Supplementary Information for

Identifying climate model structural inconsistencies allows for tight constraint of aerosol radiative forcing

Leighton A. Regayre[1,2], Lucia Deaconu[3,4], Daniel P. Grosvenor[1,2,5], David M. H. Sexton[2], Christopher Symonds[5], Tom Langton[3], Duncan Watson-Paris[3], Jane P. Mulcahy[2], Kirsty J. Pringle[1,5,6], Mark Richardson[5], Jill S. Johnson[1,7], John W. Rostron[2], Hamish Gordon[1,8], Grenville Lister[9,10], Philip Stier[3] and Ken S. Carslaw[1]

[1]Institute for Climate and Atmospheric Science, School of Earth and Environment, University of Leeds, Leeds, LS2 9JT, UK
[2]Met Office Hadley Centre, Exeter, Fitzroy Road, Exeter, Devon, EX1 3PB, UK
[3]Atmospheric, Oceanic and Planetary Physics Department, University of Oxford, Oxford, OX1 3PU
[4]Faculty of Environmental Science and Engineering, Babes-Bolyai University, Cluj, Romania, 400294
[5]Centre for Environmental Modelling and Computation, School of Earth and Environment, University of Leeds, Leeds, LS2 9JT, UK
[6]Edinburgh Parallel Computing Centre, Bayes Centre, University of Edinburgh, EH8 9BT
[7]School of Mathematics and Statistics, University of Sheffield, Sheffield, S3 7RH, UK
[8]Department of Chemical Engineering and Center for Atmospheric Particle Studies, Carnegie Mellon University, Pittsburgh, PA 15213,USA
[9]Department of Meteorology, University of Reading, RG6 6AH, UK
[10]National Centre for Atmospheric Science, Reading, RG6 6AH, UK

Leighton Regayre
Email: L.A.Regayre@leeds.ac.uk

This PDF file includes:

Supplementary Information Text

Model version

We used the atmosphere-only configuration of version 1 of the UK Earth System Model (UKESM1) (1) to create our perturbed parameter ensembles (PPEs). UKESM1 was the model version submitted to the 6th Coupled Model Intercomparison Project (CMIP6) (2). UKESM1 is based on the HADGEM3-GC3.1 physical climate model (3) with additional coupling to key Earth System processes (1), including the United Kingdom Chemistry and Aerosol (UKCA) model (4). The atmosphere-only configuration as used here consists of the GA7.1 atmosphere (5, 6), with additional aerosol, cloud and physical atmosphere structural updates as implemented in UKESM1 (6). GA7.1 includes several structural advancements to the aerosol component of the model which significantly affect anthropogenic aerosol radiative forcing (7). We refer to this model version as UKESM1-A.

Horizontal wind fields above around 2km in our simulations (model vertical level 17) were nudged towards ERA-Interim values for the period December 2016 to November 2017. Nudging largely removes the effects of differences in large-scale meteorology from our PPE members, meaning we can attribute differences between model variants to perturbed parameter values. We do not nudge winds within the boundary layer, as many of our parameters are intended to affect meteorological conditions in this part of the atmosphere.

The model was forced using anthropogenic SO_2 emissions, for the years 2014 and 1850, as prescribed in CMIP6 simulations. We calculated aerosol effective radiative forcing (ΔF_{aer}) as the difference in top-of-the-atmosphere radiative fluxes between these two periods. We accounted for above-cloud aerosol in our calculation of the components of ΔF_{aer} (8) and aerosol-cloud interactions (9).

Carbonaceous aerosol from fossil fuel and residential sources match those used in CMIP6 in our early-industrial simulations. However, in our present-day simulations (2014 anthropogenic SO_2 emissions) we prescribed carbonaceous aerosol from biomass burning sources using emissions generated using Copernicus Atmospheric Monitoring Service Information (December 2016 to November 2017) (10) and spread these emissions between the surface and around 3km. We used emissions for the same period as prescribed wind fields, for the closest possible comparison to observed values. In our early-industrial simulations (1850 anthropogenic SO_2 emissions) we similarly scaled CMIP6 carbonaceous aerosol from biomass burning over model levels between the surface and around 3km.

We also prescribed, rather than simulated, sea surface temperatures and sea ice fraction for the December 2016 to November 2017 period. We prescribed land surface quantities, ocean surface concentrations of dimethylsulfide (DMS) and chlorophyll, and atmospheric concentrations of gas species (including oxidants OH and O_3, which we then perturb), using monthly mean output values from a fully-coupled version of the UKESM model, averaged over the 1979 to 2014 period. Additionally, we prescribe volcanic SO_2 emissions for continuously emitting and sporadically erupting volcanoes (11) and for explosive volcanic eruptions (12).

We use an N96 horizontal resolution, which is 1.875×1.25^{o} (208 ×139 km) at the equator, with 85 vertical levels between the surface and 85km in altitude. Model vertical levels use a stretched grid such that the vertical resolution is around 13 m near the surface and around 150 to 200 m at the top of the boundary layer. We chose this resolution since it is the same as that used for long climate runs in CMIP6.

Aerosol number concentrations are treated prognostically with the GLOMAP multi-modal scheme (13, 14), which uses five log-normal aerosol size modes and includes sulfate, sea-salt, black carbon and organic carbon chemical components that are internally mixed within each size mode. Mineral dust is simulated separately using the CLASSIC dust scheme (15). GLOMAP simulates new particle formation, coagulation, gas-to-particle transfer, cloud processing and deposition of gases and aerosols. The activation of aerosols into cloud droplets is calculated using distributions of sub-grid vertical velocities based on available turbulent kinetic energy (16) and the removal of cloud droplets by autoconversion to rain is calculated by the host model. Aerosols are also removed by impaction scavenging of falling raindrops according to the collocation of clouds and precipitation (17, 18).

We modified some aspects of UKESM1-A in our PPE. Firstly, we define an ice mass fraction threshold (cloud_ice_thresh; table S1) above which no nucleation scavenging occurs, to allow sufficient aerosol to be transported to the Arctic (19). We assume that the wet scavenging of all aerosol particles (soluble and insoluble) is set to zero in large-scale raining clouds if the simulated ice to total water mass fraction is higher than this fixed value. This first structural change replicates the model change we implemented in (20) which is not yet in the release version of the model. We evaluated the climatic importance of this parameter as a cause of uncertainty in (21–24). Secondly, we implemented a version of look-up tables for aerosol optical properties (25) that includes optical properties for mineral dust (26) and higher-resolution increments of the imaginary part of the refractive indices, to better resolve the absorption coefficient of aerosols, especially at the low-absorption end of the spectrum. Finally, we included an organically-mediated boundary layer nucleation parametrisation (27) to enhance remote marine and early-industrial aerosol concentrations in the model.

Perturbed Parameter Ensembles (PPEs)

We created a new PPE of 221 UKESM1-A model simulations for this study. Each member of the PPE has a distinct combination of 37 aerosol and physical atmosphere parameter values, spanning expert elicited ranges (table S1). Parameters perturbed in previous PPEs using older versions of our model (20, 28) and identified as important causes of uncertainty in cloud active aerosol concentrations and/or aerosol forcing (21, 29, 30) are perturbed here, alongside parameters associated with structural model developments (5–7). Many parameters are described in table S1 as ‘scale factors’, which indicates we scaled the corresponding process parameter up or down over the indicated range. Other parameters are specific components of process parametrizations.

Multi-stage experimental design

We created our PPE in two stages, following ‘history matching’ conventions (31, 32). In the first stage, , the 221 member ensemble was made by combining a simulation using median values for each parameter with 220 additional parameter combinations were drawn from a Latin hypercube optimized to ensure design points were distributed as evenly as possible across the 37-dimensional parameter space, using the ‘optimumLHS’ R function (33). We output monthly mean data for 4 months for each ensemble member and analysed output from the final month, which corresponded to anthropogenic emissions for May 2014 and horizontal wind fields for 2015.

We created statistical Gaussian process emulators (34) of multiple monthly mean output variables. For each variable, we sampled one million model variants (parameter combinations), from the corresponding emulator, that uniformly spanned the uncertain parameter space, in keeping with efforts to constrain aerosol radiative forcing uncertainty using large ensembles (22, 23). We then ruled out implausible parameter combinations that compared poorly to observations within known emulator uncertainty and assumed observational uncertainty bounds. Observations included global mean shortwave and longwave top-of-the-atmosphere radiative fluxes from the Clouds and the Earth’s Radiant Energy System experiment (35) and global mean precipitation amount from version 2 of the Global Precipitation Climatology Project (36). Additionally, we used North Pacific and North Atlantic marine only data between 10º and 60º N for low- and total-cloud fraction from the Moderate Resolution Imaging Spectroradiometer (37) and LWP from the Multi-Sensor

Advanced Climatology of Liquid Water Path data set (38). We assumed errors of 8%, 2%, 30%, 20%, 20% and 40% respectively for these observations. Of the retained model variants, we started with the parameter combination central to the retained space, then iteratively identified an additional 220 parameter combinations with the greatest Euclidean distance from existing points, until we had a new and diverse set of 221 members that span the uncertain parameter space retained from the first observational filter. Thus, the simulation to perturbed parameter ratio in our PPE is around six.

We created full year simulations forced using 1850 and 2014 anthropogenic aerosol and precursor gas emissions for each of these 221 PPE members. We held greenhouse gas concentrations constant at 2014 levels. These 221 model simulationa are the second stage of our PPE creation process. As in the first stage, we created and validated statistical emulators of global mean and regional mean variables using these 221 members, then created a sample of one million model variants from these emulators.

Measurements

Measurements: Regional mean cloud and radiative properties

Satellites carry instruments that measure atmospheric properties, then geophysical quantities are calculated using retrieval algorithms and inverse modelling methods. We compare values derived from MODIS instruments (39) to model output calculated using the Cloud Feedback Model Intercomparison MODIS satellite simulator (40, 41) where available. This simulator minimizes errors in model comparisons to MODIS retrieval data, by recreating as near as possible what the satellite would retrieve given the model-simulated atmospheric conditions.

We use MODIS retrievals of of liquid water path (LWP), liquid cloud fraction (f_c), cloud optical depth (τ_c) and cloud droplet effective radius (r_e) at 1° by 1° resolution and use τ_c and r_e values to calculate cloud droplet number concentration (N_d). We assume constant N_d throughout cloud layers, which is a good approximation for stratocumulus clouds (9, 42). We compare all cloud properties to satellite-simulator output and compare N_d to values calculated at model–simulated cloud tops. We use outgoing top-of-the-atmosphere shortwave radiative flux (F_{SW}) measurements from the Clouds and the Earth's Radiant Energy Systems instrument (35).

We degrade all satellite-derived measurements to match our model resolution, then identify regions with high cloud fraction across the year (table S2). We evaluate constraint variables at the regional level, since there are no clear relationships between aerosol forcing and observations of global mean values (SI Fig. S26). These regions are dominated by stratocumulus cloud, have relatively high multi-model diversity in cloud amount in CMIP6 models (43) and are the most important regions for understanding the role of aerosol-cloud interactions (44). We only used values corresponding to model grid boxes with at least 50% ocean coverage in our area-weighted regional mean calculations.

Measurements: Hemispheric difference in N_d

The contrast between marine N_d in the polluted Northern Hemisphere and relatively pristine Southern Hemisphere (H_d) can act as a proxy for the difference in N_d between the early-industrial and present-day atmospheres (45). We calculate H_d as the difference in hemispheric mean marine N_d values, using MODIS τ_c and r_e values.

Measurements: Transects from stratocumulus- to cumulus-dominated regions

We calculate the changes in multiple measurement values along transects from regions dominated by stratocumulus cloud to those dominated by cumulus. Cloud physical and radiative properties are sensitive to changes in aerosol concentrations in these transition regions (46). We chose transects on the Eastern side of major ocean basins (Fig. S12, table S3) where air is advected from the sub-tropics towards the equator. We used data from July 2017 for Northern Hemisphere transects and for November 2017 for Southern Hemisphere transects, to evaluate relatively strong transitions in warmer months.

The gradients of linear relationships between observed values and distances (in meters) along these transects are used as constraint variables. We evaluate gradients of individual measurement types including N_d, r_e, f_c and LWP, calculated using values that informed our regional mean calculations. Additionally, we calculate gradients of aerosol index (AI; the total aerosol optical depth at 550 nm multiplied by the Ȧngstrȍm exponent) using MODIS aerosol optical depth retrieval data. We additionally include gradients of ratios of observation types along each transect as constraint variables. We calculate gradients of ratios using natural logarithms following (47). We include the ratios of N_d to AI, r_e to N_d, LWP to N_d and f_c to N_d. We compare satellite-derived values to probability distributions of corresponding output from our PPE members in SI Fig. S27-30.

Relative Importance of Parameters

For each constraint variable, we calculated the relative importance of parameters as causes of model uncertainty using Pearson partial correlations (48). Partial correlations control for the effects of all other perturbed parameters on the variable of interest in the correlation calculations. A partial correlation between some variable and a chosen parameter is the correlation between the residuals from a) linear regression of the variable on the remaining 36 parameters and b) linear regression of the chosen parameter on the remaining 36. For each of the 37 model parameters, we define the relative importance metric, for any chosen variable, as the proportion of its partial correlation with the chosen variable to the total of the 37 partial correlations, multiplied by the sign of the gradient of the linear regression of the variable on the parameter in question. We include the sign of the gradient in the relative importance metric to convey the effect of changing parameter values on the variable, which helps develop a process-based understanding of model behavior within the uncertainty framework. Relative importance metrics are used here as a guide to our choice of variables for model constraint and inform our understanding of how they relate to ΔF_{aer}. Variance-based sensitivity analyses (49) can be used to robustly quantify the percentage of variance caused by each parameter. However, the multi-stage design of our PPE leaves gaps in the parameter space that limits the interpretability of variance-based methods. Therefore, we approximate the relative importance of parameters as causes of uncertainty using a method that is suited to our data structure and purpose. We calculate relative importance metrics using 1 million model variants for Fig. 2 and 221 PPE members for SI Fig. S1-11.

Constraint process

We identified over 450 constraint variables for consideration as potential constraints on the ΔF_{aci} component of ΔF_{aer}. This total includes monthly mean values, annual means and seasonal amplitudes of H_d and regional mean constraint variables. Gradients along transects from stratocumulus to cumulus regions were also included as constraint variables.

We previously used 'implausibility metrics' that quantify the implausibility of each model variant with reference to an observed value, accounting for emulator uncertainty, observational uncertainty, inter-annual variability and representation errors (22, 23). Implausibility metrics were calculated for one million model variants across more than 9000 distinct measurements and we used these implausibility values to rule out model variants as observationally implausible if they did not compare well to the full set of observations. In practice, observations associated with relatively large uncertainties had little-to-no impact on ruling out model variants. Using this approach, we constrained ΔF_{aer} and our parameter space, but could not readily isolate the role of individual constraint variables on the resulting ΔF_{aer} constraint and could not quantify the efficacy of total constraint in terms of improved model skill, only in terms of reduced ΔF_{aer} uncertainty range.

Here, we calculated root mean squared error (RMSE) values for every model variant in our one million member sample for each of the 450 plus constraint variables. For each constraint variable, we then normalized the one million RMSE values and ranked model variants according to their normalized RMSE (NRMSE) values, to identify which model variants we could rule out as observationally implausible. To avoid over-constraining our model, we set NRMSE values to zero where the uncertainty in our emulators was large relative to the difference between observed and emulated values. For this step, we defined the emulator uncertainty as the square root of the emulator variance at that specific combination of model parameters. In this way, individual constraints were stronger for constraint variables where parameter perturbations clearly defined the response surface of the associated statistical emulators.

We did not account for inter-annual variability because we ensured large-scale meteorological features of our model variants were very similar to observed conditions. We did not include (largely unquantified) observational errors in our constraint because we compared satellite data to model output from satellite simulators, which significantly reduced the importance of this source of uncertainty in observation to model comparisons. We also neglected the effects of representation errors (50) because they are unquantified for the satellite-derived observations used here. Instead, we compared mean values with stratocumulus-dominated regions to reduce the magnitude of these errors. Thus, observational and representation errors did not influence our method to identify which model variants to reject as implausible. Instead, we retained a proportion of model variants of the same order of magnitude as earlier constraint efforts that used constraint variables with more readily quantifiable sources of model-observation comparison uncertainty (22, 23). In this way, our method avoided over-constrain the model, yet allowed us to identify model structural inconsistencies without the masking effects of additional uncertainties.

For each of the 450 individual constraint variables we retained the 5000 model variants (0.5% of our original sample) with the lowest NRMSE values. However, the number of variants retained was larger than 5000 in many cases where the standard deviation from the associated emulator is larger than the difference in observed and emulated values (NRMSEs set to zero) for multiple model variants. For combinations of constraint variables, we calculated the average NRMSE value across all variables, for each model variant, prior to ranking and rejecting model variants with the highest average NRMSE values across variables. The number of constraint variables needed to optimally constrain ΔF_{aer} in our structurally imperfect model was affected by the number of model variants retained (SI Fig. S25 and table S4) because reducing the efficacy of individual constraint variables affects the potential for additional observations to further reduce the ΔF_{aci} uncertainty. However, the strength of constraint (quantified as a reduction in the 90% credible interval) was largely unaffected by the number of model variants retained at each step. The constraint was improved by only around 4 percentage points when we significantly increased the constraint criteria to retain 1k variants at each step (rather than 5k, as in the main article), and decreased by only around 3 percentage points when we significantly relaxed our constraint criteria to retain 20k variants. The constrained ΔF_{aer} bounds were largely unaffected by the number of variants retained, shifting by only around 0.1 W m^{-2} (table S4). Thus, the choice of retaining 5k model variants at each step was arbitrary and did not affect our interpretation of results.

We removed constraint variables from our constraint process where the associated emulator average standard deviation across our sample of points was larger than the standard deviation of emulated values. That is, we discounted constraint variables where the emulator uncertainty was larger than the changes in the emulated response surface. This was the case for a small number of transect constraint variables and for the seasonal amplitude of f_c in the Southern Ocean. Additionally, we removed transect measurements from our set of constraint variables where the observed values were outside of the 90% credible interval of corresponding values in our sample, since such discrepancies are indicative of structural model inadequacies and/or unaccounted for observational errors (SI Fig. S27-30).

In each region we identified a subset of constraint variables as being pairwise consistent with N_d. Individual monthly mean N_d values in each region were used to identify which other constraint variables could be considered pairwise consistent. The months used were September, October, December, March and the annual mean for the North Atlantic, North Pacific, South Atlantic, South Pacific and Southern Ocean respectively. In these months, N_d was determined to be most consistent on average with N_d in other months, as represented by the effect on average NRMSE in the associated constraint (Fig. 4 and SI Fig. S18-21). At this stage, we assumed constraint variables that are consistent with N_d in these specific months in these regions were also consistent with N_d (and other selected constraint variables) in other regions. Our strategy here was to rule out constraint variables that are clearly inconsistent, rather than to assure internal consistency between all remaining constraint variables. Across all regions, we retained 225 constraint variables (out of more than 450) which we considered consistent with N_d.

We identified an optimal set of constraint variables by first identifying the individual constraint variable with the greatest impact on ΔF_{aci} uncertainty (our target model variable), then progressively added constraint variables that most improved the overall constraint. We continued to add constraint variables to the optimal set, that weakened the ΔF_{aci} constraint the least, in case our constraint was a local maximum. The effects on ΔF_{aci} uncertainty are shown in Fig. 5 of the main article. At each of the more than twenty thousand steps in this process, we evaluated the average NRMSE values for each of the one million model variants, for every possible additional constraint. The blue and purple lines in Fig. 5 are synthetic examples of how our constraint may be improved with fewer, or no, remaining structural model inadequacies. These values used to create these lines are chosen to exemplify our point and do not correspond to actual constraints of our model.

The order these constraint variables were chosen may affect the outcome. That is, a stronger constraint may have been achieved using a different set of 'optimal' constraint variables. However, we could not calculate NRMSE values for one million model variants across all possible combinations of 225 consistent constraint variables. Instead, we tested the effect of starting with all 225 consistent constraint variables and progressively removing one variable at a time. This is the most distinct test of reordering the constraint variables, from the method we used in the main article. This approach yielded a similar constraint on ΔF_{aer} as achieved by progressively adding constraint variables (90% CI between -1.4 and -0.2 W m^{-2}, or -1.2 to -0.0 W m^{-2} depending on which local maxima is used) and very similar constraints on marginal parameter distributions (equivalent to SI Fig. S23, 24). These tests revealed there are multiple ways to combine sets of consistent constraint variables to achieve a similar constraint on ΔF_{aer}, highlighting the degree of redundancy in using multiple observations of the same variable for constraint.

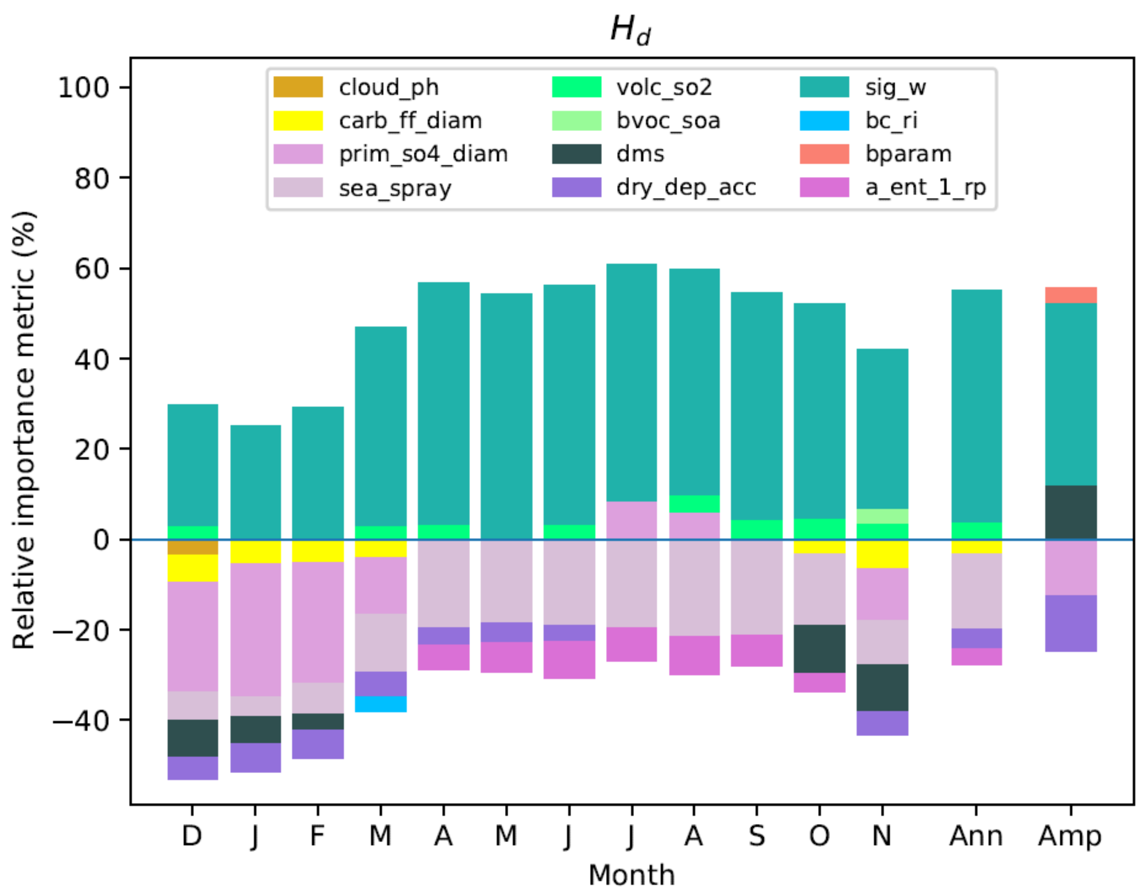
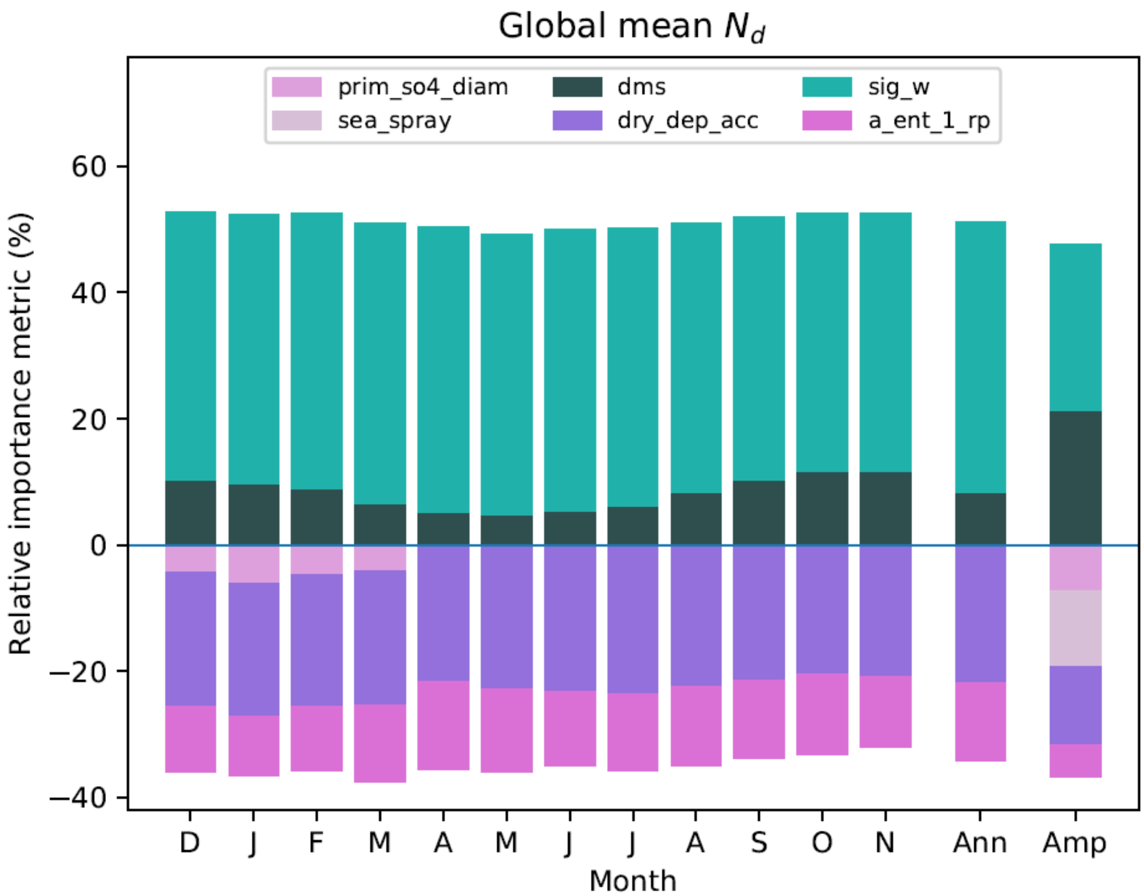


Fig. S1. Relative importance of model parameters as causes of uncertainty in H_d. Relative importance metrics are calculated for each month (December 2016 to November 2017), for the annual mean (Ann) and the seasonal amplitude (Amp). Relative importance metrics lower than 4% are not shown.



335

Fig. S2. Relative importance of model parameters as causes of uncertainty in global mean N_d. Figure features are identical to Fig. S1.

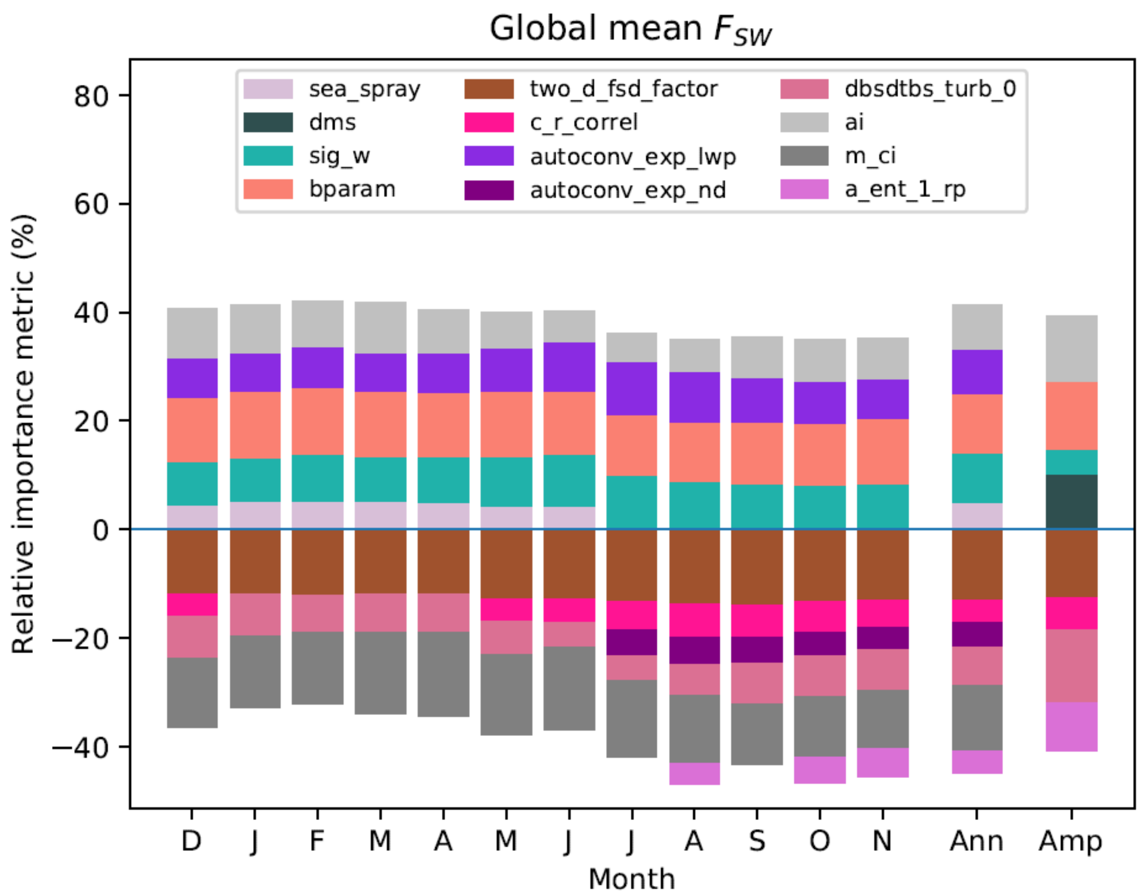


Fig. S3. Relative importance of model parameters as causes of uncertainty in global mean F_{SW}.
340 Figure features are identical to Fig. S1.

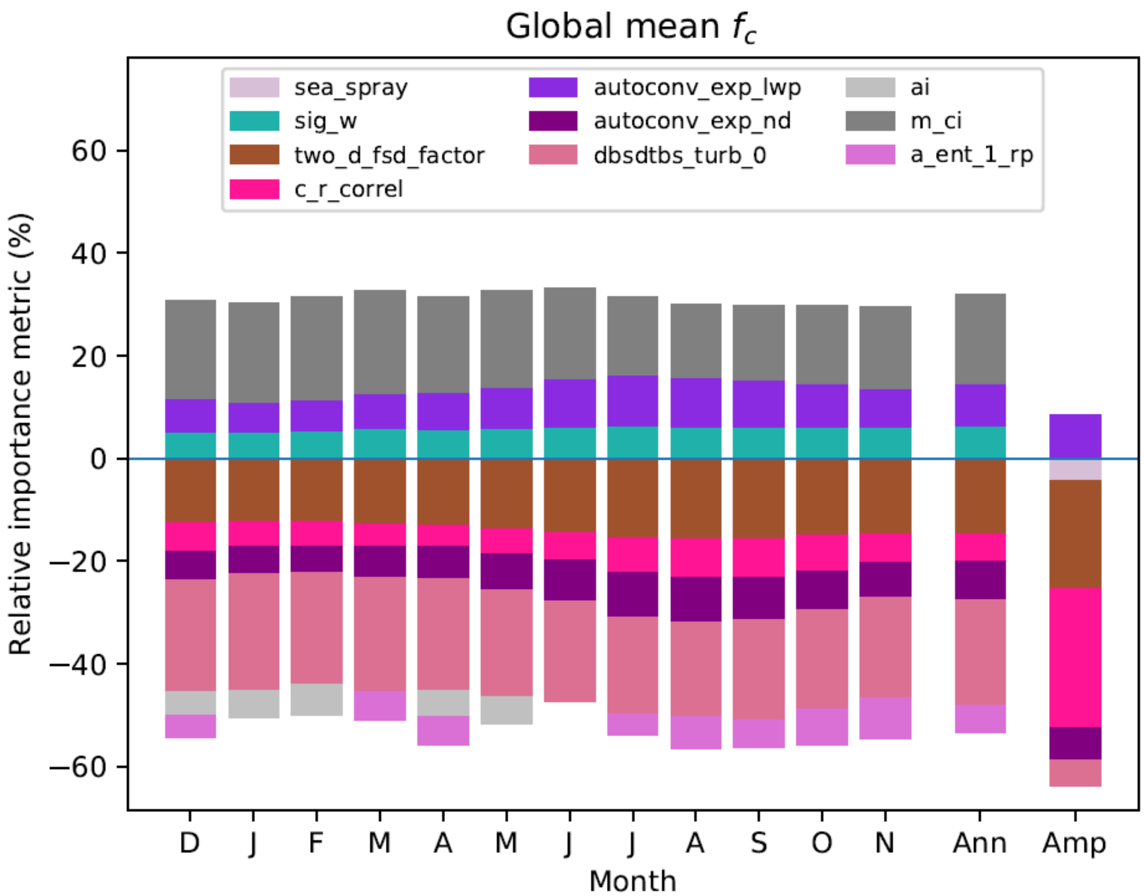


Fig. S4. Relative importance of model parameters as causes of uncertainty in global mean f_c. Figure features are identical to Fig. S1.

345

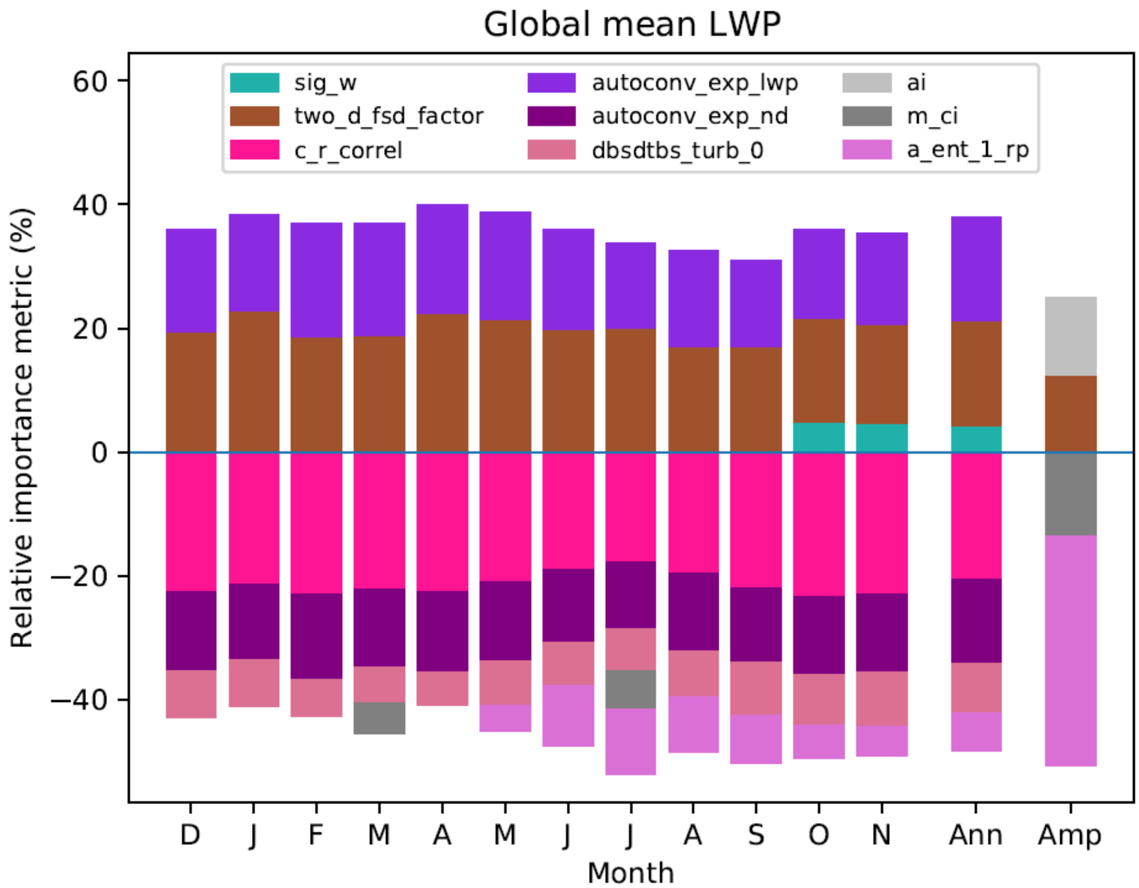
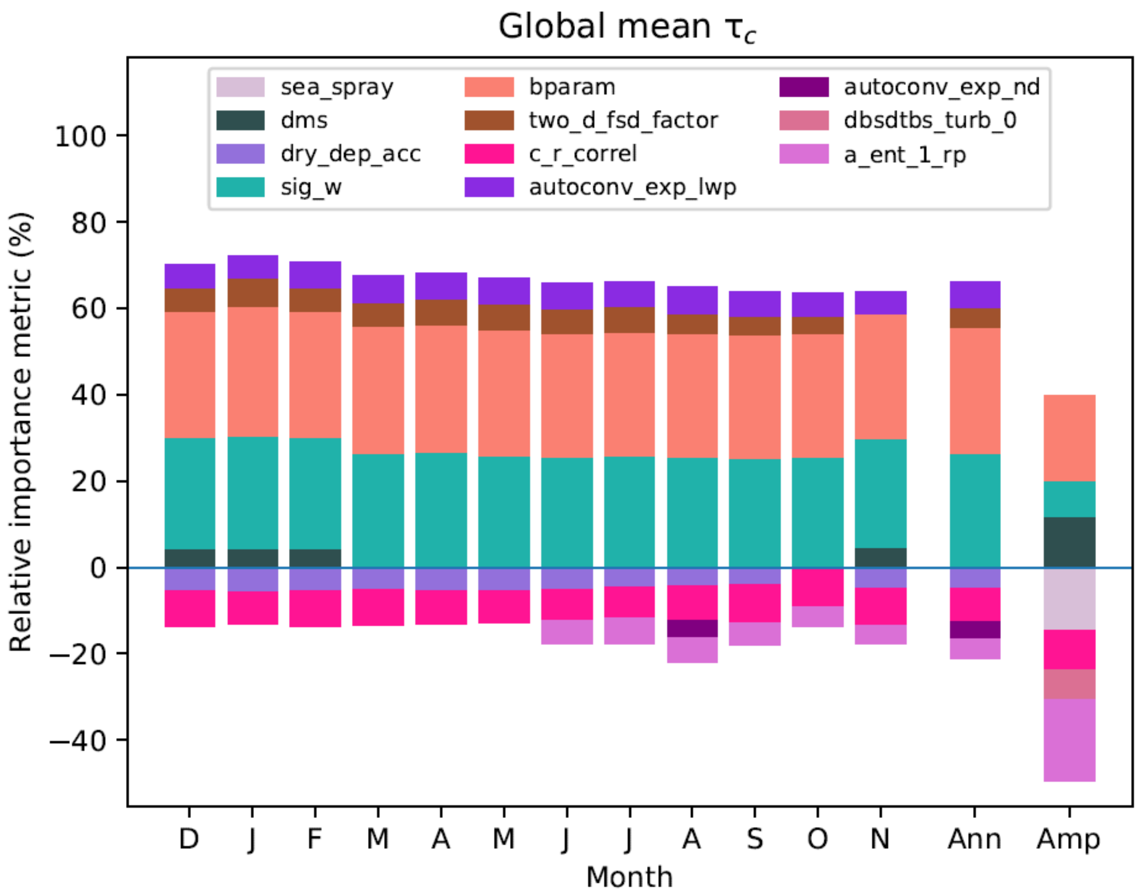
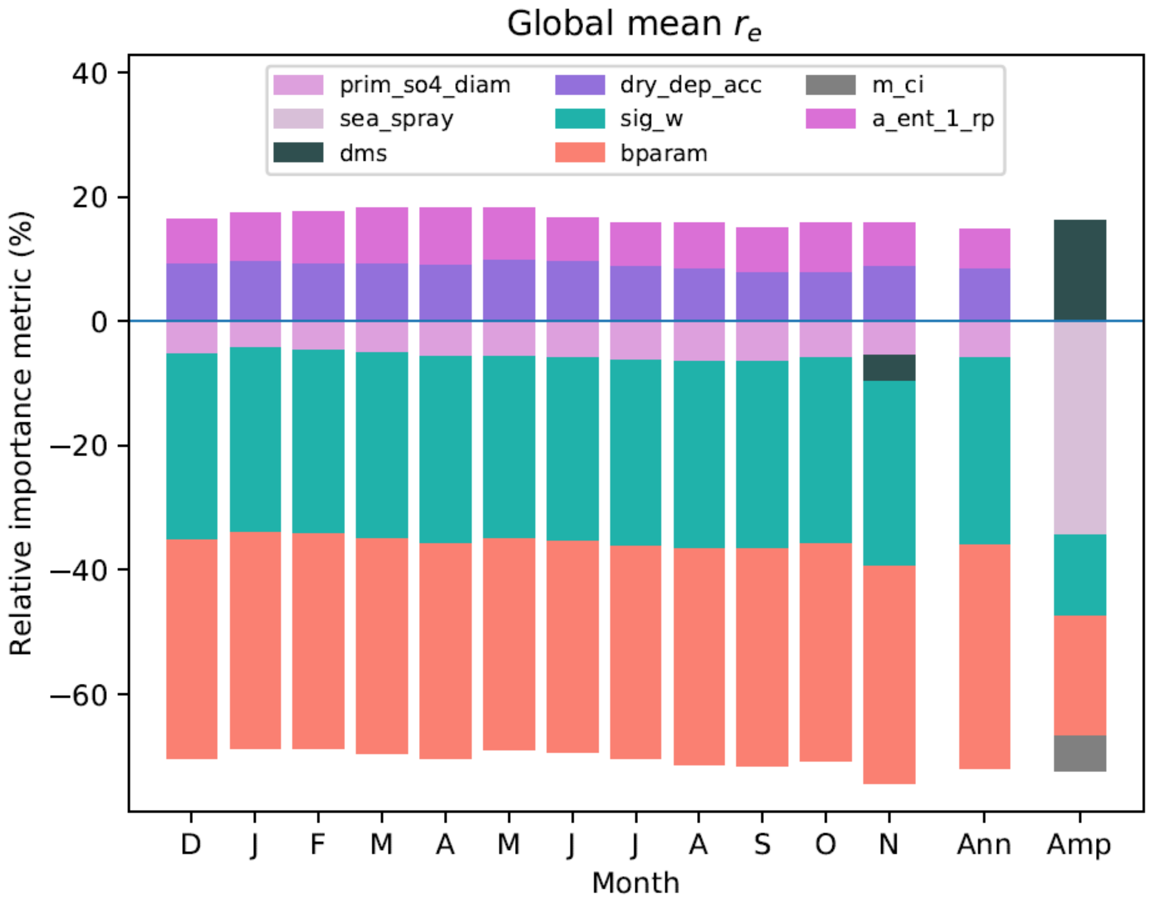


Fig. S5. Relative importance of model parameters as causes of uncertainty in global mean LWP. Figure features are identical to Fig. S1.



350 **Fig. S6.** Relative importance of model parameters as causes of uncertainty in global mean τ_c. Figure features are identical to Fig. S1.



355 **Fig. S7.** Relative importance of model parameters as causes of uncertainty in global mean r_e. Figure features are identical to Fig. S1.

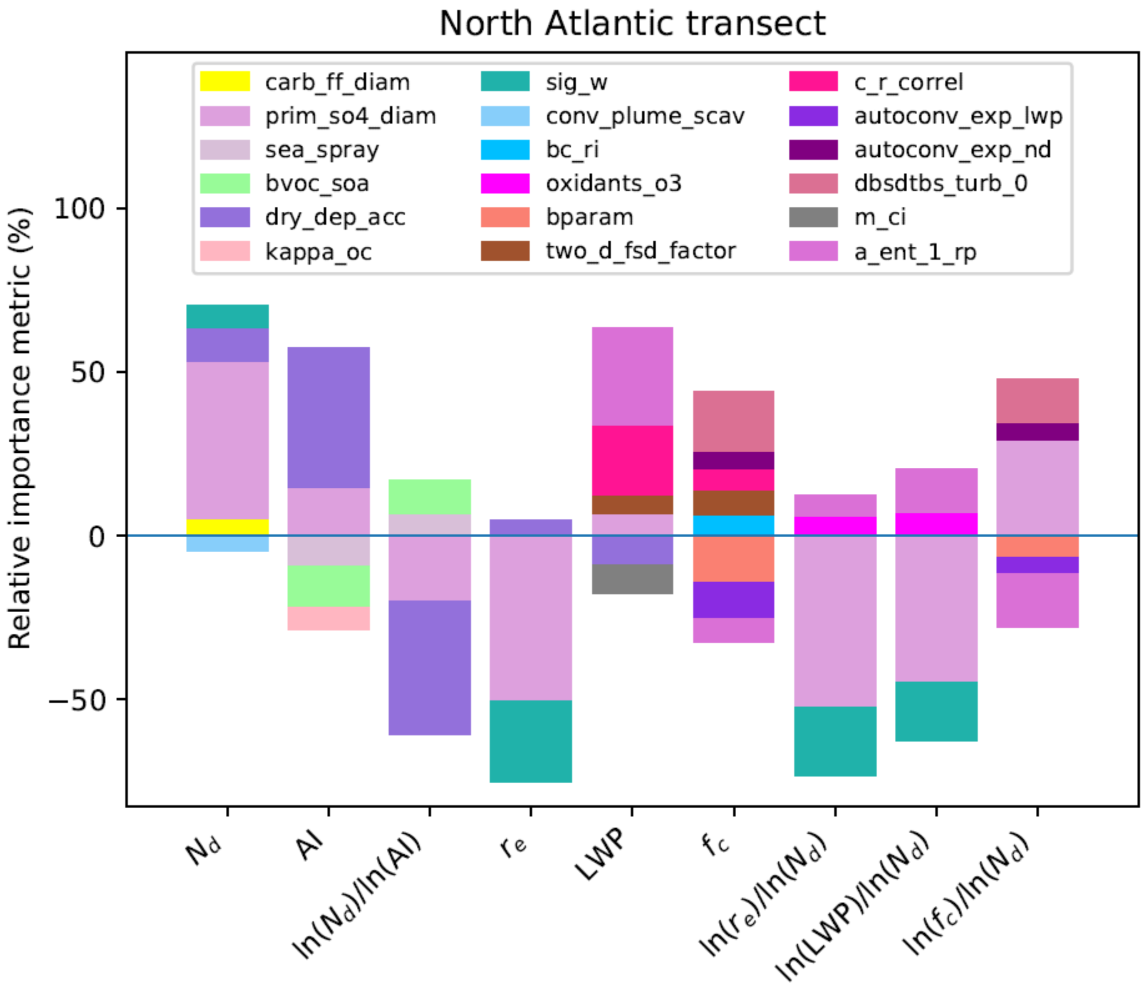


Fig. S8. Relative importance of model parameters as causes of uncertainty in North Atlantic
360 transect constraint variables. Figure features are identical to Fig. S1.

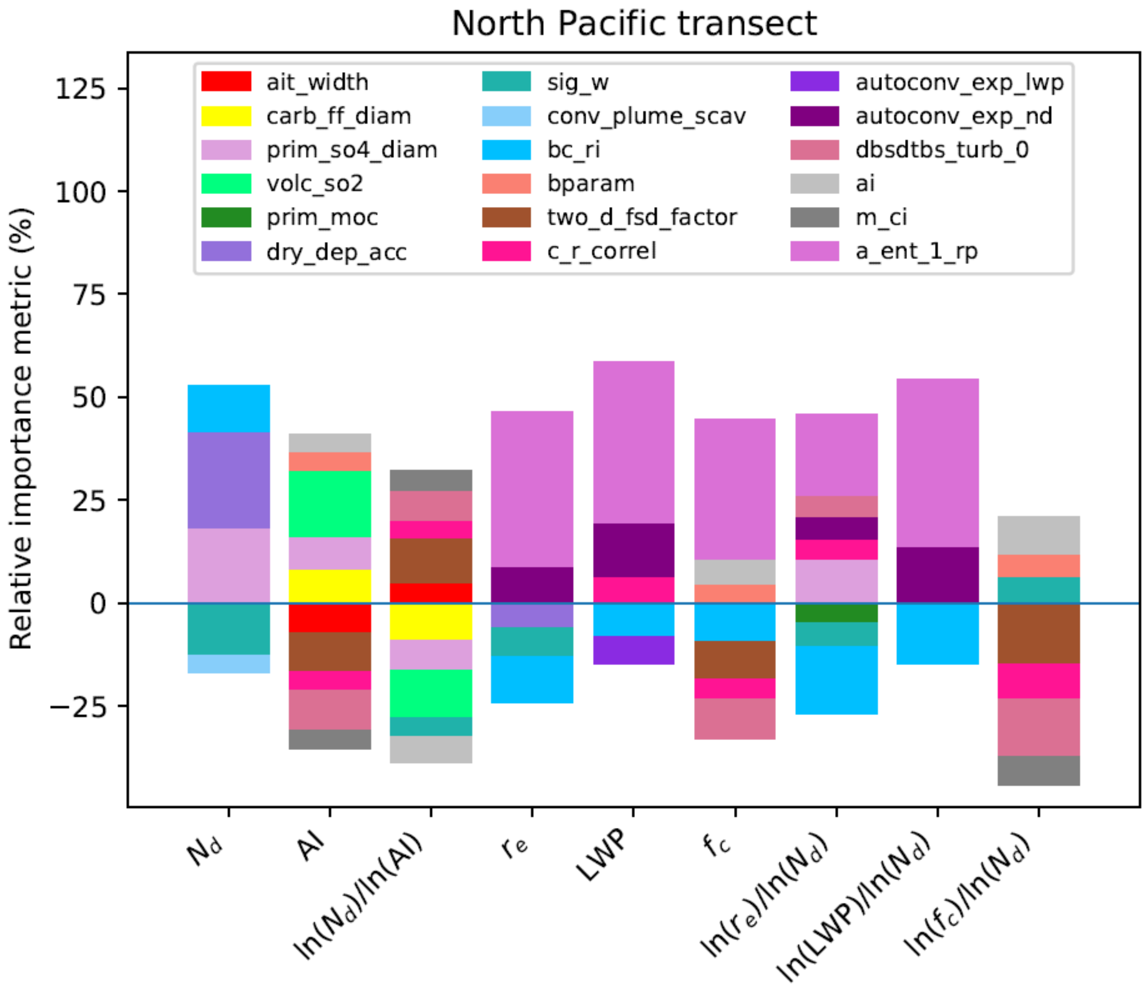


Fig. S9. Relative importance of model parameters as causes of uncertainty in North Pacific transect constraint variables. Figure features are identical to Fig. S1.

365

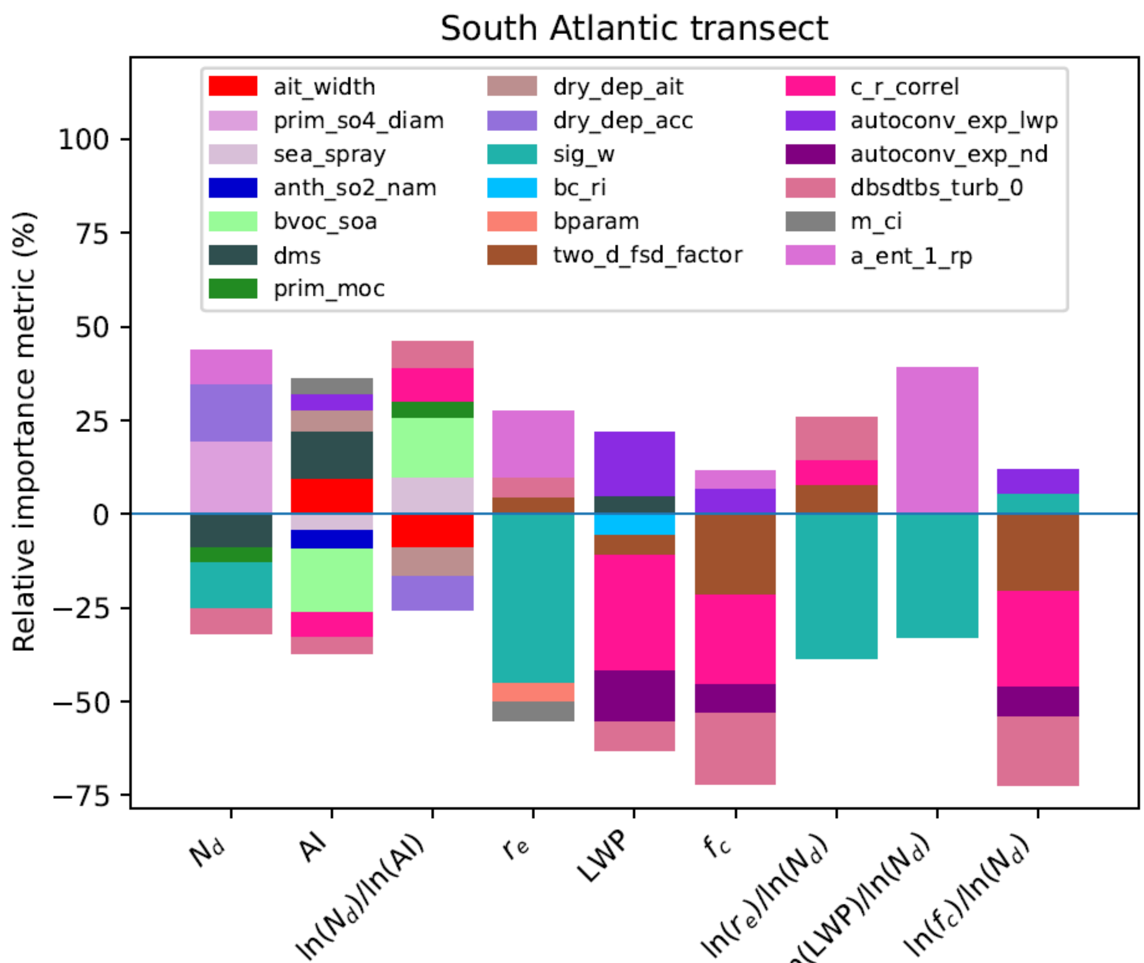
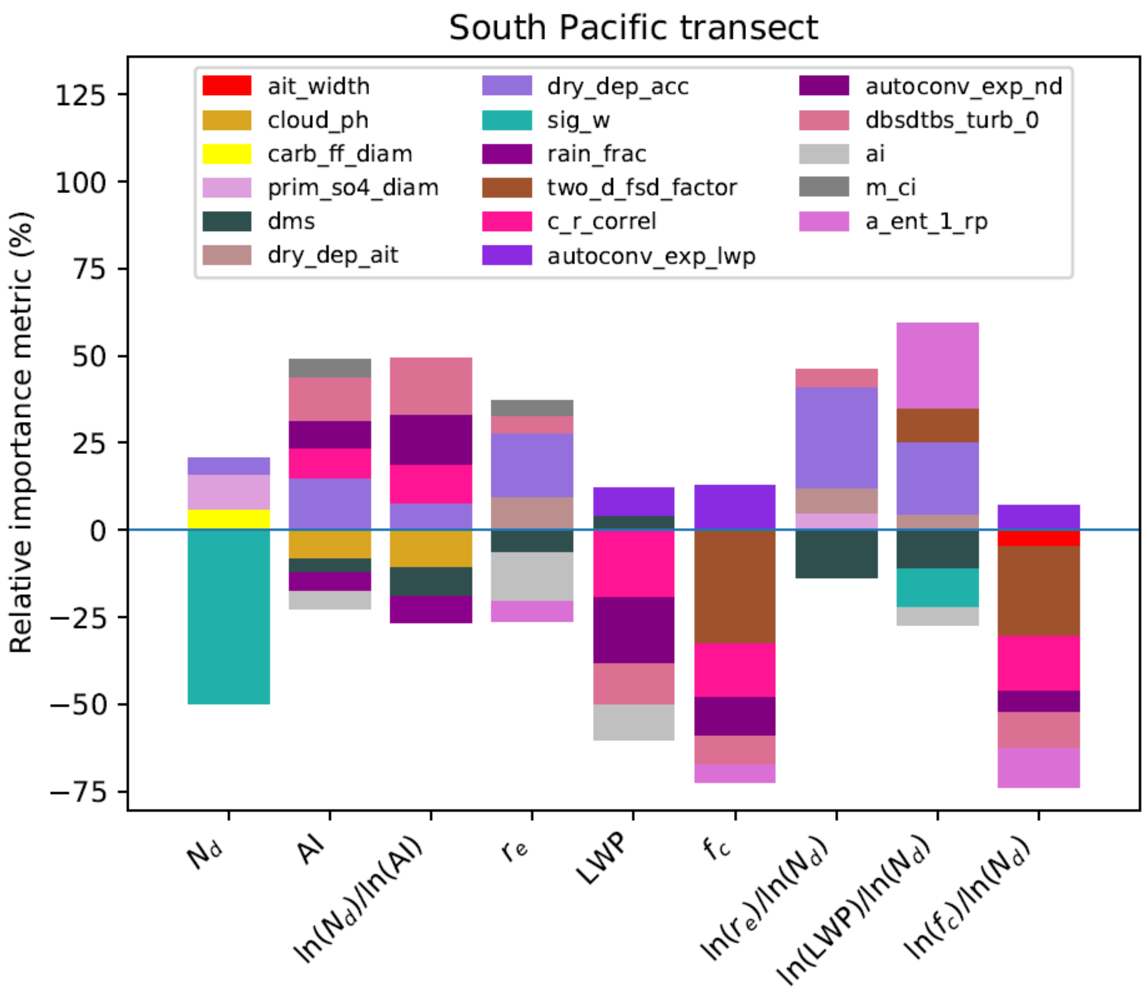
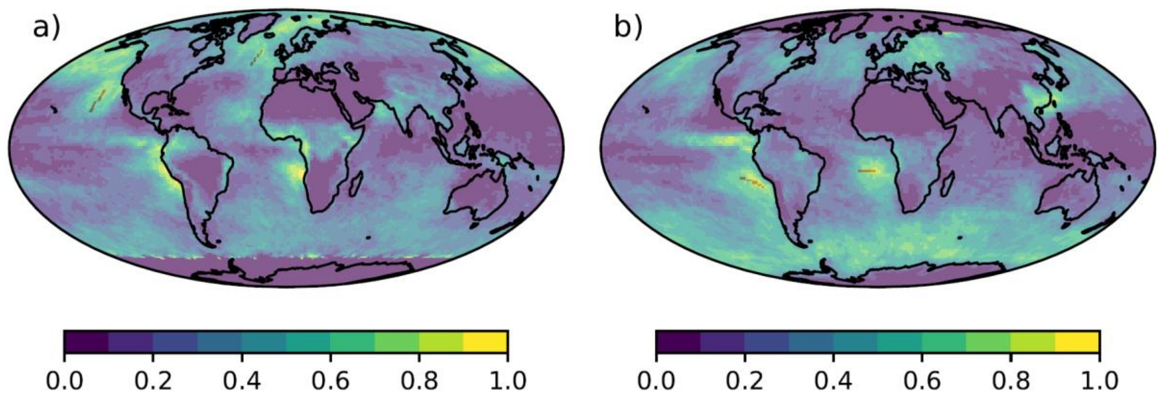


Fig. S10. Relative importance of model parameters as causes of uncertainty in South Atlantic transect constraint variables. Figure features are identical to Fig. S1.



370

Fig. S11. Relative importance of model parameters as causes of uncertainty in South Pacific transect constraint variables. Figure features are identical to Fig. S1.



375 **Fig. S12**. Transects from stratocumulus to cumulus cloud dominated regions in a) July and b) November, superimposed on MODIS liquid cloud fraction values for the corresponding month.

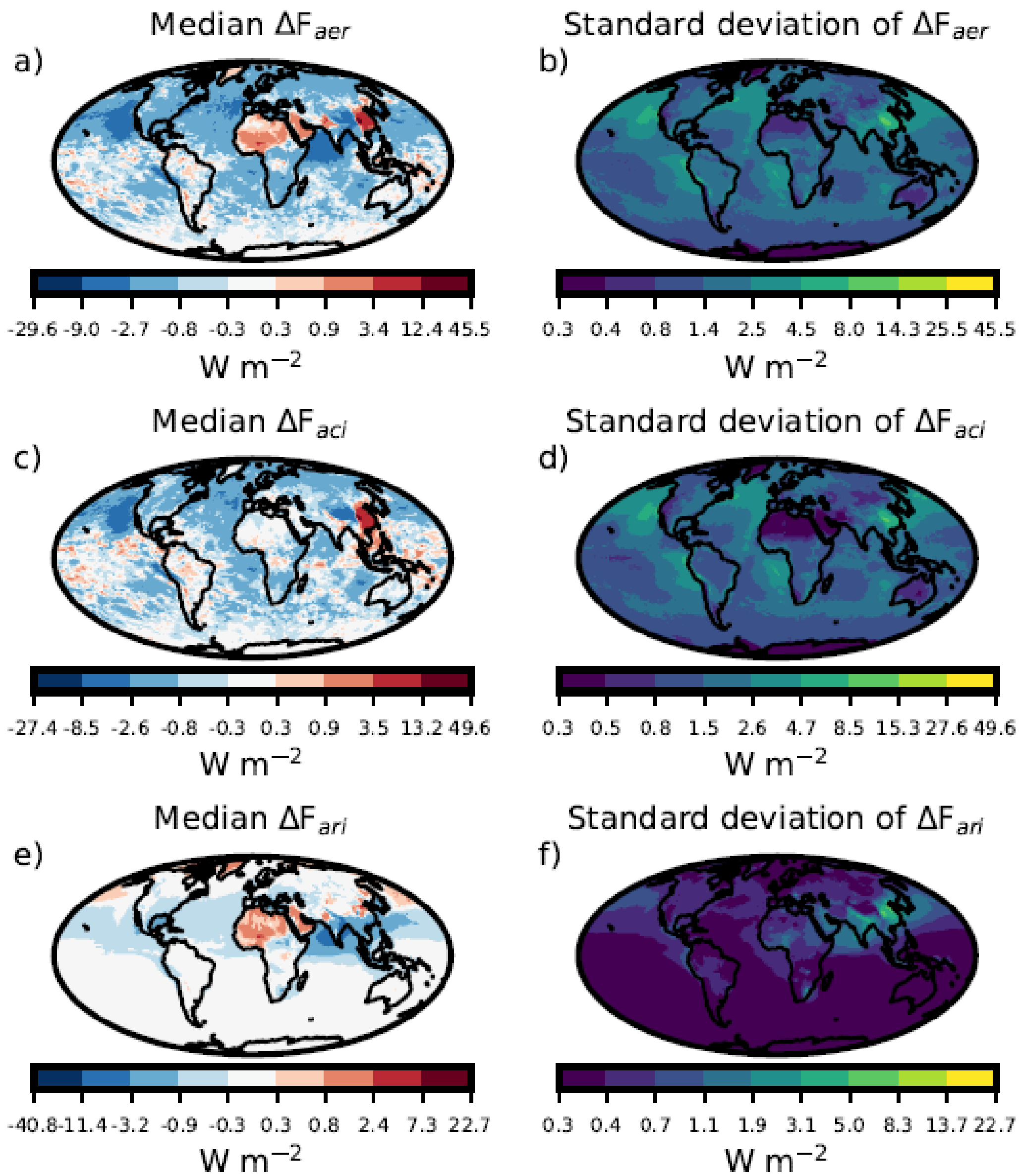
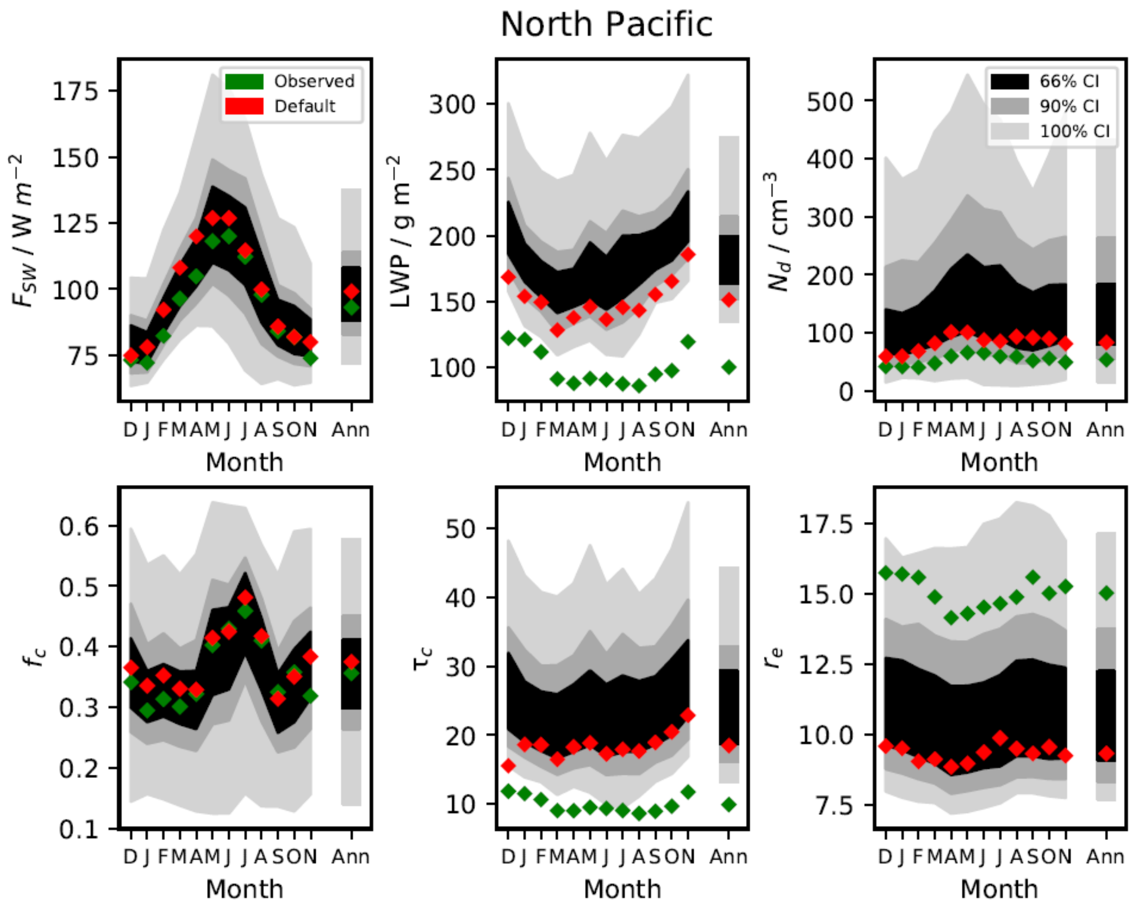


Fig. S13. Median and standard deviations of annual mean ΔF_{aer}, ΔF_{aci} and ΔF_{ari}, across the 221 PPE members. Values are calculated in each model grid box independently.



385 **Fig. S14.** Probability distributions of North Pacific regional mean output from our sample of model variants, satellite-derived measurements and the default UKESM1-A model, for individual months spanning December 2016 to November 2017 and the annual mean.

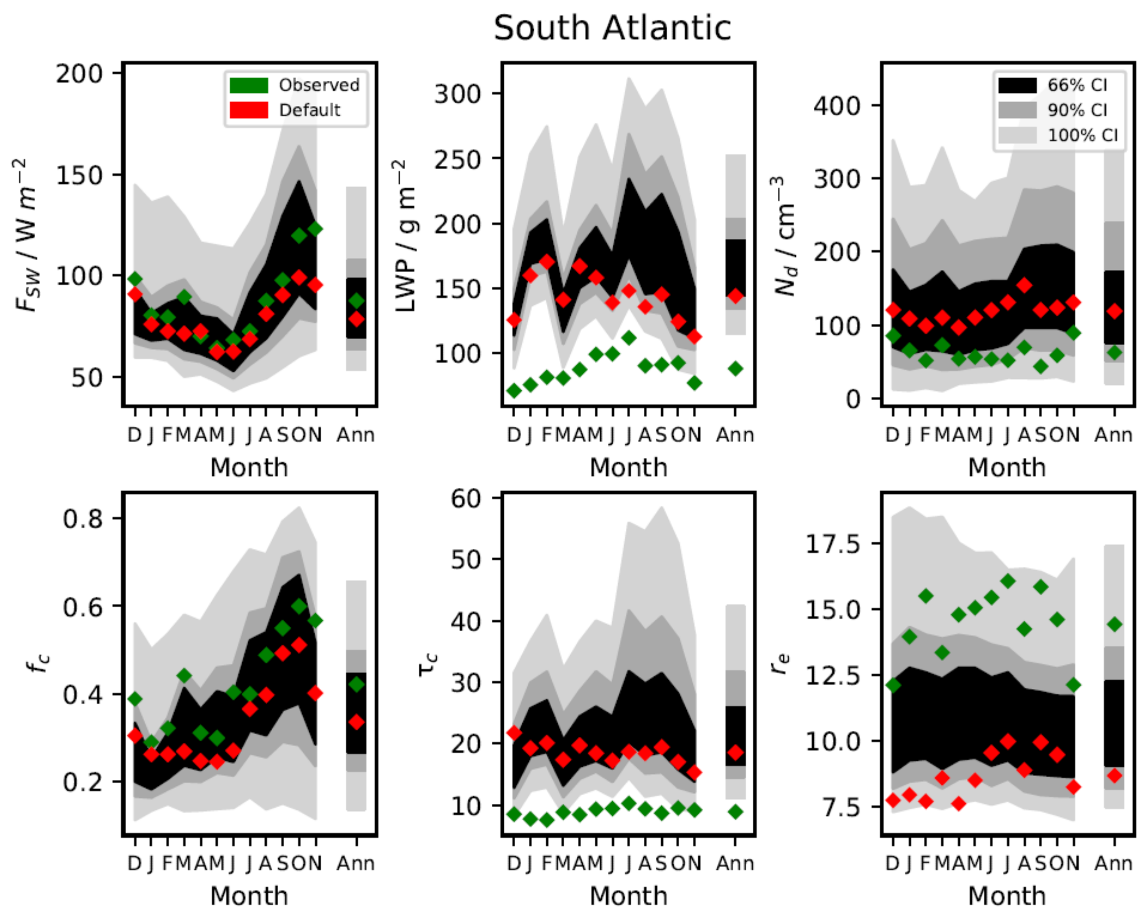
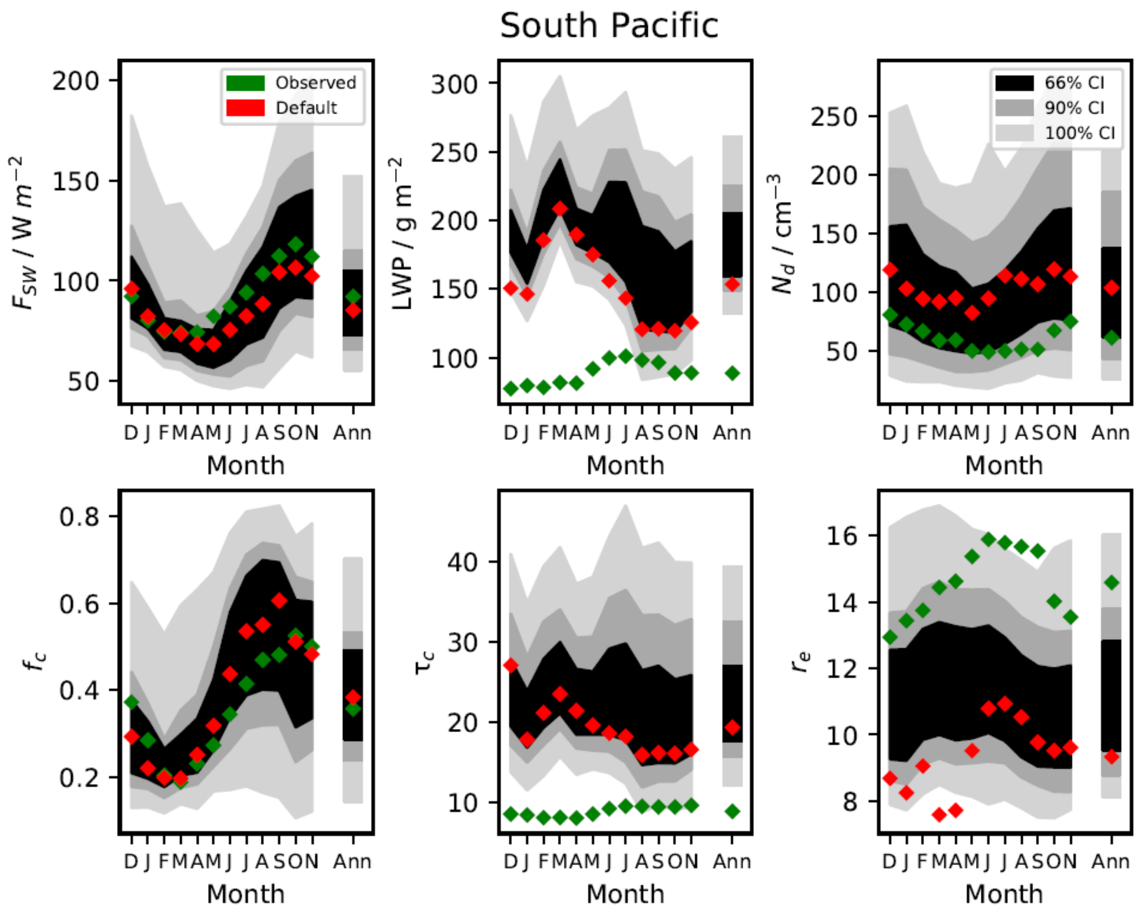


Fig. S15. Probability distributions of South Atlantic regional mean output from our sample of model variants, satellite-derived measurements and the default UKESM1-A model, for individual
390 months spanning December 2016 to November 2017 and the annual mean.



395 **Fig. S16.** Probability distributions of South Pacific regional mean output from our sample of model variants, satellite-derived measurements and the default UKESM1-A model, for individual months spanning December 2016 to November 2017 and the annual mean.

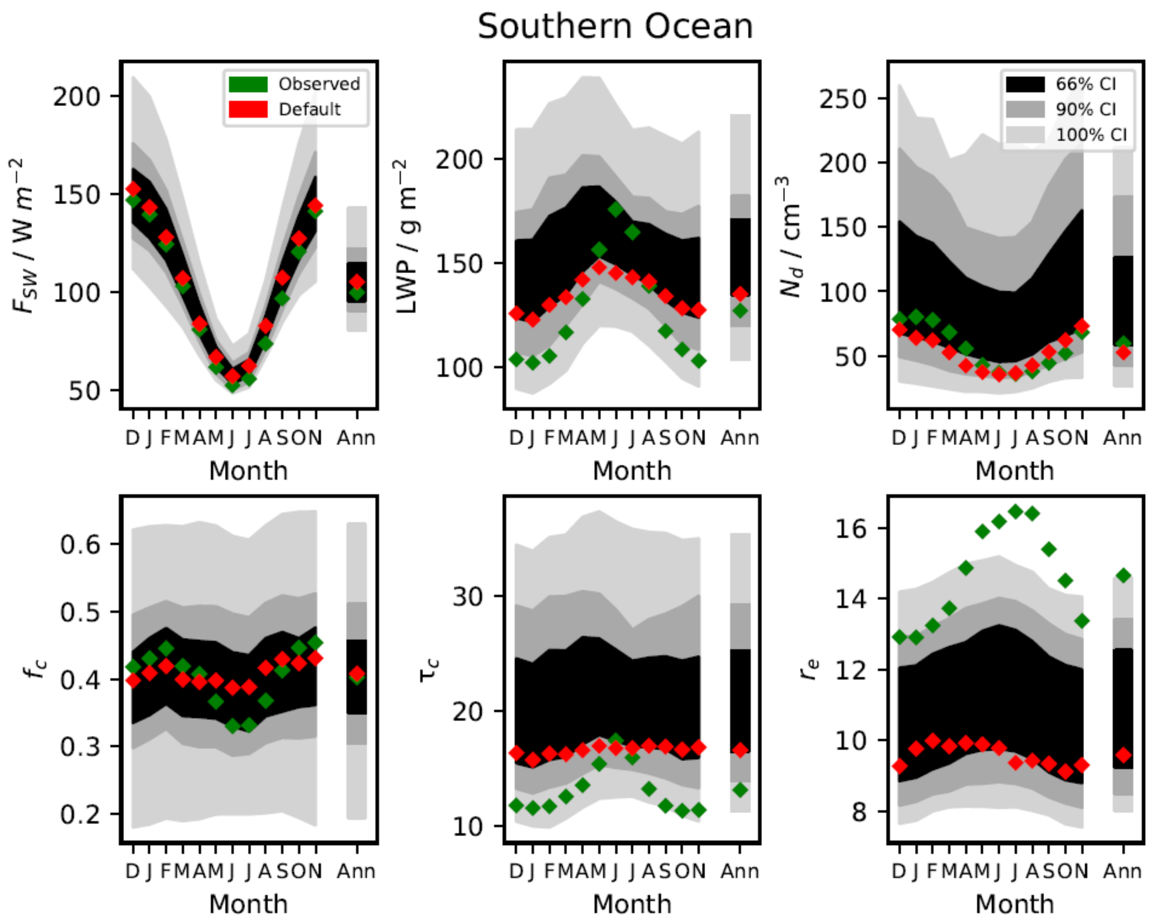


Fig. S17. Probability distributions of Southern Ocean regional mean output from our sample of model variants, satellite-derived measurements and the default UKESM1-A model, for individual
400 months spanning December 2016 to November 2017 and the annual mean.

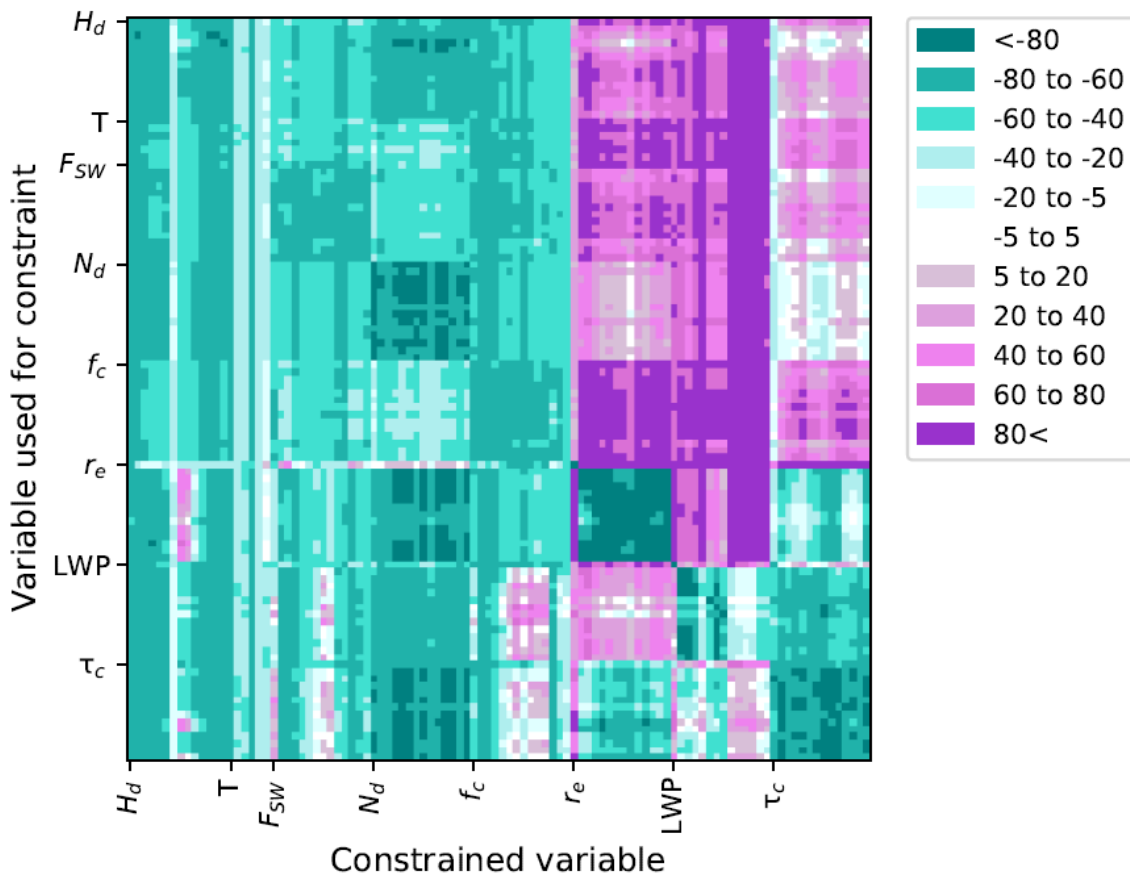
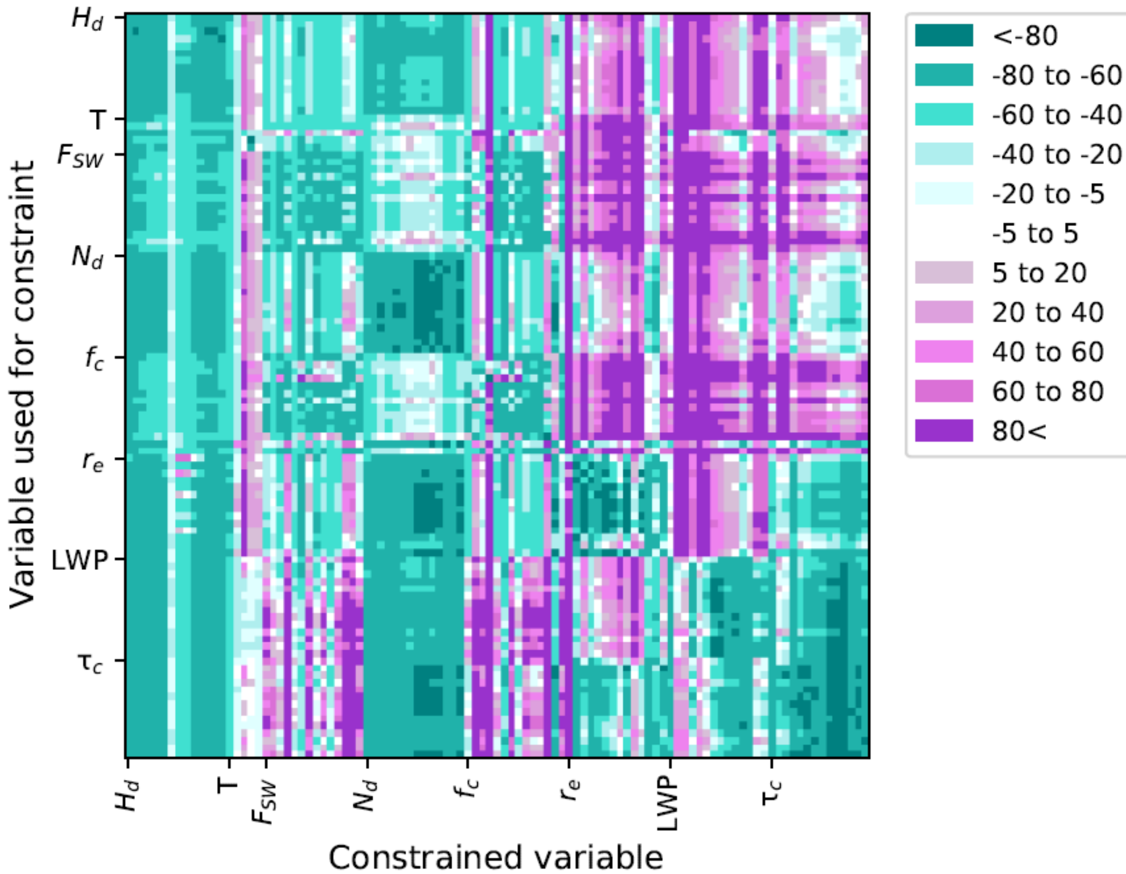
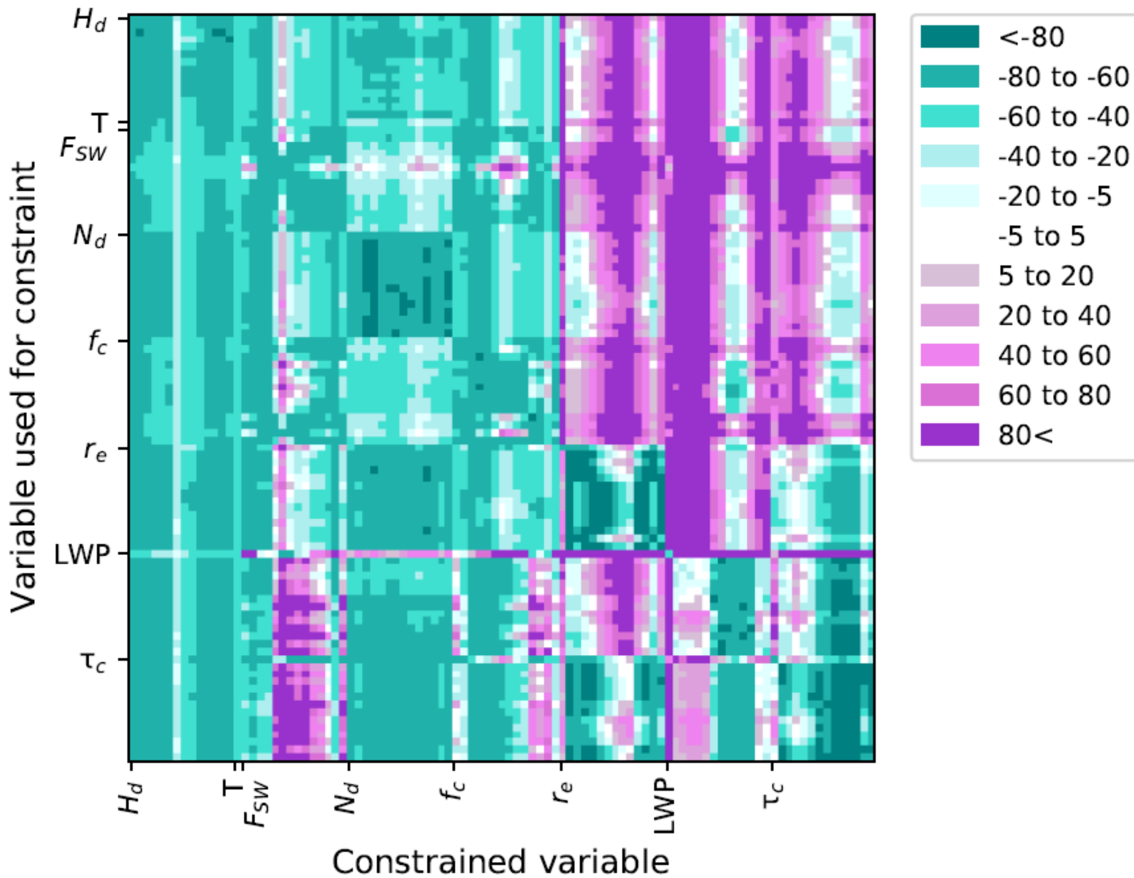


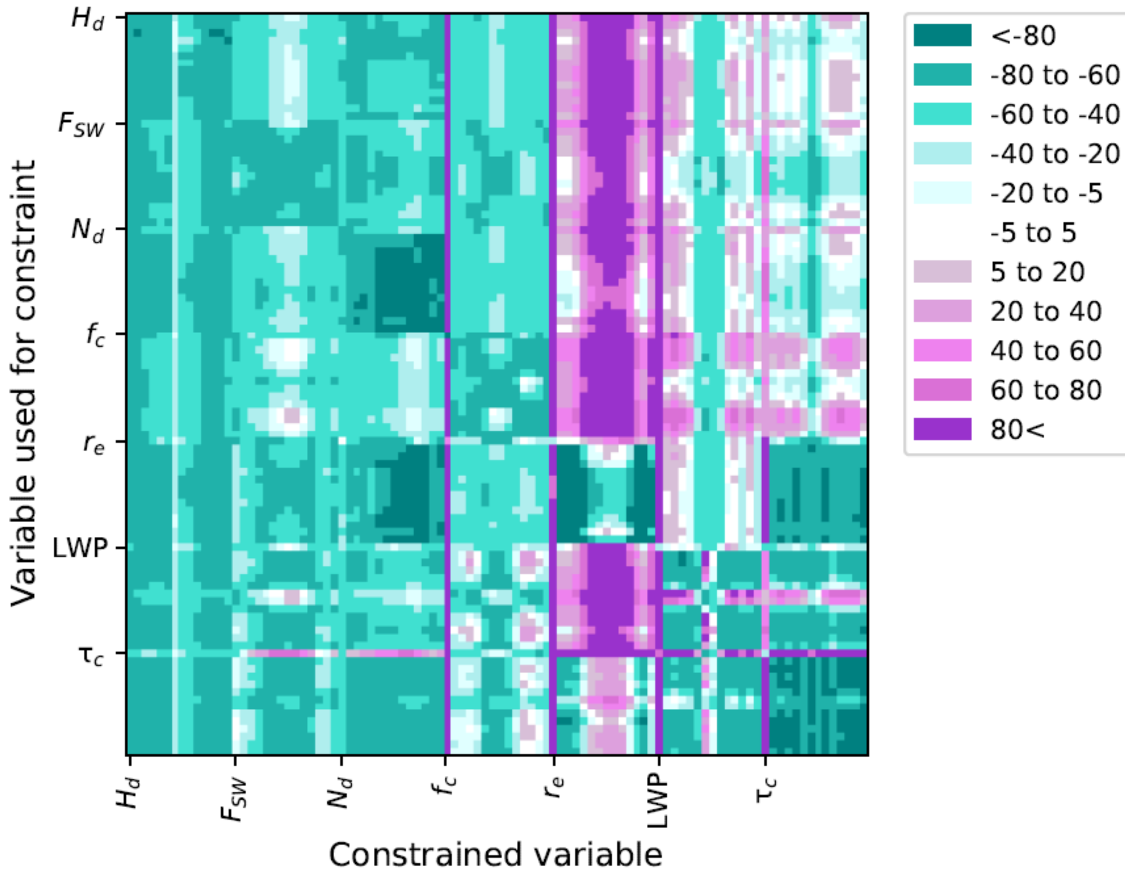
Fig. S18. Pairwise comparisons of North Pacific and H_d constraint variables. Figure features are
405 identical with Fig. 3.



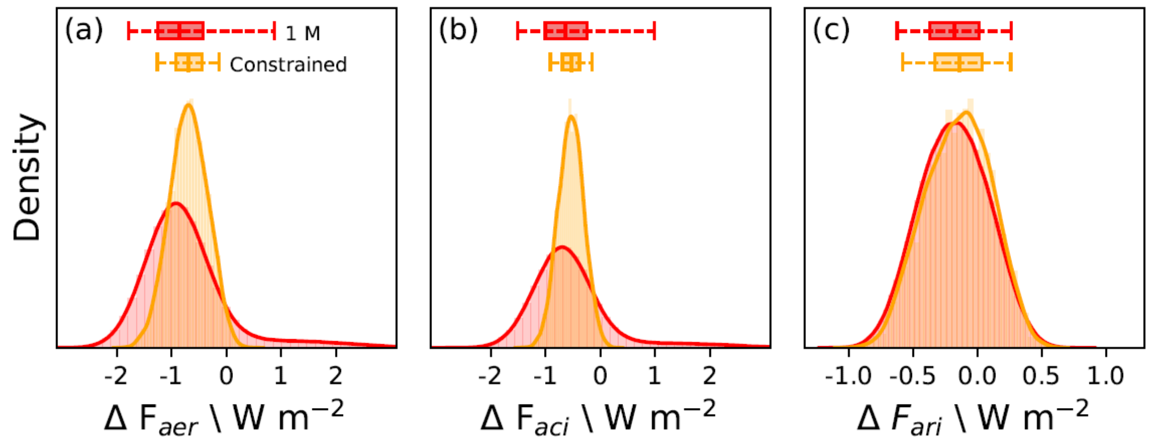
410 **Fig. S19.** Pairwise comparisons of South Atlantic and H_d constraint variables. Figure features are identical with Fig. 3.



415 **Fig. S20.** Pairwise comparisons of South Pacific and H_d constraint variables. Figure features are identical with Fig. 3.



420 **Fig. S21.** Pairwise comparisons of Southern Ocean and H_d constraint variables. Figure features are identical with Fig. 3.



425 **Fig. S22.** Probability density functions for global, annual mean a) ΔF_{aer}, b) ΔF_{aci} and c) ΔF_{ari} in the original one million member sample and after optimal constraint.

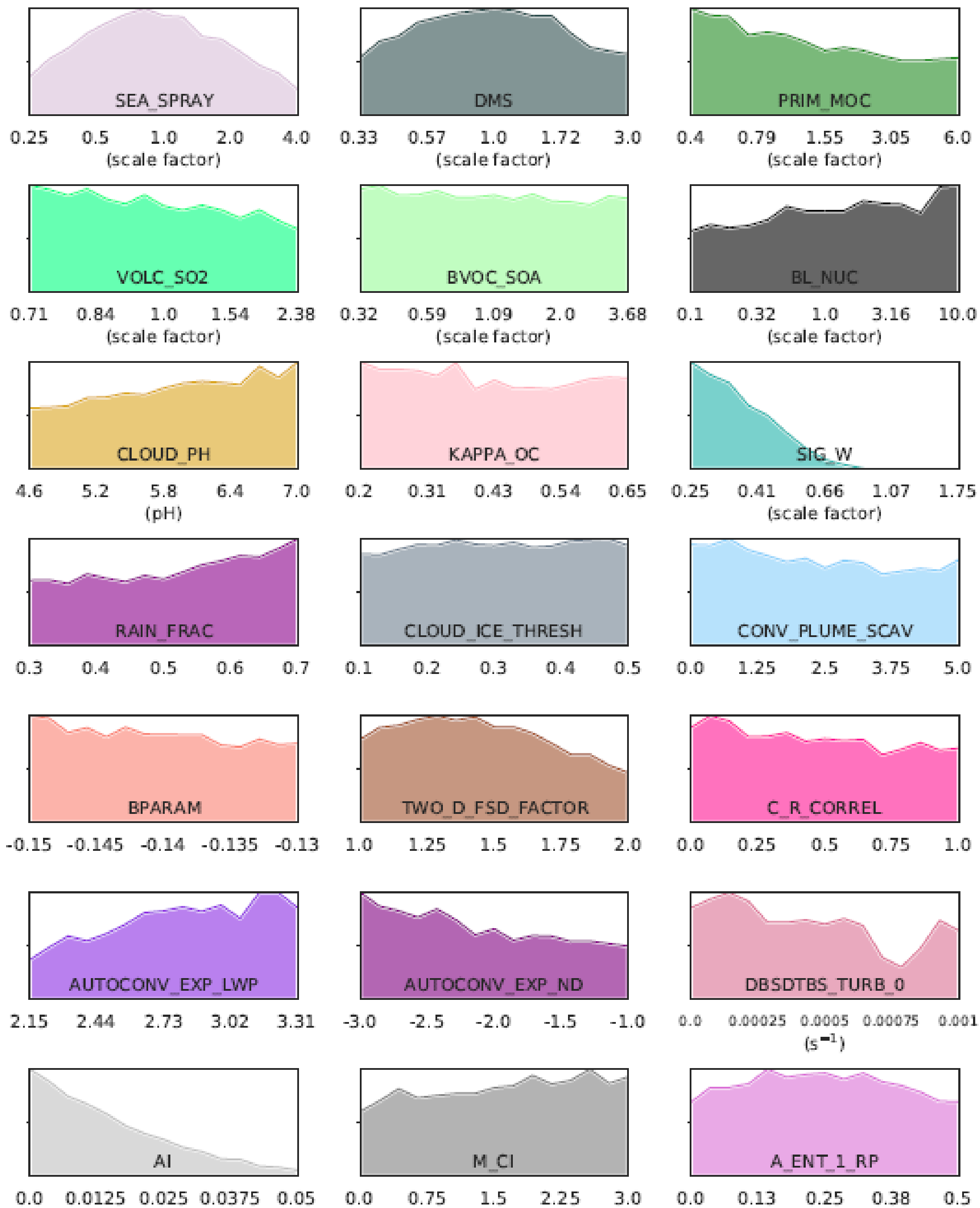
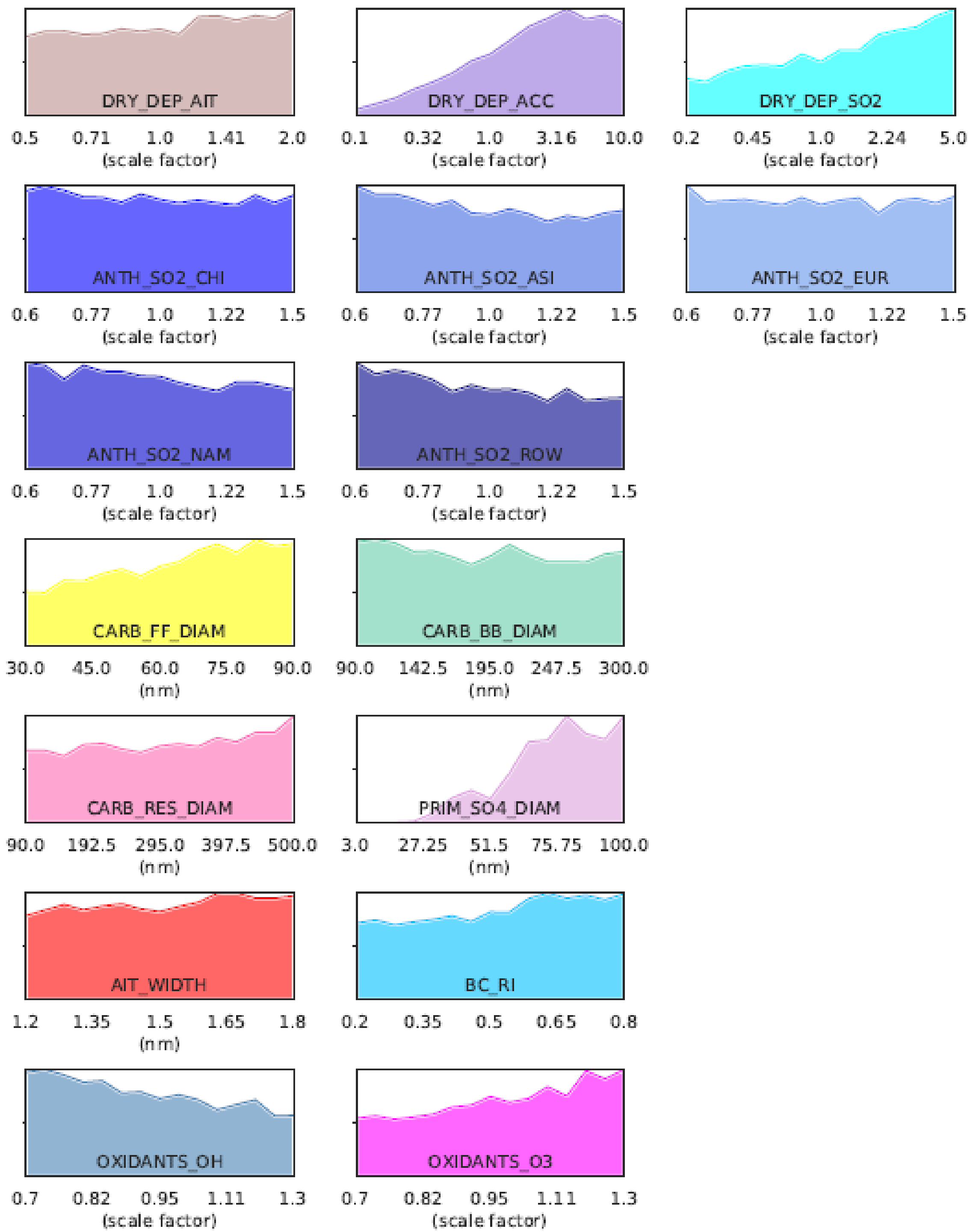


Fig. S23. Probability density functions of model parameters after constraint using our optimal set of constraint variables. In the original sample of 1 million model variants, these pdfs would be uniformly distributed on this scale. Non-shaded sections indicate a proportion of model variants with corresponding parameter values have been ruled out as implausible.



435 **Fig. S24.** Probability density functions of model parameters after constraint using our optimal set of constraint variables. Features are identical to Fig. S23.

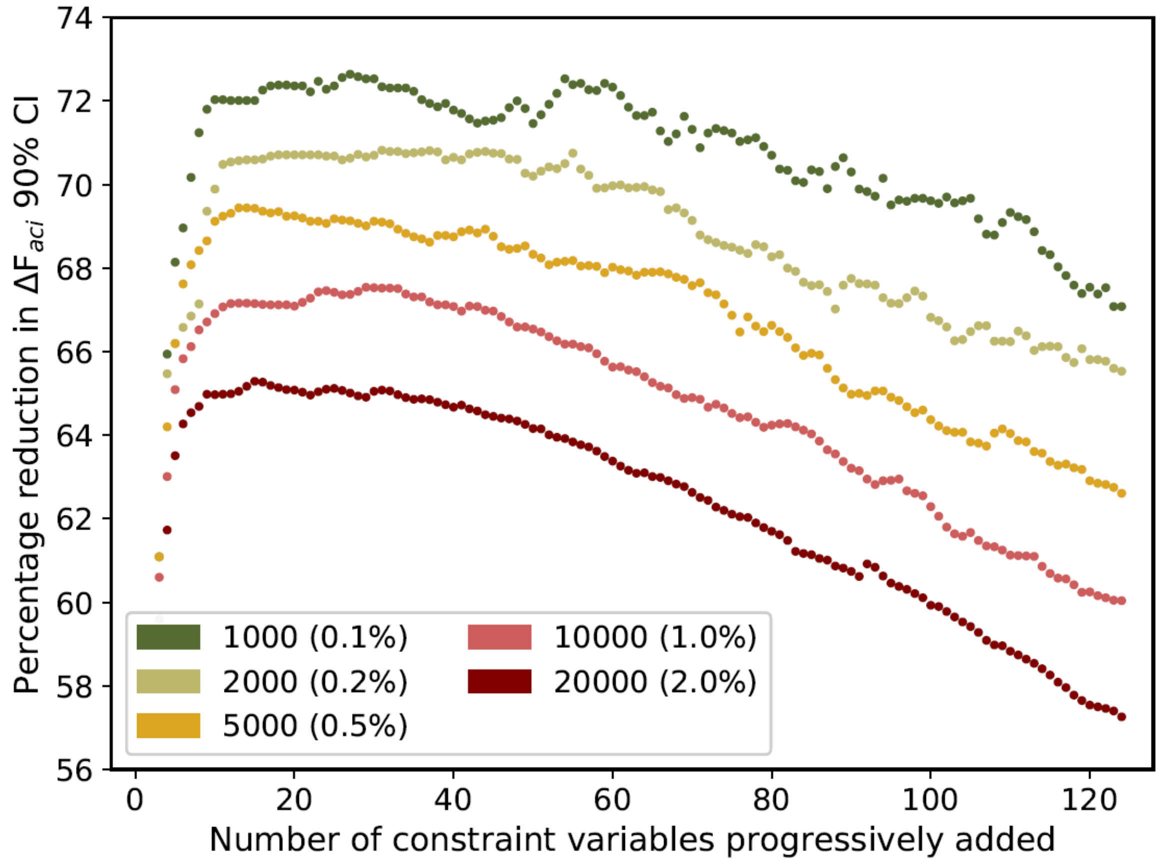
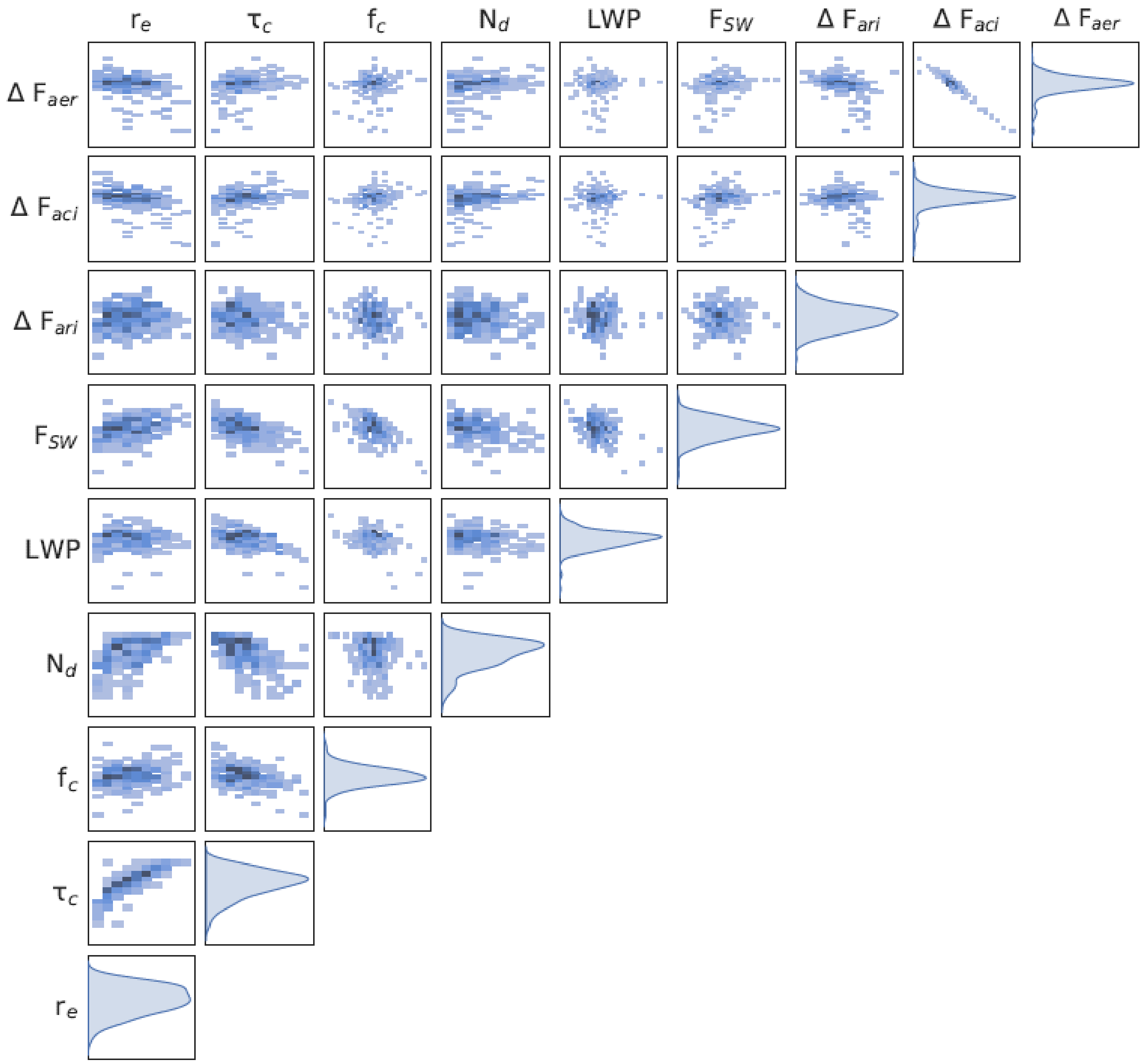
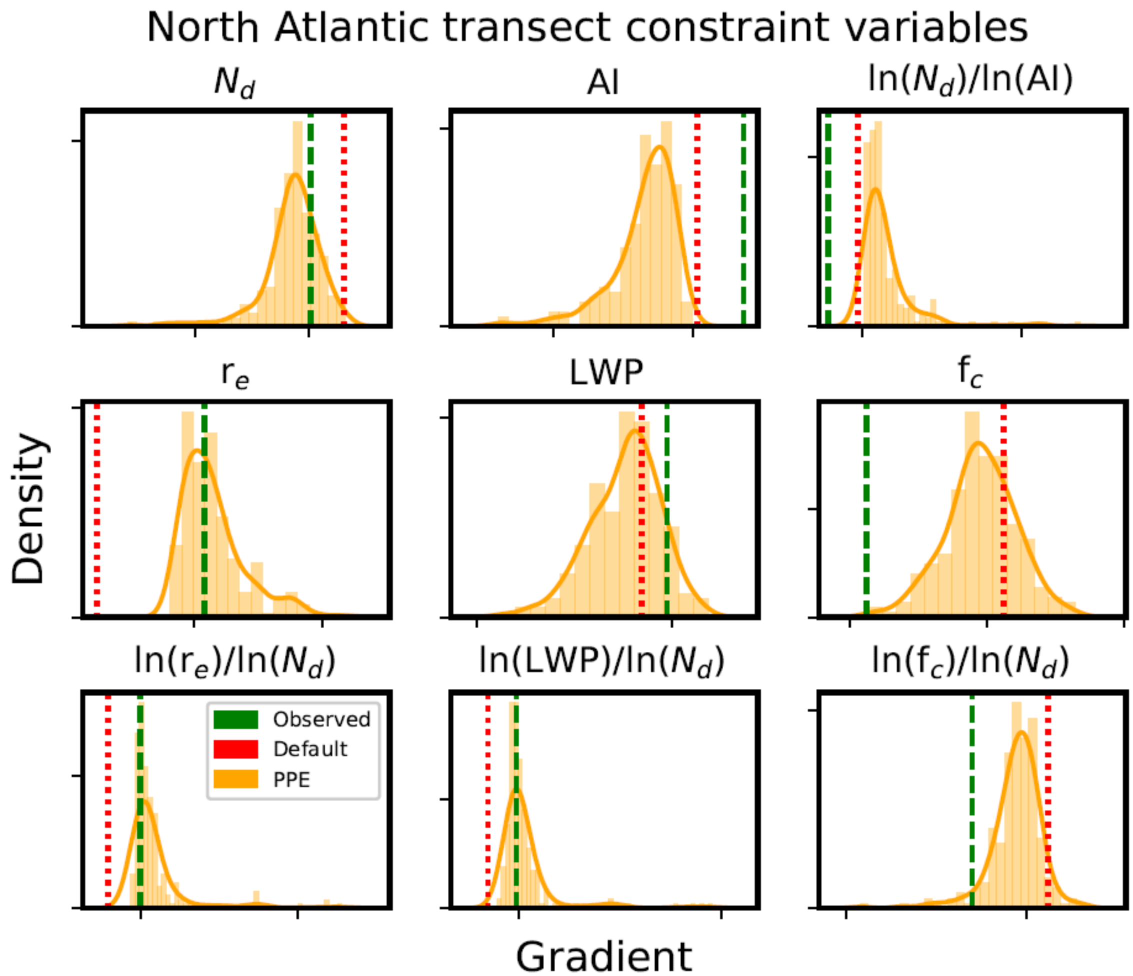


Fig. S25. Constraint of ΔF_{aci} and the effect of varying the number of constraint variables used and
440 the number of model variants retained (percentage of original 1 million) at each stage of the constraint (legend). The constraints achieved by retaining 5000 model variants at each stage is identical to the constraints shown in Fig. 5.



445 **Fig. S26.** Density plots of global, annual mean output from 221 PPE members for ΔF_{aer}, ΔF_{aci}, ΔF_{ari}, and global mean F_{SW}, LWP, N_d, f_c,τ_c and r_e,. Diagonal panels show probability density functions for individual variables.



450 **Fig. S27.** Probability density functions of North Atlantic transect constraint variables.

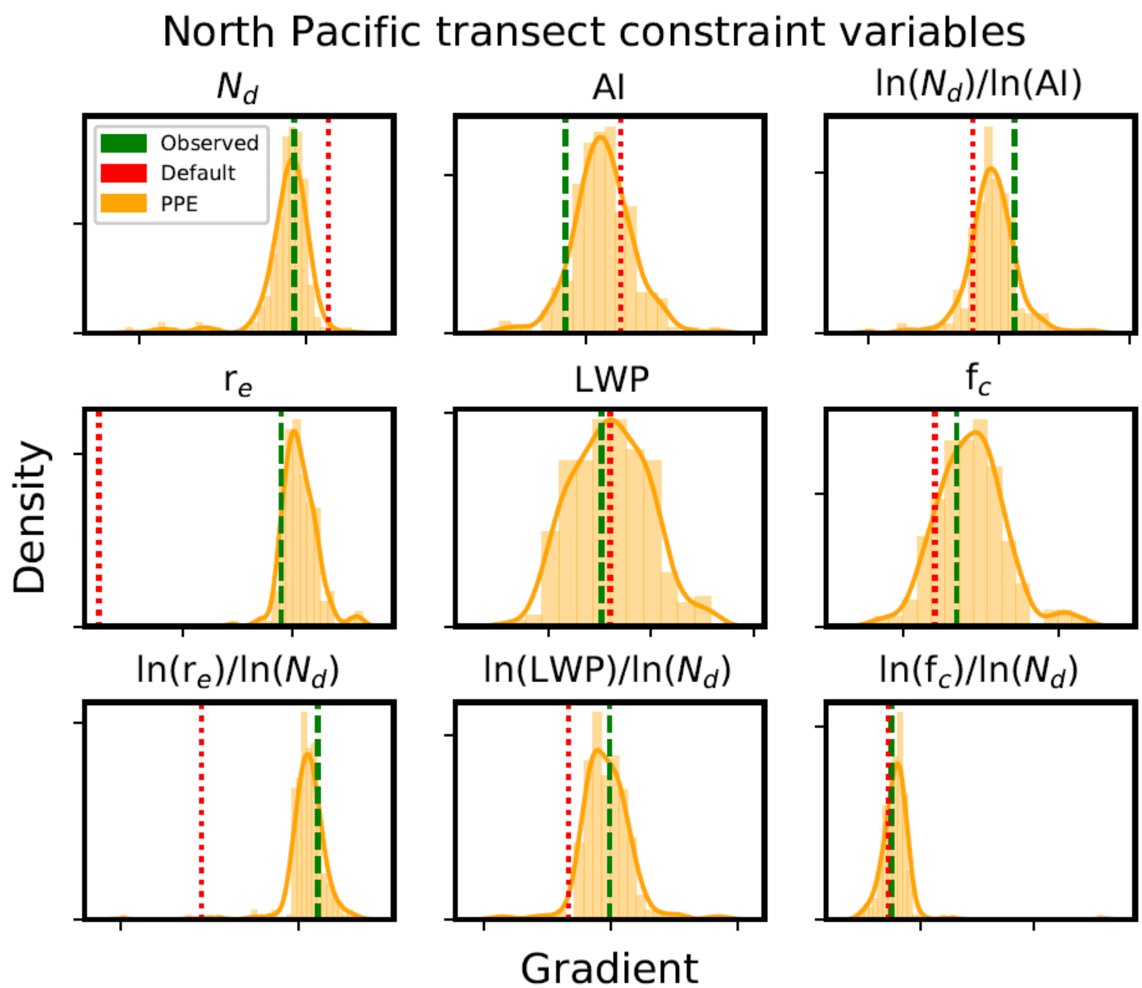
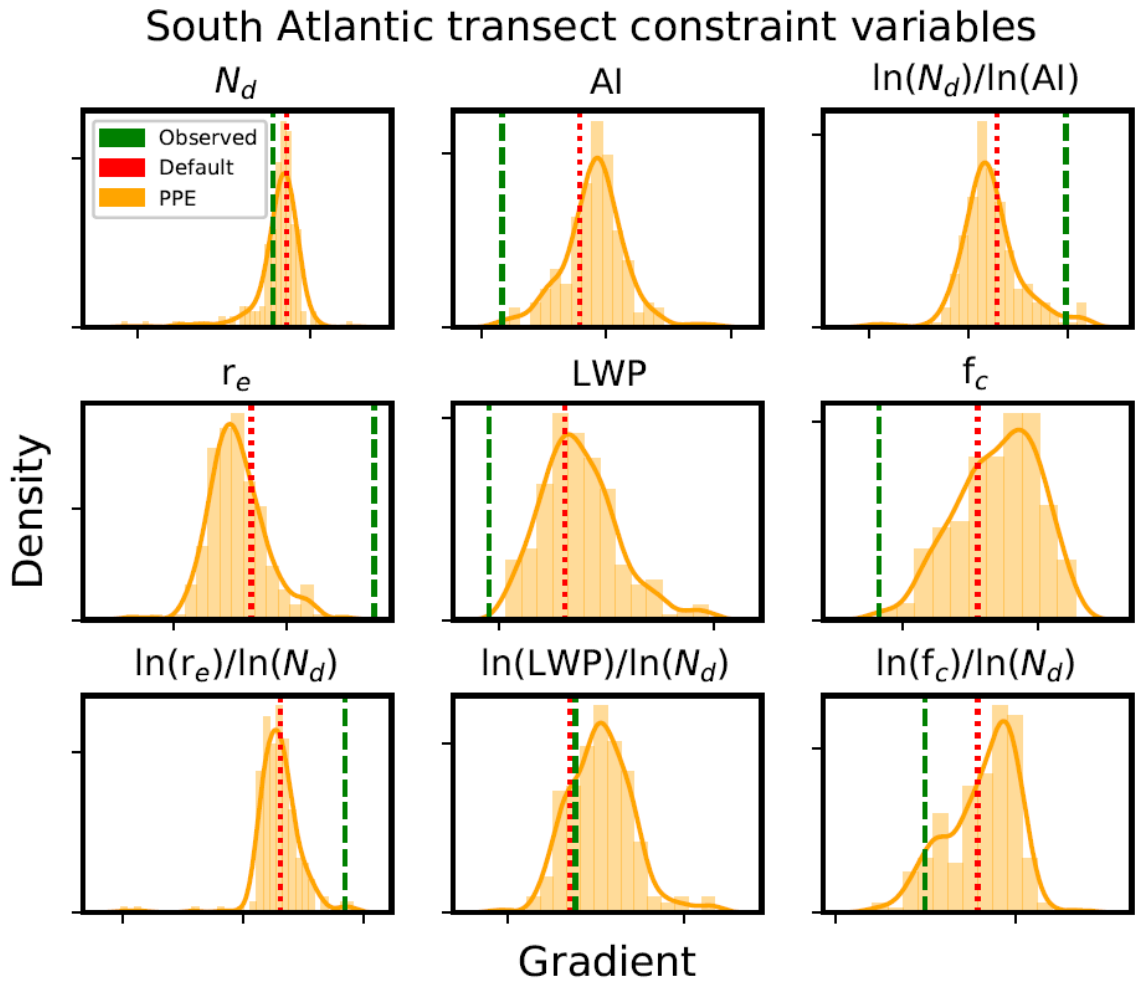


Fig. S28. Probability density functions of North Pacific transect constraint variables.



455 **Fig. S29.** Probability density functions of South Atlantic transect constraint variables.

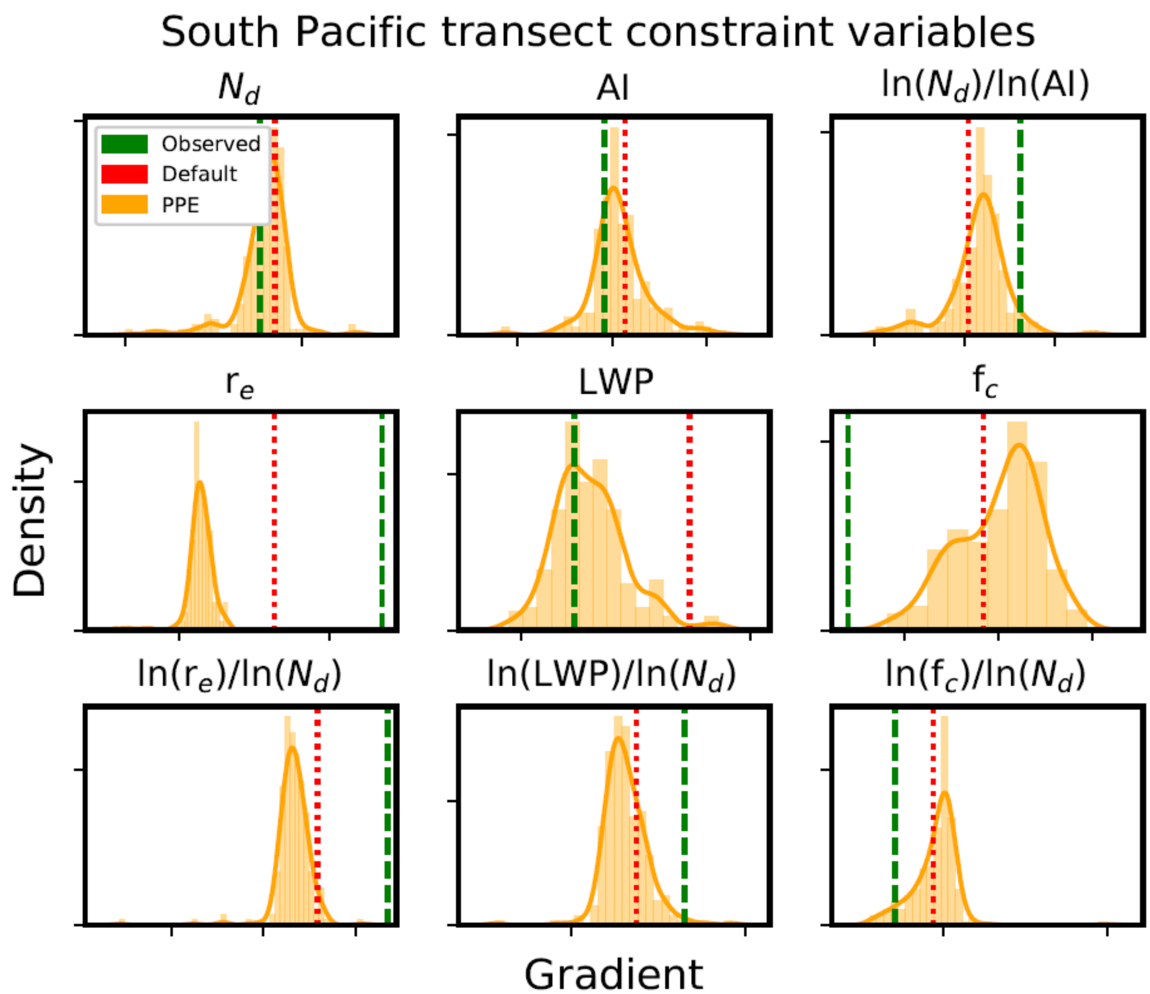


Fig. S30. Probability density functions of South Pacific transect constraint variables.

460

Table S1. Parameters perturbed in our PPE, the ranges they were perturbed over and default values as prescribed in the release version of the model.

Parameter	Minimum	Maximum	Default	Parameter Description
bl_nuc	0.1	10	1	Boundary layer nucleation rate scale factor
ait_width	1.2	1.8	1.59	Modal width of Aitken modes (nm)
cloud_ph	1e-7	2.51e-5	1e-5	Cloud droplet pH
carb_ff_diam	30	90	60	Emission diameter of carbonaceous aerosol from fossil fuel sources (nm)
carb_bb_diam	90	300	110	Emission diameter of carbonaceous aerosol from biomass burning sources (nm)
carb_res_diam	90	500	150	Emission diameter of carbonaceous aerosol from residential sources (nm)
prim_so4_diam	3	100	150	Emission diameter of 50% of new sub-grid sulfate particles (nm). Remaining 50% emitted into the larger coarse mode (nm)
sea_spray	0.25	4	1	Sea spray emission flux scale factor
anth_so_2	0.6	1.5	1	Anthropogenic SO2 emission flux scale factors. Applied independently to European, North American, Chinese, Asian regions and the rest of the world
volc_so2	0.71	2.38	1	Volcanic SO_2 emission flux scale factor
bvoc_soa	0.32	3.68	1	Biogenic monoterpene production rate of secondary organic aerosol scale factor
dms	0.33	3	1	Dimethyl-sulfide emission flux scale factor
prim_moc	0.4	6	1	Primary marine organic carbon emission flux scale factor
dry_dep_ait	0.5	2	1	Dry deposition velocity of Aitken mode aerosol
dry_dep_acc	0.1	10	1	Dry deposition velocity of accumulation mode aerosol
dry_dep_so2	0.2	5	1	Dry deposition velocity of SO_2
kappa_oc	0.2	0.65	0.65	Hygroscopicity parameter κ for organic aerosol – affects wet diameter and clear-sky radiative flux
sig_w	0.25	1.75	1	Standard deviation of shallow-cloud updraft velocity scale factor
rain_frac	0.3	0.7	0.3	Fraction of cloud covered area where rain forms
cloud_ice_thresh	0.1	0.5	N/A	Threshold of cloud ice water fraction for scavenging
conv_plume_scav	0	0.5	0.5	Scavenging efficieny (as a fraction of total aerosol removed) of Aitken mode aerosol in convective clouds
bc_ri	0.2	0.8	0.565	Imaginary part of the black carbon refractive index
oxidant_oh	0.7	1.3	1	Offline oxidant OH concentration scale factor

oxidants_o3	0.7	1.3	1	Offline oxidant O_3 concentration scale factor
bparam	-0.15	-0.13	-0.14	Coefficient of the spectral shape parameter β for effective radius
two_d_fsd_factor	1	2	1.4	Scale factor for the 2D relationship between cloud condensate variance, cloud cover and convection. Cotrols sub-grid cloud heterogeneity
c_r_correl	0	1	0.9	Cloud and rain sub-grid horizontal spatial colocation
autoconv_exp_lwp	2.15	3.31	2.47	Exponent of liquid water path in the power law for initiating autoconversion
autoconv_exp_nd	-3	-1	-1.79	Exponent of cloud droplet concentration (N_d) in the power law for initiating autoconversion
dbsdtbd_turb_0	0	1e-3	1.5e-4	Cloud erosion rate (s^{-1})
ai	0	5e-2	2.57e-2	Scaling coefficient for the dependence of ice mass on diameter
m_ci	0	3	1	Ice fallspeed scale factor
a_ent_1_rp	0	0.5	0.23	Cloud top entrainment rate scale factor

Table S2. Regions of persistent stratocumulus cloud used to calculate regional mean constraint variables.

Region	Latitude range	Longitude range
North Atlantic	34.4^{o} to 54.4^{o} N	329.1^{o} to 347.8^{o} E
North Pacific	14.4^{o} to 48.1^{o} N	197.8^{o} to 231.6^{o} E
South Atlantic	30.6^{o} to 10.6^{o} S	347.8^{o} to 2.8^{o} E
South Pacific	30.6^{o} to 15.6^{o} S	254.1^{o} to 284.1^{o} E
Southern Ocean	30.6^{o} to 50.6^{o} S	0^{o} to 360^{o} E

Table S3. Transects from stratocumulus- to cumulus-dominated regions.

Region	Start position	End position
North Atlantic	54.4º N, 336.6º E	45.6º N, 330.9º E
North Pacific	30.6º N, 229.7º E	19.4º N, 227.8º E
South Atlantic	11.9º S, 357.2º E	11.9º S, 345.9º E
South Pacific	20.6º S, 282.2º E	15.6º S, 269.1º E

Table S4. Effect of varying the number of model variants retained at each stage of constraint. We show the number of measurements needed to optimally constrain ΔF_{aer} and the 90% CI in each case.

Number of model variants retained	Number of measurements used	Lower, negative ΔF_{aer} bound	Upper ΔF_{aer} bound
1000	27	-1.15	-0.07
2000	31	-1.23	-0.10
5000	13	-1.26	-0.13
10000	29	-1.30	-0.13
20000	15	-1.33	-0.13

SI References

1. A. A. Sellar, *et al.*, UKESM1: Description and Evaluation of the U.K. Earth System Model. *J. Adv. Model. Earth Syst.* **11**, 4513–4558 (2019).

2. V. Eyring, *et al.*, Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.* **9**, 1937–1958 (2016).

3. K. D. Williams, *et al.*, The Met Office Global Coupled Model 3.0 and 3.1 (GC3.0 and GC3.1) Configurations. *J. Adv. Model. Earth Syst.* **10**, 357–380 (2018).

4. A. T. Archibald, *et al.*, "Description and evaluation of the UKCA stratosphere-troposphere chemistry scheme (StratTrop vn 1.0) implemented in UKESM1" (Atmospheric Sciences, 2019) https:/doi.org/10.5194/gmd-2019-246 (April 1, 2022).

5. D. Walters, *et al.*, The Met Office Unified Model Global Atmosphere 7.0/7.1 and JULES Global Land 7.0 configurations. *Geosci. Model Dev.* **12**, 1909–1963 (2019).

6. J. P. Mulcahy, *et al.*, Description and evaluation of aerosol in UKESM1 and HadGEM3-GC3.1 CMIP6 historical simulations. *Geosci. Model Dev.* **13**, 6383–6423 (2020).

7. J. P. Mulcahy, *et al.*, Improved Aerosol Processes and Effective Radiative Forcing in HadGEM3 and UKESM1. *J. Adv. Model. Earth Syst.* **10**, 2786–2805 (2018).

8. S. Ghan, *et al.*, Challenges in constraining anthropogenic aerosol effects on cloud radiative forcing using present-day spatiotemporal variability. *Proc Natl Acad Sci USA* **113**, 5804–5811 (2016).

9. D. P. Grosvenor, K. S. Carslaw, The decomposition of cloud–aerosol forcing in the UK Earth System Model (UKESM1). *Atmos. Chem. Phys.* **20**, 15681–15724 (2020).

10. GFAS, CAMS global biomass burning emissions based on fire radiative power (GFAS): data documentation, https://confluence.ecmwf.int/display/CKB/CAMS+global+biomass+burning+emissions+based+on+fire+radiative+power+%28GFAS%29%3A+data+documentation, (2022).

11. R. J. Andres, A. D. Kasgnoc, A time-averaged inventory of subaerial volcanic sulfur emissions. *J. Geophys. Res.* **103**, 25251–25261 (1998).

12. M. M. Halmer, H.-U. Schmincke, H.-F. Graf, The annual volcanic gas input into the atmosphere, in particular into the stratosphere: a global data set for the past 100 years. *Journal of Volcanology and Geothermal Research* **115**, 511–528 (2002).

13. G. W. Mann, *et al.*, Description and evaluation of GLOMAP-mode: a modal global aerosol microphysics model for the UKCA composition-climate model. *Geosci. Model Dev.* **3**, 519–551 (2010).

14. G. W. Mann, *et al.*, Intercomparison of modal and sectional aerosol microphysics representations within the same 3-D global chemical transport model. *Atmos. Chem. Phys.* **12**, 4449–4476 (2012).

15. S. Woodward, Modeling the atmospheric life cycle and radiative impact of mineral dust in the Hadley Centre climate model. *J. Geophys. Res.* **106**, 18155–18166 (2001).

16. R. E. L. West, *et al.*, The importance of vertical velocity variability for estimates of the indirect aerosol effects. *Atmos. Chem. Phys.* **14**, 6369–6393 (2014).

17. M. Lebsock, H. Morrison, A. Gettelman, Microphysical implications of cloud-precipitation covariance derived from satellite remote sensing: CLOUD-PRECIPITATION COVARIANCE. *J. Geophys. Res. Atmos.* **118**, 6521–6533 (2013).

18. I. A. Boutle, S. J. Abel, P. G. Hill, C. J. Morcrette, Spatial variability of liquid cloud and rain: observations and microphysical effects: Cloud and Rain Variability. *Q.J.R. Meteorol. Soc.* **140**, 583–594 (2014).

19. J. Browse, K. S. Carslaw, S. R. Arnold, K. Pringle, O. Boucher, The scavenging processes controlling the seasonal cycle in Arctic sulphate and black carbon aerosol. *Atmos. Chem. Phys.* **12**, 6775–6798 (2012).

20. M. Yoshioka, *et al.*, Ensembles of Global Climate Model Variants Designed for the Quantification and Constraint of Uncertainty in Aerosols and Their Radiative Forcing. *J. Adv. Model. Earth Syst.* **11**, 3728–3754 (2019).

21. L. A. Regayre, *et al.*, Aerosol and physical atmosphere model parameters are both important sources of uncertainty in aerosol ERF. *Atmos. Chem. Phys.* **18**, 9975–10006 (2018).

22. L. A. Regayre, *et al.*, The value of remote marine aerosol measurements for constraining radiative forcing uncertainty. *Atmos. Chem. Phys.* **20**, 10063–10072 (2020).

23. J. S. Johnson, *et al.*, Robust observational constraint of uncertain aerosol processes and emissions in a climate model and the effect on aerosol radiative forcing. *Atmos. Chem. Phys.* **20**, 9491–9524 (2020).

24. A. H. Peace, *et al.*, Effect of aerosol radiative forcing uncertainty on projected exceedance year of a 1.5 °C global temperature rise. *Environ. Res. Lett.* **15**, 0940a6 (2020).

25. N. Bellouin, *et al.*, Impact of the modal aerosol scheme GLOMAP-mode on aerosol forcing in the Hadley Centre Global Environmental Model. *Atmos. Chem. Phys.* **13**, 3027–3044 (2013).

26. Y. Balkanski, M. Schulz, T. Claquin, S. Guibert, Reevaluation of Mineral aerosol radiative forcings suggests a better agreement with satellite and AERONET data. *Atmos. Chem. Phys.* **7**, 81–95 (2007).

27. A. Metzger, *et al.*, Evidence for the role of organics in aerosol particle formation under atmospheric conditions. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 6646–6651 (2010).

28. D. M. H. Sexton, *et al.*, A perturbed parameter ensemble of HadGEM3-GC3.05 coupled model projections: part 1: selecting the parameter combinations. *Clim Dyn* **56**, 3395–3436 (2021).

29. K. S. Carslaw, *et al.*, Large contribution of natural aerosols to uncertainty in indirect forcing. *Nature* **503**, 67–71 (2013).

30. L. A. Regayre, *et al.*, The Climatic Importance of Uncertainties in Regional Aerosol–Cloud Radiative Forcings over Recent Decades. *Journal of Climate* **28**, 6589–6607 (2015).

31. P. S. Craig, M. Goldstein, A. H. Seheult, J. A. Smith, "Pressure Matching for Hydrocarbon Reservoirs: A Case Study in the Use of Bayes Linear Strategies for Large Computer Experiments" in *Case Studies in Bayesian Statistics*, Lecture Notes in Statistics., C. Gatsonis, *et al.*, Eds. (Springer New York, 1997), pp. 37–93.

32. D. Williamson, *et al.*, History matching for exploring and reducing climate model parameter space using observations and a large perturbed physics ensemble. *Clim Dyn* **41**, 1703–1729 (2013).

33. R. Stocki, A method to improve design reliability using optimal Latin hypercube sampling. *Comp. Ass. Mech. Eng. Sci.* **12**, 87–105 (2005).

34. A. O'Hagan, Bayesian analysis of computer code outputs: A tutorial. *Reliability Engineering & System Safety* **91**, 1290–1300 (2006).

35. CERES, CERES_SYN1deg_Ed4A Data Quality Summary, https://ceres.larc.nasa.gov/documents/DQ_summaries/CERES_SYN1deg_Ed4A_DQS.pdf, (2022).

36. GCPC, GCPC, https://psl.noaa.gov/data/gridded/data.gpcp.html, (2022).

37. MODIS, MODIS Cloud Product, https://modis.gsfc.nasa.gov/data/dataprod/mod06.php, (2022).

38. G. S. Elsaesser, *et al.*, The Multisensor Advanced Climatology of Liquid Water Path (MAC-LWP). *J. Climate* **30**, 10193–10210 (2017).

39. M. D. King, *et al.*, Cloud and aerosol properties, precipitable water, and profiles of temperature and water vapor from MODIS. *IEEE Trans. Geosci. Remote Sensing* **41**, 442–458 (2003).

40. A. Bodas-Salcedo, *et al.*, COSP: Satellite simulation software for model assessment. *Bulletin of the American Meteorological Society* **92**, 1023–1043 (2011).

41. G. Saponaro, *et al.*, Evaluation of aerosol and cloud properties in three climate models using MODIS observations and its corresponding COSP simulator, as well as their application in aerosol–cloud interactions. *Atmos. Chem. Phys.* **20**, 1607–1626 (2020).

42. D. Painemal, P. Zuidema, Assessment of MODIS cloud effective radius and optical thickness retrievals over the Southeast Pacific with VOCALS-REx in situ measurements: MODIS VALIDATION DURING VOCALS-REx. *J. Geophys. Res.* **116**, n/a-n/a (2011).

43. P. Vignesh, *et al.*, Assessment of CMIP6 Cloud Fraction and Comparison with Satellite Observations. *Earth and Space Science* **7** (2020).

44. T. Langton, P. Stier, D. Watson-Parris, J. P. Mulcahy, Decomposing Effective Radiative Forcing Due to Aerosol Cloud Interactions by Global Cloud Regimes. *Geophys Res Lett* **48** (2021).

45. I. L. McCoy, *et al.*, The hemispheric contrast in cloud microphysical properties constrains aerosol forcing. *Proceedings of the National Academy of Sciences* **117**, 189980–19006 (2020).

46. M. W. Christensen, W. K. Jones, P. Stier, Aerosols enhance cloud lifetime and brightness along the stratus-to-cumulus transition. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 17591–17598 (2020).

47. A. McComiskey, *et al.*, An assessment of aerosol-cloud interactions in marine stratus clouds based on surface remote sensing. *J. Geophys. Res.* **114**, D09203 (2009).

48. S. Kim, pcor: An R Package for a Fast Calculation to Semi-partial Correlation Coefficients. *Communications for Statistical Applications and Methods* **22**, 665–674 (2015).

49. L. A. Lee, K. S. Carslaw, K. J. Pringle, G. W. Mann, Mapping the uncertainty in global CCN using emulation. *Atmos. Chem. Phys.* **12**, 9739–9751 (2012).

50. N. Schutgens, *et al.*, On the spatio-temporal representativeness of observations. *Atmos. Chem. Phys.* **17**, 9761–9780 (2017).