


RESEARCH ARTICLE

10.1029/2023MS003796

Key Points:

- We present Cloud Botany, an ensemble of idealized large-eddy simulations of the winter trade wind regions, controlled by six varied parameters
- The parameter ranges are chosen to match the climatology of the trade wind region
- The simulations show a variety of cloud organization patterns: small cumulus, stripes, cold pools, cloud arcs, and anvils

Correspondence to:

F. Jansson,
fjansson@abo.fi

Citation:

Jansson, F., Janssens, M., Grönqvist, J. H., Siebesma, A. P., Glassmeier, F., Attema, J., et al. (2023). Cloud Botany: Shallow cumulus clouds in an ensemble of idealized large-domain large-eddy simulations of the trades. *Journal of Advances in Modeling Earth Systems*, 15, e2023MS003796. <https://doi.org/10.1029/2023MS003796>

Received 3 MAY 2023

Accepted 29 SEP 2023

Author Contributions:

Conceptualization: Fredrik Jansson, Martin Janssens, Johanna H. Grönqvist, A. Pier Siebesma, Franziska Glassmeier, Masaki Satoh, Yousuke Sato

Data curation: Fredrik Jansson, Martin Janssens, Hauke Schulz, Tobias Kölling

Funding acquisition: A. Pier Siebesma

Methodology: Fredrik Jansson, Martin Janssens, Johanna H. Grönqvist, A. Pier Siebesma, Franziska Glassmeier

Resources: Fredrik Jansson, Masaki Satoh, Yousuke Sato

Software: Fredrik Jansson, Jisk Attema, Victor Azizi, Masaki Satoh, Yousuke Sato

© 2023 The Authors. *Journal of Advances in Modeling Earth Systems* published by Wiley Periodicals LLC on behalf of American Geophysical Union. This is an open access article under the terms of the [Creative Commons Attribution License](#), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Cloud Botany: Shallow Cumulus Clouds in an Ensemble of Idealized Large-Domain Large-Eddy Simulations of the Trades

Fredrik Jansson¹ , Martin Janssens^{1,2} , Johanna H. Grönqvist^{3,4} , A. Pier Siebesma^{1,5} , Franziska Glassmeier¹ , Jisk Attema⁶ , Victor Azizi⁶, Masaki Satoh⁷ , Yousuke Sato^{8,9} , Hauke Schulz^{10,11} , and Tobias Kölling¹⁰

¹Department of Geoscience & Remote Sensing, Delft University of Technology, Delft, The Netherlands, ²Department of Meteorology & Air Quality, Wageningen University & Research, Wageningen, The Netherlands, ³Department of Physics, Åbo Akademi University, Turku, Finland, ⁴Institute of Physics, University of Amsterdam, Amsterdam, The Netherlands, ⁵Royal Netherlands Meteorological Institute, De Bilt, The Netherlands, ⁶Netherlands eScience Center, Amsterdam, The Netherlands, ⁷Atmosphere and Ocean Research Institute, University of Tokyo, Kashiwa, Japan, ⁸Faculty of Science, Hokkaido University, Sapporo, Japan, ⁹RIKEN Center for Computational Science, Kobe, Japan, ¹⁰Atmosphere in the Earth System, Max Planck Institute for Meteorology, Hamburg, Germany, ¹¹Cooperative Institute for Climate, Ocean, and Ecosystem Studies (CICOES), University of Washington, Seattle, WA, USA

Abstract Small shallow cumulus clouds (<1 km) over the tropical oceans appear to possess the ability to self-organize into mesoscale (10–100 km) patterns. To better understand the processes leading to such self-organized convection, we present Cloud Botany, an ensemble of 103 large-eddy simulations on domains of 150 km, produced by the Dutch Atmospheric Large Eddy Simulation model on supercomputer Fugaku. Each simulation is run in an idealized, fixed, larger-scale environment, controlled by six free parameters. We vary these over characteristic ranges for the winter trades, including parameter combinations observed during the EUREC⁴A (Elucidating the role of clouds–circulation coupling in climate) field campaign. In contrast to simulation setups striving for maximum realism, Cloud Botany provides a platform for studying idealized, and therefore more clearly interpretable causal relationships between conditions in the larger-scale environment and patterns in mesoscale, self-organized shallow convection. We find that any simulation that supports cumulus clouds eventually develops mesoscale patterns in their cloud fields. We also find a rich variety in these patterns as our control parameters change, including cold pools lined by cloudy arcs, bands of cross-wind clouds and aggregated patches, sometimes topped by thin anvils. Many of these features are similar to cloud patterns found in nature. The published data set consists of raw simulation output on full 3D grids and 2D cross-sections, as well as post-processed quantities aggregated over the vertical (2D), horizontal (1D) and all spatial dimensions (time-series). The data set is directly accessible from Python through the use of the EUREC⁴A intake catalog.

Plain Language Summary The organization of shallow cumulus clouds over the tropical ocean has recently received a lot of attention. This type of organization is potentially important for how the clouds are affected by a changing climate and also for how they modulate further warming. We present a collection of 103 detailed simulations of shallow cumulus clouds in idealized atmospheric environments. These environments are described by six parameters, and our collection is formed by systematically simulating different parameter combinations. This way an ensemble is created that spans up a multidimensional phase space of environmental conditions typical for the wintertime subtropical Atlantic Ocean. This approach allows us to form a picture of how the environmental conditions relate to the cloud organization that develops in the simulations. At a glance, most simulations evolve similarly: They quickly form small cumulus clouds, which then grow in size and organize into patterns. Often this leads to rainfall, which then causes further heterogeneity and pattern formation. The data is openly available online, and will serve future studies of cumulus clouds, their organization, and how they interact with the climate.

1. Introduction

According to the Encyclopedia Britannica, botany is the “branch of biology that deals with the study of plants, including their structure, properties, and biochemical processes. Also included are plant classification [...] and interactions with the environment” (Pelczar et al., 2022). While conceived by biologists, this definition

Visualization: Fredrik Jansson, Johanna H. Grönqvist

Writing – original draft: Fredrik Jansson, Martin Janssens, Johanna H. Grönqvist

Writing – review & editing: Fredrik Jansson, Martin Janssens, Johanna H. Grönqvist, A. Pier Siebesma, Franziska Glassmeier, Jisk Attema, Victor Azizi, Masaki Satoh, Yousuke Sato, Hauke Schulz, Tobias Kölling

fits curiously well with how meteorologists think about clouds. In fact, Luke Howard's cloud taxonomy (Howard, 1803) seems to have been explicitly inspired by Linnean nomenclature (Pedgley, 2003). Meteorologists, like botanists, to this day use this taxonomy to facilitate our study of cloud features, underlying processes and interactions with their atmospheric environment.

A recent example, with which we will concern ourselves here, focuses on cloudiness over the swaths of the tropical oceans known as the trades. During winter, this region is inhabited by shallow cumulus clouds, which in small-domain large-eddy simulations (LES, domain sizes $O(10)$ km) appear homogeneously organized over the horizontal plane (e.g., Siebesma et al., 2003), and which have historically remained unresolved by models of global scale (resolution $O(100)$ km). Thus, cloud structures in the range of scales in between ($O(10 - 100)$ km, which we will refer to collectively as the mesoscales) have been rather sparsely studied (Nuijens & Siebesma, 2019). Yet, satellite observations of the trade wind region reveal that shallow clouds are organized into a rich spectrum of patterns at these scales (Agee, 1984; Stevens et al., 2020). Simple, botanical descriptions of such mesoscale cloud patterns, for example, through classification (Stevens et al., 2020) or characterization (Denby, 2020; Janssens et al., 2021), are at present guiding our understanding of how cloud patterns interact with their environment (Schulz et al., 2021), and revealing their importance in setting the trade-wind contribution to Earth's energy balance and its sensitivity to changes in our climate (Bony et al., 2020).

The goal of improving our understanding of the mesoscale, marine trades has mobilized an entire community, centered around the EUREC⁴A (Elucidating the role of clouds–circulation coupling in climate) field campaign (Stevens et al., 2021). Fortunately, advances in computational capabilities now allow these observations to be complemented by (a) global and regional models running at a sufficiently fine resolution to begin resolving shallow convection (e.g., Stevens et al., 2019) and (b) detailed process models—“classical” LES codes—running on sufficiently large domains to capture the mesoscale (e.g., Lamaakel & Matheou, 2022; Narenpitak et al., 2021; Seifert et al., 2015). In particular, these models facilitate understanding of the degree to which mesoscale cloud patterns originate in larger-scale dynamics, which set the environment in which clouds form, or small-scale processes, which govern individual cumulus structures. Regional simulations running at less than a kilometer resolution are beginning to appear (Schulz, 2021; Schulz & Stevens, 2023); these attain a detailed representation of the larger scale and are therefore well-suited to investigate the importance of those scales. However, we still miss a systematic exploration of large-domain (>100 km) LES that maintains a simple representation of the larger-scale environment, but does not compromise on its turbulence-resolving resolution of around 100 m.

To bridge that gap, this paper presents Cloud Botany, an ensemble of 103 simulations on domains of 150 km at 100 m horizontal resolution, enabled by the computing capabilities of supercomputer Fugaku. With Cloud Botany, we take a step back from the pursuit of realistic regional or global simulations. Instead, we hypothesize that if we wish to understand the role played by cumulus convection in organizing the tropical mesoscale, it is helpful to begin by idealizing and fixing the larger-scale environment and boundary forcings on a mesoscale domain, and study the response of freely developing cloud patterns to variations in these idealized forcings. Therefore, we will parameterize the vertical structure of the trade-wind environment with six parameters. We then co-vary these parameters across the range of typically observed conditions in the trades, which results in the ensemble of initial conditions and boundary forcings that our simulations run under. Such ensembles successfully explain parameter-dependencies in small-domain simulations of the subtropics (e.g., Bellon & Stevens, 2012; Feingold et al., 2016; Glassmeier et al., 2021; Nuijens & Stevens, 2012; Schalkwijk et al., 2013; Shen et al., 2022); we designed Cloud Botany to test if extending this approach to large LES domains can help understand the origins of mesoscale cloud patterns.

The construction of the simulation ensemble and description of the resulting data products are the main focus of the present manuscript. We aim to use the data to investigate targeted questions, such as how the smallest energetic scales of motion self-organize into mesoscale structures (e.g., Bretherton & Blossey, 2017; Janssens et al., 2022; Narenpitak et al., 2021; Seifert et al., 2015) under varying conditions. However, the simulations also come forth from a general curiosity as to which trade-wind cloud structures our LES model can actually produce (and which not), and how we might describe and classify these. It is in this sense that our exploration comes closest to paralleling the botanist's quest. Most importantly, we hope the data set is useful to a community with a broad range of research questions pertaining to the understanding of the detailed dynamics of the mesoscale trades.

The paper is organized as follows. We begin by describing the creation of the initial and boundary conditions that define our simulation ensemble (Section 2). Running each simulation still requires the choice of several other parameters which we hold fixed over the ensemble. These are outlined in Section 3. Section 4 describes the workflow of setting up and running the simulations on Fugaku, and how its output is translated into accessible data sets. Section 5 describes the salient features of these data products, before Section 6 gives a brief overview of some frequently recurring cloud patterns. A conclusion is offered in Section 7.

2. Creating an LES Ensemble in a Parameter Space

To study how self-organized cloud patterns in LES respond to variations in the larger-scale environment, we will initialize and force LESs with simple, functional representations of the vertical structure of the trade-wind environment (“profiles”). The parameters that control these profiles will span a “parameter space,” which we will explore by co-varying the parameters. To cover this space with around 100 simulations, we must keep its dimensionality as low as possible. Therefore, we wish to find a set of profiles which is controlled by a minimal number of parameters. At the same time, we want these profiles retain enough realism to remain useful for comparing variability over our simulation ensemble to variability in the real-world sub-tropics.

In this section, we will elaborate on how we design a parameter space that strikes this balance. We will first present our chosen set of idealized profiles and their free parameters (Section 2.1). We will then judge the realism of these profiles by analyzing how well we can fit them to reanalysis and observations (Section 2.2). Finally, we will use the variability in the parameters as fitted to observations to inform the ranges we will co-vary our parameters over, resulting in the set of initial conditions and forcings that make up our ensemble (Section 2.3).

2.1. Idealizations of the Trade-Wind Environment

Cloud Botany is based on simulations conducted with the Dutch Large Eddy Simulation (DALES; Heus et al., 2010; Ouwersloot et al., 2017). In the configuration used here, DALES solves numerical approximations of the anelastic equations of atmospheric motion in a three-dimensional domain over a sea surface with a homogeneous temperature. The domain is discretized by a staggered grid. To initialize our idealized DALES simulations, we specify vertical profiles for five of its prognostic quantities: Liquid-water potential temperature θ_l , total specific humidity q_t , horizontal velocity in west–east (u) and south–north (v) directions, and sub-filter scale (SFS) turbulent kinetic energy e ; vertical velocities w are zero when horizontally averaged and do not require initialization. Similarly, we will parameterize scales larger than the simulation domain with idealized profiles for (a) geostrophic horizontal wind (u_g, v_g), (b) a large scale vertical velocity (w_{ls}), and (c) large scale tendencies of moistening and heating, which we keep constant over 2.5 days of simulation. We will model these profiles of initial conditions and large-scale forcings using profiles that capture basic aspects of the trade-wind environment’s expected, physical structure with at most two free parameters. Thus, our parameter space will contain both parameters that set the initial state of the atmosphere in our simulations, and parameters that explicitly force the atmospheric state; their common denominator is that they all explain an appreciable amount of variability in the environment, and are thought to be important cloud-controlling variables. Parameters that are kept fixed over the ensemble are listed in Table 1.

We set both the initial profiles and geostrophic wind profiles of horizontal velocities u and v to

$$u(z) = u_0 + u_z z, \quad v(z) = 0 \quad (1)$$

where u_0 is the initial near-surface wind and $u_z = \partial u / \partial z$ denotes the initial vertical shear of horizontal wind speed. The geostrophic wind is assumed to remain constant in time during each simulation. Except for a few exceptions, all simulations will be initialized with the same zonal shear strength. As our analysis is positioned in the downstream trades, we assume $v_0 = 0, v_z = 0$ for all our experiments, that is, the geostrophic wind is predominantly east–west.

Profiles of the initial liquid water potential temperature θ_l follow a similar, linear approximation. However, the lowest θ_l levels are expected to co-vary with the surface conditions. Indeed, upon consulting the global ERA5 reanalysis (Hersbach et al., 2020), we find a Pearson correlation of $r = 0.57$ between θ_l at the lowest ERA5 level and the surface. To avoid long model spinups where surface fluxes attempt to re-calibrate an out-of-equilibrium

Table 1
Parameters Held Constant in the Experiment Setup

Parameter (unit)	Value	Description
u_z (1/s)	0.00222	Initial zonal wind shear
$\Delta\theta_{l0}$ (K)	1.25	Initial difference in θ_l between surface and first atmospheric layer
z_{ml} (m)	500	Initial mixed layer height
w_∞ (cm/s)	-0.45	Background subsidence velocity
h_{w_∞} (m)	2,500	Scale height of background subsidence
h_{w_1} (m)	5,300	Scale height of first additional mode of imposed vertical velocity
$\partial_l\theta_{l,ls,0}$ (K/day)	-0.5	Large scale temperature tendency in first model level
$\partial_l\theta_{l,ls,z}$ (K/day/m)	$2.5 \cdot 10^{-4}$	Large scale temperature tendency slope
$\partial_lq_{l,ls,0}$ (g/kg/day)	-1.49	Large scale humidity tendency at surface
$\partial_lq_{l,ls,z}$ (g/kg/day/m)	$3.73 \cdot 10^{-4}$	Large scale humidity tendency slope
τ_∞ (hr)	6	Nudging time scale at top of domain
z_{\max} (m)	3,000	Height around which the transition from strong ($z > z_{\max}$) to weak ($z < z_{\max}$) nudging is centered
a	2	Constant for setting nudging time scale
b	3	Constant for setting nudging time scale
c	7.4	Constant for setting nudging time scale

mixed- and cloud layer, we therefore initialize θ_l with a residual layer of constant height $z_{ml} = 500$ m. Having chosen a (potential) sea-surface temperature (SST), θ_{l0} , we simply set the residual layer's value to the sea surface temperature minus the reanalysis-mean difference in θ_l between the lowest ERA5 level and the surface, $\Delta\theta_{l0}$. This gives the following definition for θ_l :

$$\theta_l(z) = \begin{cases} \theta_{l0} - \Delta\theta_{l0} & \text{if } z < z_{ml} \\ \theta_{l0} - \Delta\theta_{l0} + \Gamma(z - z_{ml}) & \text{if } z \geq z_{ml} \end{cases} \quad (2)$$

Hence, the initial profile of θ_l is fully determined by setting θ_{l0} and Γ . In observations, u_0 and Γ seem to be important control parameters for the size and degree of clustering of trade-wind clouds (Bony et al., 2020; Schulz et al., 2021). To test whether similar dependencies can be observed in our LES setup, we have deliberately chosen u and θ_l to be specified by these parameters.

Profiles of the total humidity q_t are modeled with a similar initial well-mixed layer, but drop off exponentially above z_{ml} , following Vogel et al. (2020):

$$q_t(z) = \begin{cases} q_{t,ml} & \text{if } z < z_{ml} \\ q_{t,ml} e^{-\frac{z-z_{ml}}{h_{q_t}}} & \text{if } z \geq z_{ml} \end{cases} \quad (3)$$

The free parameters of this parameterization are the initial mixed-layer moisture $q_{t,ml}$ and the moisture scale height h_{q_t} . The surface moisture is assumed to be at saturation, and thus follows from θ_{l0} and the surface pressure, and the difference in moisture between the first model level and the surface may be diagnosed in turn.

Finally, we will impose profiles of the large scale vertical velocity w_{ls} that includes two terms: (a) a term representing the downwelling branch of the Hadley cell, modeled by exponential decay with height following for

Table 2
Properties of Environmental Control Parameters, Fit to the ERA5 Database, and Selected for Cloud Botany

	θ_{10} (K)	Γ (K/km)	$q_{t,ml}$ (g/kg)	h_{q_t} (m)	u_0 (m/s)	w_1 (cm/s)
R^2		0.99		0.95	0.72	0.28
Mean	299	5	14.1	1,810	-10.6	0.0393
10–90	298–300	4.54–5.28	12.8–15.4	1,180–2,510	-14.2 to -6.93	-0.984 to 1.14
Sel.	297.5–299.5	4.5–5.5	13.5–15.0	1,200–2,500	-15 to -5	-0.350 to 0.180

Note. R^2 is the average coefficient of determination of the non-linear least squares fits per variable, used to qualitatively gauge the quality of fit for Equations 2–4. Mean indicates the average parameter values over all fits, while 10–90 refers to the value of the 10th and 90th percentile of each parameter over the fits. The ranges over which the parameters in Cloud Botany are varied are reported in the table's bottom row.

example, Bellon and Stevens (2012), and (b) a sinusoidal term, a single period of which represents mesoscale circulations, as frequently observed during EUREC⁴A (George et al., 2023):

$$w_{ls}(z) = -w_{\infty} \left(1 - e^{-\frac{z}{h_{w_{\infty}}}} \right) + \begin{cases} w_1 \sin\left(\frac{2\pi}{h_{w_1}} z\right) & \text{if } z < h_{w_1} \\ 0 & \text{if } z \geq h_{w_1} \end{cases} \quad (4)$$

Varying w_1 captures a substantial amount of the mesoscale variability in vertical velocity in the trades (George et al., 2023). Therefore, we fix the free-tropospheric, asymptotic subsidence w_{∞} and its scale height $h_{w_{\infty}}$. Furthermore, we assume (a) that the vertical depth of the circulations, encapsulated by h_{w_1} , scales with the boundary-layer height, which LES studies of the phenomenon indicate to be reasonable (Bretherton & Blossey, 2017; Janssens et al., 2022; Narenpitak et al., 2021), and (b) that it to first order is constant in time. This leaves the strength of the sinusoidal term w_1 as a free parameter in our large-scale vertical velocity profiles.

Importantly, we do not fix the large scale vertical velocity profiles to satisfy a Weak Temperature Gradient (WTG) constraint on the mean flow in the free troposphere, in which horizontally averaged vertical motion is diagnosed given a radiative heating rate and Γ (Bellon & Stevens, 2012; Nuijens & Stevens, 2012). Not enforcing WTG allows richer responses of the boundary layer to its forcing (Betts & Ridgway, 1989), and is more representative of the real trades, where free-tropospheric tendencies in heating and moistening are not usually small (e.g., Nitta & Esbensen, 1974). This choice prevents the free-troposphere from acquiring quasi-equilibrium, and requires us to add a subtle nudging to prevent the tendencies from becoming overly adventurous; we return to this in Section 3.

In all, our idealized framework has six free parameters that set the environment we launch our simulations in, spanning a six-dimensional parameter space: surface wind u_0 , surface temperature θ_{10} , temperature lapse rate Γ , surface humidity $q_{t,ml}$, humidity scale height h_{q_t} and large-scale vertical velocity variability w_1 .

2.2. Quality of Fits

To assess how idealized the chosen functional forms are with respect to the vertical structure of the real trade-wind environment, we will compare them with the ERA5 global reanalysis, sampled every 3 hr between 9.8–16.8°N and 62.22–54.22°W between 1 January 2020 and 31 March 2020. This domain is representative for the trades in general (Medeiros & Nuijens, 2016) and the period spans the winter during which the EUREC⁴A campaign was conducted (20 January 2020 to 20 February 2020). We complement use of ERA5 with the JOANNE (Joint dropsonde Observations of the Atmosphere in tropical North atlantic meso-scale Environments) data set (George et al., 2021), gathered during the campaign by launching densely spaced meteorological dropsondes from an aircraft along the perimeter of a 200 km circle. This spatial scale roughly fits that of our horizontal domain size. Therefore, we will directly use this data at the spatial scale of the circle and time scale of a day's flights. We fit all profiles in our ERA5 database with Equations 1–4, using a non-linear least squares algorithm. The results are shown in the top three rows of Table 2.

Quality of fit is assessed in terms of the coefficient of determination (R^2), averaged over all fits for θ_l , q_t , u , and w_{ls} . While R^2 cannot formally be interpreted as the explained variance for our non-linear least-squares regressions, the top row of Table 2 still illustrates that the fits of θ_l are excellent, those of q_t and u are adequate, and those of w_1 are inadequate. The poor fit of w_1 reflects both significant deviations from the prescribed functional form, and variability in higher-order modes in the ERA5 database than our simple approximation captures. Since ERA5 assimilates the JOANNE data, and agrees well with the observed velocity even when it is not assimilated (George et al., 2023; Savazzi, Nuijens, Sandu, et al., 2022), these higher-frequency w fluctuations are unlikely to be spurious. Therefore, we revisit the design of w_{ls} below.

By excluding v , we artificially remove momentum from our simulated environment. To investigate the consequences, we have fit profiles of v in the same manner as for u . The resulting meridional surface wind (v_0) is on average around 15% the strength of the zonal surface wind (u_0), while the meridional shear $v_z \approx 0$. We compensate for this general lack of momentum in the simulations by also investigating marginally broader ranges in u_0 in our parameter sweeps (see below).

2.3. Chosen Parameter Ranges

To keep our simulation number manageable while capturing as much of the variability that occurs in the winter trades as possible, Cloud Botany consists primarily of simulations conducted at the corners of a hypercube in our six-dimensional parameter space, that is, $2^6 = 64$ simulations. These stem from considering all possible combinations of our environmental control parameters, at a minimum and a maximum point informed by the 10th and 90th percentile of each parameter's variability over the ERA5 fits (second-to-last row in Table 2). This choice makes our simulations indicative of the envelope of conditions observed in the trades; they are thus not to be confused with the climatology that would have resulted from sampling the multivariate probability distribution functions of the fitted parameters. To still capture parameter dependencies in more typically observed conditions, we supplement the hypercube corners with “sweeps” of every parameter: Runs that span the range between the extrema in several steps for one control parameter, with all other parameters held at the center of the hypercube. The sweep runs thus explore the sensitivity of the central point in the parameter space to perturbations in a single parameter.

Since the chosen parameters will be varied independently of each other, it is prudent to quantify their independence in observations, that is, whether they each capture a unique aspect of the environment's variability. Pairwise Pearson correlations of our ERA5 fits broadly confirm this: All coefficients are below 0.4, with the largest correlations existing between θ_{l0} and Γ (-0.31), θ_{l0} and $q_{t,ml}$ (0.33), Γ and h_{q_t} (-0.26), $q_{t,ml}$ and w_1 (0.32), and $q_{t,ml}$ and h_{q_t} (0.35). All other correlations are below 0.25.

The final ranges over which we run each control parameter are given in the bottom row of Table 2. For Γ , h_{q_t} and u_0 , these directly result from rounding the 10th and 90th percentile values. Variability in θ_{l0} subsumes both variability in surface pressure and SST. Since we keep the surface pressure over our ensemble fixed at 1,016.05 hPa, we adjust the rounded range over which we vary θ_{l0} to better match the variability in SST. This results in a downwards adjustment of 0.5 K. In preliminary experiments, combinations of high-end free-tropospheric moisture and free-running free-tropospheric tendencies would sometimes produce clouds near our domain tops, which after spurious boundary interactions with our radiation scheme would yield temperatures exceeding the local boiling point and crash our thermodynamics scheme. Conversely, simulations with less cloud-layer moisture than the ERA5 envelope would often not even develop clouds. To avoid these situations, we narrow the envelope of $q_{t,ml}$ slightly to avoid unrealistically dry and moist free-tropospheric moisture profiles and initial profiles that exceed a relative humidity of 100%. As we shall see in Section 6, even the final ensemble still contains some runs that fail in this manner.

There are certain inherent limitations to modeling variability in w_{ls} with a framework as simple as ours: it does not adequately represent high-frequency vertical modes, nor does prescribing w_{ls} allow the convection developing in our simulations to interact with vertical velocity structures of scales larger than our domain. Our compromise aims to (a) capture sufficient w variability to satisfy our main objective—studying environmental dependencies—and (b) ensure that the variability we capture is more representative of the reanalysis than traditional exponential (Bellon & Stevens, 2012; Blossey et al., 2013) or linear (Siebesma et al., 2003; Stevens et al., 2001; Yamaguchi et al., 2019) approximations. Therefore, we set w_∞ to a number characteristic of the ERA5 mean in

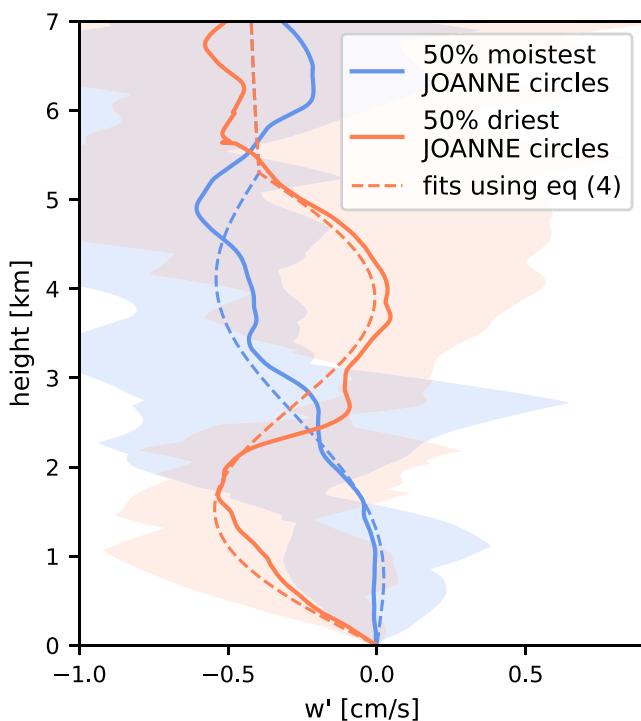


Figure 1. Envelopes and mean of daily-averaged vertical velocity in the JOANNE data set (George et al., 2021). Daily averages are obtained from the six circles flown on a day by the HALO aircraft during the EUREC⁴A campaign. The data is separated into the 50% moistest (blue) and driest (red) flight days. Dashed lines indicate the profiles used in the simulation, constructed with Equation 4 and the parameters reported in Tables 1 and 2.

prescription of a model grid, a precipitation model, a radiation model, a prescribed large-scale advective forcing, and a free tropospheric nudging. These are all kept the same for all simulations; we briefly describe them in turn below.

Our simulations run on horizontally square domains of 153.6 km, with a height of 7 km. The domains have periodic boundary conditions in the two horizontal directions. To discretize this cuboid, we use a grid with a horizontal spacing of 100 m, and vertical spacing of 20 m in our first model level, stretched by 1% in each level above. This yields 1,536 grid points on a horizontal side, and 175 vertical grid levels. The grid's size limits our simulations' integration time to 60 hr, which is substantially shorter than the 20 days that for example, M. Zhang et al. (2013) run their small-domain simulations to come into equilibrium with their larger-scale environment. Our simulations do not reach such an equilibrium. We choose to accept this limitation, since it is not obvious whether 150 km fields of trade cumuli are in equilibrium with their larger-scale environment in nature. Furthermore, we will show in Section 6 that the liquid-water path (LWP) is essentially in equilibrium after two simulated days, by extending a single simulation to 9 days.

Advection of momentum, θ_l and e is discretized with a sixth order scheme, advection of q_l and precipitation species with a fifth order scheme which is nearly monotonous (Wicker & Skamarock, 2002). The sources and sinks of precipitation are modeled with a warm microphysics scheme based on Seifert and Beheng (2001), whose two moments we prognose. We prescribe a (fixed) cloud-droplet number concentration of $7 \cdot 10^7 \text{ m}^{-3}$.

Radiative heating rates are calculated interactively with RRTMG (Iacono et al., 2008; Pincus et al., 2003). As the importance of diurnal, radiative variability in the downstream trades has recently been emphasized (Albright et al., 2021; Narenpitak et al., 2023; Vial et al., 2019, 2021, 2023), we include in the model's shortwave component the diurnal cycle representative for Feb-01-2020 at 13.1°N and 52°W (representative of the EUREC⁴A domain and period). Required input profiles for water vapor and temperature above the LES domain were derived

the free troposphere, where its variation is not expected to be important for the current study, and vary w_1 according to how it varied between the moistest and driest 50% of circles flown by the HALO aircraft during EUREC⁴A (George et al., 2023).

We separate the vertical velocity variability by moisture variability and not by the vertical velocity itself. This results in a smaller w_{ls} variability than observed during EUREC⁴A and in the reanalysis. We justify this by noting that large values in w_{ls} are not expected to remain constant over the multiple days we simulate, since they arise from shorter-lived circulations (George et al., 2023). For instance, applying a constant 2 cm/s cloud-layer subsidence velocity (the 10th percentile in ERA5), would imply a subsidence heating rate of roughly 10 K/day and a drying rate of 10 g/kg/day, which heat and dry our cloud layers beyond conditions that can sustain cumulus convection within a few hours. Indeed, we found that spanning a wider range of the w_1 parameter led to many corner points in the parameter hypercube becoming either too dry for clouds to form, or too moist, resulting in deep convection and model crashes. Multiple-day cloud-free or deeply-convecting situations are hardly observed in the winter trades, nor are they our primary interest. Thus, we elect to fit w_1 on the more stable, larger-scale water-vapor variability, resulting in the profiles shown in Figure 1. We performed additional tests varying w_∞ with $w_1 = 0$, which indicate that w_{ls} in the cloud layer controls the cloud formation, and that the shape of w_{ls} is less important.

The remaining parameters needed to complete Equations 1–4 are reported in Table 1, and the complete ensemble of initial and boundary conditions that emerges is plotted in Figure 2.

3. Design of Fixed LES Parameters

While Section 2 describes the set of initial and lower boundary conditions that vary over our simulation ensemble, running a simulation still requires the prescription of a model grid, a precipitation model, a radiation model, a prescribed large-scale advective forcing, and a free tropospheric nudging. These are all kept the same for all simulations; we briefly describe them in turn below.

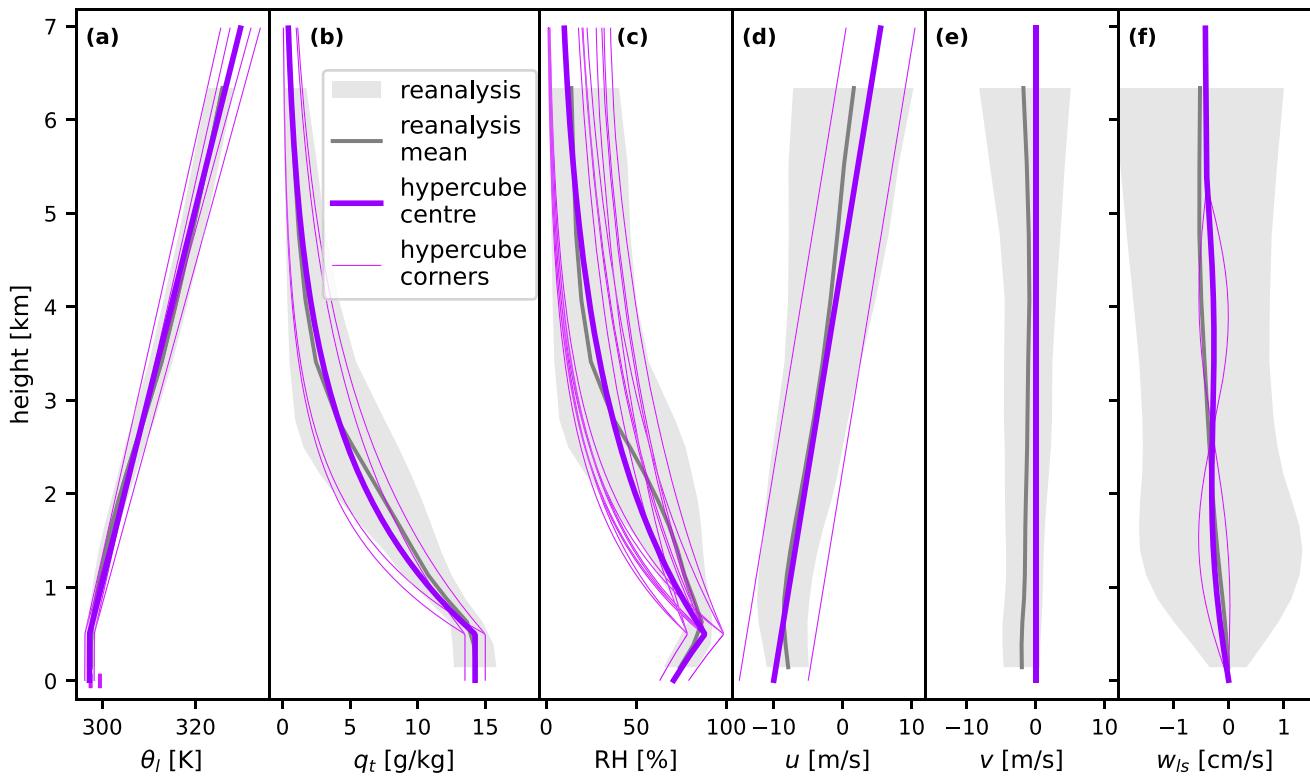


Figure 2. Profiles of (a) θ_l , (b) q_t , (c) relative humidity (RH), (d) u , (e) v (kept at zero in the ensemble), and (f) w_{ls} over the 10th–90th percentile envelope in ERA5 (reanalysis, shading), its mean (gray), and for initial and large-scale forcing of Cloud Botany simulations: the center (purple) and corners (pink) of the six-dimensional hypercube.

from ERA5, averaged over the EUREC⁴A region and period. These were stitched to the prognosed profiles of temperature and water vapor within our numerical domain from the 7 km domain top until a height corresponding to the 100 Pa pressure level (which we refer to as the top of the atmosphere—TOA). Default profiles supplied with RRTMG were adopted for other gases (O_3 , CO_2 , CH_4 , N_2O , and O_2).

We add two large-scale forcings to the simulations. The first are (horizontally constant) tendencies that aim to be representative of the typical drying (for q_t) and cooling (for θ_l) of our region of interest through advection on a horizontal scale larger than we simulate. We estimate these tendencies from JOANNE following a linear approximation, held at zero once they cross the ordinate (Figure 3):

$$\partial_t \theta_{l,ls} = \min(0, \partial_t \theta_{l,ls,0} + \partial_t \theta_{l,ls,z} z) \quad (5)$$

$$\partial_t q_{t,ls} = \min(0, \partial_t q_{t,ls,0} + \partial_t q_{t,ls,z} z) \quad (6)$$

These tendencies display variability around the fixed, approximate state we have chosen, which would have made their inclusion in our parameter space interesting. We excluded such variations to keep the required simulation number tractable, but recommend investigating their importance in future extensions.

Finally, our rich ensemble of initial conditions combine with our variation of w_{ls} to form a rather broad variety of w_{ls} -induced heating and drying tendencies forced on our slab-averaged prognostic variables in the free troposphere. To prevent these tendencies from driving the initial state outside the ERA5 envelope, we impose a nudging tendency on our prognostic variables (u , v , θ_l , and q_t) that forces them back toward their initial state with a height-dependent nudging time scale τ :

$$\tau(z) = \tau_\infty + \left(b \left[\frac{\pi}{2} + \arctan \left\{ a \frac{\pi}{2} \left(1 - \frac{z}{z_{\max}} \right) \right\} \right] \right)^c \quad (7)$$

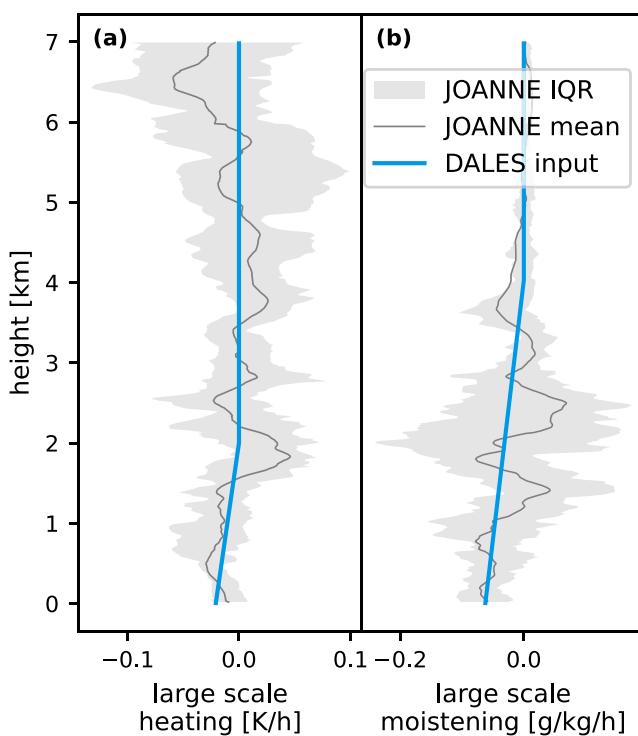


Figure 3. Inter-quartile range (IQR, shading) and mean (gray line) of JOANNE-derived tendencies of (a) heating and (b) moistening, and idealized fit used to force Cloud Botany simulations (blue line).

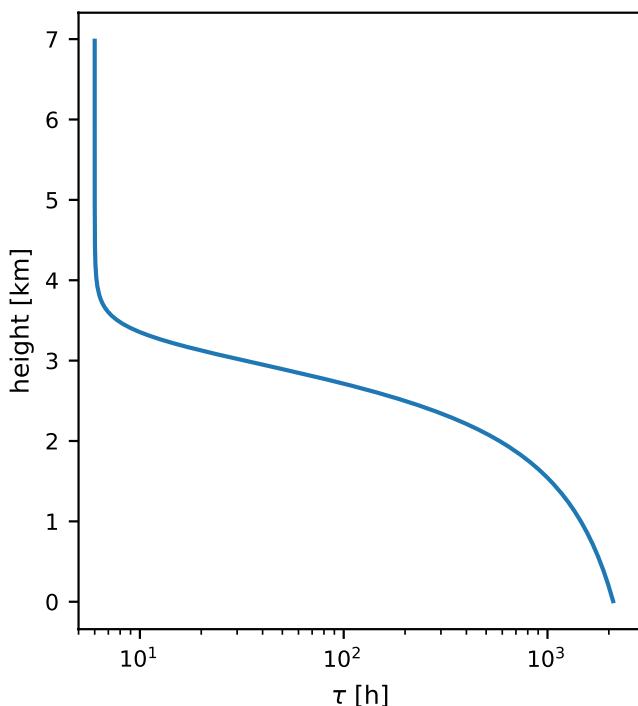


Figure 4. Height-dependence of the nudging time scale τ in Equation 7.

In this relation, the inverse tangent is centered around the top of the cloud layer: $z_{\max} = 3,000$ m. Below this height we wish the convection to develop freely, so we set the free parameters a , b , and c such that τ increases to around 3 months near the surface. In the free-tropospheric limit, where we would like to exercise some control over the profiles, we let the τ profile return to $\tau_{\infty} = 6$ hr. The fixed parameters of Equations 5–7 are listed in Table 1, and a profile of the resulting τ is shown in Figure 4.

4. Workflow to Create the Data Set

To turn the LES ensemble design into accessible data products, four steps need to be taken: (a) creating a set of input files for each ensemble member, (b) running each simulation, (c) converting simulation output to an easily accessible format, and (d) uploading the data set to a data repository. In this section, we briefly document how we carry out these steps.

To produce the input files required to run each ensemble member, we used a Python script and EasyVVUQ (Groen et al., 2021), a framework for uncertainty quantification. EasyVVUQ can sample a parameter space using different sampling strategies, for example, based on quadrature methods suitable for uncertainty quantification methods. EasyVVUQ then produces model input files, using a template where the varied parameters are substituted. We use this mechanism to produce a Fortran namelist, which is the main DALES configuration file, for each ensemble member. The input files for the initial vertical profiles of the prognostic variables are produced with a Python script. The setup for using EasyVVUQ to run DALES experiments was presented in Jansson et al. (2021).

All simulations were run with DALES on supercomputer Fugaku. Fugaku is based on the Fujitsu A64FX CPU, built on the ARM architecture with Scalable Vector Extension. Each node of Fugaku has 48 CPU cores and 32 GB RAM, and is characterized by a high memory bandwidth and fast node interconnect. Fugaku is a CPU-only system, that is, it does not rely on accelerators such as GPUs (graphics processing units). These properties seem to be a good fit for DALES, a CPU-only code, which in our experience is often memory-bandwidth limited and able to benefit from vectorized floating point mathematical operations. Porting DALES to Fugaku did not require extensive changes to the program, and was mostly a matter of adding the option to use the Fujitsu Fortran compiler. For improved performance, the possibility to store the prognostic fields in single precision was implemented. The single precision version is faster and requires less memory for storing the prognostic fields. The latter is particularly important on Fugaku which has relatively little RAM memory per core (around 600 MB). Further optimizations included rewriting and simplifying some loops for better vectorization, based on profiling the program. These modifications have been found to be beneficial on other architectures as well, enabling us to maintain a single version of the code for all architectures. See Data Availability Statement section for details of the DALES version used and for accessing the code.

DALES is parallelized using Message Passing Interfaces (MPIs) in x and y , the two horizontal directions. Each simulation was run on 24 nodes, with 24×48 MPI processes. The simulations lasted around 5 days (wall-clock time of running the simulation) per ensemble member. More details on the computational requirements of one specific ensemble member, are shown in Table 3, compared with a similar run on Snellius, the Dutch national supercomputer. The results show that at the beginning of the simulation,

Table 3

Computational Resources Used for the Simulation of Ensemble Member 1, the Central Point in the Parameter Hypercube, on Supercomputer Fugaku (One 2.0 GHz 48-Core A64FX CPU Per Node), and on the Dutch National Supercomputer Snellius (Two 2.6 GHz 64-Core AMD Rome 7H12 CPUs Per Node)

System	Time of simulation (hr)	Nodes	Cores	Wall clock time (s)	Time per grid point and time step (μs)
Fugaku	0–12	24	1,152	46,760	2.8
	36–48	24	1,152	110,391	5.6
Snellius	0–12	8	1,024	73,717	4.0
	36–48	8	1,024	92,094	4.5

DALES runs faster on Fugaku, comparing time required per grid point and time step. Further into the simulation, DALES runs slower on both systems, with a larger slowdown seen on Fugaku. This behavior is result of the cloud microphysics and precipitation scheme which is activated when precipitation occurs. This scheme has not been tuned for Fugaku yet, and appears to vectorize poorly. What the table doesn't show is that the scaling efficiency for using more nodes is better on Fugaku.

Each MPI process writes the output data for its own part of the simulation domain in the netCDF format. We used the uncompressed netCDF3 format, since it was found to require less RAM memory than netCDF4 during simulation. These netCDF tiles were then merged and converted to compressed netCDF4 using CDO 2.0.4 (Schulzweida, 2021).

Finally, the netCDF files were converted to the Zarr format (Miles et al., 2022) and uploaded to the German Climate Computing Center (DKRZ)'s SWIFT object storage for easy access, as described further in Section 5. As a backup, the netCDF files are kept on the tape archive of the European Center for Medium-Range Weather Forecasts (ECMWF).

5. Data Set Description

Cloud Botany contains a rich set of idealized large-eddy simulations that provide valuable resources to study the dependency of shallow cumulus convection to environmental conditions. In addition to the vast range of environments, the large domain-size itself allows an investigation of scales that remain uncaptured in previous simulation studies of trade-wind cumuli centered around the RICO (Rain in Cumulus over the Ocean) (vanZanten et al., 2011) and BOMEX (Barbados Oceanographic and Meteorological Experiment) (Siebesma et al., 2003) campaigns. Due to these large opportunities, we put additional effort into providing an easy and free access to these simulations.

We acknowledge that the download of 40 TB of simulation output is a burden and most users will only access portions of this data set, for example, specific timesteps, specific members or height levels. To allow for a more modular access, the data set has been chunked along all its dimensions and saved as Zarr files which support these chunks. The Zarr file format allows hosting Cloud Botany on the DKRZ SWIFT object storage. The combination of the Zarr format with an object storage leads to faster access rates compared to traditional file system based hosted data sets and make the Cloud Botany data set analysis ready.

An analysis in Python can be started by accessing the EUREC⁴A intake catalog (https://howto.eurec4a.eu/botany_dales.html):

```
import eurec4a
cat = eurec4a.get_intake_catalog()
botany_cat = cat.simulations.DALES.botany
```

Further details on how to visualize and analyze this data set can be found in the interactive How-To-EUREC⁴A book at <https://howto.eurec4a.eu> among other EUREC⁴A related data sets.

All the simulations in the Cloud Botany ensemble are listed in Table A1 together with their parameters. Run 1 is at the center of the parameter hypercube, runs 2 to 65 are at its corners. The remaining runs 66 to 103, labeled “sweep,” lie on lines through the center of the hypercube, where one parameter at a time is varied. The remark column gives subjective description of the clouds and cloud organization (from the second day on, when patterns and organization generally occur) based on visual inspection.

The data is divided into several data sets, according to output frequency and dimensionality. Each data set is indexed by ensemble member, time and spatial coordinates. The data sets and their variables are summarized in Tables A2–A9. In general, we have stored 3D fields and 2D radiation fields hourly, and 2D fields such as the liquid water path as well as horizontal cross sections of the prognostic variables every 5 min.

As an aid to navigating the ensemble, we have prepared a web page with a set of plots and animations for each member. This page and the images and animations can be downloaded and used offline (Jansson, Janssens, Grönqvist, Siebesma, et al., 2023).

6. Results

In this section, we include a preliminary exposition of the simulation results, in terms of horizontally averaged cloud properties, and the development of mesoscale cloud patterns. Our aim will be to give a brief impression of the ensemble's variability, both within a single simulation and between ensemble members, to motivate future quantitative, in-depth analyses of the data set.

6.1. Time-Evolution in a Single Ensemble Member

We begin with Figure 5, which shows the evolution of several quantities of interest and snapshots of the cloud cover and precipitation in simulation 1, the center point of our parameter space. Its evolution is qualitatively similar to that of many ensemble members, and the Lagrangian simulations by Narenpitak et al. (2021). All simulations depart from cloud-free states at midnight UTC. The first 10 hr are characterized by the onset of convection and the development of small, unorganized cumuli. These non-precipitating clouds then gradually cluster into larger structures. This evolution is modulated by the diurnal cycle of shortwave radiation. After sunrise, shortwave heating gradually stabilizes the domain to convection, reducing both the cloud fraction and horizontally averaged LWP. After sunset, the cloud structures rapidly grow vertically and begin to precipitate around 24 hr from the start of the simulation. The second diurnal cycle is then dominated by larger, precipitating convection cells, organized along cold pools and frequently topped by thin inversion clouds.

The cloud cover reached by simulation 1 (0.12, equal to the ensemble-average) is akin to earlier LESs of the trades (0.13 in Radtke et al. (2021) or 0.19 in vanZanten et al. (2011)). However, it is lower than what was observed during the EUREC⁴A field campaign (0.42 including optically thin clouds, based on both satellites and aircraft-borne lidar measurements, Mieslinger et al., 2022). This underprediction of cloud cover remains across the ensemble (Figure 6). Understanding the causes of this bias, and whether it affects the sensitivities of the cloudiness to changing the environment, are worthwhile topics of follow-up research.

6.2. Variability in Cloudiness, Rain and Mesoscale Organization Between Simulations

While most simulations develop qualitatively similarly, they produce rather different cloud fractions, and, somewhat independently, different surface rain rates (Figure 6). Hence, changing our environmental control parameters influences both the shallow convection's coverage and depth. This is further illustrated by Figure 7, which shows how vertical profiles of cloud fraction vary under parameter sweeps. At a glance, many of these sensitivities appear to match earlier smaller-domain LES sensitivity studies: the cloud fraction profiles are rather insensitive to changes in θ_{10} , in agreement with for example Bretherton et al. (2013), and stronger surface winds give deeper cloud layers with more inversion cloud, in line with Nuijens and Stevens (2012). Following Bellon and Stevens (2012), we observe that stabilizing the free troposphere through increasing Γ lowers the inversion and increases its cloudiness, while moistening the free troposphere through increasing h_q lowers the inversion and cloud base. Finally, reducing the subsidence strength through increasing w_1 gives rise to increased inversion cloud, in line with Vogel et al. (2020). The apparent similarities between the sensitivity of cloud fraction in our large-domain simulations, which include mesoscale organization, and the small-domain simulations in literature, which do not, motivate more rigorous analysis, to understand the specific role played by self-organization in setting the cloudiness in the trades. Therefore, we plan to repeat all simulations presented herein in small (10 km) domains.

Figure 8 shows a few examples of cloud patterns that develop under different parameter combinations. As anticipated by Figure 6, most of these develop precipitating convection, almost always paired with cold pools. When they occur, such cold pools visually appear to substantially modulate the cloud patterning. We find at least three different ways in which this happens. First, we sometimes find cold pools lined by arcs of cumuli (e.g., Figure 8a). Second, in simulations with strong surface wind, large lapse rates, small moisture scale heights and positive large-scale vertical velocity (e.g., runs 37, 45 in Figures 8b and 8c and 13, 79, not shown), cold pools are produced by sufficiently vigorous convective cells that they produce large (>50 km) sheets of thin, stratiform

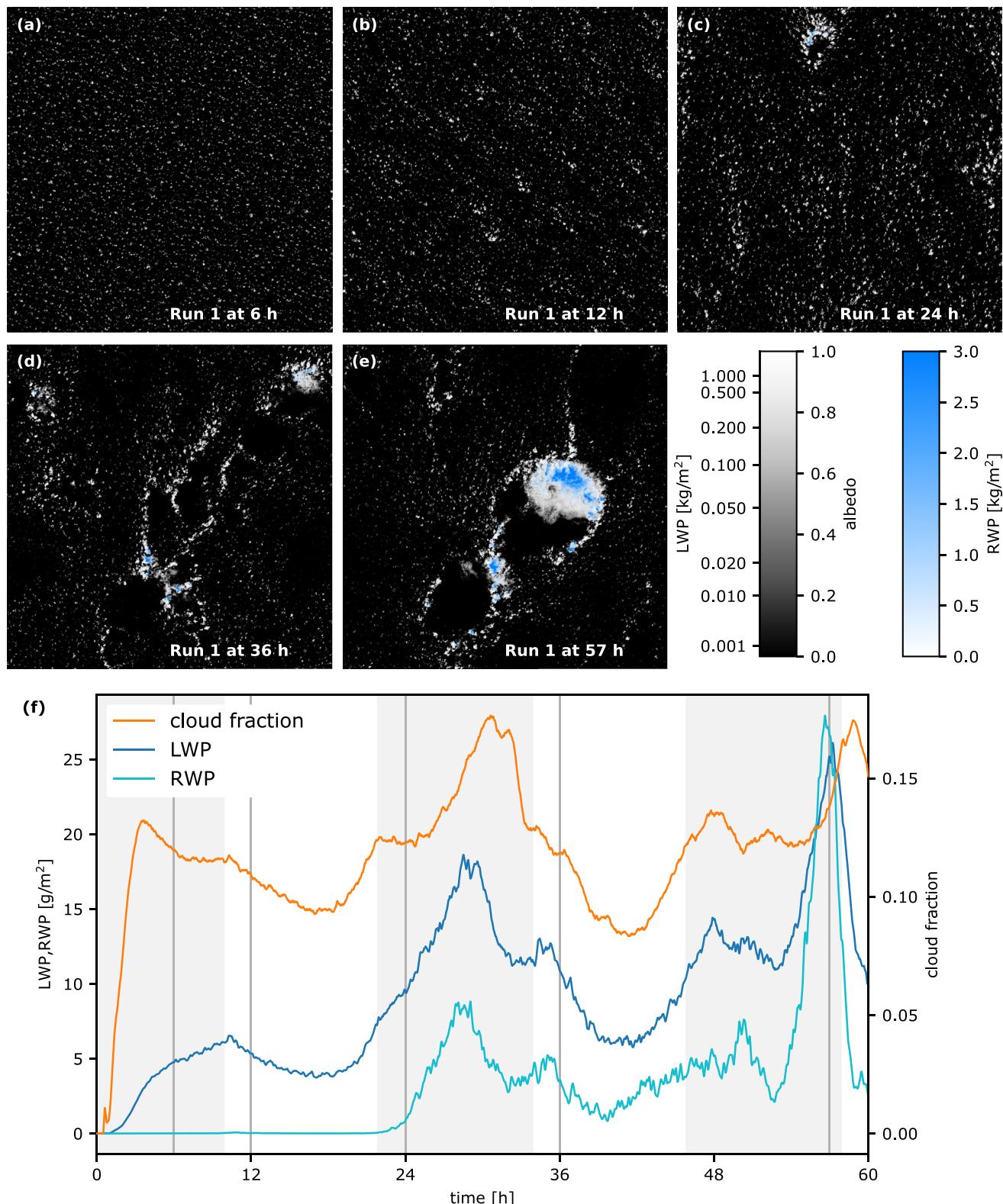


Figure 5. Time series of simulation 1, the central point of the parameter hypercube. The snapshots show cloud albedo in white (as parameterized by Y. Zhang et al. (2005)) and rain water path in blue. The time series curves show the liquid water path (LWP), rain water path (RWP) and cloud fraction over time. The shaded background shows the diurnal cycle, the darker regions are night (18–06 hr in local time). The times of the snapshots are indicated by gray vertical lines.

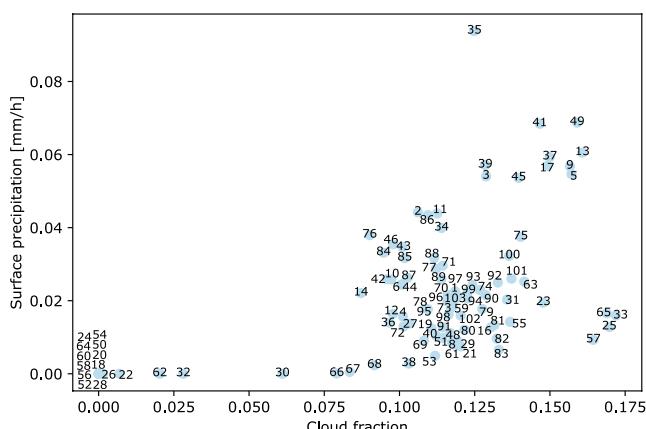


Figure 6. Scatter plot of the cloud fraction and surface precipitation of the (numbered) ensemble members, averaged over the last 24 hr of the simulations. Ensemble members that did not reach 60 hr are excluded.

outflow layers, reminiscent of the structures termed “Flowers” by Stevens et al. (2020). At a glance, the appearance of such structures under stronger stratification and higher surface winds appears consistent with the observations by Bony et al. (2020), Schulz et al. (2021), and follows from Figures 7b and 7c. Third, in runs 84 to 91 the wind shear is varied. At strong wind shears of both positive and negative signs, cold pools are observed to deform into bands (Figure 8d); such features are also found in runs 4 and 100.

We also find a set of simulations with less vigorous, at most weakly precipitating convection (lower left corner in Figure 6), often at lower surface winds. When the winds blow weakly, the large-scale vertical velocity has a strong switching effect on the cloud formation: Negative w_1 often results in very weak, sometimes cloud-free convection (e.g., runs 18 and 50); merely switching w_1 to its positive counterpart (runs 19 and 51) makes them produce deeper, precipitating convection. Yet strikingly, all non- or weakly precipitating simulations that support a cumulus layer still see their convection organize into mesoscale patterns (e.g., runs 8, 66–68), such as bands aligned with the mean wind (Figure 8e) or into quasi-circular clusters (Figure 8f). Figure 9 shows 3D renderings of parts of scenes b and e of Figure 8.

In summary, we can identify at least five visually distinct forms of convective patterning in our ensemble, several of which appear to match visually identified categories of cloud patterns in nature: “Sugar,” “Gravel,” the aforementioned “Flowers” and “Fish” (Stevens et al., 2020). In order of increasing visual complexity, we find simulations with (a) no clouds, (b) small, randomly spaced cumulus, (c) clustered, non-precipitating cumulus (both ii and iii seem to fit the “Sugar” category), (d) precipitating convection and subsequent cold pools (“Gravel”), and mesoscale convective systems topped by thin stratiform clouds (“Flowers”). Patterns larger than our domain size, such as “Fish” we cannot simulate; our data set therefore cannot shed light on their formation. However, since Schulz et al. (2021); Aemisegger et al. (2021) show that “Fish” originate in extratropical, synoptic disturbances, there is also no reason to expect such structures to form spontaneously in even larger-domain simulations forced by conditions that characterize the trades.

6.3. Length of Integration Period

To see the longer-time development we have repeated the simulation for the central point of the parameter space (run 1) for a longer period of 9 days, shown in Figure 10. Hours 24 to 60 of the simulation appear similar to the longer time behavior. As the long simulation was done on a different computer (Snellius, x86 architecture) and the model was compiled with a different compiler, the simulation trajectory is different, and thus comparing the two simulations indicates the randomness or internal variability in these simulations. The timing in the liquid water path peaks differ slightly, though their shapes and amplitudes are similar, and the mean statistics are unaffected (mean liquid water path and cloud fraction over 24–60 hr are within 1% of each other).

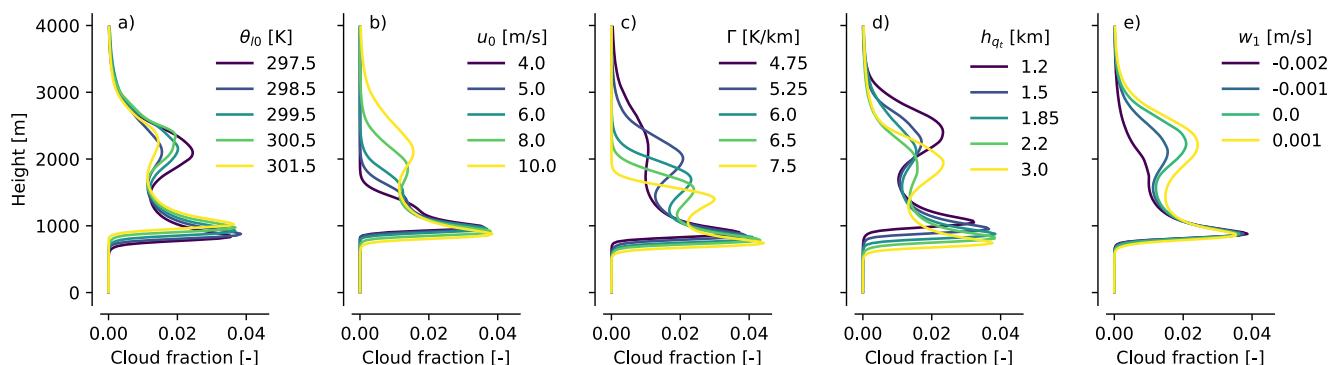


Figure 7. Variation of vertical profiles of cloud fraction with sweeps of the environmental control parameters (a) surface liquid water potential temperature, (b) near-surface wind, (c) temperature lapse rate, (d) humidity height scale, and (e) the subsidence parameter (colors), averaged over the last 24 hr of the simulations.

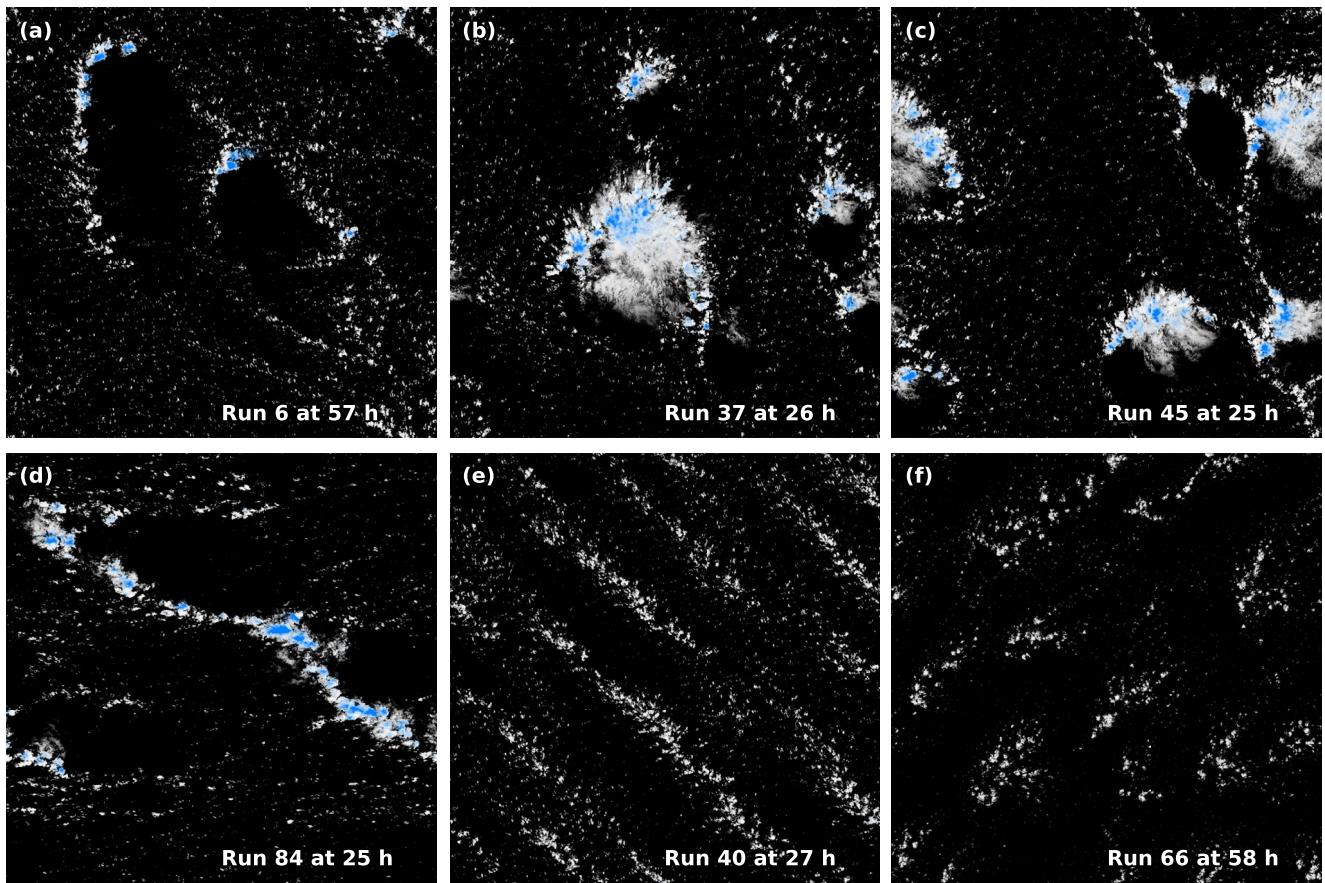


Figure 8. Different types of cloud organization seen in the Cloud Botany ensemble. (a) Run 6, cold pools, (b) run 37, large cloud cluster topped by stratiform outflow, (c) run 45, multiple such clusters, on the edges of cold pools, (d) run 84, line of precipitation, (e) run 40, non-precipitating cumulus in bands, and (f) run 66, non-precipitating cumulus in aggregated in quasi-circular clusters. The wind is easterly, that is, from the right side of the image.

6.4. Truncated Simulations

Runs 7, 15, 38, 39, and 47 did not finish due to a crash in the thermodynamics routine, when temperature and moisture reach non-physical values. All these runs have a low Γ , sometimes allowing the convection to permeate through to our domain top, where their spurious interactions with our boundary conditions makes them fail (see Section 2.3). Since we do not expect such deep convection to frequently occur during the suppressed conditions we aim to study, we recommend disregarding these runs. Additionally, runs 11, 14, 43, 46, and 87 only span 48 hr due to their computing jobs being interrupted.

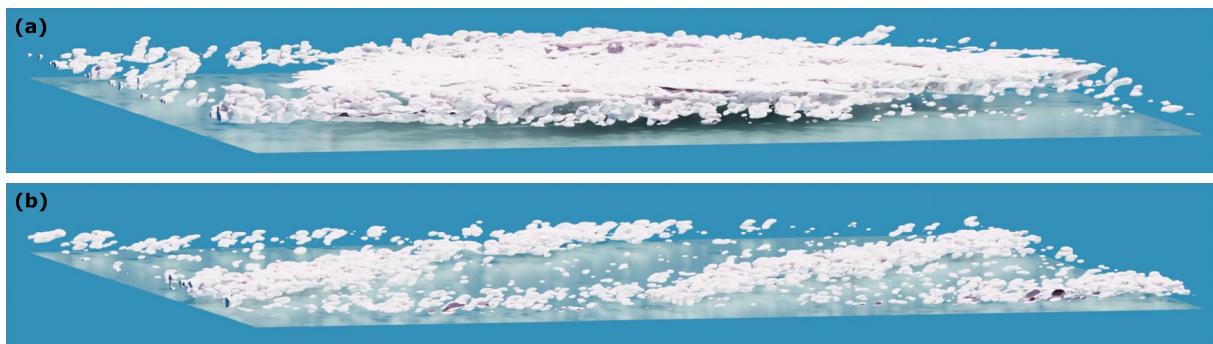


Figure 9. Rendered 3D view of (a) the central, large cloud structure in Figure 8b and of (b) the stripes in Figure 8b, above a reflecting plane representing the ocean. The rendered domain is 70 km \times 70 km. The rendering shows an isosurface of $q_i = 2 \cdot 10^{-5}$ kg/kg.

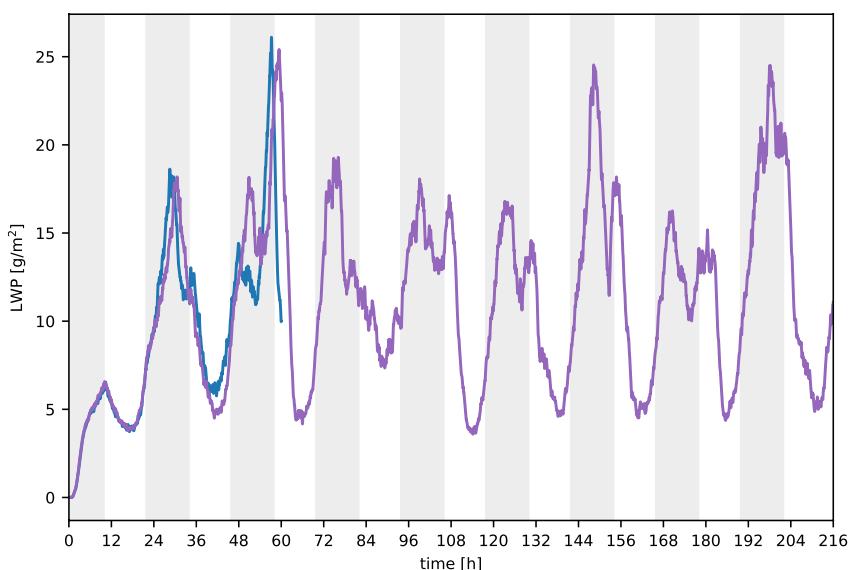


Figure 10. A 9-day simulation of the central point of the parameter space (purple), compared to the 60 hr simulation which is part of the ensemble (blue).

7. Conclusions

There are several approaches to improve understanding of the processes that underpin the rich spectrum of cloud patterns over the tropical ocean. Many attempts rest on the construction of models that strive for maximum realism across the entire relevant scale range, from the synoptics to the large-eddy scales of turbulence. In this paper, we have presented Cloud Botany, an ensemble of large-eddy simulations on 150 km domains that instead represents the larger-scale environment in a highly idealized manner. We do this to elucidate the processes through which shallow convection can self-organize into mesoscale cloud patterns, and to study systematically how these processes vary as the larger-scale environment changes.

We design our idealized large-scale environment by fitting functional forms of the vertical structure of liquid-water potential temperature, total specific humidity and horizontal wind from reanalysis, and vertical velocity from observations. For most of these, reasonable fits can be attained with very simple approximations, allowing us to span the range of observed conditions by varying only six parameters: these span a parameter space that we explore by simulating (a) all possible combinations of high and low values in the parameters that are representative for observed variability over the boreal winter of 2020, and (b) sweeps of single parameters.

In the Cloud Botany simulations, 93 out of 103 runs support cumulus-topped boundary layers. Strikingly, all those that do also self-organize into mesoscale cloud patterns. We typically first observe small, randomly spaced cumuli, which quickly begin self-aggregating into mesoscale clusters. After a marked diurnal cycle, we often observe the onset of precipitation after around 24 hr of simulation; subsequent cloud pattern varieties are dominated by cold pools and layers of thin inversion cloud. We also observe ample variability in the self-organized cloud patterns when we vary the parameters controlling the large-scale environment, all of which are closely reminiscent of cloud patterns observed in nature. We take these results to be early indications that parameter ensembles will prove fruitful for understanding the processes that govern the variability of the mesoscale trades, under a range of larger-scale conditions.

We hope this makes Cloud Botany a valuable community resource for studies that simultaneously require the resolution of individual cloud structures, a mesoscale environment and variability over a range of conditions characteristic for the trades. It also serves as a point of departure for using parameter ensembles to study variability in convective clouds in other regions of the world, or in warmer climates. Another use of the ensemble is as a benchmark for convective parameterizations for GCMs, since the task of such parameterizations is to compute cloud properties from vertical profiles of the model's prognostic variables. Finally, we see Cloud Botany as sitting on the abstract side of a spectrum of modeling approaches, which include simulation setups under time-varying forcings derived from a numerical weather prediction model (Savazzi, Nuijens, de Rooy, & Siebesma, 2022), on

the lateral boundaries of open domains (Dauhut et al., 2022), Lagrangian LES (Narenpitak et al., 2021), mesoscale models with parameterized convection (Beucher et al., 2022) and regional and global models with partially resolved convection (Schulz & Stevens, 2023; Stevens et al., 2019). All these will be needed to fully elucidate the subtleties that govern the interactions between clouds, their environment and climate at the trade-wind mesoscales.

Appendix A

This appendix contains tables describing the Cloud Botany data set. Table A1 list the parameters of the ensemble members. Tables A2–A9 give details of the variables stored in the different data sets.

Table A1
Summary of All Simulations in the Cloud Botany Ensemble

Run	θ_{10} (K)	u_0 (m/s)	$q_{t,ml}$ (g/kg)	h_{q_f} (m)	Γ (K/km)	w_i (cm/s)	u_z (s ⁻¹)	Location	Remark
1	298.5	-10	14.25	1,850	5	-0.085	0.0022	Center	Cold pools
2	297.5	-15	13.5	1,200	4.5	-0.35	0.0022	Corner	Cold pools
3	297.5	-15	13.5	1,200	4.5	0.18	0.0022	Corner	Cold pools
4	297.5	-15	13.5	1,200	5.5	-0.35	0.0022	Corner	Cold pools
5	297.5	-15	13.5	1,200	5.5	0.18	0.0022	Corner	Cold pools
6	297.5	-15	13.5	2,500	4.5	-0.35	0.0022	Corner	Cold pools
7	297.5	-15	13.5	2,500	4.5	0.18	0.0022	Corner	Thermo. crash
8	297.5	-15	13.5	2,500	5.5	-0.35	0.0022	Corner	Weak precip.
9	297.5	-15	13.5	2,500	5.5	0.18	0.0022	Corner	Cold pools
10	297.5	-15	15	1,200	4.5	-0.35	0.0022	Corner	Cold pools
11	297.5	-15	15	1,200	4.5	0.18	0.0022	Corner	Cold pools
12	297.5	-15	15	1,200	5.5	-0.35	0.0022	Corner	Cold pools
13	297.5	-15	15	1,200	5.5	0.18	0.0022	Corner	Cold pools, flower
14	297.5	-15	15	2,500	4.5	-0.35	0.0022	Corner	Cold pools
15	297.5	-15	15	2,500	4.5	0.18	0.0022	Corner	Thermo. crash
16	297.5	-15	15	2,500	5.5	-0.35	0.0022	Corner	Cold pools
17	297.5	-15	15	2,500	5.5	0.18	0.0022	Corner	Cold pools
18	297.5	-5	13.5	1,200	4.5	-0.35	0.0022	Corner	No clouds
19	297.5	-5	13.5	1,200	4.5	0.18	0.0022	Corner	Cold pools
20	297.5	-5	13.5	1,200	5.5	-0.35	0.0022	Corner	No clouds
21	297.5	-5	13.5	1,200	5.5	0.18	0.0022	Corner	Cold pools
22	297.5	-5	13.5	2,500	4.5	-0.35	0.0022	Corner	Small cumulus
23	297.5	-5	13.5	2,500	4.5	0.18	0.0022	Corner	Cold pools
24	297.5	-5	13.5	2,500	5.5	-0.35	0.0022	Corner	No clouds
25	297.5	-5	13.5	2,500	5.5	0.18	0.0022	Corner	Cold pools
26	297.5	-5	15	1,200	4.5	-0.35	0.0022	Corner	Small cumulus
27	297.5	-5	15	1,200	4.5	0.18	0.0022	Corner	Cold pools
28	297.5	-5	15	1,200	5.5	-0.35	0.0022	Corner	Small cumulus
29	297.5	-5	15	1,200	5.5	0.18	0.0022	Corner	Cold pools
30	297.5	-5	15	2,500	4.5	-0.35	0.0022	Corner	Organizing cumulus
31	297.5	-5	15	2,500	4.5	0.18	0.0022	Corner	Cold pools
32	297.5	-5	15	2,500	5.5	-0.35	0.0022	Corner	Small organizing cumulus
33	297.5	-5	15	2,500	5.5	0.18	0.0022	Corner	Cold pools
34	299.5	-15	13.5	1,200	4.5	-0.35	0.0022	Corner	Cold pools

Table A1
Continued

Run	θ_{10} (K)	u_0 (m/s)	$q_{t,ml}$ (g/kg)	h_{q_t} (m)	Γ (K/km)	w_1 (cm/s)	u_z (s ⁻¹)	Location	Remark
35	299.5	-15	13.5	1,200	4.5	0.18	0.0022	Corner	Cold pools
36	299.5	-15	13.5	1,200	5.5	-0.35	0.0022	Corner	Cold pools
37	299.5	-15	13.5	1,200	5.5	0.18	0.0022	Corner	Cold pools, aggr. clouds, flower
38	299.5	-15	13.5	2,500	4.5	-0.35	0.0022	Corner	Cold pools, thermo. crash
39	299.5	-15	13.5	2,500	4.5	0.18	0.0022	Corner	Cold pools, thermo. crash
40	299.5	-15	13.5	2,500	5.5	-0.35	0.0022	Corner	Arcs
41	299.5	-15	13.5	2,500	5.5	0.18	0.0022	Corner	Cold pools
42	299.5	-15	15	1,200	4.5	-0.35	0.0022	Corner	Cold pools
43	299.5	-15	15	1,200	4.5	0.18	0.0022	Corner	Cold pools
44	299.5	-15	15	1,200	5.5	-0.35	0.0022	Corner	Cold pools
45	299.5	-15	15	1,200	5.5	0.18	0.0022	Corner	Cold pools, aggr. clouds, flower
46	299.5	-15	15	2,500	4.5	-0.35	0.0022	Corner	Cold pools
47	299.5	-15	15	2,500	4.5	0.18	0.0022	Corner	Thermo. crash
48	299.5	-15	15	2,500	5.5	-0.35	0.0022	Corner	Cold pools
49	299.5	-15	15	2,500	5.5	0.18	0.0022	Corner	Cold pools
50	299.5	-5	13.5	1,200	4.5	-0.35	0.0022	Corner	No clouds
51	299.5	-5	13.5	1,200	4.5	0.18	0.0022	Corner	Cold pools
52	299.5	-5	13.5	1,200	5.5	-0.35	0.0022	Corner	No clouds
53	299.5	-5	13.5	1,200	5.5	0.18	0.0022	Corner	Cold pools
54	299.5	-5	13.5	2,500	4.5	-0.35	0.0022	Corner	No clouds
55	299.5	-5	13.5	2,500	4.5	0.18	0.0022	Corner	Cold pools
56	299.5	-5	13.5	2,500	5.5	-0.35	0.0022	Corner	No clouds
57	299.5	-5	13.5	2,500	5.5	0.18	0.0022	Corner	Cold pools
58	299.5	-5	15	1,200	4.5	-0.35	0.0022	Corner	No clouds
59	299.5	-5	15	1,200	4.5	0.18	0.0022	Corner	Cold pools
60	299.5	-5	15	1,200	5.5	-0.35	0.0022	Corner	No clouds
61	299.5	-5	15	1,200	5.5	0.18	0.0022	Corner	Cold pools
62	299.5	-5	15	2,500	4.5	-0.35	0.0022	Corner	Small cumulus
63	299.5	-5	15	2,500	4.5	0.18	0.0022	Corner	Cold pools
64	299.5	-5	15	2,500	5.5	-0.35	0.0022	Corner	No clouds
65	299.5	-5	15	2,500	5.5	0.18	0.0022	Corner	Cold pools
66	298.5	-4	14.25	1,850	5	-0.085	0.0022	Sweep u_0	Organizing cumulus, weak precip.
67	298.5	-5	14.25	1,850	5	-0.085	0.0022	Sweep u_0	Organizing cumulus, weak precip.
68	298.5	-6	14.25	1,850	5	-0.085	0.0022	Sweep u_0	Organizing cumulus, cold pools
69	298.5	-8	14.25	1,850	5	-0.085	0.0022	Sweep u_0	Cold pools
70	298.5	-12	14.25	1,850	5	-0.085	0.0022	Sweep u_0	Cold pools
71	298.5	-15	14.25	1,850	5	-0.085	0.0022	Sweep u_0	Cold pools
72	298.5	-10	14.25	1,850	5	-0.2	0.0022	Sweep w_1	Cold pools
73	298.5	-10	14.25	1,850	5	-0.1	0.0022	Sweep w_1	Cold pools
74	298.5	-10	14.25	1,850	5	0	0.0022	Sweep w_1	Cold pools
75	298.5	-10	14.25	1,850	5	0.1	0.0022	Sweep w_1	Cold pools
76	298.5	-10	14.25	1,850	4	-0.085	0.0022	Sweep Γ	Cold pools
77	298.5	-10	14.25	1,850	4.5	-0.085	0.0022	Sweep Γ	Cold pools

Table A1
Continued

Run	θ_{10} (K)	u_0 (m/s)	$q_{t,ml}$ (g/kg)	h_{q_t} (m)	Γ (K/km)	w_1 (cm/s)	u_z (s ⁻¹)	Location	Remark
78	298.5	-10	14.25	1,850	4.75	-0.085	0.0022	Sweep Γ	Cold pools
79	298.5	-10	14.25	1,850	5.25	-0.085	0.0022	Sweep Γ	Cold pools, aggregated clouds
80	298.5	-10	14.25	1,850	5.5	-0.085	0.0022	Sweep Γ	Cold pools
81	298.5	-10	14.25	1,850	6	-0.085	0.0022	Sweep Γ	Cold pools
82	298.5	-10	14.25	1,850	6.5	-0.085	0.0022	Sweep Γ	Cold pools
83	298.5	-10	14.25	1,850	7.5	-0.085	0.0022	Sweep Γ	Cold pools
84	298.5	-10	14.25	1,850	5	-0.085	-0.0044	Sweep u_z	Precip. and bands
85	298.5	-10	14.25	1,850	5	-0.085	-0.0033	Sweep u_z	Precip. and bands
86	298.5	-10	14.25	1,850	5	-0.085	-0.0022	Sweep u_z	Bands and arcs
87	298.5	-10	14.25	1,850	5	-0.085	-0.0011	Sweep u_z	Cold pools
88	298.5	-10	14.25	1,850	5	-0.085	0	Sweep u_z	Cold pools
89	298.5	-10	14.25	1,850	5	-0.085	0.0011	Sweep u_z	Cold pools
90	298.5	-10	14.25	1,850	5	-0.085	0.0033	Sweep u_z	Cold pools
91	298.5	-10	14.25	1,850	5	-0.085	0.0044	Sweep u_z	Arcs, bands
92	297.5	-10	14.25	1,850	5	-0.085	0.0022	Sweep θ_{10}	Cold pools
93	299.5	-10	14.25	1,850	5	-0.085	0.0022	Sweep θ_{10}	Cold pools
94	300.5	-10	14.25	1,850	5	-0.085	0.0022	Sweep θ_{10}	Cold pools
95	301.5	-10	14.25	1,850	5	-0.085	0.0022	Sweep θ_{10}	Cold pools
96	298.5	-10	14.25	800	5	-0.085	0.0022	Sweep h_{q_t}	Cold pools
97	298.5	-10	14.25	1,200	5	-0.085	0.0022	Sweep h_{q_t}	Cold pools
98	298.5	-10	14.25	1,500	5	-0.085	0.0022	Sweep h_{q_t}	Cold pools
99	298.5	-10	14.25	2,200	5	-0.085	0.0022	Sweep h_{q_t}	Cold pools
100	298.5	-10	14.25	2,500	5	-0.085	0.0022	Sweep h_{q_t}	Cold pools, arcs
101	298.5	-10	14.25	3,000	5	-0.085	0.0022	Sweep h_{q_t}	Cold pools
102	298.5	-10	13.5	1,850	5	-0.085	0.0022	Sweep $q_{t,ml}$	Cold pools
103	298.5	-10	15	1,850	5	-0.085	0.0022	Sweep $q_{t,ml}$	Cold pools

Table A2
Variables in the Timeseries Data Set, Sampled Every Minute

Variable	Units	Description
cfrac	—	Cloud fraction
lmax	kg/kg	Maximum liquid water specific humidity
lwp_bar	kg/m ²	Slab-averaged liquid-water path
lwp_max	kg/m ²	Maximum liquid-water path
obukh	m	Obukhov length
qtstr	K	Turbulent humidity scale
rwp_bar	kg/m ²	Rain water path
thlskin	K	Surface liquid water potential temperature
tstr	K	Turbulent temperature scale
twp_bar	kg/m ²	Total water path
ustar	m/s	Surface friction velocity
we	m/s	Entrainment velocity

Table A2
Continued

Variable	Units	Description
wmax	m/s	Maximum vertical velocity
wq	kg/kg m/s	Surface kinematic moisture flux
wtheta	K m/s	Surface kinematic potential temperature flux
wthetav	K m/s	Surface kinematic virtual potential temperature flux
z0	m	Roughness height
zb	m	Cloud-base height
zc_av	m	Average Cloud-top height
zc_max	m	Maximum Cloud-top height
zi	m	Boundary layer height

Note. Dimensions: (member, time).

Table A3
Variables in the Profiles Data Set Containing Horizontally Averaged Profiles, Sampled Every 5 Min, Part 1

Variable	Units	Description
cfrac	—	Cloud fraction
cs	—	Smagorinsky constant
dvrmn	m	Precipitation mean diameter
lwd	W/m ²	Long wave downward radiative flux
lwdca	W/m ²	Long wave clear air downward radiative flux
lwu	W/m ²	Long wave upward radiative flux
lwuca	W/m ²	Long wave clear air upward radiative flux
npaccr	#/m ³ /s	Accretion rain drop tendency
npauto	#/m ³ /s	Autoconversion rain drop tendency
npevap	#/m ³ /s	Evaporation rain drop tendency
npsed	#/m ³ /s	Sedimentation rain drop tendency
nptot	#/m ³ /s	Total rain drop tendency
nrrain	#/m ³	Rain droplet number concentration
preccount	—	Precipitation flux area fraction
precmn	W/m ²	Rain rate
presh	Pa	Pressure at cell center
ql	kg/kg	Liquid water specific humidity
ql2r	(kg/kg) ²	Resolved liquid water variance
qrmn	kg/kg	Precipitation specific humidity
qrpaccr	kg/kg/s	Accretion rain water content tendency
qrpauto	kg/kg/s	Autoconversion rain water content tendency
qrpevap	kg/kg/s	Evaporation rain water content tendency
qrpsed	kg/kg/s	Sedimentation rain water content tendency
qrptot	kg/kg/s	Total rain water content tendency
qt	kg/kg	Total water specific humidity
qt2D	kg ² /kg ² /s	Dissipation of <i>qt</i> variance
qt2Pr	kg ² /kg ² /s	Resolved production of <i>qt</i> variance
qt2Ps	kg ² /kg ² /s	SFS production of <i>qt</i> variance

Table A3
Continued

Variable	Units	Description
qt2Res	$\text{kg}^2/\text{kg}^2/\text{s}$	Residual of qt budget
qt2S	$\text{kg}^2/\text{kg}^2/\text{s}$	Source of qt variance
qt2Tr	$\text{kg}^2/\text{kg}^2/\text{s}$	Resolved transport of qt variance
qt2r	$(\text{kg}/\text{kg})^2$	Resolved total water variance
qt2tendf	$\text{kg}^2/\text{kg}^2/\text{s}$	Tendency of qt variance

Note. Dimensions: (member, time, z).

Table A4
Variables in the Profiles Data Set Containing Horizontally Averaged Profiles, Sampled Every 5 Min, Part 2

Variable	Units	Description
raincount	—	Rain water content area fraction
rainrate	W/m^2	Echo rain rate
rhobf	kg/m^3	Full level base-state density
robh	kg/m^3	Half level base-state density
rhof	kg/m^3	Full level slab averaged density
skew	—	Vertical velocity skewness
sv001	(kg/kg)	Scalar 001 specific mixing ratio, rain droplet number
sv0012r	$(\text{kg}/\text{kg})^2$	Resolved scalar 001 variance
sv002	(kg/kg)	Scalar 002 specific mixing ratio, specific rain water content
sv0022r	$(\text{kg}/\text{kg})^2$	Resolved scalar 002 variance
svp001	$(\text{kg}/\text{kg}/\text{s})$	Scalar 001 tendency
svp002	$(\text{kg}/\text{kg}/\text{s})$	Scalar 002 tendency
svpt001	$(\text{kg}/\text{kg}/\text{s})$	Scalar 001 turbulence tendency
svpt002	$(\text{kg}/\text{kg}/\text{s})$	Scalar 002 turbulence tendency
swd	W/m^2	Short wave downward radiative flux
swdca	W/m^2	Short wave clear air downward radiative flux
swu	W/m^2	Short wave upward radiative flux
swuca	W/m^2	Short wave clear air upward radiative flux
th2r	K^2	Resolved theta variance
thl	K	Liquid water potential temperature
thl2D	K^2/s	Dissipation of thl variance
thl2Pr	K^2/s	Resolved production of thl variance
thl2Ps	K^2/s	SFS production of thl variance
thl2Res	K^2/s	Residual of thl budget
thl2S	K^2/s	Source of thl variance
thl2Tr	K^2/s	Resolved transport of thl variance
thl2r	K^2	Resolved thl variance
thl2tendf	K^2/s	Tendency of thl variance
thllwtend	K/s	Long wave radiative tendency
thlradls	K/s	Prescribed large scale radiative tendency

Note. Dimensions: (member, time, z).

Table A5

Variables in the Profiles Data Set Containing Horizontally Averaged Profiles, Sampled Every 5 Min, Part 3

Variable	Units	Description
thlswtend	K/s	Short wave radiative tendency
thltend	K/s	Total radiative tendency
thv	K	Virtual potential temperature
thv2r	K ²	Resolved buoyancy variance
u	m/s	West–East velocity
u2r	m ² /s ²	Resolved horizontal velocity variance (u)
uwr	m ² /s ²	Resolved momentum flux (uw)
uws	m ² /s ²	SFS-momentum flux (uw)
uwt	m ² /s ²	Total momentum flux (uw)
v	m/s	South–North velocity
v2r	m ² /s ²	Resolved horizontal velocity variance (v)
vwr	m ² /s ²	Resolved momentum flux (vw)
vws	m ² /s ²	SFS-momentum flux (vw)
vwt	m ² /s ²	Total momentum flux (vw)
w2r	m ² /s ²	Resolved vertical velocity variance
w2s	m ² /s ²	SFS-TKE
wqlr	kg/kg m/s	Resolved liquid water flux
wqls	kg/kg m/s	SFS-liquid water flux
wqlt	kg/kg m/s	Total liquid water flux
wqtr	kg/kg m/s	Resolved moisture flux
wqts	kg/kg m/s	SFS-moisture flux
wqtt	kg/kg m/s	Total moisture flux
wsv001r	kg/kg m/s	Resolved scalar 001 flux
wsv001s	kg/kg m/s	SFS scalar 001 flux
wsv001 t	kg/kg m/s	Total scalar 001 flux
wsv002r	kg/kg m/s	Resolved scalar 002 flux
wsv002s	kg/kg m/s	SFS scalar 002 flux
wsv002 t	kg/kg m/s	Total scalar 002 flux
wthlr	Km/s	Resolved Theta_1 flux
wthls	Km/s	SFS-Theta_1 flux
wthlt	Km/s	Total Theta_1 flux
wthvr	Km/s	Resolved buoyancy flux
wthvs	Km/s	SFS-buoyancy flux
wthvt	Km/s	Total buoyancy flux

Note. Dimensions: (member, time, z).

Table A6

Variables in the 2D Data Set, Containing Horizontal Fields Sampled Every 5 Min, 2D

Variable	Units	Description
cldtop	m	xy cross sections cloud top height
hinvsrf	m	Height of surface inversion

Table A6
Continued

Variable	Units	Description
hmix	m	Mixed layer height
lwp	kg/m ²	xy cross sections liquid water path
rwp	kg/m ²	xy cross sections rain water path
surfprec	—	Surface precipitation
thetavmix	K	theta_v averaged over mixed layer
twp	kg/m ²	Total water path
umix	m/s	u averaged over mixed layer
vmix	m/s	v averaged over mixed layer

Note. Dimensions: (member, time, y, x).

Table A7

Variables in the 3D Data Set, the Full 3D Fields of the Model Sampled Every Hour

Variable	Units	Description
ql	kg/kg	Liquid water specific humidity
qt	kg/kg	Total water specific humidity
qr	kg/kg	Rain water specific humidity
thl	K	Liquid water potential temperature
u	m/s	West–East velocity
v	m/s	South–North velocity
w	m/s	Vertical velocity

Note. Dimensions: (member, time, z, y, x).

Table A8

Variables in the Cross_xy Data Set, Horizontal Cross-Sections of the Prognostic Variables Sampled Every 5 Min

Variable	Units	Description
qlxy	kg/kg	xy cross sections of the Liquid water specific humidity
qrxy	kg/kg	xy cross sections of the Rain water specific humidity
qtxy	kg/kg	xy cross sections of the Total water specific humidity
thlxy	K	xy cross sections of the Liquid water potential temperature
uxy	m/s	xy cross sections of the West–East velocity
vxy	m/s	xy cross sections of the South–North velocity
wxy	m/s	xy cross sections of the vertical velocity

Note. Dimensions: (z, y, x).

Table A9

Variables in the Radiation Data Set, 2D Radiation Fluxes Sampled Every Hour

Variable	Units	Description
clwvi	kg/m ²	Condensed water path
hf1s	W/m ²	Surface upward latent heat flux
hfss	W/m ²	Surface upward sensible heat flux

Table A9
Continued

Variable	Units	Description
prw	kg/m ²	Water vapor path
rlds	W/m ²	Surface downwelling longwave flux
rldscs	W/m ²	Surface downwelling longwave flux—clear sky
rlus	W/m ²	Surface upwelling longwave flux
rluscs	W/m ²	Surface upwelling longwave flux—clear sky
rlut	W/m ²	TOM outgoing longwave flux
rlutes	W/m ²	TOM outgoing longwave flux—clear sky
rlutoa	W/m ²	TOA outgoing longwave flux
rlutoacs	W/m ²	TOA outgoing longwave flux—clear sky
rsds	W/m ²	Surface downwelling shortwave flux
rsds_dif	W/m ²	Surface downwelling shortwave diffuse flux
rsds_dir	W/m ²	Surface downwelling shortwave direct flux
rsdscs	W/m ²	Surface downwelling shortwave flux—clear sky
rsdt	W/m ²	TOM incoming shortwave flux
rsdtoa	W/m ²	TOA incoming shortwave flux
rsus	W/m ²	Surface upwelling shortwave flux
rsuscs	W/m ²	Surface upwelling shortwave flux—clear sky
rsut	W/m ²	TOM outgoing shortwave flux
rsutes	W/m ²	TOM outgoing shortwave flux—clear sky
rsutoa	W/m ²	TOA outgoing shortwave flux
rsutoacs	W/m ²	TOA outgoing shortwave flux—clear sky
tabot	K	Air temperature at lowest model level
uabot	m/s	Eastward wind at lowest model level
vabot	m/s	Northward wind at lowest model level

Note. TOA stands for top of atmosphere, TOM for top of model. Dimensions: (member, time, y, x).

Data Availability Statement

The main data product in this article is the Cloud Botany ensemble, accessible through the EUREC⁴A intake catalog (EUREC4A Community, 2023), see <https://howto.eurec4a.eu/intro.html>. The version of DALES used for the Cloud Botany experiment is based on DALES v4.3 <https://doi.org/10.5281/zenodo.4604726> (Arabas et al., 2021), with modifications for running on Fugaku and for optimization. The exact version used is available on GitHub, <https://github.com/dalesteam/dales/tree/fugaku>, commit ca69c, and also archived as <https://doi.org/10.5281/zenodo.7405654> (Arabas et al., 2022). Support for running on Fugaku and most of the optimizations have subsequently been merged into DALES v4.4. The DALES input files for the ensemble, the Python scripts used to generate them, and Jupyter notebooks producing the article figures are archived at <https://doi.org/10.5281/zenodo.7709435> (Jansson, Janssens, & Grönqvist, 2023). The scripts generating Figures 5 and 8 serve as examples of accessing the Cloud Botany data through the intake catalog. An offline webpage containing basic profile and time-series plots as well as animations of all the ensemble members is available at <https://doi.org/10.5281/zenodo.7692270> (Jansson, Janssens, Grönqvist, Siebesma, et al., 2023).

References

Aemisegger, F., Vogel, R., Graf, P., Dahinden, F., Villiger, L., Jansen, F., et al. (2021). How Rossby wave breaking modulates the water cycle in the North Atlantic trade wind region. *Weather and Climate Dynamics*, 2(1), 281–309. <https://doi.org/10.5194/wcd-2-281-2021>

Agee, E. M. (1984). Observations from space and thermal convection: A historical perspective. *Bulletin of the American Meteorological Society*, 65(9), 938–949. [https://doi.org/10.1175/1520-0477\(1984\)065<0938:OFSATC>2.0.CO;2](https://doi.org/10.1175/1520-0477(1984)065<0938:OFSATC>2.0.CO;2)

Albright, A. L., Fildier, B., Touzé-Peiffer, L., Pincus, R., Vial, J., & Muller, C. (2021). Atmospheric radiative profiles during EUREC⁴A. *Earth System Science Data*, 13(2), 617–630. <https://doi.org/10.5194/essd-13-617-2021>

Arabas, S., Axelsen, S., Attema, J., Beets, C., Boeing, S. J., de Bruine, M., et al. (2021). *dalesteam/dales: Dales 4.3*. Zenodo. <https://doi.org/10.5281/zenodo.4604726>

Arabas, S., Axelsen, S., Attema, J., Beets, C., Boeing, S. J., de Bruine, M., et al. (2022). *dalesteam/dales: v4.3 for Cloud Botany on Fugaku*. Zenodo. <https://doi.org/10.5281/zenodo.7405654>

Bellon, G., & Stevens, B. (2012). Using the sensitivity of large-eddy simulations to evaluate atmospheric boundary layer models. *Journal of the Atmospheric Sciences*, 69(5), 1582–1601. <https://doi.org/10.1175/jas-d-11-0160.1>

Betts, A. K., & Ridgway, W. (1989). Climatic equilibrium of the atmospheric convective boundary layer over a tropical ocean. *Journal of the Atmospheric Sciences*, 46(17), 2621–2641. [https://doi.org/10.1175/1520-0469\(1989\)046<2621:ceotac>2.0.co;2](https://doi.org/10.1175/1520-0469(1989)046<2621:ceotac>2.0.co;2)

Beucher, F., Couvreux, F., Bouniol, D., Faure, G., Favot, F., Dauhut, T., & Ayet, A. (2022). Process-oriented evaluation of the oversea AROME configuration: Focus on the representation of cloud organisation. *Quarterly Journal of the Royal Meteorological Society*, 148(749), 3429–3447. <https://doi.org/10.1002/qj.4354>

Blossey, P. N., Bretherton, C. S., Zhang, M., Cheng, A., Endo, S., Heus, T., et al. (2013). Marine low cloud sensitivity to an idealized climate change: The CGILS LES intercomparison. *Journal of Advances in Modeling Earth Systems*, 5(2), 234–258. <https://doi.org/10.1002/jame.20025>

Bony, S., Schulz, H., Vial, J., & Stevens, B. (2020). Sugar, gravel, fish, and flowers: Dependence of mesoscale patterns of trade-wind clouds on environmental conditions. *Geophysical Research Letters*, 47(7), e2019GL085988. <https://doi.org/10.1029/2019gl085988>

Bretherton, C. S., & Blossey, P. N. (2017). Understanding mesoscale aggregation of shallow cumulus convection using large-eddy simulation. *Journal of Advances in Modeling Earth Systems*, 9(8), 2798–2821. <https://doi.org/10.1002/2017ms000981>

Bretherton, C. S., Blossey, P. N., & Jones, C. R. (2013). Mechanisms of marine low cloud sensitivity to idealized climate perturbations: A single-LES exploration extending the CGILS cases. *Journal of Advances in Modeling Earth Systems*, 5(2), 316–337. <https://doi.org/10.1002/jame.20019>

Dauhut, T., Couvreux, F., Bouniol, D., Beucher, F., Volkmer, L., Pörtge, V., et al. (2022). Flower trade-wind clouds are shallow mesoscale convective systems. *Quarterly Journal of the Royal Meteorological Society*, 149(750), 325–347. <https://doi.org/10.1002/qj.4409>

Denby, L. (2020). Discovering the importance of mesoscale cloud organization through unsupervised classification. *Geophysical Research Letters*, 47(1), e2019GL085190. <https://doi.org/10.1029/2019gl085190>

EUREC4A Community. (2023). EUREC4A (Version v1.0.0) [Dataset]. Zenodo. <https://doi.org/10.5281/ZENODO.8422321>

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

George, G., Stevens, B., Bony, S., Pincus, R., Fairall, C., Schulz, H., et al. (2021). JOANNE: Joint dropsonde Observations of the Atmosphere in tropical North atlantic meso-scale Environments. *Earth System Science Data*, 13(11), 5253–5272. <https://doi.org/10.5194/essd-13-5253-2021>

George, G., Stevens, B., Bony, S., Vogel, R., & Naumann, A. K. (2023). Widespread shallow mesoscale circulations observed in the trades. *Nature Geoscience*, 16(7), 584–589. <https://doi.org/10.1038/s41561-023-01215-1>

Glassmeier, F., Hoffmann, F., Johnson, J. S., Yamaguchi, T., Carslaw, K. S., & Feingold, G. (2021). Aerosol-cloud-climate cooling overestimated by ship-track data. *Science*, 371(6528), 485–489. <https://doi.org/10.1126/science.abd3980>

Groen, D., Arabnejad, H., Jancauskas, V., Edeling, W. N., Jansson, F., Richardson, R. A., et al. (2021). VECMAtk: A scalable verification, validation and uncertainty quantification toolkit for scientific simulations. *Philosophical Transactions of the Royal Society A: Mathematical, Physical & Engineering Sciences*, 379(2197), 20200221. <https://doi.org/10.1098/rsta.2020.0221>

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730), 1999–2049. <https://doi.org/10.1002/qj.3803>

Heus, T., van Heerwaarden, C. C., Jonker, H. J., Siebesma, A. P., Axelsen, S., van den Dries, K., et al. (2010). Formulation of and numerical studies with the Dutch Atmospheric Large-Eddy Simulation (DALES). *Geoscientific Model Development*, 3(2), 415–444. <https://doi.org/10.5194/gmd-3-415-2010>

Howard, L. (1803). I. On the modifications of clouds, and on the principles of their production, suspension, and destruction; being the substance of an essay read before the Askesian Society in the session 1802–3. *Philosophical Magazine*, 17(65), 5–11. <https://doi.org/10.1080/14786440308676365>

Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., & Collins, W. D. (2008). Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *Journal of Geophysical Research*, 113(D13), D13103. <https://doi.org/10.1029/2008jd009944>

Janssens, M., Vilà, J., van Heerwaarden, C. C., de Roode, S. R., Siebesma, A. P., & Glassmeier, F. (2022). Non-precipitating shallow cumulus convection is intrinsically unstable to length-scale growth. *Journal of the Atmospheric Sciences*, 80(3), 849–870. <https://doi.org/10.1175/JAS-D-22-0111.1>

Janssens, M., Vilà-Guerau de Arellano, J., Scheffer, M., Antonissen, C., Siebesma, A. P., & Glassmeier, F. (2021). Cloud patterns in the trades have four interpretable dimensions. *Geophysical Research Letters*, 48(5), e2020GL091001. <https://doi.org/10.1029/2020GL091001>

Jansson, F., Edeling, W., Attema, J., & Crommelin, D. (2021). Assessing uncertainties from physical parameters and modelling choices in an atmospheric large eddy simulation model. *Philosophical Transactions of the Royal Society A: Mathematical, Physical & Engineering Sciences*, 379(2197), 20200073. <https://doi.org/10.1098/rsta.2020.0073>

Jansson, F., Janssens, M., & Grönqvist, J. H. (2023). *fjansson/cloudbotanyscripts: v1.0.0*. Zenodo. <https://doi.org/10.5281/zenodo.7709435>

Jansson, F., Janssens, M., Grönqvist, J. H., Siebesma, P., Glassmeier, F., Attema, J., et al. (2023). *Cloud Botany LES ensemble visualizations*. Zenodo. <https://doi.org/10.5281/zenodo.7692270>

Lamaakel, O., & Matheou, G. (2022). Organization development in precipitating shallow cumulus convection: Evolution turbulence characteristics. *Journal of the Atmospheric Sciences*, 79(9), 2419–2433. <https://doi.org/10.1175/jas-d-21-0334.1>

Medeiros, B., & Nuijens, L. (2016). Clouds at Barbados are representative of clouds across the trade wind regions in observations and climate models. *Proceedings of the National Academy of Sciences*, 113(22), E3062–E3070. <https://doi.org/10.1073/pnas.1521494113>

Mieslinger, T., Stevens, B., Kölling, T., Brath, M., Wirth, M., & Buehler, S. A. (2022). Optically thin clouds in the trades. *Atmospheric Chemistry and Physics*, 22(10), 6879–6898. <https://doi.org/10.5194/acp-22-6879-2022>

Miles, A., jakirkham, Bussonnier, M. B., Moore, J., Orfanos, D. P., Fulton, A., et al. (2022). *zarr-developers/zarr-python: v2.13.3*. Zenodo. <https://doi.org/10.5281/zenodo.7174882>

Narenpitak, P., Kazil, J., Yamaguchi, T., Quinn, P. K., & Feingold, G. (2021). From sugar to flowers: A transition of shallow cumulus organization during ATOMIC. *Journal of Advances in Modeling Earth Systems*, 13(10), e2021MS002619. <https://doi.org/10.1029/2021ms002619>

Narenpitak, P., Kazil, J., Yamaguchi, T., Quinn, P. K., & Feingold, G. (2023). The sugar-to-flower shallow cumulus transition under the influences of diel cycle and free-tropospheric mineral dust. *Journal of Advances in Modeling Earth Systems*, 15(1), e2022MS003228. <https://doi.org/10.1029/2022MS003228>

Nitta, T., & Esbensen, S. (1974). Heat and moisture budget analyses using BOMEX data. *Monthly Weather Review*, 102(1), 17–28. [https://doi.org/10.1175/1520-0493\(1974\)102<0017:hambau>2.0.co;2](https://doi.org/10.1175/1520-0493(1974)102<0017:hambau>2.0.co;2)

Nuijens, L., & Siebesma, A. P. (2019). Boundary layer clouds and convection over subtropical oceans in our current and in a warmer climate. *Current Climate Change Reports*, 5(2), 80–94. <https://doi.org/10.1007/s40641-019-00126-x>

Nuijens, L., & Stevens, B. (2012). The influence of wind speed on shallow marine cumulus convection. *Journal of the Atmospheric Sciences*, 69(1), 168–184. <https://doi.org/10.1175/jas-d-11-02.1>

Ouwersloot, H., Moene, A., Attema, J., & De Arellano, J. V.-G. (2017). Large-eddy simulation comparison of neutral flow over a canopy: Sensitivities to physical and numerical conditions, and similarity to other representations. *Boundary-Layer Meteorology*, 162(1), 71–89. <https://doi.org/10.1007/s10546-016-0182-5>

Pedgley, D. (2003). Luke Howard and his clouds. *Weather*, 58(2), 51–55. <https://doi.org/10.1256/wea.157.02>

Pelczar, R. M., Pelczar, M. J., & Steere, W. C. (2022). Botany. Retrieved from <https://www.britannica.com/science/botany>

Pincus, R., Barker, H. W., & Morcrette, J.-J. (2003). A fast, flexible, approximate technique for computing radiative transfer in inhomogeneous cloud fields. *Journal of Geophysical Research*, 108(D13), 4376. <https://doi.org/10.1029/2002JD003322>

Radtke, J., Mauritsen, T., & Hohenegger, C. (2021). Shallow cumulus cloud feedback in large eddy simulations—bridging the gap to storm-resolving models. *Atmospheric Chemistry and Physics*, 21(5), 3275–3288. <https://doi.org/10.5194/acp-21-3275-2021>

Savazzi, A. C. M., Nuijens, L., de Rooy, W., & Siebesma, P. (2022). Unveiling convective momentum transport at different scales during EUREC⁴A (Tech. Rep.). Copernicus Meetings. <https://doi.org/10.5194/egusphere-egu22-8166>

Savazzi, A. C. M., Nuijens, L., Sandu, I., George, G., & Bechtold, P. (2022). The representation of the trade winds in ECMWF forecasts and reanalyses during EUREC⁴A. *Atmospheric Chemistry and Physics*, 22(19), 13049–13066. <https://doi.org/10.5194/acp-22-13049-2022>

Schalkwijk, J., Jonker, H. J., & Siebesma, A. P. (2013). Simple solutions to steady-state cumulus regimes in the convective boundary layer. *Journal of the Atmospheric Sciences*, 70(11), 3656–3672. <https://doi.org/10.1175/jas-d-12-0312.1>

Schulz, H. (2021). *Meso-scale patterns of shallow convection in the trades* (Doctoral dissertation). Universität Hamburg. <https://doi.org/10.17617/2.3357904>

Schulz, H., Eastman, R., & Stevens, B. (2021). Characterization and evolution of organized shallow convection in the downstream North Atlantic trades. *Journal of Geophysical Research: Atmospheres*, 126(17), e2021JD034575. <https://doi.org/10.1029/2021JD034575>

Schulz, H., & Stevens, B. (2023). On the representation of shallow convection in the trades by large-domain, hecto-meter, large-eddy simulations. *eartharxiv.org*. <https://doi.org/10.31223/X5H651>

Schulzweida, U. (2021). CDO user guide. <https://doi.org/10.5281/zenodo.5614769>

Seifert, A., & Beheng, K. D. (2001). A double-moment parameterization for simulating autoconversion, accretion and selfcollection. *Atmospheric Research*, 59, 265–281. [https://doi.org/10.1016/s0169-8095\(01\)00126-0](https://doi.org/10.1016/s0169-8095(01)00126-0)

Seifert, A., Heus, T., Pincus, R., & Stevens, B. (2015). Large-eddy simulation of the transient and near-equilibrium behavior of precipitating shallow convection. *Journal of Advances in Modeling Earth Systems*, 7(4), 1918–1937. <https://doi.org/10.1002/2015ms000489>

Shen, Z., Sridhar, A., Tan, Z., Jaruga, A., & Schneider, T. (2022). A library of large-eddy simulations forced by global climate models. *Journal of Advances in Modeling Earth Systems*, 14(3), e2021MS002631. <https://doi.org/10.1029/2021MS002631>

Siebesma, A. P., Bretherton, C. S., Brown, A., Chlond, A., Cuxart, J., Duynkerke, P. G., et al. (2003). A large eddy simulation intercomparison study of shallow cumulus convection. *Journal of the Atmospheric Sciences*, 60(10), 1201–1219. [https://doi.org/10.1175/1520-0469\(2003\)60<1201:alesis>2.0.co;2](https://doi.org/10.1175/1520-0469(2003)60<1201:alesis>2.0.co;2)

Stevens, B., Ackerman, A. S., Albrecht, B. A., Brown, A. R., Chlond, A., Cuxart, J., et al. (2001). Simulations of trade wind cumuli under a strong inversion. *Journal of the Atmospheric Sciences*, 58(14), 1870–1891. [https://doi.org/10.1175/1520-0469\(2001\)058<1870:sotwcu>2.0.co;2](https://doi.org/10.1175/1520-0469(2001)058<1870:sotwcu>2.0.co;2)

Stevens, B., Bony, S., Brogniez, H., Hentgen, L., Hohenegger, C., Kiemele, C., et al. (2020). Sugar, gravel, fish and flowers: Mesoscale cloud patterns in the trade winds. *Quarterly Journal of the Royal Meteorological Society*, 146(726), 141–152. <https://doi.org/10.1002/qj.3662>

Stevens, B., Bony, S., Farrell, D., Ament, F., Blyth, A., Fairall, C., et al. (2021). EUREC⁴A. *Earth System Science Data*, 13(8), 4067–4119. <https://doi.org/10.5194/esd-13-4067-2021>

Stevens, B., Satoh, M., Auger, L., Biercamp, J., Bretherton, C. S., Chen, X., et al. (2019). DYAMOND: The DYnamics of the Atmospheric general circulation Modeled On Non-hydrostatic Domains. *Progress in Earth and Planetary Science*, 6(1), 1–17. <https://doi.org/10.1186/s40645-019-0304-z>

vanZanten, M. C., Stevens, B., Nuijens, L., Siebesma, A. P., Ackerman, A. S., Burnet, F., et al. (2011). Controls on precipitation and cloudiness in simulations of trade-wind cumulus as observed during RICO. *Journal of Advances in Modeling Earth Systems*, 3(2), M06001. <https://doi.org/10.1029/2011MS000056>

Vial, J., Albright, A. L., Vogel, R., Musat, I., & Bony, S. (2023). Cloud transition across the daily cycle illuminates model responses of trade cumuli to warming. *Proceedings of the National Academy of Sciences*, 120(8), e2209805120. <https://doi.org/10.1073/pnas.2209805120>

Vial, J., Vogel, R., Bony, S., Stevens, B., Winker, D. M., Cai, X., et al. (2019). A new look at the daily cycle of trade wind cumuli. *Journal of Advances in Modeling Earth Systems*, 11(10), 3148–3166. <https://doi.org/10.1029/2019ms001746>

Vial, J., Vogel, R., & Schulz, H. (2021). On the daily cycle of mesoscale cloud organization in the winter trades. *Quarterly Journal of the Royal Meteorological Society*, 147(738), 2850–2873. <https://doi.org/10.1002/qj.4103>

Vogel, R., Nuijens, L., & Stevens, B. (2020). Influence of deepening and mesoscale organization of shallow convection on stratiform cloudiness in the downstream trades. *Quarterly Journal of the Royal Meteorological Society*, 146(726), 174–185. <https://doi.org/10.1002/qj.3664>

Wicker, L. J., & Skamarock, W. C. (2002). Time-splitting methods for elastic models using forward time schemes. *Monthly Weather Review*, 130(8), 2088–2097. [https://doi.org/10.1175/1520-0493\(2002\)130<2088:tsmfem>2.0.co;2](https://doi.org/10.1175/1520-0493(2002)130<2088:tsmfem>2.0.co;2)

Yamaguchi, T., Feingold, G., & Kazil, J. (2019). Aerosol-cloud interactions in trade wind cumulus clouds and the role of vertical wind shear. *Journal of Geophysical Research: Atmospheres*, 124(22), 12244–12261. <https://doi.org/10.1029/2019jd031073>

Zhang, M., Bretherton, C. S., Blossey, P. N., Austin, P. H., Bacmeister, J. T., Bony, S., et al. (2013). 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. <https://doi.org/10.1002/2013MS000246>

Zhang, Y., Stevens, B., & Ghil, M. (2005). On the diurnal cycle and susceptibility to aerosol concentration in a stratocumulus-topped mixed layer. *Quarterly Journal of the Royal Meteorological Society*, 131(608), 1567–1583. <https://doi.org/10.1256/qj.04.103>