Cloud Botany: Shallow cumulus clouds in an ensemble of idealized large-domain large-eddy simulations of the trades

Fredrik Jansson[1], Martin Janssens[1,2], Johanna H. Grönqvist[3,4], A. Pier Siebesma[1,5], Franziska Glassmeier[1], Jisk Attema[6], Victor Azizi[6], Masaki Satoh[7], Yousuke Sato[8,9], Hauke Schulz[10,11], Tobias Kölling[10]

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

Key Points:

- We present Cloud Botany, an ensemble of idealized LES 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.

Corresponding author: Fredrik Jansson, `fjansson@abo.fi`

Abstract

Small shallow cumulus clouds (<1 km) over the tropical oceans appear to possess the ability to self-organise 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 Large Eddy Simulation (DALES) 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 EUREC^{4}A (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 EUREC^{4}A 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 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. Often this leads to rainfall, which then causes further heterogeneity. 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 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., 2003a), 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, e.g. 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 EUREC^{4}A field campaign (Stevens et al., 2021). Fortunately, advances in computational capabilities now allow these observations to be complemented by i) global and regional models running at a sufficiently fine resolution to begin resolving shallow convection (e.g., Stevens et al., 2019) and ii) detailed process models - "classical" LES codes - running on sufficiently large domains to capture the mesoscale (e.g., Seifert et al., 2015; Lamaakel & Matheou, 2022). 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 kilometre resolution are beginning to appear (Schulz & Stevens, 2023; Schulz, 2021); 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 100m.

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 trades (e.g., Bellon & Stevens, 2012; Nuijens & Stevens, 2012; Schalkwijk et al., 2013; Feingold et al., 2016; Glassmeier et al., 2021; 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., Seifert et al., 2015; Bretherton & Blossey, 2017; Janssens et al., 2022) 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 east–west (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 i) geostrophic horizontal wind (u_g, v_g), ii) a large scale vertical velocity (w_{ls}) and iii) 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

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

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]	2500	scale height of background subsidence
h_{w_1} [m]	5300	scale height of first additional mode of imposed vertical velocity
$\partial_t\theta_{l,ls,0}$ [K/day]	-0.5	large scale temperature tendency in first model level
$\partial_t\theta_{l,ls,z}$ [K/day/m]	$2.5{\cdot}10^{-4}$	large scale temperature tendency slope
$\partial_t q_{t,ls,0}$ [g/kg/day]	-1.49	large scale humidity tendency at surface
$\partial_t q_{t,ls,z}$ [g/kg/day/m]	$3.73{\cdot}10^{-4}$	large scale humidity tendency slope
τ_∞ [h]	6	nudging time scale at top of domain
z_{max} [m]	3000	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

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 initialised 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, i.e. the geostrophic wind is predominantly east–west.

Profiles of the initial liquid water potential temperature θ_l follow a similar, linear approximation. However, they are slightly modified to account for their lowest levels covarying with the surface conditions (Pearson correlation $r = 0.57$ 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 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, θ_{l0}, we simply set the residual layer's value to 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:

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

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 modelled with a similar initial well-mixed layer, but drop off exponentially above z_{ml}, following Vogel et al. (2020):

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

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: i) a term representing the downwelling branch of the Hadley cell, modelled by exponential decay with height following e.g. Bellon and Stevens (2012), and ii) a sinusoidal term, a single period of which represents mesoscale circulations, as frequently observed during EUREC^{4}A (George et al., 2022):

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

Varying w_1 captures a substantial amount of the mesoscale variability in vertical velocity in the trades (George et al., 2022). Therefore, we fix the free-tropospheric, asymptotic subsidence w_∞ and its scale height h_{w_∞}. Furthermore, we assume i) 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; Narenpitak et al., 2021; Janssens et al., 2022), and ii) 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_{l0}, 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 (Hersbach et al., 2020), sampled every 3 hours between 9.8–16.8 N and 62.22–54.22 W between Jan-01-2020 and Mar-31-2020. This domain and period are representative for the trades in general (Medeiros & Nuijens, 2016) and span the winter during which the EUREC^{4}A campaign was conducted (Jan-20-2020 to Feb-20-2020). We complement use of ERA5 with the JOANNE 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 central rows of Table 2. Quality of fit is assessed in terms of the signal-to-noise ratio of each parameter, averaged over all fits, with the noise taken as the mean standard error of the least squares fit. Based on these numbers, we subjectively judge the fits of θ_l to be excellent, those of q_t and u adequate, and those of w_1 inadequate. The poor fit of w_1 reflects both significant deviations from

Table 2. Properties of environmental control parameters, fit to the ERA-5 database, and selected for Cloud Botany. Mean SNR denotes the signal to noise ratio averaged over all fits; 10–90% refers to the value of the 10th and 90th percentile of each parameter over the fits. The range over which the parameters in Cloud Botany are varied is reported in the table's bottom row. Temp. abbreviates temperature; hum. stands for humidity, and ML for mixed layer.

	θ_{l0} [K] surf. temp.	Γ [K/km] temp. lapse rate	$q_{t,ml}$ [g/kg] ML hum.	h_{q_t} [m] hum. scale	u_0 [m/s] surf. wind	w_1 [cm/s] large-scale w
Mean	299	5	14.1	1810	-10.6	0.0393
Mean st. error	0.432	0.147	0.553	175	0.782	0.331
Mean SNR	821	40.1	28.4	11.7	16.0	0.034
10–90%	298 – 300	0.00454 – 0.00528	12.8 – 15.4	1180 – 2510	-14.2 – -6.93	-0.984 – 1.14
Selected range	297.5 – 299.5	0.0045 – 0.0055	13.5 – 15.0	1200 – 2500	-15 – -5	-0.350 – 0.180

the prescribed functional form, and variability in higher-order modes in the ERA-5 database than our simple approximation captures. Since ERA5 agrees well with the JOANNE data (George et al., 2022), these higher-frequency fluctuations are unlikely entirely 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, i.e. $2^6 = 64$ simulations. These stem from considering all possible combinations of our environmental control parameters, at a minimum and a maximum point informed by the 10^{th} and 90^{th} 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": Runs that span the range between the extrema in several steps for each control parameter, with all other parameters held at the center of the hypercube.

Since the chosen parameters will be varied independently of each other, it is prudent to quantify their independence in observations, i.e. 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 (0.340), θ_{l0} and $q_{t,ml}$ (0.353), Γ and h_{q_t}(0.356), and $q_{t,ml}$ and h_{q_t} (0.396). 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 results from rounding the 10^{th} and 90^{th} percentile values. Variability in θ_{l0} subsumes both variability in surface pressure and Sea-Surface Temperature (SST). Since we keep the surface pressure over our ensemble fixed at 1016.05 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 0.5K. 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 modelling 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 i) capture sufficient w variability to satisfy our main objective – studying environmental dependencies – and ii) ensure that the variability we capture is more representative of the reanalysis than traditional exponential (Bellon & Stevens, 2012; Blossey et al., 2013) or linear (Stevens et al., 2001; Siebesma et al., 2003a; Yamaguchi et al., 2019) approximations. Therefore, we set w_∞ to a number characteristic of the ERA5 mean in 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 EUREC^{4}A (George et al., 2022). We separate the vertical velocity variability by moisture variability (and not by the vertical velocity itself), since the moisture variability tends to co-vary with the degree to which vertical velocity patterns lead to aggregated cloud structures (Bretherton & Blossey, 2017; George et al., 2022), and we are in search of such variability in the cloudiness. The resulting fits are shown in Figure 1.

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, and two larger-scale advective forcings. These are all kept the same for all simulations; we briefly describe them in turn below.

Our simulations run for 60 hours 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 1536 grid points on a horizontal side, and 175 vertical grid levels. Advection of momentum, θ_l and e is discretized with a sixth order scheme, advection of q_t and precipitation species with a fifth order scheme (Wicker & Skamarock, 2002). The sources and sinks of precipitation are modelled 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). As the importance of diurnal, radiative variability in the downstream trades has recently been emphasized (Vial et al., 2019, 2021; Albright et al., 2021), we include in

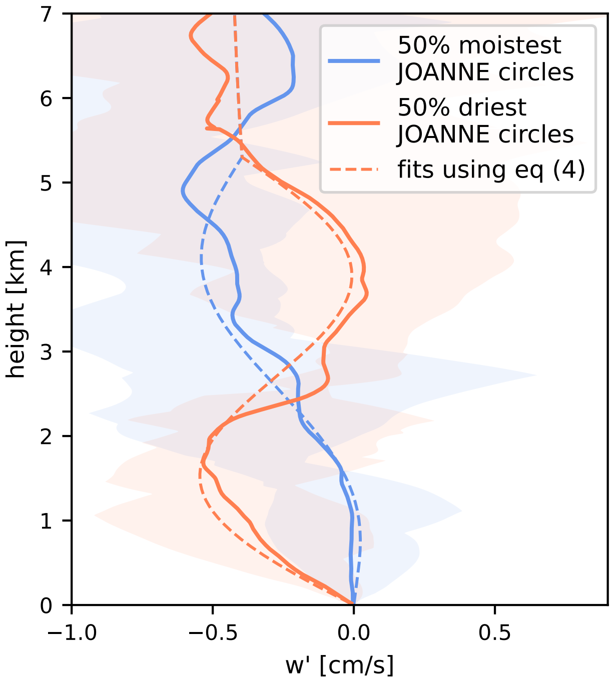

Figure 1. Envelopes and mean of the vertical velocity in the JOANNE data set (George et al., 2021), over the 50% moistest (blue) and driest (red) circles flown by the HALO aircraft during the EUREC^{4}A field campaign. Dashed lines indicate the profiles constructed with Equation 4 and the parameters reported in Tables 2 and 1.

the model's shortwave component the diurnal cycle representative for Feb-01-2020 at 13.1 N and 52 W. Required input profiles for ozone, water vapor and temperature derive from ERA5, averaged over the EUREC^{4}A region and period. These are prescribed over the entire modelled column for ozone, and 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 are adopted for all other trace gases.

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):

$$\partial_t \theta_{l,ls} = \min(0, \quad \partial_t \theta_{l,ls,0} + \partial_t \theta_{l,ls,z} z) \tag{5}$$
$$\partial_t q_{t,ls} = \min(0, \quad \partial_t q_{t,ls,0} + \partial_t q_{t,ls,z} z) \tag{6}$$

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 ten-

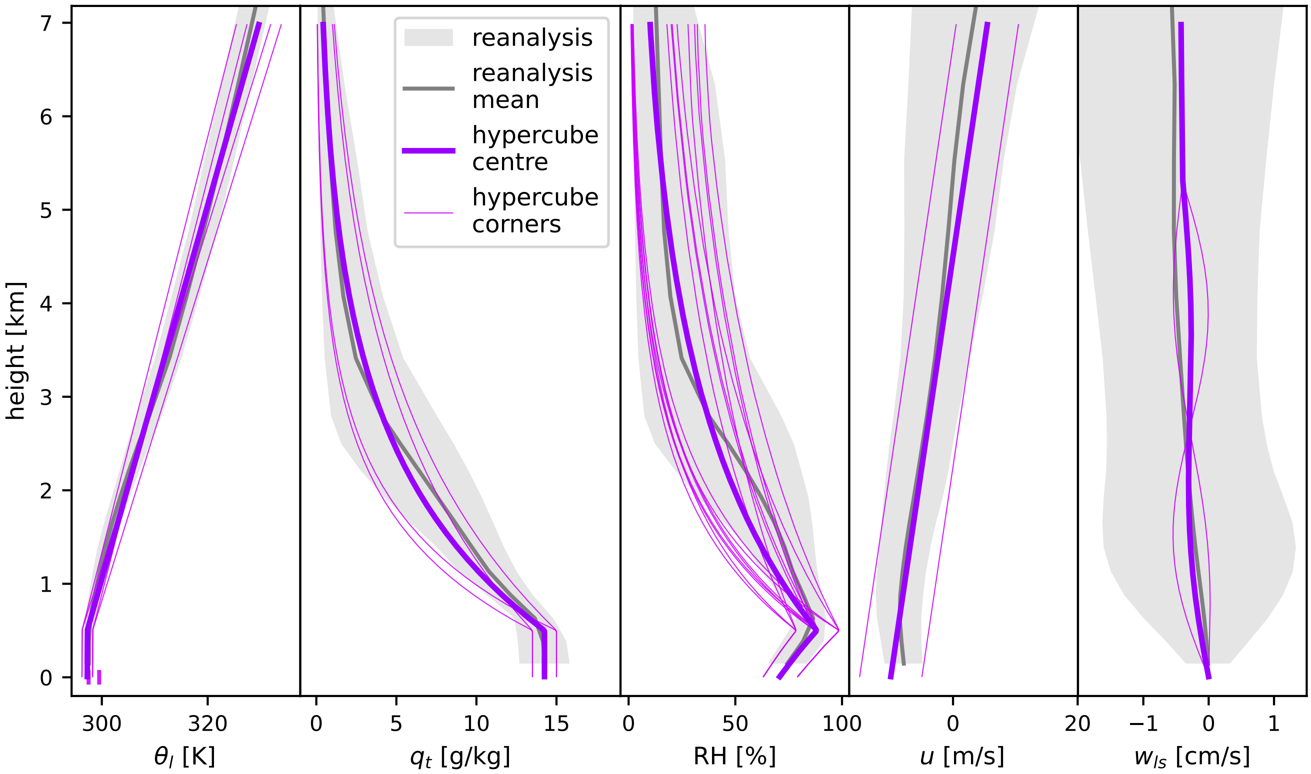

Figure 2. Profiles of θ_l, q_t, relative humidity (RH), u and w_{ls} over the 10^{th}–90^{th} percentile envelope in ERA-5 (reanalysis, shading), its mean (gray), and for initial and large-scale forcing of Cloud Botany simulations: the centre (purple) and corners (pink) of the six-dimensional hypercube.

dency on our prognostic variables (u, v, θ_l, q_t) that forces them back towards their initial state with a height-dependent nudging time scale τ:

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

In this relation, the inverse tangent is centered around the top of the cloud layer: $z_{\max} =$ 3000 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 profile return to $\tau_\infty = 6$ h. The fixed parameters of Equations 5 and 7 are listed in Table 1.

4 Workflow to create the data set

To turn the LES ensemble design into accessible data products, four steps need to be taken: i) creating a set of input files for each ensemble member, ii) running each simulation, iii) converting simulation output to an easily accessible format and iv) 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.

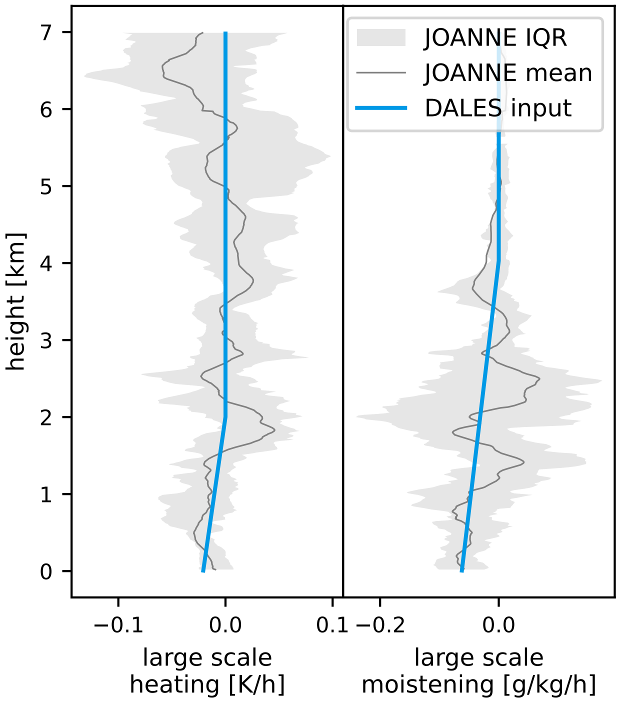

Figure 3. Inter-quartile range (IQR, shading) and mean (gray line) of JOANNE-derived tendencies of heating and moistening, and idealized fit used to force Cloud Botany simulations (blue line).

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, i.e. it does not rely on accelerators such as GPUs. 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 Section Appendix A for details of the DALES version used and for accessing the code.

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	Nodes	Cores	Wall clock time	Time per grid point and time step
Fugaku	0–12h	24	1152	46760 s	2.8 μs
	36–48h	24	1152	110391 s	5.6 μs
Snellius	0–12h	8	1024	73717 s	4.0 μs
	36–48h	8	1024	92094 s	4.5 μs

DALES is parallelized using Message Passing Interfaces (MPI) in x and y, the two horizontal directions. Each simulation was run on 24 nodes, with 24 $\times$ 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, 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 Centre (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 Centre 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 (vanZanten et al., 2011) and BOMEX (Siebesma et al., 2003b) 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 40TB of simulation output is a burden and most users will only access portions of this data set, e.g. 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 fileformat allowed further to host 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 filesystem based hosted data sets and make the Cloud Botany data set analysis ready.

An analysis in Python can be started by accessing the EUREC^{4}A 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 analyse 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 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 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 A4 – A11. 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 minutes.

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 exploration of the development of mesoscale cloud patterns in the Cloud Botany ensemble. We begin with Figure 4, which shows the evolution of several quantities of interest and snapshots of the cloud cover and precipitation in simulation 1, the centre point of our parameter space. Its evolution is qualitatively similar to that of many ensemble members. All simulations depart from cloud-free states at midnight UTC. The first 10 hours 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, this gradually heats the domain, reducing both the cloud fraction and horizontally averaged liquid-water path (LWP). After sunset, the cloud structures rapidly grow vertically and begin to precipitate around 24 hours 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.

Figure 5 shows a few examples of cloud patterns that develop under different parameter combinations. Many of these develop precipitating convection, almost always paired with cold pools. When they appear, such cold pools typically dominate the cloud patterning. We find at least three different ways in which this happens. First, cold pools lined by arcs of cumuli (e.g. Figure 5 a) are ubiquitous across our precipitating simulations. 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 Figure 5 b and c and 79, not shown), cold pools are produced by sufficiently vigorous convective cells that they produce large (> 50 km) sheets of thin, stratiform outflow layers, reminiscent of the structures termed "Flowers" by Stevens et al. (2020). At a glance, the ap-

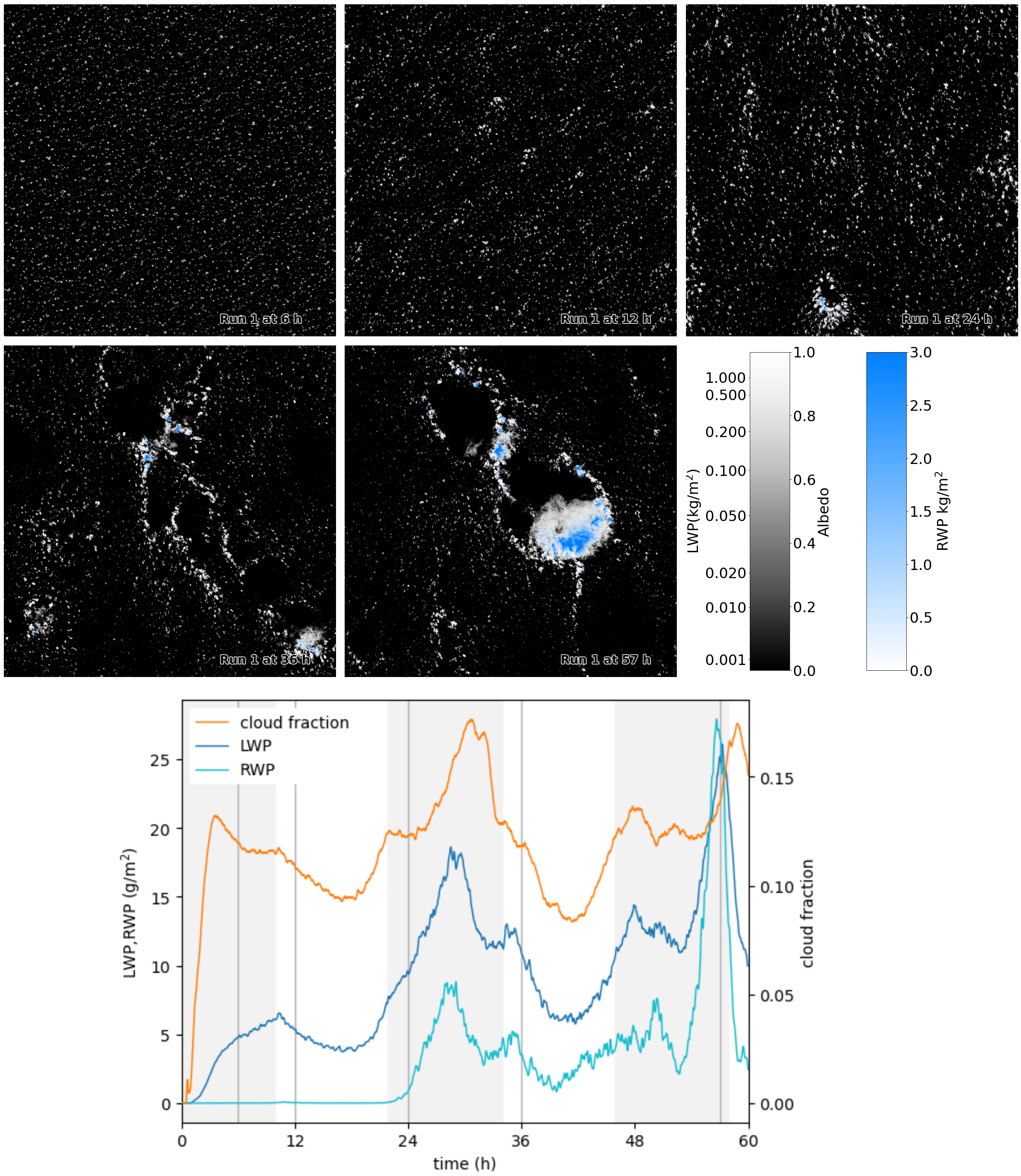

Figure 4. Time series of simulation 1, the central point of the parameter hypercube. The snapshots show cloud albedo in white (as parameterized by Zhang et al. (2005)) and rain water path in blue. The time series curves show the liquid water path (LWP), rain water path (RWP) and cloud fraction over time. The shaded background shows the diurnal cycle, the darker regions are night (18h to 06h in local time). The times of the snapshots are indicated by gray vertical lines.

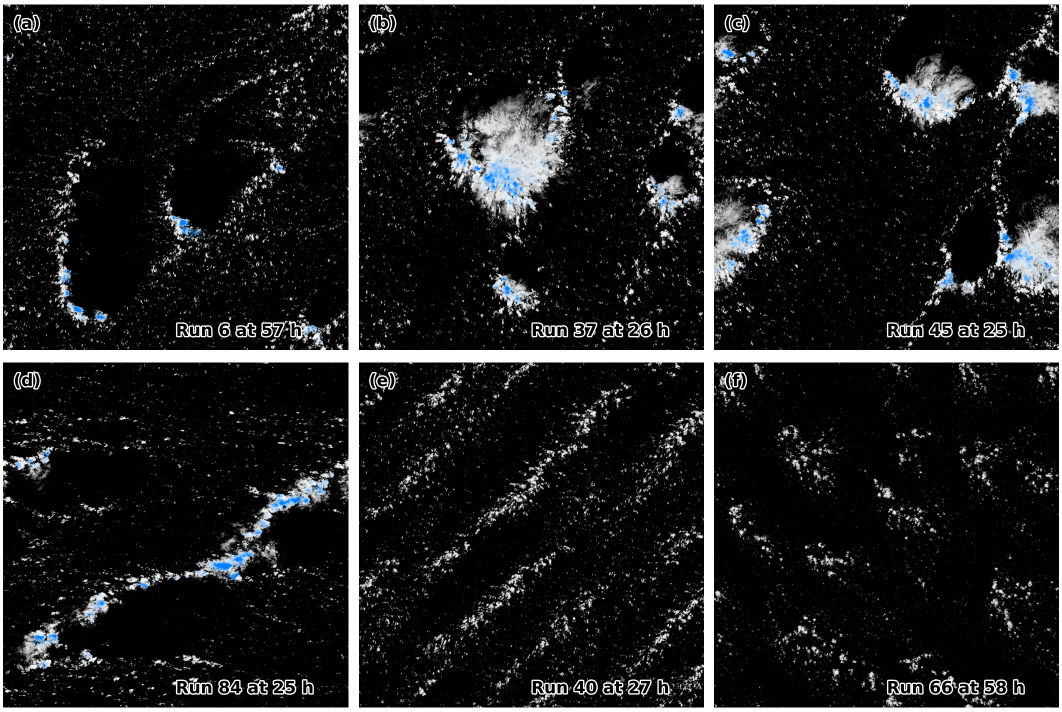

Figure 5. Different types of cloud organization seen in the Cloud Botany ensemble. a) Run 6, cold pools, b) run 37, large cloud cluster topped by stratiform outflow, c) run 45, multiple such clusters, on the edges of cold pools, d) run 84, line of precipitation, e) run 40, non-precipitating cumulus in bands and f) run 66, non-precipitating cumulus in aggregated in quasi-circular clusters. The wind is easterly, i.e. from the right side of the image.

pearance of such structures under stronger stratifications and higher surface winds appears consistent with the observations by Bony et al. (2020); Schulz et al. (2021). 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 5 d); 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, 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 in the simulations (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 5 e) or into quasi-circular clusters (Figure 5 f). Figure 6 shows 3D renderings of parts of scenes b and e of figure 5.

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 lapse rate, sometimes allowing single plumes 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. Addi-

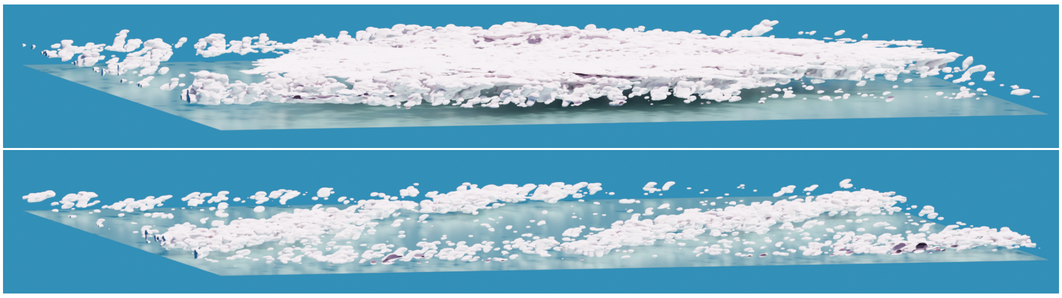

Figure 6. Rendered 3D view of the central, large cloud structure in figure 5 b and of the stripes in figure 5 e, above a reflecting plane representing the ocean. The rendered domain is 70×70 km. The rendering shows an isosurface of $q_l = 2 \cdot 10^{-5}$ kg/kg.

tionally, runs 11, 14, 43, 46 and 87 only span 48h due to their computing jobs being interrupted.

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 i) no clouds, ii) small, randomly spaced cumulus, iii) clustered, non-precipitating cumulus (both ii and iii seem to fit the "Sugar" category), iv) 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 dataset therefore cannot shed light on their formation. However, since Schulz 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 characterise the trades.

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 i) all possible combinations of high and low values in the parameters that are representative for observed variability over the boreal winter of 2020, and ii) 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 cumulus, which quickly begin self-aggregating into mesoscale clusters. After a marked diurnal cycle, we often observe the onset of precipitation after around 24 hours 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. 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 et al., 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 Tables

Acknowledgments

The research was supported by the European Union's Horizon 2020 research and innovation program under grant agreement no. 820829 (CONSTRAIN project). FG acknowledges support from The Branco Weiss Fellowship - Society in Science, administered by ETH Zürich, and from an NWO Veni grant. DALES performance improvements were made with support from the ESIWACE2 project, which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823988.

Access to supercomputer Fugaku was provided by RIKEN through the HPCI System Research Project (project hp200321). Furthermore, we acknowledge the use of ECMWF's computing and archive facilities in the research reported here. We thank the staff of the Fugaku helpdesk at RIKEN for their friendly and fast assistance and DKRZ, the German Climate Computing Centre, for hosting the dataset. We thank SURF (www.surf.nl) for the support in using the National Supercomputer Snellius.

Code and data availability

The main data product in this article is the Cloud Botany ensemble, accessible through the EUREC^{4}A intake catalog, see `https://howto.eurec4a.eu/intro.html`.

The version of DALES used for the Cloud Botany experiment is based on DALES v4.3 DOI: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 DOI: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 DOI:10.5281/zenodo.7709435

Table A1. Summary of all simulations in the Cloud Botany ensemble, part 1

run	θ_{l0} [K]	u_0 [m/s]	$q_{t,ml}$ [g/kg]	h_{q_t} [m]	Γ [K/m]	w_1 [cm/s]	u_z [s^{-1}]	location	remark
1	298.5	-10	14.25	1850	5	-0.085	0.0022	center	cold pools
2	297.5	-15	13.5	1200	4.5	-0.35	0.0022	corner	cold pools
3	297.5	-15	13.5	1200	4.5	0.18	0.0022	corner	cold pools
4	297.5	-15	13.5	1200	5.5	-0.35	0.0022	corner	cold pools
5	297.5	-15	13.5	1200	5.5	0.18	0.0022	corner	cold pools
6	297.5	-15	13.5	2500	4.5	-0.35	0.0022	corner	cold pools
7	297.5	-15	13.5	2500	4.5	0.18	0.0022	corner	thermo. crash
8	297.5	-15	13.5	2500	5.5	-0.35	0.0022	corner	weak precip
9	297.5	-15	13.5	2500	5.5	0.18	0.0022	corner	cold pools
10	297.5	-15	15	1200	4.5	-0.35	0.0022	corner	cold pools
11	297.5	-15	15	1200	4.5	0.18	0.0022	corner	cold pools
12	297.5	-15	15	1200	5.5	-0.35	0.0022	corner	cold pools
13	297.5	-15	15	1200	5.5	0.18	0.0022	corner	cold pools
14	297.5	-15	15	2500	4.5	-0.35	0.0022	corner	cold pools
15	297.5	-15	15	2500	4.5	0.18	0.0022	corner	thermo. crash
16	297.5	-15	15	2500	5.5	-0.35	0.0022	corner	cold pools
17	297.5	-15	15	2500	5.5	0.18	0.0022	corner	cold pools
18	297.5	-5	13.5	1200	4.5	-0.35	0.0022	corner	no clouds
19	297.5	-5	13.5	1200	4.5	0.18	0.0022	corner	cold pools
20	297.5	-5	13.5	1200	5.5	-0.35	0.0022	corner	no clouds
21	297.5	-5	13.5	1200	5.5	0.18	0.0022	corner	cold pools
22	297.5	-5	13.5	2500	4.5	-0.35	0.0022	corner	small cumulus
23	297.5	-5	13.5	2500	4.5	0.18	0.0022	corner	cold pools
24	297.5	-5	13.5	2500	5.5	-0.35	0.0022	corner	no clouds
25	297.5	-5	13.5	2500	5.5	0.18	0.0022	corner	cold pools
26	297.5	-5	15	1200	4.5	-0.35	0.0022	corner	small cumulus
27	297.5	-5	15	1200	4.5	0.18	0.0022	corner	cold pools
28	297.5	-5	15	1200	5.5	-0.35	0.0022	corner	small cumulus
29	297.5	-5	15	1200	5.5	0.18	0.0022	corner	cold pools
30	297.5	-5	15	2500	4.5	-0.35	0.0022	corner	organizing cumulus
31	297.5	-5	15	2500	4.5	0.18	0.0022	corner	cold pools
32	297.5	-5	15	2500	5.5	-0.35	0.0022	corner	small organizing cumulus
33	297.5	-5	15	2500	5.5	0.18	0.0022	corner	cold pools
34	299.5	-15	13.5	1200	4.5	-0.35	0.0022	corner	cold pools
35	299.5	-15	13.5	1200	4.5	0.18	0.0022	corner	cold pools

Table A2. Summary of all simulations in the Cloud Botany ensemble, part 2

run	θ_{l0} [K]	u_0 [m/s]	$q_{t,ml}$ [g/kg]	h_{q_t} [m]	Γ [K/m]	w_1 [cm/s]	u_z [s^{-1}]	location	remark
36	299.5	-15	13.5	1200	5.5	-0.35	0.0022	corner	cold pools
37	299.5	-15	13.5	1200	5.5	0.18	0.0022	corner	cold pools, aggregated clouds
38	299.5	-15	13.5	2500	4.5	-0.35	0.0022	corner	cold pools, thermo. crash
39	299.5	-15	13.5	2500	4.5	0.18	0.0022	corner	cold pools, thermo. crash
40	299.5	-15	13.5	2500	5.5	-0.35	0.0022	corner	arcs
41	299.5	-15	13.5	2500	5.5	0.18	0.0022	corner	cold pools
42	299.5	-15	15	1200	4.5	-0.35	0.0022	corner	cold pools
43	299.5	-15	15	1200	4.5	0.18	0.0022	corner	cold pools
44	299.5	-15	15	1200	5.5	-0.35	0.0022	corner	cold pools
45	299.5	-15	15	1200	5.5	0.18	0.0022	corner	cold pools, aggregated clouds
46	299.5	-15	15	2500	4.5	-0.35	0.0022	corner	cold pools
47	299.5	-15	15	2500	4.5	0.18	0.0022	corner	thermo. crash
48	299.5	-15	15	2500	5.5	-0.35	0.0022	corner	cold pools
49	299.5	-15	15	2500	5.5	0.18	0.0022	corner	cold pools
50	299.5	-5	13.5	1200	4.5	-0.35	0.0022	corner	no clouds
51	299.5	-5	13.5	1200	4.5	0.18	0.0022	corner	cold pools
52	299.5	-5	13.5	1200	5.5	-0.35	0.0022	corner	no clouds
53	299.5	-5	13.5	1200	5.5	0.18	0.0022	corner	cold pools
54	299.5	-5	13.5	2500	4.5	-0.35	0.0022	corner	no clouds
55	299.5	-5	13.5	2500	4.5	0.18	0.0022	corner	cold pools
56	299.5	-5	13.5	2500	5.5	-0.35	0.0022	corner	no clouds
57	299.5	-5	13.5	2500	5.5	0.18	0.0022	corner	cold pools
58	299.5	-5	15	1200	4.5	-0.35	0.0022	corner	no clouds
59	299.5	-5	15	1200	4.5	0.18	0.0022	corner	cold pools
60	299.5	-5	15	1200	5.5	-0.35	0.0022	corner	no clouds
61	299.5	-5	15	1200	5.5	0.18	0.0022	corner	cold pools
62	299.5	-5	15	2500	4.5	-0.35	0.0022	corner	small cumulus
63	299.5	-5	15	2500	4.5	0.18	0.0022	corner	cold pools
64	299.5	-5	15	2500	5.5	-0.35	0.0022	corner	no clouds
65	299.5	-5	15	2500	5.5	0.18	0.0022	corner	cold pools
66	298.5	-4	14.25	1850	5	-0.085	0.0022	sweep u0	organizing cumulus, weak precip
67	298.5	-5	14.25	1850	5	-0.085	0.0022	sweep u0	organizing cumulus, weak precip
68	298.5	-6	14.25	1850	5	-0.085	0.0022	sweep u0	organizing cumulus, cold pools
69	298.5	-8	14.25	1850	5	-0.085	0.0022	sweep u0	cold pools
70	298.5	-12	14.25	1850	5	-0.085	0.0022	sweep u0	cold pools

Table A3. Summary of all simulations in the Cloud Botany ensemble, part 3

run	θ_{l0} [K]	u_0 [m/s]	$q_{t,ml}$ [g/kg]	h_{q_t} [m]	Γ [K/m]	w_1 [cm/s]	u_z [s^{-1}]	location	remark
71	298.5	-15	14.25	1850	5	-0.085	0.0022	sweep u0	cold pools
72	298.5	-10	14.25	1850	5	-0.2	0.0022	sweep wpamp	cold pools
73	298.5	-10	14.25	1850	5	-0.1	0.0022	sweep wpamp	cold pools
74	298.5	-10	14.25	1850	5	0	0.0022	sweep wpamp	cold pools
75	298.5	-10	14.25	1850	5	0.1	0.0022	sweep wpamp	cold pools
76	298.5	-10	14.25	1850	4	-0.085	0.0022	sweep Γ	cold pools
77	298.5	-10	14.25	1850	4.5	-0.085	0.0022	sweep Γ	cold pools
78	298.5	-10	14.25	1850	4.75	-0.085	0.0022	sweep Γ	cold pools
79	298.5	-10	14.25	1850	5.25	-0.085	0.0022	sweep Γ	cold pools, aggregated clouds
80	298.5	-10	14.25	1850	5.5	-0.085	0.0022	sweep Γ	cold pools
81	298.5	-10	14.25	1850	6	-0.085	0.0022	sweep Γ	cold pools
82	298.5	-10	14.25	1850	6.5	-0.085	0.0022	sweep Γ	cold pools
83	298.5	-10	14.25	1850	7.5	-0.085	0.0022	sweep Γ	cold pools
84	298.5	-10	14.25	1850	5	-0.085	-0.0044	sweep u_z	precip and bands
85	298.5	-10	14.25	1850	5	-0.085	-0.0033	sweep u_z	precip and bands
86	298.5	-10	14.25	1850	5	-0.085	-0.0022	sweep u_z	bands and arcs
87	298.5	-10	14.25	1850	5	-0.085	-0.0011	sweep u_z	cold pools
88	298.5	-10	14.25	1850	5	-0.085	0	sweep u_z	cold pools
89	298.5	-10	14.25	1850	5	-0.085	0.0011	sweep u_z	cold pools
90	298.5	-10	14.25	1850	5	-0.085	0.0033	sweep u_z	cold pools
91	298.5	-10	14.25	1850	5	-0.085	0.0044	sweep u_z	arcs, bands
92	297.5	-10	14.25	1850	5	-0.085	0.0022	sweep thls	cold pools
93	299.5	-10	14.25	1850	5	-0.085	0.0022	sweep thls	cold pools
94	300.5	-10	14.25	1850	5	-0.085	0.0022	sweep thls	cold pools
95	301.5	-10	14.25	1850	5	-0.085	0.0022	sweep thls	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	1200	5	-0.085	0.0022	sweep h_{q_t}	cold pools
98	298.5	-10	14.25	1500	5	-0.085	0.0022	sweep h_{q_t}	cold pools
99	298.5	-10	14.25	2200	5	-0.085	0.0022	sweep h_{q_t}	cold pools
100	298.5	-10	14.25	2500	5	-0.085	0.0022	sweep h_{q_t}	cold pools, arcs
101	298.5	-10	14.25	3000	5	-0.085	0.0022	sweep h_{q_t}	cold pools
102	298.5	-10	13.5	1850	5	-0.085	0.0022	sweep qt0	cold pools
103	298.5	-10	15	1850	5	-0.085	0.0022	sweep qt0	cold pools

Table A4. Variables in the timeseries data set, sampled every minute. Dimensions: (member, time)

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
vtke	kg/s	Vertical integral of e
we	m/s	Entrainment velocity
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

(Jansson, Janssens, & Grönqvist, 2023). The scripts generating figures 4 and 5 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 DOI:10.5281/zenodo.7692270 (Jansson, Janssens, Grönqvist, Siebesma, et al., 2023).

Author Contribution

FJ, JHG, MJ, PS, FG, YS, MS formulated the project. VA, JA, FJ ported and optimized DALES for Fugaku and implemented single-precision floating point support. MJ analyzed the ERA5 data to obtain parameter ranges. MS and YS provided early access to Fugaku and the chance to test DALES there in advance of the project. FJ ran the simulations. HS, TK, FJ prepared the dataset for online access. FJ, MJ, and JHG wrote the article text, in collaboration with all the authors.

References

Agee, E. M. (1984). Observations from space and thermal convection: A historical perspective. *Bulletin of the American Meteorological Society*, *65*(9), 938 - 949. doi: 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.

Arabas, S., Axelsen, S., Attema, J., Beets, C., Boeing, S. J., de Bruine, M., ...

Table A5. Variables in the profiles data set containing horizontally averaged profiles, sampled every 5 minutes, part 1. Dimensions: (member, time, z)

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 qt variance
qt2Pr	$kg^2/kg^2/s$	Resolved production of qt variance
qt2Ps	$kg^2/kg^2/s$	SFS production of qt variance
qt2Res	$kg^2/kg^2/s$	Residual of qt budget
qt2S	$kg^2/kg^2/s$	Source of qt variance
qt2Tr	$kg^2/kg^2/s$	Resolved transport of qt variance
qt2r	$(kg/kg)^2$	Resolved total water variance
qt2tendf	$kg^2/kg^2/s$	Tendency of qt variance

Table A6. Variables in the profiles data set containing horizontally averaged profiles, sampled every 5 minutes, part 2. Dimensions: (member, time, z)

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
rhobh	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
sv0012r	$(kg/kg)^2$	Resolved scalar 001 variance
sv002	(kg/kg)	Scalar 002 specific mixing ratio
sv0022r	$(kg/kg)^2$	Resolved scalar 002 variance
svp001	(kg/kg/s)	Scalar 001 tendency
svp002	(kg/kg/s)	Scalar 002 tendency
svpt001	(kg/kg/s)	Scalar 001 turbulence tendency
svpt002	(kg/kg/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	Large scale radiative tendency

Table A7. Variables in the profiles data set containing horizontally averaged profiles, sampled every 5 minutes, part 3. Dimensions: (member, time, z)

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 (vw)
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
wsv001t	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
wsv002t	kg/kg m/s	Total scalar 002 flux
wthlr	Km/s	Resolved Theta_l flux
wthls	Km/s	SFS-Theta_l flux
wthlt	Km/s	Total Theta_l flux
wthvr	Km/s	Resolved buoyancy flux
wthvs	Km/s	SFS-buoyancy flux
wthvt	Km/s	Total buoyancy flux

Table A8. Variables in the 2D data set, containing horizontal fields sampled every 5 minutes. 2D. Dimensions: (member, time, y, x)

variable	units	description
cldtop	m	xy cross sections cloud top height
hinvsrf	m	height of surface inversion
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

Table A9. Variables in the 3D data set, the full 3D fields of the model sampled every hour. Dimensions: (member, time, z, y, x)

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

Table A10. Variables in the cross_xy data set, horizontal cross-sections of the prognostic variables sampled every 5 minutes. Dimensions: (z, y, x)

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

van Zanten, M. (2021, March). *dalesteam/dales: Dales 4.3.* Zenodo. doi: 10.5281/zenodo.4604726

Arabas, S., Axelsen, S., Attema, J., Beets, C., Boeing, S. J., de Bruine, M., … van Zanten, M. (2022, December). *dalesteam/dales: v4.3 for cloud botany on fugaku.* Zenodo. doi: 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.

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.

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*.

Blossey, P. N., Bretherton, C. S., Zhang, M., Cheng, A., Endo, S., Heus, T., … Xu, K.-M. (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.

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.

Bretherton, C., & Blossey, P. (2017). Understanding mesoscale aggregation of shallow cumulus convection using large-eddy simulation. *Journal of Advances in Modeling Earth Systems*, *9*(8), 2798–2821.

Dauhut, T., Couvreux, F., Bouniol, D., Beucher, F., Volkmer, L., Pörtge, V., … others (2022). Flower trade-wind clouds are shallow mesoscale convective

Table A11. Variables in the radiation data set, 2D radiation fluxes sampled every hour. TOA stands for Top of Atmosphere, TOM for Top of Model. Dimensions: (member, time, y, x)

variable	units	description
clwvi	kg/m^2	condensed water path
hfls	W/m^2	surface upward latent heat flux
hfss	W/m^2	surface upward sensible heat flux
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
rlutcs	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
rsutcs	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

systems. *Quarterly Journal of the Royal Meteorological Society*.
Denby, L. (2020). Discovering the importance of mesoscale cloud organization through unsupervised classification. *Geophysical Research Letters*, *47*(1), e2019GL085190.
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. doi: 10.1073/pnas.1514035112
George, G., Stevens, B., Bony, S., Pincus, R., Fairall, C., Schulz, H., ... others (2021). Joanne: Joint dropsonde observations of the atmosphere in tropical north atlantic meso-scale environments. *Earth System Science Data*, *13*(11), 5253–5272.
George, G., Stevens, B., Bony, S., Vogel, R., & Naumann, A. K. (2022). *The ubiquity of shallow circulations in the trades* (Tech. Rep.). Copernicus Meetings. doi: 10.5194/egusphere-egu22-13444
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. doi: 10.1126/science.abd3980
Groen, D., Arabnejad, H., Jancauskas, V., Edeling, W. N., Jansson, F., Richardson, R. A., ... Coveney, P. V. (2021). VECMAtk: a scalable verification, validation and uncertainty quantification toolkit for scientific simulations. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *379*(2197), 20200221. doi: 10.1098/rsta.2020.0221
Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., ... others (2020). The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049.
Heus, T., van Heerwaarden, C. C., Jonker, H. J., Siebesma, A. P., Axelsen, S., van den Dries, K., ... others (2010). Formulation of and numerical studies with the Dutch Atmospheric Large-Eddy Simulation (DALES). *Geosci. Model Dev*, *3*, 415–444.
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. *The Philosophical Magazine*, *17*(65), 5–11.
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: Atmospheres*, *113*(D13).
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*. doi: 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. (e2020GL091001 2020GL091001) doi: 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 and Engineering Sciences*, *379*(2197), 20200073. doi: 10.1098/rsta.2020.0073
Jansson, F., Janssens, M., & Grönqvist, J. (2023, March). *fjansson/cloudbotanyscripts: v1.0.0.* Zenodo. doi: 10.5281/zenodo.7709435
Jansson, F., Janssens, M., Grönqvist, J., Siebesma, P., Glassmeier, F., Attema, J., ... Kölling, T. (2023, March). *Cloud Botany LES ensemble visualizations.*

Zenodo. doi: 10.5281/zenodo.7692270

Lamaakel, O., & Matheou, G. (2022). Organization development in precipitating shallow cumulus convection: Evolution turbulence characteristics. *Journal of the Atmospheric Sciences*.

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.

Miles, A., jakirkham, Bussonnier, M., Moore, J., Orfanos, D. P., Fulton, A., … Banihirwe, A. (2022, October). *zarr-developers/zarr-python: v2.13.3.* Zenodo. doi: 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*, e2021MS002619.

Nitta, T., & Esbensen, S. (1974). Heat and moisture budget analyses using bomex data. *Monthly Weather Review*, *102*(1), 17–28.

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.

Nuijens, L., & Stevens, B. (2012). The influence of wind speed on shallow marine cumulus convection. *Journal of the atmospheric sciences*, *69*(1), 168–184.

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.

Pedgley, D. (2003). Luke howard and his clouds. *Weather*, *58*(2), 51–55.

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

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. doi: 10.5194/egusphere-egu22-8166

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.

Schulz, H. (2021). *Meso-scale patterns of shallow convection in the trades* (Doctoral dissertation, Universität Hamburg). doi: 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.

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

Schulzweida, U. (2021, October). *Cdo user guide.* Retrieved from `https://doi.org/10.5281/zenodo.5614769` doi: 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.

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.

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. doi: 10.1029/2021MS002631

Siebesma, A. P., Bretherton, C. S., Brown, A., Chlond, A., Cuxart, J., Duynkerke, P. G., … others (2003a). A large eddy simulation intercomparison study of shallow cumulus convection. *Journal of the Atmospheric Sciences*, *60*(10), 1201–1219.

Siebesma, A. P., Bretherton, C. S., Brown, A., Chlond, A., Cuxart, J., Duynkerke, P. G., ... Stevens, D. E. (2003b). A large eddy simulation intercomparison study of shallow cumulus convection. *Journal of the Atmospheric Sciences*, *60*(10), 1201-1219. doi: 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., ... others (2001). Simulations of trade wind cumuli under a strong inversion. *Journal of the Atmospheric Sciences*, *58*(14), 1870–1891.

Stevens, B., Bony, S., Brogniez, H., Hentgen, L., Hohenegger, C., Kiemle, C., ... others (2020). Sugar, gravel, fish and flowers: Mesoscale cloud patterns in the trade winds. *Quarterly Journal of the Royal Meteorological Society*, *146*(726), 141–152.

Stevens, B., Bony, S., Farrell, D., Ament, F., Blyth, A., Fairall, C., ... others (2021). EUREC4A. *Earth System Science Data*, *13*(8), 4067–4119. doi: 10.5194/essd-13-4067-2021

Stevens, B., Satoh, M., Auger, L., Biercamp, J., Bretherton, C. S., Chen, X., ... others (2019). DYAMOND: the dynamics of the atmospheric general circulation modeled on non-hydrostatic domains. *Progress in Earth and Planetary Science*, *6*(1), 1–17.

vanZanten, M. C., Stevens, B., Nuijens, L., Siebesma, A. P., Ackerman, A. S., Burnet, F., ... Wyszogrodzki, A. (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). doi: 10.1029/2011MS000056

Vial, J., Vogel, R., Bony, S., Stevens, B., Winker, D. M., Cai, X., ... Brogniez, H. (2019). A new look at the daily cycle of trade wind cumuli. *Journal of advances in modeling earth systems*, *11*(10), 3148–3166.

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.

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.

Wicker, L. J., & Skamarock, W. C. (2002). Time-splitting methods for elastic models using forward time schemes. *Monthly weather review*, *130*(8), 2088–2097.

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.

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. doi: 10.1256/qj.04.103