Check for updates

JAMES | Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE
10.1029/2022MS003231

Special Section:
Using radiative-convective equilibrium to understand convective organization, clouds, and tropical climate

A Dimensionless Parameter for Predicting Convective Self-Aggregation Onset in a Stochastic Reaction-Diffusion Model of Tropical Radiative-Convective Equilibrium

Giovanni Biagioli[1,2] **and Adrian Mark Tompkins**[2]

[1]University of Trieste, Trieste, Italy, [2]Abdus Salam International Center for Theoretical Physics (ICTP), Trieste, Italy

Abstract We introduce a minimal stochastic lattice model for the column relative humidity (R) in the tropics, which incorporates convective moistening, horizontal transport and subsidence drying. The probability of convection occurring in a location increases with R, based on Tropical Rainfall Measuring Mission observations, providing a positive feedback that could lead to aggregation. We show that the simple model reproduces many aspects of full-physics cloud resolving model experiments. Depending on model parameter settings and domain size and resolution choices, it can produce both random and aggregated equilibrium states. Clustering occurs more readily with larger domains and coarser resolutions, in agreement with full-physics models. Using dimensional arguments and fits from empirical data, we derive a dimensionless parameter which we call the aggregation number, N_{ag}, that predicts whether a specific model and experiment setup will result in an aggregated or random state. The parameter includes the moistening feedback strength, the horizontal moisture transport efficiency, the subsidence timescale, the domain size and spatial resolution. Using large ensembles of experiments, we show that the transition between random and aggregated states occurs at a critical value of N_{ag}. We argue that N_{ag} could help to understand the differences in aggregation states between full-physics, cloud resolving models, which show little consensus about the robustness of self-organized patterns, whose emergence appears to be sensitive to the model setup, physics and parameterizations.

Plain Language Summary Dynamical models of convection including detailed representation of microphysics and radiative processes can simulate idealized states of radiative-convective equilibrium. Sometimes in these simulations the tropical atmosphere shows a behavior whereby all convective storms evolve from a state where they occur randomly in space, to one where they are clustered together, with large areas of the simulation domain free from convection. This phenomenon is important to understand due to its implications for climate sensitivity, but the problem is that these full-physics models do not agree on when or how clustering occurs. We therefore introduce a much simpler stochastic model of the tropical atmosphere, whose minimal representation of the physics is nevertheless adequate to reproduce much of the behavior of the full-physics models. Convection can aggregate or remain random, depending on the model parameter settings as well as the domain size and resolution chosen. The simplicity of the model allows us to derive a dimensionless parameter, which we call the aggregation number and incorporates all the model parameters and the domain size and resolution. Convection is found to aggregate when this parameter falls below a critical threshold. We suggest that this parameter can help to explain the differences between the full-physics models.

Key Points:

- A stochastic reaction-diffusion model of the tropics can simulate aggregated and random convective states
- The model produces a similar sensitivity to domain size and resolution as full-physics models
- A dimensionless parameter called the aggregation number combines model and experiment configuration parameters to predict aggregation onset

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

A. M. Tompkins,
tompkins@ictp.it

Citation:

Biagioli, G., & Tompkins, A. M. (2023). A dimensionless parameter for predicting convective self-aggregation onset in a stochastic reaction-diffusion model of tropical radiative-convective equilibrium. *Journal of Advances in Modeling Earth Systems*, *15*, e2022MS003231. https://doi.org/10.1029/2022MS003231

Received 3 JUN 2022
Accepted 29 APR 2023

Author Contributions:

Conceptualization: Giovanni Biagioli, Adrian Mark Tompkins
Data curation: Giovanni Biagioli
Formal analysis: Giovanni Biagioli

© 2023 The Authors. Journal of Advances in Modeling Earth Systems published by Wiley Periodicals LLC on behalf of American Geophysical Union. This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

1. Introduction

When computing resources first permitted the multi-week, convection-resolving simulations of radiative-convective equilibrium (RCE), an interesting behavior was noted. After an initial period in which deep convection was randomly distributed throughout the domain of the cloud resolving model (CRM), a transition occurred to a state in which the convection was clustered in one part of the domain, surrounded by dry, clear-sky regions. This behavior, termed self-aggregation, was observed to occur in both 2D (Held et al., 1993) and 3D (Tompkins & Craig, 1998a) modeling frameworks, and can be interpreted as a phase transition in moist convective systems.

The drier mean atmosphere resulting from the self-aggregation (Bretherton et al., 2005) implies that the occurrence and strength of the phenomenon may have implications for tropical climate sensitivity (Mauritsen & Stevens, 2015). Indeed, it has been proposed that the aggregation can act as a regulator of tropical climate, as the associated drying allows the system to efficiently cool through reduced water vapor absorption (Wing et al., 2017). Emanuel et al. (2014) hypothesized that the self-aggregated state could be the preferred equilibrium

Investigation: Giovanni Biagioli
Methodology: Giovanni Biagioli, Adrian Mark Tompkins
Project Administration: Adrian Mark Tompkins
Software: Giovanni Biagioli, Adrian Mark Tompkins
Supervision: Adrian Mark Tompkins
Validation: Giovanni Biagioli
Visualization: Giovanni Biagioli
Writing – original draft: Giovanni Biagioli, Adrian Mark Tompkins
Writing – review & editing: Giovanni Biagioli, Adrian Mark Tompkins

state of tropical convection under warm sea-surface temperatures (SSTs), which implies that global warming may lead to the tropics switching to this regime in future climates. However, much uncertainty remains about the temperature dependence of self-aggregation, as its occurrence has been detected in models at temperatures well below the current tropical SSTs (Abbot, 2014; Holloway & Woolnough, 2016; Wing & Cronin, 2016; Wing et al., 2020).

Many studies have used a range of mechanism denial/suppression experiments to demonstrate that a range of diabatic feedback processes all contribute to convective aggregation to various degrees. This includes the feedback between convection and water vapor, radiative feedbacks with the water vapor field and clouds, and contributions from surface fluxes (Held et al., 1993; Holloway & Woolnough, 2016; Muller & Bony, 2015; Stephens et al., 2008; Tompkins, 2001; Tompkins & Craig, 1998a; Wing & Emanuel, 2014; Wing et al., 2017). The role of advective processes has also been pointed out by several studies, in particular the development of a radiatively driven, shallow circulation transporting moist static energy (MSE) upgradient has been recognized to be peculiar of aggregated runs and can help sustain the clustering (Bretherton et al., 2005; Muller & Bony, 2015; Muller & Held, 2012). Nonetheless, there is still debate in the literature about which feedback is the most important in driving the aggregation. Moreover, most of the feedbacks also change in strength and there is not a single leading feedback throughout the evolution of the aggregation process (Wing & Emanuel, 2014), to the point that the physical mechanisms that trigger the clustering of convection have been found to differ from those that maintain it once established (Muller & Bony, 2015; Muller & Held, 2012). It is also possible that some processes first favor and then even oppose (or vice versa) the organization, that is, the feedbacks change in sign between the pre- and post-aggregated states (Tompkins & Semie, 2021; Wing & Emanuel, 2014).

Despite the fact that models agree on the general role that the diabatic processes play in driving self-aggregation, there is much disparity between models regarding their aggregated states as measured by a range of aggregation indices, the sensitivity of these indices to underlying SST, and in some case even whether models aggregate at all for a given experimental framework (Wing et al., 2020). This is due to the sensitivity of aggregation to the specifics of the CRM's subgrid parameterizations, such as the details of the microphysical schemes and subgrid-scale turbulence and mixing schemes used (Huang & Wu, 2022; Shi & Fan, 2021; Tompkins & Semie, 2017). In addition, the aggregation is sensitive to the details of the experimental framework, with aggregation delayed by the use of an interactive lower boundary (Bretherton et al., 2005; Hohenegger & Stevens, 2016; Shamekh et al., 2020; Tompkins & Semie, 2021) and prevented altogether when small domains and finer grid resolutions are employed (Muller & Held, 2012), indicating a lack of numerical convergence regarding aggregation.

To better understand the differences between cloud resolving models, it is beneficial to examine simpler models of the atmosphere which may mimic aspects of the full-physics systems. Within this framework, the occurrence of self-aggregation has long been regarded as caused by an instability of the spatially uniform RCE state of tropical convection (Bretherton et al., 2005; Emanuel et al., 2014; Shi & Fan, 2021). For example, Raymond and Zeng (2000) coupled two single column models using the weak temperature gradient (WTG) approximation later formalized by Sobel et al. (2001), and showed instabilities representing the onset of aggregation. Single-column experiments by Sobel et al. (2007) subsequently demonstrated that, under certain circumstances, two stable equilibria can coexist whose occurrence depends on the initial moisture profile. These results provide a direct link to CRM findings to the extent that the non-convective equilibrium may correspond to the dry patch of a self-organized climate and the moist equilibrium to the deep convective areas. Emanuel et al. (2014) produced a two layer model and showed how an instability due to infrared radiation could occur in warmer, moister atmospheres.

In contrast to these column models, other models are spatially explicit in that they represent the horizontal variability of convection and/or humidity on spatial grids, but simplify the physics of the system to a highly conceptualized representation. The first such model was a 2D stochastic representation of cumulus self-aggregation of Randall and Huffman (1980). More recently, Böing (2016), Haerter (2019) have introduced 2D horizontal idealized models of convective cold pools to show how they could contribute to clustering while Yang (2021) used a 1D linear shallow water model to investigate aggregation.

Some simple, very idealized models have been equipped with stochastic components to make such models more relevant for the real tropical atmosphere. In the more general context of tropical convection modeling, stochastic approaches have been used for parameterization purposes and to investigate its impacts on atmospheric spatiotemporal variability (Khouider, 2014; Khouider et al., 2003, 2010; Lin & Neelin, 2000, 2002; Majda & Khouider, 2002; Plant & Craig, 2008). Another recent research strand has focused on the analogy between the

19422466, 2023, 5, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022MS003231 by CochraneItalia, Wiley Online Library on [16/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

19422466, 2023, 5, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022MS003231 by CochraneItalia, Wiley Online Library on [16/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

onset of strong precipitation as a function of column water vapor in the tropics and the theory of continuous phase transitions, and to this aim stochastic models have been developed that are able to capture the salient features of tropical convection (Hottovy & Stechmann, 2015a, 2015b; Stechmann & Neelin, 2011). In particular, the model by Hottovy and Stechmann (2015b) considers a single prognostic equation for column water vapor supplemented with both space- and time-uncorrelated noise (white noise) field. Ahmed and Neelin (2019) further elaborated on this model and introduced a more complex formulation in terms of the representation of physics and the addition of a temporal red noise to include the variability induced by the absent processes, which is assumed to possess a temporal auto-correlation. Their model is capable of closely reproducing the observed statistics of tropical precipitation and can also sustain self-aggregation if the radiative feedbacks are strong enough and the amplitude of the stochastic forcing is properly reduced. Their moisture budget equation is of the advection-reaction-diffusion type, and it is in fact not uncommon to study the intriguing convective self-organization through simple differential problems that can be even more minimal in terms of physical complexity, often consisting of a single reaction-diffusion equation (e.g., Li, 2021; Shi & Fan, 2021; Windmiller & Craig, 2019).

In another spatially explicit approach of the reaction-diffusion type, a 2D model with a representation of physics controlling the vertically integrated tropical water vapor (W) budget was introduced by Craig and Mack (2013) (CM13 hereafter) to examine aggregation of convection. The prognostic equation for W consists of three terms. Convection locally moistens the atmosphere and the moisture is then advected laterally using a diffusive mixing approximation, while the troposphere is subject to subsidence drying. Clustering of convection is driven by a function that dictates greater convective moistening where the atmosphere is more humid, basing this positive feedback on the observed exponential increase of tropical precipitation as a function of W documented in Tropical Rainfall Measuring Mission (TRMM) precipitation data (Bretherton et al., 2004; Rushley et al., 2018). The model has three physical parameters that describe the efficiency of the horizontal vapor transport, the subsidence drying timescale and a parameter that determines the strength of the convective-water vapor feedback. Starting from a homogeneous state with random fluctuations, CM13 found that the model reproduces the phenomenon of self-aggregation, termed "coarsening," for most parameter ranges, except for values of the convection-water vapor feedback strength far below those reported by Bretherton et al. (2004), Rushley et al. (2018). They described the conditions for instability in terms of the subsidence timescale and the feedback parameter, but without considering the impact of the horizontal transport, which we will show is also relevant. The theory also did not incorporate the domain size or resolution and thus could not explain the sensitivity to these simulation parameters that have been documented in the full-physics CRMs.

In this paper, we adapt the model of CM13 to investigate convective aggregation onset in CRM-like experiments which resolve convection and use spatially limited domains of size $\mathcal{O}\left(10^2 - 10^3 \text{ km}\right)$ with periodic boundaries. In order to do this, the model presented here differs from CM13 in several respects, the key difference regarding the spatial resolution employed. The water budget equation of CM13 was integrated on a 40 km climate-model sized grid in which convection was treated deterministically. This means that the convective moistening term operated continuously in all locations, since the coarse resolution implied convection could be always occurring somewhere within a cell at a rate determined by the cell's humidity. Here we instead use a cloud resolving grid resolution of $\mathcal{O}$(1 km) and treat the spatiotemporal occurrence of convective activity as a stochastic binary process, either on or off. Using a weighted random variable, a subset of cells will be selected at each time step to develop new convective activity to supplement the existing convection. This stochastic approach with its dependence on grid spacing will allow us to incorporate both the model resolution and the simulation domain size into the theory for aggregation onset.

We will demonstrate that the model presented here can produce equilibrium states of both random or aggregated convection, depending on the exact parameter settings and experiment set up employed, with the transition to organization occurring at model parameter values that are reasonable approximations of the present tropical atmosphere. We will use dimensional arguments to derive a dimensionless parameter that is a function of the three model processes (horizontal transport, subsidence and convective moistening) as well as the domain size and resolution, and successfully predicts which simulation configurations result in convective aggregation.

2. Methods

2.1. Description of the Model

The model presented here is closely related to the model presented in CM13. The model of CM13 introduces a prognostic partial differential equation for the vertically integrated water budget W in the tropical troposphere,

19422466, 2023, 5, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022MS003231 by CochraneItalia, Wiley Online Library on [16/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

whose derivation follows from arguments by Yanai et al. (1973). Instead of using W as a prognostic variable, we write the model in terms of the column total water relative humidity $R = R(\mathbf{x}, t)$, where $\mathbf{x}$ denotes the spatial coordinates and t is time. R is defined as the sum of the density-weighted, column-integrated ice water (q_i), liquid water (q_l) and water vapor (q_v) specific humidities normalized by the column-integrated saturation value (q_s),

$$R = \frac{\int \rho(q_v + q_l + q_i) dz}{\int \rho q_s dz}. \tag{1}$$

Changes in R primarily reflect those in the vertically integrated water content, owing to the horizontal temperature gradients being small in the tropical troposphere, the WTG approximation (Sobel et al., 2001). R is dimensionless and less sensitive to temperature than W, which shows exponential dependence on temperature through the Clausius-Clapeyron relationship. We also assume time-invariance of temperature, so that the system consists of a single prognostic equation for R.

In our model, as in CM13, three processes affect R:

$$\frac{\partial R}{\partial t} = R_{\text{conv}} + R_{\text{trans}} + R_{\text{sub}}. \tag{2}$$

The model assumes that column relative humidity is rapidly increased in locations where deep convection occurs (R_{conv}), these moisture sources are subsequently redistributed horizontally by lateral vapor transport (R_{trans}), while subsidence (R_{sub}), that balances the convective mass flux, dries the atmosphere.

Considering the horizontal transport first, we follow CM13 in approximating the lateral transport of water vapor by a down-gradient diffusion, $R_{\text{trans}} = K\nabla^2 R$, which is parameterized using a simple fixed value for the diffusivity K. This is an oversimplification as it neglects the enhanced mixing expected near updrafts compared to stably stratified subsidence regions (Tompkins & Semie, 2017) and also neglects temporal variations. However, Windmiller and Craig (2019) demonstrated that a diffusive treatment of lateral transport can reasonably represent the evolution of water vapor at least in the early stages of self-aggregation. The use of a down-gradient mixing implies horizontal transport always acts to reduce spatial variance of R and thus opposes aggregation onset.

The treatment of subsidence also follows CM13, as subsidence is modeled as a relaxation process toward a completely dry atmosphere, thus $R_{\text{sub}} = -R/\tau_{\text{sub}}$, with a characteristic timescale τ_{sub} set to be uniform throughout the domain. In assuming this, we have implicitly hypothesized that the adjustment due to gravity wave propagation from convective events is instantaneous, which is still reasonable since this propagation is fast relative to other processes and subsidence is a superposition of drying from all convective events. The relaxation representation of subsidence implies that this term will also always reduce spatial variance of R, since it ultimately leads to a homogeneous dry atmosphere.

The interesting behavior of the model, namely, its ability to represent convection in both random and aggregated configurations, derives from the specification of the convective moistening term. The key novelty with respect to CM13 is the adoption of a stochastic treatment for the convective moistening term, so that the model lends itself to be run at convection resolving resolutions. This is particularly suited to mimic the typical CRM experimental setup, with the modified governing equation integrated on a 2D mesh of grid cells using a $\Delta x \sim \Delta y \sim \mathcal{O}(1\text{ km})$ horizontal resolution and $L = \mathcal{O}\left(10^2 - 10^3\text{ km}\right)$ domain sizes. In some previous idealized modeling studies (Hottovy & Stechmann, 2015b), the grid spacing was chosen to be the minimum scale of a single tropical deep convective cell, but we will not similarly constrain the resolution here.

Convective moistening is modeled as a fast relaxation with characteristic time $\tau_c \ll \tau_{\text{sub}}$ toward R_c, the total water relative humidity in convective columns, which exceeds unity due to the detrainment of cloud condensate: $R_{\text{conv}} = \mathcal{I}_c(R_c - R)/\tau_c$. This term contains a diagnostic indicator random variable $\mathcal{I}_c = \mathcal{I}_c(\mathbf{x}, t)$ that maps the domain to convecting ($\mathcal{I}_c(\mathbf{x}, t) = 1$) and non-convecting (0) locations according to a non-uniform, humidity-dependent probability function $p_c(R(t))$. Thus, unlike the horizontal transport and subsidence terms that operate continuously in all cells, the moistening only occurs in locations occupied by convective updrafts, where $\mathcal{I}_c = 1$. In all other cells $\mathcal{I}_c = 0$ and the convective moisture source is zero. The moistening term represents the sole stochastic contribution to the humidity budget. The details of the sampling procedure are outlined below. Finally, no large-scale dynamical forcing is imposed and the Coriolis effect is excluded. Diurnal and seasonal cycle representations are also omitted.

19422466, 2023, 5, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022MS003231 by CochraneItalia, Wiley Online Library on [16/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The continuous form of the budget equation for R is thus given by

$$\frac{\partial R}{\partial t} = \frac{(R_c - R)}{\tau_c} \mathcal{I}_c + K\nabla^2 R - \frac{R}{\tau_{\text{sub}}}, \tag{3}$$

where the first term on the RHS represents the humidity source associated with convection, the second term expresses the lateral moisture transport, while the third term describes the drying action of subsidence. To mimic the behavior of the full-physics models, we will solve this equation on a square discretized grid with periodic boundary conditions of equal resolution in the x and y directions.

To initialize the model a specific number $N_c(t = 0)$ of cells are chosen at random to be convective, according to a weighted probability which will be discussed below. There is a memory for convection, and on subsequent time steps, each convective cell has a fixed probability of dying, to give an average convective duration of 30 min. Locations are then chosen for initiating new convective events, to ensure that the desired total population size $N_c(t)$, imposed as an external constraint, is maintained. We require the time-averaged value of convective population size $\left(\overline{N_c}\right)$, and thus also the convective fraction, be prescribed by a simple mass conservation argument proposed by Tompkins and Craig (1998a). Specifically,

$$\overline{N_c} = N_{\text{xy}} \frac{|w_{\text{sub}}|}{w_c} = N_{xy} \frac{h}{\tau_{\text{sub}} w_c} = \left(\frac{L}{\Delta x}\right)^2 \frac{h}{\tau_{\text{sub}} w_c}, \tag{4}$$

where N_{xy} is the total number of grid points in the computational domain, w_{sub} and w_c indicate the subsidence and the convective updraft vertical velocities, respectively, h is the approximate depth of the troposphere. We emphasize that, as $\overline{N_c}$ is a function of the domain size, horizontal resolution and imposed subsidence rate, it remains invariant throughout simulations.

Rather than imposing a constant convective fraction (i.e., $N_c(t) = \overline{N_c}$), we assume that the temporal variation of $N_c(t)$ follows a Poisson distribution with parameter $\overline{N_c}$, subjected to a running mean with a window length equal to the assumed convective lifetime of 30 min, to ensure that the convective birth-rate on any time step is (nearly always) zero or positive. Occasional negative rates are subject to a correction procedure to ensure the time-averaged convective population is precisely $\overline{N_c}$ in all experiments. We note that increasing (decreasing) the average convective lifetime would make self-aggregation more (less) likely, but this aspect of the model sensitivity is not investigated further.

When choosing the locations for new convective cells, all cells in the domain with $\mathcal{I}_c = 0$ are sampled without replacement (i.e., no two convective cells can occupy the same location), using a non-uniform probability distribution $p_c(R(t))$ that depends on the column relative humidity R at current time t. We base the probability of a cell being chosen as convective on the observations of the non-linear moisture-precipitation relationship by Bretherton et al. (2004) and Rushley et al. (2018) using TRMM data, which gives surface precipitation P increasing exponentially with R:

$$P(R) = P_0 e^{a_d R}, \tag{5}$$

where P_0 and a_d are constant coefficients quantifying the horizontal mean RCE rain rate and the sensitivity of precipitation to column humidity, respectively. This form was also confirmed independently by Holloway and Neelin (2010).

An increase in precipitation could result from an increase in the occurrence of convective events and/or an increase in precipitation intensity per event. To allow us to use Equation 5 to define the probabilities $p_c(R)$, we make the assumption that the increase in precipitation rate as a function of R is solely due to the more frequent occurrence of convection in moister atmospheres. In other words, we assume that the rainfall intensity per event is constant. This assumption, also adopted by previous idealized modeling studies (Hottovy & Stechmann, 2015a; Stechmann & Neelin, 2011, 2014), seems to be reasonable according to a recent analysis of (mid-latitude) station data by Yano and Manzato (2022) and neglects the limited contribution of increased humidity to increased precipitation efficiency (Narsey et al., 2019). It would be straightforward to include a relationship for this latter effect, but while it would change the critical threshold for aggregation onset, it would not affect the conclusions of the work and is omitted for simplicity. Making the above assumption means we can apply Equation 5 to give the probability of convective occurrence associated with the random sampling as

$$p_c(R_{j,k}) = C(t)e^{a_d R_{j,k}}, \tag{6}$$

where the subscripts $j, k = 1, \ldots, \frac{L}{\Delta x}$ refer to the spatial grid and the relationship is normalized each timestep by $C(t)$ to ensure the sum of probabilities across all convective-free cells is unity. CM13 modified the fit of Bretherton et al. (2004) by subtracting unity from the exponential (i.e., Equation 5 would become $P(R) = P_0\left(e^{a_d R} - 1\right)$) to give the limit of zero precipitation when $R = 0$. This assumption introduces a second dry equilibrium state as convection can not remoisten a completely dry domain where $R = 0$ everywhere. We instead retain the original form of Bretherton et al. (2004) to allow our model to trigger convection in dry columns, and our closure of specifying $\overline{N_c}$ thus differs from that adopted by CM13, who instead constrained the domain-mean accumulated precipitation by assuming that the latent heat release balances the radiative cooling above the boundary layer. In practice, the difference to CM13 in the *P*-*R* relationship has very limited effect at the values of R found in the domain.

Through Equation 6 the model incorporates the observed correlation between convection and water vapor. The convective term, unlike the transport and subsidence terms, can act to increase or decrease the spatial variance of R, depending on the spatial distribution of R itself and the choice of convective locations. If convection occurs in the moistest regions of the domain the impact is to increase spatial variance, possibly leading to aggregation. Thus the model will be sensitive to a_d, which describes the correlation between convection and R. CRM studies associate this correlation with the moisture-memory feedback (Muller & Bony, 2015), which has been shown to be key in organizing convection (Grabowski & Moncrieff, 2004; Tompkins, 2001). Deep convection moistens its local environment, while enhanced humidity encourages persistence of convection owing to reduced cooling due to entrainment mixing, thereby yielding stronger convection (Derbyshire et al., 2004). In addition, CRM studies have also acknowledged the important role of the radiative and surface flux contribution to the localization of convection (Stephens et al., 2008; Tompkins & Craig, 1998a; Wing & Emanuel, 2014). This primarily occurs through enhanced longwave heating and shortwave absorption in the moist areas due to the high infrared opacity and the presence of deep convective cloudiness and through increased fluxes due to the strengthening of the storm-induced gustiness. These effects jointly act to amplify the positive MSE anomalies in the moist patches. In this perspective, the parameter a_d can be thought of as representing the net positive feedback of all diabatic processes driving aggregation.

In summary, the net effect of convection as represented by the model is to locally moisten around the sites of convection through column saturation and detrainment of cloud condensate and to dry the far-field through compensating subsidence. Note that this treatment is more simplistic than the stochastic model of Ahmed and Neelin (2019) who modeled microphysics processes directly, and also included a treatment of horizontal vapor divergence and large-scale dynamics in addition to local diffusive mixing.

2.2. Numerical Solution

Adequate numerical treatment is needed to ensure the results are not time step sensitive. We use an implicit solution technique to ensure stability. The use of operator splitting schemes (e.g., Hundsdorfer & Verwer, 2007) was invoked, with the RHS of Equation 3 additively decomposed into two terms (the first including subsidence and horizontal transport, the second only convection), and the adoption of a second-order accurate Strang-type strategy (Strang, 1968). To solve the differential problem corresponding to the first term of the decomposition, we employed second-order finite difference approximations in space and we developed a modified version of the classical Peaceman-Rachford Alternating Direction Implicit method in time (Peaceman & Rachford, 1955), whereas the analytical solution was derived for the problem with the convection term solely. A full description of the numerical solver and a number of idealized experiments to demonstrate numerical robustness is contained in the Supporting Information S1.

2.3. Choice of the Model Parameters, Constants and Setup

In the paper large ensembles of several thousand simulations are carried out in order to find the combination of model parameters and simulation configuration that lead to random or aggregated convection. All experiments are run for at least 120 days (with some extended to 180 days), a period long enough such that there is a long-term steady state of variables indicating equilibrium has been achieved. As a metric of clustering or

19422466, 2023, 5, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022MS003231 by CochraneItalia, Wiley Online Library on [16/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 1
Parameters (Default and Ranges) Used in the Simulations

	Default value	Range
K ($m^2\ s^{-1}$)	10^4	$10^3 - 4 \times 10^4$
τ_{sub} (days)	16	5–40
a_d	14.72	10–30
	16.12	
L (km)	300	200–1,000
Δx (km)	2	0.5–4

random convection we mostly use the domain spatial standard deviation of the column relative humidity R averaged over the last 20 days of simulation $(\overline{\sigma}_{R,20})$. Low values of $\overline{\sigma}_{R,20}$ indicate random convection while high values indicate convection is aggregated. In addition we also use the I_{org} parameter of organization described in the appendix of Tompkins and Semie (2017), a more quantitative metric of aggregation as it allows one to classify scenes as random or aggregated (or even regular, where inter-convective spacing is larger than expected from a random distribution).

Details of these ensembles with the default values and ranges used for each of the parameters are reviewed below and summarized in Table 1. The subsidence timescale τ_{sub} is derived assuming that in subsidence areas, in the absence of large-scale convergence, subsidence heating approximately balances the net radiative cooling, Q_{rad}:

$$\tau_{sub} = \frac{h\frac{d\theta}{dz}}{eQ_{rad}}, \tag{7}$$

which, inserting characteristic values for the depth of the free troposphere ($h \approx 10$ km), the mean environmental lapse rate of potential temperature $\frac{d\theta}{dz} \approx 6.5$ K km^{-1} and the vertically integrated radiative cooling rate $Q_{rad} \approx 1.5$ K day^{-1}, gives $\tau_{sub} = 16$ days, with the ensembles spanning 5–40 days. We note that this timescale is much longer than that used in CM13 of 2 days.

Analysis of TRMM data by Rushley et al. (2018) gave the convection sensitivity factor a_d values of 14.72 and 16.12, depending on the TRMM retrieval version, and our ensembles span values of 10–30. The default values of a_d used here refer to the daily-mean precipitation-R relationships, but the same exponential law Equation 5 applies to the monthly-mean relationship, with a coefficient of 11.4 that was estimated by Bretherton et al. (2004) and employed by CM13 in their simulations. We note that this latter value is also included in our experimentation (see Table 1).

A reasonable estimate for the horizontal moisture transport efficiency K can be calculated by defining it as a function of characteristic length and velocity scales, ℓ_0 and v_0, associated with convective motions:

$$K = \epsilon \ell_0 v_0, \tag{8}$$

where ϵ is a eddy-size related coefficient set to $\epsilon = 0.1$. Typical scales are the free tropospheric depth, $\ell_0 = 10$ km, and $v_0 = w_c = 10$ m s^{-1} (updraft velocity observed in convective cores), implying that reliable values for K are on the order of 10^4 m^2 s^{-1}, but our experiments evaluate values from 10^3 to 4×10^4 m^2 s^{-1}.

The convective moistening characteristic time τ_c is set to a very fast timescale of 1 min to lead to almost instantaneous saturation. We did find some sensitivity of the model to the choice of τ_c but using slower adjustment times did not change the conclusions derived from the model. We set R_c accounting for column cloud water detrainment using estimates from CRM simulations to give $R_c = 1.05$.

To keep the total simulation size tractable while exploring the parameter space, we constructed series of ensembles of $\mathcal{O}(1000)$ members that investigate two parameters while keeping others fixed. Two ensembles of experiments using a domain size $L = 300$ km and resolution $\Delta x = 2$ km cover combinations of τ_{sub} and K (results shown in Figure 10), and K and a_d (Figure 11). An additional ensemble of experiments employed a limited range of fixed values for the three model key parameters, combined with a range of domain sizes (L = 200, 300, 400, 1,000 km, Figure 4) and spatial resolutions (Δx = 0.5, 1, 1.5, 2, 4 km, Figure 6), see also Figures S6 and S7 in Supporting Information S1. To construct the final analysis exploring the five-parameter space (K, τ_{sub}, a_d, L, Δx), a sub-sampled ensemble of 1,160 members was used.

Simulations are initialized with R field assumed to be completely horizontally homogeneous with $R = 0.8$ everywhere. There is sensitivity to the initial conditions with the model exhibiting a weak hysteresis, but this is not investigated here. No perturbations are imposed on the initial R distribution since stochasticity is already accounted for in the model through the convective location function. Periodic lateral boundary conditions are applied.

19422466, 2023, 5, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022MS003231 by CochraneItalia, Wiley Online Library on [16/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

19422466, 2023, 5, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022MS003231 by CochraneItalia, Wiley Online Library on [16/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Table 2
Summary of the Simulations of Figure 3

Simulation name	Parameters		
	K (m^2 s^{-1})	τ_{sub} (days)	a_d
CTRL	10^4	16	14.72
0.5K	5×10^3	16	14.72
τ_{sub}10	10^4	10	14.72
a_d16.12	10^4	16	16.12

3. Results

Before analyzing the large ensembles it is useful to demonstrate how the model can produce both random and aggregated convective states depending on the parameter settings chosen. We start by showing two experiments, one with the default values of K, a_d, and τ_{sub} (CTRL, see Table 2) and the second with a reduced value of the horizontal transport efficiency (0.5K). Five time slice panels show the evolution of the horizontal R field (Figure 1 and Movie S1). In the default experiment (upper panels), the convective sources remain randomly distributed throughout the domain, even on day 180, and the domain-mean R remains moist. In contrast, halving the strength of the lateral transport of water vapor (lower panels) causes the model to evolve toward a dramatically different state. After an initial period of random convection, the variability of R in the domain increases during the transition toward a spatially organized atmospheric state, characterized by the emergence of a single, almost circular, intensely convecting area surrounded by a dry environment. Close examination shows many examples of localized moist cells caused by the stochastic convective selection in those locations. Once aggregation has established, the dry patch is very rarely disrupted by moistening processes from local sources, but it is not guaranteed that deep convective events necessarily trigger in the wettest cells and occasionally drier cells are chosen. This behavior would be missing from a deterministic formulation of the model. These experiments highlight the ability of the simple model to mimic both random and aggregated equilibrium states, with results resembling those yielded by more complex, full-physics CRMs, at least from a qualitative point of view (e.g., Bretherton et al., 2005; Muller & Held, 2012).

Also in accordance with the full-physics CRMs (e.g., Bretherton et al., 2005; Wing & Emanuel, 2014), the mean state is much drier in the aggregated simulation relative to the random case, and column relative humidity has a higher spatial variability, clear from the temporal evolution of the probability density function (PDF) of the spatial moisture field (Figure 2). In the control experiment with higher horizontal moisture transport efficiency, the PDF stays essentially unimodal throughout the simulation, although a second minor mode corresponding to saturated cells is in evidence and is directly due to the externally imposed constraint (Equation 4) on the number N_c of convectively active columns per time step. The primary unimodal feature of the PDF is to be ascribed to larger diffusive effects (combined with relatively slow drying tendencies), which prevent the domain from developing some drier-than-average background region surrounding moist patches. A transition toward a broader distribution is apparent in the lower diffusion experiment which undergoes aggregation, since the action

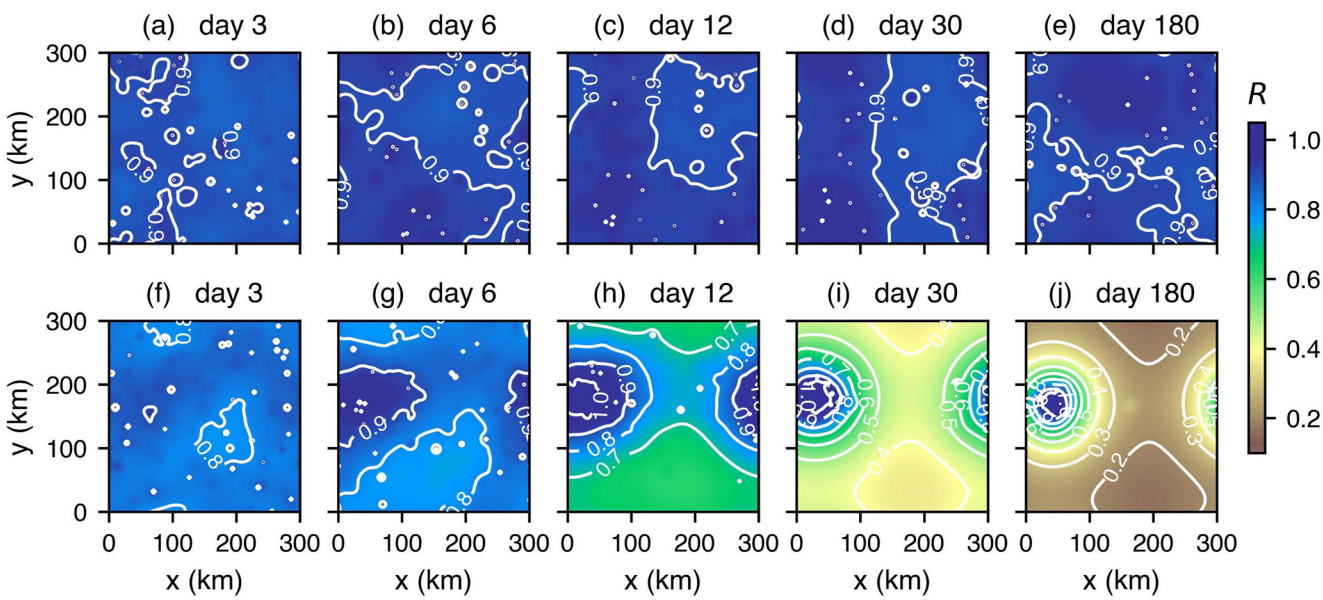

Figure 1. Evolution of the spatial column relative humidity field R (color shading and contours, with intervals of 0.1) for simulations with $K = 10^4$ m^2 s^{-1} (a–e), $K = 5 \times 10^3$ m^2 s^{-1} (f–j), $\tau_{sub} = 16$ days and $a_d = 14.72$. The domain size and the grid resolution are kept at their default values, $L = 300$ km and $\Delta x = 2$ km.

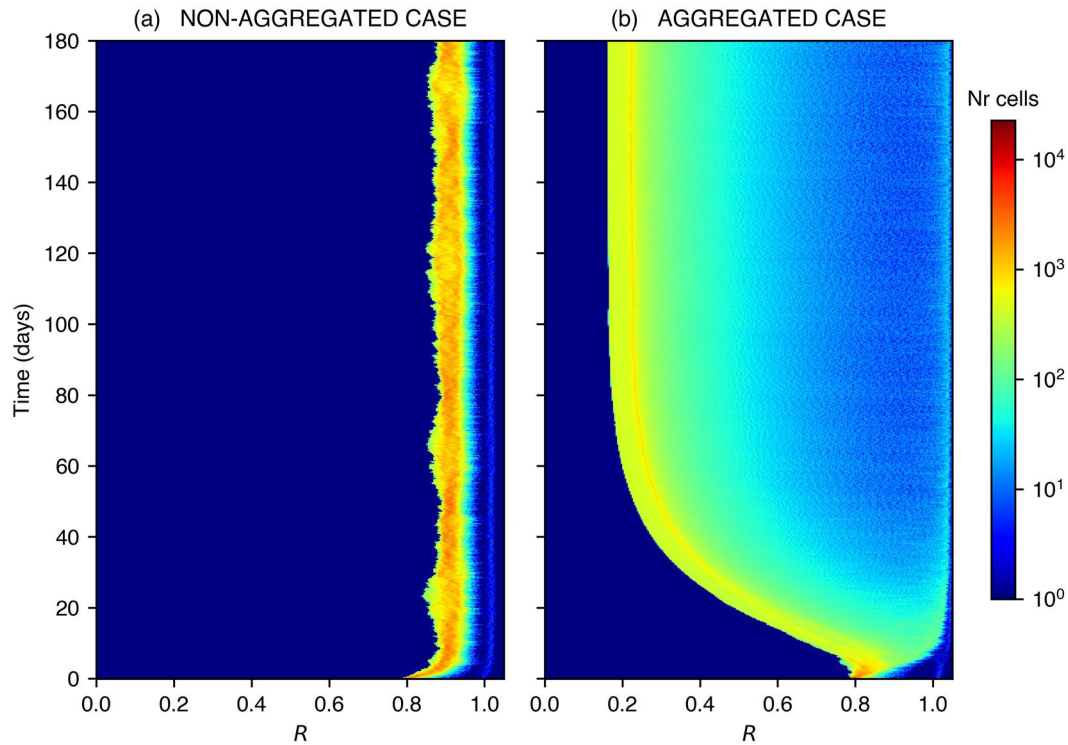

Figure 2. Time evolution of the absolute frequency of occurrence of R values for the simulations reported in Figure 1, namely (a) non-aggregating case with $K = 10^4$ m^2 s^{-1}, and (b) aggregating case with $K = 5 \times 10^3$ m^2 s^{-1}. The other model settings are $\tau_{sub} = 16$ days and $a_d = 14.72$.

of moistening processes is able to overcome the counter-gradient smoothing by subsidence and diffusion. As self-aggregation progresses and the dry and humid regions are increasingly separated, a bimodal PDF develops reminiscent of tropical observations (Mapes et al., 2018; Zhang et al., 2003). The dry mode here is linked to the long diffusive tail of the distribution and the moist mode is related to detrainment area, possibly exaggerated by

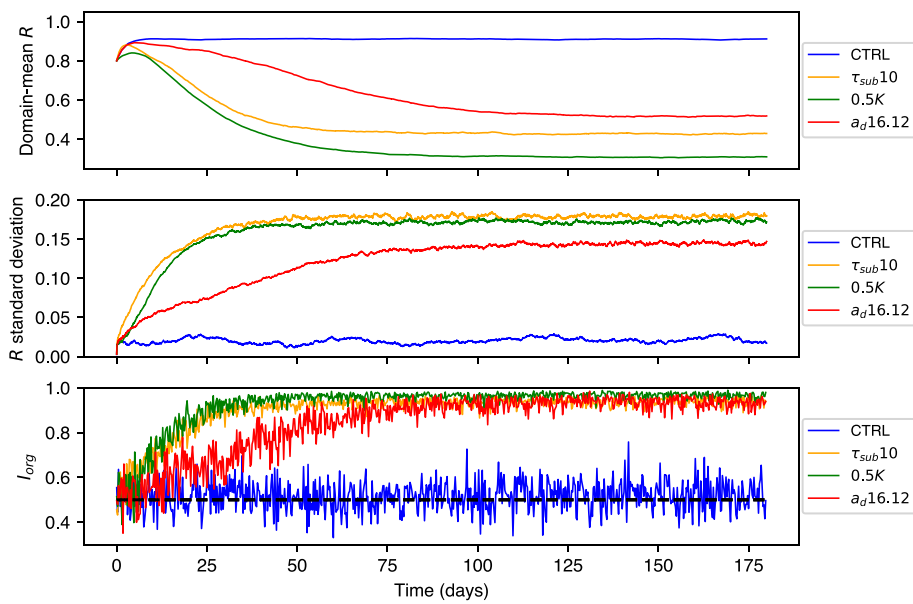

Figure 3. Temporal evolution of the spatial R distribution, in terms of domain mean (upper panel) and standard deviation (middle panel), and the organization index I_{org} (lower panel) for the simulations CTRL (blue line), 0.5K (green line), τ_{sub}10 (orange line), a_d16.12 (red line). The dashed line in the I_{org} plot marks $I_{org} = 0.5$, which is the value for a random distribution of convective cells. For details on the experimental setup, refer to Table 2.

19422466, 2023, 5, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022MS003231 by CochraneItalia, Wiley Online Library on [16/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

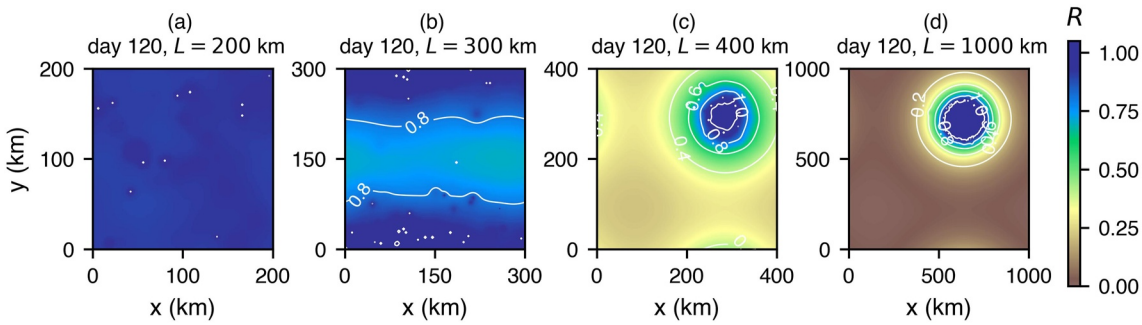

Figure 4. Plan views of the spatial field of column relative humidity R (shading and contours, interval 0.2) after 120 days of simulated time for runs with $K = 10^4$ $m^2\ s^{-1}$, $\tau_{sub} = 15$ days, $a_d = 14.72$ and domain sizes $L = 200$ km (a), 300 km (b), 400 km (c), 1,000 km (d).

the use of a single detrainment value. This behavior is almost identical to that shown in coarser resolution deterministic experiments of CM13.

Time series of R show the impact of aggregation on the mean humidity field (Figure 3) in four simulations including the control run (CTRL) and three perturbation experiments, which alter the horizontal transport efficiency ($0.5K$), the subsidence rate ($\tau_{sub}10$) or the convective-humidity feedback strength ($a_d16.12$) in turn. A brief overview of these runs is reported in Table 2. These simulations show that it is possible to generate self-aggregation in the model by reducing the diffusive humidity transport, increasing the subsidence rate or strengthening the convective-moisture feedback. It is interesting to note that the two a_d values corresponding to different TRMM retrieval versions can produce either random or aggregated states, all else being kept fixed. The existence of two characteristic timescales is apparent, the first associated with the initial fast adjustment on the convective timescale, and the second representing the time of adjustment to equilibrium related to the overturning timescale determined by the subsidence rate. This is also in agreement with previous CRM experiments using fixed surface temperatures (Cohen & Craig, 2004; Tompkins & Craig, 1998b) although Cronin and Emanuel (2013) highlight that longer timescales are possible if an interactive lower boundary is used. After the equilibrium state is reached, temporal fluctuations in the field are limited to shorter timescale variability associated with the relative position of convective events. The temporal variability is restrained by the condition that the convective population variation in time is limited (see Methods). The non-aggregated case, conversely, after the very first transient phase where initial convective events increase the humidity variance, reaches an equilibrium rapidly with a low spatial variance associated with the domain that is moistened throughout by local convective sources. The time evolution of the organization index I_{org} introduced by Tompkins and Semie (2017) shows that the convection remains random in the control run, with a time-average value of 0.5, while in the three perturbation experiments it increases toward values exceeding 0.9, indicating highly aggregated convection.

3.1. Sensitivity to Domain Size

CRM simulations show that self-aggregation is facilitated by large domains, with abrupt transition to clustered convection taking place when the domain size L exceeds a certain threshold, typically $L \gtrsim 200$–300 km (Bretherton et al., 2005; Jeevanjee & Romps, 2013; Muller & Bony, 2015; Muller & Held, 2012; Patrizio & Randall, 2019). Li (2021) provided an analytical argument to explain the domain-size dependence of self-aggregation within the framework of a conceptual, two-dimensional, stochastic reaction-diffusion model for the column moist static energy (CMSE) budget. In particular, such a dependence is found to result from the competing influences of vertical and horizontal advective transports on the CMSE anomalies.

Here too the occurrence of aggregated states is found to be sensitive to the domain size (Figure 4; another example is shown in Movie S2). Convection in the smallest domain of size $L = 200$ km remains in a random state for these parameter choices (see caption). For $L = 300$ km, there is no aggregation, but some variance of moisture over the scale of the domain is apparent, and the moist patch is elongated. This simulation was extended to 150 days which confirmed that this state is a quasi-stationary equilibrium. Extending the domain to 400 km results in aggregation with a single center.

19422466, 2023, 5, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022MS003231 by CochraneItalia, Wiley Online Library on [16/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

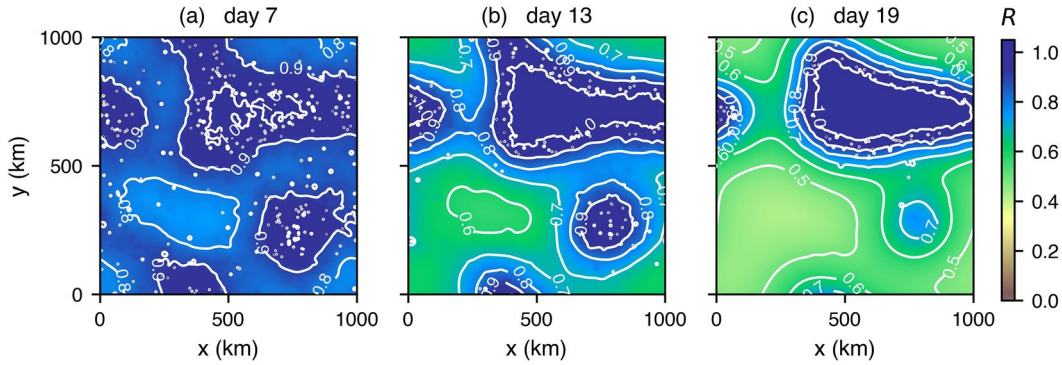

Figure 5. Horizontal maps of R (shading and contours every 0.1) after 7 (a), 13 (b) and 19 (c) days in the experiment with domain size $L = 1{,}000$ km.

The largest domain with $L = 1{,}000$ km exhibits an interesting behavior in that the convection originally organizes into two distinct convective clusters that last until around day 20, at which point the larger of the two centers starts to dominate and the first center dies out (Figure 5). This behavior is reminiscent of the 2D simulations of Held et al. (1993) which show two competing centers of convection for a period of time before collapsing to a single convective center, although this was on smaller domains. In our simple diffusive model, we hypothesize that the convection will always collapse to a single center due to the fact that the subsidence term is treated as a uniform relaxation toward zero R and does not account for the location of convection events, in contrast to the transport term which diffuses moisture out from the centers. In the real atmosphere, the subsidence occurs through the propagation of gravity waves from the convective centers, and thus the aggregated convective clusters would be separated by a Rossby deformation radius determined by the Coriolis effect off the equator and by diffusive dissipation, which would give a cluster spacing on $\mathcal{O}$(1,000 km) scales on the equator (Bretherton & Smolarkiewicz, 1989). Wing and Cronin (2016) offered an alternative mechanism for both the cluster separation distance and the spatial scale of aggregation based on boundary layer recovery through surface fluxes, which would also be a physical process missing in this simple model, that does not account for surface fluxes. Yang (2018) proposed that the characteristic horizontal scale of self-aggregation is determined by the boundary layer height and the density variations between moist and dry regions in the boundary layer, the latter owing to the virtual effect of water vapor. Additionally Beucler and Cronin (2019) recently used a new diagnostic to interpret the role of different diabatic forcings on the spatial scale of aggregation. In any case it remains that the formulation of the simple model presented here will always lead eventually to a single convective center in the cases where aggregation occurs.

3.2. Sensitivity to Horizontal Resolution

Aggregation in CRM studies is also resolution sensitive, with coarser grids favoring the occurrence of clustered convection. For instance, Muller and Held (2012) found that, for spacings $\Delta x < 2$ km, self-aggregation

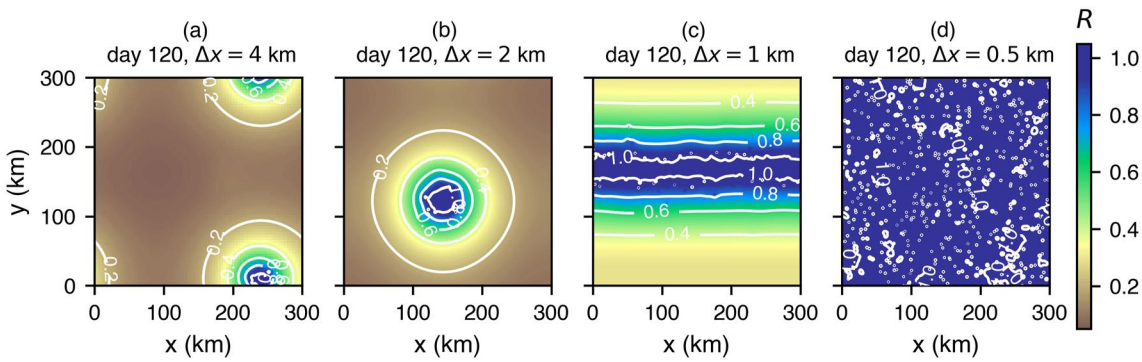

Figure 6. Snapshots of the R field (shading and contours, interval of 0.2) after 120 days of simulation in the case $K = 5 \times 10^3$ m^2 s^{-1}, $\tau_{sub} = 10$ days, $a_d = 14.72$, with $L = 300$ km, $\Delta x = 4$ km (a), 2 km (b), 1 km (c), 500 m (d).

19422466, 2023, 5, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022MS003231 by CochraneItalia, Wiley Online Library on [16/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

19422466, 2023, 5, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022MS003231 by CochraneItalia, Wiley Online Library on [16/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

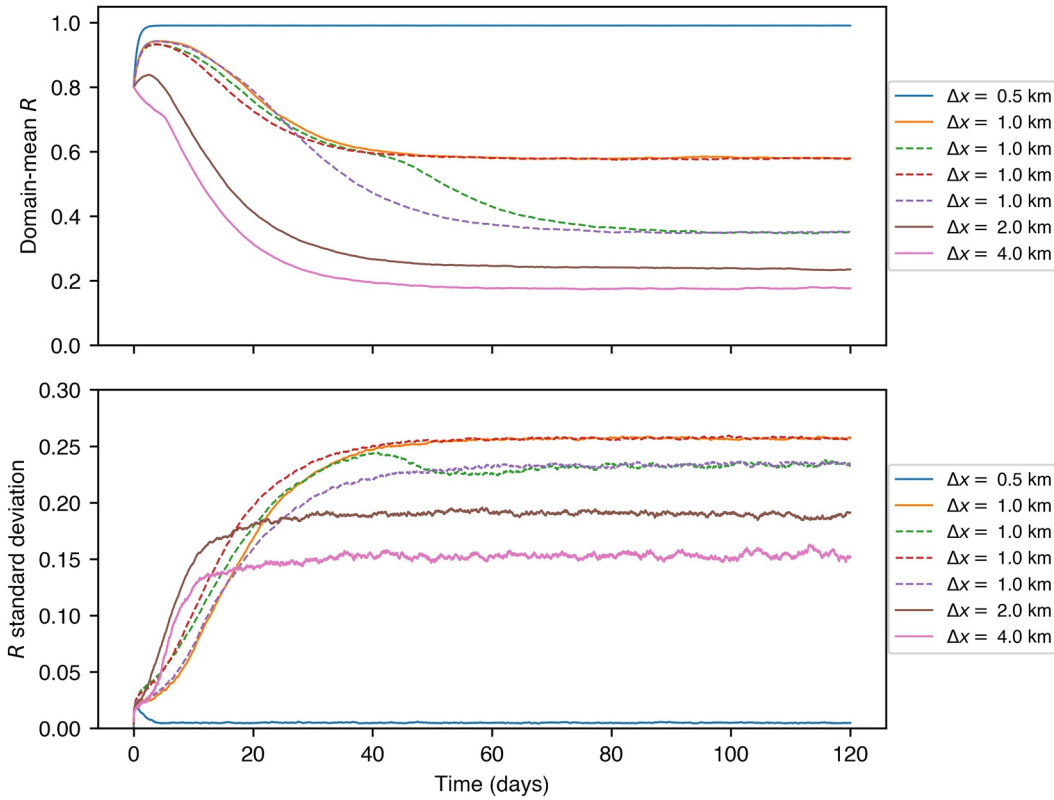

Figure 7. Time evolution of spatial R mean and standard deviation for the simulations presented in Figure 6 (solid lines) and a 4-member ensemble performed in the case $\Delta x = 1$ km (dashed lines).

never develops when starting from homogeneous initial conditions (but, when an aggregated initial profile is prescribed, it manages to persist even at resolutions as fine as $\Delta x = 500$ m if the domain size is sufficiently large, namely $L \gtrsim 200$ km). Similar results are found here, examining the atmospheric states at day 120 for simulations with the numerical grid successively refined (halved), with parameters K, τ_{sub}, a_d, and L invariant (Figure 6 and Movie S3). For a grid resolution of 4 and 2 km, the convection aggregates into a single center. Refining the resolution to 1 km, the aggregated state takes on the form of an elongated band, instead of the usual circular shape, spanning one horizontal dimension entirely, while using a resolution of 500 m leads to random convection that does not undergo aggregation at all.

Holloway and Woolnough (2016) provided a geometric argument to explain the preferred shape taken by self-aggregated convection in doubly-periodic RCE simulations, suggesting the structure of wet patches is such as to minimize their perimeter-to-area ratio, because lateral mixing acts to reduce any horizontal moisture gradient. In particular, if the area A_{cl} of the cluster is $A_{\text{cl}} > A_{\text{cl,crit}} \equiv \frac{L^2}{\pi}$ (i.e., the moist patch occupies roughly more than one third of the computational domain), then a band-like arrangement is likely to appear, as observed in the first 3-dimensional simulations of RCE by Tompkins and Craig (1998a), which only used a 100 × 100 km domain. In extremely large domain experiments, however, Patrizio and Randall (2019) actually show a transition from circular clusters toward elongated bands in the largest $\mathcal{O}$(6,000 km) domain experiments. For smaller ratios, and indeed over an infinite plane, the preferred form would be a circle in all cases, as in the seminal study of Bretherton et al. (2005).

For $\Delta x = 2$ km and $\Delta x = 4$ km, the time series of the spatial R mean show contrasting behavior at the simulation outset, with the initial adjustment in the R-mean profile completely absent in the 4 km case (Figure 7, top panel), as the initial phase involves the development of larger, but fewer, convection cells, while most columns start to be progressively dried by the subsidence. This prevents R from increasing at the beginning of the 4 km simulation when starting from these relatively moist initial conditions.

The 1 km simulation was repeated three times to ascertain any eventual, additional stochastic contribution to the final self-aggregated shape and indeed the results of the multi-run ensemble simulation, shown in Figure 7

19422466, 2023, 5, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022MS003231 by CochraneItalia, Wiley Online Library on [16/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

(orange solid line and dashed lines), manifest various evolutions. For the same parameter set and experimental design, the simulation may end up either with the usual spatial pattern typical of convective clustering, marked by a pronounced reduction in domain-mean R and a slightly lower variance, or with convective centers being aligned in a band. This indicates proximity to a critical cluster area $A_{\mathrm{cl,\ crit}}$ beyond which the wet spot arranges itself in a banded structure, and whether or not the corresponding radius is reached depends on the large stochastic effects present in the modeled system. The temporal evolution of one run (green dashed line in Figure 7) even shows an initial banded equilibrium state, which transitions to a circular cluster around day 40–45. Wing and Emanuel (2014) found similar behavior in their CRM simulations, pointing out that, in some runs, convection was confined to a single band maintained for tens of days before collapsing into a circular clump, the evolution of the spatial orientation of the cluster being thus attributed to the largely stochastic nature of self-aggregation. For $\Delta x = 500$ m, the profile is extraordinarily moist from the very beginning and so it persists throughout the run (Figures 6d and 7, blue line).

4. A Dimensionless Parameter to Predict Aggregation Onset

In the previous section it was shown that the occurrence of aggregation is sensitive to the settings of the three model parameters, K, τ_{sub}, and a_d, representing the strength of the horizontal transport, subsidence and the convective indicator function, respectively, as well as the domain size L and resolution Δx. Here we wish to derive a method to predict when aggregation will occur as a function of these five parameters. As a first step, we will use dimensional analysis to empirically derive a dimensionless quantity that predicts the onset of aggregation. We discuss the five parameters in turn to understand their impact on aggregation, and then construct the dimensionless parameter.

4.1. Sensitivity to K and τ_{sub}

The occurrence of the self-aggregated state is sensitive to the value of the horizontal moisture transport efficiency and subsidence strength. In this model, convection locally moistens its environment while drying the far-field instantaneously through subsidence. Thus the onset of aggregation will depend on how quickly moisture sources are communicated relative to the subsidence drying. Stronger diffusive transport reduces the spatial variance of humidity and makes aggregation less likely. Indeed, in the limit of infinite diffusion, convective moisture sources would be communicated instantaneously throughout the domain resulting in random convection. Likewise, stronger subsidence drying would act to promote aggregation. The competing influences of subsidence and horizontal transport are fundamental.

On dimensional grounds, the subsidence timescale (units s) and the horizontal transport efficiency ($\mathrm{m^2\ s^{-1}}$) can be combined together to give an area of influence ($K\tau_{\mathrm{sub}}$) on the moisture field of an individual convective cell. Such a quantity (or related ones) would naturally appear in the context of reaction-diffusion problems. For example, the stationary solution $\overline{R}(x)$ of the one-dimensional heat equation with linear sink term

$$\frac{\partial R}{\partial t} = K\frac{\partial^2 R}{\partial x^2} - \frac{R}{\tau_{\mathrm{sub}}},\ x > 0,\ t > 0, \tag{9}$$

with conditions $R(0, t) = R_c$ and $\lim_{x\to\infty} R(x, t) = 0$, is given by

$$\overline{R}(x) = R_c e^{-\frac{x}{\sqrt{K\tau_{\mathrm{sub}}}}}, \tag{10}$$

which shows that $(K\tau_{\mathrm{sub}})^{\frac{1}{2}}$ is the e-folding length scale of the steady-state R in this simple problem. Therefore, the area of influence can be intuitively viewed as a measure of the potential (rather than the actual, due to the finite cloud lifetime) maximum area impacted by an individual deep convective event.

As highlighted by the sensitivity studies (Figure 4 and Movie S2), when the diffusion-based communication of moisture from the sources acts over scales comparable to the domain size, aggregation may be easily prevented. This is consistent with results from previous literature. The stability analyses conducted by Shi and Fan (2021) proved that, if the diffusive tendencies are strong, large-scale humidity perturbations would be required to destabilize the spatially homogeneous state of tropical convection, up to completely inhibiting the clustering. By means of a conceptual model, Li (2021) analytically determined a threshold domain size below which the homogenizing

19422466, 2023, 5, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022MS003231 by CochraneItalia, Wiley Online Library on [16/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

effect of diffusive transport dominates, thereby preventing incipient CMSE anomalies from amplifying and leading to self-aggregation. Thus one could scale $K\tau_{sub}$ by the area of the computational domain, L^2. This would provide a dimensionless quantity, but does not yet account for the dependence of the resolution or the convective sensitivity to the moisture field, which will be considered below.

Moreover, the sensitivity of the occurrence of self-organization to τ_{sub} is more subtle, because the mean number $\overline{N_c}$ of convective points active at each time step introduces an additional dependence on the subsidence characteristic time through Equation 4. Since the number of updraft centers is inversely proportional to the subsidence timescale τ_{sub}, stronger subsidence, while reducing the area of influence $K\tau_{sub}$, also increases the density of convective events within the domain, reducing the mean inter-convective spacing. This means that experiments with different values of K and τ_{sub}, but the same product $K\tau_{sub}$, may exhibit different behavior; experiments with larger K and smaller τ_{sub}, and hence higher number of convective cells, are more likely not to organize.

These arguments and the evolution shown in Movie S4 undoubtedly motivate the necessity of including N_c in the dimensional analysis, either explicitly or implicitly. It seems reasonable to represent the contribution from N_c in terms of the distribution of spatial distances between convective towers, recalling that convection is initially randomly distributed prior to aggregation (or remains random in non-aggregating experiments).

4.2. Sensitivity to Resolution and Domain Size

Regarding first the domain-size sensitivity, it is intuitive that small domains may prevent aggregation especially when the moisture diffusion starts to act over scales on the order of the domain size, as already anticipated. Conversely, for large domains, even though the number of grid points occupied by convection increases accordingly as specified by the argument Equation 4, the maximum inter-convective distance will also increase as expected with a Poisson process. In the construction of the dimensionless parameter therefore, we shall heuristically argue that the key parameter is a measure of the expected maximum distance from the nearest convection, that is, a measure of the largest convective-free area, which will determine the magnitude of the spatial humidity variance in the pre-aggregated state. Larger distances from convection imply greater dry perturbations and humidity variance in the domain, more likely to lead to aggregation through the indicator random variable. The choice of this distance metric as a relevant one to the clustering onset is motivated by the evolution observed in the simple model (cf. Figure 1 and Movie S1) and also by the findings from previous literature, which, unless very few exceptions (Holloway & Woolnough, 2016), documents the formation, expansion and amplification of a dry patch with suppressed convection to be crucial in initiating the aggregation (e.g., Coppin & Bony, 2015; Wing & Emanuel, 2014).

Considering the resolution dependence, while the prominent sensitivity of self-aggregation to the horizontal resolution might be attributed to numerical artifacts, for instance the possibility of lateral mixing being resolution-dependent, grid refinement studies conducted to evaluate the spatial convergence properties of the numerical solver excluded this eventuality (cf. Figure S2 in Supporting Information S1). Instead, the resolution sensitivity here is a direct result of the number of convective sources. The scaling closure Equation 4 only constrains the cumulus *fraction* and not the number nor the size of convective points, and, as the resolution is refined, the convective fraction is the result of more convective centers. Put another way, with a resolution of 2 km, the minimum convective size is 4 km^2, but if Δx is halved, that same area now consists of four separate convective towers of 1 km^2 in different locations, since the model does not impose a horizontal scale on the updraft. This reduces the maximum distance between the convective cores and makes convective aggregation less likely. If a fixed area were set for a single convective updraft core, in order to avoid that the convection centers could become unrealistically small when moving to finer resolution below $\mathcal{O}$(1 km), we predict that no sensitivity to horizontal resolution would be found.

Although this explanation for resolution sensitivity seems simplistic, it is supported by recent experiments using an ensemble of CRM simulations of a mesoscale convective system at different resolutions (Prein et al., 2021). The study shows that the updraft dimension decreases monotonically with decreasing resolution and has still not converged even when the horizontal grid size reaches 250 m. Additionally, Sueki et al. (2019) show that the nearest neighbor distance between updraft cores reduces with finer resolutions and no convergence is reached at 200 m, directly supporting the hypothesized mechanism for resolution sensitivity represented in the simple

19422466, 2023, 5, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022MS003231 by CochraneItalia, Wiley Online Library on [16/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

stochastic model. Note that some stochastic approaches instead impose a specific fixed scale for convective events and would not exhibit this facet of resolution sensitivity (Fu & O'Neill, 2021; Showman, 2007; Yang, 2021).

4.3. A Distance Scaling in a Discrete Domain

The above findings further motivate the definition of a relevant distance for aggregation, which will account for the contribution from N_c and will also allow the incorporation of the resolution and domain size into the theory. In particular, in the discussion of the resolution and domain size dependence, it was heuristically argued that a relevant distance would be one that describes the largest distance from convection within the domain, which would determine the magnitude of the driest perturbation.

If the initial R distribution is horizontally homogeneous, as prescribed in all the experiments presented in this work, the convection locations are random at the simulation outset. In an infinite domain with a homogeneous planar Poisson point process, the cumulative distribution function of nearest neighbor distances between points is given by the Weibull distribution (Stoyan et al., 1987; Weger et al., 1992) as

$$\mathrm{NNCDF} = 1 - e^{-\lambda \pi r^2} \tag{11}$$

where λ is the density of the points (convective cells) and r is a radius. However, this approach is not appropriate here, as we need to consider the finite nature of the periodic domain, and treat convection as a binary occurrence on a discrete grid, that is, we consider cells to be either convective or non-convective.

During the entire pre-onset phase, the positions of the updraft centers can still be regarded as the restriction of a Poisson process to a compact set, the computational domain, and it is well known (e.g., Stoyan et al., 1987; Illian et al., 2008) that the resulting process obeys a binomial law. For a finite domain consisting of discrete cells we consider the probability, $p_{\mathrm{clr}}(n, N_c)$, of not finding any of N_c convective events within a square window of size $n\Delta x$ (consisting of n^2 grid boxes), centered at an arbitrary non-convective cell in the domain. This is termed *void probability* and can be approximated by

$$p_{\mathrm{clr}}(n, N_c) \approx \left(1 - \left(\frac{n\Delta x}{L}\right)^2\right)^{N_c}, n \in \mathbb{N}, n \leq \frac{L}{\Delta x}, \tag{12}$$

with $\mathbb{N}$ denoting the set of non-negative integers. If the base point has instead been chosen as convective, the void probabilities would simply be $p_{\mathrm{clr}}(n, N_c - 1)$. Owing to the imposed periodicity, no corrections are required if the central cell is in proximity to the edges of the domain. This relationship is an approximation since we should account for the fact that the sampling of convective grid boxes is *without* replacement (i.e., we should consider $\left(n^2 - 1\right)\left(\frac{\Delta x}{L}\right)^2$ in place of $n^2\left(\frac{\Delta x}{L}\right)^2$ in the previous formula), but this is negligible if the convective fraction is small (i.e., $N_c \ll N_{xy} = (L/\Delta x)^2$) as is the case here.

We consider two related metrics of the spacing of convective cells relevant to the onset of aggregation, which are illustrated in a schematic (Figure 8). The first distance metric is the size $d_{\mathrm{max,clr}}$ of the largest convective-free box, which would describe the greatest dry perturbation. The second metric instead considers a measure of the largest inter-convection nearest neighbor distance, specifically the dimension $d_{\mathrm{max,nn}}$ of the maximum box surrounding a convective cell that is devoid of further convective sources. The behavior of these two length scales is anti-correlated over the long term as convection starts to aggregate, since the size of the maximum convective-free region grows with the onset of aggregation, while the maximum inter-convective nearest neighbor spacing reduces, as shown comparing the left and right panels of Figure 8. This is also confirmed diagnosing the two quantities directly from the model simulations in Figure S4 in Supporting Information S1.

Using Equation 12, we can derive the distribution of $d_{\mathrm{max,nn}}$, by considering the central point of the search box to be each of the convective cells in turn. The probability that the maximum size of convection-free box centered at one of the convective towers is less than $n\Delta x$ is thus

$$p(d_{\mathrm{max,nn}} \leq n\Delta x) \approx (1 - p_{\mathrm{clr}}(n, N_c - 1))^{N_c}, n \in \mathbb{N}, n \leq \frac{L}{\Delta x}. \tag{13}$$

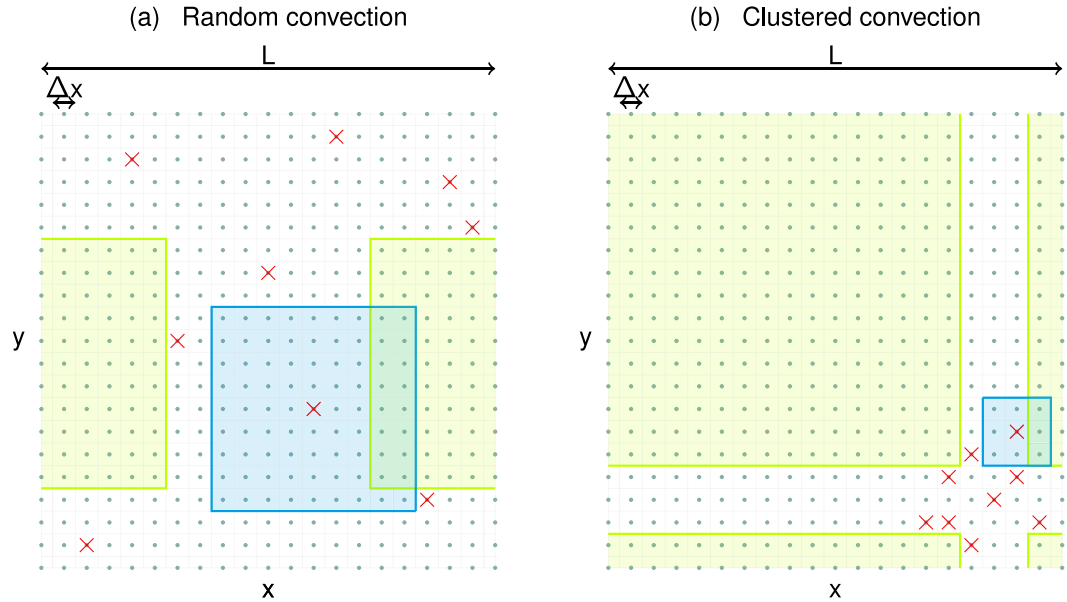

Figure 8. Sketch of two potential metrics of convective spacing relevant for aggregation onset, namely the size of the largest clear-sky, convective-free box ($d_{\mathrm{max,clr}}$, green boxes) and the maximum inter-convective nearest neighbor spacing ($d_{\mathrm{max,nn}}$, blue boxes), in a random convective situation (left) and highly aggregated situation (right). The cell centroids are represented as gray dots, the convective grid boxes as red crosses and the doubly-periodic nature of the domain is accounted for.

Equation 13 defines a cumulative distribution function, from which it is straightforward to calculate the percentiles and the expected value, $\overline{d}_{\mathrm{max,nn}}$, which represents, for a given density of randomly distributed convective sources, the average dimension of the maximum box surrounding any tower that is free from further events:

$$\overline{d}_{\mathrm{max,nn}} = \sum_{i=1}^{\frac{L}{\Delta x}} i\Delta x \left(\left(1 - \left(1 - \frac{i^2 \Delta x^2}{L^2} \right)^{N_c - 1} \right)^{N_c} - \left(1 - \left(1 - \frac{(i-1)^2 \Delta x^2}{L^2} \right)^{N_c - 1} \right)^{N_c} \right). \tag{14}$$

In the Supporting Information S1, we present the result of 70,000 artificially generated random convective scenes with varying N_c in order to show that the theoretical estimate for $\overline{d}_{\mathrm{max,nn}}$ presented in Equation 14 fits the numerical data perfectly (Figure S5 in Supporting Information S1).

One might consider the metric $d_{\mathrm{max,clr}}$ to be a more relevant metric related to the spatial variance of water vapor in the initial random convection phase, and thus to aggregation onset. An approximation for this metric is given by

$$p(d_{\mathrm{max,clr}} \leq n\Delta x) \approx (1 - p_{\mathrm{clr}}(n, N_c))^{N_{xy} - N_c}, n \in \mathbb{N}, n \leq \frac{L}{\Delta x}. \tag{15}$$

However, this analytical formula somewhat over-estimates the size of the maximum clear-sky square when tested with numerical data as it considers the test at each cell in the domain to be independent, which is not the case. The trials can instead be safely assumed independent in the derivation of Equation 13 due to the constraint $N_c \ll N_{xy}$. Additionally, the fact that N_{xy} is very large can lead to precision issues in the calculation of Equation 15. In any case, during the very early phase (first day) of the simulations, when convection is still random, an analysis of scene snapshots from the large ensembles shows that $d_{\mathrm{max,clr}}$ and $d_{\mathrm{max,nn}}$ are strongly linearly related (Figure 9), and thus either can be used in the scale analysis. We therefore choose to use $d_{\mathrm{max,nn}}$, also because it relates more closely to the more familiar nearest neighbor metrics adopted in the derivation of the widely used I_{org} aggregation metric. In the following sections, $\overline{d}_{\mathrm{max,nn}}$ will be referred to as $\overline{d}$ for brevity.

19422466, 2023, 5, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022MS003231 by CochraneItalia, Wiley Online Library on [16/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

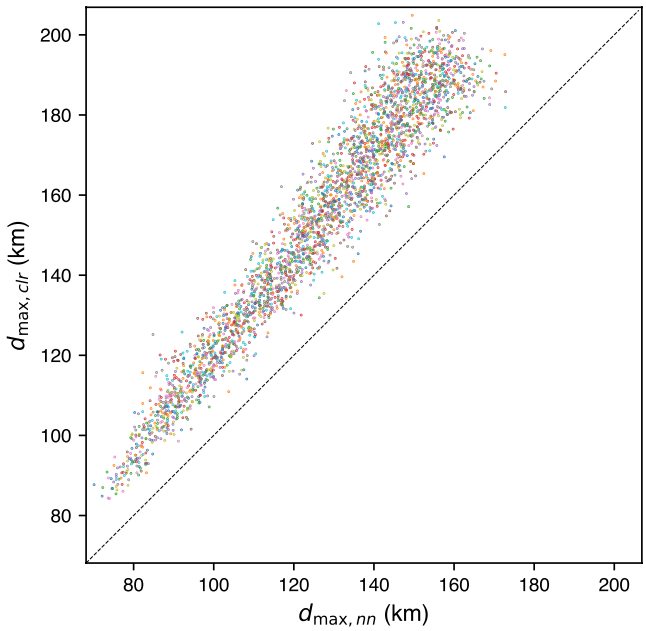

Figure 9. Scatter plot of daily averaged diagnostics for $d_{max,nn}$ versus $d_{max,clr}$ from hourly snapshots of scenes taken from a large ensemble in the first day of each experiment when convection is still randomly distributed. The identity line is shown as a black dashed line for better visualization.

4.4. Initial Dimensional Analysis

Combining the above considerations, let us introduce the following dimensionless parameter to explain the transition between homogeneous and aggregated regimes:

$$\gamma = f(a_d)\frac{K\tau_{sub}}{L^2}\frac{L}{\bar{d}} = f(a_d)\frac{K\tau_{sub}}{L\bar{d}}, \tag{16}$$

where $\bar{d}$ is given by Equation 14. The parameter γ consists of the normalized area of influence divided by the expected maximum inter-convective nearest neighbor distance, rescaled by the domain size L. Low values of γ (in turn corresponding to either lower K or τ_{sub} or larger L or $\bar{d}$) are supposed to represent aggregated states.

In addition to the four factors of domain size, resolution, horizontal transport efficiency and subsidence rate that were discussed above, we have also incorporated the sensitivity of convection to water vapor through a generic function $f(a_d)$, where f expresses the (unknown) functional dependence on a_d. As a_d is dimensionless, the functional form of f will be derived empirically using an ensemble of numerical experiments.

However, treating a_d and the other parameters in Equation 16 separately seems reasonable since we speculate that the evolution to aggregation requires the generation of R anomalies in the initial random phase, which are then magnified by the indicator function. As discussed, the formation of R anomalies is favored by large convection-free areas and results from the subsequent interplay between subsidence and horizontal transport, while

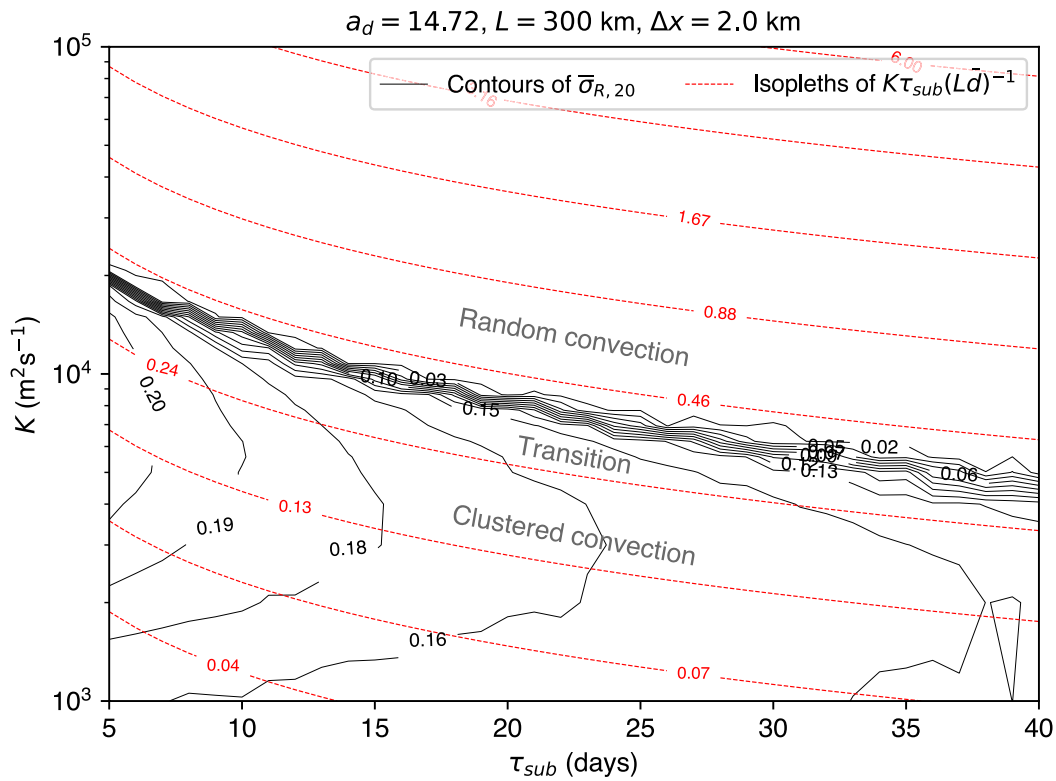

Figure 10. Contours of $\overline{\sigma}_{R,20}$ (black solid curves) along with the isopleths of $K\tau_{sub}\left(L\bar{d}\right)^{-1}$ (red dashed curves) for single realizations of the system in a large set of simulations with different K and τ_{sub} and fixed a_d, L and Δx.

19422466, 2023, 5, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022MS003231 by CochraneItalia, Wiley Online Library on [16/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

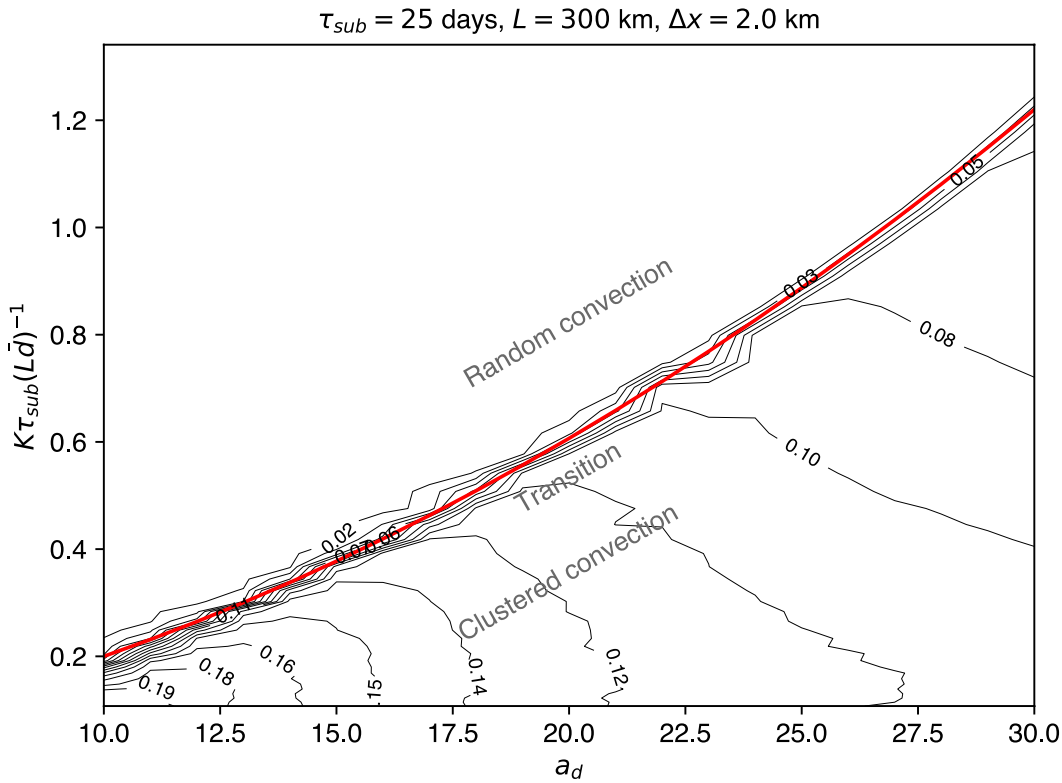

Figure 11. Contours of $\overline{\sigma}_{R,20}$ (black solid curves) for an ensemble of runs carried out with different values of $K\tau_{\text{sub}}\left(L\bar{d}\right)^{-1}$ (here obtained varying K and keeping τ_{sub}, L, and Δx fixed) and a_d. The red solid line represents the polynomial (quadratic) empirical fit for the transition regime.

a_d does not contribute at the beginning and its role can be decoupled from the rest of the analysis. Indeed, if we decompose the column humidity field as $R(\mathbf{x},t) = \overline{R}(t) + R'(\mathbf{x},t)$, $\overline{R}(t)$ and $R'(\mathbf{x}, t)$ being the domain-mean R and the local departure from the mean, respectively, it is apparent that $e^{a_d R(\mathbf{x},t)} = e^{a_d \overline{R}(t)} e^{a_d R'(\mathbf{x},t)} \approx e^{a_d \overline{R}(t)}$ if $\overline{R}(t) \gg |R'(\mathbf{x},t)| \approx 0$, as is the case prior to aggregation onset. At this stage, the weights of the selection process are dominated by the uniform term $e^{a_d \overline{R}(t)}$. Conversely, if substantial R anomalies are created, they are easily amplified by the exponential shape of the convection-vapor feedback. In this respect, we note that the exponential form of the moisture-convection feedback is not necessary for aggregation to occur, but a nonlinear relationship is still needed to retain the key behavior of the model (cf. Figure S9 in Supporting Information S1).

Thus, setting aside the functionality f for the moment and assuming $a_d = 14.72$, and using the default domain size and resolution (L = 300 km, Δx = 2 km), we evaluate the ensemble experiments that vary K and τ_{sub} to see if the dimensionless quantity (16) correctly predicts the final state to be clustered or random. Figure 10 shows contours of $\overline{\sigma}_{R,20}$, which we recall is the spatial standard deviation of R in the last 20 days. The region of dense contour lines marks the abrupt transition between those experiments that result in aggregated convection (high values of $\overline{\sigma}_{R,20}$) and those with random convection (low $\overline{\sigma}_{R,20}$) equilibrium states. Below the transition zone, on the left, the pronounced curvature of the contours is due to increasingly weak diffusive effects that encourage convection to (re)develop in a very restricted number of points, thus limiting the size of the cluster (hence the variance of the spatial R distribution). The slope of the transition zone in (τ_{sub}, K) space is almost exactly parallel to the isopleths of $\frac{K\tau_{\text{sub}}}{L\bar{d}}$ (recalling that a_d is fixed here), represented as red dashed curves. Further sets of simulations from the grand ensembles were examined for other values of L and Δx, with the fit still holding for fixed a_d, and the critical threshold value is the same as in this default case L = 300 km, Δx = 2 km (Figures S6 and S7 in Supporting Information S1). This means that there is a critical value that predicts the onset of aggregated convection. The critical isopleth that fits the transition will depend on a_d and thus the final task is to determine the functional dependence on a_d in the specification of γ.

19422466, 2023, 5, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022MS003231 by CochraneItalia, Wiley Online Library on [16/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

19422466, 2023, 5, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022MS003231 by CochraneItalia, Wiley Online Library on [16/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

4.5. The Role of the Parameter a_d

Intuitively, the relationship Equation 6 may strongly impact the aggregation of convection, via the steepness a_d of the exponential function, which governs the choice of convective locations: low values of a_d indicate that convection is very insensitive to water vapor anomalies and stochasticity of the convection choice may dominate, whereas high values produce organization as essentially only the moistest columns are likely to be selected after the initial perturbations are introduced in the water field. In fact, in the limit $a_d \simeq 0$, Equation 6 is homogeneous in the interval $[R_{\min}, R_{\max}]$, $R_{\min}$ and $R_{\max}$ being the minimum and maximum R values throughout the domain, and convection is rendered completely random by definition. It is thus expected that, as a_d increases, the critical isopleth will be shifted upwards in the (τ_{sub}, K) space (Figure S8 in Supporting Information S1).

The functional dependence of the transition on a_d is determined empirically (Figure 11), using $\overline{\sigma}_{R,20}$ for simulations performed with a range of values of K and a_d and fixed τ_{sub}, L, and Δx. The fit from empirical data shows that the position of the transition regime in the parameter space increases quadratically with a_d. The changes of $\overline{\sigma}_{R,20}$ for the simulations with aggregated convection are due to the absence of monotonicity of $\overline{\sigma}_{R,20}$ with a_d. Indeed, for organized runs, the size of the moist, convectively active region is reduced for high values of a_d. Owing to the increasingly steep shape of the exponential function Equation 6, the larger a_d gets, the more likely is for convection to reactivate at the same spots (which are the moistest ones), thus shrinking the wet patch, enlarging the area occupied by subsiding air and therefore reducing the spatial R variance beyond the onset point of aggregation. This nonlinear behavior of $\overline{\sigma}_{R,20}$ in the clustered state seen in both Figures 10 and 11, with the spatial variance of humidity increasing sharply with aggregation onset but reducing as the degree of aggregation strengthens, implies that humidity variance can be used to determine whether aggregation has occurred or not, but is not an effective metric of the degree of aggregation for model inter-comparison studies such as Wing et al. (2020).

4.6. The Aggregation Number

Knowing the quadratic dependence of a_d allows us to construct the full dimensionless quantity that incorporates all three model parameters and the experiment domain size and resolution, which will be referred to as the *aggregation number* N_{ag}:

$$N_{ag} = \frac{K\tau_{\text{sub}}}{a_d^2 L\overline{d}}. \tag{17}$$

We make an evaluation of the final dimensionless parameter using a complete ensemble of experiments which investigate the full 5-dimensional parameter space of changing K, τ_{sub}, a_d and the domain size L and resolution Δx. The resulting scatter plot in Figure 12 shows that the dimensionless quantity N_{ag} as specified in Equation 17 predicts the transition from random to aggregated states when the combination of these five parameters gives a N_{ag} value below a critical threshold, $N_{\text{ag,crit}}$, of approximately 1.72×10^{-3}. This estimate (i.e., the vertical line in Figure 12) has been obtained with an iterative procedure which yields equal number of misses on either sides of the vertical line itself. A threshold of $\overline{\sigma}_{R,20} = 0.05$ was imposed to distinguish between aggregated and non-aggregated runs. There is some variation in the transition zone which we attribute to the stochastic nature of the model. Indeed, repeating some of the experiments with configurations such that $N_{\text{ag}} \sim N_{\text{ag,crit}}$, with small initial random perturbations, showed that these could end up in either a random or aggregated state.

5. Discussion and Conclusions

Simulations of RCE that are run on $\mathcal{O}$(1,000 km) domains and convective permitting resolutions can often, but not always, undergo a transition from initially randomly distributed convection to end up in an equilibrium state in which all the convective events are aggregated into a moist zone. This phenomenon has been termed self-aggregation as it spontaneously occurs due to local diabatic feedbacks, despite homogeneous initial and boundary conditions or forcing. It is important to understand as it could have implications for our assessment of tropical climate sensitivity, since aggregated states are drier and thus lose energy to space more efficiently. Whether or not a particular model undergoes organization has been shown to depend on the resolution and domain configuration. Moreover, it is likely to be sensitive to the parameterization schemes used, such as the microphysics and sub-grid scale turbulence schemes, and recent model inter-comparison studies have shown little consensus between models concerning the details of aggregated states, the sensitivity of aggregation to lower boundary temperature, or even whether a particular experiment configuration undergoes aggregation or not.

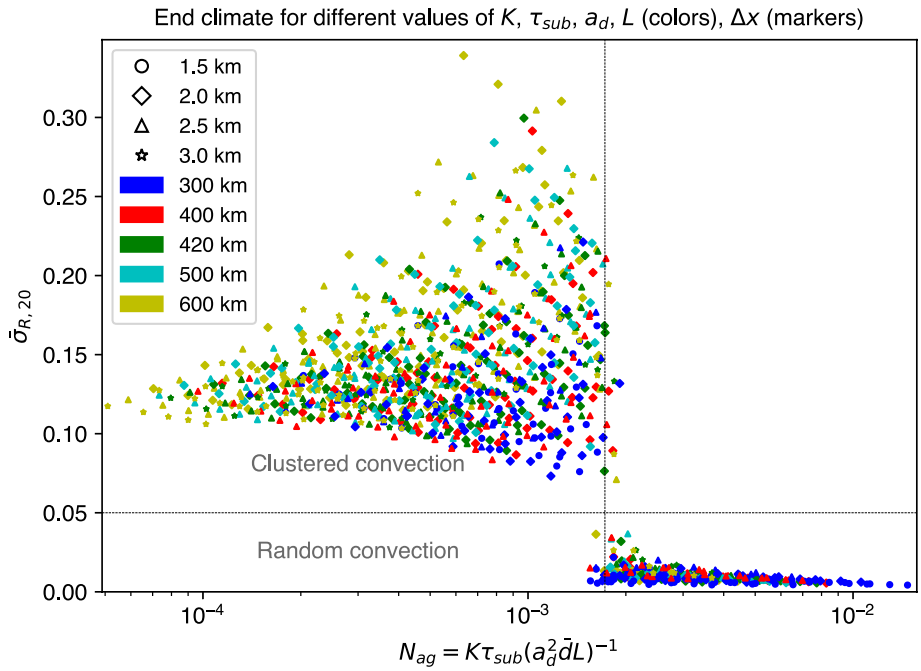

Figure 12. Scatter plot referring to a collection of simulations with different values of the parameter N_{ag}, as defined in Equation 17, each associated with the corresponding value of $\overline{\sigma}_{R,20}$. The horizontal dashed line $\overline{\sigma}_{R,20} = 0.05$ separates aggregated and non-aggregated runs, the vertical dashed line represents the threshold value of N_{ag} obtained as specified in the text.

Here we have attempted further understanding of these differences by introducing a stochastic reaction-diffusion model of the tropical atmosphere that uses similar domain sizes and resolutions to the full-physics, cloud resolving models. In our model, which is a development of the model previously presented by Craig and Mack (2013), convection towers are located according to a weighted random selection process, which makes convection more likely in moist areas using an observed functional form. The towers then rapidly moisten their local environment for the entirety of their life span, which averages 30 min. This local moistening is spread laterally by a local diffusive transport term, while subsidence drying balances the moistening uniformly throughout the domain, mimicking the action of fast spreading gravity waves in a highly idealized way as they are effectively assumed to have infinite group velocity. The model thus has three key parameters that describe the efficiency of the horizontal transport, the strength of the subsidence drying and the sensitivity of convection to humidity. Two additional parameters are the experiment domain size and resolution.

The model is found to produce both randomly distributed and aggregated states, depending on the five parameter settings. While over larger domains, two or more convective clusters can survive for a limited period in runs that aggregate, they always ultimately collapse to a single center, due to the fact that compensating subsidence occurs uniformly throughout the domain, that is, there is no explicit deformation radius. Sensitivities to the domain size and resolution are found in the simple model which mimic those found in the full-physics models, with aggregation more likely using larger domains and coarser spatial resolutions. We argue that the horizontal transport efficiency and subsidence rate can be dimensionally combined to give an "area of influence" of convection. Large areas would inhibit convective aggregation by enlarging the humidity "halo" around convective events.

Concerning the domain size and resolution, we heuristically argue that the important factor is a measure of the maximum convective-free distance prior to clustering onset, as this would determine the size of the humidity fluctuations in the pre-aggregated state. Indeed, a takeover of subsidence in the field far from convection could promote the formation of some drier-than-average region with suppressed convective activity, which may further develop and eventually lead to organization. Finer resolutions lead to more (smaller) convective centers, making aggregation less likely. We note that this is different from the suggestion of Tompkins and Semie (2017), who instead attributed resolution dependence to the reduction of explicit entrainment. Our simple model here permits a reinterpretation of the sensitivity of aggregation to diffusion scheme found in Tompkins and Semie (2017) not

19422466, 2023, 5, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022MS003231 by CochraneItalia, Wiley Online Library on [16/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

19422466, 2023, 5, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022MS003231 by CochraneItalia, Wiley Online Library on [16/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

as the effect of humidity entrainment into updrafts, but instead as the impact on mean updraft size, and therefore inter-convective distance.

Using these arguments and fits from experimental data, we were able to combine the domain size, resolution, horizontal transport efficiency, subsidence rate and the parameter that describes the sensitivity of convection to humidity into a single dimensionless parameter, N_{ag}, which we refer to as the *aggregation number*. Using super-ensembles of experiments that comprehensively explore the 5-dimensional parameter space of the model and experiment configurations, we demonstrate that the aggregation number N_{ag} is able to predict almost exactly if a particular model and domain setup will lead to aggregation, with the transition occurring at a specific critical value of the aggregation number, subject to a small amount of uncertainty due to the stochastic nature of the model.

Despite the simplistic nature of the model, it could help to explain differences between the full-physics CRM simulations seen in model inter-comparison projects such as RCEMIP (Wing et al., 2020). Models that mix humidity laterally efficiently, through higher numerical diffusion or the generation of vertical wind shear during the simulation, would be less likely to aggregate (Tompkins, 2000). Likely of more relevance is sensitivity of convection location to past convective events. In our model, and that of Craig and Mack (2013), the feedback is presented as one between convection and water vapor, demonstrated to play a role in Tompkins (2001) and Grabowski and Moncrieff (2004). In full-physics models, a number of additional diabatic processes act in tandem to promote or prevent aggregation, including radiative feedbacks with the cloud and moisture fields, surface fluxes, and the action of cold pools. The parameter a_d in the simple model, which essentially describes how likely convection is to occur in the vicinity of previous events, can be viewed as a proxy for all these feedbacks.

Ideally, the next step in this work is to devise a methodology to take consecutive CRM outputs and, using the auto-correlation of water vapor field and the locations selected for new convective events, to derive estimates for the three parameters of the simple model, and thus N_{ag}. While N_{ag} will likely evolve during the simulation, its calculation in the initial phase of the simulation when the convection is randomly distributed may predict if this random model state is unstable and will ultimately undergo aggregation. Moreover, if a given model is found to have a more complicated convective auto-correlation function, perhaps due to the mutual exclusivity of cold pools operating at scales smaller than 15 or 20 km, then this, conversely, could be incorporated into the simple model to explore the impact on aggregation in a wide parameter space.

While useful, the aggregation number does not tell the complete story. In fact, it would also be desirable to introduce a more theoretical framework to predict when the instability of the RCE state that leads to self-aggregation is expected to occur. A reasonable analytical approximation of the stochastic formulation presented here could help investigate some unexplored features of the simple model. For instance, the experiments conducted here all start from identical homogeneous moist conditions of 80% relative humidity, but, similar to CRM studies, the simple model is also found to be sensitive to the initial conditions, with aggregation more likely starting from drier and/or more heterogeneous conditions. That is, the model displays a (weak) hysteresis that can not be explored using the simple dimensionless parameter. To achieve this, a stability analysis of the system's variance equation is required, which will be a topic of future work.

Data Availability Statement

The numerical model code used for this work is freely available on github at https://github.com/adriantompkins/toy_diffusion and the version used in this paper is tagged v1.1.JAMES. The numerical model output is available in netcdf format at https://samodel.dmg.units.it/ and will be maintained for a minimum period of 5 years.

Acknowledgments
The authors acknowledge the valuable feedback and extensive suggestions of three anonymous reviewers, whose input greatly improved this manuscript. GB is supported by a MIUR/University of Trieste PhD scholarship.

References

Abbot, D. S. (2014). Resolved snowball Earth clouds. *Journal of Climate*, *27*(12), 4391–4402. https://doi.org/10.1175/JCLI-D-13-00738.1

Ahmed, F., & Neelin, J. D. (2019). Explaining scales and statistics of tropical precipitation clusters with a stochastic model. *Journal of the Atmospheric Sciences*, *76*(10), 3063–3087. https://doi.org/10.1175/JAS-D-18-0368.1

Beucler, T., & Cronin, T. (2019). A budget for the size of convective self-aggregation. *Quarterly Journal of the Royal Meteorological Society*, *145*(720), 947–966. https://doi.org/10.1002/qj.3468

Böing, S. J. (2016). An object-based model for convective cold pool dynamics. *Mathematics of Climate and Weather Forecasting*, *2*(1). https://doi.org/10.1515/mcwf-2016-0003

Bretherton, C. S., Blossey, P. N., & Khairoutdinov, M. (2005). An energy-balance analysis of deep convective self-aggregation above uniform SST. *Journal of the Atmospheric Sciences*, *62*(12), 4273–4292. https://doi.org/10.1175/jas3614.1

Bretherton, C. S., Peters, M. E., & Back, L. E. (2004). Relationships between water vapor path and precipitation over the tropical oceans. *Journal of Climate*, *17*(7), 1517–1528. https://doi.org/10.1175/1520-0442(2004)017<1517:rbwvpa>2.0.co;2

Bretherton, C. S., & Smolarkiewicz, P. K. (1989). Gravity waves, compensating subsidence and detrainment around cumulus clouds. *Journal of the Atmospheric Sciences*, *46*(6), 740–759. https://doi.org/10.1175/1520-0469(1989)046<0740:gwcsad>2.0.co;2

Cohen, B. G., & Craig, G. C. (2004). The response time of a convective cloud ensemble to a change in forcing. *Quarterly Journal of the Royal Meteorological Society*, *130*(598), 933–944. https://doi.org/10.1256/qj.02.218

Coppin, D., & Bony, S. (2015). Physical mechanisms controlling the initiation of convective self-aggregation in a General Circulation Model. *Journal of Advances in Modeling Earth Systems*, *7*(4), 2060–2078. https://doi.org/10.1002/2015ms000571

Craig, G. C., & Mack, J. M. (2013). A coarsening model for self-organization of tropical convection. *Journal of Geophysical Research: Atmospheres*, *118*(16), 8761–8769. https://doi.org/10.1002/jgrd.50674

Cronin, T. W., & Emanuel, K. A. (2013). The climate time scale in the approach to radiative-convective equilibrium. *Journal of Advances in Modeling Earth Systems*, *5*(4), 843–849. https://doi.org/10.1002/jame.20049

Derbyshire, S. H., Beau, I., Bechtold, P., Grandpeix, J.-Y., Piriou, J.-M., Redelsperger, J.-L., & Soares, P. M. M. (2004). Sensitivity of moist convection to environmental humidity. *Quarterly Journal of the Royal Meteorological Society*, *130*(604), 3055–3079. https://doi.org/10.1256/qj.03.130

Emanuel, K. A., Wing, A. A., & Vincent, E. M. (2014). Radiative-convective instability. *Journal of Advances in Modeling Earth Systems*, *6*(1), 75–90. https://doi.org/10.1002/2013ms000270

Fu, H., & O'Neill, M. (2021). The role of random vorticity stretching in tropical depression Genesis. *Journal of the Atmospheric Sciences*, *78*(12), 4143–4168. https://doi.org/10.1175/JAS-D-21-0087.1

Grabowski, W. W., & Moncrieff, M. W. (2004). Moisture-convection feedback in the tropics. *Quarterly Journal of the Royal Meteorological Society*, *130*(604), 3081–3104. https://doi.org/10.1256/qj.03.135

Haerter, J. O. (2019). Convective self-aggregation as a cold pool-driven critical phenomenon. *Geophysical Research Letters*, *46*(7), 4017–4028. https://doi.org/10.1029/2018GL081817

Held, I. M., Hemler, R. S., & Ramaswamy, V. (1993). Radiative-convective equilibrium with explicit two-dimensional moist convection. *Journal of the Atmospheric Sciences*, *50*(23), 3909–3927. https://doi.org/10.1175/1520-0469(1993)050⟨3909:RCEWET⟩2.0.CO;2

Hohenegger, C., & Stevens, B. (2016). Coupled radiative convective equilibrium simulations with explicit and parameterized convection. *Journal of Advances in Modeling Earth Systems*, *8*(3), 1468–1482. https://doi.org/10.1002/2016ms000666

Holloway, C. E., & Neelin, J. D. (2010). Temporal relations of column water vapor and tropical precipitation. *Journal of the Atmospheric Sciences*, *67*(4), 1091–1105. https://doi.org/10.1175/2009jas3284.1

Holloway, C. E., & Woolnough, S. J. (2016). The sensitivity of convective aggregation to diabatic processes in idealized radiative-convective equilibrium simulations. *Journal of Advances in Modeling Earth Systems*, *8*(1), 166–195. https://doi.org/10.1002/2015MS000511

Hottovy, S., & Stechmann, S. N. (2015a). Threshold models for rainfall and convection: Deterministic versus stochastic triggers. *SIAM Journal on Applied Mathematics*, *75*(2), 861–884. https://doi.org/10.1137/140980788

Hottovy, S., & Stechmann, S. N. (2015b). A spatiotemporal stochastic model for tropical precipitation and water vapor dynamics. *Journal of the Atmospheric Sciences*, *72*(12), 4721–4738. https://doi.org/10.1175/JAS-D-15-0119.1

Huang, J.-D., & Wu, C.-M. (2022). A framework to evaluate convective aggregation: Examples with different microphysics schemes. *Journal of Geophysical Research: Atmospheres*, *127*(5), e2021JD035886. https://doi.org/10.1029/2021JD035886

Hundsdorfer, W., & Verwer, J. G. (2007). *Numerical solution of time-dependent advection-diffusion-reaction equations*. Springer Science & Business Media.

Illian, J., Penttinen, A., Stoyan, H., & Stoyan, D. (2008). *Statistical analysis and modelling of spatial point patterns*. Wiley.

Jeevanjee, N., & Romps, D. M. (2013). Convective self-aggregation, cold pools, and domain size. *Geophysical Research Letters*, *40*(5), 994–998. https://doi.org/10.1002/grl.50204

Khouider, B. (2014). A coarse grained stochastic multi-type particle interacting model for tropical convection: Nearest neighbour interactions. *Communications in Mathematical Sciences*, *12*(8), 1379–1407. https://doi.org/10.4310/cms.2014.v12.n8.a1

Khouider, B., Biello, J., & Majda, A. J. (2010). A stochastic multicloud model for tropical convection. *Communications in Mathematical Sciences*, *8*(1), 187–216. https://doi.org/10.4310/cms.2010.v8.n1.a10

Khouider, B., Majda, A. J., & Katsoulakis, M. A. (2003). Coarse-grained stochastic models for tropical convection and climate. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(21), 11941–11946. https://doi.org/10.1073/pnas.1634951100

Li, Z. (2021). *Understanding the characteristics of precipitation and their response to climate change (unpublished doctoral dissertation)*. MIT.

Lin, J. W. B., & Neelin, J. D. (2000). Influence of a stochastic moist convective parameterization on tropical climate variability. *Geophysical Research Letters*, *27*(22), 3691–3694. https://doi.org/10.1029/2000gl011964

Lin, J. W. B., & Neelin, J. D. (2002). Considerations for stochastic convective parameterization. *Journal of the Atmospheric Sciences*, *59*(5), 959–975. https://doi.org/10.1175/1520-0469(2002)059<0959:cfscp>2.0.co;2

Majda, A. J., & Khouider, B. (2002). Stochastic and mesoscopic models for tropical convection. *Proceedings of the National Academy of Sciences*, *99*(3), 1123–1128. https://doi.org/10.1073/pnas.032663199

Mapes, B. E., Chung, E. S., Hannah, W. M., Masunaga, H., Wimmers, A. J., & Velden, C. S. (2018). The meandering margin of the meteorological moist tropics. *Geophysical Research Letters*, *45*(2), 1177–1184. https://doi.org/10.1002/2017GL076440

Mauritsen, T., & Stevens, B. (2015). Missing iris effect as a possible cause of muted hydrological change and high climate sensitivity in models. *Nature Geoscience*, *8*(5), 346–351. https://doi.org/10.1038/ngeo2414

Muller, C. J., & Bony, S. (2015). What favors convective aggregation and why? *Geophysical Research Letters*, *42*(13), 5626–5634. https://doi.org/10.1002/2015GL064260

Muller, C. J., & Held, I. M. (2012). Detailed investigation of the self-aggregation of convection in cloud-resolving simulations. *Journal of the Atmospheric Sciences*, *69*(8), 2551–2565. https://doi.org/10.1175/JAS-D-11-0257.1

Narsey, S., Jakob, C., Singh, M. S., Bergemann, M., Louf, V., Protat, A., & Williams, C. (2019). Convective precipitation efficiency observed in the tropics. *Geophysical Research Letters*, *46*(22), 13574–13583. https://doi.org/10.1029/2019GL085031

Patrizio, C. R., & Randall, D. A. (2019). Sensitivity of convective self-aggregation to domain size. *Journal of Advances in Modeling Earth Systems*, *11*(7), 1995–2019. https://doi.org/10.1029/2019MS001672

Peaceman, D. W., & Rachford, H. H., Jr. (1955). The numerical solution of parabolic and elliptic differential equations. *Journal of the Society for Industrial and Applied Mathematics*, *3*(1), 28–41. https://doi.org/10.1137/0103003

Plant, R. S., & Craig, G. C. (2008). A stochastic parameterization for deep convection based on equilibrium statistics. *Journal of the Atmospheric Sciences*, *65*(1), 87–105. https://doi.org/10.1175/2007JAS2263.1

19422466, 2023, 5, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022MS003231 by CochraneItalia, Wiley Online Library on [16/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Prein, A. F., Rasmussen, R. M., Wang, D., & Giangrande, S. E. (2021). Sensitivity of organized convective storms to model grid spacing in current and future climates. *Philosophical Transactions of the Royal Society A: Mathematical, Physical & Engineering Sciences*, *379*(2195), 20190546. https://doi.org/10.1098/rsta.2019.0546

Randall, D. A., & Huffman, G. J. (1980). A stochastic model of cumulus clumping. *Journal of the Atmospheric Sciences*, *37*(9), 2068–2078. https://doi.org/10.1175/1520-0469(1980)037<2068:asmocc>2.0.co;2

Raymond, D. J., & Zeng, X. (2000). Instability and large-scale circulations in a two-column model of the tropical troposphere. *Quarterly Journal of the Royal Meteorological Society*, *126*(570), 3117–3135. https://doi.org/10.1002/qj.49712657007

Rushley, S. S., Kim, D., Bretherton, C. S., & Ahn, M.-S. (2018). Reexamining the nonlinear moisture-precipitation relationship over the tropical oceans. *Geophysical Research Letters*, *45*(2), 1133–1140. https://doi.org/10.1002/2017GL076296

Shamekh, S., Muller, C., Duvel, J.-P., & D'Andrea, F. (2020). How do ocean warm anomalies favor the aggregation of deep convective clouds? *Journal of the Atmospheric Sciences*, *77*(11), 3733–3745. https://doi.org/10.1175/JAS-D-18-0369.1

Shi, X., & Fan, Y. (2021). Modulation of the bifurcation in radiative-convective equilibrium by gray-zone cloud and turbulence parameterizations. *Journal of Advances in Modeling Earth Systems*, *13*(10), e2021MS002632. https://doi.org/10.1029/2021MS002632

Showman, A. P. (2007). Numerical simulations of forced shallow-water turbulence: Effects of moist convection on the large-scale circulation of Jupiter and Saturn. *Journal of the Atmospheric Sciences*, *64*(9), 3132–3157. https://doi.org/10.1175/JAS4007.1

Sobel, A. H., Bellon, G., & Bacmeister, J. (2007). Multiple equilibria in a single-column model of the tropical atmosphere. *Geophysical Research Letters*, *34*(22), L22804. https://doi.org/10.1029/2007gl031320

Sobel, A. H., Nilsson, J., & Polvani, L. M. (2001). The weak temperature gradient approximation and balanced tropical moisture waves. *Journal of the Atmospheric Sciences*, *58*(23), 3650–3665. https://doi.org/10.1175/1520-0469(2001)058<3650:twtgaa>2.0.co;2

Stechmann, S. N., & Neelin, J. D. (2011). A stochastic model for the transition to strong convection. *Journal of the Atmospheric Sciences*, *68*(12), 2955–2970. https://doi.org/10.1175/JAS-D-11-028.1

Stechmann, S. N., & Neelin, J. D. (2014). First-passage-time prototypes for precipitation statistics. *Journal of the Atmospheric Sciences*, *71*(9), 3269–3291. https://doi.org/10.1175/JAS-D-13-0268.1

Stephens, G. L., van den Heever, S., & Pakula, L. (2008). Radiative–convective feedbacks in idealized states of radiative–convective equilibrium. *Journal of the Atmospheric Sciences*, *65*(12), 3899–3916. https://doi.org/10.1175/2008JAS2524.1

Stoyan, D., Kendalla, W. S., & Mecke, J. (1987). *Stochastic geometry and its applications*. John Wiley.

Strang, G. (1968). On the construction and comparison of difference schemes. *SIAM Journal on Numerical Analysis*, *5*(3), 506–517. https://doi.org/10.1137/0705041

Sueki, K., Yamaura, T., Yashiro, H., Nishizawa, S., Yoshida, R., Kajikawa, Y., & Tomita, H. (2019). Convergence of convective updraft ensembles with respect to the grid spacing of atmospheric models. *Geophysical Research Letters*, *46*(24), 14817–14825. https://doi.org/10.1029/2019GL084491

Tompkins, A. M. (2000). The impact of dimensionality on long-term cloud resolving model simulations. *Monthly Weather Review*, *128*(5), 1521–1535. https://doi.org/10.1175/1520-0493(2000)128<1521:tiodol>2.0.co;2

Tompkins, A. M. (2001). Organization of tropical convection in low vertical wind shears: The role of water vapor. *Journal of the Atmospheric Sciences*, *58*(6), 529–545. https://doi.org/10.1175/1520-0469(2001)058<0529:ootcil>2.0.co;2

Tompkins, A. M., & Craig, G. C. (1998a). Radiative-convective equilibrium in a three-dimensional cloud ensemble model. *Quarterly Journal of the Royal Meteorological Society*, *124*(550), 2073–2097. https://doi.org/10.1002/qj.49712455013

Tompkins, A. M., & Craig, G. C. (1998b). Time-scales of adjustment to radiative-convective equilibrium in the tropical atmosphere. *Quarterly Journal of the Royal Meteorological Society*, *124*(552), 2693–2713. https://doi.org/10.1002/qj.49712455208

Tompkins, A. M., & Semie, A. G. (2017). Organization of tropical convection in low vertical wind shears: Role of updraft entrainment. *Journal of Advances in Modeling Earth Systems*, *9*(2), 1046–1068. https://doi.org/10.1002/2016MS000802

Tompkins, A. M., & Semie, A. G. (2021). Impact of a mixed ocean layer and the diurnal cycle on convective aggregation. *Journal of Advances in Modeling Earth Systems*, *13*(12), e2020MS002186. https://doi.org/10.1029/2020MS002186

Weger, R. C., Lee, J., Zhu, T., & Welch, R. M. (1992). Clustering, randomness and regularity in cloud fields: 1. Theoretical considerations. *Journal of Geophysical Research*, *97*(D18), 20519–20536. https://doi.org/10.1029/92jd02038

Windmiller, J. M., & Craig, G. C. (2019). Universality in the spatial evolution of self-aggregation of tropical convection. *Journal of the Atmospheric Sciences*, *76*(6), 1677–1696. https://doi.org/10.1175/JAS-D-18-0129.1

Wing, A. A., & Cronin, T. W. (2016). Self-aggregation of convection in long channel geometry. *Quarterly Journal of the Royal Meteorological Society*, *142*(694), 1–15. https://doi.org/10.1002/qj.2628

Wing, A. A., Emanuel, K., Holloway, C. E., & Muller, C. (2017). Convective self-aggregation in numerical simulations: A review. In *Shallow clouds, water vapor, circulation, and climate sensitivity* (pp. 1–25). Springer.

Wing, A. A., & Emanuel, K. A. (2014). Physical mechanisms controlling self aggregation of convection in idealized numerical modeling simulations. *Journal of Advances in Modeling Earth Systems*, *5*, 1–14. https://doi.org/10.1002/2013MS000269

Wing, A. A., Stauffer, C. L., Becker, T., Reed, K. A., Ahn, M.-S., Arnold, N. P., et al. (2020). Clouds and convective self-aggregation in a multimodel ensemble of radiative-convective equilibrium simulations. *Journal of Advances in Modeling Earth Systems*, *12*(9), e2020MS002138. https://doi.org/10.1029/2020MS002138

Yanai, M., Esbensen, S., & Chu, J.-H. (1973). Determination of bulk properties of tropical cloud clusters from large-scale heat and moisture budgets. *Journal of the Atmospheric Sciences*, *30*(4), 611–627. https://doi.org/10.1175/1520-0469(1973)030<0611:dobpot>2.0.co;2

Yang, D. (2018). Boundary layer height and buoyancy determine the horizontal scale of convective self-aggregation. *Journal of the Atmospheric Sciences*, *75*(2), 469–478. https://doi.org/10.1175/JAS-D-17-0150.1

Yang, D. (2021). A shallow-water model for convective self-aggregation. *Journal of the Atmospheric Sciences*, *78*(2), 571–582. https://doi.org/10.1175/JAS-D-20-0031.1

Yano, J.-I., & Manzato, A. (2022). Does more moisture in the atmosphere lead to more intense rains? *Journal of the Atmospheric Sciences*, *79*(3), 663–681. https://doi.org/10.1175/JAS-D-21-0117.1

Zhang, C., Mapes, B. E., & Soden, B. J. (2003). Bimodality in tropical water vapour. *Quarterly Journal of the Royal Meteorological Society*, *129*(594), 2847–2866. https://doi.org/10.1256/qj.02.166

19422466, 2023, 5, Downloaded from https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022MS003231 by CochraneItalia, Wiley Online Library on [16/05/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License