

# Pair Correlations and Spatial Statistics of Deep Convection over the Tropical Atlantic

FABIAN SENF

*Leibniz Institute for Tropospheric Research, Leipzig, Germany*

MATTHIAS BRUECK

*Max Planck Institute for Meteorology, Hamburg, Germany*

DANIEL KLOCKE

*Hans Ertel Center for Weather Research, Deutscher Wetterdienst, Offenbach, Germany*

(Manuscript received 2 November 2018, in final form 10 July 2019)

## ABSTRACT

Over the tropical oceans, the large-scale, meridional circulation drives the accumulation of moist and warm air, leading to a relatively narrow, convectively active band. Therein, deep moist convection interacts with its heterogeneous environment—the intertropical convergence zone (ITCZ)—and organizes into multiscale structures that strongly impact Earth’s hydrological cycle and radiation budget. Understanding the spatial correlations and interactions among deep convective clouds is important, but challenging. These clouds are investigated in this study with the help of large-domain, storm-resolving simulations over the tropical Atlantic. Based on vertically integrated mass flux fields, deep convective updraft cells are identified with object-based techniques and analyzed with respect to their structural behavior and spatial arrangement. The pair-correlation method, which compares simulated pair numbers as a function of pair distance to an appropriately chosen reference, is applied and extended to allow for spatial statistics in a heterogeneous environment (i.e., the ITCZ). Based on pair-correlation analysis, the average probability is enhanced to find an updraft cell pair within 100 km compared to a random distribution. Additionally, the spatial arrangement of larger or stronger cells deviates more from randomness compared to smaller or weaker cells, which might be related to their stronger dynamical interaction mechanisms. Using simplified equilibrium statistics of interacting cells, several spatial characteristics of the storm-resolving simulations can be reproduced.

## 1. Introduction

The inherent complexity of our nature can be experienced when we look at the fascinating beauty of clouds. They can appear in very irregular patches with fractal characteristics over a vast size range (Lovejoy 1982), sometimes being spatially clustered and other

times possessing a higher degree of randomness (Nair et al. 1998). Over the oceans, tropical deep convection arranges itself in a relatively narrow, convectively active band located in the intertropical convergence zone (ITCZ; Waliser and Gautier 1993; Schneider et al. 2014). Planetary and synoptic-scale waves cause undulations in the deep convection band and organize convective activity on the larger scales (Machado et al. 1993; Kiladis et al. 2009). On the mesoscales (i.e., from several hundred kilometers down to kilometer scale), several interaction mechanisms between clouds have been discussed. Preexisting clouds influence new cloud formation by modifying environmental properties due to local moistening (Randall and Huffman 1980; Hohenegger and Stevens 2013; Kumar et al. 2014) or due to changing atmospheric stability

---

Denotes content that is immediately available upon publication as open access.

---

Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/JAS-D-18-0326.s1>.

---

Corresponding author: Fabian Senf, [senf@tropos.de](mailto:senf@tropos.de)

DOI: 10.1175/JAS-D-18-0326.1

© 2019 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](#) ([www.ametsoc.org/PUBSReuseLicenses](http://www.ametsoc.org/PUBSReuseLicenses)).

(Ramirez et al. 1990; Ramirez and Bras 1990). Competing circulations across cloud edges have been identified to induce dynamical feedbacks among clouds (Garrett et al. 2018).

A generic prototype of cloud organization found in idealized simulations has stimulated scientific interest in the recent years (see reviews of Holloway 2017; Wing et al. 2017, and references therein). Depending on environmental conditions, isolated convective cells start to spontaneously aggregate into bigger cloud clusters in which most of the moisture is concentrated (Bretherton et al. 2005; Muller and Held 2012). The so-called convective self-aggregation can be sensitive to several factors of physical origin, like ice microphysics and radiation (Bretherton et al. 2005) and subgrid-scale mixing (Tompkins and Semie 2017), or to factors of numerical origin like simulation domain size and horizontal resolution (Muller and Held 2012). It has been speculated that convective aggregation could be important for current and future climate (Tobin et al. 2012; Bony et al. 2015; Mauritsen and Stevens 2015) by modifying radiation and precipitation budgets. It is far from being trivial to assess if this mechanism can play a role in more realistic situations where tropical deep convection develops within a very heterogeneous environment even though progress toward more realism has been made more recently (Holloway 2017; Muller and Romps 2018).

Therefore, the current study pursues two general goals: (i) to investigate the spatial distribution of marine tropical deep convection in storm-resolving (no convection parameterization) model simulations and (ii) to identify possible analogies between the spatial statistics of deep convective cells and statistical physics approaches. For the *first goal*, we focus on configurational statistics of deep convective cells (i.e., how these cells organize in space). The so-called pair-correlation method (e.g., Rasp et al. 2018) is extended to allow for an application in the highly heterogeneous ITCZ. The pair-correlation functions can provide a spatial length scale at which the probability of cell pair occurrences is enhanced compared to a random reference. Thus, a link to spatial interaction mechanisms needed for the parameterization of convective organization can be established. Similarly, Tan et al. (2015) discussed the consequences of temporal and spatial localization of deep convection. They stated that statistical convection parameterizations should account for a nonlocal influence of deep moist convection to realistically represent spatial coherence. Local random interactions have been shown to be important for macroscopic cloud organization (Yuan 2011; Bengtsson et al. 2013; Khouider 2014).

Many of the most recently developed parameterization approaches have their roots in modern statistical physics (e.g., Majda and Khouider 2002; Craig and Cohen 2006; Yuan 2011). Motivated by this, the *second goal* of our study is to discuss the results from spatial statistics of cells in relation to simple statistical physics approaches that allow for clustered cell configurations. Challenges arise for describing and quantifying the way how properties of the cell populations are changed by the effects of cell interactions. For instance, Craig and Cohen (2006) analyzed statistical equilibrium fluctuations of a field of noninteracting cumulus clouds under homogeneous large-scale forcing in analogy to an ideal gas. As a possible route for extending the existing formulations, one could keep the equilibrium assumption, but include effects of external forcing and cell interactions leading to the statistical mechanics of nonuniform fluids (Henderson 1992). Such an approach would be similar to the lattice-based model of convective inhibition fluctuations by Majda and Khouider (2002), who specified nearest-neighbor interactions between active sites, and an external potential depending on the grid-scale environment.

A further extension toward nonequilibrium approaches might be inevitable to cover the full wealth of interaction mechanisms. In this way, Peters and Neelin (2006) and Peters et al. (2009) suggested that the tropical precipitation system undergoes a second-order phase transition with the total column water vapor (TCWV) as control parameter. Feedbacks in the moisture dynamics (Craig and Mack 2013) then lead to self-organized criticality. Wang (2014) further emphasized the importance of moisture and high TCWV values for tropical cyclone formation. Advanced multitype stochastic models typically based on Markov chain methods (Khouider et al. 2010; Peters et al. 2013; Khouider 2014; Ragone et al. 2015; Dorrestijn et al. 2015, 2016; Gottwald et al. 2016) could have the capability to integrate elements of nonequilibrium behavior as well as spatial coherence into a joint framework. However, the challenge remains to constrain the parameters that govern formation, transition and interaction mechanisms.

The structure of the paper is outlined in the following: Large-domain storm-resolving simulations over the tropical Atlantic are utilized to derive updraft cells as described in section 2. Background on statistical methods including configurational analysis and recipes for the calculation of total and partial pair-correlation functions are provided in section 3. In the results section 4, we present and discuss the analysis of cell characteristics and pair correlations. We further consider simplified equilibrium statistics of interacting cells using the “convective organization potential” introduced by

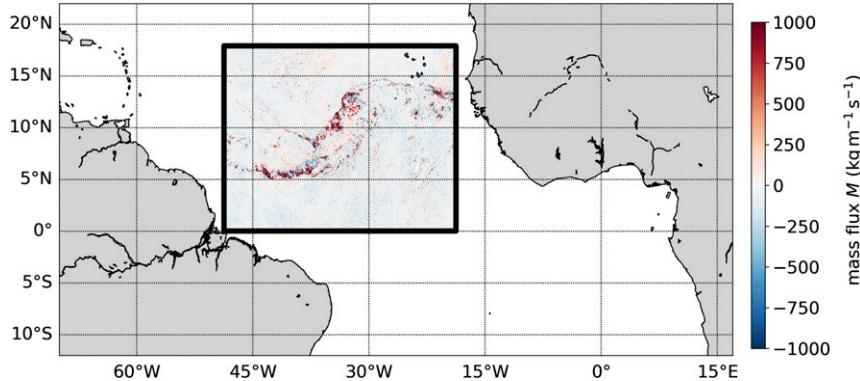


FIG. 1. Overview of the full ICON simulation domain and the analyzed subdomain in the tropical Atlantic (marked by a black bounding box). The field inside the bounding box is an example of the vertically integrated mass flux  $M$  originating from a simulation initiated at 0000 UTC 16 Aug 2016 after 24 h into the forecast.

White et al. (2018) and assess the results of this simple model. A summary and an outlook are provided in section 5, and an appendix adds details about application of the statistical model.

## 2. Data and object-based analysis

### a. Storm-resolving simulations

We use data from simulations with the Icosahedral Nonhydrostatic (ICON) model (Zängl et al. 2015) that have already been presented in previous studies (Klocke et al. 2017; Senf et al. 2018). The average horizontal resolution on the icosahedral simulation grid is 2.5 km without any parameterization for cumulus convection. All of the convective mass flux is explicitly transported at the grid scale. We refer to the simulation configuration as “storm resolving” to remain consistent with our previous studies. The simulations were conducted in support of the Next-Generation Aircraft Remote Sensing for Validation Studies II (NARVAL II) campaign (Stevens et al. 2016; Klepp et al. 2015). They cover the tropical Atlantic and parts of the adjacent continents (cf. Fig. 1) for August 2016. Each day at 0000 UTC, a new simulation is initialized from ECMWF analysis and runs out to +36 h. Three-hourly data from ECMWF analysis is provided at the lateral boundaries. The first 12 h of each simulation are disregarded as spinup period of the model. Details about the setup are further discussed in Klocke et al. (2017), where low wind conditions in the ITCZ region are investigated, and in Senf et al. (2018), where the model simulations are evaluated against Meteosat observations.

In the following, we consider a subdomain over the tropical Atlantic as indicated in Fig. 1. Deep convection is characterized by high, positive mass flux

regions. Mass flux, in general, is one of the key components of convective parameterizations (Arakawa 2004). From the simulated mass flux profiles, the vertically integrated mass flux  $M = \int dz \rho w$  is calculated, where  $\rho$  is the air density and  $w$  denotes the vertical velocity. An example of a simulated  $M$  field is also given in Fig. 1 and a more detailed view is provided in Fig. 2. Large positive values of  $M$  indicate strong upward motion in deep convective cores; negative  $M$  values appear for pronounced downdrafts, which are usually connected to convective precipitation processes. In the following analysis,  $M$  is abbreviated as “mass flux” to simplify notation. The positive and negative parts of  $M$  are examined separately. The areas covered by upward and downward motion are denoted by  $\mathcal{A}_+$  and  $\mathcal{A}_-$ , respectively, covering 32% and 68% of the Atlantic domain for the scene in Fig. 1. Taking the areal integral of  $M$  over  $\mathcal{A}_+$  gives the total mass lift

$$M_{\text{tot},+}(t) = \int_{\mathcal{A}_+} d\mathbf{x} M(\mathbf{x}, t), \quad (1)$$

which is the volume integral of the total upward-directed vertical momentum. In that way,  $M_{\text{tot},+}$  can be directly associated with the magnitude of the ascending branch of the tropical circulation. For the scene in Fig. 1, we have  $M_{\text{tot},+} = 4.5 \times 10^{14} \text{ kg m s}^{-1}$ , and  $M_{\text{tot},-} = -4.3 \times 10^{14} \text{ kg m s}^{-1}$  for the downward counterpart. Thus, not surprisingly, upward motion is mainly compensated by a downward mass transport in the surrounding regions. (The time series of  $M_{\text{tot},+}$  is shown in Fig. 4a.) Around 75% of the temporal variance in  $M_{\text{tot},+}(t)$  is induced by multiday variability. On shorter time scales,  $M_{\text{tot},+}(t)$  shows a pronounced diurnal cycle with a minimum at 0100 UTC and a maximum at 1800 UTC.

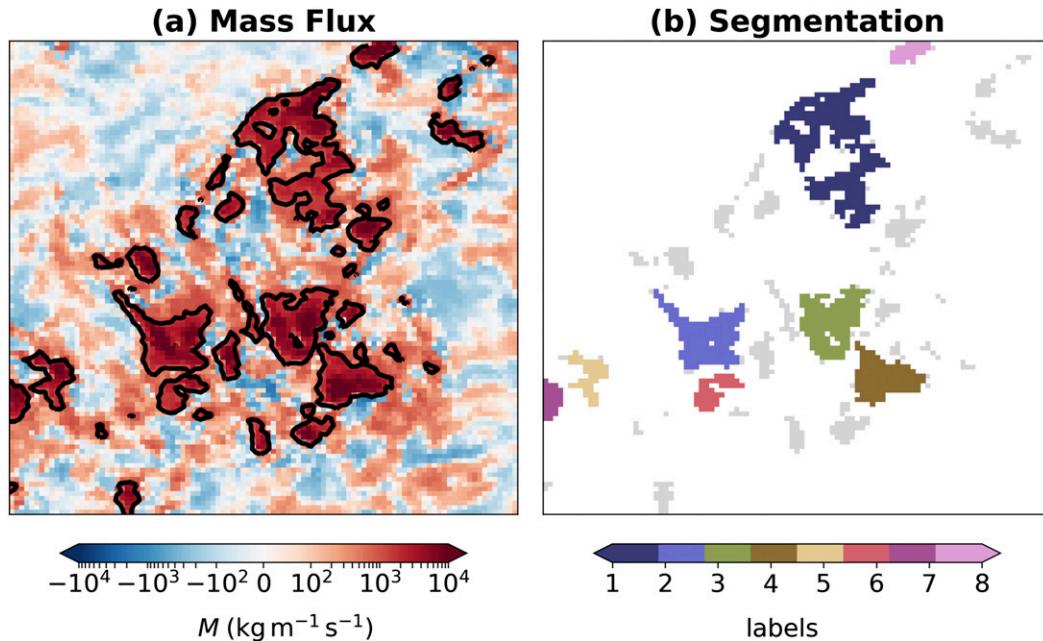


FIG. 2. Detailed view of (a) the vertically integrated mass flux field  $M$  and (b) the cell label field derived from threshold-based segmentation. The same scene as in Fig. 1 is shown for a small subregion (indicated in Fig. 3) with a side length of 275 km. The color scale in (a) is linear within the range between  $-10^2$  and  $10^2 \text{ kg m}^{-1} \text{ s}^{-1}$  and logarithmic outside. The threshold for segmentation is chosen to be  $M = 1231 \text{ kg m}^{-1} \text{ s}^{-1}$ , cell sizes decrease with increasing label numbers, and cells with diameters smaller than 20 km have been removed from the final result [light gray regions in (b)].

### b. Cell definition

Our aim is to identify updraft cells that contribute to a certain fraction  $f$  of the total mass lift  $M_{\text{tot},+}$ . Considering the instantaneous  $M$  field, we see that spatial integration over areas

$$\mathcal{A}_f = \{\mathbf{x}: M(\mathbf{x}, t) > M_{\text{thresh},f}(t)\}, \quad (2)$$

which contain values larger than a certain threshold, gives the fraction

$$f = \int_{\mathcal{A}_f} d\mathbf{x} M / M_{\text{tot},+} \quad (3)$$

of the total mass lift that is realized by the chosen updraft areas. For small  $f$ , only the most intense updraft cores are selected. For  $f$  close to one, the full upward-directed mass-flux field is recovered. Considering  $M$  and  $f$  to be given, Eq. (3) provides an implicit formula from which  $M_{\text{thresh},f}(t)$  can be derived. Keeping  $f$  fixed over time leads to a time-dependent threshold, a route that we will choose for our analysis. Conversely, a constant  $M_{\text{thresh}}$  leads to a cell population that has varying relative contribution to the instantaneous  $M_{\text{tot},+}$ .

For the scene in Figs. 1 and 2, the relative fraction was set to  $f = 0.5$  (i.e., 50% of the total mass lift), and

the resulting  $M$  threshold is at  $1231 \text{ kg m}^{-1} \text{ s}^{-1}$ . Grid boxes with larger  $M$  values only cover 1% of the Atlantic domain. The mass-flux distribution is skewed (not shown). Half of the downward mass transport is enclosed by the  $-135 \text{ kg m}^{-1} \text{ s}^{-1}$  contour, which has a smaller magnitude and covers a larger area fraction of 14%. This asymmetry is a common feature of circulations induced by deep moist convective motions (Stevens 2005). Selecting  $f = 0.5$  for the whole analysis period, the time series of  $M_{\text{thresh},50\%}$  results (also shown in Fig. 4a). In general, it follows the curve of the total mass lift. For instance, lower threshold values are obtained in situations where the ascending circulation is weak.

The cell identification is done on  $M$  fields via a threshold-based segmentation algorithm. The algorithm is similar to the one described in Senf et al. (2018), but here applied to  $M$  fields with variable thresholds  $M_{\text{thresh},f}(t)$  in contrast to brightness temperatures with a fixed threshold. In a nutshell, a binary map is generated by setting grid points to one (zero) for  $M$  values above (below) the considered threshold. The binary field is separated into distinct objects in two steps. First, contiguous subregions of category “one” are labeled with different cell indices where only connections between nearest neighbors sharing the same grid edge are

considered (four-connectivity). Second, subsequent object erosion is used, which removes cell parts that are located close to the cell edges. The remaining field of eroded objects defines the marker positions from which a watershed segmentation is started. The watershed algorithm can be interpreted as a rising water level and fills all areas above the selected threshold with a certain label value starting from the individual marker positions. As a consequence, different objects now could have a common interface. This interface is also assessed in the second step. If an interface length of two objects is greater than 50% of the size of one of the considered objects, then the two objects are merged. This split-and-merge procedure was invented for situations in which rather circular-shaped anvils start to touch each other (Senf et al. 2018). Due to the very irregular shape of updraft objects (see Fig. 2), this procedure is rather inactive and differences in the segmentation results due to the second step (i.e., compared to a standard connected compound method) are rather small.

A set of different properties is derived for each cell  $i$ ; the most important ones are the center-of-mass positions  $\mathbf{x}_i$ , the equivalent diameters  $D_i$  as diameters of area-equal circles of area  $a_i$ , and the cell-integrated upward mass lift values  $m_i$ . The cell-average mass flux  $w_i = m_i/a_i$  can be interpreted as total column ascent speeds when divided by the typical total air mass load of  $\int dz\rho \approx 10^4 \text{ kg m}^2$ . The  $w_i$  also provides a measure of how efficient a convective cell is transporting mass in the vertical. For the object-based analysis, we filter the cell set for  $D_i > 20 \text{ km}$ , which corresponds roughly to the effective model resolution (Dipankar et al. 2015; Bley et al. 2017). Due to this size filter, a significant part of updraft cores is excluded from our current analysis. The motivation for this step is that characteristics of smaller updrafts might be strongly influenced by numerical diffusion effects and a structural analysis might be inadequate. For  $f = 0.5$ , the remaining cell population with  $D_i > 20 \text{ km}$  still contributes on average to around 33% to the total mass lift, which also means that 17% has been disregarded by the size filter. Environmental characteristics of updraft cells are also collected. We consider average TCWV, average sea surface temperature (SST), and maximum convective available potential energy (CAPE) values in the vicinity of each identified cell using a box of  $50 \times 50$  grid points ( $\sim 120\text{-km}$  side length). CAPE is derived from average potential energy of a set of parcels coming from the mean surface layer. These environmental values already include modifications due to the convection itself, especially for larger organized convective systems.

### 3. Background on statistical methods

#### a. Configurational analysis

We are interested in how different updraft cells are distributed in space and how cells are affected by their neighbors. These aspects are investigated using configurational analysis. At a given time, a set of cell positions defines a spatial configuration. At the lowest order, the cell configurations are characterized by the average cell density  $n(\mathbf{x})$  where the explicit dependence on spatial coordinates was kept to allow for heterogeneous situations. On monthly average, updraft cells organize themselves into a narrow band over the ocean (see Fig. 3a for fixed  $f = 0.5$ ) at which  $n(\mathbf{x})$  shows a pronounced maximum around 900 km north of the equator with a slight southward tilt toward the west. The convectively active zone shows longitudinal gradients of the cell density along the ITCZ region with a maximum density of more than 30 cells per 1 million  $\text{km}^2$  in the eastern part and a vanishing number density in the western part of the tropical Atlantic. However, this feature might be an artifact in the storm-resolving simulations. In Senf et al. (2018), a deficit of simulated large cold cells in the western tropical Atlantic was identified based on comparisons with Meteosat observations.

We are further interested in how cell pairs behave and therefore increase the dimensionality of configurational space. The time-average pair number density  $n^{(2)}(\mathbf{x}_1, \mathbf{x}_2)$  gives the average number of pairs having cell 1 at position  $\mathbf{x}_1$  and cell 2 at position  $\mathbf{x}_2$  (see, e.g., Chandler 1987, p. 195). The function is already four-dimensional, and thus hard to visualize and to robustly estimate. Therefore, the pair number density is decomposed into single-cell number densities and a term that accounts for pair correlations

$$n^{(2)}(\mathbf{x}_1, \mathbf{x}_2) = n(\mathbf{x}_1) n(\mathbf{x}_2) g(\mathbf{x}_1, \mathbf{x}_2). \quad (4)$$

The term  $g$  is called pair correlation function (PCF; or radial distribution function) and is equal to one if all cells are noninteracting. The PCF is typically plotted against the distance between cells  $r = |\mathbf{x}_2 - \mathbf{x}_1|$  to investigate the clustering behavior of cell pairs (e.g., Rasp et al. 2018).

In the practical application,  $g$  values are approximated by counting the actual number of neighboring cells (at varying  $\mathbf{x}_1$ ) in a certain range around a selected target cell (at fixed  $\mathbf{x}_2$ ). This absolute frequency of pairs is then divided by the expectation value for noninteracting cells in the same range. In mathematical terms, we have

### Spatial Arrangement of Updraft Cells in the Tropical Atlantic

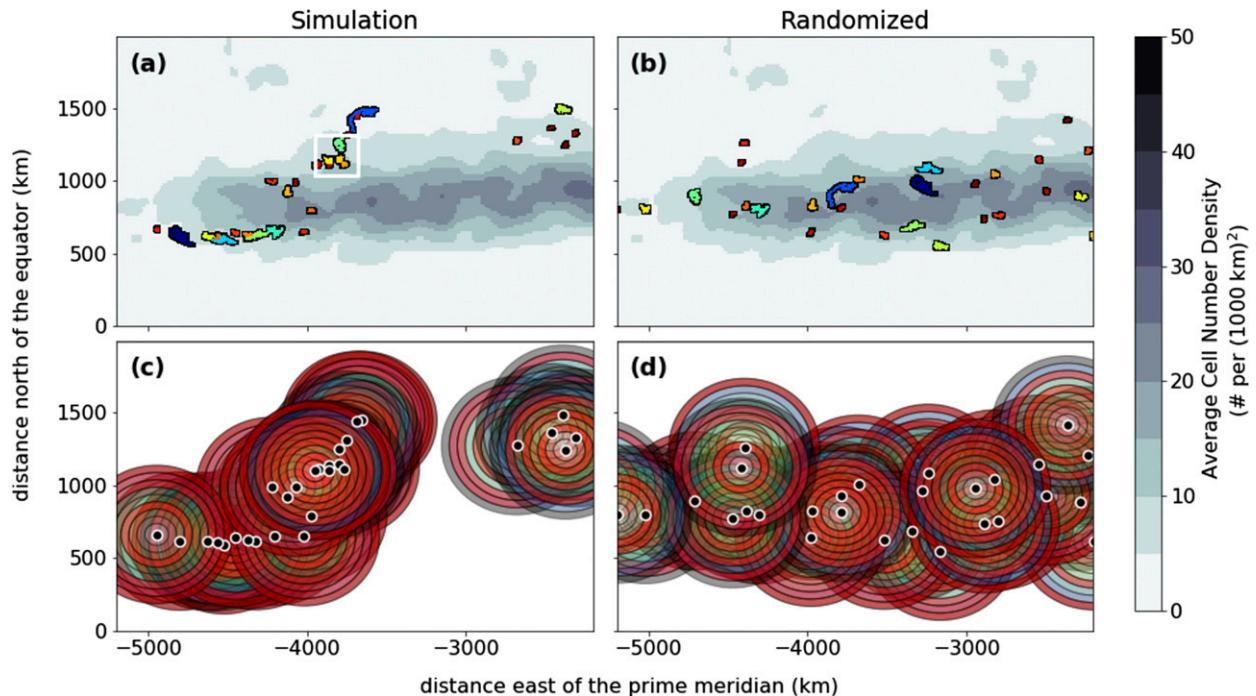


FIG. 3. (a),(b) The monthly average cell number density as gray background shadings for the subdomain shown in Fig. 1 and setting  $f = 0.5$ . The segmented updraft cells derived from the ICON simulation (rainbow colors with black edges) in (a) are randomly rearranged avoiding overlap in (b). (c),(d) The corresponding centroid positions are shown with black circles. Colored range rings are centered around each cell with red colors indicating at least one neighbor cell at the highlighted distance. A ring size of 50 km has been chosen for visualization. The white rectangle in (a) marks the region chosen for the detailed view in Fig. 2.

$$N_{|2}(r) = \int_{\mathcal{R}_2(r)} d\mathbf{x}_1 \frac{n^{(2)}(\mathbf{x}_1, \mathbf{x}_2)}{n(\mathbf{x}_2)} \quad (5)$$

for the realized number of cell pairs over a range ring  $\mathcal{R}_2(r) = \{\mathbf{x}_1: r - \delta r/2 \leq |\mathbf{x}_2 - \mathbf{x}_1| \leq r + \delta r/2\}$ , where  $\delta r$  denotes the ring width, and

$$N_2^{(0)}(r) = \int_{\mathcal{R}_2(r)} d\mathbf{x}_1 n(\mathbf{x}_1) \quad (6)$$

for the reference number of cell pairs. We approximate PCF by

$$g(r) \approx \frac{\overline{N_{|2}(r, t_k)}^{i,k}}{\overline{N_2^{(0)}(r, t_k)}^{i,k}}, \quad (7)$$

where the overbar is a shortcut for different averaging strategies, for example, average over the configuration  $\{i\}$  and then average for 1 month  $\{k\}$ . This illustrates that the determination of  $g(r)$  is just a counting exercise. Values of  $g(r)$  larger than one indicate an accumulation of cells in a certain distance range around  $r$ , which might be connected to an attractive force between the cells,

whereas  $g$  values smaller than one indicate cell inhibition possibly due to repulsive forces between cells within a certain distance. Figure 3c gives an illustration of derived cell centroid positions and range rings around the cells that are used for counting neighboring cells.

#### b. Construction of a reference

The PCF calculation in Eq. (7) involves a reference pair number  $N_i^{(0)}$  as function of pair distance. The PCF can only be interpreted as probability of enhanced or decreased pair occurrences if the reference is chosen in an appropriate way. Any oversimplification can introduce artificial pair correlations.

In the following, we will first discuss the main problems that arise from the incomplete knowledge of the reference state and then provide a practical solution that uses repeated randomization of the updraft cell positions. The three aspects being especially problematic for the analysis of updraft cells in the tropics are as follows:

- (i) Spatial heterogeneity: Applications in statistical physics (see, e.g., Chandler 1987) as well as in atmospheric science (Nair et al. 1998; Rasp et al. 2018) typically assume a spatially constant number

density as reference. In that case, the reference pair number is proportional to the range-ring area and the total number of cells. For tropical convection, this is however an invalid oversimplification as the environment is heterogeneous and gradients in the cell-density field  $n(\mathbf{x})$  are nonnegligible (see, e.g., Fig. 3a).

- (ii) Nonstationarity: Temporal patterns, like diurnal cycle and easterly waves, modulate the probability for certain cell configurations (see, e.g., Machado et al. 1993). Moreover, the meridional shift of the ITCZ with the annual cycle limits the time range to gather statistics.
- (iii) Finite cell size: Tropical convection cells organize into connected cloud clusters that can grow up to a size of several hundred kilometers (see, e.g., Houze 2004). The finite size of these large cells introduces an apparent inhibition because different cells are not allowed to overlap [see Nair et al. (1998) for an introduction into the problem].

To address the problems outlined above, we apply and extend the randomization approach of Nair et al. (1998). In its original formulation, cells have been approximated by circles to treat the finite-size problem and reinserted at new random positions without overlap to generate a reference population. This randomization strategy can already deal with a temporally changing total number of cells and thus partially addresses the nonstationary problem. As a first extension, our method draws new cell positions from an inhomogeneous cell distribution. We utilize the smooth, monthly average cell density field  $n(\mathbf{x})$  (see, e.g., Fig. 3a). This still ignores coherent spatial patterns connected to temporal variability like tropical waves—a limitation that cannot be solved here. As second extension, the circle approximation of Nair et al. (1998) is relaxed and each cell is individually cutout and reinserted preserving its size, shape and orientation. In the random rearrangement procedure, we start with the largest cells and subsequently proceed to smaller ones. All cells are randomly inserted under the condition of no overlap and a predefined minimal edge-to-edge distance. If this condition is not fulfilled, new cell positions are drawn and the insertion procedure is repeated until successful. All intrinsic cell characteristics,  $a_i$ ,  $m_i$ , and  $w_i$ , are preserved with the cells. A realization of a randomized cell field is shown in Figs. 3b and 3d. In total, 20 realizations of randomized cell populations are computed. The different realizations are analyzed in the same fashion as their unmodified counterparts. Ensemble average values are used to approximate the characteristics of the reference population (e.g., the reference number of cell pairs).

### c. Partial pair correlations

A further extension to the standard pair-correlation method is developed to assess the dependency of PCF on cell characteristics. For the sake of simplicity, we apply a categorization of cell properties using a predefined set of bins and then apply the calculation of so-called partial PCFs to the different category combinations. This is very similar to the analysis of multicomponent systems in statistical physics where each category corresponds to a different component (see, e.g., Chandler 1987). For instance, using cell diameter as categorization variable and selecting diameter bins that correspond to small, medium-size and large cells, the multicomponent system would consist of three categories. The computation of partial PCFs would only involve a subset of cell pairs, where the target cell has to belong to one category (e.g., small size) and only neighboring cells from a second category (e.g., large size) are counted within a certain range around the selected target cell. Analyzing all possible category combinations, a matrix of partial PCFs ( $3 \times 3$  for our example) is constructed where the diagonal terms result from a connection of pairs with same categories and the off-diagonal terms give the cross-category connections. The reference number of cell pairs for the partial PCFs is determined with the same randomization procedure described above. This ignores a possible dependence of the reference number density on the categorical variable, an effect that is discussed later in the next section.

## 4. Results and discussion

### a. Cell characteristics and pair correlations

We start with a discussion of the properties of a simulated updraft cell population that contributes to 50% of the total mass lift (i.e., setting a fixed  $f = 0.5$ ). The resulting cell size distributions  $dN/dD$ , defined here as total number of cells  $N$  in a certain diameter interval  $dD$ , show rather scale-free behavior across a size range between 20 and 200 km (Fig. 4b). This is in good agreement with previous studies of size distributions of clouds (Machado et al. 1992, 1993; Wood and Field 2011) and precipitation cores (Peters et al. 2009; Peters and Neelin 2006) over tropical oceans. Fitting a function of the shape  $dN/dD \sim D^{-\beta}$ , a power-law exponent  $\beta$  of larger than 3 arises. Due to the short range of spatial scales, the slope is uncertain and we cannot exclude that the size distributions are governed by a different functional dependency. Figure 4b also illustrates that the average aspect ratios [defined as ratio between minor and major axis of an equivalent ellipse, i.e.,  $(D_i/D_{\max,i})^2$  with

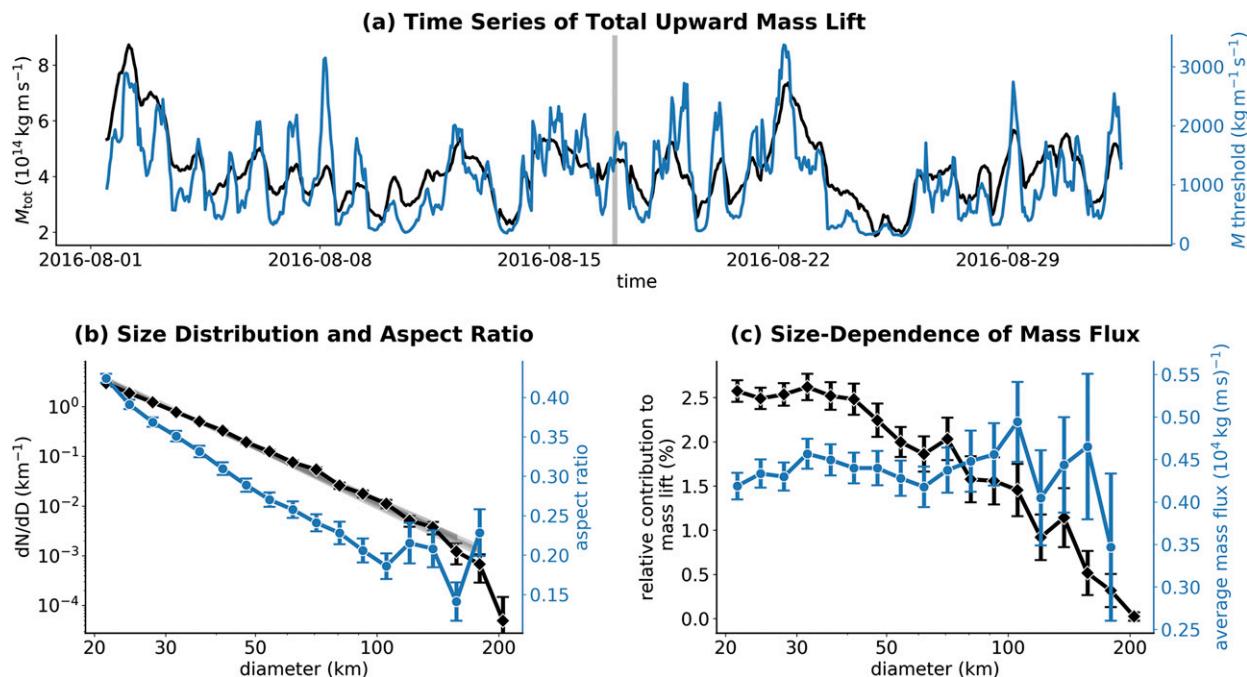


FIG. 4. (a) Time series of total upward mass lift  $M_{\text{tot},+}$  (black line) and the threshold  $M_{\text{thresh},50\%}$  (blue line). The gray vertical line marks the time shown in Figs. 1–3. (b),(c) Cell characteristics conditioned on a certain property range ( $x$  values) also derived for  $f = 0.5$ . Each cell configuration (in total  $31 \times 24 = 744$ ) is considered separately and statistics are averaged thereafter. Symbols (connected by lines) denote monthly mean statistics and error bars indicate standard errors of the monthly average values. (b) The cell number size distribution (CNSD, black) and the cell aspect ratio (blue). The gray shading behind the black line indicates a set of linear least squares fits that used different random CSND realizations calculated from a normal distribution using CSND mean and standard error. Outliers have been removed prior to the linear fit. (c) The relative contribution of each diameter interval to the total updraft mass lift (black) and the cell-average mass flux (blue). All blue curves have a separate  $y$  axis on the right side of each plot.

maximum cell dimension  $D_{\text{max},i}$ ] significantly decrease for increasing size with larger updraft cells being more elongated than smaller ones. For the largest cells with  $D_i > 100 \text{ km}$ , the major axis is 4–5 times larger than an area-equivalent minor axis.

Figure 4c shows the relative contribution of each specified diameter interval (logarithmic spacing) to the total mass lift. This relative contribution is increasing with decreasing cell size meaning that relatively more mass is lifted by small cells. As already indicated by the steep distribution slope, a significant contribution to the mass lift comes from cells below the chosen analysis limit of 20 km. For even smaller cells, the relative contribution to the total mass lift is decreasing (not shown). This signature, however, might result rather from numerical smoothing than being a physical effect. It therefore remains open at which scales the dominant contribution to the tropical mass lift and thus the forcing of the upward-directed tropical circulation is realized. This is in contrast to the scaling behavior of cold cloud coverage analyzed in Senf et al. (2018) for the same simulations in which the largest convective cloud clusters dominated the total cold cloud coverage.

The cell-average mass fluxes remain rather constant for cell diameters between 20 and 100 km (also Fig. 4c). These mass-flux values can also be interpreted as a cell-average efficiency for vertical mass transport. For larger cells, we find a much higher statistical uncertainty about the functional behavior indicated by the much larger standard errors. If the final drop of average mass-flux values at large diameters is not a statistical artifact (which cannot be excluded), then these cells would possess a lower average efficiency for vertical mass transport. In this scenario, a size of 100 km would mark a transition point where changes in the cell-internal convective organization lead to larger updraft regions with on average lower ascent rates, a pattern typically found in mesoscale convective systems (Moncrieff 1992; Houze 2004).

Monthly averaged number distributions in dependence of different environmental properties are shown in Fig. 5. Peak values are found for maximum CAPE around  $1400 \text{ J kg}^{-1}$ , for SST around 301 K, and for TCWV at 59 mm. Furthermore, variability of the size distributions is increased in the peak regions. Figure 5 also provides number distributions of the reference cell

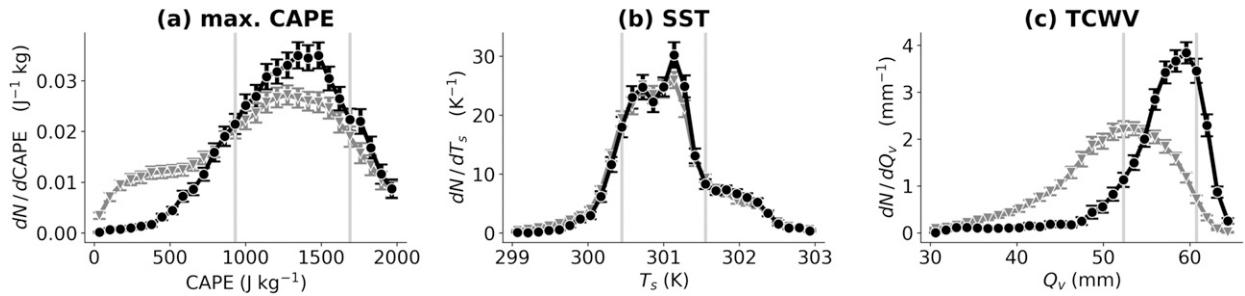


FIG. 5. Cell number distributions conditioned on different environmental properties: (a) CAPE, (b) sea surface temperature  $T_s$ , and (c) total-column water vapor  $Q_v$ . Black lines and symbols indicate characteristics directly derived from ICON simulations by setting  $f = 0.5$ . Gray curves and symbols are derived for reference cell configurations that arise due to spatial rearrangement of cells by the randomization method. Symbols are monthly averages and standard errors are shown by error bars. The thin gray vertical lines indicate the 16th and 84th percentiles of the simulated distributions.

populations that arise from random rearrangement (see section 3b). In the randomization procedure, we only took the time-average spatial cell densities into account ignoring temporal fluctuations. From Fig. 5, it is apparent that updraft cells have been partly reinserted in more unfavorable environments with largest deviations for TCWV. This clearly shows the limitations of a time-constant reference cell density and the importance of temporal fluctuations of TCWV and CAPE.

Moreover, average environmental characteristics of cells are presented in Fig. 6. The average TCWV is significantly higher for larger cells while SST and CAPE seem to be rather independent of cell size showing only a slight reduction for very large cells. Based on satellite observations of precipitation cores, Peters et al. (2009) showed that larger precipitating cells are found for higher TCWV. They discussed that a rapid change in structural cell properties is related to a critical behavior of the tropical precipitation system when crossing a certain TCWV threshold. Updraft cells with increasing cell-average mass fluxes experience higher CAPE environments (see Fig. 6b), which reflects the transfer of potential into kinetic energies. SST and TCWV slightly increase from small to medium mass-flux values and become rather uncertain thereafter.

As discussed earlier, several cell characteristics depend on cell size and updraft strength. However, the organizational state of tropical convection is not only determined on a level of single cells (and therefore influenced by its size), but also by the spatial arrangement and interaction of cells. To quantify the statistics of these aspects, we apply configurational analysis (see section 3) and compute the average cell densities  $n$  and PCFs for different cell categories. We start with a PCF analysis that concentrates on a cell population that contributes 50% to the total mass lift, that is,  $f = 0.5$  (see Figs. 7 and 8). Figure 7a shows the PDFs of the meridional cell occurrences, which appear centered

around 900 km north of the equator. The width of the cell distribution is reduced for the larger cells causing a slightly higher peak of the PDF. The average total and partial PCFs for different cell sizes are presented in Fig. 7b. At a distance between 50 and 100 km, updraft cells are 3 times more frequent than by chance. The total PCF exhibits a peak at very short pair distances. At this range, PCF estimates are affected by the

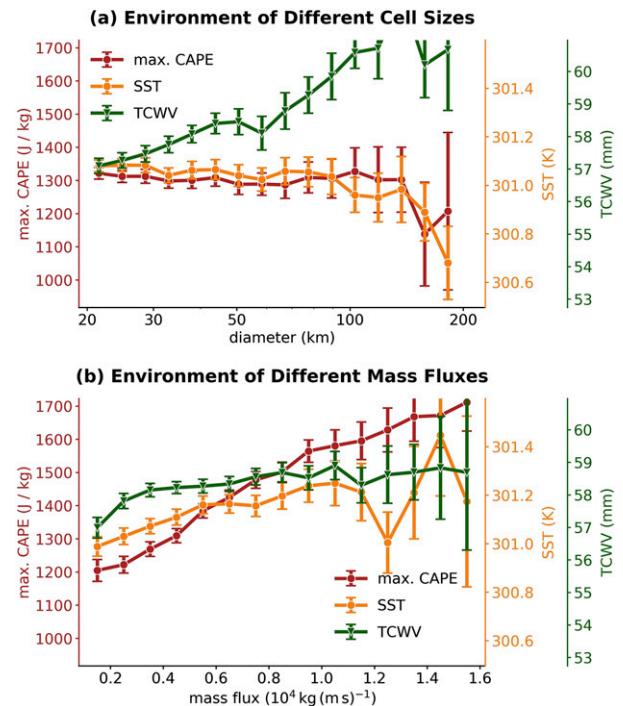


FIG. 6. As in Fig. 4c, but for average environmental properties conditioned on (a) diameter intervals and (b) cell-average mass flux intervals. Average values of CAPE,  $T_s$ , and  $Q_v$  are shown in red, orange, and green, respectively. Separate colored y axes are provided for each quantity. For an objective comparison, the y-axis range expands from the 16th to the 84th percentile of the distributions of environmental properties shown in Fig. 5.

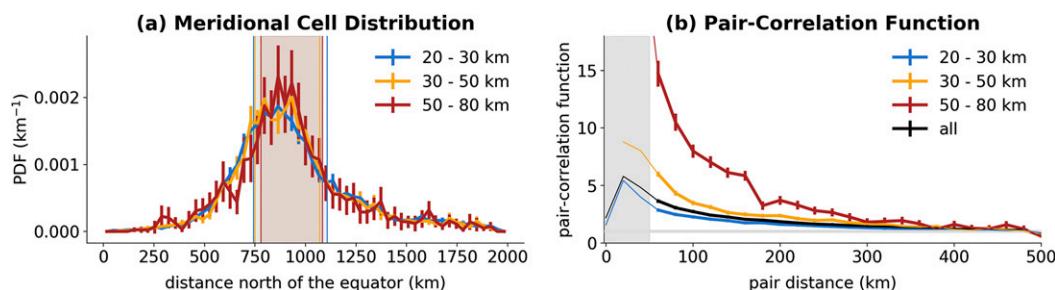


FIG. 7. Configurational statistics of updraft cells simulated by ICON and setting  $f = 0.5$ : (a) average meridional cell distributions and (b) total and partial PCFs. The statistics has been gathered for all (black), small (blue), medium-size (yellow), and large (red) cells, separately. The solid lines give monthly averages and the error bars show their standard error. The vertical lines and colored ranges in (a) mark the lower and upper quartiles of the cell distribution. The gray line in (b) shows the reference pair correlation of a randomly distributed field and the gray range between 0 and 50 km masks the highly uncertain PCF range.

randomization strategy and the reduction could be an artifact of the analysis method and not an indication for real inhibition. The total PCF decays back to one (random configuration) for large pair distances. The partial PCFs depend on cell size and have different peak values in the range between 50 and 100 km: 3 for small cells, 6 for medium-size cells, and 15 for large cells. This means that the largest cells deviate the most from randomness.

Figure 8 compares the average partial PCFs for different categorization variables (listed in Table 1). In contrast to Fig. 7, the PCFs are shown on logarithmic scale. We obtain increased PCF values for cells with large diameters, mass-flux and integrated mass-lift values. Hence, also the spatial arrangement of stronger tropical cells deviates more from randomness than the one of weaker cells. We speculate that circulation feedbacks of the larger cell objects induce a higher degree of organization. Moreover, structural changes, which updraft cells experience during their life cycle from formation to dissipation,

might also strongly impact their spatial configuration. Future research is needed to elucidate such mechanisms in more detail.

The cell identification was repeated for different choices of  $f$ , which defines how much a certain cell population is contributing to the total mass lift. Results are presented in Fig. 9 for  $f$  values chosen between 30% and 60%. For smaller  $f$  values, the cell population is more strictly constrained to intense updraft cores. Conversely, also weaker updrafts are included when larger  $f$  values are selected, and thus larger contributions to the total mass lift arise. The additional analysis supports the conclusion made earlier based on partial PCFs with a single threshold of  $f = 0.5$  (see Figs. 7b and 8). The width of the cell distributions is decreasing for more intense updrafts. At the same time, PCF values increase, confirming that the spatial arrangement of more intense updraft cells deviates more from randomness. Moreover, Figs. 9b and 9d provide further analysis of the sensitivity of our results to the threshold definition in the

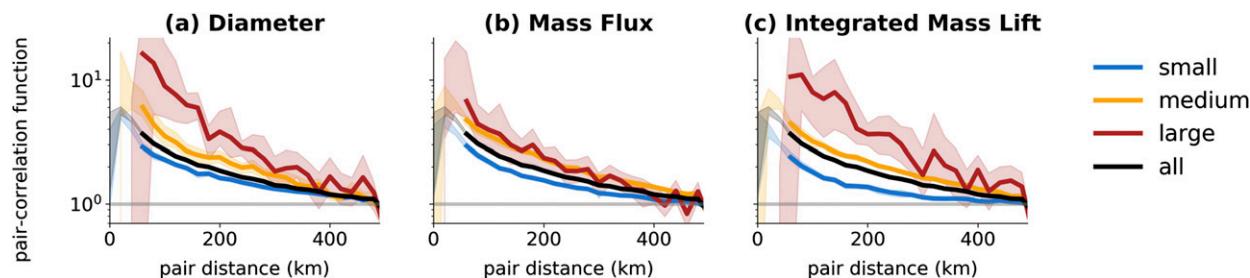


FIG. 8. Average pair correlations as function of pair distance for different combinations of cell characteristics. Similar to the analysis in Fig. 7,  $f$  is set to 0.5 and PCFs are first calculated as averages for a certain output time. Thereafter, monthly averaging is applied (thick solid lines) and the standard error is estimated (shading). Black curves give the total pair-correlation function of the full cell set (same in each panel). Colored curves give the partial pair correlations as described in section 3 where the shortcuts “small,” “medium,” and “large” refer to a categorization based on the variables (a) diameter, (b) mass flux, and (c) integrated mass lift listed in Table 1. The uncertain data region with pair distances smaller than 50 km has been masked.

TABLE 1. An overview of variables, their units, and ranges, which have been selected for categorized computation of partial pair correlations.

Variable	Name	Unit	Small	Medium	Large
$D$	Cell diameter	km	(20, 30)	(30, 50)	>50
$w$	Cell-average mass flux	$10^4 \text{ kg (m s)}^{-1}$	<0.3	(0.3, 0.8)	>0.8
$m$	Cell-integral mass lift	$10^{11} \text{ kg m s}^{-1}$	<9.42	(9.42, 94.2)	>94.2

cell identification. Additional calculation were made with fixed  $M$  thresholds of 3507, 2417, 1632, 1101, 759, 542, and  $402 \text{ kg m}^{-1} \text{ s}^{-1}$ . On time average, the so-defined cell populations contribute 30%–60% (in steps of 5%) to the total time-average mass lift. This makes the fixed  $M$  approach comparable to the fixed  $f$  approach. The results from the fixed  $M$  approach is qualitatively similar to the fixed  $f$  approach. Again, distribution widths are decreasing, and PCF values at short pair distances are increasing for more intense updraft cores. The rate of PCF increase, however, seems to be smaller in the  $M$  approach.

*b. Discussion of statistical frameworks*

In the following, we argue that the statistical behavior of tropical convection shares interesting similarities with that of ensembles of interacting particles in an external field like a nonuniform fluid. Even if this kind of considerations is a simplification, insights about the behavior of updraft ensembles and its interaction with the environmental forcing can be inferred. The similarities are as follows:

- (i) Large-scale convergence and moisture supply to the ITCZ zone causes the development of deep moist

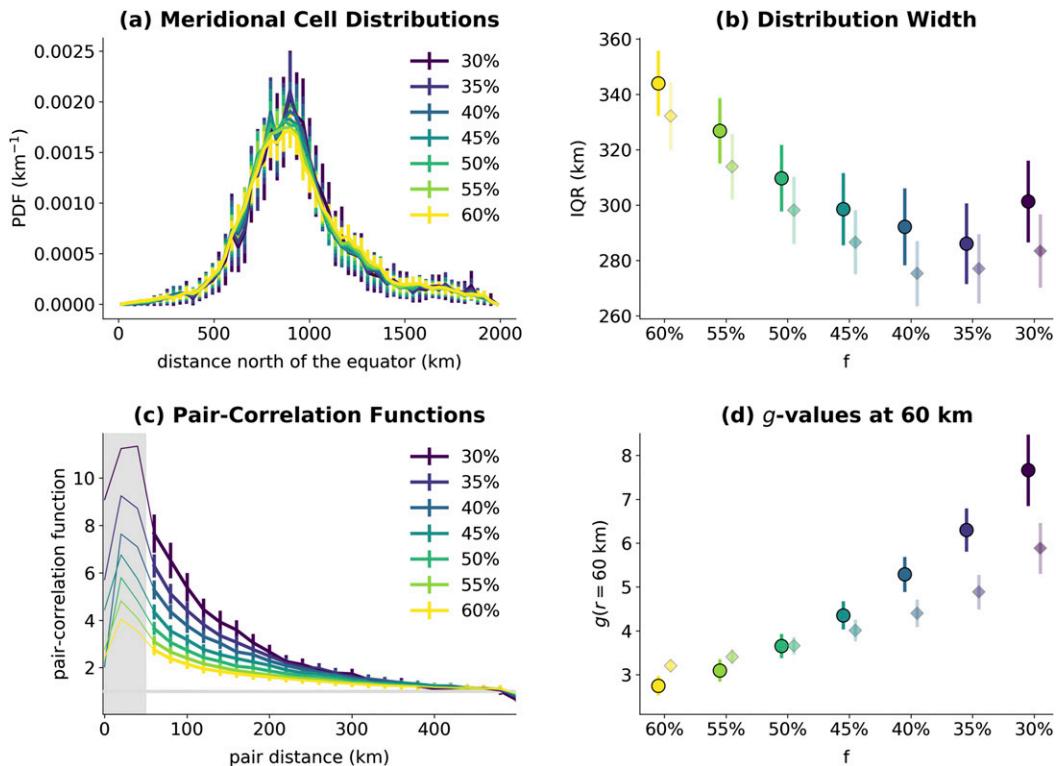


FIG. 9. Dependency of spatial cell characteristics on different threshold choices. The threshold time series  $M_{\text{thresh},f}(t)$  have been successively calculated for increasing, but fixed  $f$  values ranging from 30% (violet) to 60% (yellow). (a),(c) As in Figs. 7a and 7b. (b) The width of the meridional cell distributions measured by their interquartile range (IQR). (d) The  $g$  values at pair distances of 60 km as a function of different  $f$  values (filled circles). Furthermore, (b) and (d) present the results of additional segmentation analysis (diamonds in lighter colors) in which  $M_{\text{thresh}}$ , and not  $f$ , was set to a fixed value. The value of  $M_{\text{thresh}}$  was chosen such that the selected cell population contributes a certain time-average fraction  $f$  to the total mass lift (in contrast to an instantaneous contribution set by fixed  $f$ ).

convection in a narrow band (Waliser and Gautier 1993), similar to an external field leading to gradients in particle densities (Henderson 1992).

- (ii) Object-based approaches become increasingly important to assess the skill of cloud and precipitation simulations (Gilleland et al. 2009; Ebert et al. 2013). For these methods, a set of objects or cells is defined and cell property statistics are analyzed. A certain macroscopic configuration of convective cells can therefore be seen in analogy with a microscopic configuration of particles. If the cells carry properties in addition to their position, like size or intensity, these can be considered. Then the heterogeneous collection of cells can be interpreted as mixture of identical particles with inner degrees of freedom or as a multicomponent mixture.
- (iii) The local interactions between deep convective cells can organize tropical cell structures on mesoscales. Based on simulations, this self-aggregation behavior might result from an increased availability of moisture close to active convective cells, and furthermore influenced by convergence lines in the convective outflow and downwelling air in the far field of the convectively induced circulation (Wing et al. 2017; Holloway et al. 2017). All these processes depend on the distance to an active convective core, which reveals the analogy to a distance-dependent particle interaction potential like the Lennard–Jones potential found in simple fluids (Chandler 1987).
- (iv) The inherent nonlinear nature of developing deep moist convection limits its predictability. The resulting stochastic behavior demands a statistical description of the convective cell ensemble (Craig and Cohen 2006; Cohen and Craig 2006). Methods from statistical physics of nonuniform fluids are well suited to meet this requirement. Furthermore, there could be a multitude of physical origins leading to cell interactions. Simplified statistics could provide an effective description of the impact of underlying mechanisms. This opens routes to approximate the complex dynamical intertwining with simple means.

The major challenge lies in the identification of a conceptual model of updraft ensembles for which a standard statistical physics framework can be recovered. In a closed system, the number of cells is fixed. The spatial arrangement of cells could be either governed by an external forcing that would induce gradients in cell densities and/or by cell interactions, which could lead to range-dependent changes in the cell configuration. Both effects would act like effective forces on cells pushing the cell ensemble into a more favorable

energetic state. This kind of behavior is rather unrealistic for deep convective cells.

A more complex choice could be an open system. In addition to effective forces between cells, exchanges with a hypothetical cell reservoir could be considered. In statistical physics terms, this would mean a transition from canonical to grand canonical statistics. The exchange of cells would be then described by analogs of chemical potentials, which also might be defined to be heterogeneous and cell-distance dependent. In equilibrium, such an open system would dissipate cells with a rate similar to cell formation even for variable chemical potentials, not leading to aggregated cell clusters. Hence, temporal fluctuations of updraft cell numbers could be approximately represented in such a system, but not their spatial statistics.

A further transition to nonequilibrium systems might be inevitable to allow the convective cell population to be out of balance with external forcings or cell interactions. For example, the cell formation might be modeled by chemical potentials that depend on environmental properties, like increasing formation probabilities for more moist and unstable environments, also depending on the distance to already existing cells. In contrast, the decay of deep convective cells can be stochastically simulated by imposed lifetime distributions, which would completely decouple cell dissipation from cell formation. A Lagrangian framework could result in which an ensemble of interacting convective cells at different life cycle stages is followed in time—a method that also has potential for parameterization development.

Following above arguments, an accurate assessment of temporal cell characteristics is needed in order to model the spatial statistics of an open system of updraft cells. This concerns (i) formation probabilities in dependence of environmental factors and local cell densities and (ii) a meaningful and robust representation of cell life cycles and lifetimes. However, with the current set of simulation data with a rather large output frequency of one hour and a daily simulation restart at 0000 UTC, our analysis is restricted to temporal snapshots. Future research is needed to address the listed limitations with focus on the development of “online” object identification and tracking capabilities. Therefore, we analyze a closed system of interacting cells with simplified equilibrium statistics in the following.

### c. Simplified equilibrium statistics

In equilibrium statistics, a balance is assumed between different factors that aim to bring the system toward an organized state. The task is to find a simple representation of the balancing factors in terms of

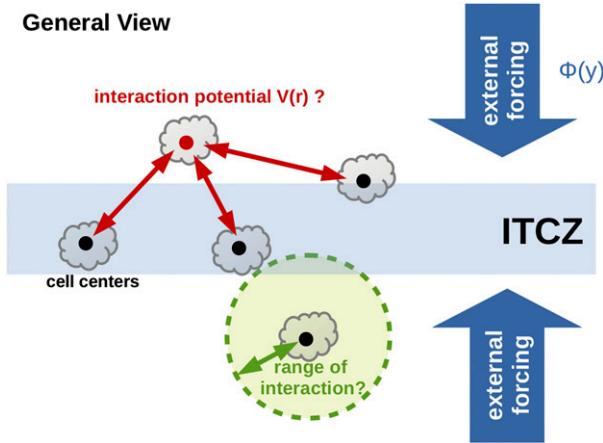


FIG. 10. Sketch of a canonical view of an interacting ensemble of tropical deep convective clouds in the ITCZ.

energy functions from which configurational characteristics similar to our updraft cell population can be reproduced. The situation is illustrated in Fig. 10: An external forcing is mainly active in the meridional direction  $\Phi = \Phi(y)$ . Cells are characterized by the centroid locations (and sizes) and interact with each other via a range-dependent interaction potential  $V = V(r)$ . All the interaction terms sum up and contribute to each particle's energy. It might also be reasonable that the interaction only occurs for a certain range involving a smaller set of neighboring cells.

For further calculations, we investigate a further simplified setup that we use in a toy model. We consider a system that experiences an external forcing of the form

$$\Phi(y) = \frac{(y - y_0)^2}{2\sigma^2}, \quad (8)$$

where  $y_0$  is the meridional offset and the parameter  $\sigma$  determines the width of the quadratic forcing potential. For a formulation of size-dependent cell interactions, we analyze the so-called ‘‘convective organization potential’’ (COP) introduced by White et al. (2018) as object-based organization metric. It was defined as

$$V_{\text{COP}}(r_{ij}; D_i, D_j) = \frac{D_i + D_j}{2r_{ij}}, \quad (9)$$

where  $r_{ij} = |\mathbf{x}_i - \mathbf{x}_j|$ . It might go far beyond the initial intention of White et al. (2018) to use COP directly as cell interaction potential, but we explore this idea and investigate the effect of such a potential on the spatial configuration. For application,  $V_{\text{COP}}$  is scaled with a negative factor to obtain an attractive potential (i.e.,  $V = cV_{\text{COP}}$ ). The scaling factor  $c$  is essentially a free

tuning parameter that scales the magnitude of the pair correlation maxima. The interaction potential is further discretized for small, medium-size, and large cells (see ranges in Table 1). In the near-distance range up to 50 km, we introduce a repulsive potential to simulate the inhibition of the finite-size cells. The resulting functions  $V$  are shown in Fig. 11a. Larger cells exhibit a more negative interaction potential that has its negative peak close to the inhibition distance. A larger inhibition energy is used for larger cells, which however does not matter for the qualitative behavior.

The functions  $\Phi$  and  $V$  are utilized in a so-called Monte Carlo algorithm (see the appendix for further details) to simulate the configuration characteristics of an equivalent canonical system. A fixed number of 50 cells has been chosen for the same domain as given in Fig. 1. This average cell number arises in the configuration with  $f = 0.5$ . We distribute the cells across the different size categories such that we have 30 small cells, 15 medium-size cells, and 5 large cells. This corresponds to a power-law size distribution with a slope of about 3. Using the parameters  $\sigma = 300$  km and  $c = -1.1$ , we obtain the average spatial cell densities  $n$  and pair correlations  $g$  presented in Figs. 11b–d. The meridional cell density is symmetric around  $y_0 = 900$  km. The width of the cell density is reduced for larger cells. This reduction shows the imprint of a stronger mean-field forcing of the larger cells. In general, a mean-field forcing arises because the external forcing pushes the cells close enough together to make the interaction more effective. This effect then leads to an additional accumulation of cells close to the peak of the external forcing. Thus, a stronger mean-field effect causes a reduction of the cell density width (seen in Figs. 11b and 11c). In addition, the mean field of cells makes the cell population more resistant against changes, like fluctuations, in the external field. The parameters  $\sigma$  and  $c$  have been selected to arrive at a width reduction similar to the one observed for the realistic ICON simulations. However, the physical origin of this effect in the ICON simulations is unclear. Could there be a similar effect in nature by which the interaction between convective cells impacts the average cell density?

Moreover, total and partial pair correlations derived from the simple toy model (shown in Fig. 11d) possess some similarities with their more realistic counterpart shown in Fig. 7b. The total pair correlation function shows values around two between 50- and 100-km distance. Due to the greater cell interactions, the partial pair correlations are larger for larger cells reaching 7 for the largest cell category. Compared to their counterparts from the ICON simulations (see Fig. 7b), the modeled pair correlation values are however significantly lower.

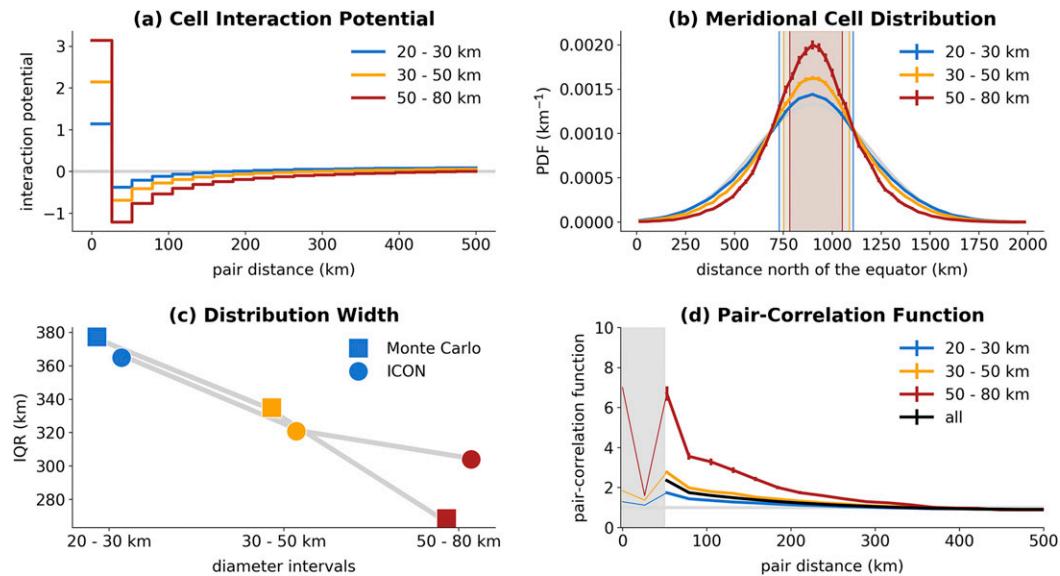


FIG. 11. Configurational statistics of the Monte Carlo simulations of a canonical ensemble of interacting cells in a heterogeneous environment: (a) cell interaction potentials, (b) meridional cell distributions, (c) distribution widths, and (d) the total and partial pair correlations for small (blue), medium-size (yellow), and large (red) cells. The plots in (b) and (d) are comparable to Figs. 7a and 7b. The distribution width in (c) has been derived from the interquartile range (IQR) of the meridional cell distributions and is plotted for ICON simulations (circles) and for the simple Monte Carlo model (squares).

Using a different setup with stronger attraction,  $\sigma = 350$  km and  $c = -1.2$ , the quantitative agreement between the pair-correlation calculations becomes much better with the disadvantage that the earlier-discussed width reduction is then overestimated (not shown). Hence, we see that several interesting configurational characteristics obtained from the ICON simulations are also described by the toy model based on equilibrium statistics, but also its limitations become apparent.

## 5. Conclusions and outlook

Deep moist convection is one of the most important phenomena of tropical meteorology (Stevens 2005). Over the oceans, large-scale convergence leads to the formation of a relatively narrow, convectively active band setting up a highly heterogeneous environment for newly forming convection (Waliser and Gautier 1993). Interactions across a vast range of scales introduce complexity in the tropical dynamics (Moncrieff et al. 2012). The scientific understanding of interactions mechanisms in the tropical belt and the impact of multiscale processes on Earth's climate is however limited (Bony et al. 2015). This motivates the present study in which we focus on pair correlations and spatial statistics of tropical deep convection in a marine environment.

Storm-resolving simulations using the ICON model with a horizontal grid spacing of 2.5 km and explicitly

resolved convection have been examined over a large domain covering the tropical Atlantic. Marine deep convective cells have been analyzed based on simulation data for one boreal summer month. Threshold-based segmentation and object analysis methods have been applied to simulated mass flux fields. Internal and environmental properties of updraft cells have been characterized. Furthermore, the spatial arrangement of cells has been investigated with the help of a pair-correlation technique, which counts the number of pair distances for a certain spatial configuration and compares this to a carefully chosen reference. Improvements over existing methods have been developed that take the heterogeneity of the convectively active region over the tropical oceans and the complex shapes and sizes of updraft cells into account.

Based on the object-based analysis of updraft cells, the following main conclusions can be formulated:

- (i) The number size distributions of simulated updraft cells possess a rather scale-free behavior in the analyzed size range between 20 and 200 km. Larger cells are more elongated than smaller cells and form in moister environments. In contrast, average CAPE and SST do not differ significantly across the cell size range.
- (ii) The relative contribution to the total domain-integrated mass lift is increasing with decreasing

cell size and a significant contribution comes from cells smaller than 20 km. As this is already below the effective model resolution, it remains unanswered at which scales the dominant contribution to the upward-directed tropical circulation is found.

- (iii) Strongest updrafts with the highest mass flux are predominant in the highest-CAPE environments. Indications are found that cells larger than 100 km have smaller cell-average mass fluxes. This might relate to changes in the efficiency of vertical mass transport due to changes in the cell-internal organization state.
- (iv) Comparing simulated cell distributions as function of environmental properties to reference distributions based on time-average conditions, it is found that temporal fluctuations in TCWV have a large impact on deviations of the spatial distribution of updraft cells.
- (v) With a pair-correlation analysis it was shown that tropical updraft cells, which contribute to half of the total mass lift, possess 3 times the probability to occur at pair distances between 50 and 100 km relative to random arrangement. This clustering effect is larger for larger and stronger updrafts indicating that dynamical feedbacks could have an impact on the configurational state of tropical convection.

An analogy between tropical cells and different approaches from statistical physics has been discussed. It was argued that an appropriate description of an open system of updraft cells would need a nonequilibrium approach in which cell formation is decoupled from cell dissipation. With the lack of reasonably accurate estimates of cell life cycle properties and their dependencies, this approach has to be postponed, and simplified equilibrium statistics based on a closed system of interacting cells, similar to a nonuniform fluid, has been explored. Using a quadratic function for external forcing and the area- and distance-dependent interaction potential of [White et al. \(2018\)](#), several configurational characteristics of the more complex cell population can be recovered, especially the larger pair correlations for bigger cells. Stronger interactions of the larger cells induce a more pronounced mean-field effect causing a reduction of meridional cell density width. An at least qualitatively similar reduction is also found in the more realistic ICON simulations, which opens room for speculations of a possible connection between the strength of convective aggregation and the average width of the convectively active zone.

*Acknowledgments.* We thank three reviewers and the editor for the helpful comments that led to significant

improvement of the manuscript. FS and MB acknowledge funding within the High Definition Clouds and Precipitation for advancing Climate Prediction [HD(CP)<sup>2</sup>] project funded by the German Ministry for Education and Research (BMBF) under respective Grants 01LK1507C and 01LK1501B. DK is supported by the Hans Ertel Center for Weather Research (HErZ), a German research network of universities, research institutions, and DWD, which is funded by the Federal Ministry of Transport and Digital Infrastructure (BMVI). FS thanks Julia Mack and Georg Craig for their advertisement of the pair-correlation method and support. In addition, FS acknowledges his colleagues from the Institute of Theoretical Physics, University Leipzig for their helpful feedback. Concerning data availability, the simulation data used in this study is stored on the supercomputer of the German Climate Computing Center (DKRZ) and can be made available from the authors upon request.

## APPENDIX

### Equilibrium Model of Interacting Updraft Cells

We start with a basic statistical physics description of a closed equilibrium system: for a canonical ensemble, the probability  $P_v^{(N)}$  to find  $N$  particles at a certain energetic state  $E_v$  is determined by the Boltzmann distribution; that is,

$$P_v^{(N)} = Q^{-1} \exp(-\beta E_v), \quad (\text{A1})$$

when the system is in equilibrium with an external heat reservoir ([Chandler 1987](#)). However, distributions can have a similar form for stationary nonequilibrium systems that are driven by material or energy fluxes ([Haken 2004](#)). The variables  $\beta$  and  $Q$  denote the inverse temperature and the canonical partition function. The total system's energy  $E_v$  for a certain configuration  $v$  is in general the sum of total kinetic energy due to particle motion and the potential energy from interaction between particles and/or an external field as well as inner degrees of freedom. For ideal gases of structureless particles, the potential energy terms vanish and only kinetic energy contributes to the distributional characteristics of the particle ensemble. In the case of classical fluids in an external field, the system's potential energy is only a function of the configuration space  $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$  for  $N$  particles, and a factorization of the distribution in momentum and configuration space applies. Hence, the configurational distribution functions obey

$$P^{(N)}(\mathbf{X}) = \hat{Q}^{-1} \exp[-E_{\text{tot}}(\mathbf{X})], \quad (\text{A2})$$

where the total potential energy  $E_{\text{tot}}$  of the particle system is measured in thermal energy units, and where  $\bar{Q}$  denotes the configurational partition sum. A certain energy configuration depends on the vector of all particle positions where the subscript  $v$  is replaced by the explicit dependence on  $2N$ -dimensional  $\mathbf{X}$  in Eq. (A2).

Next, the statistical toy model and its setup is described. We apply a so-called off-lattice Metropolis Monte Carlo algorithm that is implemented after Landau and Binder (2000). It is assumed that  $N$  cells form a canonical Gibbs ensemble and are distributed according to Boltzmann [see Eq. (A2)]. We only consider a constant cell number that is estimated from the average ICON cell characteristics. We ignore the kinetic energy contributions (i.e., only looking at potential energies and configurational space). Therefore, we define that each cell  $i$  at position  $\mathbf{x}_i$  experiences forces by an external potential from Eq. (8) and interaction forces induced by the presence of all other cells from

$$\varphi_i(\mathbf{X}) = \frac{1}{2} \sum_{j \neq i} V(|\mathbf{x}_i - \mathbf{x}_j|), \quad (\text{A3})$$

where cell-interaction potential  $V(r)$  is based on Eq. (9). The full potential energy of each cell is  $E_i = \Phi(\mathbf{x}_i) + \varphi_i(\mathbf{X})$  and the total potential energy of the cell configuration is  $E_{\text{tot}} = \sum_i E_i$ .

In the Monte Carlo algorithm, the following sequence is repeated (see, e.g., Landau and Binder 2000, p. 184): First, an arbitrary cell  $i$  is shifted by some random amount  $\Delta \mathbf{x}_i$ . Second, the total difference in potential energy  $\Delta E_{\text{tot}}$  is calculated. Finally, the resulting change in probability due to cell shift is compared to a uniformly distributed random number  $Z$  and accepted if

$$\exp(-\Delta E_{\text{tot}}) \geq Z. \quad (\text{A4})$$

The above sequence assures that cell configurations are distributed according to Eq. (A2) and is repeated for a certain number of iterations

The Monte Carlo algorithm is run in ensemble mode; that is,  $N_{\text{ens}} = 1000$  independent sets, each having  $N$  interacting cells, are run forward in time. The trial displacements  $\Delta \mathbf{x}_i$  are automatically scaled to yield an average rate of 30% at which Eq. (A4) is accepted. We perform a spinup run of the Monte Carlo ensemble using  $10^5$  trial displacements. Thereafter, 20 iterations with  $10^3$  trial displacements are carried out for each run. After each iteration, meridional cell distributions and partial pair number counts are calculated and saved. For the computation of the reference needed for the calculation of pair correlations, we run a similar setup, but with attraction terms originating from Eq. (9) switched

off. Furthermore, the external potential is estimated from the attractive runs to allow for a fair comparison. Otherwise, it cannot be ensured that the meridional cell distributions are equal for the attractive and the reference run. The results of the 20 iterations of the attractive and noninteracting runs are averaged for each ensemble member and ensemble mean values and standard deviations are computed and shown.

## REFERENCES

- Arakawa, A., 2004: The cumulus parameterization problem: Past, present, and future. *J. Climate*, **17**, 2493–2525, [https://doi.org/10.1175/1520-0442\(2004\)017<2493:RATCPP>2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017<2493:RATCPP>2.0.CO;2).
- Bengtsson, L., M. Steinheimer, P. Bechtold, and J.-F. Geleyn, 2013: A stochastic parametrization for deep convection using cellular automata. *Quart. J. Roy. Meteor. Soc.*, **139**, 1533–1543, <https://doi.org/10.1002/qj.2108>.
- Bley, S., H. Deneke, F. Senf, and L. Scheck, 2017: Metrics for the evaluation of warm convective cloud fields in a large-eddy simulation with Meteosat images. *Quart. J. Roy. Meteor. Soc.*, **143**, 2050–2060, <https://doi.org/10.1002/qj.3067>.
- Bony, S., and Coauthors, 2015: Clouds, circulation and climate sensitivity. *Nat. Geosci.*, **8**, 261–268, <https://doi.org/10.1038/ngeo2398>.
- Bretherton, C. S., P. N. Blossey, and M. Khairoutdinov, 2005: An energy-balance analysis of deep convective self-aggregation above uniform SST. *J. Atmos. Sci.*, **62**, 4273–4292, <https://doi.org/10.1175/JAS3614.1>.
- Chandler, D., 1987: *Introduction to Modern Statistical Mechanics*. Oxford University Press, 288 pp.
- Cohen, B. G., and G. C. Craig, 2006: Fluctuations in an equilibrium convective ensemble. Part II: Numerical experiments. *J. Atmos. Sci.*, **63**, 2005–2015, <https://doi.org/10.1175/JAS3710.1>.
- Craig, G. C., and B. G. Cohen, 2006: Fluctuations in an equilibrium convective ensemble. Part I: Theoretical formulation. *J. Atmos. Sci.*, **63**, 1996–2004, <https://doi.org/10.1175/JAS3709.1>.
- , and J. M. Mack, 2013: A coarsening model for self-organization of tropical convection. *J. Geophys. Res. Atmos.*, **118**, 8761–8769, <https://doi.org/10.1002/JGRD.50674>.
- Dipankar, A., B. Stevens, R. Heinze, C. Moseley, G. Zängl, M. Giorgetta, and S. Brdar, 2015: Large eddy simulation using the general circulation model icon. *J. Adv. Model. Earth Syst.*, **7**, 963–986, <https://doi.org/10.1002/2015MS000431>.
- Dorrestijn, J., D. T. Crommelin, A. P. Siebesma, H. J. J. Jonker, and C. Jakob, 2015: Stochastic parameterization of convective area fractions with a multcloud model inferred from observational data. *J. Atmos. Sci.*, **72**, 854–869, <https://doi.org/10.1175/JAS-D-14-0110.1>.
- , —, —, —, and F. Selten, 2016: Stochastic convection parameterization with Markov chains in an intermediate-complexity GCM. *J. Atmos. Sci.*, **73**, 1367–1382, <https://doi.org/10.1175/JAS-D-15-0244.1>.
- Ebert, E., and Coauthors, 2013: Progress and challenges in forecast verification. *Meteor. Appl.*, **20**, 130–139, <https://doi.org/10.1002/met.1392>.
- Garrett, T. J., I. B. Glenn, and S. K. Krueger, 2018: Thermodynamic constraints on the size distributions of tropical clouds. *J. Geophys. Res. Atmos.*, **123**, 8832–8849, <https://doi.org/10.1029/2018JD028803>.

- Gilleland, E., D. Ahijevych, B. G. Brown, B. Casati, and E. E. Ebert, 2009: Intercomparison of spatial forecast verification methods. *Wea. Forecasting*, **24**, 1416–1430, <https://doi.org/10.1175/2009WAF2222269.1>.
- Gottwald, G. A., K. Peters, and L. Davies, 2016: A data-driven method for the stochastic parametrisation of subgrid-scale tropical convective area fraction. *Quart. J. Roy. Meteor. Soc.*, **142**, 349–359, <https://doi.org/10.1002/qj.2655>.
- Haken, H., 2004: *Synergetics: Introduction and Advanced Topics*. Springer, 758 pp.
- Henderson, D., 1992: *Fundamentals of Inhomogeneous Fluids*. CRC Press, 616 pp.
- Hohenegger, C., and B. Stevens, 2013: Preconditioning deep convection with cumulus congestus. *J. Atmos. Sci.*, **70**, 448–464, <https://doi.org/10.1175/JAS-D-12-089.1>.
- Holloway, C. E., 2017: Convective aggregation in realistic convective-scale simulations. *J. Adv. Model. Earth Syst.*, **9**, 1450–1472, <https://doi.org/10.1002/2017MS000980>.
- , A. A. Wing, S. Bony, C. Muller, H. Masunaga, T. S. L'Ecuyer, D. D. Turner, and P. Zuidema, 2017: Observing convective aggregation. *Surv. Geophys.*, **38**, 1199–1236, <https://doi.org/10.1007/S10712-017-9419-1>.
- Houze, R. A., Jr., 2004: Mesoscale convective systems. *Rev. Geophys.*, **42**, RG4003, <https://doi.org/10.1029/2004RG000150>.
- Khouider, B., 2014: A coarse grained stochastic multi-type particle interacting model for tropical convection: Nearest neighbour interactions. *Commun. Math. Sci.*, **12**, 1379–1407, <https://doi.org/10.4310/CMS.2014.v12.n8.a1>.
- , J. Biello, and A. J. Majda, 2010: A stochastic multicloud model for tropical convection. *Commun. Math. Sci.*, **8**, 187–216, <https://doi.org/10.4310/CMS.2010.v8.n1.a10>.
- Kiladis, G. N., M. C. Wheeler, P. T. Haertel, K. H. Straub, and P. E. Roundy, 2009: Convectively coupled equatorial waves. *Rev. Geophys.*, **47**, RG2003, <https://doi.org/10.1029/2008RG000266>.
- Klepp, C., F. Ament, S. Bakan, L. Hirsch, and B. Stevens, 2015: NARVAL campaign report. Max Planck Institute for Meteorology Tech. Rep. 164, 218 pp.
- Klocke, D., M. Brueck, C. Hohenegger, and B. Stevens, 2017: Rediscovery of the doldrums in storm-resolving simulations over tropical Atlantic. *Nat. Geosci.*, **10**, 891–896, <https://doi.org/10.1038/S41561-017-0005-4>.
- Kumar, V. V., A. Protat, C. Jakob, and P. T. May, 2014: On the atmospheric regulation of the growth of moderate to deep cumulonimbus in a tropical environment. *J. Atmos. Sci.*, **71**, 1105–1120, <https://doi.org/10.1175/JAS-D-13-0231.1>.
- Landau, D. P., and K. Binder, 2000: *A Guide to Monte Carlo Simulations in Statistical Physics*. Cambridge University Press, 398 pp.
- Lovejoy, S., 1982: Area-perimeter relation for rain and cloud areas. *Science*, **216**, 185–187, <https://doi.org/10.1126/science.216.4542.185>.
- Machado, L. A. T., M. Desbois, and J.-P. Duvel, 1992: Structural characteristics of deep convective systems over tropical Africa and the Atlantic Ocean. *Mon. Wea. Rev.*, **120**, 392–406, [https://doi.org/10.1175/1520-0493\(1992\)120<0392:SCODCS>2.0.CO;2](https://doi.org/10.1175/1520-0493(1992)120<0392:SCODCS>2.0.CO;2).
- , J.-P. Duvel, and M. Desbois, 1993: Diurnal variations and modulation by easterly waves of the size distribution of convective cloud clusters over West Africa and the Atlantic Ocean. *Mon. Wea. Rev.*, **121**, 37–49, [https://doi.org/10.1175/1520-0493\(1993\)121<0037:DVAMBE>2.0.CO;2](https://doi.org/10.1175/1520-0493(1993)121<0037:DVAMBE>2.0.CO;2).
- Majda, A. J., and B. Khouider, 2002: Stochastic and mesoscopic models for tropical convection. *Proc. Natl. Acad. Sci. USA*, **99**, 1123–1128, <https://doi.org/10.1073/pnas.032663199>.
- Mauritsen, T., and B. Stevens, 2015: Missing iris effect as a possible cause of muted hydrological change and high climate sensitivity in models. *Nat. Geosci.*, **8**, 346–351, <https://doi.org/10.1038/ngeo2414>.
- Moncrieff, M. W., 1992: Organized convective systems: Archetypal dynamical models, mass and momentum flux theory, and parametrization. *Quart. J. Roy. Meteor. Soc.*, **118**, 819–850, <https://doi.org/10.1002/qj.49711850703>.
- , D. E. Waliser, M. J. Miller, M. A. Shapiro, G. R. Asrar, and J. Caughey, 2012: Multiscale convective organization and the YOTC virtual global field campaign. *Bull. Amer. Meteor. Soc.*, **93**, 1171–1187, <https://doi.org/10.1175/BAMS-D-11-00233.1>.
- Muller, C. J., and I. M. Held, 2012: Detailed investigation of the self-aggregation of convection in cloud-resolving simulations. *J. Atmos. Sci.*, **69**, 2551–2565, <https://doi.org/10.1175/JAS-D-11-0257.1>.
- , and D. M. Romps, 2018: Acceleration of tropical cyclogenesis by self-aggregation feedbacks. *Proc. Natl. Acad. Sci. USA*, **115**, 2930–2935, <https://doi.org/10.1073/PNAS.1719671115>.
- Nair, U. S., R. C. Weger, K. S. Kuo, and R. M. Welch, 1998: Clustering, randomness, and regularity in cloud fields: 5. The nature of regular cumulus cloud fields. *J. Geophys. Res.*, **103**, 11 363–11 380, <https://doi.org/10.1029/98JD00088>.
- Peters, K., C. Jakob, L. Davies, B. Khouider, and A. J. Majda, 2013: Stochastic behavior of tropical convection in observations and a multicloud model. *J. Atmos. Sci.*, **70**, 3556–3575, <https://doi.org/10.1175/JAS-D-13-031.1>.
- Peters, O., and J. D. Neelin, 2006: Critical phenomena in atmospheric precipitation. *Nat. Phys.*, **2**, 393–396, <https://doi.org/10.1038/nphys314>.
- , —, and S. W. Nesbitt, 2009: Mesoscale convective systems and critical clusters. *J. Atmos. Sci.*, **66**, 2913–2924, <https://doi.org/10.1175/2008JAS2761.1>.
- Ragone, F., K. Fraedrich, H. Borth, and F. Lunkeit, 2015: Coupling a minimal stochastic lattice gas model of a cloud system to an atmospheric general circulation model. *Quart. J. Roy. Meteor. Soc.*, **141**, 37–51, <https://doi.org/10.1002/qj.2331>.
- Ramirez, J. A., and R. L. Bras, 1990: Clustered or regular cumulus cloud fields: The statistical character of observed and simulated cloud fields. *J. Geophys. Res.*, **95**, 2035–2045, <https://doi.org/10.1029/JD095iD03p02035>.
- , —, and K. Emanuel, 1990: Stabilization functions of unforced cumulus clouds: Their nature and components. *J. Geophys. Res.*, **95**, 2047–2059, <https://doi.org/10.1029/JD095iD03p02047>.
- Randall, D., and G. Huffman, 1980: Stochastic model of cumulus clumping. *J. Atmos. Sci.*, **37**, 2068–2078, [https://doi.org/10.1175/1520-0469\(1980\)037<2068:ASMOCC>2.0.CO;2](https://doi.org/10.1175/1520-0469(1980)037<2068:ASMOCC>2.0.CO;2).
- Rasp, S., T. Selz, and G. C. Craig, 2018: Variability and clustering of midlatitude summertime convection: Testing the Craig and Cohen theory in a convection-permitting ensemble with stochastic boundary layer perturbations. *J. Atmos. Sci.*, **75**, 691–706, <https://doi.org/10.1175/JAS-D-17-0258.1>.
- Schneider, T., T. Bischoff, and G. H. Haug, 2014: Migrations and dynamics of the intertropical convergence zone. *Nature*, **513**, 45–53, <https://doi.org/10.1038/nature13636>.
- Senf, F., D. Klocke, and M. Brueck, 2018: Size-resolved evaluation of simulated deep tropical convection. *Mon. Wea. Rev.*, **146**, 2161–2182, <https://doi.org/10.1175/MWR-D-17-0378.1>.
- Stevens, B., 2005: Atmospheric moist convection. *Annu. Rev. Earth Planet. Sci.*, **33**, 605–643, <https://doi.org/10.1146/annurev.earth.33.092203.122658>.

- , and Coauthors, 2016: The Barbados Cloud Observatory: Anchoring investigations of clouds and circulation on the edge of the ITCZ. *Bull. Amer. Meteor. Soc.*, **97**, 787–801, <https://doi.org/10.1175/BAMS-D-14-00247.1>.
- Tan, J., C. Jakob, and T. P. Lane, 2015: The consequences of a local approach in statistical models of convection on its large-scale coherence. *J. Geophys. Res. Atmos.*, **120**, 931–944, <https://doi.org/10.1002/2014JD022680>.
- Tobin, I., S. Bony, and R. Roca, 2012: Observational evidence for relationships between the degree of aggregation of deep convection, water vapor, surface fluxes, and radiation. *J. Climate*, **25**, 6885–6904, <https://doi.org/10.1175/JCLI-D-11-00258.1>.
- Tompkins, A. M., and A. G. Semie, 2017: Organization of tropical convection in low vertical wind shears: Role of updraft entrainment. *J. Adv. Model. Earth Syst.*, **9**, 1046–1068, <https://doi.org/10.1002/2016MS000802>.
- Waliser, D. E., and C. Gautier, 1993: A satellite-derived climatology of the ITCZ. *J. Climate*, **6**, 2162–2174, [https://doi.org/10.1175/1520-0442\(1993\)006<2162:ASDCOT>2.0.CO;2](https://doi.org/10.1175/1520-0442(1993)006<2162:ASDCOT>2.0.CO;2).
- Wang, Z., 2014: Role of cumulus congestus in tropical cyclone formation in a high-resolution numerical model simulation. *J. Atmos. Sci.*, **71**, 1681–1700, <https://doi.org/10.1175/JAS-D-13-0257.1>.
- White, B. A., A. M. Buchanan, C. E. Birch, P. Stier, and K. J. Pearson, 2018: Quantifying the effects of horizontal grid length and parameterized convection on the degree of convective organization using a metric of the potential for convective interaction. *J. Atmos. Sci.*, **75**, 425–450, <https://doi.org/10.1175/JAS-D-16-0307.1>.
- Wing, A. A., K. Emanuel, C. E. Holloway, and C. Muller, 2017: Convective self-aggregation in numerical simulations: A review. *Surv. Geophys.*, **38**, 1173–1197, <https://doi.org/10.1007/S10712-017-9408-4>.
- Wood, R., and P. R. Field, 2011: The distribution of cloud horizontal sizes. *J. Climate*, **24**, 4800–4816, <https://doi.org/10.1175/2011JCLI4056.1>.
- Yuan, T., 2011: Cloud macroscopic organization: Order emerging from randomness. *Atmos. Chem. Phys.*, **11**, 7483–7490, <https://doi.org/10.5194/acp-11-7483-2011>.
- Zängl, G., D. Reinert, P. Rípodas, and M. Baldauf, 2015: The ICON (Icosahedral Non-Hydrostatic) modelling framework of DWD and MPI-m: Description of the non-hydrostatic dynamical core. *Quart. J. Roy. Meteor. Soc.*, **141**, 563–579, <https://doi.org/10.1002/qj.2378>.