A dimensionless parameter for predicting convective self-aggregation onset in a stochastic reaction-diffusion model of tropical radiative-convective equilibrium.

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

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

November 24, 2022

Abstract

We introduce a minimal stochastic lattice model for the column relative humidity (R) in the tropics, which incorporates convective moistening, lateral mixing and subsidence drying. The probability of convection occurring in a location increases with R, based on TRMM 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 or 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 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 diffusion, 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.

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

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.

Corresponding author: Adrian Tompkins, `Earth System Physics, ICTP, Strada Costiera 11, 34151 Trieste, Italy, Email: tompkins@ictp.it`

Abstract

We introduce a minimal stochastic lattice model for the column relative humidity (R) in the tropics, which incorporates convective moistening, lateral mixing and subsidence drying. The probability of convection occurring in a location increases with R, based on TRMM 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 or 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 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 diffusion, 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.

Plain Language Summary

Full physics models of the tropical atmosphere sometimes show a behavior whereby all the 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 ramifications for climate sensitivity, but the problem is that the 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, we call it the aggregation number, which 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.

1 Introduction

When computing resources first permitted the multiple-week integrations of convection-permitting, cloud resolving models (CRMs) needed to simulate a state of radiative-convective equilibrium, an interesting behavior was noted. After a period in which deep convection was randomly distributed throughout the domain, 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 has been termed self-aggregation, and was observed to occur in both 2D (Held et al., 1993) and 3D (Tompkins & Craig, 1998a) modeling frameworks. 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).

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 surface fluxes (Held et al., 1993; Tompkins & Craig, 1998a; Tompkins, 2001; Stephens et al., 2008; Wing & Emanuel, 2014; Muller & Bony, 2015; Holloway & Woolnough, 2016; Wing et al., 2017). Most of these feedbacks change in strength, and for surface fluxes even the sign, between the pre and post aggregated states (Wing & Emanuel, 2014; Tompkins & Semie, 2021). 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 upgradient has been recognized to be peculiar of aggregated runs and can help sustain the clustering (Bretherton et al., 2005; Muller & Held, 2012; Muller & Bony, 2015).

Despite the fact that models agree on the general role that the diabatic processes play in driving self-aggregation, there is a large disparity between models in the nature of the aggregated states and 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 cloud resolving model's subgrid parameterizations, such as the details of the microphysical schemes and subgrid-scale turbulence and mixing schemes used (Tompkins & Semie, 2017; Huang & Wu, 2022). 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 perhaps somewhat disturbingly finer grid resolutions are employed (Muller & Held, 2012).

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. For example, Raymond and Zeng (2000) coupled two single column models using the weak temperature gradient approximation of 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. Emanuel et al. (2014) produced a two layer model to show how an instability due to infrared radiation could occur in warmer, moister atmospheres. Other models are spatially explicit in the horizontal dimension, but simplify the physics of the system to a highly conceptualized representation. The first such model of this type was a 2D stochastic representation of cumulus self-aggregation of Randall and Huffman (1980). More recently, Böing (2016); Haerter (2019) have introduced 2D 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.

In another spatially explicit approach, a 2D model with a minimal 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 spatially explicit 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 exponential increase of tropical precipitation as a function of W observed in TRMM precipitation data (Bretherton et al., 2004; Rushley et al., 2018). While described in terms of the positive feedback between convection and water vapor, this function can be interpreted as representing the net positive feedback of all diabatic processes driving aggregation. The model thus has three physical parameters that describe the strength of the 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 ag-

gregation, termed "coarsening", for most parameter ranges, except for very weak feedback values far below those reported by Rushley et al. (2018). They described the instability in terms of the subsidence timescale and the feedback parameter, but without considering the impact of the transport, which we will show is also relevant.

In this paper, we aim to develop the model of CM13 further to investigate the stability of the system and predict whether the system will produce aggregation or random convection. The model presented here differs from CM13 in several respects. The key difference regards the spatial resolution. The water budget equation of CM13 was integrated on a 40km climate-model sized grid in which convection was treated deterministically. Here we instead use a cloud resolving grid resolution of $\mathcal{O}(1$ km) and treat convective activity as a stochastic process. This will allow us to incorporate the domain size and resolution into the theory for aggregation onset. Our modified model, also with a revised closure, can produce equilibrium states of random or aggregated convection, depending on the exact parameter settings and experiment set up employed, with the transition 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, transport, subsidence and convective moistening, as well as the domain size and resolution, that successfully predicts which simulation configurations result in convective aggregation.

2 Methods

2.1 Description of the model

The model presented here is a development of the model presented in CM13. The model of CM13 introduces a prognostic PDE for the vertically integrated water budget W. Here, we assume time-invariance of temperature and write the model in terms of the column total water relative humidity (CRH), $R = R(\mathbf{x}, t)$, defined as the sum of the density-weighted, column-integrated ice water (r_i), liquid water(r_l) and water vapor (r_v) mass mixing ratios normalized by the column-integrated saturation value (r_s),

$$R = \frac{\int \rho(r_v + r_l + r_i)dz}{\int \rho r_s dz}. \tag{1}$$

As in CM13, the model assumes that R in the tropical troposphere is rapidly increased by local deep convection. These moisture sources are then redistributed horizon-

tally by lateral mixing, while the convective moistening is balanced by subsidence drying throughout the simulation domain. Thus the effect of convection consists in moistening its local field through detrainment of water vapor and cloud condensate and drying the far-field through balancing subsidence. No large-scale dynamical forcing is imposed and the Coriolis effect is excluded. Diurnal and seasonal cycle representations are also omitted.

One key difference to CM13 is that the modified governing equation will be integrated on a 2D mesh of grid cells using a $\Delta x = \mathcal{O}(1 \text{ km})$ horizontal resolution and $L = \mathcal{O}(10^2\text{-}10^3 \text{ km})$ domain sizes to mimic the typical CRM experimental set up, with convection occurrence treated stochastically.

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

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

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.

Considering the diffusion and subsidence terms first, we follow CM13 in approximating the lateral transport of water vapor by a down-gradient diffusion, which is parameterized using a simple fixed value for the diffusivity K. This is obviously an oversimplification, but Windmiller and Craig (2019) demonstrated that a diffusive treatment of transport can reasonably represent the evolution of water vapor at least in the early stages of self-aggregation. The treatment of subsidence also follows CM13, as subsidence is modeled as a relaxation process towards a completely dry atmosphere, with a characteristic timescale τ_{sub} set to be uniform throughout the domain. We similarly neglect the complications of representing subsidence drying as a gravity wave propagation from convective events, as this propagation is fast relative to other processes and subsidence is a supposition of drying from all convective events.

The interesting behavior of the model, and its ability to represent convection in both random and aggregated configurations, derives from the specification of the convective moistening term. This is modeled as a (fast) relaxation towards R_c, the total water relative humidity in convective columns, which exceeds unity due to the detrainment of cloud

condensate. The moistening only occurs in locations occupied by convective updrafts, assigned by the indicator function $\mathcal{I}$, which is unity inside clouds and zero elsewhere.

To mimic the behaviour 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. We thus introduce the following semidiscrete formulation of the budget equation for R:

$$\frac{\partial R}{\partial t} = \frac{(R_c - R)}{\tau_c}\mathcal{H}(p_c(R) - X) + K\left(\delta_x^2 + \delta_y^2\right)R - \frac{R}{\tau_{\text{sub}}}, \tag{3}$$

with $\delta_{\cdot}^2$ denoting a suitable discretization of the second derivatives. In this form the indicator function is written in terms of the Heaviside function $\mathcal{H}(p_c(R)-X)$ which maps the domain to convecting (1) and non-convecting (0) cells based on a humidity-dependent probability function $p_c(R)$, while X is a uniform random variable. This term represents the sole stochastic contribution to the humidity budget, as the diffusion and subsidence terms operate continuously in all cells. This stochastic treatment allows experiments to use domain sizes and grid resolutions similar to those employed for the idealized CRM studies.

The probability p_c of a cell being chosen as a convective location depends on the column humidity. We follow CM13 in basing this 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 as a function of R as:

$$P(R) = P_0 e^{a_d R}, \tag{4}$$

where P_0 and a_d are constant coefficients quantifying the horizontal mean radiative–convective equilibrium rain rate and the sensitivity of precipitation to column humidity, respectively. This form was also confirmed independently by Holloway and Neelin (2010). To allow us to use this to specify $p_c(R)$, we make the assumption that the increase in precipitation as a function of R is solely due to more frequent occurrence; that is, we assume that the precipitation rate from each event is constant. While to zero order this seems to be a reasonable assumption according to a recent analysis of (mid-latitude) station data by Yano and Manzato (2022), this is an oversimplification, as increased humidity also leads to increased precipitation efficiency (Narsey et al., 2019). It would be straightforward to include a relationship for this 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 this assumption means we can apply (4) to give the probability of convective occurrence as

$$p_c(R_{j,k}) = C(t)e^{a_d R_{j,k}}, \tag{5}$$

where the subscripts $j, k = 1, \ldots, \dfrac{L}{\Delta x}$ 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 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. In effect, this change is moot as it would only affect convection at very dry values, which are not attained in the simulations here.

At each model timestep, a number of locations for the birth of new convective cells are chosen at random (without replacement) using the probability given in eqn. (5). There is a memory for convective cell locations and once a cell is selected as convective, at each subsequent time step it has a fixed probability of dying, to give an average convective duration τ_t set to 30 minutes. Increasing this average convective lifetime would make self-aggregation more likely, but this aspect of the model sensitivity is not investigated here. Thus at each time step, a specific number cells are "born" to replace dying cells and keep the population size at a specified number N_c.

The time averaged value of convective population size ($\bar{N_c}$) is prescribed as an external constraint by a simple mass conservation argument proposed by Tompkins and Craig (1998a). Specifically,

$$\bar{N_c} = N_{xy}\frac{|w_{\text{sub}}|}{w_c} = N_{xy}\frac{h}{\tau_{\text{sub}}}\frac{1}{w_c} = \left(\frac{L}{\Delta x}\right)^2 \frac{h}{\tau_{\text{sub}}}\frac{1}{w_c}, \tag{6}$$

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. One could simply set N_c as a constant in time, but we add a degree of stochasticity by selecting $N_c(t)$ as a Poisson process, which is then subjected to a box car average with time window of τ_t such that the birth-rate on any timestep is positive and we ensure the time-averaged convective population is precisely $\bar{N_c}$ in all experiments. It would be straightforward to additionally implement other time-varying functions on the population size to mimic the diurnal cycle or forcing by the large-scale flow.

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 (3) additively decomposed into two terms (the first including subsidence and diffusion, the second only convection), and the adoption of a second-order accurate Strang-type strategy (Strang, 1968). To solve the diffusion-reaction problem we employed second-order finite difference approximations in space and we developed a modified version of the classical Peaceman-Rachford Alternating Direction Implicit (ADI) 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 supplementary material.

2.3 Choice of the model parameters, constants and setup

In the paper large ensembles of integrations are carried out in order to find the combinations of parameters that lead to random or aggregated convection. All experiments are run for at least 120 days (with the 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 random convection we mostly use the domain spatial standard deviation of R averaged over the last 20 days of simulation ($\bar{\sigma}_{R,20}$). Low values of $\bar{\sigma}_{R,20}$ 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, aggregated or regular.

Table 1 contains details of these ensembles with the default values and ranges used for the ensembles. The e-folding time $\tau_{\rm sub}$ is derived assuming that in subsidence areas, in the absence of large-scale convergence, subsidence heating approximately balances the net radiative cooling, $Q_{\rm rad}$:

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

which, inserting characteristic values for the depth of the free troposphere ($h \approx 10$ km), the mean environmental lapse rate of potential temperature $\frac{{\rm d}\theta}{{\rm d}z} \approx 6.5$ Kkm^{-1} and the

vertically integrated radiative cooling rate $Q_{\rm rad} \approx 1.5$ Kday^{-1}, gives $\tau_{sub} = 16$ days, with the ensembles spanning 5-40 days. We note that this is much longer than the very short timescale used in CM13 of just 2 days. Analysis of TRMM data gave the convection sensitivity factor a_d values of 14.72 and 16.12, depending on the TRMM retrieval version (Rushley et al., 2018), and our ensembles span values of 10-30. A reasonable estimate for the horizontal moisture diffusivity 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 $v_0 = w_c = 10$ ms^{-1} (updraft velocity observed in convective cores) imply 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 time scale of 1 minute 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 cloud resolving model 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 of 300 km and $\Delta x = 2$ km cover combinations of $\tau_{\rm sub}$ and K, (results shown in Fig. 10), and K and a_d (Fig. 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 ($L = 200, 400, 1000$ km, Fig. 4) and spatial resolutions (0.5, 1, 1.5, 2 and 4 km, Fig. 6), see also Fig. S6 and Fig. S7 in the supporting information. To construct the final analysis exploring the 5 parameter space $(K, \tau_{\rm sub}, a_d, L, \Delta x)$, a sub-sampled ensemble of 1160 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 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.

Table 1. Parameters (default and ranges) used in the simulations.

	Default value	Range
K ($\mathrm{m^2 s^{-1}}$)	10^4	$10^3 - 4 \times 10^4$
τ_{sub} (days)	16	$5 - 40$
a_d	14.72 16.12	$10 - 30$
L (km)	300	$200 - 1000$
Δx (km)	2	$0.5 - 4$

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 diffusion ($0.5K$). Five time slice panels show the evolution of the horizontal R field (Fig. 1). 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 diffusive transport of water vapor (lower panels) causes the model to evolve towards a dramatically different state. After an initial period of random convection, the variability of R in the domain increases during the transition towards 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

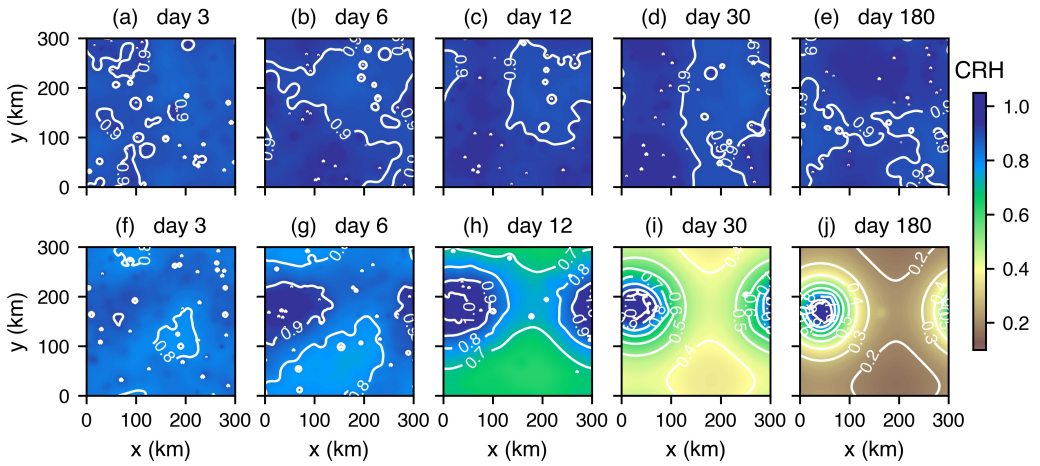

Figure 1. Evolution of the spatial CRH field for simulations with $K = 10^4$ m^{2}s^{-1} (a-e), $K = 5 \times 10^3$ m^{2}s^{-1} (f-j), $\tau_{\rm sub} = 16$ days and $a_d = 14.72$. The domain size and the grid resolution are kept at their default values, $L = 300$ km and $\Delta x = 2$ km.

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 (Fig. 2). In the control experiment with higher horizontal moisture diffusion coefficient, 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 (eqn. 6) on the number N_c of convectively active columns per time step. This behavior 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 towards a broader distribution is apparent in the lower diffusion experiment which undergoes aggregation, since the action 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

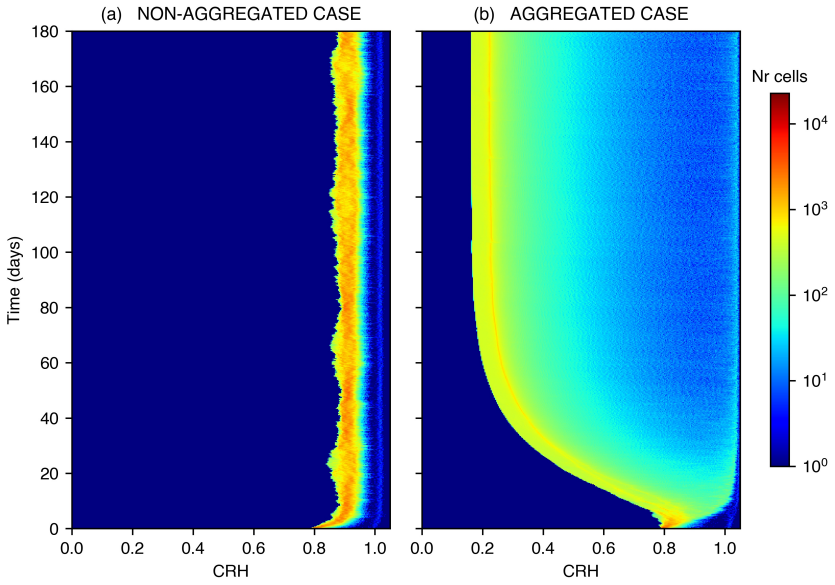

Figure 2. Time evolution of the absolute frequency of occurrence of CRH values for the simulations reported in Figure 1, namely (a) non aggregating case with $K = 10^4$ m^{2}s^{-1}, and (b) aggregating case with $K = 5 \times 10^3$ m^{2}s^{-1}, The other model settings are $\tau_{\text{sub}} = 16$ days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

of tropical observations (Zhang et al., 2003; Mapes et al., 2018). 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 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 (Fig. 3) in four simulations including the control run (CTRL) and three perturbation experiments, which alter the diffusion ($0.5K$), the subsidence rate ($\tau_{\text{sub}}10$) 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 time scales 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 (Tompkins & Craig, 1998b; Cohen & Craig, 2004) 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). In the non-aggregated case, conversely, after the very first transient phase where initial convective events increase the humidity variance, it 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 towards values exceeding 0.9, indicating highly aggregated convection.

Table 2. Summary of the simulations of Figure 3.

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

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; Muller & Held, 2012; Jeevanjee & Romps, 2013; Muller & Bony, 2015; Patrizio & Randall, 2019). Here too the occurrence of aggregated states is found to be sensitive to the domain size (Fig. 4). Convection in the smallest domain of 200 km remains in a random state for these

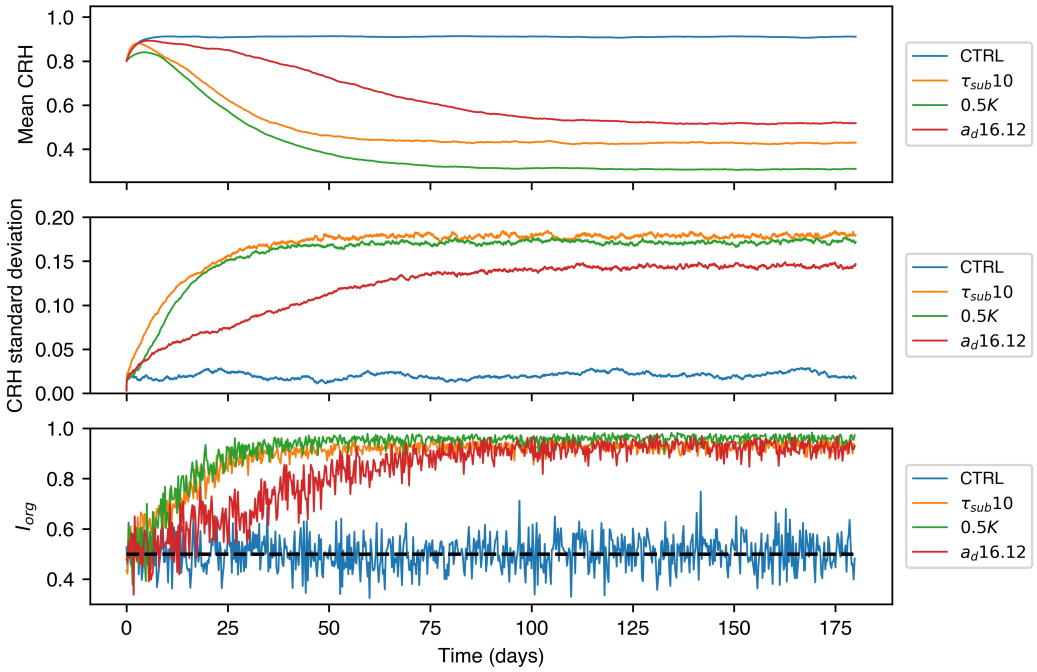

Figure 3. Temporal evolution of the spatial CRH distribution, in terms of domain mean (upper panel) and standard deviation (middle panel), and the organization index I_{org} (lower panel) for the simulations CTRL (blue line), $0.5K$ (green line), $\tau_{\mathrm{sub}}10$ (orange line), $a_d16.12$ (red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

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 quasi-stationary equilibrium. Extending the domain to 400 km results in aggregation with a single center.

The largest domain with $L = 1000$ km exhibits an interesting behavior in that the convection originally organizing 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 (Fig. 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 collapse 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 towards zero 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 $\mathcal{O}(1000$ km) scales on the equator (Bretherton & Smolarkiewicz, 1989). Wing and Cronin (2016) offered an alternative mechanism for cluster spacing 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. 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 never develops when starting from homogeneous initial conditions (but, when 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

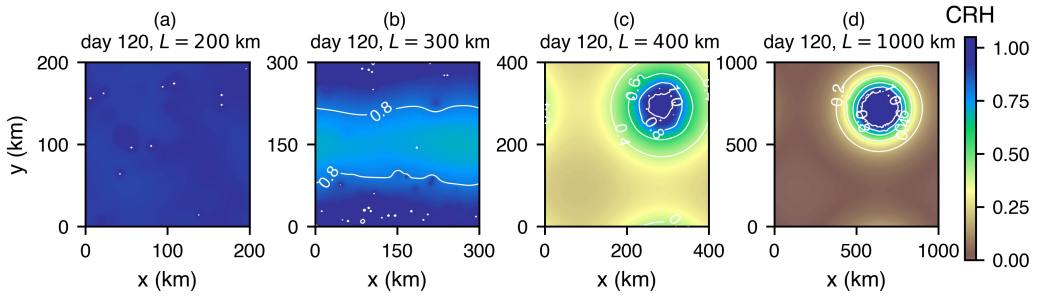

Figure 4. Plan views of the spatial CRH field after 120 days of simulated time for runs with $K = 10^4$ m^{2}s^{-1}, $\tau_{\rm sub} = 15$ days, $a_d = 14.72$ and domain sizes $L = 200$ km (a), 300 km (b), 400 km (c), 1000 km (d).

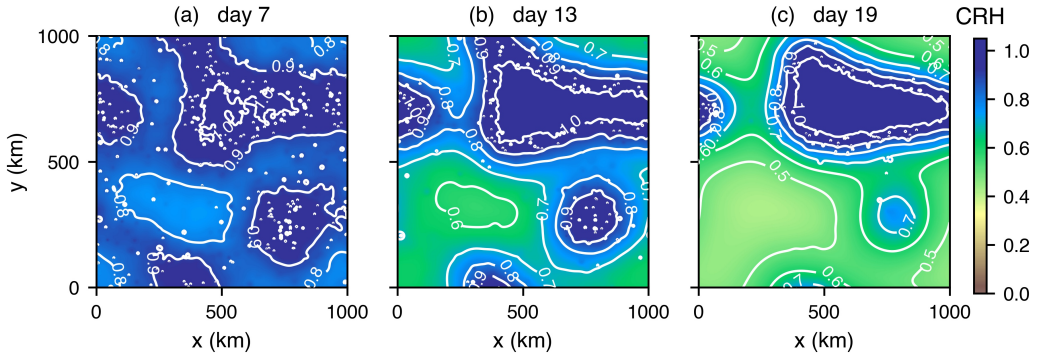

Figure 5. Horizontal maps of CRH after 7 (a), 13 (b) and 19 (c) days in the experiment with domain size $L = 1000$ km.

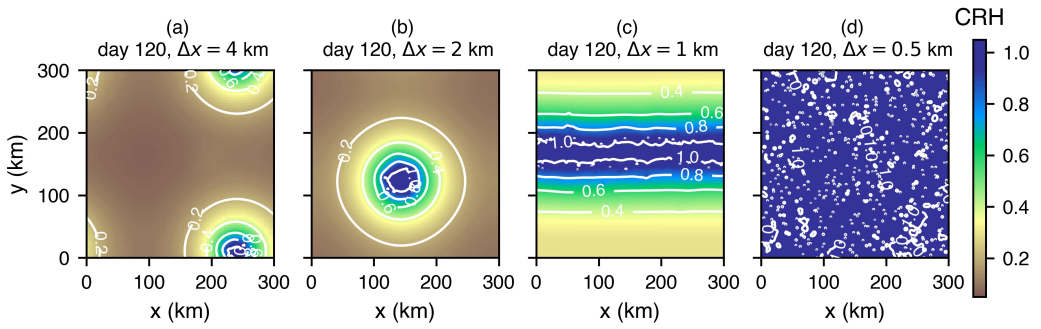

Figure 6. Snapshots of the CRH field after 120 days of simulation in the case $K = 5 \times 10^3$ m^{2}s^{-1}, $\tau_{\rm sub} = 10$ days, $a_d = 14.72$, with $L = 300$ km, $\Delta x = 4$ km (a), 2 km (b), 1 km (c), 500 m (d).

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, $\tau_{\rm sub}$, a_d and L invariant (Fig. 6). For a grid resolution of 4 km 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 $A_{\rm cl}$ of the cluster is $A_{\rm cl} > A_{\rm cl,\ crit} \equiv \frac{L^2}{\pi}$ (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 radiative-convective equilibrium 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 towards elongated bands in the largest $\mathcal{O}$(6000 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 R mean and variance show contrasting behavior at the simulation outset, with the initial adjustment in the R-mean profile completely absent in the 4 km case (Fig. 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 Fig. 7 (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 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 Fig. 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 relegated 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 (Fig. 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 diffusion, 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. To achieve this, one could derive a budget equation for the spatial variance of column relative humidity, previously identified as a suitable metric to detect aggregation. However, the presence of the stochastic indicator function in eqn. (3) complicates this approach, and thus as a first step, we

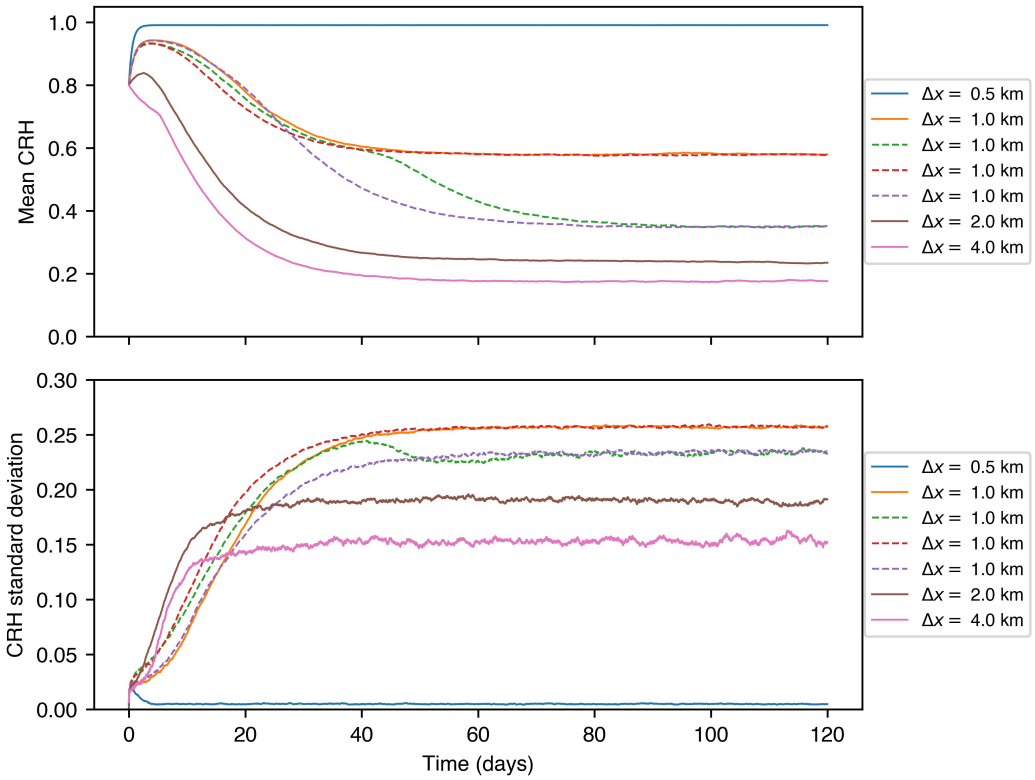

Figure 7. Time evolution of CRH mean and standard deviation for the simulations presented in Figure 6 (solid lines) and a 4-member ensemble performed in the case $\Delta x = 1$ km (dashed lines).

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 to 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 diffusion coefficient K 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 diffusion 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 diffusion are fundamental.

On dimensional grounds, the subsidence timescale (units s) and the diffusion coefficient ($\mathrm{m^2 s^{-1}}$) can be combined together to give an area of influence ($K\tau_{\mathrm{sub}}$) on the moisture field of an individual convective cell. As highlighted by the sensitivity study (Fig. 4), when the diffusion-based communication of moisture from the sources acts over scales comparable to the domain size, aggregation may be easily prevented. Thus one could scale $K\tau_{\mathrm{sub}}$ 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 $\bar{N}_c$ of convective points active at each time step introduces an additional dependence on the subsidence characteristic time through eqn. (6). Since the number of updraft centers is inversely proportional to the subsidence timescale τ_{sub}, stronger subsidence, while reducing the area of influence $K\tau_{\mathrm{sub}}$, also increases the density of convective events within the domain, reducing the mean convective spacing. This means that experiments with different values of K and τ_{sub} but the same product $K\tau_{\mathrm{sub}}$, 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 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 eqn. (6), 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, i.e., 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 function.

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. Fig. S2 in the supporting information). Instead, the resolution sensitivity here is a direct result of the number of convective sources. The scaling closure eqn. (6) 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 an MCS 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 resolution 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 200m resolutions, directly supporting the mechanism represented in the simple stochastic model down to these resolutions.

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 Poisson point process, the cumulative distribution function of nearest neighbor distances between points (NNCDF) is given by the Weibull distribution (Stoyan et al., 1987; Weger et al., 1992) as

$$\mathrm{NNCDF} = 1 - e^{-\lambda \pi r^2} \tag{9}$$

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. We consider cells to be either convective or non convective, thus the updraft centers can be regarded as the restriction of a Poisson process to a compact set, the computational domain. 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 win-

dow of size $n\Delta x$ (consisting of n^2 grid boxes), centred at an arbitrary non-convective cell in the domain. This is termed *void probability* and can be approximated by

$$p_{\mathrm{clr}}(n, N_c) \approx \left(1 - \left(\frac{n\Delta x}{L}\right)^2\right)^{N_c}, n \in \mathbb{N}, n \leq \frac{L}{\Delta x}, \tag{10}$$

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 (e.g., no two convective cells can occupy the same location), 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 the cells relevant to the onset of aggregation, which are illustrated in a schematic (Fig. 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 Fig. 8. This is also confirmed diagnosing the two quantities directly from the model simulations in Fig. S4.

Using (10), 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

$$p(d_{\mathrm{max,nn}} \leq n\Delta x) \approx (1 - p_{\mathrm{clr}}(n, N_c - 1))^{N_c}, n \in \mathbb{N}, n \leq \frac{L}{\Delta x}. \tag{11}$$

Eqn. (11) defines a cumulative distribution function, from which it is straightforward to calculate the percentiles and the expected value, $\bar{d}_{\mathrm{max,nn}}$, which represents, for a given density of randomly distributed convective sources, the average dimension of the

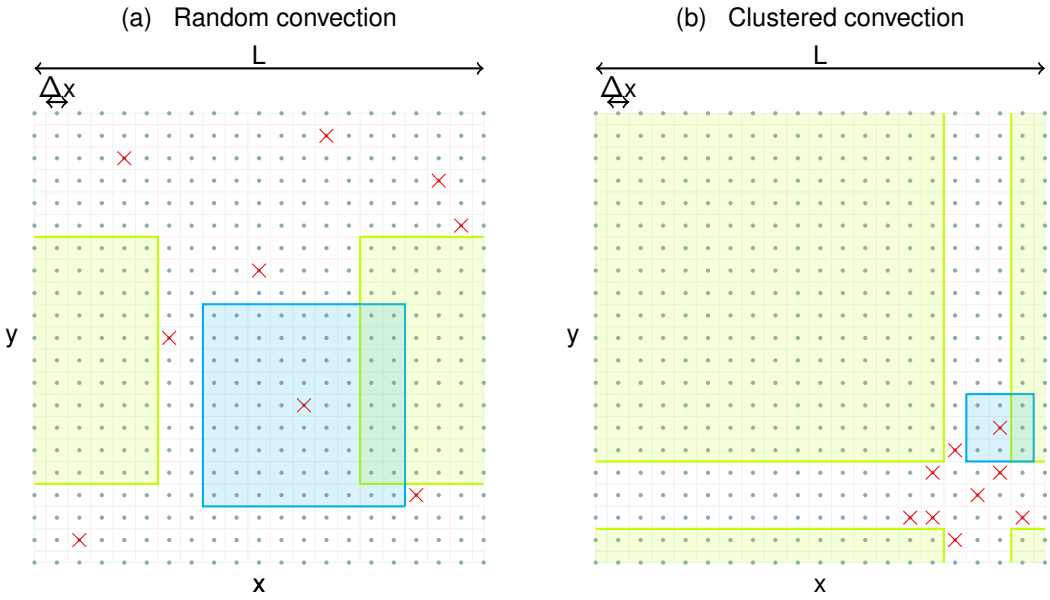

Figure 8. Sketch of two potential metrics of convective spacing relevant for aggregation onset, namely the size of the largest clear-sky, convective-free box ($d_{\mathrm{max,clr}}$, yellow boxes) and the maximum inter-convective spacing ($d_{\mathrm{max,nn}}$, blue boxes), in a random convective situation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

maximum box surrounding any tower that is free from further events:

$$\bar{d}_{\mathrm{max,nn}} = \sum_{i=1}^{\frac{L}{\Delta x}} i\Delta x \left(\left(1 - \left(1 - \frac{i^2 \Delta x^2}{L^2} \right)^{N_c - 1} \right)^{N_c} - \left(1 - \left(1 - \frac{(i-1)^2 \Delta x^2}{L^2} \right)^{N_c - 1} \right)^{N_c} \right). \tag{12}$$

In the supplementary material, we present the result of a 70000 artificially generated random convective scenes with varying N_c in order to show that the theoretical estimate for $\bar{d}_{\mathrm{max,nn}}$ presented in eqn. (12) fits the numerical data perfectly (Fig. S5).

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

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

However, this analytical formula somewhat over-estimates 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 (11) 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 (13). 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 (Fig. 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, $\bar{d}_{\mathrm{max,nn}}$ will be referred to as $\bar{d}$ for brevity.

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:

$$\gamma = f(a_d) \frac{K \tau_{\mathrm{sub}}}{L^2} \frac{L}{\bar{d}} = f(a_d) \frac{K \tau_{\mathrm{sub}}}{L \bar{d}}, \tag{14}$$

where $\bar{d}$ is given by eqn. (12) and is normalized by the domain size L.

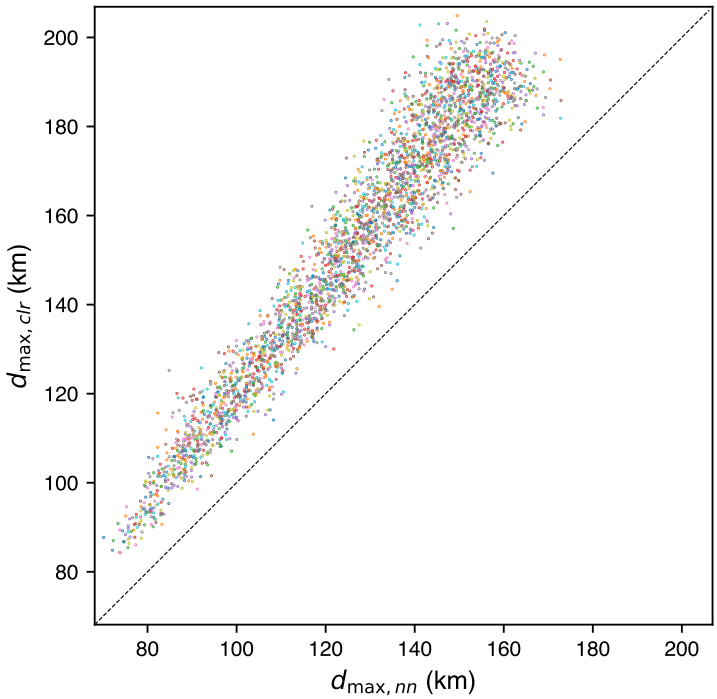

Figure 9. Scatter plot of daily averaged diagnostics for $d_{\mathrm{max,nn}}$ versus $d_{\mathrm{max,clr}}$ from hourly snapshots of scenes taken from a large ensemble in the first day of each experiment when convection is still randomly distributed. The identity line is shown as a black dashed line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

In addition to the four factors of domain size, resolution, diffusivity 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, which in principle can be either linear or non-linear, due to a_d being dimensionless. As a_d is dimensionless, the functional form of f will be derived empirically using an ensemble of numerical experiments.

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, $\Delta x = 2$ km), we evaluate the ensemble experiments that vary K and τ_{sub} to see if the dimensionless quantity correctly predicts the final state to be clustered or random. Fig. 10 shows contours of $\bar{\sigma}_{R,20}$, which we recall is the 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 $\bar{\sigma}_{R,20}$) and those with random convection (low $\bar{\sigma}_{R,20}$) 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 $\frac{K\tau_{\text{sub}}}{Ld}$ (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, $\Delta x = 2$ km (Figs. S6 and S7 in the supporting information). 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.

4.5 The role of the parameter a_d

Intuitively, the relationship eqn. (4) 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

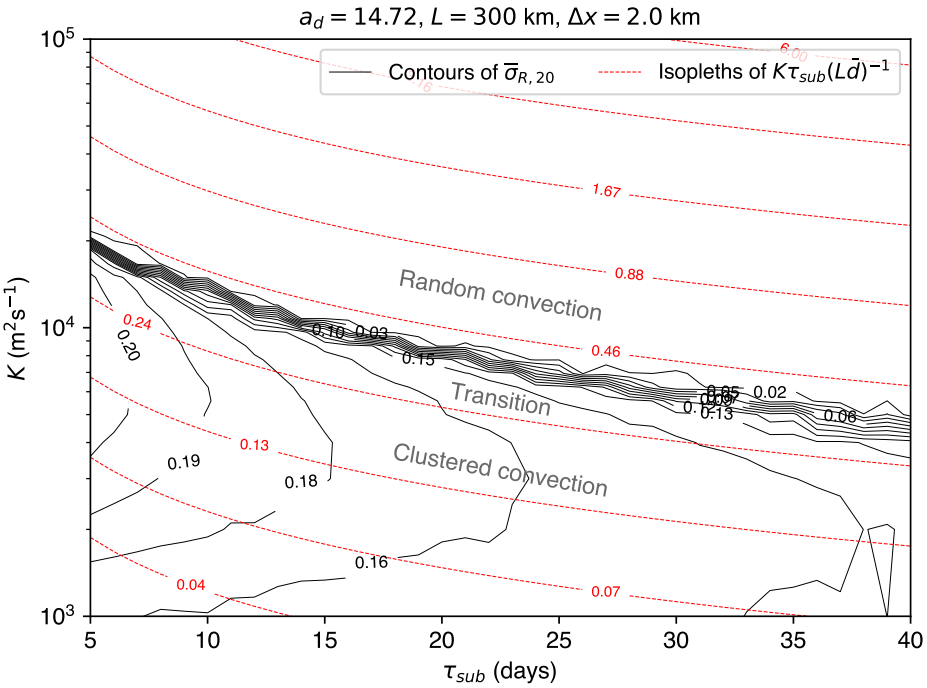

Figure 10. Contours of $\bar{\sigma}_{R,20}$ (black solid curves) along with the isopleths of $K\tau_{\rm sub}(L\bar{d}^{-1})$ for single realizations of the system in a large set of simulations with different K and $\tau_{\rm sub}$ and fixed a_d, L and Δx.

$a_d \simeq 0$, eqn. (4) is homogeneous in the interval $[R_{\min}, R_{\max}]$ 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 $(\tau_{\rm sub}, K)$ space (cf. Fig. 10).

The functional dependence of the transition on a_d is determined empirically (Fig. 11), using $\bar{\sigma}_{R,20}$ for simulations performed with a range of values of K and a_d and fixed $\tau_{\rm sub}$, 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 $\bar{\sigma}_{R,20}$ for the simulations with aggregated convection are due to the absence of monotonicity of $\bar{\sigma}_{R,20}$ 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 eqn. (4), the larger a_d gets, the more likely is for convection to reactivate at the same spots (which are the moistest ones), thus enlarging the area occupied by subsiding air, shrinking the wet moist patch, and therefore reducing spatial R variance beyond the onset point of aggregation. This nonlinear behavior of $\bar{\sigma}_{R,20}$ in the clustered state seen in both Figs. 10 and 11, with the spatial variance of humidity increasing sharply with aggregation onset but reducing as the degree of aggregation strengthens, implies that it is possibly not an effective metric of aggregation for model inter-comparison.

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}. This is:

$$N_{ag} = \frac{K\tau_{\rm sub}}{a_d^2 L \bar{d}}. \tag{15}$$

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, $\tau_{\rm sub}$, a_d and the domain size L and resolution Δx. The resulting scatter plot in Fig. 12 shows that the dimensionless parameter N_{ag} as specified in eqn. (15) predicts the transition from random to aggregated states when the combination of these five parameters gives a N_{ag} value below a critical threshold, $N_{ag,\rm crit}$, of approximately 1.72×10^{-3}. This estimate (that is, the vertical line in Fig. 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 $\bar{\sigma}_{R,20} = 0.05$ was imposed to distinguish between aggregated and

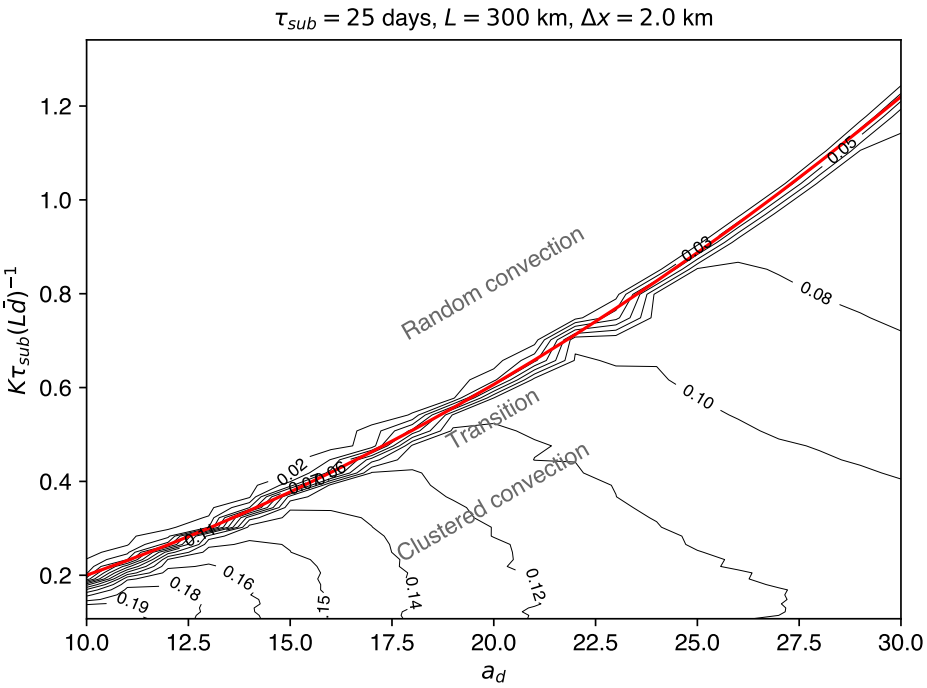

Figure 11. Contours of $\bar{\sigma}_{R,20}$ for an ensemble of runs carried out with different values of $K\tau_{\mathrm{sub}}\left(L\bar{d}\right)^{-1}$ (here obtained varying K and keeping τ_{sub}, L and Δx fixed) and a_d. The red solid line represents the polynomial (quadratic) empirical fit for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

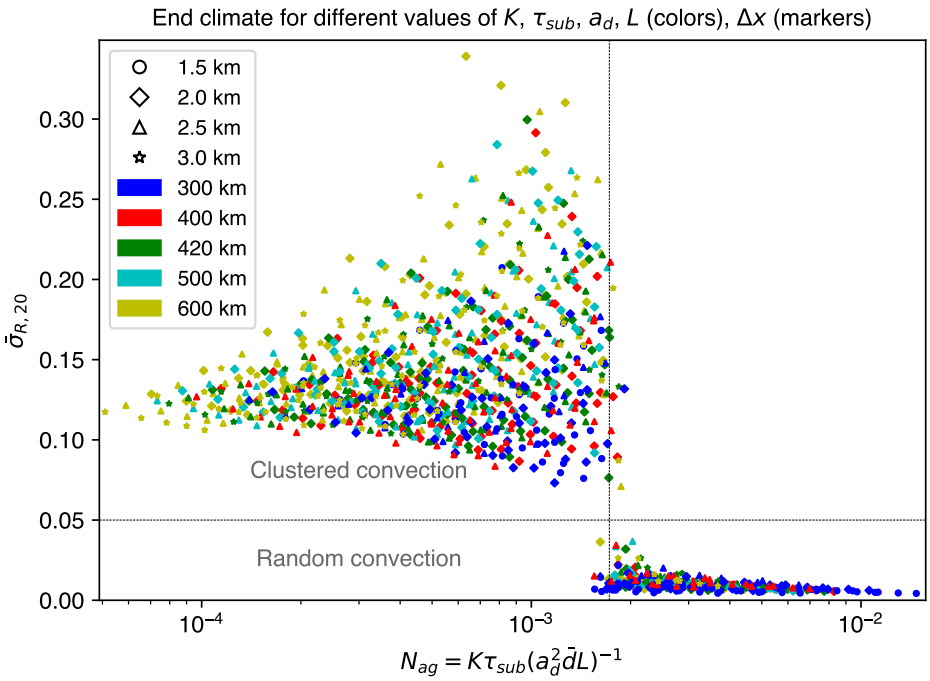

Figure 12. Scatter plot referring to a collection of simulations with different values of the parameter N_{ag}, as defined in (15), each associated with the corresponding value of $\bar{\sigma}_{R,20}$. The horizontal dashed line $\bar{\sigma}_{R,20} = 0.05$ separates aggregated and non-aggregated runs, the vertical dashed line represents the threshold value of N_{ag} obtained as specified in the text.

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 $N_{ag} \sim N_{ag,\mathrm{crit}}$, with small initial random perturbations, showed that these could end up in either a random or aggregated state.

5 Discussion and conclusions

Simulations of radiative-convective equilibrium that are run on $\mathcal{O}(1000$ 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, termed self-aggregation as it spontaneously occurs due to local diabatic feedbacks, despite homogeneous 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.

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 minutes. This local moistening is spread laterally by a local diffusive 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 strength of the diffusion, 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, i.e., 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 resolution. We argue that the diffusion 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.

Using these arguments and fits from experimental data, we were able to combine the domain size, resolution, diffusion, 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 cloud resolving model 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. Thus, ideally, the next step in this work is to devise a methodology to take consecutive cloud resolving model outputs, and using the autocorrelation of water vapor field and the location selected for new convective events to derive estimates for the three parameters of the simple model. If this can be achieved, a short simulation of a CRM in a given con-

figuration may suffice to predict whether it will ultimately aggregate using the aggregation number N_{ag}, given in eqn. (15). 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 radiative-convective equilibrium 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 cloud resolving model 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 possibly required, which will be a topic of future work.

Acknowledgments

GB is supported by a MIUR/University of Trieste PhD scholarship. 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.

References

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. doi: 10.1002/qj.3468

Böing, S. J. (2016, December). An object-based model for convective cold pool dynamics. *Mathematics of Climate and Weather Forecasting*, *2*(1). doi: 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. *J. Atmos. Sci.*, *62*, 4273–4292.

Bretherton, C. S., Peters, M. E., & Back, L. E. (2004). Relationships between water vapor path and precipitation over the tropical oceans. *J. Climate*, *17*(7), 1517–1528.

Bretherton, C. S., & Smolarkiewicz, P. K. (1989). Gravity waves, compensating subsidence and detrainment around cumulus clouds. *J. Atmos. Sci.*, *46*, 740–759.

Cohen, B. G., & Craig, G. C. (2004). The response time of a convective cloud ensemble to a change in forcing. *Q. J. R. Meteorol. Soc.*, *130*, 933–944.

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

Cronin, T. W., & Emanuel, K. A. (2013). The climate time scale in the approach to radiative-convective equilibrium. *J. Adv. Mod. Earth Sys.*, *5*(4), 843–849.

Emanuel, K. A., Wing, A. A., & Vincent, E. M. (2014). Radiative-convective instability. *J. Adv. Mod. Earth Sys.*, *6*(1), 75–90.

Grabowski, W. W., & Moncrieff, M. W. (2004). Moisture-convection feedback in the Tropics. *Q. J. R. Meteorol. Soc.*, *130*, 3081–3104.

Haerter, J. O. (2019). Convective Self-Aggregation As a Cold Pool-Driven Critical Phenomenon. *Geophysical Research Letters*, *46*(7), 4017–4028. doi: 10.1029/2018GL081817

Held, I. M., Hemler, R. S., & Ramaswamy, V. (1993, November). Radiative-Convective Equilibrium with Explicit Two-Dimensional Moist Convection. *Journal of the Atmospheric Sciences*, *50*(23), 3909–3927. doi: 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. *J. Adv. Mod. Earth Sys.*, *8*, 1468–1482.

Holloway, C. E., & Neelin, J. D. (2010). Temporal relations of column water vapor and tropical precipitation. *J. Atmos. Sci.*, *67*(4), 1091–1105.

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. doi: 10.1002/2015MS000511

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. doi: 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*. Wiley.

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

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. doi: 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.

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

Muller, C. J., & Held, I. M. (2012, April). Detailed Investigation of the Self-Aggregation of Convection in Cloud-Resolving Simulations. *Journal of the Atmospheric Sciences*, *69*(8), 2551–2565. doi: 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. doi: 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. doi: 10.1029/2019MS001672

Peaceman, D. W., & Rachford, J., H. H. (1955). The Numerical Solution of

Parabolic and Elliptic Differential Equations. *Journal of the Society for Industrial and Applied Mathematics*, *3*(1), 28–41. doi: 10.1137/0103003

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

Randall, D. A., & Huffman, G. J. (1980). A stochastic model of cumulus clumping. *J. Atmos. Sci.*, *37*, 2068–2078.

Raymond, D. J., & Zeng, X. (2000). Instability and large-scale circulations in a two-column model of the tropical troposphere. *Q. J. R. Meteorol. Soc.*, *126*, 3117–3135.

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. doi: 10.1002/2017GL076296

Shamekh, S., Muller, C., Duvel, J.-P., & D'Andrea, F. (2020, October). How Do Ocean Warm Anomalies Favor the Aggregation of Deep Convective Clouds? *Journal of the Atmospheric Sciences*, *77*(11), 3733–3745. doi: 10.1175/JAS-D-18-0369.1

Sobel, A. H., Bellon, G., & Bacmeister, J. (2007). Multiple equilibria in a single-column model of the tropical atmosphere. *Geophys. Res. Lett.*, *34*(22).

Sobel, A. H., Nilsson, J., & Polvani, L. M. (2001). The weak temperature gradient approximation and balanced tropical moisture waves. *J. Atmos. Sci.*, *58*(23), 3650–3665.

Stephens, G. L., van den Heever, S., & Pakula, L. (2008, December). Radiative–Convective Feedbacks in Idealized States of Radiative–Convective Equilibrium. *Journal of the Atmospheric Sciences*, *65*(12), 3899–3916. doi: 10.1175/2008JAS2524.1

Stoyan, D., Kendalla, W. S., & Mecke, J. (1987). *Stochastic Geometry and Its Applications.* John Wiley,New York.

Strang, G. (1968, September). On the Construction and Comparison of Difference Schemes. *SIAM Journal on Numerical Analysis*, *5*(3), 506–517. doi: 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. doi: 10.1029/2019GL084491

Tompkins, A. M. (2000). The impact of dimensionality on long-term cloud resolving model simulations. *Mon. Wea. Rev.*, *128*, 1521–1535.

Tompkins, A. M. (2001). Organization of tropical convection in low vertical wind shears: The role of water vapor. *J. Atmos. Sci.*, *58*, 529–545.

Tompkins, A. M., & Craig, G. C. (1998a). Radiative-convective equilibrium in a three-dimensional cloud ensemble model. *Q. J. R. Meteorol. Soc.*, *124*, 2073–2097.

Tompkins, A. M., & Craig, G. C. (1998b). Time-scales of adjustment to radiative-convective equilibrium in the tropical atmosphere. *Q. J. R. Meteorol. Soc.*, *124*, 2693–2713.

Tompkins, A. M., & Semie, A. G. (2017). Organization of tropical convection in low vertical wind shears: Role of updraft entrainment. *J. Adv. Mod. Earth Sys.*, http://dx.doi.org/10.1002/2016MS000802. doi: 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*, *n/a*(n/a), e2020MS002186. doi: 10.1029/2020MS002186

Weger, R. C., Lee, J., Zhu, T., & Welch, R. M. (1992). Clustering, randomness and regularity in cloud fields: 1. Theoretical considerations. *J. Geophys. Res.*, *97*, 20519–20536.

Windmiller, J. M., & Craig, G. C. (2019, April). Universality in the Spatial Evolution of Self-Aggregation of Tropical Convection. *Journal of the Atmospheric Sciences*, *76*(6), 1677–1696. doi: 10.1175/JAS-D-18-0129.1

Wing, A. A., & Cronin, T. W. (2016). Self-aggregation of convection in long channel geometry. *Q. J. R. Meteorol. Soc.*, *142*(694), 1–15, doi 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. *J. Adv. Mod.*

Earth Sys., *5*, 1–14, doi:10.1002/2013MS000269.

Wing, A. A., Stauffer, C. L., Becker, T., Reed, K. A., Ahn, M.-S., Arnold, N. P., ... Zhao, M. (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. doi: 10.1029/2020MS002138

Yang, D. (2021, February). A Shallow-Water Model for Convective Self-Aggregation. *Journal of the Atmospheric Sciences*, *78*(2), 571–582. doi: 10.1175/JAS-D-20-0031.1

Yano, J.-I., & Manzato, A. (2022, March). Does More Moisture in the Atmosphere Lead to More Intense Rains? *Journal of the Atmospheric Sciences*, *79*(3), 663–681. doi: 10.1175/JAS-D-21-0117.1

Zhang, C., Mapes, B. E., & Soden, B. J. (2003). Bimodality in tropical water vapour. *Q. J. R. Meteorol. Soc.*, *129*(594), 2847–2866.

Supporting Information for "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

Contents of this file

1. Figures S1 to S7

Introduction This file contains an overview of the numerical procedure designed to solve the diffusion-reaction prognostic equation for R, eqn. (3) in the paper. A full description of idealized tests is also included and the corresponding results are shown in Figs. S1-S3.

An example of time evolution of diagnostics for the two metrics $d_{\mathrm{max,clr}}$ and $d_{\mathrm{max,nn}}$ introduced in the main manuscript, calculated directly from the numerical routine, is then reported in Fig. S4 for the same set of simulations presented in Fig. 3. In particular, the theoretical model (12), which is used in the definition of the dimensionless parameter (15), is seen to exactly match the maximum inter-convective nearest neighbour distance to be expected when randomly throwing points onto a bounded region of the plane (Fig. S5).

Additional figures (Figs. S6-S7) are finally to be compared with Fig. 10 and motivate the reliability of the quantity γ, defined in the main manuscript (eqn. 14), in capturing the transition between aggregated and non-aggregated states in the (τ_{sub}, K) space independent of the domain size (Fig. S6) or the horizontal grid spacing (Fig. S7). In these cases the a_d-dependency is not considered, as a_d is kept at its default value, $a_d = 14.72$, but apparently the critical isopleth of $K\tau_{\mathrm{sub}}(L\bar{d})^{-1}$ that fits the transition regime is the same across these ensembles of simulations.

1. Numerics

The numerical solution of the governing equation (3) uses second-order finite differences in space and a Strang-type operator splitting scheme in time (Strang, 1968).

Splitting techniques are commonly advocated in meteorological applications with multiple time scales (e.g., Beljaars et al., 2018) and, in general, when a differential problem involves many physical processes. Commonly, it is not possibly or numerically efficient to attempt the integration of the equations by means of a single solver method. A computationally sustainable alternative is thus offered by decomposing the system into sub-groups of processes and using different suitable and advantageous methods for each group (e.g., Hundsdorfer & Verwer, 2007), with tendencies from each treated sequentially in time. In detail, given a generic, scalar partial differential equation

$$\frac{\partial u}{\partial t} = f(u),$$

where f can be regarded as a spatial partial differential operator, a two-term decomposition of the RHS is considered:

$$f(u) = f_1(u) + f_2(u).$$

The Strang splitting approach is to perform half a time step with the operator f_2, followed by a full time step with f_1 and another half step with f_2 (or vice versa), with the tendency from the previous process added to provide the initial value of the subsequent process(es). In formulae,

$$\begin{aligned}
&\frac{\partial}{\partial t}u^* = f_2(u^*), \quad t_n < t \le t_{n+\frac{1}{2}}, \quad u^*(t_n) = u_n, \\
&\frac{\partial}{\partial t}u^{**} = f_1(u^{**}), \quad t_n < t \le t_{n+1}, \quad u^{**}(t_n) = u^*\left(t_{n+\frac{1}{2}}\right), \\
&\frac{\partial}{\partial t}u^{***} = f_2(u^{***}), \; t_{n+\frac{1}{2}} < t \le t_{n+1}, \; u^{***}\left(t_{n+\frac{1}{2}}\right) = u^{**}(t_{n+1}),
\end{aligned}$$

the subscript n refers to the temporal discretization, and the overall solution is given by $u_{n+1} = u^{***}(t_{n+1})$. This sequential procedure normally introduces an error at each integration step. The Strang splitting is second-order accurate for sufficiently smooth solutions (e.g., LeVeque, 2007), provided that each subproblem is treated with a method of such accuracy at least.

In the stochastic model presented here, we separate the diffusion and subsidence components of eqn. (3) from the convective source term, i.e.,

$$f_1(R) = K\left(\delta_x^2 + \delta_y^2\right)R - \frac{R}{\tau_{\text{sub}}}, \;\; f_2(R) = \frac{(R_c - R)}{\tau_c}\mathcal{H}(p_c(R) - X),$$

as this will not produce any splitting error in the non-convective grid points, where the second operator vanishes. The difference operator $\delta_\cdot^2$ is the second-order centered difference approximation of the second derivative. For the problem involving f_2, the analytical solution is derived, in order to reduce as much as possible any integration error, related to the application of numerical methods, which can be incurred in addition to the splitting error.

The diffusion-reaction semidiscrete problem

$$\frac{\partial R}{\partial t} = K\left(\delta_x^2 + \delta_y^2\right) R - \frac{R}{\tau_{\text{sub}}} \tag{1}$$

is solved by means of a properly modified version of the classical Peaceman-Rachford Alternating Direction Implicit (ADI) method (Peaceman & Rachford, 1955). The scheme consists of splitting 2D problems into two separate steps, treating implicitly only one spatial operator at a time and therefore performing line-by-line solution of smaller, generally structured, independent sets of equations. It can be regarded as a perturbed formulation of the Crank-Nicholson scheme, whose application to eqn. (1) would yield

$$\left(1 - \beta\delta_x^2 - \beta\delta_y^2 + \omega\right) R_{j,k}^{n+1} = \left(1 + \beta\delta_x^2 + \beta\delta_y^2 - \omega\right) R_{j,k}^{n}, \tag{2}$$

where the superscript n indicates discrete time steps, the subscripts j, k refer to the horizontal square grid, $\beta = K\dfrac{\Delta t}{2\Delta x^2}$, $\omega = \dfrac{\Delta t}{2\tau_{\text{sub}}}$, Δt denoting the time step and Δx the horizontal spacing. Eqn. (2) can be factorized and rearranged as

$$\left(1 - \beta\delta_x^2 + \omega\right)\left(1 - \beta\delta_y^2\right) R_{j,k}^{n+1} = \left(1 + \beta\delta_x^2 - \omega\right)\left(1 + \beta\delta_y^2\right) R_{j,k}^{n} + \left(\beta^2\delta_x^2\delta_y^2 - \omega\beta\delta_y^2\right)\left(R_{j,k}^{n+1} - R_{j,k}^{n}\right),$$

where the last term on the RHS can be proved to be $\mathcal{O}(\Delta t^3)$ and is therefore negligible for small Δt. Thus, the application of the ADI method implies the sequential solution of the systems

$$\left(1 - \beta\delta_x^2 + \omega\right) R_{j,k}^{n+\frac{1}{2}} = \left(1 + \beta\delta_y^2\right) R_{j,k}^{n}, \tag{3}$$

$$\left(1 - \beta\delta_y^2\right) R_{j,k}^{n+1} = \left(1 + \beta\delta_x^2 - \omega\right) R_{j,k}^{n+\frac{1}{2}}, \tag{4}$$

and the two-step scheme (3)-(4) can be shown to be unconditionally stable and second order in both space and time, hence convergent.

If periodic boundary conditions are assigned, the linear systems resulting from the discretizations turn out to be circulant tridiagonal and can be easily solved by using Fast

Fourier Transform algorithms (e.g., Cooley & Tukey, 1965). Note that, if the splitting procedure had not been adopted, the matrices would not have been circulant or constant with time owing to the triggering of convective events, whose corresponding terms contribute to the entries of the main diagonal.

The convergence properties of the new ADI solver are assessed against the assumption of initial top hat or Gaussian profiles. The results of convergence tests under time step and grid refinements are shown in Figs. S1 and S2, respectively. In the first case (Fig. S1), the problem in eqn. (1) is considered, with $K = 10^4$ m^{2}s^{-1}, $\tau_{\rm sub} = 10$ days, over a domain with size $L = 300$ km and spacing $\Delta x = 2$ km, and initial condition given by

$$R_0(\mathbf{x}) = R(\mathbf{x}, 0) = \begin{cases} 1 & \text{for } \mathbf{x} \in [x_1, x_2] \times [y_1, y_2] \\ 0.8 & \text{elsewhere} \end{cases}, \tag{5}$$

where $x_1, y_1 = 140$ km and $x_2, y_2 = 160$ km. The top-hat configuration is a good test bench due to presence of very sharp discontinuities. It is well known that, if implicit discretization is adopted, there is no stability constraint on the time step, as it is for explicit numerical solver, e.g., explicit Euler, for which the following relationship (diffusive stability criterion) is to be satisfied to ensure stability:

$$\Delta t \leq \frac{\Delta x^2}{4K}. \tag{6}$$

Nevertheless, for implicit schemes, a limit of the type (6) still serves as a measure of accuracy (e.g., Ferziger et al., 2002), hence, labelling $\xi = 4K\dfrac{\Delta t}{\Delta x^2}$, we impose several time steps corresponding to a range of values of ξ. Convergence is apparent in the time slices of Fig. S1 (dashed and dash-dotted lines), even though the approximation for $\xi = 6$ exhibits a spurious oscillation at beginning, which is then damped and rapidly disappears.

Spatial convergence properties are examined, as also demanded by a resolution sensitivity study mentioned in the paper. The results of a grid refinement analysis performed on the problem eqn. (1) with $K = 5 \times 10^3$ m^{2}s^{-1}, $\tau_{\text{sub}} = 10$ days and Gaussian initial distribution with $\mu = 150$ km and $\sigma = 5$ km are shown in Fig. S2. The horizontal spacing is successively halved ranging from $\Delta x = 2$ km to 250 m, and the time step Δt is such that $\xi = 0.25$ in all cases. At $t = 600$ s (dashed lines), the profiles are almost insensitive to the resolution, then any error is eventually smoothed down and, at $t = 3600$ s (dash-dotted), the curves are nearly indistinguishable.

In case profiles with sharper discontinuities are prescribed, the method is still able to provide reasonably good approximations, despite exhibiting a more pronounced sensitivity to both the time step size and the resolution.

In the full system, the action of convection is to continuously introduce local delta function perturbations into the R distribution, with sharp gradient at between the convective point and the surrounding grid cells. This is particularly challenging for the numerics and could possibly amplify numerical errors. Further sets of tests are thus conducted, aimed at quantifying the impact of the errors associated with the operator splitting. Fig. S3 charts the results obtained for different values of K, $K = 10^4$ m^{2}s^{-1} (solid lines) and $K = 2.5 \times 10^4$ m^{2}s^{-1} (dashed lines), and ξ (colors), all else being kept fixed ($\tau_{\text{sub}} = 12$ days, $a_d = 14.72$, $L = 300$ km, $\Delta x = 2$ km, $R_0 = 0.8$). Convergence to the same statistically steady solutions is apparent, even though, in the low-diffusion case, the approximation for $\xi = 3$ ($\Delta t = 300$ s) yields some differences in both the R spatial mean and standard deviation final equilibrium state. Interestingly, both the transition to the self-aggregated state and the following evolution do not vary monotonically with Δt, and we attribute this

effect to the large stochastic component present in the modelled system. The time-step dependency almost entirely disappears for higher values of K, as they require the use of a smaller time step (via eqn. 6), which also leads to less severe splitting errors. For the experiments presented in the paper, in general, the time step is chosen so that $\xi < 1$.

References

Beljaars, A., Balsamo, G., Bechtold, P., Bozzo, A., Forbes, R., Hogan, R. J., ... Wedi, N. (2018). The numerics of physical parametrization in the ECMWF model. *frontiers*, *6*, doi:10.3389/feart.2018.00137.

Cooley, J. W., & Tukey, J. W. (1965). An Algorithm for the Machine Calculation of Complex Fourier Series. *Mathematics of Computation*, *19*(90), 297–301. doi: 10.2307/2003354

Ferziger, J. H., Peric, M., & Street, R. L. (2002). *Computational methods for fluid dynamics*. Berlin Springer.

Hundsdorfer, W., & Verwer, J. G. (2007). *Numerical Solution of Time-Dependent Advection-Diffusion-Reaction Equations*. Springer Science & Business Media.

LeVeque, R. J. (2007). *Finite Difference Methods for Ordinary and Partial Differential Equations.* https://epubs.siam.org/doi/book/10.1137/1.9780898717839.

Peaceman, D. W., & Rachford, J., H. H. (1955). The Numerical Solution of Parabolic and Elliptic Differential Equations. *Journal of the Society for Industrial and Applied Mathematics*, *3*(1), 28–41. doi: 10.1137/0103003

Strang, G. (1968, September). On the Construction and Comparison of Difference Schemes. *SIAM Journal on Numerical Analysis*, *5*(3), 506–517. doi: 10.1137/0705041

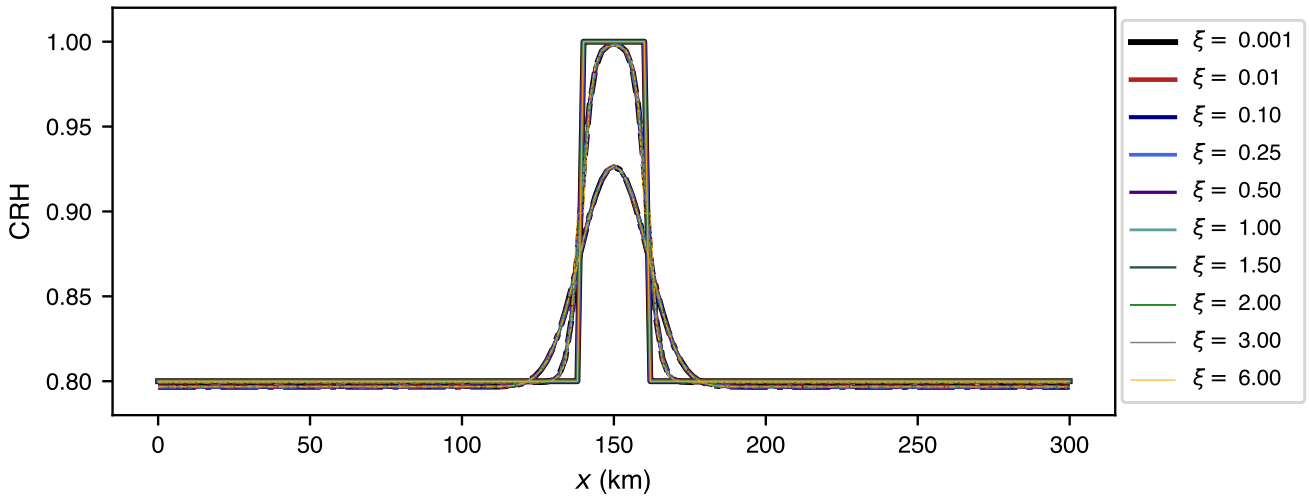

Figure S1. Computed solutions along the section $y = 150$ km for the problem (1) with $K = 10^4$ m^{2}s^{-1}, $\tau_{\text{sub}} = 10$ days, $R_0(\mathbf{x})$ as specified in (5). Shown are the initial profile (solid lines) and the numerical approximations for different time step choices at $t = 600$ s (dashed) and 3600 s (dash-dotted). $\xi = 1$ corresponds to $\Delta t = 100$ s.

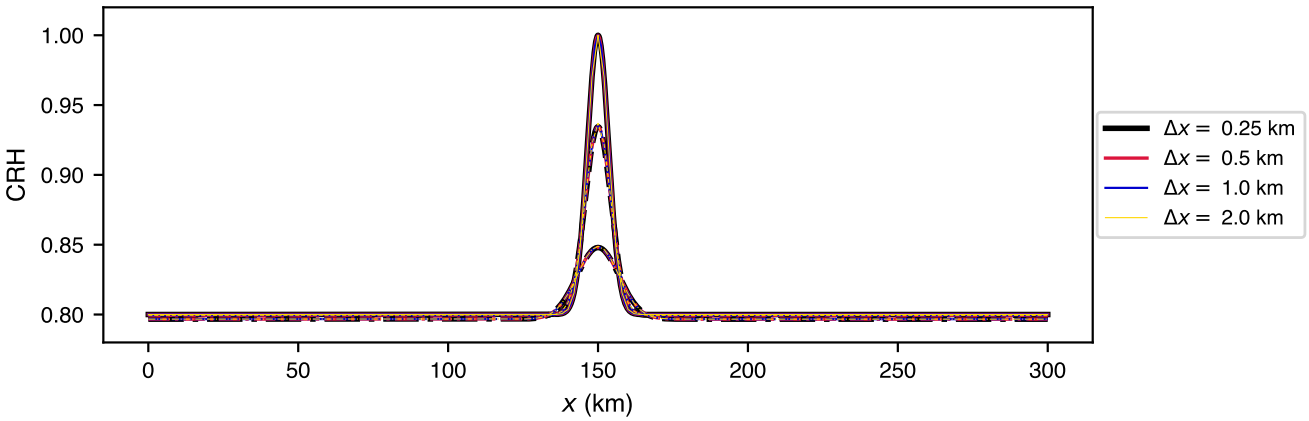

Figure S2. Results of a grid refinement study conducted on the problem (1) with $K = 5 \times 10^3$ m^{2}s^{-1}, $\tau_{\text{sub}} = 10$ days and a Gaussian initialization. Shown are the solutions along the section $y = 150$ km, at times $t = 0$ (solid lines) $t = 600$ s (dashed) and 3600 s (dash-dotted).

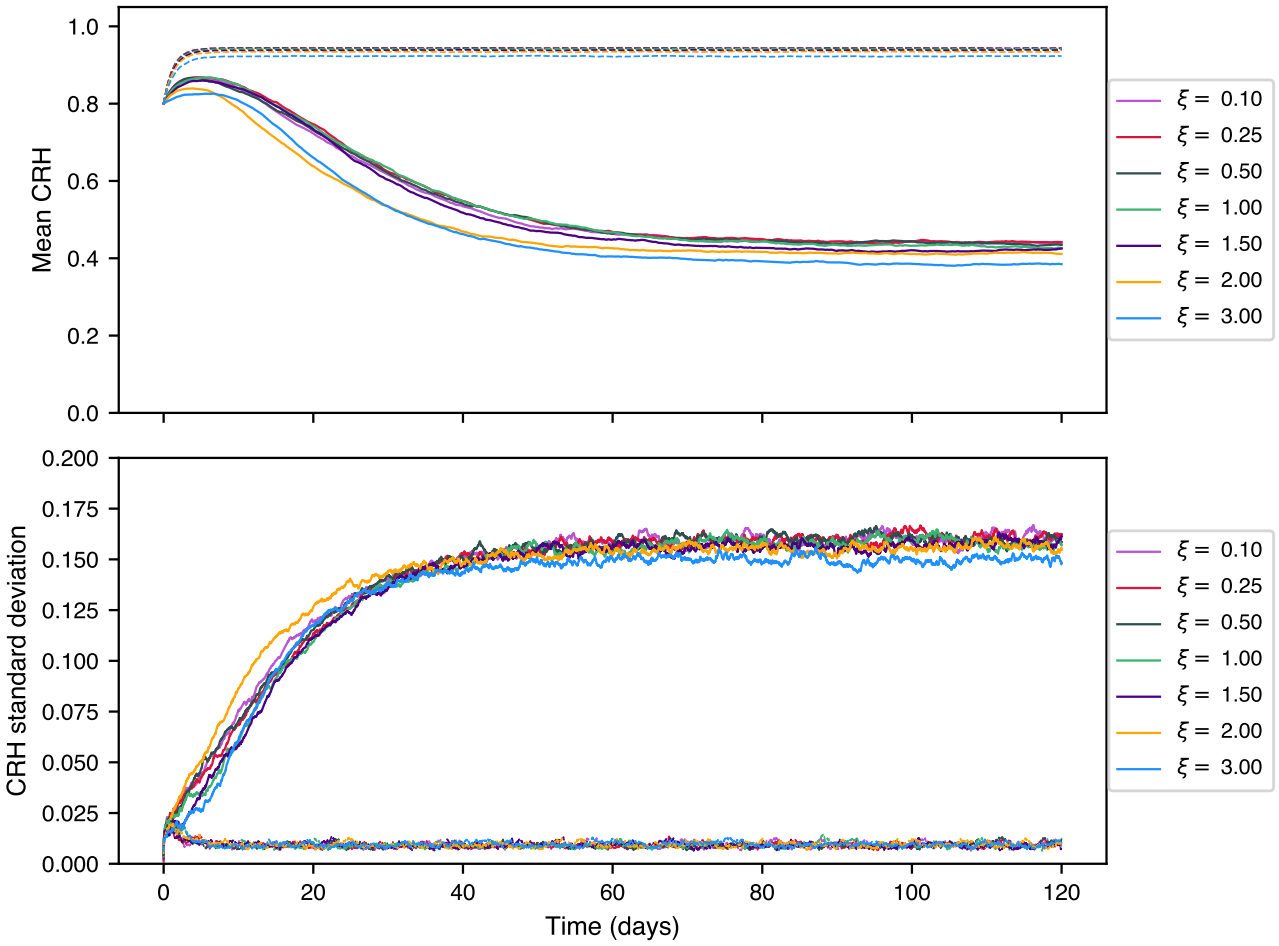

Figure S3. Statistics of the CRH distribution for different runs of the model in terms of Δt in the cases $K = 10^4$ m^{2}s^{-1} (solid lines), 2.5×10^4 m^{2}s^{-1} (dashed lines), with $\tau_{\text{sub}} = 12$ days and $a_d = 14.72$. $\xi = 1$ corresponds to $\Delta t = 100$ s and $\Delta t = 40$ s, respectively.

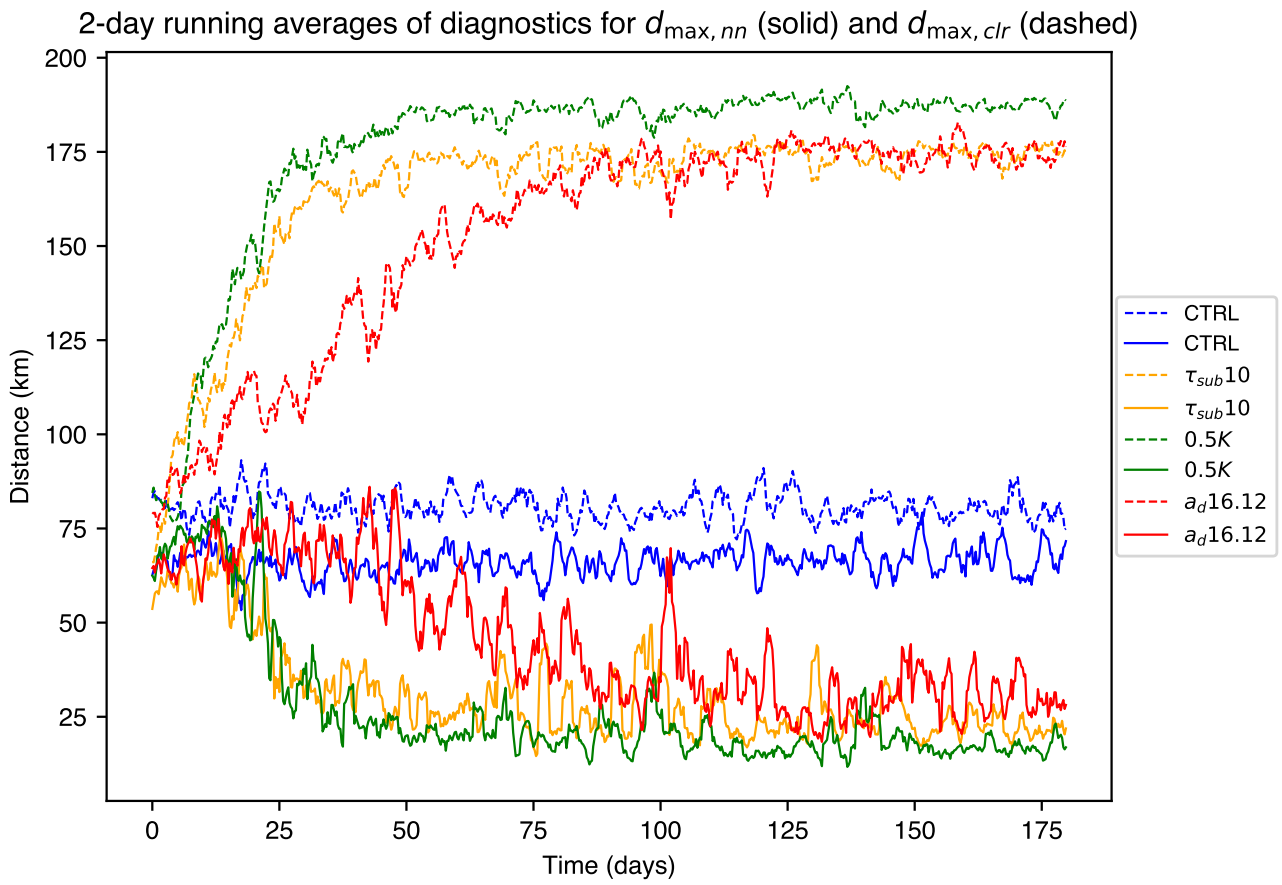

Figure S4. Evolution of the maximum inter-convection nearest neighbour distance (solid lines) and the largest distance from a non-convective to the nearest convective grid cell (dashed), as diagnosed from the numerical model, for the runs of Figure 3. A 2-day running mean has been applied.

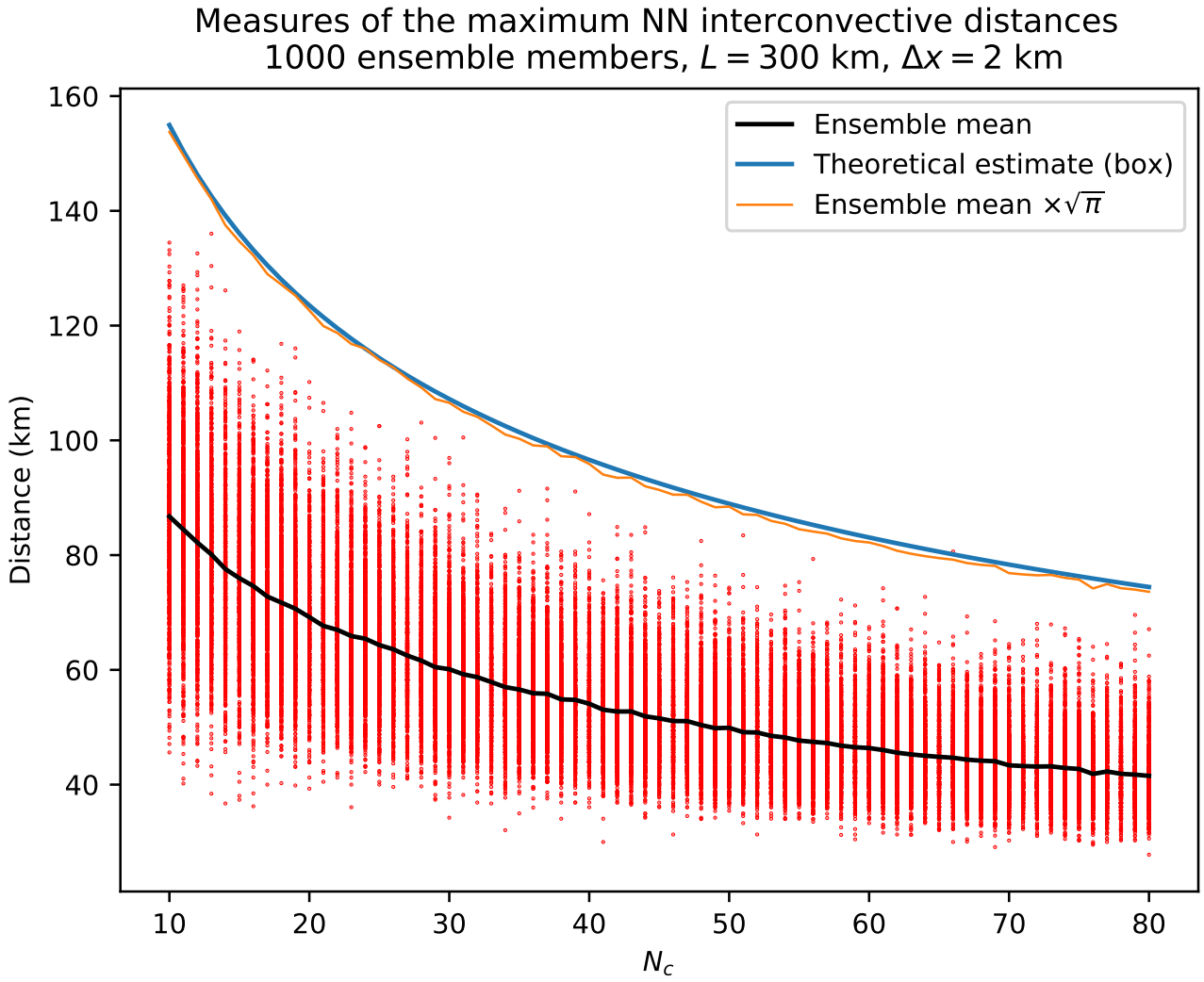

Figure S5. Comparison between the results of a multi-run ensemble of experiments and the theoretical estimate (12). Red dots illustrate the maximum nearest neighbour distances between N_c objects thrown onto a 300 × 300 km domain with 2 km resolution, whereas the black line represents the ensemble mean distance. Examining nearest neighbour distances implies that there are no events within a radius r of the base point, whereas the theory in the paper involves the void probabilities for a square box. The ratio of the area of a circle of radius r to a box of size d is $\sqrt{\pi}$, and we see that multiplying the ensemble mean by this factor (orange line) reproduces exactly the theoretical curve for $\bar{d}_{\mathrm{max,nn}}$ (blue line).

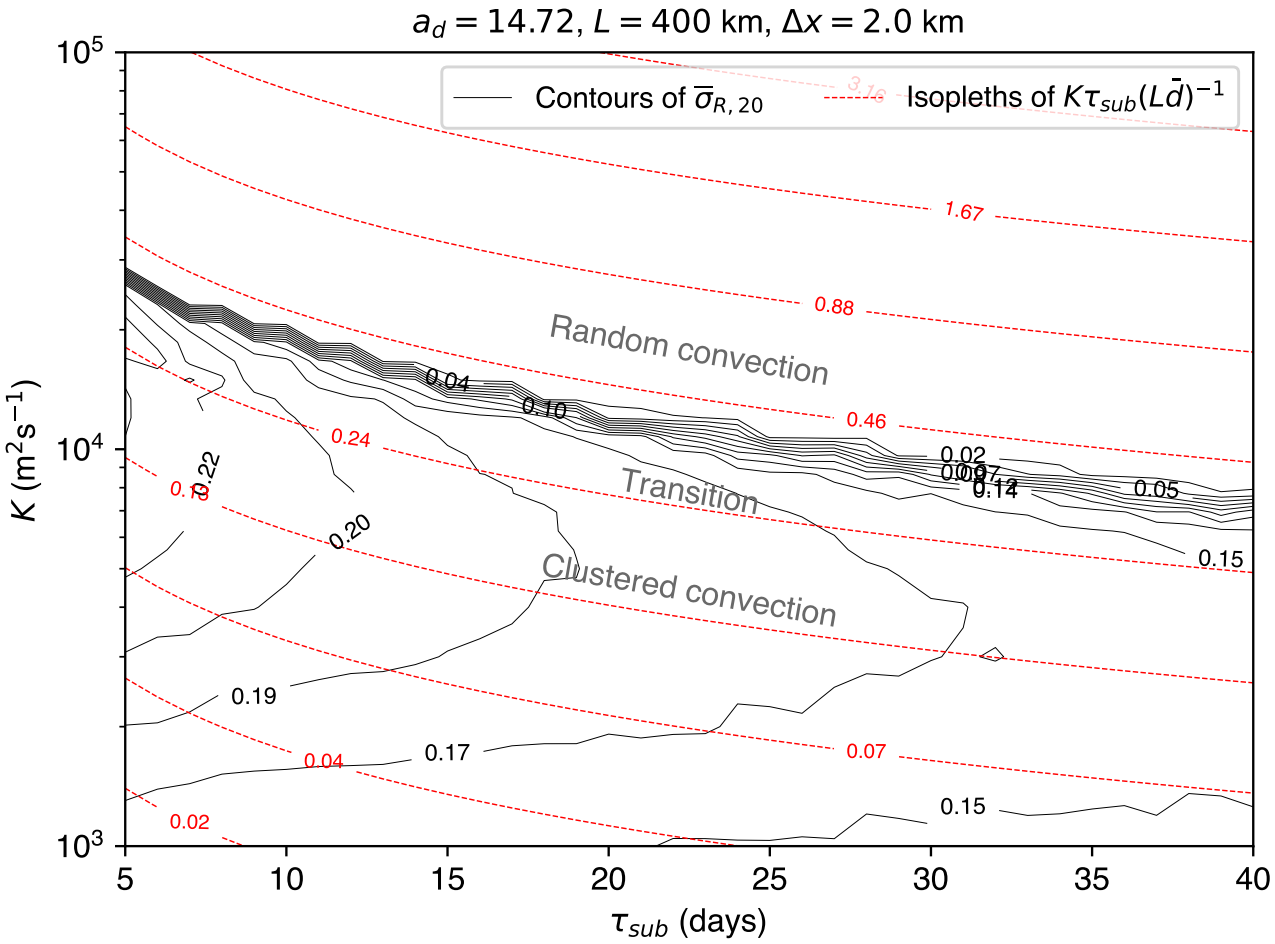

Figure S6. As in Figure 10 but with larger domain, $L = 400$ km.

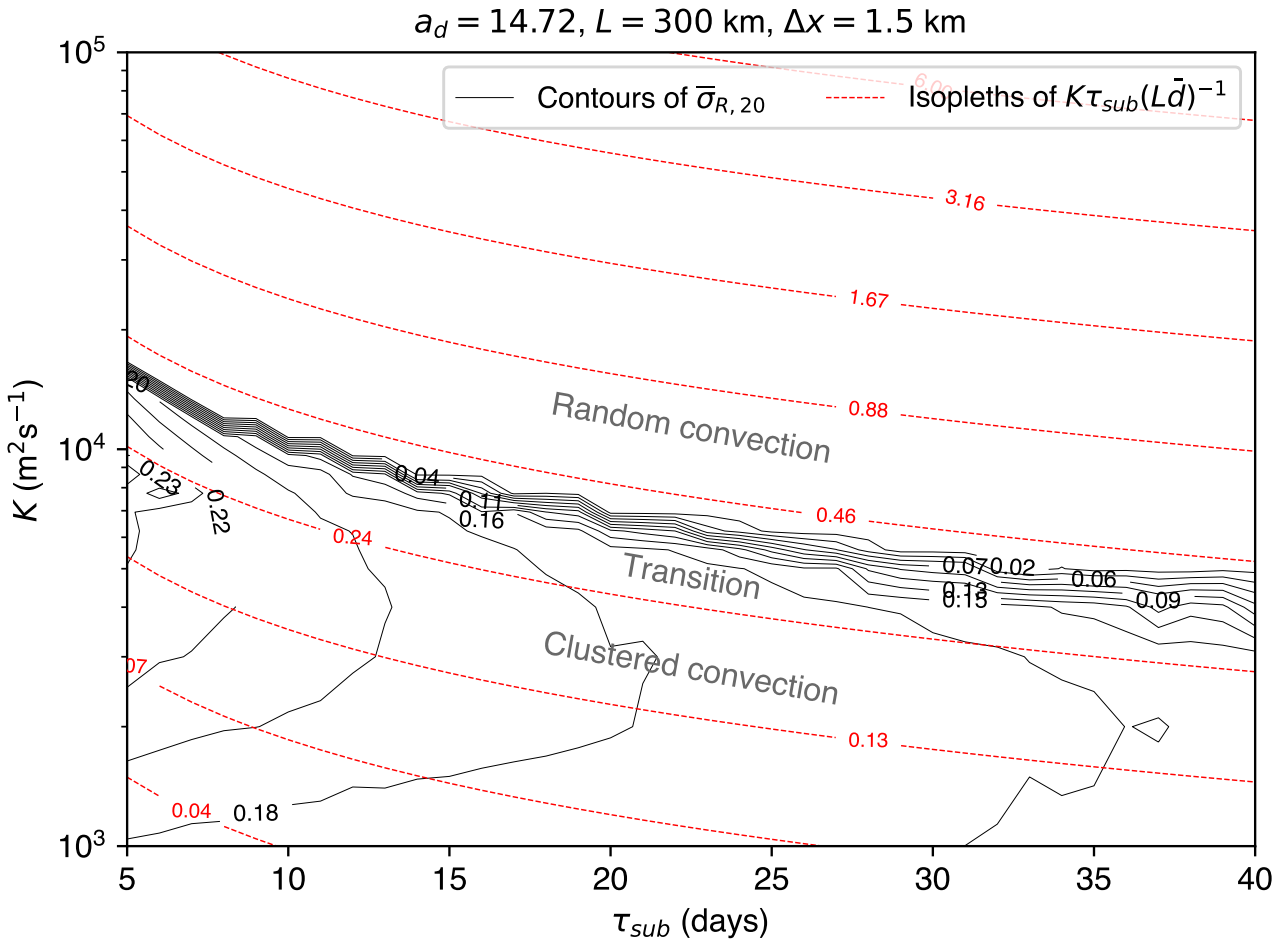

Figure S7. As in Figure 10 but with higher resolution, $\Delta x = 1.5$ km.