An emulator of stratocumulus cloud response to two cloud-controlling factors accounting for natural variability.

Rachel W. N. Sansom[1], Ken S Carslaw[2], Lindsay Lee[3], and Jill S. Johnson[4]

[1]University of Leeds
[2]Institute for Climate and Atmospheric Science
[3]Advanced Manufacturing Research Centre
[4]University of Sheffield

January 16, 2024

Abstract

Large uncertainties persist in modeling shallow, low clouds because there are many interacting nonlinear processes and multiple cloud-controlling environmental factors. In addition, sharp changes in behavior can occur when environmental thresholds are met. Model studies that follow a traditional approach of exploring the effects of factors "one-at-a-time" are unable to capture interactions between factors. We simulate a stratocumulus cloud based on the Second Dynamics and Chemistry of Marine Stratocumulus field study using a large-eddy simulation model coupled with a two-moment cloud microphysics scheme. The simulations are used to train a Gaussian process emulator, which we then use to visualize the relationships between two cloud-controlling factors and domain-averaged cloud properties. Only 29 model simulations were required to train the emulators, which then predicted cloud properties at thousands of new combinations of the two factors. Emulator response surfaces of cloud liquid water path and cloud fraction show two behavioral regimes, one of thin and patchy yet steady stratocumulus and one of thick, growing stratocumulus with cloud fraction near 1. Natural variability (initial-condition uncertainty) creates unrealistic "bumpy" response surfaces. However, we show that the variability causing the bumpiness can be characterized in an emulator "nugget term" that is adjusted to match the distribution of a small number of initial-condition ensemble simulations at various points on the surface, thereby allowing a smoother, deterministic response surface to be constructed. Accounting for variability leads to the firm conclusion that there is a smooth but steep change in cloud behavior between regimes, but not a sharp transition.

An emulator of stratocumulus cloud response to two cloud-controlling factors accounting for natural variability.

Rachel W. N. Sansom[1], Ken S. Carslaw[1], Jill S. Johnson[2], Lindsay Lee[3]

[1]School of Earth and Environment, University of Leeds, Leeds, UK
[2]School of Mathematics and Statistics, University of Sheffield, Sheffield, UK
[3]Advanced Manufacturing Research Centre, University of Sheffield, Sheffield, UK

Key Points:

- Gaussian process emulation of large eddy simulations can be used to visualize shallow cloud response to cloud-controlling factors.
- The emulator response surface shows how above-cloud temperature and moisture have a combined effect on stratocumulus cloud liquid water path.
- Cloud natural variability can be accounted for so that the emulator smoothly captures the deterministic rather variation in cloud properties.

Corresponding author: Rachel W. N. Sansom, `r.sansom@leeds.ac.uk`

Abstract

Large uncertainties persist in modeling shallow, low clouds because there are many interacting nonlinear processes and multiple cloud-controlling environmental factors. In addition, sharp changes in behavior can occur when environmental thresholds are met. Model studies that follow a traditional approach of exploring the effects of factors "one-at-a-time" are unable to capture interactions between factors. We simulate a stratocumulus cloud based on the Second Dynamics and Chemistry of Marine Stratocumulus field study using a large-eddy simulation model coupled with a two-moment cloud microphysics scheme. The simulations are used to train a Gaussian process emulator, which we then use to visualize the relationships between two cloud-controlling factors and domain-averaged cloud properties. Only 29 model simulations were required to train the emulators, which then predicted cloud properties at thousands of new combinations of the two factors. Emulator response surfaces of cloud liquid water path and cloud fraction show two behavioral regimes, one of thin and patchy yet steady stratocumulus and one of thick, growing stratocumulus with cloud fraction near 1. Natural variability (initial-condition uncertainty) creates unrealistic "bumpy" response surfaces. However, we show that the variability causing the bumpiness can be characterized in an emulator "nugget term" that is adjusted to match the distribution of a small number of initial-condition ensemble simulations at various points on the surface, thereby allowing a smoother, deterministic response surface to be constructed. Accounting for variability leads to the firm conclusion that there is a smooth but steep change in cloud behavior between regimes, but not a sharp transition.

Plain Language Summary

Modeling shallow clouds is important because these clouds have an overall cooling effect on the planet, but the magnitude of this effect is very uncertain. Shallow cloud behaviors are made up of many interacting processes that work at a large range of scales, and we are not able to fully describe all these processes in large climate models. We use a machine learning technique called Gaussian Process emulation to approximate output from the high-resolution cloud model so that we can study the cloud behavior at a much-reduced computational cost. By perturbing two cloud-controlling factors, we create a set of training data to train the emulator how to approximate the relationship between these factors and cloud properties of interest. We produce color maps that visually describe this relationship and we show that there are two regimes of cloud behavior. We assess the best way to account for cloud natural variability in the approximation. Gaussian Process emulation is a vital tool for cloud modeling because it allows us to inspect whole cloud processes and the relationships between different inputs as they influence the output of interest. In this way we can better understand cloud processes and their uncertainties.

1 Introduction

Shallow, low clouds cover a larger area of the Earth than any other cloud type, with stratocumulus clouds alone covering one-fifth of the surface. They increase Earth's albedo in most regions because they reflect more solar radiation than the underlying surfaces (Wood, 2012), while having only a small effect on emission of terrestrial radiation. Therefore, globally, they have a net cooling effect (Hartmann et al., 1992). These clouds are important for the global radiation budget and how it changes over time in response to warming (cloud feedback: Ceppi et al., 2017; Schneider et al., 2019; Bretherton, 2015; Shen et al., 2022) and changes in aerosols (radiative forcing: Bellouin et al., 2020; J. Smith et al., 2020; Douglas & L'Ecuyer, 2020; Malavelle et al., 2017). However, their responses to changes in aerosols and the thermodynamic environment (cloud-controlling factors) are very uncertain (Myhre et al., 2013). Consequently the corresponding aerosol-cloud radiative forcing (Seinfeld et al., 2016; Lund et al., 2019; Bellouin et al., 2020) and cloud feedbacks (Bony & Dufresne, 2005; Zhang et al., 2013; Blossey et al., 2016; Nuijens & Siebesma, 2019) are not well understood and significantly contribute to the uncertainties that persist in climate change projections (Peace et al., 2020; Dufresne & Bony, 2008). It is crucial that we efficiently use the modeling tools available to narrow this uncertainty in the outcomes of perturbations and climate feedbacks.

Much of the uncertainty in simulating clouds comes from the large number of interacting cloud-controlling factors. Key factors that affect the state and evolution of shallow clouds are local meteorology, large-scale forcings, radiative feedbacks and aerosols. Some of these factors, such as thermodynamic properties, can change on short timescales (hours) and shallow clouds respond quickly because internal changes in cloud microphysics and precipitation operate on similar timescales. Such cloud-controlling factors can have a dramatic effect on cloud properties, such as the rapid change from closed- to open-cell cloud structures (Stechmann & Hottovy, 2016). Other factors, such as large-scale divergence, operate on longer timescales and it can take 2 to 5 days for the cloud to adjust (Bellon & Stevens, 2013). Many of these factors covary and they also have joint effects on cloud processes, with counteracting effects creating a "buffered system" (Stevens & Feingold, 2009). In such a complex interacting system, changing one factor at a time to test cloud responses to various drivers cannot fully capture joint effects and interactions.

Our study has a similar focus to Dal Gesso et al. (2015), a model intercomparison that explored how stratocumulus cloud properties depend on two cloud-controlling factors (henceforth used interchangeably with "parameters"): the temperature and humidity differences between the surface and the free troposphere. The initial profiles of these factors were perturbed over a range of values at discrete Cartesian grid points across the 2-dimensional parameter space to study the effect on model outputs, such as cloud cover and liquid water content. This array of discrete model outputs across the parameter space allowed the model response to be partially visualized. However, this grid-point method restricts the information available from the simulation ensemble and, the number of simulations required to explore n factors also rises with 2^n. Additionally, as shown in Feingold et al. (2016), such a design of simulations may misrepresent the joint effects of factors.

To overcome the limitations of one-at-a-time sensitivity testing and to understand the joint effects of factors, we use Gaussian process emulation to generate "response surfaces" that describe how cloud properties respond to the joint effects of multiple cloud-controlling factors. Gaussian process emulation is a machine learning method to approximate the relationship between a set of model input parameters and a model output (O'Hagan, 2006). Compared to other machine learning methods, this requires only a small number of well-designed model simulations as training data. The emulator function (the approximated relationship between model outputs and inputs) can then be sampled millions of times at a fraction of the computational cost of running the model for the equivalent points in parameter space. From this dense sampling, we can produce a response surface with an associated uncertainty at any point in parameter space. The power of emulation is in the ability to study how large numbers of parameters interact to influence the output of interest and also to visualize all combinations of parameters within their realistic ranges at comparatively low computational cost. In previous emulation work, the parameters were often related to uncertain processes in the model, but here the parameters are cloud-controlling factors.

Gaussian process emulation has been widely applied in aerosol and aerosol-cloud science. First, response surfaces are an effective tool for visualizing the combined effects of the uncertain input parameters and an output of interest, such as in Marshall et al. (2019, 2021) for volcanic eruptions. Transformations from parameter space to state space (Glassmeier et al., 2019; Hoffmann et al., 2020) or selection of a few key parameters at once allows higher dimensions to be visualized (Lee et al., 2011). Second, variance-based sensitivity analysis based on a large number of emulator data points rather than the sparse training data is used to understand which parameters contribute most to the variance in the output of interest (Saltelli et al., 2000; Johnson et al., 2015; Regayre et al., 2014, 2015, 2018; Lee et al., 2011, 2013). Third, the uncertain parameter ranges can be constrained using observations of the model outputs (Johnson et al., 2018; Regayre et al., 2018, 2020; Marshall et al., 2021), which can lead to constraint of additional outputs for which observations are not available.

The first cloud model emulation study was Johnson et al. (2015). They perturbed initial aerosol concentrations and nine microphysical model parameters in a deep convective cloud microphysics model. Sensitivity analysis showed that the cloud properties considered were most sensitive to aerosol concentrations and graupel collection efficiency. This demonstrated the insight that can be gained from emulating cloud models, where buffering can obscure relationships between input parameters and cloud responses. Perturbing multiple input parameters together reveals how they jointly affect an output and under what conditions certain parameters have a larger effect than others. Following this work, emulation has been used to analyze the sensitivity of deep convective cloud properties (Wellmann et al., 2018, 2020) and sea breeze convection (Igel et al., 2018; Park et al., 2020) to initial meteorological conditions. Additionally, Glassmeier et al. (2019) and Hoffmann et al. (2020) have used emulation of state variables to explore cloud-processes in stratocumulus. Here, we use emulation to study the covariance of initial meteorological conditions in stratocumulus and, like Johnson et al. (2015) and Park et al. (2020), we identify regimes of cloud behavior in parameter space.

Shallow clouds often display sharp changes in behavior (between regimes) as cloud-controlling factors change. This can make Gaussian Process emulation challenging because of the required assumptions about smoothness. Feingold et al. (2016) found a steep gradient in a study of nocturnal marine stratocumulus clouds in which six parameters were perturbed. Pope et al. (2021) demonstrated that the steep gradient in this dataset could be emulated using a non-stationary method, where Voronoi tessellations defined regions of the 6-dimensional parameter space where separate, stationary emulators could be applied, which followed the assumption of smoothness. The discontinuity was primarily caused by perturbations in aerosol concentration, but the high dimensionality of the parameter space made visualizing the discontinuity difficult. Here, we have visualized a steep gradient in two dimensions and used adaptive sampling to explore it, but we found that it emulates reasonably so stationary methods sufficed.

Another challenge in visualizing cloud behavior as a response surface is that cloud models exhibit a high degree of natural variability, which may obscure the deterministic behavior that an emulator is designed to represent. In a purely deterministic model, the emulator function can interpolate exactly through all the training data. However, cloud models represent the non-deterministic behavior of clouds through initial small random temperature perturbations, so the simulated cloud properties at each point in parameter space also depend on these initial conditions. Such variability can be averaged out by running initial-condition ensembles at each point in parameter space and using the ensemble mean as training data (Johnson et al., 2011; Oyebamiji et al., 2017; Henderson et al., 2009). For global climate models, which are resource intensive, this variability is usually estimated using maximum likelihood methods (Williamson & Blaker, 2014; Pope et al., 2021). Here, we show the natural variability of our cloud model can be approximated based on initial-condition ensembles at just a few points in parameter space.

In this study, we assess the ability of statistical emulation to capture the transition between two regimes of shallow cloud behavior as initial vertical profiles of two cloud-controlling factors (parameters) are varied. We also explore a method to quantify natural variability and account for it when training emulators. We start from a homogeneous stratocumulus-topped boundary layer that has steady cloud properties despite environmental conditions that make the cloud prone to breaking up, as hypothesized by Lilly

(1968), Randall (1980) and Deardorff (1980). Two parameters are perturbed to identify where cloud breakup occurs across the parameter space. We will answer the following questions. First, does the hypothetical cloud-breakup threshold separate two cloud regimes? Second, how well can we characterize the change in cloud behavior using statistical emulation? Is there a discontinuity or a smooth change? Third, how can we account for the model's natural variability in the emulators so that the response surfaces represent deterministic cloud behavior rather than noise?

The remainder of this paper is laid out as follows. Section 2 gives context to the cloud-breakup region and section 3 describes the model simulation setup, the initial simulation and the parameter perturbations. Section 4 discusses the cloud behavior displayed across the perturbed parameter ensemble and exploring the model's behavior around the cloud-breakup threshold. The model's natural variability will be quantified and included in the emulation method in section 5. The results are discussed further in section 6.

2 Theoretical context

The simulations are based on observations from the first research flight (RF01) of the Second Dynamics and Chemistry of Marine Stratocumulus field study (DYCOMS-II) (Stevens et al., 2003), which took place off the west coast of California in July 2001. Flight RF01 observed a homogeneous, non-drizzling stratocumulus cloud deck over nine hours through the night. Dropsondes measured a well-mixed boundary layer up to 850 m initially, but the boundary layer and cloud layer deepened by 50 m over the course of the flight, resulting in a 250 m thick cloud. The well-mixed stratocumulus-topped boundary layer was capped by a temperature inversion, where the potential temperature, θ, increased by 8.5 K and the total water mass mixing ratio, q_{t}, decreased by 7.5 g kg^{-1} within a few tens of meters of cloud top.

Stevens et al. (2005) conducted a large-eddy simulation (LES) model intercomparison study based on RF01 to compare ten models. Many of the models simulated a more broken cloud than observed, with lower cloud fraction and lower liquid water path (the vertically integrated liquid water content), and some simulated boundary layer decoupling, which was not observed. A decoupled boundary layer is no longer well-mixed and cloud water content tends to decrease because ocean moisture no longer reaches the cloud layer. Stevens et al. (2005) suggested that differences between the models and observations might partially be because the temperature and humidity properties of the inversion made the simulations particularly sensitive to a cloud-dissipating mechanism, cloud-top entrainment instability, described below. Other LES studies that have simulated DYCOMS-II RF01 generally fall within the multi-model range of the intercomparison study (Yamaguchi & Randall, 2008; Xiao et al., 2011; Ghonima et al., 2015; Pressel et al., 2017).

The stratocumulus-topped marine boundary layer can persist as a uniform cloud field for days before transitioning to a cumulus state or breaking up (sometimes entirely) within a couple of hours. Lilly (1968) proposed a theoretical mechanism for this rapid change where warm, dry air mixed into the cloud from above (entrainment) leads to evaporative cooling and enhanced mixing, which may create a positive feedback that can rapidly dissipate the cloud. However, as with DYCOMS-II, many observations and LES studies have found stratocumulus clouds persisting within this theoretical region of cloud dissipation (Kuo & Schubert, 1988; Siems et al., 1990; Moeng, 2000; Stevens et al., 2005). Mellado (2017) summarized recent studies that found the feedback is generally not strong enough under realistic conditions to dissipate marine stratocumulus clouds, especially alongside other confounding factors.

Randall (1980) and Deardorff (1980) derived an inversion instability parameter, κ, with a threshold beyond which the cloud-dissipating feedback occurs,

$$\kappa = 1 + \frac{c_p}{L_\mathrm{V}} \frac{\Delta\theta_\mathrm{l}}{\Delta q_\mathrm{t}}, \tag{1}$$

where c_p is the specific heat of air, L_V is the latent heat of vaporization, $\Delta\theta_\mathrm{l}$ and Δq_t are the changes in potential temperature (for liquid water) and in total water mass mixing ratio, both at the inversion. Several studies since have made alternative derivations and attempted to map out the dependence of κ on these two parameters, $\Delta\theta_\mathrm{l}$ and Δq_t, using one-at-a-time model sensitivity simulations (Kuo & Schubert, 1988; MacVean & Mason, 1990; Siems et al., 1990; Yamaguchi & Randall, 2008; Xiao et al., 2011; Van Der Dussen et al., 2014; Dal Gesso et al., 2015).

Here we simulate DYCOMS-II RF01 and perturb $\Delta\theta_\mathrm{l}$ and Δq_t across a range of values to map out cloud behavior in their joint parameter space.

3 Experiment design

3.1 Model Description

The LES model used here is the UK Met Office/Natural Environment Research Council (NERC) Cloud (MONC) model (N. Brown et al., 2020). All simulations were nocturnal and used a longwave cooling parameterisation based on Bretherton et al. (1999). Horizontal resolution was 30 m and vertical resolution varied between 7.5 m around the inversion and 10 - 20 m elsewhere in the boundary layer. The domain size was 250 by 250 grid boxes with 110 vertical layers up to 1500 m. The subgrid mixing scheme for unresolved turbulence, diffusion and viscosity is an extension of the Smagorinsky-Lilly model (detailed in A. R. Brown et al., 1994).

The microphysics scheme used is the Cloud AeroSol Interacting Microphysics (CASIM) model, which is a bulk scheme that can use up to three moments for each hydrometeor (Shipway & Hill, 2012; Hill et al., 2015; Dearden et al., 2018; Field et al., 2023). Here we define cloud liquid and rain droplets by two moments: number concentration and mass mixing ratio.

The particle size distribution is defined as,

$$N(r) = N_0 r^\mu e^{-\gamma r}, \tag{2}$$

where r is a measure of size, N_0 is the distribution intercept parameter, μ is the shape parameter and γ is the slope parameter (Shipway & Hill, 2012). The k^{th} moment is then defined by,

$$M_k = \int r^k N(r) dr, \tag{3}$$

giving the number concentration (zeroth moment) as $\mathrm{M}_0 = \mathrm{N}_0$ and the mass mixing ratio (third moment) is $\mathrm{M}_3 = \frac{4}{3}\pi r^3 N_0 exp(\frac{9}{2} ln^2\sigma)$ for a lognormal distribution.

Condensation and evaporation were calculated by a saturation adjustment scheme, where any surplus water vapor in the cloud condenses onto the fixed number of cloud droplets and any deficit evaporates from the droplets, keeping the relative humidity within the cloud at 100%. Cloud droplets can be autoconverted and collected into rain droplets, and rain droplets can precipitate and either reach the surface or evaporate in sub-saturated air below the cloud base. Condensation of water vapor onto rain cannot occur due to the saturation adjustment scheme (Gray et al., 2001). Sedimentation was switched on

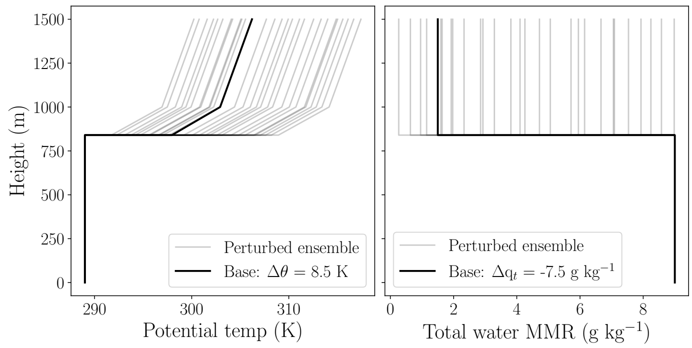

Figure 1. Initial profiles of potential temperature and total water mass mixing ratio for all simulations. Solid lines show the base simulation values taken from the DYCOMS-II observational campaign flight RF01, while gray lines show the profiles for the perturbed ensemble members.

for both cloud droplets and rain, which advects water mass downwards through the boundary layer. Autoconversion and collection are dependent on cloud water mass mixing ratio, rain water mass mixing ratio and droplet number concentration (Khairoutdinov & Kogan, 2000). New air is homogeneously mixed from above the cloud into the cloud layer, which means clear air is mixed into the cloudy air before evaporation is calculated so all droplets are evaporated equally until saturation is reached. Thus, there is a reduction in cloud droplet radius, but cloud droplet number is not affected.

3.2 Perturbed parameter ensemble

A base simulation was initialized to match the DYCOMS-II RF01 setup in Stevens et al. (2005). The simulation was run for 8 hours with initial surface sensible and latent heat fluxes of 15 W m^{-2} and 115 W m^{-2}. The initial profiles of θ and q$_t$ are shown in Figure 1.

We perturbed the initial $\Delta\theta$ and Δq_{t} to explore the joint effect of these two parameters, creating a perturbed parameter ensemble (PPE) in the 2-dimensional parameter space (Figure 1). The initial profiles were kept the same up to the inversion at 840 m, where the magnitude of the jump was varied for both. The ranges for these parameters were chosen based on the ranges outlined in Van Der Dussen et al. (2014):

$$2 \text{ K} \leq \quad \Delta\theta \quad \leq 20 \text{ K} \tag{4}$$
$$-9 \text{ g kg}^{-1} \leq \quad \Delta q_{\mathrm{t}} \quad \leq 0 \text{ g kg}^{-1}. \tag{5}$$

Theoretically, cloud thickening occurs for conditions below the κ threshold which is in the region of parameter space where $\Delta\theta \rightarrow 20$ K and Δ q$_{\mathrm{t}}$ $\rightarrow 0$ g kg^{-1}. Cloud thinning occurs above the κ threshold where $\Delta\theta \rightarrow 2$ K and Δ q$_{\mathrm{t}}$ $\rightarrow$ -9 g kg^{-1}. For a cloud fraction≈1, cloud thickening is roughly analogous to an increasing liquid water path throughout the simulation - a positive liquid water path tendency.

The PPE simulation data were used as training data for Gaussian process emulation. The combination of joint values of $\Delta\theta$ and Δq_{t} were defined using a "maximin" Latin hypercube algorithm comprised of 20 simulations, which has been shown to be sufficient to create an emulator over a 2-dimensional parameter space (Morris & Mitchell, 1995; Loeppky et al., 2009; Lee et al., 2011). The Latin hypercube (Figure 2) is a space-

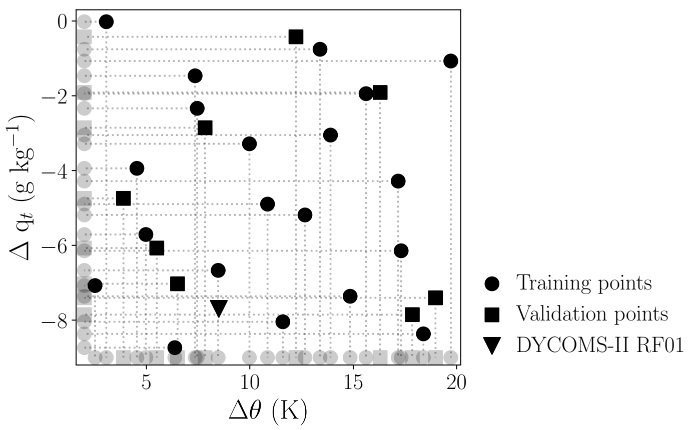

Figure 2. Latin hypercube design for the PPE. Circles show training points and squares show validation points. Faded markers show the distribution along each dimension. The triangle marker is the base simulation.

filling design that samples the parameter space efficiently to provide as much information as possible about the model to an emulator. In comparison, regular gridded (Cartesian) designs are an inefficient way of sampling a high-dimensional space. This becomes crucial if the sensitivity to some parameter perturbations is much greater than for others. When this type of design is projected to lower dimensions, as in Figure 2, design points are not replicated.

3.3 Gaussian process emulation

Gaussian process emulation is a Bayesian statistical approach to generate a mapping between a multi-dimensional input space (the parameters) and an output of interest (O'Hagan, 2006). This mapping can be used to predict the model's output for thousands of new parameter input combinations at a considerably reduced computational cost. It requires training data, consisting of input settings and corresponding output data from multiple model simulations over the perturbed parameter ranges. Under the Bayesian paradigm, the approach is initiated with a prior Gaussian process specification for this mapping, which encapsulates any prior knowledge about the model output. The prior is updated using the training data to create a better estimation of the function representing the model output response (also of Gaussian form) to the perturbed inputs. This better estimation is a posterior Gaussian process specification, and is our emulator which can be thoroughly sampled. The uncertainty surrounding each emulator-predicted output value is also calculated assuming a Gaussian error structure and based on proximity to the training data. A second smaller set of simulations is used to validate the emulator to ensure that it is producing a reasonable representation of the model's behavior. The emulation in this work has been conducted in R using the DiceKriging package (R Core Team, 2018; Roustant et al., 2012).

4 Results

We focus on the response of in-cloud liquid water path and cloud fraction to perturbations in cloud-controlling factors. The cloud fraction is the fraction of columns with liquid water content greater than 0.01 g m^{-3}. The liquid water path and cloud fraction values are calculated as domain means over the final 2 hours of simulation time. The tendencies of both are the rates of change over the 8-hr simulation after disregarding the 2-hour spin-up period.

4.1 Base case simulation

The base simulation has mean liquid water path, L, of 38 g m^{-2} (Figure 3a), similar to the multi-model mean in Stevens et al. (2005) of 40 g m^{-2}. The top-down liquid water path snapshots in Figure 3c (inset) show that the cloud initializes as homogeneous stratocumulus and the cells grow and thicken slightly over 6 hours. The cloud boundaries are mostly constant during the simulation, with cloud base at around 600 m and cloud top at 840 m. L increases through the simulation up to 56 g m^{-2}, so the L tendency is 2.2 g m^{-2} hr^{-1}, whereas it is slightly negative in Stevens et al. (2005). Cloud fraction, f_{c}, starts at 0.87 (Figure 3c) and decreases to 0.72 giving a f_{c} tendency of -0.02 hr^{-1}. The initial f_{c} is in the lowest quartile of the multi-model range in Stevens et al. (2005). The multi-model mean f_{c} begins near 1 and decreases to approximately 0.8, with the majority following similar behavior, but a small number of models simulated a decrease to around 0.2.

4.2 Perturbed cloud behavior

Figures 3a and c show that there is considerable spread across the PPE for both L and f_{c}. The space is split into two regions: A where κ is above the threshold, B where κ is below the threshold. In region A, two simulations do not show any substantial stratocumulus, with $L < 25$ g m^{-2} and $f_{\text{c}} < 0.5$. Four simulations show a very thin stratocumulus with $25 < L < 50$ g m^{-2} and initial $f_{\text{c}} > 0.7$. These clouds thicken slightly through the simulation with L increasing up to 65 g m^{-2} and f_{c} decreasing by 0.1-0.2 as the cloud water aggregates. One point in this region, at coordinates $\Delta\theta = 5$ K and Δq_{t} = -4 g kg^{-1}, is better described alongside the simulations in region B. For those simulations, L begins in the range of 50-80 g m^{-2} and increases by 30-120 g m^{-2}. These clouds all have initial stratocumulus with $f_{\text{c}} > 0.9$ and most remain in that region or decrease to 0.8.

Low $\Delta\theta$ (a weak temperature inversion) generally produces low L, shown in Figure 3b, which reaches a minimum of 21 g m^{-2} at $\Delta\theta = 2.5$ K. With a stronger inversion L also generally increases, up to 160 g m^{-2} at $\Delta\theta = 20$ K. For high Δq_{t} (a moist free troposphere) L is high, up to 185 g m^{-2} for $\Delta q_{\text{t}} = 0$ g kg^{-1}. With a drier free troposphere L is generally lower, down to 20 g m^{-2} for Δq_{t} = -8.7 g kg^{-1}. The two parameters have a combined effect such that L is lowest for weak inversions with a dry free troposphere and highest for strong inversions with humidity similar to the boundary layer.

Low $\Delta\theta$ generally produces low f_{c}, shown in Figure 3d, down to 0.3 at $\Delta\theta = 2.5$ K. As the inversion gets stronger, f_{c} generally increases up to 0.9 at $\Delta\theta = 20$ K. At high Δq_{t} f_{c} approaches 1 for $\Delta q_{\text{t}} = 0$ g kg^{-1}. With a drier free troposphere f_{c} is generally lower, down to about 0.5 for Δq_{t} = -8.7 g kg^{-1}. As with L, the two parameters have a combined effect such that f_{c} is lowest for weak inversions with a dry free troposphere and highest for strong inversions with humidity similar to the boundary layer.

The spatial distribution of L is shown in Figure 4 at the end of each simulation. The two simulations that do not form stratocumulus can be seen to the lower left of the figure as small cumulus clouds. Moving towards higher Δq_{t} and $\Delta\theta$ the plots show stratocumulus with higher L and f_{c}, and in the top row they become quite thick.

None of the simulations are drizzling significantly, but most region A simulations drizzle two to three orders of magnitude less than those in region B. The exception is the point at $\Delta\theta = 5$ K and Δq_{t} = -4 g kg^{-1} previously identified.

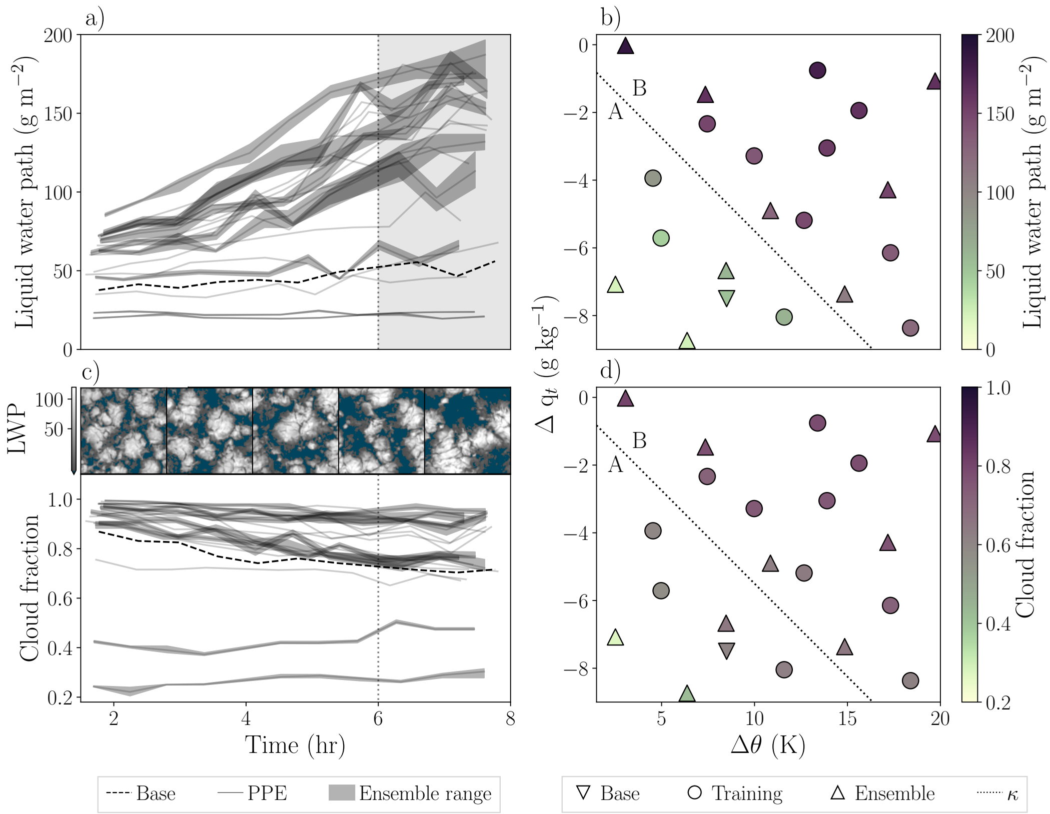

Figure 3. Ensemble model output. Liquid water path a) timeseries post-spinup to the end of the simulation, the last two hours of which is averaged (shaded area) to produce b) the training data in joint parameter space, $\Delta\theta$ vs Δq_t. Cloud fraction c) timeseries and d) training data. The inset in c shows top-down snapshots of the liquid water path for the base simulation. The κ line is the theoretical threshold described in section 2, which splits the regions into A and B.

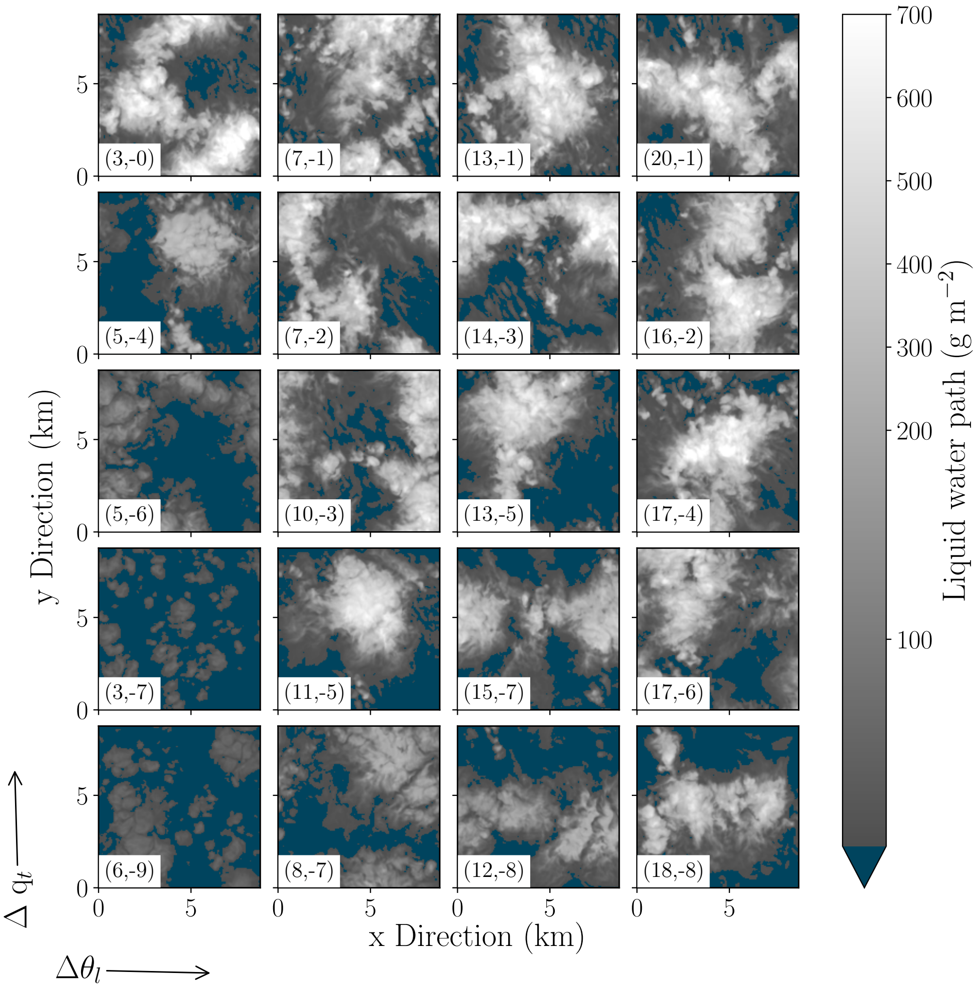

Figure 4. Liquid water path for the last timestep of the first 20 training simulations. Plots are ordered approximately by location in parameter space. The dark blue shows the areas with liquid cloud droplet mass mixing ratio < 0.01 g kg^{-1} at all levels.

In summary, the PPE simulations show that cloud behavior across parameter space falls into two behavioral regimes that are approximately separated by the theoretical κ parameter threshold. Above the κ threshold, in region A, the simulations generally have very thin stratocumulus cloud or small cumulus that show little to no growth throughout the simulation. Below the κ threshold, in region B, the simulations have stratocumulus cloud with a high f_{c} and a medium L that increases throughout the simulation. There is one point in region A at $\Delta\theta = 5$ K and $\Delta q_{\mathrm{t}} = -4$ g kg^{-1} that behaves more like the simulations in region B.

4.3 Response surface analysis

We ran 6 additional simulations to fill gaps in parameter space near the κ threshold and in regions with more extreme model output values. Here we show emulator results from the 26-point training set. We used the PPE to build Gaussian process emulators of L, f_{c}, L tendency and f_{c} tendency. To validate the emulators, we compared the emulator predictions against model output from the validation runs. For all the emulators, the model values are within the emulator prediction 95% uncertainty ranges. The validation results can be found in Figure S1.

The L response surface in Figure 5a follows the behavior described previously by the training data, with low L for low Δq_{t}, low $\Delta\theta$ (dry, cool free troposphere) and high L for high Δq_{t}, high $\Delta\theta$ (moist, warm free troposphere). But the response surface reveals that Δq_{t} has the largest effect on L and for high Δq_{t} L becomes invariant to $\Delta\theta$. There is a local maximum at $\Delta\theta = 15$ K and $\Delta q_{\mathrm{t}} = -4.5$ g kg^{-1}, which we will discuss in section 5. The L tendency response surface in Figure 5b follows a similar pattern to L. The tendency is most positive where the L is high, i.e., where there is most growth. The tendency is very close to zero where the liquid water path is low. It also shows a higher dependency on Δq_{t} and has a local maximum in a similar location. Additionally, the emulator predicts some slightly negative values in the corner of parameter space with low L, however the emulator has limited information at extremities so large uncertainties exist here.

The f_{c} response surface in Figure 5c also follows the behavior described previously, with low f_{c} for low Δq_{t}, low $\Delta\theta$ (dry, cool free troposphere) and high f_{c} for high Δq_{t}, high $\Delta\theta$ (moist, warm free troposphere). As with L, Δq_{t} has a larger effect than $\Delta\theta$, but it is not as stark as in L. The f_{c} tendency is mostly negative across the parameter space, with only a slightly positive region at low values of Δq_{t} and $\Delta\theta$. This is because there are only small cumulus clouds at start of the simulation (Figure 4) and these are mostly unchanging throughout the simulations, but increase in cloud cover slightly. Where f_{c} is approximately 1, f_{c} tendency is close to zero and slightly negative. The rest of the f_{c} tendency surface is very uneven (noisy) since there are only small changes in f_{c} throughout the simulations (Figure 3c). Some of the validation points are close in value to the predicted surface, but a few points are quite contrasting suggesting that this is not very representative of the model behavior.

The κ threshold separating regions A and B approximately follows the surface contours, except for f_{c} tendency. The surfaces show a smooth gradient between these regions of parameter space rather than a discontinuity. In the cloud behavior analysis in section 4.2, there was a single point in region A at $\Delta\theta = 5$ K and $\Delta q_{\mathrm{t}} = -4$ g kg^{-1} that did not fit with the other points in terms of behavior. We can now see in Figure 5a and b that the contours curve round in this part of parameter space.

5 Natural variability

Cloud properties are sensitive to small variations in initial conditions, not just the parameter perturbations we have discussed so far but also slight differences in turbulence

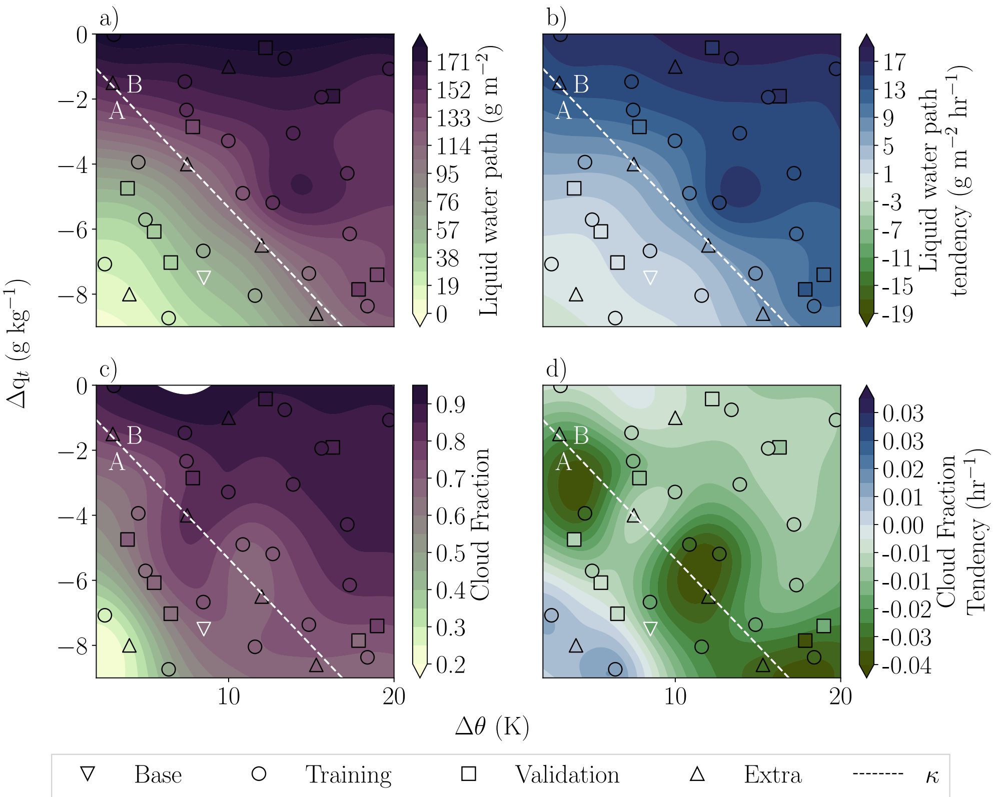

Figure 5. Response surfaces produced from emulator mean predictions. a) liquid water path, b) liquid water path tendency, c) cloud fraction and d) cloud fraction tendency. The base simulation is the inverse white triangle, the training data are the black circles, the validation data are the black squares, and the extra simulation points are the black triangles. The dashed white line is the theoretical κ threshold, which divides parameter space into regions A and B.

structure. Each training datapoint represents only one possible cloud state for those particular parameter settings – i.e., it is effectively a random draw from an unquantified distribution representing natural variability. The effect of this variability is to create a "bumpy" response surface as the emulator interpolates through each model output exactly rather than allowing for the range of possible values at that point (see Figure 6a). In attempting to fit to the training data, the mean function may also distort away from the training points creating additional extrema that are not based on the model's mean deterministic behavior. An example is the maximum around $\Delta\theta = 15$ and Δq_{t} = -4.5 in Figure 5a. Bumps from training points and additional extrema do not allow the emulator to accurately represent the deterministic cloud behavior that we aim to capture with response surfaces.

A much smoother response surface could be created by running an initial-condition ensemble at each point in parameter space and building an emulator of the initial-condition-ensemble mean. The ensembles can be created by randomizing the small temperature perturbations that are imposed at the beginning of each simulation, which initiate turbulence and cloud formation in the boundary layer. However, these ensembles would become very computationally expensive for a large number of parameters. Here, we therefore explore to use a small number of initial-condition ensembles to estimate natural variability and produce smooth, deterministic emulator surfaces by exploiting a hyperparameter within the emulator called a "nugget term". This term allows a smoother response surface because it no longer has to interpolate the points exactly.

In the posterior Gaussian process, the covariance function estimates the uncertainty for any predicted point depending on its distance from the training data. In this study, the covariance between any two points is,

$$V(x_j, x_k) = \sigma^2 K(x_j, x_k), \tag{6}$$

where σ^2 is the variance of the Gaussian process and in this case $K(x_j, x_k)$ represents the Matérn class of covariance functions. The covariance function can be extended to include the nugget term, σ_N^2,

$$V(x_j, x_k) = \sigma^2 K(x_j, x_k) + \sigma_N^2 \delta_{j,k}, \tag{7}$$

where $\delta_{\cdot,\cdot}$ is the Kronecker delta function, which equals 1 for $j = k$ and equals 0 otherwise. The nugget term is often included to alleviate numerical issues in deterministic models, but there are additional benefits to adding one (Andrianakis & Challenor, 2012; Gramacy & Lee, 2012). Practically, the nugget term is a variance that is added at each training point allowing the mean function to vary within that range and no longer interpolate exactly through that point.

For the response surface to most-realistically represent the model's deterministic behavior, we hypothesize that the variance added with the nugget term should be equal to the variance of the model's natural variability. In Figure 6a, the emulator mean function interpolates exactly through the training points resulting in a bumpy response surface, where the residuals between the emulator mean function (the response surface) and model outputs are zero. However, each training point is a single draw from the initial-condition ensemble (Figure 6b). With a nugget term that represents the variance of the initial-condition ensemble, the residuals become spread around zero. Following our hypothesis, we aim to create a surface for which the distribution of emulator residuals has a similar spread to the distribution of the initial-condition members, which represent the natural variability.

We use the variance of the "model" residuals (difference between each ensemble member and the ensemble mean) to estimate an appropriate nugget term for the L emula-

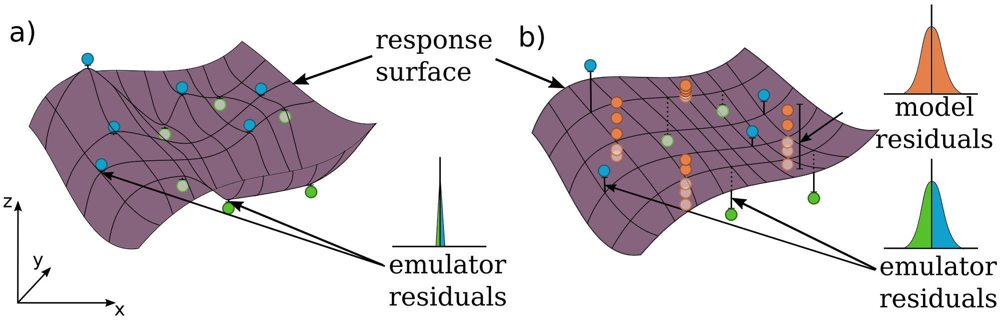

Figure 6. Schematic of the effect of adding a nugget term on the response surface smoothness. a) The purple response surface interpolates exactly through the blue and green training points so the emulator residuals are zero. b) The response surface is smooth after adding a nugget term, so the surface interpolates through a prescribed buffer around the blue and green training points. The nugget term is appropriately large when the distribution of model residuals matches the emulator residuals. Initial-condition ensembles have been run at a selection of points (orange) to gauge the nugget term.

tor as follows. For each of the training points, Z_i, in the training data set $Z_1, ..., Z_l$, running initial-condition ensembles gives a set of estimates, $(Z_i^{(1)}, ..., Z_i^{(k)})$, for k ensemble members at Z_i. Values of L averaged over the last 2 hours in Figure 7a show that the variance increases with the mean value. The ensemble means are calculated as

$$\bar{Z}_i = \frac{1}{k}\sum_{j=1}^{k} Z_i^{(j)}. \tag{8}$$

The variances of the model residuals, σ_i^2, also increase with mean L (Figure 7b). To combine these into one distribution we normalize the model residuals and assume that the normalized variance is constant across the response surface. Note that other data may require a different normalization process, or may already be normal. Here we normalize by dividing by the ensemble means to obtain the normalized model residuals as:

$$r_i^{(j)} = \frac{Z_i^{(j)} - \bar{Z}_i}{\bar{Z}_i}, \tag{9}$$

as shown in Figure 7c. We can then assume that each normalized residual is drawn from a normal distribution, R, with mean μ (=0) and standard deviation σ,

$$R \sim \mathcal{N}(\mu, \sigma_R^2).$$

Our hypothesis is that using the residual distribution's variance, σ_R^2, is an appropriate substitute for using the variance for each initial-condition ensemble, σ_i^2, which we could only know by running ensembles at every training point. We can use the variance of the sample of model residuals to estimate the population variance

$$\sigma_R^2 = \frac{\sum(r-\mu)^2}{N_R}, \tag{10}$$

for the number, N_R, of residuals, r, in the distribution. However, since we normalized the residuals by the ensemble means, we need to multiply by a factor on the same or-

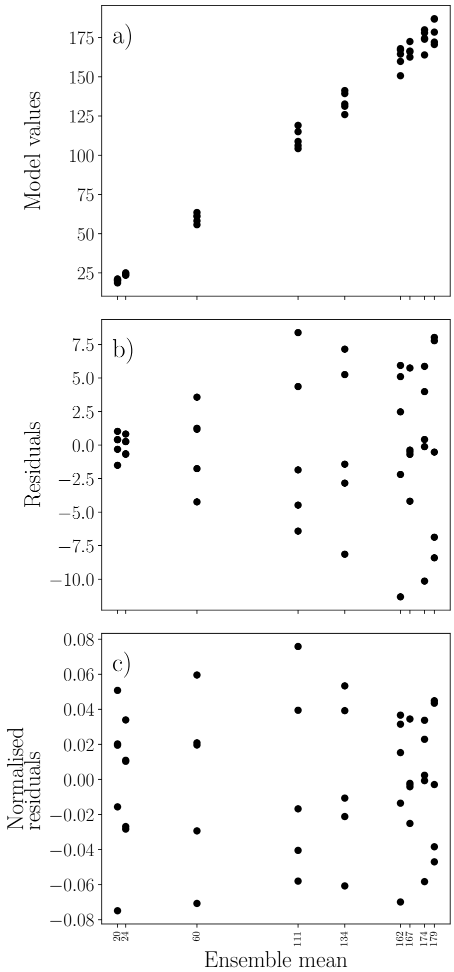

Figure 7. Initial-condition ensemble variance. The nine ensemble mean values against a) the 5-member ensemble model values b) the residual values, and c) the residual values each normalized by the ensemble mean.

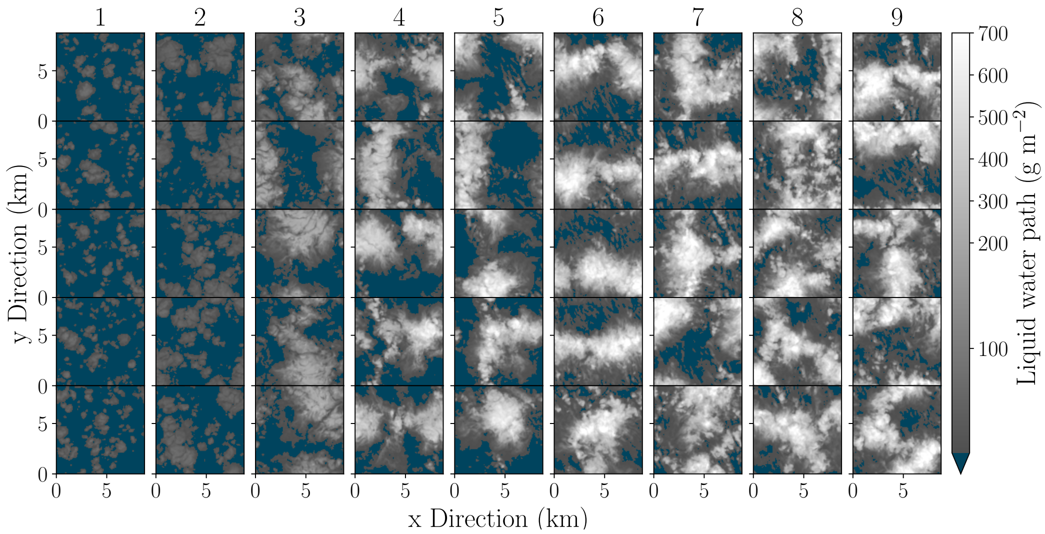

Figure 8. In-cloud liquid water path in the final timestep for each five-member ensemble simulation.

der magnitude as $\bar{Z}_i^2$ to use this estimate in the emulation process (see appendix Appendix A).

We ran 5-member ensembles at nine training points to estimate the natural variability. The timeseries evolution of the ensembles is shown in Figure 3 and the intra-ensemble differences in the simulated final cloud fields are shown in Figure 8. We explored how the estimated variance depends on the number of initial-condition ensemble points and their location in parameter space. A similar result was achieved from initial-condition ensembles at just three well-spaced points, rather than the full nine.

The nugget term is applied as a vector, $V = (v_1, ..., v_l)$, where v_i is an estimate for the variance, σ_i^2, that should be applied at each Z_i. We trialed multiplying our estimate of σ_R^2 by three different squared factors: the value of the cloud property variable at each training point (proportional); the mean value over all training points in the PPE; and the maximum output value from the PPE over the whole parameter space. Two additional noise vectors were tested using arbitrary large numbers of the maximum value $\times 10$ and $\times 100$.

Figure 9 shows that the distribution of emulator residuals compares best with the residual distribution from the model residuals when we use the maximum or $\times 10$ multipliers. The largest overlap of the two distributions (0.78) comes from the maximum multiplier in column d. We used the Kolmogorov-Smirnov (KS) two-sample test to test whether these samples are statistically likely to be from the same distribution. For those with a p-value less than 0.05 we must reject the null hypothesis that the samples are drawn from the same distribution, so only the maximum multiplier and the $\times 10$ terms (columns d and e) fulfill this criteria. Column f shows that the nugget term can be too large and the emulator will default to the underlying prior linear function. This appears as a smooth linearly increasing surface across parameter space and does not fit well with the training data. For smaller nugget variances (columns a, b, c) the distribution of the emulator residuals is narrower than the distribution of the model residuals, showing that the emulator surface is still forced to pass too closely to the training points.

Figure 10 shows how adding an appropriately sized nugget term to the other emulators removes some of the bumpy behavior created by natural variability. We found

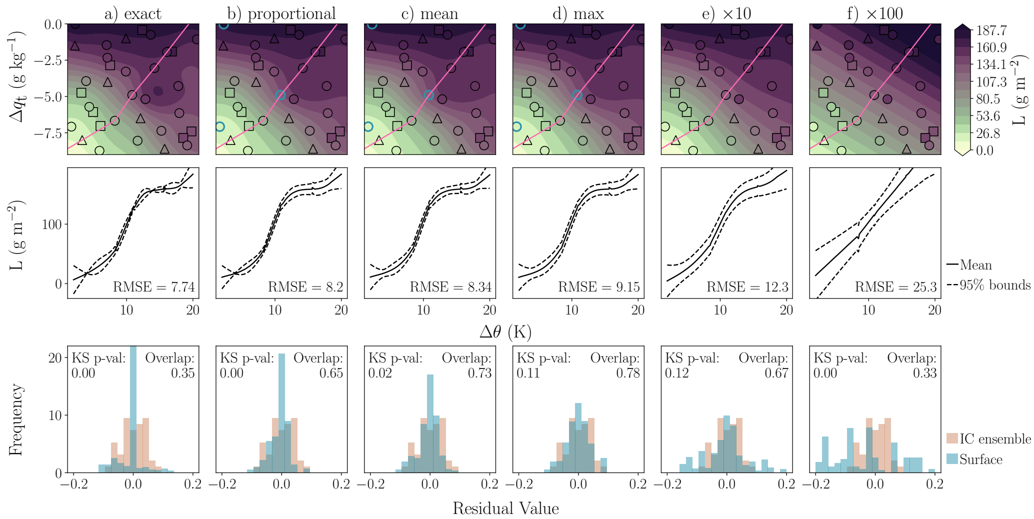

Figure 9. The mean liquid water path emulator with nugget terms applied. a) No nugget term. With a nugget term and multiplying factor a) proportional to the training point value, c) using mean PPE value, d) using maximum PPE value, e) maximum factor ×10, f) maximum factor ×100. Top row: emulator predicted response surfaces with transect line. Middle row: transects along pink line showing mean emulator function and associated uncertainty. Bottom row: histogram comparison of emulator and model residuals (Figure 6). The RMSE, Kolmogorov-Smirnov p-values, and overlap fractions are given for each nugget term.

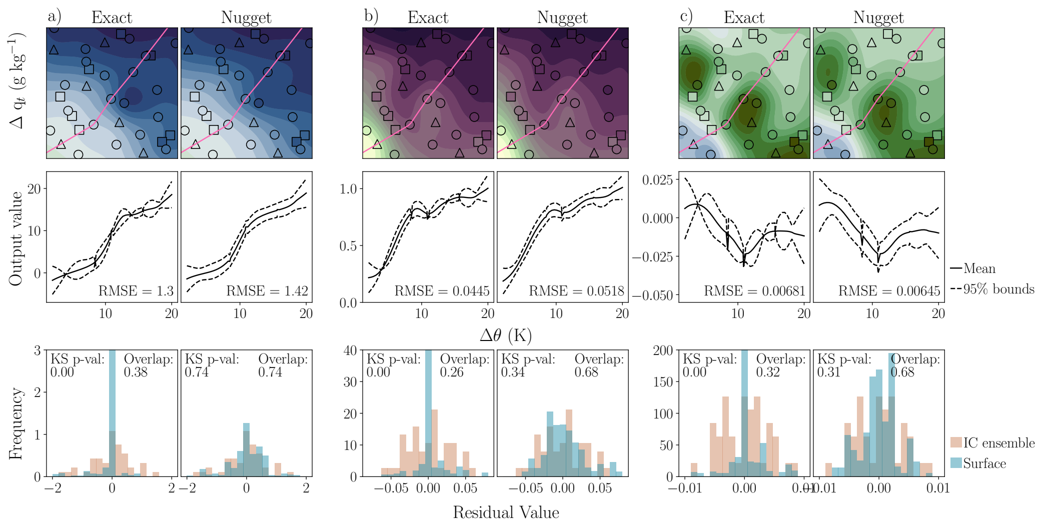

Figure 10. Cloud property emulators with and without nugget terms applied. a) liquid water path tendency, b) cloud fraction and c) cloud fraction tendency. Top row: emulator predicted response surfaces. Middle row: transects along the pink line in top row showing mean emulator function and the associated uncertainty. Bottom row: histogram comparison of emulator and model residuals (Figure 6). The RMSE, Kolmogorov-Smirnov p-values, and overlap fractions are given for each nugget term.

the procedure for calculating the nugget term depended on the output of interest. The residuals from both tendencies could be combined without normalization, whereas L and f_C both required normalization by dividing by the mean. For L tendency, an appropriate nugget term could be derived from only three points. But for the f_C and f_C tendency, all nine points were required for an appropriate nugget term. This was most likely because three points that cover the range of behavior in one cloud property do not necessarily cover the full range of behavior in another cloud property. For all of these response surfaces with the nugget term added, the behavior across parameter space remains approximately as described in section 4.2, but it is now smoother and represents the model's general behavior better. Figure S2 shows the validation results of these response surfaces.

6 Discussion and conclusions

We have used Gaussian process emulation to analyze and visualize the dependence of stratocumulus clouds on the initial profiles of two cloud-controlling factors. Using an emulator in this way helps to visualize the relationships between cloud-controlling factors and the model output at a much-reduced computational cost.

We found there are two distinct behavioral cloud regimes in the explored parameter space, with a smooth transition between them, rather than a discontinuity. We do not find a distinct point beyond which the cloud rapidly breaks up, as hypothesized by Lilly (1968), Randall (1980) and Deardorff (1980). Our findings agree with Mellado (2017), in that cloud-top entrainment instability is not strong enough to break up the cloud by itself. However the κ parameter does approximately mark a behavior change between region A (Figure 5), which has thin stratocumulus cloud (or small cumulus) and very little growth, and region B, which has stratocumulus that starts with high cloud fraction, moderate liquid water path, and grows throughout the simulation. Emulation allowed us to densely sample the parameter space to map out cloud behavior based on model output, and the addition of the nugget term more smoothly captures the underlying deterministic behavior.

Natural variability presents a challenge for training an emulator describing the deterministic response of cloud fields to cloud-controlling factors. Failure to account for variability resulted in bumpy response surfaces. We used initial-condition ensembles at a small number of points across the parameter space to define an appropriate emulator nugget term, which resulted in a smooth emulator response surface. The size of the nugget term was defined as appropriate when the distribution of the residuals of the training data around the smooth surface was approximately the same as the distribution of the initial-condition ensembles. Although previous studies have used initial-condition-ensemble means as training data (Johnson et al., 2011) this is not feasible with cloud or climate models due to the expense of running ensembles. Our approach of effectively tuning the size of the nugget term provides an efficient alternative, although it would be difficult to apply if the variability varied in a complicated way across the parameter space.

Without including aerosol processes in the simulation some cloud breakup mechanisms are not accounted for in our simulations, such as rain-depletion feedbacks (Goren et al., 2019). We also used a lower fixed droplet number concentration than Stevens et al. (2005), however we repeated the simulations with a higher fixed concentration and found the general behaviors were not altered in each region. The lack of rapid cloud breakup fits with the conclusion in Mellado (2017) that the feedback mechanism is too weak to break up the cloud by itself.

One unexpected benefit of producing a response surface from a PPE and emulator was the ability to identify outliers in the data. Against the backdrop of the PPE and the emulated surface these simulations clearly stand out, allowing further investigation into why they do not fit with the rest of the data. In some cases, this could identify an interesting region of parameter space in the real world, or a natural variability extremity that could be investigated with a small ensemble, or perhaps a collection of parameter settings for which the model is unstable. If none of those are true and it is simply a corrupted value in the model, it has been caught and can be discarded.

As climate models get more complex, machine learning is invaluable tool for understanding processes. The number of simulations required to fully explore uncertainties and certain aspects of models is already infeasible, particularly with the nonlinear behavior of clouds. Statistical emulation has already proven to be immensely useful for sensitivity analysis of model parameters in climate studies. Here we have shown it in a different capacity by perturbing cloud-controlling factors, which has only recently begun to be explored. We believe this is a unique method for exploring cloud processes, and it can be expanded to include changes in aerosol concentrations, parameterisation coefficients and more meteorological parameters.

7 Open Research

Data availability statement: The data from the perturbed parameter ensemble can be found on Zenodo at https://doi.org/10.5281/zenodo.10036710 (Sansom, 2023). All code used in the analysis can be found at https://github.com/eers1/dycoms_analysis.

Appendix A Approximating variance

Because we normalized the residuals by the mean, a multiplying factor is required to make the variance the correct order of magnitude before being used in the emulation process. This is proven by simplifying the variance of the normalized residuals,

$$\begin{aligned}
\sigma_R^2 &= \frac{\sum_{i=1}^{l}\sum_{j=1}^{k}[(\frac{Z_i^{(j)}-\bar{Z}_i}{\bar{Z}_i})-\mu]^2}{N_R}, \mu = 0 && \text{(A1)}\\
&= \frac{\sum_{i=1}^{l}\sum_{j=1}^{k}[(\frac{Z_i^{(j)}-\bar{Z}_i}{\bar{Z}_i})]^2}{N_R}, && \text{(A2)}\\
&= \frac{\sum_{i=1}^{l}\frac{1}{\bar{Z}_i{}^2}\sum_{j=1}^{k}[(Z_i^{(j)}-\bar{Z}_i)]^2}{N_R}, N_R = \sum_{i=1}^{l} N_i && \text{(A3)}\\
&= \frac{\sum_{i=1}^{l}\frac{1}{\bar{Z}_i{}^2}\sum_{j=1}^{k}[(Z_i^{(j)}-\bar{Z}_i)]^2}{\sum_{i=1}^{l} N_i}, && \text{(A4)}\\
&= \sum_{i=1}^{l}\frac{1}{\bar{Z}_i{}^2}\frac{\sum_{j=1}^{k}[(Z_i^{(j)}-\bar{Z}_i)]^2}{N_i}, && \text{(A5)}\\
&= \sum_{i=1}^{l}\frac{1}{\bar{Z}_i{}^2}\sigma_i^2, && \text{(A6)}
\end{aligned}$$

where σ_i^2 is the variance of each ensemble, Z_i, and N_i is the number of members in each ensemble. Thus, from this normalization process, the variance of the residuals needs to be multiplied by a factor on the same order of magnitude as $\bar{Z}_i^2$ to be used in the emulation process. Note that we are not using a summation of the ensemble variances, as equation A6 indicates, because we are simply estimating from a small sample. The variance we are estimating is the lower bound, since any distribution is likely to be wider than what we have sampled.

Acknowledgments

R Sansom received funding from the EPSRC DTP (grant no. 2114653). J Johnson was supported by the Newton Fund (grant no. UK–China Research and Innovation Partnership Fund through the Met Office Climate Science for Service Partnership (CSSP) China) and by the PROMOTE project (Process analysis, observations and modelling - Integrated solutions for cleaner air for Delhi, grant no. NE/P016421/1). R Sansom is grateful for the use of the Met Office/NERC cloud model and the assistance from Adrian Hill, Adrian Lock at Met Office and Steef Böing, Craig Poku and Chris Dearden. Additionally, this work used the ARCHER UK National Supercomputing Service (2013 to 2021) and JASMIN, the UK collaborative data analysis facility.

References

Andrianakis, I., & Challenor, P. G. (2012, 12). The effect of the nugget on Gaussian process emulators of computer models. *Computational Statistics and Data Analysis*, *56*(12), 4215–4228. doi: 10.1016/j.csda.2012.04.020

Bellon, G., & Stevens, B. (2013, 4). Time scales of the trade wind boundary layer adjustment. *Journal of the Atmospheric Sciences*, *70*(4), 1071–1083. doi: 10.1175/JAS-D-12-0219.1

Bellouin, N., Quaas, J., Gryspeerdt, E., Kinne, S., Stier, P., Watson-Parris, D., ... Stevens, B. (2020, 3). *Bounding Global Aerosol Radiative Forcing of Climate Change* (Vol. 58) (No. 1). Blackwell Publishing Ltd. doi: 10.1029/2019RG000660

Blossey, P. N., Bretherton, C. S., Cheng, A., Endo, S., Heus, T., Lock, A. P., & van der Dussen, J. J. (2016). CGILS Phase 2 LES intercomparison of response of subtropical marine low cloud regimes to CO2 quadrupling and a CMIP3 composite forcing change. *Journal of Advances in Modeling Earth Systems*, *8*(4), 1714–1726. doi: 10.1002/2016MS000765

Bony, S., & Dufresne, J.-L. (2005). Marine boundary layer clouds at the heart of tropical cloud feedback uncertainties in climate models. *Geophysical Research Letters*, *32*(20). doi: https://doi.org/10.1029/2005GL023851

Bretherton, C. S. (2015, 11). *Insights into low-latitude cloud feedbacks from high-resolution models* (Vol. 373) (No. 2054). Royal Society of London. doi: 10.1098/rsta.2014.0415

Bretherton, C. S., Macvean, M. K., Bechtold, P., Chlond, A., Cotton, W. R., Cuxart, J., ... Wyant, M. C. (1999, 1). An intercomparison of radiatively driven entrainment and turbulence in a smoke cloud, as simulated by different numerical models. *Quarterly Journal of the Royal Meteorological Society*, *125*(554), 391–423. doi: 10.1002/qj.49712555402

Brown, A. R., Derbyshire, S. H., & Mason, P. J. (1994, 10). Large-eddy simulation of stable atmospheric boundary layers with a revised stochastic subgrid model. *Quarterly Journal of the Royal Meteorological Society*, *120*(520), 1485–1512. doi: https://doi.org/10.1002/qj.49712052004

Brown, N., Weiland, M., Hill, A., Shipway, B., Allen, T., Maynard, C., & Rezny, M. (2020, 9). *A highly scalable met office NERC cloud model.*

Ceppi, P., Brient, F., Zelinka, M. D., & Hartmann, D. L. (2017, 7). Cloud feedback mechanisms and their representation in global climate models. *Wiley Interdisciplinary Reviews: Climate Change*, *8*(4), e465. doi: 10.1002/wcc.465

Dal Gesso, S., Van Der Dussen, J. J., Siebesma, A. P., De Roode, S. R., Boutle, I. A., Kamae, Y., ... Vial, J. (2015, 6). A single-column model intercomparison on the stratocumulus representation in present-day and future climate. *Journal of Advances in Modeling Earth Systems*, *7*(2), 617–647. doi: 10.1002/2014MS000377

Dearden, C., Hill, A., Coe, H., & Choularton, T. (2018, 10). The role of droplet sedimentation in the evolution of low-level clouds over southern West Africa. *At-*

mospheric Chemistry and Physics, *18*(19), 14253–14269. doi: 10.5194/acp-18-14253-2018

Deardorff, J. W. (1980, 1). Cloud top entrainment instability. *Journal of the Atmospheric Sciences*, *37*(1), 131–147. doi: 10.1175/1520-0469(1980)037⟨0131:CTEI⟩2.0.CO;2

Douglas, A., & L'Ecuyer, T. (2020, 5). Quantifying cloud adjustments and the radiative forcing due to aerosol-cloud interactions in satellite observations of warm marine clouds. *Atmospheric Chemistry and Physics*, *20*(10), 6225–6241. doi: 10.5194/acp-20-6225-2020

Dufresne, J. L., & Bony, S. (2008, 10). An assessment of the primary sources of spread of global warming estimates from coupled atmosphere-ocean models. *Journal of Climate*, *21*(19), 5135–5144. doi: 10.1175/2008JCLI2239.1

Feingold, G., McComiskey, A., Yamaguchi, T., Johnson, J. S., Carslaw, K. S., & Schmidte, K. S. (2016, 5). New approaches to quantifying aerosol influence on the cloud radiative effect. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(21), 5812–5819. doi: 10.1073/pnas.1514035112

Field, P. R., Hill, A., Shipway, B., Furtado, K., Wilkinson, J., Miltenberger, A., ... Van Weverberg, K. (2023, 1). Implementation of a Double Moment Cloud Microphysics Scheme in the UK Met Office Regional Numerical Weather Prediction Model. *Quarterly Journal of the Royal Meteorological Society*, *n/a*(n/a). Retrieved from `https://doi.org/10.1002/qj.4414` doi: https://doi.org/10.1002/qj.4414

Ghonima, M. S., Norris, J. R., Heus, T., & Kleissl, J. (2015, 5). Reconciling and validating the cloud thickness and liquid water path tendencies proposed by R. Wood and J. J. van der Dussen et al. *Journal of the Atmospheric Sciences*, *72*(5), 2033–2040. doi: 10.1175/JAS-D-14-0287.1

Glassmeier, F., Hoffmann, F., Johnson, J. S., Yamaguchi, T., Carslaw, K. S., & Feingold, G. (2019, 8). An emulator approach to stratocumulus susceptibility. *Atmospheric Chemistry and Physics*, *19*(15), 10191–10203. doi: 10.5194/acp-19-10191-2019

Goren, T., Kazil, J., Hoffmann, F., Yamaguchi, T., & Feingold, G. (2019). Anthropogenic Air Pollution Delays Marine Stratocumulus Breakup to Open Cells. *Geophysical Research Letters*, *46*(23), 14135–14144. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019GL085412` doi: https://doi.org/10.1029/2019GL085412

Gramacy, R. B., & Lee, H. K. (2012, 5). Cases for the nugget in modeling computer experiments. *Statistics and Computing*, *22*(3), 713–722. doi: 10.1007/s11222-010-9224-x

Gray, M. E. B., Petch, J., Derbyshire, S. H., Brown, A. R., Lock, A. P., Swann, H. A., & Brown, P. R. A. (2001). Version 2.3 of the Met Office Large Eddy Model: Part II. Scientific Documentation. *Met O (APR) Turbulence and Diffusion Note No. 276*.

Hartmann, D. L., Ockert-Bell, M. E., & Michelsen, M. L. (1992, 11). The Effect of Cloud Type on Earth's Energy Balance: Global Analysis. *Journal of Climate*, *5*(11), 1281–1304. doi: 10.1175/1520-0442(1992)005⟨1281:teocto⟩2.0.co;2

Henderson, D. A., Boys, R. J., Krishnan, K. J., Lawless, C., & Wilkinson, D. J. (2009). Bayesian Emulation and Calibration of a Stochastic Computer Model of Mitochondrial DNA Deletions in Substantia Nigra Neurons. *Journal of the American Statistical Association*, *104*, 76–87. Retrieved from `https://www.tandfonline.com/action/journalInformation?journalCode=uasa20` doi: 10.1198/jasa.2009.0005

Hill, A. A., Shipway, B. J., & Boutle, I. A. (2015, 9). How sensitive are aerosol-precipitation interactions to the warm rain representation? *Journal of Advances in Modeling Earth Systems*, *7*(3), 987–1004. doi: 10.1002/

2014MS000422

Hoffmann, F., Glassmeier, F., Yamaguchi, T., & Feingold, G. (2020, 6). Liquid Water Path Steady States in Stratocumulus: Insights from Process-Level Emulation and Mixed-Layer Theory. *Journal of the Atmospheric Sciences*, *77*(6), 2203–2215. doi: 10.1175/jas-d-19-0241.1

Igel, A. L., van den Heever, S. C., & Johnson, J. S. (2018, 1). Meteorological and Land Surface Properties Impacting Sea Breeze Extent and Aerosol Distribution in a Dry Environment. *Journal of Geophysical Research: Atmospheres*, *123*(1), 22–37. doi: 10.1002/2017JD027339

Johnson, J. S., Cui, Z., Lee, L. A., Gosling, J. P., Blyth, A. M., & Carslaw, K. S. (2015, 3). Evaluating uncertainty in convective cloud microphysics using statistical emulation. *Journal of Advances in Modeling Earth Systems*, *7*(1), 162–187. doi: 10.1002/2014MS000383

Johnson, J. S., Gosling, J. P., & Kennedy, M. C. (2011, 5). Gaussian process emulation for second-order Monte Carlo simulations. *Journal of Statistical Planning and Inference*, *141*(5), 1838–1848. doi: 10.1016/j.jspi.2010.11.034

Johnson, J. S., Regayre, L. A., Yoshioka, M., Pringle, K. J., Lee, L. A., Sexton, D. M., … Carslaw, K. S. (2018, 9). The importance of comprehensive parameter sampling and multiple observations for robust constraint of aerosol radiative forcing. *Atmospheric Chemistry and Physics*, *18*(17), 13031–13053. doi: 10.5194/acp-18-13031-2018

J. Smith, C., J. Kramer, R., Myhre, G., Alterskjr, K., Collins, W., Sima, A., … M. Forster, P. (2020, 8). Effective radiative forcing and adjustments in CMIP6 models. *Atmospheric Chemistry and Physics*, *20*(16), 9591–9618. doi: 10.5194/acp-20-9591-2020

Khairoutdinov, M., & Kogan, Y. (2000, 1). A New Cloud Physics Parameterization in a Large-Eddy Simulation Model of Marine Stratocumulus. *Monthly Weather Review*, *128*(1). doi: 10.1175/1520-0493(2000)128⟨0229:ANCPPI⟩2.0.CO;2

Kuo, H.-C., & Schubert, W. H. (1988, 7). Stability of cloud-topped boundary layers. *Quarterly Journal of the Royal Meteorological Society*, *114*(482), 887–916. doi: https://doi.org/10.1002/qj.49711448204

Lee, L. A., Carslaw, K. S., Pringle, K. J., Mann, G. W., & Spracklen, D. V. (2011, 12). Emulation of a complex global aerosol model to quantify sensitivity to uncertain parameters. *Atmospheric Chemistry and Physics*, *11*(23), 12253–12273. doi: 10.5194/acp-11-12253-2011

Lee, L. A., Pringle, K. J., Reddington, C. L., Mann, G. W., Stier, P., Spracklen, D. V., … Carslaw, K. S. (2013, 9). The magnitude and causes of uncertainty in global model simulations of cloud condensation nuclei. *Atmospheric Chemistry and Physics*, *13*(17), 8879–8914. doi: 10.5194/acp-13-8879-2013

Lilly, D. K. (1968, 7). Models of cloud-topped mixed layers under a strong inversion. *Quarterly Journal of the Royal Meteorological Society*, *94*(401), 292–309. doi: 10.1002/qj.49709440106

Loeppky, J. L., Sacks, J., & Welch, W. J. (2009, 11). Choosing the sample size of a computer experiment: A practical guide. *Technometrics*, *51*(4), 366–376. doi: 10.1198/TECH.2009.08040

Lund, M. T., Myhre, G., & Samset, B. H. (2019, 11). Anthropogenic aerosol forcing under the Shared Socioeconomic Pathways. *Atmospheric Chemistry and Physics*, *19*(22), 13827–13839. doi: 10.5194/ACP-19-13827-2019

MacVean, M. K., & Mason, P. J. (1990). Cloud-Top Entrainment Instability through Small-Scale Mixing and Its Parameterization in Numerical Models. *Journal of Atmospheric Sciences*, *47*(8), 1012–1030. doi: 10.1175/1520-0469(1990)047⟨1012:CTEITS⟩2.0.CO;2

Malavelle, F. F., Haywood, J. M., Jones, A., Gettelman, A., Clarisse, L., Bauduin, S., … Thordarson, T. (2017, 6). Strong constraints on aerosol-cloud interactions from volcanic eruptions. *Nature*, *546*(7659), 485–491. doi:

10.1038/nature22974

Marshall, L. R., Johnson, J. S., Mann, G. W., Lee, L., Dhomse, S. S., Regayre, L., . . . Schmidt, A. (2019, 1). Exploring How Eruption Source Parameters Affect Volcanic Radiative Forcing Using Statistical Emulation. *Journal of Geophysical Research: Atmospheres*, *124*(2), 964–985. doi: 10.1029/2018JD028675

Marshall, L. R., Schmidt, A., Johnson, J. S., Mann, G. W., Lee, L. A., Rigby, R., & Carslaw, K. S. (2021). Unknown Eruption Source Parameters Cause Large Uncertainty in Historical Volcanic Radiative Forcing Reconstructions. *Journal of Geophysical Research: Atmospheres*, *126*(13), e2020JD033578. doi: https://doi.org/10.1029/2020JD033578

Mellado, J. P. (2017, 1). *Cloud-Top Entrainment in Stratocumulus Clouds* (Vol. 49). Annual Reviews Inc. doi: 10.1146/annurev-fluid-010816-060231

Moeng, C.-H. (2000, 11). Entrainment Rate, Cloud Fraction, and Liquid Water Path of PBL Stratocumulus Clouds. *Journal of the Atmospheric Sciences*, *57*(21), 3627–3643. doi: 10.1175/1520-0469(2000)057⟨3627:ERCFAL⟩2.0.CO;2

Morris, M. D., & Mitchell, T. J. (1995, 2). Exploratory designs for computational experiments. *Journal of Statistical Planning and Inference*, *43*(3), 381–402. doi: 10.1016/0378-3758(94)00035-T

Myhre, G., Shindell, D., Bréon, F.-m., Collins, W., Fuglestvedt, J., Huang, J., . . . Zhang, H. (2013). *Anthropogenic and Natural Radiative Forcing. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Tech. Rep.). Cambridge: Cambridge University Press. doi: 10.1017/CBO9781107415324

Nuijens, L., & Siebesma, A. P. (2019, 5). Boundary Layer Clouds and Convection over Subtropical Oceans in our Current and in a Warmer Climate. *Current Climate Change Reports 2019 5:2*, *5*(2), 80–94. doi: 10.1007/S40641-019-00126-X

O'Hagan, A. (2006, 10). Bayesian analysis of computer code outputs: A tutorial. *Reliability Engineering and System Safety*, *91*(10-11), 1290–1300. doi: 10.1016/j.ress.2005.11.025

Oyebamiji, O. K., Wilkinson, D. J., Jayathilake, P. G., Curtis, T. P., Rushton, S. P., Li, B., & Gupta, P. (2017, 9). Gaussian process emulation of an individual-based model simulation of microbial communities. *Journal of Computational Science*, *22*, 69–84. doi: 10.1016/j.jocs.2017.08.006

Park, J. M., van den Heever, S. C., Igel, A. L., Grant, L. D., Johnson, J. S., Saleeby, S. M., . . . Reid, J. S. (2020, 3). Environmental Controls on Tropical Sea Breeze Convection and Resulting Aerosol Redistribution. *Journal of Geophysical Research: Atmospheres*, *125*(6). doi: 10.1029/2019JD031699

Peace, A. H., Carslaw, K. S., Lee, L. A., Regayre, L. A., Booth, B. B., Johnson, J. S., & Bernie, D. (2020, 9). Effect of aerosol radiative forcing uncertainty on projected exceedance year of a 1.5 °C global temperature rise. *Environmental Research Letters*, *15*(9), 0940a6. Retrieved from `https://iopscience.iop.org/article/10.1088/1748-9326/aba20chttps://iopscience.iop.org/article/10.1088/1748-9326/aba20c/meta` doi: 10.1088/1748-9326/ABA20C

Pope, C. A., Gosling, J. P., Barber, S., Johnson, J. S., Yamaguchi, T., Feingold, G., & Blackwell, P. G. (2021). Gaussian Process Modeling of Heterogeneity and Discontinuities Using Voronoi Tessellations. *Technometrics*, *63*(1), 53–63. doi: 10.1080/00401706.2019.1692696

Pressel, K. G., Mishra, S., Schneider, T., Kaul, C. M., & Tan, Z. (2017, 6). Numerics and subgrid-scale modeling in large eddy simulations of stratocumulus clouds. *Journal of Advances in Modeling Earth Systems*, *9*(2), 1342–1365. doi: 10.1002/2016MS000778

R Core Team. (2018). *R: A language and environment for statistical computing.* Vi-

enna, Austria: R Foundation for Statistical Computing.
Randall, D. A. (1980, 1). Conditional instability of the first kind up-side down. *Journal of the Atmospheric Sciences*, *37*(1), 125–130. doi: 10.1175/1520-0469(1980)037⟨0125:CIOTFK⟩2.0.CO;2
Regayre, L. A., Johnson, J. S., Yoshioka, M., Pringle, K. J., Sexton, D. M., Booth, B. B., ... Carslaw, K. S. (2018, 7). Aerosol and physical atmosphere model parameters are both important sources of uncertainty in aerosol ERF. *Atmospheric Chemistry and Physics*, *18*(13), 9975–10006. doi: 10.5194/acp-18-9975-2018
Regayre, L. A., Pringle, K. J., Booth, B. B., Lee, L. A., Mann, G. W., Browse, J., ... Carslaw, K. S. (2014, 12). Uncertainty in the magnitude of aerosol-cloud radiative forcing over recent decades. *Geophysical Research Letters*, *41*(24), 9040–9049. doi: 10.1002/2014GL062029
Regayre, L. A., Pringle, K. J., Lee, L. A., Rap, A., Browse, J., Mann, G. W., ... Woodhouse, M. T. (2015, 9). The climatic importance of uncertainties in regional aerosol-cloud radiative forcings over recent decades. *Journal of Climate*, *28*(17), 6589–6607. doi: 10.1175/JCLI-D-15-0127.1
Regayre, L. A., Schmale, J., Johnson, J. S., Tatzelt, C., Baccarini, A., Henning, S., ... Carslaw, K. S. (2020). The value of remote marine aerosol measurements for constraining radiative forcing uncertainty. *Atmospheric Chemistry and Physics*, *20*(16), 10063–10072. doi: 10.5194/acp-20-10063-2020
Roustant, O., Ginsbourger, D., & Deville, Y. (2012, 10). DiceKriging, DiceOptim: Two R Packages for the Analysis of Computer Experiments by Kriging-Based Metamodeling and Optimization. *Journal of Statistical Software*, *51*(1), 1 - 55. Retrieved from `https://www.jstatsoft.org/index.php/jss/article/view/v051i01` doi: 10.18637/jss.v051.i01
Saltelli, A., Chan, K., & Scott, E. M. (2000). *Sensitivity Analysis*. Chichester, England: Wiley.
Sansom, R. W. N. (2023). An LES perturbed parameter ensemble of free-tropospheric cloud-controlling factors on stratocumulus [Data set]. *Zenodo*.
Schneider, T., Kaul, C. M., & Pressel, K. G. (2019, 3). Possible climate transitions from breakup of stratocumulus decks under greenhouse warming. *Nature Geoscience*, *12*(3), 164–168. doi: 10.1038/s41561-019-0310-1
Seinfeld, J. H., Bretherton, C., Carslaw, K. S., Coe, H., DeMott, P. J., Dunlea, E. J., ... Wood, R. (2016, 5). Improving our fundamental understanding of the role of aerosol-cloud interactions in the climate system. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(21), 5781–5790. Retrieved from `https://www.pnas.org/doi/abs/10.1073/pnas.1514043113` doi: 10.1073/PNAS.1514043113/SUPPL{_}FILE/PNAS.201514043SI.PDF
Shen, Z., Sridhar, A., Tan, Z., Jaruga, A., & Schneider, T. (2022, 3). A Library of Large-Eddy Simulations Forced by Global Climate Models. *Journal of Advances in Modeling Earth Systems*, *14*(3), e2021MS002631. doi: https://doi.org/10.1029/2021MS002631
Shipway, B. J., & Hill, A. A. (2012, 10). Diagnosis of systematic differences between multiple parametrizations of warm rain microphysics using a kinematic framework. *Quarterly Journal of the Royal Meteorological Society*, *138*(669), 2196–2211. doi: 10.1002/qj.1913
Siems, S. T., Bretherton, C. S., Baker, M. B., Shy, S., & Breidenthal, R. E. (1990, 4). Buoyancy reversal and cloud-top entrainment instability. *Quarterly Journal of the Royal Meteorological Society*, *116*(493), 705–739. doi: 10.1002/qj.49711649309
Stechmann, S. N., & Hottovy, S. (2016). Cloud regimes as phase transitions. *Geophysical Research Letters*, *43*(12), 6579–6587. doi: 10.1002/2016GL069396
Stevens, B., & Feingold, G. (2009, 10). *Untangling aerosol effects on clouds and*

precipitation in a buffered system (Vol. 461) (No. 7264). Nature Publishing Group. doi: 10.1038/nature08281

Stevens, B., Lenschow, D. H., Faloona, I., Moeng, C.-H., Lilly, D. K., Blomquist, B., ... Thornton, D. (2003, 10). On entrainment rates in nocturnal marine stratocumulus. *Quarterly Journal of the Royal Meteorological Society*, *129*(595), 3469–3493. Retrieved from `http://doi.wiley.com/10.1256/qj.02.202` doi: 10.1256/qj.02.202

Stevens, B., Moeng, C. H., Ackerman, A. S., Bretherton, C. S., Chlond, A., de Roode, S., ... Zhu, P. (2005). Evaluation of large-eddy simulations via observations of nocturnal marine stratocumulus. *Monthly Weather Review*, *133*(6), 1443–1462. doi: 10.1175/MWR2930.1

Van Der Dussen, J. J., De Roode, S. R., & Siebesma, A. P. (2014, 2). Factors controlling rapid stratocumulus cloud thinning. *Journal of the Atmospheric Sciences*, *71*(2), 655–664. doi: 10.1175/JAS-D-13-0114.1

Wellmann, C., Barrett, A. I., Johnson, J. S., Kunz, M., Vogel, B., Carslaw, K. S., & Hoose, C. (2018, 12). Using Emulators to Understand the Sensitivity of Deep Convective Clouds and Hail to Environmental Conditions. *Journal of Advances in Modeling Earth Systems*, *10*(12), 3103–3122. doi: 10.1029/2018MS001465

Wellmann, C., I Barrett, A., S Johnson, J., Kunz, M., Vogel, B., S Carslaw, K., & Hoose, C. (2020, 2). Comparing the impact of environmental conditions and microphysics on the forecast uncertainty of deep convective clouds and hail. *Atmospheric Chemistry and Physics*, *20*(4), 2201–2219. doi: 10.5194/acp-20-2201-2020

Williamson, D., & Blaker, A. T. (2014, 1). Evolving Bayesian Emulators for Structured Chaotic Time Series, with Application to Large Climate Models. *SIAM/ASA Journal on Uncertainty Quantification*, *2*(1), 1–28. doi: 10.1137/120900915

Wood, R. (2012, 8). *Stratocumulus clouds* (Vol. 140) (No. 8). doi: 10.1175/MWR-D-11-00121.1

Xiao, H., Wu, C. M., & Mechoso, C. R. (2011, 9). Buoyancy reversal, decoupling and the transition from stratocumulus to shallow cumulus topped marine boundary layers. *Climate Dynamics*, *37*(5), 971–984. doi: 10.1007/s00382-010-0882-3

Yamaguchi, T., & Randall, D. A. (2008, 5). Large-eddy simulation of evaporatively driven entrainment in cloud-topped mixed layers. *Journal of the Atmospheric Sciences*, *65*(5), 1481–1504. doi: 10.1175/2007JAS2438.1

Zhang, M., Bretherton, C. S., Blossey, P. N., Austin, P. H., Bacmeister, J. T., Bony, S., ... Zhao, M. (2013, 12). CGILS: Results from the first phase of an international project to understand the physical mechanisms of low cloud feedbacks in single column models. *Journal of Advances in Modeling Earth Systems*, *5*(4), 826–842. doi: 10.1002/2013MS000246