An emulator approach to stratocumulus susceptibility

Franziska Glassmeier[1,4], Fabian Hoffmann[1,2], Jill S. Johnson[3], Takanobu Yamaguchi[1,2], Ken S. Carslaw[3], and Graham Feingold[1]

[1]NOAA Earth Systems Research Laboratory, Chemical Sciences Division, 325 Broadway, Boulder, CO 80302, USA
[2]Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, CO 80309, USA
[3]University of Leeds, School of Earth and Environment, Woodhouse Lane Leeds, LS2 9JT, UK
[4]Wageningen University, Department of Environmental Sciences, Droevendaalsesteeg 4, 6708PB Wageningen, NL

Correspondence: Franziska Glassmeier (franziska.glassmeier@noaa.gov)

Abstract. The climatic relevance of aerosol-cloud interactions depends on the sensitivity of the radiative effect of clouds to cloud droplet number N and liquid water path LWP. We derive the dependence of cloud fraction CF, cloud albedo A_{C} and the relative cloud radiative effect $\mathrm{rCRE} = \mathrm{CF} \cdot A_{\mathrm{C}}$ on N and LWP from 159 large-eddy simulations of nocturnal stratocumulus. These simulations vary in their initial conditions for temperature, moisture, boundary-layer height and aerosol concentration but share boundary conditions for surface fluxes and subsidence. Our approach is based on Gaussian process emulation, a statistical technique related to machine learning. We succeed in building emulators that accurately predict simulated values of CF, A_{C} and rCRE for given values of N and LWP. Emulator-derived susceptibilities $\partial \ln \mathrm{rCRE} / \partial \ln N$ and $\partial \ln \mathrm{rCRE} / \partial \ln \mathrm{LWP}$ cover the non-drizzling, fully-overcast regime as well as the drizzling regime with broken cloud cover. Theoretical results, which are limited to the non-drizzling regime, are reproduced. The susceptibility $\partial \ln \mathrm{rCRE} / \partial \ln N$ captures the strong sensitivity of the cloud radiative effect to cloud fraction, while the susceptibility $\partial \ln \mathrm{rCRE} / \partial \ln \mathrm{LWP}$ describes the influence of cloud amount on cloud albedo irrespective of cloud fraction. Our emulation-based approach provides a powerful tool for summarizing complex data in a simple framework that captures the sensitivities of cloud field properties over a wide range of states.

1 Introduction

15   Aerosol perturbations can lead to changes in cloud brightness and amount via the influence of aerosol on cloud formation and various aerosol-cloud interaction (ACI) processes. Our process understanding of ACI has improved greatly over recent decades, however the radiative forcing due to ACI continues to dominate the uncertainty margin of the total anthropogenic forcing of the climate system. Due to their abundance and location, forcing and forcing uncertainty are dominated by shallow, warm clouds in marine boundary layers (Myhre et al., 2013; Boucher et al., 2013).

20   ACIs are notoriously hard to quantify because they pose a multi-scale problem, not only in terms of spatial and temporal scales but also in terms of the effective degrees of freedom used to describe ACI in different settings. The multi-scale spectrum of approaches to the ACI problem has been described to range from "Darwinian" (low-level, high-dimensional, complex, *reductionist*) to "Newtonian" (high-level, low-dimensional, effective, *emergent*) descriptions (Harte, 2002; Feingold et al., 2016; Mülmenstädt and Feingold, 2018).

We illustrate this notion by discussing the *relative cloud radiative effect* (rCRE) as quantified by

$$\mathrm{rCRE} = \frac{F_{\mathrm{clr}} - F_{\mathrm{all}}}{F_{\mathrm{clr}}} \approx A_{\mathrm{C}} \cdot \mathrm{CF}, \tag{1}$$

where F denotes downwelling SW radiative fluxes at the surface under clear-sky (index clr) and all-sky (index all) conditions, and A_{C} and CF denote cloud albedo and cloud fraction, respectively (Xie and Liu, 2013). By describing their effect on the radiation budget, rCRE captures the effect of clouds on climate. In the spirit of Platnick and Twomey (1994), ACI can be quantified based on the *susceptibility*, or normalized sensitivity, of the rCRE to the cloud droplet number concentration N:

$$\frac{\mathrm{d}\ln\mathrm{rCRE}}{\mathrm{d}\ln N} = \frac{\partial\ln\mathrm{rCRE}}{\partial\ln N} + \frac{\partial\ln\mathrm{rCRE}}{\partial\ln\mathrm{LWP}} \frac{\mathrm{d}\ln\mathrm{LWP}}{\mathrm{d}\ln N}. \tag{2}$$

The decomposition on the right-hand side (RHS) is motivated by distinguishing cloud micro-structure as captured by N from macro-structure as represented by the liquid water path, LWP. Note that this decomposition does not necessarily align with the contributions of A_{C} and CF to rCRE. Aerosol effects on cloud micro-structure are associated with cloud brightness (Twomey, 1977, 1974; Boers and Mitchell, 1994; Feingold et al., 1997; Christensen and Stephens, 2011; McGibbon and Bretherton, 2017), while aerosol effects on the macro-structure of the cloud correspond to cloud amount (Albrecht, 1989; Matheson et al., 2005; Kaufman et al., 2005; Small et al., 2009; Stevens and Feingold, 2009; Zheng et al., 2010; Christensen and Stephens, 2011; Chen et al., 2014; Feingold et al., 2015; McGibbon and Bretherton, 2017).

The reductionist approach to deriving rCRE(N,LWP) and the partial derivatives in Equation 2 starts by asking how $A_{\mathrm{C}}(\tau)$ depends on cloud optical thickness τ, followed by deriving the dependence of τ(LWP, N) on LWP and N, which are, in turn, functions of the meteorological and aerosol conditions. Even more complex chains of dependencies can be derived for CF(N,LWP) and LWP(N) because these relationships are shaped by the entire cloud field. The advantage of this approach is that each link can in principle be deduced from detailed process understanding. The disadvantage is that a large number of variables and processes need to be quantified.

The emergence-based alternative is to subsume process complexity in low-dimensional relationships that effectively describe rCRE as a function of a small set of controlling variables. In other words, this means searching for a simple relationship rCRE(LWP, N). Or it may mean abandoning the idea of disentangling aerosol effects on cloud microstructure (N) and macrostructure (LWP) altogether – similar to the definition of effective radiative forcing in Myhre et al. (2013) – and quantifying d ln rCRE/d ln N directly. While the direct nature of this approach is an obvious advantage, the challenge of the emergence-based approach lies in its data-mining, exploratory nature, and lack of a priori guidance for finding and justifying emergent relationships.

In this paper, we aim to combine the reductionist and emergence-based approaches to determine the partial derivatives in Equation 2; the LWP adjustment $\mathrm{d}\ln\mathrm{LWP}/\mathrm{d}\ln N$ will be the topic of an upcoming paper. We demonstrate how a statistical method related to machine learning can be applied to derive system-wide relationships from the detailed process representation that is ingrained in a set of model simulations. Our contribution addresses an increasing interest in machine learning approaches within the atmospheric sciences, especially in the context of parameterizing shallow clouds (Krasnopolsky et al., 2013; Schneider et al., 2017; Brenowitz and Bretherton, 2018; Gentine et al., 2018; O'Gorman and Dwyer, 2018). This interest

in utilizing modern computational-statistical methods illustrates a community need to address a certain mismatch between traditional process-based cloud research and synthesizing approaches, especially for representing clouds in climate models.

We specifically apply the tool of *Gaussian process emulation* (Rasmussen and Williams, 2006; O'Hagan, 2006). Emulation is an established method used to extract multi-dimensional relationships from sparse data. It can be considered a form of kernel-based supervised machine learning. In the atmospheric sciences, emulation has so far been used to investigate the relationship between model response and uncertain parameters associated with physical parameterizations and to a lesser extent boundary conditions, e. g., Lee et al. (2011), Lee et al. (2013), Johnson et al. (2015) and Posselt et al. (2016). We adapt this method to quantitatively derive relationships between cloud field properties, namely rCRE(LWP, N), A_{C}(LWP, N) and CF(LWP, N), that evolve over the course of the simulation, from simulation data.

We present state-of-the-art large-eddy simulations (LES) of stratocumulus (Section 2) and demonstrate that our approach (Section 3) successfully translates process understanding captured by the LES into a quantification of the rCRE, A_{C} and CF and their relationships to LWP and N (Section 4). As an application, we then derive and discuss the partial susceptibilities in Equation 2 (Section 5), before we conclude (Section 6).

2   Simulations

We perform LES with the System for Atmospheric Modeling (SAM) (Khairoutdinov and Randall, 2003). Our domain measures $48 \times 48\,\mathrm{km}^2$ at a horizontal resolution of $200\,\mathrm{m}$ and a vertical resolution of $10\,\mathrm{m}$. The timestep is $1\,\mathrm{s}$. Simulations are nocturnal, and of $12\,\mathrm{h}$ duration. Sea surface temperature, subsidence and surface fluxes are fixed to the values of Ackerman et al. (2009). The model activates aerosol particles based on the prognosed supersaturation and simulates condensation/evaporation and collision-coalescence using a bin-emulating 2-moment approach (Feingold et al., 1998). Particles are removed by collision-coalescence scavenging and wet deposition. We assume a surface source of aerosol particles of $70\,\mathrm{cm}^2\,\mathrm{s}^{-1}$ (Yamaguchi et al., 2017; Kazil et al., 2011). Integrated properties such as cloud water path (CWP), rain water path (RWP) and their sum, the LWP, are calculated directly from cloud and rain water mass mixing ratios. Drop number concentration is calculated from the prognosed cloud and rain number concentrations and is usually dominated strongly by the former. For more details on the model setup see Yamaguchi et al. (2017).

Following Feingold et al. (2016), we simulate 191 stratocumulus (Sc) cases with different initial conditions. We simultaneously vary six initial conditions of the Sc field: liquid water potential temperature ($284 < \theta_l/\mathrm{K} < 294$), the total mixing ratio ($6.5 < q_t/(\mathrm{g\,kg}^{-1}) < 10.5$) and the aerosol concentration ($30 < N_a/\mathrm{cm}^{-3} < 500$) in a mixed layer, as well as the initial height ($500 < H_{\mathrm{mix}}/\mathrm{m} < 1300$) of that mixed layer, and the jumps ($6 < \Delta\theta_l/\mathrm{K} < 10$, $-10 < \Delta q_t/(\mathrm{g\,kg}^{-1}) < -6$) at the inversion above the mixed layer.

To generate our ensemble of model simulations, we sample well-spaced combinations of these initial conditions over the 6-dimensional space spanned by the given ranges using a maximin Latin-hypercube design algorithm (Morris and Mitchell, 1995). This Latin-hypercube sampling ensures good coverage across the 6-d space of the initial conditions and prevents any spurious aerosol-meteorology co-variability due to sampling (Feingold et al., 2016).

We create a Latin-hypercube sampling of 200 points. Not all of these correspond to conditions for which cloud formation is expected. Based on applying saturation adjustment as a posteriori condition, we identify 191 suitable initial conditions. For these initial conditions, actual LES are performed. From these simulations, we remove all those that do not sustain cloud. We also remove simulations that would form rain within the first hour while collision-coalescence and sedimentation are

5   still switched off for spin-up. These simulations are characterized by unrealistically strong rain once collisions-coalescence is allowed. We identify such early-precipitating simulations based on the criterion that the maximum domain-averaged surface precipitation rate over the timeseries $(0 - 12\,\text{h})$ exceeds $10\,\text{mm}\,\text{day}^{-1}$.

After this filtering, 159 of the initial 191 simulations remain. We discard two hours of spin-up and build our analysis on hours 2 to 12, at output intervals of $10\,\text{min}$, which means that the timeseries from each simulation consists of 60 domain-averaged

10   values. The total number of data points amounts to $60 \cdot 159 = 9540$.

Figure 1 summarizes our dataset as a function of the domain-averaged liquid water path LWP, which we define as the sum of cloud and rain water path, $\text{LWP} = \text{CWP} + \text{RWP}$, and the vertically-averaged cloud droplet number concentration N in cloud columns with optical depth $\tau_{500\,\text{nm}} > 1$, where the index indicates the considered wavelength. To illustrate the system evolution, we discuss the four labeled trajectories in the figure. Figure 2 illustrates the spatial arrangement of these trajectories. The clouds

15   of trajectory A deepen in response to longwave radiative cooling and attendant condensation. In contrast, the thick clouds of trajectory B feature strong entrainment that leads to cloud thinning. Cloud deepening and thinning approximately balance each other in a region indicated by the solid blue line in Figure 1. Trajectory C shows a cloud system with large droplets whose adiabatic volume-mean droplet radius at cloud top quickly reaches a critical value of about $12\,\mu\text{m}$ associated with the onset of precipitation (Gerber, 1996). This rapidly reduces the cloud droplet number (Figure 1) and leads to the break-up of the

20   cloud field (Figure 2). Trajectory D initially features droplet sizes that are too small for precipitation formation. Through cloud deepening similar to trajectory A, droplets grow until their size crosses the threshold value for precipitation formation. As for trajectory C, this means that the cloud droplet number starts to decrease.

To guide further discussion, we partition the state space into three regions: The upper right quadrant (1$^{\text{st}}$ quadrant Q1 in the following) is characterized by the absence of drizzle and evolution from the initial state towards decreasing LWP; the upper left

25   quadrant (Q2) features no drizzle and increasing LWP; the lower part of the state space (Q34) features drizzle and decreasing droplet number.

We distinguish the 2-dimensional *state space* spanned by LWP and N from the *parameter space* of the system. Parameters are set externally and do not evolve in time. The parameters of our simulations are its boundary conditions, especially sea surface temperature, subsidence and surface fluxes. Initial conditions could also be considered parameters. Their role for the

30   system's evolution is, however, somewhat ill-defined due to the spin-up process. In real systems, a distinction between state variables and parameters is only approximate. It requires a timescale separation so that slowly evolving variables can be considered as fixed and parameter-like in comparison to fast evolving variables. While previous applications of emulators, e. g. Lee et al. (2013) and Johnson et al. (2015), have explored how the behavior of the system varies across the parameter space, we explore how it varies across the resulting state space. Our choice of LWP and N as state variables is motivated by

35   Equation 2. It is not a priori clear that the properties of cloud fields that arise from a 6-dimensional set of initial conditions can

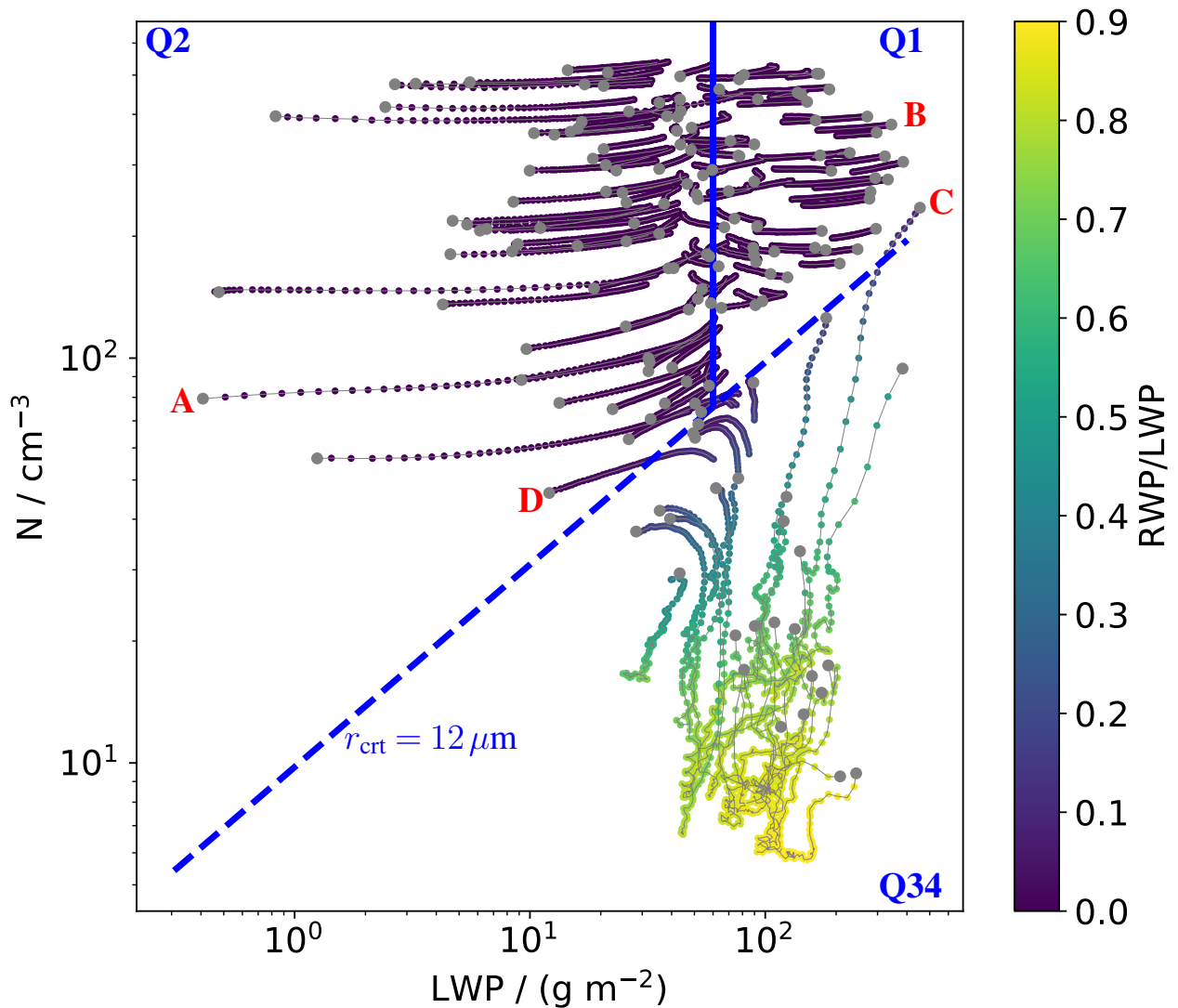



Figure 1. Temporal evolution (hours 2-12) of 159 LES with varying initial conditions (see text) in an LWP-N state space, colored by fraction of rain water path (RWP) to the total liquid water path (LWP). Individual simulations are indicated by gray lines and start at the location of gray circles. The dashed blue line corresponds to an adiabatic volume-mean droplet radius at cloud top of $12\,\mu$m (adiabatic condensation rate $\gamma = 2.5 \cdot 10^{-6}\,\text{kg}\,\text{m}^{-4}$). Together with the solid blue line, it defines three regions, which are labeled Q1 (first quadrant), Q2 (second quadrant) and Q34 (combination of third and fourth quadrants). Red letters indicate trajectories discussed in the main text.

be described as 2-dimensional functions. For such a reduction in dimensionality to be successful, it is required that multiple initial conditions in the 6-dimensional space map onto individual points on the 2-dimensional state space. In hydrology, this

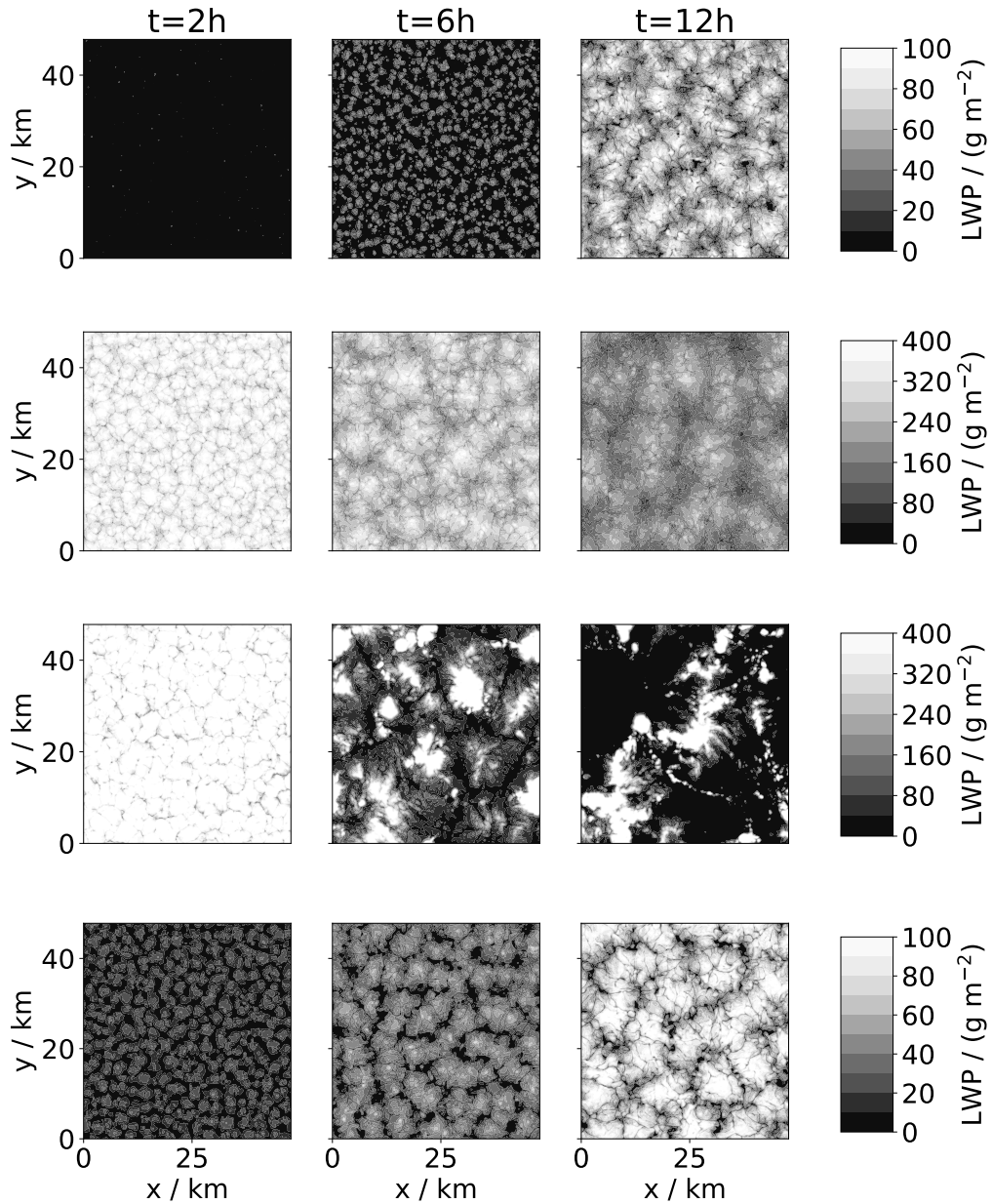



Figure 2. Horizontal arrangement of liquid water path (sum of rain and cloud water) for trajectories (top) A to (bottom) D from Figure 1. The left column shows $t = 2\,\mathrm{h}$, which corresponds to the gray circles in Figure 1. The right column shows the last time step ($t = 12\,\mathrm{h}$) and the central column shows $t = 6\,\mathrm{h}$.

circumstance is known as *equifinality* (Beven, 2005). Our data does not perfectly, but only approximately, collapse onto the 2-dimensional state space. This imperfection manifests as noise on our data when presented in two dimensions.

3 Building ensembles of emulators

We analyze our data based on the assumption that the relationship between the state-space and the cloud output variables can be modeled as a Gaussian process, and employ the technique of *Gaussian process emulation* (Rasmussen and Williams, 2006). The historical origins of this method lie in predicting the distribution of gold in South Africa, based on a small sample of carefully located test drillings. Gaussian process emulation is a preferred interpolation technique for sparse data (that is ideally well-spaced) of variables that vary smoothly (no discontinuities) across the dimensions of interest. Our data samples the LWP-N state space partially in a sparse manner (when considering different runs) and partially in a reasonably dense manner (within the timeseries of individual runs). In contrast to the 6-dimensional initial conditions, which we sampled using a space-filling Latin hypercube design, the data associated with the system evolution is not Latin-hypercube sampled. As the system evolves and moves towards the solid blue line in Figure 1, the coverage of the space becomes less evenly distributed, which can lead to issues of instability in parameter estimation when fitting a Gaussian process emulator. We therefore adapt the emulation approach to our situation. To fulfill the methodological requirements of sparsity and sampling, we generate a set of approximately Latin-hypercube sampled subsets of the data and construct a Gaussian process emulator for each subset (see Section 3.1). Together, this set of fitted emulators form an emulator ensemble that takes into account most of the available data (see Section 3.2).

3.1 Sub-sampling

To build and validate the emulators in the emulator ensemble, we split the total dataset randomly into two equally large subsets. We use one of these to build the emulator (*training dataset*) and the other to independently test the emulator (*validation dataset*). The random splitting is based on equal, unconditioned probabilities for each data point to belong to either of the two datasets. It especially does not take into account to which timeseries a data point belongs and which data points from the same timeseries may already be in the same dataset.

Gaussian process-emulation is designed for reasonably sparse and well-spaced data over the dimensions of interest. As discussed, our data does not take this form because it consists of many densely-sampled timeseries that are themselves sparsely distributed in state space. We therefore sub-sample from the total training data in the following way: We create a virtual set of n_{trn} Latin-hypercube sampled points in the LWP-N state space and replace each point in the Latin-hypercube sample by the geometrically closest point from our dataset. Data points are added when their normalized Euclidean distance from the Latin-hypercube sample point does not exceed $5/n_{trn}$. We do not use data points twice and training and validation data are treated as completely separate such that a data point cannot be selected for both training and validation. Figure 3 illustrates two such sub-samplings. We proceed in the same way to select a sub-sample of $n_{vld} = 2n_{trn}$ validation data points from the validation data set. The Latin-hypercubes underlying the samplings for the training and the validation dataset augment each other so that their combination is also Latin-hypercube sampled. We achieve the Latin-hypercube samples using the R package `lhs` (R Core Team, 2018; Carnell, 2018).

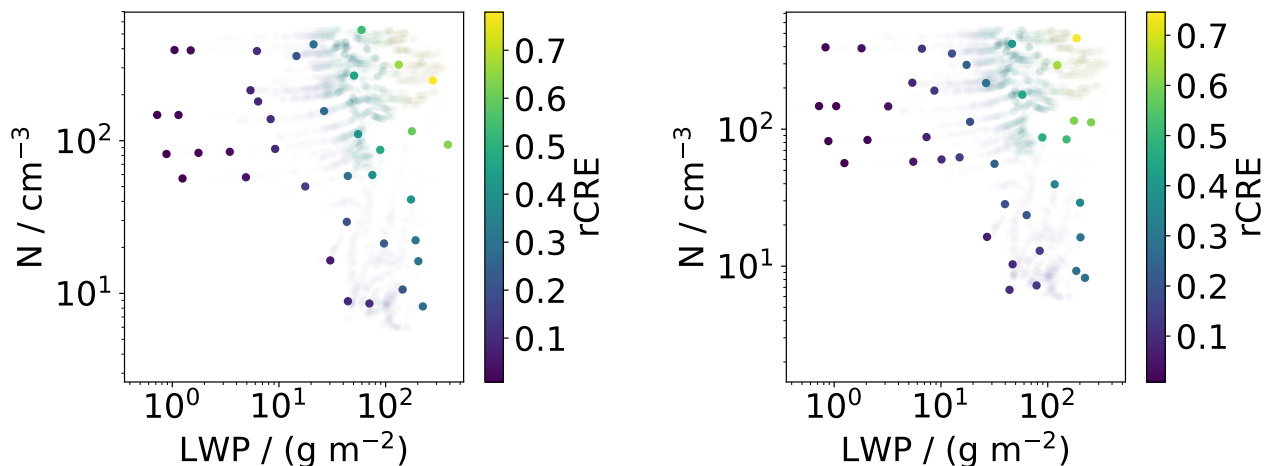



Figure 3. Sub-sampling (solid circles) from the total training data (opaque points) for the rCRE emulator ensemble members with the (left) lowest and (right) highest root mean square error (rmse) in predicting the validation data. The rmse values are 0.010 and 0.024, respectively.

3.2 Ensemble emulation and averaging

By varying the random seed of the initial Latin-hypercube sampling, we create an ensemble of sub-samplings. To each sub-sampling we apply Gaussian process emulation, constructing and validating a separate emulator model for each sub-sample in turn. For a general overview of the mathematical details of the emulator model, we refer to Johnson et al. (2015). It is based on a Bayesian statistical framework, where we select an underlying mean and co-variance structure that is then fitted given information from the training data. We specifically assume a linear combination of LWP and N as the underlying mean function, and use a Matérn covariance structure (Rasmussen and Williams, 2006). We account for noise in our data (*nugget effect*). Our data is noisy because our simulations do not perfectly, but only approximately, collapse onto the LWP-N state space. This is illustrated by the fact that individual data points in region Q34 of Figure 1 may differ in their value of RWP/LWP, i.e., their color, from closely neighboring points (cf. Figure 5 to see the same for rCRE instead of RWP/LWP). Emulators are fitted using the function `km()` in the R package `DiceKriging` (Roustant et al., 2012; R Core Team, 2018).

As the data points chosen for different Latin-hypercube samplings are not necessarily different, the emulator ensemble members that we obtain in this way are not independent. To limit the repeated use of data points, we relate the number of ensemble members n_ensbl, or random seeds, respectively, to the number of training data points for the individual emulators n_trn such that no more than 50% of the total available training data n_trn^tot will be used:

$$n_\text{ensbl} = \frac{n_\text{trn}^\text{tot}}{2n_\text{trn}}.$$

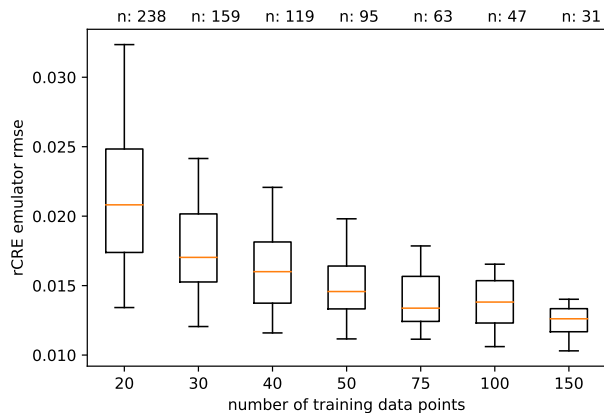



Figure 4. Standard deviation of individual rCRE emulator ensemble members obtained using different sub-samplings, as a function of the number of training data points. The number of ensemble members is indicated at the top of the plots. Orange lines, boxes and whiskers correspond to the median, upper and lower quartile, and 5th and 95th percentile of the distribution of standard deviations in the emulator ensemble.

We characterize an individual emulator within the emulator ensemble by its root mean square error (rmse) in predicting the validation data. We characterize the emulator ensemble as a whole by a weighted mean μ, where the weighting w depends on the rmse of the individual emulators:

$$\mu = \frac{\sum_i w_i x_i}{\sum_i w_i} \quad \text{with} \quad w_i = 1 - \frac{\text{rmse}_i - \min(\text{rmse})}{\max(\text{rmse}) - \min(\text{rmse})}. \tag{3}$$

For the example of the rCRE emulator shown in Figure 5, Figure 3 illustrates the best and worst sampling within the ensemble as quantified by the rmse of the corresponding individual emulators.

Figure 4 shows the spread of emulator rmse that are obtained when varying the number of training data points n_{trn} used to build an emulator ensemble. The median rmse tends to decrease with an increasing number of data points, while the number of ensemble members decreases commensurately. As a compromise between the quality of individual emulators and ensemble statistics we choose $n_{\text{trn}} = 50$.

4 Emulators for rCRE, cloud albedo and cloud fraction

Figures 5 and 6 demonstrate the successful application of our emulator ensemble technique to derive surfaces of rCRE, A_c and CF (all determined using a threshold of $\tau > 1$) as a function of LWP and N from the multi-timeseries data shown in Figure 1. The emulated surfaces successfully predict simulation outcomes (validation data) that were not used in creating the emulator (training data) as shown by scatter plots and data points in the figures.

In accordance with Equation 1, the shape of the emulated rCRE surface follows from the surfaces of cloud albedo and cloud fraction (Figure 6, a,c). Cloud fraction generally decreases with decreasing LWP as the Sc deck entrains and thins until the

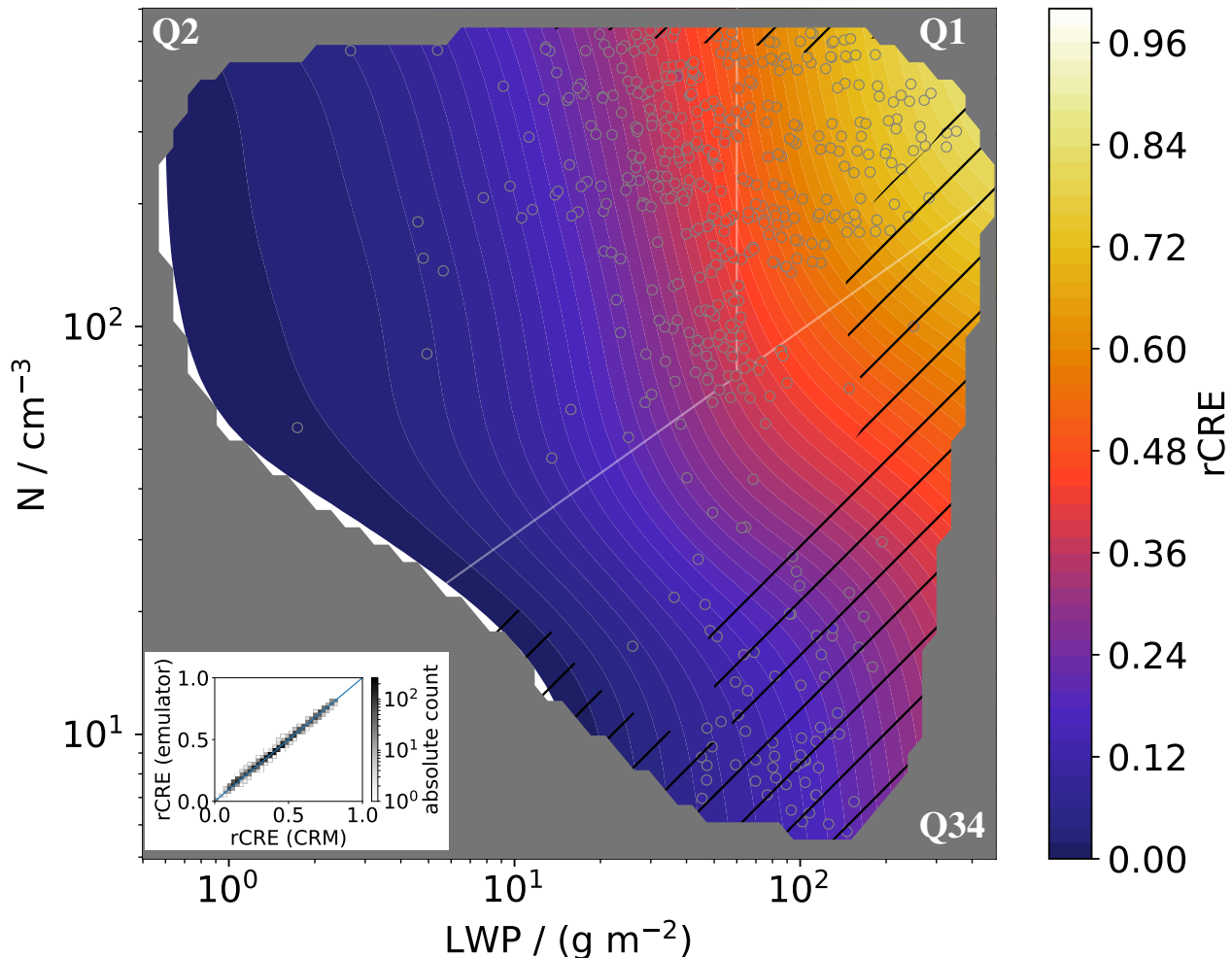
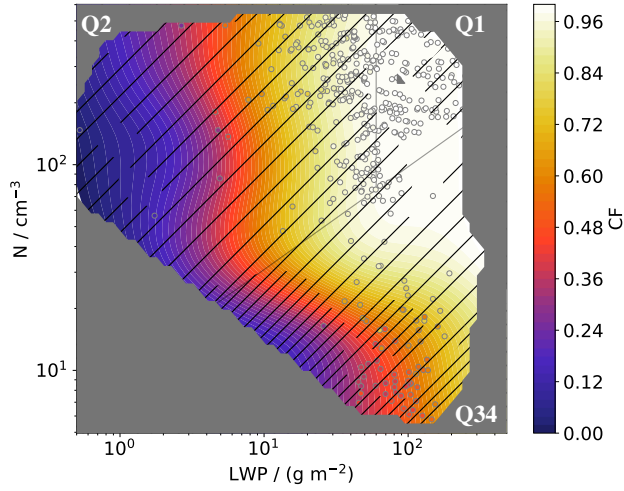
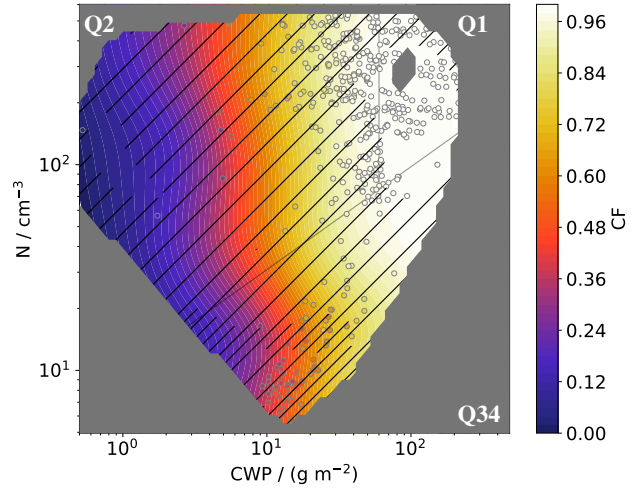



Figure 5. Emulated rCRE surface (ensemble mean Equation 3) as a function of LWP and droplet number N as color contours. The standard deviation σ of the emulated rCRE over the ensemble (Equation 4) is indicated by hatching, with $\sigma < 0.01$ and $0.01 < \sigma < 0.05$ for non-hatched and hatched regions, respectively. Emulated values outside $[0, 1]$ are masked. Color-filled circles with gray outline show the validation subset of the total dataset shown in Figure 1 (see Section 3.1). For visibility only every 10th validation data point is shown. The colorscale for the data points is the same as for the contours and the gray outline of their edges has been added to distinguish validation data from the emulated surface (see Figure 9 for an example for this type of plot where values of data points and surface differ more). The scatter plots in the insets compare the complete validation dataset to the emulated values. White lines indicate quadrants as in Figure 1.

detection threshold $\tau = 1$ occurs. Its surface is dominated by a gradual shift of its isolines from Q2 to Q34 as rain formation sets in. In the shift region (Q3), cloud fraction depends strongly on N. Elsewhere, isolines run mostly in the vertical direction

a) cloud fraction CF(LWP, N)   b) cloud fraction CF(CWP, N)

c) cloud albedo A_C(LWP, N)   d) cloud albedo A_C(CWP, N)

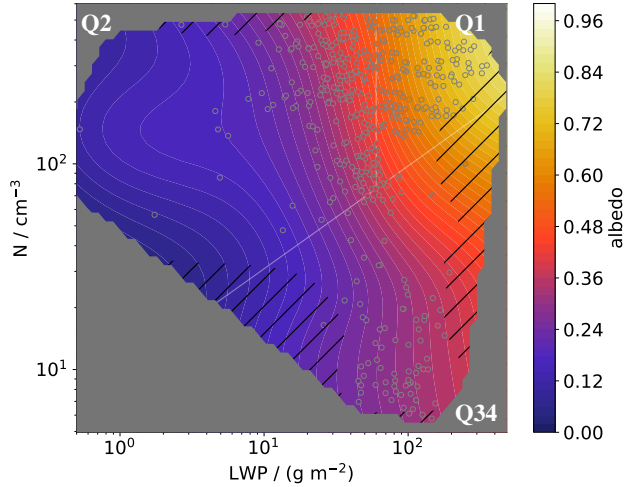
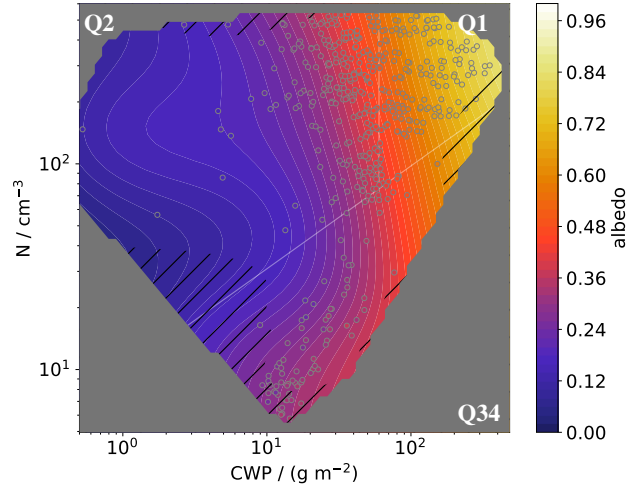



Figure 6. As Figure 5 but showing (a, b) cloud fraction and (c, d) cloud albedo as a function of (a, c) total liquid water path LWP and cloud droplet number concentration N and (b, d) cloud water path CWP and N. The standard deviation σ of the emulated variable over the ensemble (Equation 4) is indicated by hatching, with $\sigma < 0.01$, $0.01 < \sigma < 0.05$ and $0.05 < \sigma < 0.5$ for non-hatched, single-hatched, and double-hatched regions, respectively.

such that CF is largely independent of N. Cloud albedo is characterized by negatively-sloped isolines so that cloud albedo tends to increase with both, LWP and N. The isolines, and thus the dependence of cloud albedo on LWP and N, are distorted in the drizzling region (Q34).

This distortion in Q3 is caused by the bimodality of the drop-size distributions, so that cloud albedo and fraction are influenced by the radiative effects of cloud droplets as well as rain drops. Figure 6 (b, d) considers the cloud water path (CWP) instead of LWP as x-axis. This transformation leads to a shift in the isolines: Cloud fraction isolines become approximately vertical as $\tau > 1$ is controlled by CWP and hardly influenced by the additional contribution of RWP to LWP. The tilt of the cloud albedo isolines is reversed, indicating a contribution of RWP to total cloud albedo.

4.1 Uncertainty

Our approach features five different kinds of uncertainties, or errors. First, the Gaussian process emulation returns a random variable, i.e. a probability distribution of possible surfaces. This uncertainty depends on the training data used to build an individual emulator. It can be quantified by a standard deviation and its values can be inferred from Figure 8.

Second, the quality of the emulator mean function is quantified by the rmse of an emulator in predicting the validation data. This measure of uncertainty depends on the training as well as the validation data that is used for a specific emulator, i.e., one member of the emulator ensemble. It is not spatially resolved but averages the error of the emulator over the whole LPW-N state space. As indicated in the caption of Figure 3, rmse lies between 0.01 and 0.02 for the rCRE emulators.

Third, we have the error due to the noise that arises when collapsing our dataset onto the LWP-N state space. This uncertainty is the most fundamental uncertainty because it cannot be reduced by increasing the amount of data available. It is inherent to our modeling of rCRE as 2-dimensional function of LWP and N.

Fourth, we have the uncertainty of the emulator ensemble. Following Equation 3, we quantify this uncertainty using a weighted ensemble standard deviation,

$$\sigma = \sqrt{\frac{\sum_i w_i (x_i - \mu)^2}{\sum_i w_i}}, \tag{4}$$

where the weights w_i are defined in Equation 3 and depend on the rmse. Hatchings in Figure 5 show that the ensemble uncertainty (weighted standard deviation) for values of the rCRE is mostly smaller than 0.05 and in large regions of the state space smaller than 0.01. Comparing the different emulators, we find that the cloud-fraction emulator ensemble is more uncertain than that of the cloud albedo (Figure 6). This reflects that cloud albedo only depends on the local properties of the cloud field that are directly represented by LWP and N, while the cloud fraction is a cloud-field property. The cloud-fraction uncertainty is mitigated by considering the combined quantity of the rCRE.

Lastly, we have a sampling uncertainty illustrated in Figure 7. We indicate the level of sampling uncertainty by counting the number of trajectories that sample a specific part of the phase space. To this end, we partition the LWP-N state space into 60×50 bins and for each bin we count how many of our 159 trajectories contain a data point within. This uncertainty arises because our data, especially when projected onto the 2-dimensional LWP-N state-space, is noisy. Regions of the state space that are sampled by a single trajectory are thus less reliably represented than regions where the consideration of different trajectories attenuates the noise. Note that the nugget effect assumed for building individual emulators cannot account for this sampling uncertainty: an undersampled region does not appear noisy.

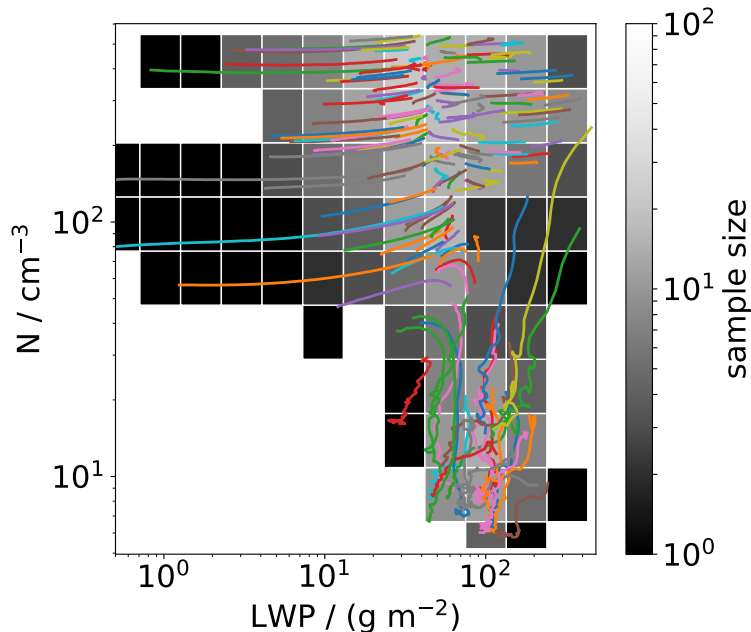



Figure 7. Illustration of sampling by trajectories. The sample size (greyscale) is determined by counting the number of trajectories (colored lines) per grid box. Note that the trajectories shown are the same as in Figure 1.

Figure 8 compares the three spatially resolved types of error for the rCRE emulator. The standard deviation of the ensemble is mostly smaller than the standard deviation of the individual emulator. This reflects the larger amount of data and information considered when building the emulator ensemble. Comparison with the sampling uncertainty indicates that the ensemble standard deviation may be overconfident in poorly sampled regions because it cannot capture the true level of noise in these

5   regions. Therefore, because the ensemble uncertainty is generally small, we will use the sampling uncertainty to guide the interpretation of the rCRE emulator in the following.

4.2   Comparison to bilinear regression and effective degrees of freedom

Previous studies have determined partial susceptibilities as in Equation 2 from binned linear (Sena et al., 2016, e.g.,) or bilinear regression (Jiang et al., 2010; Glassmeier and Lohmann, 2018). We therefore add a brief comparison of our method to bilinear

10   regression. Figure 9 demonstrates that bilinear regression does not capture the simulation data as well as the emulator surface. While the coefficient of determination, $r^2 = 0.95$, and rmse$= 0.05$ are acceptable, the regression surfaces cannot account for the tilt of isolines discussed in Section 4. As a consequence, the regression surface predicts a large region of unphysically negative rCRE. The reason behind the poor performance of bilinear regression is that its 3 free parameters are insufficient to capture the complexity of the emulated surface. The number of degrees of freedom required to adequately capture the

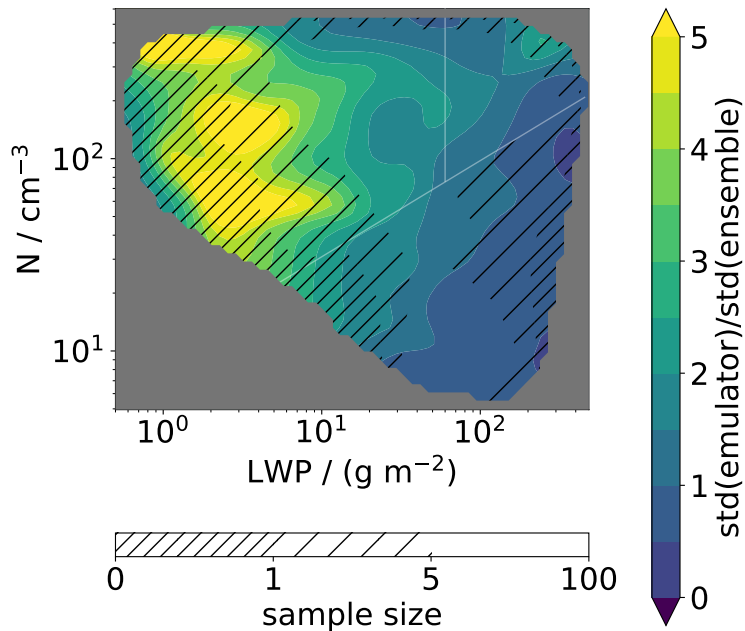



Figure 8. Comparison of errors from emulation, ensemble, and sampling. Contours show the ratio of emulator and ensemble standard deviation, hatchings show the sample size as in Figure 7. Note that sample sizes smaller than one trajectory can occur due to interpolation.

complexity of the simulation data can be estimated from Figure 4. The value of n_{trn} at which the emulator rmse levels off can be interpreted as the number of degrees of freedom of the fitted surface, in our case about 50. Performing bilinear regression for specific bins increases the total number of free parameters in the regression. In contrast to emulation, however, this approach is limited by the requirement of a sufficient number of data points per bin. We therefore consider emulation to be a powerful
5   and superior alternative to binned regression studies.

5   Partial susceptibilities of rCRE to droplet number and LWP

While partial susceptibilities (Equation 2) are directly obtained as the coefficients of a bilinear regression, the emulated rCRE surface requires their derivation as finite differences of the array that represents the surface. Figure 10 shows the logarithmic derivatives of the surface shown in Figure 5.
10   In the upper part of the state space (Q1 and Q2), emulator-derived susceptibilities in Figure 10 (a, b) compare reasonably well to theoretical results for non-drizzling conditions (black line contours), which assume a unimodal droplet-size distribution and high-cloud fraction (Boers and Mitchell, 1994; Sena et al., 2016):

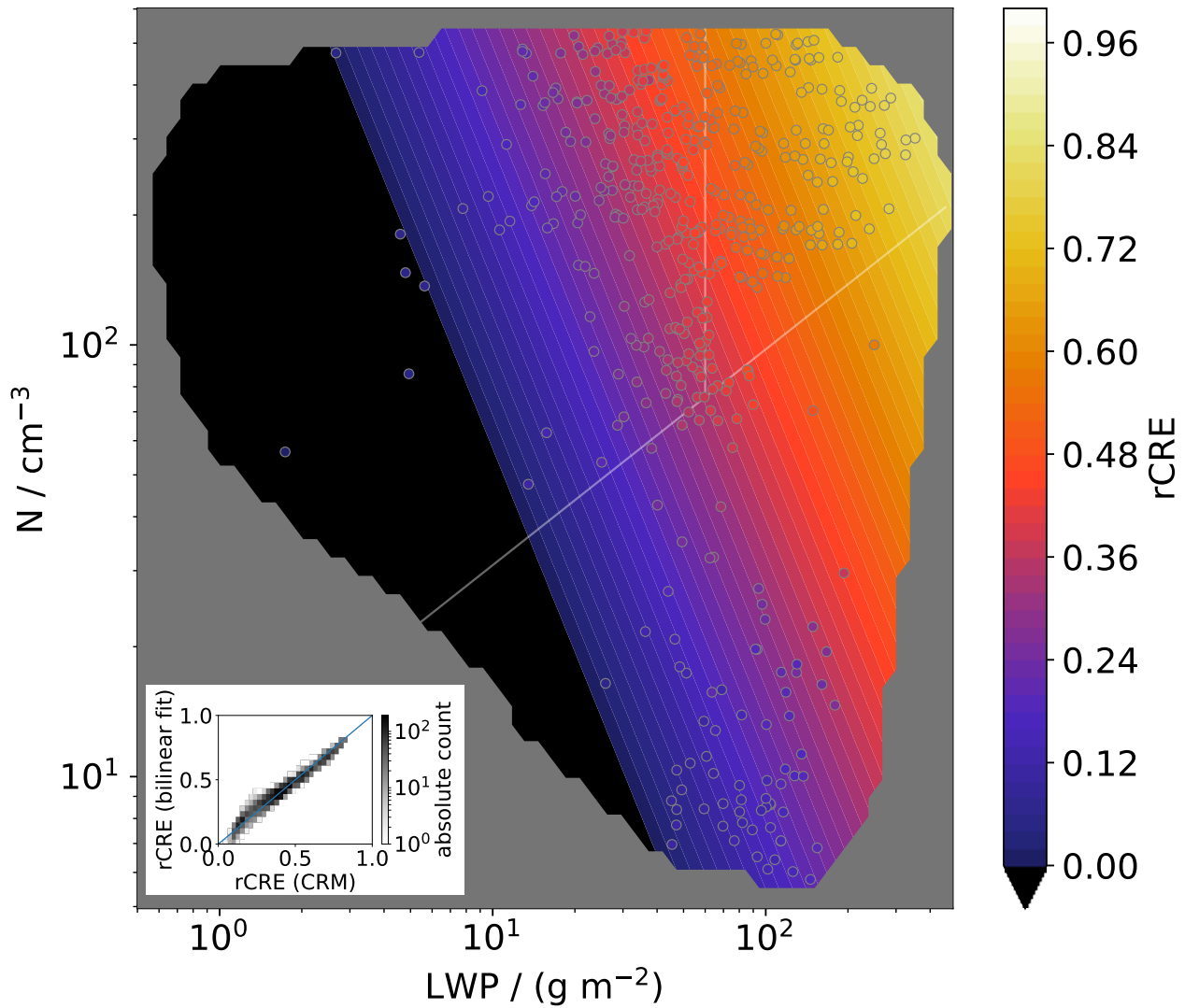



Figure 9. As Figure 5 but using bilinear regression $\mathrm{rCRE} = a \cdot \log_{10}(\mathrm{LWP}) + b \cdot \log_{10}(N) + c$ instead of emulation to obtain the surface. From the regression, we obtain $a = 0.40$, $b = 0.26$, $c = -0.86$ with a coefficient of determination of $r^2 = 0.95$ and rmse$= 0.05$.

$$\frac{d\ln \mathrm{rCRE}}{d\ln \mathrm{LWP}} = \frac{5}{6}\left(1 - \mathrm{rCRE}\right), \qquad\qquad \frac{d\ln \mathrm{rCRE}}{d\ln N} = \frac{1}{3}\left(1 - \mathrm{rCRE}\right). \tag{5}$$

In accordance with the different pre-factors, rCRE is thus more susceptible to LWP than to N. Susceptibilities decrease with increasing LWP, reflecting the saturation behavior of A_C for high LWP.

While Feingold et al. (1997) discuss the effect of drizzle initiation on A_C and precipitation susceptibility, the authors are not aware of studies addressing susceptibilities of rCRE, A_C, CF or related quantities under drizzling conditions. Our emulator approach enables us to do so. As discussed in the context of Figure 6, the contribution of rain water to total LWP leads to a shift of rCRE isolines and makes cloud fraction at constant LWP a function of N. This creates a maximum in $\partial \ln \mathrm{rCRE}/\partial \ln N$ for fixed LWP in the vicinity of the isoline distortion (Figure 10a). It also explains why isolines of $\partial \ln \mathrm{rCRE}/\partial \ln \mathrm{LWP}$ are tilted in the drizzling region. This indicates that the partial susceptibility of rCRE to N not only captures the radiative effects of droplet size but also accounts for cloud fraction changes, while the partial susceptibility to LWP is comparably insensitive to the latter.

We abstain from interpreting the susceptibility in the left half of Q2; due to the low sample size in this region, the observed substructure is likely not physical.

6   Conclusions

We present a new method to summarize the detailed process representation ingrained in LES into a simple picture of aerosol-cloud interactions (Equation 2). We have constructed ensembles of Gaussian process emulators to extract how cloud albedo A_C, cloud fraction CF, and the relative cloud radiative effect rCRE (Equation 1) depend on the domain-averaged liquid water path LWP and vertically-averaged cloud droplet number concentration N from a set of 159 cloud-resolving simulations of nocturnal stratocumulus (Sc) with different initial conditions (Figure 1). The initial conditions were Latin-hypercube sampled from a 6-dimensional space that took into account variations of moisture and temperature profiles, including their jumps at the inversion, as well as aerosol concentration and boundary layer height. The emulator-ensemble approach has enabled us to accurately capture A_C, CF and rCRE as a function of LWP and N over a wide range of LWP and N (Figures 5 and 6). Our results are based on an idealized set of simulations that currently do not account for varying boundary conditions like subsidence and surface fluxes. Taking such differences into account may lead to a broader and/or denser sampling of the LWP-N state space. This would extend and improve the emulated surfaces.

Emulation provides a viable and more powerful alternative to multivariate linear regression for deriving cloud-state-dependent partial susceptibilities (Equation 2). We demonstrate this for the partial susceptibilities of rCRE to LWP and N (Figure 10). We reproduce theoretical results for full cloud cover and monomodal droplet size distributions and extend the known relationships into the drizzling regime. As cloud fraction remains controlled by cloud water, the contribution of rain water to total LWP leads to a strong dependence of cloud fraction on N, for fixed LWP. This dependence corresponds to a strong susceptibility of rCRE to N in the transition region from solid to broken cloud cover.

Our results confirm the expectation that rCRE is most susceptible to microphysical perturbations in the transition region between the high- and low-cloud fraction regime of Sc. Our new approach allows us to clarify the interpretation of Equation 2:

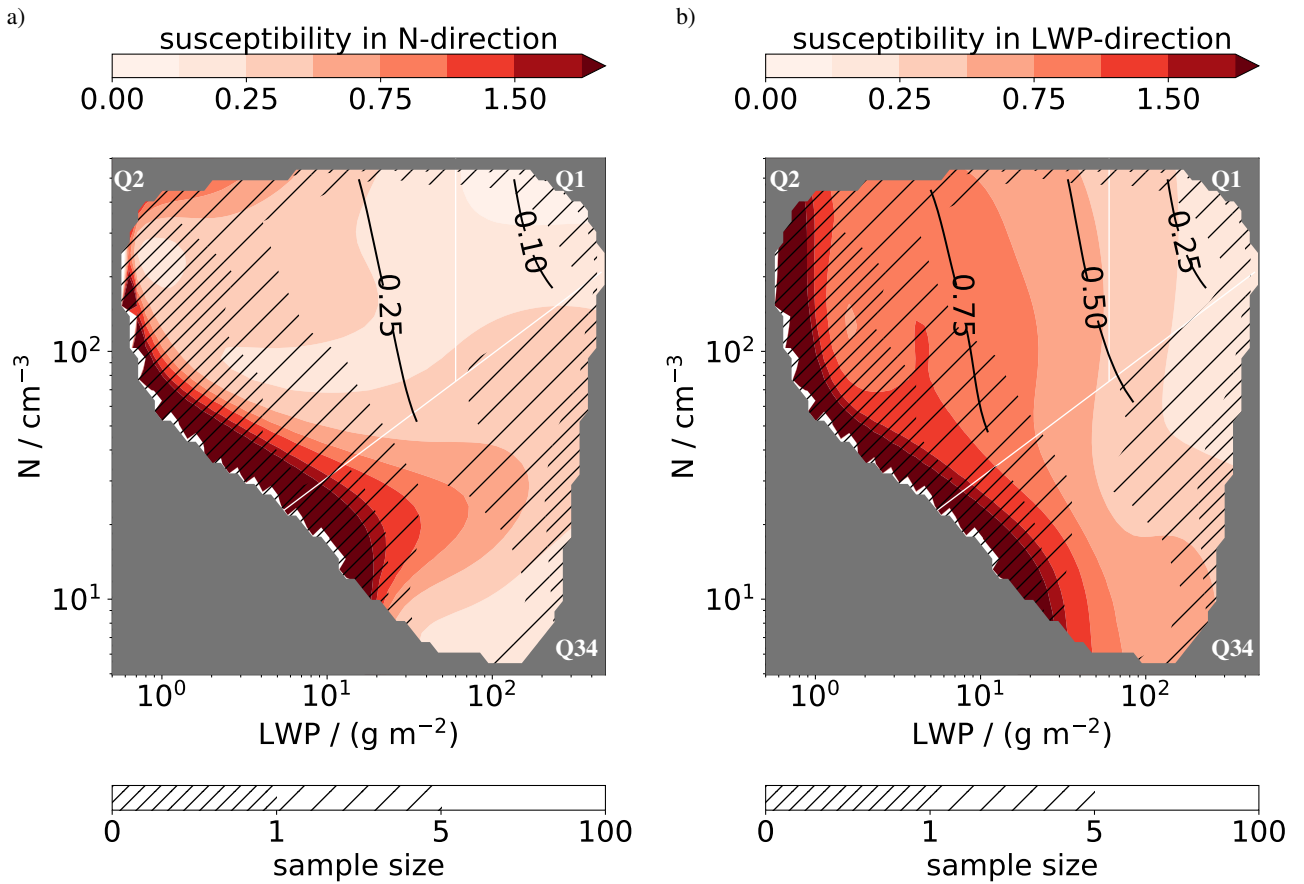



Figure 10. Partial susceptibilities of rCRE to (a) N and (b) LWP as a function of N and LWP as derived from the emulated rCRE surface (color contours). Solid lines show theoretical susceptibility values following Boers and Mitchell (1994), restricted to the non-drizzling region for which they apply. Hatchings indicate the sample size as in Figure 7. Note that sample sizes smaller than one trajectory can occur due to interpolation. White lines indicate quadrants as in Figure 1.

The direct contribution of droplet number changes to rCRE, $\partial \ln \mathrm{rCRE}/\partial \ln N$, captures the effect of droplet number on cloud brightness as well as the effect on cloud fraction. This is possible because N controls the rain water fraction RWP/LWP at constant LWP. The adjustment contribution, $\partial \ln \mathrm{rCRE}/\partial \ln \mathrm{LWP} \cdot \mathrm{d} \ln \mathrm{LWP}/\mathrm{d} \ln N$, captures the effect of cloud amount on rCRE, irrespective of its distribution onto fewer or more droplets, or thicker or thinner clouds at low or high cloud cover.

5   The methodology presented provides a powerful tool for synthesizing detailed data into a simple predictive framework. In this paper we have demonstrated its versatility for studying the sensitivities of cloud field properties over a wide range of states. Subsequent work will focus on employing this emulator approach to gain a deeper understanding of LWP adjustments $\mathrm{d} \ln \mathrm{LWP}/\mathrm{d} \ln N$. In general, computation-statistical approaches like the one discussed here have broad potential. They enable process modelers to explore their models beyond case studies, while at the same time they empower empiricists to better

account for state dependence. This combination of approaches provides a promising avenue for improving our understanding of the uncertainties associated with the representation of shallow clouds in climate models.

Code and data availability. Input files and the model code for reproducing the simulation data of this study are available from the authors upon request.

5 *Author contributions.* FG carried out the analysis. All authors contributed to developing the basic ideas, discussing the results, and preparing the manuscript.

Competing interests. KC is executive editor, GF a co-editor of ACP. Other than this, the authors declare that they have no conflict of interests.

Acknowledgements. FG acknowledges support by a National Research Council Research Associateship award at the National Oceanic and Atmospheric Administration (NOAA) and by The Branco Weiss Fellowship – Society in Science, administered by the ETH Zürich. FH holds
10 a visiting fellowship of the Cooperative Institute for Research in Environmental Sciences (CIRES) at the University of Colorado Boulder and the NOAA/Earth System Research Laboratory. TY is supported by the U.S. National Oceanic and Atmospheric Administration (NOAA) Climate Program Office and by NOAA's Climate Goal. JJ and KC were supported by the Natural Environment Research Council (NERC) under grant NE/I020059/1 (ACID-PRUF) and the UK-China Research and Innovation Partnership Fund through the Met Office Climate Science for Service Partnership (CSSP) China as part of the Newton Fund. KC is currently a Royal Society Wolfson Merit Award holder.
15 This research was partially supported by the Office of Biological and Environmental Research of the U.S. Department of Energy Atmospheric System Research Program Interagency Agreement DE-SC0016275. Marat Khairoutdinov graciously provided the SAM model.

References

Ackerman, A. S., vanZanten, M. C., Stevens, B., Savic-Jovcic, V., Bretherton, C. S., Chlond, A., Golaz, J.-C., Jiang, H., Khairoutdinov, M., Krueger, S. K., and et al.: Large-Eddy Simulations of a Drizzling, Stratocumulus-Topped Marine Boundary Layer, Mon. Weather Rev., 137, 1083–1110, https://doi.org/10.1175/2008mwr2582.1, 2009.

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

Beven, K.: A manifesto for the equifinality thesis, J Hydrol, 320, 18–36, https://doi.org/doi.org/10.1016/j.jhydrol.2005.07.007, 2005.

Boers, R. and Mitchell, R. M.: Absorption feedback in stratocumulus clouds Influence on cloud top albedo, Tellus A: Dynamic Meteorology and Oceanography, 46, 229–241, https://doi.org/10.3402/tellusa.v46i3.15476, 1994.

10 Boucher, O., Randall, D., Artaxo, P., Bretherton, C., Feingold, G., Forster, P., Kerminen, V.-M., Kondo, Y., Liao, H., Lohmann, U., Rasch, P., Satheesh, S., Sherwood, S., Stevens, B., and Zhang, X.: Clouds and Aerosols, in: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to IPCC AR5, edited by Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P., Cambridge, https://doi.org/10.1017/CBO9781107415324, 2013.

Brenowitz, N. D. and Bretherton, C. S.: Prognostic Validation of a Neural Network Unified Physics Parameterization, Geophys. Res. Lett.,
15 45, 6289–6298, 2018.

Carnell, R.: lhs: Latin Hypercube Samples, https://CRAN.R-project.org/package=lhs, 2018.

Chen, Y.-C., Christensen, M. W., Stephens, G. L., and Seinfeld, J. H.: Satellite-based estimate of global aerosol-cloud radiative forcing by marine warm clouds, Nature Geosci., 7, 643–646, 2014.

Christensen, M. W. and Stephens, G. L.: Microphysical and macrophysical responses of marine stratocumulus polluted by underlying ships:
20 evidence of cloud deepening, J. Geophys. Res., 116, 2011.

Feingold, G., Boers, R., Stevens, B., and Cotton, W. R.: A modeling study of the effect of drizzle on cloud optical depth and susceptibility, J. Geophys. Res., 102, 13 527–13 534, 1997.

Feingold, G., Walko, R. L., Stevens, B., and Cotton, W. R.: Simulations of marine stratocumulus using a new microphysical parameterization scheme, Atmos. Res., 47–48, 505–528, 1998.

25 Feingold, G., Koren, I., Yamaguchi, T., and Kazil, J.: On the reversibility of transitions between closed and open cellular convection, Atmos. Chem. Phys., 15, 7351–7367, https://doi.org/10.5194/acp-15-7351-2015, 2015.

Feingold, G., McComiskey, A., Yamaguchi, T., Johnson, J. S., Carslaw, K. S., and Schmidt, K. S.: New approaches to quantifying aerosol influence on the cloud radiative effect, Proc Natl Acad Sci U S A, https://doi.org/10.1073/pnas.1514035112, 2016.

Gentine, P., Pritchard, M., Rasp, S., Reinaudi, G., and Yacalls, G.: Could Machine Learning Break the Convection Parameterization Dead-
30 lock?, Geophys. Res. Lett., 45, 5742–5751, 2018.

Gerber, H.: Microphysics of Marine Stratocumulus Clouds with Two Drizzle Modes, J. Atmos. Sci., 53, 1649–1662, 1996.

Glassmeier, F. and Lohmann, U.: Precipitation susceptibility and aerosol buffering of warm and mixed-phase orographic clouds in idealized simulations, J. Atmos. Sci., 75, 1173–1194, https://doi.org/10.1175/JAS-D-17-0254.1, 2018.

Harte, J.: Toward a Synthesis of the Newtonian and Darwinian Worldviews, Phys. Today, 55, 29, 2002.

35 Jiang, H., Feingold, G., and Sorooshian, A.: Effect of Aerosol on the Susceptibility and Efficiency of Precipitation in Warm Trade Cumulus Clouds, Journal of the Atmospheric Sciences, 67, 3525–3540, https://doi.org/10.1175/2010jas3484.1, 2010.

Johnson, J. S., Cui, Z., Lee, L. A., Gosling, J. P., Blyth, A. M., and Carslaw, K. S.: Evaluating uncertainty in convective cloud microphysics using statistical emulation, J. Adv. Model. Earth Syst., 7, 162–187, https://doi.org/10.1002/ 2014MS000383, 2015.

Kaufman, Y. J., Koren, I., Remer, L. A., Rosenfeld, D., and Rudich, Y.: The effect of smoke, dust, and pollution aerosol on shallow cloud development over the Atlantic Ocean, Proceedings of the National Academy of Sciences, 102, 11 207–11 212, https://doi.org/10.1073/pnas.0505191102, http://dx.doi.org/10.1073/pnas.0505191102, 2005.

Kazil, J., Wang, H., Feingold, G., Clarke, A. D., Snider, J. R., and Bandy, A. R.: Modeling chemical and aerosol processes in the transition from closed to open cells during VOCALS-REx, Atmos. Chem. Phys., 11, 7491–7514, https://doi.org/10.5194/acp-11-7491-2011, 2011.

Khairoutdinov, M. F. and Randall, D. A.: Cloud resolving modeling of the ARM Summer 1997 IOP: Model formulation, results, uncertainties, and sensitivities, J. Atmos. Sci., 60, 607–625, https://doi.org/10.1175/1520-0469(2003)060<0607:CRMOTA>2.0.CO;2, 2003.

Krasnopolsky, V. M., Fox-Rabinovitz, M. S., and Belochitski, A. A.: Using ensemble of neural networks to learn stochastic convection parameterizations from climate and numerical weather prediction models from data simulated by a cloud resolving model, Advances in Artificial Neural Systems, 2013.

Lee, L. A., Carslaw, K. S., Pringle, K. J., Mann, G. W., and Spracklen, D. V.: Emulation of a complex global aerosol model to quantify sensitivity to uncertain parameters, Atmos. Chem. Phys., 11, 12 253–12 273, https://doi.org/10.5194/acp-11-12253-2011, 2011.

Lee, L. A., Pringle, K. J., Reddington, C. L., Mann, G. W., Stier, P., Spracklen, D. V., Pierce, J. R., and Carslaw, K. S.: The magnitude and causes of uncertainty in global model simulations of cloud condensation nuclei, Atmospheric Chemistry and Physics, 13, 8879–8914, https://doi.org/10.5194/acp-13-8879-2013, 2013.

Matheson, M. A., Coakley, J. A., and Tahnk, W. R.: Aerosol and cloud property relationships for summertime stratiform clouds in the northeastern Atlantic from Advanced Very High Resolution Radiometer observations, J. Geophys. Res., 110, 2005.

McGibbon, J. and Bretherton, C. S.: Skill of ship-following large-eddy simulations in reproducing MAGIC observations across the northeast Pacific stratocumulus to cumulus transition region, J Adv Model Earth Syst, 2017.

Morris, M. D. and Mitchell, T. J.: Exploratory designs for computer experiments, J. Stat. Plann. Inference, 43, 381–402, https://doi.org/10.1016/0378–3758(94)00035-T, 1995.

Mülmenstädt, J. and Feingold, G.: The radiative forcing of aerosol-cloud interactions in liquid clouds: wrestling and embracing uncertainty, Current Climate Change Reports, https://doi.org/doi.org/10.1007/s40641-018-0089-y, 2018.

Myhre, G., Shindell, D., Bréon, F.-M., Collins, W., Fuglestvedt, J., Huang, J., Koch, D., Lamarque, J.-F., Lee, D., Mendoza, B., Nakajima, T., Robock, A., Stephens, G., Takemura, T., and Zhang, H.: Anthropogenic and Natural Radiative Forcing, in: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to IPCC AR5, edited by Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P., Cambridge, https://doi.org/10.1017/CBO9781107415324, 2013.

O'Gorman, P. A. and Dwyer, J. G.: Using machine learning to parameterize moist convection: potential for modeling of climate change and extreme events, J Adv Model Earth Syst, 10, 2548–2563, https://doi.org/doi.org/10.1029/2018MS001351, 2018.

O'Hagan, A.: Bayesian analysis of computer code outputs: A tutorial, Reliab. Eng. Syst. Safe., 38, 1290–1300, https://doi.org/10.1016/j.ress.2005.11.025, 2006.

Platnick, S. and Twomey, S.: Determining the Susceptibility of Cloud Albedo to Changes in Droplet Concentration with the Advanced Very High Resolution Radiometer, J. Appl. Meteor., 33, 334 – 347, 1994.

Posselt, D. J., Fryxell, B., Molod, A., and Williams, B.: Quantitative Sensitivity Analysis of Physical Parameterizations for Cases of Deep Convection in the NASA GEOS-5, Journal of Climate, 29, 455–479, https://doi.org/10.1175/jcli-d-15-0250.1, 2016.

R Core Team: R: A Language and Environment for Statistical Computing,, R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/, 2018.

Rasmussen, C. E. and Williams, C. K. I.: Gaussian Processes for Machine Learning, MIT press, London, 2006.

Roustant, O., Ginsbourger, D., and Deville, Y.: DiceKriging, DiceOptim: Two R Packages for the Analysis of Computer Experiments by Kriging-Based Metamodeling and Optimization, Journal of Statistical Software, 51, 1–55, http://www.jstatsoft.org/v51/i01/, 2012.

Schneider, T., Lan, S., Stuart, A., and Teixeira, J.: Earth System Modeling 2.0: A Blueprint for models that learn from observations and targeted high-resolution simulations, Geophys. Res. Lett., 2017.

Sena, E. T., McComiskey, A., and Feingold, G.: A long-term study of aerosol–cloud interactions and their radiative effect at the Southern Great Plains using ground-based measurements, Atmospheric Chemistry and Physics, 16, 11 301–11 318, https://doi.org/10.5194/acp-16-11301-2016, 2016.

Small, J. D., Chuang, P. Y., Feingold, G., and Jiang, H.: Can aerosol decrease cloud lifetime?, Geophys. Res. Lett., 36, 2009.

Stevens, B. and Feingold, G.: Untangling aerosol effects on clouds and precipitation in a buffered system, Nature, 461, 607–613, https://doi.org/10.1038/nature08281, 2009.

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

Twomey, S.: The Influence of Pollution on the Shortwave Albedo of Clouds, J. Atmos. Sci., 34, 1149 – 1152, 1977.

Xie, Y. and Liu, Y.: A new approach for simultaneously retrieving cloud albedo and cloud fraction from surface-based shortwave radiation measurements, Environmental Research Letters, 8, 044 023, https://doi.org/10.1088/1748-9326/8/4/044023, 2013.

Yamaguchi, T., Feingold, G., and Kazil, J.: Stratocumulus to Cumulus Transition by Drizzle, Journal of Advances in Modeling Earth Systems, https://doi.org/10.1002/2017ms001104, 2017.

Zheng, X., Albrecht, B., Minnis, P., Ayers, K., and Jonson, H. H.: Observed aerosol and liquid water path relationships in marine stratocumulus, Geophys. Res. Lett., 37, 2010.