhar et al., 2021)
SOCRATES	15 Jan – 24 Feb, 2018	Aircraft (14 flights)	Same as above	(McFarquhar et al., 2021)

99 * Full names of the listed field campaigns:
 100 HI-SCALE: Holistic Interactions of Shallow Clouds, Aerosols and Land Ecosystems
 101 ACE-ENA: Aerosol and Cloud Experiments in the Eastern North Atlantic
 102 MAGIC: Marine ARM GCSS Pacific Cross-section Intercomparison (GPCI) Investigation of Clouds
 103 CSET: Cloud System Evolution in the Trades
 104 MARCUS: Measurements of Aerosols, Radiation and Clouds over the Southern Ocean
 105 SOCRATES: Southern Ocean Cloud Radiation and Aerosol Transport Experimental Study
 106

107 The collection and processing of observations are the most time-consuming part of developing ESMAC
 108 Diags, which also impacts the reliability of conclusions drawn from the model diagnostics. In this section,
 109 we introduce the data used in ESMAC Diags v2, existing quality issues in some datasets, and treatments
 110 to address these quality issues. Some variables are difficult to directly measure or have limited in-situ
 111 sampling and thus must be derived from remote sensing measurements using retrieval algorithms. In
 112 Section 3, we further discuss the uncertainty and reliability of some cloud retrieval products via
 113 comparisons with in-situ aircraft measurements.

114 2.1. Data availability

115 All measurements, instruments, and data products used in the six field campaigns and two long-term sites
 116 in ESMAC Diags v2 are shown in Table 2. Further details of the measurements, data product names, and
 117 DOIs are given in Tables S1 to S6 (for field campaigns) and Tables S7 and S8 (for SGP and ENA sites) in
 118 the supplementary material. To allow maximum overlapping of key measurements while also ensuring a
 119 long enough period for statistical evaluation, we select the periods of 1 Jan 2011 – 31 Dec 2020 for SGP
 120 and 1 Jan 2016 – 31 Dec 2018 for ENA for long-term analyses. In addition to the aerosol measurements
 121 discussed in Tang et al. (2022a), we incorporate more cloud and radiation measurements, as well as
 122 geostationary satellite retrievals using Visible Infrared Solar-Infrared Split Window Technique (VISST)
 123 (Minnis et al., 2008; Minnis et al., 2011) algorithm. The VISST products archived by ARM cover
 124 approximately 10° by 10° regions in 0.5° by 0.5° resolution centered over ARM sites. Moreover, ARM
 125 recently released products consisting of merged aerosol particle and cloud droplet size distributions from
 126 aircraft measurements for HI-SCALE and ACE-ENA campaigns. These data are now used in ESMAC
 127 Diags v2.

128 Table 2: List of instruments and measurements used in ESMAC Diags v2.



Platform	Measurements	Instruments / data products	Available campaigns
Ground	Surface temperature, relative humidity, wind, pressure, precipitation; upper-level temperature, relative humidity, wind	Surface meteorological station (MET), ARM best estimate (ARMBE) products	HI-SCALE, ACE-ENA, SGP, ENA
	Longwave and shortwave radiation, cloud fraction	ARM best estimate (ARMBE) products	HI-SCALE, ACE-ENA, SGP, ENA
	Aerosol number concentration	Condensation particle counter (CPC), Condensation particle counter – fine (CPCF), Condensation particle counter – ultrafine (CPCU), Ultra-high sensitivity aerosol spectrometer (UHSAS), Scanning mobility particle sizer (SMPS)	HI-SCALE, ACE-ENA, SGP, ENA
	Aerosol size distribution	Ultra-high sensitivity aerosol spectrometer (UHSAS), Scanning mobility particle sizer (SMPS), Nano scanning mobility particle sizer (nanoSMPS)	HI-SCALE, ACE-ENA, SGP, ENA
	Aerosol composition	Aerosol chemical speciation monitor (ACSM)	HI-SCALE, ACE-ENA, SGP, ENA
	CCN number concentration	Cloud condensation nuclei (CCN) counter	HI-SCALE, ACE-ENA, SGP, ENA
	Cloud optical depth	Multifilter rotating shadowband radiometer (MFRSR)	HI-SCALE, ACE-ENA, SGP, ENA
	Cloud droplet number concentration	Cloud droplet number concentration retrieval (Ndrop), cloud retrieval from Wu et al. (2020)	HI-SCALE, ACE-ENA, SGP, ENA
	Cloud droplet effective radius	Multifilter rotating shadowband radiometer (MFRSR), cloud retrieval from Wu et al. (2020)	HI-SCALE, ACE-ENA, SGP, ENA
	Cloud liquid water path	Microwave radiometer (MWR), ARM best estimate (ARMBE) products	HI-SCALE, ACE-ENA, SGP, ENA
	Cloud base height, cloud top height	Active remote sensing of clouds (ARSCl)	HI-SCALE, ACE-ENA, SGP, ENA
Satellite	TOA shortwave and longwave radiation	Geostationary satellite-based retrievals using Visible Infrared Solar-Infrared Split Window Technique (VISST) algorithm	HI-SCALE, ACE-ENA, MAGIC, MARCUS, SGP, ENA
	cloud fraction; height, pressure and temperature at cloud top	Geostationary satellite-based retrievals using Visible Infrared Solar-Infrared Split Window Technique (VISST) algorithm	HI-SCALE, ACE-ENA, MAGIC, MARCUS, SGP, ENA
	liquid water path; cloud optical depth; droplet effective radius	Geostationary satellite-based retrievals using Visible Infrared Solar-Infrared Split Window Technique (VISST) algorithm	HI-SCALE, ACE-ENA, MAGIC, MARCUS, SGP, ENA
	Cloud droplet number concentration	Retrieved from VISST data using the algorithm in Bennartz (2007)	HI-SCALE, ACE-ENA, MAGIC, MARCUS, SGP, ENA
Aircraft	Navigation information and meteorological parameters	Interagency working group for airborne data and telemetry systems (IWG)	HI-SCALE, ACE-ENA
	Aerosol number concentration	Condensation particle counter (CPC), Condensation particle counter – ultrafine (CPCU), Condensation nuclei counter (CNC), Ultra-high sensitivity aerosol spectrometer (UHSAS), Passive cavity aerosol spectrometer (PCASP)	HI-SCALE, ACE-ENA, CSET, SOCRATES
	Aerosol size distribution	Ultra-high sensitivity aerosol spectrometer (UHSAS), Fast integrated mobility spectrometer (FIMS), Passive cavity aerosol spectrometer (PCASP), Best estimate aerosol size distribution (BEASD)	HI-SCALE, ACE-ENA, CSET, SOCRATES
	Aerosol composition	High-resolution time-of-flight aerosol mass spectrometer (AMS)	HI-SCALE, ACE-ENA



	CCN number concentration	Cloud condensation nuclei (CCN) counter	HI-SCALE, ACE-ENA, SOCRATES
	Cloud liquid water content	Water content measuring system (WCM), PMS-King Liquid Water Content (LWC)	HI-SCALE, ACE-ENA, CSET, SOCRATES
	Cloud droplet number size distribution	1DC, 2DC, 2DS, CDP, Cloud probe merged size distribution (mergedSD)	HI-SCALE, ACE-ENA, CSET, SOCRATES
Ship	Navigation information and meteorological parameters	Meteorological station (MET)	MAGIC, MARCUS
	Aerosol number concentration	Condensation particle counter (CPC), Ultra-high sensitivity aerosol spectrometer (UHSAS)	MAGIC, MARCUS
	Aerosol size distribution	Ultra-high sensitivity aerosol spectrometer (UHSAS)	MAGIC, MARCUS
	CCN number concentration	Cloud condensation nuclei (CCN) counter	MAGIC, MARCUS
	Cloud liquid water path	Microwave radiometer (MWR)	MAGIC, MARCUS
	Cloud droplet number concentration, cloud effective radius	Cloud retrieval from Wu et al. (2020)	MAGIC

129

130 All the observational data are quality controlled with their time resolution re-scaled to that suitable for
 131 evaluating E3SM. Currently, ground, ship and satellite measurements are re-scaled to a 1-hour frequency
 132 which is approximately consistent with 1-degree resolution E3SM output. Rescaling consists of
 133 computing either the median, mean or interpolated value depending on the original data frequency and
 134 variable properties. For most aerosol and cloud microphysics measurements, the median value is
 135 computed to remove occasional spikes or zeros resulting from data contamination or measurement error.
 136 For some bulk cloud properties (e.g., cloud fraction, liquid water path (LWP)), the mean value is
 137 computed to be consistent with grid-mean E3SM output. Interpolation is only used when the input
 138 frequency is equal to or coarser than the frequency of model output. For aircraft measurements, 1-minute
 139 resolution is used to retain high variability and allow matching samples of aerosol and cloud at the same
 140 time. To compare with high-frequency aircraft data, E3SM output is down-scaled to 1-minute resolution
 141 using the nearest grid cell and time slice. The rescale resolution can be adjusted in ESMAC Diags data
 142 preparation code for ESMs running at higher resolution (e.g., kilometer scale grid spacing). All processed
 143 data are saved in a standardized NetCDF format (NETCDF, 2022) and available for downloading (see
 144 data availability section) and direct use.

145 2.2 Data quality issues and treatments

146 Many observation datasets used in ESMAC Diags are ARM level-b (quality-controlled) or level-c (value-
 147 added) products, which include quality control (QC) flags to indicate data quality issues. For most
 148 datasets, a QC treatment is applied to remove all data with questionable flags. However, there are certain
 149 datasets or circumstances in which a QC flag is overly strict (too many good data are removed) or not
 150 strict enough (some bad data are not removed). Here we document some of these situations and how we
 151 handle them in our data processing.

152 2.2.1 ARM Condensation Particle Counter (CPC) measurements

153 ARM CPC data have several QC values representing failure of different quality checks. One of them
 154 checks if the concentration is greater than a maximum allowable value, which is set to $8,000 \text{ cm}^{-3}$ for



155 model 3010 (CPC, size detection limit 10 nm), 10,000 cm⁻³ for model 3772 (CPCF, size detection limit 10
156 nm), and 50,000 cm⁻³ for model 3776 (CPCU, size detection limit 3 nm). At SGP, new particle formation
157 (NPF) events occur frequently when CPC and CPCF measurements can exceed 30,000 cm⁻³. This is much
158 higher than the maximum allowable value but physically reasonable. Simply removing these large values
159 results in an underestimation of aerosol number concentration and produces unrealistic diurnal cycle since
160 they usually occur during the daytime (Tang et al., 2022a). By consulting with the ARM instrument
161 mentor, we only remove data with critical QC flags, but keep data with this QC flag that is overly
162 restrictive.

163 2.2.2 NCAR research flight aerosol number concentration (CN) measurements

164 NCAR research flight (RF) data used in ESMAC Diags do not include QC flags but occasionally show
165 suspiciously large or negative aerosol counts. The following minimum and maximum thresholds are
166 applied to remove suspicious data:

- 167 • Total CN from a Condensation Nucleation Counter (CNC, reported as CONCN): minimum = 0,
168 maximum = 25,000 cm⁻³.
- 169 • Total CN from an Ultra-High-Sensitivity Aerosol Spectrometer (UHSAS, reported as
170 UHSAS100): minimum = 0, maximum = 5,000 cm⁻³.
- 171 • Aerosol number size distribution from an UHSAS (reported as CUHSAS_RWOOU or
172 CUHSAS_LWII): minimum = 0, maximum = 500 cm⁻³ per size bin.

173 2.2.3 Ship-measured aerosol properties

174 Aerosol instruments on ships are occasionally contaminated by ship emissions, which present as large
175 spikes in aerosol and CCN number concentrations. For ARM MARCUS measurements, Humphries
176 (2020) published reprocessed CN and CCN data to remove ship exhaust contamination using method
177 described in Humphries et al. (2019). This data is used in this diagnostics package. For MAGIC, we could
178 not find any ship exhaust contamination information. By visually examining the dataset, a simple
179 maximum threshold (25,000 cm⁻³ for CPC, 5,000 cm⁻³ for UHSAS100, 2,000 cm⁻³ for CCN at 0.1%
180 supersaturation and 4,000 cm⁻³ for CCN at 0.5% supersaturation) is applied to remove likely
181 contamination from ship emissions.

182 2.2.4 CCN measurements

183 There are different supersaturation (SS) setting strategies for CCN measurements. Some aircraft
184 campaigns measured CCN with constant SS (ACE-ENA, HI-SCALE). Some other campaigns measured
185 CCN with time-varying (scanning) SS (SOCRATES, surface CCN counters at SGP and ENA). However,
186 the actual SS in a scanning strategy has fluctuations that are different than the target SS. For the latter,
187 CCN for each SS (0.1%, 0.2%, 0.3% and 0.5%) are obtained by selecting CCN measured within ± 0.05%
188 of the SS target.

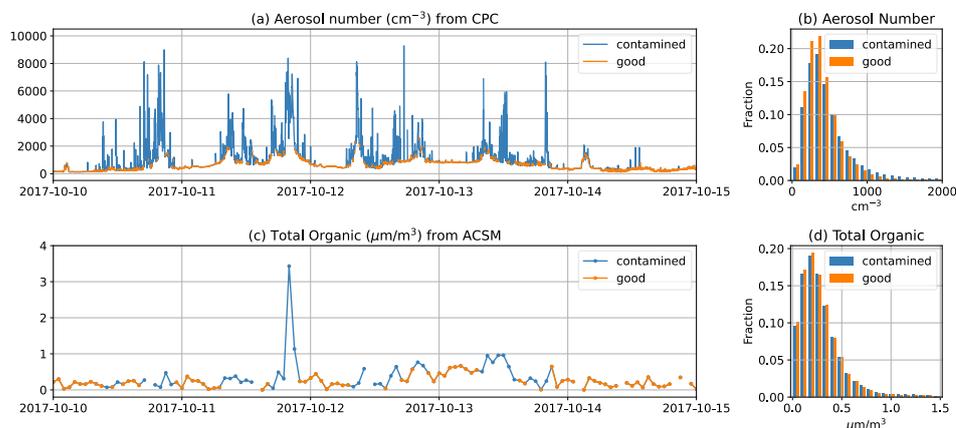
189 For long-term measurements at SGP and ENA, near-hourly CCN spectra data are available, and a
190 quadratic polynomial is fit to the spectra such that CCN number concentration can be estimated at any SS
191 between the measured minimum and maximum SS values. We calculate and output CCN number



192 concentration from these fits at three target supersaturations (0.1%, 0.2% and 0.5%). The fitted spectra
193 data provides CCN number concentration at the exact target supersaturations, but the sample number is
194 slightly smaller due to occasional failure of polynomial fitting.

195 2.2.5 Contaminated surface aerosol measurements at ENA

196 The ARM ENA site is located at a local airport. Aerosol measurements at ENA are sometimes
197 contaminated by aircraft and vehicle emissions, rendering the measurements not representative of the
198 background environment. Gallo et al. (2020) identified periods when CPC measurements were likely
199 contaminated from localized emissions (Figure 2a). Their aerosol mask data has 1-min resolution. When
200 we rescale the data to 1-hr resolution and apply the mask on other coarse time-resolution aerosol
201 measurements (e.g., ACSM, Figure 2c), we mask hours in which more than half of the hour is flagged by
202 the aerosol mask. The masking slightly increases the occurrence fraction of small values due to removing
203 many large values, but it does not change the overall distribution (Figure 2b and 2d). A sensitivity
204 analysis was performed, showing that 50% is a reasonable threshold to balance removal of contamination
205 with keeping reasonable data (not shown).



206
207 Figure 2: (a) CPC-measured CN from 10 to 15 October 2017 (1-minute resolution) with
208 local contamination flagged by Gallo et al. (2020). (b) histogram of CPC-measured CN for
209 all data from 2016-2018. (c) ACSM measured total organic matter from 10 to 15 October
210 2017 (1-hour resolution). Hours with more than half or the hour flagged in 1-minute CPC
211 data are masked as contaminated. (d) histogram of ACSM-measured total organic matter
212 for all data from 2016-2018.

213 3. Verification of cloud retrievals with in-situ measurements

214 Cloud microphysical properties such as droplet number concentration (N_d) and effective radius (R_{eff}) are
215 important variables that connect clouds to other aspects in the climate system such as aerosols and
216 radiation. Except in field campaigns where in-situ aircraft measurements are available, remote sensing
217 retrieval algorithms are usually needed to derive these quantities. Several cloud retrieval products from
218 ground and satellite measurements with different algorithms are used in ESMAC Diags v2. This section



219 compares these cloud retrievals with in-situ aircraft measurements to assess retrieval limitation,
 220 uncertainty, and utility. N_d and R_{eff} from aircraft measurements taken during HI-SCALE and ACE-ENA
 221 field campaigns are calculated from merged cloud droplet number size distributions (mergedSD) from
 222 three different cloud probes with different size ranges.

223 Table 3 lists R_{eff} and N_d retrieval products used in ESMAC Diags v2. We retrieved N_{d_sat} with input
 224 data from VISST products using the algorithms described in Bennartz (2007), but assuming a ratio of the
 225 drop volume mean radius to R_{eff} (commonly referred to as k) of 0.74 and a cloud adiabaticity of 80%
 226 (Varble et al., 2023). Other datasets are all available as released products. All retrievals assume a
 227 horizontally homogeneous single-layer liquid phase cloud with constant N_d throughout the cloud layer.
 228 However, retrieval algorithms are usually run for all conditions whenever they return valid values. When
 229 assumptions are not satisfied, retrieved properties may contain large errors and likely alter statistics such
 230 as increasing the occurrence frequency of small N_d as will be shown next.

231 Table 3: Cloud droplet effective radius R_{eff} and number concentration N_d retrievals

Variable	Dataset	Platform	Campaign/site	Retrieved from	Reference
R_{eff}	MFRSRCLDOD	Ground	HI-SCALE, ACE-ENA, SGP, ENA	SW diffuse flux, LWP	(Min and Harrison, 1996; Turner et al., 2021)
	VISST	Satellite	HI-SCALE, ACE-ENA, MAGIC, MARCUS, SGP, ENA	Brightness temperature	(Minnis et al., 2011)
	Wu_etal	Ground	ACE-ENA, MAGIC, ENA	Radar reflectivity, LWP	(Wu et al., 2020)
N_d	Ndrop	Ground	HI-SCALE, ACE-ENA, SGP, ENA	LWP, COD, cloud height	(Riihimaki et al., 2021; Lim et al., 2016)
	N_{d_sat} (calculated from VISST)	Satellite	HI-SCALE, ACE-ENA, MAGIC, MARCUS, SGP, ENA	LWP, COD, CTT	(Bennartz, 2007)
	Wu_etal	Ground	ACE-ENA, MAGIC, ENA	Radar reflectivity, LWP	(Wu et al., 2020)

232 MFRSRCLDOD: Cloud Optical Properties from the MultiFilter Shadowband Radiometer (MFRSR)

233 SW: shortwave

234 COD: cloud optical depth

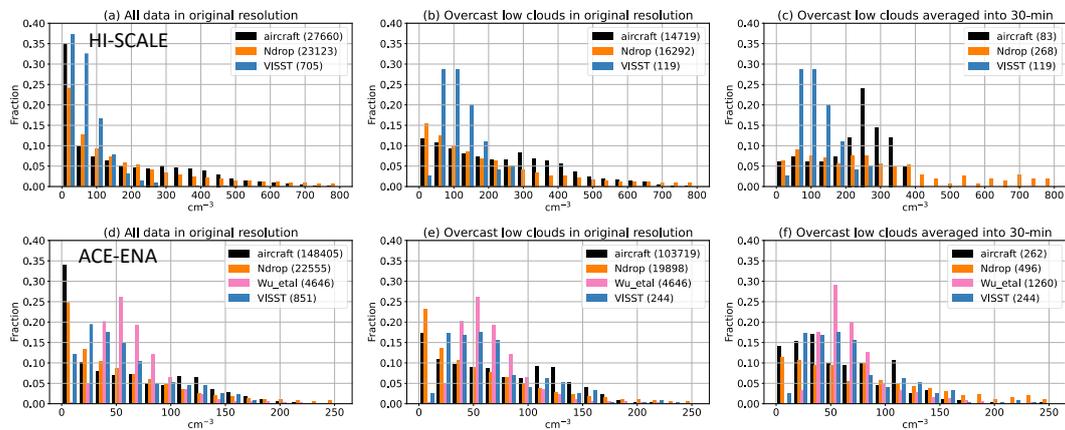
235 CTT: cloud top temperature

236

237 Figures 3 shows the probability density function (PDF) of N_d retrievals with aircraft measurements for
 238 HI-SCALE and ACE-ENA field campaigns, with the comparison of original temporal resolution versus
 239 30-minute mean, and the use of all available samples and samples that are filtered as overcast (cloud
 240 fraction > 90%) low-level (cloud top height < 4 km) clouds. Figure 4 shows similar plots but for R_{eff} .
 241 We also selected two cases with single-layer boundary layer stratus or stratocumulus clouds and plotted
 242 their timeseries of original-resolution and 30-min averaged R_{eff} and N_d in Figure S1. The high-frequency
 243 aircraft measurements and MFRSR/Ndrop retrievals exhibit much larger variability than coarse-frequency
 244 retrievals of Wu_etal and VISST. They frequently sample cloud edges or cloud top/base (for aircraft),

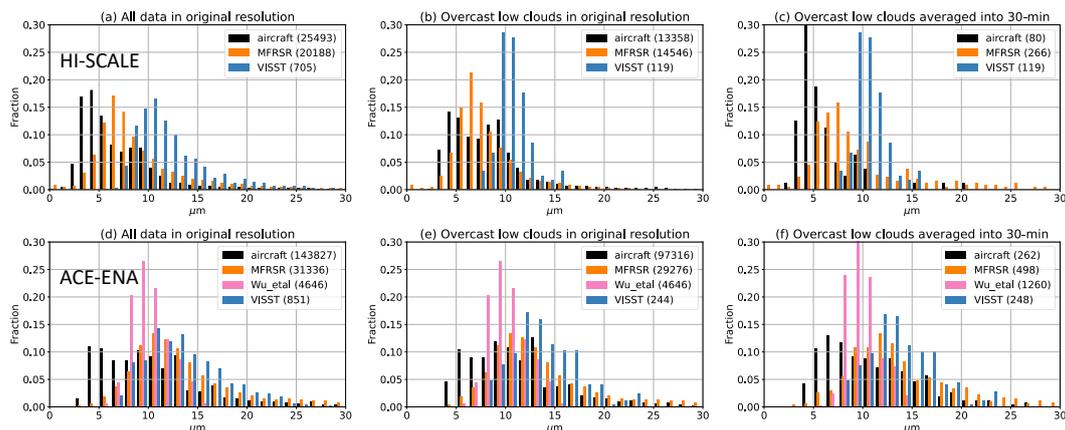


245 where N_d is typically less than further into the cloud. This causes large occurrence fractions in the lowest
 246 few bins in the N_d PDFs (Figure 3a and 3d). The 30-min VISST products also show large occurrence
 247 fraction in the lowest N_d bin for HI-SCALE (Figure 3a), likely due to high frequency of partial cloudy
 248 condition over continental U.S. Filtering conditions to only include overcast low-level clouds (Figure 3b,
 249 e) and averaging into a coarser resolution (Figure 3c, f) both contribute to the reduction of occurrence
 250 fraction in small- N_d bins, and make the measurements from different instruments more comparable.



251

252 Figure 3: Histogram of N_d from different measurements/retrievals in (top) HI-SCALE and
 253 (bottom) ACE-ENA field campaigns, with total sample numbers in the parentheses. (a) and
 254 (d) use data samples in their original resolution (1 s for aircraft measurements, 20 s for
 255 Ndrip data, 5 min for Wu_etal data, and 30 min for VISST data). (b) and (e) include only
 256 overcast low-cloud situations. For aircraft data, this means N_d is $> 1 \text{ cm}^{-3}$ for 5 s before
 257 and after the sampling time; for Ndrip and VISST data, it means cloud fraction $> 90\%$ and
 258 cloud top height $< 4\text{km}$. (c) and (f) include only overcast low-cloud situations, and
 259 average into 30-min resolution. For all the plots, VISST data with solar zenith angle $> 65^\circ$
 260 are removed to avoid artifact from sunlight.



261

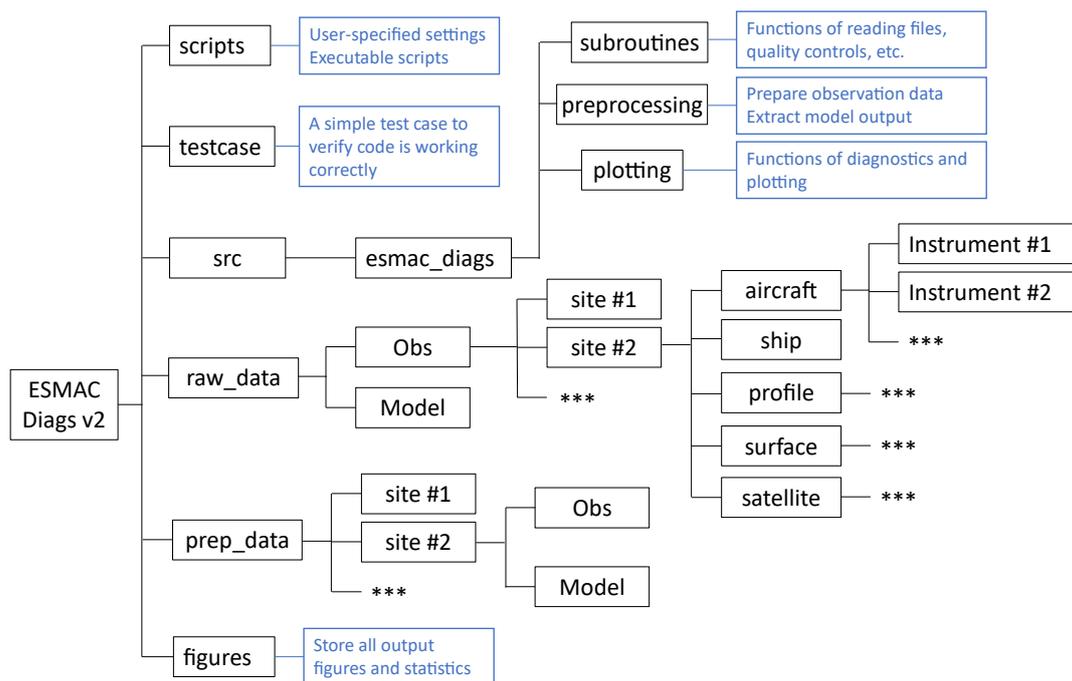


262 Figure 4: similar as in Figure 3 but for R_{eff} .

263 Overall, the remote sensing retrievals and aircraft measurements produce reasonable ranges of N_d and
264 R_{eff} . Marine clouds (ACE-ENA) have smaller N_d (Figure 3) and larger R_{eff} (Figure 4) than continental
265 clouds (HI-SCALE). Different retrievals are more consistent with each other for marine clouds than
266 continental clouds. Different N_d datasets generally agree in mean value, but aircraft and Ndrop data
267 exhibit broader distributions, likely due to their high sampling frequency that may capture more extreme
268 conditions with very high or low N_d . Moreover, the assumption of a fixed adiabaticity (0.8) in satellite
269 retrieval will also narrow N_d distribution. For R_{eff} , we do not expect different datasets to be perfectly
270 agree with each other, as cloud droplet size grows with height in the cloud. All remote sensing retrievals
271 have larger R_{eff} values than aircraft measurements, potentially because remote sensors weight more
272 towards the upper cloud where droplet size and liquid water content (LWC) are larger. Wu_etal retrieves
273 vertical profile of R_{eff} , and a median value of the R_{eff} profile is used to represent the entire cloud. This
274 makes Wu_etal retrieval weight less toward large droplets thus its R_{eff} is less than MFRSR and VISST.
275 VISST data have the largest R_{eff} values, likely because satellite retrievals reflect conditions at the cloud
276 top. Given the spread in retrieved cloud properties, the limitations and uncertainties of cloud microphysics
277 retrievals clearly need to be considered when they are used to evaluate model performances.

278 4. Structure of diagnostics package

279 Figure 5 shows the directory structure of ESMAC Diags v2. It is substantially changed from ESMAC
280 Diags v1 (Tang et al., 2022a). First, we save all data separately as *raw_data*, which stores all input
281 datasets collected from field campaigns, and *prep_data*, which stores preprocessed data with standardized
282 time resolution and quality controls as described in Section 2. The structure is still designed to be flexible
283 for future extension with additional measurements and/or functionality. Second, the diagnostics functions
284 now give users more freedom to modify analyses, such as selecting different time periods, performing
285 additional data filtering or treatments, and examining ACI relationships in specified variable
286 combinations (for scatter plots, joint histograms or heatmaps). We provide a set of example scripts to
287 assist users design their own diagnostics based on their needs.



288

289 Figure 5: Directory structure of ESMAC Diags v2. Blue boxes describe the functions of the
 290 directory. Asterisks represent boxes that follow the same format as those shown in
 291 parallel.

292 ESMAC Diags v1 included diagnostics of aerosol mean statistics (mean, bias, RMSE, correlation),
 293 timeseries, diurnal cycle, vertical profiles, mean particle number size distribution, percentiles by
 294 height/latitude, and pie/bar charts (Tang et al., 2022a). ESMAC Diags v2 now includes the following new
 295 diagnostics that include cloud variables:

- 296 - 5th, 25th, 50th, 75th and 95th percentiles,
- 297 - Seasonal cycle at SGP and ENA,
- 298 - Histograms for individual variables,
- 299 - Scatter plots,
- 300 - Joint histograms of two variables, and
- 301 - Heatmaps of three variables (mean of one variable binned by two other variables).

302 The inclusion of two-variable scatter plots, joint histograms, and three-variable heatmaps provides the
 303 functionality to study ACI-related relationships. We present a few examples in the next section to
 304 demonstrate these new diagnostics.

305

306 5. Diagnostics Examples

307 In this section, we show some examples of diagnostics applied to E3SM version 2 (E3SMv2) (Golaz et
 308 al., 2022). Compared to the aerosol and cloud parameterizations in E3SMv1 (Rasch et al., 2019; Golaz et



309 al., 2019), E3SMv2 updated the treatments on dust particles, incorporated recalibration of parameters (Ma
310 et al., 2022), changed the call order and refactored the code of the Cloud Layers Unified By Binormals
311 (CLUBB) parameterization, and retuned some parameters (Golaz et al., 2022). We constrain the model
312 simulations by nudging the horizontal winds towards the 3-hourly Modern-Era Retrospective analysis for
313 Research and Applications, Version 2 (MERRA-2, Gelaro et al., 2017) with a nudging time scale of 6
314 hour. Previous studies have shown that with nudging, E3SM can well simulate the large-scale circulations
315 in reanalyses (Sun et al., 2019; Zhang et al., 2022). The model was run for individual field campaigns
316 (Table 1) and from 2010 to 2020 for long-term diagnostics at SGP and ENA sites, with hourly model
317 output saved over the field campaign regions for detail evaluation. As described in Section 2, all
318 diagnostics for ground and ship campaigns are in 1-hour resolution while diagnostics for aircraft
319 campaigns are in 1-minute resolution. For aerosol and cloud variables, model raw output variables (not
320 from instrument simulators) are used in this paper to reveal the intrinsic ACI relationships in E3SM.
321 However, as can be seen later in this section, instrument simulators can be better used in some diagnostics
322 to ensure more consistent comparison. Users may choose whether or not to use simulators in their
323 diagnostics depending on their purpose.

324 **5.1. Single-variable diagnostics**

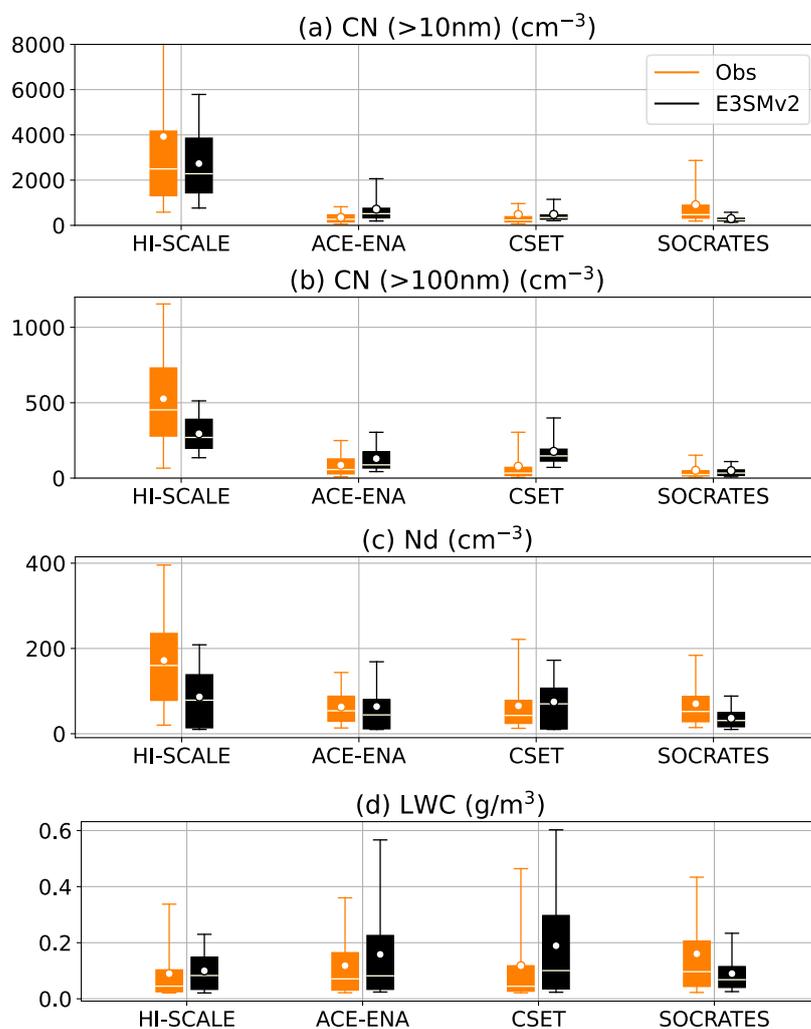
325 Figures 6 and 7 show mean and percentile values of aerosol and cloud properties measured from field
326 campaigns in the four geographical regions: CUS, ENA, NEP and SO. Figure 6 is for aircraft platforms
327 and Figure 7 is for ground or ship platforms with satellite data included when available. Note that the
328 aircraft and ground/ship campaigns may cover different time periods (Table 1), thus some differences
329 seen between aircraft and ship measurements may be caused by seasonal variation. As cloud
330 microphysical properties are usually retrieved with assumptions (Section 3), for ground/ship/satellite data,
331 we only focus on overcast low-level liquid cloud condition here (cloud fraction > 90%, cloud top height <
332 4 km and ice water path < 0.01 mm). E3SM does not output cloud top height, which is derived using a
333 weighting integration method as described in Varble et al. (2023).

334 From both aircraft and ground/ship data, HI-SCALE has much larger aerosol and cloud droplet number
335 concentrations with smaller droplet sizes compared to other campaigns, which is expected for a
336 continental environment compared to a marine environment. The cloud optical depth is also greater for
337 HI-SCALE than other campaigns, which is driven by smaller droplet sizes rather than LWP differences.
338 Satellite retrievals generally produce smaller N_d , LWP, and cloud optical depth with greater R_{eff} than
339 surface retrievals. As discussed in Section 3, retrieval uncertainties need to be kept in mind when these
340 retrieved microphysical properties are used to evaluate models.

341 E3SMv2 overestimates CN (> 10 nm) over CUS, ENA and NEP. Larger particle concentration (CN > 100
342 nm) is generally underestimated over CUS and overestimated over ENA and NEP. Over SO, E3SMv2
343 produces fewer small aerosol particles (CN > 10 nm) and about the same number of large aerosol
344 particles (CN > 100 nm) compared to the observations. These results are confirmed by both aircraft and
345 ground/ship campaigns, except for the HI-SCALE aircraft campaign where small particles from local
346 emissions were occasionally observed but unable to be simulated. These results are consistent with our
347 previous diagnostics for E3SMv1 (Tang et al., 2022a). E3SMv2 also underestimates N_d over CUS and
348 SO, which corresponds with the underestimation of accumulation mode (> 100 nm) CN over CUS but
349 underestimation of Aitken mode (> 10 nm) CN over SO. It is possible that over very clean regions such as

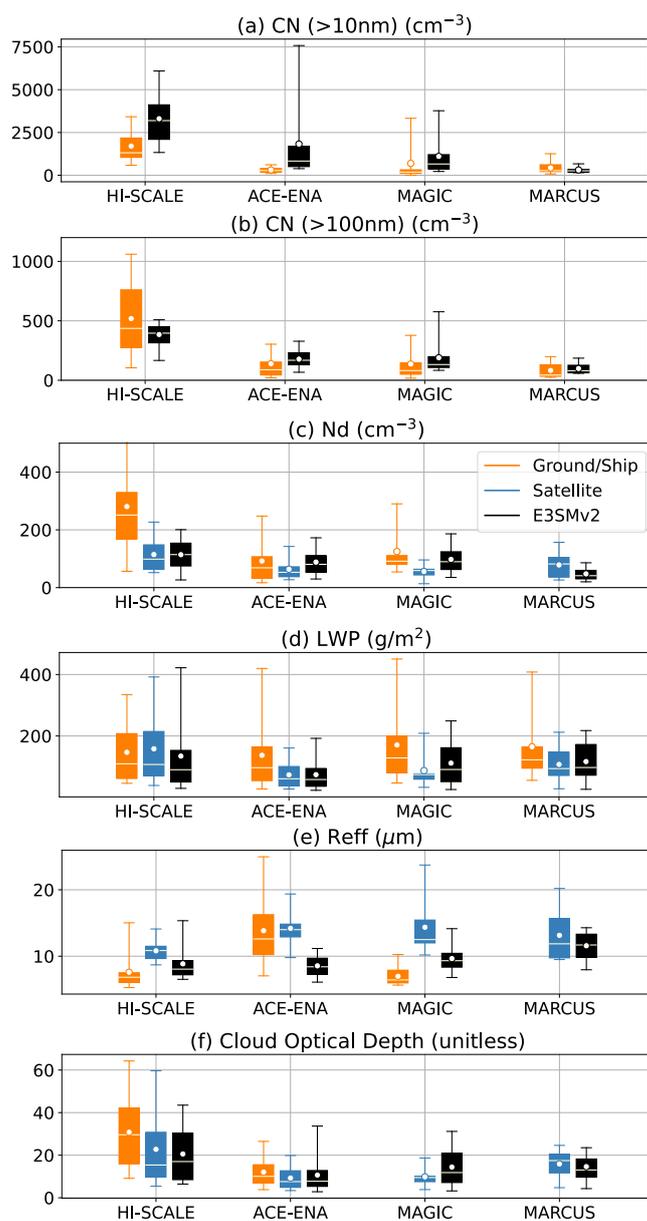


350 SO, small particles are more important in cloud formation than over continental regions such as CUS.
351 Simulated LWP (LWC) is generally consistent with satellite (aircraft) measurements, but smaller than
352 ground/ship measurements, which may be partly caused by rain contamination of ground/ship retrievals.
353 R_{eff} evaluation is less certain given large discrepancies between satellite and ground retrievals.



354

355 Figure 6: Box-whisker plots of (a) CN for size > 10 nm, (b) CN for size > 100 nm, (c) in-
356 cloud N_d , (d) LWC for all data from aircraft field campaigns at CUS, ENA, NEP and SO
357 regions from left to right. Boxes denote 25th and 75th percentiles, whiskers denote 5th and
358 95th percentiles, the white horizontal line represents median values, and the white dot
359 represents mean values.

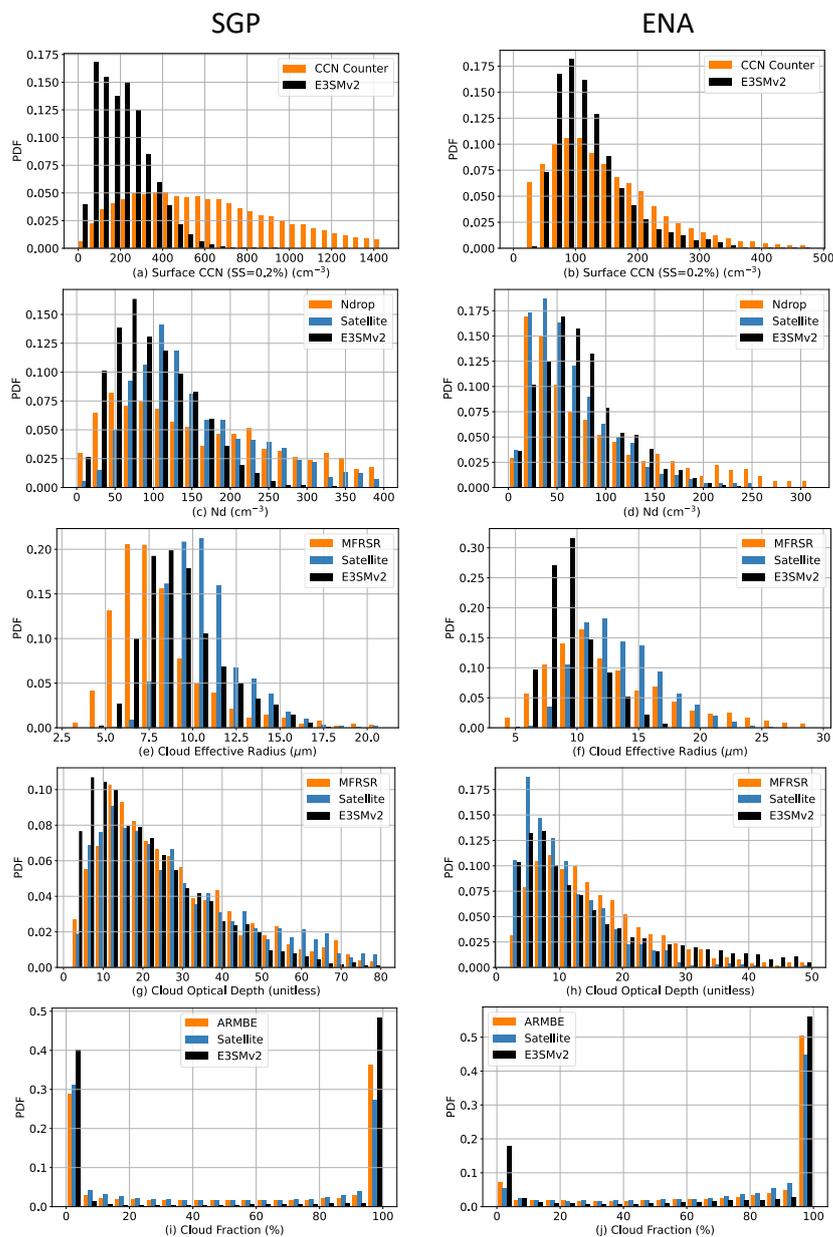


360

361 Figure 7: Box-whisker plots of (a) CN for size > 10 nm, (b) CN for size > 100 nm, (c) layer-
 362 mean N_d , (d) LWP, (e) R_{eff} , (f) cloud optical depth for overcast low-level liquid cloud
 363 conditions (cloud top height < 4 km, cloud fraction > 90% and ice water path < 0.01 mm)
 364 in ground and ship field campaigns at CUS, ENA, NEP and SO regions from left to right.
 365 Boxes denote 25th and 75th percentiles, whiskers denote 5th and 95th percentiles, the



366 white horizontal line represents median values, and the white dot represents mean
 367 values.



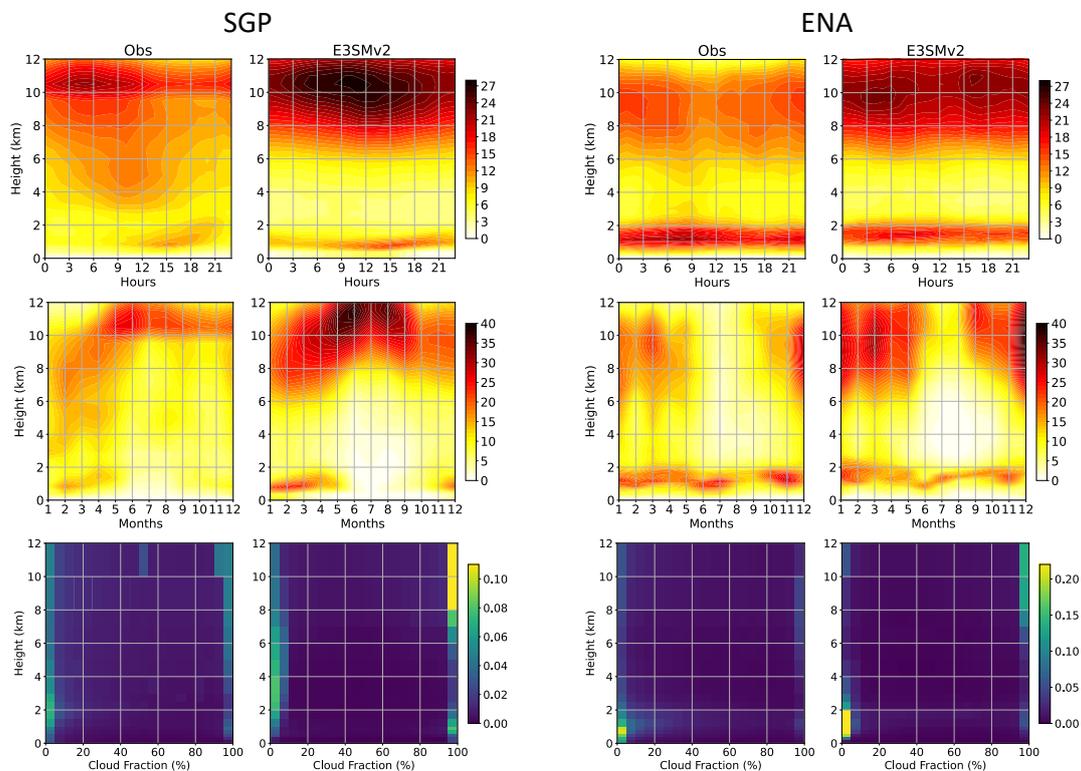
368

369 Figure 8: histogram of (from top to bottom) surface CCN number concentration, layer-
 370 mean N_d , R_{eff} , cloud optical depth and total cloud fraction at (left) SGP from 2011 to
 371 2020 and (right) ENA from 2016 to 2018. Surface CCN and total cloud fraction are using



372 all-condition samples while N_d , R_{eff} , cloud optical depth data are filtered for overcast
 373 low-level liquid clouds (cloud top height < 4 km, cloud fraction > 90%, ice water path <
 374 0.01 mm).

375 Figure 8 shows PDFs of surface CCN number concentration in 0.2% supersaturation, cloud layer mean
 376 N_d , R_{eff} , cloud optical depth and total cloud fraction for long-term diagnostics at SGP (year 2011–2020)
 377 and ENA (year 2016–2018) sites. E3SMv2 fails to reproduce the long tail of large values in CCN and N_d ,
 378 especially over SGP. This is consistent with the underestimation of CN (> 100 nm) during the HI-SCALE
 379 field campaign shown in Figures 6 and 7. Compared with ground retrievals, E3SMv2 R_{eff} is larger at
 380 SGP but smaller at ENA. However, satellite-retrieved R_{eff} has larger values than E3SMv2 at SGP. As
 381 discussed before, discrepancies between satellite and ground retrievals can be substantial for some
 382 locations and variables, and considering both in evaluating model performance gives a sense for how
 383 uncertain comparisons are. E3SMv2 generally captures the PDFs of cloud optical depth and total cloud
 384 fraction, although it underestimates the frequency of partial-cloudy conditions and overestimates the
 385 frequency of clear-sky and overcast conditions.



386

387 Figure 9: (top) Diurnal cycle, (middle) seasonal cycle, and (bottom) occurrence frequency
 388 of vertical cloud fraction at (left) SGP from 2011 to 2020 and (right) ENA from 2016 to
 389 2018.



390 Figure 9 shows the long-term diagnostics of mean diurnal cycles, seasonal cycles and PDFs of cloud
391 fraction by height at SGP and ENA sites. At SGP, observations show formation of low clouds in the
392 afternoon and in late winter through springtime. High clouds peak overnight into the early morning and in
393 the spring to summer, corresponding to nocturnal deep convective systems common over SGP (Tang et
394 al., 2022b; Tang et al., 2021; Jiang et al., 2006). These features are reasonably well represented in
395 E3SMv2, although low-level cloud deepening in the afternoon is not well predicted, and high-level clouds
396 peak in the late rather than early morning. At ENA, marine stratus or stratocumulus clouds occur in any
397 month and at any time of the day, but with less frequency in late summer and in afternoon. High clouds
398 are more frequent in winter months than in summer months and occur throughout the diurnal cycle with a
399 slight mid-day minimum. These features are well captured by E3SMv2. At both sites, high clouds usually
400 occur with high fraction (> 95%) while low clouds are more likely associated with small fraction (< 5%)
401 (bottom row). At SGP, high occurrence of low cloud fraction extends vertically up to the tropopause,
402 representing frequently occurring deep convection. At ENA, low clouds have less vertical extension but
403 are more likely to expand to greater fraction. E3SMv2 reproduces these cloud features in occurrence
404 frequency, with overestimation of occurrence frequency in high (>95%) and low (<5%) cloud fraction
405 consistent with Figure 8.

406 Overall, the mean fraction of high clouds is overestimated in E3SMv2. This overestimation has been
407 reported in many previous studies in the Community Earth System Model (CESM)-E3SM model family
408 (e.g., Song et al., 2012; Cheng and Xu, 2013; Xu and Cheng, 2013a, b; Tang et al., 2016; Zhang et al.,
409 2020). However, this is not an apple-to-apple comparison, as cloud fraction in ESMs includes clouds that
410 are optically very thin that cannot be detected by satellite passive sensors or cloud radar. When satellite
411 simulators are used, slight underestimation of high cloud fraction by E3SM is seen over most tropical
412 deep convection regions (Zhang et al., 2019; Xie et al., 2018; Rasch et al., 2019). Unfortunately, our
413 model does not output cloud vertical profiles from satellite simulators, which prevents a direct apple-to-
414 apple comparison. Thus, caution should be taken when direct model output is used to compare with
415 observed cloud fraction.

416 5.2. Multi-variable relationships related to ACI

417 The effective radiative forcing due to ACI processes are complex, nonlinear, and highly uncertain despite
418 their significant impact on climate. ACI studies are usually conducted by examining relationships
419 between aerosols, clouds, and radiation variables that are known to interact with one another. Given so
420 many variable combinations related to ACI, ESMAC Diags v2 provides a framework for users to examine
421 relationships between the variables they choose with joint histograms, scatter plots and heatmaps. Here
422 we show a few examples to assess relationships between CCN, N_d , LWP, and top of atmosphere (TOA)
423 albedo.

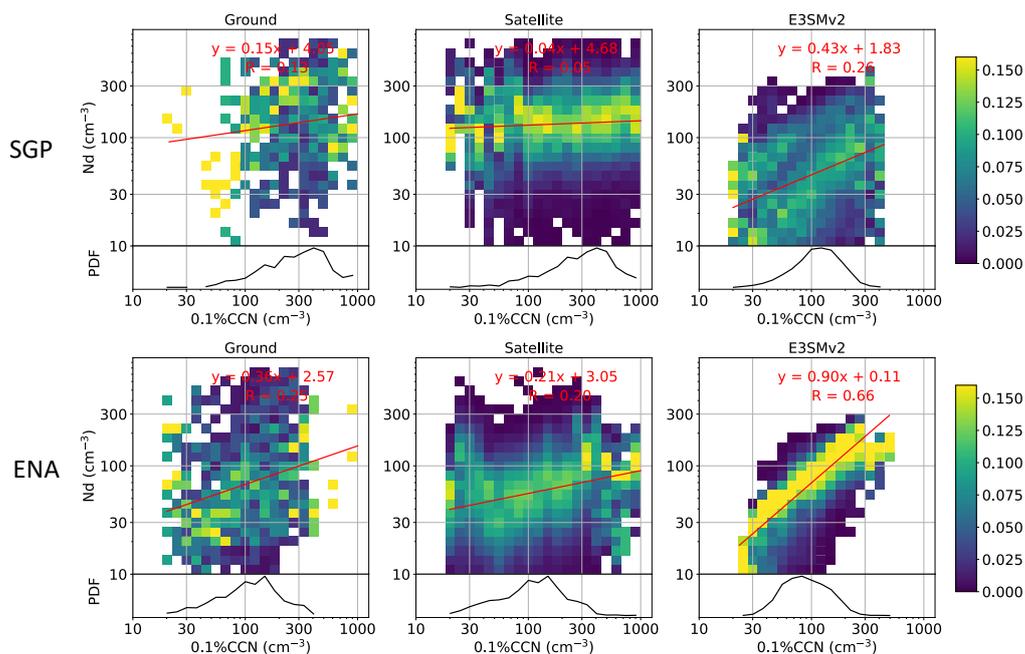
424 The dependence of TOA albedo on CCN number concentration for stratiform warm clouds can be
425 decomposed (e.g., following Quaas et al. (2008)) as:

$$426 \quad \frac{dA}{d\ln CCN} = \left(\frac{\partial A}{\partial \ln N_d} + \frac{\partial A}{\partial \ln LWP} \frac{d\ln LWP}{d\ln N_d} \right) \frac{d\ln N_d}{d\ln CCN} \quad (1)$$



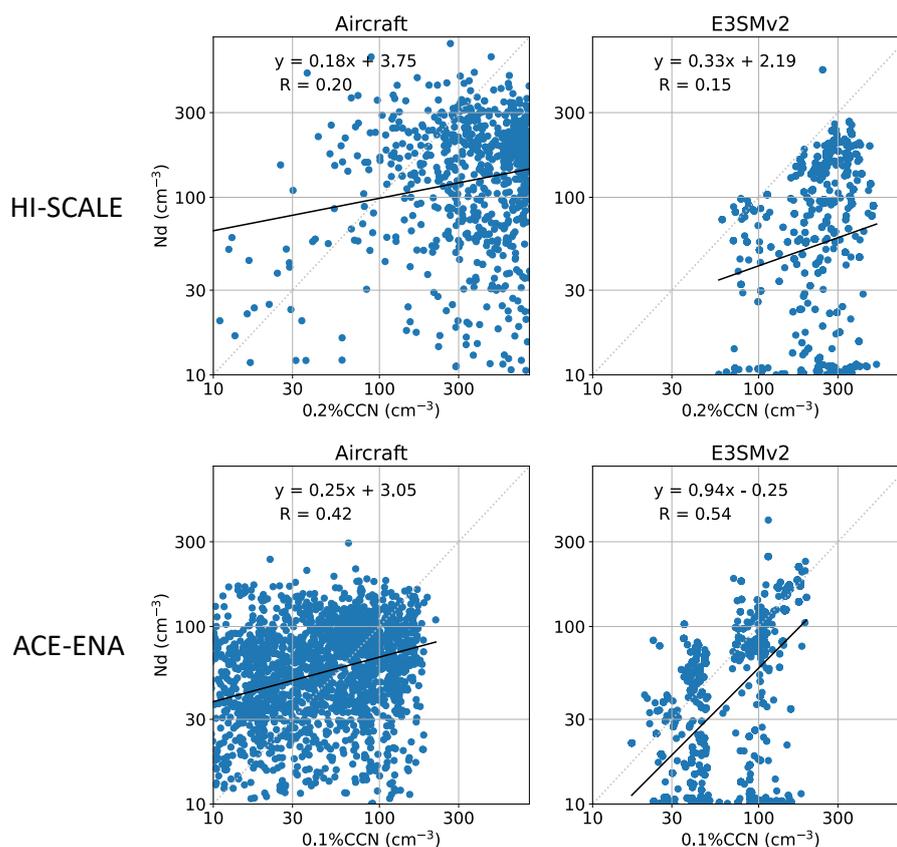
427 which allows isolation of “Twomey effect” $\left(\frac{\partial A}{\partial \ln N_d}\right) \left(\frac{d \ln N_d}{d \ln CCN}\right)$ and “LWP adjustment” $\left(\frac{d \ln LWP}{d \ln N_d}\right)$ associated
428 with specific ACI processes. Here we use joint histograms and heatmaps to evaluate each component,
429 $\frac{d \ln N_d}{d \ln CCN}$, $\frac{d \ln LWP}{d \ln N_d}$, $\frac{\partial A}{\partial \ln N_d}$ and $\frac{\partial A}{\partial \ln LWP}$ based on long-term ground and satellite measurements at SGP (2011-
430 2020) and ENA (2016-2018) sites. The analysis in this section (except Figure 11) is limited to overcast
431 (cloud fraction > 90%), low-level (cloud top height < 4 km) liquid (ice water path < 0.01 mm) clouds.
432 Since there is no direct measurement of cloud base CCN concentration from remote sensors, surface CCN
433 concentration is used in this study and only clouds that are most likely to be affected by surface
434 conditions are examined. These clouds are identified as having cloud base potential temperature minus
435 surface potential temperature smaller than 2 K. For satellite measurements, samples with solar zenith
436 angle greater than 65° are removed to avoid N_d retrieval biases (Grosvenor et al., 2018). The sample
437 number of (ground, satellite, E3SM) for overcast low-level liquid clouds are (1766, 1217, 6369) at SGP
438 and (3450, 1345, 2884) at ENA, respectively. To increase sample size for more robust statistics, satellite
439 retrievals and E3SM outputs over a 5°×5° domain centered on SGP and ENA sites are included. This
440 increases the sample number to (1766, 71942, 15231) at SGP and (3450, 104260, 28184) at ENA.
441 Analyses of all-sky conditions and overcast low-level liquid clouds for a single grid point over each site
442 are shown in Figures S2-S7 in the supplementary material. Increasing sample domain for satellite and
443 E3SM data does not change the over statistics shown here.

444 The change of N_d in response to a change of surface CCN number concentration $\left(\frac{d \ln N_d}{d \ln CCN}\right)$ is heavily
445 influenced by processes such as aerosol activation. Figure 10 shows the joint PDFs of N_d and surface
446 CCN number concentration at 0.1% supersaturation normalized within each CCN bin. Ground and
447 satellite observations show similar linear fit of $\ln N_d - \ln CCN$ relation, although ground-based plots have
448 much smaller sample number. E3SMv2 shows more sensitive $N_d - CCN$ relationships than observations
449 at both SGP and ENA sites, with the relationship tighter at ENA and more scattered at SGP. As a cross
450 validation, Figure 11 shows the $N_d - CCN$ relationships from short-term aircraft campaign during HI-
451 SCALE and ACE-ENA. The comparison with in-situ aircraft measurements confirms that E3SMv2 has
452 more sensitive N_d to CCN than observations. These results indicate that aerosol activation in E3SMv2
453 may be too weak in low CCN conditions and too strong in high CCN conditions, which may be related to
454 the differences in simulated and observed updraft velocity and supersaturation (Varble et al., 2023). Note
455 that E3SMv2 produces a significant number of small N_d (< 20 cm⁻³) samples (Figure 11). This feature is
456 reported in Golaz et al. (2022) and is partially removed by setting a minimum threshold of $N_d = 10$ cm⁻³.
457 However, as seen in Figure 11, there are still a large number of N_d between 10 and 20 cm⁻³. Further
458 investigation is underway to diagnose the causes of the abundant low- N_d values. The diagnostics shown
459 here indicate that a more physical method should be applied to improve the simulated N_d .



460

461 Figure 10: Joint histogram of layer-mean N_d versus surface CCN number concentration at
 462 0.1% supersaturation, normalized within each CCN number concentration bin (PDF of CCN
 463 shown in the bottom of each panel). Samples are constrained to likely surface-coupled,
 464 overcast low-level liquid clouds (cloud top height < 4 km, cloud fraction > 90%, ice water
 465 path < 0.01 mm and potential temperature difference between cloud base and surface < 2
 466 K). Available samples within a $5^\circ \times 5^\circ$ region centered on SGP (top) and ENA (bottom) for
 467 satellite and E3SMv2 datasets are included. Linear fits and R values are shown in red.



468

469 Figure 11: Scatter plots for N_d versus CCN along the flight tracks from (top) HI-SCALE and
 470 (bottom) ACE-ENA campaigns. Note that CCN number concentration measurements are
 471 taken under $\sim 0.2\%$ supersaturation for HI-SCALE and under $\sim 0.1\%$ supersaturation for
 472 ACE-ENA. Linear fits and R values are shown in each panel. $R = 0.34$ (SGP) and 0.74 (ENA)
 473 for E3SMv2 if a minimum $N_d = 20 \text{ cm}^{-3}$ is applied.

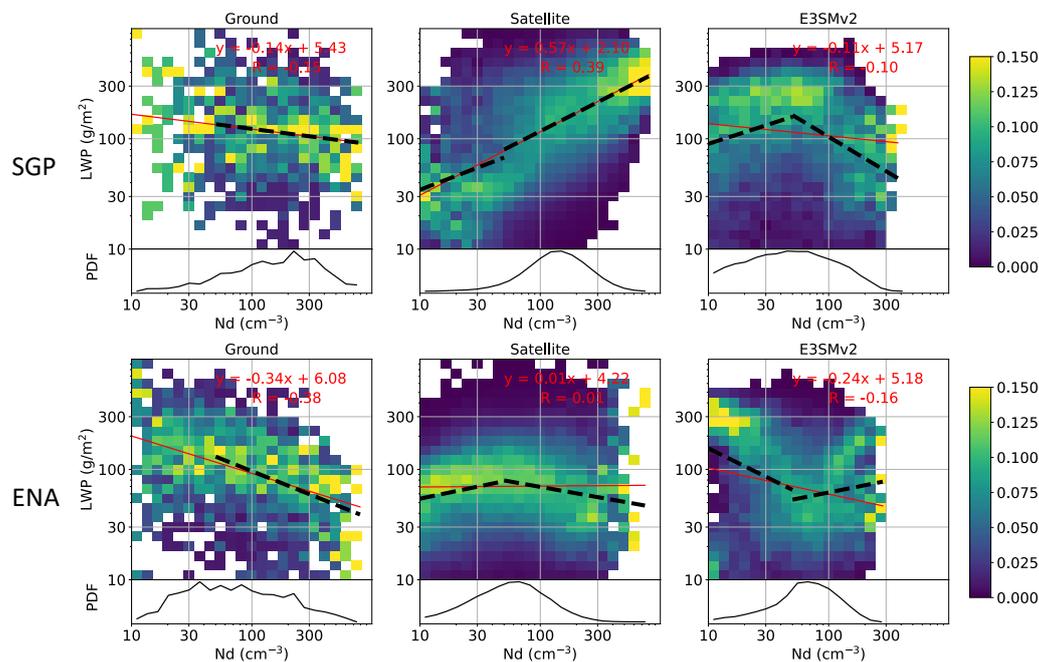
474

475 The term $\frac{d \ln LWP}{d \ln N_d}$ is commonly interpreted as the adjustment of LWP to a perturbation in N_d tied to
 476 suppression of precipitation (increase LWP) or enhancement of evaporation (decrease LWP) (e.g.,
 477 Glassmeier et al., 2019). Gryspeerd et al. (2019) show that the satellite retrieved LWP over ocean
 478 increases with N_d when $N_d < \sim 30 \text{ cm}^{-3}$ and decreases when $N_d > \sim 30 \text{ cm}^{-3}$. This relation is also seen
 479 in satellite retrievals at ENA (Figure 12) when using a higher threshold $N_d = 50 \text{ cm}^{-3}$ to perform linear
 480 fits (black dashed lines). The linear fit is insignificant for $N_d < 50 \text{ cm}^{-3}$ in surface retrievals at both
 481 sites, partly due to small sample number, and also potentially related to drizzle contamination of LWP.
 482 The slope of the LWP – N_d relation in satellite retrievals at SGP is positive for both N_d ranges. This is
 483 opposed to slope shown in the ground retrievals and indicates that retrieval biases may cause opposite

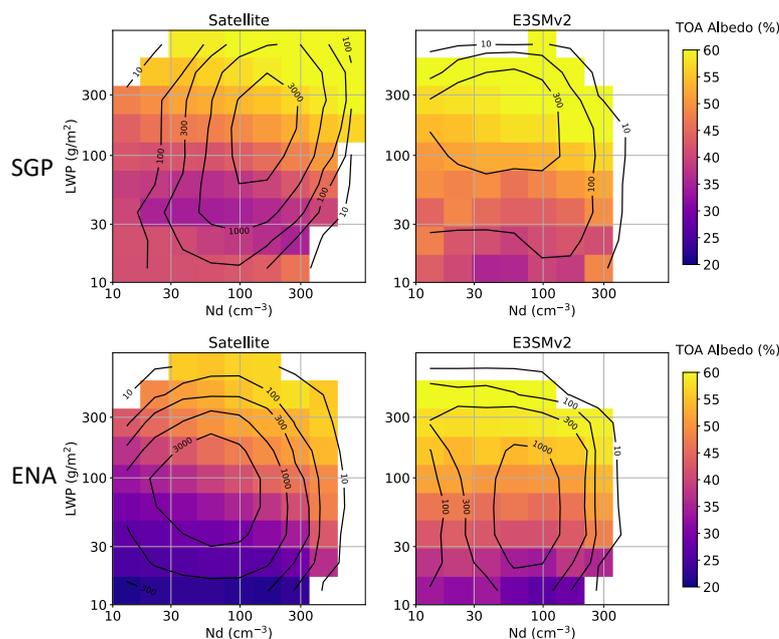


484 results in ACI studies. The reason why satellite retrievals show positive LWP – N_d relation at SGP is
 485 subject to further investigation.

486 The E3SMv2 simulated LWP – N_d relation is quite different from satellite retrievals at both sites. At
 487 SGP, it generates a positive slope for $N_d < 50 \text{ cm}^{-3}$, and a negative slope for $N_d > 50 \text{ cm}^{-3}$. At ENA, it
 488 shows an opposite relation, with LWP decreases for small N_d and increases for large N_d . We examined a
 489 few other oceanic regions with frequent stratus or stratocumulus clouds in E3SMv2 and saw similar
 490 behavior (not shown). However, LWP – N_d relation in E3SMv1 performs quite differently, as shown in
 491 Varble et al. (2023). The causes of the different LWP – N_d relation behaviors in E3SM are under further
 492 investigation. Varble et al. (2023) discussed potential physical mechanisms that may affect the different
 493 LWP adjustments in observation and simulation, such as different atmospheric states in E3SM and
 494 observations. Our user-friendly diagnostics package allows these analyses to be routinely performed for
 495 the purpose of better understanding critical model behaviors at process- and mechanistic-levels, providing
 496 observational constraints to facilitate model development efforts.



497
 498 Figure 12: Following Figure 10, but for the N_d bin-normalized joint histogram of LWP
 499 versus N_d . Red lines and equations are linear fits for all data samples and black dashed
 500 lines are linear fits for $N_d < 50 \text{ cm}^{-3}$ and $N_d > 50 \text{ cm}^{-3}$ when the fits are statistically
 501 significant ($p < 0.01$).



502

503 Figure 13: Heatmaps of mean TOA albedo versus LWP and N_d for likely surface-coupled,
504 overcast low-level liquid clouds (cloud top height < 4 km, cloud fraction > 90%, ice water
505 path < 0.01 mm and potential temperature difference between cloud base and surface < 2
506 K). Data include samples within a $5^\circ \times 5^\circ$ region centered on SGP (top) and ENA (bottom).
507 Valid sample number is shown in black contour lines. Grids with valid sample number <
508 10 are not filled. Ground data is not included, since the TOA albedo is not available.

509 Figure 13 shows heatmaps of mean TOA albedo with respect to LWP and N_d from which $\frac{\partial A}{\partial \ln N_d}$ and
510 $\frac{\partial A}{\partial \ln LWP}$ can be derived. At both ENA and SGP, TOA albedo generally increases with increases of LWP
511 and N_d , except at SGP when LWP is small. The increasing albedo in small LWP may be due to retrieval
512 artifact as uncertainty becomes large when LWP is small (e.g., < 20 g/m²), solar zenith angle is large
513 (e.g., > 55°), or cloud optical depth is small (e.g., < 5) (Grosvenor et al., 2018). TOA albedo at SGP is
514 generally higher than at ENA, which is expected for clouds with smaller droplet sizes. Increasing TOA
515 albedo with increases of LWP is also seen in E3SMv2, but the dependence with N_d is weak. This can be
516 impacted by correlation between solar zenith angle and N_d in E3SM simulation, as discussed in Varble et
517 al. (2023). For a given LWP and N_d , TOA albedo is generally higher in E3SMv2 than in satellite
518 observations, indicating that shallow clouds may be too reflective in the model, possibly due to smaller
519 cloud R_{eff} (Figure 8).

520 The above illustration of single-variable and multi-variable diagnostics present examples to demonstrate
521 the capability of ESMAC Diags v2. More analyses, such as selecting other variables, performing
522 additional data filtering or treatments, and examining ACI relationships with other variable combinations,
523 can be conducted through user-specified settings. A detailed user guide and a collection of example



524 scripts are included in the diagnostics package to assist users design customized diagnostics suited to their
525 specific needs.

526 5. Summary

527 We developed an Earth System Model aerosol-cloud diagnostics package (ESMAC Diags) to facilitate
528 routine evaluation of aerosols, clouds and ACI in the U.S. DOE's E3SM model using multiple platforms
529 of observations. As an updated version of ESMAC Diags v1 (Tang et al., 2022a) which mainly focuses on
530 aerosol properties, this paper described ESMAC Diags v2 that focuses on both aerosols, clouds, as well as
531 their interactions. In addition to the short-term field campaigns included in ESMAC Diags v1, long-term
532 diagnostics from two permanent ARM sites (SGP and ENA, each represents continental and maritime
533 conditions, respectively) are now conducted to provide more robust evaluation. The newly added multi-
534 variable joint histograms, scatter plots and heatmaps allow users to examine correlations between
535 variables that are relevant to the study of ACI.

536 Ground- and ship-based aerosol measurements are frequently impacted by local-scale emissions sources
537 such as those from airport or ship exhaust. These local sources are not resolved by coarse-resolution
538 ESMs, which usually represent an environment averaged within a region of tens to hundreds of kilometers
539 in size. In ESMAC Diags, we used available contamination-removed aerosol data, such as those from
540 Gallo et al. (2020) for ENA, and Humphries (2020) for MARCUS, and applied data filtering for other
541 field campaigns. The observations are harmonized into a uniform data format and temporal resolution that
542 are comparable with ESMs. Aircraft measurements retain higher resolution (currently 1-min) to preserve
543 high spatiotemporal variability, although ESMs have to be downscaled for evaluation with aircraft
544 measurements. This limitation of scale mismatch must be accepted to perform evaluation in current
545 coarse-resolution ESMs. Nevertheless, as ESM grid spacing approaches a few kilometers via regional
546 refinement (Tang et al., 2019) or global convection-permitting configuration (Caldwell et al., 2021), the
547 scale inconsistency between models and observations is reduced. ESMAC Diags can easily adjust the
548 preprocessing output resolution to facilitate the evaluation of high-resolution model output.

549 Cloud microphysical properties heavily rely on remote sensing measurements to achieve more robust
550 sampling, with imperfect retrieval algorithms needed to estimate these variables. Microphysical retrievals
551 are more uncertain than typical atmospheric state measurements due to the need for many assumptions
552 related to cloud dynamical and physical processes. We have shown (in Section 3) that ground- and
553 satellite-based retrievals of N_d and R_{eff} are overall consistent with each other and with in-situ aircraft
554 measurements, with some systematic differences such as smaller N_d and larger R_{eff} in satellite retrievals.
555 The discrepancies between different retrievals can be larger for individual days (e.g., Figure S1) but can
556 be mitigated to some degrees when considering broader statistics (Figures 3 and 4). The usage of multiple
557 retrieval datasets is critical to understand the robustness of evaluation results, as the spread between
558 different datasets indicates how robust model-observation differences are and guides interpretations of
559 model biases to support model development.

560 Finally, this paper presents a few examples of how well E3SMv2 simulates aerosols, clouds and ACI. We
561 showed that ESMAC Diags can be used to target further investigation into specific parameterization
562 components. For example, the analysis of $N_d - \text{CCN}$ correlation indicates that E3SMv2 may exhibit too
563 weak aerosol activation in low CCN conditions and too strong in high CCN conditions; the analysis of
564 $\text{LWP} - N_d$ correlation indicates that either the precipitation suppression and cloud evaporation



565 mechanisms are not well represented, or there are other mechanisms dominating LWP – N_d correlation in
566 E3SMv2. These diagnostic analyses provide insights into areas in aerosols, clouds and ACI that warrant
567 special attention in future model development efforts. As ESMs continuously improve its physical
568 parameterizations, resolution, and numerical schemes, ESMAC Diags offers a valuable tool for
569 systematically evaluating the performance of the newer versions of a model in simulating aerosol, clouds
570 and ACI.



571 **Code availability:**

572 *The current version of ESMAC Diags is publicly available through GitHub ([https://github.com/eagles-](https://github.com/eagles-project/ESMAC_diags)*
573 *[project/ESMAC_diags](https://github.com/eagles-project/ESMAC_diags)) under the new BSD license. The exact version (2.1.2) of the code used to produce*
574 *the results used in this paper is archived on Zenodo (<https://doi.org/10.5281/zenodo.7696871>). The model*
575 *simulation used in this paper is version 2.0 (<https://doi.org/10.11578/E3SM/dc.20210927.1>) of E3SM.*

576 **Data availability:**

577 *Measurements from the HI-SCALE, ACE-ENA, MAGIC, and MARCUS campaigns as well as the SGP*
578 *and ENA sites are supported by the DOE Atmospheric Radiation Measurement (ARM) user facility and*
579 *available at <https://adc.arm.gov/discovery/>. Measurements from the CSET and SOCRATES campaigns*
580 *are supported by National Science Foundation (NSF) and obtained from NCAR Earth Observing*
581 *Laboratory at https://data.eol.ucar.edu/master_lists/generated/cset/ and*
582 *https://data.eol.ucar.edu/master_lists/generated/socrates/, respectively. DOI numbers or references of*
583 *individual datasets are given in Tables S1-S8. All the preprocessed observational and model data used to*
584 *produce the results used in this paper is archived on Zenodo (<https://doi.org/10.5281/zenodo.7478657>).*

585 **Author contribution:**

586 *ST, JDF and PM designed the diagnostics package; ST and ACV wrote the code and performed the*
587 *analysis; PW, XD, FM and MP processed the field campaign datasets and provided discussions on the*
588 *data quality issues; KZ contributed to the model simulation; JCH contributed to the package design and*
589 *setup; ST wrote the original manuscript; all authors reviewed and edited the manuscript.*

590 **Competing interests:**

591 *Po-Lun Ma is a Topical Editor of Geoscientific Model Development. Other authors declare that they have*
592 *no conflict of interest.*

593 **Acknowledgements:**

594 *This study was supported by the Enabling Aerosol-cloud interactions at GLocal convection-permitting*
595 *scalES (EAGLES) project (74358), funded by the U.S. Department of Energy, Office of Science, Office of*
596 *Biological and Environmental Research, Earth System Model Development (ESMD) program area. We*
597 *thank the numerous instrument mentors for providing the data. This research used resources of the*
598 *National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of*
599 *Science User Facility operated under Contract No. DE-AC02-05CH11231, using NERSC awards ALCC-*
600 *ERCAP0016315, BER-ERCAP0015329, BER-ERCAP0018473, and BER-ERCAP0020990. Pacific*
601 *Northwest National Laboratory (PNNL) is operated for DOE by Battelle Memorial Institute under*
602 *contract DE-AC05-76RL01830.*

603



604 **References:**

- 605 Albrecht, B., Ghate, V., Mohrmann, J., Wood, R., Zuidema, P., Bretherton, C., Schwartz, C., Eloranta, E.,
606 Glienke, S., Donaher, S., Sarkar, M., McGibbon, J., Nugent, A. D., Shaw, R. A., Fugal, J., Minnis, P.,
607 Paliknoda, R., Lussier, L., Jensen, J., Vivekanandan, J., Ellis, S., Tsai, P., Rilling, R., Haggerty, J.,
608 Campos, T., Stell, M., Reeves, M., Beaton, S., Allison, J., Stossmeister, G., Hall, S., and Schmidt, S.:
609 Cloud System Evolution in the Trades (CSET): Following the Evolution of Boundary Layer Cloud
610 Systems with the NSF–NCAR GV, *Bull. Amer. Meteor. Soc.*, 100, 93–121, [https://doi.org/10.1175/bams-](https://doi.org/10.1175/bams-d-17-0180.1)
611 [d-17-0180.1](https://doi.org/10.1175/bams-d-17-0180.1), 2019.
- 612 AMWG Diagnostic Package: [https://www.cesm.ucar.edu/working_groups/Atmosphere/amwg-](https://www.cesm.ucar.edu/working_groups/Atmosphere/amwg-diagnostics-package/)
613 [diagnostics-package/](https://www.cesm.ucar.edu/working_groups/Atmosphere/amwg-diagnostics-package/), last access: 2 November 2021. 2021.
- 614 Bennartz, R.: Global assessment of marine boundary layer cloud droplet number concentration from
615 satellite, *Journal of Geophysical Research: Atmospheres*, 112, <https://doi.org/10.1029/2006JD007547>,
616 2007.
- 617 Caldwell, P. M., Terai, C. R., Hillman, B., Keen, N. D., Bogenschutz, P., Lin, W., Beydoun, H., Taylor,
618 M., Bertagna, L., Bradley, A. M., Clevenger, T. C., Donahue, A. S., Eldred, C., Foucar, J., Golaz, J.-C.,
619 Guba, O., Jacob, R., Johnson, J., Krishna, J., Liu, W., Pressel, K., Salinger, A. G., Singh, B., Steyer, A.,
620 Ullrich, P., Wu, D., Yuan, X., Shpund, J., Ma, H.-Y., and Zender, C. S.: Convection-Permitting
621 Simulations With the E3SM Global Atmosphere Model, *Journal of Advances in Modeling Earth Systems*,
622 13, e2021MS002544, <https://doi.org/10.1029/2021MS002544>, 2021.
- 623 Cheng, A. and Xu, K.-M.: Evaluating Low-Cloud Simulation from an Upgraded Multiscale Modeling
624 Framework Model. Part III: Tropical and Subtropical Cloud Transitions over the Northern Pacific, *J.*
625 *Climate*, 26, 5761–5781, <https://doi.org/10.1175/JCLI-D-12-00650.1>, 2013.
- 626 E3SM Diagnostics: <https://e3sm.org/resources/tools/diagnostic-tools/e3sm-diagnostics/>, last access: 2
627 November 2021. 2021.
- 628 Eyring, V., Righi, M., Lauer, A., Evaldsson, M., Wenzel, S., Jones, C., Anav, A., Andrews, O., Cionni, I.,
629 Davin, E. L., Deser, C., Ehbrecht, C., Friedlingstein, P., Gleckler, P., Gottschaldt, K. D., Hagemann, S.,
630 Juckes, M., Kindermann, S., Krasting, J., Kunert, D., Levine, R., Loew, A., Mäkelä, J., Martin, G.,
631 Mason, E., Phillips, A. S., Read, S., Rio, C., Roehrig, R., Senftleben, D., Sterl, A., van Ulft, L. H.,
632 Walton, J., Wang, S., and Williams, K. D.: ESMValTool (v1.0) – a community diagnostic and
633 performance metrics tool for routine evaluation of Earth system models in CMIP, *Geosci. Model Dev.*, 9,
634 1747–1802, <https://doi.org/10.5194/gmd-9-1747-2016>, 2016.
- 635 Fast, J. D., Berg, L. K., Alexander, L., Bell, D., D’Ambro, E., Hubbe, J., Kuang, C., Liu, J., Long, C.,
636 Matthews, A., Mei, F., Newsom, R., Pekour, M., Pinterich, T., Schmid, B., Schobesberger, S., Shilling, J.,
637 Smith, J. N., Springston, S., Suski, K., Thornton, J. A., Tomlinson, J., Wang, J., Xiao, H., and Zelenyuk,
638 A.: Overview of the HI-SCALE Field Campaign: A New Perspective on Shallow Convective Clouds,
639 *Bull. Amer. Meteor. Soc.*, 100, 821–840, <https://doi.org/10.1175/bams-d-18-0030.1>, 2019.
- 640 Gallo, F., Uin, J., Springston, S., Wang, J., Zheng, G., Kuang, C., Wood, R., Azevedo, E. B.,
641 McComiskey, A., Mei, F., Theisen, A., Kyrouac, J., and Aiken, A. C.: Identifying a regional aerosol
642 baseline in the eastern North Atlantic using collocated measurements and a mathematical algorithm to
643 mask high-submicron-number-concentration aerosol events, *Atmos. Chem. Phys.*, 20, 7553–7573,
644 <https://doi.org/10.5194/acp-20-7553-2020>, 2020.
- 645 Gelaro, R., McCarty, W., Suarez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C., Darmenov, A.,
646 Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V.,
647 Conaty, A., da Silva, A., Gu, W., Kim, G. K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. E.,
648 Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert, S. D., Sienkiewicz, M., and Zhao, B.: The



- 649 Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2), *J. Climate*,
650 30, 5419-5454, <https://doi.org/10.1175/JCLI-D-16-0758.1>, 2017.
- 651 Glassmeier, F., Hoffmann, F., Johnson, J. S., Yamaguchi, T., Carslaw, K. S., and Feingold, G.: An
652 emulator approach to stratocumulus susceptibility, *Atmos. Chem. Phys.*, 19, 10191-10203,
653 <https://doi.org/10.5194/acp-19-10191-2019>, 2019.
- 654 Gleckler, P. J., Doutriaux, C., Durack, P. J., Taylor, K. E., Zhang, Y., Williams, D. N., Mason, E., and
655 Servonnat, J.: A more powerful reality test for climate models, *Eos, Trans. Amer. Geophys. Union*, 97,
656 <https://doi.org/10.1029/2016EO051663>, 2016.
- 657 Golaz, J.-C., Van Roekel, L. P., Zheng, X., Roberts, A. F., Wolfe, J. D., Lin, W., Bradley, A. M., Tang,
658 Q., Maltrud, M. E., Forsyth, R. M., Zhang, C., Zhou, T., Zhang, K., Zender, C. S., Wu, M., Wang, H.,
659 Turner, A. K., Singh, B., Richter, J. H., Qin, Y., Petersen, M. R., Mamatjanov, A., Ma, P.-L., Larson, V.
660 E., Krishna, J., Keen, N. D., Jeffery, N., Hunke, E. C., Hannah, W. M., Guba, O., Griffin, B. M., Feng,
661 Y., Engwirda, D., Di Vittorio, A. V., Dang, C., Conlon, L. M., Chen, C.-C.-J., Brunke, M. A., Bisht, G.,
662 Benedict, J. J., Asay-Davis, X. S., Zhang, Y., Zhang, M., Zeng, X., Xie, S., Wolfram, P. J., Vo, T.,
663 Veneziani, M., Tesfa, T. K., Sreepathi, S., Salinger, A. G., Jack Reeves Eyre, J. E., Prather, M. J.,
664 Mahajan, S., Li, Q., Jones, P. W., Jacob, R. L., Huebler, G. W., Huang, X., Hillman, B. R., Harrop, B. E.,
665 Foucar, J. G., Fang, Y., Comeau, D. S., Caldwell, P. M., Bartoletti, T., Balaguru, K., Taylor, M. A.,
666 McCoy, R. B., Leung, L. R., and Bader, D. C.: The DOE E3SM Model Version 2: Overview of the
667 physical model and initial model evaluation, *Journal of Advances in Modeling Earth Systems*, n/a,
668 e2022MS003156, <https://doi.org/10.1029/2022MS003156>, 2022.
- 669 Golaz, J.-C., Caldwell, P. M., Van Roekel, L. P., Petersen, M. R., Tang, Q., Wolfe, J. D., Abeshu, G.,
670 Anantharaj, V., Asay-Davis, X. S., Bader, D. C., Baldwin, S. A., Bisht, G., Bogenschutz, P. A.,
671 Branstetter, M., Brunke, M. A., Brus, S. R., Burrows, S. M., Cameron-Smith, P. J., Donahue, A. S.,
672 Deakin, M., Easter, R. C., Evans, K. J., Feng, Y., Flanner, M., Foucar, J. G., Fyke, J. G., Griffin, B. M.,
673 Hannay, C., Harrop, B. E., Hoffman, M. J., Hunke, E. C., Jacob, R. L., Jacobsen, D. W., Jeffery, N.,
674 Jones, P. W., Keen, N. D., Klein, S. A., Larson, V. E., Leung, L. R., Li, H.-Y., Lin, W., Lipscomb, W. H.,
675 Ma, P.-L., Mahajan, S., Maltrud, M. E., Mamatjanov, A., McClean, J. L., McCoy, R. B., Neale, R. B.,
676 Price, S. F., Qian, Y., Rasch, P. J., Reeves Eyre, J. E. J., Riley, W. J., Ringler, T. D., Roberts, A. F.,
677 Roesler, E. L., Salinger, A. G., Shaheen, Z., Shi, X., Singh, B., Tang, J., Taylor, M. A., Thornton, P. E.,
678 Turner, A. K., Veneziani, M., Wan, H., Wang, H., Wang, S., Williams, D. N., Wolfram, P. J., Worley, P.
679 H., Xie, S., Yang, Y., Yoon, J.-H., Zelinka, M. D., Zender, C. S., Zeng, X., Zhang, C., Zhang, K., Zhang,
680 Y., Zheng, X., Zhou, T., and Zhu, Q.: The DOE E3SM Coupled Model Version 1: Overview and
681 Evaluation at Standard Resolution, *J. Adv. Model. Earth Syst.*, 11, 2089-2129,
682 <https://doi.org/10.1029/2018ms001603>, 2019.
- 683 Grosvenor, D. P., Sourdeval, O., Zuidema, P., Ackerman, A., Alexandrov, M. D., Bennartz, R., Boers, R.,
684 Cairns, B., Chiu, J. C., Christensen, M., Deneke, H., Diamond, M., Feingold, G., Fridlind, A., Hünerbein,
685 A., Knist, C., Kollias, P., Marshak, A., McCoy, D., Merk, D., Painemal, D., Rausch, J., Rosenfeld, D.,
686 Russchenberg, H., Seifert, P., Sinclair, K., Stier, P., van Diedenhoven, B., Wendisch, M., Werner, F.,
687 Wood, R., Zhang, Z., and Quaas, J.: Remote Sensing of Droplet Number Concentration in Warm Clouds:
688 A Review of the Current State of Knowledge and Perspectives, *Reviews of Geophysics*, 56, 409-453,
689 <https://doi.org/10.1029/2017RG000593>, 2018.
- 690 Gryspeerdt, E., Goren, T., Sourdeval, O., Quaas, J., Mülmenstädt, J., Dipu, S., Unglaub, C., Gettelman,
691 A., and Christensen, M.: Constraining the aerosol influence on cloud liquid water path, *Atmos. Chem.*
692 *Phys.*, 19, 5331-5347, <https://doi.org/10.5194/acp-19-5331-2019>, 2019.
- 693 Humphries, R.: MARCUS ARM CN and CCN data reprocessed to remove ship exhaust influence (v2)
694 [dataset], <https://doi.org/10.25919/ezp0-em87>, 2020. Accessed 8 March 2022.



- 695 Humphries, R. S., McRobert, I. M., Ponsonby, W. A., Ward, J. P., Keywood, M. D., Loh, Z. M.,
696 Krummel, P. B., and Harnwell, J.: Identification of platform exhaust on the RV Investigator, Atmos.
697 Meas. Tech., 12, 3019-3038, <https://doi.org/10.5194/amt-12-3019-2019>, 2019.
- 698 IPCC: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth
699 Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press,
700 Cambridge, United Kingdom and New York, NY, USA, 2391 pp.,
701 <https://doi.org/10.1017/9781009157896>, 2021.
- 702 Jiang, X., Lau, N.-C., and Klein, S. A.: Role of eastward propagating convection systems in the diurnal
703 cycle and seasonal mean of summertime rainfall over the U.S. Great Plains, Geophys. Res. Lett., 33,
704 <https://doi.org/10.1029/2006gl027022>, 2006.
- 705 Lewis, E. R. and Teixeira, J.: Dispelling clouds of uncertainty, Eos, Trans. Amer. Geophys. Union, 96,
706 <https://doi.org/10.1029/2015eo031303>, 2015.
- 707 Lim, K.-S. S., Riihimaki, L., Comstock, J. M., Schmid, B., Sivaraman, C., Shi, Y., and McFarquhar, G.
708 M.: Evaluation of long-term surface-retrieved cloud droplet number concentration with in situ aircraft
709 observations, Journal of Geophysical Research: Atmospheres, 121, 2318-2331,
710 <https://doi.org/10.1002/2015JD024082>, 2016.
- 711 Ma, P. L., Harrop, B. E., Larson, V. E., Neale, R. B., Gettelman, A., Morrison, H., Wang, H., Zhang, K.,
712 Klein, S. A., Zelinka, M. D., Zhang, Y., Qian, Y., Yoon, J. H., Jones, C. R., Huang, M., Tai, S. L., Singh,
713 B., Bogenschütz, P. A., Zheng, X., Lin, W., Quaas, J., Chepfer, H., Brunke, M. A., Zeng, X.,
714 Mühlmenstädt, J., Hagos, S., Zhang, Z., Song, H., Liu, X., Pritchard, M. S., Wan, H., Wang, J., Tang, Q.,
715 Caldwell, P. M., Fan, J., Berg, L. K., Fast, J. D., Taylor, M. A., Golaz, J. C., Xie, S., Rasch, P. J., and
716 Leung, L. R.: Better calibration of cloud parameterizations and subgrid effects increases the fidelity of the
717 E3SM Atmosphere Model version 1, Geosci. Model Dev., 15, 2881-2916, <https://doi.org/10.5194/gmd-15-2881-2022>, 2022.
- 719 Maloney, E. D., Gettelman, A., Ming, Y., Neelin, J. D., Barrie, D., Mariotti, A., Chen, C. C., Coleman, D.
720 R. B., Kuo, Y.-H., Singh, B., Annamalai, H., Berg, A., Booth, J. F., Camargo, S. J., Dai, A., Gonzalez, A.,
721 Hafner, J., Jiang, X., Jing, X., Kim, D., Kumar, A., Moon, Y., Naud, C. M., Sobel, A. H., Suzuki, K.,
722 Wang, F., Wang, J., Wing, A. A., Xu, X., and Zhao, M.: Process-Oriented Evaluation of Climate and
723 Weather Forecasting Models, Bull. Amer. Meteor. Soc., 100, 1665-1686, <https://doi.org/10.1175/bams-d-18-0042.1>, 2019.
- 725 McFarquhar, G. M., Bretherton, C. S., Marchand, R., Protat, A., DeMott, P. J., Alexander, S. P., Roberts,
726 G. C., Twohy, C. H., Toohey, D., Siems, S., Huang, Y., Wood, R., Rauber, R. M., Lasher-Trapp, S.,
727 Jensen, J., Stith, J. L., Mace, J., Um, J., Järvinen, E., Schnaiter, M., Gettelman, A., Sanchez, K. J.,
728 McCluskey, C. S., Russell, L. M., McCoy, I. L., Atlas, R. L., Bardeen, C. G., Moore, K. A., Hill, T. C. J.,
729 Humphries, R. S., Keywood, M. D., Ristovski, Z., Cravigan, L., Schofield, R., Fairall, C., Mallet, M. D.,
730 Kreidenweis, S. M., Rainwater, B., D'Alessandro, J., Wang, Y., Wu, W., Saliba, G., Levin, E. J. T., Ding,
731 S., Lang, F., Truong, S. C. H., Wolff, C., Haggerty, J., Harvey, M. J., Klekociuk, A. R., and McDonald,
732 A.: Observations of Clouds, Aerosols, Precipitation, and Surface Radiation over the Southern Ocean: An
733 Overview of CAPRICORN, MARCUS, MICRE, and SOCRATES, Bull. Amer. Meteor. Soc., 102, E894-
734 E928, <https://doi.org/10.1175/bams-d-20-0132.1>, 2021.
- 735 Min, Q. and Harrison, L. C.: Cloud properties derived from surface MFRSR measurements and
736 comparison with GOES results at the ARM SGP Site, Geophysical Research Letters, 23, 1641-1644,
737 <https://doi.org/10.1029/96GL01488>, 1996.
- 738 Minnis, P., Nguyen, L., Palikonda, R., Heck, P. W., Spangenberg, D. A., Doelling, D. R., Ayers, J. K.,
739 Smith, J. W. L., Khaiyer, M. M., Trepte, Q. Z., Avey, L. A., Chang, F.-L., Yost, C. R., Chee, T. L., and
740 Szedung, S.-M.: Near-real time cloud retrievals from operational and research meteorological satellites,



- 741 Proc. SPIE Europe Remote Sens., Cardiff, Wales, UK., 15-18 September, 710703,
742 <https://doi.org/10.1117/12.800344>, 2008.
- 743 Minnis, P., Sun-Mack, S., Young, D. F., Heck, P. W., Garber, D. P., Chen, Y., Spangenberg, D. A.,
744 Arduini, R. F., Trepte, Q. Z., Smith, W. L., Ayers, J. K., Gibson, S. C., Miller, W. F., Hong, G.,
745 Chakrapani, V., Takano, Y., Liou, K. N., Xie, Y., and Yang, P.: CERES Edition-2 Cloud Property
746 Retrievals Using TRMM VIRS and Terra and Aqua MODIS Data—Part I: Algorithms, IEEE
747 Transactions on Geoscience and Remote Sensing, 49, 4374-4400,
748 <https://doi.org/10.1109/TGRS.2011.2144601>, 2011.
- 749 Myhre, G., Samset, B. H., Schulz, M., Balkanski, Y., Bauer, S., Bernsten, T. K., Bian, H., Bellouin, N.,
750 Chin, M., Diehl, T., Easter, R. C., Feichter, J., Ghan, S. J., Hauglustaine, D., Iversen, T., Kinne, S.,
751 Kirkevåg, A., Lamarque, J. F., Lin, G., Liu, X., Lund, M. T., Luo, G., Ma, X., van Noije, T., Penner, J. E.,
752 Rasch, P. J., Ruiz, A., Seland, Ø., Skeie, R. B., Stier, P., Takemura, T., Tsigaridis, K., Wang, P., Wang,
753 Z., Xu, L., Yu, H., Yu, F., Yoon, J. H., Zhang, K., Zhang, H., and Zhou, C.: Radiative forcing of the
754 direct aerosol effect from AeroCom Phase II simulations, Atmos. Chem. Phys., 13, 1853-1877,
755 <https://doi.org/10.5194/acp-13-1853-2013>, 2013.
- 756 NETCDF: Introduction and Overview: <https://www.unidata.ucar.edu/software/netcdf/docs/index.html>,
757 last access: 12 November 2022. 2022.
- 758 Quaas, J., Boucher, O., Bellouin, N., and Kinne, S.: Satellite-based estimate of the direct and indirect
759 aerosol climate forcing, Journal of Geophysical Research: Atmospheres, 113,
760 <https://doi.org/10.1029/2007JD008962>, 2008.
- 761 Rasch, P. J., Xie, S., Ma, P.-L., Lin, W., Wang, H., Tang, Q., Burrows, S. M., Caldwell, P., Zhang, K.,
762 Easter, R. C., Cameron-Smith, P., Singh, B., Wan, H., Golaz, J.-C., Harrop, B. E., Roesler, E.,
763 Bacmeister, J., Larson, V. E., Evans, K. J., Qian, Y., Taylor, M., Leung, L. R., Zhang, Y., Brent, L.,
764 Branstetter, M., Hannay, C., Mahajan, S., Mامتjanov, A., Neale, R., Richter, J. H., Yoon, J.-H., Zender,
765 C. S., Bader, D., Flanner, M., Foucar, J. G., Jacob, R., Keen, N., Klein, S. A., Liu, X., Salinger, A. G.,
766 Shrivastava, M., and Yang, Y.: An Overview of the Atmospheric Component of the Energy Exascale
767 Earth System Model, J. Adv. Model. Earth Syst., 11, 2377-2411, <https://doi.org/10.1029/2019ms001629>,
768 2019.
- 769 Reddington, C. L., Carslaw, K. S., Stier, P., Schutgens, N., Coe, H., Liu, D., Allan, J., Browse, J., Pringle,
770 K. J., Lee, L. A., Yoshioka, M., Johnson, J. S., Regayre, L. A., Spracklen, D. V., Mann, G. W., Clarke,
771 A., Hermann, M., Henning, S., Wex, H., Kristensen, T. B., Leaitch, W. R., Pöschl, U., Rose, D., Andreae,
772 M. O., Schmale, J., Kondo, Y., Oshima, N., Schwarz, J. P., Nenes, A., Anderson, B., Roberts, G. C.,
773 Snider, J. R., Leck, C., Quinn, P. K., Chi, X., Ding, A., Jimenez, J. L., and Zhang, Q.: The Global Aerosol
774 Synthesis and Science Project (GASSP): Measurements and Modeling to Reduce Uncertainty, Bull.
775 Amer. Meteor. Soc., 98, 1857-1877, <https://doi.org/10.1175/bams-d-15-00317.1>, 2017.
- 776 Riihimäki, L., McFarlane, S., and Sivaraman, C.: Droplet Number Concentration Value-Added Product,
777 ARM Research Facility, Report number: DOE/SC-ARM-TR-140, 2021.
- 778 Schulz, M., Textor, C., Kinne, S., Balkanski, Y., Bauer, S., Bernsten, T., Berglen, T., Boucher, O.,
779 Dentener, F., Guibert, S., Isaksen, I. S. A., Iversen, T., Koch, D., Kirkevåg, A., Liu, X., Montanaro, V.,
780 Myhre, G., Penner, J. E., Pitari, G., Reddy, S., Seland, Ø., Stier, P., and Takemura, T.: Radiative forcing
781 by aerosols as derived from the AeroCom present-day and pre-industrial simulations, Atmos. Chem.
782 Phys., 6, 5225-5246, <https://doi.org/10.5194/acp-6-5225-2006>, 2006.
- 783 Song, X., Zhang, G. J., and Li, J. L. F.: Evaluation of Microphysics Parameterization for Convective
784 Clouds in the NCAR Community Atmosphere Model CAM5, J. Climate, 25, 8568-8590,
785 <https://doi.org/10.1175/JCLI-D-11-00563.1>, 2012.



- 786 Sun, J., Zhang, K., Wan, H., Ma, P.-L., Tang, Q., and Zhang, S.: Impact of Nudging Strategy on the
787 Climate Representativeness and Hindcast Skill of Constrained EAMv1 Simulations, *J. Adv. Model. Earth*
788 *Syst.*, 11, 3911-3933, <https://doi.org/10.1029/2019MS001831>, 2019.
- 789 Tang, Q., Klein, S. A., Xie, S., Lin, W., Golaz, J. C., Roesler, E. L., Taylor, M. A., Rasch, P. J., Bader, D.
790 C., Berg, L. K., Caldwell, P., Giangrande, S. E., Neale, R. B., Qian, Y., Riihimaki, L. D., Zender, C. S.,
791 Zhang, Y., and Zheng, X.: Regionally refined test bed in E3SM atmosphere model version 1 (EAMv1)
792 and applications for high-resolution modeling, *Geosci. Model Dev.*, 12, 2679-2706,
793 <https://doi.org/10.5194/gmd-12-2679-2019>, 2019.
- 794 Tang, S., Zhang, M., and Xie, S.: An ensemble constrained variational analysis of atmospheric forcing
795 data and its application to evaluate clouds in CAM5, *J. Geophys. Res. Atmos.*, 121, 33-48,
796 <https://doi.org/10.1002/2015JD024167>, 2016.
- 797 Tang, S., Gleckler, P., Xie, S., Lee, J., Ahn, M.-S., Covey, C., and Zhang, C.: Evaluating the Diurnal and
798 Semidiurnal Cycle of Precipitation in CMIP6 Models Using Satellite- and Ground-Based Observations, *J.*
799 *Climate*, 34, 3189-3210, <https://doi.org/10.1175/jcli-d-20-0639.1>, 2021.
- 800 Tang, S., Fast, J. D., Zhang, K., Hardin, J. C., Varble, A. C., Shilling, J. E., Mei, F., Zawadowicz, M. A.,
801 and Ma, P. L.: Earth System Model Aerosol–Cloud Diagnostics (ESMAC Diags) package, version 1:
802 assessing E3SM aerosol predictions using aircraft, ship, and surface measurements, *Geosci. Model Dev.*,
803 15, 4055-4076, <https://doi.org/10.5194/gmd-15-4055-2022>, 2022a.
- 804 Tang, S., Xie, S., Guo, Z., Hong, S.-Y., Khouider, B., Klocke, D., Köhler, M., Koo, M.-S., Krishna, P.
805 M., Larson, V. E., Park, S., Vaillancourt, P. A., Wang, Y.-C., Yang, J., Daleu, C. L., Homeyer, C. R.,
806 Jones, T. R., Malap, N., Neggers, R., Prabhakaran, T., Ramirez, E., Schumacher, C., Tao, C., Bechtold,
807 P., Ma, H.-Y., Neelin, J. D., and Zeng, X.: Long-term single-column model intercomparison of diurnal
808 cycle of precipitation over midlatitude and tropical land, *Quarterly Journal of the Royal Meteorological*
809 *Society*, 148, 641-669, <https://doi.org/10.1002/qj.4222>, 2022b.
- 810 Turner, D. D., Lo, C., Min, Q., Zhang, D., and Gaustad, K.: Cloud Optical Properties from the Multifilter
811 Shadowband Radiometer (MFRSRCLDOD): An ARM Value-Added Product, ARM Research Facility,
812 Report number: DOE/SC-ARM-TR-047, 2021.
- 813 Varble, A. C., Ma, P. L., christensen, M. W., Mülmenstädt, J., Tang, S., and Fast, J. D.: Evaluation of
814 Liquid Cloud Albedo Susceptibility in E3SM Using Coupled Eastern North Atlantic Surface and Satellite
815 Retrievals, *Atmos. Chem. Phys.*, submitted, 2023.
- 816 Wang, J., Wood, R., Jensen, M. P., Chiu, J. C., Liu, Y., Lamer, K., Desai, N., Giangrande, S. E., Knopf,
817 D. A., Kollias, P., Laskin, A., Liu, X., Lu, C., Mechem, D., Mei, F., Starzec, M., Tomlinson, J., Wang, Y.,
818 Yum, S. S., Zheng, G., Aiken, A. C., Azevedo, E. B., Blanchard, Y., China, S., Dong, X., Gallo, F., Gao,
819 S., Ghate, V. P., Glienke, S., Goldberger, L., Hardin, J. C., Kuang, C., Luke, E. P., Matthews, A. A.,
820 Miller, M. A., Moffet, R., Pekour, M., Schmid, B., Sedlacek, A. J., Shaw, R. A., Shilling, J. E., Sullivan,
821 A., Suski, K., Veghte, D. P., Weber, R., Wyant, M., Yeom, J., Zawadowicz, M., and Zhang, Z.: Aerosol
822 and Cloud Experiments in the Eastern North Atlantic (ACE-ENA), *Bull. Amer. Meteor. Soc.*, 1-51,
823 <https://doi.org/10.1175/bams-d-19-0220.1>, 2021.
- 824 Watson-Parris, D., Schutgens, N., Reddington, C., Pringle, K. J., Liu, D., Allan, J. D., Coe, H., Carslaw,
825 K. S., and Stier, P.: In situ constraints on the vertical distribution of global aerosol, *Atmos. Chem. Phys.*,
826 19, 11765-11790, <https://doi.org/10.5194/acp-19-11765-2019>, 2019.
- 827 Wu, P., Dong, X., Xi, B., Tian, J., and Ward, D. M.: Profiles of MBL Cloud and Drizzle Microphysical
828 Properties Retrieved From Ground-Based Observations and Validated by Aircraft In Situ Measurements
829 Over the Azores, *Journal of Geophysical Research: Atmospheres*, 125, e2019JD032205,
830 <https://doi.org/10.1029/2019JD032205>, 2020.



- 831 Xie, S., Lin, W., Rasch, P. J., Ma, P.-L., Neale, R., Larson, V. E., Qian, Y., Bogenschütz, P. A., Caldwell,
832 P., Cameron-Smith, P., Golaz, J.-C., Mahajan, S., Singh, B., Tang, Q., Wang, H., Yoon, J.-H., Zhang, K.,
833 and Zhang, Y.: Understanding Cloud and Convective Characteristics in Version 1 of the E3SM
834 Atmosphere Model, *J. Adv. Model. Earth Syst.*, 10, 2618-2644, <https://doi.org/10.1029/2018MS001350>,
835 2018.
- 836 Xu, K.-M. and Cheng, A.: Evaluating Low-Cloud Simulation from an Upgraded Multiscale Modeling
837 Framework Model. Part I: Sensitivity to Spatial Resolution and Climatology, *J. Climate*, 26, 5717-5740,
838 <https://doi.org/10.1175/JCLI-D-12-00200.1>, 2013a.
- 839 Xu, K.-M. and Cheng, A.: Evaluating Low-Cloud Simulation from an Upgraded Multiscale Modeling
840 Framework Model. Part II: Seasonal Variations over the Eastern Pacific, *J. Climate*, 26, 5741-5760,
841 <https://doi.org/10.1175/JCLI-D-12-00276.1>, 2013b.
- 842 Zhang, C., Xie, S., Tao, C., Tang, S., Emmenegger, T., Neelin, J. D., Schiro, K. A., Lin, W., and Shaheen,
843 Z.: The ARM Data-Oriented Metrics and Diagnostics Package for Climate Models: A New Tool for
844 Evaluating Climate Models with Field Data, *Bull. Amer. Meteor. Soc.*, 101, E1619-E1627,
845 <https://doi.org/10.1175/bams-d-19-0282.1>, 2020.
- 846 Zhang, S., Zhang, K., Wan, H., and Sun, J.: Further improvement and evaluation of nudging in the E3SM
847 Atmosphere Model version 1 (EAMv1): simulations of the mean climate, weather events, and
848 anthropogenic aerosol effects, *Geosci. Model Dev.*, 15, 6787-6816, [https://doi.org/10.5194/gmd-15-6787-](https://doi.org/10.5194/gmd-15-6787-2022)
849 [2022](https://doi.org/10.5194/gmd-15-6787-2022), 2022.
- 850 Zhang, Y., Xie, S., Lin, W., Klein, S. A., Zelinka, M., Ma, P.-L., Rasch, P. J., Qian, Y., Tang, Q., and Ma,
851 H.-Y.: Evaluation of Clouds in Version 1 of the E3SM Atmosphere Model With Satellite Simulators,
852 *Journal of Advances in Modeling Earth Systems*, 11, 1253-1268, <https://doi.org/10.1029/2018MS001562>,
853 2019.
- 854 Zhou, X., Kollias, P., and Lewis, E. R.: Clouds, Precipitation, and Marine Boundary Layer Structure
855 during the MAGIC Field Campaign, *J. Climate*, 28, 2420-2442, <https://doi.org/10.1175/jcli-d-14-00320.1>,
856 2015.
- 857